Chapter 5: Genetic Spatial Structure in Murray Cod

Alan Couch

Mon Aug 31 11:05:11 AM 2015

## Loading required package: knitr  
## Loading project configuration  
## Autoloading helper functions  
## Running helper script: helpers.R  
## Running helper script: otoPartChem.R  
## Autoloading cache  
## Autoloading data  
## Loading data set: allOtoChemData  
## Loading data set: CladeNamesToMerge  
## Loading data set: cnData  
## Loading data set: CopyOfDMac14.1567snps  
## Loading data set: DMac14.1567DistMatrix  
## Loading data set: DMac14.1567snps  
## Loading data set: qslAgeData  
## Loading data set: qslAllLarvaInfo  
## Loading data set: qslGeneticsForNestChapter  
## Loading data set: qslLarvaeAgePlus  
## Loading data set: siteGroupings  
## Munging data  
## Running preprocessing script: 01MungeGeneticsData.R  
## Running preprocessing script: 02MungeChemAverages.R

This document includes methods, results and possibly some discussion dot points for the genomics and larval dispersal chapter.

source("http://addictedtor.free.fr/packages/A2R/lastVersion/R/code.R")# load code of A2R function  
library(ggplot2)  
library(ggdendro)  
library(ape)  
library(dendextend)  
library(Hmisc)  
library(ade4)

## Method

The larvae used in the dispersal analysis were collected in 2013 from 6 sites.

Age was determined:

Species seperation

Clades and r apps.

DaRT as described...

PCR would bias representation. p67 LFC

Wilmott Skill Metrioc p.60 LFC - any relevence?

## Results

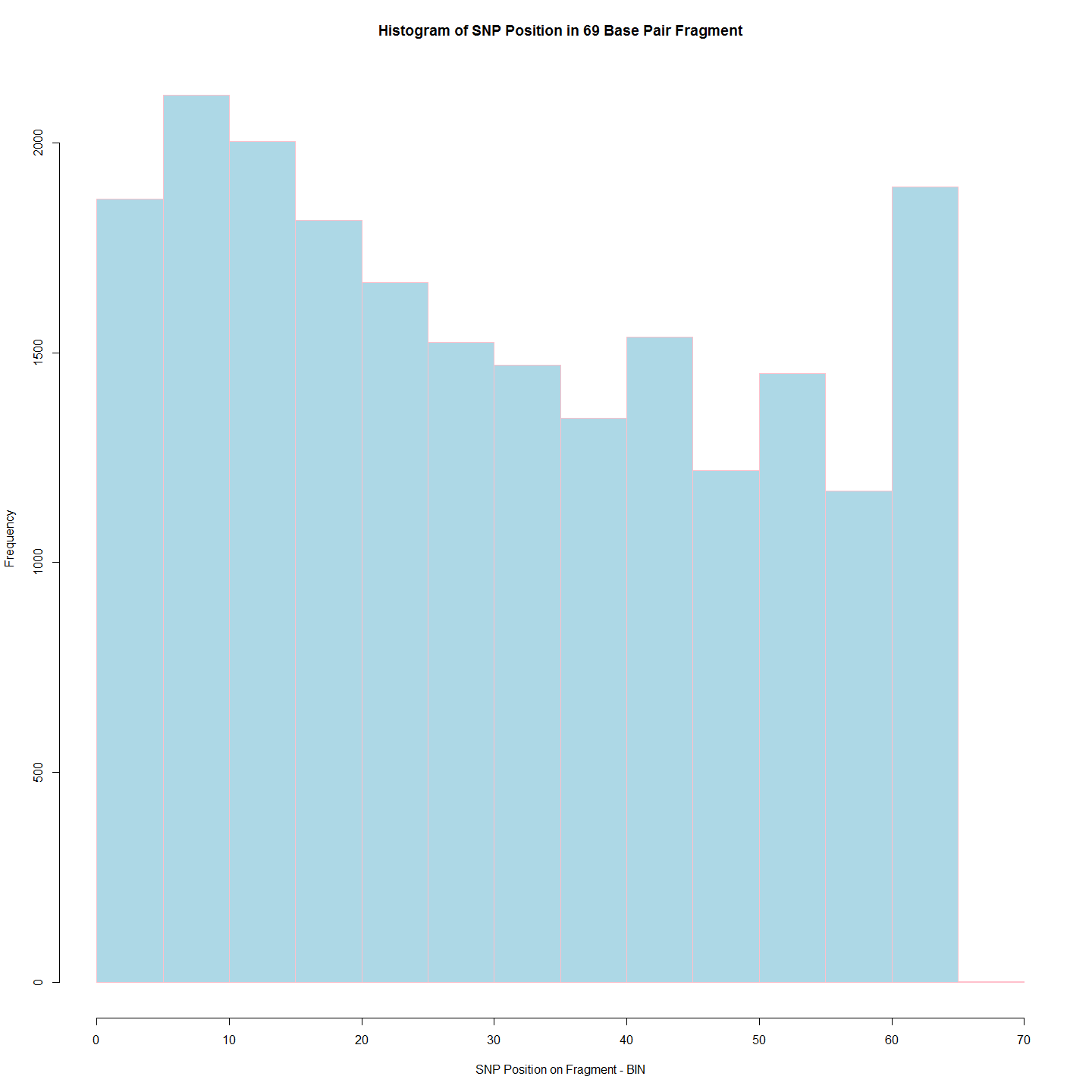
### Polymorphisms Examined

Dart sequencing on Next Generation Sequencing (NGS) platforms was used to reduce genome complexity allowing an intelligent selection of genome fraction corresponding predominantly to active genes. This selection is achieved through the use of a combination of Restriction Enzymes which separate low copy sequences (most informative for marker discovery and typing) from the repetitive fraction of the genome. The advantage of DArTseq is the very high marker densities (tens of thousands of markers) which allows high resolution mapping and detailed genetic dissection of traits.

Number of unique Maccullochella snps analysed in the DaRT sequences was 9871. This provide high resolution of the genotypes of the larvae. The base pairs that had mutated we as follows: Became  
Was | A | C | G | T | Totals| ----|-----|-----|-----|-----|-------| A | |424 |1629 |422 |2475 | C |491 | |447 |1893 |2831 | G |1790 |432 | |541 |2763 | T |457 |1570 |442 | |2469 | ----|-----|-----|-----|-----|-------| Tot |2738 |2426 |2518 |2856 | 10538 | of 21076

The polymorphisms examined included SNPs and small indels in restriction enzyme recognition sites, and larger insertions/deletions in restriction fragments. The polymorphism selected for genotyping were all snps.

The length of the fragments of DNA in this sequencing was 69 base pairs. The position of the polymorphism along the 69 base pair DNA fragment can be seen in the following histogram



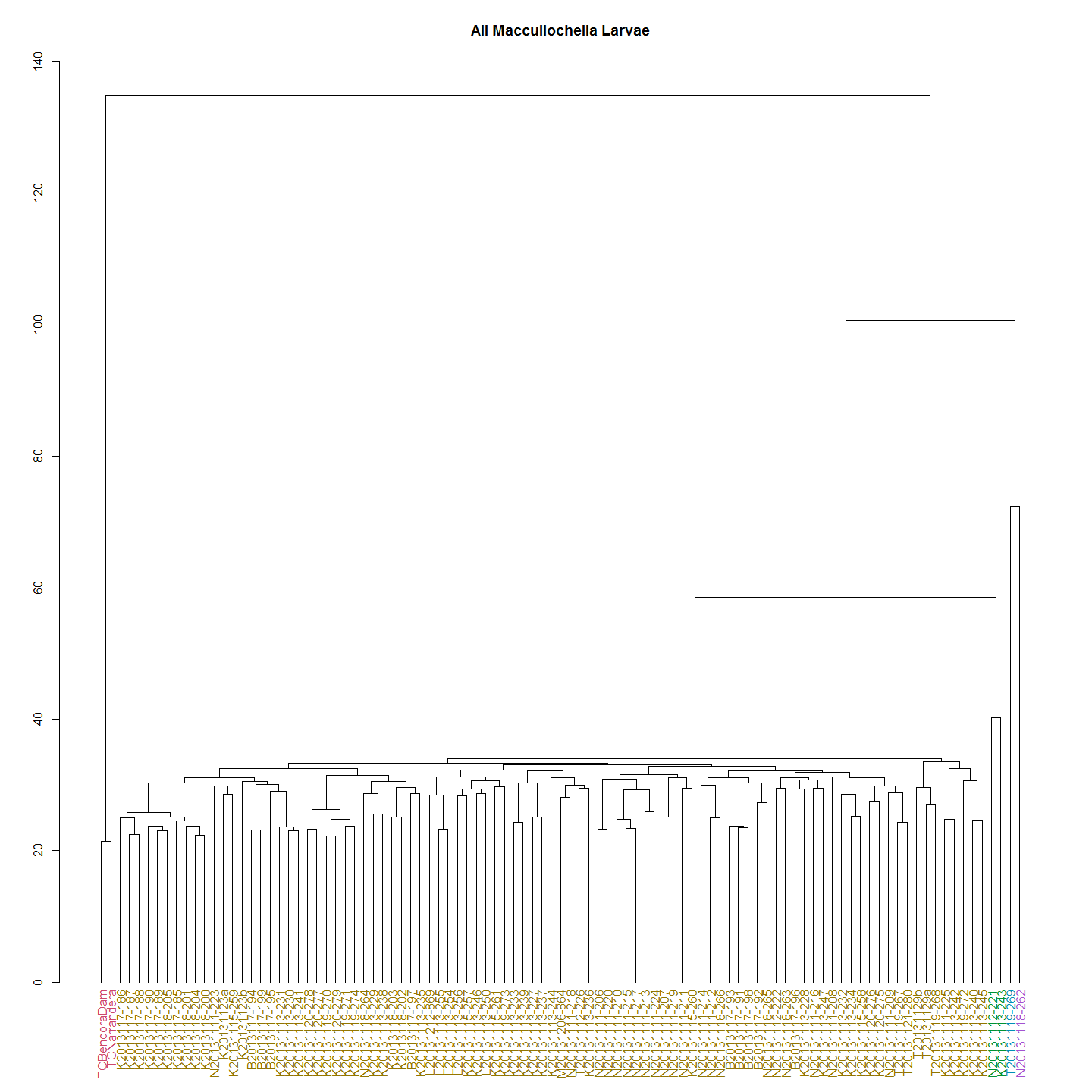
## [1] 13:G>A 7:T>C 32:A>G  
## 768 Levels: \* 10:A>C 10:A>G 10:A>T 10:C>A 10:C>G 10:C>T 10:G>A 10:G>C 10:G>T 10:T>A 10:T>C 10:T>G 11:A>C 11:A>G 11:A>T 11:C>A 11:C>G 11:C>T 11:G>A 11:G>C 11:G>T 11:T>A 11:T>C 11:T>G 12:A>C ... SNP

## dfBPchange  
## A>C A>G A>T C>A C>G C>T G>A G>C G>T T>A T>C T>G   
## 0 424 1629 422 491 447 1893 1790 432 541 457 1570 442

### All *Maccullochella* Larvae

A dendrogram of all the larvae allows examination of the relationships between all the larvae. In the first instance this is neccessary to ensure there are no species other than Murray cod in the subsequent analysis.

