

草鱼肝胰脏组织部分 EST 的鉴定及分类

陈 韬¹, 陈立祥¹, 苏建明¹, 章怀云²

(1. 湖南农业大学 动物科技学院, 湖南 长沙 410128; 2. 中南林业科技大学 生命科学与技术学院, 湖南 长沙 412006)

摘要:本研究构建了草鱼(*Ctenopharyngodon idellus*)肝胰脏组织的 cDNA 文库, 并对 768 个随机挑选的克隆进行部分测序, 获得了 567 个 ESTs。BLAST 结果表明, 457 个 ESTs 与 GenBank 中已经鉴定的 130 个基因具有较高的同源性, 其中有 5 个基因是已经报道的, 125 个是草鱼中首次鉴定的基因; 110 个和未鉴定的基因具有相对较低的同源性, 称为假定蛋白。根据编码产物的功能可以将这些基因分为: 代谢(41), 基因表达及其蛋白质合成(56), 细胞及机体防御(18), 细胞信号传导与细胞通讯(11), 细胞结构与运动(4)。[中国水产科学, 2006, 13(5): 714-722]

关键词:草鱼; 肝胰脏; cDNA 文库; 表达序列标签

中图分类号:Q959.483 **文献标识码:**A **文章编号:**1005-8737-(2006)05-0714-09

鲤科鱼类的肝脏和胰脏呈弥散型, 它们在组织形态上没有明显的区别, 一般合称为肝胰脏。肝细胞在糖、脂类、甾类、氨基酸、维生素等多种物质的代谢过程中具有重要作用。胰腺具有两种分泌功能, 一个是外分泌功能, 其产物主要是一些消化酶, 如胰蛋白酶, 胰淀粉酶, 胰脂肪酶等; 另一个是内分泌功能, 主要分泌胰岛素、胰高血糖素等激素, 以调节糖、蛋白质和脂肪的代谢。在哺乳动物中, 肝脏主要由肝细胞组成(78%), 肝细胞承担了肝脏的主要功能, 如产生大量的血浆蛋白, 解毒和生成 ATP^[1]。由于功能的重要性, 以及其细胞组成的单一性, 肝细胞成为在转录水平研究哺乳动物基因表达的模式^[2]。

和哺乳动物相比, 目前人们对鱼类基因表达的研究尚不够深入。如中国主要的淡水养殖鱼类草鱼(*Ctenopharyngodon idellus*), 以前的研究主要侧重于对单个基因克隆及表达的研究, 如生长激素基因^[3-5]、antiquitin^[6]、GC-LRG^[7]、delta-9 desaturase^[8]、partner of Nob1^[9]、CITED3^[10]等。其中大部分的研究是与机体的代谢和免疫有关, 但是肝胰脏作为在代谢和免疫过程中起重要作用的器官, 其表达基因的研究尚没有专门的报道。利用 cDNA 部分测序, 构建表达序列标签文库是一种快速、有效的鉴定新基因的方法, 目前已广泛地应用到许多动植物的基因鉴定中^[11]。本研究通过构建草鱼肝胰脏

表达序列标签文库, 鉴定了大量的肝胰脏的表达基因, 并对肝胰脏的基因表达特点进行分析, 旨在探讨草鱼的代谢及免疫等重要生理过程的分子生物学特征。

1 材料与方法

二龄草鱼(*Ctenopharyngodon idellus*)取自湖南农业大学动物科技学院实验渔场。解剖后取其肝胰脏, 用 DEPC 处理的无菌蒸馏水清洗样品 3 次, 利用 TRIzol 试剂(GIBCOBRL)提取样品的总 RNA; mRNA 的分离则采用 mRNA 纯化试剂盒(Promega); cDNA 文库构建方法(cDNA Synthesis Kit, ZAP-cDNA₃ Synthesis Kit, ZAP-cDNA₃ Gigapeak₃ III Gold Cloning Kit 均购自于 Stratagene 公司)同以前的报道^[12-13]。构建好的 cDNA 文库采用蓝白斑筛选来确定重组率和 cDNA 文库库容量的大小; 通过大规模载体切除方法将 cDNA 文库由 λ-噬菌体载体系统转变为质粒载体系统并转化大肠杆菌 SOLR; 随机挑选 10 个克隆, 利用载体引物 T3(5'-ATTAACCCCTCACTAAAGGGA-3')和 T7(5'-GCGTAATACGACTCACTA TA-3')进行 PCR 扩增, 确定 cDNA 插入片段的长度; 并采用同样的方法来筛选 cDNA 克隆, 片段长度大于 0.6 kb 的用于测序; 随机挑选 PCR 筛选后的克隆, 37℃ 液体培养过

收稿日期: 2005-12-05; 修订日期: 2006-03-07。

基金项目: 国家留学基金项目(BK0497); 湖南农业大学稳定人才研究基金(2000-09)。

作者简介: 陈 韬(1969-), 男, 博士, 副教授, 主要从事分子生物学研究。E-mail: chentao_114@163.com

夜后,提取质粒(Qiagene)。利用载体引物 T7 为测序引物,用 ABI377 XLDNA 自动测序仪进行单向部分测序,测序工作由中南大学湘雅医学院中国医学遗传学国家重点实验室完成。

