

草鱼肠道 cDNA 文库构建及部分 ESTs 分析

张学俊^{1,2} 屈刚¹ 朱文漓¹ 张黎¹ 王娟¹ 吴江²
刘汉元² 陈放¹ 徐恒^{1,2}

(1. 生物资源与生态环境教育部重点实验室, 四川大学生命科学学院, 成都 610064; 2. 通威股份有限公司, 成都 610081)

摘要: 本研究以雌性草鱼肠道为实验材料, 采用非均一化的 Oligo-dT 引物定向克隆技术构建了草鱼肠道组织的 cDNA 文库, 对文库质量的分析结果表明: cDNA 文库的库容量至少为 2.3×10^5 , 重组率达 95%, 平均插入片断长度大于 1000bp。挑取 cDNA 克隆进行 5' 端测序, 总共进行了 1571 个成功反应, 其中 1411 条 ESTs 长度大于 100bp, 初步拼接得到 939 个单基因簇 (Unigene), 其中包括 188 个重叠群 (Contigs), 751 个单拷贝 EST (Singletons)。使用 BLAST 软件将这些序列同 GenBank 等数据库进行比对、查询和注释, 结果显示 418 条序列有相关同源性, 其他的 521 (55.5%) 条序列没有明显的同源性 ($E\text{-value} \leq 1.00E-10$), 但是也同时揭示出在这些 ESTs 中能够发现新功能基因的相当可能性。本研究结果对草鱼以及其他鲤科鱼类的消化系统功能基因的筛选和发现具有指导意义。此外, 草鱼肠道 cDNA 文库的构建和 EST 测序工作将为草鱼基因组计划的实施奠定基础。

关键词: 草鱼; 肠道; cDNA 文库; EST; 测序**中图分类号:** Q343.1 **文献标识码:** A **文章编号:** 1000-3207(2007)02-0251-08

草鱼 (*Ctenopharyngodon idellus*, 英文名 grass carp): 鲤科、草鱼属、草鱼种, 又名鲩鱼, 是我国特有的鲤科大型经济鱼类, 广泛分布在我国南北水域, 是我国淡水养殖的当家鱼。据统计, 2004 年草鱼的产量约占全国淡水鱼类养殖产量的 20%, 产量达 369.84 万吨, 市场销售额超过 300 亿元人民币, 为全球产量和消费量最大的鱼类之一。同时, 其独特的草食习性可将植物蛋白最高效的转化为优质动物蛋白, 被誉为“水中的牛羊”。

从理论上说, EST 是部分的基因编码序列, 因此与 EST 对应的是全长的 cDNA, 即为特异性组织的一个基因, 深入研究这个全长 cDNA, 就能得知它的功能^[1,2]。目前, EST 测序分析技术广泛应用于基因识别、绘制基因表达图谱、寻找新基因等研究领域, 并取得了巨大的成效^[3]。

由于草鱼肠道是营养消化主体, 肠道功能基因就与草鱼食物和饲料的高效转化密切相关, 经最新 NCBI-dBEST 数据库检索, 目前草鱼的 EST 总共有

531 条, 其中 2003 年湖南农业大学提交的来源于肝、脾组织 514 条 (CK232812-CK233207; CF887792-CF887911; CX129925), 中科院武汉水生生物所提交的来源于腮、肝组织的 17 条 (CF751831-CF751844), 没有消化道组织的 EST。本课题组构建了国内第一个草鱼肠道 cDNA 文库, 并对 EST 测序的成功率和 EST 的生物信息学进行分析, 得到的相关科研数据对草鱼消化、营养、代谢等相关功能基因阐释使我们更客观、更精确地了解草鱼的营养消化机理和特性。本研究将进一步提高我国草鱼基础科研的水平, 并建立起草鱼基因组计划与功能饲料产业化之间的纽带。

1 材料和方法

1.1 材料及试剂 从四川成都通威水产科技有限公司采集的实验材料—健康雌性草鱼, 体长 570mm, 约 1.5 冬龄 (18 个月), 2.4kg。载体 pBluescript II SK (+) 购自 Stratagene 公司; DH10B 大肠杆菌超级感受

收稿日期: 2006-07-12; 修订日期: 2006-12-10

基金项目: 通威股份公司科技发展项目——草鱼基因组计划及产业化项目基金资助

作者简介: 张学俊 (1981—), 男, 汉, 江西九江人; 在读硕士研究生; 主要从事分子生物学和生物信息学方面的研究。E-mail: hotbird007@163.com

通讯作者: 徐恒 (1964—), 教授; E-mail: xuheng64@email.scu.edu.cn

态细胞购自 Promega 公司。DEPC 购自 Sigma 公司, TRIzol 试剂购自 Invitrogen 公司, PolyAtract mRNA Isolation Systems 购自 Promega 公司, Superscript II-RT 逆转录酶购自 Invitrogen 公司, DNA 聚合酶 I 和 RNase H 购自 Promega 公司, *EcoR* I 和 *Xho* I 购自 Stratagene 公司, QIAEXII Gel Extraction kit 胶回收试剂盒购自 QIAGEN 公司, 琼脂糖购自 OXIOD 公司, DNA Marker 购自 TaKaRa 公司, 实验中的其他试剂均为进口或国产分析纯度试剂。常规试剂按《分子克隆实验指南》^[4] 配制。

