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/OldResults01152
015/"
#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)

  para.list <- c(2:20)
  for (i in para.list){
    image.name <- paste(path, 'Rscripts/', target.summary, '_',
                        row.names[i], '.png', sep = '')
    png(image.name)
    plot(summary_mat[1, ], summary_mat[i, ],
         xlab = row.names[1], ylab = row.names[i],
         main = target.summary)
    dev.off()
    }
```

```
}
```

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 YDR502C_YLR180W YHR106W_YDR353W YIL057C_YER067W
##           45.21          209.16          165.03           41.56
## YPL087W_YBR183W YNL069C_YIL133C YGR043C_YLR354C YPR157W_YGR141W
##           49.82           90.20           58.85           75.49
## YDR099W_YER177W YBR024W_YBR037C YPR159W_YGR143W YGL133W_YPL216W
##          161.47           42.12          242.58           40.79
## YNL049C_YIL109C YPL232W_YMR183C YIR033W_YKL020C YMR243C_YOR316C
##          153.35          115.06           93.93           82.18
## YAL056W_YOR371C YDR438W_YML018C
##           26.78           24.13
```

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

```
## YLR406C_YDL075W YDR502C_YLR180W YHR106W_YDR353W YIL057C_YER067W
##           45.93          168.38          144.87           40.62
## YPL087W_YBR183W YNL069C_YIL133C YGR043C_YLR354C YPR157W_YGR141W
##           42.28           78.01           39.97           62.68
## YDR099W_YER177W YBR024W_YBR037C YPR159W_YGR143W YGL133W_YPL216W
##          154.92           29.14          243.60           21.22
## YNL049C_YIL109C YPL232W_YMR183C YIR033W_YKL020C YMR243C_YOR316C
##          118.15           94.26           75.01           69.15
## YAL056W_YOR371C YDR438W_YML018C
##           30.20           11.08
```

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

```
## YLR406C_YDL075W YER131W_YGL189C YDR502C_YLR180W YML026C_YDR450W
##         17.0098         15.5082         44.6676        199.1628
## YHR106W_YDR353W YIL057C_YER067W YNL069C_YIL133C YGR043C_YLR354C
##         22.1329         17.1557         59.6840          5.3063
## YDR099W_YER177W YMR143W_YDL083C YJR048W_YEL039C YBR191W_YPL079W
##         16.2328         37.2294         21.4012         63.4052
## YDR418W_YEL054C YPL232W_YMR183C YLR284C_YOR180C YBL087C_YER117W
##         31.1399         28.1146          0.7524         45.4327
## YGL062W_YBR218C YER102W_YBL072C YDR438W_YML018C
##        446.2657        138.0125          2.9400
```

