HMM_Summary

Xiang Ji

8/29/2017

This R markdown file summarizes the MG94+IS-IGC+HMM analysis results.

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

Now read in estimated parameter values

```r
HMM.Yeast.1D <- NULL
HMM.Yeast.all <- NULL
for(pair in filtered.pairs){
  summary.file.1D.name <- paste("HMM", pair, "MG94_nonclock_1D_summary.txt", sep = "_")
  summary.file.1D <- paste("./summary/", summary.file.1D.name, sep = "")
  summary.file.all <- paste("./summary/HMM", pair, "MG94_nonclock_all_summary.txt", sep = "_")
  all <- readLines(summary.file.1D, n = -1)
  row.names <- strsplit(all[length(all)], ' ')[[1]][-1]
  col.name <- pair
  summary_mat <- as.matrix(read.table(summary.file.1D,
                                      row.names = row.names,
                                      col.names = col.name))
  HMM.Yeast.1D <- cbind(HMM.Yeast.1D, summary_mat)
  summary_mat <- as.matrix(read.table(summary.file.all,
                                      row.names = row.names,
                                      col.names = col.name))
  HMM.Yeast.all <- cbind(HMM.Yeast.all, summary_mat)
}
```
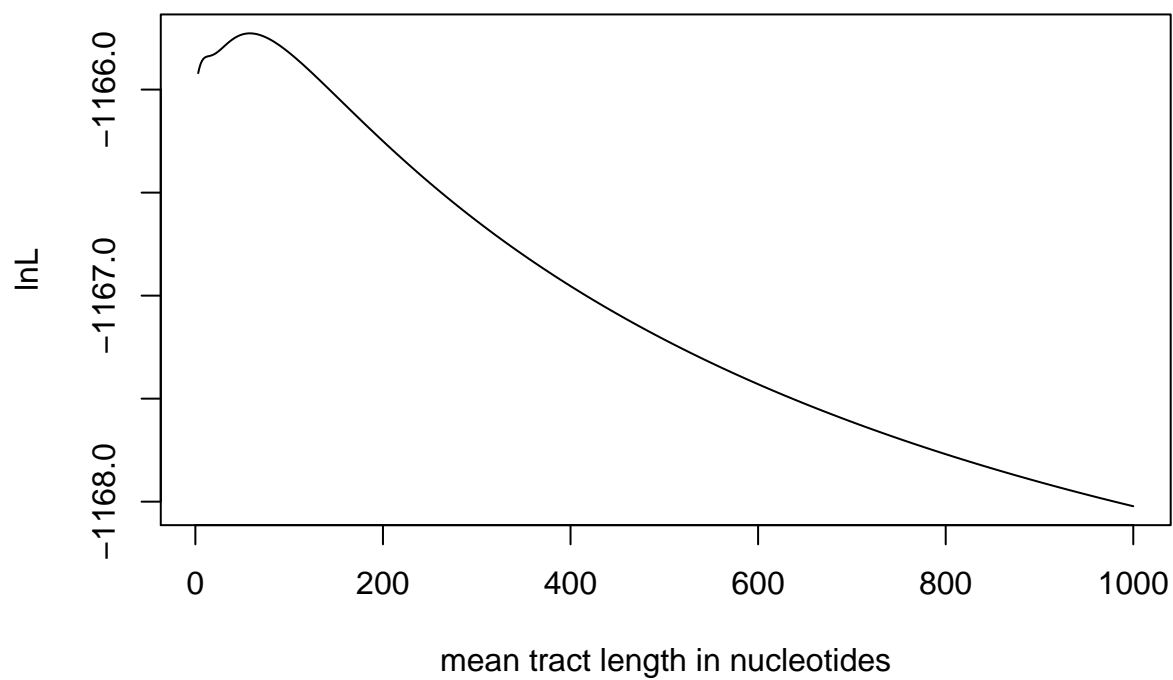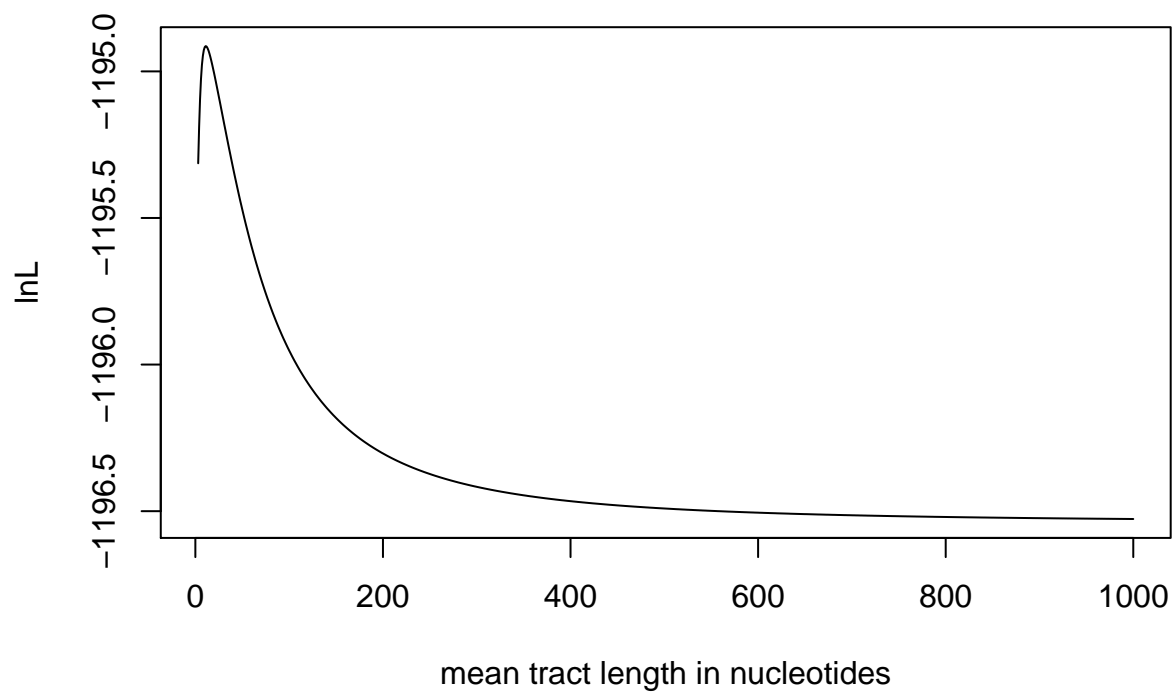
Plot the lnL surface

```r
tract.length.list <- 3:1000
# Now plot individual lnL surface along mean tract length of each gene pair
# all other parameters are fixed at MLE
for (pair in filtered.pairs){
  plot.file.name <- paste("HMM", pair, "lnL_1D_surface.txt", sep = "_")
  plot_file <- paste("./plot/", plot.file.name, sep = "")
  summary_mat <- read.table(plot_file)
  assign(paste(pair, "lnL_1D_surface", sep = "_"), summary_mat)
  plot(tract.length.list, summary_mat[, 2], type = "l",
       main = paste(pair, "lnL_surface", sep = "_"),
       xlab = "mean tract length in nucleotides",
       ylab = "lnL")
  print(3.0 / HMM.Yeast.1D["tract_p", pair])
}
```
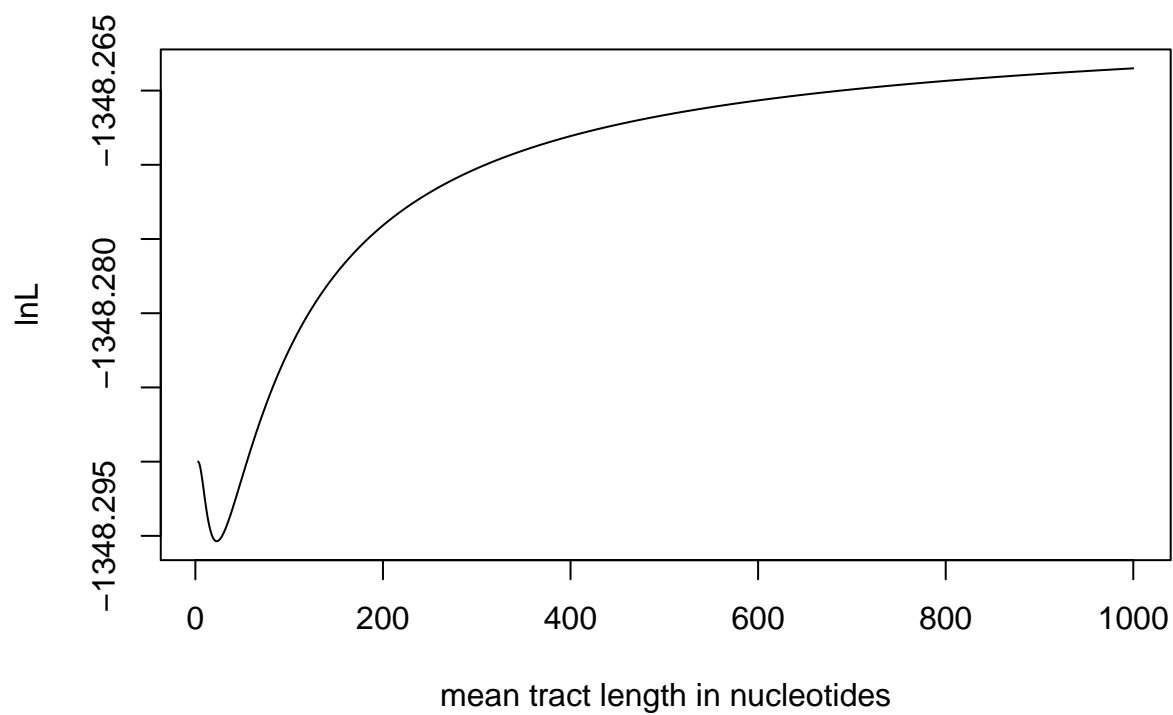
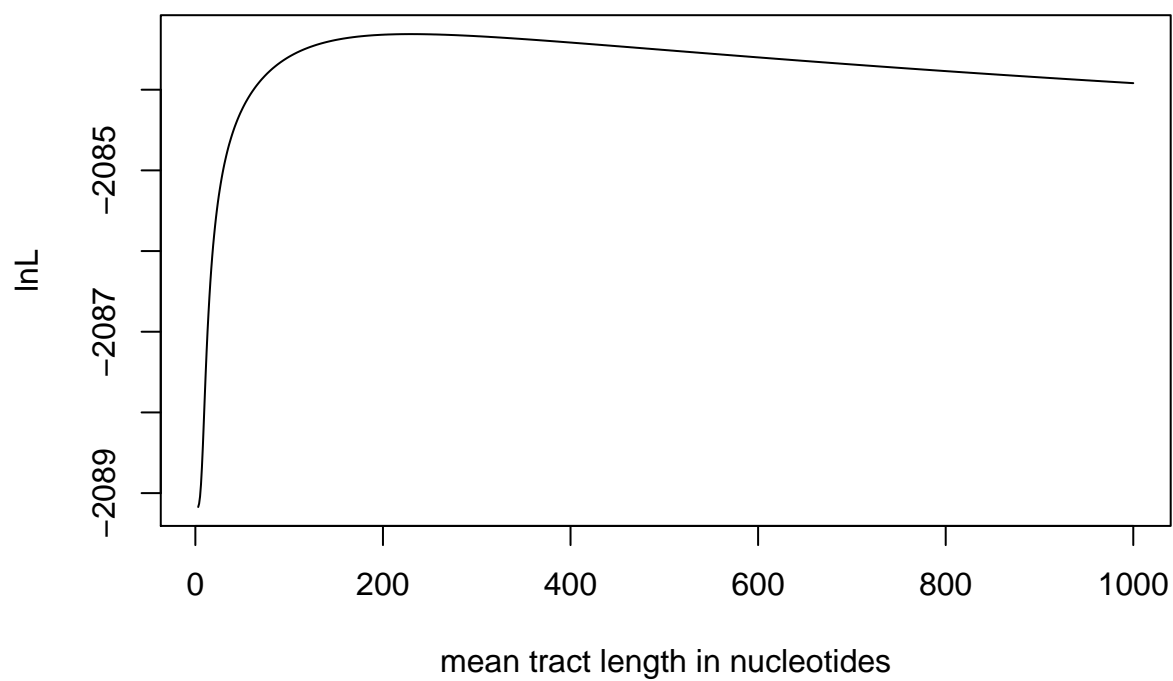**YLR406C_YDL075W_lnL_surface**

