DOI: 10.11758/yykxjz.20160322003

http://www.yykxjz.cn/

# 海蜇(*Rhopilema esculentum*)Notch 基因的 cDNA 克隆和表达<sup>\*</sup>

骆晓蕊<sup>1,2</sup> 朱 玲<sup>20</sup> 周春娅<sup>2</sup> 庄志猛<sup>2</sup> 范艳君<sup>2</sup>

(1. 上海海洋大学水产与生命学院 上海 201306; 2. 农业部海洋渔业可持续发展重点实验室 山东省渔业资源与生态环境重点实验室 中国水产科学研究院黄海水产研究所 青岛 266071)

**摘要** 基于转录组 454 GS FLX 测序结果,利用 RACE 和 RT-PCR 技术克隆了海蜇(*Rhopilema esculentum*) Notch 基因的 cDNA 全长,并分析了其 mRNA 在海蜇不同发育阶段的表达差异,以探讨 Notch 基因对海蜇无性生殖的影响。结果显示,海蜇 Notch 基因的 cDNA 全长为 6768 bp,包括 90 bp 的 5'非编码区、6066 bp 的开放阅读框及 612 bp 包含 AATAAA 加尾信号的 3'非翻译区。SMART 分析表明,海蜇 Notch 为分泌蛋白,其信号肽由 21 个氨基酸组成。成熟肽由 2000 个氨基酸组成,包括 37 个结构域、26 个表皮生长因子样结构域、3 个富含半胱氨酸的 Notch/Lin-12(NL)结构域、1 个 NOD 结构域、1 个 NOD 结构域、1 个 PS膜结构域和 6 个锚蛋白结构域 ANK,并含有 25 个 N-糖基化位点。同源分析表明,海蜇 Notch 与来自刺胞动物门的海葵(*Nematostella vectensis*)的 Notch 相似性为 39%,与无脊椎动物和脊椎动物的氨基酸相似度分别为 35%—37%和 34%—38%。RT-PCR 分析表明,Notch 基因在海蜇无性繁殖的 4 个发育时期均有表达,螅状体阶段的表达量最高,横裂体阶段表达量最低,螅状体的表达量是横裂体的 1.85 倍。这些研究结果为进一步研究 Notch 信号通路在海蜇无性繁殖中的调控作用奠定了基础。

关键词 海蜇; Notch; cDNA; 实时荧光定量 PCR 中图分类号 Q71 文献标识码 A 文章编号 2095-9869(2017)04-0134-12

Notch 途径是一条进化上十分保守的信号通路, 通过受体 Notch 与配体 DSL 蛋白的相互作用转导信 号,从而调控细胞增殖、分化、凋亡并决定细胞的命 运(Lai, 2004;梁洁等, 2008)。作为 Notch 信号途径的 重要起始蛋白, Notch 基因最初在研究果蝇(*Drosophila melanogaster*)胚胎发育过程中被识别,它的突变导致 果蝇翅膀边缘出现缺口,因此被命名为 Notch (Mohr, 1919、1922)。随后,果蝇的 Notch 基因被克隆出来 (Allman *et al*, 2002)。目前,已有 Notch 的同源蛋白从 线虫(*Caenorhabditis elegans*)、爪蟾(*Xenopus laevis*)、斑 马鱼(Barchydanio rerio var)、鸡(Gallus gallus)、小鼠 (Mus musculus)和人(Homo sapiens)等生物中被分离 出来(齐润姿等, 2002; Anant et al, 1998)。研究表明, 从低等动物到高等动物,Notch 基因家族在结构上都 具有高度的保守性,均是单链跨膜蛋白,由胞外区、 跨膜区和胞内区组成,胞外区均具有多个 EGF-like 重复结构域、3 个 NL 结构域及 1 个 NOD 结构域和 1 个 NODP 结构域,而胞内区都具有 6 个 ANK 结构域 和 2 个核定位信号(NSL)(李宝园等, 2009; Kortschak et al, 2001; Maine et al, 1995)。在功能上,Notch 家族

<sup>\*</sup>国家自然科学基金项目(31372507)、青岛海洋科学与技术国家实验室鳌山科技创新计划项目(2016ASKJ02)和中央级 公益性科研院所基本科研业务费项目(20603022016001)共同资助 [This work was supported by the National Natural Science Foundation of China (31372507), the Aoshan S&T Innovation Project from Qingdao National Laboratory of Marine Science and Technology (2016ASKJ02), and the Central Public-Interest Scientific Institution Basal Research Fund (20603022016001)]. 骆晓 蕊, E-mail: luoxiaorui168@163.com

① 通讯作者:朱 玲,副研究员, E-mail: zhuling@ysfri.ac.cn 收稿日期: 2016-03-22、收修改稿日期: 2016-05-04

成员也具有高度的相似性,Notch 不但参与了骨骼、 神经和体节形成等多个发育过程(Zanotti *et al*, 2010; de Oliveira-Carlos *et al*, 2013; Artavanis- Tsakonas *et al*, 1999),而且还调控个体胚胎发育过程中胚层细胞的增 殖、分化和凋亡(Coffman *et al*, 1993; Conboy *et al*, 1995; Fortini *et al*, 1993),影响器官形成和形态发生等(郭政 等, 2008; 鲁茁壮等, 2004)。

Notch 信号途径在高等动物中已经被广泛研究, 但在刺胞动物(Cnidaria)中的研究很少,仅在水螅和海 葵(Nematostella vectensis)中有相关报道(Bottger et al, 2002; Kasbauer et al, 2007; Marlow et al, 2012)。刺胞动 物作为最原始的真后生动物,只有外胚层和内胚层的分 化,是其他高等多细胞动物的起点(周春娅等, 2013)。 因此,研究刺胞动物 Notch 信号途径,解析 Notch 基 因的分子结构和不同发育过程的表达特点,将为了解 Notch 信号途径的起源和功能演化等研究提供重要参考。

海蜇(Rhopilema esculentum)隶属于刺胞动物门、 钵水母纲、根口水母目、海蜇属,其人工繁育和增养 殖技术成熟,是研究刺胞动物无性繁殖调控途径的理 想模式生物。本研究基于转录组 454 GS FLX 测序结 果,利用 RACE 和 RT-PCR 技术,克隆了海蜇 Notch 基因的 cDNA 全长,并分析了其 mRNA 在海蜇无性 繁殖不同发育阶段的表达差异,这些研究结果将为进 一步研究 Notch 信号通路在海蜇无性繁殖中的调控 作用奠定了基础。

# 1 材料与方法

## 1.1 实验动物

野生性成熟海蜇于2015年9月采于浙江杭州湾, 运回山东荣成海蜇养殖场进行人工繁育,以获取海蜇 螅状体(Scyphistoma)、横裂体(Strobila)、碟状体(Ephyra) 和幼蛰样品。

## 1.2 海蜇转录组 454 GS FLX 测序与 EST 分析

海蜇转录组文库构建、454 GS FLX 测序见周春娅 等(2013)方法。利用生物信息学方法检索海蜇转录组 文库,寻找与已知 Notch 基因同源的 EST 序列。

### 1.3 Notch 基因 cDNA 的全长克隆

Blast 分析表明,海蜇 EST<sub>(isotig11938)</sub>与海葵 Notch 基因具有高度的相似性,进一步分析发现,这段序列具有 Notch 基因特有的 3 个 NL 结构域、1 个 NOD 结构域和 1 个 NODP 结构域。根据 EST<sub>(isotig11938)</sub>序列,设计特异 性引物扩增海蜇 Notch 基因 cDNA 全长(表 1)。Notch 基因的 3'末端利用 pBluescript SK(+/-)载体上的通用 引物 T7 与特异性引物 Notch-F2 扩增; Notch 基因的

5′末端用载体通用引物 T3 与特异性引物 Notch-R2 扩 增。25 μl 的 PCR 反应体系: 2.5 μl 10×PCR Buffer, 2 μl dNTP (25 mmol/L), 1.5 μl MgCl<sub>2</sub> (25 mmol/L), 1 μl 引物(10 μmol/L), 0.2 μl (1U)*Taq* 酶, 1 μl cDNA 模板, 15.8 μl PCR 水。扩增产物用琼脂糖凝胶电泳进行分离 检测,切胶回收后连接到 pMD18-T (TaKaRa)载体, 然 后转化至大肠杆菌感受态细胞 Top10 中,通过蓝白斑 筛选阳性克隆并送到上海英潍捷基测序。

#### 1.4 生物信息学分析

用 Bioedit 软件对所获得的测序结果序列进行全 长拼接;用 BLAST(http://www.ncbi.nlm.nih.gov/blast) 和蛋白分析系统(http://www.expasy.org/)分析海蜇 Notch的 cDNA 和氨基酸序列。用 SMART(http://smart. emblheidelberg.de/)及 SingalP 4.1 信号肽预测软件(http: //www.cbs.dtu.dk/services/SignalP/)对海蜇 Notch 进行 蛋白结构域分析及信号肽预测。利用 ClustalX1.83 将海 蜇 Notch 与其他物种 Notch 氨基酸序列进行多序列比 对。根据多序列比对结果,用 MEGA5.0 软件采用邻接 法(Neighbor-joining) 构建 Notch 的系统进化树。

## 1.5 荧光实时定量 PCR

利用 Trizol 法分别提取海蜇 4 个不同发育阶段: 螅 状体、横裂体、碟状体、水母体(Medusa)的总 RNA, 每个发育时期 15 个个体为 1 组,每组设 3 个重复。 然后分别反转录合成 cDNA,反应体系及反应条件按 说明书要求操作(Invitrogen)。

PCR 反应在 ABI 7500 Real-time PCR system (Applied Biosystems)上进行, 20 μl 反应体系: 10 μl SYBR<sup>®</sup> Premix Ex *Taq*<sup>TM</sup> II, 0.4 μl ROX Reference Dye II, 0.4 μl 特异性引物(Notch-F1 和 Notch-R1 见表 1), 2 μl cDNA 模板。PCR 反应程序如下: 95℃预变性 30 s, 95℃ 5 s, 60℃ 34 s, 共 40 个循环。样品和内参分别设 3 个重 复。采用  $\beta$ -actin 基因作为实时定量 PCR 的内参基因。反应结束后的分解步骤(95℃ 15 s, 60℃ 1 min, 95℃ 15 s)用于绘制熔解曲线,确认 Notch 基因引物的特异 性,采用 2<sup>-ΔΔCt</sup> 法进行数据分析。获得的数据采用 SPSS 18.0 统计软件进行处理,各组间平均数比较采 用单因素方差分析(One-way ANOVA), P<0.05 为差异 显著, P<0.01 为差异极显著。

