Gene xxx (2008) xxx–xxx

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Effect of site-specific heterogeneous evolution on phylogenetic reconstruction: A simple evaluation

Qiqun Cheng [a,b], Zhixi Su [a,c], Yang Zhong [c,*], Xun Gu [a,c,d,*]

[a] Institute of Biomedical Sciences, Fudan University, Shanghai 200433, China
[b] Key Laboratory of Marine and Estuarine Fisheries, Ministry of Agriculture, East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Shanghai 200090, China
[c] School of Life Sciences and Center for Evolutionary Biology, Fudan University, Shanghai 200433, China
[d] Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA 50011, USA

ARTICLE INFO

Article history:
Received 13 April 2008
Received in revised form 27 June 2008
Accepted 5 August 2008
Available online xxxx

Received by A. Bernardi

Keywords:
DIVERGE
Heterogeneous evolution
Phylogenetic reconstruction

ABSTRACT

Recent studies have shown that heterogeneous evolution may mislead phylogenetic analysis, which has been neglected for a long time. We evaluate the effect of heterogeneous evolution on phylogenetic analysis, using 18 fish mitogenomic coding sequences as an example. Using the software DIVERGE, we identify 198 amino acid sites that have experienced heterogeneous evolution. After removing these sites, the rest of sites are shown to be virtually homogeneous in the evolutionary rate. There are some differences between phylogenetic trees built with heterogeneous sites ("before tree") and without heterogeneous sites ("after tree"). Our study demonstrates that for phylogenetic reconstruction, an effective approach is to identify and remove sites with heterogeneous evolution, and suggests that researchers can use the software DIVERGE to remove the influence of heterogeneous evolution before reconstructing phylogenetic trees.

1. Introduction

Since Darwin (1859), reconstructing phylogenetic trees of species has been one of the major foci in evolutionary biology. Large-scale utilization of sequence data, promoted by the PCR technique and more recent genome sequencing, has overcome some inherent flaws of fossil and morphological data, making the goal of tree of life achievable. Numerous tree-making methods are developed to infer the phylogenetic tree from sequence data (Nei and Kumar, 2000), which can be roughly classified into: Distance Method such as Neighbor-Joining (NJ), Maximum Parsimony (MP), Maximum Likelihood (ML), and Bayesian Inferences (BI). Substantial studies have been reported about advantages and disadvantages of these methods; one may refer to references (Li, 1997; Nei and Kumar, 2000; Felsenstein, 2004) for overviews and general discussions.

Kolaczkowski and Thorton (2004) investigated the performance of some tree-making methods when the process of sequence evolution is heterogeneous, i.e., the evolutionary rate of a given site is not always constant throughout time. Heterotachy is an important process of protein evolution (Lopez et al., 2002). By computer simulations, Kolaczkowski and Thorton (2004) concluded that in this case, maximum parsimony performs much better than maximum likelihood and Bayesian inference. In another simulation study, however, Gaucher and Miyamoto (2005) reached an opposite conclusion that likelihood phylogenetics is favorable even when the evolution is heterogeneous. Though it seems that different parameter sets in computer simulations may lead to different results, this debate highlighted the problem of heterogeneous evolution in the phylogenetic analysis, which has been neglected in practical systematic analysis. Although several sophisticated statistical models have been developed, e.g., see Tuffley and Steel (1998), Galtier (2001), and Huelsenbeck (2002), implementation to phylogenetic reconstruction has been shown a difficult task.

In this short communication, we take an alternative approach to addressing this issue. Instead of developing more sophisticated algorithms to account for the effect of heterogeneous evolution, we attempt to identify nucleotide or amino acid sites that have experienced heterogeneous evolution, and then remove these sites before conducting a phylogenetic analysis. We invoke the software DIVERGE (2002) to perform such analysis, based on the statistical method (Gu, 1999) to detect amino acid residues with shifted

Abbreviations: NJ, Neighbour Joining; MP, Maximum Parsimony; ML, Maximum Likelihood; BI, Bayesian Inference; TBR, Tree Bisection-Reconnection.

* Corresponding authors. X. Gu is to be contacted at Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA 50011, USA. Tel.: +1 515 294 8075; fax: +1 515 294 8457.
E-mail address: xgu@iastate.edu (X. Gu).

