中国农业科学 2003,36(10):1209 – 1212
Scientia Agricultura Sinica

猪脂肪组织表达序列标签(ESTs)的研究

李国华[1],刘兆良[2],许红韬[2],赵志辉[2],张 沅[1],李 宁[2]

([1] 中国农业大学动物科技学院;[2] 中国农业大学农业生物技术国家重点实验室,北京 100094)

摘要:对猪脂肪组织 cDNA 文库的 2 104 个克隆进行 EST 研究,其中 563 个 ESTs 在猪种中已有匹配序列,682 个在人类及其它物种中可以找到同源序列,376 个 ESTs 为未知新基因,有 298 个因序列质量差无法做分析,185 个在 EST 库中有同源序列。所分析的 376 个新 ESTs 中,其中有 159 个具有 ORF 结构,43 个具有 Poly(A)和 CpG 特征,这些结果初步显示了猪脂肪组织的功能基因表达特点。

关键词:猪;脂肪;cDNA 文库;表达序列标签

Analysis of Expressed Sequence Tags (ESTs) in White Fat Tissue of Pig

LI Guo-hua[1], LIU Zhao-liang[2], XU Hong-tao[2], ZHAO Zhi-hui[2], ZHANG Yuan[1], LI Ning[2]

([1] *College of Animal Science and Technology*; [2] *National Laboratory for Agrobiotechnology, China Agricultural University, Beijing 100094*)

Abstract: 2 104 expressed sequence tags (ESTs) from a cDNA library of porcine adipocyte were studied. The results showed that the nucleotide sequences of 563 ESTs have already been presented in GenBank database, 682 ESTs could be found homologous with those of human and other species, while 376 ESTs were not identified. 298 ESTs were eliminated for the poor quality of sequence, and 185 ESTs have their homologues in the EST database. Of the 376 ESTs, 159 have ORF structure, and 43 contain Poly(A) or/and CpG elements, and 174 had no above structures or elements. All these results preliminarily showed the expression style of functional genes in porcine adipocyte.

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

1989 年美国启动人类基因组计划的初始就面临两派不同的意见:一派主张直接测定人基因组的全部序列,另一派主张优先测定具有编码信息的基因序列即 cDNA 序列——这就是 cDNA 计划。后者认为 cDNA 仅占整个基因组的 3% ~ 5%,测序速度快、费用低、效益高。两派互不相让,结果基因组和 cDNA 测序同时开展,二者相互促进[1]。广义 cDNA 计划包括基因表达序列的鉴定、全长 cDNA 的克隆和测序、基因的染色体定位、基因的功能分析等;狭义的 cDNA 计划是指大规模地测定从 cDNA 文库挑出克隆的 5′和 3′端序列,这就是 EST 计划[2]。该计划由美国 NIH 的科学家 Venter 于 1989 年提出,并首先在人类基因组研究中应用。由于自动测序和生产策略的改进,Venter 将 EST 规模升级,使得对 cDNA 测序的投入产生了高信息的回报。NIH 对早期约 1 200个 EST 申请了专利,这引起了科学家们强烈的争议。对专利持批评态度的人士认为:不完整的 EST cDNA 绝大多数并没有确定其功能和用途,NIH 的申请终被搁浅[2]。受人类基因组科学的资助,美国基因组研究所的成立进一步扩大了 EST 测序的规模;由社团发起的 Incyte,创建了第二个类似的 EST 数据库。人类基因组科学与 Incyte 公司,都是在工业规模上建立起 EST 计划的,只有在签署许可合同的情况下才能够使用 EST 数据。1994 年,Merck

收稿日期:2002-06-04
基金项目:国家重大基础研究规划资助项目(G20000161)
作者简介 李国华(1971-),女,山西人,博士,主要从事动物遗传育种研究。李宁为通讯作者,Tel:010-62893323;Fax:010-62893904;E-Mail:ninglbau@publics.bta.net.cn

& Co.和华盛顿大学在 GenBank 上开创了一个向公众开放的 dbEST 数据库,1995 年向 EST 数据库输入了第一批数据,随后 dbEST 的数据以指数级增长[2]。

