A protein alignment partitioning method for protein phylogenetic inference

Thu Kim Le
Hanoi University of Sience and Technology
1 Dai Co Viet, Hai Ba Trung, Hanoi, Vietnam
thu.lekim@hust.edu.vn

Vinh Sy Le
VNU University of Engineering and Technology
144 Xuan Thuy, Cau Giay, 100000 Hanoi, Vietnam
vinhls@vnu.edu.vn

***Abstract*— Phylogenetic trees inferred from protein sequences are strongly affected by amino acid evolutionary models. Choosing proper models are needed to account for the heterogeneity in evolutionary patterns across sites, especially when analyzing multiple genes or whole genome datasets. Partitioning is a prominent approach to combine sites undergone similar evolutionary processes into separated groups with proper models. The partitioning scheme can be defined by using structural features of the sequences, however, determining structural features of protein sequences is not always practical. Recently, methods have been proposed to automatically cluster sites into groups based on the rates of sites. The rate of sites is a good indicator; however, it is unable to properly reflex the complex evolutionary processes of sites along the protein sequence. In this paper, we present a new algorithm to automatically determine a partitioning scheme based on the best-fit model of sites, i.e., sites belong to the same model will be classified into the same group. Comparing our proposed method with current methods on a set of empirical protein datasets showed that our method helped to build better trees than other methods tested. Our method will significantly improve protein phylogenetic inference from multiple gene or whole genome datasets.**

Keywords*— *Partitioning, model selection, likelihood

I. INTRODUCTION

Phylogenetic analysis is a powerful tool to study the evolutionary relationships among species [1]. Protein sequences are one of the main data types to construct phylogenetic trees. The accuracy of building phylogenic trees depends on a number of factors, in which choosing the right model of evolution significantly affects the constructed trees [2]. It is well known that the evolutionary processes among sites along the genome are not homologous, e.g., the evolutionary rates vary among sites and depend on the conservation of sites [3].

New sequencing technologies allow us to obtain large datasets including multiple genes or even whole genomes for analyzing the relationships among species. Handling the heterogeneity in the large datasets is a challenging task because none of current evolutionary models is proper for all sites of the dataset containing multiple genes or proteins.

Currently, two main approaches to model the heterogeneity among sites for protein sequences are mixture model approach [4], [5] and partitioning approach [6]–[8]. With mixture models, the likelihood value of each site is calculated under several models [4]. Meanwhile, each site in partitioning approach is assigned to one specific model [9]. In other words, sites assumed to have homologous evolutionary processes will be classified into one group (partition or subset) and follow the same amino acid evolutionary model. The partitioning approach is more realistic than the mixture model approach and therefore being used more frequently in practice.

Different methods can be used to group amino acid sites. The first and intuitive gene-based method is grouping sites by protein [10]. Thus, sites belong to the same protein will be grouped together. The gene-based partition method provides a better alternative compared to "no partitioning" method. Although sites in the same protein might share some common features, the assumption that all sites in one protein evolve by the same model is not biologically realistic. The amino acid sites in one protein might evolve at different rates and follow different amino acid substitution models.

Several studies have been proposed to automatically cluster amino acid sites [7], [8]. The methods use the properties of data, especially the evolution rates of amino acid sites in alignments. They use TIGER (Tree Independent Generation of Evolution Rates [11]) to compute the evolution site rates and cluster sites into groups based on the assumption that sites have similar rates of evolution should be in the same partition.

The k-means algorithm clusters sites based on their site rates. The k-mean algorithm groups all invariant sites into one partition that leads to an incorrect model selection [12]. To partly avoid the problem, the RatePartition algorithm [8] uses a similar approach to calculate evolution rates of sites by TIGER, then applies a simple formula to distribute sites into subsets following the distribution of rates. In the RatePartition method, the first subset will include all the invariant sites and some other sites with the slowest rates in order to partly avoid the pitfalls of k-mean method. The rates of sites in the next subset are greater than that in the previous one. The last subset consists of sites with the highest rates.

In this paper, we develop a new likelihood-based method that automatically partitions protein alignments. Our method is based on rates of sites as well as amino acid substitution models. Experiments on 15 empirical protein datasets showed that in overall our likelihood-based method was better than other methods in building maximum likelihood protein trees based on information-theoretic metrics: the corrected Akaike information criterion (AICc) [13], or the Bayesian information criterion (BIC) [14].

