HMM Tract Summary

Xiang Ji

April 6, 2017

Now plot posterior log likelihood ratio: $ln(\frac{Pr(S_i=1|x)}{Pr(S_i=0|x)}$.

The derivatives are $\frac{\partial \ln L}{\partial \ln p_{tract}}$ for the first order and $\frac{\partial^2 \ln L}{\partial \ln {p_{tract}}^2}$ for second order.

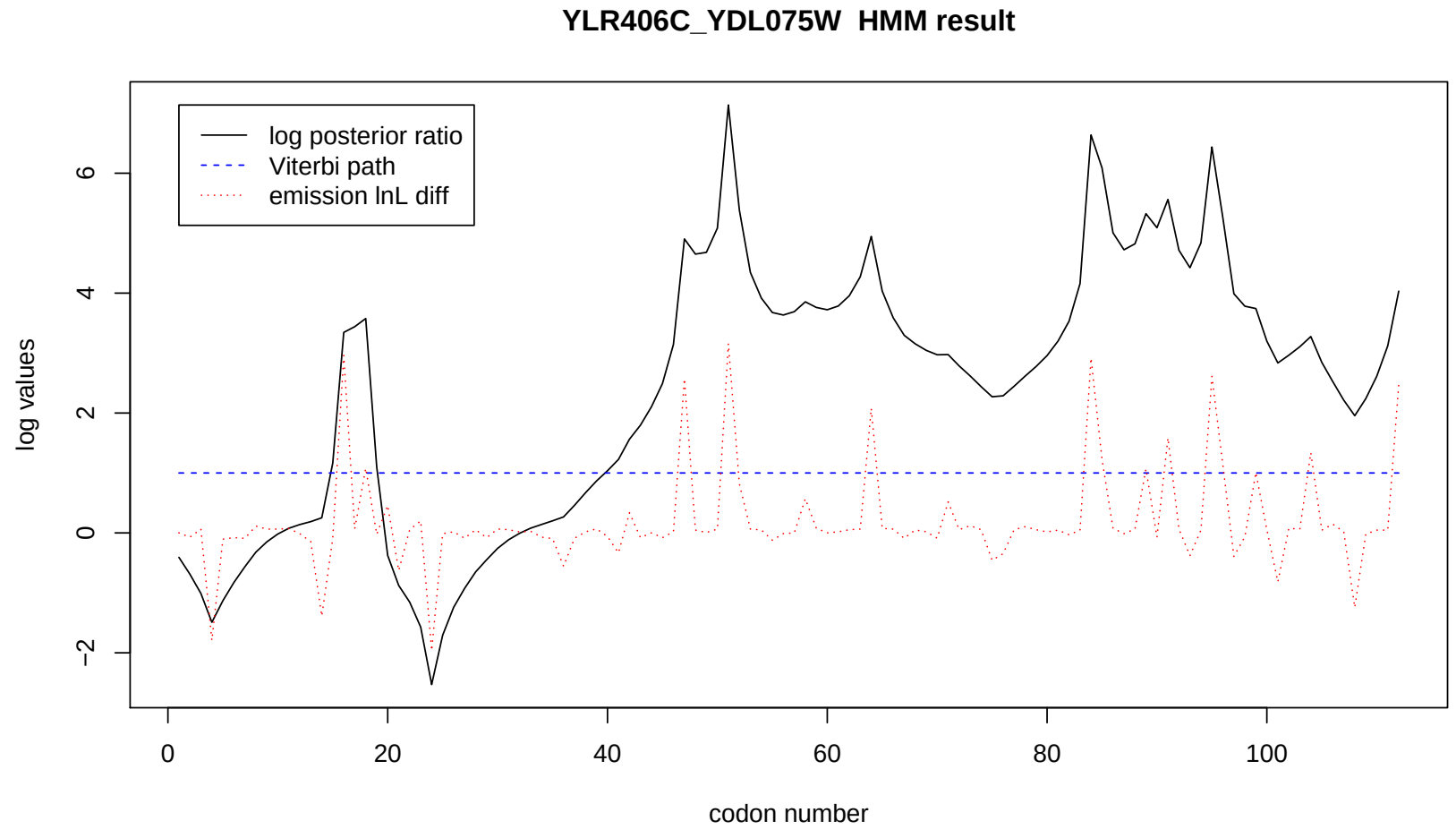
The variance is calculated by: $Var(ln(p_{tract})) = \frac{1}{I(ln(p_{tract}))} \approx -\frac{1}{\frac{\partial^2 \ln L}{\partial \ln {p_{tract}}^2}}$

95% confidence interval for $ln(p_{tract})$ is $ln(p_{tract}) \pm 1.96 * \sqrt{Var(\ln(p_{tract}))}$

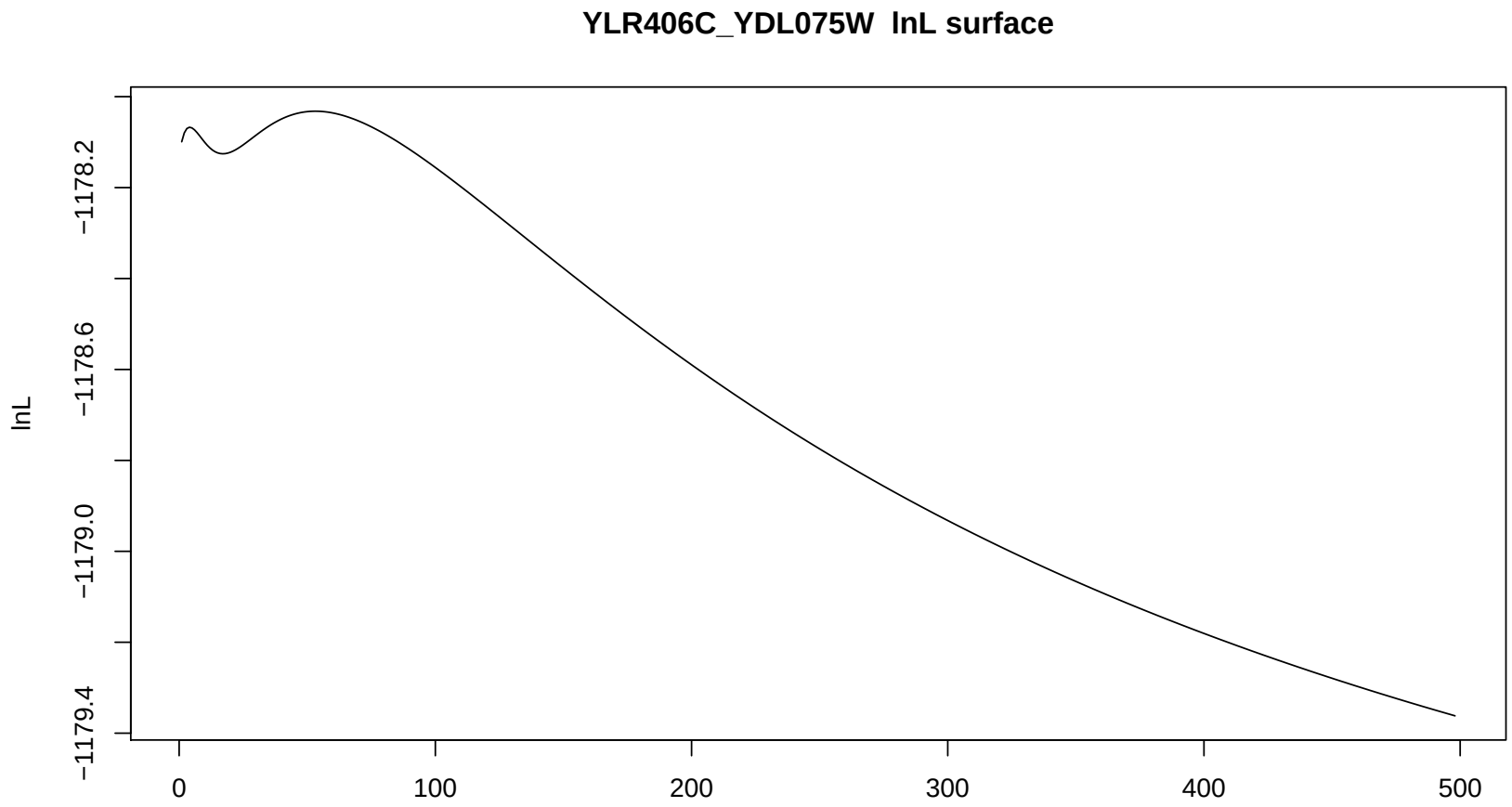
By transforming to $3.0/p_{tract}$ to get the average tract length in nucleotide.

```
rm(list=ls())  # clean up workspace
#setwd("/Users/xji3/GitFolders/YeastIGCTract/HMMTract/")
setwd("/Users/Xiang/GitFolders/YeastIGCTract/HMMTract/")
filtered.pairs <- readLines('../Filtered_pairs.txt')

summary.mat <- read.table("./HMM_tract_MG94_nonclock_summary.txt")
rownames(summary.mat) <- filtered.pairs
# Now calculate standard deviation of lnP
lnP <- log(3.0 / summary.mat[, 3])
sd.lnP <- 1.0 / sqrt(-summary.mat[, 7])
low.cf <- exp(lnP - 1.96 * sd.lnP)
up.cf  <- exp(lnP + 1.96 * sd.lnP)
up.cf[up.cf > 1] <- 1.0
summary.mat <- cbind(summary.mat, 3.0 / up.cf, 3.0 / low.cf)

colnames(summary.mat) <- c("lnL", "max lnL", "tract length",
                           "Pr(S_0)", "Pr(S_1)",
                           "df", "d^2f", "c.i. tract length", "c.i tract length")
par(mfrow=c(1, 1))
for (paralog in filtered.pairs){
  lnL.ratio <- as.vector(read.table(paste("./summary/", paralog, "_MG94_nonclock_HMM_log_posterior_ratio.txt", sep = "")))
  Viterbi.path <- as.vector(read.table(paste("./summary/", paralog, "_MG94_nonclock_HMM_Viterbi_path.txt", sep = "")))
  lnL.surface <- as.vector(read.table(paste("./summary/", paralog, "_MG94_nonclock_HMM_lnL_surface.txt", sep = "")))
  IGC.sw.lnL <- as.vector(read.table(paste("./summary/", paralog, "_MG94_nonclock_sw_lnL.txt", sep = "")))
  Force.sw.lnL <- as.vector(read.table(paste("./summary/Force_", paralog, "_MG94_nonclock_sw_lnL.txt", sep = "")))

  plot(lnL.ratio[, 1], xlab = "codon number", ylab = "log values",
       type = "l", col = "black", lty = 1,
       main = paste(paralog, " HMM result"),
       ylim = c(min(-0.5, min(lnL.ratio)), max(lnL.ratio)))
```

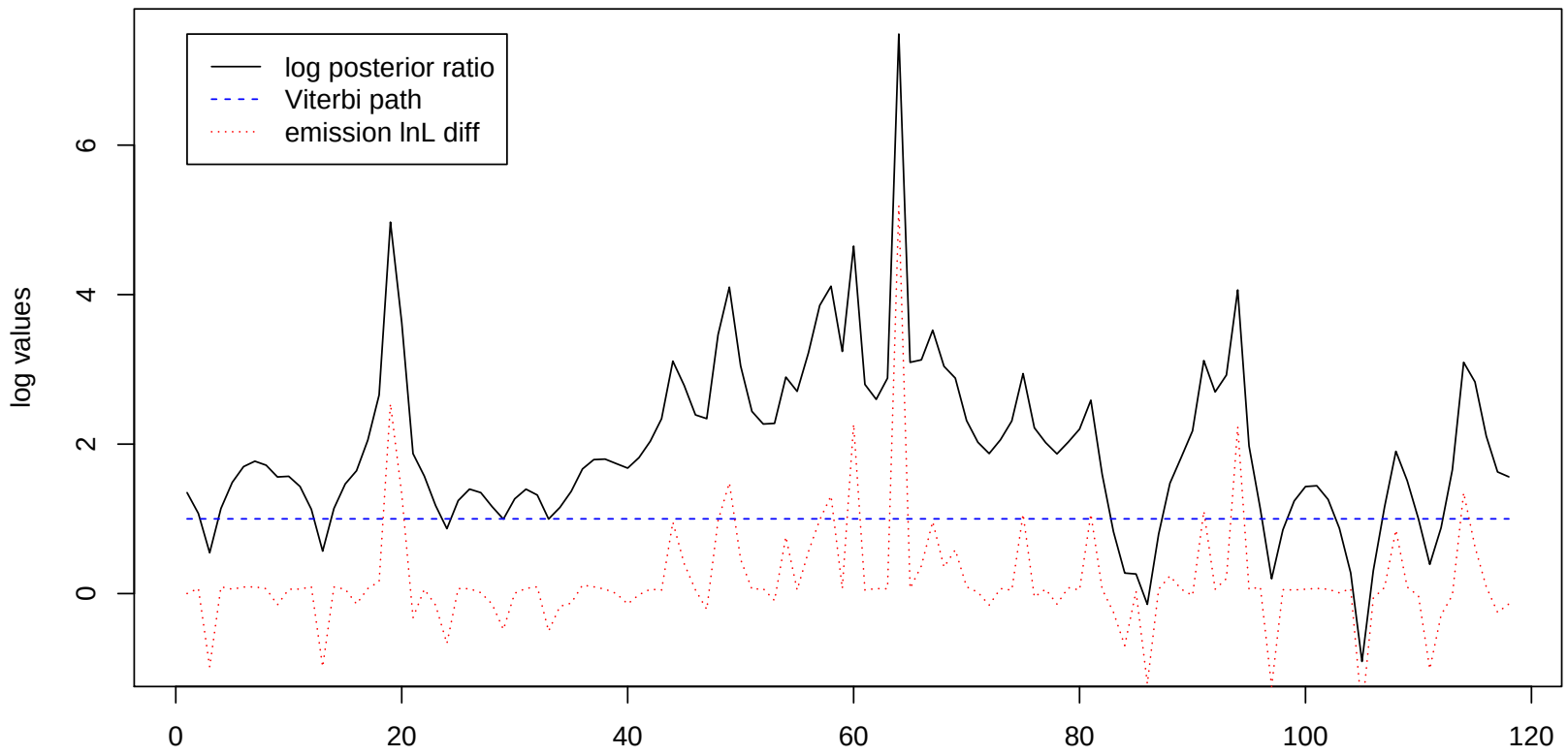
```
  lines(1:dim(Viterbi.path)[1], Viterbi.path[, 1], type = "S", lty = 2, col = "blue")
  lines(1:dim(IGC.sw.lnL)[1], IGC.sw.lnL[, 2] - Force.sw.lnL[, 2], type = "l", lty = 3, col = "red")
  legend(1, max(lnL.ratio), legend = c("log posterior ratio", "Viterbi path", "emission lnL diff"),
         lty = c(1, 2, 3), col = c("black", "blue", "red"))

  print(summary.mat[paralog, ])
  plot(-lnL.surface[, 1], xlab = "tract length in nucleotide", ylab= "lnL", type = "l", col = "black", lty = 1,
       main = paste(paralog, " lnL surface"))
  print(summary.mat[paralog, ])

}
```



