

EDN_ECP_summary

Xiang Ji

January 12, 2016

1, Read in tables

```
rm(list=ls()) # clean up workspace
path <- "/Users/xji3/GitFolders/EDN_ECP/Summary/MG94"

summary.list <- c("_clock_summary",
                  "_nonclock_summary",
                  "_Force_clock_summary",
                  "_Force_nonclock_summary"
                  )
pair = c("EDN", "ECP")
for (target.summary in summary.list){
  summary_file <- paste(path, "_EDN_ECP", target.summary, '.txt', sep = '')
  all <- readLines(summary_file, n = -1)
  row.names <- strsplit(all[length(all)], ' ')[[1]][-1]
  col.name <- paste("MG94", target.summary, sep = "")
  summary_mat <- as.matrix(read.table(summary_file,
                                     row.names = row.names,
                                     col.names = col.name))
  assign(paste("MG94", target.summary, sep = ""), summary_mat)
}
ECP.EDN.MG94 <- cbind(MG94_nonclock_summary, MG94_clock_summary,
                      MG94_Force_nonclock_summary, MG94_Force_clock_summary)
ECP.EDN.MG94
```

##	MG94_nonclock_summary	MG94_clock_summary
## length	1.570000e+02	1.570000e+02
## ll	-1.700519e+03	-1.703650e+03
## pi_a	2.910093e-01	2.916945e-01
## pi_c	2.433912e-01	2.426929e-01
## pi_g	2.069226e-01	2.068703e-01
## pi_t	2.586769e-01	2.587424e-01
## kappa	2.062436e+00	2.089065e+00
## omega	8.270292e-01	8.389169e-01
## tau	6.312271e-01	6.207583e-01
## (N0,N1)	1.990860e-01	1.960748e-01
## (N0,Tamarin)	3.252961e-01	3.268887e-01
## (N1,N2)	3.193865e-02	5.181832e-02
## (N1,Macaque)	1.564817e-01	1.250430e-01
## (N2,N3)	3.455331e-02	5.565978e-02
## (N2,Orangutan)	8.979048e-02	7.322467e-02
## (N3,Chimpanzee)	1.427697e-02	1.756489e-02
## (N3,Gorilla)	1.596838e-02	1.756489e-02
## (N0,N1,tau)	5.202655e-01	5.194297e-01
## (N0,Tamarin,tau)	0.000000e+00	0.000000e+00
## (N1,N2,tau)	4.162337e-01	4.911286e-01
## (N1,Macaque,tau)	3.747596e-01	3.373610e-01

