

# An Analysis of Single Nucleotide Substitution in Genetic Codons - Probabilities and Outcomes

Tariq Abdullah<sup>1\*</sup>, Muniba Faiza<sup>1</sup>, Prashant Pant<sup>2</sup>, Mohd Rayyan Akhtar<sup>3</sup>, Pratibha Pant<sup>4</sup>

<sup>1</sup>Department of Computer Science, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi, India; <sup>2</sup>Department of Botany, Dyal Singh College, University of Delhi, New Delhi - 110003, India; <sup>3</sup>R & D Division, UTL Inverters Pvt. Ltd., New Delhi, - 110025, India; <sup>4</sup>Department of Botany, Daulat Ram College, University of Delhi, New Delhi, India; Tariq Abdullah - Email: tariq141760@st.jmi.ac.in; Phone: +91 8527572667; \*Corresponding author

Received March 24, 2016; Revised April 12; Accepted April 14, 2016; Published June 15, 2016

## Abstract:

**Background:** Single nucleotide substitutions (SNS) in genetic codon are of prime importance due to their ability to alter an amino acid sequence as a result. Given the nature of genetic code, any SNS is expected to change the protein sequence randomly into any of the 64 possible codons. In this paper, we present a theoretical analysis of how single nucleotide substitutions in genetic codon may affect resulting amino acid residue and what is the most likely amino acid that will get selected as a result. **Methods:** A probability matrix was developed showing possible changes and routes likely being followed as a result of base substitution mutation causing changes at the translational level for the amino acid being encoded. **Results:** We observe that in event of single base pair substitution in a given amino acid; a chosen set of amino acids is theoretically more probable to be resulted suggesting a directional rather than a random change. This study also indicates that for a given amino acid coded by a number of synonymous codons, all synonymous codons will result into same list of amino acids in case of all possible SNS at three positions. **Conclusion:** The present work has resulted into development of a theoretical probability matrix which can be used to predict changes in amino acid residues in a protein sequence caused by single base substitutions.

**Keywords:** Mutation, point mutation, codon, genetic code, single base substitutions or single nucleotide substitutions

## Background:

Central dogma of molecular biology and more appropriately of life suggests that the genetic information in DNA is coded in sets of three bases called a triplet which is more evident at the mature mRNA level (after due post transcriptional and mRNA splicing/editing) [1]. The processed mRNA mobilizes out of the nucleus, into the cytoplasm and is then translated into polypeptide chains which form the functional basis of existence of life [2]. The triplets encoding amino acids show degeneracy to a certain level thereby having more than one kind of codon for a given amino acid [3, 4, 5]. The nucleotides in a DNA strand are prone to modifications, substitutions, insertions, deletions collectively referred to as mutations. These mutations can occur due to several factors out of which spontaneous mutation by environmental selection pressures, mutagens, slippage during replication, errors of DNA polymerase, errors in transcription are the most common reasons [5, 6, 7]. A mature transcript is an open reading frame consisting of codons. These codons may code for a particular amino acid and

sometimes, many codons code for the same amino acid [1]. This phenomenon of multiple codons encoding the same amino acid is called codon degeneracy. Reportedly, there is a huge bias for preferring a particular codon over others for an amino acid in an organism; a phenomenon referred to as codon (usage) bias [8]. The degenerate codon are said to have different frequency of occurrence and are subject to being favored depending on whether a genome is GC or AT rich [8]. Although being neutral at protein sequence level, occurrence of degenerate codons offers evolutionary advantages and contributes to genetic robustness. According to Archetti (2006) [9], natural selection may lead to a preference for codons reducing the impact of errors. Any single nucleotide change in such degenerate or synonymous codons is expected to result differently depending on which position is undergoing substitution [10, 11, 12, 13]. With an assumption that the net outcome would be random according to the nature of the code we attempt to envisage the possible amino acid substitutions occurring when single base

(=nucleotide) substitution are (point mutation) are applied in silico on degenerate codons.

