Evolutionary analysis of amphioxus myc gene

journal or publication title	東京海洋大学研究報告
volume	5
page range	11-16
year	2009-03-27
URL	http://id.nii.ac.jp/1342/00000347/

Journal of the Tokyo University of Marine Science and Technology, Vol. 5, pp. 11-16, 2009

Evolutionary analysis of amphioxus *myc* gene

Mikiharu MORITA[*1], Kunihiko FUTAMI[*1], Huan ZHANG[*2] , Kaoru KUBOKAWA[*3] , Yoshio OJIMA[*4] and Nobuaki OKAMOTO[*1]

（Accepted December 11, 2008）

Abstract: The proto-oncogene *myc* is one of the most important genes controlling cell proliferation. The vertebrate genome has four *myc* genes (c-, N-, L-, s-*myc*), whose evolutionary origin and relationship are unclear. Here, we isolated a *myc* gene from a protochordate, the amphioxus *Branchiostoma belcheri*, which is thought to be the nonvertebrate that is closest to the vertebrates. A 1480 bp cDNA sequence was determined and contains an ATG-initiated ORF consisting of 371 amino acids. The exon/intron structure was conserved. Southern blotting and degenerate PCR showed that the amphioxus genome contained only a single *myc* gene. A phylogenetic tree of Myc family genes based on the deduced amino acid sequences indicated that amphioxus Myc was located outside the vertebrate Myc family. These results suggest that *myc* gene duplication occurred after protochordate on phylogeny.

Key words: protochordate, evolution, genome duplication, Myc family, vertebrate

Introduction

The proto-oncogene *myc* is thought to be one of the most important genes controlling cell proliferation[1]. It is crucial for progression of the cell cycle, cell growth, and differentiation. Four *myc* genes (c-, N-, L-, s-*myc*) constitute the *myc* family[2]. The *myc* family is a subgroup of the superfamily of basic-helix-loop-helix-leucine-zipper (bHLH-LZ) transcription regulators[3]. The basic region of the bHLH-LZ domain constitutes the DNA binding motif and the contiguous HLH-LZ region mediates dimerization with other proteins that have a HLH-LZ motif. Mammalian *myc* genes have several structural similarities. They consist of three exons and two introns. The first exon is a non-coding exon that acts as a transcriptional regulator of the *myc* gene[4]. The second and third exons encode the Myc protein. The second exon contains three boxes, A, B and C that are required for the transforming activity of the Myc protein. *myc* genes are highly conserved in vertebrates, suggesting that they have important functions. However, the evolutionary origin and relationships among *myc* genes are unclear. Single *myc* genes have been isolated from invertebrates such as sea star[5], sea urchin (acc. no. L37056) and *Drosophila*[6], however, existence of *myc* family in invertebrate is unknown. We assumed that a single *myc* gene was most likely present in the stem species of vertebrates.

To better understand the origin of vertebrate *myc* genes, we examined *myc* homologs and their copy numbers in the genome of amphioxus, *Branchiostoma belcheri*, which is the nonvertebrate animal that is closest to vertebrates[7].

Materials and methods

i. Isolation of *myc* cDNA

Amphioxus were obtained from Ocean Research Institute, the University of Tokyo and were immediately frozen for extraction of total RNA. Total RNA of an amphioxus was extracted from the whole body using Sepasol RNA I Super (Nacalai Tesque) according to the supplier's method. Total RNAs were reverse-transcribed to cDNAs using an oligo(dT) primer with reverse transcriptase (SuperScript II, Invitrogen) and the cDNAs were used as a template for PCR amplification. Two degenerate primers, sense primer Amphi-*myc* Box A (5'-CCV WSY GAG GAY ATY TGG AAG-3') and antisense primer Amphi-*myc* Box C (5'-STB TCR BTB AKH GGG TAK GGG-3'), were designed from vertebrates (human[8], chicken[9], *Xenopus laevis*[10], rainbow trout[11] and common carp[12] c-*myc* sequences (Box A and Box C). PCR conditions consisted of denaturation at 94°C for 1 min, followed by 35 cycles of each of denaturation at 94°C for 30 s, annealing at 50°C for 30 s, extension at 72°C for 30 s, and a final extension step of 72°C for 10 min. A single PCR product was obtained and was subcloned into pBluescriptII SK(-) and sequenced. This sequence information was used to design primers for both 5' RACE nested primer (5'-TCT CAG CGA CCG TGG GAA TGA ATG-3') and 3' RACE nested primer (5'-CGA AGA AAA GGT TGA GAA GGC AGC-3').

