

What shapes genome-wide diversity?

A case study in *Drosophila melanogaster*

Gustavo V. Barroso and Julien Y. Dutheil

January 1st 2023

Contents

1 Drosophila-like neutral simulations of chr 2L (True Landscapes)	3
1.1 50 kb scale	3
1.1.1 Replicate 1	4
1.1.2 Replicate 2	8
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	32
1.1.11 all replicates:	35
1.2 200 kb scale	38
1.2.1 Replicate 1	38
1.2.2 Replicate 2	42
1.2.3 Replicate 3	45
1.2.4 Replicate 4	48
1.2.5 Replicate 5	51
1.2.6 Replicate 6	54
1.2.7 Replicate 7	57
1.2.8 Replicate 8	60
1.2.9 Replicate 9	63
1.2.10 Replicate 10	66
1.2.11 all replicates:	69
1.3 1 Mb scale	71
1.3.1 Replicate 1	72
1.3.2 Replicate 2	75
1.3.3 Replicate 3	79
1.3.4 Replicate 4	83
1.3.5 Replicate 5	87
1.3.6 Replicate 6	90
1.3.7 Replicate 7	93
1.3.8 Replicate 8	96
1.3.9 Replicate 9	99
1.3.10 Replicate 10	102
1.3.11 all replicates:	105

2 Drosophila-like neutral simulations of chr 2L (Inferred Landscapes)	107
2.1 50 kb scale	108
2.1.1 Replicate 1	108
2.1.2 Replicate 2	113
2.1.3 Replicate 3	118
2.1.4 Replicate 4	123
2.1.5 Replicate 5	128
2.1.6 Replicate 6	133
2.1.7 Replicate 7	138
2.1.8 Replicate 8	143
2.1.9 Replicate 9	148
2.1.10 Replicate 10	153
2.1.11 all replicates:	158
2.1.12 Plots	161
2.2 200 kb scale	162
2.2.1 Replicate 1	162
2.2.2 Replicate 2	167
2.2.3 Replicate 3	172
2.2.4 Replicate 4	177
2.2.5 Replicate 5	182
2.2.6 Replicate 6	187
2.2.7 Replicate 7	192
2.2.8 Replicate 8	197
2.2.9 Replicate 9	202
2.2.10 Replicate 10	207
2.2.11 all replicates:	212
2.2.12 Plots	215
2.3 1 Mb scale	216
2.3.1 Replicate 1	216
2.3.2 Replicate 2	221
2.3.3 Replicate 3	226
2.3.4 Replicate 4	231
2.3.5 Replicate 5	236
2.3.6 Replicate 6	241
2.3.7 Replicate 7	246
2.3.8 Replicate 8	251
2.3.9 Replicate 9	256
2.3.10 Replicate 10	261
2.3.11 all replicates:	266
2.3.12 Plots	269
3 Real Drosophila data.	271
3.1 50 kb windows	276
3.2 200 kb windows	291
3.3 1 Mb windows	306
3.4 Divergence with <i>D. yakuba</i>	321
3.5 Evolutionary (Protein) Rates	321
4 R² plot with real, simulated (true) and simulated (inferred) data.	328
5 Simulations with BGS and Gamma Mutation Landscape in SLiM	329
5.1 50 kb scale	329
5.1.1 Replicate 1	330
5.1.2 Replicate 2	333

5.1.3	Replicate 3	336
5.1.4	Replicate 4	339
5.1.5	Replicate 5	342
5.1.6	Replicate 6	345
5.1.7	Replicate 7	348
5.1.8	Replicate 8	351
5.1.9	Replicate 9	354
5.1.10	Replicate 10	357
5.2	all replicates:	360
5.3	200 kb scale	361
5.3.1	Replicate 1	361
5.3.2	Replicate 2	365
5.3.3	Replicate 3	368
5.3.4	Replicate 4	371
5.3.5	Replicate 5	374
5.3.6	Replicate 6	377
5.3.7	Replicate 7	380
5.3.8	Replicate 8	383
5.3.9	Replicate 9	386
5.3.10	Replicate 10	389
5.4	all replicates:	392
5.5	1 Mb scale	393
5.5.1	Replicate 1	393
5.5.2	Replicate 2	397
5.5.3	Replicate 3	400
5.5.4	Replicate 4	403
5.5.5	Replicate 5	406
5.5.6	Replicate 6	409
5.5.7	Replicate 7	412
5.5.8	Replicate 8	415
5.5.9	Replicate 9	418
5.5.10	Replicate 10	421
5.6	all replicates:	424
6	Plots	425

1 Drosophila-like neutral simulations of chr 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 = FALSE)
rep_1.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_1/sim.tmrca.50k.map", header = FALSE)

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

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

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

sim.lands.50kb.rep5 <- sim.lands.50kb
sim.lands.50kb.rep5$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

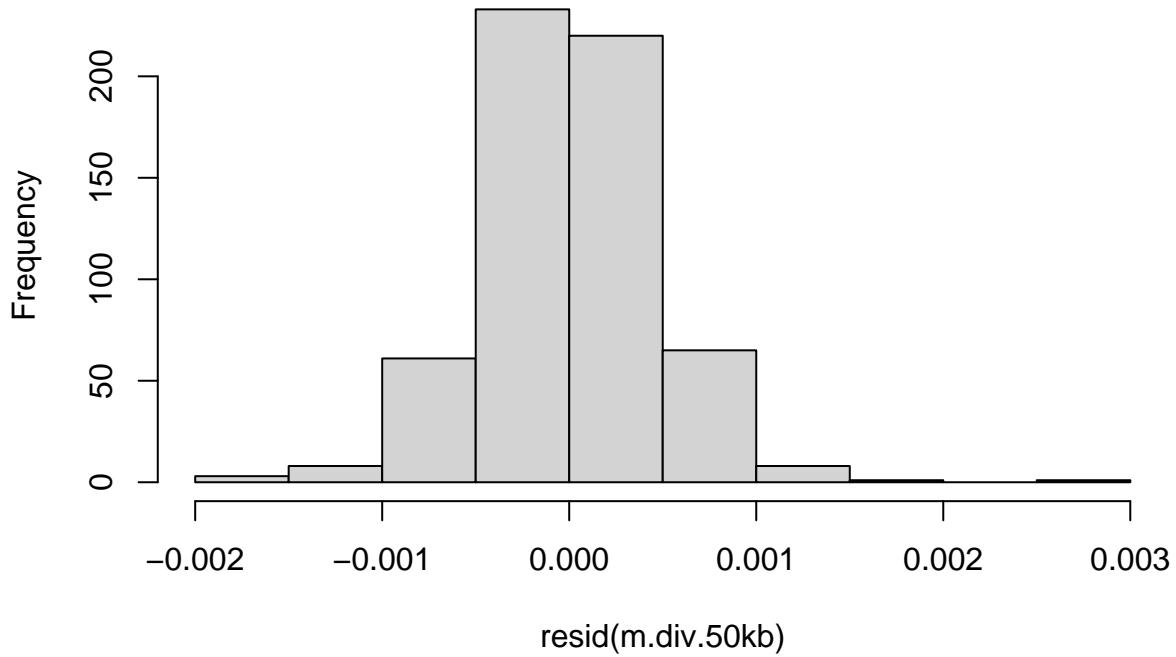
hmctest(m.div.50kb)

##
## Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.39214, p-value = 0.001

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 
## -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      1.100e-02 1.919e-05  573.279 <2e-16 ***
## rhoC        4.258e-05 1.915e-05   2.223  0.0266 *  
## tmrcaC      1.665e-03 1.910e-05   87.195 <2e-16 *** 
## thetaC:tmrcaC 8.867e-04 1.810e-05   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) 0.020690112 1.904598e-05 1086.3241 0.0000
## thetaC       0.011057434 1.990949e-05  555.3850 0.0000
## rhoC        0.000025866 1.646000e-05     1.5715 0.1166
## tmrcaC      0.001666932 1.902968e-05    87.5965 0.0000
## thetaC:tmrcaC 0.000875737 1.979066e-05   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
## -2.955335881 -0.608690299  0.006577071  0.610719662  4.719725050
## 
## 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 = FALSE)
rep_2.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_2/sim.tmrca.50k.map", header = FALSE)

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

```

```

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

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

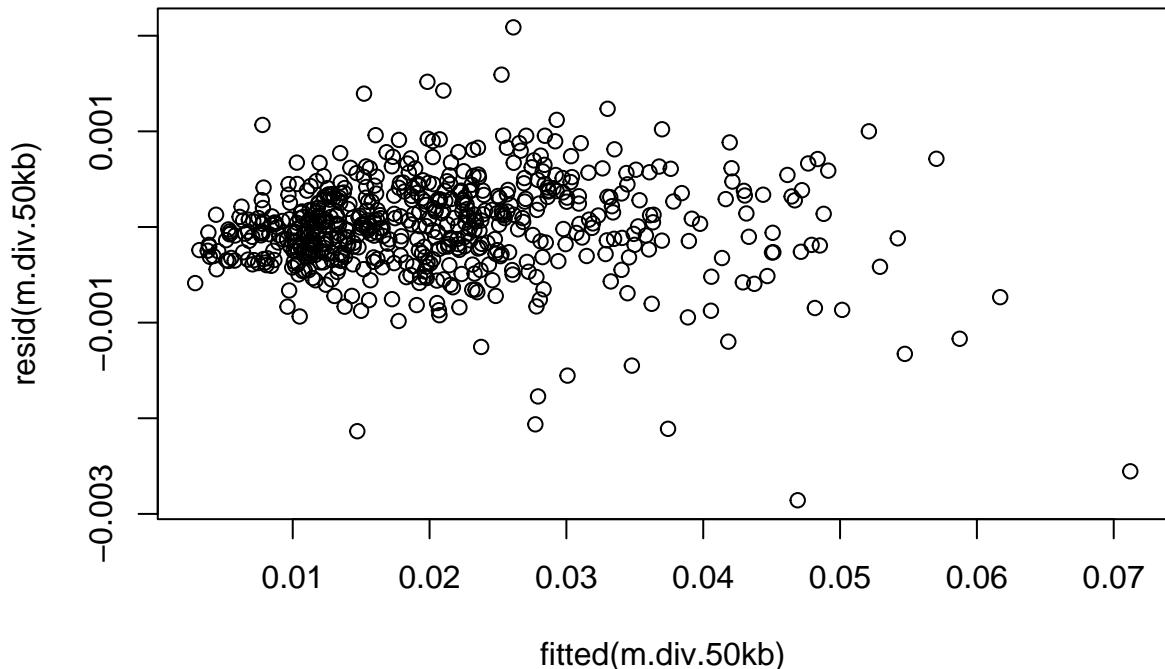
sim.lands.50kb.rep2 <- sim.lands.50kb
sim.lands.50kb.rep2$Replicate <- 2

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

```

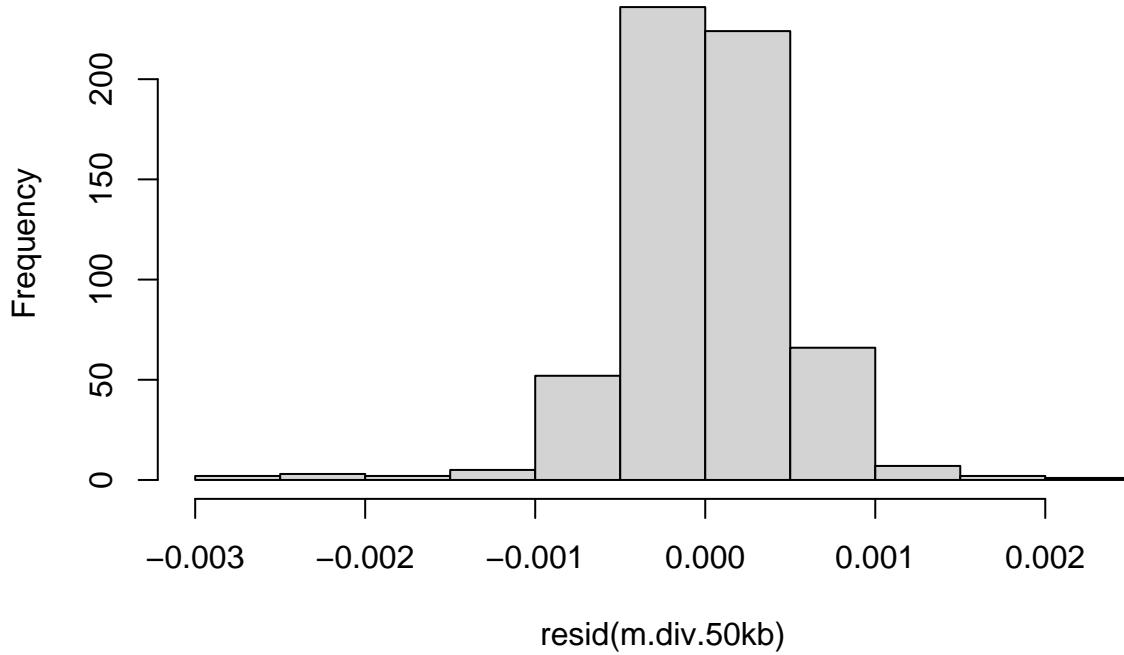
```

hmctest(m.div.50kb)

##
##  Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.51438, p-value = 0.689
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       1.092e-02 2.058e-05  530.815 <2e-16 ***
## rhoC        -2.995e-05 2.053e-05   -1.459   0.145    
## tmrcaC       1.738e-03 2.063e-05   84.273 <2e-16 ***
## thetaC:tmrcaC 9.278e-04 1.973e-05   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$VarExp[3]
r2.sim.50kb[2, 2] <- anova.diversity$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 = FALSE)
rep_3.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_3/sim.tmrca.50k.map", header = FALSE)

sim.lands.50kb <- as.data.frame(cbind(rep_3.pi.50kb$pi, sim.theta.50kb$sim, sim.rho.50kb$sim, rep_3.tmrca.50kb))
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
## -0.04613476

```

```

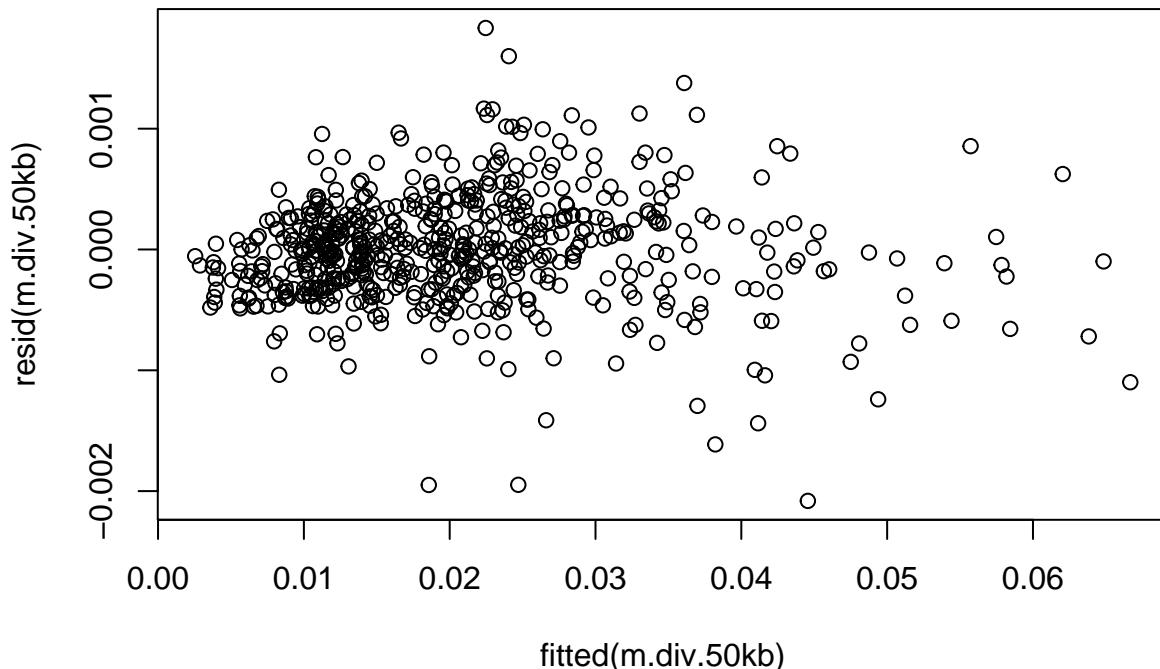
##          rho
## 0.01710883
# standardizing
sim.lands.50kb$thetaC <- (sim.lands.50kb$theta - mean(sim.lands.50kb$theta)) / sd(sim.lands.50kb$theta)
sim.lands.50kb$tmrcaC <- (sim.lands.50kb$tmrca - mean(sim.lands.50kb$tmrca)) / sd(sim.lands.50kb$tmrca)
sim.lands.50kb$rhoC <- (sim.lands.50kb$rho - mean(sim.lands.50kb$rho)) / sd(sim.lands.50kb$rho)

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

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)

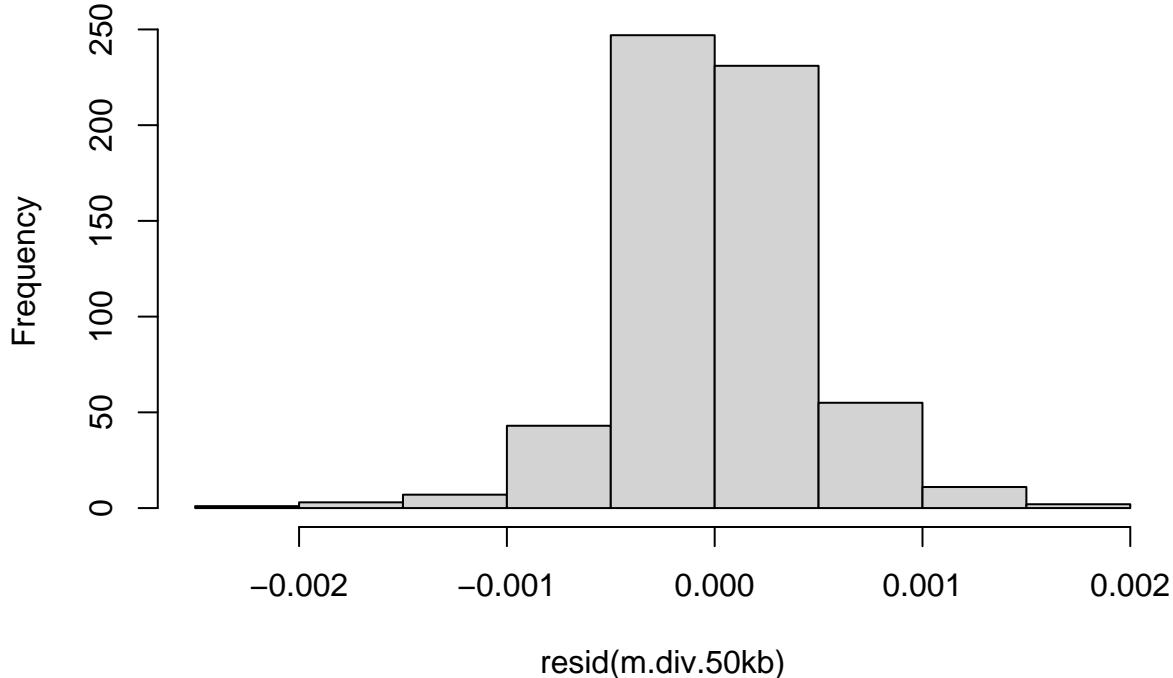
##
##  Durbin-Watson test
##
## data: m.div.50kb
## DW = 1.9214, p-value = 0.1563
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.50kb)

##
##  Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.50434, p-value = 0.578

```

```
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  
## -2.081e-03 -2.636e-04 -7.680e-06  2.600e-04  1.833e-03  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)  2.067e-02  1.870e-05 1105.47  <2e-16 ***  
## thetaC       1.101e-02  1.875e-05  587.29  <2e-16 ***  
## rhoC        1.911e-05  1.874e-05    1.02    0.308  
## tmrcaC      1.935e-03  1.882e-05  102.80  <2e-16 ***  
## thetaC:tmrcaC 1.034e-03  1.754e-05   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$VarExp[3])
r2.sim.50kb[2, 3] <- anova.diversity$VarExp[1] * 100
r2.sim.50kb[3, 3] <- anova.diversity$VarExp[2] * 100
r2.sim.50kb[4, 3] <- anova.diversity$VarExp[3] * 100
r2.sim.50kb[5, 3] <- anova.diversity$VarExp[4] * 100

```

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

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

```

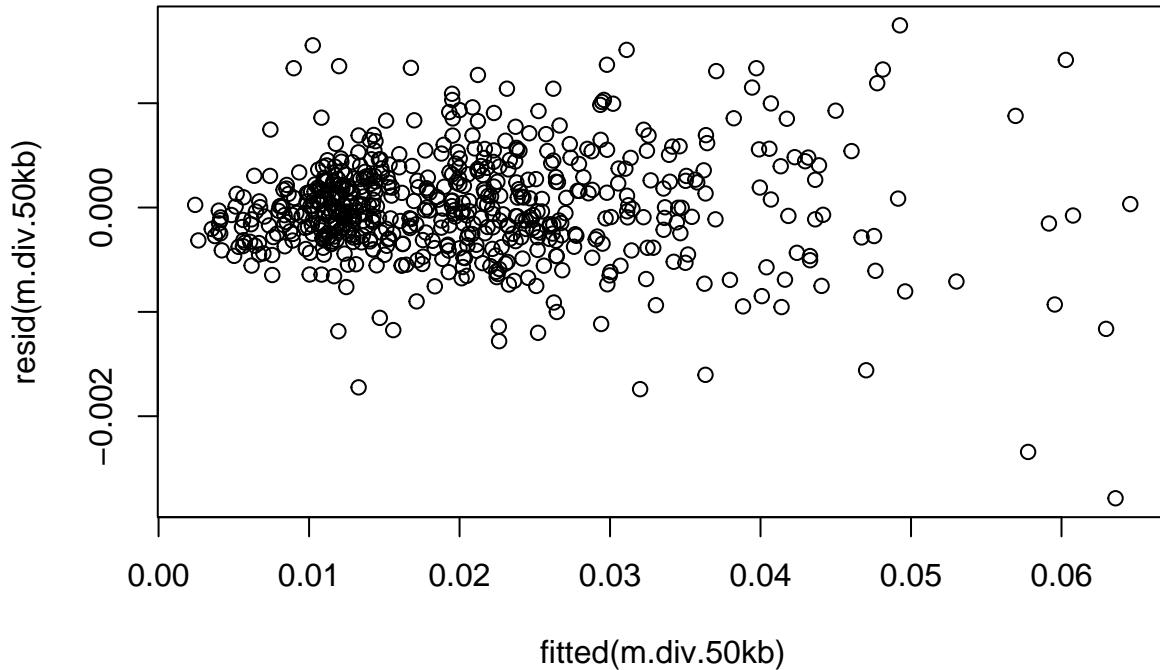
```

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

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))

```



```

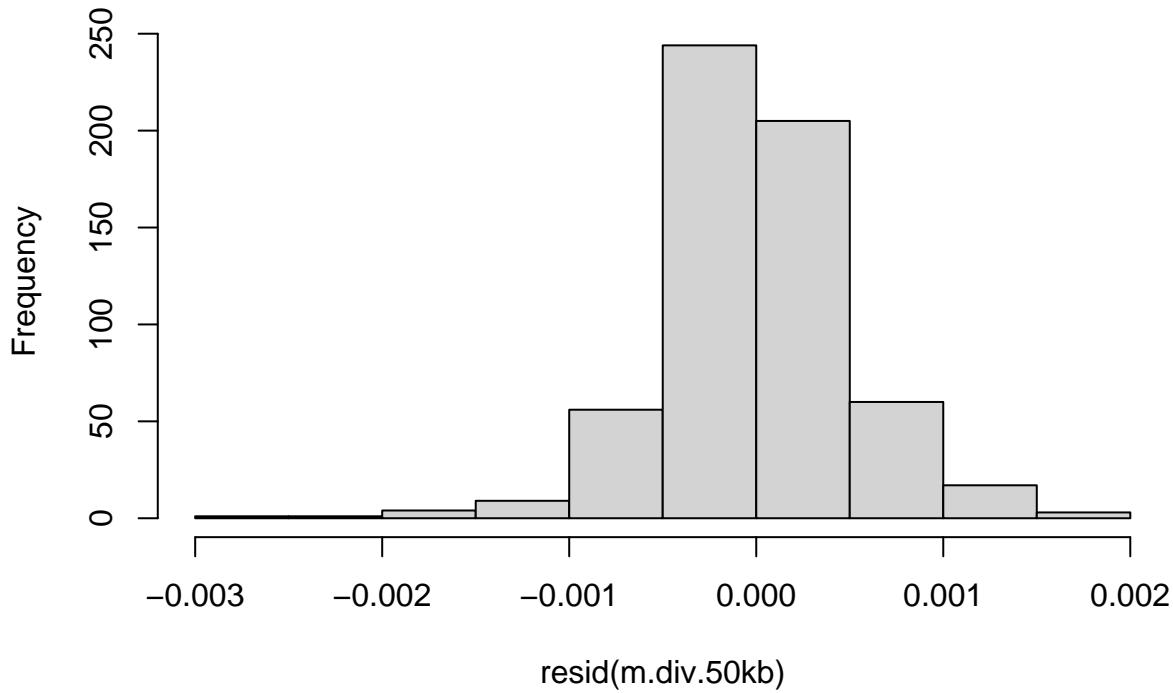
dwtest(m.div.50kb)

##
##  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 = 0.898
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
## -2.787e-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      1.093e-02 2.081e-05 525.118 <2e-16 ***
## rhoC        4.114e-05 2.082e-05   1.976  0.0486 *
## tmrcaC     1.849e-03 2.081e-05   88.857 <2e-16 ***
## thetaC:tmrcaC 9.878e-04 2.074e-05  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])
```

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

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 = FALSE)
rep_5.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_5/sim.tmrca.50k.map", header = FALSE)

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

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

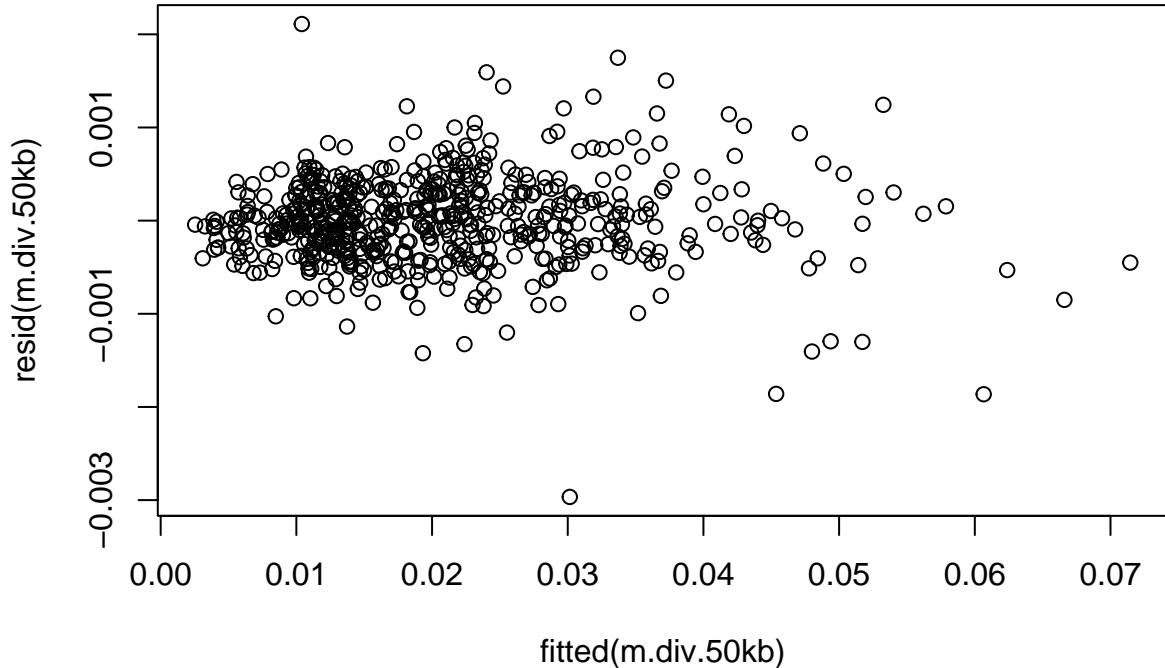
sim.lands.50kb$bin <- 1:nrow(sim.lands.50kb)
```

```

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)

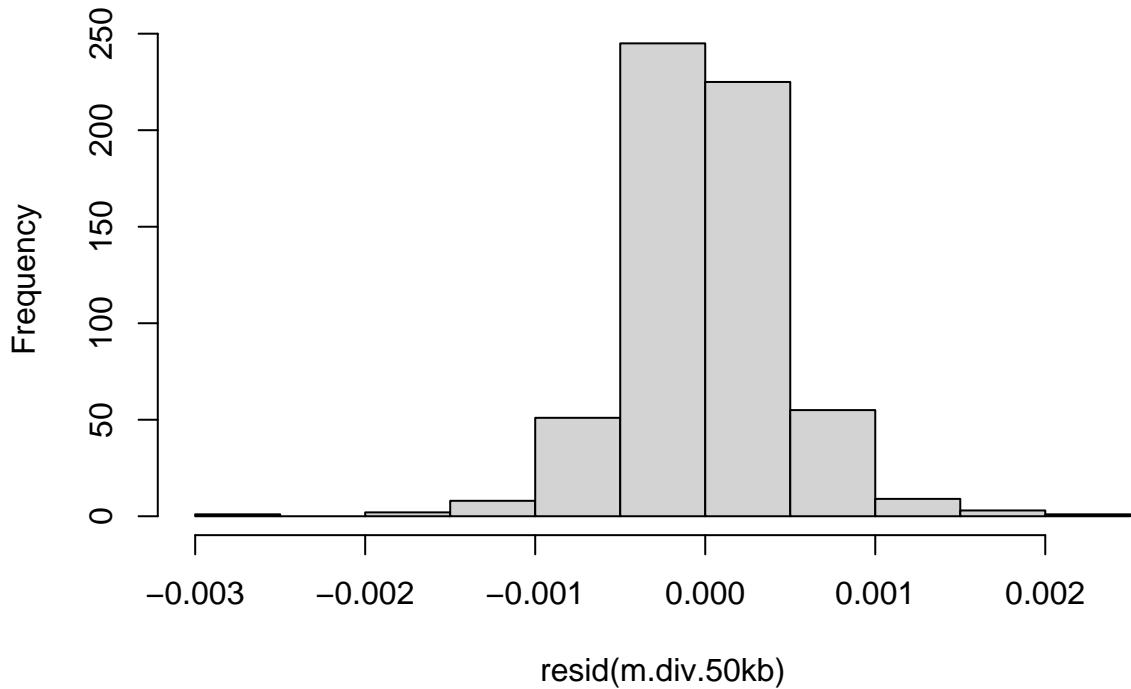
```

```

##
## Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.45962, p-value = 0.07
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
## -2.966e-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      1.099e-02 1.973e-05  557.06 <2e-16 ***
## rhoC        4.533e-06 1.972e-05    0.23   0.818
## tmrcaC     1.750e-03 1.971e-05   88.80 <2e-16 ***
## thetaC:tmrcaC 9.113e-04 1.937e-05   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 = FALSE)
rep_6.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_6/sim.tmrca.50k.map", header = FALSE)

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

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

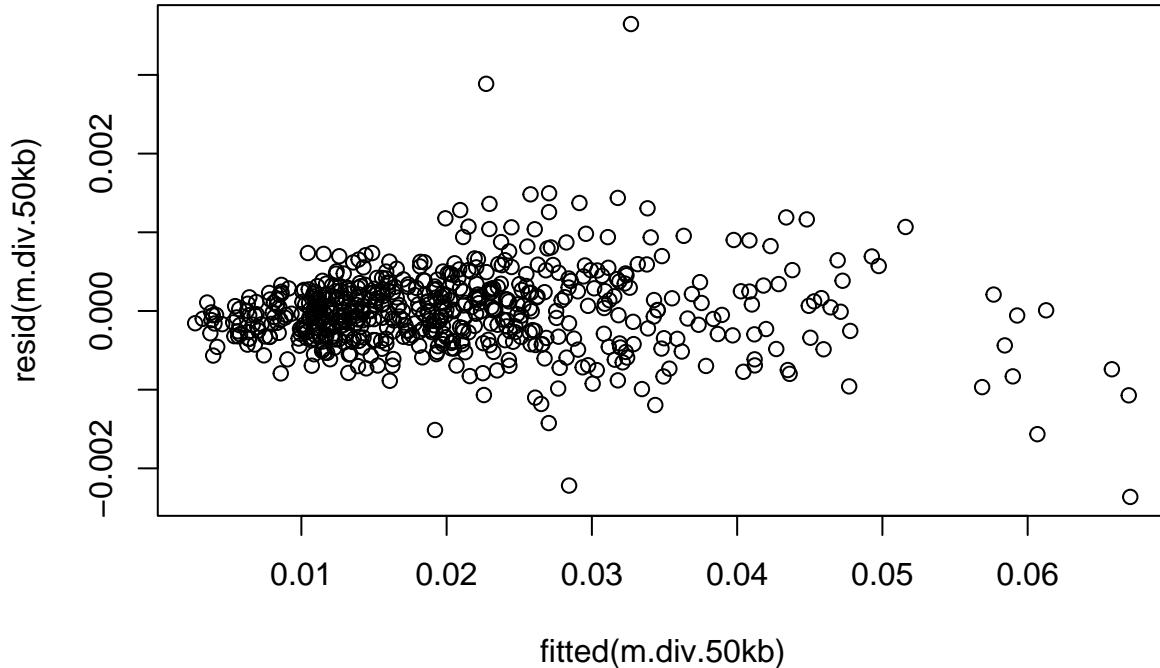
sim.lands.50kb$bin <- 1:nrow(sim.lands.50kb)
```

```

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))

```



```

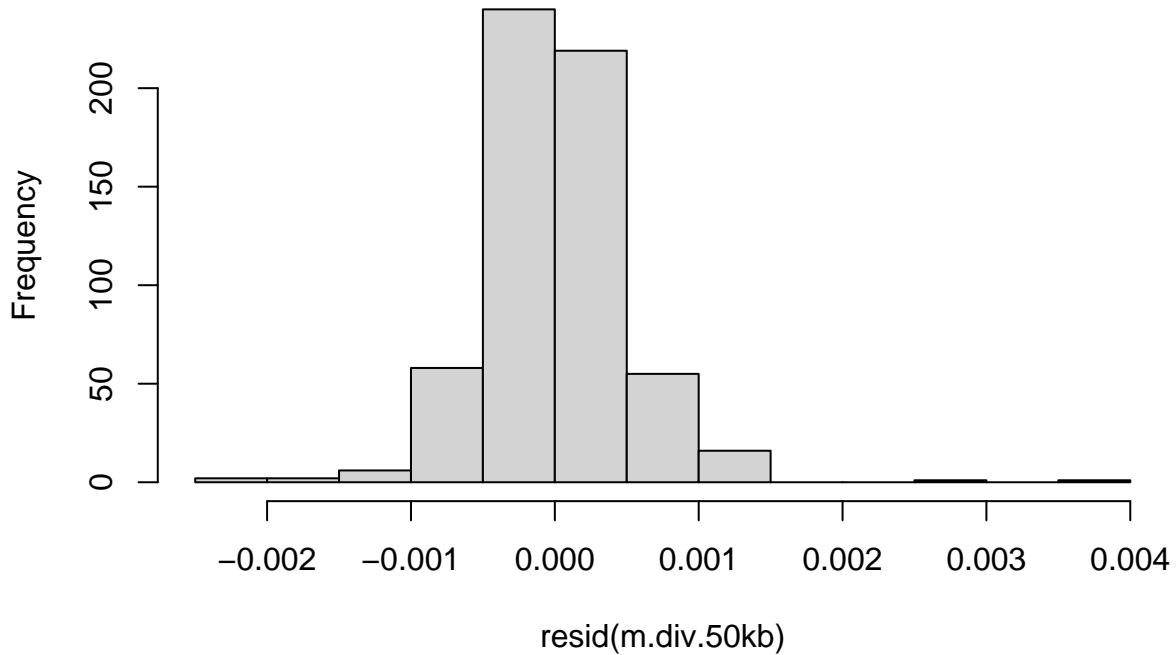
dwtest(m.div.50kb)

##
## 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 = 0.089
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
## -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      1.096e-02 2.083e-05 526.030 <2e-16 ***
## rhoC       -3.498e-05 2.078e-05 -1.683  0.0929 .
## tmrcaC     1.783e-03 2.077e-05  85.861 <2e-16 ***
## thetaC:tmrcaC 9.128e-04 2.013e-05  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 = FALSE)
rep_7.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_7/sim.tmrca.50k.map", header = FALSE)

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

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

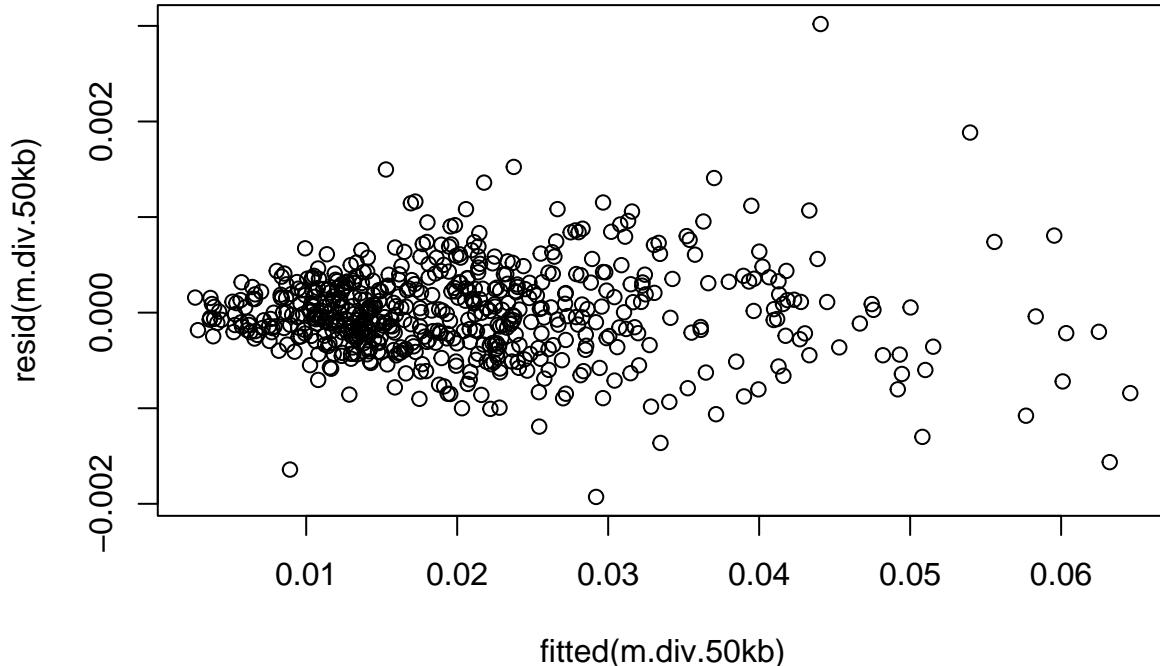
sim.lands.50kb$bin <- 1:nrow(sim.lands.50kb)
```

```

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)

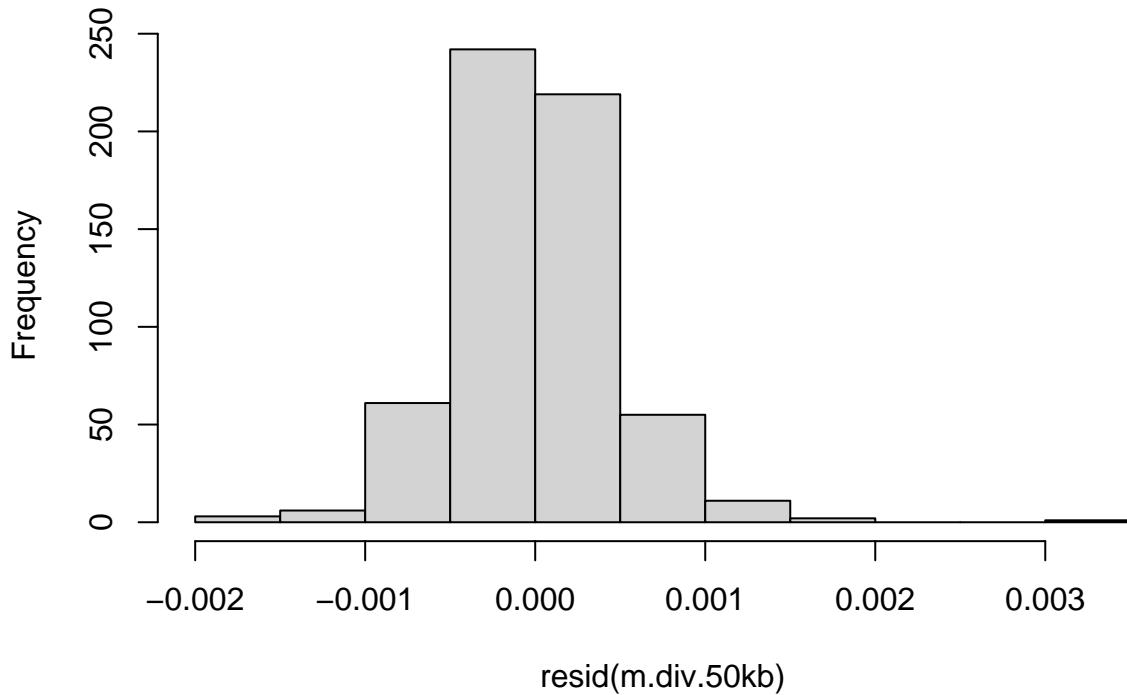
```

```

##
## Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.47957, p-value = 0.23
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
## -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      1.104e-02 1.950e-05  565.892 <2e-16 ***
## rhoC       -1.707e-05 1.947e-05   -0.877  0.381
## tmrcaC      1.829e-03 1.950e-05   93.801 <2e-16 ***
## thetaC:tmrcaC 9.594e-04 1.942e-05   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 = FALSE)
rep_8.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_8/sim.tmrca.50k.map", header = FALSE)

sim.lands.50kb <- as.data.frame(cbind(rep_8.pi.50kb$pi, sim.theta.50kb$sim, sim.rho.50kb$sim, rep_8.tmrca.50kb))
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 = mf[[1L]], y = mf[[2L]], ...): 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 = mf[[1L]], y = mf[[2L]], ...): 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
```

```

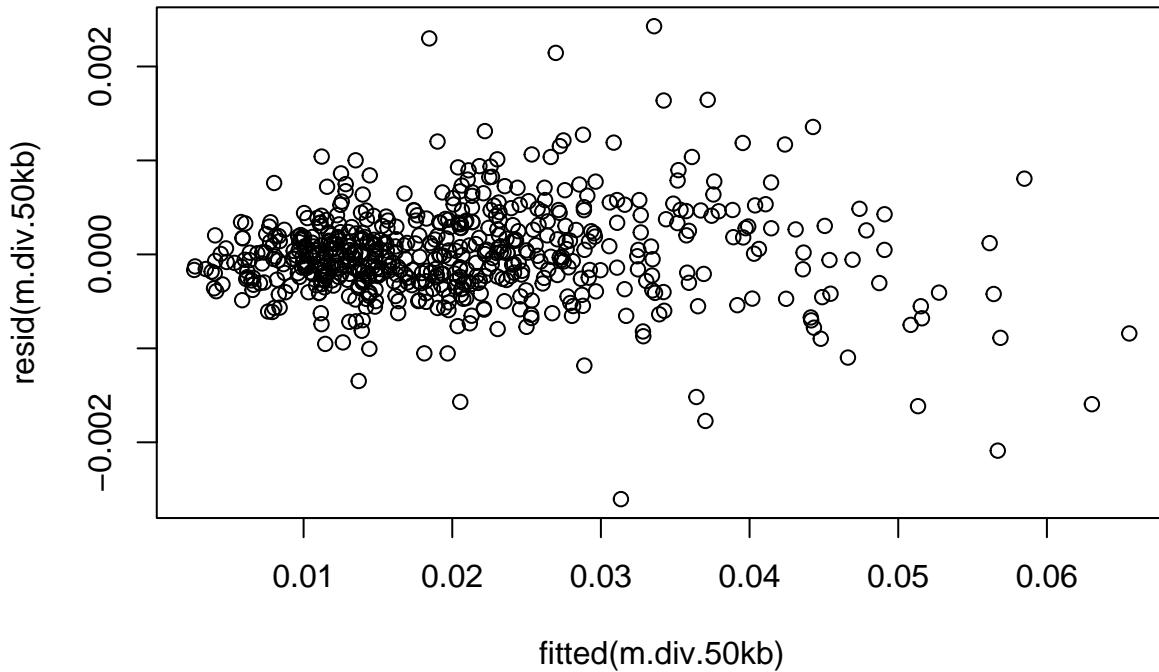
# standardizing
sim.lands.50kb$thetaC <- (sim.lands.50kb$theta - mean(sim.lands.50kb$theta)) / sd(sim.lands.50kb$theta)
sim.lands.50kb$tmrcaC <- (sim.lands.50kb$tmrca - mean(sim.lands.50kb$tmrca)) / sd(sim.lands.50kb$tmrca)
sim.lands.50kb$rhoC <- (sim.lands.50kb$rho - mean(sim.lands.50kb$rho)) / sd(sim.lands.50kb$rho)

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

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)

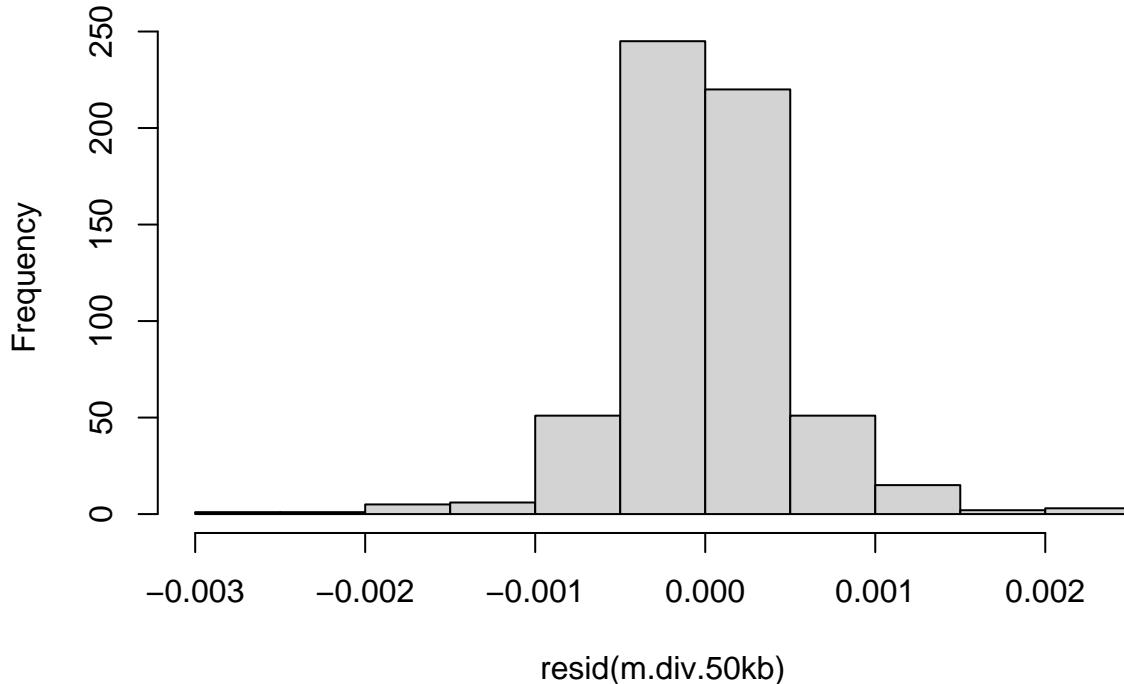
##
##  Durbin-Watson test
##
##  data: m.div.50kb
##  DW = 2.0323, p-value = 0.6384
##  alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.50kb)

##
##  Harrison-McCabe test
##
##  data: m.div.50kb
##  HMC = 0.48525, p-value = 0.308

```

```
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  
## -2.605e-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       1.095e-02 2.074e-05 527.803 <2e-16 ***  
## rhoC        -2.590e-06 2.068e-05 -0.125    0.9  
## tmrcaC       1.744e-03 2.086e-05  83.583 <2e-16 ***  
## thetaC:tmrcaC 8.513e-04 1.947e-05  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 = TRUE)
rep_9.tmrca.50kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_9/sim.tmrca.50k.map", header = TRUE)

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

# standardizing
sim.lands.50kb$thetaC <- (sim.lands.50kb$theta - mean(sim.lands.50kb$theta)) / sd(sim.lands.50kb$theta)
sim.lands.50kb$tmrcaC <- (sim.lands.50kb$tmrca - mean(sim.lands.50kb$tmrca)) / sd(sim.lands.50kb$tmrca)
sim.lands.50kb$rhoC <- (sim.lands.50kb$rho - mean(sim.lands.50kb$rho)) / sd(sim.lands.50kb$rho)
```

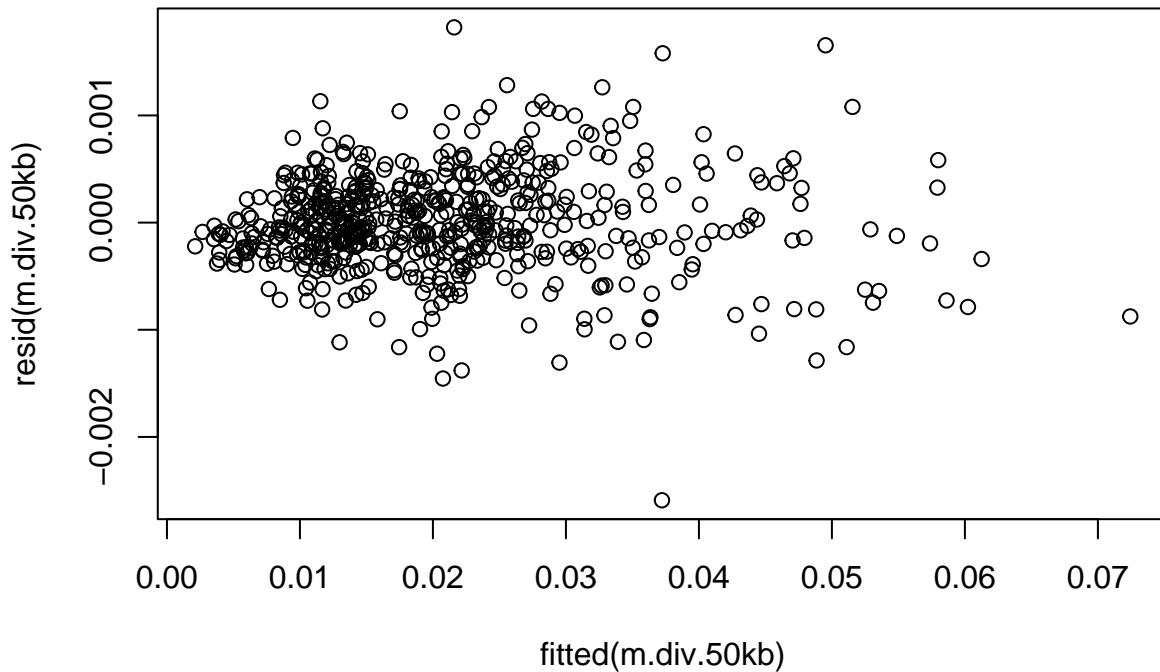
```

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

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))

```



```

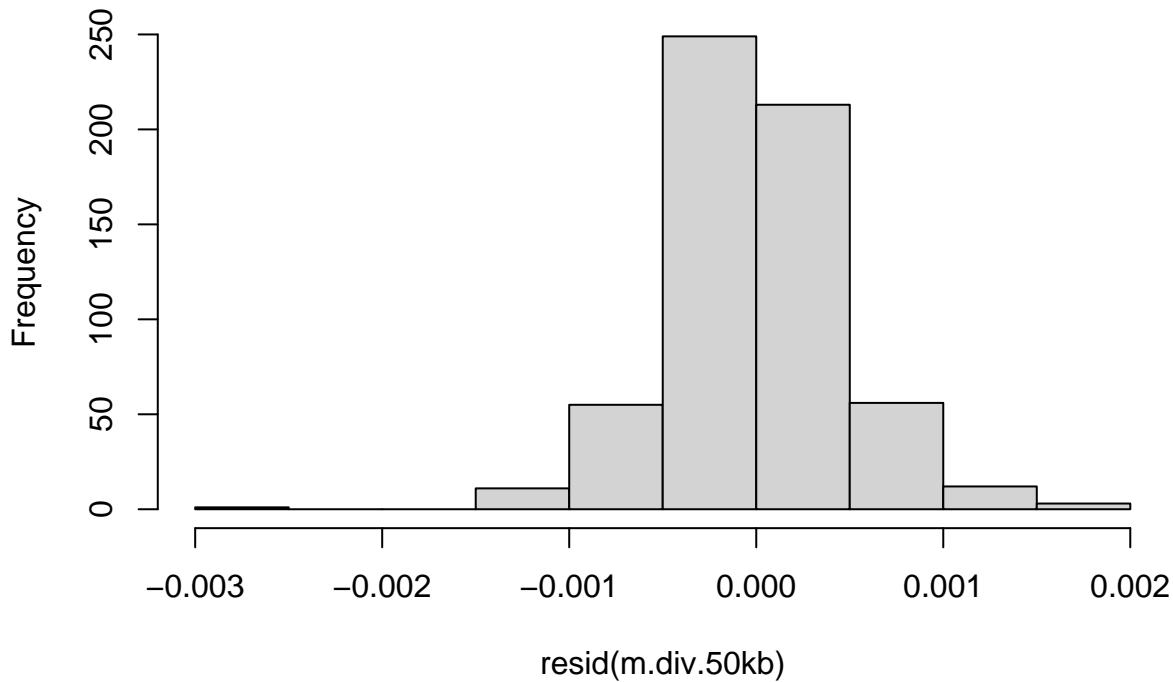
dwtest(m.div.50kb)

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

##
##  Harrison-McCabe test
##
##  data: m.div.50kb
##  HMC = 0.55455, p-value = 0.967
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      1.099e-02 1.903e-05  577.567 <2e-16 ***
## rhoC       -2.305e-05 1.902e-05  -1.212  0.226    
## tmrcaC      1.783e-03 1.902e-05   93.731 <2e-16 ***
## thetaC:tmrcaC 9.400e-04 1.800e-05   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

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

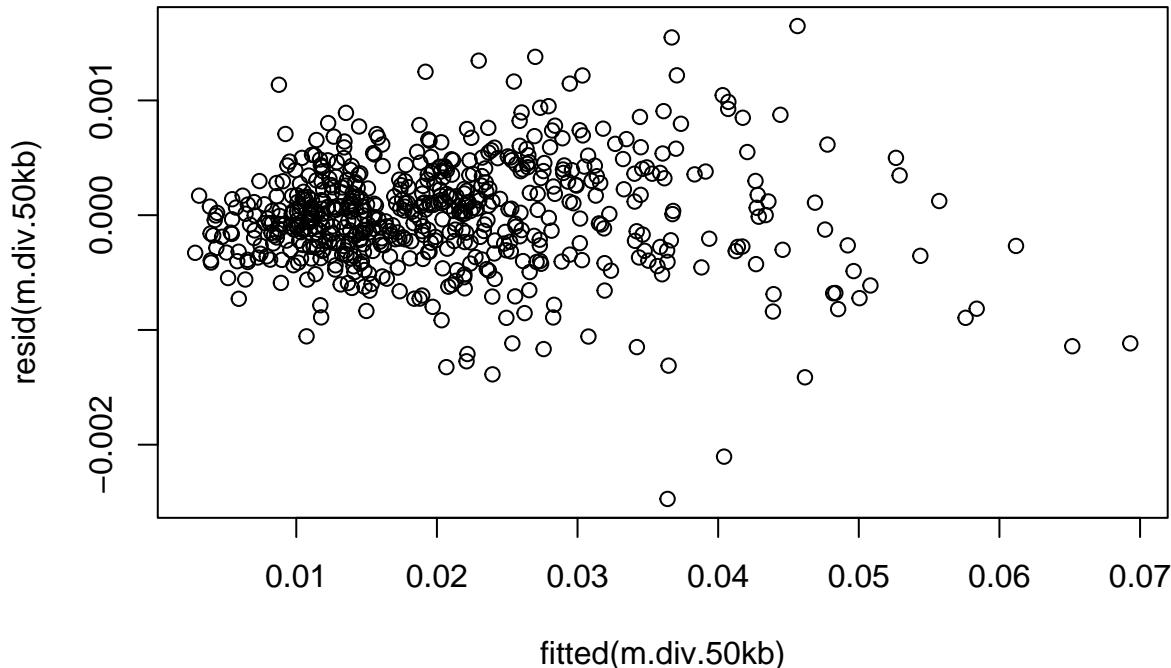
sim.lands.50kb$bin <- 1:nrow(sim.lands.50kb)
```

```

sim.lands.50kb.rep50 <- sim.lands.50kb
sim.lands.50kb.rep50$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))

```



```

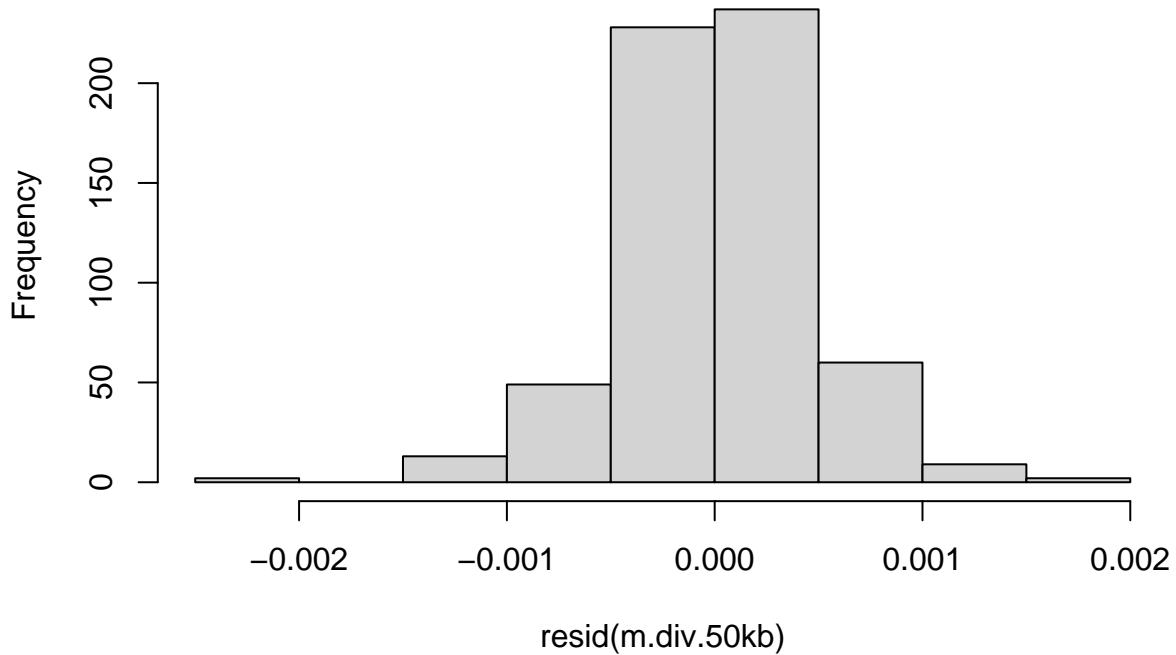
dwtest(m.div.50kb)

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

##
## Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.43153, p-value = 0.007
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 
## -2.473e-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      1.102e-02 1.937e-05  568.969 <2e-16 ***
## rhoC        2.243e-05 1.927e-05   1.164  0.245    
## tmrcaC     1.929e-03 1.929e-05  100.018 <2e-16 ***
## thetaC:tmrcaC 1.060e-03 1.803e-05   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))

```

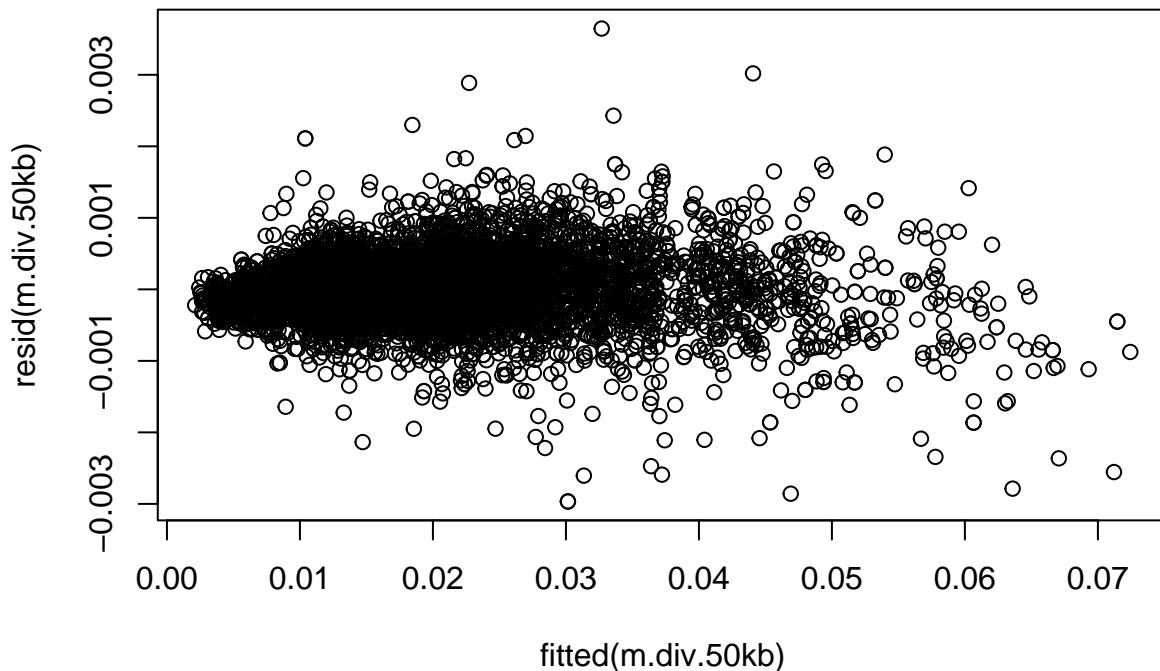
1.1.11 all replicates:

```

sim.lands.50kb.all <- rbind(sim.lands.50kb.rep5, 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)

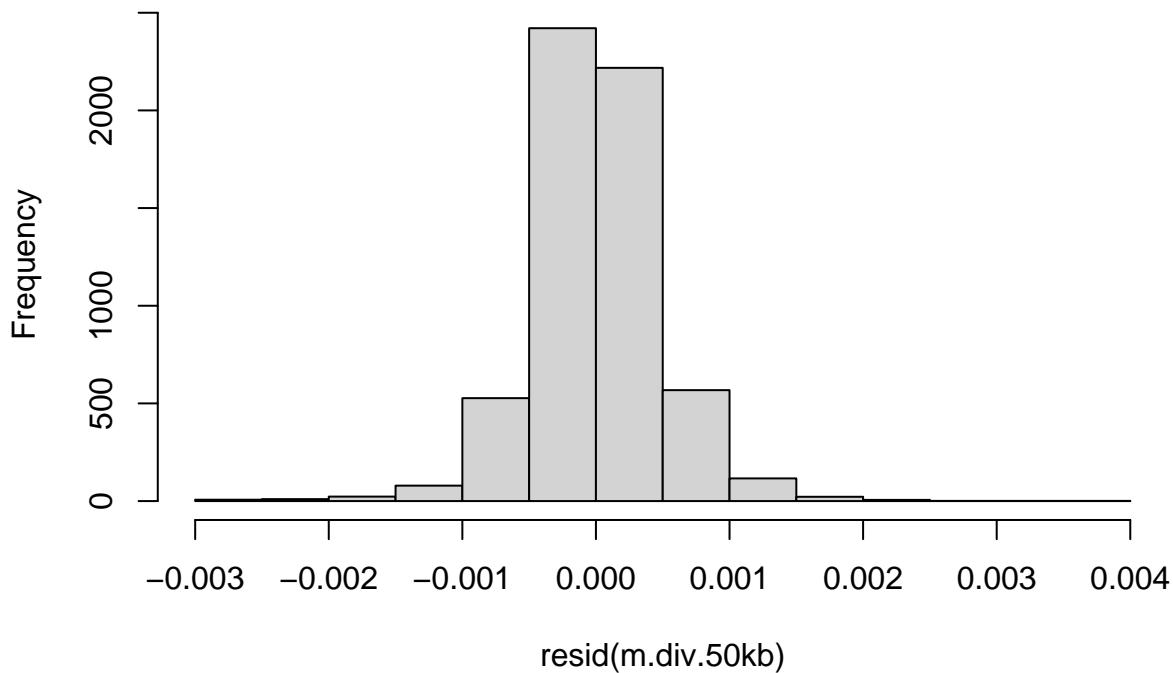
##
##  Durbin-Watson test
##
## data: m.div.50kb
## DW = 2.0313, p-value = 0.8378
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.50kb)

##
##  Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.50138, p-value = 0.547

```

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

Histogram of resid(m.div.50kb)



```
m.div.50kb.2 <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC)*as.factor(Replicate), data = sim.lands.50kb.all)
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC + rhoC:tmrcaC)*as.factor(Replicate), data = sim.lands.50kb.all)
```

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

```
##          df      AIC
## m.div.50kb   46 -74487.09
## m.div.50kb.2 55 -74476.67
## m.div.50kb.3 64 -74469.15
```

```
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 -0.0002718 -0.0000139  0.0002767  0.0036466
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.057e-02 1.983e-05 1037.362 < 2e-16
## thetaC      1.092e-02 1.994e-05  547.744 < 2e-16
## rhoC       -2.995e-05 1.990e-05   -1.505 0.132298
## tmrcaC     1.738e-03 1.999e-05   86.961 < 2e-16
## as.factor(Replicate)3 9.708e-05 2.805e-05    3.461 0.000543
```

```

## as.factor(Replicate)4      -6.998e-05  2.804e-05  -2.495  0.012614
## as.factor(Replicate)5      4.112e-05  2.429e-05  1.693  0.090555
## as.factor(Replicate)6      8.307e-06  2.804e-05  0.296  0.767075
## as.factor(Replicate)7      3.621e-05  2.806e-05  1.290  0.197011
## as.factor(Replicate)8      -2.113e-06 2.806e-05  -0.075  0.939967
## as.factor(Replicate)9      7.290e-05  2.805e-05  2.599  0.009365
## as.factor(Replicate)10     1.253e-04  2.807e-05  4.465  8.14e-06
## thetaC:tmrcaC             9.278e-04  1.913e-05  48.515 < 2e-16
## thetaC:as.factor(Replicate)3 9.166e-05  2.817e-05  3.254  0.001143
## thetaC:as.factor(Replicate)4 7.711e-06  2.815e-05  0.274  0.784112
## thetaC:as.factor(Replicate)5 7.032e-05  2.440e-05  2.882  0.003971
## thetaC:as.factor(Replicate)6 3.748e-05  2.819e-05  1.329  0.183796
## thetaC:as.factor(Replicate)7 1.140e-04  2.817e-05  4.045  5.29e-05
## thetaC:as.factor(Replicate)8 2.656e-05  2.820e-05  0.942  0.346264
## thetaC:as.factor(Replicate)9 7.181e-05  2.815e-05  2.551  0.010780
## thetaC:as.factor(Replicate)10 9.662e-05  2.824e-05  3.422  0.000626
## rhoC:as.factor(Replicate)3  4.906e-05  2.813e-05  1.744  0.081155
## rhoC:as.factor(Replicate)4  7.109e-05  2.812e-05  2.528  0.011491
## rhoC:as.factor(Replicate)5  3.448e-05  2.436e-05  1.415  0.156986
## rhoC:as.factor(Replicate)6  -5.031e-06 2.813e-05 -0.179  0.858054
## rhoC:as.factor(Replicate)7  1.288e-05  2.812e-05  0.458  0.646844
## rhoC:as.factor(Replicate)8  2.736e-05  2.813e-05  0.973  0.330721
## rhoC:as.factor(Replicate)9  6.899e-06  2.812e-05  0.245  0.806184
## rhoC:as.factor(Replicate)10 5.238e-05  2.814e-05  1.862  0.062717
## tmrcaC:as.factor(Replicate)3 1.964e-04  2.825e-05  6.951  4.02e-12
## tmrcaC:as.factor(Replicate)4 1.106e-04  2.818e-05  3.924  8.80e-05
## tmrcaC:as.factor(Replicate)5 1.172e-05  2.444e-05  0.479  0.631676
## tmrcaC:as.factor(Replicate)6 4.446e-05  2.818e-05  1.578  0.114729
## tmrcaC:as.factor(Replicate)7 9.030e-05  2.820e-05  3.202  0.001374
## tmrcaC:as.factor(Replicate)8 5.059e-06  2.832e-05  0.179  0.858203
## tmrcaC:as.factor(Replicate)9 4.405e-05  2.818e-05  1.563  0.118059
## tmrcaC:as.factor(Replicate)10 1.906e-04  2.822e-05  6.754  1.57e-11
## thetaC:tmrcaC:as.factor(Replicate)3 1.062e-04  2.668e-05  3.981  6.93e-05
## thetaC:tmrcaC:as.factor(Replicate)4 5.999e-05  2.752e-05  2.179  0.029335
## thetaC:tmrcaC:as.factor(Replicate)5 -1.653e-05 2.359e-05 -0.701  0.483489
## thetaC:tmrcaC:as.factor(Replicate)6 -1.508e-05 2.714e-05 -0.556  0.578536
## thetaC:tmrcaC:as.factor(Replicate)7 3.156e-05  2.754e-05  1.146  0.251829
## thetaC:tmrcaC:as.factor(Replicate)8 -7.658e-05 2.676e-05 -2.862  0.004227
## thetaC:tmrcaC:as.factor(Replicate)9 1.221e-05  2.681e-05  0.455  0.648890
## thetaC:tmrcaC:as.factor(Replicate)10 1.326e-04  2.669e-05  4.968  6.96e-07
##
## (Intercept)                      ***
## thetaC                                ***
## rhoC                                 ***
## tmrcaC                               ***
## 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)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)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)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)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: 0.0004856 on 5955 degrees of freedom
## Multiple R-squared:  0.9981, Adjusted R-squared:  0.9981
## F-statistic: 7.12e+04 on 44 and 5955 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
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

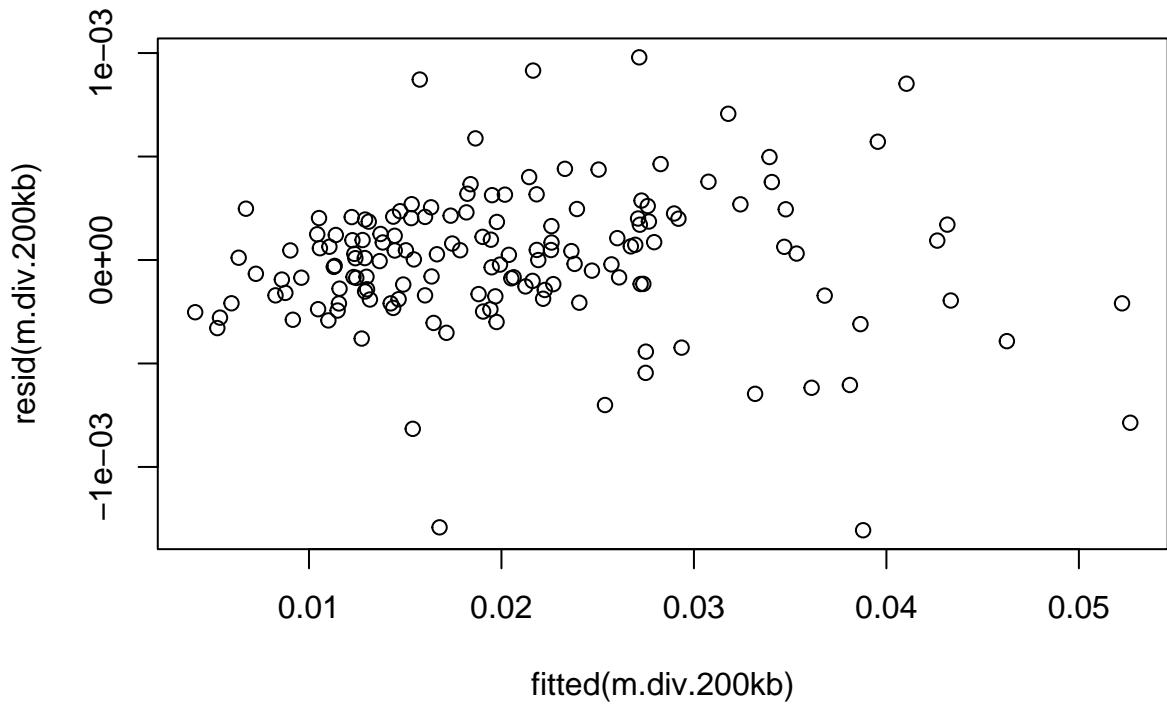
# standardizing
sim.lands.200kb$thetaC <- (sim.lands.200kb$theta - mean(sim.lands.200kb$theta)) / sd(sim.lands.200kb$theta)
sim.lands.200kb$tmrcaC <- (sim.lands.200kb$tmrca - mean(sim.lands.200kb$tmrca)) / sd(sim.lands.200kb$tmrca)
sim.lands.200kb$rhoC <- (sim.lands.200kb$rho - mean(sim.lands.200kb$rho))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

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

sim.lands.200kb.rep5 <- sim.lands.200kb
sim.lands.200kb.rep5$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))

```

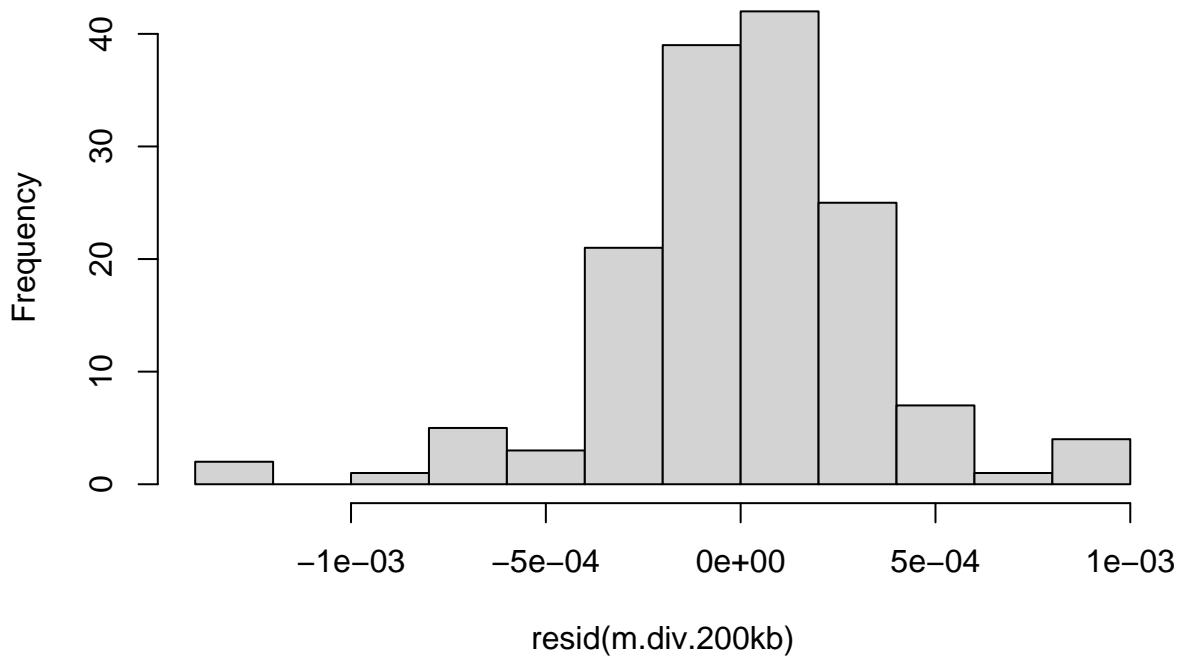


```
dwtest(m.div.200kb)
```

```
##
##  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 = 0.495
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      9.726e-03 2.807e-05 346.532 <2e-16 ***
## rhoC        4.827e-05 2.819e-05  1.712   0.089 .  
## tmrcaC      9.232e-04 2.798e-05  32.997 <2e-16 *** 
## thetaC:tmrcaC 4.181e-04 2.919e-05  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])
```

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

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

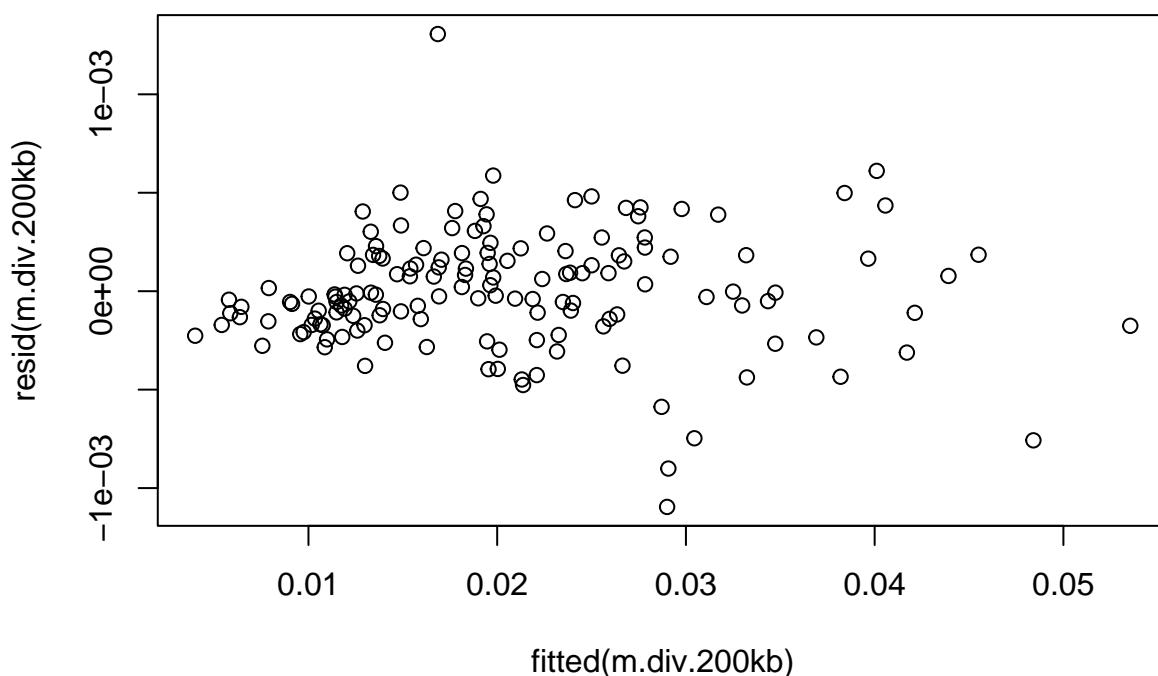
sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)
```

```

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))

```



```

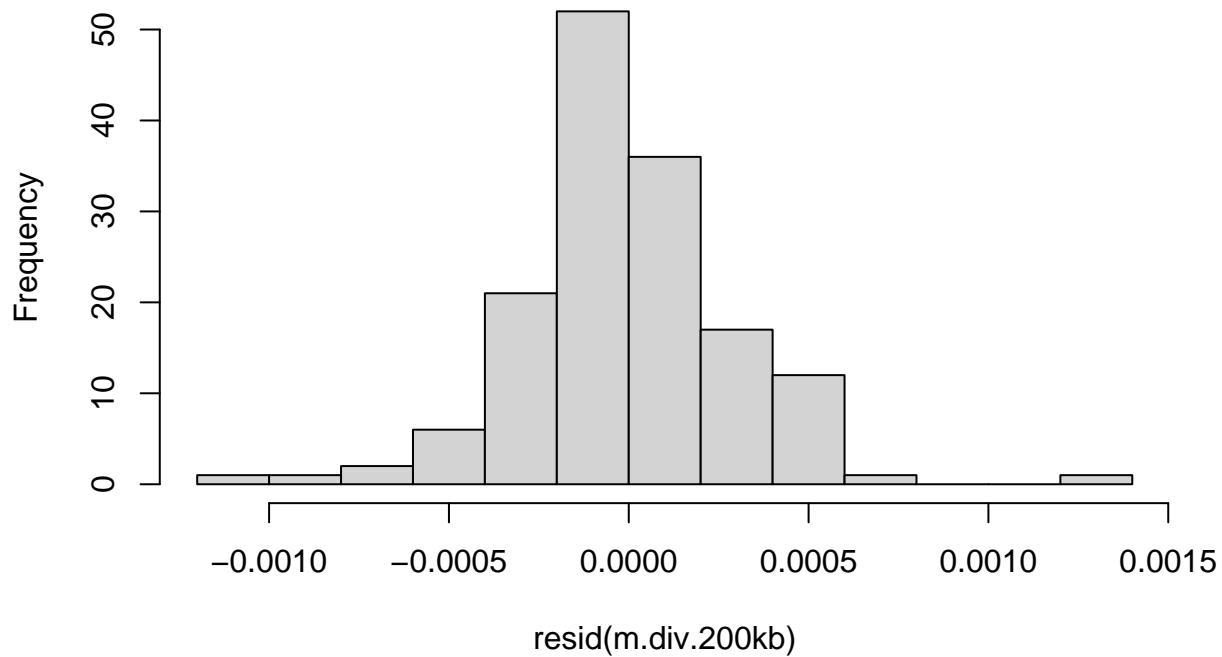
dwtest(m.div.200kb)

##
##  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 = 0.93
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
## -1.096e-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      9.659e-03 2.528e-05 382.15 <2e-16 ***
## rhoC       -4.702e-05 2.514e-05  -1.87  0.0635 .
## tmrcaC     1.008e-03 2.527e-05  39.88 <2e-16 ***
## thetaC:tmrcaC 5.023e-04 2.506e-05  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", header=TRUE)
rep_3.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_3/sim.tmrca.200k.map", header=TRUE)

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

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

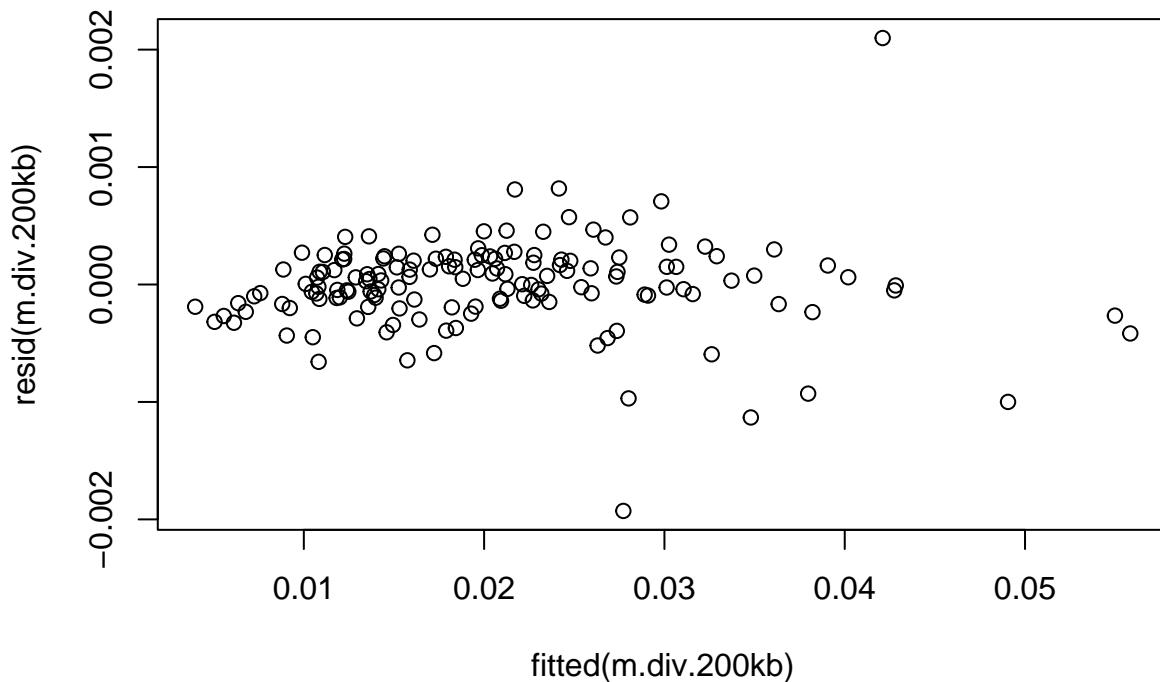
sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)
```

```

sim.lands.200kb.rep3 <- sim.lands.200kb
sim.lands.200kb.rep3$Replicate <- 3

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

```



```

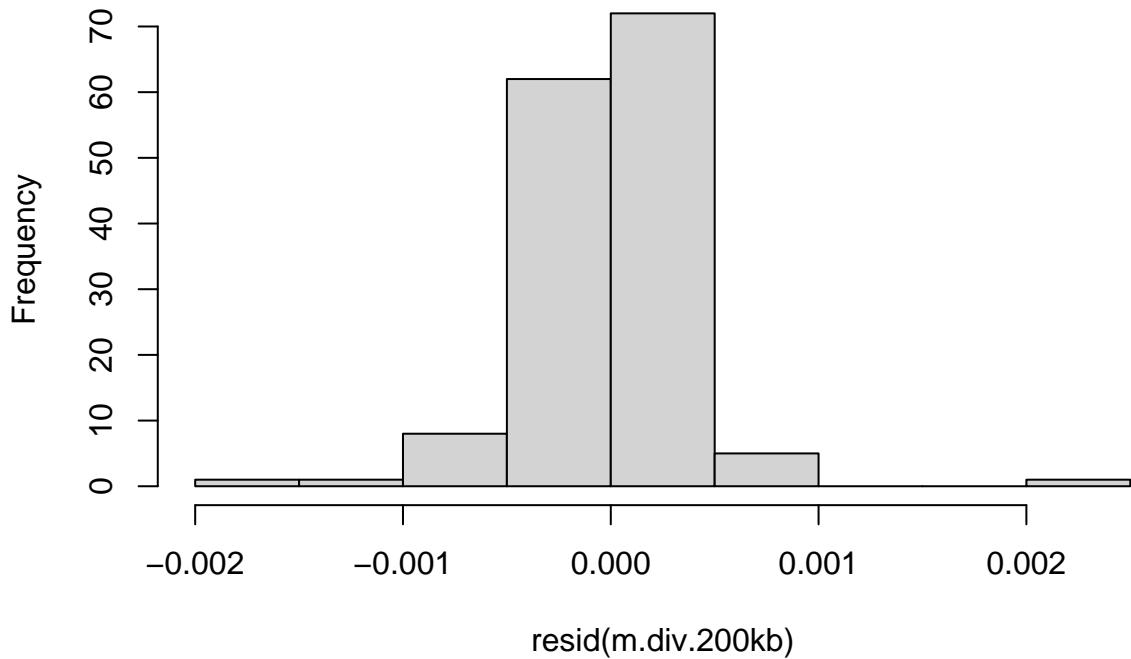
dwtest(m.div.200kb)

##
##  Durbin-Watson test
##
##  data: m.div.200kb
##  DW = 1.9205, 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 = 0.24
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
## -1.928e-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      9.746e-03 3.273e-05 297.825 <2e-16 ***
## rhoC        4.776e-05 3.277e-05   1.458   0.147
## tmrcaC     1.034e-03 3.267e-05  31.662 <2e-16 ***
## thetaC:tmrcaC 4.558e-04 3.396e-05 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))
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 = 5544941, 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

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

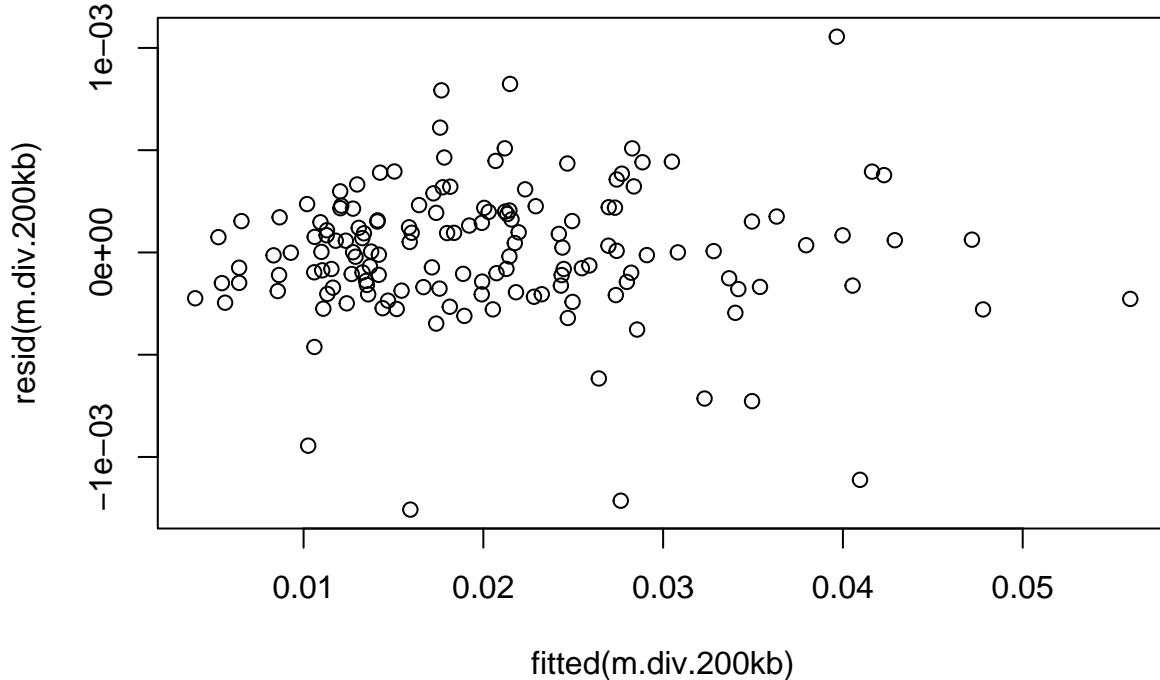
sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)
```

```

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))

```



```

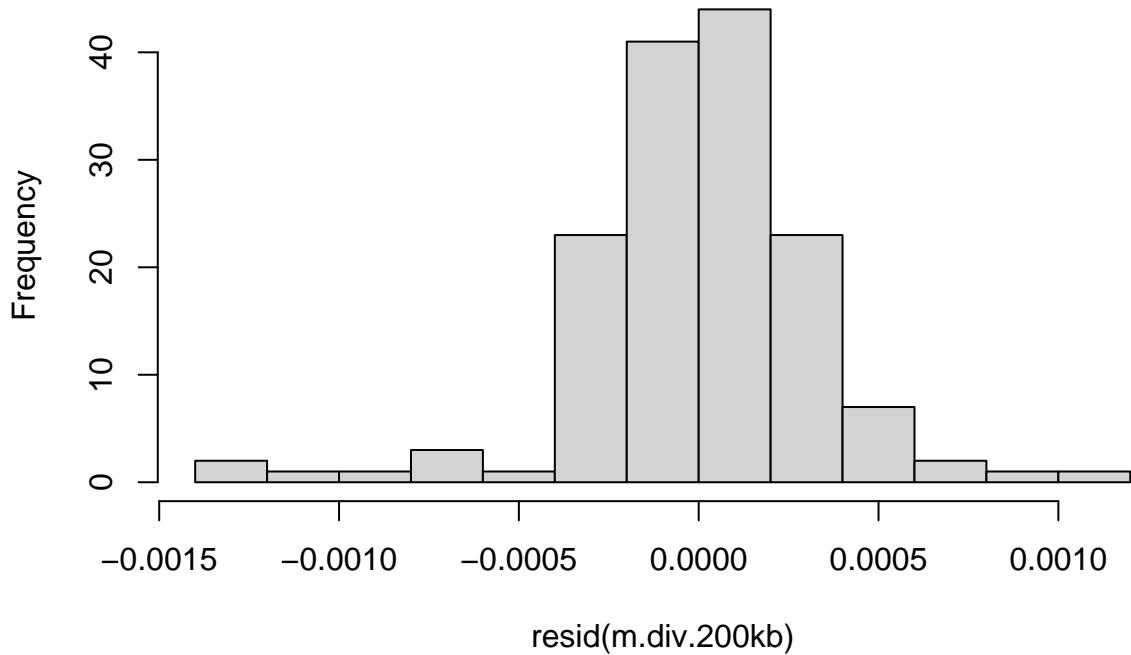
dwtest(m.div.200kb)

##
## 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 = 0.043
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
## -1.257e-03 -1.693e-04  3.530e-06  1.847e-04  1.055e-03
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.050e-02 2.710e-05 756.559 <2e-16 ***
## thetaC      9.689e-03 2.728e-05 355.236 <2e-16 ***
## rhoC        1.325e-05 2.730e-05   0.485   0.628
## tmrcaC      9.844e-04 2.744e-05  35.873 <2e-16 ***
## thetaC:tmrcaC 4.405e-04 2.620e-05 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", header=TRUE)
rep_5.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_5/sim.tmrca.200k.map", header=TRUE)

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

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

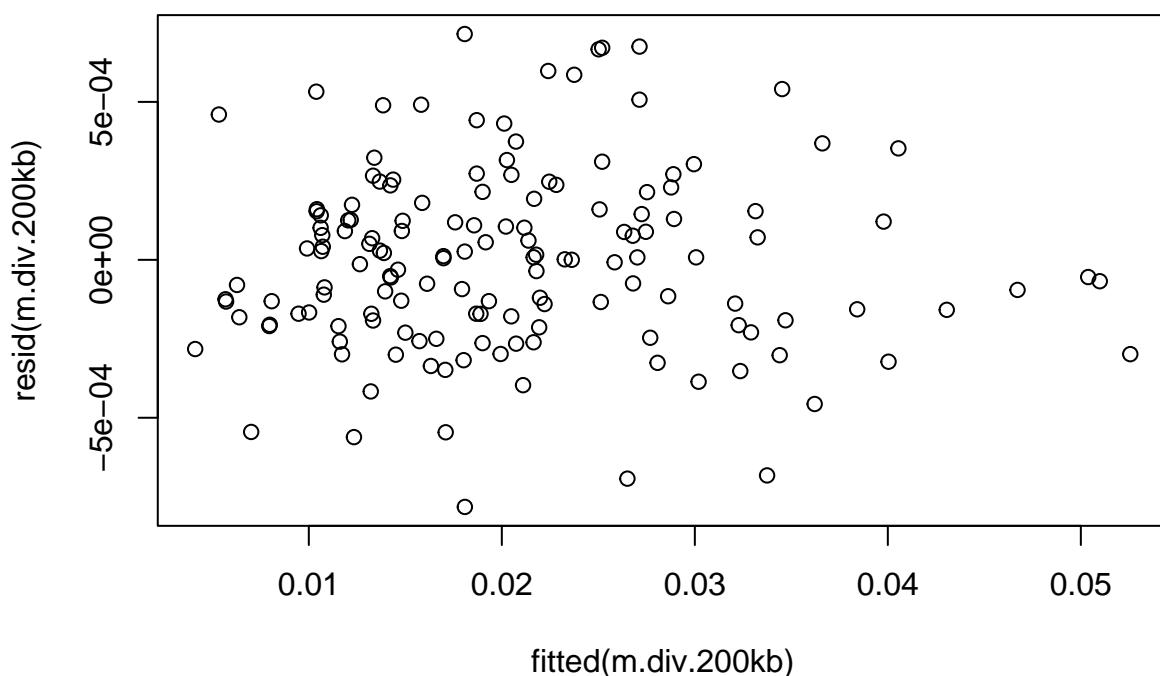
sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)
```

```

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))

```



```

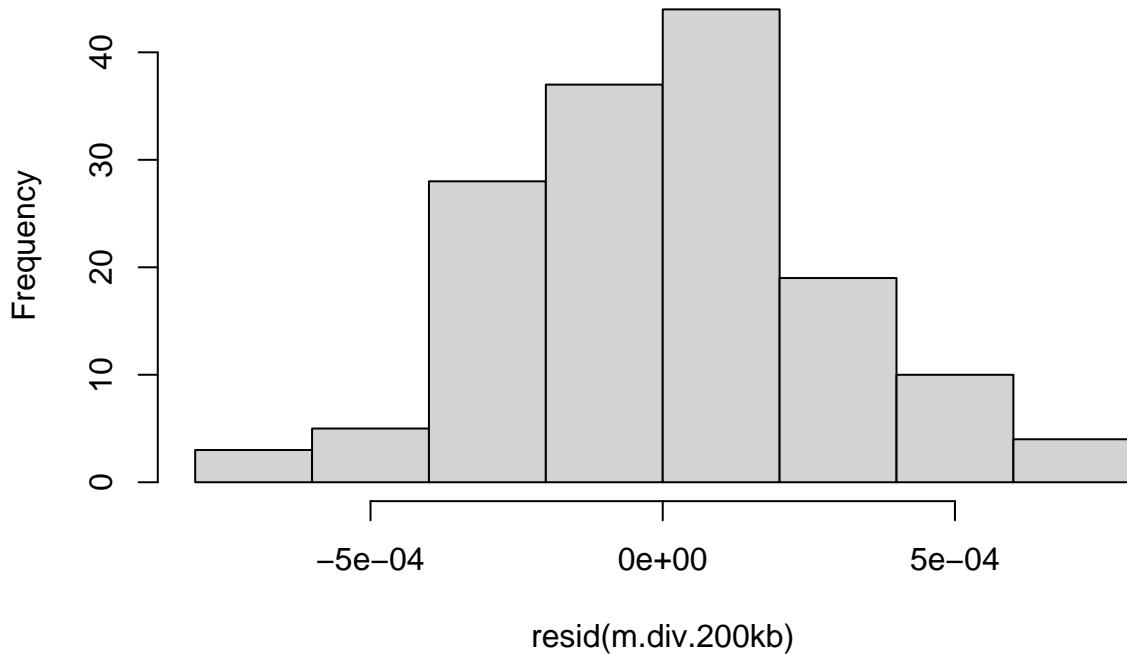
dwtest(m.div.200kb)

##
## 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 = 0.089
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 
## -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      9.728e-03 2.387e-05 407.532 <2e-16 ***
## rhoC        1.346e-05 2.392e-05  0.563  0.575    
## tmrcaC      9.268e-04 2.442e-05  37.957 <2e-16 ***
## thetaC:tmrcaC 3.965e-04 2.117e-05 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", header=TRUE)
rep_6.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_6/sim.tmrca.200k.map", header=TRUE)

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

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

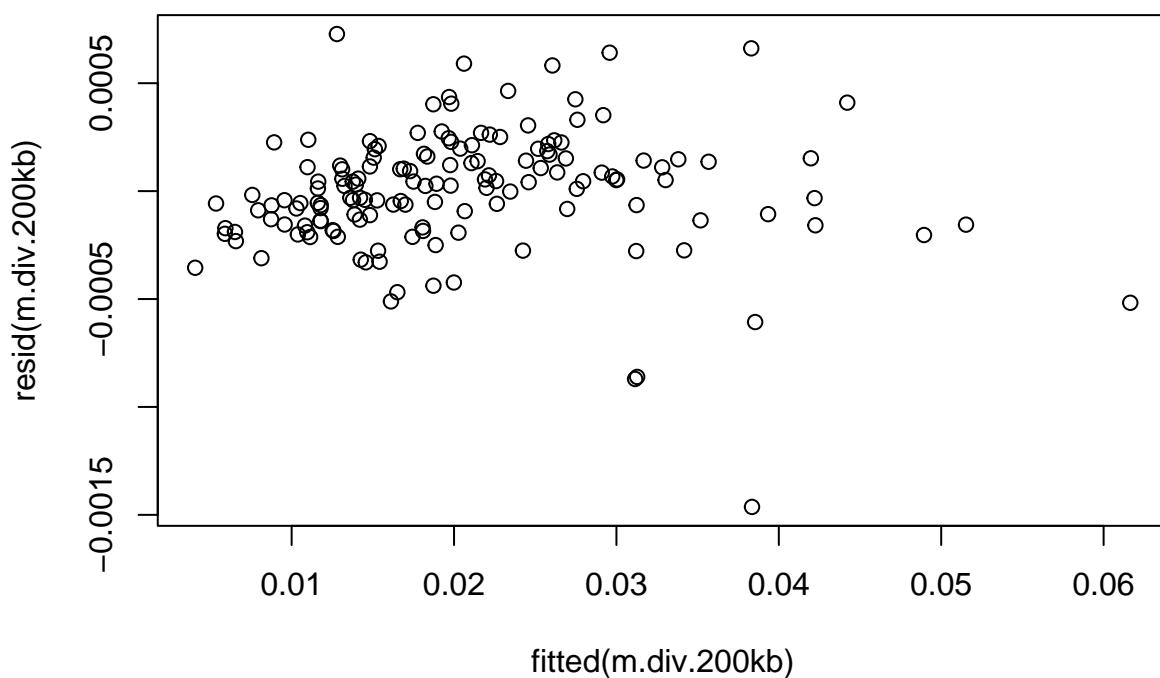
sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)
```

```

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)

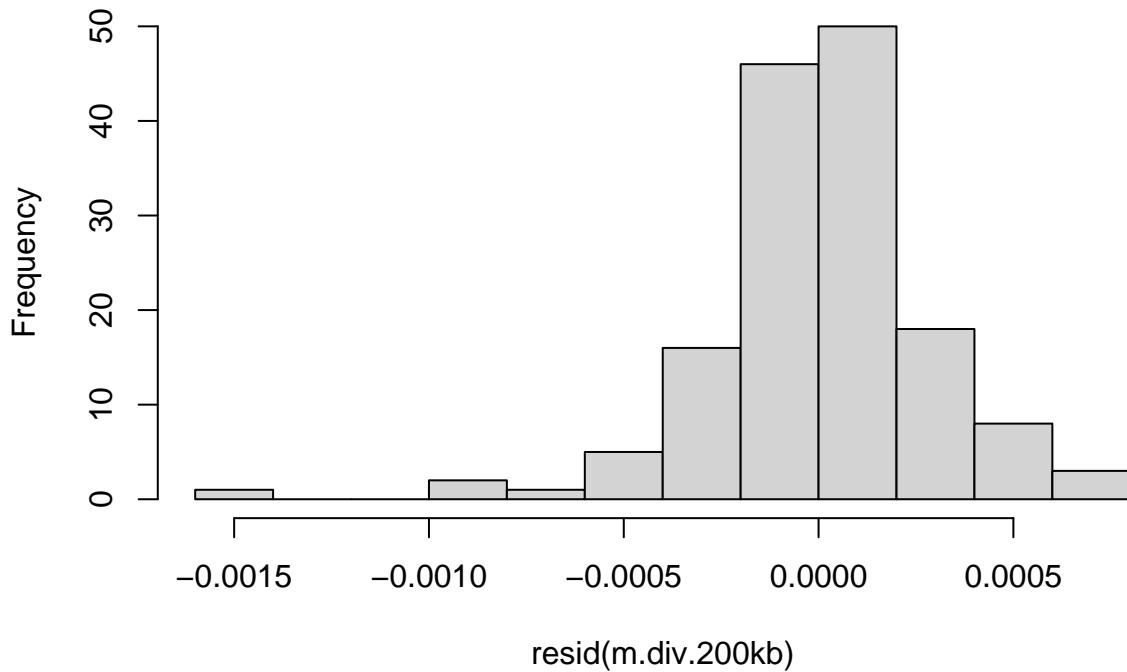
```

```

##
## Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.59616, p-value = 0.941
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
## -1.463e-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      9.676e-03 2.449e-05 395.145 <2e-16 ***
## rhoC        5.383e-06 2.351e-05   0.229    0.819
## tmrcaC     8.934e-04 2.365e-05   37.771 <2e-16 ***
## thetaC:tmrcaC 4.264e-04 1.890e-05  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", header=TRUE)
rep_7.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_7/sim.tmrca.200k.map", header=TRUE)

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

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

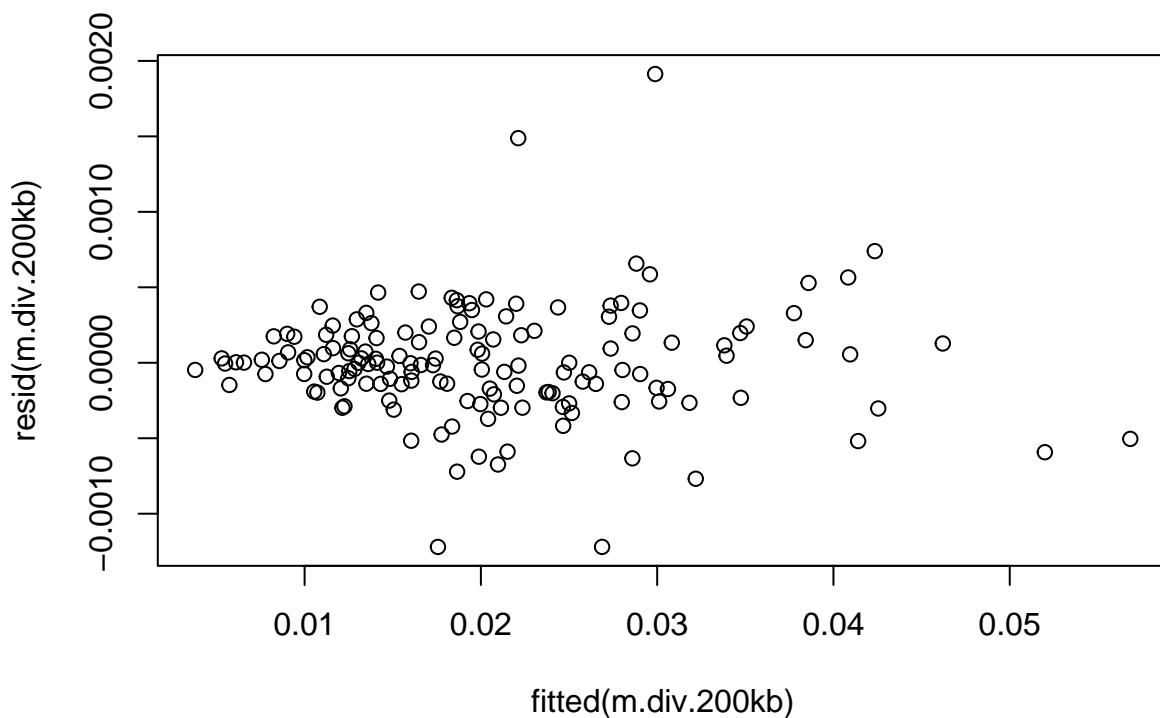
sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)
```

```

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))

```



```

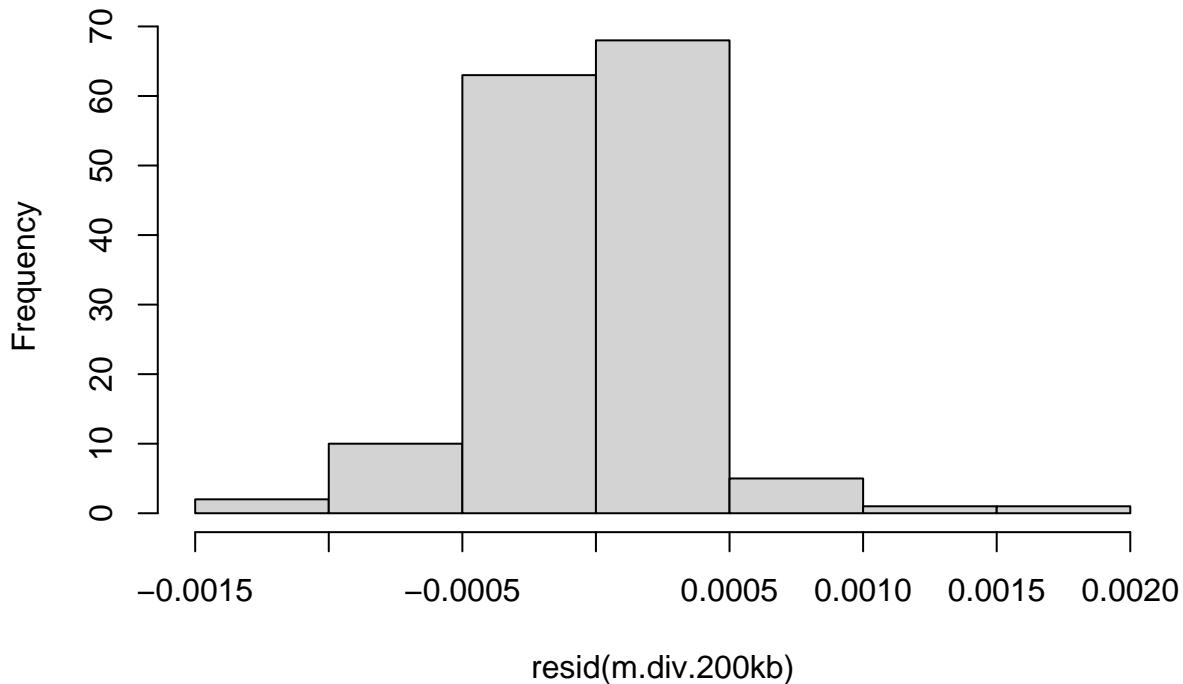
dwtest(m.div.200kb)

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

##
##  Harrison-McCabe test
##
##  data:  m.div.200kb
##  HMC = 0.40495, p-value = 0.052
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       9.795e-03 3.154e-05 310.525 <2e-16 ***
## rhoC        -2.443e-05 3.130e-05 -0.781  0.436    
## tmrcaC       8.987e-04 3.155e-05  28.486 <2e-16 ***
## thetaC:tmrcaC 4.019e-04 3.399e-05 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", header=TRUE)
rep_8.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_8/sim.tmrca.200k.map", header=TRUE)

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

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

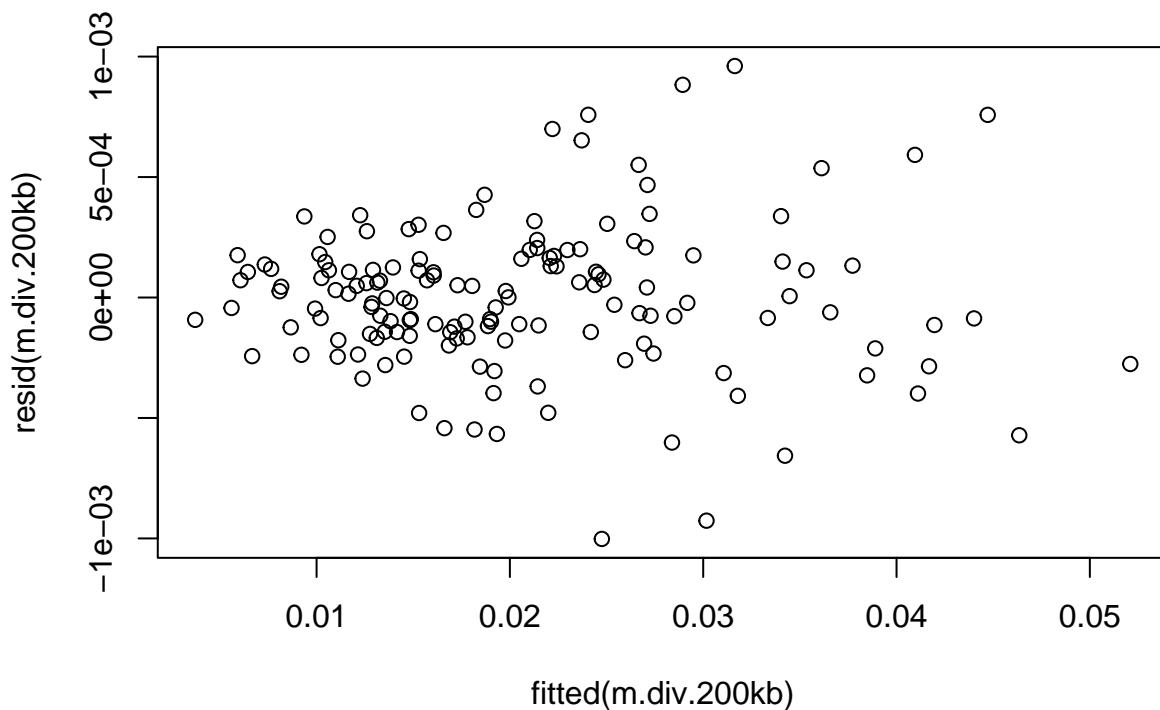
sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)
```

```

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

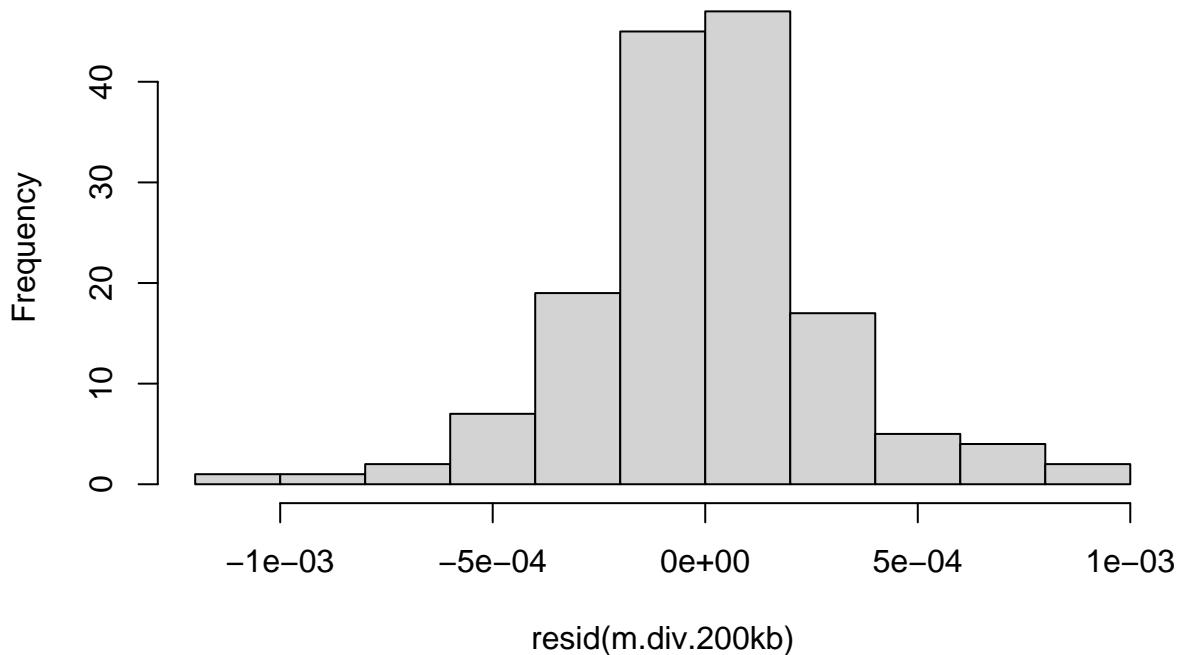
hmctest(m.div.200kb)

##
## Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.54355, p-value = 0.782

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
## -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      9.750e-03 2.590e-05 376.496 <2e-16 ***
## rhoC       -1.386e-05 2.578e-05 -0.538  0.592
## tmrcaC      9.878e-04 2.621e-05  37.681 <2e-16 ***
## thetaC:tmrcaC 4.413e-04 2.304e-05 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])
```

```
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))
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

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

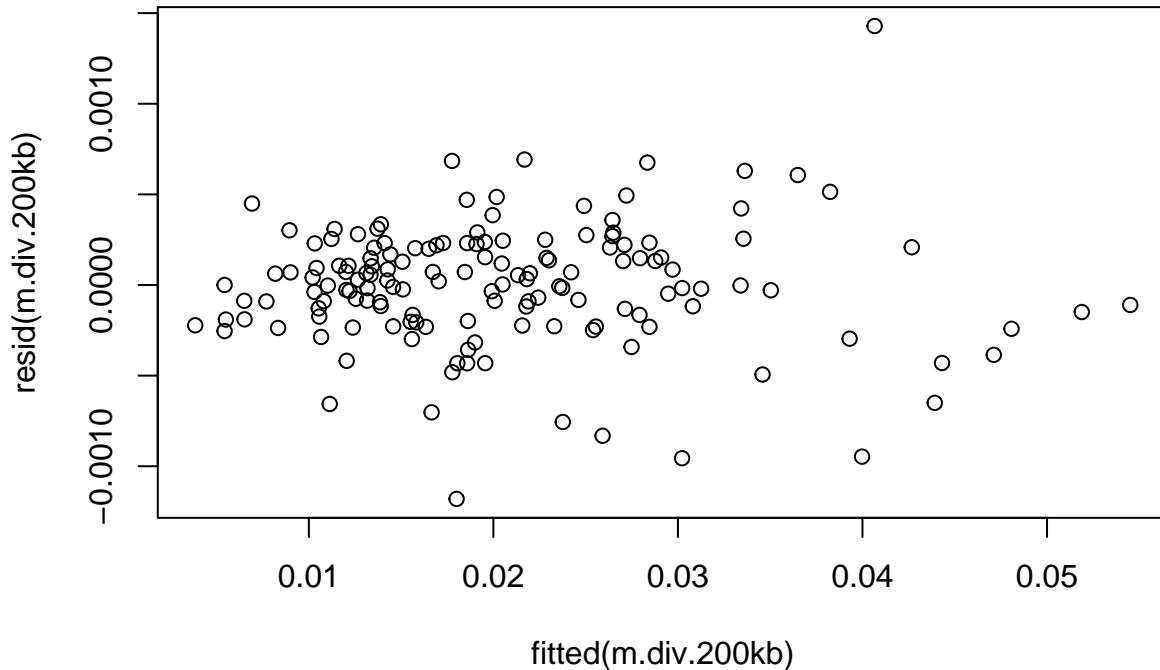
sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)
```

```

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))

```



```

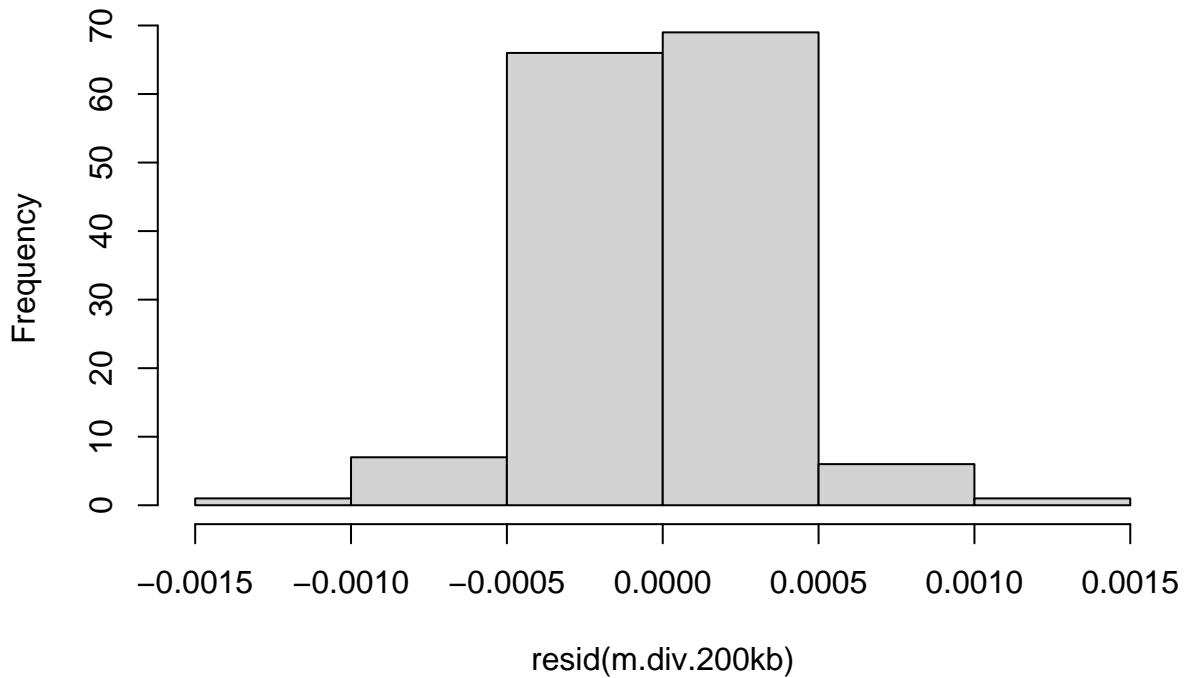
dwtest(m.div.200kb)

##
## 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))

```

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 
## -1.180e-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       9.742e-03 2.870e-05 339.390 <2e-16 ***
## rhoC        9.161e-06 2.834e-05   0.323   0.747    
## tmrcaC      9.056e-04 2.854e-05  31.733 <2e-16 ***
## thetaC:tmrcaC 4.515e-04 2.912e-05  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", header = TRUE)
rep_10.tmrca.200kb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_10/sim.tmrca.200k.map", header = TRUE)

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

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

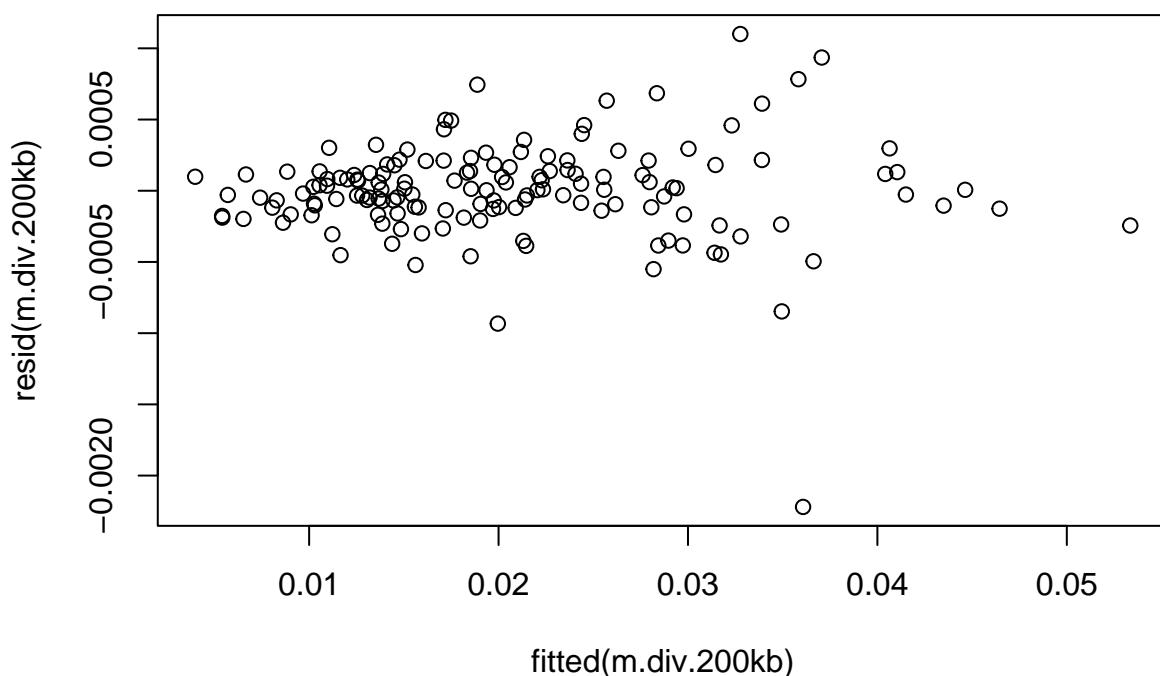
sim.lands.200kb$bin <- 1:nrow(sim.lands.200kb)
```

```

sim.lands.200kb.rep50 <- sim.lands.200kb
sim.lands.200kb.rep50$Replicate <- 10

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 = 2.1646, p-value = 0.8354
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.200kb)

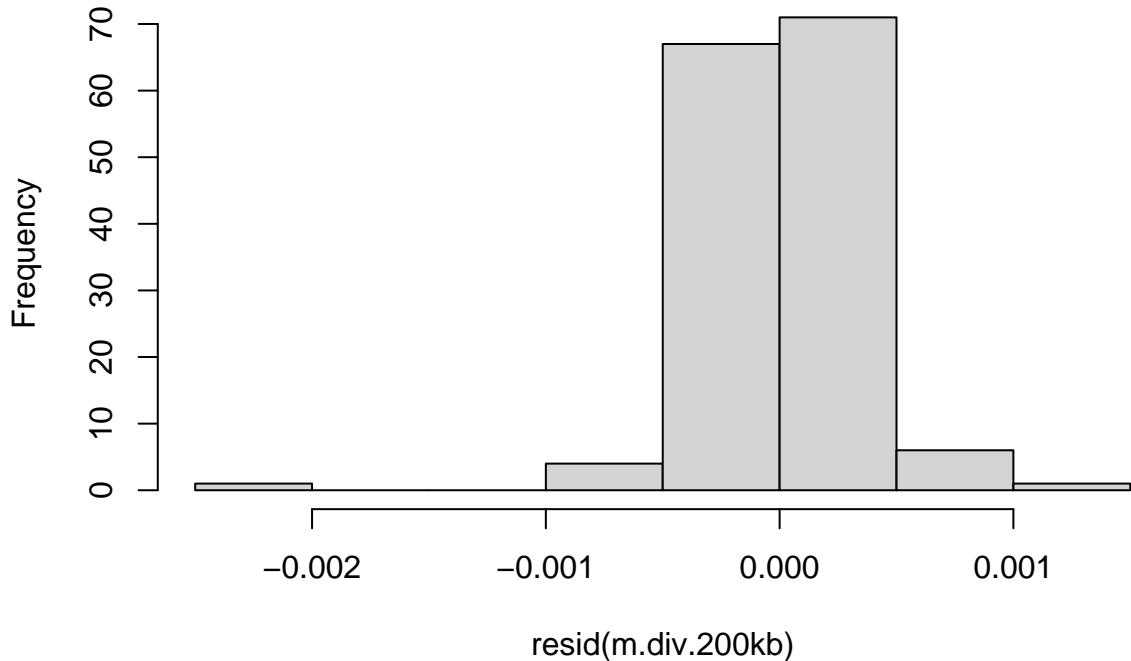
```

```

##
## Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.41358, p-value = 0.067
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 
## -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      9.806e-03 3.006e-05 326.195 <2e-16 ***
## rhoC        7.183e-06 2.907e-05  0.247   0.805    
## tmrcaC     1.083e-03 2.914e-05  37.182 <2e-16 ***
## thetaC:tmrcaC 5.833e-04 2.572e-05 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]) / 3
```

```

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))

```

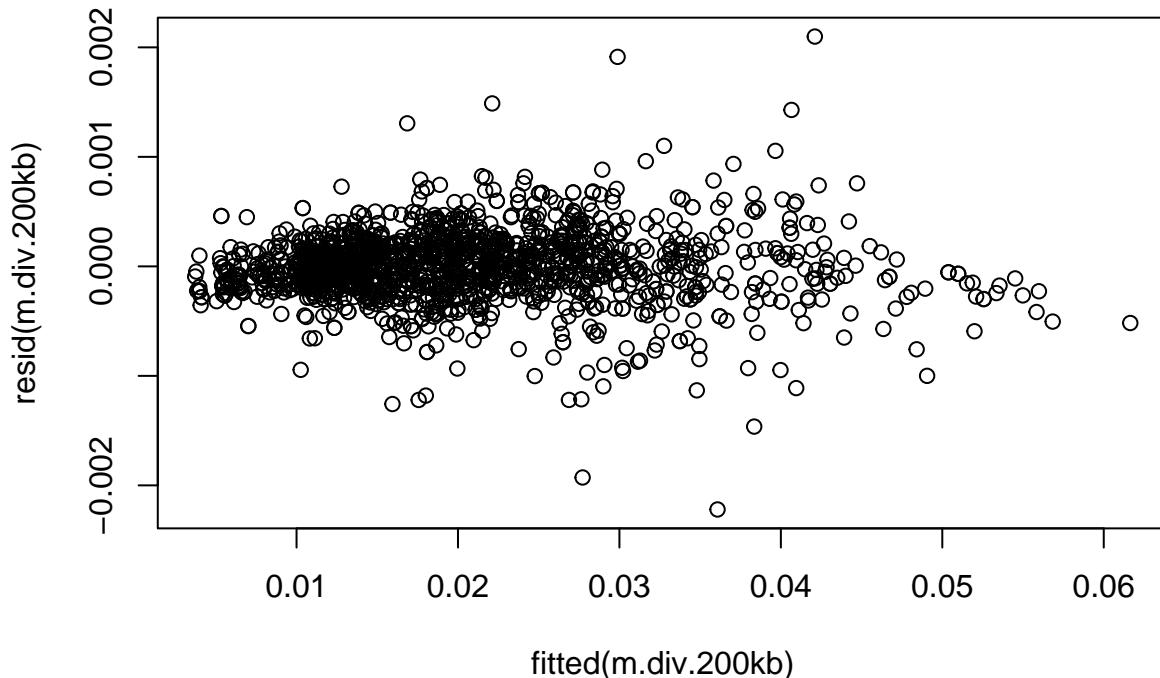
1.2.11 all replicates:

```

sim.lands.200kb.all <- rbind(sim.lands.200kb.rep5, sim.lands.200kb.rep2, sim.lands.200kb.rep3, sim.lands.200kb.rep1)

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 = 1.9357, p-value = 0.05828
##  alternative hypothesis: true autocorrelation is greater than 0

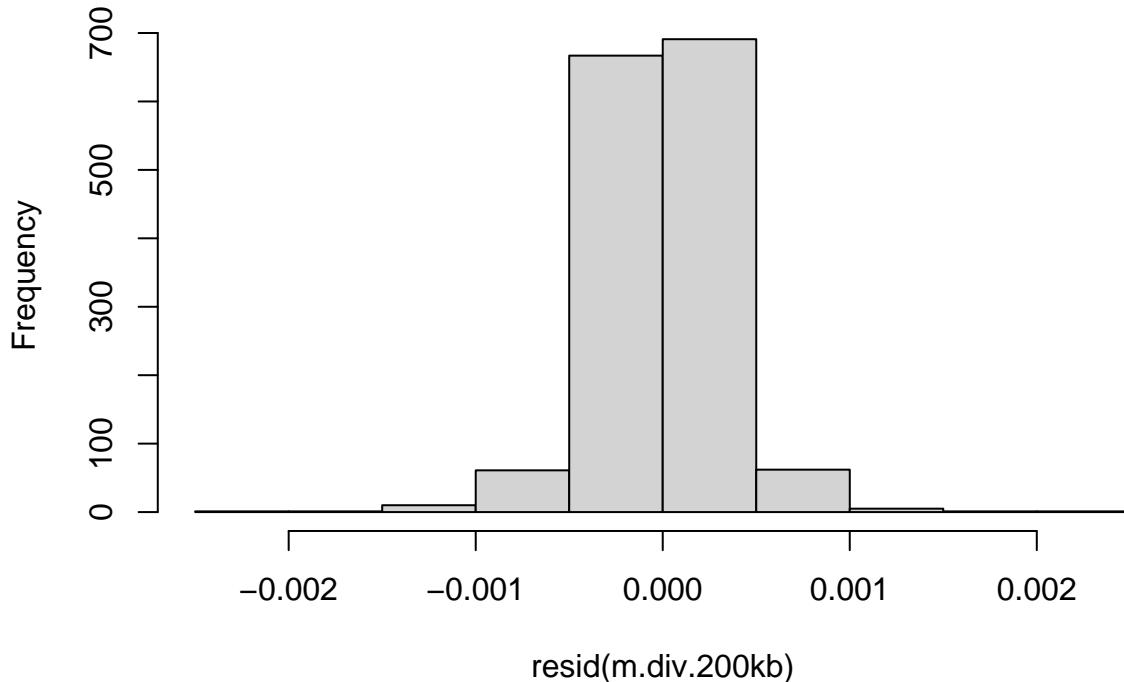
hmctest(m.div.200kb)