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

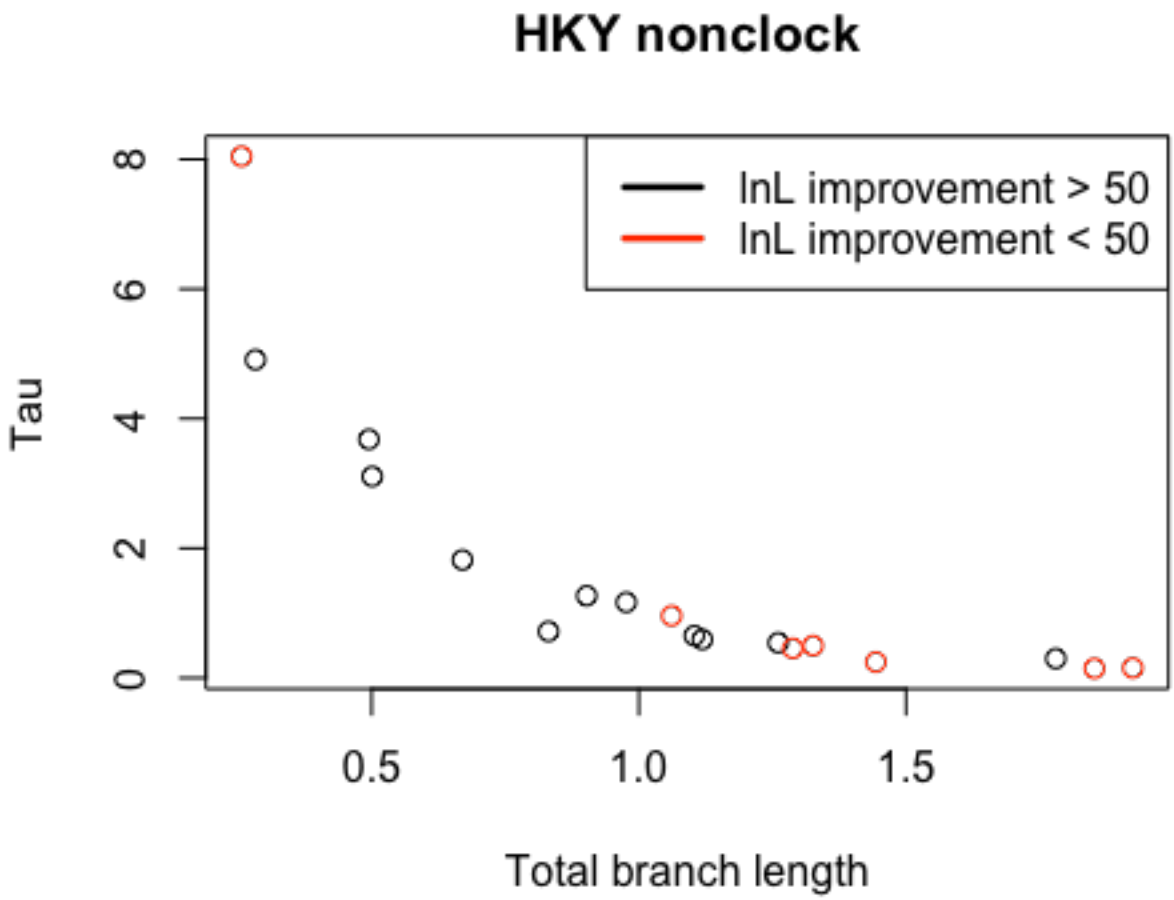
```
## YLR406C_YDL075W YER131W_YGL189C YDR502C_YLR180W YML026C_YDR450W
##           17.42           15.99          241.96          315.83
## YHR106W_YDR353W YIL057C_YER067W YNL069C_YIL133C YGR043C_YLR354C
```

```
##          521.17          592.15           54.09          722.27
## YDR099W_YER177W YMR143W_YDL083C YJR048W_YEL039C YBR191W_YPL079W
##          209.31           45.36          192.82           69.24
## YDR418W_YEL054C YPL232W_YMR183C YLR284C_YOR180C YBL087C_YER117W
##           34.20          838.06           31.09           47.32
## YGL062W_YBR218C YER102W_YBL072C YDR438W_YML018C
##         2039.77          149.04          -87.86
```

