

猪肝组织表达序列标签(ESTs)的初步分析

李 宁¹, 赵志辉¹, 刘兆良¹, 赵兴波², 连正兴², 吴常信²

(¹ 中国农业大学农业生物技术国家实验室 100094; ² 中国农业大学动物科学技术学院 100094)

摘要: 构建了猪肝组织 cDNA 文库, 并对文库中随机挑选的 438 个克隆的表达序列标签(ESTs)进行了研究。结果表明, 在所研究的 438 个 ESTs 中, 186 个在猪种中已有匹配序列, 37 个在人类及其它物种中可以找到同源序列, 215 个为未知功能基因。试验还测定了文库中 45 个 cDNA 克隆的全长插入序列, 结果表明, 19 个为已知功能基因, 11 个没有发现开放阅读框, 其它 15 个具有开放阅读框, 但基因的功能未知。这些结果初步代表了肝脏组织的功能基因表达谱。

关键词: 猪; 肝脏; cDNA 文库; 表达序列标签

Analysis of Expressed Sequence Tags from Porcine Liver Organ

LI Ning¹, ZHAO Zhi-hui¹, LIU Zhao-liang¹, ZHAO Xing-bo², LIAN Zheng-xing², WU Chang-xin²

(¹ National Laboratory for Agrobiotechnology, China Agricultural University, Beijing 100094;

² College of Animal Science and Technology, China Agricultural University, Beijing 100094)

Abstract: In order to study the expression of functional gene and its effect on metabolic control and other physiological functions in liver, 438 expressed sequence tags (ESTs) were determined, which were from a cDNA library of porcine liver tissue. The results showed that the nucleotide sequences of 186 ESTs have already presented in GenBank database, and 37 ESTs could be found the homology with human and other species, while the others were not identified. 45 full-length insertion of the clones randomly isolated from cDNA library were also completely sequenced with different size and the results showed that 19 of them were function-known genes, 11 had no open reading frame (ORF) and 15 had ORF but the function were not elucidated yet.

Key words: Swine; Liver; cDNA library; Expressed sequence tags

表达序列标签(Expressed sequence tags, ESTs)是 1991 年由美国科学家 Adams 提出的一种快速简便揭示基因组容量的方法^[1]。由于它来源于特异组织的 cDNA, 因此可以揭示不同组织不同发育阶段的基因表达情况, 同时作为遗传标记, 具有其它标记方法不可比拟的优点^[2]。

近年来, ESTs 研究进展相当迅速, 并且在人类基因组草图的构建中起到了关键的作用。随着人类基因组计划的发展和分子生物学研究的不断深入, 人们越来越认识到 ESTs 在新基因的发现、基因作图和基因组序列中编码区域的确定等方面所起到的重要作用^[3]。随着人类基因组图谱的不断完善及

物理图谱的构建成功, 有关基因组研究的重点也将随之发生转移, 即由物理图谱的构建转向功能基因组和后基因组研究, ESTs 是功能基因组学研究的重要部分之一, 并且有关全长 cDNA 的研究也越来越受到重视并逐渐成为发展方向之一。

尽管人类 ESTs 研究已经取得了巨大成绩, 但有关猪的报道仅有 18 000 条, 其中有些正在申请专利, 资源组织主要为脑及骨骼肌^[4]。肝脏是整个机体物质代谢和生理生化反应的中心, 了解肝脏 mRNA 的序列及其功能不仅是猪功能基因组研究的重要内容, 同时也将对全面提高猪的生产性能奠定分子生物学基础。本文在建立 cDNA 文库的基础上,

收稿日期: 2001-07-11

基金项目: 国家重点基础研究发展规划资助项目(G20000161)

作者简介: 李 宁(1962-), 男, 江西南昌人, 教授, 博士, 主要从事动物基因组分析和转基因的研究工作。Tel: 010-62893323; Fax: 010-62893055; E-mail: ninglbau@public3.bta.net.cn

