Deficit of cysteine in mtDNA of longlived mammals - selection or mutagenesis

Konstantin Popadin

7 June 2022

1. Read AA content, merge it with generation length of mammals (N = 650) and derive frequencies of some amino-acids for each of 13 proteins

in comparative species analyses it is possible to work just with real numbers of amino acids - withount any frequencies

```
rm(list=ls(all=TRUE))

AA = read.csv('../../Body/3Results/AminoAcidFreqsChordata.csv')
names(AA)
```

```
##  [1] "Species" "Gene"    "Ala"     "Arg"     "Asn"     "Asp"     "Cys"
##  [8] "Gln"     "Glu"     "Gly"     "His"     "Ile"     "Leu"     "Lys"
## [15] "Met"     "Phe"     "Pro"     "Ser"     "Thr"     "Trp"     "Tyr"
## [22] "Val"     "U"       "O"       "B"       "J"       "Z"       "X"
## [29] "X."      "X..1"    "X..2"    "."       "other"   "Class"
```

```
table(AA$Class)
```

```
##
##                             0 Actinopterygii      Amphibia    AncientFish
##             13                13          22997          2665           1638
##           Aves         Mammalia       Reptilia
##           5616            10244           3497
```

```
AA = AA[AA$Class == 'Mammalia',]

GenTime = read.table('../../Body/1Raw/GenerationLenghtforMammals.xlsx.txt', sep = '\t', header = TRUE)
GenTime$Species = gsub(' ','_',GenTime$Scientific_name)
GenTime = GenTime[colnames(GenTime) %in% c('Species','Genus','GenerationLength_d')]

dim(AA)
```

```
## [1] 10244    34
```

```
AA = merge(AA,GenTime)
dim(AA)
```

```
## [1] 8450    36
```

```
AA$CysFr = 0
AA$ArgFr = 0
```

```r
AA$HisFr = 0
names(AA)
```

```
##  [1] "Species"           "Gene"              "Ala"
##  [4] "Arg"               "Asn"               "Asp"
##  [7] "Cys"               "Gln"               "Glu"
## [10] "Gly"               "His"               "Ile"
## [13] "Leu"               "Lys"               "Met"
## [16] "Phe"               "Pro"               "Ser"
## [19] "Thr"               "Trp"               "Tyr"
## [22] "Val"               "U"                 "O"
## [25] "B"                 "J"                 "Z"
## [28] "X"                 "X."                "X..1"
## [31] "X..2"              "."                 "other"
## [34] "Class"             "Genus"             "GenerationLength_d"
## [37] "CysFr"             "ArgFr"             "HisFr"
```

```r
for (i in 1:nrow(AA))
{ # i = 2
AA$CysFr[i] = AA$Cys[i]/sum(as.numeric(AA[i,3:22]))  # 1/115
AA$ArgFr[i] = AA$Arg[i]/sum(as.numeric(AA[i,3:22]))  # 1/115
AA$HisFr[i] = AA$His[i]/sum(as.numeric(AA[i,3:22]))  # 1/115
}
summary(AA$CysFr)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## 0.000000 0.001946 0.006667 0.007173 0.008811 0.037736
```

```r
summary(AA$ArgFr)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## 0.000000 0.008772 0.015564 0.015542 0.021930 0.038462
```

```r
summary(AA$HisFr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.01258 0.02381 0.02416 0.03307 0.07692
```

```r
nrow(AA)/13
```

```
## [1] 650
```

2. compare Cys frequency across genes.

2A: Cys is a loser on light chain but should be gainer on heavy => higher expected frequency in ND6. Indeed ND6 is the second top gene with maximal CysFr.

