

# 甲壳动物基因研究进展\*

## REVIEW OF RESEARCH ON CRUSTACEAN GENES

吴亚君 彭宣宪

(厦门大学生命科学学院,细胞生物学与肿瘤细胞工程教育部重点实验室 361005)

基因克隆技术不断发展完善,不仅用来克隆已知功能基因,而且成功地克隆了一些未知基因,从而大大推动生物基因资源的开发。近年来,国外学者针对甲壳动物基因库,其中主要是经济类甲壳动物如虾、蟹及非经济类的卤虫(*Acartia*)基因开展了细致的研究工作,分离鉴定到约30多种功能基因,另外有线粒体基因, rRNA基因及一些重复序列。而国内在这方面的比较薄弱。研究该类动物基因,不仅拓展了已有的关于真核生物基因结构、功能、表达、进化等方面的知识,更重要的是,通过深入了解控制虾、蟹的重要经济性状如生长速度、发育调节、抗病抗逆等性状的相关基因,分析它们与动物健康、疾病、生殖、发育之间的关系,有望从根本上解决诸如品种驯化、养殖和运输过程造成的生活条件恶化、疾病暴发等因素造成的养殖业受损或受阻的问题,培育优良品种,提高养殖数量和质量。本文综合介绍近10年来国外在甲壳动物基因研究方面的主要成果。

### 1 研究方法概述

鉴于cDNA文库具备筛选比较简单易行、假阳性率低、可以直接表达蛋白质等优点,克隆目的基因多从cDNA文库出发。对文库筛选的关键是探针的选择。对于已知多肽分子的基因,可利用氨基末端序列指导合成探针或PCR引物,也

可用多肽特异性抗体筛选阳性克隆。对于已知序列基因家族的成员,可用已知分子的高保守序列指导制备同源探针筛选克隆或合成引物进行PCR克隆。若要筛选细胞分化表达基因,则采用反向生物学原理发现新基因的技术,如减数cDNA文库,差异展示PCR等。筛选过程中,往往要进行连续克隆多级筛选方可获得全长cDNA。

cDNA只能反映mRNA分子结构,所以有必要从基因组出发,利用cDNA筛选DNA克隆,研究基因结构、调控方式及组织特点。

对得到的cDNA,基因组DNA部分或全长序列进行基因结构和功能分析,在很大程度上要借助国际互联网生物信息资源,如NCBI网站BLAST程序被广泛用于序列与核酸或蛋白质数据库的同源性比较。在获得阳性克隆后,往往还要使用Northern blot, Southern blot, Western blot, PCR等技术进一步确证结果。另外,分析基因的表达情况,使用的常用方法有Northern blot, 细胞荧光定位等。

总之,研究甲壳动物基因的方法与其他生物是一致的。

### 2 甲壳动物基因研究成果

#### 2.1 功能基因

在已分离鉴定的功能基因中,生长调控相关基因和酶基因占绝大多数比例,其原因可能是生长发育调控与生长速度关系密切从而

更引起人们的兴趣;在基因编码的全部蛋白质中酶占很大比例。

#### 2.1.1 生长调控相关基因

蜕皮激素和蜕皮抑制激素相互拮抗控制甲壳动物幼体到成体的发育进程。利用果蝇蜕皮激素受体DNA结合区序列指导合成引物通过RFPCR从虾(*Metapenaeus* *er* *sis*)表皮获得1个与昆虫蜕皮激素受体基因E75A高度同源cDNA片段,以此为探针从虾泳足cDNA文库筛到1个3.6 kb cDNA,编码606肽,具有所有5个典型DNA结合区。poly(A)位点在终止信号上游400 bp处。蜕皮前在表皮、眼柄、神经索均有表达,但水平不同。蜕皮之前注射20-羟蜕皮激素可诱导mRNA水平上升<sup>[1]</sup>。Ohira T.等以类似方法获得1个日本对虾*Penaeus japonicus*蜕皮抑制激素(MH) cDNA,为814 bp,其开放阅读框(ORF)为315 bp,包含1个信号肽(28残基)和MH(77残基)。Northern blot表明该基因只在眼柄表达,mRNA水平在蜕皮周期没有显著变化<sup>[2]</sup>。另外,Chan等从蟹(*Charybdis* *fnatus*)克隆到1个4.3 kb的MH基因,3个外显子,2个内含子。第1,2个外显子编码信号肽,第2,3