BOX 1: Program to find out all possible codons due to single base pair substitution in genetic code

```
#!/usr/bin/perl
@nucleotides = ('A','U','C','G');

%aminoacids = (
  Methionine => ["AUG"],
  Tryptophan => ["UGG"],
  Tyrosine => ["UAU","UAC"],
  Cysteine => ["UGU","UGC"],
  Asparagine => ["AAU","AAC"],
  Lysine => ["AAA","AAG"],
  AsparticAcid=> ["GAU","GAC"],
  GlutamicAcid=> ["GAA","GAG"],
  Histidine => ["CAU","CAC"],
  Glutamine => ["CAA","CAG"],
  Phenylalanine=>["UUU","UUC"],
  Stop => ["UAA","UAG","UGA"],
  Isoleucine => ["AUU","AUC","AUA"],
  Proline => ["CCU","CCC","CCA","CCG"],
  Threonine => ["ACU","ACC","ACA","ACG"],
  Valine => ["GUU","GUC","GUA","GUG"],
  Alanine => ["GCU","GCC","GCA","GCG"],
  Glycine => ["GGU","GGC","GGA","GGG"],
  Leucine=> ["UUA","UUG","CUU","CUC","CUA","CUG"],
  Serine=> ["UCU","UCC","UCA","UCG","AGU","AGC"],
  Arginine=> ["CGU","CGC","CGA","CGG","AGA","AGG"]
);

open (OUTPUT,">output.txt");

foreach $aminoacid ( keys %aminoacids ){
  print OUTPUT "Combinations For $aminoacid\n";
  foreach $codon (@{$aminoacids{$aminoacid}}){
    print OUTPUT "\nCodon $codon \n";
    foreach $nucleotide (@nucleotides){
      for($i=0; $i<3;$i++){
        $cdn = $codon;
        substr($cdn, $i, 1) = $nucleotide;
        if($cdn ne $codon){
          print OUTPUT "$cdn\n";
        }
      }
    }
  }
}

print "Output file prepared";
```

Figure 1: A program code to find out all possible codons due to single nucleodite substitution in genetic code.

### Methodology:

A Perl script (PERL, <https://www.perl.org/>) using Strawberry PERL version 5.22.0.1 available at URL <http://strawberryperl.com/> was written (shown in Box 1) to generate a list of all possible amino acids that can be formed when each nucleotide position in a given codon is replaced by a given nucleotide A, U, G or C. The program works by replacing each position in a codon by every nucleotide for each codon of every amino acid (Figure 1). The resulting amino acids were determined from the output file containing codons with all the possible replacements (Supplementary file 1 & 2 - available with authors).  $P_{ij}$  value is defined as the ratio of the total number of times an amino acid 'i' is actually being replaced by amino acid 'j', to the total number of possible single base pair replacements made in an amino acid 'i'.

The probability of each amino acid resulting from single base substitutions was calculated from the following relation:

$$P_{ij} = n_j / \sum \text{possible single base pair replacements of amino acid } i$$

Where,

*i* is the amino acid being replaced;

*j* is the amino acid had been substituted in place of *i*;

$P_{ij}$  is the probability of single base pair replacement of amino acid *i* by an amino acid *j*.

$n_j$  is the number of times a particular amino acid has been replaced within its space.

Table 1 shows the matrix thus obtained containing  $P_{ij}$  values as fractions for each amino acid with numerator denoting the number of times an amino acid replacement occurs and denominator depicting the total number of single base replacements. Table 2 is a modified version of Table 1 showing the derived probability of amino acid replacement at two decimal places. All the cells with zero values (null cells) were intentionally left blank. Table 3 is matrix division of total amino acids possible for each kind of amino acid and total number of observed transitions.