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

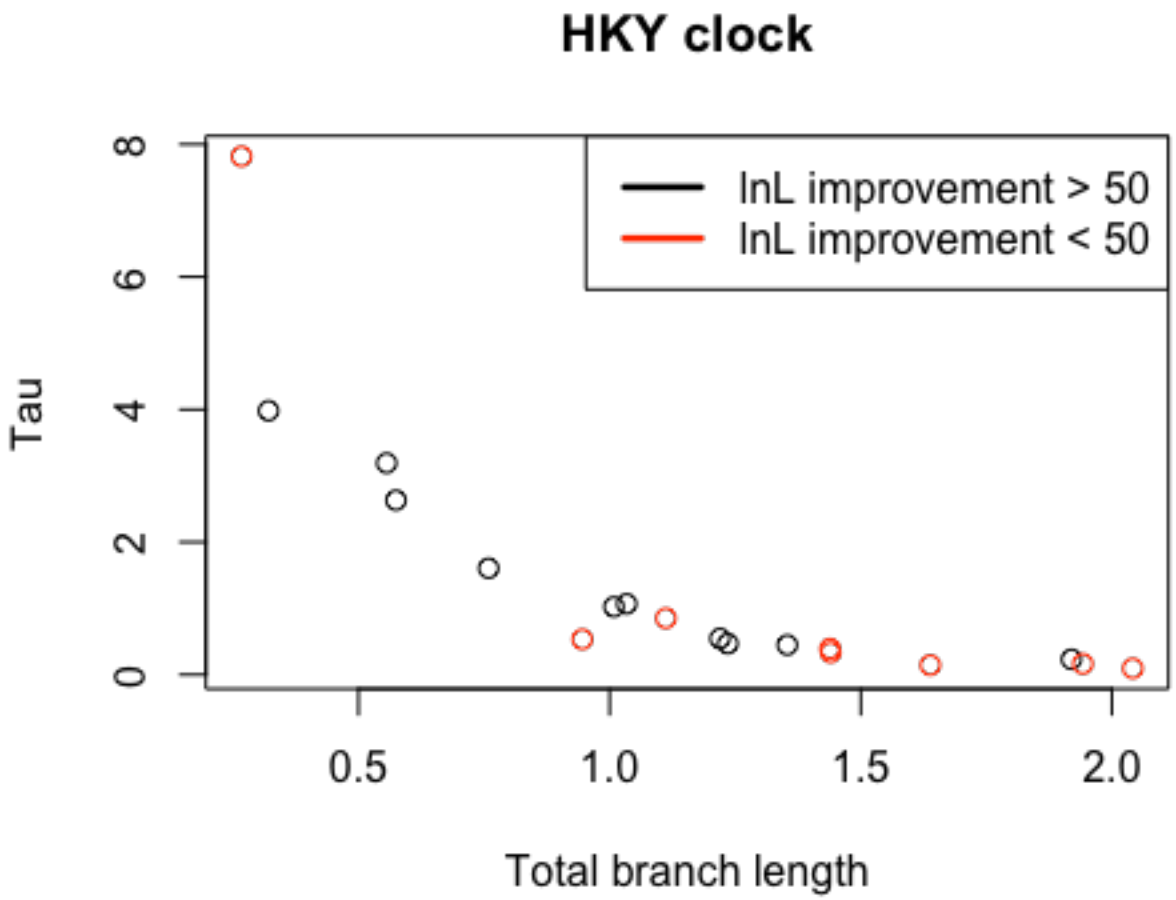
HKY nonclock case

```
plot(colSums(HKY.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.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

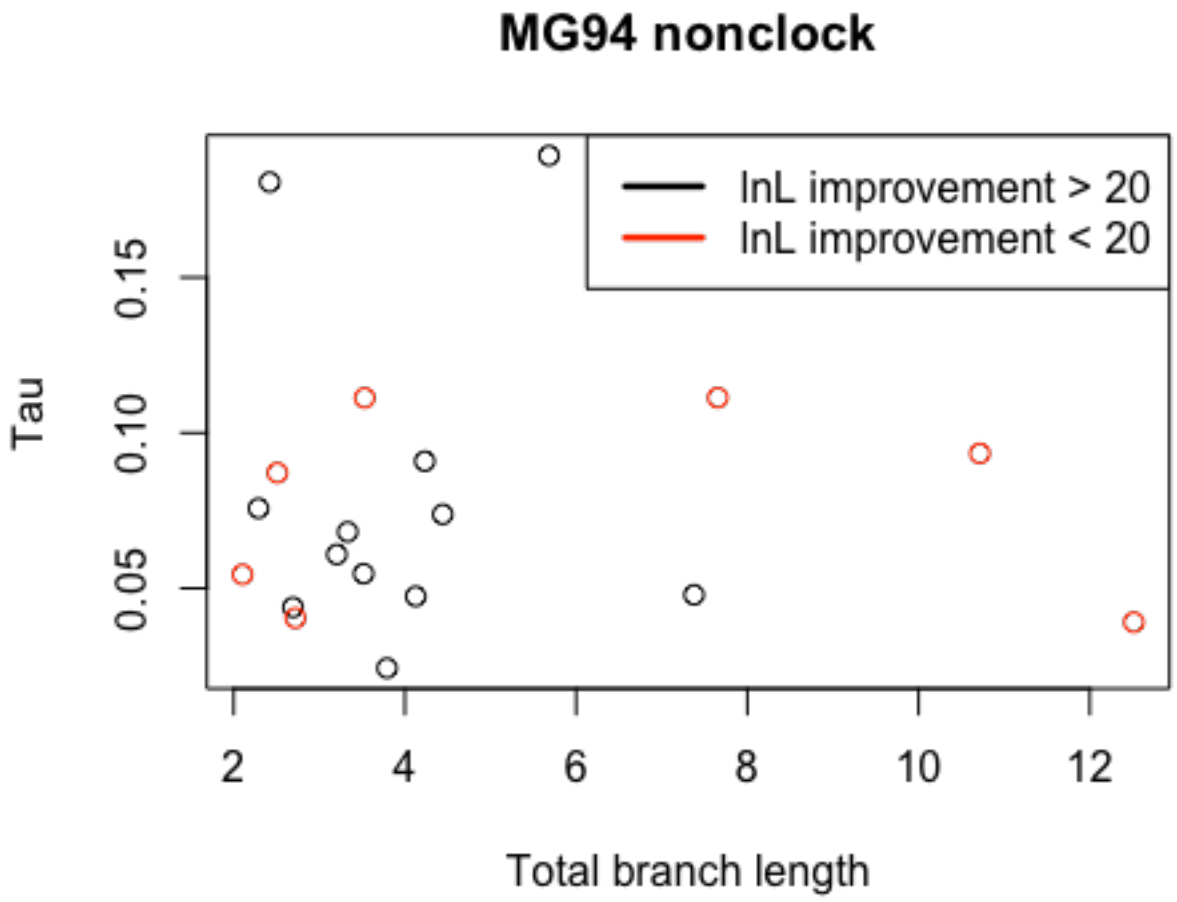
```
plot(colSums(HKY.clock.filtered[9:20, ]), HKY.clock.filtered[8, ],
     type = "n", xlab = "Total branch length", ylab = "Tau" )
col.color <- rep("black", dim(HKY.clock.filtered)[2])
col.color[(HKY.clock.filtered - HKY.Force.clock.filtered)[2,] < 50] <-
"red"
points(x = colSums(HKY.clock.filtered[9:20, ]), 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.nonclock.filtered[9:20, ]), 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.nonclock.filtered[9:20, ]), 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.clock.filtered[9:20, ]), 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.clock.filtered[9:20, ]), 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"
"YBR024W_YBR037C"
## [5] "YGL133W_YPL216W" "YAL056W_YOR371C" "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" "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" "YIL057C_YER067W" "YPL087W_YBR183W"
"YGR043C_YLR354C"
## [5] "YBR024W_YBR037C" "YGL133W_YPL216W" "YAL056W_YOR371C"
"YDR438W_YML018C"

