Tropomyosin is a nice marker gene for phylogenetic analysis of molluscs

Xiaotong Wang · Li Li · Fei Xu · Guofan Zhang

Abstract Molluscs are an extraordinarily diverse group of animals and to discriminate them based on one molecular marker/gene is very difficult because of the too fast or slow rate of nucleotide substitution. In the study, the tropomyosin cds (coding sequences) of 43 animal species were analyzed, the results of which suggested that the tropomyosin gene was a nice marker gene to phylogenetic analysis of molluscs, even for all the studied animals. In addition, InDels (insertions and deletions) in tropomyosin cds of *turbo cornutus* were also studied and one segment repeat, which probably happened recently and was of functional importance, was found.

Keywords Mollusc · Tropomyosin · Phylogenetic analysis · InDels

Introduction

Molluscs are an extraordinarily diverse group of animals with an estimated 200,000 species, second only to the Phylum Arthropoda. Some phylogenetic studies on molluscs have been reported at lower taxonomy levels, such as Genus and Family, et al. [1–3]. But at higher taxonomy levels, such as Class, molecular classification of molluscs was difficult because of the few genomic sequence information and inappropriate rate of nucleotide substitution. Though there were more complete mitochondrial genome

sequences of molluscs deposited in GenBank, the large divergency of mitochondrial genes among classes, even among families, made it almost impossible that the species were distinguished using the mitochondrial genes. So, a nice marker gene for phylogenetic analysis of molluscs probably could be chosen from nuclear genes. Here, we evaluated some nuclear genes present in GenBank for their potential phylogenetic information and found that tropomyosin is a nice marker gene for phylogenetic analysis of molluscs.

Materials and methods

Data collection

The tropomyosin cds sequences of the following 41 bilaterians, *Mizuhopecten yessoensis, Argopecten irradian, Sinonovacula constricta, Crassostrea gigas, Mytilus edulis, Solen strictus, Tresus keenae, Pseudocardium sachalinensis, Fulvia mutica, Scapharca broughtonii, Venerupis philippinarum, Chlamys nipponensis, Neptunea polycostata, Turbo cornutus, Biomphalaria glabrata, Haliotis asinina, Haliotis discus, Helix aspersa, Octopus vulgaris, Sepioteuthis lessoniana, Ommastrephes bartramii, Todarodes pacificus, Squilla aculeata, Scolopendra sp, Caenorhabditis elegans, Anisakis simplex, Schistosoma haematobium, Clonorchis sinensis, Danio rerio, Salmo salar, Thunnus thynnus, Pennahia argentata, Xenopus tropicalis, Rana catesbeiana, Ovis aries, Mus musculus, Oryctolagus cuniculus, Cervus elaphus, Homo sapiens, Gallus gallus* and *Japanese quail*, and 2 cnidarians, *Nematostella vectensis,* and *Podocoryna carnea*, were acquired from GenBank. For bilaterians, long-form transcripts (translated to 284 aa) of tropomyosin genes were

X. Wang · L. Li · F. Xu · G. Zhang (✉)
Institute of Oceanology, Chinese Academy of Sciences,
7 Nanhai Road, Qingdao 266071, China
e-mail: gfzhang@qdio.ac.cn

chosen for analysis, and for cnidarians, short-form transcripts (translated to 242 aa) were done because cnidarians only encode short-form transcripts.

Nucleotide substitution saturation detection

The stop codons of the obtained sequences were removed before further analysis. Based on the translated amino acid sequences, we conducted a comparative analysis of complete tropomyosin cds (without stop codons) of above species. According to the selected optimum replacement model, the nucleotide substitution saturation was tested using DAMBE V.4.5.2 [4].

Construction of phylogenetic tree

The above sequences were multi-aligned using Clustal W (http://www.ebi.ac.uk/Tools/clustalw) and edited manually. Aligned Sequences were studied using MEGA 4 [5] software for phylogenetic inference. The alignment file included all the sites with missing/ambiguous data and gaps. The nucleotide sequence ME tree (Model is Jukes-Cantor) was constructed using MEGA 4.0 accompanied with 1000 bootstrap tests. Two kinds of cnidarians, *Nematostella vectensis* and *Podocoryna carnea*, were used as the outgroup in this study.

