

大口黑鲈 *Myf5* 基因 cDNA 和基因组序列的克隆与分析

郭玉函^{1,2}, 白俊杰¹, 劳海华¹, 叶星¹, 罗建仁¹

(1. 中国水产科学研究院 珠江水产研究所, 中国水产科学研究院 热带亚热带鱼类选育与养殖重点开放实验室, 广东 广州 510380; 2. 上海海洋大学 生命科学与技术学院, 上海 200090)

摘要: *Myf5* 是生肌调节因子家族成员之一。为研究该基因在大口黑鲈 (*Micropterus salmoide*) 肌肉生长发育中的作用, 运用 RT-PCR 和 RACE 技术获得大口黑鲈 *Myf5* cDNA 序列 1 093 bp, 其中开放阅读框长 723 bp, 编码 240 个氨基酸, 经分析该蛋白无信号肽序列, 属核蛋白, 含有一典型的碱性螺旋—环—螺旋结构域。使用 Genomewalker 技术获得启动子区序列 2 690 bp, 预测含有肌肉特异性转录调控相关元件: E- 框、肌细胞特异增强子 2 (MEF2)、核转录因子 1 (SP1)、上游激活因子 (USF)、血清应答因子 (SRF) 等结合位点。含有早期生长应答因子 α (EGR α)、早期快反应生长应答因子 1, 2 (EGR-1, 2) 等结合位点, 它们可能是 *Myf5* 能够在体节形成早期表达的主要原因。获得 *Myf5* 3 个外显子与 2 个内含子序列, 内含子 1 中存在一个在不同鱼类中高度保守的长度为 123 bp 的序列, 推测参与调节了 *Myf5* 基因的时空和组织特异性表达。本研究旨在为进一步研究该基因对大口黑鲈肌肉生长发育的作用机理奠定基础。[中国水产科学, 2008, 15 (4): 568–576]

关键词: 肌肉; *Myf5* 基因; 调节; 大口黑鲈

中图分类号: Q959

文献标识码: A

文章编号: 1005-8737-(2008)04-0568-09

生肌调节因子家族 (Myogenic regulatory factors, MRFs) 基因的表达决定了骨骼肌细胞的命运。该家族属转录因子家族, 包括 *Myf5*、*MyoD*、*myogenin* 和 *MRF4* 4 个成员, 均存在碱性螺旋—环—螺旋结构域^[1]。通过该结构域它们能够与 E 蛋白形成二聚体, 再与肌肉特异性基因如肌球蛋白轻链和肌酸激酶等基因的上游调控序列 E-box 结合, 从而激活这些基因的表达^[2-4]。

Myf5 是该家族最早表达的基因, 在成肌细胞的特化和增殖过程中发挥了关键的调节作用。人类 *Myf5* 基因最早发现于 1989 年, 能够诱导鼠的 10T1/2C3H 成纤维细胞分化为肌纤维^[5]。*Myf5* 基因敲除实验中小鼠轴上肌肉系统发育迟缓并死于肋骨缺陷症, 说明 *Myf5* 基因起始调控了轴上肌肉的形成^[6-9]。在胚胎时期阻碍斑马鱼 (*Danio rerio*) *Myf5* 基因的表达, 体节形成与肌肉发育过程均表现出异常, 说明鱼类的 *Myf5* 基因与哺乳类有着相似的功能^[10]。目前, *Myf5* 基因在斑马鱼^[11]、条纹狼鲈 (*Morone saxatilis*)^[12]、褐牙鲈 (*Paralichthys olivaceus*)^[13]、花鲈 (*Lateolabrax japonicus*)^[14] 胚胎

时期的表达模式已经建立。在虹鳟 (*Oncorhynchus mykiss*) 的胚后生长发育中, *Myf5* 基因的表达水平与肌肉的阶段性生活生长明显的相关性^[15]。不同于哺乳动物, 大多数鱼类肌肉生长发育往往持续一生^[16], 而其肌肉生长发育的状况决定了鱼体的最终大小^[17], 深入研究鱼类肌肉生长发育的机制有利于提高鱼类的养殖效率。目前, 已有鲤 (*Cyprinus carpio*)、虹鳟、大西洋鲑 (*Salmo salar*) 和青斑河鲀 (*Tetraodon nigroviridis*) 的 *Myf5* cDNA 序列及斑马鱼、条纹狼鲈、花鲈、褐牙鲈和红鳍东方鲀 (*Takifugu rubripes*) 的 *Myf5* 基因组序列被克隆。

大口黑鲈, 隶属鲈形目 (Perciformes)、鲈亚目 (Percoidae)、太阳鱼科 (Cettrachidae)、黑鲈属 (*Micropterus*)。原产于北美密西西比河流域, 自 20 世纪 80 年代引入中国后已成为中国重要的淡水养殖优良品种之一。本研究选择大口黑鲈作为实验材料, 首先克隆了其 *Myf5* cDNA, 在此基础上得到其基因组序列, 并采用生物信息学方法对其启动子上的顺式作用元件进行了分析, 为进一步研究其对肌肉生长发育的作用机理奠定基础。

收稿日期: 2007-12-25; 修订日期: 2008-01-29.

基金项目: 国家科技支撑项目 (2005DKA21103); 国家科技基础条件平台工作项目 (2005DKA21103).

作者简介: 郭玉函 (1980-), 女, 硕士研究生, 主要从事水产生物技术研究.

通讯作者: 白俊杰. Tel: 020-81616129; E-mail: jibai@163.net

1 材料与方法

1.1 材料

实验用大口黑鲈成鱼体质量约 400 g, 取自广东省大口黑鲈良种场; SV Total RNA Isolation kit 购自 Promega 公司; RNA PCR Kit (AMV) Ver.3.0 试剂盒和 pMD19-T vectors system 购自 TaKaRa 公司; Blood & Cell Culture DNA Kit 和 GenomeWalker Universal Kit 均购自 Clontech 公司; 大肠杆菌 DH5 α 由本实验室保存。

1.2 方法

1.2.1 肌肉总 RNA 的提取 取大口黑鲈背部肌肉 30 mg, 按照 Promega 公司 SV Total RNA Isolation System 试剂盒的方法进行。

1.2.2 基因组 DNA 的提取 取大口黑鲈尾部静脉血, 按体积比 1:6 与抗凝剂混匀, 按照 Clontech 公司 Blood&Cell Culture DNA Kit 试剂盒的方法进行。

