SummaryPlot

Xiang Ji

December 29, 2014

This R script is used for generating summary plot for the Geneconv project

1, Read in tables

rm(list=ls()) # clean up workspace  
#path <- "/Users/xji3/Genconv/NewClusterPackRun/NewPackageNewRun/OldResults01152015/"  
path <- "/Users/xji3/Genconv/NewClusterPackRun/NewPackageNewRun/"  
#path <- "G:/Geneconv/NewClusterPackRun/NewPackageNewRun/"  
#HKY\_clock\_summary <- "HKY\_clock\_summary"  
summary.list <- c( "HKY\_nonclock\_summary",  
 "HKY\_clock\_summary",  
 "MG94\_clock\_summary",  
 "MG94\_nonclock\_summary",   
 "Force\_HKY\_clock\_summary",   
 "Force\_HKY\_nonclock\_summary",  
 "Force\_MG94\_clock\_summary",   
 "Force\_MG94\_nonclock\_summary"  
 )  
#summary.list <- c("Force\_MG94\_clock\_summary")  
for (target.summary in summary.list){  
 summary\_file <- paste(path, target.summary, '.txt', sep = '')  
 all <- readLines(summary\_file, n = -1)  
 col.names <- strsplit(all[1], ' ')[[1]][-1]  
 row.names <- strsplit(all[length(all)], ' ')[[1]][-1]  
 summary\_mat <- as.matrix(read.table(summary\_file,   
 row.names = row.names,   
 col.names = col.names))  
 assign(target.summary, summary\_mat)  
}

Now generate summary file of only pairs that have all cases finished in HKY or MG94 models.

# HKY  
HKY.pair.names <- intersect(intersect(intersect(colnames(HKY\_clock\_summary),   
 colnames(HKY\_nonclock\_summary)),  
 colnames(Force\_HKY\_clock\_summary)),  
 colnames(Force\_HKY\_nonclock\_summary))  
HKY.clock.filtered <- HKY\_clock\_summary[, HKY.pair.names]  
HKY.nonclock.filtered <- HKY\_nonclock\_summary[, HKY.pair.names]  
HKY.Force.clock.filtered <- Force\_HKY\_clock\_summary[, HKY.pair.names]  
HKY.Force.nonclock.filtered <- Force\_HKY\_nonclock\_summary[, HKY.pair.names]  
  
write.table(HKY.clock.filtered, paste( path, "HKY\_clock\_filtered", sep = ""))  
write.table(HKY.nonclock.filtered, paste( path, "HKY\_nonclock\_filtered", sep = ""))  
write.table(HKY.Force.clock.filtered, paste( path, "HKY\_Force\_clock\_filtered", sep = ""))  
write.table(HKY.Force.nonclock.filtered, paste( path, "HKY\_Force\_nonclock\_filtered", sep = ""))  
  
  
  
# MG94  
MG94.pair.names <- intersect(intersect(intersect(colnames(MG94\_clock\_summary),   
 colnames(MG94\_nonclock\_summary)),  
 colnames(Force\_MG94\_clock\_summary)),  
 colnames(Force\_MG94\_nonclock\_summary))  
MG94.clock.filtered <- MG94\_clock\_summary[, MG94.pair.names]  
MG94.nonclock.filtered <- MG94\_nonclock\_summary[, MG94.pair.names]  
MG94.Force.clock.filtered <- Force\_MG94\_clock\_summary[, MG94.pair.names]  
MG94.Force.nonclock.filtered <- Force\_MG94\_nonclock\_summary[, MG94.pair.names]  
  
write.table(MG94.clock.filtered, paste( path, "MG94\_clock\_filtered", sep = ""))  
write.table(MG94.nonclock.filtered, paste( path, "MG94\_nonclock\_filtered", sep = ""))  
write.table(MG94.Force.clock.filtered, paste( path, "MG94\_Force\_clock\_filtered", sep = ""))  
write.table(MG94.Force.nonclock.filtered, paste( path, "MG94\_Force\_nonclock\_filtered", sep = ""))

Now analyze the results

First, show the loglikelihood improvement for each model with/without tau

# HKY nonclock  
(HKY.nonclock.filtered - HKY.Force.nonclock.filtered)[2,]

## YLR406C\_YDL075W YER131W\_YGL189C YDR502C\_YLR180W YHR106W\_YDR353W   
## 45.32 84.73 361.25 165.28   
## YIL057C\_YER067W YPL087W\_YBR183W YNL069C\_YIL133C YDR518W\_YCL043C   
## 42.03 46.49 90.04 42.93   
## YGR043C\_YLR354C YPR157W\_YGR141W YDR099W\_YER177W YBR024W\_YBR037C   
## 60.11 73.59 156.46 44.68   
## YPL037C\_YDR252W YPR159W\_YGR143W YGL133W\_YPL216W YBR191W\_YPL079W   
## 13.36 230.93 27.77 120.41   
## YNL049C\_YIL109C YPL232W\_YMR183C YLR284C\_YOR180C YIR033W\_YKL020C   
## 158.20 109.92 16.80 94.99   
## YMR243C\_YOR316C YER102W\_YBL072C YAL056W\_YOR371C YDR438W\_YML018C   
## 85.48 232.13 22.94 17.39

