

# What shapes genome-wide diversity?

A case study in *Drosophila melanogaster*

Gustavo V. Barroso and Julien Y. Dutheil

September 1st 2021

## Contents

<b>1</b>	<b>Drosophila-like neutral simulations of 2L (True Landscapes)</b>	<b>2</b>
1.1	50 kb scale . . . . .	3
1.1.1	Replicate 1 . . . . .	3
1.1.2	Replicate 2 . . . . .	7
1.1.3	Replicate 3 . . . . .	11
1.1.4	Replicate 4 . . . . .	14
1.1.5	Replicate 5 . . . . .	17
1.1.6	Replicate 6 . . . . .	20
1.1.7	Replicate 7 . . . . .	23
1.1.8	Replicate 8 . . . . .	26
1.1.9	Replicate 9 . . . . .	29
1.1.10	Replicate 10 . . . . .	33
1.1.11	all replicates: . . . . .	36
1.2	200 kb scale . . . . .	40
1.2.1	Replicate 1 . . . . .	40
1.2.2	Replicate 2 . . . . .	43
1.2.3	Replicate 3 . . . . .	46
1.2.4	Replicate 4 . . . . .	49
1.2.5	Replicate 5 . . . . .	52
1.2.6	Replicate 6 . . . . .	55
1.2.7	Replicate 7 . . . . .	58
1.2.8	Replicate 8 . . . . .	62
1.2.9	Replicate 9 . . . . .	65
1.2.10	Replicate 10 . . . . .	68
1.2.11	all replicates: . . . . .	71
1.3	1 Mb scale . . . . .	74
1.3.1	Replicate 1 . . . . .	74
1.3.2	Replicate 2 . . . . .	78
1.3.3	Replicate 3 . . . . .	82
1.3.4	Replicate 4 . . . . .	86
1.3.5	Replicate 5 . . . . .	90
1.3.6	Replicate 6 . . . . .	93
1.3.7	Replicate 7 . . . . .	96
1.3.8	Replicate 8 . . . . .	99
1.3.9	Replicate 9 . . . . .	102
1.3.10	Replicate 10 . . . . .	105
1.3.11	all replicates: . . . . .	108

<b>2</b>	<b>Drosophila-like neutral simulations of chr 2L (Inferred Landscapes)</b>	<b>112</b>
2.1	50 kb scale . . . . .	112
2.1.1	Replicate 1 . . . . .	112
2.1.2	Replicate 2 . . . . .	117
2.1.3	Replicate 3 . . . . .	122
2.1.4	Replicate 4 . . . . .	127
2.1.5	Replicate 5 . . . . .	132
2.1.6	Replicate 6 . . . . .	137
2.1.7	Replicate 7 . . . . .	142
2.1.8	Replicate 8 . . . . .	147
2.1.9	Replicate 9 . . . . .	152
2.1.10	Replicate 10 . . . . .	157
2.1.11	all replicates: . . . . .	162
2.1.12	Plots . . . . .	165
2.2	200 kb scale . . . . .	166
2.2.1	Replicate 1 . . . . .	166
2.2.2	Replicate 2 . . . . .	172
2.2.3	Replicate 3 . . . . .	177
2.2.4	Replicate 4 . . . . .	182
2.2.5	Replicate 5 . . . . .	187
2.2.6	Replicate 6 . . . . .	192
2.2.7	Replicate 7 . . . . .	197
2.2.8	Replicate 8 . . . . .	202
2.2.9	Replicate 9 . . . . .	207
2.2.10	Replicate 10 . . . . .	212
2.2.11	all replicates: . . . . .	217
2.2.12	Plots . . . . .	220
2.3	1 Mb scale . . . . .	221
2.3.1	Replicate 1 . . . . .	221
2.3.2	Replicate 2 . . . . .	227
2.3.3	Replicate 3 . . . . .	232
2.3.4	Replicate 4 . . . . .	237
2.3.5	Replicate 5 . . . . .	242
2.3.6	Replicate 6 . . . . .	247
2.3.7	Replicate 7 . . . . .	252
2.3.8	Replicate 8 . . . . .	257
2.3.9	Replicate 9 . . . . .	262
2.3.10	Replicate 10 . . . . .	267
2.3.11	all replicates: . . . . .	272
2.3.12	Plots . . . . .	275
<b>3</b>	<b>Real Drosophila data.</b>	<b>276</b>
3.1	50 kb windows . . . . .	279
3.2	200 kb windows . . . . .	294
3.3	1 Mb windows . . . . .	309
3.4	Evolutionary (Protein) Rates . . . . .	324
<b>4</b>	<b>R<sup>2</sup> plot with real, simulated (true) and simulated (inferred) data.</b>	<b>331</b>

## 1 Drosophila-like neutral simulations of 2L (True Landscapes)

Second batch of 10 replicate datasets simulated using parameters as inferred from real Drosophila data.

Here we present results using true (simulated) genomic landscapes.

## 1.1 50 kb scale

```
r2.sim.50kb <- as.data.frame(matrix(ncol = nreps, nrow = 5))
row.names(r2.sim.50kb) <- c("Total", "Theta", "Rho", "TMRCA", "Theta:TMRCA")
colnames(r2.sim.50kb) <- reps

sim.rho.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/dm.sim.rho.50000.txt", header = T)
sim.theta.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/dm.sim.theta.50000.txt", header = T)
```

### 1.1.1 Replicate 1

```
rep_1.pi.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_1/rep_1.pi.50kb.bedgraph", header = T)
rep_1.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_1/sim.tmrca.50k.map", header = T)

sim.lands.50kb <- as.data.frame(cbind(rep_1.pi.50kb$pi, sim.theta.50kb$sim, sim.rho.50kb$sim, rep_1.tmrca.50kb$tmrca))
names(sim.lands.50kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 36665911, p-value = 0.6511
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.01850036

cor.test(~theta+rho, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 37660747, p-value = 0.2592
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.04613476

cor.test(~rho+tmrca, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 35649572, p-value = 0.8119
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.00973136

# centering
sim.lands.50kb$thetaC <- sim.lands.50kb$theta - mean(sim.lands.50kb$theta)
sim.lands.50kb$tmrcaC <- sim.lands.50kb$tmrca - mean(sim.lands.50kb$tmrca)
```

```

sim.lands.50kb$rhoC <- sim.lands.50kb$rho - mean(sim.lands.50kb$rho)

sim.lands.50kb$bin <- 1:nrow(sim.lands.50kb)

# for merging:
sim.lands.50kb.rep1 <- sim.lands.50kb
sim.lands.50kb.rep1$Replicate <- 1

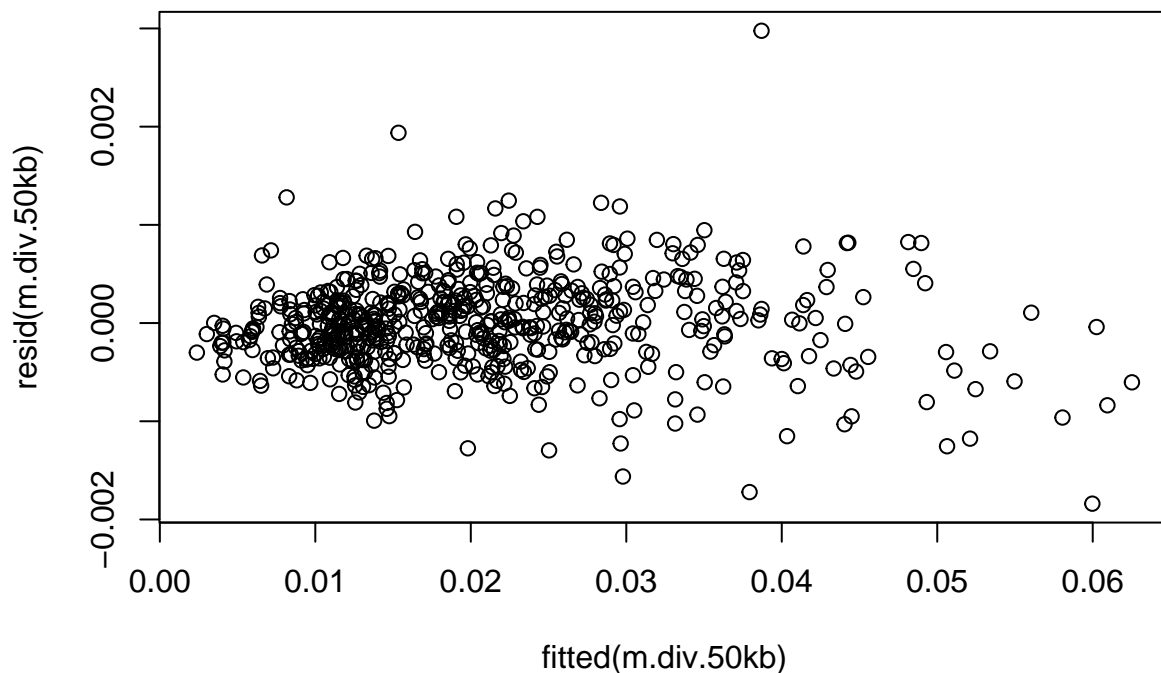
m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.50kb)
m.div.50kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = sim.lands.50kb)
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = sim.lands.50kb)

AIC(m.div.50kb, m.div.50kb.2, m.div.50kb.3)

##           df           AIC
## m.div.50kb      6 -7492.648
## m.div.50kb.2    7 -7491.365
## m.div.50kb.3    8 -7491.767

plot(resid(m.div.50kb)~fitted(m.div.50kb))

```



```

dwtest(m.div.50kb)

##
## Durbin-Watson test
##
## data:  m.div.50kb

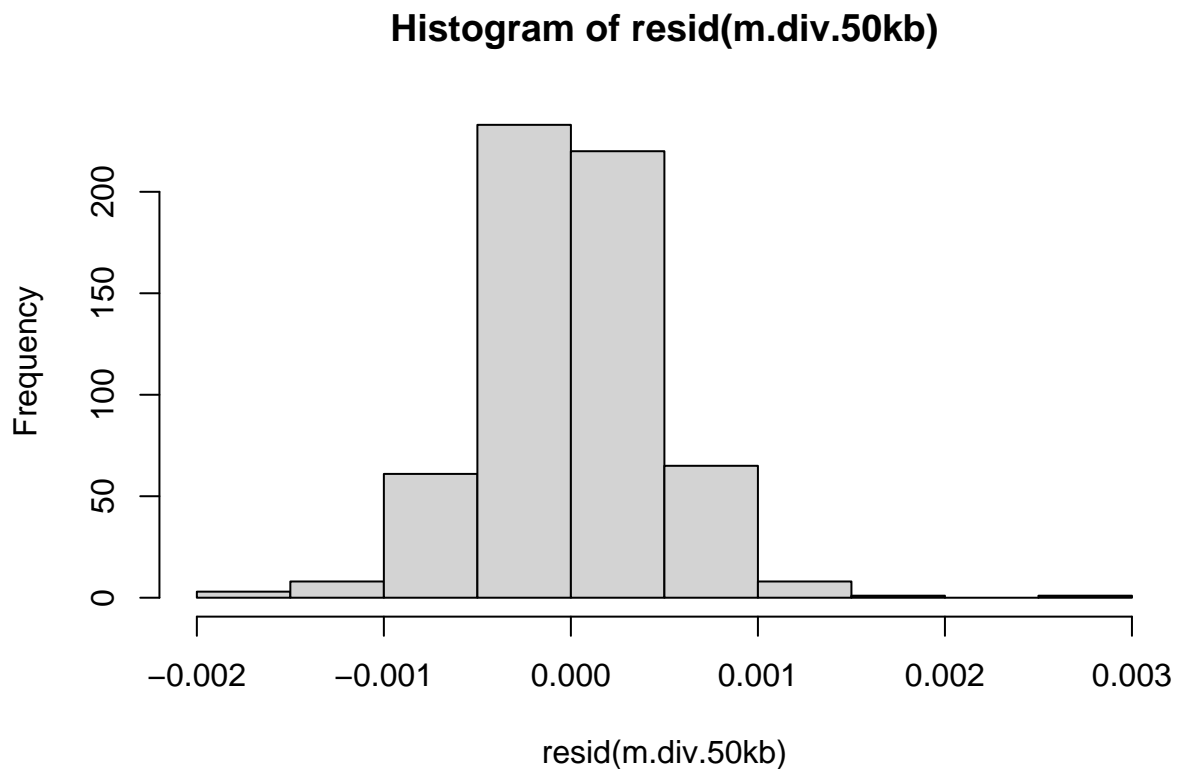
```

```
## DW = 2.1111, p-value = 0.9056
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.50kb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.39214, p-value < 2.2e-16
```

```
hist(resid(m.div.50kb))
```



```
summary(m.div.50kb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.50kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.839e-03 -2.676e-04 -1.006e-05  2.647e-04  2.976e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.069e-02  1.908e-05 1084.381  <2e-16 ***
## thetaC       1.309e+00  2.283e-03  573.279  <2e-16 ***
```

```

## rhoC          1.435e-02  6.455e-03   2.223   0.0266 *
## tmrcaC        2.342e-02  2.686e-04  87.195   <2e-16 ***
## thetaC:tmrcaC 1.483e+00  3.029e-02  48.983   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0004673 on 595 degrees of freedom
## Multiple R-squared:  0.9982, Adjusted R-squared:  0.9982
## F-statistic: 8.37e+04 on 4 and 595 DF,  p-value: < 2.2e-16

anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.50kb[1, 1] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.sim.50kb[2, 1] <- anova.diversity$VarExp[1] * 100
r2.sim.50kb[3, 1] <- anova.diversity$VarExp[2] * 100
r2.sim.50kb[4, 1] <- anova.diversity$VarExp[3] * 100
r2.sim.50kb[5, 1] <- anova.diversity$VarExp[4] * 100

g.div.50kb.1 <- gls(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
                    data = sim.lands.50kb, weights = varPower(0, ~tmrcaC), cor = corAR1(0, ~bin), method = "ML")
g.div.50kb.2 <- gls(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
                    data = sim.lands.50kb, weights = varPower(0, ~theta), cor = corAR1(0, ~bin), method = "ML")
g.div.50kb.3 <- gls(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
                    data = sim.lands.50kb, weights = varPower(0, ~theta), method = "ML")
g.div.50kb.4 <- gls(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
                    data = sim.lands.50kb, weights = varPower(0, ~tmrcaC), method = "ML")

AIC(g.div.50kb.1, g.div.50kb.2, g.div.50kb.3, g.div.50kb.4)

##           df      AIC
## g.div.50kb.1  8 -7490.623
## g.div.50kb.2  8 -7570.956
## g.div.50kb.3  7 -7572.341
## g.div.50kb.4  7 -7490.654

summary(g.div.50kb.3)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC
## Data: sim.lands.50kb
##           AIC          BIC    logLik
## -7572.341 -7541.562  3793.17
##
## Variance function:
## Structure: Power of variance covariate
## Formula: ~theta
## Parameter estimates:
##           power
## 0.4774622
##

```

```
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  0.0206901 0.00001905 1086.3241  0.0000
## thetaC      1.3155987 0.00236880  555.3850  0.0000
## rhoC        0.0087173 0.00554727   1.5715  0.1166
## tmrcaC      0.0234411 0.00026760   87.5965  0.0000
## thetaC:tmrcaC 1.4652196 0.03311231   44.2500  0.0000
##
## Correlation:
##      (Intr) thetaC rhoC   tmrcaC
## thetaC      0.497
## rhoC        0.001  0.052
## tmrcaC      0.013  0.026 -0.004
## thetaC:tmrcaC 0.026  0.025  0.036  0.494
##
## Standardized residuals:
##      Min          Q1          Med          Q3          Max
## -2.955335882 -0.608690299  0.006577071  0.610719662  4.719725049
##
## Residual standard error: 0.003393872
## Degrees of freedom: 600 total; 595 residual
```

```
vif(g.div.50kb.3)
```

```
##      thetaC      rhoC      tmrcaC thetaC:tmrcaC
##      1.003575      1.004633      1.323896      1.325415
```

### 1.1.2 Replicate 2

```
rep_2.pi.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_2/rep_2.pi.50kb.bedgraph", header = 1)
rep_2.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_2/sim.tmrca.50k.map", header = 1)

sim.lands.50kb <- as.data.frame(cbind(rep_2.pi.50kb$pi, sim.theta.50kb$sim, sim.rho.50kb$sim, rep_2.tmrca.50kb$tmrca))
names(sim.lands.50kb) <- c("diversity", "theta", "rho", "tmrca")
```

```
cor.test(~theta+tmrca, data = sim.lands.50kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data:  theta and tmrca
## S = 36499595, p-value = 0.7344
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.01388046
```

```
cor.test(~theta+rho, data = sim.lands.50kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data:  theta and rho
## S = 37660747, p-value = 0.2592
## alternative hypothesis: true rho is not equal to 0
```

```

## sample estimates:
##      rho
## -0.04613476

cor.test(~rho+tmrca, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 33177308, p-value = 0.05494
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.07840555

# centering
sim.lands.50kb$thetaC <- sim.lands.50kb$theta - mean(sim.lands.50kb$theta)
sim.lands.50kb$tmrcaC <- sim.lands.50kb$tmrca - mean(sim.lands.50kb$tmrca)
sim.lands.50kb$rhoC <- sim.lands.50kb$rho - mean(sim.lands.50kb$rho)

sim.lands.50kb$bin <- 1:nrow(sim.lands.50kb)

# for merging:
sim.lands.50kb.rep2 <- sim.lands.50kb
sim.lands.50kb.rep2$Replicate <- 2

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.50kb)
m.div.50kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = sim.lands.50kb)
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = sim.lands.50kb)

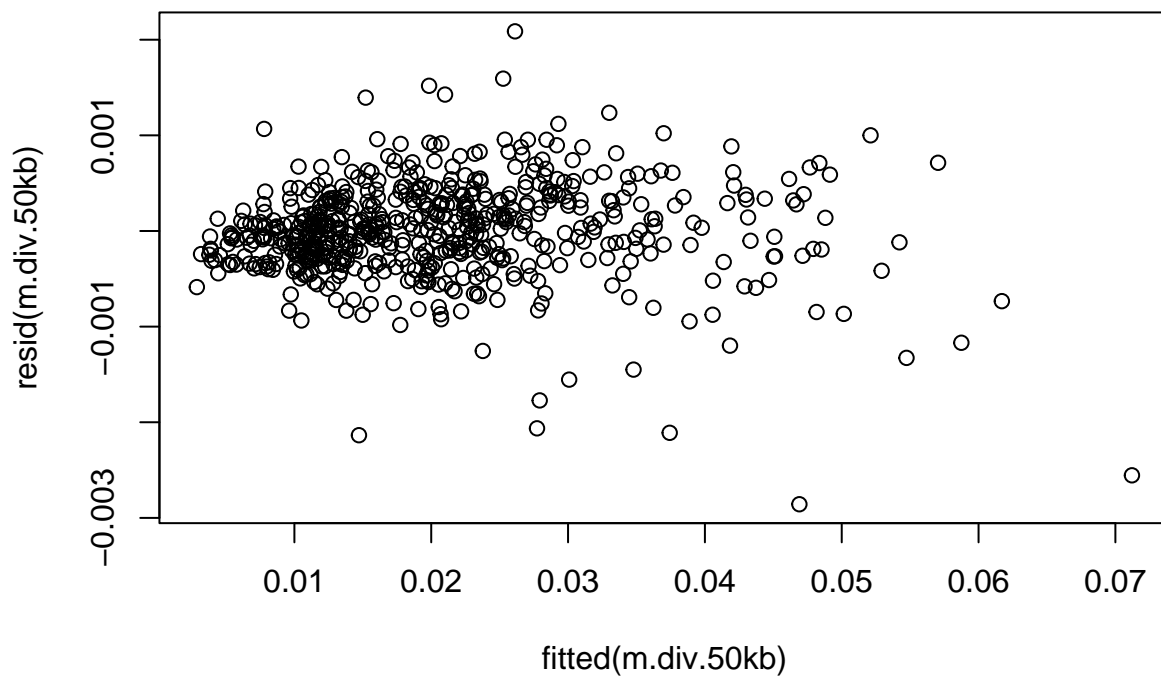
AIC(m.div.50kb, m.div.50kb.2, m.div.50kb.3)

##           df           AIC
## m.div.50kb      6 -7408.739
## m.div.50kb.2    7 -7409.225
## m.div.50kb.3    8 -7407.388

plot(resid(m.div.50kb)~fitted(m.div.50kb))

```





```
dwtest(m.div.50kb)
```

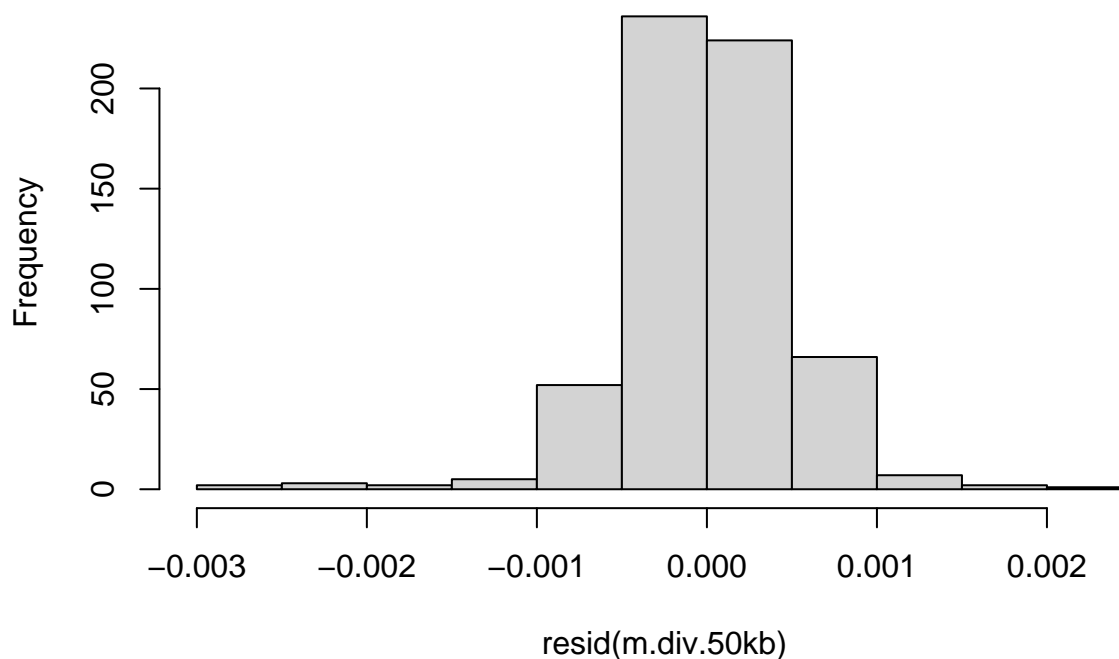
```
##  
## Durbin-Watson test  
##  
## data: m.div.50kb  
## DW = 1.9715, p-value = 0.3445  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.div.50kb)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.div.50kb  
## HMC = 0.51438, p-value = 0.69
```

```
hist(resid(m.div.50kb))
```

## Histogram of resid(m.div.50kb)



```
summary(m.div.50kb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.50kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.857e-03 -2.640e-04  1.830e-06  2.990e-04  2.087e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.057e-02  2.047e-05 1005.301  <2e-16 ***
## thetaC        1.299e+00  2.448e-03  530.815  <2e-16 ***
## rhoC         -1.009e-02  6.919e-03   -1.459    0.145
## tmrcaC        2.385e-02  2.831e-04   84.273  <2e-16 ***
## thetaC:tmrcaC  1.515e+00  3.222e-02   47.015  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005011 on 595 degrees of freedom
## Multiple R-squared:  0.9979, Adjusted R-squared:  0.9979
## F-statistic: 7.16e+04 on 4 and 595 DF, p-value: < 2.2e-16
```

```
anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
```

```
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.50kb[1, 2] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]) / 3
r2.sim.50kb[2, 2] <- anova.diversity$VarExp[1] * 100
r2.sim.50kb[3, 2] <- anova.diversity$VarExp[2] * 100
r2.sim.50kb[4, 2] <- anova.diversity$VarExp[3] * 100
r2.sim.50kb[5, 2] <- anova.diversity$VarExp[4] * 100
```

### 1.1.3 Replicate 3

```
rep_3.pi.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_3/rep_3.pi.50kb.bedgraph", header = TRUE)
rep_3.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_3/sim.tmrca.50k.map", header = TRUE)

sim.lands.50kb <- as.data.frame(cbind(rep_3.pi.50kb$pi, sim.theta.50kb$sim, sim.rho.50kb$sim, rep_3.tmrca.50kb$tmrca))
names(sim.lands.50kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 37873663, p-value = 0.203
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.05204912

cor.test(~theta+rho, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 37660747, p-value = 0.2592
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.04613476

cor.test(~rho+tmrca, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 35383984, p-value = 0.6757
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.01710883

# centering
sim.lands.50kb$thetaC <- sim.lands.50kb$theta - mean(sim.lands.50kb$theta)
sim.lands.50kb$tmrcaC <- sim.lands.50kb$tmrca - mean(sim.lands.50kb$tmrca)
```

```

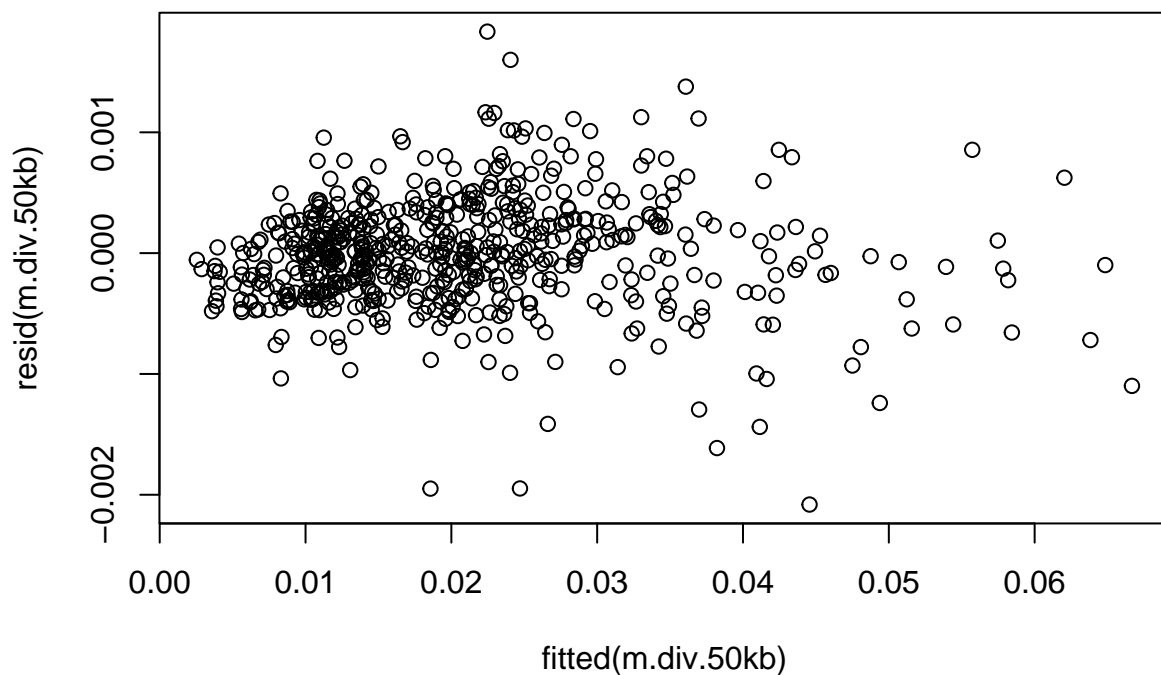
sim.lands.50kb$rhoC <- sim.lands.50kb$rho - mean(sim.lands.50kb$rho)

sim.lands.50kb$bin <- 1:nrow(sim.lands.50kb)

# for merging:
sim.lands.50kb.rep3 <- sim.lands.50kb
sim.lands.50kb.rep3$Replicate <- 3

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.50kb)
plot(resid(m.div.50kb)~fitted(m.div.50kb))

```



```

dwtest(m.div.50kb)

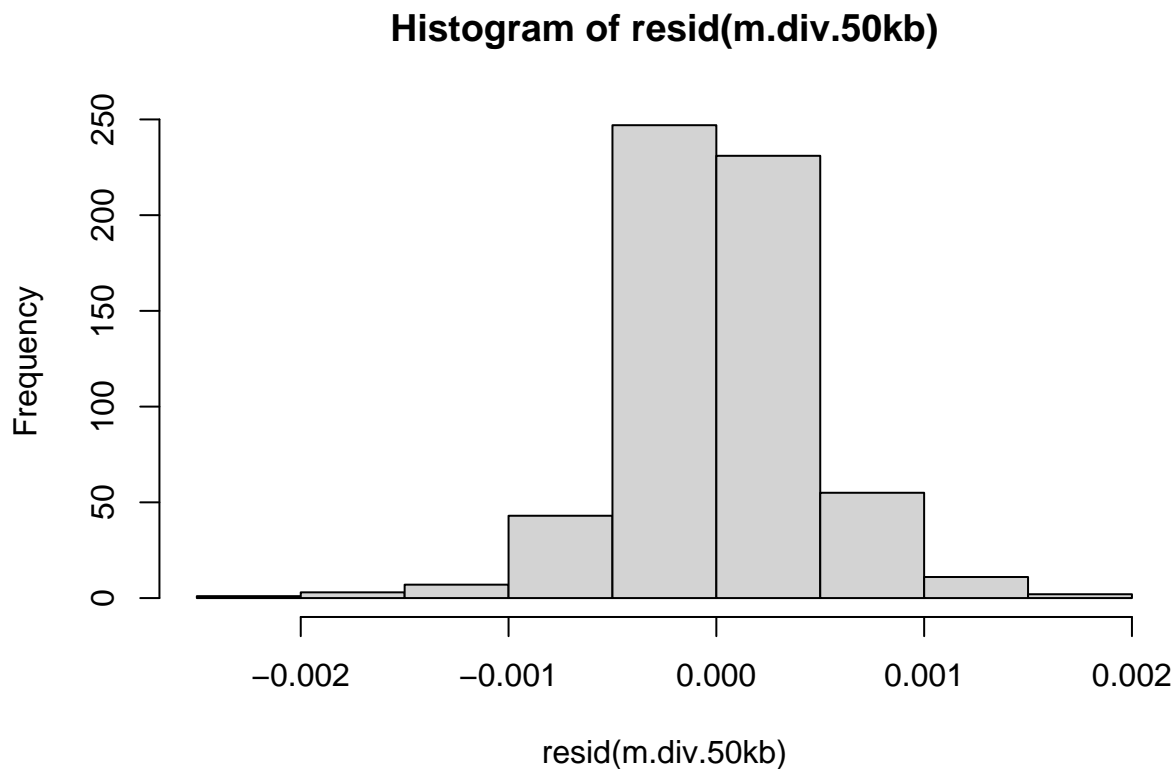
##
## Durbin-Watson test
##
## data: m.div.50kb
## DW = 1.9214, p-value = 0.1563
## alternative hypothesis: true autocorrelation is greater than 0

hmcetest(m.div.50kb)

##
## Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.50434, p-value = 0.571

```

```
hist(resid(m.div.50kb))
```



```
summary(m.div.50kb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.50kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.081e-03 -2.636e-04 -7.680e-06  2.600e-04  1.833e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0206720  0.0000187 1105.47  <2e-16 ***
## thetaC       1.3103632  0.0022312  587.29  <2e-16 ***
## rhoC         0.0064418  0.0063164    1.02   0.308
## tmrcaC       0.0236099  0.0002297  102.80  <2e-16 ***
## thetaC:tmrcaC 1.5012872  0.0254605   58.97  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0004578 on 595 degrees of freedom
## Multiple R-squared:  0.9983, Adjusted R-squared:  0.9983
## F-statistic: 8.857e+04 on 4 and 595 DF, p-value: < 2.2e-16
```

```

anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.50kb[1, 3] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]) / anova.diversity$VarExp[4]
r2.sim.50kb[2, 3] <- anova.diversity$VarExp[1] / anova.diversity$VarExp[4]
r2.sim.50kb[3, 3] <- anova.diversity$VarExp[2] / anova.diversity$VarExp[4]
r2.sim.50kb[4, 3] <- anova.diversity$VarExp[3] / anova.diversity$VarExp[4]
r2.sim.50kb[5, 3] <- anova.diversity$VarExp[4] / anova.diversity$VarExp[4]

```

#### 1.1.4 Replicate 4

```

rep_4.pi.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_4/rep_4.pi.50kb.bedgraph", header = TRUE, as.is = TRUE)
rep_4.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_4/sim.tmrca.50k.map", header = TRUE, as.is = TRUE)

sim.lands.50kb <- as.data.frame(cbind(rep_4.pi.50kb$pi, sim.theta.50kb$sim, sim.rho.50kb$sim, rep_4.tmrca.50kb$tmrca))
names(sim.lands.50kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 36135074, p-value = 0.9269
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.003754843

cor.test(~theta+rho, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 37660747, p-value = 0.2592
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.04613476

cor.test(~rho+tmrca, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 37039432, p-value = 0.4801
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.02887597

```

```

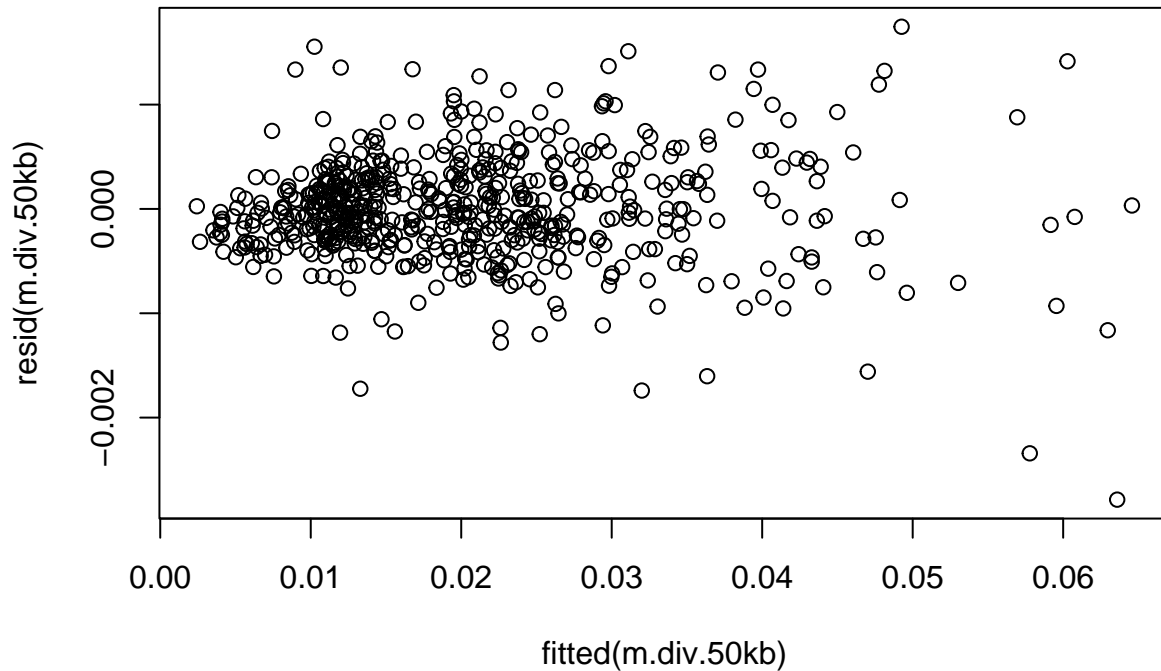
# centering
sim.lands.50kb$thetaC <- sim.lands.50kb$theta - mean(sim.lands.50kb$theta)
sim.lands.50kb$tmrcaC <- sim.lands.50kb$tmrca - mean(sim.lands.50kb$tmrca)
sim.lands.50kb$rhoC <- sim.lands.50kb$rho - mean(sim.lands.50kb$rho)

sim.lands.50kb$bin <- 1:nrow(sim.lands.50kb)

# for merging:
sim.lands.50kb.rep4 <- sim.lands.50kb
sim.lands.50kb.rep4$Replicate <- 4

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.50kb)
plot(resid(m.div.50kb)~fitted(m.div.50kb))

```



```

dwtest(m.div.50kb)

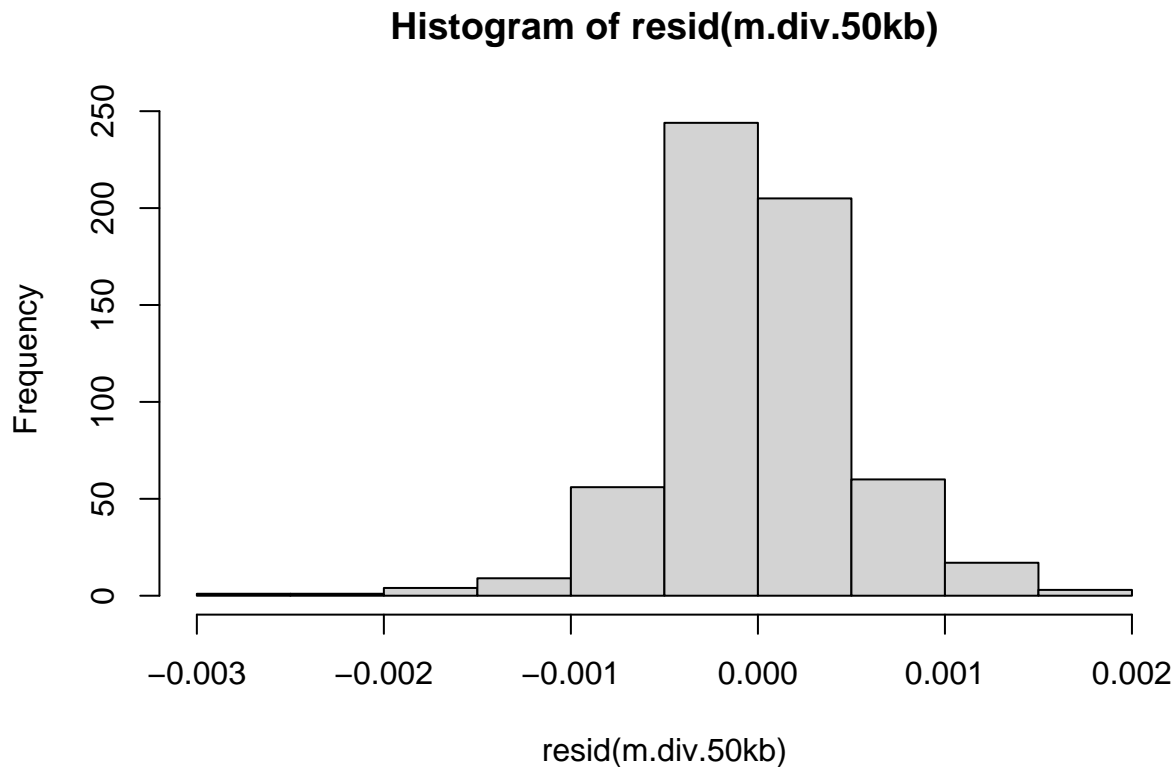
##
## Durbin-Watson test
##
## data: m.div.50kb
## DW = 2.0985, p-value = 0.8779
## alternative hypothesis: true autocorrelation is greater than 0

hmctest(m.div.50kb)

##
## Harrison-McCabe test

```

```
##
## data:  m.div.50kb
## HMC = 0.53836, p-value = 0.911
hist(resid(m.div.50kb))
```



```
summary(m.div.50kb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.50kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.787e-03 -2.851e-04 -3.163e-05  2.685e-04  1.746e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.050e-02  2.077e-05  987.094  <2e-16 ***
## thetaC       1.300e+00  2.476e-03  525.118  <2e-16 ***
## rhoC        1.386e-02  7.016e-03   1.976   0.0486 *
## tmrcaC       2.349e-02  2.643e-04  88.857  <2e-16 ***
## thetaC:tmrcaC 1.493e+00  3.134e-02  47.634  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```



```
## Residual standard error: 0.0005088 on 595 degrees of freedom
## Multiple R-squared: 0.9979, Adjusted R-squared: 0.9979
## F-statistic: 7.205e+04 on 4 and 595 DF, p-value: < 2.2e-16
```

```
anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.50kb[1, 4] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]) / anova.diversity$VarExp[4]
r2.sim.50kb[2, 4] <- anova.diversity$VarExp[1] / anova.diversity$VarExp[4]
r2.sim.50kb[3, 4] <- anova.diversity$VarExp[2] / anova.diversity$VarExp[4]
r2.sim.50kb[4, 4] <- anova.diversity$VarExp[3] / anova.diversity$VarExp[4]
r2.sim.50kb[5, 4] <- anova.diversity$VarExp[4] / anova.diversity$VarExp[4]
```

### 1.1.5 Replicate 5

```
rep_5.pi.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_5/rep_5.pi.50kb.bedgraph", header = TRUE, as.is = TRUE)
rep_5.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_5/sim.tmrca.50k.map", header = TRUE, as.is = TRUE)

sim.lands.50kb <- as.data.frame(cbind(rep_5.pi.50kb$pi, sim.theta.50kb$sim, sim.rho.50kb$sim, rep_5.tmrca.50kb$tmrca))
names(sim.lands.50kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.50kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 35425520, p-value = 0.6965
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.01595505
```

```
cor.test(~theta+rho, data = sim.lands.50kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 37660747, p-value = 0.2592
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.04613476
```

```
cor.test(~rho+tmrca, data = sim.lands.50kb, method = "spearman")
```

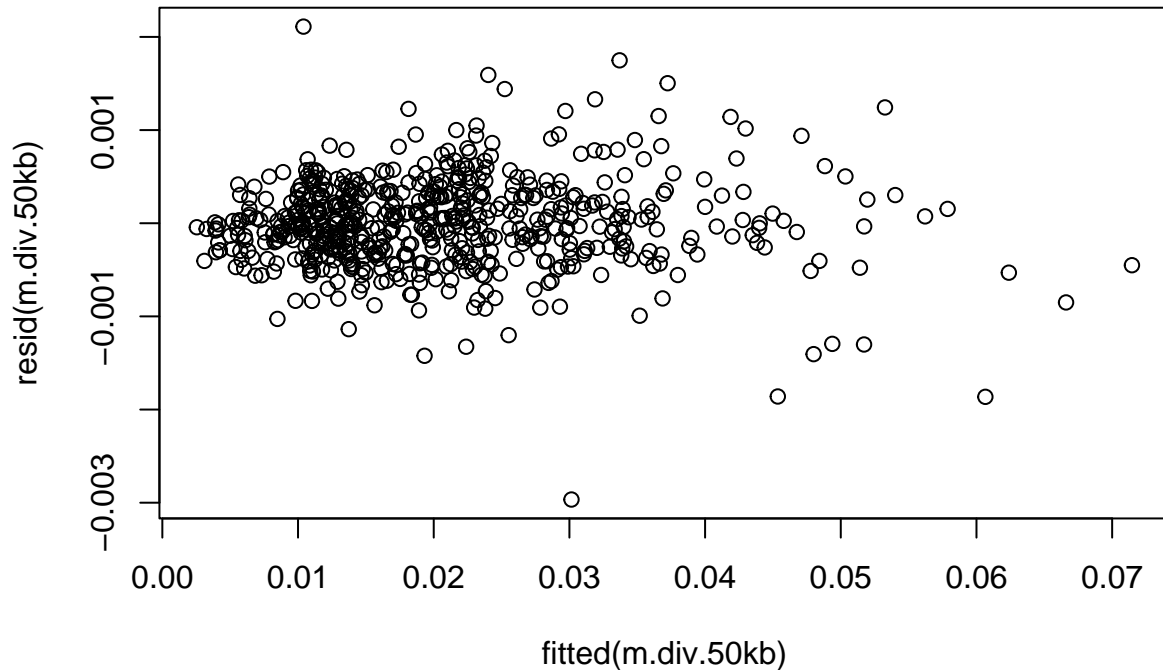
```
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 37018948, p-value = 0.4888
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
```

```
## -0.02830697
# centering
sim.lands.50kb$thetaC <- sim.lands.50kb$theta - mean(sim.lands.50kb$theta)
sim.lands.50kb$tmrcaC <- sim.lands.50kb$tmrca - mean(sim.lands.50kb$tmrca)
sim.lands.50kb$rhoC <- sim.lands.50kb$rho - mean(sim.lands.50kb$rho)

sim.lands.50kb$bin <- 1:nrow(sim.lands.50kb)

# for merging:
sim.lands.50kb.rep5 <- sim.lands.50kb
sim.lands.50kb.rep5$Replicate <- 5

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.50kb)
plot(resid(m.div.50kb)~fitted(m.div.50kb))
```



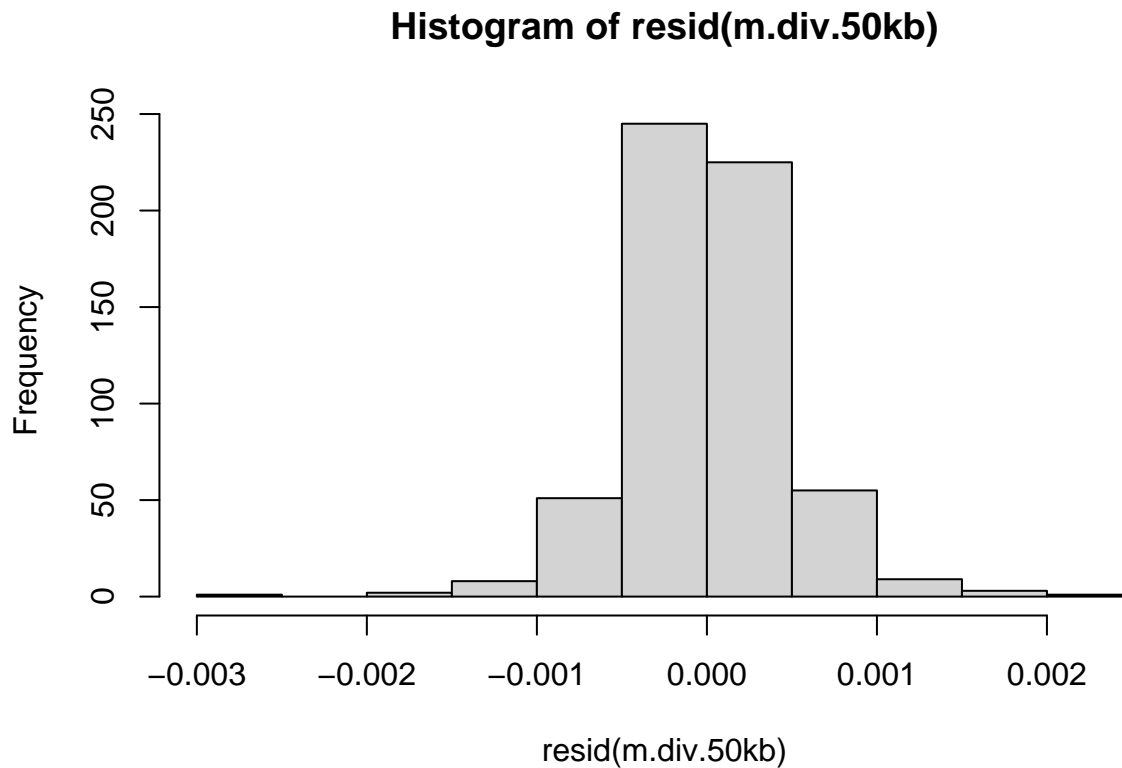
```
dwtest(m.div.50kb)

##
## Durbin-Watson test
##
## data: m.div.50kb
## DW = 2.0931, p-value = 0.864
## alternative hypothesis: true autocorrelation is greater than 0

hmctest(m.div.50kb)

##
```

```
## Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.45962, p-value = 0.069
hist(resid(m.div.50kb))
```



```
summary(m.div.50kb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.50kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.966e-03 -2.902e-04 -1.674e-05  2.879e-04  2.110e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.062e-02  1.967e-05 1048.31  <2e-16 ***
## thetaC       1.308e+00  2.348e-03  557.06  <2e-16 ***
## rhoC        1.528e-03  6.645e-03    0.23   0.818
## tmrcaC       2.374e-02  2.674e-04  88.80  <2e-16 ***
## thetaC:tmrcaC 1.471e+00  3.126e-02  47.05  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.0004816 on 595 degrees of freedom
## Multiple R-squared: 0.9981, Adjusted R-squared: 0.9981
## F-statistic: 7.932e+04 on 4 and 595 DF, p-value: < 2.2e-16

anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.50kb[1, 5] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.sim.50kb[2, 5] <- anova.diversity$VarExp[1] * 100
r2.sim.50kb[3, 5] <- anova.diversity$VarExp[2] * 100
r2.sim.50kb[4, 5] <- anova.diversity$VarExp[3] * 100
r2.sim.50kb[5, 5] <- anova.diversity$VarExp[4] * 100
```

### 1.1.6 Replicate 6

```
rep_6.pi.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_6/rep_6.pi.50kb.bedgraph", header = TRUE)
rep_6.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_6/sim.tmrca.50k.map", header = TRUE)

sim.lands.50kb <- as.data.frame(cbind(rep_6.pi.50kb$pi, sim.theta.50kb$sim, sim.rho.50kb$sim, rep_6.tmrca.50kb$tmrca))
names(sim.lands.50kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.50kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 36901608, p-value = 0.5403
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.0250475

cor.test(~theta+rho, data = sim.lands.50kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 37660747, p-value = 0.2592
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.04613476

cor.test(~rho+tmrca, data = sim.lands.50kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 37734896, p-value = 0.2384
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
```

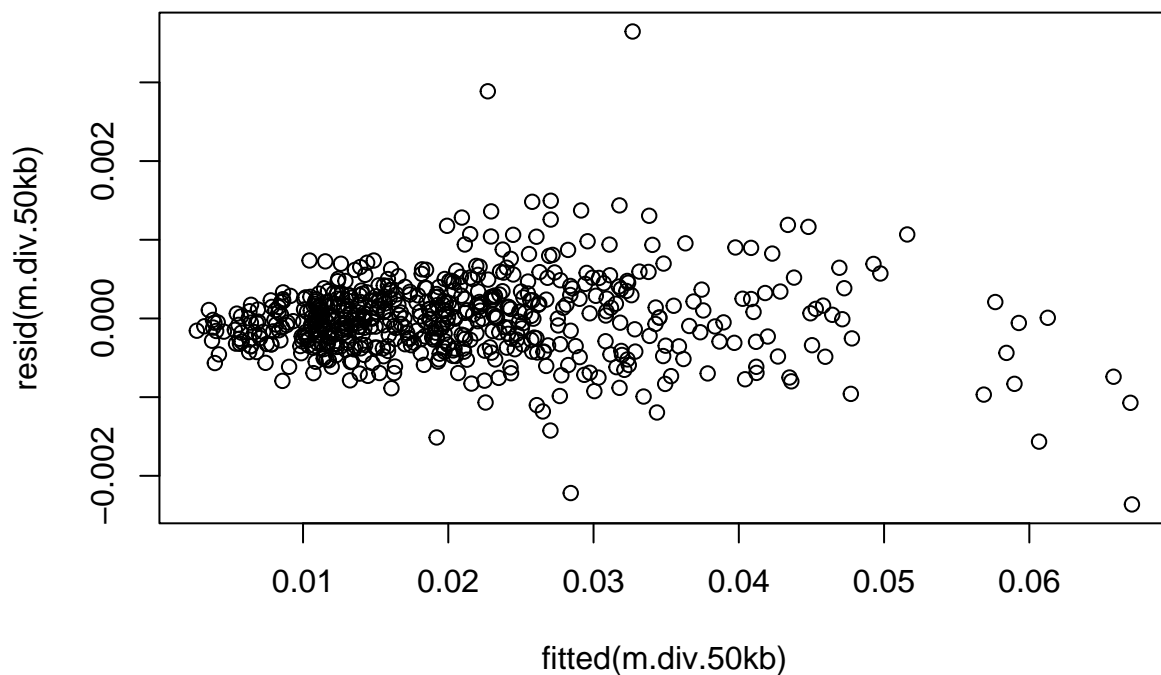
```
##          rho
## -0.04819447

# centering
sim.lands.50kb$thetaC <- sim.lands.50kb$theta - mean(sim.lands.50kb$theta)
sim.lands.50kb$tmrcaC <- sim.lands.50kb$tmrca - mean(sim.lands.50kb$tmrca)
sim.lands.50kb$rhoC <- sim.lands.50kb$rho - mean(sim.lands.50kb$rho)

sim.lands.50kb$bin <- 1:nrow(sim.lands.50kb)

# for merging:
sim.lands.50kb.rep6 <- sim.lands.50kb
sim.lands.50kb.rep6$Replicate <- 6

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.50kb)
plot(resid(m.div.50kb)~fitted(m.div.50kb))
```



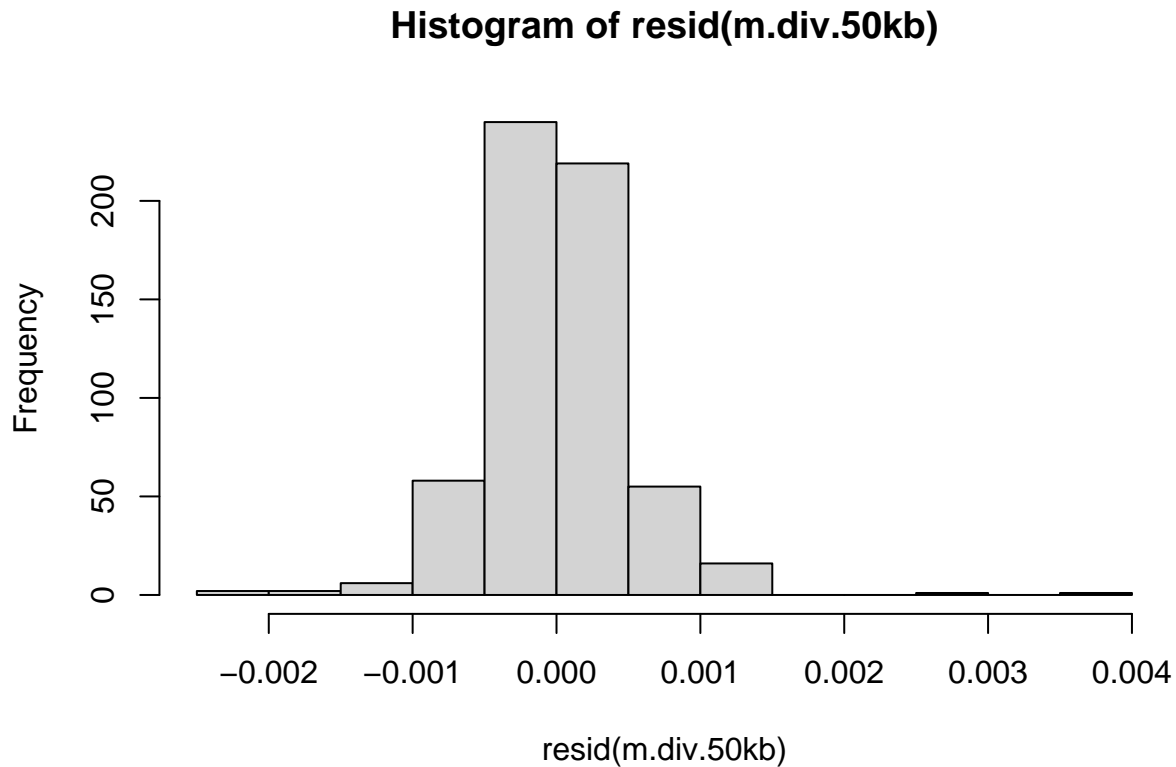
```
dwtest(m.div.50kb)

##
## Durbin-Watson test
##
## data: m.div.50kb
## DW = 2.1478, p-value = 0.9618
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.50kb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.46094, p-value = 0.081
```

```
hist(resid(m.div.50kb))
```



```
summary(m.div.50kb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.50kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0023630 -0.0002937 -0.0000107  0.0002498  0.0036466
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.058e-02  2.073e-05  993.114  <2e-16 ***
## thetaC       1.304e+00  2.479e-03  526.030  <2e-16 ***
## rhoC        -1.179e-02  7.005e-03  -1.683    0.0929 .
## tmrcaC       2.363e-02  2.752e-04  85.861  <2e-16 ***
```

```
## thetaC:tmrcaC 1.439e+00 3.174e-02 45.345 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005077 on 595 degrees of freedom
## Multiple R-squared: 0.998, Adjusted R-squared: 0.9979
## F-statistic: 7.29e+04 on 4 and 595 DF, p-value: < 2.2e-16

anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.50kb[1, 6] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.sim.50kb[2, 6] <- anova.diversity$VarExp[1] * 100
r2.sim.50kb[3, 6] <- anova.diversity$VarExp[2] * 100
r2.sim.50kb[4, 6] <- anova.diversity$VarExp[3] * 100
r2.sim.50kb[5, 6] <- anova.diversity$VarExp[4] * 100
```

### 1.1.7 Replicate 7

```
rep_7.pi.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_7/rep_7.pi.50kb.bedgraph", header =
rep_7.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_7/sim.tmrca.50k.map", header =

sim.lands.50kb <- as.data.frame(cbind(rep_7.pi.50kb$pi, sim.theta.50kb$sim, sim.rho.50kb$sim, rep_7.tmrca.50kb$tmrca))
names(sim.lands.50kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 38638560, p-value = 0.0728
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.0732963

cor.test(~theta+rho, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 37660747, p-value = 0.2592
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.04613476

cor.test(~rho+tmrca, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: rho and tmrca
```

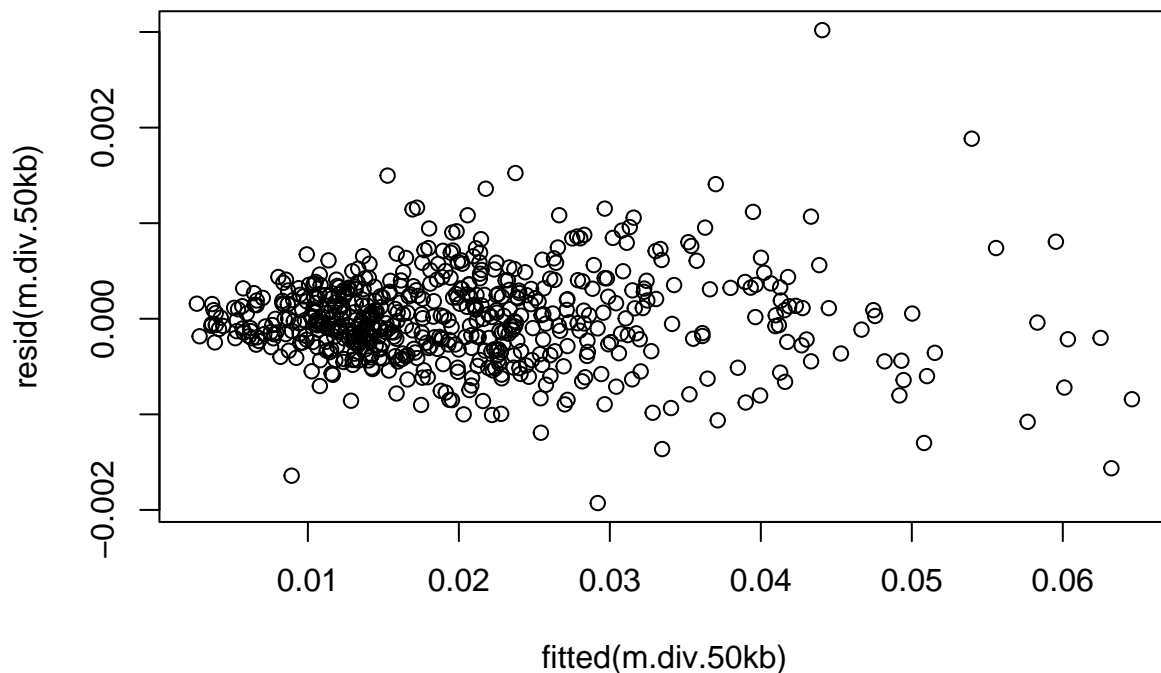
```
## S = 36196758, p-value = 0.8936
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.005468293

# centering
sim.lands.50kb$thetaC <- sim.lands.50kb$theta - mean(sim.lands.50kb$theta)
sim.lands.50kb$tmrcaC <- sim.lands.50kb$tmrca - mean(sim.lands.50kb$tmrca)
sim.lands.50kb$rhoC <- sim.lands.50kb$rho - mean(sim.lands.50kb$rho)

sim.lands.50kb$bin <- 1:nrow(sim.lands.50kb)

# for merging:
sim.lands.50kb.rep7 <- sim.lands.50kb
sim.lands.50kb.rep7$Replicate <- 7

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.50kb)
plot(resid(m.div.50kb)~fitted(m.div.50kb))
```



```
dwtest(m.div.50kb)

##
## Durbin-Watson test
##
## data: m.div.50kb
## DW = 1.959, p-value = 0.2918
```



```
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.50kb)
```

```
##
```

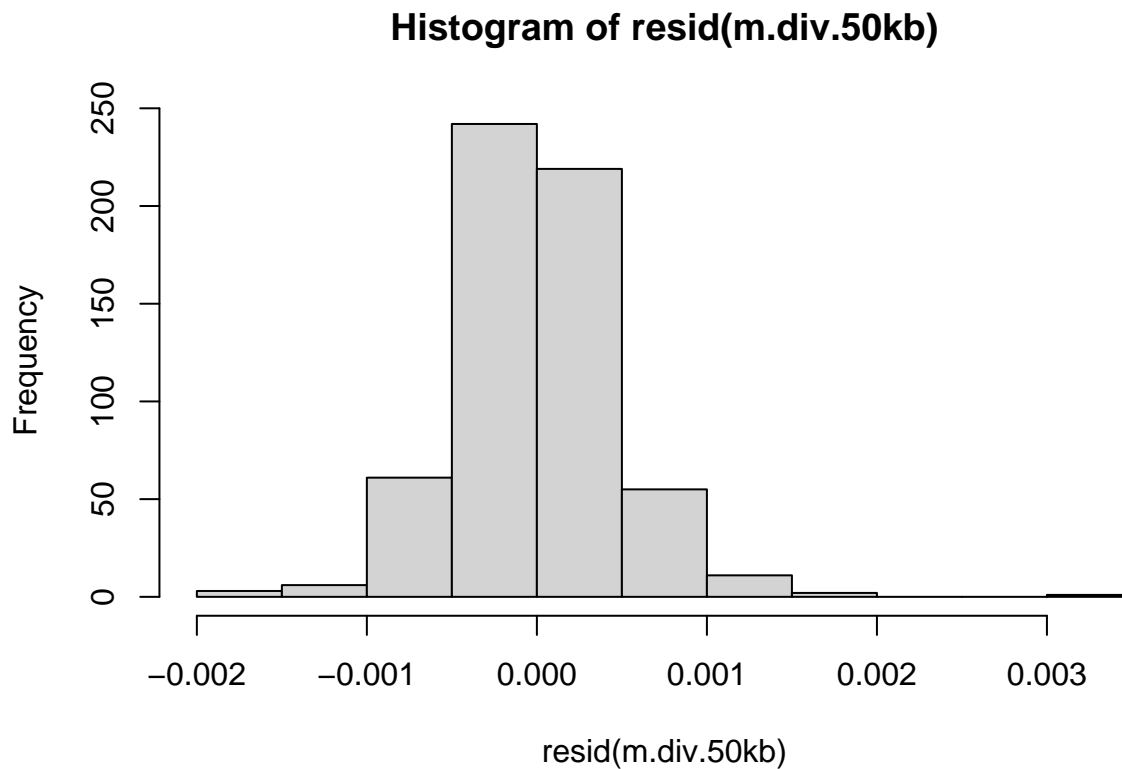
```
## Harrison-McCabe test
```

```
##
```

```
## data: m.div.50kb
```

```
## HMC = 0.47957, p-value = 0.235
```

```
hist(resid(m.div.50kb))
```



```
summary(m.div.50kb)
```

```
##
```

```
## Call:
```

```
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
```

```
## data = sim.lands.50kb)
```

```
##
```

```
## Residuals:
```

	Min	1Q	Median	3Q	Max
##	-1.928e-03	-2.610e-04	-1.731e-05	2.569e-04	3.020e-03

```
##
```

```
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	2.061e-02	1.945e-05	1059.614	<2e-16 ***
## thetaC	1.313e+00	2.320e-03	565.892	<2e-16 ***
## rhoC	-5.753e-03	6.560e-03	-0.877	0.381

```
## tmrcaC          2.389e-02  2.547e-04  93.801  <2e-16 ***
## thetaC:tmrcaC  1.491e+00  3.019e-02  49.398  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0004759 on 595 degrees of freedom
## Multiple R-squared:  0.9982, Adjusted R-squared:  0.9982
## F-statistic: 8.263e+04 on 4 and 595 DF,  p-value: < 2.2e-16

anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.50kb[1, 7] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.sim.50kb[2, 7] <- anova.diversity$VarExp[1] * 100
r2.sim.50kb[3, 7] <- anova.diversity$VarExp[2] * 100
r2.sim.50kb[4, 7] <- anova.diversity$VarExp[3] * 100
r2.sim.50kb[5, 7] <- anova.diversity$VarExp[4] * 100
```

### 1.1.8 Replicate 8

```
rep_8.pi.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_8/rep_8.pi.50kb.bedgraph", header = TRUE)
rep_8.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_8/sim.tmrca.50k.map", header = TRUE)

sim.lands.50kb <- as.data.frame(cbind(rep_8.pi.50kb$pi, sim.theta.50kb$sim, sim.rho.50kb$sim, rep_8.tmrca.50kb$tmrca))
names(sim.lands.50kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.50kb, method = "spearman")

## Warning in cor.test.default(x = c(0.0101175988992442, 0.0101175988992442, :
## Cannot compute exact p-value with ties
##
## Spearman's rank correlation rho
##
## data:  theta and tmrca
## S = 37183069, p-value = 0.4216
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.03286589

cor.test(~theta+rho, data = sim.lands.50kb, method = "spearman")

## Warning in cor.test.default(x = c(0.0101175988992442, 0.0101175988992442, :
## Cannot compute exact p-value with ties
##
## Spearman's rank correlation rho
##
## data:  theta and rho
## S = 37660747, p-value = 0.2592
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.04613476
```

```
cor.test(~rho+tmrca, data = sim.lands.50kb, method = "spearman")

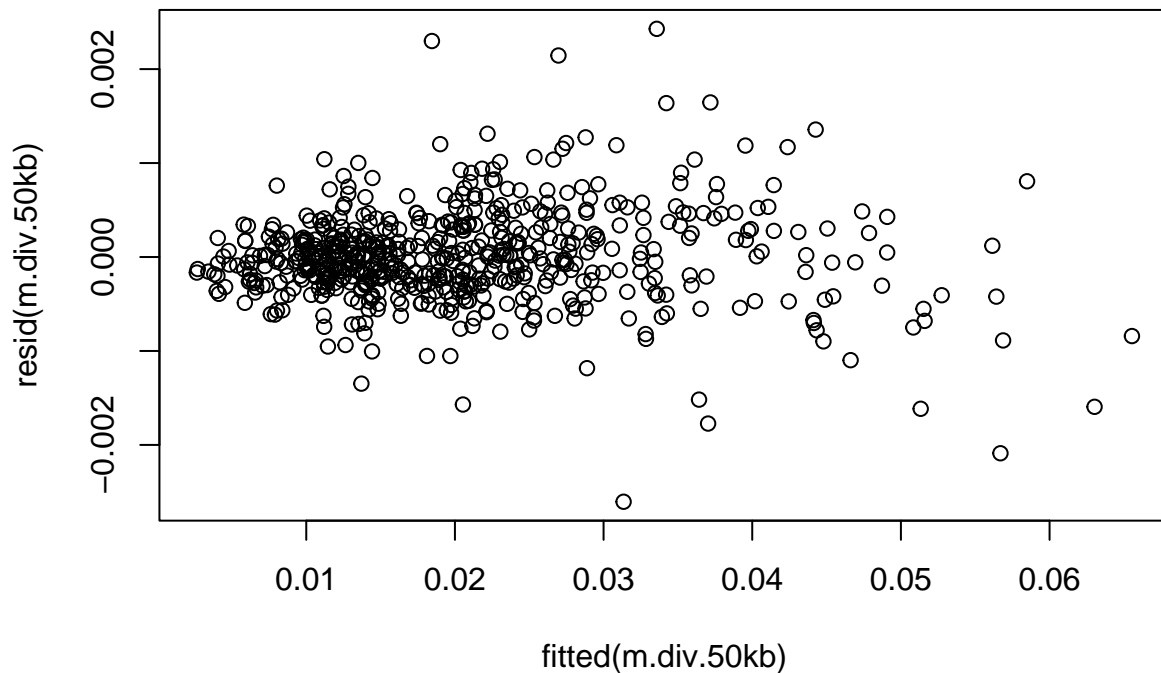
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 37073930, p-value = 0.4656
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.02983425

# centering
sim.lands.50kb$thetaC <- sim.lands.50kb$theta - mean(sim.lands.50kb$theta)
sim.lands.50kb$tmrcaC <- sim.lands.50kb$tmrca - mean(sim.lands.50kb$tmrca)
sim.lands.50kb$rhoC <- sim.lands.50kb$rho - mean(sim.lands.50kb$rho)

sim.lands.50kb$bin <- 1:nrow(sim.lands.50kb)

# for merging:
sim.lands.50kb.rep8 <- sim.lands.50kb
sim.lands.50kb.rep8$Replicate <- 8

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.50kb)
plot(resid(m.div.50kb)~fitted(m.div.50kb))
```



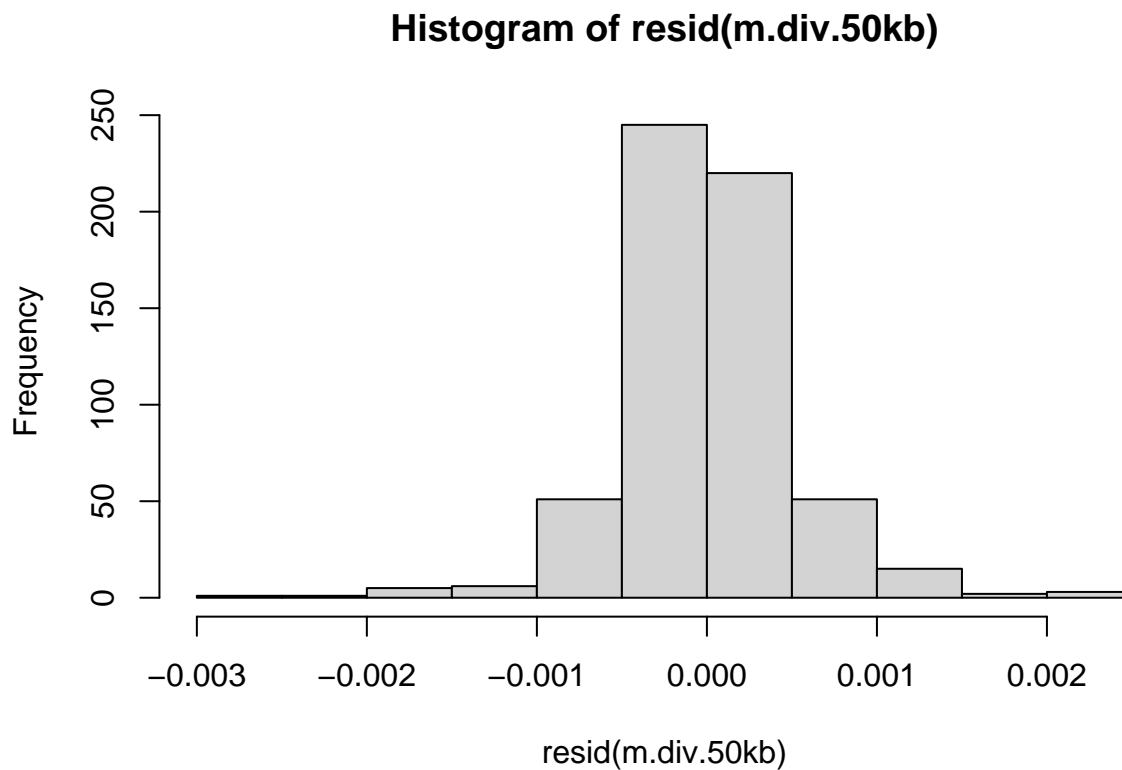
```
dwtest(m.div.50kb)
```

```
##  
## Durbin-Watson test  
##  
## data: m.div.50kb  
## DW = 2.0323, p-value = 0.6384  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.50kb)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.div.50kb  
## HMC = 0.48525, p-value = 0.3
```

```
hist(resid(m.div.50kb))
```



```
summary(m.div.50kb)
```

```
##  
## Call:  
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,  
## data = sim.lands.50kb)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max
```

```
## -2.605e-03 -2.691e-04 -1.686e-05 2.620e-04 2.428e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.057e-02  2.065e-05 996.284  <2e-16 ***
## thetaC      1.303e+00  2.468e-03 527.803  <2e-16 ***
## rhoC       -8.730e-04  6.970e-03  -0.125    0.9
## tmrcaC      2.380e-02  2.848e-04 83.583   <2e-16 ***
## thetaC:tmrcaC 1.383e+00  3.162e-02 43.722   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005051 on 595 degrees of freedom
## Multiple R-squared:  0.9979, Adjusted R-squared:  0.9979
## F-statistic: 7.078e+04 on 4 and 595 DF, p-value: < 2.2e-16
```

```
anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.50kb[1, 8] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.sim.50kb[2, 8] <- anova.diversity$VarExp[1] * 100
r2.sim.50kb[3, 8] <- anova.diversity$VarExp[2] * 100
r2.sim.50kb[4, 8] <- anova.diversity$VarExp[3] * 100
r2.sim.50kb[5, 8] <- anova.diversity$VarExp[4] * 100
```

### 1.1.9 Replicate 9

```
rep_9.pi.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_9/rep_9.pi.50kb.bedgraph", header =
rep_9.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_9/sim.tmrca.50k.map", header =

sim.lands.50kb <- as.data.frame(cbind(rep_9.pi.50kb$pi, sim.theta.50kb$sim, sim.rho.50kb$sim, rep_9.tmr
names(sim.lands.50kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 35536739, p-value = 0.7531
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.01286561

cor.test(~theta+rho, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 37660747, p-value = 0.2592
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
```

```

##          rho
## -0.04613476

cor.test(~rho+tmrca, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 37091756, p-value = 0.4583
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##          rho
## -0.03032942

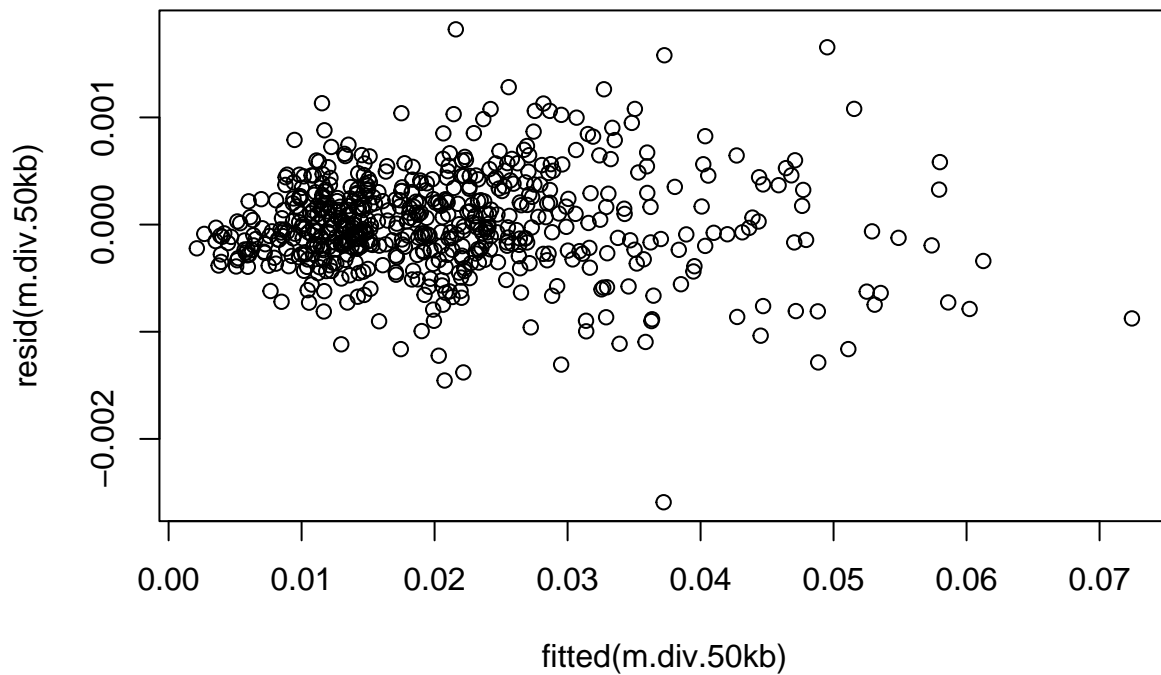
# centering
sim.lands.50kb$thetaC <- sim.lands.50kb$theta - mean(sim.lands.50kb$theta)
sim.lands.50kb$tmrcaC <- sim.lands.50kb$tmrca - mean(sim.lands.50kb$tmrca)
sim.lands.50kb$rhoC <- sim.lands.50kb$rho - mean(sim.lands.50kb$rho)

sim.lands.50kb$bin <- 1:nrow(sim.lands.50kb)

# for merging:
sim.lands.50kb.rep9 <- sim.lands.50kb
sim.lands.50kb.rep9 $Replicate <- 9

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.50kb)
plot(resid(m.div.50kb)~fitted(m.div.50kb))

```



```
dwtest(m.div.50kb)
```

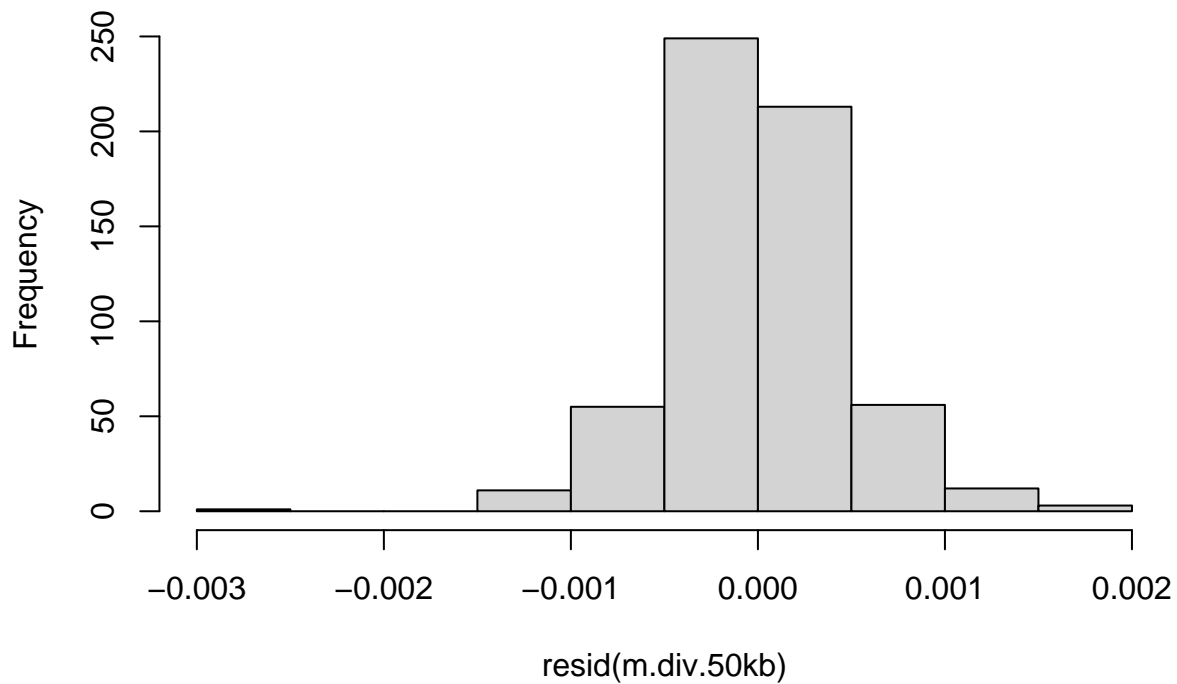
```
##  
## Durbin-Watson test  
##  
## data: m.div.50kb  
## DW = 1.9547, p-value = 0.2758  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcctest(m.div.50kb)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.div.50kb  
## HMC = 0.55455, p-value = 0.97
```

```
hist(resid(m.div.50kb))
```

## Histogram of resid(m.div.50kb)



```
summary(m.div.50kb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.50kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.591e-03 -2.409e-04 -2.982e-05  2.668e-04  1.822e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.065e-02  1.899e-05 1087.271  <2e-16 ***
## thetaC       1.308e+00  2.265e-03  577.567  <2e-16 ***
## rhoC        -7.769e-03  6.412e-03   -1.212    0.226
## tmrcaC       2.404e-02  2.565e-04   93.731  <2e-16 ***
## thetaC:tmrcaC 1.508e+00  2.888e-02   52.235  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.000465 on 595 degrees of freedom
## Multiple R-squared:  0.9983, Adjusted R-squared:  0.9983
## F-statistic: 8.777e+04 on 4 and 595 DF, p-value: < 2.2e-16

anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
```



```
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.50kb[1, 9] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.sim.50kb[2, 9] <- anova.diversity$VarExp[1] * 100
r2.sim.50kb[3, 9] <- anova.diversity$VarExp[2] * 100
r2.sim.50kb[4, 9] <- anova.diversity$VarExp[3] * 100
r2.sim.50kb[5, 9] <- anova.diversity$VarExp[4] * 100
```

### 1.1.10 Replicate 10

```
rep_10.pi.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_10/rep_10.pi.50kb.bedgraph", header=TRUE)
rep_10.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_10/sim.tmrca.50k.map", header=TRUE)

sim.lands.50kb <- as.data.frame(cbind(rep_10.pi.50kb$pi, sim.theta.50kb$sim, sim.rho.50kb$sim, rep_10.tmrca.50kb$tmrca))
names(sim.lands.50kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 37655834, p-value = 0.2606
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.0459983

cor.test(~theta+rho, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 37660747, p-value = 0.2592
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.04613476

cor.test(~rho+tmrca, data = sim.lands.50kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 36687846, p-value = 0.6403
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.01910966

# centering
sim.lands.50kb$thetaC <- sim.lands.50kb$theta - mean(sim.lands.50kb$theta)
sim.lands.50kb$tmrcaC <- sim.lands.50kb$tmrca - mean(sim.lands.50kb$tmrca)
```

```

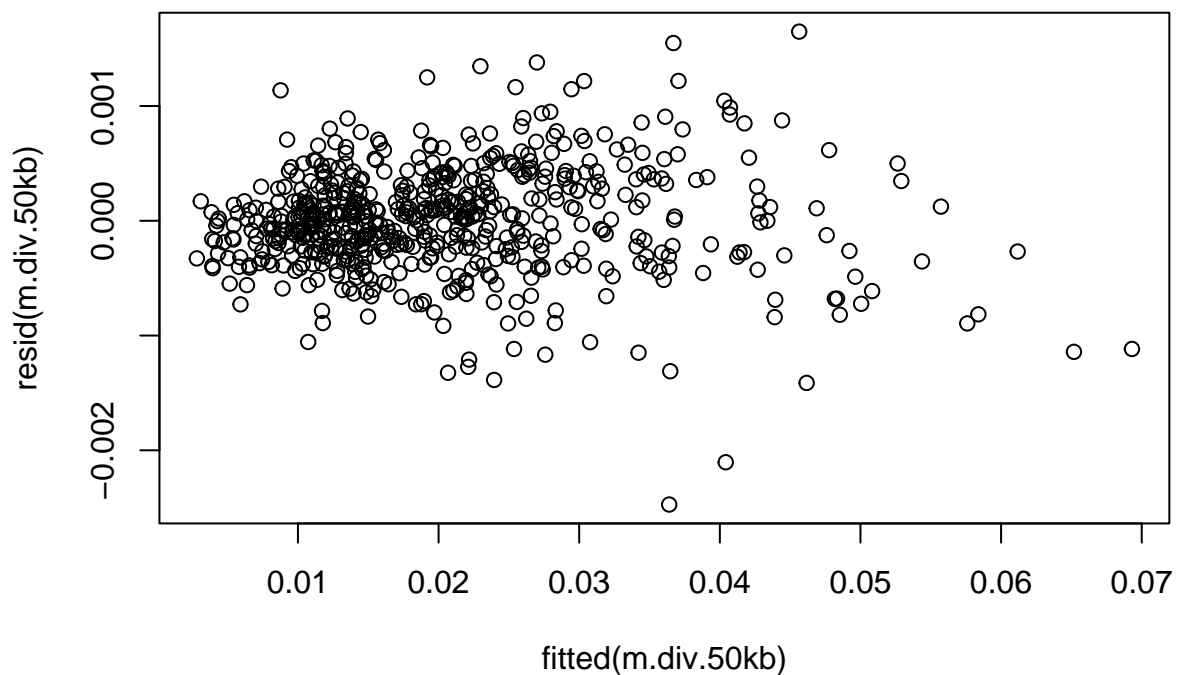
sim.lands.50kb$rhoC <- sim.lands.50kb$rho - mean(sim.lands.50kb$rho)

sim.lands.50kb$bin <- 1:nrow(sim.lands.50kb)

# for merging:
sim.lands.50kb.rep10 <- sim.lands.50kb
sim.lands.50kb.rep10$Replicate <- 10

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.50kb)
plot(resid(m.div.50kb)~fitted(m.div.50kb))

```



```

dwtest(m.div.50kb)

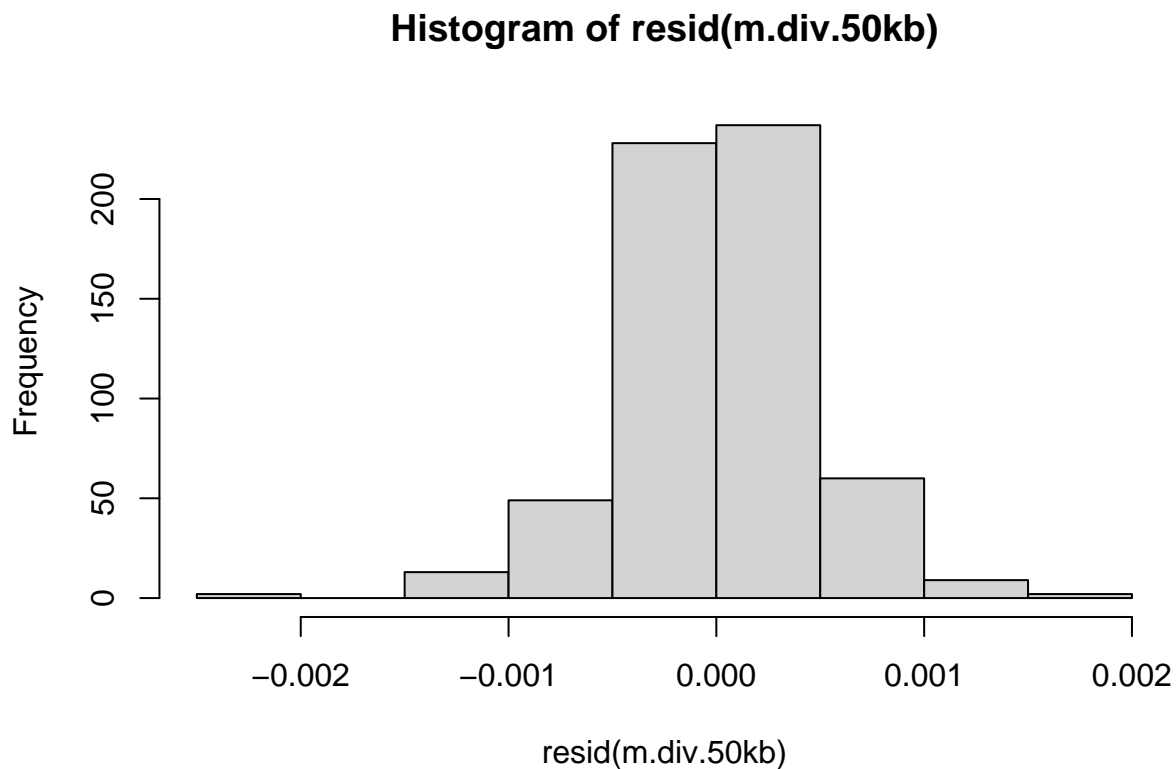
##
## Durbin-Watson test
##
## data: m.div.50kb
## DW = 1.9775, p-value = 0.3738
## alternative hypothesis: true autocorrelation is greater than 0

hmcetest(m.div.50kb)

##
## Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.43153, p-value = 0.006

```

```
hist(resid(m.div.50kb))
```



```
summary(m.div.50kb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.50kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.473e-03 -2.632e-04  1.203e-05  2.861e-04  1.648e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.070e-02  1.924e-05 1075.884  <2e-16 ***
## thetaC       1.311e+00  2.304e-03  568.969  <2e-16 ***
## rhoC         7.558e-03  6.493e-03   1.164    0.245
## tmrcaC       2.376e-02  2.376e-04 100.018  <2e-16 ***
## thetaC:tmrcaC 1.554e+00  2.642e-02  58.816   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0004703 on 595 degrees of freedom
## Multiple R-squared:  0.9982, Adjusted R-squared:  0.9982
## F-statistic: 8.211e+04 on 4 and 595 DF, p-value: < 2.2e-16
```

```

anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.50kb[1, 10] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.sim.50kb[2, 10] <- anova.diversity$VarExp[1] * 100
r2.sim.50kb[3, 10] <- anova.diversity$VarExp[2] * 100
r2.sim.50kb[4, 10] <- anova.diversity$VarExp[3] * 100
r2.sim.50kb[5, 10] <- anova.diversity$VarExp[4] * 100

r2.sim.50kb$average <- rowMeans(r2.sim.50kb)
r2.sim.50kb <- transform(r2.sim.50kb, sd=apply(r2.sim.50kb, 1, sd, na.rm = TRUE))

```

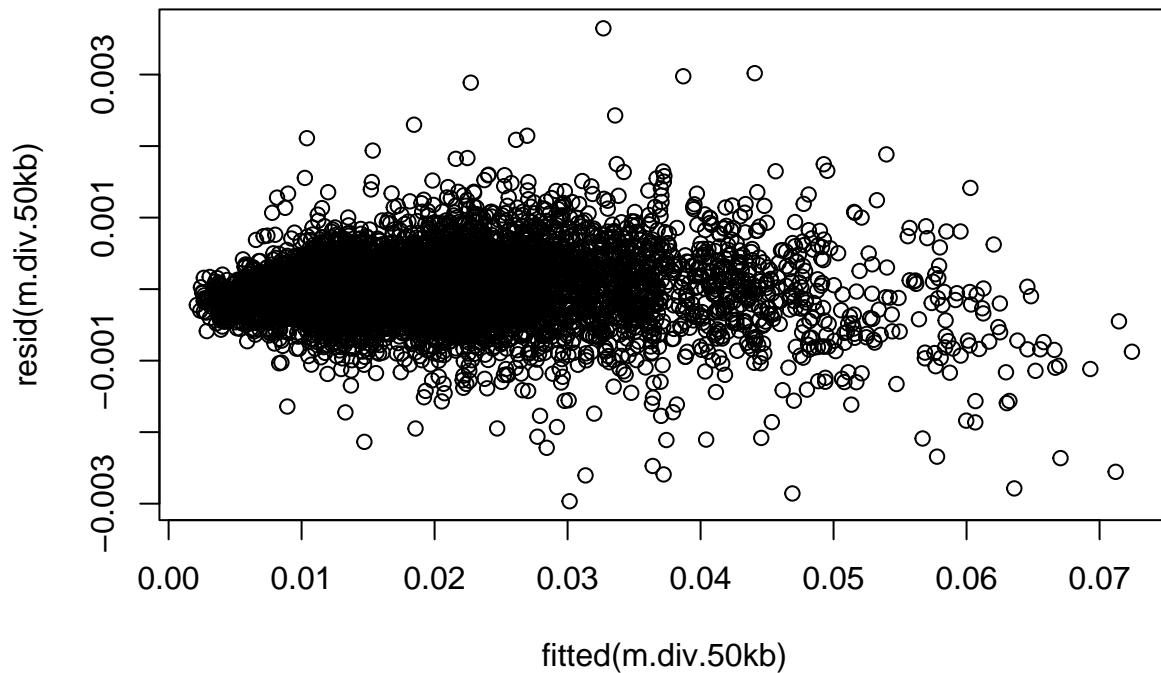
#### 1.1.11 all replicates:

```

sim.lands.50kb.all <- rbind(sim.lands.50kb.rep1, sim.lands.50kb.rep2, sim.lands.50kb.rep3, sim.lands.50kb.rep4)

m.div.50kb <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)*as.factor(Replicate), data = sim.lands.50kb.all)
plot(resid(m.div.50kb)~fitted(m.div.50kb))

```



```

dwtest(m.div.50kb)

##
## Durbin-Watson test
##
## data: m.div.50kb

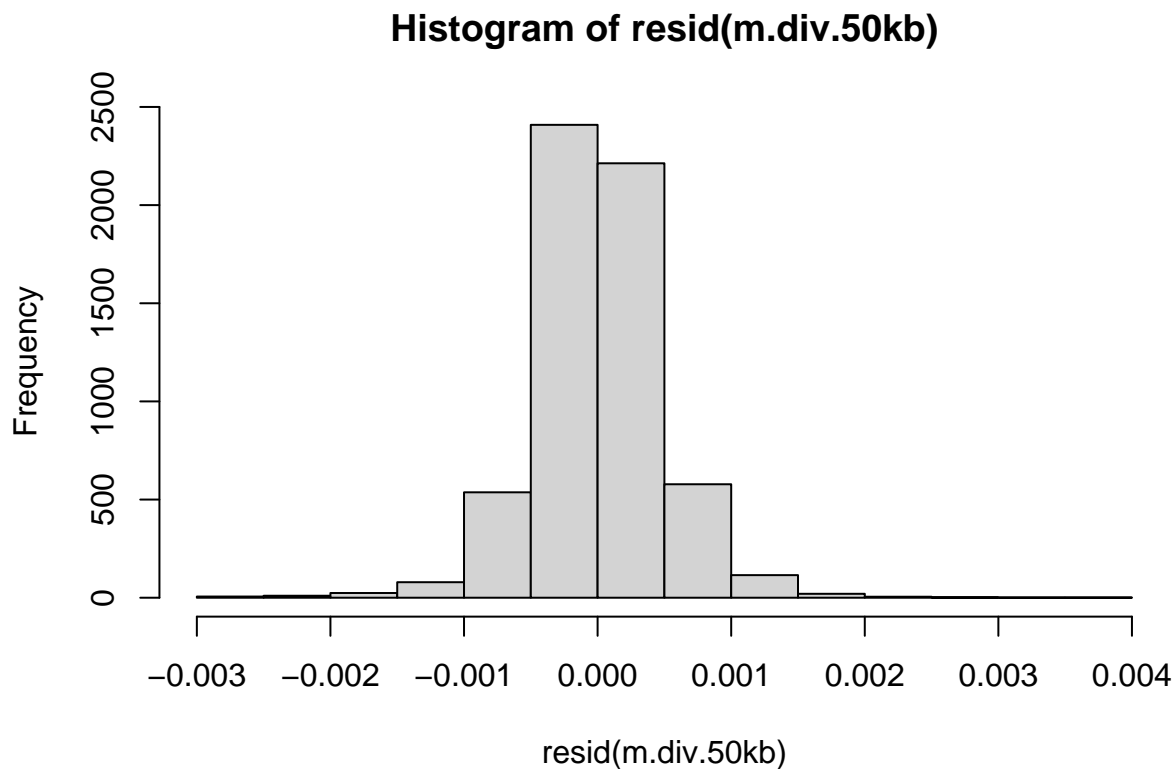
```

```
## DW = 2.0321, p-value = 0.8384
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.50kb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.4985, p-value = 0.422
```

```
hist(resid(m.div.50kb))
```



```
m.div.50kb.2 <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC)*as.factor(Replica
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC + rhoC:tmrcaC)*as.
```

```
AIC(m.div.50kb, m.div.50kb.2, m.div.50kb.3)
```

```
##           df      AIC
## m.div.50kb   51 -74511.74
## m.div.50kb.2  61 -74500.95
## m.div.50kb.3  71 -74492.20
```

```
summary(m.div.50kb)
```

```
##
## Call:
## lm(formula = diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC) *
##     as.factor(Replicate), data = sim.lands.50kb.all)
```

```

##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0029659 -0.0002709 -0.0000132  0.0002735  0.0036466
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.069e-02  1.978e-05 1045.973 < 2e-16
## thetaC           1.309e+00  2.367e-03  552.973 < 2e-16
## rhoC             1.435e-02  6.692e-03   2.145 0.032024
## tmrcaC           2.342e-02  2.784e-04  84.107 < 2e-16
## as.factor(Replicate)2 -1.161e-04  2.798e-05  -4.151 3.36e-05
## as.factor(Replicate)3 -1.905e-05  2.798e-05  -0.681 0.495974
## as.factor(Replicate)4 -1.861e-04  2.797e-05  -6.653 3.12e-11
## as.factor(Replicate)5 -7.501e-05  2.797e-05  -2.681 0.007352
## as.factor(Replicate)6 -1.078e-04  2.797e-05  -3.855 0.000117
## as.factor(Replicate)7 -7.993e-05  2.799e-05  -2.856 0.004310
## as.factor(Replicate)8 -1.182e-04  2.799e-05  -4.224 2.43e-05
## as.factor(Replicate)9 -4.323e-05  2.798e-05  -1.545 0.122345
## as.factor(Replicate)10  9.218e-06  2.800e-05   0.329 0.741996
## thetaC:tmrcaC       1.483e+00  3.140e-02  47.248 < 2e-16
## thetaC:as.factor(Replicate)2 -9.417e-03  3.347e-03  -2.813 0.004917
## thetaC:as.factor(Replicate)3  1.489e-03  3.343e-03   0.445 0.656053
## thetaC:as.factor(Replicate)4 -8.500e-03  3.341e-03  -2.544 0.010981
## thetaC:as.factor(Replicate)5 -1.050e-03  3.344e-03  -0.314 0.753425
## thetaC:as.factor(Replicate)6 -4.958e-03  3.346e-03  -1.482 0.138453
## thetaC:as.factor(Replicate)7  4.142e-03  3.344e-03   1.239 0.215471
## thetaC:as.factor(Replicate)8 -6.257e-03  3.347e-03  -1.869 0.061650
## thetaC:as.factor(Replicate)9 -8.733e-04  3.342e-03  -0.261 0.793848
## thetaC:as.factor(Replicate)10  2.079e-03  3.352e-03   0.620 0.535070
## rhoC:as.factor(Replicate)2 -2.445e-02  9.462e-03  -2.584 0.009801
## rhoC:as.factor(Replicate)3 -7.910e-03  9.458e-03  -0.836 0.403040
## rhoC:as.factor(Replicate)4 -4.870e-04  9.455e-03  -0.052 0.958925
## rhoC:as.factor(Replicate)5 -1.282e-02  9.458e-03  -1.356 0.175200
## rhoC:as.factor(Replicate)6 -2.614e-02  9.458e-03  -2.764 0.005730
## rhoC:as.factor(Replicate)7 -2.010e-02  9.454e-03  -2.127 0.033499
## rhoC:as.factor(Replicate)8 -1.522e-02  9.458e-03  -1.610 0.107517
## rhoC:as.factor(Replicate)9 -2.212e-02  9.455e-03  -2.340 0.019335
## rhoC:as.factor(Replicate)10 -6.794e-03  9.461e-03  -0.718 0.472726
## tmrcaC:as.factor(Replicate)2  4.369e-04  3.904e-04   1.119 0.263167
## tmrcaC:as.factor(Replicate)3  1.926e-04  3.696e-04   0.521 0.602310
## tmrcaC:as.factor(Replicate)4  6.863e-05  3.753e-04   0.183 0.854903
## tmrcaC:as.factor(Replicate)5  3.255e-04  3.871e-04   0.841 0.400392
## tmrcaC:as.factor(Replicate)6  2.144e-04  3.827e-04   0.560 0.575337
## tmrcaC:as.factor(Replicate)7  4.731e-04  3.804e-04   1.244 0.213699
## tmrcaC:as.factor(Replicate)8  3.839e-04  3.900e-04   0.984 0.325018
## tmrcaC:as.factor(Replicate)9  6.239e-04  3.859e-04   1.617 0.105968
## tmrcaC:as.factor(Replicate)10 3.451e-04  3.707e-04   0.931 0.351827
## thetaC:tmrcaC:as.factor(Replicate)2 3.125e-02  4.423e-02   0.707 0.479777
## thetaC:tmrcaC:as.factor(Replicate)3 1.779e-02  4.137e-02   0.430 0.667147
## thetaC:tmrcaC:as.factor(Replicate)4 9.331e-03  4.331e-02   0.215 0.829438
## thetaC:tmrcaC:as.factor(Replicate)5 -1.259e-02  4.444e-02  -0.283 0.777008
## thetaC:tmrcaC:as.factor(Replicate)6 -4.406e-02  4.363e-02  -1.010 0.312531
## thetaC:tmrcaC:as.factor(Replicate)7 7.711e-03  4.393e-02   0.176 0.860677

```

```

## thetaC:tmrcaC:as.factor(Replicate)8 -1.009e-01 4.365e-02 -2.311 0.020864
## thetaC:tmrcaC:as.factor(Replicate)9 2.500e-02 4.348e-02 0.575 0.565381
## thetaC:tmrcaC:as.factor(Replicate)10 7.066e-02 4.155e-02 1.701 0.089068
##
## (Intercept) ***
## thetaC ***
## rhoC *
## tmrcaC ***
## as.factor(Replicate)2 ***
## as.factor(Replicate)3
## as.factor(Replicate)4 ***
## as.factor(Replicate)5 **
## as.factor(Replicate)6 ***
## as.factor(Replicate)7 **
## as.factor(Replicate)8 ***
## as.factor(Replicate)9
## as.factor(Replicate)10
## thetaC:tmrcaC ***
## thetaC:as.factor(Replicate)2 **
## thetaC:as.factor(Replicate)3
## thetaC:as.factor(Replicate)4 *
## thetaC:as.factor(Replicate)5
## thetaC:as.factor(Replicate)6
## thetaC:as.factor(Replicate)7
## thetaC:as.factor(Replicate)8 .
## thetaC:as.factor(Replicate)9
## thetaC:as.factor(Replicate)10
## rhoC:as.factor(Replicate)2 **
## rhoC:as.factor(Replicate)3
## rhoC:as.factor(Replicate)4
## rhoC:as.factor(Replicate)5
## rhoC:as.factor(Replicate)6 **
## rhoC:as.factor(Replicate)7 *
## rhoC:as.factor(Replicate)8
## rhoC:as.factor(Replicate)9 *
## rhoC:as.factor(Replicate)10
## tmrcaC:as.factor(Replicate)2
## tmrcaC:as.factor(Replicate)3
## tmrcaC:as.factor(Replicate)4
## tmrcaC:as.factor(Replicate)5
## tmrcaC:as.factor(Replicate)6
## tmrcaC:as.factor(Replicate)7
## tmrcaC:as.factor(Replicate)8
## tmrcaC:as.factor(Replicate)9
## tmrcaC:as.factor(Replicate)10
## thetaC:tmrcaC:as.factor(Replicate)2
## thetaC:tmrcaC:as.factor(Replicate)3
## thetaC:tmrcaC:as.factor(Replicate)4
## thetaC:tmrcaC:as.factor(Replicate)5
## thetaC:tmrcaC:as.factor(Replicate)6
## thetaC:tmrcaC:as.factor(Replicate)7
## thetaC:tmrcaC:as.factor(Replicate)8 *
## thetaC:tmrcaC:as.factor(Replicate)9
## thetaC:tmrcaC:as.factor(Replicate)10 .