*1 Department of Marine Biosciences, Faculty of Marine Science, Tokyo University of Marine Science and Technology, 4-5-7 Konan, Minato-ku, Tokyo 108-8477, Japan（東京海洋大学海洋科学部海洋生物資源学科）
*2 Department of Marine Sciences, University of Connecticut, Groton, CT06340, USA（コネチカット大学）
*3 Ocean Research Institute, The University of Tokyo, 1-15-1 Minamidai, Nakano-ku, Tokyo 164-8639, Japan（東京大学海洋研究所）
*4 Japan Marine Science College, Osaka 551-0002, Japan（日本魚類生物科学研究所）

5' RACE was performed as described previously[13]. The PCR reaction of dC-tailed cDNA was performed using Ex Taq (TaKaRa Bio). The primers were anchor primer (5'-GGC CAC GCG TCG ACT AGT ACG GGI IGG GII GGG IIG-3') and Amphi-*myc* Box C primer. PCR conditions consisted of denaturation at 94°C for 2 min, followed by 35 cycles of each of denaturation at 94°C for 30 s, annealing at 50°C for 30 s, extension at 72°C for 30 s, and final extension step of 72°C for 3 min. The amplified fragments were separated by agarose gel electrophoresis. The predicted fragments were eluted from the gel and used as the template for the secondary amplification of nested PCR. In this secondary reaction, universal amplification primer (UAP: 5'-GGC CAC GCG TCG ACT AGT AC-3') and the 5' RACE nested primer were used. PCR conditions consisted of denaturation at 94°C for 2 min, followed by 35 cycles of each of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 30 s, and final extension step of 72°C for 3 min. The PCR fragment was subcloned and sequenced.

The 3' downstream region was determined by the 3' RACE method. Total RNA was subjected to reverse transcription by reverse transcriptase (SuperScript II, Invitrogen) using oligo(dT)-containing adapter primer (5'-GGC CAC GCG TCG ACT AGT $AC(T)_{17}$-3'). The PCR reaction was performed using AmpliTaq Gold (Applied Biosystems). The primers were universal amplification primer and Amphi-*myc* Box A primer. PCR conditions consisted of denaturation at 95°C for 10 min, followed by 35 cycles of each of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 1.5 min, and final extension step of 72°C for 10 min. The predicted fragments were separated by the same method as previously mentioned and used as the template for the secondary amplification of nested PCR. The primers were UAP and 3' RACE nested primer. PCR conditions were the same as those in the first PCR. The amplified fragments were subcloned and sequenced.

ii. Determination of the exon/intron boundaries

To determine the exon/intron structure, PCR was performed using genomic DNA as template. Gene specific primers were used to amplify Amphi-*myc* fragments containing the exon/intron borders. The following primer sets were used: Intron 1, 5'-AAC GCT CTG GAA TAT AGA GTG GAG-3' (forward); 5'-GTC TGT CTC CTC GTA GAA GTA GGG-3' (reverse). Intron 2, 5'-CAT GGA CTA CAC CCG TAC CGA CT-3' (forward); 5'-GTG GCA CAG TAG TTG TGA TGG AC-3' (reverse). PCR amplification was started with a 2 min hold at 95°C, followed by 35 cycles of 15 s at 95°C, and 2.5 min at 65°C with a post-extension of 7 min at 72°C. The PCR products were directly sequenced.

iii. Software

Comparison with known genes and proteins was carried out using BLAST (http://www.ncbi.nlm.nih.gov/BLAST/). A pairwise alignment was performed with the EMBOSS Pairwise Alignment Algorithms (http://www.ebi.ac.uk/emboss/align/index.html). Protein domains were identified by using InterProScan (http://www.ebi.ac.uk/InterProScan/). Subcellular localization was predicted by using PSORT II (http://psort.hgc.jp/form2.html).

iv. Genomic Southern blot analysis

Ten micrograms of amphioxus genomic DNA was digested with *Eco*RI, *Pst*I or *Xba*I, and electrophoresed in a 0.8% agarose gel and transferred with 0.4 N NaOH to a nylon membrane (Hybond N+, Amersham Bioscience). The blot was hybridized with a ^{32}P-labeled 342 bp probe which comprised Box A and Box C of *myc* in amphioxus. The probe was amplified by PCR using cDNA clone as a template. Low-stringency hybridization was carried out in 6 × standard saline citrate (SSC), 0.1% SDS, 10 × Denhardt's, and 1 mM EDTA overnight at 65°C. Washes were performed in 1 × SSC/0.1% SDS for 20 min four times at 55°C which allows about 30% mismatch[14].

