

Open Access

# Comparative Analysis of Non-Synonymous and Synonymous Substitution of Capsid Proteins of Human Herpes Virus

Vipan Kumar Sohpal<sup>1\*</sup>, Apurba Dey<sup>2</sup> and Amarpal Singh<sup>3</sup>

<sup>1</sup>Department of Chemical & Biotechnology, Beant College of Engineering & Technology, Gurdaspur, Punjab, India <sup>2</sup>Department of Biotechnology, National Institute of Technology, Durgapur West Bengal, India <sup>3</sup>Department of Electronics & Communication Engineering, Beant College of Engineering & Technology, Gurdaspur, Punjab, India

### Abstract

The difference between non-synonymous (dN) to synonymous (dS) changes between sequence is computed to assay the direction of evolution. In this paper we compare closely related strains HHV species, codon-by-codon using maximum likelihood analysis and the divergence time of the two sequences. We demonstrate that a substitution model provides evolution hypothesis when comparing closely related species. The effect of transition/transversion ratio and Fischer exact test are on dN-dS (p-distance), and the problems associated with these concepts for HHV, are also discussed. We applied methods on sequence capsid protein of the HHV-1 and HHV-2 for detailed analyses of capsid structures provides the best evidence for an evolution. We conclude that substitution models, dN-dS, divergence time and transition/transversion have critical role to study the evolution.

**Keywords:** HHV; Substitution models; dN-dS; Divergence time; Transition/Transversion rate; Evolution

**Abbreviations:** Fels1981: Felsenstein 1981 model; HKY: Hasegawa-Kishino-Yano model; T-Nei: Tamura-Nei model; GTR: General Time Reversible model; MNJ (P): Modified Nei-Gojobori method (Proportion); MNJ (JC): Modified Nei-Gojobori method (Jukes-Cantor); NG(P):Nei-Gojobori method (Proportion); NG (JC): Nei-Gojobori method (Jukes-Cantor); LWL: Li-Wu-Luo method (Kimura 2-parameter); PBL: Pamilo-Bianchi-Li method (Kimura 2-parameter); Kumar: Kumar method (Kimura 2-parameter)

## Introduction

The human herpes viruses (HHV-1 and HHV-2) are the widespread of the two viral strains that cause chicken pox, herpes zoster (VZV), cancers, and encephalitis in humans. Because the HHV virus rapidly evolves at a fast rate and constant moment, presently there is neither current nor appropriate vaccine against HHV infection. The HHV virus replicates its genetic material through interact with glycoprotein and subsequent DNA maturation resultant infectious diseases to a human being. Proteins of HHV viral strains (UL6, UL18, UL35, UL38 minor capsid proteins and UL19 major capsid) present an extremely high mutation rate that allows it to hinder the response of the immune system as well as the action of broad spectrum antibiotics drugs. At the same time, however, the evolution of HHV provides a powerful mechanism that reveals significant insights into its function and resistance to drugs. Selective pressures across various regions of the viral genome measure the protein evolution rate. Selective pressure (dN/dS) is an indicator used in genetics, to estimate the ratio of the nonsynonymous substitution per nonsynonymous site (dN), to the number of synonymous substitution per synonymous site (dS) (Table 1). Nonsynonymous substitutions to a nucleotide sequence cause a change in the translated amino acid sequence, whereas synonymous substitution does not alter amino sequence. A maximum likelihood method uses probability theory to complete all three steps of approximation method, simultaneously and accurately.