0378-1119/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2008.08.003

Table 1
List of 20 fish species and their GenBank accession numbers

Order	Family	Species	Common name	Accession no.
Clupeiformes	Clupeidae	*Sardinops melanostictus*	Japanese pilchard	AB032554
	Engraulidae	*Engraulis japonicus*	Japanese anchovy	AB040676
Cypriniformes	Cyprinidae	*Cyprinus carpio*	Common carp	X61010
		Carassius auratus langsdorfii	Gin-buna	AB006953
		Carassius auratus cuvieri	Japanese crucian carp	AB045144
		Danio rerio	zebrafish	NC-002333
		Sarcocheilichthys variegatus microoculus	Biwa higai	AB054124
Salmoniformes	Salmonidae	*Oncorhynchus mykiss*	Rainbow trout	L29771
		Coregonus lavaretus	Common whitefish	AB034824
		Salmo salar	Atlantic salmon	U12143
		Salvelinus fontinalis	Brook trout	NC-000860
		Salvelinus alpinus	Arctic charr	NC-000861
	Plecoglossidae	*Plecoglossus altivelis*	Sweetfish	NC-002734
	Galaxiidae	*Galaxias maculates*	Inanga	AP004104
	Retropinnidae	*Retropinna retropinna*	Common smelt	AP004108
	Salangidae	*Salangichthys microdon*	Japanese icefish	AP004109
Stephanoberyciformes	Rondeletiidae	*Rondeletia loricata*	Redmouth whalefish	AP002937
	Cetomimidae	*Cetostoma regani*	Pink flabby whalefish	AP004423
		Danacetichthys galathenus	Flabby whalefish	AP002936
	Mirapinnidae	*Eutaeniophorus sp. 033-Miya*	Festive ribbonfish	AP004424

evolutionary rates. As a first step, we illustrated this idea by the maximum parsimony analysis of fishes (Cypriniformes, Salmoniformes, and Stephanoberyciformes), using mitogenomic (complete mitochondrial) protein sequences (Inoue et al., 2001; Miya et al., 2001, 2003).

Table 2
The coefficient of functional divergence (θ) of pairwise comparisons of different fish orders

Pairwise comparison		No cutoff		Cutoff value = 0.5	
A	B	Sites[a]	$\theta_{AB} \pm SE$[b]	Sites[a]	$\theta_{AB} \pm SE$[b]
Cluster 1	Cluster 2	3776	0.266 ± 0.025	3578	0.001 ± 0.022
Cluster 1	Cluster 3	3776	0.276 ± 0.032	3578	−0.234 ± −0.042
Cluster 2	Cluster 3	3776	0.170 ± 0.028	3578	0.001 ± 0.022

Cluster 1: ((*C. carpio*,(*C. auratus*, *C. cuvieri*)), *S. variegatus*, *D. rerio*).
Cluster 2: ((*S. salar*, (*O. mykiss*,(*S. fontinalis*, S.alpinus))),(*G. maculatus*, (*R. retropinna*, (*P. altivelis*, *S. microdon*))), *C. lavaretus*).
Cluster 3: ((*C. regani*, *Eutaeniophorus* sp), *R. loricata*, *D. galathenus*).
[a] The total number of amino acid position (without gaps) in the multiple sequence alignment.
[b] The coefficient of evolutionary functional divergence between cluster A and B, and its standard error (SE).

2. Methods

2.1. Fish mitogenomic sequences data

Twenty mitogenomic sequences of fishes were used in this study, which were downloaded from the NCBI website (http://www.ncbi.nlm.nih.gov/nucleotide), including five sequences for order Cypriniformes, nine for order Salmoniformes, and four for order Stephanoberyciformes, while two mitogenomic sequences are used as the outgroup. The scientific names, taxonomy, and GenBank accession numbers are listed in Table 1.

2.2. Multiple alignment and phylogenetic analysis

The amino acid sequences of 13 protein-coding genes (i.e. ND1, ND2, COI, COII, ATPase 8, ATPase 6, COIII, ND3, ND4L, ND4, ND5, ND6, and Cytb) in mitochondrial DNA were combined and aligned by using Clustal X (Thompson et al., 1997) software with the default settings, followed by minor manual adjustment. The primary sequences of the 13 protein-coding genes were used to align, and the columns of an alignment that contained any gaps were removed from further study. Phylogenetic trees were inferred by maximum parsimony (MP) method using PAUP 4.0b10

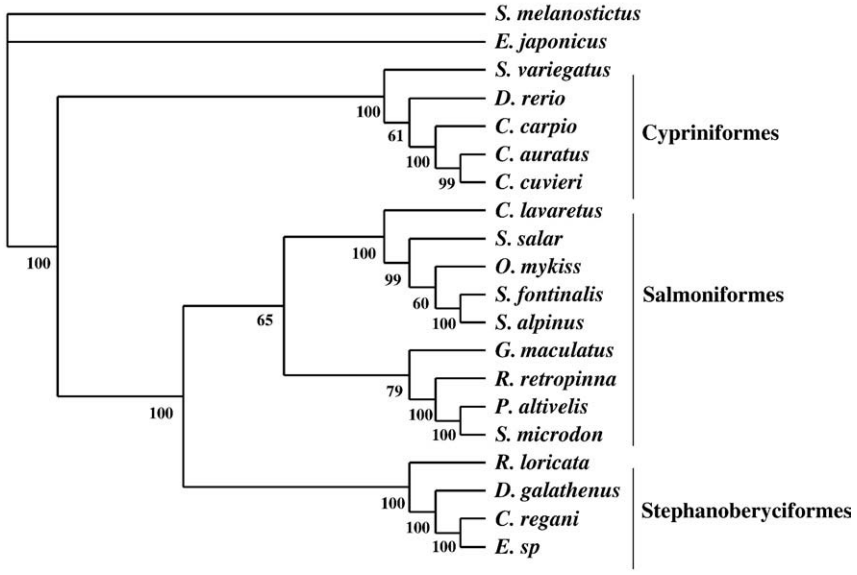

Fig. 1. Maximum parsimony phylogenetic tree of the fishes inferred from original mitogenomic protein-coding sequences with bootstrapping 1000 times (the three groups of Cypriniformes, Salmoniformes, and Stephanoberyciformes are respectively designated as cluster 1, cluster 2, and cluster 3).

