

剑尾鱼 2 种卵黄蛋白原全长 cDNA 的克隆及序列分析

刘春¹, 李凯彬¹, 耿冬雨², 聂湘平³, 王芳¹, 吴淑勤¹

(1. 中国水产科学研究院珠江水产研究所, 广东 广州 510380; 2. 大连水产学院 生命科学学院, 辽宁 大连 116023; 3. 暨南大学 水生生物研究所, 广东 广州 510630)

摘要: 卵黄蛋白原是环境雌激素效应研究的重要生物标志物, 广泛应用于水环境内分泌干扰物污染的评价。本研究利用 RT-PCR 和 cDNA 末端快速扩增法(RACE) 克隆获得了剑尾鱼(*Xiphophorus helleri*) 的 2 种卵黄蛋白原基因—VgA 和 VgB, 并对其基因结构和系统进化进行分析。VgA 基因 cDNA 序列全长 5 159 bp, 开放阅读框(ORF) 5 058 bp, 编码 1 685 个氨基酸。VgB 基因 cDNA 序列全长 5 349 bp, 开放阅读框(ORF) 5 067 bp, 编码 1 688 个氨基酸。2 种 cDNA 序列均具有与已发现的卵黄蛋白原分子相似的主要特征和功能域, 包括 Vitellogenin 结构域、VWFD 结构域和多聚丝氨酸区域, 且与已知其他鱼类卵黄蛋白原基因有较高的同源性。[中国水产科学, 2010, 17(1): 31–43]

关键词: 剑尾鱼; 卵黄蛋白原; 克隆; 序列分析

中图分类号: S91

文献标识码: A

文章编号: 1005-8737-(2010)01-0031-13

卵黄蛋白原(Vitellogenin, Vg)是卵黄蛋白的前体, 它在卵生动物卵黄发生前, 由卵巢外组织分泌到血液, 通过血液循环运输到卵巢, 经修饰后成为卵生长发育的营养物质——卵黄蛋白。卵黄蛋白和卵黄蛋白原为正在发育的胚胎提供氨基酸、脂肪、碳水化合物、维生素、磷和硫等营养物质, 对于卵生动物胚胎发育有重要意义。在卵生脊椎动物体内, Vg 由雌激素刺激肝细胞而合成。正常环境中雌性动物在卵黄生成期可合成大量 Vg, 而在其他生活期及雄性动物体内均不合成 Vg。人工雌激素或低剂量的类雌激素可诱导雄性和幼年动物体内 Vg 的生成。根据这一特性, 可通过检测雄性和幼年卵生脊椎动物肝细胞 Vg 非正常表达, 来评价环境雌激素的作用^[1-2]。

剑尾鱼(*Xiphophorus helleri*)属鳉形目花鳉科剑尾鱼属, 有体型小、繁殖力强、易辨别雌雄以及便于在实验室饲养等特点。剑尾鱼的实验动物化研究历经多年, 是中国首个通过全国水产原良种审定委员

会审定的鱼类实验动物, 已应用于水环境监测、水产药物安全性评价、动物疾病检验模型等不同领域^[3]。本实验以剑尾鱼为实验材料, 通过 RT-PCR、RACE 等方法对 2 种剑尾鱼卵黄蛋白原全基因进行克隆和序列分析, 以期剑尾鱼生殖生理研究及剑尾鱼在水环境污染检测研究中的应用提供良好平台, 为进一步进行剑尾鱼实验动物化研究打下基础。

1 材料与方法

1.1 实验动物

剑尾鱼, 4~5 月龄, 为珠江水产研究所水生实验动物中心培育。

1.2 菌株与质粒

大肠杆菌 DH5 α 由本实验室保存。pGEM-T Easy Vector System 为 Promega 公司产品。

1.3 工具酶和试剂

TaKaRa RNA PCR Kit (AMV) Ver.3.0、TaKaRa

收稿日期: 2009-02-25; 修订日期: 2009-06-19.

基金项目: 科技部公益项目(2004DIB1J029).

作者简介: 刘春(1982-), 男, 硕士, 从事水产动物病害免疫方面的研究. E-mail: lc5173483@163.com

通讯作者: 吴淑勤. E-mail: wushuqin001@21cn.com

Agarose Gel DNA Purification Kit Ver.2.0、DNA 分子 Marker 为宝生物工程(大连)有限公司产品; SV Total RNA Isolation System 为 Promega 公司 产品; Plasmid Mini Kit I 为 Omega 公司产品; SMART™ RACE cDNA Amplification Kit 购自 Clontech 公司; 限制性内切酶、T4 连接酶为 NEB 公司产品; *Taq* DNA 聚合酶、琼脂糖、二甲基亚砜(DMSO)等相关试剂购自威佳科技有限公司; 17β- 雌二醇为 Sigma 公司产品; 其他试剂为国产分析纯试剂。

1.4 剑尾鱼卵黄蛋白原的诱导及总 RNA 的提取

剑尾鱼腹腔注射 17β- 雌二醇(50 μg/g 鱼体质量,以 DMSO 为溶剂),1 周后按照 SV Total RNA Isolation System 试剂盒推荐方法提取剑尾鱼肝脏总 RNA。

1.5 剑尾鱼卵黄蛋白原 cDNA 的克隆

以提取的剑尾鱼总 RNA 为模板,按照试剂盒 TaKaRa RNA PCR Kit (AMV) Ver.3.0 方法,合成 cDNA 第一链。

1.5.1 *VgA* 基因的 PCR 扩增 参照 GenBank 已登录鱼类 *VgA* cDNA 序列: 食蚊鱼(*Gambusia affinis*) (AB181835)、底鳉(*Fundulus heteroclitus*) (U07055)、青 鳉(*Oryzias latipes*) (AB064320)、斑 马 鱼(*Danio rerio*) (AF406784),设计并合成 4 对引物 PVgA1、PVgA2、PVgA3、PVgA4 (表 1)。以 cDNA 第 1 链为模板,进行 PCR 扩增,反应条件为: 94 ℃ 预变性 2 min; 经 94 ℃ 30 s,58 ℃ 45 s,72 ℃ 90 s,30 个循环; 72 ℃ 延伸 7 min。

表 1 核苷酸引物序列
Tab. 1 Oligonucleotide primer sequences

引物名称 Primer name	上游引物 Forward primer	下游引物 Reverse primer
PVgA1	5'-TGGARYTCAGGGWGTGTGC-3'	5'-CCTCTTKGCAATKYTCCTCA-3'
PVgA2	5'-ATGAAAGCGTTGTACTAGCCCTG-3'	5'-TAGCGACTGGTTTCAAGGTGTAGC-3'
PVgA3	5'-ATCCACTTCTGCGTGAGGTTGTC-3'	5'-GGTCCAACTTTAACCTCCATCTC-3'
PVgA4	5'-ATTGAAGGTGAAGTTGTTGAGAG-3'	5'-AGGCAGAATCCAGGAGTGAGC-3'
PVgB1	5'-GAGTATGCCAATGGTGTGAT-3'	5'-CCCGCATTACCCACAAC-3'
PVgB2	5'-TCTTGCTGACCCGCTTA-3'	5'-CCATCCTGACCTGCTTCT-3'
PVgB3	5'-CCACACACTTGGTGACAGC-3'	5'-GTGGAGAAATGGTGGCTTG-3'
PVgB4	5'-CCTGAAATCCGCACTTAT-3'	5'-GTTGGACAGGTATTTGGC-3'

5'RACE 扩增: 根据已获得的 *VgA* 基因 ORF 部分序列设计了 1 对下游引物,分别为 P1: 5'-CAGGAGCAAAGATTTTACCAACAACAC-3', P2: 5'-GTTGCTGGGATTGCTGACTCGTTTGG-3'; 上游引物使用 Smart™ RACE cDNA Amplification kit 的通用引物 UPM, PCR 扩增反应按照 Takara 试剂盒操作说明进行。

3'RACE 扩增: 根据已获得的 *VgA* 基因 ORF 部分序列设计 1 对上游引物,分别为 P3: 5'-TGGA TGAAAGGAAAGACCTGTGGACT-3', P4: 5'-CTT TGCTCACTCCTGGATTCTGCCT-3'; 下游引物为通用引物 Oligo (dT)₁₆: 5'-AAGCAGTGGTATCAACG

CAGAGTACGC (T)₁₆-3'。RT-PCR 反应按照 Takara 试剂盒操作说明进行。

1.5.2 *VgB* 基因的 PCR 扩增 参照 GenBank 已登录鱼类 *VgB* cDNA 序列: 食蚊鱼(AB181836)、底鳉(U07826)、青鳉(AB074891),根据它们的保守区分别设计合成 4 对引物 PVgB1、PVgB2、PVgB3、PVgB4 (表 1)。以 cDNA 第 1 链为模板,进行 PCR 扩增,反应条件同 1.5.1。

