Upper Murrumbidgee Larval Cod Dispersal Analysis

Alan Couch

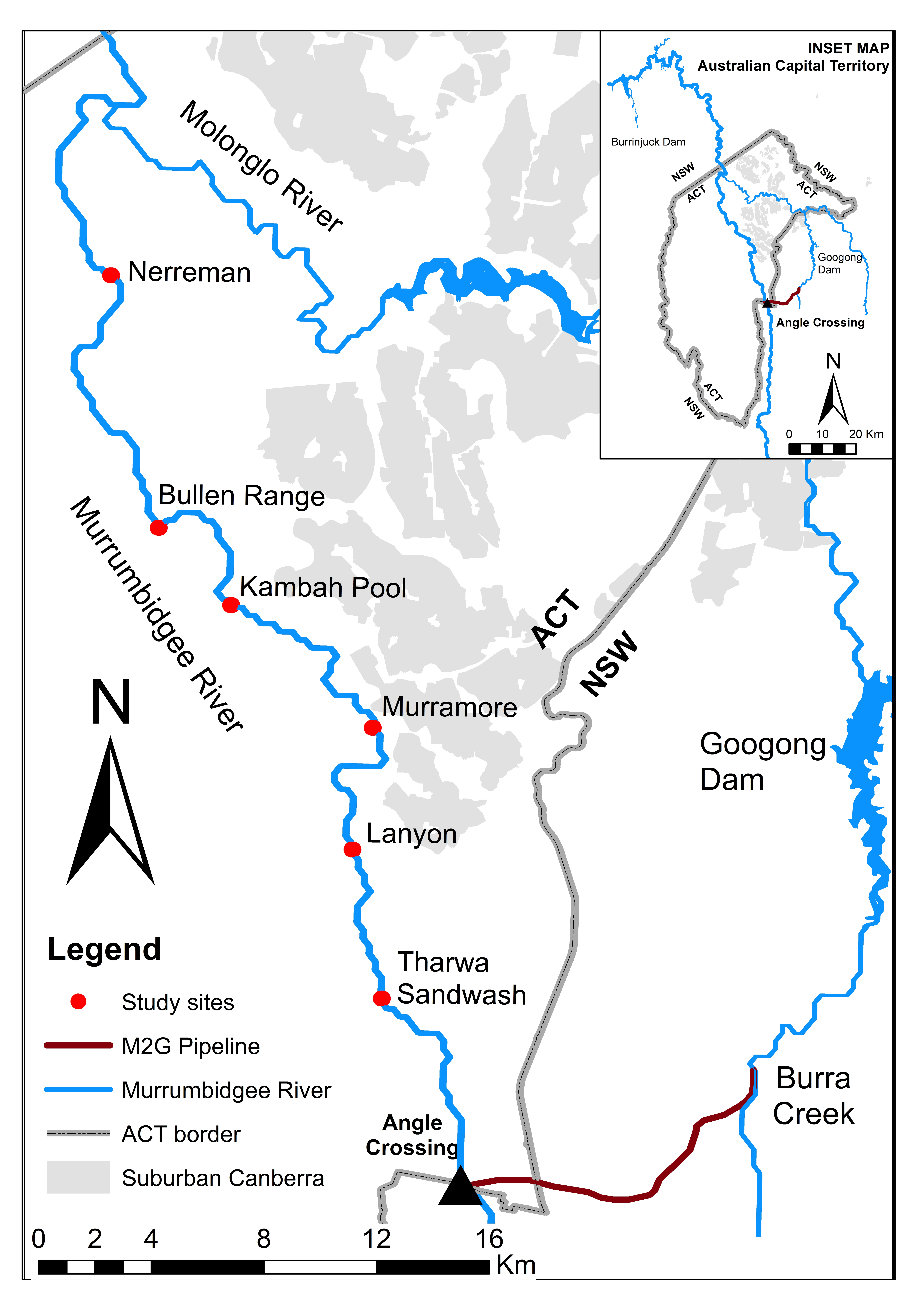
Tue Mar 31 9:58:36 AM 2015

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

This document maps out the analysis of the dispersal of Murray cod larvae in the Murriumbidgee river. It includes sampling, larval morphological characteristics, genome profile and a natural biogeochemical markers. First the relevant r libraries are loaded

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)

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



## Calculate Some Additional Parameters

This section is to add some calculated variables to the data. In particular to add:

Age from Otolith Length: 74.308\*[MeanOtolithLength]-4.44361 (Days)

Hatch date: [Day of Year Caught]-[Age From Otolith Length] (Day of the Year - DoY)

Incubation: 20.67-0.667\*[WaterTemp(DegC) Mean] (Days)

Spawning:[Hatch]-[Incubation] (Day of the Year)

These additional parameters will be used to estimate distances that the larvae have dispersed based on the number of days available to them since leaving the nest and the day and location of collection. In turn the time available will be used to model the most likely distance travelled by the larvae.

larv$ageOL<-74.308\*larv$Mean.Otolith.Length.is.in.Millimetres.for.comparison.with.Adults-4.44361  
larv$hatchdoy<-larv$Day.of.Year-larv$ageOL  
larv$incTime<-20.67-0.667\*larv$WaterTemp.DegC..Mean  
larv$spawn<-larv$hatchdoy-larv$incTime  
  
larv[c(20:30),c(119:122)] #just to see all OK

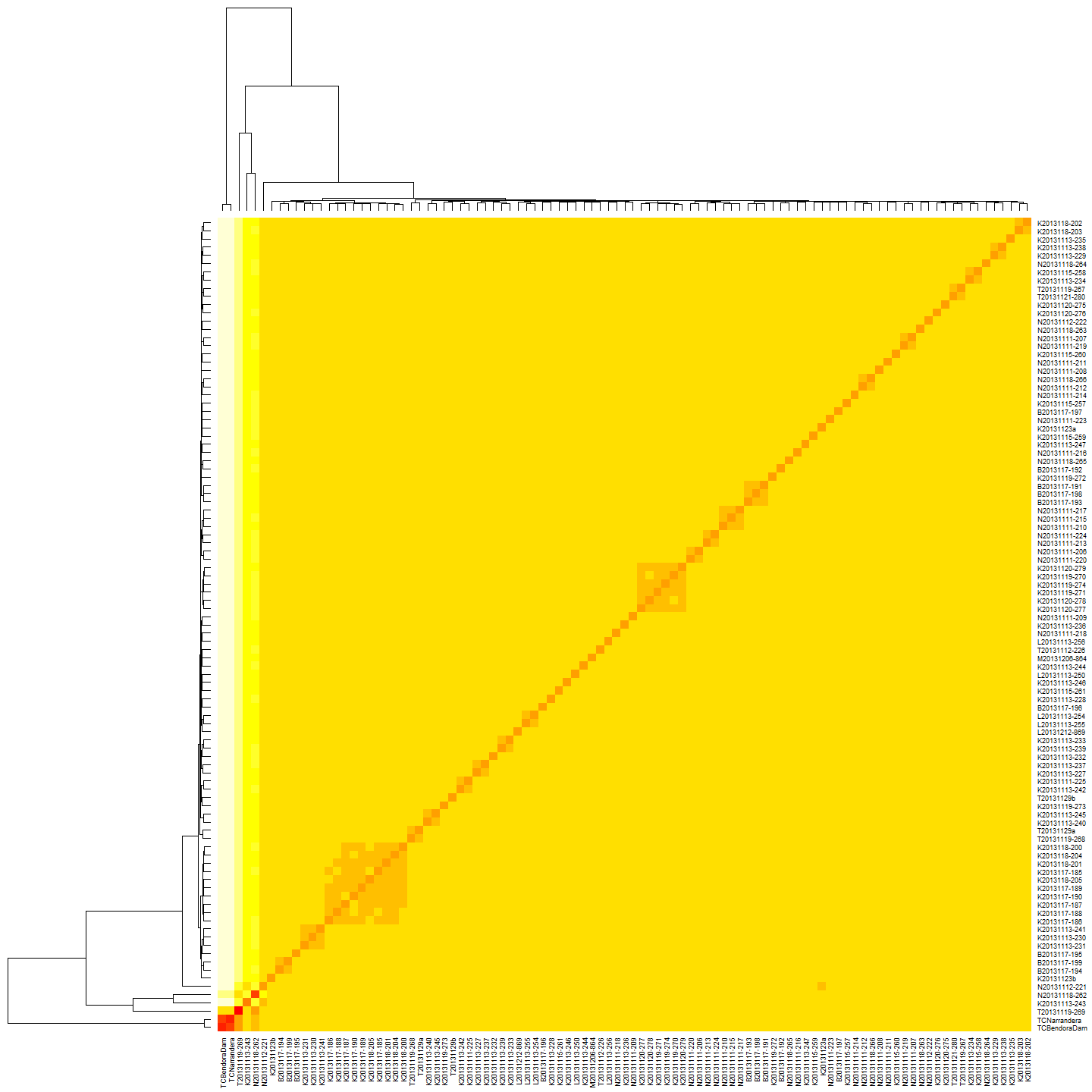
## ageOL hatchdoy incTime spawn  
## B20111111-2 16.36 297.6 6.096 291.5  
## B20111110-70 17.92 296.1 6.096 290.0  
## B20111110-71 17.48 296.5 6.096 290.4  
## B20111110-72 17.11 296.9 6.096 290.8  
## B20111110-73 18.96 295.0 6.096 288.9  
## B20111110-74 17.85 296.2 6.096 290.1  
## B20111110-75 18.96 295.0 6.096 288.9  
## B20111110-76 18.22 295.8 6.096 289.7  
## N20111115-77 18.59 300.4 5.743 294.7  
## N20111115-78 19.33 299.7 5.743 293.9  
## N20111115-79 15.62 303.4 5.743 297.6

But before that we need to establish individual larval genetic identies so that we can see if genetic distance is positively corelated with geographic distance between individuals.

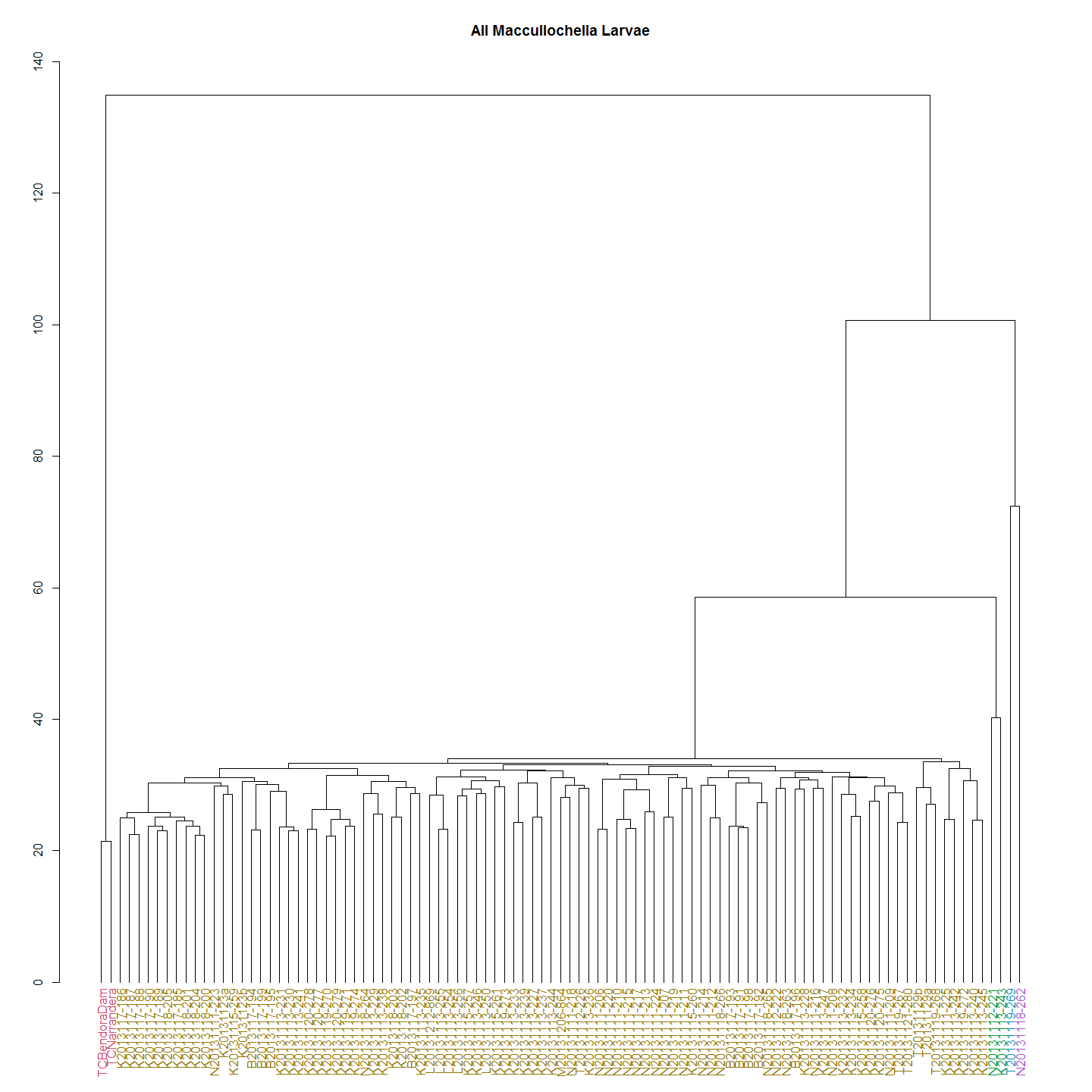
## All Maccullochella Larvae

A dendrogram of all the larvae allows examination of the relationships between all the larvae. In the first instance this is neccessry 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



All the larvae collected fall into one of four distinct clades. The above dendrograms shows 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.

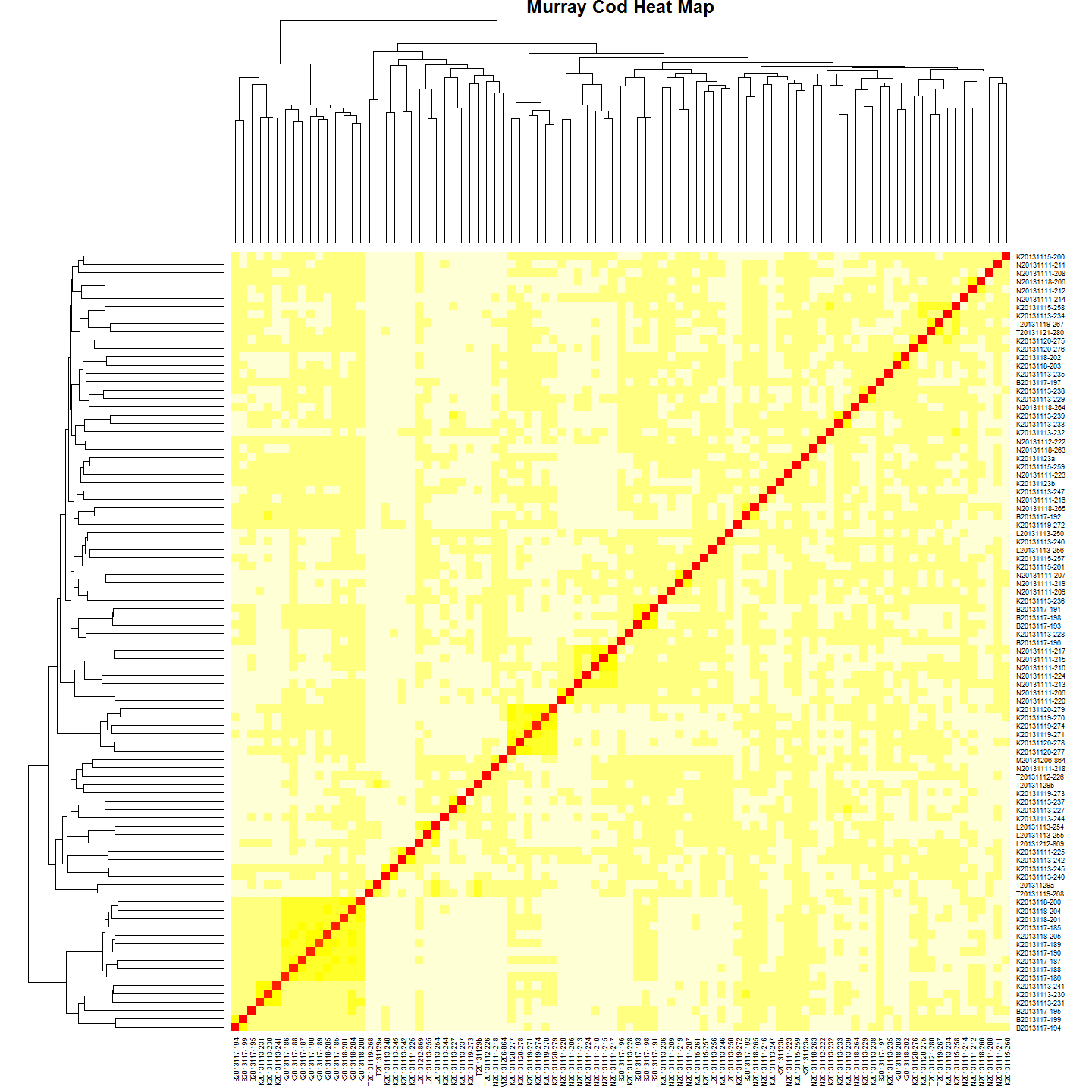
It will be interesting to mito-sequence the trout cod and determine the species of the male and female parent. It is likely that that the female is the Trout cod in the mating pair given the scarcity of trout cod compared with Murray cod and the mate pressure that must exist.

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.