[1] 58.13762

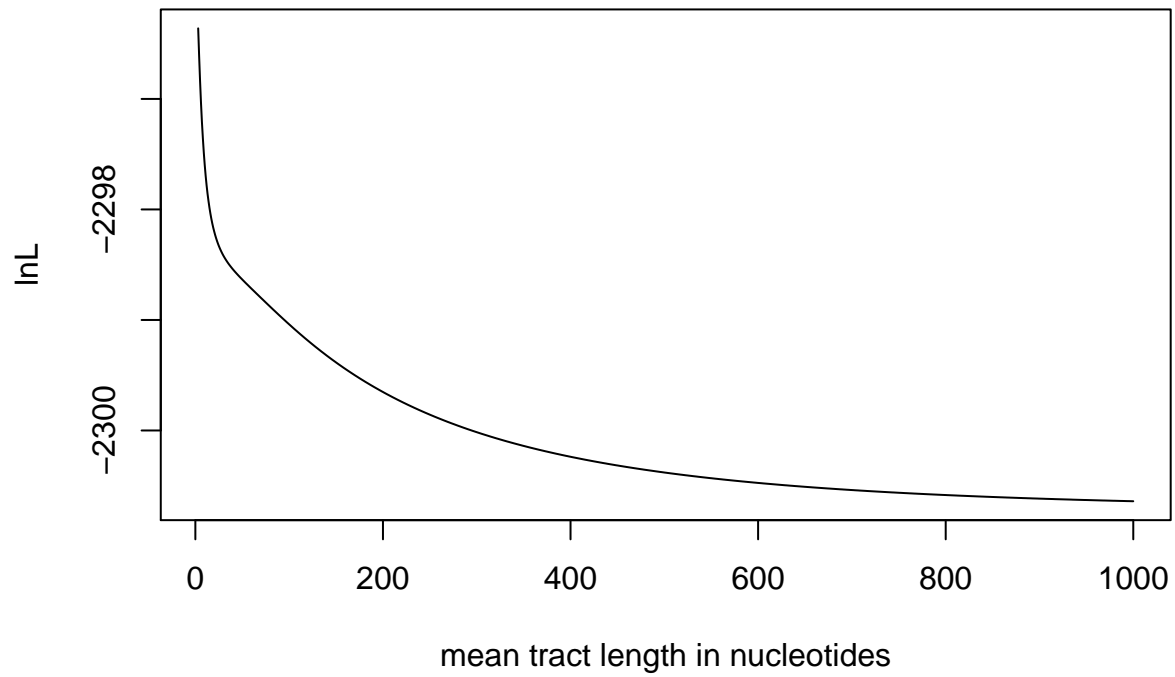**YER131W_YGL189C_lnL_surface**

[1] 11.12251

YML026C_YDR450W_lnL_surface

mean tract length in nucleotides

[1] 3

YNL301C_YOL120C_lnL_surface

mean tract length in nucleotides