# HKY clock  
(HKY.clock.filtered - HKY.Force.clock.filtered)[2,]

## YLR406C\_YDL075W YER131W\_YGL189C YDR502C\_YLR180W YHR106W\_YDR353W   
## 45.671 47.253 169.574 145.207   
## YIL057C\_YER067W YPL087W\_YBR183W YNL069C\_YIL133C YDR518W\_YCL043C   
## 41.107 39.588 77.690 34.149   
## YGR043C\_YLR354C YPR157W\_YGR141W YDR099W\_YER177W YBR024W\_YBR037C   
## 40.567 63.727 152.259 32.103   
## YPL037C\_YDR252W YPR159W\_YGR143W YGL133W\_YPL216W YBR191W\_YPL079W   
## 15.203 231.875 11.503 119.181   
## YNL049C\_YIL109C YPL232W\_YMR183C YLR284C\_YOR180C YIR033W\_YKL020C   
## 121.540 92.364 17.778 76.723   
## YMR243C\_YOR316C YER102W\_YBL072C YAL056W\_YOR371C YDR438W\_YML018C   
## 66.710 36569.553 26.628 7.088

# MG94 nonclock  
(MG94.nonclock.filtered - MG94.Force.nonclock.filtered)[2,]

## YLR406C\_YDL075W YER131W\_YGL189C YDR502C\_YLR180W YHR106W\_YDR353W   
## 16.9829 15.6925 45.7337 22.2515   
## YIL057C\_YER067W YNL069C\_YIL133C YDR518W\_YCL043C YGR043C\_YLR354C   
## 16.4041 58.6072 33.7744 5.5141   
## YDR099W\_YER177W YMR143W\_YDL083C YJR048W\_YEL039C YBR191W\_YPL079W   
## 14.2549 37.2294 21.5963 66.1023   
## YDR418W\_YEL054C YPL232W\_YMR183C YLR284C\_YOR180C YBL087C\_YER117W   
## 32.0401 30.0242 0.8918 47.6112   
## YGL062W\_YBR218C YIR033W\_YKL020C YER102W\_YBL072C YAL056W\_YOR371C   
## 442.2758 30.9337 137.7021 14.6457   
## YDR438W\_YML018C   
## 3.2045

# MG94 clock  
(MG94.clock.filtered - MG94.Force.clock.filtered)[2,]

## YLR406C\_YDL075W YER131W\_YGL189C YDR502C\_YLR180W YHR106W\_YDR353W   
## 17.42 16.95 245.34 520.78   
## YIL057C\_YER067W YNL069C\_YIL133C YDR518W\_YCL043C YGR043C\_YLR354C   
## 579.66 53.30 85.02 715.54   
## YDR099W\_YER177W YMR143W\_YDL083C YJR048W\_YEL039C YBR191W\_YPL079W   
## 213.01 45.36 179.97 71.69   
## YDR418W\_YEL054C YPL232W\_YMR183C YLR284C\_YOR180C YBL087C\_YER117W   
## 34.90 876.14 41.58 49.55   
## YGL062W\_YBR218C YIR033W\_YKL020C YER102W\_YBL072C YAL056W\_YOR371C   
## 2027.02 150.79 149.31 132.01   
## YDR438W\_YML018C   
## -96.55

Then, show the edge specific tau estimates (posterior expected number of geneconv events / posterior expected time in heterogeneous states of each branch)

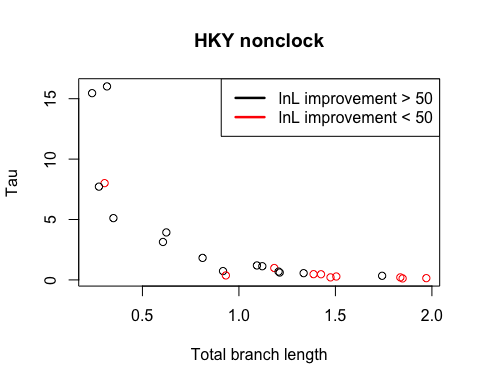
HKY.nonclock.filtered[21:32, ]

