

文章编号: 1674-5566(2013)04-0502-08

## 日本文昌鱼 ZP 蛋白氨基酸序列多样性的初步研究

王忠军<sup>1</sup>, 许强华<sup>1,2,3,4</sup>

(1. 上海海洋大学 海洋科学学院, 上海 201306; 2. 上海海洋大学 大洋渔业资源可持续开发省部共建教育部重点实验室, 上海 201306; 3. 农业部大洋渔业资源环境科学观测实验站, 上海 201306; 4. 国家远洋渔业工程技术研究中心, 上海 201306)

**摘要:** ZP 蛋白家族是一类具有相对保守的 ZP 结构域的蛋白质, 通常在精卵结合过程中起作用。从日本文昌鱼 (*Branchiostoma japonicum*) 中发现一种编码 927 个氨基酸残基的 ZP 蛋白 cDNA。利用 SMART 预测该 ZP 蛋白序列含有 3 个结构域: 透明带结构域、跨膜结构域以及低复杂度区域。利用该 ZP 基因的整个开放阅读框, 检测该基因在日本文昌鱼群体中的氨基酸序列多样性。研究结果发现, 在 9 个日本文昌鱼个体中扩增获得序列长度为 2 481 ~ 2 784 bp, 编码 826 ~ 927 个氨基酸。氨基酸序列比对显示, 存在 55 个变异位点, 占总氨基酸残基数目的 5.93% (55/927), 其中有 2 个个体的氨基酸序列分别存在 9 个和 101 个氨基酸残基的缺失, 这可能与 ZP 基因的不同剪切有关。揭示了日本文昌鱼 ZP 蛋白序列多样性可能与该蛋白的结构与功能多样性具有一定的相关性。

**研究亮点:** 通过质谱比对分析和设计引物第一次克隆得到日本文昌鱼的 ZP 基因, 分析了该 ZP 基因的氨基酸结构域, 其具有约 249 个氨基酸残基的典型 ZP 结构域。通过扩增日本文昌鱼不同个体之间的 ZP 基因, 发现该 ZP 基因编码的氨基酸序列在日本文昌鱼不同个体之间是存在差异的。

**关键词:** 日本文昌鱼; ZP 蛋白; ZP 结构域; 序列多样性

**中图分类号:** S 917

**文献标志码:** A

众所周知, 后生动物的卵细胞为卵膜 (或卵壳) 所包裹。在哺乳动物中, 称为透明带 (zona pellucida, ZP); 在非哺乳动物中, 一般称之为卵黄膜 (vitelline coat, VC)<sup>[1]</sup>。ZP 蛋白家族是一组位于卵细胞 ZP/VC 上, 具有相对保守的结构域 (ZP 结构域) 的一类蛋白质, 在后生动物精卵结合过程中起到了关键作用, 如充当精子结合的脚手架蛋白, 阻止多精子进入等<sup>[2-3]</sup>。目前对于 ZP 蛋白的研究主要集中在脊椎动物中。研究发现哺乳动物的 ZP 蛋白可分成 3 种亚型<sup>[4]</sup>; 在低等的脊椎动物中 ZP 蛋白具有更多的亚型, 如鱼类 ZP 蛋白有 5 种亚型<sup>[5-6]</sup>。

文昌鱼属于脊索动物门, 头索动物亚门, 它们终生具有脊索动物门的 3 大特征: 脊索、咽鳃裂和背神经管, 器官系统的结构简单且躯体半透

明, 其基因组长度大约只有高等脊椎动物基因组的 17%<sup>[7-8]</sup>, 是现存生物中最近似于脊椎动物亚门直接祖先的一个类群, 具有重要的进化地位。文昌鱼不仅是研究脊椎动物起源与演化的关键类群<sup>[9-11]</sup>, 而且在进化发育生物学、比较与功能基因组学, 以及性腺发育等方面的研究都具有重要意义, 是一个很有价值的实验室模式动物<sup>[12-13]</sup>。中国的文昌鱼有厦门 (白氏) 文昌鱼 (*Branchiostoma belcheri*)、日本文昌鱼 (*Branchiostoma japonicum*)、短刀侧殖文昌鱼 (*Epigonichthys cultellus*)、青岛文昌鱼 (*B. belcheri tsingtauense*) 等<sup>[14-15]</sup>。日本文昌鱼是一种分布于日本和中国的厦门海域, 与青岛文昌鱼之间的遗传距离很近<sup>[16]</sup>。本文在以日本文昌鱼成熟未受精卵为实验材料建立了一套文昌鱼卵壳的分离

收稿日期: 2013-01-13 修回日期: 2013-03-30

基金项目: 国家海洋 863 计划项目 (2008AA092602)

作者简介: 王忠军 (1988—), 男, 硕士研究生, 研究方向为海洋生物分子遗传学。E-mail: alexjun2012@126.com

通信作者: 许强华, E-mail: qhxu@shou.edu.cn

技术之后,对文昌鱼卵壳总蛋白组分进行 SDS-PAGE 分离和液相色谱-质谱串联技术(LC-MS/MS)分析,筛选获得了一条含有透明带结构域的 ZP 蛋白(GenBank 登录号:KC152940)。本研究将继续研究该 ZP 蛋白基因的序列特征以及在群体中的序列多样性。这些研究为后续文昌鱼卵壳蛋白功能探索打下了基础。