## 2 结果与分析

# 2.1 海蜇 Notch 信号通路中 Notch 基因的克隆和序 列分析

将 3'-RACE 和 5'-RACE 获得的序列与 EST<sub>(isotig11938)</sub> 序列进行拼接,获得海蜇 Notch 基因的 cDNA 全长序

	Tab.1 Primer sequences used in the study	
引物名称 Primer name	序列 Sequence (5'-3')	作用 Application
Notch-F1	ACCAGAGTCACAGAAAAGCCC	Real-time PCR
Notch-R1	CACATCCCTGGGTGTTTGCT	Real-time PCR
Actin-F	AACTGGGACGATATGGAGAAGA	Real-time PCR
Actin-R	CGACCAGAGGCGTACAATGAG	Real-time PCR
Notch-F2	TGGATACAGTTGGACAAGATGATAAT	3'Race-PCR
Notch-R2	ATTTTGAGTATGGACCCTGTAGAGA	5'Race-PCR
Τ7	GTAATACGACTCACTATAGGGC	3'Race-PCR
Т3	AATTAACCCTCACTAAAGGG	5'Race-PCR

表1 本研究所用引物序列

列。海蜇 Notch 基因的 cDNA 全长为 6768 bp,包括 5'非翻译区(5'-UTR) 90 bp,包含 1 个多腺苷酸 Poly(A)尾和 AATAAA 加尾信号,3'非翻译区(3'-UTR) 612 bp,开放阅读框(Open reading frame, ORF) 6066 bp 编码了2021个氨基酸。预测其蛋白分子量和理论 等电点分别为 218.92 kDa 和 6.24。

通过 SignalP、SMART、PredictProtein 2013 和 NetNGlyc 1.0 Server 等软件分析海蜇 Notch 的氨基酸 序列显示, Notch 为分泌蛋白, 信号肽由21个氨基酸 组成(图 1), 酶切后成熟肽的分子量和理论等电点分 别为 216.55 kDa 和 6.25。在成熟肽中共有 37 个结构 域(图2),包括 26 个表皮生长因子样(EGF-like)重复结 构域、3 个富含半胱氨酸的 Notch/Lin-12(NL)结构域、 1 个跨膜结构域、1 个 NOD 结构域、1 个 NODP 结构 域和 6 个锚蛋白 ANK 结构域,还含有 25 个潜在的 N-糖基化位点,其中,NL、NOD 和 NODP 结构域是 Notch 家族的特有结构域。