1.2 方法 取出新鲜的草鱼肠道组织, 液氮迅速冷冻后保存于 -80°C 冰箱中备用。采用 TRIzol (Invitrogen Corporation, USA) 一步法提取总 RNA, 并用甲醛变性凝胶电泳/EB 染色检测所提总 RNA 的质量。mRNA 分离用亲和素顺磁磁珠 (SA-PMPS) 法从总 RNA 中分离纯化富含 Poly(A) 的 mRNA, 操作按试剂盒的说明书进行。以 mRNA 为模板, 使用 Superscript II-RT 逆转录酶和 polyT 引物合成 cDNA 的第一条链, polyT 引物 5' 端引入 *Xho* I 酶切位点, 使用 DNA 聚合酶 I 和 RNase H 合成 cDNA 的第二条链; 通过 T_4 DNA 聚合酶 (Promega) 补平 cDNA 末端, 然后用 T_4 DNA 连接酶加 *EcoR* I adaptor, 最后用 T_4 磷酸化酶将双链 cDNA 末端磷酸化后用 *Xho* I 酶切双链 cDNA; 琼脂糖凝胶电泳后, 在紫外灯下切下大约 500—5000bp 的 cDNA 片段, 用 QIAEXII Gel Extraction kit (QIAGEN) 纯化 cDNA 片段; 利用 T_4 DNA 连接酶把加工好的双链 cDNA 连接到 pBluescript II SK (+) 载体的 *Xho* I 和 *EcoR* I 位点上; 连接产物纯化后, 用电激转化仪将重组子电转化到 DH10B, 电压为 2.1kV, 电击时间为 5ms, 取 $20\mu\text{L}$ 转化产物加 $160\mu\text{L}$ SOC, 涂板, 37°C 过夜培养, 次日查看转化结果。其余菌液加入等体积的 30% 甘油, -80°C 冻贮; 挑取阳性克隆于灭菌水, 加入 T3、T7 引物和 PCR 反应试剂, 做 PCR 检测, 根据电泳图粗略鉴定插入片段的大小及小片段率; 将合格的 cDNA 文库送检。

1.3 cDNA 克隆的大规模 EST 测序 大规模 EST 测序主要包括模板制备、测序反应、序列识别等流程。模板制备采用半自动化流水线方式, 主要步骤包括: 文库铺板培养, 大量随机挑取菌落 (cDNA 克隆), 细菌增殖, 质粒 DNA 提取, 电泳鉴定等。测序及序列初步分析工作主要在北京华大基因研究中心完成, 采用 cDNA 克隆 5' 端测序, 测序引物为 T3 (5'-ATTAACCTCACTAAAG-3'), 在 MegaBace 1000 测序仪上测序。

1.4 数据处理 使用 Cross-match 软件去掉 EST 中的低质量序列、载体序列、重复序列等, 使用 Phrap 序列拼接软件将测序得到的 EST 拼接成重叠群, 将拼接得到的重叠群和单拷贝 EST 与 GenBank 非冗余蛋白库比对并使用 GO (Gene Ontology) 分类的方法做功能注释。GO^[5] 分类是现在较常用的 EST 分类方法, 除此之外, 还进行 GC 含量、密码子使用频率等分析。

2 结果

2.1 草鱼肠道组织总 RNA 及 mRNA 的分离

取 $1\mu\text{L}$ 总 RNA 样品稀释 50 倍, 用分光光度计检测: $\text{OD}_{260} = 1.261$, $\text{OD}_{230} = 1.062$, $\text{OD}_{280} = 0.653$, $\text{OD}_{260}/\text{OD}_{280} = 1.930$, 样品浓度约为 $2.5\mu\text{g}/\mu\text{L}$ 。RNA 总量为 $750\mu\text{g}$ 。

提取的草鱼肠道组织总 RNA 在 1% 的甲醛变性凝胶电泳图上 (图 1) 能观察到很清晰的 28S 核糖体 RNA 与 18S 核糖体 RNA、以及 5S 核糖体三条带, 无明显脱尾, 亦看不见明显的基因组和蛋白质污染, 28S 和 18S 两条带的亮度比基本上呈 2:1 的关系, 由此说明此 RNA 无降解, 完整性比较好。