### Result:

At the outset, we would like to declare that the results indicated here are reserved for amino acids showing codon degeneracy. The probability matrix (Table 1) shows that when a single base pair mutation takes place for a given set of synonymous codon, the probability of an amino acid changing into another is equal to in either direction for a given single base substitutions in almost all cases leaving some instances (indicated in relation below).

$$P(S_i/S_j) = \alpha_i, \text{ where codons are synonymous}$$

It is also observed that although mutations lead to a wide range of amino acid changes in a protein sequence, but single base substitutions in genetic codons, in most cases result in no significant change at the amino acid level due to more or less equal probability of synonymous substitution (a change comprising of replacement of one amino acid with another of its own kind or group) or no change at all (Tables 2 - 3). This is of course, excluding the case where there is only one codon for an amino acid (e.g. for Methionine etc). Tables 1 & 2 also indicates that an amino acid is bound to mutate into some selected (naturally pre-determined) amino acids irrespective of how the codons mutate given that there is a single base substitutions at a time for a given codon. From the perspective of synonymous and non-synonymous substitutions due to single base substitutions, our finding suggests that whenever such a mutation occurs in a codon, an amino acid may get replaced but only into a more similar amino acids or in other words, synonymous substitutions are more favored over non-synonymous substitutions. This is generally true for high expression genes (HEG) as well as low expression genes (LEG) [14]. Table 2 indicates that the normalized probability of one kind of amino acid changing into other is same in both directions.

Another interesting fact revealed from our analyses is that when single base substitutions takes place in synonymous codon substitutions are also synonymous and are at higher frequency over non-synonymous amino acid substitutions. This implies that for a radical or significant change in the protein sequence, it is the non-degenerate codon which act as putative targets for single base substitutions as otherwise degenerate ones will largely counter the effect of mutation. For instance, Tryptophan and Methionine resist changes and replacements by single base substitutions. The value of probabilities of replacement (Pij) in case of G->G, A->A, V->V, R->R and P->P are same, i.e., 0.33. The single base pair replacement of L->L has the highest Pij value, 0.35 among all the replacements. Thus, it proves that greater is the number of codons coding a particular amino acid, higher are its chances of synonymous substitutions despite single base substitutions s (for instance Leucine and Arginine). Similarly, Threonine is encoded by four codons which results in 15 different possible replacements across various groups when substitutions occurs. Although, the results also indicates that the total number of replacements need not be in perfect correspondence with the number of

codons for a particular amino acids. For instance, while Leucine, Arginine and Serine are encoded by six codons, they still have lesser number of possible replacements as compared to Threonine (Table 2).

In Singh-Pant *et al.* (2012) [15], we have shown signatures of positive selection in AC4 gene of Begomoviruses causing Papaya leaf curl viruses detected using fixed effects likelihood (FEL) method at different p-values [15]. The amino acid residues under positive selection were detected and were predominantly of non-synonymous type. To substantiate the utility of our work, the nucleotide sequence of the AC4 gene along with other ORFs was analyzed (Figure 2) along with data described elsewhere [15]. It was found that the predicted substitution as per the probability matrix developed in this work exemplarily corresponds with the real sequence data. For instance, there is a substitution from serine to phenylalanine and cysteine in AC4 multiple sequence alignment [15]. On cross-checking the probability of this substitution in the probability matrix, it is among the highest (2/18 and 4/18) which shows our matrix to correctly model the directional substitutions.

**Table 1:** An amino acid replacement probability matrix (ARPM) calculated with the help of Perl program, showing the probability of a replacement event (raw data) of an amino acid by another due to single base replacement (SBR) mutation. Color coding of the cells was done to show the nature/group to which the amino acid belongs as shown in the table below.