The rest of the paper is organized as follows: Our method will be represented in the section II (Methods). Section III (Experiment and Results) will describe the experiments and discuss results obtained from different methods. The last section will provide discussions, remarks, and recommendations.

II. METHODS

Let $\mathbf{D} = \{D_1, D_2, \dots, D_n\}$ be a set of protein alignments. As usual, we assume that the amino acid sites are evolved independently on the same tree *T*. We use the term 'subset/partition' to represent a set of sites that have the same evolutionary process. The term 'partitioning scheme' implies

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Authorized licensed use limited to: Carleton University. Downloaded on August 06,2020 at 18:34:24 UTC from IEEE Xplore. Restrictions apply.

a collection of subsets so that every site in the alignments **D** belongs to one and only one subset. Technically, let $\mathbf{S} = \{S_1, S_2, \dots, S_k\}$ be a partitioning scheme, where $S_i = (d_i^1, d_i^2, \dots, d_i^{l_i})$ is a subset of l_i amino acid sites that are assumed to evolve under the same evolutionary model M_i. Let $\mathbf{M} = \{M_1, M_2, \dots, M_k\}$ be the set of models corresponding to k subsets.

The likelihood of a tree T is calculated as following:

$$L(T) = P(\mathbf{S}|T, \mathbf{M}) = \prod_{i=1}^{k} P(S_i|T, M_i) = \prod_{i=1}^{k} \prod_{j=1}^{l_i} P(d_i^j|T, M_i)$$

where $P(d_i^j|T, M_i)$ is the probability of amino acid site d_i^j given the tree T and model M_i. Our objective is to find a partition scheme **S** and corresponding model set **M** that help building the maximum likelihood tree T.

An evolutionary model M_i describing the amino acid evolutionary process of a partition includes two parts: the site rate model R_i and the amino acid substitution model Q_i. The amino acid substitution models are normally selected from existing empirical models that were already estimated from large datasets such as JTT [15], WAG [16] or LG [2]. If the dataset under the study is a domain-specific dataset such as viruses; models like FLU [17] or HIVs [18] can be employed.

The site rate model R_i is typically a combination of discrete Gamma distribution rate model [19] and invariant rate model. It consists of two parameters (i.e., one from the Gamma distribution rate model and another from the invariant rate model) will be directly estimated from the dataset.

The model set **M** for the non-partition scheme (original data set **D**) consists of one partition with 2 free parameters. The model set **M** for a partition scheme **S** of k partitions will consists of $2 \times k$ free parameters. The AICc score [13] and BIC score [14] can be used to compare the fitness of different partition schemes based on likelihood values of constructed trees and the number of free parameters. Note that a partition scheme with more free parameters will help increasing the likelihood of the tree, however, it will have to pay a higher penalty score for the additional free parameters.

The underlying idea of partition method is grouping amino acid sites that share the same evolutionary patterns. We propose a likelihood-based (LLB) algorithm to cluster sites based on their model preferences including not only site rate models, but also amino acid substitution models. The LLB algorithm includes three main steps: initial step, model selection step, and partitioning step. The LLB algorithm is summarized in Fig. 1.

At the initial step, the LLB algorithm determines a list of possible amino acid substitution models for the dataset under the study. The chosen models should be generally suitable for analysing the dataset. For general datasets, frequently-used general amino acid substitution models can be considered such as LG [20], JTT [15], WAG [16], BLOSUM62 [21]. This step can be reasonably accomplished by selecting potentially suitable models from a list of current existing models. We denote **Q** the set of possible amino acid substitution models. The site rate models include the none rate model (NR) and combinations of discrete Gamma distribution model G and invariant model I. We denote **R** the set of four possible site rate models, i.e., NR, G, I, G+I. All free parameters of site rate models will be directly estimated from the dataset under the study. Let **cM** be the set of possible models, each model M of **cM** consists of an amino acid substitution model Q from **Q** and a site rate model R from the **R**.

The model selection step of the LLB algorithm will assign each site to a proper model of **cM**, and consequently cluster sites of the same model into one subset. For each alignment, the model selection step starts by quickly building **|cM|** trees based on **|cM|** different models. The trees will be used to evaluate the model preference of each site of the alignment. To build trees, we can use distance-based tree reconstruction methods such as Neighbor-Joining [22], its improved version BioNJ algorithm [23], or very fast method STC [24]. For each site, the step will determine and select the most preferred model for the site based on its log-likelihood values calculated with different models from the model set **cM**.