MacDm <- dist(allsnps) #Create a distance matrix for all Maccullochella larvae  
  
#heatmap(as.matrix(MacDm)) #Make a heat map  
  
MacHC <- hclust(MacDm) #make a heirarchical cluster  
  
plot(color\_labels(MacHC, k = 5), main="All Maccullochella Larvae") #Plot the cluster dendrogram

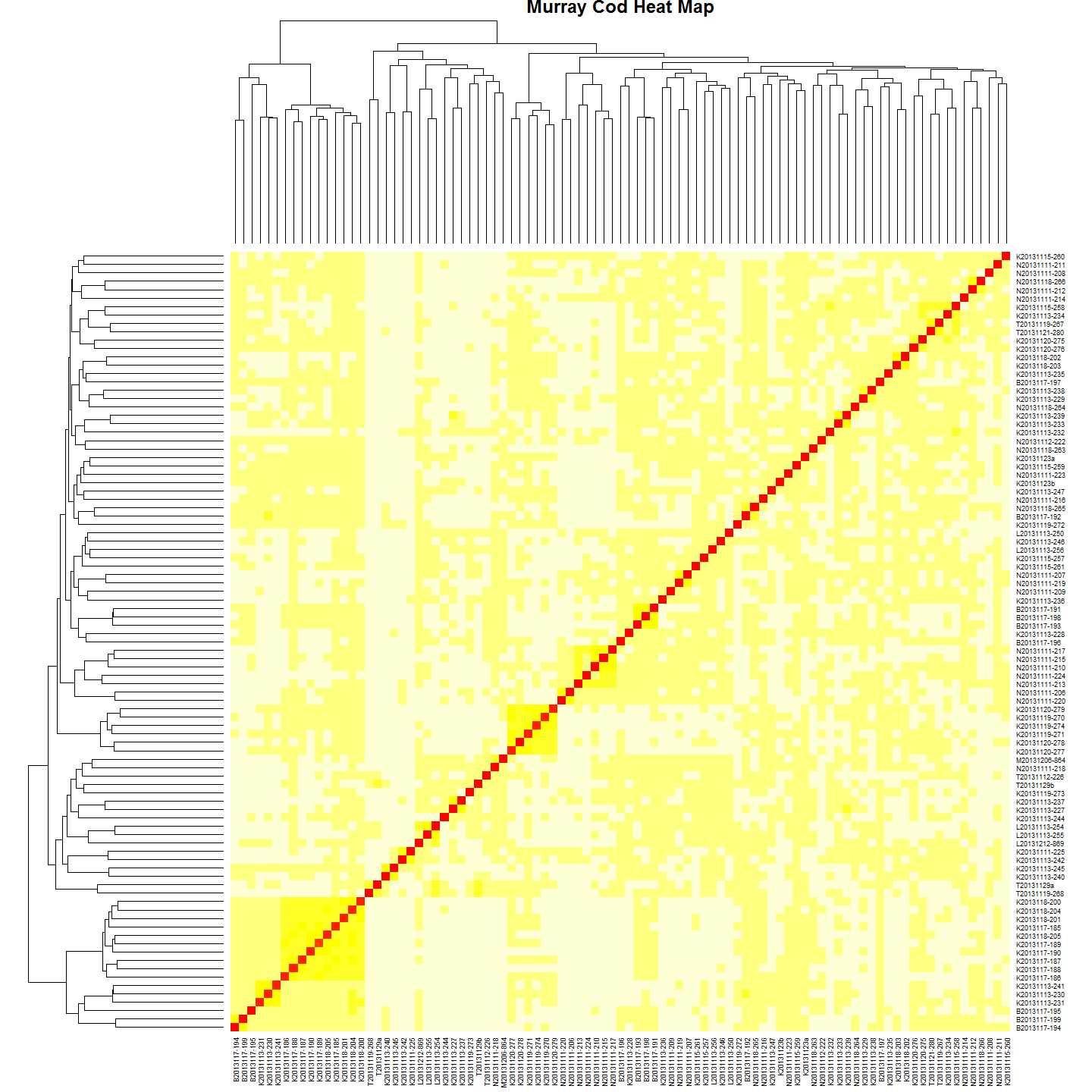
 As it turns out there are multiple *Maccullochella* species. All the larvae collected fall into one of four distinct clades. The above dendrograms shows all larvae (mostly Murray cod) and two known Trout cod controls, and what appears to be F1 and F2 hybrids between the two species. There were 4 hybrid and no pure trout cod larvae detected in the 92 larvae caught and sequenced from the river. This is an interesting finding and detail can be found in the Trout Cod Appendix if the reader is interested.

So we now use the set with the non-Murray cod removed to identify Murray cod clades and conduct the remainder of the analysis on the Upper Murrumbidgee Murray cod larvae only. The Trout cod and the hybrids are easily identified and eliminated from the data to ensure we are looking only at the Murray cod larvae.

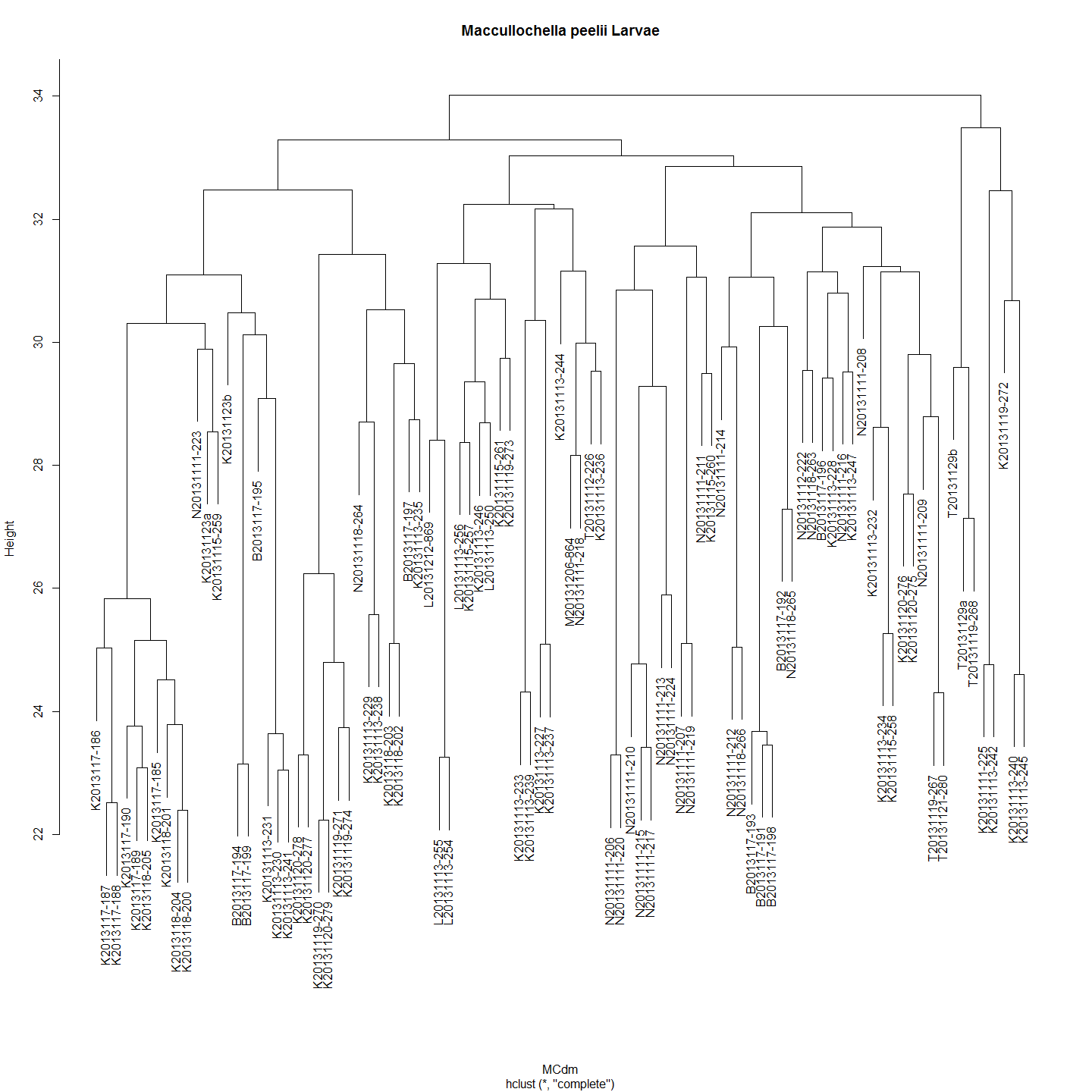
### Murray Cod Larvae

First we look at a heat map and dendrogram of Murray cod larvae having excluded Trout cod and hybrids data from the data frame.