Pond and Frost [1] considered three approaches for estimating the rates of nonsynonymous and synonymous changes at each site in a sequence alignment in order to identify sites under positive or negative selection for human immunodeficiency virus type 1 env and pol genes. Rocha et al. [2] compared the ratio of nonsynonymous (dN) to synonymous (dS) changes between taxa is frequently computed to assay the strength and direction of selection. Li et al. [3] proposed Non- degenerate, twofold degenerate or fourfold degenerate method are for estimating the number of synonymous and nonsynonymous nucleotide substitutions between homologous genes. Nei and Gojobori [4] presented Miyata and Yasunaga's two classic methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Li [5] proposed a new method that gives unbiased estimates the number of substitutions per fourfold degenerate site whereas the old method have two-thirds degenerate site as one-third synonymous and twothirds nonsynonymous. Muse and Gaut [6] constructed and applied Likelihood-ratio versions of the relative-rate

s	Number of synonymous sites
n	Number of nonsynonymous sites
S	Number of synonymous substitutions
N	Number of synonymous sites
dS	Synonymous substitution rate
dN	Nonsynonymous substitution rate
dN/dS	Selective strength of sequence
t	Divergence time between two sequences
t,	Transition /transversion ratio
dN-dS	Difference in numbers of nonsynonymous and synonymous substitutions per site
	(p-distance)
Р	Probability of rejecting the null hypothesis of neutral evolution

 Table 1: Symbols and abbreviation used for synonymous and non synonymous substitution used in Molecular evolutionary genetic analysis software.

\*Corresponding author: Vipan Kumar Sohpal, Department of Chemical & Biotechnology, Beant College of Engineering & Technology, Gurdaspur, Punjab, India, E-mail: vipan752002@gmail.com

Received June 22, 2012; Accepted August 07, 2012; Published August 09, 2012

**Citation:** Vipan Kumar S, Apurba D, Amarpal S (2012) Comparative Analysis of Non-Synonymous and Synonymous Substitution of Capsid Proteins of Human Herpes Virus. J Proteomics Bioinform 5: 172-176. doi:10.4172/jpb.1000231

**Copyright:** © 2012 Vipan Kumar S, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

test for the complete chloroplast DNA sequences. Goldman and Yang [7] presented codon-based model for the evolution of protein-coding DNA sequences, and Markov process is used to describe substitutions between codons. Comeron proposed a method for estimating the numbers of synonymous (Ks) and nonsynonymous (Ka) substitutions per site using Li's and Pamilo and Bianchi's method. Ina conducted simulation to evaluate the accuracies of the new methods, NG, MY, Li, Wu, and LWL, and PBL [8,9]. Yang and Nielsen [10] compared the approximate methods and maximum likelihood method and found that ML superior to approximate methods. Zhang et al. [11] proposed a new method (MYN), a modified version of the Yang-Nielsen algorithm (YN), for evolutionary analysis of protein-coding sequences in general. They observed that MYN has minimal deviations when parameters vary within normal ranges defined by empirical data. Zhang et al. [11] calculated the KaKs using a software package that calculates nonsynonymous (Ka) and synonymous (Ks) substitution rates through model selection and model averaging [12]. Tamura et al. [13] developed Molecular Evolutionary Genetics Analysis version 5 (MEGA5), software for mining online databases, building sequence alignments and phylogenetic trees. It is a collection of maximum likelihood (ML) analyses for inferring evolutionary trees, selecting bestfit substitution models, inferring ancestral states and sequences along with probabilities, and estimating evolutionary rates site-by-site. Corey and Wald [14] estimated 25 to 65% of pregnant women in the United States have genital infection with herpes simplex virus type 1 (HSV-1) or HSV type 2 (HSV-2). Sohpal et al. [15] compare that herpes simplex virus is a serious threat to neonatal and pregnant women. Kimberlin [16] worked on the life-threatening nature and the tremendous damage that surviving infants can incur, for differential diagnosis. McGeoch et al. [17] explained the source of herpes viruses exist, but detailed analyses of capsid structures provide the best evidence for a common origin of the three groups.

In this paper, we compare four substitution methods for detection of dN/dS and improved method, in which transitional/transversion rate is not assume to occur with an equal rate. We also review the z selection test using Fischer exact test involved in estimating dN-dS, explain differences between the two methods, and provide comparative results for divergence time of HHV strains.