## YLR406C\_YDL075W YER131W\_YGL189C YDR502C\_YLR180W  
## (N0,N1,tau) 8.019 7.160 2.941  
## (N0,kluyveri,tau) 0.000 0.000 0.000  
## (N1,N2,tau) 10.018 8.121 3.115  
## (N1,castellii,tau) 9.963 7.194 3.168  
## (N2,N3,tau) 5.498 8.104 2.829  
## (N2,bayanus,tau) 8.313 8.056 3.167  
## (N3,N4,tau) 5.518 6.568 3.246  
## (N3,kudriavzevii,tau) 4.164 7.661 3.551  
## (N4,N5,tau) 12.541 6.631 3.241  
## (N4,mikatae,tau) 7.695 8.282 3.045  
## (N5,cerevisiae,tau) 6.502 11.615 3.402  
## (N5,paradoxus,tau) 8.498 4.075 2.919  
## YHR106W\_YDR353W YIL057C\_YER067W YPL087W\_YBR183W  
## (N0,N1,tau) 1.802 0.9864 0.4710  
## (N0,kluyveri,tau) 0.000 0.0000 0.0000  
## (N1,N2,tau) 1.573 0.9380 0.4247  
## (N1,castellii,tau) 2.227 1.1657 0.5067  
## (N2,N3,tau) 1.652 0.9198 0.3093  
## (N2,bayanus,tau) 1.820 0.7903 0.4961  
## (N3,N4,tau) 1.744 0.7104 0.4495  
## (N3,kudriavzevii,tau) 1.682 0.9511 0.4387  
## (N4,N5,tau) 1.677 0.6356 0.4750  
## (N4,mikatae,tau) 1.791 1.1071 0.4834  
## (N5,cerevisiae,tau) 2.122 0.7710 0.5486  
## (N5,paradoxus,tau) 1.706 1.1633 0.4166  
## YNL069C\_YIL133C YDR518W\_YCL043C YGR043C\_YLR354C  
## (N0,N1,tau) 5.099 0.2832 0.7236  
## (N0,kluyveri,tau) 0.000 0.0000 0.0000  
## (N1,N2,tau) 5.060 0.2610 0.6623  
## (N1,castellii,tau) 6.307 0.3078 0.7191  
## (N2,N3,tau) 6.178 0.2117 0.6607  
## (N2,bayanus,tau) 5.876 0.2639 0.7236  
## (N3,N4,tau) 6.453 0.3089 0.7185  
## (N3,kudriavzevii,tau) 3.808 0.3053 0.6703  
## (N4,N5,tau) 4.843 0.2913 0.6791  
## (N4,mikatae,tau) 3.037 0.2730 0.8822  
## (N5,cerevisiae,tau) 4.258 0.3071 0.8880  
## (N5,paradoxus,tau) 4.388 0.2862 0.7330  
## YPR157W\_YGR141W YDR099W\_YER177W YBR024W\_YBR037C  
## (N0,N1,tau) 0.5579 3.854 0.4734  
## (N0,kluyveri,tau) 0.0000 0.000 0.0000  
## (N1,N2,tau) 0.5160 3.264 0.4649  
## (N1,castellii,tau) 0.5424 4.822 0.4982  
## (N2,N3,tau) 0.4070 3.452 0.4402  
## (N2,bayanus,tau) 0.5768 3.334 0.3923  
## (N3,N4,tau) 0.6953 3.518 0.6368  
## (N3,kudriavzevii,tau) 0.6129 4.400 0.4858  
## (N4,N5,tau) 0.6484 3.198 0.3822  
## (N4,mikatae,tau) 0.5480 4.732 0.4951  
## (N5,cerevisiae,tau) 0.6488 3.504 0.4750  
## (N5,paradoxus,tau) 0.5630 3.693 0.4281  
## YPL037C\_YDR252W YPR159W\_YGR143W YGL133W\_YPL216W  
## (N0,N1,tau) 0.3711 1.1321 0.1253  
## (N0,kluyveri,tau) 0.0000 0.0000 0.0000  
## (N1,N2,tau) 0.3195 1.1458 0.1281  
## (N1,castellii,tau) 0.5589 1.1930 0.1322  
## (N2,N3,tau) 0.3056 0.9359 0.1207  
## (N2,bayanus,tau) 0.3524 1.1857 0.1056  
## (N3,N4,tau) 0.4999 0.9415 0.1245  
## (N3,kudriavzevii,tau) 0.3871 1.0777 0.1234  
## (N4,N5,tau) 0.4005 1.1148 0.1329  
## (N4,mikatae,tau) 0.4121 1.1804 0.1224  
## (N5,cerevisiae,tau) 0.3479 0.9478 0.1248  
## (N5,paradoxus,tau) 0.3072 1.2585 0.1080  
## YBR191W\_YPL079W YNL049C\_YIL109C YPL232W\_YMR183C  
## (N0,N1,tau) 14.46 0.6129 1.190  
## (N0,kluyveri,tau) 0.00 0.0000 0.000  
## (N1,N2,tau) 14.12 0.6323 1.230  
## (N1,castellii,tau) 15.50 0.6125 1.207  
## (N2,N3,tau) 16.26 0.5488 1.040  
## (N2,bayanus,tau) 20.45 0.6395 1.259  
## (N3,N4,tau) 17.97 0.5647 1.225  
## (N3,kudriavzevii,tau) 12.28 0.5624 1.107  
## (N4,N5,tau) 15.62 0.5543 1.268  
## (N4,mikatae,tau) 20.28 0.6171 1.021  
## (N5,cerevisiae,tau) 11.88 0.6070 1.231  
## (N5,paradoxus,tau) 20.00 0.6765 1.354  
## YLR284C\_YOR180C YIR033W\_YKL020C YMR243C\_YOR316C  
## (N0,N1,tau) 0.2077 0.3406 0.6861  
## (N0,kluyveri,tau) 0.0000 0.0000 0.0000  
## (N1,N2,tau) 0.2003 0.3493 0.7023  
## (N1,castellii,tau) 0.1952 0.3533 0.8132  
## (N2,N3,tau) 0.2096 0.3381 0.5935  
## (N2,bayanus,tau) 0.1937 0.3502 0.6734  
## (N3,N4,tau) 0.1925 0.2873 0.6197  
## (N3,kudriavzevii,tau) 0.2322 0.3304 0.6138  
## (N4,N5,tau) 0.2711 0.3050 0.5942  
## (N4,mikatae,tau) 0.2282 0.2909 0.6334  
## (N5,cerevisiae,tau) 0.2044 0.3200 0.6758  
## (N5,paradoxus,tau) 0.2048 0.3400 0.5868  
## YER102W\_YBL072C YAL056W\_YOR371C YDR438W\_YML018C  
## (N0,N1,tau) 13.838 0.1457 0.2115  
## (N0,kluyveri,tau) 0.000 0.0000 0.0000  
## (N1,N2,tau) 15.688 0.1431 0.2220  
## (N1,castellii,tau) 15.348 0.1410 0.2395  
## (N2,N3,tau) 16.279 0.1332 0.1775  
## (N2,bayanus,tau) 19.523 0.1338 0.2195  
## (N3,N4,tau) 23.216 0.1437 0.1697  
## (N3,kudriavzevii,tau) 9.835 0.1365 0.1661  
## (N4,N5,tau) 32.514 0.1559 0.1884  
## (N4,mikatae,tau) 15.832 0.1472 0.1991  
## (N5,cerevisiae,tau) 48.916 0.2036 0.1565  
## (N5,paradoxus,tau) 23.212 0.1532 0.1921