## 1 材料与amp;方法

### 1.1 材料

日本文昌鱼(*Branchiostoma japonicum*)由渔民采集于厦门海域的潮间带。对采集到的文昌鱼活体进行分类,区分雌雄个体,按性腺发育程度不同进行发育阶段分期,之后立即分装 EP 管中,放入 -80 °C 冰箱冻存备用。性成熟的日本文昌鱼卵巢(IV 期)呈黄色或浅黄色,直径 1.0 ~ 1.19 mm,此时卵母细胞已经成熟,即将进入排卵期,卵细胞椭圆形,胞质中有大量卵黄颗粒。

### 1.2 仪器与试剂

试验仪器包括 Eppendorf centrifuge 5417R 冷冻离心机、Eppendorf PCR 仪、OLYMPUS SZ51 解剖镜、BIO-RAD Dcode™ Universal Mutation Detection System 垂直电泳系统、Alpha Innotech 凝

胶成像系统、Milli-Q 超纯水仪(美国 Millipore 公司出品)。

试剂包括 Trizol(美国 Invetrigen 公司出品),大肠杆菌(*E. coli*)DH5 $\alpha$ ,PMD18-T 载体,反转录试剂盒,Taq 酶,dNTP,DEPC,苯酚、氯仿、异戊醇、乙醇、异丙醇均为国产级试剂。

### 1.3 试验方法

#### 1.3.1 分离卵巢和总 RNA 的提取

从 -80 °C 冰箱中取出 9 条日本文昌鱼雌性 IV 期个体放在冰上解冻。在解剖镜下用细针迅速剥离出卵巢组织。按照试剂盒说明书用 Trizol 试剂分别抽提出 9 条日本文昌鱼卵巢组织的总 RNA,进行编号和琼脂糖凝胶电泳检测。

#### 1.3.2 RT-PCR

取 5  $\mu$ g 的总 RNA 样品,用 Oligo(dT)作引物分别反转录合成第一链 cDNA。取 1/10 体积的反转录产物进行 PCR 扩增。根据已得到的文昌鱼的基因序列(KC152940),采用 Primer Premier 5.0 软件设计引物,扩增该基因的整个开放阅读框序列。引物的位置见图 1。由于开放阅读框长度较大,总共设计了 3 对引物,引物名称及序列如表 1 所示。

表 1 扩增 ZP 基因的引物  
Tab.1 The primer sets for amplification of the ZP gene

引物名称	上下游引物序列	PCR 退火温度(TM)
ZP-1	ZP-1-F: ATGGCGTCCATTCCAAGCTTGCTTGTC	60 °C
	ZP-1-R: TGGACTCCTCGGCGTAGAAACAGACGAT	
ZP-2	ZP-2-F: TGACCGTCACTGGTGACATCTCATCCAC	57 °C
	ZP-2-R: GGATCTCATTCTCGTATAACAAGCTCGTC	
ZP-3	ZP-3-F: GAGATTACGGCGTGAACGCAGCTGATA	55 °C
	ZP-3-R: CTATGAGACTATAGACGGCTTTAACTGG	

PCR 反应体系为模板 cDNA 2  $\mu$ L,引物各 1.5  $\mu$ L(10 nmol/L),dNTP 4  $\mu$ L(2.5 mmol/L),10  $\times$  Taq Buffer 5  $\mu$ L,Taq 酶(5 U/ $\mu$ L)1  $\mu$ L,以去离子水补足 50  $\mu$ L。反应条件为 95 °C 5 min;95 °C 30 s,TM 55 s,72 °C 90 s,30 个循环;72 °C 10 min。每对引物的退火温度 TM 如表 1 所示。

#### 1.3.3 TA 克隆和测序

PCR 产物在 1.5% 的琼脂糖凝胶上进行电泳检测。目的片段用胶回收试剂盒(Gel Extraction Kit,天根生化科技有限公司)纯化后与 pMD18-T 载体连接,转化至 *E. coli* DH5 $\alpha$  感受态细胞,利用蓝白斑筛选阳性克隆,所获得的克隆利用原 PCR 扩增引物进行菌落 PCR 鉴定后,提取质粒,送上

海迈普生物公司进行双向测序,为了确保测序结果的准确性,每组 PCR 产物都会挑取 3 个阳性克隆进行测序。

#### 1.3.4 序列分析与数据处理

针对已得到的文昌鱼的基因序列(KC152940),运用 BioEdit 软件进行翻译,获得其氨基酸序列。使用蛋白结构预测和功能分析的在线工具 SMART (<http://smart.embl-heidelberg.de>)对该 ZP 的氨基酸序列进行结构分析。PCR 扩增的测序结果经过拼接和去除载体序列之后,得到 9 个日本文昌鱼个体的 ZP 基因序列,并利用 BioEdit 翻译软件获得其氨基酸序列。利用 Clustal W 进行序列比对分析,得到 ZP 基因在不