1.2.3 *Myf5* cDNA 序列的扩增和克隆 参照 GenBank 中已登录的条纹狼鲈 (*Morone saxatilis*, AF463525)、花鲈 (*Lateolabrax japonicus*, DQ407725)、斑马鱼 (*Danio rerio*, AF270789)、大西洋鲑 (*Salmo salar*, DQ452070)、褐牙鲈 (*Paralichthys olivaceus*, DQ872515)、虹鳟 (*Oncorhynchus mykiss*, AY751283)、红鳍东方鲀 (*Takifugu rubripes*, AY445319) 等鱼的 *Myf5* 基因的 cDNA 序列, 在同源保守区内设计引物, 上游引物 p1 位于起始密码子处: 5' -ATGGAY (T/C) GTCTTCTCV (G/A/C) M (A/C) CATCCC-3', p2: 5' -CGCCATCCAGTACATCGA GAG-3', 下游引物 p3 位于终止密码子处: 5' -TCACAGK (G/T) ACGTGGTAGACGGG-3'。反转录过程按照 TaKaRa RNA PCR Kit (AMV) Ver.3.0 试剂盒上的操作步骤进行, 反应结束后采用 p1 和上述试剂盒中的 M13 Primer M4: 5' -GTTTCCAGTCACGAC-3' 进行 PCR 扩增, 94 °C, 3 min, 1 个循环; 94 °C, 30 s; 54 °C, 30 s, 72 °C, 1 min, 共 28 个循环; 72 °C 7 min。反应结束后, 用 p1 和 p3 引物进行巢式扩增得到开放阅读框 (ORF) 序列, 用 p2 和反转录试剂盒中的 oligo dT Adaptor primer 为引物进行 3' RACE 扩增, PCR 反应程序同上。扩增产物经低熔点琼脂糖凝胶回收纯化, 与 pMD19-T 载体连接, 转化感受态大肠杆菌 DH5 α , 转化子用碱裂解法提取质粒, 酶切鉴定插入片段的大小, 挑取阳性转化子送上海英竣公司测

序。使用 Vector NTI 7.0 软件将上述两条序列进行拼接得到 ORF 与 3' 非翻译区序列。

1.2.4 *Myf5* 内含子序列的扩增和克隆 以基因组 DNA 为模板, p1、p3 为引物进行 PCR 扩增, 94 °C 4 min, 每个循环包括 94 °C 45 s, 56 °C 45 s, 72 °C 1 min 30 s, 共 30 个循环, 72 °C 10 min。扩增产物回收纯化、转化、鉴定、测序过程方法同上。并用 Vector NTI 7.0 软件进行外显子与内含子的序列拼接。

1.2.5 *Myf5* 基因组 5' 调控区域序列的扩增和克隆 在 *Myf5* 基因第一外显子处设计 2 条下游引物 GSP1: 5' -CAGCCCGTCTGCGGTCCACAAAG TTG-3', GSP2: 5' -TGCTCGTCTCCTCAGAACCA TCAAG-3'。按照 GenomeWalker Universal Kit 试剂盒的操作步骤, 使用 *EcoRV* 酶切基因组 DNA 后, 与试剂盒中的接头连接建库。使用试剂盒提供的引物 AP1: 5' -GTAATACGACTCACTATAGGGC-3' 和自备引物 GSP1 进行 PCR 扩增, 先行 2 个循环 94 °C 25 s, 72 °C 3 min, 再行 32 个循环, 94 °C 25 s, 67 °C 3 min, 最后一个循环 67 °C 7 min。取该 PCR 产物稀释 50 倍后为模板, 以 AP2: 5' -ACTATAGGGCACGCGTGGT-3' 和 GSP2 为引物进行巢式扩增, 反应程序同上。扩增产物转化、鉴定、测序过程方法同上。

1.2.6 生物信息学分析 序列同源性分析采用 NCBI 数据库 BLAST 程序。氨基酸序列结构分析采用 Vector NTI 7.0 和 Signal 3.0 server 软件。使用 MatInspector Release professional 7.4.5 软件对启动子上的顺式作用元件进行预测, 所有参数设置均使用默认值, 核心序列矩阵相似度与序列矩阵相似度比值均大于 0.8。

2 结果与分析

2.1 大口黑鲈 *Myf5* 基因 cDNA 和对应氨基酸序列

大口黑鲈总 RNA 提取结果如图 1(a) 所示, 28S、18S 和 5.8S rRNA 条带完整。*Myf5* ORF 的 PCR 和 3' RACE 扩增结果特异, 与预期片段大小相符。

克隆得到的大口黑鲈 *Myf5* cDNA 序列及推测的氨基酸序列见图 2, *Myf5* cDNA 长 1 093 bp, 其中 ORF 长 723 bp, 共编码 240 个氨基酸; 3' 非翻译区长 370 bp, 存在 Poly (A) 加尾信号 ATTAAG。该基因编码的氨基酸序列具有一典型的碱性螺旋-环-螺旋 (basic helix-loop-helix, bHLH)

结构域,位于第56~124个氨基酸处。其中,第56~80氨基酸处为富含精氨酸与赖氨酸的碱性区,易于与带负电的DNA结合。经软件预测该蛋白无信号肽序列,属核蛋白,与其只在肌细胞内特异表达的特性相符。

将15种脊椎动物*Myf5* bHLH结构域的序列进行比对,可以看到螺旋1和2区域内碱性氨基酸与酸性氨基酸(天冬氨酸和谷氨酸)并存,这种两亲性的特征使得其与具有bHLH结构域的其他蛋白聚合成为可能;同时发现不同物种中存在序列完全一致的疏水性氨基酸(苯丙氨酸、异亮氨酸和亮氨酸),它们可能在螺旋表面通过疏水作用稳定螺旋结构,与该蛋白以二聚体形式发挥作用相吻合(图3)。

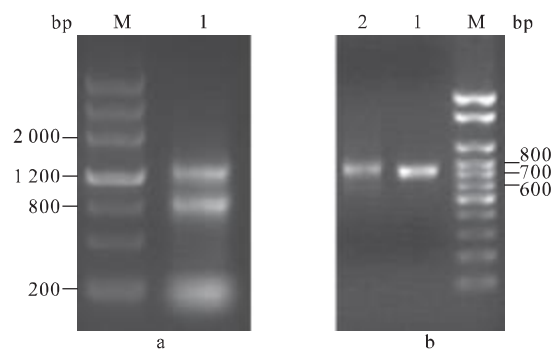


图1 RNA抽提检测结果(a)及*Myf5* ORF的PCR和3' RACE扩增结果(b)

M: 核酸分子量标准; a: 1, RNA抽提结果; b: 1, ORF的PCR扩增结果; 2, 3' RACE扩增结果

Fig.1 Extraction result of RNA (a) and PCR amplified results of *Myf5* ORF and 3' RACE (b)

