

Computational Analysis of Distance and Character based Phylogenetic Tree for Capsid Proteins of Human Herpes Virus

Vipan Kumar Sohpal^{1*}, Apurba Dey² and Amarpal Singh³

¹Department of Chemical & Biotechnology, Beant College of Engineering & Technology, Gurdaspur Punjab, India

²Department of Biotechnology, National Institute of Technology, Durgapur West Bengal, India

³Department of Electronics & Communication Engineering, Beant College of Engineering & Technology, Gurdaspur Punjab, India

Abstract

Molecular phylogenetic is a fundamental aspect of evolutionary analysis and depends on distance & character based methods. In this paper, we compare the viral capsid proteins of HHV to analyze the relationship among proteins using substitution models, phylogenetic model with exhaustive search and ME techniques. The effect of Poisson correction with shape parameter on NJ and UPGMA trees also analyze. We show by extensive computer simulation that phylogenetic tree is the reflection of substitution distance. The effect of max-mini branch & bound method and mini-mini heuristic model and log likelihood associated with character based tree also discussed. We applied ML and MP for perfectly analysis of proteins relationship. We conclude that substitution models, shape parameter, search level and SBL have a critical role to reconstruct phylogenetic tree. Molecular clock study shows that χ^2 value is higher in closely as compare distant related proteins.

Keywords: Human herpes virus (HHV); Phylogenetic methods (NJ, UPGMA, ML and MP); Sum of branch length (SBL); Molecular clock

Introduction

Molecular phylogenetic analyses consist of three stages: (1) selection and multiple sequence alignment of homologous protein (2) substitution model of amino acid evolution, (3) tree building and tree evaluation. The phylogeny of protein assess through various software such as; Paup, Phylip and Mega. These contain numbers of proteins as well as nucleotide substitution models for evolutionary analysis as mentioned in step 2 of molecular phylogenetic. Similarly, the tree construction and boot strap evaluation methods are also available in analysis tool. In this paper, we focus on MEGA software for Phylogenetic analysis of viral capsid protein of human herpes virus (HHV), which causes chicken pox, herpes zoster (VZV), cancers, and encephalitis in the human being. Previous attempts (Literature) have yielded comparative analysis of phylogenetic methods.

Alfaro and Huelsenbeck [1] used a simulation approach to assess the performance of both Bayesian and AIC-based methods and compared it. Geddes et al. [2] compared the phylogenetic trees of core erythritol catabolic genes with species phylogeny provides evidence that is consistent with these loci having been horizontally transferred from the alpha-proteobacteria into both the beta and gamma - proteobacteria. Felsenstein [3] applied the maximum likelihood techniques to the estimation of evolutionary trees from nucleic acid sequence data. Felsenstein [4] developed statistical method bootstrap and show significant evidence for a group if it is defining by three or more characters. Kenney and Rosenzweig [5] study indicates that Mbn-like compounds may be more widespread than previously thought bactins. Guindon et al. [6] introduced a new algorithm to search the tree space, and they used parsimony criterion to filter out the least promising topology. Sourdis and Nei [7] studied the relative efficiencies of the maximum parsimony (MP) and distance-matrix methods in obtaining the correct tree using computer simulation. Krause et al. [8] worked on metagenomics, which is providing striking insights into the ecology of microbial communities. They developed massively parallel 454 pyrosequencing technique that gives the opportunity to rapidly obtain metagenomic sequences at a low cost and without cloning bias. The phylogenetic analysis of the short reads produced represents a

significant computational challenge. Chen et al. [9] compared with GI, they found that the GII geno-group had four deletions and two special insertions in the VP1 region.

Posada and Buckley [10] focused on most commonly implemented model selection approach, the hierarchical likelihood ratio test, is not the optimal strategy for model selection but the Akaike Information Criterion (AIC) and Bayesian methods offer to cut edge. Saitou and Nei [11] proposed an alternative for reconstructing phylogenetic trees from evolutionary distance data. The principle of this method is to find pairs of operational taxonomic units which minimize the total branch length at each stage of clustering. Sohpal et al. [12] reviewed the bioinformatics software and analyzed that MEGA5, is user-friendly software for mining online databases, building sequence alignments and phylogenetic trees, and using methods of evolutionary bioinformatics in basic biology, biomedicine, and evolution. Kumar and Filipiski [13] compared the traditional approach to phylogeny reconstruction, in a set of homologous sequences which represent to a character. Sullivan et al. [14] conducted a direct evaluation using simulations and compared the accuracy of phylogenies estimated using full optimization of all model parameters on each tree evaluated to the accuracy of trees estimated via successive approximations. Tamura et al. [15] developed MEGA5 which is a collection of Maximum Likelihood (ML) analyses for inferring evolutionary trees, selecting best fit substitution models inferring ancestral states and sequences and estimating evolutionary rates site-by-site. Yang [16] performed the simulation with assumptions underlying the maximum parsimony (MP) method of phylogenetic tree reconstruction intuitively examine.

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

Received March 23, 2013; Accepted March 27, 2013; Published April 03, 2013

Citation: Sohpal VK, Dey A, Singh A (2013) Computational Analysis of Distance and Character based Phylogenetic Tree for Capsid Proteins of Human Herpes Virus. J Data Mining Genomics Proteomics 4: 128. doi:10.4172/2153-0602.1000128

Copyright: © 2013 Sohpal VK, 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.

The results of simulated work appeared to support the intuitive. Yang [17] show that failure to account for rate variation can have drastic effects, leading to biased dating of speciation events, biased estimation of the transition/transversion rate ratio, and incorrect reconstruction of phylogenies.

