Cloning and Sequencing of a *c-myc* Gene from Goldfish *Carassius auratus*

Huan Zhang, Hang Zhang, Nobuaki Okamoto,[†] and Yayoi Ikeda

Department of Aquatic Biosciences, Tokyo University of Fisheries, Konan, Minato, Tokyo 108, Japan

(Received June 3, 1994)

A proto-oncogene, cellular *myc* (*c-myc*) gene, has been isolated from a tetraploid fish, goldfish *Carassius auratus*. The nucleotide sequence of 5021 base pairs covering the gene and its flanking regions was determined. The goldfish *c-myc* gene, probably like those of other animals, consists of two coding exons (exon 2 and exon 3) interrupted by an intron. Exon 2 and 3 together presumably encode a protein of 399 amino acid (aa) residues. The goldfish *c-myc* gene shares higher aa identities with those of the common carp (93.0% and 95.3%) and zebra fish (90.7%) than with those of rainbow trout (71.9%) or humans (56.0%).

Key words: goldfish, tetraploidy, *c-myc* gene, cloning, nucleotide sequence, genomic organization

Polyploidy has recently been recognized as a potentially important process in the evolution of vertebrates.^{1,2} It is believed that a tetraploid event took place about 500 million years (Myr) ago in a common ancestor of all vertebrates.³ Higher vertebrates have evolved sex determination based on heteromorphic sex chromosomes, which prevent successful polyploidization.¹ Polyploidy is still common in teleosts, which seem to be sexually undifferentiated.⁴ Studies on gene duplication in tetraploid teleosts are important to understand the evolutionary processes that followed the tetraploid event.⁵

In this study, an investigation of the cellular *myc* (*c-myc*) gene in a tetraploid fish, goldfish *Carassius auratus*, was selected. The *c-myc* gene, a proto-oncogene homologous to the oncogene (*v-myc*) carried by avian retroviruses of the MC29 family, is well conserved during vertebrate evolution.^{6,7} The *c-myc* gene is thought to be one of the most important genes in controlling cell propagation. Alteration of normal *c-myc* expression by chromosomal translocation, retroviral insertion or DNA amplification is believed to play an important role in the progressive development of a variety of malignancies.^{8,9} In mammalian genomes, there are a number of *myc* genes forming the *myc* gene family. Each member is structurally similar to one another, however, the nucleotide (nt) and amino acid (aa) identities shared by them are less than 50%. The evolutionary origins and relationships of each *myc* member remain unknown. Studies of tetraploid fish *myc* duplicates should help us to understand the evolution of mammalian *myc* genes.

Two *c-myc* genes have already been isolated from a tetraploid fish, the common carp *Cyprinus carpio* (Zhang *et al.*, in preparation). The isolation of a *c-myc* gene from another tetraploid cyprinid species, the goldfish, is reported here.

Materials and Methods

Genomic DNA extracted from the hepatopancreas of goldfish was digested completely with restriction enzyme *EcoRI* or *BamHI*. Southern

blot analysis was carried out as described by Zhang *et al.*¹⁰ A carp *c-myc* cDNA clone, C1, 1334 base pairs (bp), previously isolated from a cDNA library of carp peripheral blood leukocytes,¹⁰ was used as a probe. DNA fragments used for the partial genomic library construction (corresponding to the 8 kb *EcoRI* band) were eluted from the gel by dialysis tubing¹¹ and inserted into the *EcoRI* site of the cloning vector λ EMBL4. A library with a titer of 10⁶ plaque forming unit was obtained and screened with the C1 probe. Positive clones were further analyzed by digestion of restriction enzymes (*BamHI*, *EcoRI*, *KpnI*, *MboI*, or *PstI*) and subcloned into plasmid pUC18. The nt sequences were determined by using a T7Sequence™ Kit (Pharmacia). "GENETYXMAC," a computer algorithm developed by Software Development Co., was used for sequence comparison.