M: marker; a: 1 extraction result of RNA; b: 1, PCR product of *Myf5* ORF; 2, PCR product of 3' RACE

```

1  ATG GAC GTC TTC TCA CCA TCC CAG GTC TAC TAC GAC AGA GCG TGT GCT TCG TCT CCA GAT
1  M  D  V  F  S  P  S  Q  V  Y  Y  D  R  A  C  A  S  S  P  D
61  AGC CTG GAG TTT GGA CCC GGT GTG GAG CTT GAT GGT TCT GAG GAG GAC GAG CAT GTC AGG
21  M  L  E  F  G  P  G  V  E  L  D  G  S  E  E  D  E  H  V  R
121 GTT CCC GGG GCA CCT CAC CAG CCG GGA CAC TGT CTC CAG TGG GCC TGC AAG GCC TGC AAG
41  V  P  G  A  P  H  Q  P  G  H  C  L  Q  W  A  C  K  A  C  K
181 CGC AAG TCC AAC TTT GTG GAC CGC AGA CGG GCT GCC ACC ATG CGT GAG CGC CGG CGG CTG
61  R  K  S  N  F  V  D  R  R  R  A  A  T  M  R  E  R  R  R  L
241 AAG AAG GTC AAC CAC GCG TTC GAG GCT TTG AGA CGT TGC ACC TCG GCC AAC CCC AGC CAA
81  K  K  V  N  H  A  F  E  A  L  R  R  C  T  S  A  N  P  S  Q
301 CGT CTG CCA AAG GTG GAG ATC CTG CGC AAC GCC ATC CAT TAC ATT GAG AGT CTG CAG GAC
101 R  L  P  K  V  E  I  L  R  N  A  I  H  Y  I  E  S  L  Q  D
361 CTG CTA CGA GAG CAG GTG GAA AAC TAC TAC TGC CTA CCT GGA GAG AGC AGC TCT GAG CCT
121 L  L  R  E  Q  V  E  N  Y  Y  C  L  P  G  E  S  S  S  E  P
421 GGT AGC CCA CTG TCC AGC TGC TCT GAC GGC ATG GCT GAC AGC AAC AGT CCA GTG TGG CAA
141 G  S  P  L  S  S  C  S  D  G  M  A  D  S  N  S  P  V  W  Q
481 CAT CTG AAT GCA AAC TAC AGC AAC AGA TAT TCA TAT GCG AAA AAT GAG AGT GTG GGC GAT
161 H  L  N  A  N  Y  S  N  R  Y  S  Y  A  K  N  E  S  V  G  D
541 AAG ACA GCT GGA GCC TCT AGT CTG GAG TGT CTC TCC AGC ATC GTT GAT CGC TTG TCC TCG
181 K  T  A  G  A  S  S  L  E  C  L  S  S  I  V  D  R  L  S  S
601 GTG GAG TCC AGC TGC GGA CCG GTG GCT CTG AGA GAC ATG GCC ACC TTC TCC CCT GGG AGC
201 V  E  S  S  C  G  P  V  A  L  R  D  M  A  T  F  S  P  G  S
661 TCC GAC TCG CAG CCC TGC ACG CCG GAG AGC CCC GGA TGC AGG CCC GTC TAC CAC GTC CTG TGA
2S1 S  D  S  Q  P  C  T  P  E  S  P  G  C  R  P  V  Y  H  V  L  *
AGGAACTTGCTTTAACGTGGCTATATTGCCACAGTCAGGCGCCAGCTTTACACAAACACCAGCTGCATTTGCAACAAGAAGAGATAAGA
ACTAGTTTTGTGAATTTTAAAGAGTCTGAATTTGAGGACCTGTGGCAAGTAGCTTTTTTGTACATGAGATTGTAATATGTGATGTAATT
GCCATTTATTCTATACATGCTATTATACTCAATGAGACATATTTAATTATGAGAGTACATGTAATGTTGCATAATCCACATGAAATGGT
ATTTAAGCAGTTTTTCTTACTTTCCATGTTGAATTTTCATGAACATTAAATCTTTTCACTGTTTGTGTAAAAAAGAAAGAAAAA

```

图2 大口黑鲈*Myf5* cDNA序列及推测的氨基酸序列

3'非翻译区的Poly(A)加尾信号ATTAAA用阴影标出,碱性螺旋-环-螺旋结构域用下划线标出

Fig.2 *Myf5* cDNA sequence and deduced amino acids of largemouth bass

Polyadenylation signal (ATTAAA) is shown in a gray box, and the bHLH domain is underlined.

<i>M.salmopodes</i>	CKACKRKS	NFVDRRRAATMRERRRL	KKVNHA	FEALRR	CTSANPS	QRLPK	VEILRNA	IHYIESL	QDLLRE
<i>M.saxatilis</i>	CKACKRKS	NFVDRRRAATMRERRRL	KKVNHA	FEALRR	CTSANPS	QRLPK	VEILRNA	IQYIESL	QDLLHE
<i>L.japonicus</i>	CKACKRKS	NFVDRRRAATMRERRRL	KKVNHA	FEALRR	CTSANPS	QRLPK	VEILRNA	IQYIESL	QDLLRE
<i>Polivaceus</i>	CKACKRKS	SFVDRRRAATMRERRRL	KKVNHA	FEALRR	CTSANPS	QRLPK	VEILRNA	IHYIESL	QELLRE
<i>D.erio</i>	CKACKRKA	STVDRRRAATMRERRRL	KKVNHA	FEALRR	CTSANPS	QRLPK	VEILRNA	IQYIESL	QELLRE
<i>C.carpio</i>	CKACKRKA	STVDRRRAATMRERRRL	KKVNHA	FEALRR	CTSANPS	QRLPK	VEILRNA	IQYIESL	QELLRE
<i>S.salar</i>	CKACKRKS	STVDRRRAATMRERRRL	RKVNHG	FEALRR	CTSANPS	QRLPK	VEILRNA	IQYIESL	QELLHE
<i>O.mykiss</i>	CKACKRKS	STVDRRRAATMRERRRL	KKVNHG	FEALRR	CTSANPS	QRLPK	VEILRNA	IQYIESL	QELLHE
<i>T.rubripes</i>	CKACKRKS	NFVDRRRAATMRERRRL	KKVNHA	FDALRR	CTSANSS	QRLPK	VEILRNA	IQYIESL	QELLRE
<i>T.nigroviridis</i>	CKACKRKS	NFVDRRRAATMRERRRL	KKVNHA	FDALRR	CTSANSS	QRLPK	VEILRNA	IQYIESL	QELLRE
<i>H.sapiens</i>	CKACKRK	STMDRRKAATMRERRRL	KKVNQA	FETLKR	CTTNP	NQRLPK	VEILRNA	IRYIESL	QELLRE
<i>M.musculus</i>	CKACKRK	STMDRRKAATMRERRRL	KKVNQA	FETLKR	CTTNP	NQRLPK	VEILRNA	IRYIESL	QELLRE
<i>B.taurus</i>	CKACKRK	STMDRRKAATMRERRRL	KKVNQA	FDTLKR	CTTNP	NQRLPK	VEILRNA	IRYIESL	QELLRE
<i>G.gallus</i>	CKACKRK	STMDRRKAATMRERRRL	KKVNQA	FETLKR	CTTAN	PNQRLPK	VEILRNA	IRYIESL	QELLRE
<i>X.laevis</i>	CKACKRK	SSITDRRKAATMRERRRL	KKVNQA	FETLKR	CTTNP	NQRLPK	VEILRNA	IQYIESL	QDLLRE
	Basic		helix1		loop		helix2		