5'RACE 扩增: 根据已获得的 *VgB* 基因 ORF 部分序列设计了 1 条下游引物 P5: 5'-CATTCTGAGT CTCTTGATGTTCTGCTG-3'; 上游引物使用 Smart™ RACE cDNA Amplification kit 的通用引物 UPM, PCR

扩增反应按照 Takara 试剂盒操作说明进行。

3'RACE 扩增: 根据已获得的 *VgA* 基因 ORF 部分序列设计 1 对上游引物, 分别为 P6: 5'-CAGC ACCCATCAGGCAACATACATA-3', P7: 5'-GCCAG TGAGAACCACCTCCGTC-3'; 下游引物为通用引物 Oligo (dT)₁₆。RT-PCR 反应按照 Takara 试剂盒操作说明进行。

所有扩增产物经琼脂糖凝胶回收纯化, 与 T 载体连接, 转化感受态细胞大肠杆菌 DH5 α , 转化子用碱裂解法提取质粒, 用 *EcoR* I 和 *Hind* III 酶切鉴定插入片段大小, 挑取阳性转化子送上海英骏生物技术有限公司用 ABI PRISMTM 3730 全自动荧光测序仪完成序列测定。

1.6 序列拼接及序列结构分析

将扩增获得的 ORF 部分序列、5' 序列和 3' 序列用 DNASTAR 分析软件进行全长拼接。用软件 SignalP 3.0 Server 和 ExPASy 对拼接后的 cDNA 序列进行序列分析^[4], 内容包括核苷酸序列中 ORF 的

寻找, 编码氨基酸序列的推导, 信号肽及其保守结构域的预测。

1.7 不同动物卵黄蛋白原序列的同源性比较

采用 vector VTI 10.0 软件把获得的剑尾鱼卵黄蛋白原 cDNA 序列及编码氨基酸序列与 GenBank 中登录的其他动物卵黄蛋白原基因及编码氨基酸序列进行同源性分析。

2 结果与分析

2.1 RT-PCR 扩增结果及序列分析

2.1.1 剑尾鱼 *VgA* 基因 RT-PCR 扩增结果及序列分析 应用 4 组引物对剑尾鱼 *VgA* 部分序列进行 RT-PCR 扩增。电泳结果显示, 分别在 700 bp、1 100 bp、1 800 bp、1 700 bp 有明显的目的条带, 均与预期片段大小相符(图 1)。测序结果分别得到 4 个 686 bp、1 118 bp、1 846 bp、1 701 bp 的序列。BLAST 比对结果显示所得序列均与其他鱼类 *VgA* 基因具有较高的相似性。

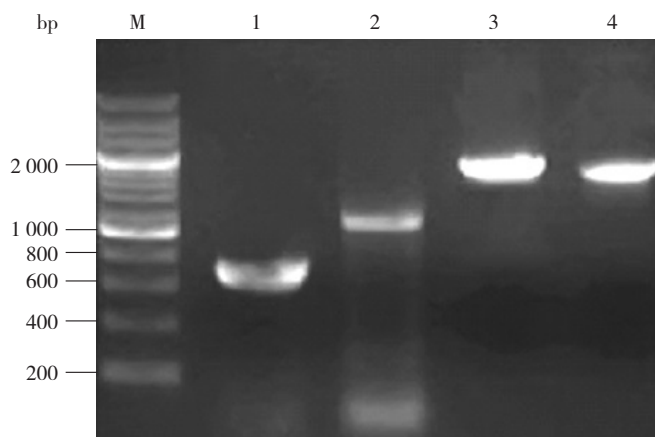


图 1 RT-PCR 扩增的 *VgA* 4 个 cDNA 片段

M: 200 bp DNA 分子量标记; 1-4: *VgA* RT-PCR 产物。

Fig. 1 Four cDNA fragments of *VgA* gene by RT-PCR

M. 200 bp DNA ladder marker; 1-4: RT-PCR products of *VgA*.

5'RACE 用通用引物 UPM 和引物 P1 进行 PCR 扩增, 在 400 bp 左右产生一明显的目的条带(图 2, 条带 1)。以该产物为模板, 以 UPM 和引物 P2 进行半巢式 PCR 扩增 5' 末端 cDNA 序列。电泳结果显示, 在 350 bp 左右得到单一特异条带(图 2, 条带 2)。测

序得到 346 bp 的序列。BLAST 比对结果显示该序列与其他鱼类卵黄蛋白原基因具有较高的相似性。

3'RACE 利用通用引物 oligo (dT)₁₆ 和引物 P3 进行 PCR 扩增, 产物在 500 bp 左右产生一明显目的条带(图 3, 条带 1), 以该产物为模板以 oligo (dT)₁₆

和引物 P4 进行半巢式 PCR 扩增,获得 400 bp 左右的片段(图 3,条带 2)。测序得到 389 bp 的序列。

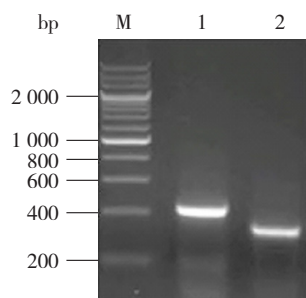


图 2 *VgA* 5' RACE 扩增结果
M: 200 bp DNA 分子量标记; 1, 2: *VgA* 5' RACE PCR 产物。

Fig. 2 5' RACE PCR product of *VgA*

M: 200 bp DNA ladder marker; 1, 2: 5' RACE PCR products of *VgA*.

将 3' 和 5' 端序列及 4 个部分序列用 DNASTar 软件拼接获得了 *VgA* 基因的 cDNA 全序列。经 ExPASy 软件分析,该序列全长 5 159 bp, 包含起始密码子 ATG、终止密码子 TAA 及转录终止信号 AATAAA。该基因的 5' 非编码区为 16 bp, 3' 非编码区为 85 bp, 开放阅读框(ORF)位于 17 ~ 5 074 bp, 长 5 058 bp, 共编码 1 685 个氨基酸,其中信号肽为 15 个氨基酸。根据推算,其编码蛋白分子量为 187.2 kD, 等电点为 9.29。该基因的核苷酸序列以及编码的氨基酸序列如图 4 所示。该编码氨基酸序列具有典型的卵黄蛋白原结构: 1 个 Vitellogenin 结构域(位于编码氨基酸肽链的 24 ~ 663 号氨基酸上)、1 个 VWFD 结构域(位于编码氨基酸肽链的 1423 ~ 1628 号氨基酸上)、1 个多聚丝氨酸区域(位于编码氨基酸肽链的 1072 ~ 1213 号氨基酸上)。另外还含有 1 个蛋白水解酶切位点(RSRR)。将该推导氨基酸序列与食蚊鱼、底鳉 *VgA* 氨基酸序列比对后发现,该氨基酸序列包含 LvH、Pv-LvL 复合物和 β' 组分 3 个部分。

2.1.2 剑尾鱼 *VgB* 基因 RT-PCR 扩增结果及序列分析 应用 4 对引物对剑尾鱼 *VgB* 部分序列进行 RT-PCR 扩增,电泳结果显示,分别在 1 100 bp、1 400 bp、1 300 bp、1 400 bp 有明显的扩增条带,均与预期的扩增片段大小相符(图 5)。测序结果分别

BLAST 比对结果显示该序列与其他鱼类卵黄蛋白原基因具有较高的相似性。

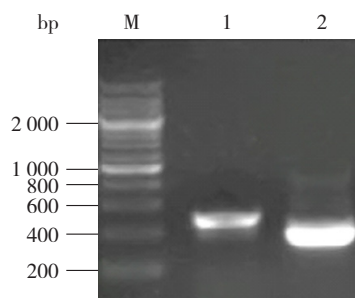


图 3 *VgA* 3' RACE 扩增结果
M: 200 bp DNA 分子量标记; 1, 2: *VgA* 3' RACE PCR 产物。

Fig. 3 3' RACE PCR product of *VgA*

M: 200 bp DNA ladder marker; 1, 2: 3' RACE PCR products of *VgA*.