Results and Discussion

Two bands were observed by Southern blot analysis, in *EcoRI* digests (16 kilo base pairs (kb) and 8 kb) and *BamHI* digests (8 kb and 5 kb), respectively (Fig. 1). These bands are thought to correspond to two *c-myc* genes existing per haploidy genome in goldfish, just like those in the common carp (Zhang *et al.*, in preparation), because both goldfish and carp are tetraploidy. In this study, DNA fragments corresponding to the 8 kb *EcoRI* band were used to construct a partial genomic library (*BamHI* fragments were not used because *BamHI* sites which exist in the 5' region of carp *c-myc* genes may also exist in those of goldfish). Three positive clones were obtained from this library. All of them showed the same restriction enzyme pattern (8 kb, Fig. 2B), indicating that they belong to the same clone, called GM2 (Fig. 2A).

Partial nt sequences of GM2, 5021 bp in length, were determined. Three thousand bp of them, covering the goldfish *c-myc* gene and its flanking regions, and the deduced aa sequences (399 aa) are shown in Fig. 3. The putative initiation codon ATG was determined by comparing the nt sequences and deduced aa sequences of the goldfish *c-myc* gene with those of carp,¹⁰ zebra fish,¹² rainbow trout,¹³ and humans,¹⁴ and the first letter, A, of ATG was numbered as +1.

The genomic organization of the goldfish *c-myc* gene is

[†] Author to whom correspondence should be addressed.

shown in Fig. 2A. The goldfish *c-myc* gene consists of two coding exons (exon 2 and exon 3) interrupted by an intron, just like those of other animals.^{10,12-14} The coding regions of exon 2 and exon 3 are 619 bp and 581 bp, respectively. The intron (406 bp) begins with GT and ends with AG, following the GT-AG rule.¹⁵ The consensus sequence for poly A addition, AATAAA, is found in two locations (+1829 and +1904).

Comparisons of the goldfish *c-myc* gene with those of carp, zebra fish, rainbow trout, and humans are shown in Table 1. In the coding region, the goldfish *c-myc* gene shares high nt identities with two *c-myc* genes of carp (90.5 and 93.0%) and relatively low identity with those of rainbow trout (71.0%) and humans (52.5%). In the intron, the goldfish *c-myc* gene also shares higher nt identities with the two carp *c-myc* genes than those of rainbow trout and humans. The aa alignments of *c-myc* genes of goldfish, carp, zebra fish, rainbow trout, and humans are shown in Fig. 4. The aa identities between goldfish and carp (93.0 and

95.3%), and goldfish and zebra fish (90.7%) are higher than those between goldfish and rainbow trout (71.9%) and goldfish and humans (56.0%) (Table 1). These results agree with morphological taxonomy in which the goldfish, common carp, and zebra fish belong to the same family, Cyprinidae, while rainbow trout belongs to a different order, Salmoniformes and humans to another class, mammalia. Boxed regions in Fig. 4 are thought to be very important for the function of *c-myc* genes (for details, see Taya,⁸ Schreiber-Agus *et al.*,¹²) and Ohmori¹⁶), and there are few aa changes among these vertebrates.

It is known that in more than 1600 species of Cyprinidae, only eight species have been identified as tetraploidy.¹⁷ Most of the studies on cyprinid tetraploids are on allozymes.¹⁷ From our results, we believe comparative studies of *c-myc* gene on different cyprinid tetraploids may significantly assist an understanding of the evolution of the *myc* gene family in cyprinid fishes as well as in other vertebrates.

Acknowledgements This study was supported in part by a Grants-in-Aid for Scientific Research (Nos. 05404014, 05044136) from the Ministry of Education, Science, and Culture of Japan and the Japanese Fisheries Agency.

Fig. 1. Southern blot analysis.

Genomic DNA of goldfish (15 µg) was digested completely with *EcoRI* or *BamHI*. After electrophoresis on a 0.8% agarose gel, DNA fragments were transferred onto a Hybond-N⁺ nylon membrane (Amersham), and hybridized with the [α -³²P]-dCTP labeled probe C1 at 65°C for 12 h in 5 × SSPE, 5 × Denhardt's reagent and 0.5% SDS. The membrane was washed in 0.2 × SSC and 0.2% SDS at 65°C, and exposed to a Kodak X-ray film. Molecular size markers are indicated on the left with short bars. Estimated sizes of hybridizing DNA fragments are indicated on the right with arrows.

Fig. 2. Genomic organization of the goldfish *c-myc* gene.