Please cite this article as: Cheng, Q., et al., Effect of site-specific heterogeneous evolution on phylogenetic reconstruction: A simple evaluation, Gene (2008), doi:10.1016/j.gene.2008.08.003

(Swofford, 2001). In the MP analysis, all characters have equal weight. We conducted heuristic searches with initial trees obtained by simple stepwise addition, followed by branch swapping using the TBR (tree bisection–reconnection) routine. The robustness of the nodes of phylogenetic trees was tested by bootstrapping methods (Felsenstein, 1985).

2.3. Evolutionary analysis on site-specific profile

Gu (1999) refers Type I functional divergence to the evolutionary process that results in altered functional constraints at some sites, leading to the site-specific evolutionary rate-shift between two monophylogenetic groups by either gene duplication or speciation. A statistical method was developed and thereafter implemented to the software DIVERGE (Gu and Vander, 2002) to estimate the coefficient of functional divergence (θ), a statistical measure for Type I functional divergence of two groups (Gu, 1999; Wang and Gu, 2001; Gu and Vander, 2002), ranging from 0 to 1. Rejection of the null hypothesis θ=0 provides statistical evidence for the shift in evolutionary rate. The software DIVERGE has been used successfully in finding specific sites and functional divergence of different protein

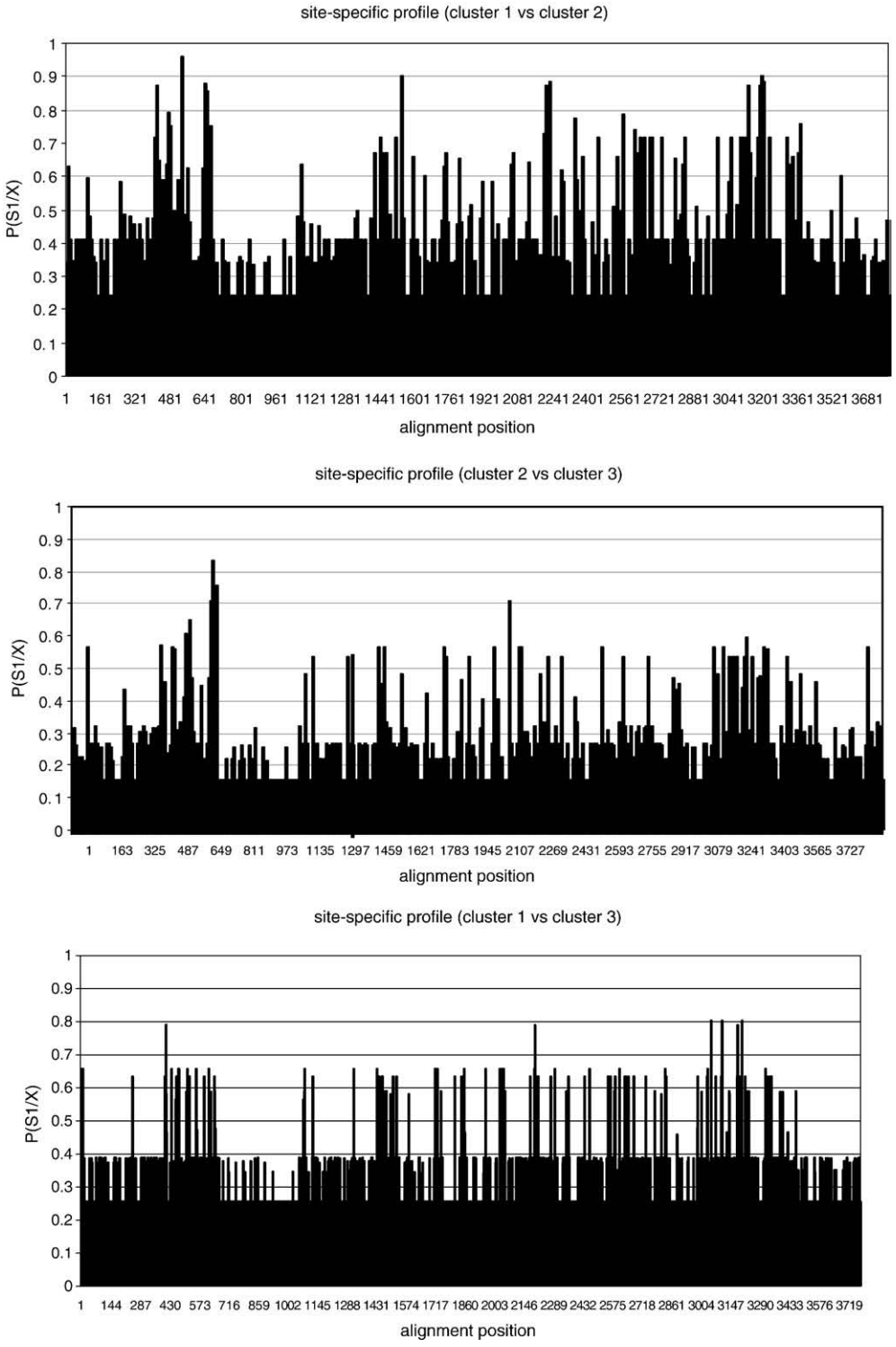

Fig. 2. Site-specific profile of posterior probability between cluster 1, cluster 2, and cluster 3 (no cutoff).

Please cite this article as: Cheng, Q., et al., Effect of site-specific heterogeneous evolution on phylogenetic reconstruction: A simple evaluation, Gene (2008), doi:10.1016/j.gene.2008.08.003

families (Gu, 1999; Wang and Gu, 2001; Oparina et al., 2005). It seems that the problem of heterogeneous evolution can be solved by the functional divergence analysis (Gaucher et al., 2002), as those sites with rate-shifts (so-called heterogeneous evolution) can be predicted by the posterior analysis.

Tentatively, we propose the following steps to solve the problem of heterogeneous evolution, which is simple and straightforward: (*i*) infer the phylogenetic tree using the whole sequence alignment; (*ii*) determine the monophyletic groups along the phylogeny, based on the biological interest and statistical reliability (the bootstrapping