Finally, the LLB algorithm clusters sites in **D** based on their preferred models to create a partition scheme **S**. Specifically, sites which have the same preferred model will be clustered into the same subset. Some subsets might contain only few sites that add more unnecessary free parameters in inferring phylogenetic trees and might distort tree structures. To overcome this problem, the LLB algorithm will merge small subsets into their highest correlated larger subsets. In this study, a subset is considered as a small subset if it contains

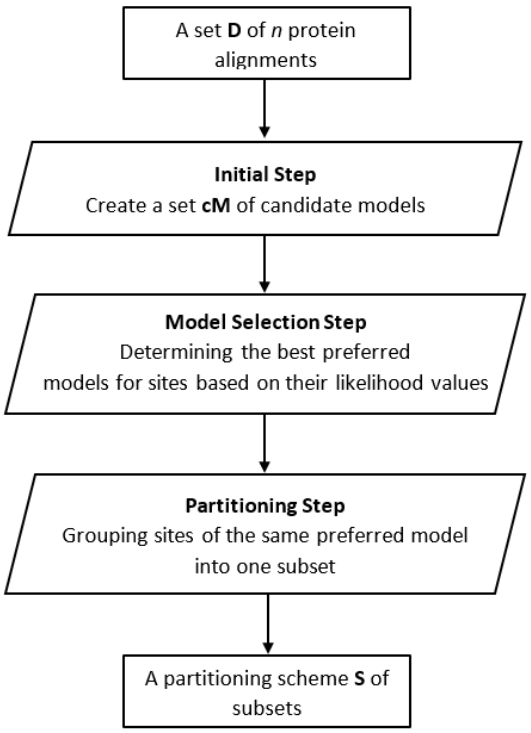

Fig. 1. THE LIKELIHOOD-BASED PARTITIONING METHOD.

less than 10% of the total number sites.

III. EXPERIMENTS AND RESULTS

We examined our proposed LLB algorithm with other partitioning methods including (1) no partitioning (NP), i.e., the partitioning scheme has only one subset that includes all the sites; (2) partitioning by gene boundaries (GP), i.e., each

Authorized licensed use limited to: Carleton University. Downloaded on August 06,2020 at 18:34:24 UTC from IEEE Xplore. Restrictions apply.

alignment is considered as a subset; (3) partitioning by RatePartition method (RP) [8]. We compared their performance on five protein benchmark datasets downloaded from *https://github.com/roblanf/BenchmarkAlignments/*. The five datasets contain protein alignments obtained from five evolutionary studies of mammals, animals, birds, jawed vertebrates, and metazoans. The number of taxa in the datasets ranges from 36 to 90 and each dataset contains thousands of loci (alignments). As it is computationally expensive to examine all partitioning methods on datasets with thousands of loci, for each dataset we randomly selected 10, 20, and 40 loci to create three different datasets. Thus, in this study we examined partitioning methods on 15 different datasets (see TABLE I.).

The initial step of LLB method will use four general amino acid substitution models LG [2], JTT [15], WAG [16], and BLOSUM62 [21] as possible amino acid substitution models for the general datasets.

The maximum likelihood software IQ-TREE [25] was used to construct distance-based trees by the BioNJ algorithm, compute site likelihoods, and build maximum likelihood trees for different partitioning schemes obtaining from partitioning methods. We used the AICc [13] and BIC [14] scores to compare the performance of different partitioning methods, i.e., the smaller AICc score (BIC score) indicates the better partitioning method.

TABLE II. presents the AICc and BIC scores of different methods. The results based on the AICc scores are similar to that based on the BIC scores. The LLB method resulted in best solutions for 10 out of 15 tests and the second-best solutions for the 5 other tests. The RP method was the second-best method. It produced the best solutions for 5 out 15 tests and the second-best solutions for the other 10 tests. The NP (no partitioning) and GP (partitioning by genes) methods did not

TABLE I. FIFTEEN DATASETS USED TO COMPARE PARTITIONING METHODS

Datasets	Clade	#Taxa	#Loci	#Sites	#Loci	#Sites	#Loci	#Sites
Borowiec [26]	Mammals	90	10	5148	20	12225	40	24423
Chen [27]	Animals	78	10	2376	20	4084	40	7893
Ran [28]	Birds	52	10	3062	20	6897	40	14749
Wu [29]	Jawed vertebrates	58	10	3967	20	6403	40	15278
Cannon [30]	Metazoans	36	10	2836	20	7618	40	15113