```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0004844 on 5950 degrees of freedom
## Multiple R-squared:  0.9981, Adjusted R-squared:  0.9981
## F-statistic: 6.421e+04 on 49 and 5950 DF,  p-value: < 2.2e-16
```

## 1.2 200 kb scale

```
r2.sim.200kb <- as.data.frame(matrix(ncol = nreps, nrow = 5))
row.names(r2.sim.200kb) <- c("Total", "Theta", "Rho", "TMRCA", "Theta:TMRCA")
colnames(r2.sim.200kb) <- reps

sim.rho.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/dm.sim.rho.2e+05.txt", header = T)
sim.theta.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/dm.sim.theta.2e+05.txt", header = T)
```

### 1.2.1 Replicate 1

```
rep_1.pi.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_1/rep_1.pi.200kb.bedgraph", header = T)
rep_1.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_1/sim.tmrca.200k.map", header = T)

sim.lands.200kb <- as.data.frame(cbind(rep_1.pi.200kb$pi, sim.theta.200kb$sim, sim.rho.200kb$sim, rep_1.tmrca.200kb$tmrca))
names(sim.lands.200kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.200kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data:  theta and tmrca
## S = 569023, p-value = 0.8876
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.01164184
```

```
cor.test(~theta+rho, data = sim.lands.200kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data:  theta and rho
## S = 593676, p-value = 0.5002
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.0554712
```

```
cor.test(~rho+tmrca, data = sim.lands.200kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data:  rho and tmrca
## S = 570624, p-value = 0.8602
```



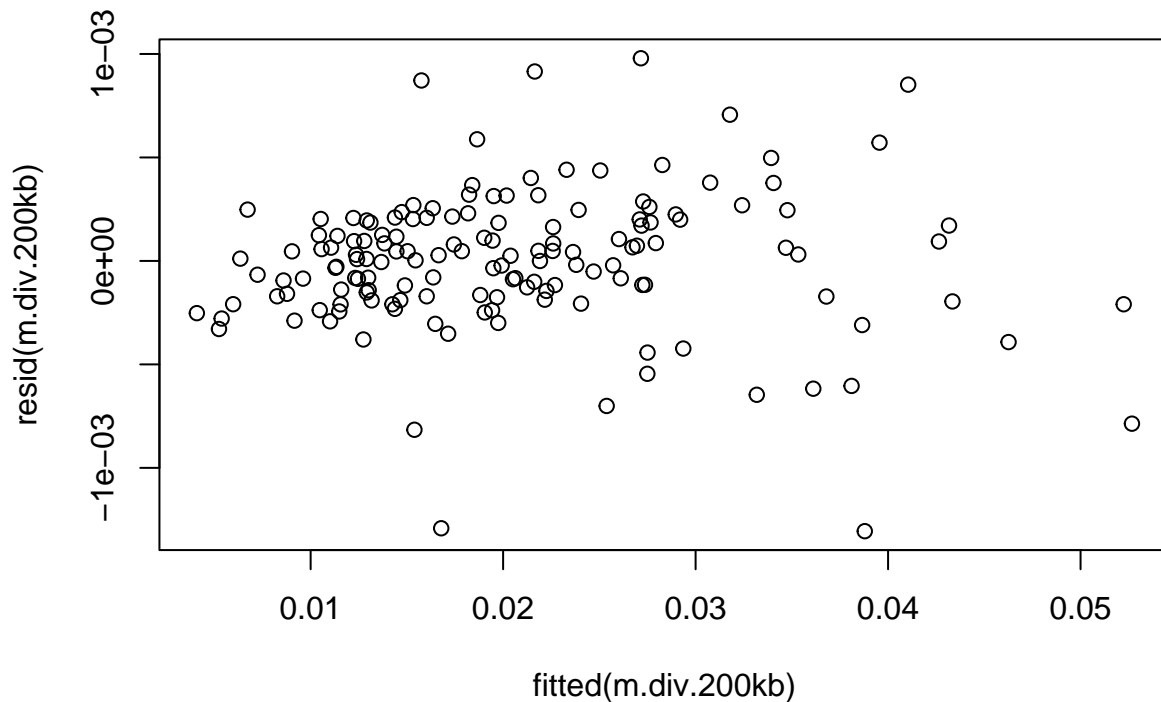
```
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.01448776

# centering
sim.lands.200kb$thetaC <- sim.lands.200kb$theta - mean(sim.lands.200kb$theta)
sim.lands.200kb$tmrcaC <- sim.lands.200kb$tmrca - mean(sim.lands.200kb$tmrca)
sim.lands.200kb$rhoC <- sim.lands.200kb$rho - mean(sim.lands.200kb$rho)

sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)

# for merging:
sim.lands.200kb.rep1 <- sim.lands.200kb
sim.lands.200kb.rep1$Replicate <- 1

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.200kb)
plot(resid(m.div.200kb)~fitted(m.div.200kb))
```



```
dwtest(m.div.200kb)

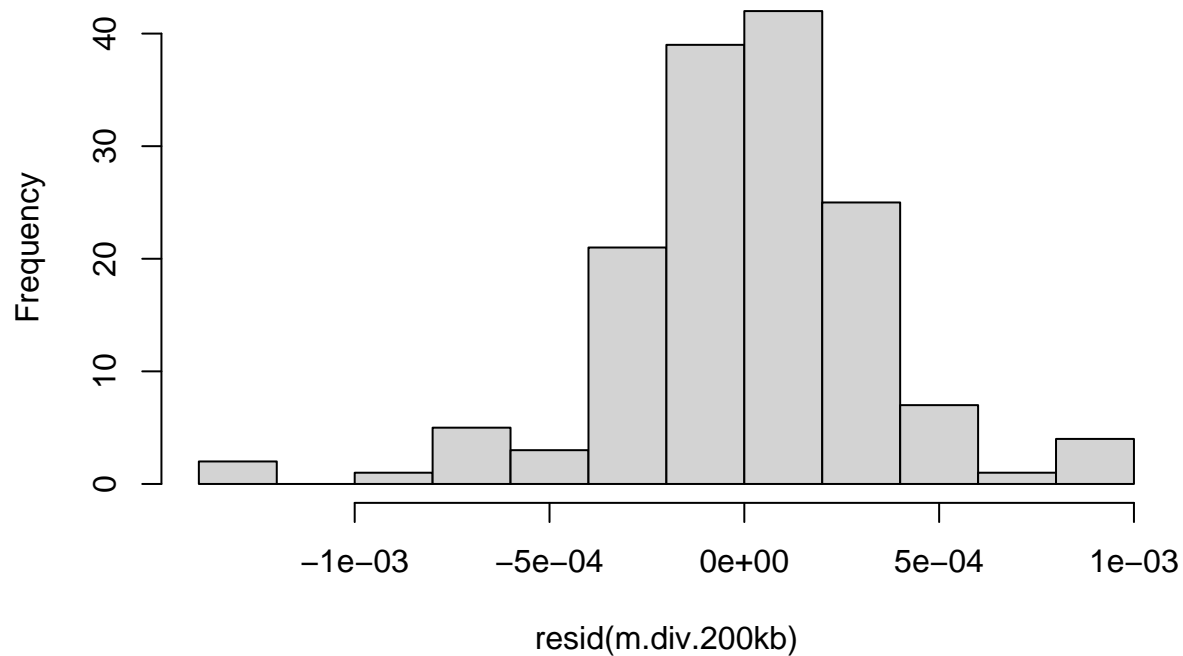
##
## Durbin-Watson test
##
## data: m.div.200kb
## DW = 1.9751, p-value = 0.4234
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.200kb)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.div.200kb  
## HMC = 0.49982, p-value = 0.5
```

```
hist(resid(m.div.200kb))
```

### Histogram of resid(m.div.200kb)



```
summary(m.div.200kb)
```

```
##  
## Call:  
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,  
##     data = sim.lands.200kb)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -1.306e-03 -1.718e-04  1.801e-05  1.985e-04  9.793e-04   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)  2.067e-02  2.788e-05  741.370  <2e-16 ***  
## thetaC       1.306e+00  3.769e-03  346.532  <2e-16 ***  
## rhoC         3.201e-02  1.870e-02   1.712    0.089 .   
## tmrcaC       2.526e-02  7.654e-04  32.997  <2e-16 ***
```

```
## thetaC:tmrcaC 1.536e+00 1.072e-01 14.325 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0003415 on 145 degrees of freedom
## Multiple R-squared: 0.9988, Adjusted R-squared: 0.9988
## F-statistic: 3.049e+04 on 4 and 145 DF, p-value: < 2.2e-16

anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.200kb[1, 1] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) * 100
r2.sim.200kb[2, 1] <- anova.diversity$VarExp[1] * 100
r2.sim.200kb[3, 1] <- anova.diversity$VarExp[2] * 100
r2.sim.200kb[4, 1] <- anova.diversity$VarExp[3] * 100
r2.sim.200kb[5, 1] <- anova.diversity$VarExp[4] * 100
```

### 1.2.2 Replicate 2

```
rep_2.pi.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_2/rep_2.pi.200kb.bedgraph", header=TRUE, as.is=TRUE)
rep_2.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_2/sim.tmrca.200k.map", header=TRUE, as.is=TRUE)

sim.lands.200kb <- as.data.frame(cbind(rep_2.pi.200kb$pi, sim.theta.200kb$sim, sim.rho.200kb$sim, rep_2.tmrca.200kb$tmrca))
names(sim.lands.200kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 551432, p-value = 0.8115
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.01963361

cor.test(~theta+rho, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 593676, p-value = 0.5002
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.0554712

cor.test(~rho+tmrca, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: rho and tmrca
```

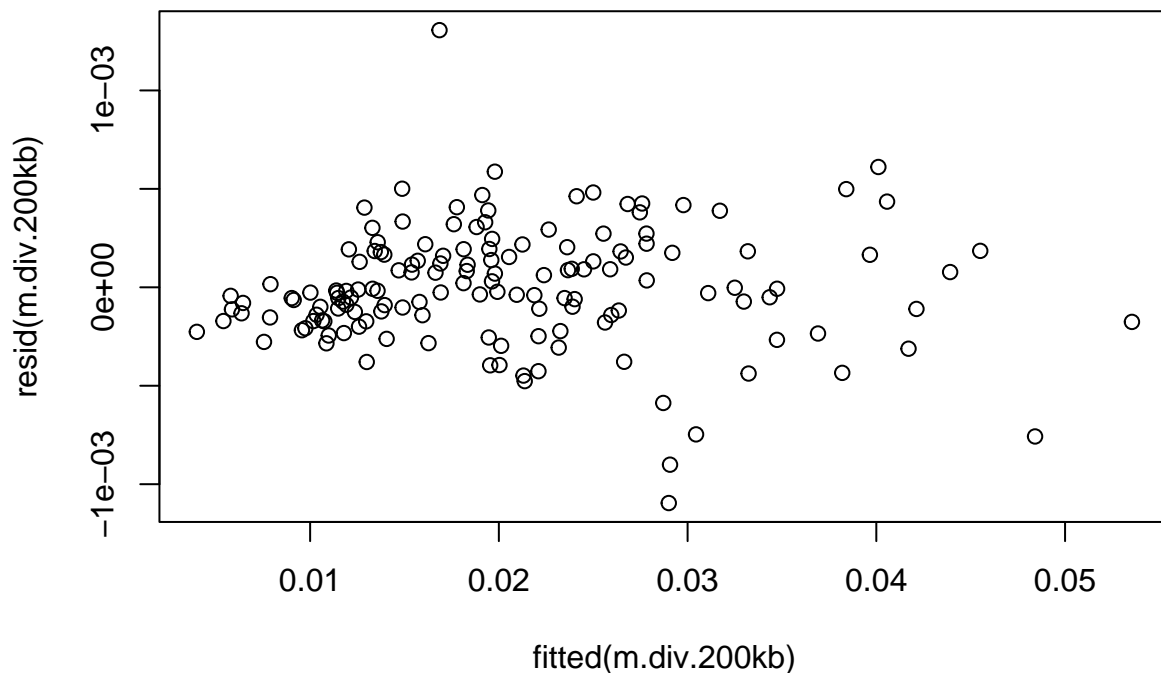
```
## S = 533372, p-value = 0.529
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.05174097

# centering
sim.lands.200kb$thetaC <- sim.lands.200kb$theta - mean(sim.lands.200kb$theta)
sim.lands.200kb$tmrcaC <- sim.lands.200kb$tmrca - mean(sim.lands.200kb$tmrca)
sim.lands.200kb$rhoC <- sim.lands.200kb$rho - mean(sim.lands.200kb$rho)

sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)

# for merging:
sim.lands.200kb.rep2 <- sim.lands.200kb
sim.lands.200kb.rep2$Replicate <- 2

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.200kb)
plot(resid(m.div.200kb)~fitted(m.div.200kb))
```



```
dwtest(m.div.200kb)

##
## Durbin-Watson test
##
## data: m.div.200kb
## DW = 1.5928, p-value = 0.005174
```

```
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.200kb)
```

```
##
```

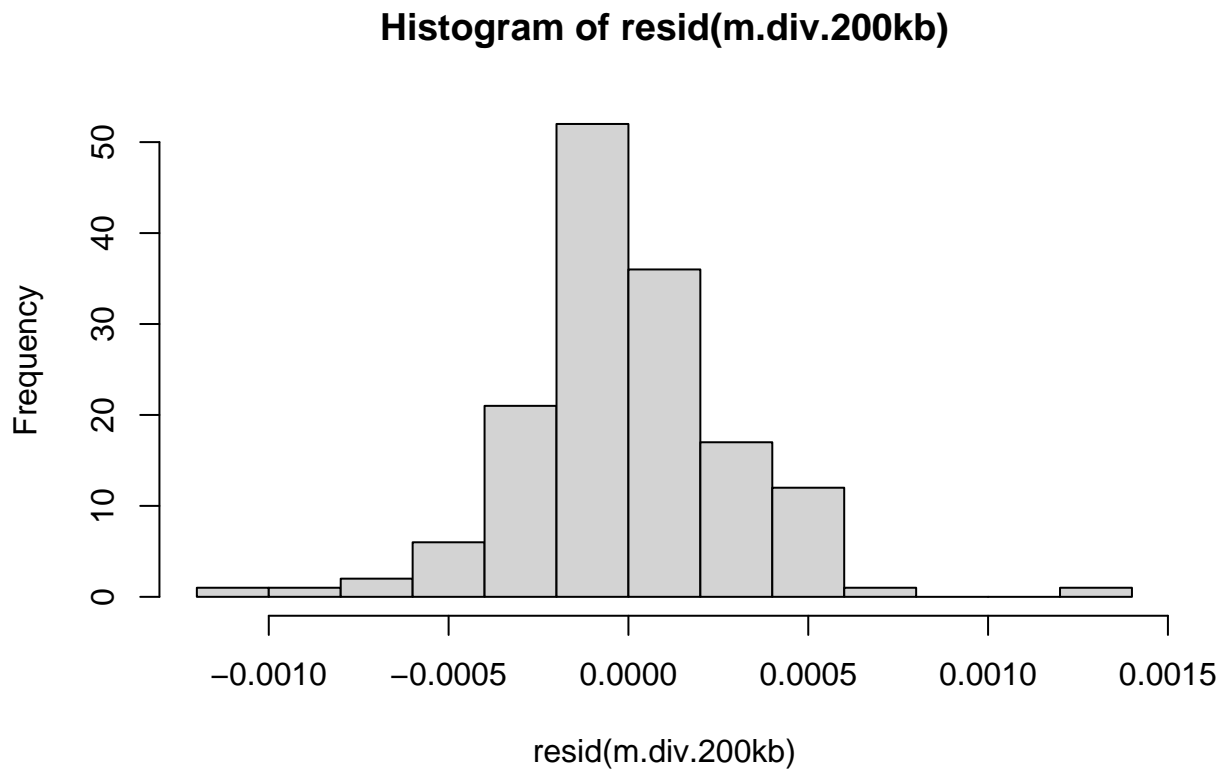
```
## Harrison-McCabe test
```

```
##
```

```
## data: m.div.200kb
```

```
## HMC = 0.59119, p-value = 0.934
```

```
hist(resid(m.div.200kb))
```



```
summary(m.div.200kb)
```

```
##
```

```
## Call:
```

```
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
```

```
## data = sim.lands.200kb)
```

```
##
```

```
## Residuals:
```

	Min	1Q	Median	3Q	Max
##	-1.096e-03	-1.706e-04	-2.456e-05	1.813e-04	1.306e-03

```
##
```

```
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	2.056e-02	2.494e-05	824.14	<2e-16 ***
## thetaC	1.297e+00	3.394e-03	382.15	<2e-16 ***
## rhoC	-3.118e-02	1.667e-02	-1.87	0.0635 .

```
## tmrcaC          2.455e-02  6.157e-04  39.88   <2e-16 ***
## thetaC:tmrcaC  1.643e+00  8.199e-02  20.04   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0003054 on 145 degrees of freedom
## Multiple R-squared:  0.999, Adjusted R-squared:  0.999
## F-statistic: 3.71e+04 on 4 and 145 DF, p-value: < 2.2e-16

anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.200kb[1, 2] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.sim.200kb[2, 2] <- anova.diversity$VarExp[1] * 100
r2.sim.200kb[3, 2] <- anova.diversity$VarExp[2] * 100
r2.sim.200kb[4, 2] <- anova.diversity$VarExp[3] * 100
r2.sim.200kb[5, 2] <- anova.diversity$VarExp[4] * 100
```

### 1.2.3 Replicate 3

```
rep_3.pi.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_3/rep_3.pi.200kb.bedgraph", head=1)
rep_3.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_3/sim.tmrca.200k.map", header=1)

sim.lands.200kb <- as.data.frame(cbind(rep_3.pi.200kb$pi, sim.theta.200kb$sim, sim.rho.200kb$sim, rep_3.tmrca.200kb$tmrca))
names(sim.lands.200kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  theta and tmrca
## S = 592174, p-value = 0.5211
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.05280076

cor.test(~theta+rho, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  theta and rho
## S = 593676, p-value = 0.5002
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.0554712

cor.test(~rho+tmrca, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
```

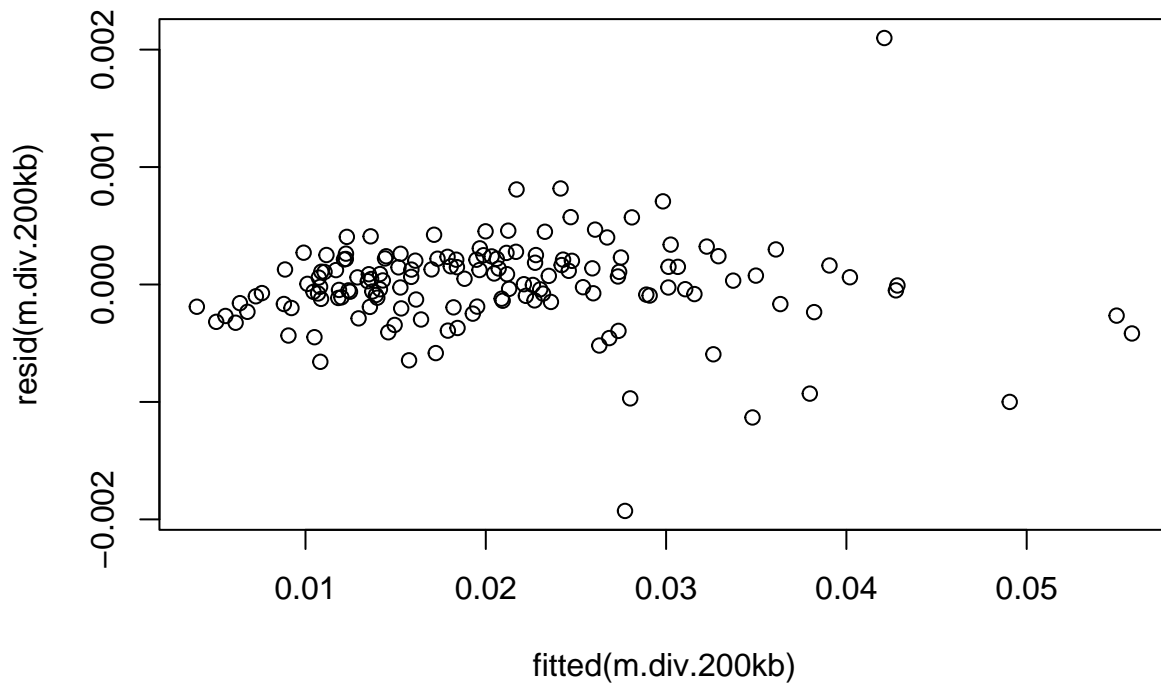
```
## data: rho and tmrca
## S = 594376, p-value = 0.4902
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.05671541

# centering
sim.lands.200kb$thetaC <- sim.lands.200kb$theta - mean(sim.lands.200kb$theta)
sim.lands.200kb$tmrcaC <- sim.lands.200kb$tmrca - mean(sim.lands.200kb$tmrca)
sim.lands.200kb$rhoC <- sim.lands.200kb$rho - mean(sim.lands.200kb$rho)

sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)

# for merging:
sim.lands.200kb.rep3 <- sim.lands.200kb
sim.lands.200kb.rep3$Replicate <- 3

m.div.200kb <- lm(多样性 ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.200kb)
plot(resid(m.div.200kb)~fitted(m.div.200kb))
```



```
dwtest(m.div.200kb)

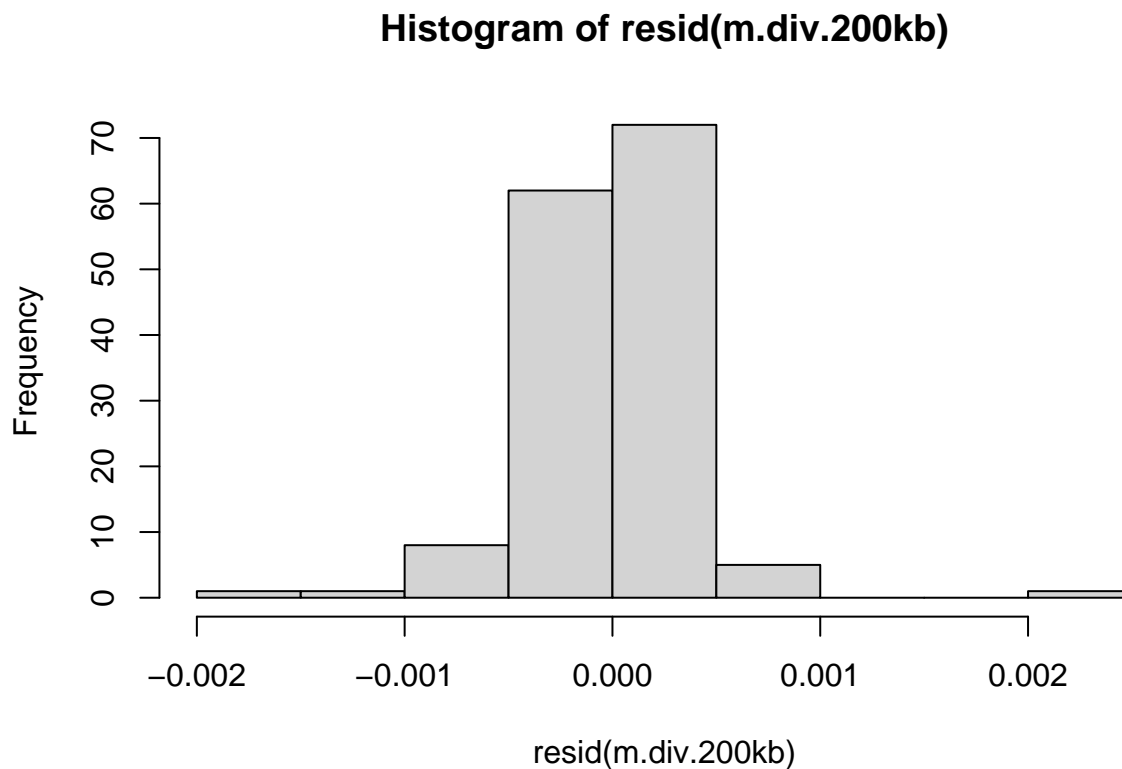
##
## Durbin-Watson test
##
## data: m.div.200kb
```

```
## DW = 1.9205, p-value = 0.2965
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.200kb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.45845, p-value = 0.22
```

```
hist(resid(m.div.200kb))
```



```
summary(m.div.200kb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.200kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.928e-03 -1.460e-04  3.044e-05  2.110e-04  2.099e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.064e-02  3.245e-05  636.263  <2e-16 ***
## thetaC       1.309e+00  4.394e-03  297.825  <2e-16 ***
```



```
## rhoC          3.167e-02  2.173e-02  1.458    0.147
## tmrcaC        2.488e-02  7.859e-04  31.662   <2e-16 ***
## thetaC:tmrcaC 1.472e+00  1.097e-01  13.422   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0003972 on 145 degrees of freedom
## Multiple R-squared:  0.9984, Adjusted R-squared:  0.9984
## F-statistic: 2.269e+04 on 4 and 145 DF,  p-value: < 2.2e-16

anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.200kb[1, 3] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.sim.200kb[2, 3] <- anova.diversity$VarExp[1] * 100
r2.sim.200kb[3, 3] <- anova.diversity$VarExp[2] * 100
r2.sim.200kb[4, 3] <- anova.diversity$VarExp[3] * 100
r2.sim.200kb[5, 3] <- anova.diversity$VarExp[4] * 100
```

#### 1.2.4 Replicate 4

```
rep_4.pi.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_4/rep_4.pi.200kb.bedgraph", header=TRUE)
rep_4.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_4/sim.tmrca.200k.map", header=TRUE)

sim.lands.200kb <- as.data.frame(cbind(rep_4.pi.200kb$pi, sim.theta.200kb$sim, sim.rho.200kb$sim, rep_4.tmrca.200kb$tmrca))
names(sim.lands.200kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  theta and tmrca
## S = 554491, p-value = 0.8631
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.01419494

cor.test(~theta+rho, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  theta and rho
## S = 593676, p-value = 0.5002
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.0554712

cor.test(~rho+tmrca, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
```

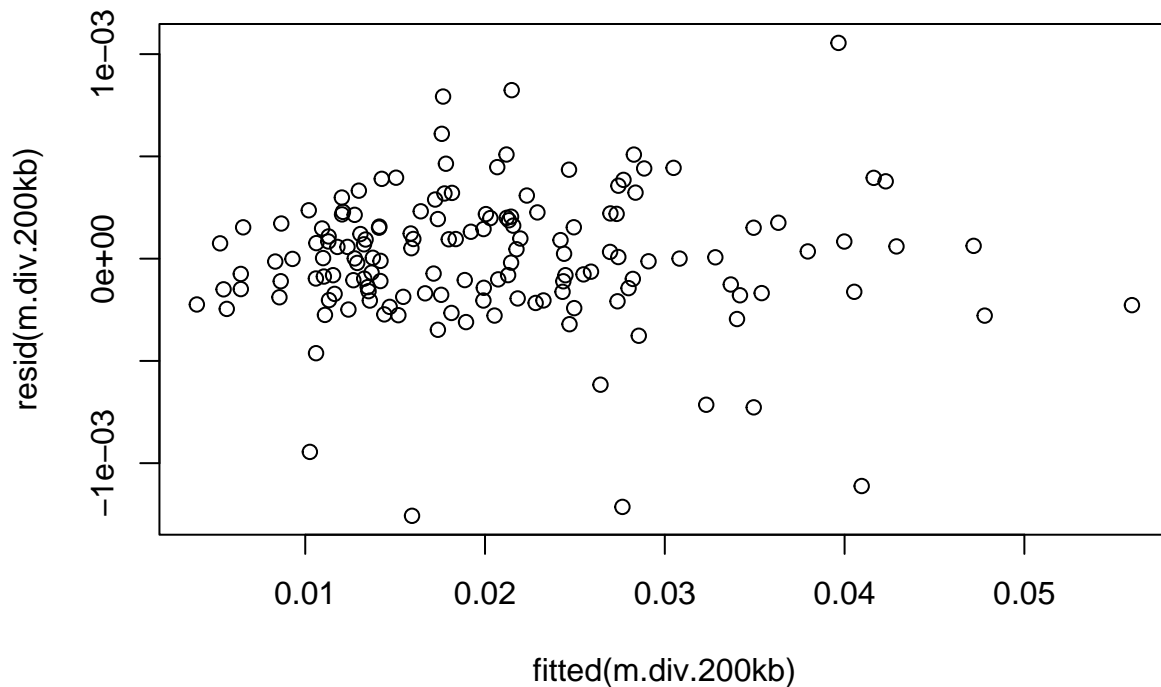
```
##
## data: rho and tmrca
## S = 559798, p-value = 0.9539
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.004759323

# centering
sim.lands.200kb$thetaC <- sim.lands.200kb$theta - mean(sim.lands.200kb$theta)
sim.lands.200kb$tmrcaC <- sim.lands.200kb$tmrca - mean(sim.lands.200kb$tmrca)
sim.lands.200kb$rhoC <- sim.lands.200kb$rho - mean(sim.lands.200kb$rho)

sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)

# for merging:
sim.lands.200kb.rep4 <- sim.lands.200kb
sim.lands.200kb.rep4$Replicate <- 4

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.200kb)
plot(resid(m.div.200kb)~fitted(m.div.200kb))
```



```
dwtest(m.div.200kb)
```

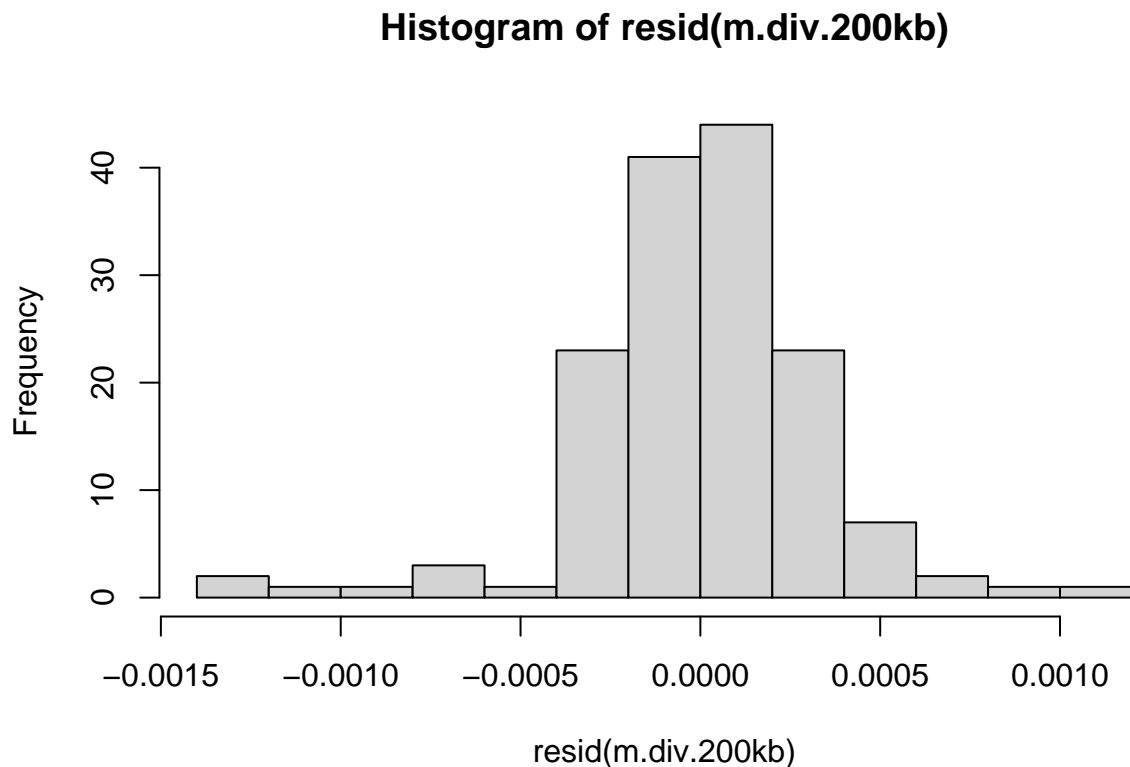
```
##
## Durbin-Watson test
##
```

```
## data: m.div.200kb
## DW = 1.9887, p-value = 0.4477
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.200kb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.4058, p-value = 0.048
```

```
hist(resid(m.div.200kb))
```



```
summary(m.div.200kb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.200kb)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-1.257e-03	-1.693e-04	3.530e-06	1.847e-04	1.055e-03

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.0205018	0.0000271	756.559	<2e-16 ***

```
## thetaC      1.3010756  0.0036626 355.236 <2e-16 ***
## rhoC        0.0087841  0.0181026  0.485  0.628
## tmrcaC      0.0244436  0.0006814 35.873 <2e-16 ***
## thetaC:tmrcaC 1.4688735  0.0873643 16.813 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0003317 on 145 degrees of freedom
## Multiple R-squared:  0.9989, Adjusted R-squared:  0.9989
## F-statistic: 3.248e+04 on 4 and 145 DF,  p-value: < 2.2e-16

anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.200kb[1, 4] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.sim.200kb[2, 4] <- anova.diversity$VarExp[1] * 100
r2.sim.200kb[3, 4] <- anova.diversity$VarExp[2] * 100
r2.sim.200kb[4, 4] <- anova.diversity$VarExp[3] * 100
r2.sim.200kb[5, 4] <- anova.diversity$VarExp[4] * 100
```

### 1.2.5 Replicate 5

```
rep_5.pi.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_5/rep_5.pi.200kb.bedgraph", head=
rep_5.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_5/sim.tmrca.200k.map", header=

sim.lands.200kb <- as.data.frame(cbind(rep_5.pi.200kb$pi, sim.theta.200kb$sim, sim.rho.200kb$sim, rep_5
names(sim.lands.200kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  theta and tmrca
## S = 570758, p-value = 0.858
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.01472654

cor.test(~theta+rho, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  theta and rho
## S = 593676, p-value = 0.5002
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.0554712

cor.test(~rho+tmrca, data = sim.lands.200kb, method = "spearman")

##
```

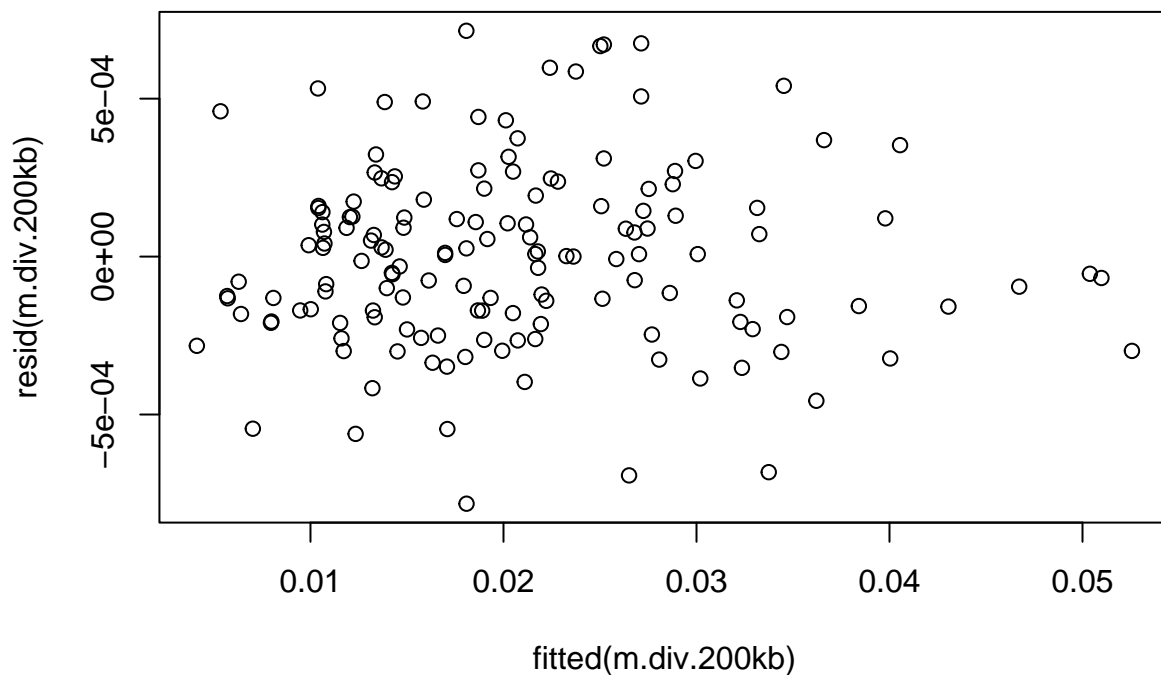
```
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 582370, p-value = 0.6671
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.03537046

# centering
sim.lands.200kb$thetaC <- sim.lands.200kb$theta - mean(sim.lands.200kb$theta)
sim.lands.200kb$tmrcaC <- sim.lands.200kb$tmrca - mean(sim.lands.200kb$tmrca)
sim.lands.200kb$rhoC <- sim.lands.200kb$rho - mean(sim.lands.200kb$rho)

sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)

# for merging:
sim.lands.200kb.rep5 <- sim.lands.200kb
sim.lands.200kb.rep5$Replicate <- 5

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.200kb)
plot(resid(m.div.200kb)~fitted(m.div.200kb))
```



```
dwtest(m.div.200kb)
```

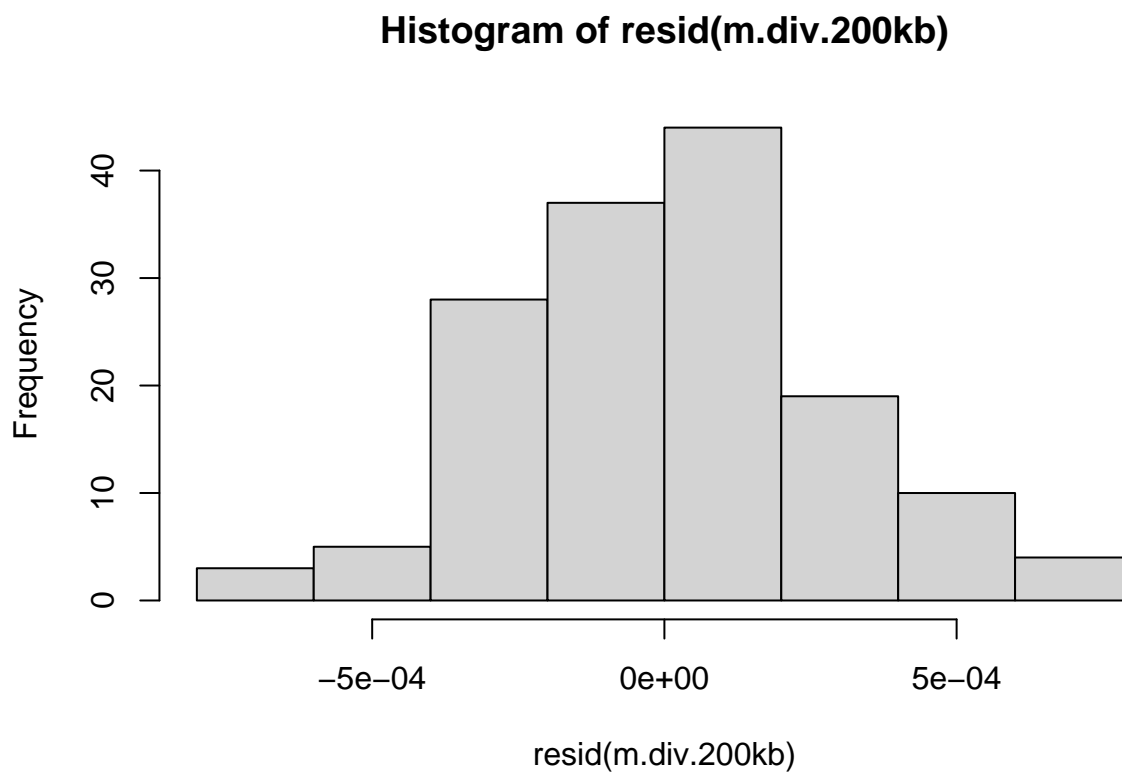
```
##
## Durbin-Watson test
```

```
##
## data: m.div.200kb
## DW = 1.9065, p-value = 0.2691
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.200kb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.42468, p-value = 0.093
```

```
hist(resid(m.div.200kb))
```



```
summary(m.div.200kb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.200kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.823e-04 -1.890e-04  3.370e-06  1.582e-04  7.148e-04
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)    2.061e-02  2.367e-05  870.811    <2e-16 ***
## thetaC        1.306e+00  3.205e-03  407.532    <2e-16 ***
## rhoC          8.924e-03  1.586e-02   0.563      0.575
## tmrcaC        2.402e-02  6.327e-04  37.957     <2e-16 ***
## thetaC:tmrcaC 1.380e+00  7.367e-02  18.728     <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0002897 on 145 degrees of freedom
## Multiple R-squared:  0.9991, Adjusted R-squared:  0.9991
## F-statistic: 4.197e+04 on 4 and 145 DF,  p-value: < 2.2e-16

anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.200kb[1, 5] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.sim.200kb[2, 5] <- anova.diversity$VarExp[1] * 100
r2.sim.200kb[3, 5] <- anova.diversity$VarExp[2] * 100
r2.sim.200kb[4, 5] <- anova.diversity$VarExp[3] * 100
r2.sim.200kb[5, 5] <- anova.diversity$VarExp[4] * 100
```

### 1.2.6 Replicate 6

```
rep_6.pi.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_6/rep_6.pi.200kb.bedgraph", head=1)
rep_6.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_6/sim.tmrca.200k.map", header=1)

sim.lands.200kb <- as.data.frame(cbind(rep_6.pi.200kb$pi, sim.theta.200kb$sim, sim.rho.200kb$sim, rep_6.tmrca.200kb$tmrca))
names(sim.lands.200kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  theta and tmrca
## S = 577585, p-value = 0.7442
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##          rho
## -0.02686264

cor.test(~theta+rho, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  theta and rho
## S = 593676, p-value = 0.5002
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##          rho
## -0.0554712
```

```
cor.test(~rho+tmrca, data = sim.lands.200kb, method = "spearman")

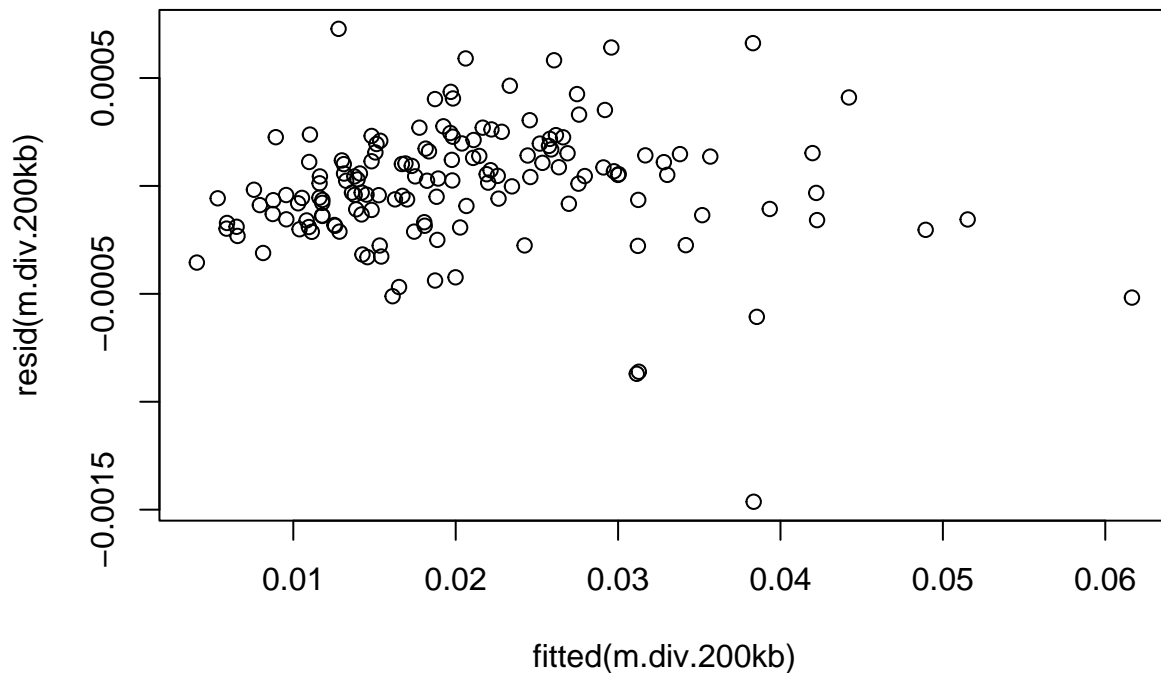
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 606554, p-value = 0.3401
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.07836615

# centering
sim.lands.200kb$thetaC <- sim.lands.200kb$theta - mean(sim.lands.200kb$theta)
sim.lands.200kb$tmrcaC <- sim.lands.200kb$tmrca - mean(sim.lands.200kb$tmrca)
sim.lands.200kb$rhoC <- sim.lands.200kb$rho - mean(sim.lands.200kb$rho)

sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)

# for merging:
sim.lands.200kb.rep6 <- sim.lands.200kb
sim.lands.200kb.rep6$Replicate <- 6

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.200kb)
plot(resid(m.div.200kb)~fitted(m.div.200kb))
```





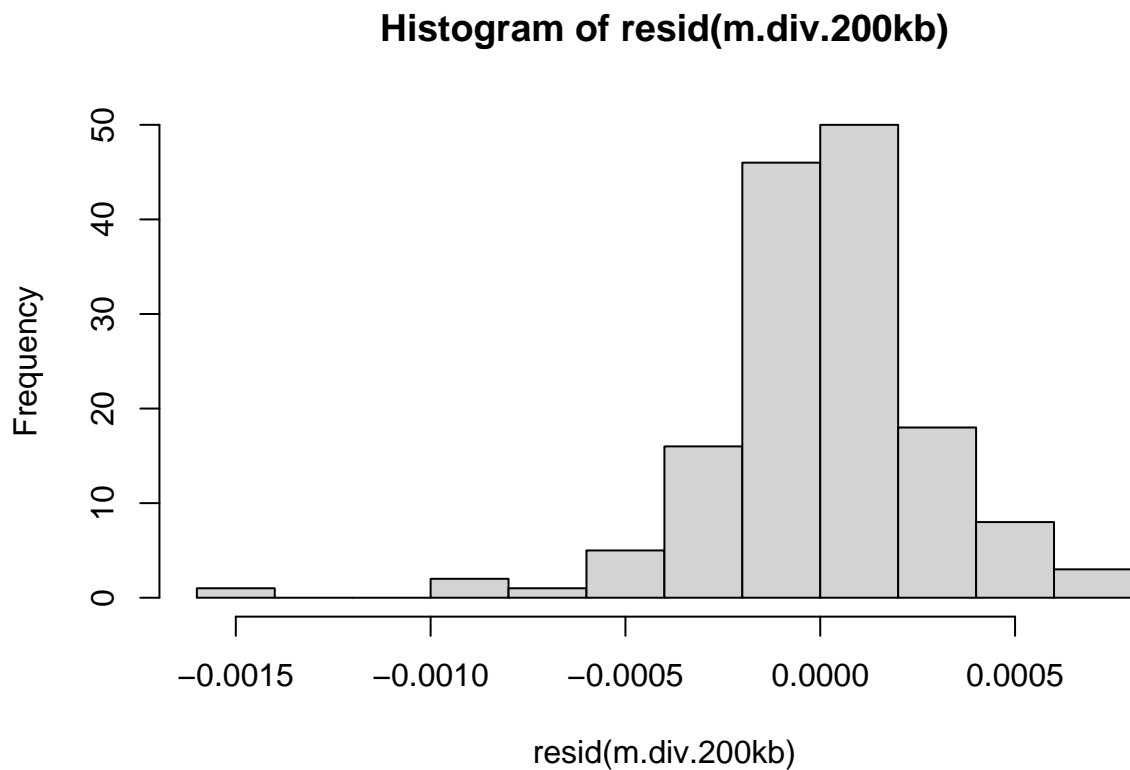
```
dwtest(m.div.200kb)
```

```
##  
## Durbin-Watson test  
##  
## data: m.div.200kb  
## DW = 2.0877, p-value = 0.6996  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.200kb)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.div.200kb  
## HMC = 0.59616, p-value = 0.952
```

```
hist(resid(m.div.200kb))
```



```
summary(m.div.200kb)
```

```
##  
## Call:  
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,  
##     data = sim.lands.200kb)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max
```

```
## -1.463e-03 -1.507e-04 2.405e-05 1.519e-04 7.279e-04
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.055e-02  2.337e-05  879.460   <2e-16 ***
## thetaC       1.299e+00  3.288e-03  395.145   <2e-16 ***
## rhoC         3.569e-03  1.559e-02   0.229    0.819
## tmrcaC       2.392e-02  6.333e-04  37.771   <2e-16 ***
## thetaC:tmrcaC 1.533e+00  6.794e-02  22.563   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.000286 on 145 degrees of freedom
## Multiple R-squared:  0.9992, Adjusted R-squared:  0.9992
## F-statistic: 4.499e+04 on 4 and 145 DF, p-value: < 2.2e-16
```

```
anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.200kb[1, 6] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.sim.200kb[2, 6] <- anova.diversity$VarExp[1] * 100
r2.sim.200kb[3, 6] <- anova.diversity$VarExp[2] * 100
r2.sim.200kb[4, 6] <- anova.diversity$VarExp[3] * 100
r2.sim.200kb[5, 6] <- anova.diversity$VarExp[4] * 100
```

### 1.2.7 Replicate 7

```
rep_7.pi.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_7/rep_7.pi.200kb.bedgraph", head=1)
rep_7.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_7/sim.tmrca.200k.map", header=1)

sim.lands.200kb <- as.data.frame(cbind(rep_7.pi.200kb$pi, sim.theta.200kb$sim, sim.rho.200kb$sim, rep_7.tmrca.200kb$tmrca))
names(sim.lands.200kb) <- c("diversity", "theta", "rho", "tmrca")
```

```
cor.test(~theta+tmrca, data = sim.lands.200kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 647408, p-value = 0.06511
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.150999
```

```
cor.test(~theta+rho, data = sim.lands.200kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 593676, p-value = 0.5002
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
```

```
##          rho
## -0.0554712
cor.test(~rho+tmrca, data = sim.lands.200kb, method = "spearman")

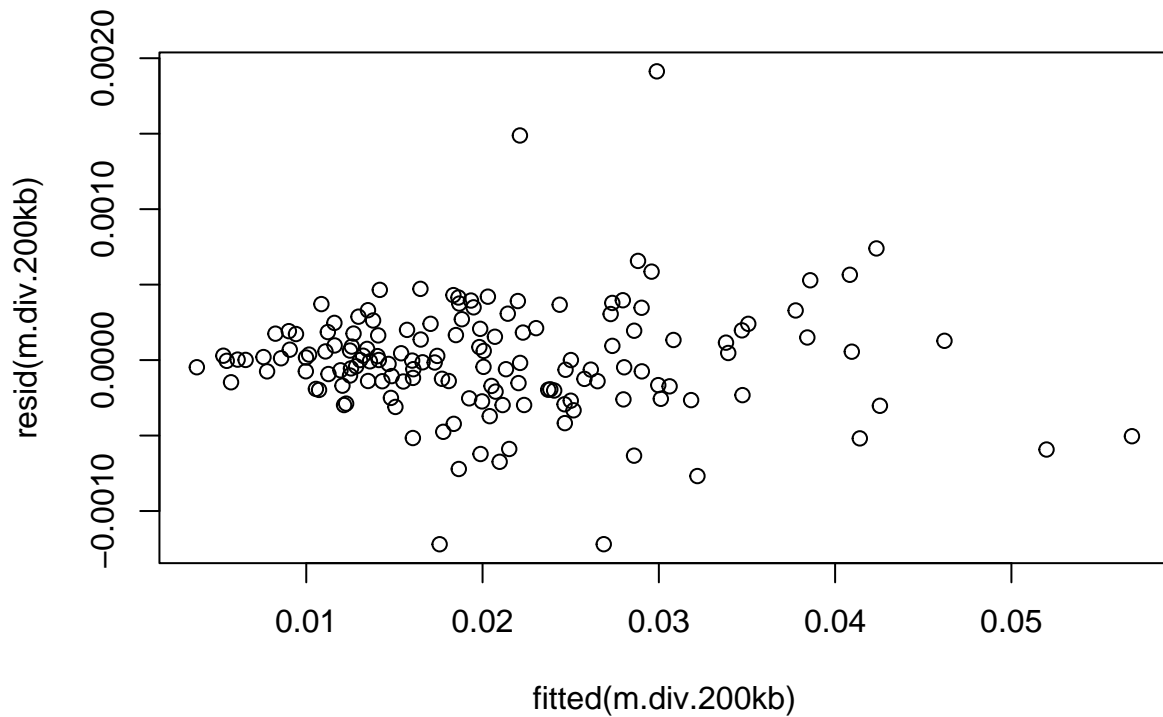
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 523746, p-value = 0.402
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##          rho
## 0.06885462

# centering
sim.lands.200kb$thetaC <- sim.lands.200kb$theta - mean(sim.lands.200kb$theta)
sim.lands.200kb$tmrcaC <- sim.lands.200kb$tmrca - mean(sim.lands.200kb$tmrca)
sim.lands.200kb$rhoC <- sim.lands.200kb$rho - mean(sim.lands.200kb$rho)

sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)

# for merging:
sim.lands.200kb.rep7 <- sim.lands.200kb
sim.lands.200kb.rep7$Replicate <- 7

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.200kb)
plot(resid(m.div.200kb)~fitted(m.div.200kb))
```



```
dwtest(m.div.200kb)
```

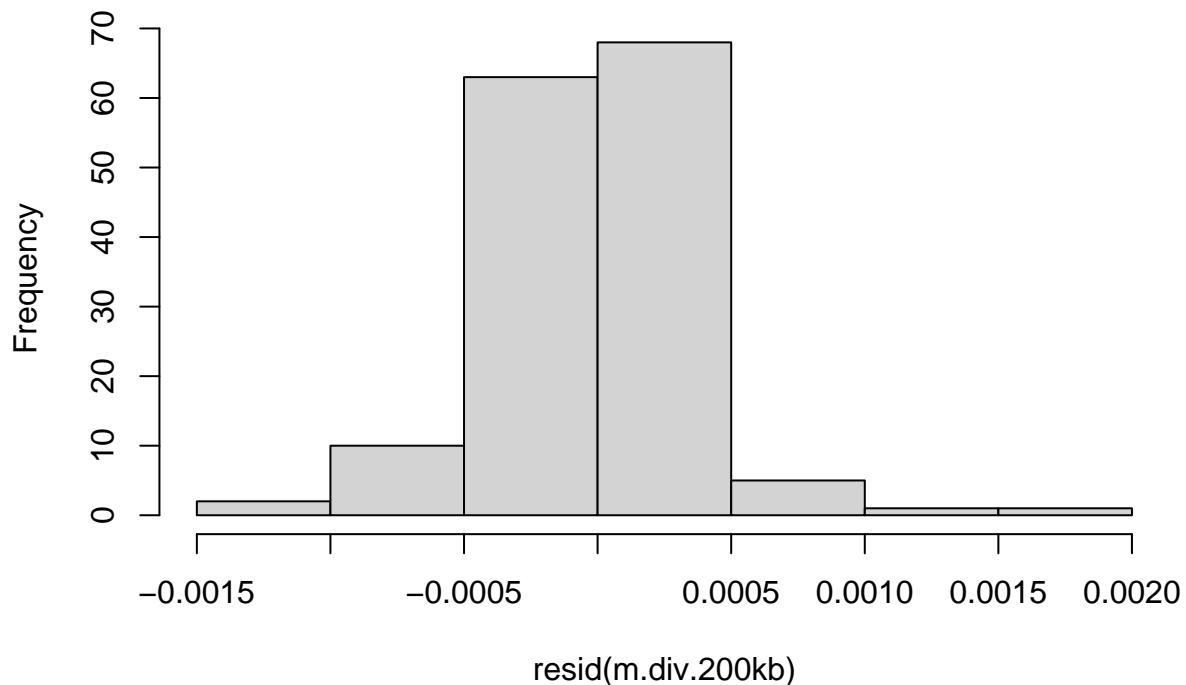
```
##  
## Durbin-Watson test  
##  
## data: m.div.200kb  
## DW = 1.9801, p-value = 0.4361  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.div.200kb)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.div.200kb  
## HMC = 0.40495, p-value = 0.043
```

```
hist(resid(m.div.200kb))
```

## Histogram of resid(m.div.200kb)



```
summary(m.div.200kb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.200kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.220e-03 -1.868e-04 -1.610e-06  1.848e-04  1.913e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.062e-02  3.141e-05  656.642  <2e-16 ***
## thetaC        1.315e+00  4.235e-03  310.525  <2e-16 ***
## rhoC         -1.620e-02  2.076e-02  -0.781    0.436
## tmrcaC        2.369e-02  8.316e-04  28.486  <2e-16 ***
## thetaC:tmrcaC  1.422e+00  1.203e-01  11.823  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0003796 on 145 degrees of freedom
## Multiple R-squared:  0.9985, Adjusted R-squared:  0.9985
## F-statistic: 2.432e+04 on 4 and 145 DF, p-value: < 2.2e-16

anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
```

```
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.200kb[1, 7] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.sim.200kb[2, 7] <- anova.diversity$VarExp[1] * 100
r2.sim.200kb[3, 7] <- anova.diversity$VarExp[2] * 100
r2.sim.200kb[4, 7] <- anova.diversity$VarExp[3] * 100
r2.sim.200kb[5, 7] <- anova.diversity$VarExp[4] * 100
```

### 1.2.8 Replicate 8

```
rep_8.pi.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_8/rep_8.pi.200kb.bedgraph", head
rep_8.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_8/sim.tmrca.200k.map", header

sim.lands.200kb <- as.data.frame(cbind(rep_8.pi.200kb$pi, sim.theta.200kb$sim, sim.rho.200kb$sim, rep_8
names(sim.lands.200kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 588671, p-value = 0.5714
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.0465727

cor.test(~theta+rho, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 593676, p-value = 0.5002
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.0554712

cor.test(~rho+tmrca, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 624456, p-value = 0.1793
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.1101933

# centering
sim.lands.200kb$thetaC <- sim.lands.200kb$theta - mean(sim.lands.200kb$theta)
sim.lands.200kb$tmrcaC <- sim.lands.200kb$tmrca - mean(sim.lands.200kb$tmrca)
```

```

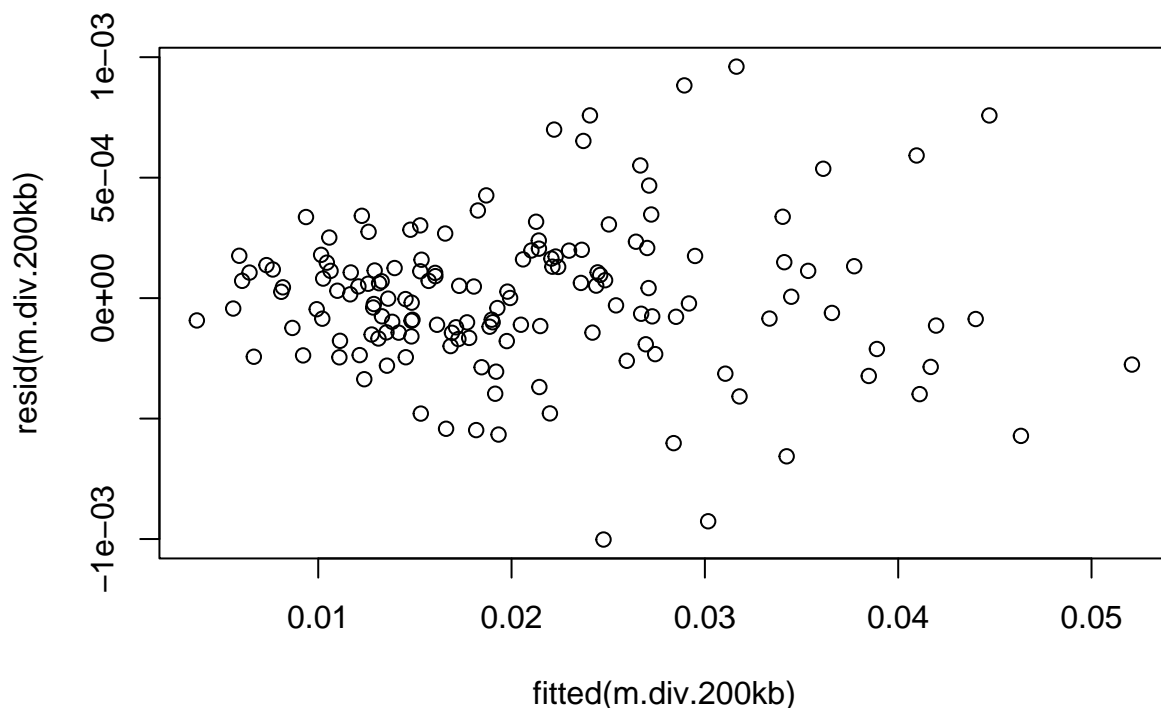
sim.lands.200kb$rhoC <- sim.lands.200kb$rho - mean(sim.lands.200kb$rho)

sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)

# for merging:
sim.lands.200kb.rep8 <- sim.lands.200kb
sim.lands.200kb.rep8$Replicate <- 8

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.200kb)
plot(resid(m.div.200kb)~fitted(m.div.200kb))

```



```

dwtest(m.div.200kb)

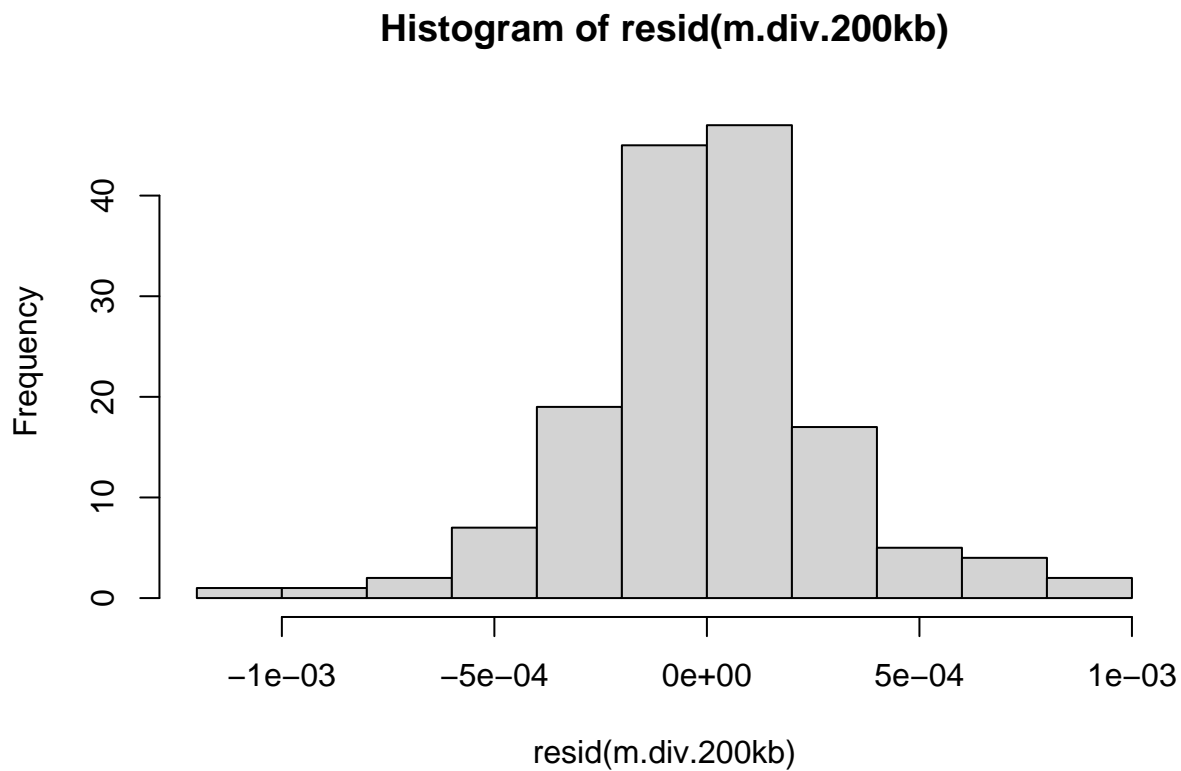
##
## Durbin-Watson test
##
## data: m.div.200kb
## DW = 2.036, p-value = 0.577
## alternative hypothesis: true autocorrelation is greater than 0

hmcetest(m.div.200kb)

##
## Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.54355, p-value = 0.776

```

```
hist(resid(m.div.200kb))
```



```
summary(m.div.200kb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.200kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0010021 -0.0001570 -0.0000008  0.0001486  0.0009608
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.057e-02  2.534e-05  811.819  <2e-16 ***
## thetaC       1.309e+00  3.478e-03  376.496  <2e-16 ***
## rhoC        -9.189e-03  1.709e-02  -0.538    0.592
## tmrcaC       2.531e-02  6.716e-04  37.681  <2e-16 ***
## thetaC:tmrcaC 1.518e+00  7.926e-02  19.154  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.000309 on 145 degrees of freedom
## Multiple R-squared:  0.999, Adjusted R-squared:  0.999
## F-statistic: 3.616e+04 on 4 and 145 DF, p-value: < 2.2e-16
```



```

anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.200kb[1, 8] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]) / anova.diversity$VarExp[4]
r2.sim.200kb[2, 8] <- anova.diversity$VarExp[1] * 100
r2.sim.200kb[3, 8] <- anova.diversity$VarExp[2] * 100
r2.sim.200kb[4, 8] <- anova.diversity$VarExp[3] * 100
r2.sim.200kb[5, 8] <- anova.diversity$VarExp[4] * 100

```

### 1.2.9 Replicate 9

```

rep_9.pi.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_9/rep_9.pi.200kb.bedgraph", header=TRUE)
rep_9.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_9/sim.tmrca.200k.map", header=TRUE)

```

```

sim.lands.200kb <- as.data.frame(cbind(rep_9.pi.200kb$pi, sim.theta.200kb$sim, sim.rho.200kb$sim, rep_9.tmrca.200kb$tmrca))
names(sim.lands.200kb) <- c("diversity", "theta", "rho", "tmrca")

```

```

cor.test(~theta+tmrca, data = sim.lands.200kb, method = "spearman")

```

```

##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 493800, p-value = 0.1366
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.1220935

```

```

cor.test(~theta+rho, data = sim.lands.200kb, method = "spearman")

```

```

##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 593676, p-value = 0.5002
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.0554712

```

```

cor.test(~rho+tmrca, data = sim.lands.200kb, method = "spearman")

```

```

##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 597682, p-value = 0.4463
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.062593

```

```

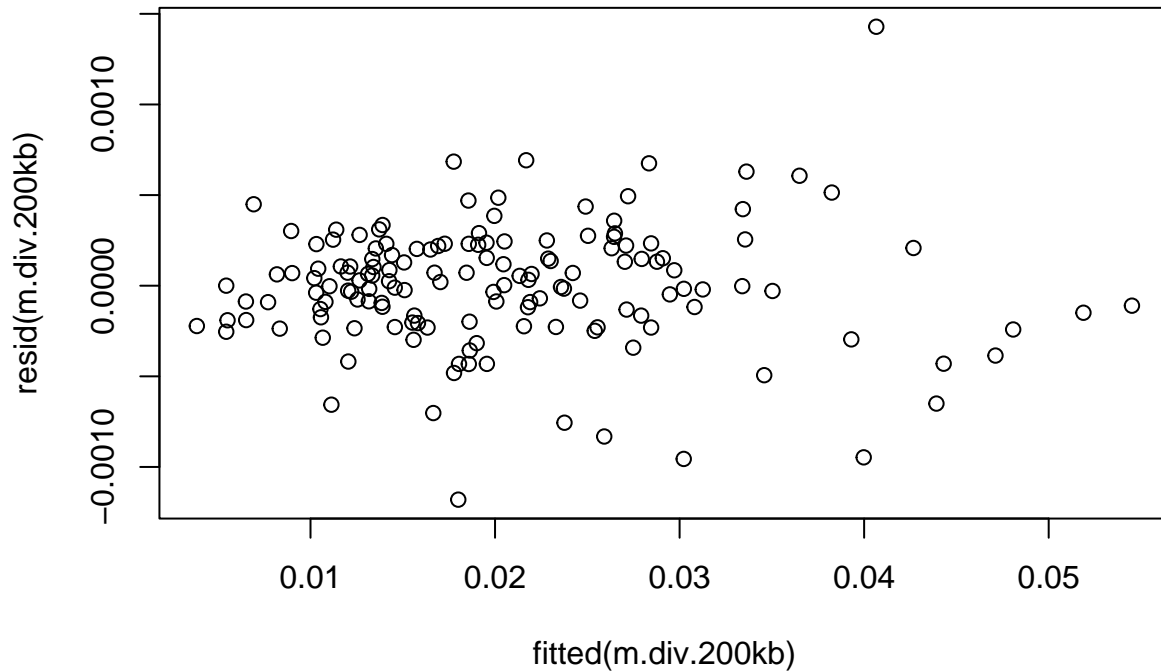
# centering
sim.lands.200kb$thetaC <- sim.lands.200kb$theta - mean(sim.lands.200kb$theta)
sim.lands.200kb$tmrcaC <- sim.lands.200kb$tmrca - mean(sim.lands.200kb$tmrca)
sim.lands.200kb$rhoC <- sim.lands.200kb$rho - mean(sim.lands.200kb$rho)

sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)

# for merging:
sim.lands.200kb.rep9 <- sim.lands.200kb
sim.lands.200kb.rep9$Replicate <- 9

m.div.200kb <- lm( diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.200kb)
plot(resid(m.div.200kb)~fitted(m.div.200kb))

```



```

dwtest(m.div.200kb)

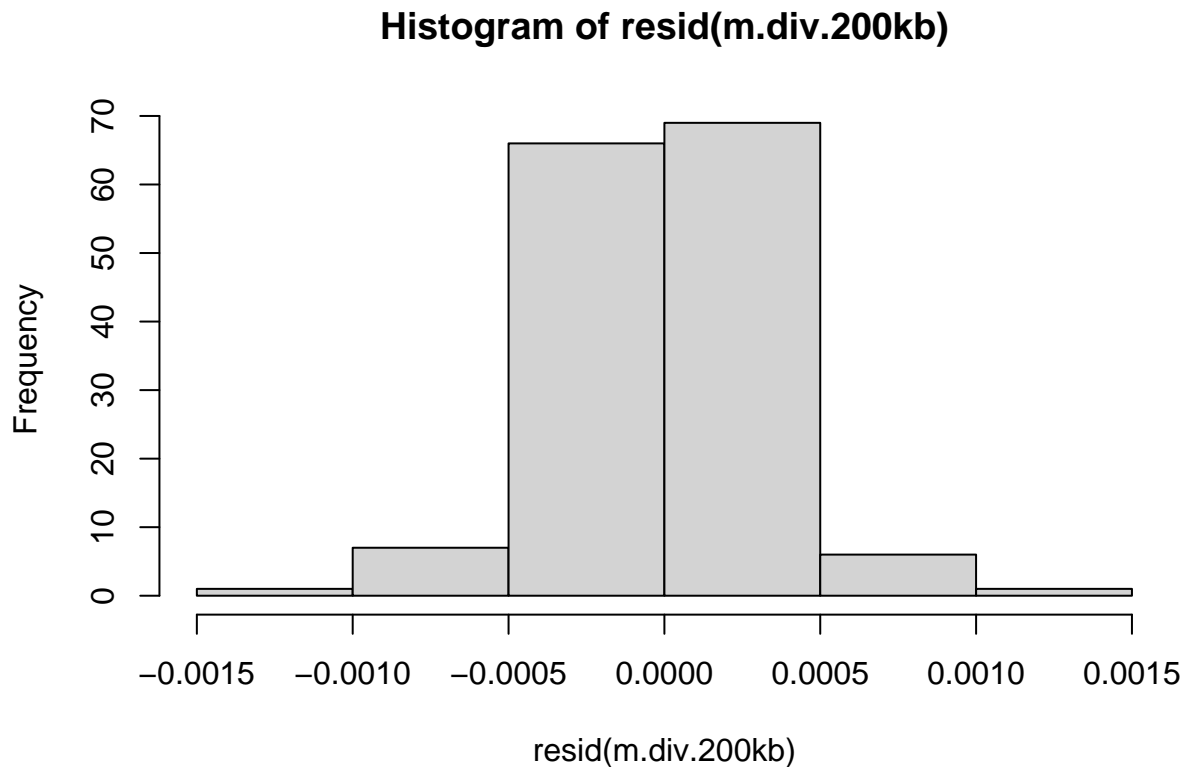
##
## Durbin-Watson test
##
## data: m.div.200kb
## DW = 1.6818, p-value = 0.022
## alternative hypothesis: true autocorrelation is greater than 0

hmctest(m.div.200kb)

##
## Harrison-McCabe test

```

```
##
## data: m.div.200kb
## HMC = 0.42403, p-value = 0.091
hist(resid(m.div.200kb))
```



```
summary(m.div.200kb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.200kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.180e-03 -1.904e-04  1.650e-06  2.078e-04  1.429e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.061e-02  2.834e-05  727.360  <2e-16 ***
## thetaC       1.308e+00  3.854e-03  339.390  <2e-16 ***
## rhoC         6.075e-03  1.879e-02   0.323    0.747
## tmrcaC       2.558e-02  8.061e-04  31.733  <2e-16 ***
## thetaC:tmrcaC 1.713e+00  1.105e-01  15.505  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.0003441 on 145 degrees of freedom
## Multiple R-squared: 0.9988, Adjusted R-squared: 0.9988
## F-statistic: 3.126e+04 on 4 and 145 DF, p-value: < 2.2e-16
```

```
anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.200kb[1, 9] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.sim.200kb[2, 9] <- anova.diversity$VarExp[1] * 100
r2.sim.200kb[3, 9] <- anova.diversity$VarExp[2] * 100
r2.sim.200kb[4, 9] <- anova.diversity$VarExp[3] * 100
r2.sim.200kb[5, 9] <- anova.diversity$VarExp[4] * 100
```

### 1.2.10 Replicate 10

```
rep_10.pi.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_10/rep_10.pi.200kb.bedgraph", 1
rep_10.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_10/sim.tmrca.200k.map", head

sim.lands.200kb <- as.data.frame(cbind(rep_10.pi.200kb$pi, sim.theta.200kb$sim, sim.rho.200kb$sim, rep_
names(sim.lands.200kb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.200kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 605678, p-value = 0.3502
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.07680806
```

```
cor.test(~theta+rho, data = sim.lands.200kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 593676, p-value = 0.5002
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.0554712
```

```
cor.test(~rho+tmrca, data = sim.lands.200kb, method = "spearman")
```

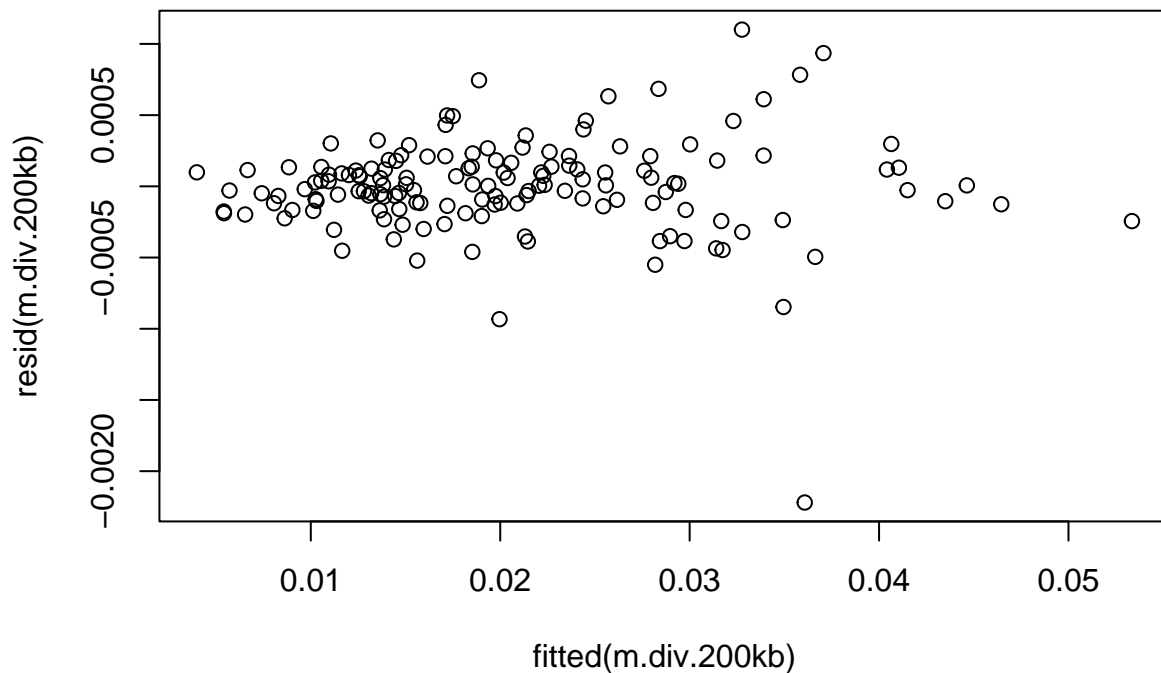
```
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 612862, p-value = 0.2753
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
```

```
## -0.08958087
# centering
sim.lands.200kb$thetaC <- sim.lands.200kb$theta - mean(sim.lands.200kb$theta)
sim.lands.200kb$tmrcaC <- sim.lands.200kb$tmrca - mean(sim.lands.200kb$tmrca)
sim.lands.200kb$rhoC <- sim.lands.200kb$rho - mean(sim.lands.200kb$rho)

sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)

# for merging:
sim.lands.200kb.rep10 <- sim.lands.200kb
sim.lands.200kb.rep10$Replicate <- 10

m.div.200kb <- lm(多样性 ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.200kb)
plot(resid(m.div.200kb)~fitted(m.div.200kb))
```



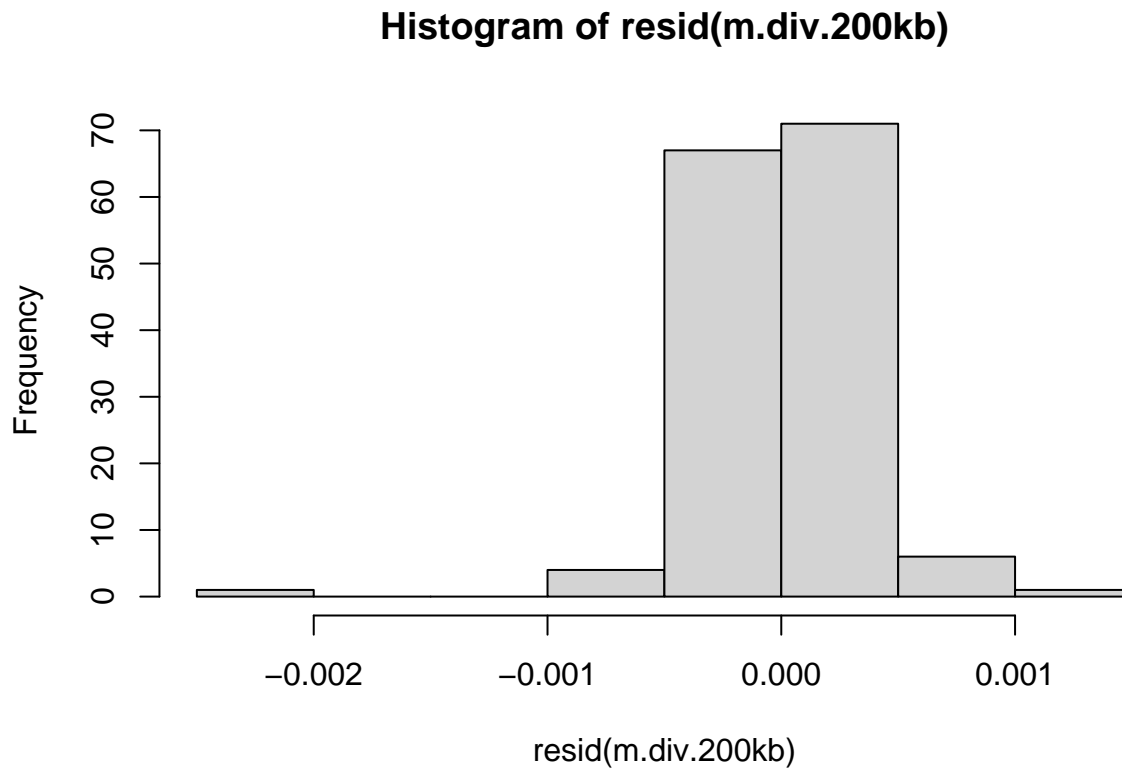
```
dwtest(m.div.200kb)

##
## Durbin-Watson test
##
## data: m.div.200kb
## DW = 2.1646, p-value = 0.8354
## alternative hypothesis: true autocorrelation is greater than 0

hmcetest(m.div.200kb)

##
```

```
## Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.41358, p-value = 0.065
hist(resid(m.div.200kb))
```



```
summary(m.div.200kb)

##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.200kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.220e-03 -1.345e-04  6.710e-06  1.375e-04  1.100e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.071e-02  2.877e-05  720.015  <2e-16 ***
## thetaC       1.317e+00  4.037e-03  326.195  <2e-16 ***
## rhoC         4.763e-03  1.928e-02   0.247    0.805
## tmrcaC       2.484e-02  6.682e-04  37.182  <2e-16 ***
## thetaC:tmrcaC 1.796e+00  7.919e-02  22.681  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.0003493 on 145 degrees of freedom
## Multiple R-squared:  0.9987, Adjusted R-squared:  0.9987
## F-statistic: 2.783e+04 on 4 and 145 DF,  p-value: < 2.2e-16

anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

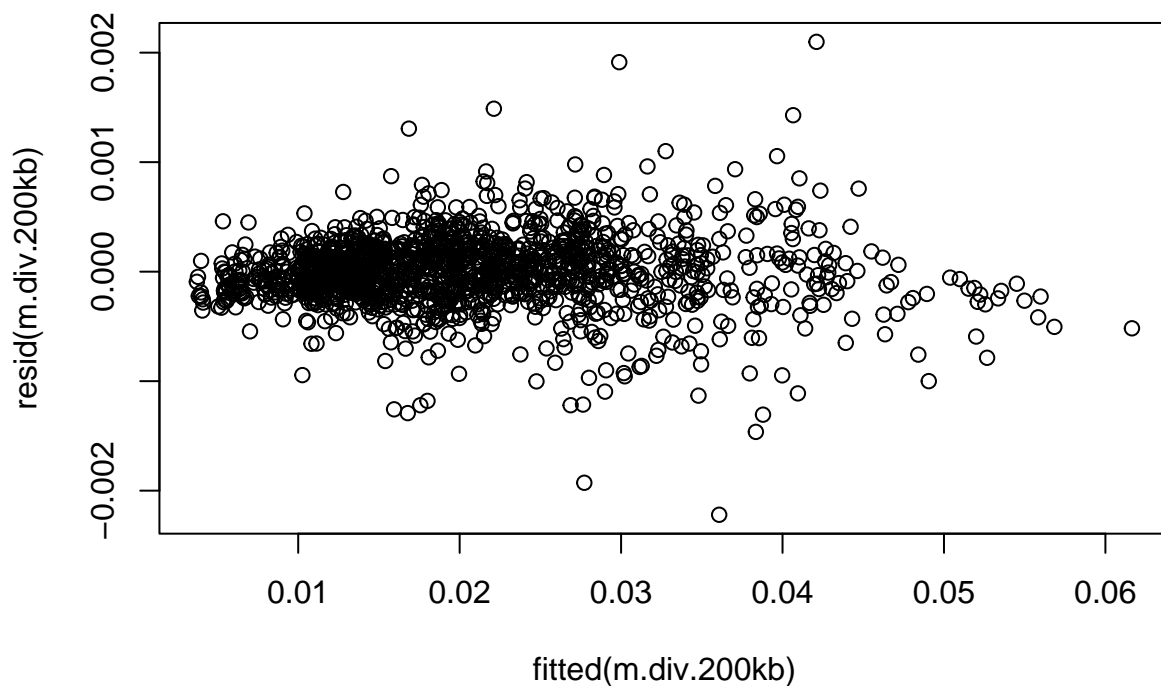
r2.sim.200kb[1, 10] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4])
r2.sim.200kb[2, 10] <- anova.diversity$VarExp[1] * 100
r2.sim.200kb[3, 10] <- anova.diversity$VarExp[2] * 100
r2.sim.200kb[4, 10] <- anova.diversity$VarExp[3] * 100
r2.sim.200kb[5, 10] <- anova.diversity$VarExp[4] * 100

r2.sim.200kb$average <- rowMeans(r2.sim.200kb)
r2.sim.200kb <- transform(r2.sim.200kb, sd=apply(r2.sim.200kb, 1, sd, na.rm = TRUE))
```

### 1.2.11 all replicates:

```
sim.lands.200kb.all <- rbind(sim.lands.200kb.rep1, sim.lands.200kb.rep2, sim.lands.200kb.rep3, sim.lands.200kb.rep4)

m.div.200kb <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)*as.factor(Replicate), data = sim.lands.200kb.all)
plot(resid(m.div.200kb)~fitted(m.div.200kb))
```



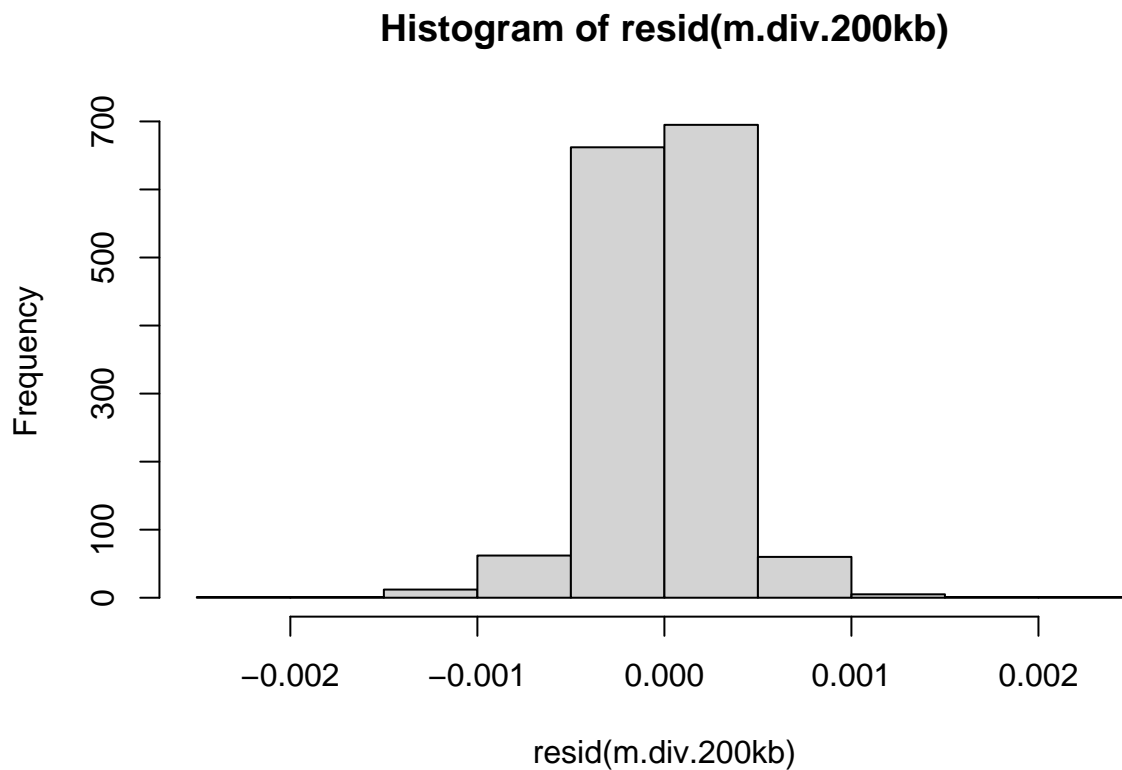
```
dwtest(m.div.200kb)
```

```
##  
## Durbin-Watson test  
##  
## data: m.div.200kb  
## DW = 1.9429, p-value = 0.07075  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.200kb)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.div.200kb  
## HMC = 0.49993, p-value = 0.507
```

```
hist(resid(m.div.200kb))
```



```
m.div.200kb.2 <- lm( diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC)*as.factor(Replica)  
m.div.200kb.3 <- lm( diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC + rhoC:tmrcaC)*as
```

```
AIC(m.div.200kb, m.div.200kb.2, m.div.200kb.3)
```

```
##           df      AIC  
## m.div.200kb  51 -19694.66  
## m.div.200kb.2 61 -19684.61  
## m.div.200kb.3 71 -19669.64
```



```
summary(m.div.200kb)
```

```
##
## Call:
## lm(formula = diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC) *
##     as.factor(Replicate), data = sim.lands.200kb.all)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.0022201	-0.0001690	0.0000028	0.0001766	0.0020986

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
## (Intercept)	2.067e-02	2.737e-05	755.296	< 2e-16	***
## thetaC	1.306e+00	3.699e-03	353.041	< 2e-16	***
## rhoC	3.201e-02	1.835e-02	1.744	0.08129	.
## tmrcaC	2.526e-02	7.513e-04	33.616	< 2e-16	***
## as.factor(Replicate)2	-1.119e-04	3.871e-05	-2.891	0.00390	**
## as.factor(Replicate)3	-2.543e-05	3.871e-05	-0.657	0.51140	
## as.factor(Replicate)4	-1.680e-04	3.871e-05	-4.340	1.52e-05	***
## as.factor(Replicate)5	-6.032e-05	3.871e-05	-1.558	0.11943	
## as.factor(Replicate)6	-1.150e-04	3.872e-05	-2.971	0.00302	**
## as.factor(Replicate)7	-4.595e-05	3.896e-05	-1.179	0.23850	
## as.factor(Replicate)8	-9.870e-05	3.879e-05	-2.545	0.01104	*
## as.factor(Replicate)9	-5.618e-05	3.887e-05	-1.445	0.14858	
## as.factor(Replicate)10	4.349e-05	3.887e-05	1.119	0.26340	
## thetaC:tmrcaC	1.536e+00	1.052e-01	14.594	< 2e-16	***
## thetaC:as.factor(Replicate)2	-8.931e-03	5.249e-03	-1.701	0.08911	.
## thetaC:as.factor(Replicate)3	2.766e-03	5.238e-03	0.528	0.59755	
## thetaC:as.factor(Replicate)4	-4.900e-03	5.233e-03	-0.936	0.34924	
## thetaC:as.factor(Replicate)5	2.876e-04	5.238e-03	0.055	0.95623	
## thetaC:as.factor(Replicate)6	-6.690e-03	5.342e-03	-1.252	0.21063	
## thetaC:as.factor(Replicate)7	9.250e-03	5.260e-03	1.759	0.07887	.
## thetaC:as.factor(Replicate)8	3.318e-03	5.283e-03	0.628	0.53009	
## thetaC:as.factor(Replicate)9	2.156e-03	5.271e-03	0.409	0.68256	
## thetaC:as.factor(Replicate)10	1.077e-02	5.356e-03	2.012	0.04443	*
## rhoC:as.factor(Replicate)2	-6.319e-02	2.591e-02	-2.439	0.01486	*
## rhoC:as.factor(Replicate)3	-3.391e-04	2.594e-02	-0.013	0.98957	
## rhoC:as.factor(Replicate)4	-2.323e-02	2.591e-02	-0.896	0.37014	
## rhoC:as.factor(Replicate)5	-2.309e-02	2.595e-02	-0.890	0.37376	
## rhoC:as.factor(Replicate)6	-2.844e-02	2.589e-02	-1.098	0.27218	
## rhoC:as.factor(Replicate)7	-4.821e-02	2.594e-02	-1.859	0.06323	.
## rhoC:as.factor(Replicate)8	-4.120e-02	2.609e-02	-1.579	0.11446	
## rhoC:as.factor(Replicate)9	-2.594e-02	2.592e-02	-1.001	0.31717	
## rhoC:as.factor(Replicate)10	-2.725e-02	2.606e-02	-1.046	0.29585	
## tmrcaC:as.factor(Replicate)2	-7.053e-04	1.010e-03	-0.698	0.48525	
## tmrcaC:as.factor(Replicate)3	-3.735e-04	1.002e-03	-0.373	0.70943	
## tmrcaC:as.factor(Replicate)4	-8.126e-04	1.019e-03	-0.797	0.42537	
## tmrcaC:as.factor(Replicate)5	-1.239e-03	1.049e-03	-1.181	0.23784	
## tmrcaC:as.factor(Replicate)6	-1.338e-03	1.056e-03	-1.267	0.20537	
## tmrcaC:as.factor(Replicate)7	-1.566e-03	1.051e-03	-1.490	0.13631	
## tmrcaC:as.factor(Replicate)8	5.125e-05	1.047e-03	0.049	0.96095	
## tmrcaC:as.factor(Replicate)9	3.244e-04	1.087e-03	0.299	0.76536	
## tmrcaC:as.factor(Replicate)10	-4.119e-04	9.877e-04	-0.417	0.67669	

```
## thetaC:tmrcaC:as.factor(Replicate)2 1.073e-01 1.385e-01 0.775 0.43829
## thetaC:tmrcaC:as.factor(Replicate)3 -6.355e-02 1.401e-01 -0.453 0.65026
## thetaC:tmrcaC:as.factor(Replicate)4 -6.695e-02 1.374e-01 -0.487 0.62605
## thetaC:tmrcaC:as.factor(Replicate)5 -1.561e-01 1.354e-01 -1.153 0.24925
## thetaC:tmrcaC:as.factor(Replicate)6 -2.867e-03 1.320e-01 -0.022 0.98267
## thetaC:tmrcaC:as.factor(Replicate)7 -1.134e-01 1.495e-01 -0.758 0.44833
## thetaC:tmrcaC:as.factor(Replicate)8 -1.763e-02 1.359e-01 -0.130 0.89677
## thetaC:tmrcaC:as.factor(Replicate)9 1.767e-01 1.505e-01 1.174 0.24045
## thetaC:tmrcaC:as.factor(Replicate)10 2.604e-01 1.298e-01 2.006 0.04505 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0003352 on 1450 degrees of freedom
## Multiple R-squared: 0.9989, Adjusted R-squared: 0.9988
## F-statistic: 2.574e+04 on 49 and 1450 DF, p-value: < 2.2e-16
```

### 1.3 1 Mb scale

```
r2.sim.1Mb <- as.data.frame(matrix(ncol = nreps, nrow = 5))
row.names(r2.sim.1Mb) <- c("Total", "Theta", "Rho", "TMRCA", "Theta:TMRCA")
colnames(r2.sim.1Mb) <- reps

sim.rho.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/dm.sim.rho.1e+06.txt", header = T)
sim.theta.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/dm.sim.theta.1e+06.txt", header = T)
```

#### 1.3.1 Replicate 1

```
rep_1.pi.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_1/rep_1.pi.1Mb.bedgraph", header = T)
rep_1.tmrca.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_1/sim.tmrca.1M.map", header = T)

sim.lands.1Mb <- as.data.frame(cbind(rep_1.pi.1Mb$pi, sim.theta.1Mb$sim, sim.rho.1Mb$sim, rep_1.tmrca.1Mb$tmrca))
names(sim.lands.1Mb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.1Mb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 5134, p-value = 0.452
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.142158

cor.test(~theta+rho, data = sim.lands.1Mb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 5654, p-value = 0.1684
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
```

```

##          rho
## -0.257842
cor.test(~rho+tmrca, data = sim.lands.1Mb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 4576, p-value = 0.9251
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##          rho
## -0.01802002

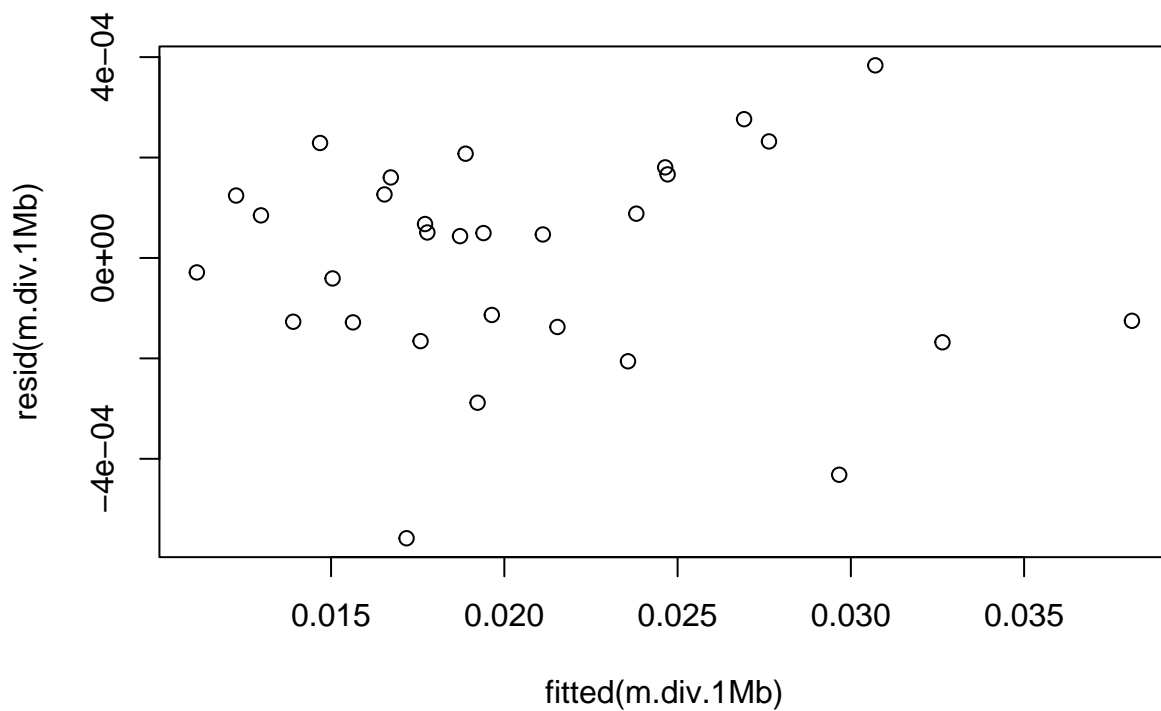
# centering
sim.lands.1Mb$thetaC <- sim.lands.1Mb$theta - mean(sim.lands.1Mb$theta)
sim.lands.1Mb$tmrcaC <- sim.lands.1Mb$tmrca - mean(sim.lands.1Mb$tmrca)
sim.lands.1Mb$rhoC <- sim.lands.1Mb$rho - mean(sim.lands.1Mb$rho)

sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)

# for merging:
sim.lands.1Mb.rep1 <- sim.lands.1Mb
sim.lands.1Mb.rep1$Replicate <- 1

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.1Mb)
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```
dwtest(m.div.1Mb)
```

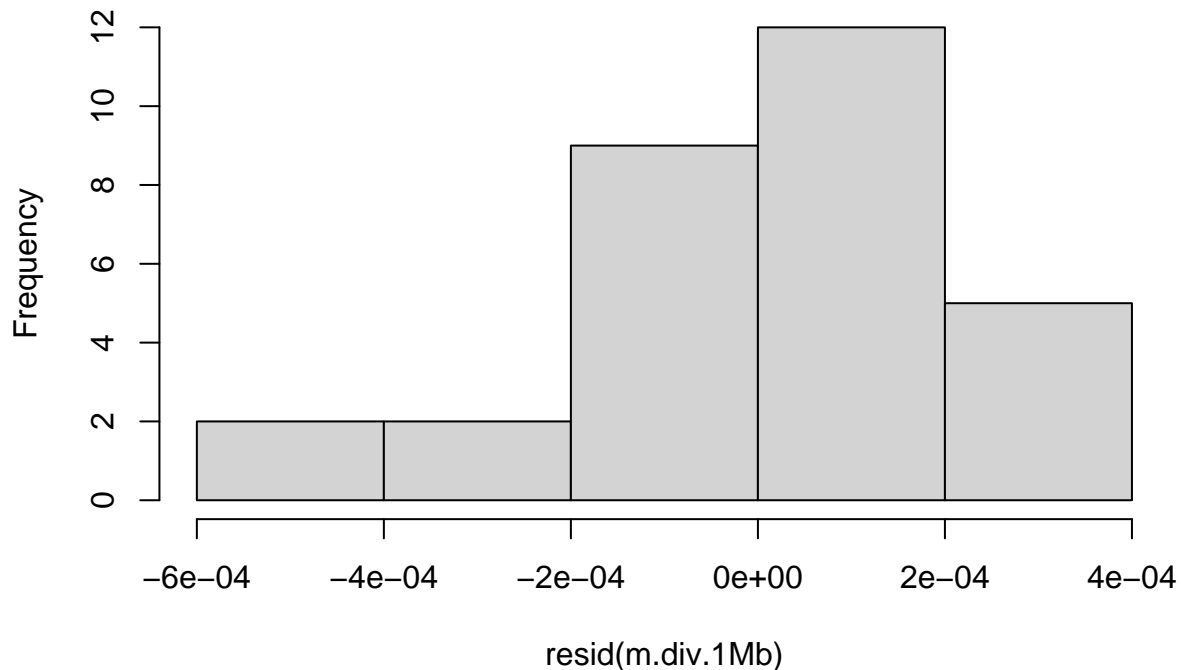
```
##
## Durbin-Watson test
##
## data: m.div.1Mb
## DW = 1.5295, p-value = 0.09374
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcctest(m.div.1Mb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.59412, p-value = 0.772
```

```
hist(resid(m.div.1Mb))
```

## Histogram of resid(m.div.1Mb)



```
summary(m.div.1Mb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.1Mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0005582 -0.0001279  0.0000482  0.0001519  0.0003836
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.067e-02  4.132e-05  500.181 < 2e-16 ***
## thetaC       1.313e+00  9.626e-03  136.418 < 2e-16 ***
## rhoC        -2.472e-03  6.889e-02  -0.036  0.9717
## tmrcaC       2.602e-02  2.881e-03   9.031 2.41e-09 ***
## thetaC:tmrcaC 1.182e+00  6.258e-01   1.889  0.0706 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0002263 on 25 degrees of freedom
## Multiple R-squared:  0.9989, Adjusted R-squared:  0.9988
## F-statistic: 5910 on 4 and 25 DF, p-value: < 2.2e-16
```

```
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
```

```
anova.diversity$VarExp <- apiss / sum(apiss)
```

```
r2.sim.1Mb[1, 1] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) * 100
r2.sim.1Mb[2, 1] <- anova.diversity$VarExp[1] * 100
r2.sim.1Mb[3, 1] <- anova.diversity$VarExp[2] * 100
r2.sim.1Mb[4, 1] <- anova.diversity$VarExp[3] * 100
r2.sim.1Mb[5, 1] <- anova.diversity$VarExp[4] * 100
```

### 1.3.2 Replicate 2

```
rep_2.pi.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_2/rep_2.pi.1Mb.bedgraph", header = TRUE, as.is = TRUE)
rep_2.tmrca.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_2/sim.tmrca.1M.map", header = TRUE, as.is = TRUE)
```

```
sim.lands.1Mb <- as.data.frame(cbind(rep_2.pi.1Mb$pi, sim.theta.1Mb$sim, sim.rho.1Mb$sim, rep_2.tmrca.1Mb$tmrca))
names(sim.lands.1Mb) <- c("diversity", "theta", "rho", "tmrca")
```

```
cor.test(~theta+tmrca, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 3982, p-value = 0.5467
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.1141268
```

```
cor.test(~theta+rho, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 5654, p-value = 0.1684
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.257842
```

```
cor.test(~rho+tmrca, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 3726, p-value = 0.3645
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.171079
```

```
cor.test(~diversity+theta, data = sim.lands.200kb, method = "spearman")
```

```
##
```

```

## Spearman's rank correlation rho
##
## data: diversity and theta
## S = 4053.2, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9927941

cor.test(~diversity+tmrca, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: diversity and tmrca
## S = 552924, p-value = 0.8364
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.01698031

cor.test(~diversity+rho, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: diversity and rho
## S = 596694, p-value = 0.4591
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.06083648

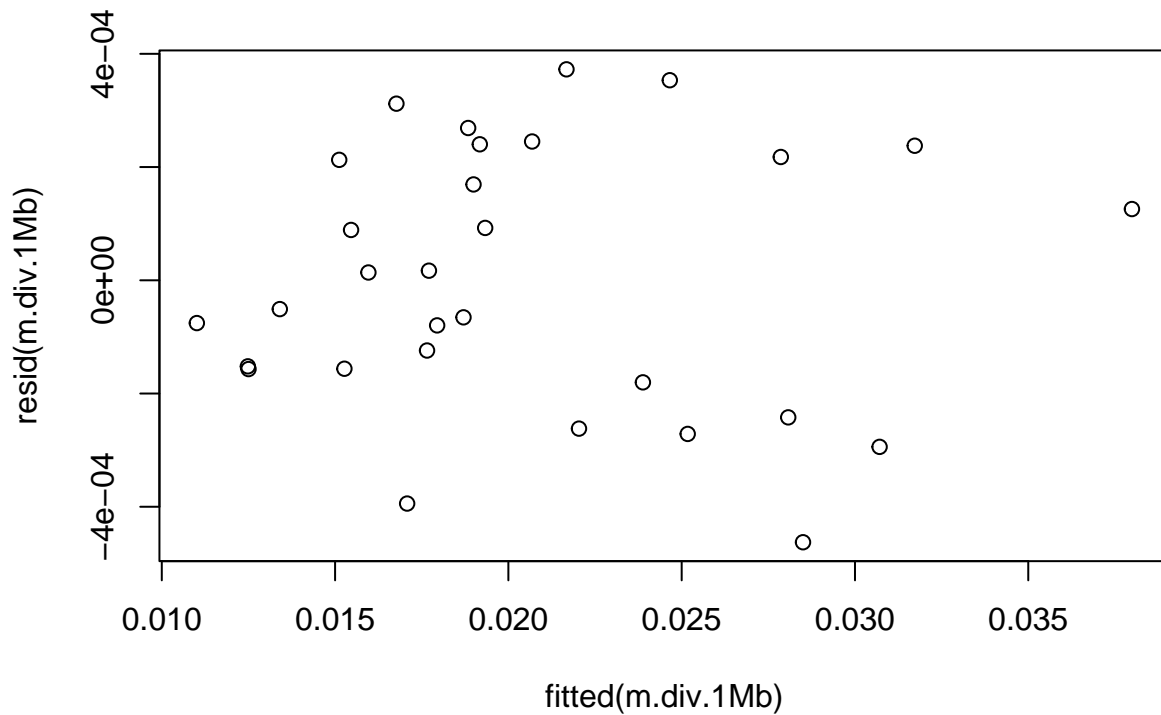
# centering
sim.lands.1Mb$thetaC <- sim.lands.1Mb$theta - mean(sim.lands.1Mb$theta)
sim.lands.1Mb$tmrcaC <- sim.lands.1Mb$tmrca - mean(sim.lands.1Mb$tmrca)
sim.lands.1Mb$rhoC <- sim.lands.1Mb$rho - mean(sim.lands.1Mb$rho)

sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)

# for merging:
sim.lands.1Mb.rep2 <- sim.lands.1Mb
sim.lands.1Mb.rep2$Replicate <- 2

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.1Mb)
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```
dwtest(m.div.1Mb)
```

```
##
## Durbin-Watson test
##
## data: m.div.1Mb
## DW = 2.6531, p-value = 0.9631
## alternative hypothesis: true autocorrelation is greater than 0
```

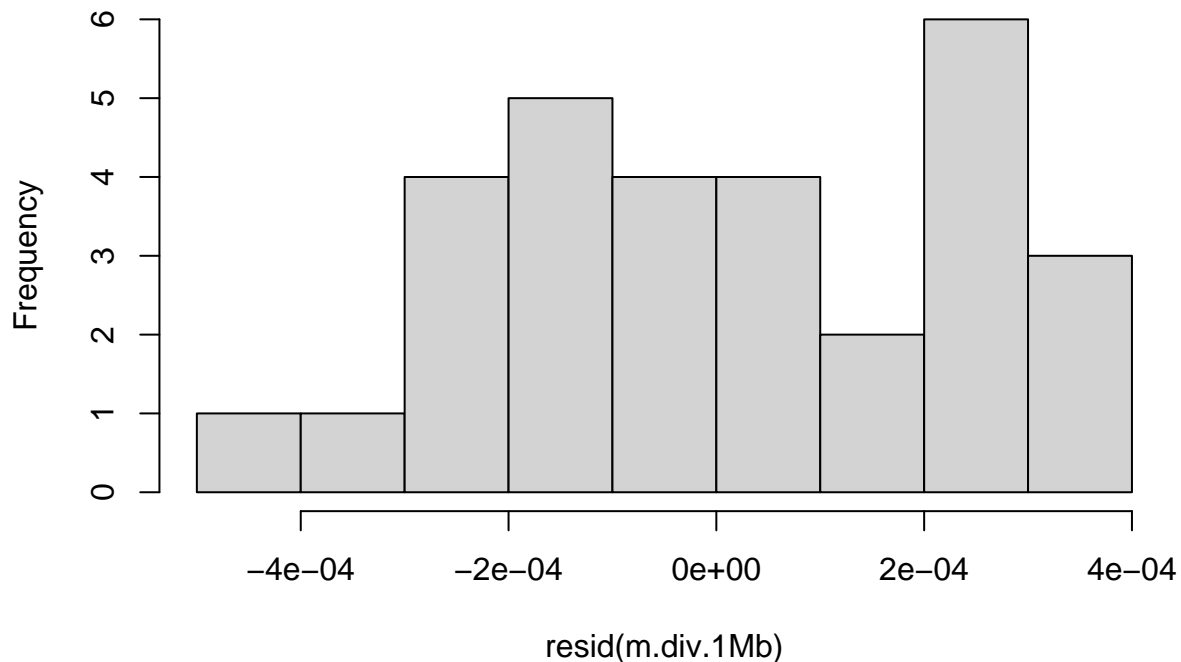
```
hmcctest(m.div.1Mb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.51119, p-value = 0.514
```

```
hist(resid(m.div.1Mb))
```



## Histogram of resid(m.div.1Mb)



```
summary(m.div.1Mb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.1Mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.627e-04 -1.567e-04 -1.856e-05  2.165e-04  3.726e-04
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.052e-02  4.606e-05  445.437 < 2e-16 ***
## thetaC       1.285e+00  1.013e-02  126.882 < 2e-16 ***
## rhoC        -4.630e-02  7.505e-02  -0.617  0.542900
## tmrcaC       2.480e-02  2.820e-03   8.796  3.99e-09 ***
## thetaC:tmrcaC 2.351e+00  5.373e-01   4.376  0.000188 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0002496 on 25 degrees of freedom
## Multiple R-squared:  0.9987, Adjusted R-squared:  0.9985
## F-statistic: 4800 on 4 and 25 DF, p-value: < 2.2e-16
```

```
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
```

```
anova.diversity$VarExp <- apiss / sum(apiss)
```

```
r2.sim.1Mb[1, 2] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) / 4
r2.sim.1Mb[2, 2] <- anova.diversity$VarExp[1] * 100
r2.sim.1Mb[3, 2] <- anova.diversity$VarExp[2] * 100
r2.sim.1Mb[4, 2] <- anova.diversity$VarExp[3] * 100
r2.sim.1Mb[5, 2] <- anova.diversity$VarExp[4] * 100
```

### 1.3.3 Replicate 3

```
rep_3.pi.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_3/rep_3.pi.1Mb.bedgraph", header = TRUE, as.is = TRUE)
rep_3.tmrca.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_3/sim.tmrca.1M.map", header = TRUE, as.is = TRUE)
```

```
sim.lands.1Mb <- as.data.frame(cbind(rep_3.pi.1Mb$pi, sim.theta.1Mb$sim, sim.rho.1Mb$sim, rep_3.tmrca.1Mb$tmrca))
names(sim.lands.1Mb) <- c("diversity", "theta", "rho", "tmrca")
```

```
cor.test(~theta+tmrca, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 6456, p-value = 0.01674
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.4362625
```

```
cor.test(~rho+tmrca, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 5654, p-value = 0.1684
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.257842
```

```
cor.test(~rho+tmrca, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 4500, p-value = 0.9963
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.001112347
```

```
cor.test(~diversity+theta, data = sim.lands.200kb, method = "spearman")
```

```
##
```

```

## Spearman's rank correlation rho
##
## data: diversity and theta
## S = 4053.2, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9927941

cor.test(~diversity+tmrca, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: diversity and tmrca
## S = 552924, p-value = 0.8364
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.01698031

cor.test(~diversity+rho, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: diversity and rho
## S = 596694, p-value = 0.4591
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.06083648

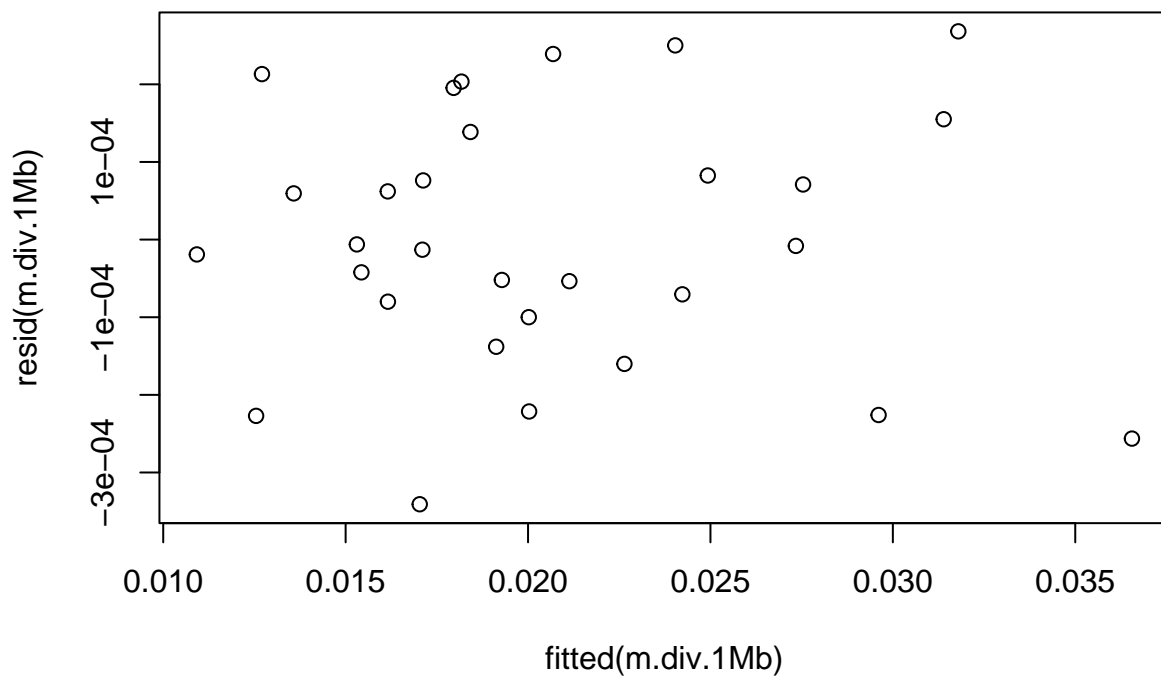
# centering
sim.lands.1Mb$thetaC <- sim.lands.1Mb$theta - mean(sim.lands.1Mb$theta)
sim.lands.1Mb$tmrcaC <- sim.lands.1Mb$tmrca - mean(sim.lands.1Mb$tmrca)
sim.lands.1Mb$rhoC <- sim.lands.1Mb$rho - mean(sim.lands.1Mb$rho)

sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)

# for merging:
sim.lands.1Mb.rep3 <- sim.lands.1Mb
sim.lands.1Mb.rep3$Replicate <- 3

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.1Mb)
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```
dwtest(m.div.1Mb)
```

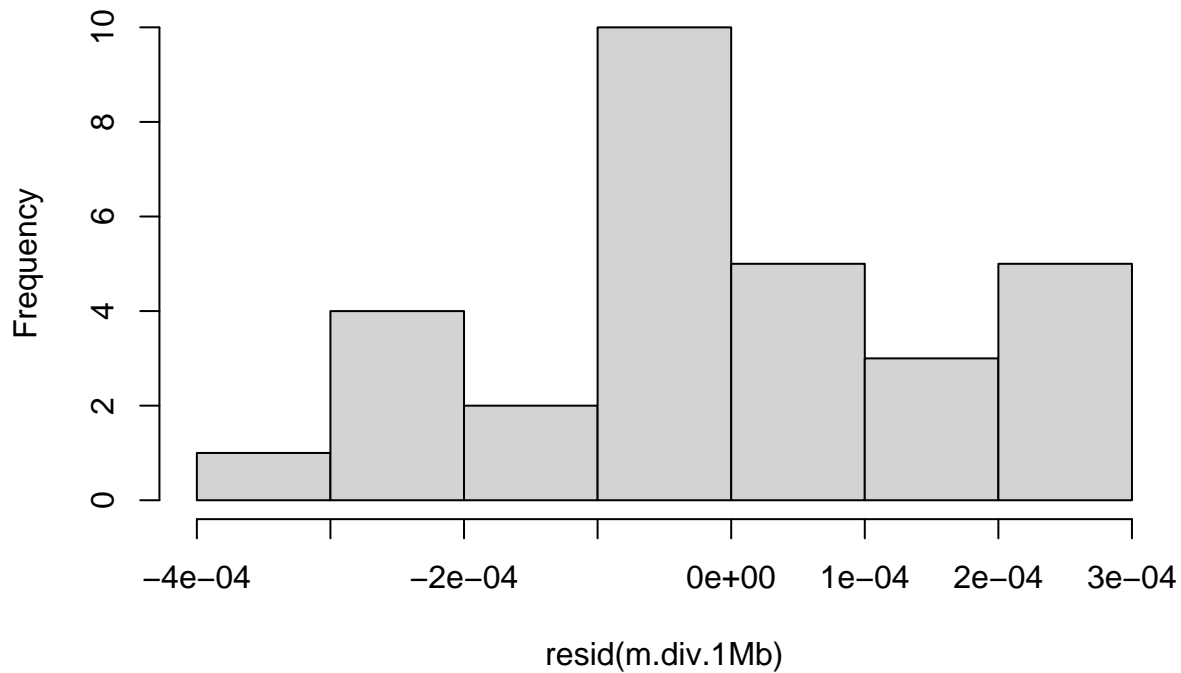
```
##
## Durbin-Watson test
##
## data: m.div.1Mb
## DW = 2.3611, p-value = 0.8304
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.1Mb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.69982, p-value = 0.937
```

```
hist(resid(m.div.1Mb))
```

## Histogram of resid(m.div.1Mb)



```
summary(m.div.1Mb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.1Mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.409e-04 -9.496e-05 -1.057e-05  1.246e-04  2.682e-04
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.070e-02  3.559e-05  581.752 < 2e-16 ***
## thetaC       1.326e+00  7.629e-03  173.822 < 2e-16 ***
## rhoC         6.225e-02  5.347e-02   1.164   0.255
## tmrcaC       3.071e-02  2.054e-03  14.949 5.67e-14 ***
## thetaC:tmrcaC 2.593e+00  5.197e-01   4.990 3.83e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001781 on 25 degrees of freedom
## Multiple R-squared:  0.9993, Adjusted R-squared:  0.9992
## F-statistic: 9011 on 4 and 25 DF, p-value: < 2.2e-16
```

```
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
```

```
anova.diversity$VarExp <- apiss / sum(apiss)
```

```
r2.sim.1Mb[1, 3] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) * 100
r2.sim.1Mb[2, 3] <- anova.diversity$VarExp[1] * 100
r2.sim.1Mb[3, 3] <- anova.diversity$VarExp[2] * 100
r2.sim.1Mb[4, 3] <- anova.diversity$VarExp[3] * 100
r2.sim.1Mb[5, 3] <- anova.diversity$VarExp[4] * 100
```