得到 1 136 bp、1 418 bp、1 260 bp、1 380 bp 的序列。BLAST 比对结果显示所得序列均与其他鱼类 *VgB* 基因具有较高的相似性。

用通用引物 UPM 和引物 P5 进行 PCR 扩增,电泳结果显示,在 500 bp 左右产生一明显的目的条带(图 6,条带 1),测序得到 493 bp 的序列。BLAST 比对结果显示该序列与其他鱼类 *VgB* 基因具有较高的相似性。

利用通用引物 Oligo (dT)₁₆ 和引物 P6 进行 PCR 扩增,产物在 800 bp 左右产生一明显目的条带(图 7,条带 1),用该产物为模板以 Oligo (dT)₁₆ 和引物 P7 进行半巢式 PCR 扩增,获得 400 bp 大小的片段(图 7,条带 2)。长片段测序得到 817 bp 的序列。BLAST 比对结果显示该序列与其他鱼类 *VgB* 基因具有较高的相似性。

将 3' 和 5' 端序列及 4 个部分序列用 DNASTar 软件拼接获得了 *VgB* 基因的 cDNA 全序列。经 ExPASy 软件分析,该序列全长 5 349 bp, 包含起始密码子 ATG、终止密码子 TAA 及转录终止信号 AATAAA,该基因的 5' 非编码区为 14 bp, 3' 非编码区为 268 bp, 开放阅读框(ORF)位于 15 ~ 5 081 bp, 长 5 067 bp, 共编码 1 688 个氨基酸,其中信号肽为 15 个氨基酸。根据推算,其编码蛋白分子量为 186.8 kD, 等电点为 9.07。该基因的核苷酸序列以及编码的氨基酸序列

1	信号肽 Signal peptide	LvH	Vitellogenin 结构域
	M K A V V L A L T L A F V A G	Q H F D L V	P E F A A G K T Y
17	ATGAAAGCGGTGTACTAGCTCTGACTCTGGCCTTCGTGGCTGGACAACATTTTGATTGGTCCCTGAATTTGCTGCCGTAAGACCTAC		
	Vitellogenin domain profile		
	V Y K Y E A L L L S G L P E E G L A R A G L K I R S K L L F		
107	GTGTACAAGTATGAAGCATTACTCCTCAGCGGTCTTCTGAGGAAGGTTTGGCAAGAGCTGGACTGAAAATCAGGTCCAACTTCTCTTC		
	G R A D Q N T L M L K L V E P E L F E Y N G I W P N D S A I		
197	GGCCGAGCTGACCAAACTTAAATGCTGAAGCTTGTTGAACCTGAGCTCTTTGAGTACAATGGCATTGGCCAAACGACTCAGCAATC		
	P A T K L T A A L A P Q F A M P I K F E Y T N S V V G K I F		
287	CCAGCAACCAAGTTGACGGCAGCCCTGGCACCTCAGTTTGCATGCCATCAAGTTCGAATACACCAATAGTGTGTGTGGTAAAACTTTT		
	A P E G I S P V V L N I H R G I L N V L Q L N I K K T H K V		
377	GCTCTGAAGGATTTGCGCTGTGGTGTGAACATCCACAGAGGCATCCTGAATGTTCTCCAGCTGAACATCAAGAAGACCCACAAAGTC		
	F D L Q E V G T Q G V C K T L Y S I S E D A R N E N I L L T		
467	TTTGACTTGCGAGAGGTTGGAACCTCAGGGTGTGTGAAGACCCCTACTCCATCAGTGAAGATGCACGTAATGAGAATCCTCTCTGACC		
	K T R D L N N C Q E R L I K D M G L A Y T E K C E K C Q E E		
557	AAGACCAGGGACCTGAACAACTGCCAGGAAAGACTCATTAAAGACATGGGATTGGCATACTGAGAAGTGTGAAAAGTGCAGGAGGAA		
	T K N L R G T T T F S Y T L K P V A N A I M I L K V D V N E		
647	ACCAAAAACCTGAGAGGTACCACAACATTAGCTACACCTTGAAACCAAGTCGTAATGCCATCATGATCCTGAAAGTGGACGTTAATGAG		
	L I Q F L P F S E D N G A T Q M R T K Q S F E F L E I Q K D		
737	CTGATCCAGTTCTTACCTTTCTCTGAGGATAATGGAGTACTCAAATGAGGACCAAGCAGTCCTTTGAGTTCCTTGAGATTGAGAAAGAC		
	P I T T P I N A V Y R H R G S L K Y E F S N E L L Q T P I K L		
827	CCCATTACACCTATCAACGCTGTATATAGACACCGTGGATCTCTCAAGTACGAGTTCTCCAACGAAGTCTTTCAGACACCCATTAAGCTT		
	V K I S N A K A Q T A E V M N Q L A T I N V E N L H E N A P		
917	GTCAAGATCAGCAATGCAAAGGCCAGACTGCAGAGGTCATGAATCAGCTTGCCACCATCAACGTGGAGAATCTTCATGAAAAATGCACCT		
	M K F L E L V Q L L R L A R Y E D L E M Y W N Q Y K K M S P		
1007	ATGAAGTTTGGAACTGGTACAGTCTCCGCTCTGCTCGCTAAGATTTGGAATGTACTGGAACCAAGTACAAAAAGATGTCCTCT		
	H R H W F L D T I P A A G T P A A F R F I K E K F M A E E I		
1097	CACAGACACTGGTTCTTGGATACCATCCCTGCTGCTGGCACTCCCGCTGCTTTTAGATTATCAAAAGAGAAGTTTATGGCTGAGGAAATA		
	N T A E A V Q A L V A A V H M V T A D P E V I K L F E N L L		
1187	AACACTGCTGAGGCTGTTCAGCATTGGTTGCAGCTGTGCACATGGTGACTGTGACCCGAGGTTATCAAGCTGTTGAGAACTGCTTGA		
	A S D K V E R N P L L R E I V F L G Y G T M V Y K Y C N E T		
1277	GCGAGCGACAAAGTAGAGAGAAATCCACTTCTGCGTGAGATTGTCTTCTTGGATACGGTACAATGGTTTACAATATTGTAATGAGACA		
	A A C P A E L I K P I Q E R L S N A V A K N E E N I I L Y		
1367	GCAGCTGCTCCAGCTGAAGTTATAAAGCCTATTCAAGAAGCACTTTCAAAATGCTGTTGCCAAGAATGAGGAAGAAAACATCATCTGTAC		
	I K V L G N A G H P S S F K G V S E I M P I H G T A A A S L		
1457	ATAAAGGTTTGGGAAATGCTGGACATCCATCTAGCTTCAAGGGGGTGAGTGAGATCATGCCATCATGGCACTGCCGCTGCATCTCTG		
	P M R I H V E A I M A L R N I A K K E P R L V Q D L V L E L		
1547	CCAATGAGAATTCATGTTGAAGCCATCATGGCTTTGAGGAACATTCGAAAGAAGGAGCCAGACTGGTCCAGGATCTGGTTCTCGAGCTC		
	Y M D K V L H P K L R M L S C I V L F E T S P S M G L V A T		
1637	TACATGGACAAGTTCTCCACCCAAAGCTCCGTATGCTTTCCTGCATTGTTCTCTTCGAGACAAGTCCTTCCATGGGTTTGGTGGCCACT		
	V A N S V K T E E N L Q V A S F T Y S H M K S L S R N H A T		
1727	GTTGCCAACTCTGTGAAAACAGAGGAGAATTTGCAAGTGGCCAGCTTCACTTACTCTACATGAAGTCCTGAGTAGGAACCATGCAACC		
	I D P D V A A A C N V A M K L L S P K L D R L S L R Y S K A		
1817	ATTGATCTGATGTTGCCGCTGCTTGCAACGTTGCCATGAAACTATTGAGCCCAAGCTGGACAGACTGAGCTGCGTTACAGCAAGACT		
	V H V D I Y K N S L M V G A A A T A F Y I N D A A N I L P K		
1907	GTTTATGTTGACATCTACAAAACTCCTTGATGGTTGGTGTGCTGCAACTGCTTTTACATCAATGATGTGCAAAATATTTGCCAAAA		
	T A V A A K T R A F L A G A T A E A L E I G A S I D G L Q E L		
1997	ACTGTGTTGCAAGACAGAGCCTTCCTCGCTGGAGCAACTGCAGAAGCCCTGGAGATCGGAGCCTCTATTGATGGACTGCAGGAGCTG		
	L L K N P A L A E N T D R I T K M K R V I K A L S E W R S L		
2087	CTTTTGAAAAACCTGCTCTCGCTGAAAACACTGACAGGATCAACCAAAATGAAGCGGTCATAAAGCCTCTGTCAGAAATGGAGATCCCTG		
	P A D K P L A S I Y V K L F G Q E V A F A S I E K P M I D Q		
2177	CCCGTGACAAACCCCTGGCTTCTATCTATGTCAAGCTCTTTGGACAGGAGGTGCGCTTTGCTAGTATCGAGAAACCCATGATCGACCAG		
	A V K Y A K D L P I Q E Y G R E I L K A L L V S G I N F K Y		
2267	GCTGTTAAGTATGCCAAGGACTTACCCATTCAAGATATGGAAGAGAGATTCTTAAGGCTCTGCTCGTGTCTGGTATCAACTTCAAGTAT		
	A K P V L A A E V R R I L P T V A G L P M E L G L F S A A V		
2357	GCTAAGCCTGTGCTGGCTGCTGAGGTGAGACGATTTCCCAACTGTCGCTGGTCTCCCGATGGAGCTTGGTCTGTTCAAGTGTGCTGTG		
	A A A S V K I K P N T S P R L P E D F P L Q K L L E T D I Q		
2447	GCTGCAGCTTCTGTAAAAATCAAGCCAAACATCACCACGCTGCCAGAAGATTCCCCCTCCAGAAGCTCTGGAACAGACATTACAG		
	L E A E V R P A V A M T T Y A V M G L N T D V F Q A F V I A		
2537	CTCGAGGCTGAGGTCAGACAGCTGTTGCCATGACCACATATGCAGTTATGGGACTCAACACTGACGCTTTTCCAGGCTTTTGTGATAGCA		
	N A N V H S V M P A K I A A R L N I K E G D F K E A L P V		
2627	AACGCCAAGCTCCACTCTGTTATGCCAGCCAAATTTGTCGAAGACTCAACATCAAGAAGGAGACTTTAAGCTTGAAGCTTCCAGTT		
	E V P E N I T F V N V T T F A G S R N I E E L S K E R I T Y		
2717	GAAGTGCCTGAAAACATCATTGTTGAATGTGACAACCTTTGCTGGGTCAAGAAACATCGAGGAACCTTTCTAAAGAGAGAATTACCTAC		
	L F P K K L L R P S S V K S S E R A S Y A D S M L S S S E L		