A shows the structure of the gene predicted by sequencing results of the isolated genomic clone GM2 and by comparing GM2 with the *c-myc* genes of the common carp, zebra fish, rainbow trout, and humans. Exons are indicated by boxes. The putative coding regions are shaded and the untranslated region is unshaded. The intron is indicated by double lines. Sequenced 5' and 3' flanking regions are indicated by single lines. B shows the restriction enzyme map. Only the enzymes used for subcloning are indicated. B, *BamHI*; E, *EcoRI*; K, *KpnI*; M, *MboI*; P, *PstI*.

Table 1. Comparison of the goldfish *c-myc* gene with those of other animals

Region	Size* ¹						Identity (%)				
	GM2	CM1	CM2	ZM	RM	HM	GM2/CM1	GM2/CM2	GM2/ZM	GM2/RM	GM2/HM
CR* ²											
(aa)	399 aa	394 aa	401 aa	406 aa	417 aa	439 aa	93.0	95.3	90.7	71.9	56.0
(nt)	1197 bp	1182 bp	1203 bp		1251 bp	1317 bp	90.5	93.0		71.0	52.5
Intron	406 bp	423 bp	420 bp		322 bp	1376 bp	78.3	84.5		50.3	<50
3'UTR* ³	231 bp	231 bp	228 bp			303 bp	89.7	95.2			<50
5'FR* ⁴	828 bp	787 bp	807 bp				74.0	80.0			

*¹ Abbreviations: GM2, goldfish *c-myc* gene 2; CM1, carp *c-myc* gene 1; CM2, carp *c-myc* gene 2; ZM, zebra fish *c-myc* gene; RM, rainbow trout *c-myc* gene; HM, human *c-myc* gene.

*² CR stands for coding region. (aa) and (nt) indicate amino acid and nucleotide residues, respectively.

*³ 3'UTR stands for 3' untranslated region (from the T of stop codon TAA to the last A of AATAAA).

*⁴ 5'FR stands for 5' flanking region (from the G of 5' *EcoRI* site GAATTC to the nt residue immediately upstream of the start codon ATG).

-828 GAATTCAG CAATACATTAACATGTCTT

-800 CTACTAATAGTAAAGACATT ACTAATTACTAAATACTAAT ATTGCTTTCCAATGCGTCTT TGTTTTTTGTGTTTTATTTCG GATGTGTGAACCTTGCGGAG

-700 GGATTTAATTCGCAACAAC CGTGCTTGTGTCGGGAAA CGTCGGTGCGCAACGTAATA CGTCAGTGATTGCTCCTCG CGGGAGCTGCAGAAATTATA

-600 CCTCCTGCGCTGTACACT CACAGACTAGTCTGCGCCAGC ATCCGAGACGTACACAGGAG CGCATCGTTGATTACATTTT CCATCACTTTCCATTGACC

-500 AGGTCTCTGATTCCACTTAT GCTGCTAGAGACTGGAGTAA ACTGTTGGATCTTACTTTTT GCTGCTGGTTTGTGTTGGTC CTCCTTTCACTGCATCGCGT

-400 CTAGAGAGAAATCATTCTGA AGCTGCATTCGTTAATTTAC ACGGTAAGGATCCGGCATTG GTTATTGCATGCATCTTTTA GACTTAAAGTGCAACTTGCG

-300 TGTTATGTGTGCAAAATGCTT TTATATTTCTAACAGGAATT TGAATGTTGCACGCTCAAAG ATTGCACAGAAAGCAAGACT GCTTTTTAATGTTTGTATAG

-200 TAAACCCCGCTTTGGGTTTA ATTGTTACACAGGATTAAGT TCCATTATTAGCAATTTTT TTCCACCCCATGCATGTTAT GCATAGTGTAATATGTGCA

-100 TTAGTAAAGCGTTTTTGTG TGATGTGTAGTGTAGCGTTG CATATGCATGTAAACGTTAG TAATTCCAATGTGTGTGTTT TTTTTCCTCTAGTGTCAA

+1 **ATG** CCG GTG AGT GCG AGT TTG GCG TAT AAA AAC TAC GAT TAC GAC TAC GAC TCA ATC CAG CCC TAT TTT TAC TTC GAC

+1 Met Pro Val Ser Ala Ser Leu Ala Tyr Lys Asn Tyr Asp Tyr Asp Tyr Asp Ser Ile Gln Pro Tyr Phe Tyr Phe Asp