In this paper, we focus on MEGA software for reconstruction of an evolutionary tree of viral capsid protein of human herpes virus (HHV). The objective of this paper to analyze the relationship between protein using distance and character based methods. We also focus on the poisson correction distance substitution with shape parameter to develop the relatedness of proteins using SBL and Log-likelihood parameter.

Results

Effect of substitution models

Evolutionary distances are a fundamental tool for the study of molecular evolution, phylogenetic reconstruction and the estimate of divergence time. Here, we use the three substitution models p-distance, number of differences and poisson model for estimating the evolutionary divergence between the viral capsid of proteins and

distance use as a parameter to compare substitution models of proteins. Table 1 gives the data for distance versus 12 viral capsid proteins of human herpes virus for three models. The analysis shows that distance between Triplex capsid protein HHV2H and HHV2G has minimum distance in all the substitution models. The distance divergences between these two strains are 0.019, 2.00 and 0.019 for the p-distance, number of amino acid and poisson models respectively. The second lowest distance values observe in major capsid protein HHV1 and HHV2H are 0.093, 10.00 and 0.108 for the p-distance, number of amino acid and poisson models. On the other hand, the highest divergence found in portal protein HHV2H and triplex capsid protein HHV2H are 0.935, 100 and 97.89 for the same evolutionary distance models. Table 1 shows that the results are qualitative consistent irrespective of model. Lower divergence (Triplex capsid protein HHV2H and HHV2G) signify the closely related proteins and a substantial increase in the estimated distance for the divergent sequences (portal protein HHV2H and triplex capsid protein HHV2H). These substitution models lead to the creation of trees with different branch lengths and topologies. The effects of substitution models of amino acid on phylogenetic trees for 12 capsid proteins detect using the neighbor-joining method, which uses the distance information that is present in table 1. The trees made using figure 1 the p-distance and figure 2 the Poisson correction having

| | [1] | [2] | [3] | [4] | [5] | [6] | [7] | [8] | [9] | [10] | [11] |
|-----------------------------------|-------|--------------|--------------|-------|-------|-------|-------|-------|-------|-------|-------|
| [1] Triplex_Capsid_Protein_HHV11 | | | | | | | | | | | |
| [2] Triplex_Capsid_Protein_HHV2G | 0.421 | | | | | | | | | | |
| [3] Triplex_Capsid_Protein_HHV2H | 0.43 | 0.019 | | | | | | | | | |
| [4] Triplex_Capsid_Protein_HHV2 | 0.879 | 0.897 | 0.888 | | | | | | | | |
| [5] Capsid_Protein-II_HHV11 | 0.888 | 0.925 | 0.925 | 0.916 | | | | | | | |
| [6] Capsid_Protein-II_HHV2H | 0.888 | 0.907 | 0.907 | 0.897 | 0.168 | | | | | | |
| [7] Capsid_Protein_HHV1 | 0.822 | 0.85 | 0.86 | 0.888 | 0.888 | 0.879 | | | | | |
| [8] Capsid_Protein_HHV2 | 0.822 | 0.813 | 0.822 | 0.86 | 0.85 | 0.841 | 0.28 | | | | |
| [9] Portal_Protein_HHV11 | 0.925 | 0.925 | 0.916 | 0.916 | 0.935 | 0.925 | 0.907 | 0.925 | | | |
| [10] Portal_Protein_HHV2H | 0.925 | 0.935 | 0.935 | 0.916 | 0.925 | 0.916 | 0.907 | 0.916 | 0.112 | | |
| [11] MajorCapsid_Protein_HHV2H | 0.897 | 0.907 | 0.916 | 0.897 | 0.804 | 0.766 | 0.907 | 0.907 | 0.907 | 0.916 | |
| [12] Major_Capsid_Protein_HHV1 | 0.907 | 0.916 | 0.925 | 0.888 | 0.794 | 0.757 | 0.897 | 0.897 | 0.879 | 0.888 | 0.093 |
| [1] Triplex_Capsid_Protein_HHV11 | | | | | | | | | | | |
| [2] Triplex_Capsid_Protein_HHV2G | 45 | | | | | | | | | | |
| [3] Triplex_Capsid_Protein_HHV2H | 46 | 2 | | | | | | | | | |
| [4] Triplex_Capsid_Protein_HHV2 | 94 | 96 | 95 | | | | | | | | |
| [5] Capsid_Protein-II_HHV11 | 95 | 99 | 99 | 98 | | | | | | | |
| [6] Capsid_Protein-II_HHV2H | 95 | 97 | 97 | 96 | 18 | | | | | | |
| [7] Capsid_Protein_HHV1 | 88 | 91 | 92 | 95 | 95 | 94 | | | | | |
| [8] Capsid_Protein_HHV2 | 88 | 87 | 88 | 92 | 91 | 90 | 30 | | | | |
| [9] Portal_Protein_HHV11 | 99 | 99 | 98 | 98 | 100 | 99 | 97 | 99 | | | |
| [10] Portal_Protein_HHV2H | 99 | 100 | 100 | 98 | 99 | 98 | 97 | 98 | 12 | | |
| [11] MajorCapsid_Protein_HHV2H | 96 | 97 | 98 | 96 | 86 | 82 | 97 | 97 | 97 | 98 | |
| [12] Major_Capsid_Protein_HHV1 | 97 | 98 | 99 | 95 | 85 | 81 | 96 | 96 | 94 | 95 | 10 |
| [1] Triplex_Capsid_Protein_HHV11 | | | | | | | | | | | |
| [2] Triplex_Capsid_Protein_HHV2G | 0.965 | | | | | | | | | | |
| [3] Triplex_Capsid_Protein_HHV2H | 1.012 | 0.019 | | | | | | | | | |
| [4] Triplex_Capsid_Protein_HHV2 | 29.43 | 40.78 | 34.41 | | | | | | | | |
| [5] Capsid_Protein-II_HHV11 | 34.41 | 75.68 | 75.68 | 60.23 | | | | | | | |
| [6] Capsid_Protein-II_HHV2H | 34.41 | 49.09 | 49.09 | 40.78 | 0.221 | | | | | | |
| [7] Capsid_Protein_HHV1 | 13.91 | 19.57 | 22.22 | 34.41 | 34.41 | 29.43 | | | | | |
| [8] Capsid_Protein_HHV2 | 13.91 | 12.56 | 13.91 | 22.22 | 19.57 | 17.36 | 0.459 | | | | |
| [9] Portal_Protein_HHV11 | 75.68 | 75.68 | 60.23 | 60.23 | 97.98 | 75.68 | 49.09 | 75.68 | | | |
| [10] Portal_Protein_HHV2H | 75.68 | 97.98 | 97.98 | 60.23 | 75.68 | 60.23 | 49.09 | 60.23 | 0.134 | | |
| [11] MajorCapsid_Protein_HHV2H | 40.78 | 49.09 | 60.23 | 40.78 | 11.39 | 7.998 | 49.09 | 49.09 | 49.09 | | |
| [12] Major_Capsid_Protein_HHV1 | 49.09 | 60.23 | 75.68 | 34.41 | 10.37 | 7.379 | 40.78 | 40.78 | 29.43 | 0.108 | 0.108 |