2807	CTCTCCCTAAAAACTCTTGCGCCCAAGCTCAGTCAAATCATCCGAGCGAGCTTCTATGTGACAGCATGCTCTCATCTCCGAGCTC
	L P E E A K Y V K A I R Q H K V R A F A K R Y C A K N K G V
2897	CTTCTGAGGAAGCAAAATATGTCAAGGCCATTTCGTCACATAAGGTTTCGAGCATTGCGAAGAGGTACTGTGTAAGAACAAGGTGTT
	G L R A C F R F A S E N G A Y I Q D G L L Y K L V G R H N F
2987	GGACTTAGGGCTGTTTCAGGTTTGCCAGTGAAATGGTGCCTACATCCAAGATGGACTCTTGTACAACTGGTTGGCCGCCACAACCTTC
	S F S V T P I E G E V V E R L E M E V K V G P N A A E K L V
3077	TCTTCTCTGTGACCAATTGAAGTGAAGTGGTTGAGAGATTGGAGATGGAGGTTAAAGTTGGACCAATGCTGCAGAGAAGCTTGTT
	K R I N L N E E E T T E E R G P V L M K L
	Pv-LvL 多聚丝氨酸区域
3167	AAGCGCATCAACCTGAACGAAGAGGAAACACAGAGGAAGAGGACCAGTTCTGATGAAGCTCAGCAAAATCCTGTCTTCAAAAAAGAAC
	Polyserine domain
	S S S S S S S S S S S S S S S S S R S S K S S S S S S S S R S S R
3257	AGCTCCTCATCTTCTCTCTAGCTCAAGCAGCTCTTCATCCAGCCGTTCTCAAAGAGCTCTCCAGCTCTTCATCTCGCTCCAGCCGT
	K I S L A A R A D N S N S S S S S S S S S S S S S R S S [R S R B]
3347	AAGATCAGTCTTGACGCCGTGCGGACAACAGCAACAGCAGCAGCAGCAGTAGCAGCAGCAGCAGCAGAGAAGCAGCAGAAGCAGGAGA
	S S R R S K S G N L S R S S S S S S S S S S S D R T S S A S K V L S D P L F
3437	AGCAGCCGAGAAGCAAGAGCGGCAACTTATCAAGATCCAGCAGCAGTTCAGATAGGACCAGCTTGCATCAAGCATCGCACCTCTCTTC
	S G S S S S S S R S S S S S S R S K Q V T E K F Q R L H N K E N P
3527	AGCGCAGTTCAAGCTCTTCTCGTTCCAGCTCTCTCGCTCTAAGCAAGTGACTGAGAAGTTCCAGAGGTTACACAATAAGGAGAAATCCC
	S K S T S R S R S T S S S F E A I Y N Q K K F L R E E E I V
3617	TCCAAATCAACTTCCAGAAGCAGGAGCACTTCTCCAGCTTGGAGCATCTATAATCAGAAAAATTCCTCAGGGAAGAGGAAATGTT
	V A L I V R A V K V D K R M L G Y Q L A A Y L D K P N R L
3707	GTGGCATTGATCGTCCGTGCTCAAAAGTTGACAAGAGGATGTGGGATACCAACTTGCTGCCTACCTTGACAAACCAATGCCAGACTT
	Q I I V S N L S A D S N W R L C A D G V V L S K H K V T T R
3797	CAGATCATTGTCTCAACCTGTCTGCTGATAGCAACTGGAGGCTCTGTGCTGATGGAGTGTGTTGAGCAAGCACAAGTTACAACTAGG
	V A W G E Q C K R Y S A N A T G E T G L V S S S P A A R L R
3887	GTTGCCTGGGCGAGCAATGCAAGAGATATAGCGCAATGCTACAGGAGAGACTGGTCTTGTTCCTCAAGCCGCGAGCTCGCTCAGA
	A S W E R L P S A L K R Y G K M V N K Y V H S K V L S D L I
3977	GCGTCTGGGAAAGACTGCCTTCTGCCCTGAAACGTTATGGAAAGATGGTCAACAAATATGTCCATTCCAAAGTACTGTCAGACTTGATT
	H T K K K N S T R K I S V I A V A T S D R T I D F I V K T P
4067	CACACAAGAAAAAGAACAGCACCAGGAAAAATCTCAGTCATTGCAGTTGCAACTTTCAGCAGGACAATGACTTCATTGTGAAGACTCCA
	M K S V Y N V T V R L P M S M P I D E I K G L T P F D E V I
4157	ATGAAATCTGTCTAATGTACTGTGCGCCTTCCCATGTCTATGCCCATTCAGCAGATCAAGGTCTCACCCCATTTGATGAAGTCATT
	β' VWFD 结构域 VWFD domain profile
	D K I H F M [V S K A A A] A E C S F F E D T L Y T F N N K S Y
4247	GACAAGATCCACTTCATGGTTTCAAAGGCTGCCGAGCTGAATGCAGCTTCTTTGAAGACACTCTACACATTCAACAACAAGAGCTAC
	K N K M P S S C Y Q V A A Q D C T D E L K F M V L L R K D S
4337	AAGAACAAGATGCCTTCTCTTGTCTACCAGGTGCGGCACAGGACTGCATGATGAAGTGAAGTTTATGGTTCTCTGAGGAAAGATTCT
	S E Q H H I N V K I S E I D I D M Y P R D N N I I V G N S E
4427	TCAGAGCAACACCATATCAATGTCAAAATTTCTGAGATTGACATCGACATGTATCCAAGGGACAACAACATCTTGTGAAGGTCAACGAA
	M E I P L T S L P Y R H P T A S I E I R Q S G E G L A V F A
4517	ATGGAAATCCCTCACCAGCCTGCCTTACCGCCACCAACAGCTTCCATTGAGATCAGACAGAGTGGAGAGGGCTTGCTGTATTGCA
	P K H G L Q E V Y F D K K T W R I K I A D W M K G K T C G L
4607	CCTAATATGGTCTCCAAGAAGTCTACTTTGACAAAAAGACATGGAGGATCAAAATTGCTGACTGGATGAAAGGAAAGACCTTGAGACTT
	C G K A D G E I R Q E Y H T P N G R V A K N S V S F A H S W
4697	TGTGAAAAGGCTGATGGAGAAATCAGACAGGAGTACCACACTCCCAACGGACGCGTGGCTAAGAAGTCAAGTCAAGCTTGGCTCCTCTGG
	I L P V E S C R D A S E C R L K L E S V Q L E K Q L T V H G
4787	ATTCTGCCTGTGAAAGCTCAGGGATGCCTCTGAGTGCCGTCTGAAGCTTGAATCTGTGCAGCTGAAAAACAATTGACCGTTTCATGGT
	D D S T C Y S V E P V P R C L P G C L P I K T T P V T V G F
4877	GACGATTCACATGCTACTCTGTTGAGCCTGTACCTCGCTGTCTGCTGGATGCTTGCCAATCAAGACCAACCCCGTCACTGTTGGTTTC
	N C W S S D S E S N V Y D R T S V D L R K T T Q A H L A C N C
4967	AACTGCTGGTCATCTGATTCTGAGCAATGTCTATGACAGAAGTGTGGACCTGAGAAAGACTACCAAGCTCACCTGGCTTGAACCTGC
	N T K C S *
5057	AACACCAAGTGCTCTTAACCTACATTTCTTTAAGTCACTATGTGTAATTTTTTCTGTAACAATAAAATAAAGTGCATCTTAAAAA
5147	AAAAAAAAAAAAA