### 1.3.4 Replicate 4

```
rep_4.pi.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_4/rep_4.pi.1Mb.bedgraph", header = TRUE, as.is = TRUE)
rep_4.tmrca.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_4/sim.tmrca.1M.map", header = TRUE, as.is = TRUE)
```

```
sim.lands.1Mb <- as.data.frame(cbind(rep_4.pi.1Mb$pi, sim.theta.1Mb$sim, sim.rho.1Mb$sim, rep_4.tmrca.1Mb$tmrca))
names(sim.lands.1Mb) <- c("diversity", "theta", "rho", "tmrca")
```

```
cor.test(~theta+tmrca, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 4678, p-value = 0.8307
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.0407119
```

```
cor.test(~theta+rho, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 5654, p-value = 0.1684
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.257842
```

```
cor.test(~rho+tmrca, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 3482, p-value = 0.2302
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.2253615
```

```
cor.test(~diversity+theta, data = sim.lands.200kb, method = "spearman")
```

```
##
```

```

## Spearman's rank correlation rho
##
## data: diversity and theta
## S = 4053.2, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9927941

cor.test(~diversity+tmrca, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: diversity and tmrca
## S = 552924, p-value = 0.8364
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.01698031

cor.test(~diversity+rho, data = sim.lands.200kb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: diversity and rho
## S = 596694, p-value = 0.4591
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.06083648

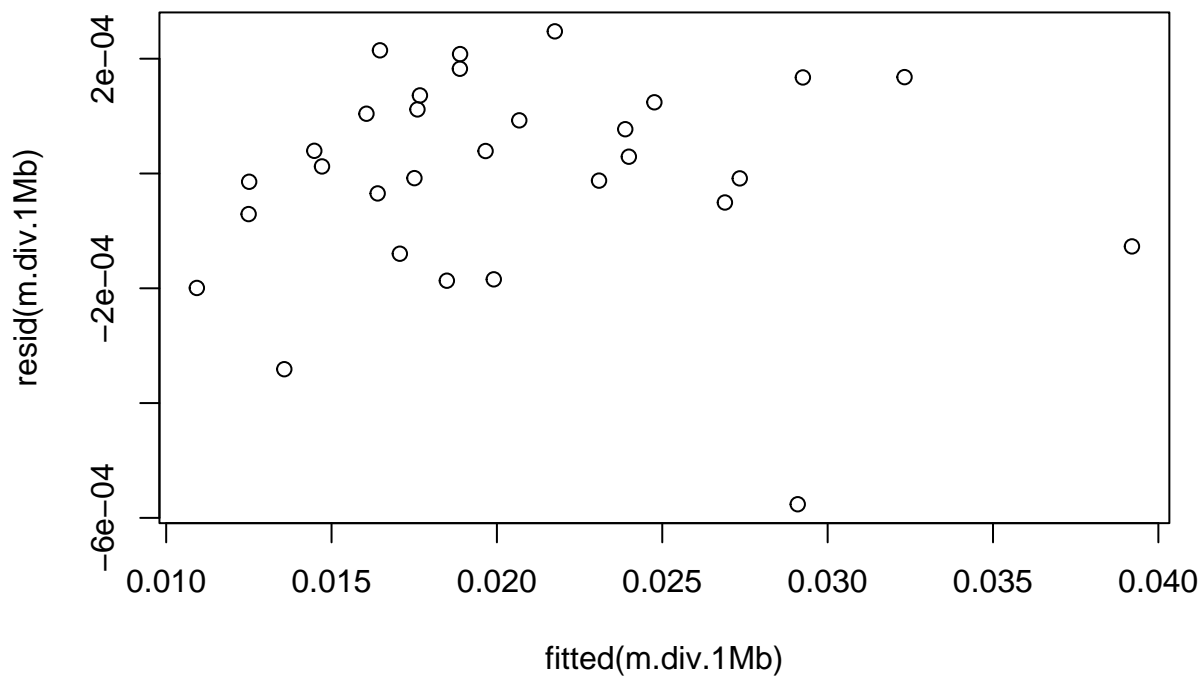
# centering
sim.lands.1Mb$thetaC <- sim.lands.1Mb$theta - mean(sim.lands.1Mb$theta)
sim.lands.1Mb$tmrcaC <- sim.lands.1Mb$tmrca - mean(sim.lands.1Mb$tmrca)
sim.lands.1Mb$rhoC <- sim.lands.1Mb$rho - mean(sim.lands.1Mb$rho)

sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)

# for merging:
sim.lands.1Mb.rep4 <- sim.lands.1Mb
sim.lands.1Mb.rep4$Replicate <- 4

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.1Mb)
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```
dwtest(m.div.1Mb)
```

```
##
## Durbin-Watson test
##
## data: m.div.1Mb
## DW = 2.4044, p-value = 0.8773
## alternative hypothesis: true autocorrelation is greater than 0
```

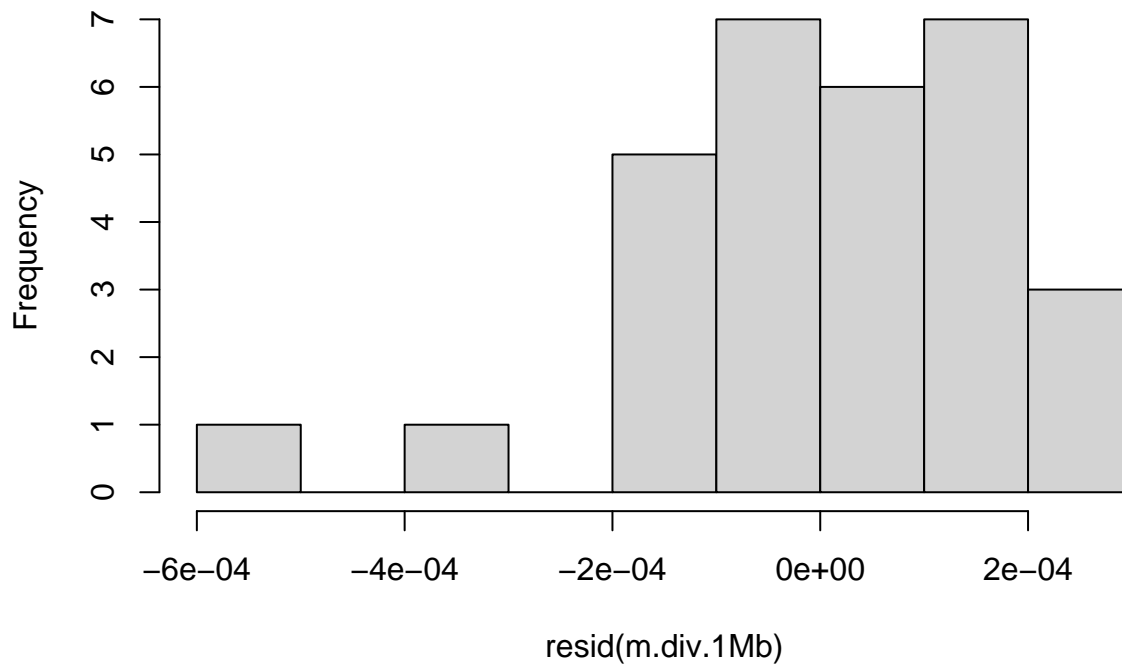
```
hmctest(m.div.1Mb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.39065, p-value = 0.179
```

```
hist(resid(m.div.1Mb))
```



## Histogram of resid(m.div.1Mb)



```
summary(m.div.1Mb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.1Mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.763e-04 -6.567e-05  2.077e-05  1.210e-04  2.477e-04
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.049e-02  3.521e-05  582.000 < 2e-16 ***
## thetaC       1.290e+00  7.811e-03  165.192 < 2e-16 ***
## rhoC         1.009e-02  5.807e-02   0.174   0.863
## tmrcaC        2.364e-02  1.989e-03  11.884 8.84e-12 ***
## thetaC:tmrcaC 2.028e+00  3.007e-01   6.744 4.55e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001916 on 25 degrees of freedom
## Multiple R-squared:  0.9992, Adjusted R-squared:  0.9991
## F-statistic: 8277 on 4 and 25 DF, p-value: < 2.2e-16
```

```
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
```

```
anova.diversity$VarExp <- apiss / sum(apiss)
```

```
r2.sim.1Mb[1, 4] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) * 100
r2.sim.1Mb[2, 4] <- anova.diversity$VarExp[1] * 100
r2.sim.1Mb[3, 4] <- anova.diversity$VarExp[2] * 100
r2.sim.1Mb[4, 4] <- anova.diversity$VarExp[3] * 100
r2.sim.1Mb[5, 4] <- anova.diversity$VarExp[4] * 100
```

### 1.3.5 Replicate 5

```
rep_5.pi.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_5/rep_5.pi.1Mb.bedgraph", header = TRUE)
rep_5.tmrca.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_5/sim.tmrca.1M.map", header = TRUE)
```

```
sim.lands.1Mb <- as.data.frame(cbind(rep_5.pi.1Mb$pi, sim.theta.1Mb$sim, sim.rho.1Mb$sim, rep_5.tmrca.1Mb$tmrca))
names(sim.lands.1Mb) <- c("diversity", "theta", "rho", "tmrca")
```

```
cor.test(~theta+tmrca, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 4438, p-value = 0.9476
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.01268076
```

```
cor.test(~theta+rho, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 5654, p-value = 0.1684
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.257842
```

```
cor.test(~rho+tmrca, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 4306, p-value = 0.8252
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.04204672
```

```
# centering
```

```
sim.lands.1Mb$thetaC <- sim.lands.1Mb$theta - mean(sim.lands.1Mb$theta)
sim.lands.1Mb$tmrcaC <- sim.lands.1Mb$tmrca - mean(sim.lands.1Mb$tmrca)
```

```

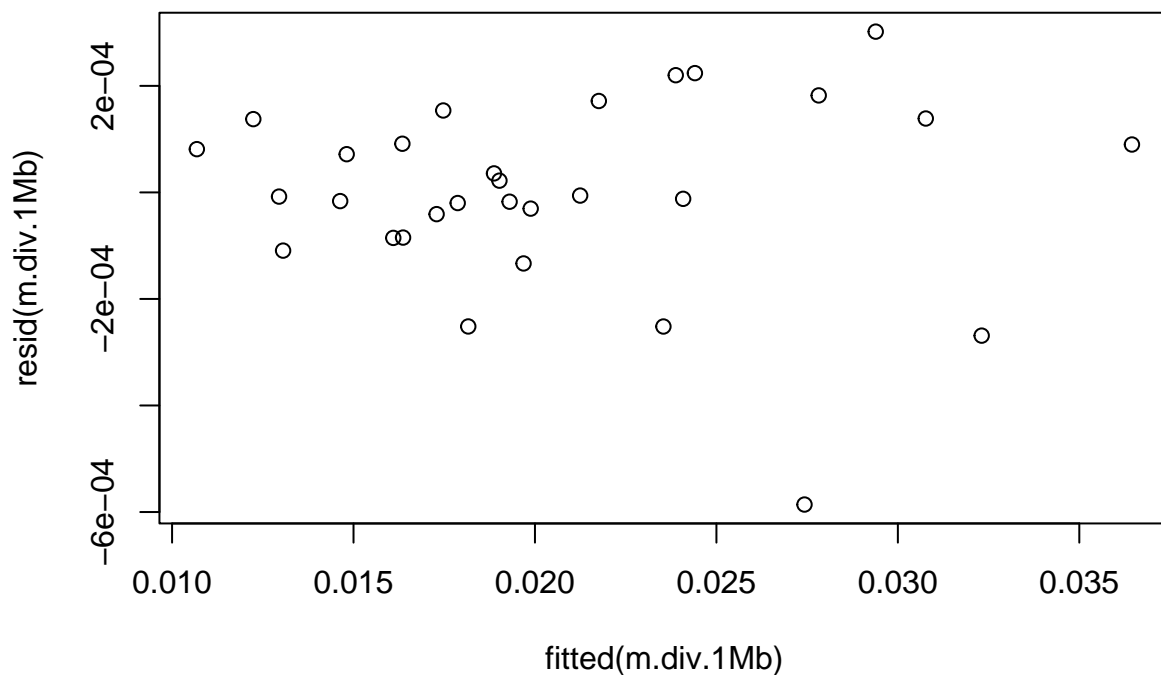
sim.lands.1Mb$rhoC <- sim.lands.1Mb$rho - mean(sim.lands.1Mb$rho)

sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)

# for merging:
sim.lands.1Mb.rep5 <- sim.lands.1Mb
sim.lands.1Mb.rep5$Replicate <- 5

m.div.1Mb <- lm(多样性 ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.1Mb)
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```

dwtest(m.div.1Mb)

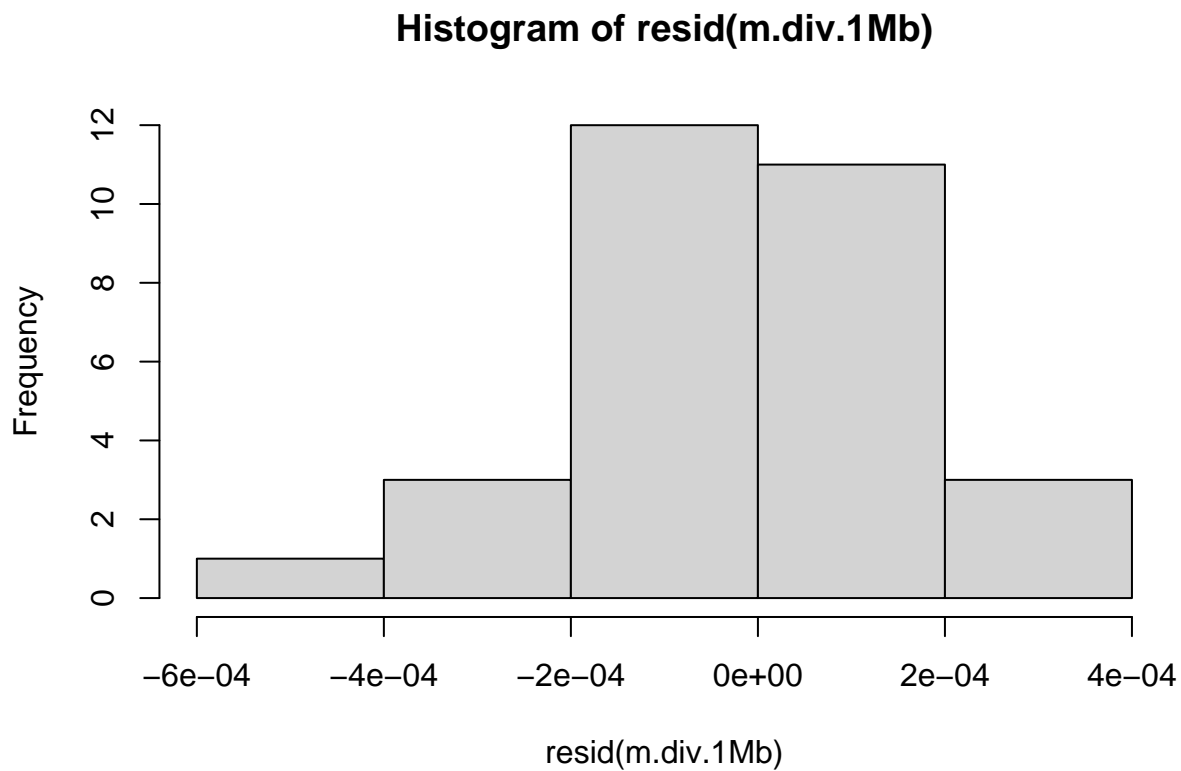
##
## Durbin-Watson test
##
## data: m.div.1Mb
## DW = 1.9571, p-value = 0.43
## alternative hypothesis: true autocorrelation is greater than 0

hmcetest(m.div.1Mb)

##
## Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.35533, p-value = 0.126

```

```
hist(resid(m.div.1Mb))
```



```
summary(m.div.1Mb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.1Mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.859e-04 -7.379e-05 -6.890e-06  1.260e-04  3.018e-04
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.060e-02  3.534e-05  582.791  < 2e-16 ***
## thetaC       1.306e+00  7.849e-03  166.389  < 2e-16 ***
## rhoC        -3.495e-02  5.720e-02  -0.611   0.5467
## tmrcaC       2.514e-02  2.157e-03  11.655  1.34e-11 ***
## thetaC:tmrcaC 1.135e+00  5.053e-01   2.246   0.0338 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001933 on 25 degrees of freedom
## Multiple R-squared:  0.9992, Adjusted R-squared:  0.9991
## F-statistic: 7809 on 4 and 25 DF,  p-value: < 2.2e-16
```

```

anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.1Mb[1, 5] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.sim.1Mb[2, 5] <- anova.diversity$VarExp[1] * 100
r2.sim.1Mb[3, 5] <- anova.diversity$VarExp[2] * 100
r2.sim.1Mb[4, 5] <- anova.diversity$VarExp[3] * 100
r2.sim.1Mb[5, 5] <- anova.diversity$VarExp[4] * 100

```

### 1.3.6 Replicate 6

```

rep_6.pi.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_6/rep_6.pi.1Mb.bedgraph", header = TRUE)
rep_6.tmrca.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_6/sim.tmrca.1M.map", header = TRUE)

sim.lands.1Mb <- as.data.frame(cbind(rep_6.pi.1Mb$pi, sim.theta.1Mb$sim, sim.rho.1Mb$sim, rep_6.tmrca.1Mb$tmrca))
names(sim.lands.1Mb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.1Mb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 3592, p-value = 0.2858
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.2008899

cor.test(~theta+rho, data = sim.lands.1Mb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 5654, p-value = 0.1684
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.257842

cor.test(~rho+tmrca, data = sim.lands.1Mb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 4268, p-value = 0.7906
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.05050056

```

```

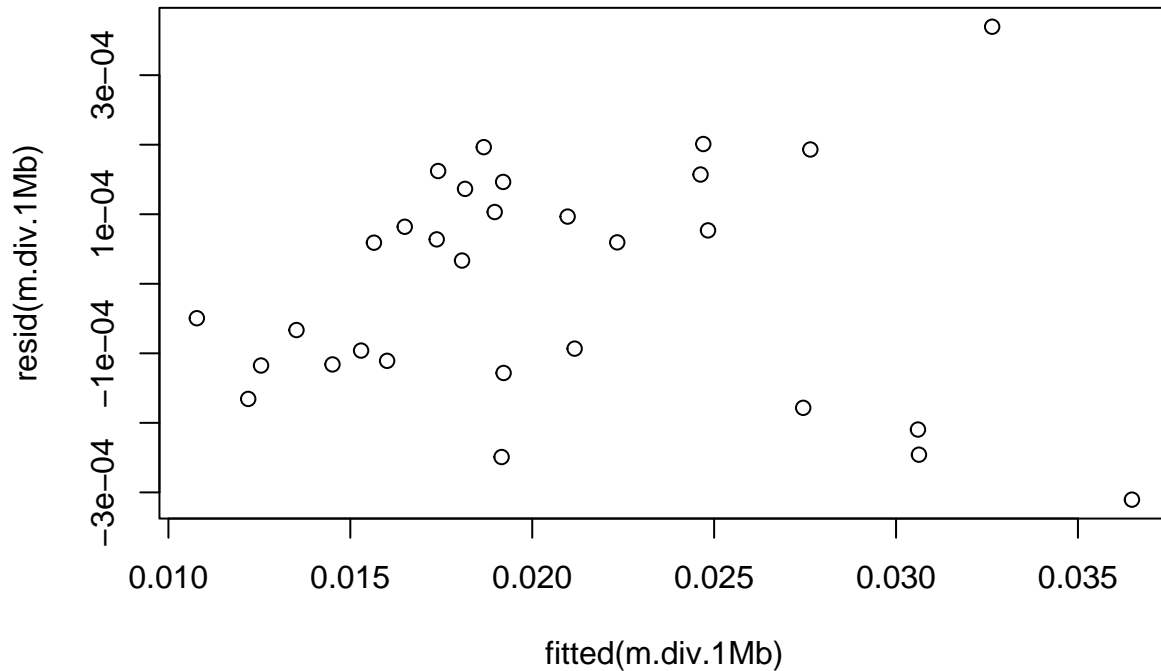
# centering
sim.lands.1Mb$thetaC <- sim.lands.1Mb$theta - mean(sim.lands.1Mb$theta)
sim.lands.1Mb$tmrcaC <- sim.lands.1Mb$tmrca - mean(sim.lands.1Mb$tmrca)
sim.lands.1Mb$rhoC <- sim.lands.1Mb$rho - mean(sim.lands.1Mb$rho)

sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)

# for merging:
sim.lands.1Mb.rep6 <- sim.lands.1Mb
sim.lands.1Mb.rep6$Replicate <- 6

m.div.1Mb <- lm(多样性 ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.1Mb)
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```

dwtest(m.div.1Mb)

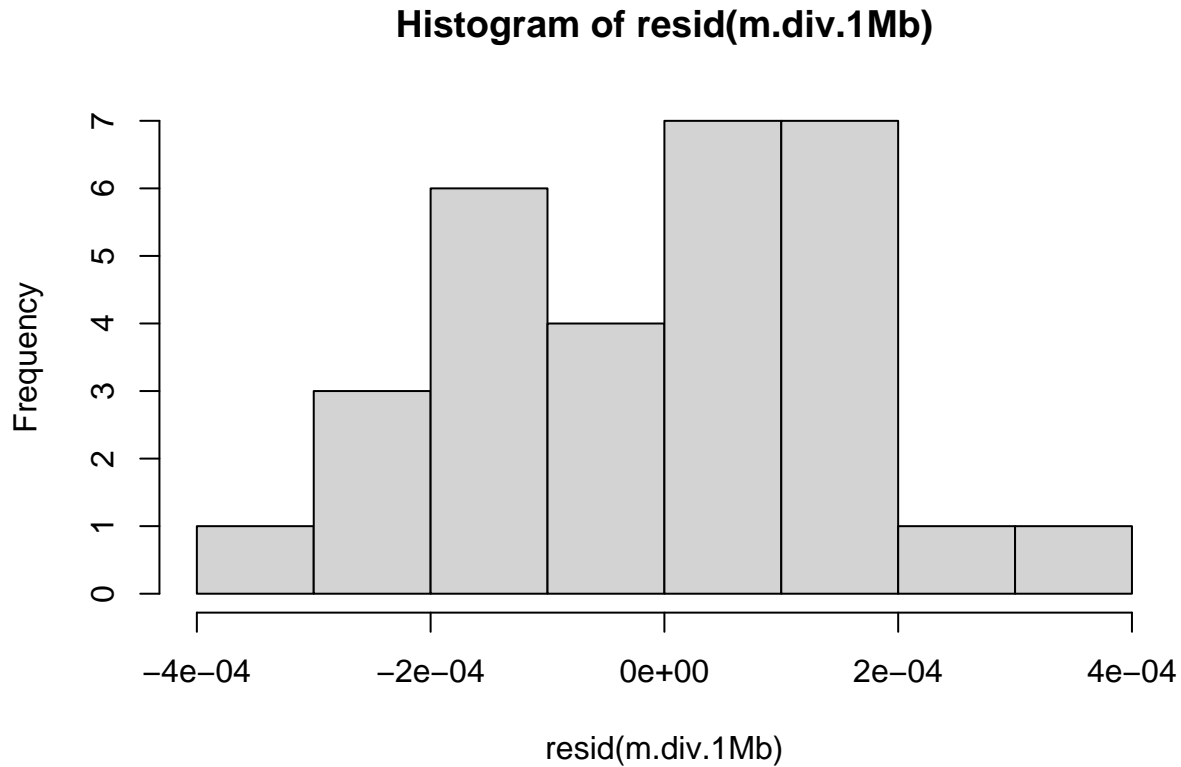
##
## Durbin-Watson test
##
## data: m.div.1Mb
## DW = 1.956, p-value = 0.4515
## alternative hypothesis: true autocorrelation is greater than 0

hmcetest(m.div.1Mb)

##
## Harrison-McCabe test

```

```
##
## data:  m.div.1Mb
## HMC = 0.62113, p-value = 0.8
hist(resid(m.div.1Mb))
```



```
summary(m.div.1Mb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.1Mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.104e-04 -1.172e-04  4.625e-05  1.282e-04  3.696e-04
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.055e-02  3.357e-05  612.052 < 2e-16 ***
## thetaC       1.305e+00  7.159e-03  182.260 < 2e-16 ***
## rhoC         6.955e-02  5.357e-02   1.298  0.20607
## tmrcaC       2.808e-02  2.754e-03  10.193 2.18e-10 ***
## thetaC:tmrcaC 2.260e+00  6.739e-01   3.353  0.00255 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.000178 on 25 degrees of freedom
## Multiple R-squared: 0.9993, Adjusted R-squared: 0.9992
## F-statistic: 9504 on 4 and 25 DF, p-value: < 2.2e-16
```

```
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
```

```
r2.sim.1Mb[1, 6] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) / anova.diversity$VarExp[5]
r2.sim.1Mb[2, 6] <- anova.diversity$VarExp[1] * 100
r2.sim.1Mb[3, 6] <- anova.diversity$VarExp[2] * 100
r2.sim.1Mb[4, 6] <- anova.diversity$VarExp[3] * 100
r2.sim.1Mb[5, 6] <- anova.diversity$VarExp[4] * 100
```

### 1.3.7 Replicate 7

```
rep_7.pi.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_7/rep_7.pi.1Mb.bedgraph", header = TRUE, as.is = TRUE)
rep_7.tmrca.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_7/sim.tmrca.1M.map", header = TRUE, as.is = TRUE)
```

```
sim.lands.1Mb <- as.data.frame(cbind(rep_7.pi.1Mb$pi, sim.theta.1Mb$sim, sim.rho.1Mb$sim, rep_7.tmrca.1Mb$tmrca))
names(sim.lands.1Mb) <- c("diversity", "theta", "rho", "tmrca")
```

```
cor.test(~theta+tmrca, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 5750, p-value = 0.135
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.2791991
```

```
cor.test(~theta+rho, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 5654, p-value = 0.1684
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.257842
```

```
cor.test(~rho+tmrca, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 3620, p-value = 0.3013
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
```

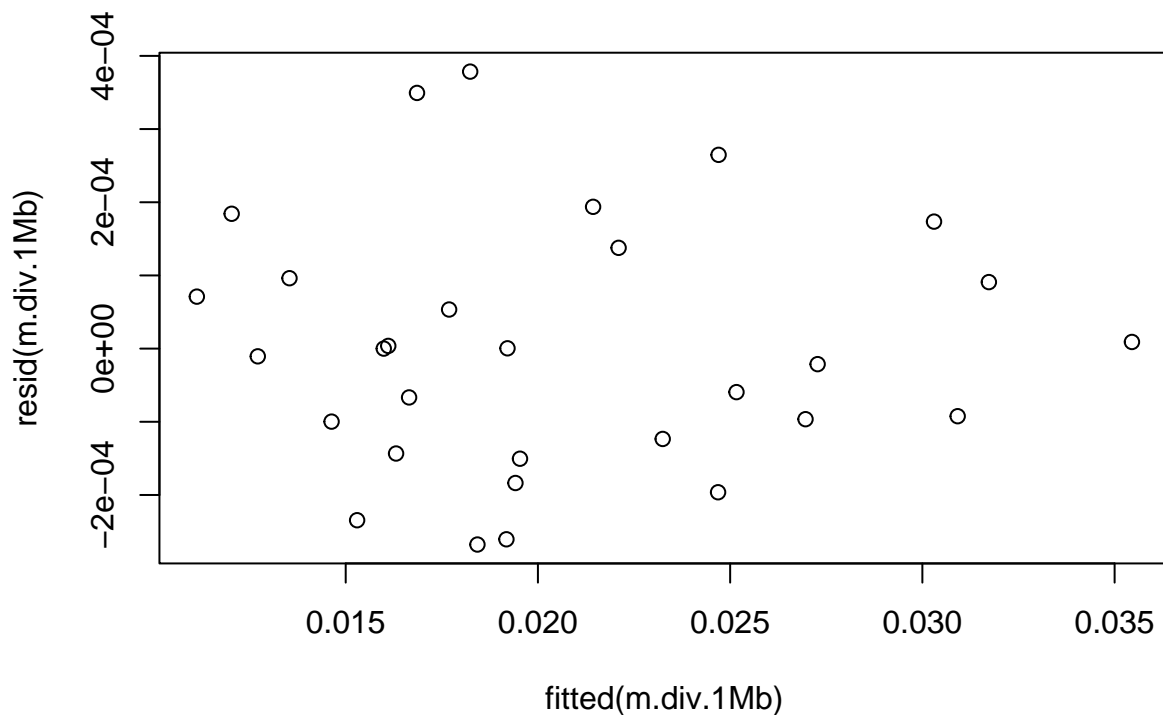


```
## 0.1946607
# centering
sim.lands.1Mb$thetaC <- sim.lands.1Mb$theta - mean(sim.lands.1Mb$theta)
sim.lands.1Mb$tmrcaC <- sim.lands.1Mb$tmrca - mean(sim.lands.1Mb$tmrca)
sim.lands.1Mb$rhoC <- sim.lands.1Mb$rho - mean(sim.lands.1Mb$rho)

sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)

# for merging:
sim.lands.1Mb.rep7 <- sim.lands.1Mb
sim.lands.1Mb.rep7$Replicate <- 7

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.1Mb)
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))
```



```
dwtest(m.div.1Mb)

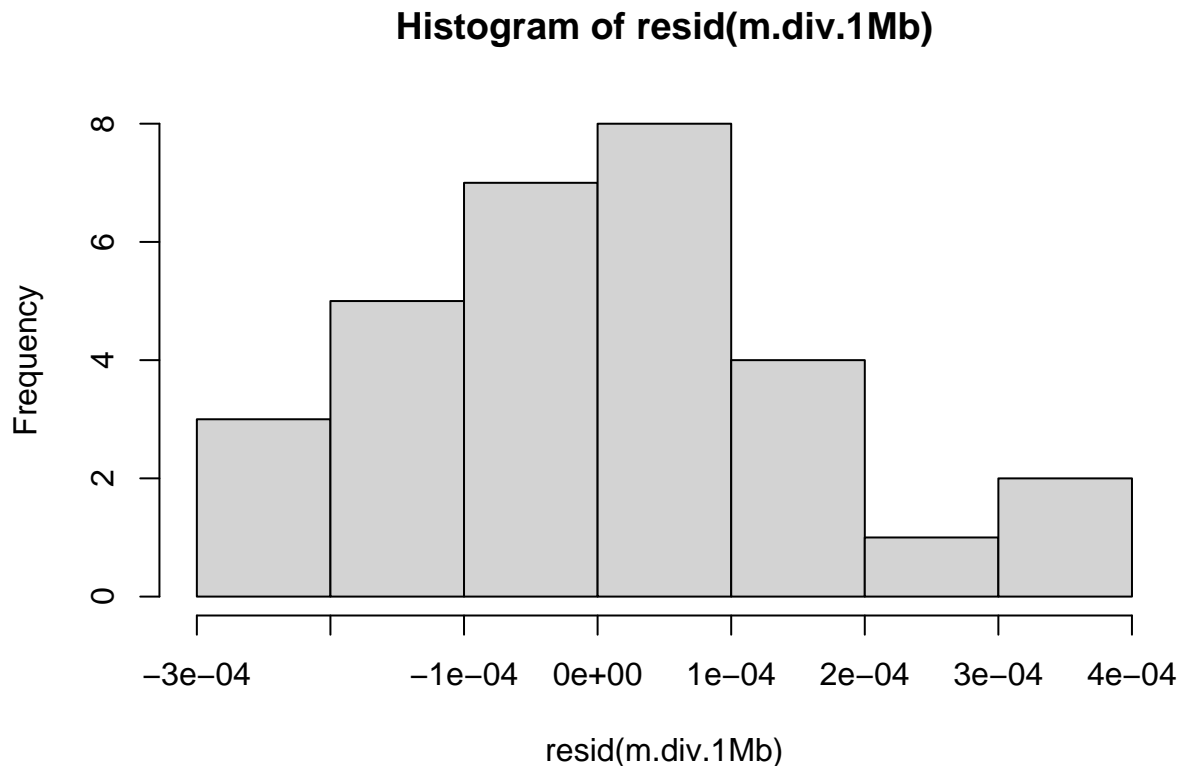
##
## Durbin-Watson test
##
## data: m.div.1Mb
## DW = 1.9852, p-value = 0.4718
## alternative hypothesis: true autocorrelation is greater than 0

hmctest(m.div.1Mb)

##
```

```
## Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.65923, p-value = 0.903
```

```
hist(resid(m.div.1Mb))
```



```
summary(m.div.1Mb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.1Mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.677e-04 -1.175e-04 -5.270e-06  9.488e-05  3.785e-04
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.060e-02  3.433e-05  599.990 < 2e-16 ***
## thetaC       1.322e+00  7.722e-03  171.218 < 2e-16 ***
## rhoC        -5.037e-02  5.447e-02  -0.925  0.363900
## tmrcaC       2.856e-02  1.841e-03  15.507  2.47e-14 ***
## thetaC:tmrcaC 1.507e+00  3.301e-01   4.565  0.000115 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.0001836 on 25 degrees of freedom
## Multiple R-squared: 0.9993, Adjusted R-squared: 0.9991
## F-statistic: 8510 on 4 and 25 DF, p-value: < 2.2e-16

anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.1Mb[1, 7] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] +
r2.sim.1Mb[2, 7] <- anova.diversity$VarExp[1] * 100
r2.sim.1Mb[3, 7] <- anova.diversity$VarExp[2] * 100
r2.sim.1Mb[4, 7] <- anova.diversity$VarExp[3] * 100
r2.sim.1Mb[5, 7] <- anova.diversity$VarExp[4] * 100
```

### 1.3.8 Replicate 8

```
rep_8.pi.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_8/rep_8.pi.1Mb.bedgraph", header = 1)
rep_8.tmrca.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_8/sim.tmrca.1M.map", header = 1)

sim.lands.1Mb <- as.data.frame(cbind(rep_8.pi.1Mb$pi, sim.theta.1Mb$sim, sim.rho.1Mb$sim, rep_8.tmrca.1Mb$tmrca))
names(sim.lands.1Mb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 4082, p-value = 0.628
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.09187987
```

```
cor.test(~theta+rho, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: theta and rho
## S = 5654, p-value = 0.1684
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.257842
```

```
cor.test(~rho+tmrca, data = sim.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 4888, p-value = 0.6448
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
```

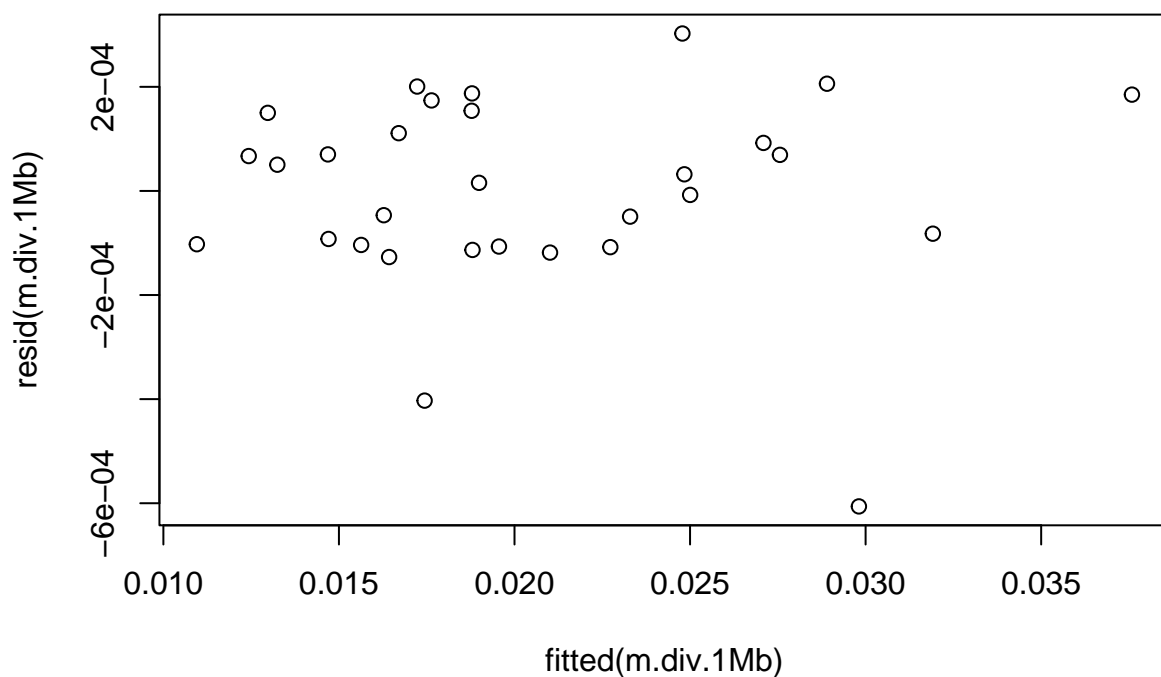
```
##          rho
## -0.08743048

# centering
sim.lands.1Mb$thetaC <- sim.lands.1Mb$theta - mean(sim.lands.1Mb$theta)
sim.lands.1Mb$tmrcaC <- sim.lands.1Mb$tmrca - mean(sim.lands.1Mb$tmrca)
sim.lands.1Mb$rhoC <- sim.lands.1Mb$rho - mean(sim.lands.1Mb$rho)

sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)

# for merging:
sim.lands.1Mb.rep8 <- sim.lands.1Mb
sim.lands.1Mb.rep8$Replicate <- 8

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.1Mb)
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))
```



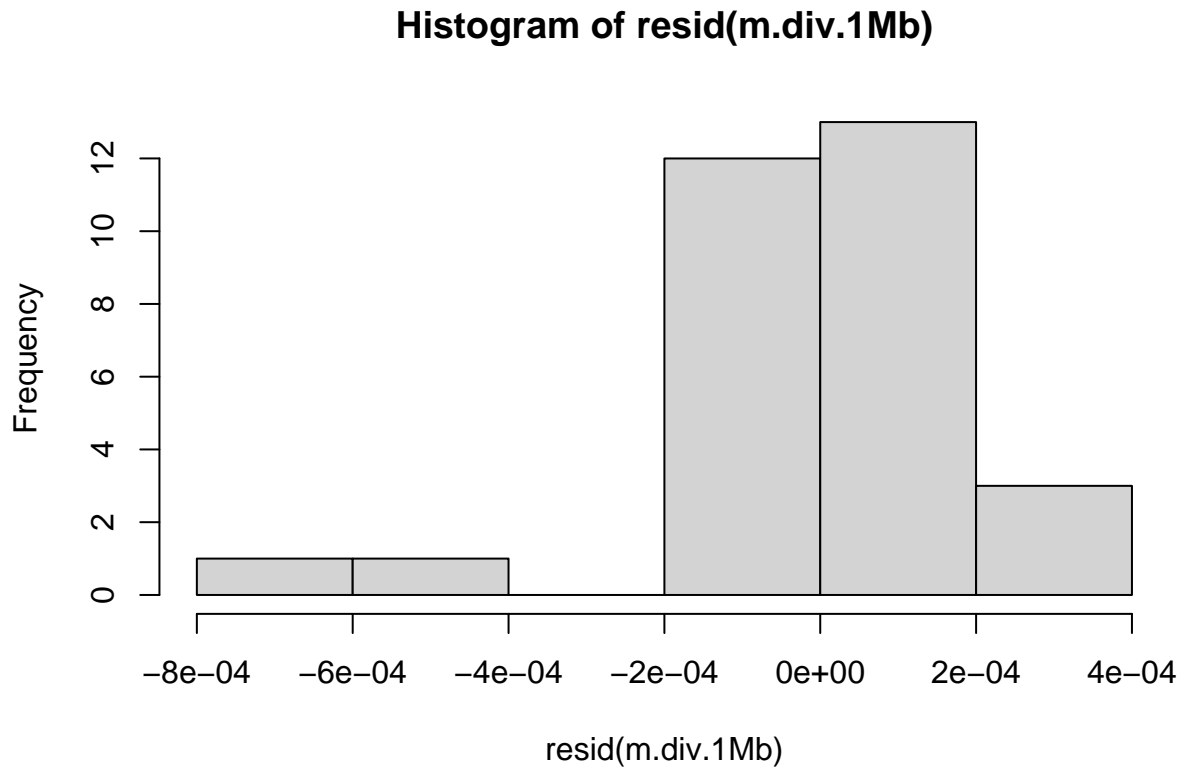
```
dwtest(m.div.1Mb)

##
## Durbin-Watson test
##
## data: m.div.1Mb
## DW = 2.4886, p-value = 0.9089
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.1Mb)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.div.1Mb  
## HMC = 0.6984, p-value = 0.939
```

```
hist(resid(m.div.1Mb))
```



```
summary(m.div.1Mb)
```

```
##  
## Call:  
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,  
##     data = sim.lands.1Mb)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -6.061e-04 -1.034e-04  2.376e-05  1.402e-04  3.024e-04   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)  2.051e-02  3.685e-05  556.403  < 2e-16 ***  
## thetaC       1.294e+00  7.915e-03  163.527  < 2e-16 ***  
## rhoC         2.309e-02  5.876e-02   0.393  0.69767      
## tmrcaC       2.659e-02  2.276e-03  11.682  1.27e-11 ***
```

```
## thetaC:tmrcaC 2.000e+00  5.608e-01   3.567  0.00149 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001996 on 25 degrees of freedom
## Multiple R-squared:  0.9992, Adjusted R-squared:  0.999
## F-statistic: 7462 on 4 and 25 DF,  p-value: < 2.2e-16

anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.1Mb[1, 8] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) / sum(apiss)
r2.sim.1Mb[2, 8] <- anova.diversity$VarExp[1] * 100
r2.sim.1Mb[3, 8] <- anova.diversity$VarExp[2] * 100
r2.sim.1Mb[4, 8] <- anova.diversity$VarExp[3] * 100
r2.sim.1Mb[5, 8] <- anova.diversity$VarExp[4] * 100
```

### 1.3.9 Replicate 9

```
rep_9.pi.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_9/rep_9.pi.1Mb.bedgraph", header = TRUE, as.is = TRUE)
rep_9.tmrca.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_9/sim.tmrca.1M.map", header = TRUE, as.is = TRUE)

sim.lands.1Mb <- as.data.frame(cbind(rep_9.pi.1Mb$pi, sim.theta.1Mb$sim, sim.rho.1Mb$sim, rep_9.tmrca.1Mb$tmrca))
names(sim.lands.1Mb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.1Mb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  theta and tmrca
## S = 3806, p-value = 0.4171
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.1532814

cor.test(~theta+rho, data = sim.lands.1Mb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  theta and rho
## S = 5654, p-value = 0.1684
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.257842

cor.test(~rho+tmrca, data = sim.lands.1Mb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  rho and tmrca
```

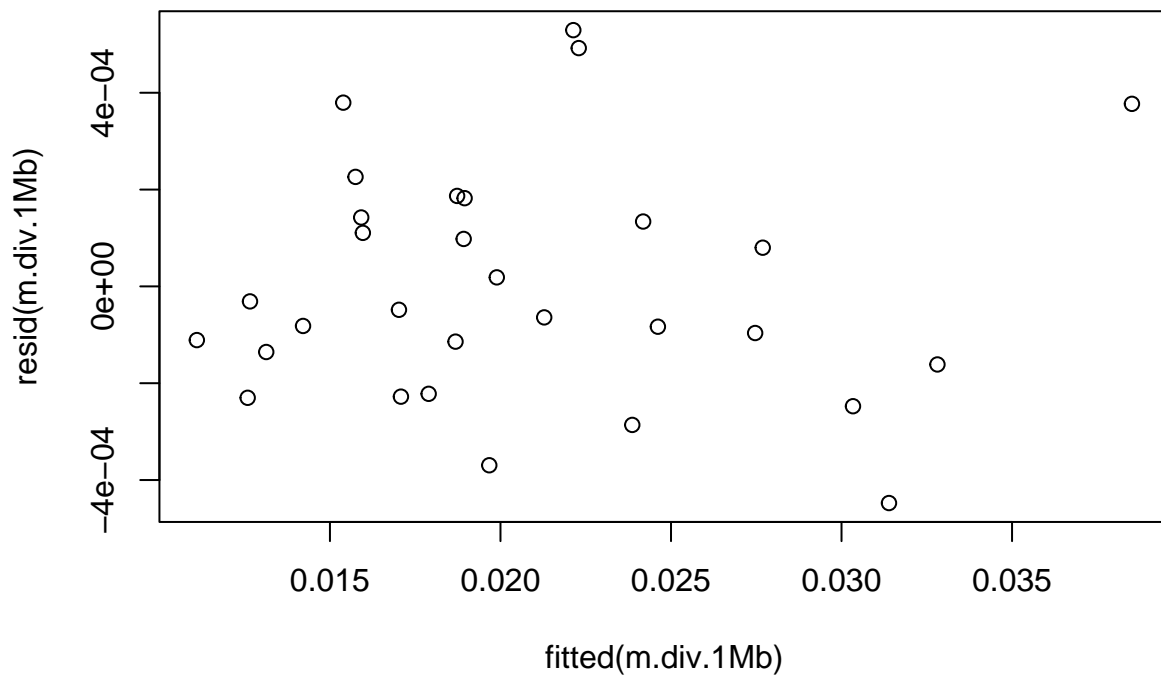
```
## S = 4996, p-value = 0.5562
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.1114572

# centering
sim.lands.1Mb$thetaC <- sim.lands.1Mb$theta - mean(sim.lands.1Mb$theta)
sim.lands.1Mb$tmrcaC <- sim.lands.1Mb$tmrca - mean(sim.lands.1Mb$tmrca)
sim.lands.1Mb$rhoC <- sim.lands.1Mb$rho - mean(sim.lands.1Mb$rho)

sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)

# for merging:
sim.lands.1Mb.rep9 <- sim.lands.1Mb
sim.lands.1Mb.rep9$Replicate <- 9

m.div.1Mb <- lm(多样性 ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.1Mb)
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))
```



```
dwtest(m.div.1Mb)
```

```
##
## Durbin-Watson test
##
## data: m.div.1Mb
## DW = 2.2409, p-value = 0.7322
```

```
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.1Mb)
```

```
##
```

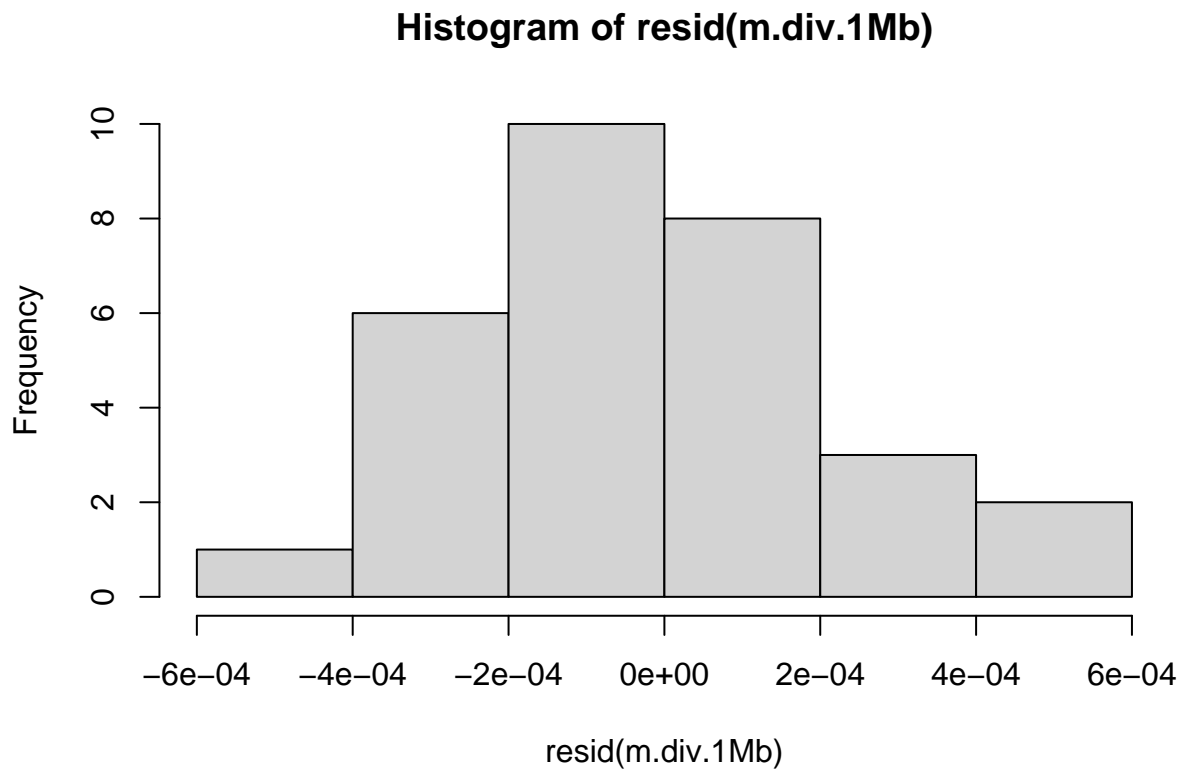
```
## Harrison-McCabe test
```

```
##
```

```
## data: m.div.1Mb
```

```
## HMC = 0.5612, p-value = 0.653
```

```
hist(resid(m.div.1Mb))
```



```
summary(m.div.1Mb)
```

```
##
```

```
## Call:
```

```
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
```

```
## data = sim.lands.1Mb)
```

```
##
```

```
## Residuals:
```

##	Min	1Q	Median	3Q	Max
##	-0.0004475	-0.0001548	-0.0000562	0.0001403	0.0005293

```
##
```

```
## Coefficients:
```

##		Estimate	Std. Error	t value	Pr(> t )
##	(Intercept)	2.064e-02	5.033e-05	410.041	< 2e-16 ***
##	thetaC	1.316e+00	1.246e-02	105.593	< 2e-16 ***
##	rhoC	3.532e-02	8.105e-02	0.436	0.6667



```
## tmrcaC          2.428e-02  2.655e-03   9.145 1.89e-09 ***
## thetaC:tmrcaC  1.211e+00  5.581e-01   2.171  0.0396 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.000263 on 25 degrees of freedom
## Multiple R-squared:  0.9986, Adjusted R-squared:  0.9984
## F-statistic:  4623 on 4 and 25 DF,  p-value: < 2.2e-16

anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.sim.1Mb[1, 9] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] +
r2.sim.1Mb[2, 9] <- anova.diversity$VarExp[1] * 100
r2.sim.1Mb[3, 9] <- anova.diversity$VarExp[2] * 100
r2.sim.1Mb[4, 9] <- anova.diversity$VarExp[3] * 100
r2.sim.1Mb[5, 9] <- anova.diversity$VarExp[4] * 100
```

### 1.3.10 Replicate 10

```
rep_10.pi.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_10/rep_10.pi.1Mb.bedgraph", header = 1)
rep_10.tmrca.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_10/sim.tmrca.1M.map", header = 1)

sim.lands.1Mb <- as.data.frame(cbind(rep_10.pi.1Mb$pi, sim.theta.1Mb$sim, sim.rho.1Mb$sim, rep_10.tmrca.1Mb$tmrca))
names(sim.lands.1Mb) <- c("diversity", "theta", "rho", "tmrca")

cor.test(~theta+tmrca, data = sim.lands.1Mb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  theta and tmrca
## S = 4496, p-value = 1
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.0002224694

cor.test(~theta+rho, data = sim.lands.1Mb, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  theta and rho
## S = 5654, p-value = 0.1684
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.257842

cor.test(~rho+tmrca, data = sim.lands.1Mb, method = "spearman")

##
## Spearman's rank correlation rho
##
```

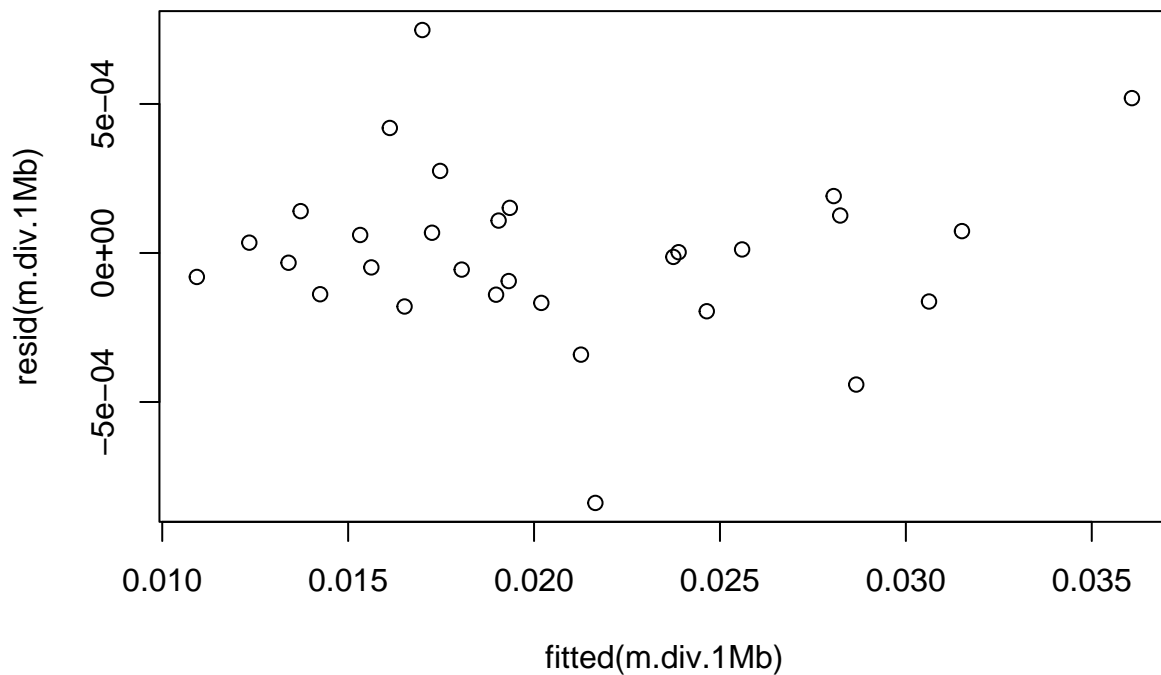
```
## data: rho and tmrca
## S = 5530, p-value = 0.22
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.2302558

# centering
sim.lands.1Mb$thetaC <- sim.lands.1Mb$theta - mean(sim.lands.1Mb$theta)
sim.lands.1Mb$tmrcaC <- sim.lands.1Mb$tmrca - mean(sim.lands.1Mb$tmrca)
sim.lands.1Mb$rhoC <- sim.lands.1Mb$rho - mean(sim.lands.1Mb$rho)

sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)

# for merging:
sim.lands.1Mb.rep10 <- sim.lands.1Mb
sim.lands.1Mb.rep10$Replicate <- 10

m.div.1Mb <- lm(多样性 ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = sim.lands.1Mb)
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))
```



```
dwtest(m.div.1Mb)
```

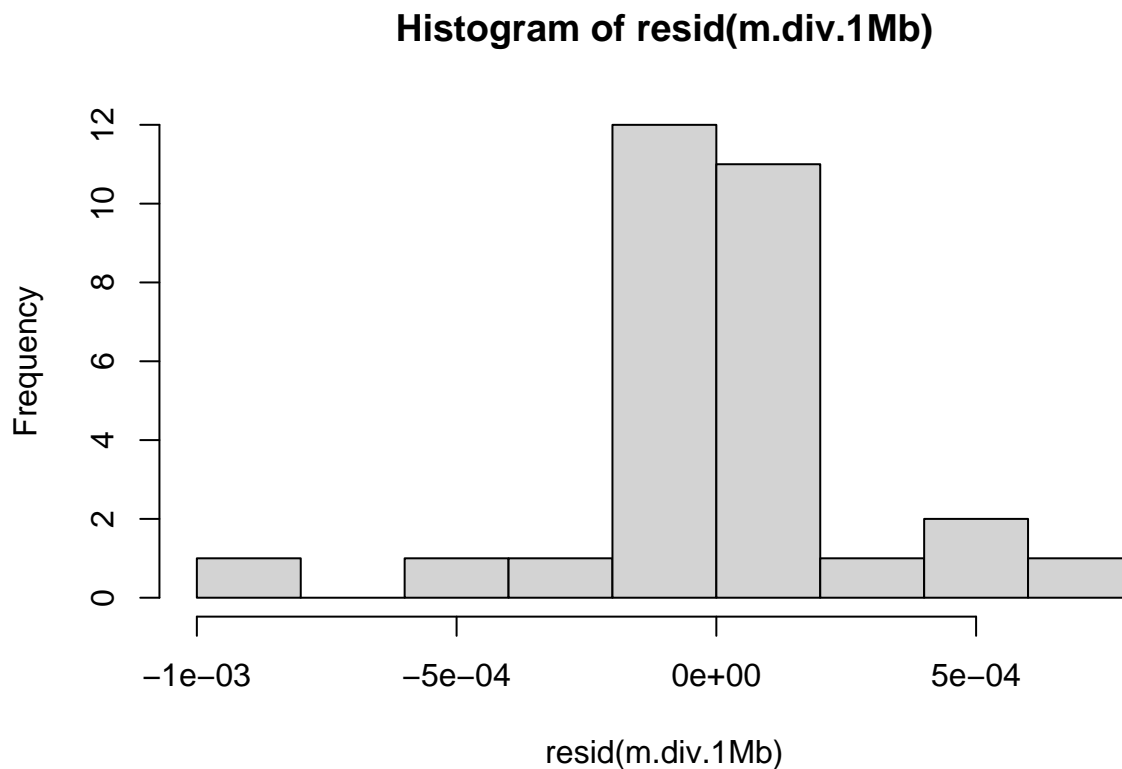
```
##
## Durbin-Watson test
##
## data: m.div.1Mb
```

```
## DW = 2.2351, p-value = 0.7114
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.1Mb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.50343, p-value = 0.505
```

```
hist(resid(m.div.1Mb))
```



```
summary(m.div.1Mb)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,
##     data = sim.lands.1Mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.386e-04 -1.396e-04 -5.280e-06  1.214e-04  7.480e-04
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.066e-02  5.729e-05  360.639 < 2e-16 ***
## thetaC       1.308e+00  1.273e-02  102.725 < 2e-16 ***
```

```
## rhoC          -9.102e-02  9.370e-02  -0.971  0.34066
## tmrcaC         2.340e-02  3.129e-03   7.479 7.84e-08 ***
## thetaC:tmrcaC  1.965e+00  5.383e-01   3.651 0.00121 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0003101 on 25 degrees of freedom
## Multiple R-squared:  0.9979, Adjusted R-squared:  0.9976
## F-statistic: 2967 on 4 and 25 DF,  p-value: < 2.2e-16
```

```
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

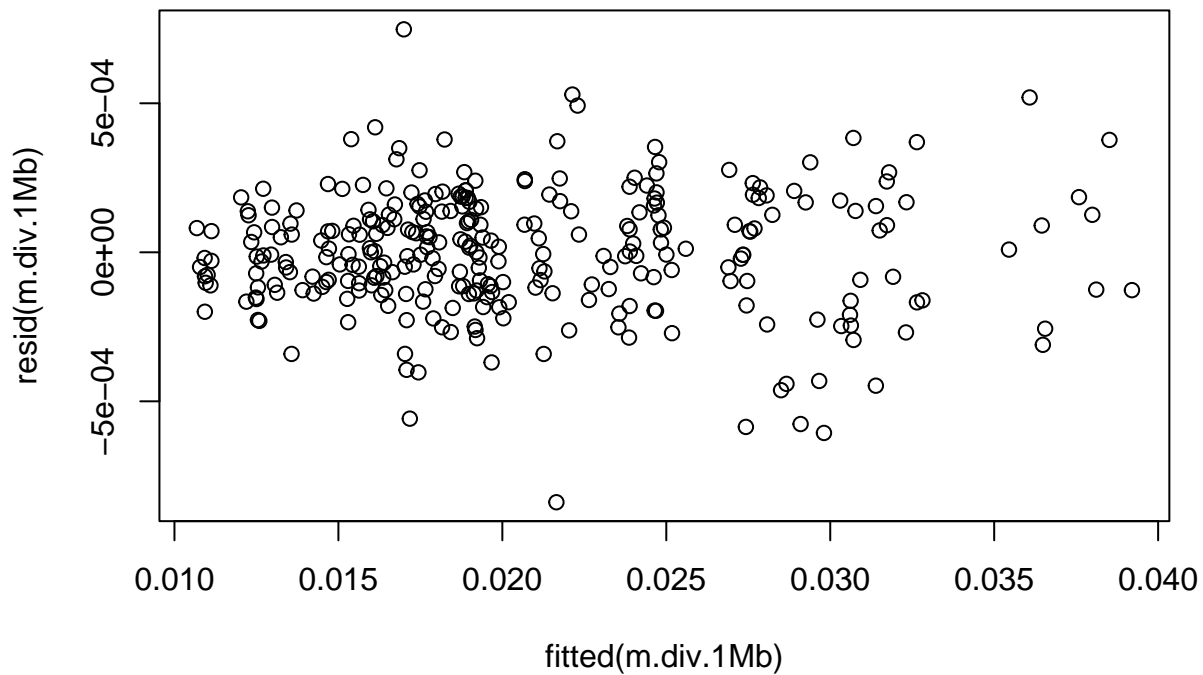
r2.sim.1Mb[1, 10] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.sim.1Mb[2, 10] <- anova.diversity$VarExp[1] * 100
r2.sim.1Mb[3, 10] <- anova.diversity$VarExp[2] * 100
r2.sim.1Mb[4, 10] <- anova.diversity$VarExp[3] * 100
r2.sim.1Mb[5, 10] <- anova.diversity$VarExp[4] * 100

r2.sim.1Mb$average <- rowMeans(r2.sim.1Mb)
r2.sim.1Mb <- transform(r2.sim.1Mb, sd=apply(r2.sim.1Mb, 1, sd, na.rm = TRUE))
```

### 1.3.11 all replicates:

```
sim.lands.1Mb.all <- rbind(sim.lands.1Mb.rep1, sim.lands.1Mb.rep2, sim.lands.1Mb.rep3, sim.lands.1Mb.rep4)

m.div.1Mb <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)*as.factor(Replicate), data = sim.lands.1Mb.all)
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))
```



```
dwtest(m.div.1Mb)
```

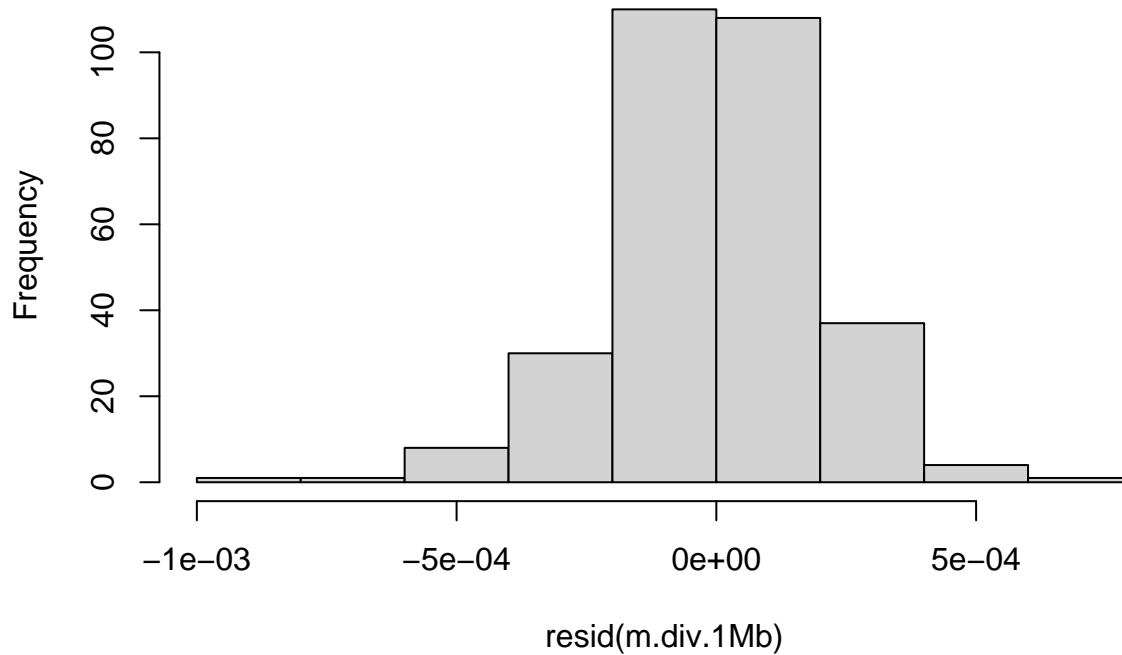
```
##
## Durbin-Watson test
##
## data: m.div.1Mb
## DW = 2.2218, p-value = 0.8988
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.div.1Mb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.44776, p-value = 0.105
```

```
hist(resid(m.div.1Mb))
```

## Histogram of resid(m.div.1Mb)



```
m.div.1Mb.2 <- lm(versity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC)*as.factor(Replicate)
m.div.1Mb.3 <- lm(versity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC + rhoC:tmrcaC)*as.f
```

```
AIC(m.div.1Mb, m.div.1Mb.2, m.div.1Mb.3)
```

```
##           df      AIC
## m.div.1Mb   51 -4150.862
## m.div.1Mb.2 61 -4156.386
## m.div.1Mb.3 71 -4142.940
```

```
summary(m.div.1Mb)
```

```
##
## Call:
## lm(formula = versity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC) *
##     as.factor(Replicate), data = sim.lands.1Mb.all)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.386e-04 -1.236e-04 -2.970e-06  1.379e-04  7.480e-04
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.067e-02  4.041e-05  511.490 < 2e-16 ***
## thetaC        1.313e+00  9.414e-03  139.502 < 2e-16 ***
## rhoC         -2.472e-03  6.736e-02  -0.037  0.97075
## tmrcaC        2.602e-02  2.818e-03   9.235 < 2e-16 ***
```

```

## as.factor(Replicate)2      -1.518e-04  5.745e-05  -2.642  0.00877 **
## as.factor(Replicate)3      3.579e-05  5.991e-05   0.597  0.55082
## as.factor(Replicate)4     -1.782e-04  5.733e-05  -3.108  0.00210 **
## as.factor(Replicate)5     -7.055e-05  5.719e-05  -1.234  0.21845
## as.factor(Replicate)6     -1.201e-04  5.811e-05  -2.066  0.03981 *
## as.factor(Replicate)7     -7.130e-05  5.783e-05  -1.233  0.21877
## as.factor(Replicate)8     -1.625e-04  5.748e-05  -2.827  0.00507 **
## as.factor(Replicate)9     -3.009e-05  5.854e-05  -0.514  0.60771
## as.factor(Replicate)10    -8.475e-06  5.749e-05  -0.147  0.88293
## thetaC:tmrcaC             1.182e+00  6.119e-01   1.931  0.05458 .
## thetaC:as.factor(Replicate)2 -2.794e-02  1.301e-02  -2.148  0.03269 *
## thetaC:as.factor(Replicate)3  1.288e-02  1.336e-02   0.964  0.33591
## thetaC:as.factor(Replicate)4 -2.283e-02  1.304e-02  -1.751  0.08119 .
## thetaC:as.factor(Replicate)5 -7.230e-03  1.301e-02  -0.556  0.57902
## thetaC:as.factor(Replicate)6 -8.429e-03  1.296e-02  -0.651  0.51596
## thetaC:as.factor(Replicate)7  8.962e-03  1.324e-02   0.677  0.49902
## thetaC:as.factor(Replicate)8 -1.890e-02  1.287e-02  -1.469  0.14321
## thetaC:as.factor(Replicate)9  2.617e-03  1.409e-02   0.186  0.85283
## thetaC:as.factor(Replicate)10 -5.483e-03  1.308e-02  -0.419  0.67554
## rhoC:as.factor(Replicate)2  -4.382e-02  9.469e-02  -0.463  0.64390
## rhoC:as.factor(Replicate)3   6.472e-02  9.463e-02   0.684  0.49461
## rhoC:as.factor(Replicate)4   1.256e-02  9.506e-02   0.132  0.89496
## rhoC:as.factor(Replicate)5  -3.248e-02  9.395e-02  -0.346  0.72985
## rhoC:as.factor(Replicate)6   7.202e-02  9.475e-02   0.760  0.44789
## rhoC:as.factor(Replicate)7  -4.790e-02  9.406e-02  -0.509  0.61103
## rhoC:as.factor(Replicate)8   2.556e-02  9.372e-02   0.273  0.78527
## rhoC:as.factor(Replicate)9   3.779e-02  9.587e-02   0.394  0.69378
## rhoC:as.factor(Replicate)10 -8.855e-02  9.493e-02  -0.933  0.35182
## tmrcaC:as.factor(Replicate)2 -1.216e-03  3.767e-03  -0.323  0.74708
## tmrcaC:as.factor(Replicate)3  4.688e-03  3.802e-03   1.233  0.21876
## tmrcaC:as.factor(Replicate)4 -2.382e-03  3.635e-03  -0.655  0.51300
## tmrcaC:as.factor(Replicate)5 -8.804e-04  3.747e-03  -0.235  0.81440
## tmrcaC:as.factor(Replicate)6  2.056e-03  4.435e-03   0.463  0.64343
## tmrcaC:as.factor(Replicate)7  2.536e-03  3.587e-03   0.707  0.48015
## tmrcaC:as.factor(Replicate)8  5.679e-04  3.783e-03   0.150  0.88078
## tmrcaC:as.factor(Replicate)9 -1.743e-03  3.596e-03  -0.485  0.62828
## tmrcaC:as.factor(Replicate)10 -2.615e-03  3.596e-03  -0.727  0.46769
## thetaC:tmrcaC:as.factor(Replicate)2 1.169e+00  7.755e-01   1.508  0.13281
## thetaC:tmrcaC:as.factor(Replicate)3 1.411e+00  8.898e-01   1.586  0.11401
## thetaC:tmrcaC:as.factor(Replicate)4 8.460e-01  7.036e-01   1.202  0.23037
## thetaC:tmrcaC:as.factor(Replicate)5 -4.715e-02  8.421e-01  -0.056  0.95539
## thetaC:tmrcaC:as.factor(Replicate)6 1.078e+00  1.038e+00   1.039  0.29994
## thetaC:tmrcaC:as.factor(Replicate)7 3.252e-01  7.299e-01   0.445  0.65635
## thetaC:tmrcaC:as.factor(Replicate)8 8.185e-01  8.725e-01   0.938  0.34910
## thetaC:tmrcaC:as.factor(Replicate)9 2.959e-02  7.714e-01   0.038  0.96944
## thetaC:tmrcaC:as.factor(Replicate)10 7.836e-01  7.226e-01   1.084  0.27923
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0002213 on 250 degrees of freedom
## Multiple R-squared:  0.999, Adjusted R-squared:  0.9988
## F-statistic: 4956 on 49 and 250 DF, p-value: < 2.2e-16

```

## 2 Drosophila-like neutral simulations of chr 2L (Inferred Landscapes)

First batch of 10 replicate datasets simulated using parameters as inferred from real Drosophila data.

Here we present results using genomic landscapes as inferred by iSMC

### 2.1 50 kb scale

```
r2.inf.50kb <- as.data.frame(matrix(ncol = nreps, nrow = 5))
row.names(r2.inf.50kb) <- c("Total", "Theta", "Rho", "TMRCA", "Theta:TMRCA")
colnames(r2.inf.50kb) <- reps

# sim landscapes
sim.rho.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims/dm.sim.rho.50000.txt")
sim.theta.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims/dm.sim.theta.50000.txt")
sim.lands.50kb <- as.data.frame(cbind(sim.theta.50kb$sim, sim.rho.50kb$sim))
names(sim.lands.50kb) <- c("theta", "rho")
```

#### 2.1.1 Replicate 1

```
p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_1/rs.pair_1.", sep = "")
pi.50kb <- read.table(paste(p, "diversity.50kb.bedgraph", sep = ""), header = T)
pi.50kb$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCA.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$avg <- apply(tmrca.50kb[4:ncol(tmrca.50kb)], 1, mean)
rho.50kb <- read.table(paste(p, "rho.50kb.bedgraph", sep = ""), header = T)
theta.50kb <- read.table(paste(p, "theta.50kb.bedgraph", sep = ""), header = T)

inf.lands.50kb.rep_1 <- as.data.frame(cbind(pi.50kb$avg, theta.50kb$sample_mean, rho.50kb$sample_mean,
names(inf.lands.50kb.rep_1) <- c("diversity", "theta", "rho", "tmrca")

cor.test(sim.theta.50kb$sim, theta.50kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.50kb$sim and theta.50kb$sample_mean
## S = 781576, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.9782895

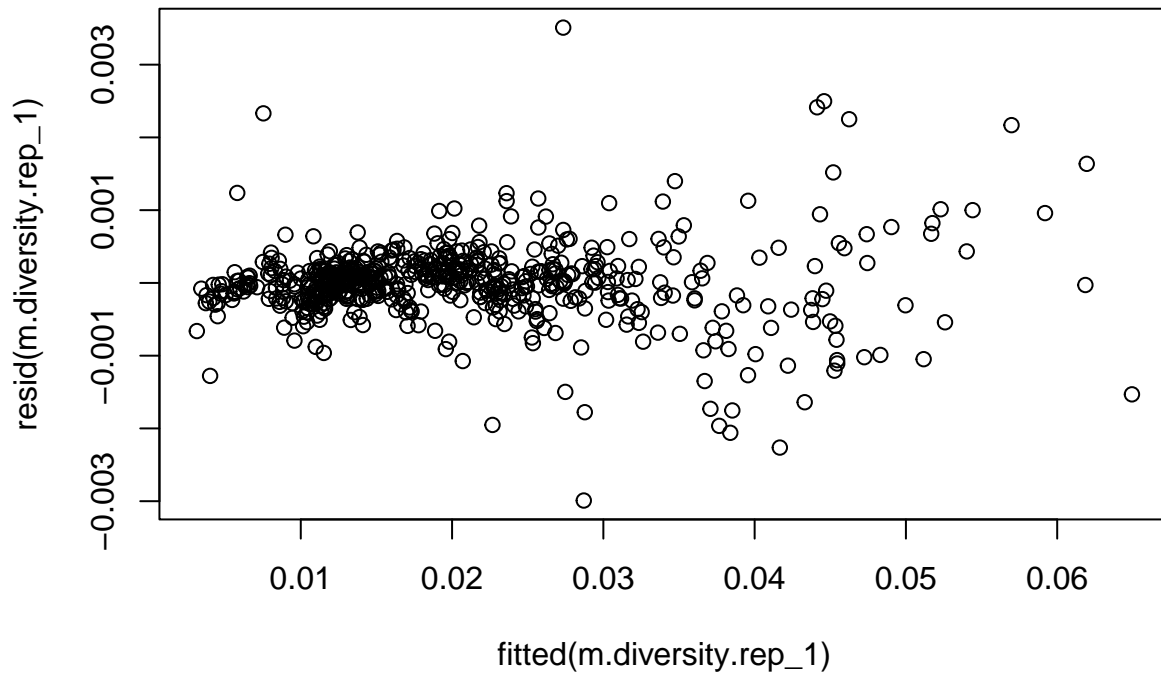
# centering
inf.lands.50kb.rep_1$thetaC <- inf.lands.50kb.rep_1$theta - mean(inf.lands.50kb.rep_1$theta)
inf.lands.50kb.rep_1$tmrcaC <- inf.lands.50kb.rep_1$tmrca - mean(inf.lands.50kb.rep_1$tmrca)
inf.lands.50kb.rep_1$rhoC <- inf.lands.50kb.rep_1$rho - mean(inf.lands.50kb.rep_1$rho)

inf.lands.50kb.rep_1$bin <- 1:nrow(inf.lands.50kb.rep_1)

# for merging:
inf.lands.50kb.rep_1$Replicate <- 1
```



```
m.diversity.rep_1 <- lm(多样性 ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.50kb.rep_1)
plot(resid(m.diversity.rep_1)~fitted(m.diversity.rep_1))
```



```
dwtest(m.diversity.rep_1)
```

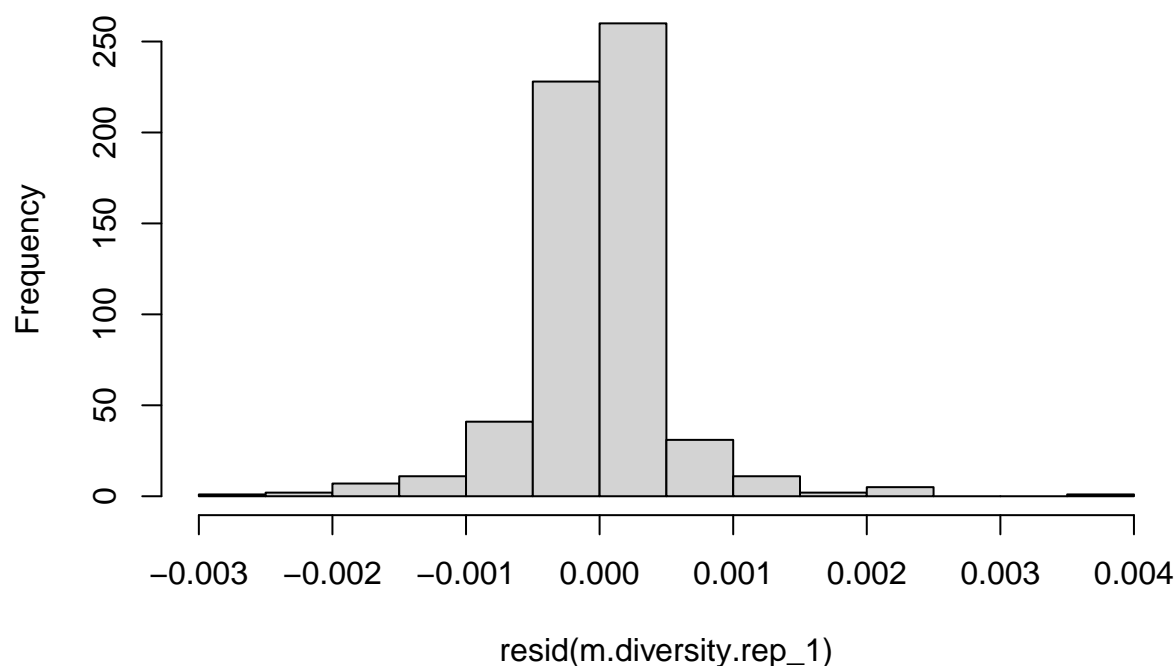
```
##
## Durbin-Watson test
##
## data: m.diversity.rep_1
## DW = 1.4157, p-value = 2.024e-13
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_1)
```

```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_1
## HMC = 0.61568, p-value = 1
```

```
hist(resid(m.diversity.rep_1))
```

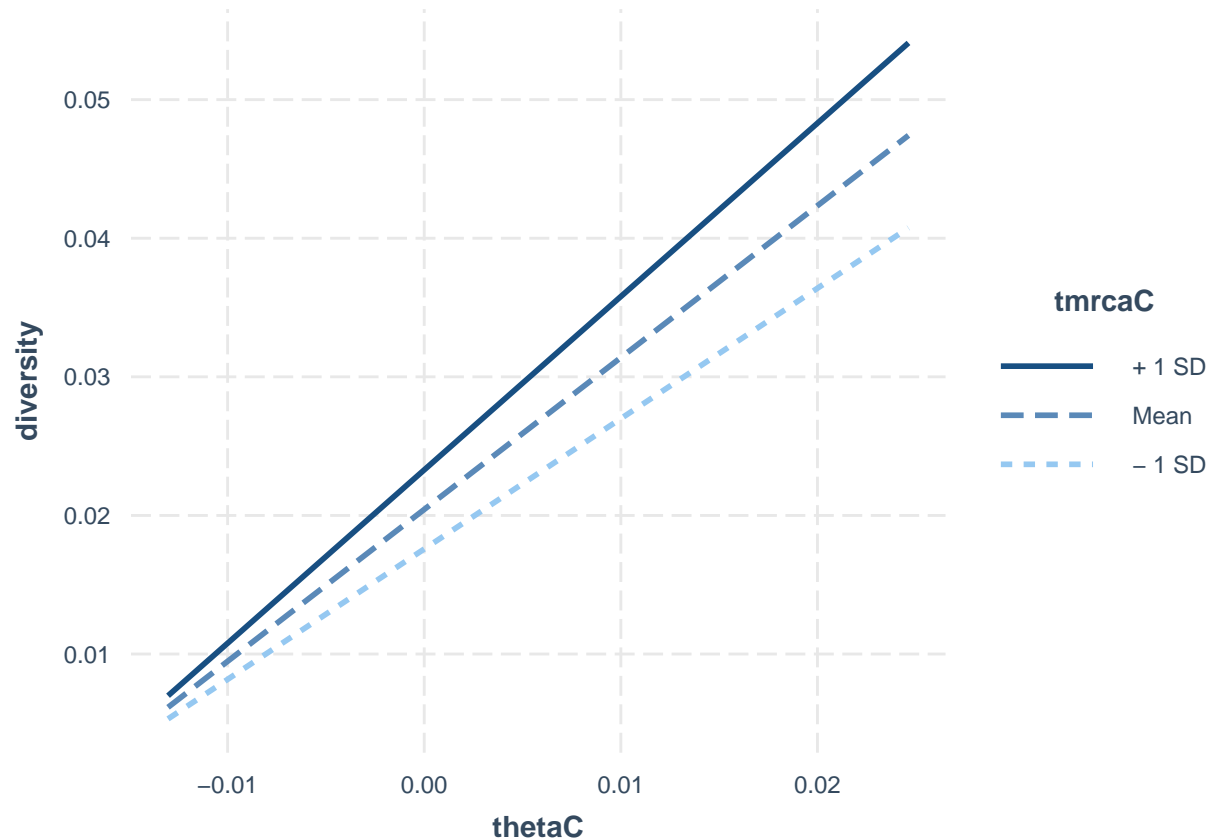
## Histogram of resid(m.diversity.rep\_1)



```
summary(m.diversity.rep_1)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.50kb.rep_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0029908 -0.0002117  0.0000130  0.0001979  0.0035089
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.043e-02  2.236e-05  913.668  <2e-16 ***
## thetaC        1.096e+00  2.364e-03  463.564  <2e-16 ***
## rhoC         -4.461e-03  1.756e-02  -0.254    0.8
## tmrcaC        2.026e-02  1.640e-04  123.483  <2e-16 ***
## thetaC:tmrcaC  1.100e+00  1.587e-02   69.290  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005436 on 595 degrees of freedom
## Multiple R-squared:  0.9977, Adjusted R-squared:  0.9977
## F-statistic: 6.386e+04 on 4 and 595 DF,  p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_1, pred = thetaC, modx= tmrcaC)
```



```
g.rep_1 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.50kb.rep_1, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_1)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.50kb.rep_1
##      AIC      BIC    logLik
## -7367.645 -7336.866 3690.822
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.3212747
##
## Coefficients:
##              Value   Std.Error   t-value p-value
## (Intercept)  0.0204388 0.000031255 653.9355  0.0000
## thetaC       1.0980191 0.002985790 367.7482  0.0000
## tmrcaC       0.0199958 0.000176128 113.5300  0.0000
## rhoC         0.0035408 0.017040672   0.2078  0.8355
## thetaC:tmrcaC 1.0653107 0.017126226  62.2035  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      0.010
## tmrcaC     -0.018 -0.155
## rhoC        -0.003  0.009  0.012
## thetaC:tmrcaC -0.096 -0.101  0.181  0.042
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -5.45341396 -0.37367657  0.05022105  0.36056906  6.43069153
##
## Residual standard error: 0.000544344
## Degrees of freedom: 600 total; 595 residual

vif(g.rep_1)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.030576      1.054551      1.001941      1.041465

g.rep_1.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.50kb.rep_1, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_1.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.50kb.rep_1
##      AIC      BIC    logLik
## -5449.253 -5427.268 2729.626
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.5079669
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept) 0.0206321 0.00021240 97.13730 0.0000
## thetaC      1.1816867 0.01703572 69.36522 0.0000
## rhoC        0.0057369 0.08137302  0.07050 0.9438
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.001  0.003
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.88116215 -0.45332037  0.04478515  0.48762545  4.98107994
##
## Residual standard error: 0.002969561
## Degrees of freedom: 600 total; 597 residual
```

```

anova.diversity <- Anova(m.diversity.rep_1)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.50kb[1, 1] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.inf.50kb[2, 1] <- anova.diversity$VarExp[1] * 100
r2.inf.50kb[3, 1] <- anova.diversity$VarExp[2] * 100
r2.inf.50kb[4, 1] <- anova.diversity$VarExp[3] * 100
r2.inf.50kb[5, 1] <- anova.diversity$VarExp[4] * 100

```

### 2.1.2 Replicate 2

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_2/rs.pair_2.", sep = "")
pi.50kb <- read.table(paste(p, "diversity.50kb.bedgraph", sep = ""), header = T)
pi.50kb$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCA.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$avg <- apply(tmrca.50kb[4:ncol(tmrca.50kb)], 1, mean)
rho.50kb <- read.table(paste(p, "rho.50kb.bedgraph", sep = ""), header = T)
theta.50kb <- read.table(paste(p, "theta.50kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.50kb$sim, theta.50kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.50kb$sim and theta.50kb$sample_mean
## S = 557174, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9845229

inf.lands.50kb.rep_2 <- as.data.frame(cbind(pi.50kb$avg, theta.50kb$sample_mean, rho.50kb$sample_mean,
names(inf.lands.50kb.rep_2) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.50kb.rep_2$thetaC <- inf.lands.50kb.rep_2$theta - mean(inf.lands.50kb.rep_2$theta)
inf.lands.50kb.rep_2$tmrcaC <- inf.lands.50kb.rep_2$tmrca - mean(inf.lands.50kb.rep_2$tmrca)
inf.lands.50kb.rep_2$rhoC <- inf.lands.50kb.rep_2$rho - mean(inf.lands.50kb.rep_2$rho)

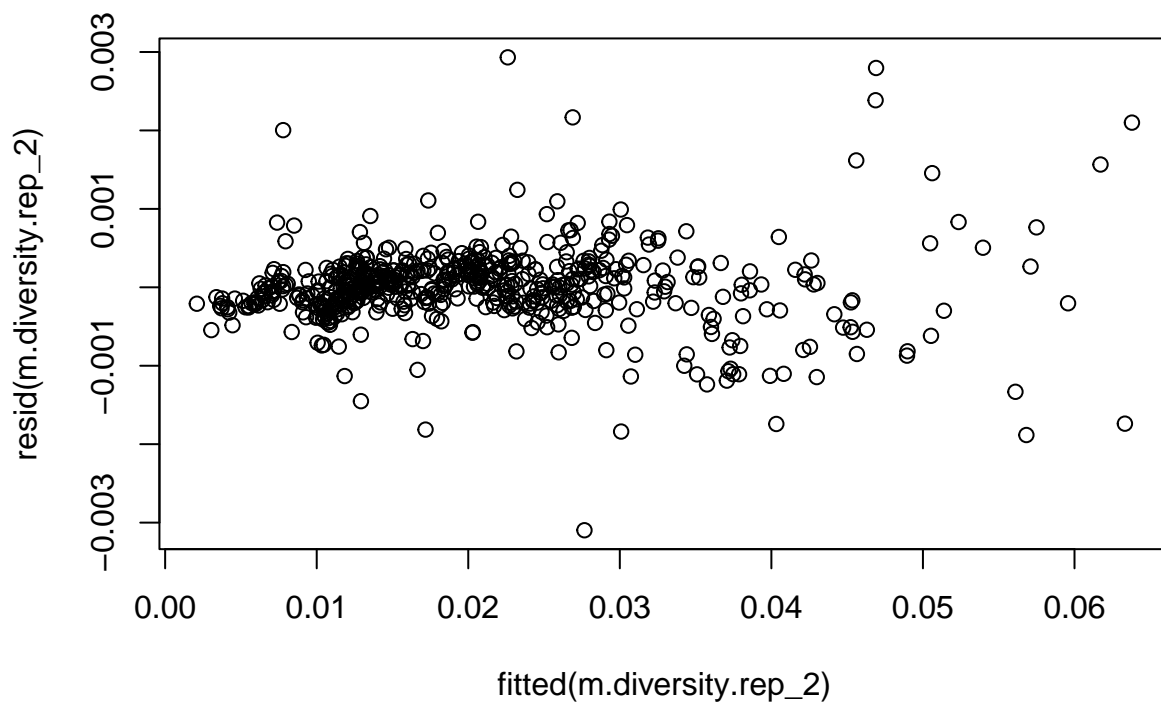
inf.lands.50kb.rep_2$bin <- 1:nrow(inf.lands.50kb.rep_2)

# for merging:
inf.lands.50kb.rep_2$Replicate <- 2

m.diversity.rep_2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.50kb.rep_2)

plot(resid(m.diversity.rep_2)~fitted(m.diversity.rep_2))

```



```
dwtest(m.diversity.rep_2)
```

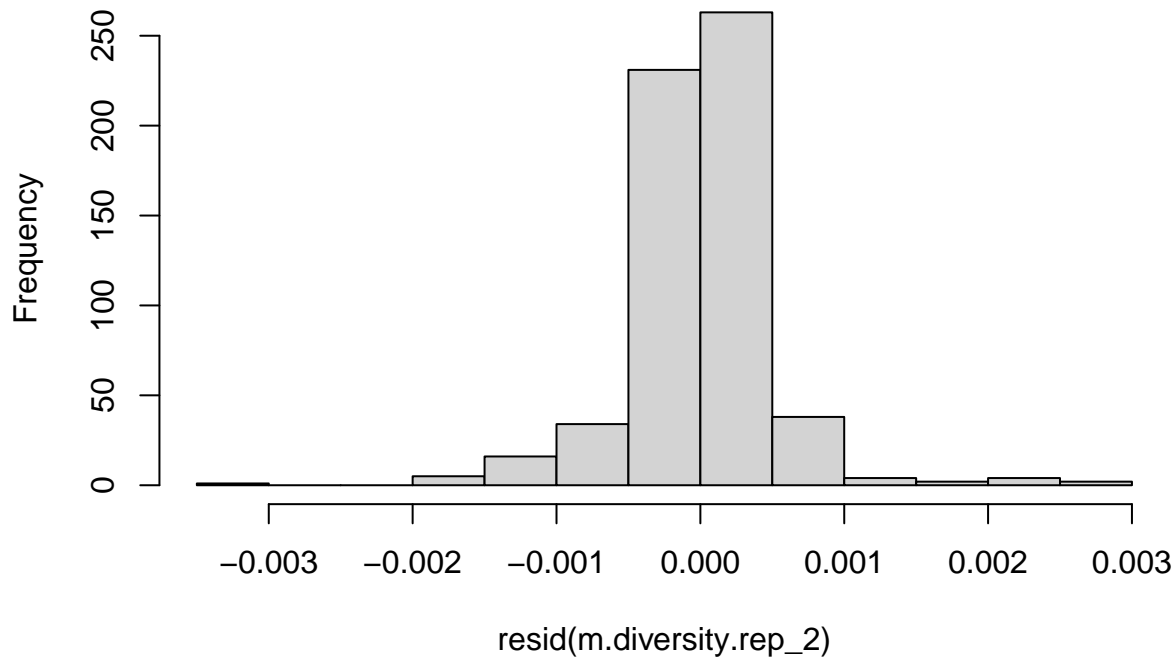
```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_2  
## DW = 1.3609, p-value = 1.173e-15  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_2)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_2  
## HMC = 0.61478, p-value = 1
```

```
hist(resid(m.diversity.rep_2))
```

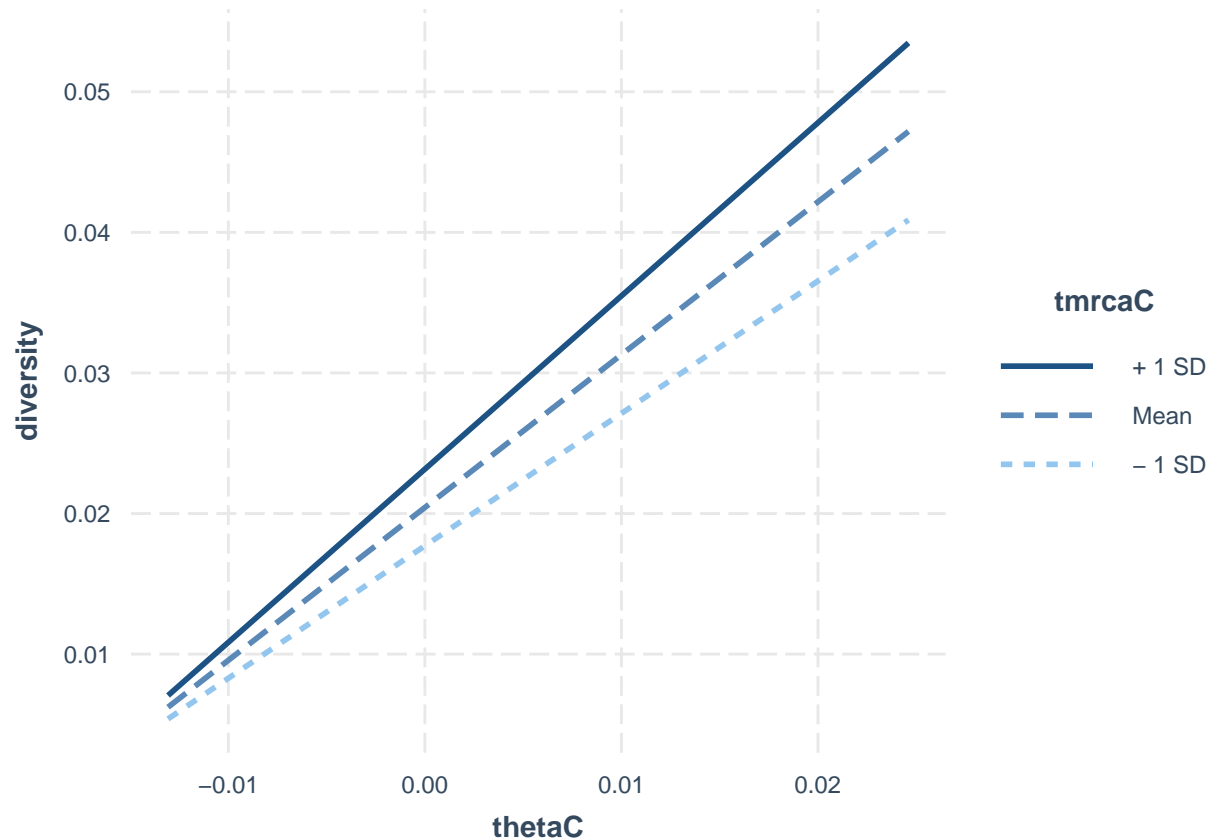
## Histogram of resid(m.diversity.rep\_2)



```
summary(m.diversity.rep_2)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.50kb.rep_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.097e-03 -2.018e-04  1.307e-05  2.192e-04  2.932e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.043e-02  2.069e-05  987.265  <2e-16 ***
## thetaC        1.087e+00  2.167e-03  501.668  <2e-16 ***
## rhoC         -1.167e-02  1.648e-02  -0.708    0.479
## tmrcaC        1.999e-02  1.568e-04  127.438  <2e-16 ***
## thetaC:tmrcaC  1.063e+00  1.365e-02   77.841  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005043 on 595 degrees of freedom
## Multiple R-squared:  0.998, Adjusted R-squared:  0.9979
## F-statistic: 7.242e+04 on 4 and 595 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_2, pred = thetaC, modx= tmrcaC)
```



```
g.rep_2 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.50kb.rep_2, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_2)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.50kb.rep_2
##      AIC      BIC    logLik
## -7474.789 -7444.011 3744.395
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.3693417
##
## Coefficients:
##              Value      Std.Error   t-value p-value
## (Intercept)  0.0204377 0.000030697  665.7833  0.0000
## thetaC       1.0884205 0.002800847  388.6041  0.0000
## tmrcaC       0.0196845 0.000163272  120.5625  0.0000
## rhoC        -0.0167357 0.015485360   -1.0807  0.2802
## thetaC:tmrcaC 1.0175254 0.014295984   71.1756  0.0000
```



```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      0.001
## tmrcaC     -0.004 -0.116
## rhoC       -0.002  0.001  0.001
## thetaC:tmrcaC -0.073 -0.014  0.052  0.045
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -5.967387960 -0.392529054  0.008556582  0.424116285  5.938323518
##
## Residual standard error: 0.0005073101
## Degrees of freedom: 600 total; 595 residual

vif(g.rep_2)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.013673      1.016230      1.002036      1.004808

g.rep_2.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.50kb.rep_2, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_2.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.50kb.rep_2
##      AIC      BIC    logLik
## -5390.022 -5368.037 2700.011
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.3415288
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept) 0.0205980 0.00016691 123.40674 0.0000
## thetaC      1.1278261 0.01545460 72.97670 0.0000
## rhoC      -0.0736804 0.08844426 -0.83307 0.4051
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.001  0.003
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -3.532949546 -0.502347038  0.001576082  0.522732374  6.220744946
##
## Residual standard error: 0.002859674
## Degrees of freedom: 600 total; 597 residual
```

```

anova.diversity <- Anova(m.diversity.rep_2)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.50kb[1, 2] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.50kb[2, 2] <- anova.diversity$VarExp[1] * 100
r2.inf.50kb[3, 2] <- anova.diversity$VarExp[2] * 100
r2.inf.50kb[4, 2] <- anova.diversity$VarExp[3] * 100
r2.inf.50kb[5, 2] <- anova.diversity$VarExp[4] * 100

```

### 2.1.3 Replicate 3

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_3/rs.pair_3.", sep = "")
pi.50kb <- read.table(paste(p, "diversity.50kb.bedgraph", sep = ""), header = T)
pi.50kb$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCA.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$avg <- apply(tmrca.50kb[4:ncol(tmrca.50kb)], 1, mean)
rho.50kb <- read.table(paste(p, "rho.50kb.bedgraph", sep = ""), header = T)
theta.50kb <- read.table(paste(p, "theta.50kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.50kb$sim, theta.50kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.50kb$sim and theta.50kb$sample_mean
## S = 789617, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9780661

inf.lands.50kb.rep_3 <- as.data.frame(cbind(pi.50kb$avg, theta.50kb$sample_mean, rho.50kb$sample_mean,
names(inf.lands.50kb.rep_3) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.50kb.rep_3$thetaC <- inf.lands.50kb.rep_3$theta - mean(inf.lands.50kb.rep_3$theta)
inf.lands.50kb.rep_3$tmrcaC <- inf.lands.50kb.rep_3$tmrca - mean(inf.lands.50kb.rep_3$tmrca)
inf.lands.50kb.rep_3$rhoC <- inf.lands.50kb.rep_3$rho - mean(inf.lands.50kb.rep_3$rho)

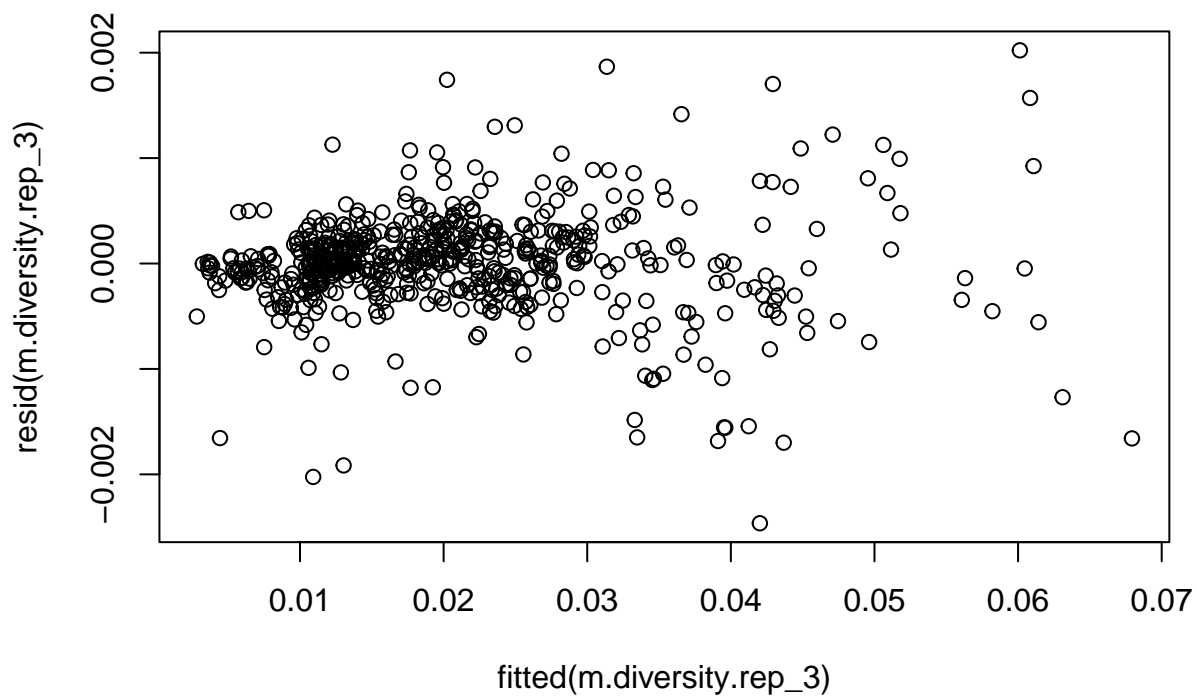
inf.lands.50kb.rep_3$bin <- 1:nrow(inf.lands.50kb.rep_3)

# for merging:
inf.lands.50kb.rep_3$Replicate <- 3

m.diversity.rep_3 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.50kb.rep_3)

plot(resid(m.diversity.rep_3)~fitted(m.diversity.rep_3))

```



```
dwtest(m.diversity.rep_3)
```

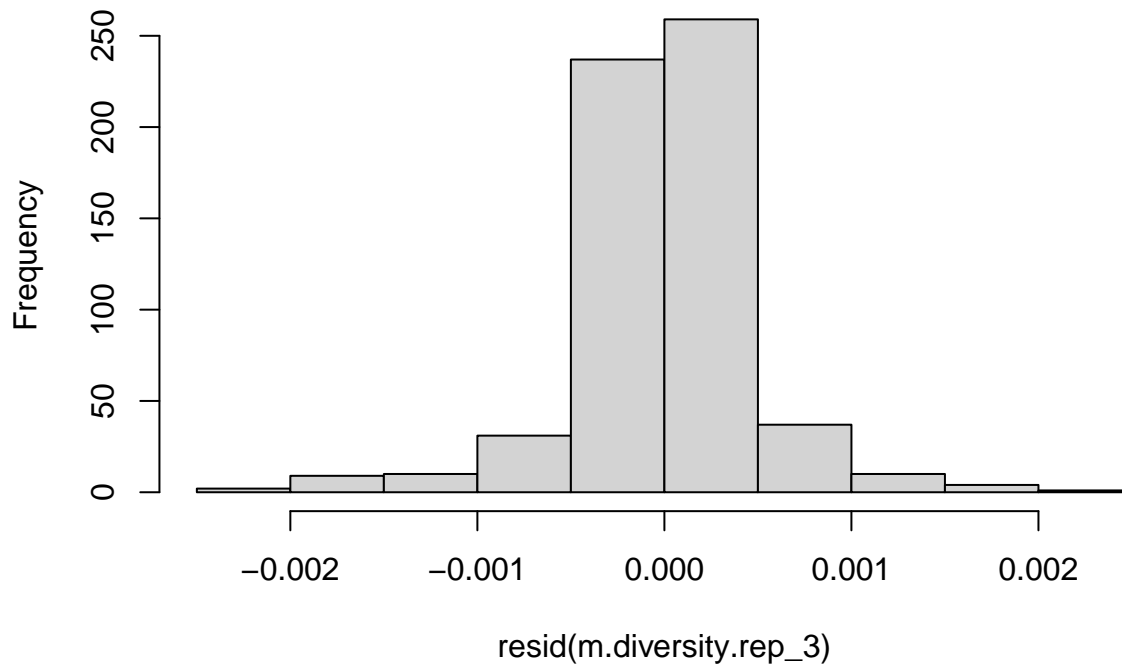
```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_3  
## DW = 1.4619, p-value = 1.133e-11  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcctest(m.diversity.rep_3)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_3  
## HMC = 0.52153, p-value = 0.789
```

```
hist(resid(m.diversity.rep_3))
```

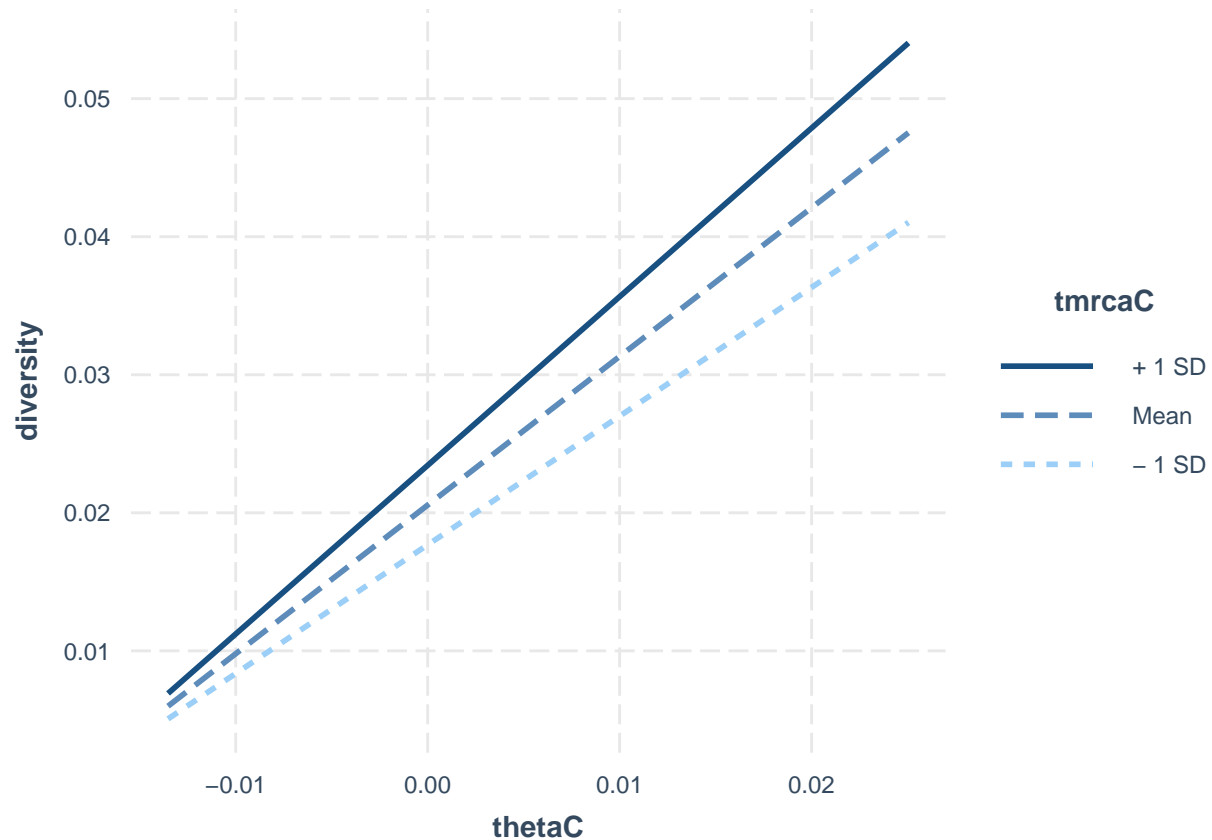
## Histogram of resid(m.diversity.rep\_3)



```
summary(m.diversity.rep_3)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.50kb.rep_3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.464e-03 -1.891e-04  1.021e-05  2.163e-04  2.022e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0205593  0.0000198 1038.471  <2e-16 ***
## thetaC       1.0769710  0.0020324  529.902  <2e-16 ***
## rhoC        -0.0138881  0.0157297   -0.883    0.378
## tmrcaC       0.0202409  0.0001408  143.768  <2e-16 ***
## thetaC:tmrcaC 1.0174596  0.0120993   84.092  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0004822 on 595 degrees of freedom
## Multiple R-squared:  0.9982, Adjusted R-squared:  0.9982
## F-statistic: 8.432e+04 on 4 and 595 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_3, pred = thetaC, modx= tmrcaC)
```



```
g.rep_3 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.50kb.rep_3, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_3)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.50kb.rep_3
##      AIC      BIC    logLik
## -7500.229 -7469.45 3757.114
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.2838818
##
## Coefficients:
##              Value      Std.Error   t-value p-value
## (Intercept)  0.0205629 0.000026507  775.7537  0.0000
## thetaC       1.0781251 0.002506188  430.1853  0.0000
## tmrcaC       0.0201223 0.000150857  133.3870  0.0000
## rhoC        -0.0178210 0.015324110   -1.1629  0.2453
## thetaC:tmrcaC 0.9985038 0.013270574   75.2419  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      0.007
## tmrcaC     -0.006 -0.122
## rhoC        -0.004  0.009  0.029
## thetaC:tmrcaC -0.087 -0.075  0.058  0.056
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -5.02048902 -0.38805559  0.03745702  0.44738982  4.48985459
##
## Residual standard error: 0.0004813794
## Degrees of freedom: 600 total; 595 residual

vif(g.rep_3)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.020199      1.018281      1.004142      1.011269

g.rep_3.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.50kb.rep_3, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_3.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.50kb.rep_3
##      AIC      BIC    logLik
## -5335.691 -5313.707 2672.846
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.4625776
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept) 0.0207420 0.00021384 96.99954 0.0000
## thetaC      1.1398975 0.01768070 64.47130 0.0000
## rhoC       -0.1017491 0.09007408 -1.12962 0.2591
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.001  0.011
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.53093269 -0.44179911 -0.01021147  0.44918244  5.33441211
##
## Residual standard error: 0.003171671
## Degrees of freedom: 600 total; 597 residual
```

```

anova.diversity <- Anova(m.diversity.rep_3)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.50kb[1, 3] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.50kb[2, 3] <- anova.diversity$VarExp[1] * 100
r2.inf.50kb[3, 3] <- anova.diversity$VarExp[2] * 100
r2.inf.50kb[4, 3] <- anova.diversity$VarExp[3] * 100
r2.inf.50kb[5, 3] <- anova.diversity$VarExp[4] * 100

```

#### 2.1.4 Replicate 4

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_4/rs.pair_4.", sep = "")
pi.50kb <- read.table(paste(p, "diversity.50kb.bedgraph", sep = ""), header = T)
pi.50kb$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCA.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$avg <- apply(tmrca.50kb[4:ncol(tmrca.50kb)], 1, mean)
rho.50kb <- read.table(paste(p, "rho.50kb.bedgraph", sep = ""), header = T)
theta.50kb <- read.table(paste(p, "theta.50kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.50kb$sim, theta.50kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.50kb$sim and theta.50kb$sample_mean
## S = 807402, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9775721

inf.lands.50kb.rep_4 <- as.data.frame(cbind(pi.50kb$avg, theta.50kb$sample_mean, rho.50kb$sample_mean,
names(inf.lands.50kb.rep_4) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.50kb.rep_4$thetaC <- inf.lands.50kb.rep_4$theta - mean(inf.lands.50kb.rep_4$theta)
inf.lands.50kb.rep_4$tmrcaC <- inf.lands.50kb.rep_4$tmrca - mean(inf.lands.50kb.rep_4$tmrca)
inf.lands.50kb.rep_4$rhoC <- inf.lands.50kb.rep_4$rho - mean(inf.lands.50kb.rep_4$rho)

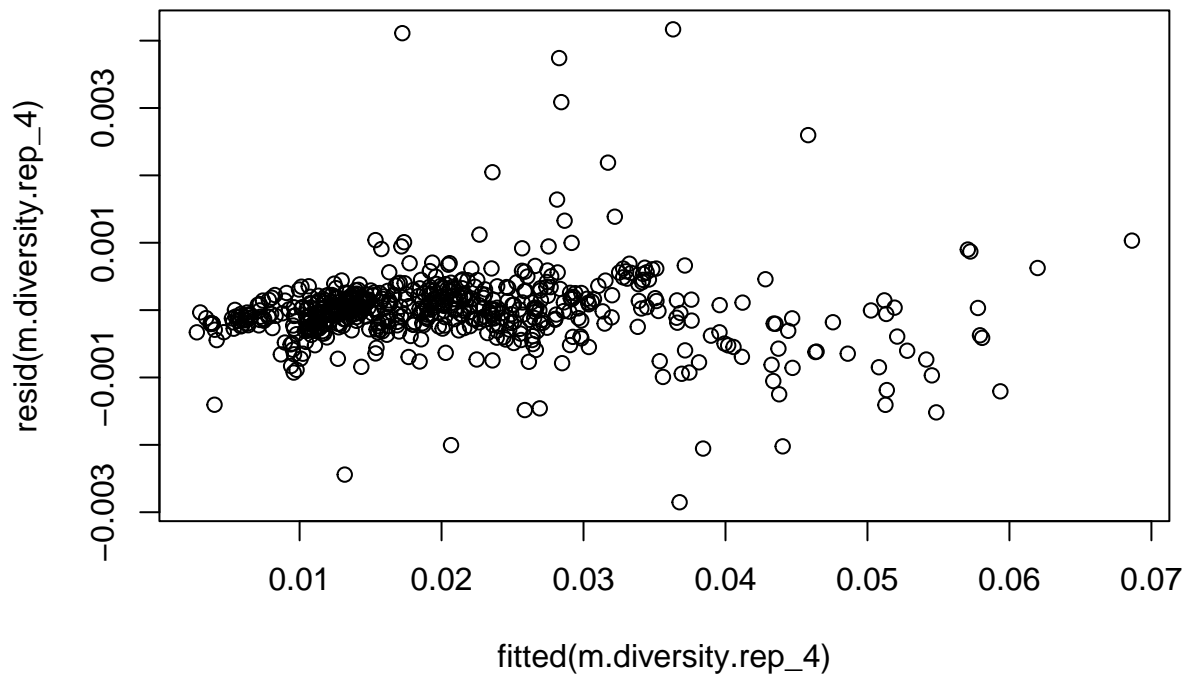
inf.lands.50kb.rep_4$bin <- 1:nrow(inf.lands.50kb.rep_4)