图3 脊椎动物中 *Myf5* 碱性螺旋—环—螺旋结构域氨基酸序列比较

注: 从上至下依次为大口黑鲈 (*Micropterus salmoide*) 与条纹狼鲈 (*Morone saxatilis*, AF463525)、花鲈 (*Lateolabrax japonicus*, DQ407725)、褐牙鲈 (*Paralichthys olivaceus*, DQ872515)、斑马鱼 (*Danio rerio*, AF270789)、鲤 (*Cyprinus carpio*, AB012883)、大西洋鲑 (*Salmo salar*, DQ452070)、虹鳟 (*Oncorhynchus mykiss*, AY751283)、红鳍东方鲀 (*Takifugu rubripes*, AY445319)、黑青斑河鲀 (*Tetraodon nigroviridis*, DQ453127)、人 (*Homo sapiens*, X14894)、鼠 (*Mus musculus*, NM_008656)、牛 (*Bos taurus*, M95684)、鸡 (*Gallus gallus*, X73250) 和非洲爪蟾 (*Xenopus laevis*, X56738) 碱性螺旋—环—螺旋结构域的氨基酸序列比较, 完全一致序列用深色区域表示, 部分一致序列用浅色区域表示。

Fig.3 Comparison of vertebrate *Myf5* bHLH domain

Note: The putative amino acid sequence of largemouth *Myf5* bHLH domain is compared to the amino acid sequences of striped bass (*Morone saxatilis*, AF463525); sea perch (*Lateolabrax japonicus*, DQ407725); flounder (*Paralichthys olivaceus*, DQ872515); zebrafish (*Danio rerio*, AF270789); carp (*Cyprinus carpio*, AB012883); Atlantic salmon (*Salmo salar*, DQ452070); rainbow trout (*Oncorhynchus mykiss*, AY751283); Fugu rubripes (*Takifugu rubripes*, AY445319); pufferfish (*Tetraodon nigroviridis*, DQ453127); human (*Homo sapiens*, X14894); mouse (*Mus musculus*, NM_008656); cow (*Bos taurus*, M95684); chicken (*Gallus gallus*, X73250) and African clawed frog (*Xenopus laevis*, X56738). Dark and light gray boxes represent identical and similar amino acid sequences respectively.

2.2 大口黑鲈 *Myf5* 基因组序列及启动子区的调控元件

以加州鲈基因组为模板, 以 p1 和 p3 为引物, 扩增内含子序列得到约 2 000 bp 的特异条带, 结果如图 4-a 所示。以 *EcoR* V 酶切建库的基因组为模板, AP1 和 GSP1 为引物, 进行 Genomewalker 第一轮 PCR 扩增, 出现弥散条带, 如图 4-b 1 所示。再以产物为模板, AP2 和 GSP2 为引物, 进行第二轮巢式 PCR 扩增, 出现 1 条约 2 700 bp 的特异条带, 如图 4-b 2 所示。

经序列拼接可知大口黑鲈 *Myf5* 基因共存在 3 个外显子和 2 个内含子。外显子 1、2、3 分别长 453 bp, 73 bp 和 197 bp。内含子 1、2 分别长 920 bp 和 328 bp, 内含子与外显子边界均符合 GT-AG 规则。通过大口黑鲈、条纹鲈、褐牙鲈和斑马鱼的基因组序列比对, 发现内含子 1 中存在 123 bp 同源性为 86% 的高度保守序列 (图 5), LIN 等已

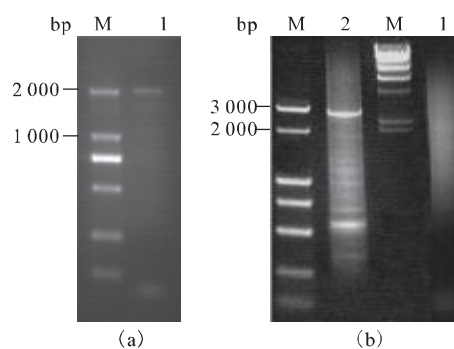


图4 内含子序列扩增结果 (a) 和 genomewalker 扩增 5' 调控区序列结果 (b)

M: 核酸分子量标准; a: 1, 内含子序列扩增结果; b: 1, Genomewalker 第一轮 PCR 扩增结果; 2, Genomewalker 巢式 PCR 扩增结果

Fig.4 PCR amplified result of introns (a) and Genomewalker amplified results of *Myf5* promoter sequence (b)

M: marker; a: 1, amplified PCR product of introns; b: 1, Genomewalker first PCR product of *Myf5* promoter; 2, Genomewalker nested PCR product of *Myf5* promoter

经证实斑马鱼中的该段序列能够显著抑制 *Myf5* 的表达^[18]。经软件预测该段序列含有肌源蛋白结合因子 1 (COMP1, Cooperates with myogenic proteins 1) 结合位点、羧基端锌指蛋白结合结构

域 (Carboxy-terminal zinc finger domain)、转录增强子 (Transcriptional enhancer factors 1, 3, 4, 5)、CCAAT 框等, 它们可能在 *Myf5* 基因的时空和组织特异性表达过程中起到调节作用。

<i>M.salmipodes</i> (1003)	CTGCATTCCATCCTGACTGCATGTTCTGCACCGCTGCCAGGATTTTCTGCAGATTACTTGGAGAAA-G
<i>M.saxatilis</i> (1046)	CTGCATTCCATCCTGACTGCATGTTCTGCACCGCTGCCAGGATTTTCTGCAGATTACTTGGAGAAA-G
<i>P.olivaceus</i> (908)	CTGCATTCCATCCTGCTGCATGTTCTGCACCGCTGCCAGGATTTTCTGCAGATTACTGGGAGAAA-G
<i>D.erio</i> (2133)	CTGCATTCCATGACTGACTGCATGAGCTGCGCTGCTGCCAGGATTTTCTGCAGATTACTCCGAAAAAAG
<i>M.salmipodes</i>	TCATCTTGACAAACTGGTGTGGCTTCCACTGATACAATCGATTGCGCCACGC (1225)
<i>M.saxatilis</i>	TCATCTTGACAAACTGGTGTGGCTTCCACTGATACAATCAATTGCGCCACGC (1168)
<i>P.olivaceus</i>	TCATCTTGACAAACTGGTGTGGCTTCCACTGATACAATCGATTGCGCCATGC (1030)
<i>D.erio</i>	TCATCTTGACAAACTGGTGTGGCTTCCACTGACTCAATCAATTGCGCTTGC (2225)

图 5 4 种鱼类 *Myf5* 基因内含子 1 中高度保守序列比较

注: 比对序列依次为加州鲈 (*Micropterus salmoide*)、条纹狼鲈 (*Morone saxatilis*, AF463525)、褐牙鲈 (*Paralichthys olivaceus*, DQ872515)、斑马鱼 (*Danio rerio*, AF270789)。完全一致序列用深色区域表示, 部分一致序列用浅色区域表示。

Fig.5 Sequence comparison of highly conserved region of intron-1 in four fish

Note: The nucleotide sequence in intron 1 of largemouth bass (*Micropterus salmoide*) is compared to stripped bass (*Morone saxatilis*, AF463525), flounder (*Paralichthys olivaceus*, DQ872515) and zebrafish (*Danio rerio*, AF270789). Dark and light gray boxes represent identical and similar nucleotide sequences respectively.