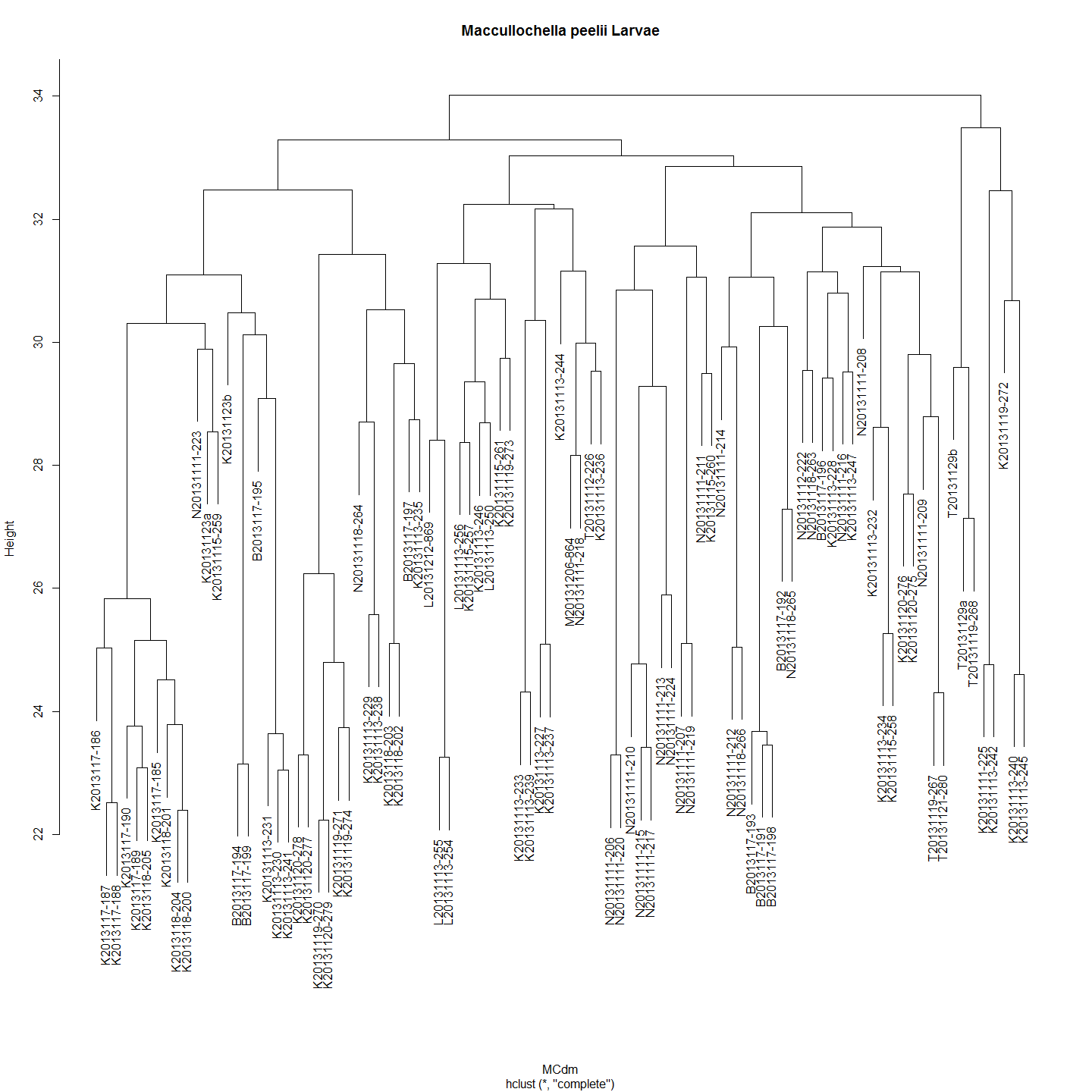
## Murray Cod Larvae

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

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

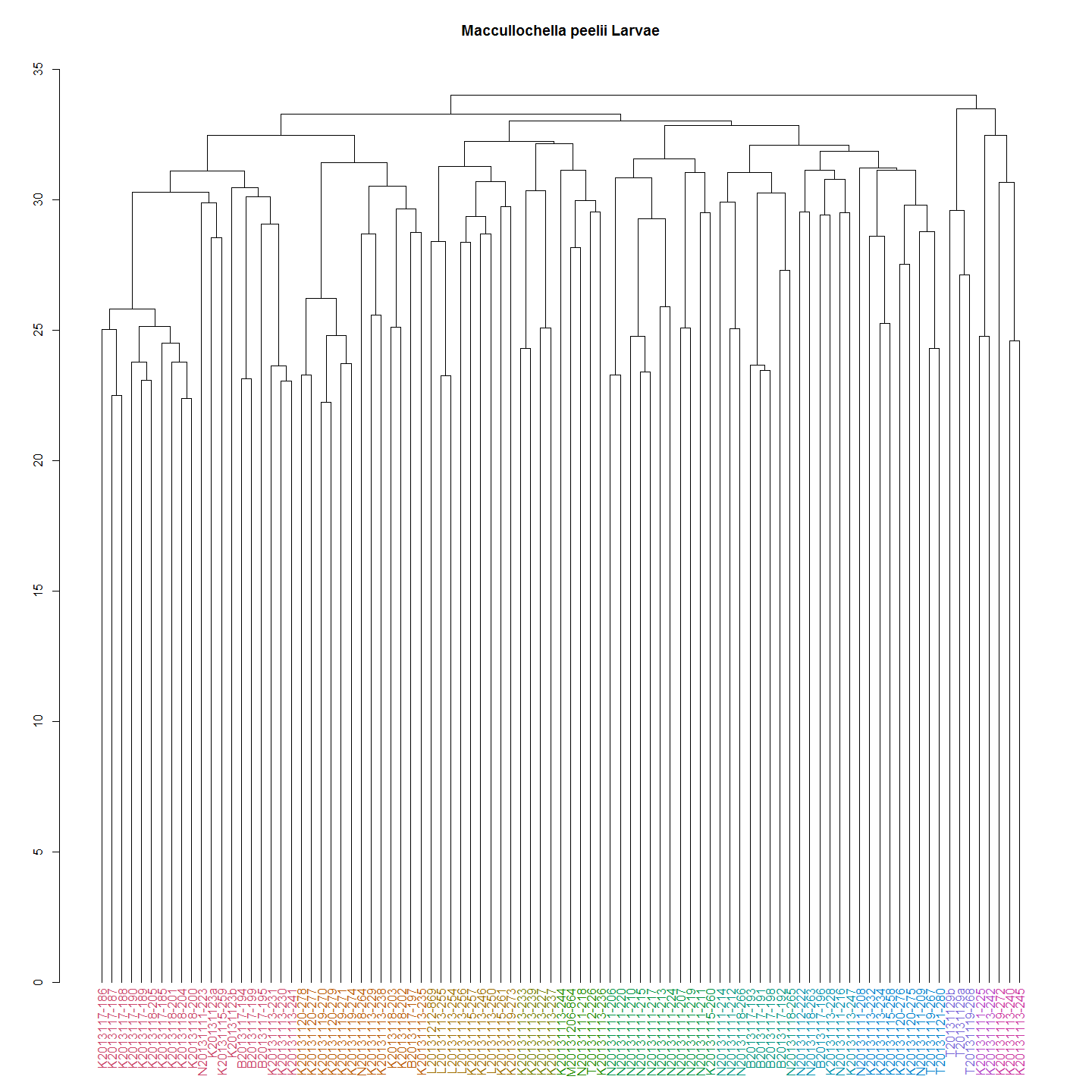


MChc <- hclust(MCdm) #Clutser  
#Plot it  
plot(MChc,main="Maccullochella peelii Larvae")



In most cases very closely related larvae have been collected at the same spot over the same period of a few of days. 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?



The Murray cod larvae can now be resolved into 12 clades.

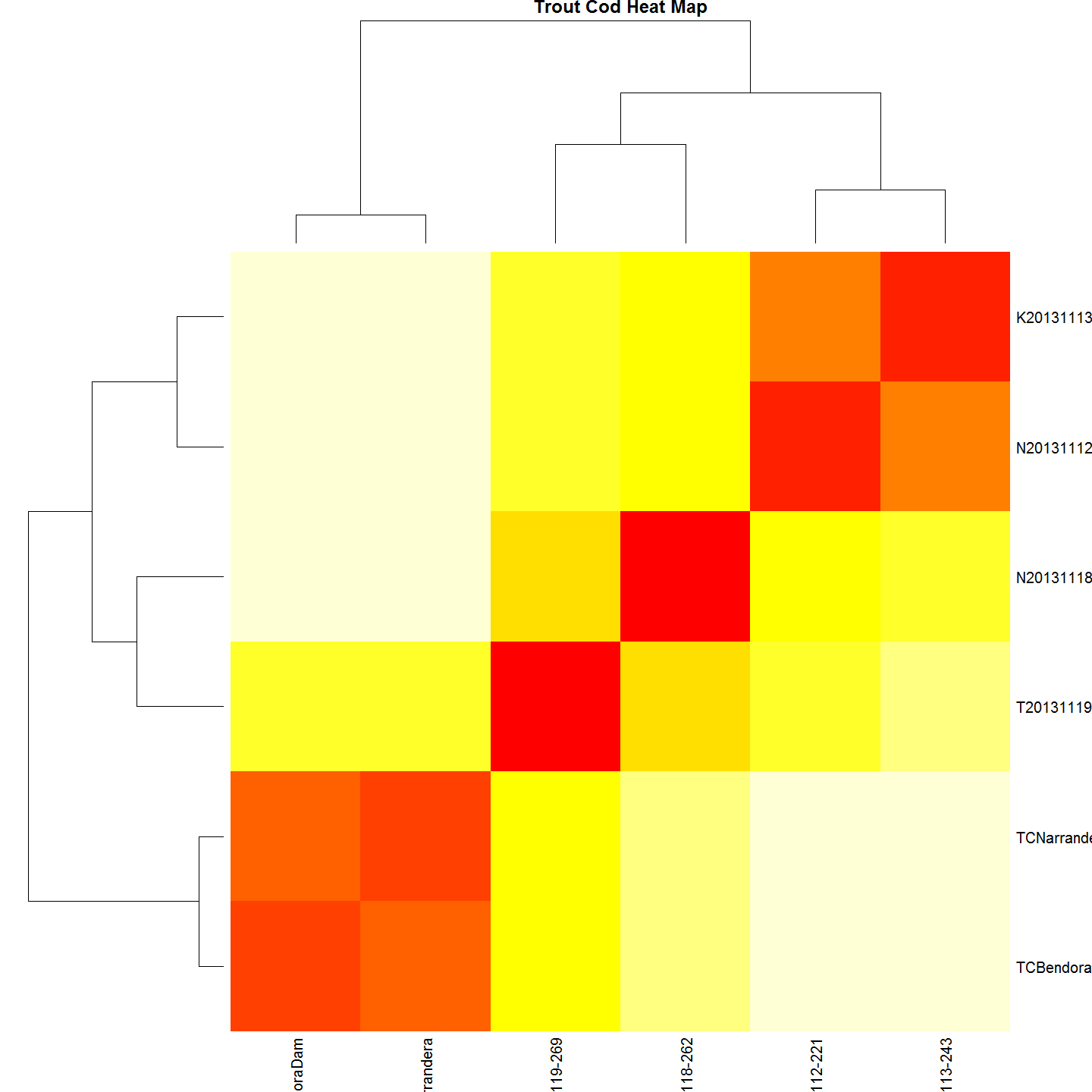
cladeNo<-cutree(dend1,k=12) #This is like using:dendextend:::cutree.dendrogram(dend1,h=70) h or k can be specified  
cladeNo<-as.data.frame(cladeNo)  
#For example:  
cladeNo[c(20:30),] #The clades are numbered by default. So I need to name them something sensible for subsequent analysis.

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

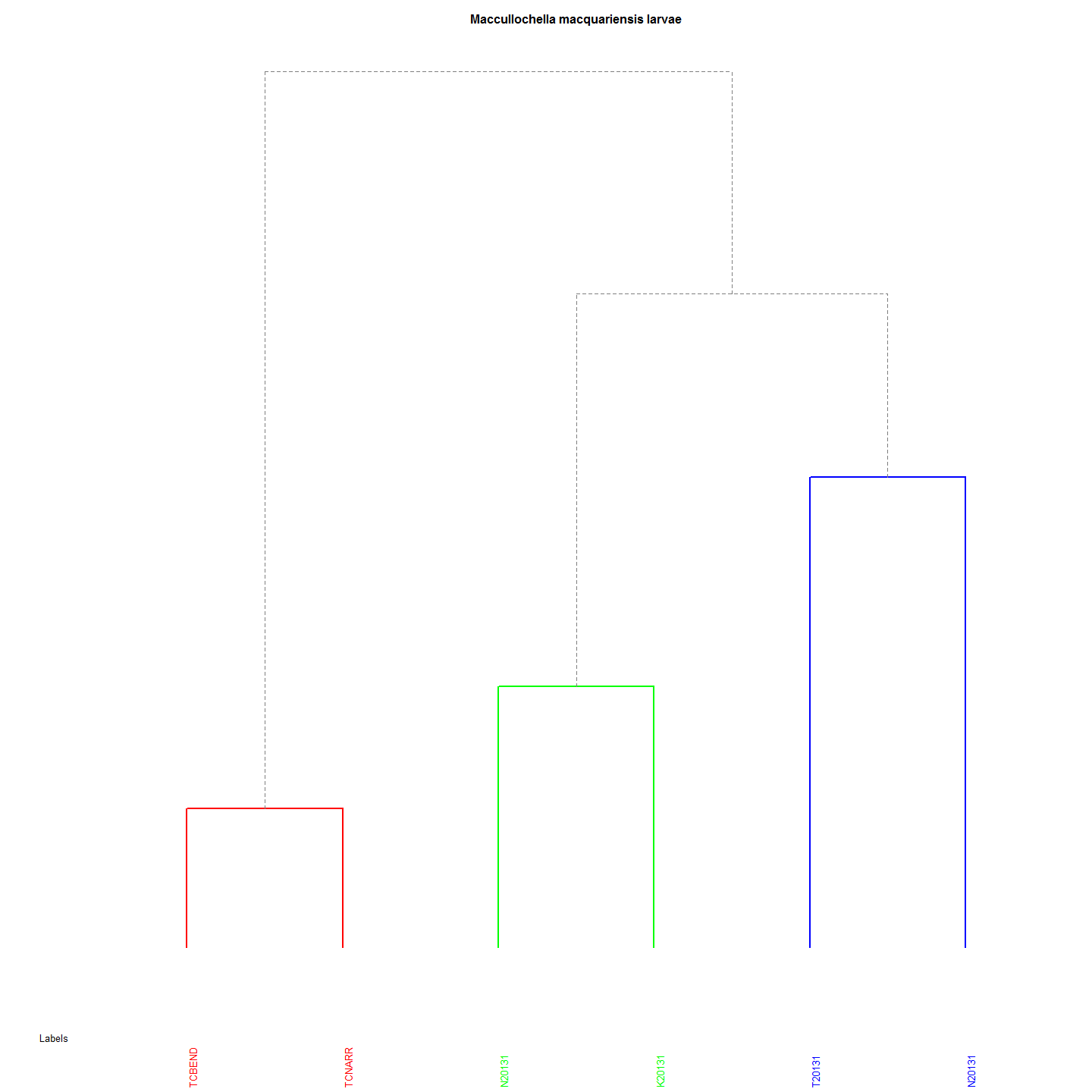
## Trout Cod Larvae

Although Trout cod and hybrid data are excluded from further analysis, the heatmap and dendrogram are provided here for completness. In any case it is not an insignificant finding that the two species do in fact hybridise in this natural riverine environment even though hybrids have been seen in hatcheries and impoundments previoulsy. Furthermore this is the first time fertile F1 have been recorded as evidenced by the finding of an F2 hybrid. That said there is some question as to the providence of the Trout cod in this case as they were believe extirpated and restocked too recently (2006?) for 2 generations to have occured. FACT check this.

#A heatmap and dendrogram for Trout Cod shows three distinct clades.  
TCdm <- dist(TCsnps)  
#Heat map  
dataMatrix <- as.matrix(TCdm)  
heatmap(dataMatrix, main="Trout Cod Heat Map")



#cluster  
TChc <- hclust(TCdm)  
#Plot it  
#plot(TChc, main="Maccullochella macquariensis larvae")  
A2Rplot(TChc, k =3, boxes = FALSE, col.up = "gray50",main = "Maccullochella macquariensis larvae")

 Both the heatmap and dendrogram for Trout Cod shows three distinct clades.