Table 1: Distance data versus 12 viral capsid proteins of human herpes virus for three substitution models.

gamma distribution 0.52. Branch lengths are in units of evolutionary distances used to infer NJ tree. The sum of the branch lengths was 313.93 in figure 1 and 3.156 in figure 2 with bootstrap consensus. Figure 1 of NJ (p-distance) tree shows that 100% of the bootstrap trials, in entire capsid proteins support as being a clade. Figure 2, the clade containing portal protein HHV11 and HHV2H support 98% of the bootstrap replicates. While the capsid proteins of HHV11 and HHV2H have 28% of bootstrap tree, this means that in 72% of the bootstrap trees, triplex capsid protein of HHV2 joined that group of proteins. The minimum divergence observed in table 1, figures 1 and 2 between triplex capsid protein HHV2H and HHV2G and distant related proteins are portal protein HHV2H & triplex capsid protein HHV2H. From that we can infer that there is not strong support for a distinct, closely relate major/capsid group that shared an ancestor with the portal protein. So further analysis of distance and character based phylogenetic trees reconstruction is required.

Distance based phylogenetic tree

The distance based phylogenetic methods measure the distance from the dissimilarity observe between pairs of sequences, computed on the basis of sequence alignment. The main assumption of distance-based methods is homologous and that tree branches can sum-up. The UPGMA method and neighbor joining methods based on the clustering algorithm use in this paper for 12 viral capsid proteins of HHV. The tree shown in figure 3 is output of the neighbor joining approach, and for which the Poisson-corrected distances shown in table 1. On comparing, the both (Figure 3 and Table 1) distance matrix, the two closest OTUs of the distance matrix (triplex capsid from HHV2G and HHV2H) has the lowest taxon separation in the

neighbor joining tree. The second lowest distance separation group (major capsid protein HHV1 and HHV2H) has the next narrow taxon separation on 100% bootstrap consensus. Those 12 OTUs distance relationships are visualize in the phylogenetic tree. The Sum of Branch Length (SBL) is the second parameter for analysis of distance based phylogenetic tree. Neighbor joining method minimizes the sum of branch length at each stage of clustering OTU's that is a function of the gamma distribution (shape parameter). We construct the NJ tree under different gamma distribution range from 0.5 to 10. Figure 3a-3d shows the phylogenetic tree with poisson correction and shape parameter of 1.0, 2.5, 4.0 and 5.0. Quantitatively lowest taxon separation observes in triplex capsid protein HHV2G and HHV2H in the entire phylogenetic tree irrespective of the shape parameter values. Contrary to that bootstrap replicate reach to 100% from 22% on increasing the gamma distribution. The higher or 100% bootstrap consensus allows measuring the distance more accurately than low bootstrapping value. From figure 3d on the basis of taxon separation, the proteins can be arranged as triplex capsid<major capsid<portal<capsid. SBL for four shape parameter is 24.20, 10.82, 9.096 and 8.605. The sum of branch length decreases with an increase of gamma parameter. Lowest SBL & highest bootstrap analysis is in the favour the neighbor joining method.

From figure 4 similar and consistent results observe for taxon separation as depicted using neighbor joining method. Lowest taxon separation 0.0095 found in triplex capsid proteins OTU while highest observe in the capsid protein OTU's. The bootstrap consensus 100% in entire phylogenetic trees of UPGMA because the assumption that all taxa evolve at a constant rate and that they are equally distant from the root. The overall branch length also goes down with the rise in the gamma parameters. The SBL for the four shape parameter 24.83, 10.82, 9.096 and 8.605 which are also close to neighbor joining method results. The closeness in the SBL motivates to focus on comparative analysis of both distance based methods. The NJ and UPGMA method's SBL varied with gamma distribution in close fashion. Therefore, the sum of branch length is inversely proportional to gamma parameter for viral capsid proteins. At lowest shape parameter (0.5), highest SBL (137.35 and 127.47) observe for UPGMA and NJ respectively. On the other hand at highest shape parameter (10), lowest SBL for UPGMA and NJ detect (7.801 and 7.72). Overall the SBL values in both methods overlap each other as shown in figure 5. These results are in support of closeness between distance methods.