克隆得到的大口黑鲈 *Myf5* 基因启动子区序列长 2 690 bp, 存在基础转录调控元件 TATA 框、GC 框、CCAAT 框 (核转录因子 NFY 能对该序列进行调控) 和八聚体转录因子 1 (Otc-1) 结合位点。同时发现与肌肉特异性转录调控相关元件有 E-box 18 个、肌细胞特异增强因子 2 (Myocyte specific enhancer factor2, MEF2) 结合位点 3 个、上游激活因子 (Upstream stimulating factor, USF) 结合位点 2 个、血清应答因子 (Serum response factor, SRF) 结合位点 2 个、核转录因子 sp1 结合位点 8 个、肌肉特异性金属硫蛋白 MTBF (Muscle-specific Mt binding site factor) 结合位点 3 个。与早期诱导肌肉分化信号有关的调控元件有早期生长应答因子 α (Early growth response gene α , EGR α /TIEG) 结合位点、早期快反应生长应答因子 1 (Early growth response gene 1, Egr-1/Krox-24) 结合位点、早期快反应生长应答因子 2 (Early growth response gene 2, Egr-2/ Krox-20) 结合位点、生长依赖因子 1 (Growth factor independence 1) 结合位点、T 细胞因子 TCF/LEF-1 结合位点, 它们可能是调控 *Myf5* 基因在体节形成早期表达的主要因子。具体序列见图 6。

3 讨论

3.1 大口黑鲈 *Myf5* 基因氨基酸序列与其他脊椎动物同源性比较

大口黑鲈 *Myf5* 基因氨基酸序列与 14 种脊椎动物相比同源性介于 56%~93% 之间, 其中 bHLH 结构域与其他鱼类的同源性可达到 90% 以上, 相比其他脊椎动物也有 82% 以上的同源性 (表 1)。序列的高度保守性与其重要的生物学功能有关。HLH 结构域能够介导 *Myf5* 蛋白与 E 蛋白 (如 E12、E47、HEB 和 ITF 等) 结合形成二聚体, 再通过碱性区的作用与肌肉特异性基因如肌球蛋白轻链基因、肌酸激酶基因等的上游调控序列中的 E-box (CANNTG) 结合从而启动这些基因的表达, 促进成肌细胞的分化^[2-4]。曾有报道认为该蛋白的氨基端和羧基端大部分序列同样能够作为内在的转录激活域与碱性区共同激活肌肉特异性基因的转录^[19]。本实验将大口黑鲈氨基酸序列的 1~54 处氨基端与 132~240 处的羧基端序列与其他脊椎动物进行比较, 发现保守性低于 bHLH 区 (表 1), 提示, 这两段序列可能在转录激活过程中起着辅助作用。