# Methodology

Synonymous and non synonymous substitutions are use to be to compare the evolution of protein-coding sequences of human herpes simplex virus. The performance and comparison of substitution models evaluate through generate of simulated model under different conditions. We simulated sequences according to species/proteins that were derive from UniProt database. Major capsid protein of B9VQE6\_HHV1, D3YP82\_HHV1, D3YPH6\_HHV1, P89442\_HHV2H and PRTP\_HHV2 retrieved from UniProt protein database. Similar way the sequence retrieved for human herpes simplex virus (UL-35/18, UL38/6 and UL-19) of mutant species. We first used protein database that contain codons of major capsid protein, assuming that an approximate method should not deviate too far from of dN/dS ratio. In addition to dN/dS ratio, other parameters were set in simulation, including the transition/transversion ratio (t<sub>r</sub>), divergence time and normalized p value.

# Results

# dN/dS and dN-dS

Selective strength, reflected in dN/dS, differs from protein to

protein sequence, some substitution model use for major capsid protein analyses for the selection at codons level (via HyPhy). We examined the effects of substitution model (Fels1981, HKY, T-Nei and GTR) with a set of UL-19 protein of human herpes simplex virus strains of sequences, three ( $\omega$ ) average values estimate (1.2463, 1.3001 and 1.2977), and fourth value of GTR model is same as observed in T-Nei. The average estimates of dN, dS, and normalized P values compute with same four substitution models and as well as expected values of dN and dS (Table 2). High dN/dS ( $\omega$ ) ratio found on the comparison in the entire model used for selection at codons level is greater than one implies positive darwinian selection, the values lie in the region of 1.2463 to 1.3001 (Figure 1). Similarly, the higher and positive value of (dN-dS) for the test statistic indicates an overabundance of nonsynonymous substitutions in MEGA. The highest value dN-dS observe 190.97 (GTR and T-Nei model) lowest 150.99 (Fels1981) and intermediate 186.14 (HKY model). Similar results readily obtain for Normalized dN-dS, which measure in expected substitutions per site. It is useful for making comparisons across data sets of human herpes simplex virus. dN/dS ratio, dN-dS, normalized dN-dS favour the positive selection hypothesis.

# Z-test of selection

Higher dN/dS and dN-dS, substantial proof to evaluate positive selection is operating on a gene are to compare the relative abundance

$\begin{array}{c} \text{Parameters} \rightarrow \\ \text{Model} \downarrow \end{array}$	Syn sites (S)	Nonsyn sites (N)	dS	dN	dN-dS	Normal- ized dN-dS
Felsenstein 1981	328.80	937.65	649.69	809.69	159.99	45.72583
Hasegawa-Kishino-Yano	334.64	931.83	620.28	806.43	186.14	53.73541
Tamura-Nei	327.92	937.20	641.32	832.27	190.94	52.07271
General Time Reversible	327 92	937 20	641.32	832 27	190 94	52 07271

**Table 2:** Estimate of S, N, dS, dN and normalized (dN-dS) value selection at codons (via HyPhy) using maximum likelihood approach of major capsid proteins of HHV strain.



of nonsynonymous and synonymous substitutions that have occurred in the sequences. In MEGA, codon based test of positive selection perform for analysis between sequences using the mathematical formula Z=(dN - dS) /  $\sqrt{(Var^*(dS) + Var^*(dN))}$ . We compared five different methods for Z test under constant tr. NG (P), NG (JC), LWL, PBL and Kumar models use to simulate the all the viral capsid protein of human herpes simplex virus for positive (dN>dS) and null (dN=dS) hypothesis. Tables 3 and 4 show the results of minor and major capsid protein of HHV strains on the positive and the null hypothesis sets. The results from table 3 are qualitative similar for the UL-35, UL-38 UL-18 and proteins, giving confidence that they indicate a general trend of positive evolution. Although the values for UL-06 and UL-19 (430 codons) exceed than 0.05 but comparatively lower than the corresponding value found in the null hypothesis, conditions in table 4. The results show the probability of rejecting the null hypothesis of strict-neutrality in favor of the alternative hypothesis (dN > dS) because values of P less than 0.05 consider significant in MEGA.