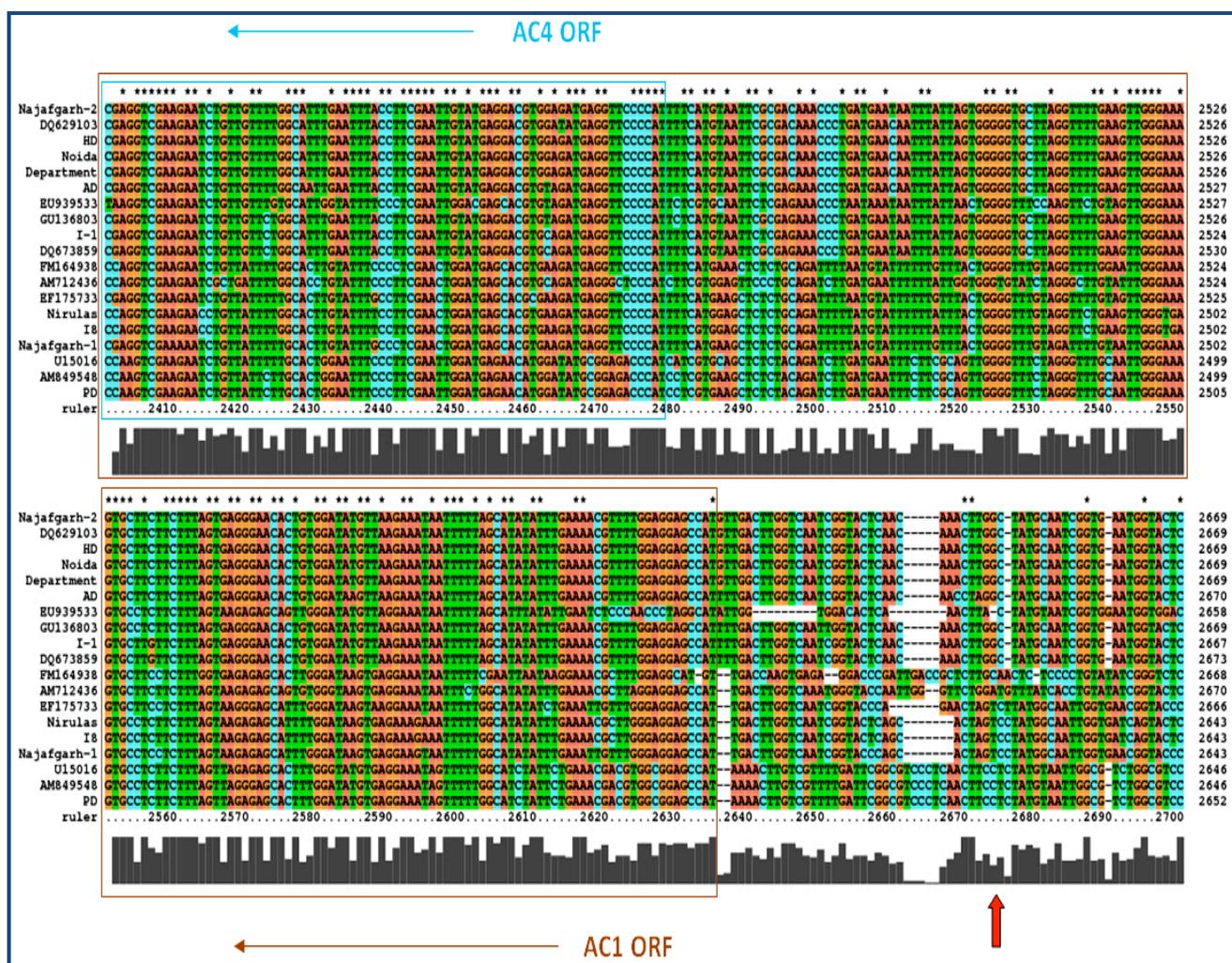
	G	A	V	L	I	P	F	Y	W	S	T	C	M	N	Q	K	R	H	D	E	ST
G	12/36	4/36	4/36	0/36	0/36	0/36	0/36	0/36	1/36	2/36	0/36	2/36	0/36	0/36	0/36	0/36	6/36	0/36	2/36	2/36	1/36
A	4/36	12/36	4/36	0/36	0/36	4/36	0/36	0/36	0/36	4/36	4/36	0/36	0/36	0/36	0/36	0/36	0/36	0/36	2/36	2/36	0/36
V	4/36	4/36	12/36	6/36	3/36	0/36	2/36	0/36	0/36	0/36	0/36	0/36	1/36	0/36	0/36	0/36	0/36	0/36	2/36	2/36	0/36
L	0/54	0/54	6/54	19/54	4/54	4/54	6/54	0/54	1/54	2/54	0/54	0/54	2/54	0/54	4/54	4/54	2/54	0/54	0/54	3/54	0/54
I	0/27	0/27	3/27	4/27	6/27	0/27	2/27	0/27	0/27	2/27	3/27	0/27	3/27	2/27	0/27	1/27	1/27	0/27	0/27	0/27	0/27
P	0/36	4/36	0/36	4/36	0/36	12/36	0/36	0/36	0/36	4/36	4/36	0/36	0/36	0/36	2/36	0/36	4/36	2/36	0/36	0/36	0/36
F	0/18	0/18	2/18	6/18	2/18	0/18	2/18	2/18	0/18	2/18	0/18	2/18	0/18	0/18	0/18	0/18	0/18	0/18	0/18	0/18	0/18
Y	0/18	0/18	0/18	0/18	0/18	0/18	2/18	2/18	0/18	2/18	0/18	2/18	0/18	0/18	0/18	0/18	0/18	2/18	2/18	0/18	4/18
W	1/9	0/9	0/9	1/9	0/9	0/9	0/9	0/9	1/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9	2/9	0/9	0/9	0/9	2/9
S	2/54	4/54	0/54	2/54	2/54	4/54	2/54	2/54	1/54	14/54	6/54	4/54	0/54	2/54	0/54	0/54	6/54	0/54	0/54	0/54	3/54