v. Degenerate PCR

Two degenerate primers were used that are capable of recognizing all vertebrate *myc* genes. The sense primer was based on Box A (5'-GAR GAY ATH TGG AAR AAR TTY G-3') and the antisense primer was based on Box B (5'-WRA AIC CIS WCC ACA TRC ART C-3'). The PCR was performed for 35 cycles at an annealing temperature of 50°C. The amplified fragments were subcloned and 16 independent clones were sequenced.

vi. Phylogenetic analysis

Amino acid sequences were aligned with ClustalX[15], and manually adjusted to increase the similarity. A phylogenetic tree for most of the members in the Myc family was constructed from amino acid data by the neighbor-joining method[16]. Positions with gaps were excluded. The reliability of the tree topology was evaluated by bootstrap analyses with 500 replicates.

Results

i. Identification and structure of a *myc* gene from amphioxus

Using consensus vertebrate *myc* primers, a 342-bp cDNA fragment was obtained from amphioxus total RNA by RT-PCR and cloned. The fragment contained sequences that had high homologies with boxes A, B and C of the vertebrate *myc* family genes, indicating that it was part of the amphioxus *myc* gene. Following

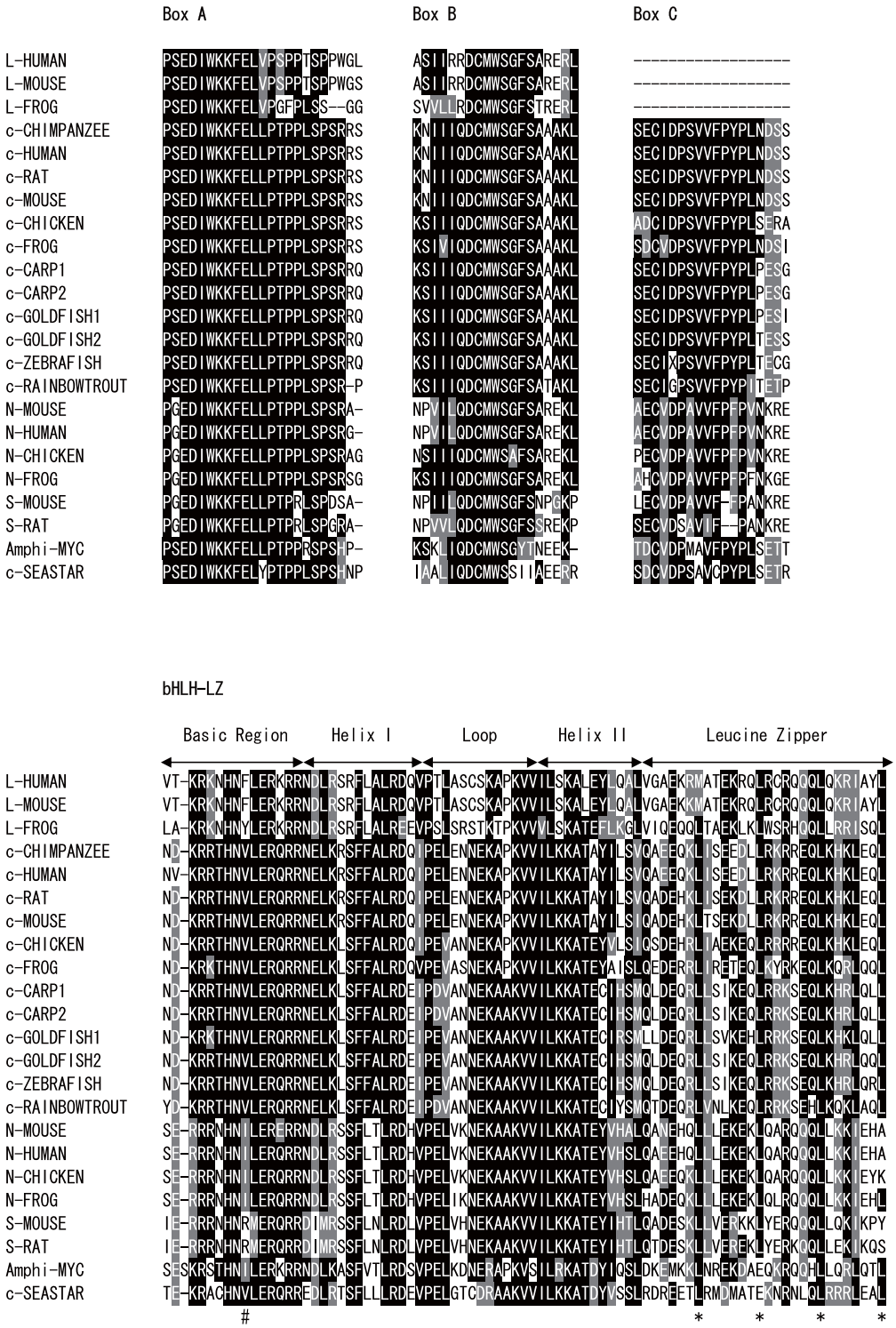

Fig. 1. Alignment of three boxes (Box A, B and C) and bHLH-LZ domain of the Myc family. Box C is not found in L-Myc. Identical and similar residues are highlighted in black and gray, respectively. The conserved leucines of the leucine zipper domain are starred. Gaps (-) were introduced to optimize identity. An isoleucine in the basic region (#) is uniquely conserved in all vertebrate N-Myc, in contrast to valine in all vertebrate c-Myc.

3' RACE and 5' RACE, a 1480-bp full-length cDNA was obtained (DDBJ acc. no. AB175932). The gene, called Amphi-*myc*, contains an ATG-initiated ORF consisting of 371 amino acids. It shares 36.2, 32.6 and 32.0% amino acid identities with human c-, N- and L-Myc, respectively, and 50.5, 44.3 and 43.4% similarities, respectively. Genomic PCR revealed that Amphi-*myc* shared a three-exon structure, and the junction between exon 2 and exon 3 was located at the same codon as in vertebrate *myc* genes (data not shown).

