

基于表达序列标签(EST)的基因克隆和 基因表达分析研究进展*

杨克强, 王跃进, 张今今, 王西平, 张剑侠, 万怡震

(西北农林科技大学 园艺学院, 陕西 杨陵 712100)

[摘要] 表达序列标签(expressed sequence tag, EST)是在生物体特定时空表达基因的一段 cDNA 序列。随着生物信息学的发展, EST 已成为基因定位、基因克隆、基因表达分析的有力工具。文章对应用 EST 克隆基因和分析基因表达的研究进展进行了综述, 并讨论了利用 EST 进行基因克隆和表达分析时应注意的问题和发展趋势。

[关键词] 表达序列标签; 基因克隆; 基因表达

[中图分类号] Q 78

[文献标识码] A

[文章编号] 1000-2782(2002)04-0141-05

在人类基因组计划(Human Genome Project, HGP)实施过程中, 美国国立卫生研究院(National Institutes of Health, NIH)生物学家 Venter J C 于 1991 年提出了发现表达基因的新战略, 即用表达序列标签(Expressed Sequence Tag, EST)克隆和定位新基因^[1]。这一构想的具体过程是将 mRNA 在体外反转录成 cDNA 后克隆到载体上, 构建 cDNA 文库, 再从 cDNA 文库中大规模随机挑选 cDNA 克隆, 对其 3' 端或 5' 端进行一步法测序(Single-run sequencing)获得 EST。EST 就是一段长约 150~500 bp 的基因表达的外元序列片段。利用 EST 克隆基因和分析基因表达, 使得克隆和定位新基因的策略发生了革命性的变革^[2,3]。近年来, 三大国际一级生物信息数据库, 即美国国家信息中心(National Center of Biotechnology Information, NCB I)的 GenBank (<http://www.ncbi.nlm.nih.gov/web/GenBank/index.html>)、欧洲分子生物学实验室(European Molecular Biology Laboratory-European Bioinformatics Institute, EMBL-EBI)的 EMBL (<http://www.ebi.ac.uk/databases/index.html>)和日本 DNA 数据库(DNA Data Bank of Japan, DDBJ) (<http://www.ddbj.nig.ac.jp/>)新收录的核酸序列数据中, EST 占 65% 以上^[4]。随着生物信息学(Bioinformatics)的发展, EST 在新基因的发现、基因定位、基因表达和重组蛋白表达等方面起到了重要作用, 已成为人们认识生物个体生长发

育、繁殖分化、遗传变异、疾病发生、衰老死亡等生命过程的有力工具^[5~7]。

1 基于 EST 的基因克隆

1.1 EST 电子杂交基因克隆

目前, 通过数据库查询 cDNA 文库直接测序、mRNA 差别显示(DDRT-PCR)、代表性差异分析(RDA-PCR)和抑制差减杂交(SSH)等方法获得的 EST 数据越来越庞大。GenBank 数据库中收录的 EST 序列有数百万个之多, 如: 拟南芥的 EST 有 10 万余个^[8], 水稻有 4 万余个, 玉米有 7 万余个^[9]。我国也计划在未来的几年中获得水稻的 EST 10 万个^[10]。与此同时, 随着人类基因组框架图的发表^[11,12]和秀丽线虫(*Caenorhabditis elegans*)^[13]、黑腹果蝇(*Drosophila melanogaster*)^[14]、拟南芥(*Arabidopsis thaliana*)^[15]等模式生物基因组测序的完成, 这些生物的基因得以初步确定。据估计, 人类约有 3 万余个基因, 线虫有 1.9 万个基因, 果蝇有 1.36 万个基因, 拟南芥有 2.5 万多个基因。这些基因都已被收录到公共数据库中。由于 EST 代表着一段表达基因序列, 这样就可利用其与公共数据库的基因进行“电子杂交”, 克隆有编码功能的基因。这个过程理论上是一个虚拟的 RNA 印迹(Virtual northern blots)。

利用 EST 电子杂交基因克隆的具体步骤可总结为: 首先, 用已知的 EST 序列为起始序列, 采用

* [收稿日期] 2001-12-06

[基金项目] 国家自然科学基金资助项目(30170653)