Now plot Total blen v.s. Tau into different groups (differ by color)

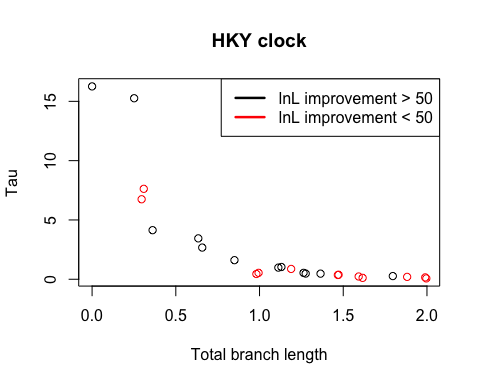
HKY nonclock case

plot(colSums(HKY.Force.nonclock.filtered[9:20, ]), HKY.nonclock.filtered[8, ],  
 type = "n", xlab = "Total branch length", ylab = "Tau" )  
col.color <- rep("black", dim(HKY.nonclock.filtered)[2])  
col.color[(HKY.nonclock.filtered - HKY.Force.nonclock.filtered)[2,] < 50] <- "red"  
points(x = colSums(HKY.Force.nonclock.filtered[9:20, ]), y = HKY.nonclock.filtered[8, ],  
 type = "p", col = col.color, bg = col.color)  
legend("topright", c("lnL improvement > 50", "lnL improvement < 50"),  
 lty = c(1, 1),   
 lwd = c(2.5, 2.5),  
 col = c("black", "red"))  
title("HKY nonclock")



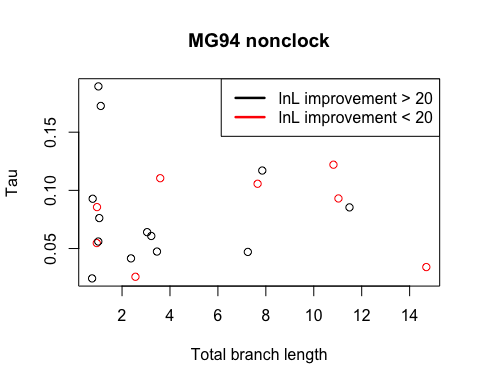
HKY clock case

plot(colSums(HKY.Force.clock.filtered[9:20, ]), HKY.clock.filtered[8, ],  
 type = "n", xlab = "Total branch length", ylab = "Tau" )  
col.color <- rep("black", dim(HKY.clock.filtered)[2])  
col.color[(HKY.clock.filtered - HKY.Force.clock.filtered)[2,] < 50] <- "red"  
points(x = colSums(HKY.Force.clock.filtered[9:20, ]), y = HKY.clock.filtered[8, ],  
 type = "p", col = col.color, bg = col.color)  
legend("topright", c("lnL improvement > 50", "lnL improvement < 50"),  
 lty = c(1, 1),   
 lwd = c(2.5, 2.5),  
 col = c("black", "red"))  
title("HKY clock")