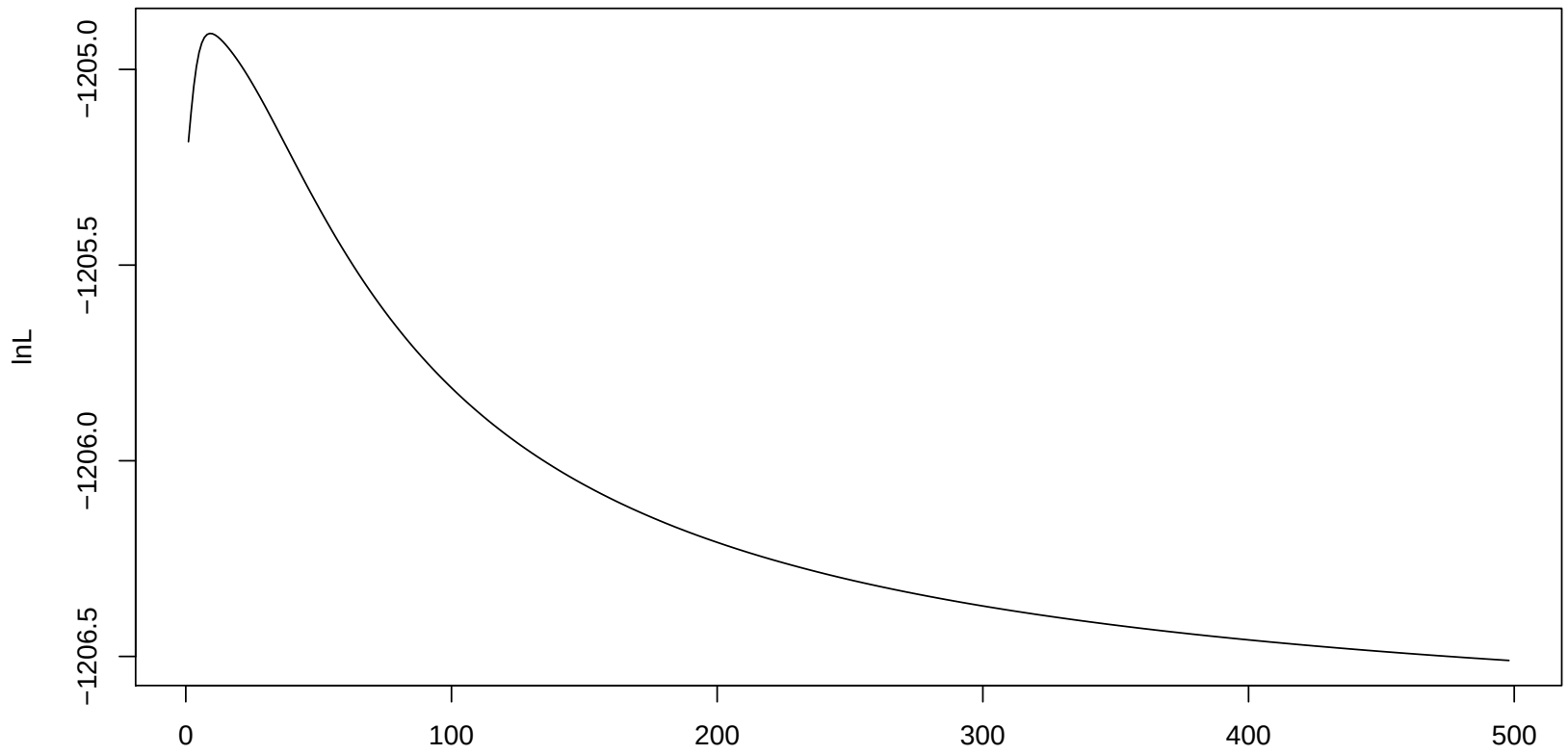

```
##                         lnL   max lnL tract length   Pr(S_0)   Pr(S_1)
## YLR406C_YDL075W -1178.099 -1178.032     55.39456 0.1281115 0.8718885
##                            df       d^2f c.i. tract length c.i tract length
## YLR406C_YDL075W 9.317345e-06 -0.5431516          3.876795         791.5192
```



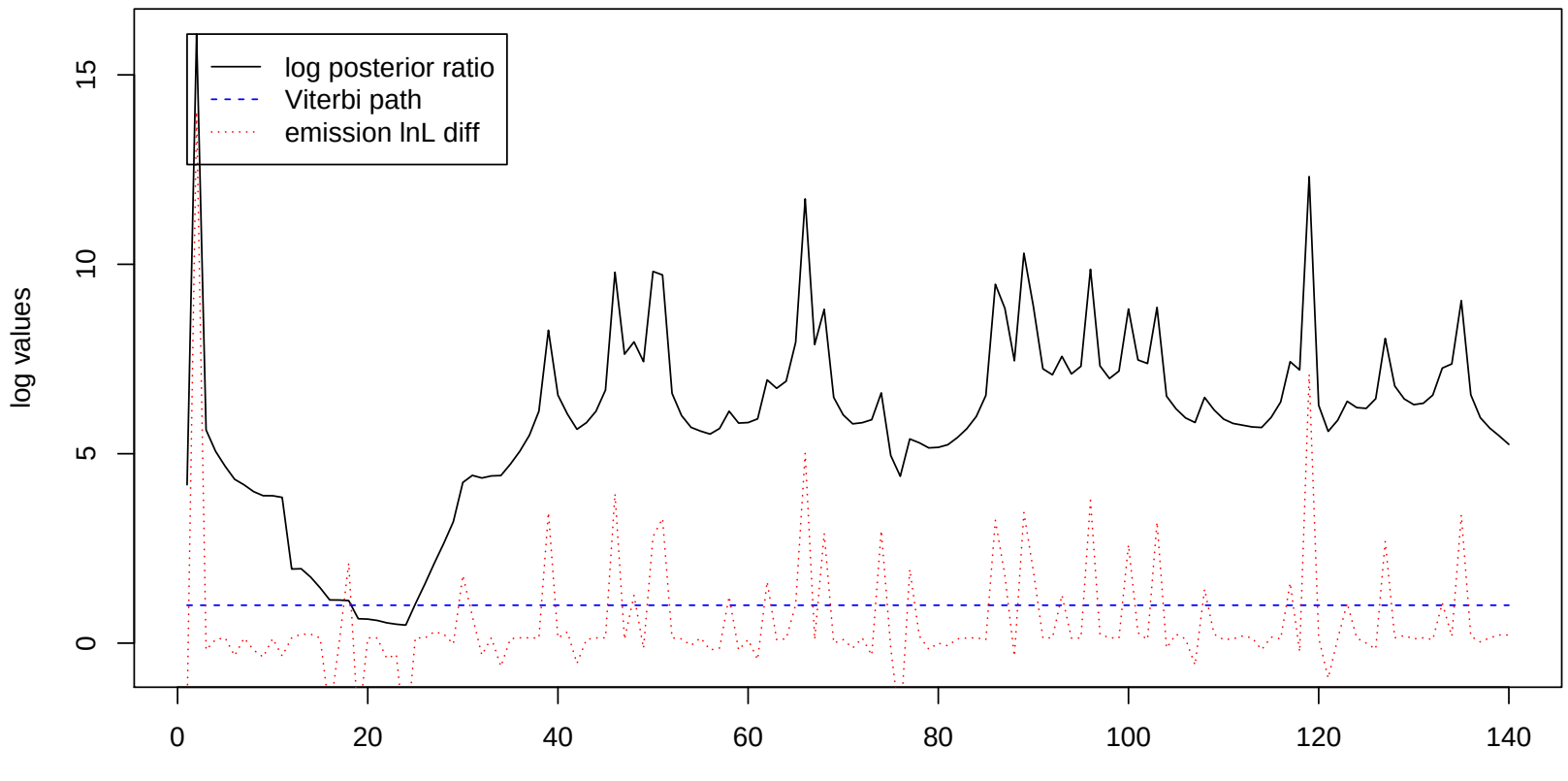

```
##                         lnL   max lnL tract length   Pr(S_0)   Pr(S_1)
## YLR406C_YDL075W -1178.099 -1178.032     55.39456 0.1281115 0.8718885
##                            df       d^2f c.i. tract length c.i tract length
## YLR406C_YDL075W 9.317345e-06 -0.5431516          3.876795         791.5192
```



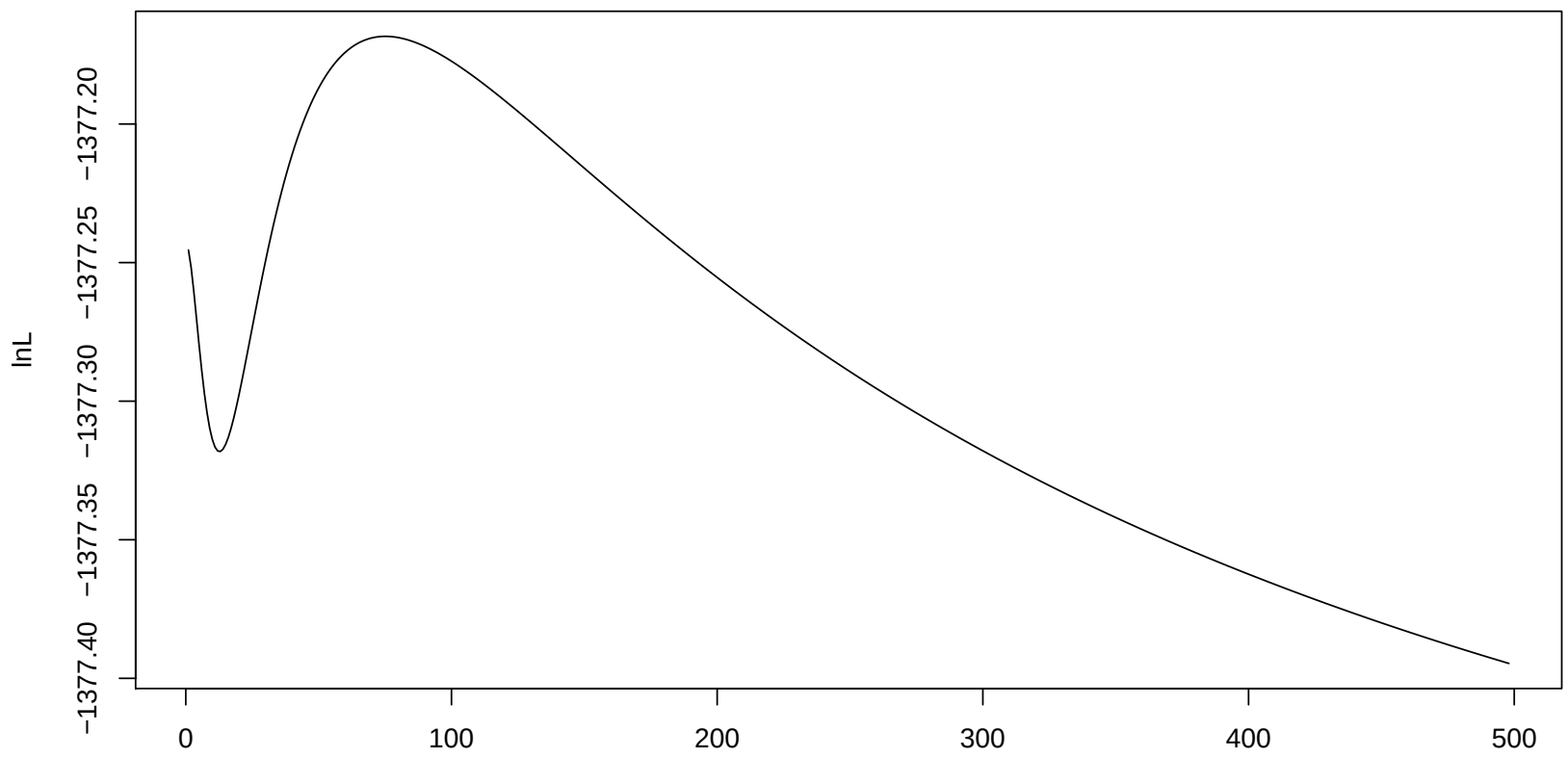

```
##                         lnL   max lnL tract length   Pr(S_0)   Pr(S_1)
## YER131W_YGL189C -1205.185 -1204.908     11.28837 0.1495886 0.8504114
##                             df       d^2f c.i. tract length c.i tract length
## YER131W_YGL189C 0.0002496522 -0.3883131                 3         262.1966
```



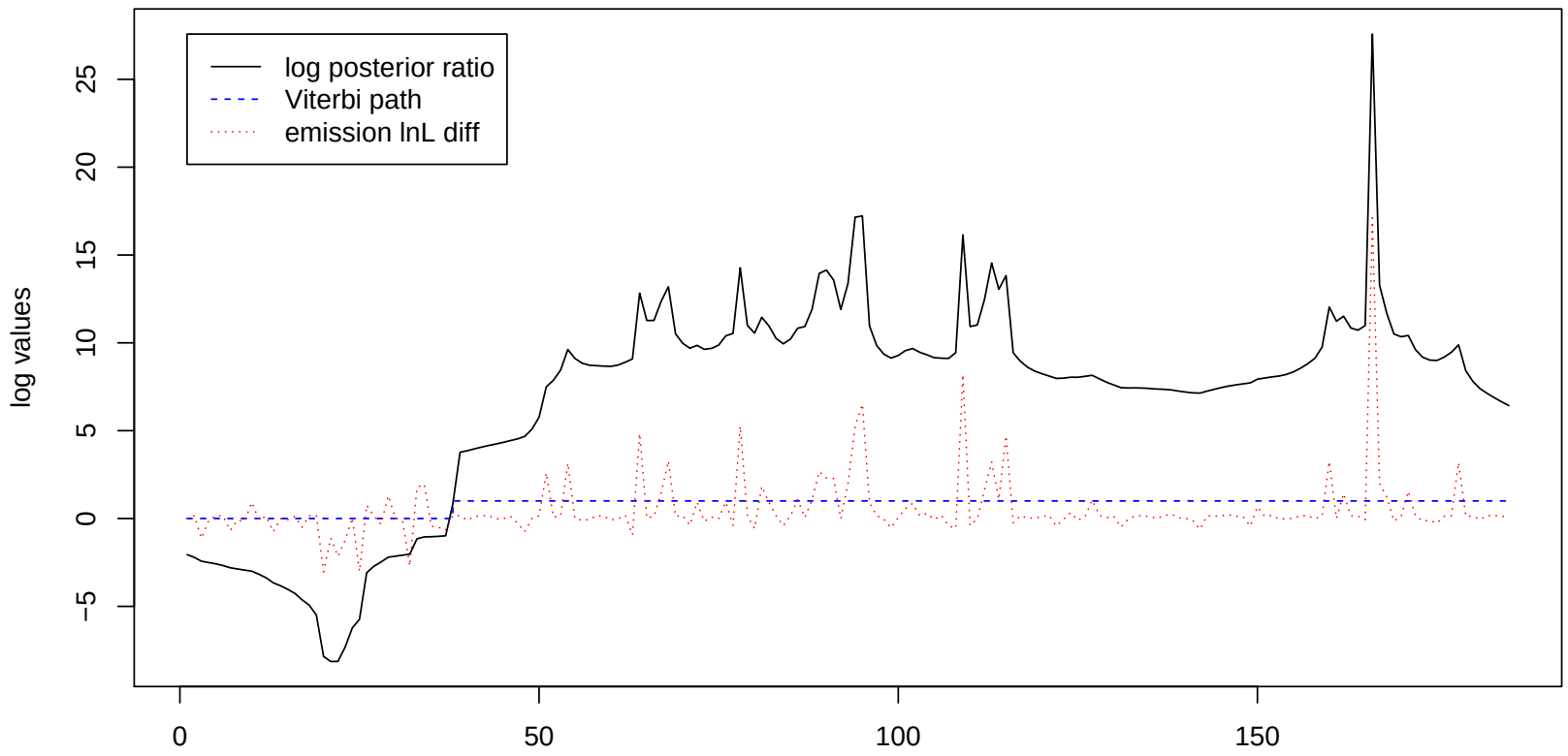

```
##                         lnL   max lnL tract length   Pr(S_0)   Pr(S_1)
## YER131W_YGL189C -1205.185 -1204.908     11.28837 0.1495886 0.8504114
##                             df       d^2f c.i. tract length c.i tract length
## YER131W_YGL189C 0.0002496522 -0.3883131                 3         262.1966
```