# for merging:
inf.lands.50kb.rep_4$Replicate <- 4

m.diversity.rep_4 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.50kb.rep_4)

plot(resid(m.diversity.rep_4)~fitted(m.diversity.rep_4))

```



```
dwtest(m.diversity.rep_4)
```

```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_4  
## DW = 0.96213, p-value < 2.2e-16  
## alternative hypothesis: true autocorrelation is greater than 0
```

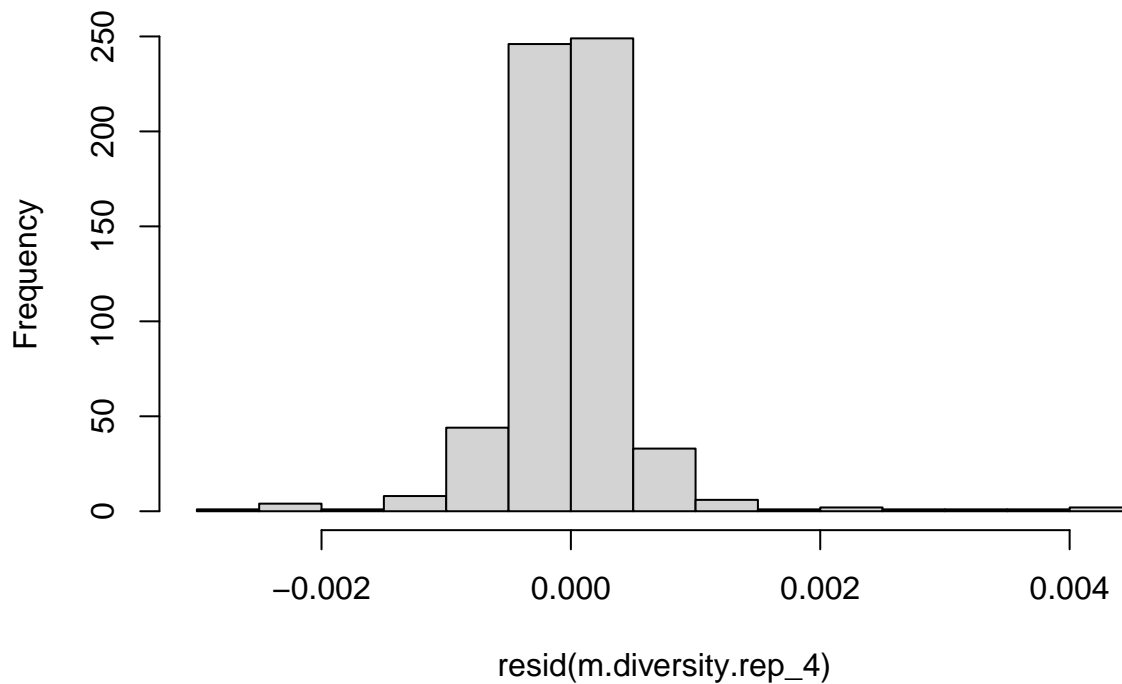
```
hmcetest(m.diversity.rep_4)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_4  
## HMC = 0.70321, p-value = 1
```

```
hist(resid(m.diversity.rep_4))
```



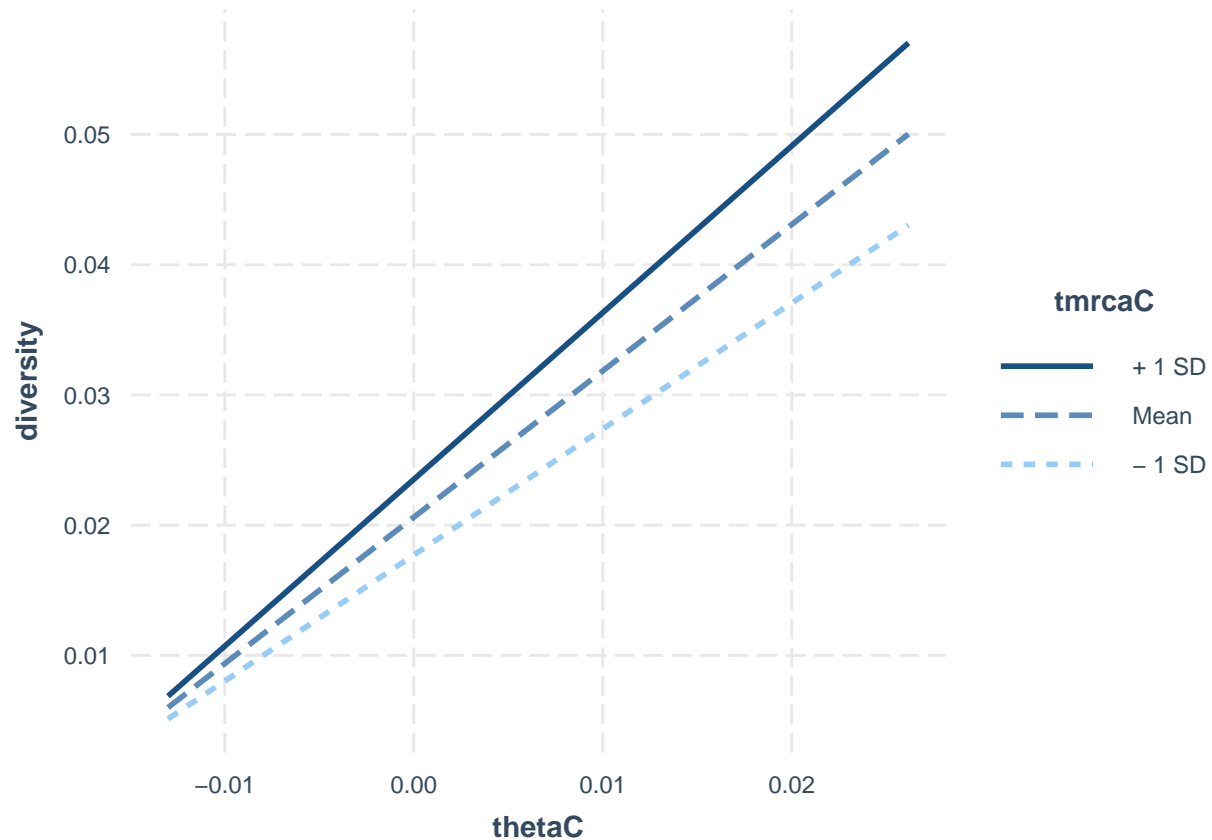
## Histogram of resid(m.diversity.rep\_4)



```
summary(m.diversity.rep_4)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.50kb.rep_4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0028517 -0.0002016 -0.0000049  0.0001933  0.0041675
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.061e-02  2.304e-05  894.744  <2e-16 ***
## thetaC       1.124e+00  2.418e-03  464.707  <2e-16 ***
## rhoC        -1.112e-02  1.845e-02  -0.603    0.547
## tmrcaC       1.960e-02  1.576e-04  124.378  <2e-16 ***
## thetaC:tmrcaC 1.055e+00  1.499e-02   70.356  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005637 on 595 degrees of freedom
## Multiple R-squared:  0.9975, Adjusted R-squared:  0.9975
## F-statistic: 5.993e+04 on 4 and 595 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_4, pred = thetaC, modx= tmrcaC)
```



```
g.rep_4 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.50kb.rep_4, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_4)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.50kb.rep_4
##      AIC      BIC   logLik
## -7464.493 -7433.715 3739.247
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.5423451
##
## Coefficients:
##              Value      Std.Error   t-value p-value
## (Intercept)  0.0206153 0.000042516 484.8841  0.0000
## thetaC       1.1223093 0.003267770 343.4480  0.0000
## tmrcaC       0.0195383 0.000157223 124.2717  0.0000
## rhoC        -0.0073321 0.015064681  -0.4867  0.6266
## thetaC:tmrcaC 1.0070650 0.014339665  70.2293  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      0.001
## tmrcaC     -0.001 -0.069
## rhoC        0.000 -0.009  0.081
## thetaC:tmrcaC -0.024 -0.065 -0.008  0.006
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -5.20520625 -0.37375535 -0.02131367  0.35668752  7.22457562
##
## Residual standard error: 0.0005658744
## Degrees of freedom: 600 total; 595 residual

vif(g.rep_4)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.009184      1.011566      1.006630      1.004416

g.rep_4.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.50kb.rep_4, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_4.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.50kb.rep_4
##      AIC      BIC    logLik
## -5330.615 -5308.631 2670.308
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.4710929
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept) 0.0206892 0.00021819 94.82308 0.0000
## thetaC      1.1585830 0.01825400 63.47010 0.0000
## rhoC       -0.1715986 0.09086145 -1.88857 0.0594
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.000  0.003
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -5.45522216 -0.42729625  0.05682276  0.47857205  5.82296892
##
## Residual standard error: 0.00320132
## Degrees of freedom: 600 total; 597 residual
```

```

anova.diversity <- Anova(m.diversity.rep_4)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.50kb[1, 4] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.inf.50kb[2, 4] <- anova.diversity$VarExp[1] * 100
r2.inf.50kb[3, 4] <- anova.diversity$VarExp[2] * 100
r2.inf.50kb[4, 4] <- anova.diversity$VarExp[3] * 100
r2.inf.50kb[5, 4] <- anova.diversity$VarExp[4] * 100

```

### 2.1.5 Replicate 5

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_5/rs.pair_5.", sep = "")
pi.50kb <- read.table(paste(p, "diversity.50kb.bedgraph", sep = ""), header = T)
pi.50kb$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCA.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$avg <- apply(tmrca.50kb[4:ncol(tmrca.50kb)], 1, mean)
rho.50kb <- read.table(paste(p, "rho.50kb.bedgraph", sep = ""), header = T)
theta.50kb <- read.table(paste(p, "theta.50kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.50kb$sim, theta.50kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.50kb$sim and theta.50kb$sample_mean
## S = 655949, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9817791

inf.lands.50kb.rep_5 <- as.data.frame(cbind(pi.50kb$avg, theta.50kb$sample_mean, rho.50kb$sample_mean,
names(inf.lands.50kb.rep_5) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.50kb.rep_5$thetaC <- inf.lands.50kb.rep_5$theta - mean(inf.lands.50kb.rep_5$theta)
inf.lands.50kb.rep_5$tmrcaC <- inf.lands.50kb.rep_5$tmrca - mean(inf.lands.50kb.rep_5$tmrca)
inf.lands.50kb.rep_5$rhoC <- inf.lands.50kb.rep_5$rho - mean(inf.lands.50kb.rep_5$rho)

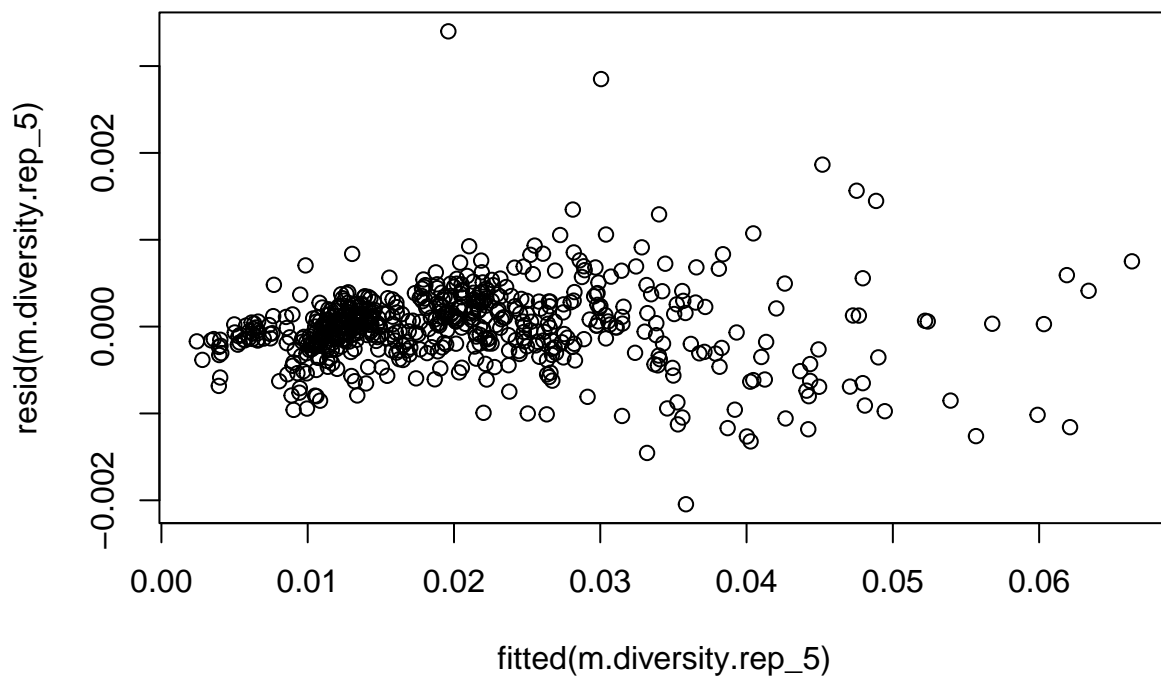
inf.lands.50kb.rep_5$bin <- 1:nrow(inf.lands.50kb.rep_5)

# for merging:
inf.lands.50kb.rep_5$Replicate <- 5

m.diversity.rep_5 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.50kb.rep_5)

plot(resid(m.diversity.rep_5)~fitted(m.diversity.rep_5))

```



```
dwtest(m.diversity.rep_5)
```

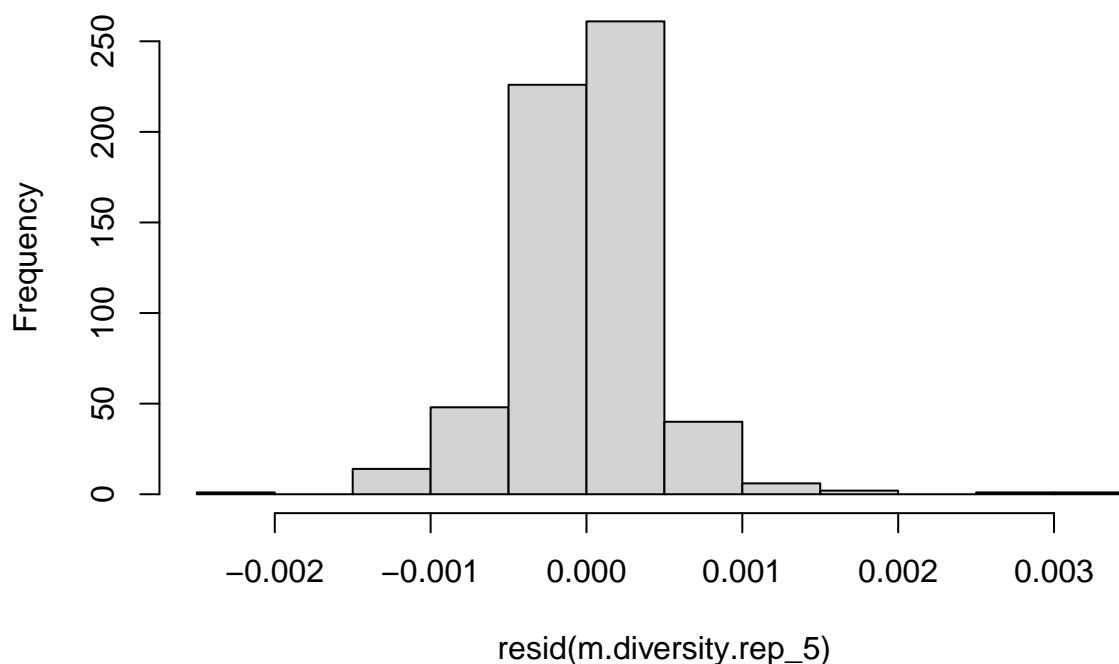
```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_5  
## DW = 1.4275, p-value = 5.886e-13  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_5)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_5  
## HMC = 0.55418, p-value = 0.966
```

```
hist(resid(m.diversity.rep_5))
```

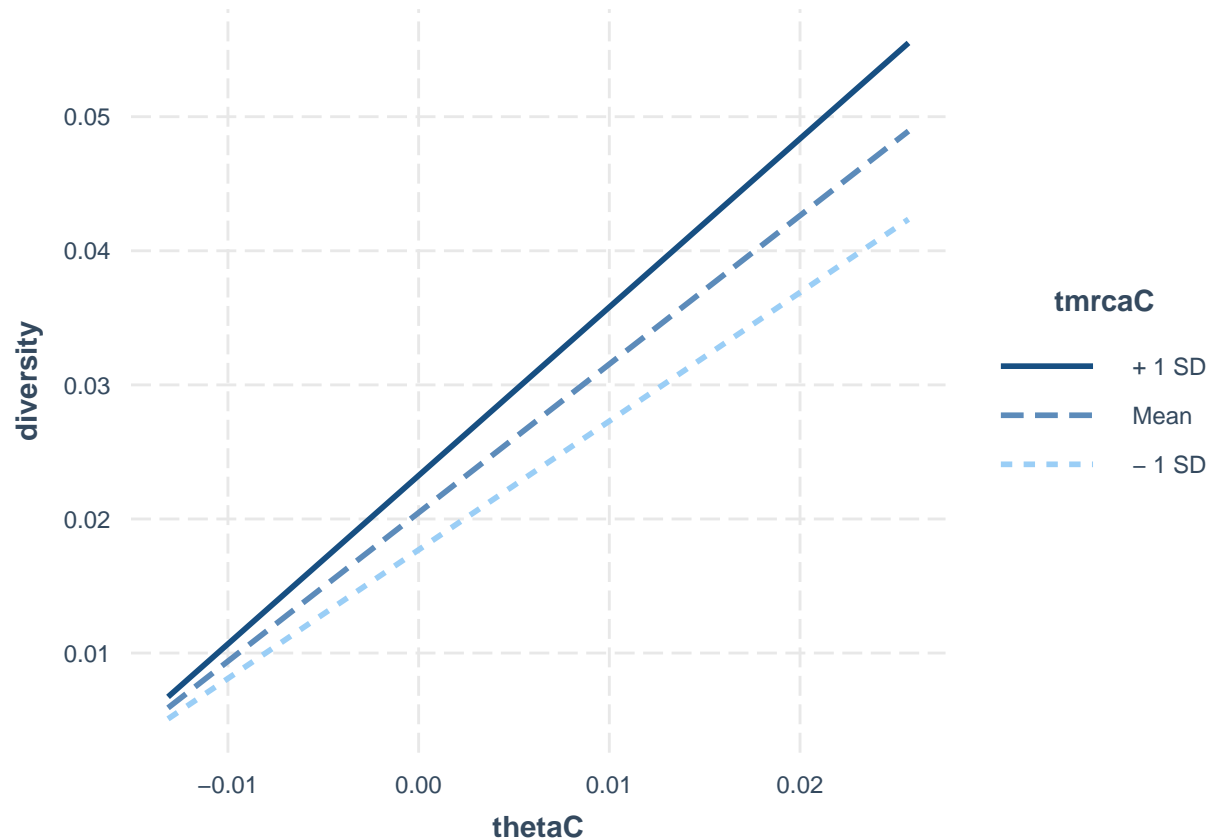
## Histogram of resid(m.diversity.rep\_5)



```
summary(m.diversity.rep_5)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.50kb.rep_5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0020455 -0.0002004  0.0000140  0.0002247  0.0033998
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.0204679   0.0000189 1082.787 <2e-16 ***
## thetaC        1.1078321   0.0019892  556.934 <2e-16 ***
## rhoC         -0.0069343   0.0156805   -0.442   0.658
## tmrcaC        0.0198380   0.0001375  144.318 <2e-16 ***
## thetaC:tmrcaC  1.0621447   0.0118730   89.459 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0004621 on 595 degrees of freedom
## Multiple R-squared:  0.9983, Adjusted R-squared:  0.9983
## F-statistic: 8.634e+04 on 4 and 595 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_5, pred = thetaC, modx= tmrcaC)
```



```
g.rep_5 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.50kb.rep_5, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_5)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.50kb.rep_5
##      AIC      BIC   logLik
## -7570.892 -7540.114 3792.446
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.3687999
##
## Coefficients:
##              Value      Std.Error   t-value p-value
## (Intercept)  0.0204733 0.000028264  724.3584  0.0000
## thetaC       1.1067703 0.002603042  425.1835  0.0000
## tmrcaC       0.0195333 0.000148186  131.8159  0.0000
## rhoC        -0.0117411 0.014941182   -0.7858  0.4323
## thetaC:tmrcaC 1.0119684 0.012760307   79.3060  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.002
## tmrcaC       0.002 -0.094
## rhoC        -0.001  0.002  0.058
## thetaC:tmrcaC -0.045  0.037 -0.040  0.028
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.3488500193 -0.4439991672  0.0002670608  0.4775851240  7.1344633176
##
## Residual standard error: 0.0004681579
## Degrees of freedom: 600 total; 595 residual

vif(g.rep_5)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.010139      1.013921      1.004356      1.003709

g.rep_5.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.50kb.rep_5, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_5.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.50kb.rep_5
##      AIC      BIC logLik
## -5329.279 -5307.294 2669.64
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.4006009
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept)  0.0205709 0.00019283 106.68153 0.0000
## thetaC      1.1335447 0.01727411  65.62102 0.0000
## rhoC      -0.1554325 0.09617437  -1.61615 0.1066
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.001  0.007
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.31295794 -0.44285975  0.05270967  0.47530087  5.66926609
##
## Residual standard error: 0.003085565
## Degrees of freedom: 600 total; 597 residual
```



```

anova.diversity <- Anova(m.diversity.rep_5)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.50kb[1, 5] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.inf.50kb[2, 5] <- anova.diversity$VarExp[1] * 100
r2.inf.50kb[3, 5] <- anova.diversity$VarExp[2] * 100
r2.inf.50kb[4, 5] <- anova.diversity$VarExp[3] * 100
r2.inf.50kb[5, 5] <- anova.diversity$VarExp[4] * 100

```

### 2.1.6 Replicate 6

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_6/rs.pair_6.", sep = "")
pi.50kb <- read.table(paste(p, "diversity.50kb.bedgraph", sep = ""), header = T)
pi.50kb$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCA.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$avg <- apply(tmrca.50kb[4:ncol(tmrca.50kb)], 1, mean)
rho.50kb <- read.table(paste(p, "rho.50kb.bedgraph", sep = ""), header = T)
theta.50kb <- read.table(paste(p, "theta.50kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.50kb$sim, theta.50kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.50kb$sim and theta.50kb$sample_mean
## S = 576502, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.983986

inf.lands.50kb.rep_6 <- as.data.frame(cbind(pi.50kb$avg, theta.50kb$sample_mean, rho.50kb$sample_mean,
names(inf.lands.50kb.rep_6) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.50kb.rep_6$thetaC <- inf.lands.50kb.rep_6$theta - mean(inf.lands.50kb.rep_6$theta)
inf.lands.50kb.rep_6$tmrcaC <- inf.lands.50kb.rep_6$tmrca - mean(inf.lands.50kb.rep_6$tmrca)
inf.lands.50kb.rep_6$rhoC <- inf.lands.50kb.rep_6$rho - mean(inf.lands.50kb.rep_6$rho)

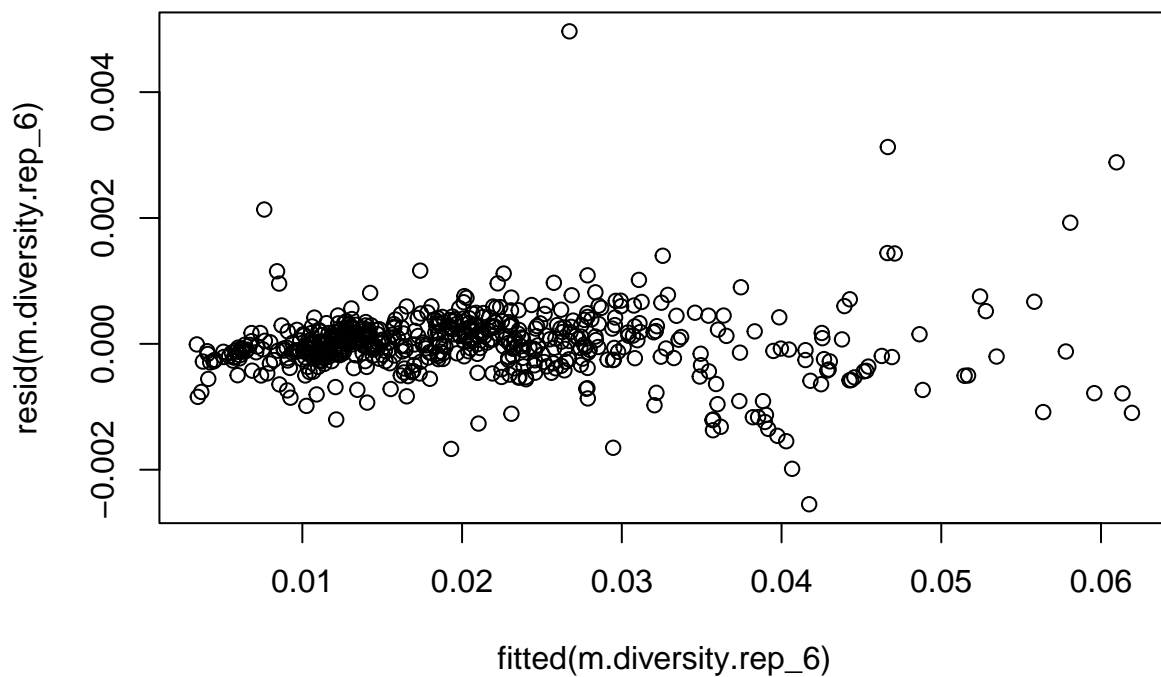
inf.lands.50kb.rep_6$bin <- 1:nrow(inf.lands.50kb.rep_6)

# for merging:
inf.lands.50kb.rep_6$Replicate <- 6

m.diversity.rep_6 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.50kb.rep_6)

plot(resid(m.diversity.rep_6)~fitted(m.diversity.rep_6))

```



```
dwtest(m.diversity.rep_6)
```

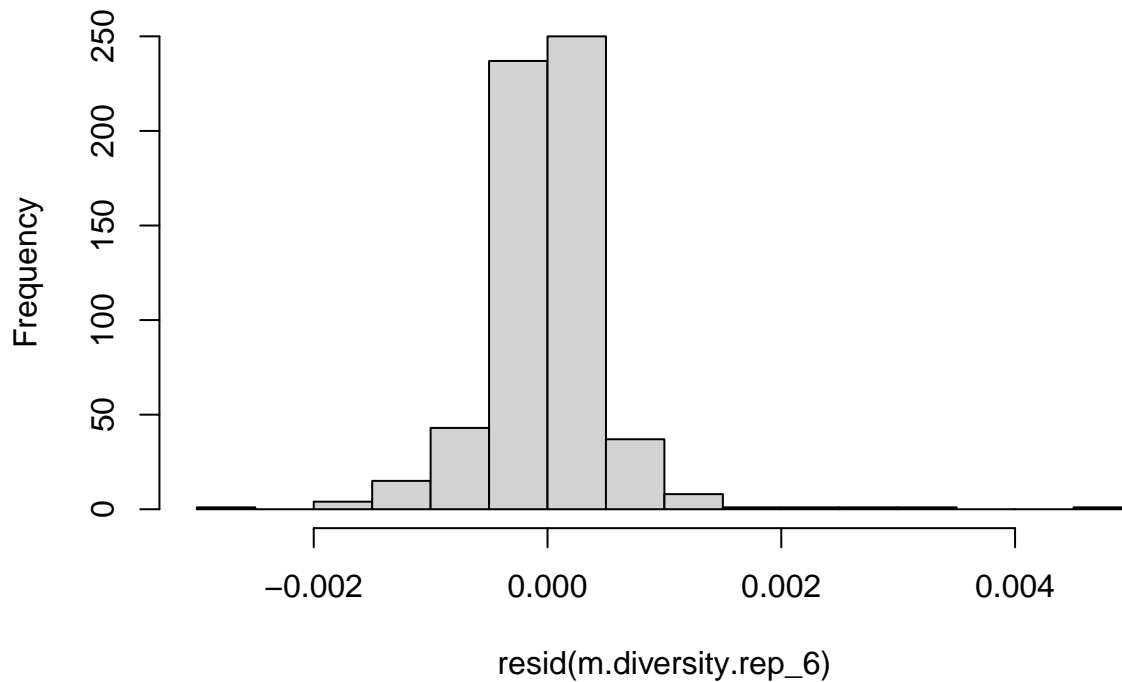
```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_6  
## DW = 1.4152, p-value = 1.956e-13  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_6)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_6  
## HMC = 0.5501, p-value = 0.952
```

```
hist(resid(m.diversity.rep_6))
```

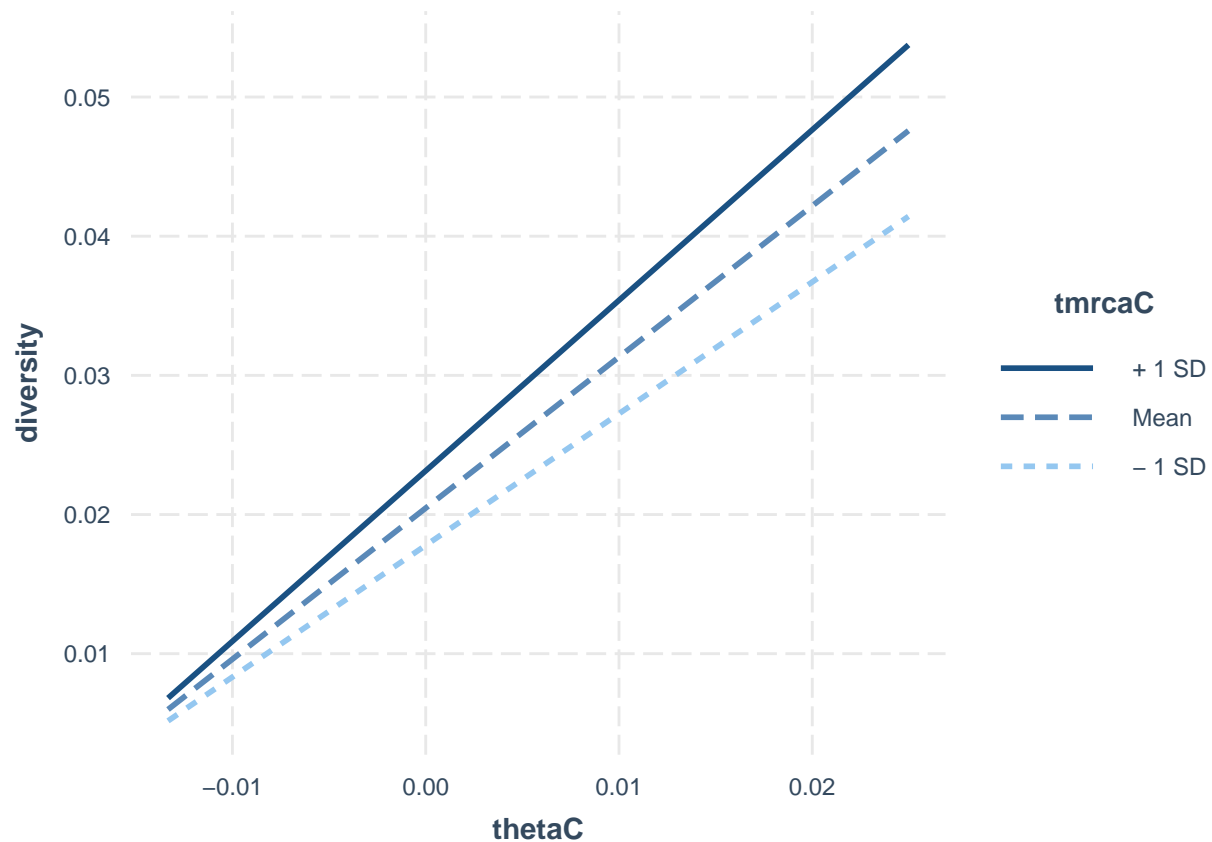
## Histogram of resid(m.diversity.rep\_6)



```
summary(m.diversity.rep_6)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.50kb.rep_6)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0025487 -0.0002073 -0.0000008  0.0002258  0.0049661
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.046e-02  2.183e-05  937.468  <2e-16 ***
## thetaC       1.086e+00  2.249e-03  482.709  <2e-16 ***
## rhoC        9.344e-03  1.682e-02   0.555    0.579
## tmrcaC       2.016e-02  1.691e-04  119.259  <2e-16 ***
## thetaC:tmrcaC 1.050e+00  1.470e-02   71.423  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005333 on 595 degrees of freedom
## Multiple R-squared:  0.9977, Adjusted R-squared:  0.9977
## F-statistic: 6.529e+04 on 4 and 595 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_6, pred = thetaC, modx= tmrcaC)
```



```
g.rep_6 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.50kb.rep_6, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_6)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.50kb.rep_6
##      AIC      BIC logLik
## -7391.16 -7360.382 3702.58
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.3219522
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0204681 0.000030586 669.1903  0.0000
## thetaC       1.0858244 0.002832026 383.4091  0.0000
## tmrcaC       0.0199376 0.000176983 112.6524  0.0000
## rhoC         0.0002293 0.016283078   0.0141  0.9888
## thetaC:tmrcaC 1.0151180 0.015909552  63.8056  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      0.000
## tmrcaC     -0.006 -0.082
## rhoC        0.000  0.014  0.075
## thetaC:tmrcaC -0.056  0.002  0.082  0.013
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.664990405 -0.377078774 -0.001461415  0.411936942  9.302435018
##
## Residual standard error: 0.0005339103
## Degrees of freedom: 600 total; 595 residual

vif(g.rep_6)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.007216      1.019473      1.006158      1.006869

g.rep_6.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.50kb.rep_6, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_6.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.50kb.rep_6
##      AIC      BIC    logLik
## -5414.912 -5392.927 2712.456
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.4107098
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept) 0.0205846 0.00018261 112.72154 0.0000
## thetaC      1.1129166 0.01585204  70.20654 0.0000
## rhoC       -0.1062152 0.08342243  -1.27322 0.2034
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.001  0.014
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.246615322 -0.484784795 -0.004012755  0.502253353  5.520109925
##
## Residual standard error: 0.002887164
## Degrees of freedom: 600 total; 597 residual
```

```

anova.diversity <- Anova(m.diversity.rep_6)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.50kb[1, 6] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.inf.50kb[2, 6] <- anova.diversity$VarExp[1] * 100
r2.inf.50kb[3, 6] <- anova.diversity$VarExp[2] * 100
r2.inf.50kb[4, 6] <- anova.diversity$VarExp[3] * 100
r2.inf.50kb[5, 6] <- anova.diversity$VarExp[4] * 100

```

### 2.1.7 Replicate 7

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_7/rs.pair_7.", sep = "")
pi.50kb <- read.table(paste(p, "diversity.50kb.bedgraph", sep = ""), header = T)
pi.50kb$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCA.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$avg <- apply(tmrca.50kb[4:ncol(tmrca.50kb)], 1, mean)
rho.50kb <- read.table(paste(p, "rho.50kb.bedgraph", sep = ""), header = T)
theta.50kb <- read.table(paste(p, "theta.50kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.50kb$sim, theta.50kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.50kb$sim and theta.50kb$sample_mean
## S = 715696, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9801195

inf.lands.50kb.rep_7 <- as.data.frame(cbind(pi.50kb$avg, theta.50kb$sample_mean, rho.50kb$sample_mean,
names(inf.lands.50kb.rep_7) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.50kb.rep_7$thetaC <- inf.lands.50kb.rep_7$theta - mean(inf.lands.50kb.rep_7$theta)
inf.lands.50kb.rep_7$tmrcaC <- inf.lands.50kb.rep_7$tmrca - mean(inf.lands.50kb.rep_7$tmrca)
inf.lands.50kb.rep_7$rhoC <- inf.lands.50kb.rep_7$rho - mean(inf.lands.50kb.rep_7$rho)

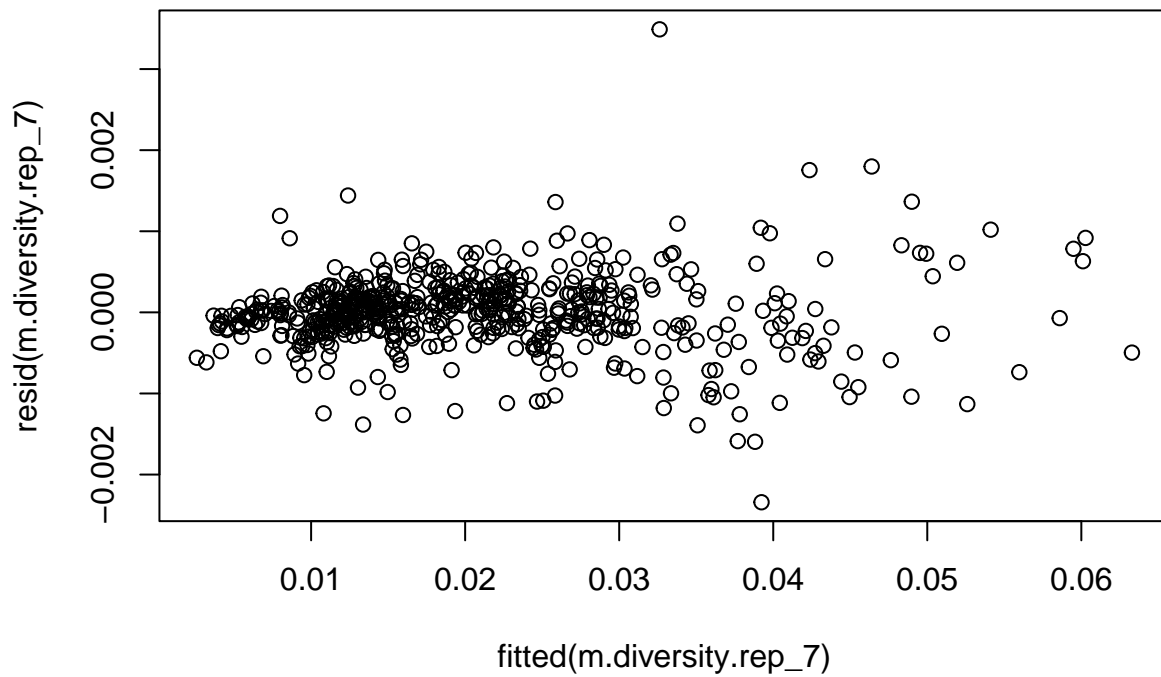
inf.lands.50kb.rep_7$bin <- 1:nrow(inf.lands.50kb.rep_7)

# for merging:
inf.lands.50kb.rep_7$Replicate <- 7

m.diversity.rep_7 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.50kb.rep_7)

plot(resid(m.diversity.rep_7)~fitted(m.diversity.rep_7))

```



```
dwtest(m.diversity.rep_7)
```

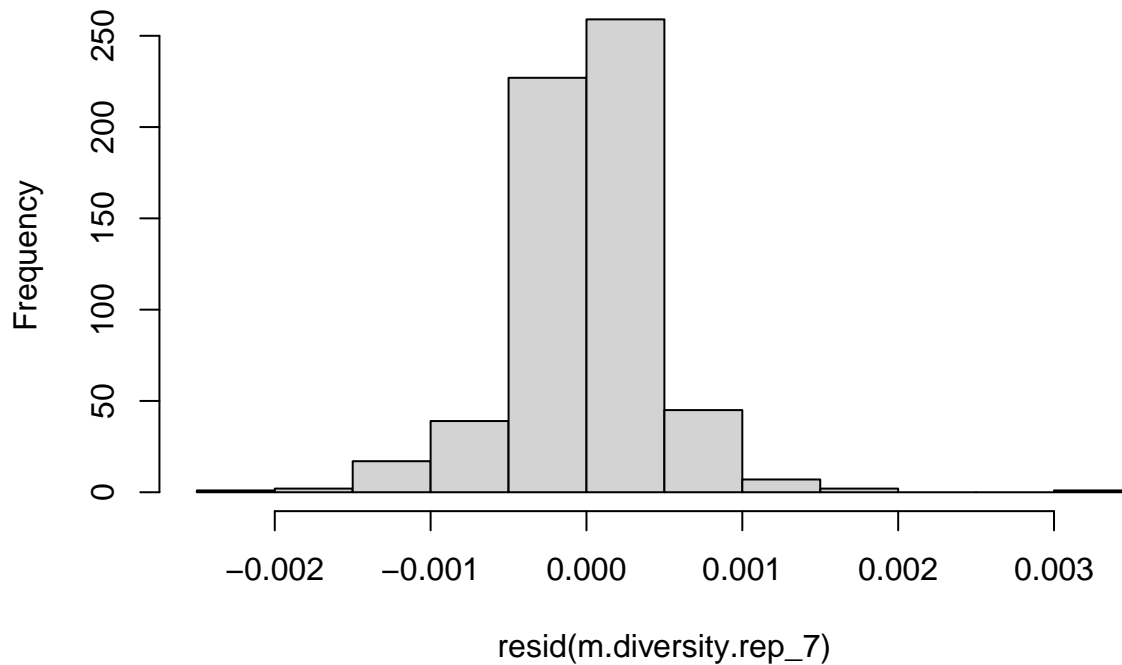
```
##
## Durbin-Watson test
##
## data: m.diversity.rep_7
## DW = 1.2084, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcctest(m.diversity.rep_7)
```

```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_7
## HMC = 0.58586, p-value = 0.999
```

```
hist(resid(m.diversity.rep_7))
```

## Histogram of resid(m.diversity.rep\_7)

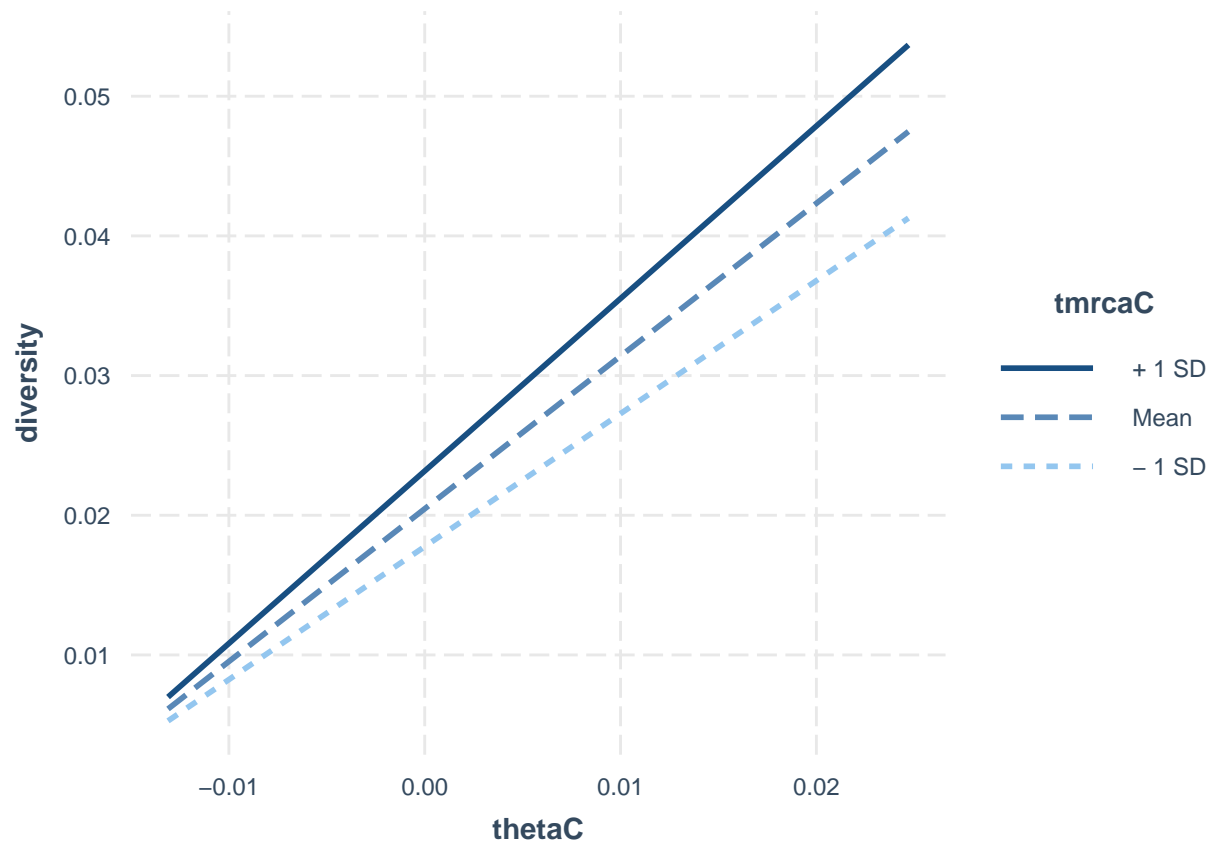


```
summary(m.diversity.rep_7)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.50kb.rep_7)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0023413 -0.0001931  0.0000183  0.0002113  0.0034898
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.046e-02  1.906e-05 1073.378 < 2e-16 ***
## thetaC       1.093e+00  1.994e-03  548.323 < 2e-16 ***
## rhoC        -3.984e-02  1.517e-02   -2.627  0.00885 **
## tmrcaC       2.021e-02  1.462e-04  138.265 < 2e-16 ***
## thetaC:tmrcaC 1.056e+00  1.344e-02   78.551 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0004657 on 595 degrees of freedom
## Multiple R-squared:  0.9982, Adjusted R-squared:  0.9982
## F-statistic: 8.313e+04 on 4 and 595 DF, p-value: < 2.2e-16
```



```
interact_plot(m.diversity.rep_7, pred = thetaC, modx= tmrcaC)
```



```
g.rep_7 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.50kb.rep_7, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_7)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.50kb.rep_7
##      AIC      BIC    logLik
## -7617.29 -7586.512 3815.645
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.4690905
##
## Coefficients:
##              Value      Std.Error   t-value p-value
## (Intercept)  0.0204701 0.000032313  633.4967  0.0000
## thetaC       1.0964625 0.002720215  403.0793  0.0000
## tmrcaC       0.0197764 0.000149360  132.4076  0.0000
## rhoC        -0.0310546 0.013470566   -2.3054  0.0215
## thetaC:tmrcaC 0.9959236 0.014228586   69.9946  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.001
## tmrcaC      -0.006 -0.099
## rhoC         0.002  0.005 -0.022
## thetaC:tmrcaC -0.047  0.014  0.100 -0.014
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.79375373 -0.41801231  0.05547631  0.45600293  7.65648237
##
## Residual standard error: 0.0004740043
## Degrees of freedom: 600 total; 595 residual

vif(g.rep_7)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.010511      1.020984      1.000614      1.010927

g.rep_7.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.50kb.rep_7, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_7.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.50kb.rep_7
##      AIC      BIC    logLik
## -5472.449 -5450.464 2741.224
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.469828
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept) 0.0205801 0.00019340 106.41030 0.0000
## thetaC      1.1283864 0.01620181  69.64571 0.0000
## rhoC        0.0166423 0.08056264   0.20658 0.8364
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.001  0.003
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.47464746 -0.42695326  0.04565358  0.50179301  5.14163869
##
## Residual standard error: 0.002842278
## Degrees of freedom: 600 total; 597 residual
```

```

anova.diversity <- Anova(m.diversity.rep_7)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.50kb[1, 7] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.inf.50kb[2, 7] <- anova.diversity$VarExp[1] * 100
r2.inf.50kb[3, 7] <- anova.diversity$VarExp[2] * 100
r2.inf.50kb[4, 7] <- anova.diversity$VarExp[3] * 100
r2.inf.50kb[5, 7] <- anova.diversity$VarExp[4] * 100

```

### 2.1.8 Replicate 8

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_8/rs.pair_8.", sep = "")
pi.50kb <- read.table(paste(p, "diversity.50kb.bedgraph", sep = ""), header = T)
pi.50kb$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCA.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$avg <- apply(tmrca.50kb[4:ncol(tmrca.50kb)], 1, mean)
rho.50kb <- read.table(paste(p, "rho.50kb.bedgraph", sep = ""), header = T)
theta.50kb <- read.table(paste(p, "theta.50kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.50kb$sim, theta.50kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.50kb$sim and theta.50kb$sample_mean
## S = 682855, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9810318

inf.lands.50kb.rep_8 <- as.data.frame(cbind(pi.50kb$avg, theta.50kb$sample_mean, rho.50kb$sample_mean,
names(inf.lands.50kb.rep_8) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.50kb.rep_8$thetaC <- inf.lands.50kb.rep_8$theta - mean(inf.lands.50kb.rep_8$theta)
inf.lands.50kb.rep_8$tmrcaC <- inf.lands.50kb.rep_8$tmrca - mean(inf.lands.50kb.rep_8$tmrca)
inf.lands.50kb.rep_8$rhoC <- inf.lands.50kb.rep_8$rho - mean(inf.lands.50kb.rep_8$rho)

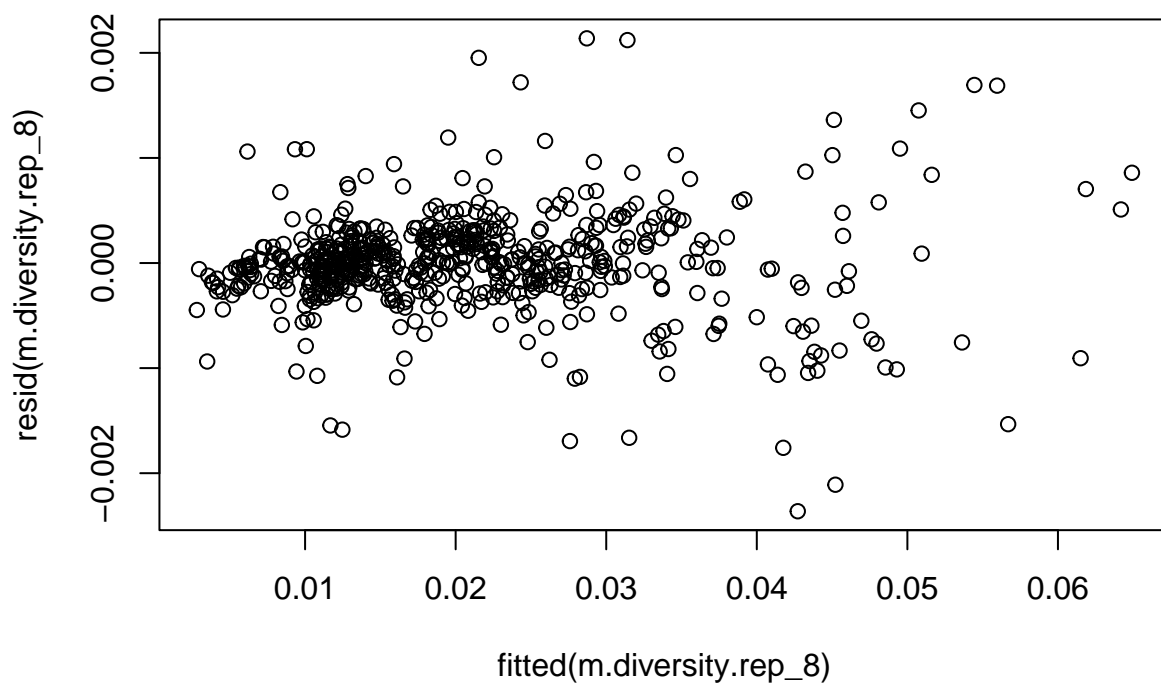
inf.lands.50kb.rep_8$bin <- 1:nrow(inf.lands.50kb.rep_8)

# for merging:
inf.lands.50kb.rep_8$Replicate <- 8

m.diversity.rep_8 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.50kb.rep_8)

plot(resid(m.diversity.rep_8)~fitted(m.diversity.rep_8))

```



```
dwtest(m.diversity.rep_8)
```

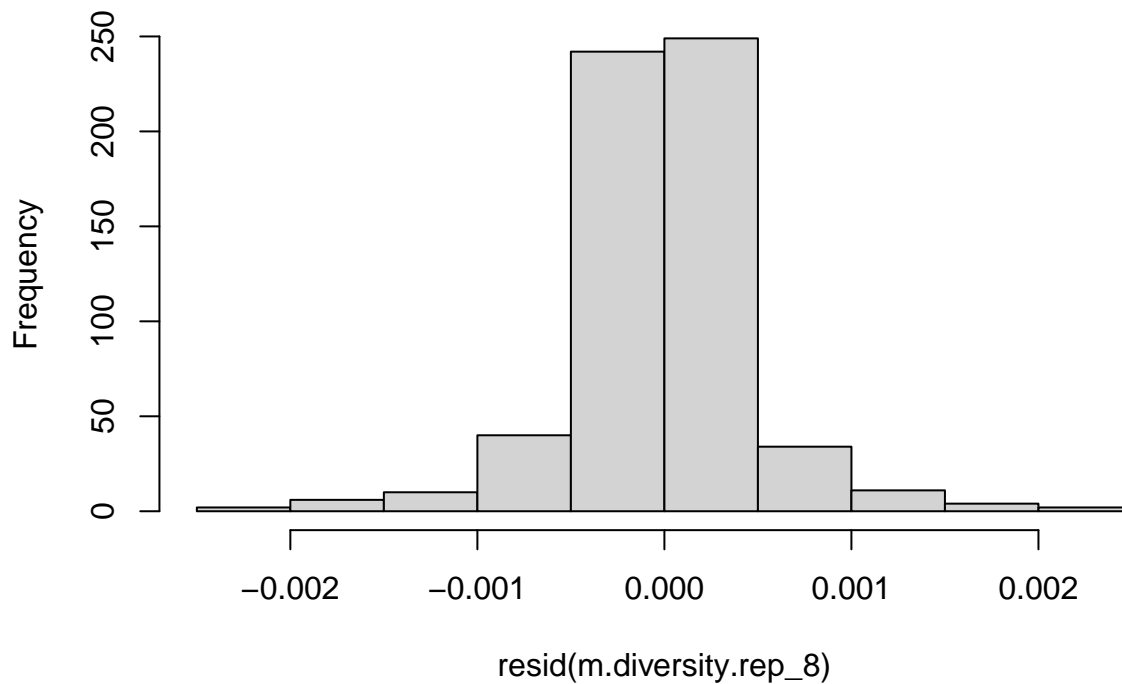
```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_8  
## DW = 1.3796, p-value = 6.877e-15  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.diversity.rep_8)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_8  
## HMC = 0.62278, p-value = 1
```

```
hist(resid(m.diversity.rep_8))
```

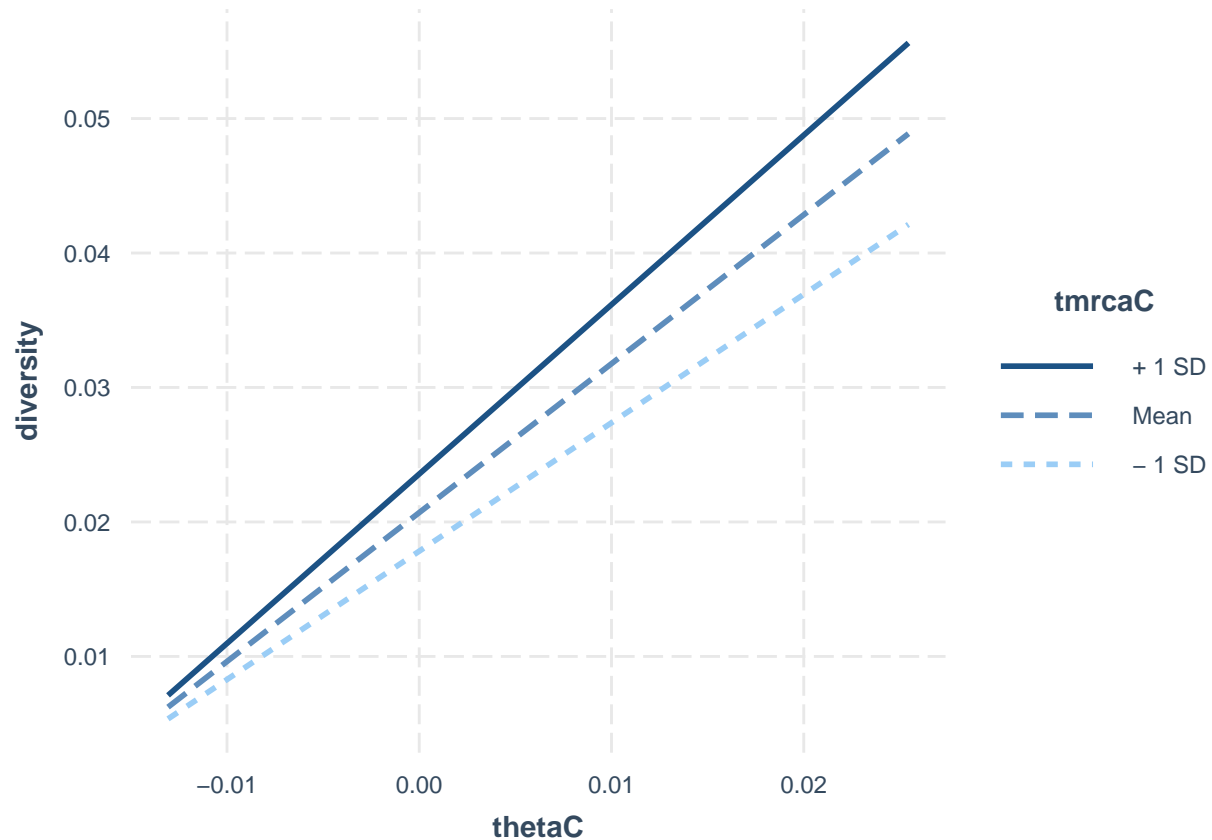
## Histogram of resid(m.diversity.rep\_8)



```
summary(m.diversity.rep_8)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.50kb.rep_8)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.362e-03 -1.893e-04  3.500e-07  2.086e-04  2.138e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.070e-02  1.938e-05 1067.98  <2e-16 ***
## thetaC       1.107e+00  2.023e-03  547.27  <2e-16 ***
## rhoC        -6.776e-03  1.539e-02   -0.44    0.66
## tmrcaC       2.013e-02  1.393e-04  144.50  <2e-16 ***
## thetaC:tmrcaC 1.071e+00  1.222e-02   87.61  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0004737 on 595 degrees of freedom
## Multiple R-squared:  0.9982, Adjusted R-squared:  0.9982
## F-statistic: 8.292e+04 on 4 and 595 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_8, pred = thetaC, modx= tmrcaC)
```



```
g.rep_8 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.50kb.rep_8, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_8)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.50kb.rep_8
##      AIC      BIC   logLik
## -7548.618 -7517.839 3781.309
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.3796214
##
## Coefficients:
##              Value   Std.Error   t-value p-value
## (Intercept)  0.0207031 0.000029299  706.6176  0.0000
## thetaC       1.1082548 0.002641688  419.5253  0.0000
## tmrcaC       0.0198182 0.000148295  133.6401  0.0000
## rhoC        0.0047060 0.014548476   0.3235  0.7465
## thetaC:tmrcaC 1.0223058 0.013460362  75.9494  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.001
## tmrcaC      -0.004 -0.069
## rhoC         0.001  0.012  0.067
## thetaC:tmrcaC -0.048  0.026  0.075 -0.015
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.79143530 -0.40162824  0.00873902  0.44671770  4.52028563
##
## Residual standard error: 0.0004791766
## Degrees of freedom: 600 total; 595 residual

vif(g.rep_8)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.006009      1.015515      1.005156      1.006994

g.rep_8.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.50kb.rep_8, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_8.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.50kb.rep_8
##      AIC      BIC logLik
## -5369.959 -5347.975 2689.98
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.4676478
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept)  0.0208143 0.00020979 99.21612  0.0000
## thetaC      1.1333322 0.01746884 64.87738  0.0000
## rhoC      -0.0900602 0.08770082 -1.02690  0.3049
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.000  0.012
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -7.41084899 -0.48902470 -0.01613762  0.49491223  5.48392430
##
## Residual standard error: 0.003091654
## Degrees of freedom: 600 total; 597 residual
```

```

anova.diversity <- Anova(m.diversity.rep_8)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.50kb[1, 8] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.inf.50kb[2, 8] <- anova.diversity$VarExp[1] * 100
r2.inf.50kb[3, 8] <- anova.diversity$VarExp[2] * 100
r2.inf.50kb[4, 8] <- anova.diversity$VarExp[3] * 100
r2.inf.50kb[5, 8] <- anova.diversity$VarExp[4] * 100

```

### 2.1.9 Replicate 9

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_9/rs.pair_9.", sep = "")
pi.50kb <- read.table(paste(p, "diversity.50kb.bedgraph", sep = ""), header = T)
pi.50kb$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCA.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$avg <- apply(tmrca.50kb[4:ncol(tmrca.50kb)], 1, mean)
rho.50kb <- read.table(paste(p, "rho.50kb.bedgraph", sep = ""), header = T)
theta.50kb <- read.table(paste(p, "theta.50kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.50kb$sim, theta.50kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.50kb$sim and theta.50kb$sample_mean
## S = 609380, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9830727

inf.lands.50kb.rep_9 <- as.data.frame(cbind(pi.50kb$avg, theta.50kb$sample_mean, rho.50kb$sample_mean,
names(inf.lands.50kb.rep_9) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.50kb.rep_9$thetaC <- inf.lands.50kb.rep_9$theta - mean(inf.lands.50kb.rep_9$theta)
inf.lands.50kb.rep_9$tmrcaC <- inf.lands.50kb.rep_9$tmrca - mean(inf.lands.50kb.rep_9$tmrca)
inf.lands.50kb.rep_9$rhoC <- inf.lands.50kb.rep_9$rho - mean(inf.lands.50kb.rep_9$rho)

inf.lands.50kb.rep_9$bin <- 1:nrow(inf.lands.50kb.rep_9)

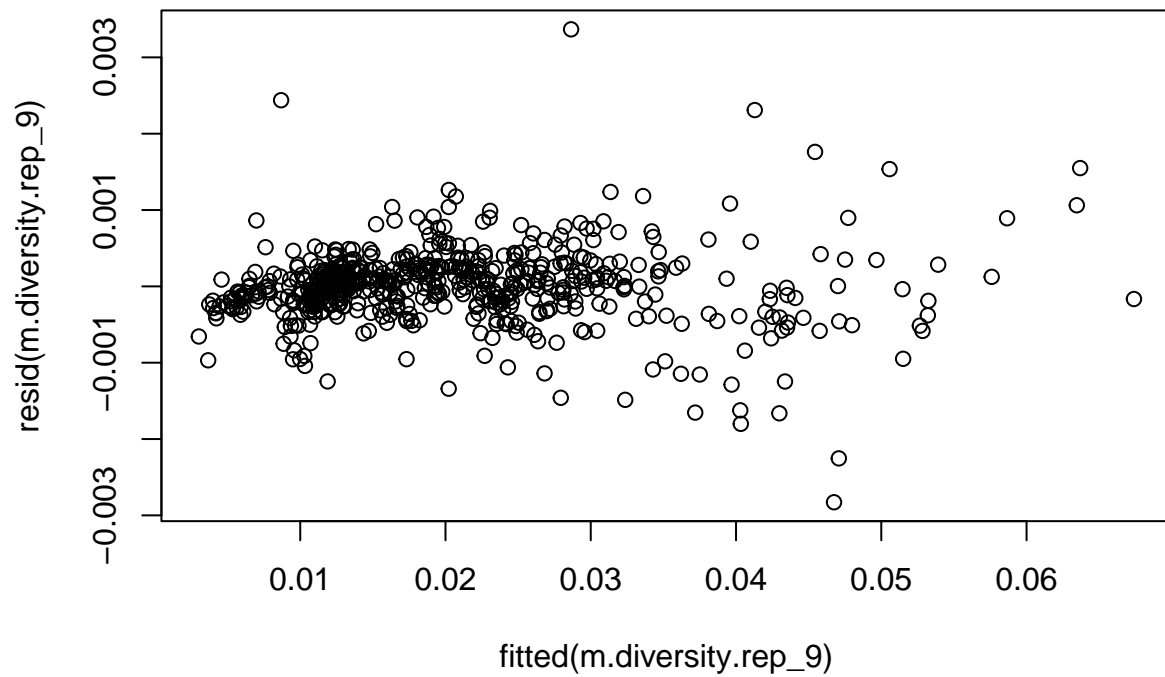
# for merging:
inf.lands.50kb.rep_9$Replicate <- 9

m.diversity.rep_9 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.50kb.rep_9)

plot(resid(m.diversity.rep_9)~fitted(m.diversity.rep_9))

```





```
dwtest(m.diversity.rep_9)
```

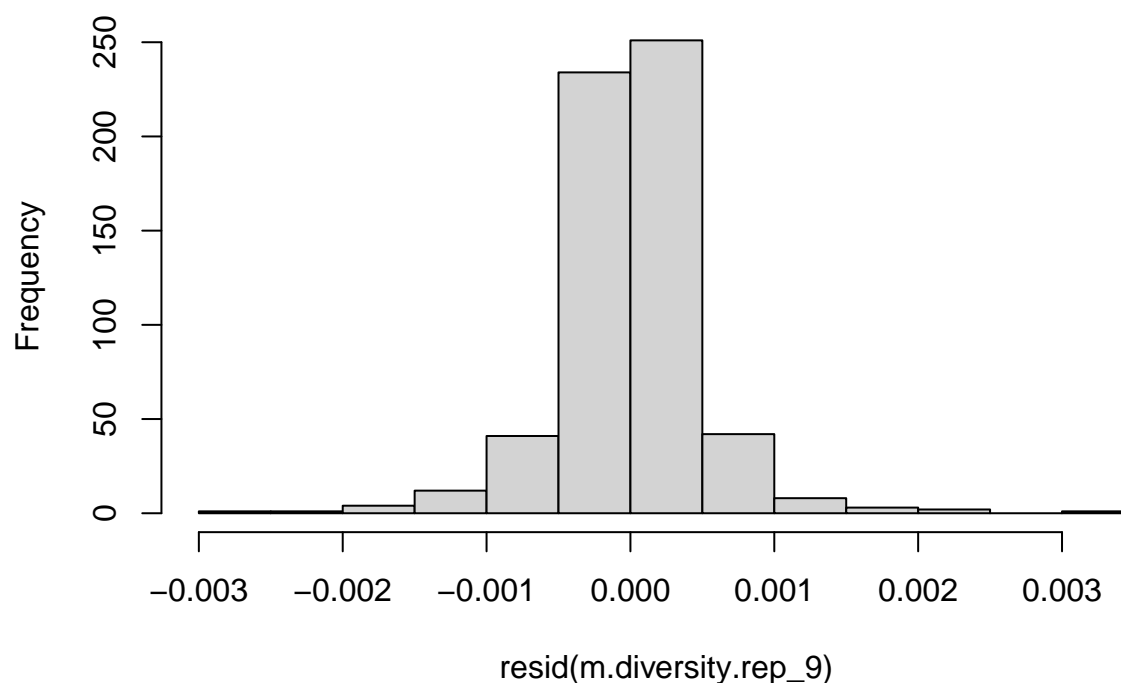
```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_9  
## DW = 1.3455, p-value = 2.541e-16  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_9)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_9  
## HMC = 0.52947, p-value = 0.83
```

```
hist(resid(m.diversity.rep_9))
```

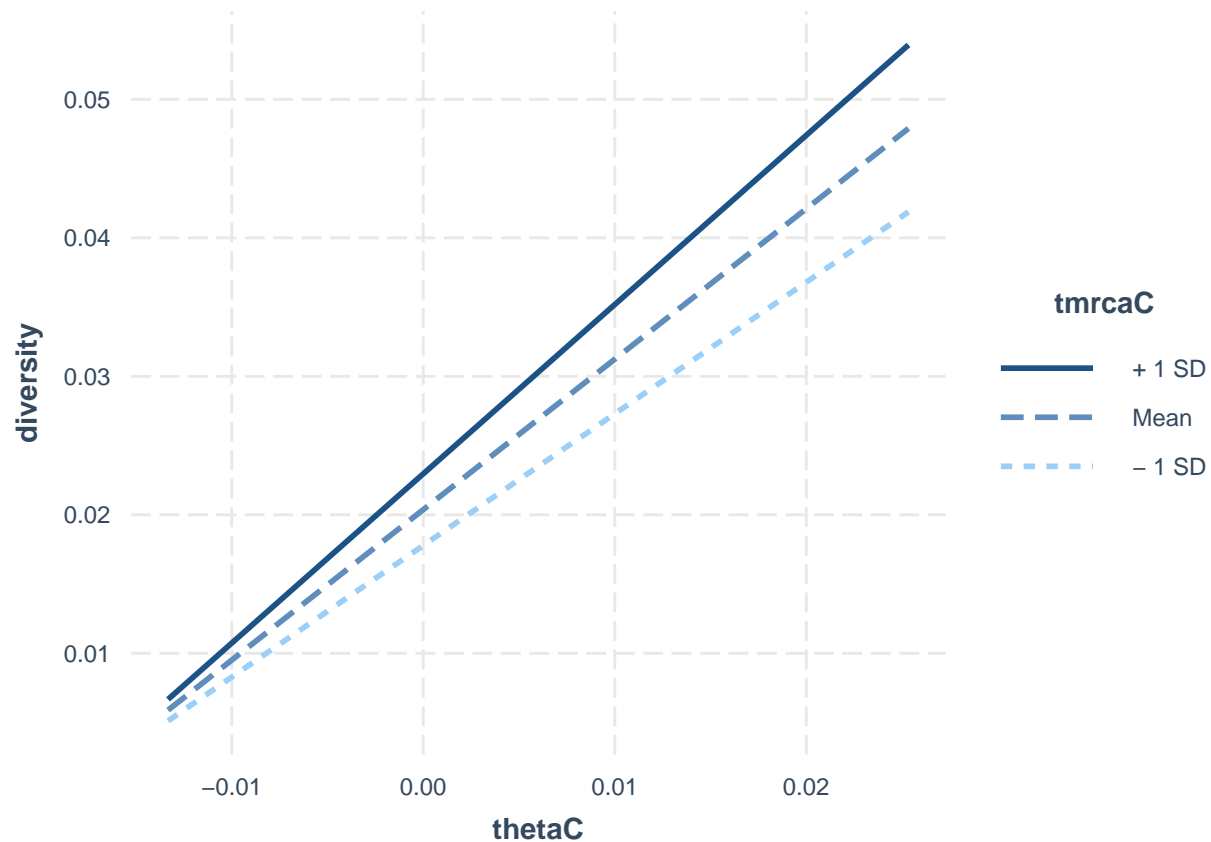
## Histogram of resid(m.diversity.rep\_9)



```
summary(m.diversity.rep_9)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.50kb.rep_9)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0028280 -0.0002274  0.0000078  0.0002235  0.0033670
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.037e-02  2.069e-05  984.748  <2e-16 ***
## thetaC       1.086e+00  2.147e-03  505.918  <2e-16 ***
## rhoC         1.171e-02  1.718e-02   0.682    0.496
## tmrcaC        1.990e-02  1.617e-04  123.090  <2e-16 ***
## thetaC:tmrcaC 1.045e+00  1.408e-02   74.239  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005043 on 595 degrees of freedom
## Multiple R-squared:  0.998, Adjusted R-squared:  0.998
## F-statistic: 7.306e+04 on 4 and 595 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_9, pred = thetaC, modx= tmrcaC)
```



```
g.rep_9 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.50kb.rep_9, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_9)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.50kb.rep_9
##      AIC      BIC   logLik
## -7487.186 -7456.408 3750.593
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.412105
##
## Coefficients:
##              Value   Std.Error   t-value p-value
## (Intercept)  0.0203835 0.000032571 625.8123  0.0000
## thetaC       1.0864837 0.002863160 379.4701  0.0000
## tmrcaC       0.0194098 0.000164202 118.2072  0.0000
## rhoC        0.0014542 0.015946130   0.0912  0.9274
## thetaC:tmrcaC 0.9860233 0.014806396 66.5944  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      0.003
## tmrcaC      0.001 -0.125
## rhoC        -0.004  0.017  0.041
## thetaC:tmrcaC -0.067 -0.040 -0.014  0.076
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.86419788 -0.43551103  0.02063926  0.43146966  6.60989139
##
## Residual standard error: 0.0005120851
## Degrees of freedom: 600 total; 595 residual

vif(g.rep_9)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.018252      1.018249      1.008251      1.008039

g.rep_9.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.50kb.rep_9, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_9.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.50kb.rep_9
##      AIC      BIC      logLik
## -5406.246 -5384.261 2708.123
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.3833499
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept) 0.0205299 0.00017580 116.77810 0.0000
## thetaC      1.1341848 0.01570929 72.19836 0.0000
## rhoC        -0.1665026 0.09061323 -1.83751 0.0666
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.001  0.026
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.88951873 -0.46241592  0.05059189  0.47484598  6.24257165
##
## Residual standard error: 0.002870881
## Degrees of freedom: 600 total; 597 residual
```

```

anova.diversity <- Anova(m.diversity.rep_9)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.50kb[1, 9] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.50kb[2, 9] <- anova.diversity$VarExp[1] * 100
r2.inf.50kb[3, 9] <- anova.diversity$VarExp[2] * 100
r2.inf.50kb[4, 9] <- anova.diversity$VarExp[3] * 100
r2.inf.50kb[5, 9] <- anova.diversity$VarExp[4] * 100

```

### 2.1.10 Replicate 10

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_10/rs.pair_10.", sep = "")
pi.50kb <- read.table(paste(p, "diversity.50kb.bedgraph", sep = ""), header = T)
pi.50kb$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCA.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$avg <- apply(tmrca.50kb[4:ncol(tmrca.50kb)], 1, mean)
rho.50kb <- read.table(paste(p, "rho.50kb.bedgraph", sep = ""), header = T)
theta.50kb <- read.table(paste(p, "theta.50kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.50kb$sim, theta.50kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.50kb$sim and theta.50kb$sample_mean
## S = 558257, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9844928

inf.lands.50kb.rep_10 <- as.data.frame(cbind(pi.50kb$avg, theta.50kb$sample_mean, rho.50kb$sample_mean,
names(inf.lands.50kb.rep_10) <- c("diversity", "theta", "rho", "tmrca")

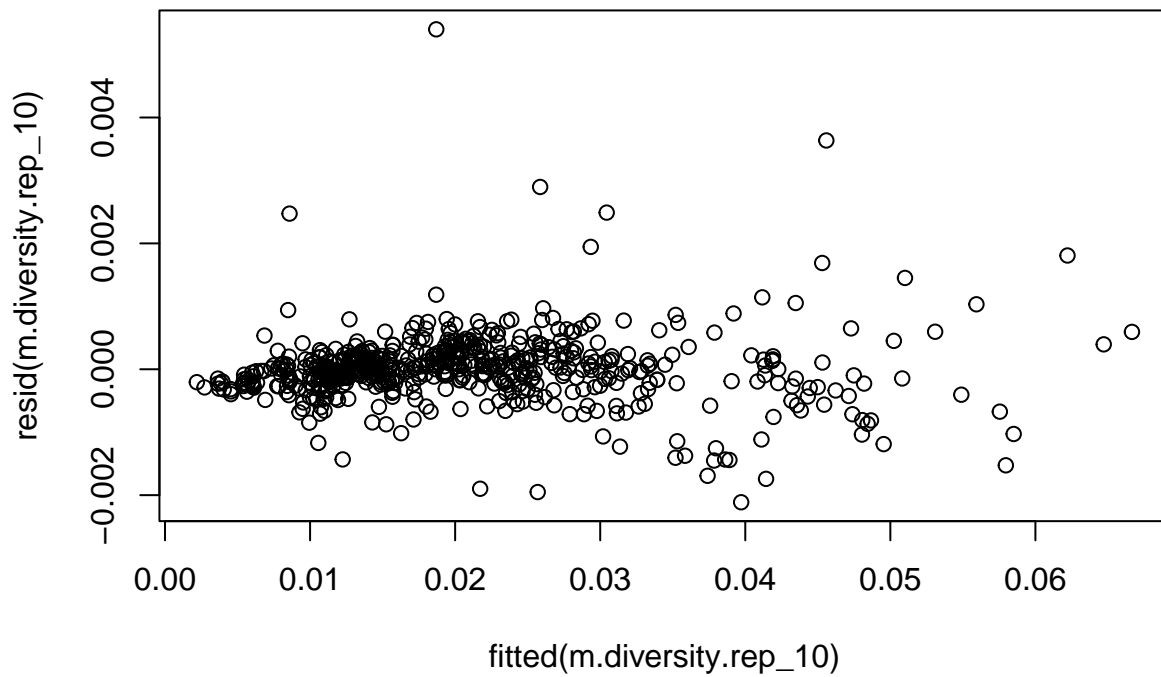
# centering
inf.lands.50kb.rep_10$thetaC <- inf.lands.50kb.rep_10$theta - mean(inf.lands.50kb.rep_10$theta)
inf.lands.50kb.rep_10$tmrcaC <- inf.lands.50kb.rep_10$tmrca - mean(inf.lands.50kb.rep_10$tmrca)
inf.lands.50kb.rep_10$rhoC <- inf.lands.50kb.rep_10$rho - mean(inf.lands.50kb.rep_10$rho)

inf.lands.50kb.rep_10$bin <- 1:nrow(inf.lands.50kb.rep_10)

# for merging:
inf.lands.50kb.rep_10$Replicate <- 10

m.diversity.rep_10 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.50kb.rep_10)
plot(resid(m.diversity.rep_10)~fitted(m.diversity.rep_10))

```



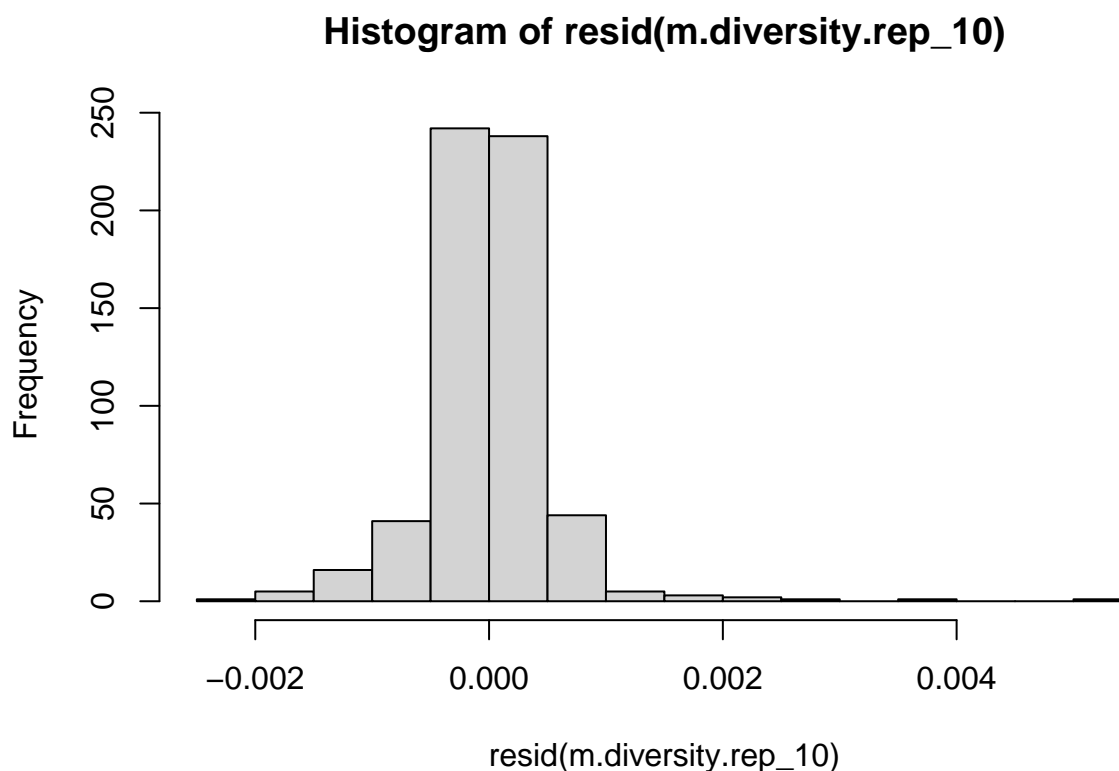
```
dwtest(m.diversity.rep_10)
```

```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_10  
## DW = 1.3449, p-value = 2.3e-16  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_10)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_10  
## HMC = 0.67092, p-value = 1
```

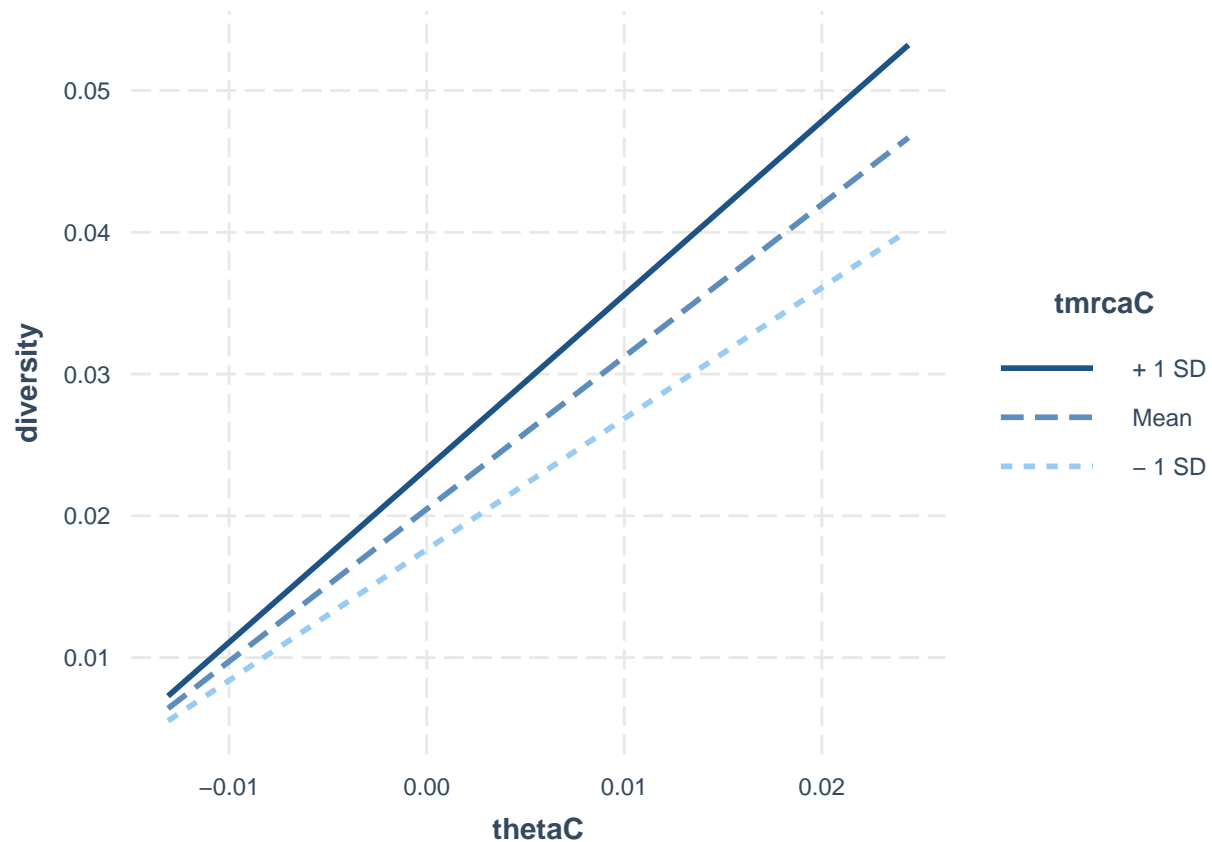
```
hist(resid(m.diversity.rep_10))
```



```
summary(m.diversity.rep_10)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.50kb.rep_10)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0021127 -0.0002121 -0.0000085  0.0001956  0.0054007
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.048e-02  2.313e-05  885.243  <2e-16 ***
## thetaC       1.074e+00  2.407e-03  446.391  <2e-16 ***
## rhoC        -2.139e-02  1.771e-02  -1.208    0.227
## tmrcaC       2.070e-02  1.767e-04  117.152  <2e-16 ***
## thetaC:tmrcaC 1.096e+00  1.562e-02   70.181  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.000561 on 595 degrees of freedom
## Multiple R-squared:  0.9975, Adjusted R-squared:  0.9975
## F-statistic: 6.053e+04 on 4 and 595 DF,  p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_10, pred = thetaC, modx= tmrcaC)
```



```
g.rep_10 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
  data = inf.lands.50kb.rep_1, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_10)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.50kb.rep_1
##      AIC      BIC   logLik
## -7367.645 -7336.866 3690.822
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.3212747
##
## Coefficients:
##              Value   Std.Error   t-value p-value
## (Intercept)  0.0204388 0.000031255  653.9355  0.0000
## thetaC       1.0980191 0.002985790  367.7482  0.0000
## tmrcaC       0.0199958 0.000176128  113.5300  0.0000
## rhoC         0.0035408 0.017040672   0.2078  0.8355
## thetaC:tmrcaC 1.0653107 0.017126226   62.2035  0.0000
```



```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      0.010
## tmrcaC     -0.018 -0.155
## rhoC        -0.003  0.009  0.012
## thetaC:tmrcaC -0.096 -0.101  0.181  0.042
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -5.45341396 -0.37367657  0.05022105  0.36056906  6.43069153
##
## Residual standard error: 0.000544344
## Degrees of freedom: 600 total; 595 residual

vif(g.rep_10)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.030576      1.054551      1.001941      1.041465

g.rep_10.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.50kb.rep_10, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_10.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.50kb.rep_10
##      AIC      BIC    logLik
## -5437.978 -5415.994 2723.989
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.4549714
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept) 0.0207082 0.00019363 106.94544 0.0000
## thetaC      1.1606422 0.01631116 71.15634 0.0000
## rhoC       -0.1660896 0.08076784 -2.05638 0.0402
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.001  0.000
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -3.86544315 -0.50573567  0.01449108  0.52601610  6.72401412
##
## Residual standard error: 0.002899699
## Degrees of freedom: 600 total; 597 residual
```

```

anova.diversity <- Anova(m.diversity.rep_10)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.50kb[1, 10] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) * 100
r2.inf.50kb[2, 10] <- anova.diversity$VarExp[1] * 100
r2.inf.50kb[3, 10] <- anova.diversity$VarExp[2] * 100
r2.inf.50kb[4, 10] <- anova.diversity$VarExp[3] * 100
r2.inf.50kb[5, 10] <- anova.diversity$VarExp[4] * 100

r2.inf.50kb$average <- rowMeans(r2.inf.50kb)
r2.inf.50kb <- transform(r2.inf.50kb, sd=apply(r2.inf.50kb, 1, sd, na.rm = TRUE))

```

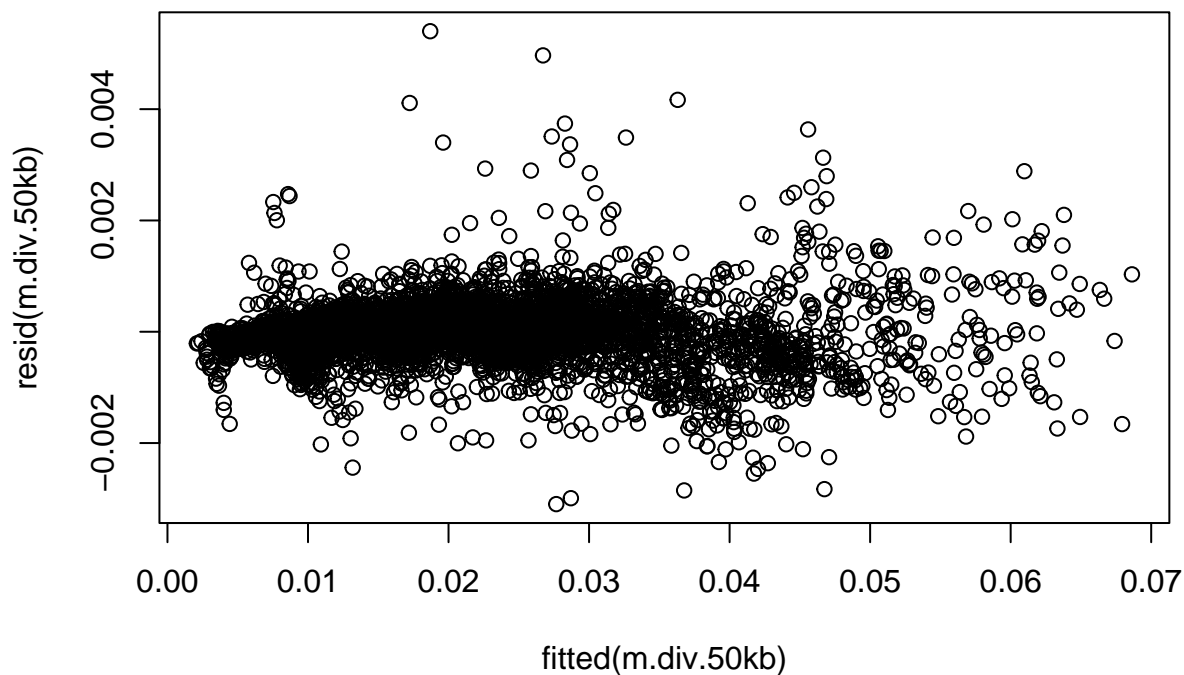
### 2.1.11 all replicates:

```

inf.lands.50kb.all <- rbind(inf.lands.50kb.rep_1, inf.lands.50kb.rep_2, inf.lands.50kb.rep_3, inf.lands.50kb.rep_4, inf.lands.50kb.rep_5, inf.lands.50kb.rep_6, inf.lands.50kb.rep_7, inf.lands.50kb.rep_8, inf.lands.50kb.rep_9, inf.lands.50kb.rep_10)

m.div.50kb <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)*as.factor(Replicate), data = inf.lands.50kb.all)
plot(resid(m.div.50kb)~fitted(m.div.50kb))

```



```
dwtest(m.div.50kb)
```

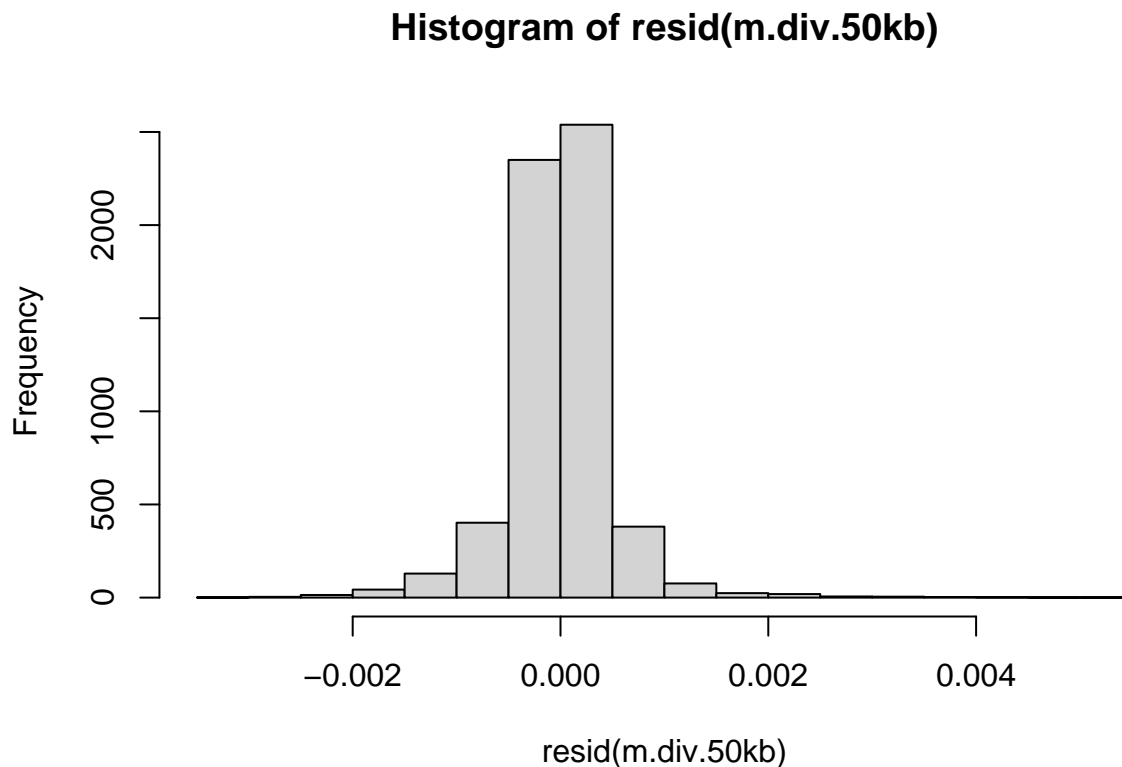
```
##
## Durbin-Watson test
##
```

```
## data: m.div.50kb
## DW = 1.3249, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.50kb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.50366, p-value = 0.642
```

```
hist(resid(m.div.50kb))
```



```
m.div.50kb.2 <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC)*as.factor(Replica
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC + rhoC:tmrcaC)*as.
```

```
AIC(m.div.50kb, m.div.50kb.2, m.div.50kb.3)
```

```
##           df      AIC
## m.div.50kb  51 -73877.32
## m.div.50kb.2 61 -73868.89
## m.div.50kb.3 71 -73992.94
```

```
summary(m.div.50kb)
```

```
##
## Call:
## lm(formula = diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC) *
```

```
##      as.factor(Replicate), data = inf.lands.50kb.all)
##
## Residuals:
##      Min        1Q      Median        3Q        Max
## -0.0030973 -0.0002036  0.0000076  0.0002133  0.0054007
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.043e-02  2.101e-05  972.548 < 2e-16 ***
## thetaC           1.096e+00  2.221e-03  493.438 < 2e-16 ***
## rhoC            -4.461e-03  1.650e-02  -0.270  0.786909
## tmrcaC           2.026e-02  1.541e-04  131.441 < 2e-16 ***
## as.factor(Replicate)2 -2.363e-06  2.968e-05  -0.080  0.936530
## as.factor(Replicate)3  1.266e-04  2.968e-05   4.266  2.02e-05 ***
## as.factor(Replicate)4  1.793e-04  2.961e-05   6.054  1.50e-09 ***
## as.factor(Replicate)5  3.524e-05  2.963e-05   1.189  0.234326
## as.factor(Replicate)6  3.133e-05  2.964e-05   1.057  0.290526
## as.factor(Replicate)7  3.036e-05  2.964e-05   1.024  0.305728
## as.factor(Replicate)8  2.651e-04  2.963e-05   8.947 < 2e-16 ***
## as.factor(Replicate)9 -5.825e-05  2.967e-05  -1.963  0.049696 *
## as.factor(Replicate)10  4.479e-05  2.975e-05   1.506  0.132197
## thetaC:tmrcaC       1.100e+00  1.491e-02  73.755 < 2e-16 ***
## thetaC:as.factor(Replicate)2 -8.685e-03  3.122e-03  -2.782  0.005420 **
## thetaC:as.factor(Replicate)3 -1.871e-02  3.093e-03  -6.049  1.54e-09 ***
## thetaC:as.factor(Replicate)4  2.796e-02  3.119e-03   8.963 < 2e-16 ***
## thetaC:as.factor(Replicate)5  1.215e-02  3.125e-03   3.889  0.000102 ***
## thetaC:as.factor(Replicate)6 -1.002e-02  3.094e-03  -3.240  0.001202 **
## thetaC:as.factor(Replicate)7 -2.283e-03  3.117e-03  -0.732  0.463905
## thetaC:as.factor(Replicate)8  1.151e-02  3.113e-03   3.697  0.000220 ***
## thetaC:as.factor(Replicate)9 -9.252e-03  3.108e-03  -2.977  0.002925 **
## thetaC:as.factor(Replicate)10 -2.131e-02  3.120e-03  -6.832  9.20e-12 ***
## rhoC:as.factor(Replicate)2 -7.206e-03  2.347e-02  -0.307  0.758849
## rhoC:as.factor(Replicate)3 -9.427e-03  2.345e-02  -0.402  0.687674
## rhoC:as.factor(Replicate)4 -6.662e-03  2.349e-02  -0.284  0.776732
## rhoC:as.factor(Replicate)5 -2.473e-03  2.393e-02  -0.103  0.917680
## rhoC:as.factor(Replicate)6  1.381e-02  2.306e-02   0.599  0.549449
## rhoC:as.factor(Replicate)7 -3.538e-02  2.343e-02  -1.510  0.131105
## rhoC:as.factor(Replicate)8 -2.315e-03  2.340e-02  -0.099  0.921183
## rhoC:as.factor(Replicate)9  1.617e-02  2.398e-02   0.674  0.500029
## rhoC:as.factor(Replicate)10 -1.693e-02  2.307e-02  -0.734  0.462966
## tmrcaC:as.factor(Replicate)2 -2.689e-04  2.213e-04  -1.215  0.224342
## tmrcaC:as.factor(Replicate)3 -1.546e-05  2.144e-04  -0.072  0.942513
## tmrcaC:as.factor(Replicate)4 -6.536e-04  2.101e-04  -3.111  0.001874 **
## tmrcaC:as.factor(Replicate)5 -4.184e-04  2.164e-04  -1.933  0.053249 .
## tmrcaC:as.factor(Replicate)6 -9.362e-05  2.235e-04  -0.419  0.675348
## tmrcaC:as.factor(Replicate)7 -4.557e-05  2.224e-04  -0.205  0.837659
## tmrcaC:as.factor(Replicate)8 -1.258e-04  2.152e-04  -0.584  0.558918
## tmrcaC:as.factor(Replicate)9 -3.572e-04  2.248e-04  -1.589  0.112145
## tmrcaC:as.factor(Replicate)10  4.449e-04  2.228e-04   1.997  0.045880 *
## thetaC:tmrcaC:as.factor(Replicate)2 -3.725e-02  2.033e-02  -1.832  0.067009 .
## thetaC:tmrcaC:as.factor(Replicate)3 -8.236e-02  1.966e-02  -4.189  2.84e-05 ***
## thetaC:tmrcaC:as.factor(Replicate)4 -4.519e-02  2.017e-02  -2.241  0.025093 *
## thetaC:tmrcaC:as.factor(Replicate)5 -3.768e-02  1.986e-02  -1.897  0.057874 .
## thetaC:tmrcaC:as.factor(Replicate)6 -4.980e-02  2.051e-02  -2.428  0.015191 *
```

```
## thetaC:tmrcaC:as.factor(Replicate)7 -4.418e-02 2.097e-02 -2.107 0.035148 *
## thetaC:tmrcaC:as.factor(Replicate)8 -2.915e-02 1.990e-02 -1.465 0.142953
## thetaC:tmrcaC:as.factor(Replicate)9 -5.477e-02 2.063e-02 -2.655 0.007954 **
## thetaC:tmrcaC:as.factor(Replicate)10 -3.768e-03 2.060e-02 -0.183 0.854904
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005107 on 5950 degrees of freedom
## Multiple R-squared:  0.9979, Adjusted R-squared:  0.9979
## F-statistic: 5.859e+04 on 49 and 5950 DF,  p-value: < 2.2e-16
```

## 2.1.12 Plots

```
scale.3d <- function(x) sprintf("%.3f", x)

rho.plot <- as.data.frame(cbind(inf.lands.50kb.rep_1$bin,
                                sim.rho.50kb$sim,
                                inf.lands.50kb.rep_1$rho,
                                inf.lands.50kb.rep_2$rho,
                                inf.lands.50kb.rep_3$rho,
                                inf.lands.50kb.rep_4$rho,
                                inf.lands.50kb.rep_5$rho,
                                inf.lands.50kb.rep_6$rho,
                                inf.lands.50kb.rep_7$rho,
                                inf.lands.50kb.rep_8$rho,
                                inf.lands.50kb.rep_9$rho,
                                inf.lands.50kb.rep_10$rho))

rho.plot[,3:ncol(rho.plot)] <- 1.53 * rho.plot[,3:ncol(rho.plot)] # scaling because of SCRM NO vs Ne

names(rho.plot) <- c("bin", "sim", reps)
molten.rho <- melt(rho.plot, id.vars = "bin")
rho.map.50kb <- ggplot(data = molten.rho, aes(x = bin * 50, y = value, colour = variable)) + theme_bw()
rho.map.50kb <- rho.map.50kb + geom_line(data = molten.rho, aes(size = variable)) + scale_size_manual(values = c(1, 2, 3, 4, 5, 6, 7, 8, 9))
rho.map.50kb <- rho.map.50kb + scale_color_manual(values = c("black", brewer.pal(n = 9, name = "Blues")))
rho.map.50kb <- rho.map.50kb + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
rho.map.50kb <- rho.map.50kb + labs(title = NULL, x = NULL, y = expression(rho))
rho.map.50kb <- rho.map.50kb + theme(text = element_text(size = 20), axis.title.x = element_text(size = 20))

theta.plot <- as.data.frame(cbind(inf.lands.50kb.rep_1$bin,
                                sim.theta.50kb$sim,
                                inf.lands.50kb.rep_1$theta,
                                inf.lands.50kb.rep_2$theta,
                                inf.lands.50kb.rep_3$theta,
                                inf.lands.50kb.rep_4$theta,
                                inf.lands.50kb.rep_5$theta,
                                inf.lands.50kb.rep_6$theta,
                                inf.lands.50kb.rep_7$theta,
                                inf.lands.50kb.rep_8$theta,
                                inf.lands.50kb.rep_9$theta,
                                inf.lands.50kb.rep_10$theta))

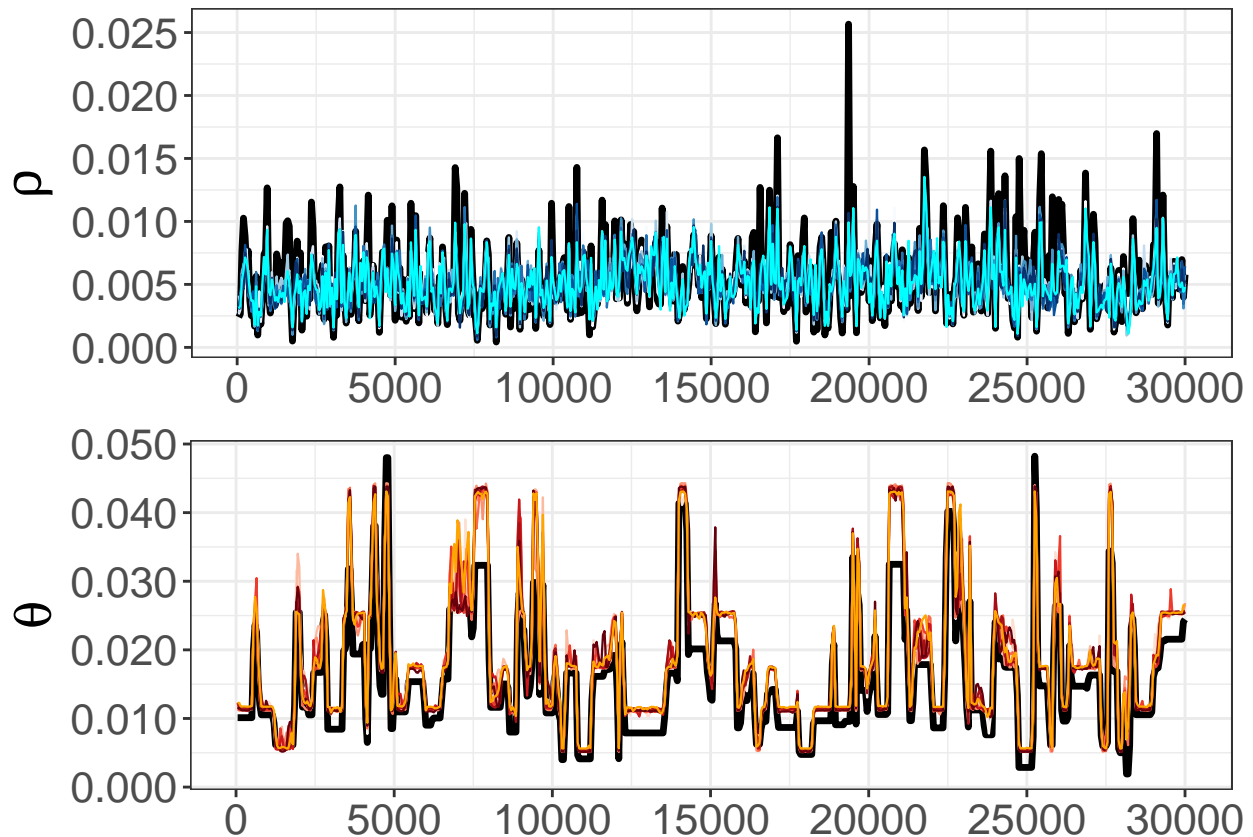
names(theta.plot) <- c("bin", "sim", reps)
molten.theta <- melt(theta.plot, id.vars = "bin")
```

```

theta.map.50kb <- ggplot(data = molten.theta, aes(x = bin * 50, y = value, colour = variable)) + theme_l
theta.map.50kb <- theta.map.50kb + geom_line(data = molten.theta, aes(size = variable)) + scale_size_ma
theta.map.50kb <- theta.map.50kb + scale_color_manual(values = c("black", brewer.pal(n = 9, name = "Red
theta.map.50kb <- theta.map.50kb + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(bro
theta.map.50kb <- theta.map.50kb + labs(title = NULL, x = NULL, y = expression(theta))
theta.map.50kb <- theta.map.50kb + theme(text = element_text(size = 20), axis.title.x = element_text(si

plot_grid(rho.map.50kb, theta.map.50kb, nrow = 2, ncol = 1)

```



## 2.2 200 kb scale

```

r2.inf.200kb <- as.data.frame(matrix(ncol = nreps, nrow = 5))
row.names(r2.inf.200kb) <- c("Total", "Theta", "Rho", "TMRCA", "Theta:TMRCA")
colnames(r2.inf.200kb) <- reps

# sim landscapes
sim.rho.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims/dm.sim.rho.2e+05.txt")
sim.theta.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims/dm.sim.theta.2e+05.txt")
sim.lands.200kb <- as.data.frame(cbind(sim.theta.200kb$sim, sim.rho.200kb$sim))
names(sim.lands.200kb) <- c("theta", "rho")

```

### 2.2.1 Replicate 1

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_1/rs.pair_1.", sep = "")
pi.200kb <- read.table(paste(p, "diversity.200kb.bedgraph", sep = ""), header = T)

```

```

pi.200kb$avg <- apply(pi.200kb[4:ncol(pi.200kb)], 1, mean)
tmrca.200kb <- read.table(paste(p, "TMRCA.200kb.bedgraph", sep = ""), header = T)
tmrca.200kb$avg <- apply(tmrca.200kb[4:ncol(tmrca.200kb)], 1, mean)
rho.200kb <- read.table(paste(p, "rho.200kb.bedgraph", sep = ""), header = T)
theta.200kb <- read.table(paste(p, "theta.200kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.200kb$sim, theta.200kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.200kb$sim and theta.200kb$sample_mean
## S = 8264.3, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.9853072

inf.lands.200kb.rep_1 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_mean,
names(inf.lands.200kb.rep_1) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.200kb.rep_1$thetaC <- inf.lands.200kb.rep_1$theta - mean(inf.lands.200kb.rep_1$theta)
inf.lands.200kb.rep_1$tmrcaC <- inf.lands.200kb.rep_1$tmrca - mean(inf.lands.200kb.rep_1$tmrca)
inf.lands.200kb.rep_1$rhoC <- inf.lands.200kb.rep_1$rho - mean(inf.lands.200kb.rep_1$rho)

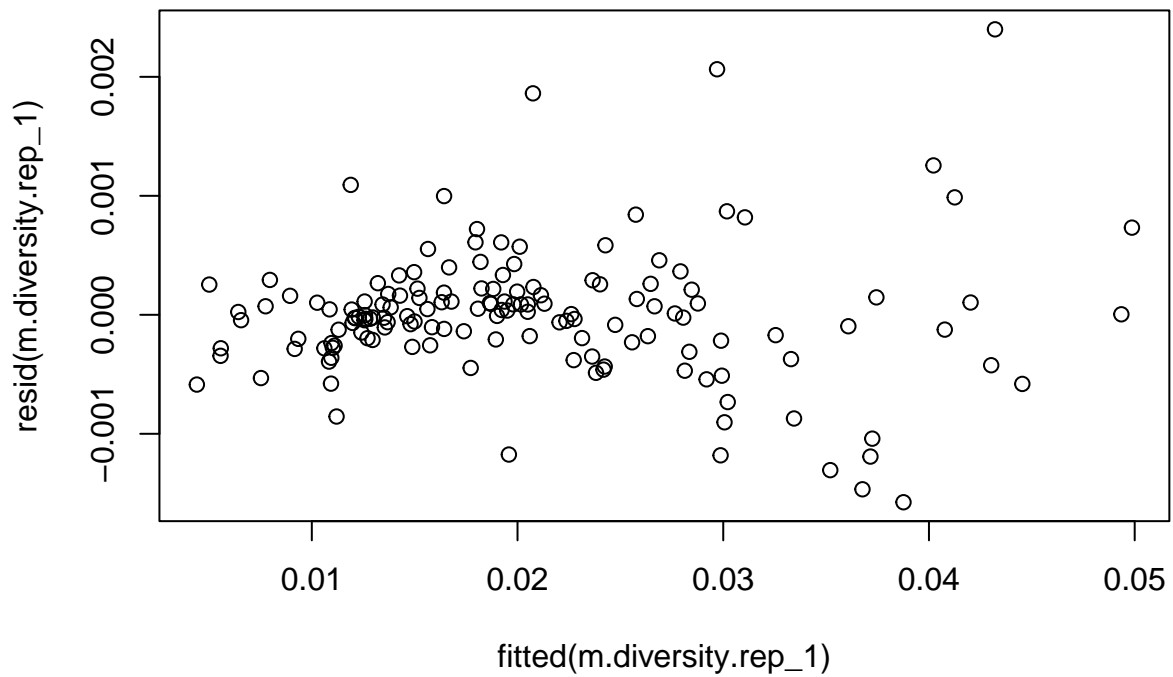
inf.lands.200kb.rep_1$bin <- 1:nrow(inf.lands.200kb.rep_1)

# for merging:
inf.lands.200kb.rep_1$Replicate <- 1

m.diversity.rep_1 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.200kb.rep_1)

plot(resid(m.diversity.rep_1)~fitted(m.diversity.rep_1))

```



```
dwtest(m.diversity.rep_1)
```

```
##
## Durbin-Watson test
##
## data: m.diversity.rep_1
## DW = 1.8429, p-value = 0.1407
## alternative hypothesis: true autocorrelation is greater than 0
```

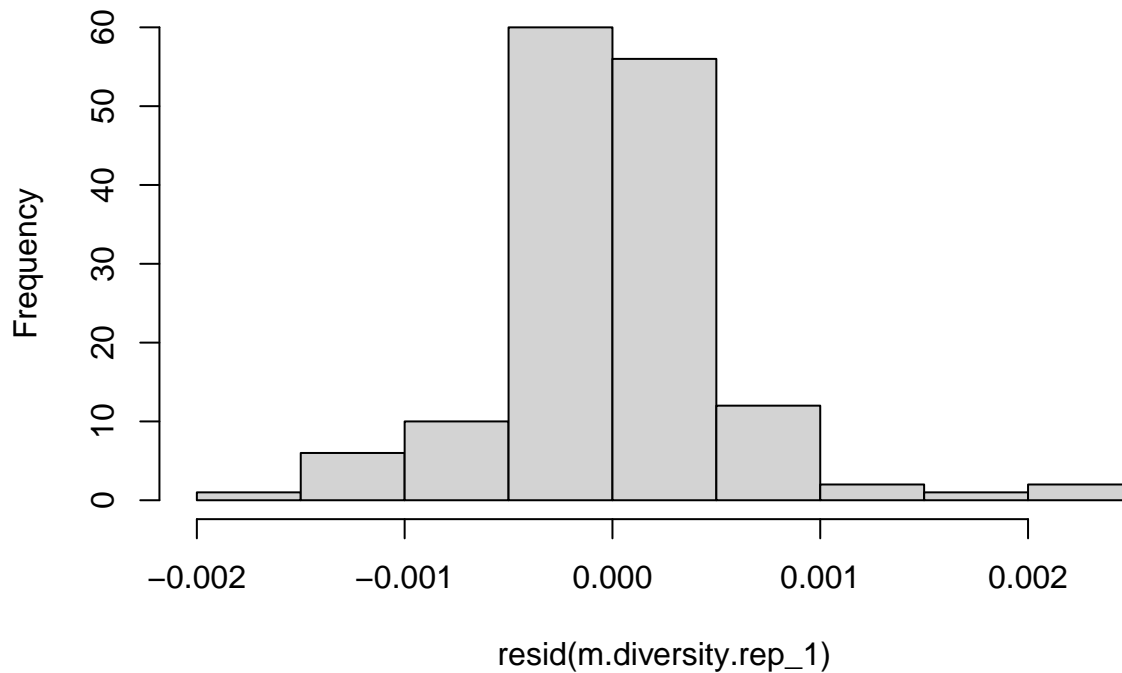
```
hmcetest(m.diversity.rep_1)
```

```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_1
## HMC = 0.49224, p-value = 0.436
```

```
hist(resid(m.diversity.rep_1))
```



