Upper Murrumbidgee Larval Cod Dispersal Analysis

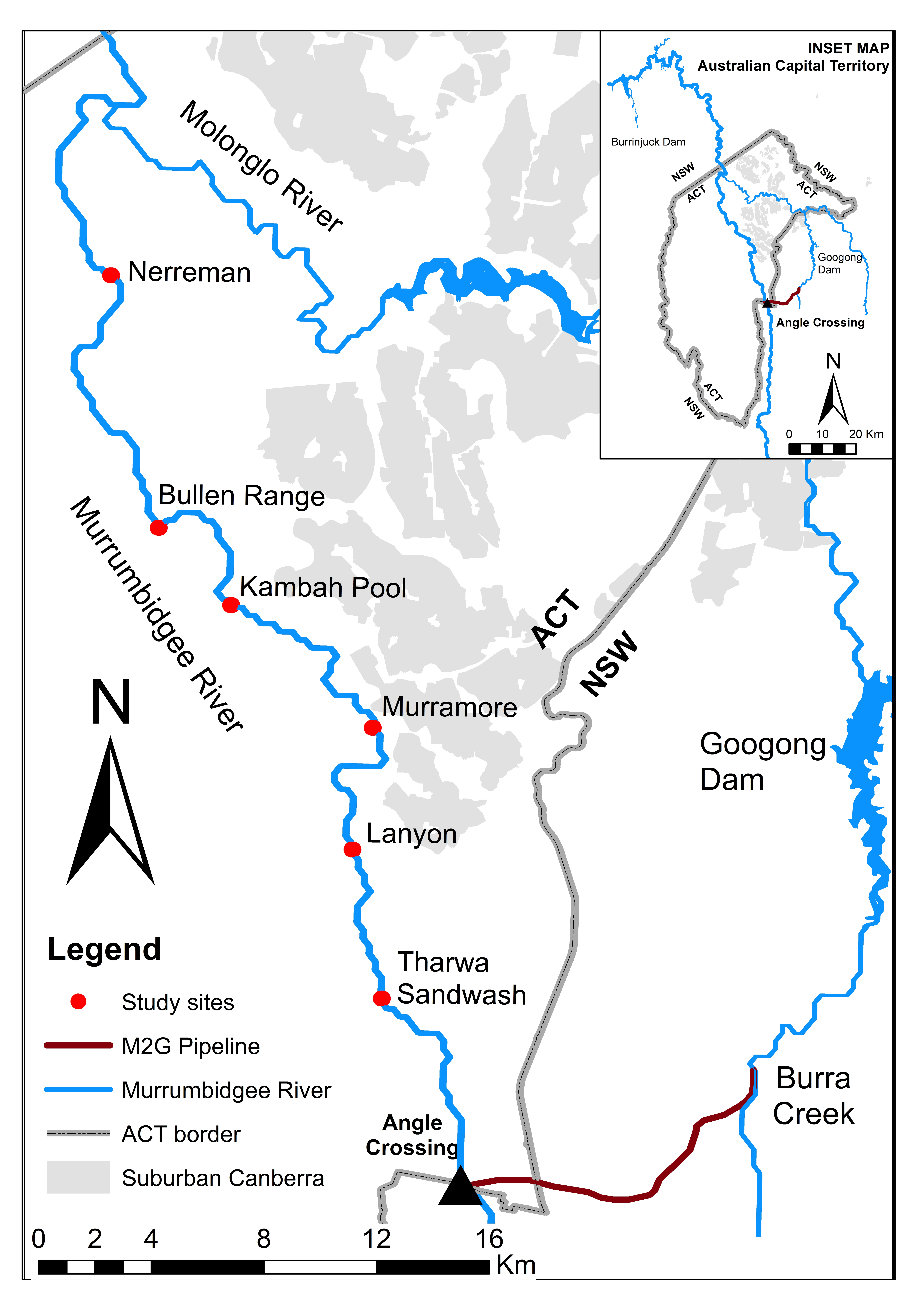
Alan Couch

Tue Mar 17 10:45:35 AM 2015

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

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)

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. The larvae used in the dispersal analysis were collected in 2013 from 6 sites. The data used for ageing were collected over the three years 2011-2013.



## 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

Hatch DoY : [Day of Year Caught]-[Age From Otolith Length]

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

Spawning:[Hatch]-[Incubation]

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.

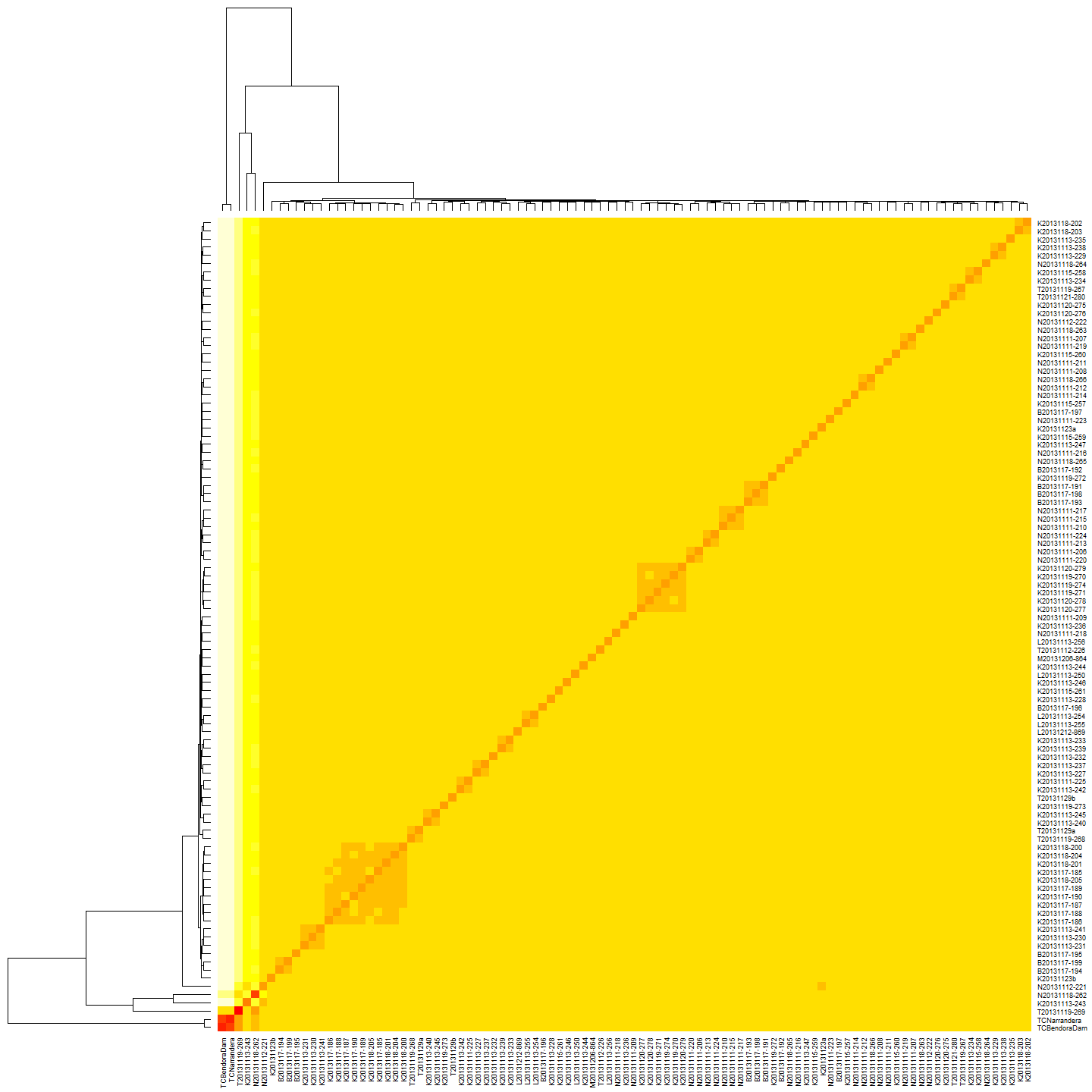
larv$ageOL<-74.308\*larv$Mean.Otolith.Length.is.in.Millimetres.for.comparison.with.Adults-4.44361  
larv$hatchdoy<-larv$Day.of.Year-larv$ageOL  
larv$incTime<-20.67-0.667\*larv$WaterTemp.DegC..Mean  
larv$spawn<-larv$hatchdoy-larv$incTime  
  
  
#larv[,119:123] #just to see all OK

But before that we need to establish the genetic identies so that we can test the assumption that genetic distance is positively corelated with geographic distance between individuals.