##
##  Harrison-McCabe test
##
##  data:  m.div.200kb
##  HMC = 0.48495, p-value = 0.195

```

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

Histogram of resid(m.div.200kb)



```
m.div.200kb.2 <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC)*as.factor(Replicate), data = sim.lands.200kb.all)
m.div.200kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC + rhoC:tmrcaC)*as.factor(Replicate), data = sim.lands.200kb.all)
```

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

```
##                df      AIC
## m.div.200kb    46 -19748.95
## m.div.200kb.2  55 -19741.46
## m.div.200kb.3  64 -19728.72
```

```
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
## -2.220e-03 -1.705e-04  1.390e-06  1.741e-04  2.099e-03
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.056e-02 2.693e-05 763.480 < 2e-16 ***
## thetaC      9.659e-03 2.728e-05 354.022 < 2e-16 ***
## rhoC       -4.702e-05 2.714e-05 -1.733 0.083394 .
## tmrcaC     1.008e-03 2.728e-05 36.941 < 2e-16 ***
## as.factor(Replicate)3 8.648e-05 3.808e-05 2.271 0.023308 *
```

```

## as.factor(Replicate)4      -5.611e-05  3.808e-05 -1.473 0.140907
## as.factor(Replicate)5      5.158e-05  3.298e-05  1.564 0.118060
## as.factor(Replicate)6     -3.139e-06  3.809e-05 -0.082 0.934334
## as.factor(Replicate)7      6.595e-05  3.833e-05  1.721 0.085525 .
## as.factor(Replicate)8      1.320e-05  3.816e-05  0.346 0.729361
## as.factor(Replicate)9      5.572e-05  3.824e-05  1.457 0.145329
## as.factor(Replicate)10     1.554e-04  3.824e-05  4.064 5.09e-05 ***
## thetaC:tmrcaC              5.023e-04  2.705e-05 18.565 < 2e-16 ***
## thetaC:as.factor(Replicate)3 8.710e-05  3.850e-05  2.263 0.023812 *
## thetaC:as.factor(Replicate)4 3.002e-05  3.846e-05  0.780 0.435225
## thetaC:as.factor(Replicate)5 6.865e-05  3.337e-05  2.057 0.039834 *
## thetaC:as.factor(Replicate)6 1.669e-05  3.926e-05  0.425 0.670855
## thetaC:as.factor(Replicate)7 1.354e-04  3.867e-05  3.502 0.000476 ***
## thetaC:as.factor(Replicate)8 9.122e-05  3.883e-05  2.349 0.018958 *
## thetaC:as.factor(Replicate)9 8.256e-05  3.874e-05  2.131 0.033240 *
## thetaC:as.factor(Replicate)10 1.467e-04  3.936e-05  3.728 0.000200 ***
## rhoC:as.factor(Replicate)3   9.478e-05  3.842e-05  2.467 0.013739 *
## rhoC:as.factor(Replicate)4   6.026e-05  3.837e-05  1.570 0.116550
## rhoC:as.factor(Replicate)5   6.047e-05  3.327e-05  1.818 0.069319 .
## rhoC:as.factor(Replicate)6   5.240e-05  3.835e-05  1.366 0.172044
## rhoC:as.factor(Replicate)7   2.259e-05  3.841e-05  0.588 0.556651
## rhoC:as.factor(Replicate)8   3.316e-05  3.864e-05  0.858 0.390926
## rhoC:as.factor(Replicate)9   5.618e-05  3.839e-05  1.463 0.143599
## rhoC:as.factor(Replicate)10  5.420e-05  3.859e-05  1.404 0.160426
## tmrcaC:as.factor(Replicate)3 2.663e-05  3.846e-05  0.692 0.488820
## tmrcaC:as.factor(Replicate)4 -2.333e-05  3.858e-05 -0.605 0.545332
## tmrcaC:as.factor(Replicate)5 -8.087e-05  3.362e-05 -2.406 0.016273 *
## tmrcaC:as.factor(Replicate)6 -1.144e-04  3.857e-05 -2.965 0.003075 **
## tmrcaC:as.factor(Replicate)7 -1.090e-04  3.867e-05 -2.818 0.004892 **
## tmrcaC:as.factor(Replicate)8 -1.991e-05  3.907e-05 -0.509 0.610508
## tmrcaC:as.factor(Replicate)9 -1.021e-04  3.862e-05 -2.643 0.008296 **
## tmrcaC:as.factor(Replicate)10 7.568e-05  3.874e-05  1.954 0.050926 .
## thetaC:tmrcaC:as.factor(Replicate)3 -4.651e-05  3.907e-05 -1.190 0.234055
## thetaC:tmrcaC:as.factor(Replicate)4 -6.175e-05  3.755e-05 -1.644 0.100311
## thetaC:tmrcaC:as.factor(Replicate)5 -1.058e-04  3.197e-05 -3.308 0.000963 ***
## thetaC:tmrcaC:as.factor(Replicate)6 -7.588e-05  3.473e-05 -2.185 0.029084 *
## thetaC:tmrcaC:as.factor(Replicate)7 -1.004e-04  4.004e-05 -2.508 0.012264 *
## thetaC:tmrcaC:as.factor(Replicate)8 -6.097e-05  3.655e-05 -1.668 0.095556 .
## thetaC:tmrcaC:as.factor(Replicate)9 -5.077e-05  3.886e-05 -1.306 0.191672
## thetaC:tmrcaC:as.factor(Replicate)10 8.104e-05  3.635e-05  2.230 0.025928 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0003297 on 1455 degrees of freedom
## Multiple R-squared:  0.9989, Adjusted R-squared:  0.9989
## F-statistic: 2.96e+04 on 44 and 1455 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

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

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

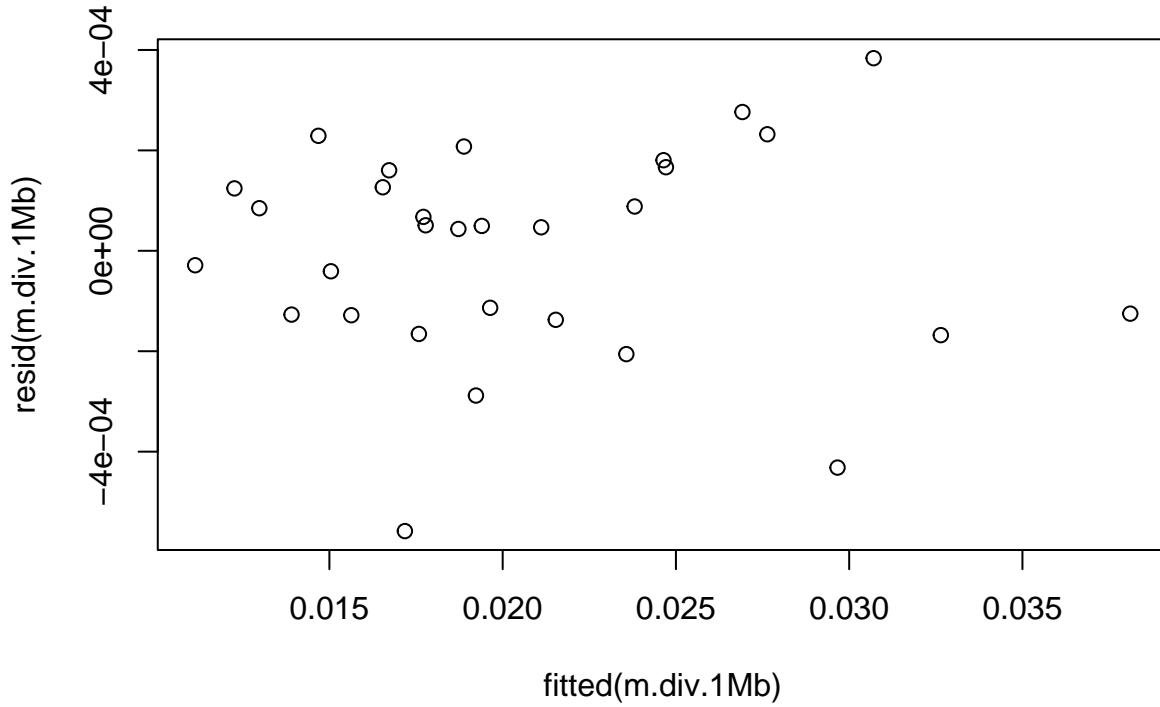
```

```

sim.lands.1Mb.rep5 <- sim.lands.1Mb
sim.lands.1Mb.rep5$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))

```



```

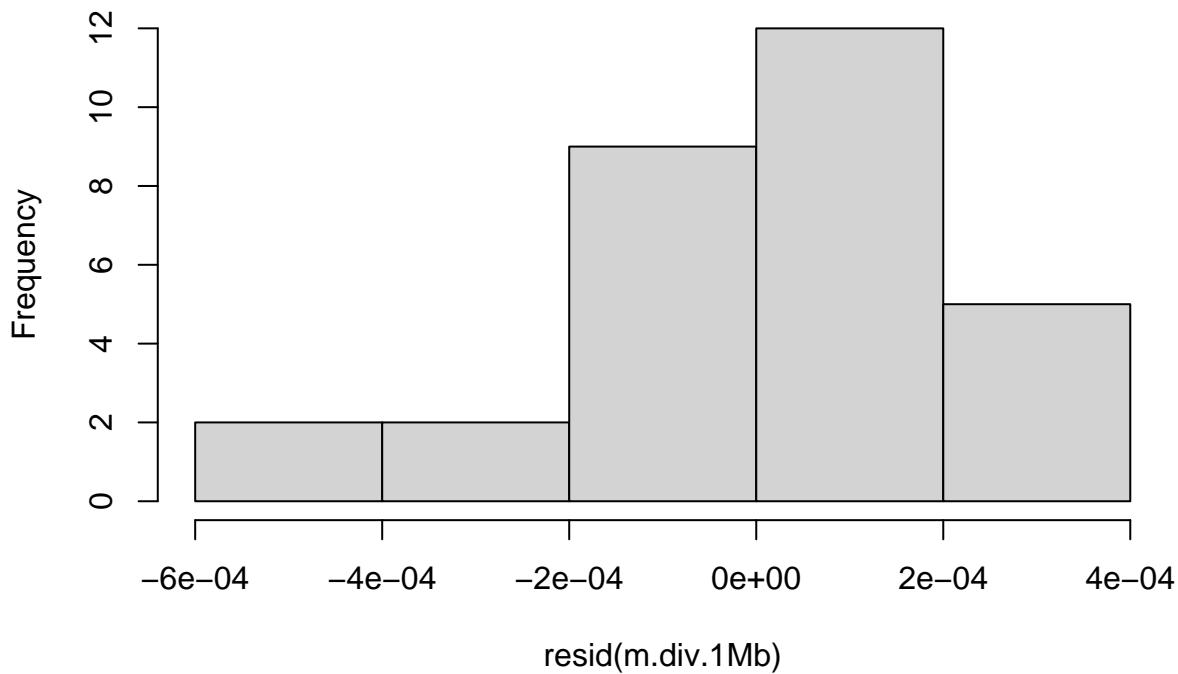
dwtest(m.div.1Mb)

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

##
##  Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.59412, p-value = 0.773
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      6.406e-03 4.696e-05 136.418 < 2e-16 ***
## rhoC       -1.598e-06 4.452e-05 -0.036  0.9717    
## tmrcaC      3.863e-04 4.278e-05  9.031 2.41e-09 ***
## thetaC:tmrcaC 8.560e-05 4.532e-05  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] +
```

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

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

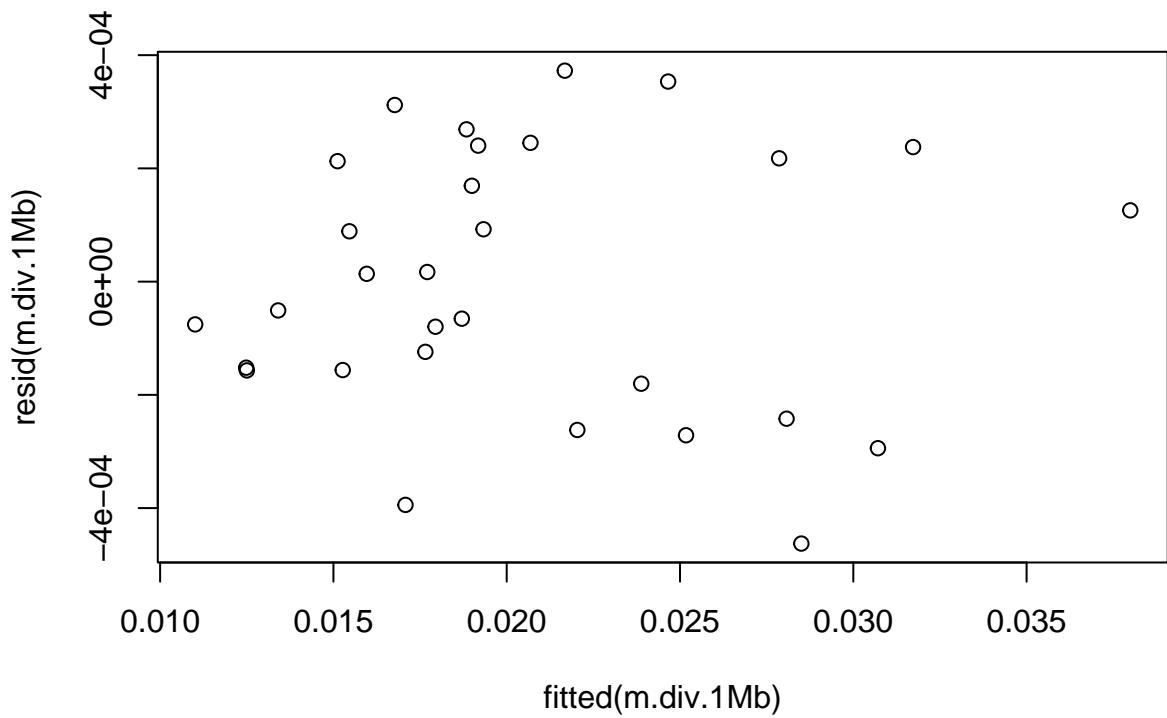
# standardizing
sim.lands.1Mb$thetaC <- (sim.lands.1Mb$theta - mean(sim.lands.1Mb$theta)) / sd(sim.lands.1Mb$theta)
sim.lands.1Mb$tmrcaC <- (sim.lands.1Mb$tmrca - mean(sim.lands.1Mb$tmrca)) / sd(sim.lands.1Mb$tmrca)
sim.lands.1Mb$rhoC <- (sim.lands.1Mb$rho - mean(sim.lands.1Mb$rho)) / sd(sim.lands.1Mb$rho)

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

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

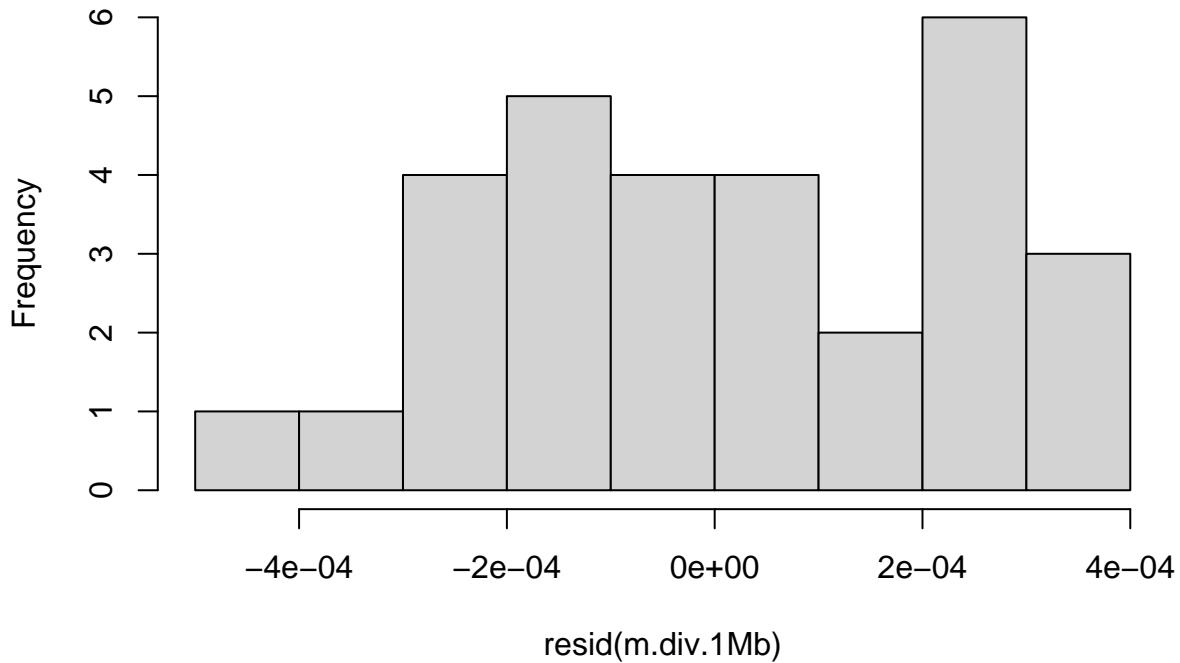
hmctest(m.div.1Mb)

##
##  Harrison-McCabe test
##
##  data:  m.div.1Mb
##  HMC = 0.51119, p-value = 0.539

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       6.270e-03 4.942e-05 126.882 < 2e-16 ***
## rhoC        -2.992e-05 4.850e-05 -0.617 0.542900  
## tmrcaC       4.395e-04 4.996e-05  8.796 3.99e-09 ***
## thetaC:tmrcaC 2.033e-04 4.644e-05  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])
```

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

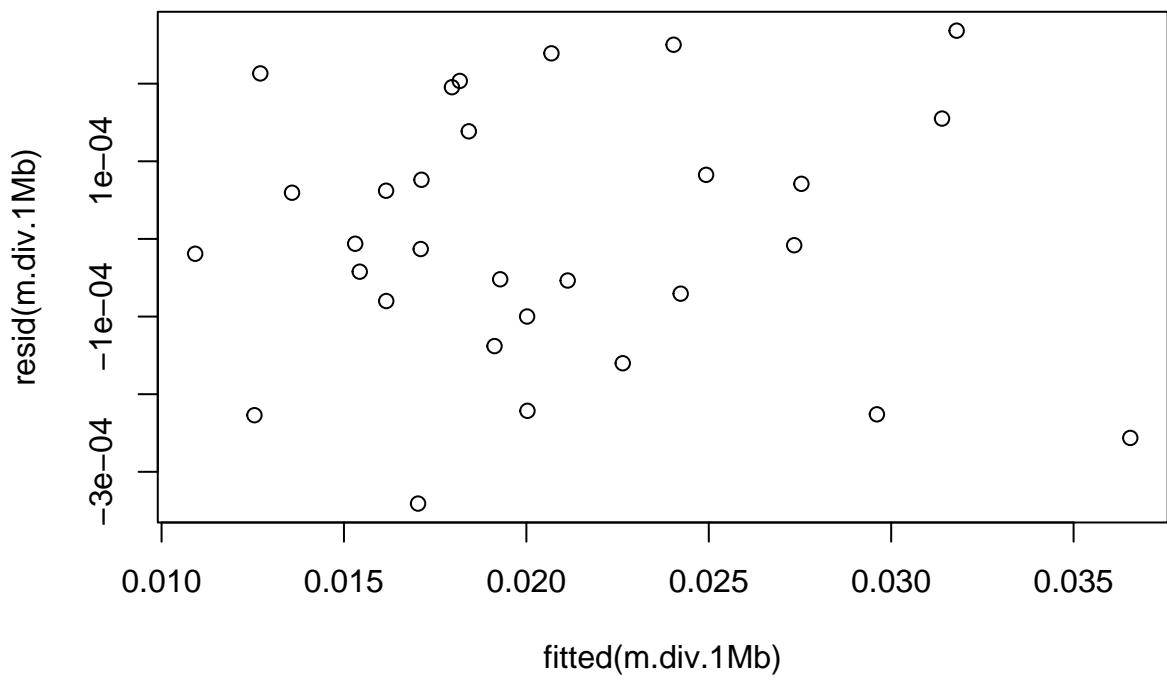
# standardizing
sim.lands.1Mb$thetaC <- (sim.lands.1Mb$theta - mean(sim.lands.1Mb$theta)) / sd(sim.lands.1Mb$theta)
sim.lands.1Mb$tmrcaC <- (sim.lands.1Mb$tmrca - mean(sim.lands.1Mb$tmrca)) / sd(sim.lands.1Mb$tmrca)
sim.lands.1Mb$rhoC <- (sim.lands.1Mb$rho - mean(sim.lands.1Mb$rho)) / sd(sim.lands.1Mb$rho)

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

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))

```

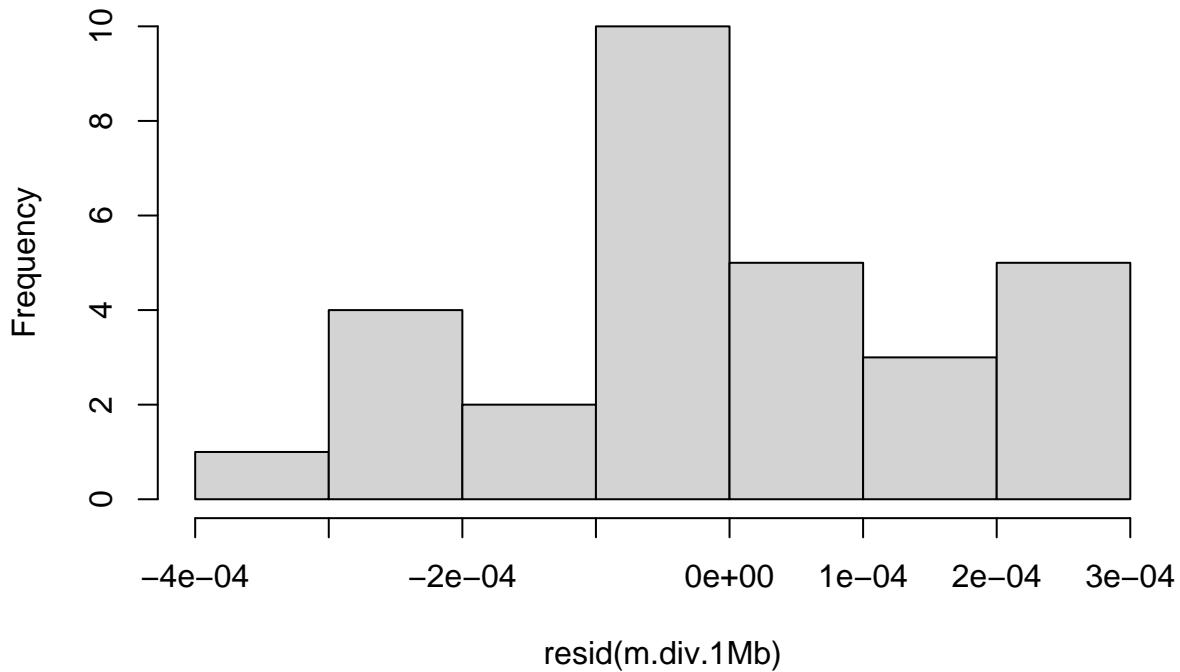


```
dwtest(m.div.1Mb)
```

```
##
##  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 = 0.93
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      6.469e-03 3.722e-05 173.822 < 2e-16 ***
## rhoC        4.023e-05 3.455e-05   1.164   0.255
## tmrcaC     5.563e-04 3.721e-05  14.949 5.67e-14 ***
## thetaC:tmrcaC 2.292e-04 4.592e-05    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])
```

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

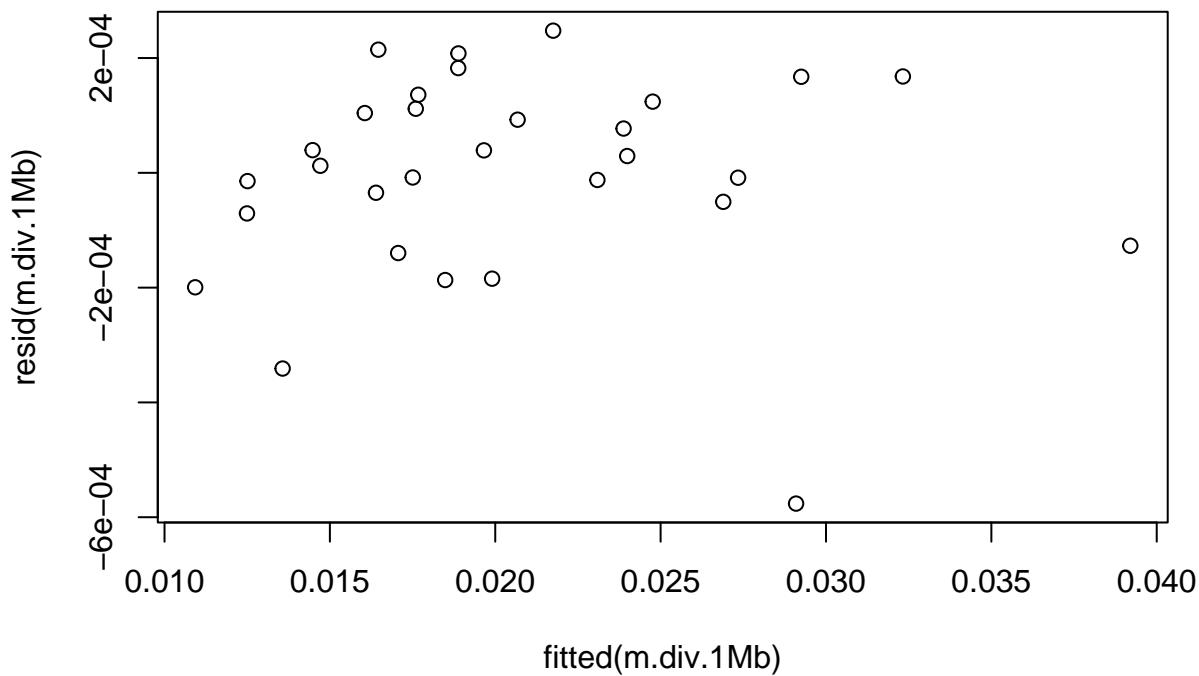
# standardizing
sim.lands.1Mb$thetaC <- (sim.lands.1Mb$theta - mean(sim.lands.1Mb$theta)) / sd(sim.lands.1Mb$theta)
sim.lands.1Mb$tmrcaC <- (sim.lands.1Mb$tmrca - mean(sim.lands.1Mb$tmrca)) / sd(sim.lands.1Mb$tmrca)
sim.lands.1Mb$rhoC <- (sim.lands.1Mb$rho - mean(sim.lands.1Mb$rho)) / sd(sim.lands.1Mb$rho)

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

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))

```

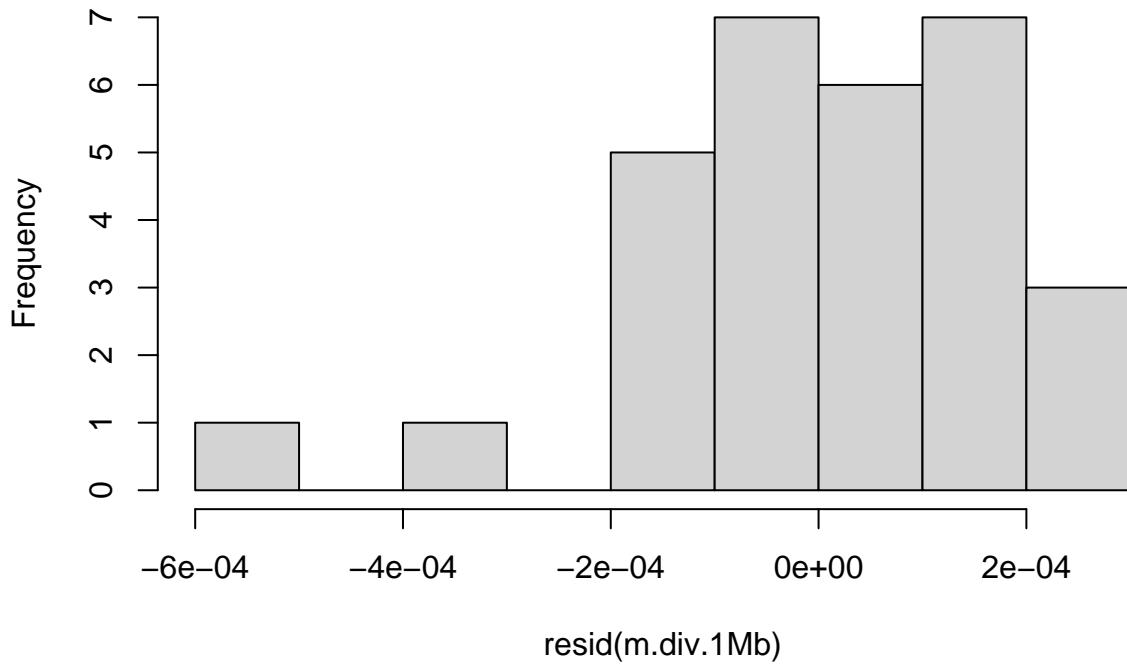


```
dwtest(m.div.1Mb)
```

```
##
##  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 = 0.237
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       6.295e-03 3.811e-05 165.192 < 2e-16 ***
## rhoC        6.522e-06 3.753e-05  0.174   0.863  
## tmrcaC      4.575e-04 3.850e-05 11.884 8.84e-12 ***
## thetaC:tmrcaC 1.915e-04 2.839e-05  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] +
```

```
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

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

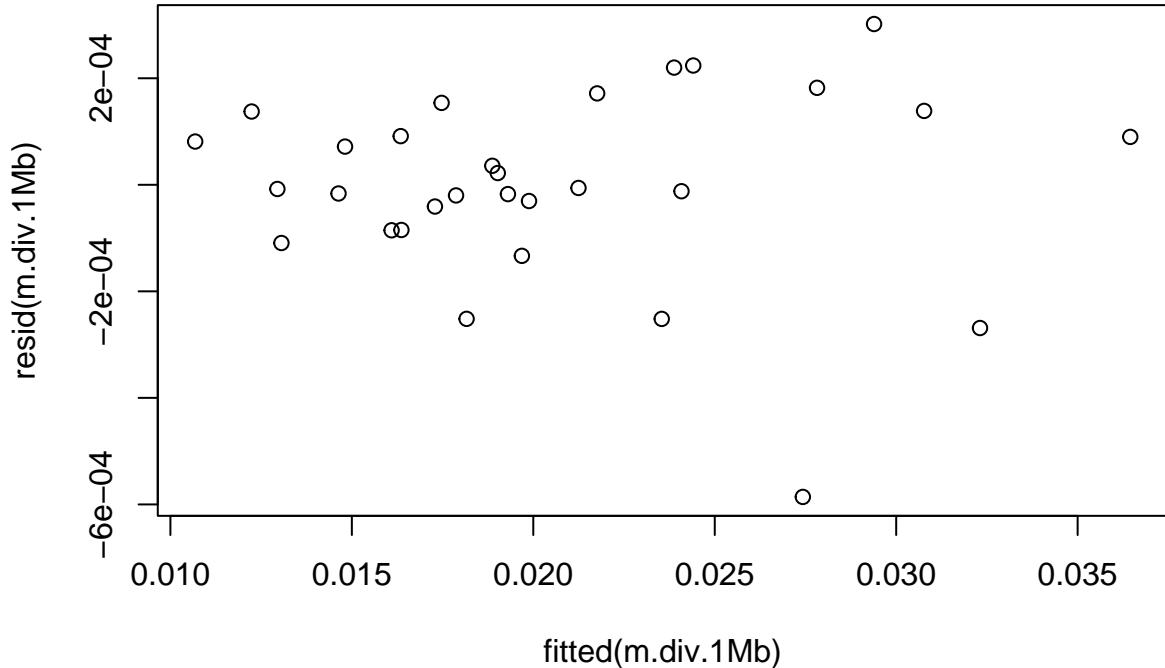
sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)
```

```

sim.lands.1Mb.rep5 <- sim.lands.1Mb
sim.lands.1Mb.rep5$Replicate <- 5

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

```



```

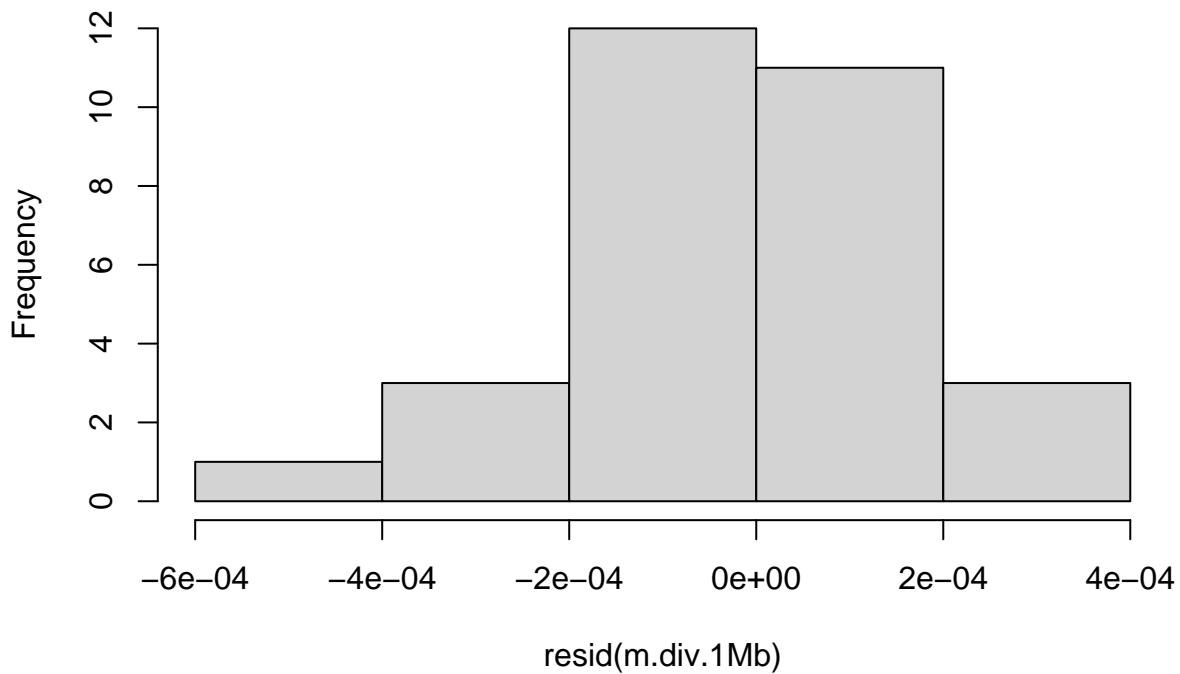
dwtest(m.div.1Mb)

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

##
## Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.35533, p-value = 0.116
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
## -5.859e-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      6.371e-03 3.829e-05 166.389 < 2e-16 ***
## rhoC       -2.259e-05 3.696e-05 -0.611   0.5467
## tmrcaC     4.195e-04 3.599e-05 11.655 1.34e-11 ***
## thetaC:tmrcaC 9.237e-05 4.113e-05  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

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

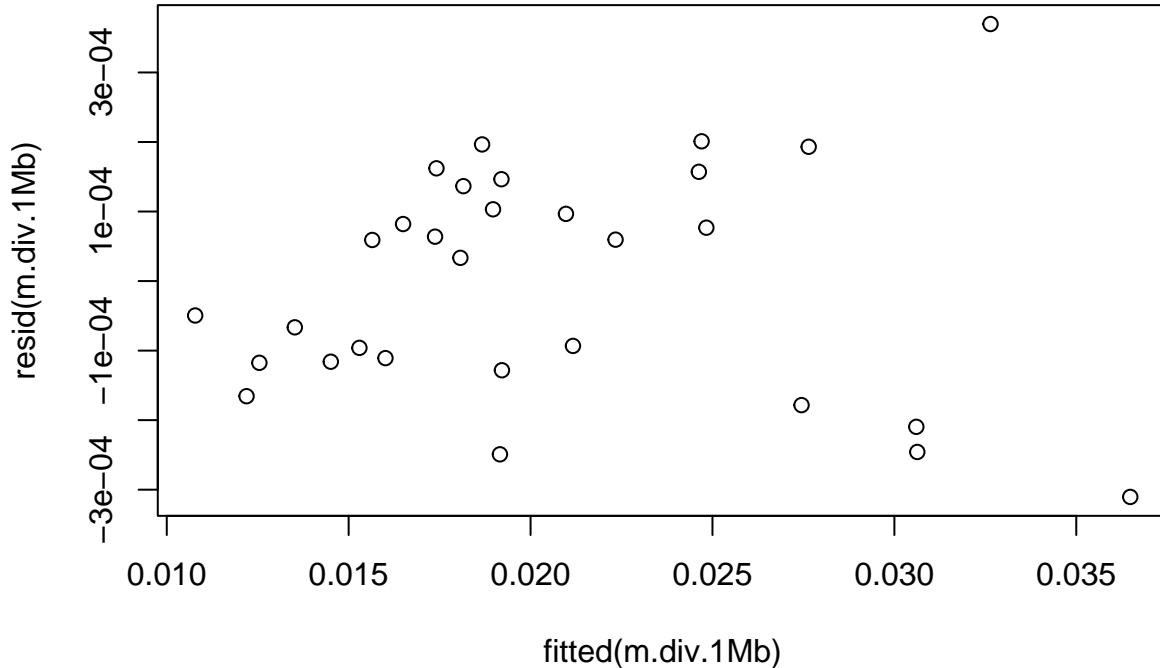
sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)
```

```

sim.lands.1Mb.rep6 <- sim.lands.1Mb
sim.lands.1Mb.rep6$Replicate <- 6

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

```



```

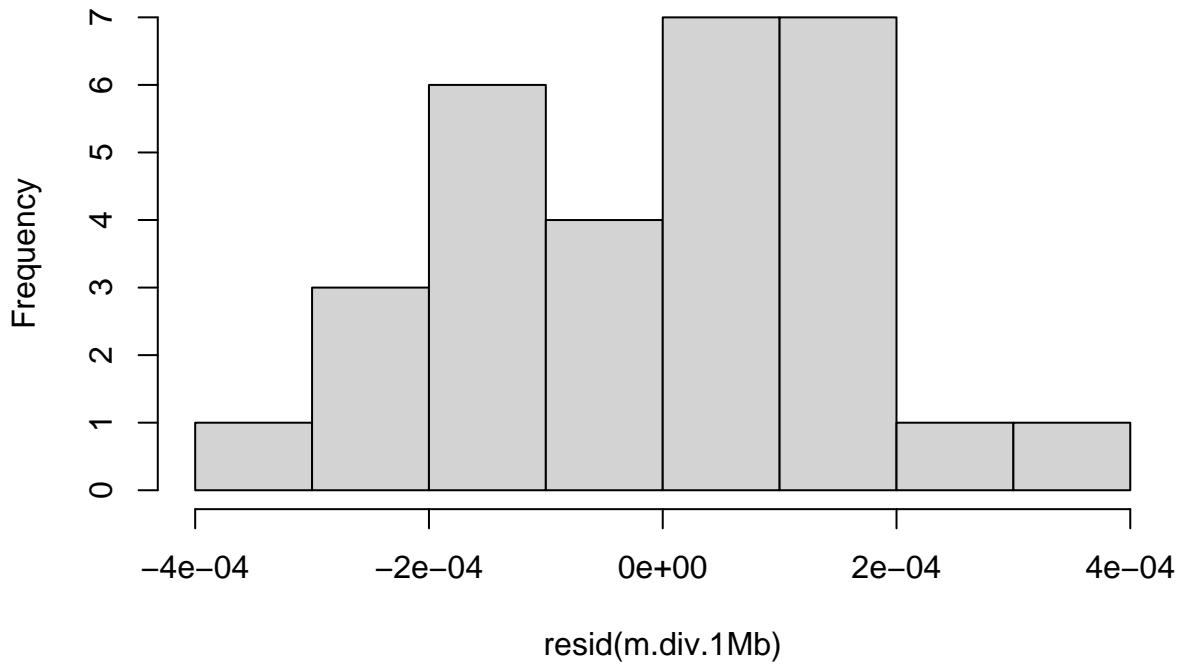
dwtest(m.div.1Mb)

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

##
## Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.62113, p-value = 0.834
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 
## -3.104e-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      6.365e-03 3.492e-05 182.260 < 2e-16 ***
## rhoC        4.494e-05 3.462e-05  1.298  0.20607  
## tmrcaC      3.544e-04 3.476e-05 10.193 2.18e-10 ***
## thetaC:tmrcaC 1.391e-04 4.149e-05  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] +
```

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

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

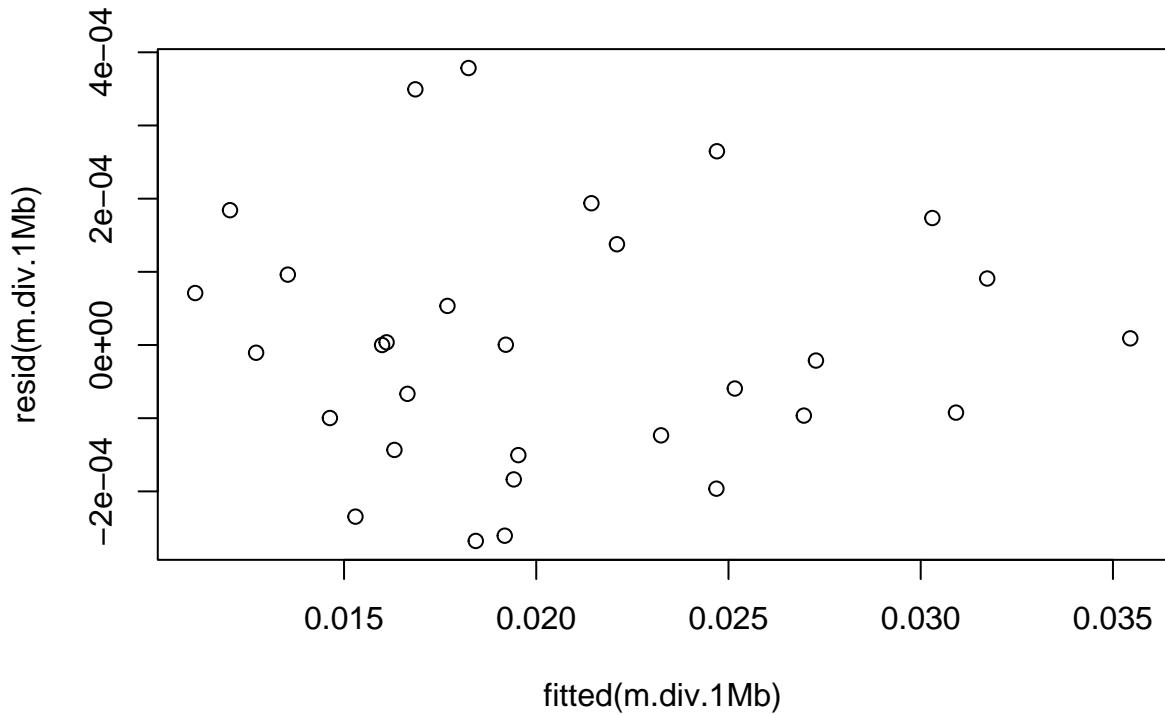
sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)
```

```

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))

```



```

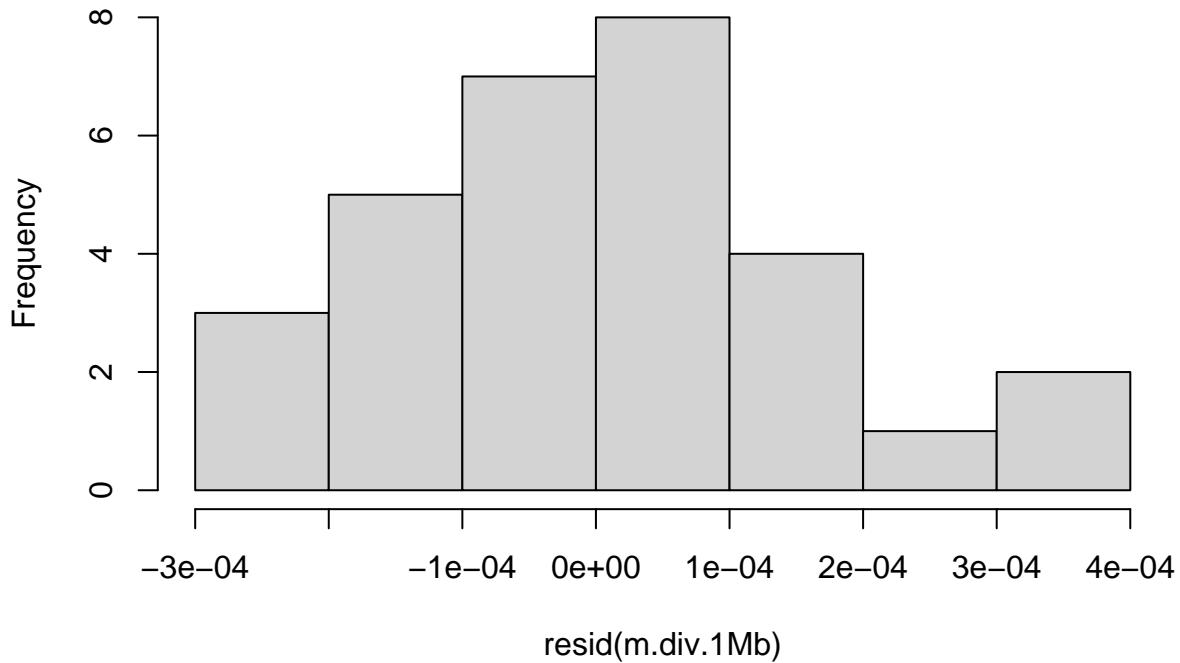
dwtest(m.div.1Mb)

##
##  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 = 0.881
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
## -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      6.450e-03 3.767e-05 171.218 < 2e-16 ***
## rhoC       -3.255e-05 3.520e-05 -0.925 0.363900
## tmrcaC      5.480e-04 3.534e-05 15.507 2.47e-14 ***
## thetaC:tmrcaC 1.411e-04 3.091e-05  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 = TRUE)
rep_8.tmrca.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_8/sim.tmrca.1M.map", header = TRUE)

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

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

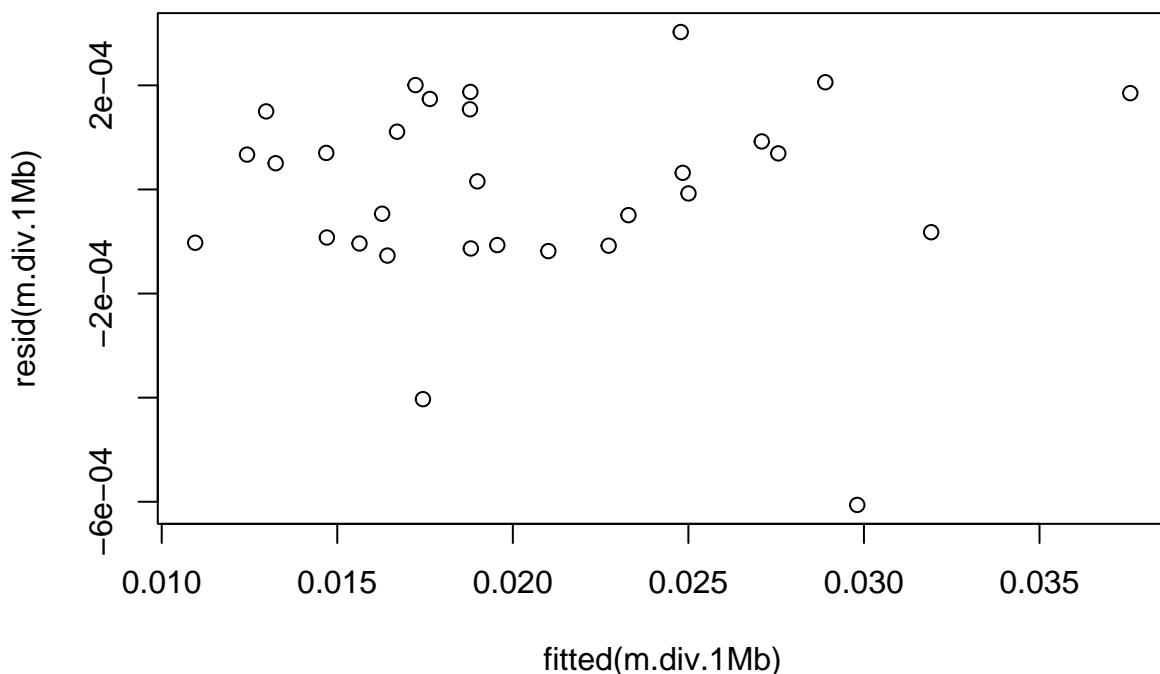
sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)
```

```

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))

```



```

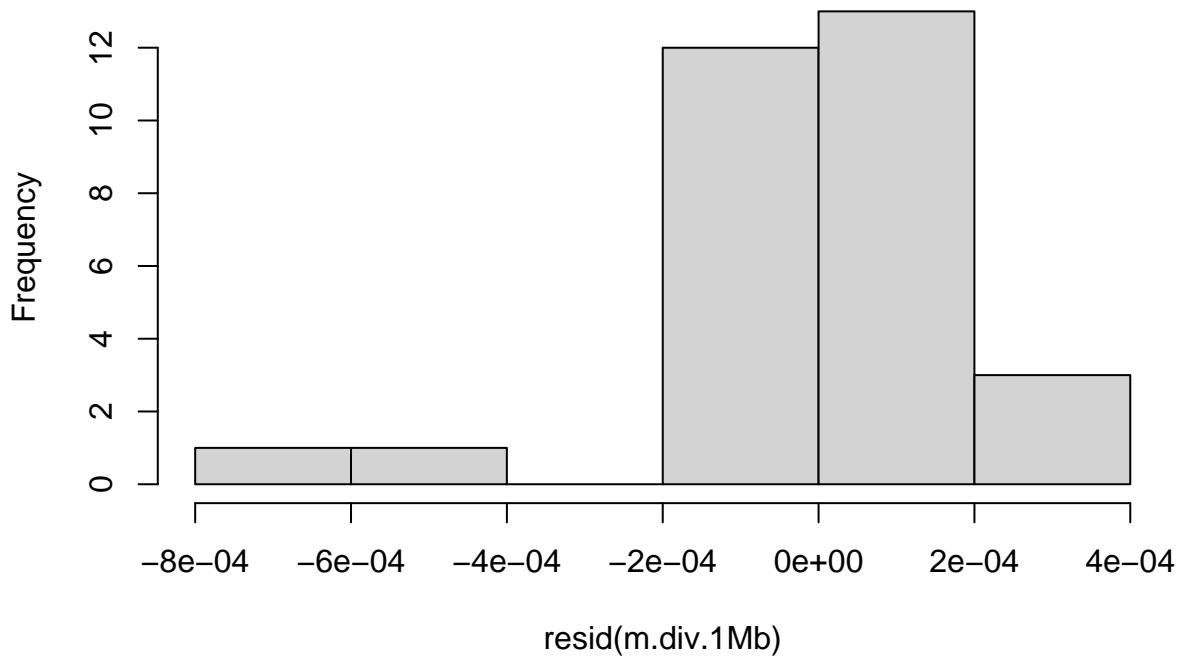
dwtest(m.div.1Mb)

##
## 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 = 0.946
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 
## -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      6.314e-03 3.861e-05 163.527 < 2e-16 ***
## rhoC        1.492e-05 3.797e-05  0.393  0.69767  
## tmrcaC      4.378e-04 3.748e-05 11.682 1.27e-11 ***
## thetaC:tmrcaC 1.607e-04 4.505e-05  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] +
```

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

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

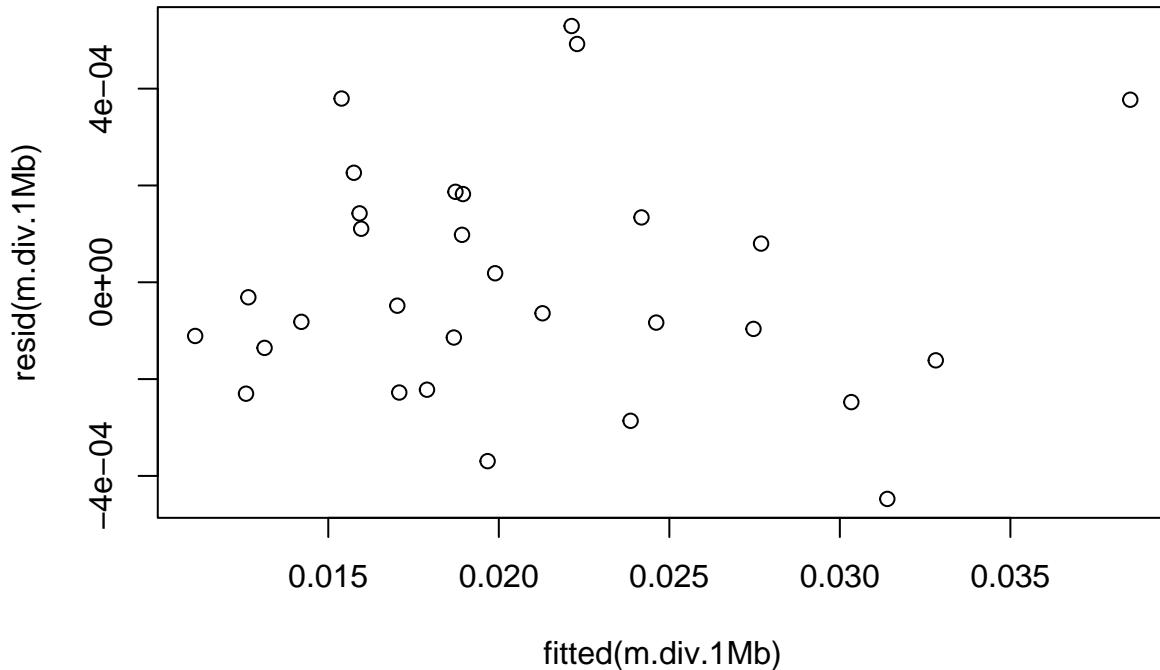
sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)
```

```

sim.lands.1Mb.rep9 <- sim.lands.1Mb
sim.lands.1Mb.rep9$Replicate <- 9

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

```



```

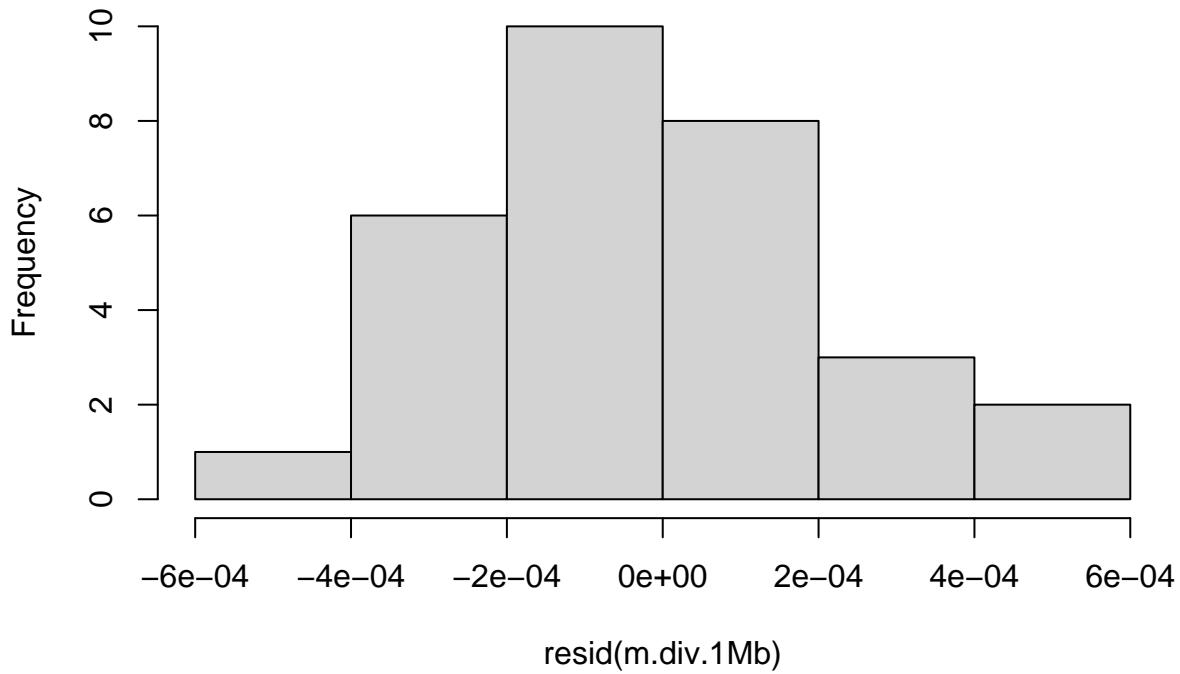
dwtest(m.div.1Mb)

##
## Durbin-Watson test
##
## data: m.div.1Mb
## DW = 2.2409, 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 = 0.709
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
## -0.0004475 -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      6.419e-03 6.079e-05 105.593 < 2e-16 ***
## rhoC        2.282e-05 5.237e-05   0.436  0.6667
## tmrcaC      4.840e-04 5.292e-05   9.145 1.89e-09 ***
## thetaC:tmrcaC 1.178e-04 5.427e-05    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 = FALSE)
rep_10.tmrca.1Mb <- read.table("dm_data/dm_chr_maps/2L/dm_coal_sims_2/rep_10/sim.tmrca.1M.map", header = FALSE)

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

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

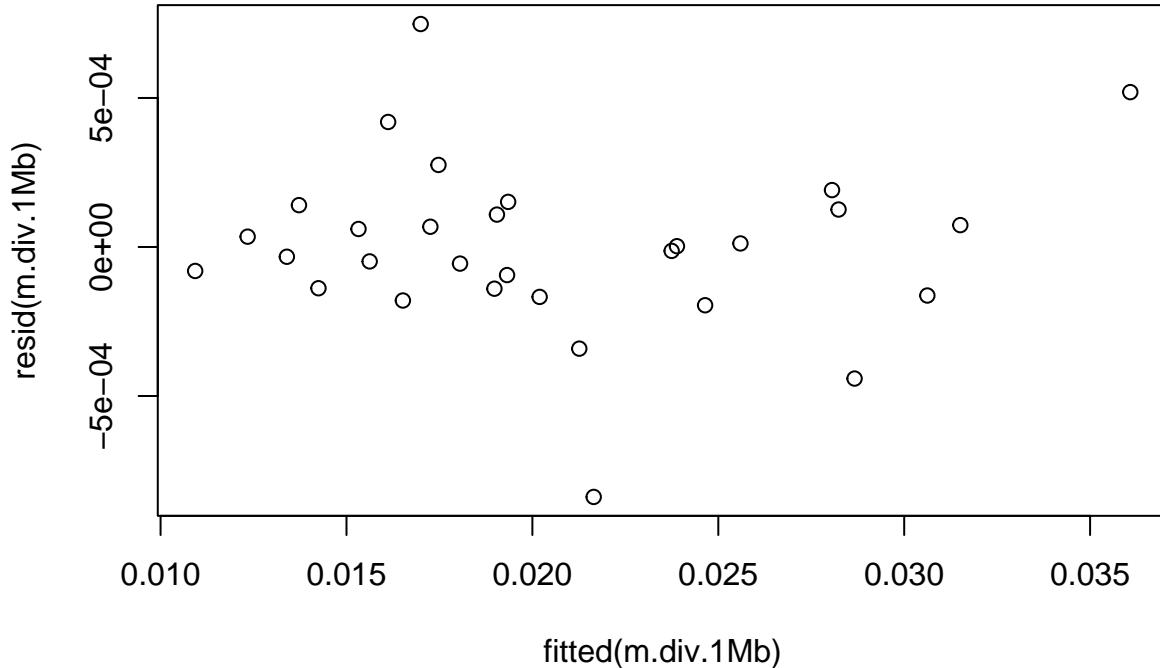
sim.lands.1Mb$bin <- 1:nrow(sim.lands.1Mb)
```

```

sim.lands.1Mb.rep50 <- sim.lands.1Mb
sim.lands.1Mb.rep50$Replicate <- 10

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 = 2.2351, p-value = 0.7114
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.1Mb)

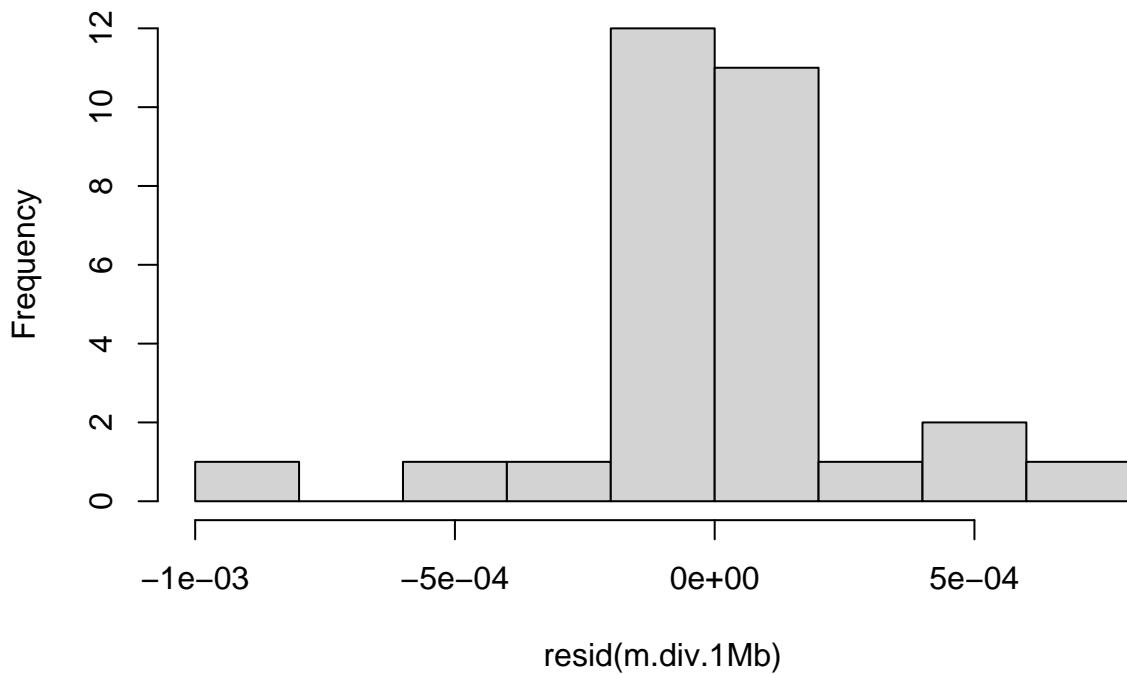
```

```

##
## Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.50343, p-value = 0.547
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
## -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       6.380e-03 6.210e-05 102.725 < 2e-16 ***
## rhoC        -5.882e-05 6.055e-05 -0.971  0.34066
## tmrcaC       4.478e-04 5.987e-05   7.479 7.84e-08 ***
## thetaC:tmrcaC 1.834e-04 5.024e-05   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])
```

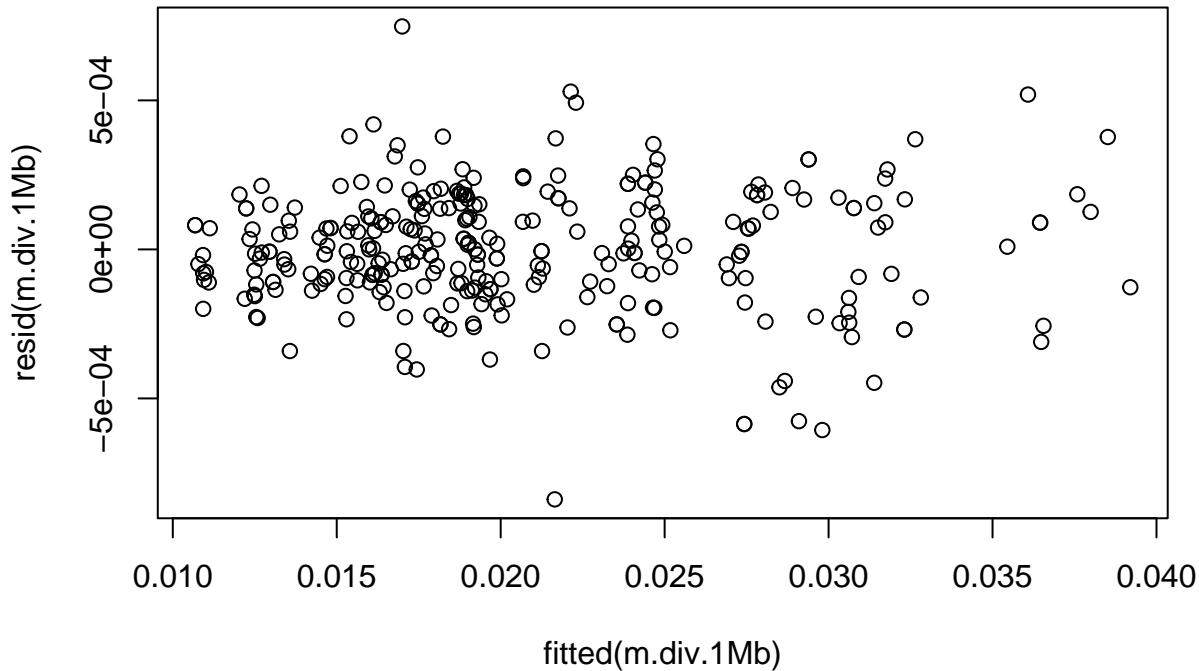
```
r2.sim.1Mb[2, 10] <- anova.diversity$VarExp[1] * 100
r2.sim.1Mb[3, 10] <- anova.diversity$VarExp[2] * 100
r2.sim.1Mb[4, 10] <- anova.diversity$VarExp[3] * 100
r2.sim.1Mb[5, 10] <- anova.diversity$VarExp[4] * 100
```

```
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.rep5, 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))
```



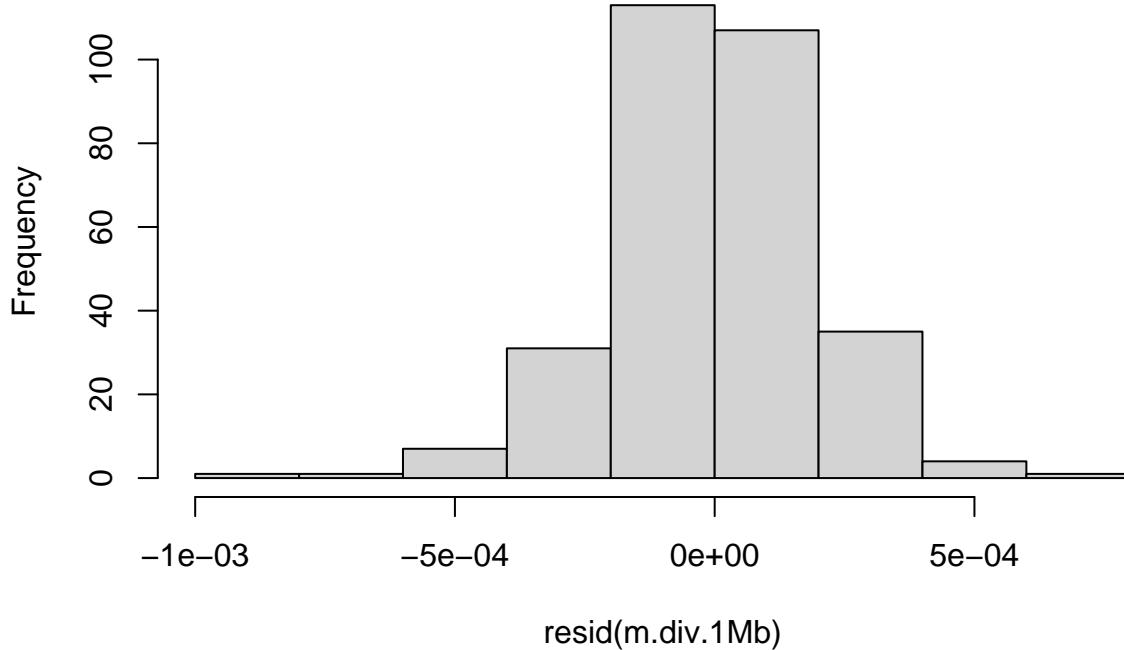
```
dwtest(m.div.1Mb)
```

```
##
##  Durbin-Watson test
##
##  data:  m.div.1Mb
##  DW = 2.2764, p-value = 0.9657
##  alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.1Mb)
```

```
##
##  Harrison-McCabe test
##
##  data:  m.div.1Mb
##  HMC = 0.43169, p-value = 0.044
```

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

Histogram of 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.factor(Replicate)
```

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

```
##          df      AIC
## m.div.1Mb 46 -4169.467
## m.div.1Mb.2 55 -4178.038
## m.div.1Mb.3 64 -4167.214
```

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

```
##
## Call:
## lm(formula = diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC) *
##     as.factor(Replicate), data = sim.lands.1Mb.all)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.386e-04 -1.136e-04 -6.930e-06  1.376e-04  7.480e-04
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.052e-02 3.986e-05 514.704 < 2e-16 ***
## thetaC      6.270e-03 4.277e-05 146.613 < 2e-16 ***
## rhoC       -2.992e-05 4.197e-05 -0.713  0.47663
## tmrcaC      4.395e-04 4.324e-05 10.164 < 2e-16 ***
## as.factor(Replicate)3 1.876e-04 5.876e-05  3.192  0.00159 **
```

```

## as.factor(Replicate)4      -2.640e-05  5.625e-05 -0.469  0.63918
## as.factor(Replicate)5      8.122e-05  4.867e-05  1.669  0.09640 .
## as.factor(Replicate)6      3.170e-05  5.701e-05  0.556  0.57872
## as.factor(Replicate)7      8.047e-05  5.674e-05  1.418  0.15738
## as.factor(Replicate)8      -1.073e-05 5.639e-05 -0.190  0.84927
## as.factor(Replicate)9      1.217e-04  5.743e-05  2.119  0.03508 *
## as.factor(Replicate)10     1.433e-04  5.641e-05  2.540  0.01167 *
## thetaC:tmrcaC              2.032e-04  4.019e-05  5.057  8.14e-07 ***
## thetaC:as.factor(Replicate)3 1.992e-04  6.219e-05  3.203  0.00153 **
## thetaC:as.factor(Replicate)4 2.495e-05  6.062e-05  0.412  0.68094
## thetaC:as.factor(Replicate)5 1.011e-04  5.239e-05  1.929  0.05483 .
## thetaC:as.factor(Replicate)6 9.521e-05  6.022e-05  1.581  0.11512
## thetaC:as.factor(Replicate)7 1.800e-04  6.159e-05  2.924  0.00377 **
## thetaC:as.factor(Replicate)8 4.412e-05  5.980e-05  0.738  0.46136
## thetaC:as.factor(Replicate)9 1.491e-04  6.575e-05  2.268  0.02419 *
## thetaC:as.factor(Replicate)10 1.096e-04  6.084e-05  1.801  0.07285 .
## rhoC:as.factor(Replicate)3   7.015e-05  5.932e-05  1.183  0.23810
## rhoC:as.factor(Replicate)4   3.644e-05  5.960e-05  0.611  0.54146
## rhoC:as.factor(Replicate)5   7.331e-06  5.113e-05  0.143  0.88611
## rhoC:as.factor(Replicate)6   7.486e-05  5.940e-05  1.260  0.20868
## rhoC:as.factor(Replicate)7   -2.634e-06 5.896e-05 -0.045  0.96440
## rhoC:as.factor(Replicate)8   4.484e-05  5.875e-05  0.763  0.44600
## rhoC:as.factor(Replicate)9   5.274e-05  6.010e-05  0.877  0.38106
## rhoC:as.factor(Replicate)10  -2.890e-05 5.951e-05 -0.486  0.62761
## tmrcaC:as.factor(Replicate)3 1.168e-04  6.251e-05  1.868  0.06287 .
## tmrcaC:as.factor(Replicate)4 1.801e-05  6.126e-05  0.294  0.76898
## tmrcaC:as.factor(Replicate)5 -2.001e-05 5.175e-05 -0.387  0.69938
## tmrcaC:as.factor(Replicate)6 -8.515e-05 6.042e-05 -1.409  0.15997
## tmrcaC:as.factor(Replicate)7 1.085e-04  5.998e-05  1.809  0.07170 .
## tmrcaC:as.factor(Replicate)8 -1.682e-06 5.929e-05 -0.028  0.97739
## tmrcaC:as.factor(Replicate)9 4.445e-05  6.132e-05  0.725  0.46914
## tmrcaC:as.factor(Replicate)10 8.298e-06  6.008e-05  0.138  0.89026
## thetaC:tmrcaC:as.factor(Replicate)3 2.590e-05  6.870e-05  0.377  0.70647
## thetaC:tmrcaC:as.factor(Replicate)4 -1.178e-05 5.138e-05 -0.229  0.81880
## thetaC:tmrcaC:as.factor(Replicate)5 -1.109e-04 5.169e-05 -2.145  0.03288 *
## thetaC:tmrcaC:as.factor(Replicate)6 -6.411e-05 6.444e-05 -0.995  0.32074
## thetaC:tmrcaC:as.factor(Replicate)7 -6.217e-05 5.419e-05 -1.147  0.25238
## thetaC:tmrcaC:as.factor(Replicate)8 -4.256e-05 6.319e-05 -0.674  0.50123
## thetaC:tmrcaC:as.factor(Replicate)9 -8.544e-05 6.002e-05 -1.423  0.15584
## thetaC:tmrcaC:as.factor(Replicate)10 -1.980e-05 5.330e-05 -0.372  0.71055
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.000216 on 255 degrees of freedom
## Multiple R-squared:  0.999, Adjusted R-squared:  0.9988
## F-statistic:  5772 on 44 and 255 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 are 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

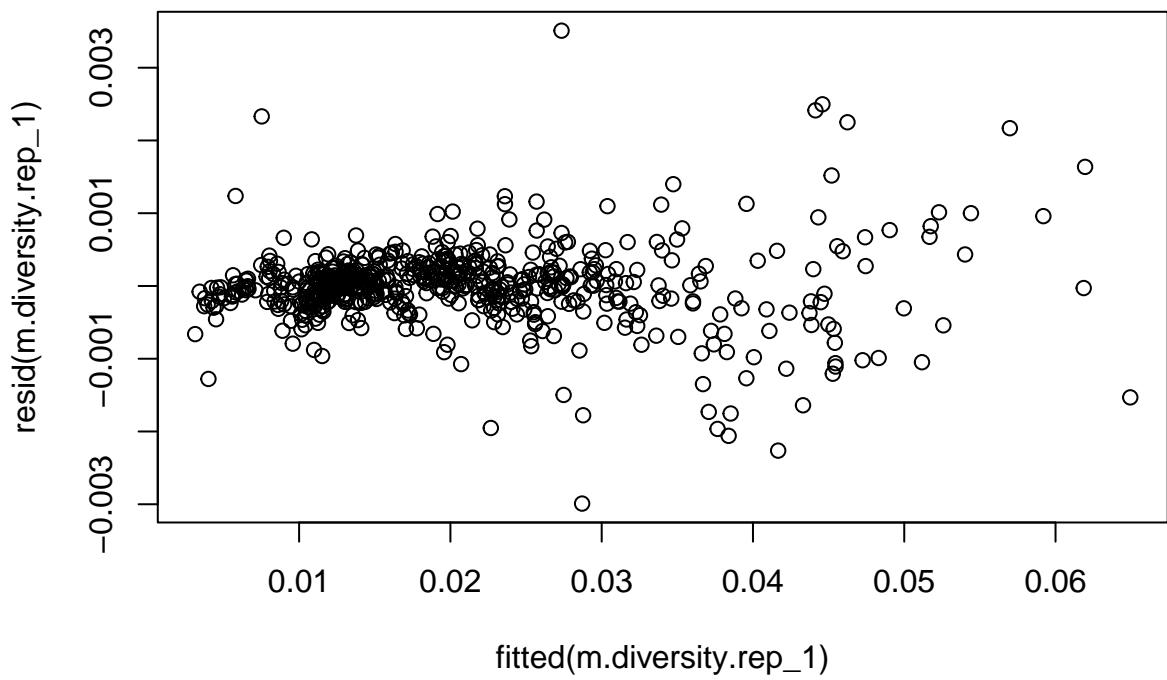
# standardizing
inf.lands.50kb.rep_1$thetaC <- (inf.lands.50kb.rep_1$theta - mean(inf.lands.50kb.rep_1$theta)) / sd(inf.lands.50kb.rep_1$theta)
inf.lands.50kb.rep_1$tmrcaC <- (inf.lands.50kb.rep_1$tmrca - mean(inf.lands.50kb.rep_1$tmrca)) / sd(inf.lands.50kb.rep_1$tmrca)
inf.lands.50kb.rep_1$rhoC <-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

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

inf.lands.50kb.rep_1$Replicate <- 1

m.diversity.rep_1 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC*tmrcaC, data = inf.lands.50kb.rep_1)

plot(resid(m.diversity.rep_1)~fitted(m.diversity.rep_1))
```

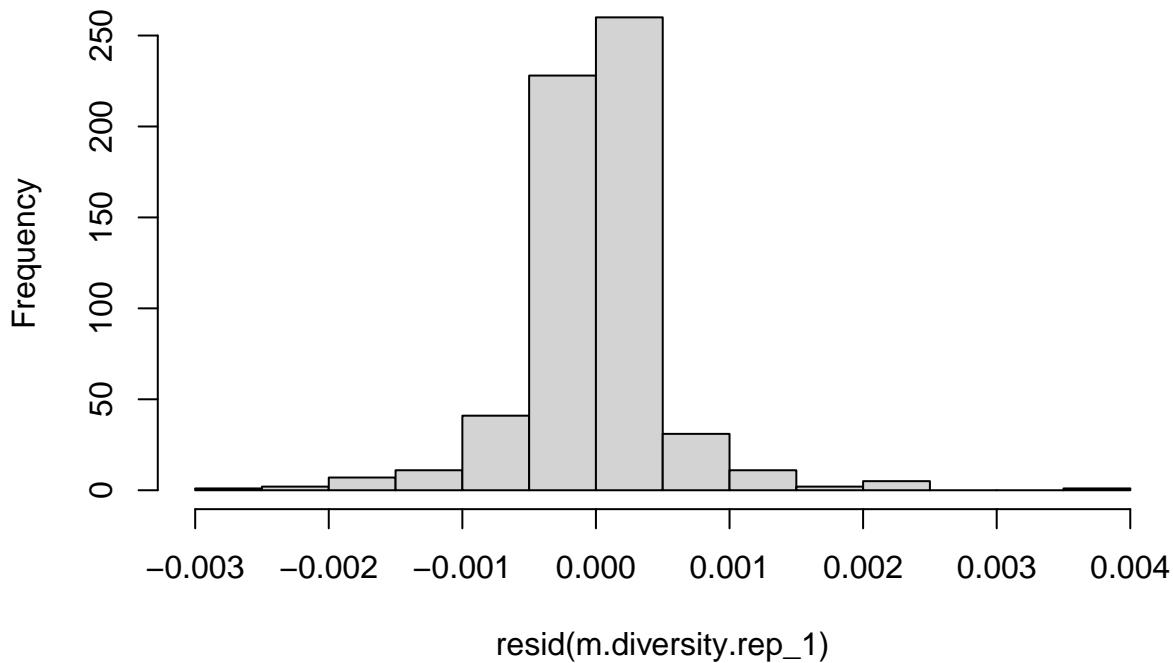


```
dwtest(m.diversity.rep_1)
```

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

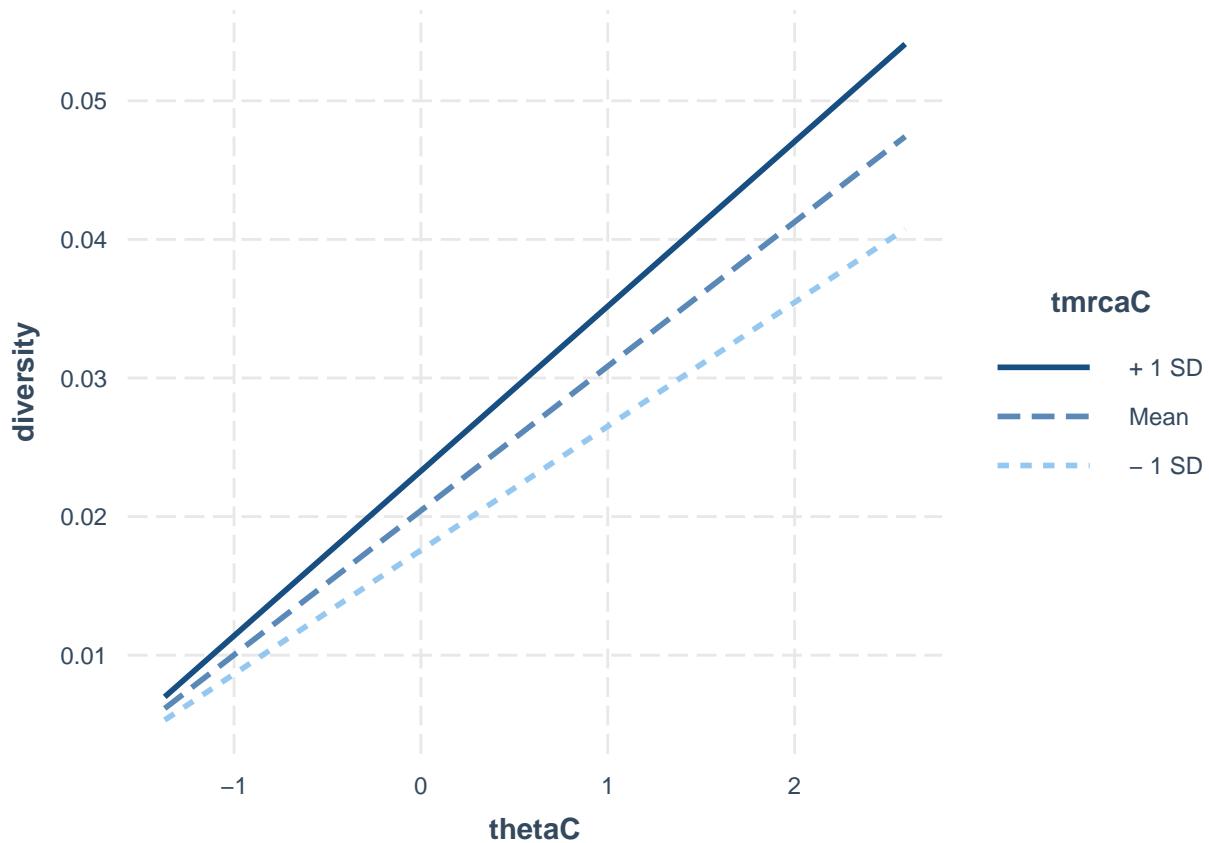
```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_1
## HMC = 0.61568, p-value = 1
hist(resid(m.diversity.rep_1))
```

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      1.041e-02 2.246e-05 463.564 <2e-16 ***
## rhoC       -5.658e-06 2.228e-05 -0.254     0.8
## tmrcaC      2.854e-03 2.312e-05 123.483 <2e-16 ***
## thetaC:tmrcaC 1.473e-03 2.126e-05  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) 0.020438810 3.125509e-05 653.9355 0.0000
## thetaC      0.010435258 2.837609e-05 367.7482 0.0000
## tmrcaC      0.002817637 2.481844e-05 113.5300 0.0000
## rhoC        0.000004491 2.161341e-05    0.2078 0.8355
## thetaC:tmrcaC 0.001426645 2.293513e-05   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) 0.020632135 0.0002124018 97.13730 0.0000
## thetaC       0.011230411 0.0001619026 69.36522 0.0000
## rhoC        0.000007276 0.0001032089  0.07050 0.9438
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.001  0.003
##
## Standardized residuals:
##      Min       Q1       Med       Q3       Max
## -4.88116216 -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$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCAs.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$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")

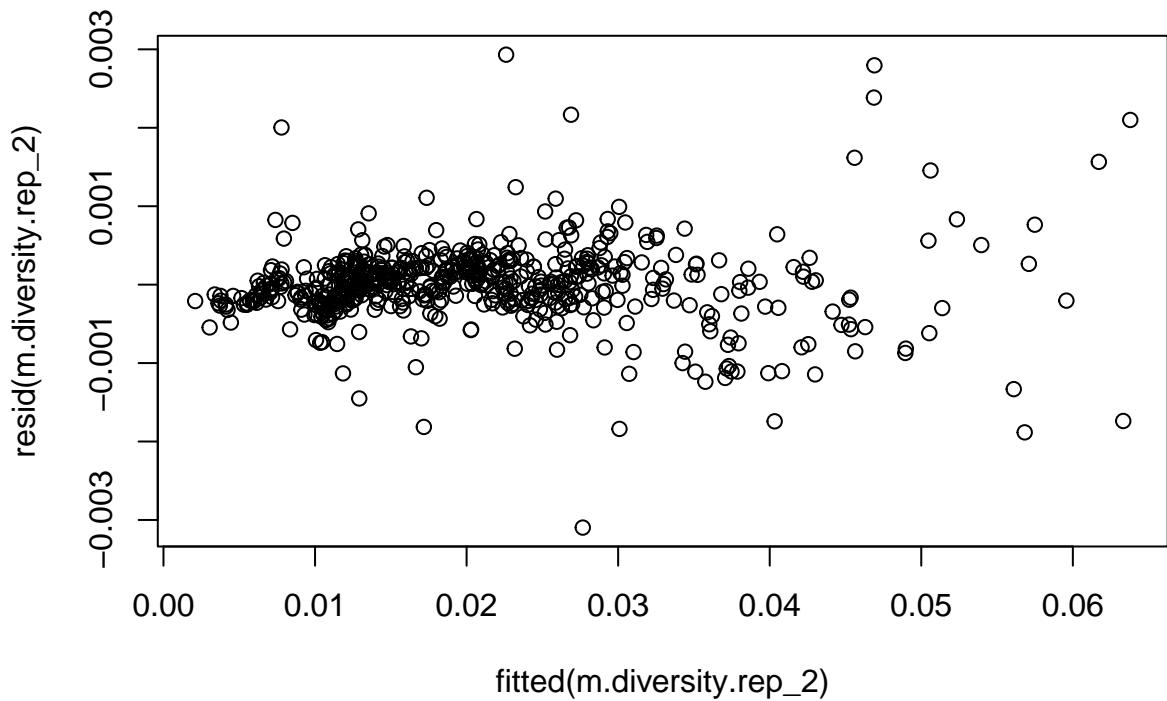
# standardizing
inf.lands.50kb.rep_2$thetaC <- (inf.lands.50kb.rep_2$theta - mean(inf.lands.50kb.rep_2$theta)) / sd(inf.lands.50kb.rep_2$theta)
inf.lands.50kb.rep_2$tmrcaC <- (inf.lands.50kb.rep_2$tmrca - mean(inf.lands.50kb.rep_2$tmrca)) / sd(inf.lands.50kb.rep_2$tmrca)
inf.lands.50kb.rep_2$rhoC <- (inf.lands.50kb.rep_2$rho - mean(inf.lands.50kb.rep_2$rho)) / sd(inf.lands.50kb.rep_2$rho)

inf.lands.50kb.rep_2$bin <- 1:nrow(inf.lands.50kb.rep_2)

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))
```

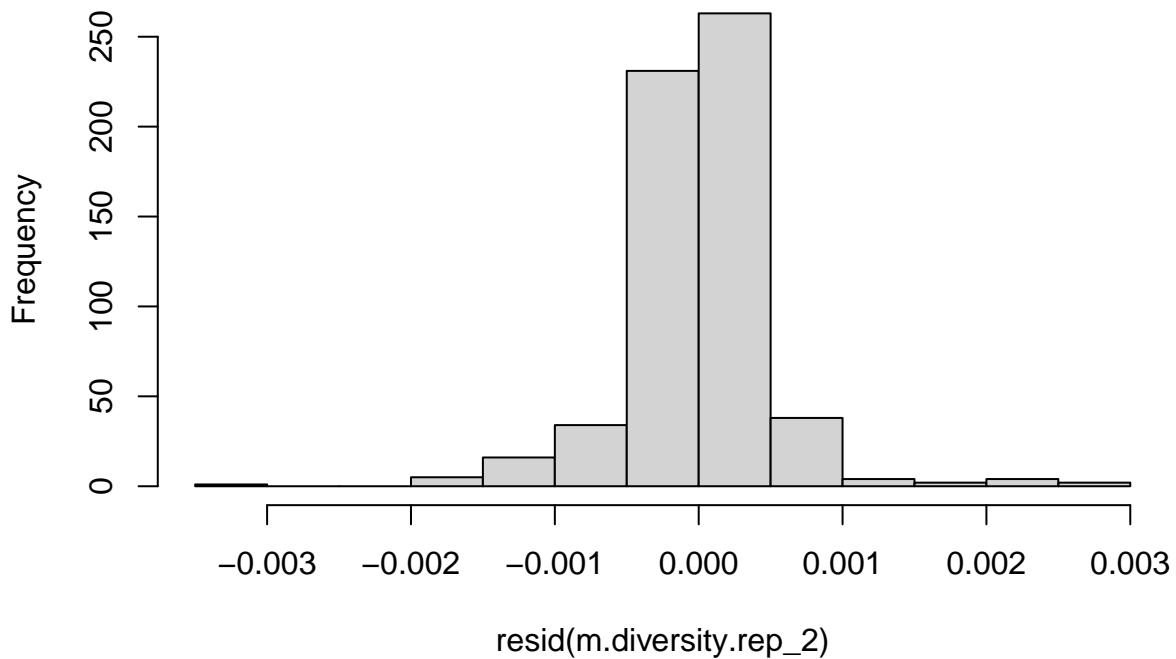


```
dwtest(m.diversity.rep_2)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_2
## DW = 1.3609, p-value = 1.173e-15
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_2)
```

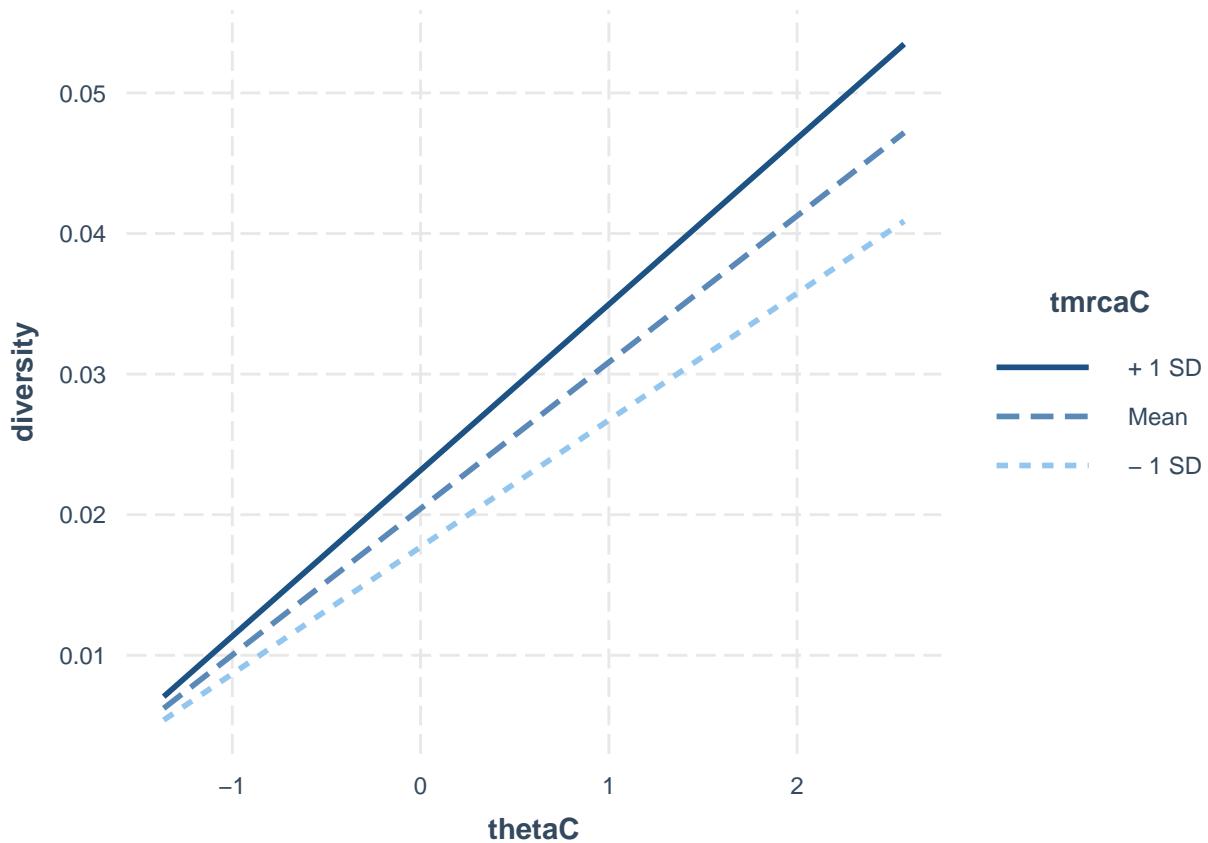
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_2
## HMC = 0.61478, p-value = 1
hist(resid(m.diversity.rep_2))
```

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       1.041e-02 2.075e-05 501.668 <2e-16 ***
## rhoC        -1.464e-05 2.069e-05 -0.708   0.479    
## tmrcaC       2.728e-03 2.141e-05 127.438 <2e-16 ***
## thetaC:tmrcaC 1.389e-03 1.785e-05 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) 0.020437731 3.069727e-05 665.7833 0.0000
## thetaC       0.010425586 2.682830e-05 388.6041 0.0000
## tmrcaC       0.002686625 2.228409e-05 120.5625 0.0000
## rhoC        -0.000021005 1.943549e-05 -1.0807 0.2802
## thetaC:tmrcaC 0.001330244 1.868961e-05  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
## -5.967387961 -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) 0.020597993 0.0001669114 123.40674 0.0000
## thetaC       0.010803037 0.0001480341  72.97670 0.0000
## rhoC        -0.000092475 0.0001110053  -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  6.220744947
##
## 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$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCAs.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$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")

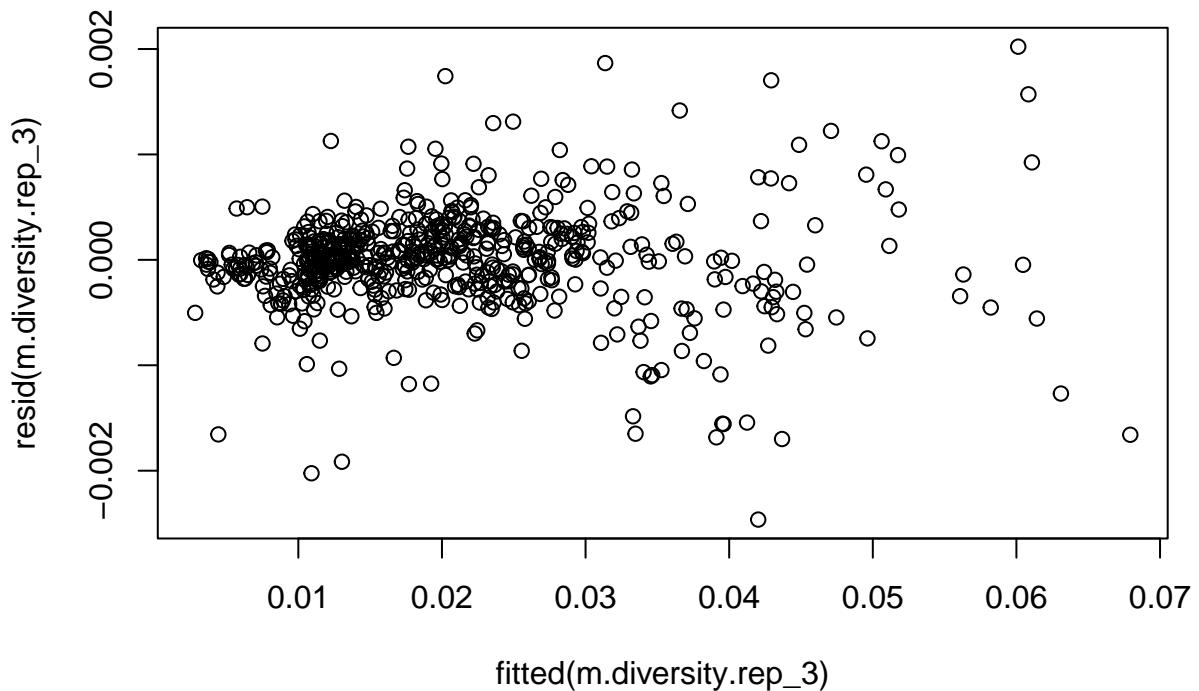
# standardizing
inf.lands.50kb.rep_3$thetaC <- (inf.lands.50kb.rep_3$theta - mean(inf.lands.50kb.rep_3$theta)) / sd(inf.lands.50kb.rep_3$theta)
inf.lands.50kb.rep_3$tmrcaC <- (inf.lands.50kb.rep_3$tmrca - mean(inf.lands.50kb.rep_3$tmrca)) / sd(inf.lands.50kb.rep_3$tmrca)
inf.lands.50kb.rep_3$rhoC <- (inf.lands.50kb.rep_3$rho - mean(inf.lands.50kb.rep_3$rho)) / sd(inf.lands.50kb.rep_3$rho)

inf.lands.50kb.rep_3$bin <- 1:nrow(inf.lands.50kb.rep_3)

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))
```

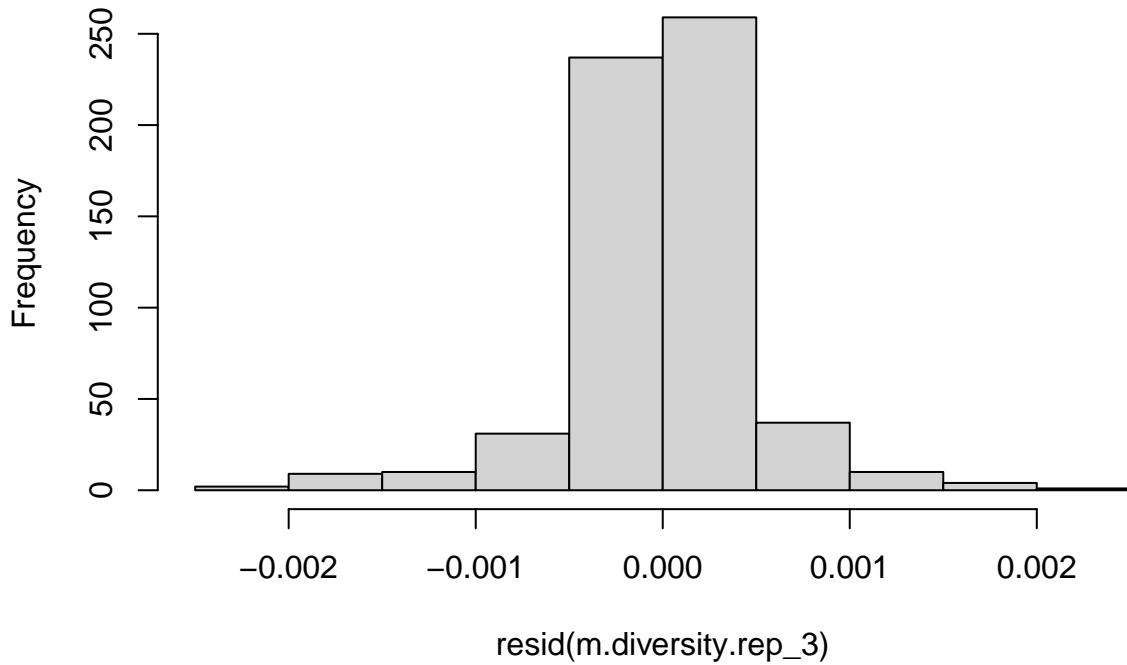


```
dwtest(m.diversity.rep_3)
```

```
##
## Durbin-Watson test
##
## data: m.diversity.rep_3
## DW = 1.4619, p-value = 1.133e-11
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_3)
```

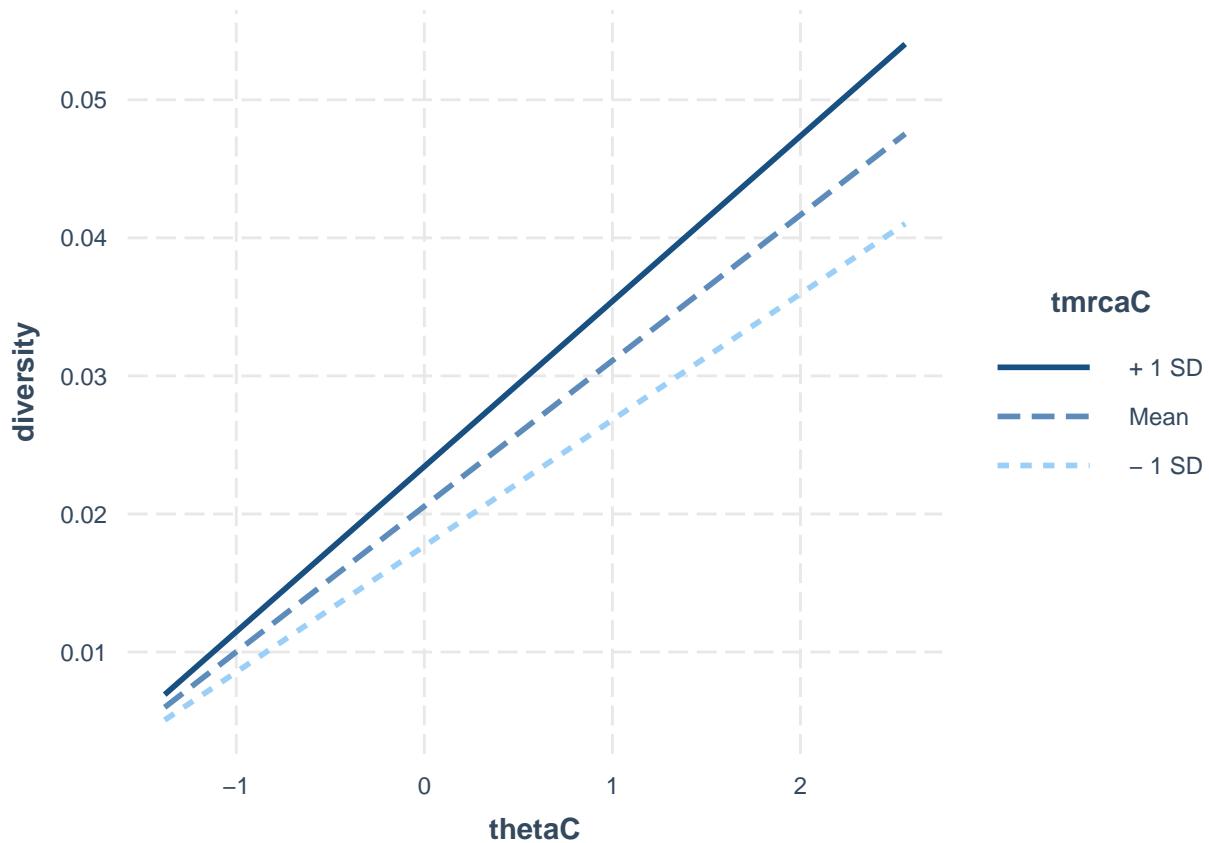
```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_3
## HMC = 0.52153, p-value = 0.773
hist(resid(m.diversity.rep_3))
```

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) 2.056e-02 1.980e-05 1038.471 <2e-16 ***
## thetaC       1.055e-02 1.991e-05  529.902 <2e-16 ***
## rhoC        -1.745e-05 1.976e-05  -0.883  0.378    
## tmrcaC       2.878e-03 2.002e-05  143.768 <2e-16 ***
## thetaC:tmrcaC 1.417e-03 1.685e-05   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) 0.020562880 2.650697e-05 775.7537 0.0000
## thetaC      0.010560835 2.454950e-05 430.1853 0.0000
## tmrcaC      0.002861485 2.145250e-05 133.3870 0.0000
## rhoC        -0.000022393 1.925540e-05 -1.1629 0.2453
## thetaC:tmrcaC 0.001390888 1.848554e-05  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) 0.020741983 0.0002138359 96.99954 0.0000
## thetaC       0.011165930 0.0001731923 64.47130 0.0000
## rhoC        -0.000127852 0.0001131819 -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$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCAs.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$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")

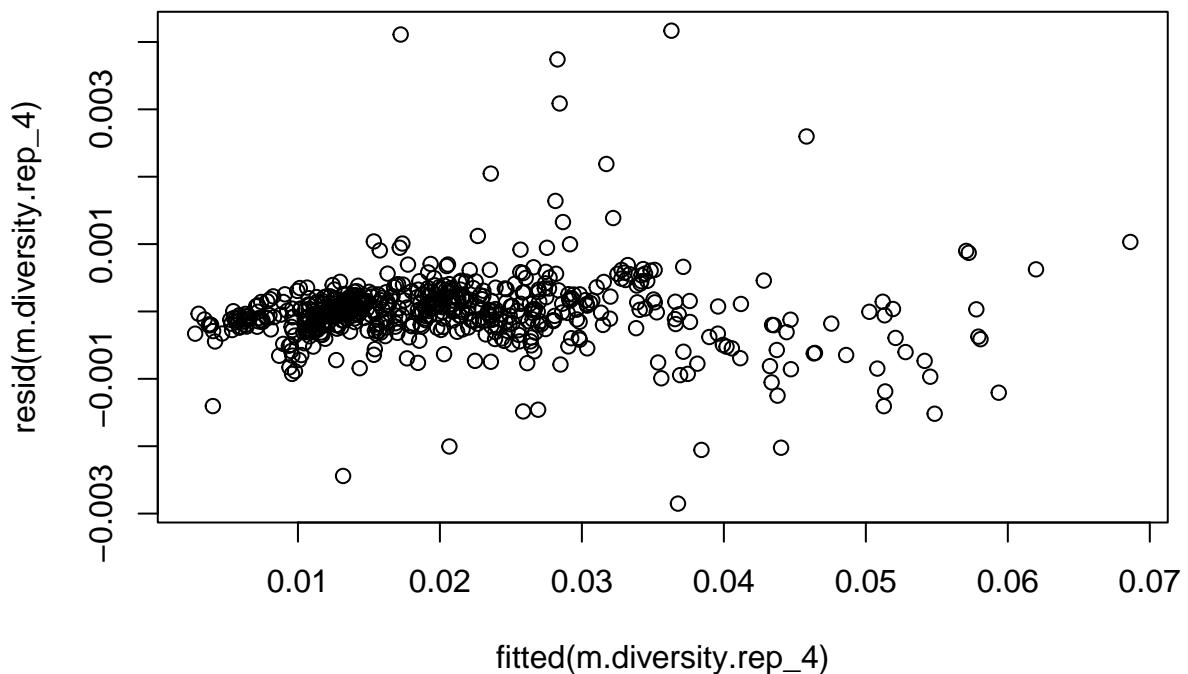
# standardizing
inf.lands.50kb.rep_4$thetaC <- (inf.lands.50kb.rep_4$theta - mean(inf.lands.50kb.rep_4$theta)) / sd(inf.lands.50kb.rep_4$theta)
inf.lands.50kb.rep_4$tmrcaC <- (inf.lands.50kb.rep_4$tmrca - mean(inf.lands.50kb.rep_4$tmrca)) / sd(inf.lands.50kb.rep_4$tmrca)
inf.lands.50kb.rep_4$rhoC <- (inf.lands.50kb.rep_4$rho - mean(inf.lands.50kb.rep_4$rho)) / sd(inf.lands.50kb.rep_4$rho)

inf.lands.50kb.rep_4$bin <- 1:nrow(inf.lands.50kb.rep_4)

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))
```

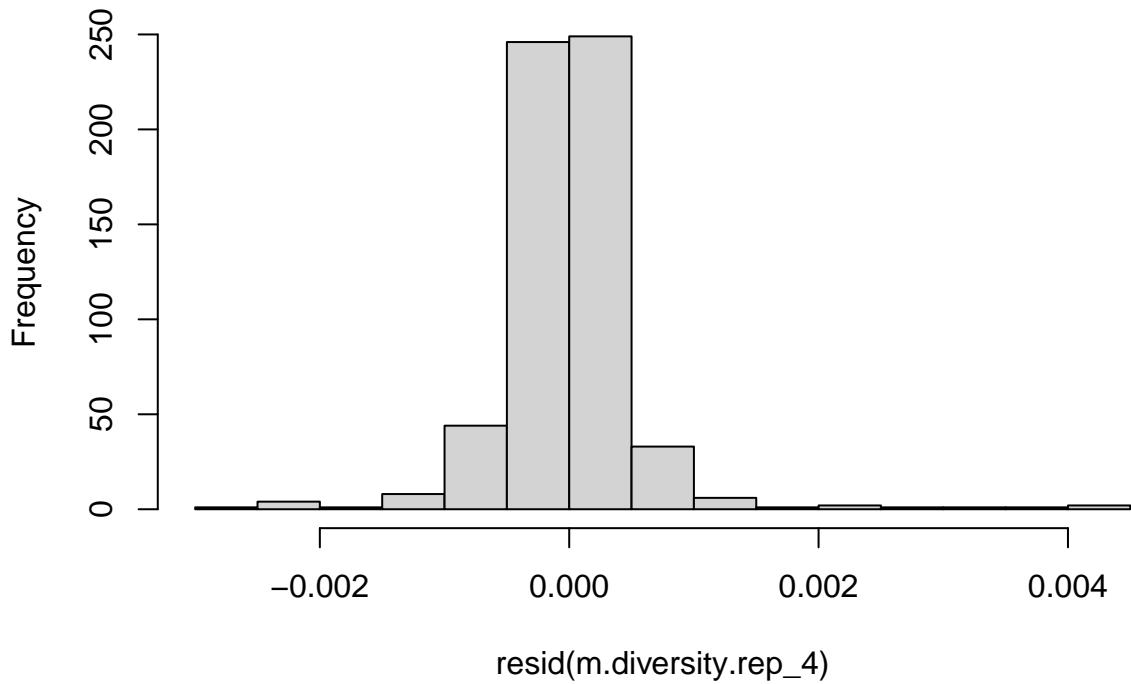


```
dwtest(m.diversity.rep_4)
```

```
##
## Durbin-Watson test
##
## data: m.diversity.rep_4
## DW = 0.96213, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_4)
```

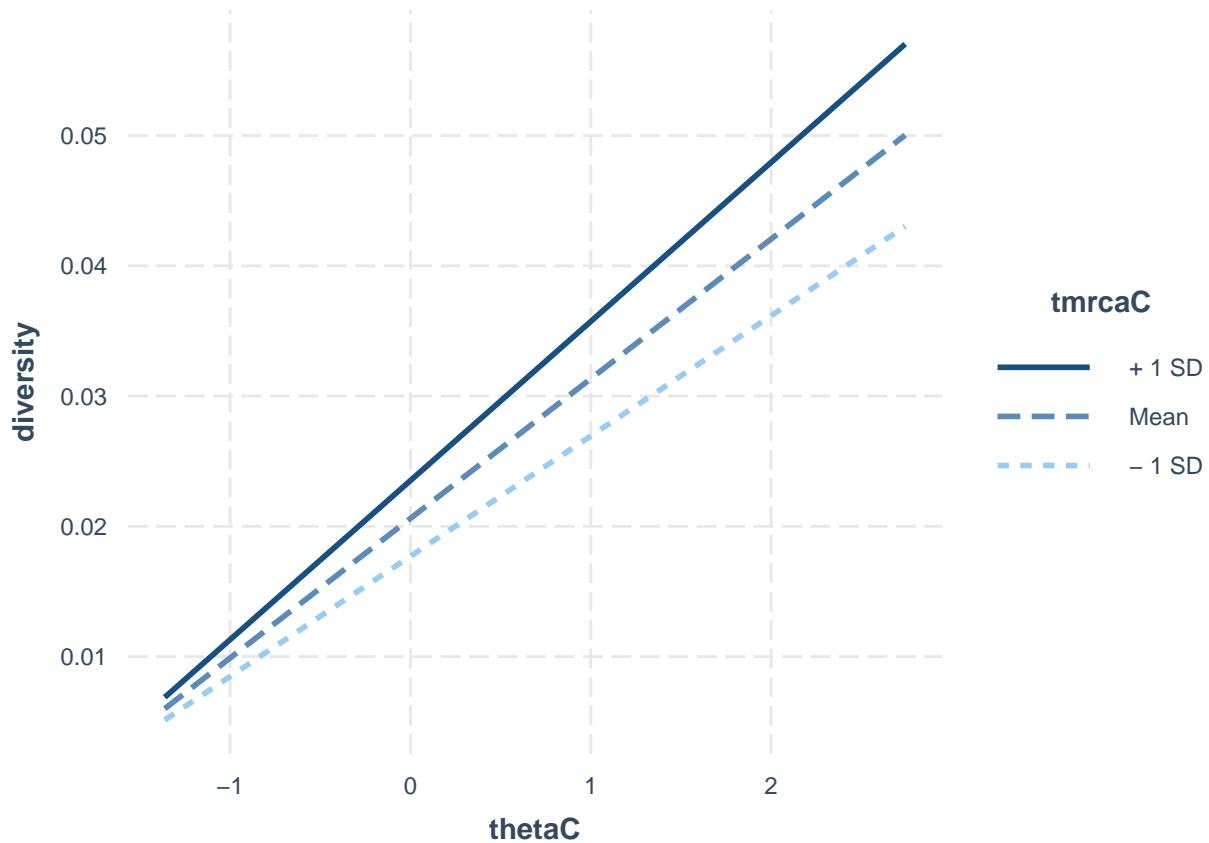
```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_4
## HMC = 0.70321, p-value = 1
hist(resid(m.diversity.rep_4))
```

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      1.072e-02 2.308e-05 464.707 <2e-16 ***
## rhoC       -1.398e-05 2.319e-05 -0.603   0.547
## tmrcaC      2.905e-03 2.336e-05 124.378 <2e-16 ***
## thetaC:tmrcaC 1.492e-03 2.120e-05 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) 0.020615250 4.251583e-05 484.8841 0.0000
## thetaC       0.010711538 3.118823e-05 343.4480 0.0000
## tmrcaC       0.002895620 2.330073e-05 124.2717 0.0000
## rhoC        -0.000009215 1.893368e-05  -0.4867 0.6266
## thetaC:tmrcaC 0.001424462 2.028301e-05  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
## -5.20520626 -0.37375535 -0.02131367  0.35668752  7.22457563
##
## 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) 0.02068915 0.0002181869 94.82308 0.0000
## thetaC       0.01105774 0.0001742197 63.47010 0.0000
## rhoC        -0.00021567 0.0001141970 -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$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCAs.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$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")

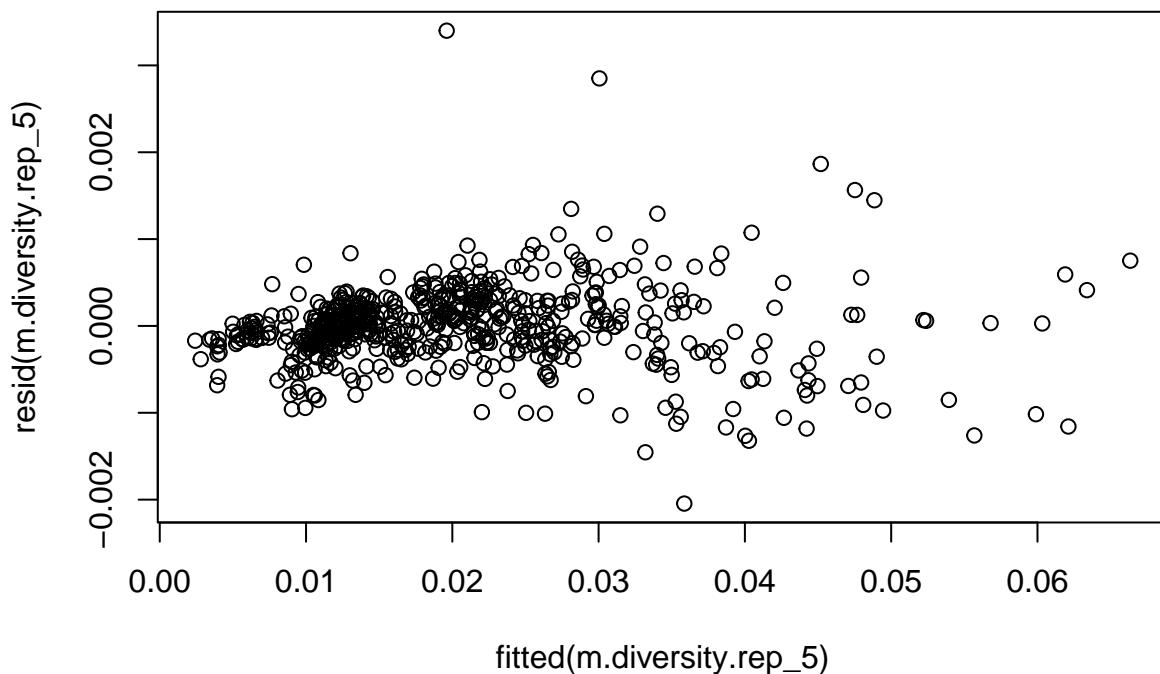
# standardizing
inf.lands.50kb.rep_5$thetaC <- (inf.lands.50kb.rep_5$theta - mean(inf.lands.50kb.rep_5$theta)) / sd(inf.lands.50kb.rep_5$theta)
inf.lands.50kb.rep_5$tmrcaC <- (inf.lands.50kb.rep_5$tmrca - mean(inf.lands.50kb.rep_5$tmrca)) / sd(inf.lands.50kb.rep_5$tmrca)
inf.lands.50kb.rep_5$rhoC <- (inf.lands.50kb.rep_5$rho - mean(inf.lands.50kb.rep_5$rho)) / sd(inf.lands.50kb.rep_5$rho)

inf.lands.50kb.rep_5$bin <- 1:nrow(inf.lands.50kb.rep_5)

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))
```

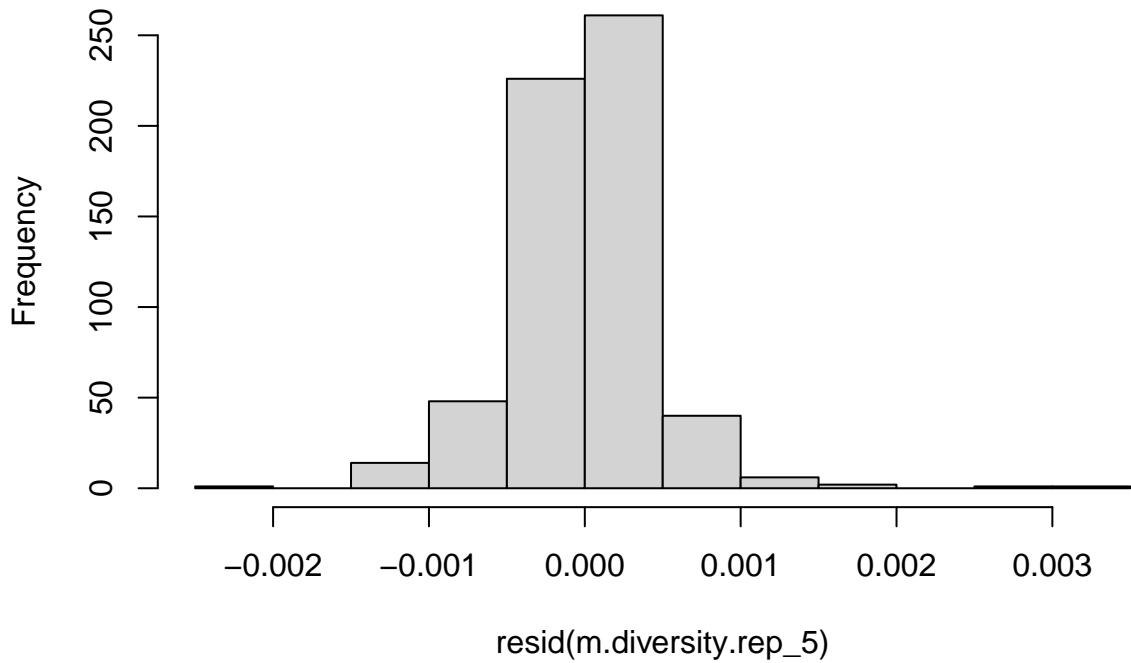


```
dwtest(m.diversity.rep_5)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_5
## DW = 1.4275, p-value = 5.886e-13
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_5)
```

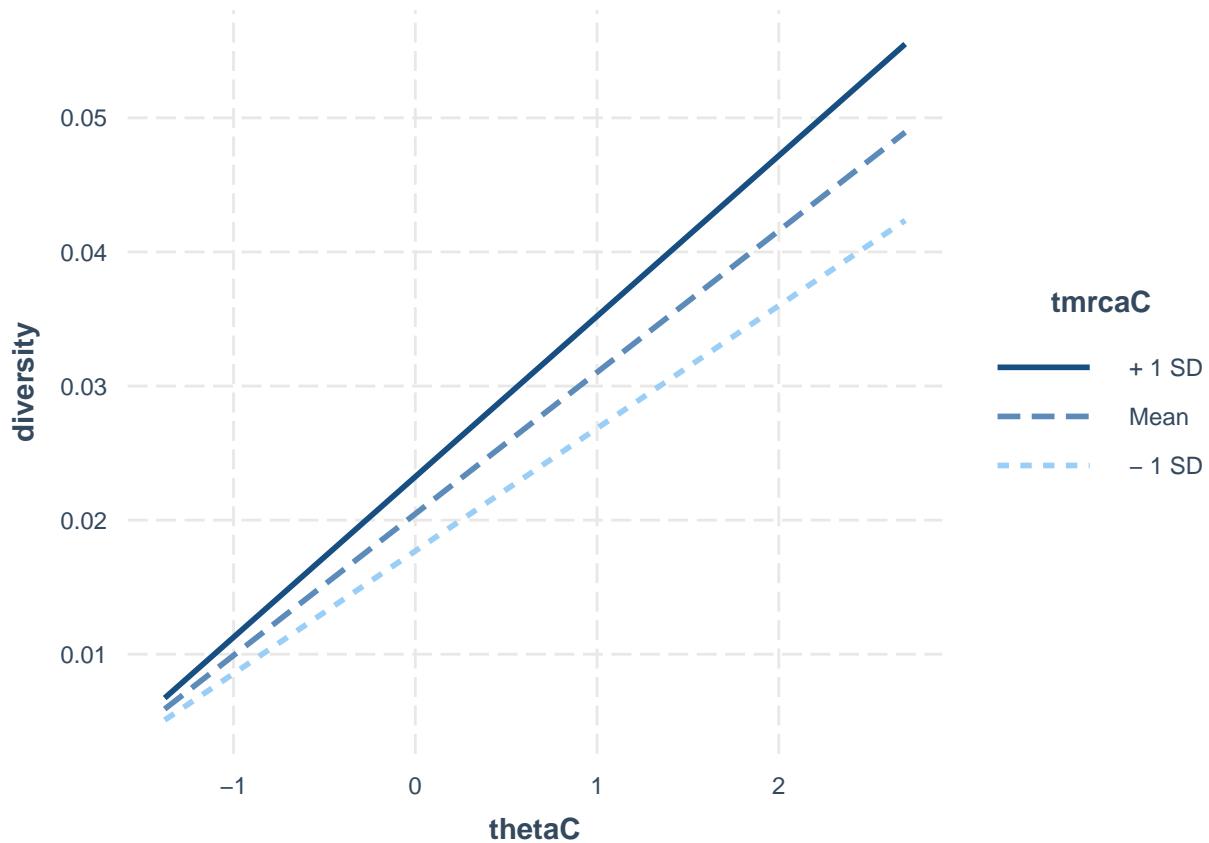
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_5
## HMC = 0.55418, p-value = 0.963
hist(resid(m.diversity.rep_5))
```

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) 2.047e-02 1.890e-05 1082.787 <2e-16 ***
## thetaC      1.056e-02 1.896e-05  556.934 <2e-16 ***
## rhoC       -8.403e-06 1.900e-05   -0.442  0.658
## tmrcaC      2.768e-03 1.918e-05   144.318 <2e-16 ***
## thetaC:tmrcaC 1.413e-03 1.579e-05    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) 0.020473277 2.826402e-05 724.3584 0.0000
## thetaC      0.010548590 2.480950e-05 425.1835 0.0000
## tmrcaC      0.002725546 2.067692e-05 131.8159 0.0000
## rhoC        -0.000014228 1.810647e-05  -0.7858 0.4323
## thetaC:tmrcaC 0.001345803 1.696976e-05  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
## -4.3488500237 -0.4439991689  0.0002670609  0.4775851247  7.1344633251
##
## 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) 0.020570880 0.0001928251 106.68153 0.0000
## thetaC       0.010803776 0.0001646389  65.62102 0.0000
## rhoC        -0.000188361 0.0001165489  -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$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCAs.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$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")

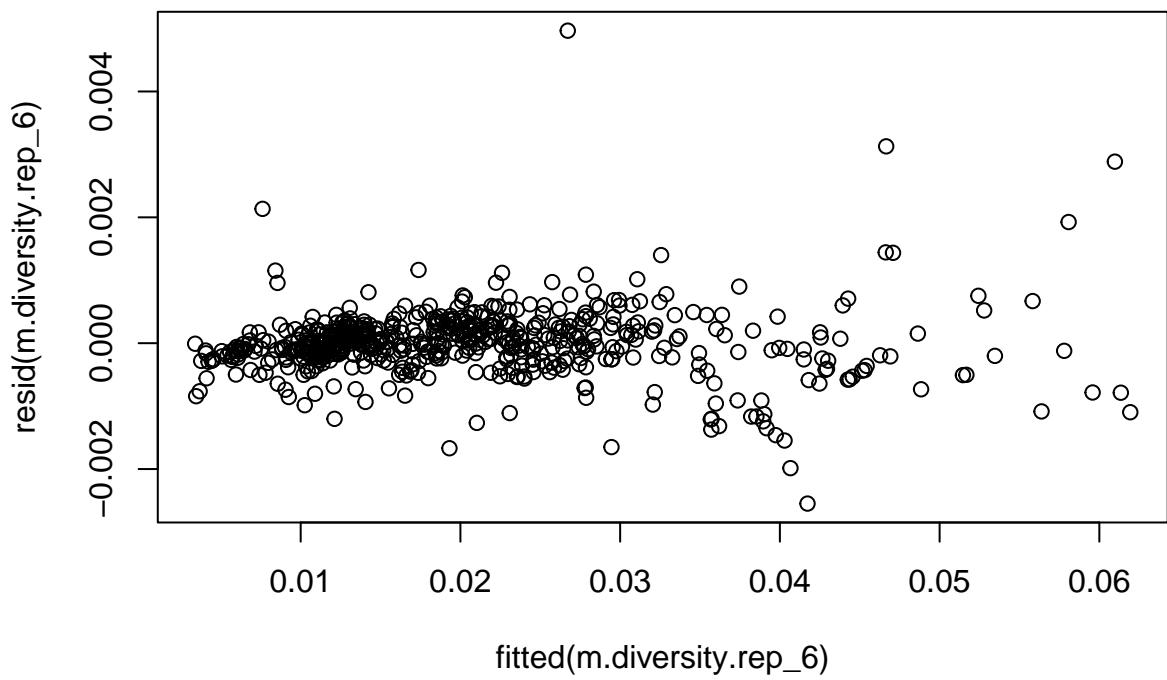
# standardizing
inf.lands.50kb.rep_6$thetaC <- (inf.lands.50kb.rep_6$theta - mean(inf.lands.50kb.rep_6$theta)) / sd(inf.lands.50kb.rep_6$theta)
inf.lands.50kb.rep_6$tmrcaC <- (inf.lands.50kb.rep_6$tmrca - mean(inf.lands.50kb.rep_6$tmrca)) / sd(inf.lands.50kb.rep_6$tmrca)
inf.lands.50kb.rep_6$rhoC <- (inf.lands.50kb.rep_6$rho - mean(inf.lands.50kb.rep_6$rho)) / sd(inf.lands.50kb.rep_6$rho)

inf.lands.50kb.rep_6$bin <- 1:nrow(inf.lands.50kb.rep_6)

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))
```

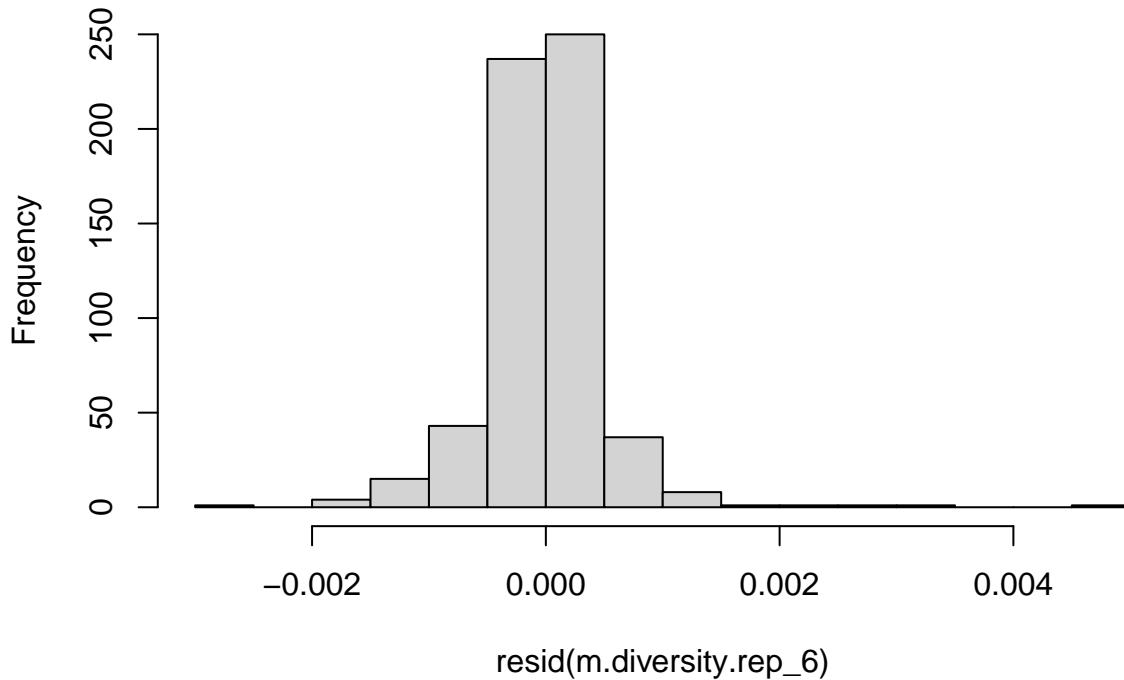


```
dwtest(m.diversity.rep_6)
```

```
##
## Durbin-Watson test
##
## data: m.diversity.rep_6
## DW = 1.4152, p-value = 1.956e-13
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_6)
```