The three boxes and the bHLH-LZ domain are conserved (Fig. 1). The bHLH domain and leucine zipper were also identified by InterProScan (data not shown). PSORT II predicts that Amphi-Myc is a nuclear protein (data not shown). These results suggest that Amphi-Myc, like vertebrate Myc proteins, functions as a transcription factor.

ii. Copy number in the amphioxus genome

Fig. 1 indicates that the only Myc protein with an isoleucine in the basic region is N-Myc. In fact, an isoleucine at this position is uniquely conserved in all vertebrate N-Myc, in contrast to valine in all vertebrate c-Myc. In the absence of firm evidence that there are no other genes, there was a possibility that we had cloned amphioxus N-*myc*. Therefore, we next investigated how many *myc* genes existed in the amphioxus genome. Using the 342 bp PCR fragment of amphioxus *myc* as a probe, a single band was observed in a Southern blot (Fig. 2). To look for similar *myc* genes, we carried out a genomic PCR with degenerate primers that were designed to amplify fragments between *myc* Box A and *myc* Box B of all vertebrate *myc* genes. Only one product was obtained and it was identical to the corresponding sequence in Amphi-*myc*. These results indicate that Amphi-*myc* is the sole member of the *myc* family present in the genome of amphioxus. Moreover, the genome of the sea squirt *Ciona intestinalis*, which is the closest known animal to amphioxus, has been fully sequenced and contains only a single *myc*-like sequence (Ensembl, ENSCING00000005036).

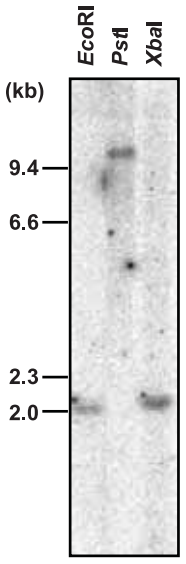

Fig. 2. Southern blot analysis of amphioxus genomic DNA digested with *Eco*RI, *Pst*I or *Xba*I. The blot was probed at low stringency with a 342 bp stretch of Amphi-*myc*.

iii. Phylogenetic analysis of Myc family

In a phylogenetic tree based on Myc amino acid sequences, Amphi-Myc was located outside the vertebrate Myc family (Fig. 3). The topology of the vertebrate Mycs in this tree is similar to that reported by Atchley and Fitch[3], in which c-Myc, N-Myc and L-Myc were grouped separately, and s-Myc clustered with N-Myc. Clearly, the *myc* family genes in vertebrates have evolutionarily originated from a single copy of the gene in an invertebrate sister group like Amphi-*myc*.