MCdm <- dist(MCsnps)#Create a Murray Cod Only distance matrix  
  
heatmap(as.matrix(MCdm), main="Murray Cod Heat Map")#Heat map

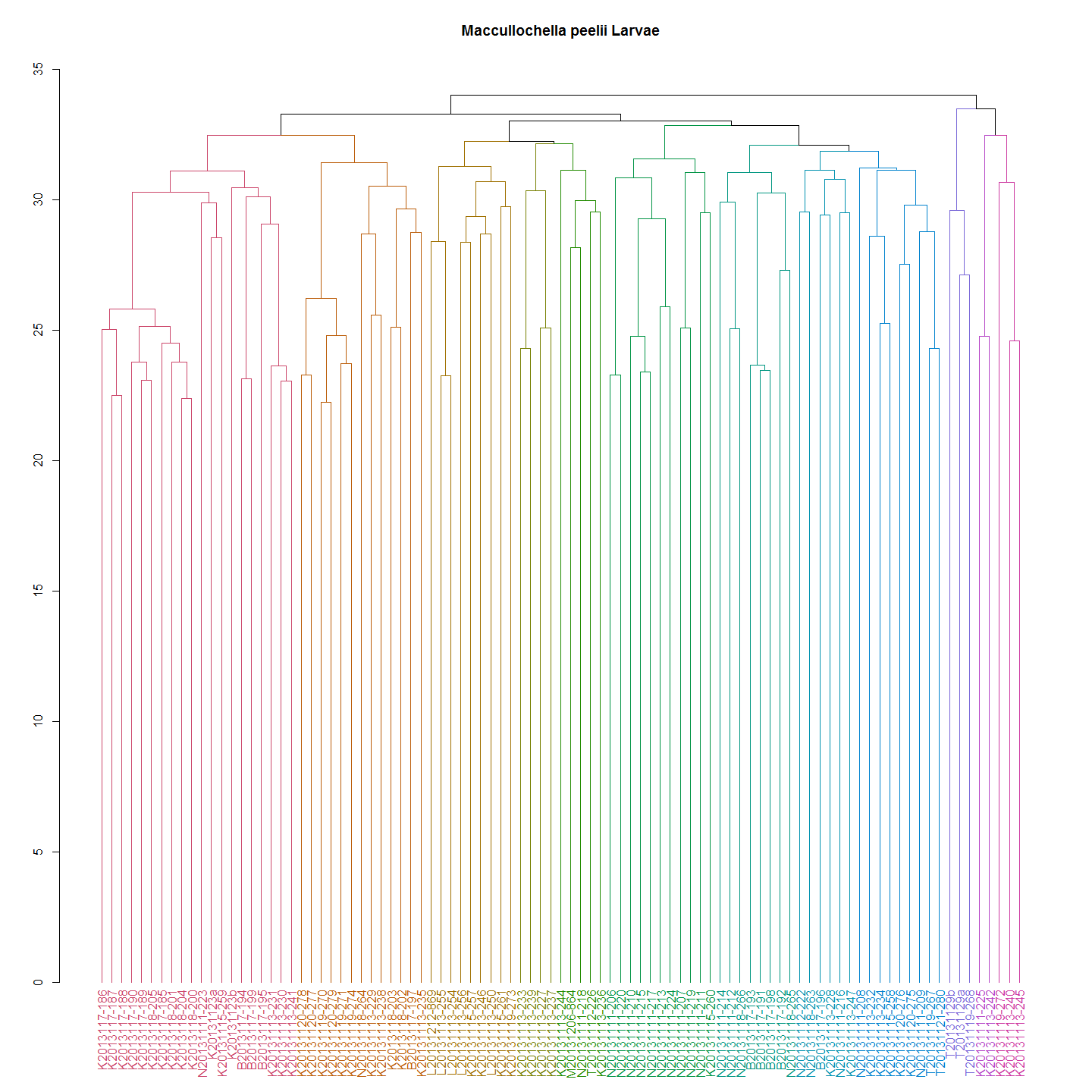


MChc <- hclust(MCdm) #Clutser  
#Plot it  
plot(MChc,main="Maccullochella peelii Larvae")



In most cases very closely related larvae have been collected at the same spot over the same period of a few of days. Pairs or clusters of larvae with a height of less than 26.5 are likely siblings. But in some cases very closely related pairs of larvae have turned up at different sites. Sometimes over time frames that seem unlikely or suggest the larvae have travelled upstream. However, must remember that it is siblings - not the same fish - that is caught so this might just represent the 'smear' of larvae along the river after dispersing from the nest.

The distance matrix suggests a very low genetic diversity in the population of Murray cod sampled with most of the distances around 0.03. Given this, can we be sure that the most closely related larvae are siblings?



While the Murray cod larvae are all very closely related, they can now be resolved into 12 clades.These clades represent 12 'extended family groups' and the 3 clades represent the three higher level (race) clades in the river.

cladeNo<-cutree(dend1,k=12) #This is like using:dendextend:::cutree.dendrogram(dend1,h=70) h or k can be specified  
cladeNo<-as.data.frame(cladeNo)  
#For example:  
cladeNo[c(20:30),] #The clades are numbered by default. So I need to name them something sensible for subsequent analysis.

## [1] 7 3 6 8 9 9 3 3 8 8 3

Extract three higher level clades

dend2 <- color\_labels(MChc, k = 3)#Use h or k  
raceCladeNumber<-cutree(dend2,k=3) #This is like using:dendextend:::cutree.dendrogram(dend1,h=70) h or k can be specified  
raceCladeNumber<-as.data.frame(raceCladeNumber)  
raceCladeNumber[,] #The clades are numbered by default. So I need to name them something sensible for subsequent analysis.

## [1] 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 3 1 1 3 1 1 1 1 1 1 1 3 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3

larvalClades<-merge(cladeNo,raceCladeNumber, by="row.names")  
row.names(larvalClades)<-larvalClades$Row.names  
larvalClades$Row.names<-NULL  
  
#Add Clade Names  
#So the list of larval clades (two levels) is:  
#larvalClades

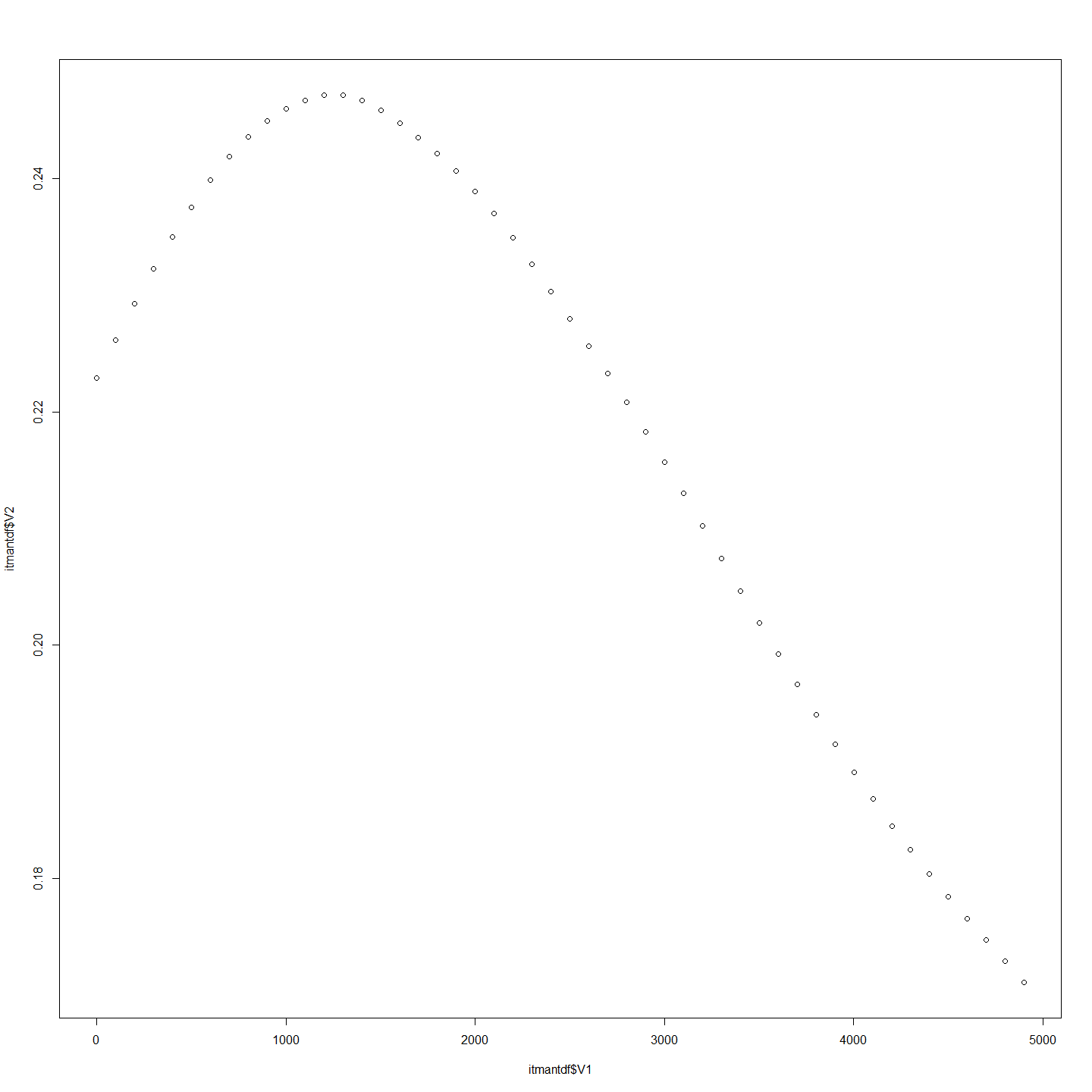
## Upper Murrumbidgee Larval Murray Cod Genetic and Geographic Distances

This test correlates the physical distances of the nests of the larvae with genetic distance of larvae from the Murrumbidgee collected in 2013 from 6 sites.