2B: Probably CysFr among 12 genes, coded on light chain, is the lowest in low-constrained genes (loser should go to zero unless there is selection): we can find dn/ds for each mtDNA gene, but from my experience it seems that ATP (6 and 8) are low constrained and exactly they have zero Cys. Need to continue this analysis


```r
AGG = aggregate(AA$CysFr,by = list(AA$Gene), FUN = mean)
names(AGG) = c('Gene','MeanCysFr')
```

```
AGG = AGG[order(AGG$MeanCysFr),]
AGG
```

```
##    Gene     MeanCysFr
## 1  ATP6 0.0003696879
## 8   ND2 0.0008446867
## 2  ATP8 0.0016966041
## 3  COX1 0.0020422143
## 7   ND1 0.0040020541
## 5  COX3 0.0070248765
## 10  ND4 0.0075319192
## 12  ND5 0.0081226613
## 9   ND3 0.0092158600
## 4  COX2 0.0095727329
## 6  CytB 0.0095985779
## 13  ND6 0.0119403233
## 11 ND4L 0.0212871122
```
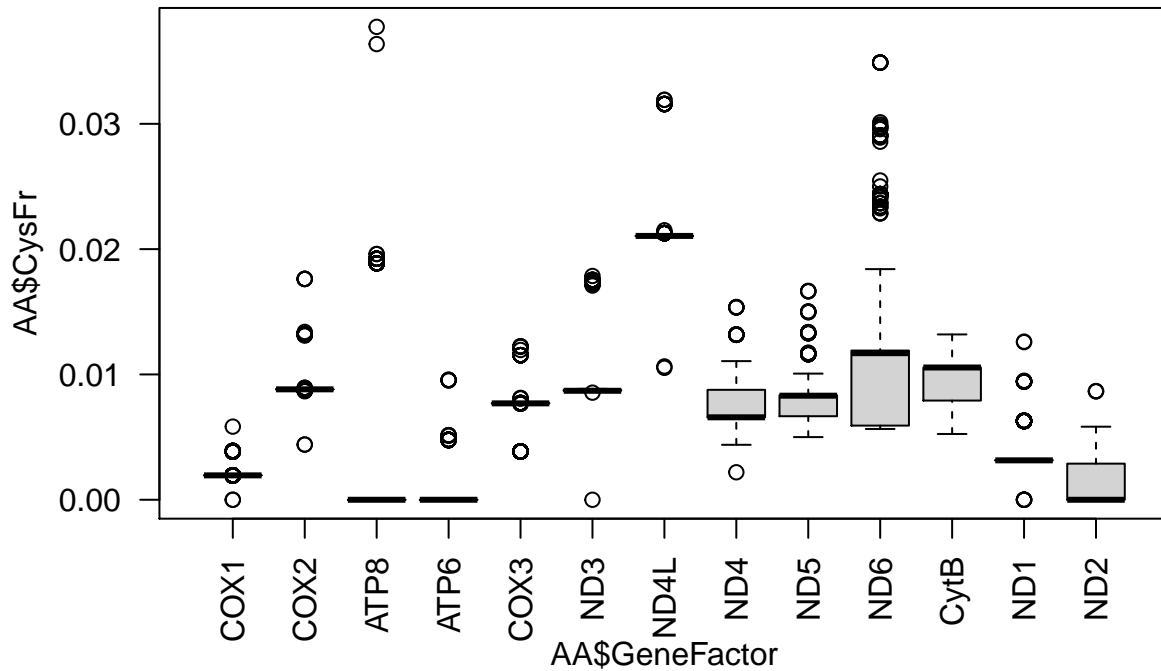
```
boxplot(AA$CysFr ~ AA$Gene, las = 2)
```
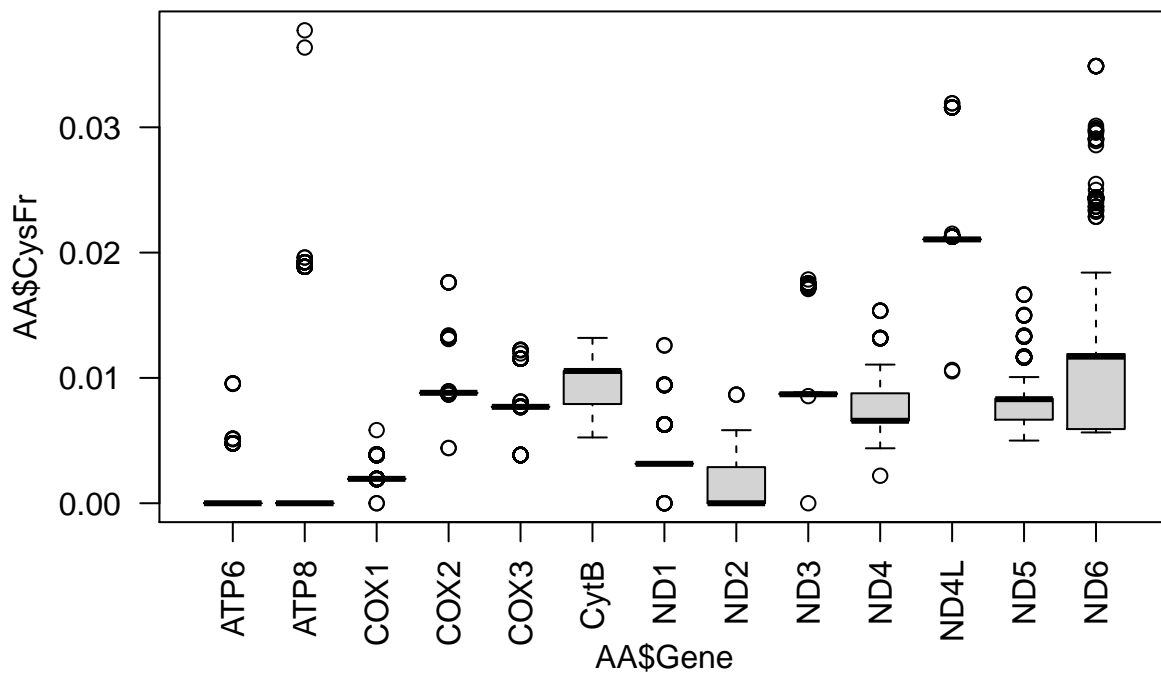


```
AA$GeneFactor <- factor(AA$Gene, levels=c("COX1", "COX2", "ATP8", "ATP6","COX3","ND3","ND4L","ND4","ND5
table(AA$GeneFactor)
```

```
##
## COX1 COX2 ATP8 ATP6 COX3  ND3 ND4L  ND4  ND5  ND6 CytB  ND1  ND2
##  650  650  650  650  650  650  650  650  650  650  650  650  650
```

```r
boxplot(AA$CysFr ~ AA$GeneFactor, las = 2)  # boxplot arranged according to TBSS (from)
```



```r
# only several genes show some variation in CysFr
```

