

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Tajima’s D | - | 0.106 | 0.163 | 0.342 | -1.343 | -2.539 |

Figure 1 Site frequency spectra for different classes of sites. The first three columns correspond to three subtypes of synonymous changes, i.e. No-Change (NoChg), Preferred-to-Unpreferred (P->U) and Unpreferred-to-Preferred (U->P), which were classified based on the expected effect of the synonymous change on codon usage (Haddrill et al. 2008). The No-Change class is closest to being a proxy for a neutral reference. A theoretical neutral spectrum is also shown for comparison (see Methods). The four bins within each class correspond to singletons (1/150), low frequency (2-15 / 150), intermediate frequency (16-119 / 150) and high frequency (120-149 / 150). Sites with less than 150 sequenced lines were discarded while those with more than 150 lines were down sampled using a hyper-geometric distribution. The table beneath the bar plot shows the Tajima’s D statistic, which summarizes the site frequency spectrum, with negative values indicating skew towards rare variants and a rule-of-thumb for determining significance being +2/-2. Negative Tajima’s D can be caused by population expansion or purifying selection.

Another way to look at the constraint is to compare pi (per site). For this purpose, I counted the total number of “one-step” sites, which accounts for the mutational opportunities to generate PTC. What I got is the following:

Table 1 Nucleotide diversity estimate for different classes of variation

|  |  |  |
| --- | --- | --- |
| 4 fold degenerate | Nonsynonymous | PTC |
| 9.1x10-3 | 9.3x10-4 | 5.0x10-5 |