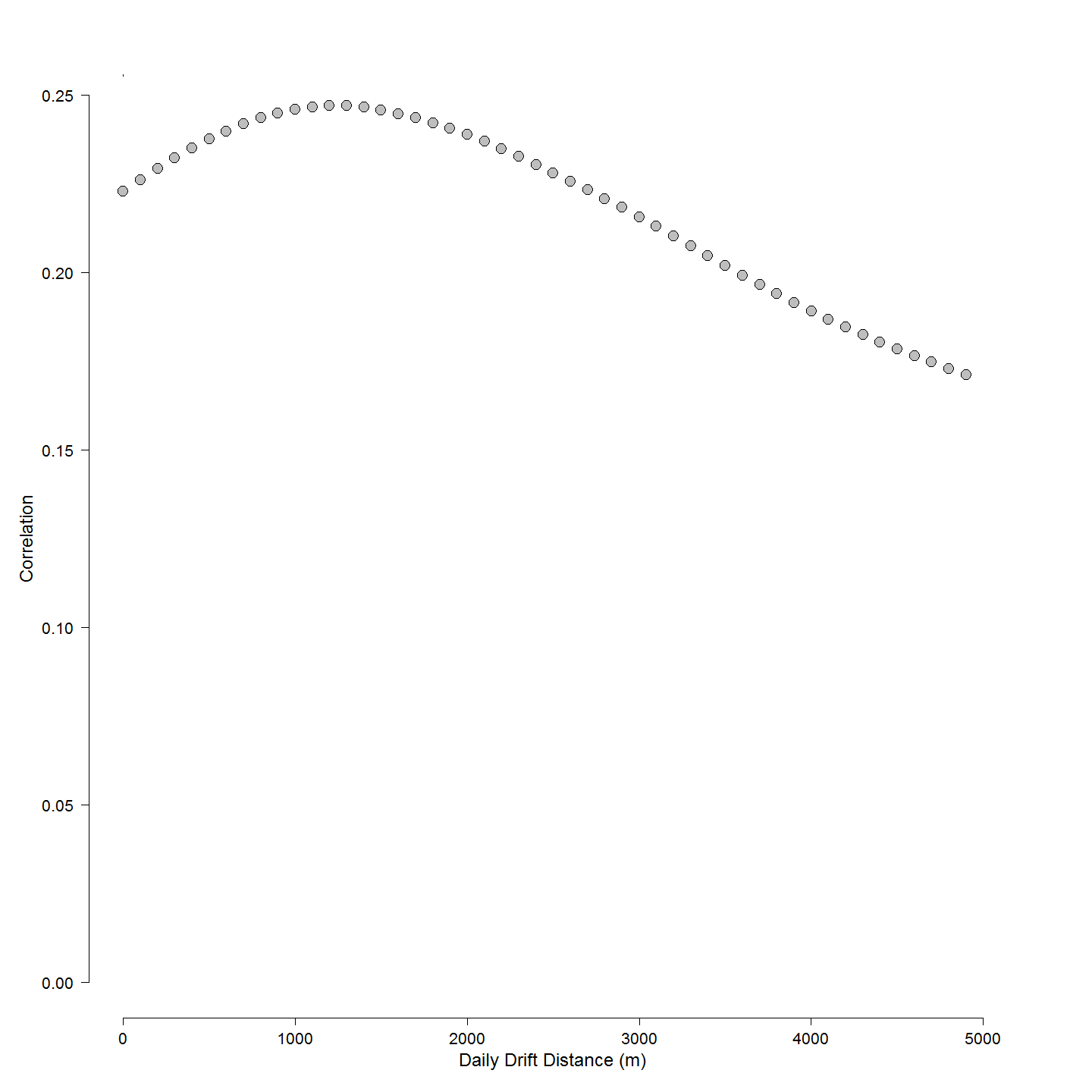
By iterating the mantel test using distance matrices generated for nest distances based on a range of larval dispersal velocities and the time available to those larvae for dispersal, we can identify the most highly correlated - the peak - which then allows an estimate of the average distance a nest site is from the collection site. The iterations of the Mantel test used a larval dispersal velocity range from -1000 m to 5000 metres per day. The highest correlation represents the the average distance that larvae disperse. The assumption is that drift is downstream, not upstream. The curve produced from the estimations is as follows.

### Iterate through Mantel test using a range of possible larval dispersal velocities.

library(Hmisc)  
library(ade4)  
  
#This file is to calculate the data as before but mainly to iterate the mantel test from 1 to 5000 metres at 100m increments (to save time).  
  
# The calculations that can happen outside the iteration are:  
  
#Age from Otolith Length: 0.0591584\*larv$MeanOtolithLength + 0.0331233  
#Hatch DoY : [Day of Year Caught]-[Age From Otolith Length]  
#Incubation: 20.67-0.667\*[WaterTemp(DegC) Mean]  
#Spawnin:[Hatch]-[Incubation]  
#larv$nestdist<-larv$Distance.to.Angle.Crossing..m.-(300\*(larv$Day.of.Year-(larv$hatchDoY+7)))  
  
#Create a MCsnps set with row names as a column.  
MCchecklist<-row.names(MCsnps)  
MCchecklist<-as.data.frame(MCchecklist)# 93 records  
  
#remove a few more anomolies  
MCchecklist1 <- as.data.frame(MCchecklist[-c(1:7), ])  
# Keep every record in larv that is also in MCchecklist (i.e., the intersection).  
  
larv\_intersection <- larv[larv$Label %in% MCchecklist$MCchecklist,]  
#Thanks: https://heuristically.wordpress.com/2009/10/08/delete-rows-from-r-data-frame/  
  
larv<-larv\_intersection  
larv\_intersection<-NULL  
  
itmant <- matrix(nrow=5000, ncol=3) #Is 5000 DF to store result but NA omited later. They result from the increment 100 in the for loop below.  
  
#Iteration begins here:  
for (nd in seq(1,5000, by=100)){#To be 0:5000 eventually for(i in seq(1, 10, by = 2))   
   
 larv$nestdist<-larv$Distance.to.Angle.Crossing..m.-(nd\*(larv$Day.of.Year-(larv$hatchDoY+7)))   
   
 ###########  
 # Create GenDist from code in the Murray Cod SNPS table  
 MCdm<-MCsnps[-c(1:7),] #remove non-numeric variables  
 MCdm <- dist(MCdm) # Create a Murray Cod distance matrix  
 MCdm<-as.matrix(MCdm)  
 MCdm<-as.data.frame(MCdm)  
 #This is to be used for plotting  
 ###########  
 #Create Geographic Distance Matrix using Nest Distance  
 geodist<-data.frame(larv$Label,larv$nestdist)  
 row.names(geodist)<-geodist[,1]  
 geodist$larv.Label<-NULL  
 geodist<-na.omit(geodist)  
 #geodist<-geodist[complete.cases(geodist),]  
   
 GeoDistMat<-dist(geodist)  
 GeoDistMathm <- as.matrix(GeoDistMat)  
   
   
 #make sure both matrices are in correct order - rows and cols  
 #First sort MCdm  
   
 MCdm<-as.data.frame(MCdm)  
 MCdm$sort<-row.names(MCdm)  
 MCdm <- MCdm[order(MCdm$sort),]#sort row order  
 MCdm$sort<-NULL  
 MCdm<-MCdm[,order(names(MCdm))]#sort column order  
 MCdm<-as.matrix(MCdm)  
   
 #Second sort GeoDist  
 GeoDistMathm<-as.data.frame(GeoDistMathm)  
 GeoDistMathm$sort<-row.names(GeoDistMathm)  
 GeoDistMathm <- GeoDistMathm[order(GeoDistMathm$sort),]#sort row order  
 GeoDistMathm$sort<-NULL  
 GeoDistMathm<-GeoDistMathm[,order(names(GeoDistMathm))]#sort column order  
 GeoDistMathm<-as.matrix(GeoDistMathm)  
   
 mant<-mantel.rtest(as.dist(GeoDistMathm), as.dist(MCdm), nrepet = 9999)  
 #print(nd)  
 #print(mant$obs)  
 #print(mant$pvalue)  
   
 itmant[nd,] <- c(nd, mant$obs, mant$pvalue)  
}  
  
itmant<-na.omit(itmant)  
itmantdf<-as.data.frame(itmant)  
plot(itmantdf$V1,itmantdf$V2)



###########  
require(plotrix)  
op <- par(cex.main = 1.5, mar = c(5, 6, 4, 5) + 0.1, mgp = c(3.5, 1, 0), cex.lab = 1.5 , font.lab = 2, cex.axis = 1.3, bty = "n", las=1)  
plot(itmantdf$V1, itmantdf$V2, col="black", pch=21, bg = "grey", cex = 2,  
 xlim=c(0,5000), ylim=c(0,.25), ylab="", xlab="", axes=F)  
axis(1)  
axis(2)   
reg1 <- lm(itmantdf$V2~itmantdf$V1)  
ablineclip(reg1, lwd=2,x1 = .9, x2 = 1.2)   
par(las=0)  
mtext("Daily Drift Distance (m)", side=1, line=2.5, cex=1.5)  
mtext("Correlation", side=2, line=3.7, cex=1.5)