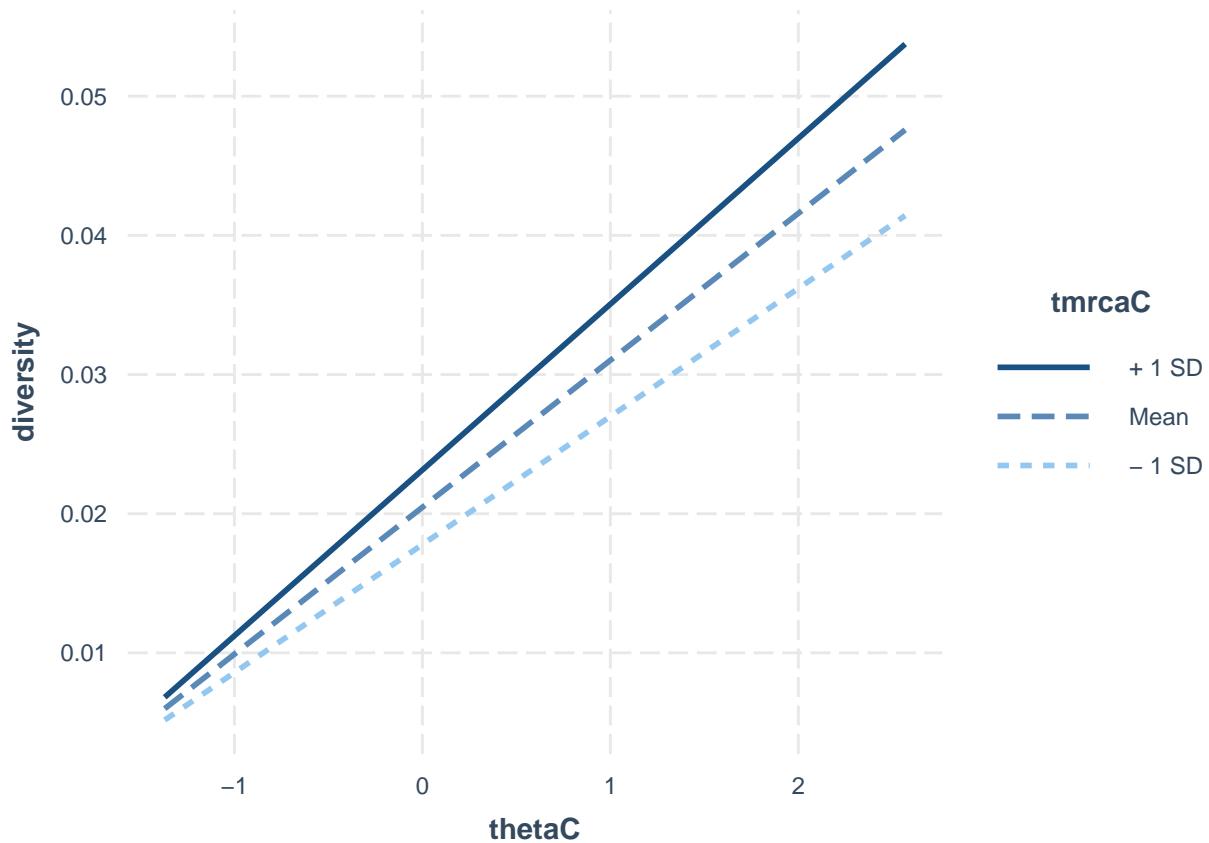
```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_6
## HMC = 0.5501, p-value = 0.964
hist(resid(m.diversity.rep_6))
```

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      1.056e-02 2.187e-05 482.709 <2e-16 ***  
## rhoC       1.225e-05 2.206e-05   0.555   0.579  
## tmrcaC     2.678e-03 2.246e-05 119.259 <2e-16 ***  
## thetaC:tmrcaC 1.356e-03 1.899e-05  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) 0.020468098 3.058636e-05 669.1903 0.0000
## thetaC      0.010559959 2.754228e-05 383.4091 0.0000
## tmrcaC     0.002648334 2.350889e-05 112.6524 0.0000
## rhoC        0.000000301 2.134923e-05   0.0141 0.9888
## thetaC:tmrcaC 0.001311352 2.055231e-05  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) 0.020584561 0.0001826143 112.72153 0.0000
## thetaC       0.010823438 0.0001541657  70.20654 0.0000
## rhoC        -0.000139262 0.0001093776  -1.27322 0.2034
##
## Correlation:
##      (Intr) thetaC
## thetaC 0.000
## rhoC   0.001  0.014
##
## Standardized residuals:
##      Min       Q1       Med       Q3       Max
## -4.246615320 -0.484784795 -0.004012756  0.502253352  5.520109922
##
## 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$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCAs.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$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")

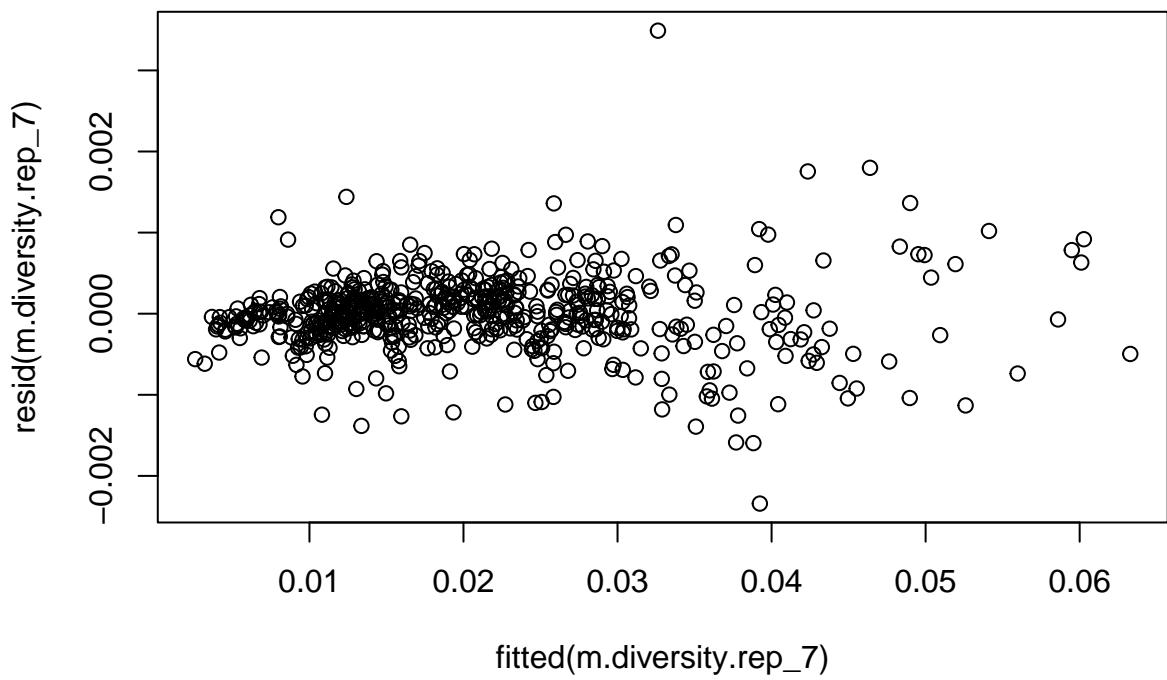
# standardizing
inf.lands.50kb.rep_7$thetaC <- (inf.lands.50kb.rep_7$theta - mean(inf.lands.50kb.rep_7$theta)) / sd(inf.lands.50kb.rep_7$theta)
inf.lands.50kb.rep_7$tmrcaC <- (inf.lands.50kb.rep_7$tmrca - mean(inf.lands.50kb.rep_7$tmrca)) / sd(inf.lands.50kb.rep_7$tmrca)
inf.lands.50kb.rep_7$rhoC <- (inf.lands.50kb.rep_7$rho - mean(inf.lands.50kb.rep_7$rho)) / sd(inf.lands.50kb.rep_7$rho)

inf.lands.50kb.rep_7$bin <- 1:nrow(inf.lands.50kb.rep_7)

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))
```

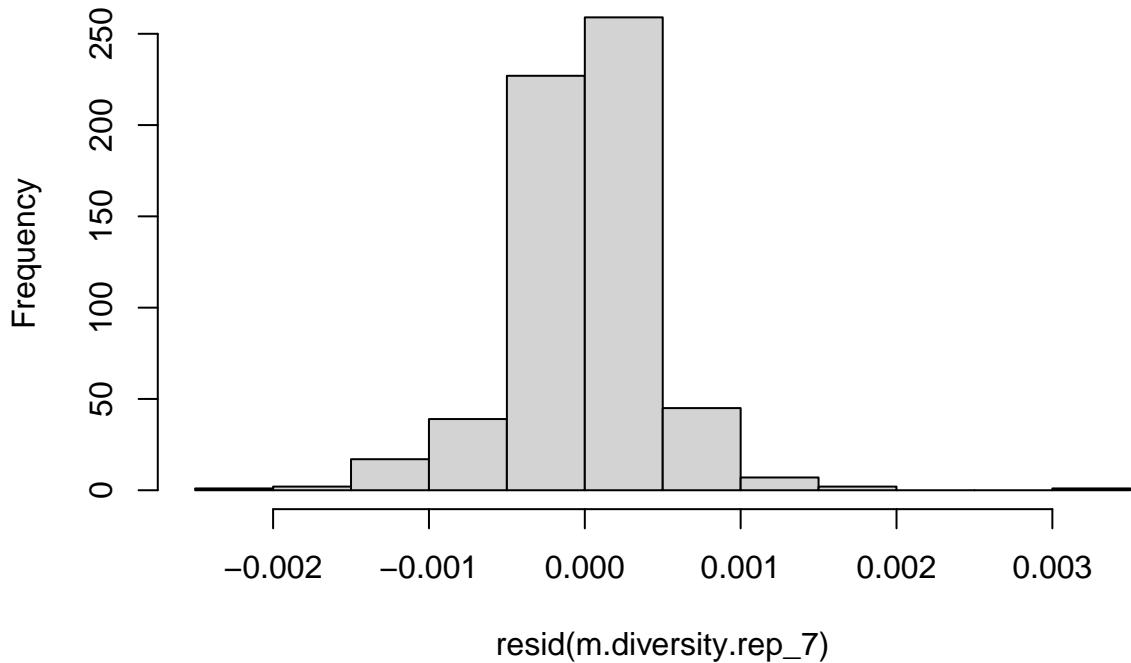


```
dwtest(m.diversity.rep_7)
```

```
##  
## Durbin-Watson test  
##  
## data: m.diversity.rep_7  
## DW = 1.2084, p-value < 2.2e-16  
## alternative hypothesis: true autocorrelation is greater than 0  
hmctest(m.diversity.rep_7)
```

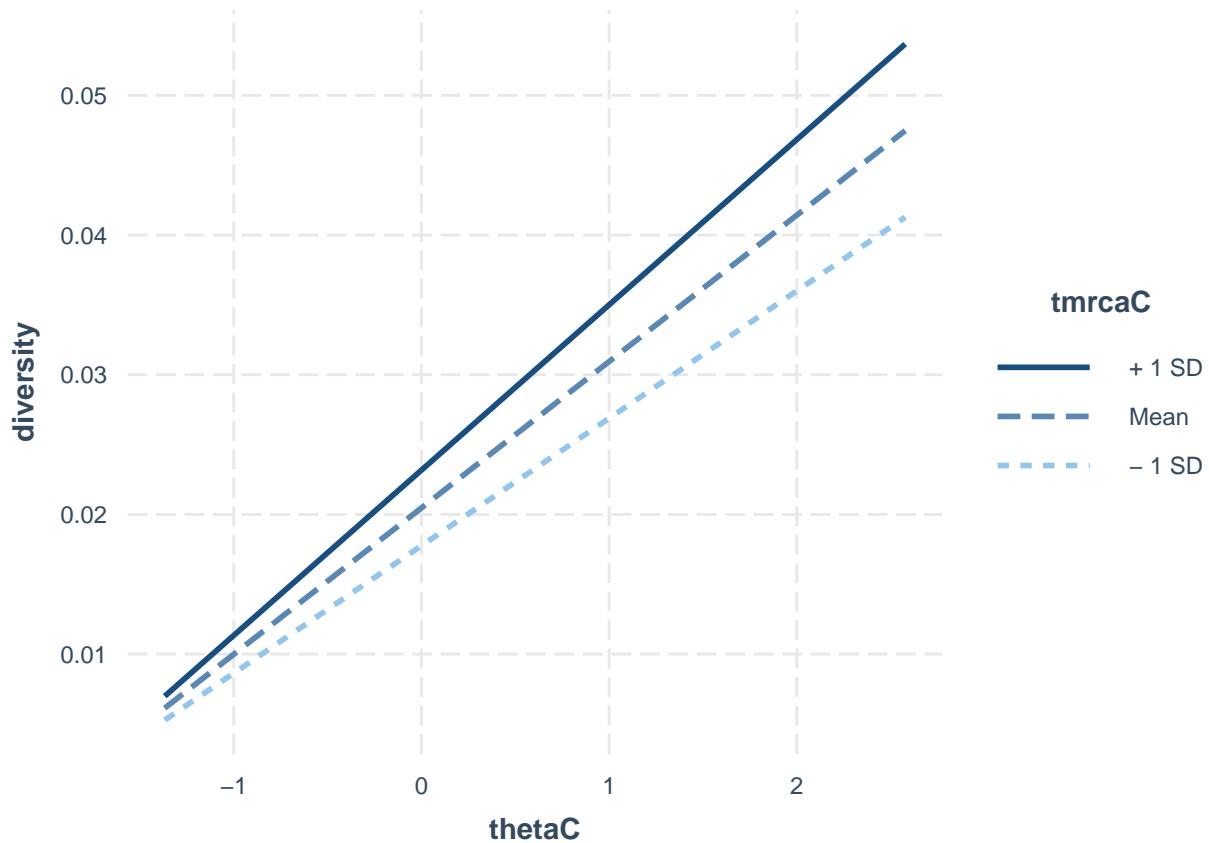
```
##  
## Harrison-McCabe test  
##  
## data: m.diversity.rep_7  
## HMC = 0.58586, p-value = 1  
hist(resid(m.diversity.rep_7))
```

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      1.048e-02 1.911e-05  548.323 < 2e-16 ***  
## rhoC       -5.001e-05 1.904e-05   -2.627 0.00885 **  
## tmrcaC      2.706e-03 1.957e-05   138.265 < 2e-16 ***  
## thetaC:tmrcaC 1.355e-03 1.725e-05    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) 0.020470105 3.231288e-05 633.4967 0.0000
## thetaC      0.010509441 2.607289e-05 403.0793 0.0000
## tmrcaC      0.002648119 1.999976e-05 132.4076 0.0000
## rhoC        -0.000038977 1.690711e-05  -2.3054 0.0215
## thetaC:tmrcaC 0.001278210 1.826157e-05  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) 0.020580105 0.0001934033 106.41030 0.0000
## thetaC       0.010815427 0.0001552921  69.64571 0.0000
## rhoC        0.000020888 0.0001011154   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$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCAs.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$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")

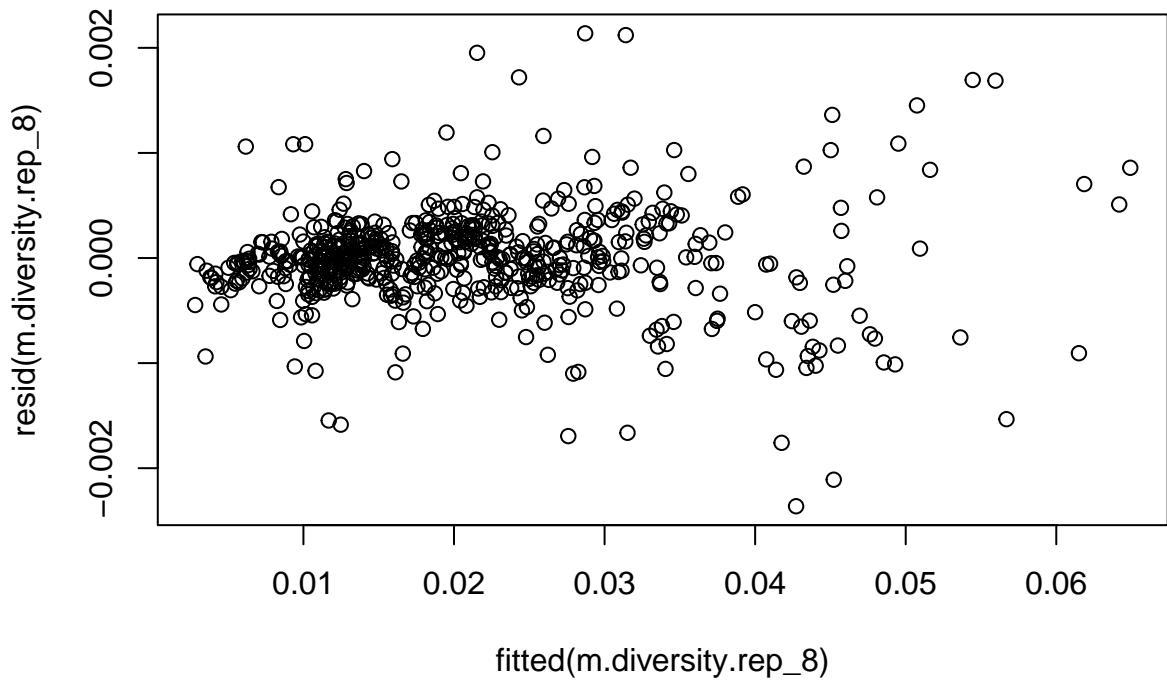
# standardizing
inf.lands.50kb.rep_8$thetaC <- (inf.lands.50kb.rep_8$theta - mean(inf.lands.50kb.rep_8$theta)) / sd(inf.lands.50kb.rep_8$theta)
inf.lands.50kb.rep_8$tmrcaC <- (inf.lands.50kb.rep_8$tmrca - mean(inf.lands.50kb.rep_8$tmrca)) / sd(inf.lands.50kb.rep_8$tmrca)
inf.lands.50kb.rep_8$rhoC <- (inf.lands.50kb.rep_8$rho - mean(inf.lands.50kb.rep_8$rho)) / sd(inf.lands.50kb.rep_8$rho)

inf.lands.50kb.rep_8$bin <- 1:nrow(inf.lands.50kb.rep_8)

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))
```

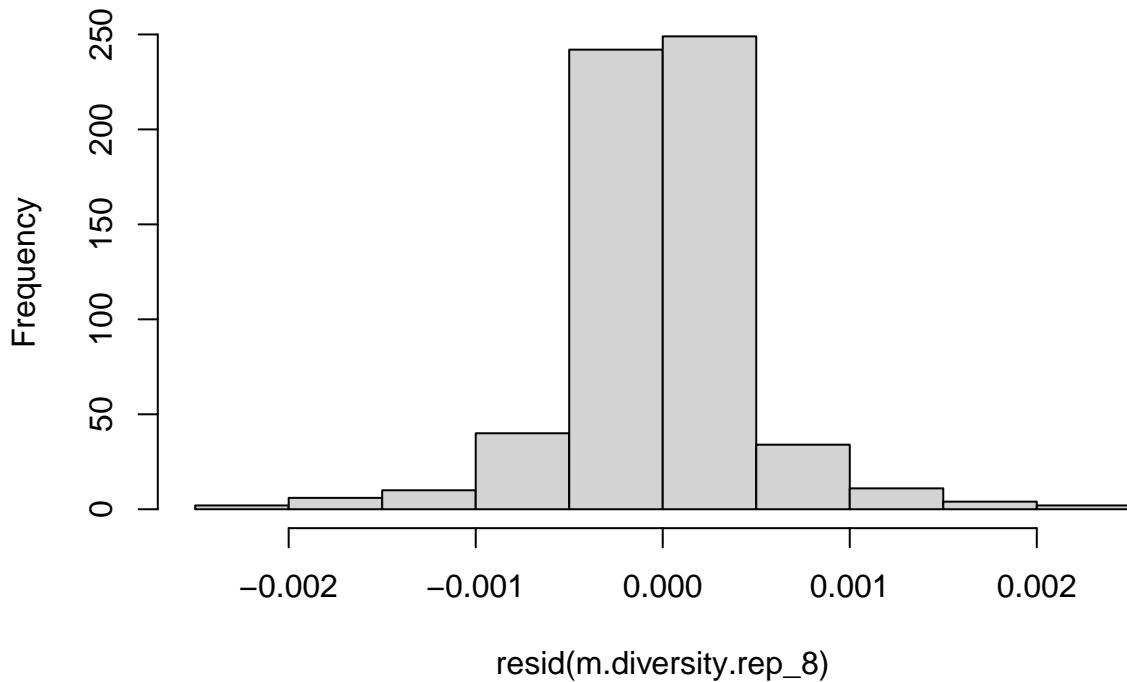


```
dwtest(m.diversity.rep_8)
```

```
##
## 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)
```

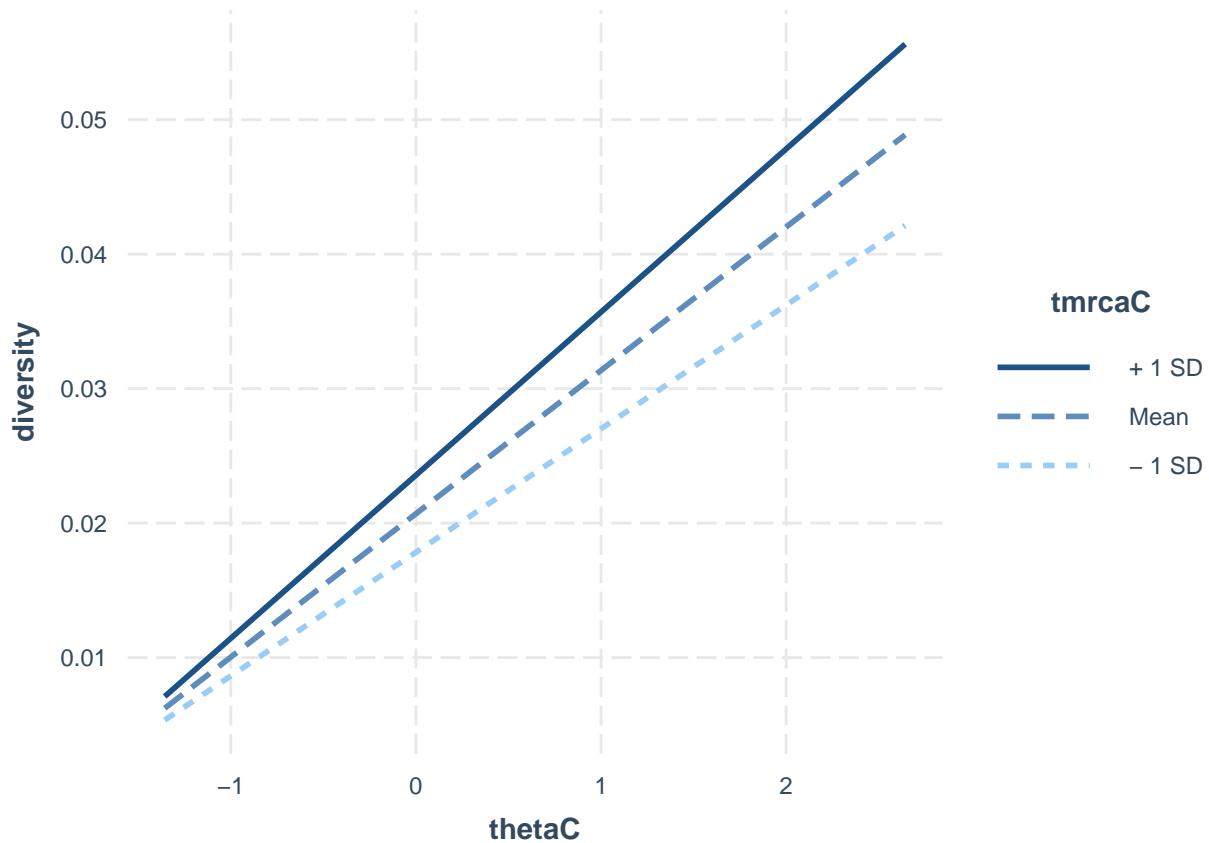
```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_8
## HMC = 0.62278, p-value = 1
hist(resid(m.diversity.rep_8))
```

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      1.066e-02 1.948e-05  547.27 <2e-16 ***
## rhoC       -8.583e-06 1.949e-05   -0.44    0.66
## tmrcaC      2.868e-03 1.985e-05   144.50 <2e-16 ***
## thetaC:tmrcaC 1.469e-03 1.676e-05    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) 0.020703134 2.929892e-05 706.6176 0.0000
## thetaC       0.010670227 2.543405e-05 419.5253 0.0000
## tmrcaC       0.002823421 2.112706e-05 133.6401 0.0000
## rhoC         0.000005961 1.842746e-05   0.3235 0.7465
## thetaC:tmrcaC 0.001402255 1.846302e-05  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) 0.020814330 0.0002097878 99.21612 0.0000
## thetaC       0.010911671 0.0001681892 64.87738 0.0000
## rhoC        -0.000114073 0.0001110840 -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$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCAs.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$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")

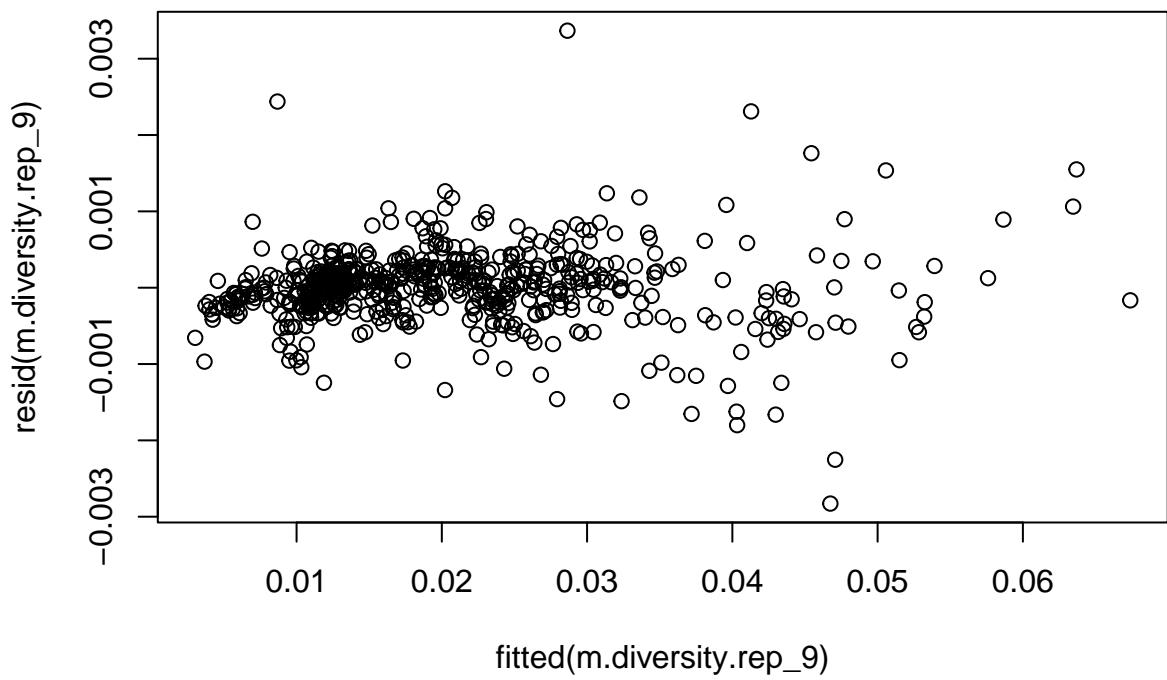
# standardizing
inf.lands.50kb.rep_9$thetaC <- (inf.lands.50kb.rep_9$theta - mean(inf.lands.50kb.rep_9$theta)) / sd(inf.lands.50kb.rep_9$theta)
inf.lands.50kb.rep_9$tmrcaC <- (inf.lands.50kb.rep_9$tmrca - mean(inf.lands.50kb.rep_9$tmrca)) / sd(inf.lands.50kb.rep_9$tmrca)
inf.lands.50kb.rep_9$rhoC <- (inf.lands.50kb.rep_9$rho - mean(inf.lands.50kb.rep_9$rho)) / sd(inf.lands.50kb.rep_9$rho)

inf.lands.50kb.rep_9$bin <- 1:nrow(inf.lands.50kb.rep_9)

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))
```

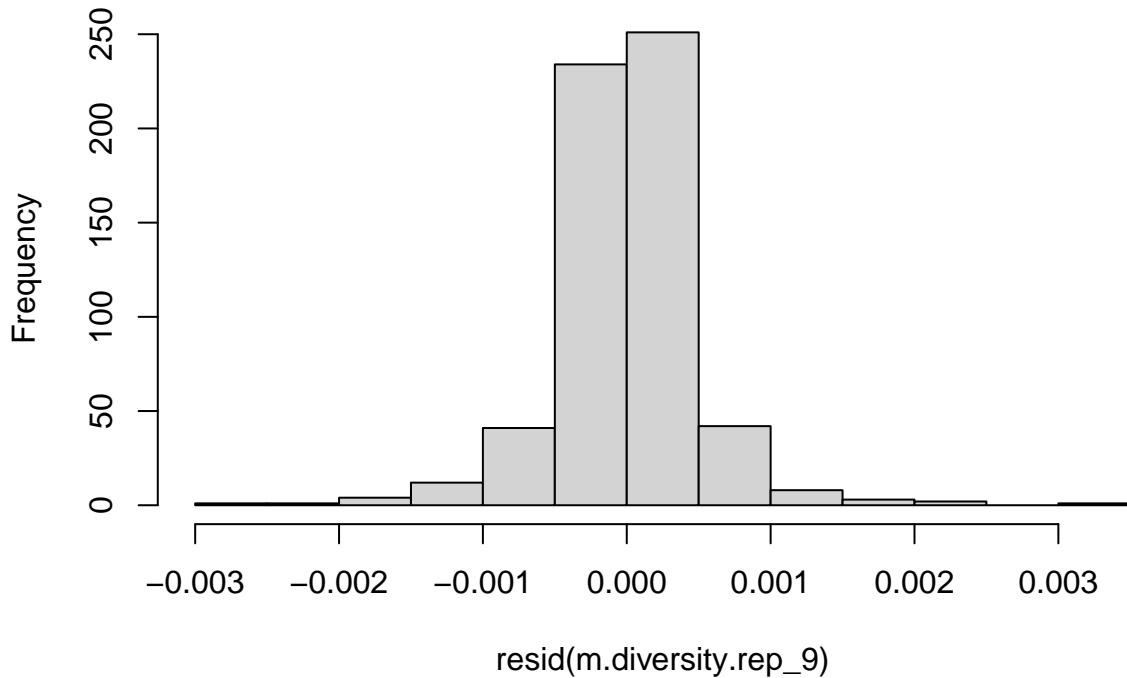


```
dwtest(m.diversity.rep_9)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_9
## DW = 1.3455, p-value = 2.541e-16
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_9)
```

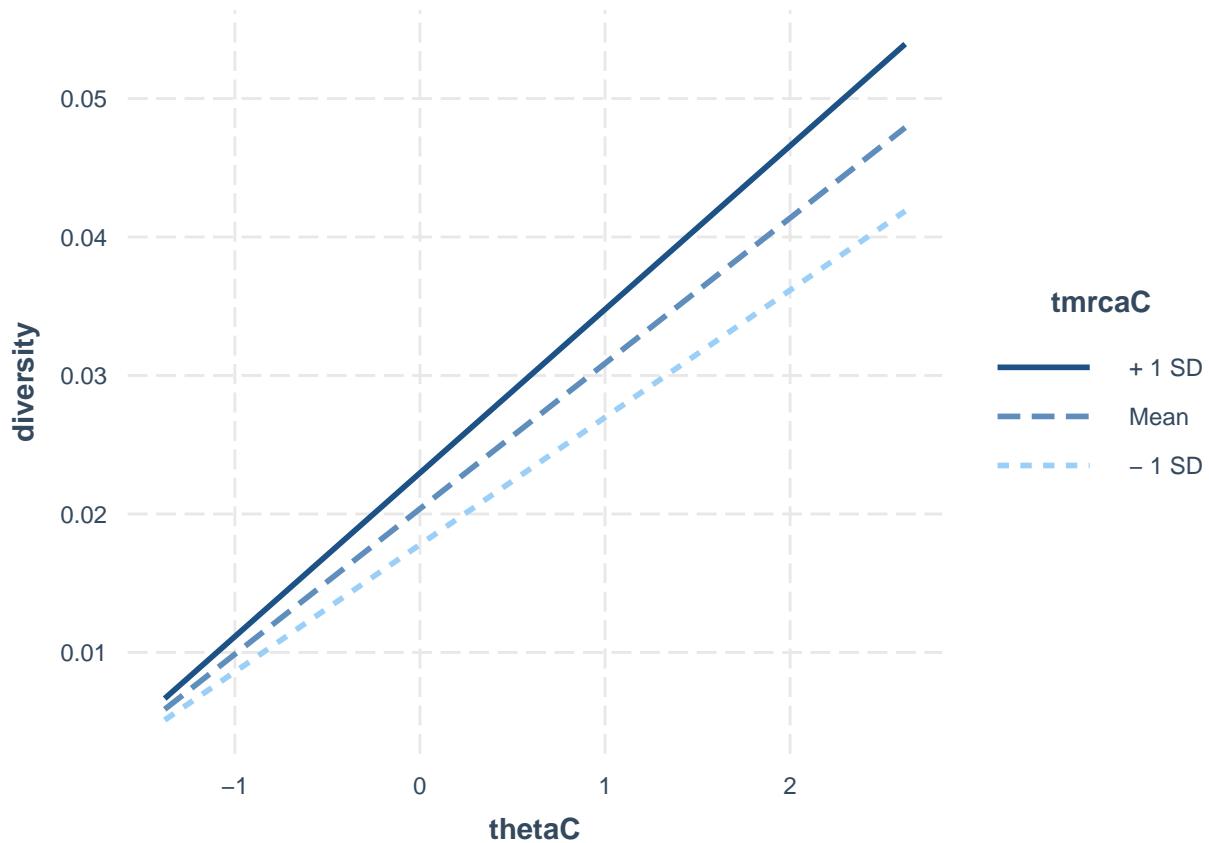
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_9
## HMC = 0.52947, p-value = 0.842
hist(resid(m.diversity.rep_9))
```

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      1.050e-02 2.075e-05 505.918 <2e-16 ***
## rhoC        1.416e-05 2.077e-05   0.682   0.496
## tmrcaC      2.587e-03 2.102e-05 123.090 <2e-16 ***
## thetaC:tmrcaC 1.313e-03 1.768e-05  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) 0.020383482 3.257124e-05 625.8123 0.0000
## thetaC      0.010499421 2.766864e-05 379.4701 0.0000
## tmrcaC      0.002523327 2.134664e-05 118.2072 0.0000
## rhoC        0.000001758 1.928100e-05   0.0912 0.9274
## thetaC:tmrcaC 0.001238743 1.860130e-05  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
## -4.86419787 -0.43551103  0.02063926  0.43146966  6.60989138
##
## 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) 0.020529891 0.0001758026 116.77810 0.0000
## thetaC       0.010960389 0.0001518094  72.19836 0.0000
## rhoC        -0.000201324 0.0001095635  -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$avg <- apply(pi.50kb[4:ncol(pi.50kb)], 1, mean)
tmrca.50kb <- read.table(paste(p, "TMRCAs.50kb.bedgraph", sep = ""), header = T)
tmrca.50kb$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

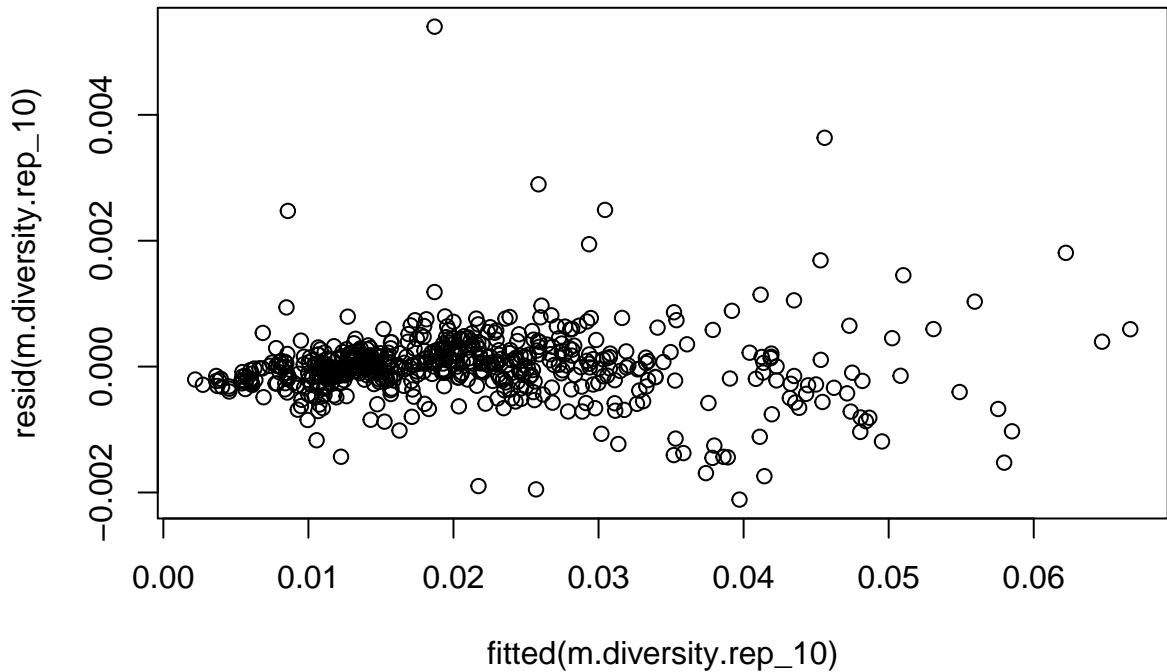
inf.lands.50kb.rep_10 <- as.data.frame(cbind(pi.50kb$avg, theta.50kb$sample_mean, rho.50kb$sample_mean,
names(inf.lands.50kb.rep_10) <- c("diversity", "theta", "rho", "tmrca")

# standardizing
inf.lands.50kb.rep_10$thetaC <- (inf.lands.50kb.rep_10$theta - mean(inf.lands.50kb.rep_10$theta)) / sd(inf.lands.50kb.rep_10$theta)
inf.lands.50kb.rep_10$tmrcaC <- (inf.lands.50kb.rep_10$tmrca - mean(inf.lands.50kb.rep_10$tmrca)) / sd(inf.lands.50kb.rep_10$tmrca)
inf.lands.50kb.rep_10$rhoC <- (inf.lands.50kb.rep_10$rho - mean(inf.lands.50kb.rep_10$rho)) / sd(inf.lands.50kb.rep_10$rho)

inf.lands.50kb.rep_10$bin <- 1:nrow(inf.lands.50kb.rep_10)

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))
```

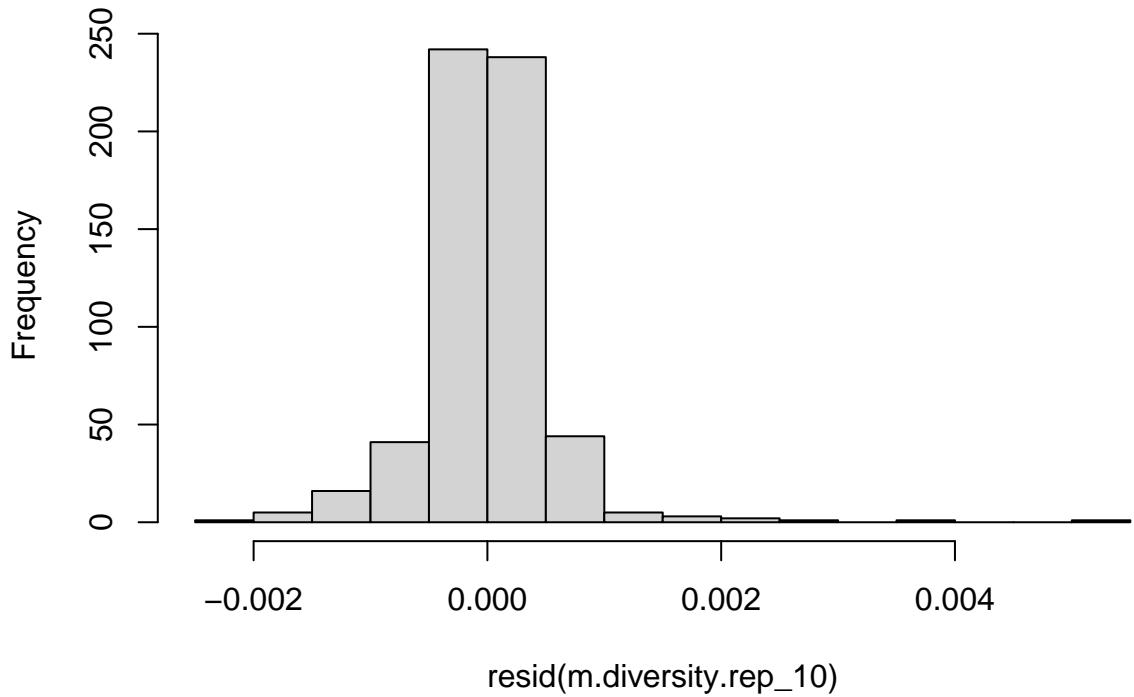


```
dwtest(m.diversity.rep_10)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_10
## DW = 1.3449, p-value = 2.3e-16
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_10)
```

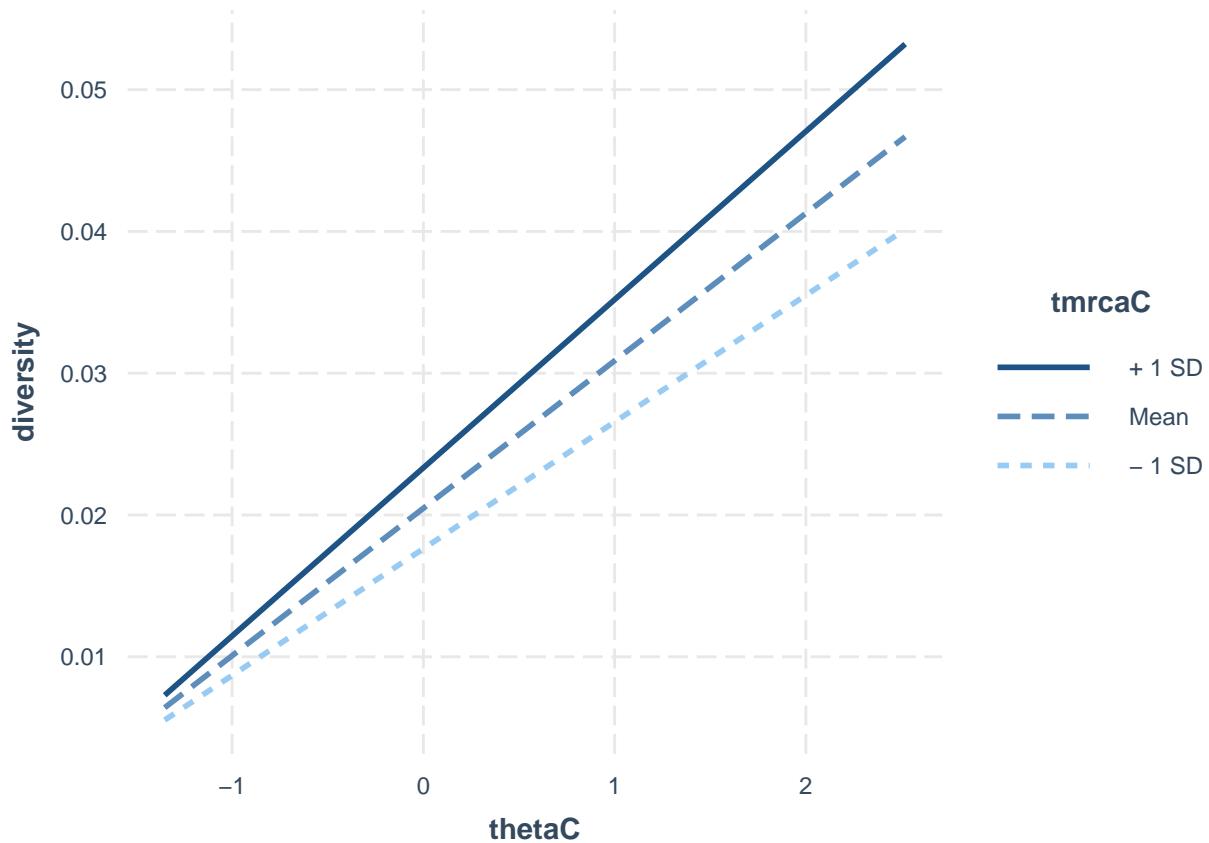
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_10
## HMC = 0.67092, p-value = 1
hist(resid(m.diversity.rep_10))
```

Histogram of resid(m.diversity.rep_10)



```
summary(m.diversity.rep_10)
```

```
##  
## Call:  
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,  
##      data = inf.lands.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      1.040e-02 2.330e-05 446.391 <2e-16 ***  
## rhoC       -2.795e-05 2.314e-05 -1.208   0.227  
## tmrcaC      2.853e-03 2.435e-05 117.152 <2e-16 ***  
## thetaC:tmrcaC 1.462e-03 2.084e-05 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) 0.020438810 3.125509e-05 653.9355 0.0000
## thetaC      0.010435258 2.837609e-05 367.7482 0.0000
## tmrcaC      0.002817637 2.481844e-05 113.5300 0.0000
## rhoC        0.000004491 2.161341e-05   0.2078 0.8355
## thetaC:tmrcaC 0.001426645 2.293513e-05  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) 0.020708204 0.0001936334 106.94544 0.0000
## thetaC       0.011234806 0.0001578890  71.15634 0.0000
## rhoC        -0.000216985 0.0001055177  -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$VarExp[3]
r2.inf.50kb[2, 10] <- anova.diversity$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))

```

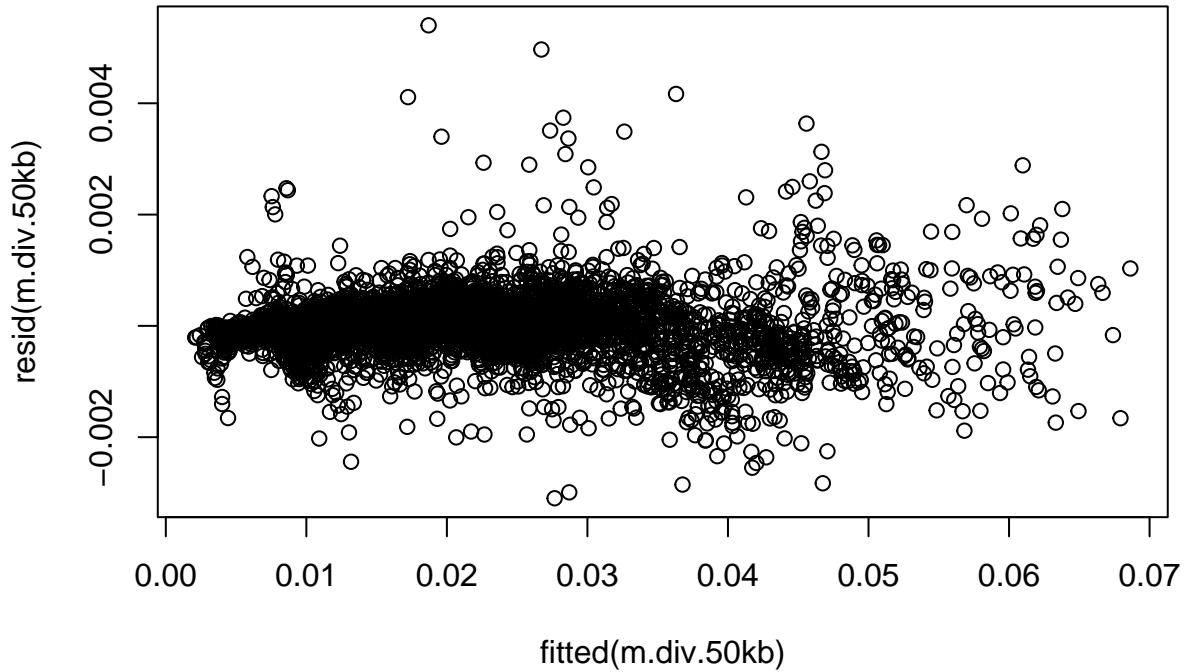
2.1.11 all replicates:

```

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)

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)

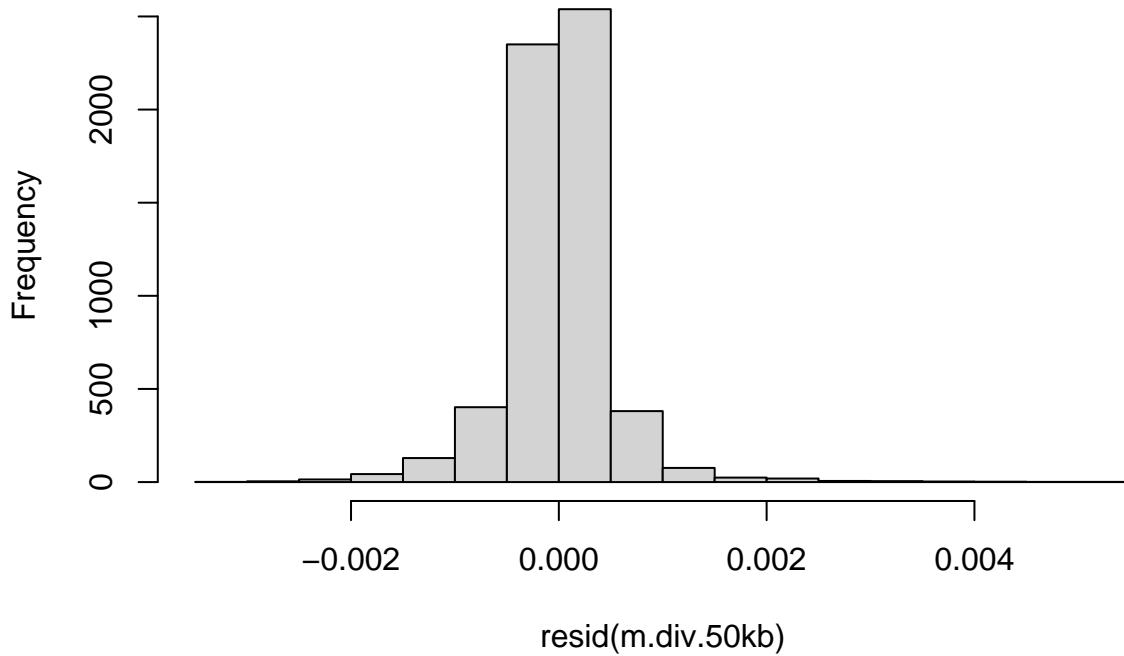
##
##  Durbin-Watson test
##
##  data:  m.div.50kb
##  DW = 1.3249, p-value < 2.2e-16
##  alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.50kb)

##
##  Harrison-McCabe test
##
##  data:  m.div.50kb
##  HMC = 0.50366, p-value = 0.631

```

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

Histogram of resid(m.div.50kb)



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

```
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      1.041e-02 2.110e-05 493.438 < 2e-16 ***
## rhoC       -5.658e-06 2.093e-05  -0.270  0.78691
## tmrcaC      2.854e-03 2.172e-05 131.441 < 2e-16 ***
## as.factor(Replicate)2 -2.363e-06 2.968e-05  -0.080  0.93653
```

```

## 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  0.23433 
## as.factor(Replicate)6      3.133e-05  2.964e-05  1.057  0.29053 
## as.factor(Replicate)7      3.036e-05  2.964e-05  1.024  0.30573 
## 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  0.04970 *  
## as.factor(Replicate)10     4.479e-05  2.975e-05  1.506  0.13220 
## thetaC:tmrcaC             1.473e-03  1.997e-05  73.755 < 2e-16 *** 
## thetaC:as.factor(Replicate)2 -1.097e-06 2.979e-05 -0.037  0.97063 
## thetaC:as.factor(Replicate)3 1.365e-04  2.983e-05  4.576  4.84e-06 *** 
## thetaC:as.factor(Replicate)4 3.112e-04  2.971e-05  10.476 < 2e-16 *** 
## thetaC:as.factor(Replicate)5 1.457e-04  2.974e-05  4.899  9.88e-07 *** 
## thetaC:as.factor(Replicate)6 1.453e-04  2.973e-05  4.887  1.05e-06 *** 
## thetaC:as.factor(Replicate)7 6.703e-05  2.975e-05  2.254  0.02426 *  
## thetaC:as.factor(Replicate)8 2.469e-04  2.977e-05  8.293  < 2e-16 *** 
## thetaC:as.factor(Replicate)9 8.585e-05  2.978e-05  2.883  0.00396 ** 
## thetaC:as.factor(Replicate)10 -1.335e-05 2.992e-05 -0.446  0.65540 
## rhoC:as.factor(Replicate)2 -8.986e-06 2.962e-05 -0.303  0.76159 
## rhoC:as.factor(Replicate)3 -1.179e-05 2.960e-05 -0.398  0.69036 
## rhoC:as.factor(Replicate)4 -8.321e-06 2.966e-05 -0.281  0.77904 
## rhoC:as.factor(Replicate)5 -2.745e-06 2.965e-05 -0.093  0.92623 
## rhoC:as.factor(Replicate)6 1.791e-05  2.974e-05  0.602  0.54700 
## rhoC:as.factor(Replicate)7 -4.435e-05 2.956e-05 -1.500  0.13364 
## rhoC:as.factor(Replicate)8 -2.925e-06 2.966e-05 -0.099  0.92144 
## rhoC:as.factor(Replicate)9 1.982e-05  2.968e-05  0.668  0.50422 
## rhoC:as.factor(Replicate)10 -2.229e-05 2.969e-05 -0.751  0.45282 
## tmrcaC:as.factor(Replicate)2 -1.264e-04 3.069e-05 -4.119  3.86e-05 *** 
## tmrcaC:as.factor(Replicate)3 2.399e-05  3.035e-05  0.791  0.42926 
## tmrcaC:as.factor(Replicate)4 5.082e-05  3.032e-05  1.676  0.09379 . 
## tmrcaC:as.factor(Replicate)5 -8.629e-05 3.035e-05 -2.844  0.00448 ** 
## tmrcaC:as.factor(Replicate)6 -1.761e-04 3.056e-05 -5.762  8.72e-09 *** 
## tmrcaC:as.factor(Replicate)7 -1.481e-04 3.054e-05 -4.849  1.27e-06 *** 
## tmrcaC:as.factor(Replicate)8 1.358e-05  3.049e-05  0.445  0.65611 
## tmrcaC:as.factor(Replicate)9 -2.674e-04 3.041e-05 -8.794  < 2e-16 *** 
## tmrcaC:as.factor(Replicate)10 -1.141e-06 3.103e-05 -0.037  0.97068 
## thetaC:tmrcaC:as.factor(Replicate)2 -8.373e-05 2.693e-05 -3.109  0.00189 ** 
## thetaC:tmrcaC:as.factor(Replicate)3 -5.557e-05 2.678e-05 -2.075  0.03805 *  
## thetaC:tmrcaC:as.factor(Replicate)4 1.888e-05  2.771e-05  0.681  0.49559 
## thetaC:tmrcaC:as.factor(Replicate)5 -6.033e-05 2.652e-05 -2.275  0.02294 *  
## thetaC:tmrcaC:as.factor(Replicate)6 -1.164e-04 2.701e-05 -4.310  1.66e-05 *** 
## thetaC:tmrcaC:as.factor(Replicate)7 -1.180e-04 2.751e-05 -4.290  1.81e-05 *** 
## thetaC:tmrcaC:as.factor(Replicate)8 -4.269e-06 2.693e-05 -0.158  0.87407 
## thetaC:tmrcaC:as.factor(Replicate)9 -1.600e-04 2.683e-05 -5.963  2.61e-09 *** 
## thetaC:tmrcaC:as.factor(Replicate)10 -1.056e-05 2.754e-05 -0.383  0.70146 
## --- 
## 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

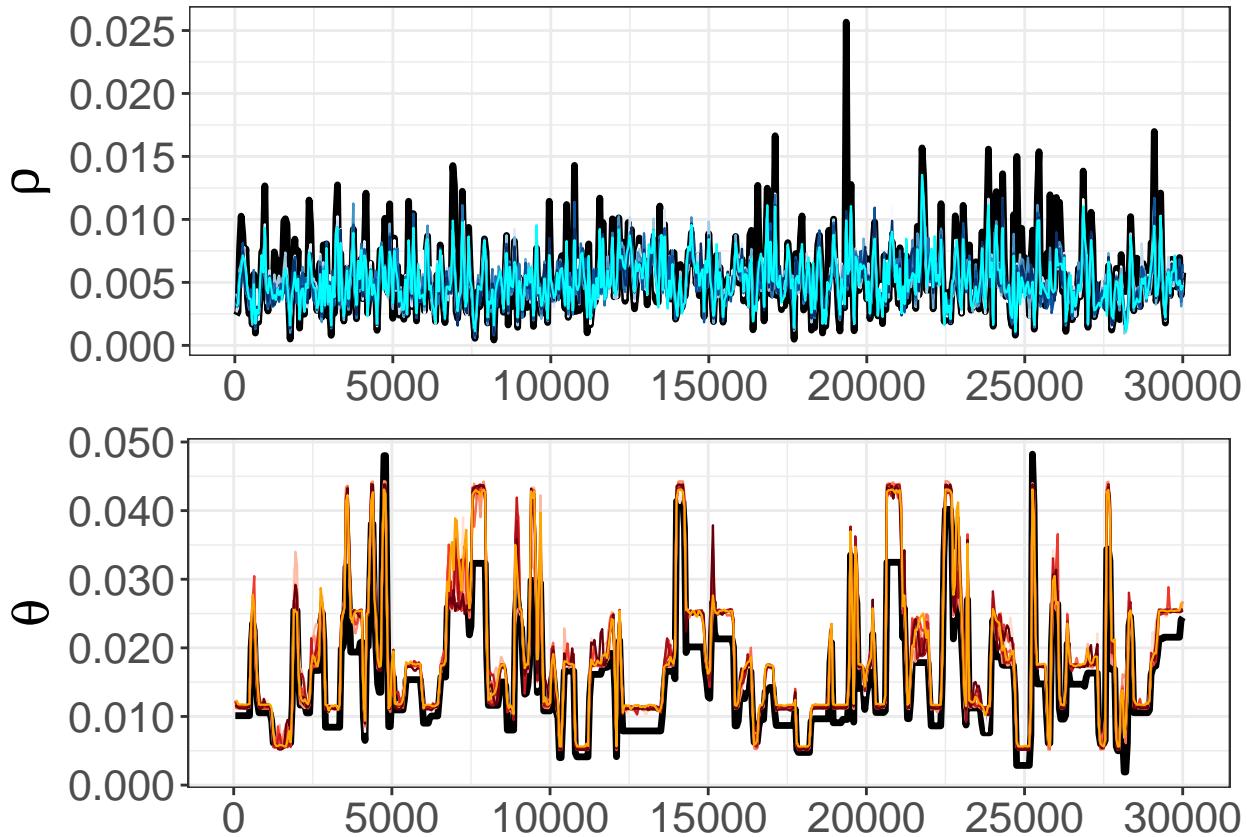
names(rho.plot) <- c("bin", "sim", reps)
molten.rho <- melt(rho.plot, id.vars = "bin")
rho.map.50kb <- ggplot(data = molten.rho, aes(x = bin * 50, y = value, colour = variable)) + theme_bw()
rho.map.50kb <- rho.map.50kb + geom_line(data = molten.rho, aes(size = variable)) + scale_size_manual(values = c("black", brewer.pal(n = 9, name = "Blues")))
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_bw()
theta.map.50kb <- theta.map.50kb + geom_line(data = molten.theta, aes(size = variable)) + scale_size_manual(values = c("black", brewer.pal(n = 9, name = "Reds")))
theta.map.50kb <- theta.map.50kb + scale_color_manual(values = c("black", brewer.pal(n = 9, name = "Reds")))
theta.map.50kb <- theta.map.50kb + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
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(size = 20))

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")  

# standardizing  

inf.lands.200kb.rep_1$thetaC <- (inf.lands.200kb.rep_1$theta - mean(inf.lands.200kb.rep_1$theta)) / sd(inf.lands.200kb.rep_1$theta)  

inf.lands.200kb.rep_1$tmrcaC <- (inf.lands.200kb.rep_1$tmrca - mean(inf.lands.200kb.rep_1$tmrca)) / sd(inf.lands.200kb.rep_1$tmrca)  

inf.lands.200kb.rep_1$rhoC <- (inf.lands.200kb.rep_1$rho - mean(inf.lands.200kb.rep_1$rho)) / sd(inf.lands.200kb.rep_1$rho)  

inf.lands.200kb.rep_1$bin <- 1:nrow(inf.lands.200kb.rep_1)  

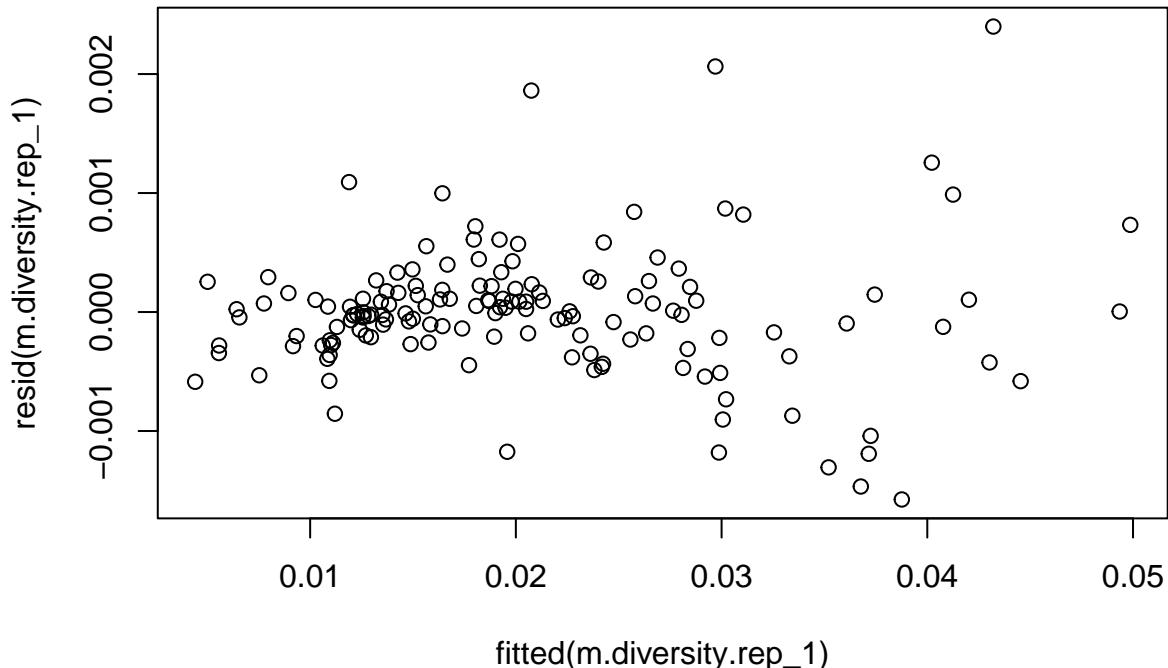
inf.lands.200kb.rep_1$Replicate <- 1  

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

```

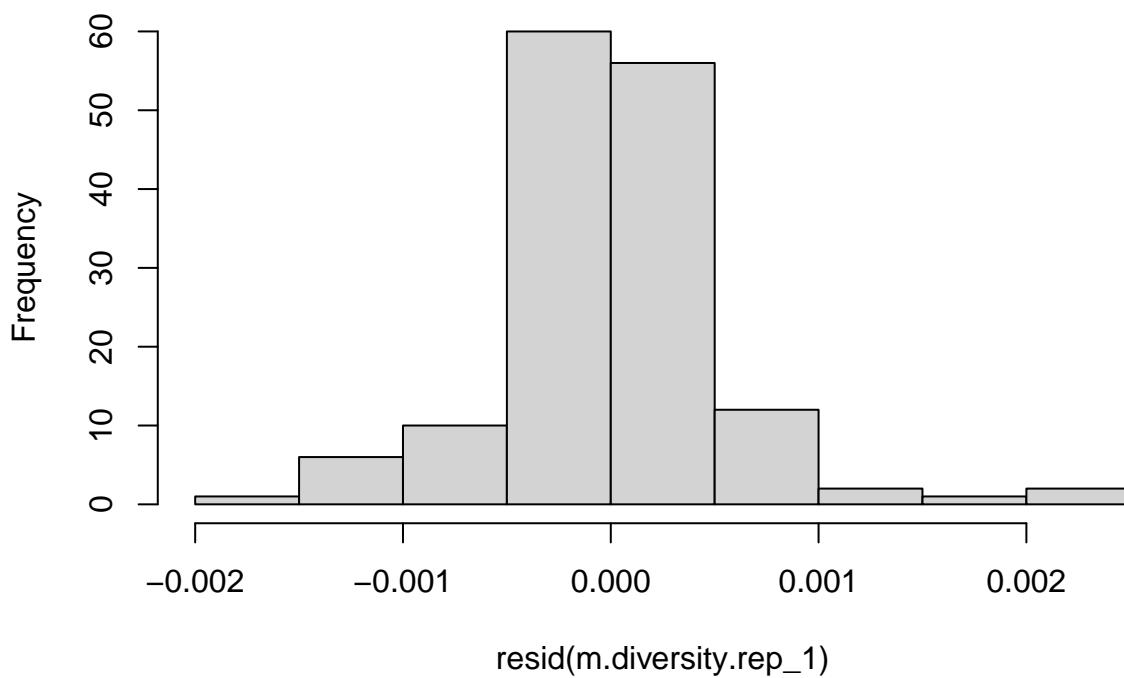
```

hmctest(m.diversity.rep_1)

##
##  Harrison-McCabe test
##
## data: m.diversity.rep_1
## HMC = 0.49224, p-value = 0.454
hist(resid(m.diversity.rep_1))

```

Histogram of resid(m.diversity.rep_1)



```

summary(m.diversity.rep_1)

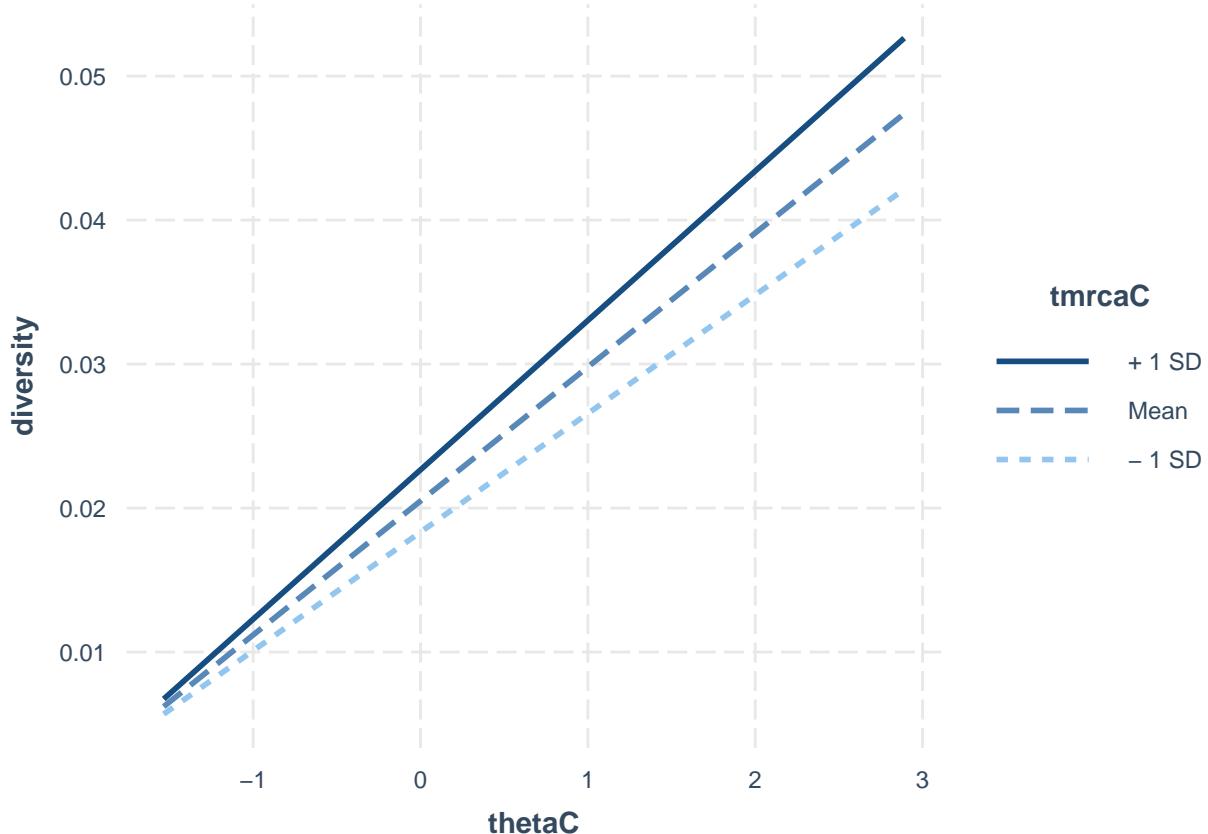
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##      data = inf.lands.200kb.rep_1)
##
## Residuals:
##      Min        1Q    Median        3Q       Max
## -1.575e-03 -2.359e-04 -1.035e-05  1.834e-04  2.399e-03
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.049e-02  4.583e-05 447.178 <2e-16 ***
## thetaC       9.304e-03  4.600e-05 202.263 <2e-16 ***
## rhoC        1.063e-05  4.731e-05   0.225   0.823
## tmrcaC      2.164e-03  5.156e-05  41.966 <2e-16 ***
## thetaC:tmrcaC 1.069e-03  4.530e-05  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) 0.020495842 4.988827e-05 410.8349 0.0000
## thetaC      0.009311061 4.778895e-05 194.8371 0.0000

```

```

## tmrcaC      0.002162321 5.209260e-05 41.5092  0.0000
## rhoC        0.000016367 4.723723e-05  0.3465  0.7295
## thetaC:tmrcaC 0.001058085 4.650120e-05 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) 0.020629702 0.0002057241 100.27847 0.0000
## thetaC      0.009546181 0.0001792629  53.25239 0.0000
## rhoC      -0.000425486 0.0001636252  -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$VarExp[3])
r2.inf.200kb[2, 1] <- anova.diversity$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))
names(inf.lands.200kb.rep_2) <- c("diversity", "theta", "rho", "tmrca")

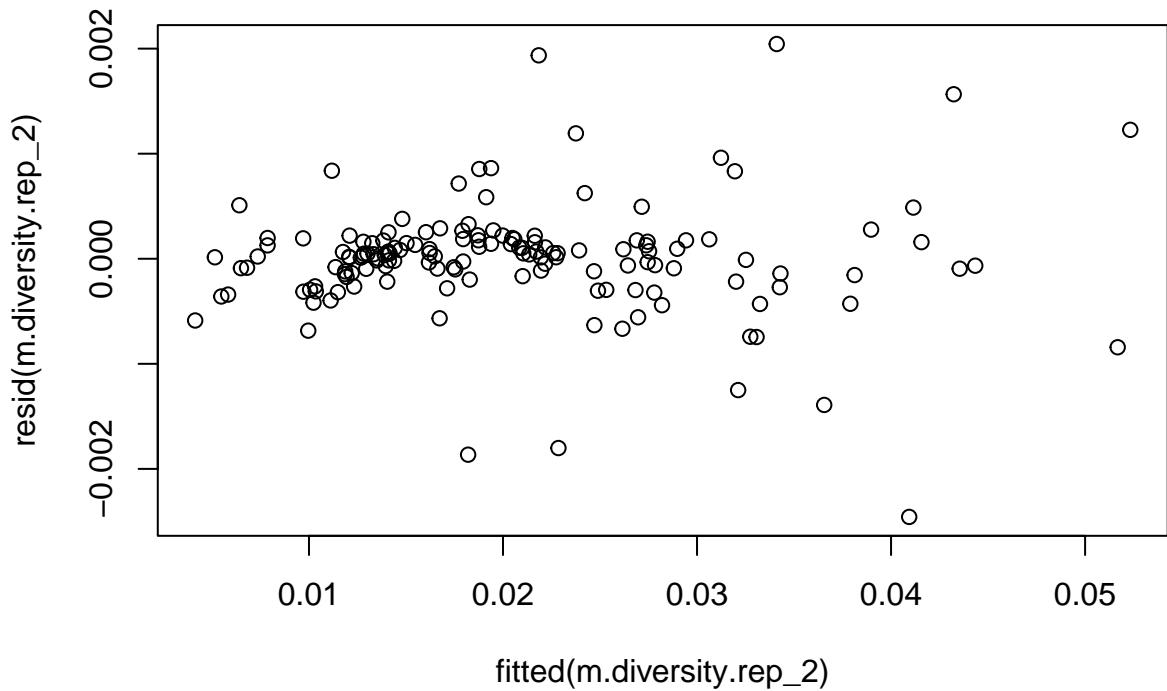
# standardizing
inf.lands.200kb.rep_2$thetaC <- (inf.lands.200kb.rep_2$theta - mean(inf.lands.200kb.rep_2$theta)) / sd(inf.lands.200kb.rep_2$theta)
inf.lands.200kb.rep_2$tmrcaC <- (inf.lands.200kb.rep_2$tmrca - mean(inf.lands.200kb.rep_2$tmrca)) / sd(inf.lands.200kb.rep_2$tmrca)
inf.lands.200kb.rep_2$rhoC <- (inf.lands.200kb.rep_2$rho - mean(inf.lands.200kb.rep_2$rho)) / sd(inf.lands.200kb.rep_2$rho)

inf.lands.200kb.rep_2$bin <- 1:nrow(inf.lands.200kb.rep_2)

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))

```



```

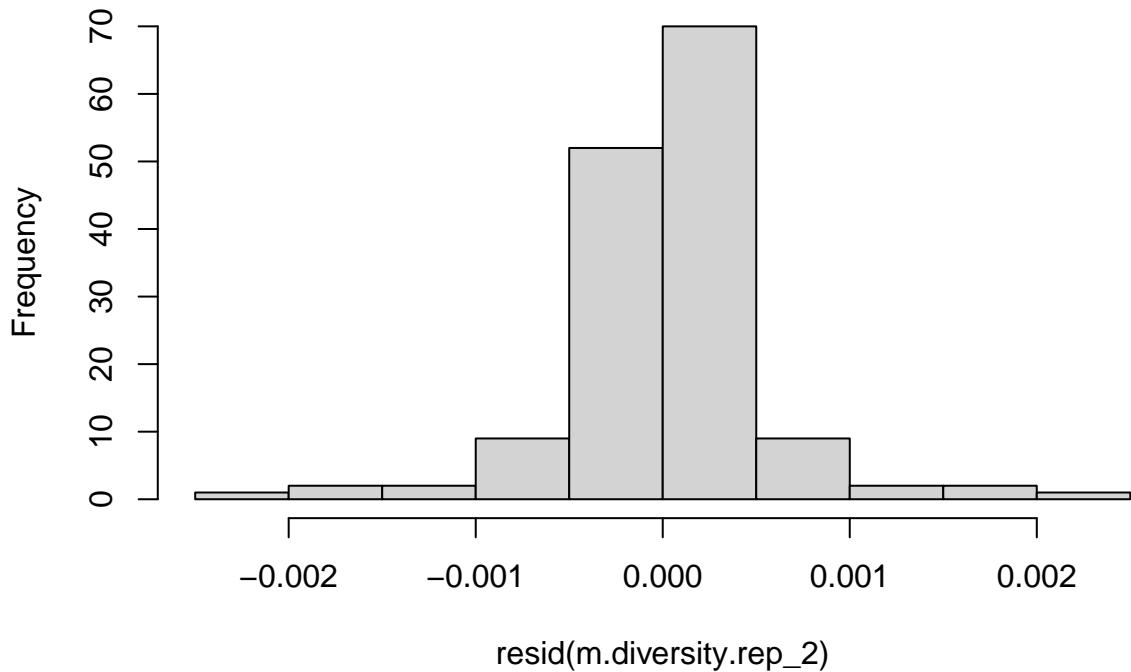
dwtest(m.diversity.rep_2)

##
##  Durbin-Watson test
##
##  data: m.diversity.rep_2
##  DW = 1.7156, p-value = 0.03159
##  alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_2)

##
##  Harrison-McCabe test
##
##  data: m.diversity.rep_2
##  HMC = 0.50531, p-value = 0.527
hist(resid(m.diversity.rep_2))

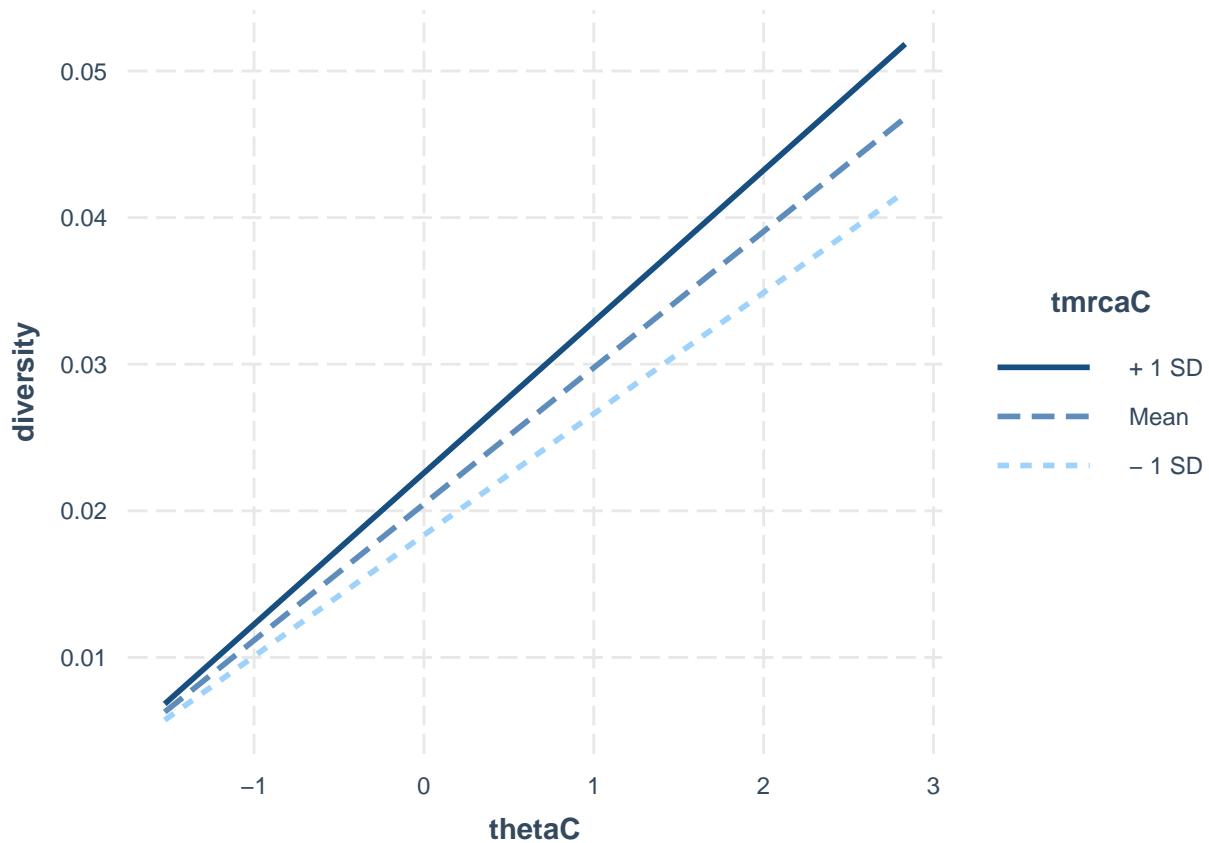
```

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       9.296e-03 4.536e-05 204.913 <2e-16 ***
## rhoC        6.479e-06 4.578e-05   0.142   0.888
## tmrcaC      2.119e-03 5.353e-05  39.582 <2e-16 ***
## thetaC:tmrcaC 1.028e-03 4.397e-05  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) 0.020468080 5.328964e-05 384.0912 0.0000
## thetaC       0.009307885 4.854920e-05 191.7207 0.0000
## tmrcaC       0.002110354 5.491646e-05  38.4284 0.0000
## rhoC        -0.000004254 4.460197e-05  -0.0954 0.9242
## thetaC:tmrcaC 0.000998188 4.498116e-05  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) 0.020600840 0.0002121094 97.12366 0.0000
## thetaC       0.009645554 0.0001663662 57.97785 0.0000
## rhoC        -0.000279070 0.0001401935 -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

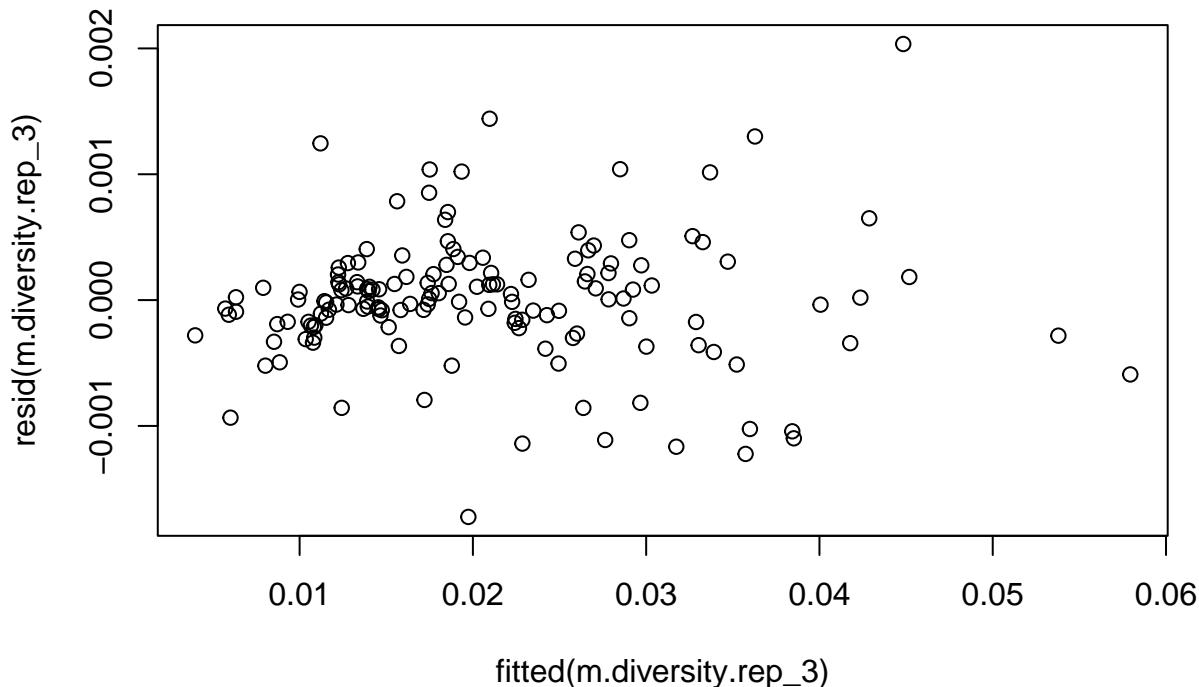
inf.lands.200kb.rep_3 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_mean))
names(inf.lands.200kb.rep_3) <- c("diversity", "theta", "rho", "tmrca")

# standardizing
inf.lands.200kb.rep_3$thetaC <- (inf.lands.200kb.rep_3$theta - mean(inf.lands.200kb.rep_3$theta)) / sd(inf.lands.200kb.rep_3$theta)
inf.lands.200kb.rep_3$tmrcaC <- (inf.lands.200kb.rep_3$tmrca - mean(inf.lands.200kb.rep_3$tmrca)) / sd(inf.lands.200kb.rep_3$tmrca)
inf.lands.200kb.rep_3$rhoC <- (inf.lands.200kb.rep_3$rho - mean(inf.lands.200kb.rep_3$rho)) / sd(inf.lands.200kb.rep_3$rho)

inf.lands.200kb.rep_3$bin <- 1:nrow(inf.lands.200kb.rep_3)

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))
```



```

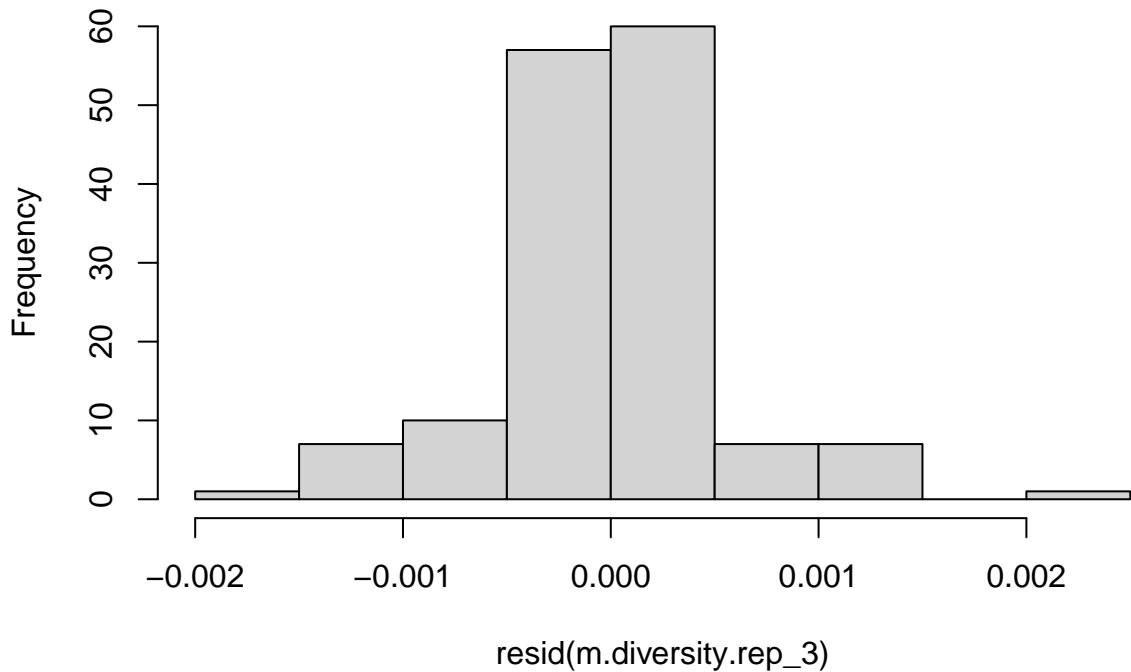
dwtest(m.diversity.rep_3)

##
## Durbin-Watson test
##
## data: m.diversity.rep_3
## DW = 1.7271, p-value = 0.03626
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_3)

##
## Harrison-McCabe test
##
## data: m.diversity.rep_3
## HMC = 0.49269, p-value = 0.464
hist(resid(m.diversity.rep_3))

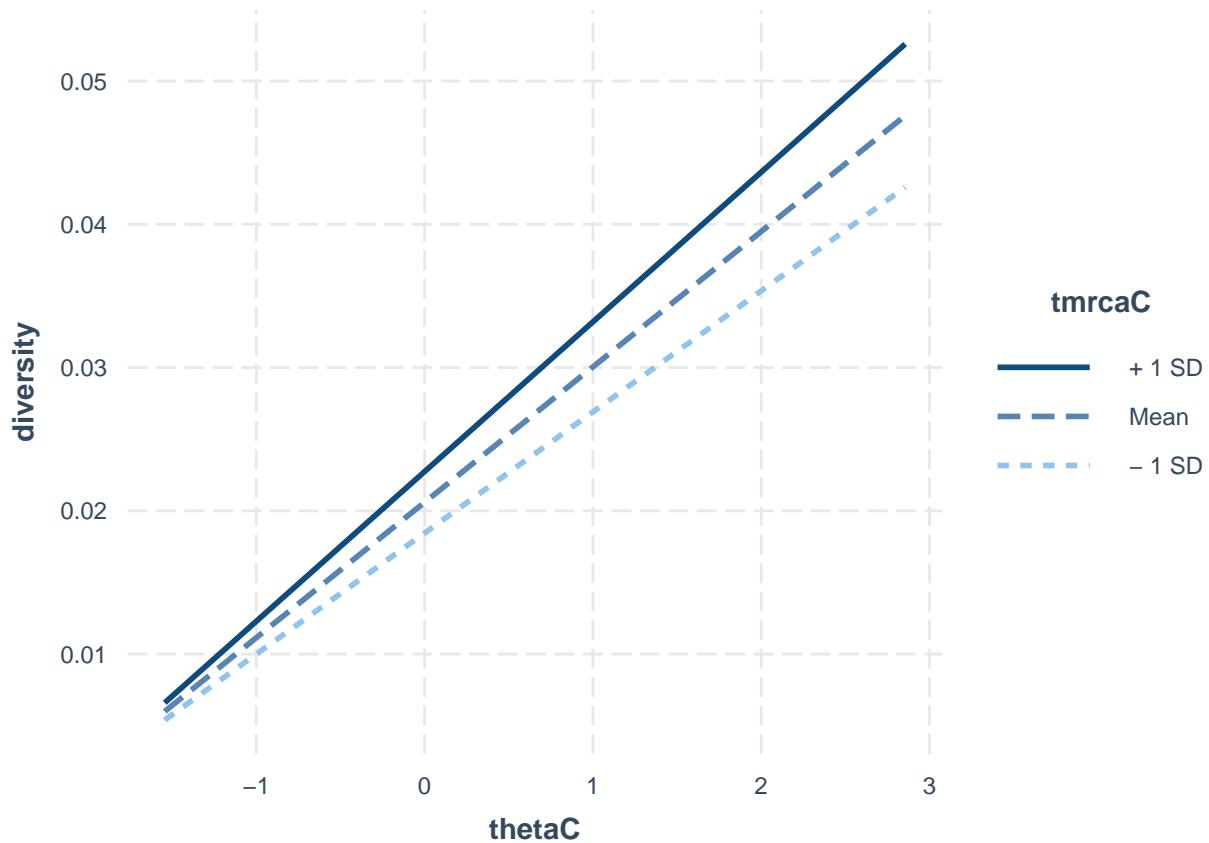
```

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       9.450e-03 4.305e-05 219.494 <2e-16 ***  
## rhoC        2.539e-05 4.324e-05   0.587   0.558  
## tmrcaC      2.146e-03 4.450e-05  48.218 <2e-16 ***  
## thetaC:tmrcaC 9.984e-04 3.487e-05  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) 0.020600045 4.978633e-05 413.7691 0.0000
## thetaC      0.009461799 4.584821e-05 206.3723 0.0000
## tmrcaC      0.002149192 4.552731e-05  47.2067 0.0000
## rhoC        0.000030344 4.317566e-05    0.7028 0.4833
## thetaC:tmrcaC 0.000979013 3.607400e-05  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) 0.020742949 0.0002477221 83.73477 0.000
## thetaC       0.009869059 0.0002025119 48.73324 0.000
## rhoC        -0.000360078 0.0001797499 -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

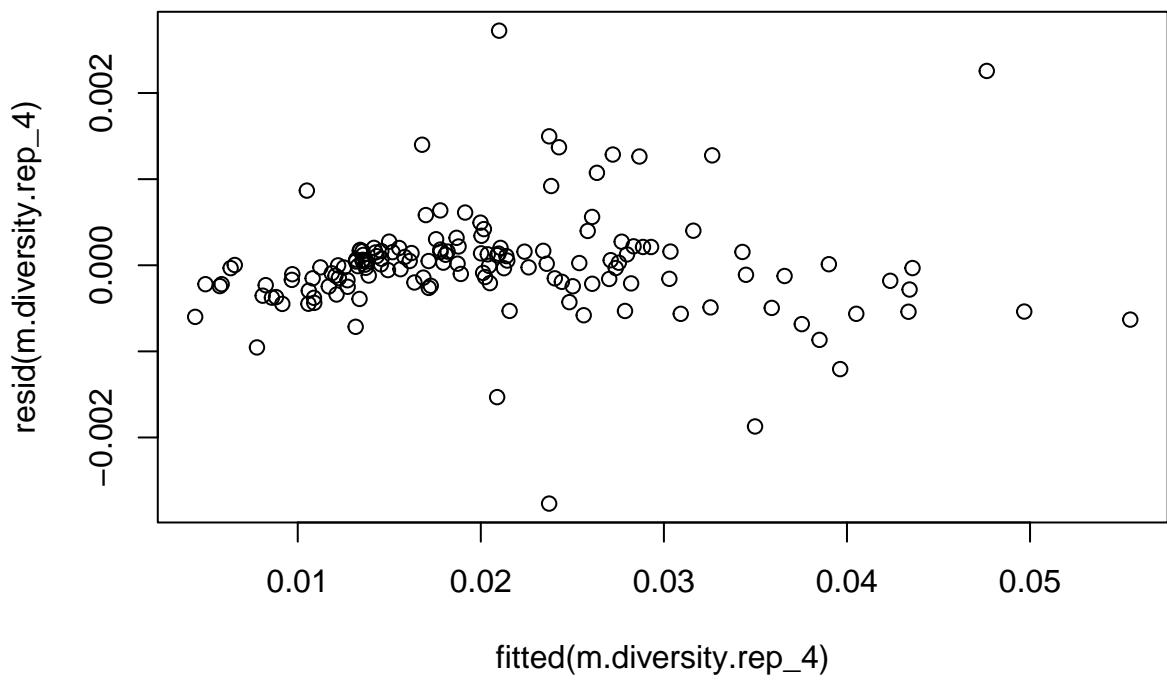
inf.lands.200kb.rep_4 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_mean))
names(inf.lands.200kb.rep_4) <- c("diversity", "theta", "rho", "tmrca")

# standardizing
inf.lands.200kb.rep_4$thetaC <- (inf.lands.200kb.rep_4$theta - mean(inf.lands.200kb.rep_4$theta)) / sd(inf.lands.200kb.rep_4$theta)
inf.lands.200kb.rep_4$tmrcaC <- (inf.lands.200kb.rep_4$tmrca - mean(inf.lands.200kb.rep_4$tmrca)) / sd(inf.lands.200kb.rep_4$tmrca)
inf.lands.200kb.rep_4$rhoC <- (inf.lands.200kb.rep_4$rho - mean(inf.lands.200kb.rep_4$rho)) / sd(inf.lands.200kb.rep_4$rho)

inf.lands.200kb.rep_4$bin <- 1:nrow(inf.lands.200kb.rep_4)

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))
```

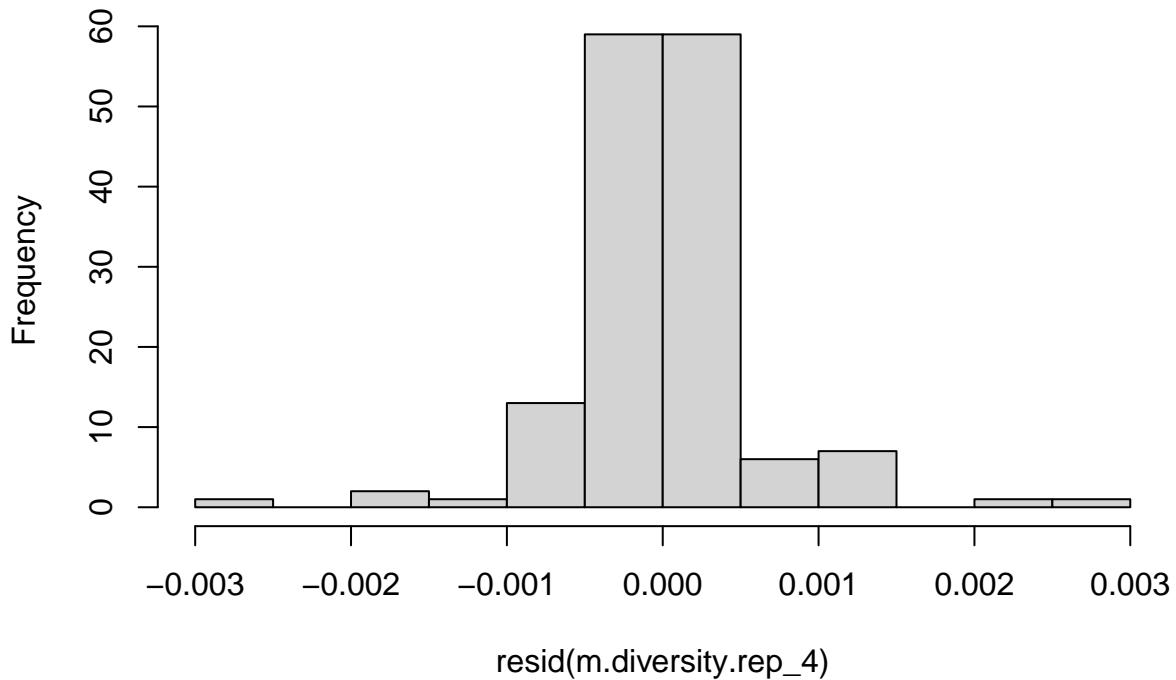


```
dwtest(m.diversity.rep_4)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_4
## DW = 1.6283, p-value = 0.00823
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_4)
```

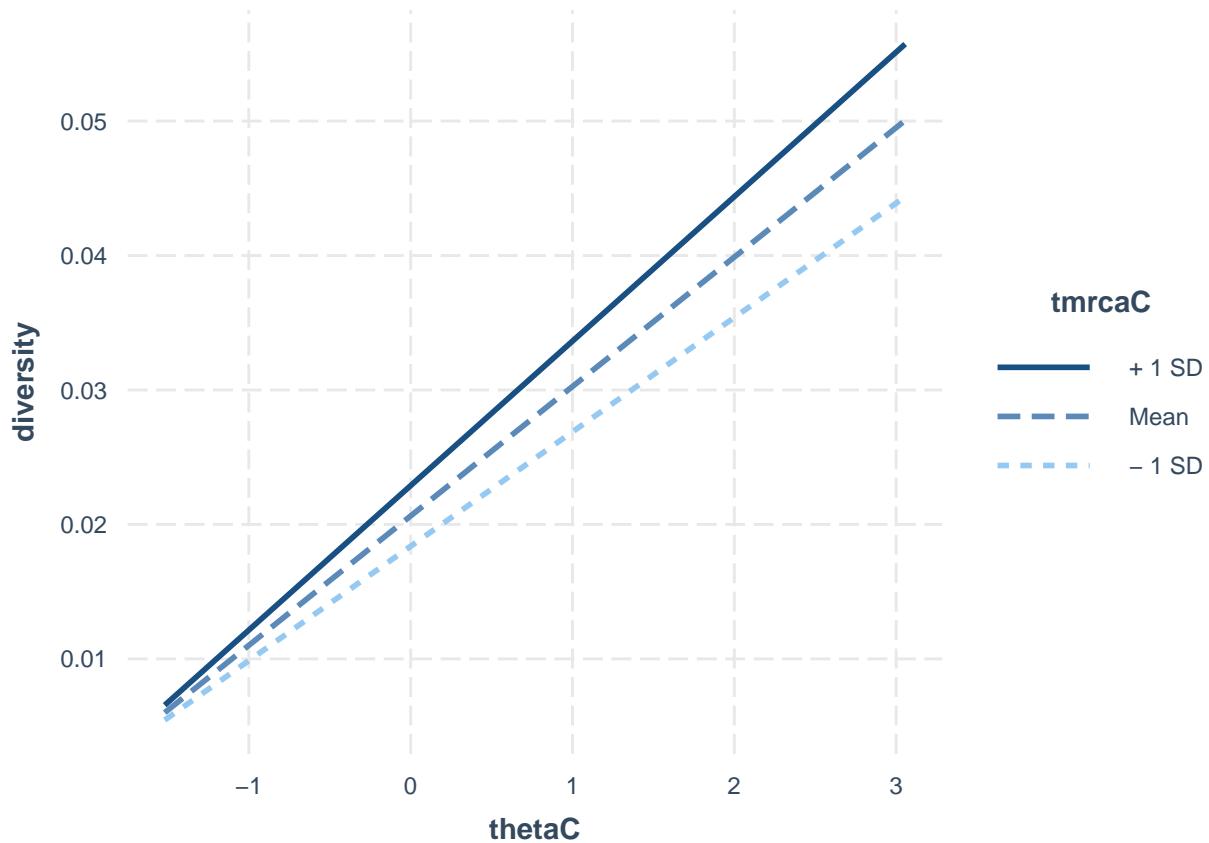
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_4
## HMC = 0.54628, p-value = 0.772
hist(resid(m.diversity.rep_4))
```

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       9.627e-03 5.010e-05 192.138 <2e-16 ***  
## rhoC        4.448e-05 5.099e-05   0.872   0.384  
## tmrcaC      2.255e-03 5.329e-05  42.319 <2e-16 ***  
## thetaC:tmrcaC 1.120e-03 5.049e-05  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) 0.020634635 6.076133e-05 339.6014 0.0000
## thetaC      0.009603855 5.361677e-05 179.1204 0.0000
## tmrcaC      0.002263942 5.453402e-05  41.5143 0.0000
## rhoC        0.000049037 4.906171e-05   0.9995 0.3192
## thetaC:tmrcaC 0.001095939 5.249914e-05  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) 0.020689764 0.0002459711 84.11461 0.0000
## thetaC       0.009677540 0.0002002974 48.31587 0.0000
## rhoC        -0.000283273 0.0001734135 -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$VarExp[3]
r2.inf.200kb[2, 4] <- anova.diversity$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

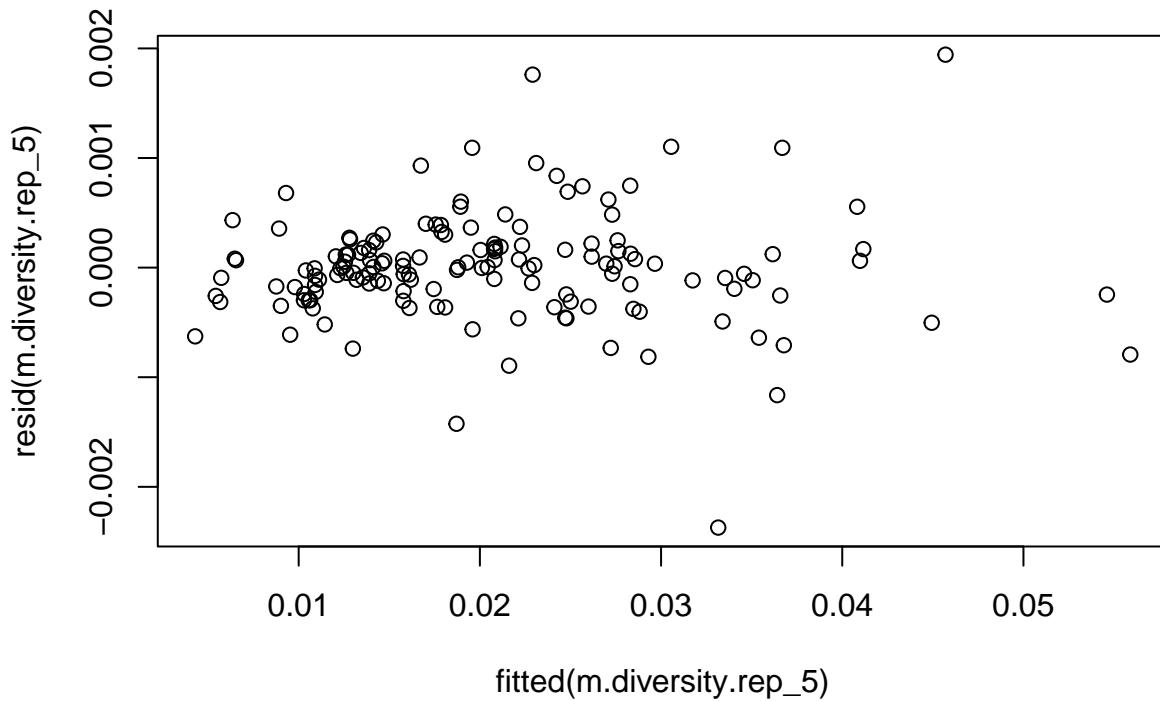
inf.lands.200kb.rep_5 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_mean))
names(inf.lands.200kb.rep_5) <- c("diversity", "theta", "rho", "tmrca")

# standardizing
inf.lands.200kb.rep_5$thetaC <- (inf.lands.200kb.rep_5$theta - mean(inf.lands.200kb.rep_5$theta)) / sd(inf.lands.200kb.rep_5$theta)
inf.lands.200kb.rep_5$tmrcaC <- (inf.lands.200kb.rep_5$tmrca - mean(inf.lands.200kb.rep_5$tmrca)) / sd(inf.lands.200kb.rep_5$tmrca)
inf.lands.200kb.rep_5$rhoC <- (inf.lands.200kb.rep_5$rho - mean(inf.lands.200kb.rep_5$rho)) / sd(inf.lands.200kb.rep_5$rho)

inf.lands.200kb.rep_5$bin <- 1:nrow(inf.lands.200kb.rep_5)

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))
```

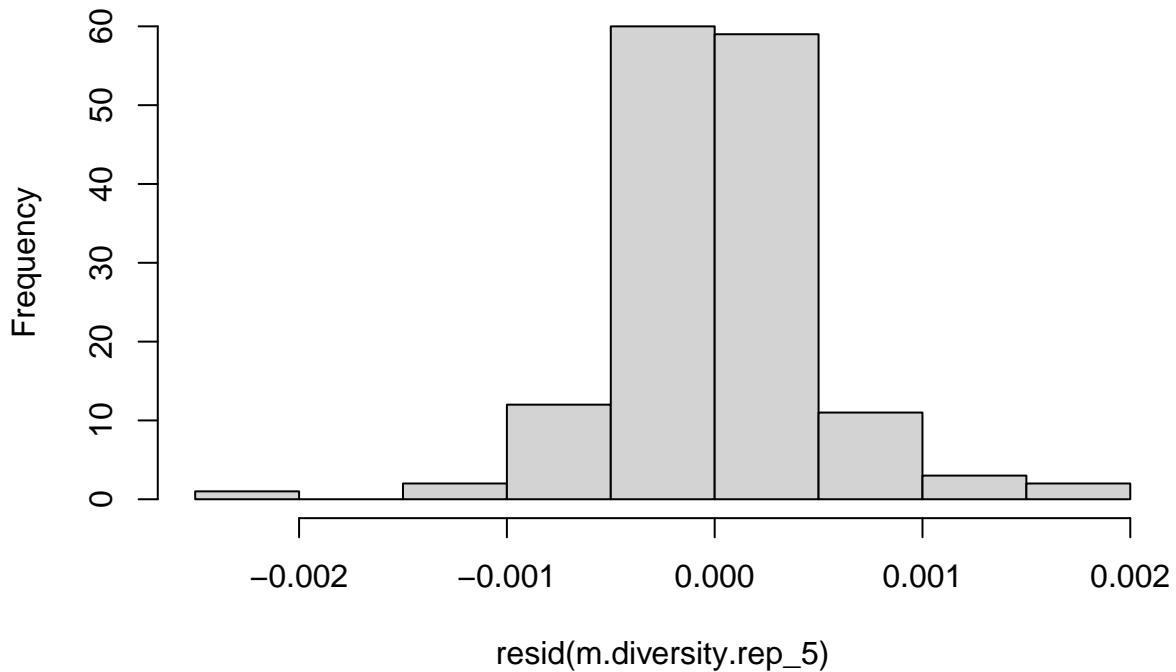


```
dwtest(m.diversity.rep_5)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_5
## DW = 1.473, p-value = 0.0003904
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_5)
```

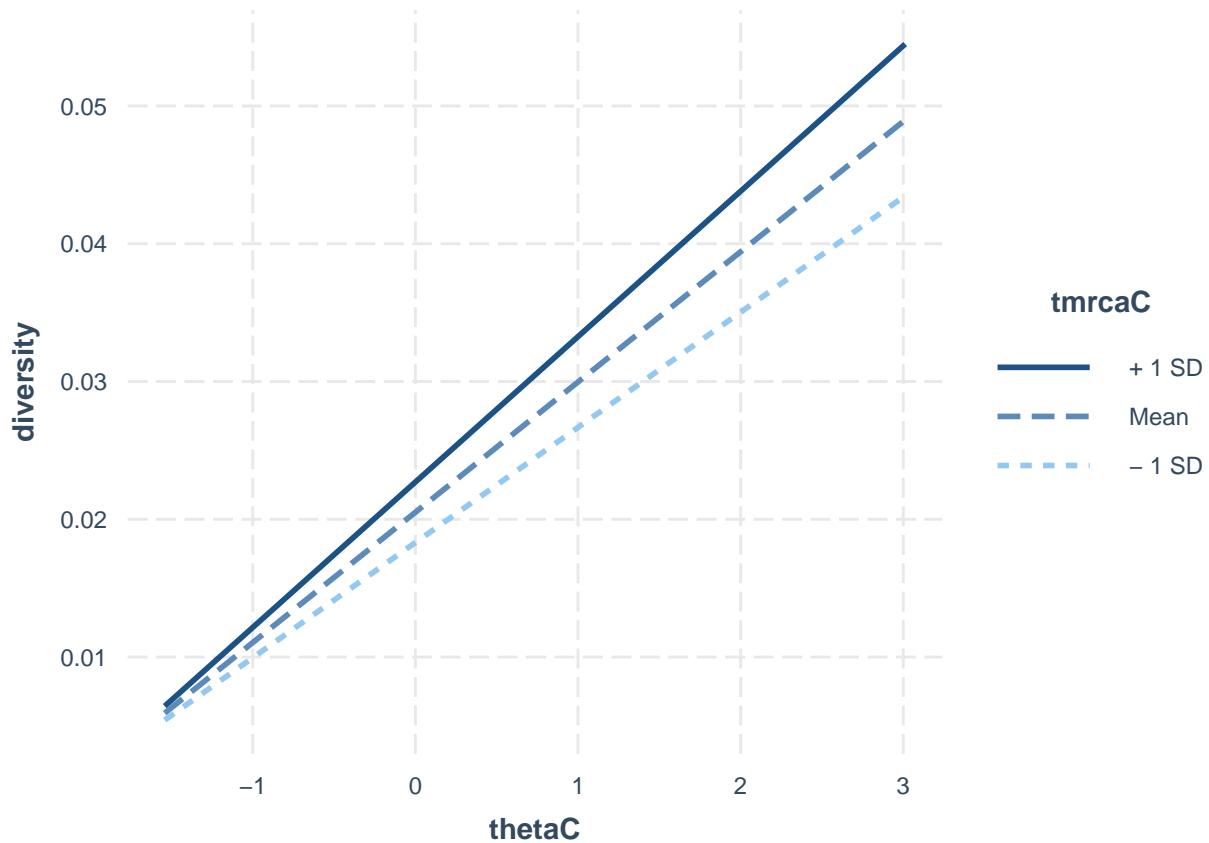
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_5
## HMC = 0.42356, p-value = 0.082
hist(resid(m.diversity.rep_5))
```

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       9.457e-03 4.173e-05 226.607 <2e-16 ***  
## rhoC        5.876e-05 4.270e-05   1.376  0.171  
## tmrcaC      2.196e-03 4.469e-05  49.148 <2e-16 ***  
## thetaC:tmrcaC 1.098e-03 3.566e-05  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) 0.020509139 5.581452e-05 367.4517 0.0000
## thetaC      0.009435033 4.516016e-05 208.9238 0.0000
## tmrcaC      0.002213089 4.538677e-05  48.7607 0.0000
## rhoC        0.000062103 4.105603e-05   1.5126 0.1325
## thetaC:tmrcaC 0.001068248 3.556924e-05 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) 0.020571613 0.0002379126 86.46709 0.0000
## thetaC       0.009596944 0.0001963043 48.88811 0.0000
## rhoC        -0.000424637 0.0001764299 -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$VarExp[3]
r2.inf.200kb[2, 5] <- anova.diversity$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))
names(inf.lands.200kb.rep_6) <- c("diversity", "theta", "rho", "tmrca")

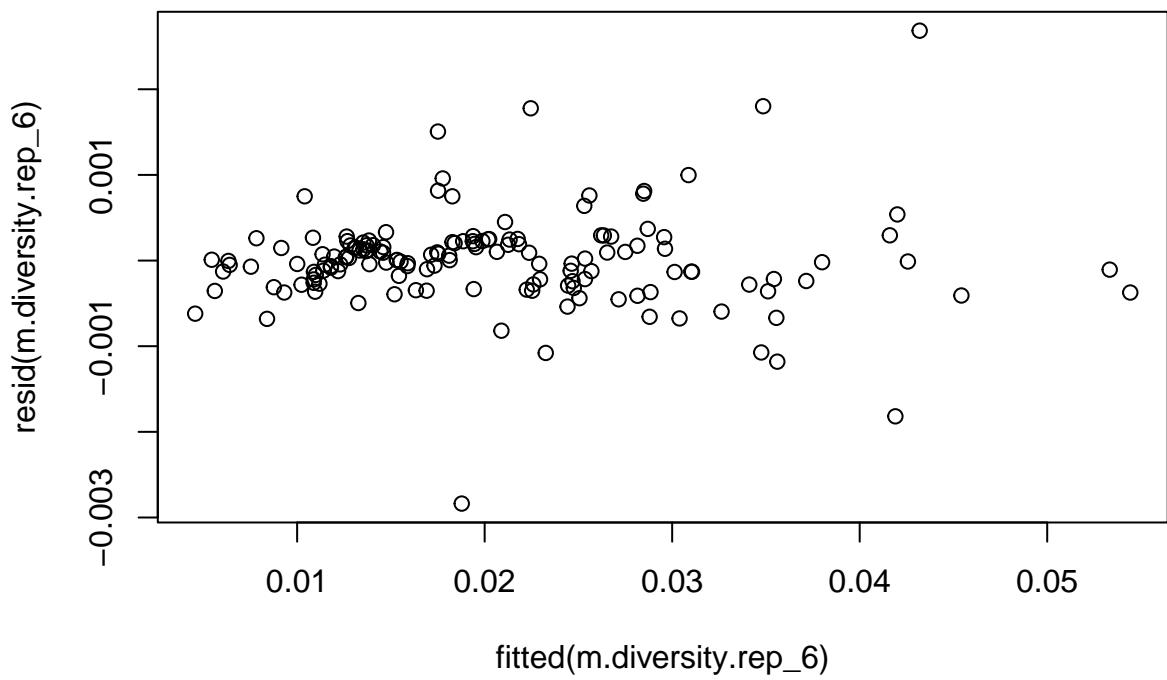
# standardizing
inf.lands.200kb.rep_6$thetaC <- (inf.lands.200kb.rep_6$theta - mean(inf.lands.200kb.rep_6$theta)) / sd(inf.lands.200kb.rep_6$theta)
inf.lands.200kb.rep_6$tmrcaC <- (inf.lands.200kb.rep_6$tmrca - mean(inf.lands.200kb.rep_6$tmrca)) / sd(inf.lands.200kb.rep_6$tmrca)
inf.lands.200kb.rep_6$rhoC <- (inf.lands.200kb.rep_6$rho - mean(inf.lands.200kb.rep_6$rho)) / sd(inf.lands.200kb.rep_6$rho)

inf.lands.200kb.rep_6$bin <- 1:nrow(inf.lands.200kb.rep_6)

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))
```

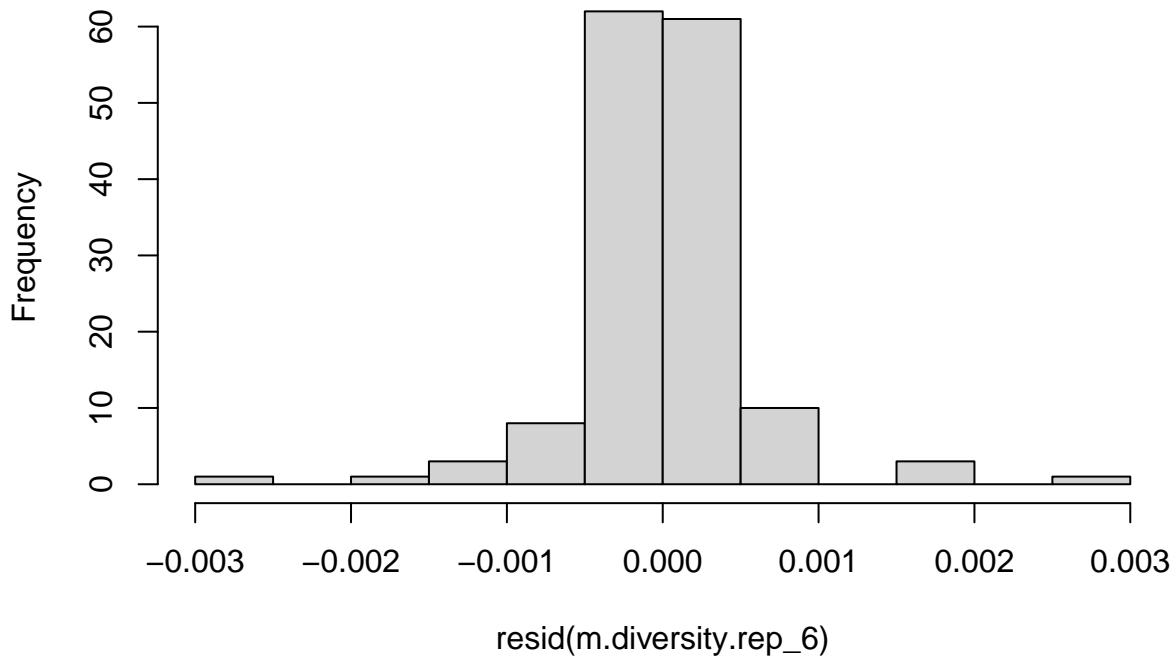


```
dwtest(m.diversity.rep_6)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_6
## DW = 1.6828, p-value = 0.01928
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_6)
```

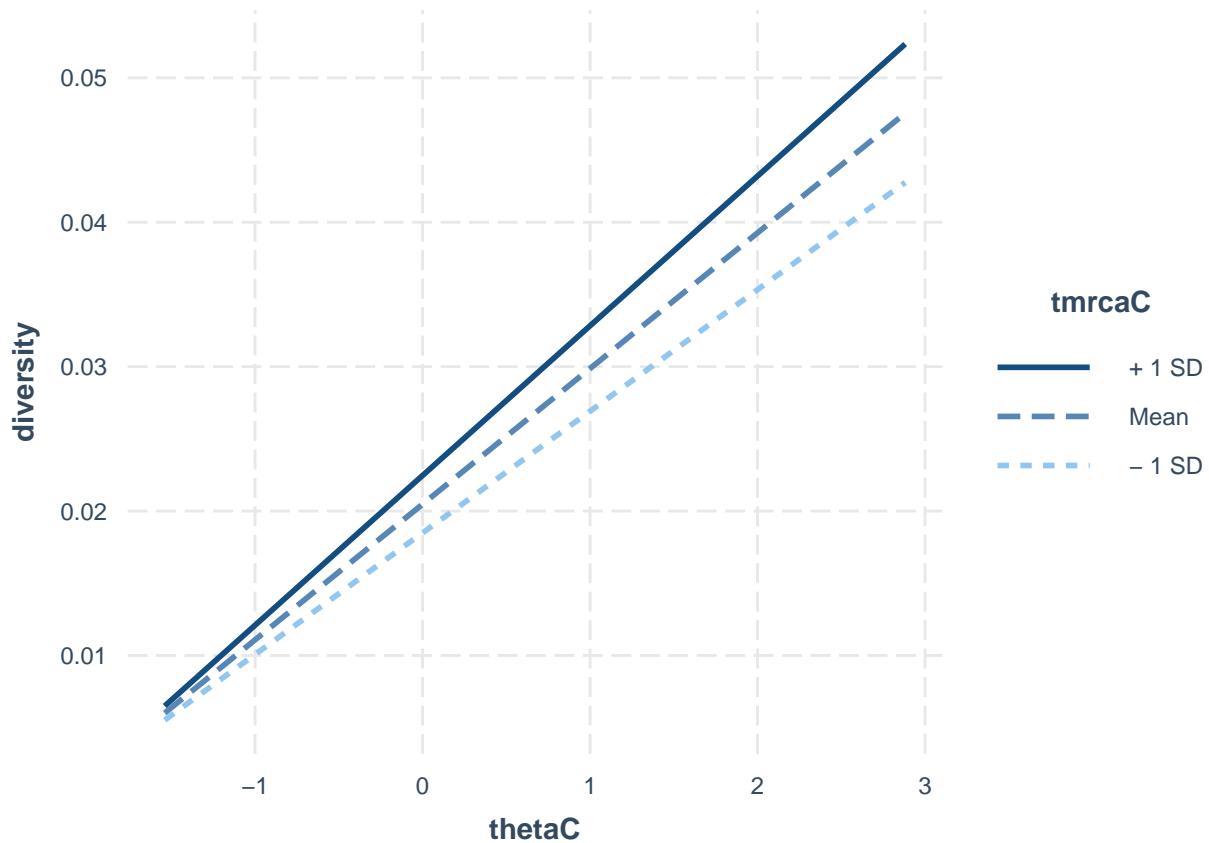
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_6
## HMC = 0.48734, p-value = 0.422
hist(resid(m.diversity.rep_6))
```

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      9.395e-03 4.679e-05 200.771 <2e-16 ***
## rhoC        3.816e-05 4.999e-05   0.763   0.447
## tmrcaC     1.986e-03 5.227e-05  37.991 <2e-16 ***
## thetaC:tmrcaC 9.736e-04 4.107e-05  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) 0.020479216 5.540422e-05 369.6328 0.0000
## thetaC      0.009400111 4.997396e-05 188.1002 0.0000
## tmrcaC      0.001993622 5.376207e-05  37.0823 0.0000
## rhoC        0.000034238 4.917537e-05   0.6962 0.4874
## thetaC:tmrcaC 0.000951305 4.232596e-05  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  4.863421008
##
## 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) 0.020583999 0.0002201173 93.51378 0e+00
## thetaC       0.009631973 0.0001737792 55.42651 0e+00
## rhoC        -0.000574664 0.0001525197 -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$VarExp[3]) * 100
r2.inf.200kb[2, 6] <- anova.diversity$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

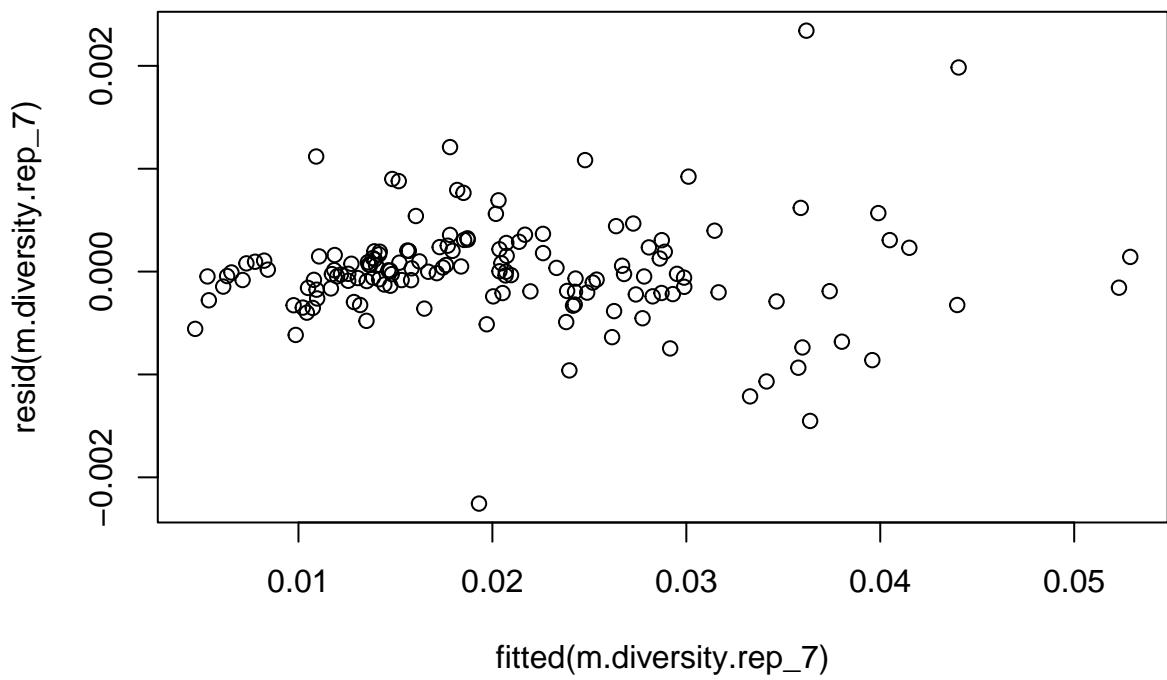
inf.lands.200kb.rep_7 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_mean))
names(inf.lands.200kb.rep_7) <- c("diversity", "theta", "rho", "tmrca")

# standardizing
inf.lands.200kb.rep_7$thetaC <- (inf.lands.200kb.rep_7$theta - mean(inf.lands.200kb.rep_7$theta)) / sd(inf.lands.200kb.rep_7$theta)
inf.lands.200kb.rep_7$tmrcaC <- (inf.lands.200kb.rep_7$tmrca - mean(inf.lands.200kb.rep_7$tmrca)) / sd(inf.lands.200kb.rep_7$tmrca)
inf.lands.200kb.rep_7$rhoC <- (inf.lands.200kb.rep_7$rho - mean(inf.lands.200kb.rep_7$rho)) / sd(inf.lands.200kb.rep_7$rho)

inf.lands.200kb.rep_7$bin <- 1:nrow(inf.lands.200kb.rep_7)

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))
```

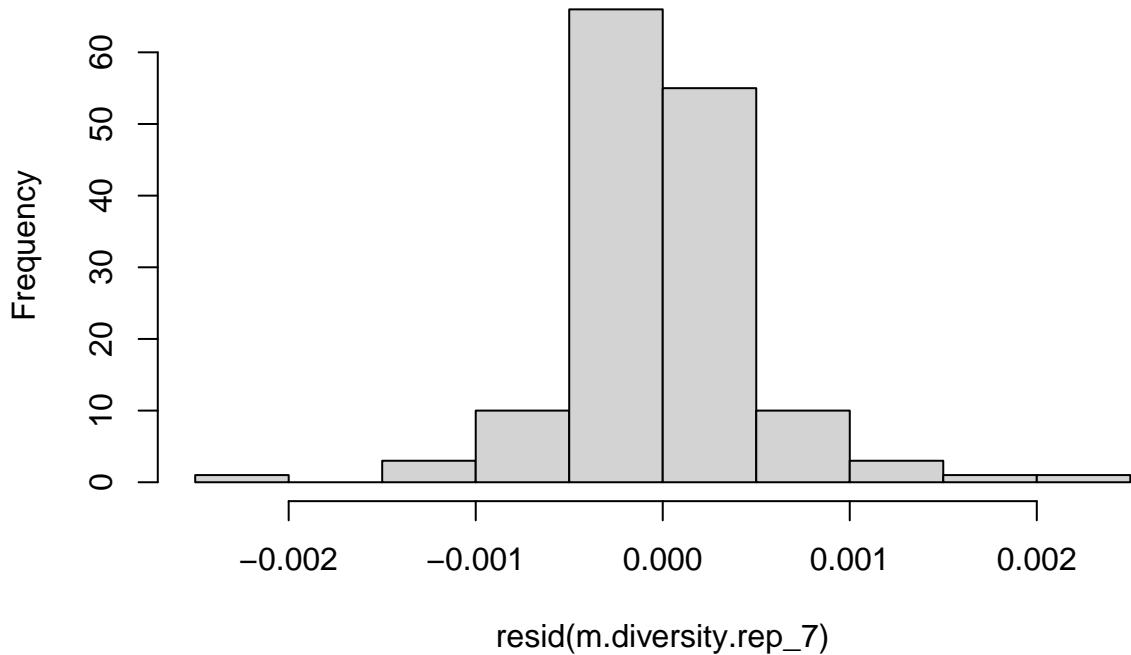


```
dwtest(m.diversity.rep_7)
```

```
##
## Durbin-Watson test
##
## data: m.diversity.rep_7
## DW = 1.6673, p-value = 0.01553
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_7)
```

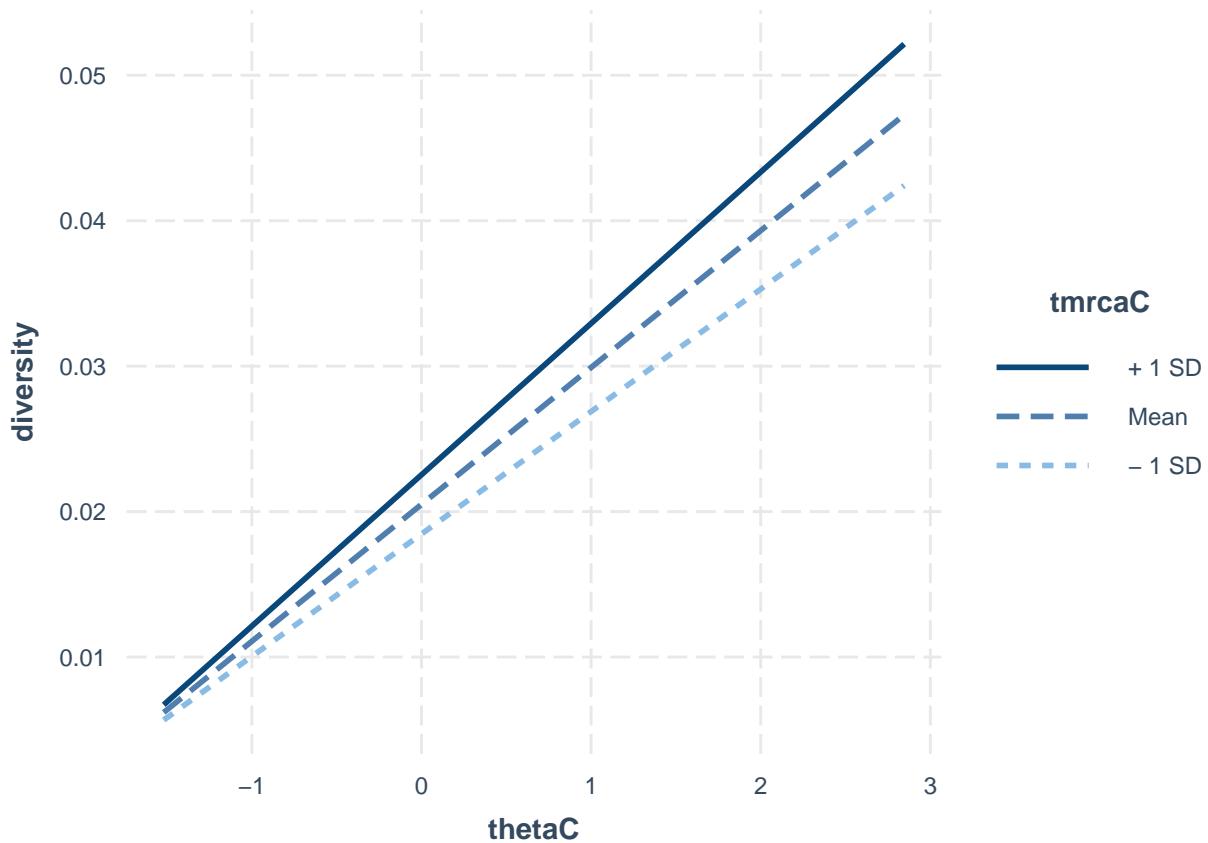
```
##
## Harrison-McCabe test
##
## data: m.diversity.rep_7
## HMC = 0.43605, p-value = 0.145
hist(resid(m.diversity.rep_7))
```