Comparison of tree topologies from clustering based algorithm with optimality-based algorithms to select one that has the best fit between estimated distances in the tree and the actual evolutionary distances. Experiment results shows that at lower gamma parameter, the (0.5) build the difference between SBL (clustering algorithm and ME) is approximately 12.54%, but at a higher value of gamma it, reduce to less than one percent. These indicate that difference between SBL at a higher value of the shape parameter in both the algorithms is like for viral capsid proteins.

Character based phylogenetic tree

Character-based methods are a measure the sequence characters rather than on pairwise distances. In this section, we analyze the viral capsid data with the maximum parsimony (MP) and maximum likelihood (ML) methods. In MP method, we aligned the sequence for informative, conserved and variable sites in all the 12 OTU's of proteins and observe that they have 43.75%, 50.12% and 50.00% sites respectively. On the basis of informative and non informative site, we focus max-mini branch-&-bound heuristic algorithm to reconstruct character based phylogenetic tree. The MEGA software using this

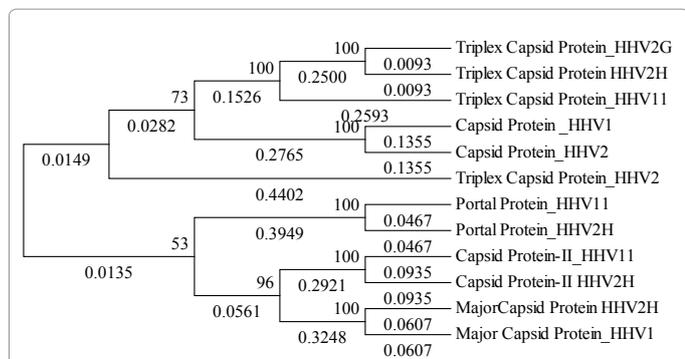


Figure 1: Neighbor joining tree with poisson correction for 12 viral capsid proteins of HHV.

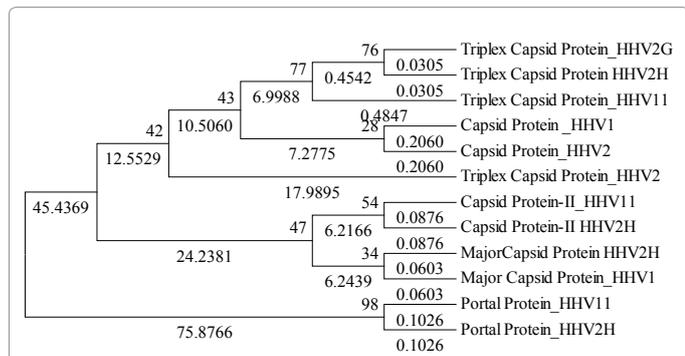


Figure 2: Neighbor joining tree with poisson correction for 12 viral capsid proteins of HHV.