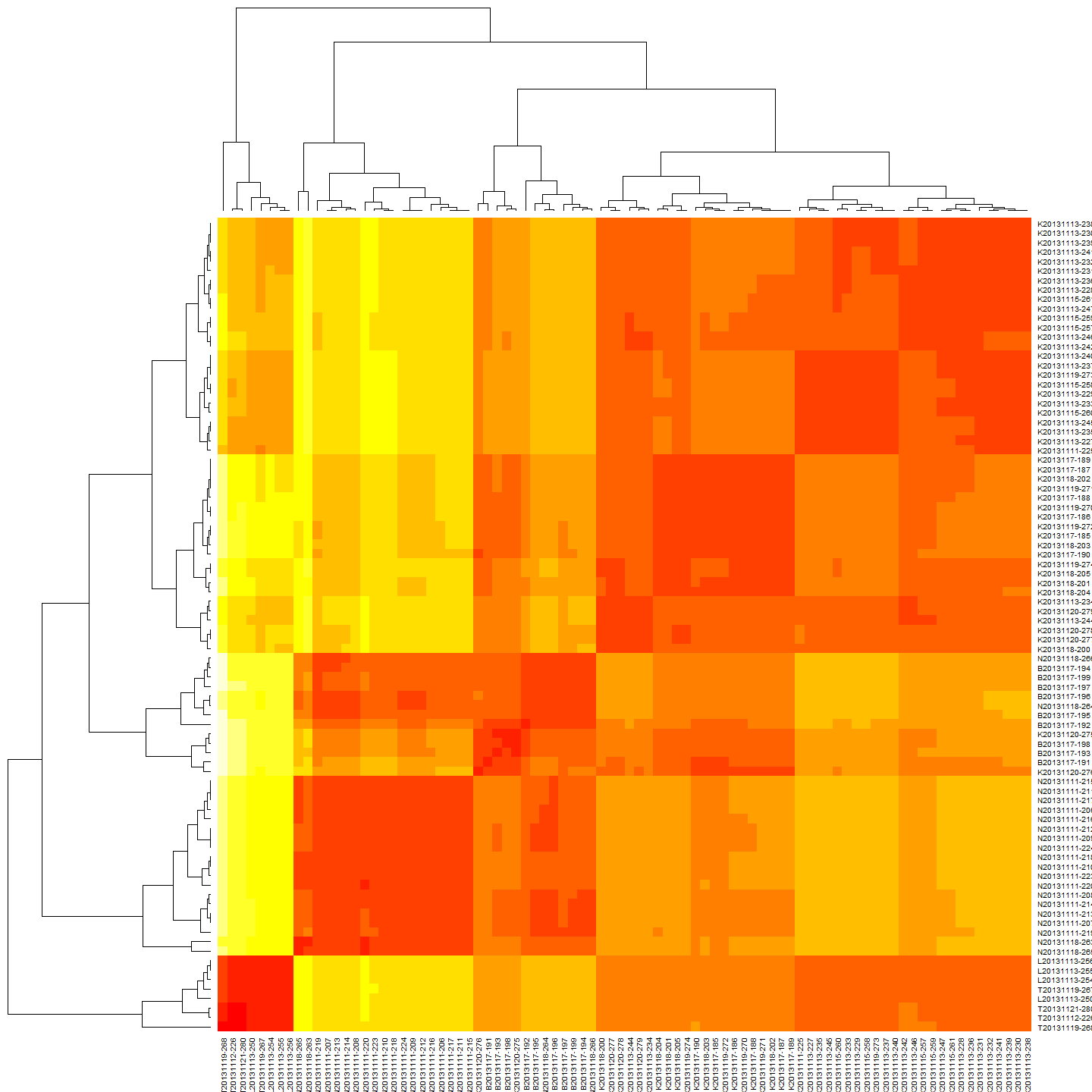


###########  
  
dispersalVelocity<- subset(itmantdf, V2==max(V2) , select = V1)  
dispersalVelocity<-as.numeric(dispersalVelocity)

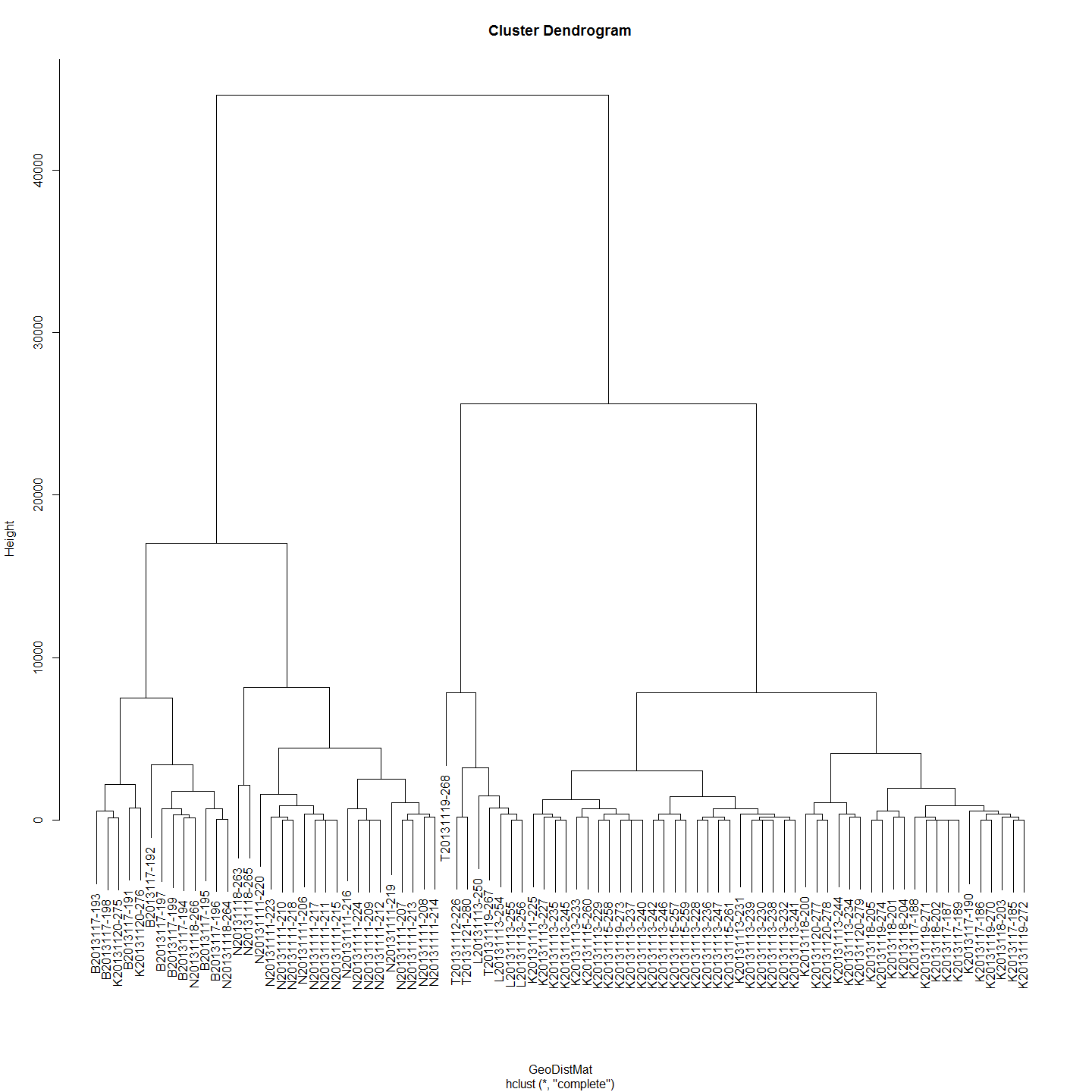
The distance above the collection site as the larvae disperse at 1201 metres per day between leaving the nest and being collected at sampling site also allows an estimate of the Murray cod pelagic larval drift duration to be made.

This average pelagic larval dispersal (PLD) value is an average for larvae along the entire river reach. It tells us nothing about the variance of that dispersal (or can it)??? It is most likely a distribution of larvae drifting (dispersing) varying distances below the nest. The obvious thing to do would be to take the approach on a site by site basis over a number of years, because the river speeds vary at each site and between years but at present there are too few samples from most sites and other years to use such an approach with confidence.

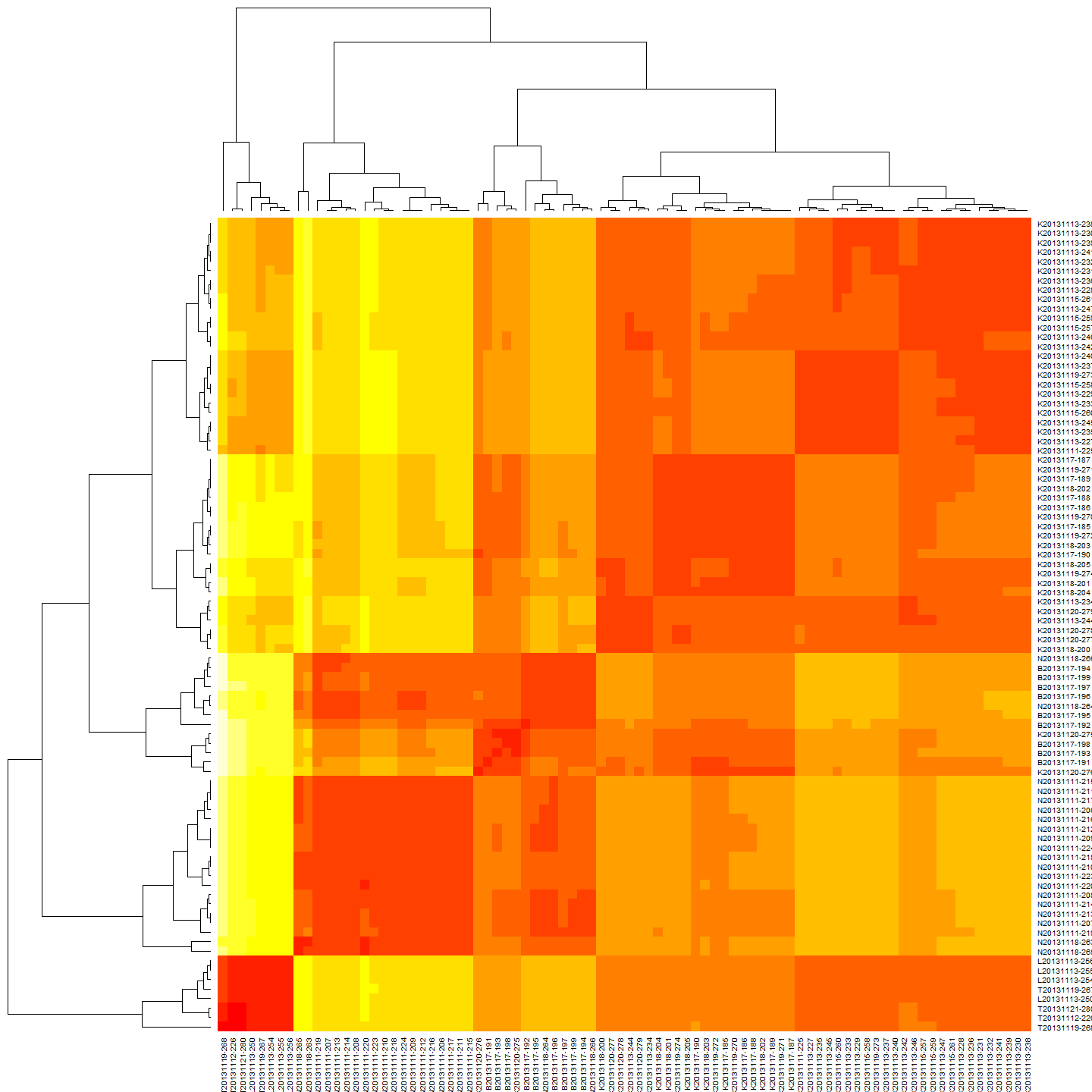
# To create a distance using the previously calculated best estimate of drift velocity (m/d available since leaving brood care)  
larv$nestdist<-larv$Distance.to.Angle.Crossing..m.-(dispersalVelocity\*(larv$Day.of.Year-(larv$hatchDoY+7)))  
#write.csv(format(larv), file="./Tableau/larvForTableau.csv", row.names=FALSE)  
#remove larvae that do not have genetic analysis done.  
#Creat a MCsnps set with row names as a column.  
MCchecklist<-row.names(MCsnps)  
MCchecklist<-as.data.frame(MCchecklist)# 93 records  
  