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) 2.050e-02 4.270e-05 480.036 <2e-16 ***
## thetaC      9.411e-03 4.296e-05 219.050 <2e-16 ***
## rhoC       -1.376e-05 4.337e-05 -0.317   0.752
## tmrcaC     2.036e-03 4.668e-05  43.612 <2e-16 ***
## thetaC:tmrcaC 9.940e-04 3.869e-05  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) 0.020499661 5.196105e-05 394.5197 0.0000
## thetaC      0.009426538 4.614902e-05 204.2630 0.0000
## tmrcaC      0.002022035 4.763455e-05 42.4489 0.0000
## rhoC        -0.000009825 4.282366e-05 -0.2294 0.8189
## thetaC:tmrcaC 0.000962183 3.908333e-05 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) 0.020579587 0.0002178324 94.47441 0.0000
## thetaC       0.009591179 0.0001768106 54.24550 0.0000
## rhoC        -0.000326427 0.0001562165 -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  4.0615521
##
## 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$VarExp[3]) * 100
r2.inf.200kb[2, 7] <- anova.diversity$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
## 0.98777626

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")

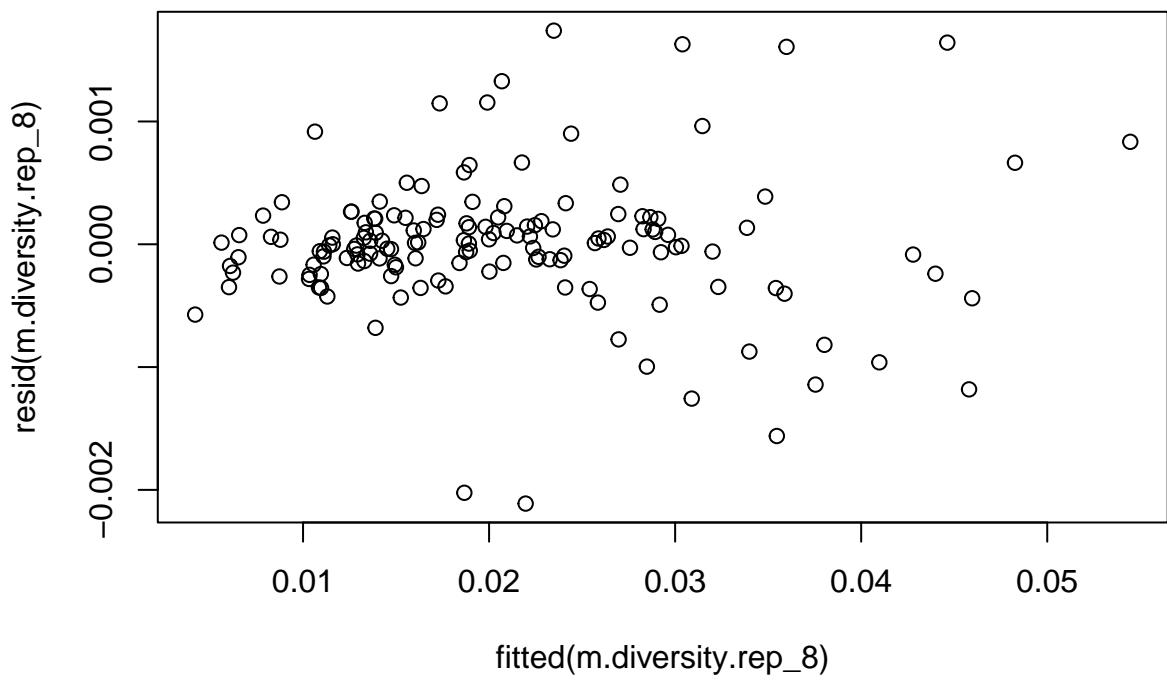
# standardizing
inf.lands.200kb.rep_8$thetaC <- (inf.lands.200kb.rep_8$theta - mean(inf.lands.200kb.rep_8$theta)) / sd(inf.lands.200kb.rep_8$theta)
inf.lands.200kb.rep_8$tmrcaC <- (inf.lands.200kb.rep_8$tmrca - mean(inf.lands.200kb.rep_8$tmrca)) / sd(inf.lands.200kb.rep_8$tmrca)
inf.lands.200kb.rep_8$rhoC <- (inf.lands.200kb.rep_8$rho - mean(inf.lands.200kb.rep_8$rho)) / sd(inf.lands.200kb.rep_8$rho)

inf.lands.200kb.rep_8$bin <- 1:nrow(inf.lands.200kb.rep_8)

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))
```

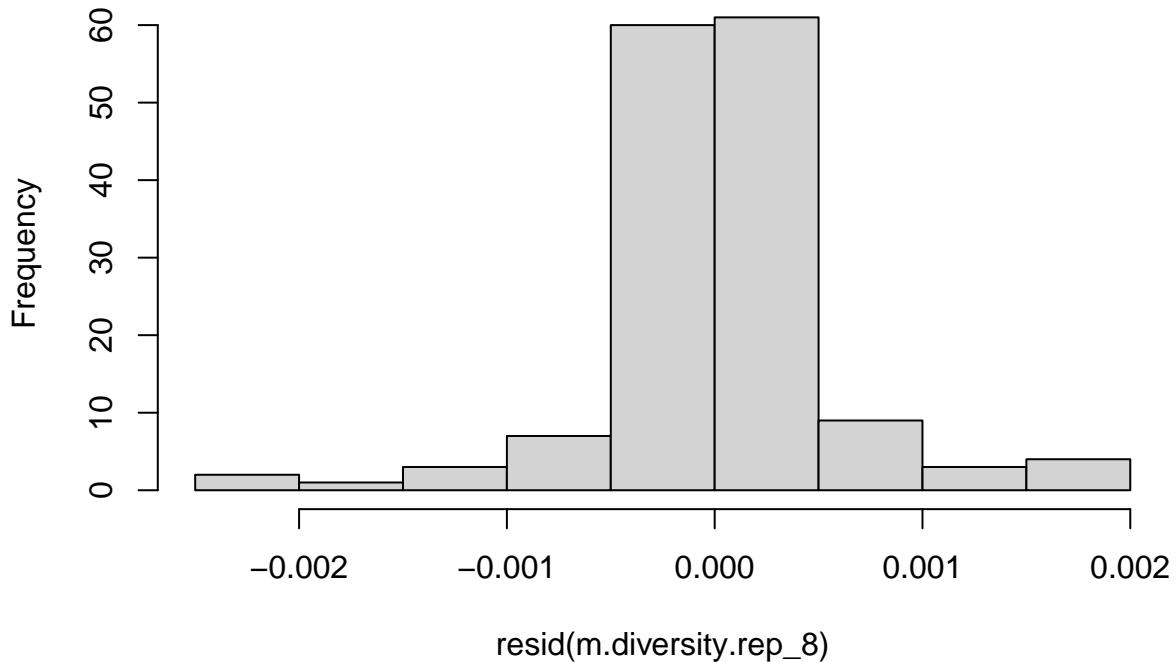


```
dwtest(m.diversity.rep_8)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_8
## DW = 1.5909, p-value = 0.004238
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_8)
```

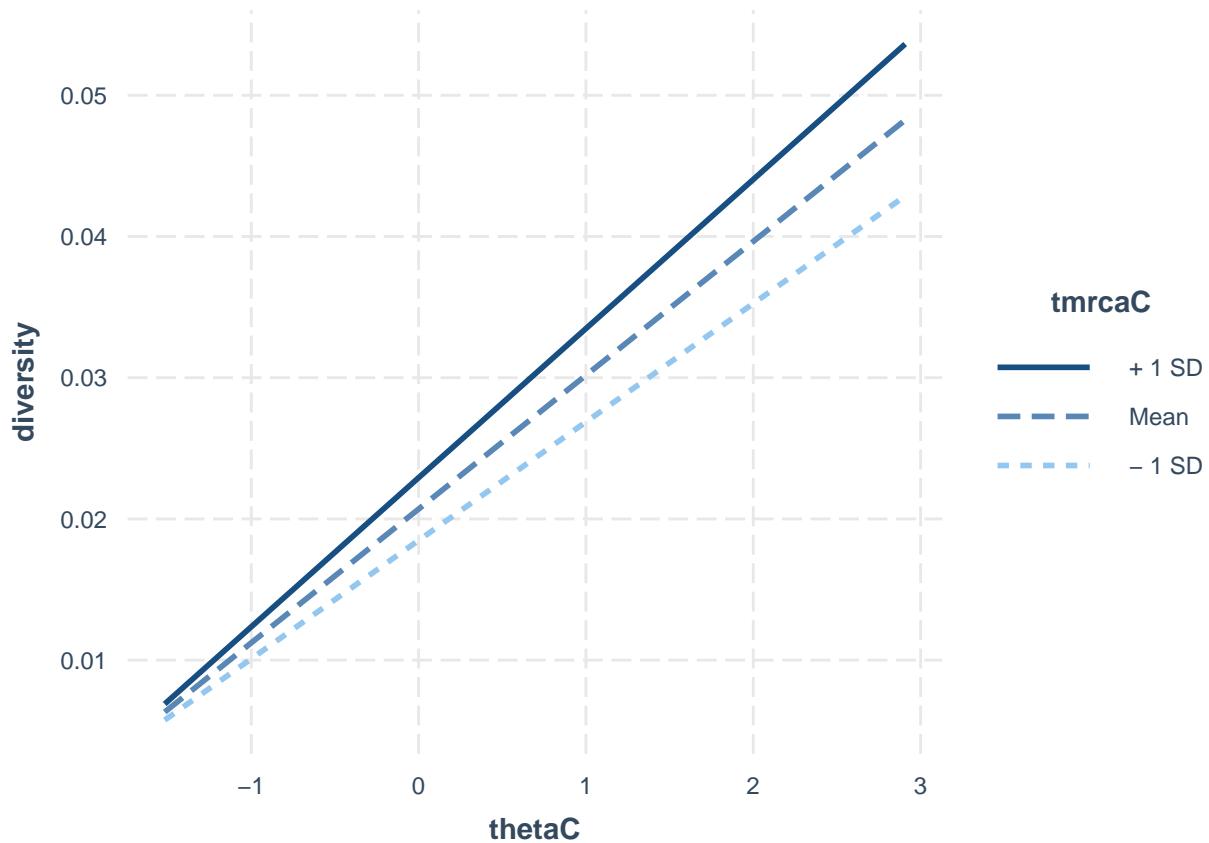
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_8
## HMC = 0.45552, p-value = 0.222
hist(resid(m.diversity.rep_8))
```

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       9.474e-03 4.651e-05 203.69 <2e-16 ***  
## rhoC        4.644e-05 4.737e-05   0.98  0.329  
## tmrcaC      2.219e-03 5.200e-05   42.68 <2e-16 ***  
## thetaC:tmrcaC 1.087e-03 4.238e-05   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) 0.020704224 5.858643e-05 353.3963 0.0000
## thetaC       0.009464706 5.036703e-05 187.9147 0.0000
## tmrcaC       0.002214528 5.337776e-05  41.4878 0.0000
## rhoC         0.000039361 4.591151e-05   0.8573 0.3927
## thetaC:tmrcaC 0.001046093 4.285662e-05  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) 0.020811072 0.0002351765 88.49129 0.0000
## thetaC       0.009729211 0.0001864523 52.18069 0.0000
## rhoC        -0.000292104 0.0001632246 -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")

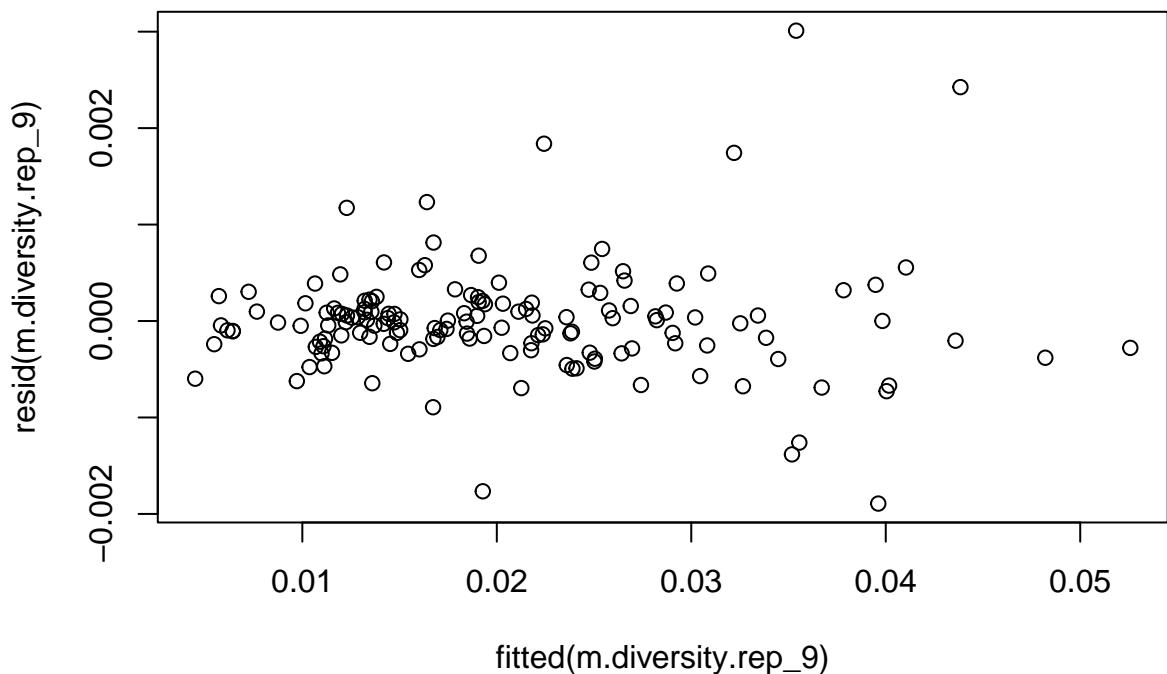
# standardizing
inf.lands.200kb.rep_9$thetaC <- (inf.lands.200kb.rep_9$theta - mean(inf.lands.200kb.rep_9$theta)) / sd(inf.lands.200kb.rep_9$theta)
inf.lands.200kb.rep_9$tmrcaC <- (inf.lands.200kb.rep_9$tmrca - mean(inf.lands.200kb.rep_9$tmrca)) / sd(inf.lands.200kb.rep_9$tmrca)
inf.lands.200kb.rep_9$rhoC <- (inf.lands.200kb.rep_9$rho - mean(inf.lands.200kb.rep_9$rho)) / sd(inf.lands.200kb.rep_9$rho)

inf.lands.200kb.rep_9$bin <- 1:nrow(inf.lands.200kb.rep_9)

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))
```

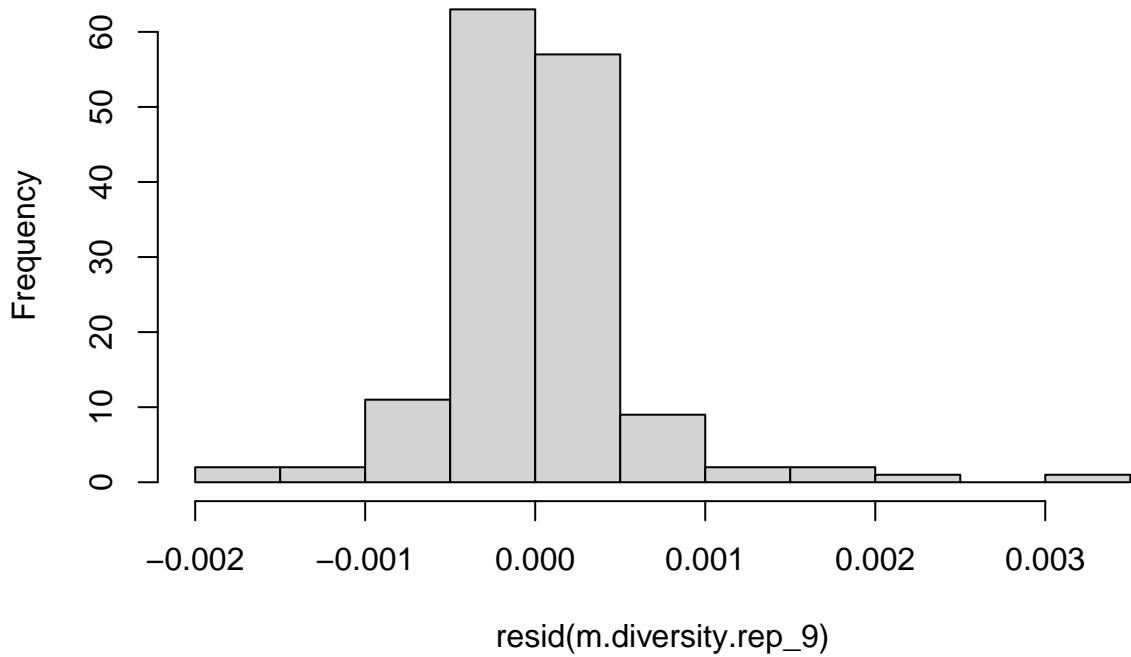


```
dwtest(m.diversity.rep_9)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_9
## DW = 1.6772, p-value = 0.01787
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_9)
```

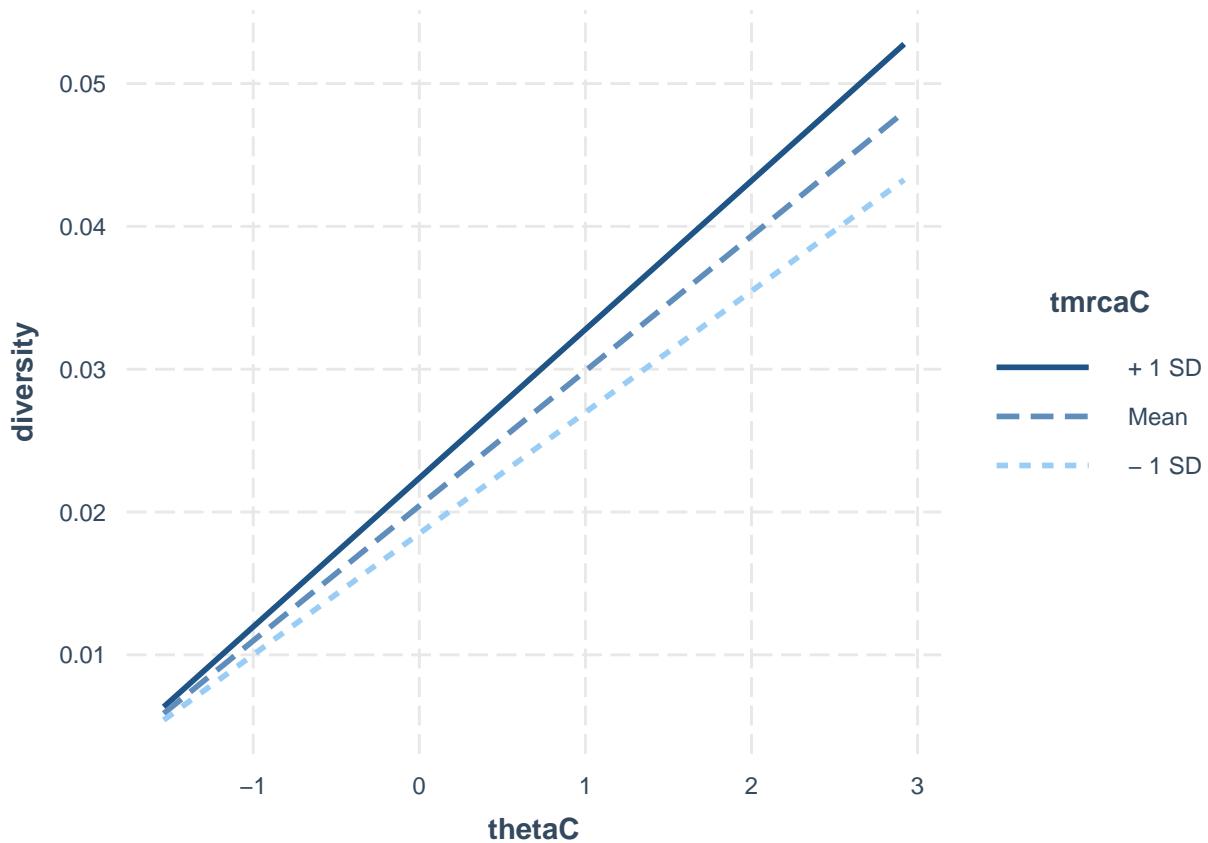
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_9
## HMC = 0.38369, p-value = 0.02
hist(resid(m.diversity.rep_9))
```

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      9.447e-03 4.835e-05 195.371 <2e-16 ***
## rhoC        8.234e-05 5.010e-05   1.643   0.102
## tmrcaC     1.941e-03 5.332e-05  36.402 <2e-16 ***
## thetaC:tmrcaC 9.618e-04 4.467e-05  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) 0.020439448 5.911181e-05 345.7760 0.0000
## thetaC       0.009454656 5.195185e-05 181.9888 0.0000
## tmrcaC       0.001951847 5.500071e-05  35.4877 0.0000
## rhoC         0.000106064 4.916687e-05   2.1572 0.0326
## thetaC:tmrcaC 0.000919495 4.629566e-05 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) 0.020537857 0.0002229473 92.11977 0.0000
## thetaC       0.009582837 0.0001677256 57.13401 0.0000
## rhoC        -0.000329055 0.0001450777 -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

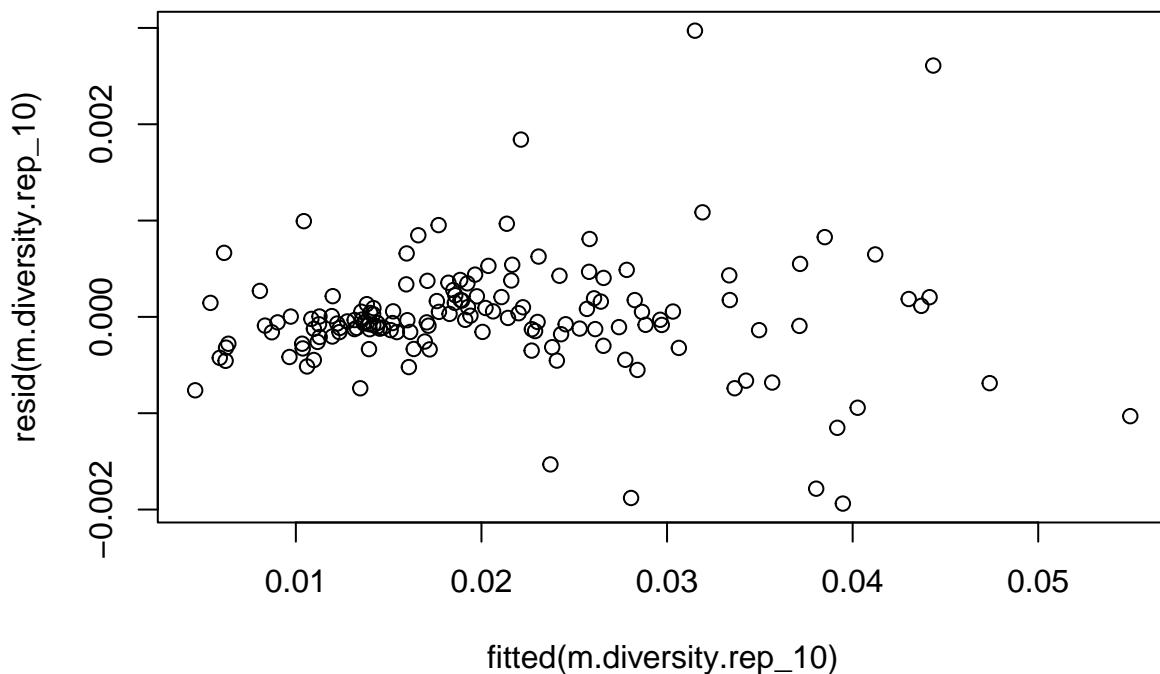
inf.lands.200kb.rep_10 <- as.data.frame(cbind(pi.200kb$avg, theta.200kb$sample_mean, rho.200kb$sample_mean))
names(inf.lands.200kb.rep_10) <- c("diversity", "theta", "rho", "tmrca")

# standardizing
inf.lands.200kb.rep_10$thetaC <- (inf.lands.200kb.rep_10$theta - mean(inf.lands.200kb.rep_10$theta)) / sd(inf.lands.200kb.rep_10$theta)
inf.lands.200kb.rep_10$tmrcaC <- (inf.lands.200kb.rep_10$tmrca - mean(inf.lands.200kb.rep_10$tmrca)) / sd(inf.lands.200kb.rep_10$tmrca)
inf.lands.200kb.rep_10$rhoC <- (inf.lands.200kb.rep_10$rho - mean(inf.lands.200kb.rep_10$rho)) / sd(inf.lands.200kb.rep_10$rho)

inf.lands.200kb.rep_10$bin <- 1:nrow(inf.lands.200kb.rep_10)

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))
```

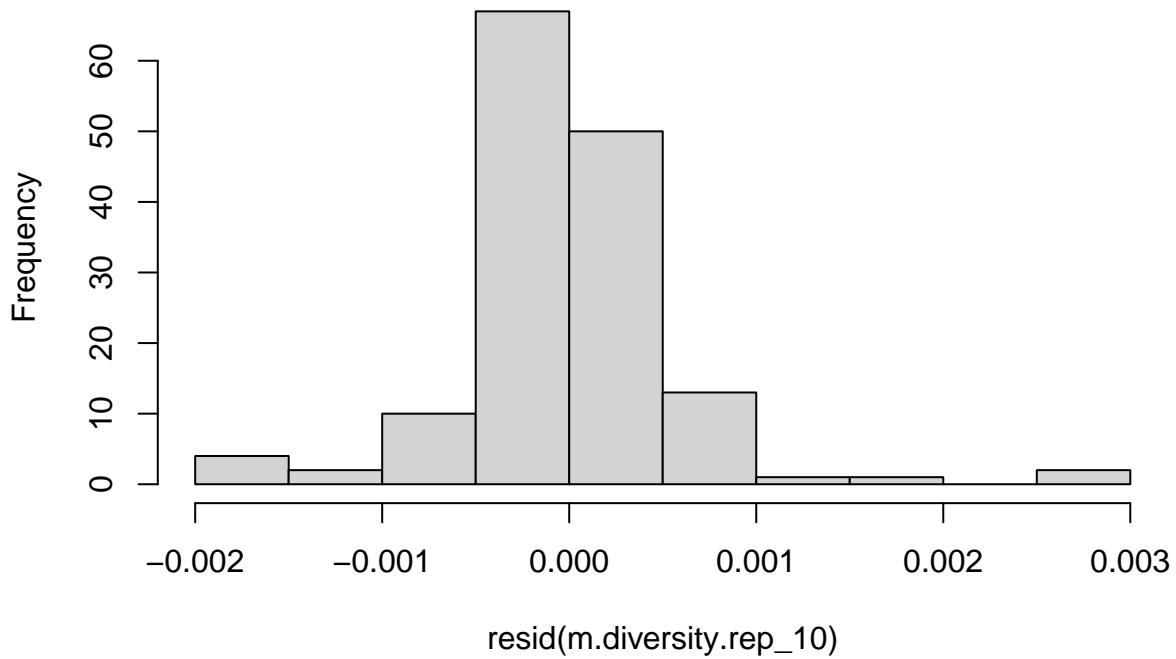


```
dwtest(m.diversity.rep_10)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_10
## DW = 1.777, p-value = 0.06912
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_10)
```

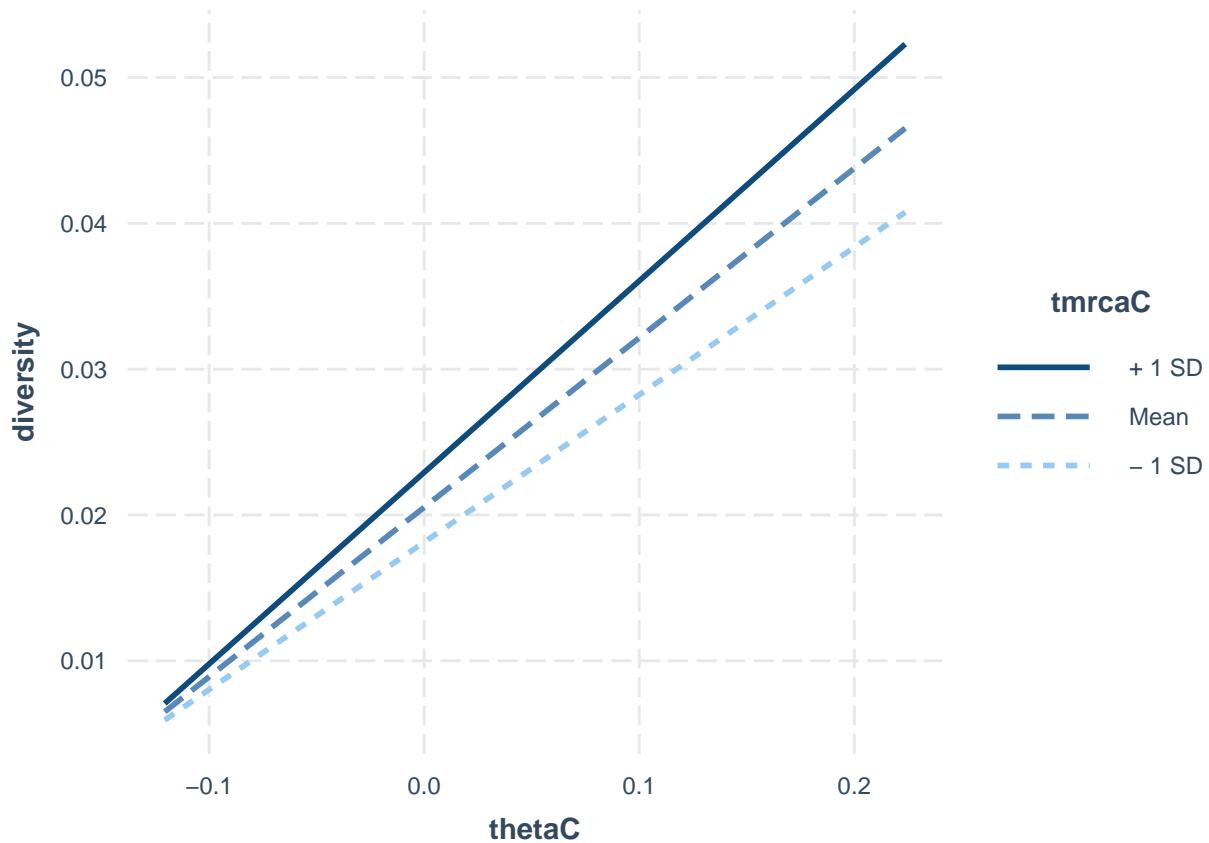
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_10
## HMC = 0.6193, p-value = 0.983
hist(resid(m.diversity.rep_10))
```

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      1.163e-01 6.297e-04 184.625 <2e-16 ***
## rhoC        2.164e-03 7.654e-03   0.283   0.778
## tmrcaC      2.391e-03 6.173e-05  38.738 <2e-16 ***
## thetaC:tmrcaC 1.511e-02 6.911e-04  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) 0.020495842 4.988827e-05 410.8349 0.0000
## thetaC      0.009311061 4.778895e-05 194.8371 0.0000
## tmrcaC     0.002162321 5.209260e-05  41.5092 0.0000
## rhoC        0.000016367 4.723723e-05    0.3465 0.7295
## thetaC:tmrcaC 0.001058085 4.650120e-05  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) 0.02071315 0.000234587 88.29634 0.0000
## thetaC       0.12318309 0.002267579 54.32360 0.0000
## rhoC        -0.05434884 0.022943808 -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$VarExp[3] * 100
r2.inf.200kb[2, 10] <- anova.diversity$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))

```

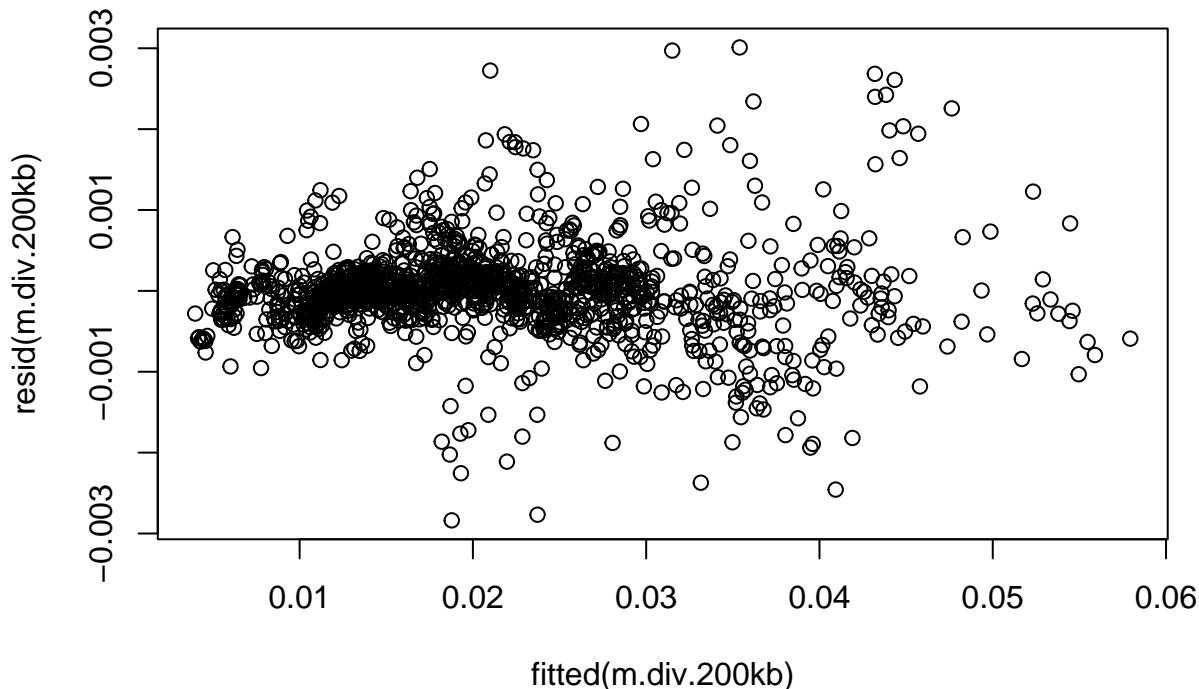
2.2.11 all replicates:

```

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)

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

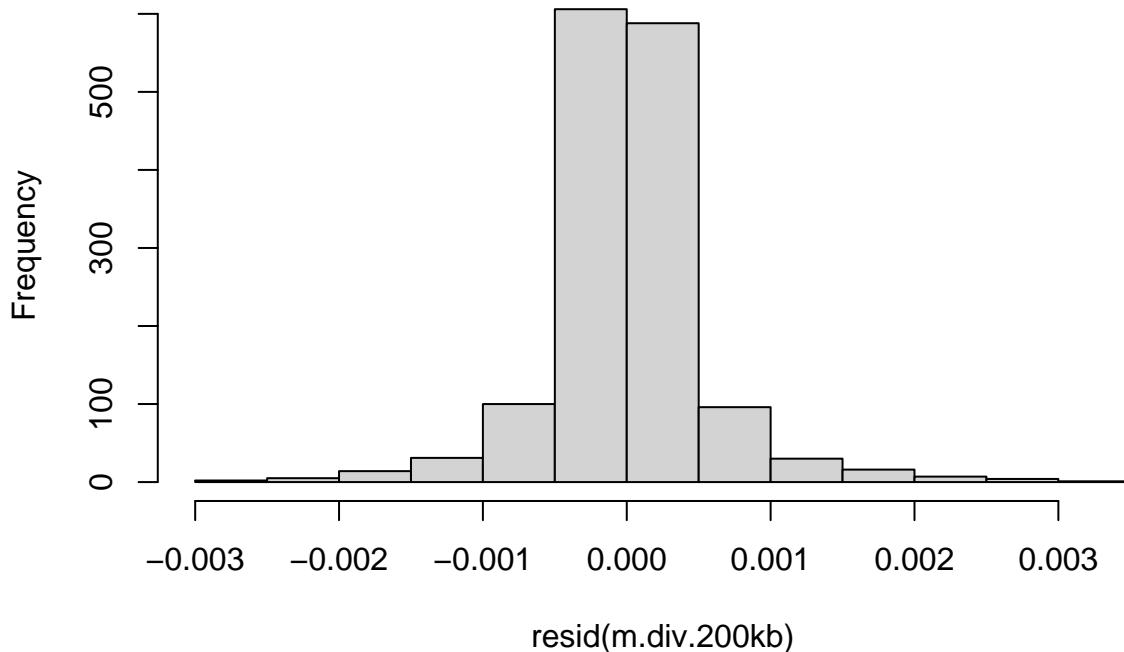
hmctest(m.div.200kb)

##
##  Harrison-McCabe test
##
##  data:  m.div.200kb
##  HMC = 0.48172, p-value = 0.159

```

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

Histogram of resid(m.div.200kb)



```
m.div.200kb.2 <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC)*as.factor(Replicate), data = inf.lands.200kb.all)
m.div.200kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC + thetaC:tmrcaC + thetaC:rhoC + rhoC:tmrcaC)*as.factor(Replicate), data = inf.lands.200kb.all)
```

```
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 = diversity ~ (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:
## (Intercept)          Estimate Std. Error t value Pr(>|t|)
## (Intercept)          2.049e-02  4.600e-05 445.509 < 2e-16 ***
## thetaC              9.304e-03  4.617e-05 201.508 < 2e-16 ***
## rhoC               1.063e-05  4.748e-05   0.224  0.82295
## tmrcaC              2.164e-03  5.175e-05  41.809 < 2e-16 ***
## as.factor(Replicate) -3.060e-05  6.507e-05  -0.470  0.63829
```

```

## as.factor(Replicate)3      1.025e-04  6.504e-05  1.576  0.11529
## as.factor(Replicate)4      1.385e-04  6.486e-05  2.136  0.03288 *
## as.factor(Replicate)5      1.180e-05  6.487e-05  0.182  0.85564
## as.factor(Replicate)6      -1.818e-05 6.496e-05 -0.280  0.77959
## as.factor(Replicate)7      2.154e-06  6.491e-05  0.033  0.97353
## as.factor(Replicate)8      2.053e-04  6.496e-05  3.160  0.00161 **
## as.factor(Replicate)9      -5.912e-05 6.496e-05 -0.910  0.36292
## as.factor(Replicate)10     3.230e-05  6.526e-05  0.495  0.62077
## thetaC:tmrcaC             1.069e-03  4.547e-05 23.507 < 2e-16 ***
## thetaC:as.factor(Replicate)2 -8.783e-06 6.535e-05 -0.134  0.89311
## thetaC:as.factor(Replicate)3 1.452e-04  6.542e-05  2.220  0.02656 *
## thetaC:as.factor(Replicate)4 3.223e-04  6.527e-05  4.937  8.86e-07 ***
## thetaC:as.factor(Replicate)5 1.526e-04  6.523e-05  2.339  0.01948 *
## thetaC:as.factor(Replicate)6 9.049e-05  6.524e-05  1.387  0.16564
## thetaC:as.factor(Replicate)7 1.069e-04  6.523e-05  1.639  0.10142
## thetaC:as.factor(Replicate)8 1.692e-04  6.527e-05  2.593  0.00962 **
## thetaC:as.factor(Replicate)9 1.425e-04  6.540e-05  2.178  0.02953 *
## thetaC:as.factor(Replicate)10 1.069e-01  5.824e-04 183.629 < 2e-16 ***
## rhoC:as.factor(Replicate)2 -4.148e-06 6.658e-05 -0.062  0.95033
## rhoC:as.factor(Replicate)3 1.476e-05  6.650e-05  0.222  0.82435
## rhoC:as.factor(Replicate)4 3.385e-05  6.678e-05  0.507  0.61229
## rhoC:as.factor(Replicate)5 4.813e-05  6.691e-05  0.719  0.47204
## rhoC:as.factor(Replicate)6 2.753e-05  6.841e-05  0.403  0.68736
## rhoC:as.factor(Replicate)7 -2.439e-05 6.647e-05 -0.367  0.71376
## rhoC:as.factor(Replicate)8 3.581e-05  6.681e-05  0.536  0.59199
## rhoC:as.factor(Replicate)9 7.171e-05  6.751e-05  1.062  0.28831
## rhoC:as.factor(Replicate)10 2.154e-03  7.058e-03  0.305  0.76028
## tmrcaC:as.factor(Replicate)2 -4.493e-05 7.521e-05 -0.597  0.55029
## tmrcaC:as.factor(Replicate)3 -1.814e-05 7.052e-05 -0.257  0.79700
## tmrcaC:as.factor(Replicate)4 9.139e-05  7.132e-05  1.281  0.20024
## tmrcaC:as.factor(Replicate)5 3.271e-05  7.150e-05  0.457  0.64743
## tmrcaC:as.factor(Replicate)6 -1.778e-04 7.300e-05 -2.435  0.01501 *
## tmrcaC:as.factor(Replicate)7 -1.279e-04 7.200e-05 -1.776  0.07596 .
## tmrcaC:as.factor(Replicate)8 5.539e-05  7.307e-05  0.758  0.44851
## tmrcaC:as.factor(Replicate)9 -2.227e-04 7.271e-05 -3.062  0.00224 **
## tmrcaC:as.factor(Replicate)10 2.275e-04  7.693e-05  2.958  0.00315 **
## thetaC:tmrcaC:as.factor(Replicate)2 -4.079e-05 6.386e-05 -0.639  0.52302
## thetaC:tmrcaC:as.factor(Replicate)3 -7.057e-05 5.897e-05 -1.197  0.23161
## thetaC:tmrcaC:as.factor(Replicate)4 5.089e-05  6.504e-05  0.783  0.43403
## thetaC:tmrcaC:as.factor(Replicate)5 2.890e-05  6.015e-05  0.480  0.63099
## thetaC:tmrcaC:as.factor(Replicate)6 -9.539e-05 6.086e-05 -1.567  0.11722
## thetaC:tmrcaC:as.factor(Replicate)7 -7.496e-05 6.156e-05 -1.218  0.22353
## thetaC:tmrcaC:as.factor(Replicate)8 1.788e-05  6.193e-05  0.289  0.77290
## thetaC:tmrcaC:as.factor(Replicate)9 -1.071e-04 6.244e-05 -1.716  0.08639 .
## thetaC:tmrcaC:as.factor(Replicate)10 1.404e-02  6.389e-04 21.982 < 2e-16 ***
## ---
## 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

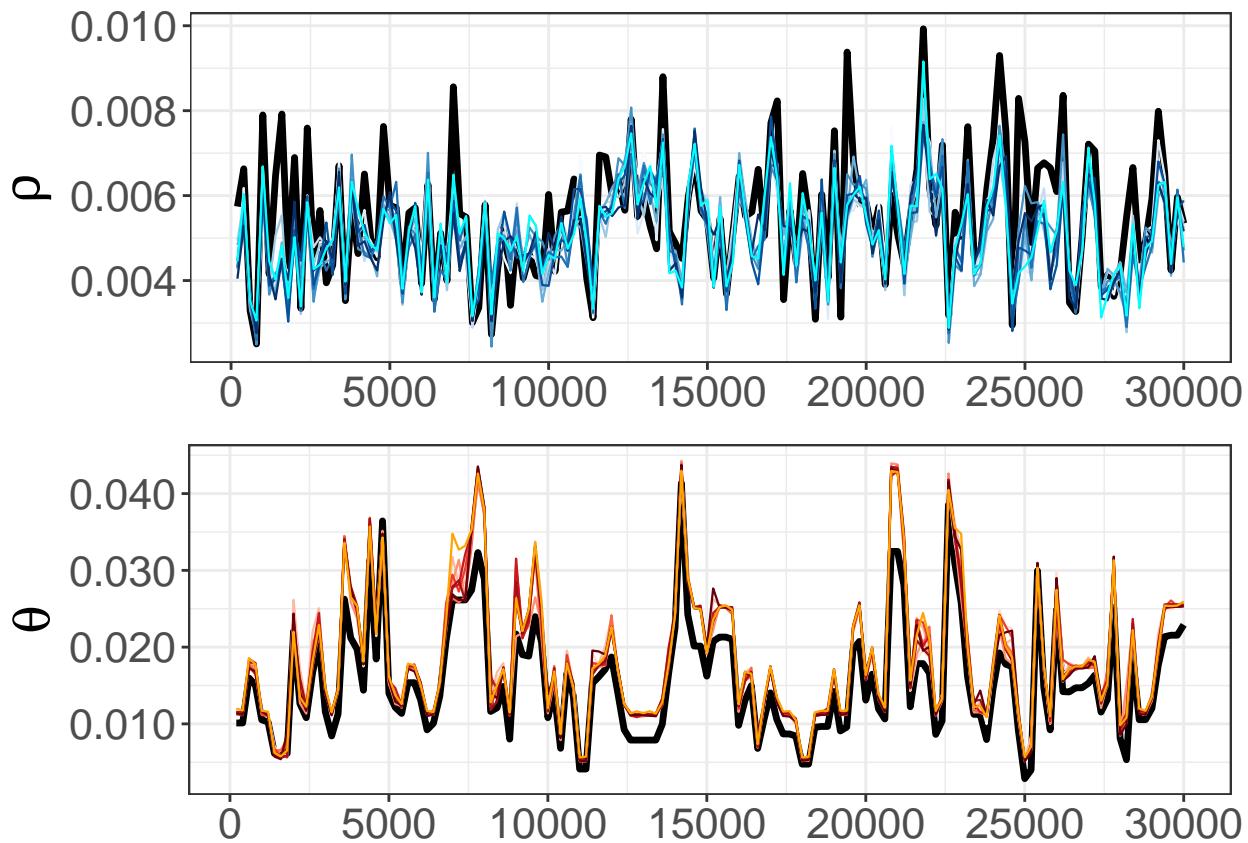
names(rho.plot) <- c("bin", "sim", reps)
molten.rho <- melt(rho.plot, id.vars = "bin")
rho.map.200kb <- ggplot(data = molten.rho, aes(x = bin * 200, y = value, colour = variable)) + theme_bw()
rho.map.200kb <- rho.map.200kb + geom_line(data = molten.rho, aes(size = variable)) + scale_size_manual()
rho.map.200kb <- 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_manual()
theta.map.200kb <- theta.map.200kb + scale_color_manual(values = c("black", brewer.pal(n = 9, name = "Reds"))
theta.map.200kb <- theta.map.200kb + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
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(size = 20))

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$sample_mean, tmrca.1Mb$sample_mean))  

names(inf.lands.1Mb.rep_1) <- c("diversity", "theta", "rho", "tmrca")  

# standardizing  

inf.lands.1Mb.rep_1$thetaC <- (inf.lands.1Mb.rep_1$theta - mean(inf.lands.1Mb.rep_1$theta)) / sd(inf.lands.1Mb.rep_1$theta)  

inf.lands.1Mb.rep_1$tmrcaC <- (inf.lands.1Mb.rep_1$tmrca - mean(inf.lands.1Mb.rep_1$tmrca)) / sd(inf.lands.1Mb.rep_1$tmrca)  

inf.lands.1Mb.rep_1$rhoC <- (inf.lands.1Mb.rep_1$rho - mean(inf.lands.1Mb.rep_1$rho)) / sd(inf.lands.1Mb.rep_1$rho)  

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

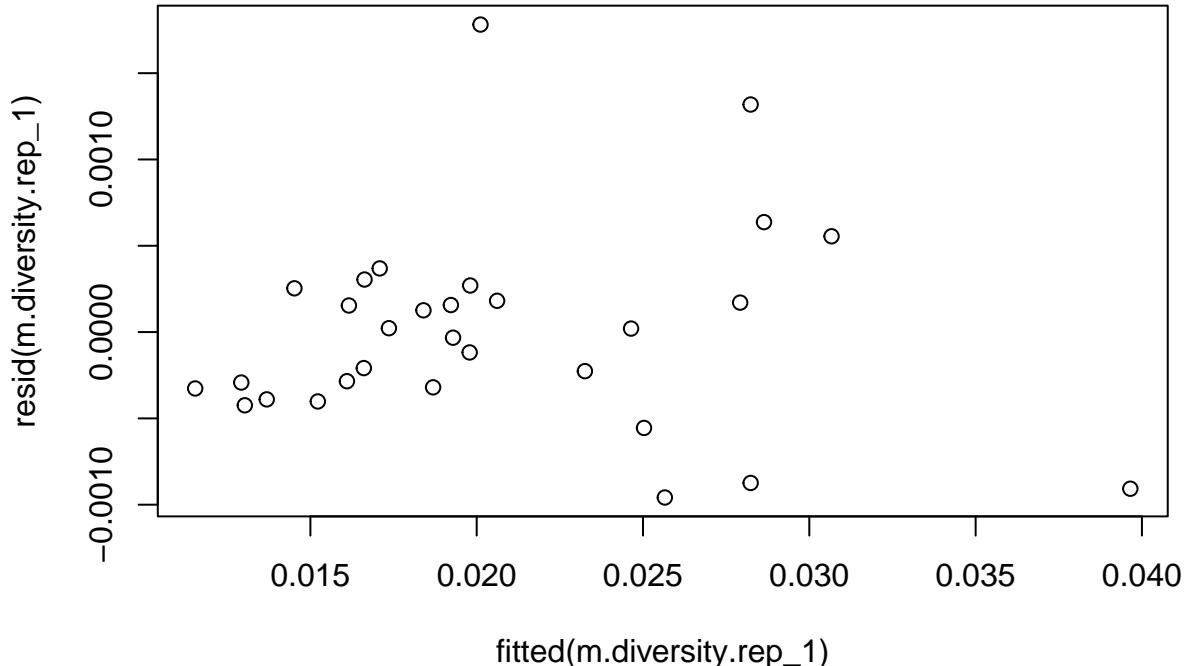
inf.lands.1Mb.rep_1$Replicate <- 1  

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

```

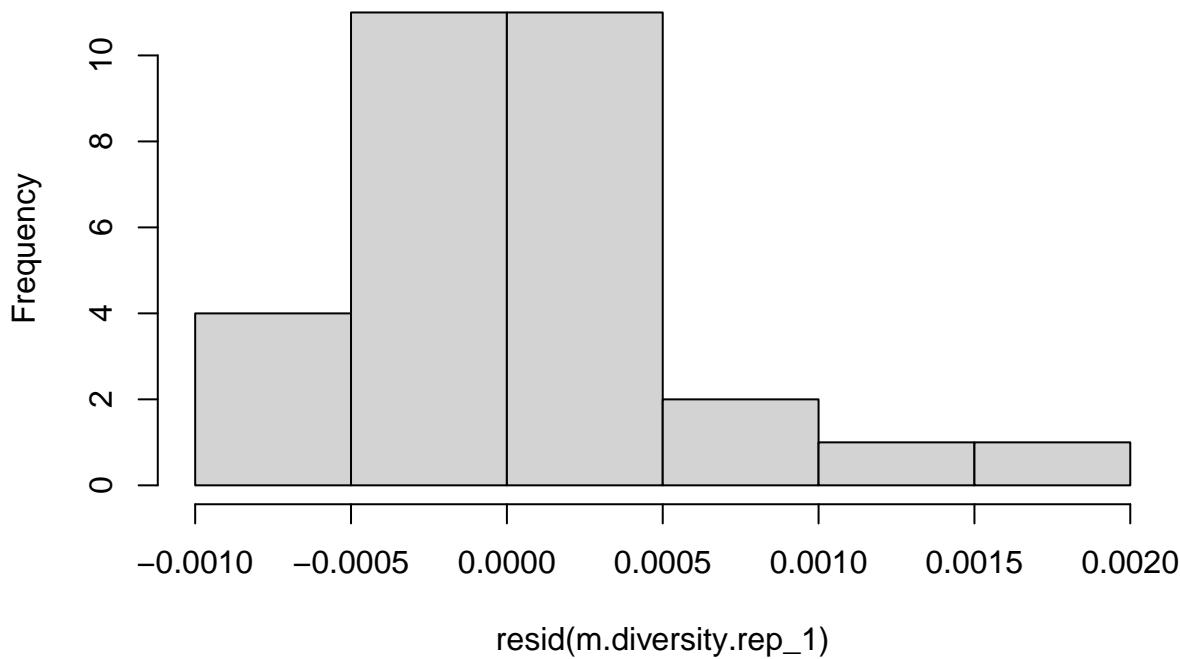
```

hmctest(m.diversity.rep_1)

##
##  Harrison-McCabe test
##
## data: m.diversity.rep_1
## HMC = 0.46606, p-value = 0.405
hist(resid(m.diversity.rep_1))

```

Histogram of resid(m.diversity.rep_1)



```

summary(m.diversity.rep_1)

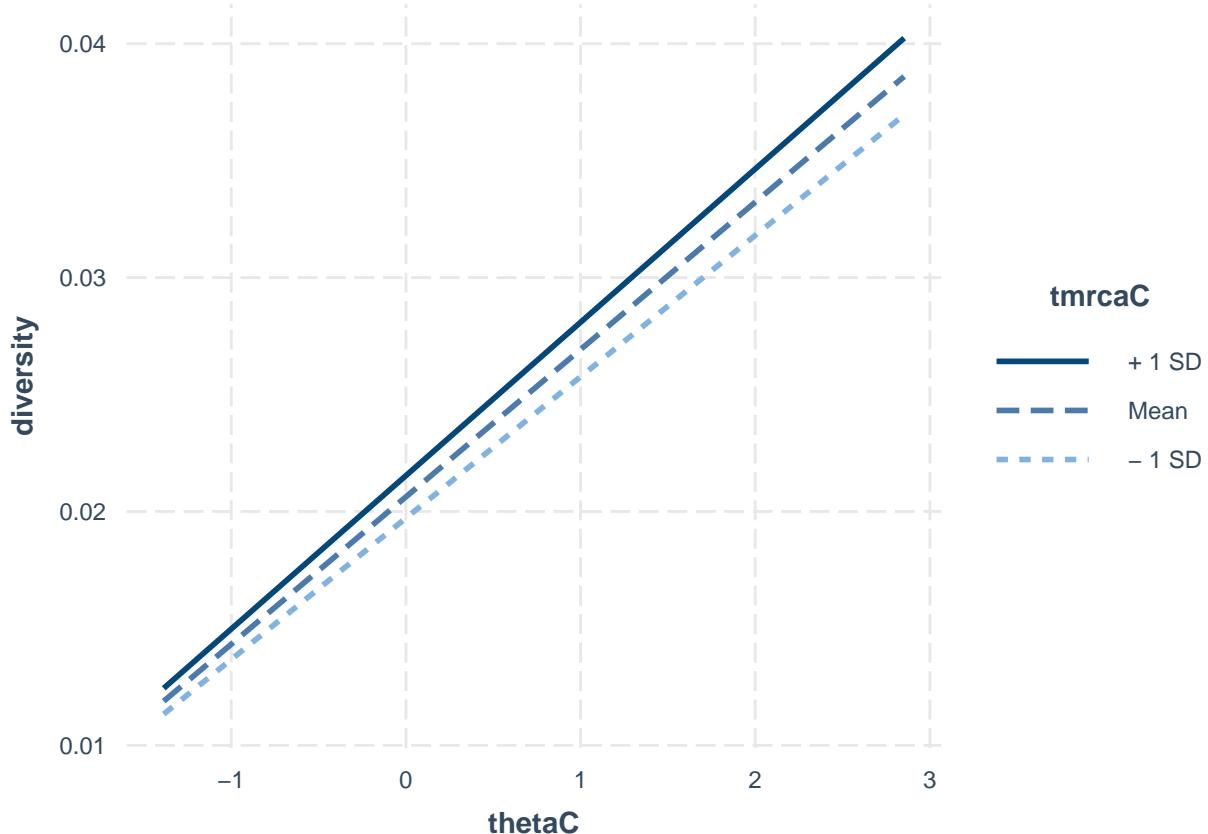
##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC * tmrcaC,
##      data = inf.lands.1Mb.rep_1)
##
## Residuals:
##       Min         1Q     Median         3Q        Max
## -0.0009583 -0.0003251 -0.0000061  0.0002355  0.0017816
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.062e-02 1.150e-04 179.401 < 2e-16 ***
## thetaC       6.300e-03 1.219e-04   51.697 < 2e-16 ***
## rhoC        1.166e-06 1.322e-04    0.009   0.993
## tmrcaC      9.079e-04 1.622e-04    5.596 8.05e-06 ***
## thetaC:tmrcaC 2.569e-04 1.562e-04    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) 0.020624763 0.0001170758 176.16584 0.0000
## thetaC      0.006298915 0.0001222271  51.53453 0.0000

```

```

## tmrcaC      0.000914620 0.0001612620  5.67164  0.0000
## rhoC        0.000000297 0.0001325695  0.00224  0.9982
## thetaC:tmrcaC 0.000263026 0.0001559369  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) 0.020608957 0.0001241807 165.95945 0.0000
## thetaC      0.006308014 0.0001520989  41.47310 0.0000
## rhoC       -0.000358265 0.0001456336  -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")

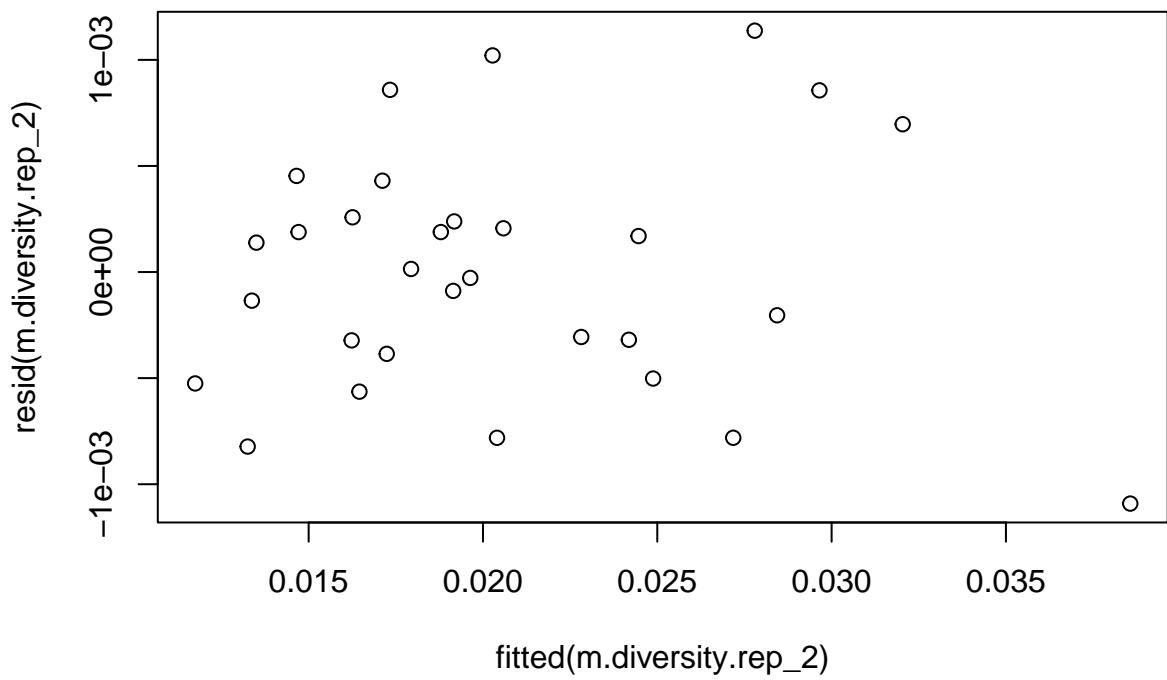
# standardizing
inf.lands.1Mb.rep_2$thetaC <- (inf.lands.1Mb.rep_2$theta - mean(inf.lands.1Mb.rep_2$theta)) / sd(inf.lands.1Mb.rep_2)
inf.lands.1Mb.rep_2$tmrcaC <- (inf.lands.1Mb.rep_2$tmrca - mean(inf.lands.1Mb.rep_2$tmrca)) / sd(inf.lands.1Mb.rep_2)
inf.lands.1Mb.rep_2$rhoC <- (inf.lands.1Mb.rep_2$rho - mean(inf.lands.1Mb.rep_2$rho)) / sd(inf.lands.1Mb.rep_2)

inf.lands.1Mb.rep_2$bin <- 1:nrow(inf.lands.1Mb.rep_2)

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))

```

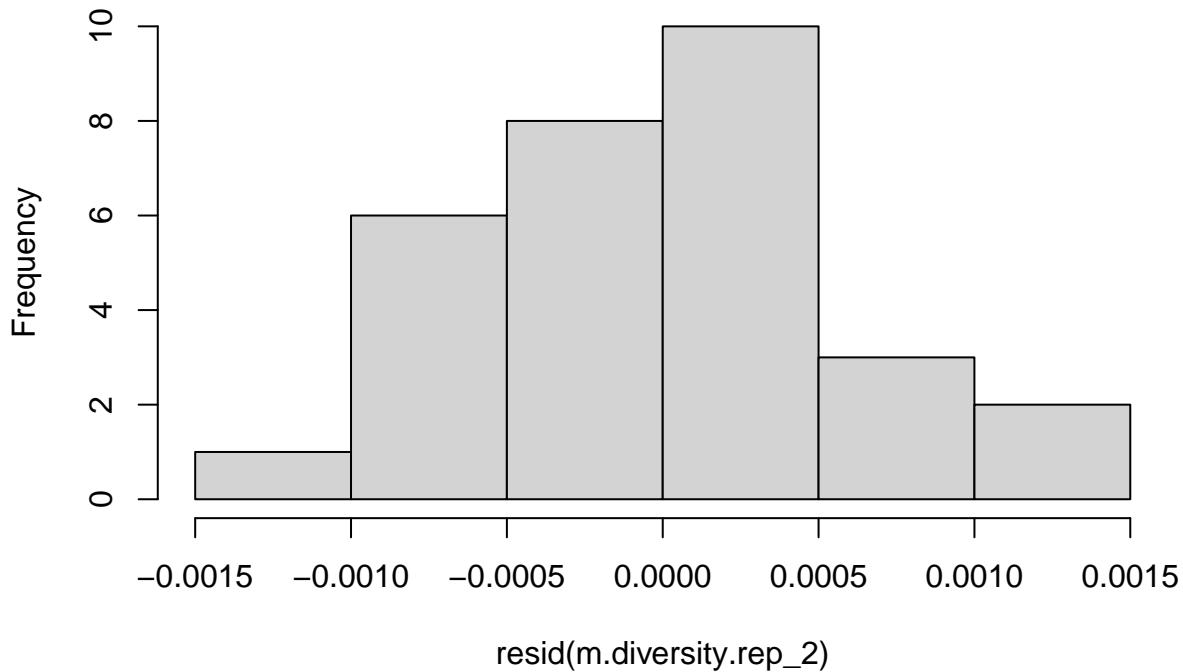


```
dwtest(m.diversity.rep_2)
```

```
##
##  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)
```

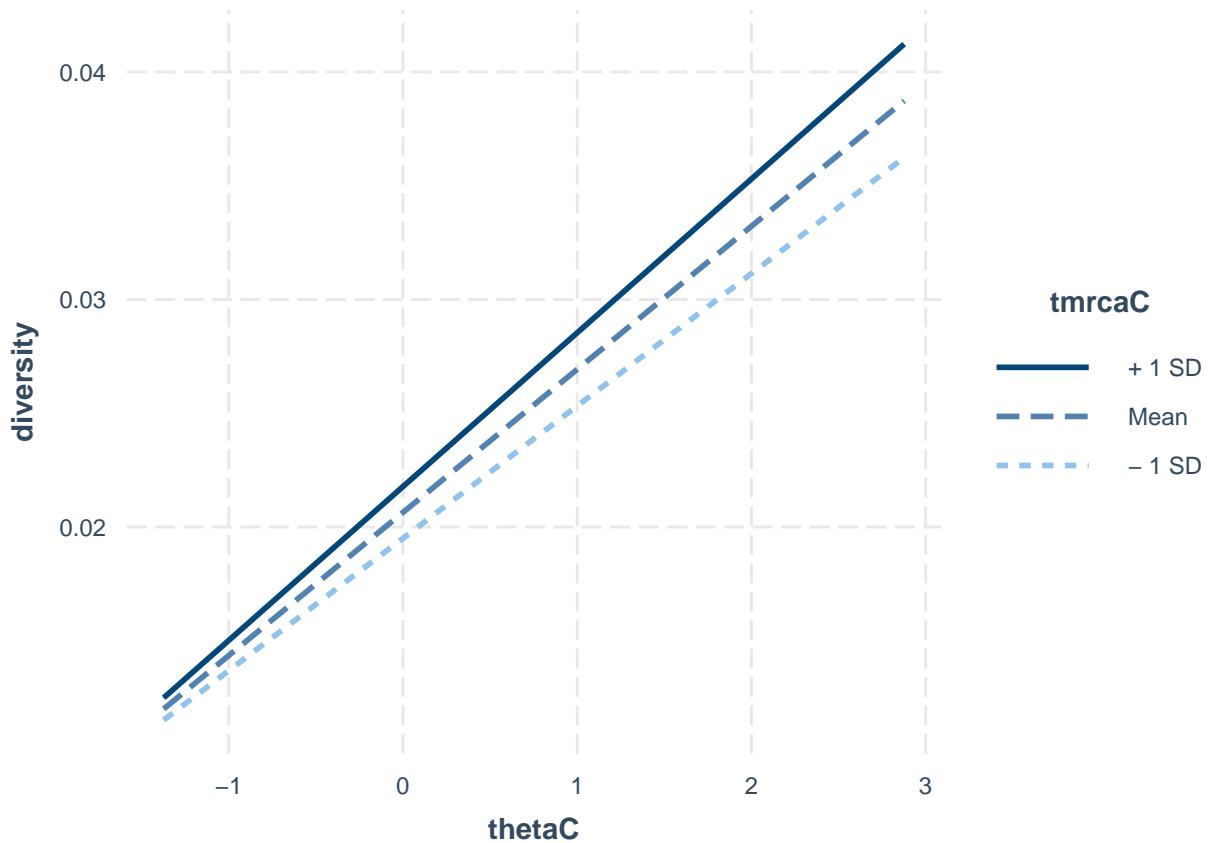
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_2
## HMC = 0.4866, p-value = 0.452
hist(resid(m.diversity.rep_2))
```

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       0.0062884  0.0001152  54.577 < 2e-16 ***
## rhoC        0.0001060  0.0001383   0.766  0.4508
## tmrcaC      0.0011363  0.0001517   7.489 7.66e-08 ***
## thetaC:tmrcaC 0.0004715  0.0001704   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) 0.020642766 0.0001148064 179.80497 0.0000
## thetaC      0.006287821 0.0001152966  54.53607 0.0000
## tmrcaC      0.001136259 0.0001515512   7.49753 0.0000
## rhoC        0.000102832 0.0001387114   0.74134 0.4654
## thetaC:tmrcaC 0.000475113 0.0001702324   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) 0.020593660 0.0001940573 106.12154 0.0000
## thetaC       0.006254517 0.0002002727  31.23000 0.0000
## rhoC        -0.000170535 0.0001996216  -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$VarExp[3] * 100)
r2.inf.1Mb[2, 2] <- anova.diversity$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")

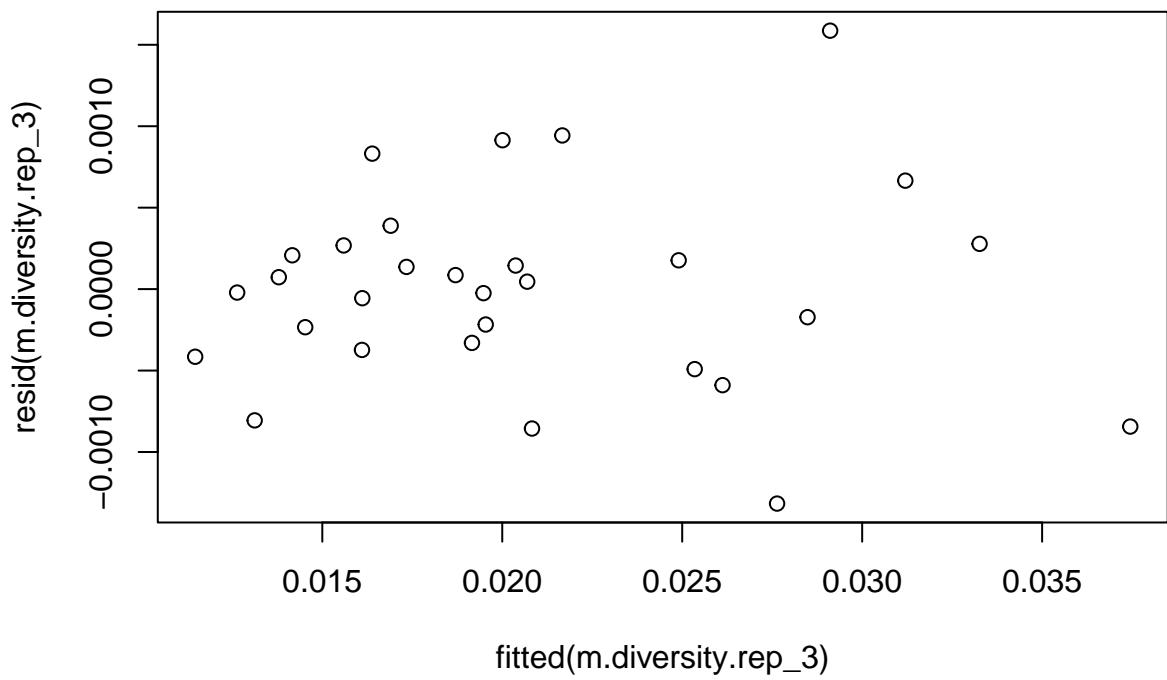
# standardizing
inf.lands.1Mb.rep_3$thetaC <- (inf.lands.1Mb.rep_3$theta - mean(inf.lands.1Mb.rep_3$theta)) / sd(inf.lands.1Mb.rep_3$theta)
inf.lands.1Mb.rep_3$tmrcaC <- (inf.lands.1Mb.rep_3$tmrca - mean(inf.lands.1Mb.rep_3$tmrca)) / sd(inf.lands.1Mb.rep_3$tmrca)
inf.lands.1Mb.rep_3$rhoC <- (inf.lands.1Mb.rep_3$rho - mean(inf.lands.1Mb.rep_3$rho)) / sd(inf.lands.1Mb.rep_3$rho)

inf.lands.1Mb.rep_3$bin <- 1:nrow(inf.lands.1Mb.rep_3)

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))
```

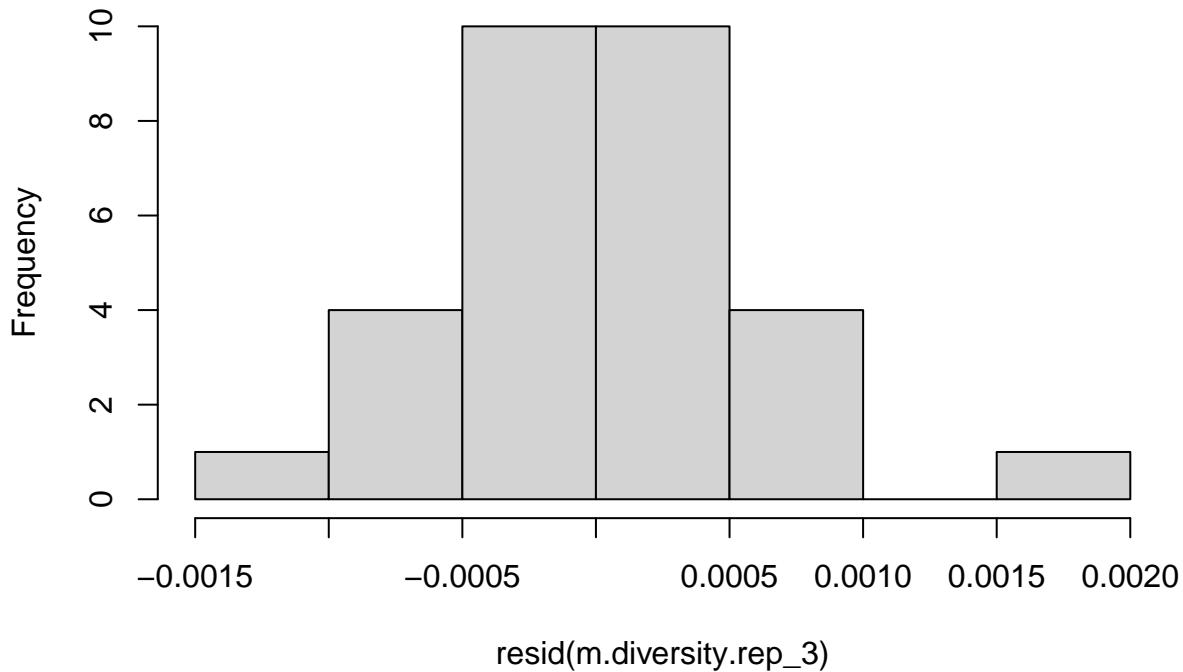


```
dwtest(m.diversity.rep_3)
```

```
##
##  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)
```

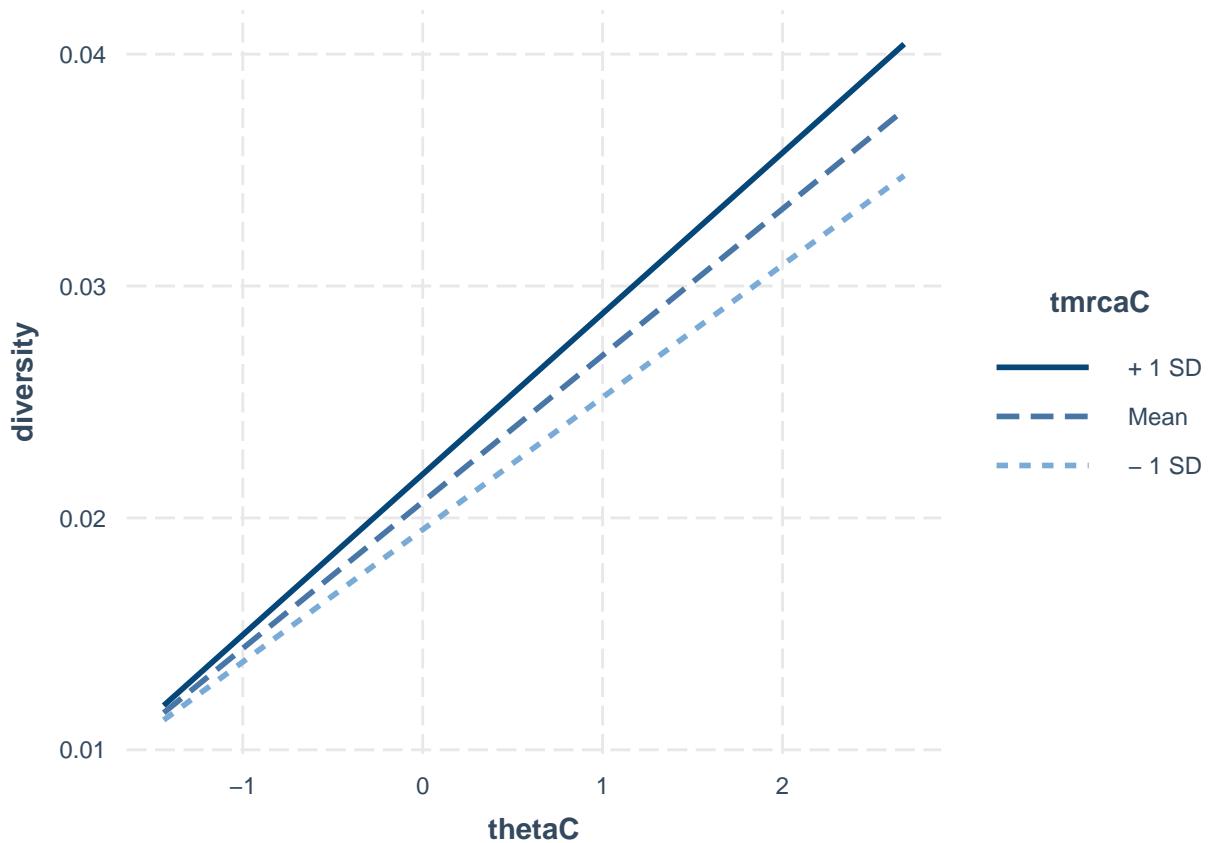
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_3
## HMC = 0.51975, p-value = 0.567
hist(resid(m.diversity.rep_3))
```

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) 2.069e-02 1.206e-04 171.565 < 2e-16 ***
## thetaC      6.318e-03 1.260e-04  50.146 < 2e-16 ***
## rhoC        5.387e-05 1.386e-04   0.389  0.70076
## tmrcaC     1.195e-03 1.461e-04   8.177 1.57e-08 ***
## thetaC:tmrcaC 6.129e-04 1.704e-04    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) 0.020686282 0.0001662473 124.43077 0.0000
## thetaC       0.006300258 0.0001217608  51.74291 0.0000
## tmrcaC       0.001233606 0.0001289727   9.56486 0.0000
## rhoC         0.000032878 0.0001293407   0.25420 0.8014
## thetaC:tmrcaC 0.000666344 0.0001545784   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) 0.020733044 0.0002082301 99.56794 0.0000
## thetaC       0.006488505 0.0002221686 29.20532 0.0000
## rhoC        -0.000114023 0.0002210716 -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$VarExp[3]) * 100
r2.inf.1Mb[2, 3] <- anova.diversity$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")

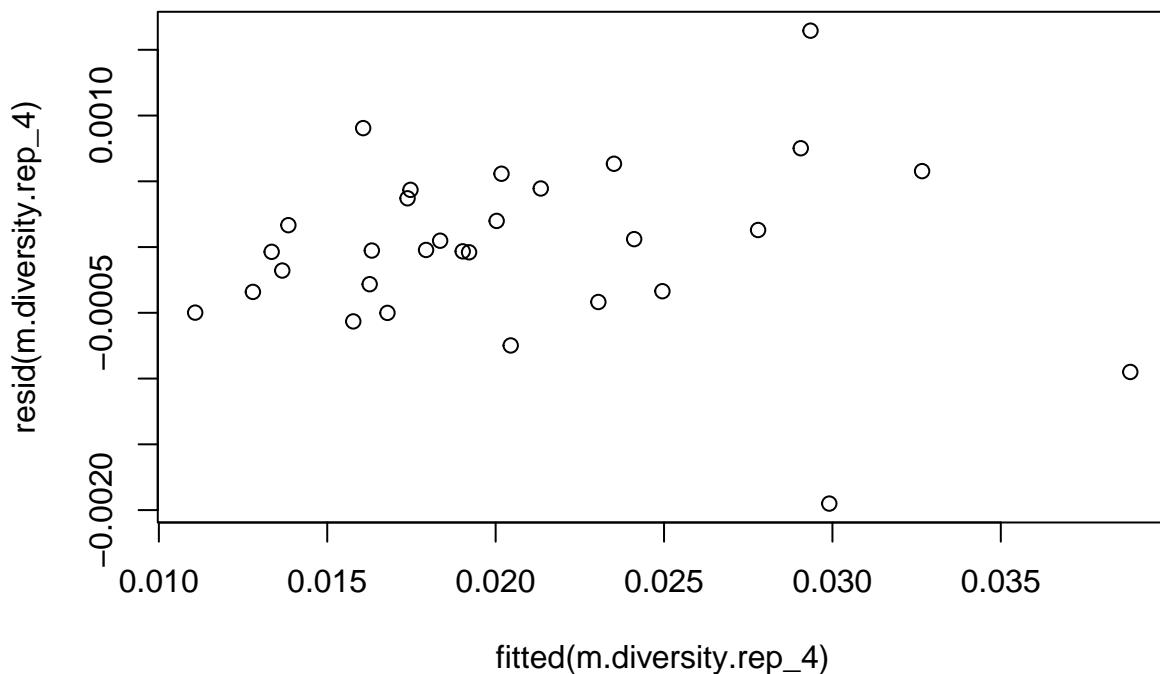
# standardizing
inf.lands.1Mb.rep_4$thetaC <- (inf.lands.1Mb.rep_4$theta - mean(inf.lands.1Mb.rep_4$theta)) / sd(inf.lands.1Mb.rep_4$theta)
inf.lands.1Mb.rep_4$tmrcaC <- (inf.lands.1Mb.rep_4$tmrca - mean(inf.lands.1Mb.rep_4$tmrca)) / sd(inf.lands.1Mb.rep_4$tmrca)
inf.lands.1Mb.rep_4$rhoC <- (inf.lands.1Mb.rep_4$rho - mean(inf.lands.1Mb.rep_4$rho)) / sd(inf.lands.1Mb.rep_4$rho)

inf.lands.1Mb.rep_4$bin <- 1:nrow(inf.lands.1Mb.rep_4)

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))
```

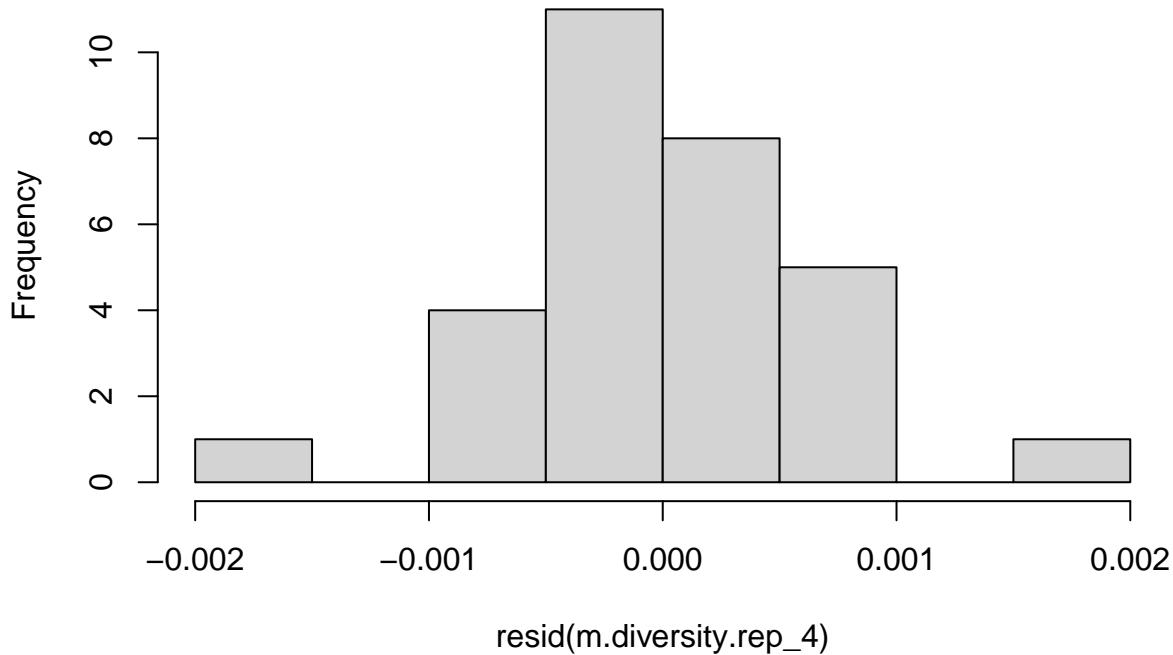


```
dwtest(m.diversity.rep_4)
```

```
##
##  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)
```

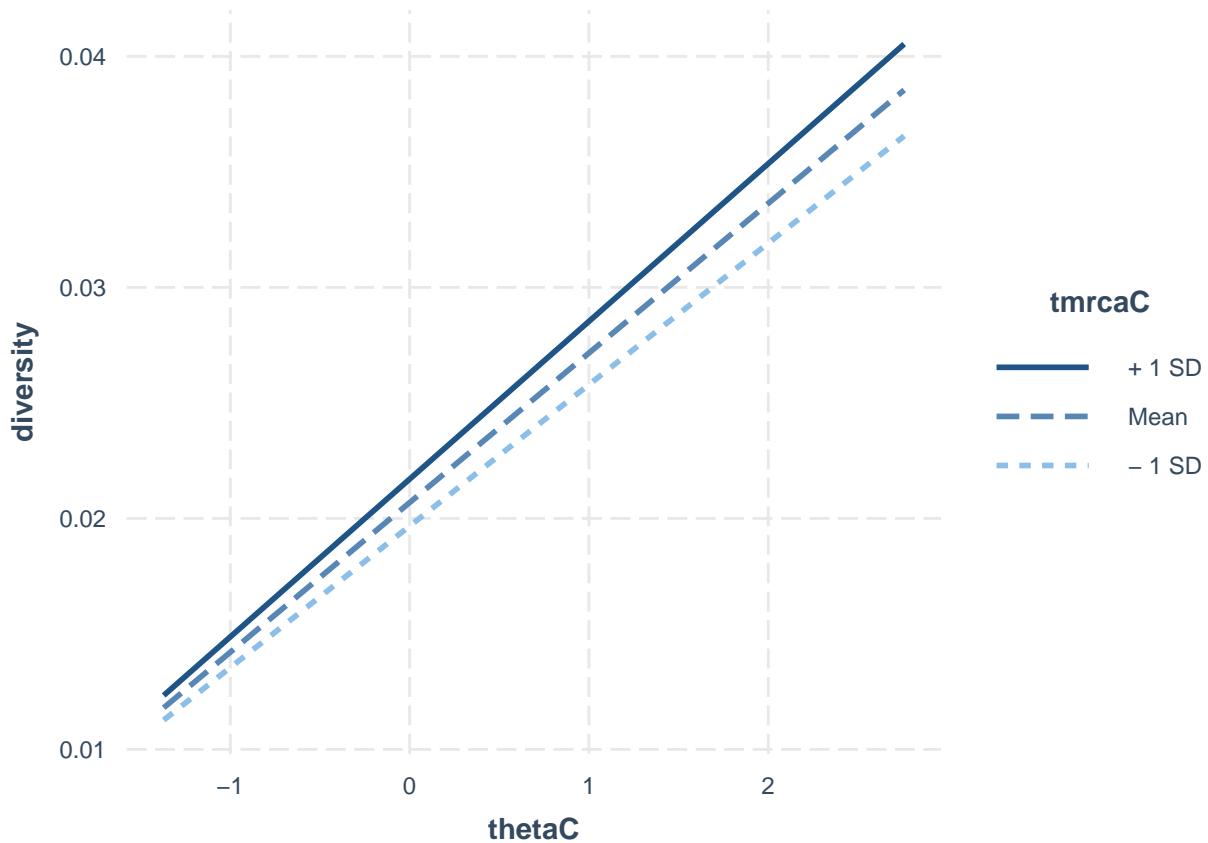
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_4
## HMC = 0.42593, p-value = 0.291
hist(resid(m.diversity.rep_4))
```

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       0.0064755 0.0001308  49.520 < 2e-16 ***
## rhoC        -0.0000109 0.0001471  -0.074   0.942
## tmrcaC       0.0010166 0.0001896   5.363 1.46e-05 ***
## thetaC:tmrcaC 0.0003526 0.0002292   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) 0.020683851 0.0001213164 170.49504 0.0000
## thetaC      0.006477694 0.0001294896  50.02482 0.0000
## tmrcaC      0.000989578 0.0001927169   5.13488 0.0000
## rhoC        -0.000015021 0.0001449330  -0.10364 0.9183
## thetaC:tmrcaC 0.000321760 0.0002315284    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) 0.020679766 0.0001347863 153.42630 0.0000
## thetaC       0.006499529 0.0001673925  38.82807 0.0000
## rhoC        -0.000371115 0.0001562284  -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$VarExp[3] * 100)
r2.inf.1Mb[2, 4] <- anova.diversity$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")

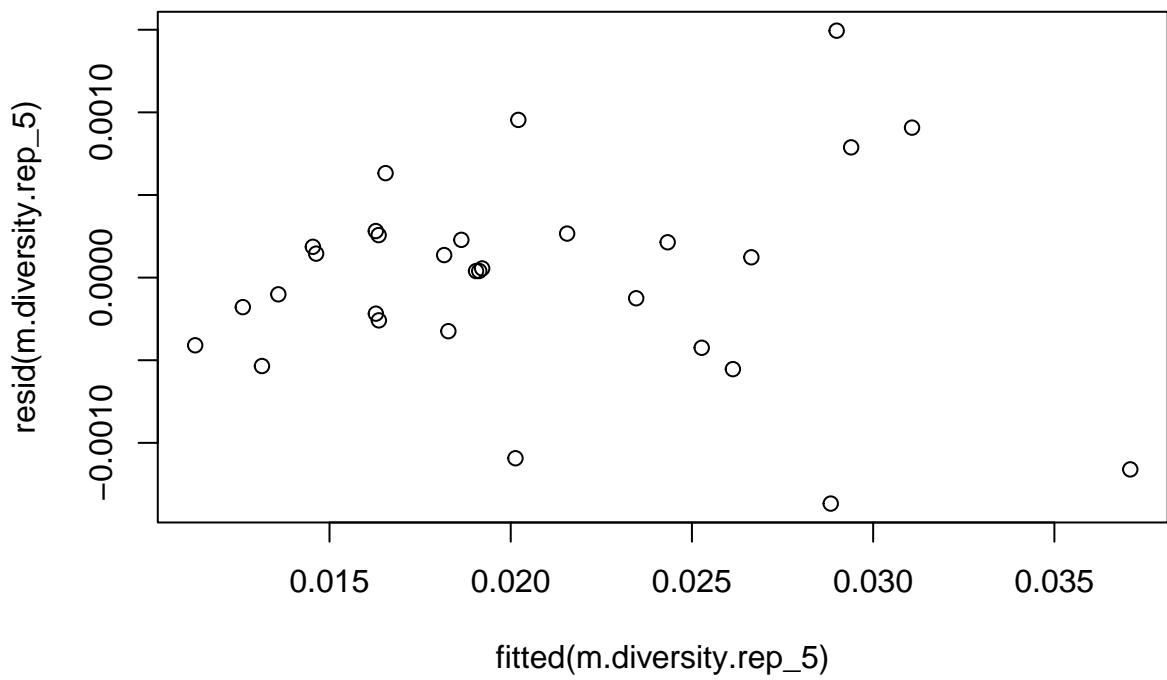
# standardizing
inf.lands.1Mb.rep_5$thetaC <- (inf.lands.1Mb.rep_5$theta - mean(inf.lands.1Mb.rep_5$theta)) / sd(inf.lands.1Mb.rep_5$theta)
inf.lands.1Mb.rep_5$tmrcaC <- (inf.lands.1Mb.rep_5$tmrca - mean(inf.lands.1Mb.rep_5$tmrca)) / sd(inf.lands.1Mb.rep_5$tmrca)
inf.lands.1Mb.rep_5$rhoC <- (inf.lands.1Mb.rep_5$rho - mean(inf.lands.1Mb.rep_5$rho)) / sd(inf.lands.1Mb.rep_5$rho)

inf.lands.1Mb.rep_5$bin <- 1:nrow(inf.lands.1Mb.rep_5)

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))
```

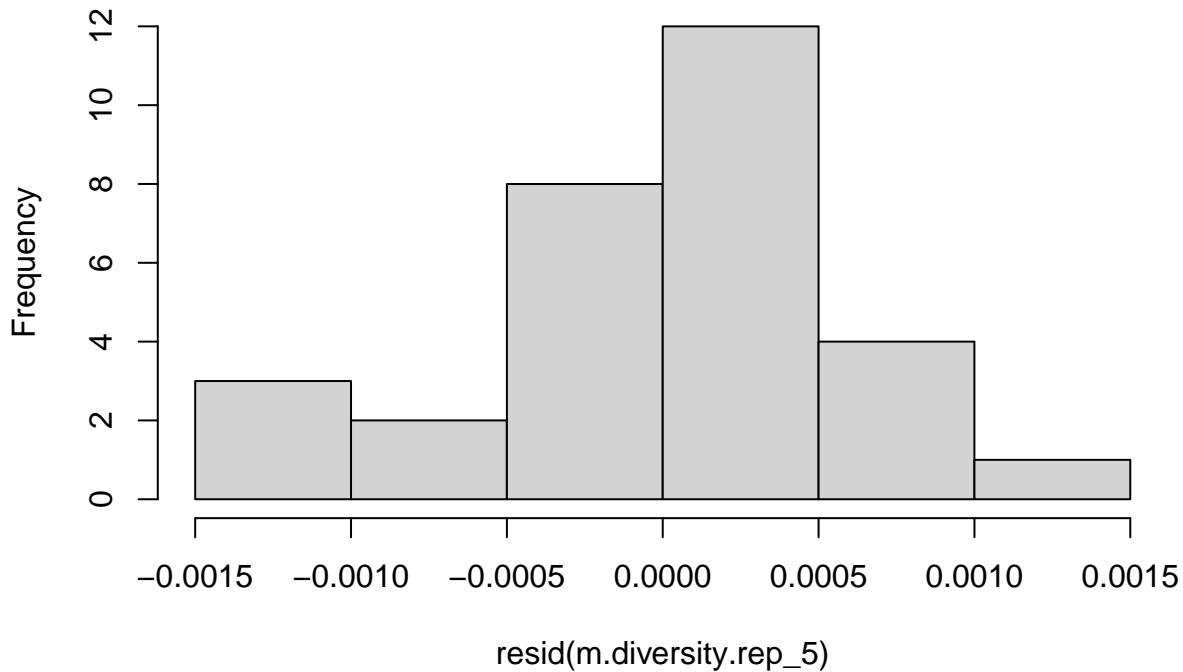


```
dwtest(m.diversity.rep_5)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_5
## DW = 1.8381, p-value = 0.2796
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_5)
```

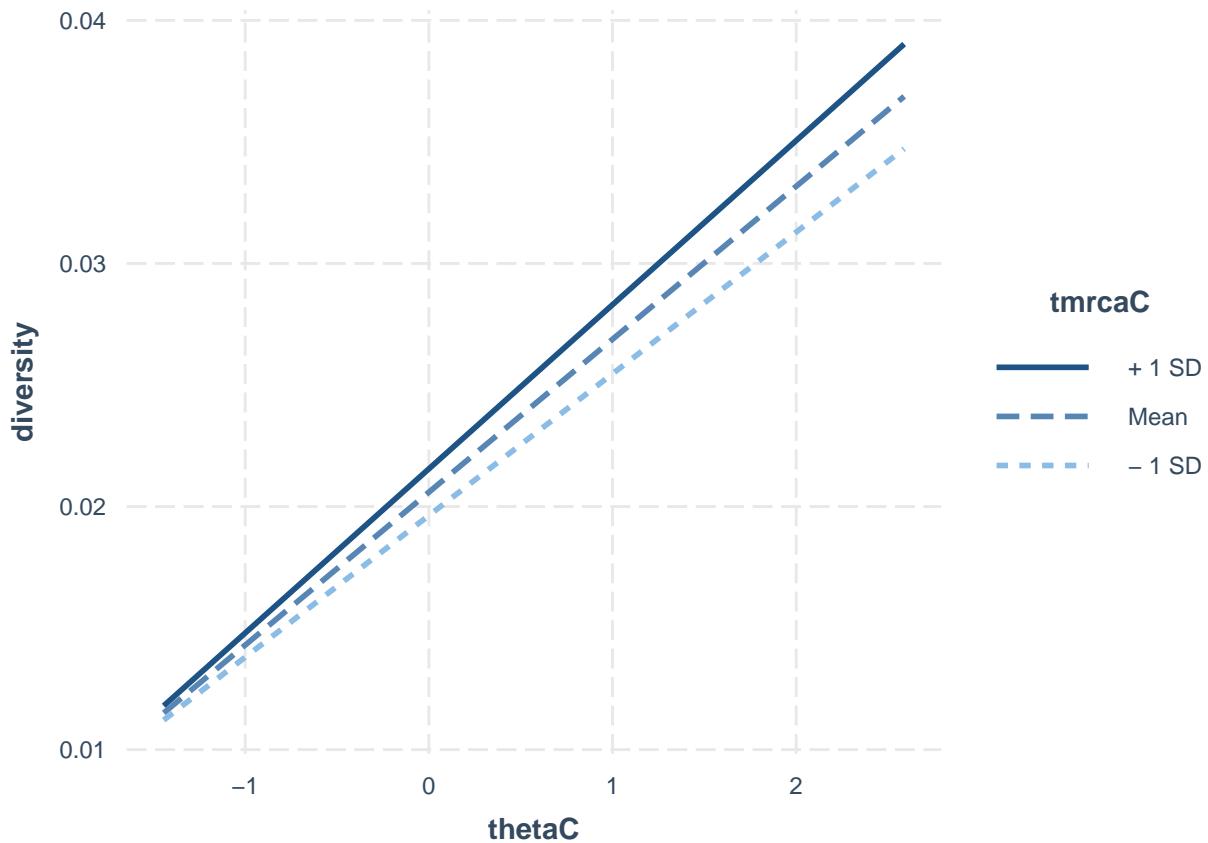
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_5
## HMC = 0.49429, p-value = 0.478
hist(resid(m.diversity.rep_5))
```

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) 2.059e-02  1.215e-04 169.475 < 2e-16 ***
## thetaC       6.291e-03  1.237e-04  50.855 < 2e-16 ***
## rhoC        -7.338e-04  2.362e-03  -0.311   0.7586
## tmrcaC       9.891e-05  1.894e-05   5.224  2.09e-05 ***
## thetaC:tmrcaC 4.759e-05  2.090e-05   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) 0.020593322 0.0001335763 154.16894 0.0000
## thetaC       0.006289699 0.0001243303  50.58862 0.0000
## tmrcaC      0.000102000 0.0000185454    5.50000 0.0000
## rhoC        -0.000671391 0.0023700424  -0.28328 0.7793
## thetaC:tmrcaC 0.000052459 0.0000206457    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) 0.020569710 0.0001476467 139.31706 0.0000
## thetaC       0.006238359 0.0001667110  37.42021 0.0000
## rhoC        -0.006901949 0.0026735091  -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$VarExp[3] * 100)
r2.inf.1Mb[2, 5] <- anova.diversity$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")

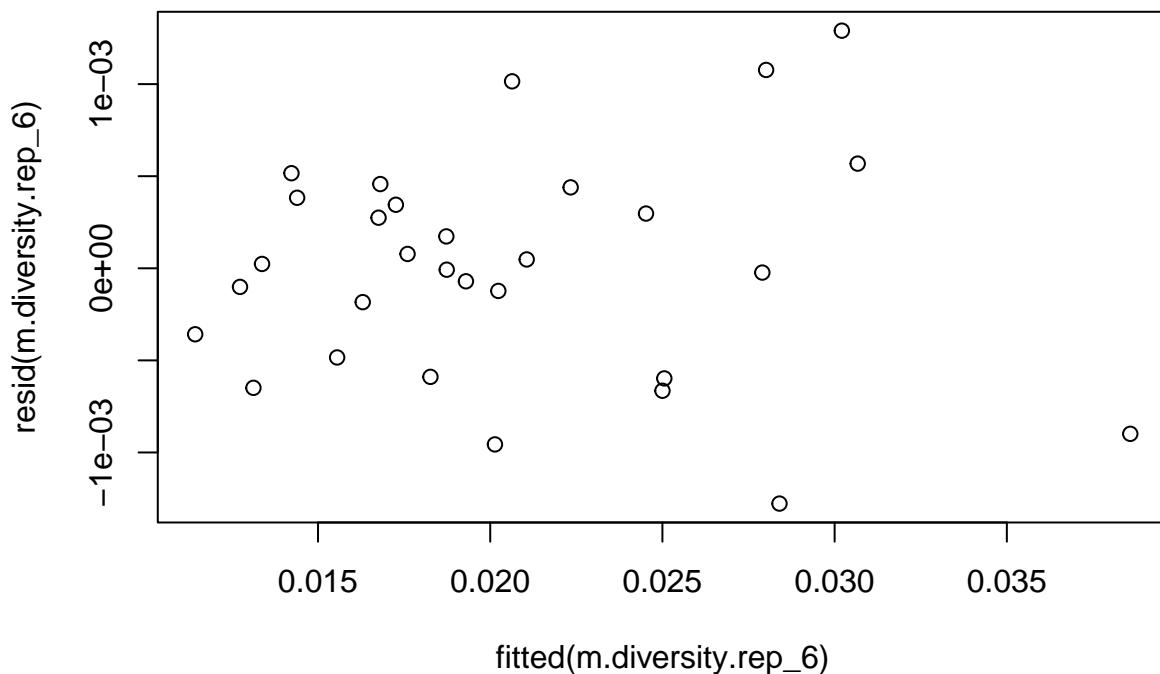
# standardizing
inf.lands.1Mb.rep_6$thetaC <- (inf.lands.1Mb.rep_6$theta - mean(inf.lands.1Mb.rep_6$theta)) / sd(inf.lands.1Mb.rep_6$theta)
inf.lands.1Mb.rep_6$tmrcaC <- (inf.lands.1Mb.rep_6$tmrca - mean(inf.lands.1Mb.rep_6$tmrca)) / sd(inf.lands.1Mb.rep_6$tmrca)
inf.lands.1Mb.rep_6$rhoC <- (inf.lands.1Mb.rep_6$rho - mean(inf.lands.1Mb.rep_6$rho)) / sd(inf.lands.1Mb.rep_6$rho)

inf.lands.1Mb.rep_6$bin <- 1:nrow(inf.lands.1Mb.rep_6)

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))
```

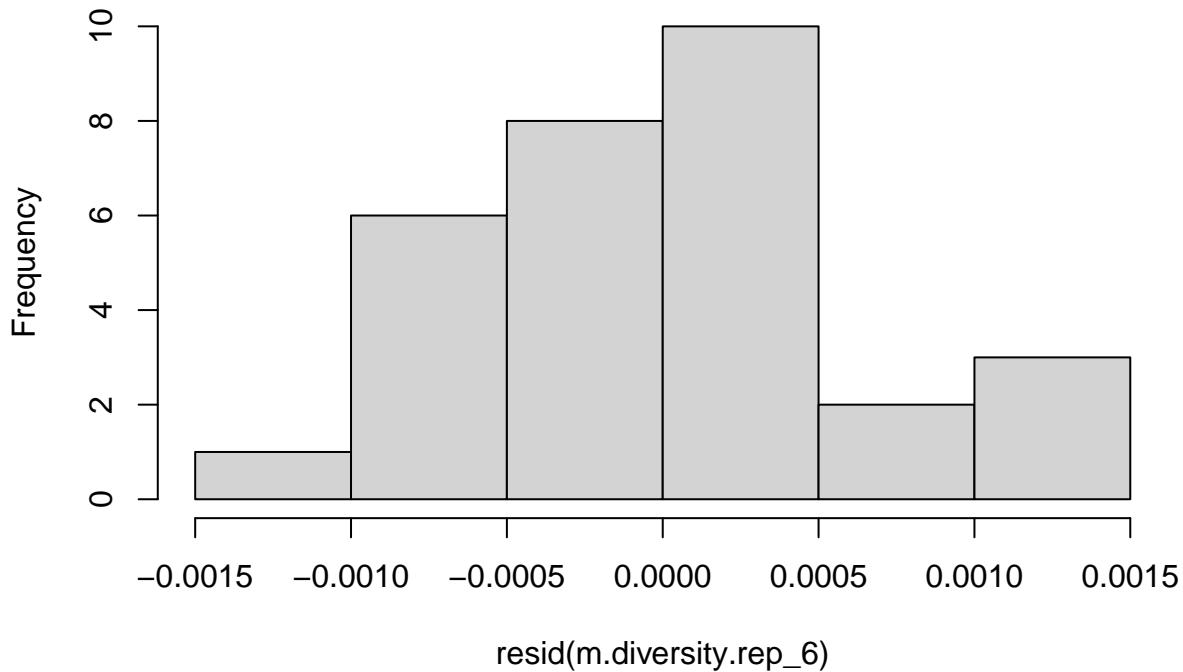


```
dwtest(m.diversity.rep_6)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_6
## DW = 1.7678, p-value = 0.2162
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_6)
```

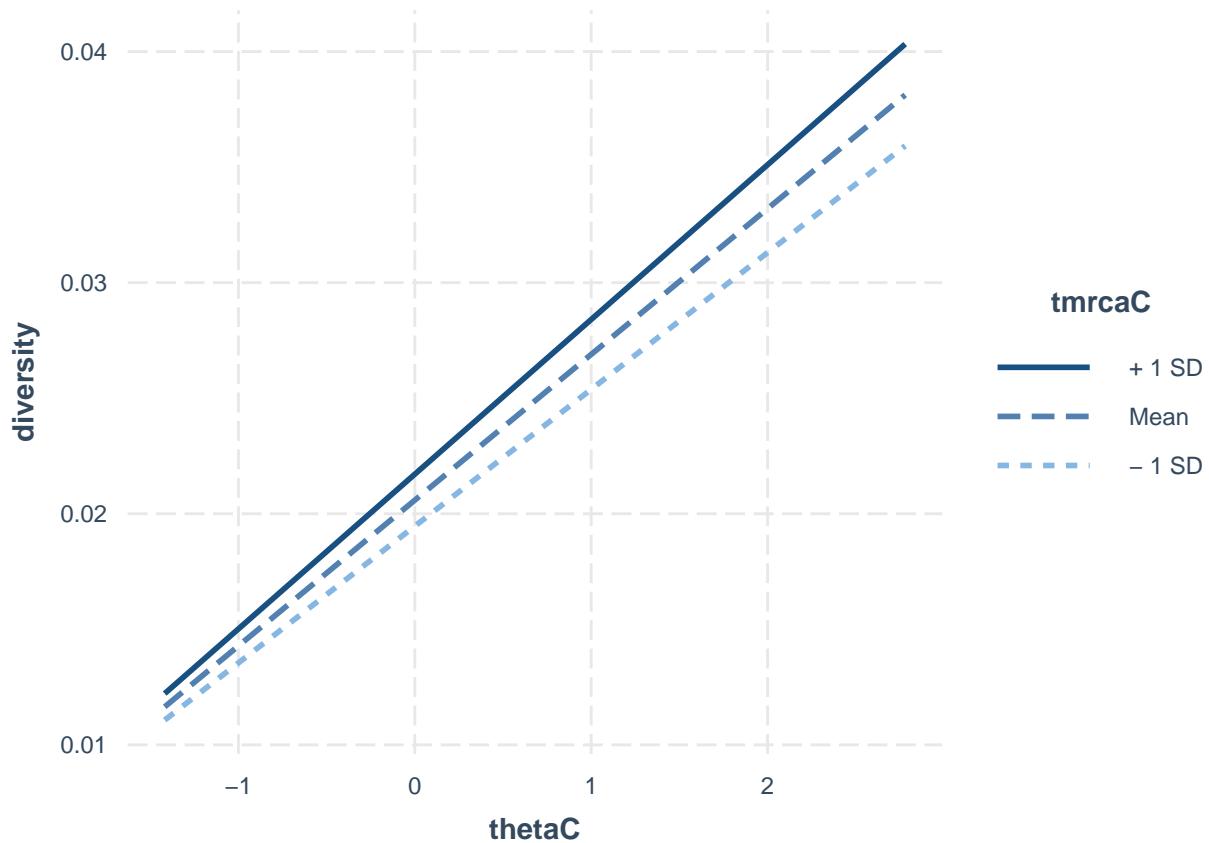
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_6
## HMC = 0.38206, p-value = 0.186
hist(resid(m.diversity.rep_6))
```

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      0.0063060  0.0001232  51.167 < 2e-16 ***
## rhoC        0.0001654  0.0001529   1.082  0.2897
## tmrcaC      0.0011233  0.0001604   7.004 2.43e-07 ***
## thetaC:tmrcaC 0.0003879  0.0001649   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) 0.020592380 0.0001362541 151.13216 0.0000
## thetaC       0.006299404 0.0001239804  50.80966 0.0000
## tmrcaC       0.001133198 0.0001570286   7.21651 0.0000
## rhoC         0.000143376 0.0001542687   0.92939 0.3616
## thetaC:tmrcaC 0.000435077 0.0001634509   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) 0.020574682 0.0001889124 108.91122 0.0000
## thetaC       0.006311497 0.0001990837  31.70273 0.0000
## rhoC        -0.000349915 0.0001977951  -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$VarExp[3] * 100)
r2.inf.1Mb[2, 6] <- anova.diversity$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")

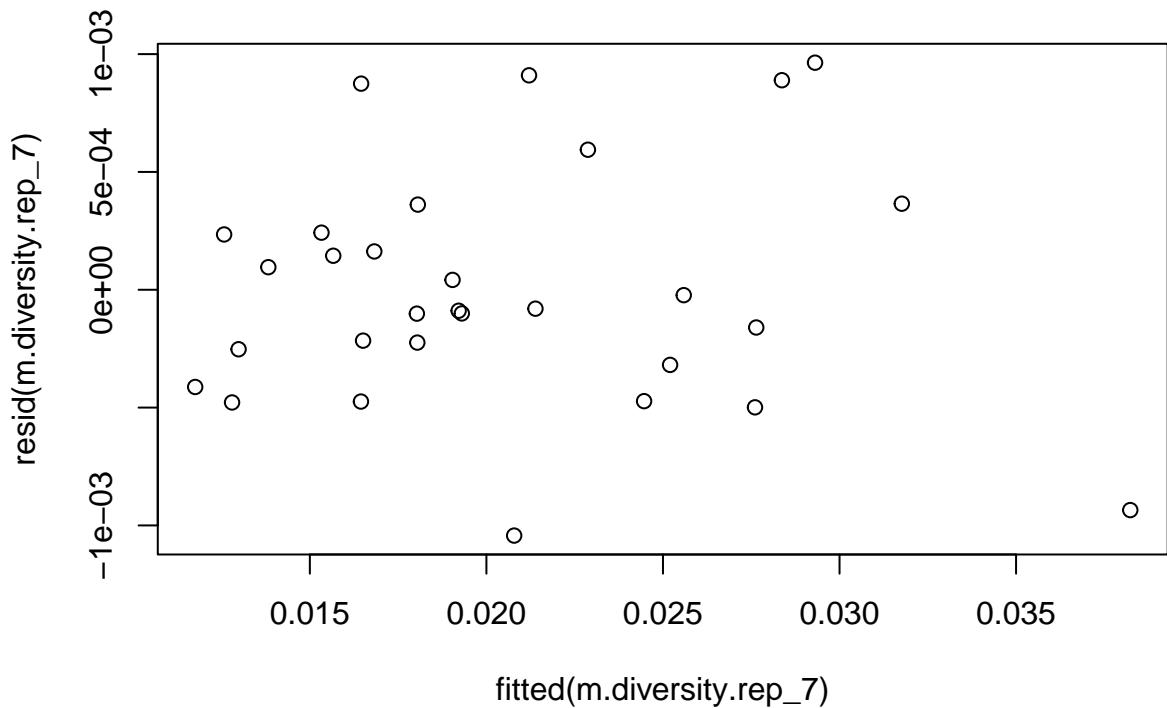
# standardizing
inf.lands.1Mb.rep_7$thetaC <- (inf.lands.1Mb.rep_7$theta - mean(inf.lands.1Mb.rep_7$theta)) / sd(inf.lands.1Mb.rep_7$theta)
inf.lands.1Mb.rep_7$tmrcaC <- (inf.lands.1Mb.rep_7$tmrca - mean(inf.lands.1Mb.rep_7$tmrca)) / sd(inf.lands.1Mb.rep_7$tmrca)
inf.lands.1Mb.rep_7$rhoC <- (inf.lands.1Mb.rep_7$rho - mean(inf.lands.1Mb.rep_7$rho)) / sd(inf.lands.1Mb.rep_7$rho)

inf.lands.1Mb.rep_7$bin <- 1:nrow(inf.lands.1Mb.rep_7)

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))
```



```

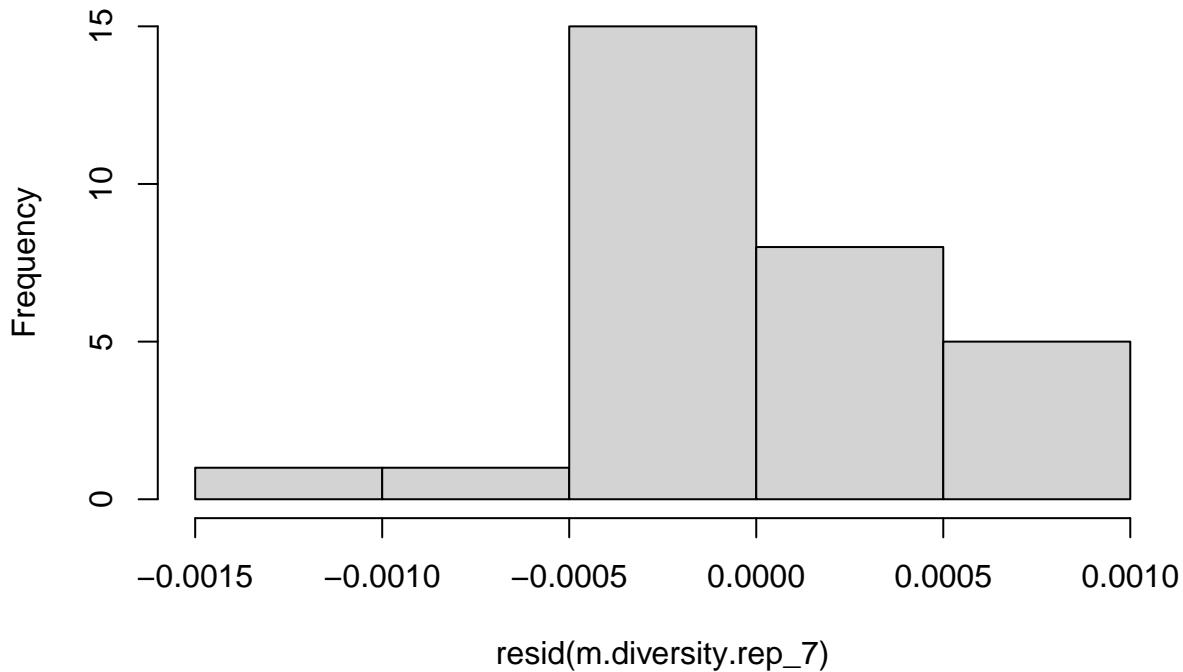
dwtest(m.diversity.rep_7)

##
##  Durbin-Watson test
##
## data: m.diversity.rep_7
## DW = 1.6938, p-value = 0.1634
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_7)

##
##  Harrison-McCabe test
##
## data: m.diversity.rep_7
## HMC = 0.40789, p-value = 0.241
hist(resid(m.diversity.rep_7))

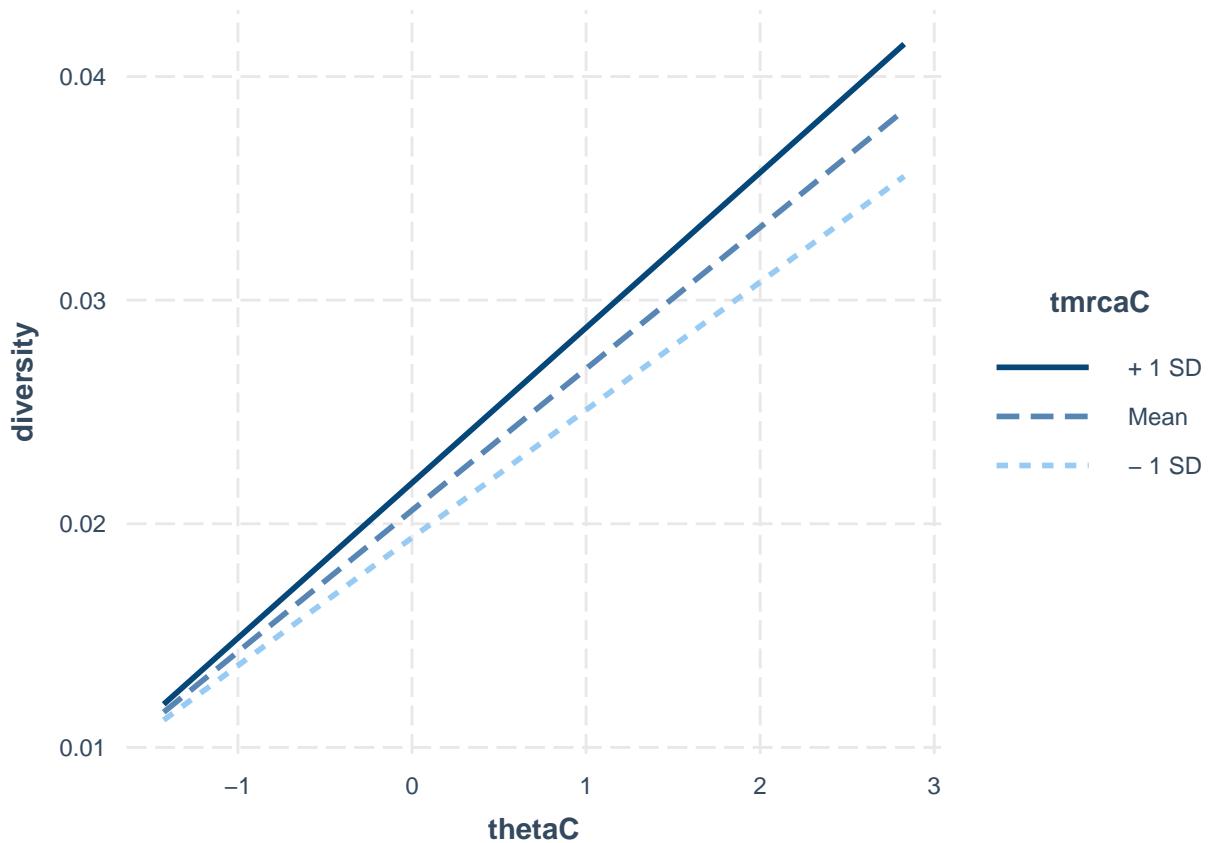
```

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) 2.060e-02 1.001e-04 205.827 < 2e-16 ***  
## thetaC       6.328e-03 1.026e-04  61.661 < 2e-16 ***  
## rhoC        2.408e-05 1.192e-04   0.202 0.841513  
## tmrcaC      1.227e-03 1.432e-04   8.568 6.58e-09 ***  
## thetaC:tmrcaC 6.121e-04 1.441e-04    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) 0.020604107 0.0001183950 174.02851 0.0000
## thetaC       0.006330465 0.0001029690  61.47931 0.0000
## tmrcaC       0.001201640 0.0001383881    8.68312 0.0000
## rhoC        -0.000015831 0.0001187358   -0.13333 0.8950
## thetaC:tmrcaC 0.000613682 0.0001380243    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) 0.020581847 0.0002148595 95.79213 0.0000
## thetaC       0.006399198 0.0001974295 32.41258 0.0000
## rhoC        -0.000287101 0.0002007978 -1.42980 0.1642
##
## Correlation:
##      (Intr) thetaC
## thetaC -0.001
## rhoC   0.002 -0.097
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.37431231 -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$VarExp[3] * 100)
r2.inf.1Mb[2, 7] <- anova.diversity$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")

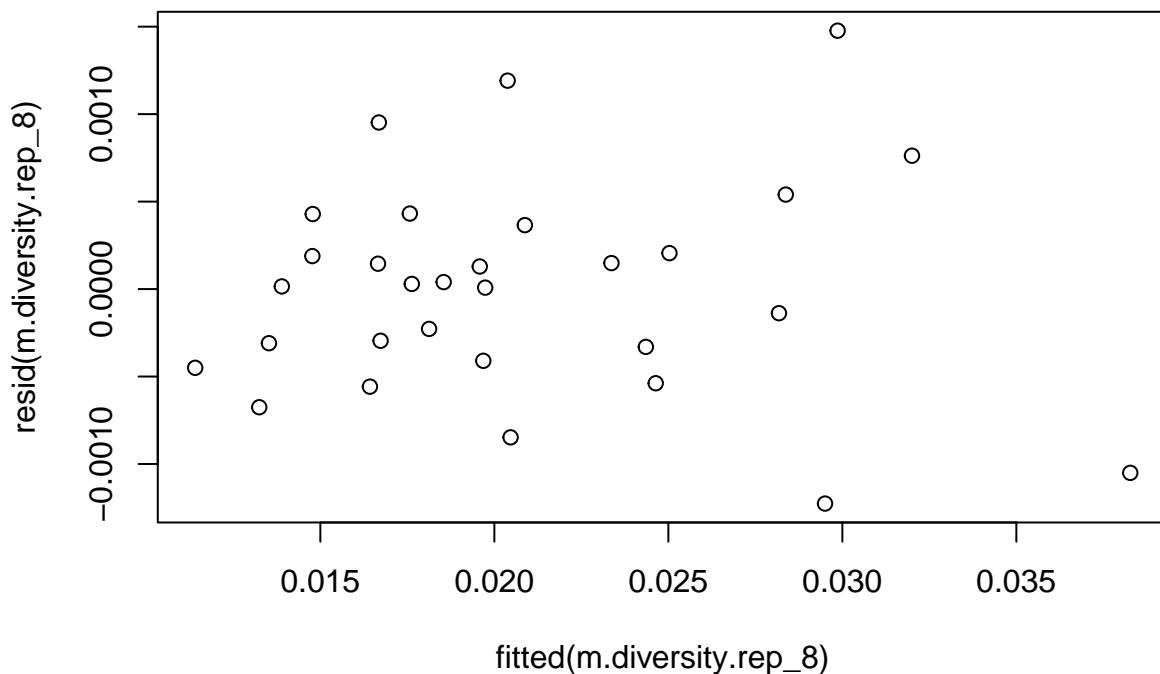
# standardizing
inf.lands.1Mb.rep_8$thetaC <- (inf.lands.1Mb.rep_8$theta - mean(inf.lands.1Mb.rep_8$theta)) / sd(inf.lands.1Mb.rep_8$theta)
inf.lands.1Mb.rep_8$tmrcaC <- (inf.lands.1Mb.rep_8$tmrca - mean(inf.lands.1Mb.rep_8$tmrca)) / sd(inf.lands.1Mb.rep_8$tmrca)
inf.lands.1Mb.rep_8$rhoC <- (inf.lands.1Mb.rep_8$rho - mean(inf.lands.1Mb.rep_8$rho)) / sd(inf.lands.1Mb.rep_8$rho)

inf.lands.1Mb.rep_8$bin <- 1:nrow(inf.lands.1Mb.rep_8)

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))
```

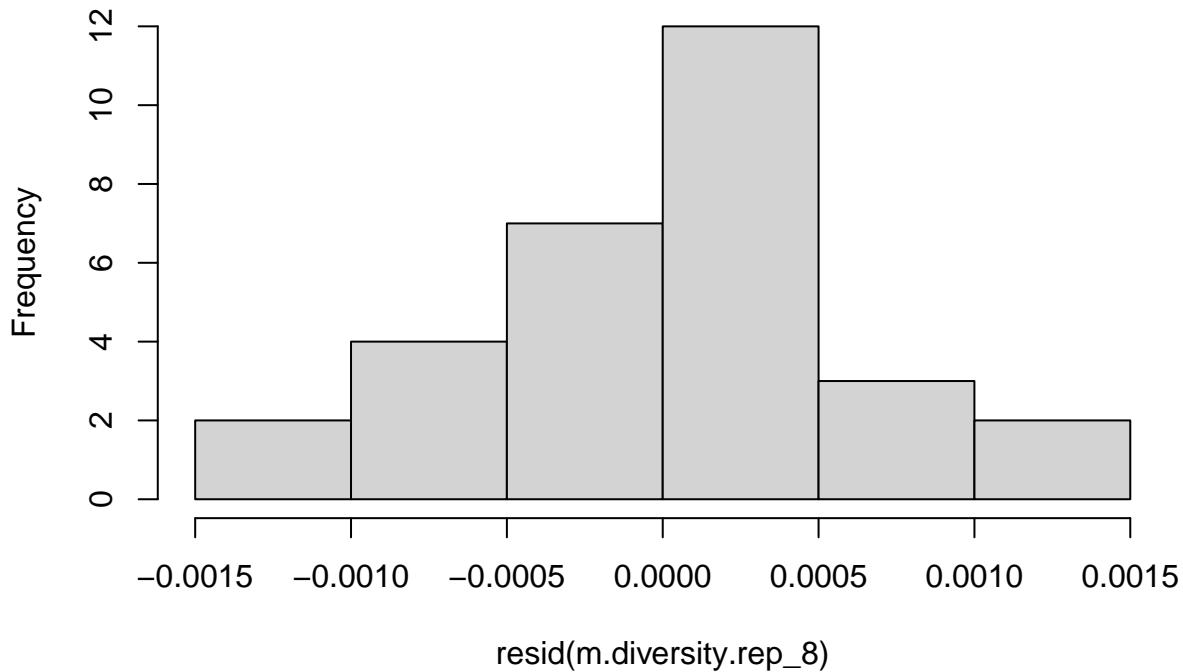


```
dwtest(m.diversity.rep_8)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_8
## DW = 1.6816, p-value = 0.17
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_8)
```

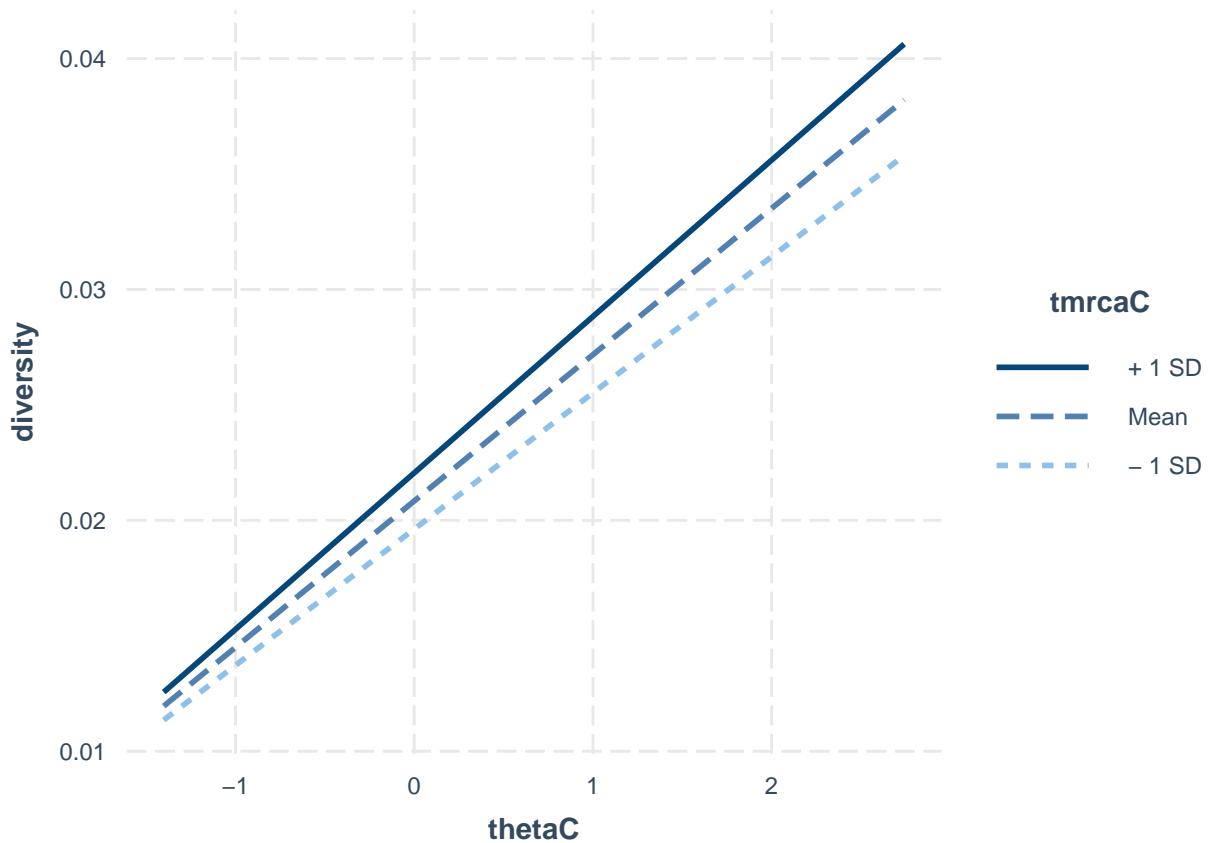
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_8
## HMC = 0.45613, p-value = 0.377
hist(resid(m.diversity.rep_8))
```

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      0.0063380  0.0001252  50.620 < 2e-16 ***  
## rhoC        0.0001281  0.0001427   0.897  0.3781  
## tmrcaC      0.0012170  0.0002058   5.915 3.58e-06 ***  
## thetaC:tmrcaC 0.0004373  0.0001878   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) 0.020838721 0.0001435030 145.21451 0.0000
## thetaC       0.006320203 0.0001245996  50.72412 0.0000
## tmrcaC       0.001246236 0.0001908528   6.52983 0.0000
## rhoC         0.000120494 0.0001459813   0.82541 0.4169
## thetaC:tmrcaC 0.000474550 0.0001759369   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) 0.020803517 0.0001699309 122.42339 0.0000
## thetaC       0.006310129 0.0001914735  32.95562 0.0000
## rhoC        -0.000199993 0.0001849373  -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$VarExp[3] * 100)
r2.inf.1Mb[2, 8] <- anova.diversity$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$avg <- apply(pi.1Mb[4:ncol(pi.1Mb)], 1, mean)
tmrca.1Mb <- read.table(paste(p, "TMRCAs.1Mb.bedgraph", sep = ""), header = T)
tmrca.1Mb$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")

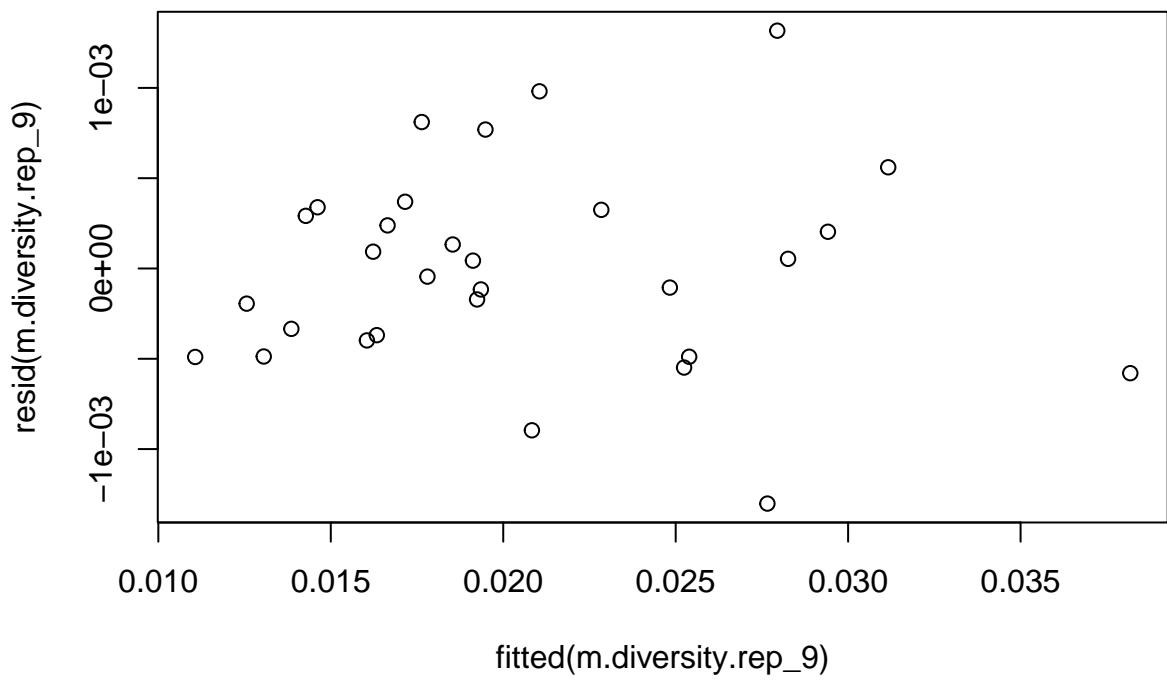
# standardizing
inf.lands.1Mb.rep_9$thetaC <- (inf.lands.1Mb.rep_9$theta - mean(inf.lands.1Mb.rep_9$theta)) / sd(inf.lands.1Mb.rep_9$theta)
inf.lands.1Mb.rep_9$tmrcaC <- (inf.lands.1Mb.rep_9$tmrca - mean(inf.lands.1Mb.rep_9$tmrca)) / sd(inf.lands.1Mb.rep_9$tmrca)
inf.lands.1Mb.rep_9$rhoC <- (inf.lands.1Mb.rep_9$rho - mean(inf.lands.1Mb.rep_9$rho)) / sd(inf.lands.1Mb.rep_9$rho)

inf.lands.1Mb.rep_9$bin <- 1:nrow(inf.lands.1Mb.rep_9)

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))
```