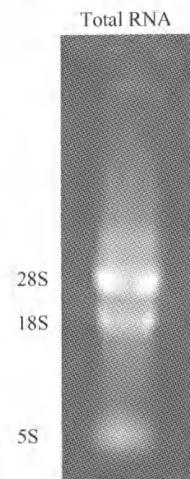


图 1 总 RNA 的甲醛变性凝胶电泳

Fig.1 Total RNA denatured formaldehyde gel analysis

mRNA 的分离亲和素顺磁磁珠 (SA-PMPS) 法, 操作按试剂盒的说明书进行, 取 $1\mu\text{L}$ mRNA 样品, 稀释 10 倍, 用分光光度计检测: $\text{OD}_{260} = 0.641$, $\text{OD}_{280} = 0.311$, $\text{OD}_{260}/\text{OD}_{280} = 2.061$, 样品浓度约为 $256.4\text{ng}/\mu\text{L}$ 。mRNA 总量约 $4.499\mu\text{g}$ 。电泳结果显示 mRNA 呈弥散分布 (图 2), 也无降解现象, mRNA 的 $A_{260\text{nm}}$ 与 $A_{280\text{nm}}$ 的比值为 2.061, 说明 mRNA 足够纯净, 完全符合逆转录要求。

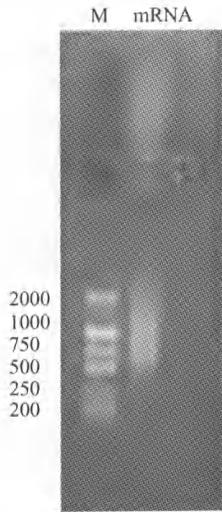


图2 mRNA 的甲醛变性凝胶电泳
Fig.2 mRNA denatured formaldehyde gelanalysis
M:DL2000 Marker;mRNA:纯化后的 mRNA
M:DNA Marker DL2000;mRNA:Isolation's mRNA

2.2 cDNA 文库的构建

2.2.1 双链 cDNA 的合成

双链 cDNA 产物电泳呈现从小到大均匀分布的拖带,大小在 0.5—5.0kb 范围内,主要集中于 500bp 以上(图 3),表明草鱼肠道组织所转录 mRNA 的成分多样复杂。在紫外灯下挖取 0.5—5.0kb 范围的胶块,用来回收目标 cDNA。用 Naro-drop 仪器检测双链 cDNA 胶回收情况:OD₂₆₀ = 0.335, OD₂₈₀ = 0.201, OD₂₃₀ = 5.371, OD₂₆₀/OD₂₈₀ = 1.667, OD₂₆₀/OD₂₃₀ = 0.062,样品浓度约为 16.8ng/μL,纯度符合要求,可以进行下一步实验。

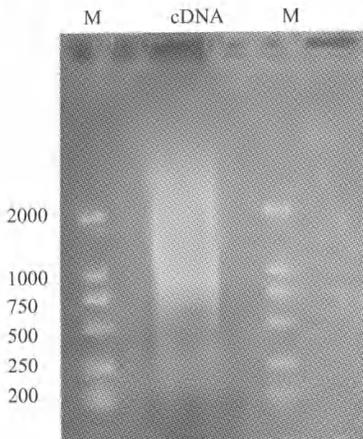


图3 草鱼肠道组织 cDNA 的 1.0% 琼脂糖凝胶电泳
Fig.3 cDNA agarose gel electrophoresis
M:DL2000 Marker;
cDNA:合成后的双链 cDNA
M:DNA Marker DL2000;
cDNA:Double-stranded cDNA

2.2.2 cDNA 与载体的重组及重组子的转化

将 100ng 的 cDNA 与质粒载体 PBK(E/X) Vector 16℃连接过夜,连接产物经 PCR 阳性检测后,再采用电转化法转化到感受态的 *E. coli* DH10B 中,经稀释铺平板检测重组率和库容。重组率在 95% 左右,原始库容平均为 2.3×10^5 ,基本达到文库的质量要求。

2.2.3 重组 cDNA 鉴定

挑取白斑转化菌,碱裂解法提取质粒 DNA 后,使用载体两端有匹配序列的引物 T3 和 T7 进行 PCR 扩增,电泳估计插入片段的大小。电泳检测表明,外源插入 cDNA 片段的大小平均在 500—2000bp 之间(图 4)。

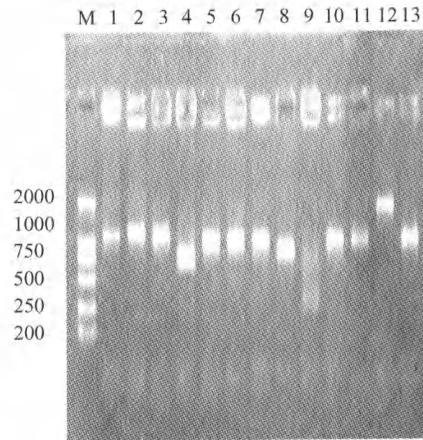


图4 草鱼 cDNA 文库插入片段的 PCR 鉴定
Fig.4 PCR analysis of 13 recombinant clones the cDNA library
M:DL2000 Marker;1—13:为随机选取的 13 个 cDNA 克隆
M:Marker DL2000;1—13:Randomly selected cDNA clones