图4 剑尾鱼 *VgA* 全长 cDNA 序列及推导出的氨基酸序列

水平下划线标记信号肽及各个卵黄蛋白组分的起始位置;方框标记蛋白水解酶切位点 RSRR; *Vitellogenin* 结构域、多聚丝氨酸区域、VWFD 结构域、加尾信号 AATAAA 用阴影标出。

Fig. 4 The cDNA and deduced amino acid sequence of *Xiphophorus helleri*'s *VgA* gene

Horizontal lines under the sequences indicate signal peptide and predicted sites for cleavage of vitellogenin into yolk proteins. The putative cleavage site of subtilisin-like convertase was boxed. Vitellogenin domain profile, polyserine domain, VWFD domain profile and polyadenylation site(AATAAA) are indicated by shade.

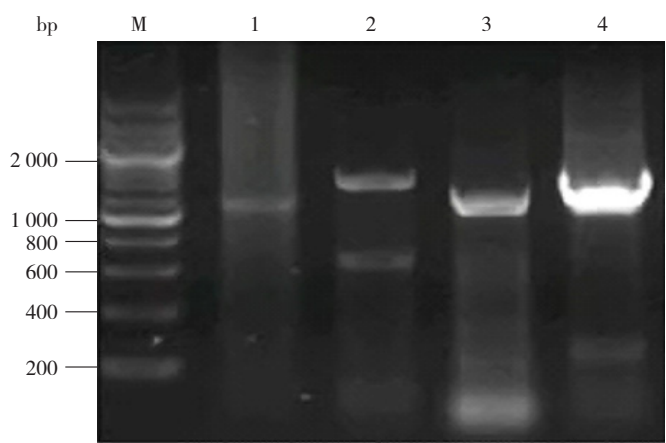


图 5 RT-PCR 扩增的 *VgB* 4 个 cDNA 片段
M: 200 bp DNA 分子量标记; 1-4: *VgB* RT-PCR 产物 .
Fig. 5 Four cDNA fragments of *VgB* by RT-PCR
M: 200 bp DNA ladder marker; 1-4: RT-PCR products of *VgB*.

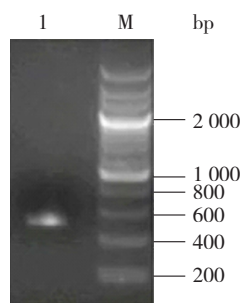


图 6 *VgB* 5' RACE 扩增结果
M: 200 bp DNA 分子量标记; 1: *VgB* 5' RACE PCR 产物 .
Fig. 6 5' RACE PCR product of *VgB*
M: 200 bp DNA ladder marker; 1: 5' RACE PCR product of *VgB*.

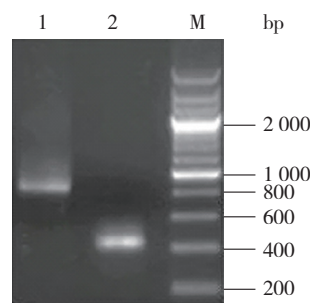


图 7 *VgB* 3' RACE 扩增结果
M. 200 bp DNA 分子量标记 1, 2: *VgB* 3' RACE PCR 产物.
Fig. 7 3' RACE PCR product of *VgB*
M. 200 bp DNA ladder marker; 1, 2: 3' RACE PCR products of *VgB*.

如图 8 所示。该编码氨基酸序列具有典型的卵黄蛋白原结构: 1 个 Vitellogenin 结构域(位于编码氨基酸肽链的 24 ~ 662 号氨基酸上)、1 个 VWFD 结构域(位于编码氨基酸肽链的 1418 ~ 1624 号氨基酸)、1 个多聚丝氨酸区域(位于编码氨基酸肽链的 1079 ~ 1207 号氨基酸)。与亲缘关系最近的食蚊鱼、底鳉 *VgB* 氨基酸序列比对后发现,该氨基酸序列包含 LvH、Pv 组分。

2.2 与其他动物卵黄蛋白原序列的同源性比较

2.2.1 不同动物 *VgA* 序列的同源性比较 将剑尾鱼 *VgA* 基因 ORF 及编码氨基酸序列与其他动物 *VgA* 基因 ORF 及编码氨基酸序列进行同源性比较,结果如表 2 所示。

与同属于花鳉科的食蚊鱼(*Gambusia affinis*)的同源性最高,核苷酸同源性为 95%,氨基酸同源性为 90%;与底鳉(*Fundulus heteroclitus*)、花鳉(*Rivulus marmoratus*)、青鳉(*Oryzias latipes*)等同属鳉形目的鱼类的同源性较高,核苷酸同源性为 71% ~ 84%,氨基酸同源性为 62% ~ 78%;与青沙鲛(*Sillago japonica*)、真鲷(*Pagrus major*)、条斑星鲷(*Verasper moseri*)、斑马鱼(*Danio rerio*)等不同目鱼类核苷酸同源性为 59% ~ 72%,氨基酸同源性为 39% ~ 61%;与鸡(*Gallus gallus*)、凡纳滨对虾(*Litopenaeus vannamei*)的核苷酸同源性为 33% ~ 51%,氨基酸同源性为 12% ~ 32%。比较结果表明,亲缘关系越近, *VgA* 的同源性越增高,同时也反映了这些物种的系统分类地位。