For UL-18, the value of the estimated P generated entire methods (dN>dS) varies from 0.065 to 0.078, and values for the same protein (dN=dS) lie in the region of 0.14 to 0.159 far away from the probability value. When UL-35 protein, simulated under both hypothesis, values lie below 5% level. UL-38 protein, provides the p-values in satisfactory zone under positive evolution, while the output for the same protein under neutral evolution on higher-side. UL-06 and UL-19 proteins p-value also moving towards positive evolution, so overall performs better, especially when dN> dS for most of the parameter combinations tested.

## Effects of Transition/Transversion ratio (t\_)

We examined the effects of transition/transversion ratio with a set of UL-35 (codons 112), UL-18 (codons 318) and UL- 6 (codons 638) protein using MEGA. We plotted p-distance that calculates with two different methods (MNJ (P) and MNJ (JC)) under two different conditions of evolution. According to the results, codon frequencies have influence on both the methods especially in UL-18 sequence. The transition/transversion effect study in the range of 0.95 to 3.55. Although neutral evolution is more biased than positive evolution, but both have nearly parallel trends with an increasing t<sub>r</sub>. These results are in substantial agreement with z-selection test for codons. MNJ(P) gives the p-distance range from 0.064 to 0.072 and 0.124 to 0.154 for positive and null evolution models. The values observe for MNJ (JC) method also exists in the domain of 0.062 to 0.078 and 0.129 to 0.151 for respective conditions. Figure 2 depicts that t\_=2.05 have maximum p value 0.078 for both methods and t=0.95 have the highest value 0.165. The results (Figure 3) for protein UL-35, t\_ratio 0.1 proved to give the lowest p-distance and observed in the range of  $\approx 0.1 - 0.2$  for all the conditions of models. With an increasing transition/transversion (lower to higher), the results are changed moderately The t\_ratio for UL-35 restricted to 0.1 to 0.7, exceed than the upper limit no influence observe on p-distance. We have made several observations to avoid divergence and to maintain consistency in evolution models. UL-6 protein has provided the uniform difference in both the models for four conditions. This is quite remarkable, that neutral and positive evolution models show a comparatively similar performance. The transition/ transversion ratio varied from 0.6 to 1.3 correspondingly of that p-distance from higher to a lower region.

# Fisher's Exact Test of Neutrality

The average number numbers of synonymous and nonsynonymous

Protein → Method ↓	UL-18	UL-35	UL-38	UL-06	UL-19
Nei-Gojobori method (Proportion)	0.073	0.001	0.057	0.563	0.514
Nei-Gojobori method (Jukes-Cantor)	0.078	0.001	0.049	0.589	0.541
Li-Wu-Luo method (Kimura 2-parameter)	0.076	0.007	0.058	0.0	0.0
Pamilo-Bianchi-Li method (Kimura 2-parameter)	0.065	0.015	0.048	0.0	0.0
Kumar method (Kimura 2-parameter)	0.065	0.011	0.072	0.438	0.438

 
 Table 3: Estimation of codons (Z test) for minor and major capsid proteins of HHV strains at constant tr & under positive evolution (dN>dS).

Protein → Method ↓	UL-18	UL-35	UL-38	UL-06	UL-19
Nei-Gojobori method (Proportion)	0.159	0.001	0.109	1.0	1.0
Nei-Gojobori method (Jukes-Cantor)	0.141	0.002	0.18	1.0	1.0
Li-Wu-Luo method (Kimura 2-parameter)	0.131	0.014	0.11	0.0	0.0
Pamilo-Bianchi-Li method (Kimura 2-parameter)	0.128	0.036	0.961	0.0	0.0
Kumar method (Kimura 2-parameter)	0.14	0.016	0.142	0.882	0.877
Table 4: Estimation of codons (7 test) for minor and major cansid protoins of					