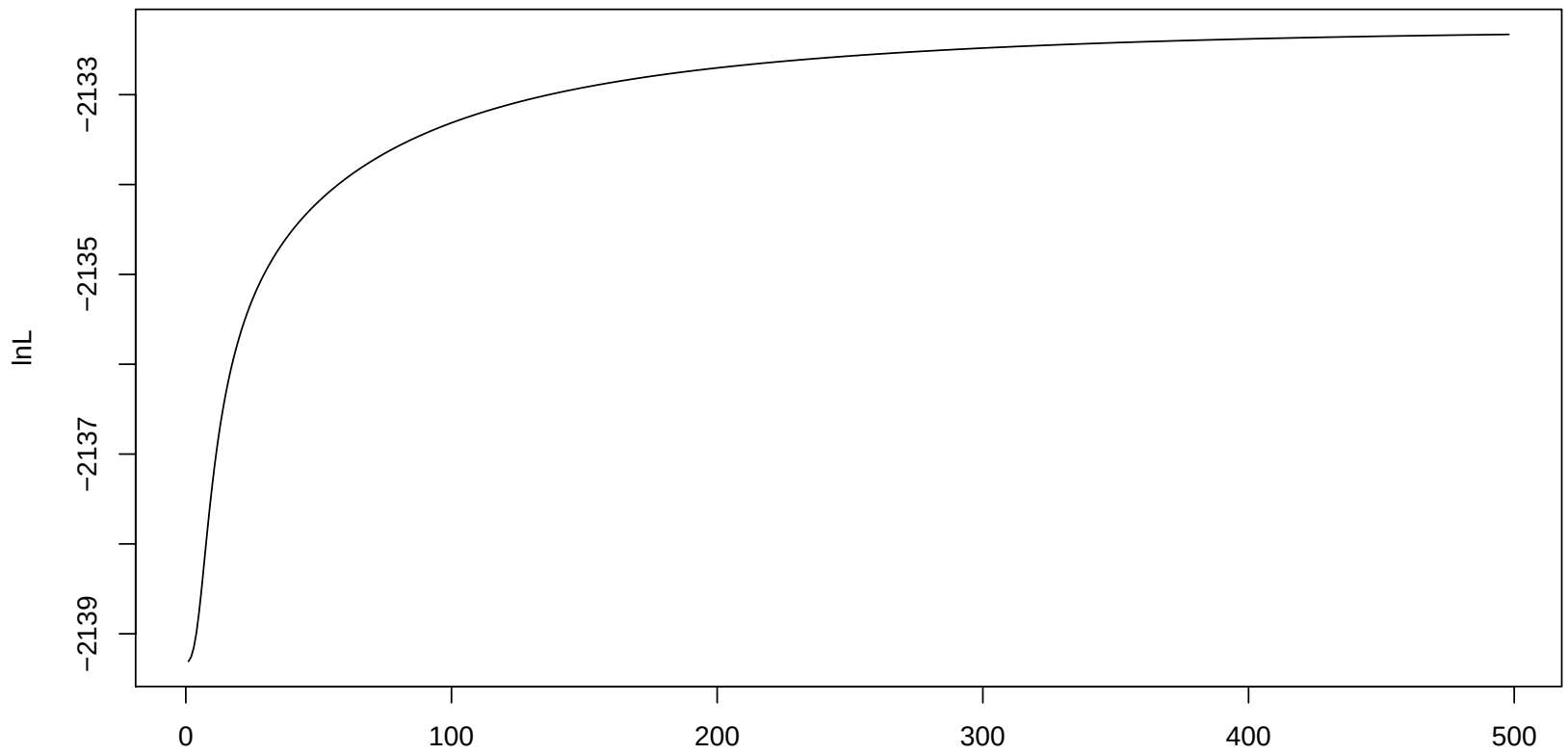

```
##                         lnL   max lnL tract length    Pr(S_0)   Pr(S_1)
## YML026C_YDR450W -1377.245 -1377.168     77.42715 0.01461544 0.9853846
##                             df       d^2f c.i. tract length
## YML026C_YDR450W -8.910021e-05 -0.2398341                 3
##                  c.i tract length
## YML026C_YDR450W          4236.76
```



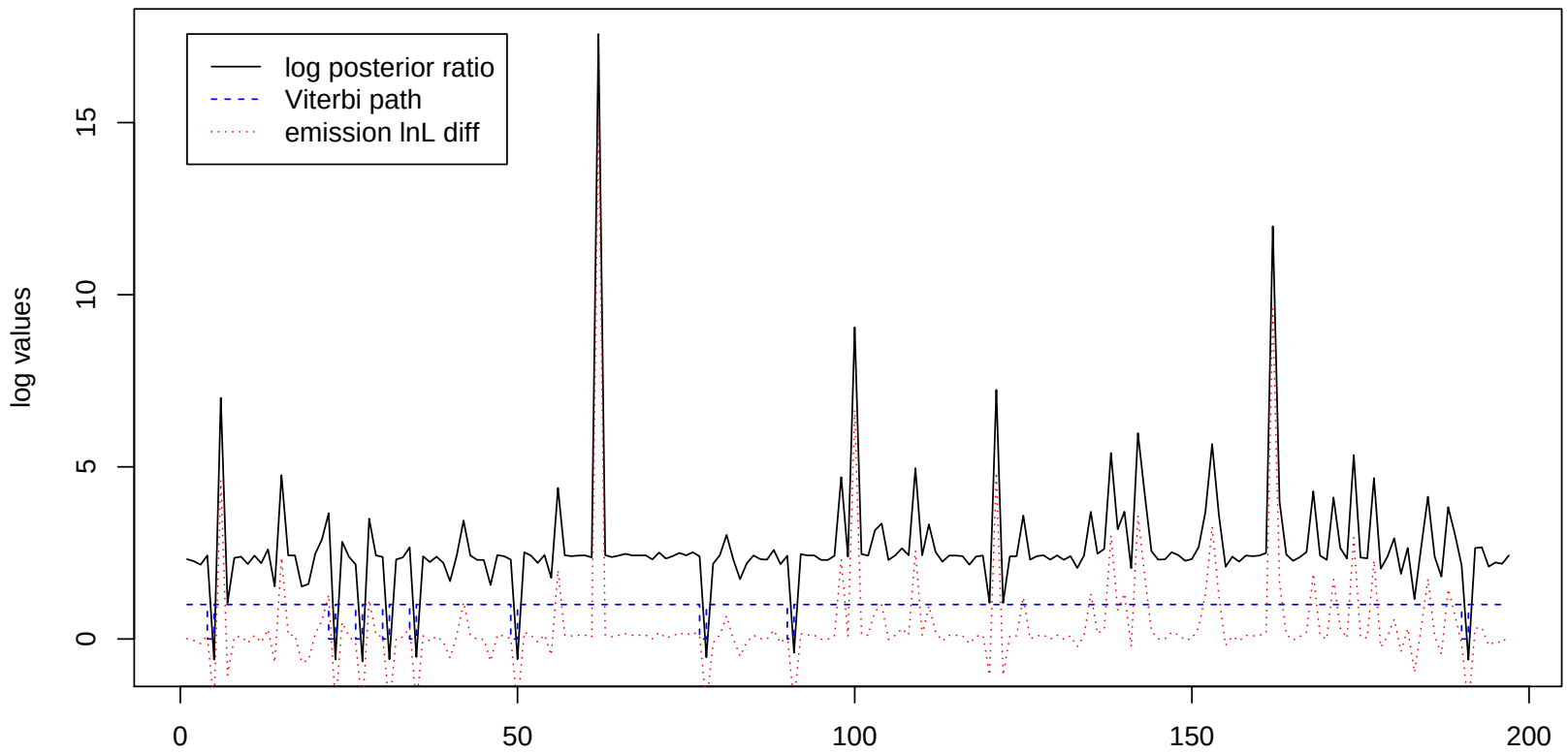

```
##                         lnL   max lnL tract length    Pr(S_0)   Pr(S_1)
## YML026C_YDR450W -1377.245 -1377.168     77.42715 0.01461544 0.9853846
##                             df       d^2f c.i. tract length
## YML026C_YDR450W -8.910021e-05 -0.2398341                 3
##                  c.i tract length
## YML026C_YDR450W          4236.76
```



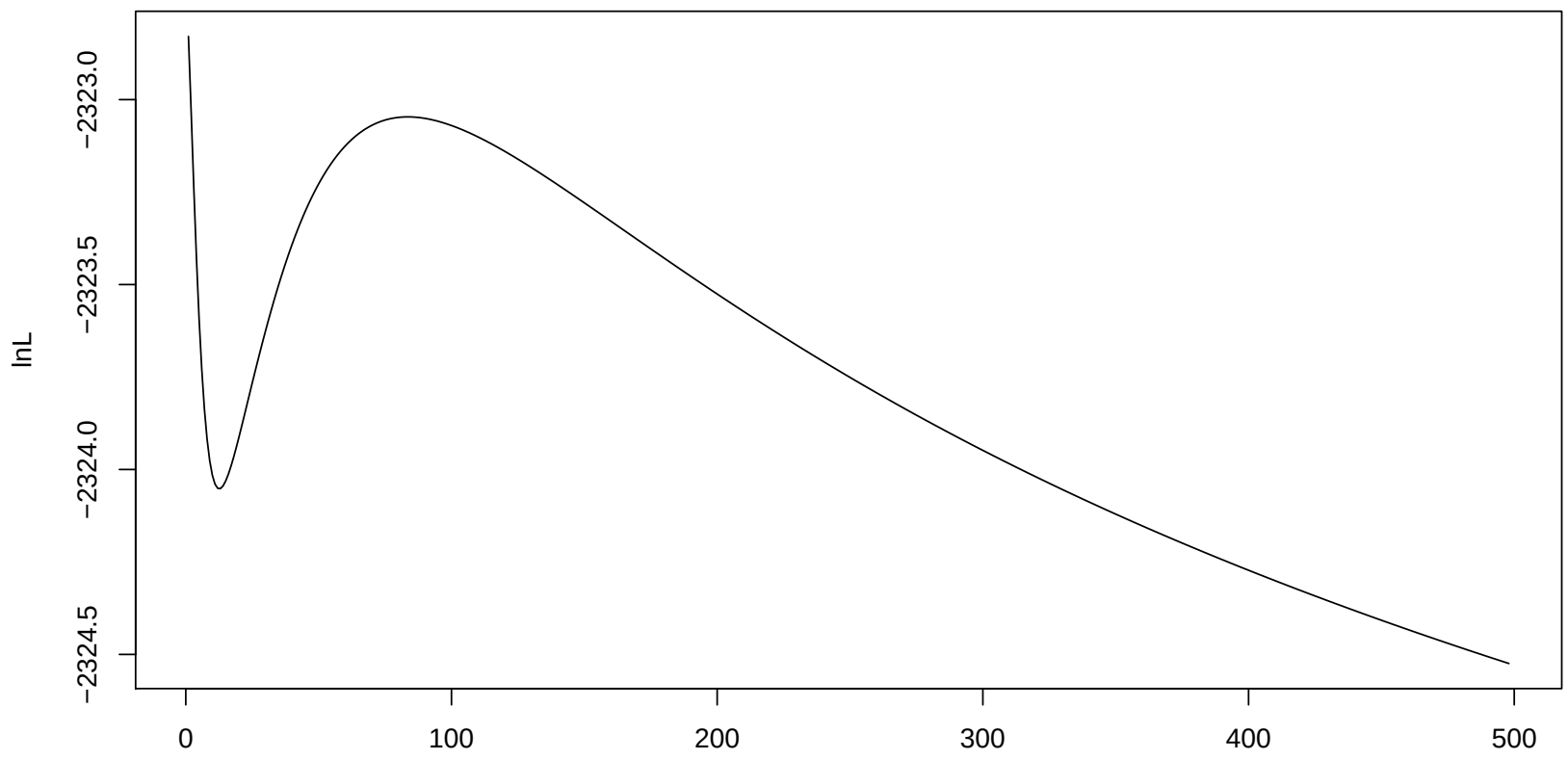

```
##                        lnL   max lnL tract length    Pr(S_0)   Pr(S_1)
## YNL301C_YOL120C -2139.308 -2132.298     715.6368 0.01965931 0.9803407
##                             df       d^2f c.i. tract length
## YNL301C_YOL120C -0.0008087931 -0.4923336            43.808
##                  c.i tract length
## YNL301C_YOL120C         11690.47
```



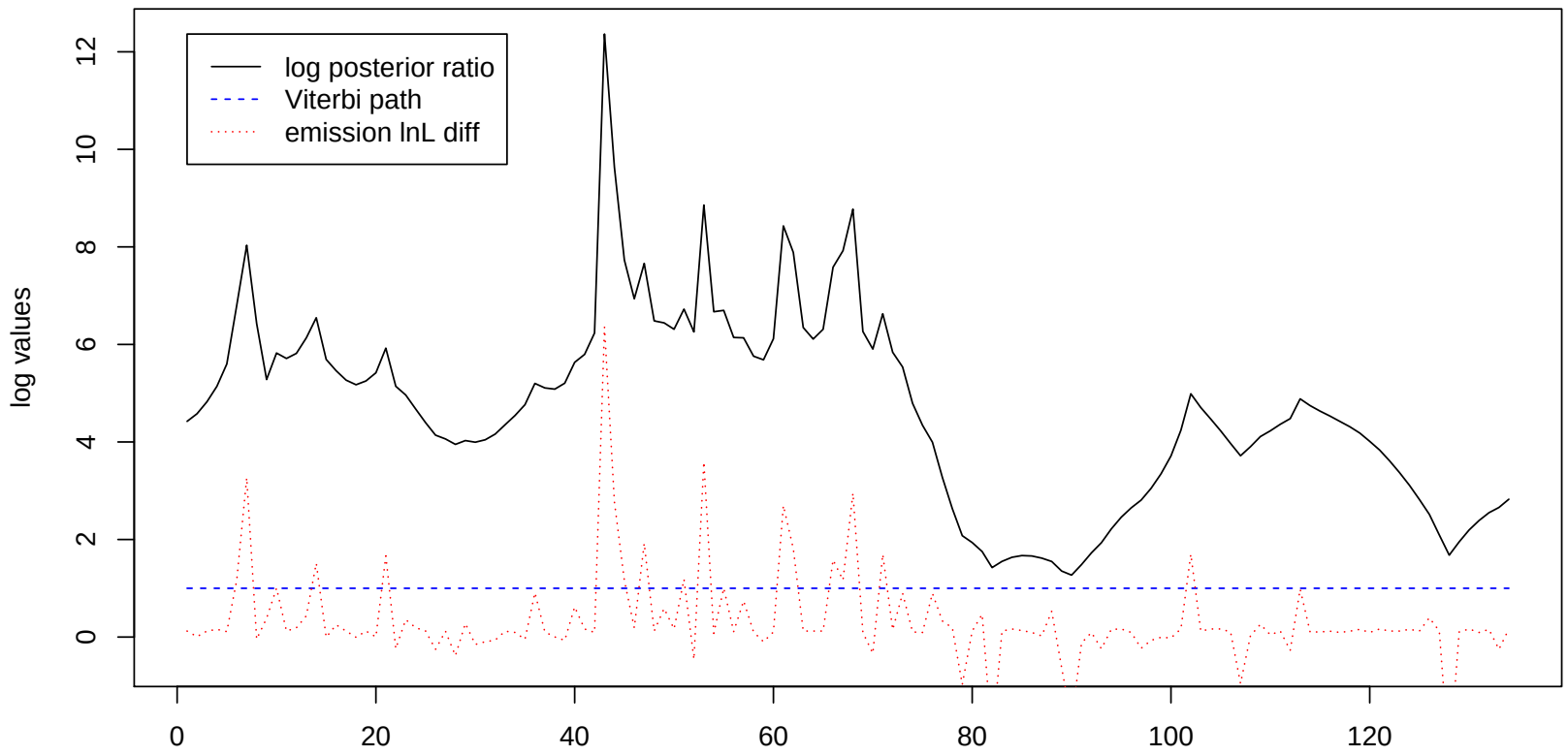

```
##                       lnL   max lnL tract length    Pr(S_0)   Pr(S_1)
## YNL301C_YOL120C -2139.308 -2132.298     715.6368 0.01965931 0.9803407
##                           df       d^2f c.i. tract length
## YNL301C_YOL120C -0.0008087931 -0.4923336            43.808
##                 c.i tract length
## YNL301C_YOL120C         11690.47
```