## 2.2 海蜇 Notch 基因的相似性和系统进化分析

多序列比对结果显示(表2),在全序列上,海蜇Notch

1	CAAC	ſĊĊĠ	AAG	GAT	GGA	GAG	TAG	TGG	AAT	GTT	GCC	GTC	CAA	GCT	GAT	TAA	TTC	TCT	TGG	CAA	TCG	AGA	TAT	ACTA
76	TGAAA	AAAG	CAA	AGA	TAT	<b>G</b> GA	TCA	TAC	AGT	GTA	TAT	TTT	ATG	TTT	TGG	ATT	GCT	GGT	TTT	GGT	TAA	TTA	TGC	ATTA
1					M	D	H	Т	V	Y	Ι	L	С	F	G	L	L	V	L	V	N	Y	A	L
151	TCAT	CTTG	TCC	AAC	TGT	ATC	CAC	CTG	CCA	GAA	TGG	AGG	AAC	ATG	TAC	TGT	AGT	AAA	TGG	AAC	AGC	AAC	TTG	CAGC
21	S, S	С	Р	Т	V	S	Т	С	Q	Ν	G	G	Т	С	Т	V	V	Ν	G	Т	А	Т	С	S
226	TGCA	CACC	TCA	GTG	GAC	TGG	GGC	GAG	ATG	TGA	.CCA	GCC	TGT	TGA	TCG	TTG	TGC	TGA	TAA	ACC	CTG	CAA	AAA	TGGA
46	С Т	Р	Q	W	Т	G	А	R	С	D	Q	Р	V	D	R	С	А	D	K	Р	С	K	Ν	G
301	GGGA	CTTG	TAT	ТСТ	ATT	TCA	TTA	TTC	GCC	TTA	TTA	TAT	ATG	СТС	ATG	TGC	TGG	CGG	GTG	GAA	GGG	AGA	AAA	CTGT
71	G T	С	Ι	L	F	Н	Y	S	Р	Y	Y	Ι	С	S	С	А	G	G	W	K	G	Е	Ν	С
376	ACGG	ГСАА	TCA	TGA	TGA	CTG	CAG	TCC	AAA	CCC	CTG	CCA	AAA	CAA	TGG	AAC	CTG	TAT	TGA	TGG	GAT	TGG	AAA	GTAC
96	T V	Ν	Н	D	D	С	S	Р	Ν	Р	С	Q	Ν	Ν	G	Т	С	Ι	D	G	Ι	G	Κ	Y
451	GAAT	GTAG	CTG	TAA	AAC	AGG	ATT	TAC	AGG	TTT	AAC	ATG	CAG	СТА	TCG	AGA	CAG	TTG	TCA	ATC	TGG	ACC	TTG	TGGC
121	E C	S	С	K	Т	G	F	Т	G	L	Т	С	S	Y	R	D	S	С	Q	S	G	Р	С	G
526	ATCCA	ATGC	ACG	TTG	TGA	AGT	GGA	TGC	СТА	CGG	TCA	ATA	TTC	TTG	CAT	TTG	TCA	CAG	TGG	СТА	TAA	CGG	TAC	CAAT
146	I H	А	R	С	Е	V	D	А	Y	G	Q	Y	S	С	Ι	С	Н	S	G	Y	Ν	G	Т	Ν
601	TGTG	ATCA	AGA	TAT	AAA	CGA	ATG	СТА	CCC	AAA	CAA	TGG	CCC	ATG	CGG	ACA	TGG	GAC	ATG	TGT	GAA	CAC	AAA	TGGC
171	C D	Q	D	Ι	Ν	Е	С	Y	Р	Ν	Ν	G	Р	С	G	Н	G	Т	С	V	Ν	Т	Ν	G
676	TCAT	ATTA	CTG	CAA	CTG	TAA	AGT	TGG	ATA	CAC	TGG	GCG	AAG	ATG	TGA	AAT	TTT	AAT	AAA	CGA	GTG	TAA	CTC	GGCA
676 196	TCATA S Y	ATTA Y	CTG. C	CAA N	CTG C	TAA K	AGT V	TGG G	ATA Y	CAC T	TGG G	GCG R	AAG R	ATG C	TGA E	AAT I	TTT L	AAT I	AAA N	CGA E	GTG C	TAA N	CTC S	GGCA A
676 196 751	$\frac{TCAT}{S} \frac{S}{CCAT}$	ATTA Y GCGT	CTG C CCA	CAA N TGG	CTG C CAC	TAA K CTG	AGT V TAT	TGG G AGA	ATA Y TGA	CAC T CAT	TGG G TGG	GCG R AAA	AAG R ATA	ATG C TAC	TGA <u>E</u> TTG	AAT I CTC	TTT L CTG	AAT I TAA	AAA N GCC	CGA E AGG	GTG C ATA	TAA N CAC	CTC S TGG	GGCA A AGTA
676 196 751 221	TCATA S Y CCATO P C	ATTA Y GCGT V	CTG C CCA H	CAA N TGG G	CTG C CAC T	TAA K CTG C	AGT V TAT I	TGG G AGA D	ATA Y TGA D	CAC T CAT I	TGG G TGG G	GCG R AAA K	AAG R ATA Y	ATG C TAC T	TGA <u>E</u> TTG C	AAT I CTC S	TTT L CTG C	AAT I TAA K	AAA N GCC P	CGA E AGG G	GTG C ATA Y	TAA N CAC T	CTC S TGG G	GGCA A AGTA V
676 196 751 221 826	TCATA S Y CCATO P C AATTO	ATTA Y GCGT V GTGA	ICTG C CCA H IGAG	CAA N TGG G TGA	CTG C CAC T AGT	TAA K CTG C	AGT V TAT I CGA	TGG G AGA D GTG	ATA Y TGA D CCT	CAC T CAT I GTC	TGG G TGG G AAG	GCG R AAA K CCC	AAG R ATA Y GTG	ATG C TAC T CAA	TGA E TTG C GAA	AAT I CTC S TGG	TTT L CTG C CCA	AAT I TAA K GTG	AAA N GCC P TAT	CGA E AGG G TGA	GTG C ATA Y CAA	TAA N CAC T AAT	CTC S TGG G CAA	GGCA A AGTA V TGGT
676 196 751 221 826 246	TCATA S Y CCATO P C AATTO N C	ATTA Y GCGT V GTGA E	CTG CCA H GAG S	CAA N TGG G TGA E	CTG CAC T AGT V	TAA K CTG C CAA N	AGT V TAT I CGA E	TGG G AGA D GTG C	ATA Y TGA D CCT L	CAC T CAT I GTC S	TGG G TGG G AAG S	GCG R AAA K CCC P	AAG R ATA Y GTG C	ATG C TAC T CAA K	TGA E TTG C GAA N	AAT I CTC S TGG G	TTT L CTG C C C Q	AAT I TAA K GTG C	AAA N GCC P TAT I	CGA E AGG G TGA D	GTG C ATA Y CAA K	TAA N CAC T AAT I	CTC S TGG G CAA N	GGCA A AGTA V TGGT G
676 196 751 221 826 246 901	TCATA <u>S</u> Y CCATO <u>P</u> C AATTO <u>N</u> C TTCCA	ATTA Y GCGT V GTGA E AGTG	CTG CCA H GAG S TAA	CAA N TGG G TGA E ATG	CTG CAC T AGT V CCA	TAA K CTG CAA N GTG	AGT V TAT I CGA E GGG	TGG G AGA D GTG C ATA	ATA Y TGA D CCT L CAC	CAC T CAT I GTC S TGG	TGG G TGG G AAG S TGC	GCG R AAA K CCC P AAC	AAG R ATA Y GTG C CTG	ATG C TAC T CAA K CGA	TGA E TTG C GAA N CAC	AAT I CTC S TGG G AAA	TTT L CTG C CCA Q TAT	AAT I TAA K GTG C AAA	AAA N GCC P TAT I CTA	CGA E AGG G TGA D .CTG	GTG C ATA Y CAA K TGC	TAA N CAC T AAT I CAT	CTC S TGG G CAA N AAA	GGCA A AGTA V TGGT G CCCC
676 196 751 221 826 246 901 271	TCATA S Y CCATO P C AATTO N C TTCCA F Q	ATTA Y GCGT V GTGA E AGTG C	CTG CCA H GAG S TAA K	CAA N TGG G TGA E ATG C	CTG CAC T AGT V CCA Q	TAA K CTG CAA N GTG W	AGT V TAT I CGA E GGG	TGG G AGA D GTG C ATA Y	ATA Y TGA D CCT L CAC T	CAC T CAT I GTC S TGG G	TGG G TGG G AAG S TGC A	GCG R AAA K CCC P AAC T	AAG R ATA Y GTG C C TG C	ATG C TAC T CAA K CGA D	TGA E TTG C GAA N CAC T	AAT I CTC S TGG G AAA N	TTT L CTG C CCA Q TAT I	AAT I TAA K GTG C AAA N	AAA N GCC P TAT I CTA Y	CGA E AGG G TGA D CTG C	GTG C ATA Y CAA K TGC A	TAA N CAC T AAT I CAT I	CTC S TGG G CAA N AAA N	GGCA A AGTA V TGGT G CCCC P
676 196 751 221 826 246 901 271 976	TCATA S Y CCATO P C AATTO N C TTCCA F Q TGTCA	ATTA Y GCGT V GTGA E AGTG C AAAA	CTG CCA H GAG S TAA K CAA	CAA N TGG TGA E ATG ATG C	CTG CAC T AGT V CCA Q CAC	TAA K CTG CAA CAA N GTG W	AGT V TAT I CGA E GGG G CAT	TGG G AGA D GTG C ATA Y TGA	ATA Y TGA D CCT L CAC T TGG	CAC T CAT I GTC S TGG G ACT	TGG G TGG G AAG S TGC A TGA	GCG R AAA K CCCC P AAC T TCG	AAG R ATA Y GTG C C C C ATA	ATG C TAC T CAA K CGA D TAC	TGA E TTG C GAA N CAC T CTG	AAT I CTC S TGG G AAA N TCG	TTT L CTG C CCA Q TAT I ATG	AAT I TAA K GTG C AAA N TTT	AAA N GCC P TAT I CTA Y GCC	CGA E AGG G TGA D CTG C C AGG	GTG C ATA Y CAA K TGC A TTA	TAA N CAC T AAT I CAT I CAC	CTC S TGG CAA N AAA N AGG	GGCA A AGTA V TGGT G CCCC P GCTT
676 196 751 221 826 246 901 271 976 296	TCATA S Y CCATC P C AATTC N C TTCCA F Q TGTCA C Q	ATTA Y GCGT V GTGA E AGTG C AAAA N	CCA CCA H GAG S TAA K K CAA N	CAA N TGG TGA E ATG ATG C TGG G	CTG CAC T AGT V CCA Q CAC T	TAA K CTG CAA CAA N GTG W CTG C	AGT V TAT I CGA E GGG G CAT I	TGG G AGA D GTG C ATA Y TGA D	ATA Y TGA D CCT L CAC T G G	CAC T CAT I GTC S TGG G ACT L	TGG G TGG G AAG S TGC A TGA D	GCG R AAA K CCCC P AAC T TCG R	AAG R ATA Y GTG C C C TG C TG ATA Y	ATG C TAC T CAA K CGA CGA TAC T	TGA E TTG GAA GAA N CAC T CTG C	AAT I CTC S TGG G AAA N TCG R	TTT L CTG C CCA Q TAT I ATG C	AAT I TAA GTG C AAA N TTT L	AAA N GCC P TAT I CTA Y GCC P	CGA E AGG G TGA D CTG C C C G	GTG C ATA Y CAA K TGC A TTA Y	TAA N CAC T AAT I CAT I CAC	CTC S TGG CAA N AAA N AGG G	GGCA A AGTA V TGGT G CCCCC P GCTT L
676 196 751 221 826 246 901 271 976 296 1051	TCATA S Y CCATC P C AATTC AATTCA TTCCA F Q TGTCA C Q GACTC	ATTA Y GCGT V GTGA E AGTG C AAAA N GTGA	CTG CCA H GAG S TAA K CAA N AAA	CAA N TGG G TGA E ATG C TGG G CGA	CTG CAC T AGT V CCA Q CAC T AAT	TAA K CTG CAA N GTG GTG CAA	AGT V TAT I CGA E GGG G G CAT I CGA	TGG G AGA D GTG C ATA Y TGA D GTG	ATA Y TGA D CCT L CAC T TGG G TGC	CAC T CAT I GTC S TGG G C ACT L GTC	TGG G TGG G AAG S TGC A TGA D	GCG R AAA K CCCC P AAC T TCG R CCC	AAG R ATA Y GTG CTG CTG CTG X TTG	ATG C TAC T CAA K CGA D TAC T TGT	TGA E TTG GAA N CAC T CTG CCA	AAT I CTC S TGG G AAA N TCG R TGG	TTT L CTG C CCA Q TAT I ATG C CAG	AAT I TAA GTG C AAAA N TTT L CTG	AAA <u>N</u> GCC <u>P</u> TAT <u>I</u> CTA <u>Y</u> GCC <u>P</u> TTT	CGA E AGG G TGA D CTG C C C C AGG G	GTG C ATA Y CAA K TGC A TTA Y CCG	TAA N CAC T AAT I CAT. I CAC. T TGT	CTC S TGG. G CAA N AAA N AGG G GAA	GGCA A AGTA V TGGT G CCCC P GCTT L CGAA
676 196 751 221 826 246 901 271 976 296 1051 321	TCAT/       S     Y       CCATC     P       P     C       AATTC       M     C       TTCC/     F       Q     TGTC/       C     Q       GACTC     D	ATTA Y GCGT V GTGA E AAGTG N N GTGA E	CTG CCCA H GAG S TAA K CAA N AAA N	CAA N TGG G TGA E ATG C G CGA E	CTG CAC T AGT V CCA Q CAC T AAT I	TAA K CTG C CAA GTG GTG C CAA N	AGT V TAT I CGA GGGG G GGGG CAT I CGA E	TGG G AGA D GTG C ATA Y TGA D GTG C	ATA Y TGA D CCT L CAC T G G TGC A	CAC T CAT I GTC S TGG G C TGG G C S	TGG G TGG G AAAG S TGC A TGA D XAAA N	GCG R AAA K CCCC P AAAC T TCG R CCCC P	AAG R ATA Y GTG C CTG C CTG C ATA Y TTG C	ATG C TAC T CAA K CGA D TAC T TGT V	TGA E TTG C GAA N CAC T CTG C CCA H	AAT I CTC S TGG G AAAA N TCG R TGG G	TTT L CTG C CCA Q CCA TAT I ATG C CAG S	AAT I TAA K GTG C AAAA N TTT L CTG C	AAA N GCC P TAT I CTA Y GCC P TTT L	CGA E AGG G TGA D CTG C C C G C AGG G D	GTG C ATA Y CAA K TGC A TTA Y CCG R	TAA N CAC T AAT I CAT. I CAC. T TGT V	CTC S TGG G CAA N AAAA <u>N</u> AGG G GAA N	GGCA A AGTA V TGGT G CCCCC P GCTT L CGAA E
676 196 751 221 826 246 901 271 976 296 1051 321 1126	TCAT/I   S Y   CCATO   P C   AATTO   N C   TTCC/I   F Q   TGTC/I   C Q   GACTO   D C   TACTO	Y Y GCGT V GTGA GTGA NAAAA N AAAAA N E GTGA GTGA	CTG CCA CCCA H GAG S TAA K CAA N CAAA N N TGT	CAA N TGG G TGA E ATG C G CGA E ATG	CTG CAC T AGT V CCA Q CAC T AAT I CCC	TAA K CTG C CAA N GTG GTG C CAA N N AAG	AGT V TAT I CGA E GGGG G CAT I CGA E TCA	TGG G AGA D GTG C ATA Y TGA D GTG C GTA	ATA Y TGA D CCT L CAC TGG G TGC A CAC	CAC T CAT GTC S TGG G ACT L GTC S TGG	TGG G TGG G AAAG S TGC A TGA D C AAAA N TAC	GCG R AAAA K CCCC P AAAC TCG R CCCC P TAA	AAG R ATA Y GTG CTG CTG ATA Y TTG C CTG	ATG C TAC T CAA K CGA K CGA T TAC T TAC T TAA	TGA <u>E</u> TTG C GAA N CAC T CTG <u>C</u> CCA H TAT	AAT I CTC S TGG G AAAA N TCG R TGG G CGA	TTT L CTG C CCA Q TTAT L ATG C CAG S TGT	AAT I TAA K GTG C AAAA N TTT L CTG C C AAAA	AAA N GCC P TAT I CTA Y GCC P TTT L AGA	CGA E AGG G TGA D CTG C C C G AAGG G AAGA D	GTG C ATA Y CAA K TGC A TTA Y CCCG R TGA	TAA N CAC T AAAT I CAAT I CAC T TGT V CAC	CTC S TGG CAA N AAAA <u>N</u> AAAA GAA N GAA	GGCA A AGTA V TGGT G CCCC P GCTT L CGAA E CCCA
676 196 751 221 826 246 901 271 976 296 1051 321 1126 346	TCAT/     S     Y       S     Y     CCATC       P     C     AATTC       M     C     TTCC/       F     Q     TGTC/       C     Q     GACTC       D     C     TACTT       Y     L     L	ATTA Y GCGT GTGA E AGTG C C GTGA E C C	CTG CCA H GAG S TAA K CAA N AAAA N TGT V	CAA N TGG G TGA E ATG C C CGA <u>E</u> C ATG C	CTG CAC T AGT V CCA CAC T AAT I CCCC P	TAA K CTG C CAA M GTG GTG C CAA N AAG S	AGT V TAT I CGA GGGG G CAT I CGA E TCA Q	TGG G AGA D GTG C ATA Y TGA D GTG C GTA Y	ATA Y TGA D CCT L CAC TGG G TGC A CAC T	CAC T CAT I GTC S TGG G C C TGG G C C	TGG G TGG G AAG S TGC A TGA D AAAA N TAC T	GCG R AAAA K CCCC P AAAC T CCC R CCCC P TAAA N	AAG R ATA Y GTG C CTG C C C C C C C C C C C C C	ATG CAA CAA CAA CGA D TAC TGT V TAA N	TGA <u>E</u> TTG C GAA N CAC T CCG <u>C</u> CCA <u>H</u> TAT I	AAT I CTC S TGG G AAAA N TCG R CGA D	TTT L CTG CCCA Q TATG CCAG S CAG S V	AAT I TAA K GTG C AAAA N TTTT L CTG C C AAAA K	AAA N GCCC P TAT I CTA GCCC P TTT L AGA E	CGA E AGG G TGA D CTG C C AGG G AGA D ATG C	GTG C ATA Y CAA K TGC A TTA Y CCCG R TGA D	TAA N CAC T AAT I CAT CAC T TGT V CAC T	CTC S TGG CAA N AAAA M AAGG G AAA N GAA N	GGCA A AGTA V TGGT G CCCC P GCTT L CGAA E CCCA P
676 196 751 221 826 246 901 271 976 296 1051 321 1126 346 1201	TCAT/   S Y   CCATO   P C   AATTO   M C   TTCC/   F Q   TGTC/   C Q   GACTO   D C   TGTATO   Y L   TGTAA	ATTA Y GCGT GTGA E AGTGA C GTGA C C C C AAAAA	CTG CCA GAG GAG S TAA K CAA N AAAA N TGT V CGA	CAA N TGG G TGA E ATG C CGA E CGA E ATG C CAA TGG C TGC	CTG CAC T AGT V CCA C CCA T CCC T AAT I CCCC P AAA	TAA K CTG CAA CAA GTG CAA CAA CAA AAG S TTG	AGT V TAT I CGA E GGGG G CAT I CGA E TCA Q TAT	TGG G AGA D GTG C ATA Y TGA D GTG C GTA Y CGA	ATA Y TGA D CCT L CAC T TGG G TGC A CAC T ACT	CAC T CAT I GTC S TGG G GTC S TGG G TTGG G TAT	TGG G TGG G AAAG S TGC A TGA D TAAA N TAC T CGG	GCG R AAAA CCCC P AACC T TCG R CCCC P TAA N AGG	AAG R ATA Y GTG C C T G C G C C G C C C G C C C C C	ATG C TAC T CAA K CGA D TAC T TAC T TAA N TAA	TGA <u>E</u> TTG C GAA T CAC C CCA C CCA H TAT I TTG	AAT I CTC S TGG G AAAA N TCG R TGG G CGA D CGA	TTT L CTG CCA Q CCA TTAT L CA CA G S TTGT V CTG	AAT I TAA K GTG C AAAA N TTTT L CTG C C C C AAAA K TAA	AAA N GCCC P TAT I CTAT GCCC P TTT L AGA E ACC	CGA E AGG G TGA D CTG C C C C AGG D ATG C TGG	GTG C ATA Y CAA TGC A TTA Y CCG R TGA D ATT	TAA N CAC T AAT I CAT I CAT T CAC T CAC T CAC T CAC	CTC S TGG CAA N AAAA N AAGG GAA N GAAA N AGG	GGCA A AGTA V TGGT G CCCC P GCTT L CGAA E CCCA P TGTC
676 196 751 221 826 246 901 271 976 296 1051 321 1126 346 1201 371	TCAT/ S Y CCATO P C AATTO N C TTCC/ F Q TGTC/ C Q GACTO D C TACT' Y L TGTA/ C K	ATTA Y GCGT GTGA E AGTGA C C C AAAAA N N	CTG CCA CCCA GAG S TAA K CAA N AAA N TGT V CGA D	CAA N TGG G TGA E ATG C CGA E CGA E CGA CGA C CAA	CTG CAC T AGT CCA CCA CCA CAC T CCC AAT I CCCC P AAA N	TAA K CTG CAA N GTG W GTG CAA K AAG S TTG C	AGT V TAT CGA E GGGG G CAT I CGA E TCA Q TAT	TGG G AGA D GTG C ATA Y TGA D GTG C GTA Y CGA E	ATA Y TGA CCT L CAC T G G TGC A CAC T CAC T A CAC L	CAC T CAT I GTC S TGG G C TGG G TGG G TAT I	TGG G TGG CAAG S TGC A TGA D CAAA N TAC T CGG G	GCG R AAAA CCCC P AACC T CCCC P TAA N AGG G	AAG R ATA Y GTG C CTG C CTG C C TTG C TTA Y	ATG C TAC T CAA CGA D TAC T TGT T TAA N TAA N	TGA <u>E</u> TTG C GAA CAC T CCA CCA H TAT I TTG C	AAT I CTC G TGG G AAA N TCG R CGA CGA S	TTT L CTG CCA Q TAT I CAG S TGT V CTG C CTG C	AAT I TAA GTG C AAAA N TTT L CTG C C AAAA K TAA K	AAA N GCC P TAT I CTA Y GCC P TTT L AGA E ACC P	CGA E AGG G TGA D CTG C C AGG G ATG C TGG G	GTG C ATA Y CAA TGC A TTA Y CCG R TGA D ATT F	TAA N CAC T AAT I CAT T CAC T TGT V CAC T CAC T TAC T	CTC S TGG CAA N AAAA N GAA S GAA N GAA N GAA S GAA S GAA S GAA S GAA S G G A G G G C C C C C C C C C C C C C C	GGCA A AGTA V TGGT G CCCC P GCTT L CGAA E CCCA P TGTC V