-2690 ATCCCTAAGCAGTAATGTGATTTAATGGACTCCTGCTCACTCACTCAATCTGATAGCAGTGTACTTCACTACATTCTGCAGCTCCATG**GGTATCT**GCT
Growth factor independence 1(-) MTBF
-2590 AATTTAATGTCTGTTTGGCAATGACTCATTTCACCAAAATGTAGACCTAGTATCAGCAATAACAATCTGTAG**AAAAATAAAGGCTT**GGTTCAATGTGG
E-Box 18 MEF2
-2490 G**TA**CTTATTACCTTCTAAATCTATATATTTAGTTGACATACTGTATACCTGCATTTTGTACTAATGGAAGACTGTATTTGTCTGTAACCAAGTCTACC
-2390 **CAGCT**TGTTTGAACCTTACTGTCTGCTCCATCAGACTGCATGCGTGCACCTCTCACAAGAAACAGCACCCGACTTTCTCTAGTCTGACTTTACAGGA
E-Box 17
-2290 TTTCGGTGTACTGATGACACCATATAATTC**CATTTC**GCTGTGAAGCAGTGTTTTAAAG**TCAAA**GCTGATGTAAGGAGAGCTGGTATAAGCCTGGAG
E-Box 16 TCF/LEF-1
-2190 GCTGAATGTTCTGATTTATCCAGACTCTCCTTGAATATCTGTAGGAAAAGAATTCAGTGCATCATGGGCTGGCAGTGAAGTGCACCTCTTTCTGTTTAC
-2090 TGCCAA**CAAA**TGTTTGGCACTCAAAATGTTTTTGTGTTTGTTCATTAATTTCTGTGTGAATGTAACATGGACTTGGACTTGT**GGGGTGTGG**
E-Box 15 EGRα / T1HG
-1990 **GGGGTGG**AICTTTGGTTTATTCACCTTTTGTATATATATATATATTTATGTAAATGTTGTGTCAGAGAGAGTATAATTAATATGGCCTTTAAATTTT
Egr-2/Krox-20
-1890 GCCTAAAGAA**AGCTTATAAAGGTATCTA**TATTGCAGGAAAGTAAGTGATATATTGCCTGTCTGCGGGGGGCTTGTGTGCACCAACAAGACATAT
SRF (-) MTBF
-1790 TAAGTCAAGCATGAGGAGTTTGAAGAATAAGGTTTTCAGACGTGAGGAAGTGGACATAAGACATCACCTCCTGCCCTCACTTAACCTGTCTCT
-1690 GACCTTTGAGAACAAACAA**CAAAAAAC**CTCTGAGAGTTCCTCAAGAGACCGAGCG**AGCGCA**CGGATCACATC**ACGGGGC**TGTGGCAACGACTGCG
MEF2 Egr-1/Krox-24 USF (-)
-1590 CCAATATAAGCTCCGGTATTCAGTAAC**CATGCT**CTTCGCAATCTAATCCAGCTTCAATTAGCTCTCGATTGTTACCAAGTTTTCTCATAGTCG
E-Box 14
-1490 ACGTTGTTCTCAGCCTGAACAGCCTTTGATCTCTCAGCAGCCACAAGCGCAGGCGGATGCTCTT**CAGCT**TGGCTAATCCAACTCCAGCGG
E-Box 13
-1390 AATTCAGAGGCAGCACTTACTGGTGTCTTCTAGGTGCGCACAGTTCACAGAGAAGGGGGCAGCCCAACGTTTGTACATTGTGAGAAATATCCGCAT
-1290 TGTATCCTTGTGAGCTGTCAACAACATTATACCTAAACATGTGAGGTGCAAGTCAAAAGCCCTTCACTATGATAATATCTGCTCCCTGAAAGAGAG
-1190 CAATTC**CGCAAT**AGAAAGACAACCTGACTCCAGTGCAGACGATGTTATGATCTTTATTTCTAGAAAATTTAGGCAATCTCTCTATTAACTGA**AGAT**
-1090 **ATTTC**CATTGTATAAAGTCTAAGTGTCTAAACCGCCTTAGTTTGTATTCTGTAAAGTAACATAACATAAGTAGCATAAGTAACAAGCAATATTTCT
MTBF E-Box 12
-990 GCAT**TATAAC**ACTTTTGCACTCTCATTAAAG**CCCCTAAAC**CCCCAGTCT**TCACCT**TGAATGAAACCTACTGCACCAATAATGA**CATTG**TGTGTGCG
MEF2 TCF/LEF-1 (-) SP1 SP1 E-Box 11 E-Box 10
-890 TGAATTGAAGTCT**CATTC**CGTCAATMACTTTAATGGTTAAATTTGGTCTTGGAGGCAAAATGGTGCAAGGAATGGGCAGTGGT**CAATT**TG
E-Box 9 E-Box 8
-790 TAATTACTGTTTAAATGAGCTTGG**CAAAC**AGGGATGGTCTCATTACAGCACCTTCAGTGATGAG**TCAGGG**TGTTAGAGGGGGCTTACCAGTGAAT
SRF SP1
-690 GTGAATTCAGAGGGCTTGAATGTTACACCTCCCTT**CAGCT**GGCT**CATCT**CAACGCTGTTGCTCCACCGTAAAAATGGCCTGCATCTCT**AT**
E47/E12 Dimer
-590 **AAGTG**GTACCCCTCTTTTCTGCACAGTTTGACTAAACAAACGCACTCTCTTAATATTTATGTCTTATAGTTAAGTGCATTTTCACTTGTGAGCTAT
E-Box 5
-490 **GACATGA**TGCTGGAGGCGGTTACAGATATATCAGGCACAATGAAGGACTTATAACAGG**TCATG**GTGGTTCAGCT**CAAA**TGTTCAACCACTC
USF E-Box 4 E-Box 3
-390 CTGAACATTC**CAAGT**TTGATTCATCATGGCTTGCAGGAGAAAGGCTCTT**GAAGGA**GAC**ATTGGTG**ATGAGTAGGAAACACCAAGT**AGT****CAGA**
E-Box 2 SP1 SP1 E-Box 1
-290 **TG**TGGAAACTCTGCAGGGTACT**GGGGGG**TGAGAGACAGTGGGTTGGGATGG**GGCAGGGAGG**GTGCTGAGCTAACCCAGCCTGCCTTCTGGCTGG**CCA**
SP1 SP1 NF-Y
-190 **AT**GGG**CATGGTAAT**TAGATCTG**CCAA**TG**GGGCCC**TGGGCTTGGCTTGGGCA**ATATA****AGGGGGGCC**TTGGGCAACAGACCCCTCATATCACTCAGA
Oct-1 NF-Y GC-Box TATA-Box SP1
-90 GGTCTCTCTCTCAACCCACTCTCTACCTCTCTTCTACCATTCGCGCCATCTTTCCATCCATTCAAGTCTCTTCTTCAACAAGCT
+1 Exons 1→
atggacgtctctctaccatccaggtctactacgacagagcgtgtgctctctccagatagcctggagtttggaccgggtgtggagctttaggtctctg
aggaggacagcatgtcagggttccgggggacclccacgcgggacclgtctccagltgggctgcaaggcctgcaagcgcaagtcacacttlttgga
ccgacagcgggttgcacacatgctgagcgccggcggtgaagaaggtcaaccacggttgcaggcttltgagacgttgcaccltggcccaacccagccaa
cgtctgccaaaggtgagatcctgcgaacgcatccattacattgagactgtcaggacgtctacgagagcaggtggaacctactactgctcactgt
Intron 1→
gagagagcagctctgagcctggtagccactgtccagctgctctgagcgcaggtTAAGACCTCGGATCCATGTTACGCTCTTCATGCAACATCTGGTT
GCTCTAAACCTAATAAATGTGAGGAGAGTAACGACAACCTGCATGCAAGACAGGACTTCATCTTGTAGAAACATCACCTGAAAGGCTTTGTGTG
TGCCATATCTCATTACCAATGATAGATTTTCAATTTCAAAATTGAGCTTGGATTCCAGTTTGAACACCTATTAATCAATAAGCTCATCTTGCCCAAGA
CCTCAATTAACATGCTTCCATGCTATCTCTTCCATTTATCTTTAATAAATAATCTATCTGTAGTCTGCATGTATATCTGCTCTTTTGTG
TGAAGCTGTATTACATTATAAGTACGATGAATTCATTGTTCTGAACACAGCTGGAATGTGACAGGGATCAGCTAATAACAAAGTACTTATAATG
TGGTGGAAATAATGGCTTTCCACCACTGGACACTTTATAGCCTTAATAACCAAACTAAATCCACCTTACTGAAGTCTGGCAGGAAAACTGGGGCTT
TTCTGGCTTCCATCTCCTGACTGATGTTCTGCACCGCTGGCAGGATTTTCTGCAGATTTACTTGGAGAAAGTGTATCTTGCACAACTGGTGTGCTTCC
ACTGATACAATCGATTGCCACGCTCAGAGCTGGCAGGGAAGGAGGCTACACTGCCACTTTGGCTTAAAGTGGATGTGAAATAAAGAGCCCATAA
CTTTTGAAATTCAGGATTTATAGCCTGTCCAGAATATCATGAAGGGCACTGGTGTGTTTGCACCTCTCTGCTCTTCTATCCCTACAAATTAAC
Exons 2→
ATAACACATATTCTGTTGTGCTTTTGTGCAAGAATTCAMGGTAAATTTGATTTTCTTCTTGTGAGGctgacagcaacagtcaggltggcga
Intron 2→
catctgaatgcaaacacagacagatattcatatgcgaagaatGTAAAGACCTTTGACAGGAAACAGCAACCTGATAGCAATATGACATAACTTTGTG
CCAACATAGTGTCTTAAACCTAATAACTGAATAGTAGATTATATTAAGATTTGTAATGTGTATTTTATTTTGGCCAGGGAGCTATATAGTAC
CAGGATTTAAGTATATCATTAATAAATAAAACCCCATACAAACCTGATAAAAGATTTATGTTTAAAAAAGTCATTTTATTAAGATGGATGTACAT
Exons 3→
ATGAAATTTGTTAAGTGCAGTTATCCCTGTATTAATCCTGACTGATCTGGACTGTCTTCTGTTATCAGagagtgtggcgataagacagctgga
gcctctagcttggagtgtctctccagcatgttgcctgttgcctgttggagtcagctgcggaccggtggctctgagagacatggccacctctccc
ctggagctcgcagctcgcagcctgcagcggagagcccgatgcagcccgcttaccacgtcctgtga

图6 大口黑鲈 *Myf5* 基因序列

TATA 框、GC 框、CCAAT 框和 Oct-1 结合位点用方框标出, E-box、MEF2、USF、SRF、MTBF、sp1 结合位点用粗线斜体表示并用下划线标出, EGRα、Egr-1、Egr-2、TCF、生长依赖因子 1 结合位点以粗线斜体表示。外显子用小写英文字母表示, 内含子用大写英文字母表示。起始密码子位置设为“+1”。

Fig. 6 Regulatory factor binding sites of largemouth bass promoter and the sequence of exons and introns

TATA-box, GC-box, CCAAT-box and Oct-1 binding sites are shown in box. E-box, MEF2, USF, SRF, MTBF and sp1 binding sites are underlined. EGRα, Egr-1, Egr-2, TCF, Growth factor independence 1 binding sites in bold and italic characters. The exons are shown in low-case character and the introns are displayed in capitalization. The start codon is designed as “+1”.