\* 国家海洋局第三海洋研究所海洋工程重点实验室开放课题基金资助项目。

收稿日期:2000-03-27;

修回日期:2001-02-08

个外显子编码成熟肽。外显子与内含子连接点符合“GTAG”规则。5'转录终止点3'下游1 kb处为poly(A)位点,5'侧翼区的启动子包含若干与某些脊椎动物神经肽基因相似的顺式作用元件<sup>[3]</sup>。Umphrey等从食用蟹 *Cancer magister* 眼柄克隆到MH基因。ORF 339 bp,编码78残基成熟肽,35残基信号肽。与虾MH只有26%~45%的同源性<sup>[4]</sup>。

高血糖激素(CHH)是1种主要调节糖类代谢的多功能激素,可能也与生育和蜕皮有关。Gu等克隆到数个 *Metapenaeus ensis* 前高血糖激素原cDNA,翻译产物包含1个信号肽,1个前体肽和CHH肽。信号肽和前体肽在所有报道CHH中长度最短。研究表明虾基因组至少存在6个拷贝CHH基因簇。CHH基因长度1.5~2.1 kb,包括3个外显子和2个内含子,5'侧翼区上游150~200 bp存在真核基因典型的启动子序列,400 bp区有数个顺式作用元件。CHH基因组织方式与MH基因类似<sup>[5]</sup>。

色素扩散激素(PDH)具有调节复眼色素光适应分布的功能。根据Doklei等1993年的报道,*Oconectes limosus* cDNA包含信号肽、前体区、C端高度保守的PDH编码序列。免疫原位杂交显示虾眼柄视神经节有4种细胞簇表达PDH。在与PDH cDNA及PDH抗体杂交试验中,有3种细胞呈阳性反应,而神经板芒神经节细胞的核周质仅与PDH抗体反应,表明该类细胞存在的是类PDH肽。

甲壳动物咀嚼器在窦腺合成的咀嚼器抑制激素(MOIH)负调控下合成并分泌保幼激素。Tang CH等从食用蟹 *Cancer pagurus* 窦腺分离到MOIH1,MOIH2(相差1个残基)。两者全长cDNA编码1个34残基信号肽及78残基成熟肽。Northern blot显示MOIH只分布于

X器官。MOIH基因存在10个拷贝。HPLC放射免疫分析发现窦腺提取物具MOIH免疫活性<sup>[6]</sup>。

Shean等于1995年报道的虾泛素cDNA,长1.7 kb,5'端缺少非翻译区。该基因与虾蜕皮期间爪部肌肉萎缩相关。Dekleijn等于1994年报道了螯虾 *Homarus americanus* 性腺抑制激素(GIH)mRNA只存在于虾眼柄,推测GIH不仅调节生殖,也与蜕皮相关激素的释放有关。从虾(*Metapenaeus ensis*)眼球cDNA文库获得1个孤儿受体家族成员FTZF1,4.3 kb,编码545肽,呈现与脊椎动物相似的核激素受体特征。研究表明,该基因对虾早期胚胎及幼体发育很重要,可能在其他生理过程也发挥作用<sup>[7]</sup>。还有学者研究了虾类Hox基因,发现Hox基因家族两个成员Ubx,AbdA表达模式的变化与虾胸部内肢演化为捕食附肢关系密切<sup>[8]</sup>。

2.1.2 酶基因 南美白对虾 *Penaeus vannamei* 组织蛋白酶L基因有3个内含子和6个长度不同的外显子(共1792 bp),基因结构与大鼠组织蛋白酶L同源,与果蝇、疟原虫完全不同。第3个内含子呈多态性,推测虾至少有3种组织蛋白酶基因<sup>[9]</sup>。该虾胰蛋白酶基因,含2个短内含子,位点与脊椎动物前2个内含子相近。已获得的3个基因属于3个不同的家族,第3个家族呈前两者嵌合体特征。3个家族之间,内含子存在高度多态性,而每个家族内部对应的内含子比较保守。同时,首次在胰蛋白酶基因发现内含子5'端GC切点<sup>[10]</sup>。Sellon等克隆到两个虾胶原纤维丝氨酸蛋白酶基因组DNA片段,转录区1522~1526 bp,含7个外显子,活性位点被两个内含子断裂。内含子数量及位点与哺乳动物胰凝乳蛋白酶或弹性蛋白酶相似。Southern blot分析表明基因组存在40~50