同个体间的核苷酸与氨基酸序列差异。

## 2 结果

### 2.1 ZP 基因的序列特征分析

本文所得到的日本文昌鱼 ZP 基因的序列全

长为 2 784 bp, 编码一条长为 927 个氨基酸的蛋白质序列 (GenBank 登录号为 KC152940, 图 1)。利用 SMART (<http://smart.embl-heidelberg.de>) 在线工具对该 ZP 蛋白的氨基酸序列进行结构分析, 结果见图 2。该氨基酸序列存在典型的信号

```

ATGGCGTCCATTCCAAGCTTGCTTGTCTGGTGTCTACATTTGTCATCCCTCTCGCCGAGGGCTCGCATTCCGAGGGGGACCATCTCT 90
M A S I P S L L V L V S T F V I P L A E G S H F R G G T I S 30
ZP-1-F
TGGCTACCTCTGGACGCTCCTTCCGAAACTTCCGCGTTTTTGTAGTTCAACCTTGGCTGGAAGAAGACGAGTGCACCGGAACCGGATGT 180
W L P L D A P S E T S A V F R F N L G W K K T S A T G T G C
GACACGGCTGCTCCAGGGTCAAACCTTGTCCAGCAATGAGGATCCCTGGGAGTGCCTGCCGATGTCGGTATCCCGGTACCGTTC 270
D T A A L Q G Q T L V T T N E D P W E C V S G C A S S G T V
TTTTGGCCAAATCAGACTACTGTACCGAATCAACGTAGTCAAACCTGGGCAAAGGCTTCAACTCCTTCGCTCACGATTTTAGA 360
F L A N Q D Y Y C T E F N V G Q N W A K A F N S F A H D F R
TCCACAGAGGACACATACGTCGTTGGCTATGAAAGTTGCTGCTGCTGTCAGATCCAGAATCAGCAGACGGGAGTCTTTCACAATCCCGCTG 450
S T E D T Y V V G Y E S C C W L Q I Q N Q Q T G V F H N P L
ATTTCCCTGAGGACGACTGTAGACTCGGGTTCAGGTGACACCGGCAAAGGCAACCGCTCCAGCCAGCATGCAAAACAACCAT 540
I S L R T T V D L G F R S D T G K G N H A P A T S M Q T T I
CGTGTCAAAAACGGGTGTCTGTCCTCCGACCTACCCCGCGGCAAGCTGGCTGTGGCGGACATCGACGGAGACACTGTGAAATGCCGCTTC 630
R V T N G C L S P T Y P A G K L A V A D I D G D T V K C R F
GCTCAGGAGCGACGAGTGCAGGCGGAGCGTGTGACAGTCCCGCGCTGGTACTGAACGAGGACTGCACGTACTCGTACAACGGCCCG 720
A Q G A T E C G E A C D Q F P G L V L N E D C T Y S Y N G
GCCTCCATCACCAGTCTGGGACTTCTCTGTCTCGTGGTGGCTCATGGTTGAAGACTTCCCGGTACAGACCATCTCCGTCAT 810
A S I T I S S G T S L F F V V A V M V E D F P V Q T I F R H
AACCAGGAAGTCCCGACGACACAGGCCCTCAGCAGCGTACCCTCCAGTTCCTGATAGAAGTCAAACCTGACCCGGGTATCCTCTGCGGC 900
N Q E V P T T Q A L S S V P L Q F L I E V K T D P G I L C G
TCTGAGCCGACTGACCGGTGCTACCGCGGCCAGGACCTGCCTGCCATCCCGGACAGTGTGAGTACACCATGACTGTGAGGCT 990
S E P V L T G A T P A Q D T C L P I P D S V E Y T M T V E A
CAACCTGTAGCACCTGCCACCTTGCAGAGATTAATGTTCTCGGACCAAAGGGCATGACTAAGAGCGCCCTGACCGTCACTGGTGACATC 1080
Q P V A P A T L A E I N V L G P K G M T K S A L T V T G D I
ZP-2-F
TCATCCAGCATCAGTGGACGCGCGCCCGGTCAGCGGACCCACATCGTCTGTTTCTACGCGGAGGACTCCAACGCGAGTGCAG 1170
S S T T I T W T P A P G Q R G P H I V C F Y A E D S N A V Q 390
ZP-1-R
TCTGACGAGCATGCGTCTCTCTCATGTCGGAGGGAGCATCACCCGCTGCTGTGTCGATGCCCTACGCTGATCCCTACACCCGGCAGG 1260
S D R T C T S L M V G S I T P P A V D A P T L I P T P G T
GCCCCGACAGACTACCCGTGTGACACGACCTCCGGTTCAGCGCCAGTTCGACAGCTGGTCTACCGCCGACTTTCGACACCTTCATC 1350
A P T D Y P C D T T L R F S A T F D Q L V L P P T S D T F I
ATCTTCTACGACTCCACCAACGCGAGTCTACAACACAACACGGAAGATTCGCGAGGGCAGCCCGCGCCGCGCCGACCAACCTAC 1440
I F Y D S T C N A E F Y K H N T K I P D G Q P A P A Q S N T Y
GTTTTTCGAGTCCCGGCGAACCTTCCAGCCCTTGGGGACTACCCATAGCCATGACAGCGGCGCGCTGGAGGTTGTGACGGGGTGC 1530
V F D V P T F Q A L S S V P L G T Y T I A M Q T G A L E G V C
GGGGCGGGCGGCGTGCAGAGCGACGCTACGAGACCGCGGAGGACCTGGAGTTTCACTGCAAGTCTACATCCCGCGCGCGGTC 1620
G A G A G V Q S D A Y E T A G G T G S F T C K F Y I P A P V
ACCCGCGCCACCTGCTCCACCTCCAGCAGCGCGGCGGACCGCGGCGAGTGCAGTGCAGGCTGCTCGGTACGCTGCTCCGCGCGG 1710
T T A P P A P P A T T A A T P G A T V R V V G T P A P G A
CTCCCACCATGCCGCTACCGCTATCGTGTCTTCTTCTTACCAGCATCAGGTTACGATTCGTTGCTGATTTACCGCGCGGTCG 1800
L P T M P V T A I V V T C S P T S I T V T I P L S E I T G V
ZP-3-F
AACCGAGTGATATCCGGTACCGGAGTGTCTCTTGTCTTACCGGTTGCTCTCCGAGACTCCGTCAGCATCACTACCGGATTTACGAGGATGT 1890
N A A D I R Y R S A P C L P V V S G D S V S I T T G F Q E C 630
GGCATTACCATGGCGACCCAGGAGACGAACTTGTATACGAGAATGAGATCCATACAGAGTTTGAAGCACCCGCTCCGAGGCAACAGC 1980
G T T M A T Q G D E L V Y E N E I H T E F A S T A I R G N S 660
ZP-2-R
GTGGATGGAAGTTGAGTGTGCGTACGACAGTAACCGGTTGACCGGAAACCGTTACGCGCTGATGGGCTCCGATTCGCGCAGG 2070
V D G K V E C A Y D C S N T V V T G K R F S A L M G S V F G R
AGGAGCAGCGGCGGAGTTCGAGTTCCTTCCGACTTCTACGCGCAACACCTTACGACGCGCTTCCAGCTACCCCGTTCCTTACCAT 2160
R S S G Q F E F S F D F Y T D N T F S T A F T S Y P V S Y H
CCTAACCAAGAGATGTTTCATCGGCGTCCACCTCGTCTCGACCAACACGGAATCTCGTGTCTTTTGGGACAACGCAAGCGGACACCCGGG 2250
P N Q E M F I G V H L V S T N T D L V L F A D N C K A T P G
GTCGAGTGGGACGCTCGCCAGTTACACCATCCGGGAGAACCGGTGTAACGTAGACCCCTCCTGACCTGTTACGAAACCGCGGACGCT 2340
V E W D S S P S Y T I R E N G C N V D P L L T W Y E P A D A
GCCCCAGCCGACGGGAGGAGAACTTCGGCATCGTGGTTCGCGTTCGCGGACTACTCAACGTTGATATCCACTGTGACCTGGTGGTG 2430
A R A A R E E N F G I V V F R F A D Y S T L Y I H C D V V V
TGTGCGGACGAGATACAGTTCGTACTGCGCCACCGCGCCGCGAGCTGCGGACCGGTTAGGAAGAGGCGAGACCGCGGACGTTGATGGAC 2520
C A A A D T S S Y C A T A P A S C G T G R K R R D A D V M D
AAACTGCACCGCTCGAGGTGACGCGGCTCAAGATCATCAGGACGACGAGGAGCATATAACAACCGTCAAGCCATGGCGGAC 2610
K L H R F E V T S G P V K I I R D D G G H N N R Q A H G D
TGGAAAGAGGCGTTTTTGGCAGTTGTGCCCTCCCGTGTGCGGGGCTGATGCTGTGTCGCGGGTGTCTTCTGTCGCGCGCGGCTG 2700
W K E A F S A V V S P P V L G V L M L C L G V I F L S A A V
TCCTGTACTACCTACCGCGCTACGAGCAGCTTACACCGGTTGTCGAGACCGACCTTAAAGCGCTTATATCTCATAG 2784
S C Y Y L R R Y Q Q L Y T T V F E T D H L K P S I L S * 927
ZP-3-R