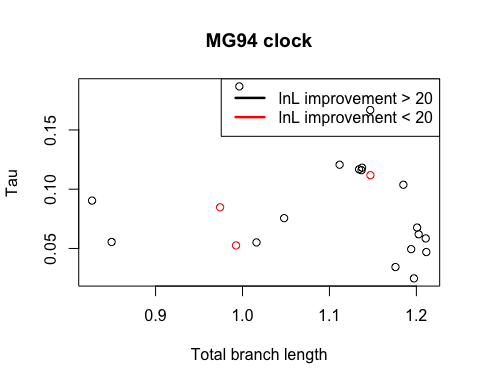
MG94 nonclock case

improvement.lmt <- 20  
plot(colSums(MG94.Force.nonclock.filtered[10:21, ]), MG94.nonclock.filtered[8, ],  
 type = "n", xlab = "Total branch length", ylab = "Tau" )  
col.color <- rep("black", dim(MG94.nonclock.filtered)[2])  
col.color[(MG94.nonclock.filtered - MG94.Force.nonclock.filtered)[2,] < improvement.lmt] <- "red"  
points(x = colSums(MG94.Force.nonclock.filtered[10:21, ]), y = MG94.nonclock.filtered[8, ],  
 type = "p", col = col.color, bg = col.color)  
legend("topright",   
 c(paste("lnL improvement > ", toString(improvement.lmt), sep = ""),  
 paste("lnL improvement < ", toString(improvement.lmt), sep = "")),  
 lty = c(1, 1),   
 lwd = c(2.5, 2.5),  
 col = c("black", "red"))  
title("MG94 nonclock")



MG94 clock case

improvement.lmt <- 20  
plot(colSums(MG94.Force.clock.filtered[10:21, ]), MG94.clock.filtered[8, ],  
 type = "n", xlab = "Total branch length", ylab = "Tau" )  
col.color <- rep("black", dim(MG94.clock.filtered)[2])  
col.color[(MG94.clock.filtered - MG94.Force.clock.filtered)[2,] < improvement.lmt] <- "red"  
points(x = colSums(MG94.Force.clock.filtered[10:21, ]), y = MG94.clock.filtered[8, ],  
 type = "p", col = col.color, bg = col.color)  
legend("topright",   
 c(paste("lnL improvement > ", toString(improvement.lmt), sep = ""),  
 paste("lnL improvement < ", toString(improvement.lmt), sep = "")),  
 lty = c(1, 1),   
 lwd = c(2.5, 2.5),  
 col = c("black", "red"))  
title("MG94 clock")



Now see if the pairs red in HKY are also red in MG94

nonclock case

HKY.pair.names[(HKY.nonclock.filtered - HKY.Force.nonclock.filtered)[2,] < 50]

## [1] "YLR406C\_YDL075W" "YIL057C\_YER067W" "YPL087W\_YBR183W"  
## [4] "YDR518W\_YCL043C" "YBR024W\_YBR037C" "YPL037C\_YDR252W"  
## [7] "YGL133W\_YPL216W" "YLR284C\_YOR180C" "YAL056W\_YOR371C"  
## [10] "YDR438W\_YML018C"

MG94.pair.names[(MG94.nonclock.filtered - MG94.Force.nonclock.filtered)[2,] < improvement.lmt]

## [1] "YLR406C\_YDL075W" "YER131W\_YGL189C" "YIL057C\_YER067W" "YGR043C\_YLR354C"  
## [5] "YDR099W\_YER177W" "YLR284C\_YOR180C" "YAL056W\_YOR371C" "YDR438W\_YML018C"

Only 3 pairs show up in both: YLR406C\_YDL075W, YIL057C\_YER067W, YDR438W\_YML018C.

clock case

HKY.pair.names[(HKY.clock.filtered - HKY.Force.clock.filtered)[2,] < 50]

## [1] "YLR406C\_YDL075W" "YER131W\_YGL189C" "YIL057C\_YER067W"  
## [4] "YPL087W\_YBR183W" "YDR518W\_YCL043C" "YGR043C\_YLR354C"  
## [7] "YBR024W\_YBR037C" "YPL037C\_YDR252W" "YGL133W\_YPL216W"  
## [10] "YLR284C\_YOR180C" "YAL056W\_YOR371C" "YDR438W\_YML018C"