#remove a few more anomolies  
MCchecklist1 <- as.data.frame(MCchecklist[-c(1:7), ])  
# Keep every record in larv that is also in MCchecklist (i.e., the intersection).  
  
larv\_intersection <- larv[larv$Label %in% MCchecklist$MCchecklist,]  
#Thanks: https://heuristically.wordpress.com/2009/10/08/delete-rows-from-r-data-frame/  
  
larv<-larv\_intersection  
larv\_intersection<-NULL  
  
###########  
# Recreate Genetic Distance  
# Create a Murray Cod distance matrix  
MCdm<-MCsnps[-c(1:7),]  
MCdm <- dist(MCdm)  
MCdm<-as.matrix(MCdm)  
MCdm<-as.data.frame(MCdm)  
#This is to be used for plotting  
MCdmForSibs<-MCdm #need this matrix for sibling analysis .r  
###########  
  
geodist<-data.frame(larv$Label,larv$nestdist)  
row.names(geodist)<-geodist[,1]  
geodist$larv.Label<-NULL  
geodist<-na.omit(geodist)  
geodist1000<-geodist #save this estimate for haplogroups distance plot (after the Iterated Mantel has changed it) my notes also say check this and next 3 lines. Needed?  
  
GeoDistMat<-dist(geodist)  
GeoDistMathm <- as.matrix(GeoDistMat)  
heatmap(GeoDistMathm)



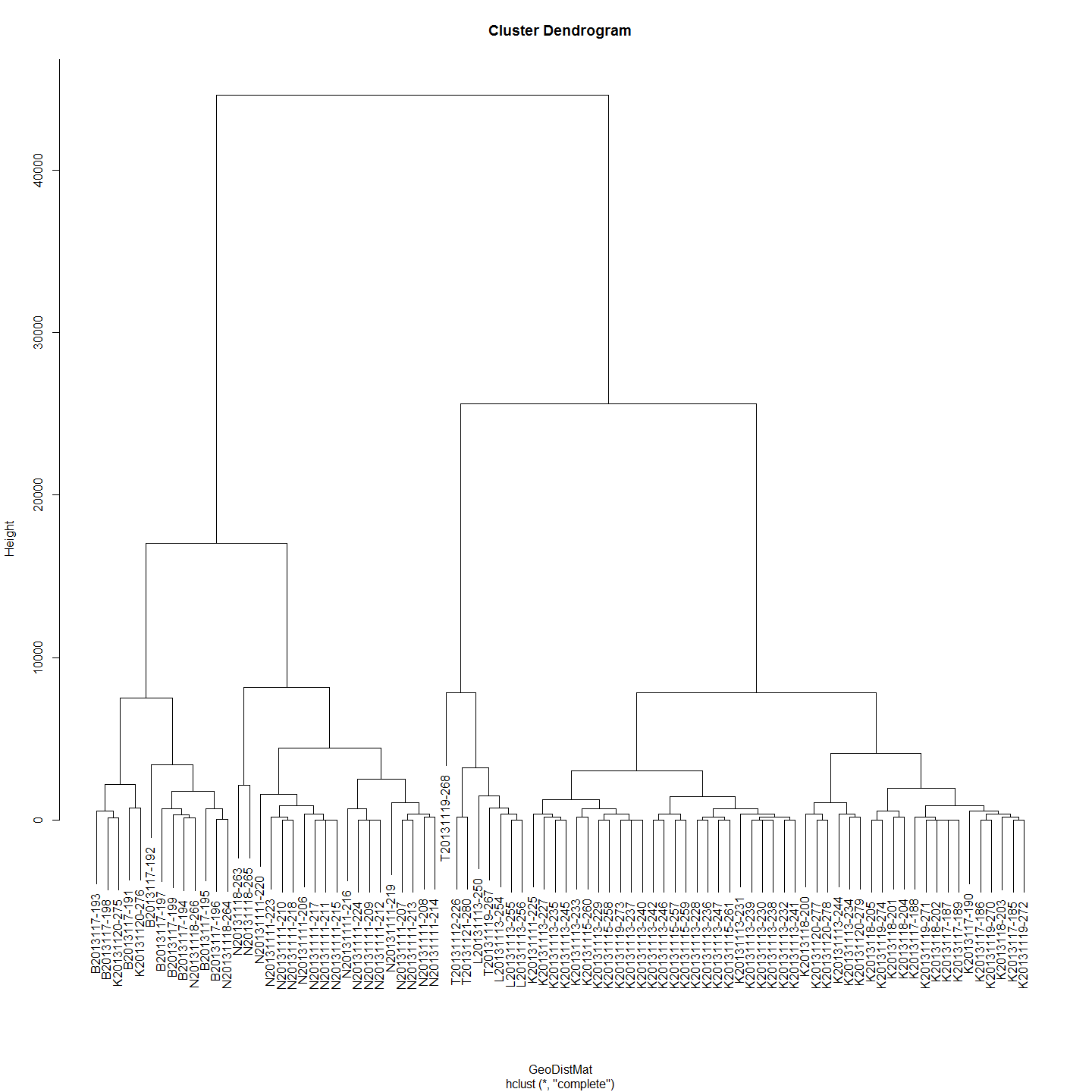
geoclust<-hclust(GeoDistMat)  
plot(geoclust)



#make sure both matrices are in correct order - rows and cols  
#First sort MCdm  
  
MCdm<-as.data.frame(MCdm)  
MCdm$sort<-row.names(MCdm)  
MCdm <- MCdm[order(MCdm$sort),]#sort row order  
MCdm$sort<-NULL  
MCdm<-MCdm[,order(names(MCdm))]#sort column order  
MCdm<-as.matrix(MCdm)  
  
#Second sort GeoDist  
GeoDistMathm<-as.data.frame(GeoDistMathm)  
GeoDistMathm$sort<-row.names(GeoDistMathm)  
GeoDistMathm <- GeoDistMathm[order(GeoDistMathm$sort),]#sort row order  
GeoDistMathm$sort<-NULL  
GeoDistMathm<-GeoDistMathm[,order(names(GeoDistMathm))]#sort column order  
GeoDistMathm<-as.matrix(GeoDistMathm)  
  
heatmap(GeoDistMathm)



geoclust<-hclust(GeoDistMat)  
plot(geoclust)



larv1<-larv#save this estimate for haplogroups distance plot (after the Iterated Mantel has changed it)

Now that these various matrices, class 'dist' objects are created we can proceed for plot.

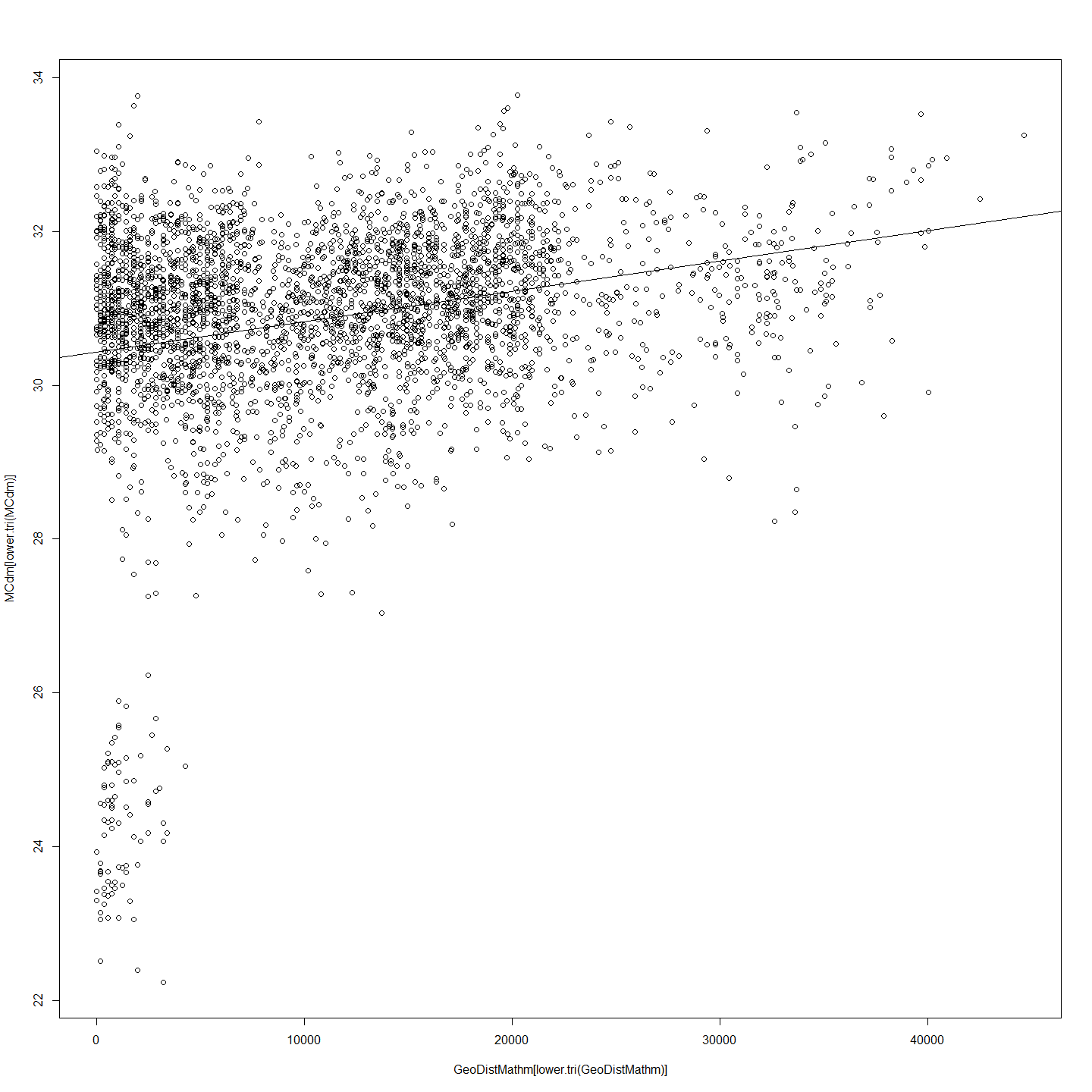
### Plot and Correlate genetic and geographic distance matrices

First a regression model is calculated and then the plot:

#Linear Regression Model  
reg=lm(MCdm[lower.tri(MCdm)]~GeoDistMathm[lower.tri(GeoDistMathm)])  
summary(reg) #does xtable do something nice with this?