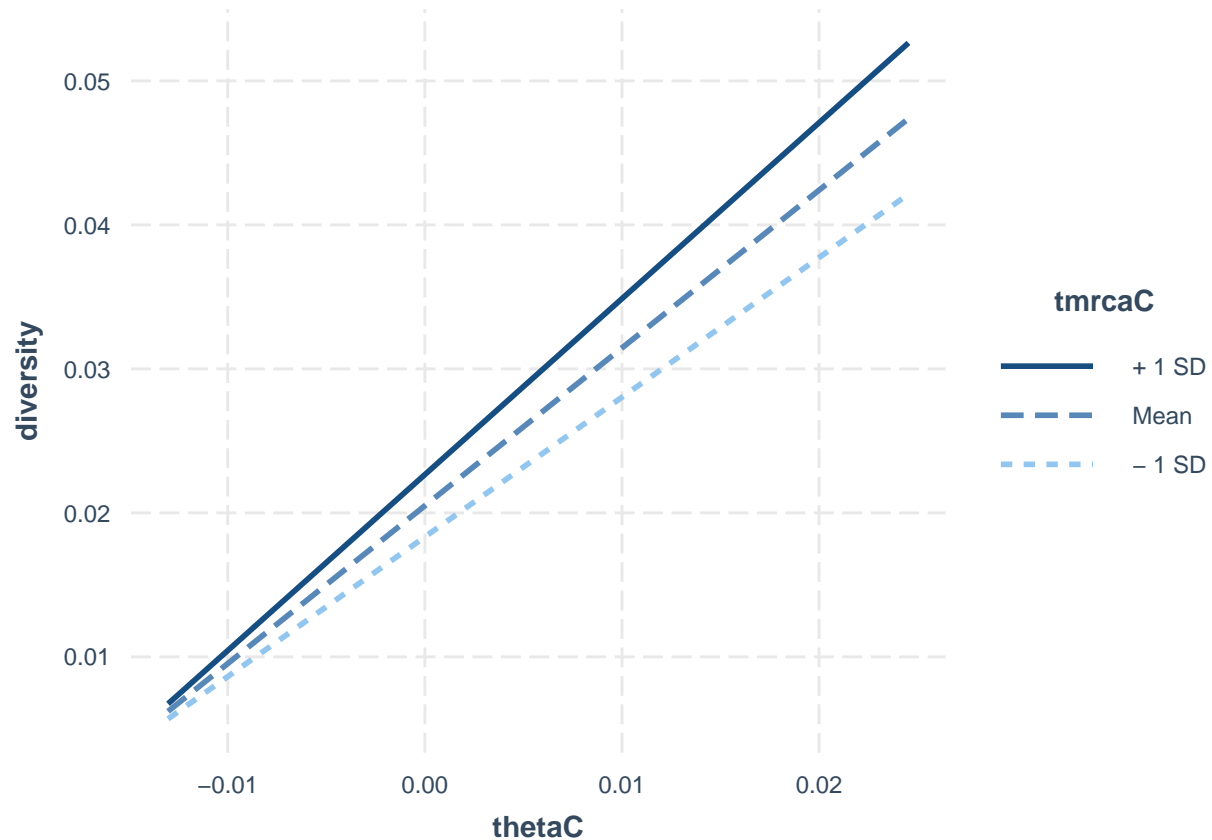
## Histogram of resid(m.diversity.rep\_1)



```
summary(m.diversity.rep_1)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.200kb.rep_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.575e-03 -2.359e-04 -1.035e-05  1.834e-04  2.399e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.049e-02  4.583e-05  447.178  <2e-16 ***
## thetaC       1.096e+00  5.418e-03  202.263  <2e-16 ***
## rhoC        1.426e-02  6.350e-02   0.225    0.823
## tmrcaC       2.103e-02  5.010e-04  41.966  <2e-16 ***
## thetaC:tmrcaC 1.224e+00  5.186e-02  23.595  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005573 on 145 degrees of freedom
## Multiple R-squared:  0.9968, Adjusted R-squared:  0.9967
## F-statistic: 1.131e+04 on 4 and 145 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_1, pred = thetaC, modx= tmrcaC)
```



```
g.rep_1 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.200kb.rep_1, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_1)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.200kb.rep_1
##      AIC      BIC   logLik
## -1814.128 -1793.054  914.0641
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.08562895
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0204958 0.00004989  410.8349  0.0000
## thetaC       1.0967285 0.00562895  194.8371  0.0000
## tmrcaC       0.0210125 0.00050621   41.5092  0.0000
## rhoC         0.0219683 0.06340360    0.3465  0.7295
## thetaC:tmrcaC 1.2110957 0.05322577   22.7539  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.001
## tmrcaC      -0.043 -0.110
## rhoC        -0.010 -0.009  0.248
## thetaC:tmrcaC -0.113  0.010  0.375  0.090
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.94983846 -0.42035716 -0.01013693  0.33115763  4.40016170
##
## Residual standard error: 0.0005481049
## Degrees of freedom: 150 total; 145 residual

vif(g.rep_1)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.015853      1.249427      1.066075      1.167458

g.rep_1.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.200kb.rep_1, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_1.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.200kb.rep_1
##      AIC      BIC    logLik
## -1432.216 -1417.163  721.1082
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.2090144
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept)  0.0206297 0.00020572 100.27847  0.0000
## thetaC      1.1244227 0.02111497  53.25239  0.0000
## rhoC       -0.5711032 0.21962397  -2.60037  0.0103
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.001  0.015
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.46691260 -0.42444799  0.03262027  0.46216734  3.67858710
##
## Residual standard error: 0.002021045
## Degrees of freedom: 150 total; 147 residual
```

```

anova.diversity <- Anova(m.diversity.rep_1)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.200kb[1, 1] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) * 100
r2.inf.200kb[2, 1] <- anova.diversity$VarExp[1] * 100
r2.inf.200kb[3, 1] <- anova.diversity$VarExp[2] * 100
r2.inf.200kb[4, 1] <- anova.diversity$VarExp[3] * 100
r2.inf.200kb[5, 1] <- anova.diversity$VarExp[4] * 100

```

## 2.2.2 Replicate 2

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_2/rs.pair_2.", sep = "")
pi.200kb <- read.table(paste(p, "diversity.200kb.bedgraph", sep = ""), header = T)
pi.200kb$avg <- apply(pi.200kb[4:ncol(pi.200kb)], 1, mean)
tmrca.200kb <- read.table(paste(p, "TMRCA.200kb.bedgraph", sep = ""), header = T)
tmrca.200kb$avg <- apply(tmrca.200kb[4:ncol(tmrca.200kb)], 1, mean)
rho.200kb <- read.table(paste(p, "rho.200kb.bedgraph", sep = ""), header = T)
theta.200kb <- read.table(paste(p, "theta.200kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.200kb$sim, theta.200kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.200kb$sim and theta.200kb$sample_mean
## S = 6148.2, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9890693

inf.lands.200kb.rep_2 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_mean, tmrca.200kb$avg))
names(inf.lands.200kb.rep_2) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.200kb.rep_2$thetaC <- inf.lands.200kb.rep_2$theta - mean(inf.lands.200kb.rep_2$theta)
inf.lands.200kb.rep_2$tmrcaC <- inf.lands.200kb.rep_2$tmrca - mean(inf.lands.200kb.rep_2$tmrca)
inf.lands.200kb.rep_2$rhoC <- inf.lands.200kb.rep_2$rho - mean(inf.lands.200kb.rep_2$rho)

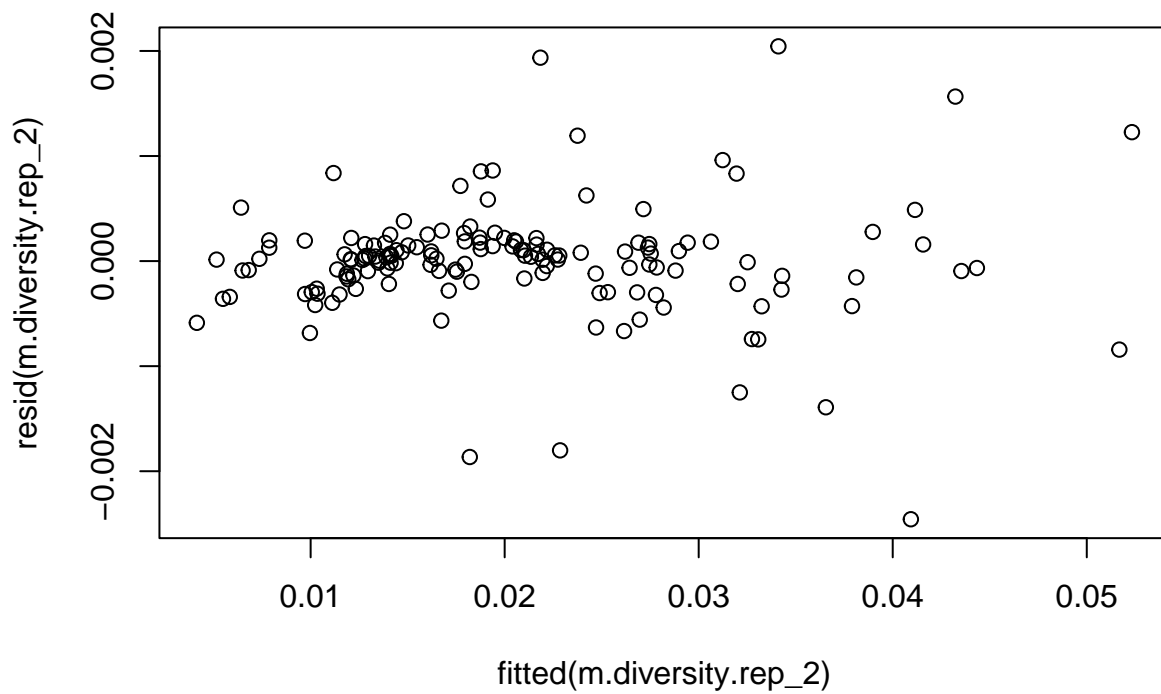
inf.lands.200kb.rep_2$bin <- 1:nrow(inf.lands.200kb.rep_2)

# for merging:
inf.lands.200kb.rep_2$Replicate <- 2

m.diversity.rep_2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.200kb.rep_2)

plot(resid(m.diversity.rep_2)~fitted(m.diversity.rep_2))

```



```
dwtest(m.diversity.rep_2)
```

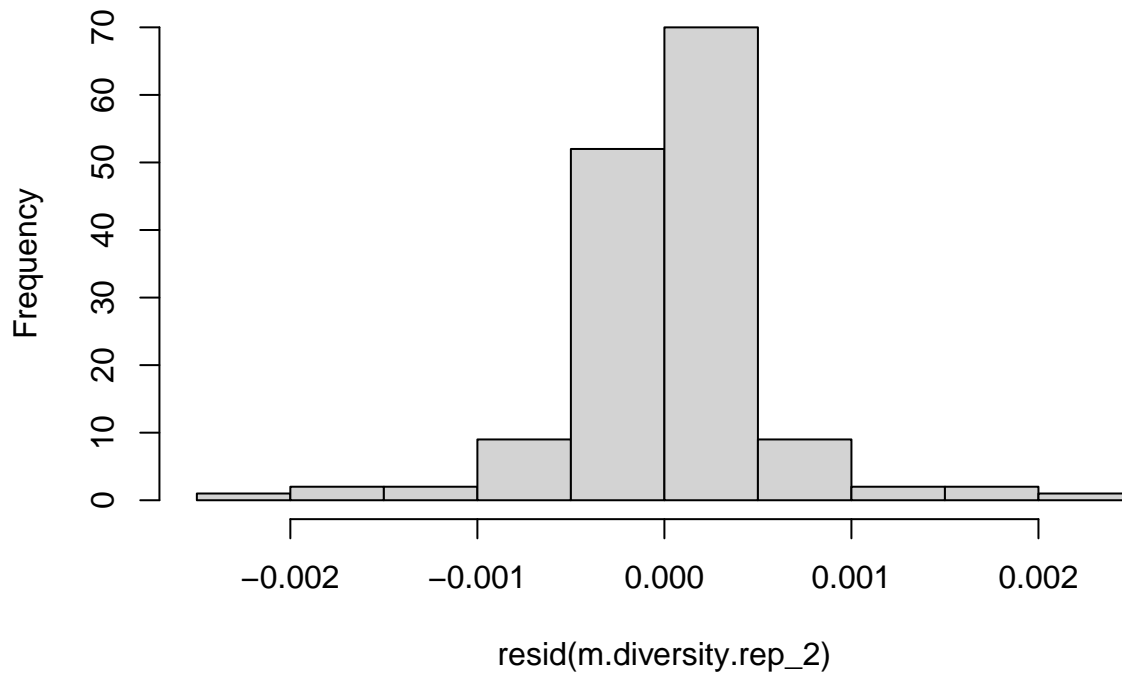
```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_2  
## DW = 1.7156, p-value = 0.03159  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_2)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_2  
## HMC = 0.50531, p-value = 0.541
```

```
hist(resid(m.diversity.rep_2))
```

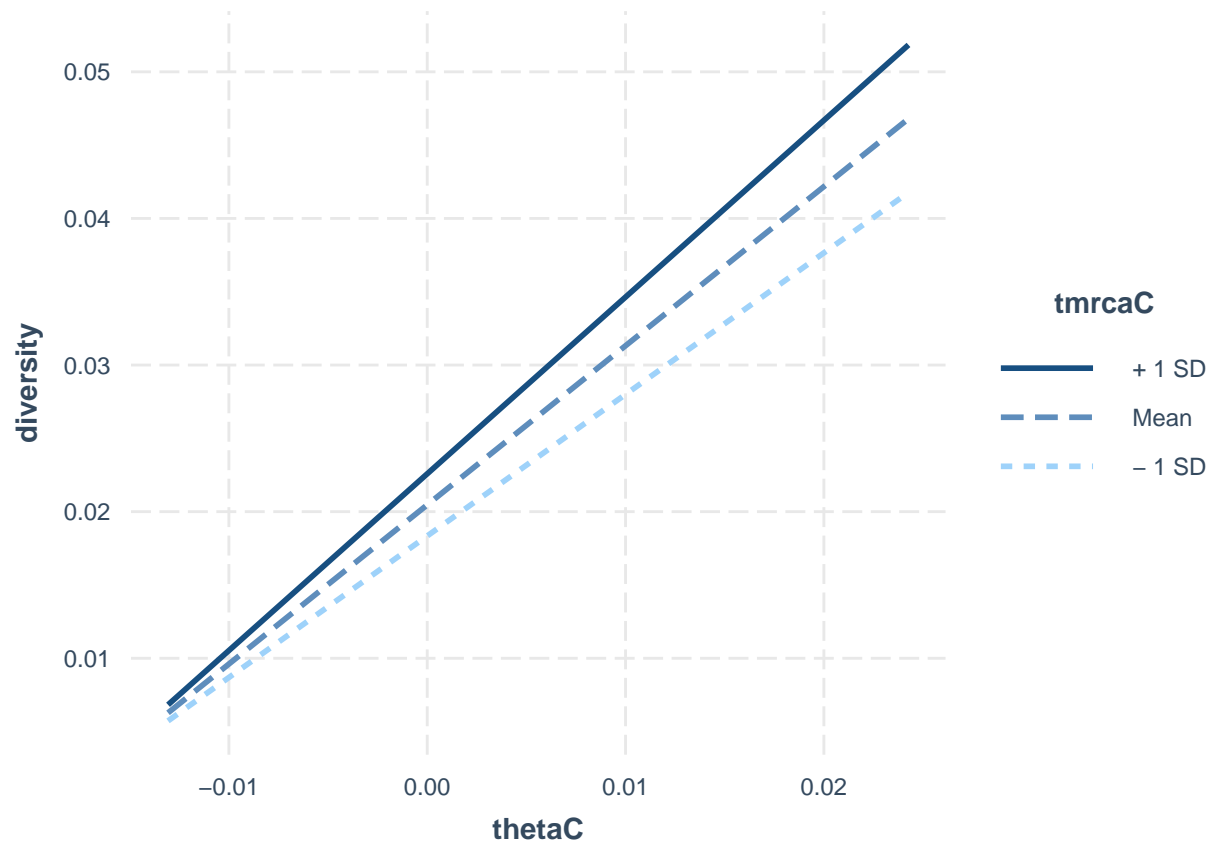
## Histogram of resid(m.diversity.rep\_2)



```
summary(m.diversity.rep_2)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.200kb.rep_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.457e-03 -1.632e-04  2.314e-05  1.700e-04  2.044e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.046e-02  4.515e-05  453.271  <2e-16 ***
## thetaC       1.085e+00  5.297e-03  204.913  <2e-16 ***
## rhoC         9.251e-03  6.537e-02   0.142    0.888
## tmrcaC       2.050e-02  5.178e-04  39.582  <2e-16 ***
## thetaC:tmrcaC 1.161e+00  4.967e-02  23.381  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005487 on 145 degrees of freedom
## Multiple R-squared:  0.9969, Adjusted R-squared:  0.9968
## F-statistic: 1.164e+04 on 4 and 145 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_2, pred = thetaC, modx= tmrcaC)
```



```
g.rep_2 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.200kb.rep_2, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_2)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.200kb.rep_2
##      AIC      BIC   logLik
## -1821.325 -1800.251  917.6627
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.1654346
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0204681 0.00005329 384.0912  0.0000
## thetaC       1.0868465 0.00566891 191.7207  0.0000
## tmrcaC       0.0204136 0.00053121 38.4284  0.0000
## rhoC        -0.0060738 0.06368644 -0.0954  0.9242
## thetaC:tmrcaC 1.1274440 0.05080579 22.1912  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      0.005
## tmrcaC     -0.054 -0.150
## rhoC       -0.013 -0.022  0.171
## thetaC:tmrcaC -0.108 -0.049  0.491  0.125
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.58086608 -0.34056969  0.06376542  0.29976087  3.77133875
##
## Residual standard error: 0.0005405587
## Degrees of freedom: 150 total; 145 residual

vif(g.rep_2)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.023970      1.366753      1.032349      1.320968

g.rep_2.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.200kb.rep_2, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_2.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.200kb.rep_2
##      AIC      BIC    logLik
## -1465.219 -1450.166  737.6095
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.3137584
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept)  0.0206008 0.00021211 97.12366  0.0000
## thetaC      1.1262748 0.01942595 57.97785  0.0000
## rhoC       -0.3984795 0.20018010 -1.99061  0.0484
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.002  0.001
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.31916797 -0.44159324  0.03266396  0.49987953  3.39180105
##
## Residual standard error: 0.001864311
## Degrees of freedom: 150 total; 147 residual
```



```

anova.diversity <- Anova(m.diversity.rep_2)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.200kb[1, 2] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.200kb[2, 2] <- anova.diversity$VarExp[1] * 100
r2.inf.200kb[3, 2] <- anova.diversity$VarExp[2] * 100
r2.inf.200kb[4, 2] <- anova.diversity$VarExp[3] * 100
r2.inf.200kb[5, 2] <- anova.diversity$VarExp[4] * 100

```

### 2.2.3 Replicate 3

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_3/rs.pair_3.", sep = "")
pi.200kb <- read.table(paste(p, "diversity.200kb.bedgraph", sep = ""), header = T)
pi.200kb$avg <- apply(pi.200kb[4:ncol(pi.200kb)], 1, mean)
tmrca.200kb <- read.table(paste(p, "TMRCA.200kb.bedgraph", sep = ""), header = T)
tmrca.200kb$avg <- apply(tmrca.200kb[4:ncol(tmrca.200kb)], 1, mean)
rho.200kb <- read.table(paste(p, "rho.200kb.bedgraph", sep = ""), header = T)
theta.200kb <- read.table(paste(p, "theta.200kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.200kb$sim, theta.200kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.200kb$sim and theta.200kb$sample_mean
## S = 9791.4, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9825924

inf.lands.200kb.rep_3 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_mean,
names(inf.lands.200kb.rep_3) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.200kb.rep_3$thetaC <- inf.lands.200kb.rep_3$theta - mean(inf.lands.200kb.rep_3$theta)
inf.lands.200kb.rep_3$tmrcaC <- inf.lands.200kb.rep_3$tmrca - mean(inf.lands.200kb.rep_3$tmrca)
inf.lands.200kb.rep_3$rhoC <- inf.lands.200kb.rep_3$rho - mean(inf.lands.200kb.rep_3$rho)

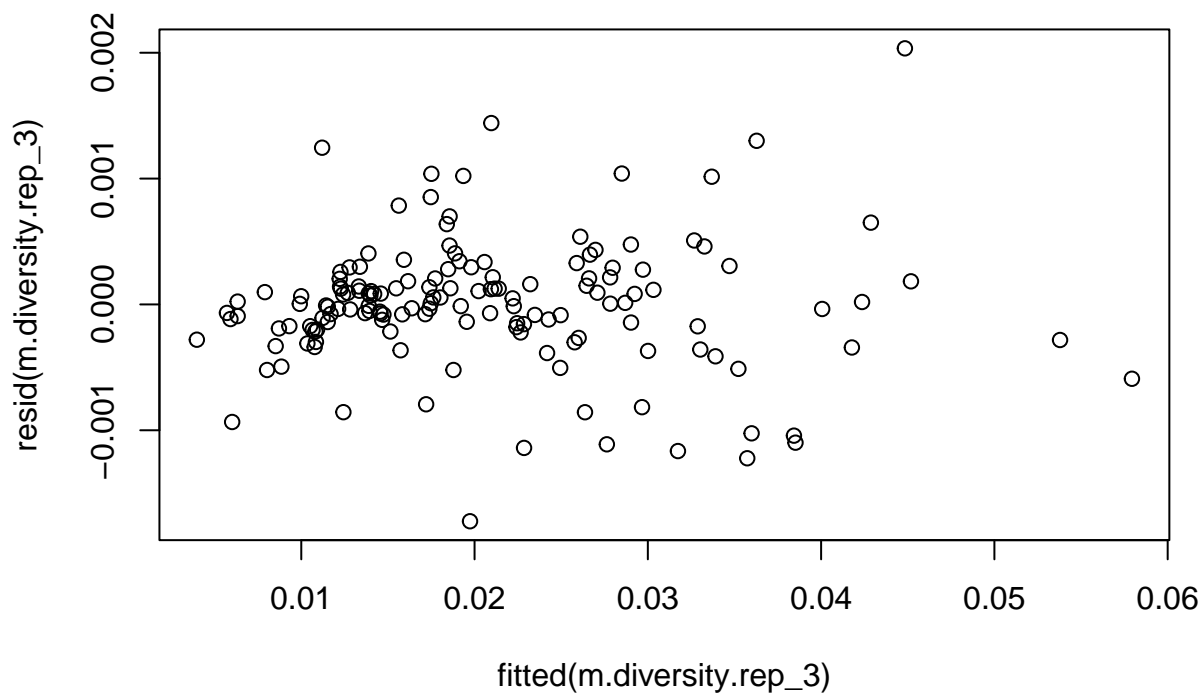
inf.lands.200kb.rep_3$bin <- 1:nrow(inf.lands.200kb.rep_3)

# for merging:
inf.lands.200kb.rep_3$Replicate <- 3

m.diversity.rep_3 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.200kb.rep_3)

plot(resid(m.diversity.rep_3)~fitted(m.diversity.rep_3))

```



```
dwtest(m.diversity.rep_3)
```

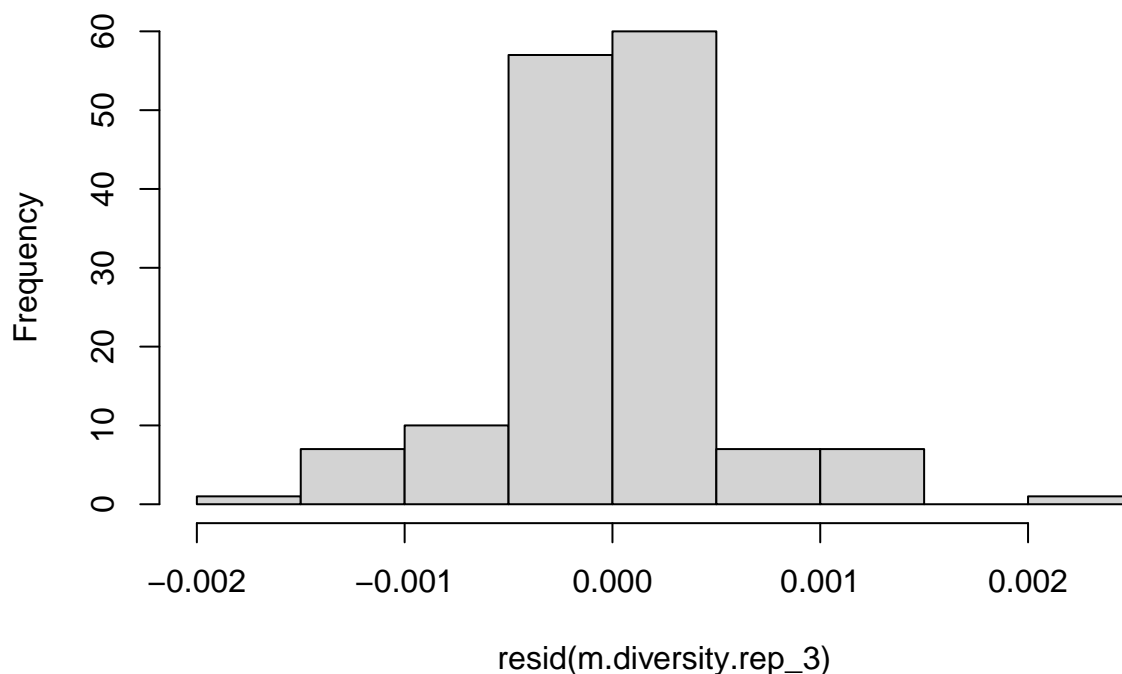
```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_3  
## DW = 1.7271, p-value = 0.03626  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_3)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_3  
## HMC = 0.49269, p-value = 0.443
```

```
hist(resid(m.diversity.rep_3))
```

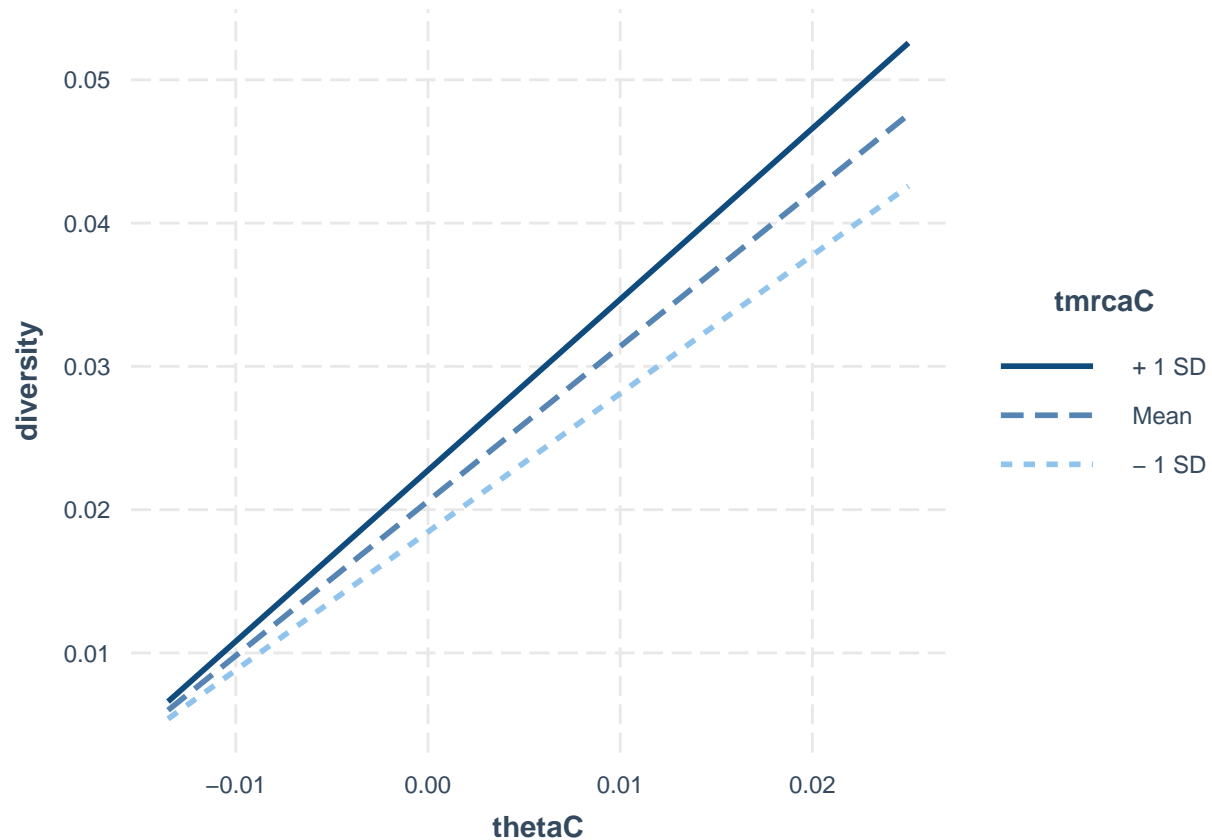
## Histogram of resid(m.diversity.rep\_3)



```
summary(m.diversity.rep_3)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.200kb.rep_3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.723e-03 -1.975e-04 -2.090e-06  2.053e-04  2.035e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.060e-02  4.271e-05  482.307  <2e-16 ***
## thetaC       1.079e+00  4.918e-03  219.494  <2e-16 ***
## rhoC        3.612e-02  6.152e-02   0.587    0.558
## tmrcaC       2.055e-02  4.262e-04  48.218  <2e-16 ***
## thetaC:tmrcaC 1.092e+00  3.816e-02  28.629  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005196 on 145 degrees of freedom
## Multiple R-squared:  0.9974, Adjusted R-squared:  0.9973
## F-statistic: 1.387e+04 on 4 and 145 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_3, pred = thetaC, modx= tmrcaC)
```



```
g.rep_3 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.200kb.rep_3, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_3)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.200kb.rep_3
##      AIC      BIC    logLik
## -1837.259 -1816.184  925.6294
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.1534496
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0206000 0.00004979  413.7691  0.0000
## thetaC       1.0808335 0.00523730  206.3723  0.0000
## tmrcaC       0.0205864 0.00043609   47.2067  0.0000
## rhoC        0.0431677 0.06142122    0.7028  0.4833
## thetaC:tmrcaC 1.0712178 0.03947149   27.1390  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      0.006
## tmrcaC     -0.020 -0.153
## rhoC        -0.009 -0.025  0.167
## thetaC:tmrcaC -0.101 -0.061  0.181  0.089
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -3.38419912 -0.35874981 -0.00849962  0.40321853  4.03279383
##
## Residual standard error: 0.0005116005
## Degrees of freedom: 150 total; 145 residual

vif(g.rep_3)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.025025      1.081264      1.032608      1.039063

g.rep_3.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.200kb.rep_3, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_3.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.200kb.rep_3
##      AIC      BIC    logLik
## -1403.367 -1388.314  706.6834
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.2774686
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept)  0.0207429 0.00024772 83.73477  0.000
## thetaC      1.1273552 0.02313319 48.73324  0.000
## rhoC       -0.5122430 0.25571026 -2.00322  0.047
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.001  0.003
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -3.57656146 -0.41478403  0.08100049  0.42358182  3.69818275
##
## Residual standard error: 0.002264567
## Degrees of freedom: 150 total; 147 residual
```

```

anova.diversity <- Anova(m.diversity.rep_3)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.200kb[1, 3] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.200kb[2, 3] <- anova.diversity$VarExp[1] * 100
r2.inf.200kb[3, 3] <- anova.diversity$VarExp[2] * 100
r2.inf.200kb[4, 3] <- anova.diversity$VarExp[3] * 100
r2.inf.200kb[5, 3] <- anova.diversity$VarExp[4] * 100

```

## 2.2.4 Replicate 4

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_4/rs.pair_4.", sep = "")
pi.200kb <- read.table(paste(p, "diversity.200kb.bedgraph", sep = ""), header = T)
pi.200kb$avg <- apply(pi.200kb[4:ncol(pi.200kb)], 1, mean)
tmrca.200kb <- read.table(paste(p, "TMRCA.200kb.bedgraph", sep = ""), header = T)
tmrca.200kb$avg <- apply(tmrca.200kb[4:ncol(tmrca.200kb)], 1, mean)
rho.200kb <- read.table(paste(p, "rho.200kb.bedgraph", sep = ""), header = T)
theta.200kb <- read.table(paste(p, "theta.200kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.200kb$sim, theta.200kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.200kb$sim and theta.200kb$sample_mean
## S = 9866.4, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.982459

inf.lands.200kb.rep_4 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_mean))
names(inf.lands.200kb.rep_4) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.200kb.rep_4$thetaC <- inf.lands.200kb.rep_4$theta - mean(inf.lands.200kb.rep_4$theta)
inf.lands.200kb.rep_4$tmrcaC <- inf.lands.200kb.rep_4$tmrca - mean(inf.lands.200kb.rep_4$tmrca)
inf.lands.200kb.rep_4$rhoC <- inf.lands.200kb.rep_4$rho - mean(inf.lands.200kb.rep_4$rho)

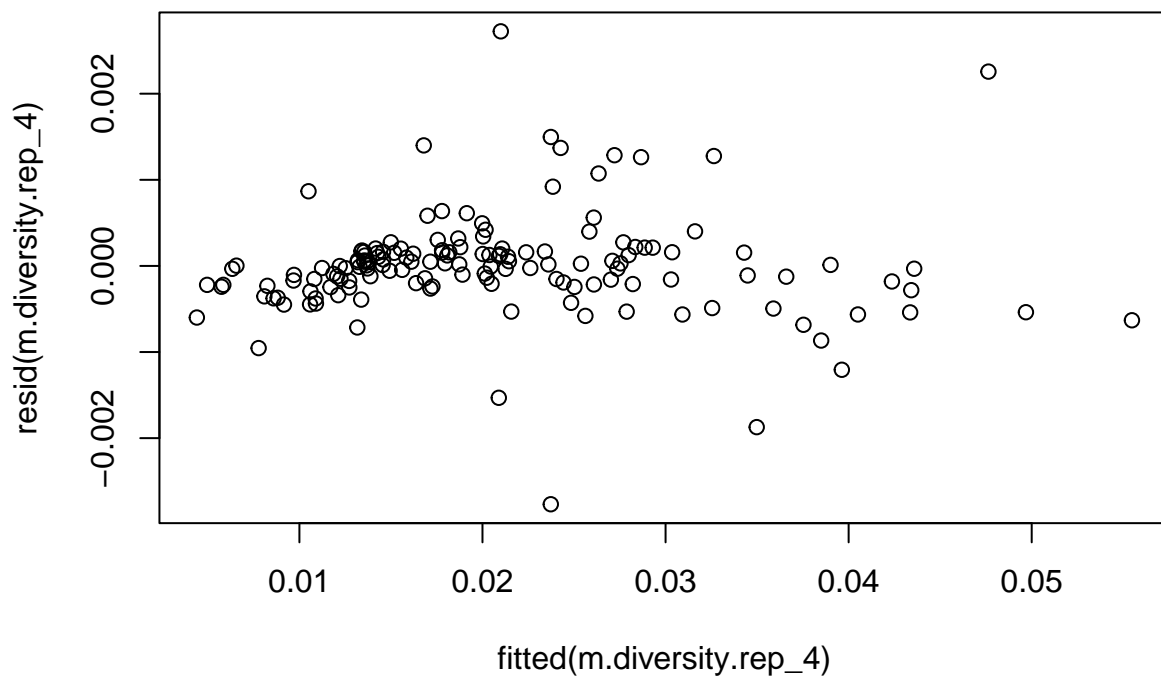
inf.lands.200kb.rep_4$bin <- 1:nrow(inf.lands.200kb.rep_4)

# for merging:
inf.lands.200kb.rep_4$Replicate <- 4

m.diversity.rep_4 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.200kb.rep_4)

plot(resid(m.diversity.rep_4)~fitted(m.diversity.rep_4))

```



```
dwtest(m.diversity.rep_4)
```

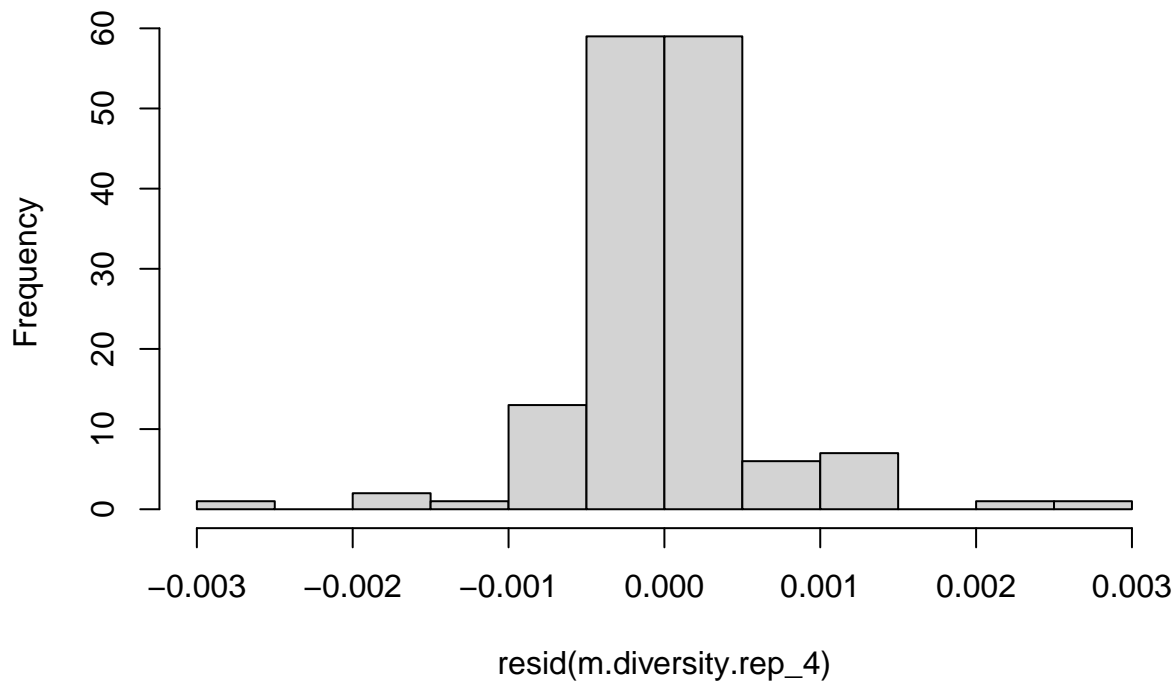
```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_4  
## DW = 1.6283, p-value = 0.00823  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_4)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_4  
## HMC = 0.54628, p-value = 0.789
```

```
hist(resid(m.diversity.rep_4))
```

## Histogram of resid(m.diversity.rep\_4)

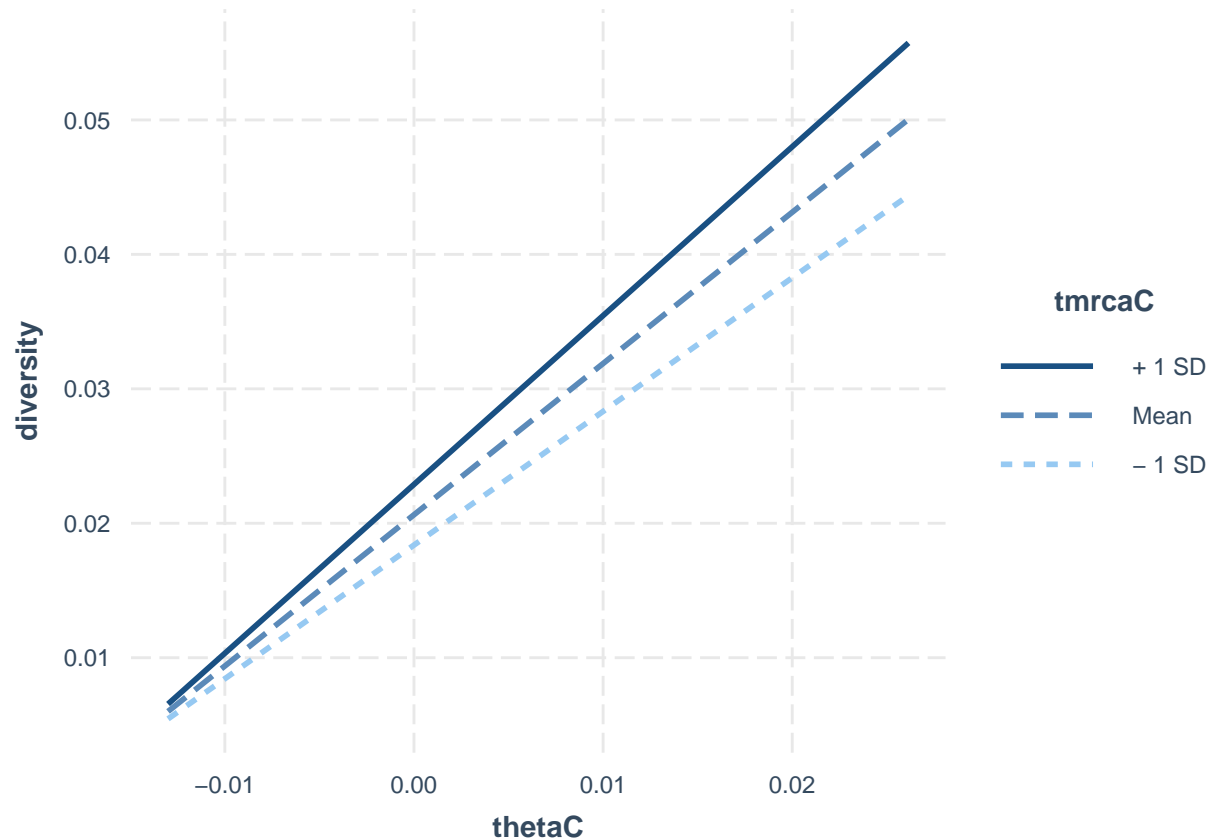


```
summary(m.diversity.rep_4)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.200kb.rep_4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.768e-03 -2.282e-04 -4.140e-06  1.588e-04  2.724e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.063e-02  4.965e-05  415.535  <2e-16 ***
## thetaC       1.125e+00  5.855e-03  192.138  <2e-16 ***
## rhoC        6.411e-02  7.349e-02   0.872   0.384
## tmrcaC       1.975e-02  4.667e-04  42.319  <2e-16 ***
## thetaC:tmrcaC 1.146e+00  5.168e-02  22.178  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0006075 on 145 degrees of freedom
## Multiple R-squared:  0.9963, Adjusted R-squared:  0.9962
## F-statistic: 9799 on 4 and 145 DF, p-value: < 2.2e-16
```



```
interact_plot(m.diversity.rep_4, pred = thetaC, modx= tmrcaC)
```



```
g.rep_4 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.200kb.rep_4, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_4)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.200kb.rep_4
##      AIC      BIC   logLik
## -1792.908 -1771.834  903.4542
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.1996946
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0206346 0.00006076 339.6014  0.0000
## thetaC       1.1223743 0.00626603 179.1204  0.0000
## tmrcaC       0.0198270 0.00047759  41.5143  0.0000
## rhoC        0.0706769 0.07071306   0.9995  0.3192
## thetaC:tmrcaC 1.1216837 0.05373238  20.8754  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.003
## tmrcaC      -0.012 -0.039
## rhoC         -0.002  0.018  0.182
## thetaC:tmrcaC -0.040  0.064  0.264  0.027
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.63571446 -0.38925930 -0.02298232  0.24825431  4.56473653
##
## Residual standard error: 0.0005980989
## Degrees of freedom: 150 total; 145 residual

vif(g.rep_4)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.008269      1.115533      1.035414      1.081806

g.rep_4.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.200kb.rep_4, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_4.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.200kb.rep_4
##      AIC      BIC    logLik
## -1406.204 -1391.151  708.1021
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.2791937
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept)  0.0206898 0.00024597 84.11461 0.0000
## thetaC      1.1309856 0.02340816 48.31587 0.0000
## rhoC       -0.4082832 0.24994231 -1.63351 0.1045
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.00
## rhoC   0.00  0.02
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.9242575 -0.4591282  0.0538119  0.4694468  4.4837294
##
## Residual standard error: 0.002244411
## Degrees of freedom: 150 total; 147 residual
```

```

anova.diversity <- Anova(m.diversity.rep_4)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.200kb[1, 4] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) * 100
r2.inf.200kb[2, 4] <- anova.diversity$VarExp[1] * 100
r2.inf.200kb[3, 4] <- anova.diversity$VarExp[2] * 100
r2.inf.200kb[4, 4] <- anova.diversity$VarExp[3] * 100
r2.inf.200kb[5, 4] <- anova.diversity$VarExp[4] * 100

```

### 2.2.5 Replicate 5

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_5/rs.pair_5.", sep = "")
pi.200kb <- read.table(paste(p, "diversity.200kb.bedgraph", sep = ""), header = T)
pi.200kb$avg <- apply(pi.200kb[4:ncol(pi.200kb)], 1, mean)
tmrca.200kb <- read.table(paste(p, "TMRCA.200kb.bedgraph", sep = ""), header = T)
tmrca.200kb$avg <- apply(tmrca.200kb[4:ncol(tmrca.200kb)], 1, mean)
rho.200kb <- read.table(paste(p, "rho.200kb.bedgraph", sep = ""), header = T)
theta.200kb <- read.table(paste(p, "theta.200kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.200kb$sim, theta.200kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.200kb$sim and theta.200kb$sample_mean
## S = 7360.3, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9869145

inf.lands.200kb.rep_5 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_mean, tmrca.200kb$avg))
names(inf.lands.200kb.rep_5) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.200kb.rep_5$thetaC <- inf.lands.200kb.rep_5$theta - mean(inf.lands.200kb.rep_5$theta)
inf.lands.200kb.rep_5$tmrcaC <- inf.lands.200kb.rep_5$tmrca - mean(inf.lands.200kb.rep_5$tmrca)
inf.lands.200kb.rep_5$rhoC <- inf.lands.200kb.rep_5$rho - mean(inf.lands.200kb.rep_5$rho)

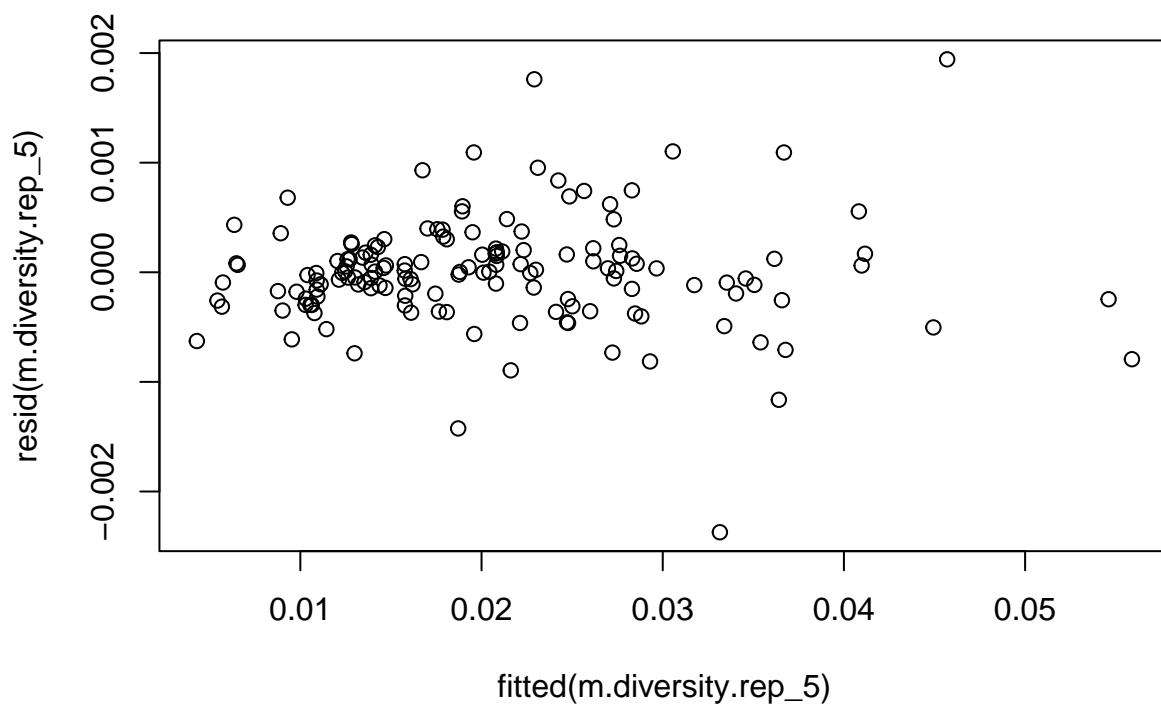
inf.lands.200kb.rep_5$bin <- 1:nrow(inf.lands.200kb.rep_5)

# for merging:
inf.lands.200kb.rep_5$Replicate <- 5

m.diversity.rep_5 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.200kb.rep_5)

plot(resid(m.diversity.rep_5)~fitted(m.diversity.rep_5))

```



```
dwtest(m.diversity.rep_5)
```

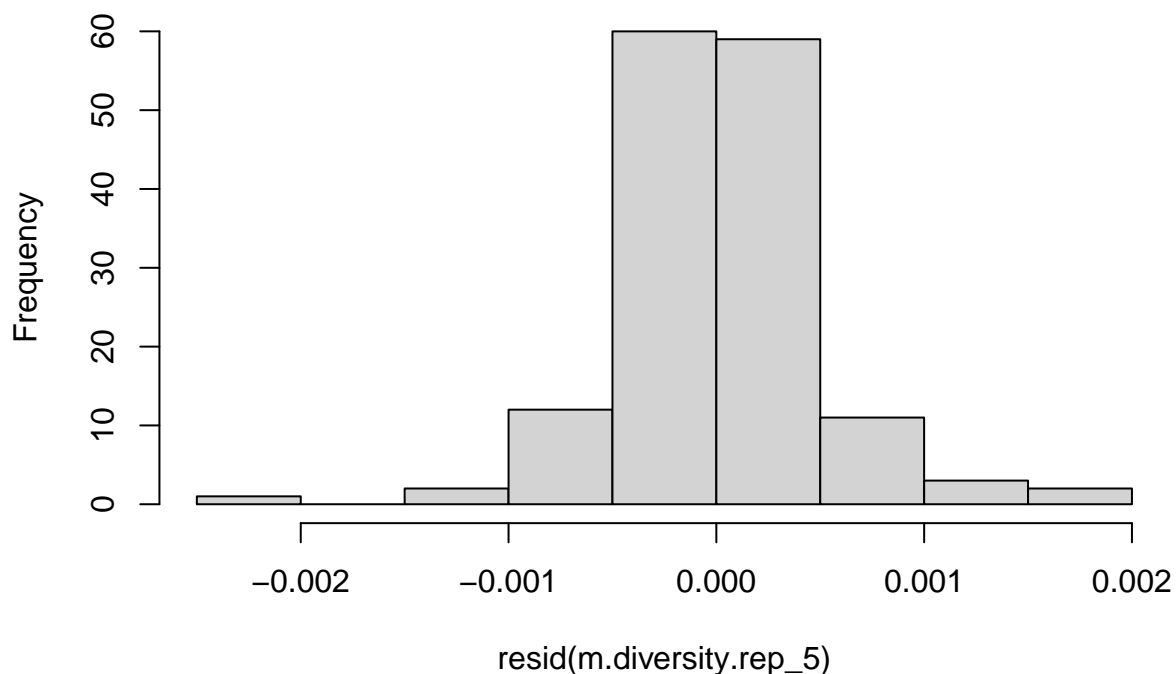
```
##
## Durbin-Watson test
##
## data: m.diversity.rep_5
## DW = 1.473, p-value = 0.0003904
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_5)
```

```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_5
## HMC = 0.42356, p-value = 0.1
```

```
hist(resid(m.diversity.rep_5))
```

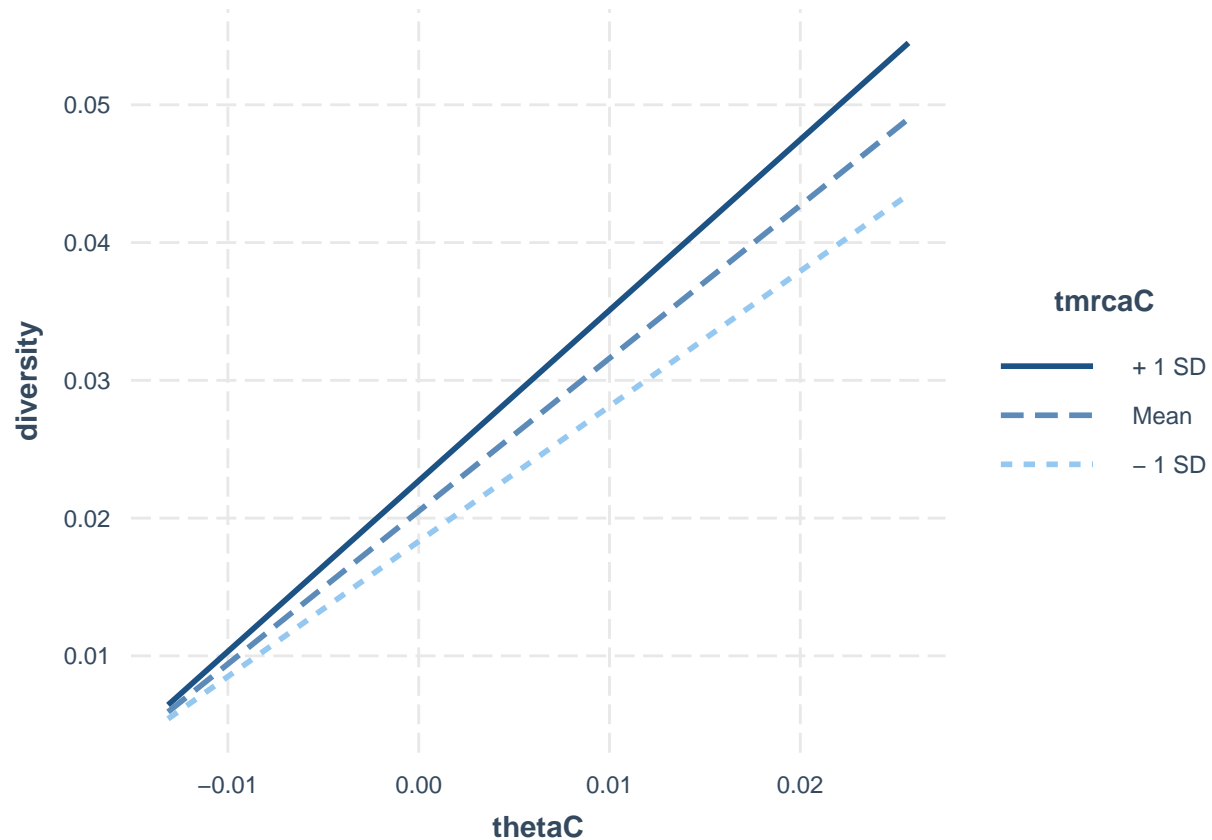
## Histogram of resid(m.diversity.rep\_5)



```
summary(m.diversity.rep_5)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.200kb.rep_5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0023718 -0.0002435  0.0000000  0.0001819  0.0019428
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.051e-02  4.143e-05  495.020  <2e-16 ***
## thetaC       1.109e+00  4.896e-03  226.607  <2e-16 ***
## rhoC         8.631e-02  6.272e-02   1.376    0.171
## tmrcaC       2.045e-02  4.162e-04   49.148  <2e-16 ***
## thetaC:tmrcaC 1.199e+00  3.896e-02   30.786  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005067 on 145 degrees of freedom
## Multiple R-squared:  0.9974, Adjusted R-squared:  0.9973
## F-statistic: 1.37e+04 on 4 and 145 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_5, pred = thetaC, modx= tmrcaC)
```



```
g.rep_5 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.200kb.rep_5, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_5)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.200kb.rep_5
##      AIC      BIC    logLik
## -1853.508 -1832.434  933.7542
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.2890306
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0205091 0.00005581 367.4517  0.0000
## thetaC       1.1068824 0.00529802 208.9238  0.0000
## tmrcaC       0.0206087 0.00042265  48.7607  0.0000
## rhoC        0.0912157 0.06030229   1.5126  0.1325
## thetaC:tmrcaC 1.1670320 0.03885841  30.0329  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.001
## tmrcaC      -0.009 -0.098
## rhoC        -0.006 -0.045  0.198
## thetaC:tmrcaC -0.038  0.025  0.256  0.166
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.69081849 -0.45570753 -0.04325154  0.33563272  4.08749773
##
## Residual standard error: 0.0005001316
## Degrees of freedom: 150 total; 145 residual

vif(g.rep_5)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.013524      1.111015      1.057373      1.089764

g.rep_5.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.200kb.rep_5, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_5.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.200kb.rep_5
##      AIC      BIC      logLik
## -1410.762 -1395.709  710.3812
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.2658747
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept)  0.0205716 0.00023791 86.46709 0.0000
## thetaC      1.1258771 0.02302967 48.88811 0.0000
## rhoC       -0.6236980 0.25913685 -2.40683 0.0173
##
## Correlation:
##      (Intr) thetaC
## thetaC  0.000
## rhoC    0.000 -0.032
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -3.2105114 -0.4827096  0.0393999  0.4735940  3.9755306
##
## Residual standard error: 0.002201973
## Degrees of freedom: 150 total; 147 residual
```

```

anova.diversity <- Anova(m.diversity.rep_5)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.200kb[1, 5] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4] + anova.diversity$VarExp[5]) / anova.diversity$VarExp[5]
r2.inf.200kb[2, 5] <- anova.diversity$VarExp[1] * 100
r2.inf.200kb[3, 5] <- anova.diversity$VarExp[2] * 100
r2.inf.200kb[4, 5] <- anova.diversity$VarExp[3] * 100
r2.inf.200kb[5, 5] <- anova.diversity$VarExp[4] * 100

```

## 2.2.6 Replicate 6

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_6/rs.pair_6.", sep = "")
pi.200kb <- read.table(paste(p, "diversity.200kb.bedgraph", sep = ""), header = T)
pi.200kb$avg <- apply(pi.200kb[4:ncol(pi.200kb)], 1, mean)
tmrca.200kb <- read.table(paste(p, "TMRCA.200kb.bedgraph", sep = ""), header = T)
tmrca.200kb$avg <- apply(tmrca.200kb[4:ncol(tmrca.200kb)], 1, mean)
rho.200kb <- read.table(paste(p, "rho.200kb.bedgraph", sep = ""), header = T)
theta.200kb <- read.table(paste(p, "theta.200kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.200kb$sim, theta.200kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.200kb$sim and theta.200kb$sample_mean
## S = 7436.3, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9867794

inf.lands.200kb.rep_6 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_mean, tmrca.200kb$avg))
names(inf.lands.200kb.rep_6) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.200kb.rep_6$thetaC <- inf.lands.200kb.rep_6$theta - mean(inf.lands.200kb.rep_6$theta)
inf.lands.200kb.rep_6$tmrcaC <- inf.lands.200kb.rep_6$tmrca - mean(inf.lands.200kb.rep_6$tmrca)
inf.lands.200kb.rep_6$rhoC <- inf.lands.200kb.rep_6$rho - mean(inf.lands.200kb.rep_6$rho)

inf.lands.200kb.rep_6$bin <- 1:nrow(inf.lands.200kb.rep_6)

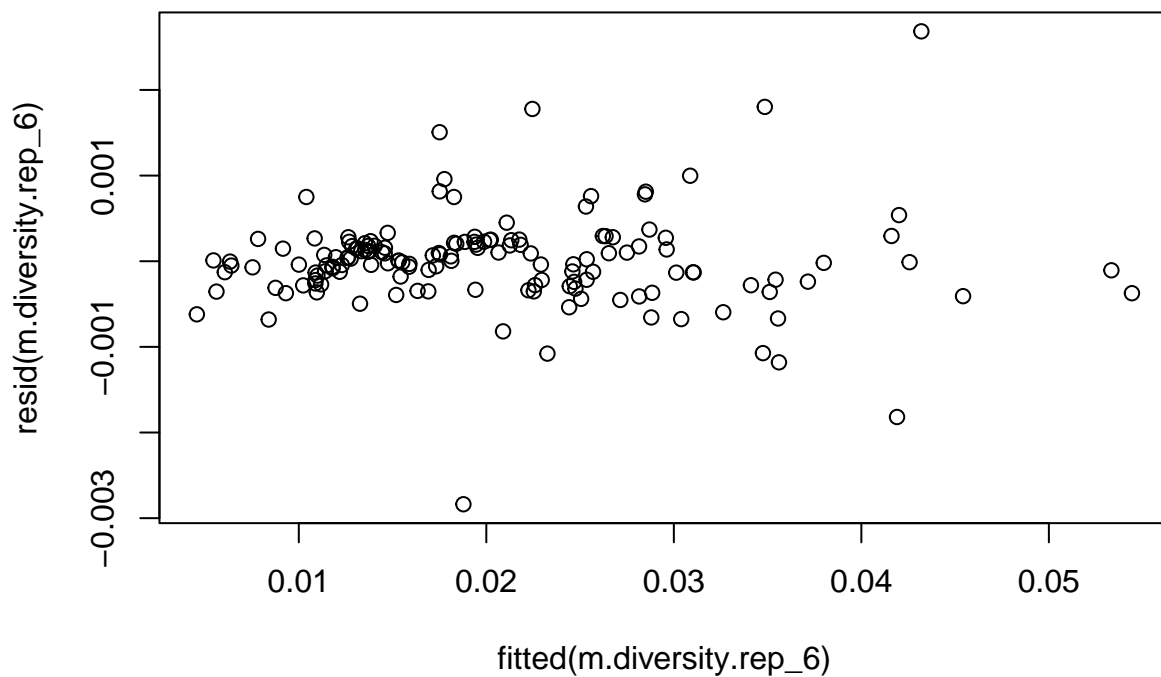
# for merging:
inf.lands.200kb.rep_6$Replicate <- 6

m.diversity.rep_6 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.200kb.rep_6)

plot(resid(m.diversity.rep_6)~fitted(m.diversity.rep_6))

```





```
dwtest(m.diversity.rep_6)
```

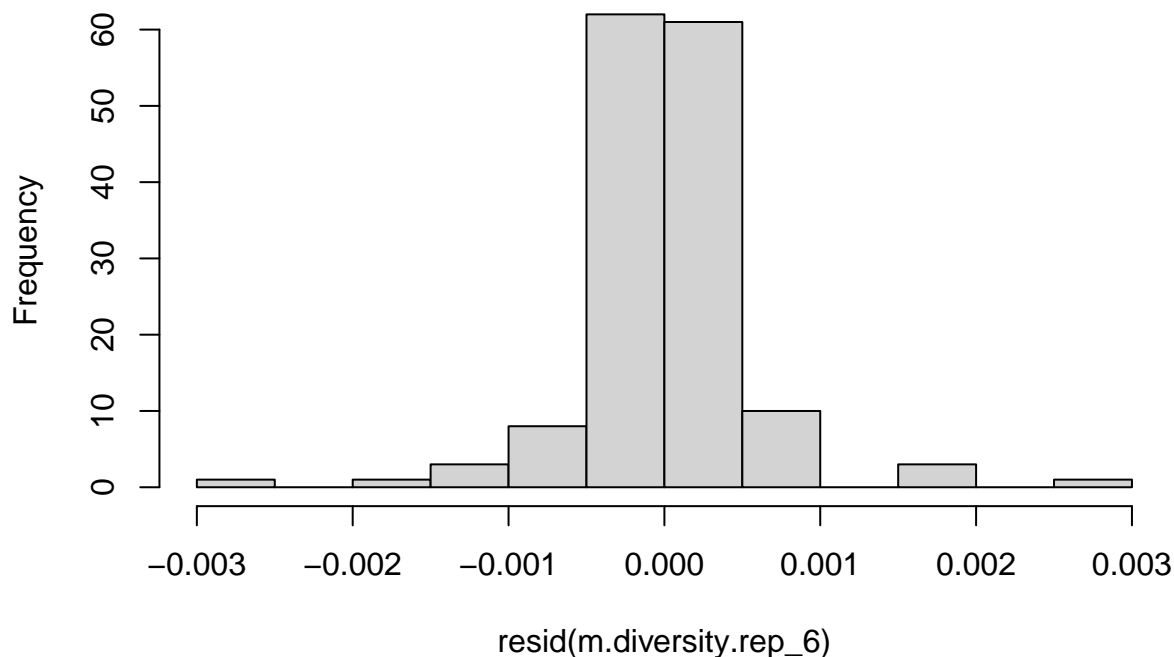
```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_6  
## DW = 1.6828, p-value = 0.01928  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_6)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_6  
## HMC = 0.48734, p-value = 0.437
```

```
hist(resid(m.diversity.rep_6))
```

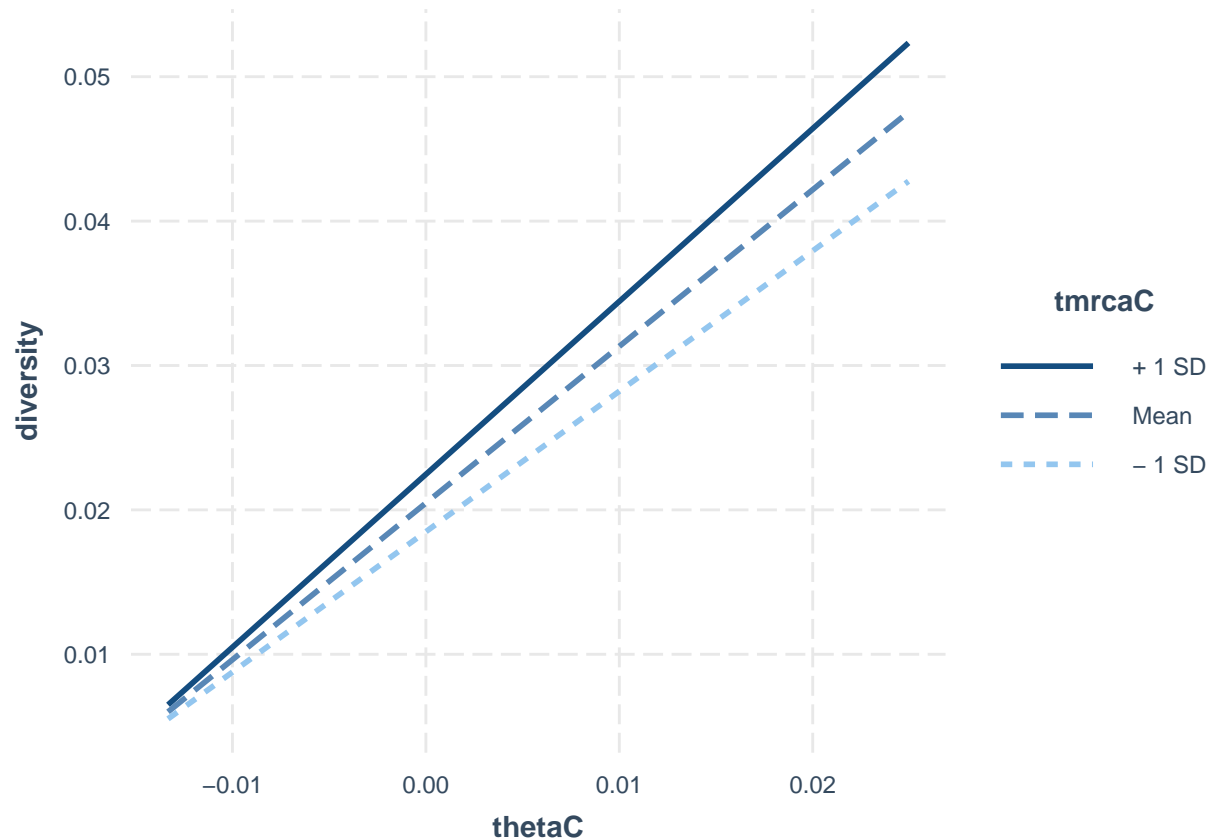
## Histogram of resid(m.diversity.rep\_6)



```
summary(m.diversity.rep_6)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.200kb.rep_6)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.837e-03 -2.547e-04  3.600e-07  2.011e-04  2.684e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.048e-02  4.657e-05  439.670  <2e-16 ***
## thetaC       1.085e+00  5.404e-03  200.771  <2e-16 ***
## rhoC        4.915e-02  6.439e-02   0.763    0.447
## tmrcaC       2.024e-02  5.329e-04  37.991  <2e-16 ***
## thetaC:tmrcaC 1.146e+00  4.835e-02  23.708  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.000568 on 145 degrees of freedom
## Multiple R-squared:  0.9967, Adjusted R-squared:  0.9966
## F-statistic: 1.106e+04 on 4 and 145 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_6, pred = thetaC, modx= tmrcaC)
```



```
g.rep_6 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.200kb.rep_6, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_6)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.200kb.rep_6
##      AIC      BIC    logLik
## -1811.57 -1790.496 912.7851
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.1730188
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0204792 0.00005540 369.6328  0.0000
## thetaC       1.0855830 0.00577130 188.1002  0.0000
## tmrcaC       0.0203230 0.00054805  37.0823  0.0000
## rhoC         0.0441005 0.06334022   0.6962  0.4874
## thetaC:tmrcaC 1.1199378 0.04982885 22.4757  0.0000
```

```
##
## Correlation:
##          (Intr) thetaC tmrcaC rhoC
## thetaC      0.002
## tmrcaC     -0.023 -0.109
## rhoC        -0.008 -0.005  0.340
## thetaC:tmrcaC -0.080 -0.023  0.278  0.116
##
## Standardized residuals:
##          Min          Q1          Med          Q3          Max
## -5.126866241 -0.430648507 -0.002042255  0.325479464  4.863421007
##
## Residual standard error: 0.0005591586
## Degrees of freedom: 150 total; 145 residual

vif(g.rep_6)

##          thetaC          tmrcaC          rhoC thetaC:tmrcaC
##          1.013289          1.224420          1.132487          1.084133

g.rep_6.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.200kb.rep_6, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_6.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.200kb.rep_6
##          AIC          BIC      logLik
## -1450.868 -1435.815 730.4341
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##          Phi
## 0.3062371
##
## Coefficients:
##          Value Std.Error t-value p-value
## (Intercept)  0.0205840 0.00022012 93.51378 0e+00
## thetaC      1.1123598 0.02006909 55.42651 0e+00
## rhoC       -0.7401946 0.19645270 -3.76780 2e-04
##
## Correlation:
##          (Intr) thetaC
## thetaC 0.000
## rhoC   0.002  0.031
##
## Standardized residuals:
##          Min          Q1          Med          Q3          Max
## -3.19333658 -0.46017960  0.02519769  0.43441985  3.88632753
##
## Residual standard error: 0.001950655
## Degrees of freedom: 150 total; 147 residual
```

```

anova.diversity <- Anova(m.diversity.rep_6)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.200kb[1, 6] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.inf.200kb[2, 6] <- anova.diversity$VarExp[1] * 100
r2.inf.200kb[3, 6] <- anova.diversity$VarExp[2] * 100
r2.inf.200kb[4, 6] <- anova.diversity$VarExp[3] * 100
r2.inf.200kb[5, 6] <- anova.diversity$VarExp[4] * 100

```

## 2.2.7 Replicate 7

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_7/rs.pair_7.", sep = "")
pi.200kb <- read.table(paste(p, "diversity.200kb.bedgraph", sep = ""), header = T)
pi.200kb$avg <- apply(pi.200kb[4:ncol(pi.200kb)], 1, mean)
tmrca.200kb <- read.table(paste(p, "TMRCA.200kb.bedgraph", sep = ""), header = T)
tmrca.200kb$avg <- apply(tmrca.200kb[4:ncol(tmrca.200kb)], 1, mean)
rho.200kb <- read.table(paste(p, "rho.200kb.bedgraph", sep = ""), header = T)
theta.200kb <- read.table(paste(p, "theta.200kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.200kb$sim, theta.200kb$sample_mean, method = "spearman")

## Warning in cor.test.default(sim.theta.200kb$sim, theta.200kb$sample_mean, :
## Cannot compute exact p-value with ties

##
## Spearman's rank correlation rho
##
## data: sim.theta.200kb$sim and theta.200kb$sample_mean
## S = 7134.3, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9873163

inf.lands.200kb.rep_7 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_mean,
names(inf.lands.200kb.rep_7) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.200kb.rep_7$thetaC <- inf.lands.200kb.rep_7$theta - mean(inf.lands.200kb.rep_7$theta)
inf.lands.200kb.rep_7$tmrcaC <- inf.lands.200kb.rep_7$tmrca - mean(inf.lands.200kb.rep_7$tmrca)
inf.lands.200kb.rep_7$rhoC <- inf.lands.200kb.rep_7$rho - mean(inf.lands.200kb.rep_7$rho)

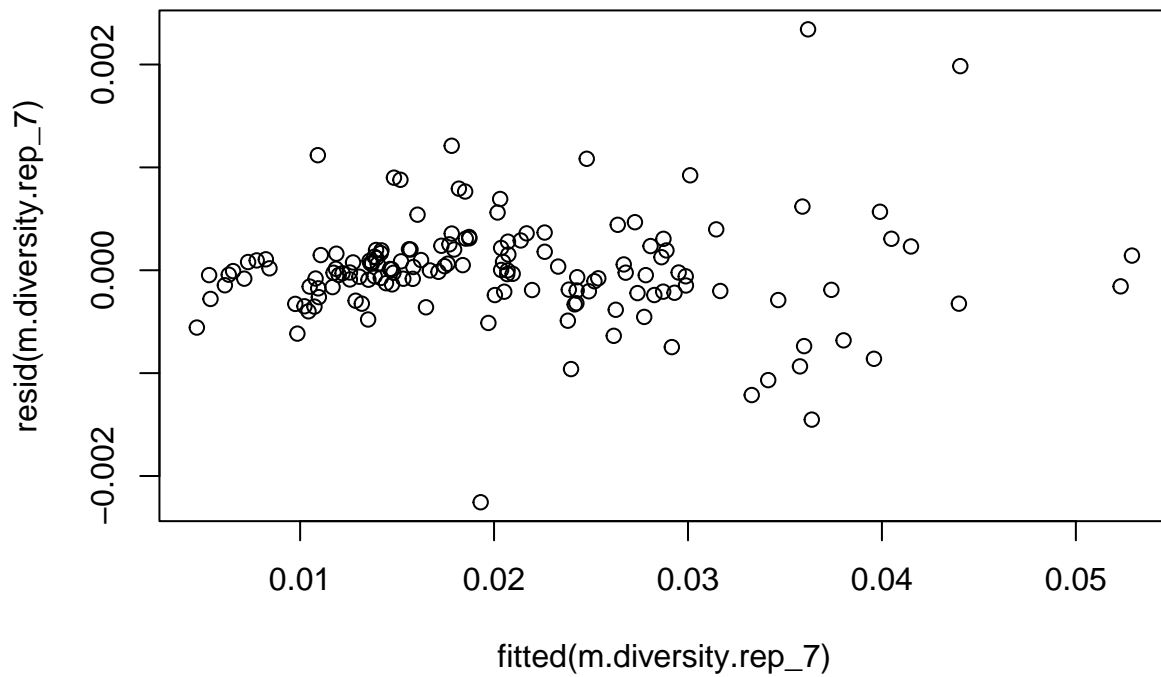
inf.lands.200kb.rep_7$bin <- 1:nrow(inf.lands.200kb.rep_7)

# for merging:
inf.lands.200kb.rep_7$Replicate <- 7

m.diversity.rep_7 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.200kb.rep_7)

plot(resid(m.diversity.rep_7)~fitted(m.diversity.rep_7))

```



```
dwtest(m.diversity.rep_7)
```

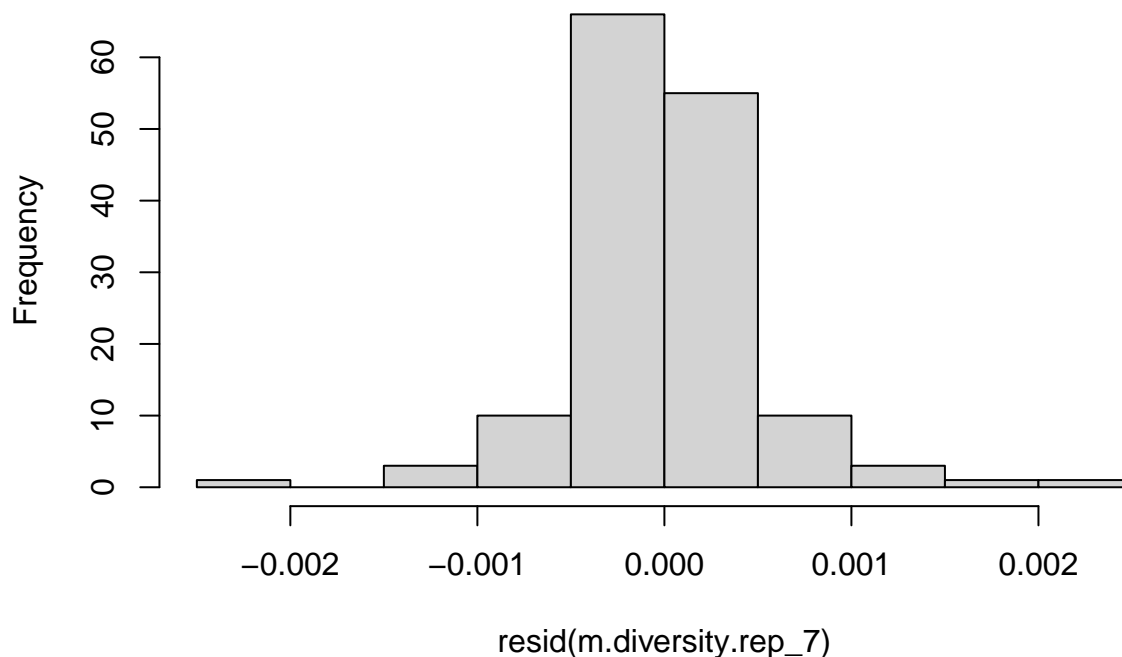
```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_7  
## DW = 1.6673, p-value = 0.01553  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_7)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_7  
## HMC = 0.43605, p-value = 0.131
```

```
hist(resid(m.diversity.rep_7))
```

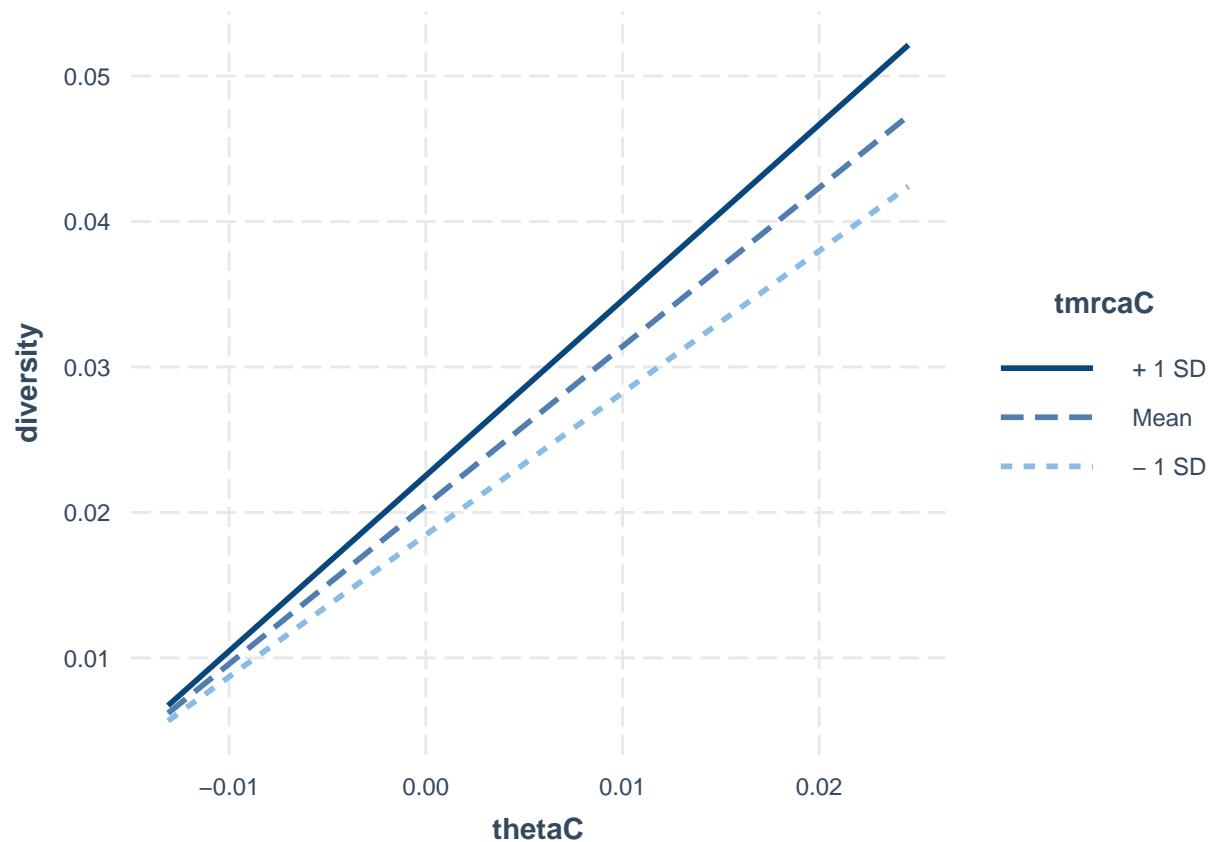
## Histogram of resid(m.diversity.rep\_7)



```
summary(m.diversity.rep_7)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.200kb.rep_7)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.255e-03 -2.038e-04 -2.189e-05  1.931e-04  2.342e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.0204967  0.0000427  480.036  <2e-16 ***
## thetaC        1.0913523  0.0049822  219.050  <2e-16 ***
## rhoC         -0.0194028  0.0611569   -0.317    0.752
## tmrcaC        0.0204993  0.0004700   43.612  <2e-16 ***
## thetaC:tmrcaC  1.1606406  0.0451718   25.694  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005216 on 145 degrees of freedom
## Multiple R-squared:  0.9972, Adjusted R-squared:  0.9971
## F-statistic: 1.281e+04 on 4 and 145 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_7, pred = thetaC, modx= tmrcaC)
```



```
g.rep_7 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.200kb.rep_7, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_7)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.200kb.rep_7
##      AIC      BIC   logLik
## -1837.873 -1816.799  925.9367
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.1935696
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0204997 0.00005196 394.5197  0.0000
## thetaC       1.0931140 0.00535150 204.2630  0.0000
## tmrcaC       0.0203603 0.00047964  42.4489  0.0000
## rhoC        -0.0138547 0.06038474  -0.2294  0.8189
## thetaC:tmrcaC 1.1234814 0.04563517  24.6188  0.0000
```



```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.002
## tmrcaC      -0.021 -0.082
## rhoC        -0.003 -0.022  0.167
## thetaC:tmrcaC -0.060  0.039  0.329  0.078
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.36585180 -0.42077188 -0.01421357  0.38381511  4.57664531
##
## Residual standard error: 0.0005142079
## Degrees of freedom: 150 total; 145 residual

vif(g.rep_7)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.011928      1.158318      1.029312      1.127670

g.rep_7.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.200kb.rep_7, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_7.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.200kb.rep_7
##      AIC      BIC      logLik
## -1444.025 -1428.971  727.0123
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.2825213
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept)  0.0205796 0.00021783 94.47441  0.0000
## thetaC      1.1122060 0.02050319 54.24550  0.0000
## rhoC       -0.4602880 0.22027764 -2.08958  0.0384
##
## Correlation:
##      (Intr) thetaC
## thetaC  0.000
## rhoC    0.002 -0.007
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -3.4561092 -0.4625831  0.1353834  0.5540699  4.0615520
##
## Residual standard error: 0.001980566
## Degrees of freedom: 150 total; 147 residual
```

```

anova.diversity <- Anova(m.diversity.rep_7)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.200kb[1, 7] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.inf.200kb[2, 7] <- anova.diversity$VarExp[1] * 100
r2.inf.200kb[3, 7] <- anova.diversity$VarExp[2] * 100
r2.inf.200kb[4, 7] <- anova.diversity$VarExp[3] * 100
r2.inf.200kb[5, 7] <- anova.diversity$VarExp[4] * 100

```

### 2.2.8 Replicate 8

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_8/rs.pair_8.", sep = "")
pi.200kb <- read.table(paste(p, "diversity.200kb.bedgraph", sep = ""), header = T)
pi.200kb$avg <- apply(pi.200kb[4:ncol(pi.200kb)], 1, mean)
tmrca.200kb <- read.table(paste(p, "TMRCA.200kb.bedgraph", sep = ""), header = T)
tmrca.200kb$avg <- apply(tmrca.200kb[4:ncol(tmrca.200kb)], 1, mean)
rho.200kb <- read.table(paste(p, "rho.200kb.bedgraph", sep = ""), header = T)
theta.200kb <- read.table(paste(p, "theta.200kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.200kb$sim, theta.200kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.200kb$sim and theta.200kb$sample_mean
## S = 6883.3, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9877626

inf.lands.200kb.rep_8 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_mean,
names(inf.lands.200kb.rep_8) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.200kb.rep_8$thetaC <- inf.lands.200kb.rep_8$theta - mean(inf.lands.200kb.rep_8$theta)
inf.lands.200kb.rep_8$tmrcaC <- inf.lands.200kb.rep_8$tmrca - mean(inf.lands.200kb.rep_8$tmrca)
inf.lands.200kb.rep_8$rhoC <- inf.lands.200kb.rep_8$rho - mean(inf.lands.200kb.rep_8$rho)

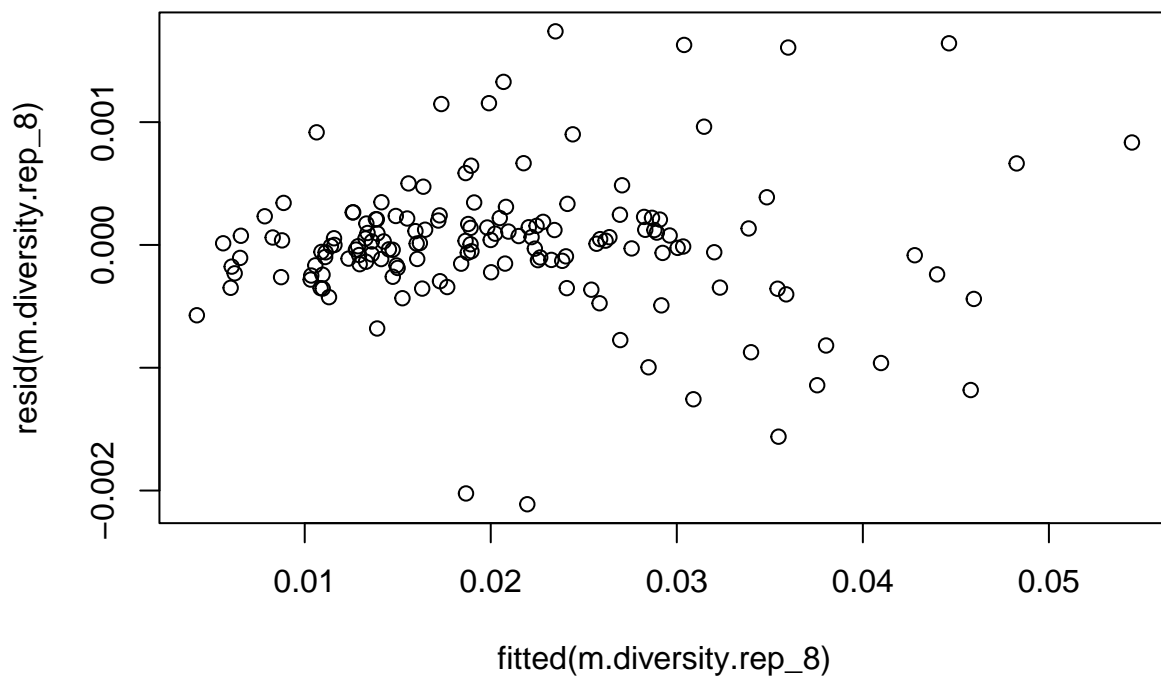
inf.lands.200kb.rep_8$bin <- 1:nrow(inf.lands.200kb.rep_8)

# for merging:
inf.lands.200kb.rep_8$Replicate <- 8

m.diversity.rep_8 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.200kb.rep_8)

plot(resid(m.diversity.rep_8)~fitted(m.diversity.rep_8))

```



```
dwtest(m.diversity.rep_8)
```

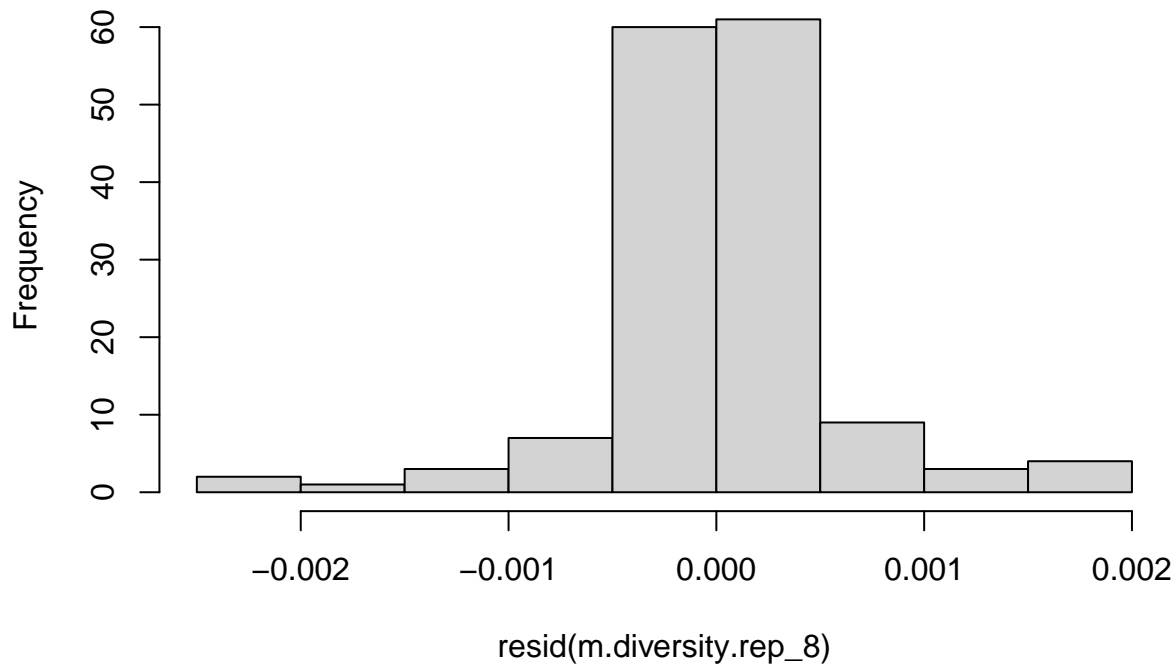
```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_8  
## DW = 1.5909, p-value = 0.004238  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_8)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_8  
## HMC = 0.45552, p-value = 0.213
```

```
hist(resid(m.diversity.rep_8))
```

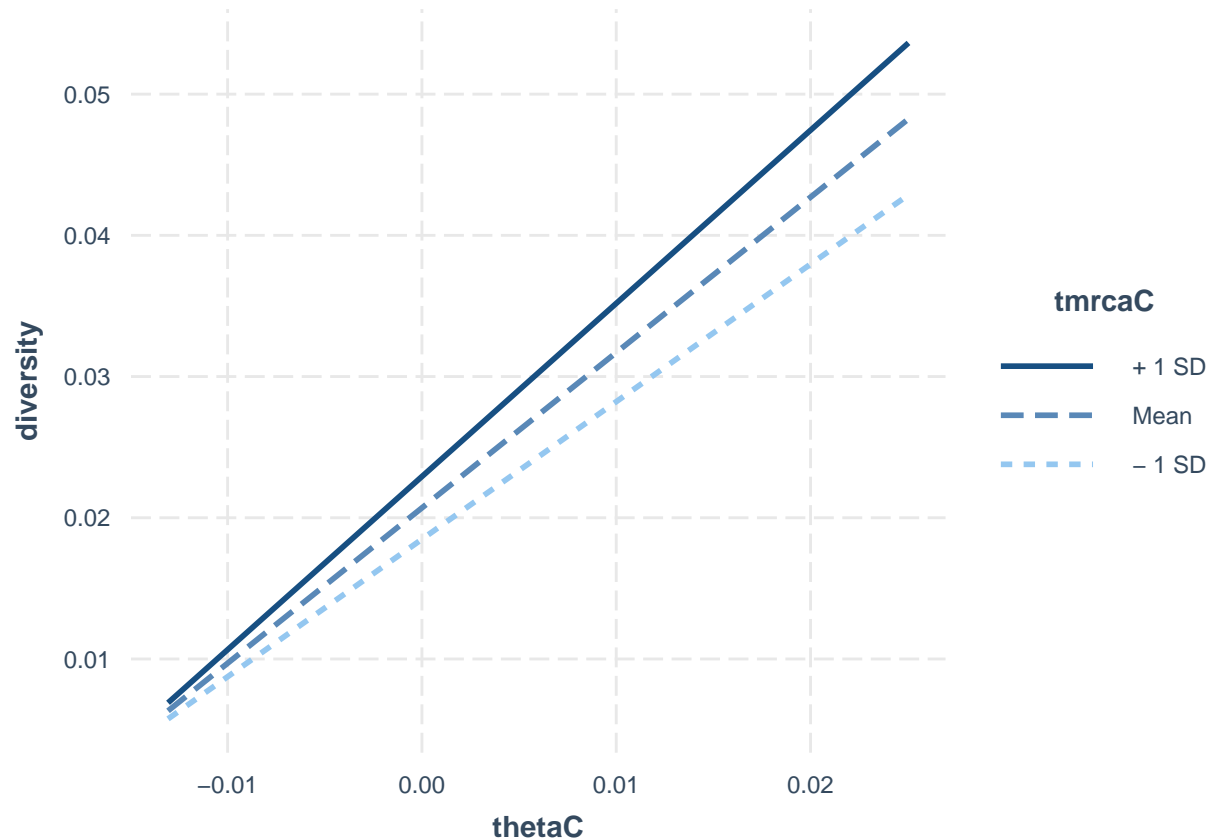
## Histogram of resid(m.diversity.rep\_8)



```
summary(m.diversity.rep_8)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.200kb.rep_8)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.112e-03 -1.846e-04  7.690e-06  1.961e-04  1.739e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.070e-02  4.624e-05  447.65  <2e-16 ***
## thetaC       1.100e+00  5.401e-03  203.69  <2e-16 ***
## rhoC         6.479e-02  6.610e-02   0.98   0.329
## tmrcaC        2.076e-02  4.865e-04  42.68  <2e-16 ***
## thetaC:tmrcaC 1.181e+00  4.604e-02  25.64  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005639 on 145 degrees of freedom
## Multiple R-squared:  0.9968, Adjusted R-squared:  0.9967
## F-statistic: 1.124e+04 on 4 and 145 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_8, pred = thetaC, modx= tmrcaC)
```



```
g.rep_8 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.200kb.rep_8, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_8)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.200kb.rep_8
##      AIC      BIC   logLik
## -1816.857 -1795.783  915.4287
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.2325221
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0207042 0.00005859  353.3963  0.0000
## thetaC       1.0990491 0.00584866  187.9147  0.0000
## tmrcaC       0.0207188 0.00049940   41.4878  0.0000
## rhoC         0.0549165 0.06405597    0.8573  0.3927
## thetaC:tmrcaC 1.1364878 0.04655992   24.4091  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.001
## tmrcaC      -0.026 -0.121
## rhoC        -0.003 -0.003  0.185
## thetaC:tmrcaC -0.074  0.014  0.351  0.043
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -3.78620355 -0.37183043 -0.03601008  0.33263464  3.19188937
##
## Residual standard error: 0.0005563072
## Degrees of freedom: 150 total; 145 residual

vif(g.rep_8)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.019264      1.202188      1.036381      1.145973

g.rep_8.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.200kb.rep_8, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_8.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.200kb.rep_8
##      AIC      BIC    logLik
## -1429.786 -1414.733  719.8932
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.3033588
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept)  0.0208111 0.00023518 88.49129  0.0000
## thetaC      1.1297636 0.02165099 52.18069  0.0000
## rhoC       -0.4075453 0.22773180 -1.78958  0.0756
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.000  0.018
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.25604299 -0.41587461  0.04359114  0.48798012  3.54428535
##
## Residual standard error: 0.002090655
## Degrees of freedom: 150 total; 147 residual
```

```

anova.diversity <- Anova(m.diversity.rep_8)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.200kb[1, 8] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.inf.200kb[2, 8] <- anova.diversity$VarExp[1] * 100
r2.inf.200kb[3, 8] <- anova.diversity$VarExp[2] * 100
r2.inf.200kb[4, 8] <- anova.diversity$VarExp[3] * 100
r2.inf.200kb[5, 8] <- anova.diversity$VarExp[4] * 100

```

### 2.2.9 Replicate 9

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_9/rs.pair_9.", sep = "")
pi.200kb <- read.table(paste(p, "diversity.200kb.bedgraph", sep = ""), header = T)
pi.200kb$avg <- apply(pi.200kb[4:ncol(pi.200kb)], 1, mean)
tmrca.200kb <- read.table(paste(p, "TMRCA.200kb.bedgraph", sep = ""), header = T)
tmrca.200kb$avg <- apply(tmrca.200kb[4:ncol(tmrca.200kb)], 1, mean)
rho.200kb <- read.table(paste(p, "rho.200kb.bedgraph", sep = ""), header = T)
theta.200kb <- read.table(paste(p, "theta.200kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.200kb$sim, theta.200kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.200kb$sim and theta.200kb$sample_mean
## S = 7233.3, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9871403

inf.lands.200kb.rep_9 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_mean,
names(inf.lands.200kb.rep_9) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.200kb.rep_9$thetaC <- inf.lands.200kb.rep_9$theta - mean(inf.lands.200kb.rep_9$theta)
inf.lands.200kb.rep_9$tmrcaC <- inf.lands.200kb.rep_9$tmrca - mean(inf.lands.200kb.rep_9$tmrca)
inf.lands.200kb.rep_9$rhoC <- inf.lands.200kb.rep_9$rho - mean(inf.lands.200kb.rep_9$rho)

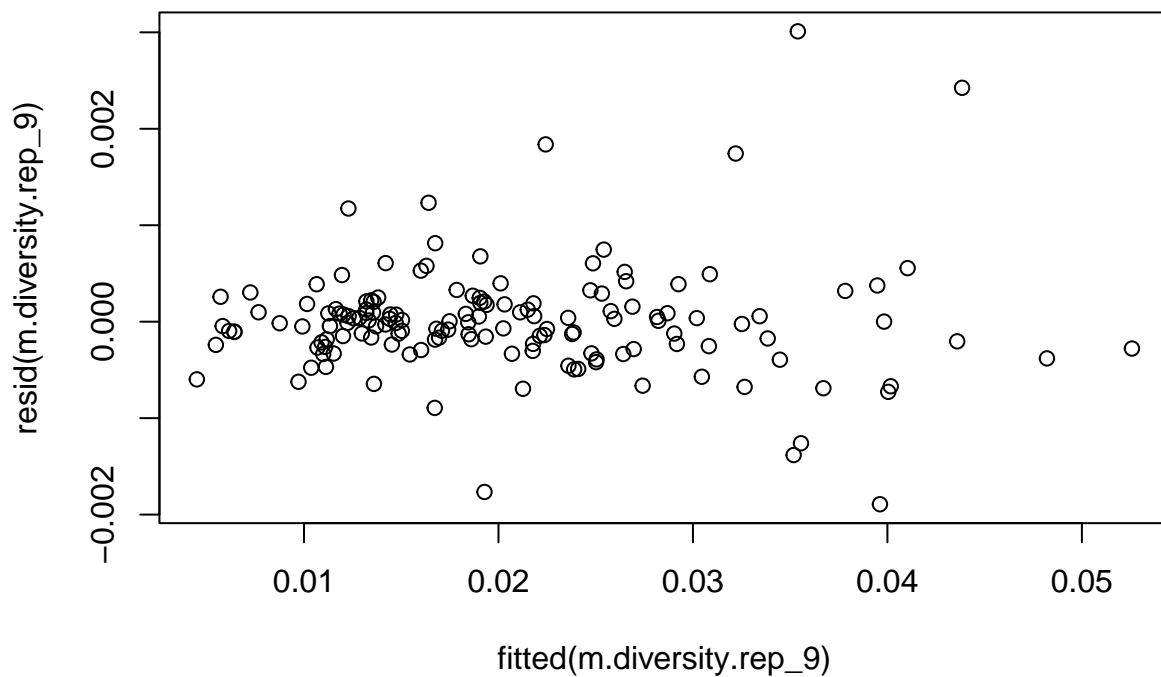
inf.lands.200kb.rep_9$bin <- 1:nrow(inf.lands.200kb.rep_9)

# for merging:
inf.lands.200kb.rep_9$Replicate <- 9

m.diversity.rep_9 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.200kb.rep_9)

plot(resid(m.diversity.rep_9)~fitted(m.diversity.rep_9))

```



```
dwtest(m.diversity.rep_9)
```

```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_9  
## DW = 1.6772, p-value = 0.01787  
## alternative hypothesis: true autocorrelation is greater than 0
```

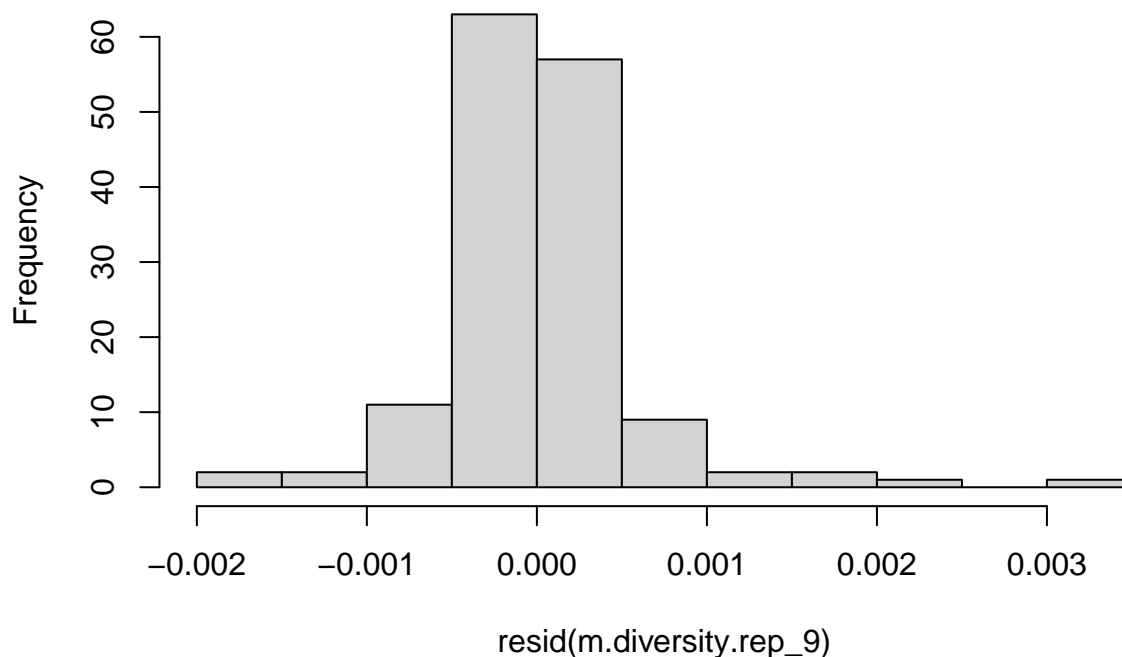
```
hmcetest(m.diversity.rep_9)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_9  
## HMC = 0.38369, p-value = 0.019
```

```
hist(resid(m.diversity.rep_9))
```



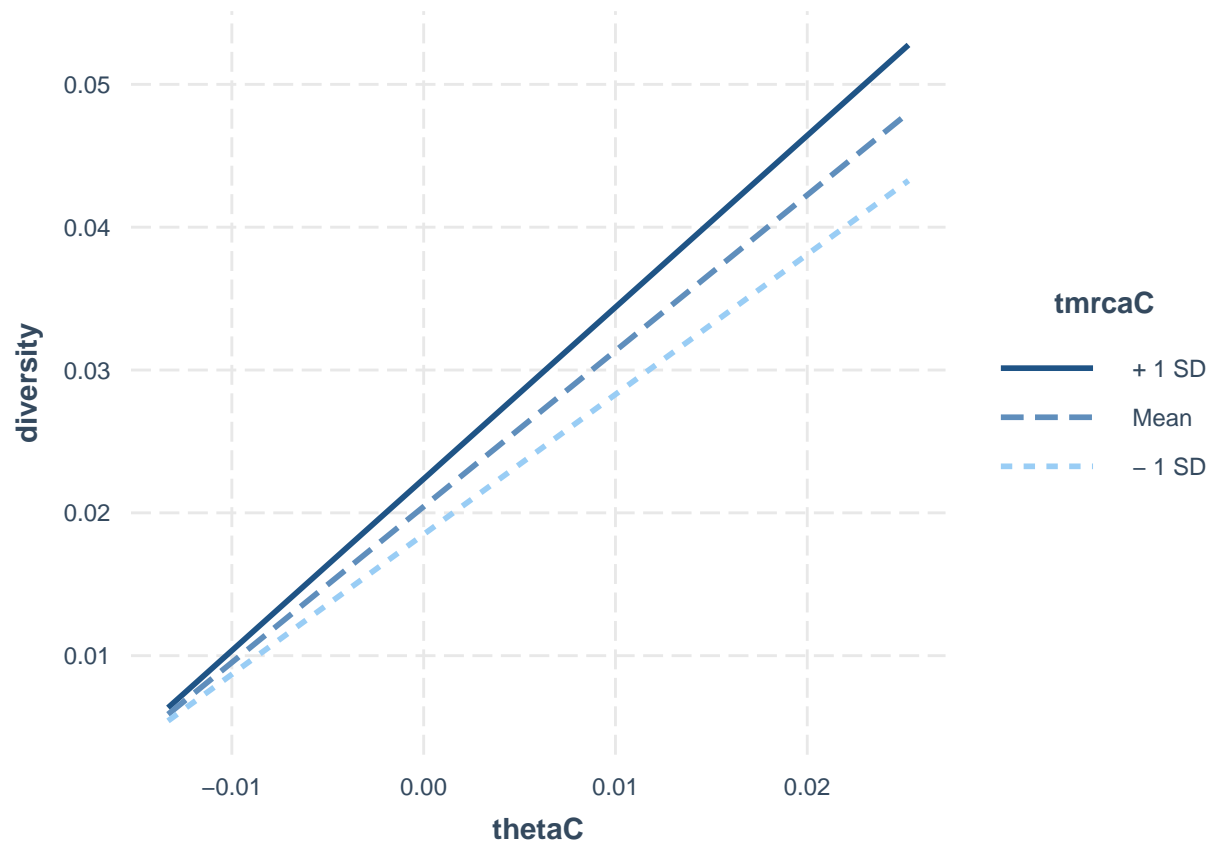
## Histogram of resid(m.diversity.rep\_9)



```
summary(m.diversity.rep_9)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.200kb.rep_9)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0018926 -0.0002394 -0.0000149  0.0001830  0.0030102
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.044e-02  4.788e-05  426.832  <2e-16 ***
## thetaC       1.091e+00  5.584e-03  195.371  <2e-16 ***
## rhoC        1.218e-01  7.409e-02   1.643    0.102
## tmrcaC       2.096e-02  5.758e-04   36.402  <2e-16 ***
## thetaC:tmrcaC 1.200e+00  5.571e-02   21.530  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.000584 on 145 degrees of freedom
## Multiple R-squared:  0.9965, Adjusted R-squared:  0.9964
## F-statistic: 1.02e+04 on 4 and 145 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_9, pred = thetaC, modx= tmrcaC)
```



```
g.rep_9 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.200kb.rep_9, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_9)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.200kb.rep_9
##      AIC      BIC   logLik
## -1804.101 -1783.026 909.0504
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.2058788
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0204394 0.00005911 345.7760  0.0000
## thetaC       1.0918715 0.00599966 181.9888  0.0000
## tmrcaC       0.0210780 0.00059395 35.4877  0.0000
## rhoC        0.1568335 0.07270121  2.1572  0.0326
## thetaC:tmrcaC 1.1467223 0.05773633 19.8614  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.006
## tmrcaC      -0.029 -0.067
## rhoC         -0.008  0.011  0.270
## thetaC:tmrcaC -0.076  0.077  0.351  0.110
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -3.22090386 -0.40412235 -0.03092443  0.30494856  5.19848696
##
## Residual standard error: 0.000576945
## Degrees of freedom: 150 total; 145 residual

vif(g.rep_9)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.017102      1.228533      1.079801      1.153832

g.rep_9.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.200kb.rep_9, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_9.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.200kb.rep_9
##      AIC      BIC logLik
## -1464.018 -1448.965 737.009
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.3448494
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept)  0.0205379 0.00022295 92.11977 0.0000
## thetaC      1.1066746 0.01936981 57.13401 0.0000
## rhoC       -0.4865608 0.21452090 -2.26813 0.0248
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.002  0.028
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.01675167 -0.35192121 -0.05519082  0.39502917  4.12501538
##
## Residual standard error: 0.00189327
## Degrees of freedom: 150 total; 147 residual
```

```

anova.diversity <- Anova(m.diversity.rep_9)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.200kb[1, 9] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]
r2.inf.200kb[2, 9] <- anova.diversity$VarExp[1] * 100
r2.inf.200kb[3, 9] <- anova.diversity$VarExp[2] * 100
r2.inf.200kb[4, 9] <- anova.diversity$VarExp[3] * 100
r2.inf.200kb[5, 9] <- anova.diversity$VarExp[4] * 100

```

## 2.2.10 Replicate 10

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_10/rs.pair_10.", sep = "")
pi.200kb <- read.table(paste(p, "diversity.200kb.bedgraph", sep = ""), header = T)
pi.200kb$avg <- apply(pi.200kb[4:ncol(pi.200kb)], 1, mean)
tmrca.200kb <- read.table(paste(p, "TMRCA.200kb.bedgraph", sep = ""), header = T)
tmrca.200kb$avg <- apply(tmrca.200kb[4:ncol(tmrca.200kb)], 1, mean)
rho.200kb <- read.table(paste(p, "rho.200kb.bedgraph", sep = ""), header = T)
theta.200kb <- read.table(paste(p, "theta.200kb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.200kb$sim, theta.200kb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.200kb$sim and theta.200kb$sample_mean
## S = 6266.2, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9888595

inf.lands.200kb.rep_10 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_m
names(inf.lands.200kb.rep_10) <- c("diversity", "theta", "rho", "tmrca")

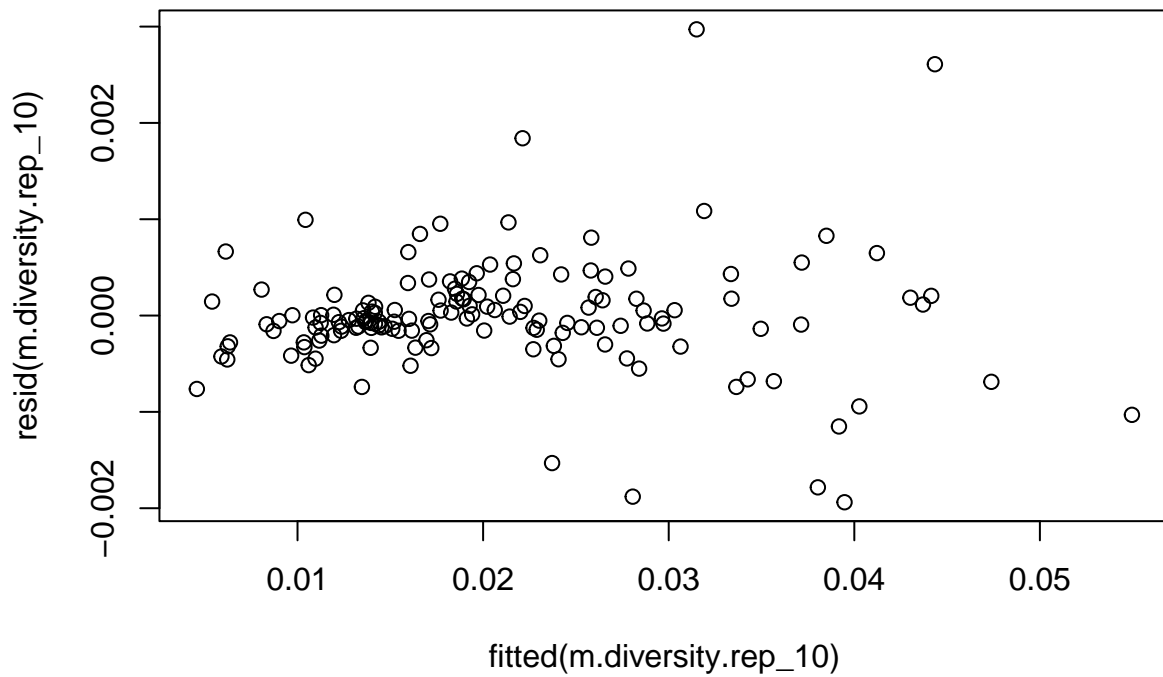
# centering
inf.lands.200kb.rep_10$thetaC <- inf.lands.200kb.rep_10$theta - mean(inf.lands.200kb.rep_10$theta)
inf.lands.200kb.rep_10$tmrcaC <- inf.lands.200kb.rep_10$tmrca - mean(inf.lands.200kb.rep_10$tmrca)
inf.lands.200kb.rep_10$rhoC <- inf.lands.200kb.rep_10$rho - mean(inf.lands.200kb.rep_10$rho)

inf.lands.200kb.rep_10$bin <- 1:nrow(inf.lands.200kb.rep_10)