## All Maccullochella Larvae

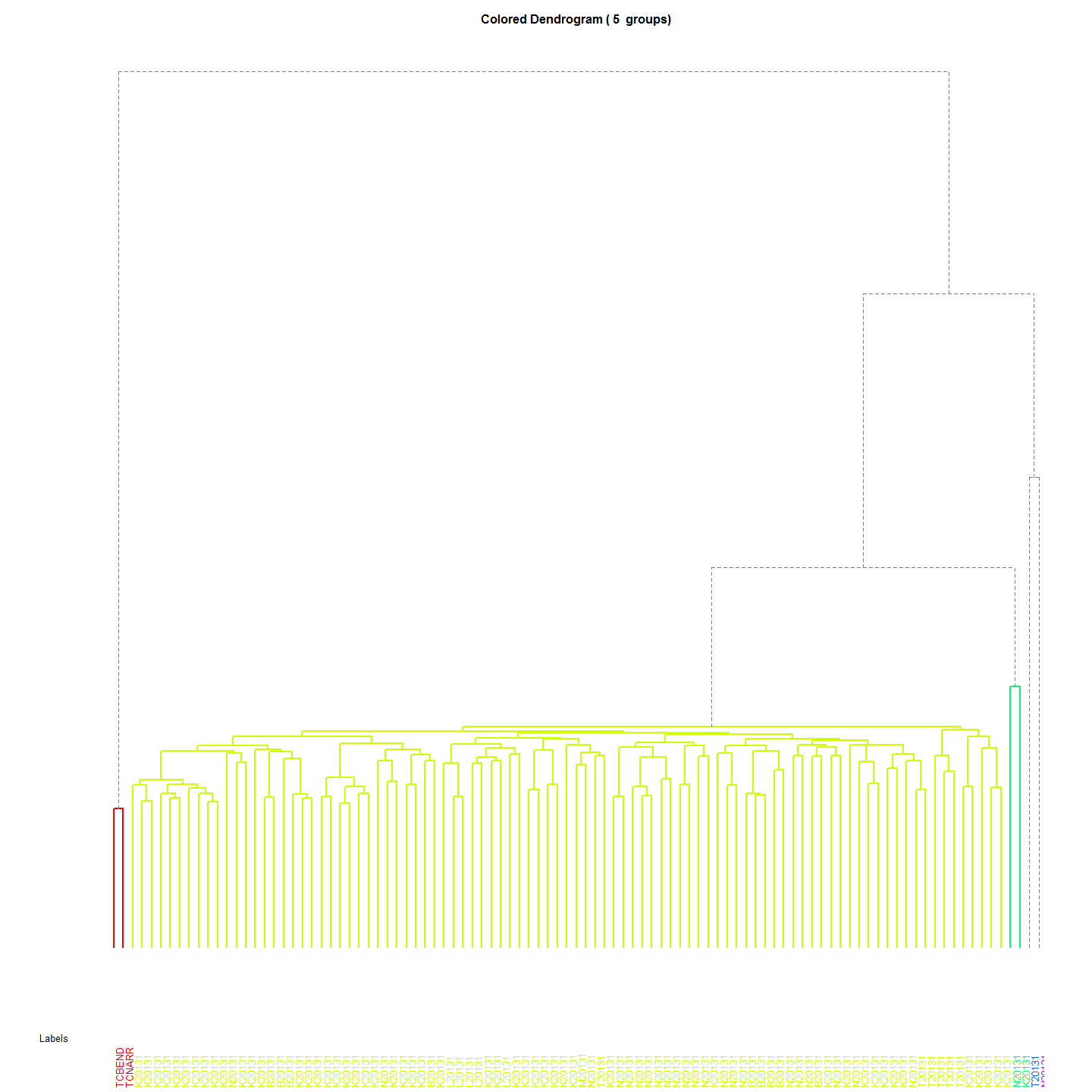
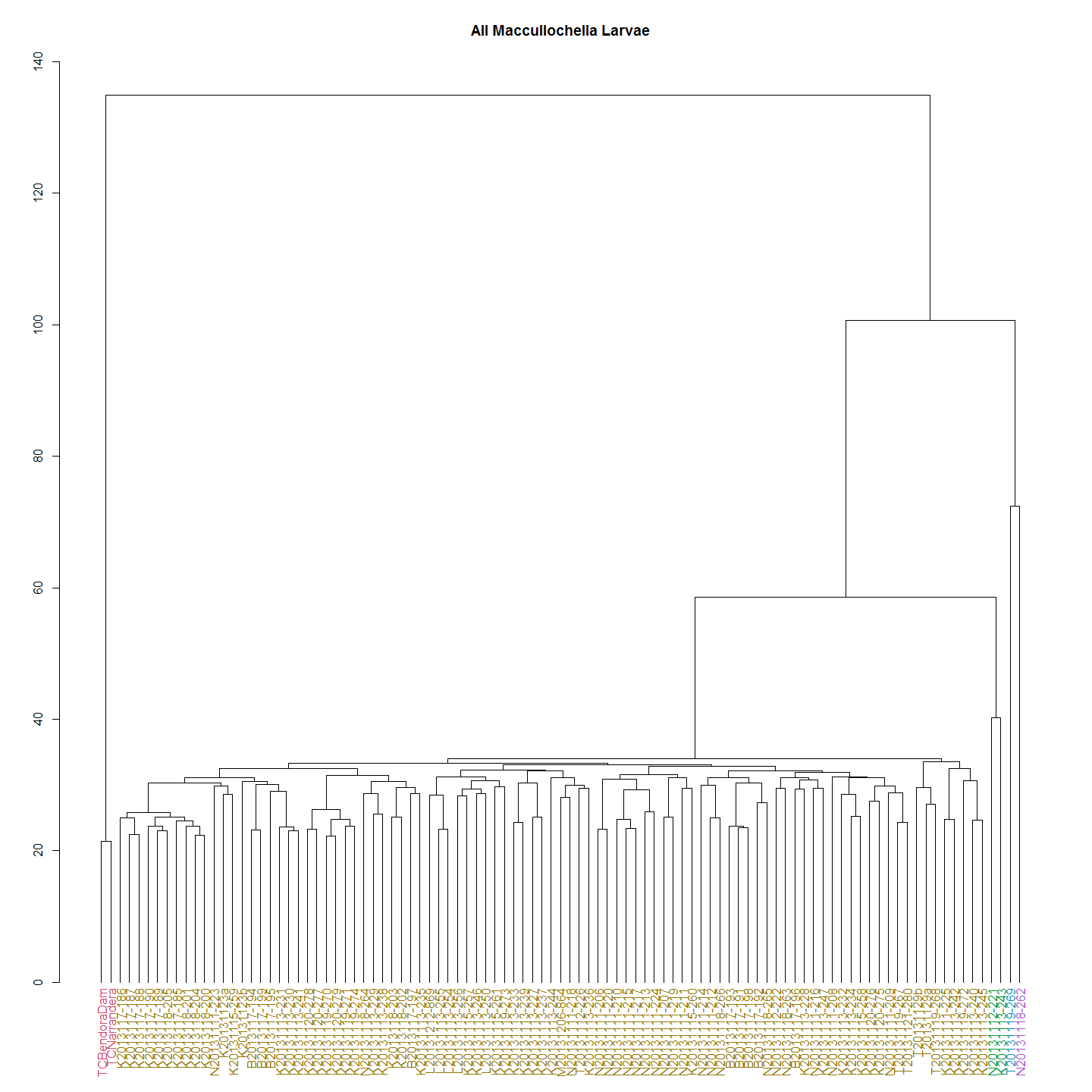
If we look at all the larvae collected we see four distinct clades.The Trout cod and the hybrids are easily identified and eliminated from the data to ensure we are looking only at the Murray cod larvae.