T	0/36	1/36	0/36	0/36	3/36	1/36	0/36	1/36	0/36	3/36	6/36	0/36	1/36	6/36	2/36	2/36	2/36	1/36	1/36	2/36	2/36
C	2/18	0/18	0/18	0/18	0/18	0/18	2/18	2/18	2/18	4/18	0/18	2/18	0/18	0/18	0/18	0/18	2/18	0/18	0/18	2/18	2/18
M	0/9	0/9	1/9	2/9	3/9	0/9	0/9	0/9	0/9	0/9	1/9	0/9	0/9	0/9	0/9	1/9	1/9	0/9	0/9	0/9	0/9
N	0/18	0/18	0/18	0/18	2/18	0/18	0/18	2/18	0/18	2/18	2/18	0/18	0/18	2/18	0/18	4/18	0/18	2/18	2/18	0/18	0/18
Q	0/18	0/18	0/18	2/18	0/18	2/18	0/18	0/18	0/18	0/18	0/18	0/18	0/18	2/18	2/18	2/18	2/18	4/18	0/18	2/18	2/18
K	0/18	0/18	0/18	0/18	1/18	0/18	0/18	0/18	0/18	0/18	2/18	0/18	1/18	4/18	2/18	2/18	2/18	0/18	0/18	2/18	2/18
R	6/54	0/54	0/54	4/54	1/54	4/54	0/54	0/54	2/54	6/54	2/54	2/54	1/54	0/54	2/54	18/54	2/54	0/54	0/54	2/54	2/54
H	0/18	0/18	0/18	2/18	0/18	2/18	0/18	2/18	0/18	0/18	0/18	0/18	0/18	2/18	4/18	0/18	2/18	2/18	2/18	0/18	0/18
D	2/18	2/18	2/18	0/18	0/18	0/18	0/18	2/18	0/18	0/18	0/18	0/18	0/18	2/18	0/18	0/18	0/18	2/18	2/18	4/18	0/18
E	2/18	2/18	2/18	0/18	0/18	0/18	0/18	0/18	0/18	0/18	0/18	0/18	0/18	2/18	2/18	0/18	0/18	4/18	2/18	2/18	2/18
ST	1/27	0/27	0/27	3/27	0/27	0/27	0/27	4/27	2/27	3/27	0/27	2/27	0/27	0/27	2/27	2/27	2/27	0/27	0/27	2/27	4/27