大部分公开的 EST 序列都来自于 Merck-WashU 的 EST 计划,这项计划的目的在于鉴定 cDNA,并对基因组的基本功能元件进行测序和鉴定。一旦 Merck-WashU EST 序列产生,就被收录到 NCBI 的 dbEST 数据库中,Merck-WashU 计划包括了许多生物的 dbEST 序列[3]。

EST(expressed sequence tags,EST)是长约 300 ~ 400 个核苷酸,来自随机选取的 cDNA 克隆的末端序列,简单地说,一个 EST 就是对应于某一种 mRNA 的一个 cDNA 克隆的一端序列。ESTs 通常是从已有的 cDNA 文库中随机取出几百到几千个克隆一次测序产生。一般长于 150 bp 的 EST 在同源查找和基因作图中的作用较大[4]。EST 方法发展几年来,使分子遗传学家识别和克隆新基因的策略发生了革命性的变化。现在 NCBI 的 EST 数据库(dbEST)中已录入的来自不同物种的不同组织的 EST 有 8 336 111 条(6 月 29 日,2001),其中,人和小鼠的 EST 占绝大部分。由于计算机和网络的普及,公开的 EST 数据库越来越多,内容也越来越全面,这就大大推进了基因组研究的进程。目前比较常用的 EST 或包括 EST 的 DNA 数据库有 dbEST (http://www.ncbi.nlm.nih.gov/dbEST),GenBank (http://www.ncbi.nlm.nih.Gov/web/GenBank), EMBL (http://www.ebi.ac.uk/databases), DDBJ (http://www.ddbj.nig.ac.jp), TDB (http://www.tigr.org/tdb/tdb.html (TIGR Database), ATCC (http://www.atcc.org/catalogs/recomb.html)。通过构建猪脂肪组织 cDNA 文库,进而对其大部分表达序列的研究,就可以了解脂肪组织中影响脂肪代谢相关的重要基因及表达机制,为进一步利用这些基因打下基础。我国是个养猪大国,提高猪种的瘦肉率,降低脂肪含量有着重要的经济意义。

1 材料与方法

1.1 材料

1.1.1 试验动物 长白猪,采自中国农业大学试验站猪场。

1.1.2 主要试剂 总 RNA 提取试剂盒 ISOGEN 来自日本,mRNA 分离纯化试剂盒购自 Pharmacia 公司,cDNA 合成试剂盒和包装蛋白来自 Stratagene 公司。

1.2 方法

1.2.1 cDNA 文库的构建 从猪脂肪组织中提取总 RNA,分离 mRNA 并合成 cDNA,经连接包装后构建脂肪 cDNA 文库,所用载体为 λUni-ZAP™,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,测序试剂盒为 Dye Terminator,测序仪为 MEGAbase 1000 和 377 DNA Sequencer。

1.2.3 序列分析 NCBI GenBank 中 BLASTn 对所测 ESTs 序列进行同源性分析,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 序列同源性达到期望值 < 1e-10 或匹配序列在 70 bp 以上,同源性 > 70% 时可判定为同源序列。

2 结果与分析

2.1 cDNA 文库的构建

制备总 RNA 后,立即进行 mRNA 的分离纯化,然后反转录生成 cDNA,经连接包装后,构建脂肪 cDNA 文库,结果表明,该文库的容量为 1.2×10^7 pfu·ml^{-1},重组与非重组克隆比例为 > 120∶1,平均插入片段大小为 1.4 kb。

2.2 序列测定

利用辅助噬菌体转化质粒后,分离单克隆后碱裂解法提取 DNA,然后以 λZAP 多克隆位点侧翼序列为测序引物,以 377 测序仪进行平板电泳测序和 MegaBASE1000 高压毛细管电泳测序。共分析了猪脂肪 cDNA 文库中 2 104 个克隆的序列,所得到的 ESTs 大部分在 400 ~ 600 bp 之间。