396 N C E I N I N E C A S N P C L Q G S R C T D G 1351 AGTTTCGTTTGCGAATGTAAACTTGGATACACTGGCAAAATTTGTGAAACACAGATCAACGAATGCCTGTCTAAT 421 <u>S F V C E C K L G Y T G K I C E</u> T <u>Q I N E C L S</u> 1426 CCCTGCATGAATGGTGCAACGTGCTTTGACCGACTGGGATTTTATGACTGCACATGCGCACCGGGCTACACTGGA 446 P C M N G A T C F D R L G F Y D C T C A P G Y T 1501 AAGCGTTGTGAAACAAGAAACGATAACTGCATATACAACCAGTGCCAAAACGGTGGAACTTGTCAAAGCAACCCG 471 <u>K R C E</u> T R N D <mark>N C I Y N Q C Q N G G T C Q S N P</mark> 1576 ACAGGGTATACTTGTACGTGCCCGTCAGGGTATTCAGGGACATTCTGCGAAAATAATATCAACGAATGCACGCCA 496 T G Y T C T C P S G Y S G T F C E N <u>N I N E</u> 1651 AACCCCTGTGTGCATGGCAACTGCACTGATCTGGCCTATGATTATCATTGTATTGTAAAGATGGCTATACCGGC 521 N P C V H G N C T D L A Y D Y H C I C K D G Y T G 546 K N C T K D V D D C Y I G A C F N N G Q C I D G 1801 AACACGTTCCGGTGCATCTGCCCTGTTGGTTACACTGGCAAGCAGTGTAGCAAGGAAATAGATGAATGCAGTCTG 571 <u>N T F R C I C P V G Y T G K Q C S K E I D E C S L</u> 1876 AGACCCTGCCAGAATGGTGGAAAATGCCACGACAAAATTGGCAGTTATGTCTGTGAATGTCCCTTCGGTTTCACG 596 R P C Q N G G K C H D K I G S Y V C E C P F G F 1951 GGAACTAATTGCGAATCTTACGTAAACTGGTGCAGTCAGGGTATATGCCAGAATGGAGCCACCTGCTCCCAAGTT 621 <u>G T N C E</u> S Y V N W C S Q G I C Q N G A T C S Q V 2026 GGTGCCAATTTTTCATGTAGCTGTAGCAATGGATGGACAGGAGCGATTTGTGATGTACCAAAAATATCTTGTGCT 646 G A N F S C S C S N G W T G A I C D V P K I S C A 2101 CAAGCACAAGCGAGAGGACTAGGCTGTCATAATGGCGGTACATGTGTAAATACCACTTATGGGCATTTCTGTAGG 671 Q A Q A R G L G C H N G G T C V N T T Y G H F C R 696 CASGYTGSHCQININECASHPCYHG 2251 GGGACATGCAAAGATCTGTTGAATGGATACTCATGCACATGTCGAAGTGGGTTCTCAGGGCCTAACTGCAACGTA 721 G T C K D L L N G Y S C T C R S G F S G P N C N V 2326 CGTATCGACGAATGCACTTCAAGGCCATGCATGAATGGTGGTACGTGCGAAGATGTTGTCAATGGTTACATATGT 746<u>RIDECTSRPCMNGGTCEDVVNGYIC</u> 2401 AGATGCCCAACAGGTTTTGAAGGGCCCAAACTGTGAACGAAGACATAATTACTGTCATAAAGGTGCATGTTACAAT 771 <u>R C P T G F E G P N C E</u> R R H N Y C H K G A C Y N 2476 AATGGAATTTGCATGTCTGGCACCAACTGGTACCGCTGCACATGTCCTGTTGGATATACTGGTCCACGTTGTGAG 796 N G I C M S G T N W Y R C T C P V G Y T G P R C E 2551 ACTGATTTTAACGAATGTCTACAAGGTCGTTGTTATGAATATGGCACAAAGCAGTGTATTAACACTCCAGGTAGT 821 T D F N E C L Q G R C Y E Y G T K Q C I N T P G S 2626 TATGCCTGTATTTGTAAAAGCGGTTTCGTTGGAGAAAACTGCGAAATTAATCAAGATGACTGTTTGAGTGAACCC 846 Y A C I C K S G F V G E N C E I N Q D D C L S E P 2701 TGCTTGAACTCTGGATATTGTATAGATGGAATAAACAATTACACATGCCACTGCCCCAGAGGTTACAGTGGAAGA 871 C L N S G Y C I D G I N N Y T C H C P R G Y S G 2776 AACTGCGAACGAGTTGTGCGCACCTGTAGTGACAACCCGTGCCAAAATGCTGCCAAATGCTCAATGGTGGGCGAC E R V V R T C S D N P C Q N A A K C S M V G D 896 N C 921 K Y M C H C K V G F E G T H C E Y N E D A C L S 2926 CCCTGTGCTCATGGCTCTTGCGTCGATACAGCGGGTGGCTATGAATGCACATGCAGACCCGGCTGGCGTGGGAAG 946 P C A H G S C V D T A G G Y E C T C R P G W R G K 3001 AACTGCTCGGAAAGTCTCAACGAATGCGCCTCGAATCCATGTTTGAACAATGGCATATGCTCTGTCTCTCAGA 971 <u>N C S E S L N E C A S N P C L N N G I C S V S L R</u> 3076 GGTTTTACATGCATGTGTAAAGCTGGATTCACGGGAGATCGATGTGAAAAAGTAAAGGCTTCAAATGGCACATCC 996 G F T C M C K A G F T G D R C E K V K A S N G T S 3151 TGTCCCCACAGCGACTGTTCAACAAAGTTTGACGGTGGACAATGCTTGGAGAAATGTAACAACCAAGAGTGTCAG 1021 C P H S D C S T K F D G G Q C L E K C N N Q E C Q 3226 TGGGACGGAACGGACTGCACACTAAGCCGTGATCCATGGCTCAACTGCAGTGTTGTTGACGTAGAATGCTGGAAA 1046 W D G T D C T L S R D P W L N C S V V D V E C W K 3301 CACTTTGATAATGGCGTTTGTGATAAAAAATGTAACAATTCCGGTTGCCTCTTCGATGGTCACGACTGCTTCAGT 1071 <u>H F D N G V C D K K C N N S G C L F D G H D C F</u> S 3376 AATGTGGGCACTTGCAAGTATCCAGTTTTTTGTGAAGCTACTTATGGCAATGGAAAATGCGACGAATATTGCAAC 1096 N V G T C K Y P V F C E A T Y G N G K C D E Y C N 3451 AACGATGCCTGAGGTTATGACAACTTAGAATGTGGAAGCAACAAGCAAAAATTGCTACCTGGAATTTTGAAAGTG 1121<u>NDATGYDNLEC</u>GSN<u>KQKLLPGILKV</u> 1146 V I S A S E E E V R R L A R L F V R N L S I K L 1171<u>THVYILKDENNRDRIRALTN</u>DVTSA 1196 R K R R D A S L F A L I S H S R V K R Q A G K <u>N R</u>