Detection of positive selection sites

The dN (nonsynonymous substitution rate), dS (synonymous substitution rate), and the ω values (dN/dS) were calculated by popular methods described in PAML package [6]. PAML 4 [7] was used to analyze the variation of selective pressures during tropomyosin evolution. This program used three models, which are all based on the comparison of two sub-models. The first model is the "branch-specific" model. One of its sub-models is one ratio model, which assumes that the values of ω of all evolutionary branches are the same, while another is free ratio model, which assumes that they are all different. LRT (Likelihood ratio test) of these two sub-models can test whether the value of ω changes in different branches. The second model is the "site-specific" model. It assumes that the value of ω can be changed among different sites. There are two pairs of sub-models for the "site-specific" model, M1a and M2a as well as M8a and M8, respectively. The third model is the "branch-sitespecific" model, which divides the branches of the evolutionary tree into the "foreground" and "background" categories. The ω values may be variable in the two categories and the value of ω of "foreground" is higher than that of "background". It further assumes that the ω value of each category is fixed, but can differ among different sites.

Analysis of sequence variation and prediction of motif change

Sequence variations, especially InDels found in the tropomyosin sequence of *turbo cornutus*, were analyzed, and the changes of motif were predicted at http://hits.isb-sib.ch/cgi-bin/PFSCAN.

Results and discussion

Nucleotide substitution saturation is little

After the nucleotide substitution saturation was detected among these species using DAMBE V.4.5.2, it was found that the observed Iss value was significantly lower than Iss.c when assuming a symmetrical topology (Table 1; $P = 0.00 < 0.01$), which hints that the nucleotide substitution saturation is little among the tropomyosin sequences of all the tested species and construction of the phylogenetic tree could be carried out according to usual methods.

A nice mollusc phylogenetic tree was constructed

For molluscs, it is very difficult to find a nice molecular marker to distinguish each other because the nucleotide substitution was usually saturated. In this study, a phylogenetic tree was constructed base on tropomyosin sequences of all the studied species (Fig. 1). Based on tropomyosin sequences, all the molluscs were arranged to the same evolutionary location with classical taxonomy, which suggested that tropomyosin was a nice marker gene for phyligenetic analysis of molluscs. In the last two decades, mitochondrial DNA was frequently used to describe the evolution of mollusks, but usually limited within the level of Genus or Family [8, 9]. The results of this study suggested that the tropomyosin sequence could classify the molluscs at lower or higher taxonomy levels. To our knowledge, it is the first time that the phyligenetic tree of entire Phylum Mollusca was successfully constructed based on only one gene sequence.

In addition, the whole phylogenetic tree based on all tested tropomyosin sequences was consistent with the

Table 1 Test of substitution saturation of the tropomyosin sequences of the tested species

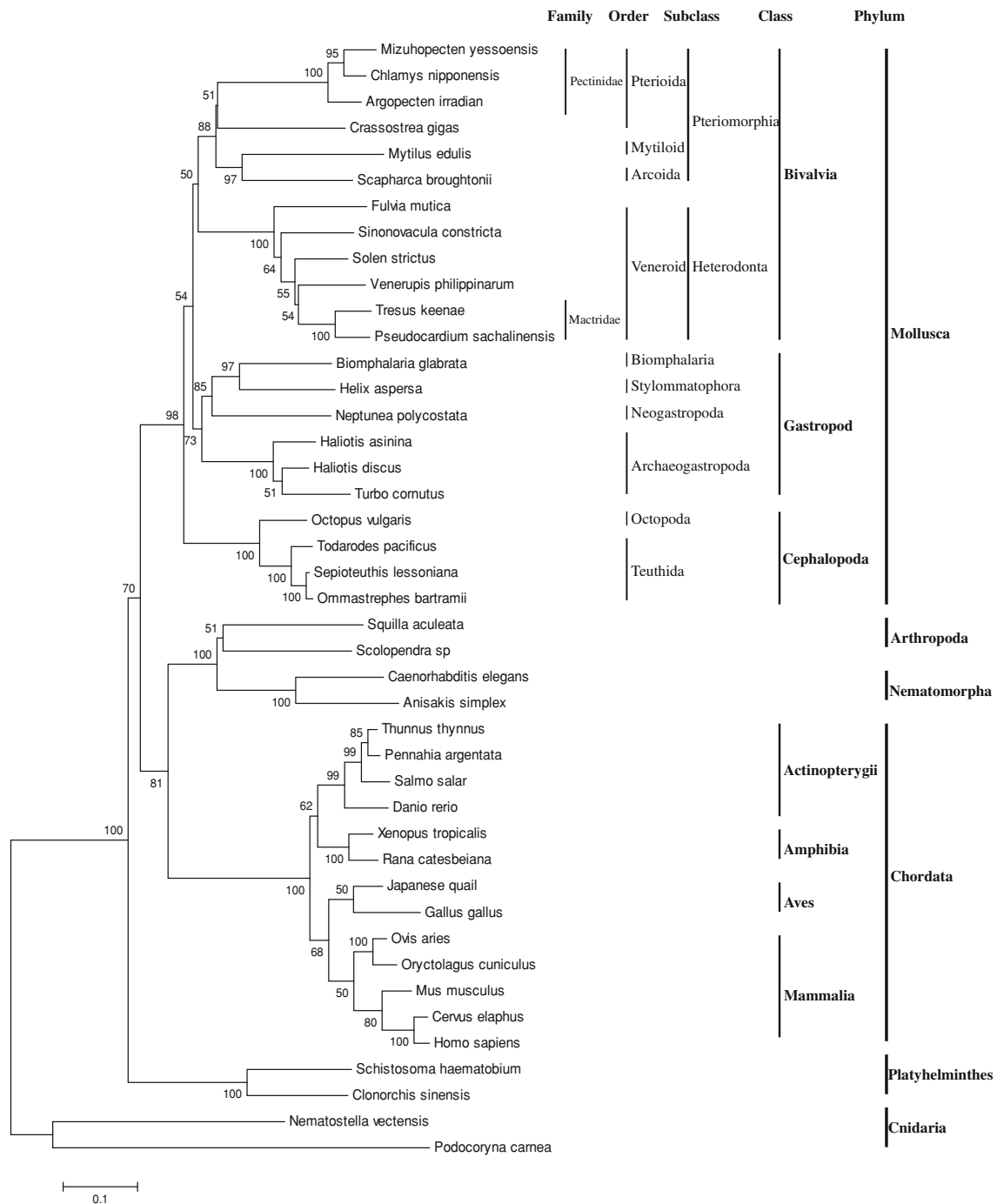| NumOTU	Iss	Iss.cSym	T value	DF	*P*
4	0.511	0.814	14.875	872	0.0000
8	0.523	0.779	10.774	872	0.0000
16	0.537	0.756	8.197	872	0.0000
32	0.552	0.738	6.109	872	0.0000