[1] 229.7465

YNL069C_YIL133C_lnL_surface

[1] 3

YMR143W_YDL083C_lnL_surface
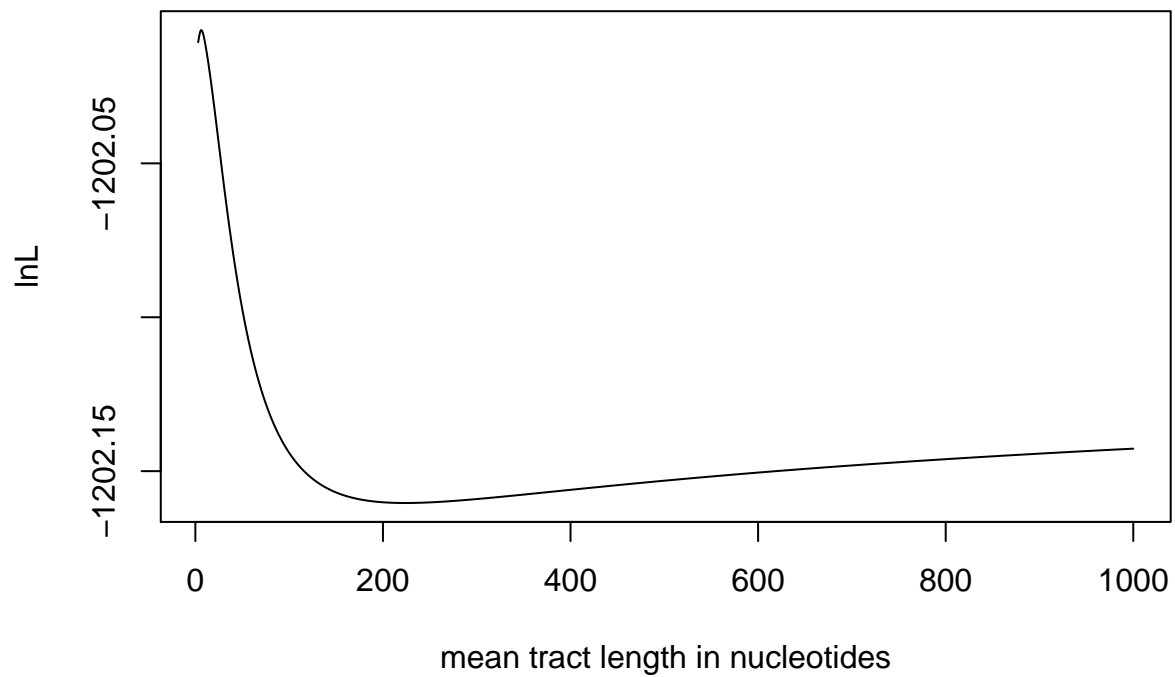
[1] 6.266257

YJL177W_YKL180W_lnL_surface
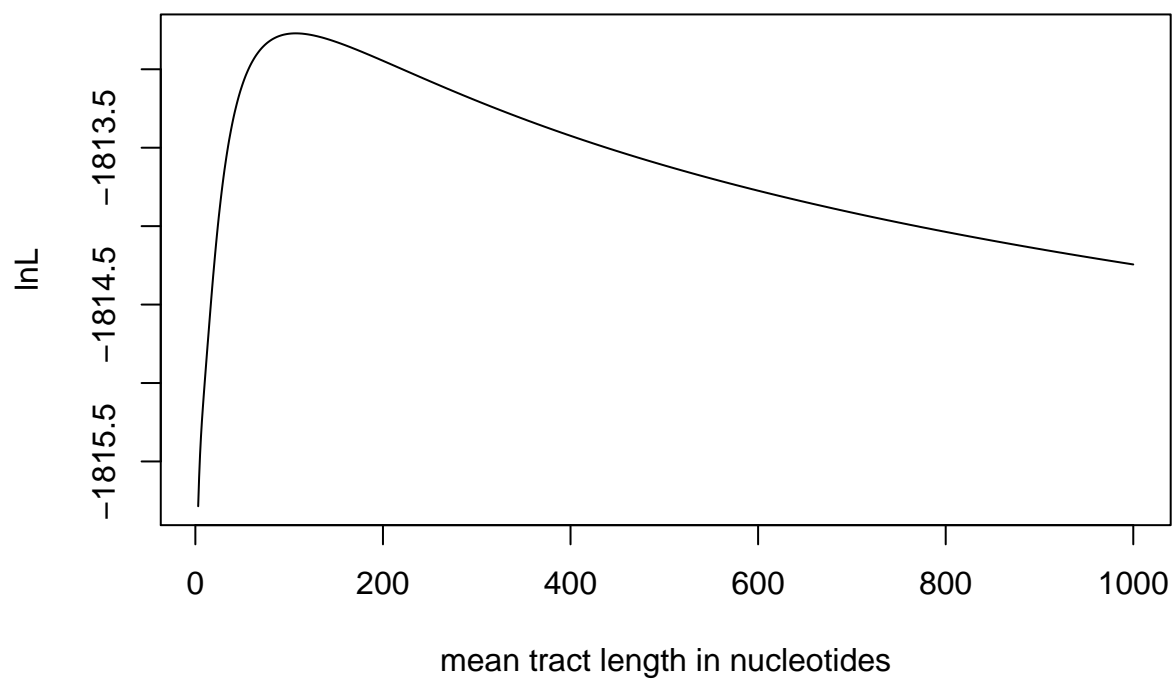

```
## [1] 106.9944
```