 $\label{eq:table_transform} \begin{array}{l} \textbf{Table 4:} \ \text{Estimation of codons} \ (Z \ \text{test}) \ \ \text{for minor and major capsid proteins of } \\ \text{HHV strains at constant } t_{\text{r}} \ \ \& \ \text{under neutral evolution} \ (dN=dS). \end{array}$ 

substitutions are medium, but codons number of minor and major capsid proteins of human herpes simplex virus (UL-19=430 codons, UL-18=318 codons, Ul-35=112 codons, UL-38=458 codons and UL-06=638) they are small. In this case, tests of selection can be conducted to examine the null hypothesis of the neutral evolution by Nei-Gojobori (NG) and Modified Nei-Gojobori (MNJ) methods. We run the simulation for MNG because it considers transition/transversion rate for the direct computation of the numbers of nonsynonymous and synonymous differences. UL-19 major capsid protein use to analysis of difference (dN-dS) by Fischer exact test. Figure 5 reveals that lower t ratio favors the highest distance, but as we move from lower to higher transition to transversion ratio, the significant t<sub>r</sub> becomes negligible. The highest value of p-distance observed at lowest t<sub>r</sub> value is equal to 0.228. Figures 2, 3 and 4 are qualitatively similar for the three proteins, giving confidence that they indicate of positive evolution. This is further confirming by Fischer exact test on fourth protein of HHV (Figure 5), which again gives similar, results, as would be expected in under dN>dS conditions.

# Effect of Divergence Time

Divergence time in synonymous and non synonymous substitution is a function of nucleotide substitutions per codon,  $t=(dS \times 3^*S+dN \times 3^*N)$ (S+N). We again used the major capsid protein for simulations having 430 codons), and tested four different models namely Fels1981, HKY, T-Nei and GTR. We plotted average estimates of with Fels 1981, HKY, T-Nei and GTR against divergence time (Figure 6). Both T-Nei and GTR have a nearly parallel overall trend. Fels1981 closely related to GTR model having variation about 2% only, while HKY model deviates 4% from T-Nei model. With t increasing, GTR and T-Nei tends to give a better estimate for dN-dS for positive evolution, and biased estimate for neutral selection, where t has no significant influence on purifying evolution Figure 1 and Figure 6 predict the co-relation that as dN/dS ratio increase, the divergence times also increase. The results of dN/dS replicate in divergence time, to strength the positive evolution.

# Discussion

The number of substitution/mutation models used in this paper

allows for comparative study of evolution in minor and major capsid protein of HHV strains, using maximum likelihood method. The main thrust is to analysis the dN-dS, divergence time and transition /transversional rate of nucleotide for the proposed antibiotics drug. Comparative analyses indicate that ignoring t<sub>r</sub> often results in closer estimates of dN/dS. HKY is more robust in simulations and real datasets by comparison with other methods, especially with Fels1981. T-Nei and GTR have quantitatively similar results. The entire models have dN/dS is greater than 1 predict, diversity of the amino acid (minor and major capsid proteins) is favour, likely due to the fitness advantage provided by the mutations. In Z -selection of codon without considering t all proteins of HHV strains pronounced positive evolution on the basis of difference in numbers of nonsynonymous and synonymous substitutions per site (p-distance). Since nonsynonymous substitutions are more likely to occur than synonymous, related to more drastic reduction of transitions than transversions this estimation becomes severer for positive selection. Therefore, it is indispensable to take account of transition/transversion rates for accurately capturing evolutionary information, when unequal transitional rates among



Figure 2: Estimates of p distance for unique long region (UL-18) of human herpes simplex virus with two different methods and conditions of evolution by varying the transition/transversion ratio.



**Figure 3:** Estimates of p distance for unique long region (UL-35) of human herpes simplex virus with two different methods and conditions of evolution by varying the transition/transversion ratio.







compare sequences exist. We use maximum likelihood method for unequal transitional/transversion rates that allow model evaluation for choosing a better suit to a dataset and therefore, MEGA allows us to obtain more reliable estimates in a single step. UL-35, UL-18 and UL-6 simulation result show that lowest transition/transversion ratio favour the positive evolution, which means non synonymous substitution is more dominant than synonymous. The closeness on the basis of p-distance between three proteins is as UL-6< UL-35< UL-18 signifies that evolution/mutation rate is higher in UL-18 as compare to UL-6, while UL-35 lies in an intermediate zone. The trend sometimes is more obvious, resulting in positive evolution. Therefore, to acquire more reliable and ultimate estimates on dN and dS Fischer exact test simulated for major capsid protein. It produces similar results of p-distance for the most of the parameter combinations. Closer estimation observed at higher transversion rate and underestimate when transition dominates in Fischer exact test. So dN-dS are more realistic method for selection of strength under lower transition/transversion ratio.