利用 Staden Package^[14]对测序获得的结果进行分析,去掉每个序列中的载体序列和不能辨认的区域。处理后的序列分别用 BlastX、BlastN 在 nr-as 和 nr-nt 中进行同源性搜索^[15],BLAST 的阈值设为低于 e^{-8} ,对于少数序列含有较长的 3'-UTR,编码序列较短,虽然 $E > e^{-8}$,但氨基酸的相似性很高,也可以作为初步鉴定的依据。

2 结果

2.1 cDNA 文库的质量

来自于肝胰脏的 mRNA 用于合成 cDNA 并克隆到 Uni-ZAP XR 载体中。将文库与受体菌 XL1-Blue MRF' 涂布在含有 IPTG 和 X-Gal 的 LB 平板上,蓝白斑结果显示 cDNA 文库库容量为 6.5×10^6 pfu/mL,这一库容量基本上可以代表草鱼肝胰脏表达的 mRNA,98.7% 的克隆含有 cDNA 插入片段。PCR 结果显示,随机挑取的 20 个克隆插入片段平均长度为 1.5 kb,片段最短的为 0.4 kb,片段最长的为 3.2 kb。

2.2 EST 文库的统计结果

共对 768 个随机挑选的 cDNA 克隆进行了测序,获得了 567 个序列(73.83%, 567/768)。同源性搜索和初步鉴定结果将 567 个序列分成 2 大类:457 个序列归类为 130 个不同的转录本,与编码已知功能蛋白的 cDNA 具有较高的同源性;其余的 110 个序列都具有 1 个单一的开放阅读框,与某些未知功能的蛋白质具有一定的同源性,被称为假定蛋白。因此,总共得到了 240 个不同的表达序列。由于 cDNA 文库是未经扩增的初级文库,并且用于测序的 cDNA 克隆也没有经过任何削减,所以得到的 EST 的丰度基本反映了对应的 mRNA 的丰度(图 1),因为不同的 mRNA 丰度不同,本研究将 EST 文库归类为不同的簇,即基因,每个簇由编码同一个蛋白质的序列组成。在已知的 130 个 EST 簇中,根据其丰度可以将 EST 簇分成 3 个不同的水平:低丰度簇:1~2 个拷贝,其中单拷贝基因有 46 个,占总数的 35.66%,双拷贝基因有 33 个,占 25.58%;中等丰度:3~9 个拷贝;高丰度簇:拷贝数大于 10,属于这一簇的基因比较少,其中丰度最高的簇是 apolipoprotein A-I、apolipoprotein、trypsinogen a 和 fatty acid binding protein,其拷贝数分别为 29、24、24 和 32。

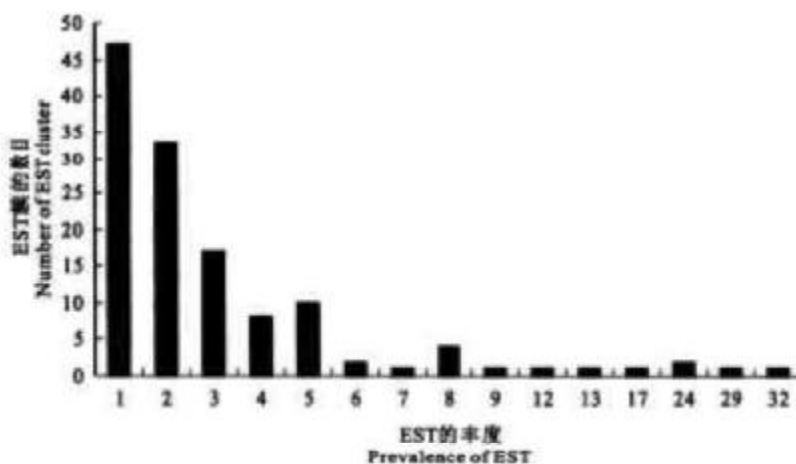


图 1 已鉴定表达序列标签的丰度分布

Fig.1 Prevalence distribution of identified ESTs

2.3 EST 的功能分类

为了更好地描述所得到的 EST 文库,按照 Expressed Gene Anatomy Database(EGAD)^[16]的分类方法,对编码已知蛋白质的 EST 进行功能分类。

(表 1 和表 2)。其中大量的 EST 编码与细胞代谢和基因表达有关的蛋白质,前者分别占 EST 簇数和总克隆数的 31.54% 和 46.83%,后者分别为 43.08% 和 34.57%,这两大类 EST 中,不仅基因的数目多,

同时很多基因的拷贝数也比较多。数量相对较少的是与细胞信号传导和细胞通讯、细胞与机体防御有关的 EST,除 α 和 β 珠蛋白以及 transferrin variant A

具有较多的拷贝数以外,绝大部分 EST 都只有1~3个拷贝。数量最少的是与结构和运动有关的,只有2个基因,各有1个拷贝。

表1 草鱼肝胰脏的表达序列标签

Tab.1 ESTs from liver and pancreas(hepatopancreas) of grass carp, *Ctenopharyngodon idella*