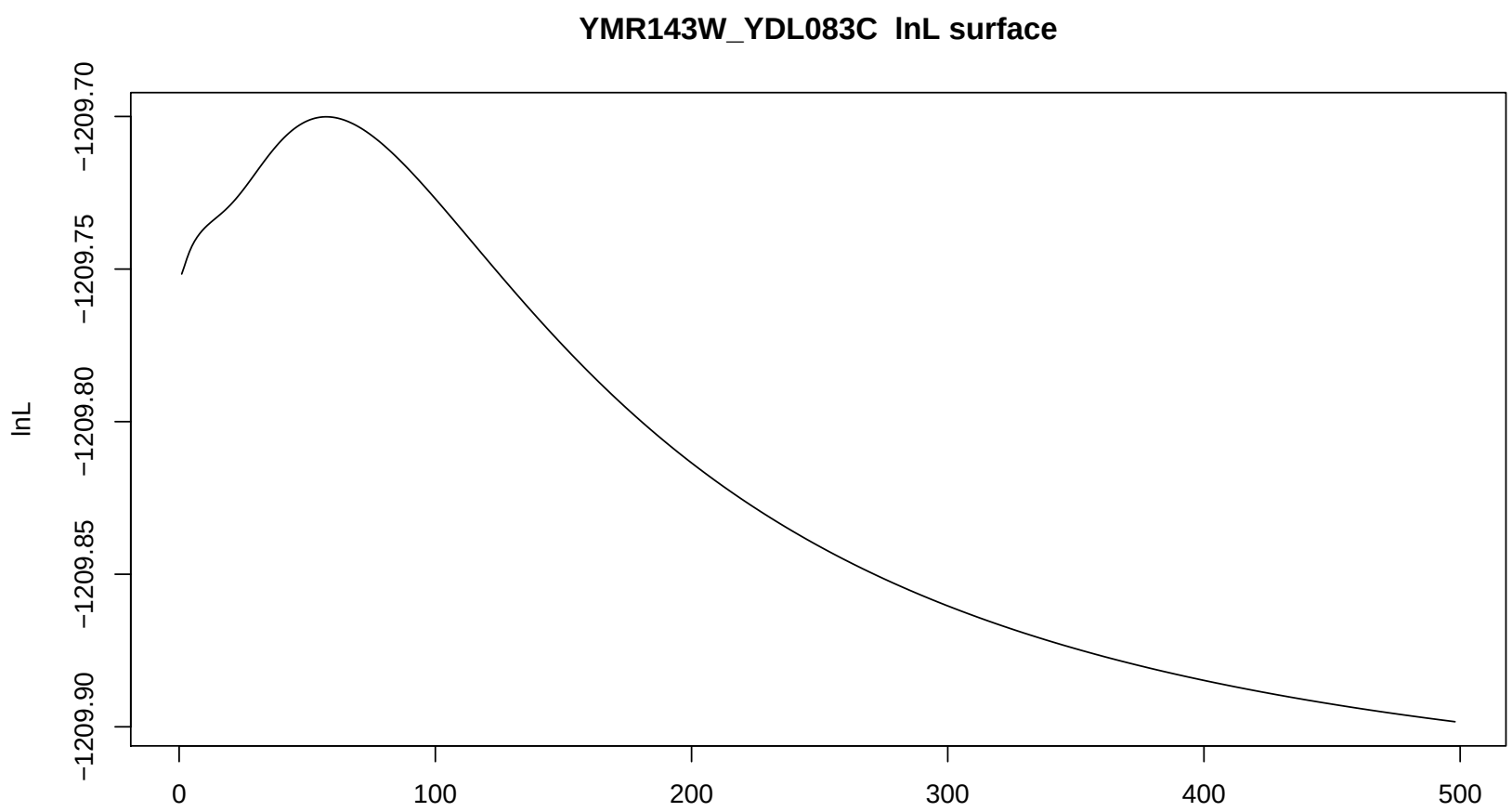

```
##                         lnL   max lnL tract length    Pr(S_0)   Pr(S_1)
## YNL069C_YIL133C -2322.829 -2322.829     3.008705 0.09060675 0.9093933
##                           df       d^2f c.i. tract length c.i tract length
## YNL069C_YIL133C 0.5614021 -0.9095127                 3         23.49291
```



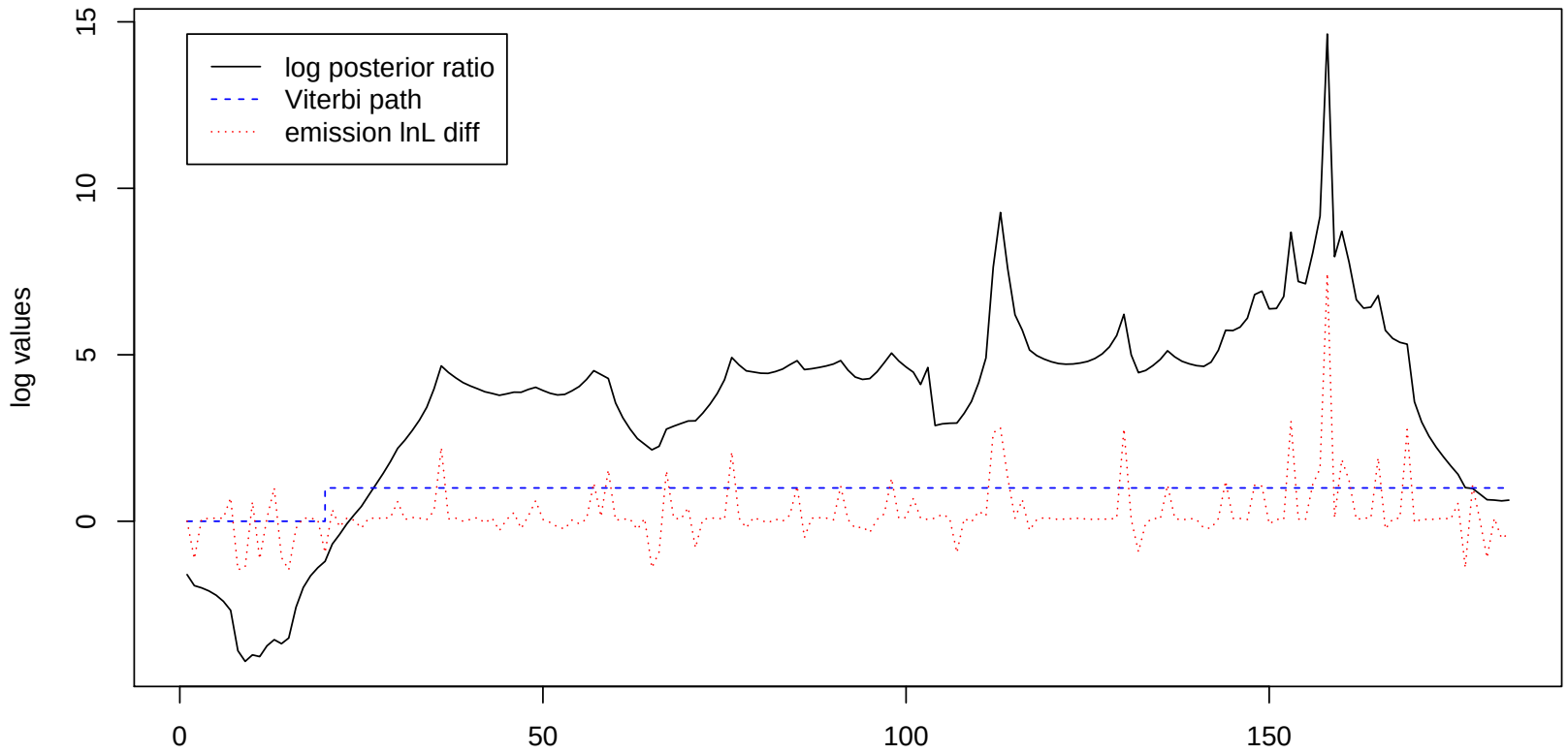

```
##                        lnL   max lnL tract length    Pr(S_0)   Pr(S_1)
## YNL069C_YIL133C -2322.829 -2322.829     3.008705 0.09060675 0.9093933
##                         df       d^2f c.i. tract length c.i tract length
## YNL069C_YIL133C 0.5614021 -0.9095127                 3         23.49291
```



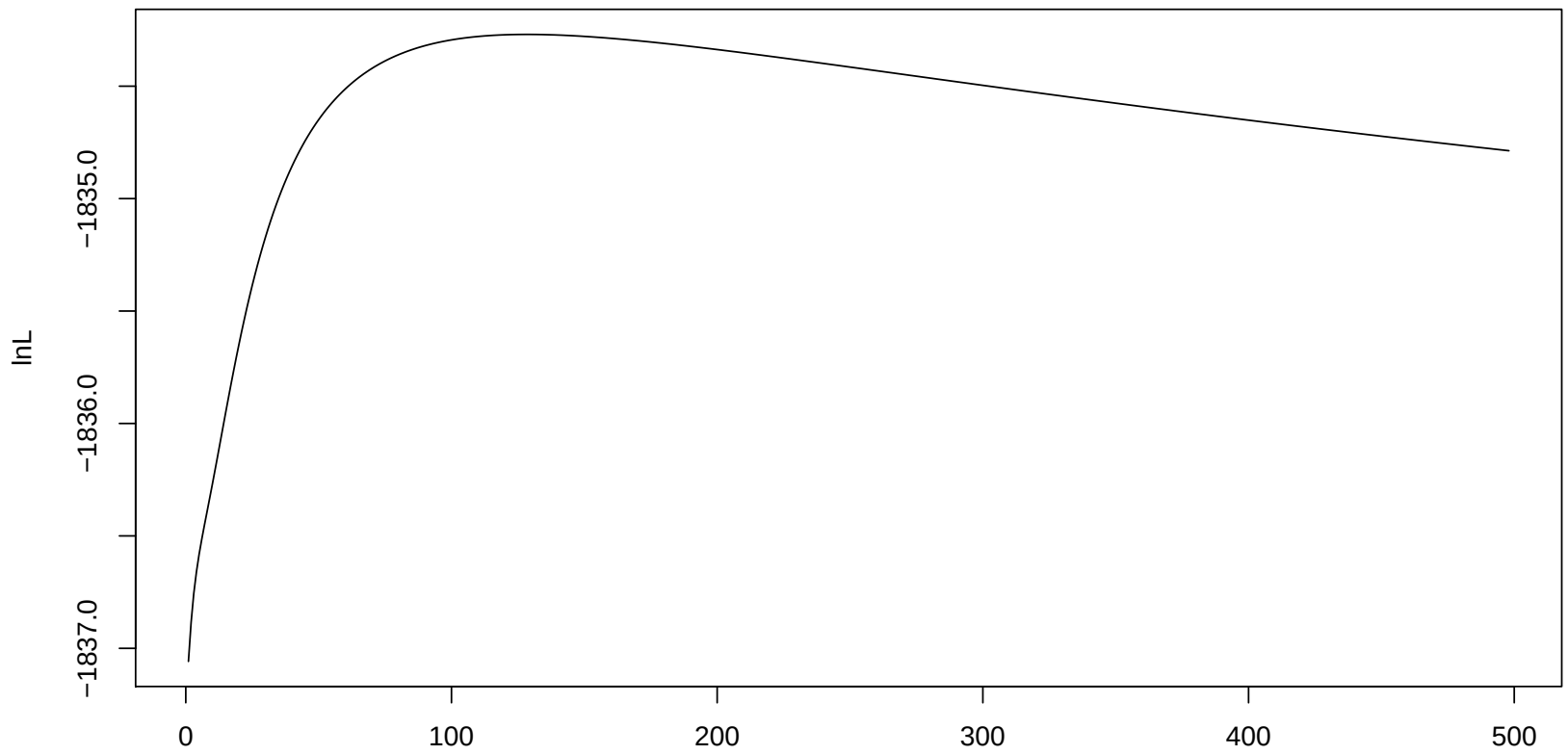

```
##                        lnL max lnL tract length    Pr(S_0)   Pr(S_1)
## YMR143W_YDL083C -1209.752 -1209.7     59.62301 0.02365226 0.9763477
##                             df       d^2f c.i. tract length
## YMR143W_YDL083C -9.777746e-05 -0.1631073                 3
##                  c.i tract length
## YMR143W_YDL083C          7639.92
```



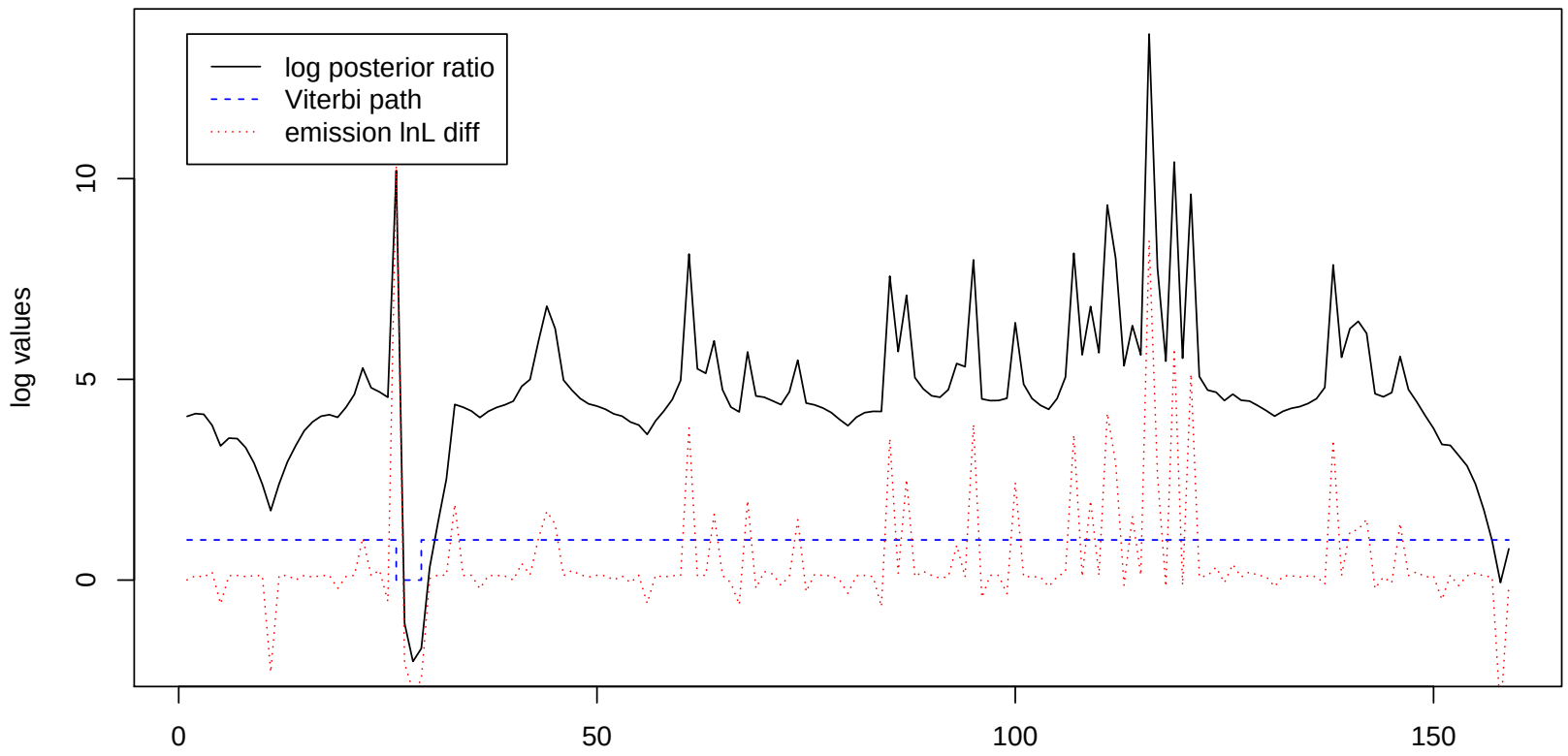

```
##                        lnL max lnL tract length    Pr(S_0)   Pr(S_1)
## YMR143W_YDL083C -1209.752 -1209.7     59.62301 0.02365226 0.9763477
##                             df       d^2f c.i. tract length
## YMR143W_YDL083C -9.777746e-05 -0.1631073                 3
##                   c.i tract length
## YMR143W_YDL083C          7639.92
```



