Optimal Codons in Gilchrist(2015) vs. Stenico(1994)

The “Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases” study by Stenico, et. al. (1994) used a correspondence analysis on the genes of C. elegans to determine correlations with codon usage bias. They found that relative synonymous codon usage (RSCU) has a large correspondence with codon usage bias, which they measured by looking at Nc or the effective number of codons used in a gene.

The study looked at the difference between the coding in the top 10% of biased—the most biased— genes and the lowest 10% of biased—the least biased—unrecognized reading frames (URFs). Comparing the usage of certain codons in genes versus the usage of codons in these URFs, they found that there were 21 codons whose usage was significantly higher (p < .01) among the high bias genes than the low bias URFs. I will state below the Gilchrist (2015) Figure 6 results (for *S. cerevisiae*) and the results from Stenico (1994) table 3, both of which look at codon usage frequency, for each amino acid. Note that Gilchrist (2015) looks at a continuous set of protein expressions and this paper looks at the top 10% of genes and lowest 10% percent of URFs regarding expression levels. Also note that the Stenico (1994) study uses the nucleotide U where the Gilchrist study uses T, so I put a XXT/XXU in the Gilchrist columns.

|  |  |  |  |
| --- | --- | --- | --- |
| Amino Acid: | Gilchrist Optimal Codon: | Stenica Optimal Codon: | Studies Agree: (x=yes) |
| Ala | GCT/GCU | GCU | x |
|  | - | GCC |  |
| Cys | TGT/UGU | UGC |  |
| Asp | GAC | GAC | x |
| Glu | GAA | GAG |  |
| Phe | TTC/UUC | UUC | x |
| Gly | GGT/GGU | GGA |  |
| His | CAC | CAC | x |
| Ile | ATC/AUC | AUC | x |
| Lys | AAG | AAG | x |
| Leu | TTG/UUG | CUU |  |
|  | - | CUC |  |
| Asn | AAC | AAC | x |
| Pro | CCA | CCA | x |
| Gln | CAA | CAG2 |  |
| Arg | AGA | CGU |  |
|  | - | CGC |  |
| Ser4 | TCC/UCC | UCC | x |
|  | TCT/UCU1 | UCU2 | x |
| Ser2 | AGT/AGU | - |  |
|  | AGC1 | - |  |
| Thr | ACC | ACC | x |
|  | ACT/ACU1 | - |  |
| Val | GTC/GUC | GUC | x |
|  | GTT/GUU1 | - |  |
| Tyr | TAC/UAC | UAC | x |

1 This indicates “codons that have been previously identified as optimal but our ROC SEMPPR model fit indicates these codons actually are the second most efficient codons” –Gilchrist (2016)

2 This indicates that the codon was bordering the significance criterion (p < .01) for being named an optimal codon. Glutamine’s CAG has a p = .005 while Ser4’s UCU has a p = .027.

Specific Comments on Each Amino Acid:

\*Alanine (Ala): ROC only identified one optimal codon (GCT/GCU) but it may be worth noting that the graph shows that GCC initially increases with increased gene expression. However, the GCC codon saw far more use in the Stenico study than the GCU.

\*Cysteine (Cys): Whereas ROC has TGT/UGU being both mutationally and selectively favored, the study shows that UGC is the optimal codon. It does appear from the study that UGU may indeed be the more mutationally favored though.

\*Aspartic acid (Asp): Follows same general trend for each paper. Including the “crossover” where one codon that is mutationally favored (in this case GAT/GAU) is more frequent in low bias RUFs but the more selectively advantageous codon GAC has significantly more frequency in high bias genes.

\*Glutamic Acid (Glu): Whereas ROC has GAA being both mutationally and selectively favored, the study shows that GAG is the optimal codon. It does appear from the study that GAA may indeed be the more mutationally favored though.

\*Phenylalanine (Phe): Follows same general trend for each paper. This, like Aspartic acid, shows the trend of the crossover, although with less extremity regarding the low bias mutation- favoring of UUU.

\*Glycine (Gly): Data does not at all match.

\*Histidine (His): Follows same general trend for each paper, including the crossover trend.

\*Isoleucine (Ile): Follows same general trend for each paper.

\*Lysine (Lys): Follows same general trend for each paper.

\*Leucine (Leu): Data does not at all match.

\*Asparagine (Asn): Follows same general trend for each paper, including the crossover trend.

\*Proline (Pro): Follows same general trend for each paper. CCA is pretty dominant in both high and low bias cases, and this makes sense as it is show by ROC to be the most mutationally and selectively favored.