YBR191W_YPL079W_lnL_surface


```
## [1] 21.08593
```

5

YER074W_YIL069C_lnL_surface

mean tract length in nucleotides
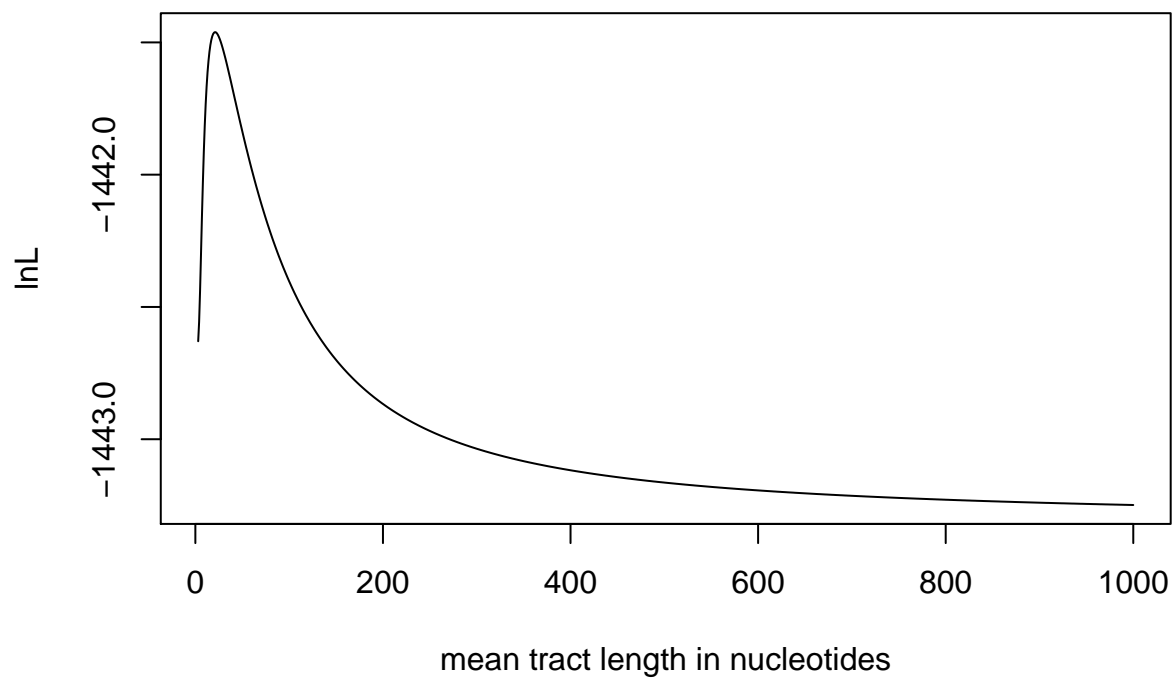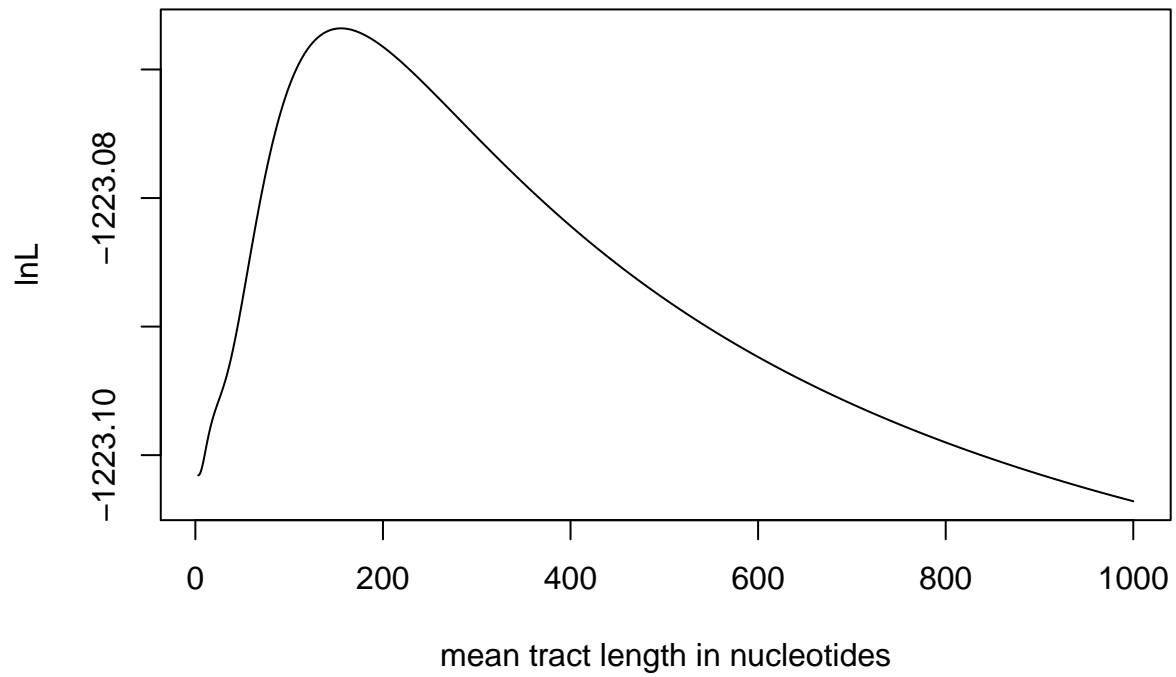
```
## [1] 3.000034
```