##   
## Call:  
## lm(formula = MCdm[lower.tri(MCdm)] ~ GeoDistMathm[lower.tri(GeoDistMathm)])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.3204 -0.4746 0.1832 0.7561 3.2567   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.043e+01 3.631e-02 838.15 <2e-16 \*\*\*  
## GeoDistMathm[lower.tri(GeoDistMathm)] 3.946e-05 2.560e-06 15.41 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.338 on 3653 degrees of freedom  
## Multiple R-squared: 0.06107, Adjusted R-squared: 0.06081   
## F-statistic: 237.6 on 1 and 3653 DF, p-value: < 2.2e-16

plot(GeoDistMathm[lower.tri(GeoDistMathm)],MCdm[lower.tri(MCdm)])  
abline(reg)



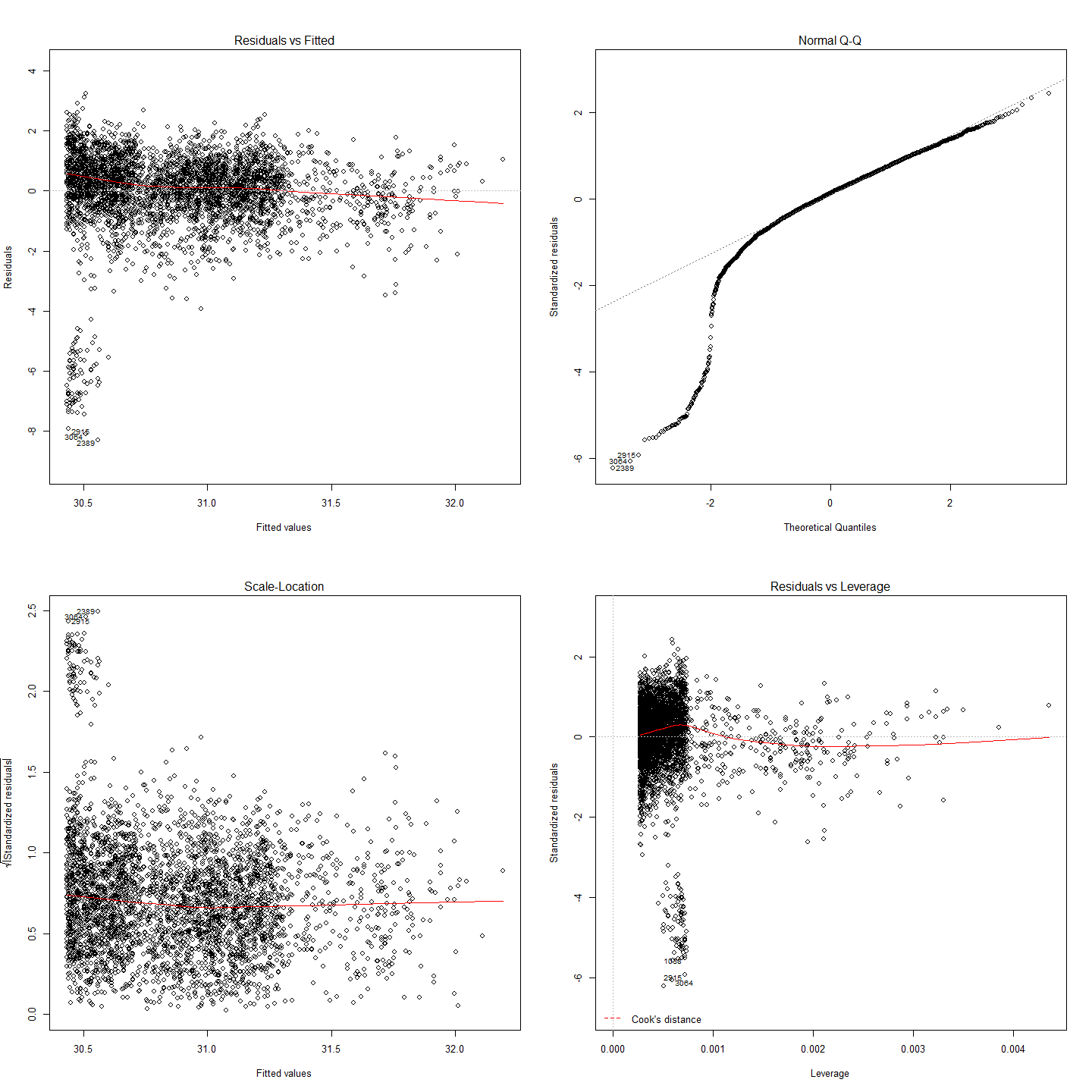
# Correlations with significance levels  
rcorr(GeoDistMathm[lower.tri(GeoDistMathm)],MCdm[lower.tri(MCdm)])#(x, type="pearson") # type can be pearson or spearman

## x y  
## x 1.00 0.25  
## y 0.25 1.00  
##   
## n= 3655   
##   
##   
## P  
## x y   
## x 0  
## y 0

#ANOVA  
anova(reg)

## Analysis of Variance Table  
##   
## Response: MCdm[lower.tri(MCdm)]  
## Df Sum Sq Mean Sq F value Pr(>F)   
## GeoDistMathm[lower.tri(GeoDistMathm)] 1 425.4 425.41 237.59 < 2.2e-16 \*\*\*  
## Residuals 3653 6540.8 1.79   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Plot residuals  
par(mfrow=c(2,2))  
plot(reg)



So there is some small but significant correlation between genetic distance and geographic distance in the Murray cod sampled. Isolation by distance is evident, even over this very small spatial scale.

### Mantel Test

mant<-mantel.rtest(as.dist(GeoDistMathm), as.dist(MCdm), nrepet = 9999)  
mant

## Monte-Carlo test  
## Observation: 0.2471171   
## Call: mantel.rtest(m1 = as.dist(GeoDistMathm), m2 = as.dist(MCdm),   
## nrepet = 9999)  
## Based on 9999 replicates  
## Simulated p-value: 1e-04

#Check all is in order  
#as.matrix(GeoDistMathm)[1:5, 1:5]  
#as.matrix(MCdm)[1:5, 1:5]

Based on these results, we can reject the null hypothesis that these two matrices, spatial distance and genetic distance, are unrelated with alpha = 10^{-4}. The observed correlation, 0.2471171, suggests that the matrix entries are positively associated. This means that smaller differences in genotype are generally seen among pairs of larvae that are from nests geographically close to each other, rather than nests which are estimated to be further away from each other. Note that since this test is based on random permutations, the code will always arrive at the same observed correlation but rarely at exactly the same p-value.

The most likely positon of nests can be estimated based on the larval dispersal velocity can now be mapped onto the river topology.

AMOVA: compare catch, nets sites. Also is it suitable for chem data too?

## Siblings and What they can tell us about Larval Dispersal

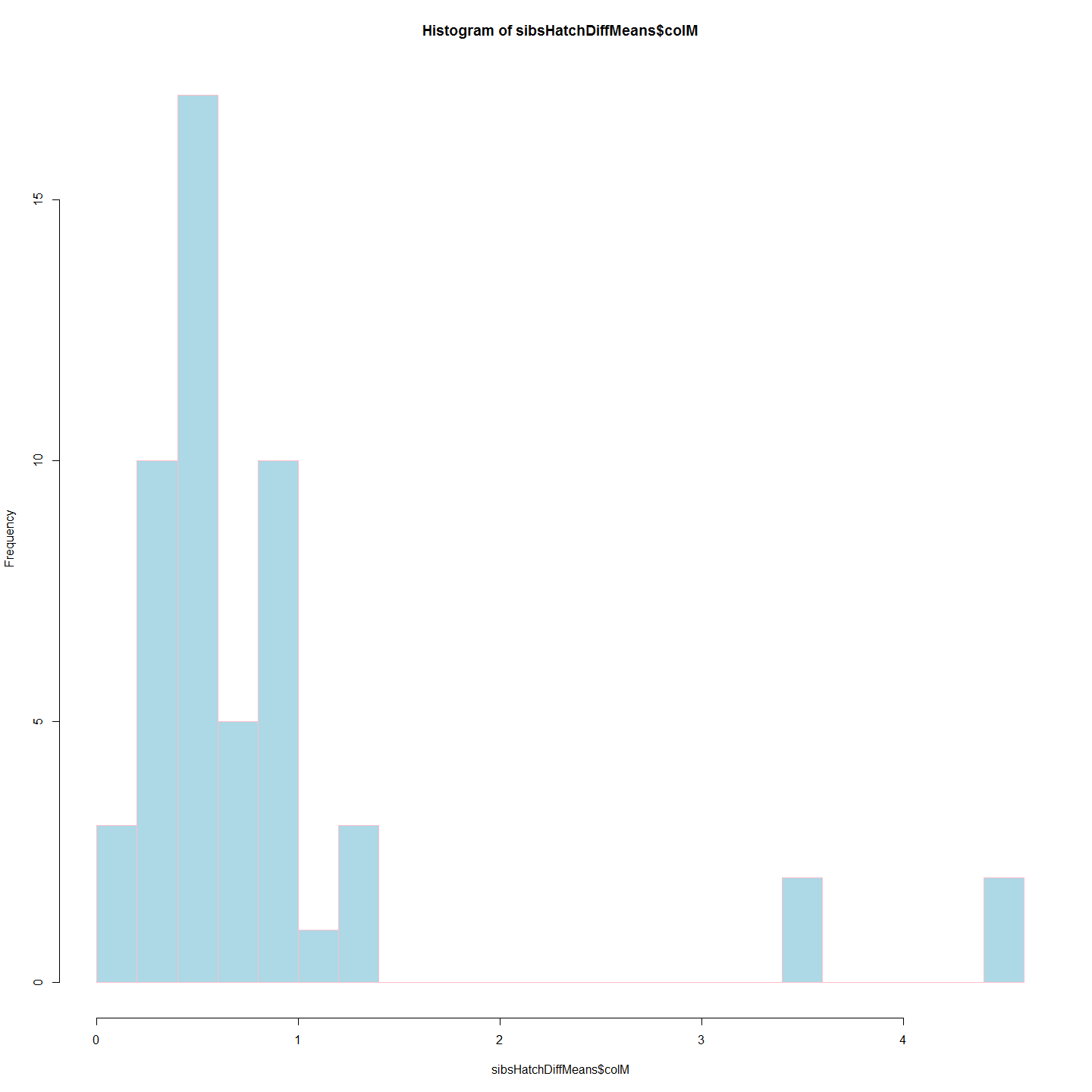
The identification of various sibling pairs and groups in the data (they can be seen in the bottom left of the genetic distance v. distance scatterplot) can potentially be useful in exploring some angles of larval dispersal. For example:

* do siblings dispere in schools or alone?
* How long does the hatch period in the wild last? It is known in hatcheries to last x days but not in the wild. Siblings also facilitate accurate allocation of larvae to nest because, by definition, they come from the same nest. If siblings appear between years then this would suggest repeat adult pairings over years. To date we do not know if this happens.