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

## [1] "YLR406C_YDL075W" "YER131W_YGL189C" "YDR438W_YML018C"
```

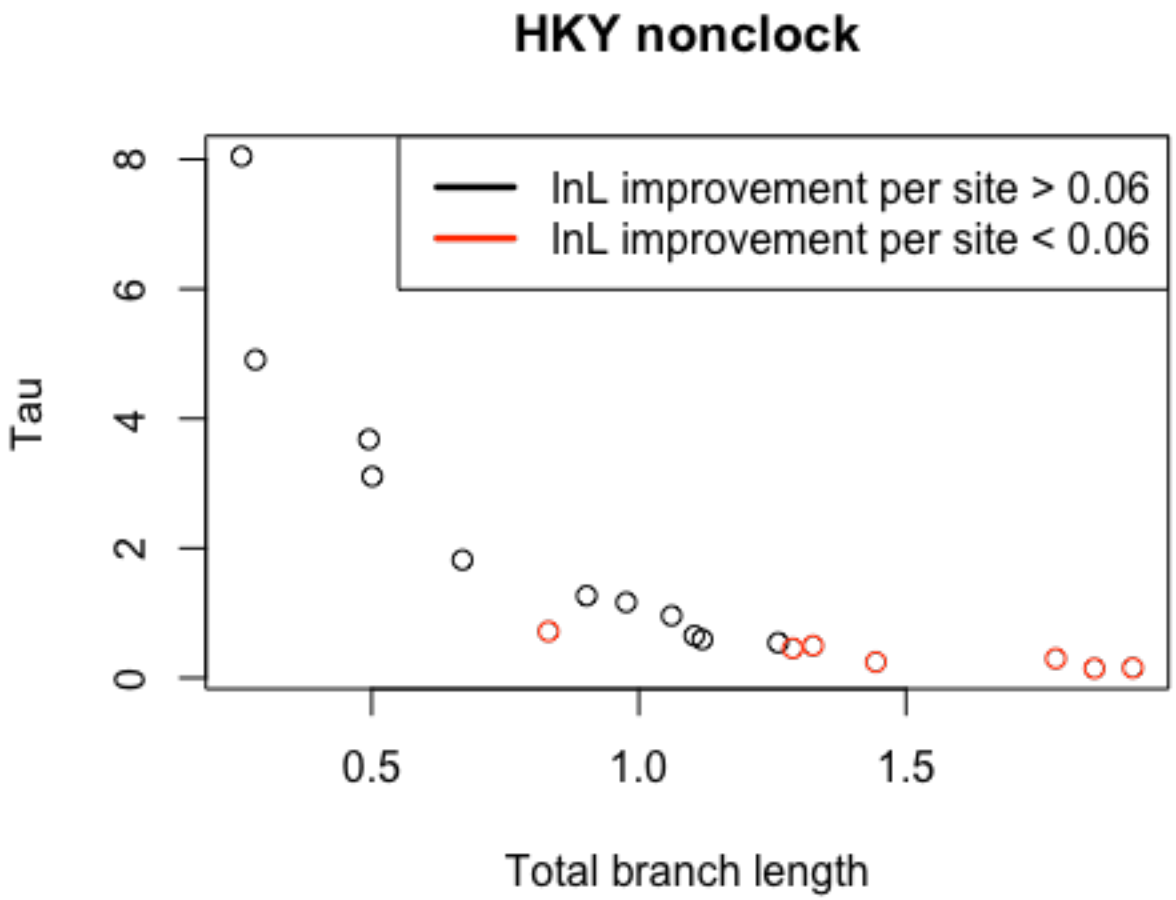
2 pairs show up in both: YLR406C_YDL075W, YDR438W_YML018C

==========