Table 3
Specific sites with posterior probability higher than 0.5

Site	1 vs. 2	1 vs 3	2 vs 3	Site	1 vs 2	1 vs 3	2 vs 3	Site	1 vs 2	1 vs 3	2 vs 3
2	0.338	**0.634**	0.105	1531	0.069	**0.634**	0.305	2679	0.062	**0.590**	0.320
7	0.485	**0.657**	0.037	1538	**0.902**	0.378	0.127	2680	**0.719**	**0.634**	0.155
8	**0.629**	**0.634**	0.022	1539	**0.630**	0.256	0.446	2683	**0.668**	0.112	0.271
13	0.348	**0.657**	0.098	1591	**0.658**	**0.581**	0.155	2685	**0.719**	0.036	**0.536**
70	0.412	0.110	**0.564**	1648	**0.603**	0.256	0.422	2737	**0.719**	**0.634**	0.155
74	0.412	0.108	**0.536**	1717	0.339	**0.657**	0.108	2781	0.072	**0.590**	0.298
100	**0.595**	0.389	0.060	1723	0.465	**0.657**	0.041	2795	**0.652**	0.256	0.467
245	**0.586**	0.389	0.060	1728	0.243	**0.657**	**0.564**	2811	0.144	**0.581**	0.224
253	0.062	**0.634**	0.323	1735	**0.628**	0.256	0.444	2827	**0.637**	0.256	0.453
409	**0.719**	**0.634**	0.155	1741	0.359	0.157	**0.536**	2830	**0.637**	**0.657**	0.021
414	**0.873**	**0.790**	0.155	1745	**0.668**	**0.590**	0.155	2836	**0.719**	**0.634**	0.155
416	0.033	**0.581**	**0.570**	1809	**0.651**	0.256	0.467	2889	**0.506**	0.459	0.155
429	**0.646**	0.256	0.461	1812	0.461	**0.634**	0.042	2986	**0.528**	0.477	0.155
442	0.341	**0.657**	0.106	1845	0.243	**0.634**	**0.536**	2988	**0.719**	**0.634**	0.155
446	**0.590**	0.378	0.065	1858	0.348	**0.657**	0.098	2989	0.412	0.110	**0.564**
462	**0.637**	**0.564**	0.155	1861	**0.514**	0.466	0.155	3006	0.243	**0.588**	0.481
466	**0.652**	**0.634**	0.021	1907	**0.584**	0.256	0.405	3033	0.487	**0.634**	0.037
470	0.366	0.153	**0.564**	1959	**0.587**	0.378	0.065	3037	0.243	**0.657**	**0.564**
472	**0.793**	0.378	0.087	1961	0.243	**0.657**	**0.564**	3044	**0.581**	0.374	0.067
473	**0.641**	0.378	0.068	1962	0.412	0.108	**0.536**	3049	**0.719**	0.127	0.260
474	0.497	**0.657**	0.035	2031	0.336	**0.657**	0.112	3054	0.296	**0.803**	0.218
476	**0.595**	**0.634**	0.025	2037	0.024	**0.634**	**0.707**	3057	0.359	0.157	**0.536**
477	**0.749**	**0.657**	0.016	2041	0.457	**0.657**	0.043	3074	0.243	**0.634**	**0.536**
479	**0.738**	0.256	**0.557**	2048	**0.633**	**0.657**	0.021	3075	**0.514**	0.120	0.257
513	0.472	**0.588**	0.040	2055	**0.668**	**0.590**	0.155	3092	**0.719**	0.132	0.257
518	**0.589**	0.256	0.409	2084	0.412	0.110	**0.564**	3094	0.412	0.108	**0.536**
519	0.339	**0.657**	0.108	2087	0.412	0.110	**0.564**	3095	**0.658**	0.125	0.257
520	0.344	**0.588**	0.090	2125	**0.640**	0.389	0.062	3102	**0.719**	**0.634**	0.155
525	**0.514**	0.466	0.155	2177	0.337	**0.634**	0.105	3106	0.098	**0.803**	0.298
528	**0.958**	**0.634**	0.017	2178	0.243	**0.588**	0.481	3127	**0.719**	0.116	0.271
532	0.190	0.346	**0.610**	2198	**0.727**	**0.657**	0.017	3128	**0.528**	0.062	**0.536**
544	0.207	0.346	**0.647**	2201	**0.873**	**0.789**	0.155	3129	**0.514**	0.466	0.155