2.3 EST 测序结果初步分析

测序得到 1571 个草鱼肠道组织 EST,去除低质量的序列和载体序列后,长度大于 100bp 的有 1411 个 EST,使用 phrap 默认参数拼接(表 1),拼出的重叠群有 188 个,单拷贝 EST 有 751 个,总计单基因簇 939 个。

表 1 草鱼 cDNA 文库的 ESTs 初步统计数据

Tab.1 Summary of ESTs from cDNA library of Grass Carp

描述项 Descriptive category	数目 Number
ESTs 总数 Total number of ESTs (去除载体后 ≥ 100bp)	1411
ESTs 平均长度 Average length (bp)	450
包含 contigs 的 ESTs 数目 ESTs covered by contigs	660

续表

描述项 Descriptive category	数目 Number
重叠群 Contigs	188
单拷贝 EST singletons	751
单基因簇 Unigene	939
已知功能的 ESTs Known genes	195
推测功能的 ESTs Hypothetical genes	223
非同源性的 ESTs No homology	521

使用 BlastX 软件和 GenBank 的非冗余蛋白库比对,并进一步在 EMBL 中利用 FASTA 进行比较。当得到的匹配序列与所研究的 EST 序列同源性达到 $E\text{-value} \leq 1.00E-10$; 或匹配序列在 70bp 以上,同源性 $\geq 70\%$ 时可判定为同源序列。939 个重叠群 (contigs) 和单拷贝 EST (singletons) 一共有 418 个比对上,占 44.5%,未能比对上的 EST 占 55.5%。对 418 个功能的重叠群和单拷贝 EST 做 GO 分类。418 个重叠群和单拷贝 EST 中有 195 个可归属于细胞组分 (Cellular component)、分子功能 (Molecular function) 和生物学过程 (Biological process) 三个大类,而其余 223 个重叠群和单拷贝 EST 则为未知功能的 EST。195 个 EST 的 GO 分类结果见图 5。绝大部分 EST 与营养代谢、生理学过程、细胞结构、酶活性及细胞过程有关,少数与转录调节活性、信号传导活性以及细胞分裂分化等功能有关。

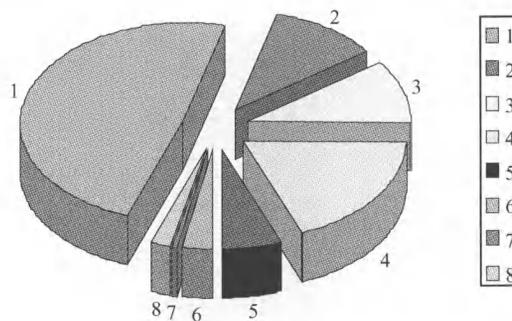


图 5 ESTs 依照功能分类百分比图示

Fig. 5 Percentage distribution of ESTs by functional classes

1.95 条同翻译,核糖体结构和生物合成功能相关序列 (48.72%); 2.21 条同蛋白质折叠和修饰功能相关序列 (10.77%); 3.20 条被预测为仅仅具有预测功能序列 (10.26%); 4.37 条序列同物质代谢功能相关序列 (18.97%); 5.11 条同细胞结构和运动功能相关序列 (5.64%); 6.6 条同细胞信号和传导功能相关 (3.08%); 7.1 条同细胞分裂功能相关序列 (0.51%); 8.4 条同转录功能相关序列 (2.05%)

1. 95 for Translation, ribosomal structure and biogenesis (48.72%);
2. 21 for Posttranslational modification, protein turnover, chaperones (10.77%);
3. 20 matched to genes of General function prediction only (10.26%);
4. 37 were for genes related to metabolism (18.97%);
5. 11 for cell structure/motility (5.64%);
6. 6 for cell signaling/communication (3.08%);
7. 1 for cell division (0.51%);
8. 4 for Transcription (2.05%)

2.4 营养消化功能相关的 ESTs 实例

根据比对的类型将营养消化功能相关的 ESTs 展示出来 (表 2), 方便后续研究工作的开展。相关的功能 ESTs 初步统计有 7 类, 37 种: 能量产生和转化类 12 种, 碳水化合物运输和代谢类 9 种, 无机离子运输和代谢类 5 种, 核苷运输和代谢类 4 种, 氨基酸运输和代谢类 4 种, 脂类运输和代谢类 2 种, 次级代谢产物合成、运输和异化作用类 1 种。

表 2 草鱼 cDNA 文库的营养消化相关的 ESTs 的统计数据

Tab. 2 Summary statistics of ESTs genes related to metabolism from cDNA library of Grass Carp