3. compare correlation of CysFr with generation length for each gene

why there is a variation in the strength of the correlations?

ND6 shows the best negative correlation with GT - why? (if MutSpec ~ increased Ah>Gh in long-lived ~ the main factor, we expect positive correlation with GT, something is not so simple)

```r
a = cor.test(AA[AA$Gene == 'COX1',]$CysFr,AA[AA$Gene == 'COX1',]$GenerationLength_d); a$p.value; a$esti
```

```
## [1] 0.9821301
```

```
##           cor
## -0.0008802327
```

```r
cor.test(AA[AA$Gene == 'COX2',]$CysFr,AA[AA$Gene == 'COX2',]$GenerationLength_d)
```

```
##
##  Pearson's product-moment correlation
##
```

```
## data:  AA[AA$Gene == "COX2", ]$CysFr and AA[AA$Gene == "COX2", ]$GenerationLength_d
## t = -5.6438, df = 648, p-value = 2.489e-08
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2885504 -0.1419117
## sample estimates:
##        cor
## -0.2164515
```

```r
cor.test(AA[AA$Gene == 'ATP8',]$CysFr,AA[AA$Gene == 'ATP8',]$GenerationLength_d)
```

```
##
##  Pearson's product-moment correlation
##
## data:  AA[AA$Gene == "ATP8", ]$CysFr and AA[AA$Gene == "ATP8", ]$GenerationLength_d
## t = -1.216, df = 648, p-value = 0.2244
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.12415990  0.02929602
## sample estimates:
##         cor
## -0.04771347
```

```r
cor.test(AA[AA$Gene == 'ATP6',]$CysFr,AA[AA$Gene == 'ATP6',]$GenerationLength_d)
```

```
##
##  Pearson's product-moment correlation
##
## data:  AA[AA$Gene == "ATP6", ]$CysFr and AA[AA$Gene == "ATP6", ]$GenerationLength_d
## t = -3.5302, df = 648, p-value = 0.0004446
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.21202744 -0.06110876
## sample estimates:
##        cor
## -0.1373652
```

```r
cor.test(AA[AA$Gene == 'COX3',]$CysFr,AA[AA$Gene == 'COX3',]$GenerationLength_d)
```

```
##
##  Pearson's product-moment correlation
##
## data:  AA[AA$Gene == "COX3", ]$CysFr and AA[AA$Gene == "COX3", ]$GenerationLength_d
## t = -4.0183, df = 648, p-value = 6.55e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.23006624 -0.07997994
## sample estimates:
##        cor
## -0.1559229
```

```r
cor.test(AA[AA$Gene == 'ND3',]$CysFr,AA[AA$Gene == 'ND3',]$GenerationLength_d)
```

```