```

图 1 由 ZP (KC152940) 翻译得到的氨基酸序列

Fig. 1 The amino acid sequence translated from ZP (KC152940)

细下划线所示为起始密码子、终止密码子;粗下划线所示为引物的位置,引物名称标注在下划线下方。

肽 (signal peptide, SP)、透明带结构域 (ZP domain)、跨膜结构域 (transmembrane domain, TMD) 以及低复杂度区域 (low complexity region, LCR)。从第 537 位至第 566 位为低复杂度结构域;从第 582 位至 830 位为透明带结构域 ZP;从第 878 位至第 900 为跨膜结构域(图 2、表 2)。

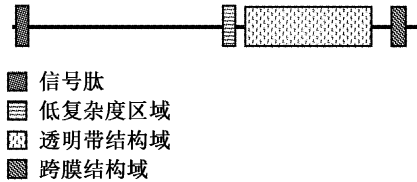


图 2 ZP 蛋白的结构域预测  
 Fig. 2 The prediction of structural domains in ZP protein

表 2 ZP 蛋白的预测结构域 E 值  
 Tab. 2 The E-value of the predicted domains in the ZP protein

名称	起点	终点	E 期望值
信号肽	1	21	N/A
低复杂度区域	537	566	-
透明带结构域	582	830	7.77e-21
跨膜结构域	878	900	-