#Create a distance matrix for all Maccullochella larvae  
MacDm <- dist(allsnps)  
  
#Make a heat map  
dataMatrix <- as.matrix(MacDm)  
heatmap(dataMatrix)



#Do a heirarchical cluster  
MacHC <- hclust(MacDm)  
#Plot the cluster dendrogram  
plot(MacHC, main="All Maccullochella Larvae")



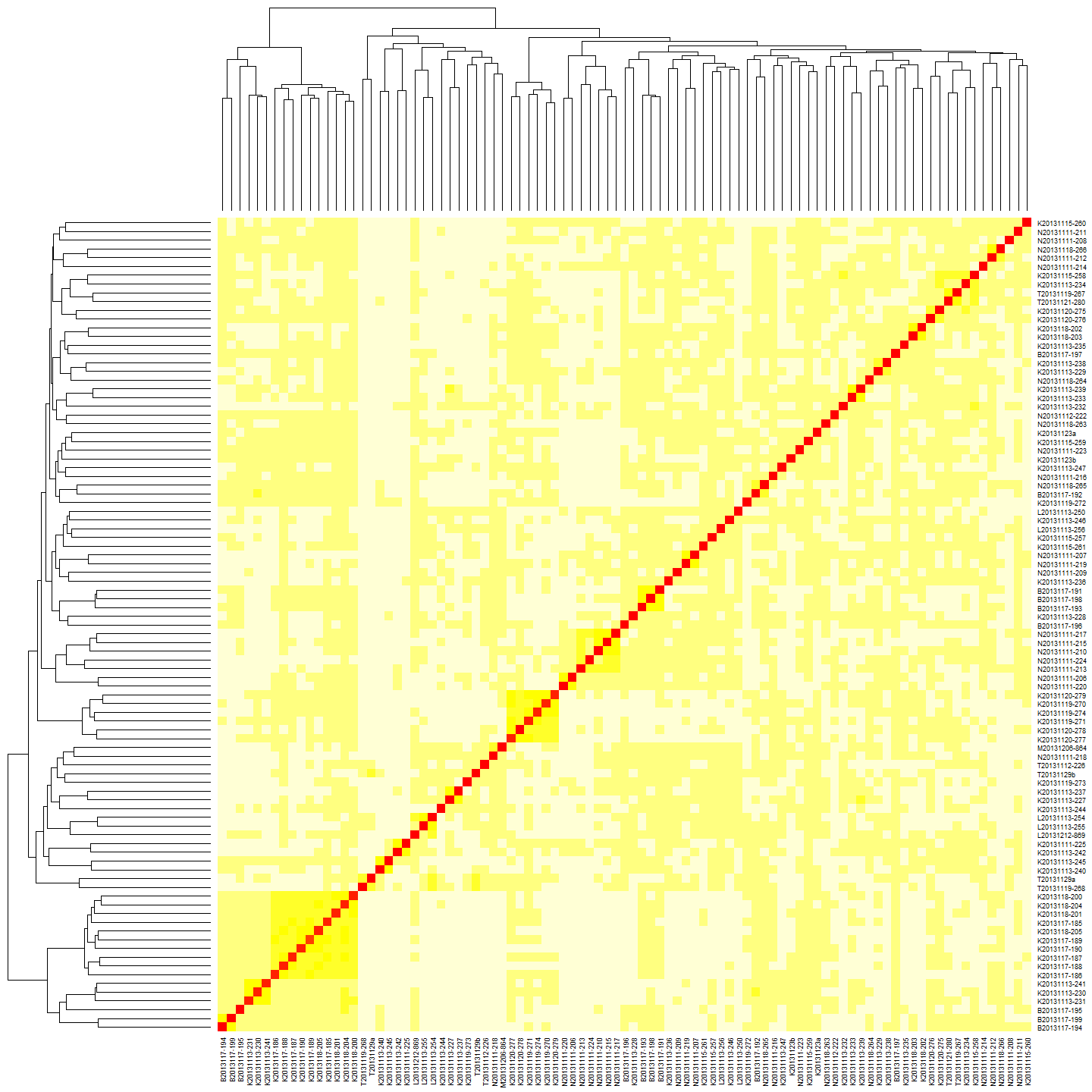
The above dendrograms shows Murray cod and two known Trout cod controls and what appears to be F1 and F2 hybrids. There were only 4 hybrid and 0 pure trout cod larvae deteced in the 92 larvae caught and sequenced.

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

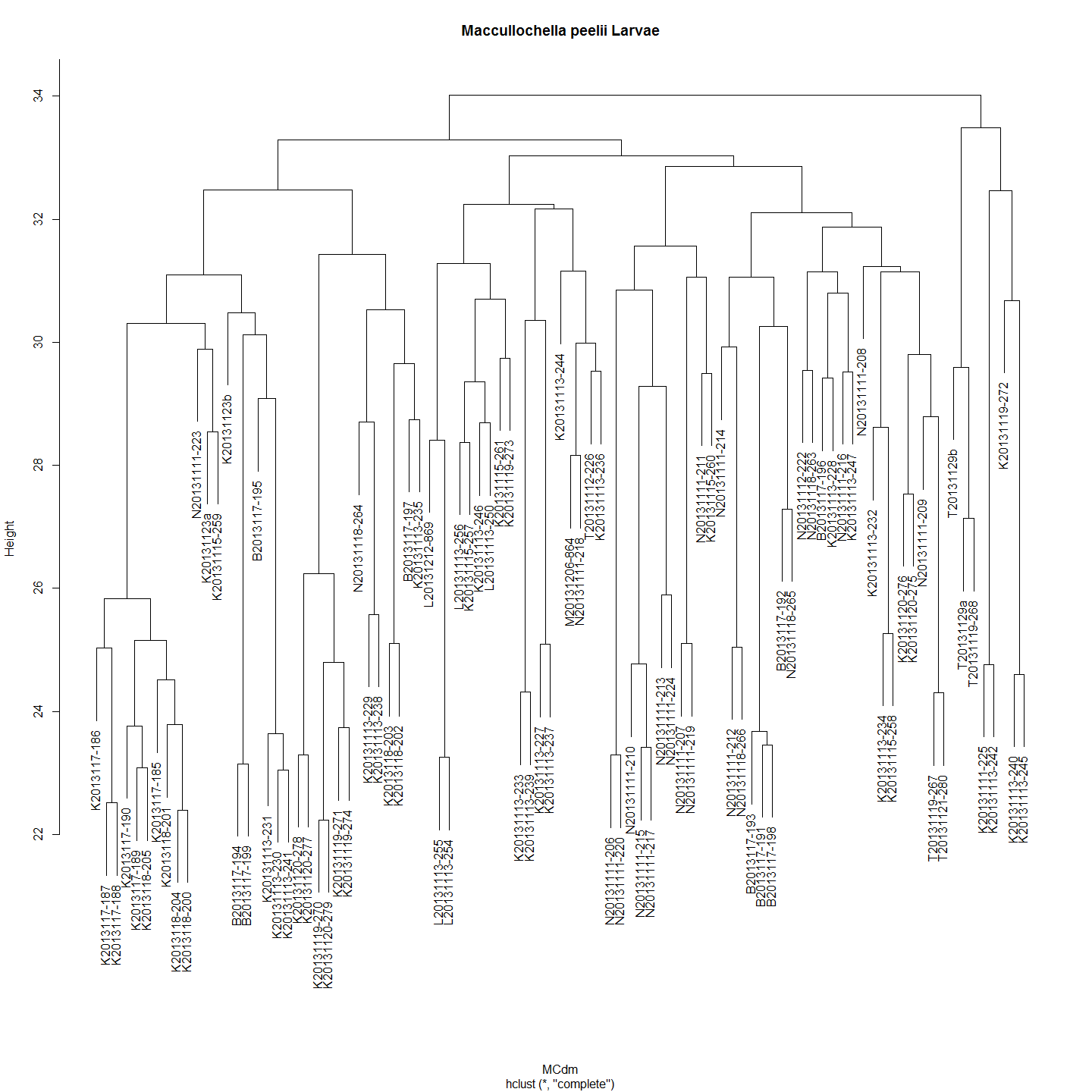
So we now remove the non- Murray cod to identify Murray cod haplogroups and conduct the remainder of the analysis on the Upper Murrumbidgee Murray cod larvae only.

## Murray Cod Larvae

#Create a Murray Cod distance matrix  
MCdm <- dist(MCsnps)  
#Heat map  
dataMatrix <- as.matrix(MCdm)  
heatmap(dataMatrix)