GenBank 接受号 GenBank AC. No.	假定功能 Putative Function	E 值 E Value	最接近种 Closest Species	拷贝数 Copies
结构与运动 Structure and motility				
CF887837	Alpha globin	2.00E-69	<i>Ctenopharyngodon idella</i>	17
CK232970	Beta globin	1.00E-80	<i>Ctenopharyngodon idella</i>	13
CK233040	Heparan sulfate proteoglycan	1.00E-11	<i>Homo sapiens</i>	1
CK233053	Myosin regulatory light chain 2	7.00E-35	<i>Gallus gallus</i>	1
代谢 Metabolism				
CK232813	3-oxoacid CoA transferase	1.00E-20	<i>Homo sapiens</i>	3
CF887834	Alcohol dehydrogenase	2.00E-16	<i>Danio rerio</i>	1
CK233146	Alcohol dehydrogenase 8a	3.00E-23	<i>Danio rerio</i>	1
CK232925	Alpha1 microglobulin/bikunin protein a	4.00E-82	<i>Danio rerio</i>	6
CF887840	Alpha1 microglobulin/bikunin protein b	6.00E-61	<i>Salmo salar</i>	2
CK232929	Amylase	8.00E-75	<i>Danio rerio</i>	2
CK232940	Apolipoprotein	1.00E-16	<i>Anguilla japonica</i>	24
CF887896	Apolipoprotein A- I	9.00E-20	<i>Cyprinus carpio</i>	29
CK232958	Apolipoprotein C- I	4.00E-06	<i>Canis familiaris</i>	2
CF887911	Apolipoprotein C II	5.00E-14	<i>Oncorhynchus mykiss</i>	4
CK233179	Apolipoprotein Eb	2.00E-57	<i>Danio rerio</i>	1
CK232963	ATP synthase coupling factor 6	3.00E-31	<i>Xenopus laevis</i>	4
CK232967	ATP synthase epsilon	1.00E-10	<i>Xenopus laevis</i>	1
CK232980	Carboxylester lipase	9.00E-75	<i>Danio rerio</i>	2
CK232982	Carboxypeptidase A1	1.00E-79	<i>Danio rerio</i>	1
CK232987	Chymotrypsinogen a	2.00E-72	<i>Danio rerio</i>	1
CK232988	Chymotrypsinogen b	2.00E-70	<i>Danio rerio</i>	8
CK232992	Chymotrypsinogen c	5.00E-25	<i>Gadus morhua</i>	2
CK232998	Cytochrome c oxidase subunit I	1.00E-80	<i>Lefua echigonia</i>	2
CK233000	Cytochrome c oxidase subunit II	3.00E-69	<i>Cyprinus carpio</i>	1
CK233002	Cytochrome c oxidase subunit III	1.00E-83	<i>Carassius auratus</i>	8
CF887853	Cytochrome P450	2.00E-56	<i>Danio rerio</i>	1
CK233008	Cytochrome P450 51A1	4.00E-26	<i>Homo sapiens</i>	2
CK233010	Delta-5/delta-6 fatty acid desaturase	2.00E-64	<i>Danio rerio</i>	2
CK233012	Dihydrofolate reductase	6.00E-82	<i>Danio rerio</i>	2
CK233019	Elastase a	1.00E-74	<i>Paralichthys olivaceus</i>	5
CF887858	Elastase b	2.00E-70	<i>Paralichthys olivaceus</i>	3
CF887855	Elastase c	3.00E-70	<i>Gadus morhua</i>	3
CF887854	Elastase d	3.00E-82	<i>Danio rerio</i>	9
CK233197	Fatty acid-binding protein	1.00E-61	<i>Danio rerio</i>	32
CF887869	Formiminotransferase cyclodextrinase	5.00E-75	<i>Danio rerio</i>	2
CK233175	Glucose transporter X	1.00E-25	<i>Ctenopharyngodon idella</i>	2
CK233045	Inositol oxygenase	1.00E-17	<i>Mus musculus</i>	1
CK233050	Macroglobulin	9.00E-54	<i>Ctenopharyngodon idella</i>	1
CF887876	Methyltransferase	3.00E-14	<i>Rattus norvegicus</i>	2
CF887880	NADH-ubiquinone oxidoreductase MLRQ subunit	2.00E-39	<i>Danio rerio</i>	2
CF887883	Quiescent cell Pro dipeptidase	5.00E-64	<i>Homo sapiens</i>	2