2.2 ZP 蛋白的氨基酸序列多样性分析

提取日本文昌鱼卵巢组织的总 RNA,经 Oligo(dT)作引物反转录之后得到 cDNA,用设计的引物进行 PCR 扩增(表 1)。引物 ZP-3-F/R 扩

增得到的第三个片段(ZP-3)为从第 596 位至末尾 927 位片段,恰恰横跨了 ZP 和 TMD 两个结构域。利用引物 ZP-3-F/R 进行 PCR 扩增得到 ZP-3 片段的大小在不同个体之间呈现差异,电泳图谱见图 3,扩增产物长度范围为 700 ~ 1 000 bp。ZP-1、ZP-2 片段的电泳结果显示大小基本一致,之后的测序结果也验证了这点。测序结果经分析去除载体序列后翻译成氨基酸序列,9 条氨基酸序列比对结果见图 4。9 个日本文昌鱼个体中,扩增获得的 ZP 序列长度介于 2 481 ~ 2 784 bp,编码 826 ~ 927 个氨基酸。5 条氨基酸序列含有 927 个氨基酸残基,2 条含有 926 个氨基酸残基,另 2 条序列的氨基酸残基数分别是 918 和 826。氨基酸序列比对后发现序列中存在 55 个变异位点,占总氨基酸残基数的 5.93% (55/927)。另外,还发现在 5 号与 9 号个体中存在序列缺失(图 4)。

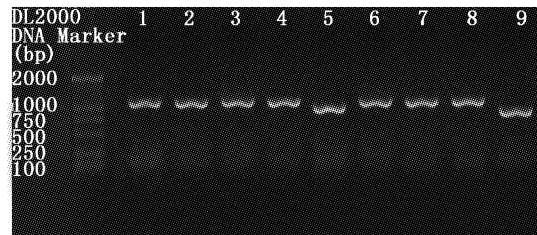


图 3 文昌鱼 ZP 基因扩增凝胶电泳图(ZP-3)  
 Fig. 3 Amplification of ZP gene by PCR(ZP-3)

```

1 MASIPSLLVLVSTFVIPLAEGSHFRGGTISWLPLDAPSETSAVFRFNLGWKKTSAATGTGCDTAAALQGQTLVTTNEDPWEC 80
2 ..... 80
3 ..... 80
4 ..... 80
5 ..... 80
6 ..... 80
7 ..... 80
8 ..... 80
9 ..... 80

1 VSGCASSGTVFLANQDYICTEFNVGQNWAKAFNSFAHDFRSTEDTYVVGYESCCWLQIQNQQTGVFHNPLISLRTTVDLG 160
2 .....F.....T..... 160
3 .....F.....T..... 160
4 .....F.....T..... 160
5 .....F.....T..... 160
6 .....F.....T..... 160
7 .....F.....T..... 160
8 .....F.....T..... 160
9 .....F.....T..... 160

1 FRSDTGKGNHAPATSMQTTIRVTNGCLSPTYPAGKLAVADIDGDTVKCRFAQGATECGEACDQFPGLVLNEDCTYSYNGP 240
2 .....A..... 240
3 .....A·D..... 240
4 ··N·····A·D·····V···T···P···· 240
5 .....A·D·····H···· 240
6 .....A·D·····PA···MR···· 240
7 .....N·····ARD···· 240
8 ··N·····A·D···· 240
9 ··N·····A·D·····PA···MR···· 240
    
```

1	ASITISSGTSLFFVAVMVEDFPVQTI FRHNQEVPTTQALSSVPLQFLIEVKTDPGILCGSEPVLTGATPAQDTCLPIPD	320
2	.....	320
3	.....-N-V.....	320
4	.....-N-V.....	320
5	.....-N-V.....	320
6	·L·····F·····N·····-N-V.....	320
7	.....V.....	320
8	.....-N-V.....	320
9	·L·····-N-V.....	320
1	SVEYTMVVEAQFVAPATLAEINVLGPKGMTKSALTVTGDISSTTITWTTPAPGQRGPHIVCFYAEDSNAVQSDRTCVSLMV	400
2	.....	400
3	.....	400
4	.....A.....	400
5	.....-T.....	400
6	.....A.....	400
7	.....E.....	400
8	.....	400
9	.....	400
1	GGSI T PPAVDAPTLIPTFGTAPTDPDCTTLRF SATFDQLVLPPTS DTFI I FYDSTNAEFYKHNTKI PDGQ PAPAQSNTY	480
2	.....	480
3	.....	480
4	.....	480
5	.....	480
6	.....	480
7	.....	480
8	.....	480
9	.....	480
1	VFDVPANTFQPLGTYTIAMQTGALEGVTCGAGAGVQSDAYETAGGTWSFTCKFYI PAPVTTAPPAPPATTAATPGATV	560
2	.....	560
3	.....	560
4	.....	560
5	.....	560
6	.....	560
7	.....	560
8	.....	560
9	.....	560
1	RVVGT PAFGALPTMPVTAIVVTC SPTSITVTI PLSEITGVNAADIRYRSAPCLPVVSGDSVSITTFQECGTMTTQGD E	640
2	.....-V.....	640
3	.....	640
4	.....	640
5	.....A.....	640
6	.....	640
7	.....	640
8	.....	640
9	.....	640
1	LVYENEIHTEFASTAIRGNSVDGKVECA YDSNTVVTKRFSALMGSVFGRSSGQFEFSDFYTDNTFSTAFTSY PVS YH	720
2	.....T.....	720
3	.....T.....Q.....	720
4	.....	720
5	.....	720
6	.....	720
7	.....	720
8	.....F.....	720
9	.....T.....	720
1	PNQEMFIGVHLVSTNTDLVLFADNCKATPGVEWDSSPSYTIRENGCNVDPLLTWYEPADAARAAREENFGIVVFRFADYS	800
2	.....	800
3	.....Q.....	800
4	.....I.....Q.....	800
5	.....	800
6	.....	800
7	.....-VP.....S.....	800
8	S···I·····K·····A·LH···D·····F·····	799
9	.....Q.....	782
1	TLYIHC DVVCAAADTSSYCATAFASCGTGRKRRDADVMDKLRHFEVTS GPVKI IRDDGGHNNNRQAHGDWKEAFSAVVS	880
2	.....	879
3	.....-V.....	880
4	.....A.....-V.....	880
5	.....-V.....	880
6	.....-V.....	880
7	.....A.....	880
8	.....V.....-C.....	799
9	-----	782
1	PPVLGVL MVCLGVIFLSAAVSCYYLRRYQQLYTTVFETDHLKPSILS*	927
2	.....T.....	926
3	.....	927
4	.....	927
5	.....L.....H·R-----HQKQ.....	918
6	.....V.....	927
7	.....	927
8	.....P.....	926
9	---V.....	826