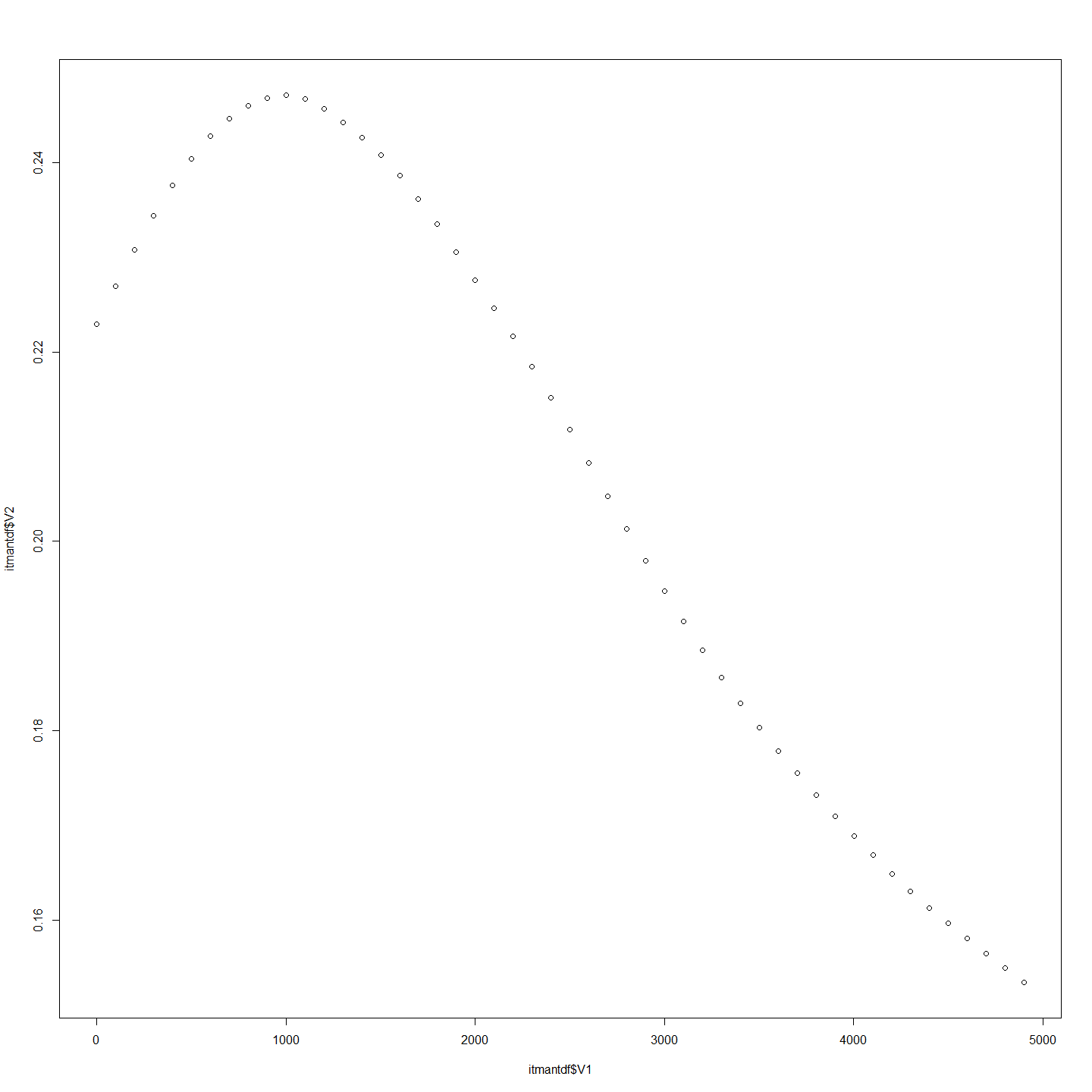
# Upper Murrumbidgee Larval Murray Cod Genetic and Geographic Distances

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

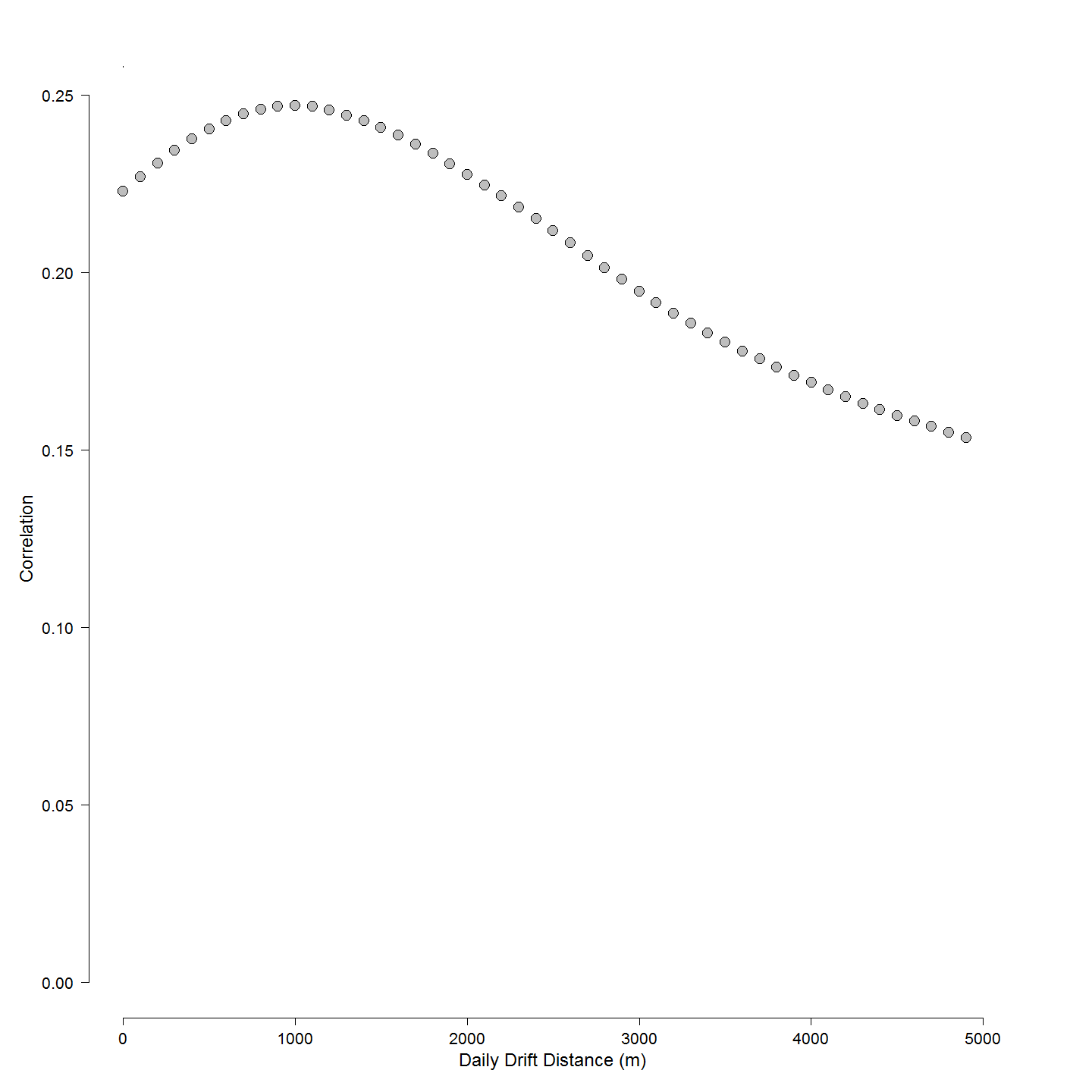
By iterating the mantel test using distance matrices generated for nest distances based on a range of larval 'drift' velocity and the time available to those larvae for 'drifting' calculates the most highly correlated at the asymtote which allows an estimate of the average distance a nest site is from the collection site. The iterations of the Mantel test used 'larval dispersal' velocity range from 1 m to 5000 metres per day available. The highest corellation suggests that it is the distance that best represents 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 drift 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: 74.308\*[MeanOtolithLength]-4.44361  
 #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)))  
#  
#  
  
larv$ageOL<-74.308\*larv$Mean.Otolith.Length.is.in.Millimetres.for.comparison.with.Adults-4.44361  
larv$hatchdoy<-larv$Day.of.Year-larv$ageOL  
larv$incTime<-20.67-0.667\*larv$WaterTemp.DegC..Mean  
larv$spawn<-larv$hatchdoy-larv$incTime  
#larv$nestdist<-larv$Distance.to.Angle.Crossing..m.-(0\*(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)

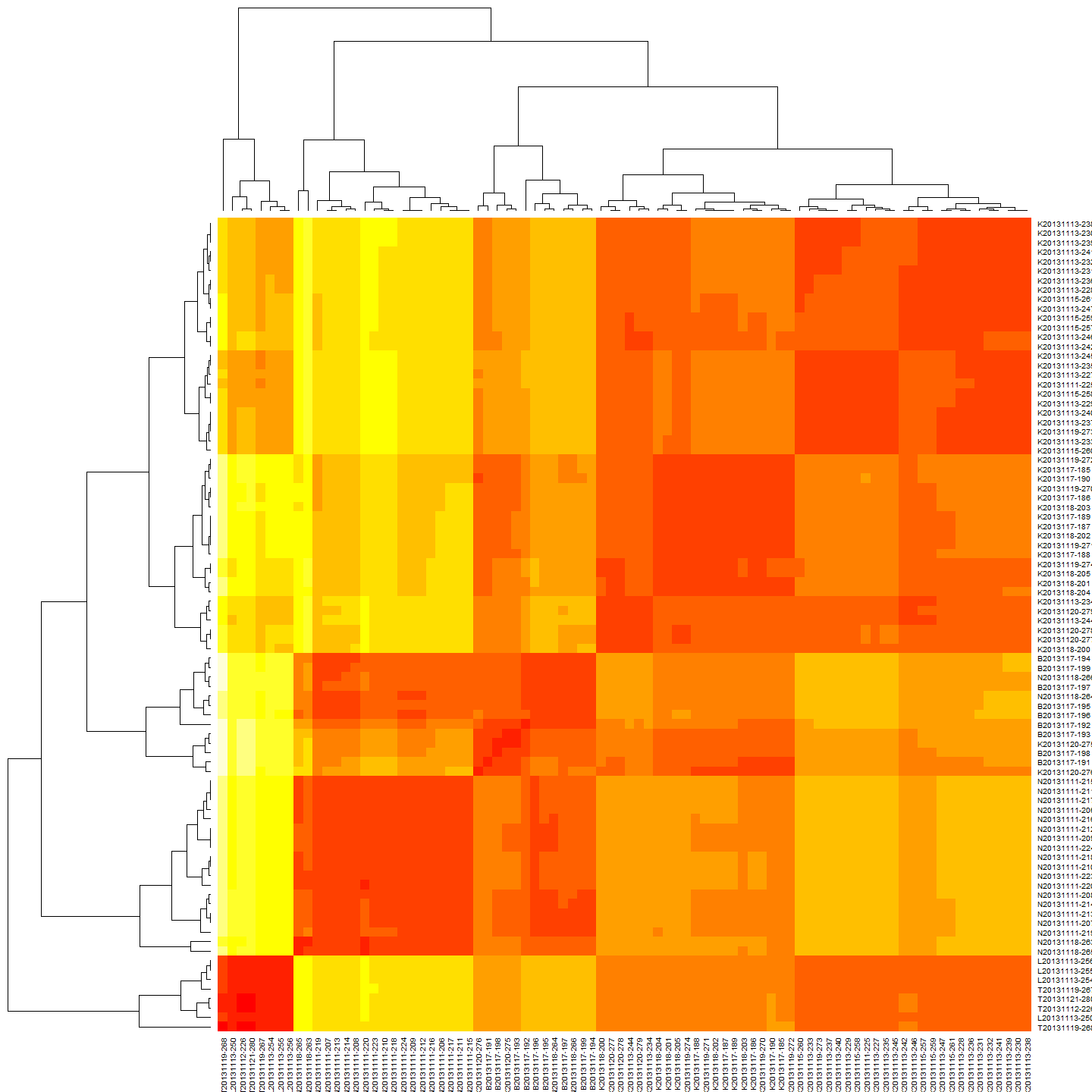


###########  
  
BestNestEst<- subset(itmantdf, V2==max(V2) , select = V1)  
BestNestEst<-as.numeric(BestNestEst)

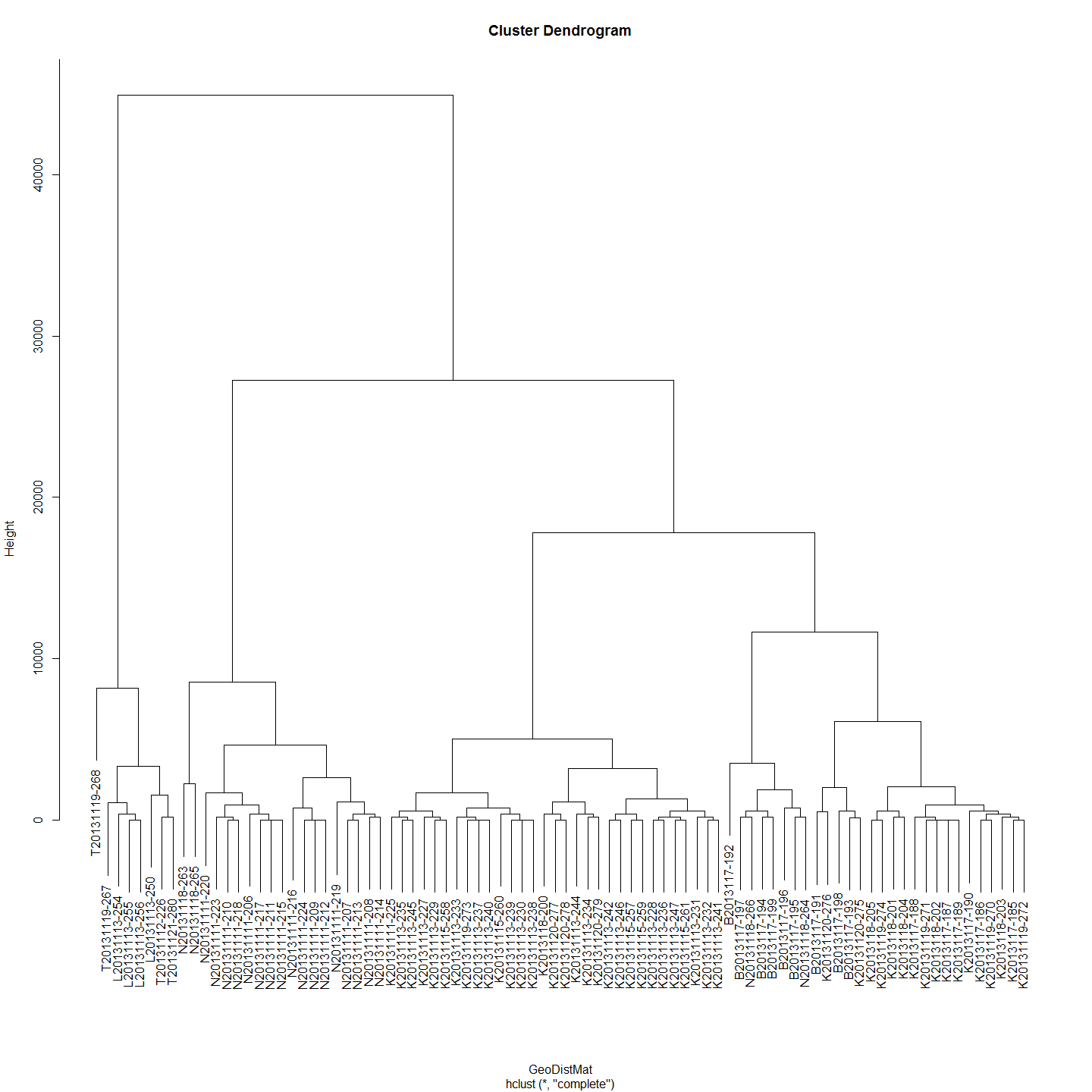
In this case it is the distance above the collection site if the larvae 'drift' 1001 metres per day between leaving the nest and being collected at sampling site.

Of course it is important to keep in mind this is an average for larvae along the entire river reach. It is most likely a distribution of larvae drifting 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 brrod care)  
larv$nestdist<-larv$Distance.to.Angle.Crossing..m.-(BestNestEst\*(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  
###########  
  
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)  
  
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)  
  
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.

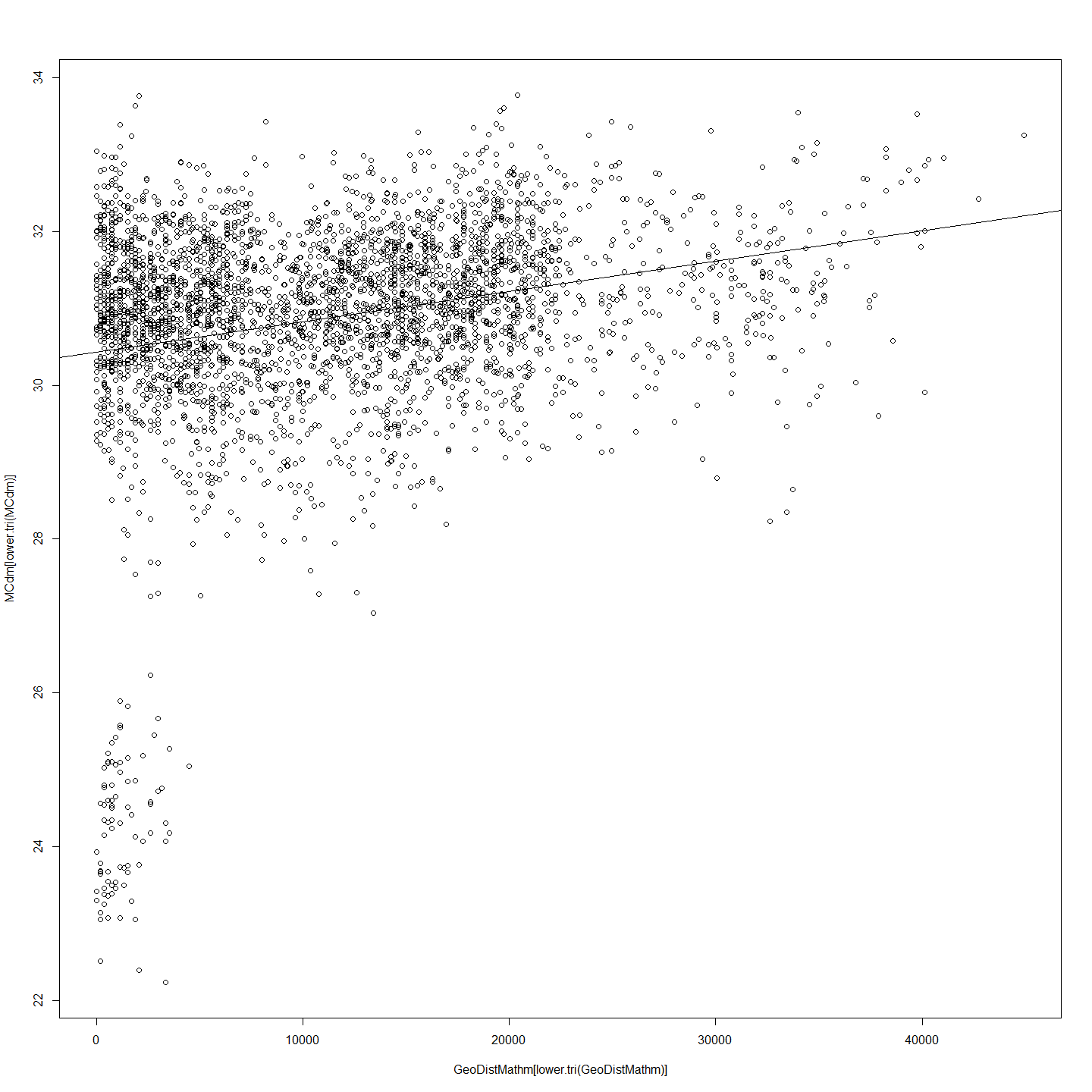
## Plot and Correlate genetic and geographic distance matrices