# for merging:
inf.lands.200kb.rep_10$Replicate <- 10

m.diversity.rep_10 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.200kb.rep_10)
plot(resid(m.diversity.rep_10)~fitted(m.diversity.rep_10))

```



```
dwtest(m.diversity.rep_10)
```

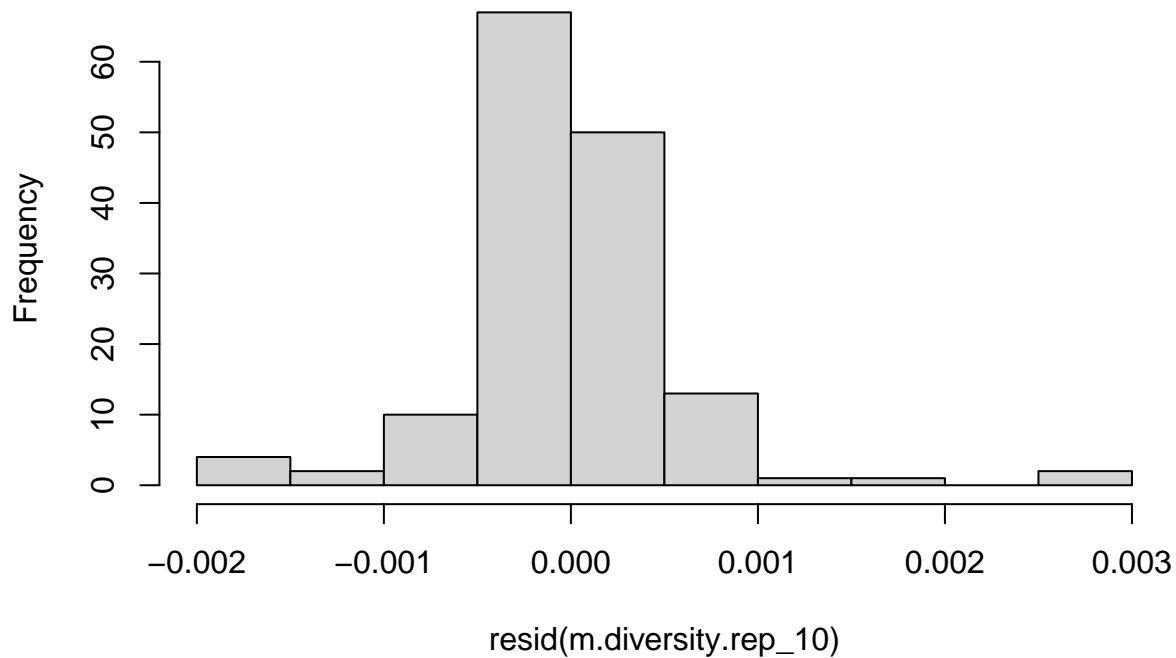
```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_10  
## DW = 1.777, p-value = 0.06912  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcctest(m.diversity.rep_10)
```

```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_10  
## HMC = 0.6193, p-value = 0.983
```

```
hist(resid(m.diversity.rep_10))
```

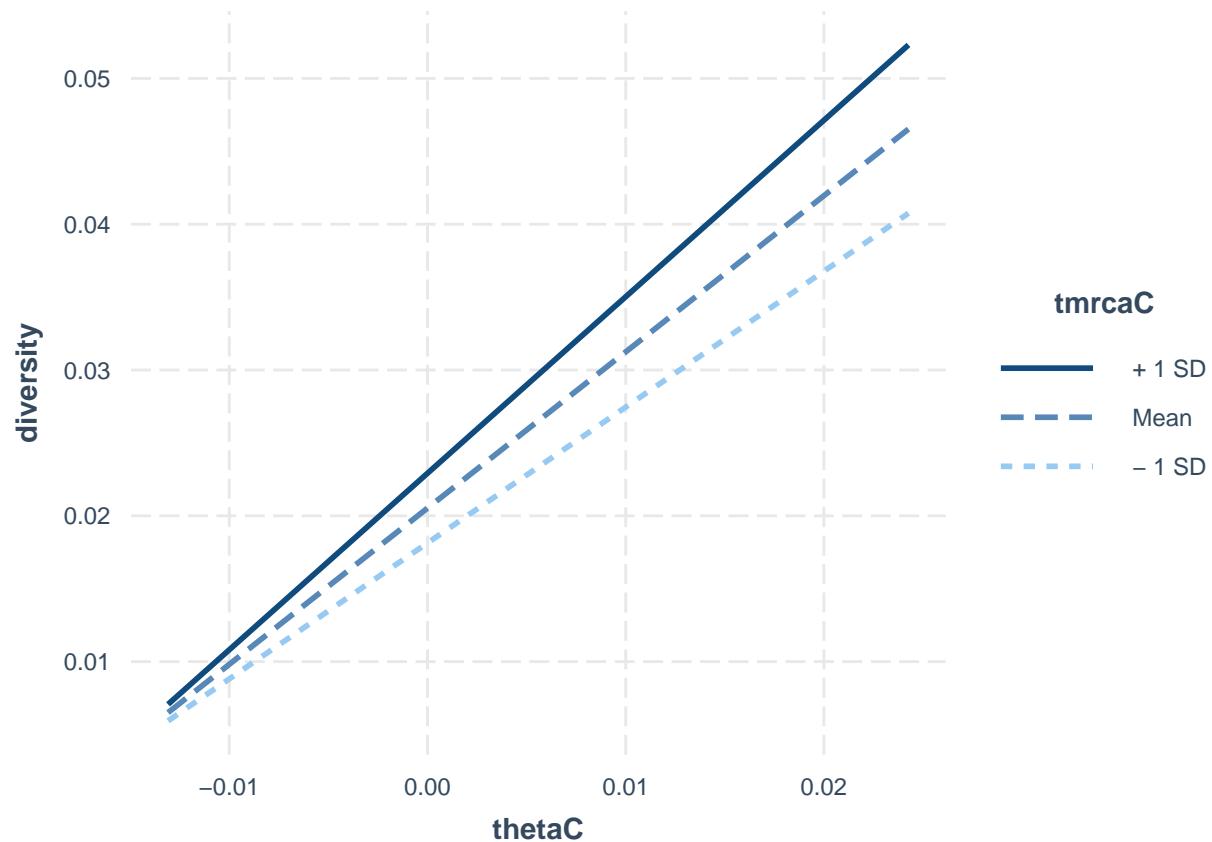
## Histogram of resid(m.diversity.rep\_10)



```
summary(m.diversity.rep_10)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.200kb.rep_10)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.938e-03 -1.748e-04 -4.971e-05  1.828e-04  2.971e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.053e-02  5.021e-05  408.838  <2e-16 ***
## thetaC       1.071e+00  5.803e-03  184.625  <2e-16 ***
## rhoC        1.995e-02  7.054e-02   0.283    0.778
## tmrcaC       2.204e-02  5.689e-04  38.738  <2e-16 ***
## thetaC:tmrcaC 1.283e+00  5.870e-02  21.866  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0006067 on 145 degrees of freedom
## Multiple R-squared:  0.9963, Adjusted R-squared:  0.9962
## F-statistic: 9827 on 4 and 145 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_10, pred = thetaC, modx= tmrcaC)
```



```
g.rep_10 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
  data = inf.lands.200kb.rep_1, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_10)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.200kb.rep_1
##      AIC      BIC   logLik
## -1814.128 -1793.054  914.0641
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.08562895
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0204958 0.00004989  410.8349  0.0000
## thetaC       1.0967285 0.00562895  194.8371  0.0000
## tmrcaC       0.0210125 0.00050621   41.5092  0.0000
## rhoC         0.0219683 0.06340360    0.3465  0.7295
## thetaC:tmrcaC 1.2110957 0.05322577   22.7539  0.0000
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.001
## tmrcaC      -0.043 -0.110
## rhoC        -0.010 -0.009  0.248
## thetaC:tmrcaC -0.113  0.010  0.375  0.090
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.94983846 -0.42035716 -0.01013693  0.33115763  4.40016170
##
## Residual standard error: 0.0005481049
## Degrees of freedom: 150 total; 145 residual

vif(g.rep_10)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.015853      1.249427      1.066075      1.167458

g.rep_10.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.200kb.rep_10, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_10.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.200kb.rep_10
##      AIC      BIC      logLik
## -1442.284 -1427.23 726.1418
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.3304014
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept) 0.0207131 0.00023459 88.29634 0.0000
## thetaC      1.1352063 0.02089711 54.32360 0.0000
## rhoC       -0.5008572 0.21144100 -2.36878 0.0191
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.002 -0.030
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.75464196 -0.53844834  0.02521541  0.52547583  4.31476850
##
## Residual standard error: 0.002024422
## Degrees of freedom: 150 total; 147 residual
```



```

anova.diversity <- Anova(m.diversity.rep_10)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.200kb[1, 10] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) * 100
r2.inf.200kb[2, 10] <- anova.diversity$VarExp[1] * 100
r2.inf.200kb[3, 10] <- anova.diversity$VarExp[2] * 100
r2.inf.200kb[4, 10] <- anova.diversity$VarExp[3] * 100
r2.inf.200kb[5, 10] <- anova.diversity$VarExp[4] * 100

r2.inf.200kb$average <- rowMeans(r2.inf.200kb)
r2.inf.200kb <- transform(r2.inf.200kb, sd=apply(r2.inf.200kb, 1, sd, na.rm = TRUE))

```

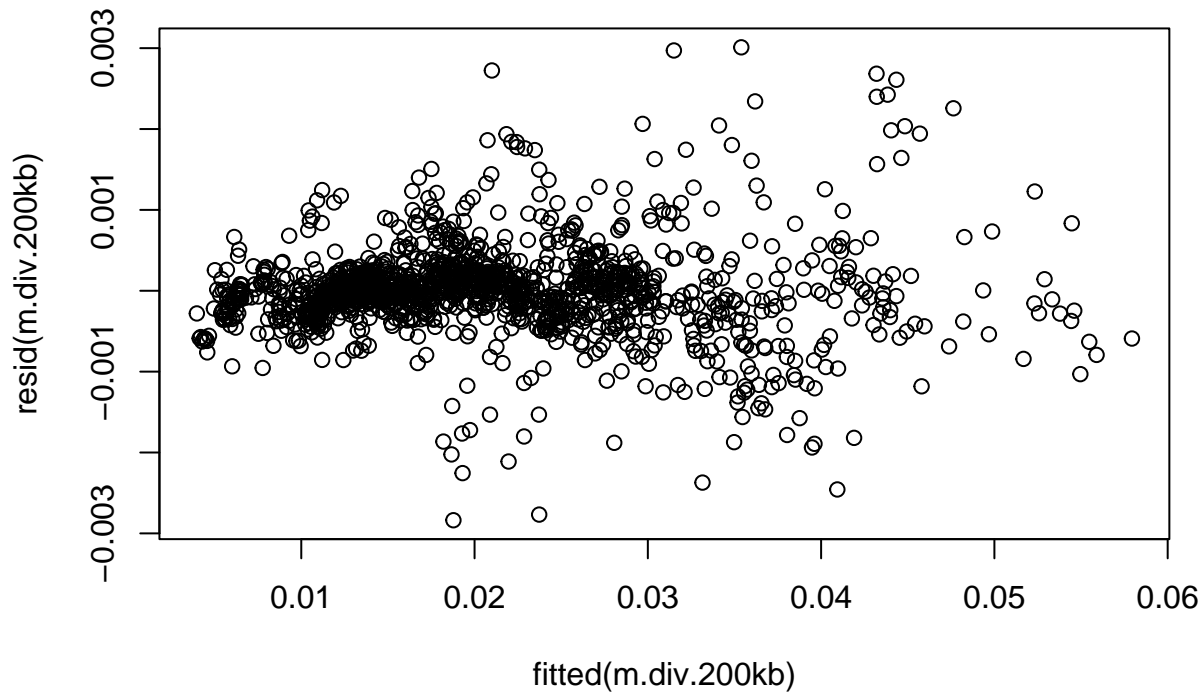
### 2.2.11 all replicates:

```

inf.lands.200kb.all <- rbind(inf.lands.200kb.rep_1, inf.lands.200kb.rep_2, inf.lands.200kb.rep_3, inf.lands.200kb.rep_4, inf.lands.200kb.rep_5, inf.lands.200kb.rep_6, inf.lands.200kb.rep_7, inf.lands.200kb.rep_8, inf.lands.200kb.rep_9, inf.lands.200kb.rep_10)

m.div.200kb <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)*as.factor(Replicate), data = inf.lands.200kb.all)
plot(resid(m.div.200kb)~fitted(m.div.200kb))

```



```

dwtest(m.div.200kb)

##
## Durbin-Watson test
##
## data: m.div.200kb

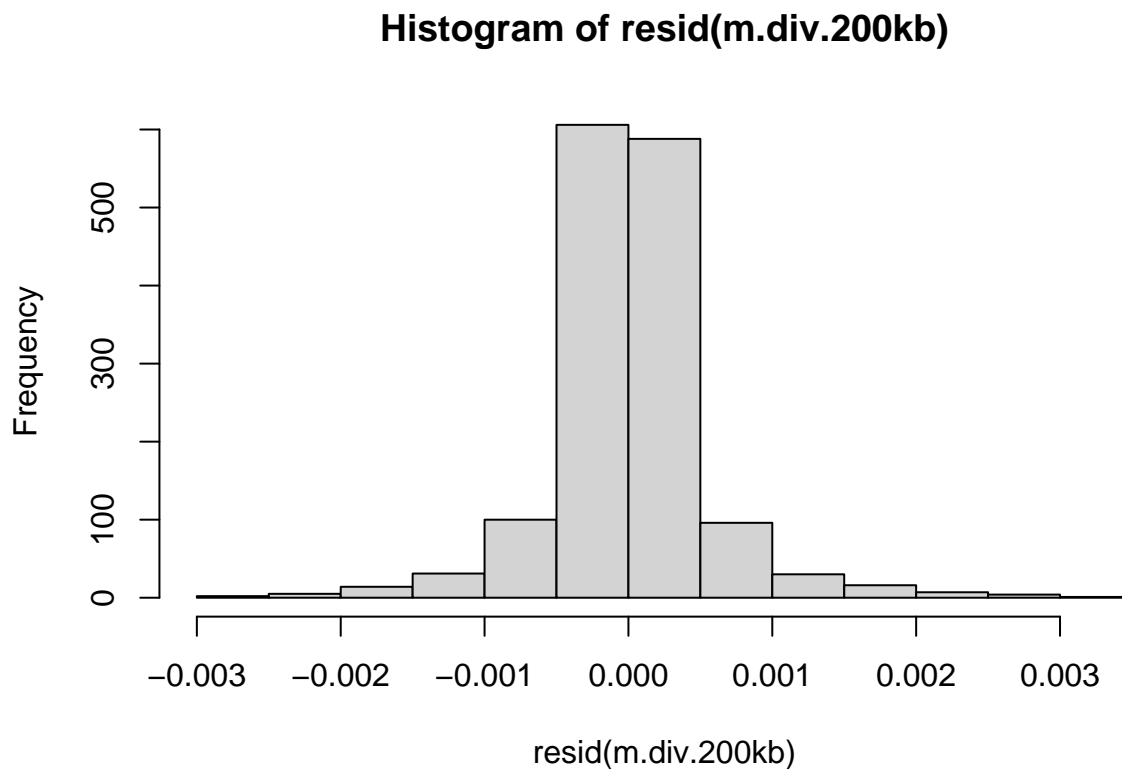
```

```
## DW = 1.6825, p-value = 8.31e-12
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.200kb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.48172, p-value = 0.158
```

```
hist(resid(m.div.200kb))
```



```
m.div.200kb.2 <- lm(versity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC)*as.factor(Replicate)
m.div.200kb.3 <- lm(versity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC + rhoC:tmrcaC)*as
```

```
AIC(m.div.200kb, m.div.200kb.2, m.div.200kb.3)
```

```
##           df      AIC
## m.div.200kb  51 -18158.03
## m.div.200kb.2 61 -18155.65
## m.div.200kb.3 71 -18154.31
```

```
summary(m.div.200kb)
```

```
##
## Call:
## lm(formula = versity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC) *
##     as.factor(Replicate), data = inf.lands.200kb.all)
```

```

##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.837e-03 -2.188e-04 -6.580e-06  1.836e-04  3.010e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.049e-02  4.600e-05  445.509 < 2e-16 ***
## thetaC           1.096e+00  5.439e-03  201.508 < 2e-16 ***
## rhoC             1.426e-02  6.373e-02   0.224 0.822950
## tmrcaC           2.103e-02  5.029e-04  41.809 < 2e-16 ***
## as.factor(Replicate)2 -3.060e-05  6.507e-05  -0.470 0.638292
## as.factor(Replicate)3  1.025e-04  6.504e-05   1.576 0.115287
## as.factor(Replicate)4  1.385e-04  6.486e-05   2.136 0.032875 *
## as.factor(Replicate)5  1.180e-05  6.487e-05   0.182 0.855637
## as.factor(Replicate)6 -1.818e-05  6.496e-05  -0.280 0.779594
## as.factor(Replicate)7  2.154e-06  6.491e-05   0.033 0.973535
## as.factor(Replicate)8  2.053e-04  6.496e-05   3.160 0.001611 **
## as.factor(Replicate)9 -5.912e-05  6.496e-05  -0.910 0.362922
## as.factor(Replicate)10 3.230e-05  6.526e-05   0.495 0.620766
## thetaC:tmrcaC       1.224e+00  5.205e-02  23.507 < 2e-16 ***
## thetaC:as.factor(Replicate)2 -1.053e-02  7.664e-03  -1.374 0.169687
## thetaC:as.factor(Replicate)3 -1.650e-02  7.590e-03  -2.174 0.029886 *
## thetaC:as.factor(Replicate)4  2.909e-02  7.659e-03   3.799 0.000151 ***
## thetaC:as.factor(Replicate)5  1.351e-02  7.668e-03   1.762 0.078268 .
## thetaC:as.factor(Replicate)6 -1.096e-02  7.610e-03  -1.441 0.149866
## thetaC:as.factor(Replicate)7 -4.595e-03  7.624e-03  -0.603 0.546846
## thetaC:as.factor(Replicate)8  4.142e-03  7.634e-03   0.543 0.587471
## thetaC:as.factor(Replicate)9 -4.972e-03  7.628e-03  -0.652 0.514662
## thetaC:as.factor(Replicate)10 -2.460e-02  7.629e-03  -3.225 0.001289 **
## rhoC:as.factor(Replicate)2 -5.012e-03  9.221e-02  -0.054 0.956658
## rhoC:as.factor(Replicate)3  2.185e-02  9.191e-02   0.238 0.812089
## rhoC:as.factor(Replicate)4  4.984e-02  9.296e-02   0.536 0.591924
## rhoC:as.factor(Replicate)5  7.204e-02  9.411e-02   0.766 0.444095
## rhoC:as.factor(Replicate)6  3.489e-02  8.991e-02   0.388 0.698042
## rhoC:as.factor(Replicate)7 -3.367e-02  9.146e-02  -0.368 0.712840
## rhoC:as.factor(Replicate)8  5.053e-02  9.144e-02   0.553 0.580613
## rhoC:as.factor(Replicate)9  1.075e-01  9.538e-02   1.127 0.259957
## rhoC:as.factor(Replicate)10 5.683e-03  9.106e-02   0.062 0.950247
## tmrcaC:as.factor(Replicate)2 -5.309e-04  7.291e-04  -0.728 0.466627
## tmrcaC:as.factor(Replicate)3 -4.744e-04  6.808e-04  -0.697 0.486026
## tmrcaC:as.factor(Replicate)4 -1.276e-03  6.615e-04  -1.930 0.053851 .
## tmrcaC:as.factor(Replicate)5 -5.725e-04  6.812e-04  -0.840 0.400788
## tmrcaC:as.factor(Replicate)6 -7.813e-04  7.269e-04  -1.075 0.282629
## tmrcaC:as.factor(Replicate)7 -5.267e-04  7.121e-04  -0.740 0.459573
## tmrcaC:as.factor(Replicate)8 -2.644e-04  6.970e-04  -0.379 0.704532
## tmrcaC:as.factor(Replicate)9 -6.463e-05  7.464e-04  -0.087 0.931011
## tmrcaC:as.factor(Replicate)10 1.011e-03  7.267e-04   1.391 0.164432
## thetaC:tmrcaC:as.factor(Replicate)2 -6.224e-02  7.262e-02  -0.857 0.391526
## thetaC:tmrcaC:as.factor(Replicate)3 -1.311e-01  6.631e-02  -1.978 0.048172 *
## thetaC:tmrcaC:as.factor(Replicate)4 -7.738e-02  7.053e-02  -1.097 0.272721
## thetaC:tmrcaC:as.factor(Replicate)5 -2.416e-02  6.752e-02  -0.358 0.720504
## thetaC:tmrcaC:as.factor(Replicate)6 -7.740e-02  7.054e-02  -1.097 0.272752
## thetaC:tmrcaC:as.factor(Replicate)7 -6.291e-02  7.111e-02  -0.885 0.376449

```

```
## thetaC:tmrcaC:as.factor(Replicate)8 -4.279e-02 6.925e-02 -0.618 0.536691
## thetaC:tmrcaC:as.factor(Replicate)9 -2.404e-02 7.454e-02 -0.322 0.747141
## thetaC:tmrcaC:as.factor(Replicate)10 5.990e-02 7.509e-02 0.798 0.425121
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005594 on 1450 degrees of freedom
## Multiple R-squared: 0.9968, Adjusted R-squared: 0.9967
## F-statistic: 9292 on 49 and 1450 DF, p-value: < 2.2e-16
```

## 2.2.12 Plots

```
scale.3d <- function(x) sprintf("%.3f", x)

rho.plot <- as.data.frame(cbind(inf.lands.200kb.rep_1$bin,
                                sim.rho.200kb$sim,
                                inf.lands.200kb.rep_1$rho,
                                inf.lands.200kb.rep_2$rho,
                                inf.lands.200kb.rep_3$rho,
                                inf.lands.200kb.rep_4$rho,
                                inf.lands.200kb.rep_5$rho,
                                inf.lands.200kb.rep_6$rho,
                                inf.lands.200kb.rep_7$rho,
                                inf.lands.200kb.rep_8$rho,
                                inf.lands.200kb.rep_9$rho,
                                inf.lands.200kb.rep_10$rho))

rho.plot[,3:ncol(rho.plot)] <- 1.53 * rho.plot[,3:ncol(rho.plot)] # scaling because of SCRM NO vs Ne

names(rho.plot) <- c("bin", "sim", reps)
molten.rho <- melt(rho.plot, id.vars = "bin")
rho.map.200kb <- ggplot(data = molten.rho, aes(x = bin * 200, y = value, colour = variable)) + theme_bw()
rho.map.200kb <- rho.map.200kb + geom_line(data = molten.rho, aes(size = variable)) + scale_size_manual()
rho.map.200kb <- rho.map.200kb + scale_color_manual(values = c("black", brewer.pal(n = 9, name = "Blues")))
rho.map.200kb <- rho.map.200kb + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
rho.map.200kb <- rho.map.200kb + labs(title = NULL, x = NULL, y = expression(rho))
rho.map.200kb <- rho.map.200kb + theme(text = element_text(size = 20), axis.title.x = element_text(size = 20))

theta.plot <- as.data.frame(cbind(inf.lands.200kb.rep_1$bin,
                                sim.theta.200kb$sim,
                                inf.lands.200kb.rep_1$theta,
                                inf.lands.200kb.rep_2$theta,
                                inf.lands.200kb.rep_3$theta,
                                inf.lands.200kb.rep_4$theta,
                                inf.lands.200kb.rep_5$theta,
                                inf.lands.200kb.rep_6$theta,
                                inf.lands.200kb.rep_7$theta,
                                inf.lands.200kb.rep_8$theta,
                                inf.lands.200kb.rep_9$theta,
                                inf.lands.200kb.rep_10$theta))

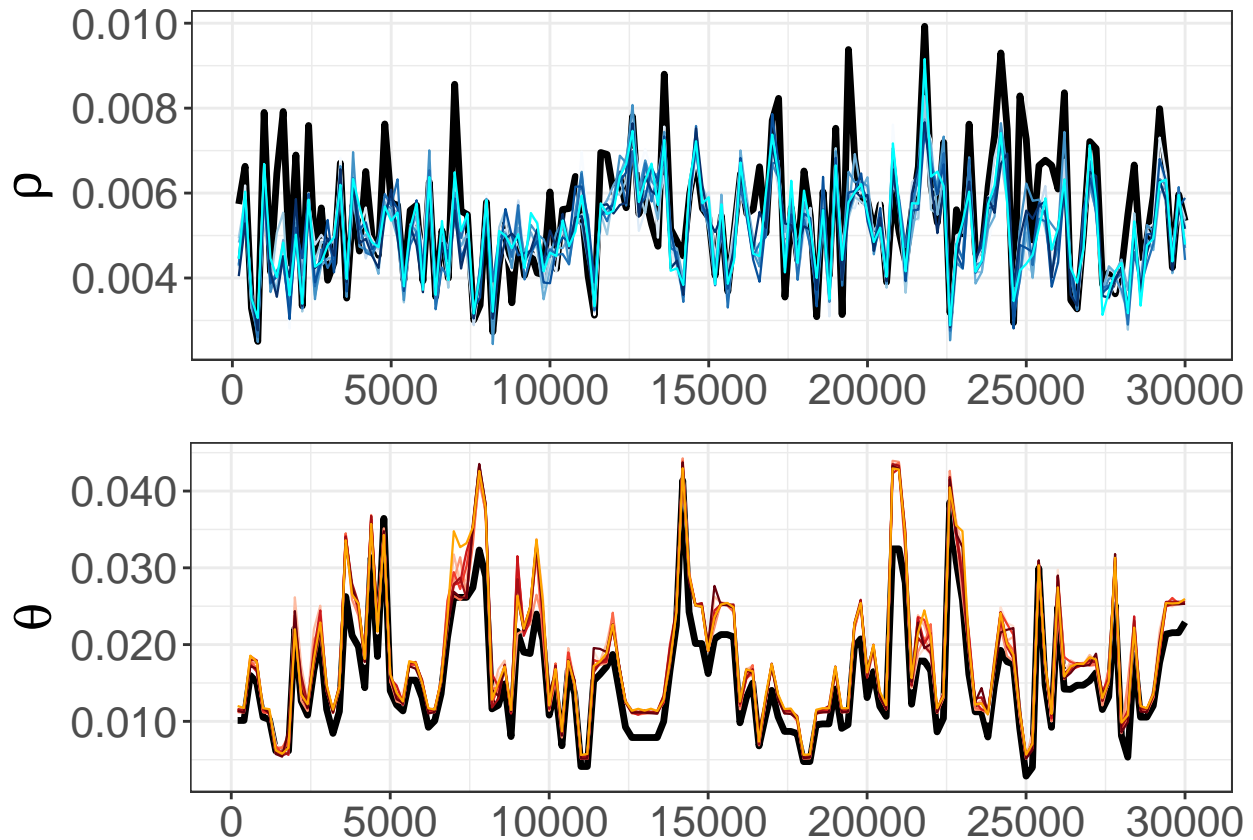
names(theta.plot) <- c("bin", "sim", reps)
molten.theta <- melt(theta.plot, id.vars = "bin")
theta.map.200kb <- ggplot(data = molten.theta, aes(x = bin * 200, y = value, colour = variable)) + theme_bw()
```

```

theta.map.200kb <- theta.map.200kb + geom_line(data = molten.theta, aes(size = variable)) + scale_size_
theta.map.200kb <- theta.map.200kb + scale_color_manual(values = c("black", brewer.pal(n = 9, name = "R
theta.map.200kb <- theta.map.200kb + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(
theta.map.200kb <- theta.map.200kb + labs(title = NULL, x = NULL, y = expression(theta))
theta.map.200kb <- theta.map.200kb + theme(text = element_text(size = 20), axis.title.x = element_text(

plot_grid(rho.map.200kb, theta.map.200kb, nrow = 2, ncol = 1)

```



## 2.3 1 Mb scale

```

r2.inf.1Mb <- as.data.frame(matrix(ncol = nreps, nrow = 5))
row.names(r2.inf.1Mb) <- c("Total", "Theta", "Rho", "TMRCA", "Theta:TMRCA")
colnames(r2.inf.1Mb) <- reps

# sim landscapes
sim.rho.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims/dm.sim.rho.1e+06.txt")
sim.theta.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims/dm.sim.theta.1e+06.txt")
sim.lands.1Mb <- as.data.frame(cbind(sim.theta.1Mb$sim, sim.rho.1Mb$sim))
names(sim.lands.1Mb) <- c("theta", "rho")

```

### 2.3.1 Replicate 1

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_1/rs.pair_1.", sep = "")
pi.1Mb <- read.table(paste(p, "diversity.1Mb.bedgraph", sep = ""), header = T)
pi.1Mb$avg <- apply(pi.1Mb[4:ncol(pi.1Mb)], 1, mean)

```

```

tmrca.1Mb <- read.table(paste(p, "TMRCA.1Mb.bedgraph", sep = ""), header = T)
tmrca.1Mb$avg <- apply(tmrca.1Mb[4:ncol(tmrca.1Mb)], 1, mean)
rho.1Mb <- read.table(paste(p, "rho.1Mb.bedgraph", sep = ""), header = T)
theta.1Mb <- read.table(paste(p, "theta.1Mb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.1Mb$sim, theta.1Mb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.1Mb$sim and theta.1Mb$sample_mean
## S = 94, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9790879

inf.lands.1Mb.rep_1 <- as.data.frame(cbind(pi.1Mb$avg, theta.1Mb$sample_mean, rho.1Mb$sample_mean, tmrca.1Mb$avg),
names(inf.lands.1Mb.rep_1) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.1Mb.rep_1$thetaC <- inf.lands.1Mb.rep_1$theta - mean(inf.lands.1Mb.rep_1$theta)
inf.lands.1Mb.rep_1$tmrcaC <- inf.lands.1Mb.rep_1$tmrca - mean(inf.lands.1Mb.rep_1$tmrca)
inf.lands.1Mb.rep_1$rhoC <- inf.lands.1Mb.rep_1$rho - mean(inf.lands.1Mb.rep_1$rho)

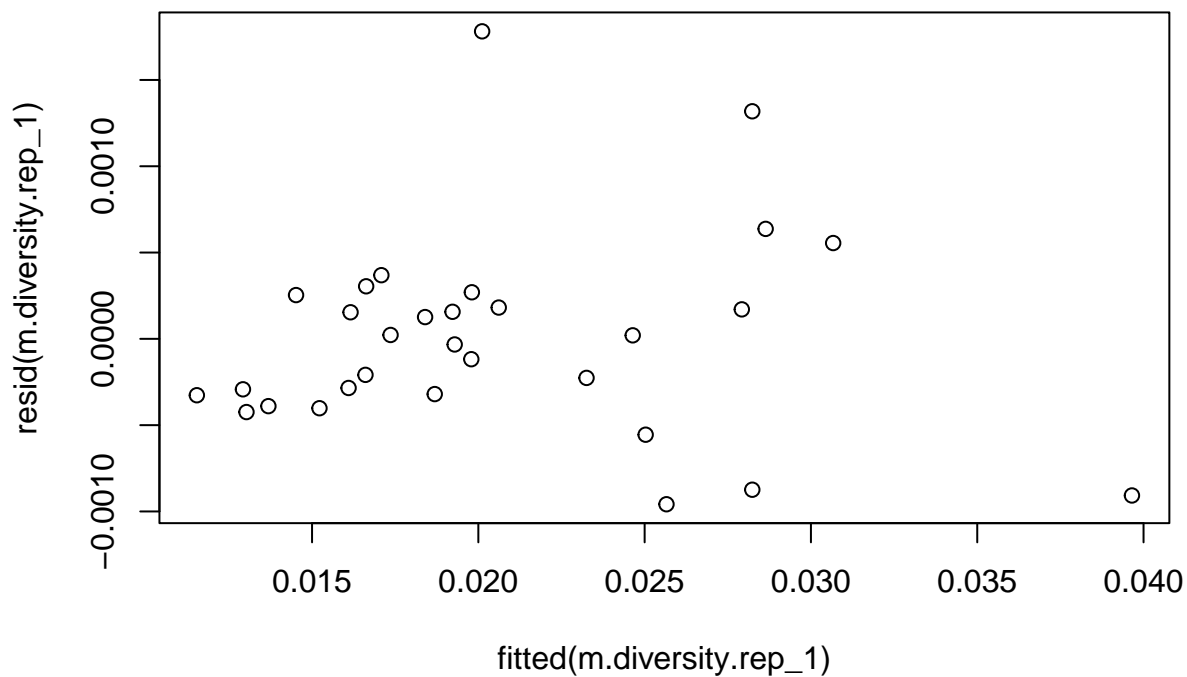
inf.lands.1Mb.rep_1$bin <- 1:nrow(inf.lands.1Mb.rep_1)

# for merging:
inf.lands.1Mb.rep_1$Replicate <- 1

m.diversity.rep_1 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.1Mb.rep_1)

plot(resid(m.diversity.rep_1)~fitted(m.diversity.rep_1))

```



```
dwtest(m.diversity.rep_1)
```

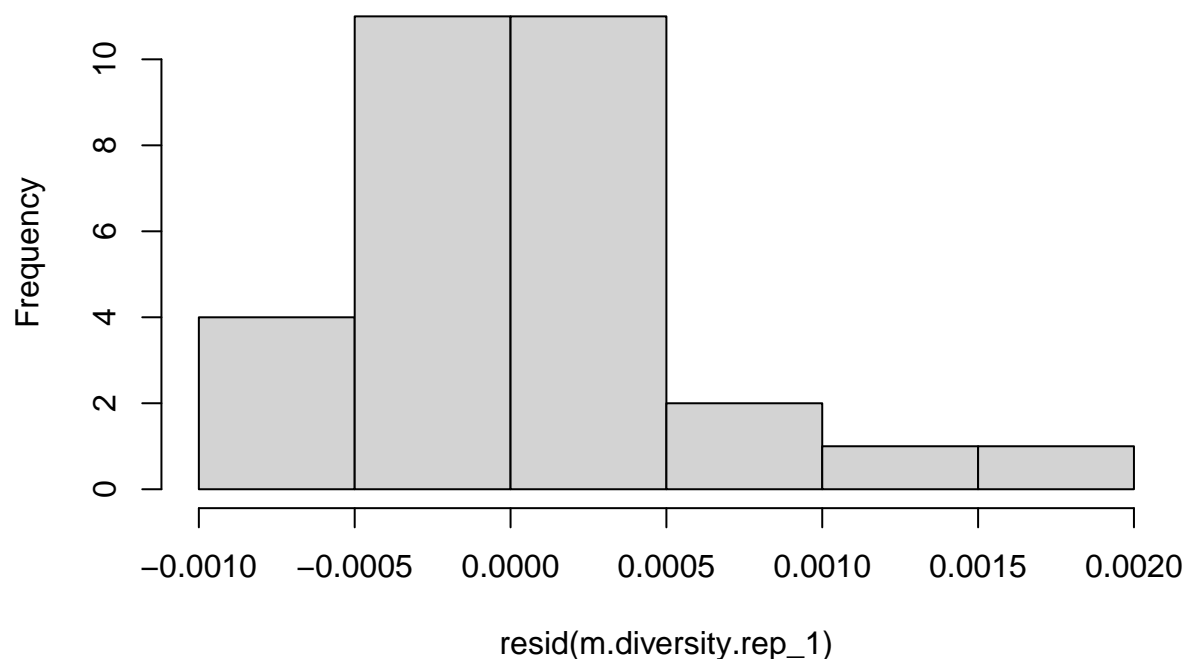
```
##
## Durbin-Watson test
##
## data: m.diversity.rep_1
## DW = 1.9667, p-value = 0.4252
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.diversity.rep_1)
```

```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_1
## HMC = 0.46606, p-value = 0.41
```

```
hist(resid(m.diversity.rep_1))
```

## Histogram of resid(m.diversity.rep\_1)

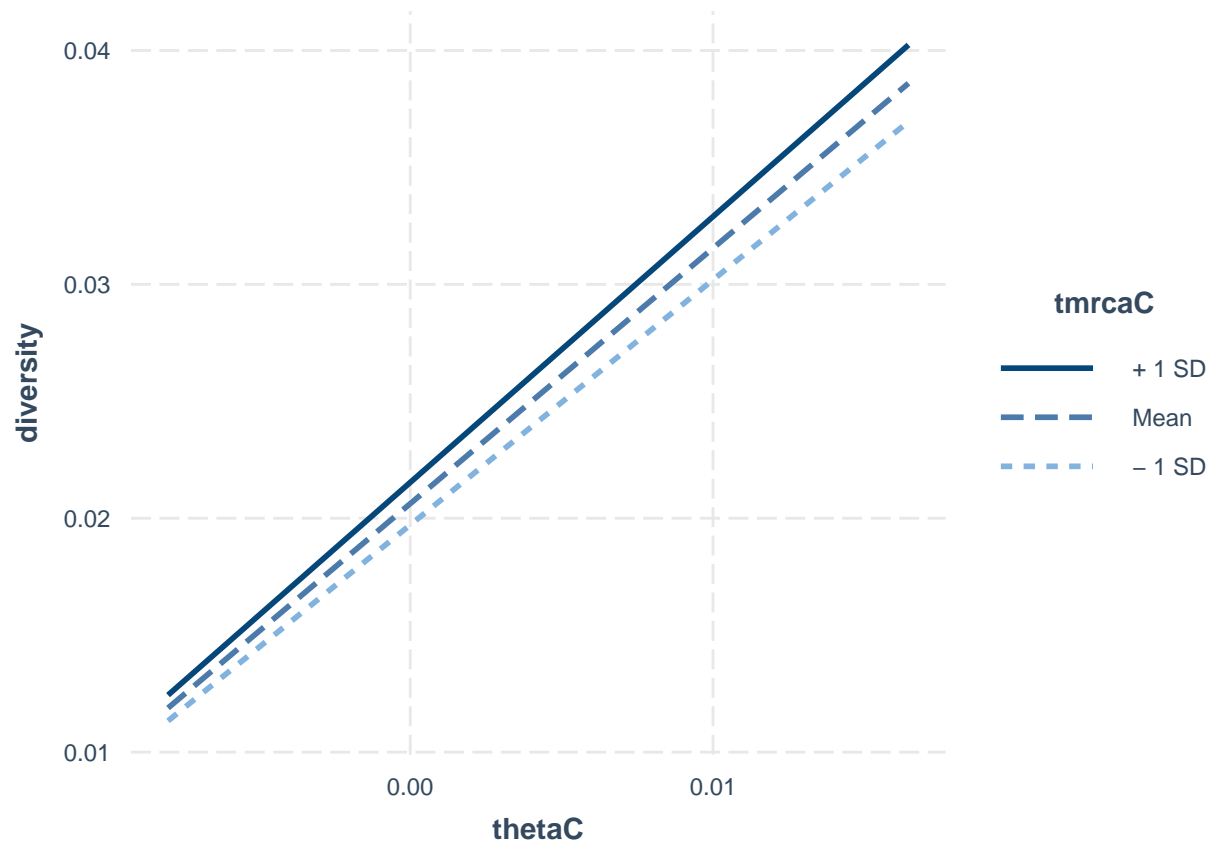


```
summary(m.diversity.rep_1)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.1Mb.rep_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0009583 -0.0003251 -0.0000061  0.0002355  0.0017816
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.020625   0.000115  179.401  < 2e-16 ***
## thetaC       1.092123   0.021125   51.697  < 2e-16 ***
## rhoC         0.003099   0.351525    0.009   0.993
## tmrcaC       0.016704   0.002985    5.596 8.05e-06 ***
## thetaC:tmrcaC 0.819285   0.498125    1.645   0.113
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0006297 on 25 degrees of freedom
## Multiple R-squared:  0.9917, Adjusted R-squared:  0.9904
## F-statistic: 749.1 on 4 and 25 DF,  p-value: < 2.2e-16
```



```
interact_plot(m.diversity.rep_1, pred = thetaC, modx= tmrcaC)
```



```
g.rep_1 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.1Mb.rep_1, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_1)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.1Mb.rep_1
##      AIC      BIC logLik
## -348.56 -338.7516 181.28
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.0188108
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  0.0206248 0.0001171 176.16584  0.0000
## thetaC       1.0920063 0.0211898  51.53453  0.0000
## tmrcaC       0.0168279 0.0029670   5.67164  0.0000
## rhoC        0.0007886 0.3523958   0.00224  0.9982
## thetaC:tmrcaC 0.8389718 0.4973907   1.68675  0.1041
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.001
## tmrcaC       0.001 -0.171
## rhoC         0.000 -0.021  0.376
## thetaC:tmrcaC 0.004 -0.283  0.575 -0.003
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.67095844 -0.56781078 -0.01108315  0.40839933  3.09944720
##
## Residual standard error: 0.0005748314
## Degrees of freedom: 30 total; 25 residual

vif(g.rep_1)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.087501      1.898249      1.271394      1.717045

g.rep_1.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.1Mb.rep_1, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_1.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.1Mb.rep_1
##      AIC      BIC logLik
## -329.2361 -322.2301 169.618
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## -0.3222673
##
## Coefficients:
##      Value Std.Error t-value p-value
## (Intercept)  0.0206090 0.0001242 165.95945  0.0000
## thetaC      1.0935838 0.0263685  41.47310  0.0000
## rhoC       -0.9523384 0.3871228  -2.46004  0.0206
##
## Correlation:
##      (Intr) thetaC
## thetaC  0.002
## rhoC   -0.001 -0.056
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.3284963 -0.5352289  0.1792210  0.6290505  1.9096098
##
## Residual standard error: 0.0008939388
## Degrees of freedom: 30 total; 27 residual
```

```

anova.diversity <- Anova(m.diversity.rep_1)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.1Mb[1, 1] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.1Mb[2, 1] <- anova.diversity$VarExp[1] * 100
r2.inf.1Mb[3, 1] <- anova.diversity$VarExp[2] * 100
r2.inf.1Mb[4, 1] <- anova.diversity$VarExp[3] * 100
r2.inf.1Mb[5, 1] <- anova.diversity$VarExp[4] * 100

```

### 2.3.2 Replicate 2

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_2/rs.pair_2.", sep = "")
pi.1Mb <- read.table(paste(p, "diversity.1Mb.bedgraph", sep = ""), header = T)
pi.1Mb$avg <- apply(pi.1Mb[4:ncol(pi.1Mb)], 1, mean)
tmrca.1Mb <- read.table(paste(p, "TMRCA.1Mb.bedgraph", sep = ""), header = T)
tmrca.1Mb$avg <- apply(tmrca.1Mb[4:ncol(tmrca.1Mb)], 1, mean)
rho.1Mb <- read.table(paste(p, "rho.1Mb.bedgraph", sep = ""), header = T)
theta.1Mb <- read.table(paste(p, "theta.1Mb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.1Mb$sim, theta.1Mb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.1Mb$sim and theta.1Mb$sample_mean
## S = 114, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9746385

inf.lands.1Mb.rep_2 <- as.data.frame(cbind(pi.1Mb$avg, theta.1Mb$sample_mean, rho.1Mb$sample_mean, tmrca.1Mb$avg))
names(inf.lands.1Mb.rep_2) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.1Mb.rep_2$thetaC <- inf.lands.1Mb.rep_2$theta - mean(inf.lands.1Mb.rep_2$theta)
inf.lands.1Mb.rep_2$tmrcaC <- inf.lands.1Mb.rep_2$tmrca - mean(inf.lands.1Mb.rep_2$tmrca)
inf.lands.1Mb.rep_2$rhoC <- inf.lands.1Mb.rep_2$rho - mean(inf.lands.1Mb.rep_2$rho)

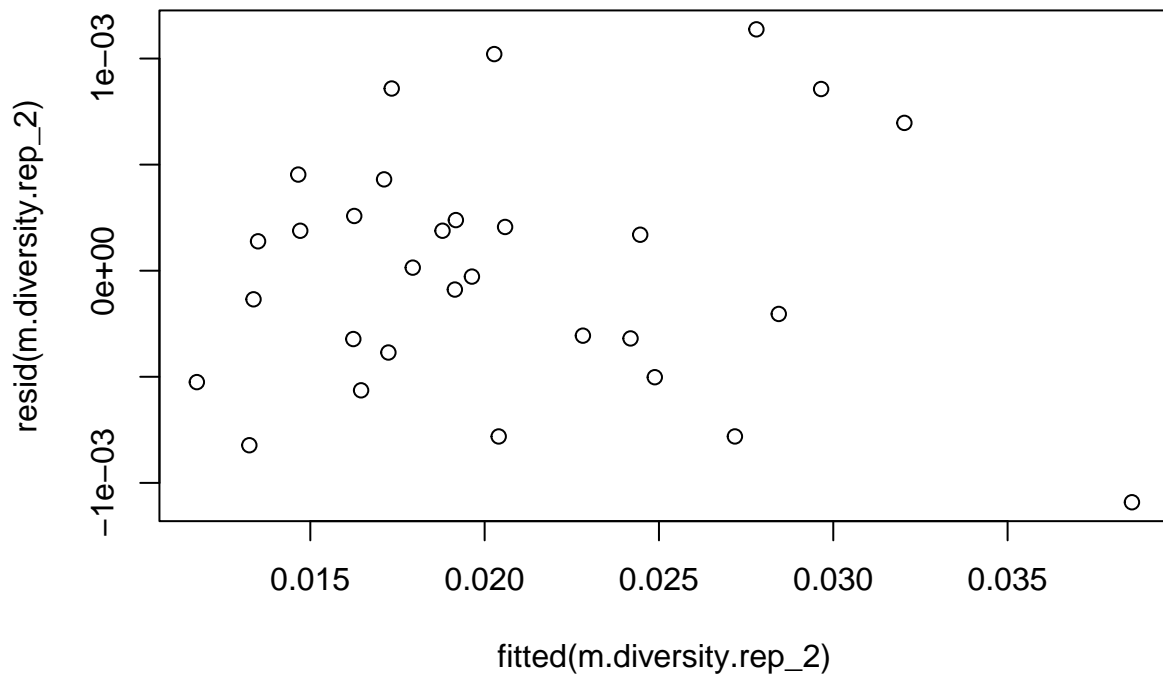
inf.lands.1Mb.rep_2$bin <- 1:nrow(inf.lands.1Mb.rep_2)

# for merging:
inf.lands.1Mb.rep_2$Replicate <- 2

m.diversity.rep_2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.1Mb.rep_2)

plot(resid(m.diversity.rep_2)~fitted(m.diversity.rep_2))

```



```
dwtest(m.diversity.rep_2)
```

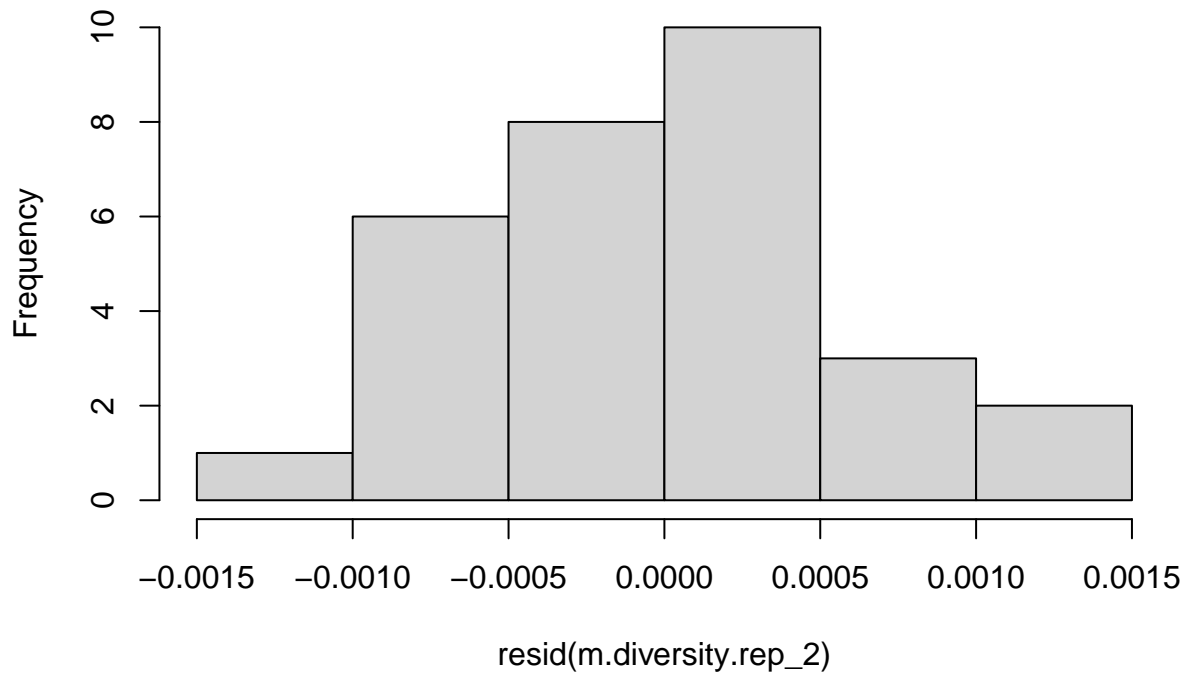
```
##
## Durbin-Watson test
##
## data: m.diversity.rep_2
## DW = 1.966, p-value = 0.4212
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.diversity.rep_2)
```

```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_2
## HMC = 0.4866, p-value = 0.45
```

```
hist(resid(m.diversity.rep_2))
```

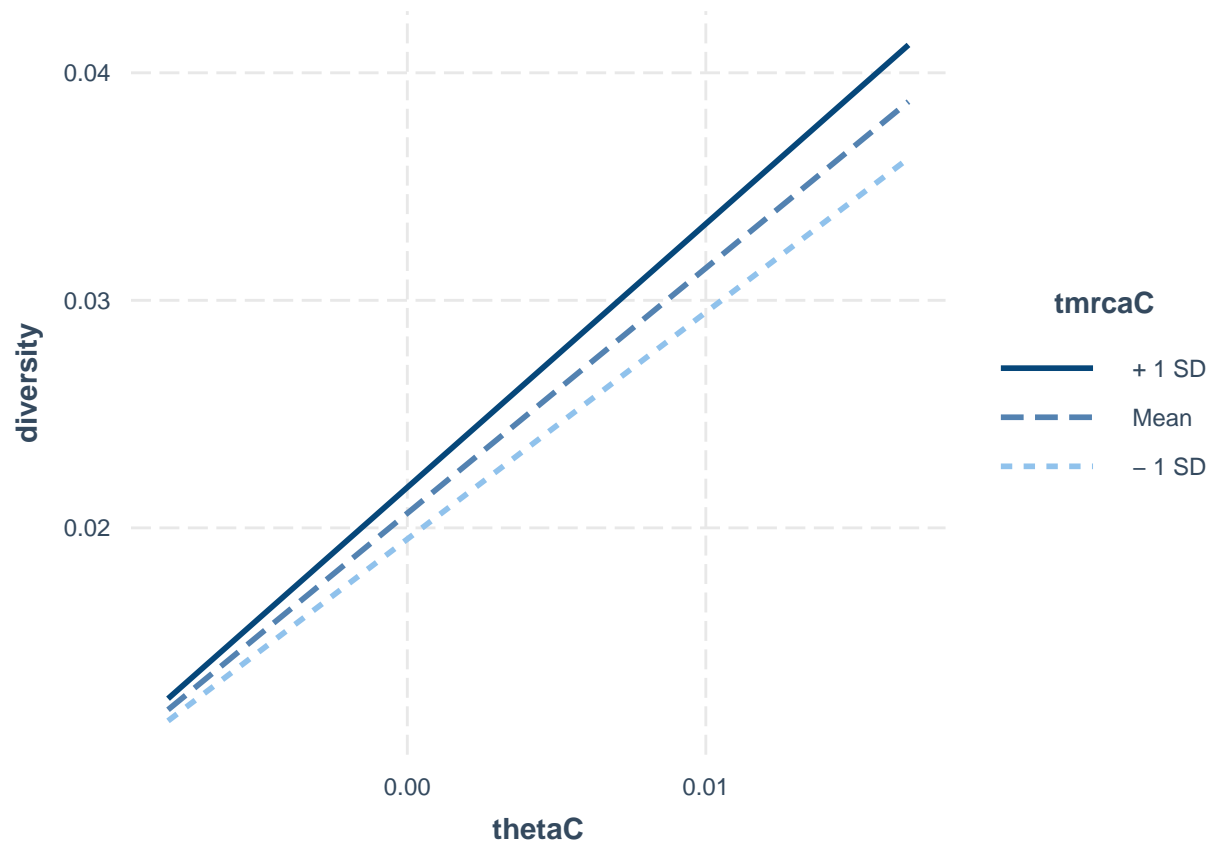
## Histogram of resid(m.diversity.rep\_2)



```
summary(m.diversity.rep_2)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.1Mb.rep_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.091e-03 -3.697e-04 -6.640e-06  2.531e-04  1.138e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0206424  0.0001134  182.048 < 2e-16 ***
## thetaC       1.0774623  0.0197421   54.577 < 2e-16 ***
## rhoC        0.3474698  0.4535646    0.766  0.4508
## tmrcaC       0.0195388  0.0026090    7.489 7.66e-08 ***
## thetaC:tmrcaC 1.3891871  0.5019294    2.768  0.0105 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0006138 on 25 degrees of freedom
## Multiple R-squared:  0.9919, Adjusted R-squared:  0.9906
## F-statistic: 765.1 on 4 and 25 DF,  p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_2, pred = thetaC, modx= tmrcaC)
```



```
g.rep_2 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.1Mb.rep_2, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_2)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.1Mb.rep_2
##      AIC      BIC   logLik
## -350.0905 -340.2821 182.0452
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.01325283
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  0.0206428 0.0001148 179.80497  0.0000
## thetaC       1.0773596 0.0197550  54.53607  0.0000
## tmrcaC       0.0195384 0.0026060   7.49753  0.0000
## rhoC        0.3372081 0.4548652   0.74134  0.4654
## thetaC:tmrcaC 1.3998089 0.5015497   2.79097  0.0099
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.015
## tmrcaC       0.078  0.012
## rhoC        -0.056  0.004  0.165
## thetaC:tmrcaC 0.150 -0.097  0.527 -0.369
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.95006224 -0.66519213 -0.01401553  0.44798577  2.03312007
##
## Residual standard error: 0.0005603058
## Degrees of freedom: 30 total; 25 residual

vif(g.rep_2)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.021369      1.764385      1.468410      2.007938

g.rep_2.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.1Mb.rep_2, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_2.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.1Mb.rep_2
##      AIC      BIC    logLik
## -318.0822 -311.0762 164.0411
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## -0.01297054
##
## Coefficients:
##      Value Std.Error   t-value p-value
## (Intercept)  0.0205937 0.0001941 106.12154  0.0000
## thetaC      1.0716533 0.0343149  31.23000  0.0000
## rhoC       -0.5592219 0.6546032  -0.85429  0.4005
##
## Correlation:
##      (Intr) thetaC
## thetaC  0.000
## rhoC    0.000 -0.081
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.0512845 -0.6085269  0.1849871  0.5460123  1.6511980
##
## Residual standard error: 0.00102108
## Degrees of freedom: 30 total; 27 residual
```

```

anova.diversity <- Anova(m.diversity.rep_2)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.1Mb[1, 2] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.1Mb[2, 2] <- anova.diversity$VarExp[1] * 100
r2.inf.1Mb[3, 2] <- anova.diversity$VarExp[2] * 100
r2.inf.1Mb[4, 2] <- anova.diversity$VarExp[3] * 100
r2.inf.1Mb[5, 2] <- anova.diversity$VarExp[4] * 100

```

### 2.3.3 Replicate 3

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_3/rs.pair_3.", sep = "")
pi.1Mb <- read.table(paste(p, "diversity.1Mb.bedgraph", sep = ""), header = T)
pi.1Mb$avg <- apply(pi.1Mb[4:ncol(pi.1Mb)], 1, mean)
tmrca.1Mb <- read.table(paste(p, "TMRCA.1Mb.bedgraph", sep = ""), header = T)
tmrca.1Mb$avg <- apply(tmrca.1Mb[4:ncol(tmrca.1Mb)], 1, mean)
rho.1Mb <- read.table(paste(p, "rho.1Mb.bedgraph", sep = ""), header = T)
theta.1Mb <- read.table(paste(p, "theta.1Mb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.1Mb$sim, theta.1Mb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.1Mb$sim and theta.1Mb$sample_mean
## S = 86, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9808676

inf.lands.1Mb.rep_3 <- as.data.frame(cbind(pi.1Mb$avg, theta.1Mb$sample_mean, rho.1Mb$sample_mean, tmrca.1Mb$avg))
names(inf.lands.1Mb.rep_3) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.1Mb.rep_3$thetaC <- inf.lands.1Mb.rep_3$theta - mean(inf.lands.1Mb.rep_3$theta)
inf.lands.1Mb.rep_3$tmrcaC <- inf.lands.1Mb.rep_3$tmrca - mean(inf.lands.1Mb.rep_3$tmrca)
inf.lands.1Mb.rep_3$rhoC <- inf.lands.1Mb.rep_3$rho - mean(inf.lands.1Mb.rep_3$rho)

inf.lands.1Mb.rep_3$bin <- 1:nrow(inf.lands.1Mb.rep_3)

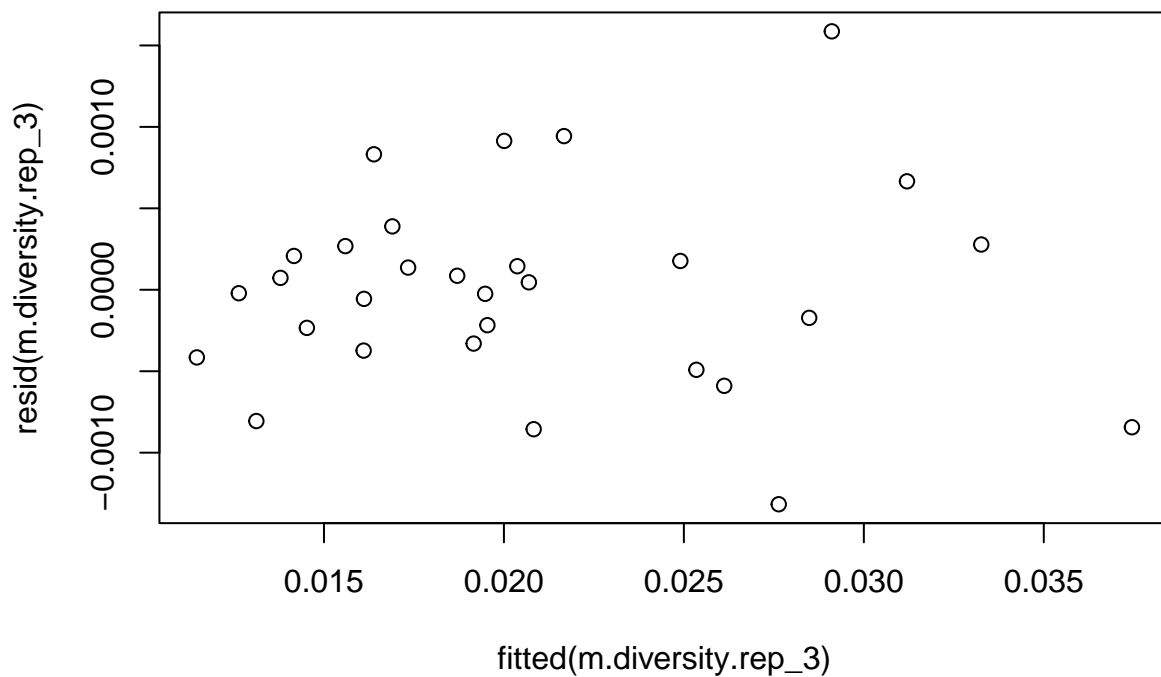
# for merging:
inf.lands.1Mb.rep_3$Replicate <- 3

m.diversity.rep_3 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.1Mb.rep_3)

plot(resid(m.diversity.rep_3)~fitted(m.diversity.rep_3))

```





```
dwtest(m.diversity.rep_3)
```

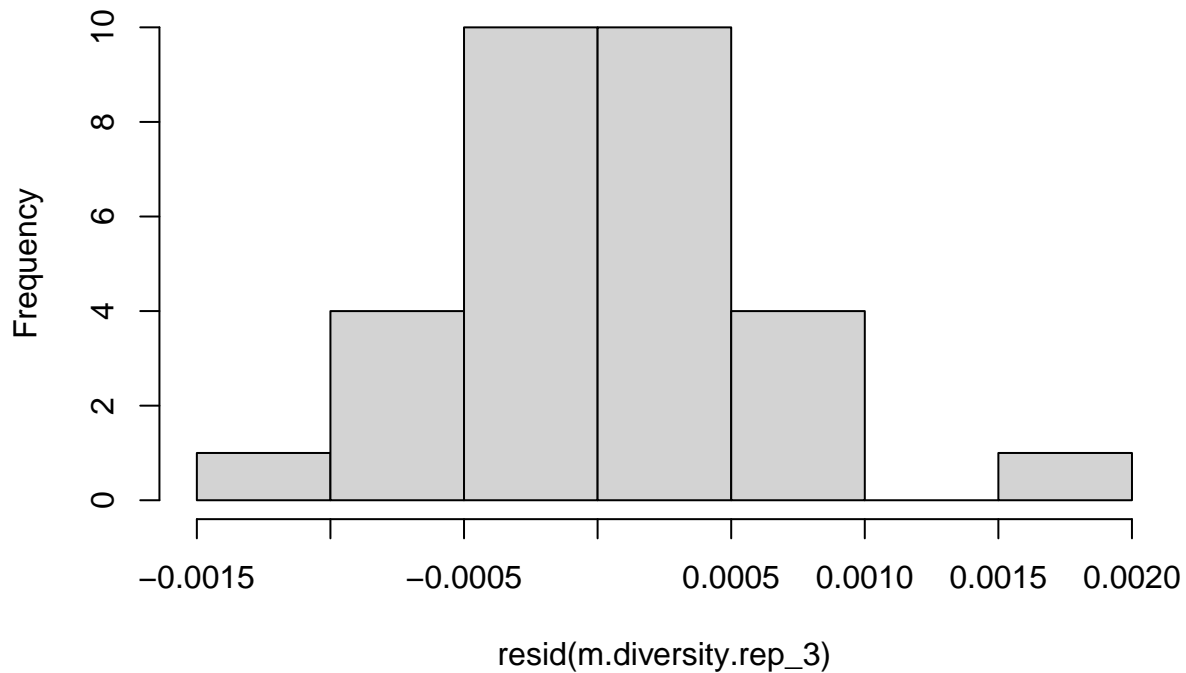
```
##
## Durbin-Watson test
##
## data: m.diversity.rep_3
## DW = 1.3502, p-value = 0.02515
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.diversity.rep_3)
```

```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_3
## HMC = 0.51975, p-value = 0.58
```

```
hist(resid(m.diversity.rep_3))
```

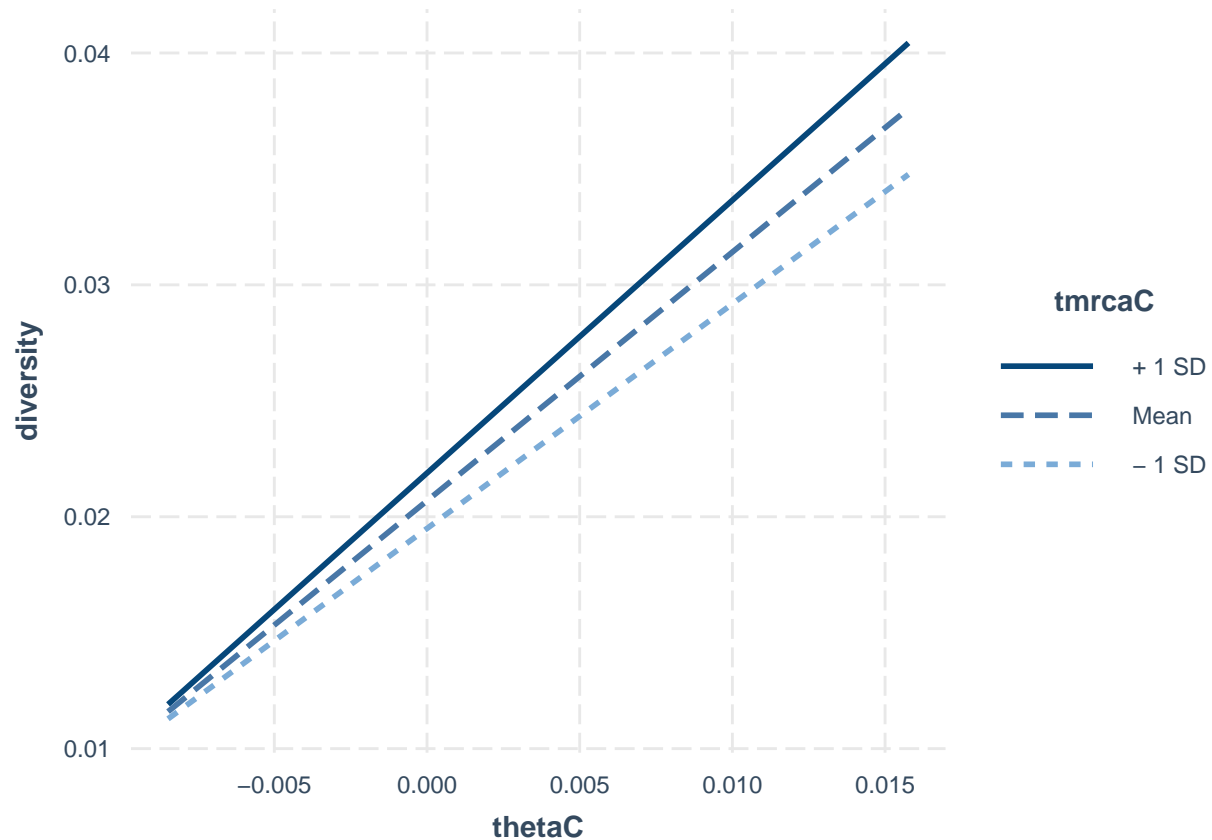
## Histogram of resid(m.diversity.rep\_3)



```
summary(m.diversity.rep_3)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.1Mb.rep_3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.317e-03 -3.626e-04  1.232e-05  2.530e-04  1.587e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0206899  0.0001206  171.565 < 2e-16 ***
## thetaC       1.0723605  0.0213849   50.146 < 2e-16 ***
## rhoC        0.1501119  0.3861502    0.389  0.70076
## tmrcaC       0.0209333  0.0025600    8.177 1.57e-08 ***
## thetaC:tmrcaC 1.8224106  0.5068011    3.596  0.00139 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0006567 on 25 degrees of freedom
## Multiple R-squared:  0.9915, Adjusted R-squared:  0.9901
## F-statistic: 728.3 on 4 and 25 DF,  p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_3, pred = thetaC, modx= tmrcaC)
```



```
g.rep_3 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.1Mb.rep_3, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_3)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.1Mb.rep_3
##      AIC      BIC   logLik
## -349.3777 -339.5693 181.6889
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.3277574
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  0.0206863 0.0001662 124.43077  0.0000
## thetaC       1.0693355 0.0206663  51.74291  0.0000
## tmrcaC       0.0216118 0.0022595   9.56486  0.0000
## rhoC        0.0916173 0.3604146   0.25420  0.8014
## thetaC:tmrcaC 1.9813808 0.4596404   4.31072  0.0002
```

```
##
## Correlation:
##          (Intr) thetaC tmrcaC rhoC
## thetaC      0.024
## tmrcaC     -0.051 -0.272
## rhoC        0.026  0.090  0.062
## thetaC:tmrcaC -0.081 -0.284  0.478 -0.312
##
## Standardized residuals:
##          Min          Q1          Med          Q3          Max
## -2.06120732 -0.59725161  0.02282195  0.36381078  2.61148857
##
## Residual standard error: 0.0005989662
## Degrees of freedom: 30 total; 25 residual

vif(g.rep_3)

##          thetaC          tmrcaC          rhoC thetaC:tmrcaC
##          1.119139          1.425723          1.185890          1.564342

g.rep_3.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.1Mb.rep_3, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_3.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.1Mb.rep_3
##          AIC          BIC      logLik
## -311.0899 -304.0839 160.5449
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##          Phi
## -0.06226692
##
## Coefficients:
##          Value Std.Error  t-value p-value
## (Intercept)  0.0207330 0.0002082 99.56794  0.0000
## thetaC       1.1012864 0.0377084 29.20532  0.0000
## rhoC        -0.3177321 0.6160273 -0.51578  0.6102
##
## Correlation:
##          (Intr) thetaC
## thetaC    0.000
## rhoC     -0.001 -0.036
##
## Standardized residuals:
##          Min          Q1          Med          Q3          Max
## -2.20029757 -0.55512347  0.03085907  0.50826091  1.89284077
##
## Residual standard error: 0.00114935
## Degrees of freedom: 30 total; 27 residual
```

```

anova.diversity <- Anova(m.diversity.rep_3)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.1Mb[1, 3] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.1Mb[2, 3] <- anova.diversity$VarExp[1] * 100
r2.inf.1Mb[3, 3] <- anova.diversity$VarExp[2] * 100
r2.inf.1Mb[4, 3] <- anova.diversity$VarExp[3] * 100
r2.inf.1Mb[5, 3] <- anova.diversity$VarExp[4] * 100

```

### 2.3.4 Replicate 4

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_4/rs.pair_4.", sep = "")
pi.1Mb <- read.table(paste(p, "diversity.1Mb.bedgraph", sep = ""), header = T)
pi.1Mb$avg <- apply(pi.1Mb[4:ncol(pi.1Mb)], 1, mean)
tmrca.1Mb <- read.table(paste(p, "TMRCA.1Mb.bedgraph", sep = ""), header = T)
tmrca.1Mb$avg <- apply(tmrca.1Mb[4:ncol(tmrca.1Mb)], 1, mean)
rho.1Mb <- read.table(paste(p, "rho.1Mb.bedgraph", sep = ""), header = T)
theta.1Mb <- read.table(paste(p, "theta.1Mb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.1Mb$sim, theta.1Mb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.1Mb$sim and theta.1Mb$sample_mean
## S = 80, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9822024

inf.lands.1Mb.rep_4 <- as.data.frame(cbind(pi.1Mb$avg, theta.1Mb$sample_mean, rho.1Mb$sample_mean, tmrca.1Mb$avg))
names(inf.lands.1Mb.rep_4) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.1Mb.rep_4$thetaC <- inf.lands.1Mb.rep_4$theta - mean(inf.lands.1Mb.rep_4$theta)
inf.lands.1Mb.rep_4$tmrcaC <- inf.lands.1Mb.rep_4$tmrca - mean(inf.lands.1Mb.rep_4$tmrca)
inf.lands.1Mb.rep_4$rhoC <- inf.lands.1Mb.rep_4$rho - mean(inf.lands.1Mb.rep_4$rho)

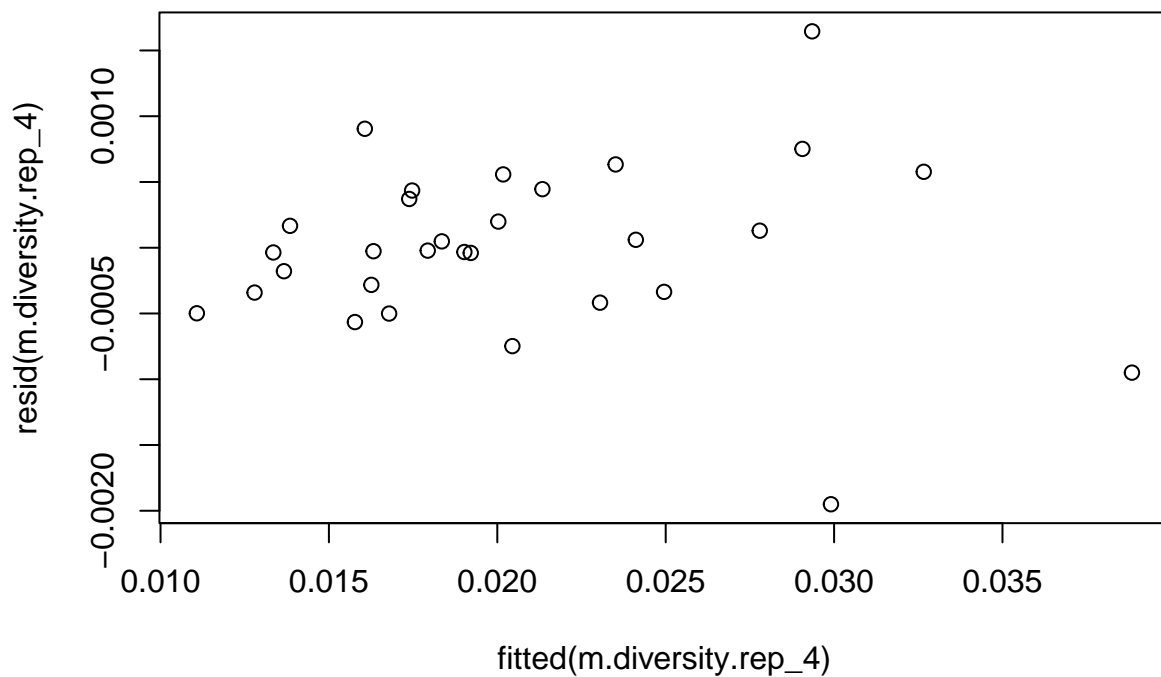
inf.lands.1Mb.rep_4$bin <- 1:nrow(inf.lands.1Mb.rep_4)

# for merging:
inf.lands.1Mb.rep_4$Replicate <- 4

m.diversity.rep_4 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.1Mb.rep_4)

plot(resid(m.diversity.rep_4)~fitted(m.diversity.rep_4))

```



```
dwtest(m.diversity.rep_4)
```

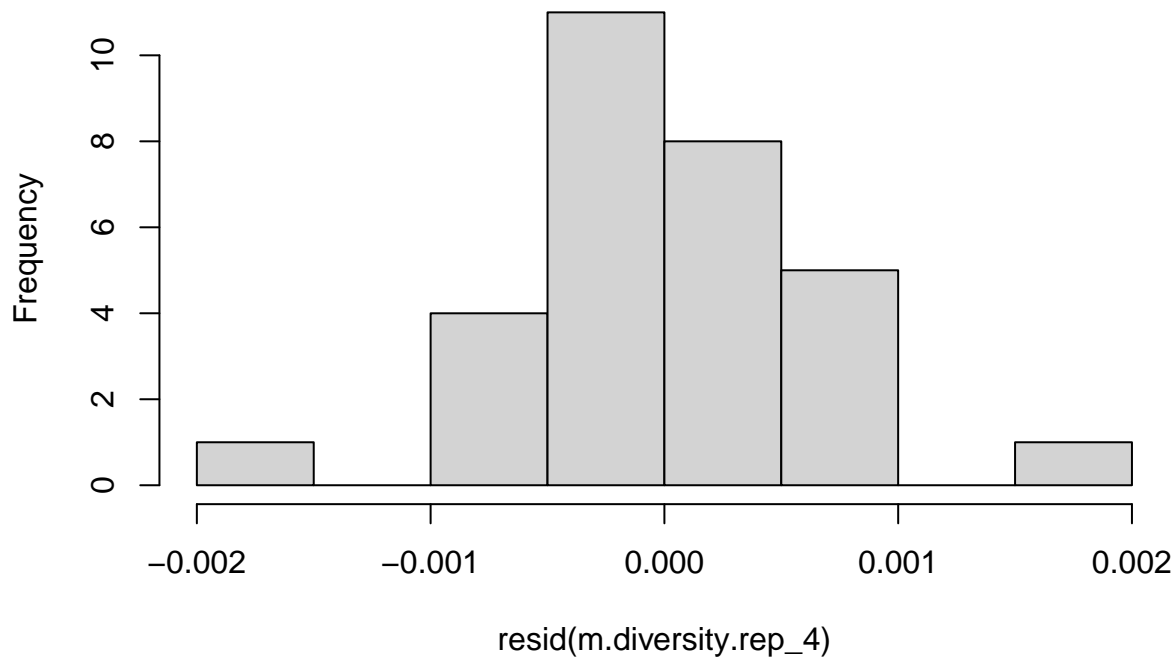
```
##
## Durbin-Watson test
##
## data: m.diversity.rep_4
## DW = 2.0871, p-value = 0.5863
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.diversity.rep_4)
```

```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_4
## HMC = 0.42593, p-value = 0.287
```

```
hist(resid(m.diversity.rep_4))
```

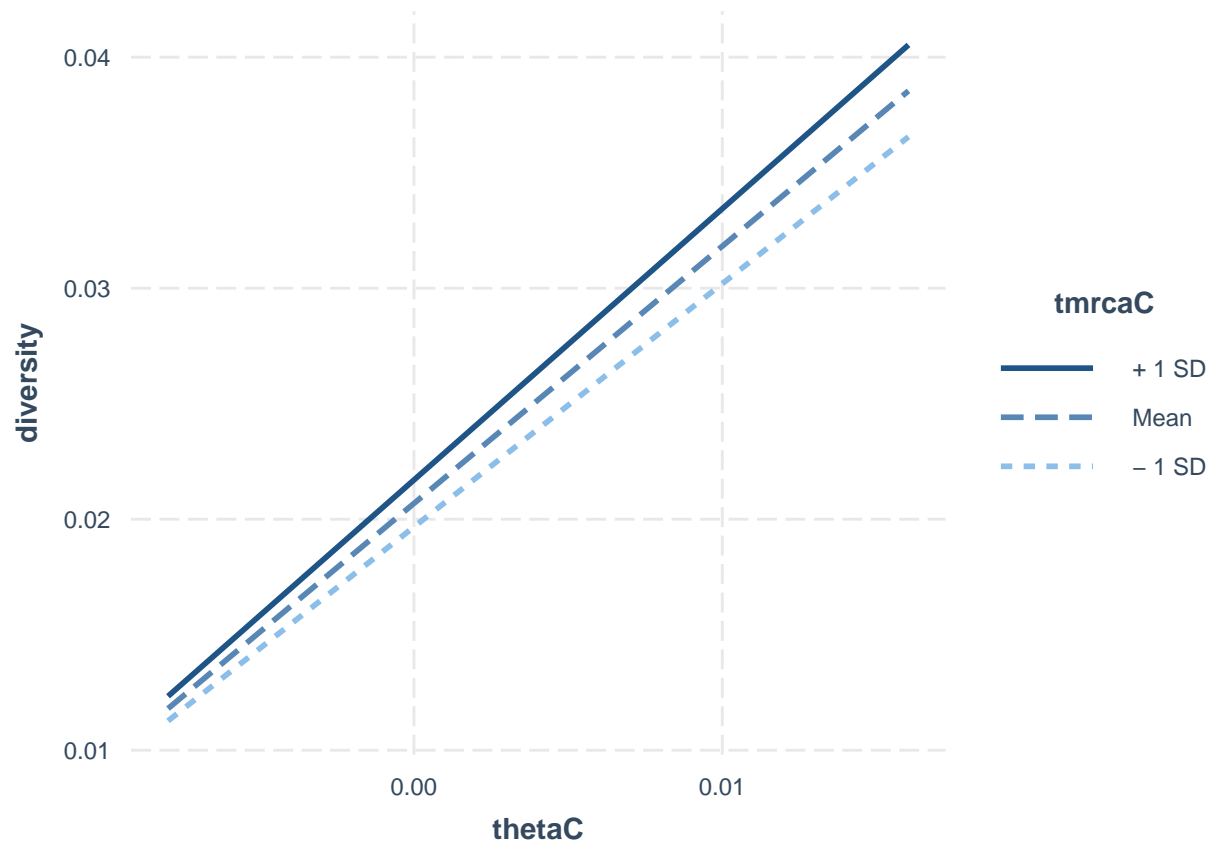
## Histogram of resid(m.diversity.rep\_4)



```
summary(m.diversity.rep_4)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.1Mb.rep_4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.950e-03 -3.397e-04 -2.417e-05  4.190e-04  1.645e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0206840  0.0001284 161.149 < 2e-16 ***
## thetaC       1.1134126  0.0224842  49.520 < 2e-16 ***
## rhoC        -0.0339570  0.4581591  -0.074   0.942
## tmrcaC       0.0178545  0.0033294   5.363 1.46e-05 ***
## thetaC:tmrcaC 1.0646819  0.6921958   1.538   0.137
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.000703 on 25 degrees of freedom
## Multiple R-squared:  0.9901, Adjusted R-squared:  0.9886
## F-statistic: 628.1 on 4 and 25 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_4, pred = thetaC, modx= tmrcaC)
```



```
g.rep_4 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.1Mb.rep_4, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_4)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.1Mb.rep_4
##      AIC      BIC    logLik
## -342.0193 -332.2109 178.0096
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## -0.05853672
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  0.0206839 0.0001213 170.49504  0.0000
## thetaC       1.1137846 0.0222646  50.02482  0.0000
## tmrcaC       0.0173804 0.0033848   5.13488  0.0000
## rhoC        -0.0467801 0.4513716  -0.10364  0.9183
## thetaC:tmrcaC 0.9716766 0.6991874   1.38972  0.1769
```



```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      0.001
## tmrcaC     -0.003 -0.036
## rhoC        0.002  0.016  0.215
## thetaC:tmrcaC -0.006 -0.054  0.653 -0.181
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -3.03132272 -0.49951188 -0.03148496  0.64957883  2.58514901
##
## Residual standard error: 0.0006419926
## Degrees of freedom: 30 total; 25 residual

vif(g.rep_4)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.002978      2.181651      1.293335      2.154008

g.rep_4.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.1Mb.rep_4, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_4.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.1Mb.rep_4
##      AIC      BIC    logLik
## -323.3904 -316.3844 166.6952
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## -0.3430849
##
## Coefficients:
##      Value Std.Error  t-value p-value
## (Intercept)  0.0206798 0.0001348 153.42630  0.0000
## thetaC      1.1175389 0.0287817  38.82807  0.0000
## rhoC       -1.1557791 0.4865494  -2.37546  0.0249
##
## Correlation:
##      (Intr) thetaC
## thetaC  0.004
## rhoC    0.003  0.002
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.3370913 -0.2969692  0.1671986  0.5302828  2.0808719
##
## Residual standard error: 0.0009928647
## Degrees of freedom: 30 total; 27 residual
```

```

anova.diversity <- Anova(m.diversity.rep_4)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.1Mb[1, 4] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.1Mb[2, 4] <- anova.diversity$VarExp[1] * 100
r2.inf.1Mb[3, 4] <- anova.diversity$VarExp[2] * 100
r2.inf.1Mb[4, 4] <- anova.diversity$VarExp[3] * 100
r2.inf.1Mb[5, 4] <- anova.diversity$VarExp[4] * 100

```

### 2.3.5 Replicate 5

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_5/rs.pair_5.", sep = "")
pi.1Mb <- read.table(paste(p, "diversity.1Mb.bedgraph", sep = ""), header = T)
pi.1Mb$avg <- apply(pi.1Mb[4:ncol(pi.1Mb)], 1, mean)
tmrca.1Mb <- read.table(paste(p, "TMRCA.1Mb.bedgraph", sep = ""), header = T)
tmrca.1Mb$avg <- apply(tmrca.1Mb[4:ncol(tmrca.1Mb)], 1, mean)
rho.1Mb <- read.table(paste(p, "rho.1Mb.bedgraph", sep = ""), header = T)
theta.1Mb <- read.table(paste(p, "theta.1Mb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.1Mb$sim, theta.1Mb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.1Mb$sim and theta.1Mb$sample_mean
## S = 76, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9830923

inf.lands.1Mb.rep_5 <- as.data.frame(cbind(pi.1Mb$avg, theta.1Mb$sample_mean, rho.1Mb$sample_mean, tmrca.1Mb$avg))
names(inf.lands.1Mb.rep_5) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.1Mb.rep_5$thetaC <- inf.lands.1Mb.rep_5$theta - mean(inf.lands.1Mb.rep_5$theta)
inf.lands.1Mb.rep_5$tmrcaC <- inf.lands.1Mb.rep_5$tmrca - mean(inf.lands.1Mb.rep_5$tmrca)
inf.lands.1Mb.rep_5$rhoC <- inf.lands.1Mb.rep_5$rho - mean(inf.lands.1Mb.rep_5$rho)

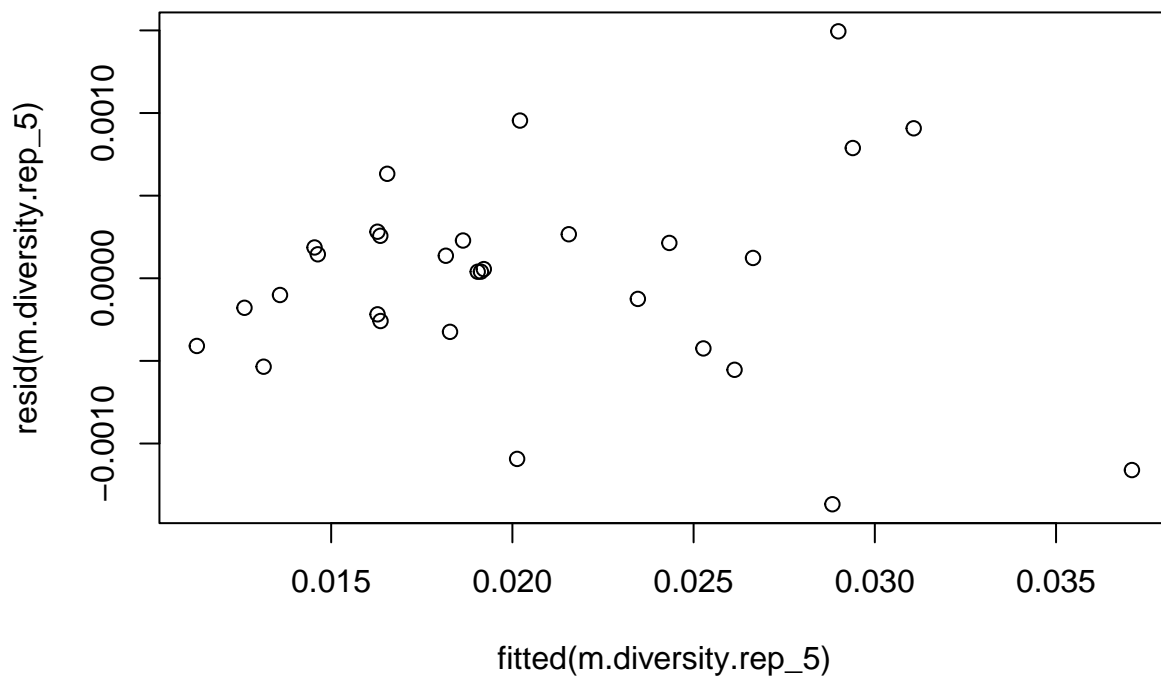
inf.lands.1Mb.rep_5$bin <- 1:nrow(inf.lands.1Mb.rep_5)

# for merging:
inf.lands.1Mb.rep_5$Replicate <- 5

m.diversity.rep_5 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.1Mb.rep_5)

plot(resid(m.diversity.rep_5)~fitted(m.diversity.rep_5))

```



```
dwtest(m.diversity.rep_5)
```

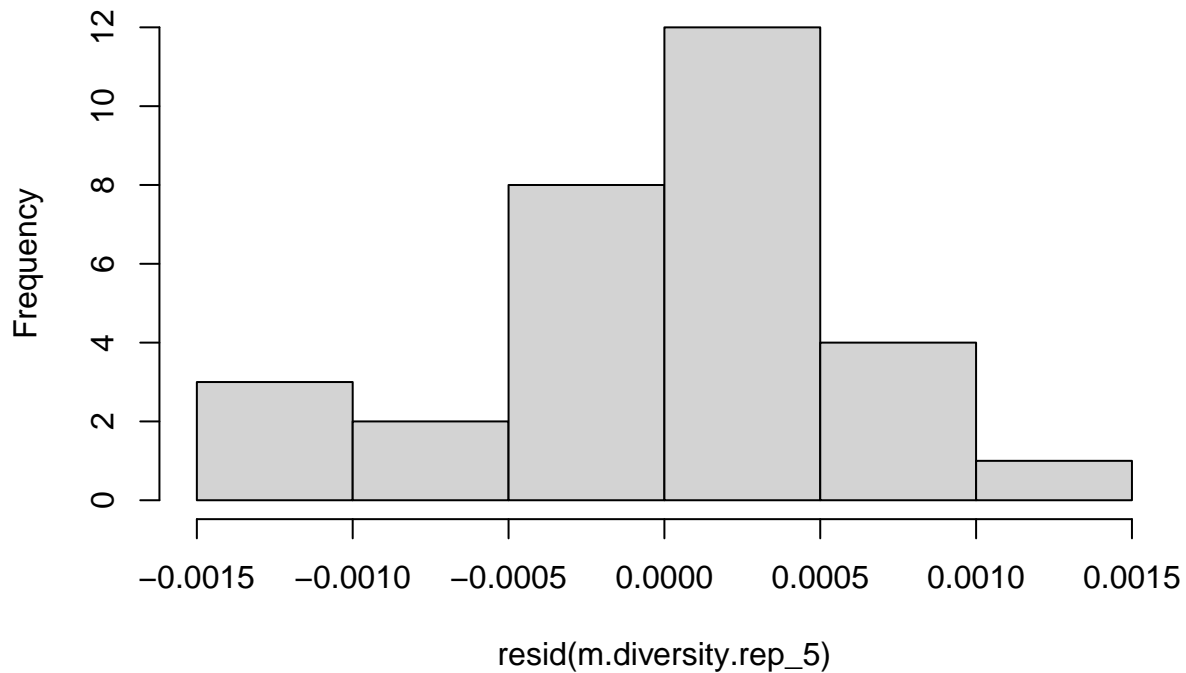
```
##
## Durbin-Watson test
##
## data: m.diversity.rep_5
## DW = 1.8381, p-value = 0.2796
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_5)
```

```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_5
## HMC = 0.49429, p-value = 0.499
```

```
hist(resid(m.diversity.rep_5))
```

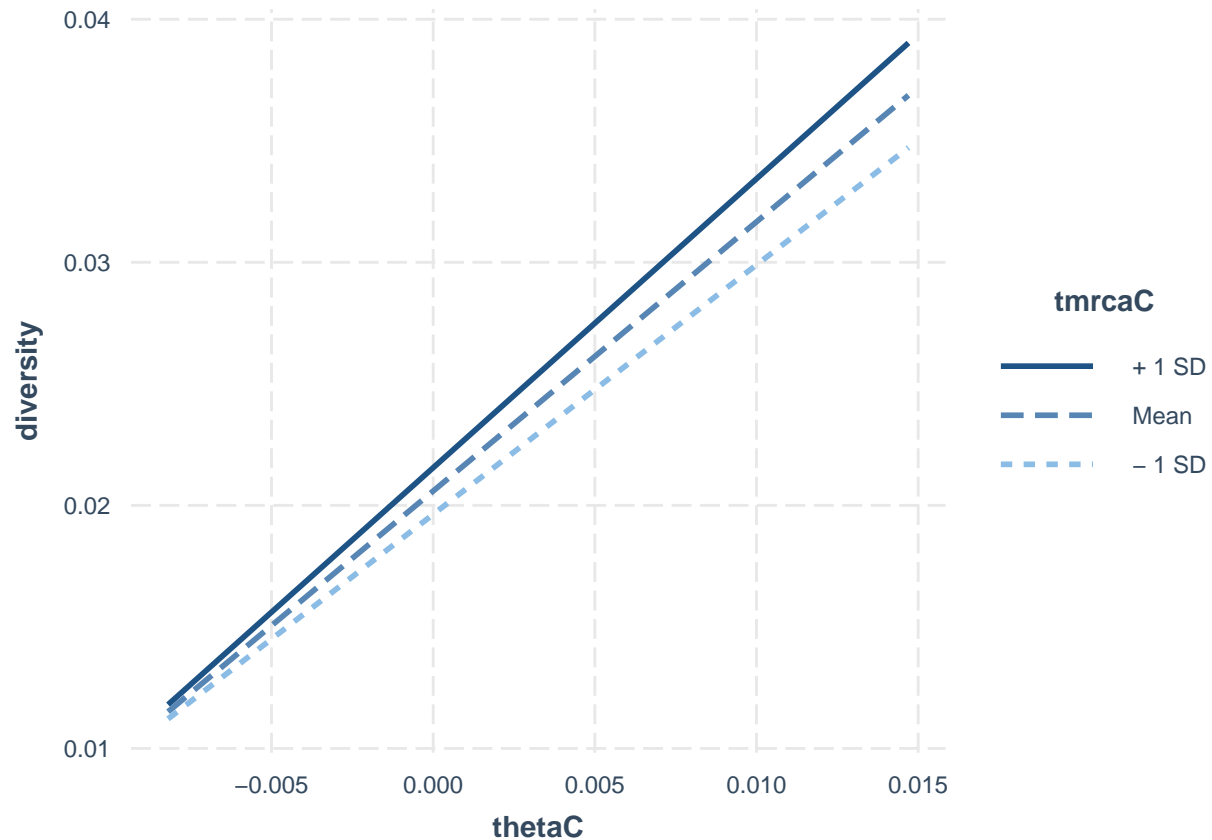
## Histogram of resid(m.diversity.rep\_5)



```
summary(m.diversity.rep_5)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.1Mb.rep_5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.367e-03 -3.078e-04  4.773e-05  2.499e-04  1.494e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.0205901  0.0001215  169.475 < 2e-16 ***
## thetaC        1.1075044  0.0217775   50.855 < 2e-16 ***
## rhoC         -0.1291914  0.4158082   -0.311  0.7586
## tmrcaC         0.0174135  0.0033337    5.224 2.09e-05 ***
## thetaC:tmrcaC  1.4750759  0.6476973    2.277  0.0316 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0006642 on 25 degrees of freedom
## Multiple R-squared:  0.9905, Adjusted R-squared:  0.989
## F-statistic: 650.9 on 4 and 25 DF,  p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_5, pred = thetaC, modx= tmrcaC)
```



```
g.rep_5 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.1Mb.rep_5, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_5)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.1Mb.rep_5
##      AIC      BIC    logLik
## -345.5754 -335.767 179.7877
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.09742974
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  0.0205933 0.0001336 154.16894  0.0000
## thetaC       1.1072954 0.0218882  50.58862  0.0000
## tmrcaC       0.0179569 0.0032649   5.50000  0.0000
## rhoC        -0.1181976 0.4172436  -0.28328  0.7793
## thetaC:tmrcaC 1.6258812 0.6398768   2.54093  0.0176
```

```
##
## Correlation:
##          (Intr) thetaC tmrcaC rhoC
## thetaC      0.002
## tmrcaC      0.033  0.055
## rhoC        -0.002  0.016  0.365
## thetaC:tmrcaC 0.058  0.056  0.609 -0.028
##
## Standardized residuals:
##          Min          Q1          Med          Q3          Max
## -2.2002408 -0.4979867  0.0723577  0.3822209  2.4009759
##
## Residual standard error: 0.0006068364
## Degrees of freedom: 30 total; 25 residual

vif(g.rep_5)

##          thetaC          tmrcaC          rhoC thetaC:tmrcaC
##          1.003874          2.071251          1.303671          1.797134

g.rep_5.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.1Mb.rep_5, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_5.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.1Mb.rep_5
##      AIC      BIC    logLik
## -327.0266 -320.0206 168.5133
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## -0.151117
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  0.0205697 0.0001476 139.31706  0.0000
## thetaC       1.0982569 0.0293493  37.42021  0.0000
## rhoC        -1.2150813 0.4706686  -2.58161  0.0156
##
## Correlation:
##          (Intr) thetaC
## thetaC    0.001
## rhoC     -0.002 -0.018
##
## Standardized residuals:
##          Min          Q1          Med          Q3          Max
## -1.9762372 -0.5677260  0.1592677  0.6160406  2.4429343
##
## Residual standard error: 0.0008894714
## Degrees of freedom: 30 total; 27 residual
```

```

anova.diversity <- Anova(m.diversity.rep_5)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.1Mb[1, 5] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.1Mb[2, 5] <- anova.diversity$VarExp[1] * 100
r2.inf.1Mb[3, 5] <- anova.diversity$VarExp[2] * 100
r2.inf.1Mb[4, 5] <- anova.diversity$VarExp[3] * 100
r2.inf.1Mb[5, 5] <- anova.diversity$VarExp[4] * 100

```

### 2.3.6 Replicate 6

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_6/rs.pair_6.", sep = "")
pi.1Mb <- read.table(paste(p, "diversity.1Mb.bedgraph", sep = ""), header = T)
pi.1Mb$avg <- apply(pi.1Mb[4:ncol(pi.1Mb)], 1, mean)
tmrca.1Mb <- read.table(paste(p, "TMRCA.1Mb.bedgraph", sep = ""), header = T)
tmrca.1Mb$avg <- apply(tmrca.1Mb[4:ncol(tmrca.1Mb)], 1, mean)
rho.1Mb <- read.table(paste(p, "rho.1Mb.bedgraph", sep = ""), header = T)
theta.1Mb <- read.table(paste(p, "theta.1Mb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.1Mb$sim, theta.1Mb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.1Mb$sim and theta.1Mb$sample_mean
## S = 84, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9813126

inf.lands.1Mb.rep_6 <- as.data.frame(cbind(pi.1Mb$avg, theta.1Mb$sample_mean, rho.1Mb$sample_mean, tmrca.1Mb$avg))
names(inf.lands.1Mb.rep_6) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.1Mb.rep_6$thetaC <- inf.lands.1Mb.rep_6$theta - mean(inf.lands.1Mb.rep_6$theta)
inf.lands.1Mb.rep_6$tmrcaC <- inf.lands.1Mb.rep_6$tmrca - mean(inf.lands.1Mb.rep_6$tmrca)
inf.lands.1Mb.rep_6$rhoC <- inf.lands.1Mb.rep_6$rho - mean(inf.lands.1Mb.rep_6$rho)

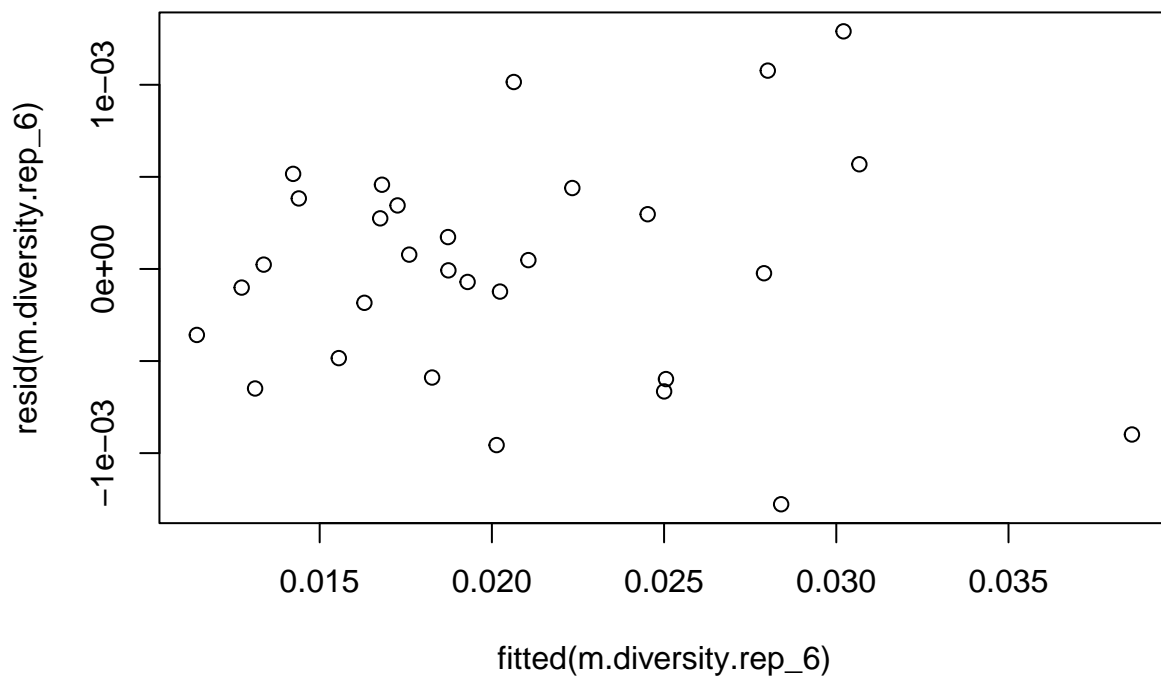
inf.lands.1Mb.rep_6$bin <- 1:nrow(inf.lands.1Mb.rep_6)

# for merging:
inf.lands.1Mb.rep_6$Replicate <- 6

m.diversity.rep_6 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.1Mb.rep_6)

plot(resid(m.diversity.rep_6)~fitted(m.diversity.rep_6))

```



```
dwtest(m.diversity.rep_6)
```

```
##
## Durbin-Watson test
##
## data: m.diversity.rep_6
## DW = 1.7678, p-value = 0.2162
## alternative hypothesis: true autocorrelation is greater than 0
```

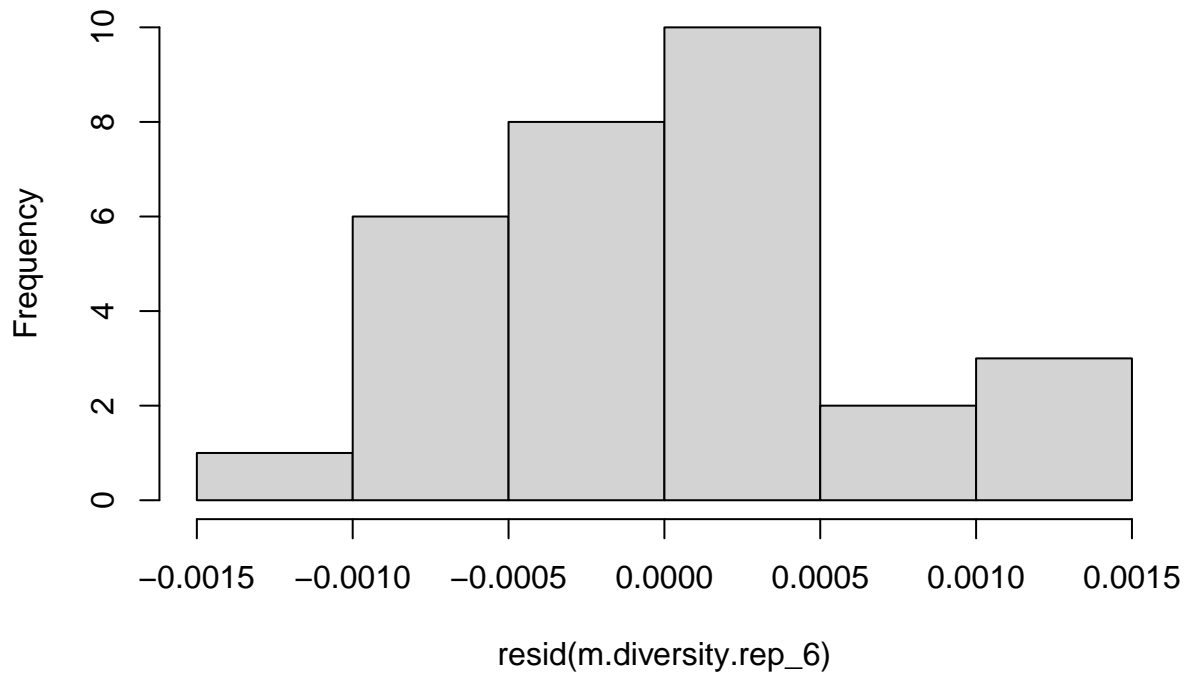
```
hmcetest(m.diversity.rep_6)
```

```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_6
## HMC = 0.38206, p-value = 0.182
```

```
hist(resid(m.diversity.rep_6))
```



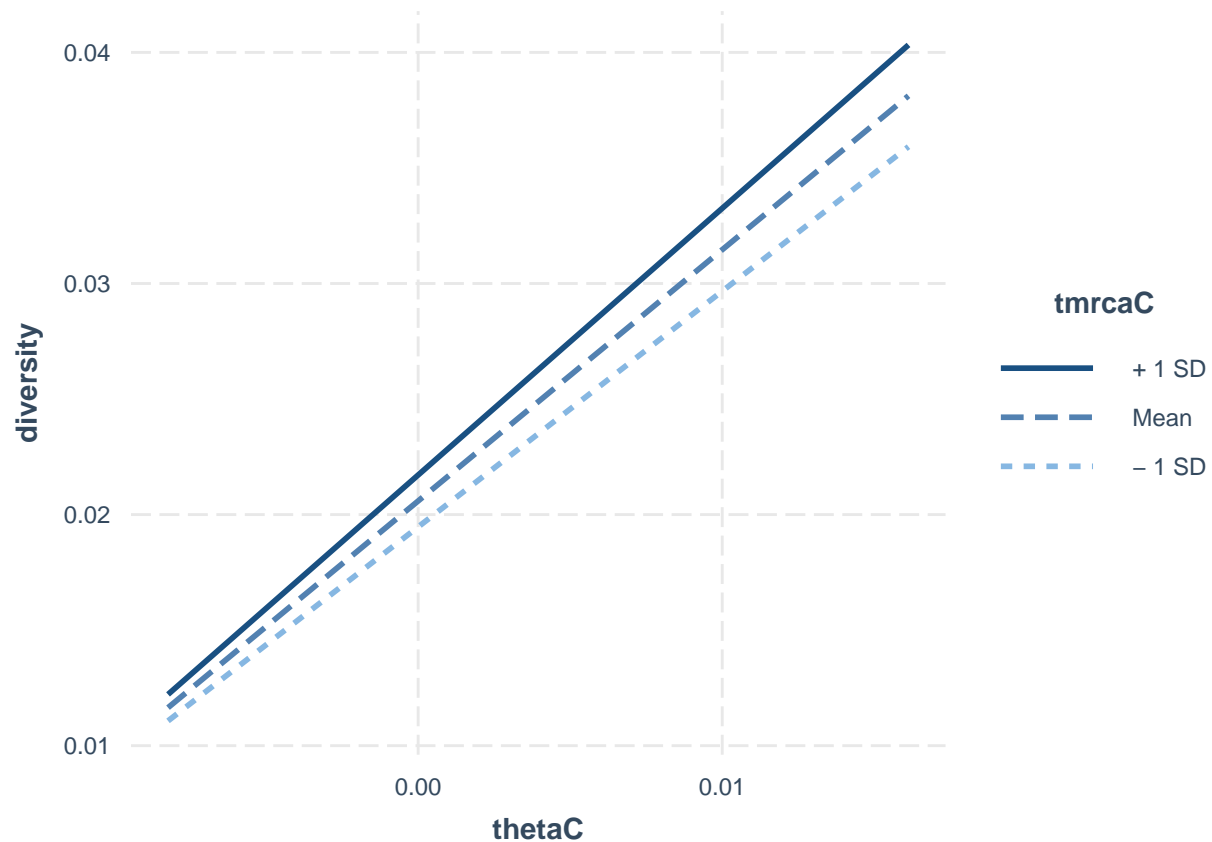
## Histogram of resid(m.diversity.rep\_6)



```
summary(m.diversity.rep_6)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.1Mb.rep_6)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.277e-03 -4.527e-04  7.960e-06  3.734e-04  1.290e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0205897  0.0001193  172.524 < 2e-16 ***
## thetaC       1.0872234  0.0212487   51.167 < 2e-16 ***
## rhoC        0.3998838  0.3697035    1.082  0.2897
## tmrcaC       0.0196732  0.0028089    7.004 2.43e-07 ***
## thetaC:tmrcaC 1.1711909  0.4979919    2.352  0.0269 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0006531 on 25 degrees of freedom
## Multiple R-squared:  0.9911, Adjusted R-squared:  0.9897
## F-statistic: 694.4 on 4 and 25 DF,  p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_6, pred = thetaC, modx= tmrcaC)
```



```
g.rep_6 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.1Mb.rep_6, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_6)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.1Mb.rep_6
##      AIC      BIC   logLik
## -346.8212 -337.0128 180.4106
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.1358209
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  0.0205924 0.0001363 151.13216  0.0000
## thetaC       1.0860890 0.0213756  50.80966  0.0000
## tmrcaC       0.0198458 0.0027501   7.21651  0.0000
## rhoC        0.3466959 0.3730361   0.92939  0.3616
## thetaC:tmrcaC 1.3137006 0.4935342   2.66182  0.0134
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.007
## tmrcaC       0.000 -0.065
## rhoC        -0.010  0.040  0.485
## thetaC:tmrcaC 0.033 -0.186  0.336 -0.157
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.09674866 -0.76764130  0.02681106  0.64010637  2.02153252
##
## Residual standard error: 0.0005969812
## Degrees of freedom: 30 total; 25 residual

vif(g.rep_6)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.036142      1.691913      1.539768      1.360141

g.rep_6.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.1Mb.rep_6, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_6.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.1Mb.rep_6
##      AIC      BIC    logLik
## -317.8306 -310.8246 163.9153
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## -0.04598612
##
## Coefficients:
##      Value Std.Error   t-value p-value
## (Intercept)  0.0205747 0.0001889 108.91122  0.0000
## thetaC      1.0881740 0.0343243  31.70273  0.0000
## rhoC      -0.8461274 0.4782870  -1.76908  0.0882
##
## Correlation:
##      (Intr) thetaC
## thetaC  0.000
## rhoC   -0.001 -0.001
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.7855991 -0.6320726  0.1751760  0.5604975  2.9148885
##
## Residual standard error: 0.001026338
## Degrees of freedom: 30 total; 27 residual
```

```

anova.diversity <- Anova(m.diversity.rep_6)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.1Mb[1, 6] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.1Mb[2, 6] <- anova.diversity$VarExp[1] * 100
r2.inf.1Mb[3, 6] <- anova.diversity$VarExp[2] * 100
r2.inf.1Mb[4, 6] <- anova.diversity$VarExp[3] * 100
r2.inf.1Mb[5, 6] <- anova.diversity$VarExp[4] * 100

```

### 2.3.7 Replicate 7

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_7/rs.pair_7.", sep = "")
pi.1Mb <- read.table(paste(p, "diversity.1Mb.bedgraph", sep = ""), header = T)
pi.1Mb$avg <- apply(pi.1Mb[4:ncol(pi.1Mb)], 1, mean)
tmrca.1Mb <- read.table(paste(p, "TMRCA.1Mb.bedgraph", sep = ""), header = T)
tmrca.1Mb$avg <- apply(tmrca.1Mb[4:ncol(tmrca.1Mb)], 1, mean)
rho.1Mb <- read.table(paste(p, "rho.1Mb.bedgraph", sep = ""), header = T)
theta.1Mb <- read.table(paste(p, "theta.1Mb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.1Mb$sim, theta.1Mb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.1Mb$sim and theta.1Mb$sample_mean
## S = 112, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9750834

inf.lands.1Mb.rep_7 <- as.data.frame(cbind(pi.1Mb$avg, theta.1Mb$sample_mean, rho.1Mb$sample_mean, tmrca.1Mb$avg))
names(inf.lands.1Mb.rep_7) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.1Mb.rep_7$thetaC <- inf.lands.1Mb.rep_7$theta - mean(inf.lands.1Mb.rep_7$theta)
inf.lands.1Mb.rep_7$tmrcaC <- inf.lands.1Mb.rep_7$tmrca - mean(inf.lands.1Mb.rep_7$tmrca)
inf.lands.1Mb.rep_7$rhoC <- inf.lands.1Mb.rep_7$rho - mean(inf.lands.1Mb.rep_7$rho)

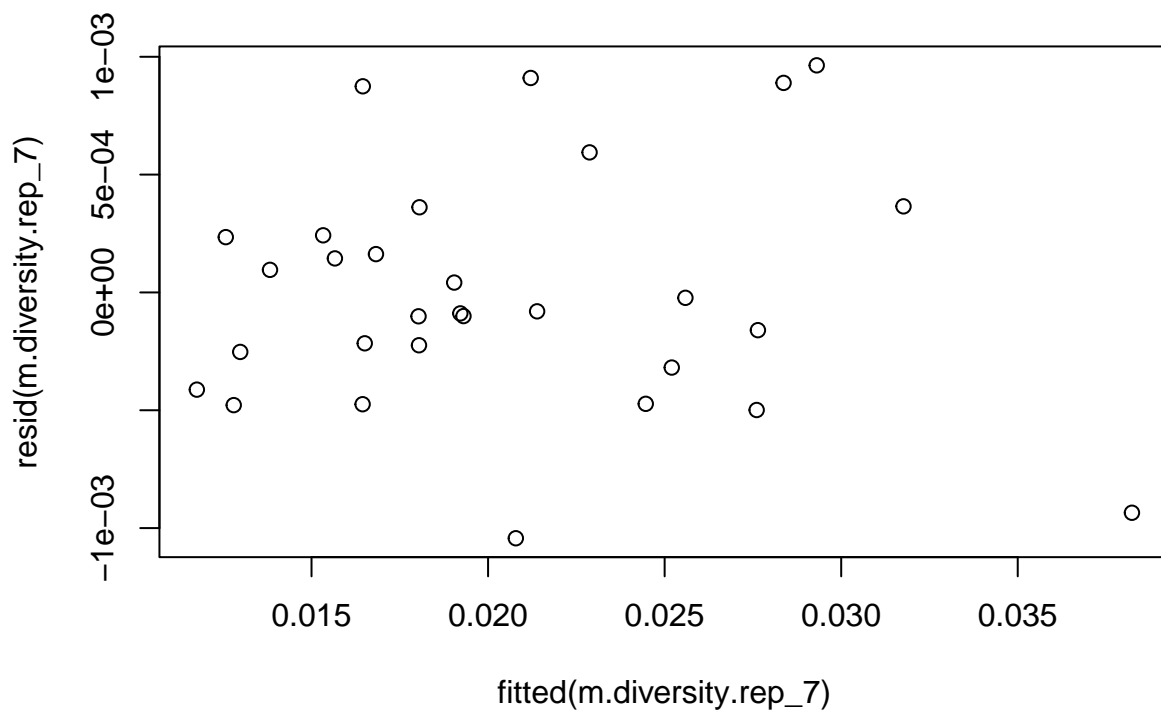
inf.lands.1Mb.rep_7$bin <- 1:nrow(inf.lands.1Mb.rep_7)

# for merging:
inf.lands.1Mb.rep_7$Replicate <- 7

m.diversity.rep_7 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.1Mb.rep_7)

plot(resid(m.diversity.rep_7)~fitted(m.diversity.rep_7))