## (N2,N3,tau)	4.485431e-01	4.106192e-01
## (N2,Orangutan,tau)	1.071400e+00	1.231183e+00
## (N3,Chimpanzee,tau)	4.868393e-02	6.296886e-02
## (N3,Gorilla,tau)	4.722111e-01	4.331299e-01
## (N0,N1,1->2)	2.592440e+00	2.515012e+00
## (N0,Tamarin,1->2)	0.000000e+00	0.000000e+00
## (N1,N2,1->2)	3.635547e-01	7.182819e-01
## (N1,Macaque,1->2)	1.512961e+00	1.002534e+00
## (N2,N3,1->2)	5.827584e-01	8.879946e-01
## (N2,Orangutan,1->2)	6.399548e+00	6.069097e+00
## (N3,Chimpanzee,1->2)	6.505905e-02	1.065007e-01
## (N3,Gorilla,1->2)	5.130258e-02	6.176839e-02
## (N0,N1,2->1)	2.592440e+00	2.515012e+00
## (N0,Tamarin,2->1)	0.000000e+00	0.000000e+00
## (N1,N2,2->1)	9.049477e-01	1.701750e+00
## (N1,Macaque,2->1)	5.101368e+00	3.726697e+00
## (N2,N3,2->1)	1.046395e+00	1.519445e+00
## (N2,Orangutan,2->1)	3.413858e+00	3.137466e+00
## (N3,Chimpanzee,2->1)	1.286481e-02	1.756355e-02
## (N3,Gorilla,2->1)	7.941026e-01	7.916351e-01
## (N0,N1,mut)	6.295535e+01	6.198371e+01
## (N0,Tamarin,mut)	5.115805e+01	5.141753e+01
## (N1,N2,mut)	1.045748e+01	1.399356e+01
## (N1,Macaque,mut)	5.231267e+01	4.851329e+01
## (N2,N3,mut)	1.122474e+01	1.245078e+01
## (N2,Orangutan,mut)	2.350807e+01	2.198678e+01
## (N3,Chimpanzee,mut)	5.280903e+00	5.203542e+00
## (N3,Gorilla,mut)	5.147585e+00	5.162176e+00
##	MG94_Force_nonclock_summary	MG94_Force_clock_summary
## length	1.570000e+02	1.570000e+02
## ll	-1.714099e+03	-1.716567e+03
## pi_a	2.927431e-01	2.927869e-01
## pi_c	2.425981e-01	2.421026e-01
## pi_g	2.076225e-01	2.078204e-01
## pi_t	2.570363e-01	2.572901e-01
## kappa	2.100482e+00	2.102322e+00
## omega	9.044276e-01	9.065773e-01
## tau	0.000000e+00	0.000000e+00
## (N0,N1)	1.440343e-01	1.400382e-01
## (N0,Tamarin)	3.556100e-01	3.579323e-01
## (N1,N2)	4.519787e-02	6.042200e-02
## (N1,Macaque)	1.777791e-01	1.480029e-01
## (N2,N3)	4.510421e-02	6.707861e-02
## (N2,Orangutan)	9.981750e-02	8.758086e-02
## (N3,Chimpanzee)	1.699996e-02	2.050224e-02
## (N3,Gorilla)	1.880088e-02	2.050224e-02
## (N0,N1,tau)	0.000000e+00	0.000000e+00
## (N0,Tamarin,tau)	0.000000e+00	0.000000e+00
## (N1,N2,tau)	0.000000e+00	0.000000e+00
## (N1,Macaque,tau)	0.000000e+00	0.000000e+00
## (N2,N3,tau)	0.000000e+00	0.000000e+00
## (N2,Orangutan,tau)	0.000000e+00	0.000000e+00
## (N3,Chimpanzee,tau)	0.000000e+00	0.000000e+00
## (N3,Gorilla,tau)	0.000000e+00	0.000000e+00

## (N0,N1,1->2)	0.000000e+00	0.000000e+00
## (N0,Tamarin,1->2)	0.000000e+00	0.000000e+00
## (N1,N2,1->2)	0.000000e+00	0.000000e+00
## (N1,Macaque,1->2)	0.000000e+00	0.000000e+00
## (N2,N3,1->2)	0.000000e+00	0.000000e+00
## (N2,Orangutan,1->2)	0.000000e+00	0.000000e+00
## (N3,Chimpanzee,1->2)	0.000000e+00	0.000000e+00
## (N3,Gorilla,1->2)	0.000000e+00	0.000000e+00
## (N0,N1,2->1)	0.000000e+00	0.000000e+00
## (N0,Tamarin,2->1)	0.000000e+00	0.000000e+00
## (N1,N2,2->1)	0.000000e+00	0.000000e+00
## (N1,Macaque,2->1)	0.000000e+00	0.000000e+00
## (N2,N3,2->1)	0.000000e+00	0.000000e+00
## (N2,Orangutan,2->1)	0.000000e+00	0.000000e+00
## (N3,Chimpanzee,2->1)	0.000000e+00	0.000000e+00
## (N3,Gorilla,2->1)	0.000000e+00	0.000000e+00
## (N0,N1,mut)	4.544445e+01	4.418651e+01
## (N0,Tamarin,mut)	5.598401e+01	5.635691e+01
## (N1,N2,mut)	1.427259e+01	1.584501e+01
## (N1,Macaque,mut)	5.608860e+01	5.462142e+01
## (N2,N3,mut)	1.422474e+01	1.573075e+01
## (N2,Orangutan,mut)	3.153865e+01	3.006394e+01
## (N3,Chimpanzee,mut)	5.352806e+00	5.299775e+00
## (N3,Gorilla,mut)	5.923543e+00	5.952135e+00