GenBank 接受号 GenBank AC. No.	假定功能 Putative Function	E 值 E Value	最接近种 Closest Species	拷贝数 Copies
代谢 Metabolism				
CK233062	Succinyl-CoA synthetase	1.00E-34	<i>Columba livia</i>	1
CF887888	Trypsinogen a	5.00E-77	<i>Danio rerio</i>	24
CK233091	Trypsinogen b	9.00E-78	<i>Danio rerio</i>	4
CK233093	Ubiquinol-cytochrome C reductase FeS protein	5.00E-79	<i>Danio rerio</i>	2
细胞信号传导与细胞通讯 Cell signaling/Cell communication				
CX129925	Acidic nuclear phosphoprotein 32 family member B	8.00E-25	<i>Danio rerio</i>	1
CK232968	Beta adaptin	6.00E-38	<i>Homo sapiens</i>	1
CF887845	Cellular retinol-binding protein	2.00E-71	<i>Danio rerio</i>	2
CK233167	Chloride intracellular channel 4	8.00E-36	<i>Danio rerio</i>	3
CK233038	Ganglioside M2 activator protein	1.00E-07	<i>Homo sapiens</i>	1
CK233039	G-protein beta subunit	2.00E-92	<i>Danio rerio</i>	1
CF887877	Muscle-specific beta 1 integrin binding protein	6.00E-53	<i>Danio rerio</i>	3
CK233177	Nuclear receptor subfamily 2, group F, member 2	1.00E-43	<i>Danio rerio</i>	1
CK233057	Phosphatidylinositol-4-phosphate 5-kinase	5.00E-67	<i>Mus musculus</i>	2
CK233060	Rho GTPase activating protein	1.00E-32	<i>Xenopus laevis</i>	1
CK233061	S100 Ca-binding protein	3.00E-34	<i>Danio rerio</i>	1
细胞与机体防御 Cell/organism defense				
CK232919	Alpha1 antitrypsin a	5.00E-62	<i>Cyprinus carpio</i>	5
CF887848	Complement C3	1.00E-63	<i>Cyprinus carpio</i>	1
CK232995	Complement C9	1.00E-36	<i>Danio rerio</i>	1
CK232997	Cystatin A1	3.00E-17	<i>Sus scrofa</i>	3
CK233013	DnaJ-like protein	1.00E-14	<i>Homo sapiens</i>	2
CK233113	Ferritin, heavy polypeptide 1	1.00E-13	<i>Oncorhynchus nerka</i>	2
CK233036	Fetuin long form	6.00E-51	<i>Cyprinus carpio</i>	4
CK233162	Fibronectin	1.00E-09	<i>Danio rerio</i>	1
CK233044	Immunoglobulin light chain	1.00E-30	<i>Cyprinus carpio</i>	3
CF887873	Inhibitor of apoptosis protein	8.00E-97	<i>Danio rerio</i>	1
CK233047	Invariant chain-like protein	5.00E-23	<i>Danio rerio</i>	3
CF887875	Lectin precursor	8.00E-22	<i>Pleurodeles waltl</i>	3
CK233056	Pentraxin C-reactive protein	5.00E-48	<i>Cyprinus carpio</i>	1
CK233154	Polymeric immunoglobulin receptor	7.00E-15	<i>Danio rerio</i>	1
CK233058	Polymyositis/scleroderma autoantigen	2.00E-80	<i>Homo sapiens</i>	2
CF887884	Transferrin variant A	7.00E-35	<i>Carassius auratus</i>	8
CF887892	Warm-temperature-acclimation-related-65	6.00E-54	<i>Oryzias latipes</i>	1
CK233096	Zuotin related factor-1	5.00E-03	<i>Rattus norvegicus</i>	1
基因与蛋白表达 Gene/protein expression				
CF887793	40S ribosomal protein S10	5.00E-78	<i>Ictalurus punctatus</i>	8
CK232819	40S ribosomal protein S14	3.00E-74	<i>Ictalurus punctatus</i>	3
CK232822	40S ribosomal protein S17	4.00E-66	<i>Ictalurus punctatus</i>	1
CF887792	40S ribosomal protein S18	2.00E-66	<i>Danio rerio</i>	1
CF887796	40S ribosomal protein S19	1.00E-73	<i>Ictalurus punctatus</i>	3
CF887798	40S ribosomal protein S2	6.00E-78	<i>Ictalurus punctatus</i>	4
CF887799	40S ribosomal protein S23	1.00E-77	<i>Ictalurus punctatus</i>	2
CF887800	40S ribosomal protein S26	8.00E-44	<i>Ictalurus punctatus</i>	3
CK232833	40S ribosomal protein S27	1.00E-34	<i>Ictalurus punctatus</i>	5
CK232835	40S ribosomal protein S27a	2.00E-62	<i>Danio rerio</i>	2
CK232836	40S ribosomal protein S3	4.00E-52	<i>Ictalurus punctatus</i>	1