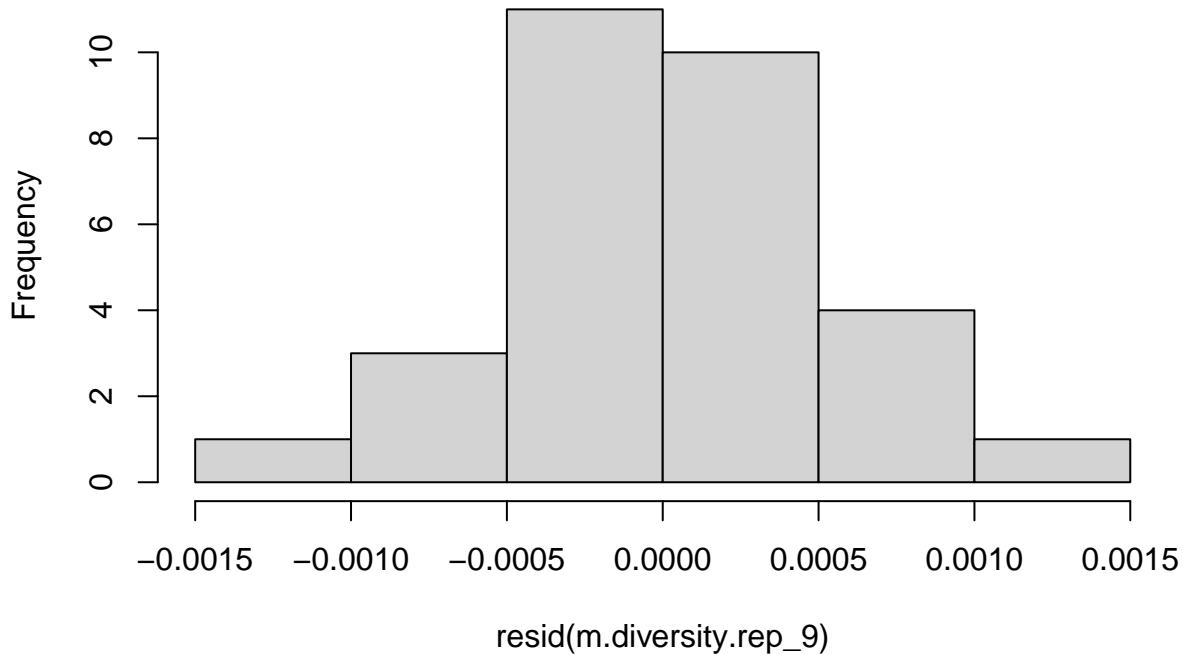


```
dwtest(m.diversity.rep_9)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_9
## DW = 1.686, p-value = 0.1501
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_9)
```

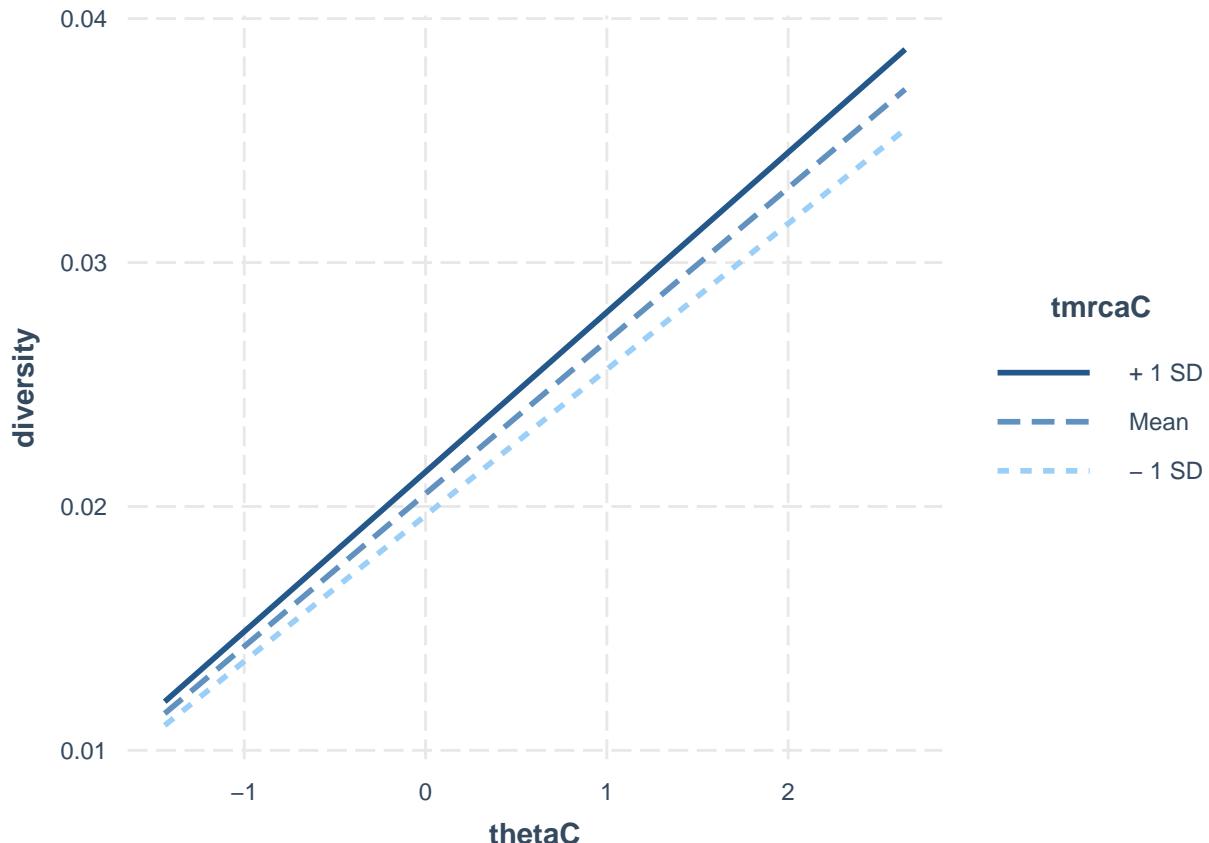
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_9
## HMC = 0.49123, p-value = 0.491
hist(resid(m.diversity.rep_9))
```

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) 2.052e-02  1.107e-04 185.445 < 2e-16 ***
## thetaC       6.265e-03  1.139e-04  55.004 < 2e-16 ***
## rhoC        6.536e-05  1.304e-04   0.501  0.6205
## tmrcaC      8.922e-04  1.363e-04   6.547 7.39e-07 ***
## thetaC:tmrcaC 2.822e-04  1.308e-04   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) 0.020523996 0.0001296981 158.24435 0.0000
## thetaC       0.006241612 0.0001134046  55.03845 0.0000
## tmrcaC       0.000904089 0.0001283011   7.04662 0.0000
## rhoC         0.000075891 0.0001311077   0.57885 0.5679
## thetaC:tmrcaC 0.000285637 0.0001311836   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  0.581067342  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) 0.020514406 0.0001444358 142.03128 0.0000
## thetaC       0.006312166 0.0001690106  37.34776 0.0000
## rhoC        -0.000327436 0.0001615530  -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$VarExp[3] * 100)
r2.inf.1Mb[2, 9] <- anova.diversity$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$avg <- apply(pi.1Mb[4:ncol(pi.1Mb)], 1, mean)
tmrca.1Mb <- read.table(paste(p, "TMRC.1Mb.bedgraph", sep = ""), header = T)
tmrca.1Mb$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

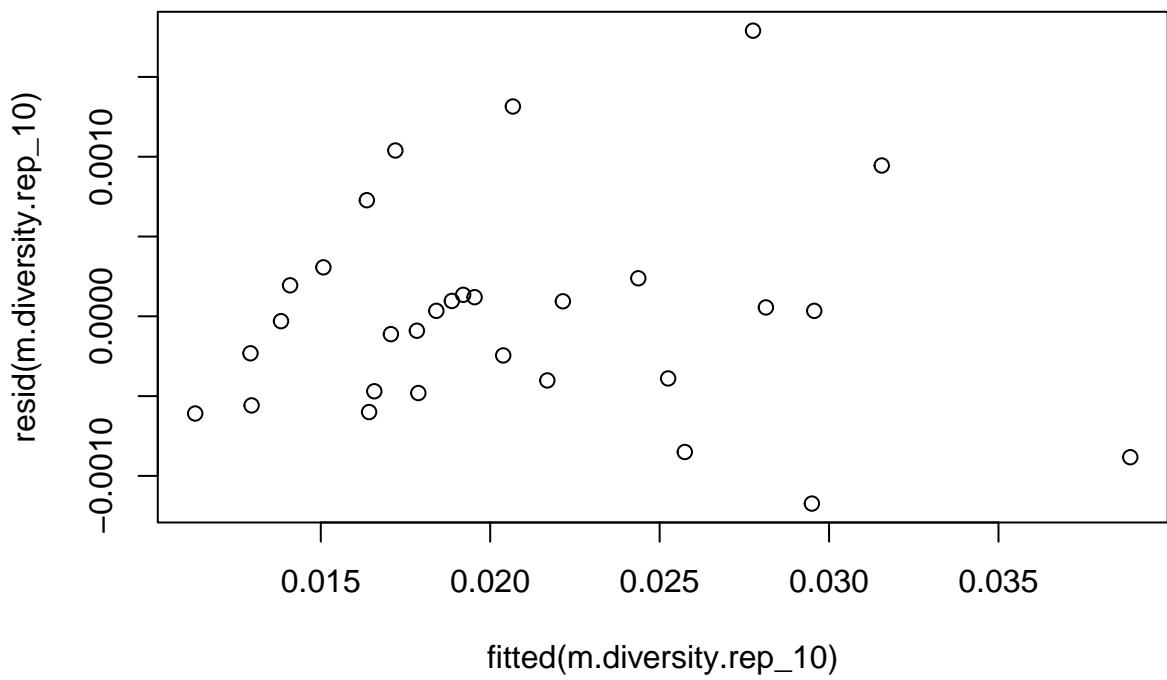
inf.lands.1Mb.rep_10 <- as.data.frame(cbind(pi.1Mb$avg, theta.1Mb$sample_mean, rho.1Mb$sample_mean, tmrca.1Mb$avg))
names(inf.lands.1Mb.rep_10) <- c("diversity", "theta", "rho", "tmrca")

# standardizing
inf.lands.1Mb.rep_10$thetaC <- (inf.lands.1Mb.rep_10$theta - mean(inf.lands.1Mb.rep_10$theta)) / sd(inf.lands.1Mb.rep_10$theta)
inf.lands.1Mb.rep_10$tmrcaC <- (inf.lands.1Mb.rep_10$tmrca - mean(inf.lands.1Mb.rep_10$tmrca)) / sd(inf.lands.1Mb.rep_10$tmrca)
inf.lands.1Mb.rep_10$rhoC <- (inf.lands.1Mb.rep_10$rho - mean(inf.lands.1Mb.rep_10$rho)) / sd(inf.lands.1Mb.rep_10$rho)

inf.lands.1Mb.rep_10$bin <- 1:nrow(inf.lands.1Mb.rep_10)

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))
```

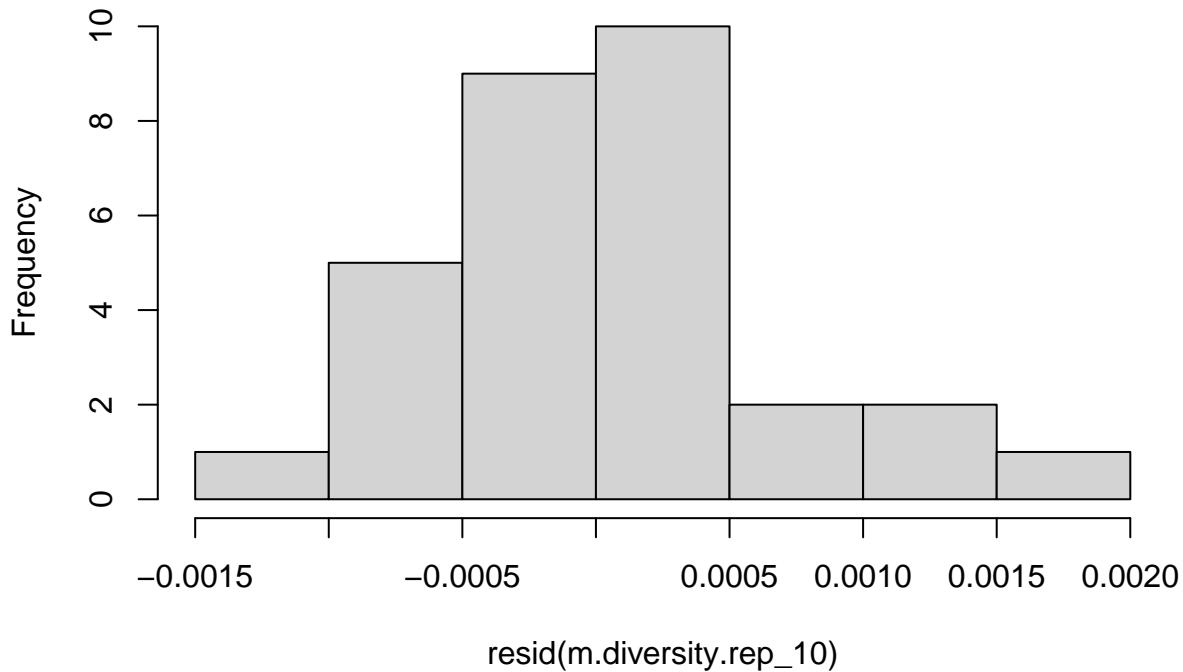


```
dwtest(m.diversity.rep_10)
```

```
##
##  Durbin-Watson test
##
## data: m.diversity.rep_10
## DW = 1.5209, p-value = 0.06939
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.diversity.rep_10)
```

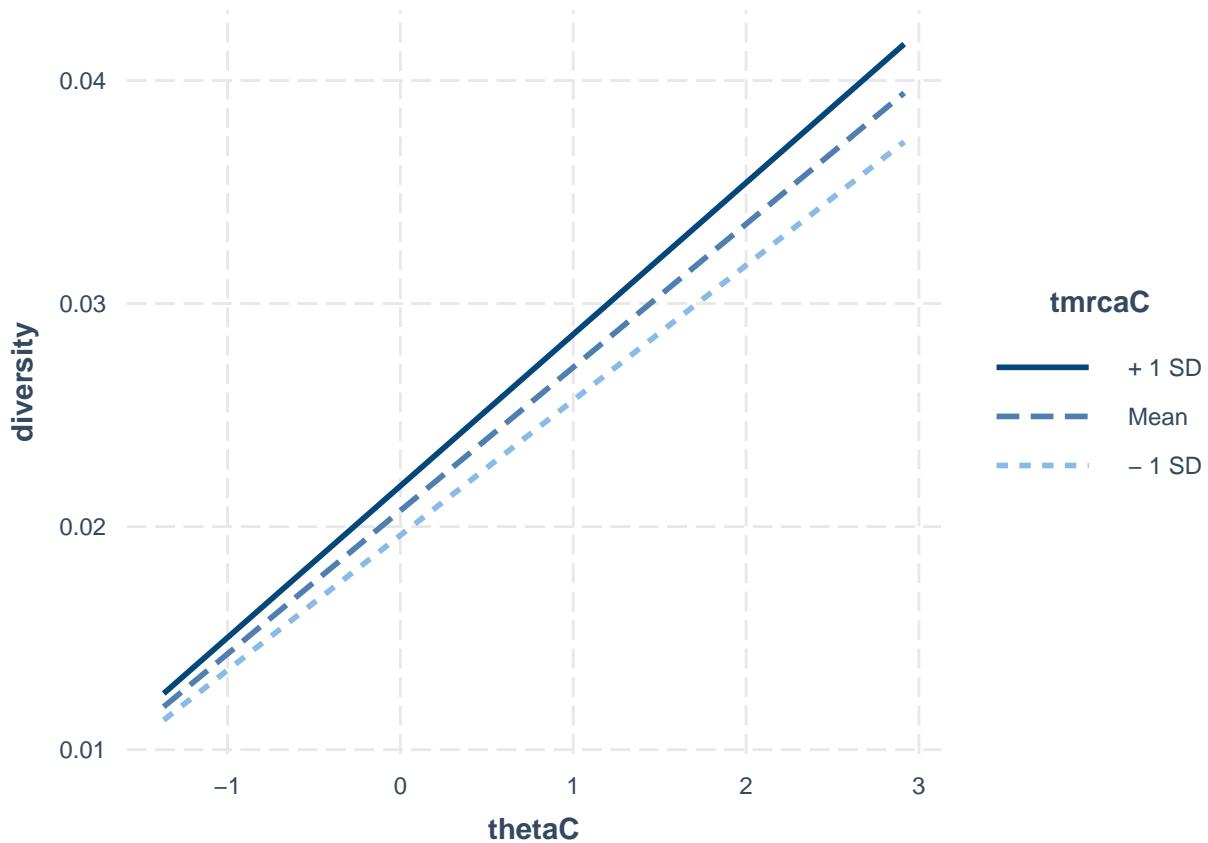
```
##
##  Harrison-McCabe test
##
## data: m.diversity.rep_10
## HMC = 0.59687, p-value = 0.765
hist(resid(m.diversity.rep_10))
```

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) 2.072e-02 1.292e-04 160.343 < 2e-16 ***
## thetaC      6.424e-03 1.321e-04  48.641 < 2e-16 ***
## rhoC        2.264e-05 1.492e-04   0.152   0.8806
## tmrcaC     1.105e-03 1.860e-04   5.940 3.35e-06 ***
## thetaC:tmrcaC 3.732e-04 1.781e-04   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) 0.020624763 0.0001170758 176.16584 0.0000
## thetaC       0.006298915 0.0001222271  51.53453 0.0000
## tmrcaC       0.000914620 0.0001612620   5.67164 0.0000
## rhoC         0.000000297 0.0001325695   0.00224 0.9982
## thetaC:tmrcaC 0.000263026 0.0001559369   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
## -0.01181332
##
## Coefficients:
##             Value Std. Error t-value p-value
## (Intercept) 0.020705049 0.0001979583 104.59296 0.0000
## thetaC       0.006429287 0.0002044684  31.44392 0.0000
## rhoC        -0.000251137 0.0002041256  -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))

```

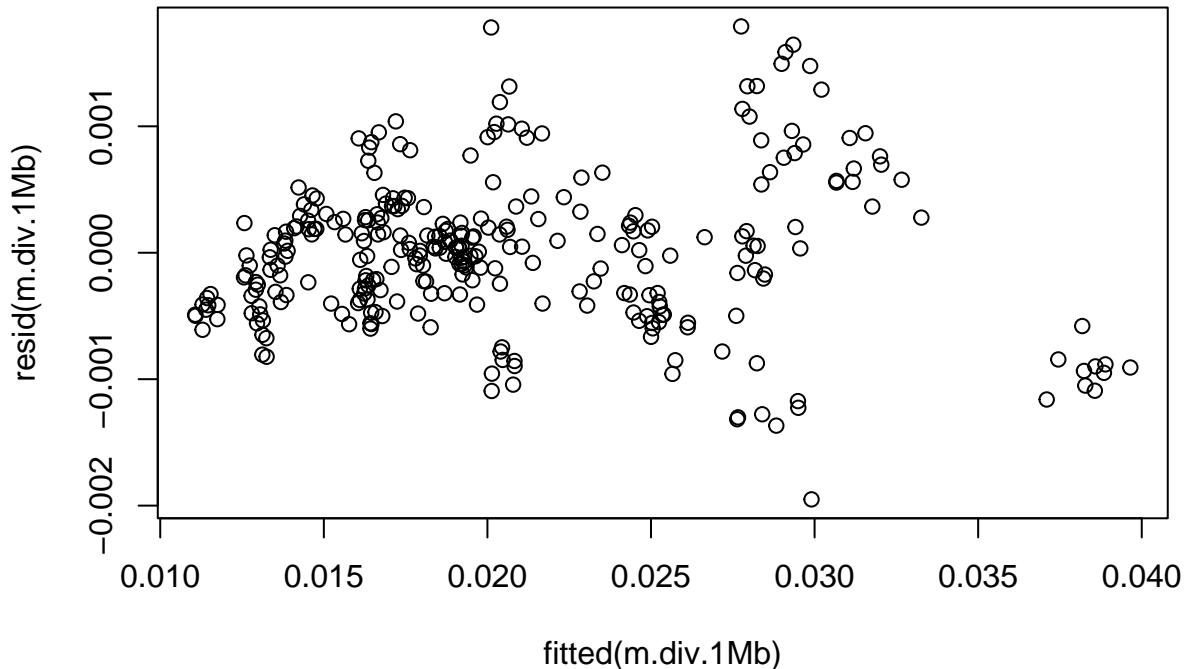
2.3.11 all replicates:

```

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)

```

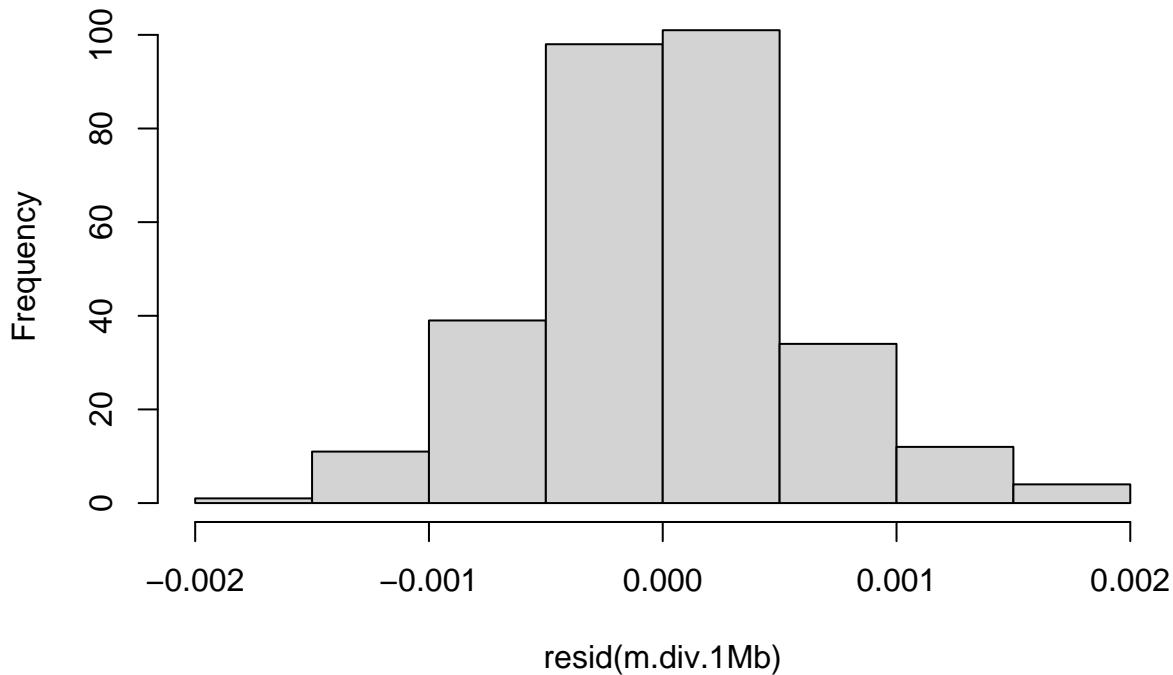
```

##
## Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.51208, p-value = 0.596

```

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

Histogram of 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.factor(Replicate)
```

```
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       6.300e-03 1.251e-04  50.357 < 2e-16 ***
## rhoC         1.166e-06 1.358e-04   0.009   0.993
## tmrcaC       9.079e-04 1.666e-04   5.451 1.20e-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     -3.448e-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            2.569e-04  1.603e-04  1.602  0.110
## thetaC:as.factor(Replicate)2  -1.117e-05  1.743e-04 -0.064  0.949
## thetaC:as.factor(Replicate)3  1.849e-05  1.762e-04  0.105  0.916
## thetaC:as.factor(Replicate)4  1.759e-04  1.735e-04  1.014  0.312
## thetaC:as.factor(Replicate)5  -8.703e-06  1.736e-04 -0.050  0.960
## thetaC:as.factor(Replicate)6  6.393e-06  1.747e-04  0.037  0.971
## thetaC:as.factor(Replicate)7  2.830e-05  1.742e-04  0.162  0.871
## thetaC:as.factor(Replicate)8  3.844e-05  1.742e-04  0.221  0.826
## thetaC:as.factor(Replicate)9  -3.451e-05  1.744e-04 -0.198  0.843
## thetaC:as.factor(Replicate)10  1.248e-04  1.739e-04  0.718  0.474
## rhoC:as.factor(Replicate)2  1.048e-04  1.991e-04  0.526  0.599
## rhoC:as.factor(Replicate)3  5.270e-05  1.925e-04  0.274  0.784
## rhoC:as.factor(Replicate)4  -1.207e-05  1.917e-04 -0.063  0.950
## rhoC:as.factor(Replicate)5  -7.350e-04  2.303e-03 -0.319  0.750
## rhoC:as.factor(Replicate)6  1.642e-04  2.033e-04  0.808  0.420
## rhoC:as.factor(Replicate)7  2.291e-05  1.956e-04  0.117  0.907
## rhoC:as.factor(Replicate)8  1.269e-04  1.937e-04  0.655  0.513
## rhoC:as.factor(Replicate)9  6.420e-05  1.943e-04  0.330  0.741
## rhoC:as.factor(Replicate)10  2.148e-05  1.924e-04  0.112  0.911
## tmrcaC:as.factor(Replicate)2  2.284e-04  2.308e-04  0.990  0.323
## tmrcaC:as.factor(Replicate)3  2.870e-04  2.201e-04  1.304  0.193
## tmrcaC:as.factor(Replicate)4  1.087e-04  2.411e-04  0.451  0.652
## tmrcaC:as.factor(Replicate)5  -8.090e-04  1.676e-04 -4.827  2.41e-06 ***
## tmrcaC:as.factor(Replicate)6  2.155e-04  2.301e-04  0.936  0.350
## tmrcaC:as.factor(Replicate)7  3.186e-04  2.374e-04  1.342  0.181
## tmrcaC:as.factor(Replicate)8  3.092e-04  2.597e-04  1.190  0.235
## tmrcaC:as.factor(Replicate)9  -1.569e-05  2.211e-04 -0.071  0.943
## tmrcaC:as.factor(Replicate)10  1.972e-04  2.380e-04  0.828  0.408
## thetaC:tmrcaC:as.factor(Replicate)2  2.147e-04  2.406e-04  0.892  0.373
## thetaC:tmrcaC:as.factor(Replicate)3  3.560e-04  2.321e-04  1.534  0.126
## thetaC:tmrcaC:as.factor(Replicate)4  9.570e-05  2.648e-04  0.361  0.718
## thetaC:tmrcaC:as.factor(Replicate)5  -2.093e-04  1.616e-04 -1.295  0.197
## thetaC:tmrcaC:as.factor(Replicate)6  1.310e-04  2.288e-04  0.573  0.567
## thetaC:tmrcaC:as.factor(Replicate)7  3.553e-04  2.339e-04  1.519  0.130
## thetaC:tmrcaC:as.factor(Replicate)8  1.804e-04  2.424e-04  0.744  0.457
## thetaC:tmrcaC:as.factor(Replicate)9  2.537e-05  2.125e-04  0.119  0.905
## thetaC:tmrcaC:as.factor(Replicate)10  1.163e-04  2.285e-04  0.509  0.611
## ---
## 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

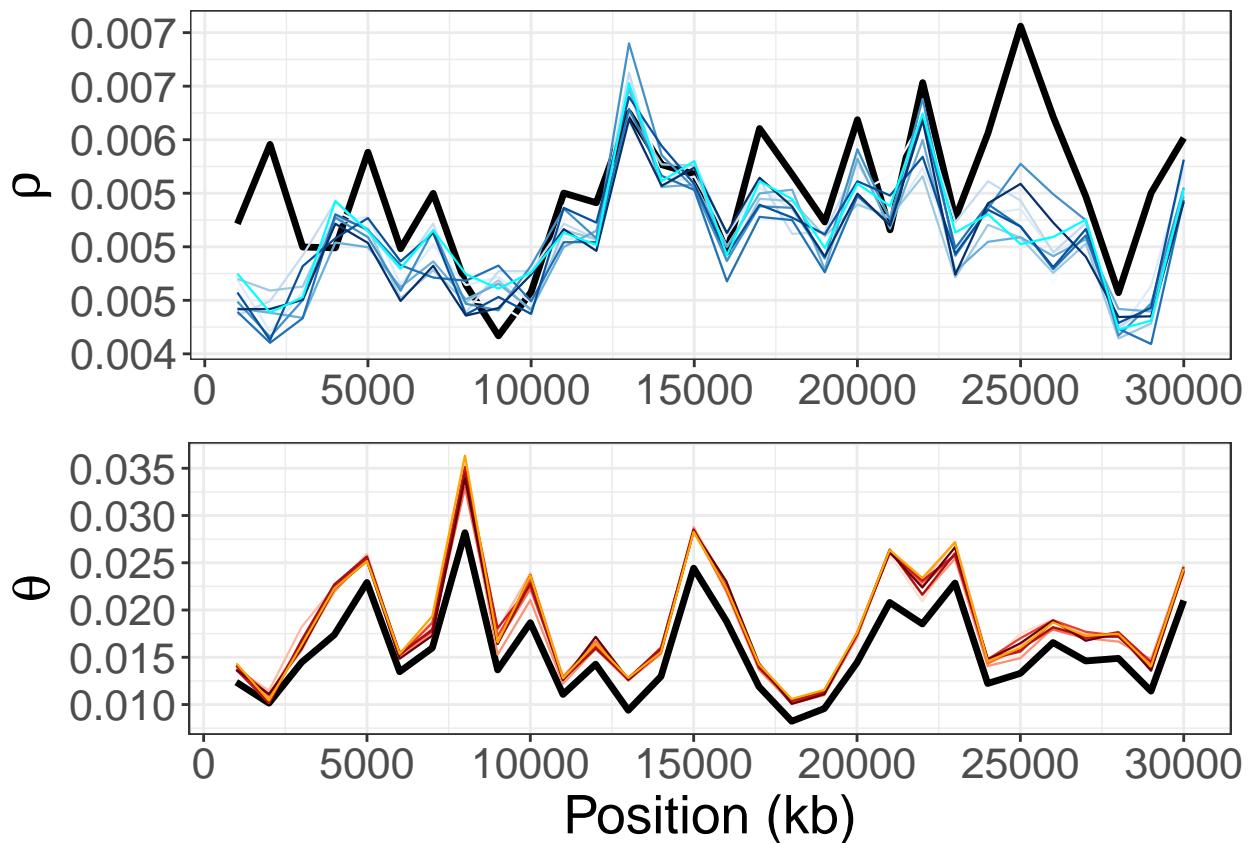
names(rho.plot) <- c("bin", "sim", reps)
molten.rho <- melt(rho.plot, id.vars = "bin")
rho.map.1Mb <- ggplot(data = molten.rho, aes(x = bin * 1000, y = value, colour = variable)) + theme_bw()
rho.map.1Mb <- rho.map.1Mb + geom_line(data = molten.rho, aes(size = variable)) + scale_size_manual(values = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10))
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))

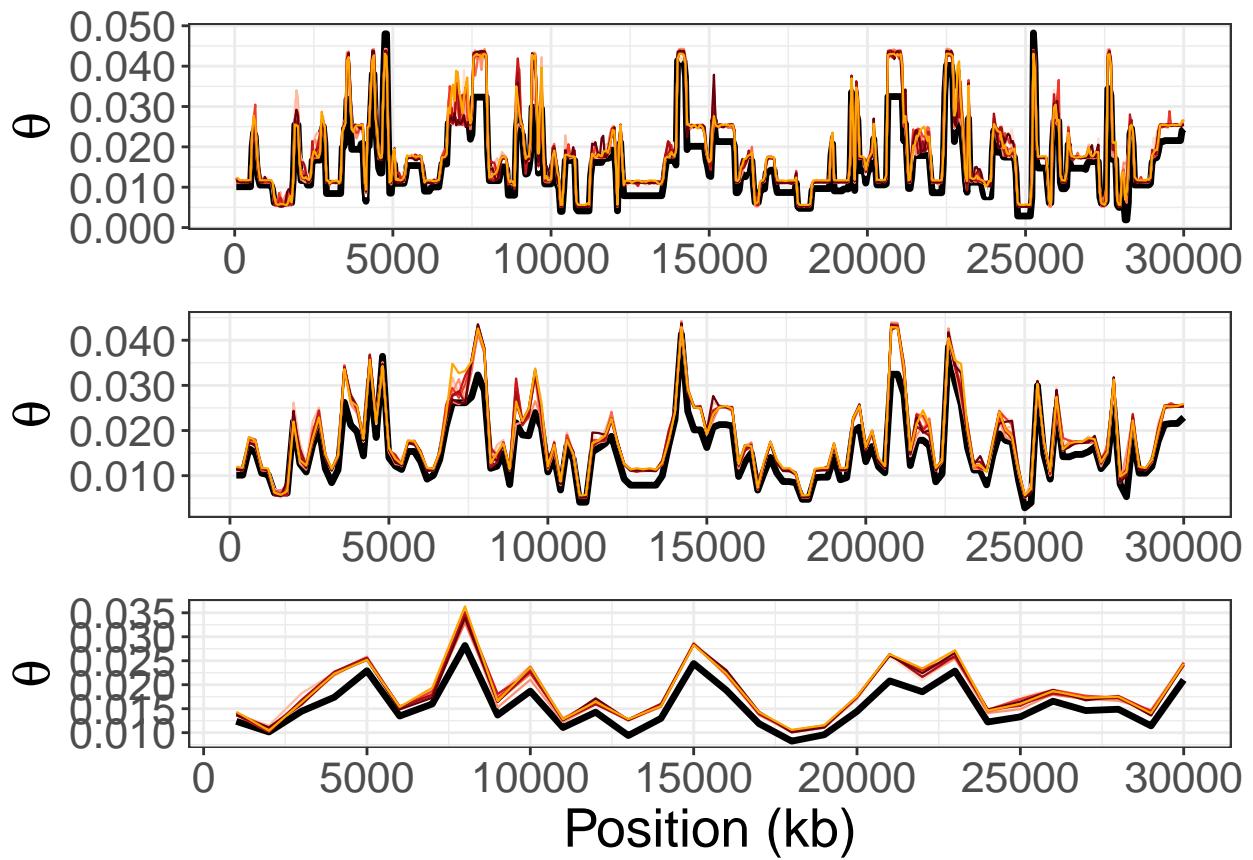
names(theta.plot) <- c("bin", "sim", reps)
molten.theta <- melt(theta.plot, id.vars = "bin")
theta.map.1Mb <- ggplot(data = molten.theta, aes(x = bin * 1000, y = value, colour = variable)) + theme_bw()
theta.map.1Mb <- theta.map.1Mb + geom_line(data = molten.theta, aes(size = variable)) + scale_size_manual(values = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10))
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)

```



```
theta.maps.coal <- plot_grid(theta.map.50kb, theta.map.200kb, theta.map.1Mb, nrow = 3, ncol = 1)
theta.maps.coal
```



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 =
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)

dm.lands.50kb <- dm.lands.50kb[which(intersect.50kb == F),] # filters based on missing data

# correlation between inferred TMRCA and Pi
cor.test(dm.lands.50kb$diversity, dm.lands.50kb$tmrca, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: dm.lands.50kb$diversity and dm.lands.50kb$tmrca
## S = 1694418, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.6983048

# correlation with Comeron map
comeron.50kb <- read.table("dm_bgs_sims/RecombinationMap_50kb.csv", sep = ",", header = T)
comeron.50kb$chrom <- "2L"
comeron.50kb_gr <- makeGRangesFromDataFrame(comeron.50kb, keep.extra.columns = T)

rho.dm.50kb$chromStart <- rho.dm.50kb$chromStart + 1
rho.dm.50kb$chrom <- "2L"
rho.dm.50kb_gr <- makeGRangesFromDataFrame(rho.dm.50kb[,c(1:3, ncol(rho.dm.50kb))], keep.extra.columns = T)

hits <- findOverlaps(query = rho.dm.50kb_gr, subject = comeron.50kb_gr, type = "equal")

comeron.50kb_gr <- comeron.50kb_gr[subjectHits(hits),]
rho.dm.50kb_gr <- rho.dm.50kb_gr[queryHits(hits),]

rec_df <- merge(as.data.frame(rho.dm.50kb_gr), as.data.frame(comeron.50kb_gr))
cor.test(rec_df[,6], rec_df[,7], method = "spearman")

##
## Spearman's rank correlation rho
##
## data: rec_df[, 6] and rec_df[, 7]
## S = 6581510, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.5942997

# OLS

# standardizing
dm.lands.50kb$thetaC <- (dm.lands.50kb$theta - mean(dm.lands.50kb$theta)) / sd(dm.lands.50kb$theta)
dm.lands.50kb$tmrcaC <- (dm.lands.50kb$tmrca - mean(dm.lands.50kb$tmrca)) / sd(dm.lands.50kb$tmrca)

```

```

dm.lands.50kb$rhoC <- (dm.lands.50kb$rho - mean(dm.lands.50kb$rho)) / sd(dm.lands.50kb$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),]

# correlation between inferred TMRCA and Pi
cor.test(dm.lands.200kb$diversity, dm.lands.200kb$tmrca, method = "spearman")

## Spearman's rank correlation rho
## data: dm.lands.200kb$diversity and dm.lands.200kb$tmrca
## S = 32240, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
```

```

##      rho
## 0.6616431
# correlation with Comeron map
comeron.200kb <- read.table("dm_bgs_sims/RecombinationMap_200kb.csv", sep = ",", header = T)
comeron.200kb$chrom <- "2L"
comeron.200kb_gr <- makeGRangesFromDataFrame(comeron.200kb, keep.extra.columns = T)

rho_dm.200kb$chromStart <- rho_dm.200kb$chromStart + 1
rho_dm.200kb$chrom <- "2L"
rho_dm.200kb_gr <- makeGRangesFromDataFrame(rho_dm.200kb[,c(1:3, ncol(rho_dm.200kb))], keep.extra.columns = T)

hits <- findOverlaps(query = rho_dm.200kb_gr, subject = comeron.200kb_gr, type = "equal")

comeron.200kb_gr <- comeron.200kb_gr[subjectHits(hits),]
rho_dm.200kb_gr <- rho_dm.200kb_gr[queryHits(hits),]

rec_df <- merge(as.data.frame(rho_dm.200kb_gr), as.data.frame(cameron.200kb_gr))
cor.test(rec_df[,6], rec_df[,7], method = "spearman")

##
## Spearman's rank correlation rho
##
## data: rec_df[, 6] and rec_df[, 7]
## S = 77773, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.6931551

# OLS

dm.lands.200kb$thetaC <- (dm.lands.200kb$theta - mean(dm.lands.200kb$theta)) / sd(dm.lands.200kb$theta)
dm.lands.200kb$tmrcaC <- (dm.lands.200kb$tmrca - mean(dm.lands.200kb$tmrca)) / sd(dm.lands.200kb$tmrca)
dm.lands.200kb$rhoC <- (dm.lands.200kb$rho - mean(dm.lands.200kb$rho)) / sd(dm.lands.200kb$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[1] * 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),]

# correlation between inferred TMRCA and Pi
cor.test(dm.lands.1Mb$diversity, dm.lands.1Mb$tmrca, method = "spearman")

##  

## Spearman's rank correlation rho  

##  

## data: dm.lands.1Mb$diversity and dm.lands.1Mb$tmrca  

## S = 156, p-value < 2.2e-16  

## alternative hypothesis: true rho is not equal to 0  

## sample estimates:  

## rho  

## 0.8631579

# correlation with Comeron map
comeron.1Mb <- read.table("dm_bgs_sims/RecombinationMap_1000kb.csv", sep = ",", header = T)
comeron.1Mb$chrom <- "2L"
comeron.1Mb_gr <- makeGRangesFromDataFrame(comeron.1Mb, keep.extra.columns = T)

rho.dm.1Mb$chromStart <- rho.dm.1Mb$chromStart + 1
rho.dm.1Mb$chrom <- "2L"
rho.dm.1Mb_gr <- makeGRangesFromDataFrame(rho.dm.1Mb[,c(1:3, ncol(rho.dm.1Mb))], keep.extra.columns = T)

hits <- findOverlaps(query = rho.dm.1Mb_gr, subject = comeron.1Mb_gr, type = "equal")

comeron.1Mb_gr <- comeron.1Mb_gr[subjectHits(hits),]
rho.dm.1Mb_gr <- rho.dm.1Mb_gr[queryHits(hits),]

rec_df <- merge(as.data.frame(rho.dm.1Mb_gr), as.data.frame(comeron.1Mb_gr))
cor.test(rec_df[,6], rec_df[,7], method = "spearman")

##
```

```

## Spearman's rank correlation rho
##
## data: rec_df[, 6] and rec_df[, 7]
## S = 274, p-value = 2.329e-06
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.8646245

# 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 =
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$theta)) / sd(dm.lands.50kb.2L$theta)
dm.lands.50kb.2L$tmrcaC <- (dm.lands.50kb.2L$tmrca - mean(dm.lands.50kb.2L$tmrca)) / sd(dm.lands.50kb.2L$tmrca)
dm.lands.50kb.2L$rhoC <- (dm.lands.50kb.2L$rho - mean(dm.lands.50kb.2L$rho)) / sd(dm.lands.50kb.2L$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), nAGQ = 100)

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) 0.009712999 1.086516e-05 893.9585 0.0000
## thetaC       0.002296955 1.174194e-05 195.6197 0.0000
## rhoC        0.000013747 1.131652e-05   1.2148 0.2253
## tmrcaC      0.000847496 1.415952e-05  59.8534 0.0000
## thetaC:tmrcaC 0.000187792 7.006312e-06  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: 0.0001704629  

## 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$theta)) / sd(dm.lands.50kb.2R$theta)  

dm.lands.50kb.2R$tmrcaC <- (dm.lands.50kb.2R$tmrca - mean(dm.lands.50kb.2R$tmrca)) / sd(dm.lands.50kb.2R$tmrca)  

dm.lands.50kb.2R$rhoC <- (dm.lands.50kb.2R$rho - mean(dm.lands.50kb.2R$rho)) / sd(dm.lands.50kb.2R$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),  

                           na.action = na.omit)  
  

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):
##     Phi1
## 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) 0.008565213 9.380579e-06 913.0793 0.0000
## thetaC       0.002258458 8.399497e-06 268.8801 0.0000
## rhoC        0.000001107 9.546538e-06   0.1160 0.9078
## tmrcaC      0.000679336 1.139247e-05   59.6303 0.0000
## thetaC:tmrcaC 0.000144456 6.673200e-06  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: 0.0001503138
## 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 =
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 =
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$theta)) / sd(dm.lands.50kb.3L$theta)
dm.lands.50kb.3L$tmrcaC <- (dm.lands.50kb.3L$tmrca - mean(dm.lands.50kb.3L$tmrca)) / sd(dm.lands.50kb.3L$tmrca)
dm.lands.50kb.3L$rhoC <- (dm.lands.50kb.3L$rho - mean(dm.lands.50kb.3L$rho)) / sd(dm.lands.50kb.3L$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),
                           na.action = na.omit)

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) 0.008955574 1.251395e-05 715.6471 0.0000
## thetaC       0.002871271 1.332006e-05 215.5600 0.0000
## rhoC        0.000024709 1.243714e-05   1.9867 0.0477
## tmrcaC      0.001241658 1.613800e-05  76.9400 0.0000
## thetaC:tmrcaC 0.000340870 9.410582e-06  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: 0.0001904737
## 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 =
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 =
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$theta)) / sd(dm.lands.50kb.3R$theta)
dm.lands.50kb.3R$tmrcaC <- (dm.lands.50kb.3R$tmrca - mean(dm.lands.50kb.3R$tmrca)) / sd(dm.lands.50kb.3R$tmrca)
dm.lands.50kb.3R$rhoC <- (dm.lands.50kb.3R$rho - mean(dm.lands.50kb.3R$rho)) / sd(dm.lands.50kb.3R$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):
##     Phi1
## 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) 0.007533415 8.949417e-06 841.7772 0.0000
## thetaC       0.002597237 8.926272e-06 290.9655 0.0000
## rhoC        -0.000014567 9.973726e-06 -1.4606 0.1449
## tmrcaC       0.001147132 1.426828e-05  80.3973 0.0000
## thetaC:tmrcaC 0.000308911 1.077443e-05  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: 0.0001618053
## 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)

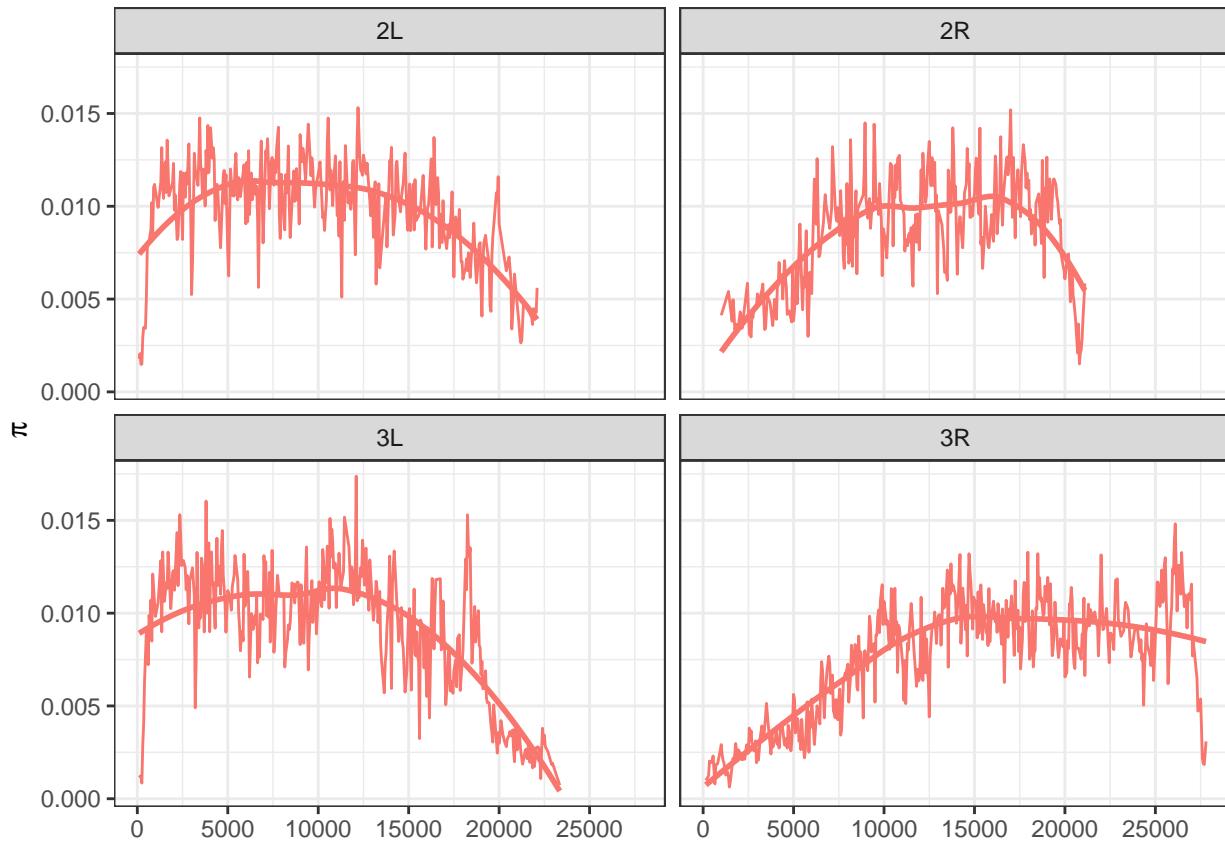
# Plots
scale.3d <- function(x) sprintf("%.3f", x) # digits shown in y axis

setTimeLimit(cpu = Inf, elapsed = Inf, transient = F)

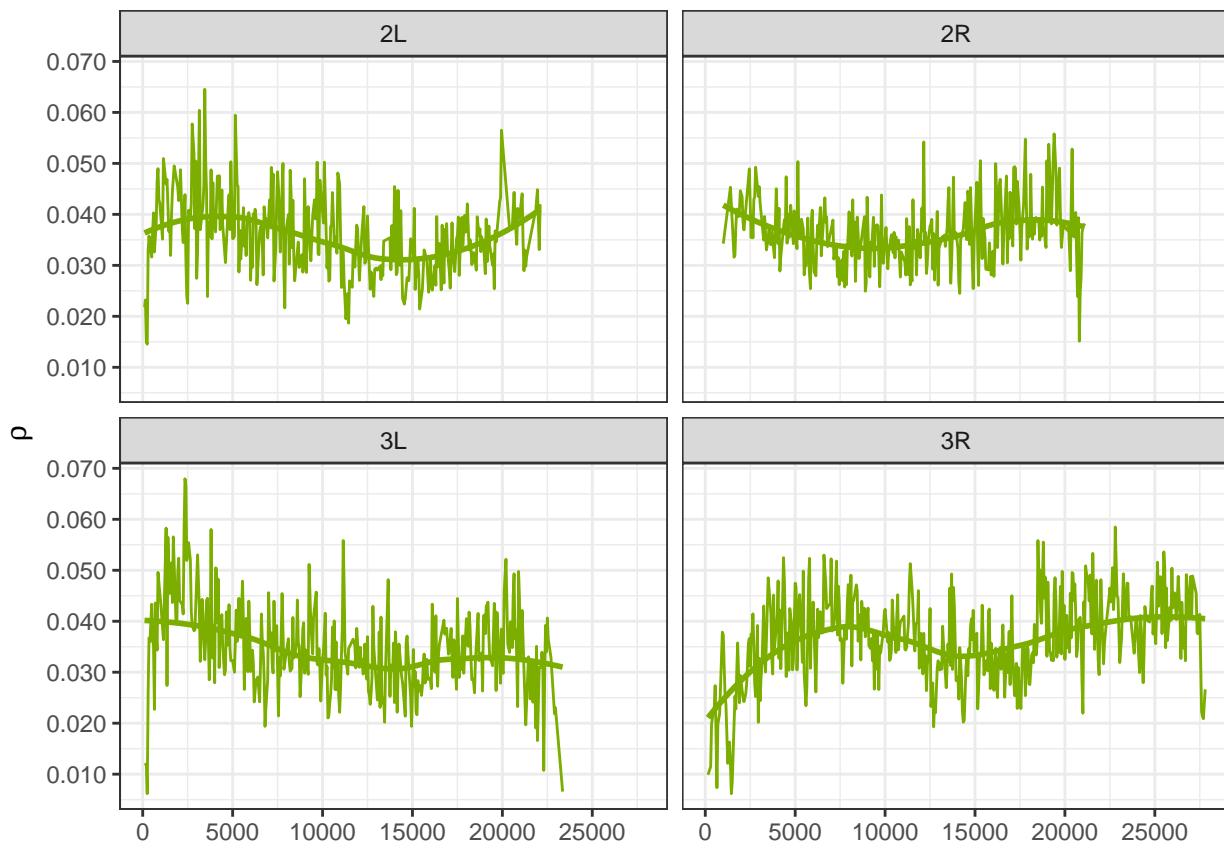
molten.diversity <- melt(dm.lands.50kb[c(3,7,8)], id.vars = c("bin", "chr"))
diversity.map <- ggplot(data = molten.diversity, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
diversity.map <- diversity.map + geom_line(data = molten.diversity, colour = "#F8766D")
diversity.map <- diversity.map + geom_smooth(method = "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
```



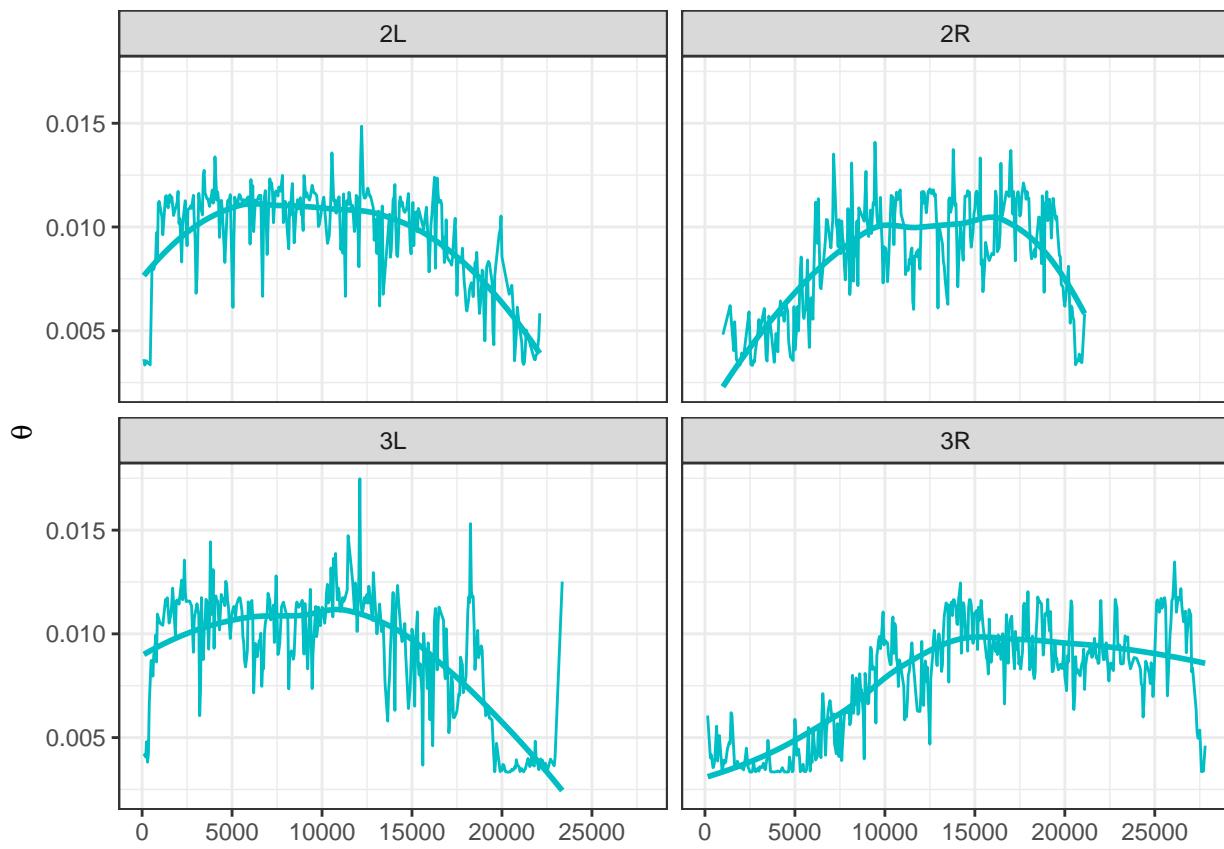
```
molten.rho <- melt(dm.lands.50kb[c(5,7,8)], id.vars = c("bin", "chr"))
rho.map <- ggplot(data = molten.rho, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
rho.map <- rho.map + geom_line(data = molten.rho, colour = "#7CAE00")
rho.map <- rho.map + geom_smooth(method = "loess", se = F, colour = "#7CAE00")
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
```



```

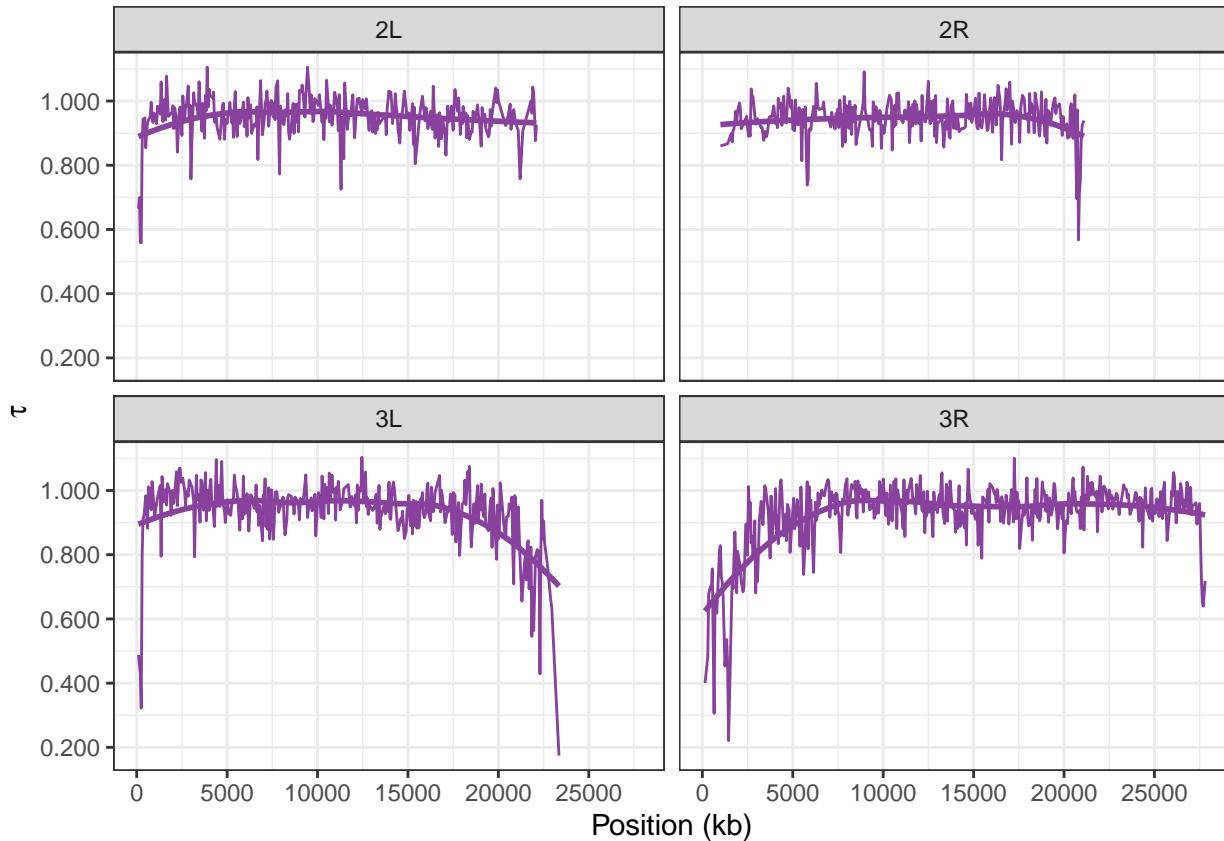
molten.theta <- melt(dm.lands.50kb[c(4,7,8)], id.vars = c("bin", "chr"))
theta.map <- ggplot(data = molten.theta, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
theta.map <- theta.map + geom_line(data = molten.theta, colour = "#00BFC4")
theta.map <- theta.map + geom_smooth(method = "loess", se = F, 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
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

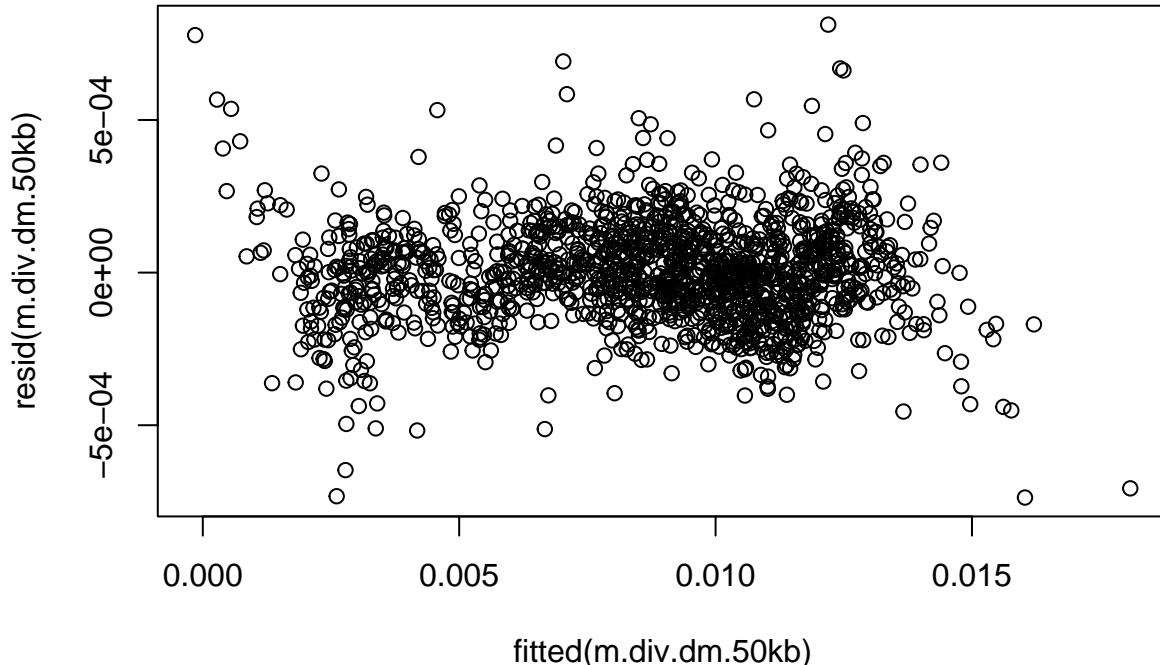
# Linear models
# standardizing
dm.lands.50kb$thetaC <- (dm.lands.50kb$theta - mean(dm.lands.50kb$theta)) / sd(dm.lands.50kb$theta)
dm.lands.50kb$tmrcaC <- (dm.lands.50kb$tmrca - mean(dm.lands.50kb$tmrca)) / sd(dm.lands.50kb$tmrca)
dm.lands.50kb$rhoC <- (dm.lands.50kb$rho - mean(dm.lands.50kb$rho)) / sd(dm.lands.50kb$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)

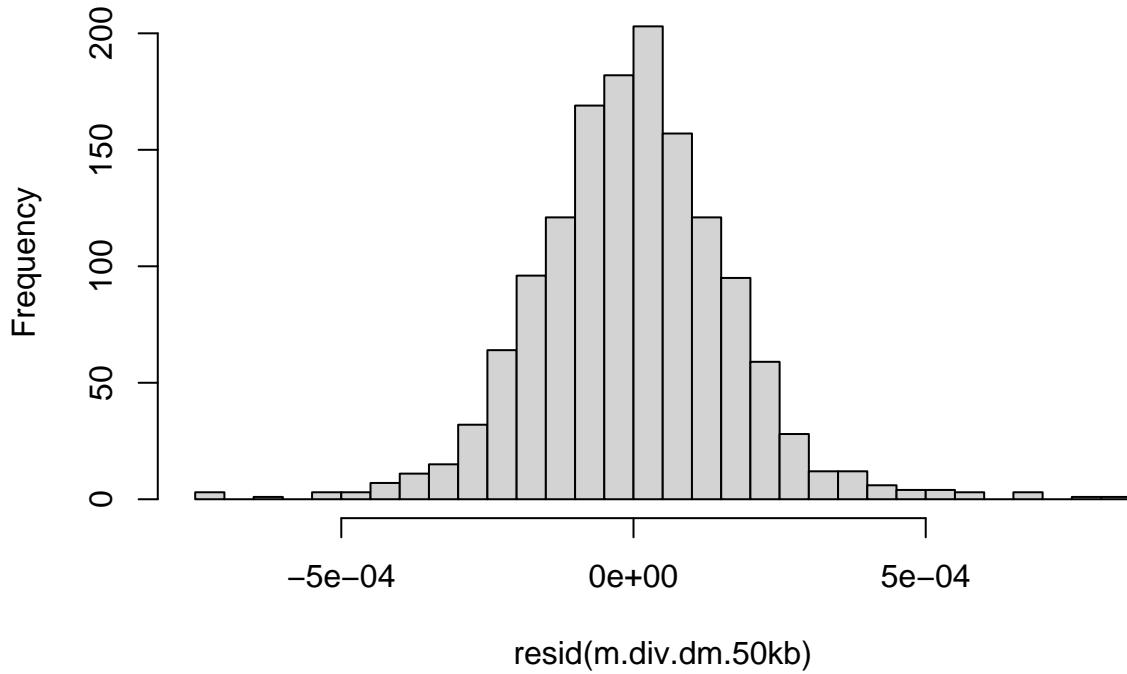
plot(resid(m.div.dm.50kb)~fitted(m.div.dm.50kb))

```



```
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
hmctest(m.div.dm.50kb, nsim = 10000)

##
##  Harrison-McCabe test
##
## data: m.div.dm.50kb
## HMC = 0.50534, p-value = 0.6103
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      2.713e-03 5.066e-06 535.514 <2e-16 ***
## rhoC       9.656e-06 5.490e-06   1.759  0.0788 .
## tmrcaC     1.027e-03 7.018e-06 146.379 <2e-16 ***
## thetaC:tmrcaC 2.670e-04 4.148e-06   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
vif(m.div.dm.50kb)

##          thetaC          rhoC          tmrcaC thetaC:tmrcaC
## 1.294662    1.520540    2.484427    1.557110

# 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 -20703.80
## g.div.dm.50kb.2 11 -20681.77
## g.div.dm.50kb.3 10 -20620.30
## g.div.dm.50kb.4 10 -20638.55

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
## -20703.8 -20645.99 10362.9
##
## Correlation Structure: AR(1)
## Formula: ~bin | chr
## Parameter estimate(s):
##     Phi
## 0.2202817
## Variance function:
## Structure: Power of variance covariate, different strata
## Formula: ~tmrcaC | chr
## Parameter estimates:
##      2L      2R      3L      3R
## 0.04460265 0.08496112 0.05902826 0.13129630
##
## Coefficients:
##             Value   Std.Error  t-value p-value
## (Intercept) 0.008591977 5.604959e-06 1532.9242 0.0000
## thetaC       0.002708023 5.808024e-06  466.2554 0.0000
## rhoC        0.000004864 5.653024e-06    0.8605 0.3897
## tmrcaC      0.001044101 7.626110e-06  136.9114 0.0000
## thetaC:tmrcaC 0.000276806 4.678853e-06   59.1612 0.0000
##
## Correlation:
##            (Intr) thetaC rhoC  tmrcaC
## thetaC      0.022
## rhoC        0.005  0.135
## tmrcaC     -0.135 -0.380 -0.482
## thetaC:tmrcaC -0.285 -0.042  0.001  0.377
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -4.25429337 -0.59755855  0.02639989  0.63307563  4.57005093
##
## Residual standard error: 0.000179039
## Degrees of freedom: 1416 total; 1411 residual
vif(g.div.dm.50kb.1)

##      thetaC      rhoC      tmrcaC thetaC:tmrcaC
## 1.195062 1.382728 1.872265 1.250111

# 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
## -16766.39 -16719.09 8392.194
##

```

```

## Correlation Structure: AR(1)
## Formula: ~bin | chr
## Parameter estimate(s):
##     Phi
## 0.2220407
## Variance function:
## Structure: Power of variance covariate, different strata
## Formula: ~thetaC | chr
## Parameter estimates:
##      2L      2R      3L      3R
## 0.15844287 0.05935656 0.14568484 0.10009402
##
## Coefficients:
##             Value Std.Error t-value p-value
## (Intercept) 0.008741224 2.144029e-05 407.7008   0
## thetaC       0.003045861 2.346944e-05 129.7798   0
## rhoC        0.000401770 1.903572e-05 21.1061   0
##
## Correlation:
##      (Intr) thetaC
## thetaC -0.028
## rhoC    0.029 -0.141
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.441445e+01 -5.435804e-01 -4.302118e-04  5.859934e-01  3.218548e+00
##
## Residual standard error: 0.0007072019
## Degrees of freedom: 1416 total; 1413 residual
vif(g.div.dm.50kb.5)

## thetaC      rhoC
## 1.020325 1.020325

```

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=TRUE)
intersect.200kb.2L <- apply(missing.prop.200kb.2L[4:ncol(missing.prop.200kb.2L)], 1, function(x) any(x))

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$theta)) / sd(dm.lands.200kb.2L)
dm.lands.200kb.2L$tmrcaC <- (dm.lands.200kb.2L$tmrca - mean(dm.lands.200kb.2L$tmrca)) / sd(dm.lands.200kb.2L)
dm.lands.200kb.2L$rhoC <- (dm.lands.200kb.2L$rho - mean(dm.lands.200kb.2L$rho)) / sd(dm.lands.200kb.2L)

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) 0.009685117 1.589668e-05 609.2540 0.0000
## thetaC       0.002127407 1.828930e-05 116.3197 0.0000
## rhoC        0.000020279 1.866681e-05   1.0864 0.2807
## tmrcaC      0.000617392 3.120432e-05  19.7855 0.0000
## thetaC:tmrcaC 0.000112009 1.103284e-05 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: 0.0001297071
## 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)

# TMRCAs 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))

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$theta)) / sd(dm.lands.200kb.2R$theta)
dm.lands.200kb.2R$tmrcaC <- (dm.lands.200kb.2R$tmrca - mean(dm.lands.200kb.2R$tmrca)) / sd(dm.lands.200kb.2R$tmrca)
dm.lands.200kb.2R$rhoC <- (dm.lands.200kb.2R$rho - mean(dm.lands.200kb.2R$rho)) / sd(dm.lands.200kb.2R$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) 0.008539183 1.263509e-05 675.8307 0.0000
## thetaC       0.002457245 1.280128e-05 191.9530 0.0000
## rhoC        0.000010548 1.124604e-05   0.9379 0.3516
## tmrcaC      0.000382626 1.512599e-05  25.2959 0.0000
## thetaC:tmrcaC 0.000082382 9.704038e-06   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: 8.89356e-05
## 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", head=TRUE)
intersect.200kb.3L <- apply(missing.prop.200kb.3L[4:ncol(missing.prop.200kb.3L)], 1, function(x) any(x))

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$theta)) / sd(dm.lands.200kb.3L$theta)
dm.lands.200kb.3L$tmrcaC <- (dm.lands.200kb.3L$tmrca - mean(dm.lands.200kb.3L$tmrca)) / sd(dm.lands.200kb.3L$tmrca)
dm.lands.200kb.3L$rhoC <- (dm.lands.200kb.3L$rho - mean(dm.lands.200kb.3L$rho)) / sd(dm.lands.200kb.3L$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) 0.008947413 1.815834e-05 492.7438 0.0000
## thetaC       0.002775488 2.084148e-05 133.1713 0.0000
## rhoC        0.000046521 1.554662e-05   2.9923 0.0037
## tmrcaC      0.000728686 2.894097e-05  25.1784 0.0000
## thetaC:tmrcaC 0.000152921 1.496396e-05 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: 0.0001102298
## 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:ncol(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:ncol(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:ncol(missing.prop.200kb.3R)], 1, function(x) any(x))

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$theta)) / sd(dm.lands.200kb.3R$theta)
dm.lands.200kb.3R$tmrcaC <- (dm.lands.200kb.3R$tmrca - mean(dm.lands.200kb.3R$tmrca)) / sd(dm.lands.200kb.3R$tmrca)
dm.lands.200kb.3R$rhoC <- (dm.lands.200kb.3R$rho - mean(dm.lands.200kb.3R$rho)) / sd(dm.lands.200kb.3R$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) 0.007566141 1.039634e-05 727.7696 0.0000
## thetaC       0.002532598 1.014172e-05 249.7209 0.0000
## rhoC        -0.000027958 1.195302e-05 -2.3390 0.0211
## tmrcaC       0.000981423 2.689006e-05  36.4976 0.0000
## thetaC:tmrcaC 0.000196288 1.830320e-05 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: 0.0001028136
## 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 = FALSE)

# Plots
scale.3d <- function(x) sprintf("%.3f", x) # digits shown in y axis

setTimeLimit(cpu = Inf, elapsed = Inf, transient = F)

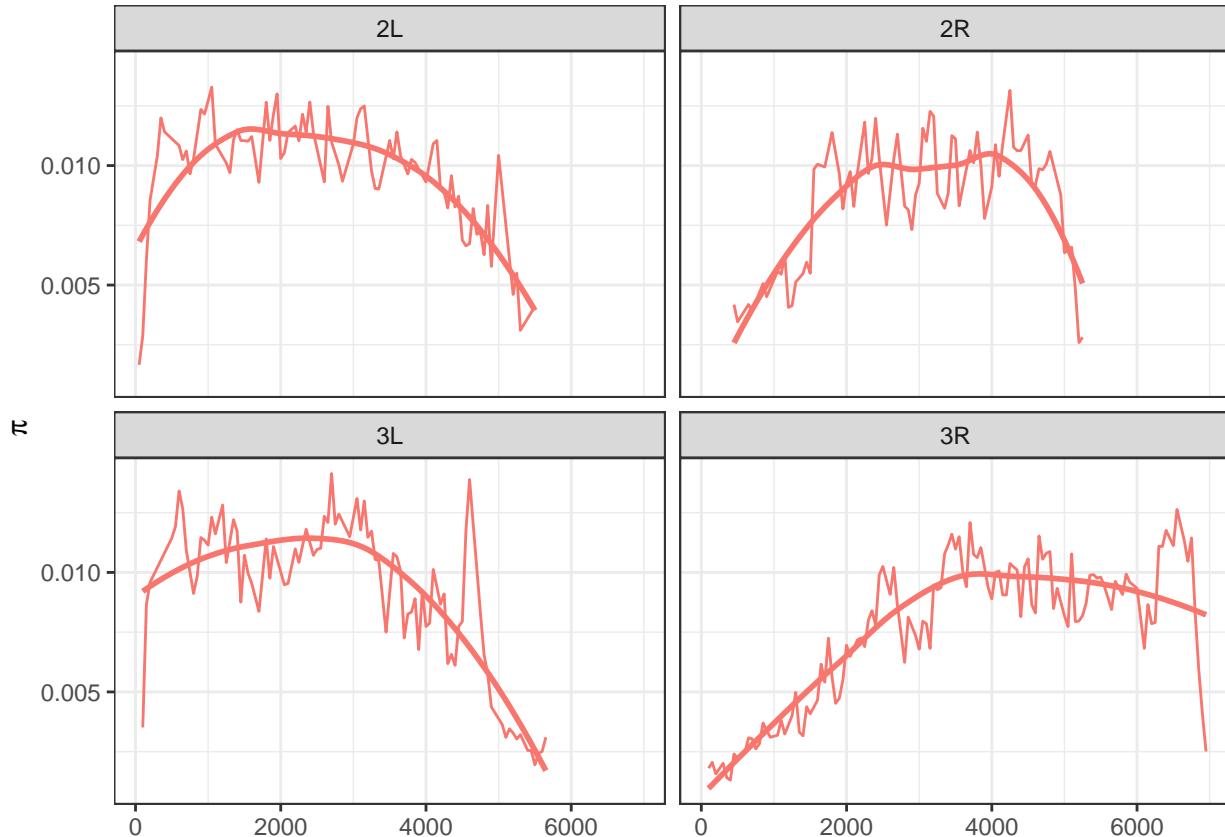
molten.diversity <- melt(dm.lands.200kb[c(3,7,8)], id.vars = c("bin", "chr"))

```

```

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

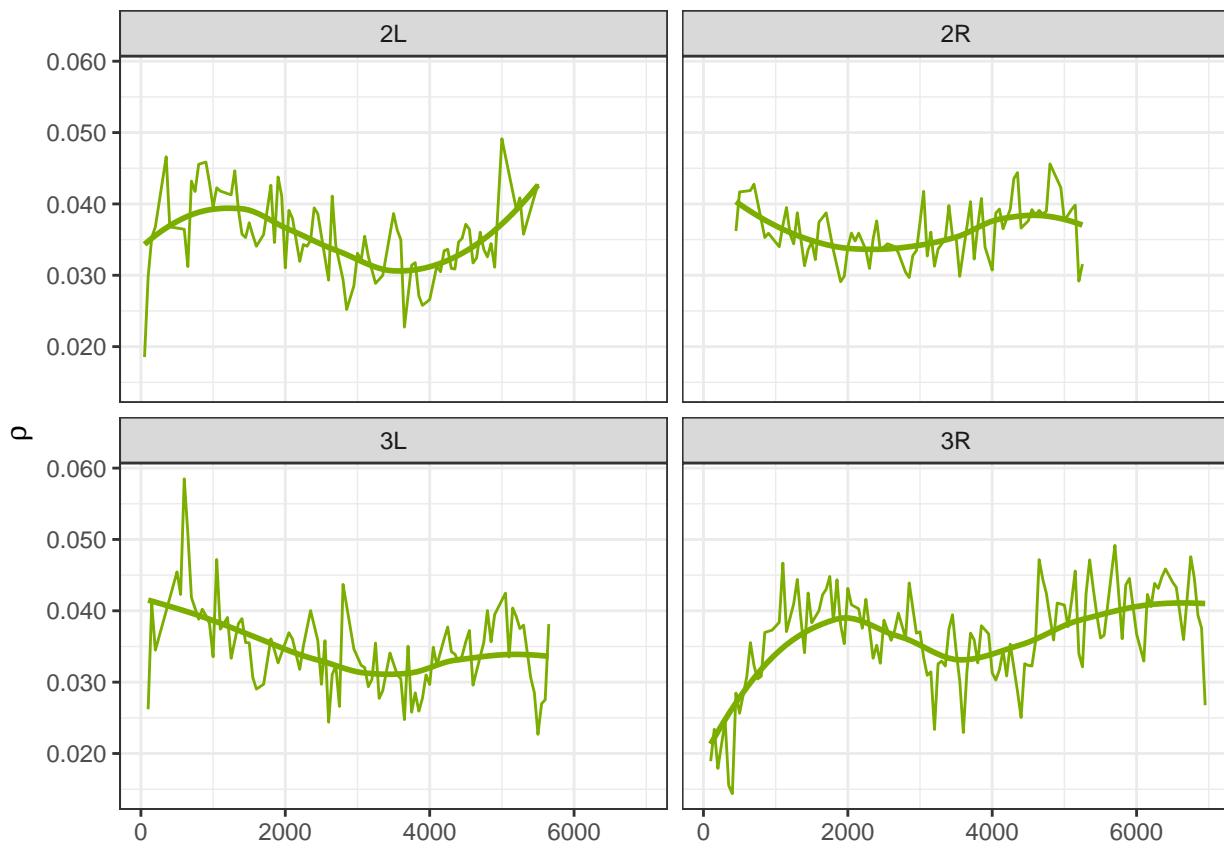
```



```

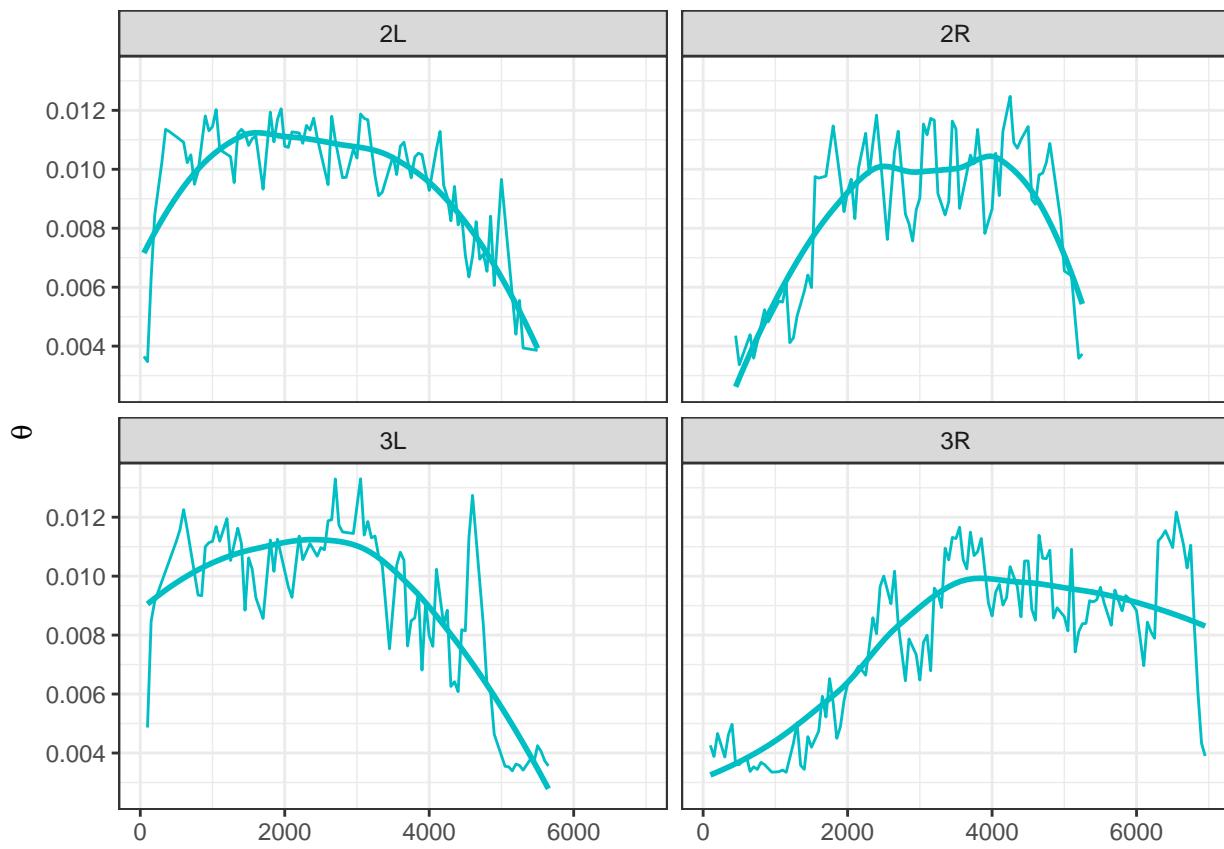
molten.rho <- melt(dm.lands.200kb[c(5,7,8)], id.vars = c("bin", "chr"))
rho.map <- ggplot(data = molten.rho, aes(x = bin * 50, y = value)) + facet_wrap(~chr)
rho.map <- rho.map + geom_line(data = molten.rho, colour = "#7CAE00")
rho.map <- rho.map + geom_smooth(method = "loess", se = F, colour = "#7CAE00")
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

```



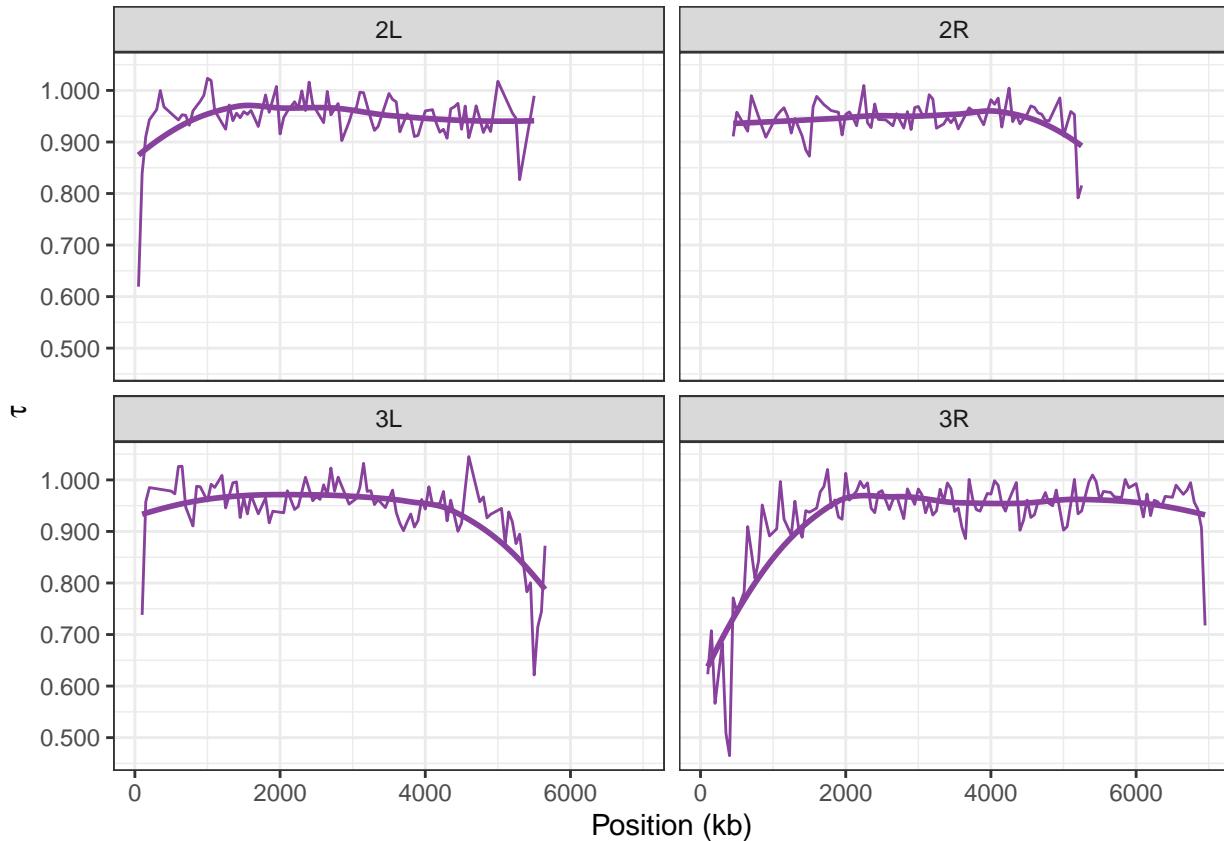
```

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

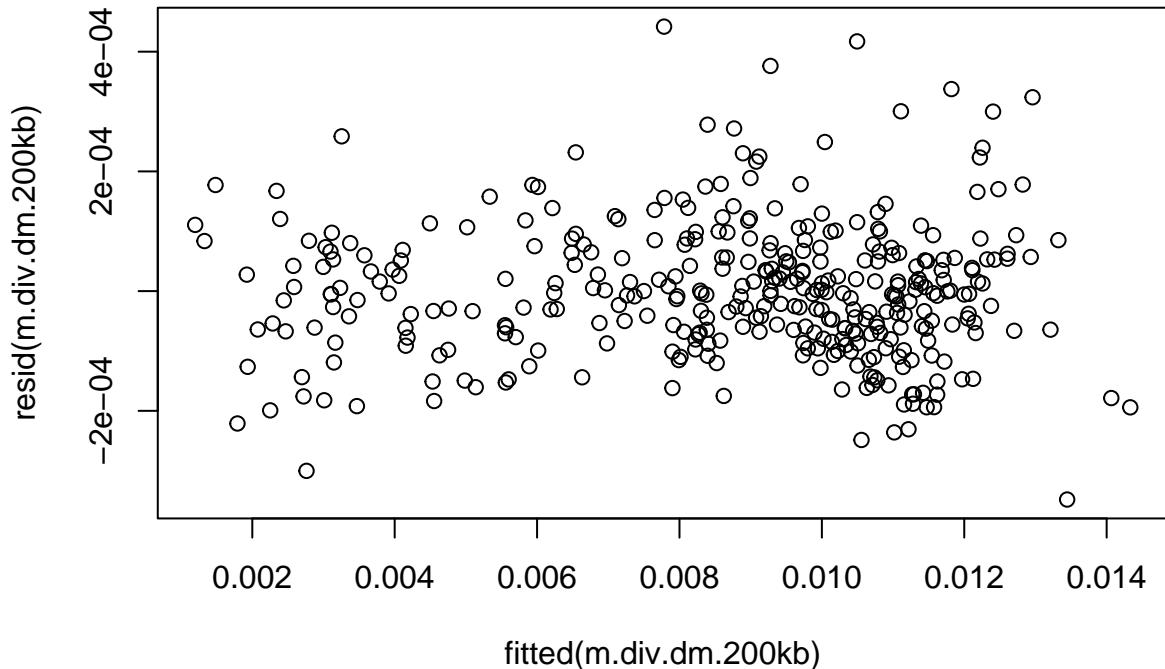
# Linear models
# standardizing
dm.lands.200kb$thetaC <- (dm.lands.200kb$theta - mean(dm.lands.200kb$theta)) / sd(dm.lands.200kb$theta)
dm.lands.200kb$tmrcaC <- (dm.lands.200kb$tmrca - mean(dm.lands.200kb$tmrca)) / sd(dm.lands.200kb$tmrca)
dm.lands.200kb$rhoC <- (dm.lands.200kb$rho - mean(dm.lands.200kb$rho)) / sd(dm.lands.200kb$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)

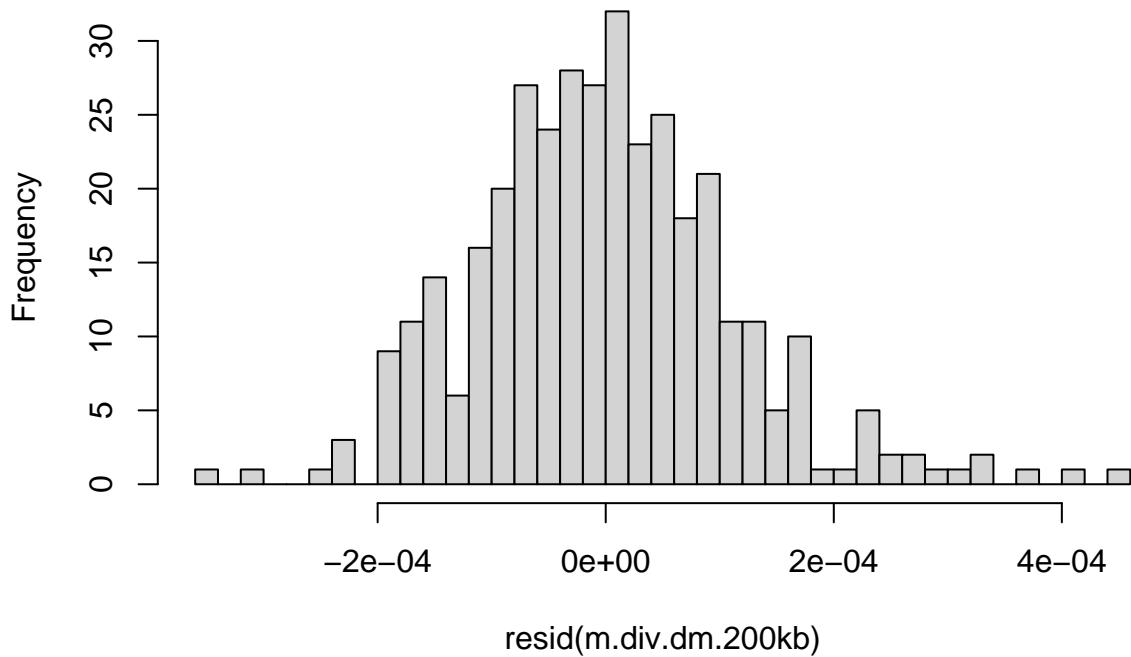
plot(resid(m.div.dm.200kb)~fitted(m.div.dm.200kb))

```



```
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  
hmctest(m.div.dm.200kb, nsim = 10000)
```

```
##
```

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

```
## data: m.div.dm.200kb  
## HMC = 0.56948, p-value = 0.9699
```

```
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      2.587e-03 7.721e-06 335.083 <2e-16 ***
## rhoC       9.685e-06 7.737e-06   1.252    0.211
## tmrcaC     7.752e-04 1.470e-05   52.753 <2e-16 ***
## thetaC:tmrcaC 1.626e-04 7.893e-06   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
vif(m.div.dm.200kb)

##          thetaC          rhoC          tmrcaC thetaC:tmrcaC
## 1.646067    1.652579    5.961897    3.721737

# 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 -5572.433
## g.div.dm.200kb.2 11 -5581.537
## g.div.dm.200kb.3 10 -5552.913
## g.div.dm.200kb.4 10 -5544.269

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
## -5572.433 -5529.625 2797.217
##
## Correlation Structure: AR(1)
## Formula: ~bin | chr
## Parameter estimate(s):
##     Phi
## 0.3007646
## Variance function:
## Structure: Power of variance covariate, different strata
## Formula: ~tmrcaC | chr
## Parameter estimates:
##      2L      2R      3L      3R
## -0.10668620 0.10590734 -0.02949358 0.09853923
##
## Coefficients:
##             Value Std.Error t-value p-value
## (Intercept) 0.008555930 8.650991e-06 989.0115 0.0000
## thetaC       0.002574166 8.570711e-06 300.3445 0.0000
## rhoC        0.000004540 8.030015e-06   0.5654 0.5721
## tmrcaC      0.000788411 1.409514e-05  55.9350 0.0000
## thetaC:tmrcaC 0.000173631 7.825353e-06 22.1882 0.0000
##
## Correlation:
##            (Intr) thetaC rhoC  tmrcaC
## thetaC      0.106
## rhoC        -0.002  0.180
## tmrcaC     -0.299 -0.475 -0.447
## thetaC:tmrcaC -0.434 -0.205 -0.075  0.705
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.72141057 -0.58127824  0.07056529  0.58993673  3.65529501
##
## Residual standard error: 0.0001151488
## Degrees of freedom: 362 total; 357 residual
vif(g.div.dm.200kb.1)

##      thetaC      rhoC      tmrcaC thetaC:tmrcaC
## 1.373145 1.485516 3.683701 2.461692

# 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) 0.008688913 3.045635e-05 285.29066      0
## thetaC       0.002853184 3.097921e-05  92.09995      0
## rhoC        0.000276861 2.255108e-05 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: 0.0004130832
## Degrees of freedom: 362 total; 359 residual
vif(g.div.dm.200kb.5)

##     thetaC     rhoC
## 1.021358 1.021358

```

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 = TRUE)
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$theta)) / sd(dm.lands.1Mb.2L$theta)
dm.lands.1Mb.2L$tmrcaC <- (dm.lands.1Mb.2L$tmrca - mean(dm.lands.1Mb.2L$tmrca)) / sd(dm.lands.1Mb.2L$tmrca)
dm.lands.1Mb.2L$rhoC <- (dm.lands.1Mb.2L$rho - mean(dm.lands.1Mb.2L$rho)) / sd(dm.lands.1Mb.2L$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), na.action = na.omit)

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) 0.010112367 2.513180e-05 402.3734 0.0000
## thetaC       0.001482486 3.056196e-05 48.5076 0.0000
## rhoC        0.000004499 2.386795e-05   0.1885 0.8532
## tmrcaC      0.000367251 3.471251e-05 10.5798 0.0000
## thetaC:tmrcaC 0.000078084 1.341728e-05    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: 6.059571e-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$theta)) / sd(dm.lands.1Mb.2R$theta)
dm.lands.1Mb.2R$tmrcaC <- (dm.lands.1Mb.2R$tmrca - mean(dm.lands.1Mb.2R$tmrca)) / sd(dm.lands.1Mb.2R$tmrca)
dm.lands.1Mb.2R$rhoC <- (dm.lands.1Mb.2R$rho - mean(dm.lands.1Mb.2R$rho)) / sd(dm.lands.1Mb.2R$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), na.action = na.omit)

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
## -0.0218456
## Variance function:
## Structure: Power of variance covariate
## Formula: ~tmrcaC
## Parameter estimates:
##     power
## -0.286626
##
## Coefficients:
##             Value    Std.Error  t-value p-value
## (Intercept) 0.009155188 2.832317e-05 323.2402 0.0000
## thetaC       0.001868053 4.094254e-05 45.6262 0.0000
## rhoC        0.000023538 2.409532e-05   0.9769 0.3479
## tmrcaC      0.0000200999 3.703936e-05   5.4266 0.0002
## thetaC:tmrcaC 0.000019395 1.900705e-05   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: 4.577664e-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 = TRUE)
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$theta)) / sd(dm.lands.1Mb.3L$theta)
dm.lands.1Mb.3L$tmrcaC <- (dm.lands.1Mb.3L$tmrca - mean(dm.lands.1Mb.3L$tmrca)) / sd(dm.lands.1Mb.3L$tmrca)
dm.lands.1Mb.3L$rhoC <- (dm.lands.1Mb.3L$rho - mean(dm.lands.1Mb.3L$rho)) / sd(dm.lands.1Mb.3L$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), na.action = na.omit)

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) 0.009181282 2.404774e-05 381.7940 0.0000
## thetaC       0.002615494 3.244090e-05  80.6233 0.0000
## rhoC        0.000034900 1.950167e-05   1.7896 0.0913
## tmrcaC      0.000517909 5.784910e-05   8.9528 0.0000
## thetaC:tmrcaC 0.000062953 2.655589e-05   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: 8.201734e-05
## 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$theta)) / sd(dm.lands.1Mb.3R$theta)
dm.lands.1Mb.3R$tmrcaC <- (dm.lands.1Mb.3R$tmrca - mean(dm.lands.1Mb.3R$tmrca)) / sd(dm.lands.1Mb.3R$tmrca)
dm.lands.1Mb.3R$rhoC <- (dm.lands.1Mb.3R$rho - mean(dm.lands.1Mb.3R$rho)) / sd(dm.lands.1Mb.3R$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), na.action = na.omit)

```

```

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) 0.006724616 2.251544e-05 298.66683 0.0000
## thetaC       0.002582348 1.759737e-05 146.74629 0.0000
## rhoC        0.000018618 1.440997e-05   1.29200 0.2173
## tmrcaC      0.000969833 5.083886e-05  19.07661 0.0000
## thetaC:tmrcaC 0.000203782 4.894035e-05   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: 6.652517e-05
## 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)

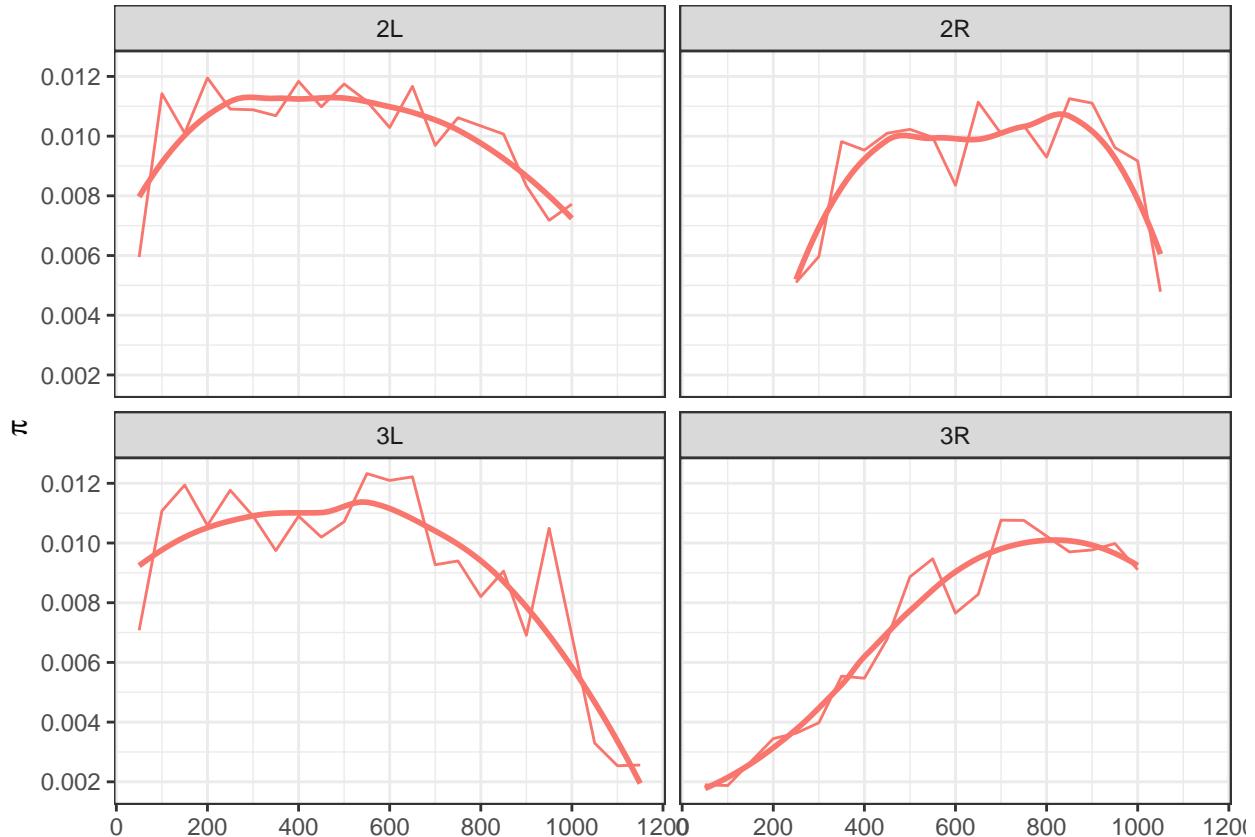
molten.diversity <- melt(dm.lands.1Mb[c(3,7,8)], id.vars = c("bin", "chr"))

```

```

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

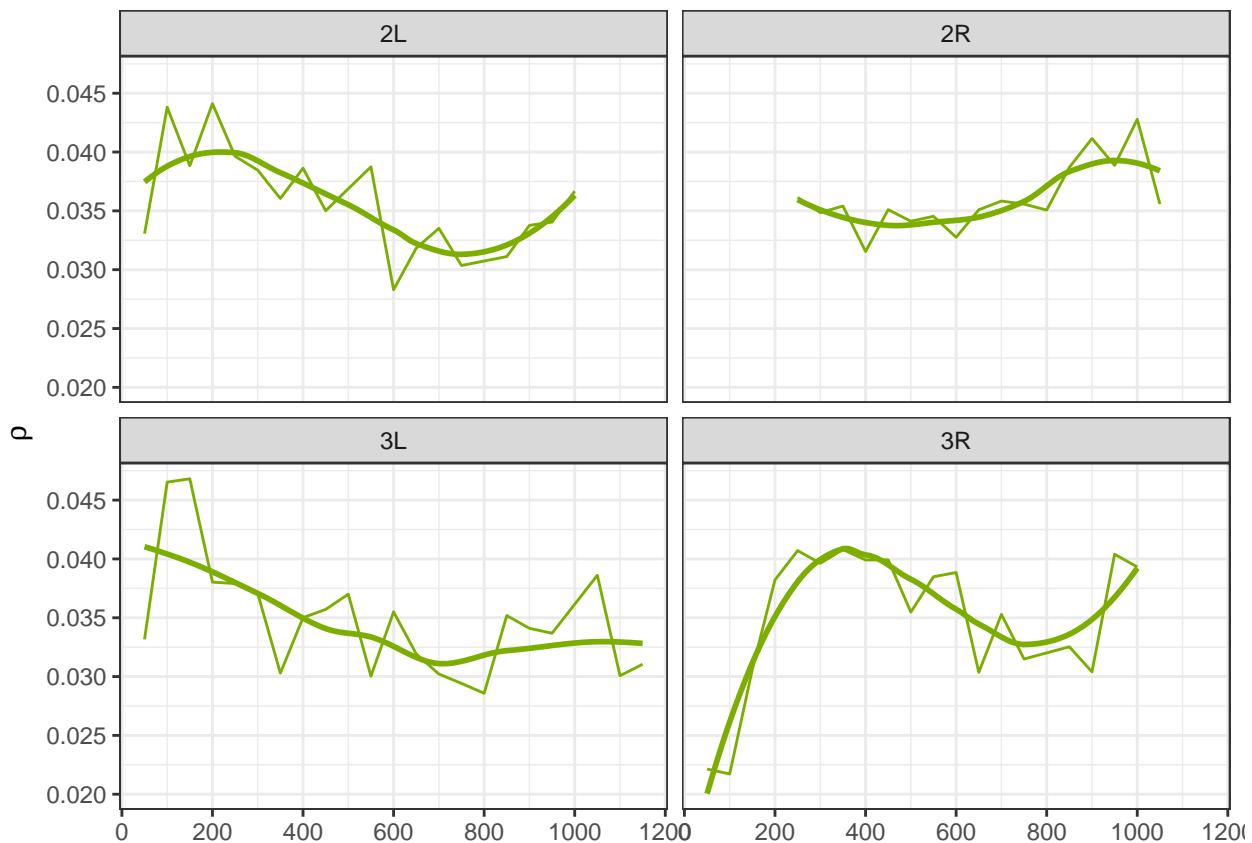
```



```

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

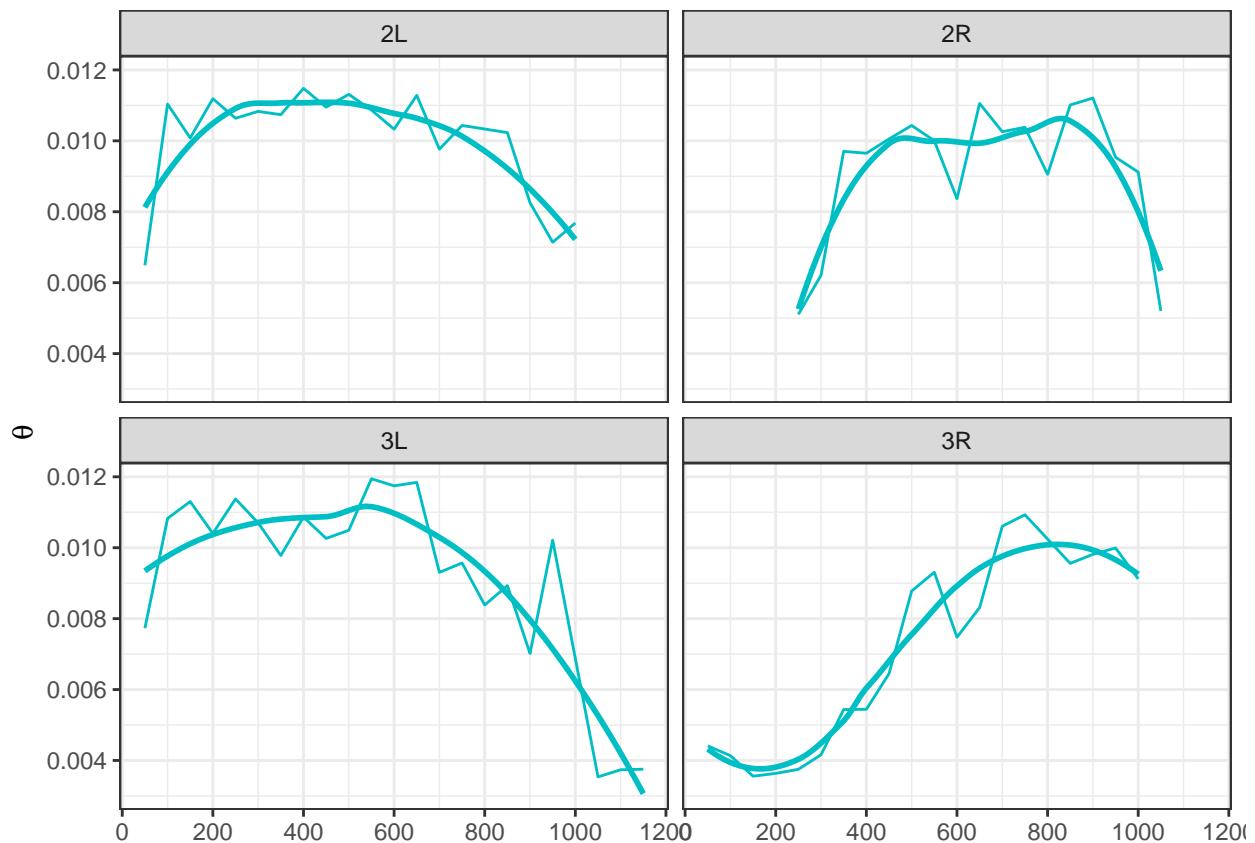
```



```

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

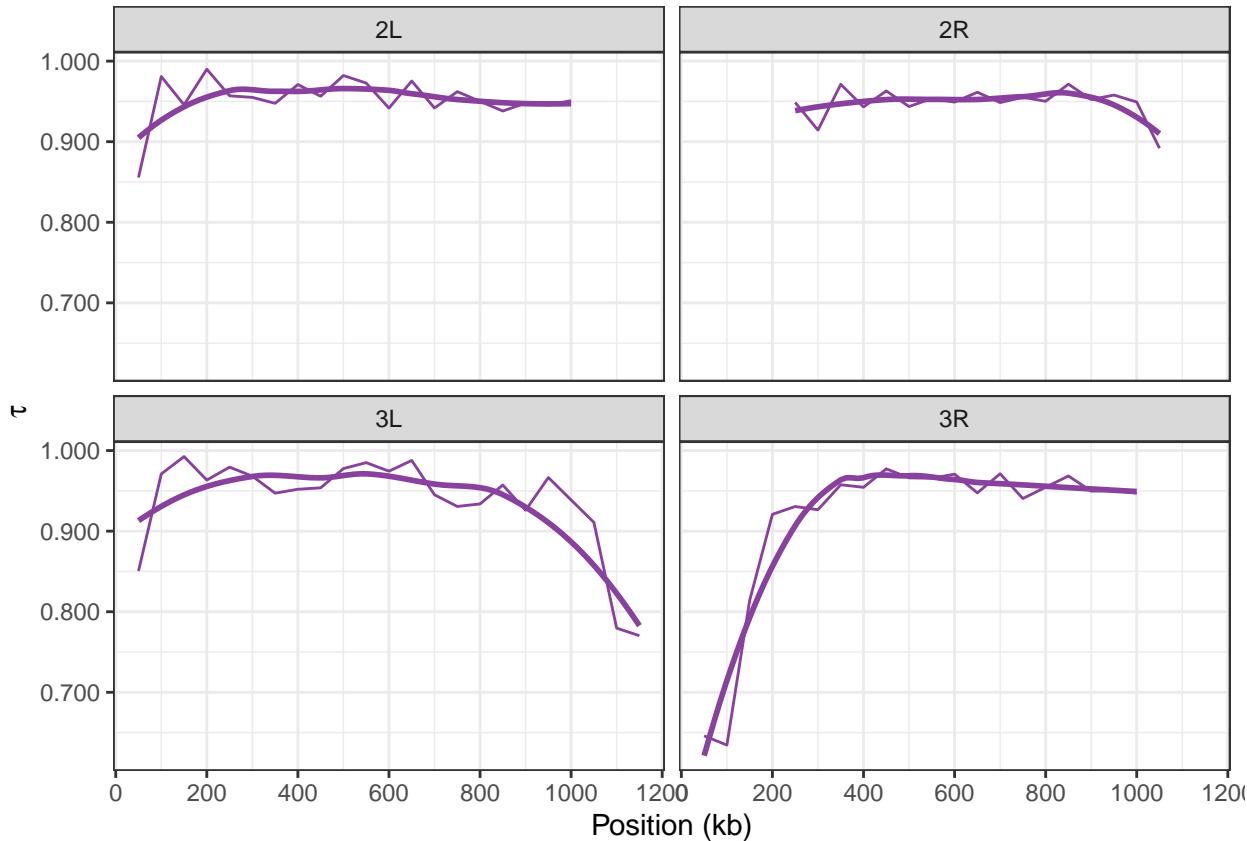
```



```

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

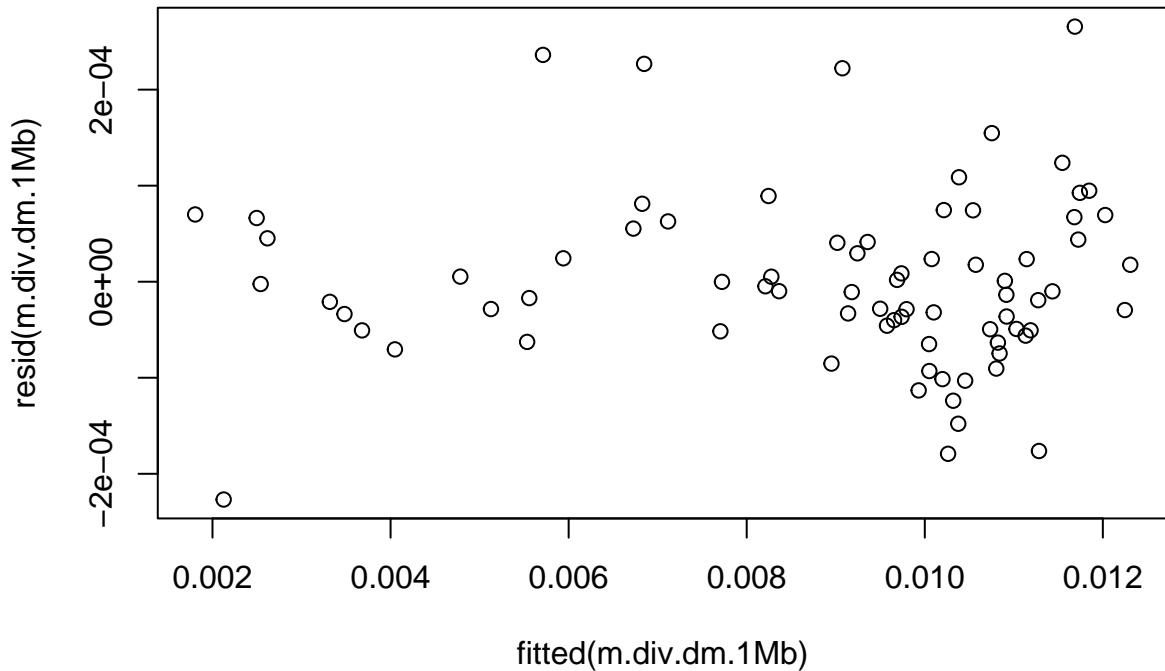
# Linear models
# standardizing
dm.lands.1Mb$thetaC <- (dm.lands.1Mb$theta - mean(dm.lands.1Mb$theta)) / sd(dm.lands.1Mb$theta)
dm.lands.1Mb$tmrcaC <- (dm.lands.1Mb$tmrca - mean(dm.lands.1Mb$tmrca)) / sd(dm.lands.1Mb$tmrca)
dm.lands.1Mb$rhoC <- (dm.lands.1Mb$rho - mean(dm.lands.1Mb$rho)) / sd(dm.lands.1Mb$rho)

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

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

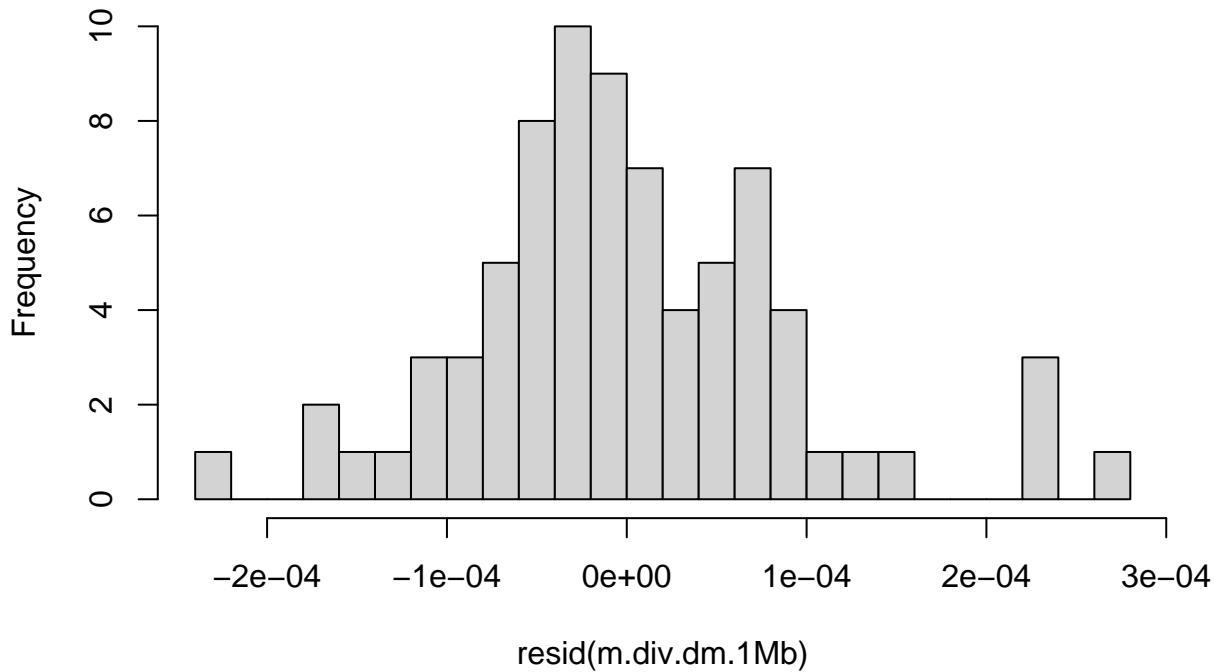
plot(resid(m.div.dm.1Mb)~fitted(m.div.dm.1Mb))

```



```
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 = 0.99993
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      2.445e-03  1.657e-05 147.497 < 2e-16 ***
## rhoC       3.015e-05  1.405e-05   2.145  0.0353 *
## tmrcaC     6.043e-04  3.829e-05 15.784 < 2e-16 ***
## thetaC:tmrcaC 7.632e-05  1.769e-05   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
vif(m.div.dm.1Mb)

##          thetaC          rhoC          tmrcaC thetaC:tmrcaC
## 2.422826    1.741910    12.927873    7.711517

# 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 -1225.918
## g.div.dm.1Mb.2 11 -1202.198
## g.div.dm.1Mb.3 10 -1202.415
## g.div.dm.1Mb.4 10 -1217.967

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
## -1225.918 -1200.136 623.9591
##
## Correlation Structure: AR(1)
## Formula: ~bin | chr
## Parameter estimate(s):
##     Phi
## 0.581682
## Variance function:
## Structure: Power of variance covariate, different strata
## Formula: ~tmrcaC | chr
## Parameter estimates:
##      2L      2R      3L      3R
## 0.37995450 0.05817476 0.46667745 0.77367312
##
## Coefficients:
##             Value   Std.Error t-value p-value
## (Intercept) 0.008752066 8.873720e-06 986.2903 0.0000
## thetaC       0.002417340 9.855870e-06 245.2691 0.0000
## rhoC        0.000002480 9.655180e-06   0.2568 0.7980
## tmrcaC      0.000648198 3.397848e-05  19.0767 0.0000
## thetaC:tmrcaC 0.000090139 2.387994e-05   3.7747 0.0003
##
## Correlation:
##            (Intr) thetaC rhoC  tmrcaC
## thetaC      0.008
## rhoC        0.477  0.584
## tmrcaC     -0.152 -0.488 -0.546
## thetaC:tmrcaC -0.114 -0.148 -0.237  0.438
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.0805313 -0.1152080  0.4338327  1.0150849  2.7175860
##
## Residual standard error: 0.0001394046
## Degrees of freedom: 77 total; 72 residual
vif(g.div.dm.1Mb.1)

##      thetaC      rhoC      tmrcaC thetaC:tmrcaC
## 1.630894 1.757698 1.789243 1.247969

# 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
## -1054.675 -1033.581 536.3375
##

```

```

## Correlation Structure: AR(1)
## Formula: ~bin | chr
## Parameter estimate(s):
##     Phi
## 0.5686675
## Variance function:
## Structure: Power of variance covariate, different strata
## Formula: ~thetaC | chr
## Parameter estimates:
##      2L      2R      3L      3R
## 0.3737074 0.1255485 0.2621324 0.5031969
##
## Coefficients:
##             Value Std.Error t-value p-value
## (Intercept) 0.008916718 3.759919e-05 237.15183      0
## thetaC       0.002695004 4.945118e-05  54.49828      0
## rhoC        0.000153498 3.164833e-05   4.85011      0
##
## Correlation:
##      (Intr) thetaC
## thetaC -0.071
## rhoC    0.043 -0.282
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -3.85658001 -0.62173921 -0.01623445  0.46158057  1.27574549
##
## Residual standard error: 0.0003189399
## Degrees of freedom: 77 total; 74 residual
vif(g.div.dm.1Mb.5)

##     thetaC     rhoC
## 1.086593 1.086593

```

3.4 Divergence with *D. yakuba*

```

lands.divergence.dm <- read.table("dm_data/dm_maps_50kb_divergence.tsv", header = T, sep = "\t")

cor.test(lands.divergence.dm$divergence, lands.divergence.dm$theta, method = "spearman")

##
## Spearman's rank correlation rho
##
## data: lands.divergence.dm$divergence and lands.divergence.dm$theta
## S = 97624952, p-value = 2.989e-09
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##     rho
## 0.1965015

```

3.5 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 <- ddply(dm.raw, .variables = "GeneID", .fun = plyr::summarize,
  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)

# standardizing
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)

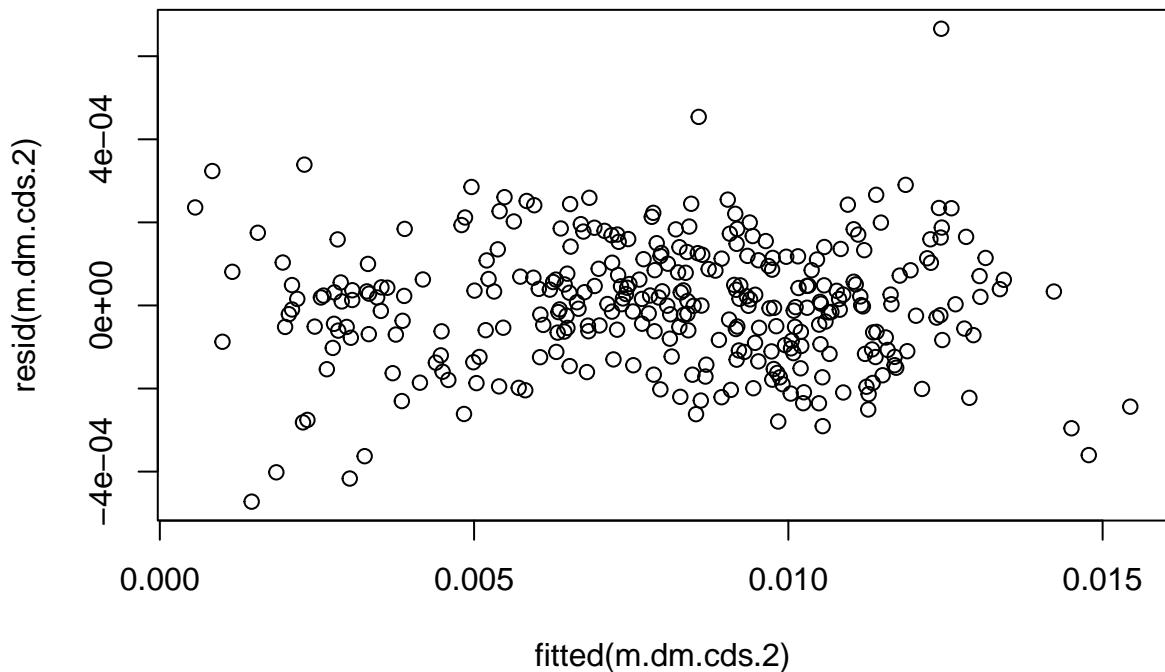
##
## Call:
## lm(formula = diversity ~ (thetaC + rhoC + tmrcaC + thetaC * tmrcaC +
##     rhoC * tmrcaC), data = dm.lands.genes)
##
## Residuals:
##      Min        1Q    Median        3Q       Max
## -4.723e-04 -9.596e-05  3.810e-06  8.994e-05  6.661e-04

```

```

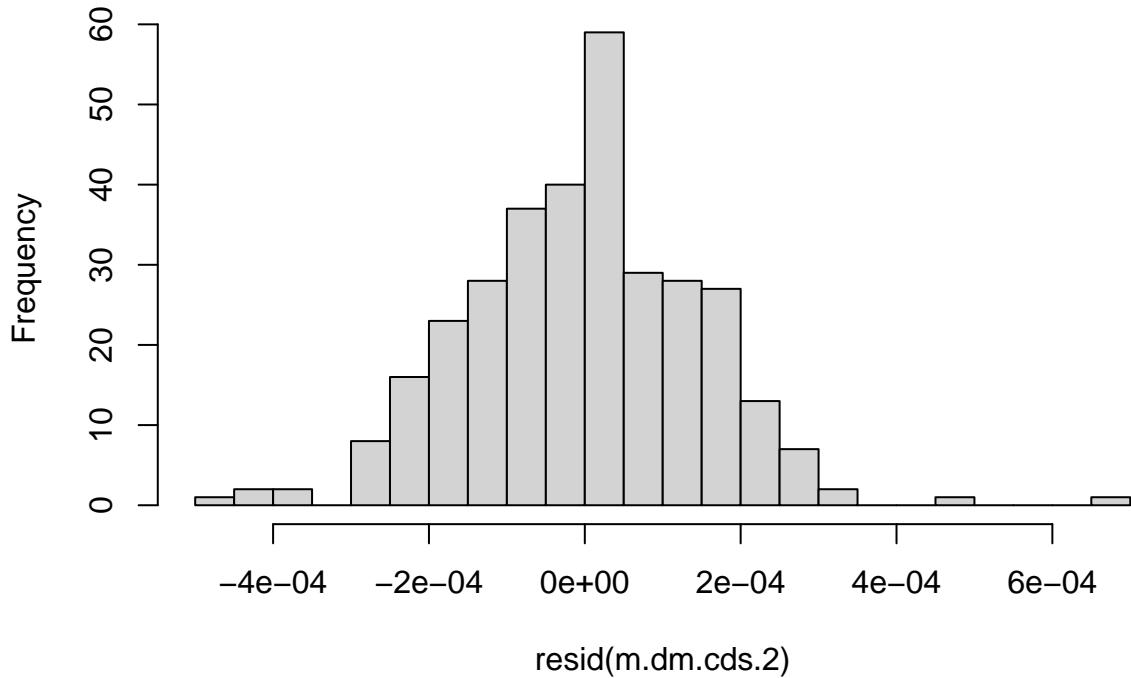
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)      7.890e-03  9.716e-06 812.013   <2e-16 ***
## thetaC          9.511e-01  3.665e-03 259.551   <2e-16 ***
## rhoC           1.053e-03  1.530e-03   0.688   0.4919    
## tmrcaC         1.120e-02  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
hmctest(m.dm.cds.2, nsim = 10000)

##
##  Harrison-McCabe test
##
## data: m.dm.cds.2
## HMC = 0.41117, p-value = 0.0117
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$the
```

```

##      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$PiN, dm.lands.evolrate.bins$theta, dm.lands.evolrate.bins$tmrca, method = "spearman")

##      estimate      p.value statistic     n gp   Method
## 1 0.2711637 6.149318e-22  9.815788 1217  1 spearman
pcor.test(dm.lands.evolrate.bins$PiS, dm.lands.evolrate.bins$theta, dm.lands.evolrate.bins$tmrca, method = "spearman")

##      estimate      p.value statistic     n gp   Method
## 1 0.6767774 1.129784e-163 32.03071 1217  1 spearman

# bootstrapping PiN and PiS correlations
boot.tbl <- dm.lands.evolrate.bins %>% dplyr::select(PiN.window, theta, PiS.window, tmrca)
boot.PiN.theta <- boot(boot.tbl,
  statistic = function(boot.tbl, i) {
    pcor.test(boot.tbl[i, "PiN.window"], boot.tbl[i, "theta"], boot.tbl[i, "tmrca"], method='spearman')
  },
  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%   ( 0.2193,  0.3245 )
## Calculations and Intervals on Original Scale

boot.PiS.theta <- boot(boot.tbl,
  statistic = function(boot.tbl, i) {
    pcor.test(boot.tbl[i, "PiS.window"], boot.tbl[i, "theta"], boot.tbl[i, "tmrca"], method='spearman')
  },
  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%   ( 0.6408,  0.7154 )
## Calculations and Intervals on Original Scale

```

4 R^2 plot with real, simulated (true) and simulated (inferred) data.

```

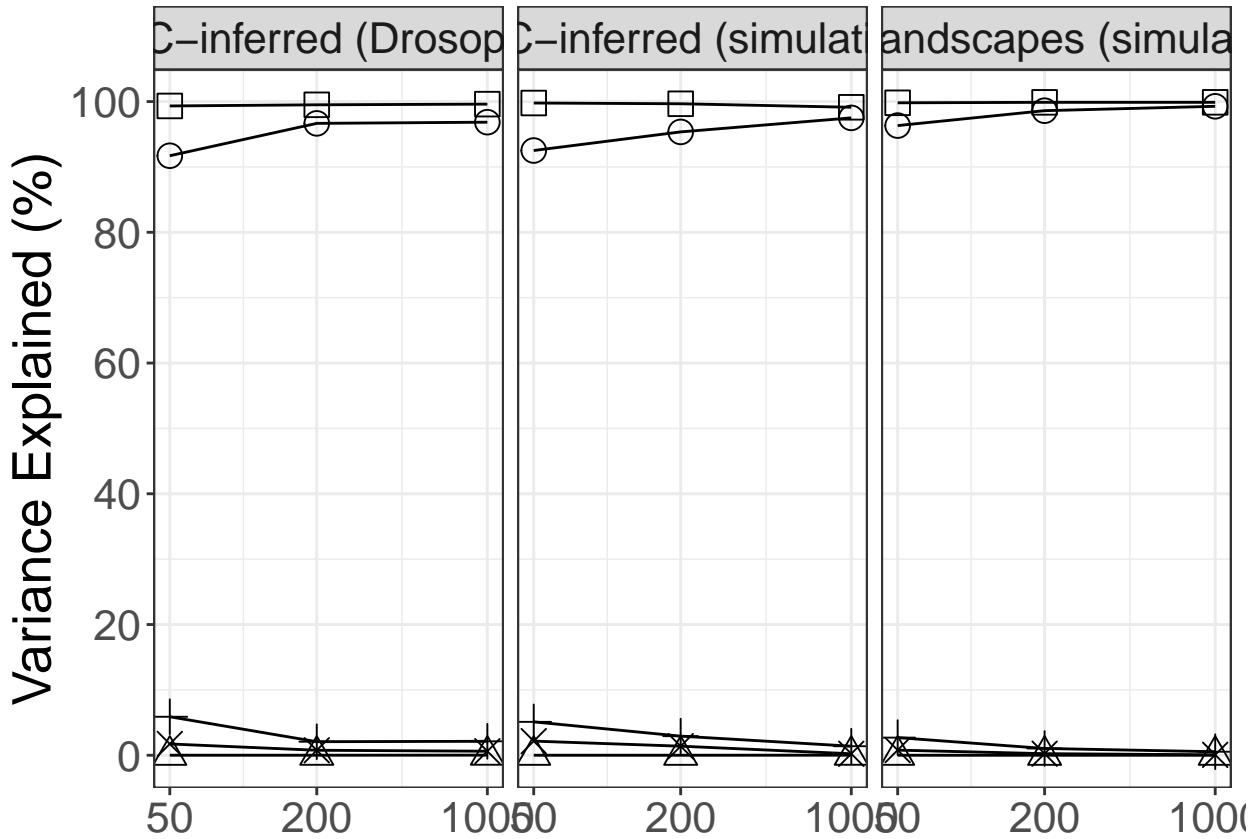
# true landscapes
r2.sim.avg <- rbind.data.frame(r2.sim.50kb$average, r2.sim.200kb$average, r2.sim.1Mb$average, make.row.names=TRUE)
colnames(r2.sim.avg) <- c("Total", "Theta", "Rho", "TMRCA", "Theta:TMRCA")
r2.sim.avg$bin.size <- c(50, 200, 1000)

# inferred landscapes
r2.inf.avg <- rbind.data.frame(r2.inf.50kb$average, r2.inf.200kb$average, r2.inf.1Mb$average, make.row.names=TRUE)
colnames(r2.inf.avg) <- c("Total", "Theta", "Rho", "TMRCA", "Theta:TMRCA")
r2.inf.avg$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("iSMC-inferred (Drosophila)", 3), rep("True landscapes (simulations)", 3), rep("True landscapes (estimates)", 3))

molten.r2 <- melt(r2.tab.comb, id.vars = c("bin.size", "type"))
r2.coal.plot <- ggplot(data = molten.r2, aes(x = bin.size, y = value, shape = variable))
r2.coal.plot <- r2.coal.plot + geom_line(data = molten.r2)
r2.coal.plot <- r2.coal.plot + geom_point(aes(shape = variable), size = 4) + facet_wrap(~type)
r2.coal.plot <- r2.coal.plot + scale_x_continuous(breaks = c(50, 200, 1000), trans="log10")
r2.coal.plot <- r2.coal.plot + scale_y_continuous(breaks = pretty_breaks())
r2.coal.plot <- r2.coal.plot + scale_shape_manual(values = c(0, 1, 2, 3, 4))
r2.coal.plot <- r2.coal.plot + labs(title = NULL, x = NULL, y = "Variance Explained (%)")
r2.coal.plot <- r2.coal.plot + theme_bw()
r2.coal.plot <- r2.coal.plot + theme(axis.title = element_text(size = 20), axis.text = element_text(size = 16),
                                         strip.text.x = element_text(size = 16), legend.position = "none")
r2.coal.plot