MG94.pair.names[(MG94.clock.filtered - MG94.Force.clock.filtered)[2,] < improvement.lmt]

## [1] "YLR406C\_YDL075W" "YER131W\_YGL189C" "YDR438W\_YML018C"

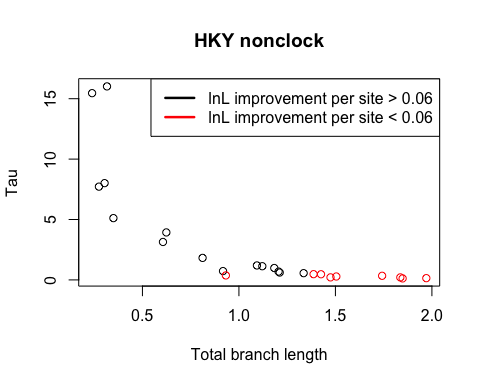
2 pairs show up in both: YLR406C\_YDL075W, YDR438W\_YML018C

=========================================

Now plot Total blen v.s. Tau into different groups (differ by lnL improvement per site)

HKY nonclock case

plot(colSums(HKY.Force.nonclock.filtered[9:20, ]), HKY.nonclock.filtered[8, ],  
 type = "n", xlab = "Total branch length", ylab = "Tau" )  
col.color <- rep("black", dim(HKY.nonclock.filtered)[2])  
col.color[((HKY.nonclock.filtered - HKY.Force.nonclock.filtered)[2,] / HKY.nonclock.filtered[1, ]) < 0.06] <- "red"  
points(x = colSums(HKY.Force.nonclock.filtered[9:20, ]), y = HKY.nonclock.filtered[8, ],  
 type = "p", col = col.color, bg = col.color)  
legend("topright", c("lnL improvement per site > 0.06",  
 "lnL improvement per site < 0.06"),  
 lty = c(1, 1),   
 lwd = c(2.5, 2.5),  
 col = c("black", "red"))  
title("HKY nonclock")

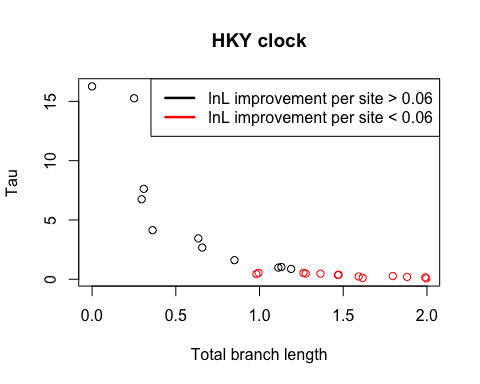


# lnL improvement per site  
((HKY.nonclock.filtered - HKY.Force.nonclock.filtered)[2,] / HKY.nonclock.filtered[1, ])

## YLR406C\_YDL075W YER131W\_YGL189C YDR502C\_YLR180W YHR106W\_YDR353W   
## 0.134895 0.239360 0.316054 0.173246   
## YIL057C\_YER067W YPL087W\_YBR183W YNL069C\_YIL133C YDR518W\_YCL043C   
## 0.090389 0.050645 0.152353 0.028735   
## YGR043C\_YLR354C YPR157W\_YGR141W YDR099W\_YER177W YBR024W\_YBR037C   
## 0.060532 0.060868 0.210301 0.053765   
## YPL037C\_YDR252W YPR159W\_YGR143W YGL133W\_YPL216W YBR191W\_YPL079W   
## 0.031368 0.115064 0.009705 0.252429   
## YNL049C\_YIL109C YPL232W\_YMR183C YLR284C\_YOR180C YIR033W\_YKL020C   
## 0.063228 0.127221 0.020901 0.037829   
## YMR243C\_YOR316C YER102W\_YBL072C YAL056W\_YOR371C YDR438W\_YML018C   
## 0.076393 0.390796 0.011672 0.017508

HKY clock case

plot(colSums(HKY.Force.clock.filtered[9:20, ]), HKY.clock.filtered[8, ],  
 type = "n", xlab = "Total branch length", ylab = "Tau" )  
col.color <- rep("black", dim(HKY.clock.filtered)[2])  
col.color[((HKY.clock.filtered - HKY.Force.clock.filtered)[2,] / HKY.clock.filtered[1, ]) < 0.06] <- "red"  
points(x = colSums(HKY.Force.clock.filtered[9:20, ]), y = HKY.clock.filtered[8, ],  
 type = "p", col = col.color, bg = col.color)  
legend("topright", c("lnL improvement per site > 0.06",  
 "lnL improvement per site < 0.06"),  
 lty = c(1, 1),   
 lwd = c(2.5, 2.5),  
 col = c("black", "red"))  
title("HKY clock")