GenBank 接受号 GenBank AC. No.	假定功能 Putative Function	E 值 E Value	最接近种 Closest Species	续表
				拷贝数 Copies
	基因与蛋白表达 Gene/protein expression			
CK232837	40S ribosomal protein S30	6.00E-52	<i>Danio rerio</i>	2
CK232839	40S ribosomal protein S3a	4.00E-67	<i>Danio rerio</i>	1
CK232842	40S ribosomal protein S4	6.00E-80	<i>Ictalurus punctatus</i>	3
CK232844	40S ribosomal protein S8	1.00E-79	<i>Ictalurus punctatus</i>	5
CF887802	40S ribosomal protein Sa	1.00E-76	<i>Danio rerio</i>	4
CF887803	60S acidic ribosomal protein P0	5.00E-64	<i>Danio rerio</i>	5
CF887804	60S acidic ribosomal protein P1	7.00E-27	<i>Ictalurus punctatus</i>	2
CF887805	60S acidic ribosomal protein P2	2.00E-16	<i>Homo sapiens</i>	1
CF887806	60S ribosomal protein L10	9.00E-77	<i>Ictalurus punctatus</i>	2
CF887807	60S ribosomal protein L10a	2.00E-90	<i>Ictalurus punctatus</i>	3
CF887809	60S ribosomal protein L11	4.00E-96	<i>Ictalurus punctatus</i>	1
CF887810	60S ribosomal protein L12	3.00E-83	<i>Ictalurus punctatus</i>	12
CF887812	60S ribosomal protein L13	7.00E-83	<i>Danio rerio</i>	2
CK232868	60S ribosomal protein L13a	3.00E-78	<i>Ictalurus punctatus</i>	5
CK232870	60S ribosomal protein L15	1.00E-71	<i>Ctenopharyngodon idella</i>	2
CF887816	60S ribosomal protein L17	1.00E-76	<i>Danio rerio</i>	4
CK232872	60S ribosomal protein L18a	4.00E-67	<i>Danio rerio</i>	1
CK232873	60S ribosomal protein L23	1.00E-72	<i>Danio rerio</i>	1
CF887818	60S ribosomal protein L23a	3.00E-66	<i>Ictalurus punctatus</i>	7
CK232877	60S ribosomal protein L24	7.00E-53	<i>Danio rerio</i>	1
CK232878	60S ribosomal protein L26	3.00E-66	<i>Danio rerio</i>	3
CK232880	60S ribosomal protein L27	3.00E-34	<i>Danio rerio</i>	1
CK232882	60S ribosomal protein L27a	6.00E-73	<i>Oncorhynchus mykiss</i>	5
CK232885	60S ribosomal protein L29	9.00E-18	<i>Macaca fascicularis</i>	4
CK232886	60S ribosomal protein L3	2.00E-67	<i>Ictalurus punctatus</i>	3
CF887826	60S ribosomal protein L31	2.00E-52	<i>Ictalurus punctatus</i>	2
CK232893	60S ribosomal protein L32	8.00E-63	<i>Epinephelus coioides</i>	5
CF887828	60S ribosomal protein L34	1.00E-59	<i>Danio rerio</i>	5
CK232897	60S ribosomal protein L37	2.00E-48	<i>Ictalurus punctatus</i>	1
CK232898	60S ribosomal protein L4	4.00E-66	<i>Danio rerio</i>	1
CK232899	60S ribosomal protein L7	1.00E-89	<i>Ictalurus punctatus</i>	6
CK232903	60S ribosomal protein L7a	9.00E-59	<i>Danio rerio</i>	1
CF887832	60S ribosomal protein L9	2.00E-65	<i>Ictalurus punctatus</i>	2
CK232927	Alpha-NAC protein	1.00E-51	<i>Danio rerio</i>	3
CK232962	Asparaginyl-tRNA synthetase, cytoplasmic	5.00E-67	<i>Mus musculus</i>	2
CF887870	Histone deacetylase 3	2.00E-04	<i>Gallus gallus</i>	1
CK233041	I-kappa-B-epsilon protein	3.00E-07	<i>Mus musculus</i>	2
CF887879	N27C7-4 protein	6.00E-15	<i>Homo sapiens</i>	1
CK233055	Partner of Nob 1	1.00E-44	<i>Oryzias latipes</i>	1
CF887882	Peptidylprolyl isomerase	5.00E-36	<i>Mus musculus</i>	1
CK233063	T-box transcription factor eomesodermin	5.00E-29	<i>Danio rerio</i>	1
CF887860	Translation elongation factor eEF-1-alpha	8.00E-47	<i>Cyprinus carpio</i>	5
CF887861	Translation elongation factor eEF-2	1.00E-11	<i>Danio rerio</i>	2
CF887886	Translation initiation factor eIF-2-gamma	7.00E-48	<i>Danio rerio</i>	2
CK233095	Zinc finger protein subfamily 1A 5	8.00E-40	<i>Danio rerio</i>	1

表 2 编码已知蛋白的 EST 的功能分类
Tab.2 Functional category of ESTs encoding known proteins

EST 功能分类 EST functional category	EST 簇 EST clusters		克隆 Clones	
	数目 No.	百分比 %	数目 No.	百分比 %
细胞结构和运动 Structure and motility	4	3.76	32	87.00
代谢 Metabolism	41	31.54	207	45.30
细胞信号传导与细胞通讯 Cell signaling and cell communication	11	8.46	17	3.76
细胞与机体防御 Cell and organism defense	18	13.85	43	9.51
基因与蛋白质表达 Gene and protein expression	56	43.08	15	34.57

3 讨论

3.1 EST 文库的代表性

肝的主要功能是合成大量的蛋白和解毒,是很多物质分解代谢的场所,胰腺则是机体最重要的消化腺内分泌腺,分泌多种消化酶和激素,与消化以及糖类、蛋白质和脂类的代谢有密切的关系。本研究获得了 567 个 EST,除 110 个是编码未知功能的蛋白质以外,其余的 457 个 EST 分别属于 129 个不同的基因。在 129 个不同的基因中,除 5 个是以前鉴定的外,124 个都是首次在草鱼中报道。这些基因主要由以下 5 个部分组成:与代谢有关,与基因和蛋白质表达有关,细胞与机体防御,细胞信号传导与细胞通讯,细胞结构和运动。和已知的哺乳动物肝脏表达基因相比^[17],一些在人类和家鼠肝脏中丰富 mRNA 的编码基因,同样也出现在草鱼肝胰脏中,这些表达基因的组成基本反映肝脏和胰脏的功能特点,如编码 apolipoprotein、apolipoproteinA-I、fatty acid-binding protein 等基因,以及一些胰腺特有的编码基因,如编码 trypsinogen、chymotrypsinogen 和 elastase 等基因,这种保守的高水平基因表达意味着与肝脏和胰脏基本功能有关的成分的进化在 4~4.5 亿年前硬骨鱼类和四足动物分歧时就已完成。