拷贝该基因,分别属于ChyA,ChyB家族<sup>[11]</sup>。

南美白对虾 *Penaeus vannamei* 至少有两种淀粉酶同工酶。1996年Vanwormhoudt等根据葡萄球菌蛋白酶消化 $\alpha$ -淀粉酶的片段指导合成探针从其肝脏cDNA文库筛选到3个克隆。AMY SK37编码511肽前体酶,具高度疏水信号肽(16残基)。AMY SK20与前者相比在3'端有57个残基的变化。比较发现,虾淀粉酶与哺乳动物有59%~63%同源性,与昆虫有52%~62%同源性。日本对虾 *Penaeus japonicus* 几丁质酶基因编码527肽,与已知几丁质酶相似。蜕皮前期尾扇和尾叶mRNA显著增加,蜕皮间期水平很低,说明该基因与蜕皮有关<sup>[12]</sup>。从卤虫 *Ateia franciscana* 胚cDNA文库筛到两个海藻糖酶cDNA,分别为2496 bp和2485 bp,具近乎相同的ORF(703残基)。编码产物有4个N糖基化位点,1个长的羧化C末端,该末端含1个跨膜区和cAMP依赖性磷酸化位点<sup>[13]</sup>。

*Panulirus argus* 细胞色素P450存在于肝脏微体,催化甲基苯基丙基苯胺、氨基嘌呤、苯并芘等单加氧反应。James等利用P450 N端39个残基序列指导合成探针通过多级筛选最终获得1个P450 cDNA,编码492肽,含典型的血红素结合位点。虾P450与哺乳动物P450具36%序列同源性。

卤虫 *Ateia* 磷脂酶c cDNA克隆PLG $\beta$ (x),长285 bp,编码489肽,与高等物种PLG $\beta$ 相比在5'端缺失几百个残基。PLG $\beta$ (x)具保守的x,y结构域,但x序列在5'端高度截短。1994年Su等通过Northern blot及Y区域上游序列PCR反应结果验证了PLG $\beta$ (x)不是其他无关可隆与PLG $\beta$ 同源区连接形成的假产物。

根据Escalante R.等1994年的

报道, 虾肌浆内质网  $\text{Ca}^{2+}$ -ATPase (SERCA) 基因克隆, 长约 65 kb, 含 18 个外显子, 17 个内含子。两个主要转录起始区相距 30 个核苷酸, 起始区有 6 核苷酸重复单元, 启动子区有数个调控位点, 包括肌肉组织特异性 CArG box, 3 MEF2 及 8 个 MyoD 家族转录因子结合位点。虾组织中存在两种该基因 mRNA, C 末端及 5' 非翻译区起始序列不同, 分别分布于附肢肌肉和幼虾所有组织, 与脊椎动物 SERCA2 mRNA 分布特点相似。分析表明这两种 mRNA 来自同一基因不同剪接的结果, 并且虾与脊椎动物剪接方式相似。进化上相差甚远的物种存在相同的剪接方式和组织特异性表达方式说明该基因存在两种同源产物具有重要的生理意义。Garciasaez 等得到 *Astrea franciscana* Na<sup>+</sup>/K<sup>+</sup>-ATPase  $\alpha_3$  亚基的数个重叠基因组克隆。分析显示该基因有 15 个外显子和 14 个内含子。其中 10 个内含子的定位与人的 Na<sup>+</sup>/K<sup>+</sup>-ATPase  $\alpha_3$  亚基基因相同。该基因 5' 侧翼区检测到转录起始子, 相邻的上游区具启动子功能<sup>[14]</sup>。

Dumas 等于 1993 年报道了用抗体免疫和多聚核苷酸探针筛选到虾 Arg 激酶 cDNA, 开放阅读框 355 bp, 序列符合 ATP: 胍基磷酸转移酶家族序列特征。这是第 1 个关于该家族非肌酸激酶类成员 cDNA 克隆的报道。