1 信号肽 Signal peptide LvH Vitellogenin 结构域 gacaaccaccaacc

15 ATGAAGGTGATTGTGCTGGCTCTGACTGTAGCCTTTGTGGCCAGCACCAGGTTAACTATGCCCCAGAAATTGCAGCTGGAAAGACCTAC

Vitellogenin domain profile

E Y K Y E G Y I L G G L P E E G L A R A G V K I Q S K V L I

105 GAGTACAAATATGAAGGATATATTCTGGGAGGGCTGCCTGAGGAGGCGCTGGCAAGGGCCGGGGTGAAAAATCCAGAGCAAAGTTTGTATT

G A A N P D N Y I L K L E D P V I S E Y S G I W P K E V F H

195 GGTGCAGCAAATCTGACAACTACATTCTGAACTTGAAGACCTGTGATCTCAGAGTACAGTGGTATCTGGCCAAAGAGGTTTCCAC

P A T K L T S A L S A Q L L T P V K F E Y A N G V I D K V Y

285 CCTGCCACAAAGCTCACCTCAGCTCTTCTGCTCAGCTCTTGACACCCGTCAGTTTGAGTATGCCAATGGTGTGATCGATAAAGTGAT

A P V G V S T T V L N I I R G I L N I F Q T N I K K T Q N V

375 GCGGCGTGGGCGTCTCTACAACTGTGCTGAATATCATCAGGGGAATCCTCAACATCTTTCAGACGAACATCAAGAAGACTCAGAATGTG

Y Y L Q E T G V K G V C K T Y Y V L S E D P K A D R L H L T

465 TATTACCTGCAAGAGACTGGAGTAAAGGTGTGTGCAAGACATACTATGTCTTAGTGAGGATCCAAAGGCTGACCGCTTCATTGTGACA

K T T D L N Q C S E K I Y M N V G M A G Y T D K C A E C E T

555 AAGACACCAACTGAATCAGTGTCTCAGAAAAATCTACATGAATGTTGGCATGGCTGGTTATACAGACAAATGTGCAGAGTGTGAGACT

R G E M F S G A M S L N Y V M K P T A S G S L I L E A T A T

645 CGGGGGGAAATGTTTTCAGGAGCAATGTCTCTAACTACGTCATGAAGCCACAGCCTCTGGCTCCTTGATCTTAGAGGCCACTGCCACT

E L L Q Y S P F N I V N G A V Q M E A K Q N I T F V D V K K

735 GAGCTCTCCAGTACTCCCTTTCAACATTGTGAACGGAGCTGTCCAGATGGAGGCAAAGCAGAATATTACCTTTGTGTGTCAAGAAG

T P L E P I K A D Y V P R G S L K Y E L G T E L L Q T P I Q

825 ACCCATTAGAGCTATCAAAGCTGACTATGTTCCCGTGGATCACTTAAGTACGAGTTAGGTACTGAACTCTCCAGACCAATTCAG

L L R I T N A E A Q I V E T L N N L V S L N L G K A H E D S

915 CTTCTGAGGATCACCAATGCTGAGGCTCAGATTGTTGAGACTCTGAACAACCTTGTGAGCCTCAATTTGGGCAAGGCCATGAAGATTCT

P L K F I E L I Q L L R V A R Y E S I E A L W G Q F K S K S

1005 CCCCTGAAGTTTATTGAGCTTATCCAGCTTCTGCGTGTGGCCAGATATGAAAGCATTGAAGCTCTCTGGGGCAATTCAAATCCAAAAGC

D H R H W M L S S I P A I G T H V A L K F I K E K F V A G E

1095 GATCATAGGCACTGGATGCTGAGCTCCCTGCCATTGGTACACATGTTGCTCTTAAGTTATCAAGGAGAAGTTGTGTGCTGGTGAA

V T A A E A A Q A I M S S T H L V T A D L E S I N L L E A L

1185 GTGACCGTGTGAAGTGTCTAGGCCATCATGTCTACACACTTGGTGACAGCTGACCTGGAGTCCATCAATCTTTGGAGGCCCTTA

A K T P K I Q E N A A L R E I T M L G F G S M I H K Y C A E

1275 GCCAAGACCCCTAAGATTCAAGAAAATGCTGCTTTGCGTGAGATCACCATGCTGGGCTTTGGCAGCATGATTACAAATACTGTGCGGAA

N P C T S E L L R P I H D I I T S A L E K R D N D E L S L A

1365 AACCATGTACATCTGAGCTTCTCAGGCTATCCATGACATTATTACAGTGCTCTTGAGAAACGTGACAATGTGAGCTCTCCCTGGCT

L K V L G N V G H P S S L K P I M K L L P G F G S S A S N L

1455 CTCAAAGTTCTGGGTAATGTGGGACATCTAGCAGCCTGAAGCCGATCATGAAGCTGTGCTGGTCTTTGGCAGCTCTGCCTCCAATCTT

E H R V H V D A A M A L R K I A K R E P R M I Q D V A I Q L

1545 GAGCACAGAGTTTATGTTGATGCCGTATGGCACTGAGAAAAATGCAAGAGAGAACCCAGGATGATTACAGGACGTGGCTATTACAGCTC

F M D S A L D Q E L R I I A A I V L F D T K L P L V M V T T

1635 TTCATGGACAGTGCTCTTGACCAAGAGCTCCGCATAATTGCTGCTATTGTGCTGTTGATACCAAGCTGCCCTTGGTTATTGGTGACCACT

V A Q S L L N E P D L Q V A S F V Y S Y M K A F T R T T T P

1725 GTAGCTCAGAGCCTCCTGAATGAACCGACCTGCAGGTGTGCTAGCTTTGTCTACTCTTACATGAAGGCTTTACAGGAGACCACCCCT

D H S S V A A A C N V A V R M L S P R L E R L S Y R Y S R A

1815 GACCATTCTCTGTAGTGTGCTGCAATGTTGCTGTGAGGATGCTCAGCCCTAGATTAGAGAGACTAAGCTACCGCTACAGTCGAGCT

F H Y D H Y Y N P W M M G A A S A F Y I N D A A T V L P R

1905 TTCCATTGACCACTATTACAATCCTTGGATGATGGAGCTGTGCGCAGCGCATTTACATCAATGATGTGCACTGTGTGCCAAGA

N V M A K A R V Y L S G V Y V D V L E F G A R T E G T V Q E A

1995 AACGTCATGGCAAAAGCTCGAGTTTACCTGTCTGGAGTGTATGTTGACGTTCTGGAGTTGCGAGCCAGAACTGAGGAGTACAGGAGGCC

L L K V R D V P G D A D R I T K M K Q A L K A L I E W R A N

2085 CTTTGAAGTCCGTGATGTTCCAGGTGATGCAGACAGGATCACCAAGATGAAGCAAGCTCAAGGCTTTAATTGAGTGGAGGCGCAAT

P S R Q P L G S L Y V K V L G Q E V A F A N I D K E M V E K

2175 CCTTCCCGCCAGCCCTTGGCTCTCTGTACGTGAAAGTTCTTGACAGGAAGTCGCATTTGCAAAACATTGACAGAAATGGTTGAAAAG

I I E F A T G P E I R T Y G K K A L E S L L S G Y S K Y S

2265 ATCATTGAGTTTGAACCGGACCTGAAATCCGCACCTTATGGCAAAAAGGCCCTTAGAGTCTCTGCTGTCTGGTTACTCTGTGAAATACTCC

K P L L A I E V R H I F P S S L G L P M E L G L Y T A A V A

2355 AAGCCACTGCTGGCTATTGAGGTCCGTACATCTTCCCGAGCTCTCTGGTTTACCATGGAGCTCGGTCTCTACACTGCTGCTGTGGCA

A A S V E V Q A T V S P P P P E N F H P V H L L R S D V S L

2445 GCCGCTTCTGTTGAAGTACAAGCCACCGTTTCTCCACACCTCTGAGAACTTCCATCTGTCCACCTCTTGAGGTCTGATGTTCCCTA

K A S V A P S V S L H T Y A V M G V N N P F I Q A T V M S R

2535 AAGGCTTCAGTCGCTCCGAGTGTGCTTTGACACACCTATGTGTTATGGGAGTAAATAATCCTTTCATTACAGGCTACTGTGATGTCAAGA

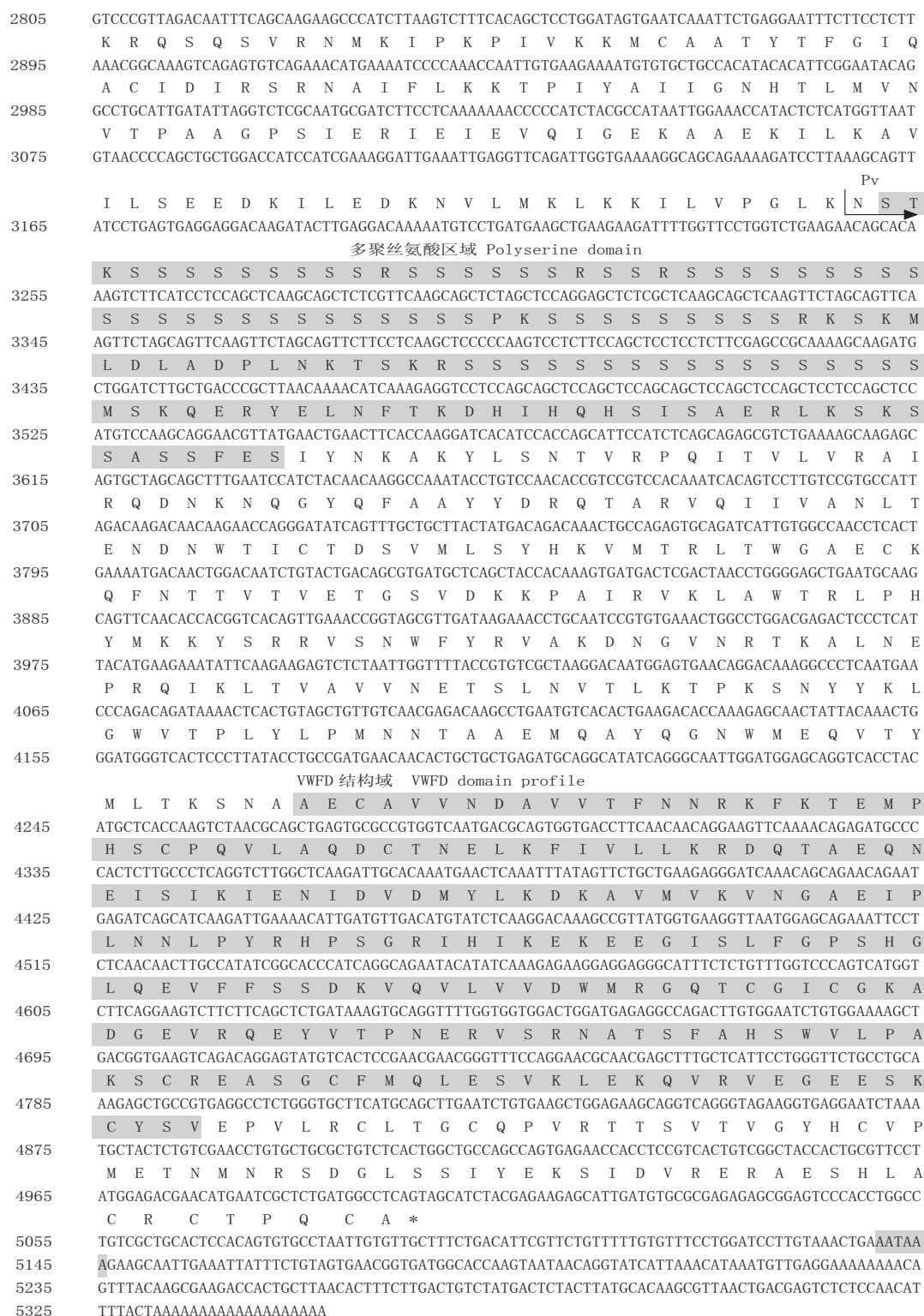
A K V H T V I P K K M E A R L D I E K G Y F N Y Q I L P I E