图 4 ZP 蛋白的氨基酸序列的比对图

Fig. 4 Alignment of ZP amino acid sequences

· 表示氨基酸残基与 1 相同; - 表示氨基酸残基缺失; \* 表示终止密码子。

### 3 讨论

在对文昌鱼的现阶段研究中,大多数学者侧重于研究文昌鱼在进化方面的地位,以及通过文昌鱼来研究脊椎动物的起源问题。而在文昌鱼卵细胞的研究上,还只停留在文昌鱼卵母细胞的形态、内部超微结构等方面的研究。早在 1989 年,两位 HOLLAND 学者在对佛罗里达文昌鱼卵细胞卵壳结构的研究中指出,文昌鱼卵细胞的卵壳从外到里可分为胶质膜(jelly layer)、卵黄膜(vitelline layer)、皮质层(cortical granules)和含有大量卵黄粒(yolk granules)的细胞质结构<sup>[17]</sup>。而迄今为止对于文昌鱼卵壳蛋白相关基因的研究还未见报道。文章以模式生物文昌鱼为研究材料,以具有重要功能的 ZP 基因为研究对象,为以后相关研究打下了基础。

ZP 蛋白是一类高度糖基化的蛋白,其中的 ZP 结构域是一个长度约为 260 个氨基酸残基的结构域<sup>[18]</sup>。自 ZP 蛋白在鼠的卵细胞中被发现以来,已经在很多真核生物中发现了带有 ZP 结构域的 ZP 蛋白,具有不同的结构和功能<sup>[19-20]</sup>。在脊椎动物中,ZP 蛋白在精卵的特异结合和防止多精子与单个卵子结合的过程中有重要的作用,如在人卵细胞中,ZP3 是精卵结合过程中精子的主要受体,ZP2 作为第二配体结合到参与顶体反应的精子上<sup>[21]</sup>。在哺乳动物中卵细胞可以合成 3 种亚型的 ZP 蛋白,它们都有相同的结构域:N 端的信号肽结构、含有约 260 个氨基酸的保守 ZP 结构域、保守的碱性氨基酸蛋白酶水解位点、C 末端的跨膜结构域(TMD)和胞质尾区<sup>[2-3,21]</sup>。笔者获得的日本文昌鱼 ZP 蛋白所含的结构域模式与哺乳动物基本相同,且与尾索动物也相似,如玻璃海鞘(*Ciona intestinalis*)的 ZP 蛋白中同样含有这些保守的结构域<sup>[22]</sup>。可以推测,头索动物(日本文昌鱼)、脊椎动物(家鼠、人类<sup>[2-3,21]</sup>)、尾索动物(玻璃海鞘<sup>[22]</sup>)的 ZP 蛋白都具有保守的结构域,这也说明了它们在进化上的同源性。鉴于 ZP 蛋白结构的独特性,笔者克隆了 ZP 蛋白的整个开放阅读框来研究 ZP 蛋白氨基酸序列的种内遗传多样性。9 个日本文昌鱼个体中,存在 55 个变异位点,占总氨基酸数的 5.93% (55/927) (图 4)。其中,23 个氨基酸变异位点位于信号肽与低复杂度结构域之间,20 个氨基酸变异位点处于

ZP 结构域,4 个变异位点处于 TMD 结构域,CFCS 位点无变异(图 4)。已有研究表明,CFCS 位点比较保守,在蛋白分泌过程中,碱性氨基酸蛋白酶会选择性地对 ZP 蛋白前体进行剪切,这个过程在包括鱼类和鸟类的众多生物中都会存在<sup>[23-25]</sup>。该研究还在 CFCS 区域与 TMD 结构域之间发现了 6 个变异位点(图 4),该区域被报道对于调节 ZP 蛋白的分泌活动具有一定作用<sup>[26]</sup>。另外,在胞质尾区也找到 6 个变异位点(图 4),且 5 号与 9 号个体的氨基酸序列还分别存在 9 个和 101 个氨基酸的片段缺失(图 4),这可能是由于基因转录过程中出现了选择性的剪切所致。ZP 蛋白是卵壳的重要组成部分,通常在精卵细胞的识别过程中发挥重要作用。这在高等动物、尤其是哺乳动物的研究相对较多<sup>[2-3,21,25]</sup>。一些学者开始探索性地研究低等动物 ZP 蛋白的功能。如 KÜRN 等<sup>[22]</sup>对玻璃海鞘的雌性性腺组织的转录组进行 ESTs 分析,得到 4 个编码 ZP 蛋白的基因序列(分别为 Vc16、Vc20、Vc182 和 Vc569),这 4 种 ZP 蛋白同样具有上述的 4 种保守结构域,其研究发现这 4 种 ZP 基因编码的氨基酸序列不仅在种间(与相近物种 *C. savignyi*)有差异性,而且在玻璃海鞘的种内不同个体之间也有多样性,其种内多样性主要是由于核苷酸的替换和缺失。推测玻璃海鞘的 4 种 ZP 蛋白的氨基酸序列多样性可能与防止不同海鞘物种之间的交叉受精和自体受精有关<sup>[20]</sup>。与隶属于尾索动物门的海鞘不同,文昌鱼是雌雄异体,不存在自体受精的问题,但是可以推测日本文昌鱼不同个体之间的 ZP 蛋白的氨基酸序列多样性是否会更有利于整个日本文昌鱼种群的精卵结合过程,还有待研究证实。