TABLE II. AICC AND BIC SCORES OF DIFFERENT PARTITIONING METHODS FOR 15 DATASETS. THE NUMBER IN THE BRACKETS OF A DATASET INDICATES THE NUMBER OF LOCI. THE BEST SOLUTIONS ARE HIGHLIGHTED IN BOLD. LLB (LIKELIHOOD-BASED), NP (NO PARTITIONING), GP (PARTITIONING BY GENE) AND RP (RATEPARTITION)

Datasets	AICc				BIC			
	LLB	***NP***	***GP***	***RP***	***LLB***	***NP***	***GP***	***RP***
Borowiec (10)	**211699**	215506	215526	211701	**212217**	216070	216084	212410
Cannon (10)	244445	248142	247772	**243961**	245465	249059	249130	**245067**
Chen (10)	**140840**	144673	143857	141138	**141704**	145426	144830	141940
Ran (10)	111956	115336	115092	**110952**	112587	115808	115694	**111460**
Wu (10)	**187943**	194667	194025	190758	**189605**	195864	195497	192026
Borowiec (20)	**561396**	571739	570930	562392	**562299**	572410	572442	563302
Cannon (20)	**482289**	488894	489029	482465	**483635**	490038	490947	483968
Chen (20)	**262497**	269460	268279	263105	**263444**	270273	269805	263932
Ran (20)	284004	602319	291872	**281236**	284733	603699	292715	**282086**
Wu (20)	**590263**	602320	599822	590926	**591959**	603700	601935	592526
Borowiec (40)	**1111525**	1133462	1132482	1113434	**1112824**	1134208	1134508	1114362
Cannon (40)	**915019**	927756	928265	915079	**916534**	929044	931348	916827
Chen (40)	**720272**	734939	733643	720539	**721361**	736005	735982	721613
Ran (40)	600776	619515	618767	**599479**	601762	620274	620289	**600450**
Wu (40)	1308500	1332075	1328103	**1304664**	1310375	1333757	1331805	**1306733**

Authorized licensed use limited to: Carleton University. Downloaded on August 06,2020 at 18:34:24 UTC from IEEE Xplore. Restrictions apply.

result in any best solution. The results confirm that partitioning methods help constructing better phylogenetic trees in comparison to no partitioning or partitioning by genes methods. The results also show that partitioning based on the combination of both site rate models and amino acid substitution models is much better than that based on only the site rates.

We summarized the number of subsets of partitioning schemes created from two partitioning methods LLB and RP in TABLE III. The LLB method produced partitioning schemes with fewer subsets than that produced by the RP method. It could be explained by the merging strategy of LLB method to merge small subsets into large subsets to avoid adding unnecessary free parameters when inferring the phylogenetic trees.

TABLE III. THE NUMBER OF SUBSETS IN PARTITIONING SCHEMES USING LLB AND RP METHODS

Dataset name	LLB	RP
Borowiec (10)	6	13
Cannon (10)	7	14
Chen (10)	5	7
Ran (10)	5	10
Wu (10)	5	6
Borowiec (20)	6	13
Cannon (20)	6	14
Chen (20)	5	5
Ran (20)	5	9
Wu (20)	5	7
Borowiec (40)	6	13
Cannon (40)	6	14
Chen (40)	5	8
Ran (40)	5	10
Wu (40)	6	8

We also measured the distances between trees constructed from different partitioning schemes to examine if partitioning schemes affect constructed trees. The average of Robinson-Foulds distance [31] between phylogenies that constructed by four methods are present in TABLE IV. The results show that the trees constructed from four partitioning schemes are different. In other words, partitioning schemes considerably affect the tree structures.

Invariant sites play an important role in partitioning methods. The k-mean partitioning method clusters all invariant sites into one subset that might significantly increase the likelihood value of the tree, however, seriously distort the tree structure [12]. As a result, the k-mean partitioning method has been suspended by the authors and no long for use. The RP partitioning method tries to avoid the pitfall by adding some slowest rate sites into the subset of invariant sites. In our testing datasets, the Ran's datasets with 10, 20, and 40 loci consist of 30%, 27%, and 22% invariant sites, respectively. Interestingly, our LLB method clustered the invariant sites into different subsets in the partitioning scheme (see TABLE V.). This will help avoiding the pitfall of grouping all invariant sites into one subset by the both k-mean and RP methods.