另外, Linser 等于 1993 年报道了从虾嗅觉器官 cDNA 文库筛到 Gu 合成酶 cDNA, 2 045 bp, 5' 非翻译区 55 bp, 开放阅读框 1 083 bp, 3' 非翻译区 907 bp, 编码肽与鸡、人、果蝇 Gu 合成酶分别具 65%, 64%, 63% 同源性。

2.1.3 结构蛋白基因 *Artemia franciscana* 快肌和慢肌原肌球蛋白 (Tm) cDNA, STm cDNA 编码 284 肽, 具 5', 3' 非翻译区, 肽末端

具乙酰化位点。FTm 与 sTm 比较只有第 39 至第 80 个氨基酸残基的区别。有人用虾肉致敏病人血清从虾 cDNA 文库筛选到 1 个 IgE 反应阳性克隆 Met e I, 纯化该克隆在质粒中表达, 对病人血清 Ig E 呈阳性反应, 正常人血清阴性。Genbank 分析表明 Met e I 为原肌球蛋白基因<sup>[15]</sup>。1996 年 Ortega 等获得 3 种虾肌动蛋白等位基因 *A. pithenogenetia* 的 Act 211、Act 302 和 *A. franciscana* Act 403。Act 211 含 4 个外显子, Act 302 含 6 个, Act 403 含 7 个。Act 211 与 Act 403 在第 168 个密码子处有一相同内含子, Act 302 与 Act 403 在第 121~122, 第 246~247 密码子之间有相同内含子。Cotton JLS 等从螯虾 *Homarus americanus* 深层腹屈肌 cDNA 文库筛到一肌球蛋白重链 cDNA, 1.5 kb, 肽 C 端 413 个残基与果蝇高度同源 (73%), 与脊椎动物有 49% 同源性。该基因只在快肌中表达, mRNA 在胞内分布不均一, 大多存在于肌原纤维之间的空间, 肌纤维膜内质及细胞膜皱褶内, 这些区域以内的 mRNA 聚集在核周围胞质内。Demex 等为揭示虾腹足皮层腺微管蛋白的均一性是源于 1 个大的基因家族抑或少数基因转录后修饰的结果, 从 *A. franciscana* cDNA 文库筛选到  $\alpha$  微管蛋白和  $\beta$  微管蛋白 cDNA, 均编码 451 肽, 一级结构存在一些特有的变化, 估计具功能上的意义。Southern blot 显示简单的条带模式, 说明少数基因转录后修饰可以满足机体对多种不同功能微管蛋白的需求。有人研究了虾在蜕皮周期中肌原纤维蛋白、肌动蛋白、肌球蛋白基因表达的调控因子, 发现蜕皮激素和温度可能通过调节核糖体活性及翻译过程影响肌肉生长。其中温度直接影响蛋白质合成速率及肌动蛋白 mRNA 水平, 可能是虾季节性蜕皮循环的主要控制因子<sup>[16]</sup>。

卵黄发生对于甲壳动物生殖过程十分重要。卵黄蛋白主要包括卵黄生成素 (Vg) 和卵黄磷蛋白 (Vt), Vg 存在于卵母细胞外组织而 Vt 存在于卵母细胞内。研究表明, 虾类肝胰腺是 Vg 合成位点。Khayat 等从短沟对虾 *Penaeus semisulcatus* Vg 分泌型卵巢 cDNA 文库中筛到 1 个 Vt cDNA, 约 1.1 kb。该 cDNA 与 Vg 分泌型雌体卵巢及肝脏 mRNA 均有杂交信号, 与非 Vg 分泌型雌体 mRNA 有弱信号, 与肝脏 mRNA 无杂交信号。这些研究说明, Vt 和 Vg 为同一基因产物, 两者同源性很高。

Sellos 等获得南美白对虾 *Penaeus vannamei* 血蓝蛋白 cDNA, 2 095 bp, 具短 5' 非翻译区 (17 bp), ORF 1 980 bp, 13 残基信号肽和 648 残基成熟肽, 另有 89 残基 3' 非翻译区和 polyA 尾。这是首次在节肢动物血蓝蛋白发现信号肽<sup>[17]</sup>。