\*Glutamine (Gln): It should be noted the CAA showed more expression in both the high bias genes and low bias URFs, but it showed a lower frequency in the high bias genes than in the low bias URFs. This could make some sense seeing as ROC predicts that CAA is both mutationally and selectively advantageous. However, the significant difference in frequency of CAG pushed the study to establish CAG as the optimal codon. This case is one of the two that border on the criterion for significance level with p = .005, which still meets the criterion (p < .01). The other codon that borders is in Ser4.

\*Arginine (Arg): Data does not match at all between the two studies.

\*Serine (Ser):

* Ser4 – The data is in agreement even to the point where TCC/UCC is the most optimal in each. The TCT/UCU is the second most optimal according to ROC and is the second case that borders the criterion according to the Stenico study with a chi square value that has a probability of .027.
* Ser2 – ROC predicts two optimal codons, but the Stenico study found no significant difference between the codon usage in high bias genes and low bias RUFs.

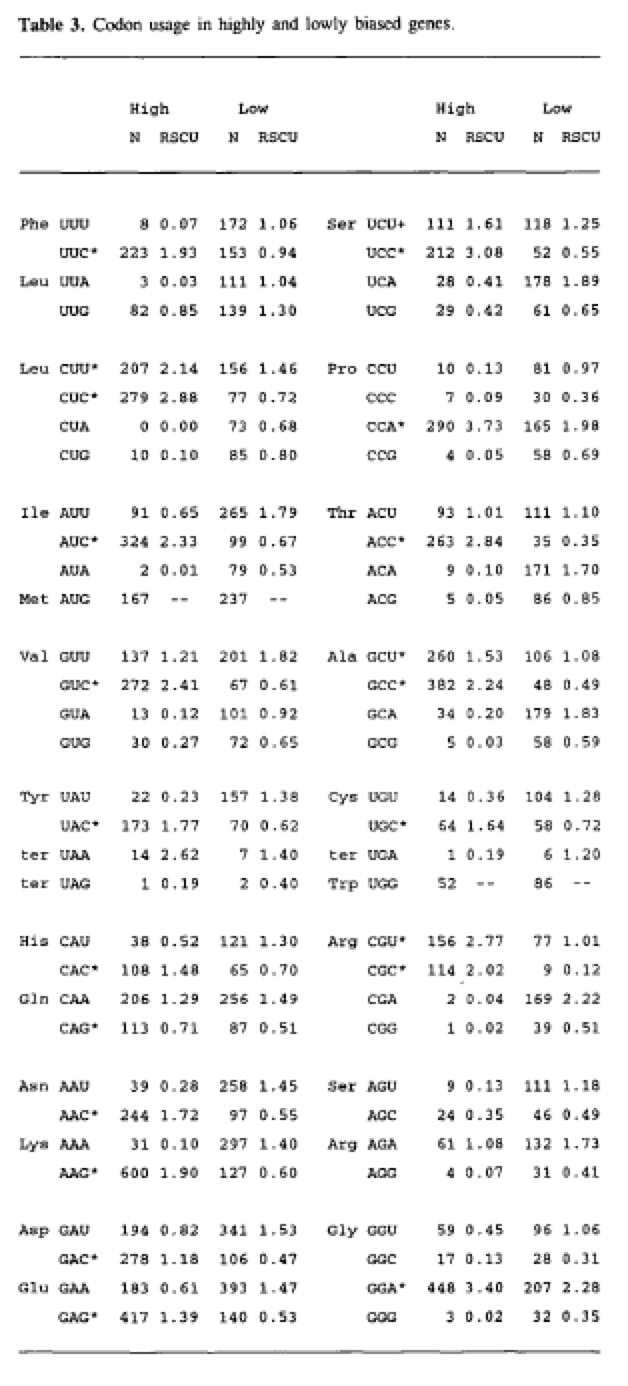
\*Threonine (The): The two studies agree that the ACC codon is the optimal codon, but ROC indicates that ACT/ACU is the second optimal, while the Stenico study finds no significant difference in the data for that codon. Still, ACT/ACU was that second most observed in the high bias genes. The difference could be caused by the high mutational value of ACU not indicating much difference between high bias and low bias RUFs.

\*Valine (Val): The two studies agree that the GTC/GUC codon is the optimal codon, but ROC

indicates that GTT/GUU is the second optimal, while the Stenico study finds no significant difference in the data for that codon. Still, GTT/GUU was the second most observed in the high bias genes. The difference could again be caused by the high mutational value of GUU not indicating much difference between high bias genes and low bias RUFs.

\*Tyrosine (Tyr): Follows the same general trend for each paper.. including the fact that TAC is less mutationally favored, and is seen less often in the low bias RUFs but more selectively favored and seen more often in high bias genes.

Stenico(1994) Table 3



Gilchrist (2015) Figure 6