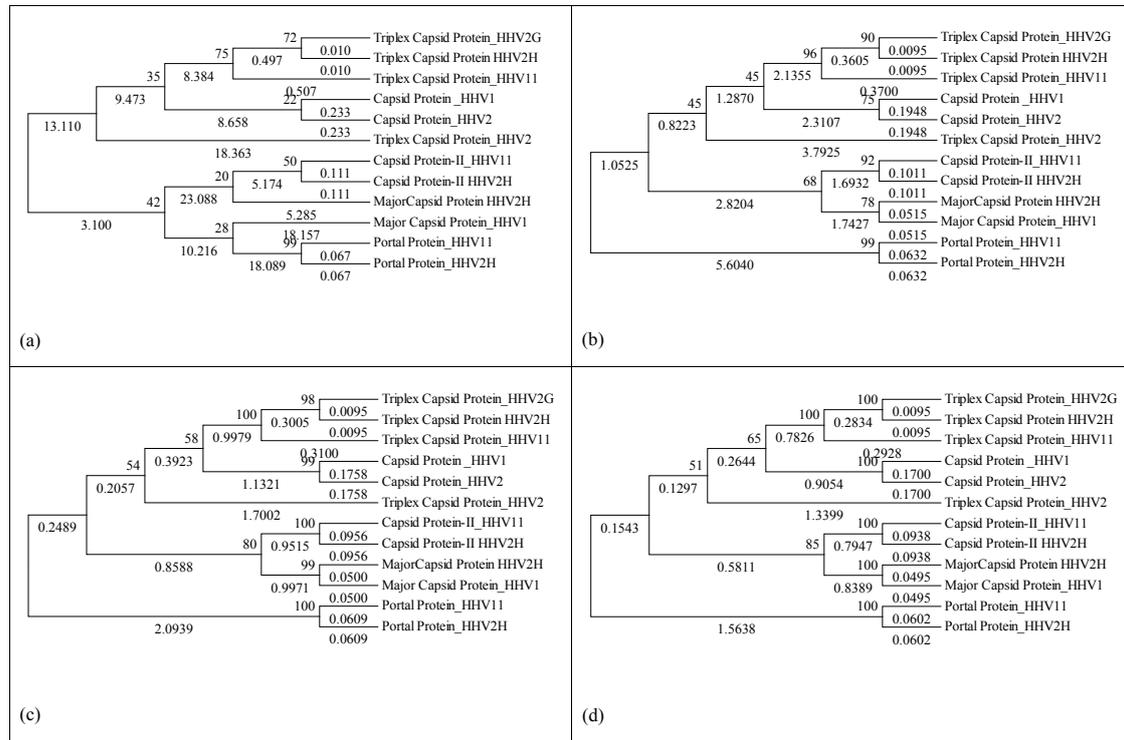


Figure 3: A dataset consisting of 12 viral capsid proteins of HHV strains and tree were generated using neighbor joining technique, the poisson correction and shape parameter of (a) 1.0, (b) 2.5, (c) 4.0 and (d) 5.0.

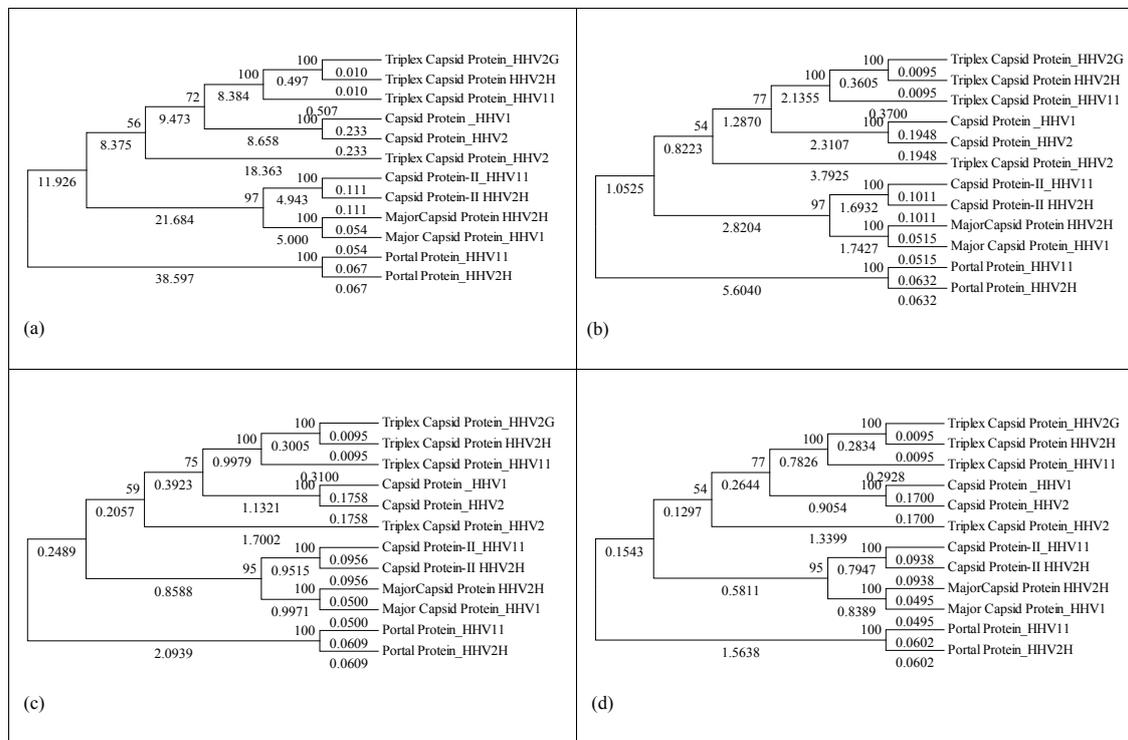


Figure 4: A dataset consisting of 12 viral capsid proteins of HHV strains and tree were generated using UPGMA, the poisson correction and shape parameter of (a) 1.0, (b) 2.5, (c) 4.0 and (d) 5.0.

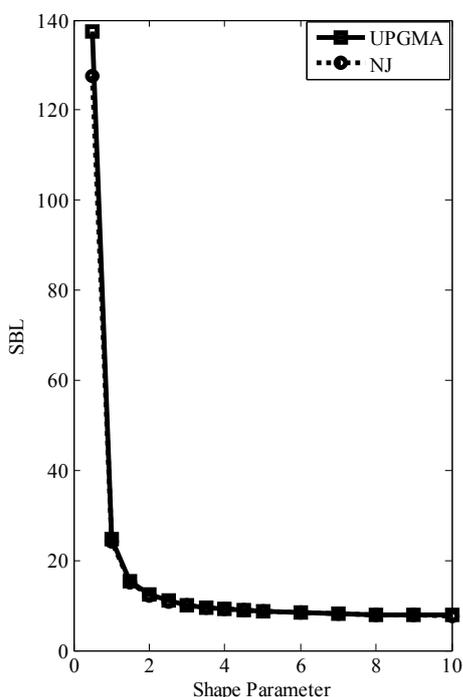


Figure 5: A dataset consisting of 12 viral capsid proteins of HHV strains and sum of branch length were generated using UPGMA and neighbor joining having shape parameter in the range of 0.5 -10.

heuristic approach restricts the branch length 497, and the tree is shown in figure 6a, and Long-branch Attraction (LBA) artifact use to analyze the phylogenetic tree. In max-mini branch-&-bound tree depict that all clade has 100% bootstrap replication consensus for 12 OTU's. Capsid protein-II of HHV11 is rapidly evolving proteins as compare to Portal protein of HHV11 and HHV2H. Major capsid and capsid proteins (HHV1 and HHV2 strains) develop at the same rate of mutation while triplex capsid proteins evolve with slowest rate.