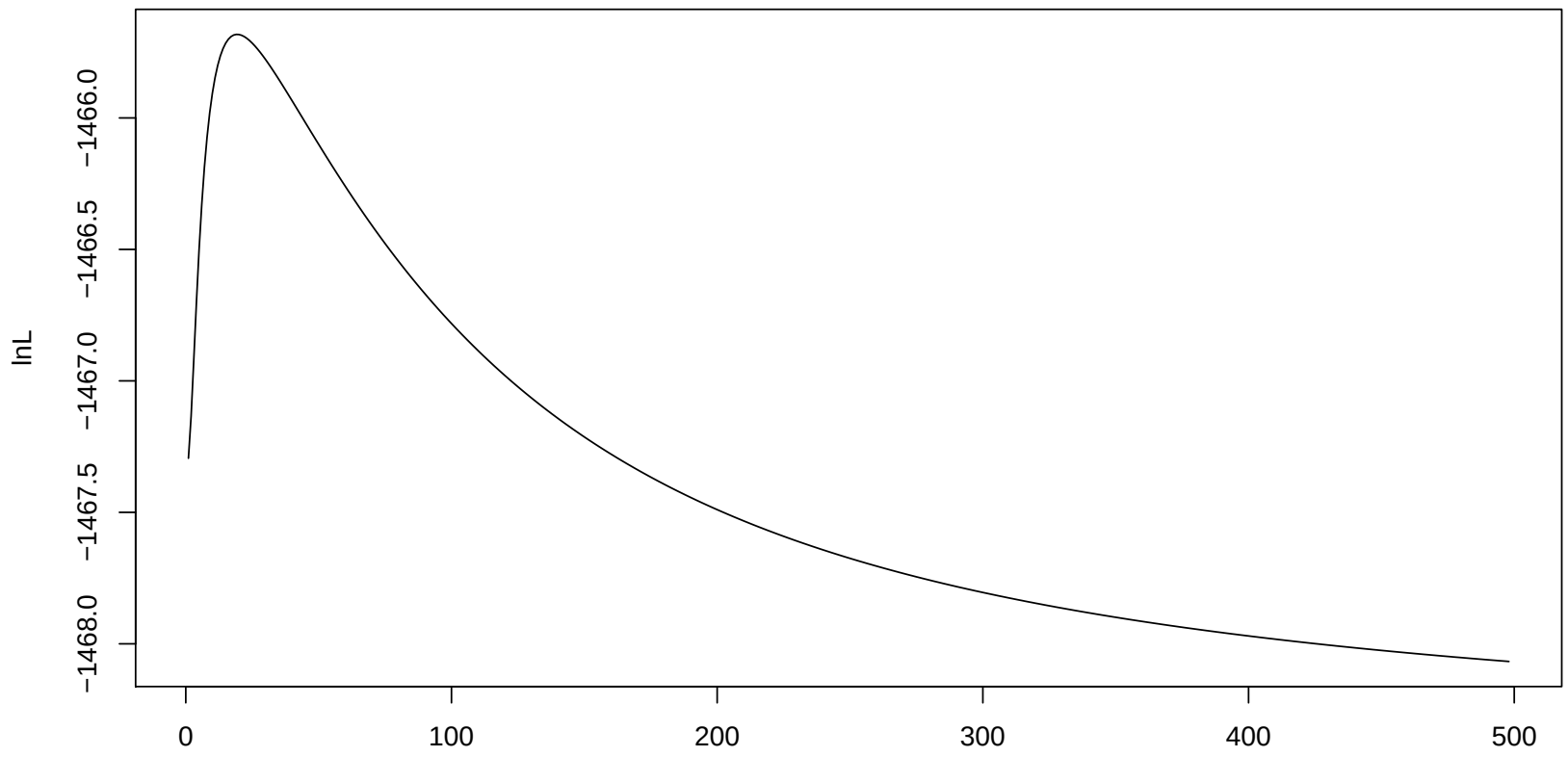

```
##                         lnL  max lnL tract length   Pr(S_0)   Pr(S_1)
## YJL177W_YKL180W -1837.058 -1834.27     130.8989 0.1163812 0.8836188
##                             df       d^2f c.i. tract length c.i tract length
## YJL177W_YKL180W -0.000453207 -0.7715419          14.05574          1219.04
```



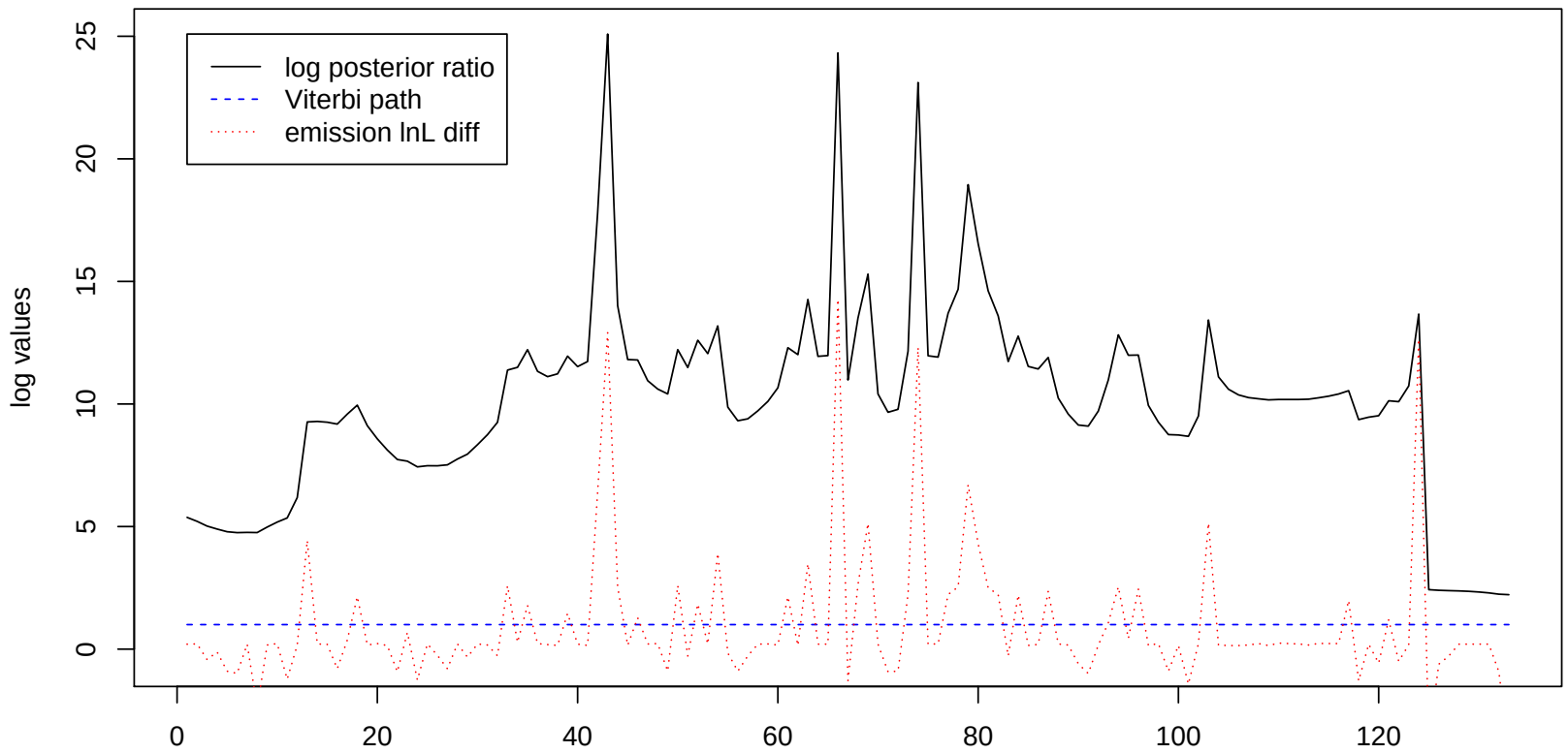

```
##                        lnL  max lnL tract length   Pr(S_0)   Pr(S_1)
## YJL177W_YKL180W -1837.058 -1834.27     130.8989 0.1163812 0.8836188
##                           df       d^2f c.i. tract length c.i tract length
## YJL177W_YKL180W -0.000453207 -0.7715419          14.05574          1219.04
```



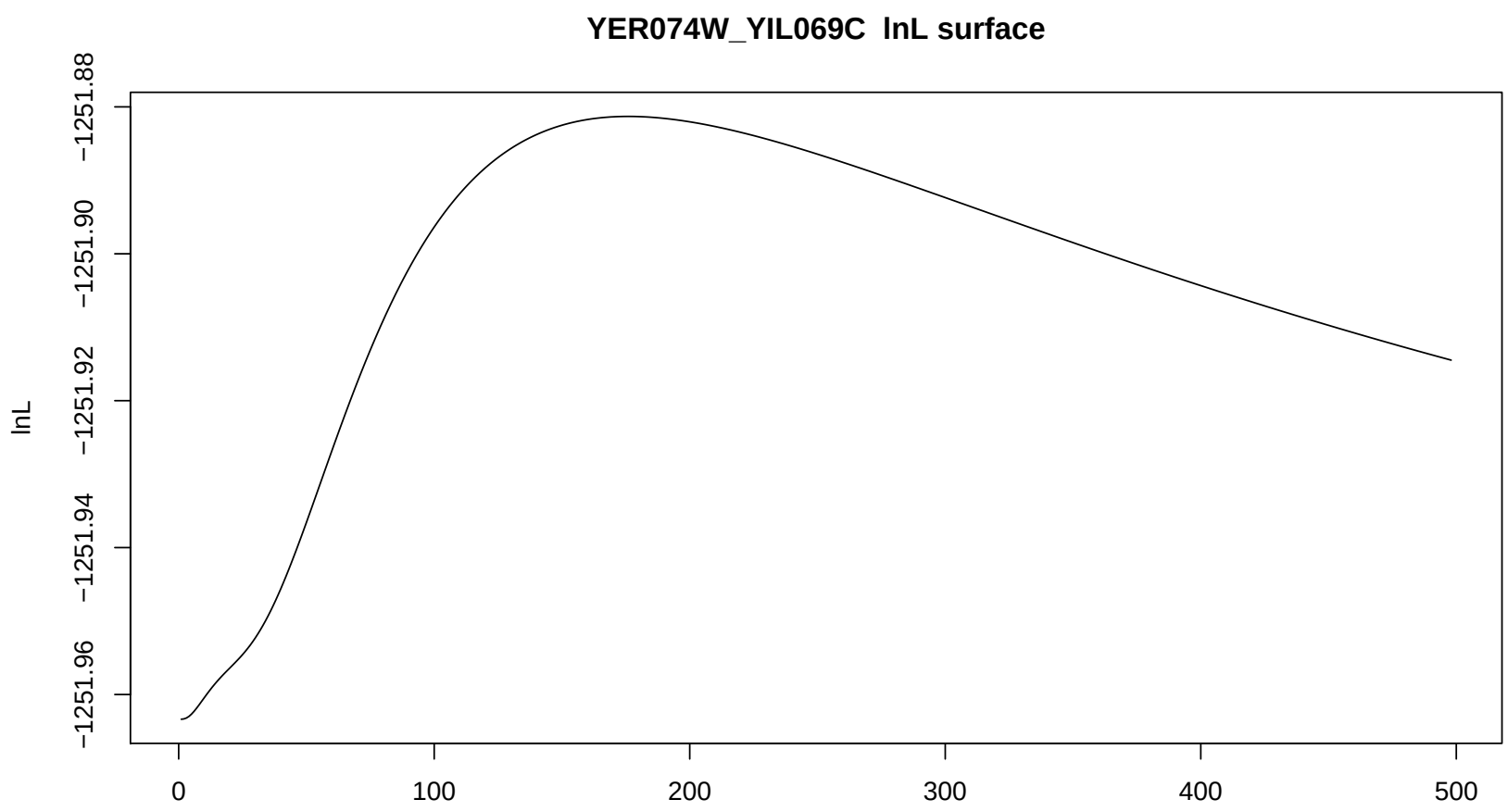

```
##                         lnL   max lnL tract length    Pr(S_0)   Pr(S_1)
## YBR191W_YPL079W -1467.294 -1465.682     21.40781 0.01817025 0.9818297
##                              df      d^2f c.i. tract length c.i tract length
## YBR191W_YPL079W -2.501933e-05 -1.253452          3.717658          123.275
```



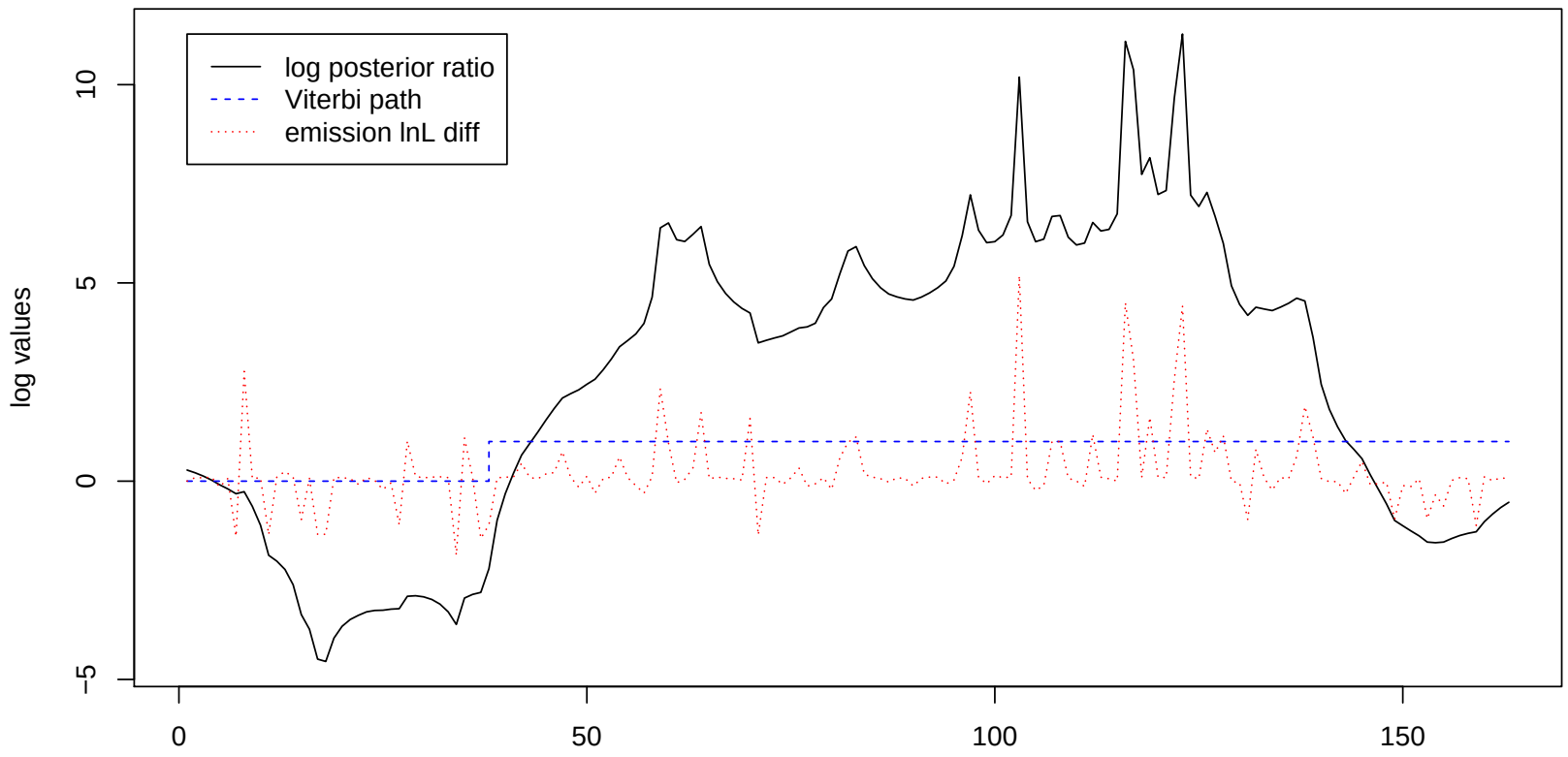

```
##                         lnL   max lnL tract length    Pr(S_0)   Pr(S_1)
## YBR191W_YPL079W -1467.294 -1465.682     21.40781 0.01817025 0.9818297
##                              df      d^2f c.i. tract length c.i tract length
## YBR191W_YPL079W -2.501933e-05 -1.253452          3.717658          123.275
```