561	**0.622**	**0.657**	0.022	2208	0.339	**0.588**	0.094	3134	**0.873**	0.170	0.260
599	0.068	**0.634**	0.311	2209	0.497	**0.634**	0.035	3140	**0.668**	**0.590**	0.155
622	0.338	**0.657**	0.109	2216	0.040	**0.634**	0.412	3142	0.032	**0.581**	**0.593**
632	**0.623**	**0.588**	0.025	2220	**0.884**	0.028	**0.536**	3163	0.366	0.148	**0.536**
640	**0.879**	0.374	0.120	2224	**0.658**	0.111	0.271	3169	**0.593**	0.256	0.412
641	**0.783**	0.374	0.088	2270	**0.618**	0.389	0.061	3179	**0.719**	**0.634**	0.155
642	**0.637**	0.389	0.062	2277	0.243	**0.634**	**0.536**	3181	**0.873**	**0.790**	0.155
643	**0.856**	0.256	**0.710**	2286	**0.586**	0.378	0.065	3188	**0.716**	**0.588**	0.021
649	0.176	**0.634**	0.218	2297	0.348	**0.657**	0.098	3190	**0.902**	**0.634**	0.015
650	0.038	**0.590**	0.461	2335	**0.641**	0.374	0.070	3197	0.040	**0.564**	0.472
654	0.083	0.477	**0.835**	2337	**0.773**	0.389	0.074	3199	0.173	**0.634**	0.219
667	0.193	0.343	**0.594**	2343	**0.590**	0.256	0.410	3200	**0.664**	0.256	0.479
668	**0.754**	0.256	**0.575**	2352	0.059	**0.590**	0.332	3202	**0.884**	**0.803**	0.155
670	0.258	0.352	**0.755**	2363	0.344	**0.634**	0.098	3217	0.066	**0.634**	0.312
1078	**0.637**	**0.564**	0.155	2367	**0.657**	0.389	0.063	3218	0.152	0.352	**0.536**
1085	0.243	**0.588**	0.481	2440	**0.719**	**0.634**	0.155	3220	0.412	0.110	**0.564**
1086	0.348	**0.657**	0.098	2465	0.243	**0.657**	**0.564**	3223	0.040	**0.590**	0.432
1126	0.243	**0.634**	**0.536**	2515	**0.506**	0.111	0.271	3232	**0.719**	0.116	0.271
1283	0.412	0.108	**0.536**	2526	**0.661**	0.374	0.072	3235	0.032	**0.590**	**0.560**
1323	0.243	**0.657**	**0.564**	2555	**0.787**	**0.634**	0.016	3312	**0.719**	0.127	0.260
1420	**0.668**	0.112	0.271	2566	0.243	**0.634**	**0.536**	3317	0.455	**0.657**	0.043
1424	0.412	0.110	**0.564**	2584	0.146	**0.590**	0.224	3320	**0.637**	0.389	0.062
1436	0.348	**0.657**	0.098	2588	0.348	**0.634**	0.094	3327	0.243	**0.634**	**0.536**
1446	**0.719**	**0.634**	0.155	2589	0.145	**0.564**	0.221	3335	**0.661**	0.374	0.072
1447	**0.637**	0.256	0.453	2590	0.152	**0.590**	0.221	3339	0.072	**0.634**	0.300
1451	0.366	0.153	**0.564**	2610	**0.740**	**0.657**	0.016	3345	0.348	**0.634**	0.094
1454	**0.668**	0.112	0.271	2630	**0.668**	0.126	0.257	3365	**0.668**	0.122	0.260
1458	**0.589**	**0.588**	0.027	2635	**0.621**	**0.634**	0.023	3366	**0.611**	0.389	0.061
1459	0.461	**0.634**	0.041	2639	**0.719**	**0.634**	0.155	3369	**0.759**	0.378	0.080
1462	0.173	**0.634**	0.219	2645	**0.609**	0.374	0.068	3388	0.243	**0.588**	0.481
1470	**0.668**	**0.590**	0.155	2652	0.178	**0.634**	0.217	3399	0.344	**0.588**	0.090
1481	0.064	**0.590**	0.315	2657	**0.668**	0.122	0.260	3463	0.038	**0.590**	0.455
1502	0.157	**0.581**	0.218	2658	**0.719**	0.127	0.260	3554	**0.601**	0.389	0.060
1513	**0.719**	**0.634**	0.155	2678	0.176	**0.634**	0.218	3702	0.359	0.163	**0.564**