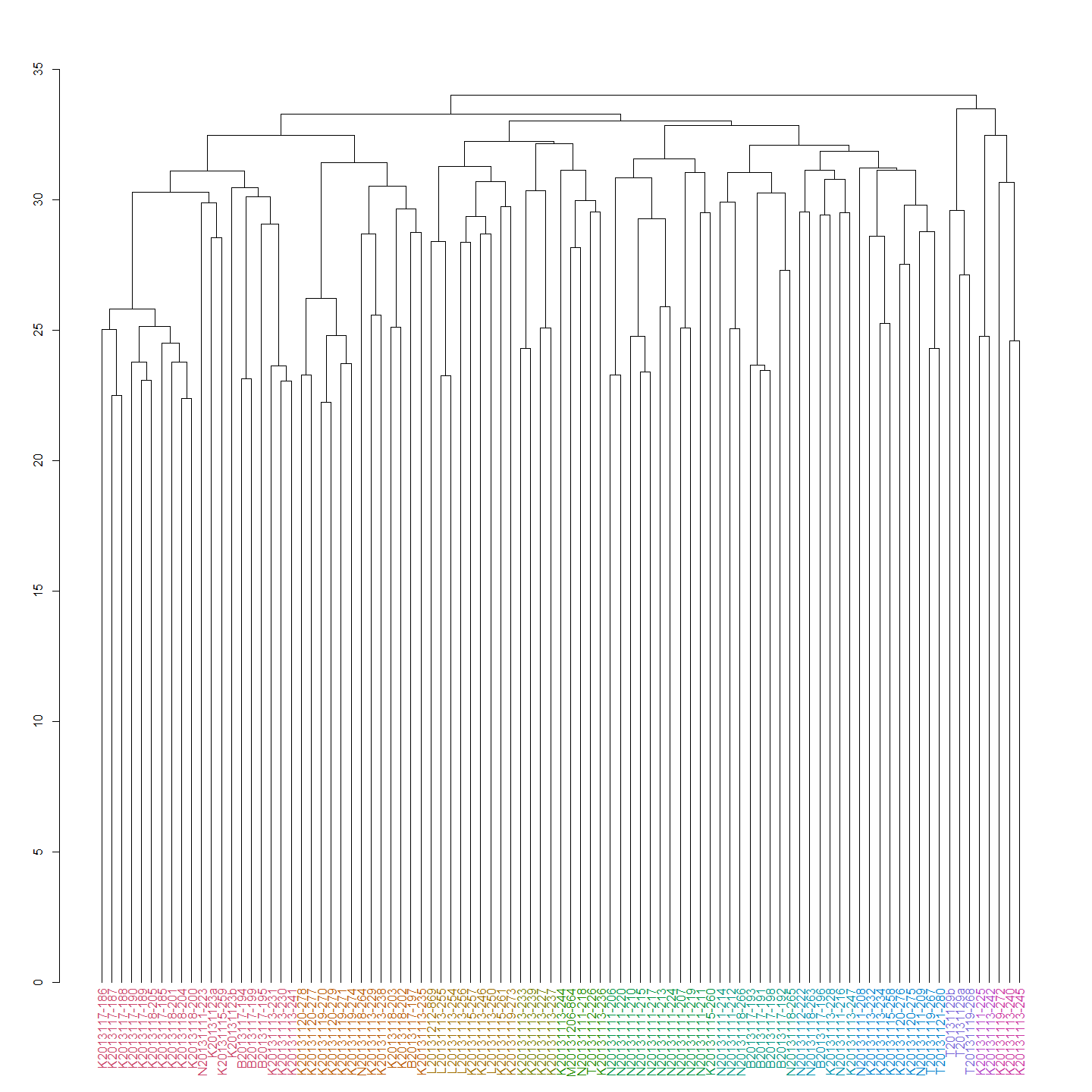


#cluster  
MChc <- hclust(MCdm)  
#Plot it  
plot(MChc,main="Maccullochella peelii Larvae")

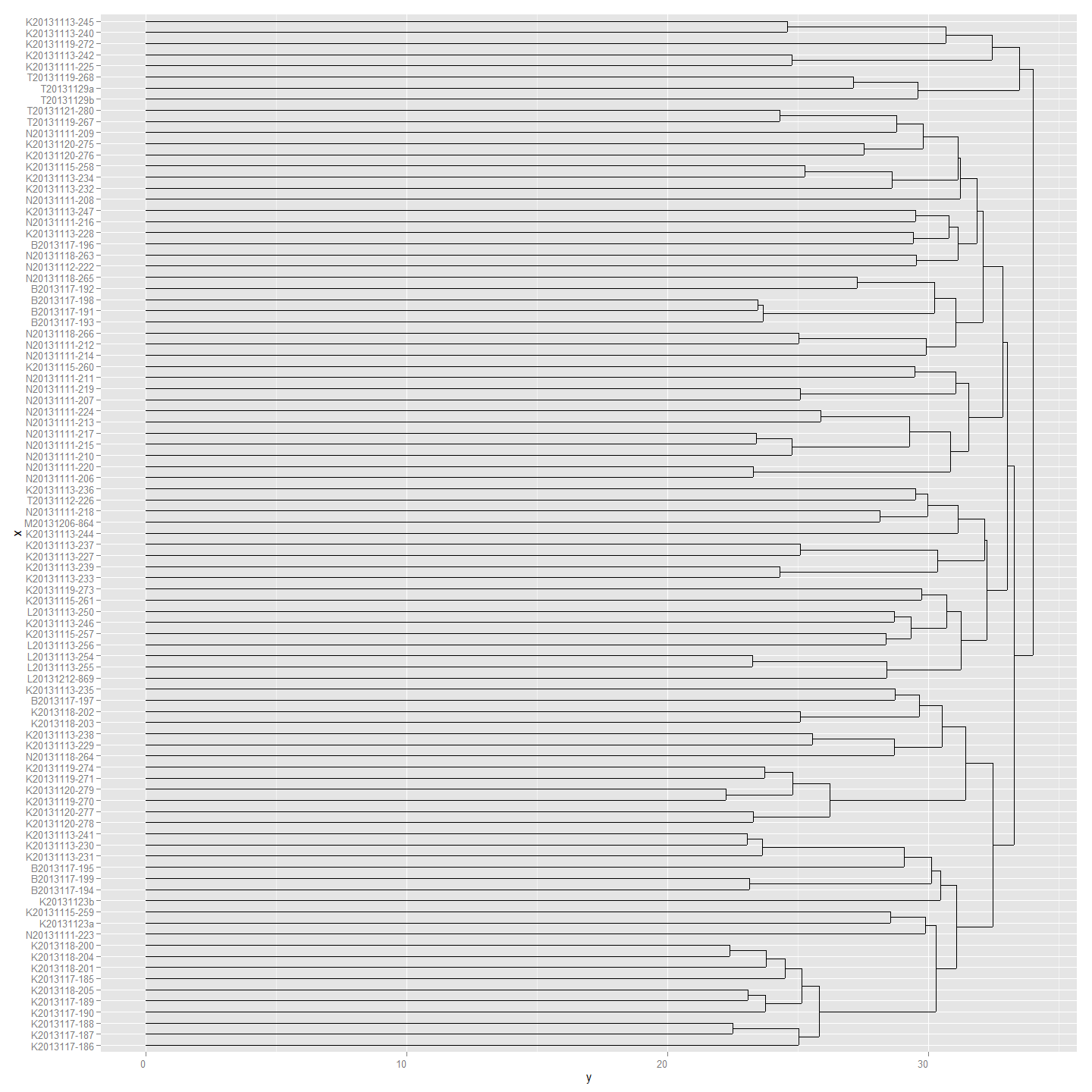


The distance matrix suggest a very low genetic diversity in the population of Murray cod sampled. Given this, how to determine if the closely related larvae are siblings?

In most cases very closely related larvae have been collected at the same spot over the same couple 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.



# basic option  
ggdendrogram(MChc, rotate = TRUE, size = 4, theme\_dendro = FALSE)

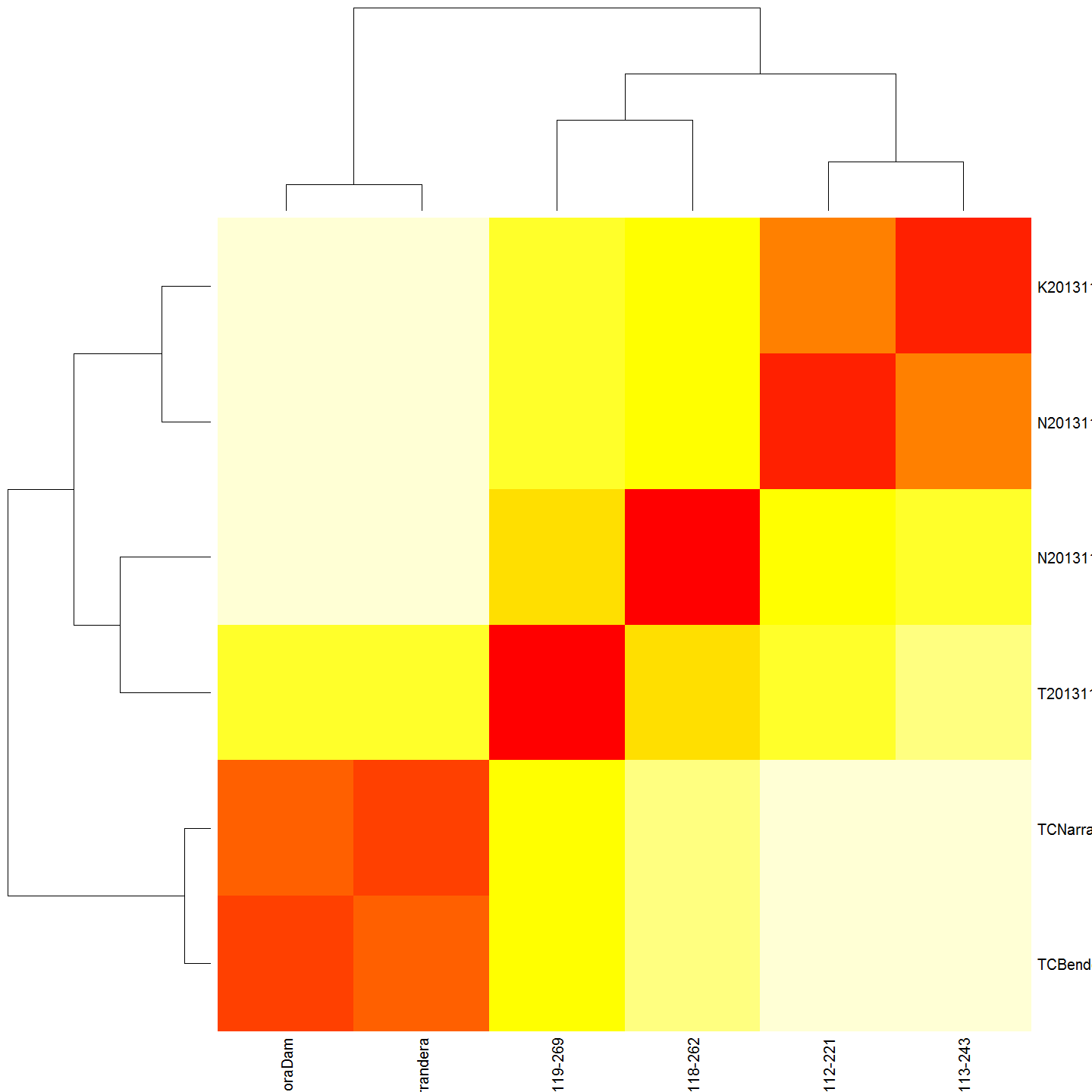
 The Murray cod can be resolved into 12 clades. Can I call these Haplogroups? Is 12 the right sort of cutoff? Should there be more?