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)

##   
## Call:  
## lm(formula = MCdm[lower.tri(MCdm)] ~ GeoDistMathm[lower.tri(GeoDistMathm)])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.324 -0.472 0.184 0.757 3.256   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.04e+01 3.65e-02 834.8 <2e-16 \*\*\*  
## GeoDistMathm[lower.tri(GeoDistMathm)] 3.95e-05 2.56e-06 15.4 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.34 on 3653 degrees of freedom  
## Multiple R-squared: 0.0611, Adjusted R-squared: 0.0608   
## F-statistic: 238 on 1 and 3653 DF, p-value: <2e-16

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



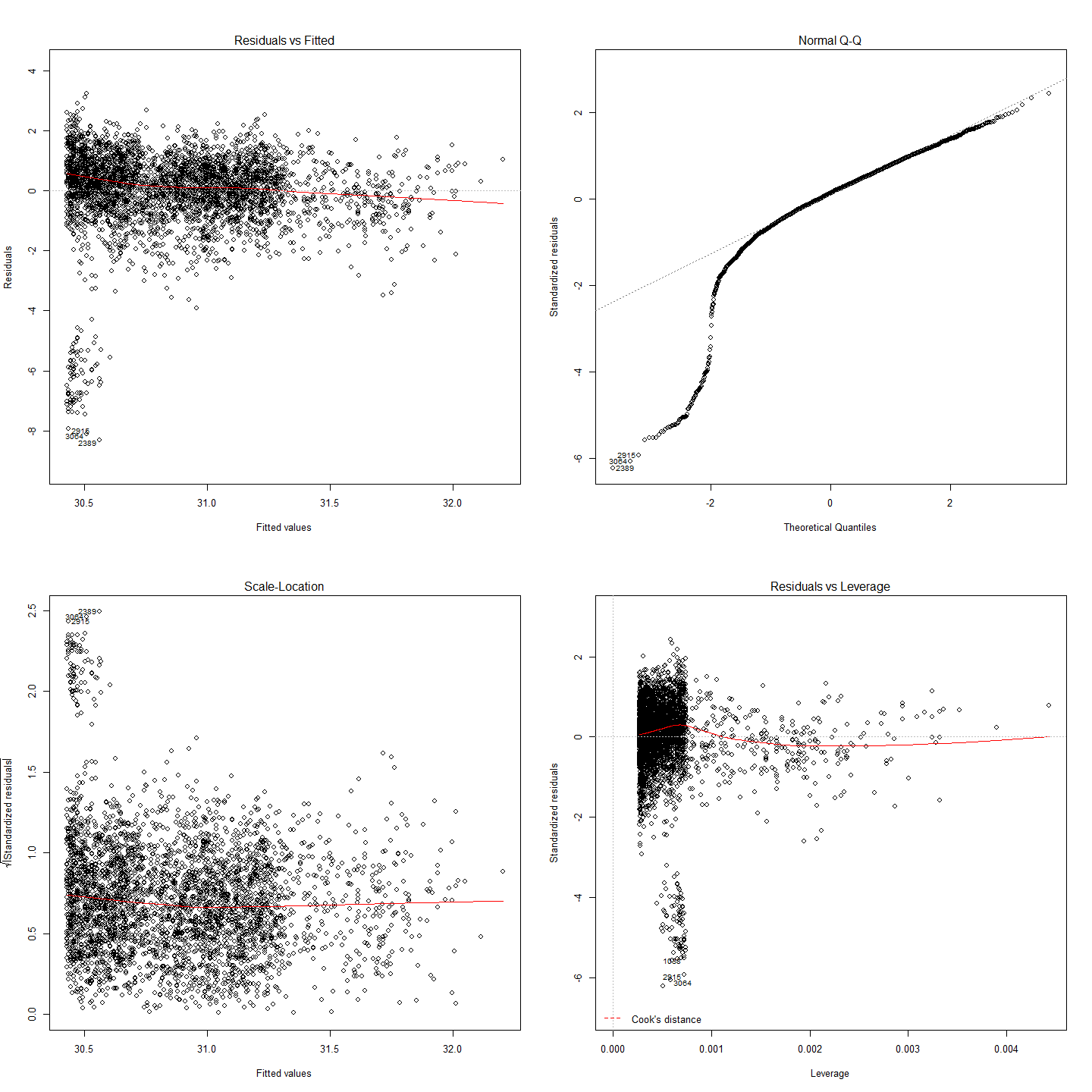
# Correlations with significance levels  
rcorr(GeoDistMathm[lower.tri(GeoDistMathm)],MCdm[lower.tri(MCdm)])#(x, type="pearson") # type can be pearson or spearman

## 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)] 1 425 425 238 <2e-16 \*\*\*  
## Residuals 3653 6541 2   
## ---  
## 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.

## Mantel Test

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

## Monte-Carlo test  
## Observation: 0.2471   
## 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]

## B2013117-191 B2013117-192 B2013117-193 B2013117-194 B2013117-195  
## B2013117-191 0 3533 1488 5764.6 6322.5  
## B2013117-192 3533 0 2046 2231.5 2789.3  
## B2013117-193 1488 2046 0 4277.0 4834.9  
## B2013117-194 5765 2231 4277 0.0 557.9  
## B2013117-195 6322 2789 4835 557.9 0.0

as.matrix(MCdm)[1:5, 1:5]

## B2013117-191 B2013117-192 B2013117-193 B2013117-194 B2013117-195  
## B2013117-191 0.00 30.08 23.66 30.13 30.33  
## B2013117-192 30.08 0.00 30.05 28.74 30.06  
## B2013117-193 23.66 30.05 0.00 30.12 29.26  
## B2013117-194 30.13 28.74 30.12 0.00 29.87  
## B2013117-195 30.33 30.06 29.26 29.87 0.00

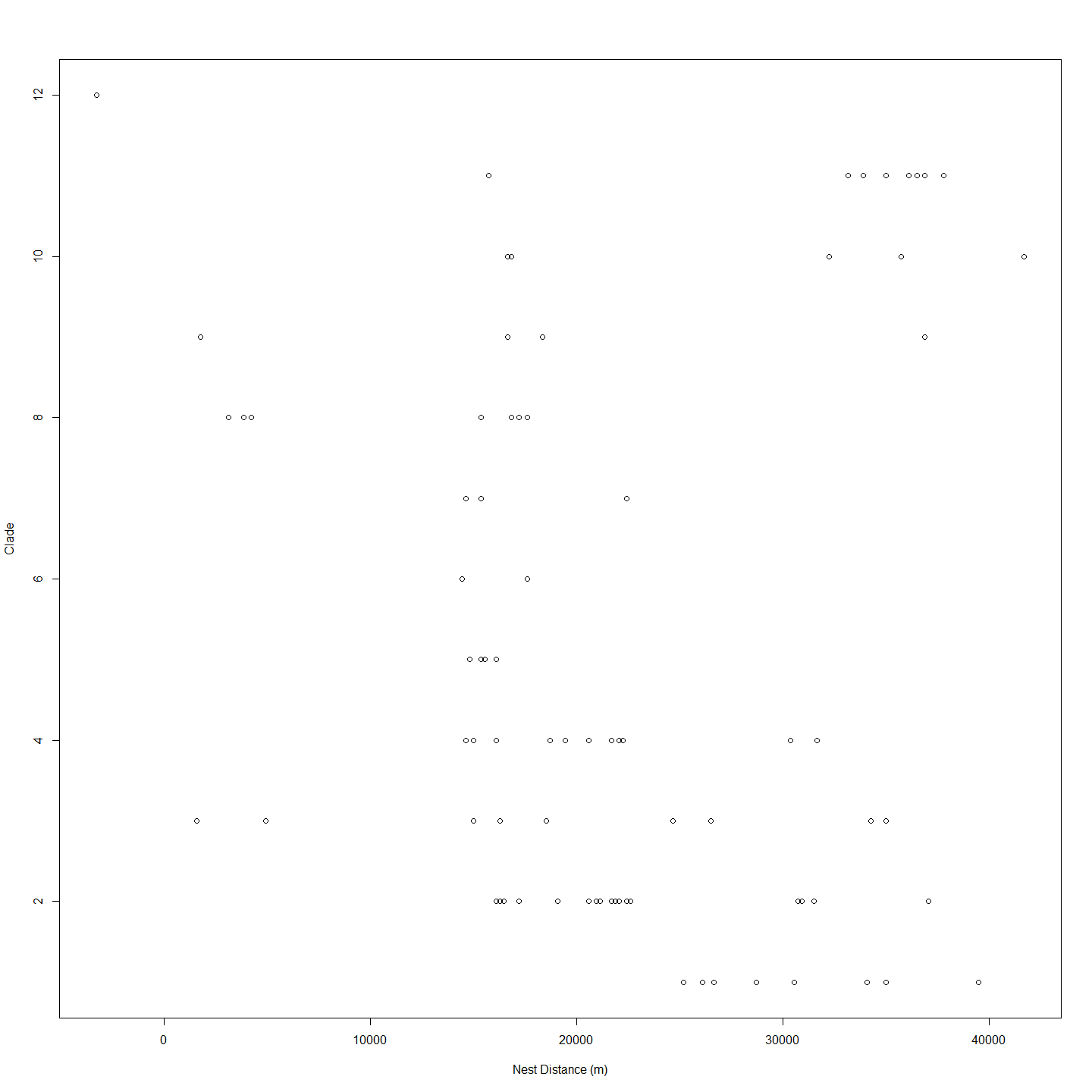
Based on these results, we can reject the null hypothesis that these two matrices, spatial distance and genetic distance, are unrelated with alpha = 10-4. The observed correlation, 0.2471, suggests that the matrix entries are positively associated. This means that smaller differences in genotype are generally seen among pairs of larvae that are from nests estimated to be 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 as estimated based on the best estimate of larval drift can now be mapped onto the river topology.

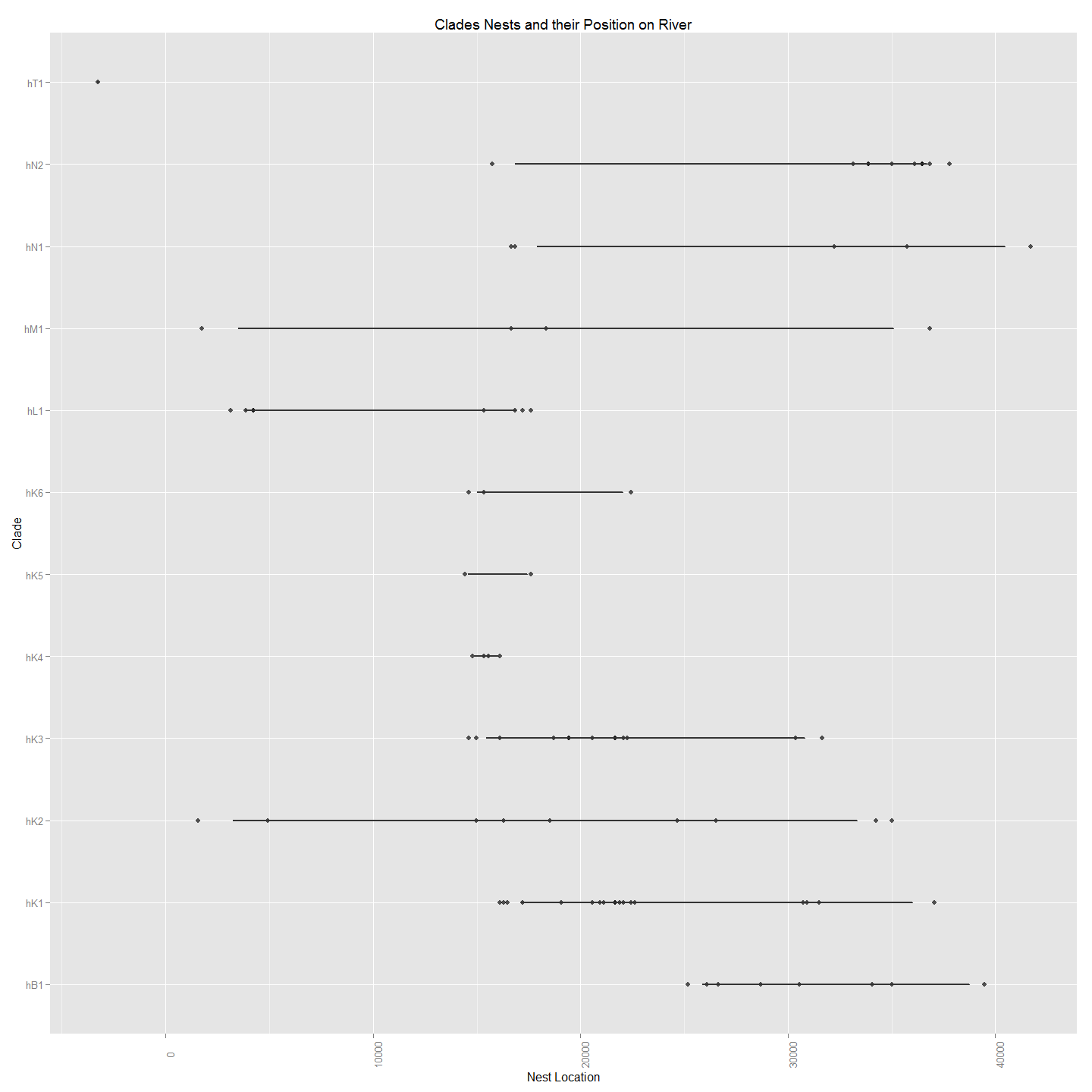
## Look at Clades Over the River Reach

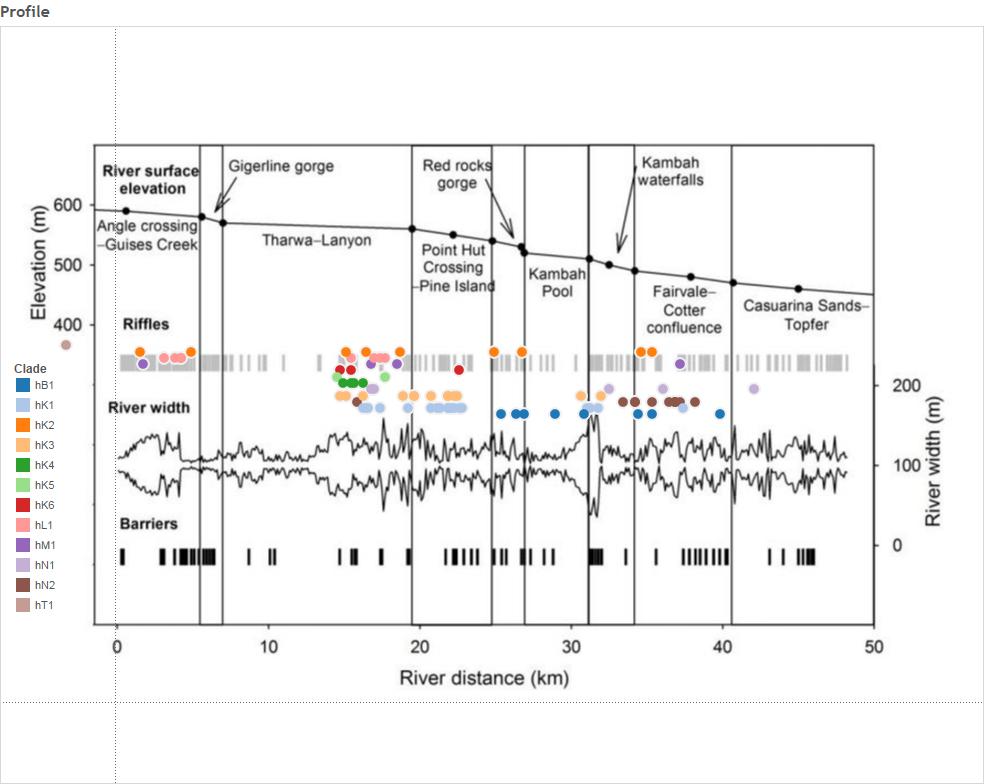
The clades distribution over the river reach suggests some structure.

larv1<-merge(larv,cladeNo, by="row.names")  
rownames(CladeNamesToMerge) <- CladeNamesToMerge[,1]  
CladeNamesToMerge$Label<-rownames(CladeNamesToMerge)  
larv2<-merge(larv1,CladeNamesToMerge, by="Label")  
write.csv(format(larv2), file="larvForTableau.csv", row.names=FALSE)  
plot(larv2$nestdist,larv2$clade, xlab="Nest Distance (m)", ylab = "Clade")



#merged<-merge(Haplogroups,geodist1000, by="row.names")   
#plot(merged$larv.nest,merged$Haplogroups)  
plot2 <- ggplot(larv2, aes(nestdist,clade))  
plot2 + geom\_point(alpha = 2/3) +geom\_boxplot()+labs(title = "Clades Nests and their Position on River")+ labs(x="Nest Location") +labs(y = "Clade")+ theme(axis.text.x=element\_text(angle=90))

 This suggests that all clades except one (hT1) exist below a barrier around 10000m but that only three clades (hM1,hL1 and hK2) exist above and below this barrier. On possible explanation that might be inferred from this is that adults migrating upstream for spawning are prevented from doing so by a barrier between the 5000 and 15000m mark but that larvae produced above the barrier are able to disperse and so are represented downstream. It is noteable that the barrier and the big gap in apparent nests from 6000m to 14000m corresponds with the Tharwa sand slug - a long stretch of sand that has previously been the subject of remedial work because it has been believed to be a barrier to cod migration (ref).