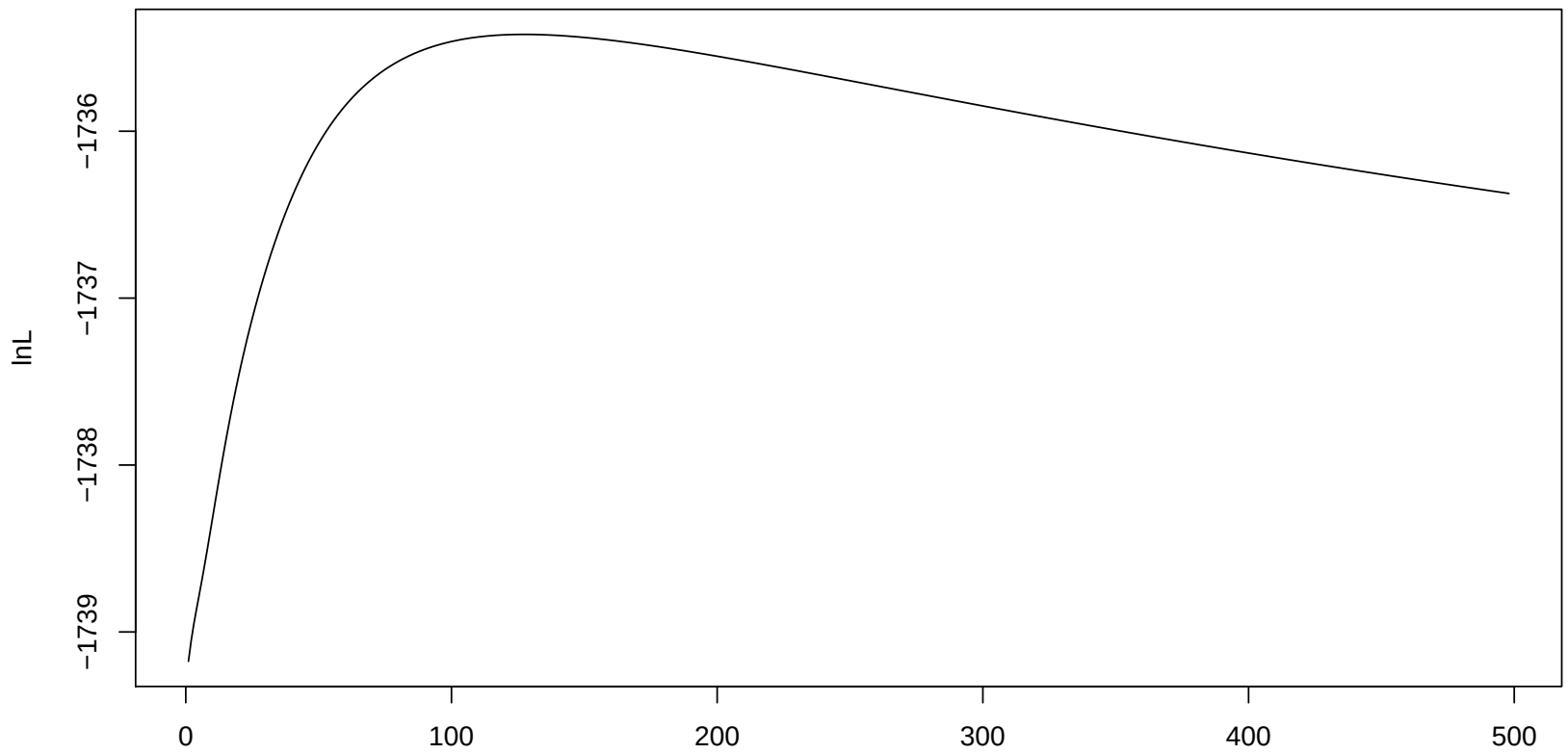

```
##                         lnL   max lnL tract length      Pr(S_0)   Pr(S_1)
## YER074W_YIL069C -1251.963 -1251.881     178.8784 0.0002647946 0.9997352
##                            df       d^2f c.i. tract length c.i tract length
## YER074W_YIL069C 6.910447e-05 -0.0940479                 3         106712.5
```



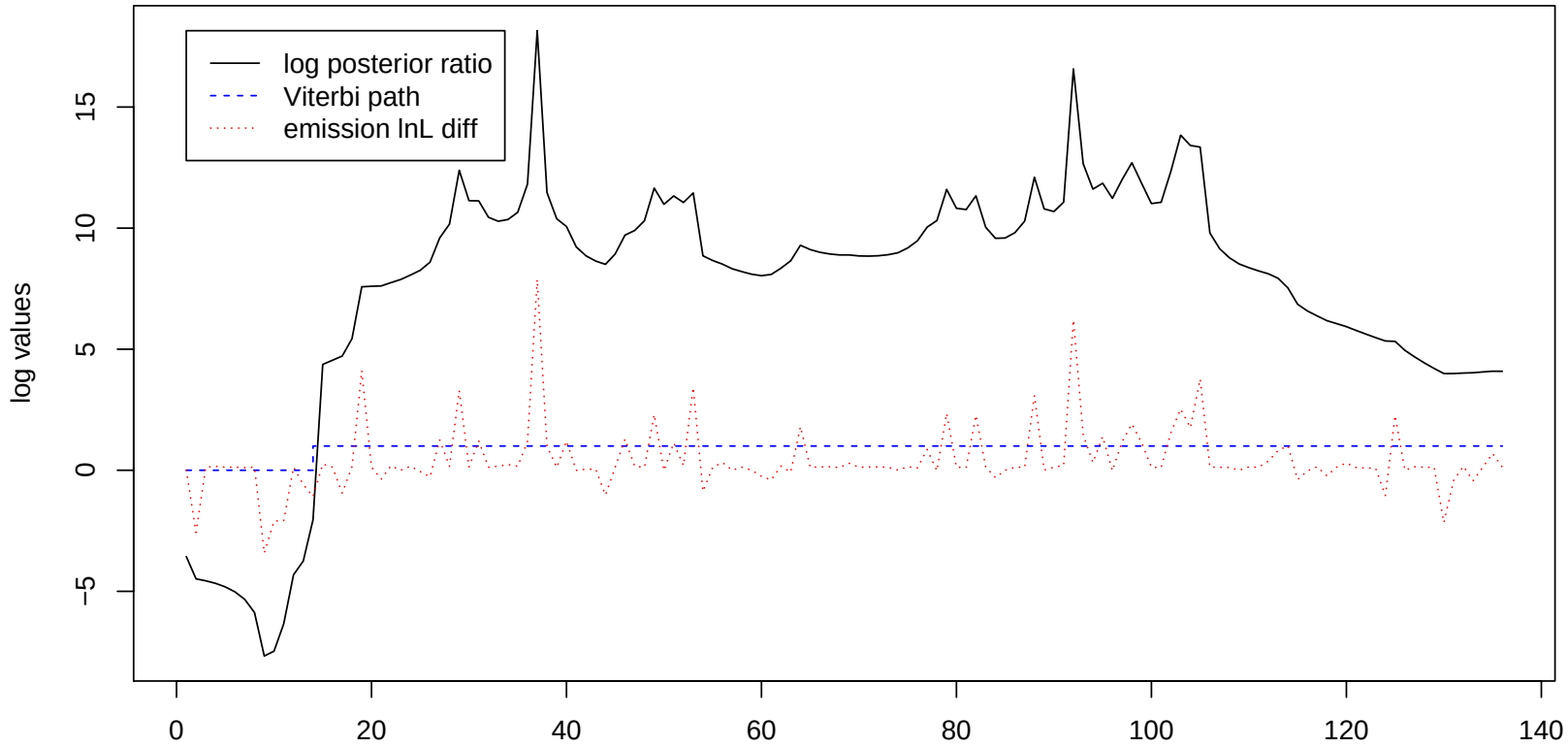

```
##                         lnL   max lnL tract length      Pr(S_0)   Pr(S_1)
## YER074W_YIL069C -1251.963 -1251.881     178.8784 0.0002647946 0.9997352
##                            df       d^2f c.i. tract length c.i tract length
## YER074W_YIL069C 6.910447e-05 -0.0940479                 3         106712.5
```



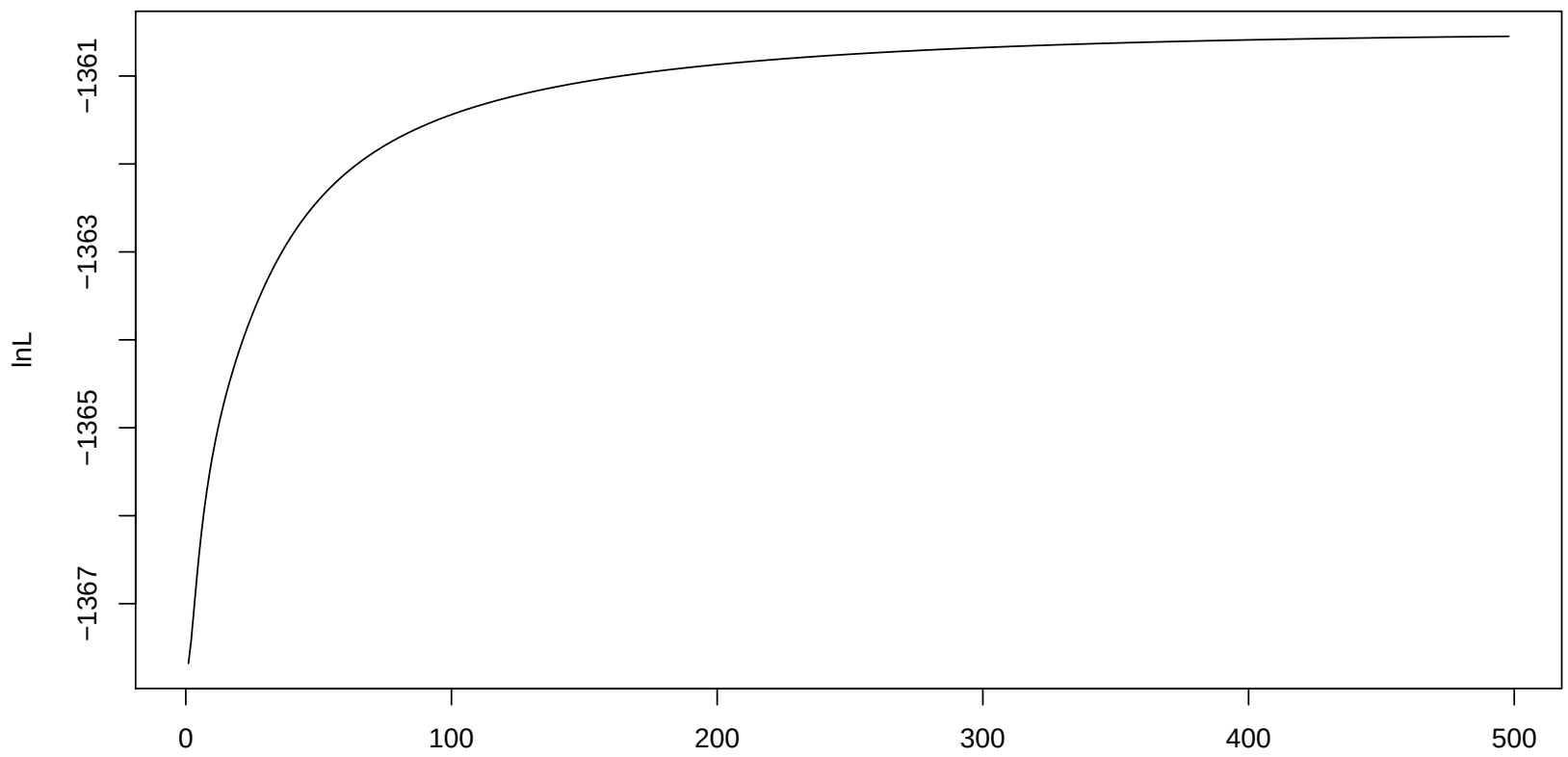

```
##                         lnL  max lnL tract length   Pr(S_0)   Pr(S_1)
## YDR418W_YEL054C -1739.176 -1735.42     129.8896 0.1269732 0.8730268
##                              df      d^2f c.i. tract length c.i tract length
## YDR418W_YEL054C -0.0006276662 -1.450723          25.51839         661.1433
```



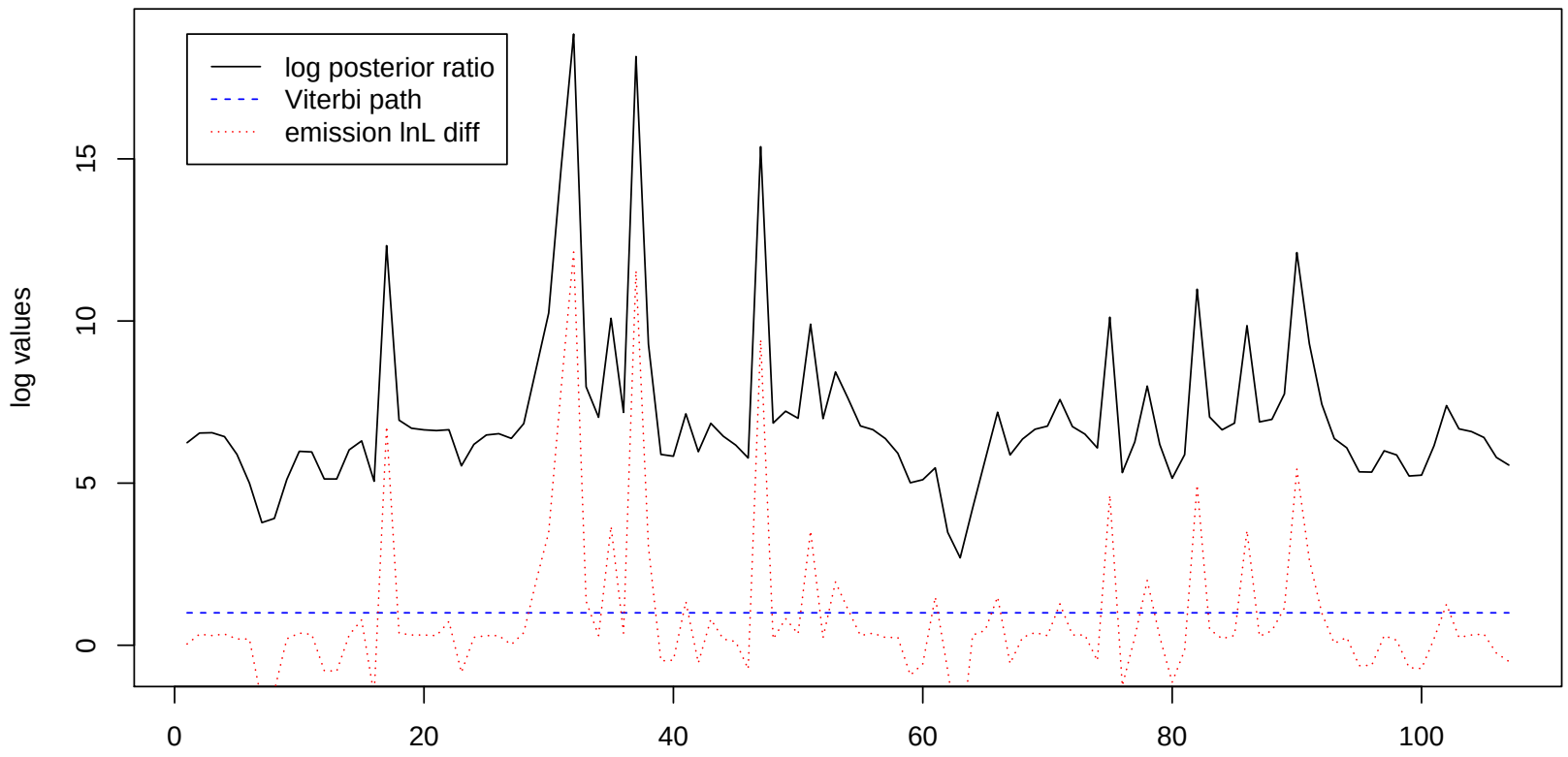

```
##                        lnL  max lnL tract length   Pr(S_0)   Pr(S_1)
## YDR418W_YEL054C -1739.176 -1735.42     129.8896 0.1269732 0.8730268
##                             df      d^2f c.i. tract length c.i tract length
## YDR418W_YEL054C -0.0006276662 -1.450723          25.51839         661.1433
```



