Questions about: Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution

Question 1

On the first page the authors introduce the problem. The authors say that 'functional importance' or the number of functionally important positions were thought to be the main determinants of the rate of evolution of a gene. What precisely is the argument here? What data speaks against this argument? What has recently been observed to be the strongest correlate of evolutionary rate?

Question 2

Identifying a dominant cause in any observed process that is influenced by many causes boils down to quantifying the relative sizes of different influences. What is the key rate that suggests mis-folding errors may be a large influence, say in comparison in comparison to errors in replication.

Question 3

Figure 1 shows the correlations of a number of features of genes across a range of species. In particular, the features whose correlations are measured are:

1. mRNA-level
2. dN
3. dS
4. F_{op}
5. ts/tv-ratio

What is the definition of each of these quantities? How is it measured for each species?

What is the overall correlation structure that is observed across most species? The authors state "Principal component analysis (PCA) of each organism's correlation matrix confirmed that a single underlying component explains 36%–60% of the variance in all five analyzed variables". What does this mean precisely?

Why is it surprising that the ts/tv-ratio correlates negatively with dS?

Question 4

What is Akashi's test? (You can simply Google it if necessary). What is the significance of the outcome of this test? Say that I wanted to argue that high-expressed genes use preferred codons because they need to be translated *fast*, and not because they need to be more accurate. Does the test address this?

Question 5

Explain the misfolding hypothesis illustrated in figure 3. The evolutionary rate of a gene is the result of a balance of evolutionary and biophysical forces including that rate of errors in translation, the rate of misfolding, the expression level of the gene, the selective cost per misfolded protein, and the fraction of genetic mutations that either increase or decrease any of these parameters. How does this balance lead to the observed patterns? Does this imply that evolution is unable to affect certain parameters substantially? What is the argument for why neurons would be particularly sensitive?

Question 6

To test if their hypothesis can explain all observed patterns, the authors turned to computer simulations. How did these work? For example, how did the authors determine whether a particular sequence folds into the desired shape? How did they decide when the simulation was done? Do they discuss initial conditions and whether they matter?

Question 7

The suprising negative correlation between ts/tv-ratio and dS is by far largest at third positions in codons (that are most often synonymous) and weakest at second positions (which are essentially always nonsynonymous). What explanation does this support? So how in this explanation does the negative correlation arise?

Question 8

It seems that highly expressed genes acquire translation accuracy and translational robustness by different means. How do these genes optimize these two independent contributions to minimizing mis-folding?

Question 9

"Because we focus on patterns of covariation, our results do not imply that all variation in molecular evolution can be explained by protein misfolding or that protein function plays a negligible role." What is the purpose of this remark? Does it make sense to you?

Question 10

What is the nature of the cost? The authors present several arguments that the costs must mainly be the cytotoxicity of the unfolded proteins and not the cost of producing and degrading these useless products. What are the arguments? Are you convinced?