[作者简介] 杨克强(1965-), 男, 甘肃张家川人, 高级讲师, 在读博士, 主要从事果树育种和果树分子生物学研究。

BLAST (Basic Local Alignment Search Tool, BLAST) 检索程序检索数据库中与其同源或有部分重叠的 EST 序列, 以确定哪些代表已知基因, 哪些代表已知 EST, 哪些代表新发现的 EST 序列; 还可以多个 EST 序列电子延伸组装为基序(contigs), 即通过与已知的 EST 序列重叠区(overlapping sequence)的配对延伸, 组装成毗邻序列; 以此基序为被检序列再进行 BLAST 检索; 重复以上过程, 直至没有更多的重叠 EST 检出, 基序不能继续延伸为止。此外, 也可直接用 EST 序列从数据库中检出未知功能的基因序列, 然后进行定性分析(characterization)。定性分析是电子杂交克隆基因的重要步骤, 主要用核酸和蛋白序列分析软件分析重复序列、转录结合位点、编码区统计特征、内含子/外显子剪接位点及氨基酸序列和蛋白质二级结构等, 确定组装或查询的序列是否具有基因特征^[16]。

常用的生物信息学分析工具有: 适于核酸序列查询的 BLASTN、BLASTX、TBLASTX (主要用于 EST 分析) 和适于蛋白质序列查询的 BLASTN、BLASTP 数据库; 另外, DBEST (Database EST) 是集成了 Genbank、欧洲分子生物学实验室数据库、日本 DNA 数据库、美国 Brookhaven 国家实验室蛋白质结构数据库(PDB) 的非冗余的 EST 数据库; Nr 既是一个汇集了瑞士蛋白质数据库(Swiss-Prot)、蛋白质信息资源数据库(PR)、日本蛋白质研究基金会蛋白质数据库(PRF) 以及从 Genbank 序列编码区中的蛋白质和 PDB 中拥有原子座标的蛋白质的非冗余蛋白质数据库, 又是一个非冗余 DNA 数据库。近年来, 生物信息学仍在不断完善和发展中^[17, 18], 我国于 1996 年在北京大学建立了生物信息中心 CBI(<http://www.chipku.edu.cn>), 引进了核酸、蛋白序列和结构等近 40 个数据库^[19], 一些新的核酸和蛋白序列相似性分析工具也在开发^[20]。这些都为利用 EST 在数据库中克隆基因提供了更为快捷和准确的手段。

利用 EST 电子杂交克隆基因已有许多成功的范例。李永青等^[21]用果蝇心脏发育基因 *zfh-1* 的 cDNA 序列从 BLASTX (网址为: <http://www.ncbi.nlm.nih.gov/blast>) 查找基因序列数据库“Nr”, 从数据库中检索出了该 cDNA 的各种人类同源序列新编码的氨基酸序列, 在其中发现了一个新的蛋白质, 用蛋白质序列检索到了 ZNF322, 这是一个与人类锌指基因 ZNF 家族同源的 cDNA, 用该段 cDNA 查询 BLASTN 发现了一个包含 ZNF322 的

基因组序列, 用 BSL 程序(网址为: <http://dot.im-gen.bcm.tmc.edu/9331>) 对该序列进行特性分析, 找到了该序列的启动子 5' 端起始密码子 ATG, 3' 端终止密码子 TGA 以及加尾信号 ATAAA, 成功克隆了一个新基因, 经国际基因命名委员会批准, 命名为 ZNF322 基因。李刚等^[22]用 mRNA 差异显示法, 在恶性转化的大鼠气管上皮细胞中, 克隆到不与任何已知开放阅读框架的基因同源的 ZA73 cDNA 序列。通过 Tigen (意大利 Telethon 遗传及医学研究所) 的服务器对人及小鼠的 UniGene 数据库进行 BLAST 检索, 检出 ZA73 与人的 Hs100293 号基因族有同源性, 该基因族由 128 条 EST 序列构成, 可组装为 1871 bp, 777 bp 2 个基序; 与小鼠的 Mm22710 号基因族有同源性; 该基因族由 22 条 EST 序列构成, 可组装为 1417 bp 的基序。对尝试性的人类一致性序列数据库(Tentative Human Consensus Sequence, THC) 进行 BLAST 检索, 检出 ZA73 与 261843 号 THC 序列同源, 该序列长度为 2797 bp, 由 97 条 EST 序列组装而成, 与人的 N-乙酰葡萄糖胺转移酶(GlcNAc) 基因有 704 bp 的完全一致性重叠区。并从基因组及其表达的整合分子分析(Integrated Molecular Analysis of Genome and their Expression, MAGE) 协定(网址为: <http://bbp.llnl.gov/bbp/image/>) 免费索取相应的 cDNA 克隆用于进一步的研究。