```



```
dwtest(m.diversity.rep_7)
```

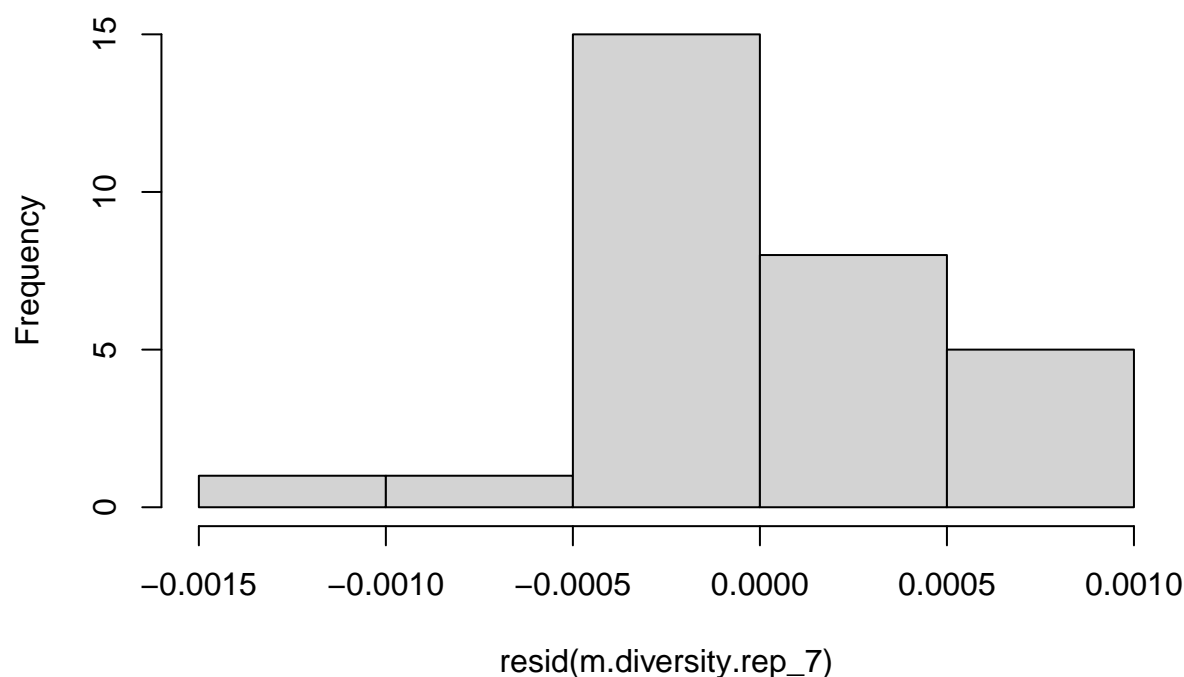
```
##
## Durbin-Watson test
##
## data: m.diversity.rep_7
## DW = 1.6938, p-value = 0.1634
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_7)
```

```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_7
## HMC = 0.40789, p-value = 0.25
```

```
hist(resid(m.diversity.rep_7))
```

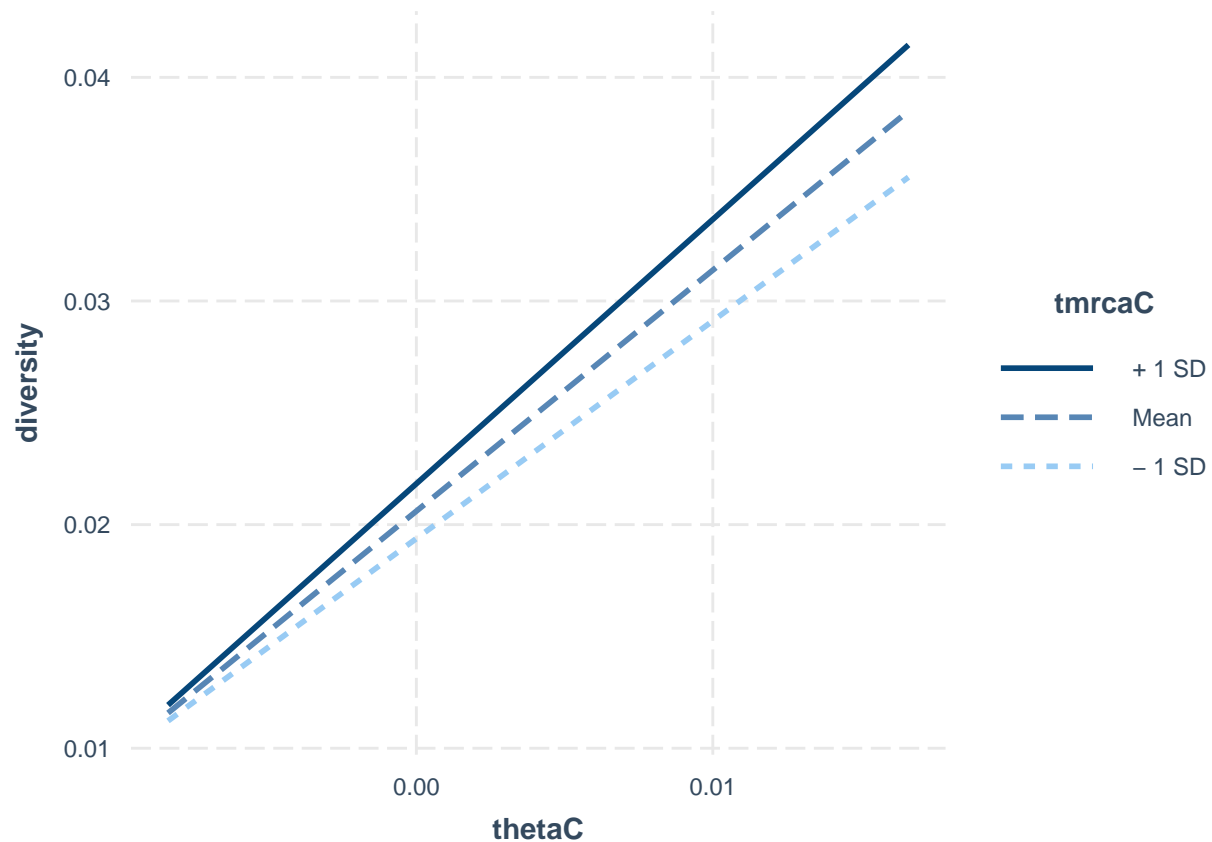
## Histogram of resid(m.diversity.rep\_7)



```
summary(m.diversity.rep_7)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.1Mb.rep_7)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0010431 -0.0003022 -0.0000846  0.0002404  0.0009636
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0206041  0.0001001  205.827  < 2e-16 ***
## thetaC       1.0776129  0.0174765   61.661  < 2e-16 ***
## rhoC         0.0680221  0.3366597    0.202  0.841513
## tmrcaC       0.0238615  0.0027851    8.568 6.58e-09 ***
## thetaC:tmrcaC 2.0280640  0.4774338    4.248 0.000262 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005471 on 25 degrees of freedom
## Multiple R-squared:  0.9937, Adjusted R-squared:  0.9927
## F-statistic: 992.7 on 4 and 25 DF,  p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_7, pred = thetaC, modx= tmrcaC)
```



```
g.rep_7 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.1Mb.rep_7, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_7)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.1Mb.rep_7
##      AIC      BIC   logLik
## -357.7791 -347.9708 185.8896
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.1726746
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  0.0206041 0.0001184 174.02851  0.0000
## thetaC       1.0780510 0.0175352  61.47931  0.0000
## tmrcaC       0.0233774 0.0026923   8.68312  0.0000
## rhoC        -0.0447197 0.3354118  -0.13333  0.8950
## thetaC:tmrcaC 2.0331470 0.4572788   4.44619  0.0002
```

```
##
## Correlation:
##          (Intr) thetaC tmrcaC rhoC
## thetaC      -0.005
## tmrcaC       0.022 -0.097
## rhoC        -0.009 -0.074  0.228
## thetaC:tmrcaC 0.046 -0.103  0.586 -0.198
##
## Standardized residuals:
##          Min          Q1          Med          Q3          Max
## -2.1497675 -0.5506932 -0.1283266  0.4700014  1.9723133
##
## Residual standard error: 0.0005001377
## Degrees of freedom: 30 total; 25 residual

vif(g.rep_7)

##          thetaC          tmrcaC          rhoC thetaC:tmrcaC
##          1.020214          1.873531          1.290005          1.862508

g.rep_7.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.1Mb.rep_7, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_7.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.1Mb.rep_7
##      AIC      BIC logLik
## -319.9699 -312.9639 164.985
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.1179429
##
## Coefficients:
##              Value Std.Error  t-value p-value
## (Intercept)  0.0205818 0.0002149 95.79213  0.0000
## thetaC       1.0897560 0.0336214 32.41258  0.0000
## rhoC        -0.8110199 0.5672255 -1.42980  0.1642
##
## Correlation:
##          (Intr) thetaC
## thetaC -0.001
## rhoC    0.002 -0.097
##
## Standardized residuals:
##          Min          Q1          Med          Q3          Max
## -2.37431230 -0.29863999  0.01070772  0.55844318  2.30950358
##
## Residual standard error: 0.0009960966
## Degrees of freedom: 30 total; 27 residual
```



```

anova.diversity <- Anova(m.diversity.rep_7)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.1Mb[1, 7] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.1Mb[2, 7] <- anova.diversity$VarExp[1] * 100
r2.inf.1Mb[3, 7] <- anova.diversity$VarExp[2] * 100
r2.inf.1Mb[4, 7] <- anova.diversity$VarExp[3] * 100
r2.inf.1Mb[5, 7] <- anova.diversity$VarExp[4] * 100

```

### 2.3.8 Replicate 8

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_8/rs.pair_8.", sep = "")
pi.1Mb <- read.table(paste(p, "diversity.1Mb.bedgraph", sep = ""), header = T)
pi.1Mb$avg <- apply(pi.1Mb[4:ncol(pi.1Mb)], 1, mean)
tmrca.1Mb <- read.table(paste(p, "TMRCA.1Mb.bedgraph", sep = ""), header = T)
tmrca.1Mb$avg <- apply(tmrca.1Mb[4:ncol(tmrca.1Mb)], 1, mean)
rho.1Mb <- read.table(paste(p, "rho.1Mb.bedgraph", sep = ""), header = T)
theta.1Mb <- read.table(paste(p, "theta.1Mb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.1Mb$sim, theta.1Mb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.1Mb$sim and theta.1Mb$sample_mean
## S = 80, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9822024

inf.lands.1Mb.rep_8 <- as.data.frame(cbind(pi.1Mb$avg, theta.1Mb$sample_mean, rho.1Mb$sample_mean, tmrca.1Mb$avg))
names(inf.lands.1Mb.rep_8) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.1Mb.rep_8$thetaC <- inf.lands.1Mb.rep_8$theta - mean(inf.lands.1Mb.rep_8$theta)
inf.lands.1Mb.rep_8$tmrcaC <- inf.lands.1Mb.rep_8$tmrca - mean(inf.lands.1Mb.rep_8$tmrca)
inf.lands.1Mb.rep_8$rhoC <- inf.lands.1Mb.rep_8$rho - mean(inf.lands.1Mb.rep_8$rho)

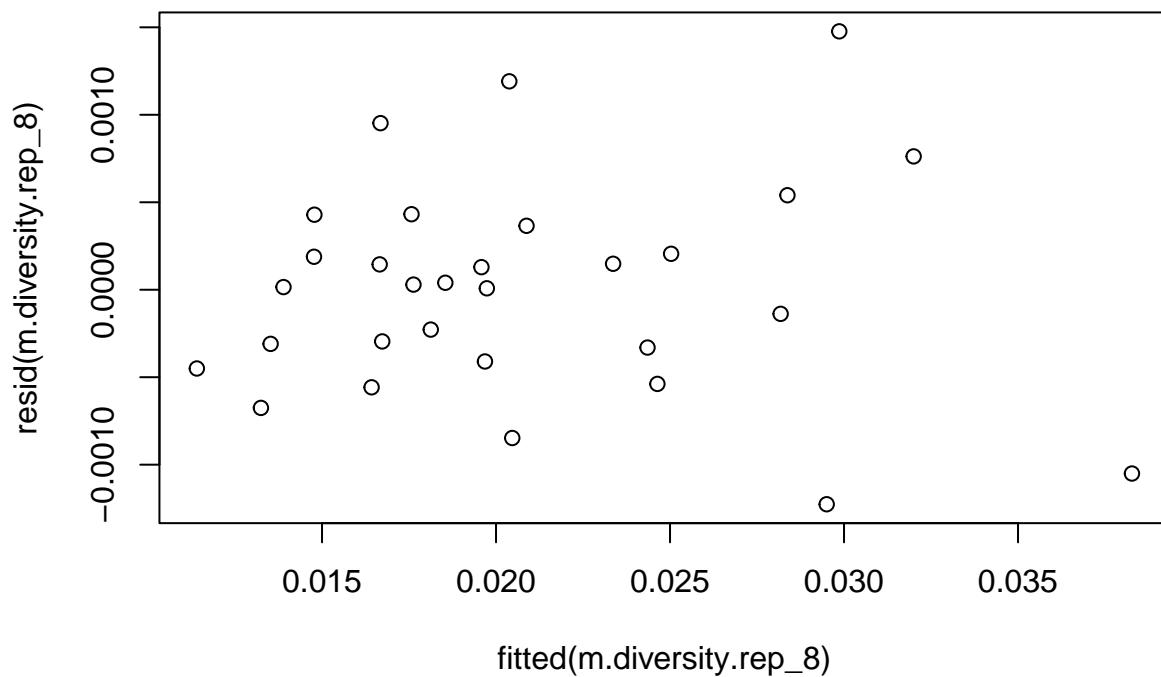
inf.lands.1Mb.rep_8$bin <- 1:nrow(inf.lands.1Mb.rep_8)

# for merging:
inf.lands.1Mb.rep_8$Replicate <- 8

m.diversity.rep_8 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.1Mb.rep_8)

plot(resid(m.diversity.rep_8)~fitted(m.diversity.rep_8))

```



```
dwtest(m.diversity.rep_8)
```

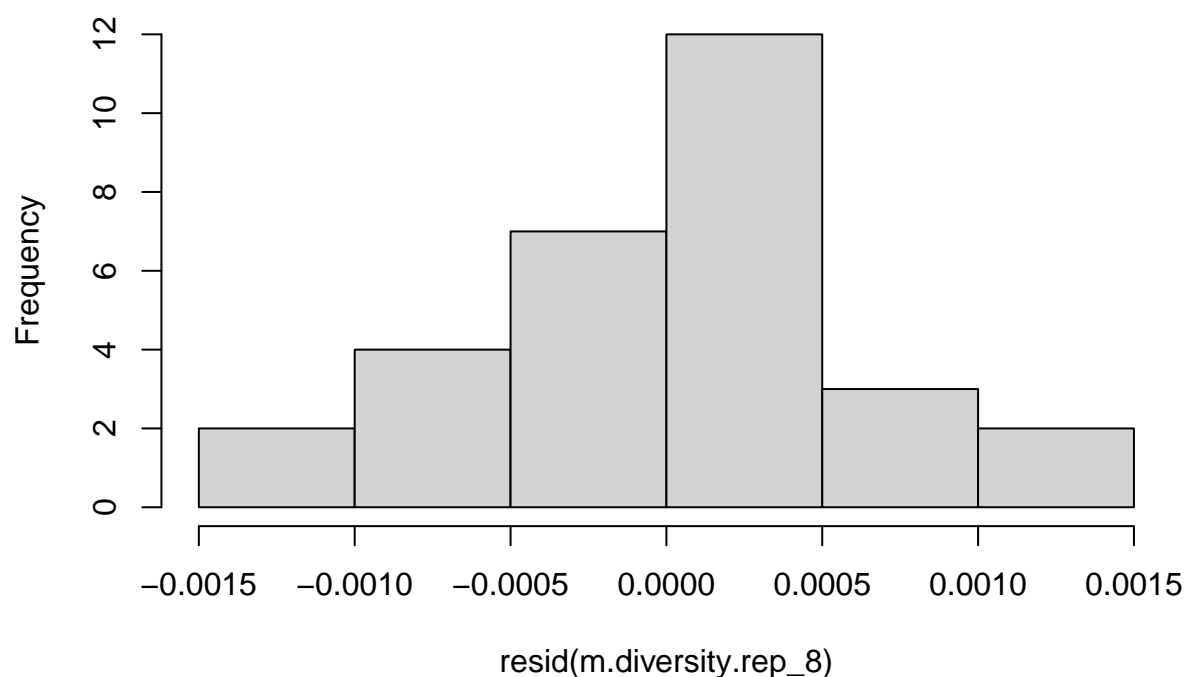
```
##
## Durbin-Watson test
##
## data: m.diversity.rep_8
## DW = 1.6816, p-value = 0.17
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_8)
```

```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_8
## HMC = 0.45613, p-value = 0.382
```

```
hist(resid(m.diversity.rep_8))
```

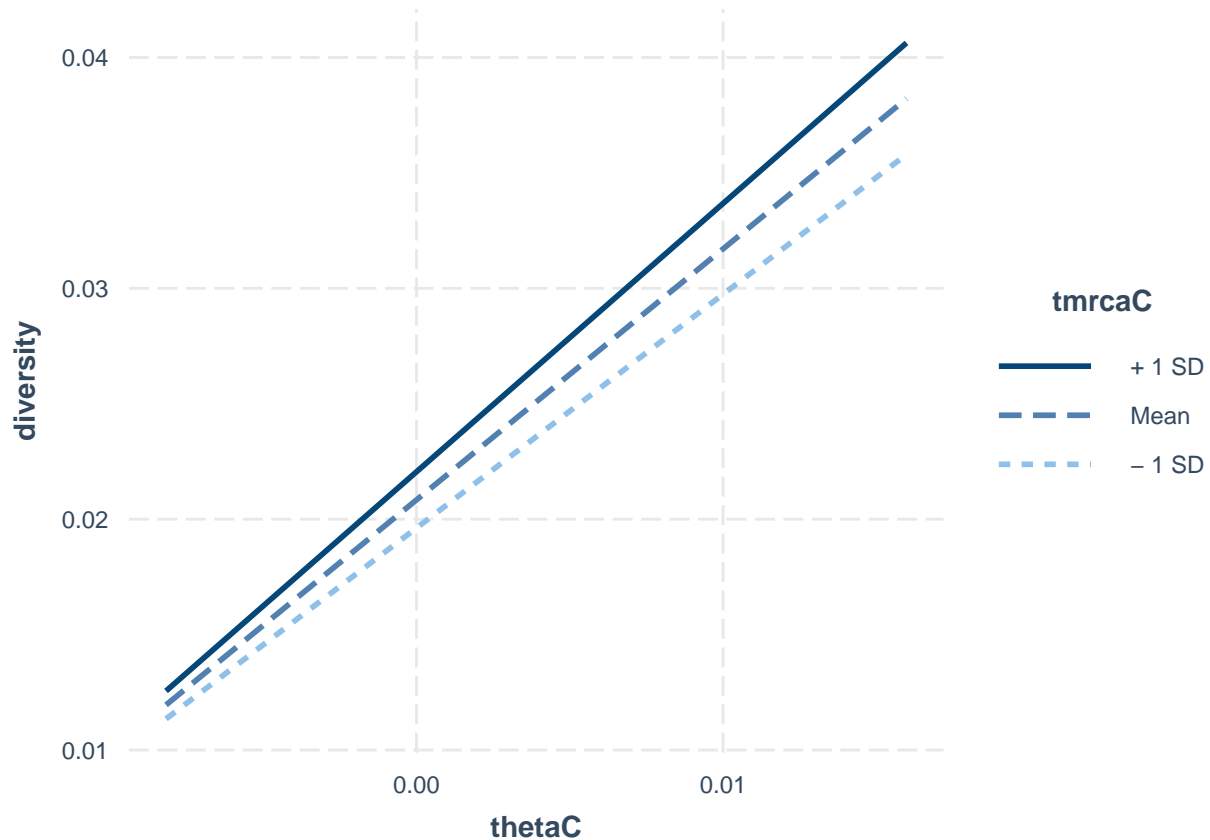
## Histogram of resid(m.diversity.rep\_8)



```
summary(m.diversity.rep_8)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.1Mb.rep_8)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.226e-03 -3.903e-04  2.222e-05  3.255e-04  1.477e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0208365  0.0001224  170.170 < 2e-16 ***
## thetaC       1.0871732  0.0214774   50.620 < 2e-16 ***
## rhoC        0.3643104  0.4059648    0.897  0.3781
## tmrcaC       0.0197018  0.0033310    5.915 3.58e-06 ***
## thetaC:tmrcaC 1.2143037  0.5213795    2.329  0.0282 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0006676 on 25 degrees of freedom
## Multiple R-squared:  0.9906, Adjusted R-squared:  0.9891
## F-statistic: 660.6 on 4 and 25 DF, p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_8, pred = thetaC, modx= tmrcaC)
```



```
g.rep_8 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.1Mb.rep_8, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_8)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.1Mb.rep_8
##      AIC      BIC   logLik
## -345.8378 -336.0294 179.9189
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.1655662
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  0.0208387 0.0001435 145.21451  0.0000
## thetaC       1.0841156 0.0213728  50.72412  0.0000
## tmrcaC       0.0201746 0.0030896   6.52983  0.0000
## rhoC        0.3427329 0.4152276   0.82541  0.4169
## thetaC:tmrcaC 1.3177433 0.4885467   2.69727  0.0123
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.012
## tmrcaC       0.046 -0.065
## rhoC        -0.024  0.082  0.056
## thetaC:tmrcaC 0.071 -0.140  0.721 -0.279
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.96269837 -0.63215676  0.03038078  0.52605973  2.37801009
##
## Residual standard error: 0.0006095428
## Degrees of freedom: 30 total; 25 residual

vif(g.rep_8)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.023443      2.452808      1.275892      2.678721

g.rep_8.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.1Mb.rep_8, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_8.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.1Mb.rep_8
##      AIC      BIC    logLik
## -319.0793 -312.0733 164.5397
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## -0.1415906
##
## Coefficients:
##      Value Std.Error   t-value p-value
## (Intercept)  0.0208035 0.0001699 122.42339  0.0000
## thetaC      1.0823875 0.0328438  32.95562  0.0000
## rhoC      -0.5688595 0.5260339  -1.08141  0.2891
##
## Correlation:
##      (Intr) thetaC
## thetaC  0.001
## rhoC    0.001 -0.012
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.87467372 -0.74418645  0.07635072  0.53300754  2.76001890
##
## Residual standard error: 0.001014048
## Degrees of freedom: 30 total; 27 residual
```

```

anova.diversity <- Anova(m.diversity.rep_8)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.1Mb[1, 8] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.1Mb[2, 8] <- anova.diversity$VarExp[1] * 100
r2.inf.1Mb[3, 8] <- anova.diversity$VarExp[2] * 100
r2.inf.1Mb[4, 8] <- anova.diversity$VarExp[3] * 100
r2.inf.1Mb[5, 8] <- anova.diversity$VarExp[4] * 100

```

### 2.3.9 Replicate 9

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_9/rs.pair_9.", sep = "")
pi.1Mb <- read.table(paste(p, "diversity.1Mb.bedgraph", sep = ""), header = T)
pi.1Mb$avg <- apply(pi.1Mb[4:ncol(pi.1Mb)], 1, mean)
tmrca.1Mb <- read.table(paste(p, "TMRCA.1Mb.bedgraph", sep = ""), header = T)
tmrca.1Mb$avg <- apply(tmrca.1Mb[4:ncol(tmrca.1Mb)], 1, mean)
rho.1Mb <- read.table(paste(p, "rho.1Mb.bedgraph", sep = ""), header = T)
theta.1Mb <- read.table(paste(p, "theta.1Mb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.1Mb$sim, theta.1Mb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.1Mb$sim and theta.1Mb$sample_mean
## S = 84, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9813126

inf.lands.1Mb.rep_9 <- as.data.frame(cbind(pi.1Mb$avg, theta.1Mb$sample_mean, rho.1Mb$sample_mean, tmrca.1Mb$avg))
names(inf.lands.1Mb.rep_9) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.1Mb.rep_9$thetaC <- inf.lands.1Mb.rep_9$theta - mean(inf.lands.1Mb.rep_9$theta)
inf.lands.1Mb.rep_9$tmrcaC <- inf.lands.1Mb.rep_9$tmrca - mean(inf.lands.1Mb.rep_9$tmrca)
inf.lands.1Mb.rep_9$rhoC <- inf.lands.1Mb.rep_9$rho - mean(inf.lands.1Mb.rep_9$rho)

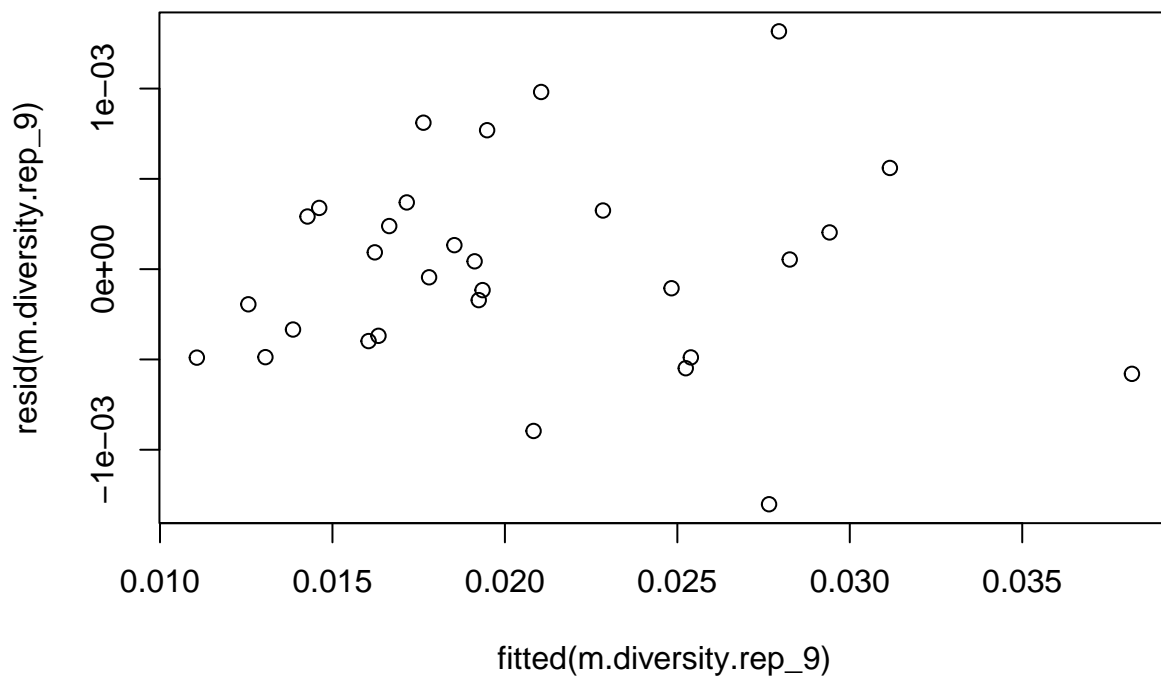
inf.lands.1Mb.rep_9$bin <- 1:nrow(inf.lands.1Mb.rep_9)

# for merging:
inf.lands.1Mb.rep_9$Replicate <- 9

m.diversity.rep_9 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.1Mb.rep_9)

plot(resid(m.diversity.rep_9)~fitted(m.diversity.rep_9))

```



```
dwtest(m.diversity.rep_9)
```

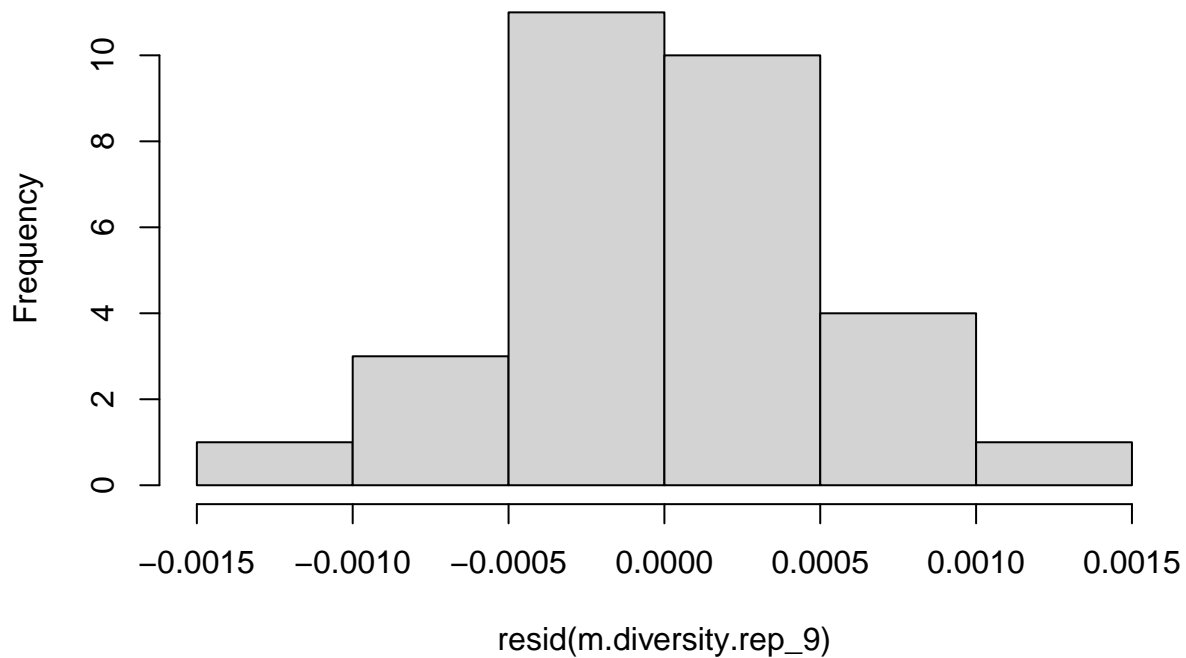
```
##
## Durbin-Watson test
##
## data: m.diversity.rep_9
## DW = 1.686, p-value = 0.1501
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_9)
```

```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_9
## HMC = 0.49123, p-value = 0.49
```

```
hist(resid(m.diversity.rep_9))
```

## Histogram of resid(m.diversity.rep\_9)

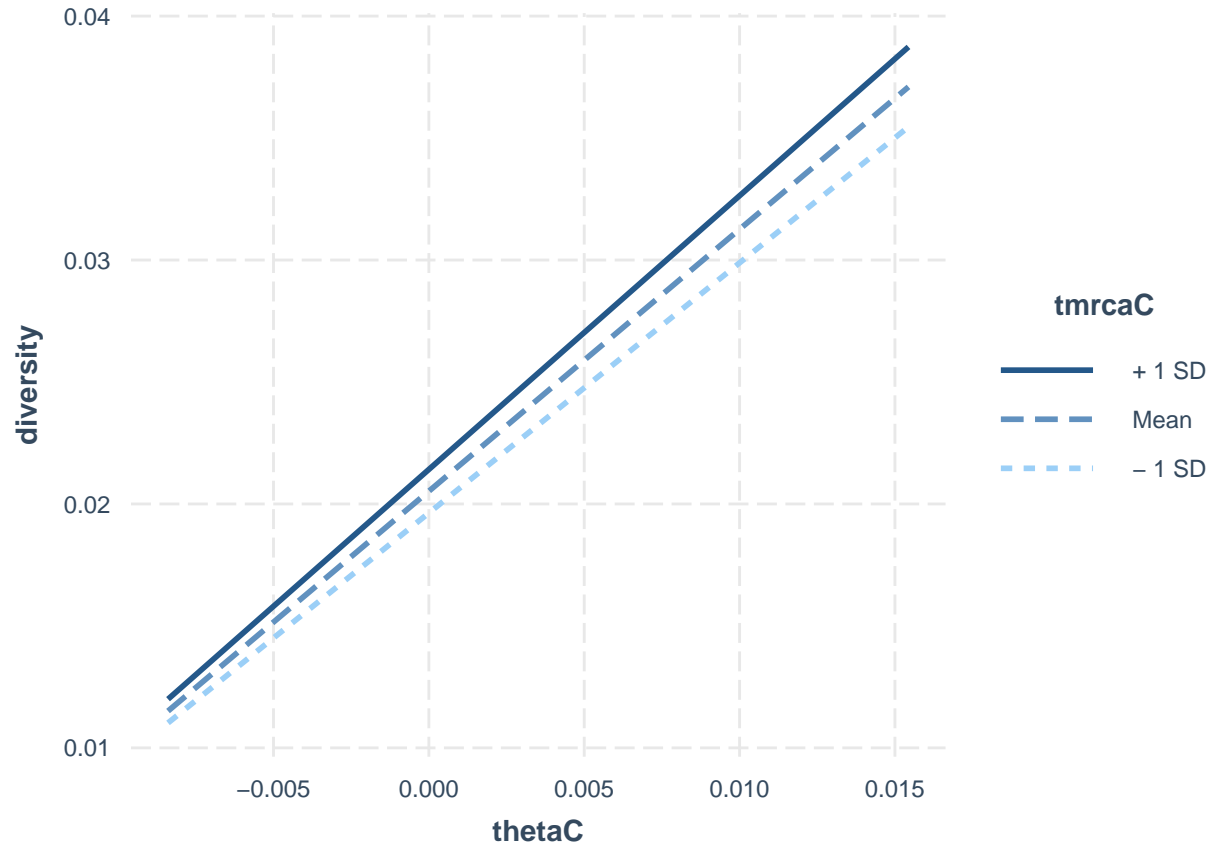


```
summary(m.diversity.rep_9)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.1Mb.rep_9)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.302e-03 -3.910e-04 -8.900e-07  3.164e-04  1.317e-03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0205233  0.0001107  185.445 < 2e-16 ***
## thetaC       1.0737054  0.0195206   55.004 < 2e-16 ***
## rhoC         0.1883059  0.3756234    0.501  0.6205
## tmrcaC        0.0183079  0.0027965    6.547 7.39e-07 ***
## thetaC:tmrcaC 0.9925142  0.4598859    2.158  0.0407 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0006061 on 25 degrees of freedom
## Multiple R-squared:  0.9922, Adjusted R-squared:  0.991
## F-statistic: 799.9 on 4 and 25 DF,  p-value: < 2.2e-16
```



```
interact_plot(m.diversity.rep_9, pred = thetaC, modx= tmrcaC)
```



```
g.rep_9 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
               data = inf.lands.1Mb.rep_9, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_9)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.1Mb.rep_9
##      AIC      BIC   logLik
## -351.6229 -341.8146 182.8115
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.1633414
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  0.0205240 0.0001297 158.24435  0.0000
## thetaC       1.0696841 0.0194352  55.03845  0.0000
## tmrcaC       0.0185521 0.0026328   7.04662  0.0000
## rhoC        0.2186371 0.3777124   0.57885  0.5679
## thetaC:tmrcaC 1.0045124 0.4613387   2.17739  0.0391
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      0.004
## tmrcaC     -0.021 -0.112
## rhoC        0.002  0.012  0.321
## thetaC:tmrcaC -0.022 -0.171  0.296 -0.217
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.277428606 -0.752076800 -0.002716583  0.581067343  2.423831154
##
## Residual standard error: 0.0005533147
## Degrees of freedom: 30 total; 25 residual

vif(g.rep_9)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.034584      1.327310      1.266322      1.268614

g.rep_9.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                        data = inf.lands.1Mb.rep_9, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_9.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.1Mb.rep_9
##      AIC      BIC      logLik
## -325.8718 -318.8658 167.9359
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## -0.2006686
##
## Coefficients:
##      Value Std.Error   t-value p-value
## (Intercept)  0.0205144 0.0001444 142.03128  0.0000
## thetaC      1.0817758 0.0289649  37.34776  0.0000
## rhoC       -0.9433213 0.4654231  -2.02680  0.0527
##
## Correlation:
##      (Intr) thetaC
## thetaC  0.001
## rhoC   -0.004 -0.041
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.57578335 -0.68638049 -0.02735148  0.65238571  1.96492871
##
## Residual standard error: 0.0009146795
## Degrees of freedom: 30 total; 27 residual
```

```

anova.diversity <- Anova(m.diversity.rep_9)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.1Mb[1, 9] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.1Mb[2, 9] <- anova.diversity$VarExp[1] * 100
r2.inf.1Mb[3, 9] <- anova.diversity$VarExp[2] * 100
r2.inf.1Mb[4, 9] <- anova.diversity$VarExp[3] * 100
r2.inf.1Mb[5, 9] <- anova.diversity$VarExp[4] * 100

```

### 2.3.10 Replicate 10

```

p <- paste("dm_data/dm_chr_maps/2L/dm_coal_sims/rep_10/rs.pair_10.", sep = "")
pi.1Mb <- read.table(paste(p, "diversity.1Mb.bedgraph", sep = ""), header = T)
pi.1Mb$avg <- apply(pi.1Mb[4:ncol(pi.1Mb)], 1, mean)
tmrca.1Mb <- read.table(paste(p, "TMRCA.1Mb.bedgraph", sep = ""), header = T)
tmrca.1Mb$avg <- apply(tmrca.1Mb[4:ncol(tmrca.1Mb)], 1, mean)
rho.1Mb <- read.table(paste(p, "rho.1Mb.bedgraph", sep = ""), header = T)
theta.1Mb <- read.table(paste(p, "theta.1Mb.bedgraph", sep = ""), header = T)

cor.test(sim.theta.1Mb$sim, theta.1Mb$sample_mean, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: sim.theta.1Mb$sim and theta.1Mb$sample_mean
## S = 70, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9844271

inf.lands.1Mb.rep_10 <- as.data.frame(cbind(pi.1Mb$avg, theta.1Mb$sample_mean, rho.1Mb$sample_mean, tmrca.1Mb$avg))
names(inf.lands.1Mb.rep_10) <- c("diversity", "theta", "rho", "tmrca")

# centering
inf.lands.1Mb.rep_10$thetaC <- inf.lands.1Mb.rep_10$theta - mean(inf.lands.1Mb.rep_10$theta)
inf.lands.1Mb.rep_10$tmrcaC <- inf.lands.1Mb.rep_10$tmrca - mean(inf.lands.1Mb.rep_10$tmrca)
inf.lands.1Mb.rep_10$rhoC <- inf.lands.1Mb.rep_10$rho - mean(inf.lands.1Mb.rep_10$rho)

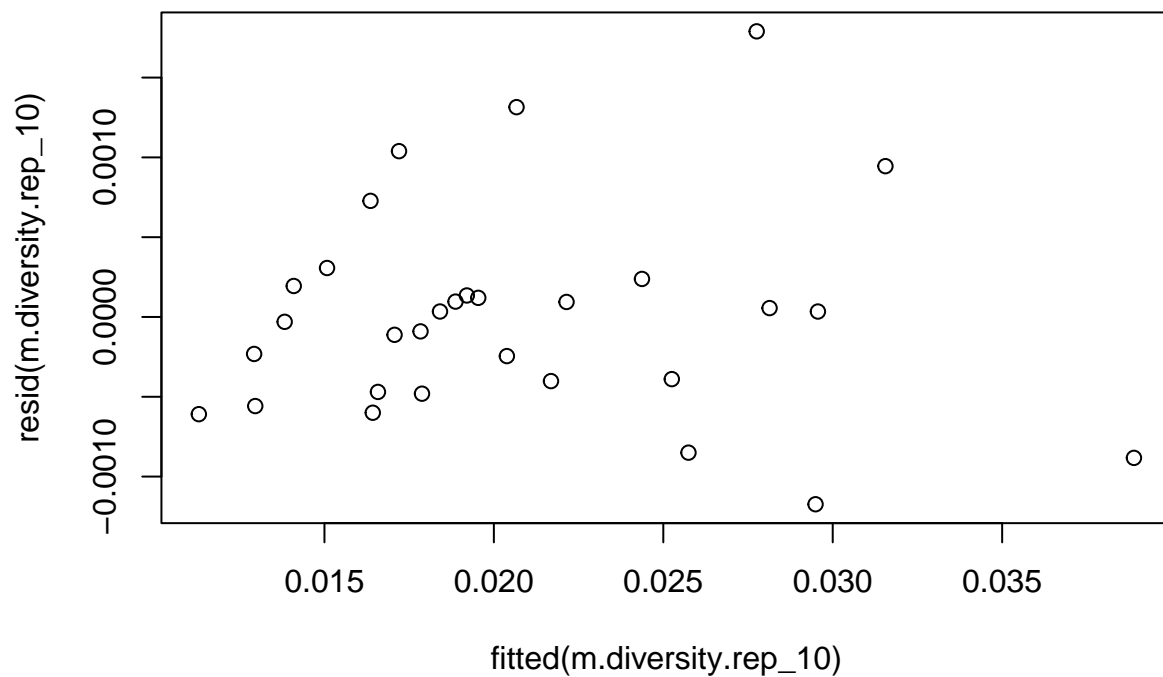
inf.lands.1Mb.rep_10$bin <- 1:nrow(inf.lands.1Mb.rep_10)

# for merging:
inf.lands.1Mb.rep_10$Replicate <- 10

m.diversity.rep_10 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.1Mb.rep_10)

plot(resid(m.diversity.rep_10)~fitted(m.diversity.rep_10))

```



```
dwtest(m.diversity.rep_10)
```

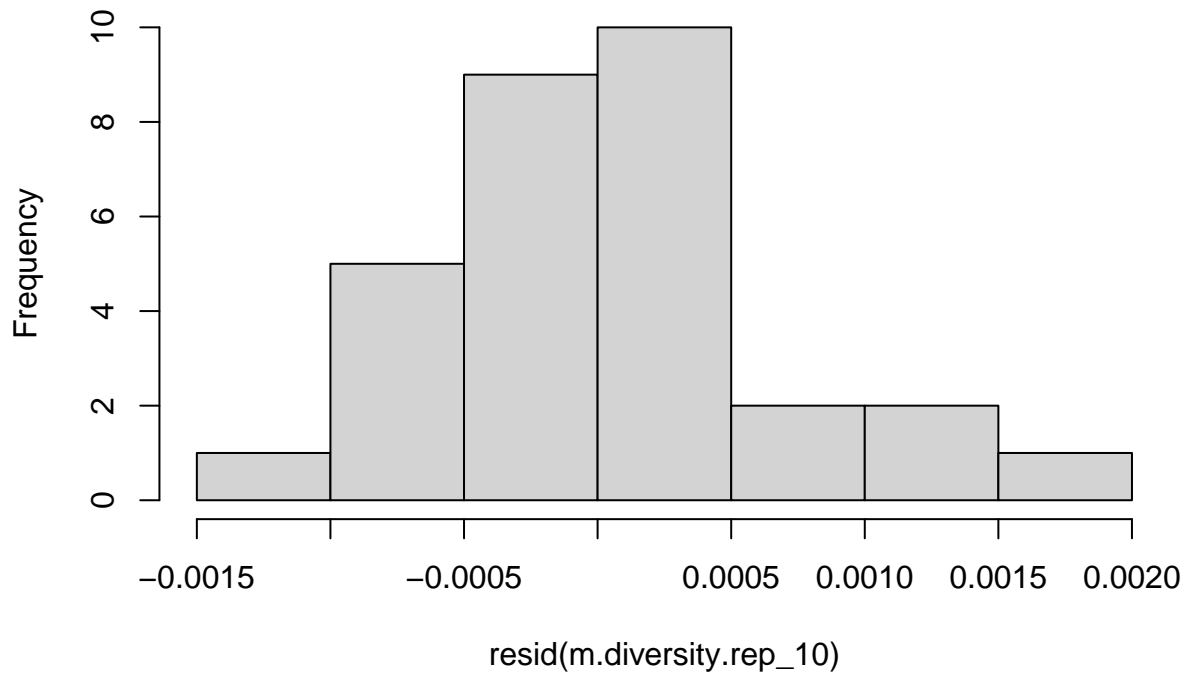
```
##
## Durbin-Watson test
##
## data: m.diversity.rep_10
## DW = 1.5209, p-value = 0.06939
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.diversity.rep_10)
```

```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_10
## HMC = 0.59687, p-value = 0.778
```

```
hist(resid(m.diversity.rep_10))
```

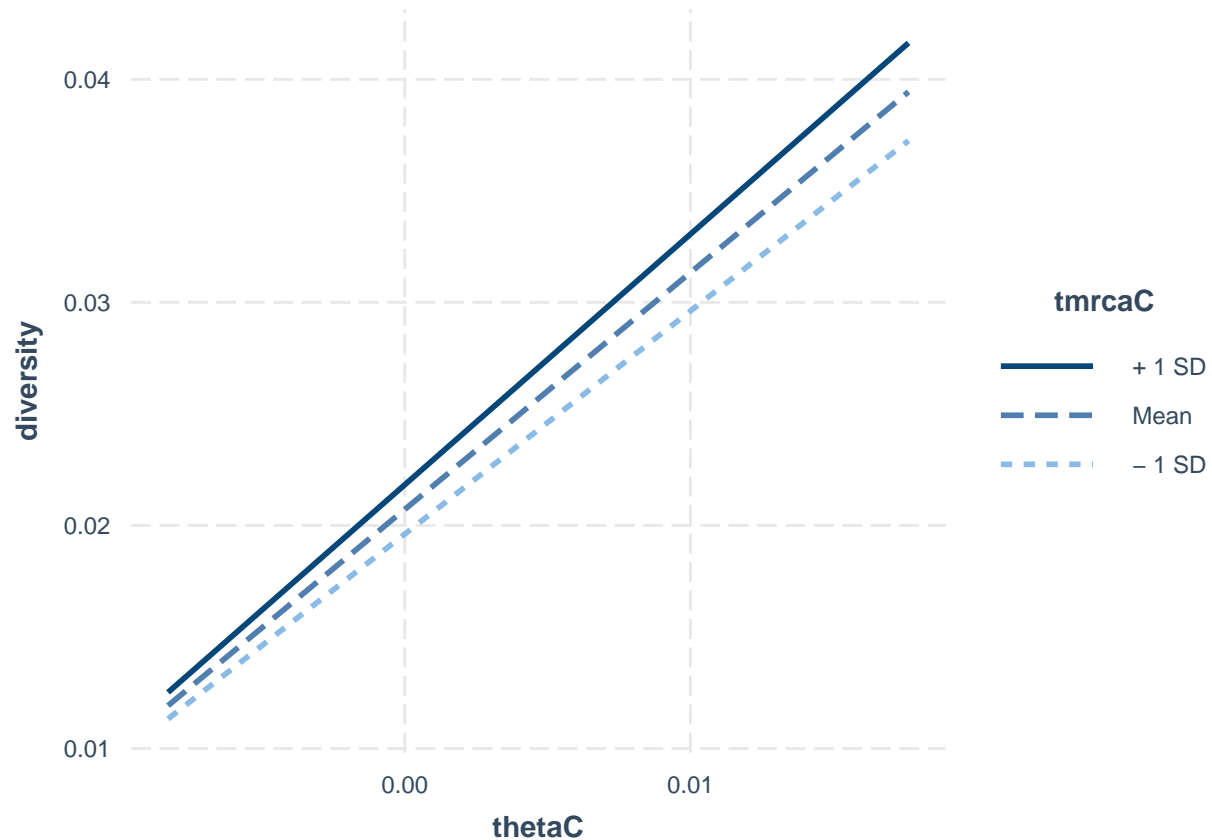
## Histogram of resid(m.diversity.rep\_10)



```
summary(m.diversity.rep_10)
```

```
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##     data = inf.lands.1Mb.rep_10)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0011734 -0.0004527  0.0000019  0.0001794  0.0017898
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0207164  0.0001292  160.343 < 2e-16 ***
## thetaC       1.0619903  0.0218331   48.641 < 2e-16 ***
## rhoC         0.0656389  0.4323685    0.152  0.8806
## tmrcaC       0.0189583  0.0031915    5.940 3.35e-06 ***
## thetaC:tmrcaC 1.0582401  0.5049627    2.096  0.0464 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0007071 on 25 degrees of freedom
## Multiple R-squared:  0.9898, Adjusted R-squared:  0.9882
## F-statistic: 605.8 on 4 and 25 DF,  p-value: < 2.2e-16
```

```
interact_plot(m.diversity.rep_10, pred = thetaC, modx= tmrcaC)
```



```
g.rep_10 <- gls(diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC,
  data = inf.lands.1Mb.rep_1, cor = corAR1(0, ~bin), method = "ML")
```

```
summary(g.rep_10)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + tmrcaC + rhoC + thetaC:tmrcaC
## Data: inf.lands.1Mb.rep_1
##      AIC      BIC logLik
## -348.56 -338.7516 181.28
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.0188108
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  0.0206248 0.0001171 176.16584  0.0000
## thetaC       1.0920063 0.0211898  51.53453  0.0000
## tmrcaC       0.0168279 0.0029670   5.67164  0.0000
## rhoC         0.0007886 0.3523958   0.00224  0.9982
## thetaC:tmrcaC 0.8389718 0.4973907   1.68675  0.1041
```

```
##
## Correlation:
##      (Intr) thetaC tmrcaC rhoC
## thetaC      -0.001
## tmrcaC       0.001 -0.171
## rhoC         0.000 -0.021  0.376
## thetaC:tmrcaC 0.004 -0.283  0.575 -0.003
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.67095844 -0.56781078 -0.01108315  0.40839933  3.09944720
##
## Residual standard error: 0.0005748314
## Degrees of freedom: 30 total; 25 residual

vif(g.rep_10)

##      thetaC      tmrcaC      rhoC thetaC:tmrcaC
##      1.087501      1.898249      1.271394      1.717045

g.rep_10.no.tmrca <- gls(diversity ~ thetaC + rhoC,
                          data = inf.lands.1Mb.rep_10, cor = corAR1(0, ~bin), method = "ML")

summary(g.rep_10.no.tmrca)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ thetaC + rhoC
## Data: inf.lands.1Mb.rep_10
##      AIC      BIC    logLik
## -316.9544 -309.9484 163.4772
##
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## -0.01181333
##
## Coefficients:
##      Value Std.Error   t-value p-value
## (Intercept)  0.020705 0.0001980 104.59296  0.0000
## thetaC      1.062802 0.0337999  31.44392  0.0000
## rhoC       -0.727976 0.5917022  -1.23031  0.2292
##
## Correlation:
##      (Intr) thetaC
## thetaC  0.000
## rhoC    0.000 -0.106
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.05588562 -0.66664272 -0.06688445  0.53908458  1.87562405
##
## Residual standard error: 0.001040441
## Degrees of freedom: 30 total; 27 residual
```

```

anova.diversity <- Anova(m.diversity.rep_10)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.inf.1Mb[1, 10] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.inf.1Mb[2, 10] <- anova.diversity$VarExp[1] * 100
r2.inf.1Mb[3, 10] <- anova.diversity$VarExp[2] * 100
r2.inf.1Mb[4, 10] <- anova.diversity$VarExp[3] * 100
r2.inf.1Mb[5, 10] <- anova.diversity$VarExp[4] * 100

r2.inf.1Mb$average <- rowMeans(r2.inf.1Mb)
r2.inf.1Mb <- transform(r2.inf.1Mb, sd=apply(r2.inf.1Mb, 1, sd, na.rm = TRUE))

```

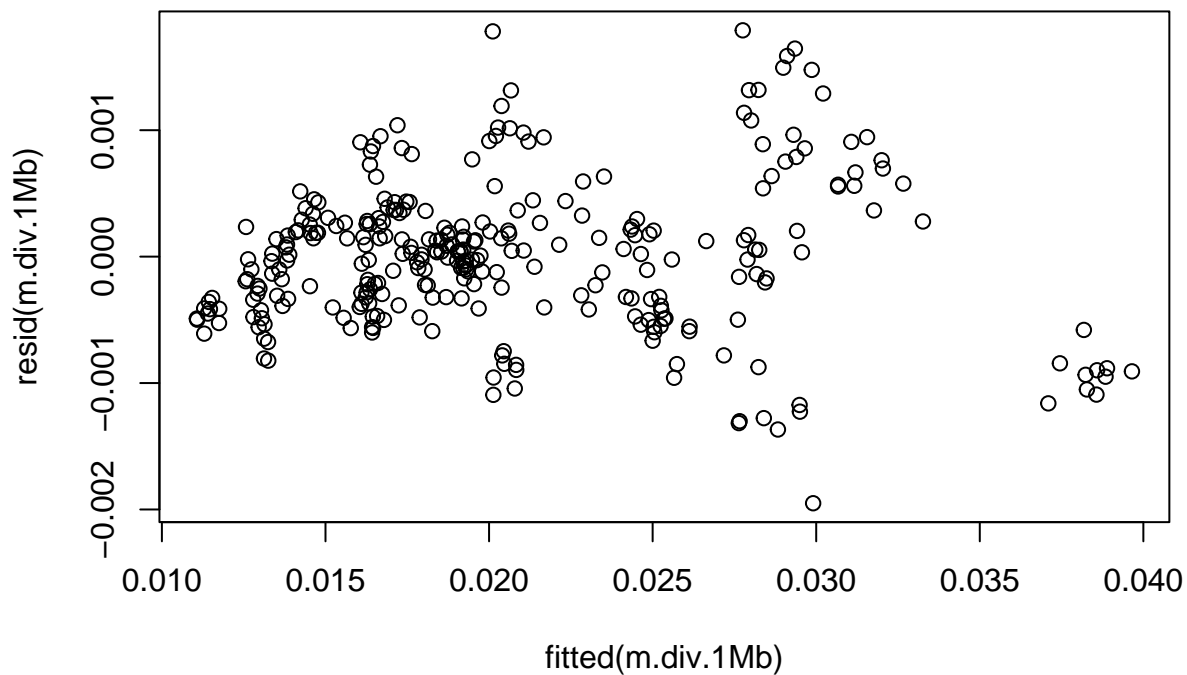
### 2.3.11 all replicates:

```

inf.lands.1Mb.all <- rbind(inf.lands.1Mb.rep_1, inf.lands.1Mb.rep_2, inf.lands.1Mb.rep_3, inf.lands.1Mb.rep_4)

m.div.1Mb <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)*as.factor(Replicate), data = inf.lands.1Mb.all)
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```
dwtest(m.div.1Mb)
```

```

##
## Durbin-Watson test
##
## data: m.div.1Mb

```

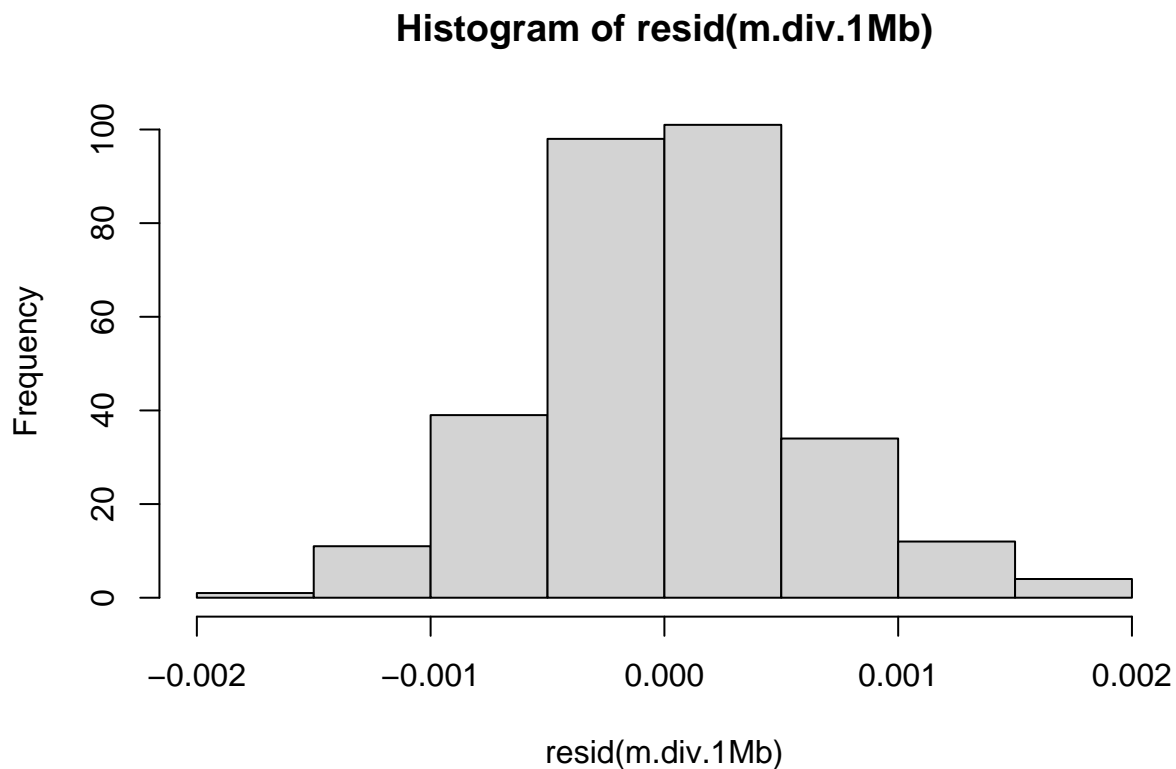


```
## DW = 1.7602, p-value = 0.001541
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmctest(m.div.1Mb)
```

```
##
## Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.51208, p-value = 0.602
```

```
hist(resid(m.div.1Mb))
```



```
m.div.1Mb.2 <- lm( diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC)*as.factor(Replicate)
m.div.1Mb.3 <- lm( diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC + rhoC:tmrcaC)*as.f
```

```
AIC(m.div.1Mb, m.div.1Mb.2, m.div.1Mb.3)
```

```
##           df      AIC
## m.div.1Mb  51 -3507.758
## m.div.1Mb.2 61 -3522.377
## m.div.1Mb.3 71 -3504.645
```

```
summary(m.div.1Mb)
```

```
##
## Call:
## lm(formula = diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC) *
##     as.factor(Replicate), data = inf.lands.1Mb.all)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0019505 -0.0003895  0.0000113  0.0002755  0.0017898
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.062e-02  1.180e-04 174.750 < 2e-16 ***
## thetaC           1.092e+00  2.169e-02  50.357 < 2e-16 ***
## rhoC             3.099e-03  3.609e-01   0.009   0.993
## tmrcaC           1.670e-02  3.065e-03   5.451 1.2e-07 ***
## as.factor(Replicate)2      1.783e-05  1.679e-04   0.106   0.915
## as.factor(Replicate)3      6.531e-05  1.674e-04   0.390   0.697
## as.factor(Replicate)4      5.942e-05  1.669e-04   0.356   0.722
## as.factor(Replicate)5     -3.449e-05  1.671e-04  -0.206   0.837
## as.factor(Replicate)6     -3.480e-05  1.670e-04  -0.208   0.835
## as.factor(Replicate)7     -2.047e-05  1.671e-04  -0.123   0.903
## as.factor(Replicate)8      2.120e-04  1.673e-04   1.267   0.206
## as.factor(Replicate)9     -1.013e-04  1.669e-04  -0.607   0.545
## as.factor(Replicate)10     9.189e-05  1.670e-04   0.550   0.583
## thetaC:tmrcaC           8.193e-01  5.114e-01   1.602   0.110
## thetaC:as.factor(Replicate)2 -1.466e-02  3.004e-02  -0.488   0.626
## thetaC:as.factor(Replicate)3 -1.976e-02  3.022e-02  -0.654   0.514
## thetaC:as.factor(Replicate)4   2.129e-02  2.996e-02   0.711   0.478
## thetaC:as.factor(Replicate)5   1.538e-02  3.032e-02   0.507   0.612
## thetaC:as.factor(Replicate)6  -4.900e-03  3.021e-02  -0.162   0.871
## thetaC:as.factor(Replicate)7  -1.451e-02  2.995e-02  -0.485   0.628
## thetaC:as.factor(Replicate)8  -4.950e-03  3.005e-02  -0.165   0.869
## thetaC:as.factor(Replicate)9  -1.842e-02  3.006e-02  -0.613   0.541
## thetaC:as.factor(Replicate)10 -3.013e-02  2.947e-02  -1.022   0.308
## rhoC:as.factor(Replicate)2   3.444e-01  5.987e-01   0.575   0.566
## rhoC:as.factor(Replicate)3   1.470e-01  5.241e-01   0.280   0.779
## rhoC:as.factor(Replicate)4  -3.706e-02  5.547e-01  -0.067   0.947
## rhoC:as.factor(Replicate)5  -1.323e-01  5.422e-01  -0.244   0.807
## rhoC:as.factor(Replicate)6   3.968e-01  5.139e-01   0.772   0.441
## rhoC:as.factor(Replicate)7   6.492e-02  5.371e-01   0.121   0.904
## rhoC:as.factor(Replicate)8   3.612e-01  5.336e-01   0.677   0.499
## rhoC:as.factor(Replicate)9   1.852e-01  5.392e-01   0.343   0.732
## rhoC:as.factor(Replicate)10  6.254e-02  5.352e-01   0.117   0.907
## tmrcaC:as.factor(Replicate)2  2.835e-03  4.116e-03   0.689   0.492
## tmrcaC:as.factor(Replicate)3  4.229e-03  3.968e-03   1.066   0.287
## tmrcaC:as.factor(Replicate)4  1.151e-03  4.332e-03   0.266   0.791
## tmrcaC:as.factor(Replicate)5  7.096e-04  4.463e-03   0.159   0.874
## tmrcaC:as.factor(Replicate)6  2.969e-03  4.138e-03   0.718   0.474
## tmrcaC:as.factor(Replicate)7  7.158e-03  4.497e-03   1.592   0.113
## tmrcaC:as.factor(Replicate)8  2.998e-03  4.449e-03   0.674   0.501
## tmrcaC:as.factor(Replicate)9  1.604e-03  4.277e-03   0.375   0.708
## tmrcaC:as.factor(Replicate)10 2.254e-03  4.231e-03   0.533   0.595
## thetaC:tmrcaC:as.factor(Replicate)2 5.699e-01  7.355e-01   0.775   0.439
## thetaC:tmrcaC:as.factor(Replicate)3 1.003e+00  7.144e-01   1.404   0.162
## thetaC:tmrcaC:as.factor(Replicate)4 2.454e-01  8.165e-01   0.301   0.764
## thetaC:tmrcaC:as.factor(Replicate)5 6.558e-01  8.117e-01   0.808   0.420
## thetaC:tmrcaC:as.factor(Replicate)6 3.519e-01  7.102e-01   0.495   0.621
## thetaC:tmrcaC:as.factor(Replicate)7 1.209e+00  7.614e-01   1.588   0.114
```

```
## thetaC:tmrcaC:as.factor(Replicate)8 3.950e-01 7.186e-01 0.550 0.583
## thetaC:tmrcaC:as.factor(Replicate)9 1.732e-01 7.086e-01 0.244 0.807
## thetaC:tmrcaC:as.factor(Replicate)10 2.390e-01 6.889e-01 0.347 0.729
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0006464 on 250 degrees of freedom
## Multiple R-squared: 0.9913, Adjusted R-squared: 0.9896
## F-statistic: 582.5 on 49 and 250 DF, p-value: < 2.2e-16
```

### 2.3.12 Plots

```
scale.3d <- function(x) sprintf("%.3f", x)

rho.plot <- as.data.frame(cbind(inf.lands.1Mb.rep_1$bin,
                                sim.rho.1Mb$sim,
                                inf.lands.1Mb.rep_1$rho,
                                inf.lands.1Mb.rep_2$rho,
                                inf.lands.1Mb.rep_3$rho,
                                inf.lands.1Mb.rep_4$rho,
                                inf.lands.1Mb.rep_5$rho,
                                inf.lands.1Mb.rep_6$rho,
                                inf.lands.1Mb.rep_7$rho,
                                inf.lands.1Mb.rep_8$rho,
                                inf.lands.1Mb.rep_9$rho,
                                inf.lands.1Mb.rep_10$rho))

rho.plot[,3:ncol(rho.plot)] <- 1.53 * rho.plot[,3:ncol(rho.plot)] # scaling because of SCRM NO vs Ne

names(rho.plot) <- c("bin", "sim", reps)
molten.rho <- melt(rho.plot, id.vars = "bin")
rho.map.1Mb <- ggplot(data = molten.rho, aes(x = bin * 1000, y = value, colour = variable)) + theme_bw()
rho.map.1Mb <- rho.map.1Mb + geom_line(data = molten.rho, aes(size = variable)) + scale_size_manual(values = c(1, 2, 3, 4, 5, 6, 7, 8, 9))
rho.map.1Mb <- rho.map.1Mb + scale_color_manual(values = c("black", brewer.pal(n = 9, name = "Blues")))
rho.map.1Mb <- rho.map.1Mb + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
rho.map.1Mb <- rho.map.1Mb + labs(title = NULL, x = NULL, y = expression(rho))
rho.map.1Mb <- rho.map.1Mb + theme(text = element_text(size = 20), axis.title.x = element_text(size = 20))

theta.plot <- as.data.frame(cbind(inf.lands.1Mb.rep_1$bin,
                                sim.theta.1Mb$sim,
                                inf.lands.1Mb.rep_1$theta,
                                inf.lands.1Mb.rep_2$theta,
                                inf.lands.1Mb.rep_3$theta,
                                inf.lands.1Mb.rep_4$theta,
                                inf.lands.1Mb.rep_5$theta,
                                inf.lands.1Mb.rep_6$theta,
                                inf.lands.1Mb.rep_7$theta,
                                inf.lands.1Mb.rep_8$theta,
                                inf.lands.1Mb.rep_9$theta,
                                inf.lands.1Mb.rep_10$theta))

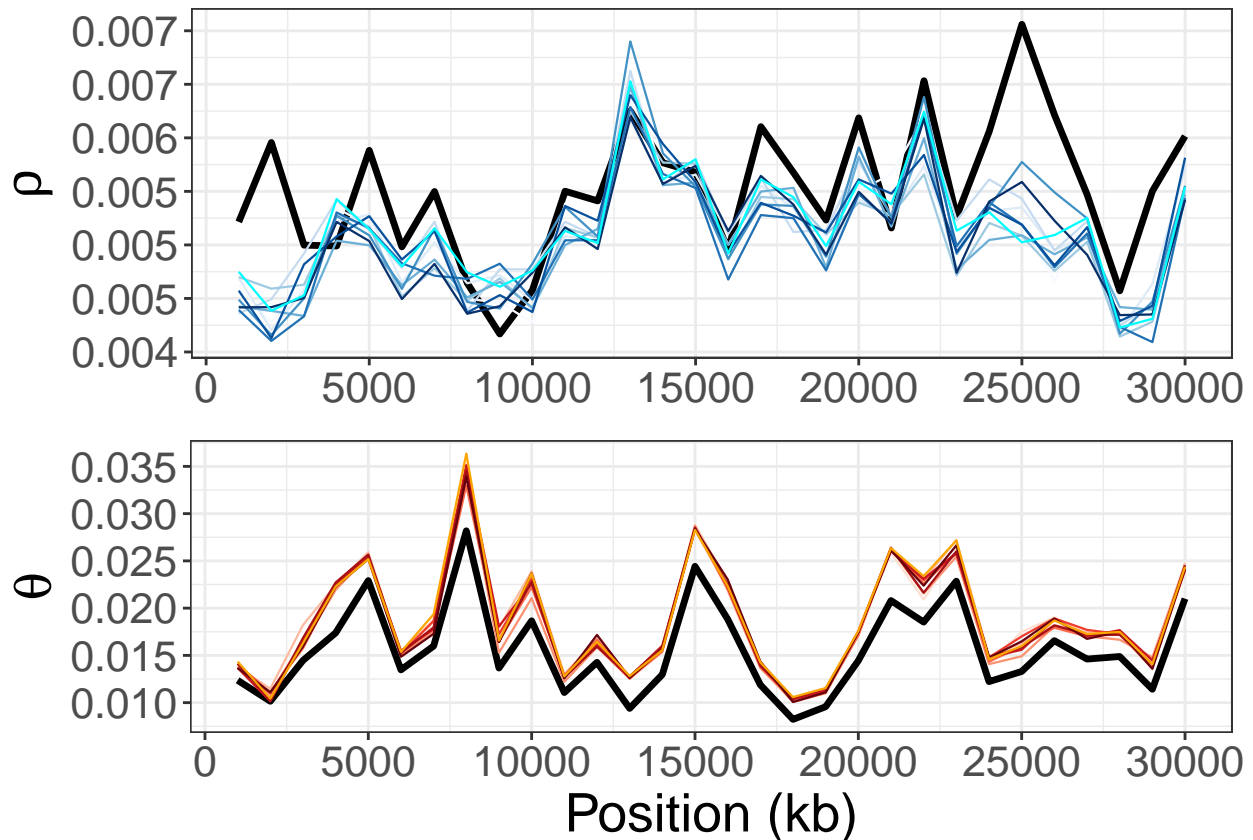
names(theta.plot) <- c("bin", "sim", reps)
molten.theta <- melt(theta.plot, id.vars = "bin")
theta.map.1Mb <- ggplot(data = molten.theta, aes(x = bin * 1000, y = value, colour = variable)) + theme_bw()
```

```

theta.map.1Mb <- theta.map.1Mb + geom_line(data = molten.theta, aes(size = variable)) + scale_size_manual()
theta.map.1Mb <- theta.map.1Mb + scale_color_manual(values = c("black", brewer.pal(n = 9, name = "Reds")))
theta.map.1Mb <- theta.map.1Mb + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
theta.map.1Mb <- theta.map.1Mb + labs(title = NULL, x = "Position (kb)", y = expression(theta))
theta.map.1Mb <- theta.map.1Mb + theme(text = element_text(size = 20), axis.title.x = element_text(size = 20))

plot_grid(rho.map.1Mb, theta.map.1Mb, nrow = 2, ncol = 1)

```



```

fig2 <- plot_grid(theta.map.50kb, theta.map.200kb, theta.map.1Mb, labels = "AUTO", nrow = 3, ncol = 1)
save_plot("fig2.pdf", fig2, base_height = 16, base_width = 12)

```

### 3 Real Drosophila data.

We first focus on chr 2L for comparing  $R^2$  with simulations

```

r2.dm.tab <- as.data.frame(matrix(ncol = 6, nrow = 3))
colnames(r2.dm.tab) <- c("Total", "Theta", "Rho", "TMRCA", "Theta:TMRCA", "bin.size")

# 50kb
# recombination landscapes
rho.dm.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.rho.50kb.bedgraph", header = T)

# diversity
diversity.dm.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.diversity.50kb.bedgraph", header = T)
diversity.dm.50kb$avg <- apply(diversity.dm.50kb[4:ncol(diversity.dm.50kb)], 1, mean)

```

```

# mutation landscapes
theta.dm.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.theta.50kb.bedgraph", header = T)

# TMRCA landscapes
tmrca.dm.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.TMRCA.50kb.bedgraph", header = T)
tmrca.dm.50kb$sample_mean <- apply(tmrca.dm.50kb[4:ncol(tmrca.dm.50kb)], 1, mean)

# missing data
missing.prop.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.missing.prop.50kb.bedgraph", header = T)
intersect.50kb <- apply(missing.prop.50kb[4:ncol(missing.prop.50kb)], 1, function(x) any(x > 0.1))

dm.lands.50kb <- as.data.frame(cbind(
  diversity.dm.50kb$avg,
  theta.dm.50kb$sample_mean,
  rho.dm.50kb$sample_mean,
  tmrca.dm.50kb$sample_mean))

names(dm.lands.50kb) <- c("diversity", "theta", "rho", "tmrca")
dm.lands.50kb$bin <- 1:nrow(dm.lands.50kb)

# filters based on missing data
dm.lands.50kb <- dm.lands.50kb[which(intersect.50kb == F),]

# OLS

# centering
dm.lands.50kb$thetaC <- dm.lands.50kb$theta - mean(dm.lands.50kb$theta)
dm.lands.50kb$tmrcaC <- dm.lands.50kb$tmrca - mean(dm.lands.50kb$tmrca)
dm.lands.50kb$rhoC <- dm.lands.50kb$rho - mean(dm.lands.50kb$rho)

m.diversity <- lm(diversity ~ thetaC + rhoC + tmrcaC + tmrcaC*thetaC, data = dm.lands.50kb)

# type 2 ANOVA
anova.diversity <- Anova(m.diversity)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.dm.tab[1,] <- c((anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] +
  anova.diversity$VarExp[1] * 100,
  anova.diversity$VarExp[2] * 100,
  anova.diversity$VarExp[3] * 100,
  anova.diversity$VarExp[4] * 100, 50))

# 200kb
# recombination landscapes
rho.dm.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.rho.200kb.bedgraph", header = T)

# diversity
diversity.dm.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.diversity.200kb.bedgraph", header = T)
diversity.dm.200kb$avg <- apply(diversity.dm.200kb[4:ncol(diversity.dm.200kb)], 1, mean)

# mutation landscapes
theta.dm.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.theta.200kb.bedgraph", header = T)

```

```

# TMRCA landscapes
tmrca.dm.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.TMRCA.200kb.bedgraph", header = T)
tmrca.dm.200kb$sample_mean <- apply(tmrca.dm.200kb[4:ncol(tmrca.dm.200kb)], 1, mean)

# missing data
missing.prop.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.missing.prop.200kb.bedgraph", header = T)
intersect.200kb <- apply(missing.prop.200kb[4:ncol(missing.prop.200kb)], 1, function(x) any(x > 0.1))

dm.lands.200kb <- as.data.frame(cbind(diversity.dm.200kb$avg,
                                     theta.dm.200kb$sample_mean,
                                     rho.dm.200kb$sample_mean,
                                     tmrca.dm.200kb$sample_mean))

names(dm.lands.200kb) <- c("diversity", "theta", "rho", "tmrca")
dm.lands.200kb$bin <- 1:nrow(dm.lands.200kb)
# filters based on missing data
dm.lands.200kb <- dm.lands.200kb[which(intersect.200kb == F),]

# OLS

dm.lands.200kb$thetaC <- dm.lands.200kb$theta - mean(dm.lands.200kb$theta)
dm.lands.200kb$tmrcaC <- dm.lands.200kb$tmrca - mean(dm.lands.200kb$tmrca)
dm.lands.200kb$rhoC <- dm.lands.200kb$rho - mean(dm.lands.200kb$rho)

m.diversity <- lm(diversity ~ thetaC + rhoC + tmrcaC + tmrcaC*thetaC, data = dm.lands.200kb)

# type 2 ANOVA
anova.diversity <- Anova(m.diversity)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.dm.tab[2,] <- c((anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] +
                  anova.diversity$VarExp[4] * 100,
                  anova.diversity$VarExp[2] * 100,
                  anova.diversity$VarExp[3] * 100,
                  anova.diversity$VarExp[4] * 100, 200))

# 1Mb
# recombination landscapes
rho.dm.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.rho.1Mb.bedgraph", header = T)

# diversity
diversity.dm.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.diversity.1Mb.bedgraph", header = T)
diversity.dm.1Mb$avg <- apply(diversity.dm.1Mb[4:ncol(diversity.dm.1Mb)], 1, mean)

# mutation landscapes
theta.dm.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.theta.1Mb.bedgraph", header = T)

# TMRCA landscapes
tmrca.dm.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.TMRCA.1Mb.bedgraph", header = T)
tmrca.dm.1Mb$sample_mean <- apply(tmrca.dm.1Mb[4:ncol(tmrca.dm.1Mb)], 1, mean)

# missing data

```

```

missing.prop.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.missing.prop.1Mb.bedgraph", header = T)
intersect.1Mb <- apply(missing.prop.1Mb[4:ncol(missing.prop.1Mb)], 1, function(x) any(x > 0.1))

dm.lands.1Mb <- as.data.frame(cbind(diversity.dm.1Mb$avg,
                                   theta.dm.1Mb$sample_mean,
                                   rho.dm.1Mb$sample_mean,
                                   tmrca.dm.1Mb$sample_mean))

names(dm.lands.1Mb) <- c("diversity", "theta", "rho", "tmrca")
dm.lands.1Mb$bin <- 1:nrow(dm.lands.1Mb)
# filters based on missing data
dm.lands.1Mb <- dm.lands.1Mb[which(intersect.1Mb == F),]

# OLS
dm.lands.1Mb$thetaC <- dm.lands.1Mb$theta - mean(dm.lands.1Mb$theta)
dm.lands.1Mb$tmrcaC <- dm.lands.1Mb$tmrca - mean(dm.lands.1Mb$tmrca)
dm.lands.1Mb$rhoC <- dm.lands.1Mb$rho - mean(dm.lands.1Mb$rho)

m.diversity <- lm(diversity ~ thetaC + rhoC + tmrcaC + tmrcaC*thetaC, data = dm.lands.1Mb)

# type 2 ANOVA
anova.diversity <- Anova(m.diversity)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)

r2.dm.tab[3,] <- c((anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] +
                    anova.diversity$VarExp[4]) * 100,
                  anova.diversity$VarExp[1] * 100,
                  anova.diversity$VarExp[2] * 100,
                  anova.diversity$VarExp[3] * 100,
                  anova.diversity$VarExp[4] * 100, 1000)

```

We now move on to whole-genome analyses

### 3.1 50 kb windows

```

# Chr 2L

# recombination landscapes
rho.dm.50kb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.rho.50kb.bedgraph", header = T)

# diversity
diversity.dm.50kb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.diversity.50kb.bedgraph", header = T)
diversity.dm.50kb.2L$avg <- apply(diversity.dm.50kb.2L[4:ncol(diversity.dm.50kb.2L)], 1, mean)

# mutation landscapes
theta.dm.50kb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.theta.50kb.bedgraph", header = T)

# TMRCA landscapes
tmrca.dm.50kb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.TMRCA.50kb.bedgraph", header = T)
tmrca.dm.50kb.2L$sample_mean <- apply(tmrca.dm.50kb.2L[4:ncol(tmrca.dm.50kb.2L)], 1, mean)

```

```

# missing data
missing.prop.50kb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.missing.prop.50kb.bedgraph", header=
intersect.50kb.2L <- apply(missing.prop.50kb.2L[4:ncol(missing.prop.50kb.2L)], 1, function(x) any(x > 0))

dm.lands.50kb.2L <- as.data.frame(cbind(
  diversity.dm.50kb.2L$chromStart,
  diversity.dm.50kb.2L$chromEnd,
  diversity.dm.50kb.2L$avg,
  theta.dm.50kb.2L$sample_mean,
  rho.dm.50kb.2L$sample_mean,
  tmrca.dm.50kb.2L$sample_mean))
names(dm.lands.50kb.2L) <- c("start", "end", "diversity", "theta", "rho", "tmrca")
dm.lands.50kb.2L$bin <- 1:nrow(dm.lands.50kb.2L)

# filters based on missing data
dm.lands.50kb.2L <- dm.lands.50kb.2L[which(intersect.50kb.2L == F),]

dm.lands.50kb.2L$chr <- "2L"

dm.lands.50kb.2L$thetaC <- dm.lands.50kb.2L$theta - mean(dm.lands.50kb.2L$theta)
dm.lands.50kb.2L$tmrcaC <- dm.lands.50kb.2L$tmrca - mean(dm.lands.50kb.2L$tmrca)
dm.lands.50kb.2L$rhoC <- dm.lands.50kb.2L$rho - mean(dm.lands.50kb.2L$rho)

g.div.dm.50kb.2L <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
  data = dm.lands.50kb.2L, weights = varPower(0, ~tmrcaC), cor = corAR1(0, ~bin), method="REML")

summary(g.div.dm.50kb.2L)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
## Data: dm.lands.50kb.2L
## AIC BIC logLik
## -4690.441 -4660.219 2353.22
##
## Correlation Structure: ARMA(1,0)
## Formula: ~bin
## Parameter estimate(s):
## Phi1
## 0.1318436
## Variance function:
## Structure: Power of variance covariate
## Formula: ~tmrcaC
## Parameter estimates:
## power
## 0.02239814
##
## Coefficients:
## Value Std.Error t-value p-value
## (Intercept) 0.0097130 0.00001087 893.9585 0.0000
## thetaC 0.9874554 0.00504783 195.6197 0.0000
## rhoC 0.0017748 0.00146095 1.2148 0.2253
## tmrcaC 0.0126819 0.00021188 59.8534 0.0000
## thetaC:tmrcaC 1.2080584 0.04507126 26.8033 0.0000
##
## Correlation:

```



```
##              (Intr) thetaC rhoC   tmrcaC
## thetaC      -0.016
## rhoC        0.027  0.026
## tmrcaC      -0.132 -0.379 -0.456
## thetaC:tmrcaC -0.306  0.061 -0.088  0.402
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -3.60864294 -0.52688547 -0.05949911  0.51229899  4.16571201
##
## Residual standard error: 0.0001811126
## Degrees of freedom: 323 total; 318 residual

# Chr 2R

# recombination landscapes
rho.dm.50kb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.rho.50kb.bedgraph", header = T)

# diversity
diversity.dm.50kb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.diversity.50kb.bedgraph", header = T)
diversity.dm.50kb.2R$avg <- apply(diversity.dm.50kb.2R[4:ncol(diversity.dm.50kb.2R)], 1, mean)

# mutation landscapes
theta.dm.50kb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.theta.50kb.bedgraph", header = T)

# TMRCA landscapes
tmrca.dm.50kb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.TMRCA.50kb.bedgraph", header = T)
tmrca.dm.50kb.2R$sample_mean <- apply(tmrca.dm.50kb.2R[4:ncol(tmrca.dm.50kb.2R)], 1, mean)

# missing data
missing.prop.50kb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.missing.prop.50kb.bedgraph", header = T)
intersect.50kb.2R <- apply(missing.prop.50kb.2R[4:ncol(missing.prop.50kb.2R)], 1, function(x) any(x > 0))

dm.lands.50kb.2R <- as.data.frame(cbind(diversity.dm.50kb.2R$chromStart,
                                       diversity.dm.50kb.2R$chromEnd,
                                       diversity.dm.50kb.2R$avg,
                                       theta.dm.50kb.2R$sample_mean,
                                       rho.dm.50kb.2R$sample_mean,
                                       tmrca.dm.50kb.2R$sample_mean))
names(dm.lands.50kb.2R) <- c("start", "end", "diversity", "theta", "rho", "tmrca")
dm.lands.50kb.2R$bin <- 1:nrow(dm.lands.50kb.2R)

# filters based on missing data
dm.lands.50kb.2R <- dm.lands.50kb.2R[which(intersect.50kb.2R == F),]

dm.lands.50kb.2R$chr <- "2R"

dm.lands.50kb.2R$thetaC <- dm.lands.50kb.2R$theta - mean(dm.lands.50kb.2R$theta)
dm.lands.50kb.2R$tmrcaC <- dm.lands.50kb.2R$tmrca - mean(dm.lands.50kb.2R$tmrca)
dm.lands.50kb.2R$rhoC <- dm.lands.50kb.2R$rho - mean(dm.lands.50kb.2R$rho)

g.div.dm.50kb.2R <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                        data = dm.lands.50kb.2R, weights = varPower(0, ~tmrcaC), cor = corAR1(0, ~bin),
```

```
summary(g.div.dm.50kb.2R)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
## Data: dm.lands.50kb.2R
##      AIC      BIC   logLik
## -4473.266 -4443.583 2244.633
##
## Correlation Structure: ARMA(1,0)
## Formula: ~bin
## Parameter estimate(s):
##      Phil
## 0.1094586
## Variance function:
## Structure: Power of variance covariate
## Formula: ~tmrcaC
## Parameter estimates:
##      power
## 0.05919674
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0085652 0.00000938  913.0793  0.0000
## thetaC       0.9709055 0.00361092  268.8801  0.0000
## rhoC         0.0001728 0.00149052    0.1160  0.9078
## tmrcaC       0.0116837 0.00019594   59.6303  0.0000
## thetaC:tmrcaC 1.0680644 0.04933969   21.6472  0.0000
##
## Correlation:
##              (Intr) thetaC rhoC   tmrcaC
## thetaC       0.044
## rhoC        -0.003  0.084
## tmrcaC      -0.103 -0.360 -0.381
## thetaC:tmrcaC -0.260 -0.114  0.022  0.341
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -3.56718360 -0.56905223  0.01512333  0.64333847  3.32352563
##
## Residual standard error: 0.0001778837
## Degrees of freedom: 302 total; 297 residual
```

```
# recombination landscapes
```

```
rho.dm.50kb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.rho.50kb.bedgraph", header = T)
```

```
# diversity
```

```
diversity.dm.50kb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.diversity.50kb.bedgraph", header = T)
diversity.dm.50kb.3L$avg <- apply(diversity.dm.50kb.3L[4:ncol(diversity.dm.50kb.3L)], 1, mean)
```

```
# mutation landscapes
```

```
theta.dm.50kb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.theta.50kb.bedgraph", header = T)
```

```
# TMRCA landscapes
```

```
tmrca.dm.50kb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.TMRCA.50kb.bedgraph", header = T)
```

```

tmrca.dm.50kb.3L$sample_mean <- apply(tmrca.dm.50kb.3L[4:ncol(tmrca.dm.50kb.3L)], 1, mean)

# missing data
missing.prop.50kb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.missing.prop.50kb.bedgraph", header=1)
intersect.50kb.3L <- apply(missing.prop.50kb.3L[4:ncol(missing.prop.50kb.3L)], 1, function(x) any(x > 0))

dm.lands.50kb.3L <- as.data.frame(cbind(
  diversity.dm.50kb.3L$chromStart,
  diversity.dm.50kb.3L$chromEnd,
  diversity.dm.50kb.3L$avg,
  theta.dm.50kb.3L$sample_mean,
  rho.dm.50kb.3L$sample_mean,
  tmrca.dm.50kb.3L$sample_mean))
names(dm.lands.50kb.3L) <- c("start", "end", "diversity", "theta", "rho", "tmrca")
dm.lands.50kb.3L$bin <- 1:nrow(dm.lands.50kb.3L)

# filters based on missing data
dm.lands.50kb.3L <- dm.lands.50kb.3L[which(intersect.50kb.3L == F),]

dm.lands.50kb.3L$chr <- "3L"

dm.lands.50kb.3L$thetaC <- dm.lands.50kb.3L$theta - mean(dm.lands.50kb.3L$theta)
dm.lands.50kb.3L$tmrcaC <- dm.lands.50kb.3L$tmrca - mean(dm.lands.50kb.3L$tmrca)
dm.lands.50kb.3L$rhoC <- dm.lands.50kb.3L$rho - mean(dm.lands.50kb.3L$rho)

g.div.dm.50kb.3L <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
  data = dm.lands.50kb.3L, weights = varPower(0, ~tmrcaC), cor = corAR1(0, ~bin),
  method="REML")

summary(g.div.dm.50kb.3L)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
## Data: dm.lands.50kb.3L
## AIC BIC logLik
## -5025.491 -4994.674 2520.746
##
## Correlation Structure: ARMA(1,0)
## Formula: ~bin
## Parameter estimate(s):
## Phi1
## 0.2696667
## Variance function:
## Structure: Power of variance covariate
## Formula: ~tmrcaC
## Parameter estimates:
## power
## 0.06685875
##
## Coefficients:
## Value Std.Error t-value p-value
## (Intercept) 0.0089556 0.000012514 715.6471 0.0000
## thetaC 0.9656448 0.004479703 215.5600 0.0000
## rhoC 0.0028522 0.001435660 1.9867 0.0477
## tmrcaC 0.0118244 0.000153683 76.9400 0.0000
## thetaC:tmrcaC 1.0917168 0.030139577 36.2220 0.0000

```

```
##
## Correlation:
##          (Intr) thetaC rhoC   tmrcaC
## thetaC      -0.004
## rhoC         0.004  0.150
## tmrcaC      -0.106 -0.427 -0.517
## thetaC:tmrcaC -0.323  0.045 -0.029  0.309
##
## Standardized residuals:
##          Min          Q1          Med          Q3          Max
## -3.70241395 -0.68659890  0.01012415  0.61472090  3.79680208
##
## Residual standard error: 0.0002214495
## Degrees of freedom: 348 total; 343 residual

# Chr 3R

# recombination landscapes
rho.dm.50kb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.rho.50kb.bedgraph", header = T)

# diversity
diversity.dm.50kb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.diversity.50kb.bedgraph", header = T)
diversity.dm.50kb.3R$avg <- apply(diversity.dm.50kb.3R[4:ncol(diversity.dm.50kb.3R)], 1, mean)

# mutation landscapes
theta.dm.50kb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.theta.50kb.bedgraph", header = T)

# TMRCA landscapes
tmrca.dm.50kb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.TMRCA.50kb.bedgraph", header = T)
tmrca.dm.50kb.3R$sample_mean <- apply(tmrca.dm.50kb.3R[4:ncol(tmrca.dm.50kb.3R)], 1, mean)

# missing data
missing.prop.50kb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.missing.prop.50kb.bedgraph", header = T)
intersect.50kb.3R <- apply(missing.prop.50kb.3R[4:ncol(missing.prop.50kb.3R)], 1, function(x) any(x > 0))

dm.lands.50kb.3R <- as.data.frame(cbind(diversity.dm.50kb.3R$chromStart,
                                       diversity.dm.50kb.3R$chromEnd,
                                       diversity.dm.50kb.3R$avg,
                                       theta.dm.50kb.3R$sample_mean,
                                       rho.dm.50kb.3R$sample_mean,
                                       tmrca.dm.50kb.3R$sample_mean))
names(dm.lands.50kb.3R) <- c("start", "end", "diversity", "theta", "rho", "tmrca")
dm.lands.50kb.3R$bin <- 1:nrow(dm.lands.50kb.3R)

# filters based on missing data
dm.lands.50kb.3R <- dm.lands.50kb.3R[which(intersect.50kb.3R == F),]

dm.lands.50kb.3R$chr <- "3R"

dm.lands.50kb.3R$thetaC <- dm.lands.50kb.3R$theta - mean(dm.lands.50kb.3R$theta)
dm.lands.50kb.3R$tmrcaC <- dm.lands.50kb.3R$tmrca - mean(dm.lands.50kb.3R$tmrca)
dm.lands.50kb.3R$rhoC <- dm.lands.50kb.3R$rho - mean(dm.lands.50kb.3R$rho)

g.div.dm.50kb.3R <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
```

```

data = dm.lands.50kb.3R, weights = varPower(0, ~tmrcaC), cor = corAR1(0, ~bin),

summary(g.div.dm.50kb.3R)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
## Data: dm.lands.50kb.3R
##      AIC      BIC    logLik
## -6576.594 -6543.846 3296.297
##
## Correlation Structure: ARMA(1,0)
## Formula: ~bin
## Parameter estimate(s):
##      Phil
## 0.2356751
## Variance function:
## Structure: Power of variance covariate
## Formula: ~tmrcaC
## Parameter estimates:
##      power
## 0.1042665
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0075334 0.00000895 841.7772  0.0000
## thetaC       0.9644922 0.00331480 290.9655  0.0000
## rhoC        -0.0017265 0.00118208  -1.4606  0.1449
## tmrcaC       0.0106635 0.00013264  80.3973  0.0000
## thetaC:tmrcaC 1.0663698 0.03719362 28.6708  0.0000
##
## Correlation:
##      (Intr) thetaC rhoC   tmrcaC
## thetaC      0.003
## rhoC       -0.022  0.107
## tmrcaC      -0.203 -0.346 -0.501
## thetaC:tmrcaC -0.384 -0.089  0.039  0.507
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -3.64020083 -0.59320311  0.03347561  0.68857658  4.67175243
##
## Residual standard error: 0.0002041515
## Degrees of freedom: 443 total; 438 residual

# all together now
dm.lands.50kb <- rbind.data.frame(dm.lands.50kb.2L, dm.lands.50kb.2R, dm.lands.50kb.3L, dm.lands.50kb.3R)

write.table(dm.lands.50kb, "dm_data/dm_chr_maps/dm.maps.50kb.tsv", quote = F, sep = "\t", row.names = F)

# Plots
scale.3d <- function(x) sprintf("%.3f", x) # digits shown in y axis

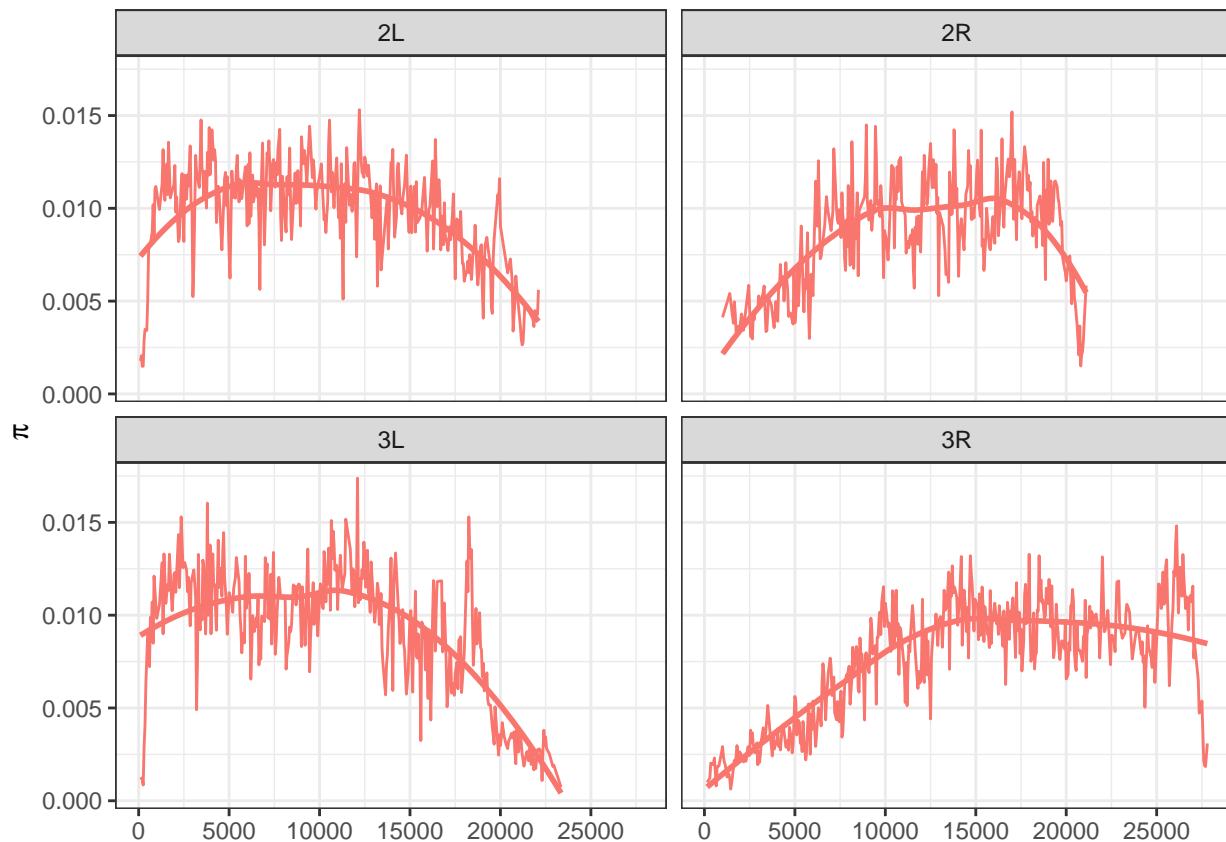
setTimeLimit(cpu = Inf, elapsed = Inf, transient = F)

```

```

molten.diversity <- melt(dm.lands.50kb[c(3,7,8)], id.vars = c("bin", "chr"))
diversity.map <- ggplot(data = molten.diversity, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
diversity.map <- diversity.map + geom_line(data = molten.diversity, colour = "#F8766D")
diversity.map <- diversity.map + geom_smooth(method = "loess", se = F, colour = "#F8766D")
diversity.map <- diversity.map + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
diversity.map <- diversity.map + labs(title = NULL, x = NULL, y = expression(pi))
diversity.map <- diversity.map + theme(axis.title.x = element_text(size = 20), axis.title.y = element_text(size = 20))
diversity.map

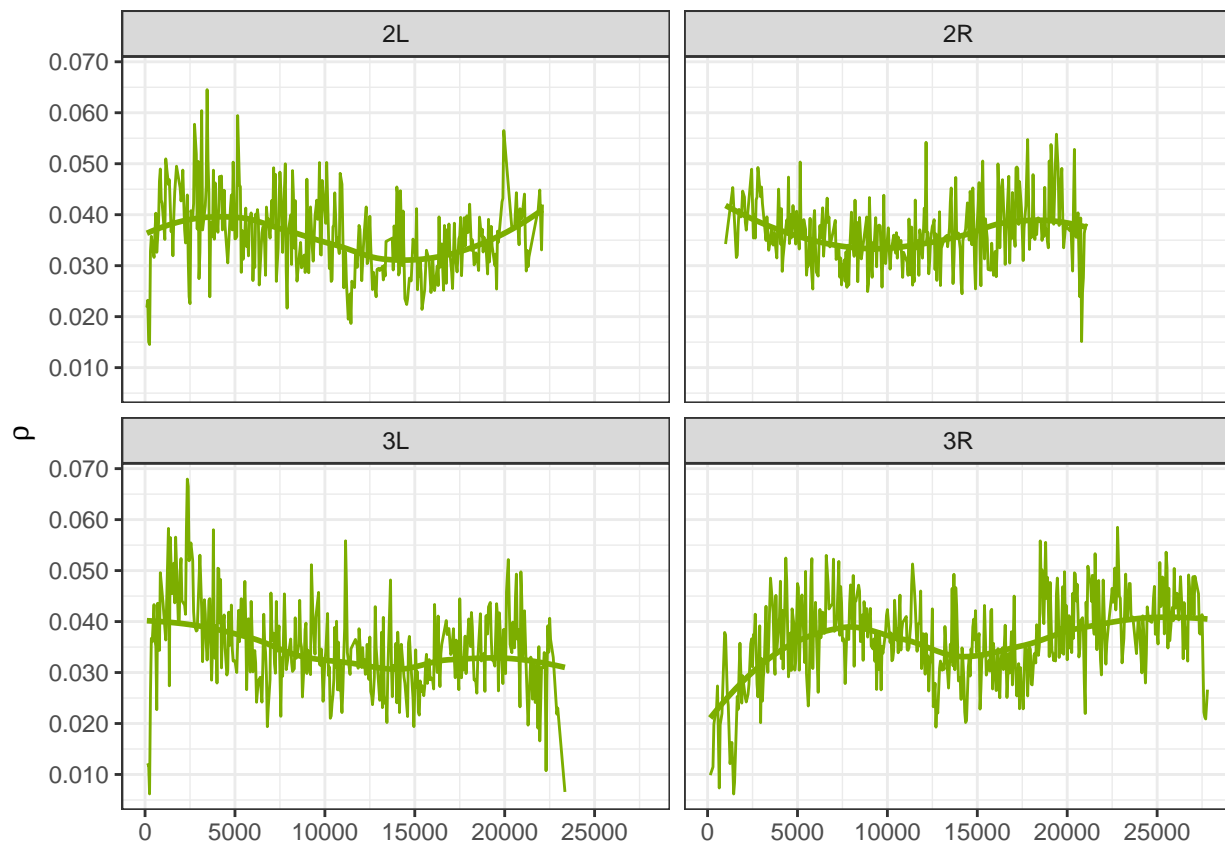
```



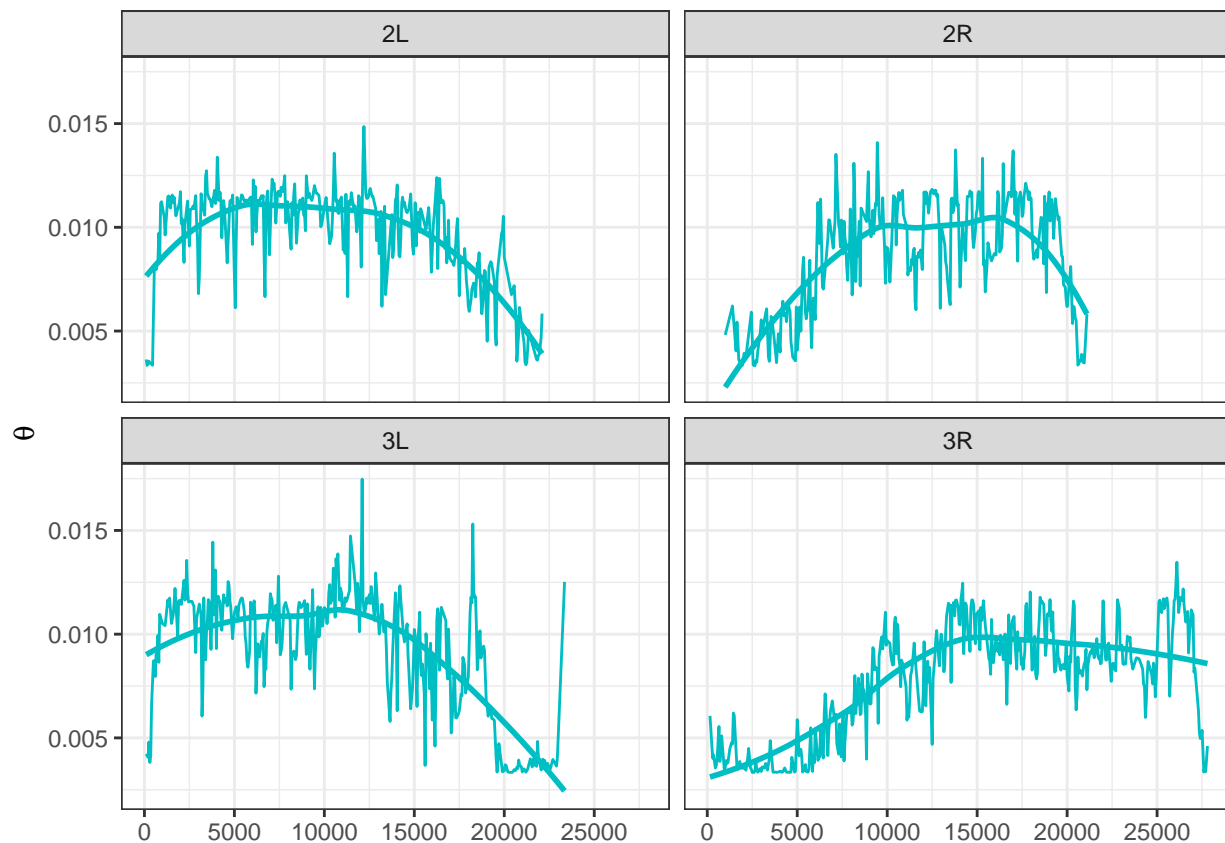
```

molten.rho <- melt(dm.lands.50kb[c(5,7,8)], id.vars = c("bin", "chr"))
rho.map <- ggplot(data = molten.rho, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
rho.map <- rho.map + geom_line(data = molten.rho, colour = "#7CAE00")
rho.map <- rho.map + geom_smooth(method = "loess", se = F, colour = "#7CAE00")
rho.map <- rho.map + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
rho.map <- rho.map + labs(title = NULL, x = NULL, y = expression(rho))
rho.map <- rho.map + theme(axis.title.x = element_text(size = 20), axis.title.y = element_text(size = 20))
rho.map

```

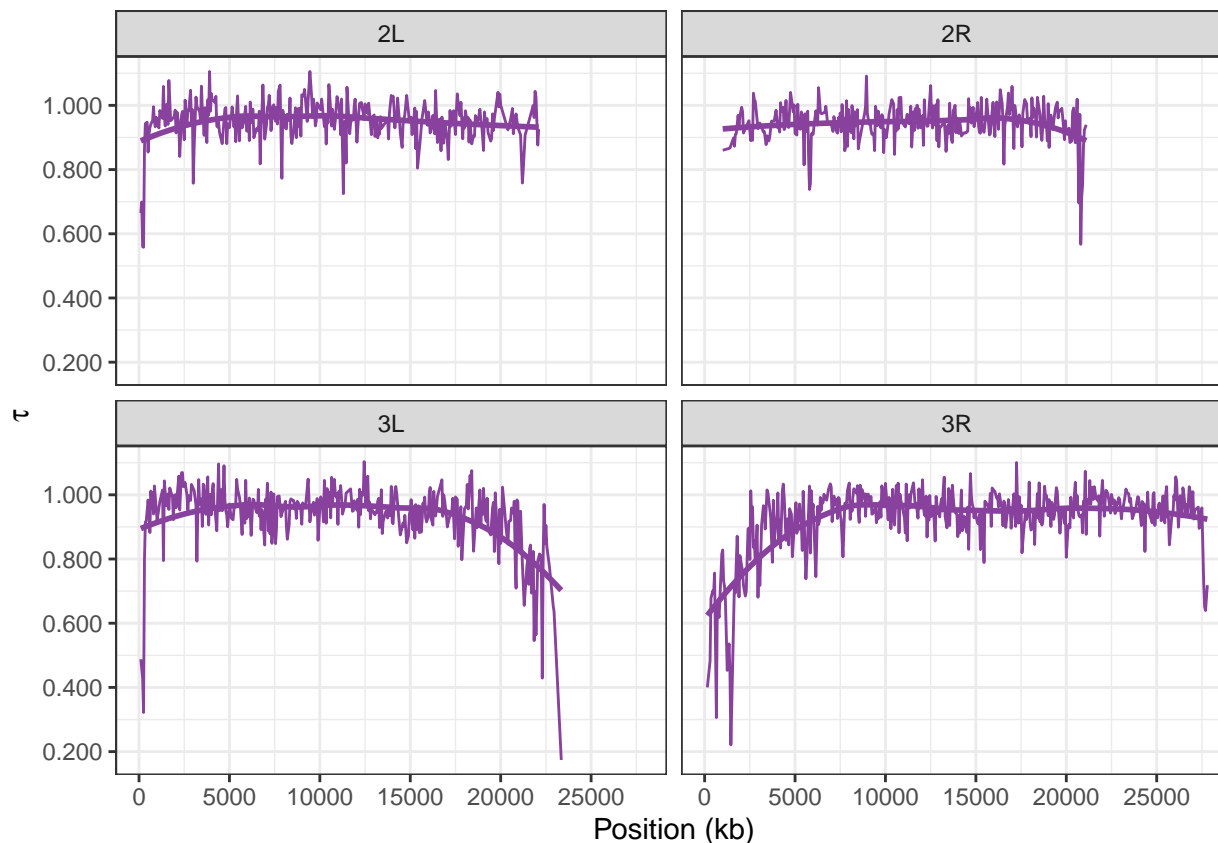


```
molten.theta <- melt(dm.lands.50kb[c(4,7,8)], id.vars = c("bin", "chr"))
theta.map <- ggplot(data = molten.theta, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
theta.map <- theta.map + geom_line(data = molten.theta, colour = "#00BFC4")
theta.map <- theta.map + geom_smooth(method = "loess", se = F, colour = "#00BFC4")
theta.map <- theta.map + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
theta.map <- theta.map + labs(title = NULL, x = NULL, y = expression(theta))
theta.map <- theta.map + theme(axis.title.x = element_text(size = 20), axis.title.y = element_text(size = 20))
theta.map
```



```
molten.tmrca <- melt(dm.lands.50kb[c(6,7,8)], id.vars = c("bin", "chr"))
tmrca.map <- ggplot(data = molten.tmrca, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
tmrca.map <- tmrca.map + geom_line(data = molten.tmrca, colour = "#88419d")
tmrca.map <- tmrca.map + geom_smooth(method = "loess", se = F, colour = "#88419d")
tmrca.map <- tmrca.map + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
tmrca.map <- tmrca.map + labs(title = NULL, x = "Position (kb)", y = expression(tau))
tmrca.map <- tmrca.map + theme(axis.title.x = element_text(size = 20), axis.title.y = element_text(size = 20))
tmrca.map
```





```
# genome-wide correlations between maps
cor.test(~rho + diversity, data = dm.lands.50kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: rho and diversity
## S = 369190884, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.2197879
```

```
cor.test(~rho + tmrca, data = dm.lands.50kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 245063950, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.4821057
```

```
cor.test(~theta + tmrca, data = dm.lands.50kb, method = "spearman")
```

```
##
```

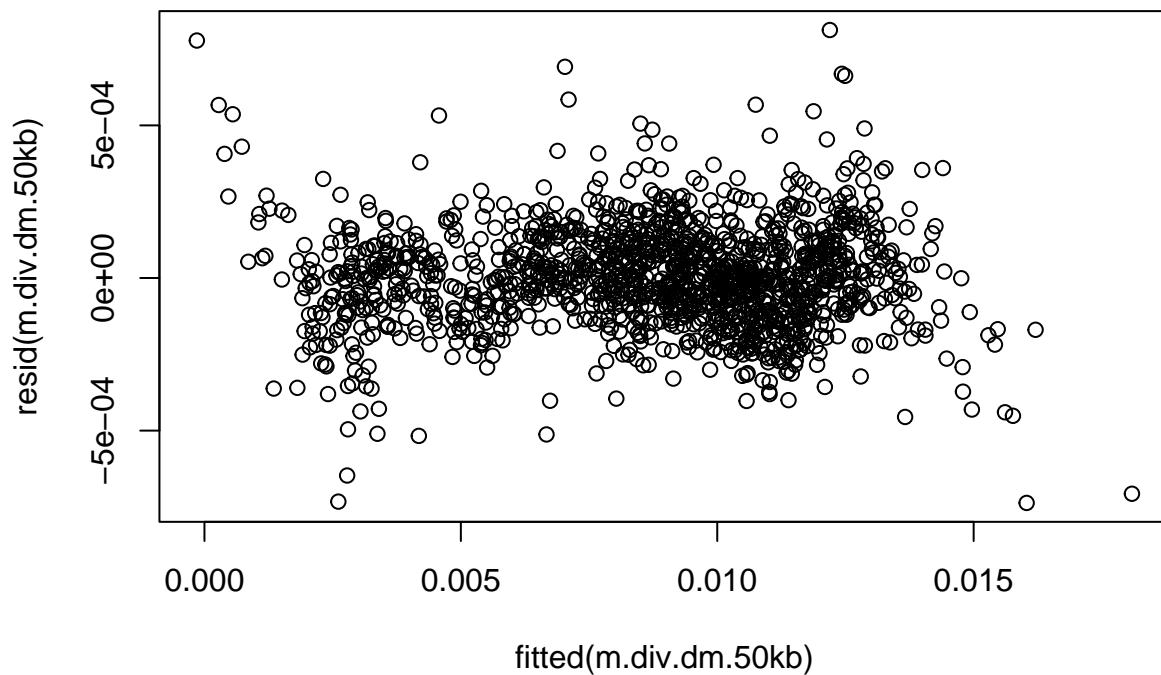
```
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 254806288, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.4615172

# Linear models
# centering
dm.lands.50kb$thetaC <- dm.lands.50kb$theta - mean(dm.lands.50kb$theta)
dm.lands.50kb$tmrcaC <- dm.lands.50kb$tmrca - mean(dm.lands.50kb$tmrca)
dm.lands.50kb$rhoC <- dm.lands.50kb$rho - mean(dm.lands.50kb$rho)

dm.lands.50kb$bin <- 1:nrow(dm.lands.50kb)

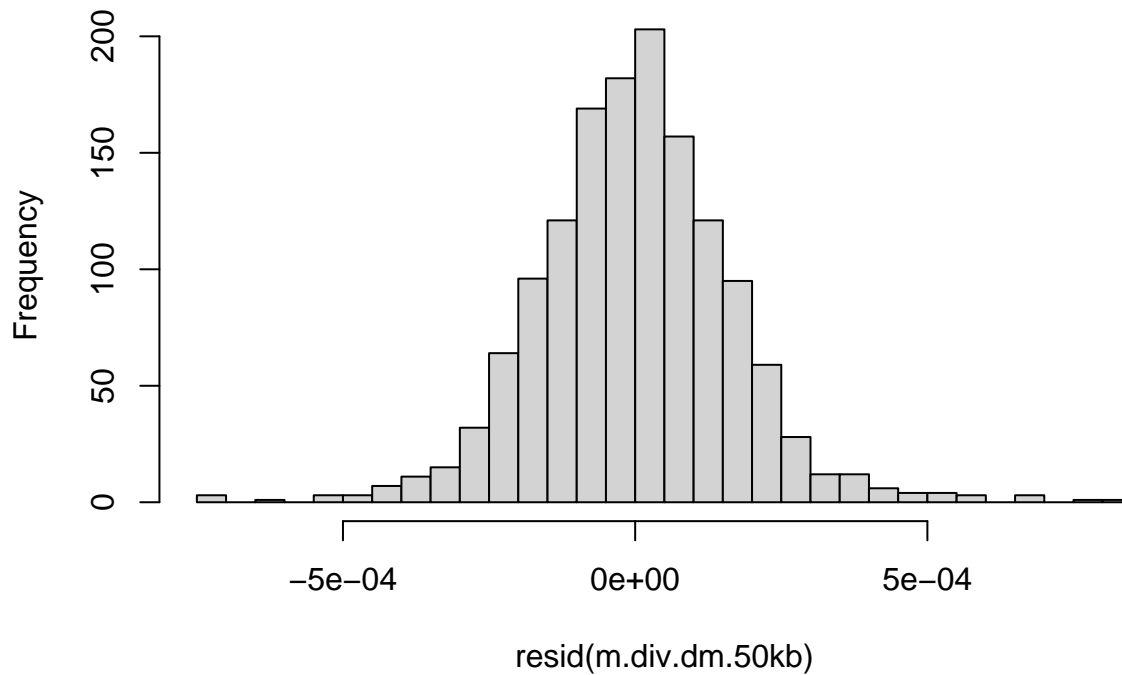
m.div.dm.50kb <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC), data = dm.lands.50kb)

plot(resid(m.div.dm.50kb)~fitted(m.div.dm.50kb))
```



```
hist(resid(m.div.dm.50kb), nclass = 30)
```