2.3 序列分析

2.3.1 ESTs 分析 经 BLASTn 与 GenBank 的 DNA 数据库中的序列进行同源性比较,结果表明,2 104 个 ESTs 序列中有 682 个在人、鼠等物种中存在同源序列,占总数的 32.4%;有 563 个 ESTs 在猪种中可以找到同源序列,占 26.76%;有 185 个是已报道的 ESTs 序列,占 8.79%;而在公众网上未找到匹配同源序列的有 376 个,占 17.87%;其余有 298 个因序列质量太差无法采用。在所测定的 2 000 多个 ESTs 中,线粒体基因 ESTs 最多为 253 个,占总数的 12%,其次是 MHC I 类 SLA 基因的序列,有 144 个 ESTs,占 6.84%(表)。在分析过程还发现脂肪特异表达基因,如脂肪高丰度基因转录物 1、脂肪特异 2、脂蛋白脂酶、推断早期前脂肪蛋白等。

表　2 104 个 ESTs 分析结果

Table　The analyzing results of 2 104 ESTs

猪中匹配序列 Matches in pig genomic DNA	数量 Numbers	猪中匹配序列 Matches in pig genomic DNA	数量 Numbers
线粒体基因组　Mitochondrion，complete genome	253	微卫星 S0375　Microsatellite S0375	1
MHC I 族 SLA 基因　MHC class I SLA genes	144	膜辅助因子蛋白　Membrane cofactor protein	1
脂肪酸结合蛋白　Fatty acid binding protein	23	谷胱甘肽 S 转移酶　Glutathione S-transferase	1
单倍型 E1 细胞色素 b　Haplotype E1 cytochrome b(cyt b)	17	克隆 SW452 微卫星　Clone SW452 microsatellite	1
铁蛋白重链　Ferritin heavy-chain	9	抗白细胞蛋白酶　Antileukoproteinase	1
Endozepine	9	谷胱甘肽过氧化物酶　Glutathione peroxidase	1
脂蛋白脱辅基 E　Apolipoprotein E	8	脂蛋白脂酶　Lipoprotein lipase	1
钙蛋白酶 I 轻亚基　Calpain I light subunit	7	微克分子钙激活神经蛋白酶 1 的同工酶 A	1
		Micromolar calcium-activated neutral protease 1 isoform A	
HSL	5	Endoglin	1
前胶原蛋白 III 型　Type III pro-collagen	4	延伸因子 1　Elongation factor 1，delta	1
乳酸脱氢酶-B　Lactate dehydrogenase-B	4	高迁移率型 1　High mobility group 1	1
类 G-beta 蛋白　G-beta like protein	4	IGFBP7	1
甘油脱氢-3-磷酸脱氢酶	4	染色体 21 长臂区　Chromosome 21q section 63/105	1
Glyceraldehyde-3-phosphate dehydrogenase			
过氧化物酶　Catalase	3	转化生长因子-beta1　Transforming growth factor-beta 1	1
GTP-结合蛋白，α 激活亚基	3	核糖体蛋白 S12　Ribosomal protein S12	1
GTP-binding protein，alpha stimulating subunit			
淀粉前体蛋白　Amyloid precursor protein	3	线粒体 ATP 酶 6　(Mitochondrial ATPase 6)	1
血浆凝胶蛋白　Plasma gelsolin	3	WO 0036143 序列 72　Sequence 72 from patent WO 0036143	1
β2-微球蛋白　Beta 2-microglobulin	3	电压依赖型阴离子通道 1 2	1
		Voltage-dependent anion channel 2	
装饰蛋白　Decorin	3	未知肝蛋白　Unidentified hepatic protein	1
类装配 DNA 结合蛋白　Rig-analog DNA binding protein	2	柠檬酸合酶　Citrate synthase	1
Cofilin	2	6-磷酸葡糖酸脱氢酶　6-phosphogluconate dehydrogenase	1
随遇蛋白/核糖体融合蛋白	2	90 kD 热激蛋白　90 kD heat shock protein	1
Ubiquitin/ribosomal fusion protein			
肌静止蛋白　Myostatin (Mstn)	2	磷酸葡糖酸酶 1　Clone g mus 4 g phosphoglucomutase 1	1
类随遇蛋白/S30 核糖体融合蛋白	2	IGFBP5	1
Ubiquitin-like/S30 ribosomal fusion protein			
线粒体粗糙氨基酸转移酶	2	S100C	1
Mitochondrial asperate amino-transferase			
铁蛋白 L 亚基　Ferritin L subunit	2	推断早期前脂肪蛋白　Putative preproadipsin	1
依赖 NADP 的细胞质苹果酸酶		热激蛋白 70　Hsp70	1
Cytosolic malic enzyme NADP-dependent			
HBp 15/L22	1	分泌型叶酸结合蛋白　Secreted folatebinding protein	1
PRO2640	1	p-选择性前体　p-selection precursor	1
蛋白酶体　Proteasome	1	硝酸氧化物合酶　Nitric oxide synthase(NOS)	1
亚精氨/精氨乙基转移酶 (SAT)	1	甘油-3-磷酸脱氢酶	1
Spermidine/spermine N1-acetyltransferase		Glycerol-3-phosphate dehydrogenase	
非肌肉肌蛋白轻链　Non-muscle myosin light chain	1		
人、鼠等物种匹配序列	682	无匹配序列　No matches in public data	376
Matches in human being，mouse and other species			
序列质量差的序列　Poor sequences	298	EST 库中有同源的序列　Matches in EST data base	185