##
##  Pearson's product-moment correlation
##
## data:  AA[AA$Gene == "ND3", ]$CysFr and AA[AA$Gene == "ND3", ]$GenerationLength_d
```

```
## t = -3.074, df = 648, p-value = 0.002201
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.19499020 -0.04338389
## sample estimates:
##        cor
## -0.1198859
```

```r
cor.test(AA[AA$Gene == 'ND4L',]$CysFr,AA[AA$Gene == 'ND4L',]$GenerationLength_d)
```

```
##
##  Pearson's product-moment correlation
##
## data:  AA[AA$Gene == "ND4L", ]$CysFr and AA[AA$Gene == "ND4L", ]$GenerationLength_d
## t = -0.36989, df = 648, p-value = 0.7116
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.09132923  0.06244249
## sample estimates:
##         cor
## -0.01452928
```

```r
cor.test(AA[AA$Gene == 'ND4',]$CysFr,AA[AA$Gene == 'ND4',]$GenerationLength_d)
```

```
##
##  Pearson's product-moment correlation
##
## data:  AA[AA$Gene == "ND4", ]$CysFr and AA[AA$Gene == "ND4", ]$GenerationLength_d
## t = -2.5052, df = 648, p-value = 0.01248
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.17353635 -0.02119908
## sample estimates:
##         cor
## -0.09794141
```

```r
cor.test(AA[AA$Gene == 'ND5',]$CysFr,AA[AA$Gene == 'ND5',]$GenerationLength_d)
```

```
##
##  Pearson's product-moment correlation
##
## data:  AA[AA$Gene == "ND5", ]$CysFr and AA[AA$Gene == "ND5", ]$GenerationLength_d
## t = -4.1531, df = 648, p-value = 3.72e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.23501089 -0.08517162
## sample estimates:
##        cor
## -0.161019
```

```r
cor.test(AA[AA$Gene == 'ND1',]$CysFr,AA[AA$Gene == 'ND1',]$GenerationLength_d)
```

```
##
##  Pearson's product-moment correlation
##
## data:  AA[AA$Gene == "ND1", ]$CysFr and AA[AA$Gene == "ND1", ]$GenerationLength_d
## t = 0.92931, df = 648, p-value = 0.3531
```

```
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.04053342  0.11306707
## sample estimates:
##        cor
## 0.03648229
```

```r
cor.test(AA[AA$Gene == 'ND2',]$CysFr,AA[AA$Gene == 'ND2',]$GenerationLength_d)
```

```
##
##  Pearson's product-moment correlation
##
## data:  AA[AA$Gene == "ND2", ]$CysFr and AA[AA$Gene == "ND2", ]$GenerationLength_d
## t = -4.2179, df = 648, p-value = 2.817e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.23738416 -0.08766633
## sample estimates:
##        cor
## -0.1634663
```