+79 AAC GAC GAT GAG GAT TTC TAT CAC CAC CAG CAA GGA CAG CCT CAG CCT CCA GCT CCC AGC GAG GAC ATT TGG AAA AAA

+27 Asn Asp Asp Glu Asp Phe Tyr His His Gln Gln Gly Gln Pro Gln Pro Pro Ala Pro Ser Glu Asp Ile Trp Lys Lys

+157 TTC GAG CTG CTG CCC ACG CCG CCC CTC TCT CCC AGC CGG AGA CAG TCG CTC TCC ACC GCC GAA CAG CTG GAG ATG GTC

+53 Phe Glu Leu Leu Pro Thr Pro Leu Ser Pro Ser Arg Arg Gln Ser Leu Ser Thr Ala Glu Gln Leu Glu Met Val

+235 AGC GAG TTC CTG GGA GAC GAC GTG GTC AAC CAG AGC TTC ATC TGC GAC GCG GAC TAC TCC CAG TCA TTC ATC AAG TCC

+79 Ser Glu Phe Leu Gly Asp Asp Val Val Asn Gln Ser Phe Ile Cys Asp Ala Asp Tyr Ser Gln Ser Phe Ile Lys Ser

+313 ATC ATC ATT CAG GAC TGC ATG TGG AGC GGC TTT TCT GCC GCC GCC AAG TTA GAG AAA GCC GTT TCT GAG AGG CTG GCG

+105 Ile Ile Ile Gln Asp Cys Met Trp Ser Gly Phe Ser Ala Ala Lys Leu Glu Lys Ala Val Ser Glu Arg Leu Ala

+391 TCC CTG CAC GCT GCA AGG AAG GAA CTG ATA TCT GAC AGC AGC TCA AAT CGA CTC AGT GCA AGT TAT TTG CAG GAT CTG

+131 Ser Leu His Ala Ala Arg Lys Glu Leu Ile Ser Asp Ser Ser Ser Asn Arg Leu Ser Ala Ser Tyr Leu Gln Asp Leu

+469 AGC ACC TCC GCA TCA GAA TGC ATT GAC CCC TCT GTG GTC TTT CCG TAT CCC CTC ACA GAG TCC AGC AAA TCT AAC AAG

+157 Ser Thr Ser Ala Ser Glu Cys Ile Asp Pro Ser Val Val Phe Pro Tyr Pro Leu Thr Glu Ser Ser Lys Ser Asn Lys

+547 GTT GCT CCA TCA CAA CCC ATG CTT GTC CTG GAC ACT CCA CCT AAC AGC TCC AGC AGC AGT GGC AGC GAT TCT Gttagct

+183 Val Ala pro Ser Gln Pro Met Leu Val Leu Asp Thr Pro Pro Asn Ser Ser Ser Ser Ser Ser Ser G

+626 tatgcatattttttgtcat ttctgacatgcagaatcttg atttagctggttactgctg cgcaaatgcttgattttct aagtgcctttacatgtcaa

+726 cggtaggtggtgtgcccgt caaggtgctttatctctcata cccctgtagcaatgattggt gtaagtgtggttgataat atcacctcccctttgcttag

+826 acccgagtcagatctgtaga aagactgaaatgtgcaata tcagctgacgagtaagtgg gcgggaactgaatgcagaga ttccccctacctcagccttt

+926 gggcaaaacaaattgcaaaa ttctgttcaaatcccatcg cctgttgaatgaactgttt ttttttaagttattaattta tttttgtaatttaattgtag

+1026 AA GAT GAA GAA GAG GAG GAA GAG GAG GAG GAA GAA GAA GAG GAA GAA GAG GAG GAA GAA GAA GAA ATT

+207 lu Asp Glu Ile

+1103 GAT GTG GTG ACT GTG GAA AAG CGT CAA AAA AGG AAT GAG GCA GAT GTG TCG GAT TCA AGG TAC CCC AGT CCC CTG GTG

+233 Asp Val Val Thr Val Glu Lys Arg Gln Lys Arg Asn Glu Ala Asp Val Ser Asp Ser Arg Tyr Pro Ser Pro Leu Val