2.3.2　ORF 的分析　在 GenBank 中没有找到匹配的 376 条新序列，利用开放阅读框查找软件(ORF Finder)进行分析，结果是其中的 43 个 ESTs 具有 Poly(A)和 CpG 结构；159 个新 ESTs 具有 ORF；而剩下的 174 个 ESTs 没有上述结构。随后采用氨基酸同源性比较软件对具有开放阅读框的 159 个 ESTs 进一步分析，结果是仅 1 个 EST 的氨基酸序列已报道，其余的 158 个 ESTs 均为具有编码功能的新序列。

3　讨论

EST 可提供不同组织、不同发育阶段的基因表达信息。通过对代表同一基因的 cDNA 克隆进行重复和全长测序，再根据 cDNA 丰度和组织分布来确定高转录基因和管家基因[5]。此项研究对脂肪文库

的 2 104 个 ESTs 的测定，包含了数目最多的线粒体基因、主组织相容复合物 I 型的 SLA 基因，以及与糖代谢、脂肪代谢有关的基因，这说明在脂肪组织内与能量代谢和脂肪代谢有关的基因进行着大量的表达，故 EST 数据可对组织器官基因多样性及表达方式进行初步的认识。

提高瘦肉率、降低脂肪含量始终是猪育种工作的一个首要环节。对脂肪组织中基因表达代谢的系统认识，了解不同阶段的基因表达情况和特点、发现与脂肪代谢有关的重要新基因，无疑可为育种工作者提供有价值的选种、育种信息。笔者研究发现的具有编码功能的 158 个新 ESTs，在公众的数据库中没有找到同源序列，代表以前未鉴定过的新序列，可采用测全长、表达特征、染色体定位、组织分布、免疫学鉴定等方法将这些新序列进一步归类[6]。

EST 数据与物理定位相结合，可以产生高分辨的染色体基因图谱[7]。截止目前，在 GenBank 中的 4 900 多条脂肪 ESTs，以牛的脂肪 ESTs 序列最多，显然本研究所分析的 2 104 个 ESTs 序列将会极大地丰富猪的 EST 库，更加深刻地理解脂肪在机体内的沉积特点，并为猪基因组计划和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, Merril 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, Fleischman 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(Suppl.28): 3 – 17.

[3] David Gerhold, Caskey C T. It's the genes! EST access to human genome content. *Bioessays*, 1996, 18(12): 973 – 981.

[4] Hartl D L. EST! EST!! EST!!!. *Bioessays*, 1996, 18(2): 1 021 – 1 023.

[5] Schmitt A O, Specht T, Beckmann G, Dahl E, Pilarsky C P, Hinzmann B, Rosenthal A. Exhaustive mining of EST libraries for genes differentially expressed in normal and tumour tissues. *Nucleic Acids Research*, 1999, 27(21): 4 251 – 4 260.

[6] De Souza S J, Camargo A A, Briones M R, Costa F F, Nagai M A, Verjovski-Almeida S, Zago M A, Andrade L E, Carrer H, El-Dorry H F. Identification of human chromosome 22 transcribed sequences with ORF expressed sequence tags. *Proceedings of the National Academy of Science of USA*, 2000, 97(23): 12 690 – 12 693.

[7] Boguske M S, Schuler G D. Establishing a human transcript map. *Nature Genetics*, 1995, 10: 369 – 371.

（责任编辑 林鉴非）