YDR418W_YEL054C_lnL_surface

mean tract length in nucleotides
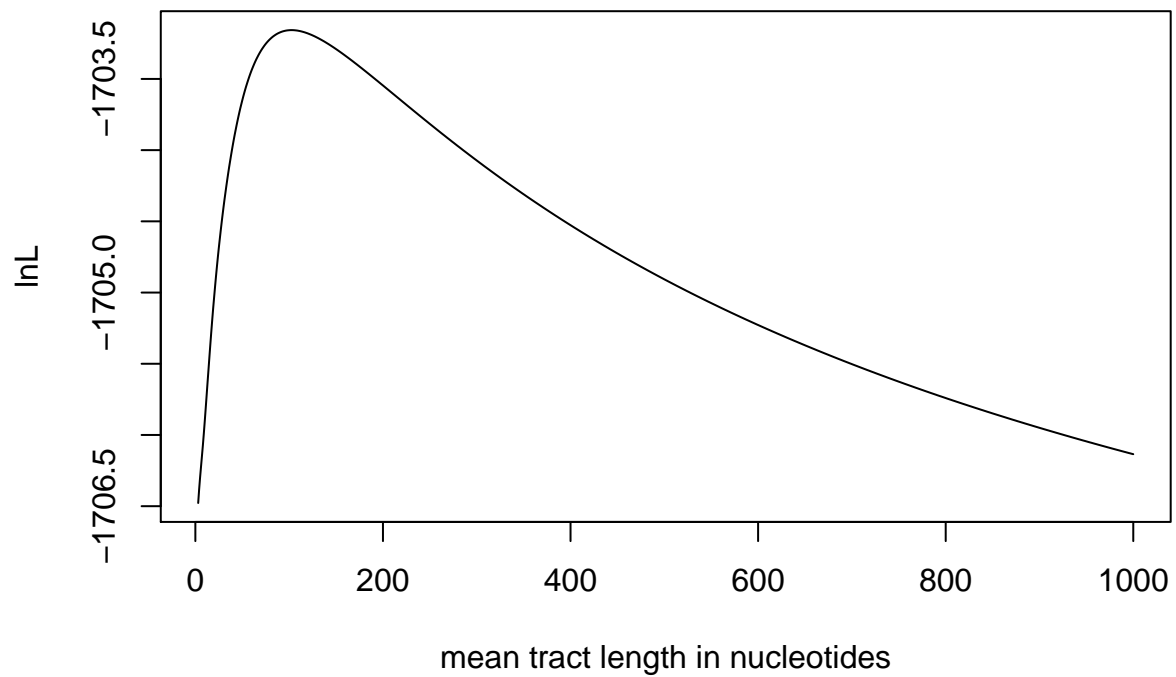
```
## [1] 102.5466
```

YBL087C_YER117W_lnL_surface

mean tract length in nucleotides
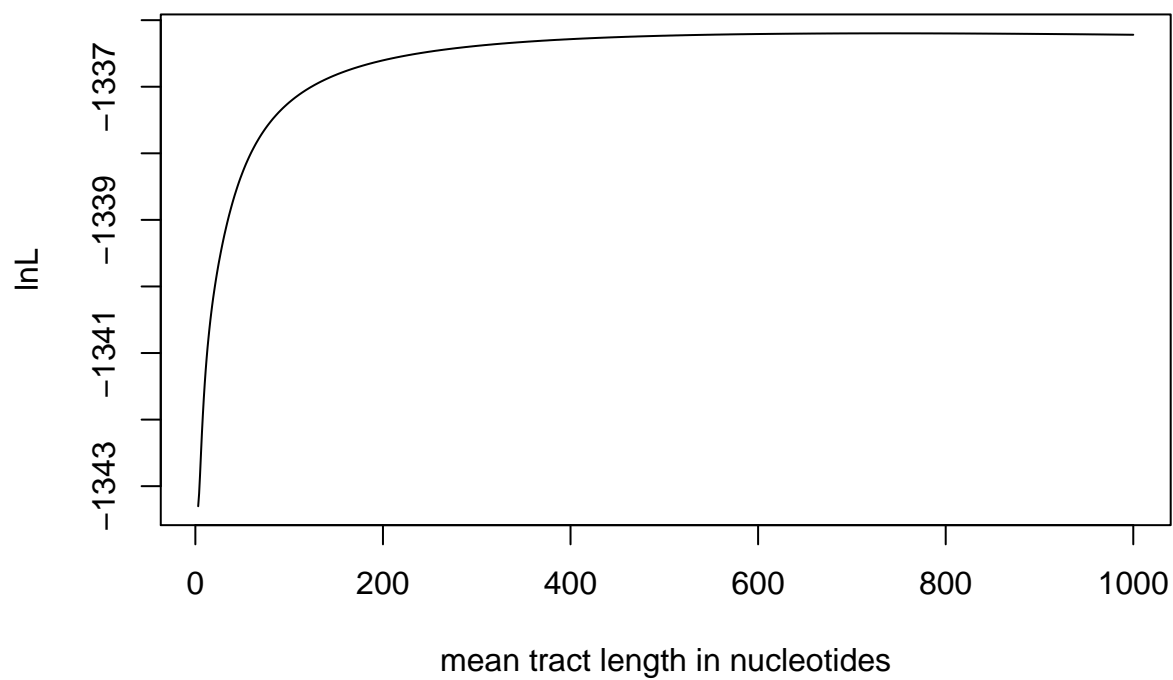
```
## [1] 740.0473
```