2625 GCCAAAGTCCACACAGTTATTCCTAAAAAGATGGAGGCGAGACTTGACATAGAAAAGGGCTACTTAACTATCAGATCTGCCTATTGAG

G V K T I A S A R L E T V A I A R D V E D L A A A K V T P V

2715 GGTGTGAAAACATCGCATCTGCTGCTTGAACAGTTGCCATTGCAAGAGATGTGGAAGACCTCGCAGCTGCCAAGGTACACACAGTT

V P L D N F S K K P I L S L S Q L L D S E S N S E E F L P L

图 8 剑尾鱼 *VgB* 全长 cDNA 序列及推导的氨基酸序列

水平下划线标记信号肽及各个卵黄蛋白组分的起始位置; *Vitellogenin* 结构域、多聚丝氨酸区域、VWFD 结构域、转录终止信号 AATAAA 用阴影标出。

Fig. 8 The cDNA and deduced amino acid sequence of *Xiphophorus helleri*'s *VgB* gene

Horizontal lines under the sequences indicate signal peptide and predicted sites for cleavage of vitellogenins into yolk proteins. Vitellogenin domain profile, polyserine domain, VWFD domain profile and polyadenylation site (AATAAA) are indicated by shade.

表 2 剑尾鱼与其他动物 *VgA* 基因 ORF 的同源性比较
Tab. 2 Homology comparison between *Xiphophorus helleri*'s *VgA* ORF and those of other animals

物种 Species	核苷酸同源性 /% Nucleotide homology	氨基酸同源性 /% Amino acid homology	编码多肽的长度 /aa Length of peptide
剑尾鱼 <i>Xiphophorus helleri</i>	100	100	1685
食蚊鱼 <i>Gambusia affinis</i>	95	90	1695
底鳉 <i>Fundulus heteroclitus</i>	84	78	1704
花鳉 <i>Rivulus marmoratus</i>	78	66	1711
青鳉 <i>Oryzias latipes</i>	71	62	1678
青沙鲛 <i>Sillago japonica</i>	71	61	1701
真鲷 <i>Pagrus major</i>	72	61	1715
条斑星鲷 <i>Verasper moseri</i>	69	58	1628
斑马鱼 <i>Danio rerio</i>	59	39	1362
鸡 <i>Gallus gallus</i>	51	32	1850
凡纳滨对虾 <i>Litopenaeus vannamei</i>	33	12	2587

2.2.2 不同鱼类 *VgB* 序列的同源性比较 将剑尾鱼 *VgB* 基因 ORF 及编码氨基酸序列与其他硬骨鱼 *VgB* 基因 ORF 及编码氨基酸序列进行同源性比较, 结果如表 3 所示。

表 3 剑尾鱼与其他硬骨鱼类 *VgB* 基因 ORF 的同源性比较
Tab. 3 Homology comparison between *Xiphophorus helleri*'s *VgB* ORF and those of other teleosts

物种 Species	核苷酸同源性 /% Nucleotide homology	氨基酸同源性 /% Amino acid homology	编码多肽的长度 /aa Length of peptide
剑尾鱼 <i>Xiphophorus helleri</i>	100	100	1688
食蚊鱼 <i>Gambusia affinis</i>	94	89	1675
底鳉 <i>Fundulus heteroclitus</i>	85	80	1687
青鳉 <i>Oryzias latipes</i>	76	65	1716
真鲷 <i>Pagrus major</i>	72	74	1725
条斑星鲷 <i>Verasper moseri</i>	74	63	1651
黑线鳉 <i>Melanogrammus aeglefinus</i>	71	55	1643

剑尾鱼 *VgB* 与 GenBank 中已登录的其他鱼类 *VgB* 相比, 与同属于花鳉科的食蚊鱼的同源性最高, 核苷酸同源性为 94%, 氨基酸同源性为 89%; 与底鳉、真鲷、青鳉、条斑星鲷、黑线鳉 (*Melanogrammus aeglefinus*) 的核苷酸同源性为 71% ~ 85%, 氨基酸同源性为 55% ~ 80%。比较结果表明, 亲缘关系越近, *VgB* 间的同源性越高, 同时也反映了这些物种的系统分类地位。

3 讨论

本实验运用 RT-PCR 和 RACE 扩增法从剑尾鱼肝脏组织中分别扩增得到了 *VgA*、*VgB* 的 3' 和 5' 末端序列及 4 个部分序列, 经过 DNASTar 软件

拼接得到 5 159 bp 和 5 349 bp 2 条 cDNA 序列。用软件 ExPASy 对拼接的序列进行分析发现, 2 条序列均具有完整的 ORF, 3' 端均含有转录终止信号 AATAAA, 2 条序列与食蚊鱼 *VgA*、*VgB* 的 ORF 比对, 同源性分别高达 95% 和 94%。经过氨基酸序列同源比对分析, 本实验得到的剑尾鱼 *VgA* 基因编码的氨基酸序列与 GenBank 中登录的 8 种鱼类 *VgA* 同源性为 39% ~ 90%, 剑尾鱼 *VgB* 基因编码的氨基酸序列与 GenBank 中登录的 6 种鱼类 *VgB* 同源性为 55% ~ 89%, 且亲缘关系越近, 同源性越高。上述分析表明, 本实验成功克隆了剑尾鱼 *VgA*、*VgB* 全长 cDNA 序列。本研究扩增得到的剑尾鱼 *VgA* 氨基酸序列推导的蛋白分子量大小为 187 kD, 笔者先前已

报道剑尾鱼 VgA 分子量大小为约为 180 kD^[5], 与本研究结果一致。

卵黄蛋白原是一种大分子量的磷酸脂糖蛋白, 在运输的过程中被酶切修饰成几种小的卵黄蛋白。在硬骨鱼中, 这些卵黄蛋白包括由重链和轻链组成的卵黄磷蛋白(Lv), 富含磷酸化丝氨酸的卵黄高磷蛋白(Pv)和 β' 组分蛋白。 β' 组分被 Campbell 等^[6]定义为与卵黄蛋白原相关的又 1 种卵黄蛋白, 它是不含有磷酸脂的小分子蛋白。笔者把克隆到的剑尾鱼卵黄蛋白原氨基酸序列与亲缘关系较近的食蚊鱼、底鳉氨基酸序列比对后发现, 所得的 2 条剑尾鱼 Vg 序列均包含卵黄磷蛋白和卵黄高磷蛋白, 其中 VgA 还含有 β' 组分。

经结构分析表明, 本研究克隆到的 2 条剑尾鱼 Vg 氨基酸序列均含有 Vitellogenin 结构域、VWFD 结构域和多聚丝氨酸区域。Vitellogenin 结构域是卵黄蛋白原 N 端的稳定结构, 是 LvH 链的一部分, 长约 670 个氨基酸, 由一定的折叠和螺旋构成^[7]。VWFD 结构域是一个高度结构化的区域, 其中第 1 个保守的半胱氨酸以二硫键形式与第 2 个保守的半胱氨酸连接。VWFD 结构域不仅是血液凝集因子 VIII 结合到 VWF 的必需区域, 也是血管性血友病因子(VWF)形成正常聚合所必需的^[8]。VWFD 结构域可以与不同的结构域连接, 如 Vitellogenin、VWFC、VWFA 和 ZP 结构域等。多聚丝氨酸区域是含有前后排列串连在一起的多个丝氨酸残基的特殊结构, 是昆虫和许多脊椎动物卵黄蛋白原中常见的特征结构, GenBank 中登录的大部分鱼类卵黄蛋白原氨基酸序列也都含有多聚丝氨酸区域。研究表明, 多聚丝氨酸区域可能是酪蛋白激酶 II 的磷酸化位点^[9-10]。在昆虫中这种特殊的结构域可能与受体介导的卵黄蛋白原进入卵母细胞的内吞作用相关^[11]; 而在脊椎动物中, 这些结构域可能做为磷酸和金属离子的载体参与骨的形成^[12]。多聚丝氨酸区域是通过氨基酸密码子连续扩增而产生, 且其存在的区域进化速率比多数基因都快^[13]。