Divergence time and number of codons influence the dN and dS parameter all are deduce (Fels 1981, HKY, GTR and T-Nei) from



fourfold-degenerate sites at the third codon positions and non-coding sites. Lower dN-dS and multiple substitutions lead to higher divergence time. The comparatively highest time observed in T-Nei and lowest HKY model while Fels 1981 have intermediate divergence time. From the results, it shows that divergence time is inversely proportional to dN-dS. It means non synonymous substitution is dominating for positive evolution. From the above discussion based on estimate selection at Codons (via HyPhy), Z-test of selection, Fisher's exact test of Selection, and results also agree well with positive evolution and distance estimation between two different strains of HHV.

# Conclusion

We compared Fels1981 with other substitution methods, especially with HKY and T-Nei, by examining major capsid protein sequences and found that dN/dS ratio exceed than one, towards positive selection. NG (JC), LWL, PBL and Kumar methods results also suggest that even in the absence of transition/transversion, synonymous and nonsynonymous substitution rates inclined to positive evolution. The comparative results deriving from consistent (Z and Fisher test) and continuous analysis throughout the paper indicate that lower transition/transversion rates lead to higher dN-dS under positive evolution and rejecting the null hypothesis. In addition, divergence time also influence by substitution models and dN-dS.

### Acknowledgment

This work was supported by Punjab Technical University of India and Dr Sudhir Kumar University of Arizona.US for MEGA 5.5 software. It is available free of charge on http://www.megasoftware.net

### **Author Contribution**

Sohpal VK performed computer simulation and comparative analysis of viral capsid proteins of HHV and drafted the manuscript. Dey A and Singh A supervised field of work and revised the manuscript. Both authors read carefully and approved the manuscript.

### References

- Kosakovsky Pond SL, Frost SD (2005) Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Mol Biol Evol 22: 1208-1222.
- Rocha EP, Smith JM, Hurst LD, Holden MT, Cooper JE, et al. (2006) Comparisons of dN/dS are time dependent for closely related bacterial genomes. J Theor Biol 239: 226-235.
- Li WH, Wu CI, Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2: 150-174.
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3: 418-426.
- Li WH (1993) Unbiased estimation of the Rates of synonymous and nonsynonymous substitution. J Mol Evol 36: 96-99.
- Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol Biol Evol 11: 715-724.
- Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11: 725-736.
- Comeron JM (1995) A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. J Mol Evol 41: 1152-1159.
- 9. Ina Y (1995) New methods for estimating the numbers of synonymous and nonsynonymous substitutions. J Mol Evol 40: 190-226.
- Yang Z, Nielsen R (2000) Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. Mol Biol Evol 17: 32-43.
- 11. Zhang Z, Li J, Yu J (2006) Computing Ka and Ks with a consideration of unequal transitional substitutions. BMC Evol Biol 6: 44.
- Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, et al. (2006) KaKs\_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics 4: 259-263.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28: 2731-2739.
- 14. Corey L, Wald A (2009) Maternal and neonatal herpes simplex virus infections. N Engl J Med 361: 1376-1385.
- Sohpal VK, Dey A, Singh A (2010) Sequence alignment and phylogenetic analysis of human herpes simplex virus (HHV) using bioinformatics tool: a review. Int J Comput Biol Drug Des 3: 68-88.
- 16. Kimberlin DW (2007) Herpes simplex virus infections of the newborn. Semin Perinatol 31: 19-25.
- McGeoch DJ, Rixon FJ, Davison AJ (2006)Topics in herpesvirus genomics and evolution. Virus Res 117: 90-104.