```r
cor.test(AA[AA$Gene == 'ND6',]$CysFr,AA[AA$Gene == 'ND6',]$GenerationLength_d)  # opposite chain!! did i
```

```
##
##  Pearson's product-moment correlation
##
## data:  AA[AA$Gene == "ND6", ]$CysFr and AA[AA$Gene == "ND6", ]$GenerationLength_d
## t = -8.7112, df = 648, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.3909417 -0.2531755
## sample estimates:
##        cor
## -0.3237737
```

```r
cor.test(AA[AA$Gene == 'CytB',]$CysFr,AA[AA$Gene == 'CytB',]$GenerationLength_d)
```

```
##
##  Pearson's product-moment correlation
##
## data:  AA[AA$Gene == "CytB", ]$CysFr and AA[AA$Gene == "CytB", ]$GenerationLength_d
## t = -6.272, df = 648, p-value = 6.515e-10
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.3104233 -0.1653727
## sample estimates:
##        cor
## -0.2392323
```

```r
plot(AA[AA$Gene == 'ND6',]$CysFr,log2(AA[AA$Gene == 'ND6',]$GenerationLength_d))
```
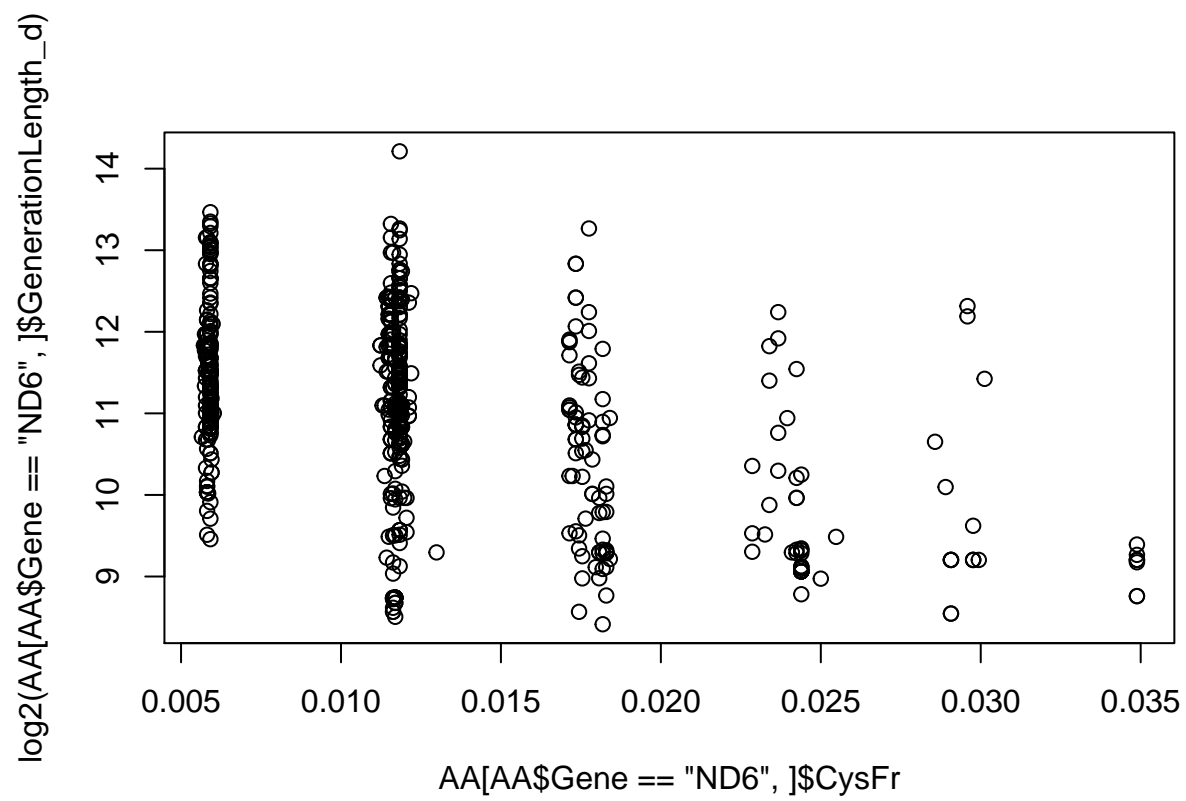
```r
plot(AA[AA$Gene == 'ND6',]$Cys,log2(AA[AA$Gene == 'ND6',]$GenerationLength_d))
```
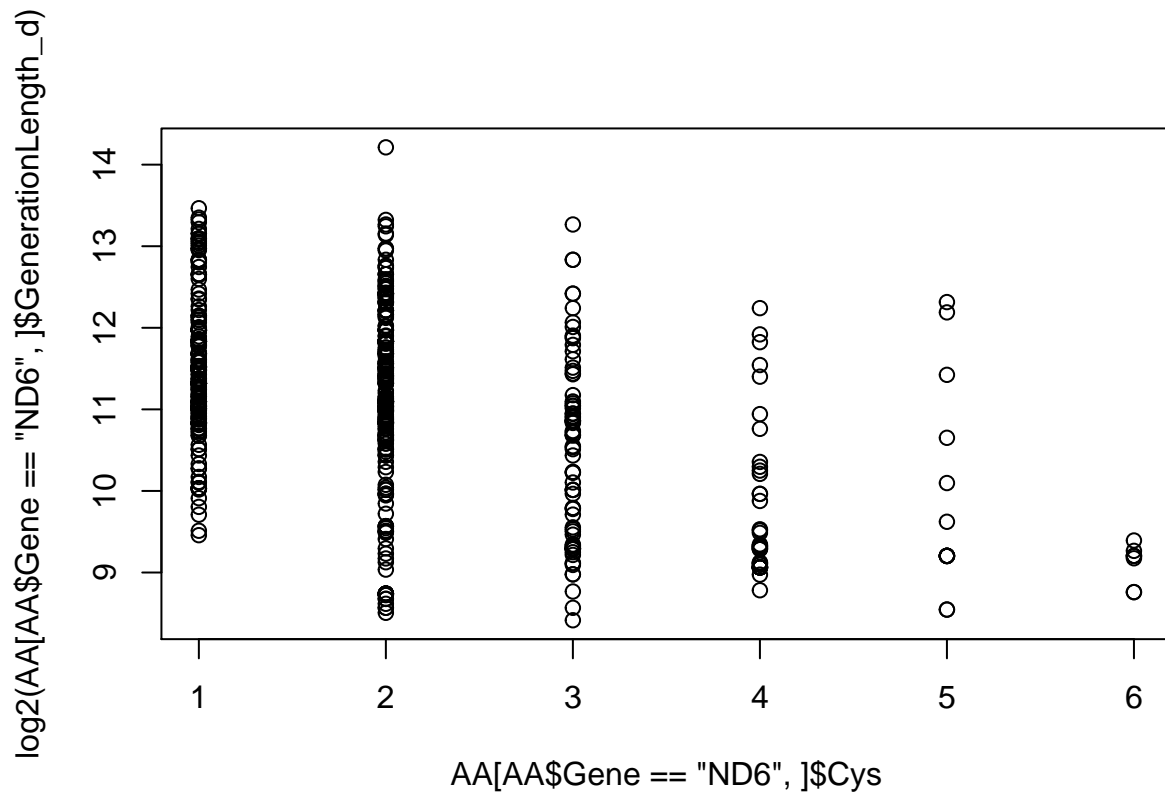
4. compare Cys frequency among light chain genes and ND6 - it is positive! unusual!

```
AGG2 = aggregate(AA[AA$Gene != 'ND6',]$CysFr, by = list(AA[AA$Gene != 'ND6',]$Species), FUN = mean)
names(AGG2) = c('Species','CysFr12genes')
AGG3 = AA[AA$Gene == 'ND6',];
AGG3= AGG3[colnames(AGG3) %in% c('Species','CysFr','GenerationLength_d')]
AGG4 = merge(AGG2,AGG3, by = 'Species')
names(AGG4)[4]='CysFrNd6'
str(AGG4)
```

```
## 'data.frame':    650 obs. of  4 variables:
##  $ Species          : chr  "Acinonyx_jubatus" "Acomys_cahirinus" "Addax_nasomaculatus" "Aepyceros_me
##  $ CysFr12genes     : num  0.0065 0.00656 0.0064 0.0064 0.00678 ...
##  $ GenerationLength_d: num  2190 650 2555 2674 5247 ...
##  $ CysFrNd6         : num  0.01183 0.02439 0.00592 0.01183 0.01212 ...
```

```
cor.test(AGG4$CysFr12genes, AGG4$CysFrNd6, method = 'spearman') # positive and significant. unusual.
```

```
## Warning in cor.test.default(AGG4$CysFr12genes, AGG4$CysFrNd6, method =
## "spearman"): Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AGG4$CysFr12genes and AGG4$CysFrNd6
## S = 26688729, p-value < 2.2e-16
```

```
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.416904
```