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

HKY nonclock case

```
plot(colSums(HKY.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.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")
```



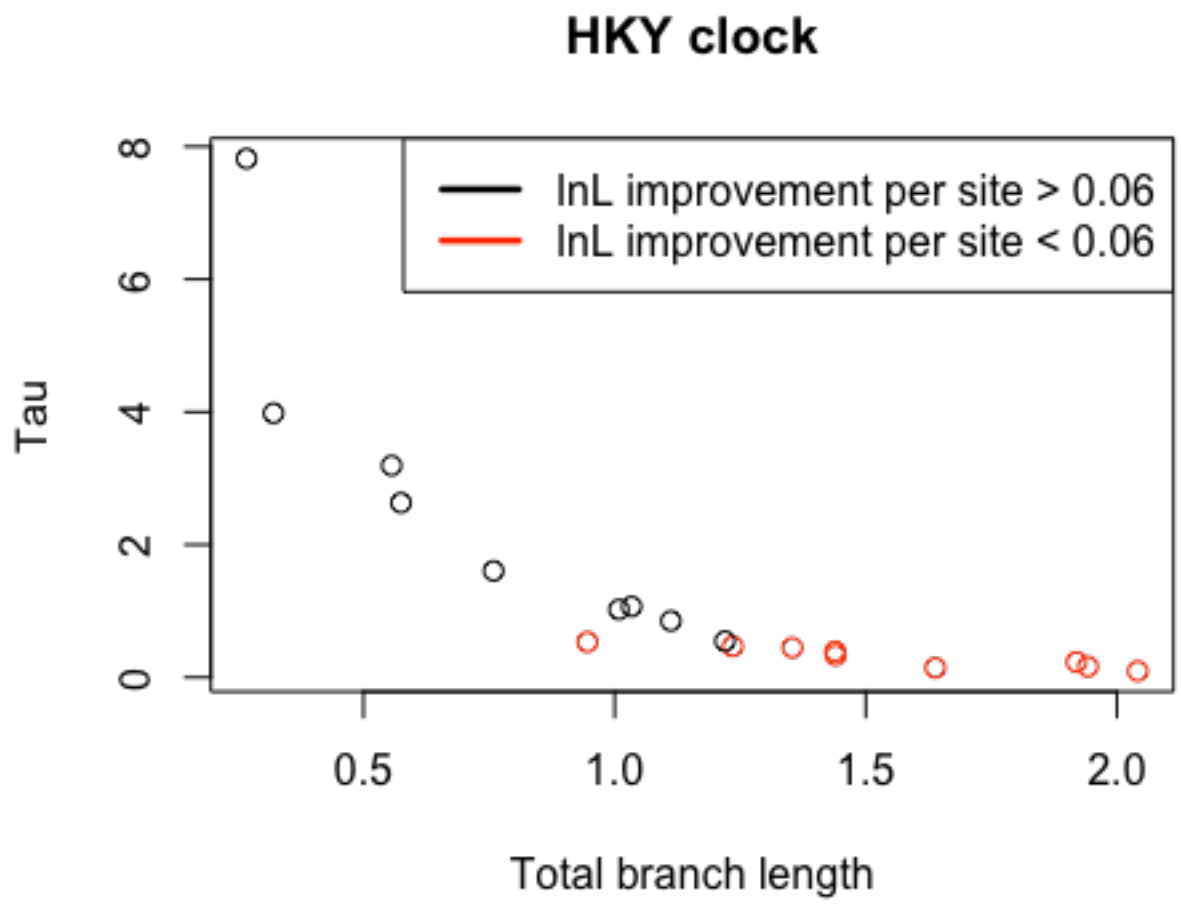