Note two-tailed tests are used

Fig. 1 The ME tree constructed based on the tropomyosin cds sequences of all the studied species

classical species phylogenetic tree based on morphology, which indicated that the tropomyosin gene could also be regarded as a useful marker gene for the phylogenesis study of most animals.

No positive selection sites was found

Both genetic factors and selection factors have effect during the evolutionary history of mollusc genes. Two parameters,

dN and dS, play an important role in estimating selection. The ratio of dN/dS provides a measure of the selection pressure to which a gene pair is subject. If dN/dS is <1, we consider that purifying selection acts on this gene; if dN/dS is >1, we consider that positive selection acts on this gene; if dN/dS = 1, then no selection acts on this gene, which means that this gene evolved neutrally; if dN/dS is <1, which means that some replacement substitutions have been purged by natural selection, presumably because of their
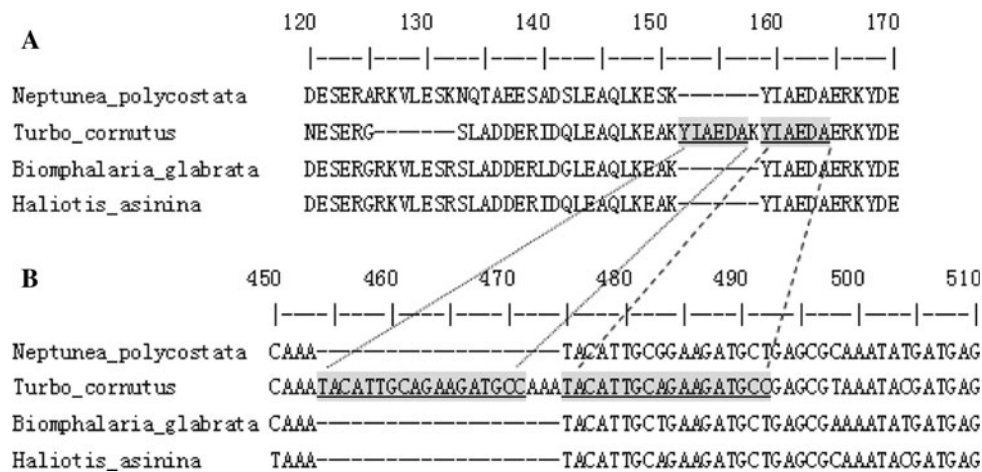
Fig. 2 The InDels and amino acid fragment repeats in tropomyosin gene of *Turbo cornutus*. There were 1 deletion (aa 126–132) and 1 insertion (aa 152–158) found in *turbo cornutus* (*Panel a*), compared with other bilaterian animals in this study. It is very interesting that the insertion (aa 152–157) is the same with the sequence (aa 159–164) (*Panel a*) and their DNA sequences are also the same (*Panel b*)