卵黄蛋白原是广泛存在于卵生动物体内的雌性特异性蛋白。研究显示, 几乎所有动物的卵黄蛋

白原基因都起源于一个共同的祖先基因。由于卵黄蛋白原既存在进化速率较快的富含丝氨酸区域, 又存在高度保守的序列结构, 因此该蛋白序列在种间显示出大的分歧度, 同时在高级分类阶元间也表现出良好的分辨率。因而卵黄蛋白原序列作为一种分子标记, 可以很好地用于进化关系较近和较远物种的系统进化研究^[14]。本研究将剑尾鱼卵黄蛋白原氨基酸序列与 GenBank 中已登录的其他动物卵黄蛋白原氨基酸序列进行比对发现, 剑尾鱼 2 条卵黄蛋白原氨基酸序列与同为花鳉科的食蚊鱼的同源性均高达 89% 以上, 与其他鱼类同源性为 40% ~ 80%, 与鸡和凡纳滨对虾的氨基酸同源性低于 35%, 很好地反映了它们的系统分类地位。

雌激素或其类似污染物在进入鱼体后, 会诱导肝脏合成 Vg, 在 Vg 合成之前其 mRNA 的表达将被启动。通过对 Vg mRNA 的检测, 可以从分子水平上反映环境内分泌干扰物的污染状况。最近几年, 国外许多实验室已开始致力于 Vg 分子水平的研究, 并取得了较大的进展。将黑头软口鲮(*Pimephales promelas*)^[15] 和加州鲈(*Micropterus salmoides*)^[16] 的 Vg mRNA 作为环境雌激素污染物的监测指标, 发现污染物诱导 4 d 就能检测到 Vg mRNA 的表达; Leusch 等^[17] 利用食蚊鱼卵黄蛋白原 mRNA 作为环境雌激素污染物监测指标, 可以检测到 250 ng/L 的 17 β -雌二醇(E2)。本研究获得了剑尾鱼的卵黄蛋白原基因的全长 cDNA 序列, 为剑尾鱼卵黄蛋白原分子水平检测方法的建立打下基础。

参考文献:

- [1] Roubal W T, Lomax D P, Willis M L, et al. Purification and partial characterization of English sole (*Pleuronectes vetulus*) vitellogenin [J]. Comp Biochem Physiol, 1997, 118B: 613-622.
- [2] Lomax D P, Roubal W T, Moore J D, et al. An enzyme-linked immunosorbent assay (ELISA) for measuring vitellogenin in English sole (*Pleuronectes vetulus*): development, validation and cross-reactivity with other pleuronectids [J]. Comp Biochem Physiol, 1998, 121B: 425-436.
- [3] 吴淑勤, 黄志斌, 石存斌, 等. 水生实验动物—剑尾鱼 [M]. 北京: 中国农业出版社, 2005.

- [4] Jannick D B, Henrik N, Gunnarvon H, et al. Improved prediction of signal peptides: SignalP 3.0 [J]. *Mol Biol*, 2004, 340: 783–795.
- [5] 刘春, 李凯斌, 王芳, 等. 剑尾鱼卵黄蛋白原基因片段克隆、表达及蛋白检测方法的建立 [J]. *中国水产科学*, 2007, 14 (6): 883–888.
- [6] Campbell C M, Idler D R. Characterization of an estradiol-induced protein from rainbow trout as vitellogenin by the cross reactivity to ovarian yolk fractions [J]. *Biol Reprod*, 1980, 22: 605–617.
- [7] Anderson T A, Levitt D G, Banaszak L J. The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein [J]. *Structure*, 1998, 6: 895–909.
- [8] Jorieux S, Fressinaud E, Goudemand J, et al. Conformational changes in the D' domain of von Willebrand factor induced by CYS 25 and CYS 95 mutations lead to factor VIII binding defect and multimeric impairment [J]. *Blood*, 2000, 95 (10): 3139–3145.
- [9] Kuenzel E A, Mulligan J A, Sommercorn J, et al. Substrate specificity determinants for casein kinase II as deduced from studies with synthetic peptides [J]. *J Biol Chem*, 1987, 262: 9136–9140.
- [10] Meggio F, Pinna L A. Phosphorylation of phosphatidylserine by casein kinase-2 provides the evidence that phosphoserines can replace carboxylic amino acids as specificity determinants [J]. *Biochim Biophys Acta*, 1988, 971 (2): 227–231.
- [11] Wahli W. Evolution and expression of vitellogenin genes [J]. *Trends Genet*, 1988, 4: 227–232.
- [12] Nardelli D, Gerber-Huber S, van het Schip F D, et al. Vertebrate and nematode genes coding for yolk proteins are derived from a common ancestor [J]. *Biochemistry*, 1987, 26: 6397–6402.
- [13] 陈家安, 唐振华. 昆虫分子科学 [M]. 北京: 科学出版社, 2001: 45–55.
- [14] 叶海燕, 黄原. 卵黄蛋白原在系统进化中的应用 [J]. *陕西农业科学*, 2005, 5: 67–70.
- [15] Korte J J, Kahl M D, Jensen K M, et al. Fathead minnow vitellogenin: complementary DNA sequence and messenger RNA and protein expression after 17 β -estradiol treatment [J]. *Environ Toxicol Chem*, 2000, 19 (4): 972–981.
- [16] Bowman C J, Denslow N D. Development and validation of a species- and gene-specific molecular biomarker: vitellogenin mRNA in largemouth bass (*Micropterus salmoides*) [J]. *Ecotoxicology*, 1999, 8 (5): 399–416.
- [17] Leusch F D, Van den Heuvel M R, Laurie A D, et al. Quantification of vitellogenin mRNA induction in mosquitofish (*Gambusia affinis*) by reverse transcription real-time polymerase chain reaction (RT-PCR) [J]. *Biomarkers*, 2005, 10 (6): 429–438.

Cloning and characterization of vitellogenin gene from swordtail fish *Xiphophorus helleri*

LIU Chun¹, LI Kaibin¹, GENG Dongyu², NIE Xiangping³, WANG Fang¹, WU Shuqin¹

(1. Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou 510380, China; 2. School of Life Science and Technology, Dalian Fisheries University, Dalian 116023, China; 3. Institute of Hydrobiology, Jinan University, Guangzhou 510632, China)

Abstract: Vitellin is the major component of egg yolk protein which comes from cleavage and process of vitellogenin (Vg). Vitellogenin is a precursor of egg yolk protein and has acted as a useful biomarker of estrogenic substances exposure. In this study, we cloned and analyzed two forms of full-length Vg cDNA sequences in swordtail fish *Xiphophorus helleri*. The two full-length Vg cDNAs were obtained by RT-PCR and RACE (rapid amplification of cDNA ends) method with total RNA extracted from liver of swordtail fish which were injected with 17 β -estradiol. The result of sequence analysis indicated that the full length of swordtail fish VgA cDNA is 5 159 bp, containing a 3' untranslated region (UTR) of 85 bp, a 5' UTR of 16 bp and an open reading frame (ORF) of 5 058 bp, encoding 1 685 amino acids. The comparison results showed that the nucleotide homology of VgA was 59%–95% and amino acid homology of VgA was 39%–90% between swordtail fish and other eight teleosts. The full length of VgB cDNA is 5 349 bp, containing a 3' untranslated region (UTR) of 268 bp, a 5' UTR of 14 bp, and an open reading frame (ORF) of 5 067 bp, encoding 1 688 amino acids. The sequence comparison results show that the nucleotide homology and amino acid homology of VgB were 71%–94% and 55%–89%, respectively between swordtail fish and other teleosts. The two cloned cDNA sequences possess all characteristic motifs of teleost Vg genes including vitellogenin domain profile, VWFD domain profile and polyserine domains. This study demonstrated the existence of the two different forms of Vg gene at cDNA level. The two Vg sequence data will help to design nucleotide probes for detecting Vg gene expressions as a biomarker in environmental estrogens monitoring. [Journal of Fishery Sciences of China, 2010, 17 (1) : 31–43]

Key words: *Xiphophorus helleri*; vitellogenin; gene cloning; sequence analysis

Corresponding author: WU Shuqin. E-mail : wushuqin001@21cn.com