

Department of Biology

Ashiana Jivraj

Mentor: Brian Morton

Codon usage patterns in nuclear genes from *Arabidopsis thaliana*, *Oryza sativa* and *Solanum tuberosum*

Evolutionary models of codon bias are traditionally defined by the selection-mutational drift hypothesis, which suggest that codon bias is the result of a balance between mutation bias and selection. Hypotheses about selection on codon usage have generally focused on increased translation efficiency in highly expressed genes and it has been proposed that this is responsible for an increased use of major codons in *Arabidopsis thaliana*. In this study we introduce a more complex mutation model to account for bias as a function of the two neighboring bases, or context. A previous analysis showed that in *Arabidopsis* major codons are actually underused in some contexts. In order to more fully examine this feature of codon use, high and low expression genes of *Solanum tuberosum*, *Oryza sativa*, and *A. thaliana* were collected and analyzed using non-coding regions to estimate context-dependent mutation biases. These data were statistically analyzed to compare the observed number of each codon to its expected number as a function of context. Across the three species, overall, codon usage within each synonymous group was biased, as expected, to those codons ending in G or C. However, in the dicots, we find that major codon usage is actually influenced by context: when G is the 3' flanking nucleotide that there is under representation of many major codons. This indicates that translational efficiency cannot solely explain codon bias patterns and suggests that there are other selective factors, such as perhaps RNA structure that are necessary to explain the codon bias observed in some plant nuclear genomes.