3.1.1 代谢 由于比较旺盛的物质和能量代谢,肝胰脏内与代谢有关的基因表达水平相对较高。在报道的 129 个不同的基因中,有 40 个基因是与机体的物质和能量代谢有关,其功能涉及以下几个方面。(1)较高丰度的具有与物质的结合和运输功能的蛋白,如 microglobulin、macroglobulin、apolipoprotein 和 fatty acid binding protein 等。其中,apolipoprotein 和 fatty acid binding protein 的丰度非常高,这些基因的产物参与脂类的结合与运输,与脂肪酸、胆固醇的合成与代谢有着密切的关系;(2)胰腺分泌的一些消化酶,如 amylase、carboxypeptidase、chymotrypsinogen、

elastase 和 trypsinogen 等;与物质和能量代谢有关的蛋白及酶,如 3-oxoacid CoA transferase、succinyl-CoA synthetase、ATP synthase 亚基与偶联因子、cytochrome c oxidase 的不同亚基、delta-5/delta-6 fatty acid desaturase、dihydrofolate reductase、formimino-transferase cyclodeaminase^[18]等等;还有一类与肝脏的解毒功能有关,如 alcohol dehydrogenase、cytochrome P450^[19]等。

3.1.2 基因与蛋白质表达 肝脏是合成许多蛋白质的主要场所,因此肝胰脏 EST 中有很大一部分与基因与蛋白质表达有关。其中大部分是核糖体蛋白质(44),分别组成核糖体的大亚基和小亚基,这些基因除 L12 有 12 个拷贝外,大部分的基因为 1~5 个拷贝;还有一些基因与转录调控有关,如 histone deacetylase 3(HDAC3)^[20-21]、I-kappa-B-epsilon protein^[22-23]、partner of Nob 1^[9]、T-box transcription factor coesodermin、zinc finger protein subfamily 1A 5 通过与转录因子的相互作用,来调节不同基因的表达;其余的基因包括翻译起始因子和延伸因子,以及与蛋白质合成后加工有关的 alpha-NAC protein。

3.1.3 细胞与机体防御 在 EST 文库中鉴定了不少与细胞和机体防御有关的基因(18),其中有的基因是与免疫作用有关,如 immunoglobulin light chain、complement C3 和 C9,作为免疫活性分子,直接参与免疫反应^[24]。polymeric immunoglobulin receptor 是一种跨膜蛋白,负责 IgA 二聚体和 IgM 五聚体的运输^[25]。pentraxin CRP 是一些与炎症反应有关细胞因子的效应物,能激活补体、与 IgG 结合、识别病原微生物以及受损细胞的核成分,协助巨噬细胞的吞噬作用等^[26]。fetuin long form 是 nephrosin 的抑制因子,与鱼类的免疫和造血有关^[27]。有的基因是蛋白酶抑制剂,如 alpha-1-antitrypsin 是主要的控制组织蛋白降解的蛋白酶的抑制剂,当它的水平下降时,可以引起胶原蛋白代谢的变化,引起很多疾

病^[28-29],cystatin A1 则是另一类蛋白酶抑制剂,主要抑制半胱氨酸蛋白酶;也有一些基因是通过调节细胞内的物质代谢过程来保护细胞或机体,如 ferritin,是存在于细胞内的一种主要的铁结合蛋白,可以保护细胞免受高铁的氧化作用带来的损害^[30-31]或清除来自细菌等微生物的游离铁,transferrin 则是一种存在于血浆中的结合 Fe^{3+} 的蛋白,并通过与 transferrin 受体结合以吞噬作用进入细胞^[32],抑制因铁参与的氧化反应产生的自由基造成的损害^[33]。

3.2 基因重复或等位基因变异

基因和基因组重复在脊椎动物,尤其在鱼类的进化中是经常发生的,对于新基因座的产生具有十分重要的意义^[34]。近年来人们在鱼类中发现了大量的重复基因^[35-36]。等位基因变异是在气候、栖息环境等多种因素影响下造成的遗传物质的改变^[37],和重复基因相比,等位基因之间可能具有更

高的同源性。在鉴定的草鱼肝胰脏 EST 中,发现胰蛋白酶原、弹性蛋白酶和糜蛋白酶原 3 个基因分别具有 2、4 和 3 个不同的拷贝(表 1)。部分序列比较结果表明,这些基因的不同成员之间存在不等的同源性(图 2)。2 个胰蛋白酶原基因 A 和 B 部分序列相似性仅为 68.67%;在 4 个弹性蛋白酶基因中,A 和 B 的相似性为 97.08%,而其他拷贝之间的相似性分别为 56.2% (A 与 C)、54.74% (A 与 D)、58.39% (B 与 C)、55.47% (B 与 D)和 68.31% (C 与 D);在 3 个糜蛋白酶原基因中,A 与 B 的相似性为 94.23%,A 与 C 为 80.36%,B 与 C 为 82.14%。根据这 3 个基因不同成员之间相似性的比较,笔者推测弹性蛋白酶的 A 和 B 以及糜蛋白酶原的 A 与 B 可能是由于等位基因的突变而产生的,而这些基因的其他成员则可能是基因组重复造成的,不过这种推测还有待于通过基因组序列来证实。



图2 部分推导氨基酸序列比较

注:黑色表示氨基酸残基相同,灰色表示部分残基相同,括号里的数字为拷贝数。

Fig.2 Alignment of partial deduced amino acid sequence

Note: Black fonts show amino acid residues are same completely, Grey fonts indicate part of residues are same. Numbers in brackets mean copies.