5. I am not sure that for ND6 we used correct aminoacids (reverse compliment from GenBank) - test it in two ways (from another dataset and GenBank): IT IS CORRECT!!!

```r
Test = read.table('../../Body/3Results/AllGenesCodonUsageNoOverlap.txt', sep = '\t', header = TRUE)
Test1 = Test[Test$Gene == 'ND6',]
Test1 = Test1[names(Test1) %in% c('Species','AminoNoOverlap')]

Test3 = Test1[Test1$Species == 'Acinonyx_jubatus',]
table(unlist(strsplit(Test3$AminoNoOverlap,''))) # it is the same as in AA data frame for the same spec
```

```
##
## A  C  D  E  F  G  I  K  L  M  N  P  R  S  T  V  W  Y
## 11  2  3 10 14 24 13  4 18  6  4  4  1 12 11 20  4  8
```

```r
# from web page https://www.ncbi.nlm.nih.gov/protein/NP_941389.1
Nd6AcinonyxJubatus = 'mmtyivfils tifvvsfvsf sskpspiygg fglivaggig cgivlnfggs flglmvfliy lggmlvvfgy ttama
table(unlist(strsplit(Nd6AcinonyxJubatus,''))) # it is the same with little deviations (probably becaus
```

```
##
##  a  c  d  e  f  g  i  k  l  m  n  p  r  s  t  v  w  y
## 15 11  2  3 10 14 24 14  4 18  8  4  4  1 12 12 21  4  9
```