## This file is to examine the siblings collected that are known to be siblings from DaRT sequencing. The idea is to learn:  
# 1. What duration is hatching in wild nests?   
# 2. Create a list of larvae with siblings (withSibs) and maybe (withoutSibs?) for checking  
# 3. Create a logical matrix of siblinf relationships to re-use (sibsLogical)  
  
#These data can then be used to inform Nest assignments in cases where clade and location are the same.  
  
#Line to extract subset of DF which include at least one sibling in the dataframe  
sibsSubset<-subset(MCdmForSibs, apply(MCdmForSibs, 1, function(MCdmForSibs){any(MCdmForSibs > 0 & MCdmForSibs < 26.5)}))  
#The siblings have a dissimilarity index of less than 26.5. This can be seen in the dendrogram and in the scatterplot for IBD.  
  
sibsLogical<-sibsSubset<26.5 #a logical DF to record T and F for sibling pairs  
sibsSubset<-sibsLogical\*sibsSubset #This turns to 0 all the genetic similarity above 26.5 (no-sibs)  
  
sibsSubset<-sibsSubset[!sapply(sibsSubset, function(x) all(x == 0))] #now remove those with no sibs  
sibsLogical<-sibsSubset>0  
  
#Make a list (data frame) with names of those with sibs (same nest)  
withSib<-row.names(sibsSubset) #make a list of those with sibs  
withSibDF<-as.data.frame(withSib)  
row.names(withSibDF)<-withSibDF[,1] # Make sensible row names (larva labels)  
withSibDF<-as.data.frame(withSibDF) #make back to a df  
  
#Elsewhere I have Make a list (data frame) with names of those withOUT sibs (as each means must be a different NestID)  
  
##Now to see how many days between hatch of sibs.  
  
withSib2<-subset(larv, select=c(Label,hatchDoY)) #get labels and hatch DoY  
withSib2<-withSib2[withSib2$Label %in% withSib,] #reduce df to the siblings only  
  
row.names(withSib2)<-withSib2[,1] #make label row names for making distance matrix  
withSib2$Label<-NULL #tidy up redundant  
  
#Then to make a distance matrix.  
withSibDM<-dist(withSib2)  
withSibDM2<-as.matrix(withSibDM)  
  
###sort rows and columns of both before we can do matrix algebra  
withSibDM2<-as.data.frame(withSibDM2)  
withSibDM2$sort<-row.names(withSibDM2)  
withSibDM2 <- withSibDM2[order(withSibDM2$sort),]#sort row order  
withSibDM2$sort<-NULL  
withSibDM2<-withSibDM2[,order(names(withSibDM2))]#sort column order  
withSibDM2<-as.matrix(withSibDM2)  
  
sibsLogical<-as.data.frame(sibsLogical)  
sibsLogical$sort<-row.names(sibsLogical)  
sibsLogical <- sibsLogical[order(sibsLogical$sort),]#sort row order  
sibsLogical$sort<-NULL  
sibsLogical<-sibsLogical[,order(names(sibsLogical))]#sort column order  
sibsLogical<-as.matrix(sibsLogical)  
####  
  
sibsHatchDiff<-sibsLogical\*withSibDM2 # need half the triangle here? as mean lowered otehrwise?  
sibsHatchDiff[sibsHatchDiff == 0] <- NA   
sibsHatchDiffMeans<-as.data.frame(colMeans(sibsHatchDiff, na.rm=TRUE))  
  
#Get rid of NAs  
sibsHatchDiffMeans<-na.omit(sibsHatchDiffMeans)  
  
#Output something  
write.csv(format(sibsHatchDiffMeans), file="sibsHatchDiffMeans.csv")  
  
#Describe and Plot Summary Statistics regading the difference in hatch day-of-year between siblings  
describe(sibsHatchDiffMeans)

## sibsHatchDiffMeans   
##   
## 1 Variables 53 Observations  
## --------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------  
## colMeans(sibsHatchDiff, na.rm = TRUE)   
## n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95   
## 53 0 31 1 0.8492 0.1890 0.2900 0.4033 0.5900 0.8900 1.3300 3.4500   
##   
## lowest : 0.145 0.150 0.215 0.220 0.290, highest: 1.058 1.330 1.388 3.450 4.520   
## --------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

#sibsHatchDiffMeans # was just printed for validation  
names(sibsHatchDiffMeans)[names(sibsHatchDiffMeans)=="colMeans(sibsHatchDiff, na.rm = TRUE)"] <- "colM"  
hist(sibsHatchDiffMeans$colM,breaks = 25, col = "lightblue", border = "pink")



#So the three objective have been achieved.   
# 1. What duration is hatching in wild nests? Description above showns mean etc.  
# 2. Create a list of larvae with siblings (withSibsDF) done  
# 3. Create a logical matrix of siblinf relationships to re-use (sibsLogical) done  
  
#Clean Up  
rm(sibsSubset)  
rm(sibsHatchDiff)  
rm(sibsHatchDiffMeans)  
rm(sibsLogical)  
rm(withSib)  
rm(withSib2)  
rm(withSibDF)  
rm(withSibDM)  
rm(withSibDM2)  
rm(MCdmForSibs)  
  
#END

### Using a sibsCorrectedNestDistance. Can this give us an idea of variance back calculating to dispersal velocity?

### Use freq histogram of age to adjust dispersal velocity?

ie: if most fish are sampled at 12 days, few at 7 or 20 then most of the dispersal happens then (doesn’t it?) so dispersal may go up from 1200 in near mean, and below 1200 at ends of the curve. Therefore the distance calculations can – at least in theory – be adjusted based on their age. If I do that does it change the variance in the distribution of (sibs?) larvae nest positions? Ie: does the estimate improve?

### Parentage Analysis

Colony etc. Page 68 LFC

half sibs: sneaker males to nests?

### AMOVA

* Bernd's data munging etc

### PLD Kernel (LFC p.64)

## Discussion (points only)

* Evidence for local recruitment. Re recruitment to local area.
* IBD quite marked compared to some marine species. See salmon over thousands of kilometres (same r squared) [ref: powerpoint slide] p.65 LFC
* Eel tramps p110 LFC

\*Natal Signatures p80 and p31 LFC

p68 and Gst and F st p72 LFC

## Code Chunks in this Document

## [1] "Project\_Template\_and\_Knitr" "Set\_Global\_Options" "LoadLibraries" "unnamed-chunk-1"   
## [5] "unnamed-chunk-2" "unnamed-chunk-3" "unnamed-chunk-4" "All\_Maccullochella\_Larvae"   
## [9] "Murray\_Cod\_Larvae\_Only" "A2R\_Dendrogram" "ExtractClades" "extractRaceClades"   
## [13] "unnamed-chunk-5" "first" "Distance\_Matrices\_and\_Ordering" "Plots\_and\_Correlation"   
## [17] "Mantel\_Test" "unnamed-chunk-6" "siblingAnalysisGo" "Include\_Chunk\_Labels\_and\_Session Information"  
## [21] "IM" "siblingAnalysis"

## R version 3.2.2 (2015-08-14)  
## Platform: x86\_64-w64-mingw32/x64 (64-bit)  
## Running under: Windows 7 x64 (build 7601) Service Pack 1  
##   
## locale:  
## [1] LC\_COLLATE=English\_Australia.1252 LC\_CTYPE=English\_Australia.1252 LC\_MONETARY=English\_Australia.1252 LC\_NUMERIC=C LC\_TIME=English\_Australia.1252   
##   
## attached base packages:  
## [1] grid stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] ade4\_1.7-2 Hmisc\_3.16-0 Formula\_1.2-1 survival\_2.38-3 lattice\_0.20-33 dendextend\_1.1.0 ape\_3.3 ggdendro\_0.1-15 ggplot2\_1.0.1   
## [10] ProjectTemplate\_0.6 knitr\_1.11   
##   
## loaded via a namespace (and not attached):  
## [1] Rcpp\_0.12.0 formatR\_1.2 RColorBrewer\_1.1-2 plyr\_1.8.3 tools\_3.2.2 rpart\_4.1-10 digest\_0.6.8 evaluate\_0.7.2 nlme\_3.1-121   
## [10] gtable\_0.1.2 yaml\_2.1.13 proto\_0.3-10 gridExtra\_2.0.0 stringr\_1.0.0 cluster\_2.0.3 nnet\_7.3-10 foreign\_0.8-65 rmarkdown\_0.7   
## [19] latticeExtra\_0.6-26 reshape2\_1.4.1 magrittr\_1.5 whisker\_0.3-2 scales\_0.2.5 htmltools\_0.2.6 MASS\_7.3-43 splines\_3.2.2 colorspace\_1.2-6   
## [28] stringi\_0.5-5 acepack\_1.3-3.3 munsell\_0.4.2