Method 1:

Stitching each gene in a given strand (direction), number of occurrences of each codon is calculated, this is followed by summing the number of codons of each type in both directions. The count for each codon is actually calculated for a segment, which gives us the breakup of the codon count along the genome length. The codon count is shown in Figure 1, Figure 2, Figure 3 and Table 1.

Figure 1: Codon counted according to position-1, seperated by count in each segment

Figure 2: Codon counted according to position-2, seperated by count in each segment

Figure 3: Codon counted according to position-3, seperated by count in each segment

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Position | TotalGene | A | T | G | C | Total |
| 1 | 1M | 61929 | 62110 | 83999 | 81149 | 289187 |
| 2M | 65541 | 68471 | 74385 | 72622 | 281019 |
| 3M | 62898 | 71223 | 68078 | 83897 | 286096 |
| 4M | 66487 | 67873 | 77435 | 79912 | 291707 |
| 4M to end | 33974 | 46161 | 49000 | 54081 | 183216 |
| Cumulative | **290829** | **315838** | **352897** | **371661** | **1331225** |
| 2 | 1M | 70563 | 72918 | 73885 | 71821 | 289187 |
| 2M | 69250 | 62923 | 74167 | 74679 | 281019 |
| 3M | 68912 | 70994 | 73888 | 72302 | 286096 |
| 4M | 70320 | 67012 | 76681 | 77694 | 291707 |
| 4M to end | 48822 | 44994 | 50370 | 39030 | 183216 |
| Cumulative | **327867** | **318841** | **348991** | **335526** | **1331225** |
| 3 | 1M | 72601 | 71566 | 76257 | 68763 | 289187 |
| 2M | 67356 | 72315 | 67464 | 73884 | 281019 |
| 3M | 75972 | 63665 | 73151 | 73308 | 286096 |
| 4M | 75996 | 76243 | 69435 | 70033 | 291707 |
| 4M to end | 47395 | 40646 | 48084 | 47091 | 183216 |
| Cumulative | **339320** | **324435** | **334391** | **333079** | **1331225** |

Table 1: Codon count per segment for method-1

The codon frequency empirically determined is shown in Table 2; comparison with codon frequency found in literature is also shown .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Provided | A | T | G | C |
| Pos1 | 262 | 160 | 345 | 231 |
| Pos2 | 296 | 296 | 179 | 227 |
| Pos3 | 193 | 272 | 278 | 255 |
| Method1 | A | T | G | C |
| Pos1 | 218 | 237 | 265 | 279 |
| Pos2 | 246 | 239 | 262 | 252 |
| Pos3 | 254 | 243 | 251 | 250 |

Table 2: Comparing codon frequency in literature and one empirically found

The breakdown of codon count along positive and negative gene are shown below in Table 3, Table 4 and nucleotide skew for positive and negative genes are shown in Figure 4, Figure 5. We notice that the codon count with nucleotide ‘G’ in any of the three position is always higher than that of ‘C’ in positive gene, while the opposite is true for negative gene. Although when we combine codon/nucleotide count of both positive and negative gene there does not seem to be any pattern available among nucleotide count, we can see clear distinction while they are seen individually.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Position | Pgene | A | T | G | C |
| 1 | 1M | 36048 | 34183 | 49605 | 40398 |
| 2M | 30890 | 34156 | 39910 | 34054 |
| 3M | 30304 | 33287 | 28462 | 29748 |
| 4M | 29126 | 31997 | 38965 | 36601 |
| 4M to end | 19066 | 22629 | 25871 | 22882 |
| Cumulative | 145434 | 156252 | **182813** | **163683** |
| 2 | 1M | 40359 | 37985 | 43473 | 38417 |
| 2M | 35262 | 30334 | 39612 | 33802 |
| 3M | 27477 | 27425 | 35600 | 31299 |
| 4M | 34598 | 28052 | 39859 | 34180 |
| 4M to end | 21462 | 18060 | 29244 | 21682 |
| Cumulative | 159158 | 141856 | **187788** | **159380** |
| 3 | 1M | 36630 | 41535 | 43449 | 38620 |
| 2M | 34966 | 38340 | 33148 | 32556 |
| 3M | 31418 | 28234 | 32322 | 29827 |
| 4M | 34965 | 37367 | 31437 | 32920 |
| 4M to end | 24321 | 23781 | 21514 | 20832 |
| Cumulative | 162300 | 169257 | **161870** | **154755** |