6. epistatic trajectories: Cys is in the corner of the genetic code (see the butterfly scheme here: https://docs.google.com/presentation/d/1KsieoHhePhRG5m1Bg32j9p6jLH7qkeukW1 jtuZcBRy8/edit?usp=sharing)

its neighbors are rare and unusual: **Arg (CGX) is highly mutated and goes to His+Gln unless it is very constrained.**

its neighbors are rare and unusual: **A stop codon from other cite!**

In ND6 the rarest amino acid is Arg (R) and the second most rare is Cys (C)

where was ND6 in invertebrates? what was his starting point in the evolution?

```r
cor.test(AA$CysFr,AA$ArgFr, method = 'spearman') # positive but weak
```

```
## Warning in cor.test.default(AA$CysFr, AA$ArgFr, method = "spearman"): Cannot
## compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA$CysFr and AA$ArgFr
## S = 9.11e+10, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
```

```
##        rho
## 0.09406347
```

```r
cor.test(AA[AA$Gene == 'ND6',]$CysFr,AA[AA$Gene == 'ND6',]$ArgFr, method = 'spearman') # NO CORR
```

```
## Warning in cor.test.default(AA[AA$Gene == "ND6", ]$CysFr, AA[AA$Gene == : Cannot
## compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "ND6", ]$CysFr and AA[AA$Gene == "ND6", ]$ArgFr
## S = 46595244, p-value = 0.6466
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##         rho
## -0.01801411
```

```r
cor.test(AA[AA$Gene != 'ND6',]$CysFr,AA[AA$Gene != 'ND6',]$ArgFr, method = 'spearman') # GOOD POSITIVE
```

```
## Warning in cor.test.default(AA[AA$Gene != "ND6", ]$CysFr, AA[AA$Gene != : Cannot
## compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene != "ND6", ]$CysFr and AA[AA$Gene != "ND6", ]$ArgFr
## S = 6.4779e+10, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## 0.1809676
```

```r
### Cys is rarer than Arg among 12 genes (as expected according to the MutSpec). Cys < Arg, Cys < His
summary(AA[AA$Gene != 'ND6',]$CysFr)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## 0.000000 0.001946 0.006667 0.006776 0.008811 0.037736
```

```r
summary(AA[AA$Gene != 'ND6',]$ArgFr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.01053 0.01751 0.01613 0.02193 0.03846
```

```r
summary(AA[AA$Gene != 'ND6',]$HisFr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.01577 0.02632 0.02611 0.03333 0.07692
```

```r
### But Cys is more common than Arg in ND6 (as expected according to the MutSpec). Cys > Arg, Cys > His
summary(AA[AA$Gene == 'ND6',]$CysFr)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## 0.005650 0.005917 0.011696 0.011940 0.011905 0.034884
```

```r
summary(AA[AA$Gene == 'ND6',]$ArgFr)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## 0.000000 0.005917 0.005917 0.008491 0.011628 0.023121
```

```
summary(AA[AA$Gene == 'ND6',]$HisFr)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000000 0.0000000 0.0000000 0.0008128 0.0000000 0.0172414
```

7. gainers and losers altogether stronger describe correlation with Generation Time as compared to Cys (see example with CytB).

so - can we say that Cys is somehow special among other losers?