2, Now show branch specific % changes due to IGC

`(ECP.EDN.MG94[26:33,] + ECP.EDN.MG94[34:41,])/(ECP.EDN.MG94[42:49,] + ECP.EDN.MG94[26:33,] + ECP.EDN.MG94[34:41,])`

##	MG94_nonclock_summary	MG94_clock_summary
## (N0,N1,1->2)	0.07609132	0.07505960
## (N0,Tamarin,1->2)	0.00000000	0.00000000
## (N1,N2,1->2)	0.10817880	0.14744073
## (N1,Macaque,1->2)	0.11224614	0.08882434
## (N2,N3,1->2)	0.12674399	0.16202747
## (N2,Orangutan,1->2)	0.29450698	0.29514515
## (N3,Chimpanzee,1->2)	0.01454122	0.02328706
## (N3,Gorilla,1->2)	0.14106566	0.14186554
##	MG94_Force_nonclock_summary	MG94_Force_clock_summary
## (N0,N1,1->2)	0	0
## (N0,Tamarin,1->2)	0	0
## (N1,N2,1->2)	0	0
## (N1,Macaque,1->2)	0	0
## (N2,N3,1->2)	0	0
## (N2,Orangutan,1->2)	0	0
## (N3,Chimpanzee,1->2)	0	0
## (N3,Gorilla,1->2)	0	0

3, % changes due to IGC in all branches

`colSums(ECP.EDN.MG94[26:33,] + ECP.EDN.MG94[34:41,])/colSums(ECP.EDN.MG94[42:49,] + ECP.EDN.MG94[26:33,] + ECP.EDN.MG94[34:41,])`

##	MG94_nonclock_summary	MG94_clock_summary
----	-----------------------	--------------------

##	0.1027710	0.1009066
## MG94_Force_nonclock_summary		MG94_Force_clock_summary
##	0.0000000	0.0000000

04212017 update

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

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

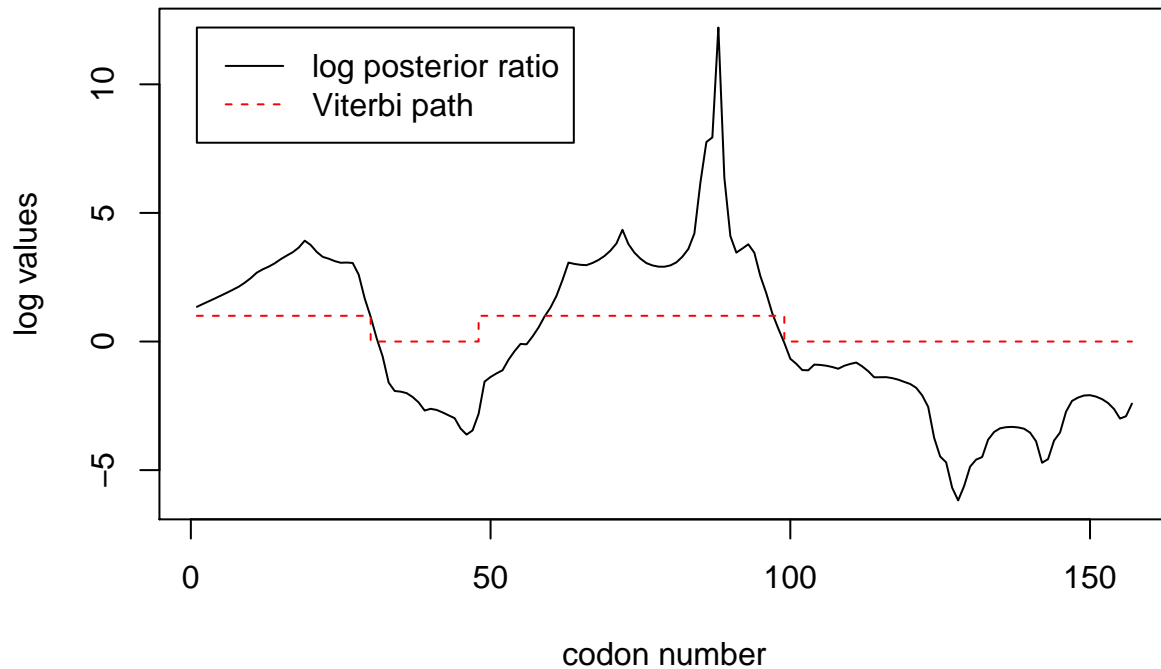
95% confidence interval for $\ln(p_{tract})$ is $\ln(p_{tract}) \pm 1.96 * \sqrt{Var(\ln(p_{tract}))}$