Please cite this article as: Cheng, Q., et al., Effect of site-specific heterogeneous evolution on phylogenetic reconstruction: A simple evaluation, Gene (2008), doi:10.1016/j.gene.2008.08.003

value); (*iii*) Use the software DIVERGE to statistically detect nucleotide or amino acid residues with rate-shifts, which may have experienced heterogeneous evolution; (*iv*) remove those sites with rate-shift under a specified cutoff, and then infer the phylogenetic tree from the remaining aligned sequences; and (*v*) compare the result from step-*iv* with step-*i* to test whether heterogeneous evolution misleads the phylogenetic inference.

3. Case-study

3.1. Maximum parsimony phylogenetic tree

The multiple alignment of 20 fish mitogenomic protein-coding sites includes 3776 sites, among of which 1079 characters are parsimony-informative. We inferred the maximum parsimony tree

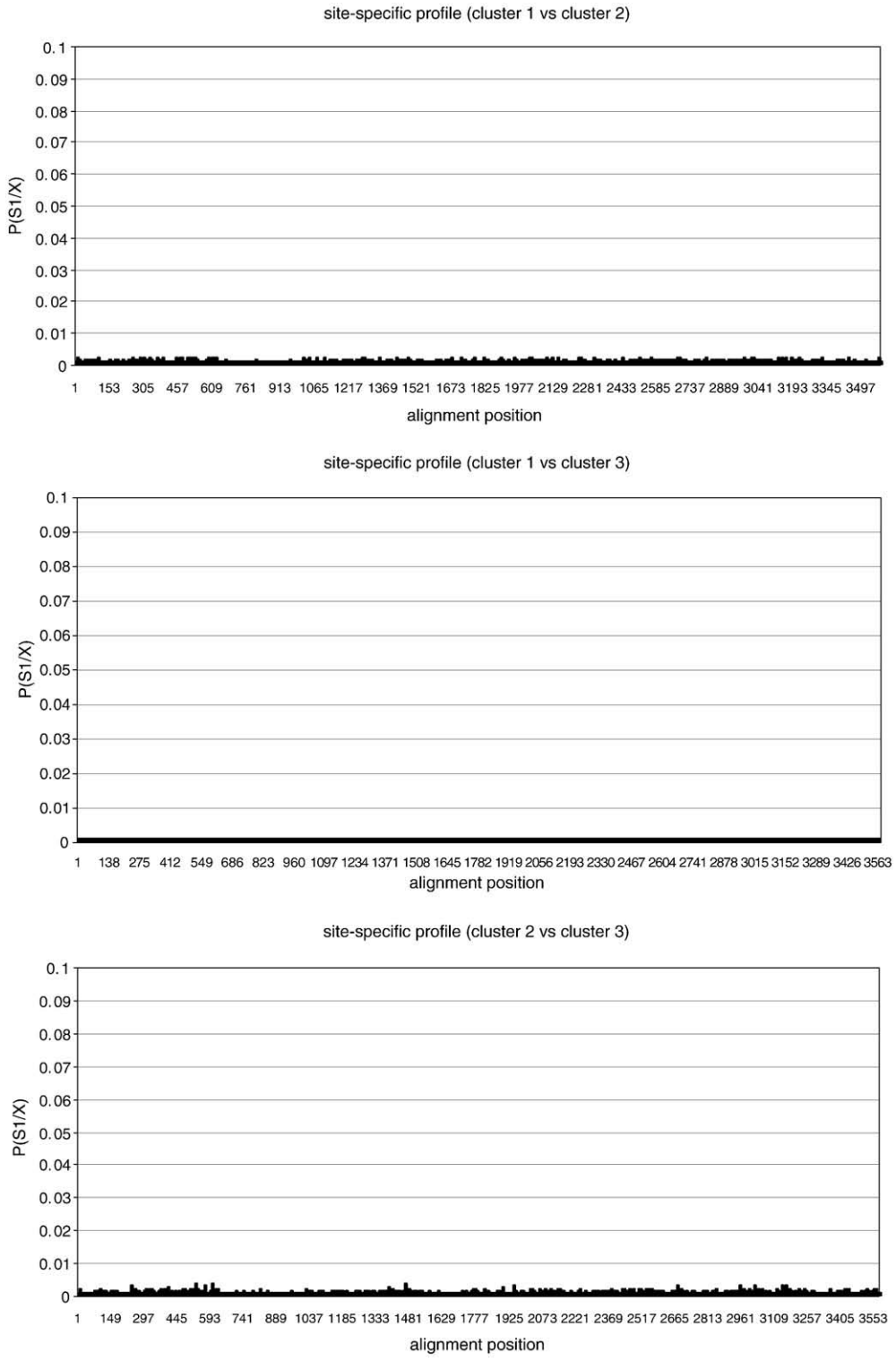

Fig. 3. Site-specific profile of posterior probability between cluster 1, cluster 2, and cluster 3 (cutoff value = 0.5).