ZP 蛋白已被证实在许多物种中有表达,在从蠕虫到人类的多样物种中,ZP 蛋白具有物种间的丰富多样性。同一物种往往也含有多种 ZP 蛋白,ANNA 和 MARKUS 研究指出:在黑腹果蝇(*Drosophila melanogaster*)中,有 18 个 ZP 基因编码 ZP 蛋白,这些 ZP 基因都会在胚胎的上皮组织中表达,有 5 个基因会只在卵子生成的过程中表达<sup>[27]</sup>。KANAMORI 等通过研究在青鳞(*Oryzias latipes*)中编码卵膜蛋白基因的基因组结构和表达图谱,来探究硬骨鱼类的 ZP 基因多样性。其研究发现:在脊椎动物中,编码 ZP 蛋白的基因有

5 类,分别为: *zpa*, *zpac*, *zpb*, *zpc* 和 *zpd*,这也是 ZP 结构域进化的表现<sup>[6]</sup>。所以,在生物进化过程中,ZP 蛋白在不同物种中存在多样性,ZP 蛋白的表达在同一物种的不同发育阶段,不同组织器官中也存在差异性<sup>[28]</sup>。可以推测,ZP 蛋白的多样性是进化的结果,有利于文昌鱼的精卵结合过程,有利于文昌鱼在地球上长达几亿年进化过程。日本文昌鱼不同个体之间 ZP 蛋白多样性研究是对 ZP 蛋白多样性进化研究的一个重要补充。

综上所述,关于 ZP 基因的研究在哺乳动物和一些低等的脊椎动物中已经有相关报道,而关于“化石生物”文昌鱼的 ZP 基因研究还不是很多。研究文昌鱼的 ZP 基因型以及不同 ZP 基因型编码的氨基酸序列在不同物种或者同一物种不同个体之间的多样性是非常有意义的,笔者发现的文昌鱼 ZP 蛋白序列在群体中的多样性推测可能与该蛋白的结构与功能多样性具有一定的相关性,此研究也为今后进一步研究文昌鱼的 ZP 基因提供了一定的基础。

#### 参考文献:

- [1] DARIE C C, JANSSEN W G, LITSCHER E S, et al. Purified trout egg vitelline envelope proteins VE $\beta$  and VE $\gamma$  polymerize into homomeric fibrils from dimers in vitro[J]. *Biochim et Biophysica Acta*,2008, 1784(2):385-392.
- [2] LITSCHER E S, WILLIAMS Z, WASSARMAN P M. Zona pellucida glycoprotein ZP3 and fertilization in mammals[J]. *Molecular Reproduction Development*, 2009,76(10):933-941.
- [3] SAMOYLOVA T I, COX N R, COCHRAN A M, et al. ZP-binding peptides identified via phage display stimulate production of sperm antibodies in dogs [J]. *Animal Reproduction Science*,2010,120(1/4):151-157.
- [4] WASSARMAN P M, JOVINE L, LITSCHER E S, et al. Egg-sperm interactions at fertilization in mammals [J]. *European Journal of Obstetrics & Gynecology and Reproductive Biology*,2004,115(s):57-60.
- [5] LITSCHER E S, WASSARMAN P M. Egg extracellular coat proteins: from fish to mammals [J]. *Histology and Histopathology*,2007,22(3):337-347.
- [6] KANAMORI A, NARUSE K, MITANI H, et al. Genomic organization of ZP domain containing egg envelope genes in medaka (*Oryzias latipes*) [J]. *Gene*,2003,305(1):35-45.
- [7] HOLLAND L Z, ALBALAT R, AZUMI K, et al. The amphioxus genome illuminates vertebrate origins and cephalochordate biology[J]. *Genome Research*,2008,18(7):1100-1111.
- [8] LOUIS A, CROLLIUS H R, RECHAVI M R. How much does the amphioxus genome represent the ancestor of chordates? [J]. *Briefings in Functional Genomics*,2012,11(2):89-95.
- [9] HOLLAND L Z. Amphioxus genomics [J]. *Briefings in Functional Genomics*,2012,11(2):87-88.
- [10] WANG Y B, CHEN S H, LIN C Y, et al. EST and transcriptome analysis of cephalochordate amphioxus-past, present and future [J]. *Briefings in Functional Genomics*, 2012,11(2):96-106.
- [11] LACALLI T C. Sensory systems in amphioxus: A window on the ancestral chordate condition [J]. *Brain Behavior and Evolution*,2004,64(3):148-162.
- [12] PAPS J, HOLLAND P W H, SHIMELD S M. A genome-wide view of transcription factor gene diversity in chordate evolution: less gene loss in amphioxus? [J]. *Briefings in Functional Genomics*,2012,11(2):177-186.
- [13] SCHUBERT M, ESCRIVA H, XAVIER-NETO J, et al. Amphioxus and tunicates as evolutionary model systems[J]. *Trends in Ecology & Evolution*,2006,21(5):269-277.
- [14] ZHANG Q J, ZHONG J, FANG S H, et al. *Branchiostoma japonicum* and *B. belcheri* are distinct Lancelets (Cephalochordata) in Xiamen waters in China [J]. *Zoological Science*,2006,23(6):573-579.
- [15] WANG L, SU H Y, WANG C L, et al. The bridge between invertebrate and vertebrate: the cephalochordate amphioxus [J]. *Transaction of Oceanology and Limnology*,2007(2):45-51.
- [16] LUO J C, XIAO Y S, SONG L, et al. Phylogenetic separation of lancelets in China revealed by mitochondrial *CO I* and *Cyt b* gene analysis [J]. *South China Fisheries Science*,2007,3(2):9-14.
- [17] HOLLAND N D, HOLLAND L Z. Fine-structural study of the cortical reaction and formation of the egg coats in a Lancelet, *Branchiostoma floridae* (Phylum Chordata: Subphylum Cephalochordata = Acrania)[J]. *Biological Bulletin*,1989,176(2):111-122.
- [18] JOVINE L, QI H, WILLIAMS Z, et al. The ZP domain is a conserved module for polymerization of extracellular proteins [J]. *Nature Cell Biology*,2002,4(6):457-461.
- [19] BORK P, SANDER C. A large domain common to sperm receptors (Zp2 and Zp3) and TGF-beta type III receptor [J]. *FEBS Lett*,1992,300(3):237-240.
- [20] WASSARMAN P M, LITSCHER E S. The multifunctional zona pellucida and mammalian fertilization [J]. *Journal of Reproductive Immunology*,2009,83(1/2):45-49.
- [21] MONNÉM, HAN L, JOVINE L. Tracking down the ZP domain: From the mammalian zona pellucida to the molluscan vitelline envelope [J]. *Seminars in Reproductive Medicine*,2006,24(4):204-216.
- [22] KÜRN U, SOMMER F, BOSCH T C G, et al. In the