Mini-mini heuristic search algorithm also uses for finding the maximum parsimony tree. Here, we use search factor up to three levels to control the extensiveness of the reconstructed phylogenetic tree. Analysis of figure 6b for 12 OTU of viral protein revealed that triplex capsid protein and portal protein evolve from monophyletic clade while major capsid and capsid protein arise from other clade. Triplex capsid protein HHV2 and capsid protein HHV1 are a combination of top and bottom lineage. LBA analysis for search factor one indicates that portal protein of HHV11 and HHV2H have high mutation during evolution and same confirm by count the number of change. Capsid protein -II of HHV11 and HHV2H have less mutation as compare to portal protein virtue of that number of change at taxon level is less than 7. Major capsid protein is not far behind the portal and capsid protein in term of change. Triplex capsid protein shows minimum value for the number of changes relative to other OTU's. On comparative analysis for search factor two and three as shown in figure 6c and 6d, that triplex capsid protein's clade length is constant and minimum mutagenic evolution. Capsid protein -II HHV-11 evolve at a faster rate relative to other capsid proteins. Portal proteins show the maximum cost on the basis of changes observes in both phylogenetic trees are 5 and 7. Capsid proteins and major capsid proteins show the variation in respect of search level. Mini-mini heuristic search algorithm assessed that the

total cost (number of changes) in MP search level (1 & 3) is equal, and 2 is approximately similar.

Maximum likelihood method is also character-based approach, which uses probabilistic models to choose an optimum tree. In this section, we incorporate gamma parameters that account rate variations across sites and measure impact of gamma parameter on the log likelihood score for viral capsid data. Poisson substitution model use with shape parameter values in the range of 2 to 7. The log likelihood score elevate with a rise in gamma parameter from 2 to 5, but the value of the log likelihood become constant from 5.25 gamma value. The extreme value observes -2467.29 with discrete rate 0.931, 0.9613, 1.0364, 1.1009 and 1.1011. Figure 7a shows that gamma value control the log likelihood value (probability) up to a limit in the original tree highest values indicate (gamma=5) that greatest likelihood producing in observed dataset. Figure 7a is the original tree without bootstrapping consensus. Now we use parametric bootstrapping analysis to remove sampling error of phylogenetic tree and reduce the deviation in sum of the branch length. SBL is also the function of the gamma distribution at the lowest rate, but it become constant at a higher rate. The minimum SBL observed in both the cases (original and bootstrap consensus tree) at lowest gamma parameter 2.0. The values of SBL vary from 8.3799 to 8.3927 in the original tree and 8.2414 to 8.3927 in case of bootstrap consensus tree. The lowest value of SBL variation is 1.65% and approximately negligible at higher gamma value. It reflects that original trees are shadow of bootstrap consensus and a similar result shown in figure 7b.

Molecular clock

We use the Tajima test of the molecular clock which basis on the equal rate (null hypothesis). The equality rate between the triplex capsid protein sequences analyzed and their unique difference in sequence A, B and C are 0, 6 and 84 respectively. χ^2 test statistic value for triplex capsid proteins 6.0 and P-value 0.01431. The P values less than 0.05 so the null hypothesis is rejecting meant equal rates between lineages does not exist. Similar simulation also performed with two more sets [1 A (Capsid Protein_HHV1), B (Capsid Protein_HHV2) and 2 A (Major Capsid Protein_HHV2H), B (Major Capsid Protein_HHV1), C (Capsid Protein_HHV2) and C (Capsid Protein-II_HHV11) and found that χ^2 and P values reduce as shown in table 2. The observation shows that in addition of outgroup lead lower χ^2 and increases p-value.

Discussion

Substitution model is a powerful tool for reconstructing a distance based phylogenetic tree. The present study also supports this argument; it has shown that NJ method is efficient to obtaining the correct tree on the distance data. One of the main reasons for this is that the table 1 produces distance matrix under three different substitution models

| Equality rate between sequences | χ^2 test statistic | P-value |
|--|-------------------------|---------|
| A (Triplex Capsid Protein_HHV2G) B (Triplex Capsid Protein_HHV2H) C (Triplex Capsid Protein_HHV11) | 6.0 | 0.01431 |
| A (Capsid Protein_HHV1) B (Capsid Protein_HHV2) C (Capsid Protein-II_HHV11) | 4.0 | 0.04550 |
| A (MajorCapsid Protein_HHV2H) B (Major Capsid Protein_HHV1) C (Capsid Protein_HHV2) | 3.0 | 0.08326 |

Table 2: Tajima's relative rate test for χ^2 test statistic, P-value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages. The analysis involved 3 amino acid sequences. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA5.