## Histogram of resid(m.div.dm.50kb)



```
dwtest(m.div.dm.50kb)
```

```
##
## Durbin-Watson test
##
## data:  m.div.dm.50kb
## DW = 1.5747, p-value = 3.694e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.div.dm.50kb, nsim = 10000)
```

```
##
## Harrison-McCabe test
##
## data:  m.div.dm.50kb
## HMC = 0.50534, p-value = 0.608
```

```
summary(m.div.dm.50kb)
```

```
##
## Call:
## lm(formula = diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
##     data = dm.lands.50kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.369e-04 -1.007e-04  1.500e-07  9.777e-05  8.124e-04
##
```

```
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  8.603e-03  4.838e-06 1778.154   <2e-16 ***
## thetaC       9.759e-01  1.822e-03  535.514   <2e-16 ***
## rhoC         1.209e-03  6.874e-04   1.759    0.0788 .
## tmrcaC       1.139e-02  7.780e-05  146.379   <2e-16 ***
## thetaC:tmrcaC 1.065e+00  1.654e-02   64.386   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001675 on 1411 degrees of freedom
## Multiple R-squared:  0.9973, Adjusted R-squared:  0.9973
## F-statistic: 1.286e+05 on 4 and 1411 DF,  p-value: < 2.2e-16

# type 2 ANOVA
anova.diversity <- Anova(m.div.dm.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
anova.diversity

## Anova Table (Type II tests)
##
## Response: diversity
##              Sum Sq   Df    F value    Pr(>F)   VarExp
## thetaC         0.0081479     1 2.9046e+05 0.000000 0.92684
## rhoC           0.0000001     1 3.0932e+00 0.078838 0.00001
## tmrcaC         0.0004872     1 1.7367e+04 0.000000 0.05542
## thetaC:tmrcaC 0.0001163     1 4.1456e+03 0.000000 0.01323
## Residuals      0.0000396 1411                0.00450

# GLS
g.div.dm.50kb.1 <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
  data = dm.lands.50kb, weights = varPower(0, ~tmrcaC|chr), cor = corAR1(0, ~bin|chr)

g.div.dm.50kb.2 <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
  data = dm.lands.50kb, weights = varPower(0, ~theta|chr), cor = corAR1(0, ~bin|chr)

g.div.dm.50kb.3 <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
  data = dm.lands.50kb, weights = varPower(0, ~theta|chr), method = "ML")

g.div.dm.50kb.4 <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
  data = dm.lands.50kb, weights = varPower(0, ~tmrcaC|chr), method = "ML")

AIC(g.div.dm.50kb.1, g.div.dm.50kb.2, g.div.dm.50kb.3, g.div.dm.50kb.4)

##              df          AIC
## g.div.dm.50kb.1 11 -20716.73
## g.div.dm.50kb.2 11 -20681.77
## g.div.dm.50kb.3 10 -20620.30
## g.div.dm.50kb.4 10 -20655.40
summary(g.div.dm.50kb.1)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
## Data: dm.lands.50kb
```

```
##           AIC           BIC    logLik
##    -20716.73 -20658.92 10369.36
##
## Correlation Structure: AR(1)
## Formula: ~bin | chr
## Parameter estimate(s):
##     Phi
## 0.21449
## Variance function:
## Structure: Power of variance covariate, different strata
## Formula: ~tmrcaC | chr
## Parameter estimates:
##           2L           2R           3L           3R
## 0.05445273 0.09431197 0.06301401 0.11582323
##
## Coefficients:
##              Value    Std.Error    t-value p-value
## (Intercept)  0.0085894 0.000005516 1557.1023  0.0000
## thetaC       0.9739854 0.002055438  473.8578  0.0000
## rhoC         0.0004325 0.000704757   0.6136  0.5396
## tmrcaC       0.0115855 0.000084597  136.9496  0.0000
## thetaC:tmrcaC 1.0997190 0.018815541   58.4474  0.0000
##
## Correlation:
##              (Intr) thetaC rhoC    tmrcaC
## thetaC       0.044
## rhoC        -0.005  0.137
## tmrcaC      -0.143 -0.375 -0.480
## thetaC:tmrcaC -0.289 -0.046  0.003  0.409
##
## Standardized residuals:
##              Min           Q1           Med           Q3           Max
## -4.04795893 -0.58185196  0.04817016  0.65222190  4.30625982
##
## Residual standard error: 0.0002182622
## Degrees of freedom: 1416 total; 1411 residual
```

```
vif(g.div.dm.50kb.1)
```

```
##           thetaC           rhoC           tmrcaC thetaC:tmrcaC
##           1.191541           1.394431           1.938119           1.306771
```

```
# Linear model without TMRCA --> rho becomes significant
```

```
g.div.dm.50kb.5 <- gls(diversity ~ (thetaC + rhoC),
                      data = dm.lands.50kb, weights = varPower(0, ~thetaC|chr), cor = corAR1(0, ~bin|chr))
```

```
summary(g.div.dm.50kb.5)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC)
## Data: dm.lands.50kb
##           AIC           BIC    logLik
##    -16799.87 -16752.56 8408.933
##
## Correlation Structure: AR(1)
```

```
## Formula: ~bin | chr
## Parameter estimate(s):
##      Phi
## 0.2247158
## Variance function:
## Structure: Power of variance covariate, different strata
## Formula: ~thetaC | chr
## Parameter estimates:
##      2L      2R      3L      3R
## 0.09783222 0.09779700 0.05144593 0.09035535
##
## Coefficients:
##              Value   Std.Error  t-value p-value
## (Intercept) 0.0087332 0.000021244 411.0950      0
## thetaC      1.0944710 0.008181052 133.7812      0
## rhoC        0.0499971 0.002361411  21.1725      0
##
## Correlation:
##      (Intr) thetaC
## thetaC -0.012
## rhoC   -0.003 -0.148
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -12.5524266 -0.5303105  0.0136743  0.6108370  3.0172248
##
## Residual standard error: 0.001130581
## Degrees of freedom: 1416 total; 1413 residual
```

## 3.2 200 kb windows

```
# Chr 2L

# recombination landscapes
rho.dm.200kb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.rho.200kb.bedgraph", header = T)

# diversity
diversity.dm.200kb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.diversity.200kb.bedgraph", header = T)
diversity.dm.200kb.2L$avg <- apply(diversity.dm.200kb.2L[,4:ncol(diversity.dm.200kb.2L)], 1, mean)

# mutation landscapes
theta.dm.200kb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.theta.200kb.bedgraph", header = T)

# TMRCA landscapes
tmrca.dm.200kb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.TMRCA.200kb.bedgraph", header = T)
tmrca.dm.200kb.2L$sample_mean <- apply(tmrca.dm.200kb.2L[,4:ncol(tmrca.dm.200kb.2L)], 1, mean)

# missing data
missing.prop.200kb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.missing.prop.200kb.bedgraph", header = T)
intersect.200kb.2L <- apply(missing.prop.200kb.2L[,4:ncol(missing.prop.200kb.2L)], 1, function(x) any(x != 0))

dm.lands.200kb.2L <- as.data.frame(cbind(diversity.dm.200kb.2L$chromStart,
                                         diversity.dm.200kb.2L$chromEnd,
                                         diversity.dm.200kb.2L$avg,
```

```

theta.dm.200kb.2L$sample_mean,
rho.dm.200kb.2L$sample_mean,
tmrca.dm.200kb.2L$sample_mean))
names(dm.lands.200kb.2L) <- c("start", "end", "diversity", "theta", "rho", "tmrca")
dm.lands.200kb.2L$bin <- 1:nrow(dm.lands.200kb.2L)

# filters based on missing data
dm.lands.200kb.2L <- dm.lands.200kb.2L[which(intersect.200kb.2L == F),]

dm.lands.200kb.2L$chr <- "2L"

dm.lands.200kb.2L$thetaC <- dm.lands.200kb.2L$theta - mean(dm.lands.200kb.2L$theta)
dm.lands.200kb.2L$tmrcaC <- dm.lands.200kb.2L$tmrca - mean(dm.lands.200kb.2L$tmrca)
dm.lands.200kb.2L$rhoC <- dm.lands.200kb.2L$rho - mean(dm.lands.200kb.2L$rho)

g.div.dm.200kb.2L <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                        data = dm.lands.200kb.2L, weights = varPower(0, ~tmrcaC), cor = corAR1(0, ~bin))

summary(g.div.dm.200kb.2L)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
## Data: dm.lands.200kb.2L
## AIC BIC logLik
## -1244.082 -1224.731 630.041
##
## Correlation Structure: ARMA(1,0)
## Formula: ~bin
## Parameter estimate(s):
## Phi1
## 0.1002962
## Variance function:
## Structure: Power of variance covariate
## Formula: ~tmrcaC
## Parameter estimates:
## power
## 0.05640964
##
## Coefficients:
## Value Std.Error t-value p-value
## (Intercept) 0.0096851 0.00001590 609.2540 0.0000
## thetaC 0.9912075 0.00852140 116.3197 0.0000
## rhoC 0.0035710 0.00328708 1.0864 0.2807
## tmrcaC 0.0122532 0.00061930 19.7855 0.0000
## thetaC:tmrcaC 1.0357572 0.10202136 10.1524 0.0000
##
## Correlation:
## (Intr) thetaC rhoC tmrcaC
## thetaC 0.034
## rhoC 0.100 0.195
## tmrcaC -0.244 -0.449 -0.572
## thetaC:tmrcaC -0.337 -0.112 -0.276 0.696
##
## Standardized residuals:

```

```
##           Min           Q1           Med           Q3           Max
## -1.8896987 -0.6840157 -0.1976799  0.4999174  3.1380216
##
## Residual standard error: 0.0001535204
## Degrees of freedom: 83 total; 78 residual

# Chr 2R

# recombination landscapes
rho.dm.200kb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.rho.200kb.bedgraph", header = T)

# diversity
diversity.dm.200kb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.diversity.200kb.bedgraph", header = T)
diversity.dm.200kb.2R$avg <- apply(diversity.dm.200kb.2R[4:ncol(diversity.dm.200kb.2R)], 1, mean)

# mutation landscapes
theta.dm.200kb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.theta.200kb.bedgraph", header = T)

# TMRCA landscapes
tmrca.dm.200kb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.TMRCA.200kb.bedgraph", header = T)
tmrca.dm.200kb.2R$sample_mean <- apply(tmrca.dm.200kb.2R[4:ncol(tmrca.dm.200kb.2R)], 1, mean)

# missing data
missing.prop.200kb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.missing.prop.200kb.bedgraph", header = T)
intersect.200kb.2R <- apply(missing.prop.200kb.2R[4:ncol(missing.prop.200kb.2R)], 1, function(x) any(x == 0))

dm.lands.200kb.2R <- as.data.frame(cbind(diversity.dm.200kb.2R$chromStart,
                                         diversity.dm.200kb.2R$chromEnd,
                                         diversity.dm.200kb.2R$avg,
                                         theta.dm.200kb.2R$sample_mean,
                                         rho.dm.200kb.2R$sample_mean,
                                         tmrca.dm.200kb.2R$sample_mean))
names(dm.lands.200kb.2R) <- c("start", "end", "diversity", "theta", "rho", "tmrca")
dm.lands.200kb.2R$bin <- 1:nrow(dm.lands.200kb.2R)

# filters based on missing data
dm.lands.200kb.2R <- dm.lands.200kb.2R[which(intersect.200kb.2R == F),]

dm.lands.200kb.2R$chr <- "2R"

dm.lands.200kb.2R$thetaC <- dm.lands.200kb.2R$theta - mean(dm.lands.200kb.2R$theta)
dm.lands.200kb.2R$tmrcaC <- dm.lands.200kb.2R$tmrca - mean(dm.lands.200kb.2R$tmrca)
dm.lands.200kb.2R$rhoC <- dm.lands.200kb.2R$rho - mean(dm.lands.200kb.2R$rho)

g.div.dm.200kb.2R <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                        data = dm.lands.200kb.2R, weights = varPower(0, ~tmrcaC), cor = corAR1(0, ~bin))

summary(g.div.dm.200kb.2R)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
## Data: dm.lands.200kb.2R
##      AIC      BIC    logLik
## -1165.932 -1147.499 590.9659
##
```



```
## Correlation Structure: ARMA(1,0)
## Formula: ~bin
## Parameter estimate(s):
##     Phi1
## 0.2372428
## Variance function:
## Structure: Power of variance covariate
## Formula: ~tmrcaC
## Parameter estimates:
##     power
## 0.06275131
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0085392 0.00001264 675.8307  0.0000
## thetaC       0.9699152 0.00505288 191.9530  0.0000
## rhoC         0.0027465 0.00292834   0.9379  0.3516
## tmrcaC       0.0110299 0.00043604  25.2959  0.0000
## thetaC:tmrcaC 0.9373743 0.11041693   8.4894  0.0000
##
## Correlation:
##              (Intr) thetaC rhoC   tmrcaC
## thetaC       0.030
## rhoC        -0.063  0.097
## tmrcaC      -0.153 -0.376 -0.219
## thetaC:tmrcaC -0.290 -0.069  0.120  0.497
##
## Standardized residuals:
##              Min          Q1          Med          Q3          Max
## -2.13836840 -0.68198223  0.08032991  0.80898896  2.57448459
##
## Residual standard error: 0.0001098196
## Degrees of freedom: 74 total; 69 residual

# recombination landscapes
rho.dm.200kb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.rho.200kb.bedgraph", header = T)

# diversity
diversity.dm.200kb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.diversity.200kb.bedgraph", header = T)
diversity.dm.200kb.3L$avg <- apply(diversity.dm.200kb.3L[4:ncol(diversity.dm.200kb.3L)], 1, mean)

# mutation landscapes
theta.dm.200kb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.theta.200kb.bedgraph", header = T)

# TMRCA landscapes
tmrca.dm.200kb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.TMRCA.200kb.bedgraph", header = T)
tmrca.dm.200kb.3L$sample_mean <- apply(tmrca.dm.200kb.3L[4:ncol(tmrca.dm.200kb.3L)], 1, mean)

# missing data
missing.prop.200kb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.missing.prop.200kb.bedgraph", header = T)
intersect.200kb.3L <- apply(missing.prop.200kb.3L[4:ncol(missing.prop.200kb.3L)], 1, function(x) any(x != 0))

dm.lands.200kb.3L <- as.data.frame(cbind(diversity.dm.200kb.3L$chromStart,
                                         diversity.dm.200kb.3L$chromEnd,
```

```

diversity.dm.200kb.3L$avg,
theta.dm.200kb.3L$sample_mean,
rho.dm.200kb.3L$sample_mean,
tmrca.dm.200kb.3L$sample_mean))
names(dm.lands.200kb.3L) <- c("start", "end", "diversity", "theta", "rho", "tmrca")
dm.lands.200kb.3L$bin <- 1:nrow(dm.lands.200kb.3L)

# filters based on missing data
dm.lands.200kb.3L <- dm.lands.200kb.3L[which(intersect.200kb.3L == F),]

dm.lands.200kb.3L$chr <- "3L"

dm.lands.200kb.3L$thetaC <- dm.lands.200kb.3L$theta - mean(dm.lands.200kb.3L$theta)
dm.lands.200kb.3L$tmrcaC <- dm.lands.200kb.3L$tmrca - mean(dm.lands.200kb.3L$tmrca)
dm.lands.200kb.3L$rhoC <- dm.lands.200kb.3L$rho - mean(dm.lands.200kb.3L$rho)

g.div.dm.200kb.3L <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                        data = dm.lands.200kb.3L, weights = varPower(0, ~tmrcaC), cor = corAR1(0, ~bin))

summary(g.div.dm.200kb.3L)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
## Data: dm.lands.200kb.3L
## AIC BIC logLik
## -1307.956 -1288.321 661.9781
##
## Correlation Structure: ARMA(1,0)
## Formula: ~bin
## Parameter estimate(s):
## Phi1
## 0.2105682
## Variance function:
## Structure: Power of variance covariate
## Formula: ~tmrcaC
## Parameter estimates:
## power
## -0.01470136
##
## Coefficients:
## Value Std.Error t-value p-value
## (Intercept) 0.0089474 0.00001816 492.7438 0.0000
## thetaC 0.9811147 0.00736731 133.1713 0.0000
## rhoC 0.0078270 0.00261567 2.9923 0.0037
## tmrcaC 0.0106115 0.00042145 25.1784 0.0000
## thetaC:tmrcaC 0.7871946 0.07703046 10.2193 0.0000
##
## Correlation:
## (Intr) thetaC rhoC tmrcaC
## thetaC 0.133
## rhoC 0.079 0.300
## tmrcaC -0.394 -0.661 -0.463
## thetaC:tmrcaC -0.587 -0.238 -0.153 0.680
##

```

```

## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.89079240 -0.54320978  0.05844326  0.51224633  2.31374278
##
## Residual standard error: 0.0001059737
## Degrees of freedom: 86 total; 81 residual

# Chr 3R

# recombination landscapes
rho.dm.200kb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.rho.200kb.bedgraph", header = T)

# diversity
diversity.dm.200kb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.diversity.200kb.bedgraph", header = T)
diversity.dm.200kb.3R$avg <- apply(diversity.dm.200kb.3R[4:nrow(diversity.dm.200kb.3R)], 1, mean)

# mutation landscapes
theta.dm.200kb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.theta.200kb.bedgraph", header = T)

# TMRCA landscapes
tmrca.dm.200kb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.TMRCA.200kb.bedgraph", header = T)
tmrca.dm.200kb.3R$sample_mean <- apply(tmrca.dm.200kb.3R[4:nrow(tmrca.dm.200kb.3R)], 1, mean)

# missing data
missing.prop.200kb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.missing.prop.200kb.bedgraph", header = T)
intersect.200kb.3R <- apply(missing.prop.200kb.3R[4:nrow(missing.prop.200kb.3R)], 1, function(x) any(x == 0))

dm.lands.200kb.3R <- as.data.frame(cbind(diversity.dm.200kb.3R$chromStart,
                                         diversity.dm.200kb.3R$chromEnd,
                                         diversity.dm.200kb.3R$avg,
                                         theta.dm.200kb.3R$sample_mean,
                                         rho.dm.200kb.3R$sample_mean,
                                         tmrca.dm.200kb.3R$sample_mean))
names(dm.lands.200kb.3R) <- c("start", "end", "diversity", "theta", "rho", "tmrca")
dm.lands.200kb.3R$bin <- 1:nrow(dm.lands.200kb.3R)

# filters based on missing data
dm.lands.200kb.3R <- dm.lands.200kb.3R[which(intersect.200kb.3R == F),]

dm.lands.200kb.3R$chr <- "3R"

dm.lands.200kb.3R$thetaC <- dm.lands.200kb.3R$theta - mean(dm.lands.200kb.3R$theta)
dm.lands.200kb.3R$tmrcaC <- dm.lands.200kb.3R$tmrca - mean(dm.lands.200kb.3R$tmrca)
dm.lands.200kb.3R$rhoC <- dm.lands.200kb.3R$rho - mean(dm.lands.200kb.3R$rho)

g.div.dm.200kb.3R <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                        data = dm.lands.200kb.3R, weights = varPower(0, ~tmrcaC), cor = corAR1(0, ~bin))

summary(g.div.dm.200kb.3R)

## Generalized least squares fit by maximum likelihood
##      Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
##      Data: dm.lands.200kb.3R
##           AIC           BIC    logLik

```

```
## -1883.368 -1861.135 949.684
##
## Correlation Structure: ARMA(1,0)
## Formula: ~bin
## Parameter estimate(s):
## Phi1
## 0.09974947
## Variance function:
## Structure: Power of variance covariate
## Formula: ~tmrcaC
## Parameter estimates:
## power
## 0.1960113
##
## Coefficients:
## Value Std.Error t-value p-value
## (Intercept) 0.0075661 0.00001040 727.7696 0.0000
## thetaC 0.9683546 0.00387775 249.7209 0.0000
## rhoC -0.0041012 0.00175339 -2.3390 0.0211
## tmrcaC 0.0103602 0.00028386 36.4976 0.0000
## thetaC:tmrcaC 0.7922675 0.07387641 10.7242 0.0000
##
## Correlation:
## (Intr) thetaC rhoC tmrcaC
## thetaC 0.283
## rhoC 0.214 0.306
## tmrcaC -0.574 -0.559 -0.546
## thetaC:tmrcaC -0.625 -0.374 -0.215 0.781
##
## Standardized residuals:
## Min Q1 Med Q3 Max
## -2.78869608 -0.72931864 -0.06801681 0.64535064 2.75174311
##
## Residual standard error: 0.0001631813
## Degrees of freedom: 119 total; 114 residual

# all together now
dm.lands.200kb <- rbind.data.frame(dm.lands.200kb.2L, dm.lands.200kb.2R, dm.lands.200kb.3L, dm.lands.200kb.3R)

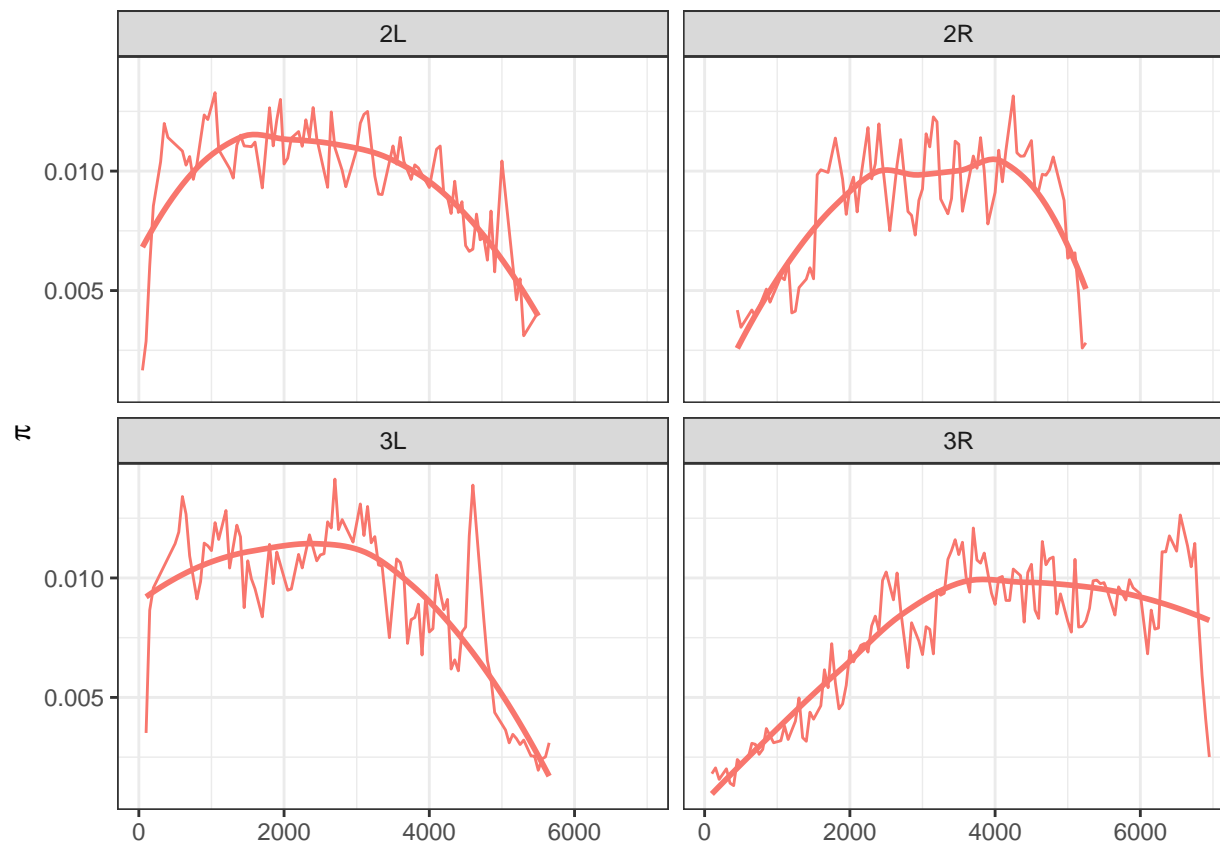
write.table(dm.lands.200kb, "dm_data/dm_chr_maps/dm.maps.200kb.tsv", quote = F, sep = "\t", row.names = F)

# Plots
scale.3d <- function(x) sprintf("%.3f", x) # digits shown in y axis

setTimeLimit(cpu = Inf, elapsed = Inf, transient = F)

molten.diversity <- melt(dm.lands.200kb[c(3,7,8)], id.vars = c("bin", "chr"))
diversity.map <- ggplot(data = molten.diversity, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
diversity.map <- diversity.map + geom_line(data = molten.diversity, colour = "#F8766D")
diversity.map <- diversity.map + geom_smooth(method = "loess", se = F, colour = "#F8766D")
diversity.map <- diversity.map + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
diversity.map <- diversity.map + labs(title = NULL, x = NULL, y = expression(pi))
diversity.map <- diversity.map + theme(axis.title.x = element_text(size = 20), axis.title.y = element_text(size = 20))
```

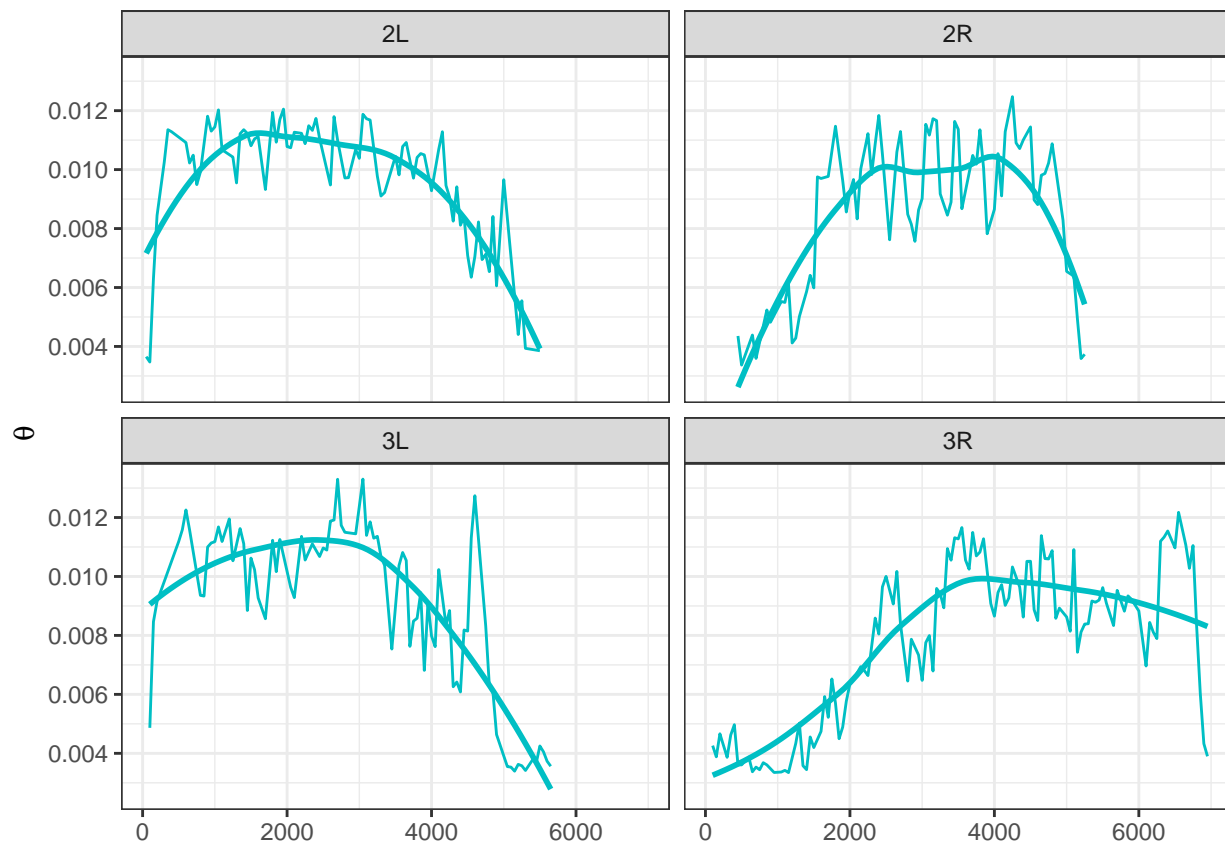
diversity.map



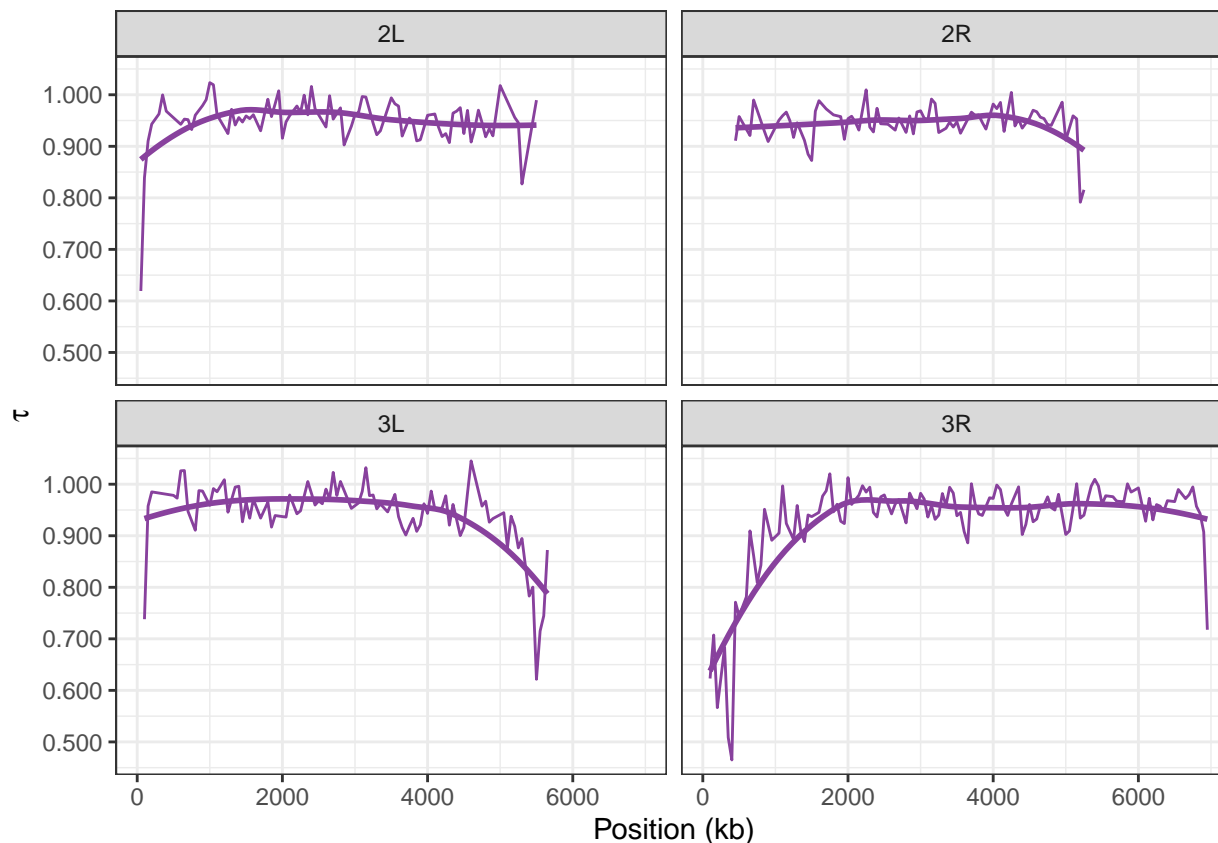
```
molten.rho <- melt(dm.lands.200kb[c(5,7,8)], id.vars = c("bin", "chr"))
rho.map <- ggplot(data = molten.rho, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
rho.map <- rho.map + geom_line(data = molten.rho, colour = "#7CAE00")
rho.map <- rho.map + geom_smooth(method = "loess", se = F, colour = "#7CAE00")
rho.map <- rho.map + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
rho.map <- rho.map + labs(title = NULL, x = NULL, y = expression(rho))
rho.map <- rho.map + theme(axis.title.x = element_text(size = 20), axis.title.y = element_text(size = 20))
rho.map
```



```
molten.theta <- melt(dm.lands.200kb[c(4,7,8)], id.vars = c("bin", "chr"))
theta.map <- ggplot(data = molten.theta, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
theta.map <- theta.map + geom_line(data = molten.theta, colour = "#00BFC4")
theta.map <- theta.map + geom_smooth(method = "loess", se = F, colour = "#00BFC4")
theta.map <- theta.map + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
theta.map <- theta.map + labs(title = NULL, x = NULL, y = expression(theta))
theta.map <- theta.map + theme(axis.title.x = element_text(size = 20), axis.title.y = element_text(size = 20))
theta.map
```



```
molten.tmrca <- melt(dm.lands.200kb[c(6,7,8)], id.vars = c("bin", "chr"))
tmrca.map <- ggplot(data = molten.tmrca, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
tmrca.map <- tmrca.map + geom_line(data = molten.tmrca, colour = "#88419d")
tmrca.map <- tmrca.map + geom_smooth(method = "loess", se = F, colour = "#88419d")
tmrca.map <- tmrca.map + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
tmrca.map <- tmrca.map + labs(title = NULL, x = "Position (kb)", y = expression(tau))
tmrca.map <- tmrca.map + theme(axis.title.x = element_text(size = 20), axis.title.y = element_text(size = 20))
tmrca.map
```



```
# genome-wide correlations
cor.test(~rho + diversity, data = dm.lands.200kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: rho and diversity
## S = 6652998, p-value = 0.002515
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.1585153
```

```
cor.test(~rho + tmrca, data = dm.lands.200kb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 4384696, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.4454147
```

```
cor.test(~theta + tmrca, data = dm.lands.200kb, method = "spearman")
```

```
##
```



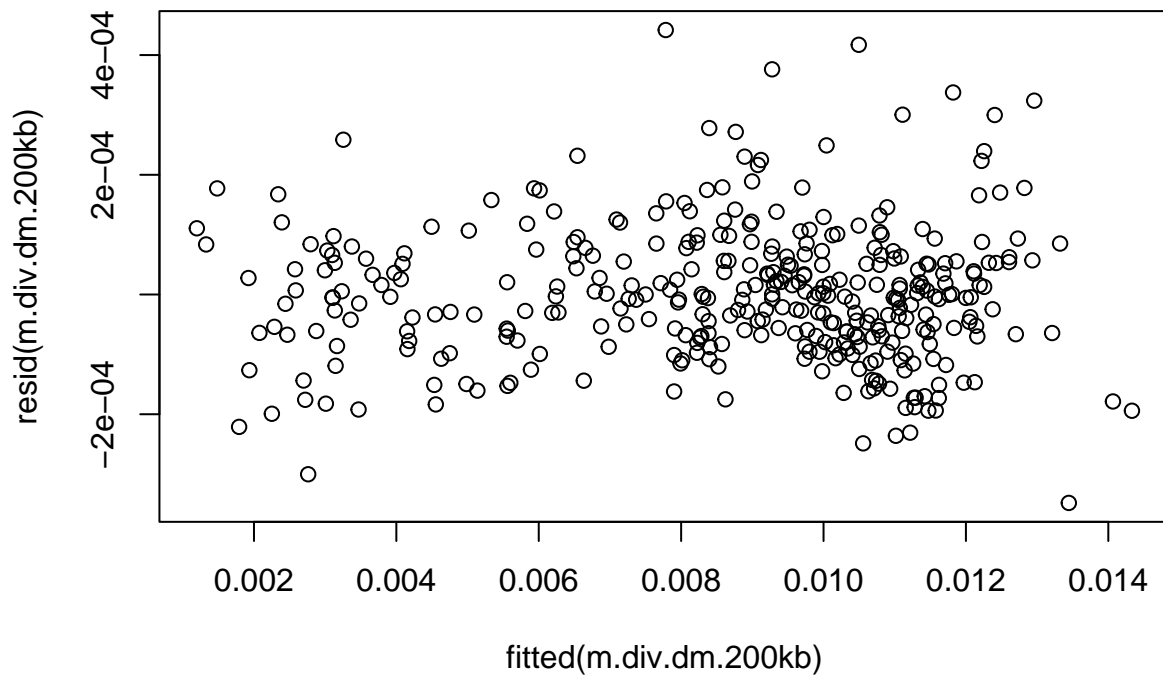
```
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 3844174, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.513781

# Linear models
# centering
dm.lands.200kb$thetaC <- dm.lands.200kb$theta - mean(dm.lands.200kb$theta)
dm.lands.200kb$tmrcaC <- dm.lands.200kb$tmrca - mean(dm.lands.200kb$tmrca)
dm.lands.200kb$rhoC <- dm.lands.200kb$rho - mean(dm.lands.200kb$rho)

dm.lands.200kb$bin <- 1:nrow(dm.lands.200kb)

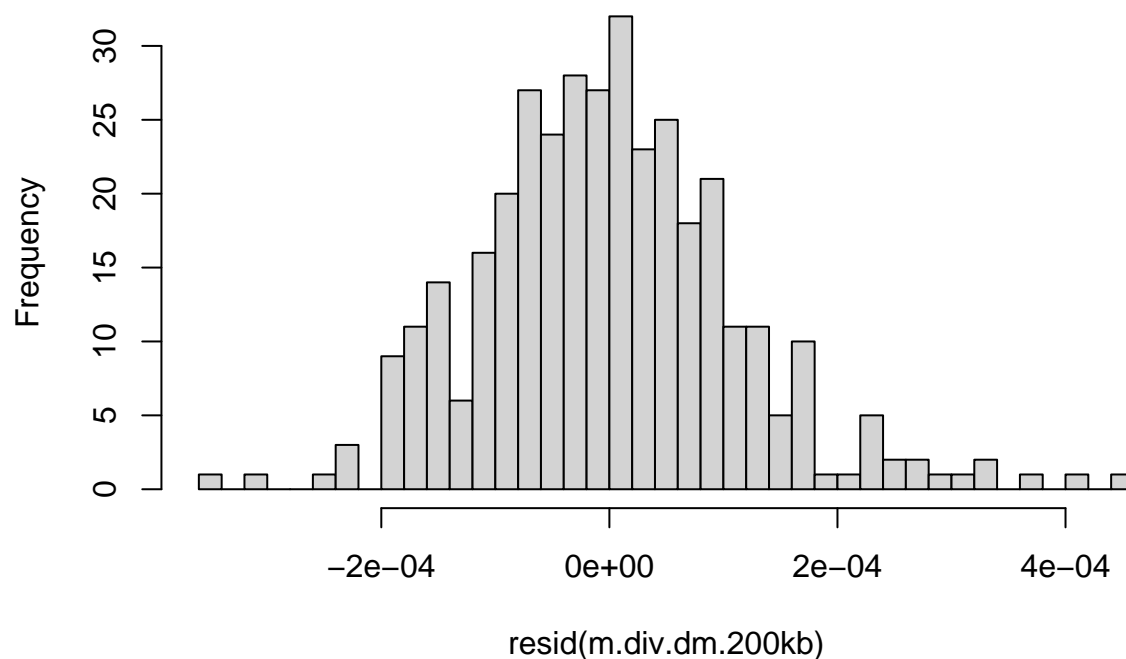
m.div.dm.200kb <- lm( diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC), data = dm.lands.200kb)

plot(resid(m.div.dm.200kb)~fitted(m.div.dm.200kb))
```



```
hist(resid(m.div.dm.200kb), nclass = 30)
```

## Histogram of resid(m.div.dm.200kb)



```
dwtest(m.div.dm.200kb)
```

```
##
## Durbin-Watson test
##
## data:  m.div.dm.200kb
## DW = 1.4463, p-value = 3.114e-08
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.div.dm.200kb, nsim = 10000)
```

```
##
## Harrison-McCabe test
##
## data:  m.div.dm.200kb
## HMC = 0.56948, p-value = 0.9695
```

```
summary(m.div.dm.200kb)
```

```
##
## Call:
## lm(formula = diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
##     data = dm.lands.200kb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.481e-04 -6.976e-05 -4.240e-06  6.351e-05  4.418e-04
##
```

```
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  8.571e-03  7.361e-06 1164.402  <2e-16 ***
## thetaC      9.802e-01  2.925e-03  335.083  <2e-16 ***
## rhoC        1.652e-03  1.319e-03   1.252    0.211
## tmrcaC      1.104e-02  2.092e-04  52.753  <2e-16 ***
## thetaC:tmrcaC 8.772e-01  4.258e-02  20.604  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001143 on 357 degrees of freedom
## Multiple R-squared:  0.9985, Adjusted R-squared:  0.9985
## F-statistic: 6.074e+04 on 4 and 357 DF,  p-value: < 2.2e-16

# type 2 ANOVA
anova.diversity <- Anova(m.div.dm.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
anova.diversity

## Anova Table (Type II tests)
##
## Response: diversity
##              Sum Sq Df    F value    Pr(>F)    VarExp
## thetaC      0.00165091  1 126265.729 0.00000 0.96688
## rhoC        0.00000002  1    1.567 0.21147 0.00001
## tmrcaC      0.00004631  1  3541.894 0.00000 0.02712
## thetaC:tmrcaC 0.00000555  1   424.530 0.00000 0.00325
## Residuals    0.00000467 357                0.00273

# GLS
g.div.dm.200kb.1 <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                        data = dm.lands.200kb, weights = varPower(0, ~tmrcaC|chr), cor = corAR1(0, ~bin|chr))
g.div.dm.200kb.2 <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                        data = dm.lands.200kb, weights = varPower(0, ~thetaC|chr), cor = corAR1(0, ~bin|chr))
g.div.dm.200kb.3 <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                        data = dm.lands.200kb, weights = varPower(0, ~thetaC|chr), method = "ML")
g.div.dm.200kb.4 <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                        data = dm.lands.200kb, weights = varPower(0, ~tmrcaC|chr), method = "ML")

AIC(g.div.dm.200kb.1, g.div.dm.200kb.2, g.div.dm.200kb.3, g.div.dm.200kb.4)

##              df          AIC
## g.div.dm.200kb.1 11 -5583.726
## g.div.dm.200kb.2 11 -5586.917
## g.div.dm.200kb.3 10 -5561.165
## g.div.dm.200kb.4 10 -5559.348

summary(g.div.dm.200kb.1)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
## Data: dm.lands.200kb
```

```
##           AIC           BIC    logLik
##    -5583.726 -5540.918 2802.863
##
## Correlation Structure: AR(1)
## Formula: ~bin | chr
## Parameter estimate(s):
##      Phi
## 0.2843839
## Variance function:
## Structure: Power of variance covariate, different strata
## Formula: ~tmrcaC | chr
## Parameter estimates:
##           2L           2R           3L           3R
## -0.041145299 0.074090879 0.001954927 0.075773630
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept) 0.0085528 0.00000826 1035.3783 0.0000
## thetaC      0.9746366 0.00308268 316.1653 0.0000
## rhoC        0.0005062 0.00134208 0.3772 0.7063
## tmrcaC      0.0111855 0.00019829 56.4104 0.0000
## thetaC:tmrcaC 0.9085934 0.04279035 21.2336 0.0000
##
## Correlation:
##              (Intr) thetaC rhoC tmrcaC
## thetaC      0.133
## rhoC        -0.012 0.189
## tmrcaC      -0.315 -0.455 -0.439
## thetaC:tmrcaC -0.440 -0.196 -0.068 0.727
##
## Standardized residuals:
##           Min           Q1           Med           Q3           Max
## -2.9288615 -0.5661763 0.1198206 0.6423921 3.2163661
##
## Residual standard error: 0.0001234785
## Degrees of freedom: 362 total; 357 residual
```

```
vif(g.div.dm.200kb.1)
```

```
##           thetaC           rhoC           tmrcaC thetaC:tmrcaC
##           1.344036           1.505618           3.925968           2.712061
```

```
# Linear model without TMRCA --> rho becomes significant
```

```
g.div.dm.200kb.5 <- gls(diversity ~ (thetaC + rhoC),
                        data = dm.lands.200kb, weights = varPower(0, ~thetaC), cor = corAR1(0, ~bin), me
```

```
summary(g.div.dm.200kb.5)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC)
## Data: dm.lands.200kb
##           AIC           BIC    logLik
##    -4723.291 -4699.941 2367.645
##
## Correlation Structure: AR(1)
```

```
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## 0.4168254
## Variance function:
## Structure: Power of variance covariate
## Formula: ~thetaC
## Parameter estimates:
##      power
## 0.13171
##
## Coefficients:
##              Value      Std.Error    t-value p-value
## (Intercept) 0.0086889 0.000030456 285.29066      0
## thetaC      1.0809727 0.011736952  92.09995      0
## rhoC        0.0472191 0.003846123  12.27706      0
##
## Correlation:
##      (Intr) thetaC
## thetaC -0.016
## rhoC    0.048 -0.145
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -5.445432702 -0.523327124 -0.009243674  0.572229138  2.972069432
##
## Residual standard error: 0.0009029211
## Degrees of freedom: 362 total; 359 residual
```

### 3.3 1 Mb windows

```
# Chr 2L

# recombination landscapes
rho.dm.1Mb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.rho.1Mb.bedgraph", header = T)

# diversity
diversity.dm.1Mb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.diversity.1Mb.bedgraph", header = T)
diversity.dm.1Mb.2L$avg <- apply(diversity.dm.1Mb.2L[4:ncol(diversity.dm.1Mb.2L)], 1, mean)

# mutation landscapes
theta.dm.1Mb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.theta.1Mb.bedgraph", header = T)

# TMRCA landscapes
tmrca.dm.1Mb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.TMRCA.1Mb.bedgraph", header = T)
tmrca.dm.1Mb.2L$sample_mean <- apply(tmrca.dm.1Mb.2L[4:ncol(tmrca.dm.1Mb.2L)], 1, mean)

# missing data
missing.prop.1Mb.2L <- read.table("dm_data/dm_chr_maps/2L/dm_30x5x5.missing.prop.1Mb.bedgraph", header = T)
intersect.1Mb.2L <- apply(missing.prop.1Mb.2L[4:ncol(missing.prop.1Mb.2L)], 1, function(x) any(x > 0.1))

dm.lands.1Mb.2L <- as.data.frame(cbind(diversity.dm.1Mb.2L$chromStart,
                                       diversity.dm.1Mb.2L$chromEnd,
                                       diversity.dm.1Mb.2L$avg,
```

```

theta.dm.1Mb.2L$sample_mean,
rho.dm.1Mb.2L$sample_mean,
tmrca.dm.1Mb.2L$sample_mean))
names(dm.lands.1Mb.2L) <- c("start", "end", "diversity", "theta", "rho", "tmrca")
dm.lands.1Mb.2L$bin <- 1:nrow(dm.lands.1Mb.2L)

# filters based on missing data
dm.lands.1Mb.2L <- dm.lands.1Mb.2L[which(intersect.1Mb.2L == F),]

dm.lands.1Mb.2L$chr <- "2L"

dm.lands.1Mb.2L$thetaC <- dm.lands.1Mb.2L$theta - mean(dm.lands.1Mb.2L$theta)
dm.lands.1Mb.2L$tmrcaC <- dm.lands.1Mb.2L$tmrca - mean(dm.lands.1Mb.2L$tmrca)
dm.lands.1Mb.2L$rhoC <- dm.lands.1Mb.2L$rho - mean(dm.lands.1Mb.2L$rho)

g.div.dm.1Mb.2L <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
  data = dm.lands.1Mb.2L, weights = varPower(0, ~tmrcaC), cor = corAR1(0, ~bin), n

summary(g.div.dm.1Mb.2L)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
## Data: dm.lands.1Mb.2L
## AIC BIC logLik
## -299.3268 -291.7713 157.6634
##
## Correlation Structure: ARMA(1,0)
## Formula: ~bin
## Parameter estimate(s):
## Phi1
## 0.3144048
## Variance function:
## Structure: Power of variance covariate
## Formula: ~tmrcaC
## Parameter estimates:
## power
## -0.04167166
##
## Coefficients:
## Value Std.Error t-value p-value
## (Intercept) 0.0101124 0.00002513 402.3734 0.0000
## thetaC 0.9865547 0.02033816 48.5076 0.0000
## rhoC 0.0010558 0.00560117 0.1885 0.8532
## tmrcaC 0.0129654 0.00122549 10.5798 0.0000
## thetaC:tmrcaC 1.8344985 0.31522351 5.8197 0.0000
##
## Correlation:
## (Intr) thetaC rhoC tmrcaC
## thetaC -0.006
## rhoC 0.168 0.107
## tmrcaC -0.247 -0.591 -0.594
## thetaC:tmrcaC -0.447 0.063 -0.394 0.584
##
## Standardized residuals:

```

```

##           Min           Q1           Med           Q3           Max
## -1.4305539 -0.5991356 -0.1638086  0.5088896  2.4443876
##
## Residual standard error: 5.223258e-05
## Degrees of freedom: 19 total; 14 residual

# Chr 2R

# recombination landscapes
rho.dm.1Mb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.rho.1Mb.bedgraph", header = T)

# diversity
diversity.dm.1Mb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.diversity.1Mb.bedgraph", header = T)
diversity.dm.1Mb.2R$avg <- apply(diversity.dm.1Mb.2R[4:ncol(diversity.dm.1Mb.2R)], 1, mean)

# mutation landscapes
theta.dm.1Mb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.theta.1Mb.bedgraph", header = T)

# TMRCA landscapes
tmrca.dm.1Mb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.TMRCA.1Mb.bedgraph", header = T)
tmrca.dm.1Mb.2R$sample_mean <- apply(tmrca.dm.1Mb.2R[4:ncol(tmrca.dm.1Mb.2R)], 1, mean)

# missing data
missing.prop.1Mb.2R <- read.table("dm_data/dm_chr_maps/2R/dm_30x5x5.missing.prop.1Mb.bedgraph", header = T)
intersect.1Mb.2R <- apply(missing.prop.1Mb.2R[4:ncol(missing.prop.1Mb.2R)], 1, function(x) any(x > 0.1))

dm.lands.1Mb.2R <- as.data.frame(cbind(diversity.dm.1Mb.2R$chromStart,
                                     diversity.dm.1Mb.2R$chromEnd,
                                     diversity.dm.1Mb.2R$avg,
                                     theta.dm.1Mb.2R$sample_mean,
                                     rho.dm.1Mb.2R$sample_mean,
                                     tmrca.dm.1Mb.2R$sample_mean))
names(dm.lands.1Mb.2R) <- c("start", "end", "diversity", "theta", "rho", "tmrca")
dm.lands.1Mb.2R$bin <- 1:nrow(dm.lands.1Mb.2R)

# filters based on missing data
dm.lands.1Mb.2R <- dm.lands.1Mb.2R[which(intersect.1Mb.2R == F),]

dm.lands.1Mb.2R$chr <- "2R"

dm.lands.1Mb.2R$thetaC <- dm.lands.1Mb.2R$theta - mean(dm.lands.1Mb.2R$theta)
dm.lands.1Mb.2R$tmrcaC <- dm.lands.1Mb.2R$tmrca - mean(dm.lands.1Mb.2R$tmrca)
dm.lands.1Mb.2R$rhoC <- dm.lands.1Mb.2R$rho - mean(dm.lands.1Mb.2R$rho)

g.div.dm.1Mb.2R <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                      data = dm.lands.1Mb.2R, weights = varPower(0, ~tmrcaC), cor = corAR1(0, ~bin), method = "REML")

summary(g.div.dm.1Mb.2R)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
## Data: dm.lands.1Mb.2R
##      AIC      BIC    logLik
## -260.3174 -253.6517 138.1587
##

```

```
## Correlation Structure: AR(1)
## Formula: ~bin
## Parameter estimate(s):
##      Phi
## -0.02184561
## Variance function:
## Structure: Power of variance covariate
## Formula: ~tmrcaC
## Parameter estimates:
##      power
## -0.286626
##
## Coefficients:
##              Value Std.Error  t-value p-value
## (Intercept)  0.0091552 0.0000283 323.2402  0.0000
## thetaC       0.9718147 0.0212995  45.6262  0.0000
## rhoC         0.0083025 0.0084990   0.9769  0.3479
## tmrcaC       0.0103641 0.0019099   5.4266  0.0002
## thetaC:tmrcaC 0.5202501 0.5098563   1.0204  0.3277
##
## Correlation:
##      (Intr) thetaC rhoC  tmrcaC
## thetaC      -0.233
## rhoC         0.302 -0.140
## tmrcaC       -0.353 -0.638 -0.213
## thetaC:tmrcaC -0.803  0.231 -0.355  0.536
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.4712709 -0.4619228 -0.3324611  0.4352917  2.9199214
##
## Residual standard error: 1.478574e-05
## Degrees of freedom: 17 total; 12 residual

# recombination landscapes
rho.dm.1Mb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.rho.1Mb.bedgraph", header = T)

# diversity
diversity.dm.1Mb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.diversity.1Mb.bedgraph", header = T)
diversity.dm.1Mb.3L$avg <- apply(diversity.dm.1Mb.3L[4:ncol(diversity.dm.1Mb.3L)], 1, mean)

# mutation landscapes
theta.dm.1Mb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.theta.1Mb.bedgraph", header = T)

# TMRCA landscapes
tmrca.dm.1Mb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.TMRCA.1Mb.bedgraph", header = T)
tmrca.dm.1Mb.3L$sample_mean <- apply(tmrca.dm.1Mb.3L[4:ncol(tmrca.dm.1Mb.3L)], 1, mean)

# missing data
missing.prop.1Mb.3L <- read.table("dm_data/dm_chr_maps/3L/dm_30x5x5.missing.prop.1Mb.bedgraph", header = T)
intersect.1Mb.3L <- apply(missing.prop.1Mb.3L[4:ncol(missing.prop.1Mb.3L)], 1, function(x) any(x > 0.1))

dm.lands.1Mb.3L <- as.data.frame(cbind(diversity.dm.1Mb.3L$chromStart,
                                     diversity.dm.1Mb.3L$chromEnd,
```



```

diversity.dm.1Mb.3L$avg,
theta.dm.1Mb.3L$sample_mean,
rho.dm.1Mb.3L$sample_mean,
tmrca.dm.1Mb.3L$sample_mean))
names(dm.lands.1Mb.3L) <- c("start", "end", "diversity", "theta", "rho", "tmrca")
dm.lands.1Mb.3L$bin <- 1:nrow(dm.lands.1Mb.3L)

# filters based on missing data
dm.lands.1Mb.3L <- dm.lands.1Mb.3L[which(intersect.1Mb.3L == F),]

dm.lands.1Mb.3L$chr <- "3L"

dm.lands.1Mb.3L$thetaC <- dm.lands.1Mb.3L$theta - mean(dm.lands.1Mb.3L$theta)
dm.lands.1Mb.3L$tmrcaC <- dm.lands.1Mb.3L$tmrca - mean(dm.lands.1Mb.3L$tmrca)
dm.lands.1Mb.3L$rhoC <- dm.lands.1Mb.3L$rho - mean(dm.lands.1Mb.3L$rho)

g.div.dm.1Mb.3L <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
  data = dm.lands.1Mb.3L, weights = varPower(0, ~tmrcaC), cor = corAR1(0, ~bin), n

summary(g.div.dm.1Mb.3L)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
## Data: dm.lands.1Mb.3L
## AIC BIC logLik
## -348.8084 -340.0801 182.4042
##
## Correlation Structure: ARMA(1,0)
## Formula: ~bin
## Parameter estimate(s):
## Phi1
## 0.249026
## Variance function:
## Structure: Power of variance covariate
## Formula: ~tmrcaC
## Parameter estimates:
## power
## 0.3377385
##
## Coefficients:
## Value Std.Error t-value p-value
## (Intercept) 0.0091813 0.00002405 381.7940 0.0000
## thetaC 1.0032652 0.01244385 80.6233 0.0000
## rhoC 0.0070903 0.00396199 1.7896 0.0913
## tmrcaC 0.0084717 0.00094627 8.9528 0.0000
## thetaC:tmrcaC 0.3950013 0.16662491 2.3706 0.0298
##
## Correlation:
## (Intr) thetaC rhoC tmrcaC
## thetaC 0.400
## rhoC 0.531 0.320
## tmrcaC -0.622 -0.738 -0.571
## thetaC:tmrcaC -0.618 -0.348 -0.438 0.689

```

```
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.71883315 -0.45696303  0.01830762  0.69607904  1.69584501
##
## Residual standard error: 0.0002107777
## Degrees of freedom: 22 total; 17 residual

# Chr 3R

# recombination landscapes
rho.dm.1Mb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.rho.1Mb.bedgraph", header = T)

# diversity
diversity.dm.1Mb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.diversity.1Mb.bedgraph", header = T)
diversity.dm.1Mb.3R$avg <- apply(diversity.dm.1Mb.3R[4:ncol(diversity.dm.1Mb.3R)], 1, mean)

# mutation landscapes
theta.dm.1Mb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.theta.1Mb.bedgraph", header = T)

# TMRCA landscapes
tmrca.dm.1Mb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.TMRCA.1Mb.bedgraph", header = T)
tmrca.dm.1Mb.3R$sample_mean <- apply(tmrca.dm.1Mb.3R[4:ncol(tmrca.dm.1Mb.3R)], 1, mean)

# missing data
missing.prop.1Mb.3R <- read.table("dm_data/dm_chr_maps/3R/dm_30x5x5.missing.prop.1Mb.bedgraph", header = T)
intersect.1Mb.3R <- apply(missing.prop.1Mb[4:ncol(missing.prop.1Mb.3R)], 1, function(x) any(x > 0.1))

dm.lands.1Mb.3R <- as.data.frame(cbind(diversity.dm.1Mb.3R$chromStart,
                                     diversity.dm.1Mb.3R$chromEnd,
                                     diversity.dm.1Mb.3R$avg,
                                     theta.dm.1Mb.3R$sample_mean,
                                     rho.dm.1Mb.3R$sample_mean,
                                     tmrca.dm.1Mb.3R$sample_mean))
names(dm.lands.1Mb.3R) <- c("start", "end", "diversity", "theta", "rho", "tmrca")
dm.lands.1Mb.3R$bin <- 1:nrow(dm.lands.1Mb.3R)

# filters based on missing data
dm.lands.1Mb.3R <- dm.lands.1Mb.3R[which(intersect.1Mb.3R == F),]

dm.lands.1Mb.3R$chr <- "3R"

dm.lands.1Mb.3R$thetaC <- dm.lands.1Mb.3R$theta - mean(dm.lands.1Mb.3R$theta)
dm.lands.1Mb.3R$tmrcaC <- dm.lands.1Mb.3R$tmrca - mean(dm.lands.1Mb.3R$tmrca)
dm.lands.1Mb.3R$rhoC <- dm.lands.1Mb.3R$rho - mean(dm.lands.1Mb.3R$rho)

g.div.dm.1Mb.3R <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                      data = dm.lands.1Mb.3R, weights = varPower(0, ~tmrcaC), cor = corAR1(0, ~bin), method = "REML")

summary(g.div.dm.1Mb.3R)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
## Data: dm.lands.1Mb.3R
##      AIC      BIC    logLik
```

```
## -324.5017 -316.9462 170.2509
##
## Correlation Structure: ARMA(1,0)
## Formula: ~bin
## Parameter estimate(s):
## Phi1
## -0.4823526
## Variance function:
## Structure: Power of variance covariate
## Formula: ~tmrcaC
## Parameter estimates:
## power
## 0.8128992
##
## Coefficients:
## Value Std.Error t-value p-value
## (Intercept) 0.0067246 0.00002252 298.66683 0.0000
## thetaC 0.9664707 0.00658600 146.74629 0.0000
## rhoC 0.0031144 0.00241054 1.29200 0.2173
## tmrcaC 0.0094832 0.00049711 19.07661 0.0000
## thetaC:tmrcaC 0.7457601 0.17910164 4.16389 0.0010
##
## Correlation:
## (Intr) thetaC rhoC tmrcaC
## thetaC 0.911
## rhoC -0.280 0.066
## tmrcaC -0.923 -0.944 -0.023
## thetaC:tmrcaC -0.940 -0.881 0.320 0.856
##
## Standardized residuals:
## Min Q1 Med Q3 Max
## -1.7308894 -0.4913490 0.1058365 0.6374128 1.9989783
##
## Residual standard error: 0.0004245867
## Degrees of freedom: 19 total; 14 residual

# all together now
dm.lands.1Mb <- rbind.data.frame(dm.lands.1Mb.2L, dm.lands.1Mb.2R, dm.lands.1Mb.3L, dm.lands.1Mb.3R)

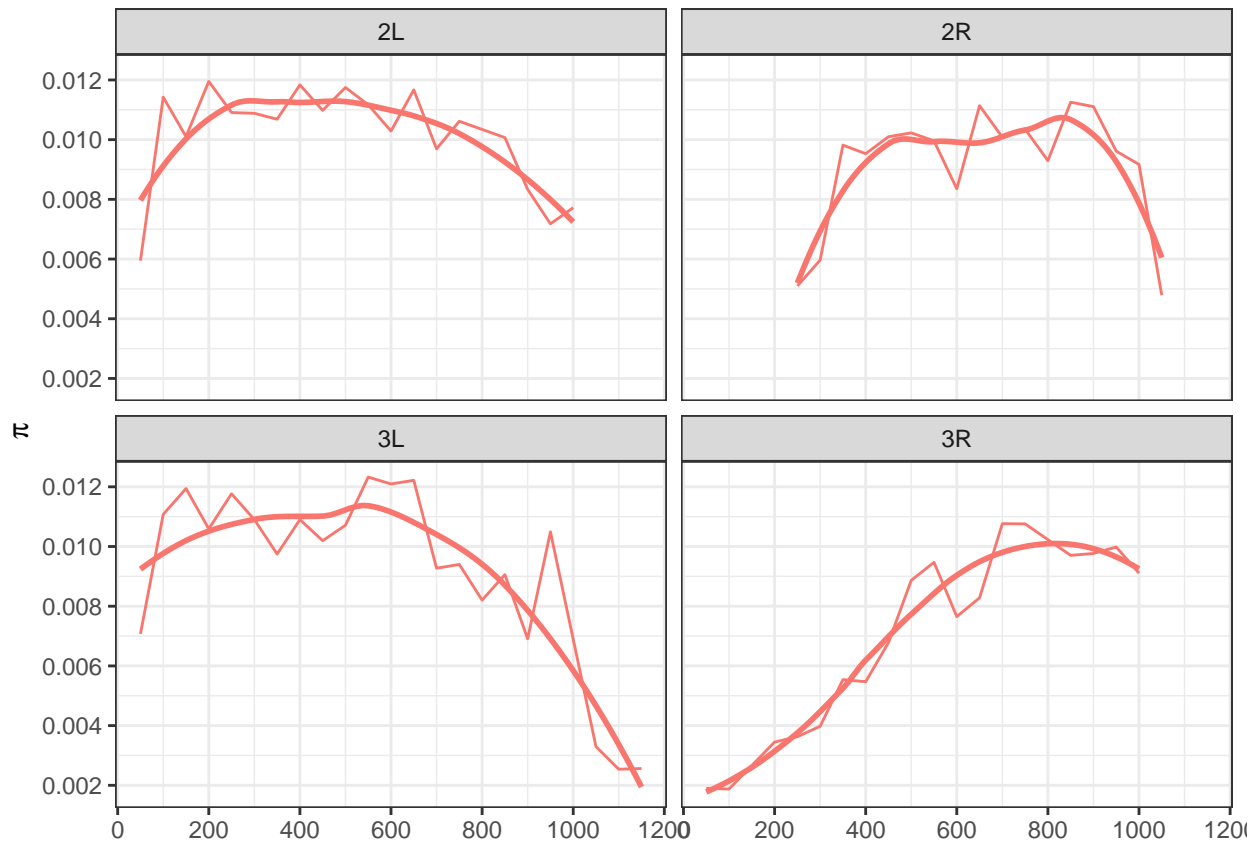
write.table(dm.lands.1Mb, "dm_data/dm_chr_maps/dm.maps.1Mb.tsv", quote = F, sep = "\t", row.names = F,

# Plots
scale.3d <- function(x) sprintf("%.3f", x) # digits shown in y axis

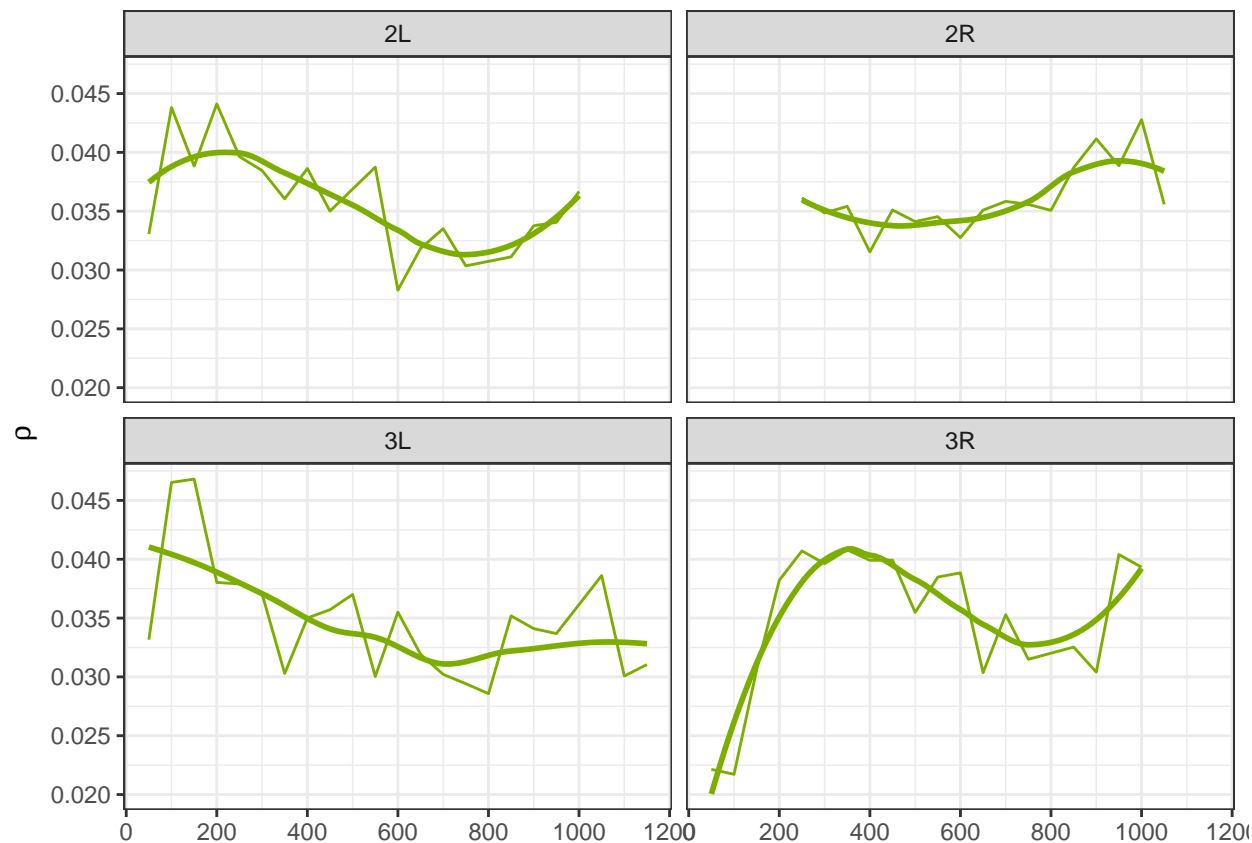
setTimeLimit(cpu = Inf, elapsed = Inf, transient = F)

molten.diversity <- melt(dm.lands.1Mb[c(3,7,8)], id.vars = c("bin", "chr"))
diversity.map <- ggplot(data = molten.diversity, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
diversity.map <- diversity.map + geom_line(data = molten.diversity, colour = "#F8766D")
diversity.map <- diversity.map + geom_smooth(method = "loess", se = F, colour = "#F8766D")
diversity.map <- diversity.map + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
diversity.map <- diversity.map + labs(title = NULL, x = NULL, y = expression(pi))
diversity.map <- diversity.map + theme(axis.title.x = element_text(size = 20), axis.title.y = element_text(size = 20))
```

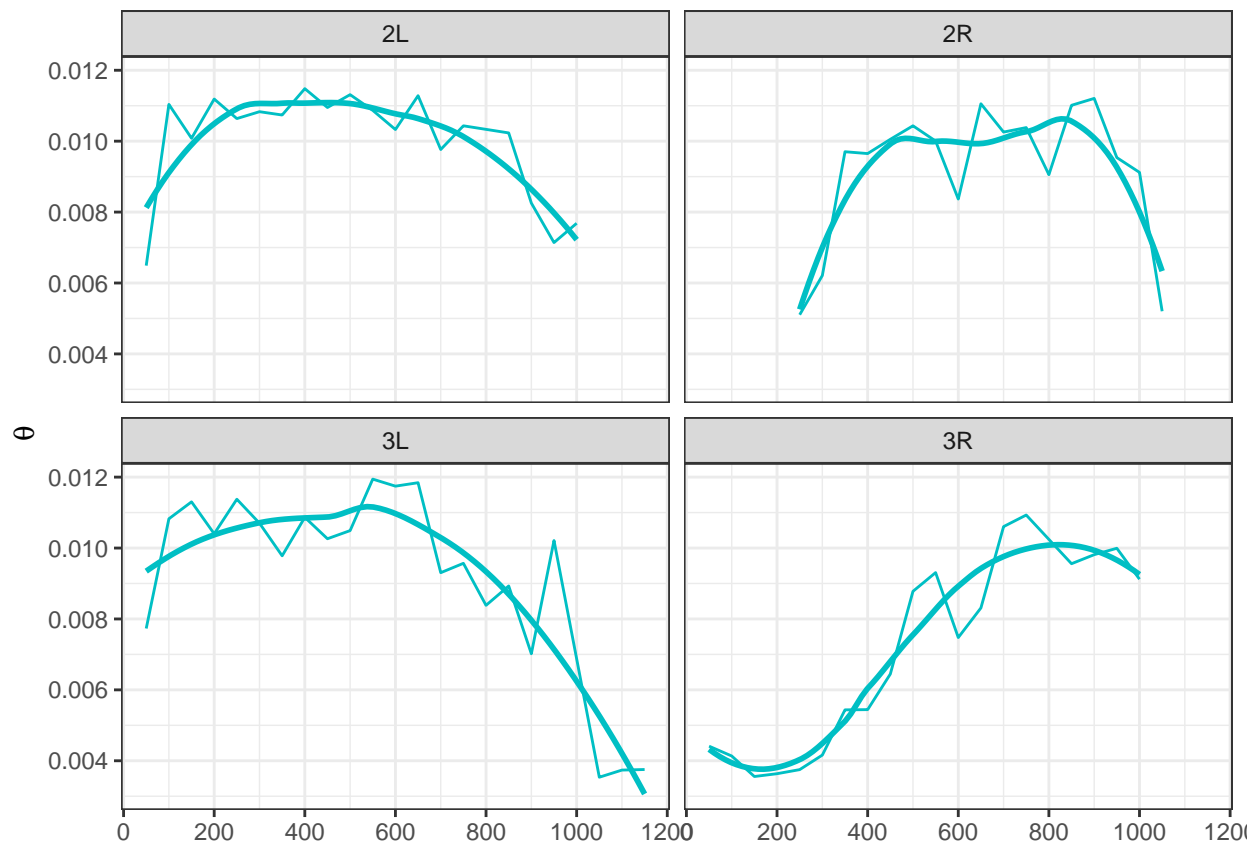
diversity.map



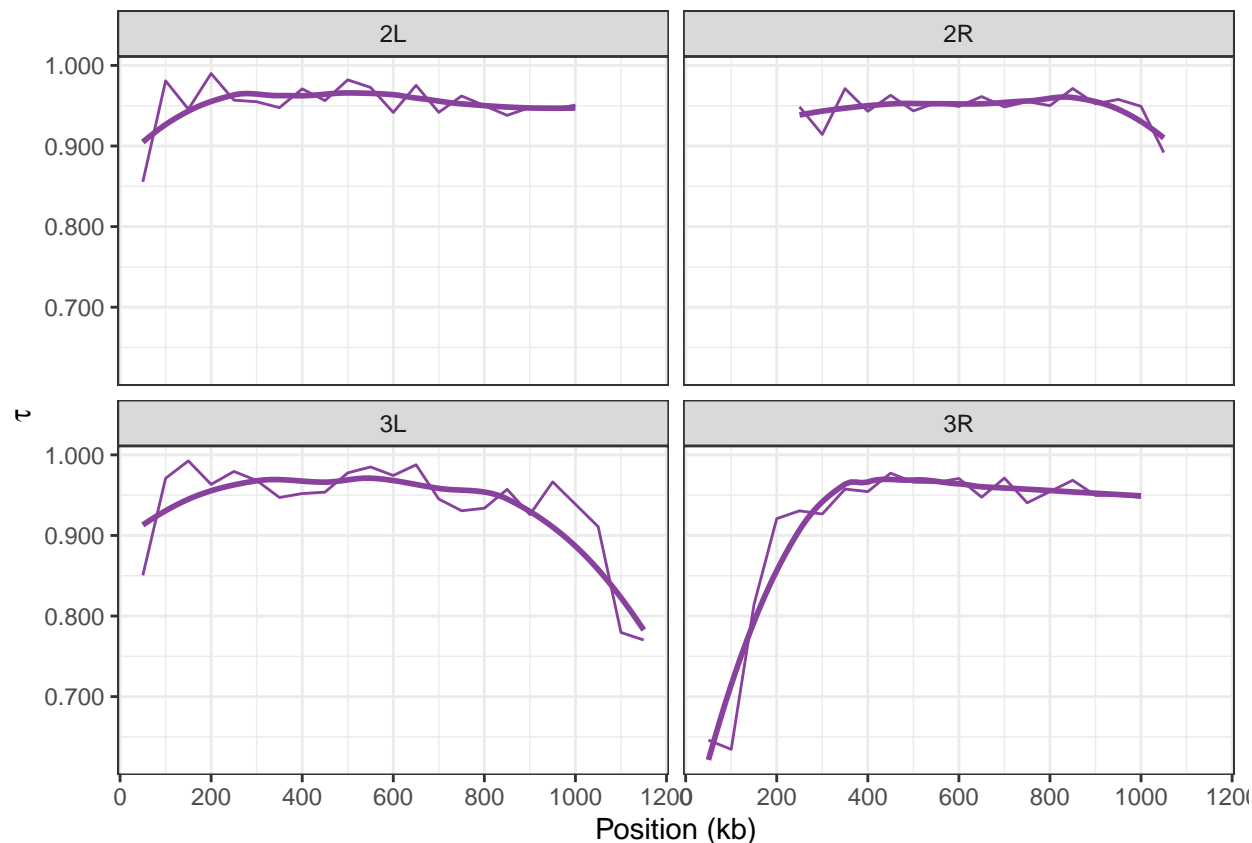
```
molten.rho <- melt(dm.lands.1Mb[c(5,7,8)], id.vars = c("bin", "chr"))
rho.map <- ggplot(data = molten.rho, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
rho.map <- rho.map + geom_line(data = molten.rho, colour = "#7CAE00")
rho.map <- rho.map + geom_smooth(method = "loess", se = F, colour = "#7CAE00")
rho.map <- rho.map + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
rho.map <- rho.map + labs(title = NULL, x = NULL, y = expression(rho))
rho.map <- rho.map + theme(axis.title.x = element_text(size = 20), axis.title.y = element_text(size = 20))
rho.map
```



```
molten.theta <- melt(dm.lands.1Mb[c(4,7,8)], id.vars = c("bin", "chr"))
theta.map <- ggplot(data = molten.theta, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
theta.map <- theta.map + geom_line(data = molten.theta, colour = "#00BFC4")
theta.map <- theta.map + geom_smooth(method = "loess", se = F, colour = "#00BFC4")
theta.map <- theta.map + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
theta.map <- theta.map + labs(title = NULL, x = NULL, y = expression(theta))
theta.map <- theta.map + theme(axis.title.x = element_text(size = 20), axis.title.y = element_text(size = 20))
theta.map
```



```
molten.tmrca <- melt(dm.lands.1Mb[c(6,7,8)], id.vars = c("bin", "chr"))
tmrca.map <- ggplot(data = molten.tmrca, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
tmrca.map <- tmrca.map + geom_line(data = molten.tmrca, colour = "#88419d")
tmrca.map <- tmrca.map + geom_smooth(method = "loess", se = F, colour = "#88419d")
tmrca.map <- tmrca.map + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
tmrca.map <- tmrca.map + labs(title = NULL, x = "Position (kb)", y = expression(tau))
tmrca.map <- tmrca.map + theme(axis.title.x = element_text(size = 20), axis.title.y = element_text(size = 20))
tmrca.map
```



```
# genome-wide correlations
```

```
cor.test(~rho + diversity, data = dm.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: rho and diversity
## S = 60404, p-value = 0.07236
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.2060045
```

```
cor.test(~rho + tmrca, data = dm.lands.1Mb, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: rho and tmrca
## S = 42930, p-value = 8.798e-05
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.4356959
```

```
cor.test(~theta + tmrca, data = dm.lands.1Mb, method = "spearman")
```

```
##
```

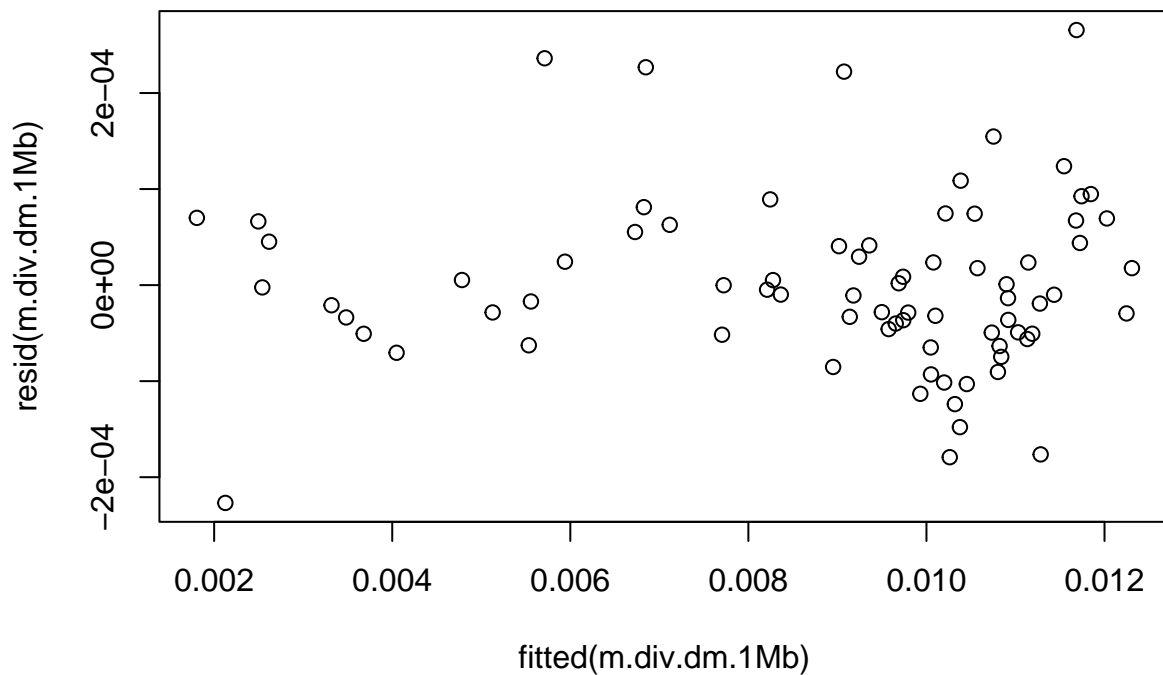
```
## Spearman's rank correlation rho
##
## data: theta and tmrca
## S = 22964, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.698144

# Linear models
# centering
dm.lands.1Mb$thetaC <- dm.lands.1Mb$theta - mean(dm.lands.1Mb$theta)
dm.lands.1Mb$tmrcaC <- dm.lands.1Mb$tmrca - mean(dm.lands.1Mb$tmrca)
dm.lands.1Mb$rhoC <- dm.lands.1Mb$rho - mean(dm.lands.1Mb$rho)

dm.lands.1Mb$bin <- 1:nrow(dm.lands.1Mb)

m.div.dm.1Mb <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC), data = dm.lands.1Mb)

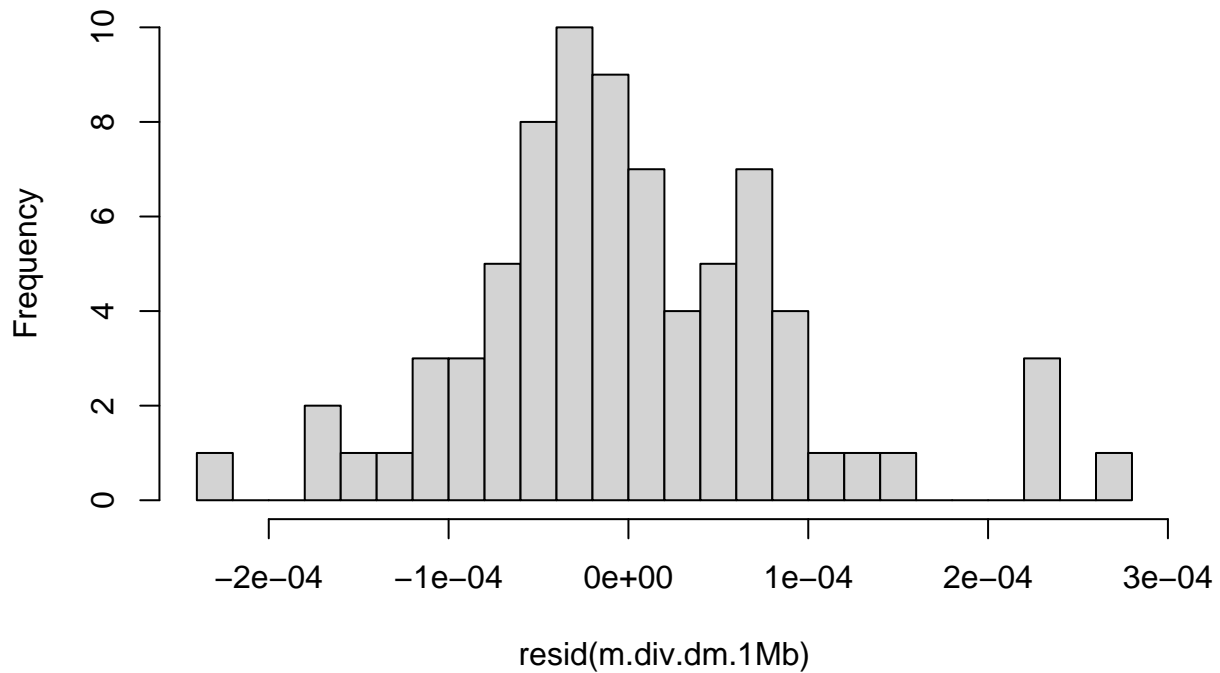
plot(resid(m.div.dm.1Mb)~fitted(m.div.dm.1Mb))
```



```
hist(resid(m.div.dm.1Mb), nclass = 30)
```



## Histogram of resid(m.div.dm.1Mb)



```
dwtest(m.div.dm.1Mb)
```

```
##
## Durbin-Watson test
##
## data: m.div.dm.1Mb
## DW = 1.6563, p-value = 0.03648
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcTest(m.div.dm.1Mb, nsim = 10000)
```

```
##
## Harrison-McCabe test
##
## data: m.div.dm.1Mb
## HMC = 0.72492, p-value = 0.9989
```

```
summary(m.div.dm.1Mb)
```

```
##
## Call:
## lm(formula = diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
##     data = dm.lands.1Mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.268e-04 -5.063e-05 -1.004e-05  4.519e-05  2.656e-04
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.808e-03  1.551e-05  568.036 < 2e-16 ***
## thetaC      9.919e-01  6.725e-03  147.497 < 2e-16 ***
## rhoC        6.510e-03  3.034e-03   2.145  0.0353 *
## tmrcaC      9.525e-03  6.035e-04  15.784 < 2e-16 ***
## thetaC:tmrcaC 4.881e-01  1.131e-01   4.315 4.99e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.283e-05 on 72 degrees of freedom
## Multiple R-squared:  0.999, Adjusted R-squared:  0.9989
## F-statistic: 1.74e+04 on 4 and 72 DF, p-value: < 2.2e-16

# type 2 ANOVA
anova.diversity <- Anova(m.div.dm.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
anova.diversity

## Anova Table (Type II tests)
##
## Response: diversity
##              Sum Sq Df    F value    Pr(>F)    VarExp
## thetaC      2.4658e-04  1 28614.4891 0.000000 0.97371
## rhoC        4.0000e-08  1    4.6028 0.035293 0.00016
## tmrcaC      5.8360e-06  1   677.2706 0.000000 0.02305
## thetaC:tmrcaC 1.6000e-07  1   18.6226 0.000050 0.00063
## Residuals    6.2000e-07 72
##              0.00245

# GLS
g.div.dm.1Mb.1 <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                      data = dm.lands.1Mb, weights = varPower(0, ~tmrcaC|chr), cor = corAR1(0, ~bin|chr))
g.div.dm.1Mb.2 <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                      data = dm.lands.1Mb, weights = varPower(0, ~theta|chr), cor = corAR1(0, ~bin|chr))
g.div.dm.1Mb.3 <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                      data = dm.lands.1Mb, weights = varPower(0, ~theta|chr), method = "ML")
g.div.dm.1Mb.4 <- gls(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC),
                      data = dm.lands.1Mb, weights = varPower(0, ~tmrcaC|chr), method = "ML")

AIC(g.div.dm.1Mb.1, g.div.dm.1Mb.2, g.div.dm.1Mb.3, g.div.dm.1Mb.4)

##              df          AIC
## g.div.dm.1Mb.1 11 -1228.719
## g.div.dm.1Mb.2 11 -1202.198
## g.div.dm.1Mb.3 10 -1202.415
## g.div.dm.1Mb.4 10 -1222.559

summary(g.div.dm.1Mb.1)

## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC)
## Data: dm.lands.1Mb
```

```
##           AIC           BIC    logLik
##    -1228.719 -1202.937 625.3593
##
## Correlation Structure: AR(1)
## Formula: ~bin | chr
## Parameter estimate(s):
##      Phi
## 0.4040379
## Variance function:
## Structure: Power of variance covariate, different strata
## Formula: ~tmrcaC | chr
## Parameter estimates:
##      2L      2R      3L      3R
## 0.3814218 0.3225539 0.4579989 0.5841944
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  0.0087611 0.00000890 984.6353 0.0000
## thetaC       0.9830140 0.00395599 248.4877 0.0000
## rhoC         0.0012460 0.00207145   0.6015 0.5494
## tmrcaC       0.0103831 0.00051388  20.2055 0.0000
## thetaC:tmrcaC 0.6061715 0.12215917   4.9621 0.0000
##
## Correlation:
##      (Intr) thetaC rhoC   tmrcaC
## thetaC      0.225
## rhoC         0.466 0.603
## tmrcaC      -0.364 -0.486 -0.568
## thetaC:tmrcaC -0.286 -0.210 -0.299 0.636
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.6371467 -0.3675723 0.2903415 0.9878831 2.4854662
##
## Residual standard error: 0.0004263998
## Degrees of freedom: 77 total; 72 residual
```

```
vif(g.div.dm.1Mb.1)
```

```
##      thetaC      rhoC      tmrcaC thetaC:tmrcaC
##      1.674634      1.863553      2.423500      1.719225
```

```
# Linear model without TMRCA --> rho becomes significant
```

```
g.div.dm.1Mb.5 <- gls(diversity ~ (thetaC + rhoC),
                      data = dm.lands.1Mb, weights = varPower(0, ~thetaC|chr), cor = corAR1(0, ~bin|chr))
```

```
summary(g.div.dm.1Mb.5)
```

```
## Generalized least squares fit by maximum likelihood
## Model: diversity ~ (thetaC + rhoC)
## Data: dm.lands.1Mb
##           AIC      BIC    logLik
##    -1055.094 -1034 536.5469
##
## Correlation Structure: AR(1)
```

```
## Formula: ~bin | chr
## Parameter estimate(s):
##      Phi
## 0.5243617
## Variance function:
## Structure: Power of variance covariate, different strata
## Formula: ~thetaC | chr
## Parameter estimates:
##      2L      2R      3L      3R
## 0.2518860 0.2065555 0.2564038 0.1788172
##
## Coefficients:
##              Value   Std.Error   t-value p-value
## (Intercept) 0.0089103 0.000042565 209.33432     0
## thetaC      1.1056720 0.018207498  60.72619     0
## rhoC        0.0395498 0.006816697   5.80191     0
##
## Correlation:
##      (Intr) thetaC
## thetaC -0.115
## rhoC    0.055 -0.292
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -3.46396737 -0.69180396 -0.06437384  0.58736967  1.68761843
##
## Residual standard error: 0.001151706
## Degrees of freedom: 77 total; 74 residual
```

### 3.4 Evolutionary (Protein) Rates

This is how we generated the table from raw data:

```
# loads data
dm.raw <- read.table("dm_data/dm_misc/dpgp2_Dsimulans_bpp.all.csv", header = TRUE, sep = "\t") %>%
  select(GeneID, MeanNumberSynPos, PiN, PiS, MeanNumberSynPosDiv, dS, dN) %>% na.omit()

# summarises
dm.tbl <- dplyr::summarize(dm.raw,
  MeanNumberNonSynPos = sum(3 - MeanNumberSynPos),
  MeanNumberSynPos = sum(MeanNumberSynPos),
  PiN = sum(PiN),
  PiS = sum(PiS),
  MeanNumberNonSynPosDiv = sum(3 - MeanNumberSynPosDiv),
  MeanNumberSynPosDiv = sum(MeanNumberSynPosDiv),
  dN = sum(dN),
  dS = sum(dS))

dm.tbl$PiN <- (dm.tbl$PiN / dm.tbl$MeanNumberNonSynPos)
dm.tbl$PiS <- (dm.tbl$PiS / dm.tbl$MeanNumberSynPos)
dm.tbl$dN <- (dm.tbl$dN / dm.tbl$MeanNumberNonSynPosDiv)
dm.tbl$dS <- (dm.tbl$dS / dm.tbl$MeanNumberSynPosDiv)

# computes ratios
dm.tbl$PiN_PiS <- with(dm.tbl, PiN / PiS)
```

```

dm.tbl$dN_dS <- with(dm.tbl, dN / dS)

# we ignore genes with ratios above 1
dm.tbl <- dm.tbl[dm.tbl$PiN_PiS < 1,]
dm.tbl <- dm.tbl[dm.tbl$dN_dS < 1,]
dm.tbl <- na.omit(dm.tbl)

dm.genes.coord <- read.table("dm_data/dm_misc/Dmel_trans_nooverlaps-final.txt", header = F) %>% select(
names(dm.genes.coord) <- c("chr", "start", "end", "GeneID")

dm.evol <- merge(dm.genes.coord, dm.tbl, by = "GeneID")
dm.evol <- arrange(dm.evol, chr, start)

# reorder maps to use GR
dm.maps.50kb <- dm.lands.50kb[,c(8,1:6)]

# mapping genes to 50kb windows in ismc maps
dm.lands.gr <- makeGRangesFromDataFrame(dm.maps.50kb)
values(dm.lands.gr) <- select(dm.maps.50kb, diversity, theta, rho, tmrca)
evolrate.gr <- makeGRangesFromDataFrame(dm.evol)
values(evolrate.gr) <- select(dm.evol, PiN, PiS, dN, dS)

hits_1 <- findOverlaps(query = evolrate.gr, subject = dm.lands.gr, type = "within")

evolrate.gr.df <- as.data.frame(evolrate.gr[queryHits(hits_1)], row.names = NULL)
dm.lands.gr.df <- as.data.frame(dm.lands.gr[subjectHits(hits_1)], row.names = NULL)

dm.lands.evolrate <- cbind.data.frame(dm.lands.gr.df[,c(1:3,6:9)], evolrate.gr.df[,c(2:4,6:9)])

names(dm.lands.evolrate)[1] <- "chr"
names(dm.lands.evolrate)[2] <- "start.window"
names(dm.lands.evolrate)[3] <- "end.window"
names(dm.lands.evolrate)[8] <- "start.gene"
names(dm.lands.evolrate)[9] <- "end.gene"
names(dm.lands.evolrate)[10] <- "length.gene"

dm.lands.evolrate <- arrange(dm.lands.evolrate, chr, start.window, start.gene)

dm.lands.evolrate.bins <- ddply(dm.lands.evolrate, .variables = c("chr", "start.window"), .fun = plyr::
  diversity = mean(diversity),
  theta = mean(theta),
  rho = mean(rho),
  tmrca = mean(tmrca),
  exonic.sites = sum(length.gene),
  PiN.window = sum(PiN*length.gene) / sum(length.gene),
  PiS.window = sum(PiS*length.gene) / sum(length.gene),
  dN.window = sum(dN*length.gene) / sum(length.gene),
  dS.window = sum(dS*length.gene) / sum(length.gene))

dm.lands.evolrate.bins$bin <- 1:nrow(dm.lands.evolrate.bins)

write.table(dm.lands.evolrate.bins, "dm_data/dm_chr_maps/dm_maps_50kb_protein_rates.tsv",

```

```
sep = "\t", quote = F, col.names = T, row.names = F)
```

We can read the table generated and move on with the analyses:

```
dm.lands.evolrate.bins <- read.table("dm_data/dm_chr_maps/dm_maps_50kb_protein_rates.tsv", header = T,

# linear model in windows with more than 20kb exonic sites
dm.lands.genes <- filter(dm.lands.evolrate.bins, exonic.sites > 2e+4)

# centering
dm.lands.genes$thetaC <- dm.lands.genes$theta - mean(dm.lands.genes$theta)
dm.lands.genes$tmrcaC <- dm.lands.genes$tmrca - mean(dm.lands.genes$tmrca)
dm.lands.genes$rhoC <- dm.lands.genes$rho - mean(dm.lands.genes$rho)

m.dm.cds.1 <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC*tmrcaC), data = dm.lands.genes)
m.dm.cds.2 <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC*tmrcaC + rhoC*tmrcaC), data = dm.lands.genes)

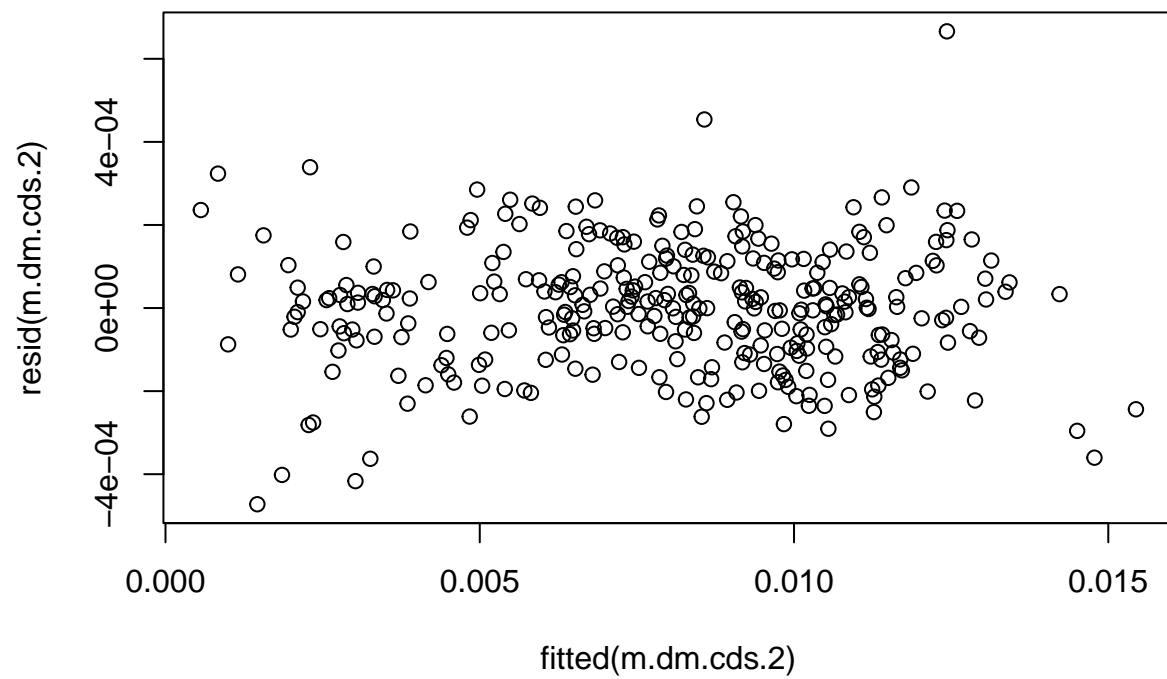
AIC(m.dm.cds.1, m.dm.cds.2)
```

```
##           df          AIC
## m.dm.cds.1  6 -4778.177
## m.dm.cds.2  7 -4782.425
```

```
summary(m.dm.cds.2)
```

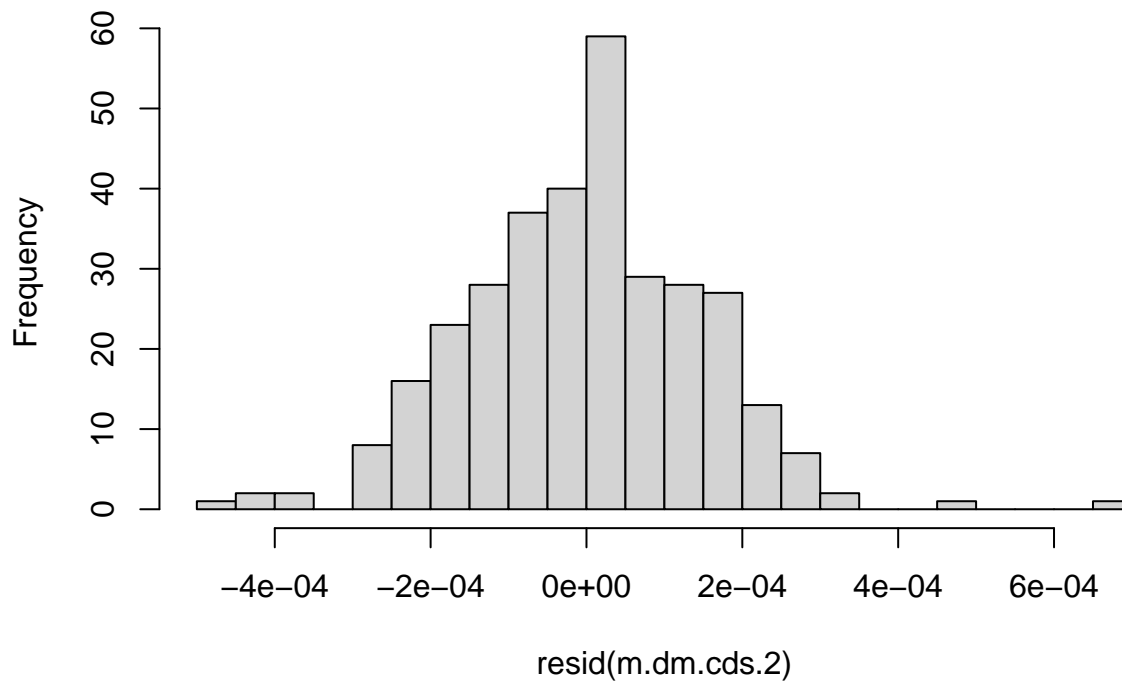
```
##
## Call:
## lm(formula = diversity ~ (thetaC + rhoC + tmrcaC + thetaC * tmrcaC +
##      rhoC * tmrcaC), data = dm.lands.genes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.723e-04 -9.596e-05  3.810e-06  8.994e-05  6.661e-04
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.890e-03  9.716e-06  812.013  <2e-16 ***
## thetaC       9.511e-01  3.665e-03  259.551  <2e-16 ***
## rhoC        1.053e-03  1.530e-03   0.688   0.4919
## tmrcaC      1.120e-02  1.964e-04  57.010  <2e-16 ***
## thetaC:tmrcaC 1.079e+00  4.299e-02  25.098  <2e-16 ***
## rhoC:tmrcaC   2.110e-02  8.481e-03   2.488   0.0133 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.000149 on 318 degrees of freedom
## Multiple R-squared:  0.9977, Adjusted R-squared:  0.9977
## F-statistic: 2.785e+04 on 5 and 318 DF, p-value: < 2.2e-16
```

```
# picking best OLS for ANOVA
plot(resid(m.dm.cds.2)~fitted(m.dm.cds.2))
```



```
hist(resid(m.dm.cds.2), nclass = 30)
```

## Histogram of resid(m.dm.cds.2)



```
dwtest(m.dm.cds.2)
```

```
##
## Durbin-Watson test
##
## data: m.dm.cds.2
## DW = 2.0035, p-value = 0.4719
## alternative hypothesis: true autocorrelation is greater than 0
```

```
hmcetest(m.dm.cds.2, nsim = 10000)
```

```
##
## Harrison-McCabe test
##
## data: m.dm.cds.2
## HMC = 0.41117, p-value = 0.0114
```

```
summary(m.dm.cds.2)
```

```
##
## Call:
## lm(formula = diversity ~ (thetaC + rhoC + tmrcaC + thetaC * tmrcaC +
##   rhoC * tmrcaC), data = dm.lands.genes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.723e-04 -9.596e-05  3.810e-06  8.994e-05  6.661e-04
##
```



```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.890e-03  9.716e-06 812.013  <2e-16 ***
## thetaC      9.511e-01  3.665e-03 259.551  <2e-16 ***
## rhoC        1.053e-03  1.530e-03   0.688   0.4919
## tmrcaC      1.120e-02  1.964e-04  57.010  <2e-16 ***
## thetaC:tmrcaC 1.079e+00  4.299e-02  25.098  <2e-16 ***
## rhoC:tmrcaC  2.110e-02  8.481e-03   2.488   0.0133 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.000149 on 318 degrees of freedom
## Multiple R-squared:  0.9977, Adjusted R-squared:  0.9977
## F-statistic: 2.785e+04 on 5 and 318 DF,  p-value: < 2.2e-16

# type 2 anova
anova.diversity.cds <- Anova(m.dm.cds.2)
apiss <- anova.diversity.cds$"Sum Sq"
anova.diversity.cds$VarExp <- apiss / sum(apiss)

anova.diversity.cds

## Anova Table (Type II tests)
##
## Response: diversity
##               Sum Sq Df    F value    Pr(>F)    VarExp
## thetaC        1.544e-03  1 69536.4716 0.000000 0.93236
## rhoC          5.000e-08  1   2.1199 0.146386 0.00003
## tmrcaC        9.078e-05  1  4088.1586 0.000000 0.05482
## thetaC:tmrcaC 1.399e-05  1   629.9096 0.000000 0.00845
## rhoC:tmrcaC   1.400e-07  1    6.1917 0.013347 0.00008
## Residuals     7.060e-06 318
##                                0.00426

# simple correlations
cor.test(dm.lands.evolrate.bins$theta, dm.lands.evolrate.bins$exonic.sites, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  dm.lands.evolrate.bins$theta and dm.lands.evolrate.bins$exonic.sites
## S = 339155743, p-value = 6.388e-06
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##          rho
## -0.1289611

cor.test(dm.lands.evolrate.bins$tmrca, dm.lands.evolrate.bins$exonic.sites, method = "spearman")

##
## Spearman's rank correlation rho
##
## data:  dm.lands.evolrate.bins$tmrca and dm.lands.evolrate.bins$exonic.sites
## S = 360421366, p-value = 2.034e-12
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##          rho
```

```

## -0.1997488
pcor.test(dm.lands.evolrate.bins$tmrca, dm.lands.evolrate.bins$exonic.sites, dm.lands.evolrate.bins$theta)

##      estimate      p.value statistic      n gp  Method
## 1 -0.1582194 2.912866e-08 -5.583083 1217 1 spearman
pcor.test(dm.lands.evolrate.bins$theta, dm.lands.evolrate.bins$exonic.sites, dm.lands.evolrate.bins$tmrca)

##      estimate      p.value statistic      n gp  Method
## 1 -0.0374759 0.1915716 -1.306672 1217 1 spearman
cor.test(dm.lands.evolrate.bins$PiN, dm.lands.evolrate.bins$theta, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: dm.lands.evolrate.bins$PiN and dm.lands.evolrate.bins$theta
## S = 192322126, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.3598097
cor.test(dm.lands.evolrate.bins$PiS, dm.lands.evolrate.bins$theta, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: dm.lands.evolrate.bins$PiS and dm.lands.evolrate.bins$theta
## S = 78488708, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.7387315
pcor.test(dm.lands.evolrate.bins$PiS, dm.lands.evolrate.bins$theta, dm.lands.evolrate.bins$dS, method = "spearman")

##      estimate      p.value statistic      n gp  Method
## 1 0.7012704 1.221182e-180 34.27434 1217 1 spearman
# bootstrapping PiN and PiS correlations
boot.tbl <- dm.lands.evolrate.bins %>% select(PiN.window, theta, PiS.window)
boot.PiN.theta <- boot(boot.tbl,
  statistic = function(boot.tbl, i) {
    cor.test(boot.tbl[i, "PiN.window"], boot.tbl[i, "theta"], method='spearman')$estimate
  },
  R = 10000
)
boot.ci(boot.PiN.theta, type = "basic")

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 10000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = boot.PiN.theta, type = "basic")
##
## Intervals :

```

```
## Level      Basic
## 95%      ( 0.3099, 0.4098 )
## Calculations and Intervals on Original Scale

boot.PiS.theta <- boot(boot.tbl,
  statistic = function(boot.tbl, i) {
    cor.test(boot.tbl[i, "PiS.window"], boot.tbl[i, "theta"], method='spearman')$estimate
  },
  R = 10000
)
boot.ci(boot.PiS.theta, type = "basic")

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 10000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = boot.PiS.theta, type = "basic")
##
## Intervals :
## Level      Basic
## 95%      ( 0.7092, 0.7719 )
## Calculations and Intervals on Original Scale
```

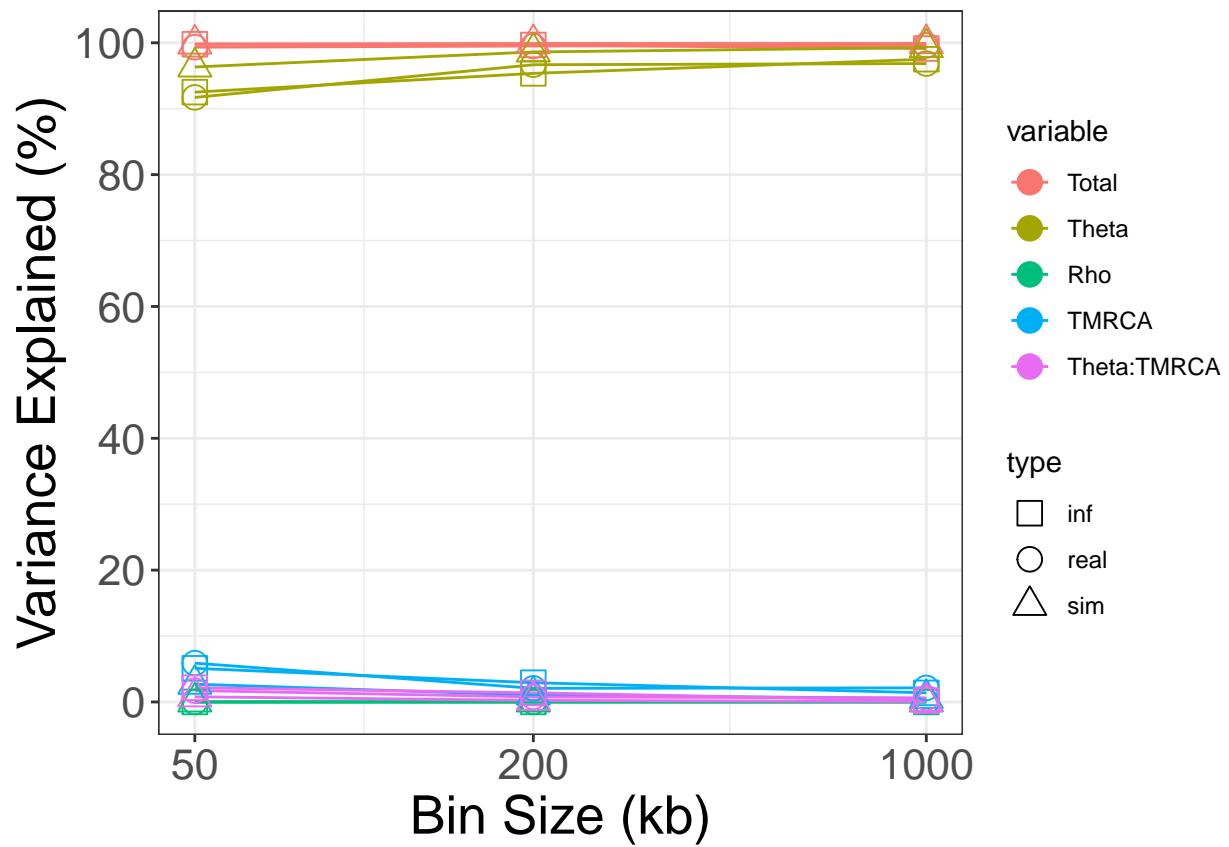
#### 4 R<sup>2</sup> plot with real, simulated (true) and simulated (inferred) data.

```
# true landscapes
r2.sim.avg <- rbind.data.frame(r2.sim.50kb$average, r2.sim.200kb$average, r2.sim.1Mb$average, make.row.names = FALSE)
colnames(r2.sim.avg) <- c("Total", "Theta", "Rho", "TMRCA", "Theta:TMRCA")
r2.sim.avg$bin.size <- c(50, 200, 1000)

# inferred landscapes
r2.inf.avg <- rbind.data.frame(r2.inf.50kb$average, r2.inf.200kb$average, r2.inf.1Mb$average, make.row.names = FALSE)
colnames(r2.inf.avg) <- c("Total", "Theta", "Rho", "TMRCA", "Theta:TMRCA")
r2.inf.avg$bin.size <- c(50, 200, 1000)

r2.tab.comb <- rbind.data.frame(r2.dm.tab, r2.sim.avg, r2.inf.avg)
r2.tab.comb$type <- c(rep("real", 3), rep("sim", 3), rep("inf", 3))

molten.r2 <- melt(r2.tab.comb, id.vars = c("bin.size", "type"))
r2.plot <- ggplot(data = molten.r2, aes(x = bin.size, y = value, colour = variable, fill = type))
r2.plot <- r2.plot + geom_line(data = molten.r2)
r2.plot <- r2.plot + geom_point(aes(shape = type, colour = variable), size = 4)
r2.plot <- r2.plot + scale_x_continuous(breaks = c(50, 200, 1000), trans="log10")
r2.plot <- r2.plot + scale_y_continuous(breaks = pretty_breaks())
r2.plot <- r2.plot + scale_shape_manual(values = c(0, 1, 2))
r2.plot <- r2.plot + labs(title = NULL, x = "Bin Size (kb)", y = "Variance Explained (%)") + theme_bw()
r2.plot <- r2.plot + theme(axis.title = element_text(size = 20), axis.text = element_text(size = 16))
r2.plot
```



```
ggsave("Fig4.pdf", r2.plot, device = "pdf", )
```

```
# Simulations with BGS
```