- urochordate Ciona intestinalis* zona pellucida domain proteins vary among individuals[J]. *Developmental and Comparative Immunology*,2007,31(12):1242–1254.
- [23] WU T T, WANG H P, QIN F, et al. Expression of zona pellucida B proteins in juvenile rare minnow (*Gobiocypris rarus*) exposed to 17 $\alpha$ -ethinylestradiol, 4-nonylphenol and bisphenol[J]. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*,2012,155(2):259–268.
- [24] HAN L, MONNÉM, OKUMURA H, et al. Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3[J]. *Cell*,2010,143(3):404–415.
- [25] GUPAT S K, BANSAL P, GANGULY A, et al. Human zona pellucida glycoproteins: functional relevance during fertilization[J]. *Journal of Reproductive Immunology*,2009,83(1/2):50–55.
- [26] JOVINE L, QI H, WILLIAMS Z, et al. A duplicated motif controls assembly of zona pellucida domain proteins [J]. *Proceedings of the National Academy of Sciences of the USA*, 2004,101(16):5922–5927.
- [27] ANNA J, MARKUS A. A family of genes encoding zona pellucida (ZP) domain proteins is expressed in various epithelial tissues during *Drosophila* embryogenesis [J]. *Gene Expression Patterns*,2004,4(4):413–421.
- [28] XU Q H, LI G, CAO L X, et al. Proteomic characterization and evolutionary analyses of zona pellucida domain-containing proteins in the egg coat of the cephalochordate, *Branchiostoma belcheri* [J]. *BMC Evolutionary Biology*, 2012,239(12):1471–2148.

## The preliminary study on the amino acid sequence diversity of ZP proteins from *Branchiostoma japonicum*

WANG Zhong-jun<sup>1</sup>, XU Qiang-hua<sup>1,2,3,4</sup>

(1. College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China; 2. Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources, Ministry of Education, Shanghai Ocean University, Shanghai 201306, China; 3. Scientific Observing and Experimental Station of Oceanic Fishery Resources, Ministry of Agriculture, Shanghai 201306, China; 4. National Distant-water Fisheries Engineering Research Center, Shanghai 201306, China)

**Abstract:** ZP protein is a family of proteins which possess a conserved ZP domain and generally play an important role in the fertilization process. In this study, a new ZP cDNA from *Branchiostoma japonicum*, which codes 927 amino acid residues, was found. By using SMART software, it was confirmed that the ZP protein contains three domains (ZP Domain, Transmembrane Domain and Low Complexity Region). Furthermore, the sequence diversity of this ZP gene in *B. japonicum* population was analyzed. As a result, the ZP sequences amplified from 9 *B. japonicum* individuals are 2 481–2 784 bp, which code 826–927 amino acids. There are 55 variable sites (5.93%, 55/927) identified by amino acid sequence alignment. 9 and 101 amino acids fragment deletion in two individuals may be due to the diverse alternative splicing of ZP gene. In conclusion, the sequence diversity of ZP proteins may be correlated with its structure and function diversity.

**Key words:** *Branchiostoma japonicum*; ZP protein; ZP domain; sequence diversity