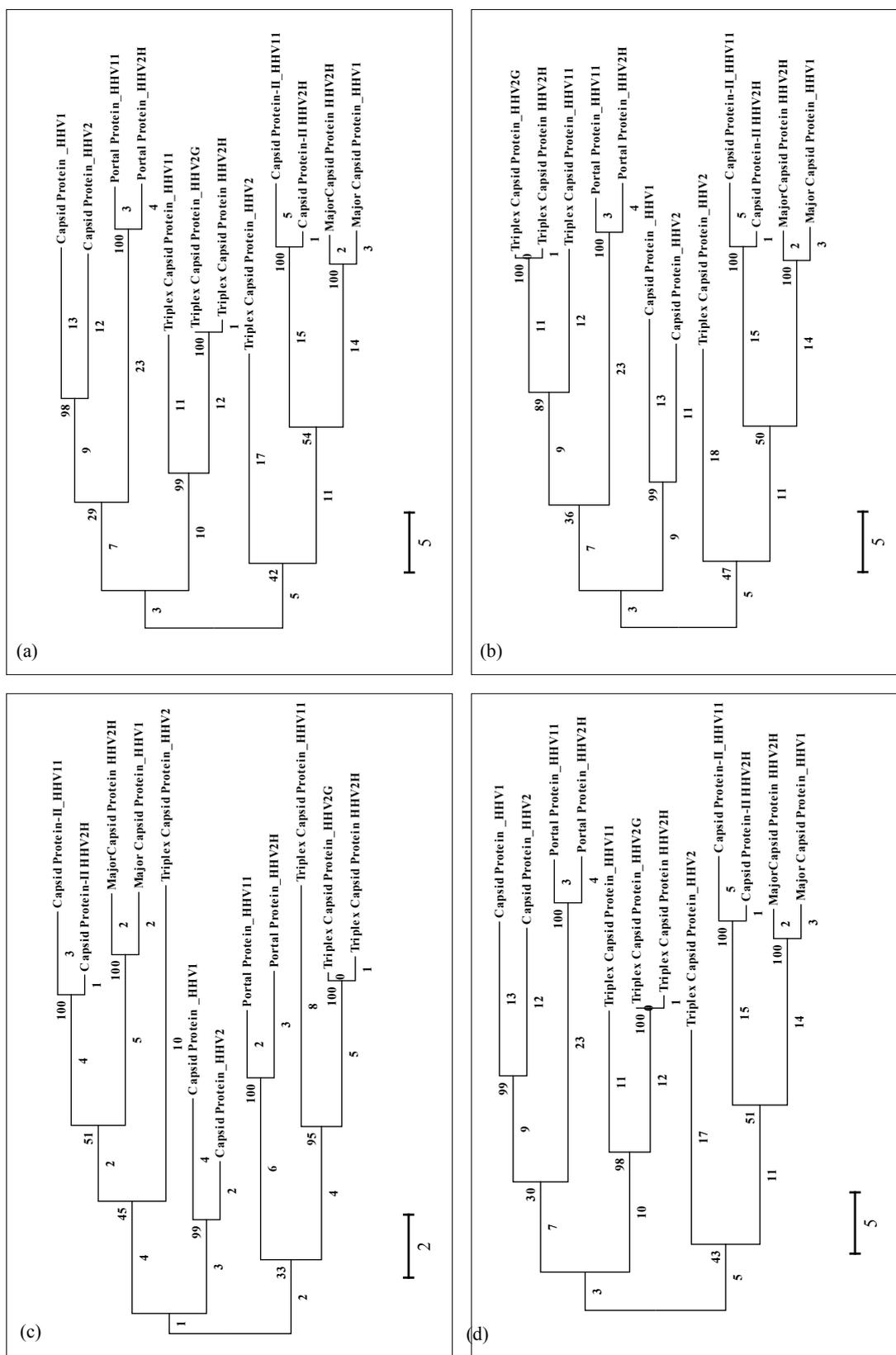


Figure 6: A dataset consisting of 12 viral capsid proteins of HHV strains and tree were generated using maximum parsimony technique and heuristic analysis (a) max-mini (b) mini-mini with search level 1.

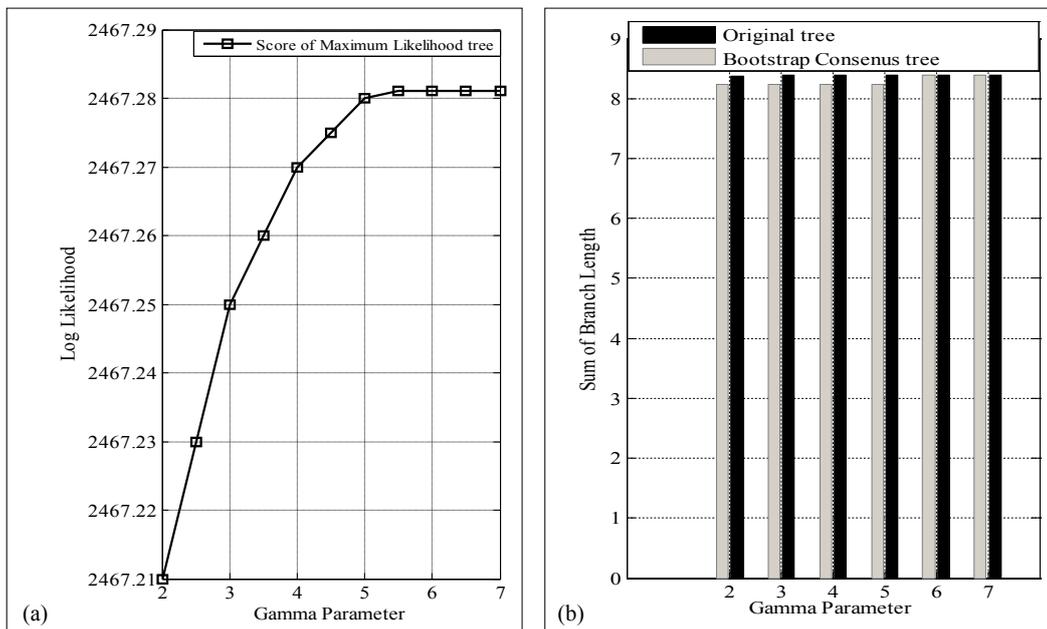


Figure 7: A dataset consisting of 12 viral capsid proteins of HHV strains and tree were generated using maximum likelihood technique using CNI analysis (a) Log likelihood (b) Sum of branch length.