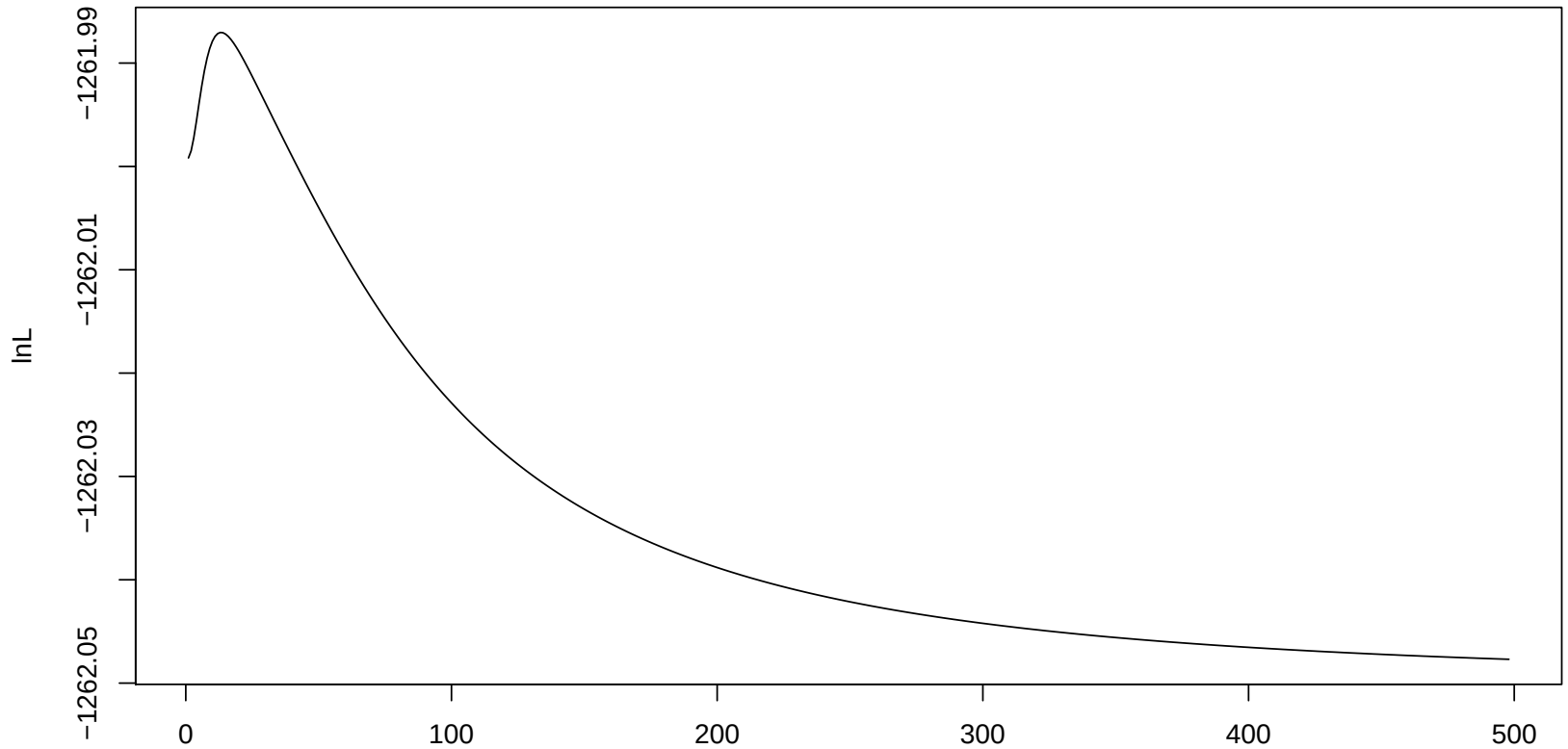

```
##                         lnL   max lnL tract length    Pr(S_0)  Pr(S_1)
## YBL087C_YER117W -1367.679 -1360.529     674.5267 0.02308803 0.976912
##                            df       d^2f c.i. tract length c.i tract length
## YBL087C_YER117W 0.000296163 -0.4636916          37.92746         11996.22
```



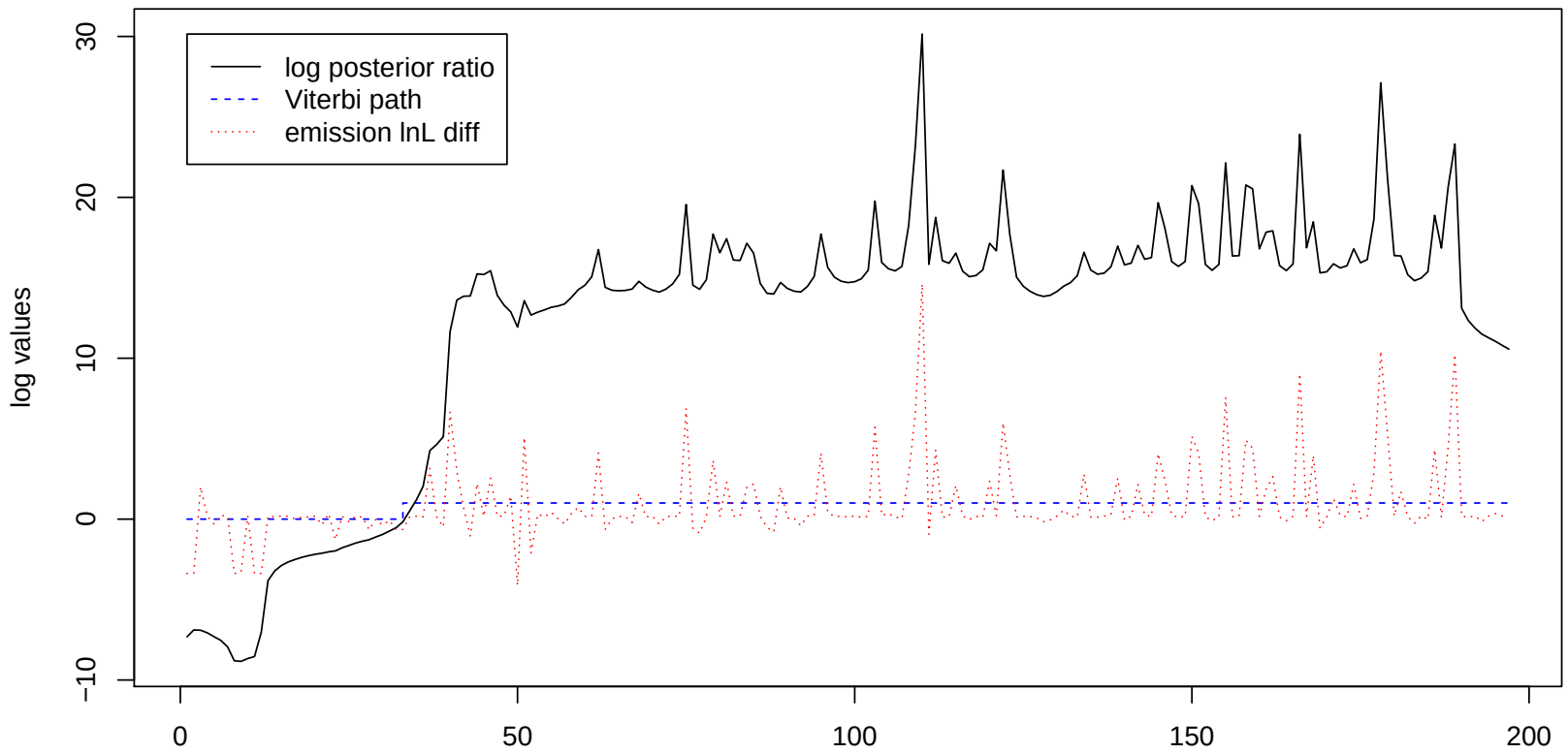

```
##                        lnL   max lnL tract length    Pr(S_0)  Pr(S_1)
## YBL087C_YER117W -1367.679 -1360.529     674.5267 0.02308803 0.976912
##                           df       d^2f c.i. tract length c.i tract length
## YBL087C_YER117W 0.000296163 -0.4636916          37.92746         11996.22
```



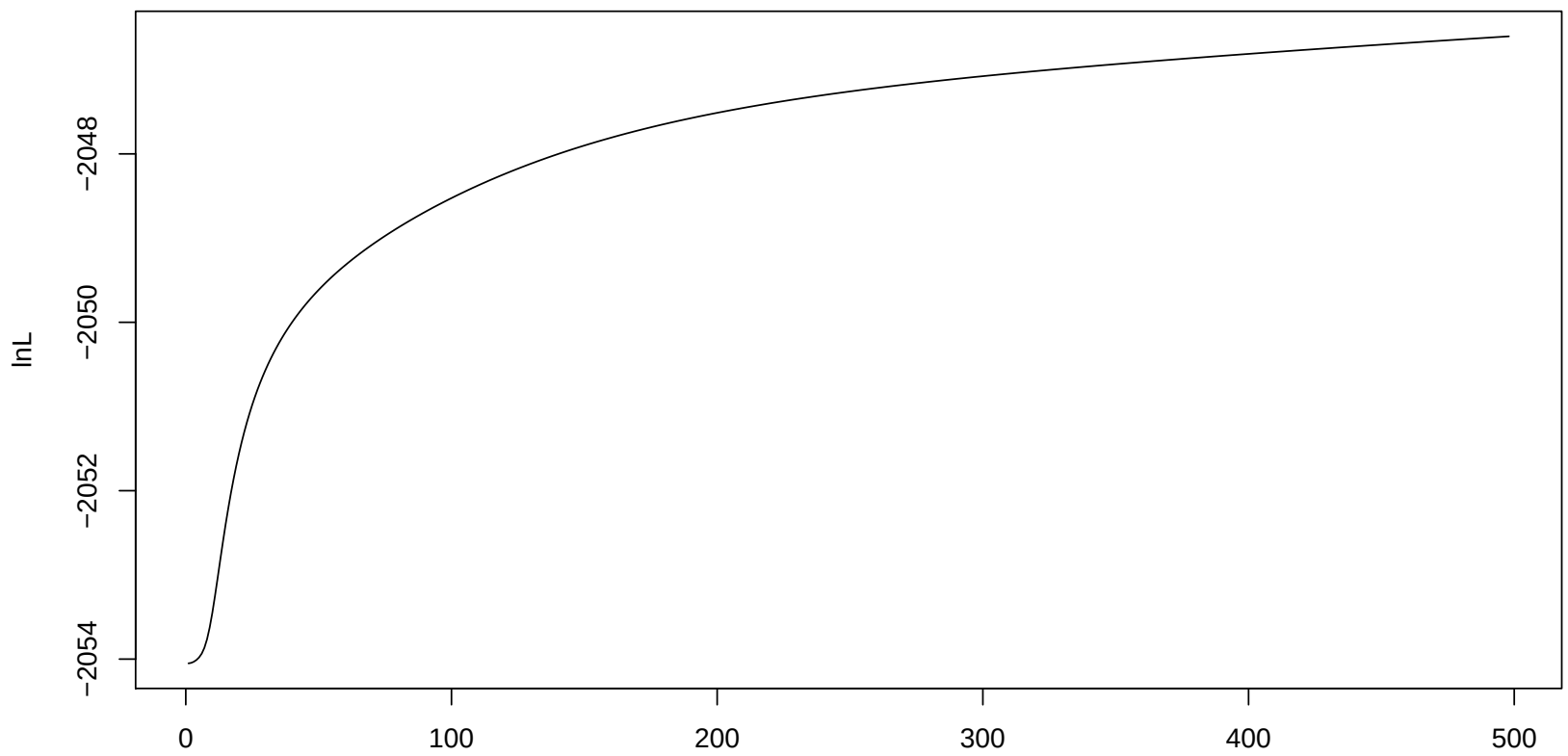

```
##                         lnL   max lnL tract length     Pr(S_0)   Pr(S_1)
## YLR333C_YGR027C -1261.999 -1261.987     15.33108 0.002249659 0.9977503
##                            df        d^2f c.i. tract length
## YLR333C_YGR027C 1.695126e-07 -0.02976511                 3
##                  c.i tract length
## YLR333C_YGR027C          1316531
```



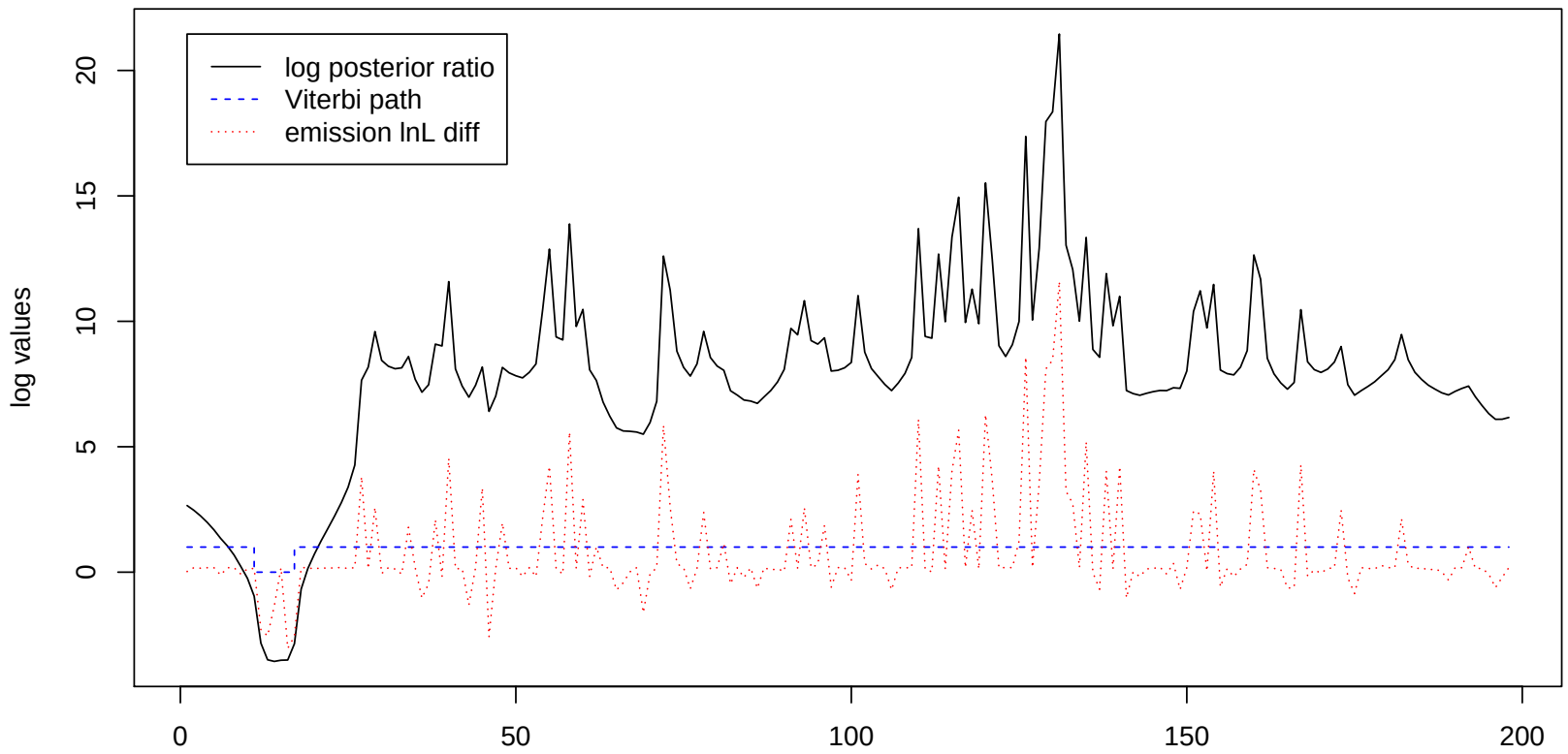

```
##                        lnL   max lnL tract length     Pr(S_0)   Pr(S_1)
## YLR333C_YGR027C -1261.999 -1261.987     15.33108 0.002249659 0.9977503
##                            df        d^2f c.i. tract length
## YLR333C_YGR027C 1.695126e-07 -0.02976511                 3
##                 c.i tract length
## YLR333C_YGR027C          1316531
```