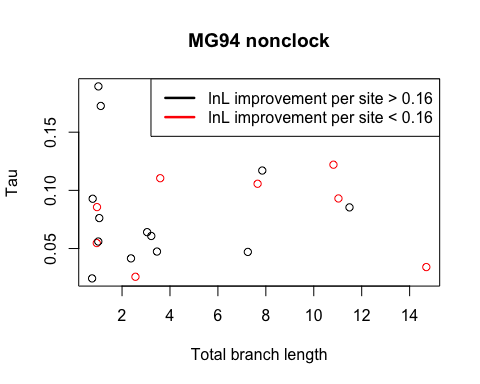


# lnL improvement per site  
((HKY.clock.filtered - HKY.Force.clock.filtered)[2,] / HKY.clock.filtered[1, ])

## YLR406C\_YDL075W YER131W\_YGL189C YDR502C\_YLR180W YHR106W\_YDR353W   
## 0.135924 0.133484 0.148359 0.152208   
## YIL057C\_YER067W YPL087W\_YBR183W YNL069C\_YIL133C YDR518W\_YCL043C   
## 0.088402 0.043125 0.131456 0.022857   
## YGR043C\_YLR354C YPR157W\_YGR141W YDR099W\_YER177W YBR024W\_YBR037C   
## 0.040853 0.052711 0.204649 0.038632   
## YPL037C\_YDR252W YPR159W\_YGR143W YGL133W\_YPL216W YBR191W\_YPL079W   
## 0.035687 0.115533 0.004019 0.249856   
## YNL049C\_YIL109C YPL232W\_YMR183C YLR284C\_YOR180C YIR033W\_YKL020C   
## 0.048577 0.106903 0.022112 0.030555   
## YMR243C\_YOR316C YER102W\_YBL072C YAL056W\_YOR371C YDR438W\_YML018C   
## 0.059616 61.564903 0.013551 0.007138

MG94 nonclock case

improvement.lmt <- 0.16  
plot(colSums(MG94.Force.nonclock.filtered[10:21, ]), MG94.nonclock.filtered[8, ],  
 type = "n", xlab = "Total branch length", ylab = "Tau" )  
col.color <- rep("black", dim(MG94.nonclock.filtered)[2])  
col.color[((MG94.nonclock.filtered - MG94.Force.nonclock.filtered)[2,] / MG94.nonclock.filtered[1]) < improvement.lmt] <- "red"  
points(x = colSums(MG94.Force.nonclock.filtered[10:21, ]), y = MG94.nonclock.filtered[8, ],  
 type = "p", col = col.color, bg = col.color)  
legend("topright",   
 c(paste("lnL improvement per site > ", toString(improvement.lmt), sep = ""),  
 paste("lnL improvement per site < ", toString(improvement.lmt), sep = "")),  
 lty = c(1, 1),   
 lwd = c(2.5, 2.5),  
 col = c("black", "red"))  
title("MG94 nonclock")

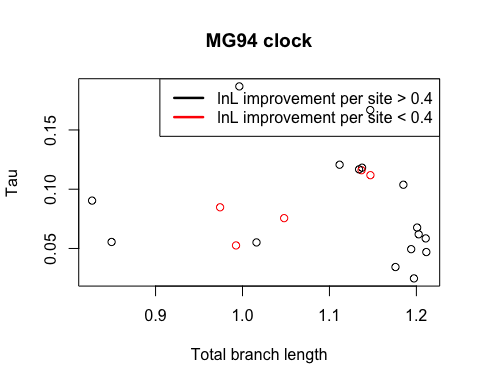


# lnL improvement per site  
((MG94.nonclock.filtered - MG94.Force.nonclock.filtered)[2,] / MG94.nonclock.filtered[1])

## YLR406C\_YDL075W YER131W\_YGL189C YDR502C\_YLR180W YHR106W\_YDR353W   
## 0.151633 0.140111 0.408336 0.198674   
## YIL057C\_YER067W YNL069C\_YIL133C YDR518W\_YCL043C YGR043C\_YLR354C   
## 0.146466 0.523279 0.301557 0.049233   
## YDR099W\_YER177W YMR143W\_YDL083C YJR048W\_YEL039C YBR191W\_YPL079W   
## 0.127276 0.332405 0.192824 0.590199   
## YDR418W\_YEL054C YPL232W\_YMR183C YLR284C\_YOR180C YBL087C\_YER117W   
## 0.286073 0.268073 0.007963 0.425100   
## YGL062W\_YBR218C YIR033W\_YKL020C YER102W\_YBL072C YAL056W\_YOR371C   
## 3.948891 0.276194 1.229483 0.130766   
## YDR438W\_YML018C   
## 0.028612