Cys and Trp sit together and both of them unusual (in Cytb Trp is positive with GT, Cys is weakly negative)

```
# Gainers versus losers
AA$NumberOfGainers = AA$Pro + AA$His + AA$Gln + AA$Asn + AA$Lys + AA$Thr
summary(AA$NumberOfGainers)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.00   25.00   68.00   72.83  109.00  197.00
```

```
AA$NumberOfLosers = AA$Phe + AA$Cys + AA$Trp + AA$Gly + AA$Val
summary(AA$NumberOfLosers)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.00   31.00   56.00   57.55   74.00  151.00
```

```
cor.test(AA[AA$Gene == 'ND6',]$NumberOfGainers,AA[AA$Gene == 'ND6',]$NumberOfLosers, method = 'spearman
```

```
## Warning in cor.test.default(AA[AA$Gene == "ND6", ]$NumberOfGainers, AA[AA$Gene
## == : Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "ND6", ]$NumberOfGainers and AA[AA$Gene == "ND6", ]$NumberOfLosers
## S = 61207752, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.3372686
```

```
cor.test(AA[AA$Gene == 'COX1',]$NumberOfGainers,AA[AA$Gene == 'COX1',]$NumberOfLosers, method = 'spearma
```

```
## Warning in cor.test.default(AA[AA$Gene == "COX1", ]$NumberOfGainers, AA[AA$Gene
## == : Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "COX1", ]$NumberOfGainers and AA[AA$Gene == "COX1", ]$NumberOfLosers
## S = 55460849, p-value = 5.064e-08
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.2117101
```

```r
cor.test(AA[AA$Gene == 'CytB',]$NumberOfGainers,AA[AA$Gene == 'CytB',]$NumberOfLosers, method = 'spearm
```

```
## Warning in cor.test.default(AA[AA$Gene == "CytB", ]$NumberOfGainers, AA[AA$Gene
## == : Cannot compute exact p-value with ties

##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "CytB", ]$NumberOfGainers and AA[AA$Gene == "CytB", ]$NumberOfLosers
## S = 70703097, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.5447231
```

```r
cor.test(AA[AA$Gene == 'ND1',]$NumberOfGainers,AA[AA$Gene == 'ND1',]$NumberOfLosers, method = 'spearman
```

```
## Warning in cor.test.default(AA[AA$Gene == "ND1", ]$NumberOfGainers, AA[AA$Gene
## == : Cannot compute exact p-value with ties

##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "ND1", ]$NumberOfGainers and AA[AA$Gene == "ND1", ]$NumberOfLosers
## S = 66244556, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.4473128
```

```r
### try CytB with generation length
cor.test(AA[AA$Gene == 'CytB',]$NumberOfGainers,AA[AA$Gene == 'CytB',]$GenerationLength_d, method = 'sp
```

```
## Warning in cor.test.default(AA[AA$Gene == "CytB", ]$NumberOfGainers, AA[AA$Gene
## == : Cannot compute exact p-value with ties

##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "CytB", ]$NumberOfGainers and AA[AA$Gene == "CytB", ]$GenerationLength_d
## S = 26205063, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.4274711
```

```r
cor.test(AA[AA$Gene == 'CytB',]$NumberOfLosers,AA[AA$Gene == 'CytB',]$GenerationLength_d, method = 'spea
```

```
## Warning in cor.test.default(AA[AA$Gene == "CytB", ]$NumberOfLosers, AA[AA$Gene
## == : Cannot compute exact p-value with ties

##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "CytB", ]$NumberOfLosers and AA[AA$Gene == "CytB", ]$GenerationLength_d
## S = 69276940, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
```

```
##        rho
## -0.5135644
```

```r
cor.test(AA[AA$Gene == 'CytB',]$Cys,AA[AA$Gene == 'CytB',]$GenerationLength_d, method = 'spearman')  #
```

```
## Warning in cor.test.default(AA[AA$Gene == "CytB", ]$Cys, AA[AA$Gene == "CytB", :
## Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "CytB", ]$Cys and AA[AA$Gene == "CytB", ]$GenerationLength_d
## S = 52130689, p-value = 0.0003808
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.1389527
```

```r
cor.test(AA[AA$Gene == 'CytB',]$Trp,AA[AA$Gene == 'CytB',]$GenerationLength_d, method = 'spearman')  #
```

```
## Warning in cor.test.default(AA[AA$Gene == "CytB", ]$Trp, AA[AA$Gene == "CytB", :
## Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "CytB", ]$Trp and AA[AA$Gene == "CytB", ]$GenerationLength_d
## S = 37471028, p-value = 3.275e-06
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.181332
```

```r
cor.test(AA[AA$Gene == 'CytB',]$Phe,AA[AA$Gene == 'CytB',]$GenerationLength_d, method = 'spearman')  #
```

```
## Warning in cor.test.default(AA[AA$Gene == "CytB", ]$Phe, AA[AA$Gene == "CytB", :
## Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "CytB", ]$Phe and AA[AA$Gene == "CytB", ]$GenerationLength_d
## S = 67251499, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.4693125
```

```r
cor.test(AA[AA$Gene == 'CytB',]$Gly,AA[AA$Gene == 'CytB',]$GenerationLength_d, method = 'spearman')  #
```

```
## Warning in cor.test.default(AA[AA$Gene == "CytB", ]$Gly, AA[AA$Gene == "CytB", :
## Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "CytB", ]$Gly and AA[AA$Gene == "CytB", ]$GenerationLength_d
## S = 61725725, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
```

```
## sample estimates:
##        rho
## -0.3485853
```

```r
cor.test(AA[AA$Gene == 'CytB',]$Val,AA[AA$Gene == 'CytB',]$GenerationLength_d, method = 'spearman')  #
```

```
## Warning in cor.test.default(AA[AA$Gene == "CytB", ]$Val, AA[AA$Gene == "CytB", :
## Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "CytB", ]$Val and AA[AA$Gene == "CytB", ]$GenerationLength_d
## S = 61571169, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.3452085
```

```r
### try ND6 with generation length
cor.test(AA[AA$Gene == 'ND6',]$NumberOfGainers,AA[AA$Gene == 'ND6',]$GenerationLength_d, method = 'spea
```

```
## Warning in cor.test.default(AA[AA$Gene == "ND6", ]$NumberOfGainers, AA[AA$Gene
## == : Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "ND6", ]$NumberOfGainers and AA[AA$Gene == "ND6", ]$GenerationLength_d
## S = 42687830, p-value = 0.08619
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## 0.06735518
```

```r
cor.test(AA[AA$Gene == 'ND6',]$NumberOfLosers,AA[AA$Gene == 'ND6',]$GenerationLength_d, method = 'spearm
```

```
## Warning in cor.test.default(AA[AA$Gene == "ND6", ]$NumberOfLosers, AA[AA$Gene
## == : Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "ND6", ]$NumberOfLosers and AA[AA$Gene == "ND6", ]$GenerationLength_d
## S = 26946192, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## 0.4112789
```

```r
cor.test(AA[AA$Gene == 'ND6',]$Cys,AA[AA$Gene == 'ND6',]$GenerationLength_d, method = 'spearman')  # -0
```

```
## Warning in cor.test.default(AA[AA$Gene == "ND6", ]$Cys, AA[AA$Gene == "ND6", :
## Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "ND6", ]$Cys and AA[AA$Gene == "ND6", ]$GenerationLength_d
```

```
## S = 63393889, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.3850313
```

```r
cor.test(AA[AA$Gene == 'ND6',]$Trp,AA[AA$Gene == 'ND6',]$GenerationLength_d, method = 'spearman')  # -0
```

```
## Warning in cor.test.default(AA[AA$Gene == "ND6", ]$Trp, AA[AA$Gene == "ND6", :
## Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "ND6", ]$Trp and AA[AA$Gene == "ND6", ]$GenerationLength_d
## S = 52268613, p-value = 0.0002825
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.141966
```

```r
cor.test(AA[AA$Gene == 'ND6',]$Phe,AA[AA$Gene == 'ND6',]$GenerationLength_d, method = 'spearman')  # -0
```

```
## Warning in cor.test.default(AA[AA$Gene == "ND6", ]$Phe, AA[AA$Gene == "ND6", :
## Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "ND6", ]$Phe and AA[AA$Gene == "ND6", ]$GenerationLength_d
## S = 43691717, p-value = 0.2475
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## 0.04542221
```

```r
cor.test(AA[AA$Gene == 'ND6',]$Gly,AA[AA$Gene == 'ND6',]$GenerationLength_d, method = 'spearman')  # 0
```

```
## Warning in cor.test.default(AA[AA$Gene == "ND6", ]$Gly, AA[AA$Gene == "ND6", :
## Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  AA[AA$Gene == "ND6", ]$Gly and AA[AA$Gene == "ND6", ]$GenerationLength_d
## S = 30634475, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## 0.3306972
```

```r
cor.test(AA[AA$Gene == 'ND6',]$Val,AA[AA$Gene == 'ND6',]$GenerationLength_d, method = 'spearman')  # 0
```

```
## Warning in cor.test.default(AA[AA$Gene == "ND6", ]$Val, AA[AA$Gene == "ND6", :
## Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
```

```
## data:  AA[AA$Gene == "ND6", ]$Val and AA[AA$Gene == "ND6", ]$GenerationLength_d
## S = 26050897, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.4308393
```