3751 ACGGAGATACTGCTGGGTGTTGACAACCAAAAGTGTACGCAAGGATGCTTTACTGATGTAGACCTTATCGCCAAC 1221 T E I L L G V D N Q K C T Q G C F T D V D L I A 3826 TACATTGGAGCATTACAATCGAAAGGAAAGTTGGATTCGAACCTTCCAAATTTATTCTGTCACAGGGGACAAGGCT 1246 Y I G A L Q S K G K L D S N L P I Y S V T G D K A 3901 AAAGTTGAAAAGGAACGAACTAAGGAGCCTCTGGCCCCACTCTGGATCTTGCTTATTTGCATAAGTGTTCCATTG 1271<u>KVEKER</u>TKEPLAP**LWILLICISVPL** 3976 CTGGCATTTTCTGTCCTTGTTGGTGCTAAGAAGTGTTCCAGTCGTCTGTGGGTGCCACGTGGATTTGTTCTTTCA 1296 <u>LAFSVLV</u>GAKKCSSRLWVPRGFVLS 4051 AGTGGATTGGATCAGGGAAACCTGCCAAGCACAGACAGATCCGATCCAGTTGGCCAGGAATATAGCTTGAAGTCG 1321 S G L D Q G N L P S T D R S D P V G Q E Y S L K S 1346 L K P G D R S S A E S I A S A A C S W N T P P S P 4201 TCGAGCAGCATTCACGATCAGAAACTGGTATGCTTAGAACCGTCACTTGAAGGATGTAAACAAATCCACGACGAC 1371 S S S I H D Q K L V C L E P S L E G C K Q I H D D 4276 AGAACTTGGACAAGCCTACACAGAGAGGCGGCAGATTCACGCCATTTGGGACTTACGCCACCGCAGGGCTCGTTT 1396 R T W T S L H R E A A D S R H L G L T P P Q G S F 4351 GCTTCTTCTGGTCGCTACCCAGATATCAATGTGAGAGGACCTGGGGGTTTGACTCCGCTTCACGTAGCCGTTTGC 1421 A S S G R Y P D I N V R G P <u>G G L T P L H V A</u> V C 4426 CGTTCGACTGATTTTAGGGATAGCTATGATGATGACAAGGACAGTGAGGATAGTGCAGAAAAGCGAAAACTCCTT 1446 R S T D F R D S Y D D D K D S E D S A E K R K L L 4501 GAAGGTGGTGCTGATGCCAACGTGCGTGATATTCATGGACGAAGTCCGTTGCATTTGGCAGTATCGGCAGATGCT 1471 E G G A D A N V R D I H G R S P L H L A V S A D A 4576 GTTGGAGCATTTAAGCTTCTGATTAAGCACCGAAGTACCATAATTGATGCTCAGAGTTTCGATGGAAGCACTCCA 1496 <u>V G A F K L L I K H R S T I I D A</u> Q S F <u>D G S T P</u> 4651 TTGATGGATGCTTGCAGGTTTGAAGTGAATTCAATGGTCGAAGATCTTATAAACACTGGAGCCAAAGTTAATATA 1521 L M D A C R F E V N S M V E D L I N T G A K V N I 4726 ACAGATAATCAAGGACGAACCGCATTGCATTGGGCCGTATCTGTTGATAATGAAATAGCGACGATGATATTGACA 1546 T D N Q G R T A L H W A V S V D N E I A T M I L T 1571 <u>R N G T K V D A</u> A D N <u>K G Q T A L Y L A A R E G S</u> 4876 TTGAAGTCTGCAAAGATTCTTCTACTTAATTTTGCCAACAGAAATGCAGCAGACAACATGGATCAATCTCCCATT 1596 <u>LKSAKILLLNFANRNA</u>ADN<u>MDQSP</u> 4951 CAAATTGCTCGGGAGCGAGGCCACCATGATATTGTCGAGTTAATATCTGATTGGACCATAGGGGCTAACTCCCCT 1621 Q I A R E R G H H D I V E L I S D W T I G A N S P 5026 CCGAAAGTTCCAGGTCCCACTTCACCAGAGTCACAGAAAAGCCCGCATATTCACCATGGCCACCCCACACCGCCG 1646 P K V P G P T S P E S Q K S P H I H H G H P T P P 5101 CGAAGTACAACTTCCCCGCCGAACATGATCCGACCAAAGATTAACAATATAGCACCAAAGATCCCAAGATCTCAT 1671 R S T T S P P N M I R P K I N N I A P K I P R S H 5176 ATACCAAGGTCTCACGCTCATGCAGCCCACACACAAGCAAACACCCAGGGATGTGCGCCAACTGTCACCGACAAA 1696 I P R S H A H A A H T Q A N T Q G C A P T V T D K 5251 TGTCTCAACGGTGGTTCAAGGCGACCACCAAACAAGCGAAAGCGCAAGTCCTGTACCAAGGCTGCTCAGCACCCT 1721 C L N G G S R R P P N K R K R K S C T K A A Q H P 5326 CCAAGGGCCCAAAATTAATGGATATGACATGAATTTGAATGGTACATCAGTAAATGGGTACGCACCAAATATATCT 1746 P.R.A.K.I.N.G.Y.D.M.N.L.N.G.T.S.V.N.G.Y.A.P.N.I.S. 5401 ATGTCACATGCATTAACCACAGCACCTACGCTGTCACCACCACGCGCAGAGGTTCCACCAAATCGCCTTGTTAAA 1771 M S H A L T T A P T L S P P R A E V P P N R L V K 5476 GTGCCTGTTGAGCGAAACTCAGGTCATGGAAACAGAGGTTCTCTTCCTGAACTTTCGGAAAAAGATATAATGGAG 1796 V P V E R N S G H G N R G S L P E L S E K D I M E 5551 GGCCTGTCGTTGTTTGCAACACATGGTTGTTTGGAAGACTTTCCGCCTAACTGGGATAACGAGGAGACTGCCTTA 1821 G L S L F A T H G C L E D F P P N W D N E E T A L 5626 AACCCCCAGGCTCATCATCAAAGTGTCCCCGAACATAAATGGGAACGATTCTTTAAGCATAAGTGCTTCACTACCA 1846 N P Q A H H Q S V P N I N G N D S L S I S A S L P 5701 AACTCGTCCATAAGGCTTAATGACGTTAATATGCAAATTGTCTCACACAATAATGGTGCTAGACCAAAGTTTCCT 1871 N S S I R L N D V N M Q I V S H N N G A R P K F P 5776 CGAAGCAGTGACAACGATCCTATGTCTGCAAACCATTATCTACACATGATGTCATCAGGTGATAATACGTGCGGG 1896 R S S D N D P M S A N H Y L H M M S S G D N T C G 5851 ACGGAAGCAATAACGTATTCACATGATATGCAGTTGCATATGTCAAACGATTTTGAGTATGGACCCTGTAGAGAT 1921 T E A I T Y S H D M Q L H M S N D F E Y G P C R D 5926 ACTCATTTACAAAGACAATTGCACCCAGAAGGGGTGTCGTTCTTGCAGCAGTTCCCCACTCCCCATCAACGCAC 1946 THLQRQLHPEGVSFLQQFPTPPSTH 6001 TCAAGTGCATATGTTTCATCTCCCGGCAAAAGTATATCACCGCAAGACGGTAACGTGGCAGCTAGTTTCTTGACA 1971 S S A Y V S S P G K S I S P Q D G N V A A S F L T 6076 CCGTCTCCTGATTCACCCAAGCGCAGTCCAGGGGGGCTGGTCCACGTCTCCACAGTCATCTTCAGAATCTTCAGTC 1996 PSPDSPKRSPGGWSTSPQSSSESSV 6151 TGTTAAGTGGTTAGACAGCTAGGCAATACTGTCTATATGCGAGGGGAGCGAACTGCAGACAGTCTCAGTGGGATC 2021 C \*

6226	TTGCAGCCAATGCTTCTGTATCCATATACAATAGAAGACATGTGCTGCTGCTGCGCGCTTAATGTTGAACGTTTTAA
6301	${\tt GACGTTCCAAGTTATTTGTACGATTTCATTTTGTTGTGAATCTCACAAAAAACGATTGGTCTTATTGAACTATTG$
6376	GAAAGTTGAAGACCAGCCTGCTTTGCAAGTCAGGATTTCTTCTGATATTCACAAGGAGAACAATTTGATATGAGA
6451	CTGAATACAATGTATATAGTTATTTTGTCGTTGCTCATATAATTTGACATCCAGCTATGTTAGACAAAGGAGTAG
6526	${\tt TTCGTCACTGCCATAACACAGAAGTGTCCATTTGGTCAGTGAATTGAATATTGCTTTTGATTGTAGTGCTACAAT}$
6601	GTTCTTACACAATGCTCCATTAAAATGCCTAATATTGTGGCCCCATTCACACTGGCATCAGGACAGCTGAAATAA
6676	TAAAAAATTGAAATAAAAAATATTATGAGAAATAATAATA
6751	АААААААААААААААА

#### 图 1 海蜇 Notch 全长 cDNA 序列及推导的氨基酸序列

Fig.1 Full-length Notch cDNA of R. esculentum and its deduced amino acid sequences

粗体为信号肽;箭头为酶切位点;灰色阴影为 Notch 结构域 EGF-like;双下划线为结构域 EGF-CA;下划波浪线为结构域 NL;下划点状线为结构域 Pfam:NOD;下划\_\_\_为结构域 Pfam:NODP;粗体并双下划线为跨膜结构域;下划实线为结构 域 ANK;方框为多聚腺苷酸加尾信号;\*为终止密码子