# Natural BioGeochemistry Markers

The first effort is based on the location of collection. Later, if there is a corelation between 'chemotype distance' and geographic distance we will look at the possibility of using an offset downstream of the origin of the mother. It is not likely possible as with larvae we could use larval 'drift' age days to get infrormation about each larvae. We can do this with the mothers but can we do something with the clade groups?

## Identify Important Chemistry Variables for Prediction of Site

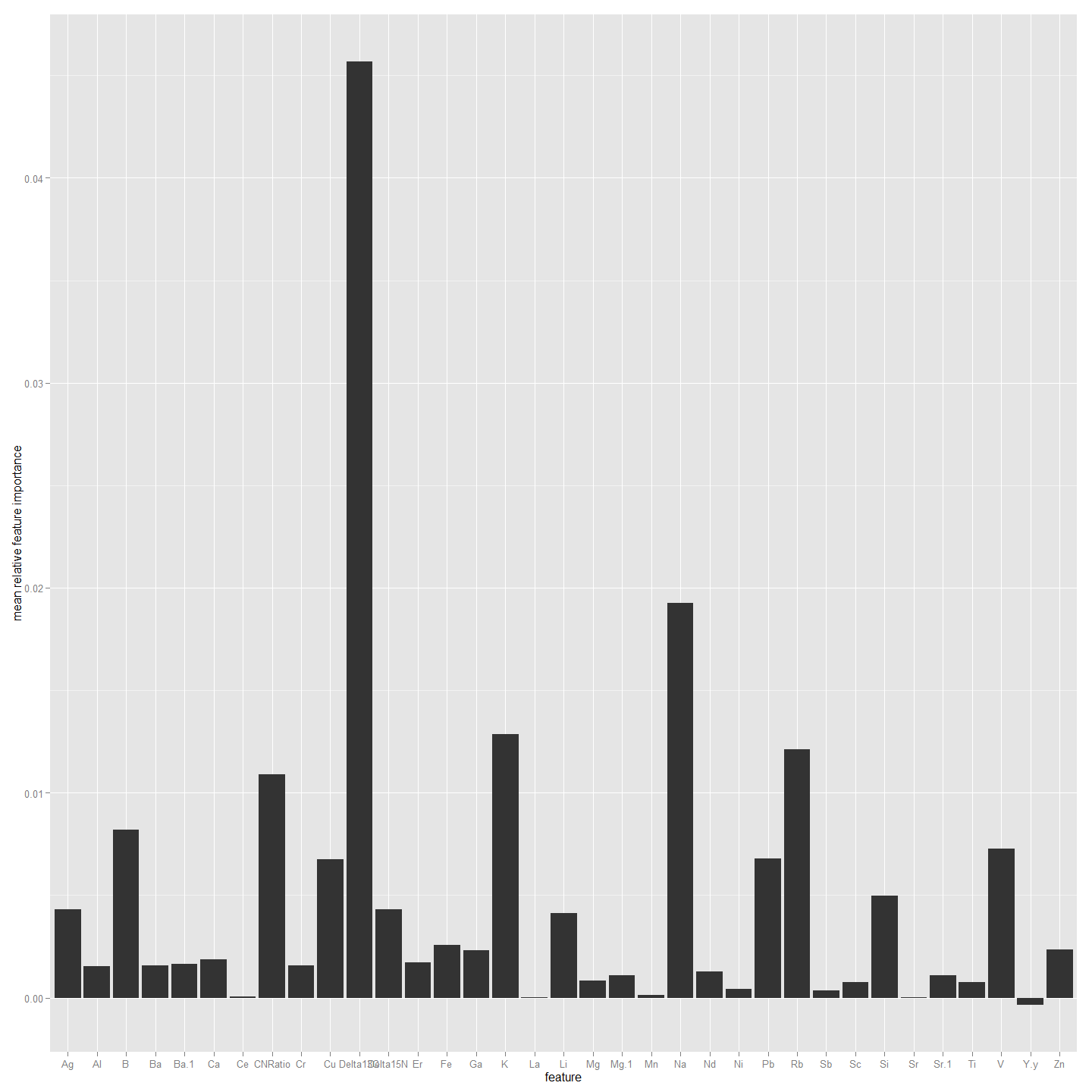
In order to identify important variables in a multivariate dataset one can utilize machine learning methods. There are many different machine learning algorithms for different tasks. One common task is to decide if a feature vector belongs to a certain class. This can be done with a random forest classifier. In order to do so, one has to train the classifier with training data first. Then the classifier can be used to predict the class of other feature vectors.<http://proven-inconclusive.com/blog/machine_learning_methods_to_identify_important_variables.html>. There is no need for other tests, such as cross-validation, to get an unbiased estimate of the test set error as each tree is created with a different bootstrap sample [2].

The classifier saves information on feature importance ("importance=TRUE"). We can use this information in order to identify potentially import variables in the data set.

#This identifies the important Chemistry factors that could be used to predict site the larva are from.  
#First using Core as the values. So need to combine first  
ImportantVars<-ChemAnalCore  
  
ImportantVars<-ImportantVars[c(84,13,14,15,120:151)] #  
ImportantVars<-ImportantVars[complete.cases(ImportantVars),] #remove any nulls  
colnames(ImportantVars)[1]<-"SiteName"  
ImportantVars$SiteName<-as.factor(ImportantVars$SiteName)  
  
ImportantVars <- droplevels(ImportantVars)#Not sure why get error without this line.  
  
library(randomForest)  
forest <- randomForest(SiteName ~.,data=ImportantVars, importance=TRUE)  
  
forest

##   
## Call:  
## randomForest(formula = SiteName ~ ., data = ImportantVars, importance = TRUE)   
## Type of random forest: classification  
## Number of trees: 500  
## No. of variables tried at each split: 5  
##   
## OOB estimate of error rate: 21.05%  
## Confusion matrix:  
## Bullen Range Kambah Pool Lanyon Nerreman Tharwa Sandwash claserror  
## Bullen Range 2 5 0 1 0 0.7500  
## Kambah Pool 0 43 0 0 0 0.0000  
## Lanyon 0 2 0 1 0 1.0000  
## Nerreman 0 4 0 15 0 0.2105  
## Tharwa Sandwash 0 2 0 1 0 1.0000

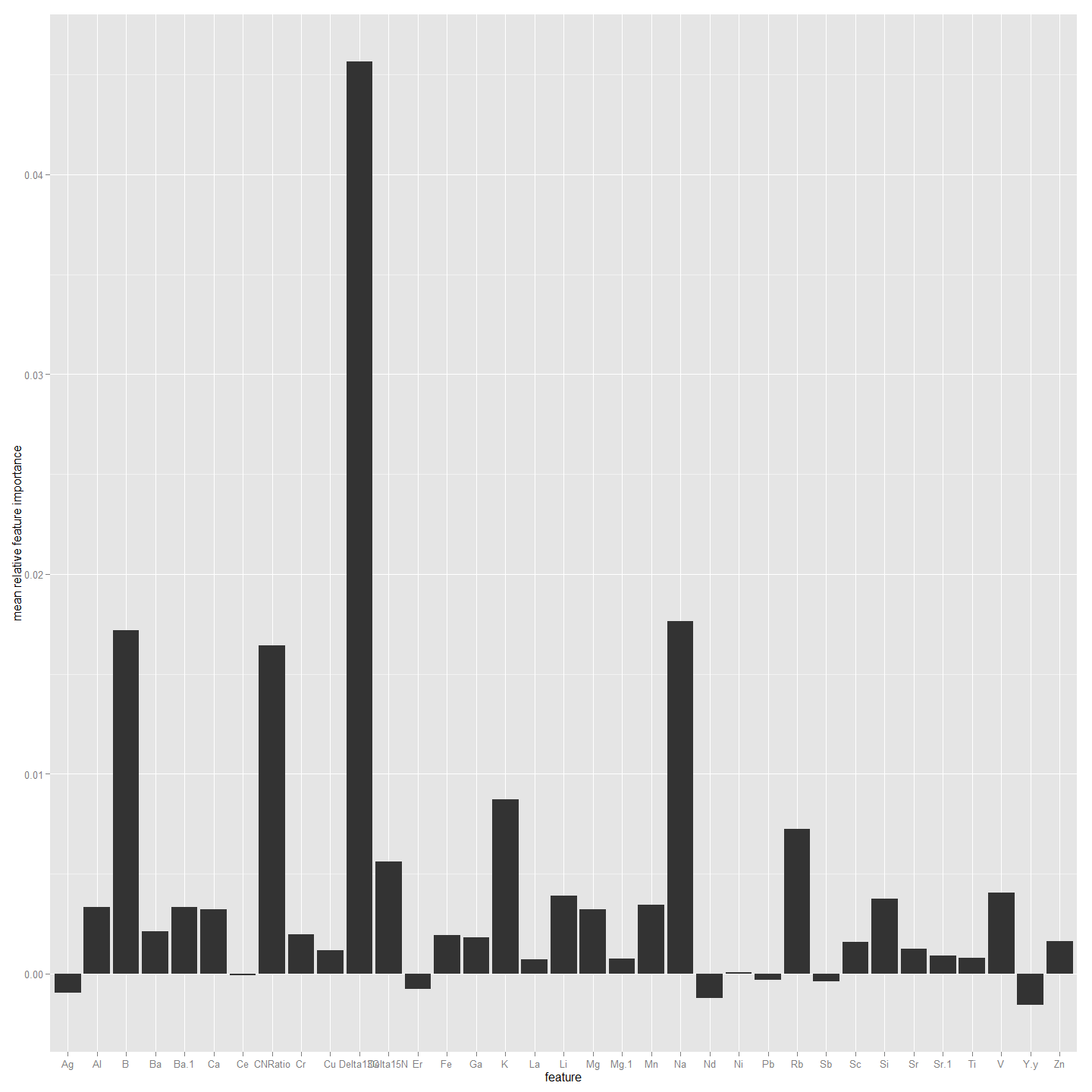
library(ggplot2)  
forest.importance = as.data.frame(importance(forest, scale=FALSE))  
forest.importance = forest.importance[,1:(ncol(forest.importance)-2)]  
forest.importance$mean = rowMeans(forest.importance)  
  
#forest.importance  
  
ggplot(forest.importance, aes(x=row.names(forest.importance), y=mean)) +  
 ylab('mean relative feature importance') +  
 xlab('feature') +  
 geom\_bar(stat='identity')



##################################################################################  
#Now to do the same but using Edge1 as the values. So need to combine first  
ImportantVars<-ChemAnalE1  
  
ImportantVars<-ImportantVars[c(97,13,14,15,120:151)] #  
ImportantVars<-ImportantVars[complete.cases(ImportantVars),] #remove any nulls  
colnames(ImportantVars)[1]<-"SiteName"  
ImportantVars$SiteName<-as.factor(ImportantVars$SiteName)  
  
ImportantVars <- droplevels(ImportantVars)#Not sure why get error without this line.  
  
library(randomForest)  
forest <- randomForest(SiteName ~.,data=ImportantVars, importance=TRUE)  
  
forest

##   
## Call:  
## randomForest(formula = SiteName ~ ., data = ImportantVars, importance = TRUE)   
## Type of random forest: classification  
## Number of trees: 500  
## No. of variables tried at each split: 5  
##   
## OOB estimate of error rate: 22.78%  
## Confusion matrix:  
## Bullen Range Kambah Pool Lanyon Nerreman Tharwa Sandwash classerror  
## Bullen Range 3 4 0 2 0 0.6667  
## Kambah Pool 0 43 0 0 0 0.0000  
## Lanyon 0 3 0 1 0 1.0000  
## Nerreman 0 5 0 15 0 0.2500  
## Tharwa Sandwash 0 2 0 1 0 1.0000

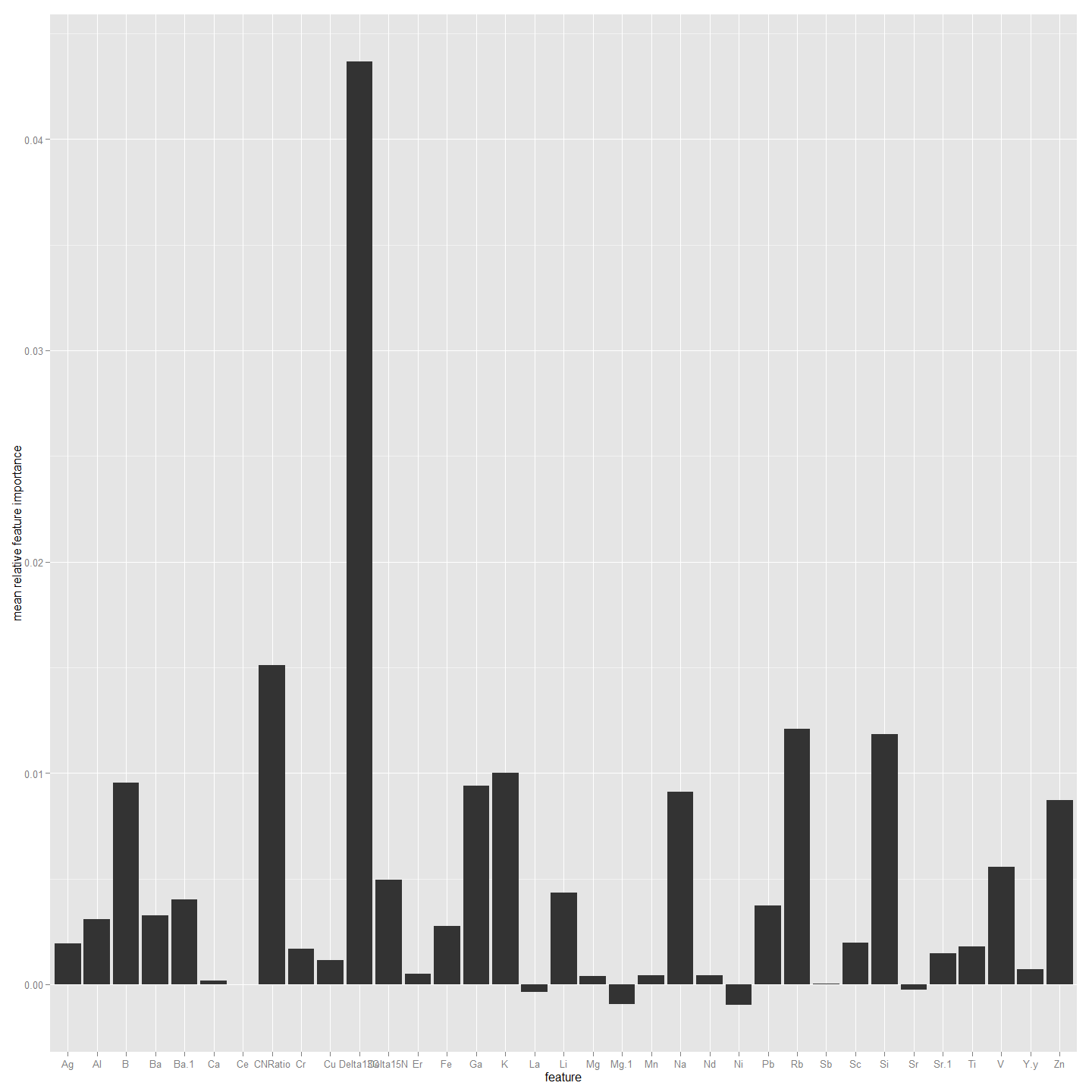
library(ggplot2)  
forest.importance = as.data.frame(importance(forest, scale=FALSE))  
forest.importance = forest.importance[,1:(ncol(forest.importance)-2)]  
forest.importance$mean = rowMeans(forest.importance)  
  
#forest.importance  
  
ggplot(forest.importance, aes(x=row.names(forest.importance), y=mean)) +  
 ylab('mean relative feature importance') +  
 xlab('feature') +  
 geom\_bar(stat='identity')



################################################################################################  
#Now to do the same but using Edge2 as the factor rather than SiteName. So need to combine first  
ImportantVars<-ChemAnalE2  
  
ImportantVars<-ImportantVars[c(97,13,14,15,120:151)] #  
ImportantVars<-ImportantVars[complete.cases(ImportantVars),] #remove any nulls  
colnames(ImportantVars)[1]<-"SiteName"  
ImportantVars$SiteName<-as.factor(ImportantVars$SiteName)  
  
ImportantVars <- droplevels(ImportantVars)#Not sure why get error without this line.  
  
library(randomForest)  
forest <- randomForest(SiteName ~.,data=ImportantVars, importance=TRUE)  
  
forest

##   
## Call:  
## randomForest(formula = SiteName ~ ., data = ImportantVars, importance = TRUE)   
## Type of random forest: classification  
## Number of trees: 500  
## No. of variables tried at each split: 5  
##   
## OOB estimate of error rate: 21.79%  
## Confusion matrix:  
## Bullen Range Kambah Pool Lanyon Nerreman Tharwa Sandwash classerror  
## Bullen Range 4 5 0 0 0 0.5556  
## Kambah Pool 0 43 0 0 0 0.0000  
## Lanyon 0 3 0 1 0 1.0000  
## Nerreman 0 5 0 14 0 0.2632  
## Tharwa Sandwash 0 2 0 1 0 1.0000