**Table 2:** Refined ARPM table with probability values in two decimal places. Replacement event with no (zero) probabilities were left blank. Color coding of the cells was done to show the nature/group to which the amino acid belong.

	G	A	V	L	I	P	F	Y	W	S	T	C	M	N	Q	K	R	H	D	E	STOP
G	0.33	0.11	0.11						0.03	0.06		0.06							0.06	0.06	0.03
A	0.11	0.33	0.11			0.11				0.11	0.11									0.06	0.06
V	0.11	0.11	0.33	0.17	0.08		0.06						0.03						0.06	0.06	
L			0.11	0.35	0.07	0.07	0.11		0.02	0.04			0.04		0.04		0.07	0.04			0.06
I			0.11	0.15	0.22		0.07			0.07	0.11		0.11	0.07		0.04	0.04				
P		0.11		0.11	0.33					0.11	0.11				0.06		0.11	0.06			
F				0.33	0.11		0.11	0.11		0.11		0.11									0.22
Y							0.11	0.11		0.11		0.11		0.11				0.11	0.11		0.22
W	0.11			0.11				0.11		0.11		0.11					0.22				0.22
S	0.04	0.07		0.04	0.04	0.07	0.04	0.04	0.02	0.26	0.11	0.07		0.04			0.11				0.06
T		0.03			0.08	0.03		0.03		0.08	0.17		0.03	0.17	0.06	0.06	0.06	0.03	0.03	0.06	0.06
C	0.11						0.11	0.11	0.11	0.22		0.11						0.11			0.11
M				0.11	0.22	0.33					0.11						0.11	0.11			
N					0.11				0.11		0.11				0.11		0.22		0.11		
Q				0.11		0.11									0.11	0.11	0.11	0.22		0.11	0.11
K					0.06						0.11		0.06	0.22	0.11	0.11	0.11			0.11	0.11
R	0.11			0.07	0.02	0.07			0.04	0.11	0.04	0.04	0.02		0.04	0.33	0.04				0.04
H				0.11		0.11			0.11					0.11	0.22		0.11	0.11	0.11		
D	0.11	0.11	0.11						0.11					0.11			0.11	0.11	0.11		0.22
E	0.11	0.11	0.11												0.11	0.11			0.11	0.11	0.11
STOP	0.04			0.11				0.15	0.07	0.11		0.07			0.07	0.07	0.07			0.07	0.15

**Table 3:** Ratio of Possible and observed replacements tabulated group-wise. Values in this table were calculated after counting the total number of possible one-group-to-other transitions in Table 2 and dividing it by the total number of cells containing non-zero values. For example, there are 36 possible non polar to non polar amino acid transition (color coded yellow) in Table 2 yet only 21 cells have a non zero value, Hence we get  $21/36 = 0.58$ .

	NPA	ARM	PUH	SC	AMDC	B+	A-	STP
NPA	0.58	0.28	0.67	0.33	0.25	0.39	0.50	0.33
ARM	0.28	0.44	0.50	0.50	0.25	0.22	0.17	0.67
PUH	0.67	0.67	1.00	0.50	0.75	0.67	0.50	1.00
SC	0.33	0.50	0.50	0.25	0.00	0.50	0.00	0.50
AMDC	0.25	0.17	0.50	0.00	0.50	0.83	0.50	0.50
B+	0.58	0.22	0.50	0.50	0.83	0.78	0.33	0.67
A-	0.50	0.17	0.00	0.00	0.50	0.33	1.00	0.50
STP	0.33	0.67	0.50	0.50	0.50	0.67	0.50	1.00



**Figure 2:** Multiple sequence Alignment (MSA) of the nucleotide of AC4 and AC1 ORFs of Papaya Leaf Curl Virus. Single Nucleotide Substitutions observed in AV4 and other ORFs (data not shown) showed similar amino acid substitutions (synonymous as well as non-synonymous) at the protein sequence level as depicted in Singh-Pant *et al.* (2012).