TABLE IV. NORMALIZED ROBINSON & FOULDS (RF) DISTANCES BETWEEN PHYLOGENIES BUILT WITH 4 PARTITIONING METHODS

	GP	NP	LLB	RP
GP		0.055974	0.048647	0.052734
NP	0.055974		0.055535	0.056771
LLB	0.048647	0.055535		0.067211
RP	0.052734	0.056771	0.067211	

TABLE V. THE NUMBER OF INVARIANT SITES IN SUBSETS OF THE PARTITIONING SCHEME OBTAINED FROM THE LLB ALGORITHM

	Subsets				
Dataset	1	2	3	4	5
Ran (10)	102	34	340	239	208
Ran (20)	473	652	79	566	88
Ran (40)	1095	879	199	890	167

IV. DISCUSSIONS AND CONCLUSIONS

The number of large datasets including multiple genes or even whole genomes have been generated. It is necessary to develop adequate methods to handle the heterogeneity in the large datasets. Partitioning data is being used as the most effective way to deal with the problem. In this paper, we present the likelihood-based algorithm LLB to automatically partition a given protein dataset into a partitioning scheme such that all sites in one subset have undergone the same evolutionary model.

The results on empirical protein datasets confirmed that proper partitioning schemes helped building better trees than no partitioning or simply partitioning by genes. The LLB method was generally better than other partitioning methods tested in terms of both AICc and BIC criteria. The RP partitioning method produced solutions with higher likelihood values than LLB method on Ran's datasets that include too many invariant sites. The higher likelihood values of RP method over LLB method on the Ran's datasets might come from the big subset of all invariant sites that might lead to incorrect inference of phylogenetic trees. We note that the LLB method clustered the invariant sites into different subsets in the partitioning scheme and avoided the pitfall.

In this paper, we tested different partitioning methods on empirical general protein datasets so the list of general amino substitution models such as JTT, WAG, LG were employed. The list of possible models should be modified when analyzing other datasets such that they can properly reflex the evolutionary processes of proteins in the datasets. For

Authorized licensed use limited to: Carleton University. Downloaded on August 06,2020 at 18:34:24 UTC from IEEE Xplore. Restrictions apply.

example, if the alignment contains proteins from viruses, we can consider including virus models such as HIV[18], FLU [17], DEN [32] in the list. A proper list of possible models will improve the accuracy of partitioning schemes.

In a nutshell, the LLB method provides a practical mean to deal with the heterogeneity in the large datasets. It enhances the quality of phylogenomic inference, especially when we do not know much about characteristics of the datasets to create proper partitioning schemes for building phylogenomic trees

ACKNOWLEDGMENT

This work was financially supported by Vietnam National Foundation for Science and Technology Development.

REFERENCES

[1] J. Felsenstein, *Inferring phytogenies*. Sunderland, MA, USA: Sinauer Associates, 2003.

[2] S. Q. Le and O. Gascuel, "An improved general amino acid replacement matrix," *Mol. Biol. Evol.*, vol. 25, no. 7, pp. 1307–1320, 2008.

[3] M. E. C. Lemmon AR, "The importance of proper model assumption in Bayesian phylogenetics," *Syst. Biol.*, vol. 53, pp. 265–27, 2004.

[4] G. O. Le SQ Dang CC, "Modeling protein evolution with several amino acid replacement matrices depending on site rates," *Mol Biol Evol*, vol. 29, pp. 2921–36, 2012.

[5] P. H. Lartillot N, "A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process," *Mol Biol Evol*, vol. 21, pp. 1095–1109, 2004.

[6] H. J. P. N.-A. J. Nylander JAA Ronquist F, "Bayesian phylogenetic analysis of combined data," *Syst Biol*, vol. 53, pp. 47–67, 2004.

[7] M. C. L. R. Frandsen PB Calcott B, "Automatic selection of partitioning schemes for phylogenetic analyses using iterative k-means clustering of site rates," *BMC Evol. Biol.*, vol. 15, 2015.

[8] C. N. P. C. W. N. Rota J Malm T, "A simple method for data partitioning based on relative evolutionary rates," *PeerJ*, vol. 6, 2018.

[9] H. S. Y. W. G. S. Lanfear R Calcott B, "PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses," *Mol. Biol. Evol.*, vol. 29, pp. 1695–1701, 2012.