#dend1 <- color\_branches(MChc, k = 12)  
# Get the package:  
#cutree(dend1,h=70) # it now works on a dendrogram. Use h or k  
# It is like using:  
#dendextend:::cutree.dendrogram(dend1,h=70)  
  
cladeNo<-cutree(dend1,k=12)  
cladeNo<-as.data.frame(cladeNo)  
cladeNo #The clades are numbered by default.

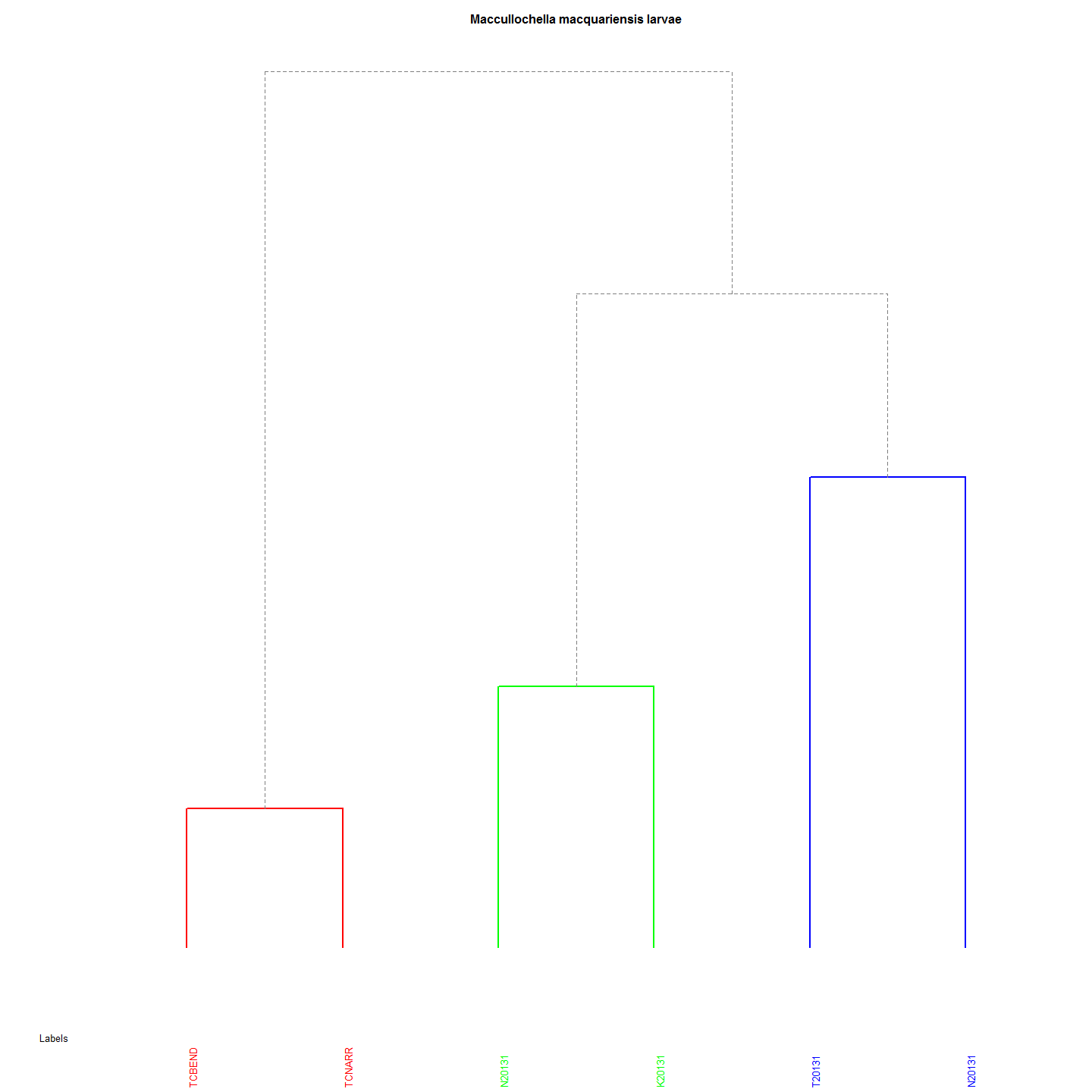
## cladeNo  
## N20131112-222 1  
## L20131212-869 2  
## K20131123a 3  
## K20131123b 3  
## M20131206-864 4  
## T20131129a 5  
## T20131129b 5  
## K2013117-185 3  
## B2013117-194 3  
## B2013117-195 3  
## B2013117-196 1  
## K2013117-186 3  
## K2013117-187 3  
## K2013117-188 3  
## K2013117-189 3  
## K2013117-190 3  
## B2013117-191 6  
## B2013117-192 6  
## B2013117-193 6  
## N20131111-208 7  
## B2013117-199 3  
## B2013117-198 6  
## B2013117-197 8  
## N20131111-207 9  
## N20131111-206 9  
## K2013118-205 3  
## K2013118-204 3  
## K2013118-203 8  
## K2013118-202 8  
## K2013118-201 3  
## K2013118-200 3  
## N20131111-209 7  
## N20131111-218 4  
## N20131111-219 9  
## N20131111-220 9  
## N20131111-210 9  
## N20131111-211 9  
## N20131111-212 6  
## N20131111-213 9  
## N20131111-214 6  
## N20131111-215 9  
## N20131111-216 1  
## N20131111-217 9  
## K20131113-232 7  
## N20131111-223 3  
## T20131119-268 5  
## K20131113-231 3  
## K20131113-230 3  
## K20131113-229 8  
## K20131113-228 1  
## K20131113-227 10  
## T20131112-226 4  
## K20131111-225 11  
## N20131111-224 9  
## K20131113-233 10  
## K20131113-242 11  
## K20131113-244 4  
## K20131113-234 7  
## K20131113-235 8  
## K20131113-236 4  
## K20131113-237 10  
## K20131113-238 8  
## K20131113-239 10  
## K20131113-240 12  
## K20131113-241 3  
## L20131113-256 2  
## K20131113-247 1  
## K20131113-246 2  
## K20131113-245 12  
## L20131113-255 2  
## L20131113-254 2  
## L20131113-250 2  
## K20131115-257 2  
## N20131118-266 6  
## T20131119-267 7  
## K20131115-258 7  
## K20131115-259 3  
## K20131115-260 9  
## K20131115-261 2  
## N20131118-263 1  
## N20131118-264 8  
## N20131118-265 6  
## T20131121-280 7  
## K20131119-271 8  
## K20131119-270 8  
## K20131120-279 8  
## K20131120-278 8  
## K20131120-277 8  
## K20131120-276 7  
## K20131120-275 7  
## K20131119-274 8  
## K20131119-273 2  
## K20131119-272 12