```
# lnL improvement per site
((HKY.nonclock.filtered - HKY.Force.nonclock.filtered)[2,] /
HKY.nonclock.filtered[1, ])
```

```
## YLR406C_YDL075W YDR502C_YLR180W YHR106W_YDR353W YIL057C_YER067W
##         0.13337         0.18252         0.17245         0.08823
## YPL087W_YBR183W YNL069C_YIL133C YGR043C_YLR354C YPR157W_YGR141W
##         0.05409         0.15185         0.05909         0.06093
## YDR099W_YER177W YBR024W_YBR037C YPR159W_YGR143W YGL133W_YPL216W
##         0.21108         0.05014         0.12123         0.01368
## YNL049C_YIL109C YPL232W_YMR183C YIR033W_YKL020C YMR243C_YOR316C
##         0.06107         0.13457         0.03502         0.07266
## YAL056W_YOR371C YDR438W_YML018C
##         0.01355         0.02331
```

HKY clock case

```
plot(colSums(HKY.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.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")
```



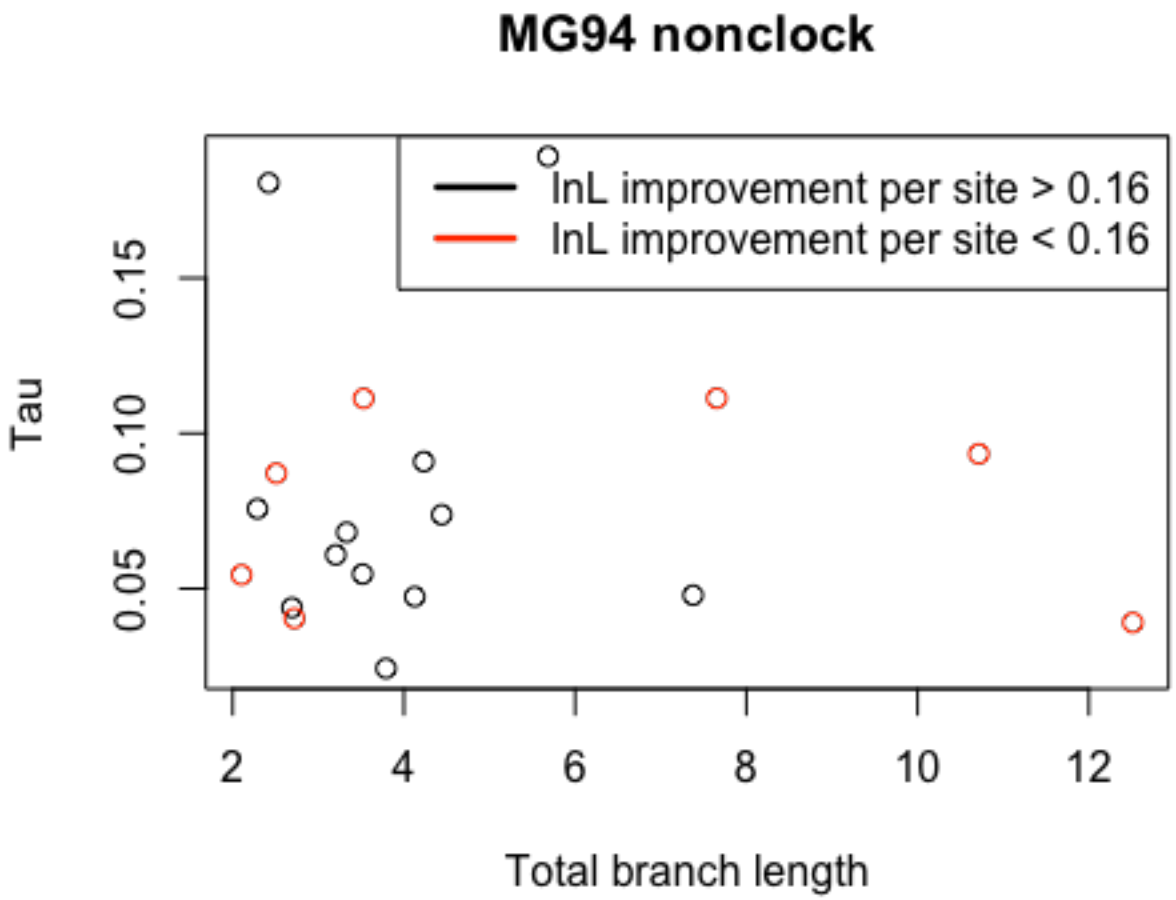

```
# lnL improvement per site
((HKY.clock.filtered - HKY.Force.clock.filtered)[2,] /
HKY.clock.filtered[1, ])
```

```
## YLR406C_YDL075W YDR502C_YLR180W YHR106W_YDR353W YIL057C_YER067W
##        0.135487        0.146931        0.151376        0.086245
## YPL087W_YBR183W YNL069C_YIL133C YGR043C_YLR354C YPR157W_YGR141W
##        0.045908        0.131323        0.040133        0.050586
```

```
## YDR099W_YER177W YBR024W_YBR037C YPR159W_YGR143W YGL133W_YPL216W
##        0.202513        0.034693        0.121737        0.007116
## YNL049C_YIL109C YPL232W_YMR183C YIR033W_YKL020C YMR243C_YOR316C
##        0.047052        0.110245        0.027969        0.061142
## YAL056W_YOR371C YDR438W_YML018C
##        0.015277        0.010706
```

MG94 nonclock case