+1181 TTG AAG CGC TGT CAC GTC TCC ACC CAC CAG CAC AAC TAC GCC GCC CAT CCC TCC ACG CGG CAT GAC CAG CCG GCT GTC

+259 Leu Lys Arg Cys His Val Ser Thr His Gln His Asn Tyr Ala Ala His Pro Ser Thr Arg His Asp Gln Pro Ala Val

+1259 AAG AGG CTA CGA CTG GAG ACC AGC AGC AGC AAC AGG CAT GGG AAG CAA CGA AAA TGT ACG AGT CCC CGG ACA TCG GAT

+285 Lys Arg Leu Arg Leu Glu Thr Ser Ser Ser Asn Arg His Gly Lys Gln Arg Lys Cys Thr Ser Pro Arg Thr Ser Asp

+1337 TCT GAG GAC AAT GAC AAA CGC AGG ACT CAC AAT GTG CTG GAG CGC CAA CGT AGA AAT GAA CTC AAA CTC AGC TTT TTT

+311 Ser Glu Asp Asn Asp Lys Arg Arg Thr His Asn Val Leu Glu Arg Gln Arg Arg Asn Glu Leu Lys Leu Ser Phe Phe

+1415 GCA TTA CGG GAC GAG ATC CCT GAG GTT GCA AAT AAC GAG AAA GCT GCT AAA GTT GTT ATT CTA AAG AAG GCG ACA GAG

+337 Ala Leu Arg Asp Glu Ile Pro Glu Val Ala Asn Asn Glu Lys Ala Ala Lys Val Val Ile Leu Lys Lys Ala Thr Glu

+1493 TGC ATT CAC AGC ATG CAG TTG GAT GAG CAA AGG TTG CTG TCC ATC AAA GAG CAG CTG AGG CGA AAG AGC GAA CAG TTA

+363 Cys Ile His Ser Met Gln Leu Asp Glu Gln Arg Leu Leu Ser Ile Lys Glu Gln Leu Arg Arg Lys Ser Glu Gln Leu

+1571 AAA CAC AGG CTG CAG CAA CTG CGG AGT TCA CAT **TAA** TGGACCTGTGTAC AAGAGTGAAAGTGAAGAAC TTTTAAAGAACCCGCTCCCC

+389 Lys His Arg Leu Gln Gln Leu Arg Ser Ser His ***

+1661 CTGATTTTTTTATGCAAGTT AAAGTGTGCTATGTATAGT TCAAGTGCATGTCCAATGC ATTTTTTTTCTTTTGAATA TTGTCAGTTCTTTTCATGTT

+1761 TAAAGTGCAGAGTTTAAAT ATATAAGGTGGTTAAATCC ATTTGTATTTAAGTCATTTA TTTTATTAATAAATGAAA AAAAAAATAGGGCTGCAC

+1861 AACCCATAAGTAAAAATTGT TTGTATGTTTTTGTATATA ATGAATAAATATTTCCTAAG AAATGATTTTTGGATCTGT ATTAGCTACTGTTCTCATA

+1961 AACCATATTCTGTTTCGCTA TGTTTTTTTTTCTTCATGTT TTGACAATATAAATGAATTT CTTTTGAATTGACTCTGG TCTGCTTAGTCTCTTTGCA

+2061 CTGAACCCCTTATACACATA GATCAGACACTTTTGCAAGA AGTTTGGGAATATTTCAGAG AAAGTTGATATTGCTTTCT TTATTACATATGATTGTTA

+2161 CTTTGATAAACT

Fig. 3. Nucleotide (nt) and deduced amino acid (aa) sequences of the goldfish *c-myc* gene.

The nt sequences in the intron are indicated by lowercase letters. Other sequences are indicated by uppercase letters. Deduced aa is shown as three italic letters below each codon and italic numbers. The putative initiation codon ATG, stop codon TAA and the consensus sequence for poly A addition AATAAA are indicated by bold letters. The nt sequence data of the goldfish *c-myc* gene reported in this paper will appear in the GSDB, DDBJ, EMBL, and NCBI nucleotide sequence databases with the accession number of D31729.