Please cite this article as: Cheng, Q., et al., Effect of site-specific heterogeneous evolution on phylogenetic reconstruction: A simple evaluation, Gene (2008), doi:10.1016/j.gene.2008.08.003

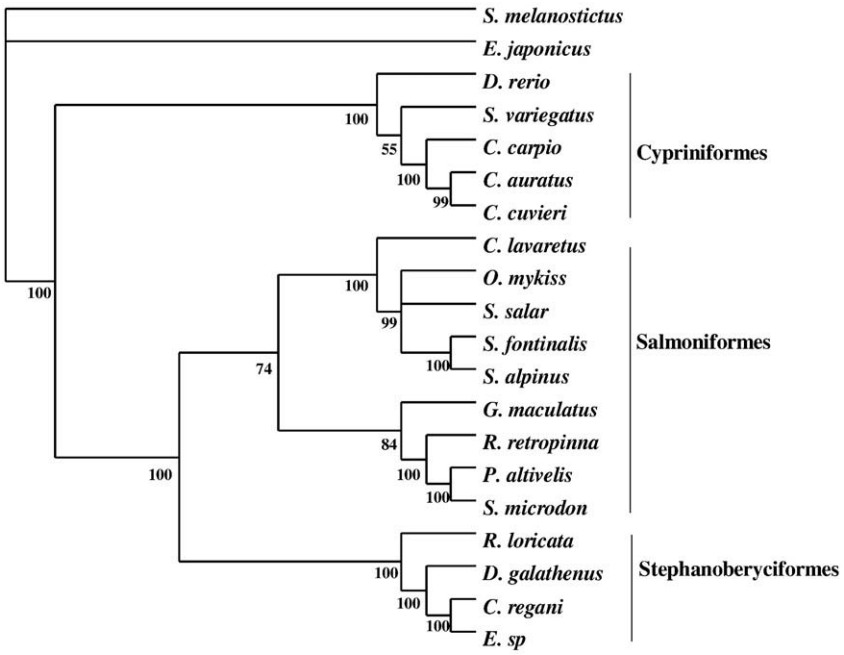

Fig. 4. Maximum parsimony phylogenetic tree of the fishes inferred from mitogenomic protein-coding sequences with specific sites being removed (bootstrap 1000 times) (the three groups of Cypriniformes, Salmoniformes, and Stephanoberyciformes are respectively designated as cluster 1, cluster 2, and cluster 3).

followed by bootstrapping (1000 replicates) to examine the statistical reliability. As shown in Fig. 1 ("before tree"), it clearly shows that three orders, Cypriniformes, Salmoniformes, and Stephanoberyciformes, are monophyletic, which, for simplicity, are denoted by clusters 1, 2, and 3, respectively, in the following analysis.

3.2. Predicting the critical residues for type I functional divergence (altered functional constraints) between clusters

Using DIVERGE, we estimated that the coefficient of functional divergence between these three clusters are $\theta = 0.266 \pm 0.025$ (cluster 1 vs cluster 2), 0.276 ± 0.032 (cluster 1 vs cluster 3), and 0.170 ± 0.028 (cluster 2 vs cluster 3), respectively (Table 2). This result implies that heterogeneous evolution between the monophyletic groups of Cypriniformes, Salmoniformes, and Stephanoberyciformes are statistically significant. Next we used the posterior probability to predict those critical amino acid residues that are responsible for the heterogeneous evolution between these clusters, as shown in Fig. 2. Among 3776 aligned sites, we found 198 sites (~5.2%) that show the posterior probability >0.5 at least in one of three pairwise comparisons (Table 3), indicating that these sites may have experienced heterogeneous evolution. The fact that the majority of amino acid residues have scores lower than 50% indicates that heterogeneous evolution may occur rarely in fish mitogenome.

Although the posterior analysis is widely used in bioinformatics, the cutoff value for residue selection is usually empirical. At any rate, we found that when these 198 sites with posterior probability >0.5 (an empirical value) were removed from the multiple alignment, the estimation of θ for each pairwise clusters is virtually 0 (Table 2). From Fig. 3, one can see clearly that the signal of sites with rate-shifts (heterogeneous evolution) vanished almost completely.