```



5 Simulations with BGS and Gamma Mutation Landscape in SLiM

```
cor.table.theta <- as.data.frame(matrix(ncol = 3, nrow = nreps))
row.names(cor.table.theta) <- reps
colnames(cor.table.theta) <- c("50kb", "200kb", "1Mb")

cor.table.rho <- as.data.frame(matrix(ncol = 3, nrow = nreps))
row.names(cor.table.rho) <- reps
colnames(cor.table.rho) <- c("50kb", "200kb", "1Mb")

cor.table.tmrca <- as.data.frame(matrix(ncol = 3, nrow = nreps))
row.names(cor.table.tmrca) <- reps
colnames(cor.table.tmrca) <- c("50kb", "200kb", "1Mb")
```

5.1 50 kb scale

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

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

sim.theta.50kb <- read.table("dm_bgs_sims/MutationMap_50kb.csv", sep = ",", header = T)
```

```

sim.rho.50kb <- read.table("dm_bgs_sims/RecombinationMap_50kb.csv", sep = "", header = T)

5.1.1 Replicate 1

rep1.sim.tmrca.50kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep1_w50000.csv", sep = "", header = T)

rep_1.pi.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep1/dm2L_bgs_rep1_5.dim")
rep_1.pi.50kb$avg <- apply(rep_1.pi.50kb[4:ncol(rep_1.pi.50kb)], 1, mean)
rep_1.rho.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep1/dm2L_bgs_rep1_5.dim")
rep_1.theta.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep1/dm2L_bgs_rep1_5.dim")
rep_1.tmrca.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep1/dm2L_bgs_rep1_5.dim")
rep_1.tmrca.50kb$avg <- apply(rep_1.tmrca.50kb[4:ncol(rep_1.tmrca.50kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.50kb.rep1 <- as.data.frame(cbind(rep_1.pi.50kb$avg, sim.theta.50kb$Rate, rep1.sim.tmrca.50kb))
names(true.lands.50kb.rep1) <- c("diversity", "theta", "tmrca", "rho")

true.lands.50kb.rep1$thetaC <- (true.lands.50kb.rep1$theta - mean(true.lands.50kb.rep1$theta)) / sd(true.lands.50kb.rep1$theta)
true.lands.50kb.rep1$tmrcaC <- (true.lands.50kb.rep1$tmrca - mean(true.lands.50kb.rep1$tmrca)) / sd(true.lands.50kb.rep1$tmrca)
true.lands.50kb.rep1$rhoC <- (true.lands.50kb.rep1$rho - mean(true.lands.50kb.rep1$rho, na.rm = T)) / sd(true.lands.50kb.rep1$rho)

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.50kb.rep1)
anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.50kb.true[1, 1] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]) * 100
r2.bgs.50kb.true[2, 1] <- anova.diversity$VarExp[1] * 100
r2.bgs.50kb.true[3, 1] <- anova.diversity$VarExp[2] * 100
r2.bgs.50kb.true[4, 1] <- anova.diversity$VarExp[3] * 100
r2.bgs.50kb.true[5, 1] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.50kb.rep1 <- as.data.frame(cbind(rep_1.pi.50kb$avg, rep_1.theta.50kb$sample_mean, rep_1.tmrca.50kb$sample_mean))
names(inf.lands.50kb.rep1) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[1, 1] <- cor.test(sim.theta.50kb$Rate, rep_1.theta.50kb$sample_mean, method = "spearman")
cor.table.rho[1, 1] <- cor.test(sim.rho.50kb$Rate, rep_1.rho.50kb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[1, 1] <- cor.test(rep1.sim.tmrca.50kb$AverageTmrca, rep_1.tmrca.50kb$avg, method = "spearman")$estimate

# standardizing
inf.lands.50kb.rep1$thetaC <- (inf.lands.50kb.rep1$theta - mean(inf.lands.50kb.rep1$theta)) / sd(inf.lands.50kb.rep1$theta)
inf.lands.50kb.rep1$tmrcaC <- (inf.lands.50kb.rep1$tmrca - mean(inf.lands.50kb.rep1$tmrca)) / sd(inf.lands.50kb.rep1$tmrca)
inf.lands.50kb.rep1$rhoC <- (inf.lands.50kb.rep1$rho - mean(inf.lands.50kb.rep1$rho)) / sd(inf.lands.50kb.rep1$rho)

inf.lands.50kb.rep1$bin <- 1:nrow(inf.lands.50kb.rep1)

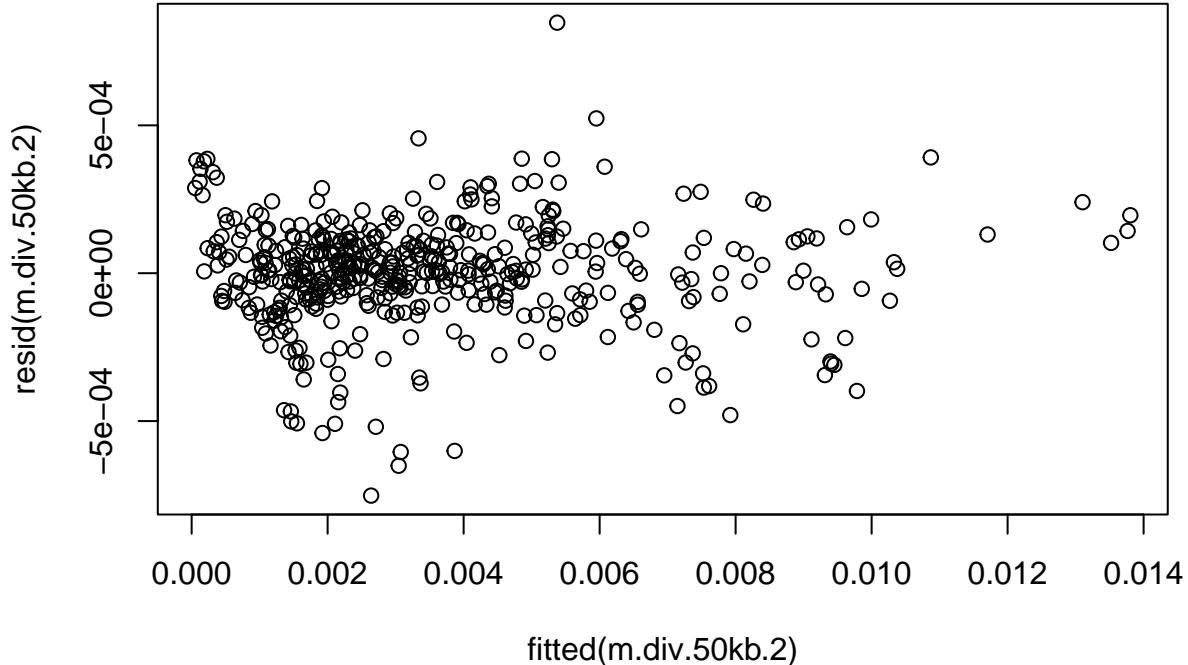
inf.lands.50kb.rep1$Replicate <- 1

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.50kb.rep1)
m.div.50kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.50kb.rep1)
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC)^2, data = inf.lands.50kb.rep1)

```

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

```
##          df      AIC
## m.div.50kb    6 -6711.225
## m.div.50kb.2  7 -6769.917
## m.div.50kb.3  8 -6773.287
plot(resid(m.div.50kb.2)~fitted(m.div.50kb.2))
```

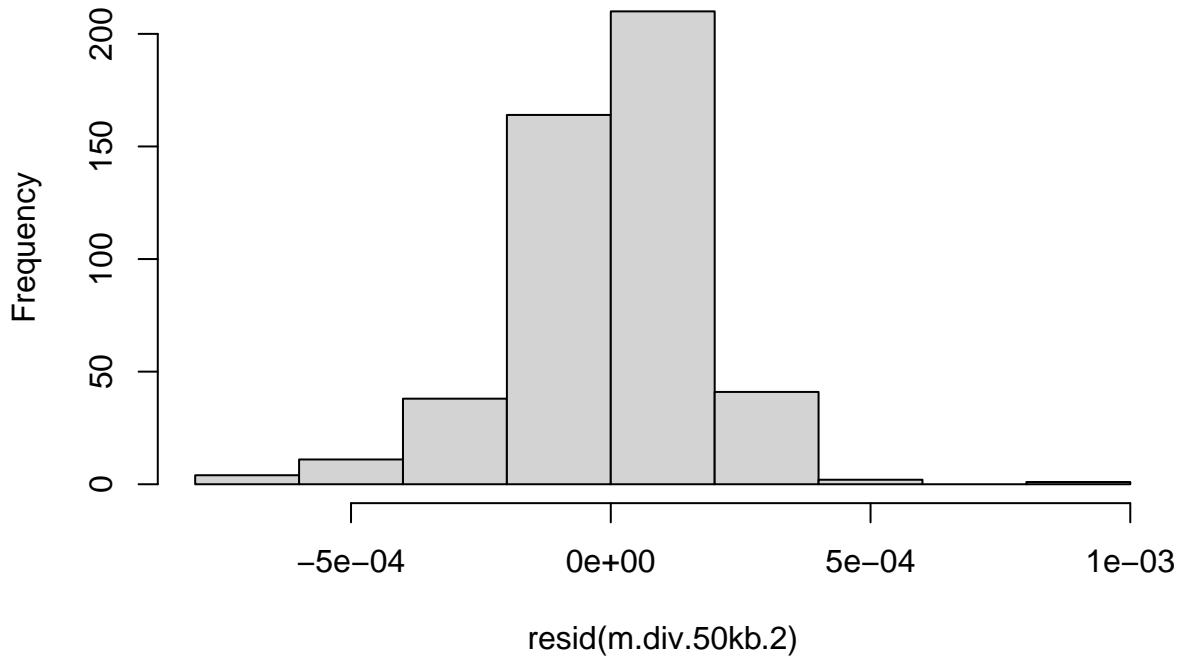


```
dwtest(m.div.50kb.2)
```

```
##
##  Durbin-Watson test
##
## data: m.div.50kb.2
## DW = 0.87081, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.50kb.2)
```

```
##
##  Harrison-McCabe test
##
## data: m.div.50kb.2
## HMC = 0.44862, p-value = 0.054
hist(resid(m.div.50kb.2))
```

Histogram of resid(m.div.50kb.2)



```
summary(m.div.50kb.2)

##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC +
##     rhoC:tmrcaC, data = inf.lands.50kb.rep1)
##
## Residuals:
##       Min         1Q      Median         3Q        Max
## -7.520e-04 -7.777e-05  9.490e-06  1.025e-04  8.472e-04
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.408e-03 9.997e-06 340.914 < 2e-16 ***
## thetaC      2.020e-03 9.343e-06 216.158 < 2e-16 ***
## rhoC        7.090e-05 9.965e-06  7.115 4.27e-12 ***
## tmrcaC      1.029e-03 1.844e-05  55.823 < 2e-16 ***
## thetaC:tmrcaC 5.722e-04 1.197e-05  47.792 < 2e-16 ***
## rhoC:tmrcaC  1.153e-04 1.442e-05   7.997 1.02e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001815 on 465 degrees of freedom
## Multiple R-squared:  0.9949, Adjusted R-squared:  0.9948
## F-statistic: 1.8e+04 on 5 and 465 DF,  p-value: < 2.2e-16

anova.diversity <- Anova(m.div.50kb.2)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
```

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

5.1.2 Replicate 2

```
rep2.sim.tmrca.50kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep2_w50000.csv", sep = ",", header = T)

rep_2.pi.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep2/dm2L_bgs_rep2_5.dim")
rep_2.pi.50kb$avg <- apply(rep_2.pi.50kb[4:ncol(rep_2.pi.50kb)], 1, mean)
rep_2.rho.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep2/dm2L_bgs_rep2_5.rho")
rep_2.theta.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep2/dm2L_bgs_rep2_5.theta")
rep_2.tmrca.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep2/dm2L_bgs_rep2_5.tmrca")
rep_2.tmrca.50kb$avg <- apply(rep_2.tmrca.50kb[4:ncol(rep_2.tmrca.50kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.50kb.rep2 <- as.data.frame(cbind(rep_1.pi.50kb$avg, sim.theta.50kb$Rate, rep2.sim.tmrca.50kb$Rate))
names(true.lands.50kb.rep2) <- c("diversity", "theta", "tmrca", "rho")

true.lands.50kb.rep2$thetaC <- (true.lands.50kb.rep2$theta - mean(true.lands.50kb.rep2$theta)) / sd(true.lands.50kb.rep2$theta)
true.lands.50kb.rep2$tmrcaC <- (true.lands.50kb.rep2$tmrca - mean(true.lands.50kb.rep2$tmrca)) / sd(true.lands.50kb.rep2$tmrca)
true.lands.50kb.rep2$rhoC <- (true.lands.50kb.rep2$rho - mean(true.lands.50kb.rep2$rho, na.rm = T)) / sd(true.lands.50kb.rep2$rho)

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.50kb.rep2)
anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.50kb.true[1, 2] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.bgs.50kb.true[2, 2] <- anova.diversity$VarExp[1] * 100
r2.bgs.50kb.true[3, 2] <- anova.diversity$VarExp[2] * 100
r2.bgs.50kb.true[4, 2] <- anova.diversity$VarExp[3] * 100
r2.bgs.50kb.true[5, 2] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.50kb.rep2 <- as.data.frame(cbind(rep_2.pi.50kb$avg, rep_2.theta.50kb$sample_mean, rep_2.tmrca.50kb$sample_mean))
names(inf.lands.50kb.rep2) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[2, 1] <- cor.test(sim.theta.50kb$Rate, rep_2.theta.50kb$sample_mean, method = "spearman")
cor.table.rho[2, 1] <- cor.test(sim.rho.50kb$Rate, rep_2.rho.50kb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[2, 1] <- cor.test(rep2.sim.tmrca.50kb$AverageTmrca, rep_2.tmrca.50kb$avg, method = "spearman")$estimate

# standardizing
inf.lands.50kb.rep2$thetaC <- (inf.lands.50kb.rep2$theta - mean(inf.lands.50kb.rep2$theta)) / sd(inf.lands.50kb.rep2$theta)
inf.lands.50kb.rep2$tmrcaC <- (inf.lands.50kb.rep2$tmrca - mean(inf.lands.50kb.rep2$tmrca)) / sd(inf.lands.50kb.rep2$tmrca)
inf.lands.50kb.rep2$rhoC <- (inf.lands.50kb.rep2$rho - mean(inf.lands.50kb.rep2$rho)) / sd(inf.lands.50kb.rep2$rho)

inf.lands.50kb.rep2$bin <- 1:nrow(inf.lands.50kb.rep2)

inf.lands.50kb.rep2$Replicate <- 2
```

```

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.50kb.rep2)
m.div.50kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.50kb.rep2)
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.50kb.rep2)

```

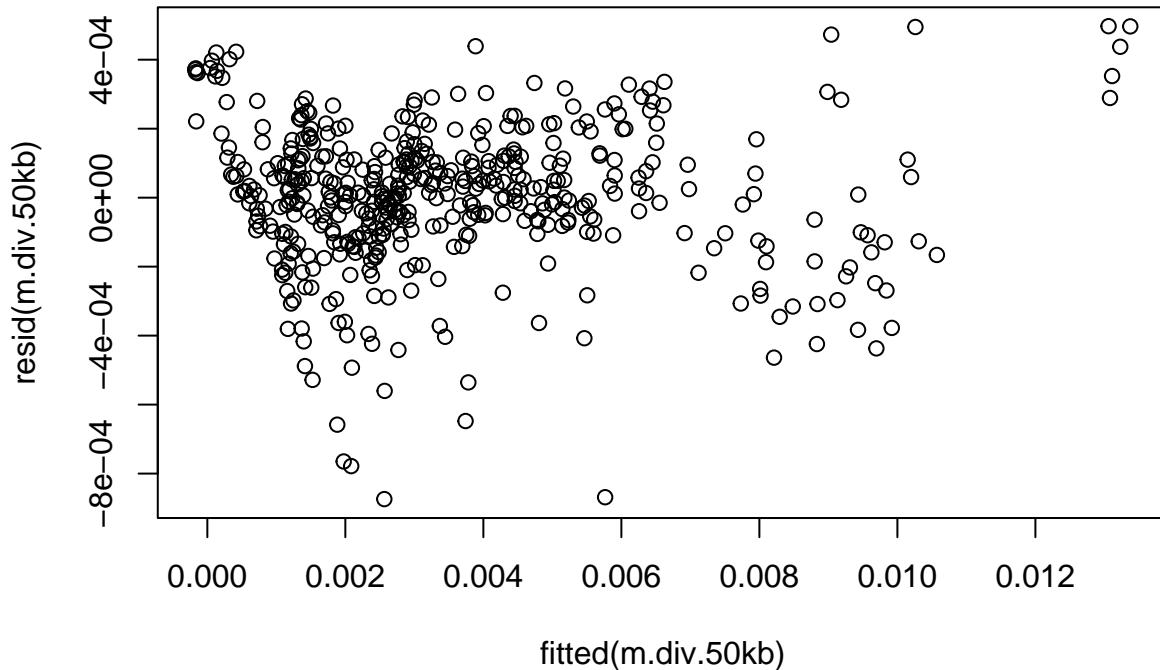
```
AIC(m.div.50kb, m.div.50kb.2, m.div.50kb.3)
```

```

##           df      AIC
## m.div.50kb    6 -6649.218
## m.div.50kb.2  7 -6767.980
## m.div.50kb.3  8 -6779.854

```

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



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

```

##
##  Durbin-Watson test
##
##  data:  m.div.50kb
##  DW = 0.92098, p-value < 2.2e-16
##  alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.50kb)

```

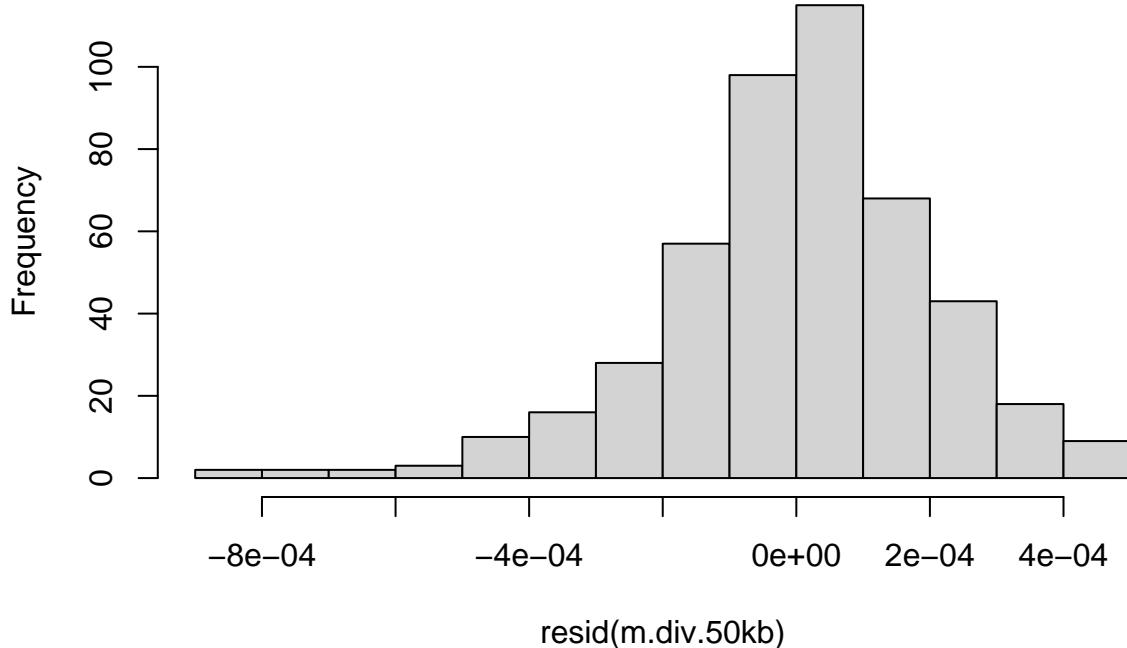
```

##
##  Harrison-McCabe test
##
##  data:  m.div.50kb
##  HMC = 0.4077, p-value = 0.001

```

```
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 = inf.lands.50kb.rep2)  
##  
## Residuals:  
##       Min        1Q     Median        3Q       Max  
## -8.740e-04 -1.030e-04  1.349e-05  1.126e-04  4.971e-04  
##  
## Coefficients:  
##             Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3.476e-03 9.569e-06 363.28 <2e-16 ***  
## thetaC      2.014e-03 1.014e-05 198.57 <2e-16 ***  
## rhoC        2.409e-05 1.052e-05   2.29  0.0225 *  
## tmrcaC      1.076e-03 1.030e-05 104.47 <2e-16 ***  
## thetaC:tmrcaC 6.254e-04 1.147e-05   54.51 <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.0002066 on 466 degrees of freedom  
## Multiple R-squared:  0.9935, Adjusted R-squared:  0.9934  
## F-statistic: 1.781e+04 on 4 and 466 DF,  p-value: < 2.2e-16  
anova.diversity <- Anova(m.div.50kb)  
apiss <- anova.diversity$"Sum Sq"  
anova.diversity$VarExp <- apiss / sum(apiss)
```

```

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

```

5.1.3 Replicate 3

```

rep3.sim.tmrca.50kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep3_w50000.csv", sep = ",", header = T)

rep_3.pi.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep3/dm2L_bgs_rep3_5.di")
rep_3.pi.50kb$avg <- apply(rep_3.pi.50kb[4:ncol(rep_3.pi.50kb)], 1, mean)
rep_3.rho.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep3/dm2L_bgs_rep3_5.r")
rep_3.theta.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep3/dm2L_bgs_rep3_5.t")
rep_3.tmrca.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep3/dm2L_bgs_rep3_5.tmrca")
rep_3.tmrca.50kb$avg <- apply(rep_3.tmrca.50kb[4:ncol(rep_3.tmrca.50kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.50kb.rep3 <- as.data.frame(cbind(rep_1.pi.50kb$avg, sim.theta.50kb$Rate, rep3.sim.tmrca.50kb))
names(true.lands.50kb.rep3) <- c("diversity", "theta", "tmrca", "rho")

true.lands.50kb.rep3$thetaC <- (true.lands.50kb.rep3$theta - mean(true.lands.50kb.rep3$theta)) / sd(true.lands.50kb.rep3$theta)
true.lands.50kb.rep3$tmrcaC <- (true.lands.50kb.rep3$tmrca - mean(true.lands.50kb.rep3$tmrca)) / sd(true.lands.50kb.rep3$tmrca)
true.lands.50kb.rep3$rhoC <- (true.lands.50kb.rep3$rho - mean(true.lands.50kb.rep3$rho, na.rm = T)) / sd(true.lands.50kb.rep3$rho)

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.50kb.rep3)
anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.50kb.true[1, 3] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.bgs.50kb.true[2, 3] <- anova.diversity$VarExp[1] * 100
r2.bgs.50kb.true[3, 3] <- anova.diversity$VarExp[2] * 100
r2.bgs.50kb.true[4, 3] <- anova.diversity$VarExp[3] * 100
r2.bgs.50kb.true[5, 3] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.50kb.rep3 <- as.data.frame(cbind(rep_3.pi.50kb$avg, rep_3.theta.50kb$sample_mean, rep_3.tmrca.50kb))
names(inf.lands.50kb.rep3) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[3, 1] <- cor.test(sim.theta.50kb$Rate, rep_3.theta.50kb$sample_mean, method = "spearman")
cor.table.rho[3, 1] <- cor.test(sim.rho.50kb$Rate, rep_3.rho.50kb$sample_mean, method = "spearman")$est
cor.table.tmrca[3, 1] <- cor.test(rep3.sim.tmrca.50kb$AverageTmrca, rep_3.tmrca.50kb$avg, method = "spearman")

# standardizing
inf.lands.50kb.rep3$thetaC <- (inf.lands.50kb.rep3$theta - mean(inf.lands.50kb.rep3$theta)) / sd(inf.lands.50kb.rep3$theta)
inf.lands.50kb.rep3$tmrcaC <- (inf.lands.50kb.rep3$tmrca - mean(inf.lands.50kb.rep3$tmrca)) / sd(inf.lands.50kb.rep3$tmrca)
inf.lands.50kb.rep3$rhoC <- (inf.lands.50kb.rep3$rho - mean(inf.lands.50kb.rep3$rho)) / sd(inf.lands.50kb.rep3$rho)

inf.lands.50kb.rep3$bin <- 1:nrow(inf.lands.50kb.rep3)

inf.lands.50kb.rep3$Replicate <- 3

```

```

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.50kb.rep3)
m.div.50kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.50kb.rep3)
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.50kb.rep3)

```

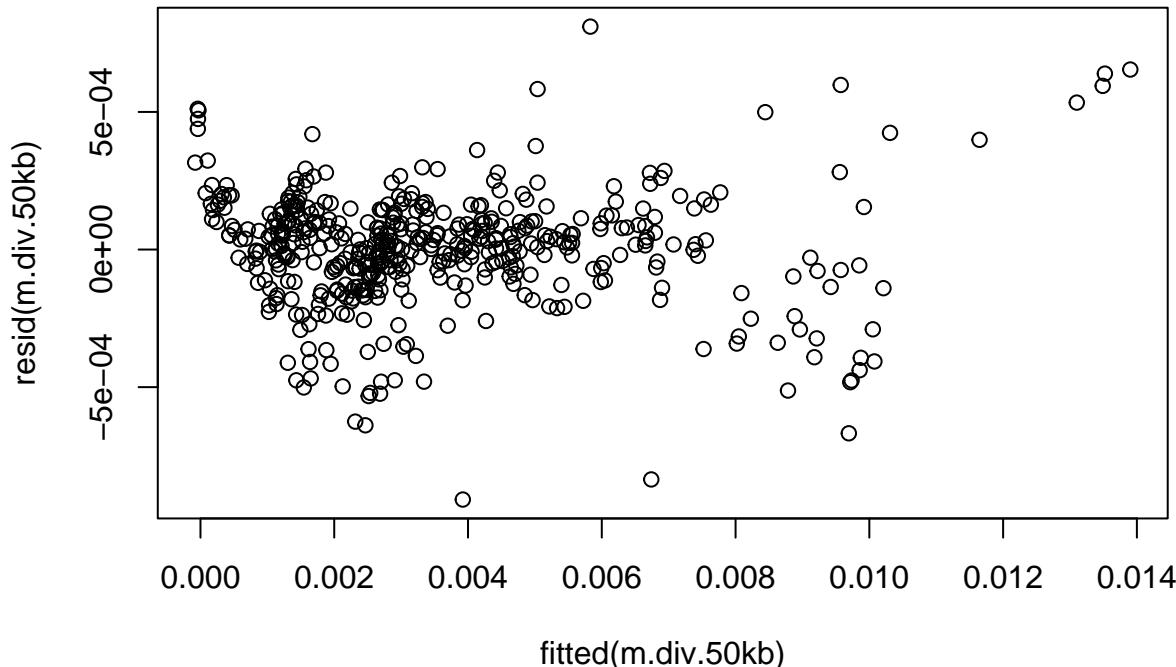
```
AIC(m.div.50kb, m.div.50kb.2, m.div.50kb.3)
```

```

##           df      AIC
## m.div.50kb    6 -6637.516
## m.div.50kb.2  7 -6780.842
## m.div.50kb.3  8 -6801.784

```

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



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

```

##
##  Durbin-Watson test
##
## data: m.div.50kb
## DW = 0.89865, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.50kb)

```

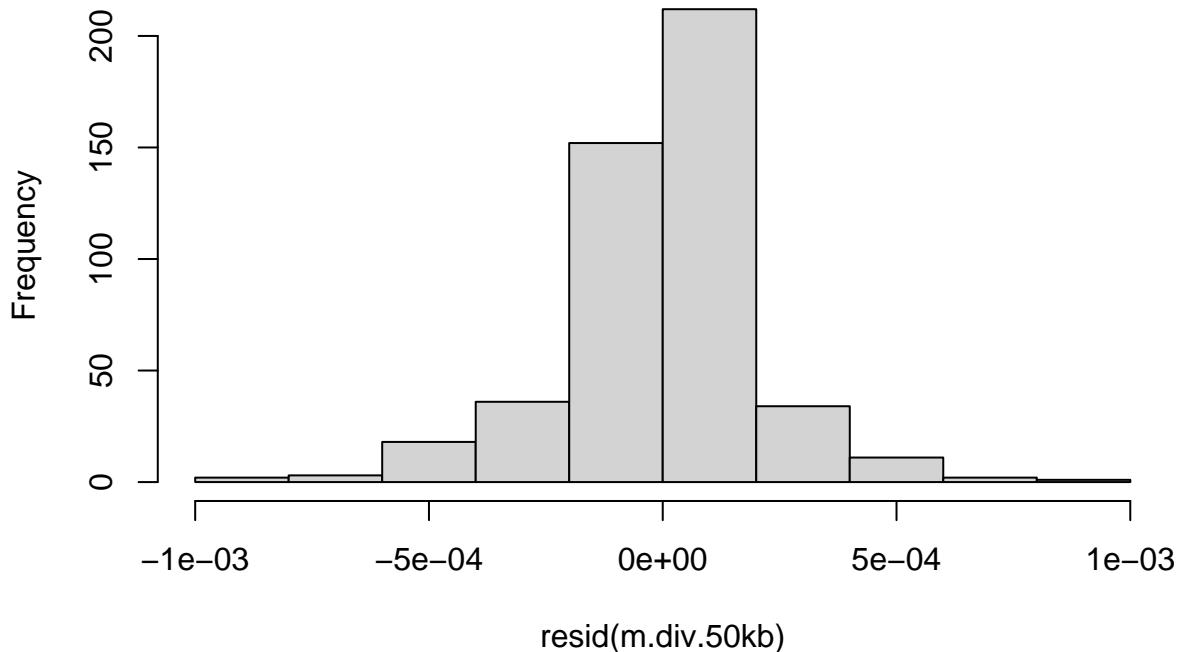
```

##
##  Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.50047, p-value = 0.538

```

```
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 = inf.lands.50kb.rep3)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max  
## -9.084e-04 -8.949e-05  1.918e-05  1.054e-04  8.099e-04  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3.505e-03 9.676e-06 362.237 <2e-16 ***  
## thetaC       2.048e-03 1.007e-05 203.462 <2e-16 ***  
## rhoC        2.280e-05 1.071e-05   2.129  0.0338 *  
## tmrcaC      9.816e-04 1.057e-05  92.838 <2e-16 ***  
## thetaC:tmrcaC 5.611e-04 1.131e-05  49.600 <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.0002091 on 466 degrees of freedom  
## Multiple R-squared:  0.9932, Adjusted R-squared:  0.9931  
## F-statistic: 1.694e+04 on 4 and 466 DF,  p-value: < 2.2e-16  
anova.diversity <- Anova(m.div.50kb)  
apiss <- anova.diversity$"Sum Sq"  
anova.diversity$VarExp <- apiss / sum(apiss)
```

```

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

```

5.1.4 Replicate 4

```

rep4.sim.tmrca.50kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep4_w50000.csv", sep = ",", header = T)

rep_4.pi.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep4/dm2L_bgs_rep4_5.di")
rep_4.pi.50kb$avg <- apply(rep_4.pi.50kb[4:ncol(rep_4.pi.50kb)], 1, mean)
rep_4.rho.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep4/dm2L_bgs_rep4_5.r")
rep_4.theta.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep4/dm2L_bgs_rep4_5.t")
rep_4.tmrca.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep4/dm2L_bgs_rep4_5.tmrca")
rep_4.tmrca.50kb$avg <- apply(rep_4.tmrca.50kb[4:ncol(rep_4.tmrca.50kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.50kb.rep4 <- as.data.frame(cbind(rep_1.pi.50kb$avg, sim.theta.50kb$Rate, rep4.sim.tmrca.50kb))
names(true.lands.50kb.rep4) <- c("diversity", "theta", "tmrca", "rho")

true.lands.50kb.rep4$thetaC <- (true.lands.50kb.rep4$theta - mean(true.lands.50kb.rep4$theta)) / sd(true.lands.50kb.rep4$theta)
true.lands.50kb.rep4$tmrcaC <- (true.lands.50kb.rep4$tmrca - mean(true.lands.50kb.rep4$tmrca)) / sd(true.lands.50kb.rep4$tmrca)
true.lands.50kb.rep4$rhoC <- (true.lands.50kb.rep4$rho - mean(true.lands.50kb.rep4$rho, na.rm = T)) / sd(true.lands.50kb.rep4$rho)

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.50kb.rep4)
anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.50kb.true[1, 4] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.bgs.50kb.true[2, 4] <- anova.diversity$VarExp[1] * 100
r2.bgs.50kb.true[3, 4] <- anova.diversity$VarExp[2] * 100
r2.bgs.50kb.true[4, 4] <- anova.diversity$VarExp[3] * 100
r2.bgs.50kb.true[5, 4] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.50kb.rep4 <- as.data.frame(cbind(rep_4.pi.50kb$avg, rep_4.theta.50kb$sample_mean, rep_4.tmrca.50kb))
names(inf.lands.50kb.rep4) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[4, 1] <- cor.test(sim.theta.50kb$Rate, rep_4.theta.50kb$sample_mean, method = "spearman")
cor.table.rho[4, 1] <- cor.test(sim.rho.50kb$Rate, rep_4.rho.50kb$sample_mean, method = "spearman")$est
cor.table.tmrca[4, 1] <- cor.test(rep4.sim.tmrca.50kb$AverageTmrca, rep_4.tmrca.50kb$avg, method = "spearman")$est

# standardizing
inf.lands.50kb.rep4$thetaC <- (inf.lands.50kb.rep4$theta - mean(inf.lands.50kb.rep4$theta)) / sd(inf.lands.50kb.rep4$theta)
inf.lands.50kb.rep4$tmrcaC <- (inf.lands.50kb.rep4$tmrca - mean(inf.lands.50kb.rep4$tmrca)) / sd(inf.lands.50kb.rep4$tmrca)
inf.lands.50kb.rep4$rhoC <- (inf.lands.50kb.rep4$rho - mean(inf.lands.50kb.rep4$rho)) / sd(inf.lands.50kb.rep4$rho)

inf.lands.50kb.rep4$bin <- 1:nrow(inf.lands.50kb.rep4)

inf.lands.50kb.rep4$Replicate <- 4

```

```

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.50kb.rep4)
m.div.50kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.50kb.rep4)
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.50kb.rep4)

```

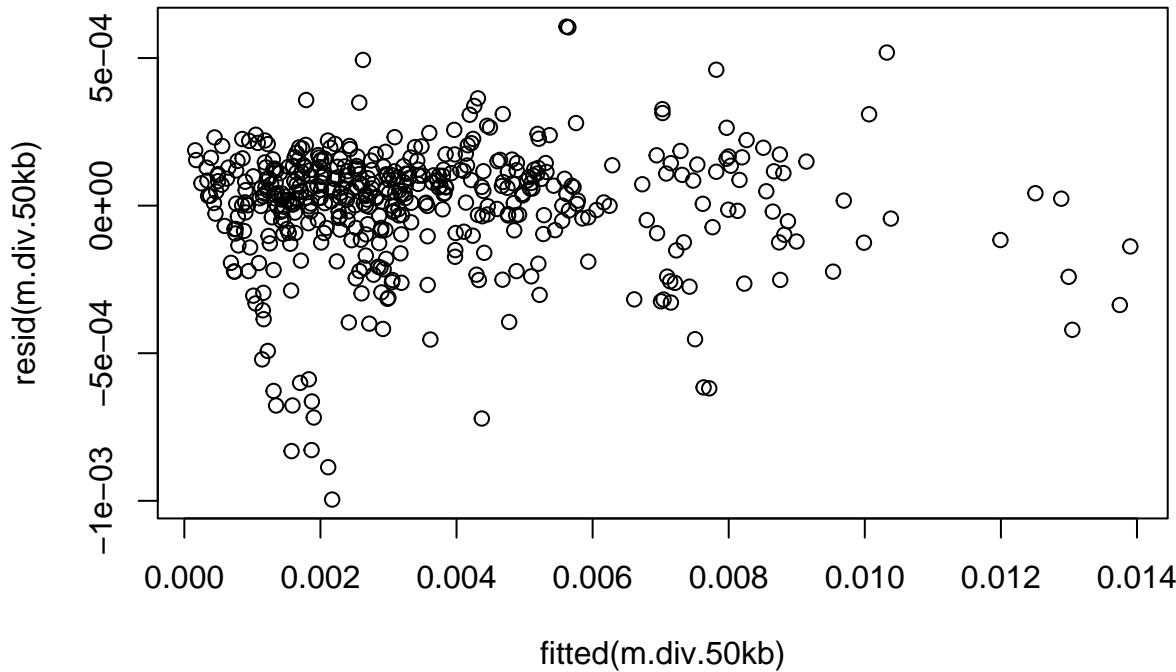
```
AIC(m.div.50kb, m.div.50kb.2, m.div.50kb.3)
```

```

##           df      AIC
## m.div.50kb    6 -6651.111
## m.div.50kb.2  7 -6649.680
## m.div.50kb.3  8 -6651.347

```

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



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

```

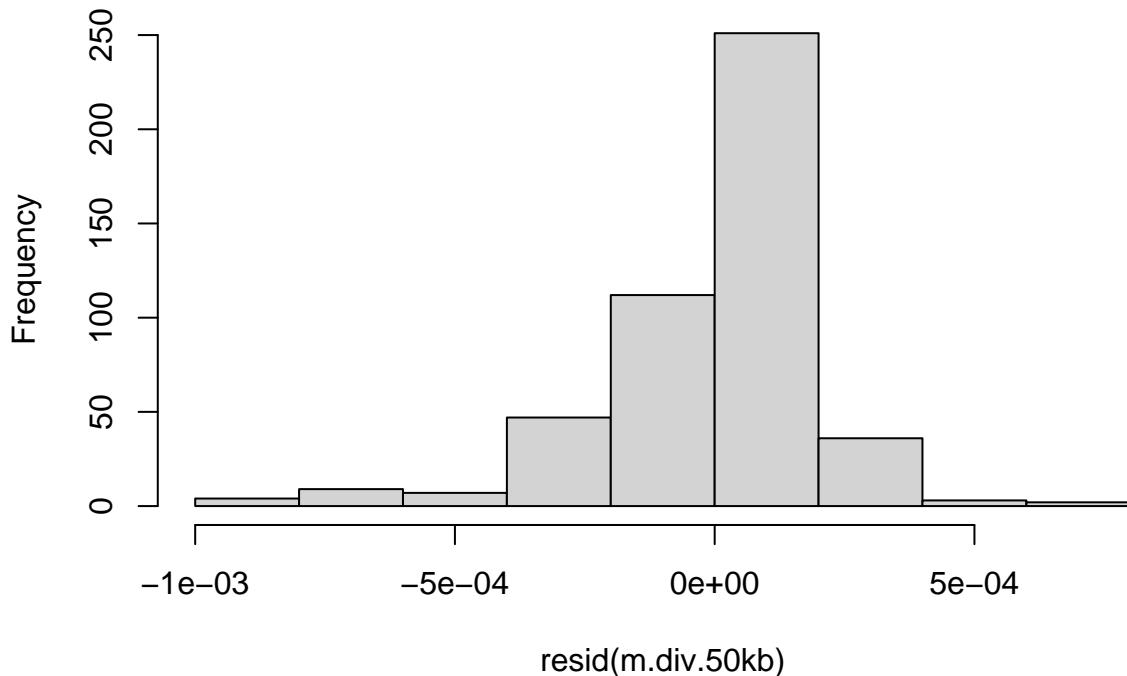
##
##  Durbin-Watson test
##
##  data:  m.div.50kb
##  DW = 0.94293, p-value < 2.2e-16
##  alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.50kb)
```

```

##
##  Harrison-McCabe test
##
##  data:  m.div.50kb
##  HMC = 0.43008, p-value = 0.017
```

```
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 = inf.lands.50kb.rep4)  
##  
## Residuals:  
##       Min         1Q     Median        3Q       Max  
## -9.957e-04 -7.736e-05  4.142e-05  1.167e-04  6.064e-04  
##  
## Coefficients:  
##             Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3.439e-03 9.716e-06 353.943 <2e-16 ***  
## thetaC      2.124e-03 1.023e-05 207.727 <2e-16 ***  
## rhoC        8.646e-05 1.011e-05   8.552 <2e-16 ***  
## tmrcaC      7.698e-04 1.005e-05  76.620 <2e-16 ***  
## thetaC:tmrcaC 4.884e-04 9.649e-06  50.617 <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.0002061 on 466 degrees of freedom  
## Multiple R-squared:  0.9934, Adjusted R-squared:  0.9933  
## F-statistic: 1.742e+04 on 4 and 466 DF,  p-value: < 2.2e-16  
anova.diversity <- Anova(m.div.50kb)  
apiss <- anova.diversity$"Sum Sq"  
anova.diversity$VarExp <- apiss / sum(apiss)
```

```

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

```

5.1.5 Replicate 5

```

rep5.sim.tmrca.50kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep5_w50000.csv", sep = ",", header = T)

rep_5.pi.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep5/dm2L_bgs_rep5_5.di")
rep_5.pi.50kb$avg <- apply(rep_5.pi.50kb[4:ncol(rep_5.pi.50kb)], 1, mean)
rep_5.rho.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep5/dm2L_bgs_rep5_5.r")
rep_5.theta.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep5/dm2L_bgs_rep5_5.t")
rep_5.tmrca.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep5/dm2L_bgs_rep5_5.tmrca")
rep_5.tmrca.50kb$avg <- apply(rep_5.tmrca.50kb[4:ncol(rep_5.tmrca.50kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.50kb.rep5 <- as.data.frame(cbind(rep_1.pi.50kb$avg, sim.theta.50kb$Rate, rep5.sim.tmrca.50kb))
names(true.lands.50kb.rep5) <- c("diversity", "theta", "tmrca", "rho")

true.lands.50kb.rep5$thetaC <- (true.lands.50kb.rep5$theta - mean(true.lands.50kb.rep5$theta)) / sd(true.lands.50kb.rep5$theta)
true.lands.50kb.rep5$tmrcaC <- (true.lands.50kb.rep5$tmrca - mean(true.lands.50kb.rep5$tmrca)) / sd(true.lands.50kb.rep5$tmrca)
true.lands.50kb.rep5$rhoC <- (true.lands.50kb.rep5$rho - mean(true.lands.50kb.rep5$rho, na.rm = T)) / sd(true.lands.50kb.rep5$rho)

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.50kb.rep5)
anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.50kb.true[1, 5] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.bgs.50kb.true[2, 5] <- anova.diversity$VarExp[1] * 100
r2.bgs.50kb.true[3, 5] <- anova.diversity$VarExp[2] * 100
r2.bgs.50kb.true[4, 5] <- anova.diversity$VarExp[3] * 100
r2.bgs.50kb.true[5, 5] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.50kb.rep5 <- as.data.frame(cbind(rep_5.pi.50kb$avg, rep_5.theta.50kb$sample_mean, rep_5.tmrca.50kb$sample_mean))
names(inf.lands.50kb.rep5) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[5, 1] <- cor.test(sim.theta.50kb$Rate, rep_5.theta.50kb$sample_mean, method = "spearman")
cor.table.rho[5, 1] <- cor.test(sim.rho.50kb$Rate, rep_5.rho.50kb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[5, 1] <- cor.test(rep5.sim.tmrca.50kb$AverageTmrca, rep_5.tmrca.50kb$avg, method = "spearman")$estimate

# standardizing
inf.lands.50kb.rep5$thetaC <- (inf.lands.50kb.rep5$theta - mean(inf.lands.50kb.rep5$theta)) / sd(inf.lands.50kb.rep5$theta)
inf.lands.50kb.rep5$tmrcaC <- (inf.lands.50kb.rep5$tmrca - mean(inf.lands.50kb.rep5$tmrca)) / sd(inf.lands.50kb.rep5$tmrca)
inf.lands.50kb.rep5$rhoC <- (inf.lands.50kb.rep5$rho - mean(inf.lands.50kb.rep5$rho)) / sd(inf.lands.50kb.rep5$rho)

inf.lands.50kb.rep5$bin <- 1:nrow(inf.lands.50kb.rep5)

inf.lands.50kb.rep5$Replicate <- 5

```

```

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.50kb.rep5)
m.div.50kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.50kb.rep5)
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.50kb.rep5)

```

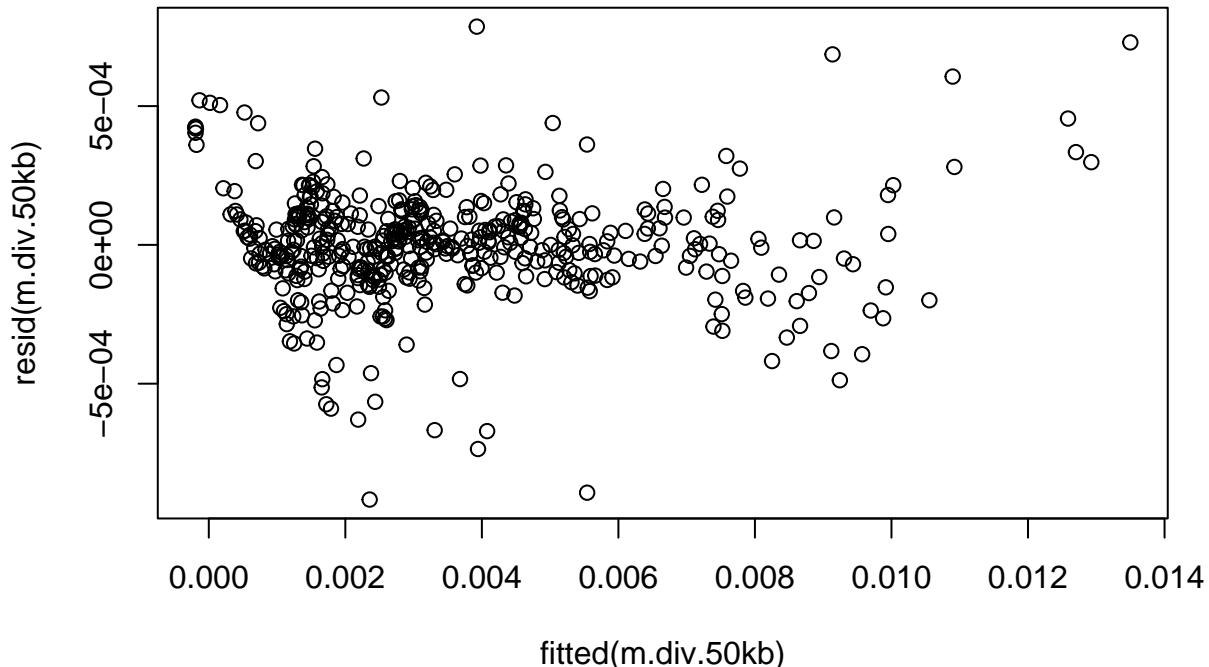
```
AIC(m.div.50kb, m.div.50kb.2, m.div.50kb.3)
```

```

##           df      AIC
## m.div.50kb    6 -6679.244
## m.div.50kb.2  7 -6802.716
## m.div.50kb.3  8 -6811.009

```

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



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

```

##
##  Durbin-Watson test
##
##  data:  m.div.50kb
##  DW = 0.98276, p-value < 2.2e-16
##  alternative hypothesis: true autocorrelation is greater than 0

```

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

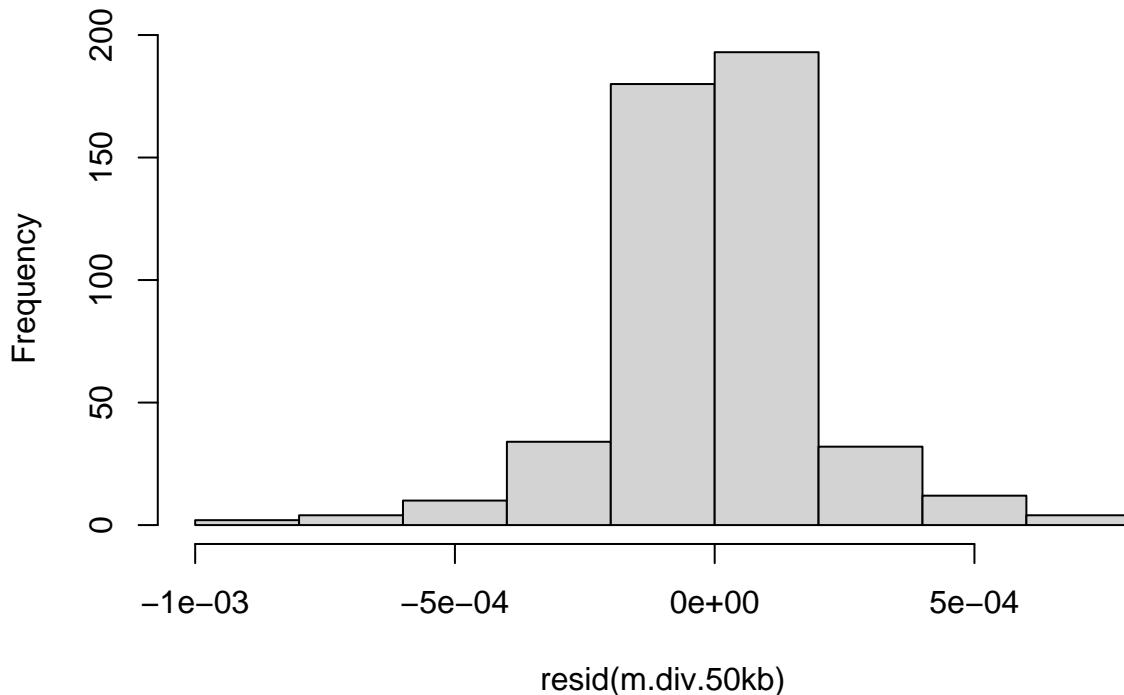
```

##
##  Harrison-McCabe test
##
##  data:  m.div.50kb
##  HMC = 0.39935, p-value = 0.004

```

```
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 = inf.lands.50kb.rep5)  
##  
## Residuals:  
##       Min        1Q     Median        3Q       Max  
## -9.174e-04 -8.788e-05  7.320e-06  9.869e-05  7.862e-04  
##  
## Coefficients:  
##             Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3.452e-03 9.312e-06 370.741 <2e-16 ***  
## thetaC      2.040e-03 9.801e-06 208.139 <2e-16 ***  
## rhoC        1.816e-05 1.023e-05   1.776  0.0764 .  
## tmrcaC      9.692e-04 1.002e-05  96.761 <2e-16 ***  
## thetaC:tmrcaC 5.711e-04 1.104e-05  51.749 <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.0002001 on 466 degrees of freedom  
## Multiple R-squared:  0.9936, Adjusted R-squared:  0.9936  
## F-statistic: 1.81e+04 on 4 and 466 DF,  p-value: < 2.2e-16  
anova.diversity <- Anova(m.div.50kb)  
apiss <- anova.diversity$"Sum Sq"  
anova.diversity$VarExp <- apiss / sum(apiss)
```

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

5.1.6 Replicate 6

```
rep6.sim.tmrca.50kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep6_w50000.csv", sep = ",", header = T)

rep_6.pi.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep6/dm2L_bgs_rep6_5.di")
rep_6.pi.50kb$avg <- apply(rep_6.pi.50kb[4:ncol(rep_6.pi.50kb)], 1, mean)
rep_6.rho.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep6/dm2L_bgs_rep6_5.r")
rep_6.theta.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep6/dm2L_bgs_rep6_5.t")
rep_6.tmrca.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep6/dm2L_bgs_rep6_5.tmrca")
rep_6.tmrca.50kb$avg <- apply(rep_6.tmrca.50kb[4:ncol(rep_6.tmrca.50kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.50kb.rep6 <- as.data.frame(cbind(rep_1.pi.50kb$avg, sim.theta.50kb$Rate, rep6.sim.tmrca.50kb))
names(true.lands.50kb.rep6) <- c("diversity", "theta", "tmrca", "rho")

true.lands.50kb.rep6$thetaC <- (true.lands.50kb.rep6$theta - mean(true.lands.50kb.rep6$theta)) / sd(true.lands.50kb.rep6$theta)
true.lands.50kb.rep6$tmrcaC <- (true.lands.50kb.rep6$tmrca - mean(true.lands.50kb.rep6$tmrca)) / sd(true.lands.50kb.rep6$tmrca)
true.lands.50kb.rep6$rhoC <- (true.lands.50kb.rep6$rho - mean(true.lands.50kb.rep6$rho, na.rm = T)) / sd(true.lands.50kb.rep6$rho)

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.50kb.rep6)
anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.50kb.true[1, 6] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.bgs.50kb.true[2, 6] <- anova.diversity$VarExp[1] * 100
r2.bgs.50kb.true[3, 6] <- anova.diversity$VarExp[2] * 100
r2.bgs.50kb.true[4, 6] <- anova.diversity$VarExp[3] * 100
r2.bgs.50kb.true[5, 6] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.50kb.rep6 <- as.data.frame(cbind(rep_6.pi.50kb$avg, rep_6.theta.50kb$sample_mean, rep_6.tmrca.50kb))
names(inf.lands.50kb.rep6) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[6, 1] <- cor.test(sim.theta.50kb$Rate, rep_6.theta.50kb$sample_mean, method = "spearman")
cor.table.rho[6, 1] <- cor.test(sim.rho.50kb$Rate, rep_6.rho.50kb$sample_mean, method = "spearman")$est
cor.table.tmrca[6, 1] <- cor.test(rep6.sim.tmrca.50kb$AverageTmrca, rep_6.tmrca.50kb$avg, method = "spearman")

# standardizing
inf.lands.50kb.rep6$thetaC <- (inf.lands.50kb.rep6$theta - mean(inf.lands.50kb.rep6$theta)) / sd(inf.lands.50kb.rep6$theta)
inf.lands.50kb.rep6$tmrcaC <- (inf.lands.50kb.rep6$tmrca - mean(inf.lands.50kb.rep6$tmrca)) / sd(inf.lands.50kb.rep6$tmrca)
inf.lands.50kb.rep6$rhoC <- (inf.lands.50kb.rep6$rho - mean(inf.lands.50kb.rep6$rho)) / sd(inf.lands.50kb.rep6$rho)

inf.lands.50kb.rep6$bin <- 1:nrow(inf.lands.50kb.rep6)

inf.lands.50kb.rep6$Replicate <- 6
```

```

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.50kb.rep6)
m.div.50kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.50kb.rep6)
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.50kb.rep6)

```

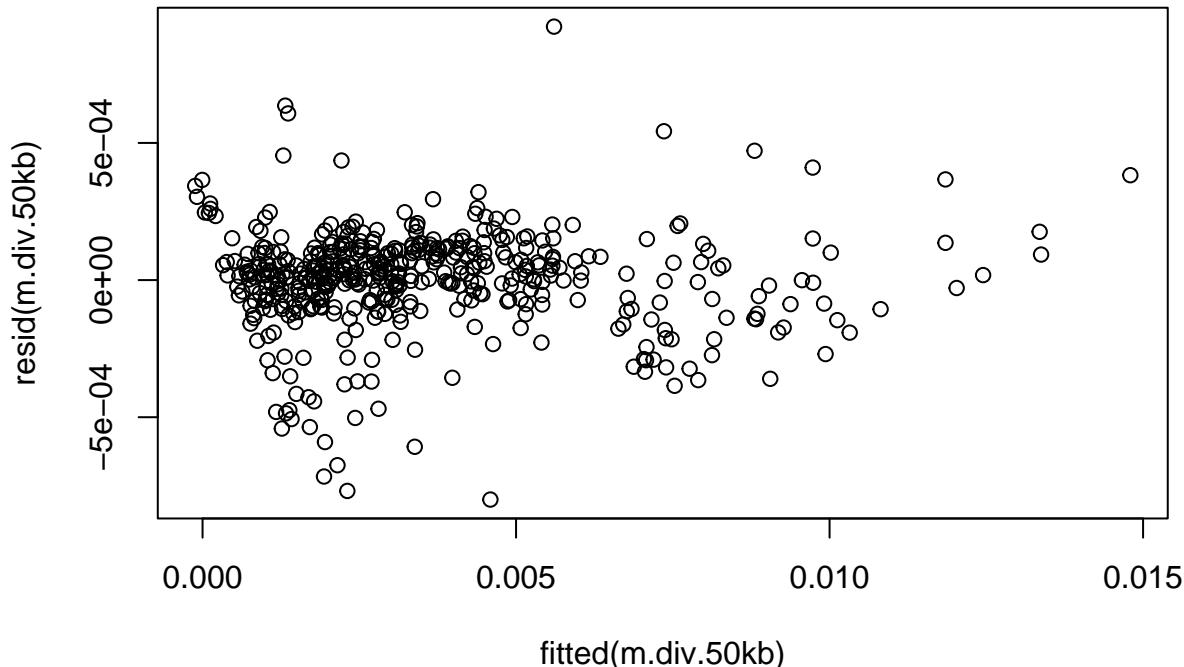
```
AIC(m.div.50kb, m.div.50kb.2, m.div.50kb.3)
```

```

##           df      AIC
## m.div.50kb    6 -6743.728
## m.div.50kb.2  7 -6802.823
## m.div.50kb.3  8 -6805.058

```

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



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

```

##
##  Durbin-Watson test
##
## data: m.div.50kb
## DW = 0.93612, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.50kb)

```

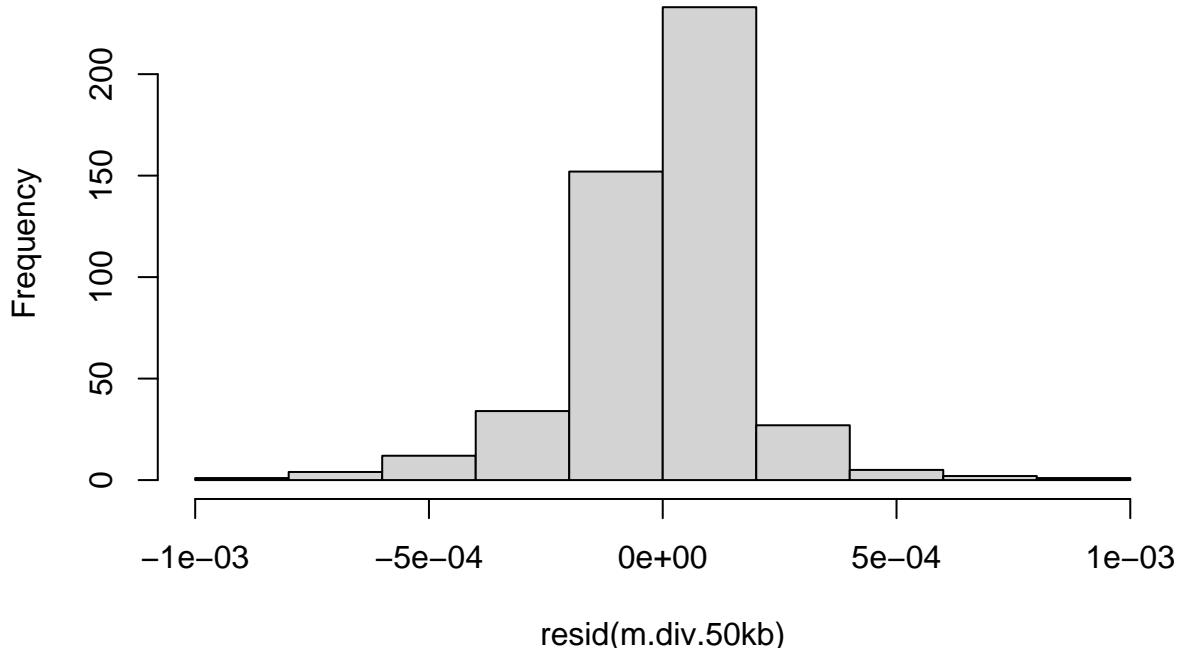
```

##
##  Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.4347, p-value = 0.024

```

```
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 = inf.lands.50kb.rep6)  
##  
## Residuals:  
##      Min        1Q     Median        3Q       Max  
## -8.002e-04 -7.326e-05  1.942e-05  9.821e-05  9.241e-04  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3.404e-03 8.832e-06 385.358 < 2e-16 ***  
## thetaC       2.044e-03 9.356e-06 218.484 < 2e-16 ***  
## rhoC         3.985e-05 9.270e-06   4.299 2.09e-05 ***  
## tmrcaC       8.965e-04 9.517e-06   94.200 < 2e-16 ***  
## thetaC:tmrcaC 5.574e-04 9.219e-06  60.461 < 2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.0001868 on 466 degrees of freedom  
## Multiple R-squared:  0.9947, Adjusted R-squared:  0.9947  
## F-statistic: 2.194e+04 on 4 and 466 DF,  p-value: < 2.2e-16  
anova.diversity <- Anova(m.div.50kb)  
apiss <- anova.diversity$"Sum Sq"  
anova.diversity$VarExp <- apiss / sum(apiss)
```

```

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

```

5.1.7 Replicate 7

```

rep7.sim.tmrca.50kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep7_w50000.csv", sep = ",", header = T)

rep_7.pi.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep7/dm2L_bgs_rep7_5.di")
rep_7.pi.50kb$avg <- apply(rep_7.pi.50kb[4:ncol(rep_7.pi.50kb)], 1, mean)
rep_7.rho.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep7/dm2L_bgs_rep7_5.r")
rep_7.theta.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep7/dm2L_bgs_rep7_5.t")
rep_7.tmrca.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep7/dm2L_bgs_rep7_5.tmrca")
rep_7.tmrca.50kb$avg <- apply(rep_7.tmrca.50kb[4:ncol(rep_7.tmrca.50kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.50kb.rep7 <- as.data.frame(cbind(rep_1.pi.50kb$avg, sim.theta.50kb$Rate, rep7.sim.tmrca.50kb))
names(true.lands.50kb.rep7) <- c("diversity", "theta", "tmrca", "rho")

true.lands.50kb.rep7$thetaC <- (true.lands.50kb.rep7$theta - mean(true.lands.50kb.rep7$theta)) / sd(true.lands.50kb.rep7$theta)
true.lands.50kb.rep7$tmrcaC <- (true.lands.50kb.rep7$tmrca - mean(true.lands.50kb.rep7$tmrca)) / sd(true.lands.50kb.rep7$tmrca)
true.lands.50kb.rep7$rhoC <- (true.lands.50kb.rep7$rho - mean(true.lands.50kb.rep7$rho, na.rm = T)) / sd(true.lands.50kb.rep7$rho)

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.50kb.rep7)
anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.50kb.true[1, 7] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.bgs.50kb.true[2, 7] <- anova.diversity$VarExp[1] * 100
r2.bgs.50kb.true[3, 7] <- anova.diversity$VarExp[2] * 100
r2.bgs.50kb.true[4, 7] <- anova.diversity$VarExp[3] * 100
r2.bgs.50kb.true[5, 7] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.50kb.rep7 <- as.data.frame(cbind(rep_7.pi.50kb$avg, rep_7.theta.50kb$sample_mean, rep_7.tmrca.50kb$AverageTmrca))
names(inf.lands.50kb.rep7) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[7, 1] <- cor.test(sim.theta.50kb$Rate, rep_7.theta.50kb$sample_mean, method = "spearman")
cor.table.rho[7, 1] <- cor.test(sim.rho.50kb$Rate, rep_7.rho.50kb$sample_mean, method = "spearman")$est
cor.table.tmrca[7, 1] <- cor.test(rep7.sim.tmrca.50kb$AverageTmrca, rep_7.tmrca.50kb$avg, method = "spearman")$est

# standardizing
inf.lands.50kb.rep7$thetaC <- (inf.lands.50kb.rep7$theta - mean(inf.lands.50kb.rep7$theta)) / sd(inf.lands.50kb.rep7$theta)
inf.lands.50kb.rep7$tmrcaC <- (inf.lands.50kb.rep7$tmrca - mean(inf.lands.50kb.rep7$tmrca)) / sd(inf.lands.50kb.rep7$tmrca)
inf.lands.50kb.rep7$rhoC <- (inf.lands.50kb.rep7$rho - mean(inf.lands.50kb.rep7$rho)) / sd(inf.lands.50kb.rep7$rho)

inf.lands.50kb.rep7$bin <- 1:nrow(inf.lands.50kb.rep7)

inf.lands.50kb.rep7$Replicate <- 7

```

```

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.50kb.rep7)
m.div.50kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.50kb.rep7)
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.50kb.rep7)

```

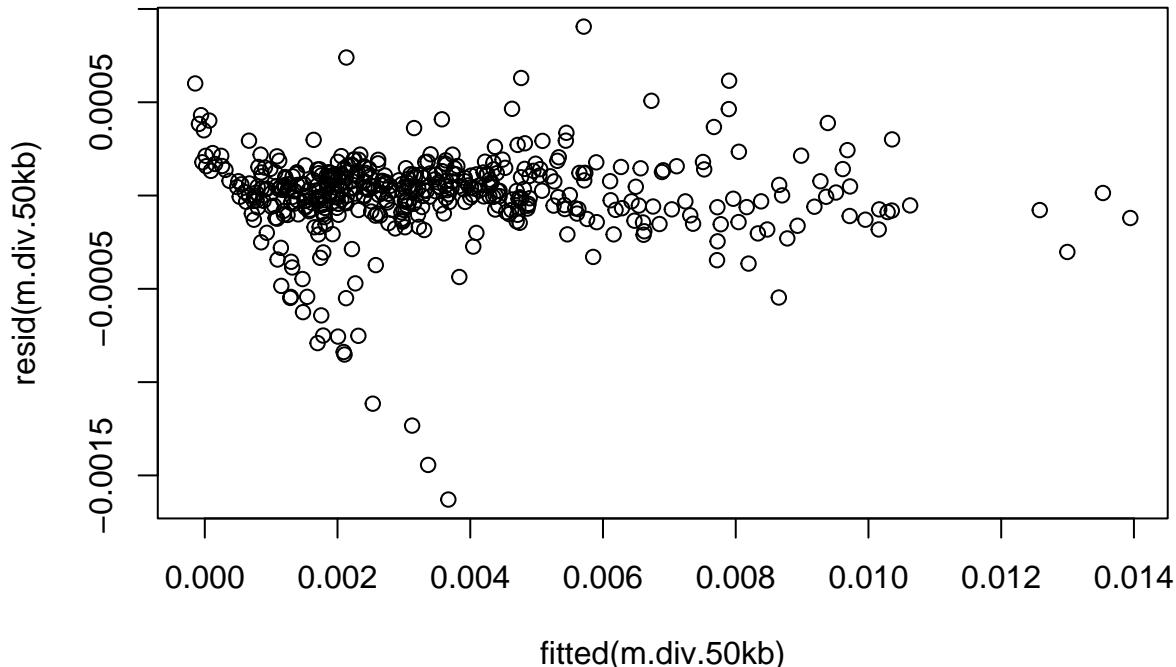
```
AIC(m.div.50kb, m.div.50kb.2, m.div.50kb.3)
```

```

##           df      AIC
## m.div.50kb    6 -6529.539
## m.div.50kb.2  7 -6563.670
## m.div.50kb.3  8 -6563.389

```

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



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

```

##
##  Durbin-Watson test
##
##  data: m.div.50kb
##  DW = 0.80944, p-value < 2.2e-16
##  alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.50kb)

```

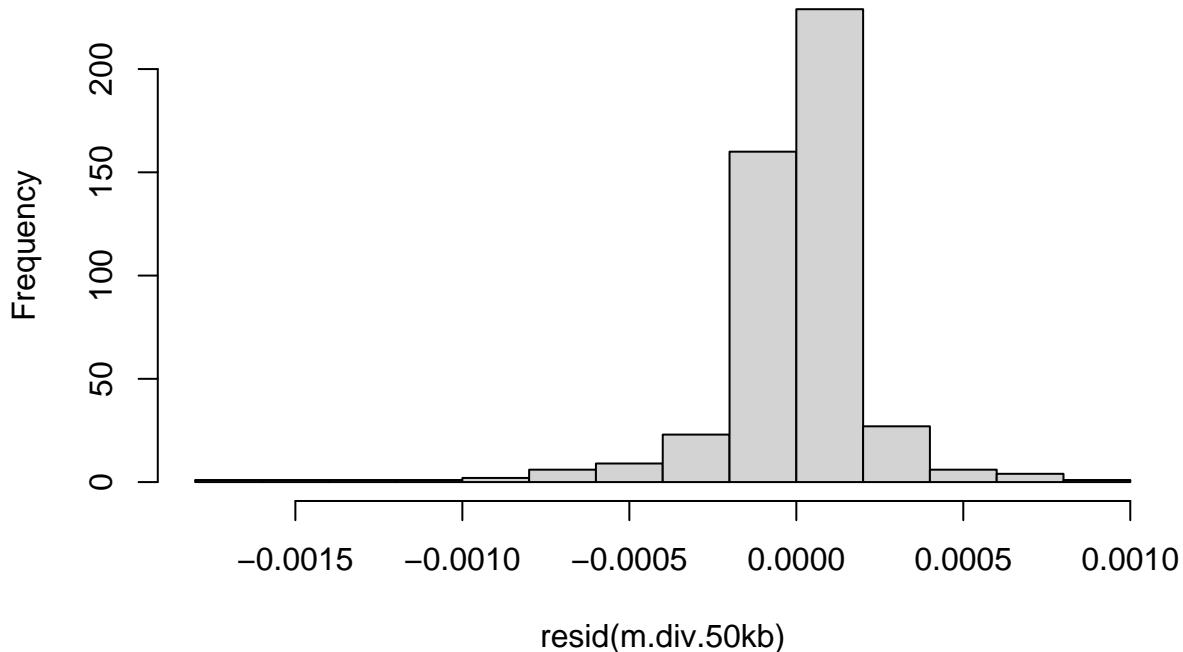
```

##
##  Harrison-McCabe test
##
##  data: m.div.50kb
##  HMC = 0.3513, p-value < 2.2e-16

```

```
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 = inf.lands.50kb.rep7)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max  
## -1.629e-03 -6.430e-05  2.768e-05  1.120e-04  9.053e-04  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3.322e-03 1.142e-05 290.772 < 2e-16 ***  
## thetaC       2.031e-03 1.241e-05 163.749 < 2e-16 ***  
## rhoC        6.334e-05 1.190e-05   5.321 1.61e-07 ***  
## tmrcaC      8.639e-04 1.326e-05  65.143 < 2e-16 ***  
## thetaC:tmrcaC 5.727e-04 1.454e-05  39.399 < 2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.0002345 on 466 degrees of freedom  
## Multiple R-squared:  0.9912, Adjusted R-squared:  0.9912  
## F-statistic: 1.317e+04 on 4 and 466 DF,  p-value: < 2.2e-16  
anova.diversity <- Anova(m.div.50kb)  
apiss <- anova.diversity$"Sum Sq"  
anova.diversity$VarExp <- apiss / sum(apiss)
```

```

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

```

5.1.8 Replicate 8

```

rep8.sim.tmrca.50kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep8_w50000.csv", sep = ",", header = T)

rep_8.pi.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep8/dm2L_bgs_rep8_5.di")
rep_8.pi.50kb$avg <- apply(rep_8.pi.50kb[4:ncol(rep_8.pi.50kb)], 1, mean)
rep_8.rho.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep8/dm2L_bgs_rep8_5.r")
rep_8.theta.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep8/dm2L_bgs_rep8_5.t")
rep_8.tmrca.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep8/dm2L_bgs_rep8_5.tmrca")
rep_8.tmrca.50kb$avg <- apply(rep_8.tmrca.50kb[4:ncol(rep_8.tmrca.50kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.50kb.rep8 <- as.data.frame(cbind(rep_1.pi.50kb$avg, sim.theta.50kb$Rate, rep8.sim.tmrca.50kb))
names(true.lands.50kb.rep8) <- c("diversity", "theta", "tmrca", "rho")

true.lands.50kb.rep8$thetaC <- (true.lands.50kb.rep8$theta - mean(true.lands.50kb.rep8$theta)) / sd(true.lands.50kb.rep8$theta)
true.lands.50kb.rep8$tmrcaC <- (true.lands.50kb.rep8$tmrca - mean(true.lands.50kb.rep8$tmrca)) / sd(true.lands.50kb.rep8$tmrca)
true.lands.50kb.rep8$rhoC <- (true.lands.50kb.rep8$rho - mean(true.lands.50kb.rep8$rho, na.rm = T)) / sd(true.lands.50kb.rep8$rho)

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.50kb.rep8)
anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.50kb.true[1, 8] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.bgs.50kb.true[2, 8] <- anova.diversity$VarExp[1] * 100
r2.bgs.50kb.true[3, 8] <- anova.diversity$VarExp[2] * 100
r2.bgs.50kb.true[4, 8] <- anova.diversity$VarExp[3] * 100
r2.bgs.50kb.true[5, 8] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.50kb.rep8 <- as.data.frame(cbind(rep_8.pi.50kb$avg, rep_8.theta.50kb$sample_mean, rep_8.tmrca.50kb))
names(inf.lands.50kb.rep8) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[8, 1] <- cor.test(sim.theta.50kb$Rate, rep_8.theta.50kb$sample_mean, method = "spearman")
cor.table.rho[8, 1] <- cor.test(sim.rho.50kb$Rate, rep_8.rho.50kb$sample_mean, method = "spearman")$est
cor.table.tmrca[8, 1] <- cor.test(rep8.sim.tmrca.50kb$AverageTmrca, rep_8.tmrca.50kb$avg, method = "spearman")

# standardizing
inf.lands.50kb.rep8$thetaC <- (inf.lands.50kb.rep8$theta - mean(inf.lands.50kb.rep8$theta)) / sd(inf.lands.50kb.rep8$theta)
inf.lands.50kb.rep8$tmrcaC <- (inf.lands.50kb.rep8$tmrca - mean(inf.lands.50kb.rep8$tmrca)) / sd(inf.lands.50kb.rep8$tmrca)
inf.lands.50kb.rep8$rhoC <- (inf.lands.50kb.rep8$rho - mean(inf.lands.50kb.rep8$rho)) / sd(inf.lands.50kb.rep8$rho)

inf.lands.50kb.rep8$bin <- 1:nrow(inf.lands.50kb.rep8)

inf.lands.50kb.rep8$Replicate <- 8

```

```

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.50kb.rep8)
m.div.50kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.50kb.rep8)
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.50kb.rep8)

```

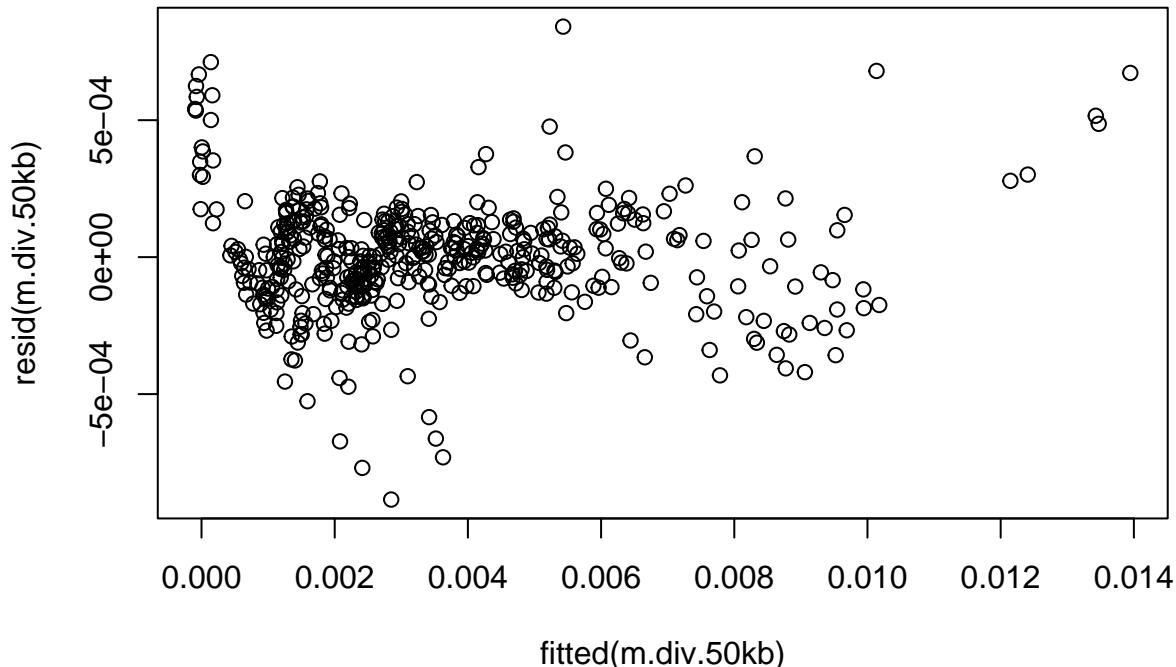
```
AIC(m.div.50kb, m.div.50kb.2, m.div.50kb.3)
```

```

##           df      AIC
## m.div.50kb    6 -6664.906
## m.div.50kb.2  7 -6820.661
## m.div.50kb.3  8 -6830.462

```

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



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

```

##
##  Durbin-Watson test
##
## data: m.div.50kb
## DW = 0.81053, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.50kb)

```

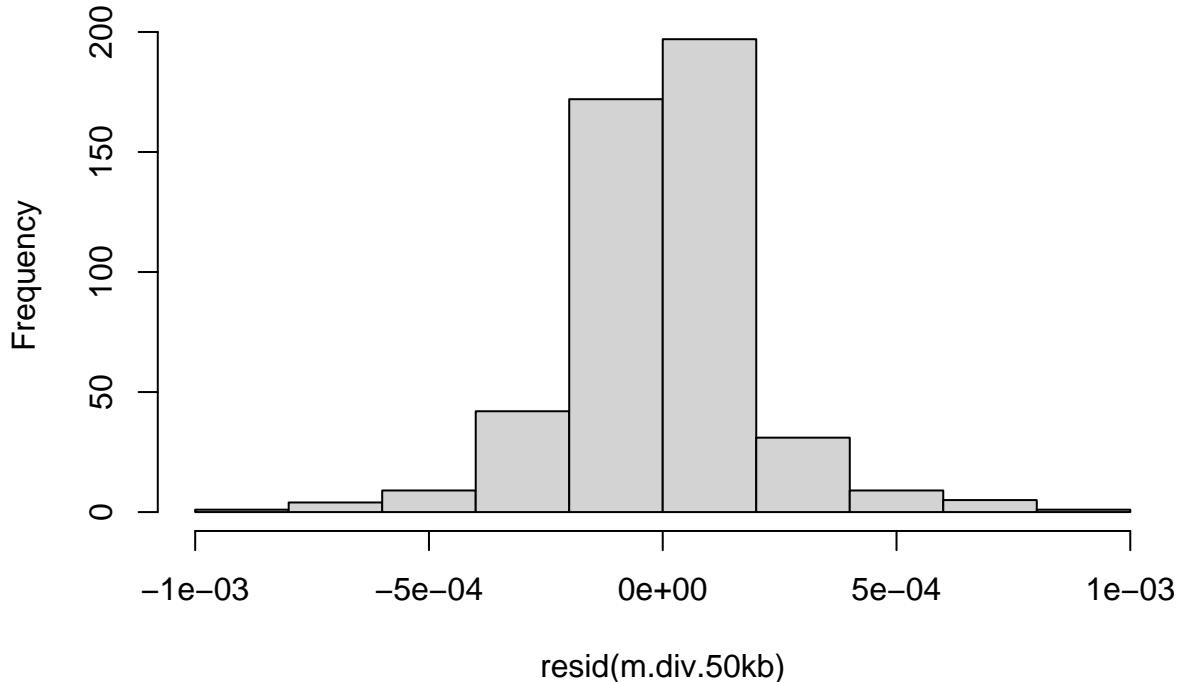
```

##
##  Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.39763, p-value < 2.2e-16

```

```
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 = inf.lands.50kb.rep8)  
##  
## Residuals:  
##       Min        1Q     Median        3Q       Max  
## -8.849e-04 -1.035e-04  4.480e-06  1.033e-04  8.415e-04  
##  
## Coefficients:  
##             Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3.424e-03 9.459e-06 361.965 <2e-16 ***  
## thetaC       1.938e-03 1.024e-05 189.277 <2e-16 ***  
## rhoC         1.664e-05 1.052e-05   1.581   0.115  
## tmrcaC       1.048e-03 1.034e-05 101.373 <2e-16 ***  
## thetaC:tmrcaC 6.299e-04 1.253e-05 50.287 <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.0002031 on 466 degrees of freedom  
## Multiple R-squared:  0.9934, Adjusted R-squared:  0.9934  
## F-statistic: 1.766e+04 on 4 and 466 DF,  p-value: < 2.2e-16  
anova.diversity <- Anova(m.div.50kb)  
apiss <- anova.diversity$"Sum Sq"  
anova.diversity$VarExp <- apiss / sum(apiss)
```

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

5.1.9 Replicate 9

```
rep9.sim.tmrca.50kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep9_w50000.csv", sep = ",", header = T)

rep_9.pi.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep9/dm2L_bgs_rep9_5.di")
rep_9.pi.50kb$avg <- apply(rep_9.pi.50kb[4:ncol(rep_9.pi.50kb)], 1, mean)
rep_9.rho.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep9/dm2L_bgs_rep9_5.r")
rep_9.theta.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep9/dm2L_bgs_rep9_5.t")
rep_9.tmrca.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep9/dm2L_bgs_rep9_5.tmrca")
rep_9.tmrca.50kb$avg <- apply(rep_9.tmrca.50kb[4:ncol(rep_9.tmrca.50kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.50kb.rep9 <- as.data.frame(cbind(rep_1.pi.50kb$avg, sim.theta.50kb$Rate, rep9.sim.tmrca.50kb))
names(true.lands.50kb.rep9) <- c("diversity", "theta", "tmrca", "rho")

true.lands.50kb.rep9$thetaC <- (true.lands.50kb.rep9$theta - mean(true.lands.50kb.rep9$theta)) / sd(true.lands.50kb.rep9$theta)
true.lands.50kb.rep9$tmrcaC <- (true.lands.50kb.rep9$tmrca - mean(true.lands.50kb.rep9$tmrca)) / sd(true.lands.50kb.rep9$tmrca)
true.lands.50kb.rep9$rhoC <- (true.lands.50kb.rep9$rho - mean(true.lands.50kb.rep9$rho, na.rm = T)) / sd(true.lands.50kb.rep9$rho)

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.50kb.rep9)
anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.50kb.true[1, 9] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.bgs.50kb.true[2, 9] <- anova.diversity$VarExp[1] * 100
r2.bgs.50kb.true[3, 9] <- anova.diversity$VarExp[2] * 100
r2.bgs.50kb.true[4, 9] <- anova.diversity$VarExp[3] * 100
r2.bgs.50kb.true[5, 9] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.50kb.rep9 <- as.data.frame(cbind(rep_9.pi.50kb$avg, rep_9.theta.50kb$sample_mean, rep_9.tmrca.50kb$sample_mean))
names(inf.lands.50kb.rep9) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[9, 1] <- cor.test(sim.theta.50kb$Rate, rep_9.theta.50kb$sample_mean, method = "spearman")
cor.table.rho[9, 1] <- cor.test(sim.rho.50kb$Rate, rep_9.rho.50kb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[9, 1] <- cor.test(rep9.sim.tmrca.50kb$AverageTmrca, rep_9.tmrca.50kb$avg, method = "spearman")$estimate

# standardizing
inf.lands.50kb.rep9$thetaC <- (inf.lands.50kb.rep9$theta - mean(inf.lands.50kb.rep9$theta)) / sd(inf.lands.50kb.rep9$theta)
inf.lands.50kb.rep9$tmrcaC <- (inf.lands.50kb.rep9$tmrca - mean(inf.lands.50kb.rep9$tmrca)) / sd(inf.lands.50kb.rep9$tmrca)
inf.lands.50kb.rep9$rhoC <- (inf.lands.50kb.rep9$rho - mean(inf.lands.50kb.rep9$rho)) / sd(inf.lands.50kb.rep9$rho)

inf.lands.50kb.rep9$bin <- 1:nrow(inf.lands.50kb.rep9)

inf.lands.50kb.rep9$Replicate <- 9
```

```

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.50kb.rep9)
m.div.50kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.50kb.rep9)
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.50kb.rep9)

```

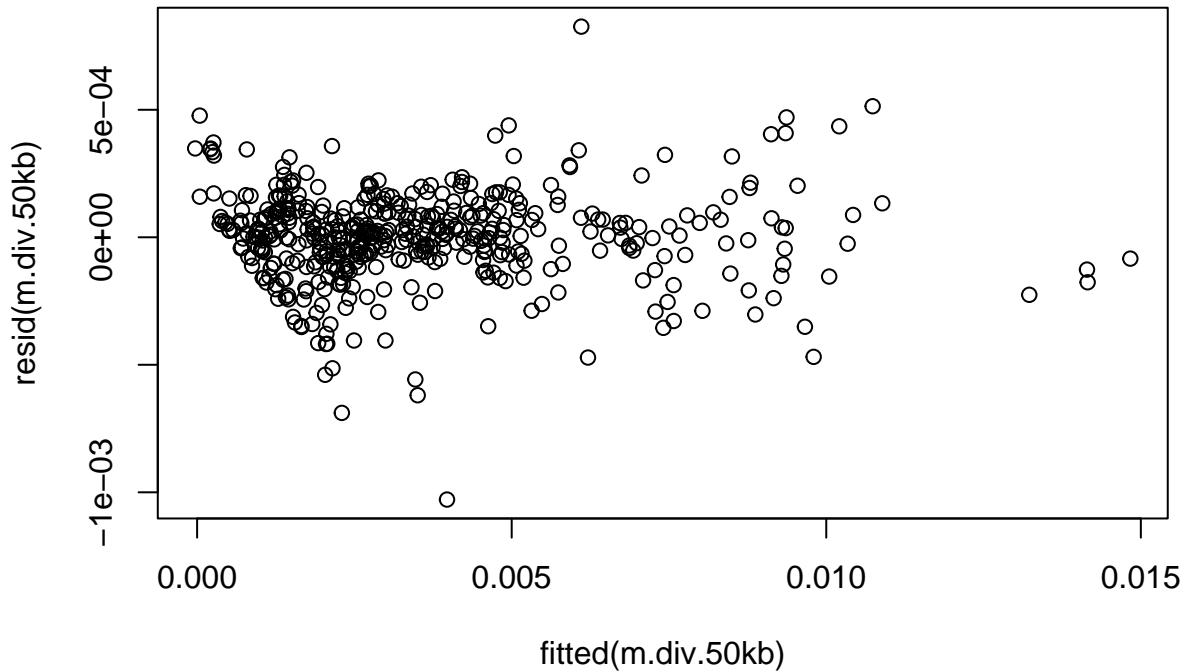
```
AIC(m.div.50kb, m.div.50kb.2, m.div.50kb.3)
```

```

##           df      AIC
## m.div.50kb    6 -6772.620
## m.div.50kb.2  7 -6788.142
## m.div.50kb.3  8 -6788.032

```

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



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

```

##
##  Durbin-Watson test
##
##  data: m.div.50kb
##  DW = 1.2553, p-value < 2.2e-16
##  alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.50kb)

```

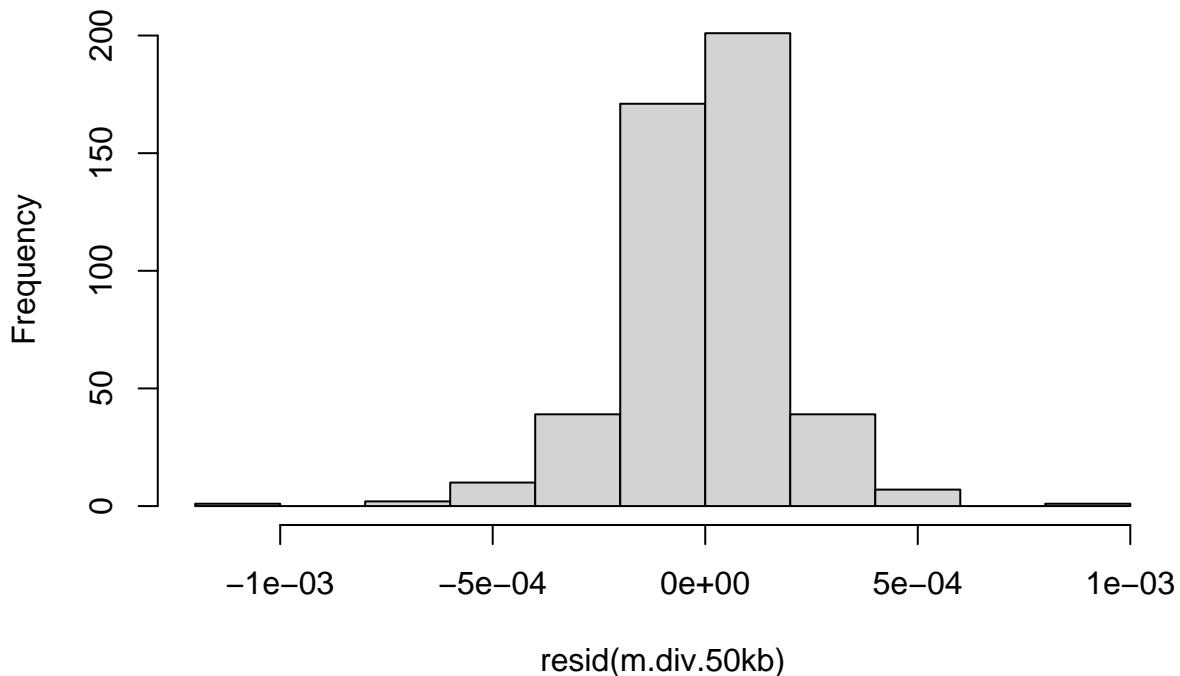
```

##
##  Harrison-McCabe test
##
##  data: m.div.50kb
##  HMC = 0.48698, p-value = 0.343

```

```
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 = inf.lands.50kb.rep9)  
##  
## Residuals:  
##       Min        1Q     Median        3Q       Max  
## -1.028e-03 -7.795e-05  1.040e-05  9.457e-05  8.260e-04  
##  
## Coefficients:  
##             Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3.386e-03 8.797e-06 384.938 < 2e-16 ***  
## thetaC       1.948e-03 9.855e-06 197.676 < 2e-16 ***  
## rhoC         3.811e-05 9.451e-06   4.032 6.45e-05 ***  
## tmrcaC       9.816e-04 1.021e-05  96.094 < 2e-16 ***  
## thetaC:tmrcaC 6.367e-04 1.178e-05  54.050 < 2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.0001812 on 466 degrees of freedom  
## Multiple R-squared:  0.9949, Adjusted R-squared:  0.9948  
## F-statistic: 2.265e+04 on 4 and 466 DF,  p-value: < 2.2e-16  
anova.diversity <- Anova(m.div.50kb)  
apiss <- anova.diversity$"Sum Sq"  
anova.diversity$VarExp <- apiss / sum(apiss)
```

```

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

```

5.1.10 Replicate 10

```

rep10.sim.tmrca.50kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep10_w50000.csv", sep = ",", header = T)

rep_10.pi.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep10/dm2L_bgs_rep10_50kb.csv")
rep_10.pi.50kb$avg <- apply(rep_10.pi.50kb[4:ncol(rep_10.pi.50kb)], 1, mean)
rep_10.rho.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep10/dm2L_bgs_rep10_50kb.csv")
rep_10.theta.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep10/dm2L_bgs_rep10_50kb.csv")
rep_10.tmrca.50kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep10/dm2L_bgs_rep10_50kb.csv")
rep_10.tmrca.50kb$avg <- apply(rep_10.tmrca.50kb[4:ncol(rep_10.tmrca.50kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.50kb.rep10 <- as.data.frame(cbind(rep_1.pi.50kb$avg, sim.theta.50kb$Rate, rep10.sim.tmrca.50kb$AverageTmrca))
names(true.lands.50kb.rep10) <- c("diversity", "theta", "tmrca", "rho")

true.lands.50kb.rep10$thetaC <- (true.lands.50kb.rep10$theta - mean(true.lands.50kb.rep10$theta)) / sd(true.lands.50kb.rep10$theta)
true.lands.50kb.rep10$tmrcaC <- (true.lands.50kb.rep10$tmrca - mean(true.lands.50kb.rep10$tmrca)) / sd(true.lands.50kb.rep10$tmrca)
true.lands.50kb.rep10$rhoC <- (true.lands.50kb.rep10$rho - mean(true.lands.50kb.rep10$rho), na.rm = T))

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.50kb.rep10)
anova.diversity <- Anova(m.div.50kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.50kb.true[1, 10] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.bgs.50kb.true[2, 10] <- anova.diversity$VarExp[1] * 100
r2.bgs.50kb.true[3, 10] <- anova.diversity$VarExp[2] * 100
r2.bgs.50kb.true[4, 10] <- anova.diversity$VarExp[3] * 100
r2.bgs.50kb.true[5, 10] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.50kb.rep10 <- as.data.frame(cbind(rep_10.pi.50kb$avg, rep_10.theta.50kb$sample_mean, rep_10.theta.50kb$sample_sd, rep_10.tmrca.50kb$sample_mean, rep_10.tmrca.50kb$sample_sd))
names(inf.lands.50kb.rep10) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[10, 1] <- cor.test(sim.theta.50kb$Rate, rep_10.theta.50kb$sample_mean, method = "spearman")
cor.table.rho[10, 1] <- cor.test(sim.rho.50kb$Rate, rep_10.rho.50kb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[10, 1] <- cor.test(rep10.sim.tmrca.50kb$AverageTmrca, rep_10.tmrca.50kb$avg, method = "spearman")$estimate

# standardizing
inf.lands.50kb.rep10$thetaC <- (inf.lands.50kb.rep10$theta - mean(inf.lands.50kb.rep10$theta)) / sd(inf.lands.50kb.rep10$theta)
inf.lands.50kb.rep10$tmrcaC <- (inf.lands.50kb.rep10$tmrca - mean(inf.lands.50kb.rep10$tmrca)) / sd(inf.lands.50kb.rep10$tmrca)
inf.lands.50kb.rep10$rhoC <- (inf.lands.50kb.rep10$rho - mean(inf.lands.50kb.rep10$rho)) / sd(inf.lands.50kb.rep10$rho)

inf.lands.50kb.rep10$bin <- 1:nrow(inf.lands.50kb.rep10)

inf.lands.50kb.rep10$Replicate <- 10

```

```

m.div.50kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.50kb.rep10)
m.div.50kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.50kb.rep10)
m.div.50kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.50kb.rep10)

```

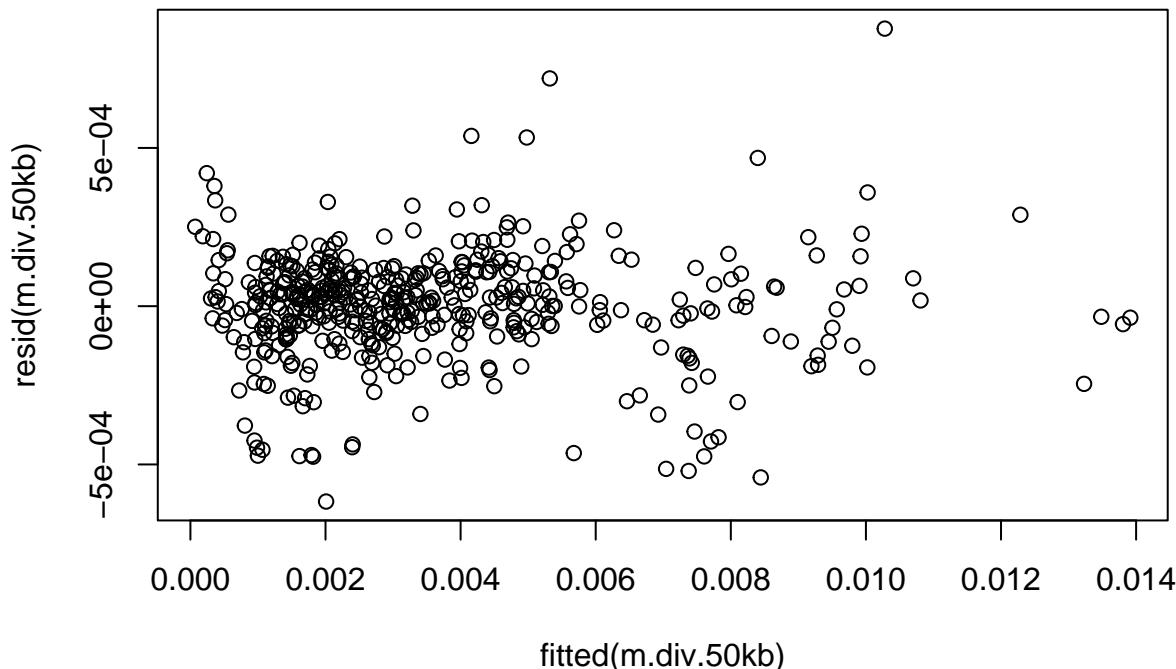
```
AIC(m.div.50kb, m.div.50kb.2, m.div.50kb.3)
```

```

##           df      AIC
## m.div.50kb    6 -6830.551
## m.div.50kb.2  7 -6838.420
## m.div.50kb.3  8 -6849.626

```

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



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

```

##
##  Durbin-Watson test
##
## data: m.div.50kb
## DW = 1.1672, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.50kb)

```

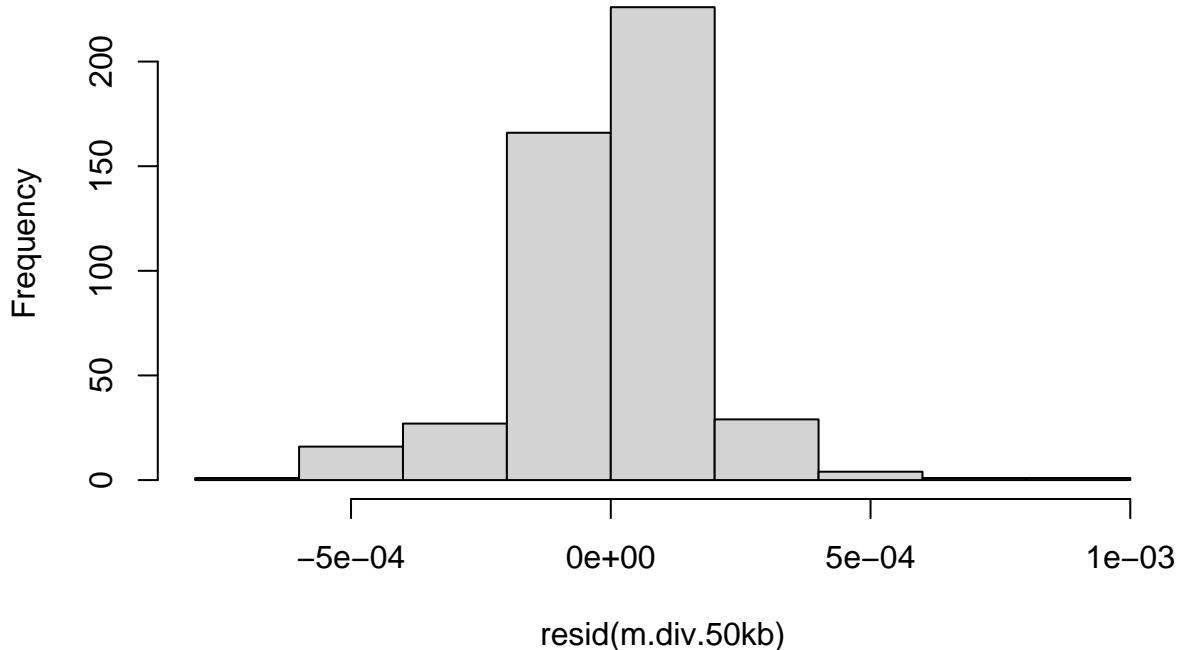
```

##
##  Harrison-McCabe test
##
## data: m.div.50kb
## HMC = 0.47978, p-value = 0.298

```

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

Histogram of resid(m.div.50kb)



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

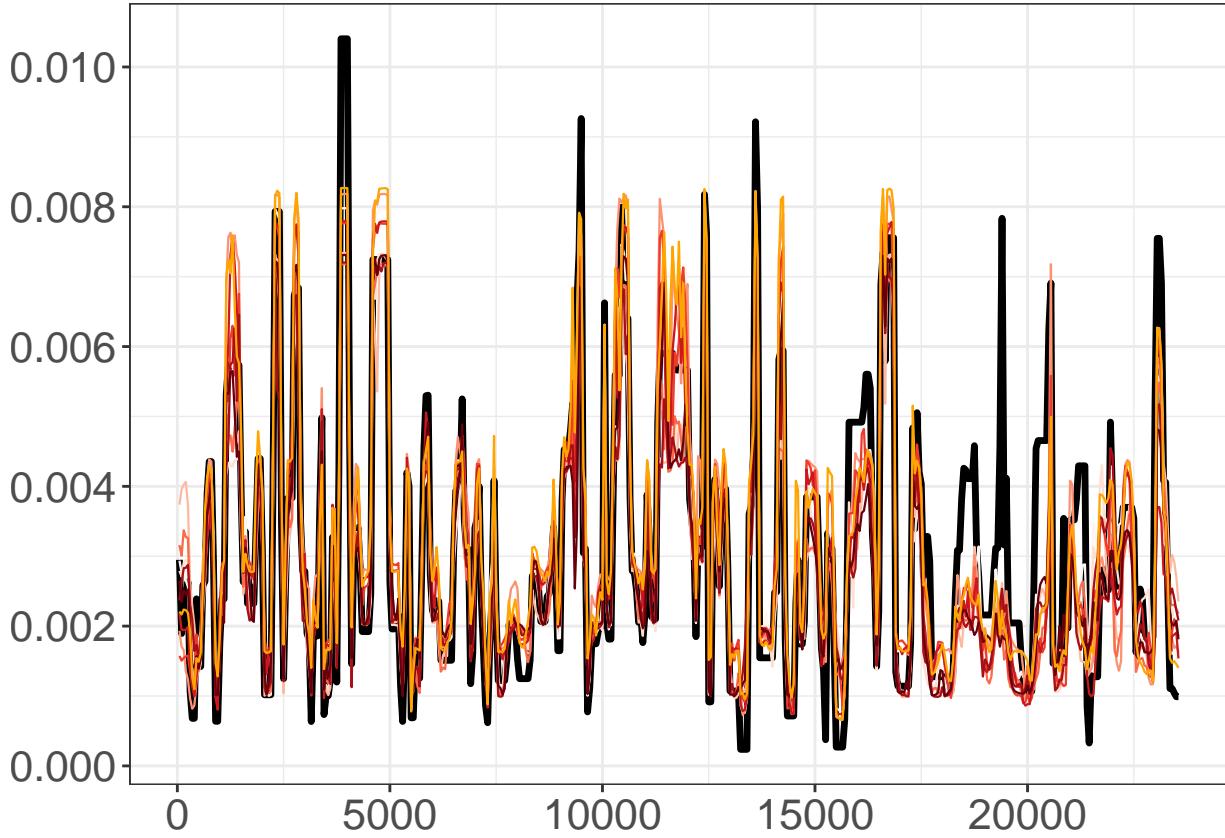
```
##  
## Call:  
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC,  
##      data = inf.lands.50kb.rep10)  
##  
## Residuals:  
##       Min        1Q     Median        3Q       Max  
## -6.172e-04 -6.760e-05  1.266e-05  9.275e-05  8.769e-04  
##  
## Coefficients:  
##             Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3.443e-03 8.020e-06 429.318 < 2e-16 ***  
## thetaC       2.093e-03 8.540e-06 245.081 < 2e-16 ***  
## rhoC         6.913e-05 9.289e-06  7.442 4.84e-13 ***  
## tmrcaC       8.255e-04 8.971e-06  92.015 < 2e-16 ***  
## thetaC:tmrcaC 5.158e-04 8.760e-06  58.877 < 2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.0001704 on 466 degrees of freedom  
## Multiple R-squared:  0.9955, Adjusted R-squared:  0.9955  
## F-statistic: 2.585e+04 on 4 and 466 DF,  p-value: < 2.2e-16  
anova.diversity <- Anova(m.div.50kb)  
apiss <- anova.diversity$"Sum Sq"  
anova.diversity$VarExp <- apiss / sum(apiss)
```

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

5.2 all replicates:

```
theta.plot <- as.data.frame(cbind(1:nrow(sim.theta.50kb),
                                    sim.theta.50kb$Rate * 1e-7 * 3e+4,
                                    rep_1.theta.50kb$sample_mean,
                                    rep_2.theta.50kb$sample_mean,
                                    rep_3.theta.50kb$sample_mean,
                                    rep_4.theta.50kb$sample_mean,
                                    rep_5.theta.50kb$sample_mean,
                                    rep_6.theta.50kb$sample_mean,
                                    rep_7.theta.50kb$sample_mean,
                                    rep_8.theta.50kb$sample_mean,
                                    rep_9.theta.50kb$sample_mean,
                                    rep_10.theta.50kb$sample_mean))

names(theta.plot) <- c("bin", "sim", "reps")
molten.theta <- melt(theta.plot, id.vars = "bin")
theta.map.50kb.bgs <- ggplot(data = molten.theta, aes(x = bin * 50, y = value, colour = variable)) + theta
theta.map.50kb.bgs <- theta.map.50kb.bgs + geom_line(data = molten.theta, aes(size = variable)) + scale_
theta.map.50kb.bgs <- theta.map.50kb.bgs + scale_color_manual(values = c("black", brewer.pal(n = 9, name =
theta.map.50kb.bgs <- theta.map.50kb.bgs + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continu
theta.map.50kb.bgs <- theta.map.50kb.bgs + labs(title = NULL, x = NULL, y = NULL)
theta.map.50kb.bgs <- theta.map.50kb.bgs + theme(text = element_text(size = 20), axis.title.x = element_
theta.map.50kb.bgs
```



5.3 200 kb scale

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

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

sim.theta.200kb <- read.table("dm_bgs_sims/MutationMap_200kb.csv", sep = ",", header = T)
sim.rho.200kb <- read.table("dm_bgs_sims/RecombinationMap_200kb.csv", sep = ",", header = T)
```

5.3.1 Replicate 1

```
rep1.sim.tmrca.200kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep1_w200000.csv", sep = ",", header = T)

rep_1.pi.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep1/dm2L_bgs_rep1_5.d")
rep_1.pi.200kb$avg <- apply(rep_1.pi.200kb[4:ncol(rep_1.pi.200kb)], 1, mean)
rep_1.rho.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep1/dm2L_bgs_rep1_5.d")
rep_1.theta.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep1/dm2L_bgs_rep1_5.d")
rep_1.tmrca.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep1/dm2L_bgs_rep1_5.d")
rep_1.tmrca.200kb$avg <- apply(rep_1.tmrca.200kb[4:ncol(rep_1.tmrca.200kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.200kb.rep1 <- as.data.frame(cbind(rep_1.pi.200kb$avg, sim.theta.200kb$Rate, rep1.sim.tmrca.200kb$avg))
```

```

names(true.lands.200kb.rep1) <- c("diversity", "theta", "tmrca", "rho")
true.lands.200kb.rep1$thetaC <- (true.lands.200kb.rep1$theta - mean(true.lands.200kb.rep1$theta)) / sd(true.lands.200kb.rep1$theta)
true.lands.200kb.rep1$tmrcaC <- (true.lands.200kb.rep1$tmrca - mean(true.lands.200kb.rep1$tmrca)) / sd(true.lands.200kb.rep1$tmrca)
true.lands.200kb.rep1$rhoC <- (true.lands.200kb.rep1$rho - mean(true.lands.200kb.rep1$rho), na.rm = T))

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.200kb.rep1)
anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.200kb.true[1, 1] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]) * 100
r2.bgs.200kb.true[2, 1] <- anova.diversity$VarExp[1] * 100
r2.bgs.200kb.true[3, 1] <- anova.diversity$VarExp[2] * 100
r2.bgs.200kb.true[4, 1] <- anova.diversity$VarExp[3] * 100
r2.bgs.200kb.true[5, 1] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.200kb.rep1 <- as.data.frame(cbind(rep_1.pi.200kb$avg, rep_1.theta.200kb$sample_mean, rep_1.tmrca.200kb$sample_mean))
names(inf.lands.200kb.rep1) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[1, 2] <- cor.test(sim.theta.200kb$Rate, rep_1.theta.200kb$sample_mean, method = "spearman")
cor.table.rho[1, 2] <- cor.test(sim.rho.200kb$Rate, rep_1.rho.200kb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[1, 2] <- cor.test(rep1.sim.tmrca.200kb$AverageTmrca, rep_1.tmrca.200kb$avg, method = "spearman")$estimate

# standardizing
inf.lands.200kb.rep1$thetaC <- (inf.lands.200kb.rep1$theta - mean(inf.lands.200kb.rep1$theta)) / sd(inf.lands.200kb.rep1$theta)
inf.lands.200kb.rep1$tmrcaC <- (inf.lands.200kb.rep1$tmrca - mean(inf.lands.200kb.rep1$tmrca)) / sd(inf.lands.200kb.rep1$tmrca)
inf.lands.200kb.rep1$rhoC <- (inf.lands.200kb.rep1$rho - mean(inf.lands.200kb.rep1$rho)) / sd(inf.lands.200kb.rep1$rho)

inf.lands.200kb.rep1$bin <- 1:nrow(inf.lands.200kb.rep1)

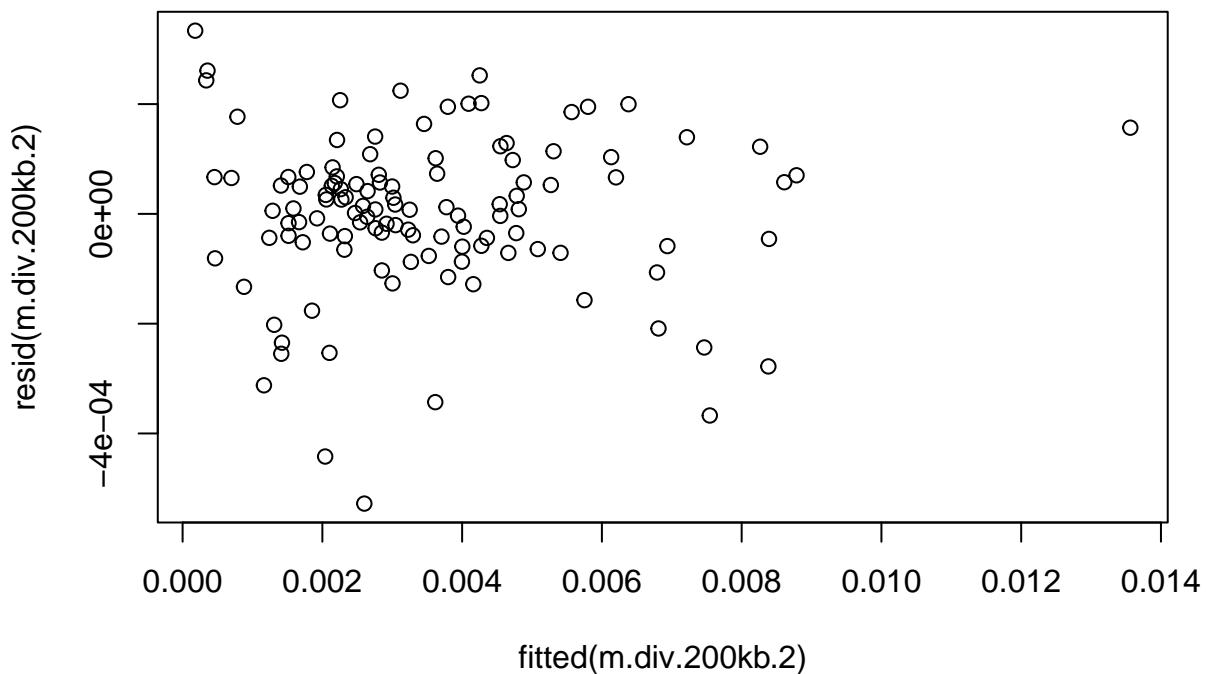
inf.lands.200kb.rep1$Replicate <- 1

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.200kb.rep1)
m.div.200kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.200kb.rep1)
m.div.200kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC)^2, data = inf.lands.200kb.rep1)

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

##          df      AIC
## m.div.200kb    6 -1724.393
## m.div.200kb.2  7 -1740.570
## m.div.200kb.3  8 -1739.520
plot(resid(m.div.200kb.2)~fitted(m.div.200kb.2))

```

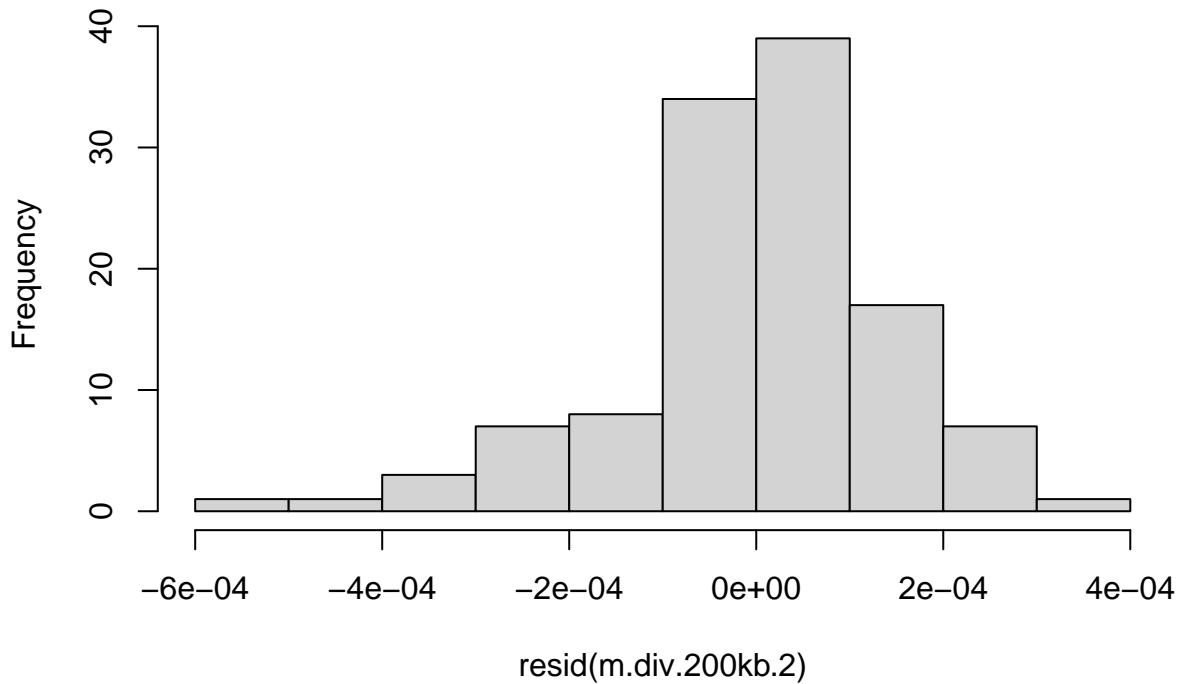


```
dwtest(m.div.200kb.2)
```

```
##
##  Durbin-Watson test
##
## data: m.div.200kb.2
## DW = 1.167, p-value = 1.002e-06
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.200kb.2)
```

```
##
##  Harrison-McCabe test
##
## data: m.div.200kb.2
## HMC = 0.38434, p-value = 0.044
hist(resid(m.div.200kb.2))
```

Histogram of resid(m.div.200kb.2)



```
summary(m.div.200kb.2)

##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC +
##     rhoC:tmrcaC, data = inf.lands.200kb.rep1)
##
## Residuals:
##      Min        1Q    Median        3Q       Max 
## -5.278e-04 -5.856e-05  9.240e-06  7.099e-05  3.337e-04
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.415e-03 1.769e-05 193.085 < 2e-16 ***
## thetaC      1.701e-03 1.584e-05 107.387 < 2e-16 ***
## rhoC        5.423e-05 1.762e-05  3.077  0.00262 ** 
## tmrcaC      9.552e-04 3.711e-05 25.740 < 2e-16 ***
## thetaC:tmrcaC 4.314e-04 1.988e-05 21.699 < 2e-16 ***
## rhoC:tmrcaC  1.178e-04 2.728e-05  4.319 3.41e-05 ***
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.0001466 on 112 degrees of freedom
## Multiple R-squared:  0.9957, Adjusted R-squared:  0.9955 
## F-statistic: 5132 on 5 and 112 DF,  p-value: < 2.2e-16

anova.diversity <- Anova(m.div.200kb.2)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
```

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

5.3.2 Replicate 2

```
rep2.sim.tmrca.200kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep2_w200000.csv", sep = ",", header = T)

rep_2.pi.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep2/dm2L_bgs_rep2_5.dim")
rep_2.pi.200kb$avg <- apply(rep_2.pi.200kb[4:ncol(rep_2.pi.200kb)], 1, mean)
rep_2.rho.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep2/dm2L_bgs_rep2_5.dim")
rep_2.theta.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep2/dm2L_bgs_rep2_5.dim")
rep_2.tmrca.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep2/dm2L_bgs_rep2_5.dim")
rep_2.tmrca.200kb$avg <- apply(rep_2.tmrca.200kb[4:ncol(rep_2.tmrca.200kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.200kb.rep2 <- as.data.frame(cbind(rep_1.pi.200kb$avg, sim.theta.200kb$Rate, rep2.sim.tmrca.200kb$AvgTmrca))
names(true.lands.200kb.rep2) <- c("diversity", "theta", "tmrca", "rho")
true.lands.200kb.rep2$thetaC <- (true.lands.200kb.rep2$theta - mean(true.lands.200kb.rep2$theta)) / sd(true.lands.200kb.rep2$theta)
true.lands.200kb.rep2$tmrcaC <- (true.lands.200kb.rep2$tmrca - mean(true.lands.200kb.rep2$tmrca)) / sd(true.lands.200kb.rep2$tmrca)
true.lands.200kb.rep2$rhoC <- (true.lands.200kb.rep2$rho - mean(true.lands.200kb.rep2$rho), na.rm = T))

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.200kb.rep2)
anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.200kb.true[1, 2] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]) * 100
r2.bgs.200kb.true[2, 2] <- anova.diversity$VarExp[1] * 100
r2.bgs.200kb.true[3, 2] <- anova.diversity$VarExp[2] * 100
r2.bgs.200kb.true[4, 2] <- anova.diversity$VarExp[3] * 100
r2.bgs.200kb.true[5, 2] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.200kb.rep2 <- as.data.frame(cbind(rep_2.pi.200kb$avg, rep_2.theta.200kb$sample_mean, rep_2.tmrca.200kb$AvgTmrca))
names(inf.lands.200kb.rep2) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[2, 2] <- cor.test(sim.theta.200kb$Rate, rep_2.theta.200kb$sample_mean, method = "spearman")
cor.table.rho[2, 2] <- cor.test(sim.rho.200kb$Rate, rep_2.rho.200kb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[2, 2] <- cor.test(rep2.sim.tmrca.200kb$AverageTmrca, rep_2.tmrca.200kb$avg, method = "spearman")$estimate

# standardizing
inf.lands.200kb.rep2$thetaC <- (inf.lands.200kb.rep2$theta - mean(inf.lands.200kb.rep2$theta)) / sd(inf.lands.200kb.rep2$theta)
inf.lands.200kb.rep2$tmrcaC <- (inf.lands.200kb.rep2$tmrca - mean(inf.lands.200kb.rep2$tmrca)) / sd(inf.lands.200kb.rep2$tmrca)
inf.lands.200kb.rep2$rhoC <- (inf.lands.200kb.rep2$rho - mean(inf.lands.200kb.rep2$rho)) / sd(inf.lands.200kb.rep2$rho)

inf.lands.200kb.rep2$bin <- 1:nrow(inf.lands.200kb.rep2)

inf.lands.200kb.rep2$Replicate <- 2

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.200kb.rep2)
```

```

m.div.200kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.200kb)
m.div.200kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC)^2, data = inf.lands.200kb.rep2)

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

##             df      AIC
## m.div.200kb    6 -1700.943
## m.div.200kb.2  7 -1724.292
## m.div.200kb.3  8 -1728.465
plot(resid(m.div.200kb)~fitted(m.div.200kb))



```

dwtest(m.div.200kb)

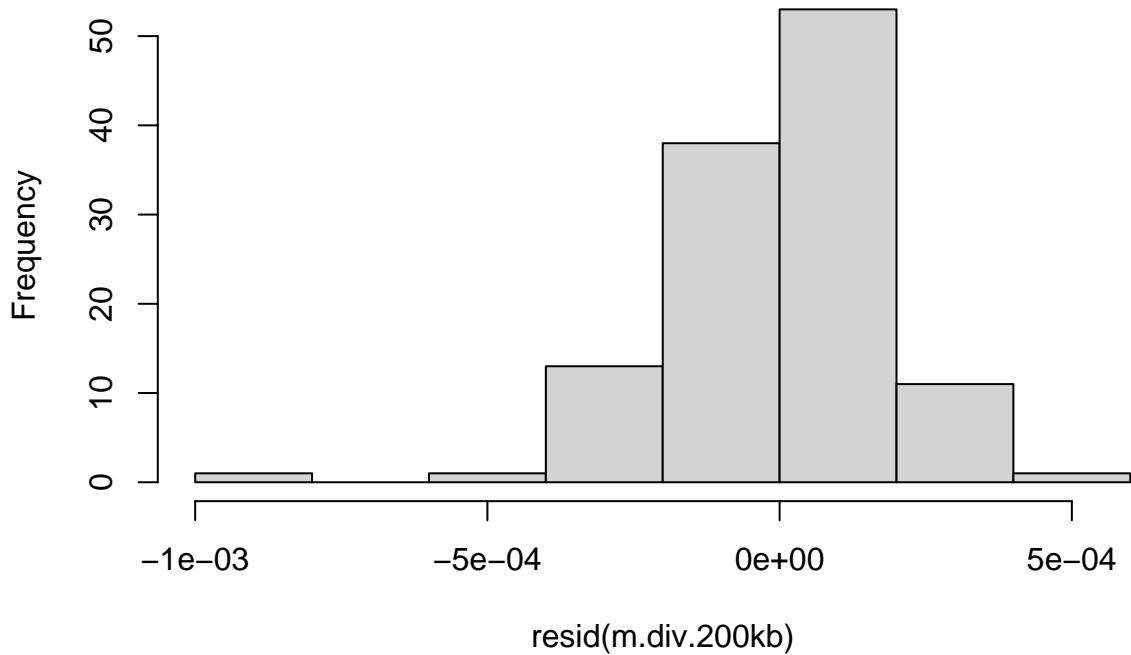
```

##
## Durbin-Watson test
##
## data: m.div.200kb
## DW = 1.4796, p-value = 0.001381
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.200kb)

##
## Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.31601, p-value < 2.2e-16
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 = inf.lands.200kb.rep2)  
##  
## Residuals:  
##       Min        1Q     Median        3Q       Max  
## -8.069e-04 -8.170e-05  1.681e-05  8.234e-05  4.681e-04  
##  
## Coefficients:  
##             Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3.490e-03 1.608e-05 217.015 <2e-16 ***  
## thetaC      1.700e-03 1.742e-05  97.564 <2e-16 ***  
## rhoC        2.854e-05 1.858e-05   1.536  0.127  
## tmrcaC      9.721e-04 1.789e-05  54.351 <2e-16 ***  
## thetaC:tmrcaC 4.837e-04 1.808e-05  26.756 <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.0001742 on 113 degrees of freedom  
## Multiple R-squared:  0.9939, Adjusted R-squared:  0.9937  
## F-statistic:  4640 on 4 and 113 DF,  p-value: < 2.2e-16  
anova.diversity <- Anova(m.div.200kb)  
apiss <- anova.diversity$"Sum Sq"  
anova.diversity$VarExp <- apiss / sum(apiss)
```

```
r2.bgs.200kb[1, 2] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
```

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

5.3.3 Replicate 3

```
rep3.sim.tmrca.200kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep3_w200000.csv", sep = ",", header = T)

rep_3.pi.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep3/dm2L_bgs_rep3_5.d")
rep_3.pi.200kb$avg <- apply(rep_3.pi.200kb[4:ncol(rep_3.pi.200kb)], 1, mean)
rep_3.rho.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep3/dm2L_bgs_rep3_5.d")
rep_3.theta.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep3/dm2L_bgs_rep3_5.d")
rep_3.tmrca.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep3/dm2L_bgs_rep3_5.d")
rep_3.tmrca.200kb$avg <- apply(rep_3.tmrca.200kb[4:ncol(rep_3.tmrca.200kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.200kb.rep3 <- as.data.frame(cbind(rep_1.pi.200kb$avg, sim.theta.200kb$Rate, rep3.sim.tmrca.200kb$AvgTmrca))
names(true.lands.200kb.rep3) <- c("diversity", "theta", "tmrca", "rho")
true.lands.200kb.rep3$thetaC <- (true.lands.200kb.rep3$theta - mean(true.lands.200kb.rep3$theta)) / sd(true.lands.200kb.rep3$theta)
true.lands.200kb.rep3$tmrcaC <- (true.lands.200kb.rep3$tmrca - mean(true.lands.200kb.rep3$tmrca)) / sd(true.lands.200kb.rep3$tmrca)
true.lands.200kb.rep3$rhoC <- (true.lands.200kb.rep3$rho - mean(true.lands.200kb.rep3$rho), na.rm = T))

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.200kb.rep3)
anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.200kb.true[1, 3] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.bgs.200kb.true[2, 3] <- anova.diversity$VarExp[1] * 100
r2.bgs.200kb.true[3, 3] <- anova.diversity$VarExp[2] * 100
r2.bgs.200kb.true[4, 3] <- anova.diversity$VarExp[3] * 100
r2.bgs.200kb.true[5, 3] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.200kb.rep3 <- as.data.frame(cbind(rep_3.pi.200kb$avg, rep_3.theta.200kb$sample_mean, rep_3.tmrca.200kb$AvgTmrca))
names(inf.lands.200kb.rep3) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[3, 2] <- cor.test(sim.theta.200kb$Rate, rep_3.theta.200kb$sample_mean, method = "spearman")
cor.table.rho[3, 2] <- cor.test(sim.rho.200kb$Rate, rep_3.rho.200kb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[3, 2] <- cor.test(rep3.sim.tmrca.200kb$AverageTmrca, rep_3.tmrca.200kb$avg, method = "spearman")$estimate

# standardizing
inf.lands.200kb.rep3$thetaC <- (inf.lands.200kb.rep3$theta - mean(inf.lands.200kb.rep3$theta)) / sd(inf.lands.200kb.rep3$theta)
inf.lands.200kb.rep3$tmrcaC <- (inf.lands.200kb.rep3$tmrca - mean(inf.lands.200kb.rep3$tmrca)) / sd(inf.lands.200kb.rep3$tmrca)
inf.lands.200kb.rep3$rhoC <- (inf.lands.200kb.rep3$rho - mean(inf.lands.200kb.rep3$rho)) / sd(inf.lands.200kb.rep3$rho)

inf.lands.200kb.rep3$bin <- 1:nrow(inf.lands.200kb.rep3)

inf.lands.200kb.rep3$Replicate <- 3

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.200kb.rep3)
m.div.200kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.200kb.rep3)
```

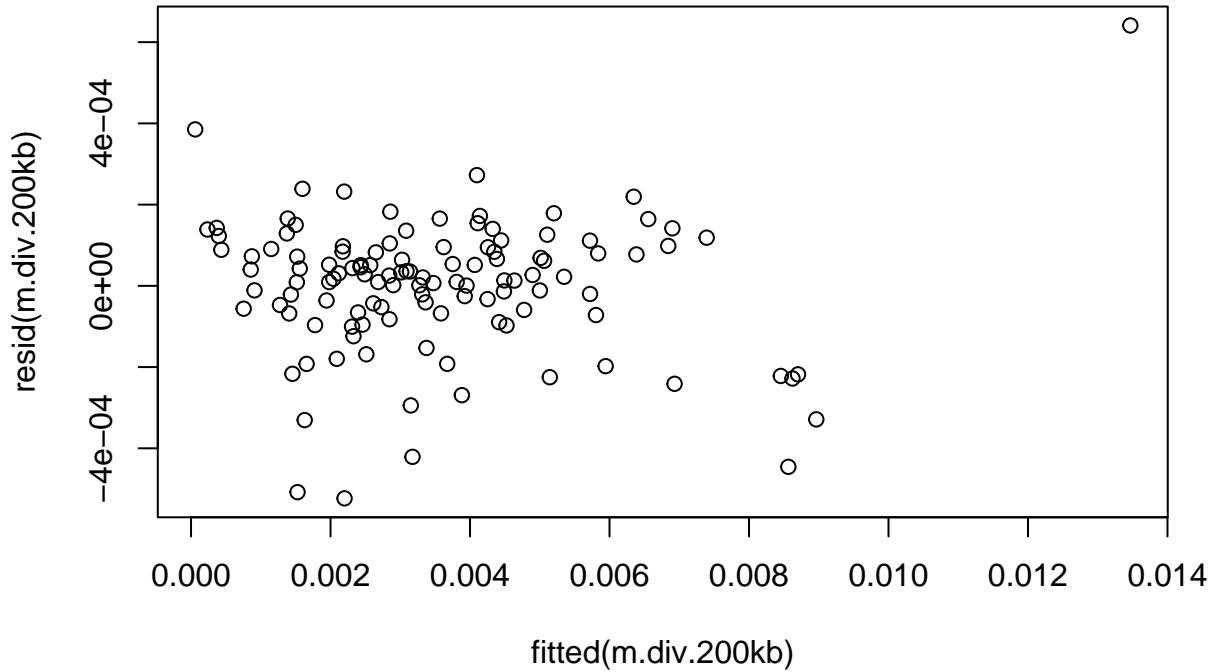
```

m.div.200kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.200kb.rep3)

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

##           df      AIC
## m.div.200kb     6 -1707.314
## m.div.200kb.2   7 -1754.542
## m.div.200kb.3   8 -1764.055
plot(resid(m.div.200kb)~fitted(m.div.200kb))