Black bold indicated signal peptied; Arrow indicated cleavage sites; EGF was printed with grey background; EGF-CA was lined with double underline; NL was underlined with wavy line; Pfam: NODP was underlined with long point line; Pfam: NOD was underlined with dotted line; Transmembrane domain was black bold and underlined with double underline; ANK was under lined with single underline; Poly (A) signals was surrounded by box; The pentagrams indicated stop codon



图 2 预测的海蜇 Notch 主要结构域

Fig.2 Predicted Notch domains of R.esculentum

## 表 2 海蜇 Notch 与其他物种的 Notch 的氨基酸序列一致性

Tab.2 Comparison of Notch amino acid identity between R. esculentum Notch and vertebrates/invertebrates

GenBank 注册号	拉丁学名	中文名	氨基酸序列的一致性
GenBank Accession	No. Latin name	Chinese name	Identity of amino acid sequence (%)
NP060087	H. sapiens	人	36
EHH23670	Macaca mulatta	猕猴	36
NP001099191	Rattus norvegicu	褐家鼠	34
EGW12778	Cricetulus griseus	中国仓鼠	34
NP032740	M. musculus	小鼠	34
XP004285023	Orcinus orca	虎鲸	38
XP00562549	Canis lupus familiaris	家犬	37
NP001099191	Erinaceus.europaeus	刺猬	36
NP001025466	G. Gallus	原鸡	38
XP005529261	Pseudopodoces humilis	地山雀	38
NP001090757	X. tropicalis	非洲爪蟾	38
BAC41349	Cynops pyrrhogaster	红腹蝾螈	35
XP008274642	Stegastes partitus	深裂眶锯雀鲷	35
XP008334545	Cynoglossus.semilaevis	半滑舌鳎	37
CDQ87512	O. mykiss	虹鳟鱼	38
XP007253900	Astyanax mexicanus	墨西哥脂鲤	38
XP006640602	Lepisosteus oculatus	眼斑雀鳝	38
XP011267598	C. floridanus	佛罗里达弓背蚁	37
XP006568052	A. mellifera	意大利蜜蜂	36
XP008204188	N. vitripennis	蝇蛹金小蜂	35
XP006818779	S. kowalevskii	囊舌虫	35
AEW42991	N. vectensis	海葵	39

展

与来自刺胞动物门的海葵(AEW42991)Notch 氨基酸全 序列相似性为 39%,与佛罗里达弓背蚁(Camponotus floridanus)、意大利蜜蜂(Apis mellifera)、蝇蛹金小蜂 (Nasonia vitripennis)、囊舌虫(Saccoglossus kowalevskii) 等无脊椎动物的氨基酸相似度为35%-37%,而与哺乳 动物、两栖动物、鱼类等脊椎动物的氨基酸相似性为 34%-38%。NL、NOD 和 NODP 结构域是 Notch 家族 特有的结构,对这 3 个结构域多序列比对结果显示, 海蜇 Notch 的 NL、NOD 和 NODP 结构域与刺胞动物、 无脊椎动物、脊椎动物的氨基酸相似性分别为 48%、 40%-46%、42%-47%; 25%、25%-28%、23%-30%; 37%、22%-31%、27%-31%(图 3)。

对海蜇Notch基因的3个NL结构域进行多序列 比对表明,不同物种之间其氨基酸序列保守性很强, 且每个 NL 结构域含有 6 个半胱氨酸(图 4)。而 NOD 和 NODP 2 个结构域的多序列比对表明,海蜇 NOD 结构域在甘氨酸(Gly)、苯丙氨酸(Phe)、精氨酸(Arg)、赖氨酸(Lys)、苏氨酸(Thr) 5 个氨基酸位点保 持一致(图 3), NODP 结构域在天冬氨酸(Asp)、天冬 酰胺(Asn)、2 个半胱氨酸(Cys)、苯丙氨酸(Phe)、丙 氨酸(Ala)、丙氨酸(Ala) 7 个氨基酸位点也保持一致(图 5)。

采用邻接法构建Notch的系统进化树(图6)显示, 海蜇与来自刺胞动物门海葵的Notch亲缘关系最近, 聚为一支,而佛罗里达弓背蚁、意大利蜜蜂、蝇蛹 金小蜂、囊舌虫等非脊椎动物聚为一支,人、原鸡、 虹鳟鱼(Oncorhynchus mykiss)等众多脊椎动物又聚 为一支。



图 3 海蜇与其他物种 Notch NOD 结构域的氨基酸多序列比对

Fig.3 Multiple sequence alignment NOD domain of Notch of *R. esculentum* and other species



图 4 海蜇与其他物种 Notch NL 结构域的氨基酸多序列比对 Fig.4 Multiple sequence alignment NL domain of Notch of *R. esculentum* and other species

#### 2.3 Notch 基因在海蜇不同发育时期的表达分析

利用 RT-PCR 技术分析了 Notch 基因在海蜇螅状体、横裂体、蝶状体、水母体 4 个不同发育时期的表达差异。结果显示,该基因在海蜇无性繁殖 4 个发育时期均有表达,螅状体阶段 Notch 的表达量最高,然后是水母体和碟状体(图 7),这 3 个发育阶段的表达量分别是是横裂体阶段表达量的 1.85、1.60、1.33 倍,而横裂体阶段 Notch 的表达量最低。

#### 3 讨论

本研究基于转录组 454 GS FLX 测序和 EST 序 列分析,利用 RACE 和 RT-PCR 技术,解析了海蜇 Notch 基因结构,发现海蜇 Notch 基因与其他物种的 Notch 基因一样都为分泌蛋白,并具有 Notch 家族特 有的 NOD、NODP 结构域和 3 个 NL 结构域,此外, 还包括 EGF-like 结构域、ANK 重复结构域和跨膜结