随机分离单克隆,证实为重组子后再测定 DNA 序列,共获得 438 个 ESTs 序列,并进一步测定了其中 45 个克隆的全长插入序列。本研究为我国猪基因组计划的开展和分析肝脏内表达基因对代谢调节及其它生理功能的影响提供了初步依据。

1 材料与方 法

1.1 材 料

1.1.1 试验动物 试验猪为杜洛克×(长白猪×大白猪)的三元杂交后代商品肥育猪,5 月龄屠宰,体重为 85 kg,取其肝脏置液氮中备用。

1.1.2 主要试剂 TRIzol Reagent (GIBCO BRL 公司),DEPC (SIGMA 公司),PolyA Ttract mRNA Isolation System (Promega 公司),Universal Ribosome cDNA Synthesis System (Promega 公司),DNA Packing Kit (Boehringer Mannheim 公司),λUni-ZAP™XR 载体(Stratagene 公司)。

1.2 方 法

1.2.1 cDNA 文库的构建 从肝脏中提取总 RNA,分离 mRNA 并合成 cDNA,经连接包装后构建肝脏 cDNA 文库,所用载体为 λZAP。利用辅助噬菌体转化质粒后,分离单克隆并进行 DNA 提取。具体步骤参照分子克隆及所用试剂盒的有关说明。

1.2.2 序列测定 利用 λZAP 多克隆位点侧翼序列作为测序引物,所用测序试剂盒为 BigDye Terminator(PE 公司),测序仪为 ABI PRISM 377(PE 公司)。

1.2.3 序列分析 所测 ESTs 序列同源性分析可通过 NCBI GenBank 中 BLASTn 进行,并进一步在 EMBL 中利用 FASTA 进行比较^[5]。当得到的匹配序列与所研究的 EST 序列期望值 $<1e-10$ 或同源性 $>70\%$ 时,可判为同源序列。在分析克隆全长插入序列时,通过网上在线分析可发现有完整开放读框(open reading frame)及部分 5' 和 3' 端非编码序列,定义为全长 cDNA^[6]。

2 结果与分析

2.1 cDNA 文库的构建

制备总 RNA 后,分离 mRNA 并合成 cDNA。经连接包装后,构建肝脏 cDNA 文库。结果表明,该文库的容量为 2.4×10^6 pfu·ml⁻¹,通过对随机挑选 105 个克隆的限制性酶切分析,重组与非重组比为 100:1,平均插入片段大小为 1.5 kb。

2.2 序列测定

利用辅助噬菌体转化质粒后,分离单克隆并提取 DNA。以 λZAP 多克隆位点侧翼序列为测序引物进行序列测定。经分析后共得到 438 个较好的 ESTs,所得到的 ESTs 长度大部分在 214 ~ 410 bp 之间,47% 以上的序列均进行了两次测序。所有测序的 ESTs 序列均已送入 GenBank 中的 ESTs 数据库。

2.3 序列分析

2.3.1 ESTs 分析 经 BLASTn 和 FASTA3.1 与 GenBank 和 EMBL 中现存的哺乳类动物 ESTs 进行同源分析后表明,所得到的 438 个 ESTs 序列中有 186 个在猪种中可以找到同源序列,在人类、牛及兔等物种中已知功能同源序列为 37 个,在 GenBank 和 EMBL 中未找到匹配同源序列的 ESTs 为 215 个(表 1)。在测定的 438 个 ESTs 中,白蛋白基因 ESTs 最多为 37 个,占测定总数的 8.5%,线粒体基因为 21 个,占测定总数的 4.8%。本研究还发现了前人已证明在肝脏特异表达的阿朴脂蛋白 B、乙醇脱氢酶等基因。因此,目前结果初步反映了肝脏组织内的基因表达谱(expression profile)。