```
improvement.lmt <- 0.16
plot(colSums(MG94.nonclock.filtered[9:20, ]), 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.nonclock.filtered[9:20, ]), 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")
```



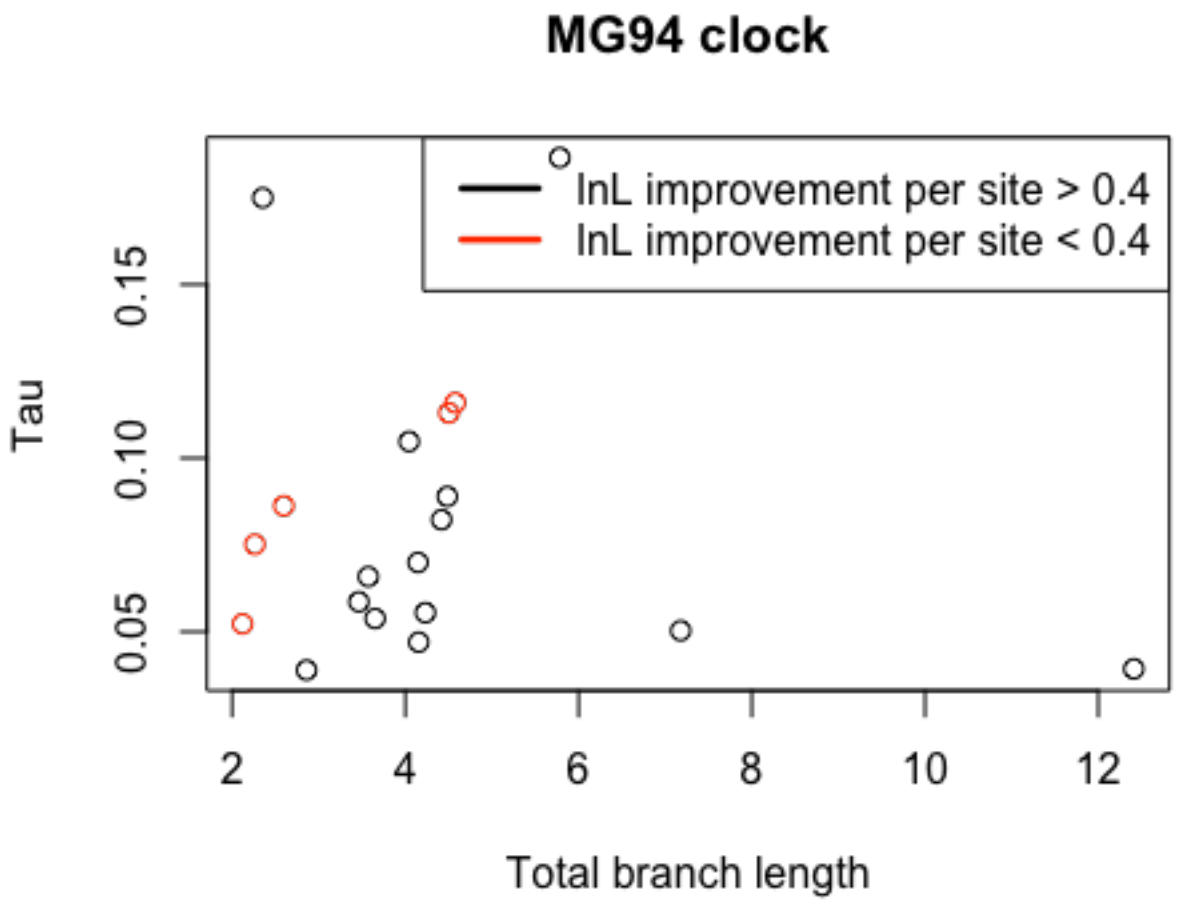

```
# lnL improvement per site
((MG94.nonclock.filtered - MG94.Force.nonclock.filtered)[2,] /
MG94.nonclock.filtered[1])
```

```
## YLR406C_YDL075W YER131W_YGL189C YDR502C_YLR180W YML026C_YDR450W
##        0.150529        0.137241        0.395289        1.762502
## YHR106W_YDR353W YIL057C_YER067W YNL069C_YIL133C YGR043C_YLR354C
##        0.195866        0.151820        0.528177        0.046959
## YDR099W_YER177W YMR143W_YDL083C YJR048W_YEL039C YBR191W_YPL079W
##        0.143653        0.329463        0.189391        0.561108
## YDR418W_YEL054C YPL232W_YMR183C YLR284C_YOR180C YBL087C_YER117W
##        0.275574        0.248802        0.006659        0.402059
## YGL062W_YBR218C YER102W_YBL072C YDR438W_YML018C
##        3.949254        1.221350        0.026018
```

MG94 clock case

```
improvement.lmt <- 0.4
plot(colSums(MG94.clock.filtered[9:20, ]), 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.clock.filtered[9:20, ]), 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 YML026C_YDR450W
##          0.1542          0.1415          2.1412          2.7950
## YHR106W_YDR353W YIL057C_YER067W YNL069C_YIL133C YGR043C_YLR354C
##          4.6121          5.2403          0.4787          6.3917
## YDR099W_YER177W YMR143W_YDL083C YJR048W_YEL039C YBR191W_YPL079W
##          1.8523          0.4014          1.7063          0.6128
## YDR418W_YEL054C YPL232W_YMR183C YLR284C_YOR180C YBL087C_YER117W
##          0.3027          7.4165          0.2752          0.4188
## YGL062W_YBR218C YER102W_YBL072C YDR438W_YML018C
##         18.0510          1.3189         -0.7776
```