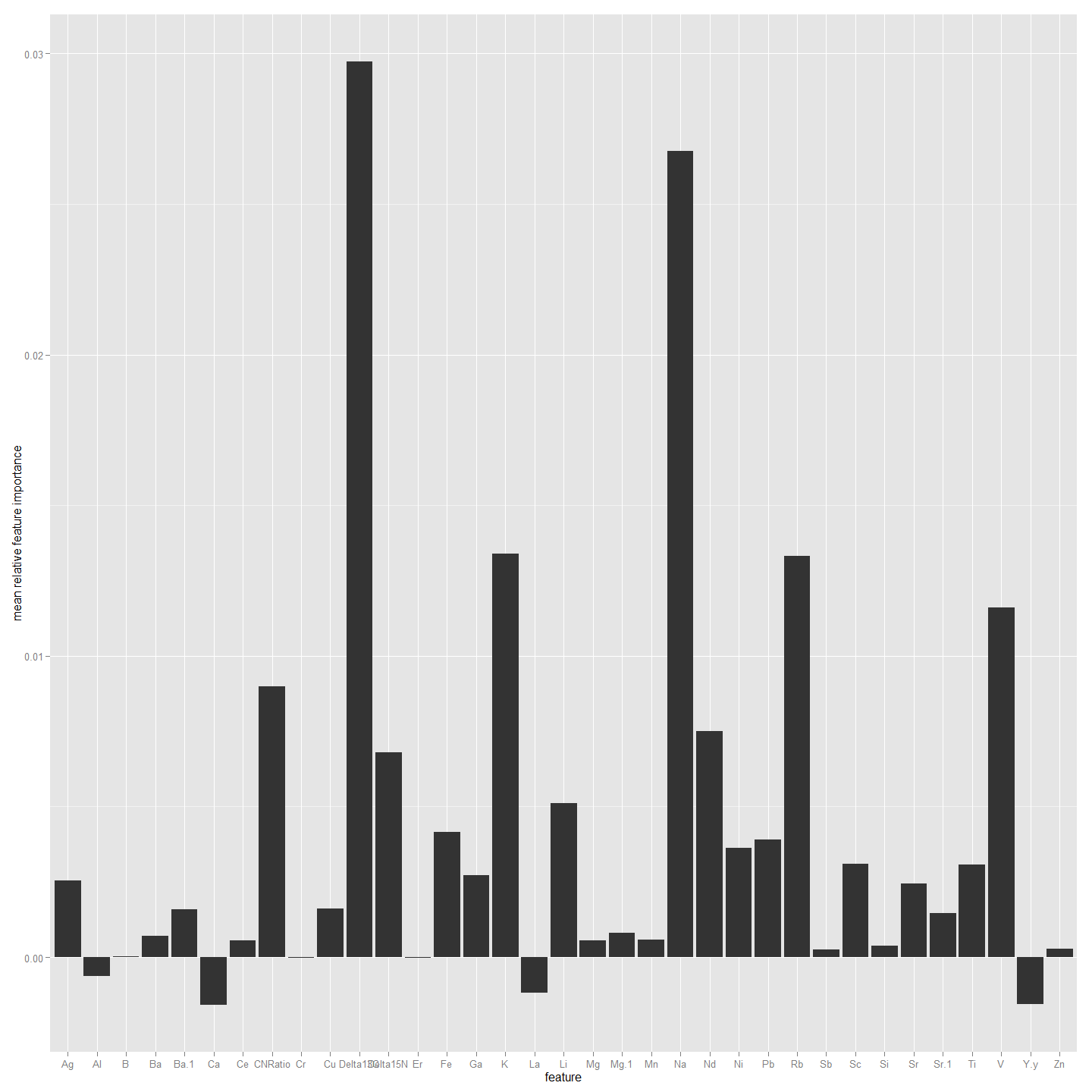
library(ggplot2)  
forest.importance = as.data.frame(importance(forest, scale=FALSE))  
forest.importance = forest.importance[,1:(ncol(forest.importance)-2)]  
forest.importance$mean = rowMeans(forest.importance)  
  
#forest.importance  
  
ggplot(forest.importance, aes(x=row.names(forest.importance), y=mean)) +  
 ylab('mean relative feature importance') +  
 xlab('feature') +  
 geom\_bar(stat='identity')



##################################################################################################  
#Now to do the same but using Edge0 values. So need to combine first  
ImportantVars<-ChemAnal0  
  
ImportantVars<-ImportantVars[c(97,13,14,15,120:151)] #  
ImportantVars<-ImportantVars[complete.cases(ImportantVars),] #remove any nulls  
colnames(ImportantVars)[1]<-"SiteName"  
ImportantVars$SiteName<-as.factor(ImportantVars$SiteName)  
  
ImportantVars <- droplevels(ImportantVars)#Not sure why get error without this line.  
  
library(randomForest)  
forest <- randomForest(SiteName ~.,data=ImportantVars, importance=TRUE)  
  
forest

##   
## Call:  
## randomForest(formula = SiteName ~ ., data = ImportantVars, importance = TRUE)   
## Type of random forest: classification  
## Number of trees: 500  
## No. of variables tried at each split: 5  
##   
## OOB estimate of error rate: 21.31%  
## Confusion matrix:  
## Bullen Range Kambah Pool Lanyon Nerreman Tharwa Sandwash classerror  
## Bullen Range 0 4 0 0 0 1.00000  
## Kambah Pool 0 34 0 2 0 0.05556  
## Lanyon 0 0 0 2 0 1.00000  
## Nerreman 1 3 0 14 0 0.22222  
## Tharwa Sandwash 0 1 0 0 0 1.00000

library(ggplot2)  
forest.importance = as.data.frame(importance(forest, scale=FALSE))  
forest.importance = forest.importance[,1:(ncol(forest.importance)-2)]  
forest.importance$mean = rowMeans(forest.importance)  
  
#forest.importance  
  
ggplot(forest.importance, aes(x=row.names(forest.importance), y=mean)) +  
 ylab('mean relative feature importance') +  
 xlab('feature') +  
 geom\_bar(stat='identity')

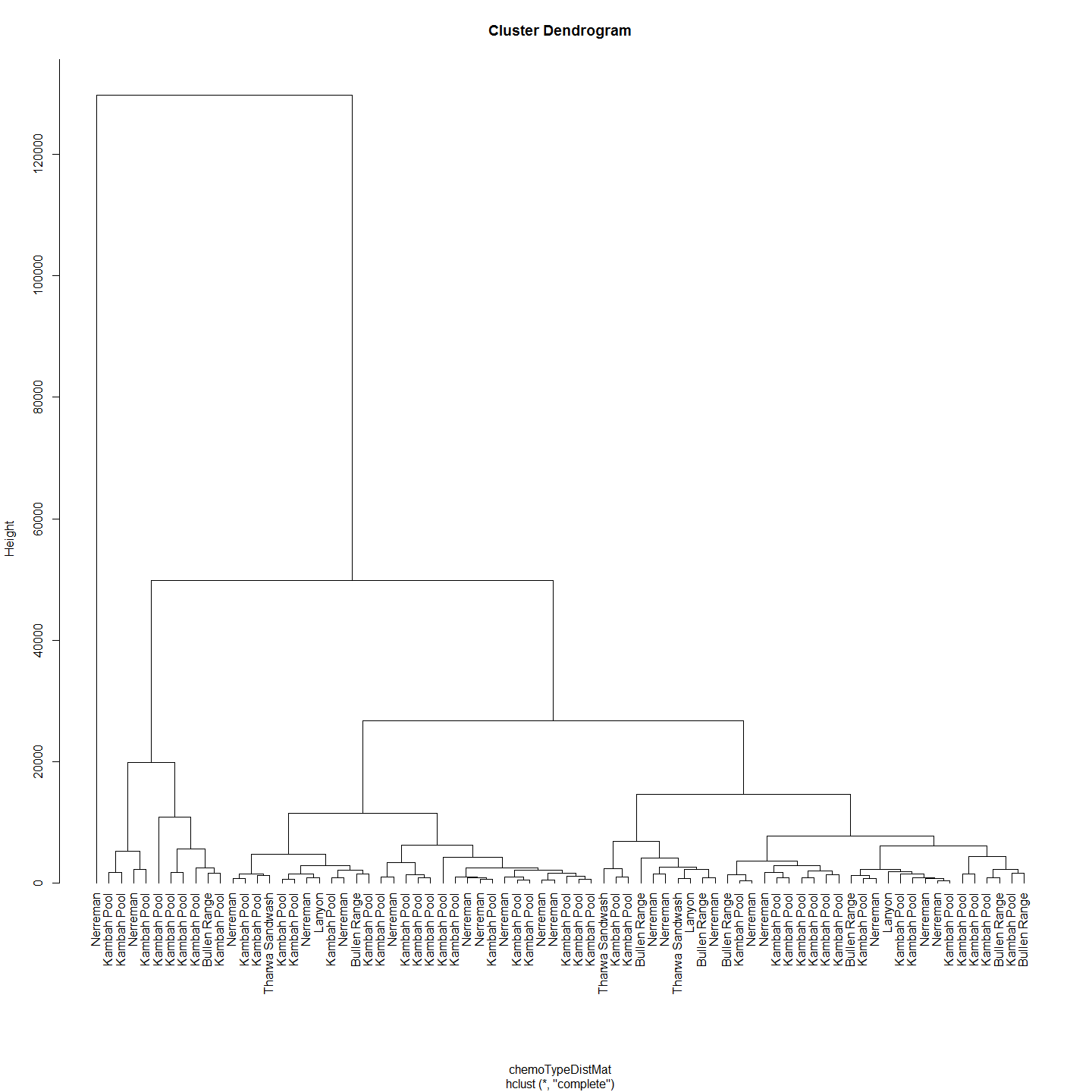


Given that each part of the otolith can provide an out-of-bag classification rate better than 75% it suggests that otolith microchemistry with delta 13 C and CN ratio, over this small spatial scale, is sufficiently consistently variable, at least within a year, to make a reasonable predictor of site of origin for the Larvae. It is worth noting that the otolith core provided the most accurate classification. Given the accuracy of the core chemistry is sub-optimal because of the inability of the operator to accuratel identify and sample the centre of the otolith in all cases with laser ablation the possibilty remains that cleaning the data in that dataset might be improved by deleting inaccurate ablates. This is possible because each otolith was assigned an accuracy score.

## A Chemotype Dendrogram and a Mantel Test.

Now that we have identified the best variables to use from the data we can create a dendrogram and use Mantel test to look for any relationship between chemotype and geographic distances. This first creates distance matrices for chemotype and geographic distance and then uses a mantel test to see correlation. The latter chemotypes are scaled.

library(clusterSim)  
#This creates distance matrices for chemotype and geographic distance and then use a mantel test to see correlation  
#First using Otolith Core chemistry.  
  
allVars<-ChemAnalCore  
  
allVars<-allVars[c(2,84,13,14,15,120:151)] #  
allVars<-allVars[complete.cases(allVars),] #remove any nulls  
row.names(allVars)<-allVars[,1]  
allVars$Label<-NULL  
colnames(allVars)[1]<-"SiteName"  
  
allVars$SiteName<-as.factor(allVars$SiteName)  
  
df<-allVars  
chemoTypeDistMat <- dist(df)  
hClusters <- hclust(chemoTypeDistMat)  
plot(hClusters,labels=(df$SiteName), hang = -1)



#Make distance matrices for geographic distance as well  
allVars<-ChemAnalCore  
allVars<-allVars[c(2,84,13,14,15,107,120:151)]  
allVars<-allVars[complete.cases(allVars),] #remove any nulls  
allVars<-allVars[c(1,6)] # Distance from Angle Crossing changes from 107 above to 6  
row.names(allVars)<-allVars[,1]  
allVars$Label<-NULL  
  
geoDist<-allVars  
geoDist<-na.omit(geoDist)  
#geoDistColl1000<-geodist #save this estimate for haplogroups distance plot (after the Iterated Mantel has changed it)  
  
geoDistMat<-dist(geoDist)  
geoDistMathm <- as.matrix(geoDistMat)  
#heatmap(geoDistMathm) #this and next two to be deleted  
#geoclust<-hclust(geoDistMat)  
#plot(geoclust)  
  
#make sure both matrices are in correct order - rows and cols  
#Check all is in order  
as.matrix(geoDistMat)[1:5, 1:5] # zero distances in the first 5

## K2013117-185 K2013117-186 K2013117-187 K2013117-188 K2013117-189  
## K2013117-185 0 0 0 0 0  
## K2013117-186 0 0 0 0 0  
## K2013117-187 0 0 0 0 0  
## K2013117-188 0 0 0 0 0  
## K2013117-189 0 0 0 0 0

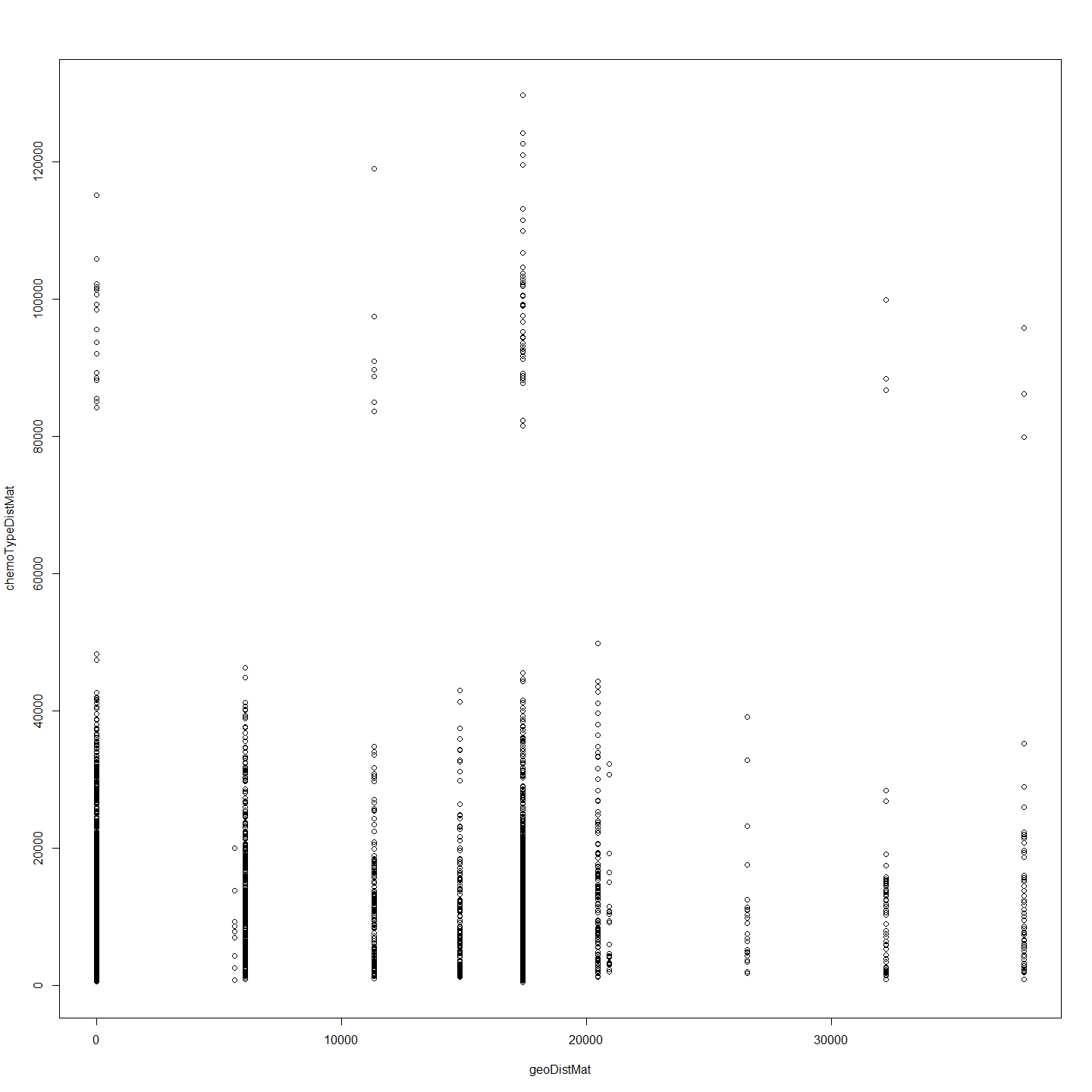
as.matrix(chemoTypeDistMat)[1:5, 1:5]

## K2013117-185 K2013117-186 K2013117-187 K2013117-188 K2013117-189  
## K2013117-185 0 35448 35951 33051 22446  
## K2013117-186 35448 0 1937 2663 13140  
## K2013117-187 35951 1937 0 3006 13513  
## K2013117-188 33051 2663 3006 0 10658  
## K2013117-189 22446 13140 13513 10658 0