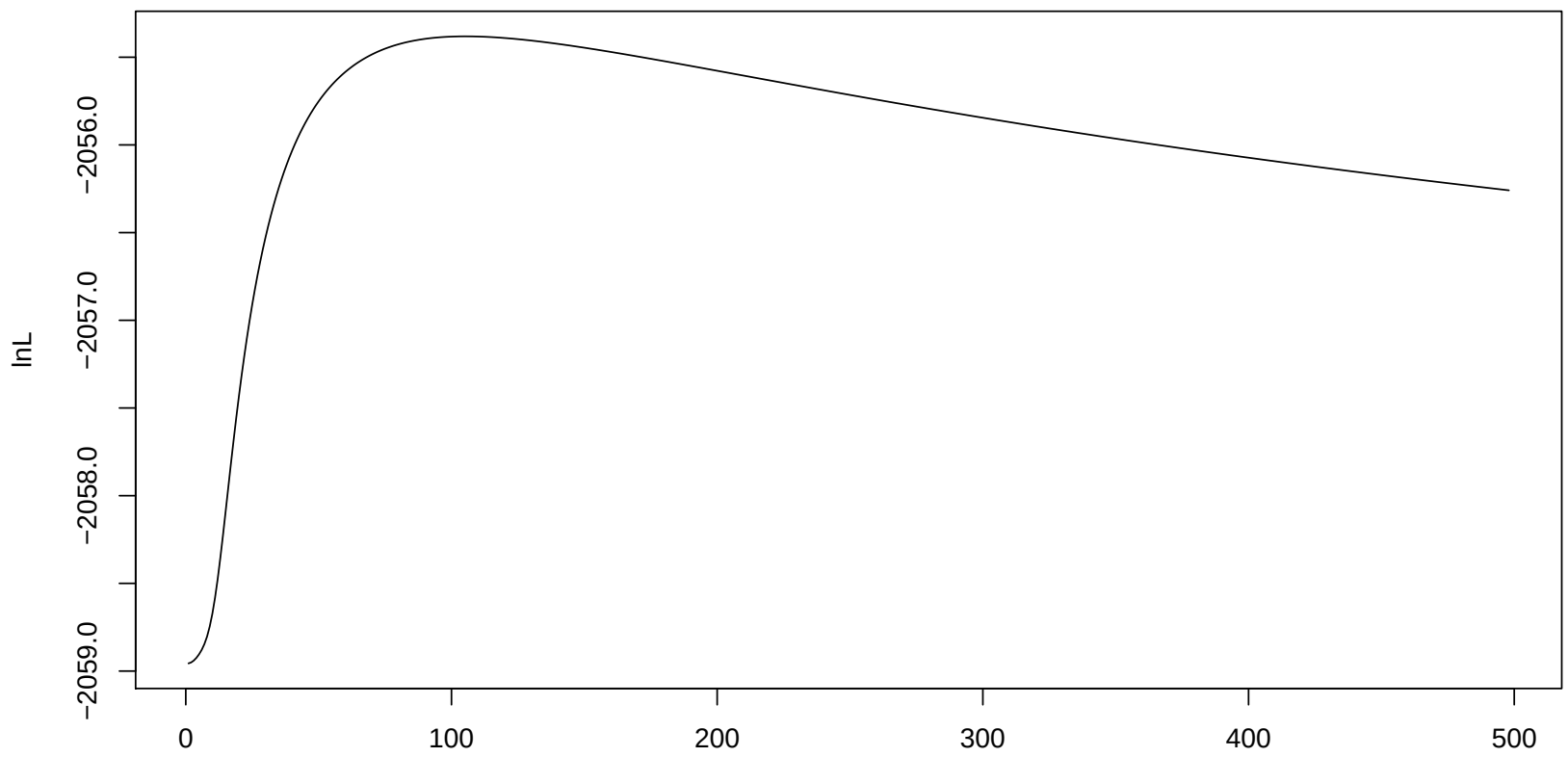

```
##                         lnL   max lnL tract length      Pr(S_0)   Pr(S_1)
## YMR142C_YDL082W -2054.051 -2045.363     2871.324 0.0009455638 0.9990544
##                            df       d^2f c.i. tract length c.i tract length
## YMR142C_YDL082W 3.500825e-05 -0.8301934          334.0864         24677.75
```




```
##                         lnL   max lnL tract length      Pr(S_0)   Pr(S_1)
## YMR142C_YDL082W -2054.051 -2045.363     2871.324 0.0009455638 0.9990544
##                            df       d^2f c.i. tract length c.i tract length
## YMR142C_YDL082W 3.500825e-05 -0.8301934          334.0864         24677.75
```



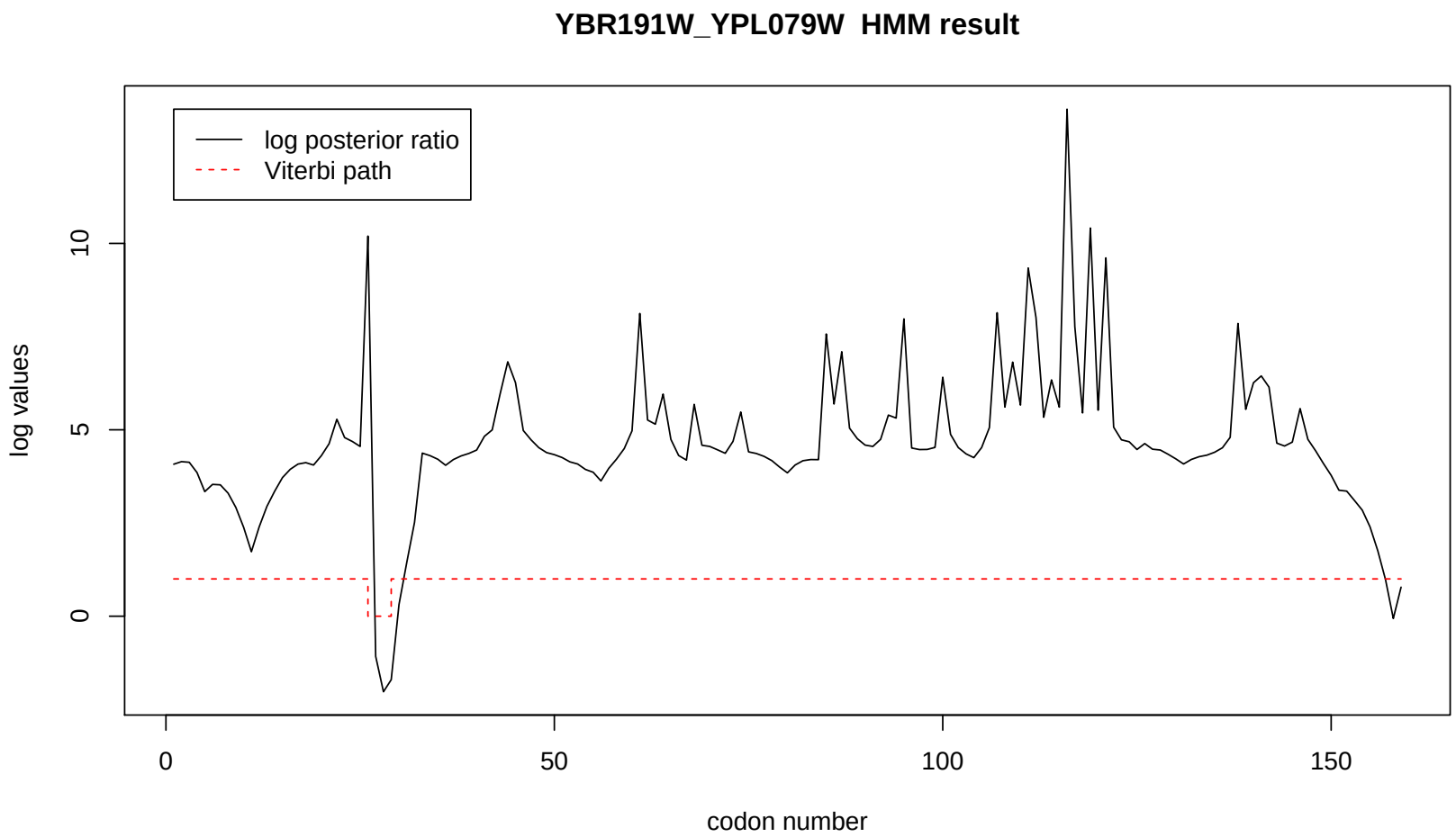

```
##                         lnL   max lnL tract length     Pr(S_0)   Pr(S_1)
## YER102W_YBL072C -2058.956 -2055.381     107.7412 0.001988219 0.9980118
##                             df     d^2f c.i. tract length c.i tract length
## YER102W_YBL072C 0.0005777144 -1.17477          17.66162         657.2537
```



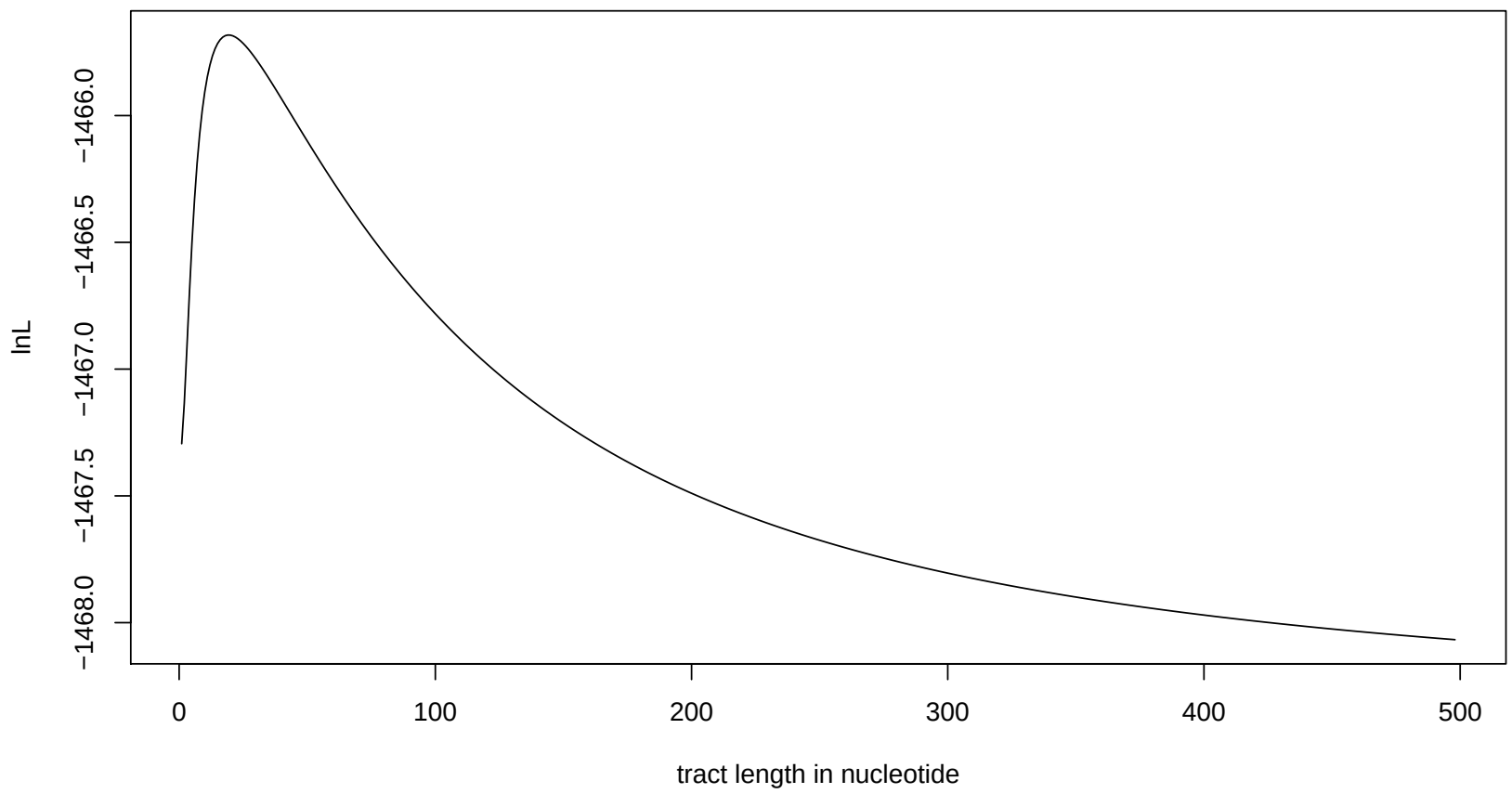

```
##                         lnL   max lnL tract length     Pr(S_0)   Pr(S_1)
## YER102W_YBL072C -2058.956 -2055.381     107.7412 0.001988219 0.9980118
##                            df     d^2f c.i. tract length c.i tract length
## YER102W_YBL072C 0.0005777144 -1.17477          17.66162         657.2537
```

```
# plot one paralog
paralog = "YER102W_YBL072C"
paralog = "YBR191W_YPL079W"
lnL.ratio <- as.vector(read.table(paste("./summary/", paralog, "_MG94_nonclock_HMM_log_posterior_ratio.txt", sep = "")))
Viterbi.path <- as.vector(read.table(paste("./summary/", paralog, "_MG94_nonclock_HMM_Viterbi_path.txt", sep = "")))
lnL.surface <- as.vector(read.table(paste("./summary/", paralog, "_MG94_nonclock_HMM_lnL_surface.txt", sep = "")))
IGC.sw.lnL <- as.vector(read.table(paste("./summary/", paralog, "_MG94_nonclock_sw_lnL.txt", sep = "")))
Force.sw.lnL <- as.vector(read.table(paste("./summary/Force_", paralog, "_MG94_nonclock_sw_lnL.txt", sep = "")))

plot(lnL.ratio[, 1], xlab = "codon number", ylab = "log values",
     type = "l", col = "black", lty = 1,
```

```
     main = paste(paralog, " HMM result"),
     ylim = c(min(-0.5, min(lnL.ratio)), max(lnL.ratio)))
lines(1:dim(Viterbi.path)[1], Viterbi.path[, 1], type = "S", lty = 2, col = "red")
#lines(1:dim(IGC.sw.lnL)[1], IGC.sw.lnL[, 2] - Force.sw.lnL[, 2], type = "l", lty = 3, col = "red")
legend(1, max(lnL.ratio), legend = c("log posterior ratio", "Viterbi path"),
       lty = c(1, 2), col = c( "black", "red"))
```




```
plot(-lnL.surface[, 1], xlab = "tract length in nucleotide", ylab= "lnL", type = "l", col = "black", lty = 1,
     main = paste(paralog, " lnL surface"))
```


YBR191W_YPL079W lnL surface
lnL
−1466.0
−1466.5
−1467.0
−1467.5
−1468.0
0
100
200
300
400
500
tract length in nucleotide