Table 3: Codon count per segment for positive gene using method-1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Position | Ngene | A | T | G | C |
| 1 | 1M | 25881 | 27927 | 34394 | 40751 |
|  | 2M | 34651 | 34315 | 34475 | 38568 |
|  | 3M | 32594 | 37936 | 39616 | 54149 |
|  | 4M | 37361 | 35876 | 38470 | 43311 |
|  | 4M to end | 14908 | 23532 | 23129 | 31199 |
|  | Cumulative | 145395 | 159586 | **170084** | **207978** |
| 2 | 1M | 30204 | 34933 | 30412 | 33404 |
|  | 2M | 33988 | 32589 | 34555 | 40877 |
|  | 3M | 41435 | 43569 | 38288 | 41003 |
|  | 4M | 35722 | 38960 | 36822 | 43514 |
|  | 4M to end | 27360 | 26934 | 21126 | 17348 |
|  | Cumulative | 168709 | 176985 | **161203** | **176146** |
| 3 | 1M | 35971 | 30031 | 32808 | 30143 |
|  | 2M | 32390 | 33975 | 34316 | 41328 |
|  | 3M | 44554 | 35431 | 40829 | 43481 |
|  | 4M | 41031 | 38876 | 37998 | 37113 |
|  | 4M to end | 23074 | 16865 | 26570 | 26259 |
|  | Cumulative | 177020 | 155178 | **172521** | **178324** |

Table 4: Codon count per segment for negative gene using method-1

The cumulative count for each segment for positive, negative gene and their sum is as shown respectively in Table 5, Table 6, Table 7. The codon/nucleotide count is significantly different in each of the segments.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Segment | CountofA | CountofT | CountofG | CountofC |
| 1M | 113037 | 113703 | 136527 | 117435 |
| 2M | 101118 | 102830 | 112670 | 100412 |
| 3M | 89199 | 88946 | 96384 | 90874 |
| 4M | 98689 | 97416 | 110261 | 103701 |
| 4M to end | 64849 | 64470 | 76629 | 65396 |
| Cumulative | 466892 | 467365 | 532471 | 477818 |

Table 5: Count of nucleotides in a segment of positive gene, regardless of position

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Segment | CountofA | CountofT | CountofG | CountofC |
| 1M | 92056 | 92891 | 97614 | 104298 |
| 2M | 101029 | 100879 | 103346 | 120773 |
| 3M | 118583 | 116936 | 118733 | 138633 |
| 4M | 114114 | 113712 | 113290 | 123938 |
| 4M to end | 65342 | 67331 | 70825 | 74806 |
| Cumulative | 491124 | 491749 | 503808 | 562448 |

Table 6: Count of nucleotides in a segment of negative gene, regardless of position

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Segment | CountofA | CountofT | CountofG | CountofC |
| 1M | 92056 | 92891 | 97614 | 104298 |
| 2M | 101029 | 100879 | 103346 | 120773 |
| 3M | 118583 | 116936 | 118733 | 138633 |
| 4M | 114114 | 113712 | 113290 | 123938 |
| 4M to end | 65342 | 67331 | 70825 | 74806 |
| Cumulative | 491124 | 491749 | 503808 | 562448 |