```



```

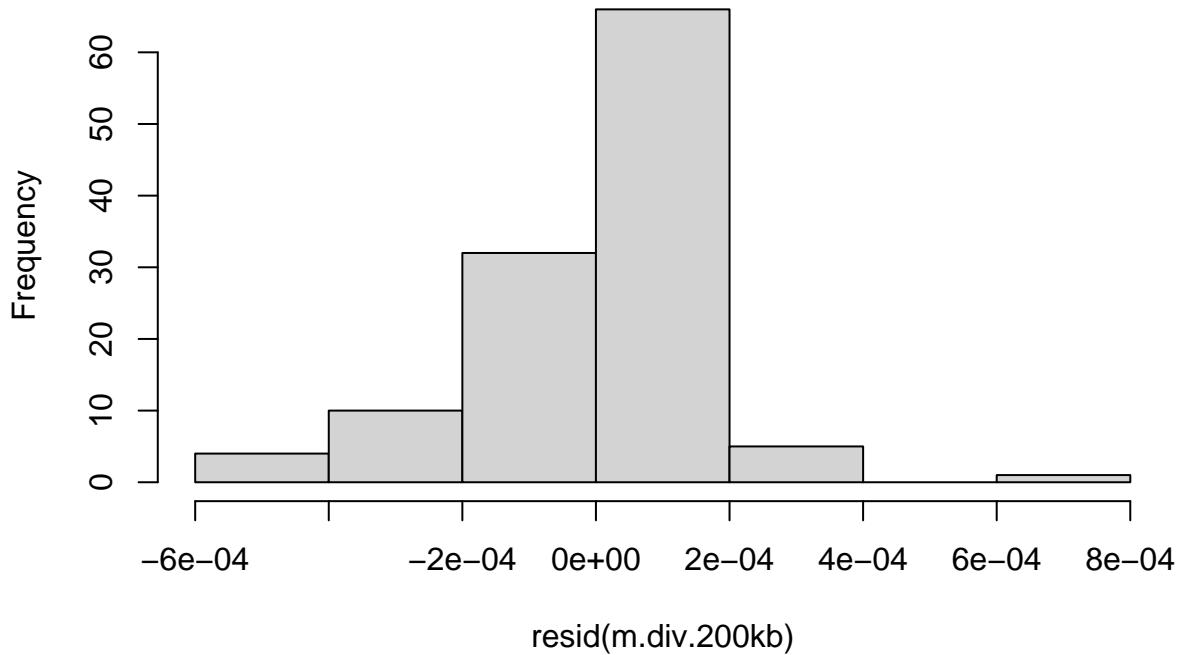
dwtest(m.div.200kb)

##
##  Durbin-Watson test
##
## data: m.div.200kb
## DW = 1.3035, p-value = 3.818e-05
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.200kb)

##
##  Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.44743, p-value = 0.219
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 = inf.lands.200kb.rep3)
##
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -5.233e-04 -6.723e-05  2.374e-05  9.394e-05  6.412e-04
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.519e-03 1.563e-05 225.098 <2e-16 ***
## thetaC      1.728e-03 1.676e-05 103.074 <2e-16 ***
## rhoC        2.999e-05 1.826e-05  1.643   0.103    
## tmrcaC      8.940e-04 1.825e-05  48.993 <2e-16 ***
## thetaC:tmrcaC 4.364e-04 1.763e-05  24.752 <2e-16 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.0001695 on 113 degrees of freedom
## Multiple R-squared:  0.9941, Adjusted R-squared:  0.9939 
## F-statistic: 4792 on 4 and 113 DF,  p-value: < 2.2e-16

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

r2.bgs.200kb[1, 3] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
```

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

5.3.4 Replicate 4

```
rep4.sim.tmrca.200kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep4_w200000.csv", sep = ",", header = T)

rep_4.pi.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep4/dm2L_bgs_rep4_5.d")
rep_4.pi.200kb$avg <- apply(rep_4.pi.200kb[4:ncol(rep_4.pi.200kb)], 1, mean)
rep_4.rho.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep4/dm2L_bgs_rep4_5.d")
rep_4.theta.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep4/dm2L_bgs_rep4_5.d")
rep_4.tmrca.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep4/dm2L_bgs_rep4_5.d")
rep_4.tmrca.200kb$avg <- apply(rep_4.tmrca.200kb[4:ncol(rep_4.tmrca.200kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.200kb.rep4 <- as.data.frame(cbind(rep_1.pi.200kb$avg, sim.theta.200kb$Rate, rep4.sim.tmrca.200kb$AvgTmrca))
names(true.lands.200kb.rep4) <- c("diversity", "theta", "tmrca", "rho")
true.lands.200kb.rep4$thetaC <- (true.lands.200kb.rep4$theta - mean(true.lands.200kb.rep4$theta)) / sd(true.lands.200kb.rep4$theta)
true.lands.200kb.rep4$tmrcaC <- (true.lands.200kb.rep4$tmrca - mean(true.lands.200kb.rep4$tmrca)) / sd(true.lands.200kb.rep4$tmrca)
true.lands.200kb.rep4$rhoC <- (true.lands.200kb.rep4$rho - mean(true.lands.200kb.rep4$rho), na.rm = T))

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.200kb.rep4)
anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.200kb.true[1, 4] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4])
r2.bgs.200kb.true[2, 4] <- anova.diversity$VarExp[1] * 100
r2.bgs.200kb.true[3, 4] <- anova.diversity$VarExp[2] * 100
r2.bgs.200kb.true[4, 4] <- anova.diversity$VarExp[3] * 100
r2.bgs.200kb.true[5, 4] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.200kb.rep4 <- as.data.frame(cbind(rep_4.pi.200kb$avg, rep_4.theta.200kb$sample_mean, rep_4.tmrca.200kb$AvgTmrca))
names(inf.lands.200kb.rep4) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[4, 2] <- cor.test(sim.theta.200kb$Rate, rep_4.theta.200kb$sample_mean, method = "spearman")
cor.table.rho[4, 2] <- cor.test(sim.rho.200kb$Rate, rep_4.rho.200kb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[4, 2] <- cor.test(rep4.sim.tmrca.200kb$AverageTmrca, rep_4.tmrca.200kb$avg, method = "spearman")$estimate

# standardizing
inf.lands.200kb.rep4$thetaC <- (inf.lands.200kb.rep4$theta - mean(inf.lands.200kb.rep4$theta)) / sd(inf.lands.200kb.rep4$theta)
inf.lands.200kb.rep4$tmrcaC <- (inf.lands.200kb.rep4$tmrca - mean(inf.lands.200kb.rep4$tmrca)) / sd(inf.lands.200kb.rep4$tmrca)
inf.lands.200kb.rep4$rhoC <- (inf.lands.200kb.rep4$rho - mean(inf.lands.200kb.rep4$rho)) / sd(inf.lands.200kb.rep4$rho)

inf.lands.200kb.rep4$bin <- 1:nrow(inf.lands.200kb.rep4)

inf.lands.200kb.rep4$Replicate <- 4

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.200kb.rep4)
m.div.200kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.200kb.rep4)
```

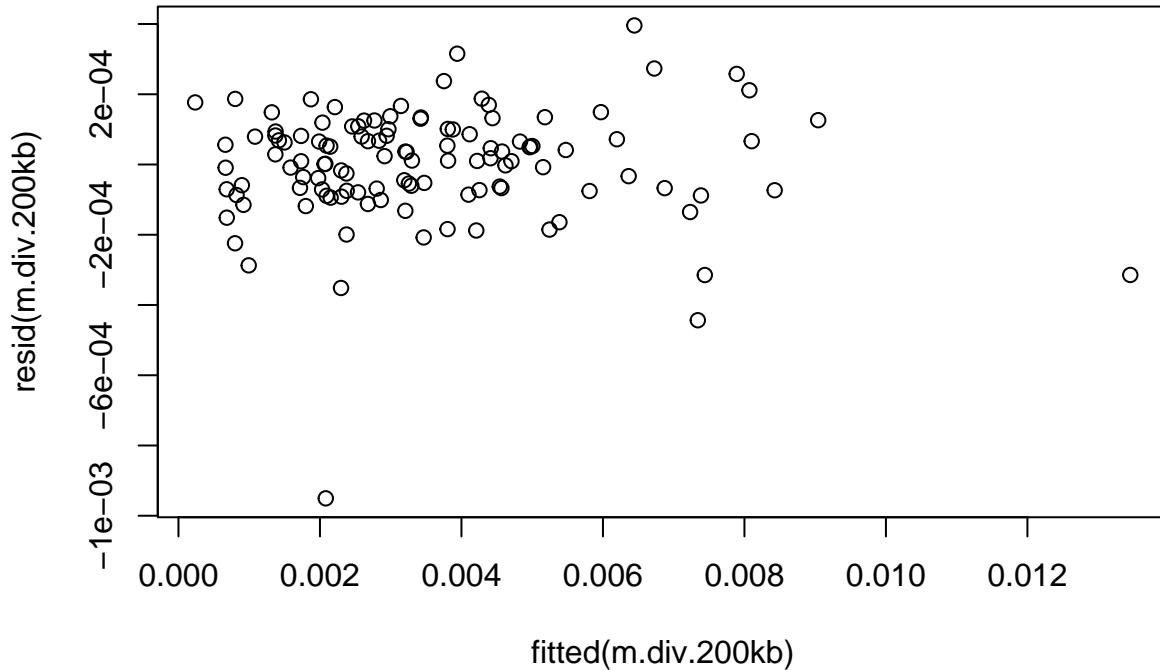
```

m.div.200kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.200kb.rep4)

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

##          df      AIC
## m.div.200kb     6 -1713.378
## m.div.200kb.2   7 -1712.955
## m.div.200kb.3   8 -1711.494
plot(resid(m.div.200kb)~fitted(m.div.200kb))

```



```

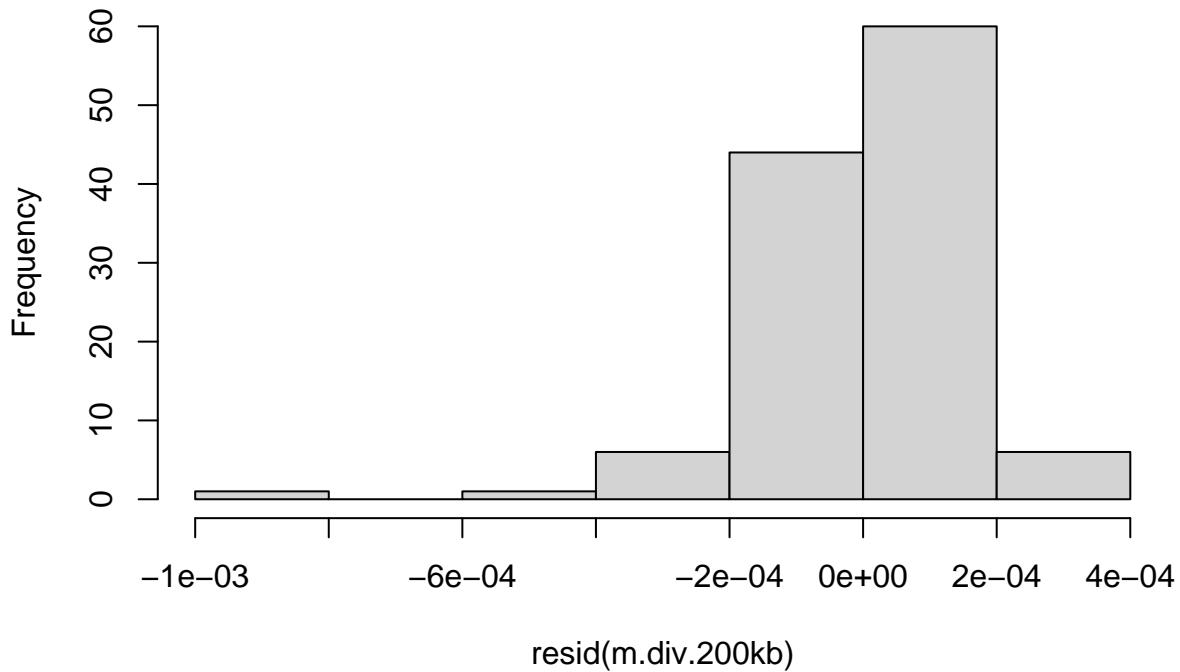
dwtest(m.div.200kb)

##
##  Durbin-Watson test
##
## data: m.div.200kb
## DW = 1.3999, p-value = 0.0002898
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.200kb)

##
##  Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.54457, p-value = 0.752
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 = inf.lands.200kb.rep4)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max  
## -9.505e-04 -7.323e-05  1.451e-05  9.198e-05  3.959e-04  
##  
## Coefficients:  
##             Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 0.0034540  0.0000156 221.409 < 2e-16 ***  
## thetaC      1.0510911  0.0098390 106.829 < 2e-16 ***  
## rhoC        0.0457967  0.0086872  5.272 6.56e-07 ***  
## tmrcaC      0.0050692  0.0001243  40.793 < 2e-16 ***  
## thetaC:tmrcaC 1.4832766  0.0629216 23.573 < 2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.0001652 on 113 degrees of freedom  
## Multiple R-squared:  0.9944, Adjusted R-squared:  0.9942  
## F-statistic: 5034 on 4 and 113 DF,  p-value: < 2.2e-16
```

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

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

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

5.3.5 Replicate 5

```
rep5.sim.tmrca.200kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep5_w200000.csv", sep = ",", header = T)

rep_5.pi.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep5/dm2L_bgs_rep5_5.d")
rep_5.pi.200kb$avg <- apply(rep_5.pi.200kb[4:ncol(rep_5.pi.200kb)], 1, mean)
rep_5.rho.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep5/dm2L_bgs_rep5_5.d")
rep_5.theta.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep5/dm2L_bgs_rep5_5.d")
rep_5.tmrca.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep5/dm2L_bgs_rep5_5.d")
rep_5.tmrca.200kb$avg <- apply(rep_5.tmrca.200kb[4:ncol(rep_5.tmrca.200kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.200kb.rep5 <- as.data.frame(cbind(rep_1.pi.200kb$avg, sim.theta.200kb$Rate, rep5.sim.tmrca.200kb$AvgTmrca))
names(true.lands.200kb.rep5) <- c("diversity", "theta", "tmrca", "rho")
true.lands.200kb.rep5$thetaC <- (true.lands.200kb.rep5$theta - mean(true.lands.200kb.rep5$theta)) / sd(true.lands.200kb.rep5$theta)
true.lands.200kb.rep5$tmrcaC <- (true.lands.200kb.rep5$tmrca - mean(true.lands.200kb.rep5$tmrca)) / sd(true.lands.200kb.rep5$tmrca)
true.lands.200kb.rep5$rhoC <- (true.lands.200kb.rep5$rho - mean(true.lands.200kb.rep5$rho), na.rm = T))

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.200kb.rep5)
anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.200kb.true[1, 5] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) * 100
r2.bgs.200kb.true[2, 5] <- anova.diversity$VarExp[1] * 100
r2.bgs.200kb.true[3, 5] <- anova.diversity$VarExp[2] * 100
r2.bgs.200kb.true[4, 5] <- anova.diversity$VarExp[3] * 100
r2.bgs.200kb.true[5, 5] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.200kb.rep5 <- as.data.frame(cbind(rep_5.pi.200kb$avg, rep_5.theta.200kb$sample_mean, rep_5.tmrca.200kb$AvgTmrca))
names(inf.lands.200kb.rep5) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[5, 2] <- cor.test(sim.theta.200kb$Rate, rep_5.theta.200kb$sample_mean, method = "spearman")
cor.table.rho[5, 2] <- cor.test(sim.rho.200kb$Rate, rep_5.rho.200kb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[5, 2] <- cor.test(rep5.sim.tmrca.200kb$AverageTmrca, rep_5.tmrca.200kb$avg, method = "spearman")$estimate

# standardizing
inf.lands.200kb.rep5$thetaC <- (inf.lands.200kb.rep5$theta - mean(inf.lands.200kb.rep5$theta)) / sd(inf.lands.200kb.rep5$theta)
inf.lands.200kb.rep5$tmrcaC <- (inf.lands.200kb.rep5$tmrca - mean(inf.lands.200kb.rep5$tmrca)) / sd(inf.lands.200kb.rep5$tmrca)
inf.lands.200kb.rep5$rhoC <- (inf.lands.200kb.rep5$rho - mean(inf.lands.200kb.rep5$rho)) / sd(inf.lands.200kb.rep5$rho)

inf.lands.200kb.rep5$bin <- 1:nrow(inf.lands.200kb.rep5)

inf.lands.200kb.rep5$Replicate <- 5

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.200kb.rep5)
m.div.200kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.200kb.rep5)
```

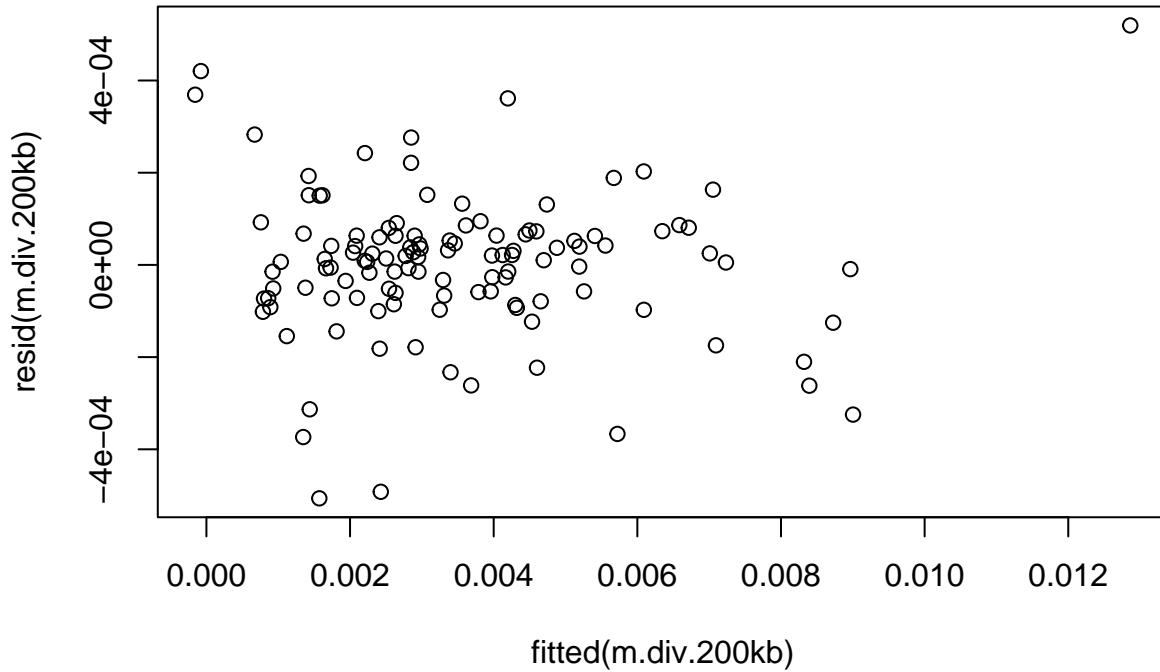
```

m.div.200kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.200kb.rep5)

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

##          df      AIC
## m.div.200kb     6 -1719.223
## m.div.200kb.2   7 -1752.807
## m.div.200kb.3   8 -1753.702
plot(resid(m.div.200kb)~fitted(m.div.200kb))

```



```

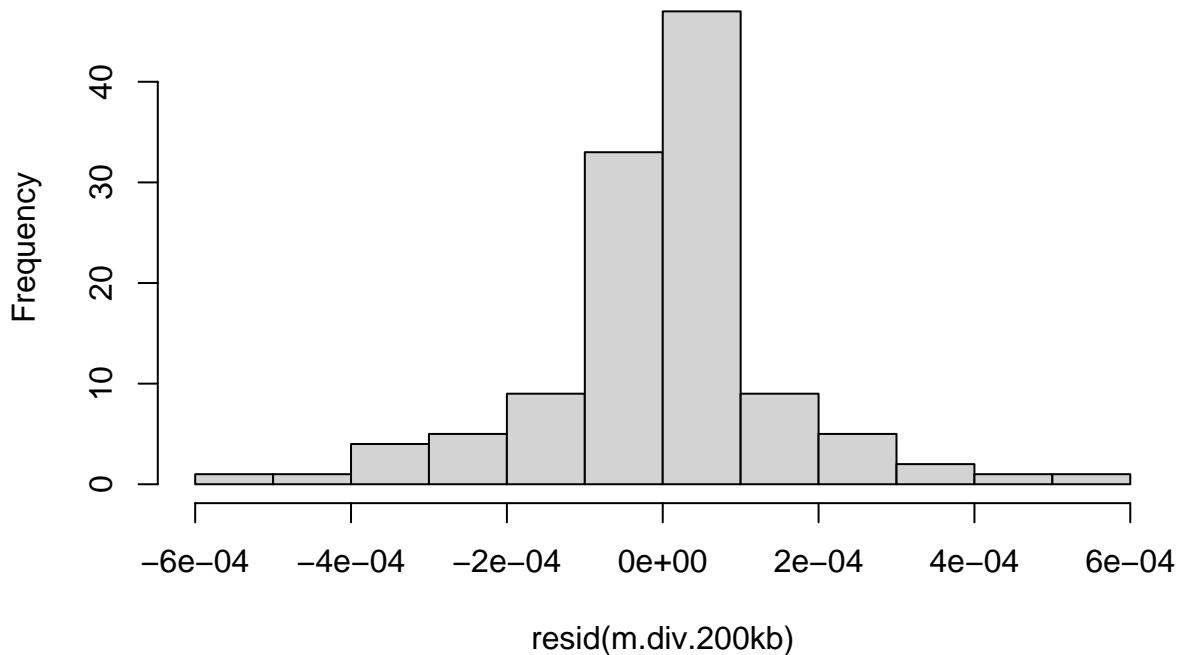
dwtest(m.div.200kb)

##
##  Durbin-Watson test
##
## data: m.div.200kb
## DW = 1.3237, p-value = 6.006e-05
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.200kb)

##
##  Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.38497, p-value = 0.035
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 = inf.lands.200kb.rep5)
##
## Residuals:
##       Min         1Q     Median        3Q       Max
## -5.064e-04 -7.206e-05  1.342e-05  6.558e-05  5.195e-04
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.461e-03  1.497e-05 231.248 <2e-16 ***
## thetaC      1.746e-03  1.628e-05 107.250 <2e-16 ***
## rhoC        1.074e-05  1.741e-05   0.617   0.539
## tmrcaC      8.733e-04  1.645e-05  53.084 <2e-16 ***
## thetaC:tmrcaC 4.289e-04  1.624e-05  26.410 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001612 on 113 degrees of freedom
## Multiple R-squared:  0.9947, Adjusted R-squared:  0.9945
## F-statistic:  5272 on 4 and 113 DF,  p-value: < 2.2e-16

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

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

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

5.3.6 Replicate 6

```
rep6.sim.tmrca.200kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep6_w200000.csv", sep = ",", header = T)

rep_6.pi.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep6/dm2L_bgs_rep6_5.d")
rep_6.pi.200kb$avg <- apply(rep_6.pi.200kb[4:ncol(rep_6.pi.200kb)], 1, mean)
rep_6.rho.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep6/dm2L_bgs_rep6_5.d")
rep_6.theta.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep6/dm2L_bgs_rep6_5.d")
rep_6.tmrca.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep6/dm2L_bgs_rep6_5.d")
rep_6.tmrca.200kb$avg <- apply(rep_6.tmrca.200kb[4:ncol(rep_6.tmrca.200kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.200kb.rep6 <- as.data.frame(cbind(rep_1.pi.200kb$avg, sim.theta.200kb$Rate, rep6.sim.tmrca.200kb$AvgTmrca))
names(true.lands.200kb.rep6) <- c("diversity", "theta", "tmrca", "rho")
true.lands.200kb.rep6$thetaC <- (true.lands.200kb.rep6$theta - mean(true.lands.200kb.rep6$theta)) / sd(true.lands.200kb.rep6$theta)
true.lands.200kb.rep6$tmrcaC <- (true.lands.200kb.rep6$tmrca - mean(true.lands.200kb.rep6$tmrca)) / sd(true.lands.200kb.rep6$tmrca)
true.lands.200kb.rep6$rhoC <- (true.lands.200kb.rep6$rho - mean(true.lands.200kb.rep6$rho), na.rm = T))

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.200kb.rep6)
anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.200kb.true[1, 6] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4])
r2.bgs.200kb.true[2, 6] <- anova.diversity$VarExp[1] * 100
r2.bgs.200kb.true[3, 6] <- anova.diversity$VarExp[2] * 100
r2.bgs.200kb.true[4, 6] <- anova.diversity$VarExp[3] * 100
r2.bgs.200kb.true[5, 6] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.200kb.rep6 <- as.data.frame(cbind(rep_6.pi.200kb$avg, rep_6.theta.200kb$sample_mean, rep_6.tmrca.200kb$AvgTmrca))
names(inf.lands.200kb.rep6) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[6, 2] <- cor.test(sim.theta.200kb$Rate, rep_6.theta.200kb$sample_mean, method = "spearman")
cor.table.rho[6, 2] <- cor.test(sim.rho.200kb$Rate, rep_6.rho.200kb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[6, 2] <- cor.test(rep6.sim.tmrca.200kb$AverageTmrca, rep_6.tmrca.200kb$avg, method = "spearman")$estimate

# standardizing
inf.lands.200kb.rep6$thetaC <- (inf.lands.200kb.rep6$theta - mean(inf.lands.200kb.rep6$theta)) / sd(inf.lands.200kb.rep6$theta)
inf.lands.200kb.rep6$tmrcaC <- (inf.lands.200kb.rep6$tmrca - mean(inf.lands.200kb.rep6$tmrca)) / sd(inf.lands.200kb.rep6$tmrca)
inf.lands.200kb.rep6$rhoC <- (inf.lands.200kb.rep6$rho - mean(inf.lands.200kb.rep6$rho)) / sd(inf.lands.200kb.rep6$rho)

inf.lands.200kb.rep6$bin <- 1:nrow(inf.lands.200kb.rep6)

inf.lands.200kb.rep6$Replicate <- 6

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.200kb.rep6)
m.div.200kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.200kb.rep6)
```

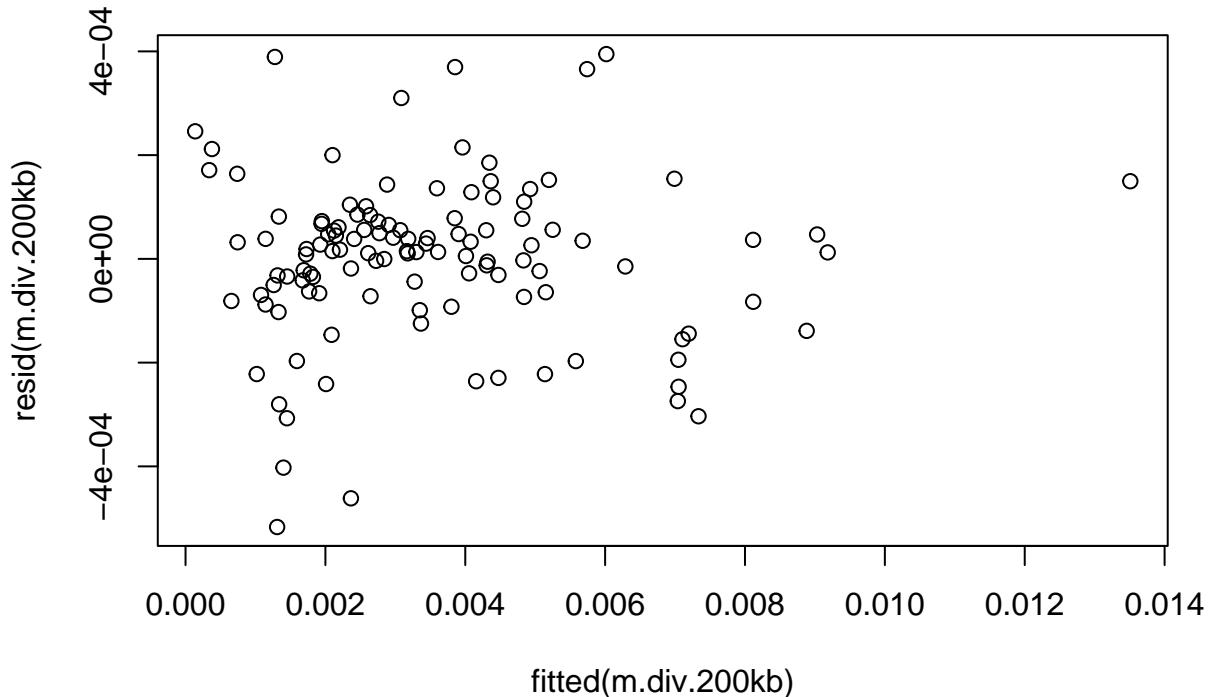
```

m.div.200kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.200kb.rep6)

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

##          df      AIC
## m.div.200kb     6 -1721.284
## m.div.200kb.2   7 -1738.537
## m.div.200kb.3   8 -1737.848
plot(resid(m.div.200kb)~fitted(m.div.200kb))

```



```

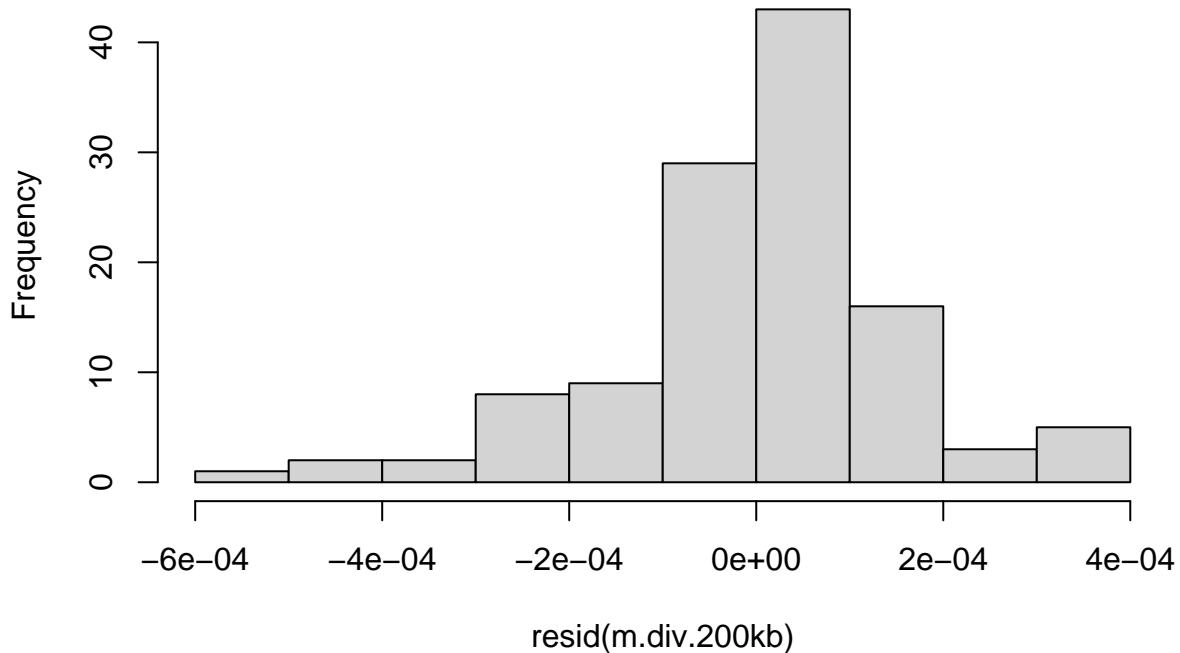
dwtest(m.div.200kb)

##
##  Durbin-Watson test
##
## data: m.div.200kb
## DW = 1.1428, p-value = 6.441e-07
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.200kb)

##
##  Harrison-M McCabe test
##
## data: m.div.200kb
## HMC = 0.37066, p-value = 0.024
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 = inf.lands.200kb.rep6)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -5.169e-04 -6.868e-05  1.496e-05  7.251e-05  3.947e-04
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.420e-03 1.512e-05 226.13 <2e-16 ***
## thetaC      1.761e-03 1.631e-05 107.98 <2e-16 ***
## rhoC        3.867e-05 1.646e-05   2.35  0.0205 *  
## tmrcaC      8.002e-04 1.654e-05   48.37 <2e-16 ***
## thetaC:tmrcaC 4.255e-04 1.545e-05   27.55 <2e-16 ***
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.0001598 on 113 degrees of freedom
## Multiple R-squared:  0.995, Adjusted R-squared:  0.9949 
## F-statistic: 5662 on 4 and 113 DF,  p-value: < 2.2e-16
anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
```

```
r2.bgs.200kb[1, 6] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
```

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

5.3.7 Replicate 7

```
rep7.sim.tmrca.200kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep7_w200000.csv", sep = ",", header = T)

rep_7.pi.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep7/dm2L_bgs_rep7_5.d")
rep_7.pi.200kb$avg <- apply(rep_7.pi.200kb[4:ncol(rep_7.pi.200kb)], 1, mean)
rep_7.rho.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep7/dm2L_bgs_rep7_5.d")
rep_7.theta.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep7/dm2L_bgs_rep7_5.d")
rep_7.tmrca.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep7/dm2L_bgs_rep7_5.d")
rep_7.tmrca.200kb$avg <- apply(rep_7.tmrca.200kb[4:ncol(rep_7.tmrca.200kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.200kb.rep7 <- as.data.frame(cbind(rep_1.pi.200kb$avg, sim.theta.200kb$Rate, rep7.sim.tmrca.200kb$AvgTmrca))
names(true.lands.200kb.rep7) <- c("diversity", "theta", "tmrca", "rho")
true.lands.200kb.rep7$thetaC <- (true.lands.200kb.rep7$theta - mean(true.lands.200kb.rep7$theta)) / sd(true.lands.200kb.rep7$theta)
true.lands.200kb.rep7$tmrcaC <- (true.lands.200kb.rep7$tmrca - mean(true.lands.200kb.rep7$tmrca)) / sd(true.lands.200kb.rep7$tmrca)
true.lands.200kb.rep7$rhoC <- (true.lands.200kb.rep7$rho - mean(true.lands.200kb.rep7$rho), na.rm = T))

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.200kb.rep7)
anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.200kb.true[1, 7] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4])
r2.bgs.200kb.true[2, 7] <- anova.diversity$VarExp[1] * 100
r2.bgs.200kb.true[3, 7] <- anova.diversity$VarExp[2] * 100
r2.bgs.200kb.true[4, 7] <- anova.diversity$VarExp[3] * 100
r2.bgs.200kb.true[5, 7] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.200kb.rep7 <- as.data.frame(cbind(rep_7.pi.200kb$avg, rep_7.theta.200kb$sample_mean, rep_7.tmrca.200kb$AvgTmrca))
names(inf.lands.200kb.rep7) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[7, 2] <- cor.test(sim.theta.200kb$Rate, rep_7.theta.200kb$sample_mean, method = "spearman")
cor.table.rho[7, 2] <- cor.test(sim.rho.200kb$Rate, rep_1.rho.200kb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[7, 2] <- cor.test(rep7.sim.tmrca.200kb$AverageTmrca, rep_7.tmrca.200kb$avg, method = "spearman")$estimate

# standardizing
inf.lands.200kb.rep7$thetaC <- (inf.lands.200kb.rep7$theta - mean(inf.lands.200kb.rep7$theta)) / sd(inf.lands.200kb.rep7$theta)
inf.lands.200kb.rep7$tmrcaC <- (inf.lands.200kb.rep7$tmrca - mean(inf.lands.200kb.rep7$tmrca)) / sd(inf.lands.200kb.rep7$tmrca)
inf.lands.200kb.rep7$rhoC <- (inf.lands.200kb.rep7$rho - mean(inf.lands.200kb.rep7$rho)) / sd(inf.lands.200kb.rep7$rho)

inf.lands.200kb.rep7$bin <- 1:nrow(inf.lands.200kb.rep7)

inf.lands.200kb.rep7$Replicate <- 7

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.200kb.rep7)
m.div.200kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.200kb.rep7)
```

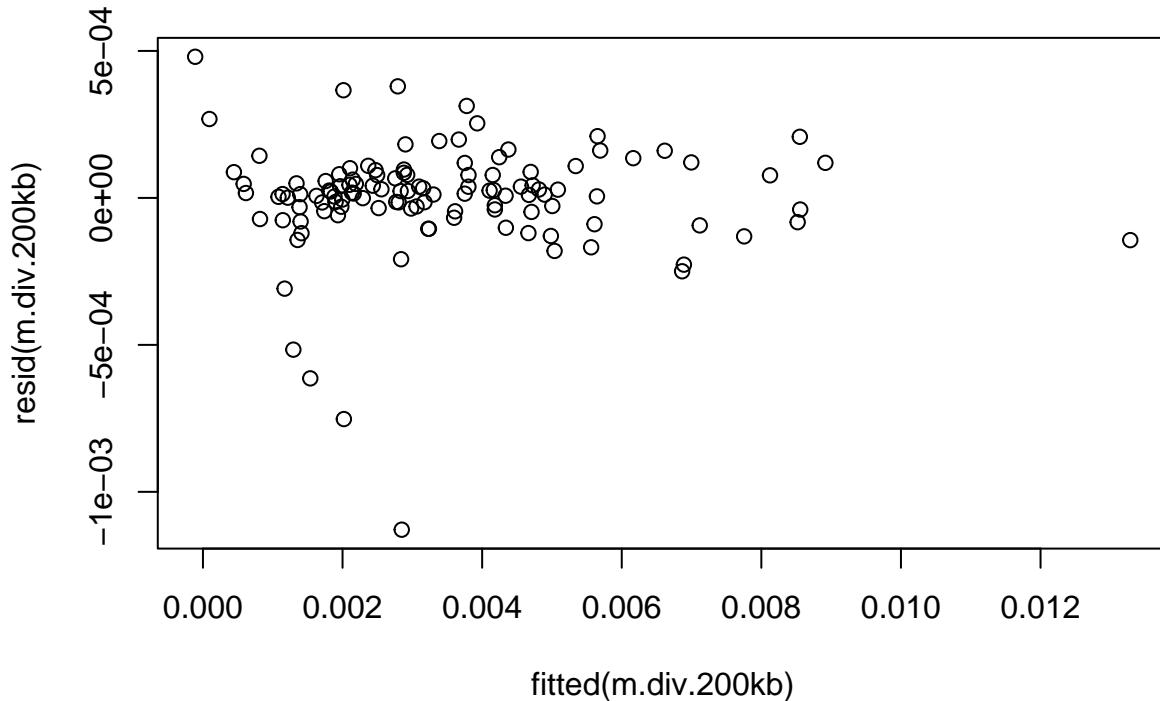
```

m.div.200kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.200kb.rep7)

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

##          df      AIC
## m.div.200kb     6 -1675.014
## m.div.200kb.2   7 -1686.944
## m.div.200kb.3   8 -1685.792
plot(resid(m.div.200kb)~fitted(m.div.200kb))

```



```

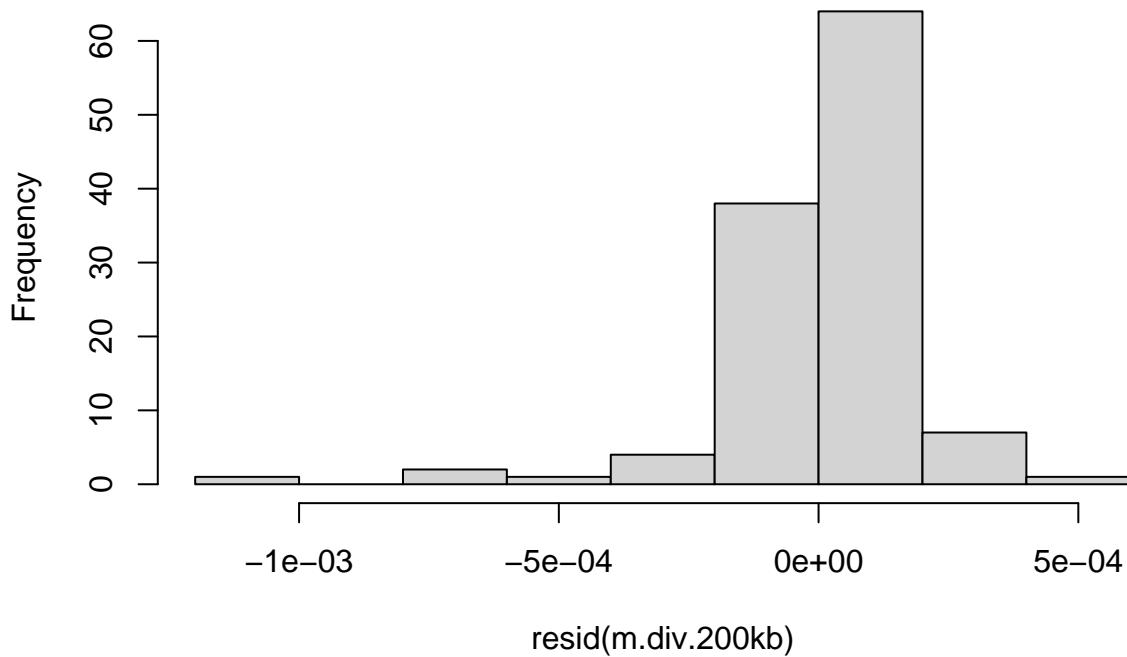
dwtest(m.div.200kb)

##
##  Durbin-Watson test
##
## data: m.div.200kb
## DW = 1.4537, p-value = 0.0008495
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.200kb)

##
##  Harrison-M McCabe test
##
## data: m.div.200kb
## HMC = 0.2727, p-value < 2.2e-16
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 = inf.lands.200kb.rep7)
##
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -1.129e-03 -4.377e-05  1.575e-05  7.812e-05  4.801e-04
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.341e-03  1.908e-05 175.137 <2e-16 ***
## thetaC      1.759e-03  2.163e-05  81.322 <2e-16 ***
## rhoC        5.173e-05  2.089e-05   2.477  0.0147 *  
## tmrcaC      7.373e-04  2.268e-05  32.515 <2e-16 ***
## thetaC:tmrcaC 4.077e-04  2.234e-05  18.247 <2e-16 ***
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.0001944 on 113 degrees of freedom
## Multiple R-squared:  0.9923, Adjusted R-squared:  0.9921 
## F-statistic: 3652 on 4 and 113 DF,  p-value: < 2.2e-16

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

r2.bgs.200kb[1, 7] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
```

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

5.3.8 Replicate 8

```
rep8.sim.tmrca.200kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep8_w200000.csv", sep = ",", header = T)

rep_8.pi.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep8/dm2L_bgs_rep8_5.d")
rep_8.pi.200kb$avg <- apply(rep_8.pi.200kb[4:ncol(rep_8.pi.200kb)], 1, mean)
rep_8.rho.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep8/dm2L_bgs_rep8_5.d")
rep_8.theta.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep8/dm2L_bgs_rep8_5.d")
rep_8.tmrca.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep8/dm2L_bgs_rep8_5.d")
rep_8.tmrca.200kb$avg <- apply(rep_8.tmrca.200kb[4:ncol(rep_8.tmrca.200kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.200kb.rep8 <- as.data.frame(cbind(rep_1.pi.200kb$avg, sim.theta.200kb$Rate, rep8.sim.tmrca.200kb$AvgTmrca))
names(true.lands.200kb.rep8) <- c("diversity", "theta", "tmrca", "rho")
true.lands.200kb.rep8$thetaC <- (true.lands.200kb.rep8$theta - mean(true.lands.200kb.rep8$theta)) / sd(true.lands.200kb.rep8$theta)
true.lands.200kb.rep8$tmrcaC <- (true.lands.200kb.rep8$tmrca - mean(true.lands.200kb.rep8$tmrca)) / sd(true.lands.200kb.rep8$tmrca)
true.lands.200kb.rep8$rhoC <- (true.lands.200kb.rep8$rho - mean(true.lands.200kb.rep8$rho), na.rm = T))

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.200kb.rep8)
anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.200kb.true[1, 8] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) * 100
r2.bgs.200kb.true[2, 8] <- anova.diversity$VarExp[1] * 100
r2.bgs.200kb.true[3, 8] <- anova.diversity$VarExp[2] * 100
r2.bgs.200kb.true[4, 8] <- anova.diversity$VarExp[3] * 100
r2.bgs.200kb.true[5, 8] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.200kb.rep8 <- as.data.frame(cbind(rep_8.pi.200kb$avg, rep_8.theta.200kb$sample_mean, rep_8.tmrca.200kb$AvgTmrca))
names(inf.lands.200kb.rep8) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[8, 2] <- cor.test(sim.theta.200kb$Rate, rep_8.theta.200kb$sample_mean, method = "spearman")
cor.table.rho[8, 2] <- cor.test(sim.rho.200kb$Rate, rep_8.rho.200kb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[8, 2] <- cor.test(rep8.sim.tmrca.200kb$AverageTmrca, rep_8.tmrca.200kb$avg, method = "spearman")$estimate

# standardizing
inf.lands.200kb.rep8$thetaC <- (inf.lands.200kb.rep8$theta - mean(inf.lands.200kb.rep8$theta)) / sd(inf.lands.200kb.rep8$theta)
inf.lands.200kb.rep8$tmrcaC <- (inf.lands.200kb.rep8$tmrca - mean(inf.lands.200kb.rep8$tmrca)) / sd(inf.lands.200kb.rep8$tmrca)
inf.lands.200kb.rep8$rhoC <- (inf.lands.200kb.rep8$rho - mean(inf.lands.200kb.rep8$rho)) / sd(inf.lands.200kb.rep8$rho)

inf.lands.200kb.rep8$bin <- 1:nrow(inf.lands.200kb.rep8)

inf.lands.200kb.rep8$Replicate <- 8

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.200kb.rep8)
m.div.200kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.200kb.rep8)
```

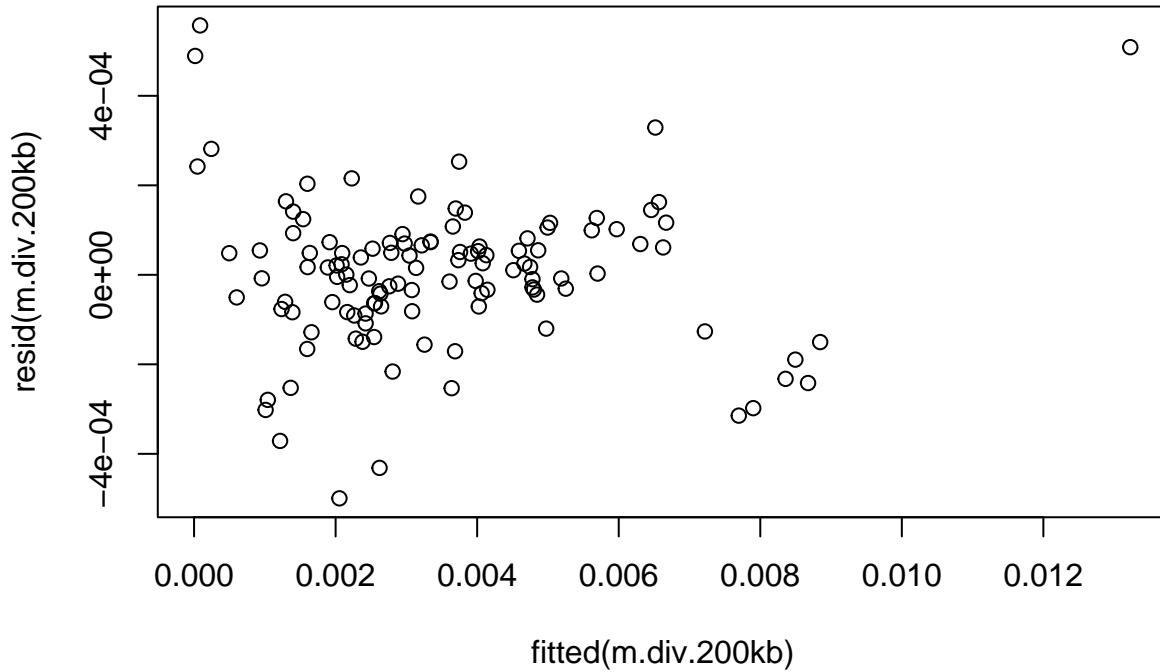
```

m.div.200kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.200kb.rep8)

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

##           df      AIC
## m.div.200kb     6 -1711.201
## m.div.200kb.2   7 -1752.777
## m.div.200kb.3   8 -1754.136
plot(resid(m.div.200kb)~fitted(m.div.200kb))

```



```

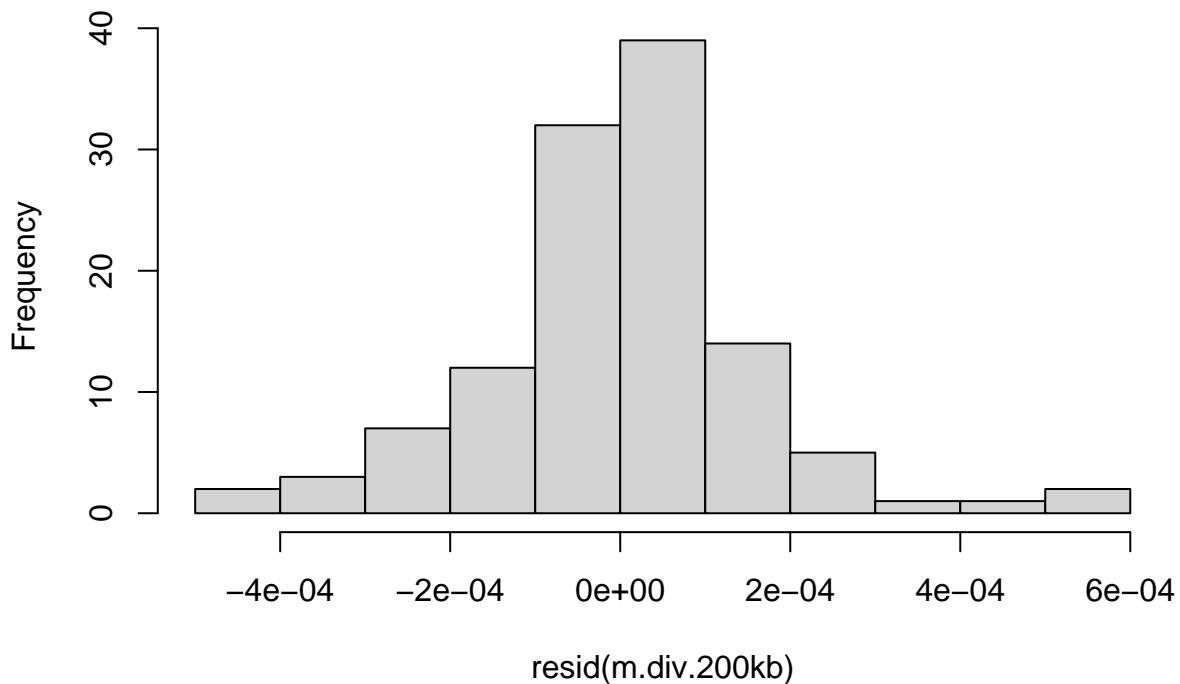
dwtest(m.div.200kb)

##
##  Durbin-Watson test
##
## data: m.div.200kb
## DW = 1.4172, p-value = 0.0004208
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.200kb)

##
##  Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.36722, p-value = 0.021
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 = inf.lands.200kb.rep8)
##
## Residuals:
##       Min         1Q     Median         3Q        Max
## -4.995e-04 -7.501e-05  1.296e-05  7.253e-05  5.573e-04
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.435e-03 1.548e-05 221.813 <2e-16 ***
## thetaC      1.640e-03 1.724e-05  95.133 <2e-16 ***
## rhoC        1.806e-05 1.814e-05   0.996   0.322
## tmrcaC      9.569e-04 1.745e-05  54.824 <2e-16 ***
## thetaC:tmrcaC 4.956e-04 1.962e-05  25.255 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001668 on 113 degrees of freedom
## Multiple R-squared:  0.9943, Adjusted R-squared:  0.9941
## F-statistic:  4932 on 4 and 113 DF,  p-value: < 2.2e-16

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

r2.bgs.200kb[1, 8] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
```

```
r2.bgs.200kb[2, 8] <- anova.diversity$VarExp[1] * 100
r2.bgs.200kb[3, 8] <- anova.diversity$VarExp[2] * 100
r2.bgs.200kb[4, 8] <- anova.diversity$VarExp[3] * 100
r2.bgs.200kb[5, 8] <- anova.diversity$VarExp[4] * 100
```

5.3.9 Replicate 9

```
rep9.sim.tmrca.200kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep9_w200000.csv", sep = ",", header = T)

rep_9.pi.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep9/dm2L_bgs_rep9_5.d")
rep_9.pi.200kb$avg <- apply(rep_9.pi.200kb[4:ncol(rep_9.pi.200kb)], 1, mean)
rep_9.rho.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep9/dm2L_bgs_rep9_5.d")
rep_9.theta.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep9/dm2L_bgs_rep9_5.d")
rep_9.tmrca.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep9/dm2L_bgs_rep9_5.d")
rep_9.tmrca.200kb$avg <- apply(rep_9.tmrca.200kb[4:ncol(rep_9.tmrca.200kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.200kb.rep9 <- as.data.frame(cbind(rep_1.pi.200kb$avg, sim.theta.200kb$Rate, rep9.sim.tmrca.200kb$AvgTmrca))
names(true.lands.200kb.rep9) <- c("diversity", "theta", "tmrca", "rho")
true.lands.200kb.rep9$thetaC <- (true.lands.200kb.rep9$theta - mean(true.lands.200kb.rep9$theta)) / sd(true.lands.200kb.rep9$theta)
true.lands.200kb.rep9$tmrcaC <- (true.lands.200kb.rep9$tmrca - mean(true.lands.200kb.rep9$tmrca)) / sd(true.lands.200kb.rep9$tmrca)
true.lands.200kb.rep9$rhoC <- (true.lands.200kb.rep9$rho - mean(true.lands.200kb.rep9$rho), na.rm = T))

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.200kb.rep9)
anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.200kb.true[1, 9] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) * 100
r2.bgs.200kb.true[2, 9] <- anova.diversity$VarExp[1] * 100
r2.bgs.200kb.true[3, 9] <- anova.diversity$VarExp[2] * 100
r2.bgs.200kb.true[4, 9] <- anova.diversity$VarExp[3] * 100
r2.bgs.200kb.true[5, 9] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.200kb.rep9 <- as.data.frame(cbind(rep_9.pi.200kb$avg, rep_9.theta.200kb$sample_mean, rep_9.tmrca.200kb$AvgTmrca))
names(inf.lands.200kb.rep9) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[9, 2] <- cor.test(sim.theta.200kb$Rate, rep_9.theta.200kb$sample_mean, method = "spearman")
cor.table.rho[9, 2] <- cor.test(sim.rho.200kb$Rate, rep_9.rho.200kb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[9, 2] <- cor.test(rep9.sim.tmrca.200kb$AverageTmrca, rep_9.tmrca.200kb$avg, method = "spearman")$estimate

# standardizing
inf.lands.200kb.rep9$thetaC <- (inf.lands.200kb.rep9$theta - mean(inf.lands.200kb.rep9$theta)) / sd(inf.lands.200kb.rep9$theta)
inf.lands.200kb.rep9$tmrcaC <- (inf.lands.200kb.rep9$tmrca - mean(inf.lands.200kb.rep9$tmrca)) / sd(inf.lands.200kb.rep9$tmrca)
inf.lands.200kb.rep9$rhoC <- (inf.lands.200kb.rep9$rho - mean(inf.lands.200kb.rep9$rho)) / sd(inf.lands.200kb.rep9$rho)

inf.lands.200kb.rep9$bin <- 1:nrow(inf.lands.200kb.rep9)

inf.lands.200kb.rep9$Replicate <- 9

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.200kb.rep9)
m.div.200kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.200kb.rep9)
```

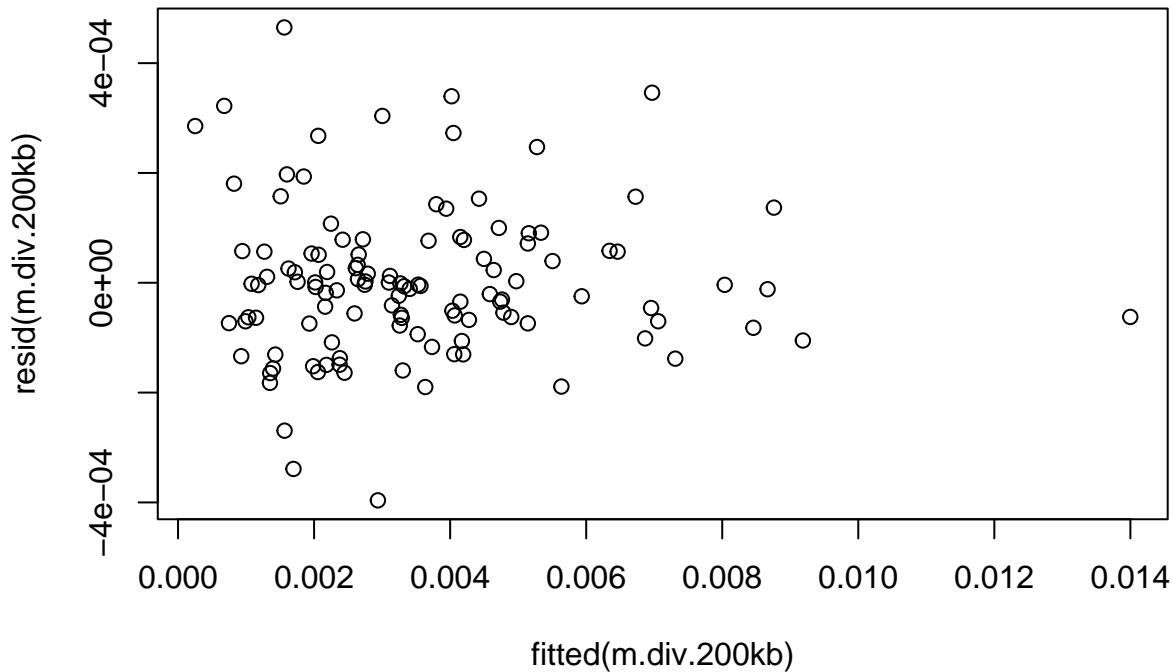
```

m.div.200kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.200kb.rep9)

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

##          df      AIC
## m.div.200kb    6 -1753.774
## m.div.200kb.2  7 -1752.849
## m.div.200kb.3  8 -1750.851
plot(resid(m.div.200kb)~fitted(m.div.200kb))

```



```

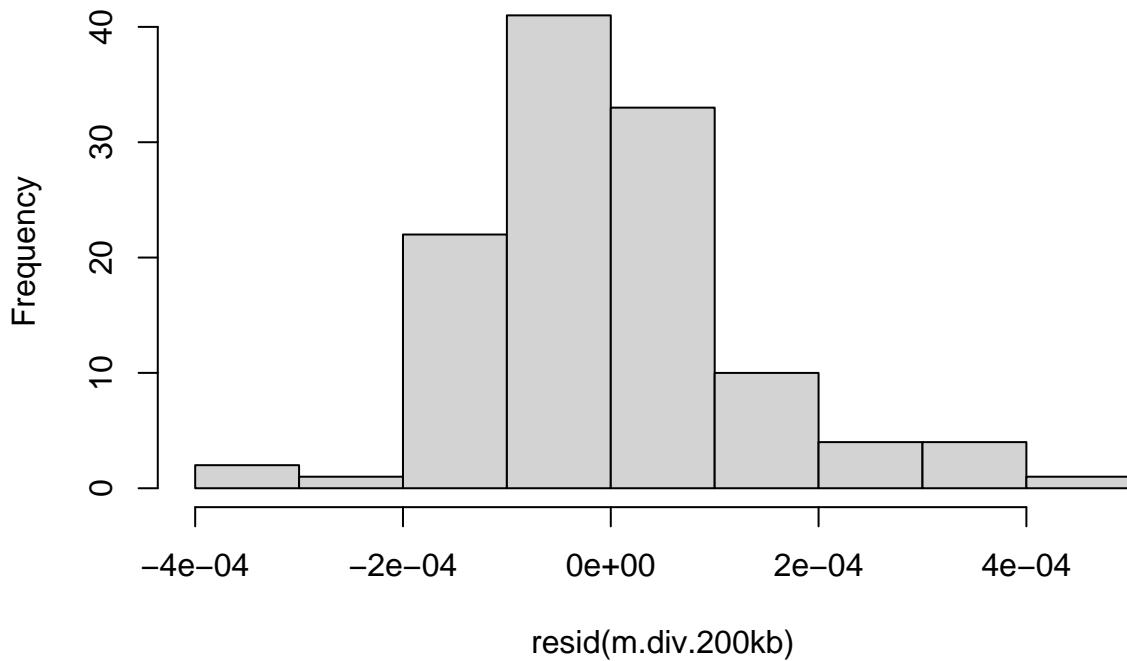
dwtest(m.div.200kb)

##
##  Durbin-Watson test
##
## data: m.div.200kb
## DW = 1.9167, p-value = 0.2774
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.200kb)

##
##  Harrison-McCabe test
##
## data: m.div.200kb
## HMC = 0.31553, p-value = 0.002
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 = inf.lands.200kb.rep9)
##
## Residuals:
##       Min         1Q     Median        3Q       Max
## -3.962e-04 -7.380e-05 -6.080e-06  5.746e-05  4.650e-04
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.408e-03 1.349e-05 252.580 < 2e-16 ***
## thetaC      1.642e-03 1.571e-05 104.514 < 2e-16 ***
## rhoC        4.060e-05 1.548e-05   2.623  0.00991 **
## tmrcaC      8.452e-04 1.551e-05  54.487 < 2e-16 ***
## thetaC:tmrcaC 4.608e-04 1.587e-05  29.035 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001392 on 113 degrees of freedom
## Multiple R-squared:  0.9961, Adjusted R-squared:  0.9959
## F-statistic: 7167 on 4 and 113 DF,  p-value: < 2.2e-16

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

r2.bgs.200kb[1, 9] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
```

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

5.3.10 Replicate 10

```
rep10.sim.tmrca.200kb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep10_w200000.csv", sep = ",", header = T)

rep_10.pi.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep10/dm2L_bgs_rep10")
rep_10.pi.200kb$avg <- apply(rep_10.pi.200kb[4:ncol(rep_10.pi.200kb)], 1, mean)
rep_10.rho.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep10/dm2L_bgs_rep10")
rep_10.theta.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep10/dm2L_bgs_rep10")
rep_10.tmrca.200kb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep10/dm2L_bgs_rep10")
rep_10.tmrca.200kb$avg <- apply(rep_10.tmrca.200kb[4:ncol(rep_10.tmrca.200kb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.200kb.rep10 <- as.data.frame(cbind(rep_1.pi.200kb$avg, sim.theta.200kb$Rate, rep10.sim.tmrca.200kb))
names(true.lands.200kb.rep10) <- c("diversity", "theta", "tmrca", "rho")
true.lands.200kb.rep10$thetaC <- (true.lands.200kb.rep10$theta - mean(true.lands.200kb.rep10$theta)) / sd(true.lands.200kb.rep10$theta)
true.lands.200kb.rep10$tmrcaC <- (true.lands.200kb.rep10$tmrca - mean(true.lands.200kb.rep10$tmrca)) / sd(true.lands.200kb.rep10$tmrca)
true.lands.200kb.rep10$rhoC <- (true.lands.200kb.rep10$rho - mean(true.lands.200kb.rep10$rho), na.rm = T)

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.200kb.rep10)
anova.diversity <- Anova(m.div.200kb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.200kb.true[1, 10] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) * 100
r2.bgs.200kb.true[2, 10] <- anova.diversity$VarExp[1] * 100
r2.bgs.200kb.true[3, 10] <- anova.diversity$VarExp[2] * 100
r2.bgs.200kb.true[4, 10] <- anova.diversity$VarExp[3] * 100
r2.bgs.200kb.true[5, 10] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.200kb.rep10 <- as.data.frame(cbind(rep_10.pi.200kb$avg, rep_10.theta.200kb$sample_mean, rep_10.rho.200kb$sample_mean))
names(inf.lands.200kb.rep10) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[10, 2] <- cor.test(sim.theta.200kb$Rate, rep_10.theta.200kb$sample_mean, method = "spearman")
cor.table.rho[10, 2] <- cor.test(sim.rho.200kb$Rate, rep_10.rho.200kb$sample_mean, method = "spearman")
cor.table.tmrca[10, 2] <- cor.test(rep10.sim.tmrca.200kb$AverageTmrca, rep_10.tmrca.200kb$avg, method = "spearman")

# standardizing
inf.lands.200kb.rep10$thetaC <- (inf.lands.200kb.rep10$theta - mean(inf.lands.200kb.rep10$theta)) / sd(inf.lands.200kb.rep10$theta)
inf.lands.200kb.rep10$tmrcaC <- (inf.lands.200kb.rep10$tmrca - mean(inf.lands.200kb.rep10$tmrca)) / sd(inf.lands.200kb.rep10$tmrca)
inf.lands.200kb.rep10$rhoC <- (inf.lands.200kb.rep10$rho - mean(inf.lands.200kb.rep10$rho)) / sd(inf.lands.200kb.rep10$rho)

inf.lands.200kb.rep10$bin <- 1:nrow(inf.lands.200kb.rep10)

inf.lands.200kb.rep10$Replicate <- 10

m.div.200kb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.200kb.rep10)
m.div.200kb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.200kb.rep10)
```

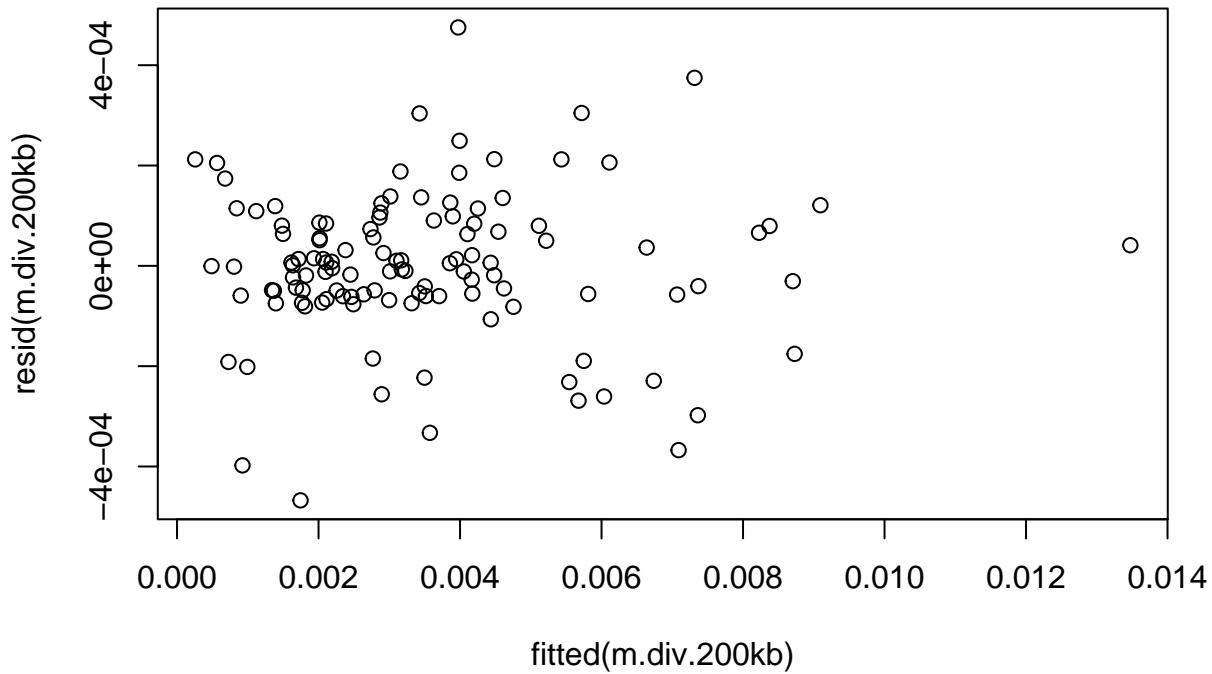
```

m.div.200kb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.200kb.rep10)

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

##          df      AIC
## m.div.200kb    6 -1734.618
## m.div.200kb.2  7 -1735.599
## m.div.200kb.3  8 -1737.127
plot(resid(m.div.200kb)~fitted(m.div.200kb))

```



```

dwtest(m.div.200kb)

##
##  Durbin-Watson test
##
##  data:  m.div.200kb
##  DW = 1.2969, p-value = 3.058e-05
##  alternative hypothesis: true autocorrelation is greater than 0

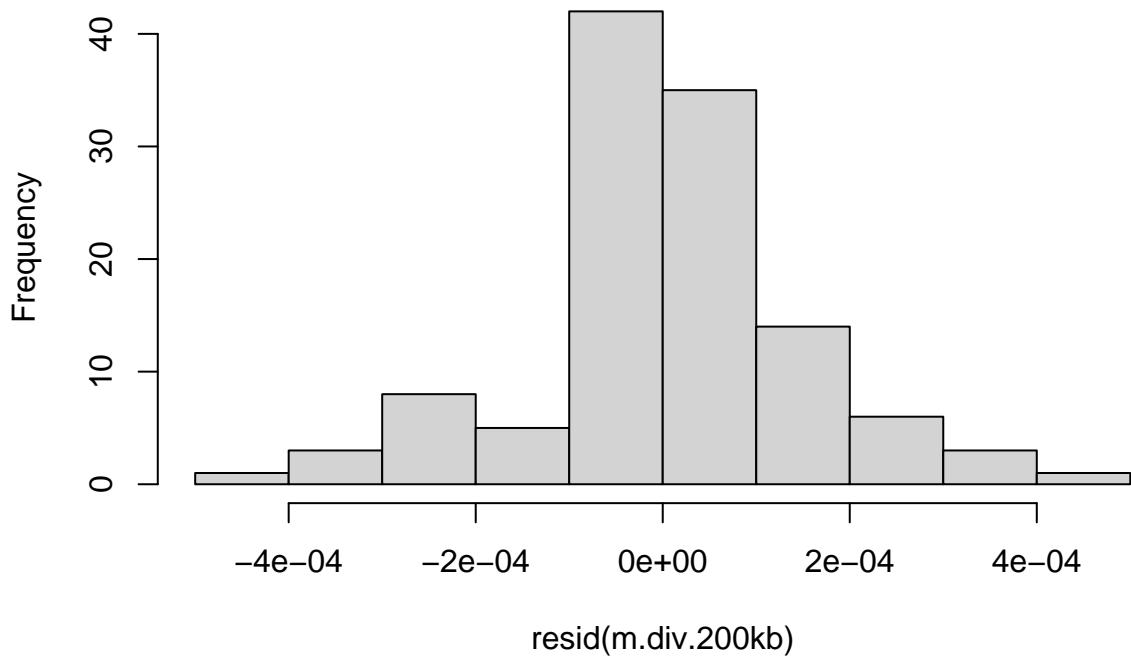
hmctest(m.div.200kb)

##
##  Harrison-McCabe test
##
##  data:  m.div.200kb
##  HMC = 0.39016, p-value = 0.031

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 = inf.lands.200kb.rep10)
##
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -4.675e-04 -5.994e-05  2.700e-07  8.305e-05  4.753e-04
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.461e-03  1.419e-05 244.019 < 2e-16 ***
## thetaC       1.794e-03  1.560e-05 115.002 < 2e-16 ***
## rhoC         5.741e-05  1.781e-05  3.224  0.00165 ** 
## tmrcaC       7.298e-04  1.665e-05 43.821 < 2e-16 ***
## thetaC:tmrcaC 3.619e-04  1.426e-05 25.371 < 2e-16 ***
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.000151 on 113 degrees of freedom
## Multiple R-squared:  0.9954, Adjusted R-squared:  0.9953 
## F-statistic: 6149 on 4 and 113 DF,  p-value: < 2.2e-16

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

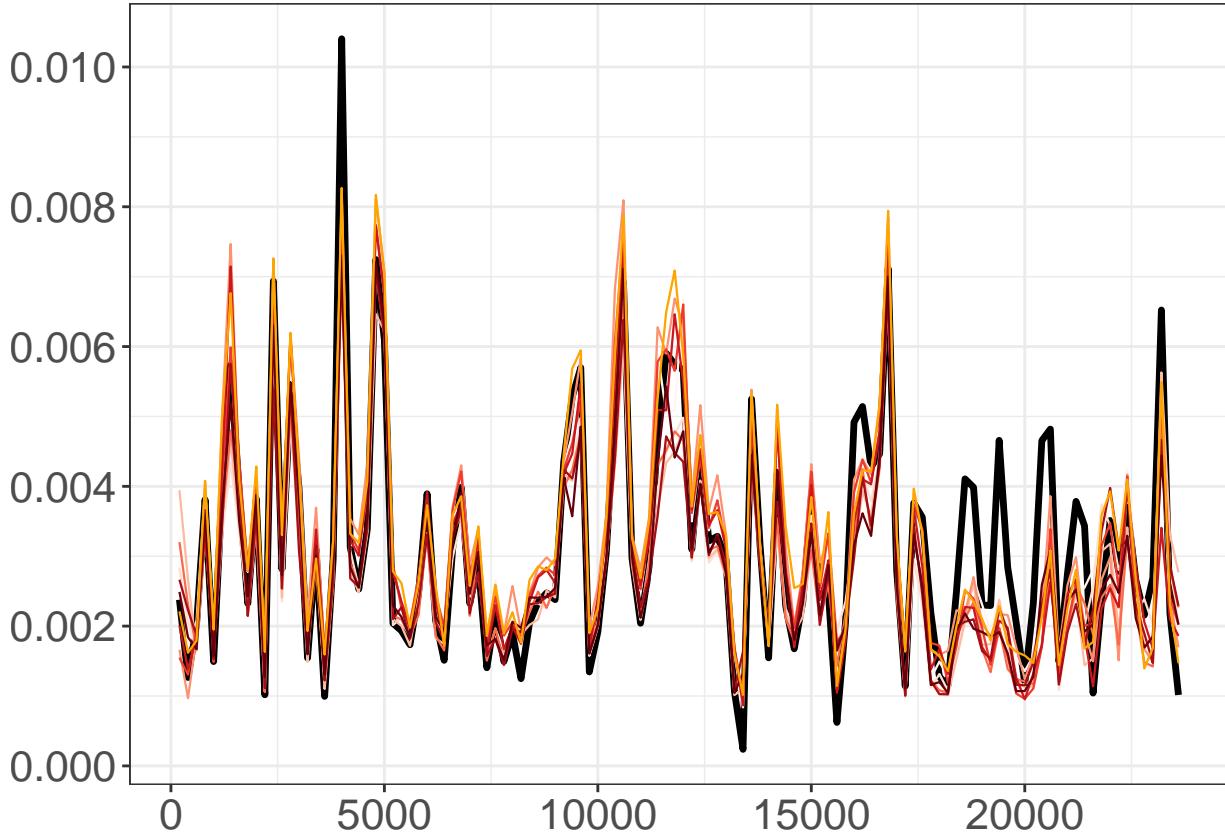
r2.bgs.200kb[1, 10] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
```

```
r2.bgs.200kb[2, 10] <- anova.diversity$VarExp[1] * 100
r2.bgs.200kb[3, 10] <- anova.diversity$VarExp[2] * 100
r2.bgs.200kb[4, 10] <- anova.diversity$VarExp[3] * 100
r2.bgs.200kb[5, 10] <- anova.diversity$VarExp[4] * 100
```

5.4 all replicates:

```
theta.plot <- as.data.frame(cbind(1:nrow(sim.theta.200kb),
                                    sim.theta.200kb$Rate * 1e-7 * 3e+4,
                                    rep_1.theta.200kb$sample_mean,
                                    rep_2.theta.200kb$sample_mean,
                                    rep_3.theta.200kb$sample_mean,
                                    rep_4.theta.200kb$sample_mean,
                                    rep_5.theta.200kb$sample_mean,
                                    rep_6.theta.200kb$sample_mean,
                                    rep_7.theta.200kb$sample_mean,
                                    rep_8.theta.200kb$sample_mean,
                                    rep_9.theta.200kb$sample_mean,
                                    rep_10.theta.200kb$sample_mean))

names(theta.plot) <- c("bin", "sim", "reps")
molten.theta <- melt(theta.plot, id.vars = "bin")
theta.map.200kb.bgs <- ggplot(data = molten.theta, aes(x = bin * 200, y = value, colour = variable)) +
  theta.map.200kb.bgs + geom_line(data = molten.theta, aes(size = variable)) + scale_color_manual(values = c("black", brewer.pal(n = 9, name = "Set1")[1:8], "#000000"))
theta.map.200kb.bgs <- theta.map.200kb.bgs + scale_color_manual(values = c("black", brewer.pal(n = 9, name = "Set1")[1:8], "#000000"))
theta.map.200kb.bgs <- theta.map.200kb.bgs + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous(breaks = pretty_breaks())
theta.map.200kb.bgs <- theta.map.200kb.bgs + labs(title = NULL, x = NULL, y = NULL)
theta.map.200kb.bgs <- theta.map.200kb.bgs + theme(text = element_text(size = 20), axis.title.x = element_text(size = 20))
theta.map.200kb.bgs
```



5.5 1 Mb scale

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

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

sim.theta.1Mb <- read.table("dm_bgs_sims/MutationMap_1000kb.csv", sep = ",", header = T)
sim.rho.1Mb <- read.table("dm_bgs_sims/RecombinationMap_1000kb.csv", sep = ",", header = T)
```

5.5.1 Replicate 1

```
rep1.sim.tmrca.1Mb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep1_w1000000.csv", sep = ",", header = T)

rep_1.pi.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep1/dm2L_bgs_rep1_5.div")
rep_1.pi.1Mb$avg <- apply(rep_1.pi.1Mb[4:ncol(rep_1.pi.1Mb)], 1, mean)
rep_1.rho.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep1/dm2L_bgs_rep1_5.rho")
rep_1.theta.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep1/dm2L_bgs_rep1_5.theta")
rep_1.tmrca.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep1/dm2L_bgs_rep1_5.tmrca")
rep_1.tmrca.1Mb$avg <- apply(rep_1.tmrca.1Mb[4:ncol(rep_1.tmrca.1Mb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.1Mb.rep1 <- as.data.frame(cbind(rep_1.pi.1Mb$avg, sim.theta.1Mb$Rate, rep1.sim.tmrca.1Mb$Avg))
```

```

names(true.lands.1Mb.rep1) <- c("diversity", "theta", "tmrca", "rho")
true.lands.1Mb.rep1$thetaC <- (true.lands.1Mb.rep1$theta - mean(true.lands.1Mb.rep1$theta)) / sd(true.lands.1Mb.rep1$theta)
true.lands.1Mb.rep1$tmrcaC <- (true.lands.1Mb.rep1$tmrca - mean(true.lands.1Mb.rep1$tmrca)) / sd(true.lands.1Mb.rep1$tmrca)
true.lands.1Mb.rep1$rhoC <- (true.lands.1Mb.rep1$rho - mean(true.lands.1Mb.rep1$rho), na.rm = T) / sd(true.lands.1Mb.rep1$rho)

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.1Mb.rep1)
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.1Mb.true[1, 1] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]) * 100
r2.bgs.1Mb.true[2, 1] <- anova.diversity$VarExp[1] * 100
r2.bgs.1Mb.true[3, 1] <- anova.diversity$VarExp[2] * 100
r2.bgs.1Mb.true[4, 1] <- anova.diversity$VarExp[3] * 100
r2.bgs.1Mb.true[5, 1] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.1Mb.rep1 <- as.data.frame(cbind(rep_1.pi.1Mb$avg, rep_1.theta.1Mb$sample_mean, rep_1.tmrca.1Mb$sample_mean))
names(inf.lands.1Mb.rep1) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[1, 3] <- cor.test(sim.theta.1Mb$Rate, rep_1.theta.1Mb$sample_mean, method = "spearman")
cor.table.rho[1, 3] <- cor.test(sim.rho.1Mb$Rate, rep_1.rho.1Mb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[1, 3] <- cor.test(rep1.sim.tmrca.1Mb$AverageTmrca, rep_1.tmrca.1Mb$avg, method = "spearman")$estimate

# standardizing
inf.lands.1Mb.rep1$thetaC <- (inf.lands.1Mb.rep1$theta - mean(inf.lands.1Mb.rep1$theta)) / sd(inf.lands.1Mb.rep1$theta)
inf.lands.1Mb.rep1$tmrcaC <- (inf.lands.1Mb.rep1$tmrca - mean(inf.lands.1Mb.rep1$tmrca)) / sd(inf.lands.1Mb.rep1$tmrca)
inf.lands.1Mb.rep1$rhoC <- (inf.lands.1Mb.rep1$rho - mean(inf.lands.1Mb.rep1$rho)) / sd(inf.lands.1Mb.rep1$rho)

inf.lands.1Mb.rep1$bin <- 1:nrow(inf.lands.1Mb.rep1)

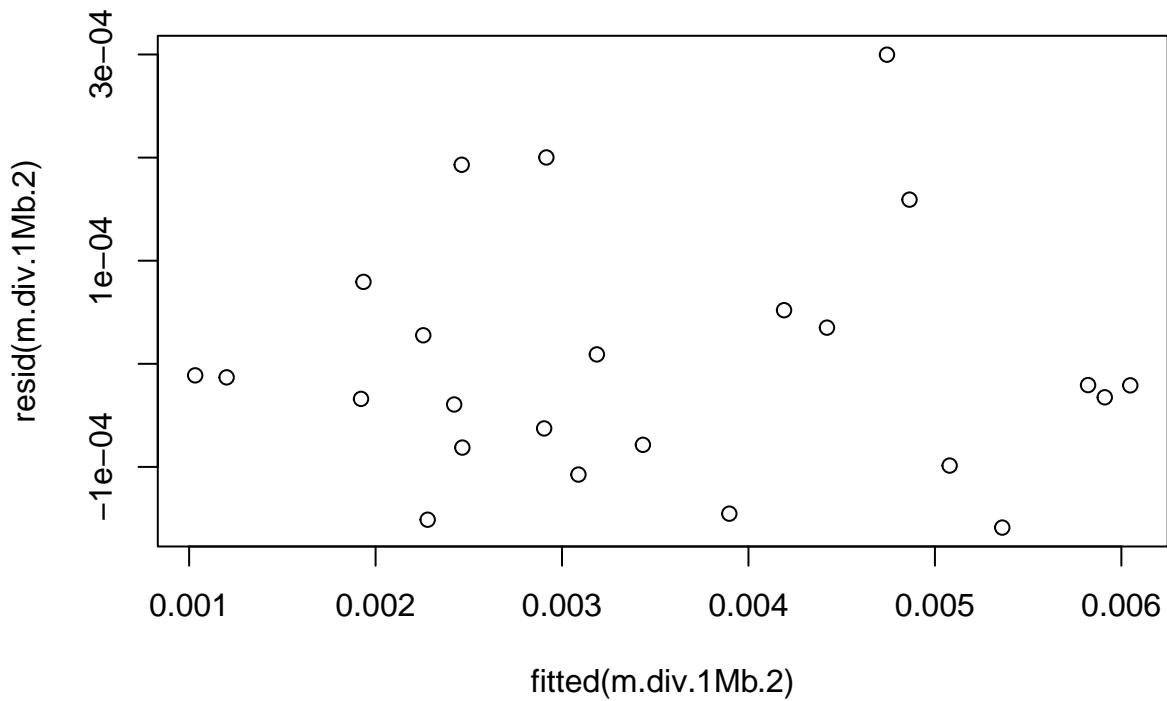
inf.lands.1Mb.rep1$Replicate <- 1

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.1Mb.rep1)
m.div.1Mb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.1Mb.rep1)
m.div.1Mb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC)^2, data = inf.lands.1Mb.rep1)

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

##          df      AIC
## m.div.1Mb    6 -350.1912
## m.div.1Mb.2   7 -353.4005
## m.div.1Mb.3   8 -351.4502
plot(resid(m.div.1Mb.2)~fitted(m.div.1Mb.2))

```

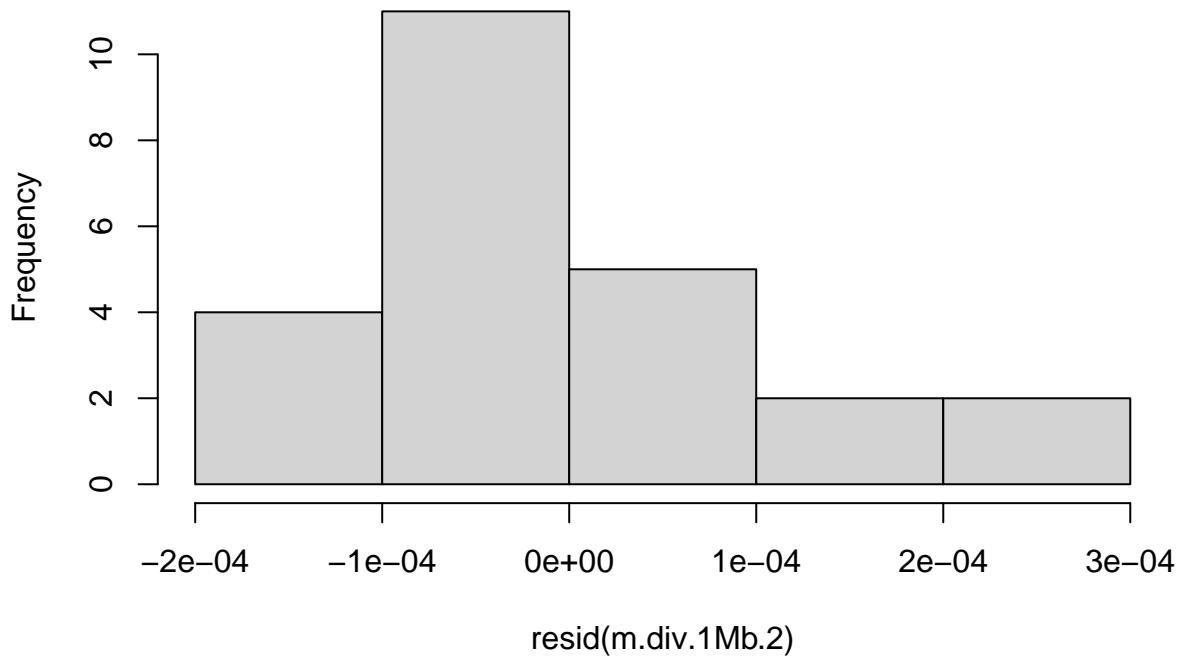


```
dwtest(m.div.1Mb.2)
```

```
##
##  Durbin-Watson test
##
## data: m.div.1Mb.2
## DW = 2.0504, p-value = 0.4914
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.1Mb.2)
```

```
##
##  Harrison-McCabe test
##
## data: m.div.1Mb.2
## HMC = 0.57197, p-value = 0.686
hist(resid(m.div.1Mb.2))
```

Histogram of resid(m.div.1Mb.2)



```
summary(m.div.1Mb.2)

##
## Call:
## lm(formula = diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC +
##     rhoC:tmrcaC, data = inf.lands.1Mb.rep1)
##
## Residuals:
##      Min        1Q    Median        3Q       Max
## -1.588e-04 -7.924e-05 -2.075e-05  3.932e-05  2.998e-04
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.342e-03 5.732e-05  58.303 < 2e-16 ***
## thetaC      9.969e-04 7.367e-05 13.532 7.13e-11 ***
## rhoC       5.945e-05 5.593e-05  1.063  0.3018
## tmrcaC     9.781e-04 1.159e-04  8.436 1.14e-07 ***
## thetaC:tmrcaC 3.895e-04 1.675e-04   2.325  0.0319 *
## rhoC:tmrcaC  1.602e-04 7.668e-05   2.089  0.0512 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001325 on 18 degrees of freedom
## Multiple R-squared:  0.994, Adjusted R-squared:  0.9923
## F-statistic: 594.7 on 5 and 18 DF,  p-value: < 2.2e-16

anova.diversity <- Anova(m.div.1Mb.2)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
```

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

5.5.2 Replicate 2

```
rep2.sim.tmrca.1Mb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep2_w1000000.csv", sep = ",", header = T)

rep_2.pi.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep2/dm2L_bgs_rep2_5.div")
rep_2.pi.1Mb$avg <- apply(rep_2.pi.1Mb[4:ncol(rep_2.pi.1Mb)], 1, mean)
rep_2.rho.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep2/dm2L_bgs_rep2_5.rho")
rep_2.theta.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep2/dm2L_bgs_rep2_5.theta")
rep_2.tmrca.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep2/dm2L_bgs_rep2_5.tmrca")
rep_2.tmrca.1Mb$avg <- apply(rep_2.tmrca.1Mb[4:ncol(rep_2.tmrca.1Mb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.1Mb.rep2 <- as.data.frame(cbind(rep_2.pi.1Mb$avg, sim.theta.1Mb$Rate, rep2.sim.tmrca.1Mb$AverageTmrca))
names(true.lands.1Mb.rep2) <- c("diversity", "theta", "tmrca", "rho")
true.lands.1Mb.rep2$thetaC <- (true.lands.1Mb.rep2$theta - mean(true.lands.1Mb.rep2$theta)) / sd(true.lands.1Mb.rep2$theta)
true.lands.1Mb.rep2$tmrcaC <- (true.lands.1Mb.rep2$tmrca - mean(true.lands.1Mb.rep2$tmrca)) / sd(true.lands.1Mb.rep2$tmrca)
true.lands.1Mb.rep2$rhoC <- (true.lands.1Mb.rep2$rho - mean(true.lands.1Mb.rep2$rho, na.rm = T)) / sd(true.lands.1Mb.rep2$rho)

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.1Mb.rep2)
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.1Mb.true[1, 2] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] * 100)
r2.bgs.1Mb.true[2, 2] <- anova.diversity$VarExp[1] * 100
r2.bgs.1Mb.true[3, 2] <- anova.diversity$VarExp[2] * 100
r2.bgs.1Mb.true[4, 2] <- anova.diversity$VarExp[3] * 100
r2.bgs.1Mb.true[5, 2] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.1Mb.rep2 <- as.data.frame(cbind(rep_2.pi.1Mb$avg, rep_2.theta.1Mb$sample_mean, rep_2.tmrca.1Mb$AverageTmrca))
names(inf.lands.1Mb.rep2) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[2, 3] <- cor.test(sim.theta.1Mb$Rate, rep_2.theta.1Mb$sample_mean, method = "spearman")
cor.table.rho[2, 3] <- cor.test(sim.rho.1Mb$Rate, rep_2.rho.1Mb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[2, 3] <- cor.test(rep2.sim.tmrca.1Mb$AverageTmrca, rep_2.tmrca.1Mb$avg, method = "spearman")

# standardizing
inf.lands.1Mb.rep2$thetaC <- (inf.lands.1Mb.rep2$theta - mean(inf.lands.1Mb.rep2$theta)) / sd(inf.lands.1Mb.rep2$theta)
inf.lands.1Mb.rep2$tmrcaC <- (inf.lands.1Mb.rep2$tmrca - mean(inf.lands.1Mb.rep2$tmrca)) / sd(inf.lands.1Mb.rep2$tmrca)
inf.lands.1Mb.rep2$rhoC <- (inf.lands.1Mb.rep2$rho - mean(inf.lands.1Mb.rep2$rho)) / sd(inf.lands.1Mb.rep2$rho)

inf.lands.1Mb.rep2$bin <- 1:nrow(inf.lands.1Mb.rep2)

inf.lands.1Mb.rep2$Replicate <- 2

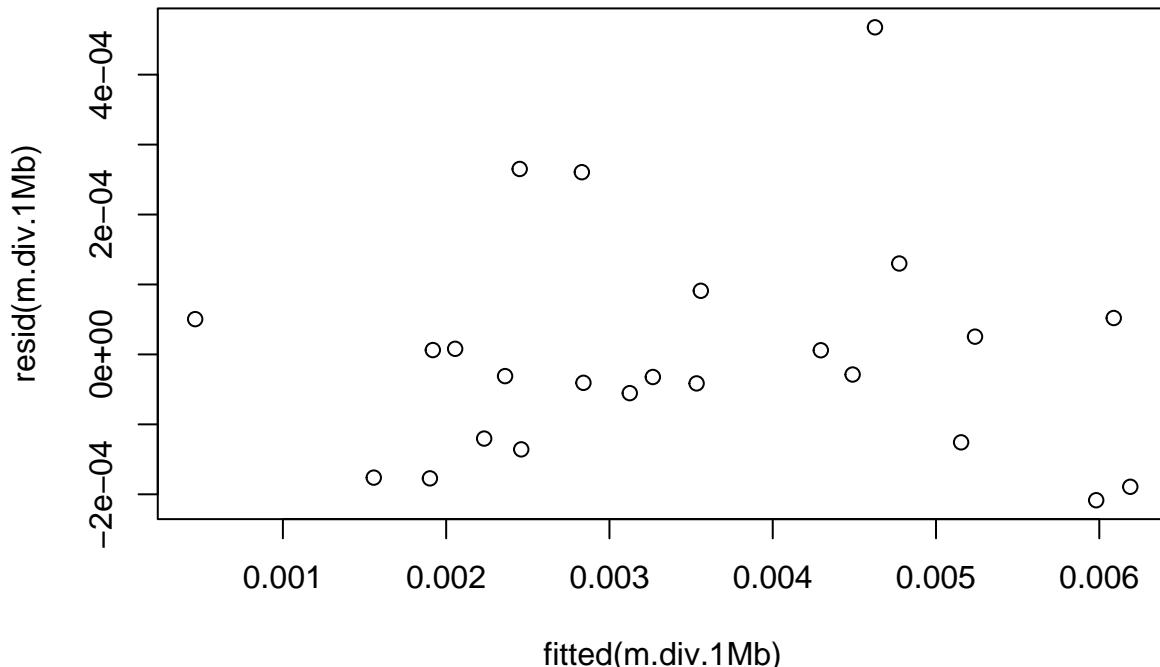
m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.1Mb.rep2)
```

```
m.div.1Mb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.1Mb)
m.div.1Mb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.1Mb.rep2)
```

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

```
##          df      AIC
## m.div.1Mb     6 -340.4890
## m.div.1Mb.2   7 -338.4960
## m.div.1Mb.3   8 -340.4696
```

```
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))
```

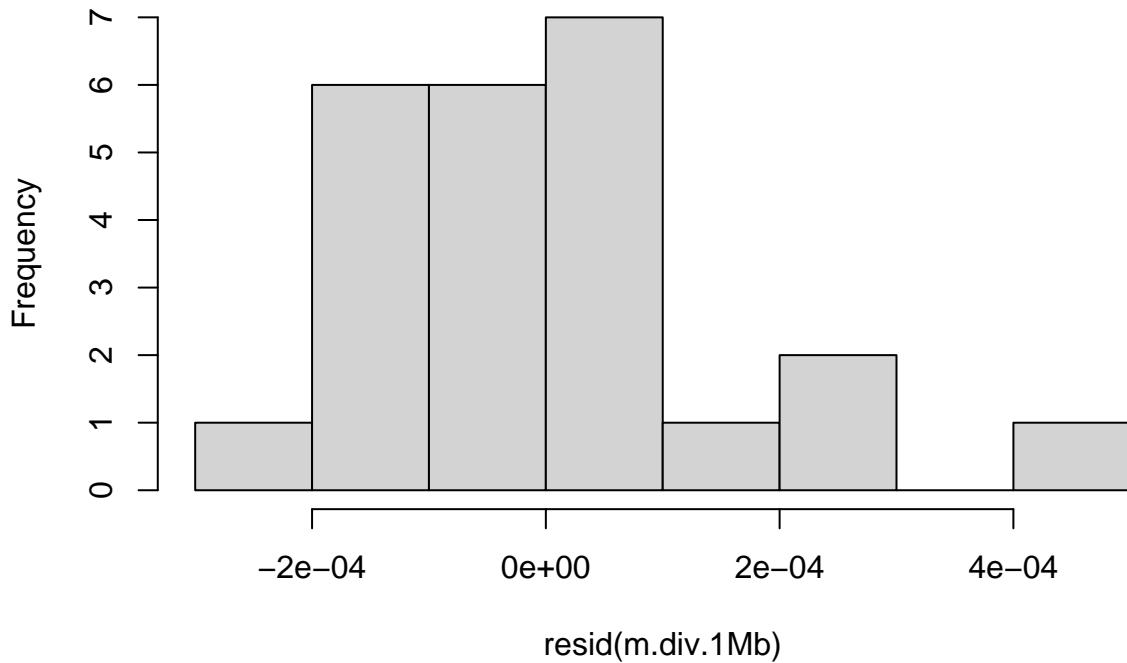


```
dwtest(m.div.1Mb)
```

```
##
##  Durbin-Watson test
##
##  data:  m.div.1Mb
##  DW = 2.2615, p-value = 0.5747
##  alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.1Mb)
```

```
##
##  Harrison-McCabe test
##
##  data:  m.div.1Mb
##  HMC = 0.50859, p-value = 0.502
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 = inf.lands.1Mb.rep2)
##
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -2.085e-04 -1.217e-04 -2.989e-05  5.085e-05  4.677e-04
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.476e-03 3.592e-05 96.792 < 2e-16 ***
## thetaC      1.130e-03 4.817e-05 23.465 1.71e-15 ***
## rhoC        1.368e-05 5.337e-05  0.256  0.8005  
## tmrcaC      9.845e-04 6.117e-05 16.094 1.59e-12 ***
## thetaC:tmrcaC 1.300e-04 6.448e-05  2.016  0.0582 .  
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.0001759 on 19 degrees of freedom
## Multiple R-squared:  0.9897, Adjusted R-squared:  0.9876 
## F-statistic: 457.9 on 4 and 19 DF,  p-value: < 2.2e-16

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

r2.bgs.1Mb[1, 2] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] +
```

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

5.5.3 Replicate 3

```
rep3.sim.tmrca.1Mb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep3_w1000000.csv", sep = ",", header = T)

rep_3.pi.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep3/dm2L_bgs_rep3_5.div")
rep_3.pi.1Mb$avg <- apply(rep_3.pi.1Mb[4:ncol(rep_3.pi.1Mb)], 1, mean)
rep_3.rho.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep3/dm2L_bgs_rep3_5.rho")
rep_3.theta.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep3/dm2L_bgs_rep3_5.theta")
rep_3.tmrca.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep3/dm2L_bgs_rep3_5.tmrca")
rep_3.tmrca.1Mb$avg <- apply(rep_3.tmrca.1Mb[4:ncol(rep_3.tmrca.1Mb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.1Mb.rep3 <- as.data.frame(cbind(rep_2.pi.1Mb$avg, sim.theta.1Mb$Rate, rep3.sim.tmrca.1Mb$Average))
names(true.lands.1Mb.rep3) <- c("diversity", "theta", "tmrca", "rho")
true.lands.1Mb.rep3$thetaC <- (true.lands.1Mb.rep3$theta - mean(true.lands.1Mb.rep3$theta)) / sd(true.lands.1Mb.rep3$theta)
true.lands.1Mb.rep3$tmrcaC <- (true.lands.1Mb.rep3$tmrca - mean(true.lands.1Mb.rep3$tmrca)) / sd(true.lands.1Mb.rep3$tmrca)
true.lands.1Mb.rep3$rhoC <- (true.lands.1Mb.rep3$rho - mean(true.lands.1Mb.rep3$rho, na.rm = T)) / sd(true.lands.1Mb.rep3$rho)

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.1Mb.rep3)
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.1Mb.true[1, 3] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]) * 100
r2.bgs.1Mb.true[2, 3] <- anova.diversity$VarExp[1] * 100
r2.bgs.1Mb.true[3, 3] <- anova.diversity$VarExp[2] * 100
r2.bgs.1Mb.true[4, 3] <- anova.diversity$VarExp[3] * 100
r2.bgs.1Mb.true[5, 3] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.1Mb.rep3 <- as.data.frame(cbind(rep_3.pi.1Mb$avg, rep_3.theta.1Mb$sample_mean, rep_3.tmrca.1Mb$Average))
names(inf.lands.1Mb.rep3) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[3, 3] <- cor.test(sim.theta.1Mb$Rate, rep_3.theta.1Mb$sample_mean, method = "spearman")
cor.table.rho[3, 3] <- cor.test(sim.rho.1Mb$Rate, rep_3.rho.1Mb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[3, 3] <- cor.test(rep3.sim.tmrca.1Mb$AverageTmrca, rep_3.tmrca.1Mb$avg, method = "spearman")

# standardizing
inf.lands.1Mb.rep3$thetaC <- (inf.lands.1Mb.rep3$theta - mean(inf.lands.1Mb.rep3$theta)) / sd(inf.lands.1Mb.rep3$theta)
inf.lands.1Mb.rep3$tmrcaC <- (inf.lands.1Mb.rep3$tmrca - mean(inf.lands.1Mb.rep3$tmrca)) / sd(inf.lands.1Mb.rep3$tmrca)
inf.lands.1Mb.rep3$rhoC <- (inf.lands.1Mb.rep3$rho - mean(inf.lands.1Mb.rep3$rho)) / sd(inf.lands.1Mb.rep3$rho)

inf.lands.1Mb.rep3$bin <- 1:nrow(inf.lands.1Mb.rep3)

inf.lands.1Mb.rep3$Replicate <- 3

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.1Mb.rep3)
m.div.1Mb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.1Mb.rep3)
```

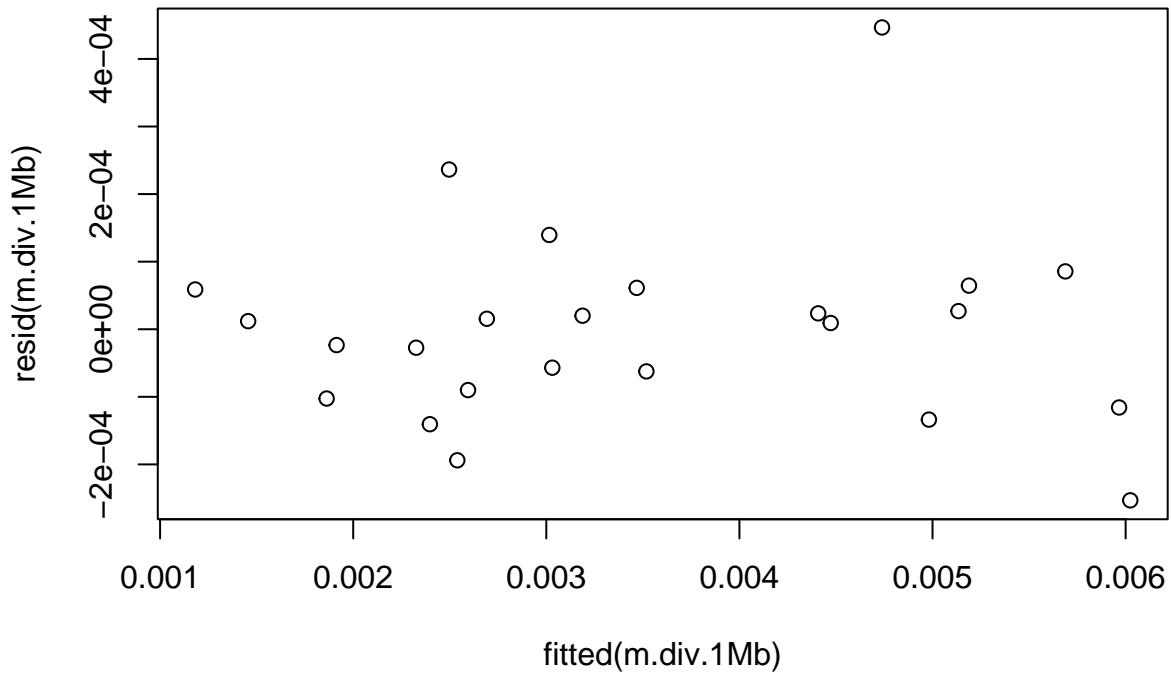
```

m.div.1Mb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.1Mb.rep3)

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

##          df      AIC
## m.div.1Mb    6 -345.5820
## m.div.1Mb.2  7 -346.9896
## m.div.1Mb.3  8 -348.3194
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```

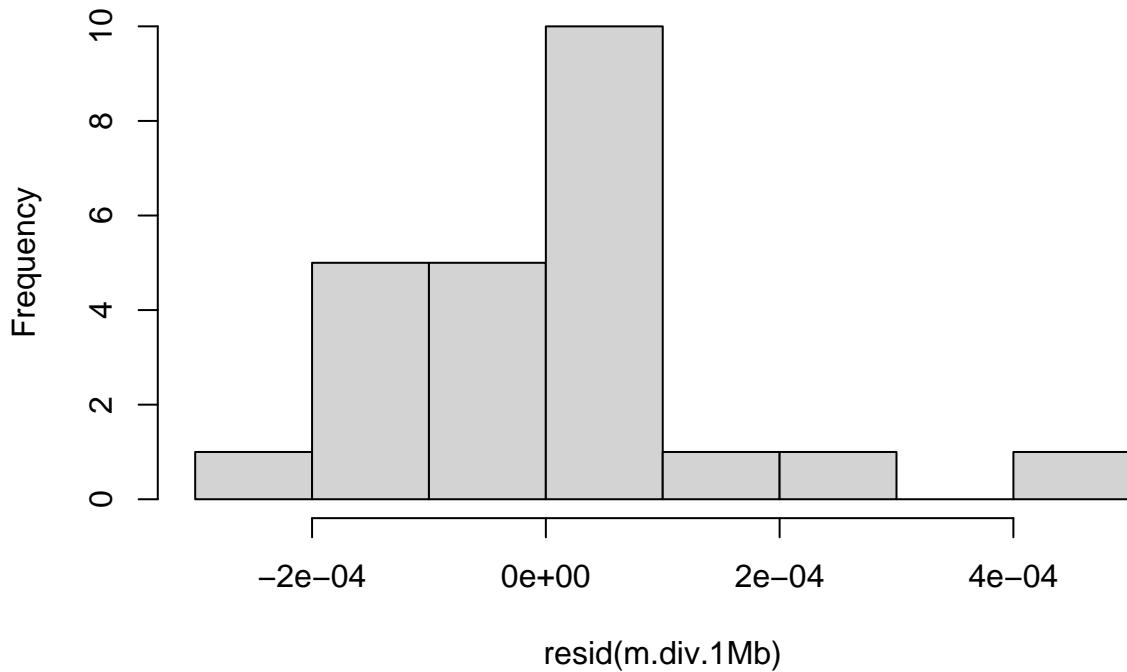
dwtest(m.div.1Mb)

##
##  Durbin-Watson test
##
## data: m.div.1Mb
## DW = 2.5052, p-value = 0.8416
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.1Mb)

##
##  Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.52869, p-value = 0.571
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 = inf.lands.1Mb.rep3)
##
## Residuals:
##       Min         1Q     Median         3Q        Max
## -2.532e-04 -9.316e-05  1.061e-05  5.944e-05  4.466e-04
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.517e-03 3.231e-05 108.850 < 2e-16 ***
## thetaC      1.114e-03 3.870e-05  28.788 < 2e-16 ***
## rhoC       -1.151e-05 4.912e-05  -0.234  0.81719
## tmrcaC      8.973e-04 5.060e-05  17.732 2.81e-13 ***
## thetaC:tmrcaC 1.546e-04 3.399e-05   4.549  0.00022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001582 on 19 degrees of freedom
## Multiple R-squared:  0.9905, Adjusted R-squared:  0.9885
## F-statistic: 495.2 on 4 and 19 DF,  p-value: < 2.2e-16

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

r2.bgs.1Mb[1, 3] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
```

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

5.5.4 Replicate 4

```
rep4.sim.tmrca.1Mb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep4_w1000000.csv", sep = ",", header = T)

rep_4.pi.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep4/dm2L_bgs_rep4_5.div")
rep_4.pi.1Mb$avg <- apply(rep_4.pi.1Mb[4:ncol(rep_4.pi.1Mb)], 1, mean)
rep_4.rho.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep4/dm2L_bgs_rep4_5.rho")
rep_4.theta.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep4/dm2L_bgs_rep4_5.theta")
rep_4.tmrca.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep4/dm2L_bgs_rep4_5.tmrca")
rep_4.tmrca.1Mb$avg <- apply(rep_4.tmrca.1Mb[4:ncol(rep_4.tmrca.1Mb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.1Mb.rep4 <- as.data.frame(cbind(rep_4.pi.1Mb$avg, sim.theta.1Mb$Rate, rep4.sim.tmrca.1Mb$Average))
names(true.lands.1Mb.rep4) <- c("diversity", "theta", "tmrca", "rho")
true.lands.1Mb.rep4$thetaC <- (true.lands.1Mb.rep4$theta - mean(true.lands.1Mb.rep4$theta)) / sd(true.lands.1Mb.rep4$theta)
true.lands.1Mb.rep4$tmrcaC <- (true.lands.1Mb.rep4$tmrca - mean(true.lands.1Mb.rep4$tmrca)) / sd(true.lands.1Mb.rep4$tmrca)
true.lands.1Mb.rep4$rhoC <- (true.lands.1Mb.rep4$rho - mean(true.lands.1Mb.rep4$rho, na.rm = T)) / sd(true.lands.1Mb.rep4$rho)

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.1Mb.rep4)
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.1Mb.true[1, 4] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]) * 100
r2.bgs.1Mb.true[2, 4] <- anova.diversity$VarExp[1] * 100
r2.bgs.1Mb.true[3, 4] <- anova.diversity$VarExp[2] * 100
r2.bgs.1Mb.true[4, 4] <- anova.diversity$VarExp[3] * 100
r2.bgs.1Mb.true[5, 4] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.1Mb.rep4 <- as.data.frame(cbind(rep_4.pi.1Mb$avg, rep_4.theta.1Mb$sample_mean, rep_4.tmrca.1Mb$Average))
names(inf.lands.1Mb.rep4) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[4, 3] <- cor.test(sim.theta.1Mb$Rate, rep_4.theta.1Mb$sample_mean, method = "spearman")
cor.table.rho[4, 3] <- cor.test(sim.rho.1Mb$Rate, rep_4.rho.1Mb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[4, 3] <- cor.test(rep4.sim.tmrca.1Mb$AverageTmrca, rep_4.tmrca.1Mb$avg, method = "spearman")

# standardizing
inf.lands.1Mb.rep4$thetaC <- (inf.lands.1Mb.rep4$theta - mean(inf.lands.1Mb.rep4$theta)) / sd(inf.lands.1Mb.rep4$theta)
inf.lands.1Mb.rep4$tmrcaC <- (inf.lands.1Mb.rep4$tmrca - mean(inf.lands.1Mb.rep4$tmrca)) / sd(inf.lands.1Mb.rep4$tmrca)
inf.lands.1Mb.rep4$rhoC <- (inf.lands.1Mb.rep4$rho - mean(inf.lands.1Mb.rep4$rho)) / sd(inf.lands.1Mb.rep4$rho)

inf.lands.1Mb.rep4$bin <- 1:nrow(inf.lands.1Mb.rep4)

inf.lands.1Mb.rep4$Replicate <- 4

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.1Mb.rep4)
m.div.1Mb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.1Mb.rep4)
```

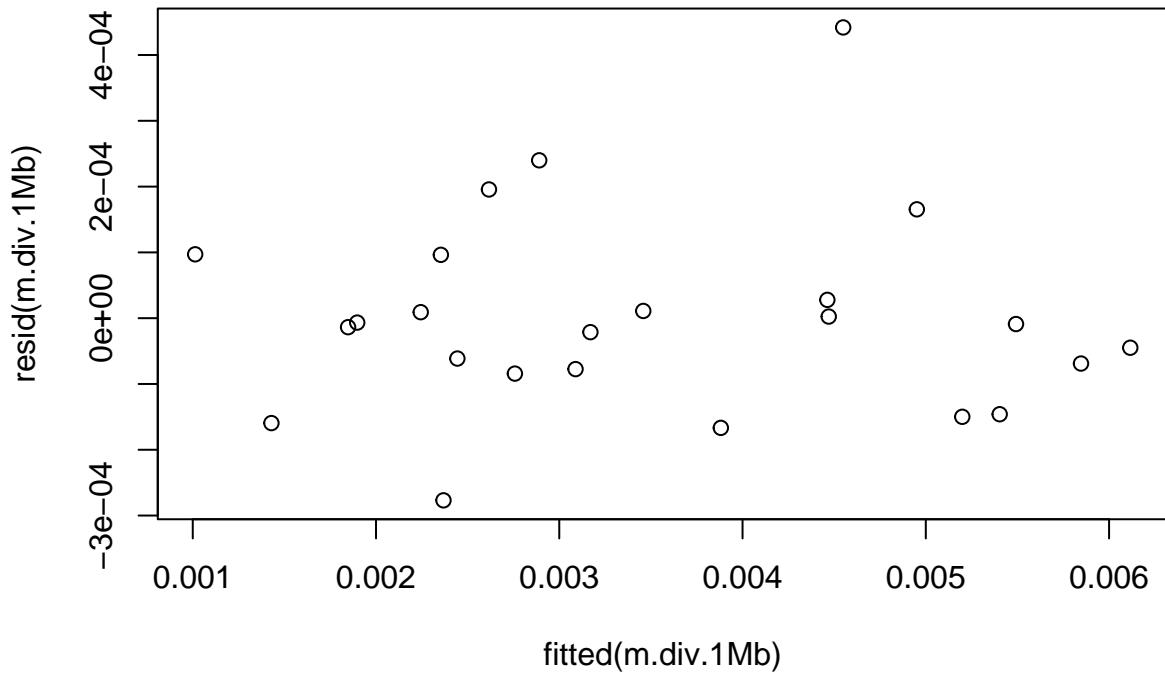
```

m.div.1Mb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcuC) ^ 2, data = inf.lands.1Mb.rep4)

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

##          df      AIC
## m.div.1Mb    6 -342.6901
## m.div.1Mb.2  7 -343.4354
## m.div.1Mb.3  8 -341.7061
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```

dwtest(m.div.1Mb)

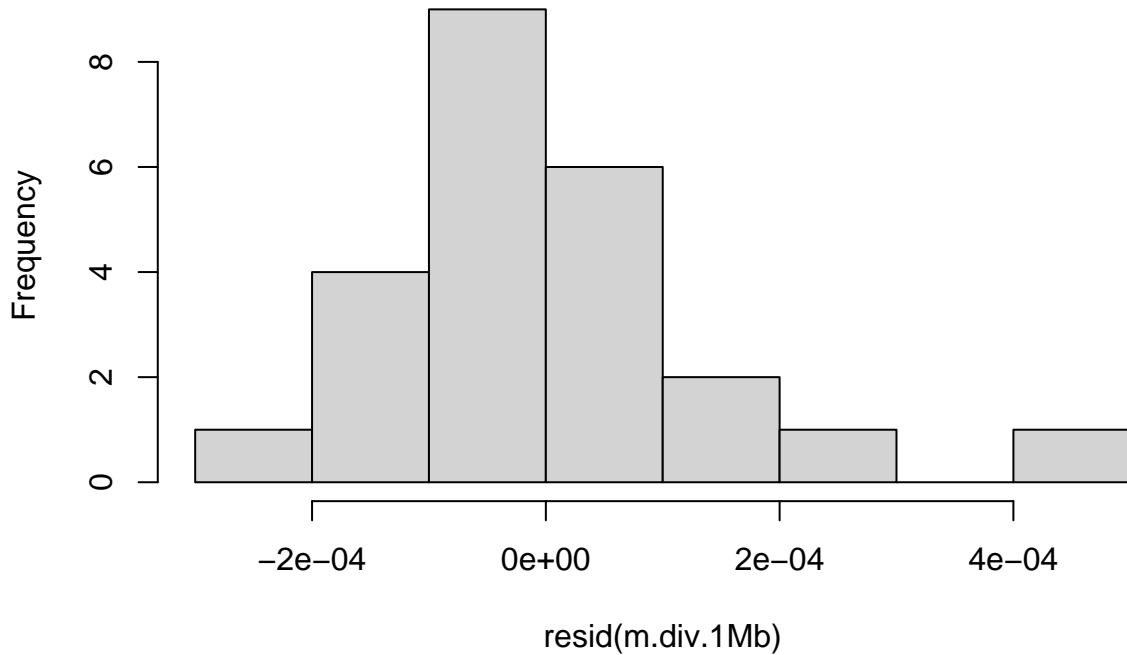
##
##  Durbin-Watson test
##
## data: m.div.1Mb
## DW = 1.8262, p-value = 0.2127
## alternative hypothesis: true autocorrelation is greater than 0

hmctest(m.div.1Mb)

##
##  Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.67938, p-value = 0.902
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 = inf.lands.1Mb.rep4)
##
## Residuals:
##      Min        1Q    Median        3Q       Max 
## -2.769e-04 -7.914e-05 -1.131e-05  4.493e-05  4.419e-04 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.458e-03 3.683e-05 93.894 < 2e-16 ***
## thetaC      1.003e+00 4.174e-02 24.022 1.11e-15 ***
## rhoC        5.858e-02 3.134e-02  1.869  0.07708  
## tmrcaC      5.291e-03 4.059e-04 13.037 6.31e-11 ***
## thetaC:tmrcaC 1.875e+00 6.290e-01   2.981  0.00768 ** 
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 0.000168 on 19 degrees of freedom
## Multiple R-squared:  0.9896, Adjusted R-squared:  0.9874 
## F-statistic: 452.5 on 4 and 19 DF,  p-value: < 2.2e-16 

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

r2.bgs.1Mb[1, 4] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] +
```

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

5.5.5 Replicate 5

```
rep5.sim.tmrca.1Mb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep5_w1000000.csv", sep = ",", header = T)

rep_5.pi.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep5/dm2L_bgs_rep5_5.div")
rep_5.pi.1Mb$avg <- apply(rep_5.pi.1Mb[4:ncol(rep_5.pi.1Mb)], 1, mean)
rep_5.rho.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep5/dm2L_bgs_rep5_5.rho")
rep_5.theta.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep5/dm2L_bgs_rep5_5.theta")
rep_5.tmrca.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep5/dm2L_bgs_rep5_5.tmrca")
rep_5.tmrca.1Mb$avg <- apply(rep_5.tmrca.1Mb[4:ncol(rep_5.tmrca.1Mb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.1Mb.rep5 <- as.data.frame(cbind(rep_4.pi.1Mb$avg, sim.theta.1Mb$Rate, rep5.sim.tmrca.1Mb$Average))
names(true.lands.1Mb.rep5) <- c("diversity", "theta", "tmrca", "rho")
true.lands.1Mb.rep5$thetaC <- (true.lands.1Mb.rep5$theta - mean(true.lands.1Mb.rep5$theta)) / sd(true.lands.1Mb.rep5$theta)
true.lands.1Mb.rep5$tmrcaC <- (true.lands.1Mb.rep5$tmrca - mean(true.lands.1Mb.rep5$tmrca)) / sd(true.lands.1Mb.rep5$tmrca)
true.lands.1Mb.rep5$rhoC <- (true.lands.1Mb.rep5$rho - mean(true.lands.1Mb.rep5$rho, na.rm = T)) / sd(true.lands.1Mb.rep5$rho)

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.1Mb.rep5)
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.1Mb.true[1, 5] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]) * 100
r2.bgs.1Mb.true[2, 5] <- anova.diversity$VarExp[1] * 100
r2.bgs.1Mb.true[3, 5] <- anova.diversity$VarExp[2] * 100
r2.bgs.1Mb.true[4, 5] <- anova.diversity$VarExp[3] * 100
r2.bgs.1Mb.true[5, 5] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.1Mb.rep5 <- as.data.frame(cbind(rep_5.pi.1Mb$avg, rep_5.theta.1Mb$sample_mean, rep_5.tmrca.1Mb$Average))
names(inf.lands.1Mb.rep5) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[5, 3] <- cor.test(sim.theta.1Mb$Rate, rep_5.theta.1Mb$sample_mean, method = "spearman")
cor.table.rho[5, 3] <- cor.test(sim.rho.1Mb$Rate, rep_5.rho.1Mb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[5, 3] <- cor.test(rep5.sim.tmrca.1Mb$AverageTmrca, rep_5.tmrca.1Mb$avg, method = "spearman")

# standardizing
inf.lands.1Mb.rep5$thetaC <- (inf.lands.1Mb.rep5$theta - mean(inf.lands.1Mb.rep5$theta)) / sd(inf.lands.1Mb.rep5$theta)
inf.lands.1Mb.rep5$tmrcaC <- (inf.lands.1Mb.rep5$tmrca - mean(inf.lands.1Mb.rep5$tmrca)) / sd(inf.lands.1Mb.rep5$tmrca)
inf.lands.1Mb.rep5$rhoC <- (inf.lands.1Mb.rep5$rho - mean(inf.lands.1Mb.rep5$rho)) / sd(inf.lands.1Mb.rep5$rho)

inf.lands.1Mb.rep5$bin <- 1:nrow(inf.lands.1Mb.rep5)

inf.lands.1Mb.rep5$Replicate <- 5

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.1Mb.rep5)
m.div.1Mb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.1Mb.rep5)
```

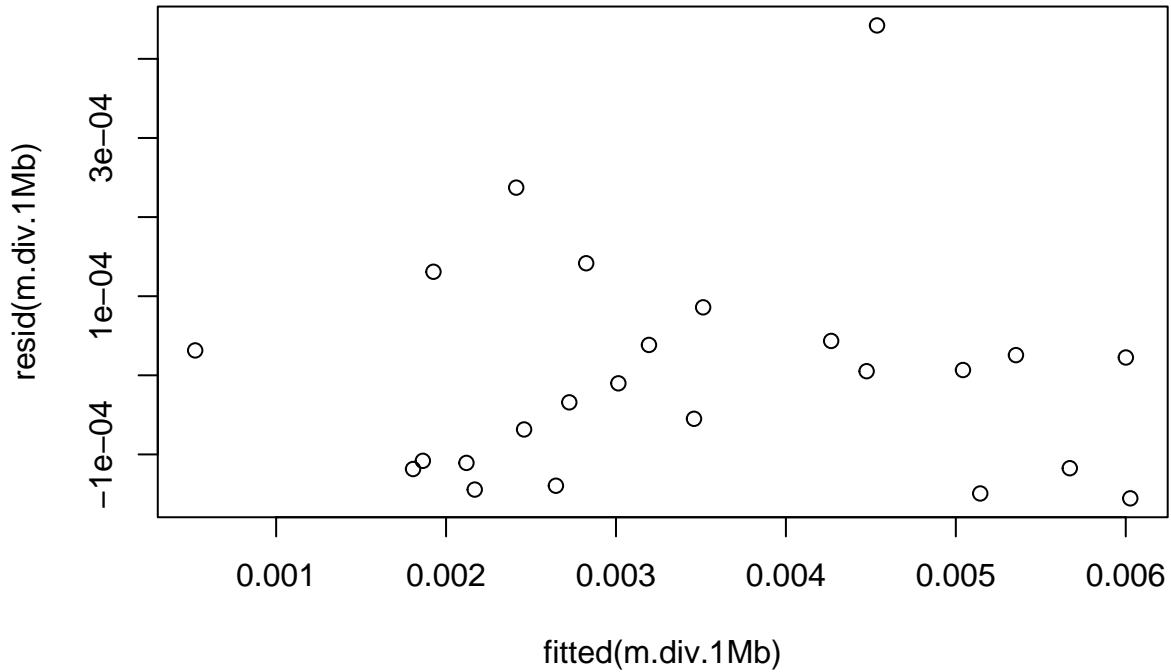
```

m.div.1Mb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.1Mb.rep5)

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

##          df      AIC
## m.div.1Mb    6 -346.9226
## m.div.1Mb.2  7 -345.5662
## m.div.1Mb.3  8 -347.9807
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```

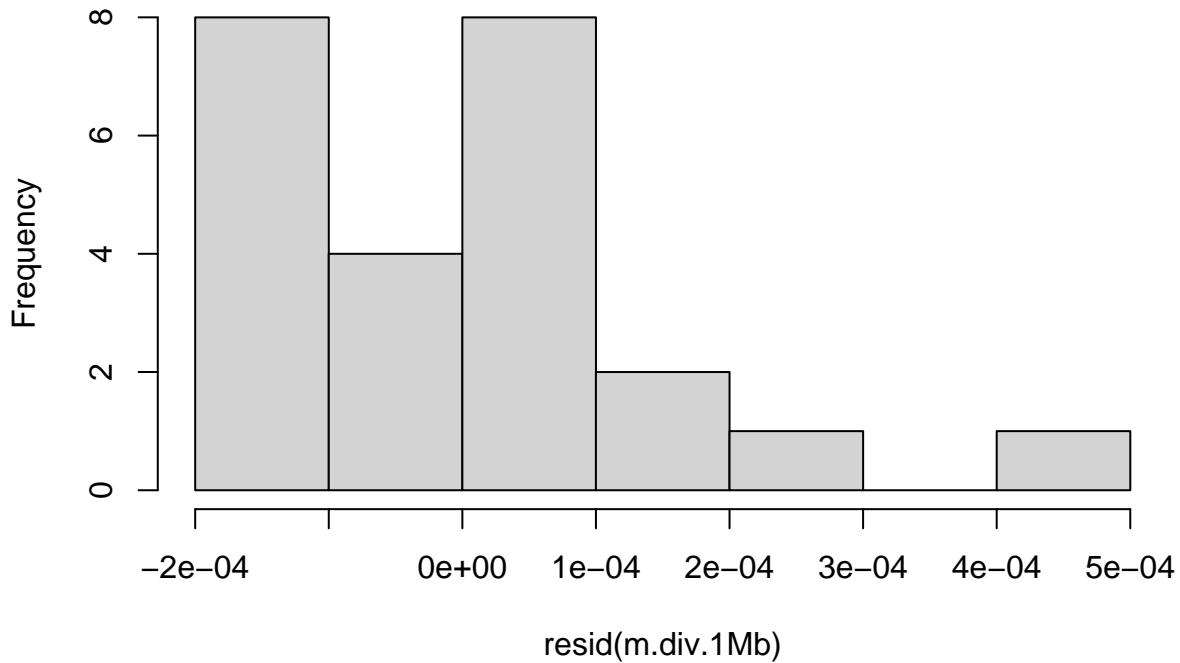
dwtest(m.div.1Mb)

##
##  Durbin-Watson test
##
## data: m.div.1Mb
## DW = 2.3054, p-value = 0.6445
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.1Mb)

##
##  Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.58123, p-value = 0.72
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 = inf.lands.1Mb.rep5)
##
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -1.555e-04 -1.124e-04 -2.460e-06  3.966e-05  4.422e-04
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.463e-03 3.141e-05 110.243 < 2e-16 ***
## thetaC      1.151e-03 4.098e-05  28.076 < 2e-16 ***
## rhoC       -1.837e-05 4.694e-05  -0.391  0.69992  
## tmrcaC      8.697e-04 4.908e-05  17.720 2.84e-13 ***
## thetaC:tmrcaC 1.694e-04 4.923e-05   3.441  0.00274 ** 
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.0001538 on 19 degrees of freedom
## Multiple R-squared:  0.9917, Adjusted R-squared:  0.9899 
## F-statistic: 564.1 on 4 and 19 DF,  p-value: < 2.2e-16

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

r2.bgs.1Mb[1, 5] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] +
```