## Trout Cod Larvae

#and a Trout Cod   
TCdm <- dist(TCsnps)  
#Heat map  
dataMatrix <- as.matrix(TCdm)  
heatmap(dataMatrix)



#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")

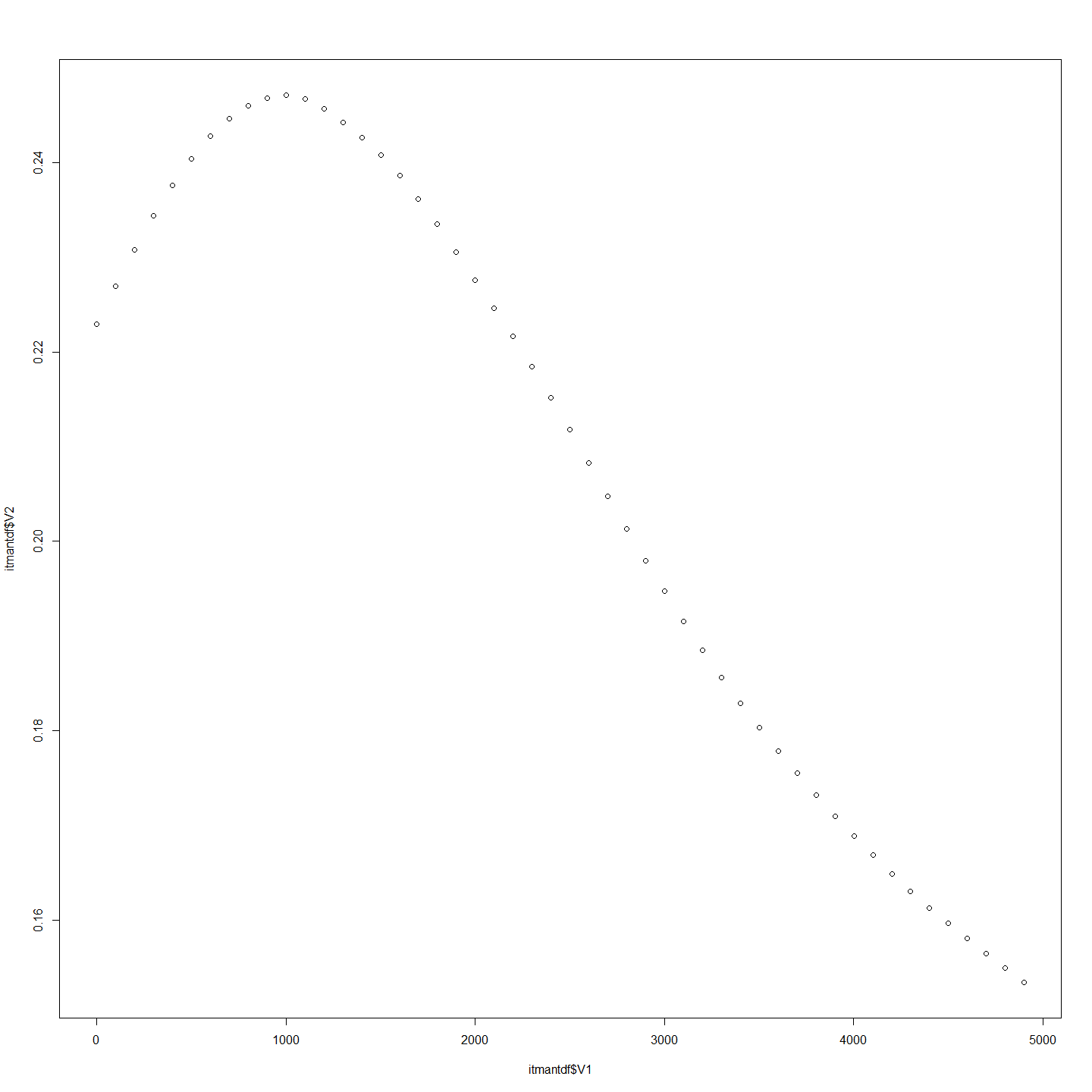


# Upper Murrumbidgee Larval Cod Genetic and Geographic Distances

read chunk (does not run code)

# run the IM chunk (all) IteratedMantel

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)

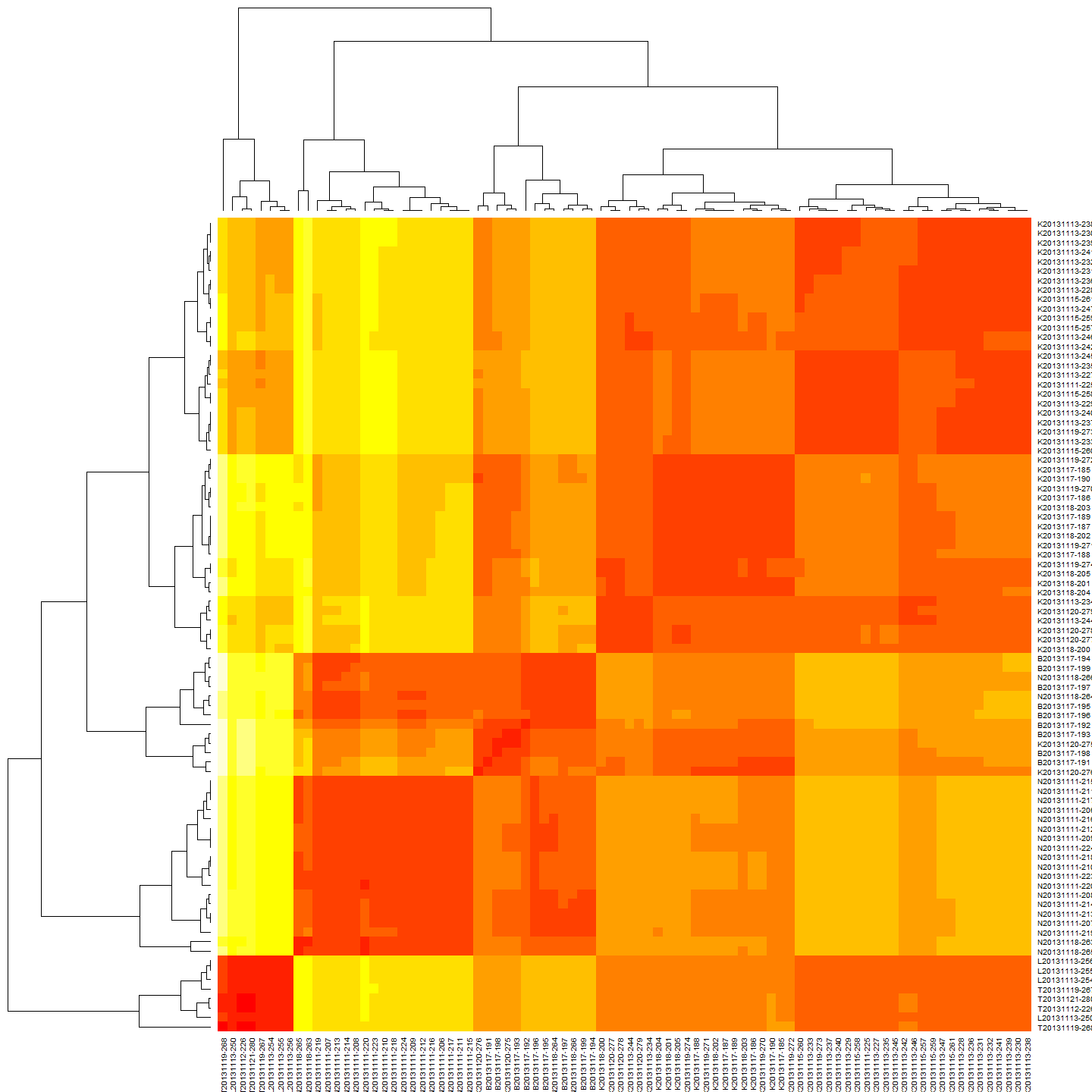


BestNestEst<- subset(itmantdf, V2==max(V2) , select = V1)  
BestNestEst<-as.numeric(BestNestEst)