编号 Number	EST 数目 ^a EST No. ^a	大小 (bp) Size (bp)	同源性登录号 Homology ID	功能 Function	E 值 E-value
65	2	360	gb AAH67170.1	醇脱氢酶 5 Alcohol dehydrogenase 5	4.00E-56
98	2	624	gb AAH45894.1	三磷酸腺苷合酶 ATP synthase	7.00E-53
157	5	1132	gb AAL23929.1	细胞色素 b cytochrome b [C. idella]	0
178	9	1576	dbj BAC67392.1	细胞色素氧化酶亚基 1 Cytochrome oxidase subunit 1	0
181	11	776	dbj BAC67409.1	细胞色素氧化酶亚基 3 Cytochrome oxidase subunit 3	1.00E-118
187	13	689	dbj BAB40351.1	细胞色素氧化酶亚基 2 基因 Cytochrome oxidase subunit II gene [C. idella]	3.00E-95
386	1	634	gb AAO26196.1	细胞质苹果酸类脱氢酶 Cytosolic malate dehydrogenase	3.00E-70

续表

编号	EST 数目 ^a	大小(bp)	同源性登录号	功能	E 值
Number	EST No. ^a	Size(bp)	Homology ID	Function	E-value
446	1	277	dbj BAC67390.1	NADH 脱氢酶亚基 1 NADH dehydrogenase subunit 1	2.00E-20
495	1	570	ref NP_001017800.1	NifU 类似的 N 端结构域蛋白 NifU-like N-terminal domain containing protein	5.00E-53
567	1	623	ref NP_989234.1	细胞色素 C1 Cytochrome c1	9.00E-67
746	1	633	gb AAQ91233.1	二氢硫辛酰胺乙酰脱氢酶 Dihydropyrimidinase	1.00E-102
784	1	467	gb AAF21270.1	NADH 辅酶 Q10 氧化还原酶 NADH ubiquinone oxidoreductase[<i>C. idella</i>]	5.00E-60
34	1	404	gb AAH83506.1	3 磷酸甘油醛脱氢酶 Glyceraldehyde-3-phosphate dehydrogenase	1.00E-56
45	2	671	gb AAH65885.1	哺乳动物几丁质酶前体 Acidic mammalian chitinase precursor	1.00E-106
68	2	792	gb AAH83506.1	推测为 3 磷酸甘油醛脱氢酶 Putative glyceraldehydes -3- phosphate dehydrogenase	1.00E-142
78	2	638	emb CAG09753.1	无名蛋白产物 Unnamed protein product	3.00E-76
93	2	1018	gb AAQ94592.1	醛缩酶 B aldolase B	1.00E-156
489	1	646	ref XP_417164.1	透性酶 Permeases	1.00E-24
498	1	671	gb AAH55561.1	丙酮酸激酶 Pyruvate kinase	2.00E-90
733	1	493	gb AAH66691.1	岩藻糖苷酶 Fucosidase	3.00E-52
780	1	459	ref XP_342332.2	肝糖原支酶 Glycogen debranching enzyme	1.00E-42
13	1	697	ref NP_001018401.1	铁蛋白 H ferritin H	5.00E-84
120	3	1202	emb CAF97355.1	肠道钠盐/双羧酸盐协同运输物质 Intestinal sodium/dicarboxylate cotransporter	5.00E-81
175	8	1879	gb AAH63947.1	依赖钠盐的双羧酸盐转运物质 Sodium-dependent dicarboxylate transporter	4.00E-70
188	17	1002	gb AAH45278.1	铁蛋白 H Ferritin H	2.00E-90
750	1	654	ref NP_001002378.1	铁蛋白 M Ferritin M	8.00E-93
143	4	883	gb AAH47855.1	肌肉特异性 β_1 黏合素蛋白 2 Muscle-specific beta 1 integrin binding protein 2	1.00E-85
242	1	434	gb AAQ97739.1	腺苷酸琥珀酸裂解酶 Adenylosuccinate lyase	1.00E-63
630	1	566	gb AAH55613.1	二磷酸核苷激酶 A2 Nucleoside diphosphate kinase A2	8.00E-80
673	1	538	gb AAH53160.1	腺苷酸激酶 2 Adenylate kinase 2	2.00E-67
76	2	659	gb AAH85630.1	肌酸激酶 Creatine kinase	5.00E-69
575	1	551	ref XP_421393.1	左旋天冬酰胺 L-asparaginase	6.00E-53
606	1	649	gb AAH91994.1	麸胺酶 Glutamyl aminopeptidase	4.00E-60
739	1	673	gb AAC32754.1	氨肽酶 N aminopeptidase N	7.00E-58
97	2	729	emb CAD60853.1	脱氢酶/还原酶 SDR 家族 1 号 Dehydrogenase/reductase SDR family member 1	1.00E-102
494	1	303	emb CAB53008.1	Δ^9 脲酶 Delta-9-desaturase	3.00E-54
133	3	652	gb AAL16897.1	细胞色素 P450 3A Cytochrome P450 3A [<i>C. idella</i>]	8.00E-64

^a用单拷贝 EST(总是 1 个)和重叠群数目代表的 EST 个数,对于重叠群只有一个 EST 登录号列出 (*The number of EST represented by singletons (always 1) and contigs. For contigs, accession number of one EST is listed)