deleterious effects. The smaller the dN/dS ratio is, the greater the number of eliminated substitutions and the greater the selective constraint under which the genes have evolved. Tropomyosin is an actin-binding protein that regulates actin mechanics. It is important, among other things, for muscle contraction. The nonsynonymous mutations of this gene can cause some severe diseasese [10, 11], suggesting it is functionally important. In this study, we found that all pairwise ratios of dN/dS were always less than 1 and 0 sites was positively selected, meaning that these genes are under purifying selection for very strong functional constraint. Once deleterious mutation (nonsynonymous mutation) occurs, this deleterious mutation can not fix in population because purifying selection may act on this mutation and remove it, so the detected dN was less than dS.

Two InDels in tropomyosin of turbo cornutus and their meaning

The InDels were found in Only three species, *turbo cornutus*, *Nematostella vectensis* and *Podocoryna carnea*. *Nematostella vectensis* and *Podocoryna carnea* are Cnidaria, belonging to Radiata, and their sequence are 42 amino acids shorter than the other animals, all belonging to Bilateria, at the sequence start in the study, which are understandable because of the very long divergency time between Radiata and Bilateria. More interesting, there were 1 deletion (aa 126–132) and 1 insertion (aa 152–158) found in *turbo cornutus*, compared with other Gastropod animals and other Bilaterias in this study. It is very interesting that the insertion (aa 152–157) is the same with the sequence (aa 159–164) and their DNA sequences are also the same (Fig. 2, Panel a, b), which indicated that segment repeating

Table 2 Two new domains caused by the InDels

Species	Domain name	Domain position
Turbo cornutus	CK2_PHOSPHO_SITE	127–130
	TYR_PHOSPHO_SITE	145–153
Biomphalaria glabrata	AMIDATION	125–128

happened recently because the dN and dS are both 0 between the insertion (aa 152–157) and the sequence (aa 159–164). Next, the domain changes were scanned at http://hits.isb-sib.ch/cgi-bin/PFSCAN and two new domains (Table 2) appeared because of the InDels, which suggested the InDels in the tropomyosin sequence of *turbo cornutus* was of functional importance. It has been reported that positive selection could also act on InDels in the evolution of a protein [12]. In this study, because of the short period from InDels appearing and their functional importance, it could be deduced that the cause for the unique InDels fixed in *turbo cornutus* genome was also positive selection.

References

1. Maeda T, Kajita T, Maruyama T, Hirano Y (2010) Molecular phylogeny of the sacoglossa, with a discussion of gain and loss of kleptoplasty in the evolution of the group. Biol Bull 219:17–26

2. Reid DG, Dyal P, Williams ST (2010) Global diversification of mangrove fauna: a molecular phylogeny of Littoraria (Gastropoda: Littorinidae). Mol Phylogenet Evol 55:185–201

3. Sengupta ME, Kristensen TK, Madsen H, Jørgensen A (2009) Molecular phylogenetic investigations of the Viviparidae (Gastropoda: Caenogastropoda) in the lakes of the Rift Valley area of Africa. Mol Phylogenet Evol 52:797–805

4. Xia X, Xie Z, Salemi M, Chen L, Wang Y (2003) An index of substitution saturation and its application. Mol Phylogenet Evol 26:1–7

5. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA 4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

6. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13: 555–556

7. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

8. Barco A, Claremont M, Reid DG, Houart R, Bouchet P, Williams ST, Cruaud C, Couloux A, Oliverio M (2010) A molecular phylogenetic framework for the Muricidae, a diverse family of carnivorous gastropods. Mol Phylogenet Evol 56:1025–1039

9. Miura O, Torchin ME, Bermingham E (2010) Molecular phylogenetics reveals differential divergence of coastal snails separated by the Isthmus of Panama. Mol Phylogenet Evol 56:40–48

10. Ochala J, Iwamoto H, Larsson L, Yagi N (2010) A myopathy-linked tropomyosin mutation severely alters thin filament conformational changes during activation. PNAS 107:9807–9812

11. Lakdawala NK, Dellefave L, Redwood CS, Sparks E, Cirino AL, Depalma S, Colan SD, Funke B, Zimmerman RS, Robinson P, Watkins H, Seidman CE, Seidman JG, McNally EM, Ho CY (2010) Familial dilated cardiomyopathy caused by an alpha-tropomyosin mutation: the distinctive natural history of sarcomeric dilated cardiomyopathy. J Am Coll Cardiol 55:320–329

12. Podlaha O, Zhang J (2003) Positive selection on protein-length in the evolution of a primate sperm ion channel. PNAS 100:12241–12246