Fig. 4. Comparison of deduced amino acid (aa) sequences of *c-myc* genes of the goldfish (GM2), common carp (CM1 and CM2), zebra fish (ZM), rainbow trout (RM), and human (HM).

The amino acids are abbreviated by the standard one-letter code. Gaps (—) are introduced to optimize identity. Dots (·) represent identical aa with the goldfish *c-myc* gene and the asterisks (*) represent the end of exon 3. The arrow indicates the boundary of exon 2 and 3. The five functionally important regions are boxed.

References

- 1) S. Ohno: Evolution by Gene Duplication, Springer Verlag Press, Heidelberg, New York, 1970, pp. 1-160.
- 2) L. G. Lundin: Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics*, **16**, 1-19 (1993).
- 3) F. W. Allendorf and G. H. Thorgaard: Tetraploidy and the evolution of salmonid fishes, in "Evolutionary Genetics of Fishes" (ed. by B. J. Turner), Plenum Press, New York and London, 1984, pp. 1-53.
- 4) Y. Ojima: Fish Cytogenetics, Suikosha Press, Tokyo, 1983, pp. 50-64 (in Japanese).
- 5) S. Ohno: Pattern in genome evolution. *Current Opinion in Genetics and Development*, **3**, 911-914 (1993).
- 6) D. K. Watson, E. P. Reddy, P. H. Duesberg, and T. S. Papas: Nucleotide sequence analysis of the chicken *c-myc* gene reveals homologous and unique coding regions by comparison with the transforming gene of avian myelocytomatosis virus MC29, *Agag-myc*. *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 2146-2150 (1983).
- 7) K. Hayashi: The *c-myc* gene. *Protein, Nucl. Acid and Enzyme*, **37**, 860-861 (1992) (in Japanese).
- 8) Y. Taya: The *myc* genes and signal transduction. *Exper. Med.*, **9**, 169-172 (1991) (in Japanese).
- 9) E. LeGouy, R. DePinho, K. Zimmerman, P. Ferrier, R. Collum, and F. W. Alt: Structure and Expression of *myc*-family Genes, in "Nuclear Oncogenes" (ed. by F. W. Alt, E. Harlow, and E. B. Ziff), Cold Spring Harbor Laboratory, New York, 1987, pp. 144-151.
- 10) H. Zhang, N. Okamoto, N. Yamamoto, and Y. Ikeda: Molecular Cloning of Carp Cellular *myc* (*c-myc*) cDNA. *Gyobyo kenkyu (Fish Pathol.)*, **28**, 111-117 (1993).
- 11) J. Sambrook, E. F. Fritsch, and T. Maniatis: Molecular Cloning, 2nd ed., Cold Spring Harbor Laboratory Press, New York, 1989, pp. 6.28-6.29.
- 12) N. Schreiber-Agus, J. Horner, R. Torres, F. C. Chiu, and R. A. Depinho: Zebra Fish *myc* Family and *max* Genes: Differential Expression and Oncogenic Activity throughout Vertebrate Evolution. *Mol. Cell. Biol.*, **13**, 2765-2775 (1993).

- 13) R. J. Van Beneden, D. K. Watson, T. T. Chen, J. A. Lautenberger, and T. S. Papas: 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. U.S.A.*, **83**, 3698–3702 (1986).
- 14) J. Battey, C. Moulding, R. Taub, W. Murphy, T. Stewart, H. Potter, G. Lenoir, and P. Leder: The human *c-myc* oncogene: structural consequences of translocation into the IgH locus in Burkitt lymphoma. *Cell*, **34**, 779–787 (1983).
- 15) R. Breathnach, C. Benoist, K. O'Hare, F. Gannon, and P. Chambon: Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. U.S.A.*, **75**, 4853–4857 (1978).
- 16) Y. Ohmori: The *myc* genes and the determination of cellular differentiation. *Exper. Med.*, **10**, 19–25 (1992) (in Japanese).
- 17) D. G. Bush: Allozymes of the cyprinid fishes, in "Evolutionary Genetics of Fishes" (ed. by B. J. Turner), Plenum Press, New York and London, 1984, pp. 561–590.