YLR333C_YGR027C_lnL_surface

mean tract length in nucleotides
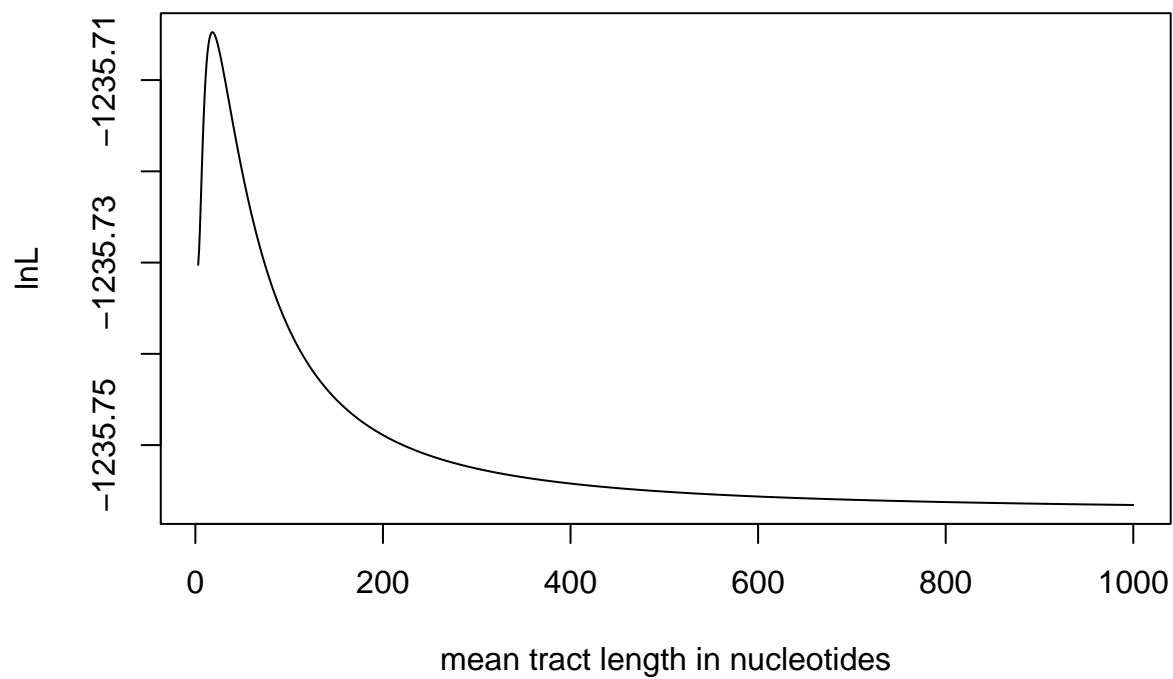
```
## [1] 18.1717
```

YMR142C_YDL082W_lnL_surface
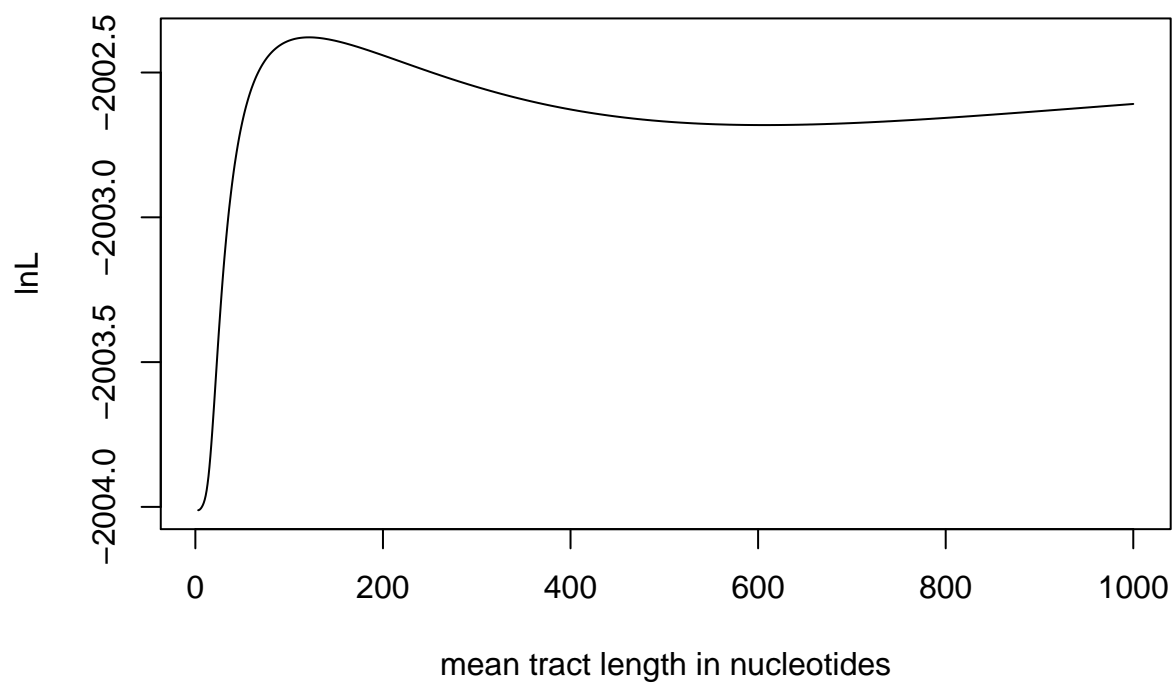

```
## [1] 3.007582
```