表 1 大口黑鲈与其他脊椎动物 *Myf5* 基因氨基酸序列的同源性比较
Tab.1 Homology analysis of largemouth bass *Myf5* amino acid sequence to those of other vertebrates

物种 Species	bHLH 结构域 / % bHLH domain	氨基端 / % Amino terminus	羧基端 / % Carboxyl terminus	全序列 / % Complete amino acids	编码氨基酸长度 / 个 Length of peptide
条纹鲈 <i>Morone saxatilis</i>	97	93	89	92	239
花鲈 <i>Lateolabrax japonicus</i>	99	93	89	93	240
褐牙鲈 <i>Paralichthys olivaceus</i>	97	93	88	92	239
鲤 <i>Cyprinus carpio</i>	93	88	59	76	237
斑马鱼 <i>Danio rerio</i>	93	83	60	75	240
红鳍东方鲀 <i>Takifugu rubripes</i>	94	90	86	89	237
黑青斑河鲀 <i>Tetraodon nigroviridis</i>	94	92	68	81	246
大西洋鲑 <i>Salmo salar</i>	88	86	70	79	239
虹鳟 <i>Oncorhynchus mykiss</i>	90	86	74	82	239
牛 <i>Bos taurus</i>	82	40	49	56	255
人 <i>Homo sapiens</i>	84	40	49	57	255
鼠 <i>Mus musculus</i>	84	36	52	57	255
鸡 <i>Gallus gallus</i>	85	35	49	56	258
非洲爪蟾 <i>Xenopus laevis</i>	85	38	51	57	255

3.2 大口黑鲈 *Myf5* 基因启动子区转录调控元件的分析

克隆得到的大口黑鲈 *Myf5* 基因上游调控序列中存在肌肉特异启动子调控元件,分别为 E box 18 个,MEF2 结合位点 3 个,SP1 结合位点 8 个,SRF 结合位点 2 个,USF 结合位点 2 个。肌肉特异性基因(如肌球蛋白轻链、肌酸激酶等基因)的启动子中的 E-box 区域是生肌调节因子家族成员的结合位点^[3],发现大口黑鲈 *Myf5* 启动子中也拥有数量较多的 E-box,从而推测该家族其他成员可能通过反馈作用抑制或激活 *Myf5* 的后续表达。在 *MyoD* 基因敲除小鼠中,*Myf5* 基因的表达量比野生型小鼠高四倍^[20],提示 *MyoD* 可能作为负调控因子,反馈作用于 *Myf5* 启动子中的 E-Box 区域,从而限制其表达水平。

在骨骼肌的发育过程中,MEF2 家族是另一类关键的转录因子,它们与 bHLH 蛋白因子家族相互作用,形成一个有助于强化和稳定生肌潜能的正反馈回路^[21]。本研究在大口黑鲈 *Myf5* 启动子区域也发现了 MEF2 结合位点,推测 MEF2 可能作为正调控因子,增加大口黑鲈 *Myf5* 的表达水平。

SP1 在 1999 年被证实是肌肉特异性转录的关键调节因子^[22];2003 年 Guo 等^[23] 研究证明,SP1 因子的过度表达可以解除其他因子对 *MyoD* 基因

的抑制作用^[23]。大口黑鲈 *Myf5* 启动子区域拥有数量较多的 SP1 结合位点,推测它们在 *Myf5* 基因的调控过程中发挥着重要的作用。

近年,上游激活因子 USF 和血清应答因子 SRF 被相继证实能够与 *MyoD* 启动子结合^[24-25],调控其转录活性,推测它们可能对同一家族的 *Myf5* 基因也有调控作用。

目前,人们对于肌肉发生相关调控因子组成的调控网络所知甚少,生肌决定因子家族相关信号通路是其重要组成部分。通过对大口黑鲈 *Myf5* 基因 cDNA 和启动子序列的克隆,分析得到了相应调控因子,这些蛋白因子与 *Myf5* 基因的相互作用和相关信号通路的研究将是今后研究的重点,以期为进一步的肌肉发生调控网络研究提供证据。

参考文献:

- [1] Murre C, McCaw P S, Baltimore D. A new binding and dimerization motif in immunoglobulin enhancer binding, daughterless, *MyoD*, and myc proteins[J]. Cell, 1989, 10 (56): 777 783.
- [2] Jaynes J B, Johnson J E, Buskin J N, et al. The muscle creatine kinase gene is regulated by multiple upstream elements, including a muscle-specific enhancer[J]. Mol Cell Biol, 1988,