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

5.5.6 Replicate 6

```
rep6.sim.tmrca.1Mb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep6_w1000000.csv", sep = ",", header = T)

rep_6.pi.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep6/dm2L_bgs_rep6_5.div")
rep_6.pi.1Mb$avg <- apply(rep_6.pi.1Mb[4:ncol(rep_6.pi.1Mb)], 1, mean)
rep_6.rho.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep6/dm2L_bgs_rep6_5.rho")
rep_6.theta.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep6/dm2L_bgs_rep6_5.theta")
rep_6.tmrca.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep6/dm2L_bgs_rep6_5.tmrca")
rep_6.tmrca.1Mb$avg <- apply(rep_6.tmrca.1Mb[4:ncol(rep_6.tmrca.1Mb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.1Mb.rep6 <- as.data.frame(cbind(rep_6.pi.1Mb$avg, sim.theta.1Mb$Rate, rep6.sim.tmrca.1Mb$Average))
names(true.lands.1Mb.rep6) <- c("diversity", "theta", "tmrca", "rho")
true.lands.1Mb.rep6$thetaC <- (true.lands.1Mb.rep6$theta - mean(true.lands.1Mb.rep6$theta)) / sd(true.lands.1Mb.rep6$theta)
true.lands.1Mb.rep6$tmrcaC <- (true.lands.1Mb.rep6$tmrca - mean(true.lands.1Mb.rep6$tmrca)) / sd(true.lands.1Mb.rep6$tmrca)
true.lands.1Mb.rep6$rhoC <- (true.lands.1Mb.rep6$rho - mean(true.lands.1Mb.rep6$rho, na.rm = T)) / sd(true.lands.1Mb.rep6$rho)

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.1Mb.rep6)
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.1Mb.true[1, 6] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) * 100
r2.bgs.1Mb.true[2, 6] <- anova.diversity$VarExp[1] * 100
r2.bgs.1Mb.true[3, 6] <- anova.diversity$VarExp[2] * 100
r2.bgs.1Mb.true[4, 6] <- anova.diversity$VarExp[3] * 100
r2.bgs.1Mb.true[5, 6] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.1Mb.rep6 <- as.data.frame(cbind(rep_6.pi.1Mb$avg, rep_6.theta.1Mb$sample_mean, rep_6.tmrca.1Mb$Average))
names(inf.lands.1Mb.rep6) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[6, 3] <- cor.test(sim.theta.1Mb$Rate, rep_6.theta.1Mb$sample_mean, method = "spearman")
cor.table.rho[6, 3] <- cor.test(sim.rho.1Mb$Rate, rep_6.rho.1Mb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[6, 3] <- cor.test(rep6.sim.tmrca.1Mb$AverageTmrca, rep_6.tmrca.1Mb$avg, method = "spearman")

# standardizing
inf.lands.1Mb.rep6$thetaC <- (inf.lands.1Mb.rep6$theta - mean(inf.lands.1Mb.rep6$theta)) / sd(inf.lands.1Mb.rep6$theta)
inf.lands.1Mb.rep6$tmrcaC <- (inf.lands.1Mb.rep6$tmrca - mean(inf.lands.1Mb.rep6$tmrca)) / sd(inf.lands.1Mb.rep6$tmrca)
inf.lands.1Mb.rep6$rhoC <- (inf.lands.1Mb.rep6$rho - mean(inf.lands.1Mb.rep6$rho)) / sd(inf.lands.1Mb.rep6$rho)

inf.lands.1Mb.rep6$bin <- 1:nrow(inf.lands.1Mb.rep6)

inf.lands.1Mb.rep6$Replicate <- 6

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.1Mb.rep6)
m.div.1Mb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.1Mb.rep6)
```

```

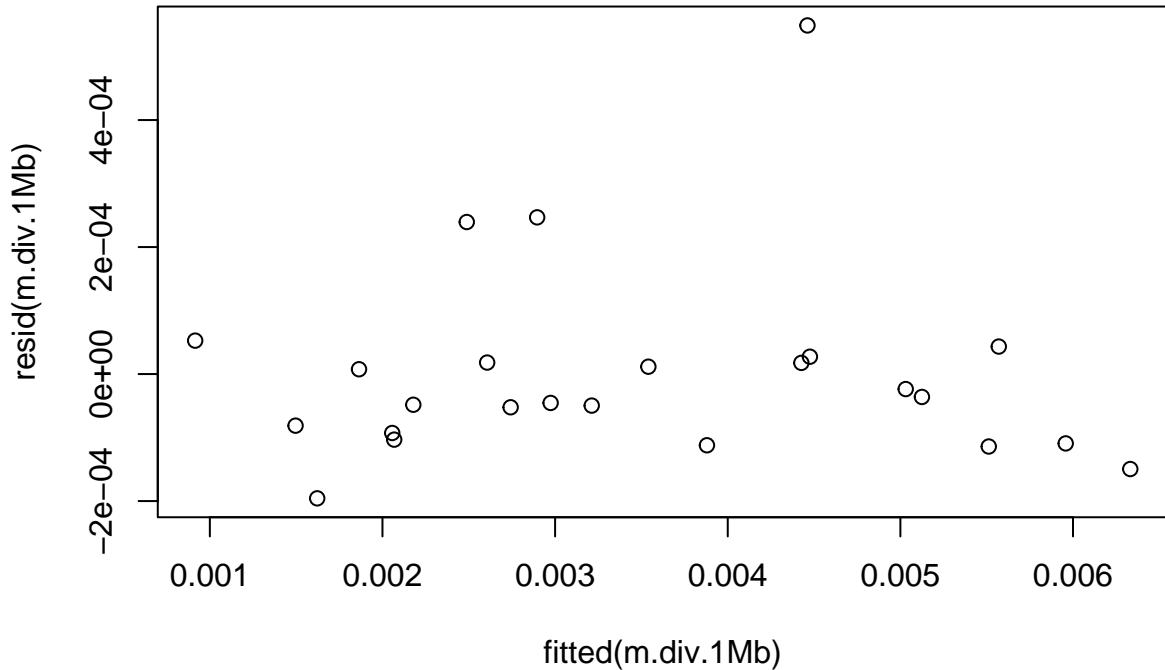
m.div.1Mb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.1Mb.rep6)

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

##          df      AIC
## m.div.1Mb    6 -341.5943
## m.div.1Mb.2  7 -342.4713
## m.div.1Mb.3  8 -343.2357

plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```

dwtest(m.div.1Mb)

##
##  Durbin-Watson test
##
## data: m.div.1Mb
## DW = 2.1636, p-value = 0.4919
## alternative hypothesis: true autocorrelation is greater than 0

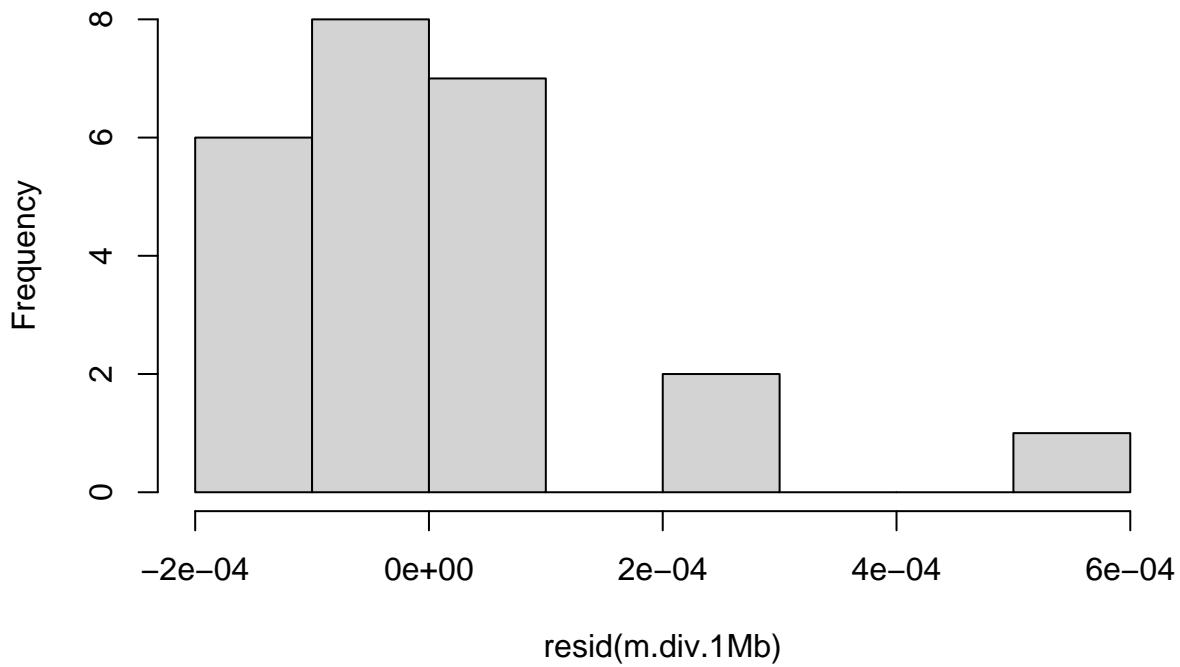
hmctest(m.div.1Mb)

##
##  Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.65357, p-value = 0.874

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 = inf.lands.1Mb.rep6)
##
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -1.957e-04 -9.541e-05 -4.075e-05  2.041e-05  5.490e-04
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.457e-03 3.707e-05 93.270 < 2e-16 ***
## thetaC      1.212e-03 4.895e-05 24.757 6.38e-16 ***
## rhoC        4.961e-05 5.086e-05  0.975  0.342    
## tmrcaC      7.178e-04 4.516e-05 15.894 1.98e-12 ***
## thetaC:tmrcaC 1.438e-04 9.074e-05   1.585   0.129  
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.0001719 on 19 degrees of freedom
## Multiple R-squared:  0.9901, Adjusted R-squared:  0.988 
## F-statistic: 475.1 on 4 and 19 DF,  p-value: < 2.2e-16

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

r2.bgs.1Mb[1, 6] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] +
```

```
r2.bgs.1Mb[2, 6] <- anova.diversity$VarExp[1] * 100
r2.bgs.1Mb[3, 6] <- anova.diversity$VarExp[2] * 100
r2.bgs.1Mb[4, 6] <- anova.diversity$VarExp[3] * 100
r2.bgs.1Mb[5, 6] <- anova.diversity$VarExp[4] * 100
```

5.5.7 Replicate 7

```
rep7.sim.tmrca.1Mb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep7_w1000000.csv", sep = ",", header = T)

rep_7.pi.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep7/dm2L_bgs_rep7_5.div")
rep_7.pi.1Mb$avg <- apply(rep_7.pi.1Mb[4:ncol(rep_7.pi.1Mb)], 1, mean)
rep_7.rho.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep7/dm2L_bgs_rep7_5.rho")
rep_7.theta.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep7/dm2L_bgs_rep7_5.theta")
rep_7.tmrca.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep7/dm2L_bgs_rep7_5.tmrca")
rep_7.tmrca.1Mb$avg <- apply(rep_7.tmrca.1Mb[4:ncol(rep_7.tmrca.1Mb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.1Mb.rep7 <- as.data.frame(cbind(rep_7.pi.1Mb$avg, sim.theta.1Mb$Rate, rep7.sim.tmrca.1Mb$Average))
names(true.lands.1Mb.rep7) <- c("diversity", "theta", "tmrca", "rho")
true.lands.1Mb.rep7$thetaC <- (true.lands.1Mb.rep7$theta - mean(true.lands.1Mb.rep7$theta)) / sd(true.lands.1Mb.rep7$theta)
true.lands.1Mb.rep7$tmrcaC <- (true.lands.1Mb.rep7$tmrca - mean(true.lands.1Mb.rep7$tmrca)) / sd(true.lands.1Mb.rep7$tmrca)
true.lands.1Mb.rep7$rhoC <- (true.lands.1Mb.rep7$rho - mean(true.lands.1Mb.rep7$rho, na.rm = T)) / sd(true.lands.1Mb.rep7$rho)

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.1Mb.rep7)
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.1Mb.true[1, 7] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] + anova.diversity$VarExp[4]) * 100
r2.bgs.1Mb.true[2, 7] <- anova.diversity$VarExp[1] * 100
r2.bgs.1Mb.true[3, 7] <- anova.diversity$VarExp[2] * 100
r2.bgs.1Mb.true[4, 7] <- anova.diversity$VarExp[3] * 100
r2.bgs.1Mb.true[5, 7] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.1Mb.rep7 <- as.data.frame(cbind(rep_7.pi.1Mb$avg, rep_7.theta.1Mb$sample_mean, rep_7.tmrca.1Mb$Average))
names(inf.lands.1Mb.rep7) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[7, 3] <- cor.test(sim.theta.1Mb$Rate, rep_7.theta.1Mb$sample_mean, method = "spearman")
cor.table.rho[7, 3] <- cor.test(sim.rho.1Mb$Rate, rep_7.rho.1Mb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[7, 3] <- cor.test(rep7.sim.tmrca.1Mb$AverageTmrca, rep_7.tmrca.1Mb$avg, method = "spearman")$estimate

# standardizing
inf.lands.1Mb.rep7$thetaC <- (inf.lands.1Mb.rep7$theta - mean(inf.lands.1Mb.rep7$theta)) / sd(inf.lands.1Mb.rep7$theta)
inf.lands.1Mb.rep7$tmrcaC <- (inf.lands.1Mb.rep7$tmrca - mean(inf.lands.1Mb.rep7$tmrca)) / sd(inf.lands.1Mb.rep7$tmrca)
inf.lands.1Mb.rep7$rhoC <- (inf.lands.1Mb.rep7$rho - mean(inf.lands.1Mb.rep7$rho)) / sd(inf.lands.1Mb.rep7$rho)

inf.lands.1Mb.rep7$bin <- 1:nrow(inf.lands.1Mb.rep7)

inf.lands.1Mb.rep7$Replicate <- 7

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.1Mb.rep7)
m.div.1Mb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.1Mb.rep7)
```

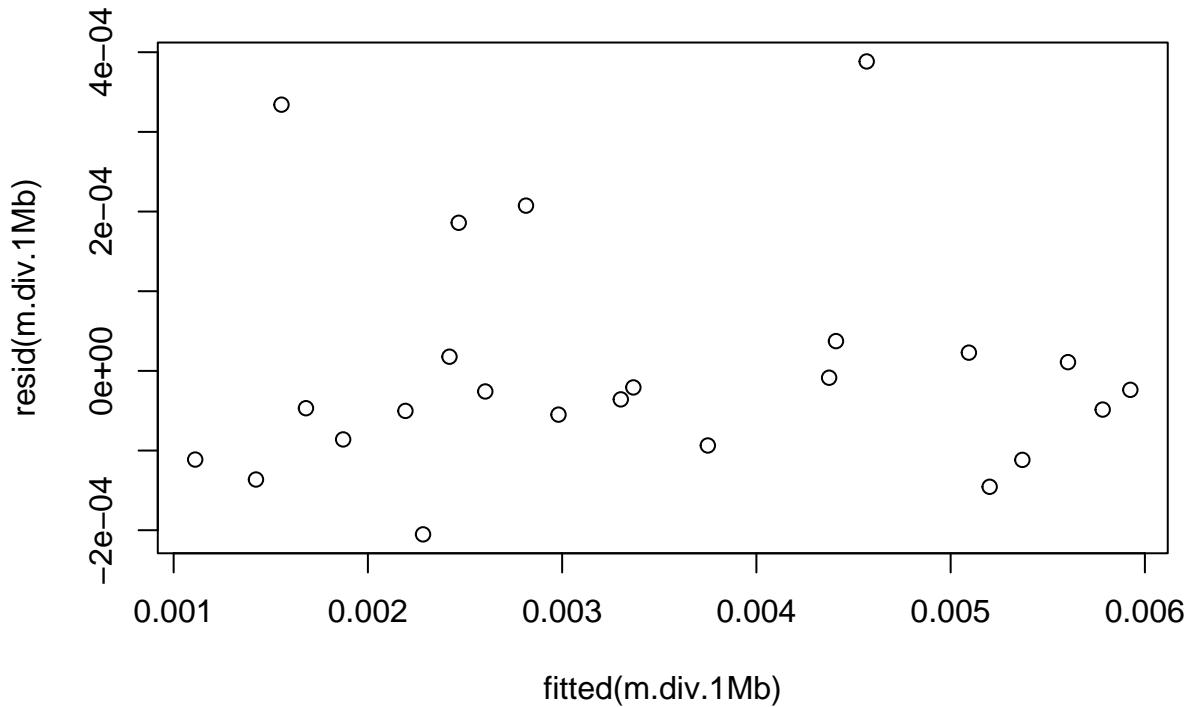
```

m.div.1Mb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.1Mb.rep7)

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

##          df      AIC
## m.div.1Mb    6 -345.4274
## m.div.1Mb.2  7 -343.4288
## m.div.1Mb.3  8 -342.2231
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```

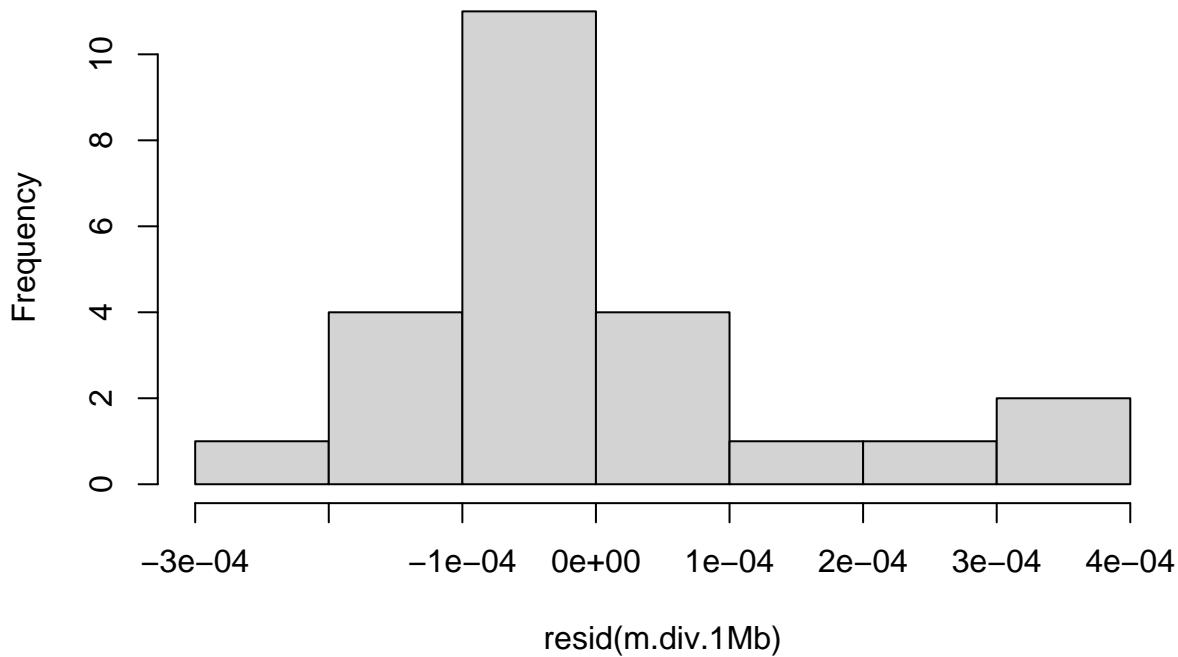
dwtest(m.div.1Mb)

##
##  Durbin-Watson test
##
##  data:  m.div.1Mb
##  DW = 2.1895, p-value = 0.5182
##  alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.1Mb)

##
##  Harrison-McCabe test
##
##  data:  m.div.1Mb
##  HMC = 0.49501, p-value = 0.476
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 = inf.lands.1Mb.rep7)
##
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -2.052e-04 -8.794e-05 -3.079e-05  1.908e-05  3.884e-04
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.351e-03 5.037e-05 66.526 < 2e-16 ***
## thetaC      1.181e-03 5.828e-05 20.259 2.52e-14 ***
## rhoC        4.537e-05 4.995e-05  0.908  0.3752  
## tmrcaC      6.409e-04 8.369e-05  7.658 3.19e-07 ***
## thetaC:tmrcaC 2.512e-04 1.342e-04   1.873  0.0766 .  
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.0001587 on 19 degrees of freedom
## Multiple R-squared:  0.9912, Adjusted R-squared:  0.9893 
## F-statistic: 534.8 on 4 and 19 DF,  p-value: < 2.2e-16

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

r2.bgs.1Mb[1, 7] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3] +
```

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

5.5.8 Replicate 8

```
rep8.sim.tmrca.1Mb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep8_w1000000.csv", sep = ",", header = T)

rep_8.pi.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep8/dm2L_bgs_rep8_5.div")
rep_8.pi.1Mb$avg <- apply(rep_8.pi.1Mb[4:ncol(rep_8.pi.1Mb)], 1, mean)
rep_8.rho.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep8/dm2L_bgs_rep8_5.rho")
rep_8.theta.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep8/dm2L_bgs_rep8_5.theta")
rep_8.tmrca.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep8/dm2L_bgs_rep8_5.tmrca")
rep_8.tmrca.1Mb$avg <- apply(rep_8.tmrca.1Mb[4:ncol(rep_8.tmrca.1Mb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.1Mb.rep8 <- as.data.frame(cbind(rep_8.pi.1Mb$avg, sim.theta.1Mb$Rate, rep8.sim.tmrca.1Mb$Average))
names(true.lands.1Mb.rep8) <- c("diversity", "theta", "tmrca", "rho")
true.lands.1Mb.rep8$thetaC <- (true.lands.1Mb.rep8$theta - mean(true.lands.1Mb.rep8$theta)) / sd(true.lands.1Mb.rep8$theta)
true.lands.1Mb.rep8$tmrcaC <- (true.lands.1Mb.rep8$tmrca - mean(true.lands.1Mb.rep8$tmrca)) / sd(true.lands.1Mb.rep8$tmrca)
true.lands.1Mb.rep8$rhoC <- (true.lands.1Mb.rep8$rho - mean(true.lands.1Mb.rep8$rho, na.rm = T)) / sd(true.lands.1Mb.rep8$rho)

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.1Mb.rep8)
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.1Mb.true[1, 8] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.bgs.1Mb.true[2, 8] <- anova.diversity$VarExp[1] * 100
r2.bgs.1Mb.true[3, 8] <- anova.diversity$VarExp[2] * 100
r2.bgs.1Mb.true[4, 8] <- anova.diversity$VarExp[3] * 100
r2.bgs.1Mb.true[5, 8] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.1Mb.rep8 <- as.data.frame(cbind(rep_8.pi.1Mb$avg, rep_8.theta.1Mb$sample_mean, rep_8.tmrca.1Mb$Average))
names(inf.lands.1Mb.rep8) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[8, 3] <- cor.test(sim.theta.1Mb$Rate, rep_8.theta.1Mb$sample_mean, method = "spearman")
cor.table.rho[8, 3] <- cor.test(sim.rho.1Mb$Rate, rep_8.rho.1Mb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[8, 3] <- cor.test(rep8.sim.tmrca.1Mb$AverageTmrca, rep_8.tmrca.1Mb$avg, method = "spearman")

# standardizing
inf.lands.1Mb.rep8$thetaC <- (inf.lands.1Mb.rep8$theta - mean(inf.lands.1Mb.rep8$theta)) / sd(inf.lands.1Mb.rep8$theta)
inf.lands.1Mb.rep8$tmrcaC <- (inf.lands.1Mb.rep8$tmrca - mean(inf.lands.1Mb.rep8$tmrca)) / sd(inf.lands.1Mb.rep8$tmrca)
inf.lands.1Mb.rep8$rhoC <- (inf.lands.1Mb.rep8$rho - mean(inf.lands.1Mb.rep8$rho)) / sd(inf.lands.1Mb.rep8$rho)

inf.lands.1Mb.rep8$bin <- 1:nrow(inf.lands.1Mb.rep8)

inf.lands.1Mb.rep8$Replicate <- 8

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.1Mb.rep8)
m.div.1Mb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.1Mb.rep8)
```

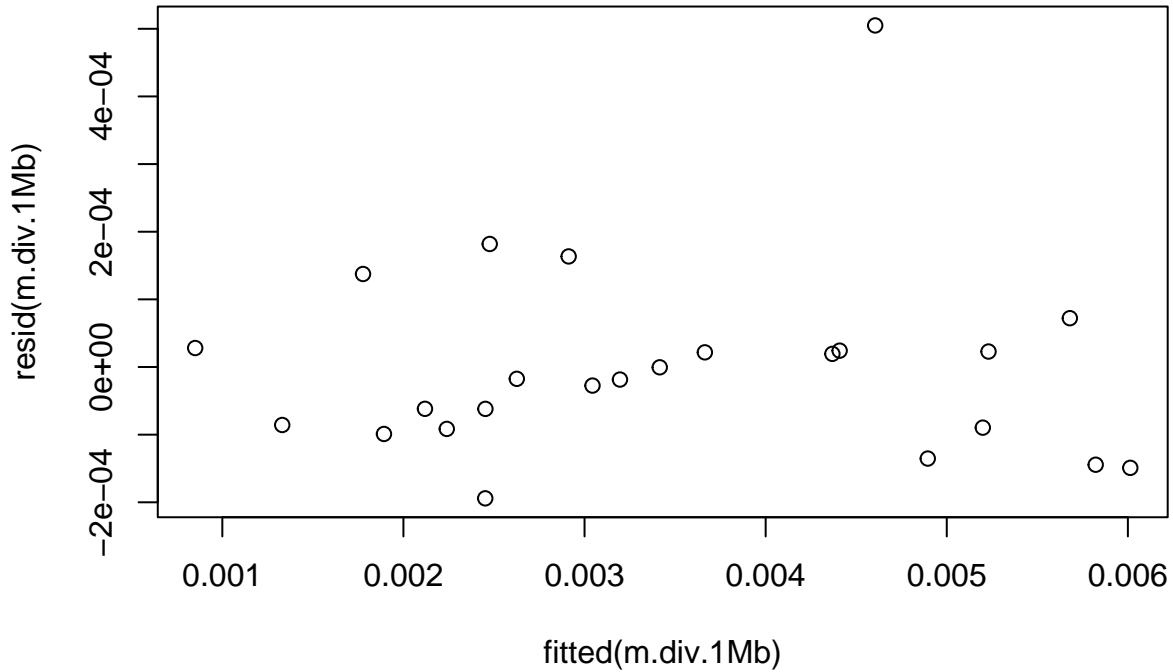
```

m.div.1Mb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.1Mb.rep8)

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

##          df      AIC
## m.div.1Mb    6 -345.1858
## m.div.1Mb.2  7 -343.4070
## m.div.1Mb.3  8 -345.1559
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```

dwtest(m.div.1Mb)

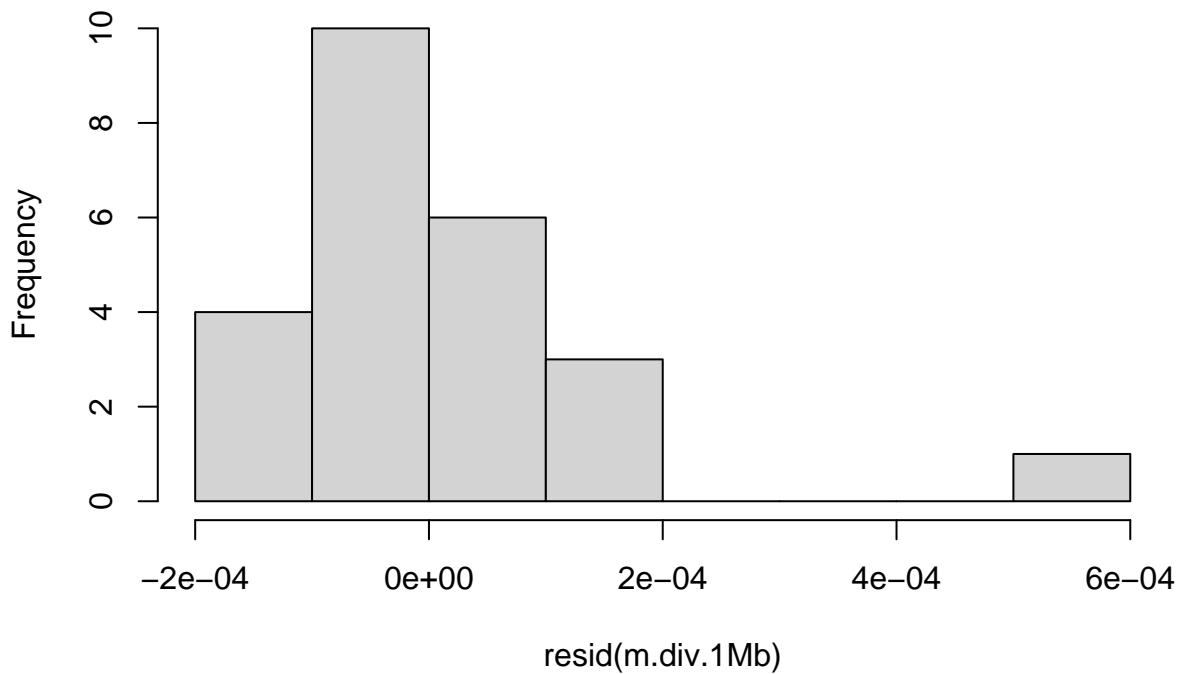
##
##  Durbin-Watson test
##
## data: m.div.1Mb
## DW = 2.3323, p-value = 0.6935
## alternative hypothesis: true autocorrelation is greater than 0

hmctest(m.div.1Mb)

##
##  Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.65145, p-value = 0.843
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 = inf.lands.1Mb.rep8)
##
## Residuals:
##       Min         1Q     Median        3Q       Max
## -1.941e-04 -9.009e-05 -1.798e-05  2.516e-05  5.049e-04
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.428e-03 3.304e-05 103.752 < 2e-16 ***
## thetaC      1.048e-03 4.550e-05 23.029 2.42e-15 ***
## rhoC       -9.642e-06 5.525e-05 -0.175  0.8633
## tmrcaC      8.935e-04 5.420e-05 16.485 1.03e-12 ***
## thetaC:tmrcaC 1.709e-04 6.009e-05   2.844  0.0104 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001595 on 19 degrees of freedom
## Multiple R-squared:  0.9909, Adjusted R-squared:  0.989
## F-statistic: 515.7 on 4 and 19 DF,  p-value: < 2.2e-16

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

r2.bgs.1Mb[1, 8] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
```

```
r2.bgs.1Mb[2, 8] <- anova.diversity$VarExp[1] * 100
r2.bgs.1Mb[3, 8] <- anova.diversity$VarExp[2] * 100
r2.bgs.1Mb[4, 8] <- anova.diversity$VarExp[3] * 100
r2.bgs.1Mb[5, 8] <- anova.diversity$VarExp[4] * 100
```

5.5.9 Replicate 9

```
rep9.sim.tmrca.1Mb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep9_w1000000.csv", sep = ",", header = T)

rep_9.pi.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep9/dm2L_bgs_rep9_5.div")
rep_9.pi.1Mb$avg <- apply(rep_9.pi.1Mb[4:ncol(rep_9.pi.1Mb)], 1, mean)
rep_9.rho.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep9/dm2L_bgs_rep9_5.rho")
rep_9.theta.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep9/dm2L_bgs_rep9_5.theta")
rep_9.tmrca.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep9/dm2L_bgs_rep9_5.tmrca")
rep_9.tmrca.1Mb$avg <- apply(rep_9.tmrca.1Mb[4:ncol(rep_9.tmrca.1Mb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.1Mb.rep9 <- as.data.frame(cbind(rep_9.pi.1Mb$avg, sim.theta.1Mb$Rate, rep9.sim.tmrca.1Mb$Average))
names(true.lands.1Mb.rep9) <- c("diversity", "theta", "tmrca", "rho")
true.lands.1Mb.rep9$thetaC <- (true.lands.1Mb.rep9$theta - mean(true.lands.1Mb.rep9$theta)) / sd(true.lands.1Mb.rep9$theta)
true.lands.1Mb.rep9$tmrcaC <- (true.lands.1Mb.rep9$tmrca - mean(true.lands.1Mb.rep9$tmrca)) / sd(true.lands.1Mb.rep9$tmrca)
true.lands.1Mb.rep9$rhoC <- (true.lands.1Mb.rep9$rho - mean(true.lands.1Mb.rep9$rho, na.rm = T)) / sd(true.lands.1Mb.rep9$rho)

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.1Mb.rep9)
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.1Mb.true[1, 9] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
r2.bgs.1Mb.true[2, 9] <- anova.diversity$VarExp[1] * 100
r2.bgs.1Mb.true[3, 9] <- anova.diversity$VarExp[2] * 100
r2.bgs.1Mb.true[4, 9] <- anova.diversity$VarExp[3] * 100
r2.bgs.1Mb.true[5, 9] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.1Mb.rep9 <- as.data.frame(cbind(rep_9.pi.1Mb$avg, rep_9.theta.1Mb$sample_mean, rep_9.tmrca.1Mb$Average))
names(inf.lands.1Mb.rep9) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[9, 3] <- cor.test(sim.theta.1Mb$Rate, rep_9.theta.1Mb$sample_mean, method = "spearman")
cor.table.rho[9, 3] <- cor.test(sim.rho.1Mb$Rate, rep_9.rho.1Mb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[9, 3] <- cor.test(rep9.sim.tmrca.1Mb$AverageTmrca, rep_9.tmrca.1Mb$avg, method = "spearman")

# standardizing
inf.lands.1Mb.rep9$thetaC <- (inf.lands.1Mb.rep9$theta - mean(inf.lands.1Mb.rep9$theta)) / sd(inf.lands.1Mb.rep9$theta)
inf.lands.1Mb.rep9$tmrcaC <- (inf.lands.1Mb.rep9$tmrca - mean(inf.lands.1Mb.rep9$tmrca)) / sd(inf.lands.1Mb.rep9$tmrca)
inf.lands.1Mb.rep9$rhoC <- (inf.lands.1Mb.rep9$rho - mean(inf.lands.1Mb.rep9$rho)) / sd(inf.lands.1Mb.rep9$rho)

inf.lands.1Mb.rep9$bin <- 1:nrow(inf.lands.1Mb.rep9)

inf.lands.1Mb.rep9$Replicate <- 9

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.1Mb.rep9)
m.div.1Mb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.1Mb.rep9)
```

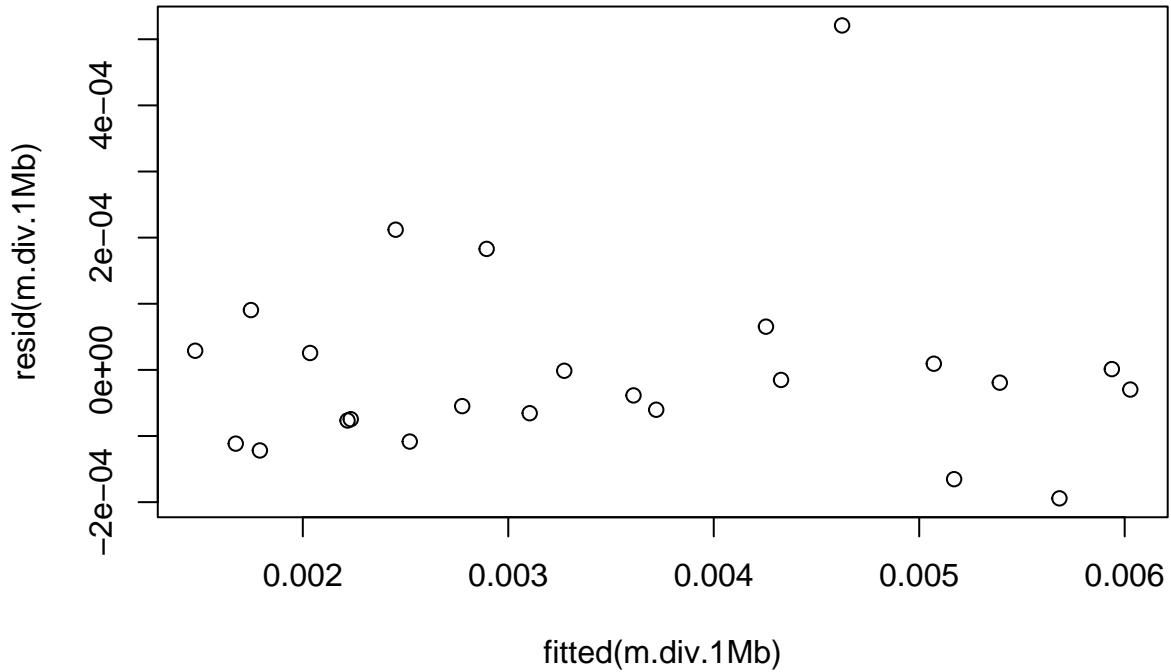
```

m.div.1Mb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.1Mb.rep9)

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

##          df      AIC
## m.div.1Mb    6 -344.7058
## m.div.1Mb.2  7 -345.4784
## m.div.1Mb.3  8 -344.8491
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```

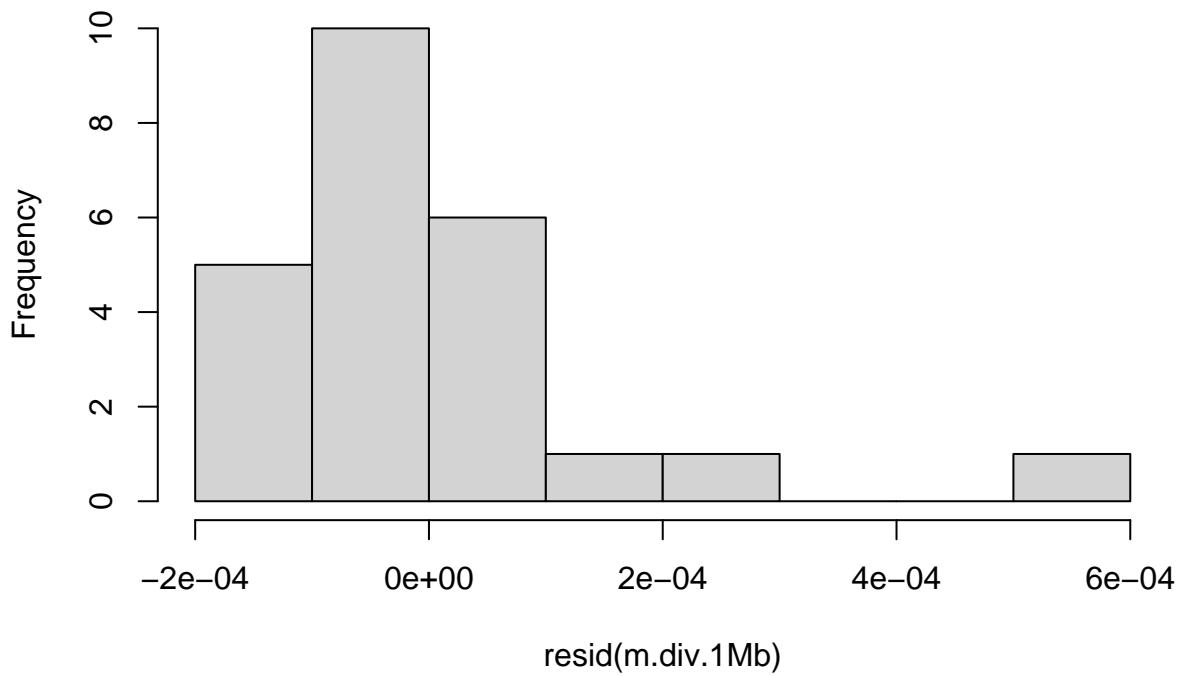
dwtest(m.div.1Mb)

##
##  Durbin-Watson test
##
## data: m.div.1Mb
## DW = 2.6206, p-value = 0.8798
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.1Mb)

##
##  Harrison-McCabe test
##
## data: m.div.1Mb
## HMC = 0.64661, p-value = 0.858
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 = inf.lands.1Mb.rep9)
##
## Residuals:
##       Min         1Q     Median        3Q       Max
## -1.942e-04 -7.503e-05 -2.451e-05  2.644e-05  5.209e-04
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.453e-03 3.751e-05 92.060 < 2e-16 ***
## thetaC      1.053e-03 4.624e-05 22.783 2.94e-15 ***
## rhoC       -9.680e-06 5.351e-05 -0.181  0.8584
## tmrcaC      7.147e-04 4.670e-05 15.304 3.87e-12 ***
## thetaC:tmrcaC 1.586e-04 6.041e-05  2.626   0.0166 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001611 on 19 degrees of freedom
## Multiple R-squared:  0.9903, Adjusted R-squared:  0.9883
## F-statistic: 486.3 on 4 and 19 DF,  p-value: < 2.2e-16

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

r2.bgs.1Mb[1, 9] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
```

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

5.5.10 Replicate 10

```
rep10.sim.tmrca.1Mb <- read.table("dm_bgs_sims/10r/dm2L_bgs_rep10_w1000000.csv", sep = ",", header = T)

rep_10.pi.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep10/dm2L_bgs_rep10_5.csv")
rep_10.pi.1Mb$avg <- apply(rep_10.pi.1Mb[4:ncol(rep_10.pi.1Mb)], 1, mean)
rep_10.rho.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep10/dm2L_bgs_rep10_5.csv")
rep_10.theta.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep10/dm2L_bgs_rep10_5.csv")
rep_10.tmrca.1Mb <- read.table("dm_bgs_sims/10r/NonHomogeneous/iSMC/NonHomogeneous/rep10/dm2L_bgs_rep10_5.csv")
rep_10.tmrca.1Mb$avg <- apply(rep_10.tmrca.1Mb[4:ncol(rep_10.tmrca.1Mb)], 1, mean)

# first the linear modeling with the true (simulated) landscapes
true.lands.1Mb.rep10 <- as.data.frame(cbind(rep_10.pi.1Mb$avg, sim.theta.1Mb$Rate, rep10.sim.tmrca.1Mb$Rate))
names(true.lands.1Mb.rep10) <- c("diversity", "theta", "tmrca", "rho")
true.lands.1Mb.rep10$thetaC <- (true.lands.1Mb.rep10$theta - mean(true.lands.1Mb.rep10$theta)) / sd(true.lands.1Mb.rep10$theta)
true.lands.1Mb.rep10$tmrcaC <- (true.lands.1Mb.rep10$tmrca - mean(true.lands.1Mb.rep10$tmrca)) / sd(true.lands.1Mb.rep10$tmrca)
true.lands.1Mb.rep10$rhoC <- (true.lands.1Mb.rep10$rho - mean(true.lands.1Mb.rep10$rho, na.rm = T)) / sd(true.lands.1Mb.rep10$rho)

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = true.lands.1Mb.rep10)
anova.diversity <- Anova(m.div.1Mb)
apiss <- anova.diversity$"Sum Sq"
anova.diversity$VarExp <- apiss / sum(apiss)
r2.bgs.1Mb.true[1, 10] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3]) * 100
r2.bgs.1Mb.true[2, 10] <- anova.diversity$VarExp[1] * 100
r2.bgs.1Mb.true[3, 10] <- anova.diversity$VarExp[2] * 100
r2.bgs.1Mb.true[4, 10] <- anova.diversity$VarExp[3] * 100
r2.bgs.1Mb.true[5, 10] <- anova.diversity$VarExp[4] * 100

# now the lm with inferred landscapes
inf.lands.1Mb.rep10 <- as.data.frame(cbind(rep_10.pi.1Mb$avg, rep_10.theta.1Mb$sample_mean, rep_10.tmrca.1Mb$sample_mean))
names(inf.lands.1Mb.rep10) <- c("diversity", "theta", "tmrca", "rho")

cor.table.theta[10, 3] <- cor.test(sim.theta.1Mb$Rate, rep_10.theta.1Mb$sample_mean, method = "spearman")
cor.table.rho[10, 3] <- cor.test(sim.rho.1Mb$Rate, rep_10.rho.1Mb$sample_mean, method = "spearman")$estimate
cor.table.tmrca[10, 3] <- cor.test(rep10.sim.tmrca.1Mb$AverageTmrca, rep_10.tmrca.1Mb$avg, method = "spearman")$estimate

# standardizing
inf.lands.1Mb.rep10$thetaC <- (inf.lands.1Mb.rep10$theta - mean(inf.lands.1Mb.rep10$theta)) / sd(inf.lands.1Mb.rep10$theta)
inf.lands.1Mb.rep10$tmrcaC <- (inf.lands.1Mb.rep10$tmrca - mean(inf.lands.1Mb.rep10$tmrca)) / sd(inf.lands.1Mb.rep10$tmrca)
inf.lands.1Mb.rep10$rhoC <- (inf.lands.1Mb.rep10$rho - mean(inf.lands.1Mb.rep10$rho)) / sd(inf.lands.1Mb.rep10$rho)

inf.lands.1Mb.rep10$bin <- 1:nrow(inf.lands.1Mb.rep10)

inf.lands.1Mb.rep10$Replicate <- 10

m.div.1Mb <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC, data = inf.lands.1Mb.rep10)
m.div.1Mb.2 <- lm(diversity ~ thetaC + rhoC + tmrcaC + thetaC:tmrcaC + rhoC:tmrcaC, data = inf.lands.1Mb.rep10)
```

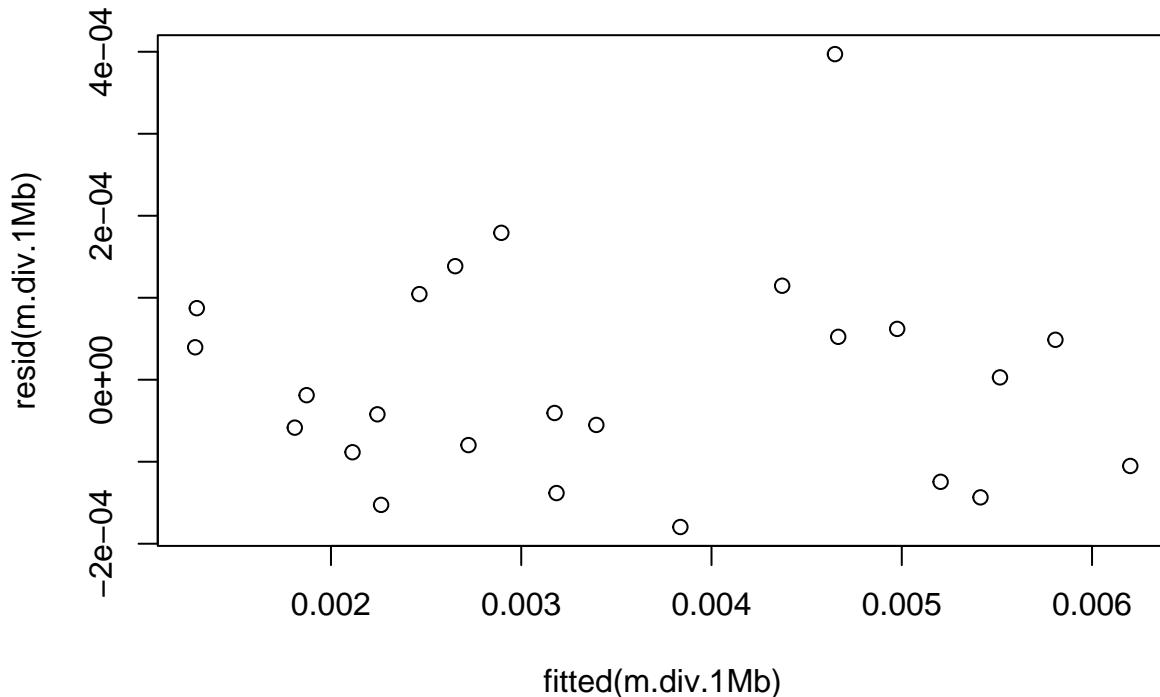
```

m.div.1Mb.3 <- lm(diversity ~ (thetaC + rhoC + tmrcaC) ^ 2, data = inf.lands.1Mb.rep10)

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

##          df      AIC
## m.div.1Mb     6 -349.8858
## m.div.1Mb.2   7 -353.7635
## m.div.1Mb.3   8 -351.7726
plot(resid(m.div.1Mb)~fitted(m.div.1Mb))

```



```

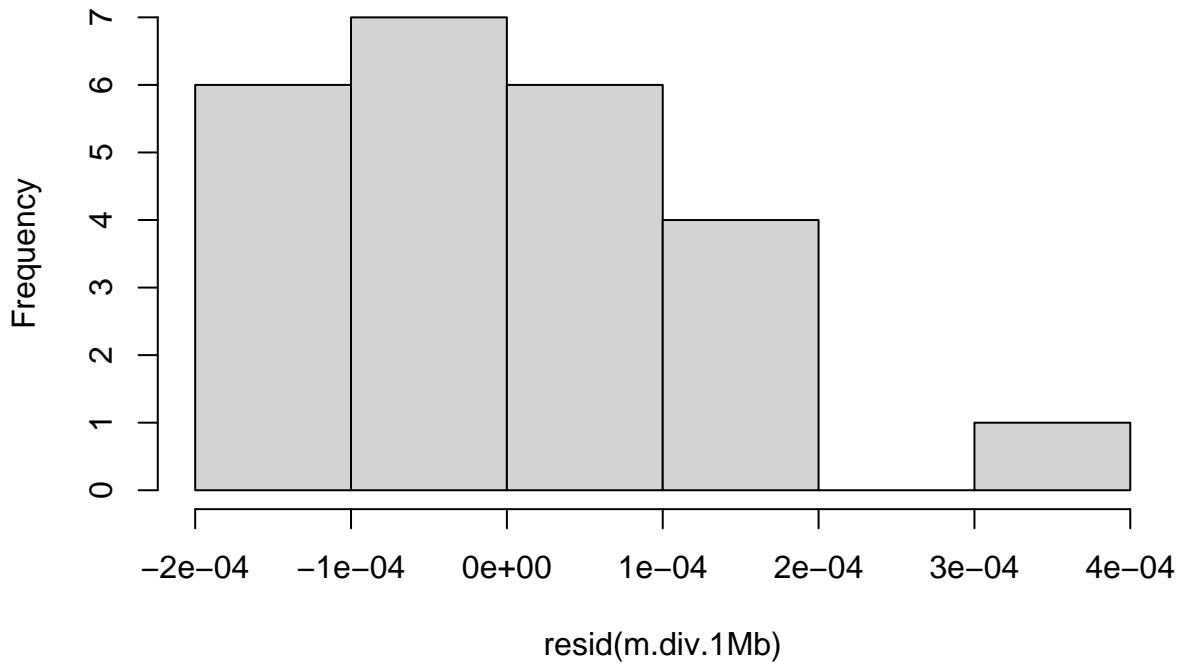
dwtest(m.div.1Mb)

##
##  Durbin-Watson test
##
## data: m.div.1Mb
## DW = 2.0448, p-value = 0.4006
## alternative hypothesis: true autocorrelation is greater than 0
hmctest(m.div.1Mb)

##
##  Harrison-M McCabe test
##
## data: m.div.1Mb
## HMC = 0.61476, p-value = 0.77
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 = inf.lands.1Mb.rep10)
##
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -1.796e-04 -9.261e-05 -2.968e-05  6.834e-05  3.971e-04
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.457e-03 3.521e-05 98.178 < 2e-16 ***
## thetaC      1.124e-03 4.650e-05 24.171 9.92e-16 ***
## rhoC        7.688e-05 5.132e-05  1.498  0.1505    
## tmrcaC      6.577e-04 4.431e-05 14.844 6.62e-12 ***
## thetaC:tmrcaC 1.995e-04 8.776e-05   2.273  0.0348 *  
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001446 on 19 degrees of freedom
## Multiple R-squared:  0.9925, Adjusted R-squared:  0.9909 
## F-statistic: 624.7 on 4 and 19 DF,  p-value: < 2.2e-16

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

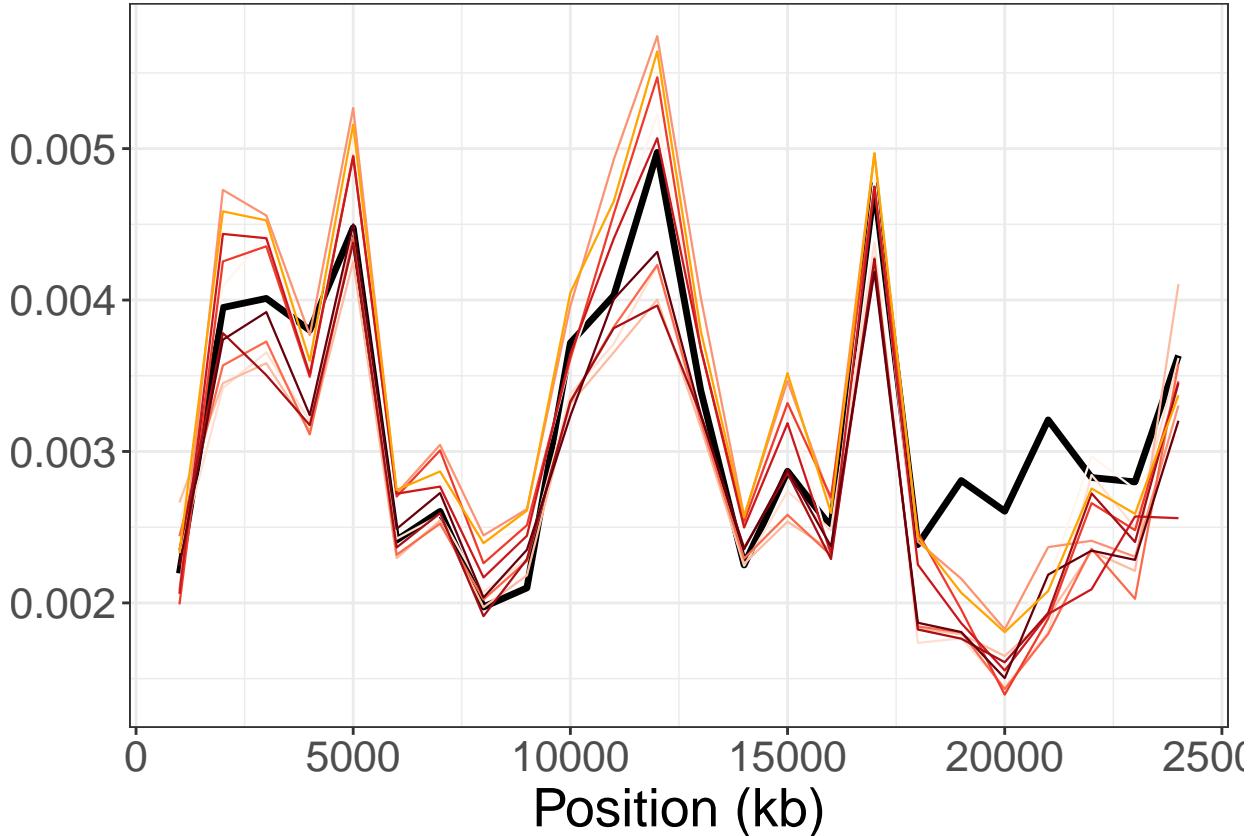
r2.bgs.1Mb[1, 10] <- (anova.diversity$VarExp[1] + anova.diversity$VarExp[2] + anova.diversity$VarExp[3])
```

```
r2.bgs.1Mb[2, 10] <- anova.diversity$VarExp[1] * 100
r2.bgs.1Mb[3, 10] <- anova.diversity$VarExp[2] * 100
r2.bgs.1Mb[4, 10] <- anova.diversity$VarExp[3] * 100
r2.bgs.1Mb[5, 10] <- anova.diversity$VarExp[4] * 100
```

5.6 all replicates:

```
theta.plot <- as.data.frame(cbind(1:nrow(sim.theta.1Mb),
                                    sim.theta.1Mb$Rate * 1e-7 * 3e+4,
                                    rep_1.theta.1Mb$sample_mean,
                                    rep_2.theta.1Mb$sample_mean,
                                    rep_3.theta.1Mb$sample_mean,
                                    rep_4.theta.1Mb$sample_mean,
                                    rep_5.theta.1Mb$sample_mean,
                                    rep_6.theta.1Mb$sample_mean,
                                    rep_7.theta.1Mb$sample_mean,
                                    rep_8.theta.1Mb$sample_mean,
                                    rep_9.theta.1Mb$sample_mean,
                                    rep_10.theta.1Mb$sample_mean))

names(theta.plot) <- c("bin", "sim", "reps")
molten.theta <- melt(theta.plot, id.vars = "bin")
theta.map.1Mb.bgs <- ggplot(data = molten.theta, aes(x = bin * 1000, y = value, colour = variable)) +
  theta.map.1Mb.bgs + geom_line(data = molten.theta, aes(size = variable)) + scale_size()
theta.map.1Mb.bgs <- theta.map.1Mb.bgs + scale_color_manual(values = c("black", brewer.pal(n = 9, name =
theta.map.1Mb.bgs <- theta.map.1Mb.bgs + scale_x_continuous(breaks = pretty_breaks()) + scale_y_continuous()
theta.map.1Mb.bgs <- theta.map.1Mb.bgs + labs(title = NULL, x = "Position (kb)", y = NULL)
theta.map.1Mb.bgs <- theta.map.1Mb.bgs + theme(text = element_text(size = 20), axis.title.x = element_text(),
theta.map.1Mb.bgs
```



6 Plots

```

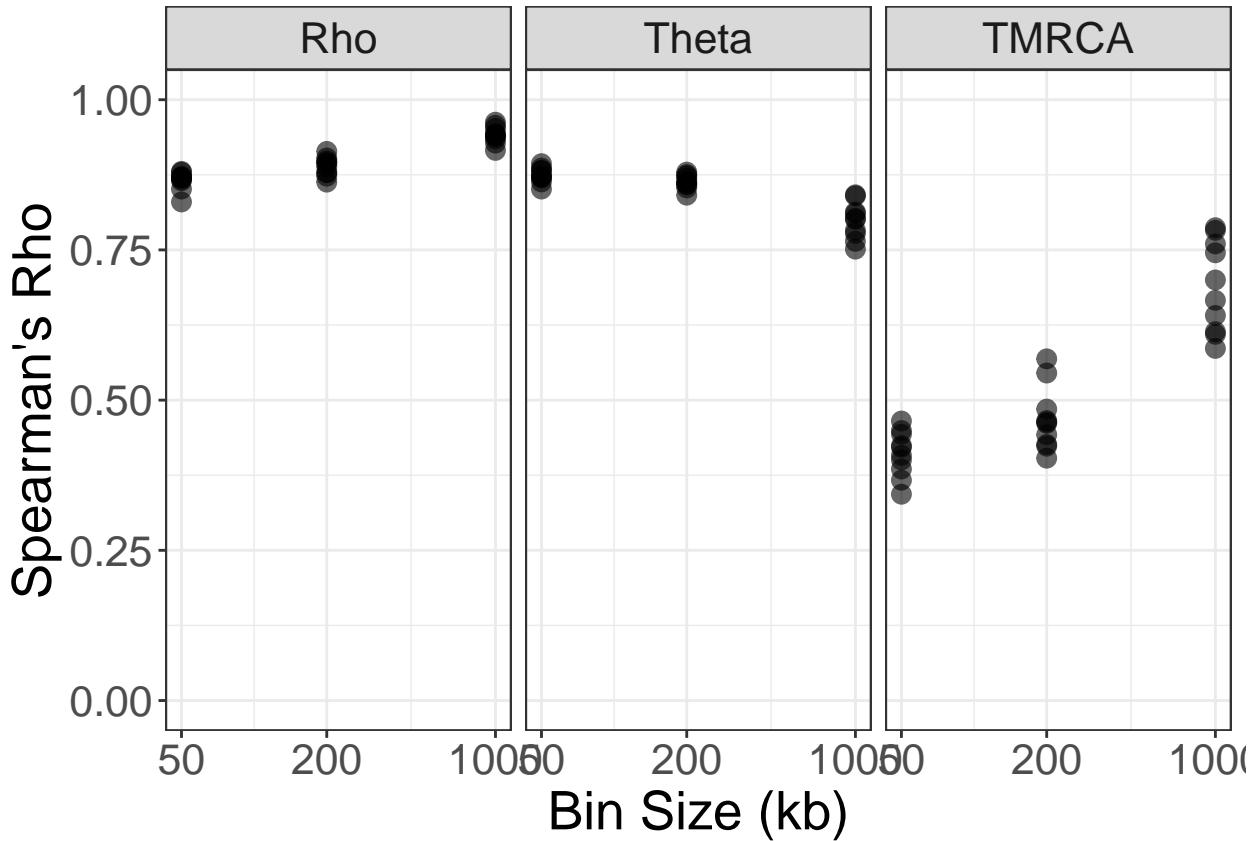
scale.2d <- function(x) sprintf("%.2f", x)

write.table(cor.table.theta, "cor.mutation.bgs.tsv", quote = F, sep = "\t")
write.table(cor.table.tmrca, "cor.tmrca.bgs.tsv", quote = F, sep = "\t")
write.table(cor.table.rho, "cor.rho.bgs.tsv", quote = F, sep = "\t")

cor.table.comb <- rbind.data.frame(t(cor.table.theta), t(cor.table.rho), t(cor.table.tmrca))
cor.table.comb$bin.size <- rep(c(50, 200, 1000), 3)
cor.table.comb$rate <- c(rep("Theta", 3), rep("Rho", 3), rep("TMRCA", 3))

molten.cor <- melt(cor.table.comb, id.vars = c("bin.size", "rate"))
cor.bgs.plot <- ggplot(data = molten.cor, aes(x = bin.size, y = value))
cor.bgs.plot <- cor.bgs.plot + geom_point(size=3, alpha = 0.6) + facet_wrap(~rate)
cor.bgs.plot <- cor.bgs.plot + scale_x_continuous(breaks = c(50, 200, 1000), trans="log10")
cor.bgs.plot <- cor.bgs.plot + scale_y_continuous(breaks = pretty_breaks())
cor.bgs.plot <- cor.bgs.plot + labs(title = NULL, x = "Bin Size (kb)", y = "Spearman's Rho")
cor.bgs.plot <- cor.bgs.plot + theme_bw() + ylim(0, 1)
cor.bgs.plot <- cor.bgs.plot + theme(axis.title = element_text(size = 20), axis.text = element_text(size = 16),
                                         strip.text.x = element_text(size = 16), legend.position = "bottom")
cor.bgs.plot

```



```

ggsave("Figure6.pdf", cor.bgs.plot, device = "pdf", height = 5, width = 12)

# true (simulated) landscapes
r2.bgs.50kb.true$average <- apply(r2.bgs.50kb.true, 1, mean)
r2.bgs.200kb.true$average <- apply(r2.bgs.200kb.true, 1, mean)
r2.bgs.1Mb.true$average <- apply(r2.bgs.1Mb.true, 1, mean)

r2.true.avg <- rbind.data.frame(r2.bgs.50kb.true$average, r2.bgs.200kb.true$average, r2.bgs.1Mb.true$average)
colnames(r2.true.avg) <- c("Total", "Theta", "Rho", "TMRCA", "Theta:TMRCA")
r2.true.avg$bin.size <- c(50, 200, 1000)

# inferred landscapes
r2.bgs.50kb$average <- apply(r2.bgs.50kb, 1, mean)
r2.bgs.200kb$average <- apply(r2.bgs.200kb, 1, mean)
r2.bgs.1Mb$average <- apply(r2.bgs.1Mb, 1, mean)

r2.inf.avg <- rbind.data.frame(r2.bgs.50kb$average, r2.bgs.200kb$average, r2.bgs.1Mb$average, make.row=rbind)
colnames(r2.inf.avg) <- c("Total", "Theta", "Rho", "TMRCA", "Theta:TMRCA")
r2.inf.avg$bin.size <- c(50, 200, 1000)

r2.tab.comb <- rbind.data.frame(r2.dm.tab, r2.true.avg, r2.inf.avg)
r2.tab.comb$type <- c(rep("iSMC-inferred (Drosophila)", 3), rep("True landscapes (simulations)", 3), rep("True landscapes (estimates)", 3))

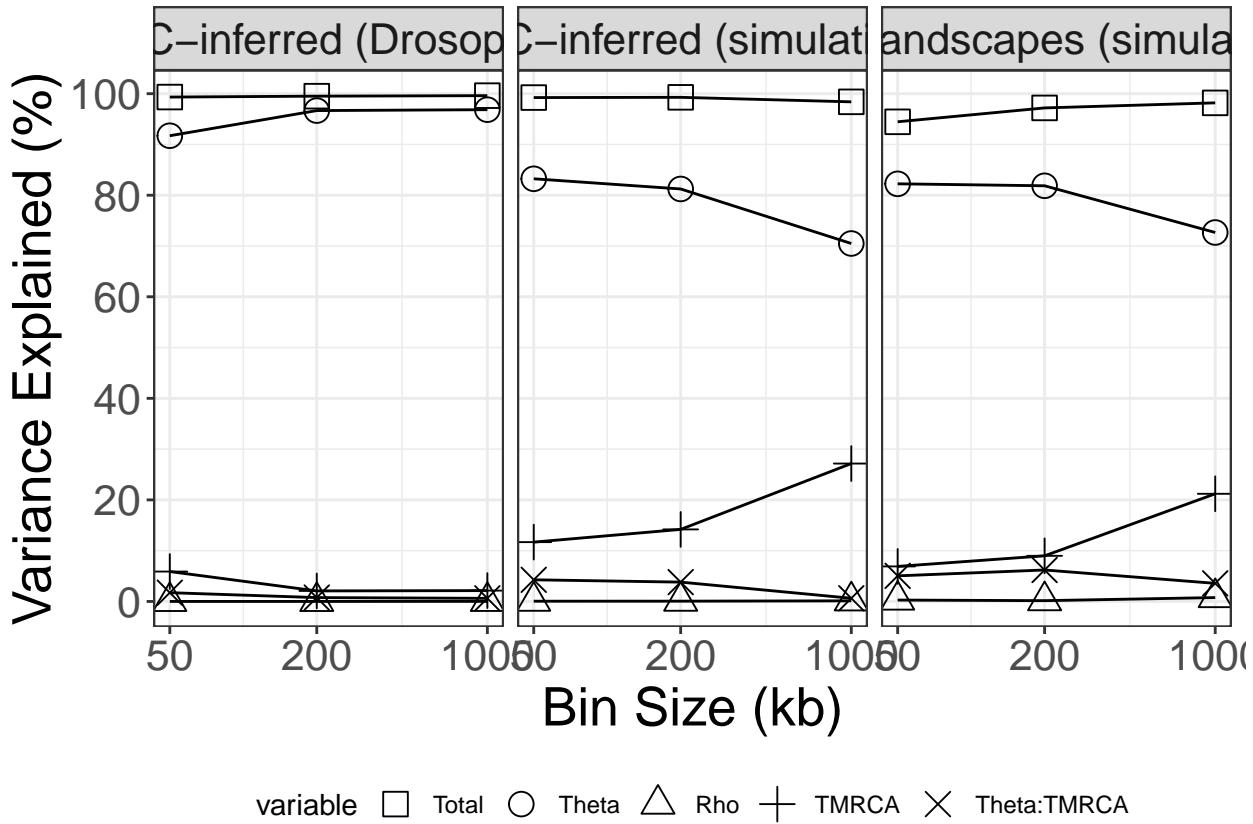
molten.r2 <- melt(r2.tab.comb, id.vars = c("bin.size", "type"))
r2.bgs.plot <- ggplot(data = molten.r2, aes(x = bin.size, y = value, shape = variable))
r2.bgs.plot <- r2.bgs.plot + geom_line(data = molten.r2)
r2.bgs.plot <- r2.bgs.plot + geom_point(aes(shape = variable), size = 4) + facet_wrap(~type)

```

```

r2.bgs.plot <- r2.bgs.plot + scale_x_continuous(breaks = c(50, 200, 1000), trans="log10")
r2.bgs.plot <- r2.bgs.plot + scale_y_continuous(breaks = pretty_breaks())
r2.bgs.plot <- r2.bgs.plot + scale_shape_manual(values = c(0, 1, 2, 3, 4))
r2.bgs.plot <- r2.bgs.plot + labs(title = NULL, x = "Bin Size (kb)", y = "Variance Explained (%)")
r2.bgs.plot <- r2.bgs.plot + theme_bw()
r2.bgs.plot <- r2.bgs.plot + theme(axis.title = element_text(size = 20), axis.text = element_text(size = 16),
                                     strip.text.x = element_text(size = 16), legend.position = "bottom")
r2.bgs.plot

```



```

variable □ Total ○ Theta △ Rho + TMRCA × Theta:TMRCA

# combine coal and BGS R2 plots
r2.comb.plot <- plot_grid(r2.coal.plot, r2.bgs.plot, nrow = 2, ncol = 1, labels = "AUTO")
cowplot::save_plot("Figure4.pdf", r2.comb.plot, device = "pdf", base_height = 9, base_width = 12)

# bench theta
theta.maps.bgs <- plot_grid(theta.map.50kb.bgs, theta.map.200kb.bgs, theta.map.1Mb.bgs, nrow = 3, ncol = 1)

theta.maps.comb <- plot_grid(theta.maps.coal, theta.maps.bgs, labels = "AUTO", nrow = 1, ncol = 2)
save_plot("Figure2.pdf", theta.maps.comb, base_height = 16, base_width = 12)

# visualise true join landscapes
dm_2L_exome <- read.table("dm_bgs_sims/GeneAnnotations/exons_chr2L.csv", header = F, sep = ",")
names(dm_2L_exome) <- c("exon_start", "exon_end")

genes_barcode <- ggplot(data = dm_2L_exome) + geom_segment(aes(x = exon_start, xend = exon_end, size = 1))

## 50kb
theta.land.50kb <- cbind.data.frame(1:nrow(sim.theta.50kb), sim.theta.50kb$Rate * 4 * 1e+4 * 1e-7)

```

```

names(theta.land.50kb) <- c("bin", "rate")

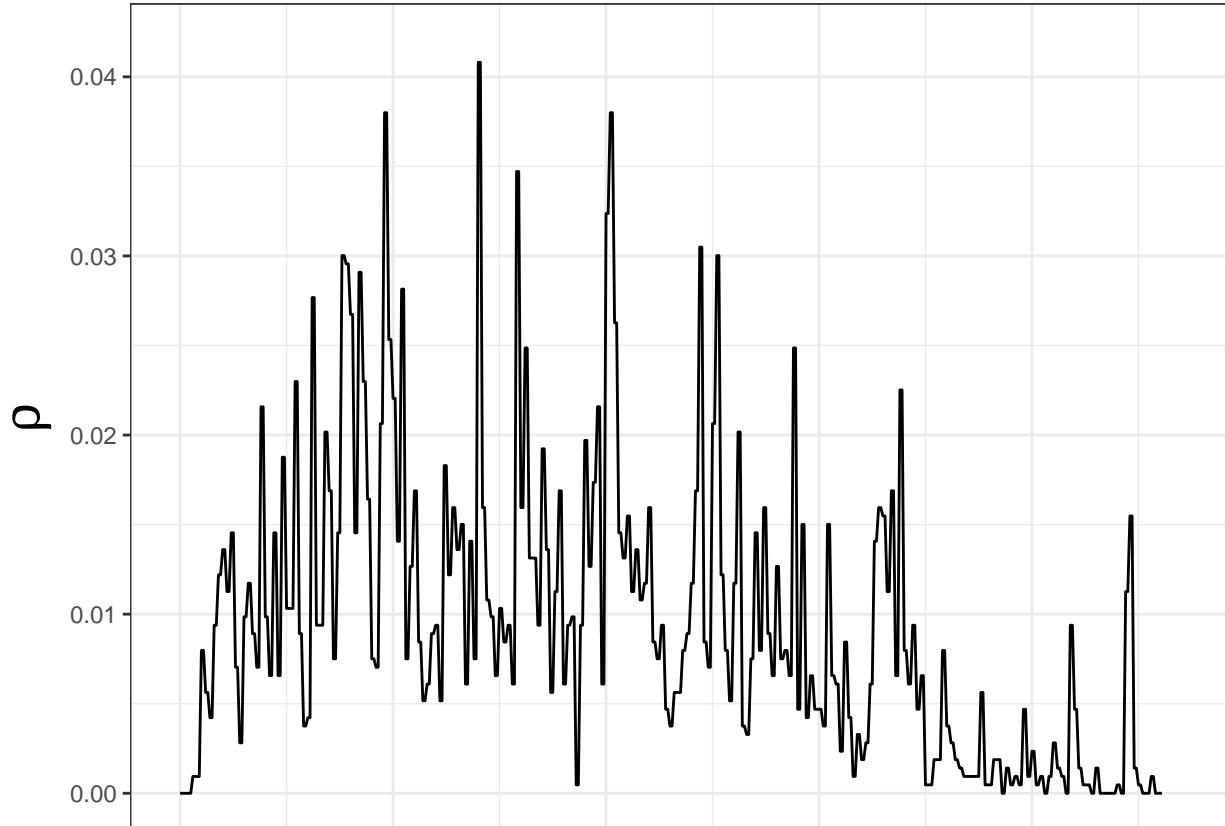
rho.land.50kb <- cbind.data.frame(1:nrow(sim.rho.50kb), sim.rho.50kb$Rate * 1e-7 * 4 * 1e+4)
names(rho.land.50kb) <- c("bin", "rate")

pi.lands.50kb <- cbind.data.frame(1:nrow(sim.theta.50kb), rep_1.pi.50kb$avg, rep_2.pi.50kb$avg, rep_3.pi.50kb$avg)
names(pi.lands.50kb) <- c("bin", "reps")

tau.lands.50kb <- cbind.data.frame(1:nrow(sim.theta.50kb), rep1.sim.tmrca.50kb$AverageTmrca, rep2.sim.tmrca.50kb$AverageTmrca)
names(tau.lands.50kb) <- c("bin", "reps")

rho.map.50kb <- ggplot(data = rho.land.50kb, aes(x = (bin - 1) * 50, y = rate))
rho.map.50kb <- rho.map.50kb + geom_line(data = rho.land.50kb) + theme_bw()
rho.map.50kb <- rho.map.50kb + scale_y_continuous(breaks = pretty_breaks(), labels = scale.2d, limits = c(0, 0.04))
rho.map.50kb <- rho.map.50kb + labs(title = NULL, x = NULL, y = expression(rho))
rho.map.50kb <- rho.map.50kb + theme(axis.title.x=element_blank(),
                                      axis.text.x=element_blank(),
                                      axis.ticks.x=element_blank(),
                                      axis.title.y=element_text(size = 20), legend.position = "none")
rho.map.50kb

```

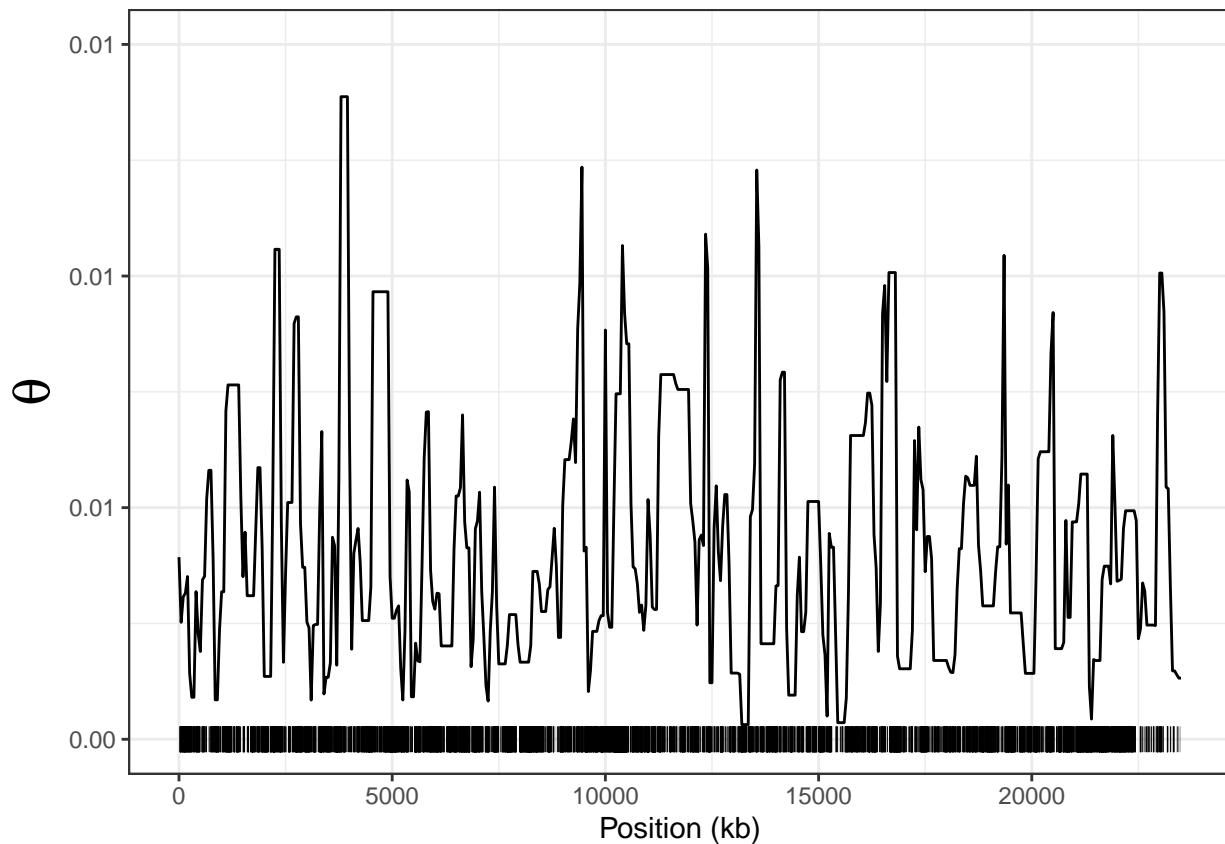


```

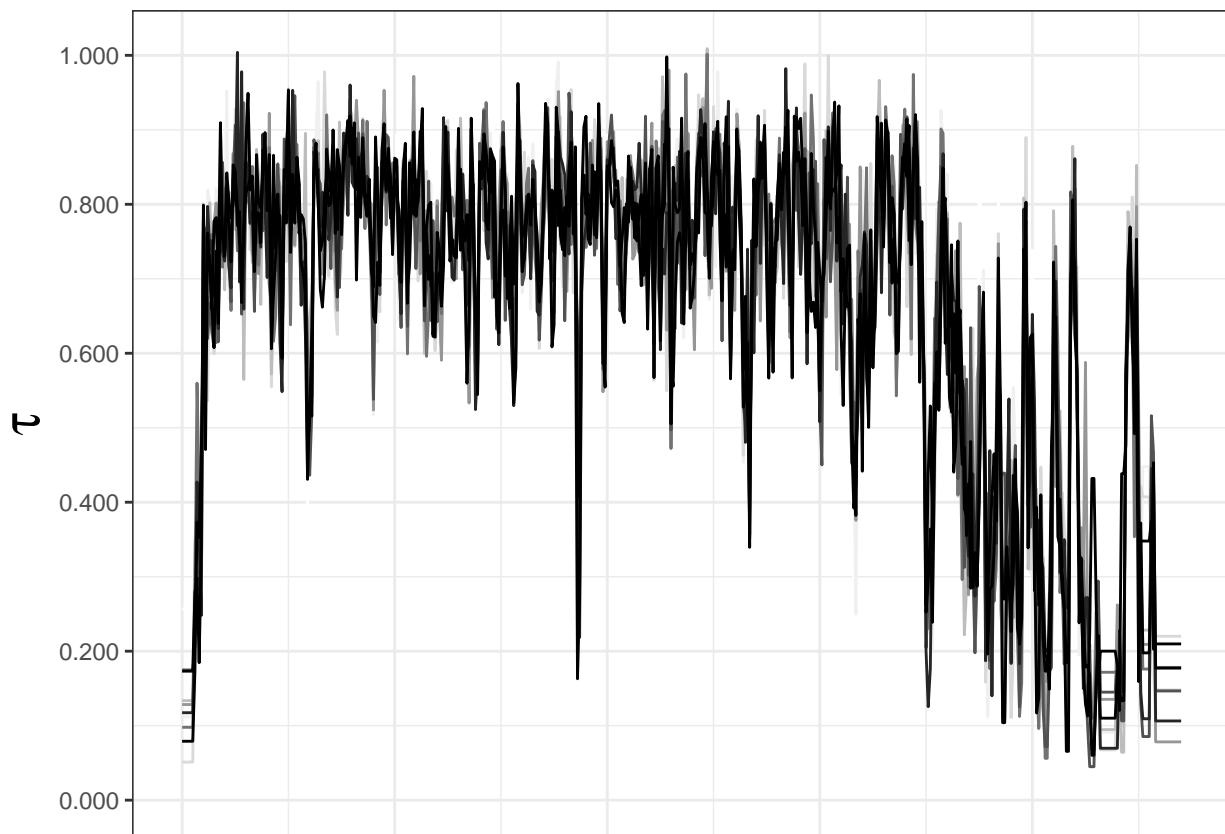
theta.map.50kb <- ggplot(data = theta.land.50kb, aes(x = (bin - 1) * 50, y = rate))
theta.map.50kb <- theta.map.50kb + geom_line(data = theta.land.50kb) + theme_bw()
theta.map.50kb <- theta.map.50kb + scale_y_continuous(breaks = pretty_breaks(), labels = scale.2d, limits = c(0, 0.04))
theta.map.50kb <- theta.map.50kb + labs(title = NULL, x = "Position (kb)", y = expression(theta))
theta.map.50kb <- theta.map.50kb + geom_segment(data = dm_2L_exome, aes(x = exon_start / 1000, xend =

```

```
theta.map.50kb <- theta.map.50kb + theme(legend.position = "none", axis.title.y=element_text(size = 20))
theta.map.50kb
```



```
molten.tau <- melt(tau.lands.50kb, id.vars = "bin")
tau.map.50kb <- ggplot(data = molten.tau, aes(x = (bin -1) * 50, y = value / (4 * 1e+4), colour = variable))
tau.map.50kb <- tau.map.50kb + geom_line(data = molten.tau) + theme_bw()
tau.map.50kb <- tau.map.50kb + scale_y_continuous(breaks = pretty_breaks(), labels = scale.3d, limits = c(0, 1))
tau.map.50kb <- tau.map.50kb + labs(title = NULL, x = NULL, y = expression(tau))
tau.map.50kb <- tau.map.50kb + scale_color_manual(values = c(brewer.pal(n = 9, name = "Greys"), "Black"))
tau.map.50kb <- tau.map.50kb + theme(axis.title.x=element_blank(),
                                      axis.text.x=element_blank(),
                                      axis.ticks.x=element_blank(),
                                      axis.title.y=element_text(size = 20), legend.position = "none")
tau.map.50kb
```

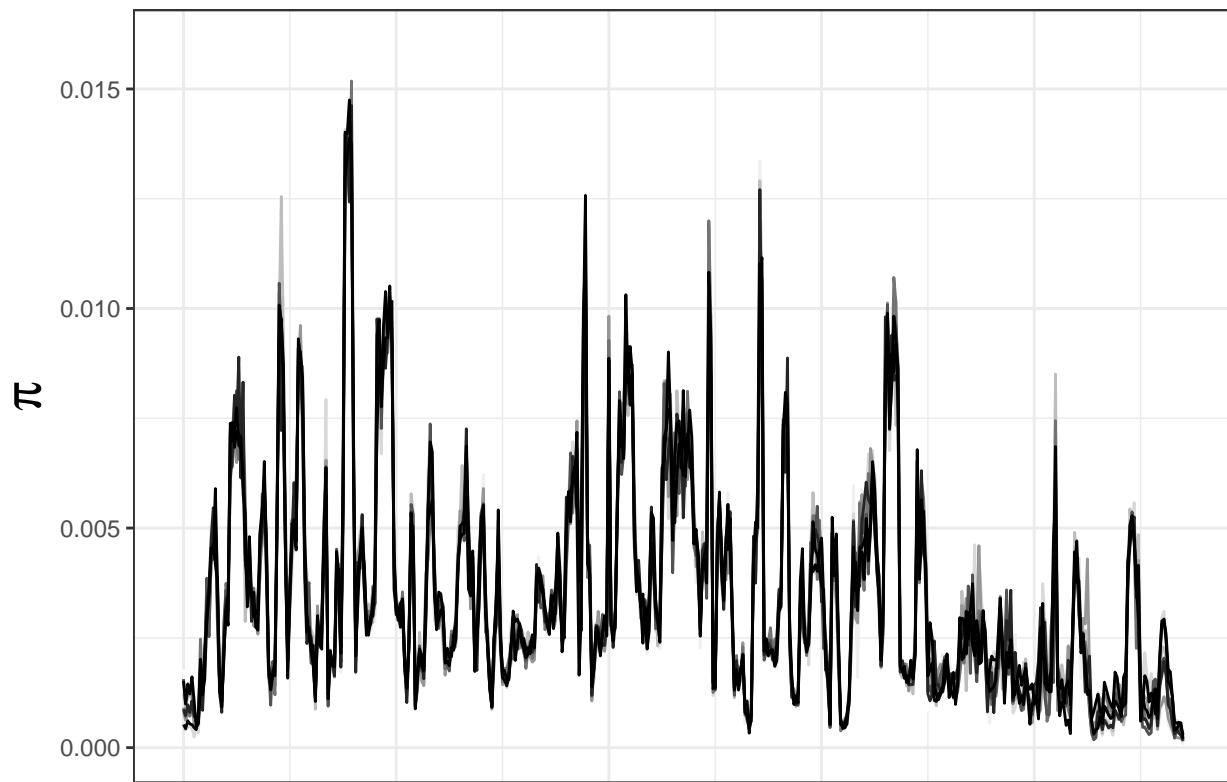


```

molten.pi <- melt(pi.lands.50kb, id.vars = "bin")
pi.map.50kb <- ggplot(data = molten.pi, aes(x = (bin -1) * 50, y = value, colour = variable))
pi.map.50kb <- pi.map.50kb + geom_line(data = molten.pi) + theme_bw()
pi.map.50kb <- pi.map.50kb + scale_y_continuous(breaks = pretty_breaks(), labels = scale.3d, limits = c(0, 1))
pi.map.50kb <- pi.map.50kb + labs(title = "50 kb", x = NULL, y = expression(pi))
pi.map.50kb <- pi.map.50kb + scale_color_manual(values = c(brewer.pal(n = 9, name = "Greys"), "Black"))
pi.map.50kb <- pi.map.50kb + theme(axis.title.x=element_blank(),
                                      axis.text.x=element_blank(),
                                      axis.ticks.x=element_blank(),
                                      axis.title.y=element_text(size = 20),
                                      plot.title = element_text(size = 16), legend.position = "none")
pi.map.50kb

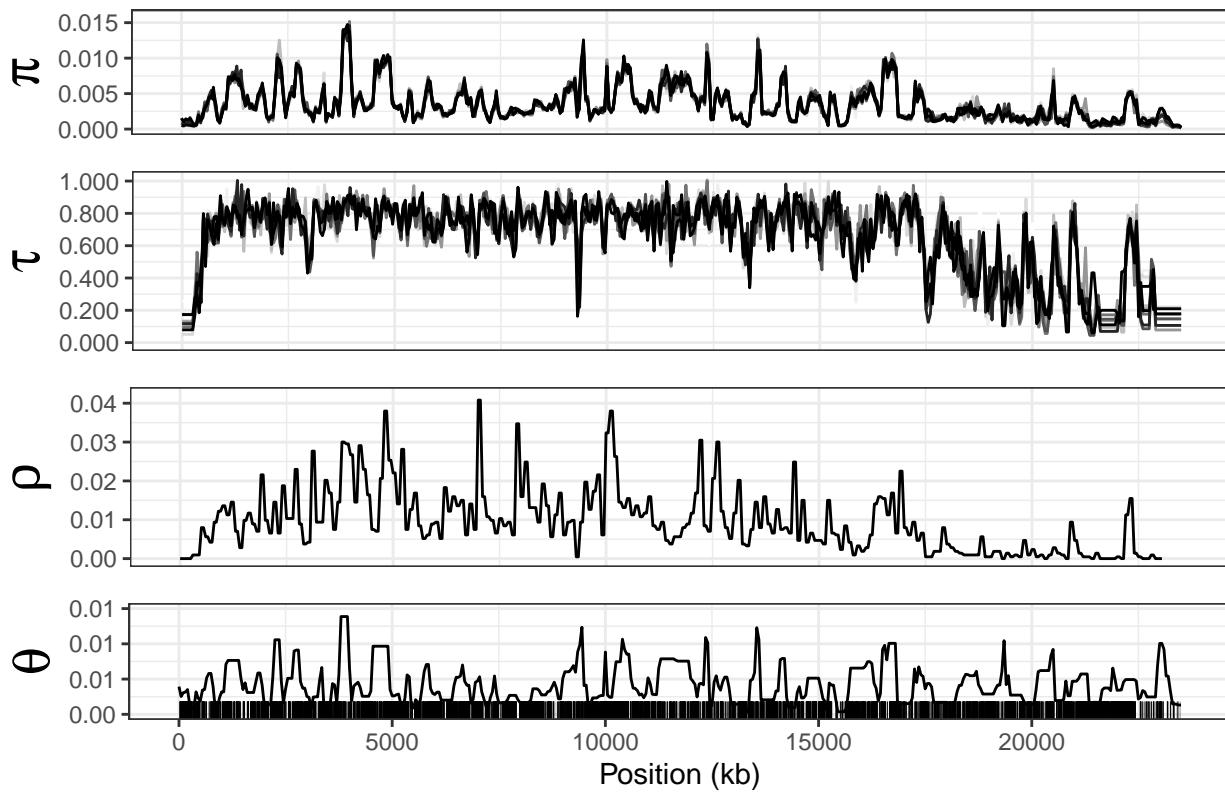
```

50 kb



```
true.maps.50kb <- plot_grid(pi.map.50kb, tau.map.50kb, rho.map.50kb, theta.map.50kb, ncol = 1)  
true.maps.50kb
```

50 kb



```

## 200kb
theta.land.200kb <- cbind.data.frame(1:nrow(sim.theta.200kb), sim.theta.200kb$Rate * 4 * 1e+4 * 1e-7)
names(theta.land.200kb) <- c("bin", "rate")

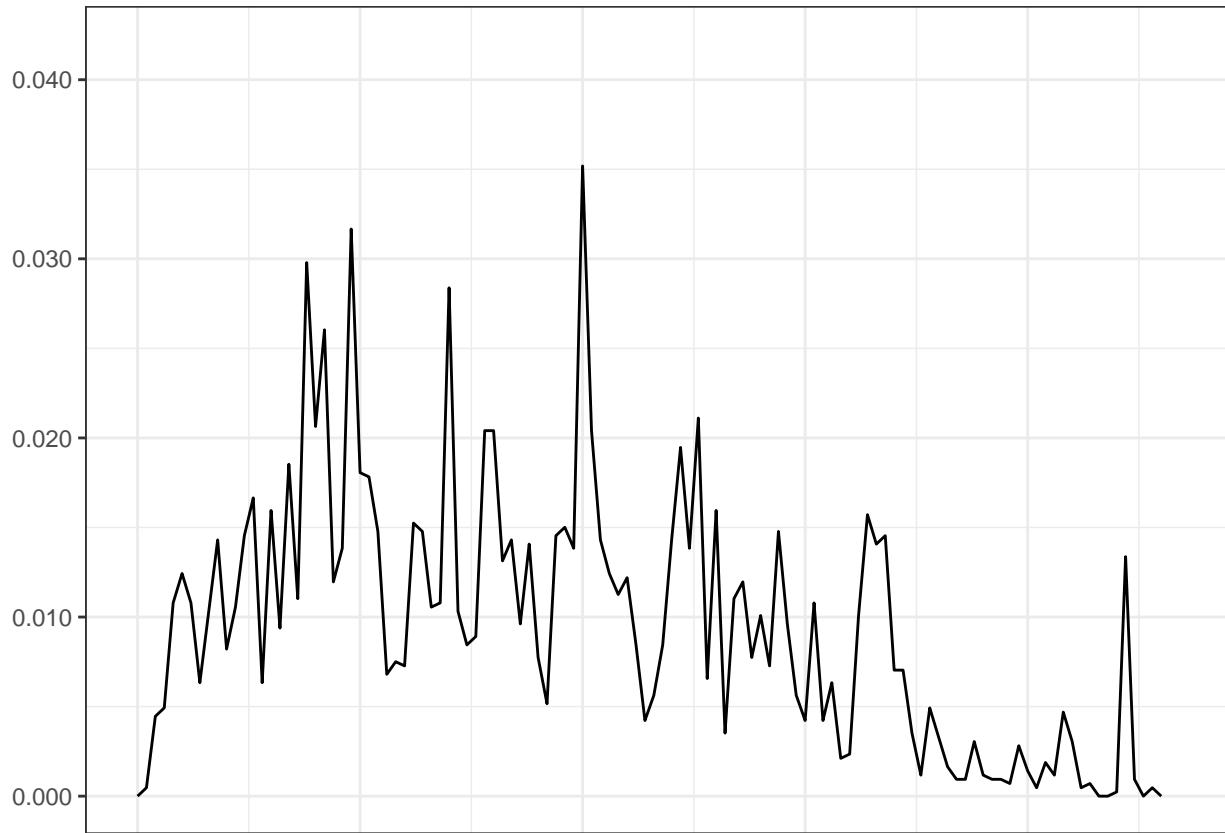
rho.land.200kb <- cbind.data.frame(1:nrow(sim.rho.200kb), sim.rho.200kb$Rate * 1e-7 * 4 * 1e+4)
names(rho.land.200kb) <- c("bin", "rate")

pi.lands.200kb <- cbind.data.frame(1:nrow(sim.theta.200kb), rep_1.pi.200kb$avg, rep_2.pi.200kb$avg, rep_
names(pi.lands.200kb) <- c("bin", "reps")

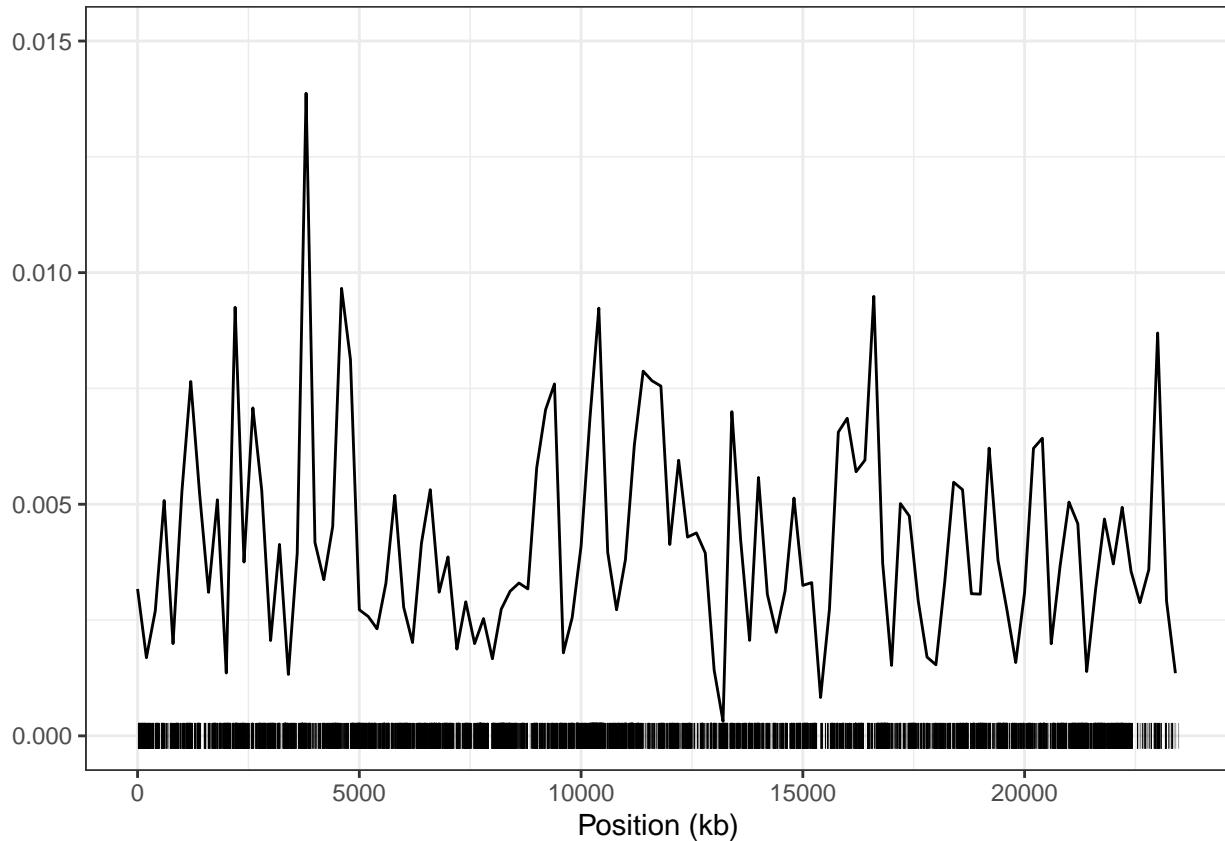
tau.lands.200kb <- cbind.data.frame(1:nrow(sim.theta.200kb), rep1.sim.tmrca.200kb$AverageTmrca, rep2.sim.
names(tau.lands.200kb) <- c("bin", "reps")

rho.map.200kb <- ggplot(data = rho.land.200kb, aes(x = (bin - 1) * 200, y = rate))
rho.map.200kb <- rho.map.200kb + geom_line(data = rho.land.200kb) + theme_bw()
rho.map.200kb <- rho.map.200kb + scale_y_continuous(breaks = pretty_breaks(), labels = scale.3d, limits
rho.map.200kb <- rho.map.200kb + labs(title = NULL, x = NULL, y = NULL)
rho.map.200kb <- rho.map.200kb + theme(axis.title.x=element_blank(),
                                         axis.text.x=element_blank(),
                                         axis.ticks.x=element_blank(),
                                         axis.title.y=element_text(size = 20), legend.position = "none")
rho.map.200kb

```



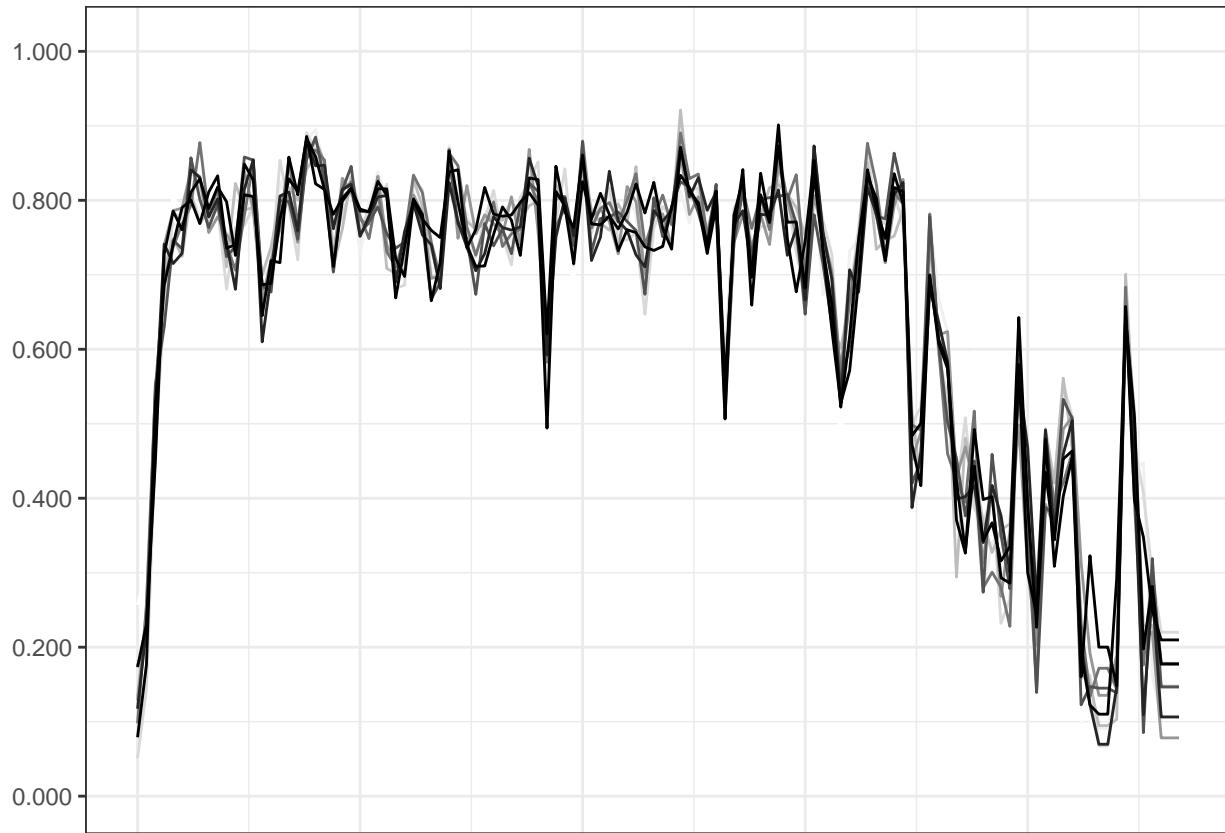
```
theta.map.200kb <- ggplot(data = theta.land.200kb, aes(x = (bin -1) * 200, y = rate))
theta.map.200kb <- theta.map.200kb + geom_line(data = theta.land.200kb) + theme_bw()
theta.map.200kb <- theta.map.200kb + scale_y_continuous(breaks = pretty_breaks(), labels = scale.3d, lin
theta.map.200kb <- theta.map.200kb + labs(title = NULL, x = "Position (kb)", y = NULL)
theta.map.200kb <- theta.map.200kb + geom_segment(data = dm_2L_exome, aes(x = exon_start / 1000, xend =
theta.map.200kb <- theta.map.200kb + theme(legend.position = "none", axis.title.y=element_text(size = 20
theta.map.200kb
```



```

molten.tau <- melt(tau.lands.200kb, id.vars = "bin")
tau.map.200kb <- ggplot(data = molten.tau, aes(x = (bin -1) * 200, y = value / (4 * 1e+4), colour = var))
tau.map.200kb <- tau.map.200kb + geom_line(data = molten.tau) + theme_bw()
tau.map.200kb <- tau.map.200kb + scale_y_continuous(breaks = pretty_breaks(), labels = scale.3d, limits = c(0, 0.015))
tau.map.200kb <- tau.map.200kb + labs(title = NULL, x = NULL, y = NULL)
tau.map.200kb <- tau.map.200kb + scale_color_manual(values = c(brewer.pal(n = 9, name = "Greys"), "Black"))
tau.map.200kb <- tau.map.200kb + theme(axis.title.x=element_blank(),
                                         axis.text.x=element_blank(),
                                         axis.ticks.x=element_blank(),
                                         axis.title.y=element_text(size = 20), legend.position = "none")
tau.map.200kb

```

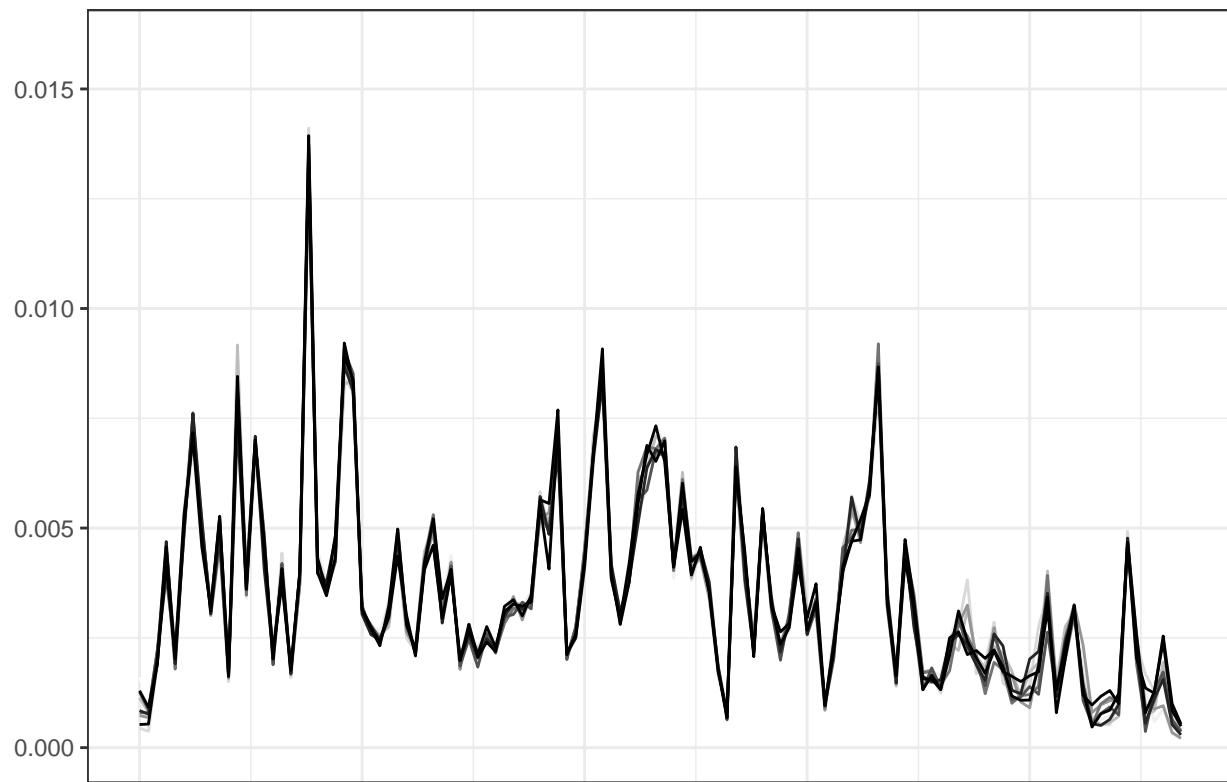


```

molten.pi <- melt(pi.lands.200kb, id.vars = "bin")
pi.map.200kb <- ggplot(data = molten.pi, aes(x = (bin -1) * 200, y = value, colour = variable))
pi.map.200kb <- pi.map.200kb + geom_line(data = molten.pi) + theme_bw()
pi.map.200kb <- pi.map.200kb + scale_y_continuous(breaks = pretty_breaks(), labels = scale.3d, limits =
pi.map.200kb <- pi.map.200kb + labs(title = "200 kb", x = NULL, y = NULL)
pi.map.200kb <- pi.map.200kb + scale_color_manual(values = c(brewer.pal(n = 9, name = "Greys"), "Black"))
pi.map.200kb <- pi.map.200kb + theme(axis.title.x=element_blank(),
                                         axis.text.x=element_blank(),
                                         axis.ticks.x=element_blank(),
                                         axis.title.y=element_text(size = 20),
                                         plot.title = element_text(size = 16), legend.position = "none")
pi.map.200kb

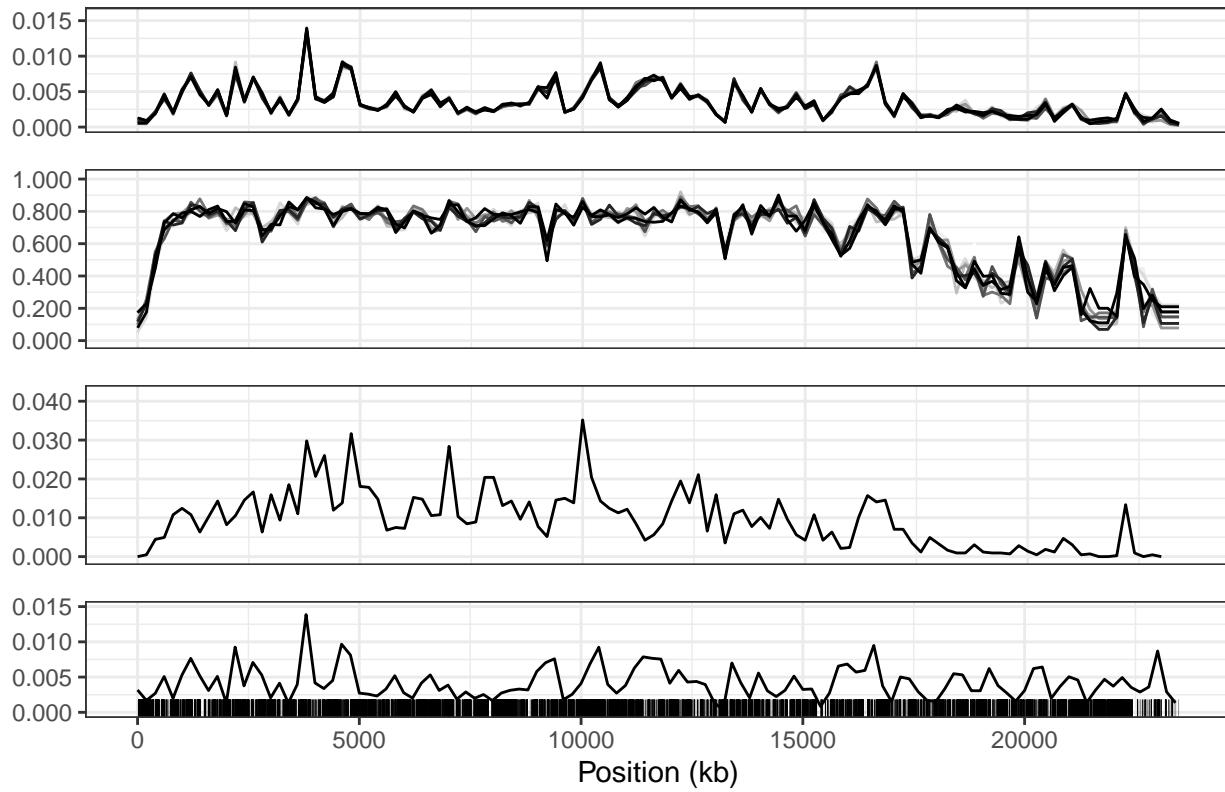
```

200 kb



```
true.maps.200kb <- plot_grid(pi.map.200kb, tau.map.200kb, rho.map.200kb, theta.map.200kb, ncol = 1)  
true.maps.200kb
```

200 kb



```

## 1Mb
theta.land.1Mb <- cbind.data.frame(1:nrow(sim.theta.1Mb), sim.theta.1Mb$Rate * 4 * 1e+4 * 1e-7)
names(theta.land.1Mb) <- c("bin", "rate")

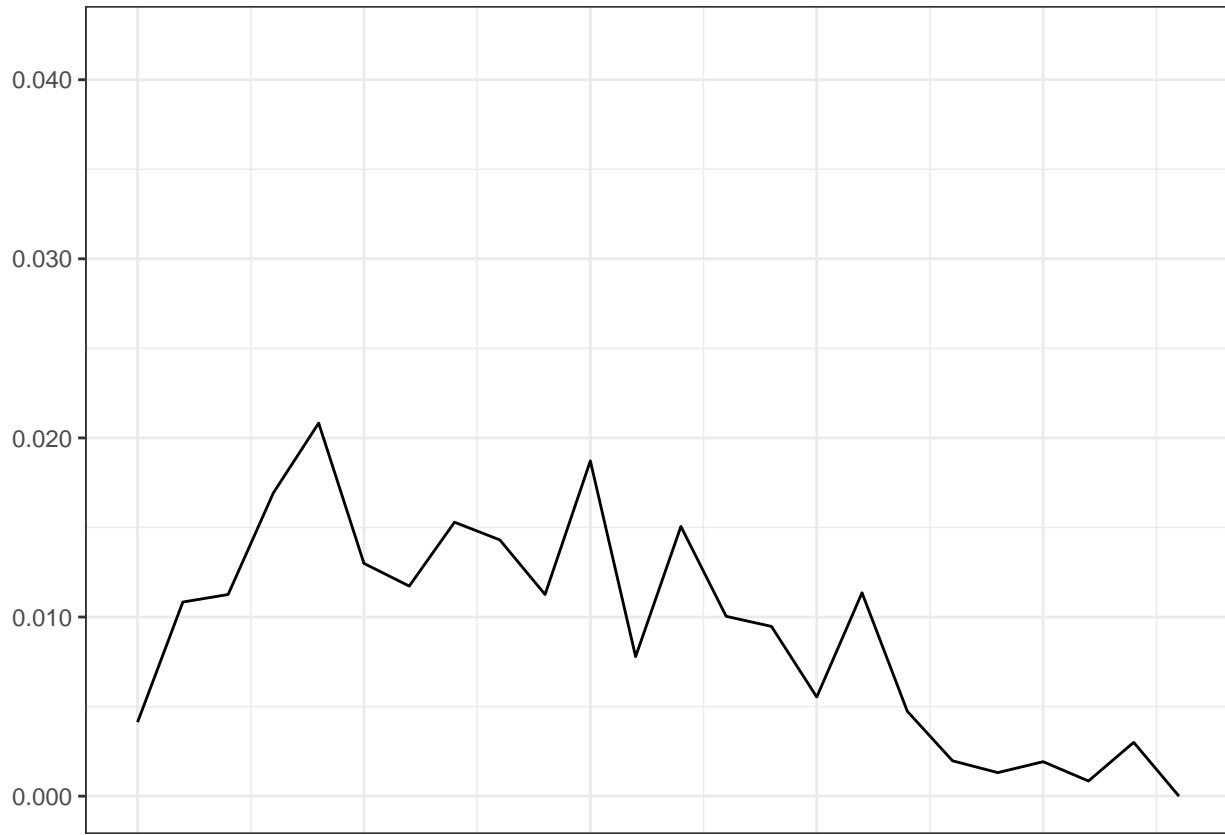
rho.land.1Mb <- cbind.data.frame(1:nrow(sim.rho.1Mb), sim.rho.1Mb$Rate * 1e-7 * 4 * 1e+4)
names(rho.land.1Mb) <- c("bin", "rate")

pi.lands.1Mb <- cbind.data.frame(1:nrow(sim.theta.1Mb), rep_1.pi.1Mb$avg, rep_2.pi.1Mb$avg, rep_3.pi.1Mb$avg)
names(pi.lands.1Mb) <- c("bin", "reps")

tau.lands.1Mb <- cbind.data.frame(1:nrow(sim.theta.1Mb), rep1.sim.tmrca.1Mb$AverageTmrca, rep2.sim.tmrca.1Mb$AverageTmrca)
names(tau.lands.1Mb) <- c("bin", "reps")

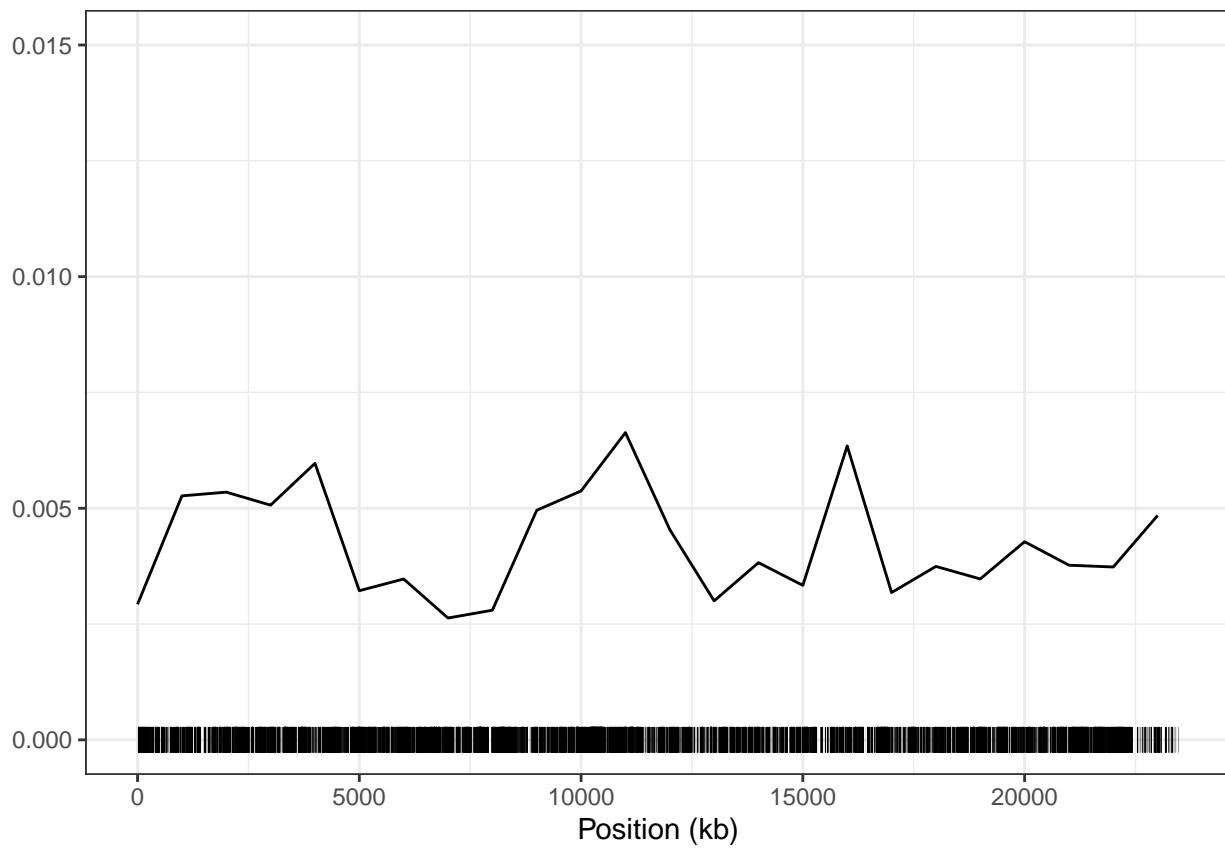
rho.map.1Mb <- ggplot(data = rho.land.1Mb, aes(x = (bin - 1) * 1000, y = rate))
rho.map.1Mb <- rho.map.1Mb + geom_line(data = rho.land.1Mb) + theme_bw()
rho.map.1Mb <- rho.map.1Mb + scale_y_continuous(breaks = pretty_breaks(), labels = scale.3d, limits = c(0, 1))
rho.map.1Mb <- rho.map.1Mb + labs(title = NULL, x = NULL, y = NULL)
rho.map.1Mb <- rho.map.1Mb + theme(axis.title.x=element_blank(),
                                     axis.text.x=element_blank(),
                                     axis.ticks.x=element_blank(),
                                     axis.title.y=element_text(size = 20), legend.position = "none")
rho.map.1Mb

```



```
theta.map.1Mb <- ggplot(data = theta.land.1Mb, aes(x = (bin -1) * 1000, y = rate))
theta.map.1Mb <- theta.map.1Mb + geom_line(data = theta.land.1Mb) + theme_bw()
theta.map.1Mb <- theta.map.1Mb + scale_y_continuous(breaks = pretty_breaks(), labels = scale.3d, limits = c(0, 0.04))
theta.map.1Mb <- theta.map.1Mb + labs(title = NULL, x = "Position (kb)", y = NULL)
theta.map.1Mb <- theta.map.1Mb + geom_segment(data = dm_2L_exome, aes(x = exon_start / 1000, xend = exon_end / 1000, y = 0.04))
theta.map.1Mb <- theta.map.1Mb + theme(legend.position = "none", axis.title.y=element_text(size = 20))

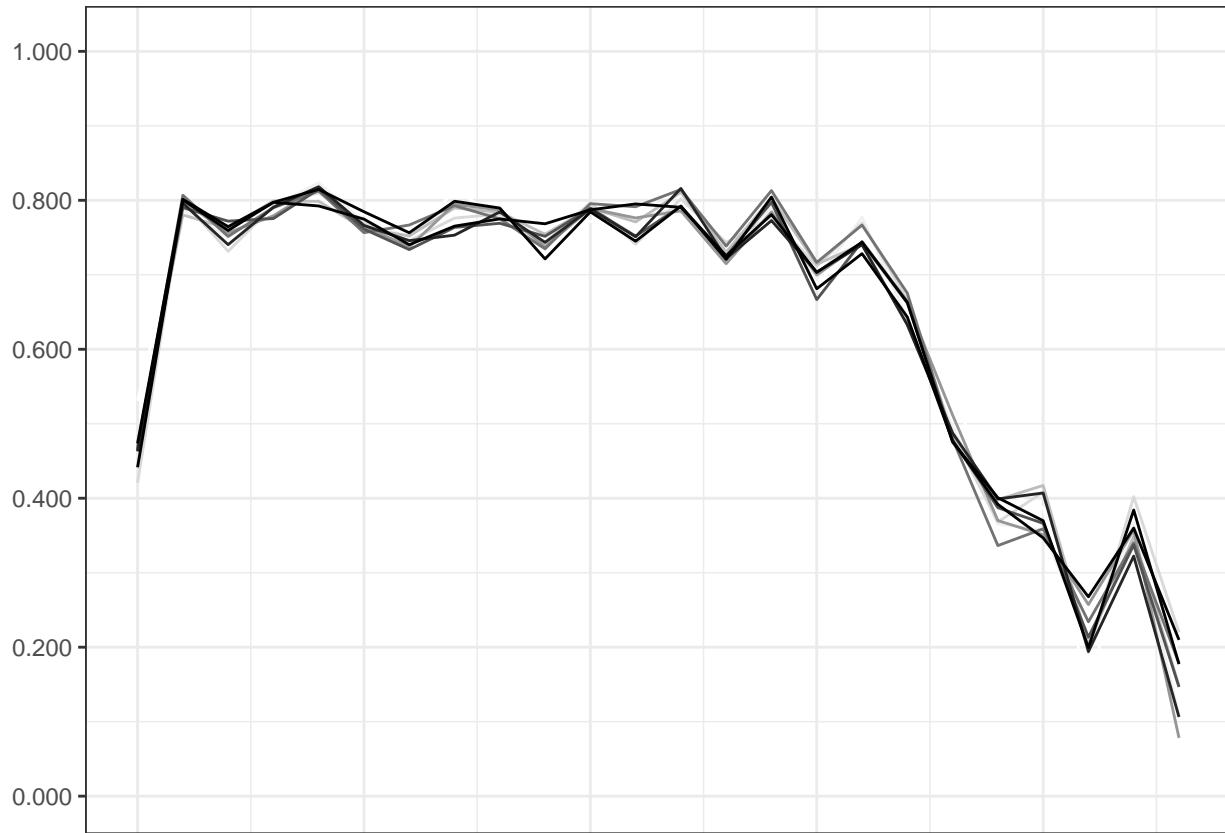
theta.map.1Mb
```



```

molten.tau <- melt(tau.lands.1Mb, id.vars = "bin")
tau.map.1Mb <- ggplot(data = molten.tau, aes(x = (bin -1) * 1000, y = value / (4 * 1e+4), colour = vari
tau.map.1Mb <- tau.map.1Mb + geom_line(data = molten.tau) + theme_bw()
tau.map.1Mb <- tau.map.1Mb + scale_y_continuous(breaks = pretty_breaks(), labels = scale.3d, limits = c(0, 0.015))
tau.map.1Mb <- tau.map.1Mb + labs(title = NULL, x = NULL, y = NULL)
tau.map.1Mb <- tau.map.1Mb + scale_color_manual(values = c(brewer.pal(n = 9, name = "Greys"), "Black"))
tau.map.1Mb <- tau.map.1Mb + theme(axis.title.x=element_blank(),
                                    axis.text.x=element_blank(),
                                    axis.ticks.x=element_blank(),
                                    axis.title.y=element_text(size = 20), legend.position = "none")
tau.map.1Mb

```

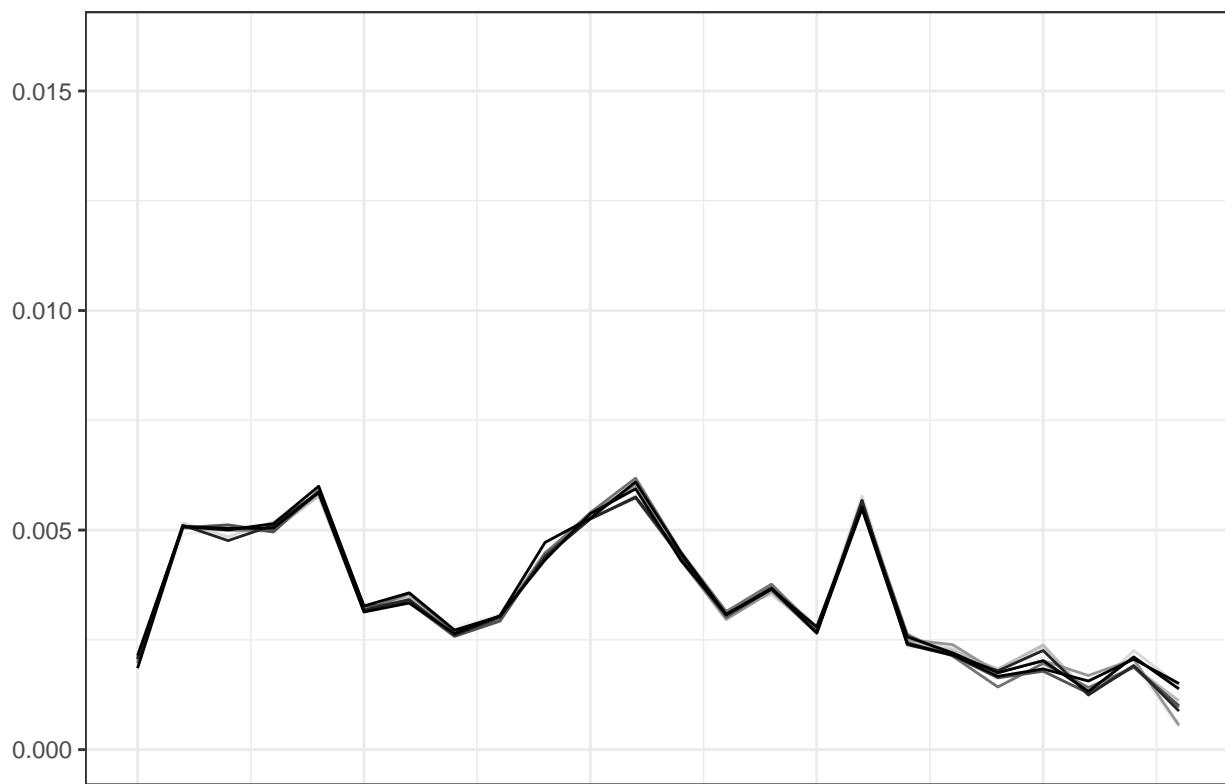


```

molten.pi <- melt(pi.lands.1Mb, id.vars = "bin")
pi.map.1Mb <- ggplot(data = molten.pi, aes(x = (bin -1) * 1000, y = value, colour = variable))
pi.map.1Mb <- pi.map.1Mb + geom_line(data = molten.pi) + theme_bw()
pi.map.1Mb <- pi.map.1Mb + scale_y_continuous(breaks = pretty_breaks(), labels = scale.3d, limits = c(0
pi.map.1Mb <- pi.map.1Mb + labs(title = "1 Mb", x = NULL, y = NULL)
pi.map.1Mb <- pi.map.1Mb + scale_color_manual(values = c(brewer.pal(n = 9, name = "Greys"), "Black"))
pi.map.1Mb <- pi.map.1Mb + theme(axis.title.x=element_blank(),
                                    axis.text.x=element_blank(),
                                    axis.ticks.x=element_blank(),
                                    axis.title.y=element_text(size = 20),
                                    plot.title = element_text(size = 16), legend.position = "none")
pi.map.1Mb

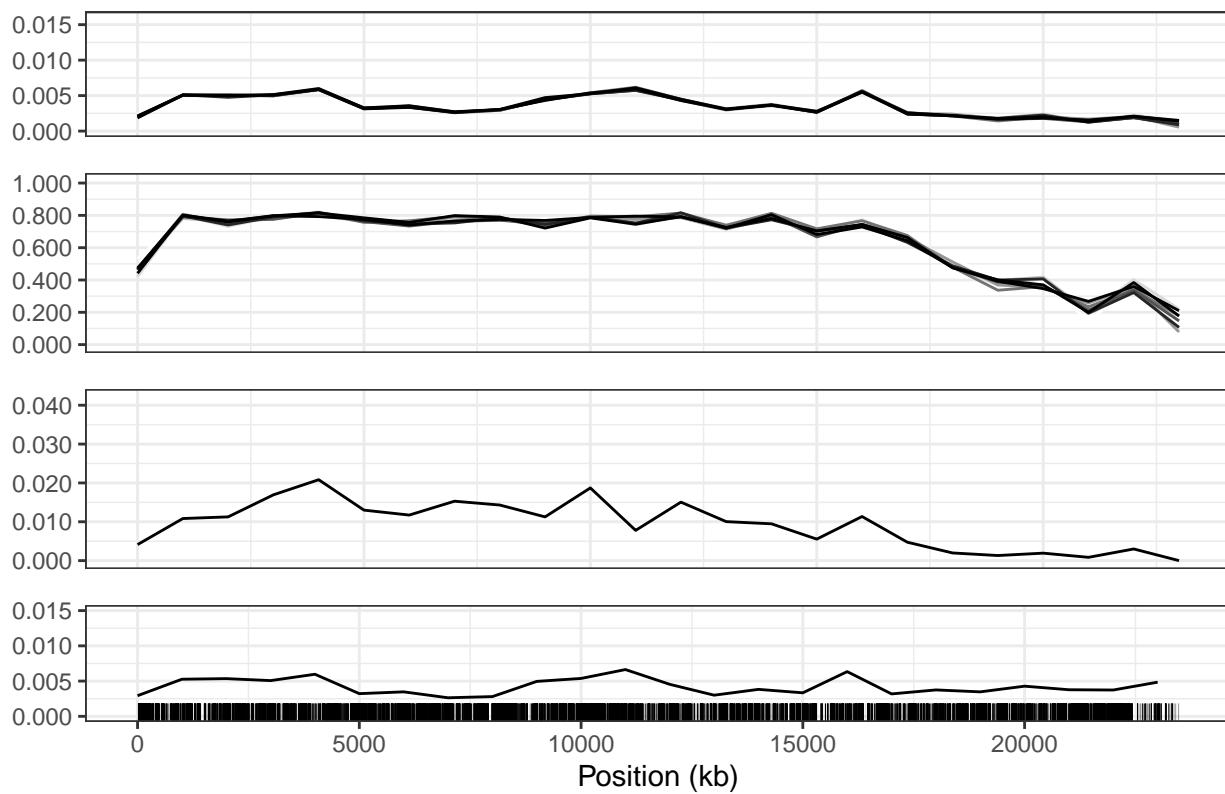
```

1 Mb

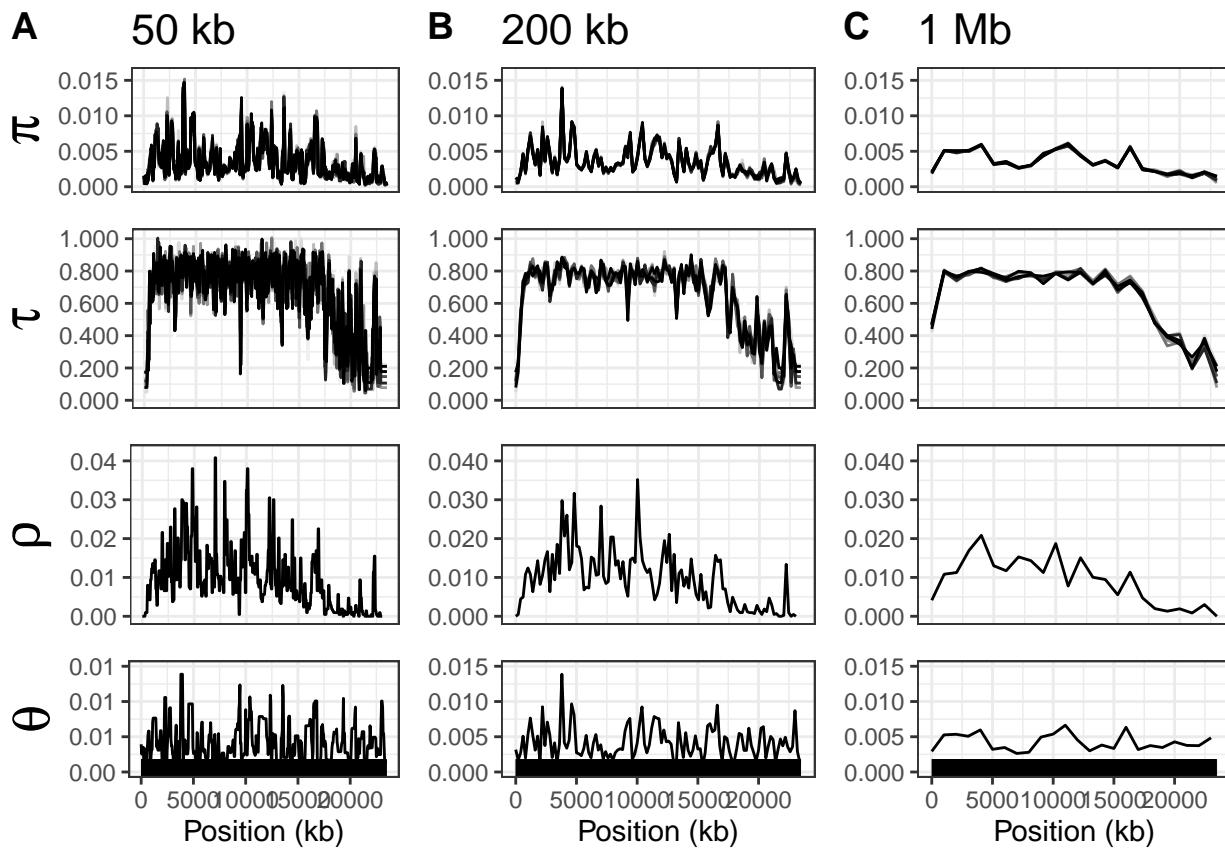


```
true.maps.1Mb <- plot_grid(pi.map.1Mb, tau.map.1Mb, rho.map.1Mb, theta.map.1Mb, ncol = 1)  
true.maps.1Mb
```

1 Mb



```
true.maps <- plot_grid(true.maps.50kb, true.maps.200kb, true.maps.1Mb, labels = "AUTO", ncol = 3, align = "center")  
true.maps
```



```
save_plot("Figure7.pdf", true.maps, base_height = 10, base_width = 20)
```

```
mean(apply(tau.lands.50kb[2:ncol(tau.lands.50kb)], 2, sd) / apply(tau.lands.50kb[2:ncol(tau.lands.50kb)]))

## [1] 0.3626566

mean(apply(tau.lands.200kb[2:ncol(tau.lands.200kb)], 2, sd) / apply(tau.lands.200kb[2:ncol(tau.lands.200kb)]))

## [1] 0.3352754

mean(apply(tau.lands.1Mb[2:ncol(tau.lands.1Mb)], 2, sd) / apply(tau.lands.1Mb[2:ncol(tau.lands.1Mb)]), 2)

## [1] 0.3202002

sd(apply(tau.lands.50kb[2:ncol(tau.lands.50kb)], 2, sd) / apply(tau.lands.50kb[2:ncol(tau.lands.50kb)]))

## [1] 0.007548117

sd(apply(tau.lands.200kb[2:ncol(tau.lands.200kb)], 2, sd) / apply(tau.lands.200kb[2:ncol(tau.lands.200kb)]))

## [1] 0.006994296

sd(apply(tau.lands.1Mb[2:ncol(tau.lands.1Mb)], 2, sd) / apply(tau.lands.1Mb[2:ncol(tau.lands.1Mb)]), 2)

## [1] 0.009614134

mean(apply(pi.lands.50kb[2:ncol(pi.lands.50kb)], 2, sd) / apply(pi.lands.50kb[2:ncol(pi.lands.50kb)]), 2)

## [1] 0.715368
```

```
mean(apply(pi.lands.200kb[2:ncol(pi.lands.200kb)], 2, sd) / apply(pi.lands.200kb[2:ncol(pi.lands.200kb)], 2, mean))
## [1] 0.6213829
mean(apply(pi.lands.1Mb[2:ncol(pi.lands.1Mb)], 2, sd) / apply(pi.lands.1Mb[2:ncol(pi.lands.1Mb)], 2, mean))
## [1] 0.437516
sd(apply(pi.lands.50kb[2:ncol(pi.lands.50kb)], 2, sd) / apply(pi.lands.50kb[2:ncol(pi.lands.50kb)], 2, mean))
## [1] 0.005943989
sd(apply(pi.lands.200kb[2:ncol(pi.lands.200kb)], 2, sd) / apply(pi.lands.200kb[2:ncol(pi.lands.200kb)], 2, mean))
## [1] 0.006465506
sd(apply(pi.lands.1Mb[2:ncol(pi.lands.1Mb)], 2, sd) / apply(pi.lands.1Mb[2:ncol(pi.lands.1Mb)], 2, mean))
## [1] 0.01168644
```