2.3.2 cDNA 克隆全长插入序列的测定及结果分析 除了对 438 个 ESTs 进行了序列测定外,还依据不同插入长度分组后,随机挑取克隆进行了全长插入序列测定。全长序列测定均进行了 2 次。所测片段长度分布及数量如表 2 所示。结果表明,本研究构建的文库,发现全长 cDNA 的比率较高,特别是插入片段愈大,获得全长 cDNA 的比率愈高(部分全长 cDNA 已经向 GenBank 数据库输送)。由于是计算机分析结果,是否具有生物学意义的全长 cDNA 还需要试验证明。

3 讨论与结论

单向 DNA 测序(single-run DNA sequencing)已经证实是获得 cDNA 克隆最初资料的有效方法。大量的资料表明,150 ~ 400 bp 长度的核苷酸片段包含了 cDNA 序列鉴定及其染色体定位所需要的足够信息^[7,8]。本文所研究的 438 个 ESTs 中,长度均在于 214 ~ 410 bp 之间,许多基因在现有的基因库中没有发现同源性高的匹配序列,说明在猪的基因组研究中,可能尚有大量新的或未确定的基因有待进一步研究;或者在猪种肝脏组织表达的基因与其它物种表达的基因同源性较低。有关这些基因的全序列特征,表达特征,功能确定,染色体定位,组织分布及其免疫特征也有待进一步了解。

表 1 438 个 ESTs 分析结果

Table 1 The analyzing results of 438 ESTs

猪中匹配序列 ESTs homology with the pig gene sequences	数量 Number	猪中匹配序列 ESTs homologous with the pig gene sequences	数量 Number
白蛋白 Albumin	37	视网膜结合蛋白 Retinal-binding protein	1
纤维蛋白原 α 、 β 、 γ Fibrinogen alpha, beta, gamma	17	补体 C5 Complement C5	1
线粒体基因细胞色素 b Mitochondrial gene, cytochrome b	21	β 还原酶 Beta reductase	1
$\alpha 1$ 抗胰凝乳蛋白酶 Alpha 1 antichymotrypsin	11	谷氨酸合成酶 Glutamine synthetase	1
亲血色球蛋白 Haptoglobin alpha	11	苯丙氨酸羟化酶 Phenylalanine hydroxylase	1
细胞色素 P-450 Cytochrome P-450	10	G β 蛋白 G beta protein	1
阿朴脂蛋白 B.E Apolipoprotein B, E	9	醛缩酶 Aldolase	1
生长粘附因子 Vitronectin	7	羧酸酯酶 Carboxylesterase	1
转铁蛋白 Transferrin	7	酰基辅酶 A 脱氢酶 Acyl-CoA dehydrogenase	1
胎球蛋白 Fetuin	5	视癌结合蛋白 Retinoblastoma binding protein	1
核糖体 RNA Ribosomal RNA	5	蛋白质 C Protein C	1
转运 RNA Transfer RNA	4	核酸结合蛋白 Nucleic acid binding protein	1
过氧化氢酶 Catalase	4	粘连蛋白 Fibronectin	1
血纤维蛋白溶酶原 Plasminogen	4	巨球蛋白 Macroglobulin	1
蛋白质抑制子 Proteinase inhibitor	3	细胞溶解抑制子 Cytolysis inhibitor	1
生长激素受体 Growth hormone receptor	2	锌糖蛋白 Zinc glycoprotein	1
类胰岛素生长因子-I IGF-I	3	乙醇脱氢酶 Alcohol dehydrogenase	1
柠檬酸合成酶 Citrate synthetase	2	粘液素 Mucin	1
淀粉 P 成分 Amyloid P component	2	α 抗凝乳蛋白酶 Alpha antichymotrypsin	1
兰诺丁受体 Rynodine receptor	2	坞蛋白 Docking protein	1
人、牛、兔中匹配序列 ESTs homologous with the gene sequences of human, cattle and rabbit but not pig	37	无匹配序列 ESTs ESTs without any homologous genes determined	215