3 讨论

3.1 构建 cDNA 文库

构建 cDNA 文库,应当注意如下几个方面:保证获得数量足够的高质量起始 RNA。如本实验中

总 RNA 提取结果与分析中,OD₂₆₀与 OD₂₈₀比值稍大于 2.0,可推测出样品中还含有少量杂质,但从电泳结果中可知 RNA 的完整性还比较好,所以这并不影响进行下一步的实验。

测序文库的库容量^[6,7]、重组率及插入片段的

大小是鉴定 cDNA 文库质量的重要指标。根据公式 $N = \ln(1-P)/\ln(1-1/n)$ (N 为实际所需克隆数, P 为要求的概率(一般为 99%), n 为某种类型的稀有 mRNA 在总 mRNA 中所占的比例), 要求克隆到某低丰度 mRNA 的概率为 99% 时, 文库的克隆数不应低于 1.7×10^5 个^[8]。本文库的总克隆数为 2.3×10^5 个, 高于所需标准, 因此从本文库筛选出低丰度目的基因在理论上是可行的。文库的重组率为 95%, 说明文库是完整的、有效的, 并且质量较高。

本研究采用了代表性组织或器官的大规模 EST 测序分析策略, 一次测序基本上获得了特定组织的较完整的表达基因信息, 为了解基因表达概况, 分析组织和发育的表达特异性, 筛选特定功能基因提供了数据资源。文库构建时所使用的载体系统是质粒, 其具有操作简单, 可塑性强的特点^[9], 能避免以往噬菌体操作的许多繁琐步骤, 这对于草鱼特异组织的大规模 EST 测序分析是非常有利的。

3.2 几种功能 ESTs 的生物学研究进展

能量产生和转化类 由于肠道器官旺盛的物质和能量代谢作用, 肠道内必然含有如 cytochrome b、Cytochrome oxidase subunit 1, 2, 3 这样的高丰度的具有与能量产生和转化的蛋白, 也含有呼吸链的其他组分如 Alcohol dehydrogenase 5、ATP synthase、cytosolic malate dehydrogenase、NADH dehydrogenase subunit 1、NADH ubiquinone oxidoreductase 等等。呼吸链在传递氢或电子的过程中, 为肠道消化吸收作用源源不断地产生 ATP 能量。其中细胞色素氧化酶 (Cytochrome oxidase, CCO) 为含铜组分的酶, 参与能量代谢也涉足细胞或神经元膜电位的维持、细胞成份的合成、神经元内的快速轴浆运输等功能。Germon^[10] 研究发现铜元素缺乏可导致 CCO 活性下降, 氧化磷酸化受到抑制, ATP 生成减少, 从而影响许多物质的合成及生物电的产生。Boyer^[11] 阐明了 ATP 酶作用机制要点: 一是 ATP 合成所需要的能量原则上是用于促进酶上紧密结合的 ATP 的释放和无机磷、ADP 的结合; 二是在净 ATP 形成过程中, 酶上的各催化部位是高度协同地顺序起作用的。 γ 亚基在 F1 ATP 酶中的旋转运动使三个催化部位构象不对称是实现结合变化的基础。

碳水化合物运输和代谢类 甘油醛-3-磷酸脱氢酶 (Glyceraldehyde-3-phosphate dehydrogenase, GAPDH) 在糖酵解及光合作用中起重要的作用。另外该基因在功能区的序列高度保守。GAPDH 基因的进化速率较慢 (5PAMs 每 100 万年)^[12], 因此在研

究生物的分子系统发育中被广泛使用。醛缩酶 B (aldolase B) 目前研究集中在遗传性果糖不耐病 (Hereditary fructose intolerance, HFI)^[13]、肝功能紊乱^[14] 等领域, 由于它是寄生虫入侵宿主过程中速殖关键因子, 对寄生虫防治领域有相当研究^[15]。

无机离子运输和代谢类 铁蛋白 (Ferritin) 是一切生物细胞内主要的铁贮藏蛋白质。每个脱铁铁蛋白 (无铁的铁蛋白) 壳由 24 条多肽链组成, 多肽链有重亚基 (H-亚基) 和轻亚基 (L-亚基) 二种。H-亚基分子量 21kDa, 电泳迁移率偏酸性, 而 L-亚基分子量 19kDa。因为 H-铁蛋白比 L-铁蛋白更易结合铁及释放铁, 通常认为 H-铁蛋白在铁的迅速解毒和细胞内铁的运输中起主要作用。铁和铁体内平衡与许多疾病的发病机理有关, 这包括与铁的结合、运输、贮藏有关的疾病 (如原发性血色病) 以及动脉粥样硬化、帕金森病、阿尔茨海默病和小腿多动综合征等常见疾病。