2.1.4 其他功能基因 实验发现虾胚可以忍受多种压迫性外条件, 如极端温度、长时间缺氧、缺能、缺水。从 *Astrea* 胚肽 cDNA 文库中筛到第 1 个甲壳类热休克蛋白 P26 cDNA, 具典型的  $\alpha$ -晶体蛋白结构域, 富  $\beta$ -折叠, 肽末端不保守, 富 Arg, Gly<sup>[18]</sup>。

Baro 等克隆到一种龙虾 *Paralithys interuptus* 胃神经节幽门神经元 shab 基因, 属于编码电压依赖 K<sup>+</sup>通道蛋白的 Shaker 家族成员。不同剪接方式产生多种转录体。组织特异性表达研究 shab 基因表达的差异导致不同类型细胞独特的电生理现象。

## 2.2 rDNA

蚤状蚤 *Daphnia pulex* 18 s 与 28 s rDNA 间隔 4 819 bp, 含两个非重复序列, 中间隔以两个重复序列。其中大重复序列包括 3 个完整、1 个截短的 330 bp 重复单位, 每个亚单位具 5'-TATAAGGGAAG 3' 序

列,与其他物种间隔区重复单位中的启动子序列相似。3'非重复区含启动子(2 689 bp)及两个 poly(GT)区域,分别为 18 bp 和 20 bp。该 GT 序列在体外可形成 Z-DNA 构象。人和鼠的 rDNA 及蛋白质基因转录单位上游亦存在 GT 区。*Proasellus coxalis* 5S rDNA 呈 589 bp 串联重复,其中有 120 bp 转录区和 469 bp 间隔区。不同种属之间 5S rDNA 序列比较显示,序列的一致程度反映种间关系<sup>[19]</sup>。Baizotti 等从等足目 *Asellus aquaticus* 基因组 PCR 扩增到 1 个 6 553 bp 串联重复单位,含 1 拷贝组蛋白 H2B, H2A, H3, H4 及 5S rRNA 基因。在卤虫基因也发现类似的基因结构。组蛋白基因 3'侧翼区具两个不同的转录终止信号,另具茎环结构和 AATAAA 顺序。原位荧光杂交显示该基因串在每个基因组约 200~300 个拷贝<sup>[20]</sup>。

### 2.3 卫星 DNA

有人分析 *Atenia franciscana* 基因组 Alu I 序列,重复单元 113 bp,是组成型异染色质主要成分,推测与异染色质凝缩有关。将 Alu I 卫星 DNA 片段与酵母 lac Z 报告基因重组,发现 Alu I DNA 的含量和方向影响报告基因表达<sup>[21]</sup>。另外,有人发现南美白虾 *Penaeus vannamei* 基因组存在约 10<sup>6</sup> 拷贝的 139~188 bp 长度微卫星 DNA 重复单元<sup>[22]</sup>。

### 3 线粒体基因组

1993 年 Okimoto R 等研究表明,虾线粒体基因组平均长度约 18 000 bp<sup>[23]</sup>,重链转录主要起始点位于 12s rDNA 5'末端,十分短小,第 2 起始点在第 1 位点上游 250 bp 处。Okimoto 等测定一段卤虫属的 *Atenia salina* mtDNA 片段序列,发现 NADH 还原酶复合物(ND1)基因与细胞色素 b(cytb)基因之间被 1 个丝氨酸 tRNA(UCN)基因分离,该

结果纠正了过去的关于 ND1 与 Cytb 基因紧密相连的说法。Garcia 等测定了 *A. franciscana* 的完整 mtDNA 序列(15 822 bp),从而为甲壳纲比较学研究提供分子标志。虾 mtDNA 与果蝇的组织方式基本相同,除了虾 ND2 基因 3'下游有两个 tRNA 基因,该区呈现与脊椎动物 mtDNA 轻链重复区相似的茎环结构。推测甲壳类与昆虫类 mtDNA 分化源于 1 个祖先 mtDNA 复制错误。虾 mtDNA 与昆虫 mtDNA 密码子相同,ATN 和 GTG 为起始密码,一些基因的终止密码为不完整的 T 或 TA。另外,有人利用 mtDNA 和 rDNA 等基因标志研究小长臂虾属的 *Palaeomonetes kadiakensis* 端始种杂交地区的基因图和基因交换的动力学特征<sup>[24]</sup>。