As far as start and stop codons are concerned which are very crucial for any protein and mRNA sequence, it is observed that stop codons have second highest probability to remain same (stop codon) due to single base (=nucleotide) substitutions. Highest (equal) probability of conversion of stop codon is into Tyrosine and Tryptophan and after that is the probability of remaining a stop codon only. In the standard genetic codon usage, start codon AUG (encoding Methionine, M) is one of the least probable codon affected by single base substitution. Only codons for phenylalanine and tryptophan have comparable transition probabilities. Nature wise, any single base substitution in start codon leads to an amino acid in one of the three groups of amino acid - Non polar - Aliphatic, Aromatic or Basic-Positively charged. It is interesting to note that while a single base substitution in AUG may lead to one of the six amino acids, with highest probability of 0.33, the probability of any other codon changing into start codon is always less than or equal to 0.11, indicating that creation of new start codon is intrinsically less favored.

## Discussion:

Codons show degeneracy for most amino acids as well as specificity for a few [5]. Owing to the equilibrium maintained in natural systems, there is symmetry in our observations too as per our results. Using a computational biology approach, we have shown that synonymous substitutions are more favored over non-synonymous ones from a bioinformatics/computational biology approach. This further validates the pre-destined or directional nature of amino acid replacement due to single base substitutions. Our results also support earlier works on viral systems indicating that amino acid of similar nature replacing each other (synonymous substitutions) is more favored over non-synonymous substitution [15]. This further emphasizes trends indicated from our results that the probability of a non-polar converting into a polar amino acid is less as compared to that of conversion of non-polar to non-polar amino acid and so on. This stands in support of various studies which depict radical changes (usually destabilizing the three dimensional structure) when in silico changes are done involving non-synonymous substitutions in amino acid sequences. For example, amino acids G, A, V, L, I, P are less probable to be replaced by S, T, C, M, N, Q which are polar in nature, i.e., amino acids of one group tend to be replaced by amino acids of similar nature (i.e. synonymous substitutions). This substantiate the fact that even though codon bias is widespread in organism but as resulting amino acids are mostly same or similar in nature, they are expected to not have any effect on protein structure and function conserving the protein family and preventing its evolution going abrupt. This also keeps the phylogenetic affinities and relationships sacrosanct.

Exceptions are also found but only in the case of Threonine. The calculated probability of SBRs in Threonine is not same in all replacements. From the calculated probabilities and the result output, it is observed that Threonine is the only amino acid with exceptions. Also, the question arises as to why only Threonine has different behavior? Threonine is encoded by only four codons and it shows the maximum replacements,

even though there are some other amino acids which are encoded by more than 4 codons and still show comparatively less number of replacements. This is the second odd behavior shown by the Threonine. The probability of replacement of amino acid by itself is same irrespective of the number of replacements that particular amino acid have. On the basis of this observation it can be assumed that the amino acids are more probable to be replaced by themselves thereby, resisting for the functional defect in the amino acids sequence. But Methionine and Tryptophan are the two amino acids which cannot be replaced by themselves when SBR occur, this is due to the fact that both of these amino acids are encoded by a single codon. As previously mentioned regarding codon bias that some of the codons are used more commonly over others, our result shows that the SBR is directional with every amino acid liable to be replaced in a specific pre-defined style, therefore this may be one of the reasons responsible for more usage of some codons in comparison to others.