The maximum parsimony phylogenetic tree inferred from the multiple aligned sequences after removing the 198 sites with heterogeneous evolution is shown in Fig. 4 ("after tree"). Comparing this MP tree with Fig. 1, we conclude that these two phylogenies are almost the same except for some slight differences in Cluster 1 and Cluster 2. *Sarcocheilichthys variegatus* is at the basal position of Cluster 1 in "before tree", while *Danio rerio* is at the basal position of Cluster 1 in "after tree". In "before tree", the phylogenetic relationship between *Salmo salar* and *Oncorhynchus mykiss* is solved in Cluster 2, while their relationship cannot be solved in "after tree", indicating that the heterogeneous evolution may mislead the phylogenetic reconstruction.

4. Discussion

Though the debate on phylogenetic inference continues, recent studies have shown that the effect of heterogeneous evolution should be considered. Our method has been widely used for detecting heterogeneous evolution in gene families or events of gene duplication (Gu, 1999; Gu and Vander, 2002), it has not been used in phylogenetic reconstruction yet. Here, we present a simple method implemented in the software DIVERGE to evaluate the effect in practical phylogenetic inference, with a case-study of 20 fish mitogenomic protein-coding sequences. We found statistical evidence for that heterogeneous evolution during the mitochondrial protein sequences. Interestingly, only a small portion (~5%) of amino acid residues was found to be involved in heterogeneous evolution. After removing these sites, the effect of heterogeneous evolution is virtually eliminated. This provides a practically feasible approach to evaluating whether the inferred phylogeny can be misled by the effect of heterogeneous evolution. In short, we suggest that for phylogenetic reconstruction, the most efficient approach is to identify and remove those sites with heterogeneous evolution.

At the same time, we shall note the limitation of this approach. If the total number of sites is small (less than 300 amino acid sites), this approach performs not well. If the number of heterogeneous sites is large (accounting for more than 15% of total sites), this approach will produce misleading results and cannot be used. Our future study will to improve this heuristic algorithm so that it can be useful in the case of large phylogenetic inference.

Acknowledgments

This work is supported by National Natural Sciences Foundation of China to Gu and Zhong (30600456), and China Postdoctoral Science Foundation to Cheng (20060400622).

References

Darwin, C., 1859. The Origin of Species. Murray, London.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Felsenstein, J., 2004. Inferring phylogenies. Sinauer Associates, Sunderland, Massachusetts.

Galtier, N., 2001. Maximum-likelihood phylogenetic analysis under a covarion-like model. Mol. Biol. Evol. 18, 866–873.

Gaucher, E.A., Miyamoto, M.M., 2005. A call for likelihood phylogenetics even when the process of sequence evolution is heterogenerous. Mol. Phylogenet. Evol. 37, 928–931.

Gaucher, E.A., Gu, X., Miyamoto, M.M., Benner, S.A., 2002. Predicting functional divergence in protein evolution by site-specifc rate shifts. Trends Biochem. Sci. 27, 315–321.

Gu, X., 1999. Statistical methods for testing functional divergence after gene duplication. Mol. Biol. Evol. 16, 1664–1674.

Gu, X., Vander, V.K., 2002. DIVERGE: phylogeny-based analysis for functional–structural divergence of a protein family. Bioinformatics 18, 500–501.

Huelsenbeck, J.P., 2002. Testing a Covariotide Model of DNA Substitution. Mol. Biol. Evol. 19, 698–707.

Inoue, J.G., Miya, M., Tsukamoto, K., Nishida, M., 2001. A mitogenomic perspective on the basal teleostean phylogeny: resolving higher-level relationships with longer DNA sequences. Mol. Phylogenet. Evol. 20, 275–285.

Kolaczkowski, B., Thorton, J.W., 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431, 980–984.

Li, W.H., 1997. Molecular Evolution. Sinauer Associates, Sunderland, Massachusetts.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford.

Lopez, P., Casane, D., Philippe, H., 2002. Heterotachy, an important process of protein evolution. Mol. Biol. Evol. 19, 1–7.

Miya, M., Kawaguchi, A., Nishida, M., 2001. Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Mol. Biol. Evol. 18, 1993–2009.

Miya, M., et al., 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial sequences. Mol. Phylogenet. Evol. 26, 121–138.

Oparina, N.J., Kalinina, O.V., Gelfand, M.S., Kisselev, L.L., 2005. Common and specific amino acid residues in the prokaryotic polypeptide release factors RF1 and RF2: possible functional implications. Nucleic Acids Res. 33, 5226–5234.

Swofford, D.L., 2001. PAUP*, Phylogenetic Analysis Using Parsimony (*and Other Methods).Version 4. Sinauer Associates, Sunderland.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal X windows interface: flexible strategies for multiple sequences alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Tuffley, C., Steel, M., 1998. Modeling the covarion hypothesis of nuclear substitution. Math. Biosci. 147, 63–91.

Wang, Y., Gu, X., 2001. Functional divergence in the caspase gene family and altered functional constraints: statistical analysis and prediction. Genetics 158, 1311–1320.

Please cite this article as: Cheng, Q., et al., Effect of site-specific heterogeneous evolution on phylogenetic reconstruction: A simple evaluation, Gene (2008), doi:10.1016/j.gene.2008.08.003