C. familiaris	– GALLNRVEAALLPG––GAGRRRRELDPMDIRGSIVYLEIDNRQCVQS–––––– 1	1734
O. orca	GASLLPG-SGGGRRRRELDPMDIRGSIVYLEIDNRQCVQS1	1677
E. europaeus	- A A L L G Q V K A V L L P A - T D G S R H R R E L D P M D I R G S I V Y L E I D N R Q C V Q S 1	1688
H. sapiens	- DALLGQVKASLLPGGSEGGRRRRELDPMDVRGSIVYLEIDNRQCVQA1	1688
M. mulatta	- E A L L G Q V K A S L L P G G G G G G R R R R E L D P M D V R G S I V Y L E I D N R Q C V Q A 1	1668
R. norvegicus	TSLLPG-TNGGRQRRELDPMDIHGSIVYLEIDNRQCVQS	1678
M. musculus	SSLLPG-TSGGRQRRELDPMDIRGSIVYLEIDNRQCVQS	1678
C. griseus	SSLLPS-TNGGRQRRELDPMDIRGSIVYLEIDNRQCVQS	1636
G. Gallus	- SSAVINKVKSSLYS-RAGRRQKRELDQMDIRGSIVYLEIDNRQCIQSI	1697
P. numilis V. tuomicalia	- SSAVINKVKSSLYS-RAGRRQKR LDQMD RGSIVYLEIDNRQCIQS	1687
A. iropicalis		1682
C. pyrrhogasier	SINVFNKVKMSLYIS-SNGRQRRELDQNE KGSIVYLEIDNRQCFQS	1085
S. parillas C. somilasvis	P G Q V L N K V K K S L Y D M A G - G K I K K L L D H L Q I K G S I V H L E V D N K Q C F Q Q	108/
C. semiluevis	PGKVLNVVKKSLIDAVAMKKKKLDHLQIKGSIVVLEVDNRQCFQQ	1002
1 morioanus		1200
A. mexicunus		1207
L. OCUltulus A mollifora		1625
A. menijeru C floridanus		1033
C. jioriaanus N vitrinennis		1247
S kowalevskii		1687
N vectensis		1170
R esculentum		1235
n. escutentum		1255
C. familiaris	SSQCFQSATDVAAFLGALASLGSLNIPYKIEAVQSETVELP-PPPQLHFMYVAVA-1	1788
C. familiaris O. orca	S	1788 1731
C. familiaris O. orca E. europaeus	S	1788 1731 1742
C. familiaris O. orca E. europaeus H. sapiens	S	1788 1731 1742 1742
C. familiaris O. orca E. europaeus H. sapiens M. mulatta	S S Q C F Q S A T D V A A F L G A L A S L G S L N I P YK I E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P YK I E A V Q S E T V E P P - P P A P L H F M Y V A V V - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P YK I E A V Q S E T V E P P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P YK I E A V Q S E T V E P P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P YK I E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P YK I E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1	1788 1731 1742 1742 1722
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus	S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V Q S E T V E P P - P P A P L H F M Y V A V V - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V Q S E T V E P P - P P P Q L H F M Y V A V V - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V Q S E T V E P P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V Q S E T V E P P P A Q L H F M Y V A A A - 1	1788 1731 1742 1742 1722 1722
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus	S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A P L H F M Y V A V V - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1	1788 1731 1742 1742 1722 1722 1732
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus	S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A P L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E P V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E P V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E P V E P P - L P S Q L H L M Y V A A A - 1	1788 1731 1742 1742 1722 1732 1732 1690
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus G. Gallus	S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A P L H F M Y V A V V - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y L A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y L A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P A - R N S Q L H L M Y V A A - 1	1788 1731 1742 1742 1722 1732 1732 1690 1751
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus G. Gallus P. humilis	S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A P L H F M Y V A V V - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1	1788 1731 1742 1742 1722 1732 1732 1690 1751 1741
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus G. Gallus P. humilis X. tropicalis	S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A P L H F M Y V A V V - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A V V - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y L A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F T S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P P - L Y S Q L Y M Y V Y A - 1	1788 1731 1742 1742 1722 1732 1732 1690 1751 1741 1737
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus G. Gallus P. humilis X. tropicalis C. pyrrhogaster	S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A P L H F M Y V A V V - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E S V E P P P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y L M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V Q A - 1 S L Q C F Q S A T D V A A F L G A L A T H G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V Q A - 1 S L Q C F Q S A T D V A A F L G A L A S H G N L N T P Y K T E A V K S E T G E P S K G P P - L Y L M M F S M L V - 1	1788 1731 1742 1742 1722 1732 1732 1690 1751 1741 1737 1739
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus G. Gallus P. humilis X. tropicalis C. pyrrhogaster S. partitus	S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A P L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P T - K N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P T - K N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P T - K N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S H G N L N T P Y K T E A V K S E T A E P T - K N S Q L Y P M Y V V A - 1 S T E C F Q S T D V A A F L G A L A S H G N L N T P Y K T E A V K S E T Q E P S K G P P L Y A M F S M V V V V - 1 S T E	1788 1731 1742 1742 1722 1732 1732 1690 1751 1741 1737 1739 1740
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus G. Gallus P. humilis X. tropicalis C. pyrrhogaster S. partitus C. semilaevis	S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A P L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P T - K N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P T - K N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S H G N L N I P Y K I E A V K S E T A E P T - K N S Q L Y P M Y V V A - 1 S T E C F Q S T N D A A A F L G A L A S H G N L N I P Y K I E A V K S E T A E P T - K N S Q L Y P M Y V V A - 1 S T E C F Q S T N D A A A F L G A L A S H G N L N I P Y K I E A V K S E T O E P S Q E L Y P M Y V V	1788 1731 1742 1742 1722 1732 1732 1732 1732 1751 1741 1737 1739 1740
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus G. Gallus P. humilis X. tropicalis C. pyrrhogaster S. partitus C. semilaevis O. mykiss	S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V Q S E T V E P P - P P A P L H F M Y V A V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N I P Y K T E A V K S E T V E P P - L P S Q L H L M Y U A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N I P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S A T	1788 1731 1742 1742 1722 1732 1732 1732 1732 1741 1737 1739 1740 1734 1734
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus G. Gallus P. humilis X. tropicalis C. pyrrhogaster S. partitus C. semilaevis O. mykiss A. mexicanus	S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A P L H F M Y V A V V - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - P P A Q L H F M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S A	1788 1731 1742 1742 1722 1732 1732 1732 1732 1741 1737 1739 1740 1734 1633 1722
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus G. Gallus P. humilis X. tropicalis C. pyrrhogaster S. partitus C. semilaevis O. mykiss A. mexicanus L. oculatus	S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A P L H F M Y V A V V - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S L Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S T E C F Q S T T D A A A F L G A L A S H G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S T E C F Q S T T D A A A F L G A L A S S G C L N T P Y K T E A V K S E T O P A - P S E L Y P T Y L V L G - 1 S T E C F Q S T T D A A A F L G A L A S S G N L N T P Y T T E A V K S E - V D P Q - P S S E L Y P T Y L V L G - 1 S T E	1788 1731 1742 1742 1722 1732 1732 1732 1739 1740 1734 1633 1722 1741
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus G. Gallus P. humilis X. tropicalis C. pyrrhogaster S. partitus C. semilaevis O. mykiss A. mexicanus L. oculatus A. mellifera	S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A P L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A A 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1 S T E C F Q S A T D V A A F L G A L A S H G N L N T P Y K T E A V K S E T O Y P A Y N Y G A - 1 S T E C F Q S T N D A A A F L G A L A S H G N L N T P Y K T E A V K S E T O Z P - P S E L Y P T Y L V L G - 1 S T E C F Q S T N D A A A F L G A L A S G K L N T P Y K T E A V K S E T O Y P A Y C S S N L Y P T Y L V L G - 1 S T E C F Q S A T D V A A F L G A L A S S G K L N T P Y T T E A V F S - V D Q V - Q S S N L Y P T Y L V L G - 1 S T E C F Q S A T D V A A F L	1788 1731 1742 1742 1732 1732 1732 1732 1732 1732 1732 173
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus G. Gallus P. humilis X. tropicalis C. pyrrhogaster S. partitus C. semilaevis O. mykiss A. mexicanus L. oculatus A. mellifera C. floridanus	S Q C F Q S AT D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1     S S Q C F Q S AT D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A P L H F M Y V A V A - 1     S S Q C F Q S AT D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1     S S Q C F Q S AT D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1     S S Q C F Q S AT D V A A F L G A L A S L G S L N T P Y K T E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1     S S Q C F Q S AT D V A A F L G A L A S L G S L N T P Y K T E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1     S Q C F Q S AT D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S Q C F Q S AT D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S AT D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S AT D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S AT D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S AT D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S AT D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S C C F Q S AT D V A A F L G A L A	1788 1731 1742 1742 1732 1732 1732 1732 1732 1732 1732 173
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus G. Gallus P. humilis X. tropicalis C. pyrrhogaster S. partitus C. semilaevis O. mykiss A. mexicanus L. oculatus A. mellifera C. floridanus N. vitripennis	S Q C F Q S AT D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1     S S Q C F Q S AT D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A P L H F M Y V A V A - 1     S S Q C F Q S AT D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1     S S Q C F Q S AT D V A A F L G A L A S L G S L N I P Y K I E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1     S S Q C F Q S AT D V A A F L G A L A S L G S L N I P Y K I E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1     S S Q C F Q S AT D V A A F L G A L A S L G S L N I P Y K I E A V Q S E S V E P P - P P A Q L H F M Y V A A A - 1     S S Q C F Q S AT D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S AT D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S AT D V A A F L G A L A S L G S L N I P Y K I E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S AT D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S AT D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P T - K N S Q L Y P M Y V V A - 1     S S Q C F Q S AT D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P T - K N S Q L Y P M Y V V A - 1     S S Q C F Q S AT D V A A F L G A L A S L G N L N I P Y K I E A V K S E T A E P T - K N S Q L Y P M Y V V A - 1     S S Q C F Q S AT D V A A F	1788 1731 1742 1742 1722 1732 1732 1732 1732 1741 1737 1739 1740 1734 1633 1722 1741 1694 1306 1534
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus G. Gallus P. humilis X. tropicalis C. pyrrhogaster S. partitus C. semilaevis O. mykiss A. mexicanus L. oculatus A. mellifera C. floridamus N. vitripennis S. kowalevskii	S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A P L H F M Y V A V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1     S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1     S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1     S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V C A - 1     S S Q C F Q S A T D V A A F L G A L A S H G N L N T P Y K T E A V K S E T A E P P - L P S Q L H L M Y W V V A - 1     S S Q C F Q S A T	1788 1731 1742 1742 1722 1732 1732 1732 1732 1741 1737 1739 1740 1734 1633 1722 1741 1694 1306 1534 1740
C. familiaris O. orca E. europaeus H. sapiens M. mulatta R. norvegicus M. musculus C. griseus G. Gallus P. humilis X. tropicalis C. pyrrhogaster S. partitus C. semilaevis O. mykiss A. mexicanus L. oculatus A. mellifera C. floridanus N. vitripennis S. kowalevskii N. vectensis P. aculatus	S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E L P - P P P Q L H F M Y V A V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A P L H F M Y V A V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V Q S E T V E P P - P P A Q L H F M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - P P A Q L H F M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G S L N T P Y K T E A V K S E T V E P P - L P S Q L H L M Y V A A A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V A - 1     S S Q C F Q S A T D V A A F L G A L A S L G N L N T P Y K T E A V K S E T A E P A - R N S Q L Y P M Y V V Q A - 1     S T E C F Q	1788 1731 1742 1742 1722 1732 1732 1732 1732 1731 1740 1734 1306 1534 1740 1220

图 5 海蜇与其他物种 Notch NODP 结构域的氨基酸多序列比对 Fig.5 Multiple sequence alignment NODP domain of Notch of *R. esculentum* and other species

黑色为相同氨基酸;灰色为相似氨基酸;黑色圆点为 NL 结构域完全一致的半胱氨酸; \*为 NOD 和 NODP 结构域完全一致氨基酸

Black background indicated the same amino acid; Gray background indicated similar amino acids, the black dots showed the identical cysteine; The pentagrams showed the same amino acid of NOD and NODP

构域(周庆军等, 2004)。Blast 分析表明(表 2),不管是 全序列还是 Notch 基因特有结构域相似性分析,海蜇 Notch 基因与刺胞动物、无脊椎动物和脊椎动物都具 有较高的同源性。系统进化树分析也显示,海蜇与来 自刺胞动物门海葵的 Notch 聚为一支,而来自无脊椎 动物和脊椎动物的 Notch 基因分别聚为 2 个分支,说 明 Notch 基因符合生物进化规律,按照从低等动物到 高等动物分别聚类。

Notch 广泛存在于各种生物中,是一种单次跨膜 蛋白,均由胞外区(NEC)、跨膜区(TM)和胞内区 (NICD)3 大部分组成(付亚娟等,2007)。典型的 Notch 结构域胞外区包括 36 个的 EGF-like 结构域、3 个 NL 结构域、NOD 和 NODP 结构域,胞内区主要包括 6 个 ANK 结构域以及 ANK 两侧的 2 个核定位信号 (NSL)(Hori et al, 1997)。研究表明,不同物种之间 EGF-like 结构域数量为 10-36 不等,差别明显。在脊 椎动物和双翅目动物的 Notch 基因中 EGF-like 结构域 一般为 36 个,而在无脊椎动物中,EGF-like 结构域 数量普遍少于 36 个,比如蝇蛹金小蜂中有 31 个 (Kortschak et al, 2001),而在本研究中,海蜇 Notch 的 EGF-like 结构域数量更少,只有 26 个,其原因可 能是低等动物 Notch 基因在进化过程中为了适应环境 对 EGF-like 结构域进行了必要的删除,类似的现象也 出现在海鞘(Pyrosomella verticilliata) Notch 基因的研究 中(Hori et al, 1997)。富含 Cys 的 NL(Notch/Lin-12)结 构域是所有 Notch 基因特有结构域之一。海蜇









Notch 基因与其他物种的 Notch 一样也含有 3 个 NL 结构域,每个结构域都由 40 个左右的氨基酸组成,并含有 6 个可以形成 3 对高度保守的二硫键的 Cys。 多序列比对也发现,从刺胞动物到哺乳动物,NL 结构域的氨基酸序列相似性高达 48%-83%。这种高度的序列相似性和结构的保守性表明,从低等动物到高 等动物,NL结构域的功能相似,都是通过与配体特 异性结合、启动 Notch 蛋白发挥作用 (Kortschak et al, 2001;李文等, 2014)。

无性繁殖是刺胞动物重要的繁殖方式,其调控 途径和分子机制的研究已经逐步开展(Trevino et al, 2011)。出芽生殖是水螅的无性繁殖方式之一, Philipp 等(2009)研究表明, Wnt 信号途径可以启动并调控出 芽生殖过程。在水螅的出芽生殖中, Wnt5、Wnt8、 Frizzled 2 和 Dsh 基因的表达被局限于水螅出芽部位 和触手上,并随着芽体的增长,这些基因的表达量 逐步升高。而横裂生殖是海蜇无性繁殖的主要方式, 即螅状体经过分节和变态 2 个紧密联系的发育阶段 产生稚水母的过程 (周春娅等, 2013)。本实验室的前 期研究也表明, Wnt 信号途径的几个关键节点基因 家族,包括 Wnt 基因家族、Frizzled 基因家族等都 参与调控了海蜇的无性繁殖过程。这些基因被局限 表达海蜇触手和横裂生殖的分节部位,并且无性繁 殖过程中,横裂体阶段的表达量显著高于螅状体、 碟状体及其水母体阶段(周春娅等, 2013)。而在本研 究中, Notch 基因在海蜇无性繁殖过程的表达模式与