[10] L. R. Kainer D, "The effects of partitioning on phylogenetic inference," *Mol. Biol. Evol.*, vol. 32, pp. 1611–1627, 2015.

[11] M. J. O. Cummins CA, "A method for inferring the rate of evolution of homologous characters that can potentially improve phylogenetic inference, resolve deep divergence and correct systematic biases," *Syst. Biol.*, vol. 60, pp. 833–844, 2011.

[12] M. K. B. S. A. E. Z. Baca SM Toussaint EFA, "Molecular phylogeny of the aquatic beetle family Noteridae (Coleoptera: Adephaga) with an emphasis on data partitioning strategies," *Mol. Phylogenet. Evol.*, vol. 107, pp. 282–292, 2017.

[13] T. C.-L. Hurvich CM, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, pp. 297–307, 1989.

[14] S. G, "Estimating the dimension of a model," *Ann Stat*, vol. 6, pp. 461–464, 1978.

[15] D. T. Jones, W. R. Taylor, and J. M. Thornton, "The rapid generation of mutation data matrices from protein sequences," *Bioinformatics*, vol. 8, pp. 275–282, 1992.

[16] S. Whelan and N. Goldman, "A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach.," *Mol. Biol. Evol.*, vol. 18, no. 5, pp. 691–699, 2001.

[17] G. O. V. Le Dang Cuong Le Quang, "FLU, an amino acid substitution model for influenza proteins," *BMC Evol. Biol.*, vol. 10, p. 99, 2010.

[18] J. M. A. G. P. B. M. J. I. K. S. L. Nickle DC Heath L, "HIV-Specific Probabilistic Models of Protein Evolution," *PLoS One*, vol. e503, 2007.

[19] Z. Yang, "Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods," *J. Mol. Evol.*, vol. 39, no. 3, pp. 306–314, 1994.

[20] L. N. Quang LS Gascuel O, "Empirical profile mixture models for phylogenetic reconstruction," *Bioinformatics*, vol. 24, pp. 2317–23, 2008.

[21] H. J. G. Henikoff S, "Amino acid substitution matrices from protein blocks," *Proc Natl Acad Sci USA*, vol. 89, pp. 10915–10919, 1992.

[22] N. Saitou and M. Nei, "The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees," *Mol Biol Evol*, vol. 24, 1987.

[23] G. Olivier, "BIONJ: An Improved Version of the NJ Algorithm Based on a Simple Model of Sequence Data. Molecular biology and evolution," *Mol. Biol. Evol.*, vol. 14, pp. 685–695, 1997.

[24] V. Le Sy and A. von Haeseler, "Shortest triplet clustering: Reconstructing large phylogenies using representative sets," *BMC Bioinformatics*, vol. 6, p. 92, 2005.

[25] von H. A. M. B. Nguyen LT Schmidt H, "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies," *Mol Biol Evol*, vol. 32, 2014.

[26] C. J. C. P. D. C. Borowiec M. L. Lee E. K., "Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa," *BMC Genomics*, vol. 16, 2015.

[27] Z. P. Chen MY Liang D, "Selecting question-specific genes to reduce incongruence in phylogenomics: a case study of jawed vertebrate backbone phylogeny," *Syst. Biol.*, vol. 64, pp. 1104–1120, 2015.

[28] W. M.-M. W. X.-Q. Ran J-H Shen T-T, "Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms," *Proc. R. Soc. B*, vol. 285(1881), 2018.

[29] E. S. L. L. Wu S., "Genome-scale DNA sequence data and the evolutionary history of placental mammals," *Data Br.*, vol. 18, pp. 1972–1975, 2018.

[30] S. J. R. F. J. U. H. A. Cannon JT Vellutini BC, "Xenacoelomorpha is the sister group to Nephrozoa," *Nature*, vol. 530, pp. 89–93, 2016.

[31] F. L. R. Robinson DF, "Comparison of phylogenetic trees," *Math. Biosci.*, vol. 53, pp. 131–147, 1981.

[32] T. Kim, C. Dang, and V. Le, "Building a Specific Amino Acid Substitution Model for Dengue Viruses," 2018, pp. 242–246.

Authorized licensed use limited to: Carleton University. Downloaded on August 06,2020 at 18:34:24 UTC from IEEE Xplore. Restrictions apply.