MG94 clock case

improvement.lmt <- 0.4  
plot(colSums(MG94.Force.clock.filtered[10:21, ]), MG94.clock.filtered[8, ],  
 type = "n", xlab = "Total branch length", ylab = "Tau" )  
col.color <- rep("black", dim(MG94.clock.filtered)[2])  
col.color[((MG94.clock.filtered - MG94.Force.clock.filtered)[2,] / MG94.clock.filtered[1]) < improvement.lmt] <- "red"  
points(x = colSums(MG94.Force.clock.filtered[10:21, ]), y = MG94.clock.filtered[8, ],  
 type = "p", col = col.color, bg = col.color)  
legend("topright",   
 c(paste("lnL improvement per site > ", toString(improvement.lmt), sep = ""),  
 paste("lnL improvement per site < ", toString(improvement.lmt), sep = "")),  
 lty = c(1, 1),   
 lwd = c(2.5, 2.5),  
 col = c("black", "red"))  
title("MG94 clock")



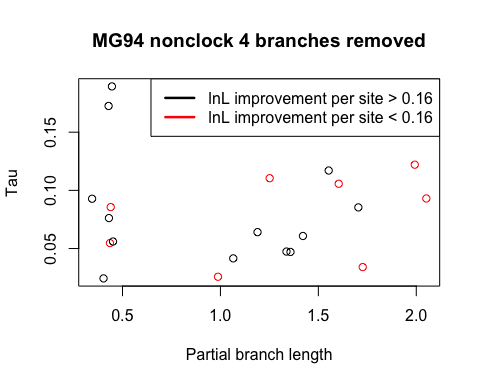
# lnL improvement per site  
((MG94.clock.filtered - MG94.Force.clock.filtered)[2,] / MG94.clock.filtered[1])

## YLR406C\_YDL075W YER131W\_YGL189C YDR502C\_YLR180W YHR106W\_YDR353W   
## 0.1556 0.1513 2.1905 4.6498   
## YIL057C\_YER067W YNL069C\_YIL133C YDR518W\_YCL043C YGR043C\_YLR354C   
## 5.1756 0.4759 0.7591 6.3887   
## YDR099W\_YER177W YMR143W\_YDL083C YJR048W\_YEL039C YBR191W\_YPL079W   
## 1.9019 0.4050 1.6068 0.6401   
## YDR418W\_YEL054C YPL232W\_YMR183C YLR284C\_YOR180C YBL087C\_YER117W   
## 0.3116 7.8226 0.3712 0.4424   
## YGL062W\_YBR218C YIR033W\_YKL020C YER102W\_YBL072C YAL056W\_YOR371C   
## 18.0984 1.3464 1.3331 1.1786   
## YDR438W\_YML018C   
## -0.8621

OK, the MG94 nonclock long branch lengths are suspecious. It seems that the first four branches carry most of the total length. What if plot without those branches?

MG94 nonclock case without first four branches: (N0, N1), (N0,kluyveri), (N1,N2), (N1,castellii) are excluded

improvement.lmt <- 0.16  
plot(colSums(MG94.Force.nonclock.filtered[14:21, ]), MG94.nonclock.filtered[8, ],  
 type = "n", xlab = "Partial branch length", ylab = "Tau" )  
col.color <- rep("black", dim(MG94.nonclock.filtered)[2])  
col.color[((MG94.nonclock.filtered - MG94.Force.nonclock.filtered)[2,] / MG94.nonclock.filtered[1]) < improvement.lmt] <- "red"  
points(x = colSums(MG94.Force.nonclock.filtered[14:21, ]), y = MG94.nonclock.filtered[8, ],  
 type = "p", col = col.color, bg = col.color)  
legend("topright",   
 c(paste("lnL improvement per site > ", toString(improvement.lmt), sep = ""),  
 paste("lnL improvement per site < ", toString(improvement.lmt), sep = "")),  
 lty = c(1, 1),   
 lwd = c(2.5, 2.5),  
 col = c("black", "red"))  
title("MG94 nonclock 4 branches removed")



# lnL improvement per site  
((MG94.nonclock.filtered - MG94.Force.nonclock.filtered)[2,] / MG94.nonclock.filtered[1])

## YLR406C\_YDL075W YER131W\_YGL189C YDR502C\_YLR180W YHR106W\_YDR353W   
## 0.151633 0.140111 0.408336 0.198674   
## YIL057C\_YER067W YNL069C\_YIL133C YDR518W\_YCL043C YGR043C\_YLR354C   
## 0.146466 0.523279 0.301557 0.049233   
## YDR099W\_YER177W YMR143W\_YDL083C YJR048W\_YEL039C YBR191W\_YPL079W   
## 0.127276 0.332405 0.192824 0.590199   
## YDR418W\_YEL054C YPL232W\_YMR183C YLR284C\_YOR180C YBL087C\_YER117W   
## 0.286073 0.268073 0.007963 0.425100   
## YGL062W\_YBR218C YIR033W\_YKL020C YER102W\_YBL072C YAL056W\_YOR371C   
## 3.948891 0.276194 1.229483 0.130766   
## YDR438W\_YML018C   
## 0.028612