Wnt等基因正好相反。虽然,Notch基因在海蜇无性 繁殖4个发育阶段均有表达,但横裂体表达量最低, 而螅状体表达量最高,并随着横裂结束表达量逐步 升高,这一结果可能与Notch信号途径在刺胞动物 无性繁殖过程中的特殊调控作用相关。Münder等 (2010)研究表明,Notch基因的表达是水螅芽体与母 体之间形成界限的必需条件,在芽体与母体分离时, Notch通路的主要靶基因-hairy/Enhancer of Split基 因在分离区开始表达,进而导致芽体与母体分离。 由此可以推断,Wnt信号途径和Notch信号途径在 刺胞动物无性繁殖过程中,均发挥重要的调控作用, 但其调控机制和模式是不同的,进一步的调控机理 值得更深入探讨。

# 参考文献

- Allman D, Aster JC, Pear WS. Notch signaling in hematopoiesis and early lymphocyte development. Immunological Reviews, 2002, 187(1): 75–86
- Anant S, Roy S, Raghavan KV. Twist and Notch negatively regulate adult muscle differentiation in *Drosophila*. Development, 1998, 125(8): 1361–1369
- Artavanis-Tsakonas S, Rand MD, Lake RJ. Notch signaling: Cell fate control and signal integration in development. Science, 1999, 284(5415): 770–776
- Bottger A, Alexandrova O, Cikala M, et al. GFP expression in Hydra: Lessons from the particle gun. Development Genes and Evolution, 2002, 212(6): 302–305
- Coffman CR, Skoglund P, Harris WA, et al. Expression of an extracellular deletion of Notch diverts cell fate in *Xenopus* embryos. Cell, 1993, 73(4): 659–671
- Conboy IM, Conboy MJ, Wagers AT, *et al*. Rejuvenation of aged progenitor cells by exposure to a young systemic environment. Nature, 2005, 433(7027): 760–764
- de Oliveira-Carlos V, Ganz J, Hans S, et al. Notch receptor expression in neurogenic regions of the adult zebrafish brain. PLoS One, 2013, 8(9): 1–21
- Fortini ME, Rebay I, Caron LA, *et al.* An activated Notch receptor blocks cell-fate commitment in the developing *Drosophila* eye. Nature, 1993, 365(6446): 555–557
- Fu YJ, YE F, Lv WG, et al. Recent researches on the Notch signaling pathway. Journal of Medical Molecular Biology. 2007, 4(5): 447–450 [付亚娟, 叶枫, 吕卫国, 等. Notch 信 号通路的研究现状. 医学分子生物学杂志, 2007, 4(5): 447–450]
- Guo Z, Huang J. Notch signaling pathway and Neuralized protein. Chinese Journal of Biochemistry and Molecular Biology, 2008, 24(11): 987–991 [郭政, 黄瑾. Notch 信号通 路与 Neuralized 蛋白. 中国生物化学与分子生物学报, 2008, 24(11): 987–991]

- Hori S, Saitoh T, Matsumoto M, *et al.* Notch homologue from *Halocynthia roretzi* is preferentially expressed in the central nervous system during ascidian embryogenesis. Development Genes and Evolution, 1997, 207(6): 371–380
- Kasbauer T, Towb P, Alexandrova, Alexandrova O, *et al.* The Notch signaling pathway in the cnidarian Hydra. Developmental Biology, 2007, 303(1): 376–390
- Kortschak DR, Tamme R, Lardelli M. Evolutionary analysis of vertebrate Notch genes. Development Genes and Evolution, 2001, 211(7): 350–354
- Lai EC. Notch signaling: Control of cell communication and cell fate. Development, 2004, 131(5): 965–973
- Li BY, Feng L, Wang YC, et al. Molecular composition, activation pathways and activity of Notch signaling Molecules in cell differentiation. Chinese Journal of Neuro medicine, 2009, 8(6): 633–636 [李宝园, 丰玲, 王雁春, 等. Notch 信号转导相关分子的构成、相关机制与调节. 中华 神经医学杂志, 2009, 8(6): 633–636]
- Li W, Li CS. VEGF-Notch signaling pathway and angiogenesis. Shandong Medical Journal, 2014, 54(13): 83-85 [李文, 李 长生. VEGF-Notch 信号通路与血管再生的关系. 山东医 药, 2014, 54(13): 83-85]
- Liang J, Han H. Notch signaling pathway and vascular development. Chinese Journal of Cellular and Molecular Immunology, 2008, 24(12): 1225–1226 [梁洁, 韩骅. Notch 信号通路与血管发育. 细胞与分子免疫学杂志, 2008, 24(12): 1225–1226]
- Lu ZZ, Wang LS, Wu ZZ. Progress of Notch signaling. Progress in Physiological Sciences, 2004, 35(2): 135–138 [鲁茁壮, 王立生, 吴祖泽. Notch 信号通路研究进展. 生理科学进 展, 2004, 35(2): 135–138]
- Maine EM, Lissemore JL, Starmer WT. A phylogenetic analysis of vertebrate and invertebrate Notch-related genes. Molecular Phylogenetics and Evolution, 1995, 4(2): 139–149
- Marlow H, Roettinger E, Boekhout M, et al. Functional roles of Notch signaling in the cnidarian Nematostella vectensis. Developmental Biology, 2012, 362(2): 295–308
- Mohr OL. Cases of mimic mutations and secondary mutations in the X-chromosome of *Drosophila melanogaster*. Molecular Genetics and Genomics, 1922, 1(28): 1–22
- Mohr OL. Character change caused by mutation of an entire region of a chromosome in *Drosophila*. Genetics, 1919, 3(4): 275–282
- Münder S, Kasbauer T, Prexl A, *et al.* Notch signalling defines critical boundary during budding in Hydra. Developmental Biology, 2010, 344(1): 331–345
- Philipp I, Aufschnaiter R, Ozbek S, *et al.* Wnt/beta-catenin and noncanonical Wnt signaling interact in tissue evagination in the simple eumetazoan Hydra. Proceeding of the National Academy of Science, 2009, 106(11): 4290–4295
- Qi RZ, Cao XT. The structure of the Notch and signal transduction. Chinese Journal of Cancer Biotherapy, 2002, 9(3): 212–214 [齐润姿, 曹雪涛. Notch 的结构和信号转导.

中国肿瘤生物治疗杂志, 2002, 9(3): 212-214]

- Sun LZ, Hou L. Structure, function and related signaling pathway of Notch. Chinese Journal of Cell Biology, 2010, 32(6): 914–921 [孙丽哲, 侯林. Notch 的结构、功能和相关 信号通路. 中国细胞生物学学报, 2010, 32(6): 914–921]
- Trevino M, Harmon S, Burton PM. Wnt signaling promotes oral fates during regeneration and embryogenesis in the cnidarian *Nematostella vectensis*. Developmental Biology, 2011, 356(1): 252–254
- Zanotti S, Canalis E. Notch and the Skeleton. Molecular and Cellular Biology, 2010, 30(4): 886–896
- Zhou CY, Zhu L, Pan Y, et al. Wnt5 gene from Rhopilema esculentum: cDNA cloning, genomic organization and mRNA expression. Oceanologia et Limnologia Sinica, 2013, 44(4): 1115–1122 [周春娅,朱玲,潘滢,等.海蜇 (Rhopilema esculentum)Wnt5 基因: cDNA 克隆、基因组结 构与表达.海洋与湖沼, 2013, 44(4): 1115–1122]
- Zhou QJ, Hu RZ, Zou JZ, et al. Notch signal transduction and regulation. Progress in Biochemistry and Biophysics, 2004, 31(3): 198–203 [周庆军, 胡若真, 邵健忠, 等. Notch 信号 转导与调控. 生物化学与生物物理进展, 2004, 31(3): 198–203]

(编辑 陈严)

# Cloning and Expression of Notch Gene in Rhopilema esculentum

LUO Xiaorui<sup>1,2</sup>, ZHU Ling<sup>20</sup>, ZHOU Chunya<sup>2</sup>, ZHUANG Zhimeng<sup>2</sup>, FAN Yanjun<sup>2</sup>

(1. College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306; 2. Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture, Shandong Provincial Key Laboratory of Fishery Resources and Eco-environment, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071)

Abstract Notch Pathway is a very conservative signaling pathway in evolution, through the interaction between Notch receptor and the ligand DSL protein which conveys signals to regulate cell proliferation, differentiation, apoptosis, and determine the fate of cells. *Rhopilema esculentum* belongs to cnidaria, it is an ideal model organism to study cnidaria asexual propagation because of the mature artificial breeding and aquaculture technology. The cDNA cloning and expression analysis of Notch were first performed in the *R. esculentum* based on 454 GS-FLX sequencing technique by RT-PCR and RACE method. The results showed that the full-length cDNA of Notch was 6768 bp, containing 5'- untranslated region (UTR) of 90 bp, an open reading frame (ORF) of 6066 bp and 3'- untranslated region (UTR) of 612 bp which contains a polyadenylation signals of AATAAA. SMART analysis showed that R. esculentum Notch, as a secretory protein, included a putative signal peptide of 21 amino acid residues; its mature peptide included 2000 amino acid residues and consisted of 37 structure domains, which includes 26 EGF-like domains, 6 ANK domains, and a transmembrane domain except 3 Notch / Lin-12(NL) structure domain, a NOD and a NODP domains which only exist in Notch family. The homology and phylogenetic analysis showed that the amino acid sequence similarity between *R.esculentum* Notch and Cnidaria Nematostella vectensis amino acid was 39%; by contrast, compared with invertebrates and vertebrates, R. esculentum shares 35%-37% and 34%-38% respectively, in the similarity of the amino acid. Phylogenetic tree of the Notch based on neighbor-joining method showed that R.esculentum was clustered with Cnidaria N. vectensis. Quantitative real-time PCR analysis revealed that the expression of Notch transcript was detected in all four developmental stages, with the highest level in scyphistoma stage and the lowest in strobili stage. The expression level of scyphistoma stage was 1.85 -fold of that in strobili. The results indicate that Notch gene is involved in the regulation of asexual reproduction of *R.esculentum*. Key words Rhopilema esculentum; Notch; cDNA; Real-time PCR

① Corresponding author: ZHU Ling, E-mail: zhuling@ysfri.ac.cn