## Conclusion:

Probability matrices for the amino acid substitutions have been made in the past by various workers but they were specific to a given protein sequence. It has been to our notice that similar, but not identical, matrices were derived in past based on experimental data which further strengthens present theoretical study. This work presents a novel matrix of amino acid substitutions as predicted by computational approach when single nucleotide substitution takes place. A probability matrix for the genetic codons and the resultant amino acids shows that synonymous changes are more favored over non-synonymous ones. However, the major finding of this work is that when a single base substitution takes place, the probability of outcome varies for different residues and as expected, the one with highest probability gets more often selected over others setting the stage for a direction change of course due to SNS rather than being random. This directional change drifts evolution of the protein structure and function toward an evolutionarily favored state. In this way, we have tried to model the nature of the substitutions whether synonymous or non-synonymous and found it to be directional rather than being very random. As the mandate of the work was to understand the standard genetic codon usage and not the mitochondrial one per se, therefore Mitochondrial Codon system was out of scope of this work. Still, from the trends of [16], we hypothesize that mitochondrial codon system will not be different, but that as of now remains yet to be figured out. We plan to take up the utility of this work beyond the standard genetic codon usage systems as a future work. This work may form ground for further research in this relatively unexplored area eventually enabling us to predict more accurate phylogenetic relationships. Also, in studies pertaining to estimation of rate and nature of recombination in plant and animal viral systems this study has a lot to offer.

## Acknowledgement:

PP<sup>2</sup> and PP<sup>4</sup> acknowledge a Senior Research Fellowship from Department of Biotechnology and CSIR respectively. Authors would also like to acknowledge the intellectual inputs by Dr. M. Zakaria Siddiqui (Australian National University)

## Author Contribution Declaration:

TA, MF and MRA conceived the idea, designed the experiments, write the perl script and performed the experimental analyses; TA, MF, MRA, PP<sup>2</sup> and PP<sup>4</sup> wrote the manuscript. PP<sup>4</sup> provided valuable suggestions and data for applicability of the work.

## Conflict of interest

The authors declare that there is no conflict of interest whatsoever

## References:

- [1] Crick FH *et al.* *Nature* 1961 **192**: 1227 [PMID: 13882203]
- [2] Crick F *Nature*. 1970 **227**: 561 [PMID: 4913914]
- [3] Khorana HG *et al.* *Cold Spring Harb Symp Quant Biol.* 1996 **31**: 39 [PMID: 5237635]
- [4] Woese CR, *Naturwissenschaften.* 1973 **60**: 447 [PMID: 4588588]
- [5] Bossi L & Roth JR, *Nature.* 1980 **286**: 123 [PMID: 7402305]
- [6] Yanofsky C *et al.* *Cold Spring Harb Symp Quant Biol.* 1966 **31**: 151 [PMID: 4866372]
- [7] Newman SA & Bhat R, *J Biosci.* 2008 **32**: 1041 [PMID: 17954964]
- [8] Ray SK *et al.* *J Genet.* 2014 **93**: 613 [PMID: 25572219]
- [9] Archett M, *J Evol Biol.* 2006 **19**: 353 [PMID: 16599911]
- [10] Li WH *et al.* *Mol Biol Evol.* 1985 **2**: 150 [PMID: 3916709]
- [11] Bulmer M, *Genetics.* 1991 **129**: 897 [PMID: 1752426]
- [12] Antezana MA & Kreitman M, *J Mol Evol.* 1999 **49**: 36 [PMID: 10368432]
- [13] Delpont W *et al.* *PLoS Comput Biol.* 2010 **6**: e1000885 [PMID: 20808876]
- [14] Hershberg R & Petrov DA, *Annu Rev Genet.* 2008 **42**: 287 [PMID: 18983258]
- [15] Singh-Pant P *et al.* *Arch Virol.* 2012 **157**: 1217 [PMID: 22437254]
- [16] Yang Z *et al.* *Mol Biol Evol.* 1998 **15**: 1600 [PMID: 9866196]

Edited by P. Kanguane

Citation: Abdullah *et al.* *Bioinformation* 12(3): 98-104 (2016)

**License statement:** This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