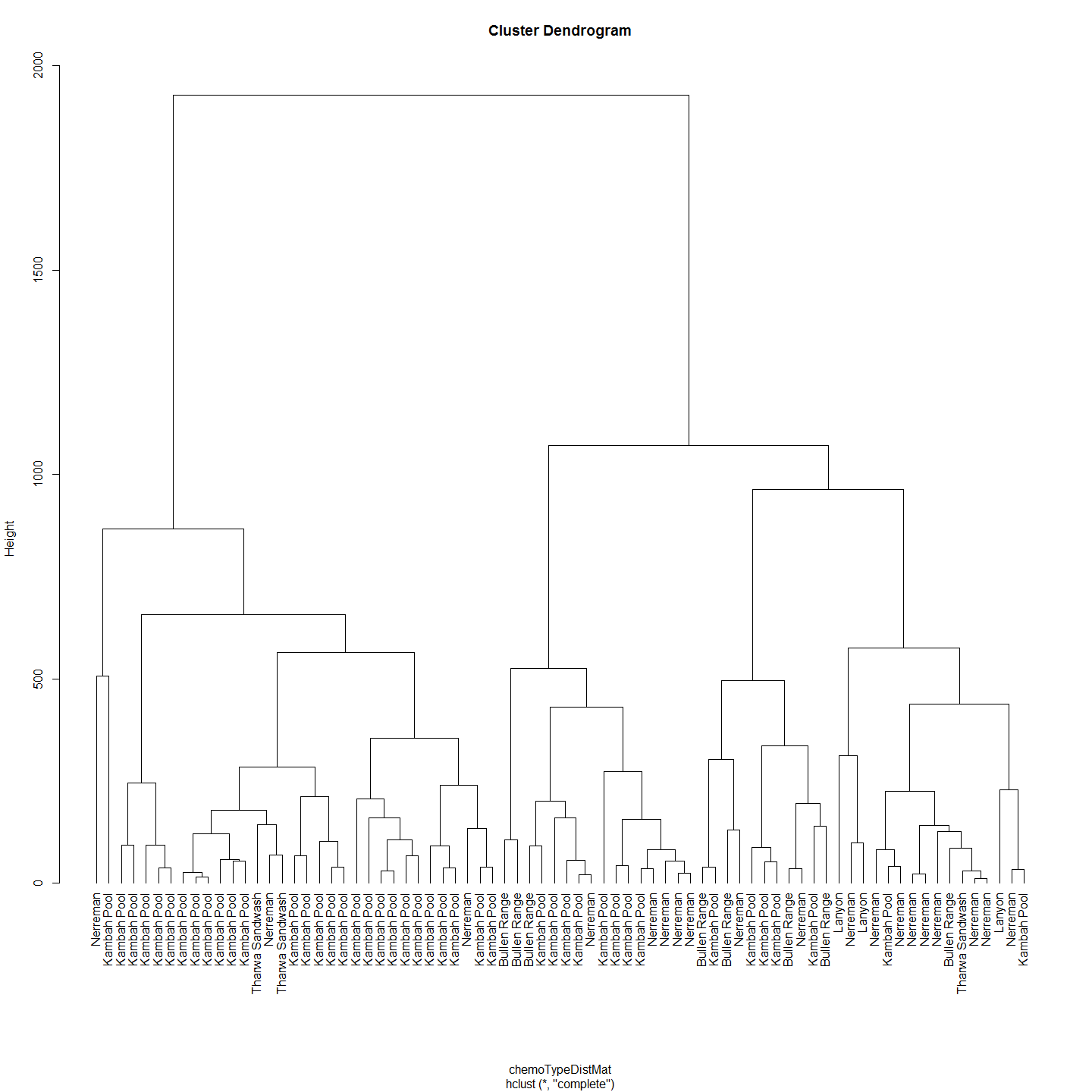
#Conduct Mantel Test on Matrices  
mant<-mantel.rtest(as.dist(geoDistMat), as.dist(chemoTypeDistMat), nrepet = 9999)  
mant

## Monte-Carlo test  
## Observation: 0.03284   
## Call: mantel.rtest(m1 = as.dist(geoDistMat), m2 = as.dist(chemoTypeDistMat),   
## nrepet = 9999)  
## Based on 9999 replicates  
## Simulated p-value: 0.2973

plot(geoDistMat,chemoTypeDistMat)



#Even if the data is scaled there is no correlation to speak of.  
#########################################  
#allVars$SiteName<-NULL  
#allVars <- droplevels(allVars)#Not sure why get error without this line  
#allVarScaled<-data.Normalization (allVars,type="n7",normalization="column")  
# allVarScaled$Label<-row.names(allVarScaled)  
# sChemAnalCore<-subset(ChemAnalCore, select=c(Label,SiteName))  
# allVars<-merge(allVarScaled, sChemAnalCore,by = "Label")  
# allVars$Label<-NULL  
# df<-allVarScaled  
# distxy <- dist(df)  
# hClusters <- hclust(distxy)  
# plot(hClusters,labels=(df$Site.Name), hang = -1)  
# df<-allVarScaled  
###################################  
  
#But now with important variables only. These include Delta13C, Delta15N, CN ratio, B, K, V, Na, Rb and were settled on by using Random Forest.  
  
ImportantVars<-ChemAnalCore  
  
ImportantVars<-ImportantVars[c(2,84,13,14,15,121,122,127,131,139)] #  
ImportantVars<-ImportantVars[complete.cases(ImportantVars),] #remove any nulls  
row.names(ImportantVars)<-ImportantVars[,1]  
ImportantVars$Label<-NULL  
colnames(ImportantVars)[1]<-"SiteName"  
  
ImportantVars$SiteName<-as.factor(ImportantVars$SiteName)  
  
#ImportantVars$SiteName<-NULL  
#ImportantVars <- droplevels(ImportantVars)#Not sure why get error without this line  
#ImpVarScaled<-data.Normalization (ImportantVars,type="n7",normalization="column")  
# ImpVarScaled$Label<-row.names(ImpVarScaled)  
# sChemAnalCore<-subset(ChemAnalCore, select=c(Label,SiteName))  
# ImportantVars<-merge(ImpVarScaled, sChemAnalCore,by = "Label")  
# ImportantVars$Label<-NULL  
# df<-ImpVarScaled  
# distxy <- dist(df)  
# hClusters <- hclust(distxy)  
# plot(hClusters,labels=(df$Site.Name), hang = -1)  
# df<-ImpVarScaled  
  
df<-ImportantVars  
chemoTypeDistMat <- dist(df)  
hClusters <- hclust(chemoTypeDistMat)  
plot(hClusters,labels=(df$SiteName), hang = -1)



#Make distance matrices for geographic distance as well  
ImportantVars<-ChemAnalCore  
ImportantVars<-ImportantVars[c(2,84,13,14,15,107,121,122,127,131,139)]  
ImportantVars<-ImportantVars[complete.cases(ImportantVars),] #remove any nulls  
ImportantVars<-ImportantVars[c(1,6)] # Distance from Angle Crossing changes from 107 above to 6  
row.names(ImportantVars)<-ImportantVars[,1]  
ImportantVars$Label<-NULL  
  
geoDist<-ImportantVars  
geoDist<-na.omit(geoDist)  
#geoDistColl1000<-geodist #save this estimate for haplogroups distance plot (after the Iterated Mantel has changed it)  
  
geoDistMat<-dist(geoDist)  
geoDistMathm <- as.matrix(geoDistMat)  
#heatmap(geoDistMathm) #this and next two to be deleted  
#geoclust<-hclust(geoDistMat)  
#plot(geoclust)  
  
#make sure both matrices are in correct order - rows and cols  
#Check all is in order  
as.matrix(geoDistMat)[1:5, 1:5] # zero distances in the first 5

## K2013117-185 K2013117-186 K2013117-187 K2013117-188 K2013117-189  
## K2013117-185 0 0 0 0 0  
## K2013117-186 0 0 0 0 0  
## K2013117-187 0 0 0 0 0  
## K2013117-188 0 0 0 0 0  
## K2013117-189 0 0 0 0 0

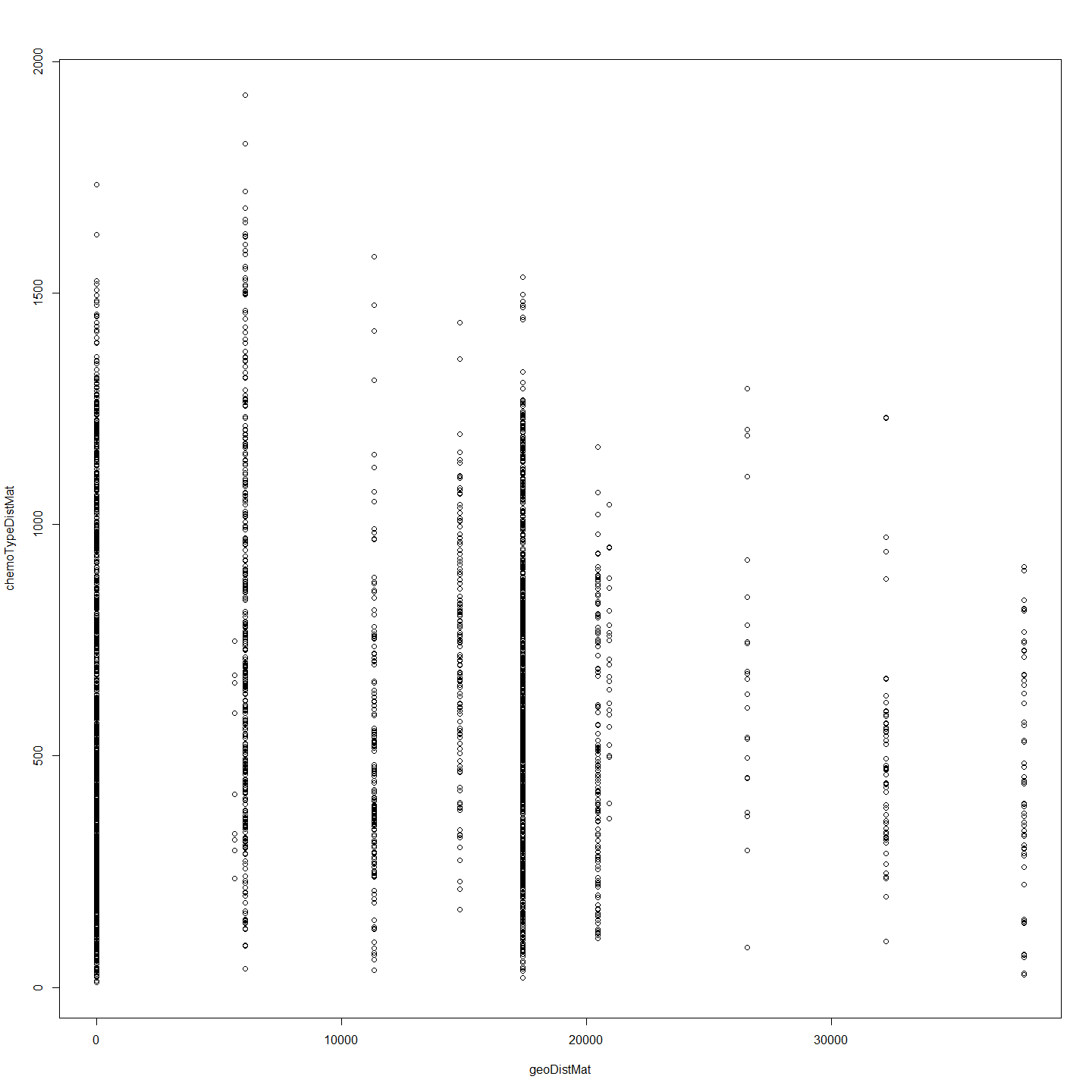
as.matrix(chemoTypeDistMat)[1:5, 1:5]

## K2013117-185 K2013117-186 K2013117-187 K2013117-188 K2013117-189  
## K2013117-185 0.00 87.63 226.6 494.6 52.71  
## K2013117-186 87.63 0.00 200.0 459.1 64.30  
## K2013117-187 226.57 199.98 0.0 268.0 176.07  
## K2013117-188 494.60 459.14 268.0 0.0 443.81  
## K2013117-189 52.71 64.30 176.1 443.8 0.00

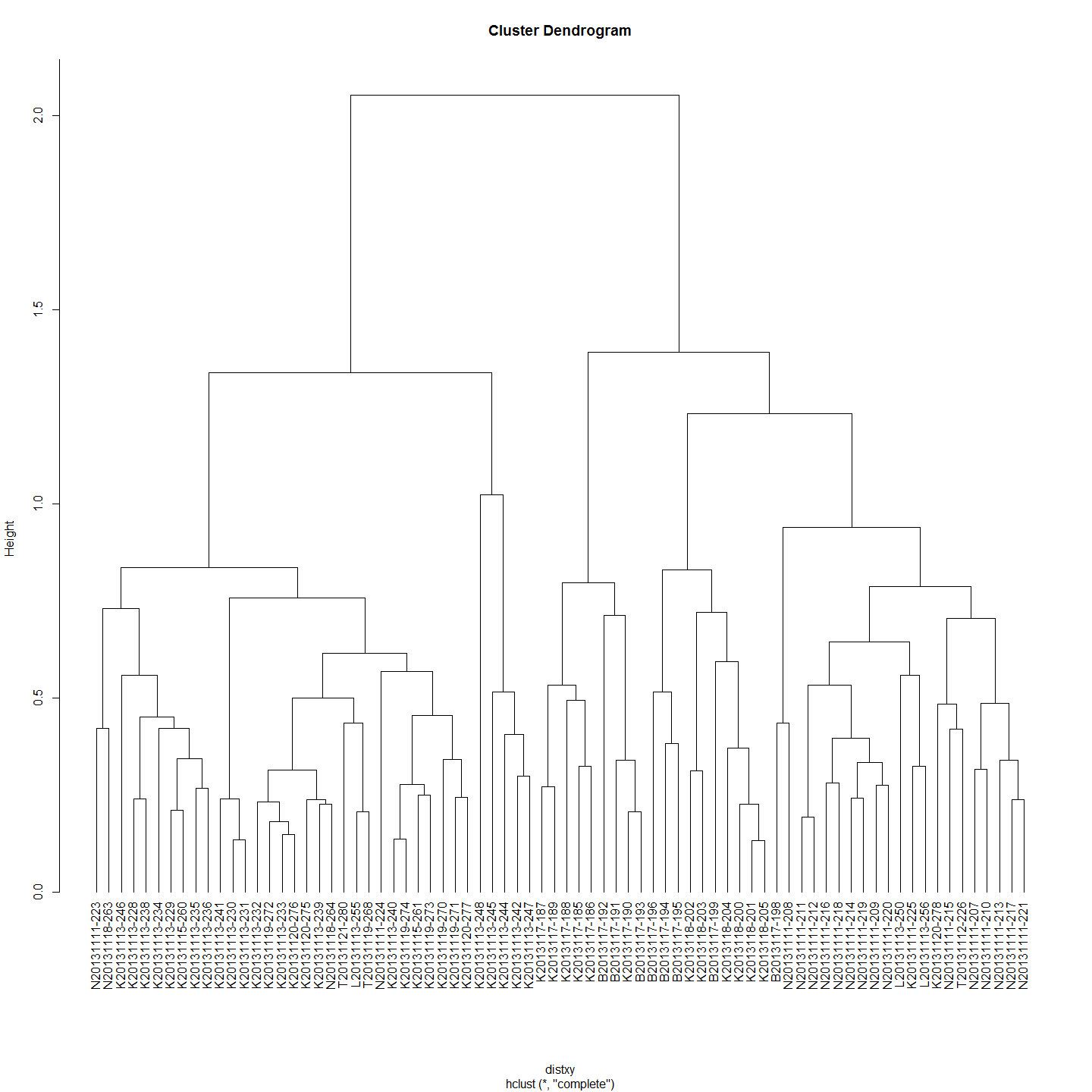
#Conduct Mantel Test on Matrices  
mantIV<-mantel.rtest(as.dist(geoDistMat), as.dist(chemoTypeDistMat), nrepet = 9999)  
mantIV

## Monte-Carlo test  
## Observation: -0.01684   
## Call: mantel.rtest(m1 = as.dist(geoDistMat), m2 = as.dist(chemoTypeDistMat),   
## nrepet = 9999)  
## Based on 9999 replicates  
## Simulated p-value: 0.6312

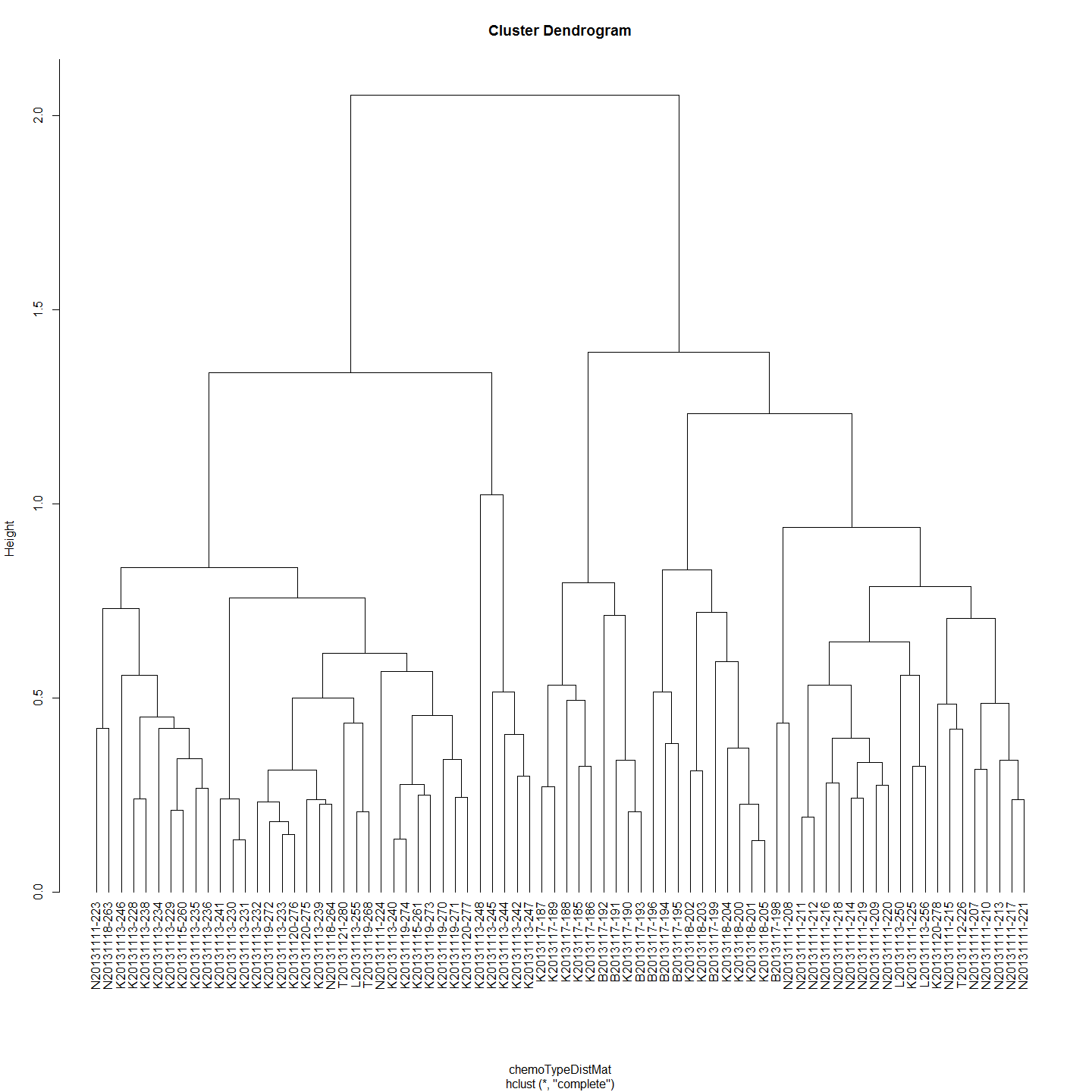
plot(geoDistMat,chemoTypeDistMat)