表 2 文库中部分 cDNA 克隆全长插入序列的测定及分析结果

Table 2 The analyzing results of the full length insertion of some cDNA clones

全长插入片段大小 The size of the full insertion in the vector (bp)	所测数量/已知基因 Numbers sequenced / numbers of the known genes	非编码 mRNA 数量(无开放阅读框) Numbers of the mRNA without ORF
300 ~ 500	8/4	4
500 ~ 1 000	18/8	4
1 000 ~ 1 500	11/5	2
1 500 ~ 2 000	5/1	1
2 000 ~	3/1	0

在所测定的 ESTs 中,包含了线粒体基因、tRNA 基因和 rRNA 基因,以及与糖代谢、脂肪代谢和蛋白质代谢有关的基因,说明在肝脏内与物质代谢和能量代谢有关的基因进行着大量的表达,也说明肝脏作为代谢中心而起着重要的作用。试验结果不但给出了所测 ESTs 的序列情况,而且还表明了肝脏组织内基因的初步表达谱,该结果将为分析基因的表达情况及各个基因在特定发育时期作用的重要程度提供依据。测定全长 cDNA 序列是 ESTs 研究的深入和发展方向,本试验中较大比例的 ESTs 未能发现完整的开放阅读框,特别是那些小片段的 ESTs,表明在文库构建中插入 cDNA 的长度或文库质量(如 mRNA 部分被降解),对于发现新的功能基因至关重要。

ESTs 的研究在人类基因组草图的顺利完成过程中起到了重要作用^[9]。随着猪基因组计划的实施,不同发育时期不同组织的 ESTs 研究将会得到很大的发展。这些 ESTs 的序列测定及其定性研究将会极大地丰富猪的 ESTs 库,同时也将有助于理解肝脏对机体物质代谢的作用过程,并将为猪基因组计划的顺利实施和草图的完成以及猪功能基因组的研究起到积极的推动作用。

References

- [1] Adams M D, Kelley J M, Gocayne J D, Dubnick M, Polymeropoulos M H, Xiao H, Merric C R, Wu A, Olde B, Moreno R F. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 1991, 252: 1 651 - 1 656.
- [2] Adams M D, Kerlavage A R, Fleischmann R D, Fuldner R A,

- Bult C J, Lee N H, Kirkness E F, Weinstock K G, Gocayne J D, White O. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature*, 1995, 377(6547 Suppl.): 3 - 174.
- [3] Boguske M S, Schuler G D. ESTablishing a human transcript map. *Nature Genetics*, 1995, 10:369 - 371.
- [4] Davoli R, Zamboni P, Bigi D, Fontanesi L, Kusso V. Analysis of expressed sequence tags of porcine skeletal muscle. *Gene*, 1999, 233:181 - 188.
- [5] Smith E, Shi L, Drummond P, Rodriguez L, Hamilton R, Powell E, Nahashon S, Ramlal S, Smith G, Foster J. Development and characterization of expressed sequence tags for the turkey (*Meleagris gallopavo*) genome and comparative sequence analysis with birds. *Animal Genetics*, 2000, 31:62 - 67.
- [6] The statistics of sequence (<http://www.ncbi.nlm.nih.gov/BLAST/tutorial/>).
- [7] Touchman J W, Bouffard G G, Weintraub L A, Idol J R, Wang L, Robbins C M, Nussbaum J C, Lovett M, Green E D. 2006 Expressed sequence tags derived from human chromosome 7-enriched cDNA libraries. *Genome Research*, 1997, 7:281 - 297.
- [8] Durkin A S, Nierman W C, Zoghbi H, Jones C, Kozak C A, Maglott D R. Chromosome assignment of human brain expressed sequence tags (ESTs) by analyzing fluorescently labeled PCR products from hybrid cell panels. *Cytogenetics Cell Genetics*, 1994, 65:86 - 91.
- [9] Schuler G D, Boguski M S, Stewart E A, Stein L D, Gyapay G, Rice K, White R E, Rodriguez-Tome P, Aggarwal A, Bajorek E. A gene map of the human genome. *Science*, 1996, 274:540 - 546.