脂类运输和代谢类 $\Delta 9$ -去饱和酶 ($\Delta 9$ -desaturase) 是在动植物不饱和脂酸 (软油酸、油酸、亚油酸、 α -亚麻酸、花生四烯酸等) 的合成代谢中的过程中的酶之一。在饲料工业上, 与共轭压油酸 (CLA) 联系在一起。Corl 等^[16] 通过奶牛真胃灌注实验表明, $\Delta 9$ -去饱和酶去饱和作用帮助 CLA 合成。Griinari 等^[17] 研究指出通过源自 trans-11 18:1 的 $\Delta 9$ -去饱和酶作用的内源合成是奶牛乳脂中 CLA 的主要来源。Glaser 等^[18] 研究指出饲喂 C18:1 反式脂肪酸的氢化脂肪的猪具有高 CLA 含量的背脂 (0.44%), 表明猪可能通过内源性 $\Delta 9$ -去饱和作用将反式脂肪酸转变为 CLA。

3.3 和大西洋鲑鱼 (*Salmo salar*) 肠道 cDNA 文库的比较

Hagen^[19] 等研究构建了大西洋鲑的腮和肠道的 cDNA 文库, 并通过 EST 测序拼接得到 2974 单基因簇, 并对其中 967 个已知功能的基因的 ESTs 做了 GO 分析。通过大西洋鲑 (肉食性) 和本实验草鱼 (草食性) 肠道 cDNA 文库的 EST 比较, 发现相似性个数极少, 只发现与 β -2 微球蛋白 (Beta-2 microglobulin) 基因有同源性。其他的 510 个属于目前研究比较多的模式生物斑马鱼, 剩下的 82 个分属于 20 个物种类型, 如: 人 (*Homo sapiens*) (2)、家鼠 (*Mus musculus*) (2)、欧洲野牛 (*Bos taurus*) (1)、热带爪蟾 (*Xenopus tropicalis*) (1)、黑猩猩 (*Pan troglodytes*) (1), 其他都是鱼类: 草鱼 (24)、鲤鱼 (*Cyprinus carpio*) (22)、黑青斑河豚 (*Tetraodon nigroviridis*) (9)、鲫鱼 (*Carassius auratus*) (5)、

斑点叉尾鲟 (*Ictalurus punctatus*) (4)、南亚野鲮 (*Labeo rohita*) (2)、虹鳟 (*Oncorhynchus mykiss*) (1)、珠星三块鱼 (*Tribolodon hakonensis*) (1)、大鳞裂尾鱼 (*Pogonichthys macrolepidotus*) (1)、间鲃 (*Barbus intermedius*) (1)、四眼鱼 (*Anableps anableps*) (1)、剑鱼 (*Xiphophorus maculatus*) (1)、三刺鱼 (*Gasterosteus aculeatus*) (1)、黑头软口鲮 (*Pimephales promelas*) (1)、西方鲃 (*Barbus bynni occidentalis*) (1)。原因分析:一是大西洋鲑和草鱼的生活环境以及食性都不同,这是造成 ESTs 相似性个数少的直接原因,另外测序的个数也对相似性个数起到间接的影响。二是模式生物斑马鱼全基因组测序已经完成,作为模式研究鱼类,统计出的数据占测序数据大部分在预期实验范围内。三是草鱼与同是鲤科鱼类和经济鱼类的鲤鱼、鲫鱼相似性个数分别是 22 和 5,这证实它们之间存在一定的同源性,同时也说明了文库潜在的经济研究价值。

参考文献:

- [1] 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
- [2] Wang X L, Li N, Zhao Z H, et al. Analysis of succeed percentage on sequencing ESTs and construction of porcine muscle cDNA library [J]. *Hereditas*, 2002, **24**(3): 263—266 [王秀利, 李宁, 赵志辉, 等. 香猪肌肉组织 cDNA 文库的构建及其 EST 测序成功率的分析. *遗传*, 2002, **24**(3): 263—266]
- [3] Pandey A, Lewitter F. Nucleotide sequence databases: a gold mine for biologists [J]. *Trends biochem Sci*, 1999, **24**(7): 276—280
- [4] Sambrook J, Fritsch E F, Maniatis T. *Molecular Cloning: A Laboratory Manual* (3rd edition) [M]. New York: Cold Spring Harbor Laboratory Press. 2002, 857—980
- [5] Ashburner M, Ball C A, Blake J A, et al. Gene Ontology: tool for the unification of biology [J]. *Nature Genetics*, 2000, **25**(1): 25—29
- [6] Yan H J, Huang X Q, Cheng Z Q. Advances of the studies on construction strategy and analysis of cDNA library [J]. *Journal of Yunnan Agricultural University*, 2002, **21**(1): 1—6 [晏慧君, 黄兴奇, 程在全. cDNA 文库构建策略及其分析研究进展. *云南农业大学学报*, 2002, **21**(1): 1—6]
- [7] Zhao Y H. *Molecular Biology Tutorial* [M]. Beijing: Academic Press. 2004 [赵亚华. 分子生物学教程. 北京: 科学出版社. 2004]
- [8] Jin X B, Wang T T, Zhu J Y. Construction and preliminary identification of the cDNA library for *Musca domestica* larvae [J]. *Chinese Journal of Zoonoses*, 2005, **21**(1): 52—55 [金小宝, 王婷婷, 朱家勇. 家蝇幼虫脂肪体 cDNA 文库的构建及初步鉴定. *中国人兽共患病杂志*, 2005, **21**(1): 52—55]
- [9] Chen L, Qiang B Q, Fang F D. *Genomics and Diseases of Human* [M]. Beijing: Academic Press. 2002 [陈竺, 强伯勤, 方福德. 基因组科学与人类疾病. 北京: 科学出版社. 2002]
- [10] Germon T J, Evans P D, Manara A R, et al. Detection of hypoxic cells with the 2-nitroimidazole, EF5, correlates with early redox changes in rat brain after perinatal hypoxia-ischemia [J]. *Neuroscience*, 1999, **89**(4): 1357—1366
- [11] Boyer P D. The ATP synthase-A splendid molecular machine [J]. *Annual Review of Biochemistry*, 1997, **66**: 717—749
- [12] Fothergill-Gilmore L A, Michels P A M. Evolution of glycolysis [J]. *Prog Biophys Mol Biol*, 1993, **59**: 105—235
- [13] Malay A D, Allen K N, Tolan D R. Structure of the thermolabile mutant Aldolase B, A149P: Molecular basis of hereditary fructose intolerance [J]. *Journal of Molecular Biology*, 2005, **347**(1): 135—144
- [14] Maniratanachote R, Shibata A, Kaneko S, et al. Detection of autoantibody to aldolase B in sera from patients with troglitazone-induced liver dysfunction [J]. *Toxicology*, 2005, **216**(1): 15—23
- [15] Yu N, He A, Li Z Y, et al. Characterization and analysis of *Toxoplasma gondii* aldolase gene and intron [J]. *Chinese Journal of Zoonoses*, 2005, **21**(7): 580—583 [余南, 何嵩, 李卓雅, 等. 刚地弓形虫醛缩酶基因及其内含子的分析. *中国人兽共患病杂志*, 2005, **21**(7): 580—583]
- [16] Dhiman T R, Satter L D, Pariza M W, et al. Examination of the importance of $\Delta 9$ -desaturase and endogenous synthesis of CLA in lactating dairy cows [J]. *J Anim. Sci*, 1999, **77**(Suppl.1): 118
- [17] Griinari J M. Conjugated linoleic acid is synthesized endogenously in lactating dairy cows by $\Delta 9$ -desaturase [J]. *Nutr*, 2000, **130**(9): 2285—2291
- [18] Glaser K R, Scheeder M R L. Dietary C18:1 trans fatty acids increase conjugated linoleic acid in adipose tissue of pigs [J]. *European Journal of Lipid Science and Technology*, 2000, **102**(11): 684—686
- [19] Hagen-Larsen H, Laerdahl JK, Panitz F, et al. An EST-based approach for identifying genes expressed in the intestine and gills of pre-smolt Atlantic salmon (*Salmo salar*) [J]. *BMC Genomics*, 2005, **6**: 171

CONSTRUCTION OF INTESTINAL cDNA LIBRARY AND ANALYSIS OF SOME EXPRESSED SEQUENCE TAGS SEQUENCING OF *CTENOPHARYNGODON IDELLUS*

ZHANG Xue-Jun^{1,2}, QU Gang¹, ZHU Wen-Li¹, ZHANG Li¹, WANG Juan¹, WU Jiang², LIU Han-Yuan²,
CHEN Fang¹ and XU Heng^{1,2}

(1. Key Laboratory of Bio-resources and Eco-environment (Ministry of Education), College of Life Science, Sichuan University, Chengdu 610064; 2. Tongwei Group Co., Ltd., Chengdu 610081)

Abstract: The technology of expressed sequence tag (EST) was broadly applied in analysis of gene function and expression pattern. Based on the construction technology of non-normalized cDNA library, the cDNA libraries of grass carp's intestinal cDNA Library was constructed. This library reaches 2.3×10^5 in capacity; the percentage of recombination is as high as 95%. PCR results showed that the average size of inserts was larger than 1000 bp. And a total of 1571 successful sequencing reactions yielded 1411 which expressed sequence tags (ESTs) longer than 100 bp. In addition, 939 non-repetitive sequences were formed including 188 contigs and 751 singletons from these ESTs after initial assembly. Basic Local Alignment Search Tool (BLAST) analysis showed that nearly 44.5% of the sequences were assigned putative functions because they showed similarity to known genes (E-value < 1.00E-10), mostly from other species, in one or more of the databases used, others suggesting a potential for further novel gene discovery. The study provides a significant number of ESTs for gene discovery and candidate genes for studying digestive system in grass carps and other carps. Furthermore, the large-scale ESTs sequencing provided a valuable resource for the functional genome research and the whole genome sequencing project of *Ctenopharyngodon idellus*.

Key words: *Ctenopharyngodon idellus*; Intestine; cDNA library; Expressed Sequence Tag; Sequencing