### 4 展望

由于动物的重要经济性状往往有数量性状位点的多基因调控,因此从基因组水平研究虾基因意义深远。有人提出虾基因组研究计划,显然只有通过基因组全序列的测定才能真正 1 个不丢地确定每个基因,了解基因组携带的全部遗传信息,从宏观水平理解基因转录和转录后调节,从三维空间角度研究基因表达调控。当然,事实上不可能等基因组 DNA 全部序列测出后再去逐一了解每个基因的功能和结构,而且基因组研究需要积累大量表达标签序列(EST)和单拷贝序列(SIS),所以对单个基因的分选鉴定工作还是很重要的。

无论是单个基因还是整个基因组开放阅读框架所提供的新基因,虽然从序列上说是新成员,但不足以阐明其生物学功能,所以需与具体的组织和系统相结合,从分子、细胞、整体 3 个水平结合基因敲除和基因打靶技术加以研究。这样,将会有更多重要因子、酶蛋白

质及共同的基因被发现。

另外,有关抗病基因的研究迄今尚未见报道,这方面的工作很值得探索。

### 参考文献

- 1 Chan S. M. . *FEBS letter* , 1998 , **436** (3) : 395 ~ 400
- 2 Ohira T. *et al.* . *Zoological Science* , 1997 , **14**(5) : 785 ~ 789
- 3 Chan S. M. *et al.* . *Gene* , 1998 , **224** (1-2) : 23 ~ 33
- 4 Umphrey H. R. *et al.* . *Molecular and Cellular Endocrinology* , 1998 , **136** (2) : 145 ~ 149
- 5 Gu PL. , Chan S. M. . *FEBS Letter* , 1998 , **441**(3) : 397 ~ 403
- 6 Abzhanov A. *et al.* . *Development* , 1999 , **126**(6) : 1 121 ~ 1 128
- 7 Chan S. M. , Chan K. M. . *FEBS Letter* , 1999 , **454**(1) : 109 ~ 114
- 8 Averof M. , Patel N. H. . *Nature* , 1997 , **388**(6443) : 682 ~ 686
- 9 Leboulay C. *et al.* . *Gene* , 1998 , **218** (1) : 77 ~ 84
- 10 Klein B. *et al.* . *Gene* , 1998 , **216** (1) : 123 ~ 129
- 11 Sellos D. , Vanwormhoudt A. . *Biochimica et Biophysica Acta* , 1999 , **1432**(2) : 419 ~ 424
- 12 Watanabe T. *et al.* . *Zoological Science* , 1997 , **14**(1) : 65 ~ 68
- 13 Tanaka S. , Kono M. . *Zoological Science* , 1999 , **16**(2) : 269 ~ 277
- 14 Garciasaez A. *et al.* . *Biochemical Journal* , 1997 , **321**(2) : 509 ~ 518
- 15 Mykles D. L. *et al.* . *Journal of Muscle Research and Cell Motility* , 1998 , **19**(2) : 105 ~ 115
- 16 Whiteley N. M. *et al.* . *Comparative Biochemistry and Physiology* , 1997 , **117**(3) : 323 ~ 331
- 17 Sellos D. *et al.* . *FEBS Letter* , 1997 , **407**(2) : 153 ~ 158
- 18 Liang P. *et al.* . *Journal of Biological Chemistry* , 1997 , **272** (30) : 19 051 ~ 19 058
- 19 Pellicia F. *et al.* . *Genome* , 1998 , **41**(1) : 129 ~ 133

- 20 Barzotti R. *et al.* . *Genome* ,2000 ,**43**  
( 2) :341 ~ 345
- 21 Bagshaw J.C., Buckholt M.A. .  
*Gene* ,1997 , **184**( 2) :56 ~ 60
- 22 Maiarano D. *et al.* . *Gene* , 1997 ,
- 23 Grandjean F., Soutygrosset C. .  
*Comptes Rendus de l'Académie Des  
Science Series 3 Sciences de La Vie et  
Sciences* ,1997 , **320**( 7) :551 ~ 556
- 24 Chen Y.N. *et al.* . *Molecular  
Reproduction and Development* ,  
1999 , **54**( 3) :475 ~ 482
- ( 本文编辑 :刘珊珊)