##############################  
  
#But now with important variables only AFTER SCALING. These include C13, N15, CN ratio, B, K, V, Na, Rb.  
ImportantVars<-ChemAnalCore  
  
ImportantVars<-ImportantVars[c(2,84,13,14,15,121,122,127,131,139)] #  
ImportantVars<-ImportantVars[complete.cases(ImportantVars),] #remove any nulls  
row.names(ImportantVars)<-ImportantVars[,1]  
ImportantVars$Label<-NULL  
colnames(ImportantVars)[1]<-"SiteName"  
  
ImportantVars$SiteName<-as.factor(ImportantVars$SiteName)  
  
ImportantVars$SiteName<-NULL  
ImportantVars <- droplevels(ImportantVars)#Not sure why get error without this line  
ImpVarScaled<-data.Normalization (ImportantVars,type="n4",normalization="column")  
ImpVarScaled$Label<-row.names(ImpVarScaled)  
sChemAnalCore<-subset(ChemAnalCore, select=c(Label,SiteName))  
ImportantVars<-merge(ImpVarScaled, sChemAnalCore,by = "Label")  
ImportantVars$Label<-NULL  
df<-ImpVarScaled  
distxy <- dist(df)  
hClusters <- hclust(distxy)  
plot(hClusters,labels=(df$Site.Name), hang = -1)



df<-ImpVarScaled  
  
#df<-ImportantVars  
chemoTypeDistMat <- dist(df)  
hClusters <- hclust(chemoTypeDistMat)  
plot(hClusters,labels=(df$SiteName), hang = -1)



#Make distance matrices for geographic distance as well  
ImportantVars<-ChemAnalCore  
ImportantVars<-ImportantVars[c(2,84,13,14,15,107,121,122,127,131,139)]  
ImportantVars<-ImportantVars[complete.cases(ImportantVars),] #remove any nulls  
ImportantVars<-ImportantVars[c(1,6)] # Distance from Angle Crossing changes from 107 above to 6  
row.names(ImportantVars)<-ImportantVars[,1]  
ImportantVars$Label<-NULL  
  
geoDist<-ImportantVars  
geoDist<-na.omit(geoDist)  
#geoDistColl1000<-geodist #save this estimate for haplogroups distance plot (after the Iterated Mantel has changed it)  
  
geoDistMat<-dist(geoDist)  
geoDistMathm <- as.matrix(geoDistMat)  
#heatmap(geoDistMathm) #this and next two to be deleted  
#geoclust<-hclust(geoDistMat)  
#plot(geoclust)  
  
#make sure both matrices are in correct order - rows and cols  
#Check all is in order  
as.matrix(geoDistMat)[1:5, 1:5] # zero distances in the first 5

## K2013117-185 K2013117-186 K2013117-187 K2013117-188 K2013117-189  
## K2013117-185 0 0 0 0 0  
## K2013117-186 0 0 0 0 0  
## K2013117-187 0 0 0 0 0  
## K2013117-188 0 0 0 0 0  
## K2013117-189 0 0 0 0 0

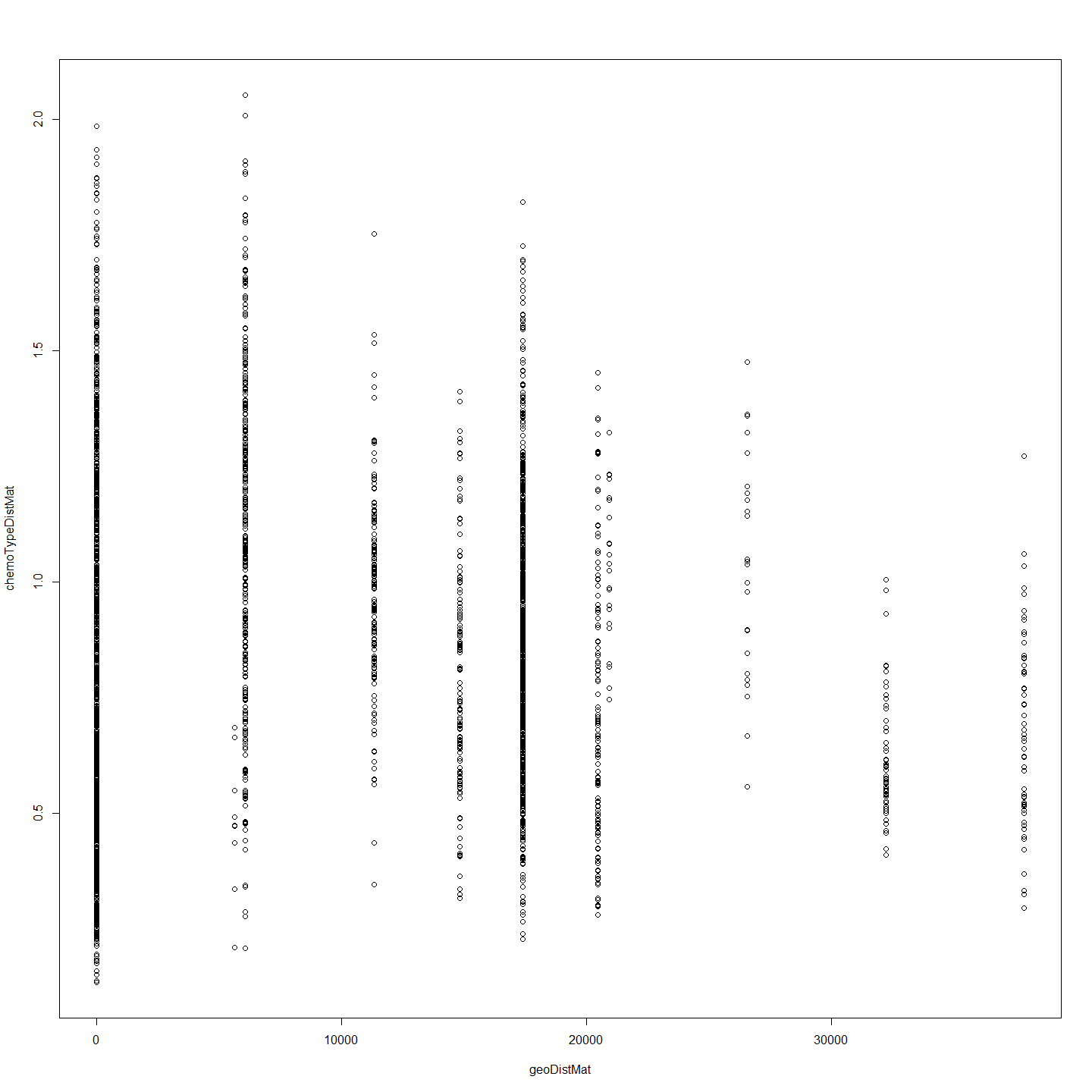
as.matrix(chemoTypeDistMat)[1:5, 1:5]

## K2013117-185 K2013117-186 K2013117-187 K2013117-188 K2013117-189  
## K2013117-185 0.0000 0.3254 0.3835 0.4947 0.5131  
## K2013117-186 0.3254 0.0000 0.3438 0.3968 0.3405  
## K2013117-187 0.3835 0.3438 0.0000 0.4501 0.2711  
## K2013117-188 0.4947 0.3968 0.4501 0.0000 0.5329  
## K2013117-189 0.5131 0.3405 0.2711 0.5329 0.0000

#Conduct Mantel Test on Matrices  
mantIVScaled<-mantel.rtest(as.dist(geoDistMat), as.dist(chemoTypeDistMat), nrepet = 9999)  
mantIVScaled

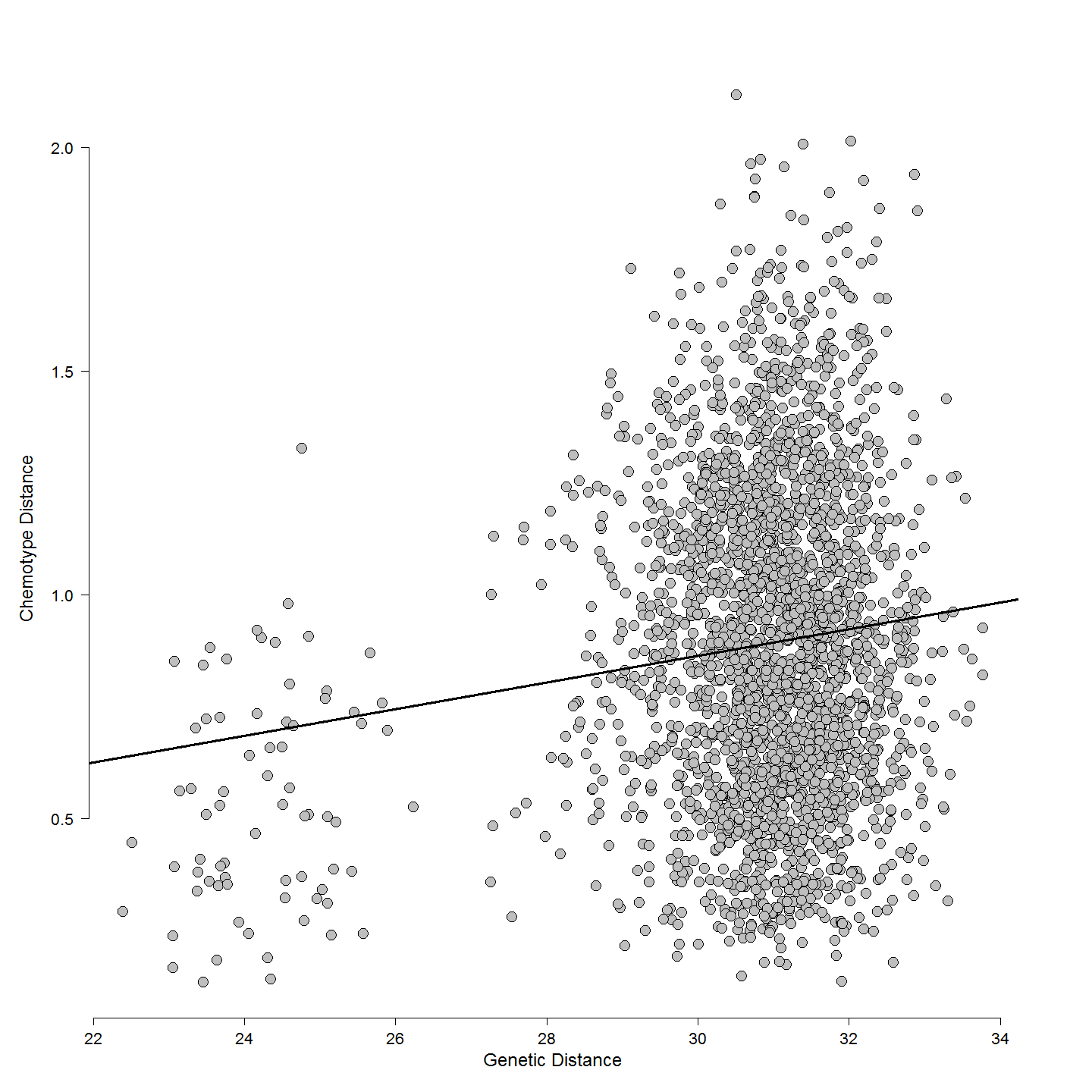
## Monte-Carlo test  
## Observation: -0.04129   
## Call: mantel.rtest(m1 = as.dist(geoDistMat), m2 = as.dist(chemoTypeDistMat),   
## nrepet = 9999)  
## Based on 9999 replicates  
## Simulated p-value: 0.7766

plot(geoDistMat,chemoTypeDistMat)

 Based on these results, we cannot reject the null hypothesis that these two matrices, spatial distance and chemotype distance, are unrelated with alpha = 0.2973. The observed correlation, 0.0328, suggests that the matrix entries are not positively associated.

### Is Chemotype Distance Correlated with Genotype Distance?

library(ade4)  
  
MCdmPrep<-MCsnps[-c(1:7),]  
  
test<-merge(MCdmPrep,ImpVarScaled,by = "row.names")  
genDist<-dist(test[,c(1:21076)])  
chemDist<-dist(test[,c(21077:21086)])  
  
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(genDist,chemDist, col="black", pch=21, bg = "grey", cex = 2,  
 ylab="", xlab="", axes=F)  
axis(1)  
axis(2)   
reg1 <- lm(chemDist~genDist)  
ablineclip(reg1, lwd=3)   
par(las=0)  
mtext("Genetic Distance", side=1, line=2.5, cex=1.5)  
mtext("Chemotype Distance", side=2, line=3.7, cex=1.5)



cor(genDist,chemDist)

## [1] 0.1202

rcorr(genDist[lower.tri(genDist)],chemDist[lower.tri(chemDist)])#(x, type="pearson") # type can be pearson or spearman

## x y  
## x 1.0 0.1  
## y 0.1 1.0  
##   
## n  
## x y  
## x 2701 1998  
## y 1998 2701  
##   
## P  
## x y   
## x 0  
## y 0

mant<-mantel.rtest(genDist, chemDist, nrepet = 9999)  
mant

## Monte-Carlo test  
## Observation: 0.1202   
## Call: mantelnoneuclid(m1 = m1, m2 = m2, nrepet = nrepet)  
## Based on 9999 replicates  
## Simulated p-value: 0.004

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

## 1 2 3 4 5  
## 1 0.00 30.08 23.66 30.13 30.33  
## 2 30.08 0.00 30.05 28.74 30.06  
## 3 23.66 30.05 0.00 30.12 29.26  
## 4 30.13 28.74 30.12 0.00 29.87  
## 5 30.33 30.06 29.26 29.87 0.00

as.matrix(chemDist)[1:5, 1:5]

## 1 2 3 4 5  
## 1 0.0000 0.7524 0.3493 1.2691 1.1965  
## 2 0.7524 0.0000 0.5107 1.0790 0.8908  
## 3 0.3493 0.5107 0.0000 1.1545 0.9933  
## 4 1.2691 1.0790 1.1545 0.0000 0.3817  
## 5 1.1965 0.8908 0.9933 0.3817 0.0000

This is interesting given that there is a relationship between genotype and chemotype but it is not one due to covariance based on geographic distance. One not unreasonable explaination for this is that suggest that chemotype is afftected by the genotype directly. That is, element deposition in otolith is variable but under some genetic control. Another more intriguing possibility is that the geochemistry fails to predict the capture site of the larvae because it is not the ultimte source of that geochemistry. The ultimate source of the geochemistry is thewater and food environment of the female when the yolk which was laid down in the developing eggs. This happens earlier in the season and well before spawning so it does leave ope the possibiltiy as suggested by some authors that the female has migrated to the spawning site. Unfortunately in this study there is no measure that might be suitable as a proxy regarding the origin of the female during oogenisis.

## Code Chunks in this Document

## [1] "Project\_Template\_and\_Knitr" "Set\_Global\_Options" "LoadLibraries" "Additional\_Calculations"   
## [5] "All\_Maccullochella\_Larvae" "Murray\_Cod\_Larvae\_Only" "A2R\_Dendrogram" "ExtractClades"   
## [9] "Trout\_Cod\_Dendrograms" "unnamed-chunk-1" "first" "Distance\_Matrices\_and\_Ordering"   
## [13] "Plots\_and\_Correlation" "Mantel\_Test" "Clades\_and\_Location" "unnamed-chunk-2"   
## [17] "IRandomForestGo" "unnamed-chunk-3" "chemotypeDistanceMatricesGo" "unnamed-chunk-4"   
## [21] "genChemDistPlotGo" "Include\_Chunk\_Labels\_and\_Session Information" "IM" "RandomForestChem"   
## [25] "chemotypeDistanceMatrices" "genChemPlot"

## R version 3.1.1 (2014-07-10)  
## Platform: x86\_64-w64-mingw32/x64 (64-bit)  
##   
## 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] splines grid stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] clusterSim\_0.44-2 MASS\_7.3-34 cluster\_1.15.2 randomForest\_4.6-10 plotrix\_3.5-11 ade4\_1.6-2 Hmisc\_3.14-4 Formula\_1.1-2   
## [9] survival\_2.37-7 lattice\_0.20-29 dendextend\_0.18.3 ape\_3.1-4 ggdendro\_0.1-15 ggplot2\_1.0.0 ProjectTemplate\_0.5-1 knitr\_1.6   
##   
## loaded via a namespace (and not attached):  
## [1] class\_7.3-11 colorspace\_1.2-4 digest\_0.6.4 e1071\_1.6-4 evaluate\_0.5.5 formatR\_1.0 gtable\_0.1.2 htmltools\_0.2.6 labeling\_0.3   
## [10] latticeExtra\_0.6-26 magrittr\_1.0.1 munsell\_0.4.2 nlme\_3.1-117 plyr\_1.8.1 proto\_0.3-10 R2HTML\_2.3.1 RColorBrewer\_1.0-5 Rcpp\_0.11.2   
## [19] reshape2\_1.4 rgl\_0.95.1201 rmarkdown\_0.2.54 scales\_0.2.4 stringr\_0.6.2 tools\_3.1.1 whisker\_0.3-2 yaml\_2.1.13