一般情况下,不同细胞在某个发育阶段表达的基因数在 $10^3 \sim 10^4$ 之间,由于本研究测定的 cDNA 序列数目有限,远远低于细胞表达的基因数,因此尚不足以揭示草鱼肝胰脏基因表达的特点,只能反映一些丰富基因的组成。但大量基因的鉴定,特别是一些与代谢、免疫作用有密切关系的基因的鉴定,更有助于对草鱼的生理学、免疫学方面的进一步研究。

参考文献:

- [1] Feldmann G. Tissue and cellular organisation of the liver [A]. Liver Gene Expression [C]. Austin: Landes R G Company, Austin, TX, 1994. 17-34.
- [2] Derman E, Krutler K, Walling L, et al. Transcriptional control in the production of liver specific mRNAs [J]. Cell, 1981, 23(3): 731-739.
- [3] Ho W K, Tsang W H, Dias N P. Cloning of the grass carp growth hormone cDNA [J]. Biochem Biophys Res Commun, 1989, 161(3): 1239-1243.
- [4] Ho W K, Wong M W, Chan A P. Cloning and sequencing of the grass carp (*Ctenopharyngodon idellus*) growth hormone gene [J]. Biochem Biophys Acta, 1991, 1090(2): 245-248.
- [5] Zhu Z, He L, Chen T T. Primary-structural and evolutionary analyses of the growth-hormone gene from grass carp (*Ctenopharyngodon idellus*) [J]. Eur J Biochem, 1992, 207(2): 643-648.
- [6] Chan W M, Tang W K, Cheng C K, et al. Purification, N-terminal sequence determination and enzymatic characterization of anti-quitin from the liver of the grass carp [J]. Comparative Biochemistry and Physiology (Part B), 2003, 136(3): 443-450.
- [7] Chang M X, Nie P, Xie H X, et al. Characterization of two genes encoding leucine-rich repeat-containing proteins in grass carp *Ctenopharyngodon idellus* [J]. Immunogenetics, 2005, 56(10): 710-721.
- [8] Fong W, Choy K. Purification and characterization of grass carp mitochondrial aldehyde dehydrogenase [J]. Chem Biol Interact, 2001, 130-132(1-3): 161-171.
- [9] Zhou G J, Zhang Y, Wang J, et al. Cloning and characterization of a novel human RNA binding protein gene PNO1 [J]. DNA Seq, 2004, 15(3): 219-224.
- [10] Ng P K, Wu R S, Zhang Z P, et al. Molecular cloning and characterization of a hypoxia-responsive CITED3 cDNA from grass carp [J]. Comp Biochem Physiol B Biochem Mol Biol, 2003, 136(2): 163-172.
- [11] Adams M D, Kelley J M, Gocayne J D, et al. Complementary DNA sequencing: expressed sequence tags and human genome project [J]. Science, 1991, 252(5013): 1651-1656.
- [12] Chen T, Reith M E, Rans N W, et al. Expressed sequence tag (EST)-based characterization of gene regulation in *Artemia* larvae [J]. Invert Reprod Dev, 2003, 44(1): 33-44.
- [13] 陈 娟, 苏建明, 文祝友, 等. 扇贝性腺和肠组织表达序列标签的初步识别 [J]. 农业生物技术学报, 2003, 11(4): 407-411.
- [14] Bonfield J K, Smith K, Staden R. A new DNA sequence assembly program [J]. Nucl Acids Res, 1995, 23(24): 4992-4999.
- [15] White O, Kerlavage A R. TDB: new databases for biological discovery [J]. Methods Enzymol, 1996, 266: 27-40.
- [16] Kawamoto S, Matsumoto Y, Mizuno K, et al. Expression profiles of active genes in human and mouse livers [J]. Gene, 1996, 174(1): 151-158.
- [17] Mao Y, Vyas N K, Vyas M N, et al. Structure of the bifunctional and Golgi-associated formiminotransferase cyclodextrinase octamer [J]. EMBO J, 2004, 23(15): 2963-2971.
- [18] Martin S A, Caplice N C, Davey G C, et al. EST-based identification of genes expressed in the liver of adult Atlantic salmon (*Salmo salar*) [J]. Biochem Biophys Res Commun, 2002, 293(1): 578-585.
- [19] Mahlknecht U, Will J, Varin A, et al. Histone deacetylase 3, a class I histone deacetylase, suppresses MAPK11-mediated activating transcription factor-2 activation and represses TNF gene expression [J]. J Immunol, 2004, 173(6): 3979-3990.
- [20] Zhang X, Wharton W, Yuan Z, et al. Activation of the growth-differentiation factor 11 gene by the histone deacetylase (HDAC) inhibitor trichostatin A and repression by HDAC3 [J]. Mol Cell Biol, 2004, 24(12): 5106-5118.
- [21] Spieker M, Darius H, Liao J K. A functional role of I kappa B-epsilon in endothelial cell activation [J]. J Immunol, 2000, 164(6): 3316-3322.
- [22] Whiteside S T, Epinat J C, Rice N R, et al. A novel member of the I kappa B family, controls RelA and c-Rel NF-kappa B activity [J]. EMBO J, 1997, 16(6): 1413-1426.
- [23] Cahir-McFarland E D, Carter K, Rosenawald A, et al. Role of NF-kappa B in cell survival and transcription of latent membrane protein 1-expressing or Epstein-Barr virus latency II-infected cells [J]. J Virol, 2004, 78(8): 4108-4119.
- [24] Ducloux T W, Mold C. C-reactive protein: an activator of innate immunity and a modulator of adaptive immunity [J]. Immunol Res, 2004, 30(3): 261-277.
- [25] Hamburger A E, West A P, Bjorkman P J. Crystal structure of a polymeric immunoglobulin binding fragment of the human polymeric immunoglobulin receptor [J]. Structure (Camb), 2004, 12(11): 1925-1935.
- [26] Garlanda C, Bottazzi B, Bastone A, et al. Pentoxons at the Crossroads Between Innate Immunity, Inflammation, Matrix Deposition and Female Fertility [J]. Annu Rev Immunol, 2005, 23(23): 337-366.
- [27] Tai P L, Chen C H, Huang C J, et al. Purification and cloning of an endogenous protein inhibitor of carp nephrokin, an atypical metalloproteinase [J]. J Biol Chem, 2004, 279(12): 11146-11155.
- [28] Hauck E W, Hauptmann A, Haug S M, et al. Alpha-1-antitrypsin levels and genetic variation of the alpha-1-antitrypsin gene in Peyronie's disease [J]. Eur Urol, 2004, 46(5): 623-628.