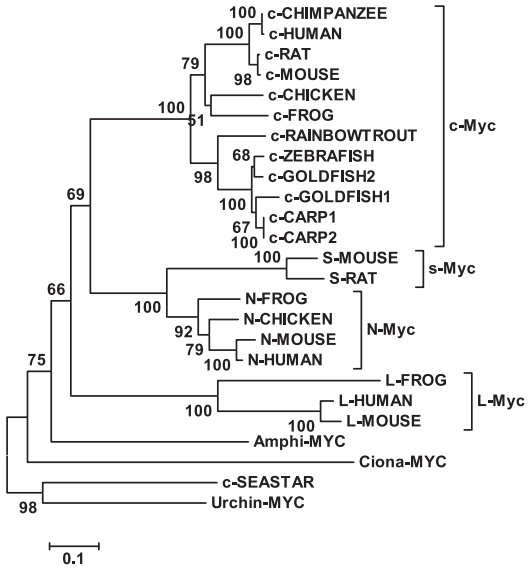

Fig. 3. A neighbor-joining tree of major members of the Myc family based on amino acid sequences. Sea star and sea urchin Myc were used as an outgroup. The scale bar is calibrated in amino acid replacements per site. Bootstrap confidence values are shown. Accession numbers are as follows: c-SEASTAR, M80364; Urchin-MYC, L37056; Ciona-MYC, ENSCING00000005036 (Ensembl); Amphi-MYC, AB175932; S-MOUSE, NM_010850; S-RAT, M29069; N-FROG, X58670; N-CHICKEN, D90071; N-MOUSE, X03919; N-HUMAN, M13241; L-FROG, L11363; L-MOUSE, X13945; L-HUMAN, M19720; c-RAINBOWTROUT, M13048; c-ZEBRAFISH, L11710; c-GOLDFISH2, D31729; c-GOLDFISH1, AB040746; c-CARP2, D37888; c-CARP1, D37887; c-FROG, M14455; c-CHICKEN, J00889; c-MOUSE, X01023; c-RAT, Y00396; c-CHIMPANZEE, M38057; c-HUMAN, J00120. Since *Drosophila* Myc makes it difficult to notice areas of high similarity, we deleted that sequence from consideration.

Discussion

We showed here that the amphioxus genome contained only a single *myc* gene. This suggests that *myc* gene duplication occurred after protochordate on phylogeny. In vertebrates, a genome duplication event that produced two gene lineages is believed to have occurred in the Cambrian period[17,18]. Duplication of genes might have been the main mechanism for the functional diversification of genes, the creation of gene families, and the increase in genomic and phenotypic complexity[19]. For example, the appearance of N-*myc* may have had a role in neurogenesis in vertebrates, because N-*myc* mRNA has been observed in the central nervous system of vertebrates[20,21].

In this study, we could not estimate the divergence time between Amphi-Myc and the vertebrate Myc family, because the evolutionary rates are not constant in different species. Therefore, it is unclear whether the *myc* family was formed by genome duplications in the Cambrian period. It will be of particular interest and significance to examine whether the genomes of jawless hagfish and lampreys have two copies of the *myc* genes to test the 2R hypothesis, i.e., the hypothesis that two rounds of genome duplication occurred in vertebrate evolution[17].