YER102W_YBL072C_lnL_surface
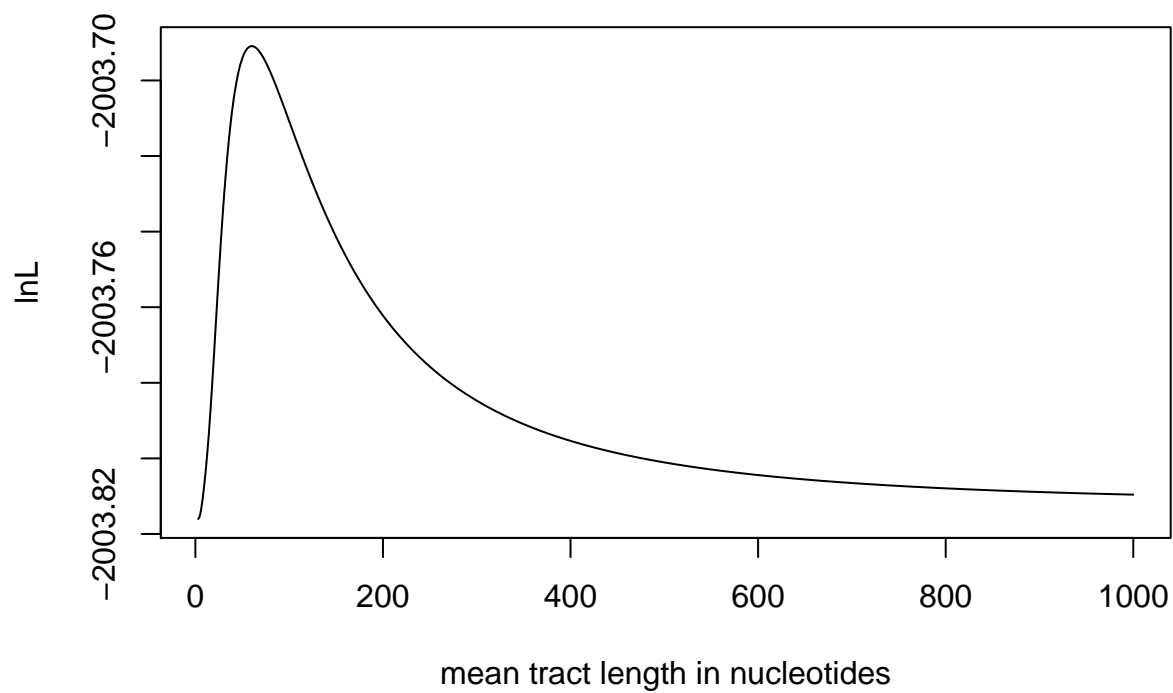

```
## [1] 3.003618
```

Now show the maximum log likelihood difference between search in only tract_p (one dimensional) with all other paramters fixed at their MLE in from the MG94+IS-IGC model and that from the search in all parameters

```
# lnL difference (minus sign because -lnL was recorded)
-(HMM.Yeast.1D["ll", ] - HMM.Yeast.all["ll", ])
```

```
## YLR406C_YDL075W YER131W_YGL189C YML026C_YDR450W YNL301C_YOL120C
##   -9.400649e-02   -7.750812e-03   -2.046363e-11   -2.528403e+00
## YNL069C_YIL133C YMR143W_YDL083C YJL177W_YKL180W YBR191W_YPL079W
##   -7.745803e-08   -2.978142e-05   -1.869835e-01   -2.368623e-01
## YER074W_YIL069C YDR418W_YEL054C YBL087C_YER117W YLR333C_YGR027C
##   -6.207301e-11   -1.234144e+00   -6.494979e-01   -1.334033e-04
## YMR142C_YDL082W YER102W_YBL072C
##    6.821210e-13   -9.333917e-09
```

Now show the estimated average tract length in 1D case

```
3.0/HMM.Yeast.1D["tract_p", ]
```

```
## YLR406C_YDL075W YER131W_YGL189C YML026C_YDR450W YNL301C_YOL120C
##       58.137616       11.122506        3.000000      229.746483
## YNL069C_YIL133C YMR143W_YDL083C YJL177W_YKL180W YBR191W_YPL079W
##        3.000000        6.266257      106.994369       21.085935
## YER074W_YIL069C YDR418W_YEL054C YBL087C_YER117W YLR333C_YGR027C
##        3.000034      102.546616      740.047313       18.171699
## YMR142C_YDL082W YER102W_YBL072C
##        3.007582        3.003618
```

Now show the estimated average tract length in 'all parameter search' case

```
3.0/HMM.Yeast.all["tract_p", ]
```

```
## YLR406C_YDL075W YER131W_YGL189C YML026C_YDR450W YNL301C_YOL120C
##       59.268871       11.161257        3.000000      449.015581
## YNL069C_YIL133C YMR143W_YDL083C YJL177W_YKL180W YBR191W_YPL079W
##        3.000000        6.266280      107.572474       22.015802
## YER074W_YIL069C YDR418W_YEL054C YBL087C_YER117W YLR333C_YGR027C
##        3.000034      143.635872      795.066492       18.171652
## YMR142C_YDL082W YER102W_YBL072C
##        3.007582        3.003618
```