(p-distance, Number of amino acid and poisson correction method) and the replica of distance into tree form shown in figures 1 and 2. Triplex capsid proteins of HHV2G and HHV2H have least matrix different in the entire substitution model and the same observation found in reconstructed phylogenetic tree. Portal protein of HHV2H has unexceptional highest distance value in substitution methods producing clade at lowest bootstrapping value of 28% in figure 3. Therefore, the efficiency of the distance based methods depends on the model and relatedness in OTU's.

Several lines of evidence from figures 3 and 4 suggest that UPGMA and NJ estimates are more robust because gamma distances (poisson correction method) also account for rate variation among sites in calculating evolutionary distances. We have used shape parameter (1.0, 2.5, 4.0 and 5.0) as a criterion because we are interest in estimating a clade relationship and taxon separation in the phylogenetic tree. NJ method results show that each protein placed in fixed position in the four phylogenetic trees, the distances between proteins vary, but bootstrapping values of the trees can be quite large, depending on gamma distribution. Contrary to NJ, gamma distribution does not influence UPGMA bootstrapping value. The clade separation distance in all the four sets of phylogenetic tree is similar in quantitative to NJ trees. This happen because UPGMA method's assumption (all taxa evolve at a constant rate and that they are equally distant from the root, implying that a molecular clock is in effect) is applicable in computer simulation for phylogenetic tree. In a comparative study of the UPGMA and NJ methods, show that significant parameter SBL maintain the original data from a distance matrix. Clustering and optimally techniques (ME) results indicate that higher gamma distribution is most suitable for reconstruction and analysis of tree.

We also use both character based approach (MP and ML) to examine the relationship among viral capsid proteins of HHV strains. Simulation of 12 OTU's of proteins using max-mini branch-and-

bound method which discard the clade of capsid protein HHV1 and HHV2 as shown in figure 6a and results are consistent with distance based approach. We apply the mini-mini heuristic technique (MP) to authenticate the results obtained in the previous simulation. The MP search level 1 and 3 having an equal number of changes (19), while MP search factor 2 has only 14 changes with the assumption that outfit discard. Moreover, the one more consistency observed that triplex capsid proteins have the lowest cost. The accuracy and robustness of mini-mini search method having value 2 is better than other tree represent in figure 6b-6d. Log-likelihood value of ML tree is also changing with gamma parameter. The simulation based analysis suggest that lower value of gamma give the transient, and higher value of gamma provide the stable log likelihood, which are most suitable for phylogenetic tree analysis. The difference between SBL in the original tree and bootstrapping tree are insignificant, indicates that closeness between viral capsid proteins.

Molecular clock's results show that the null hypothesis is fail due to $p < 0.5$ in first two cases, and evolution takes in various proteins at different rates. The results of Tajima molecular clock also favour the phylogenetic tree discussed in figures 3-5.

Conclusion

Our phylogenetic analyses of viral capsid proteins deduced from the HHV-1 and HHV-2 strongly suggests that the major capsid, triplex capsid and portal protein have common lineage. However, sequences of capsid protein HHV1 and HHV2 evolved are comparatively fast with respect to their counterpart HHV-II-1 and HHV-II-2. Poisson substitutions with shape parameter are requiring understanding the phylogenetic tree. Our results suggest that the taxon separation of proteins is less and highly closeness between viral strains. Comprehensive knowledge of evolution in HHV strain may allow the identification of targets and propose drug for therapy in human.

References

1. Alfaro ME, Huelsenbeck JP (2006) Comparative performance of Bayesian and AIC-based measures of phylogenetic model uncertainty. *Syst Biol* 55: 89-96.
2. Geddes BA, Hausner G, Oresnik IJ (2013) Phylogenetic analysis of erythritol catabolic loci within the Rhizobiales and Proteobacteria. *BMC Microbiol* 13: 46.
3. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.
4. Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 17: 368-376.
5. Kenney GE, Rosenzweig AC (2013) Genome mining for methanobactins. *BMC Biol* 11: 17.
6. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59: 307-321.
7. Sourdís J, Nei M (1988) Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. *Mol Biol Evol* 5: 298-311.
8. Krause L, Diaz NN, Goesmann A, Kelley S, Nattkemper TW, et al. (2008) Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Res* 36: 2230-2239.
9. Chen L, Wu D, Ji L, Wu X, Xu D, et al. (2013) Bioinformatics analysis of the epitope regions for norovirus capsid protein. *BMC Bioinformatics* 14: S5.
10. Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst Biol* 53: 793-808.
11. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406-425.
12. Sohpal VK, Dey A, Singh A (2010) MEGA biocentric software for sequence and phylogenetic analysis: a review. *Int J Bioinform Res Appl* 6: 230-240.
13. Kumar S, Filipski AJ (2001) *Molecular Phylogeny Reconstruction*. Encyclopedia of Life Sciences, Nature Publishing Group, London, UK.
14. Sullivan J, Abdo Z, Joyce P, Swofford DL (2005) Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation. *Mol Biol Evol* 22: 1386-1392.
15. 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.
16. Yang Z (1996) Phylogenetic analysis using parsimony and likelihood methods. *J Mol Evol* 42: 294-307.
17. Yang Z (1996) Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol Evol* 11: 367-372.