- 8(1): 62–70.
- [3] Braun T, Arnold H H. The four human muscle regulatory helix-loop-helix proteins Myf3-Myf6 exhibit similar heterodimerization and DNA binding properties[J]. Nucleic Acids Res 1991, 19(20): 5 645–5 651.
- [4] Faerman A, Shani M. The expression of the regulatory myosin light chain 2 gene during mouse embryogenesis[J]. Development, 1993, 118(3): 919–929.
- [5] Braun T, Buschhausen-Denker G, Bober E, et al. A novel human muscle factor related to but distinct from *MyoD1* induces myogenic conversion in 10T1/2 fibroblasts[J]. EMBO J, 1989, 8(3): 701–709.
- [6] Kablar B, Krastel K, Ying C, et al. *MyoD* and *Myf-5* differentially regulate the development of limb versus trunk skeletal muscle[J]. Development, 1997, 124(23): 4 729–4 738.
- [7] Kablar B, Krastel K, Tajbakhsh S, et al. *Myf5* and *MyoD* activation define independent myogenic compartments during embryonic development[J]. Dev Biol, 2003, 2(58): 307–318.
- [8] Braun T., Rudnicki M A, Arnold H H, et al. Targeted inactivation of the mouse regulatory gene *Myf5* results in abnormal distal rib development and early postnatal death in homozygous mouse mutants[J]. Cell, 1992, 71(3): 369–382.
- [9] Braun T, Bober E, Rudnicki M A, et al. *MyoD* expression marks the onset of skeletal myogenesis in homozygous *Myf5* mutant mice[J]. Development, 1994, 120(11): 3 083–3 092.
- [10] Chen Y H, Tsai H J. Treatment with *Myf5*-morpholino results in somite patterning and brain formation defects in zebrafish[J]. Differentiation, 2002, 70(8): 447–456.
- [11] Chen Y H, Lee W C, Liu C F, et al. Molecular structure, dynamic expression, and promoter analysis of zebrafish (*Danio rerio*) *myf-5* gene[J]. Genesis, 2001, 29(1): 22–35.
- [12] Tan X, Hoang L, Du S J. Characterization of muscle-regulatory genes, *Myf5* and *Myogenin*, from striped bass and promoter analysis of muscle specific expression[J]. Mar Biotechnol (NY), 2002, 4(6): 537–545.
- [13] Tan X, Zhang Y, Zhang P J, et al. Molecular structure and expression patterns of flounder (*Paralichthys olivaceus*) *Myf-5*, a myogenic regulatory factor[J]. Comp Biochem Physiol B, 2006, 145(2): 204–203.
- [14] Ye H Q, Chen S L, Xu J Y. Molecular cloning and characterization of the *Myf5* gene in sea perch (*Lateolabrax japonicus*) [J]. Comp Biochem Physiol B Biochem Mol Biol. 2007, 147(3): 452–459.
- [15] Johansen K A, Overturf K. Quantitative expression analysis of genes affecting muscle growth during development of rainbow trout (*Oncorhynchus mykiss*) [J]. Mar Biotechnol (NY), 2005, 7(6): 576–587.
- [16] Rowlerson A, Veggetti A. Cellular mechanisms of postembryonic muscle growth in aquaculture species[J]. Fish Physiology Series, 2001(18): 103–140.
- [17] Weatherley A H, Gill H S. The role of muscle in determining growth and size in teleost fish[J]. Cell Mol Life Sci, 1989, 9(4): 875–878.
- [18] Lin C Y, Chen Y H, Lee H C, et al. Novel cis-element in intron 1 represses somite expression of zebrafish *myf-5* [J]. Gene, 2004, 334(9): 63–72.
- [19] Braun T, Winter B, Bober E, et al. Transcriptional activation domain of the muscle-specific gene-regulatory protein *myf5* [J]. Nature, 1990, 346(16): 663–665.
- [20] Rudnicki M A, Braun T, Hinuma S, Jaenisch R. Inactivation of *MyoD* in mice leads to upregulation of the myogenic HLH gene *Myf5* and results in apparently normal muscle development[J]. Cell, 1992, 71(3): 383–390.
- [21] 阿里亚斯 A M, 斯图尔德 A. 樊启超, 滕俊琳译. 动物发育的分子原理 [M]. 北京: 高等教育出版社, 2006.
- [22] Hamamori Y, Kedes L, Sartorelli V. Myogenic basic helix-loop-helix proteins and Sp1 interact as components of a multiprotein transcriptional complex required for activity of the human cardiac alpha-actin promoter[J]. Mol Cell Biol, 1999, 4(19): 2 577–2 584.
- [23] Guo C S, Degnin C, Fiddler T A, et al. Regulation of *MyoD* activity and muscle cell differentiation by MDM2, pRb, and Sp1 [J]. Biol Chem, 2003, 278(25): 22 615–22 622.
- [24] Lun Y, Sawadogo M, Perry M. Autoactivation of *Xenopus MyoD* transcription and its inhibition by USF [J]. Cell Growth Differ, 1997, 8(3): 275–282.
- [25] L' honore A, Lamb N J, Vandromme M, et al. *MyoD* distal regulatory region contains an SRF binding CARG element required for *MyoD* expression in skeletal myoblasts and during muscle regeneration [J]. Mol Biol Cell, 2003, 14(5): 2 151–2 162.

Cloning and analysis of *Myf5* cDNA and genomic sequence of largemouth bass (*Micropterus salmoides*)

GUO Yu-han^{1,2}, BAI Jun-jie¹, LAO Hai-hua¹, YE Xing¹, LUO Jian-ren¹

(1. Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Key Laboratory of Tropical & Subtropical Fish Breeding & Cultivation, CAFS, Guangzhou 510380, China; 2. College of aqua-life Science and Technology, Shanghai Ocean University, Shanghai 200090, China)

Abstract: As a member of the myogenic regulatory factors (MRFs), *Myf5* which is responsible for the specification and proliferation of myoblasts, plays an essential role in the growth and development of skeletal muscle. The human *Myf5* was found firstly in 1989, capable of converting mouse 10T1/2C3H fibroblast into myosin expressing myotubes. The *Myf5* gene-knockout mice delayed development of epaxial musculature and died from severe rib abnormalities, suggesting that *Myf5* should primarily regulate the epaxial musculature formation. In fish, *Myf5* cDNA of the carp, rainbow trout, Atlantic salmon, pufferfish and the genomic sequence of zebrafish, striped bass, sea perch, flounder and *Fugu rubripes* have been cloned and their molecular mechanism on muscle development and growth have been partially revealed. Unlike mammals, most of fish skeletal muscles grow dramatically during the post-larval life, and muscle growth determines the growth and size of the fish. Uncovering the molecular basis of muscle development and growth is an important step in aquaculture. To better understand the function of *Myf5* in largemouth muscle development and growth, the cDNA and genomic sequence are characterized. Using RT-PCR and RACE techniques, a 1093-bp *Myf5* cDNA from largemouth bass (*Micropterus salmoides*) was isolated. Its ORF was 723bp, encoding 240 amino acids, with a conserved basic Helix-Loop-Helix (bHLH) domain. It has no signal peptide and belonged to nucleic protein. The amino acid sequence of largemouth bass *Myf5* bHLH domain was exactly identical to that of other fishes and exhibited 80% or greater identity with that of other vertebrates such as African clawed frog, chickens, bovine, mice, and humans. It indicates that largemouth bass *Myf5* probably acted as a transcriptional factor to regulate muscle-specific genes, like its homologs in other species. Using Genomewalker technology, a 2 690-bp promoter sequence was isolated. By bioinformatics analysis, the basic transcriptional regulatory elements such as TATA Box, GC Box, CCAAT Box, muscle specific transcriptional factors such as E-box, MEF2, SP1, USF, SRF, MTBF binding sites and the early growth related transcriptional factors such as early growth response gene α (EGR α), Egr-1 early growth response gene, Egr-2 early growth response gene product, growth factor independence 1, TCF binding sites which may be responsible for the *Myf5* expression in developing somite were found. For many E-boxes in promoter, we predict that the other MRFs such as *MyoD* may be a negative regulator to the transcription of *Myf5*. MEF2 family was reported to promote the muscle-specific gene transcription together with the MRFs family. The MEF2 binding sites were found in largemouth bass *Myf5* promoter, and we predict MEF2, as a transcription factor, raise the expression of *Myf5* in largemouth bass. SP1, USF and SRF were all revealed to regulate the transcription of *MyoD*. So we predict that they may be the potential regulators to *Myf5*. Meanwhile, *Myf5* of largemouth bass consisted of three exons and two introns, a highly conserved 123-bp region in intron-1 of the different fishes which may regulate the temporal and tissue-specific expression of *Myf5*. Conclusion: cDNA and genomic sequence of largemouth bass *Myf5* are isolated and characterized, and the regulatory elements and potential transcriptional factors are predicted. It will help us to better understand the expression regulation of *Myf5* in fish and further study the different signal paths in complex myogenesis net. [Journal of Fishery Sciences of China, 2008, 15 (4) : 568-576]

Key words: muscle; *Myf5*; gene; regulation; largemouth bass (*Micropterus salmoides*)

Corresponding author: BAI Jun-jie. E-mail: jjbai@163.ent