Table 7: Count of nucleotides in a segment of total gene, regardless of position

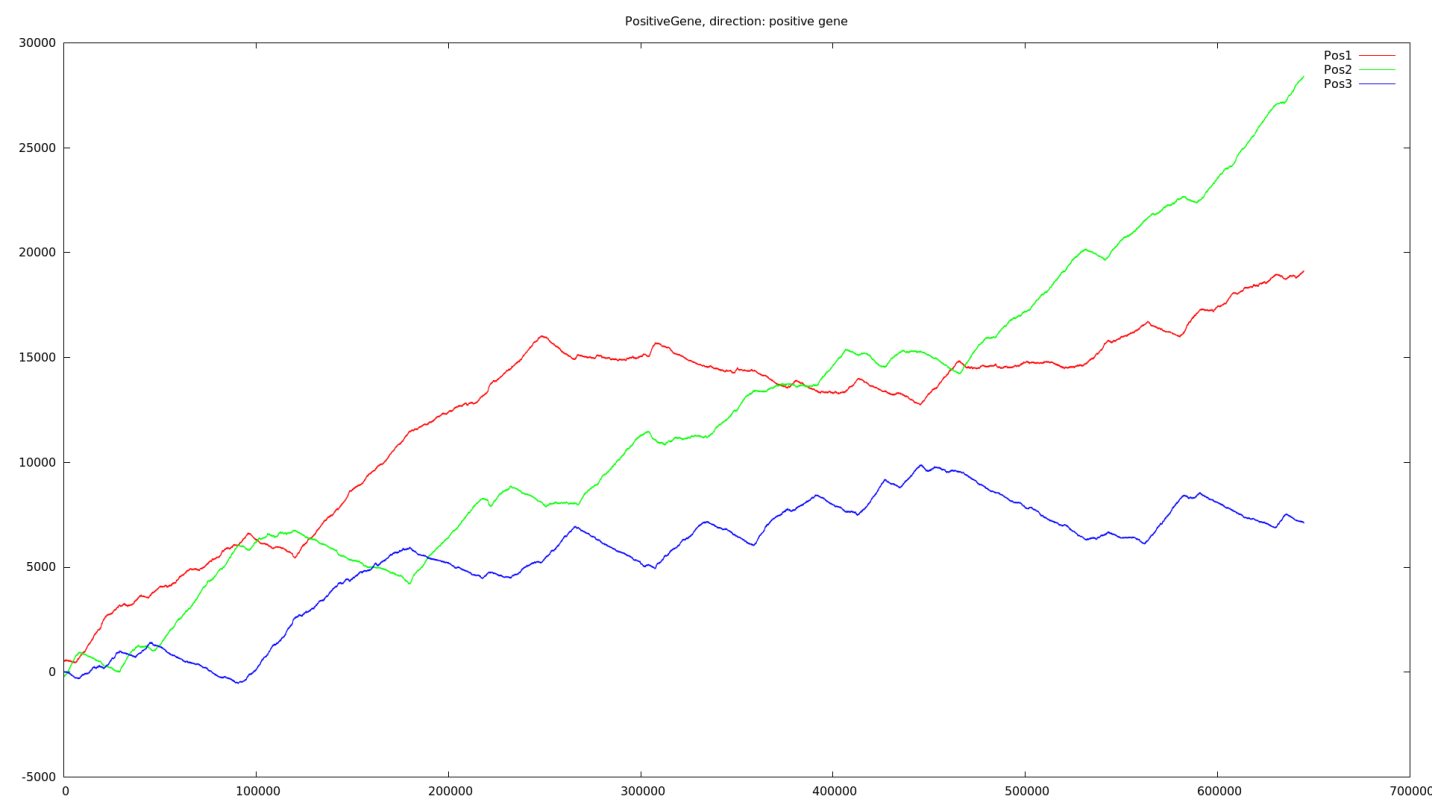


Figure 4: Nucleotide skew of positive gene

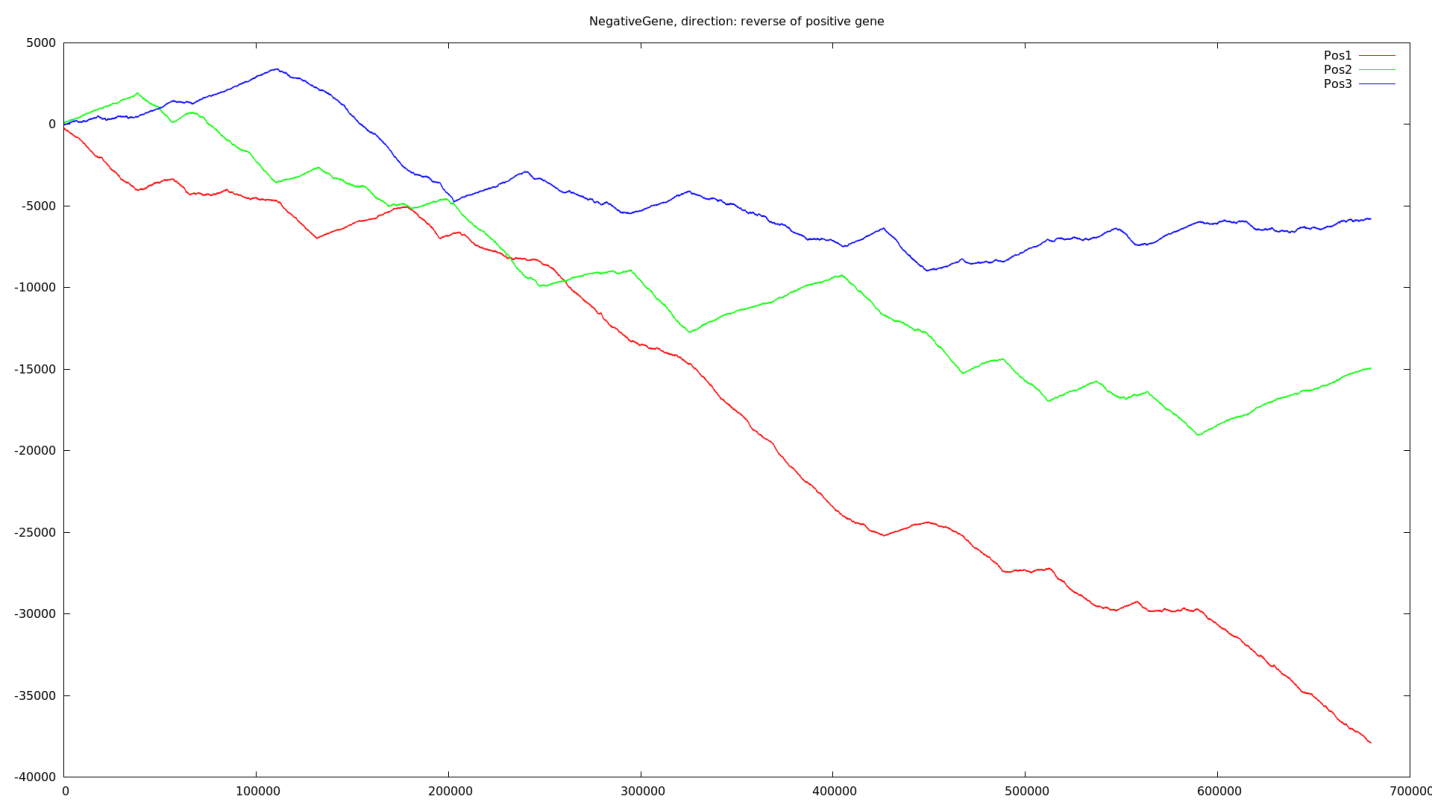


Figure 5: Nucleotide skew of Negative gene