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

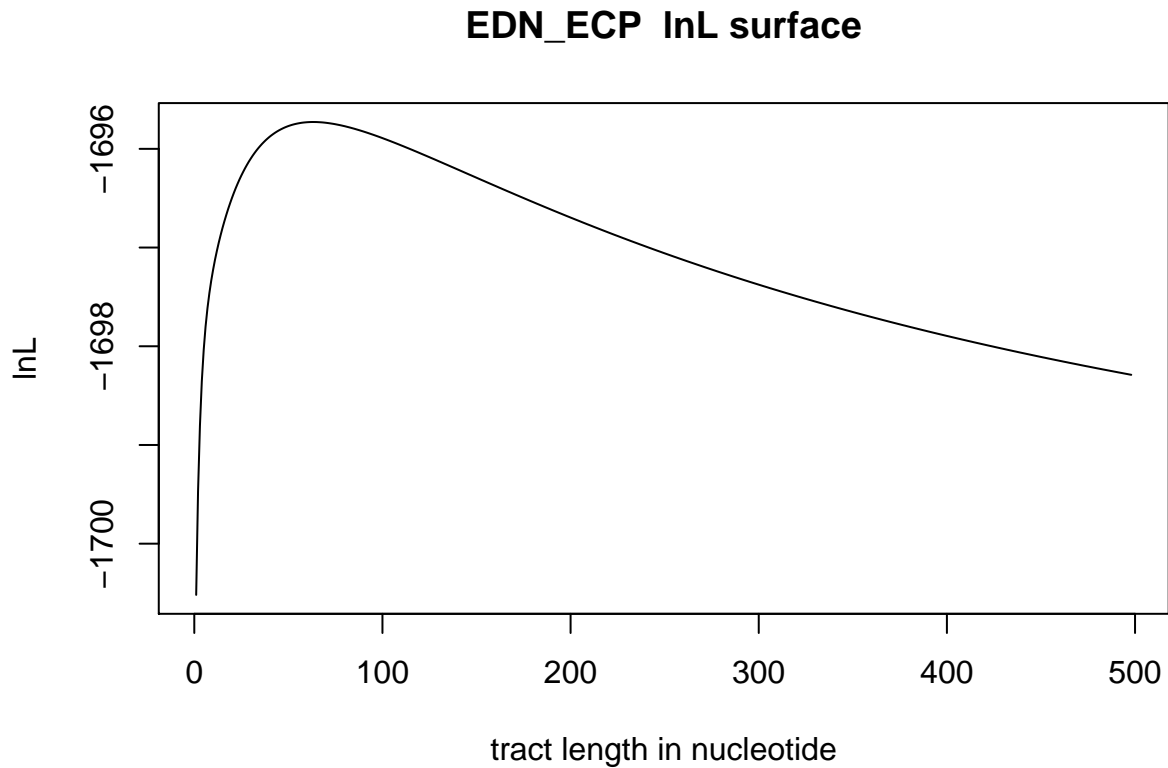
```
# plot one paralog
paralog = "EDN_ECP"
lnL.ratio <- as.vector(read.table(paste("./summary/", paralog, "_MG94_nonclock_HMM_log_posterior_ratio.txt"),
Viterbi.path <- as.vector(read.table(paste("./summary/", paralog, "_MG94_nonclock_HMM_Viterbi_path.txt"),
lnL.surface <- as.vector(read.table(paste("./summary/", 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"))
```

EDN_ECP HMM result



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



```
summary.mat <- read.table("./HMM_tract_MG94_nonclock_summary.txt")

# 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)
rownames(summary.mat) <- c("EDN_ECP")
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")

summary.mat
```

```
##          lnL    max lnL tract length  Pr(S_0)  Pr(S_1)          df
## EDN_ECP -1700.519 -1695.728    65.48743 0.5043814 0.4956186 4.592165e-06
##          d^2f c.i. tract length c.i tract length
## EDN_ECP -1.631126          14.11443          303.8453
```