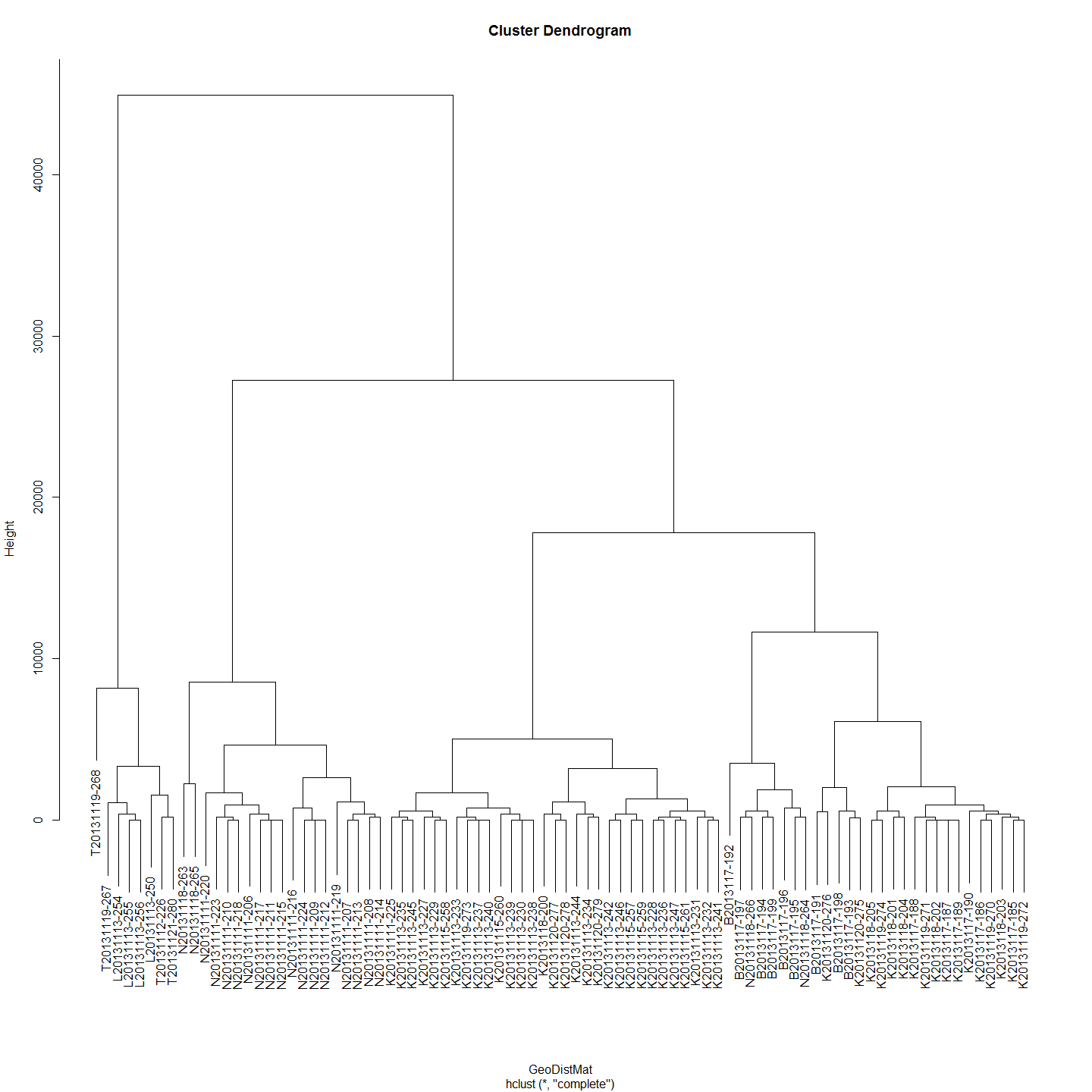
This test corelates the physical distances with genetic distance of larvae from the Murrumbidgee collected in 2013 from 6 sites. By iterating the mantel test using distance matrices generated over varying nest distances and finding the most highly corelated provides a best estimate of the distance a nest site is. In this case it is 1001 *Alan - change this code to be more interactive with the maxima of the iterated mantel.*

There is a possibiltiy that if this was done with mitochondrial sequences and downstram distance matrix the best estimate for the female upward migration might be found if iterate say: 1000-20000m. But also the chemotype might allow use to do something similar with it and geo distance.

# To create a distance  
larv$nestdist<-larv$Distance.to.Angle.Crossing..m.-(BestNestEst\*(larv$Day.of.Year-(larv$hatchdoy+7)))  
  
#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  
  
###########  
# Re creat GenDist from code in dendrograms code file  
# 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)

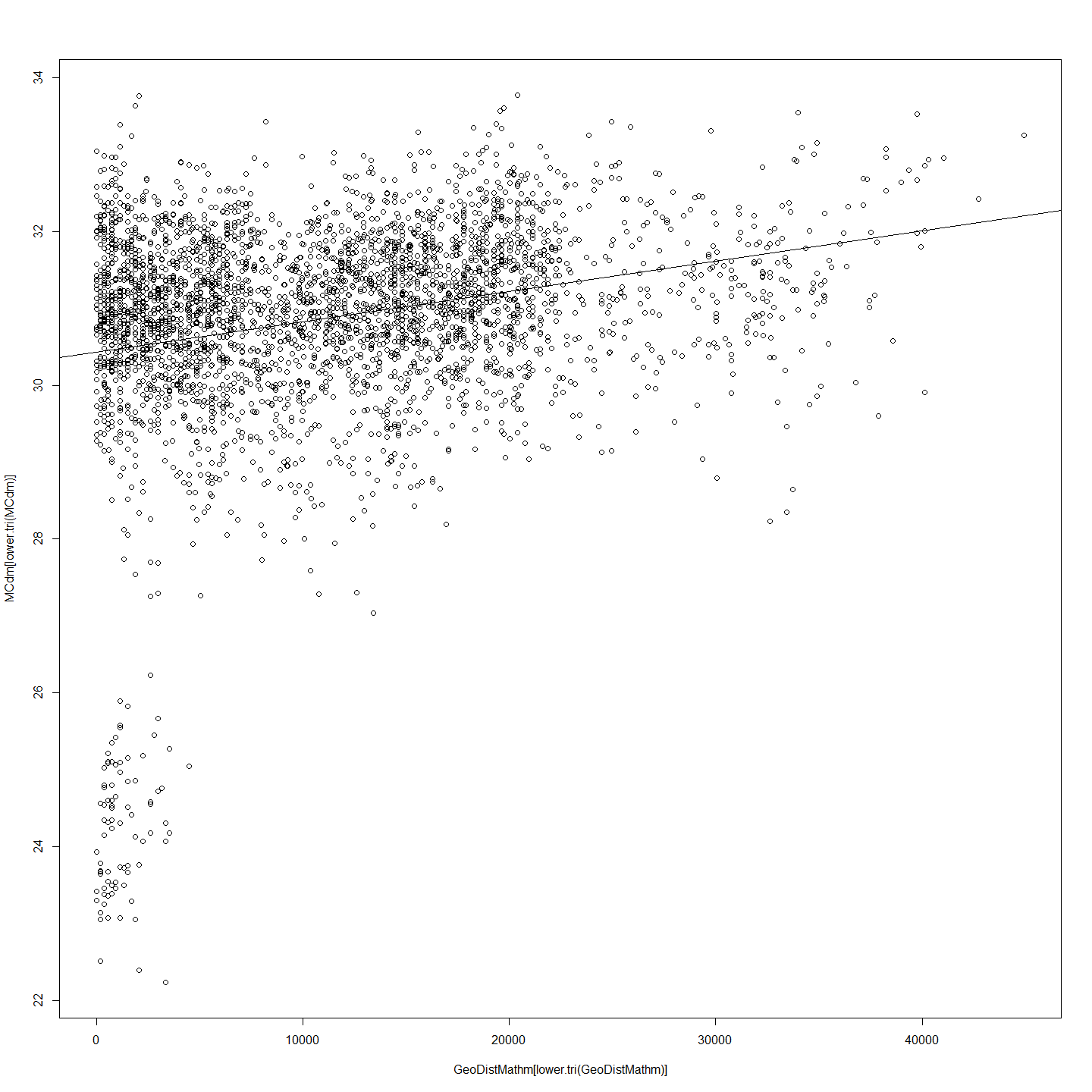
Now that these various matrices, class 'dist' objects are created we can proceed for plot.