- [29] Khoo A. Alpha-1 antitrypsin deficiency: A study from the registry[J]. *Hum Pathol*, 2004, 35(12): 1433-1434.
- [30] Epstein S, Glickstein H, Picard V, et al. H-ferritin subunit over-expression in erythroid cells reduces the oxidative stress response and induces multidrug resistance properties[J]. *Blood*, 1999, 94(10): 3593-3603.
- [31] Balla G, Jacob HS, Balla J, et al. Ferritin: a cytoprotective antioxidant strategem of endothelium[J]. *J Biol Chem*, 1992, 267(25): 18148-18153.
- [32] Cheng Y, Zak O, Aisen P, et al. Structure of the human transferrin receptor-transferrin complex[J]. *Cell*, 2004, 116(4): 565-576.
- [33] Van Rensburg S J, Varney J, Hon D, et al. Biochemical model for inflammation of the brain: the effect of iron and transferrin on monocytes and lipid peroxidation[J]. *Metab Brain Dis*, 2004, 19(1-2): 97-112.
- [34] Volff J N, Schard M. Evolution of signal transduction by gene and genome duplication in fish[J]. *J Struct Funct Genomics*, 2003, 3(1-4): 139-150.
- [35] Wang X L, Mei J, Sun M, et al. Identification of three duplicated Spin genes in medaka (*Oryzias latipes*) [J]. *Gene*, 2005, 350(2): 99-106.
- [36] Liu R Z, Sun Q, Thiesse C, et al. The cellular retinol-binding protein genes are duplicated and differentially transcribed in the developing and adult zebrafish (*Danio rerio*) [J]. *Mol Biol Evol*, 2005, 22(3): 469-477.
- [37] Pomortsev A A, Kalashnikov B A, Lialina E V. Uniformity of distributing allelic variants of three hordein-coding loci of spring barley in the Russian Federation [J]. *Genetics*, 2001, 37(11): 1522-1529.

Identification of genes expressed in hepatopancreas from grass carp, *Ctenopharyngodon idellus*

CHEN Tao¹, CHEN Li-xiang¹, SU Jian-ming¹, ZHANG Hui-yun²

(1. Animal Science and Technology College, Hunan Agricultural University, Changsha 410128, China; 2. Life Science and Technology College, Central South University of Forestry & Technology, Changsha 412006, China)

Abstract: The hepatopancreas of cyprinoid fish plays an important role in carbohydrate, lipid, amino acid, protein, and prostaglandin metabolism as well. Due to its vital function and relative lack of cell type complexity, the hepatopancreas is an ideal model for studying cyprinoid fish expressed genes relating to some important physiological functions such as metabolism and immunity. However, far less is known of gene expression in hepatopancreas of the grass carp, *Ctenopharyngodon idellus*, a commercially important species in fresh water aquaculture in China. Most previous studies of grass carp have focused on the description of individual genes. To get a basic insight on gene expression, to obtain genetical and molecular biological information to the studies of some important physiological procession in grass carp, a rapid method of novel gene identification using partial sequencing of cloned cDNA to produce expressed sequence tag (EST) was explored in this context. Total RNA was isolated from grass carp hepatopancreas using TRIzol reagent (GIBCOBRL) and mRNA was purified by oligo(dT) magnet beads (Promega). Then a cDNA library was constructed using a cDNA cloning kit (Stratagene). After massive excise, the cDNA library was converted from λ -phage system into plasmid system and plated onto LB agar plate. Single colonies were randomly picked up and screened by PCR and only the clones with cDNA inserts more than 0.6 kb were used for sequencing. 768 cDNA clones were partially sequenced and 567 ESTs were obtained. BLAST result showed that 457 of the hepatopancreas ESTs could be ascribed to the transcriptional products of 129 identified genes including 5 previously reported and 124 firstly identified in the grass carp, whereas 110 ESTs exhibited relatively low homology with unidentified sequences and were called hypothetical proteins. 129 known genes were classified into 5 function categories: 1) metabolism (41); 2) gene and protein expression (56); 3) cell signaling and communication (11); 4) cell and organism defense (18); 5) cell structure and motility (4), which provide a genetical background for our molecular biology studies on metabolism and immunology in grass carp. [Journal of Fishery Sciences of China, 2006, 13(5): 714-722]

Key words: grass carp; hepatopancreas; cDNA library; expressed sequence tag