References

1) Oster, S.K., Ho, C.S., Soucie, E.L., Penn, L.Z.: The *myc* oncogene: MarvelouslY Complex. Adv. Cancer Res. **84**, 81-154 (2002).
2) Ryan, K.M., Birnie, G.D.: Myc oncogenes: the enigmatic family. Biochem. J. **314**, 713-721 (1996).
3) Atchley, W.R., Fitch, W.M.: A natural classification of the basic helix-loop-helix class of transcription factors. Proc. Natl. Acad. Sci. USA **94**, 5172-5176 (1997).
4) Saito, H., Hayday, A.C., Wiman, K., Hayward, W.S., Tonegawa, S.: Activation of the c-*myc* gene by translocation: a model for translational control. Proc. Natl. Acad. Sci. USA **80**, 7476-7480 (1983).
5) Walker, C.W., Boom, J.D., Marsh, A.G.: First non-vertebrate member of the myc gene family is seasonally expressed in an invertebrate testis. Oncogene **7**, 2007-2012 (1992).
6) Gallant, P., Shiio,Y., Cheng, P.F., Parkhurst, S.M., Eisenman, R.N.: Myc and Max homologs in Drosophila. Science **274**, 1523-1527 (1996).
7) Wada, H., Satoh, N.: Details of the evolutionary history from invertebrates to vertebrates, as deduced from the sequences of 18S rDNA. Proc. Natl. Acad. Sci. USA **91**, 1801-1804 (1994).
8) Bernard, O., Cory, S., Gerondakis, S., Webb, E., Lipman, J.M.: Sequence of the murine and human cellular *myc* oncogenes and two modes of *myc* transcription resulting from chromosome translocation in B lymphoid tumors. EMBO J. **2**, 2375-2383 (1983).
9) Watson, D.K., Reddy, E.P., Duesberg, P.H., Papas, T.S.: Nucleotide sequence analysis of the chiken c-*myc* gene reveals homologous and unique coding regions by comparison with the transforming gene of avian myelocytimatosis virus MC29, Dgag-*myc*. Proc. Natl. Acad. Sci. USA **80**, 2146-2150 (1983).
10) King, M.W., Roberts, J.M., Eisenman, R.N.: Expression of the c-*myc* proto-oncogene during development of *Xenopus laevis*. Mol. Cell Biol. **6**, 4499-4508 (1986).
11) Van Beneden, R.J., Watson, D.K., Chen, T.T., Lautenberger, J.A., Papas, T.S.: Cellular *myc* (c-*myc*) in fish (rainbow trout): its relationship to other vertebrate *myc* genes and to the transforming genes of the MC29 of viruses. Proc. Natl. Acad. Sci. USA **83**, 3698-3702 (1986).
12) Zhang, H., Okamoto, N., Ikeda, Y.: Two c-*myc* genes from a tetraploid fish, the common carp (*Cyprinus carpio*). Gene **153**, 231-236 (1995).
13) Futami, K., Komiya, T., Zhang, H., Okamoto, N.: Differential expression of *max* and two types of c-*myc* genes in a tetraploid fish, the common carp (*Cyprinus carpio*). Gene **269**, 113-119 (2001).
14) Meinkoth, J., Wahl, G.: Hybridization of nucleic acids immobilized on solid supports. Anal. Biochem. **138**, 267-284 (1984).
15) Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G.: The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic. Acids Res. **24**, 4876-4882 (1997).
16) Saitou, N., Nei, M.: The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**, 406-425 (1987).
17) Holland, P.W., Garcia-Fernandez, J., Williams, N.A., Sidow, A.: Gene duplications and the origins of vertebrate development. Dev. Suppl., 125-133 (1994).
18) Sidow, A.: Genome duplications in the evolution of early vertebrates. Curr. Opin. Genet. Dev. **6**, 715-722 (1996).
19) Ohno S.: Evolution by gene duplication. Springer-Verlag Press, Heidelberg, New York (1970).
20) Grady, E.F., Schwab, M., Rosenau, W.: Expression of N-*myc* and c-*src* during the development of fetal human brain. Cancer Res. **47**, 2931-2936 (1987).
21) Vize, P.D., Vaughan, A., Krieg, P.: Expression of the N-*myc* proto-oncogene during the early development of *Xenopus laevis*. Development **110**, 885-896 (1990).

ナメクジウオ *myc* 遺伝子の進化的解析

森田　幹晴[*1]・二見　邦彦[*1]・張　寰[*2]・窪川　かおる[*3]・小島　吉雄[*4]・岡本　信明[*1]

[*1] 東京海洋大学海洋科学部海洋生物資源学科
[*2] コネチカット大学
[*3] 東京大学海洋研究所
[*4] 日本魚類生物科学研究所

要旨： がん原遺伝子 *myc* は細胞増殖を制御する重要な遺伝子の一つである。脊椎動物では 4 種類の *myc* が存在しているが，その起源や類縁関係は明らかとなっていない。本研究で我々は，脊椎動物に最も近縁の無脊椎動物であるナメクジウオ *Branchiostoma belcheri* から *myc* 遺伝子を単離した。全長 1480 塩基対の cDNA は 371 アミノ酸からなる ORF を含んでいた。また，エキソン／イントロン構造は保存されていた。サザンブロット法および degenerate PCR の結果, ナメクジウオのゲノムは単一の *myc* 遺伝子を持つことが明らかとなった。また，推定アミノ酸配列に基づいた系統解析の結果，ナメクジウオ Myc は脊椎動物 Myc ファミリーの外側に位置していた。これらの結果から，*myc* ファミリーは原索動物以降に形成されたものと考えられた。

キーワード： 原索動物，進化，遺伝子重複，Myc ファミリー，脊椎動物