## Plot and Correlate genetic and geographic distance matrices

#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

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]

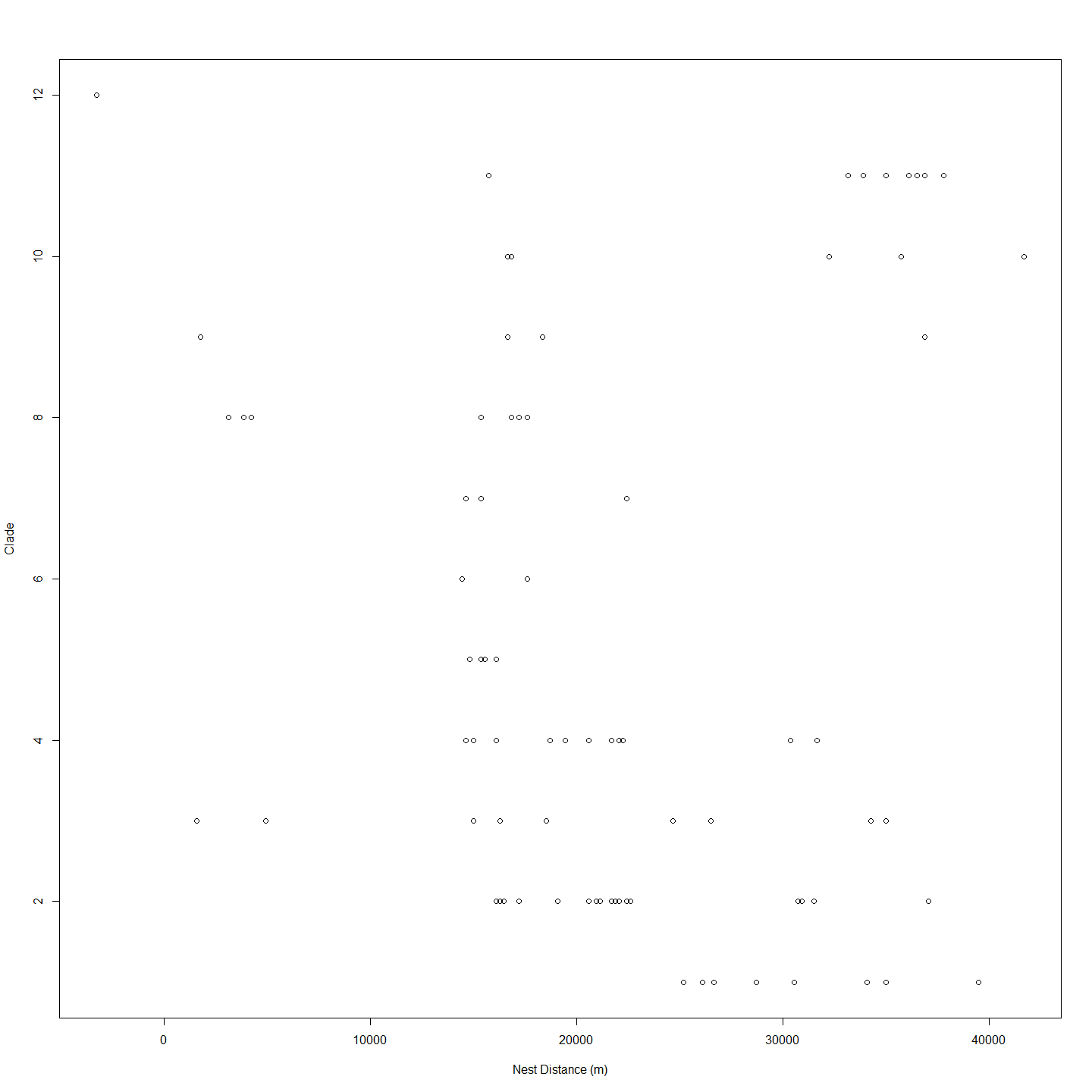
## B2013117-191 B2013117-192 B2013117-193 B2013117-194 B2013117-195  
## B2013117-191 0.00 30.08 23.66 30.13 30.33  
## B2013117-192 30.08 0.00 30.05 28.74 30.06  
## B2013117-193 23.66 30.05 0.00 30.12 29.26  
## B2013117-194 30.13 28.74 30.12 0.00 29.87  
## B2013117-195 30.33 30.06 29.26 29.87 0.00

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. So smaller differences in genotype are generally seen among pairs of larvae that are from nests estimated to be geographicallyclose 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 same code will always arrive at the same observed correlation but rarely the same p-value.

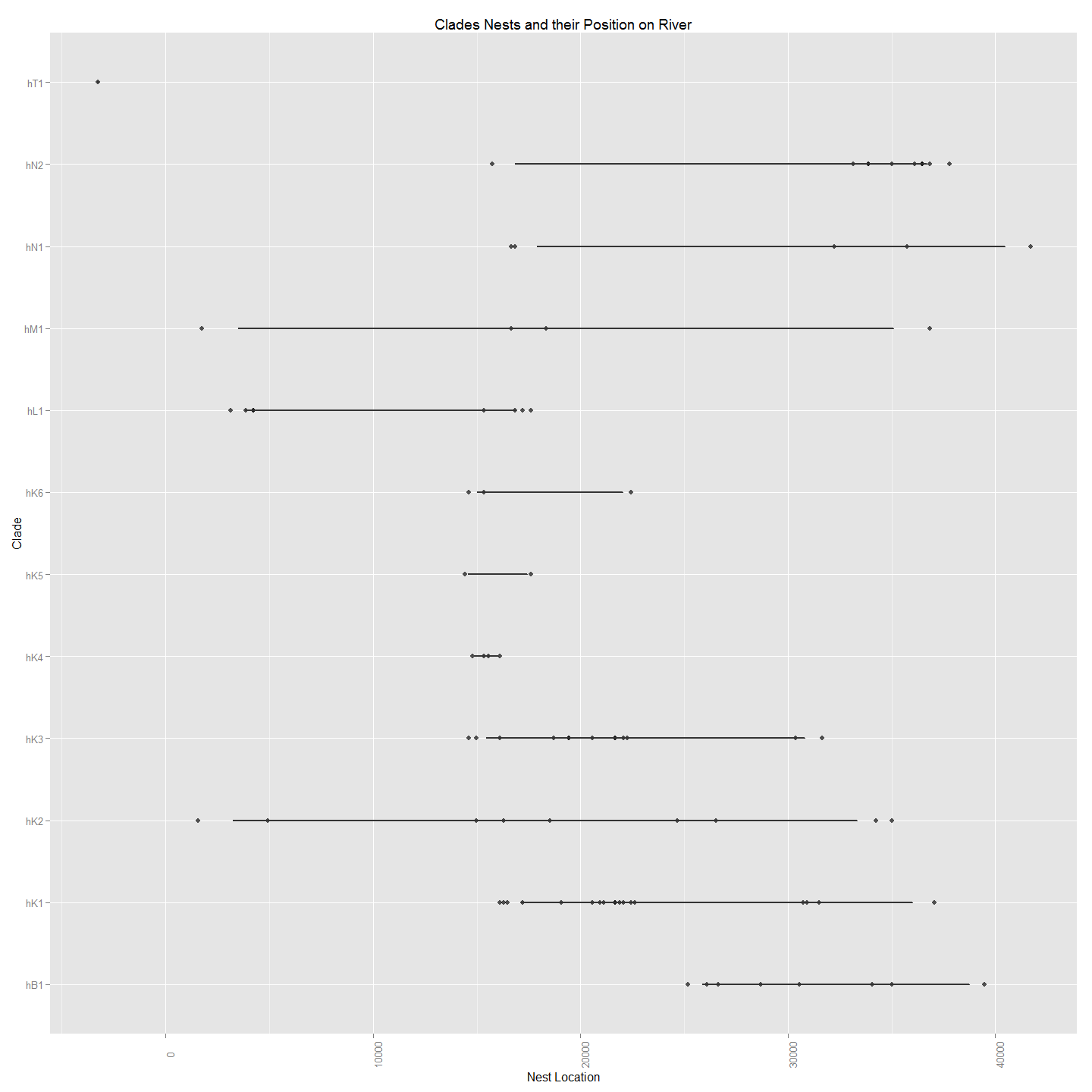
The question needs to be asked, and answered, as to how the 1000m distance used in the nest distance estimation equation was derived. This was estimated by iterating the Mantel test for each distance matrix calculated after calculating that matrix using a 'larval dispersal' distance from 1 m to 5000 metres.The highest corellation suggests that it is the distance that best represents the average distance that larvae disperse. The curve produced from the estimations is as follows.

## Look at Haplogroups Over the River Reach

larv1<-merge(larv,cladeNo, by="row.names")  
rownames(CladeNamesToMerge) <- CladeNamesToMerge[,1]  
CladeNamesToMerge$Label<-rownames(CladeNamesToMerge)  
larv2<-merge(larv1,CladeNamesToMerge, by="Label")  
plot(larv2$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 at 10000m 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).

## Code Chunks in this Document

## [1] "Project\_Template\_and\_Knitr" "Set\_Global\_Options" "LoadLibraries" "Additional\_Calculations" "All\_Maccullochella\_Larvae"   
## [6] "Addicted\_2\_R\_All" "Murray\_Cod\_Larvae\_Only" "A2R\_Dendrogram" "Another\_Dendrogram" "ExtractClades"   
## [11] "Trout\_Cod\_Dendrograms" "unnamed-chunk-1" "IteratedMantelGo" "Distance\_Matrices\_and\_Ordering" "Plots\_and\_Correlation"   
## [16] "Mantel\_Test" "Haplogroups\_and\_Location" "Include\_Labels\_for\_all\_code\_Chunks" "IM"