1.2 EST 结合 PCR 技术基因克隆

经电子杂交基因筛选后的 EST 及其延伸产物基序可作为某一基因的标签, 结合 PCR 技术克隆新基因。一般是以该标签序列提供的信息设计合成基因特异引物(gene specific primer, GSP), 在使用 Oligo(dT) 对 mRNA 进行反转录的同时加上锚定引物(anchored primer), 然后在总 RNA 中, 采用 cDNA 末端快速扩增(rapid amplification of cDNA ends, RACE) 技术^[23], 获得目的基因的全长序列。也可以用该标签序列提供的信息设计引物, 合成探针对 cDNA 文库进行筛选, 挑取最长的阳性克隆进行测序, 以期获得目的基因。邢桂春等^[24]用电子延伸结合 RACE 技术克隆出了肿瘤相关 MAGE 基因家族的一个新基因 MAGE-D1; 罗瑛等^[25]用与辐射诱导基因(Radiation induced gene 1, RIG1) 相关的两段 EST 序列, 电子基因筛选预测其外元序列, 据此设计巢式 RACE-PCR, 成功分离了 RIG1 基因的全序列; 韩为东等^[26]用已获得的一段与血友病复发相关的 DNA 序列, 在 NCBI 的 hESTs 数据库中进行

电子杂交对重叠克隆片段组装,再设计引物,进行 RACE-PCR,扩增出了 *LRP16* 基因的全序列。Qian Jun 等^[27]用一个与泛蛋白途径(Ubiquitin pathway)相关的 EST 序列,结合 RACE-PCR 技术,分离了 *UBAP1* 基因。

2 基于 EST 的基因表达分析

在拟南芥基因组研究中^[8],用已知的 EST 和蛋白质序列对基因进行精确定位,93% 的 EST 和表达序列相匹配。与此同时还将已确定的 25 498 个基因与秀丽线虫、黑腹果蝇、酿酒酵母(*J. cerevisiae*)、流感嗜血杆菌(*H. influenzae*)的基因在结构和功能上的同源性进行了比较,这不仅在基因水平上认识了物种间的差异和进化关系,而且为基因表达研究提供了思路。EST 在果蝇基因结构和定位分析中也发挥了重要作用^[14],Rubin 等^[28]还报道了通过 5 端 EST 和 3 端 EST 测序,获得非冗余的具有全长序列的 cDNA 克隆,这些克隆能够精确地揭示基因结构,包括开放阅读框架(open reading frame, ORF)和 3 端不翻译区(untranslated regions, UTRs),用这一方法确定了大约 40% 的果蝇基因结构。

EST 代表着一段表达的功能基因,根据同源同功原理,按照一组同源 EST 在不同组织不同时期的分布,利用数据库提供的信息,可在基因整体表达的水平上研究相关的功能及代谢。李永青等^[21]在电子杂交获得 *ZN F322* 基因后,用其 cDNA 序列检索人体表达数据库(Human EST)和基因组 DNA 序列库(HTGS),获得了属于该基因的 38 个 EST,根据这些 EST 在不同组织的分布,确定了该基因在人体不同组织中的表达情况。Hisada 等^[29]构建了温州蜜桔(*Citrus unshiu* Marc)果实细胞快速膨大期的幼果组织 cDNA 文库,对获得的全部 950 个 EST 序列在 GenBank 中进行了同源性分析(用 FASTA 程序),426(44.8%) 个 EST 与已知序列基因有同源性,524(55.2%) 个 EST 为新的 EST 序列。大部分与已知序列同源的 EST 与转录、翻译、细胞壁代谢和胁迫反应相关基因有关,而与已知基因同源的 EST 与持家基因、胁迫反应基因(包括金属硫蛋白、Lea 蛋白、热激蛋白基因)相关。Hisada 等^[29]对这些 EST 组装成的基序进行分析,发现最大的一个基序由 10 个 EST 组成,与拟南芥 EST 和转录序列相关。结合已有资料,分析了温州蜜桔果实细胞快速膨大期的代谢特性。Hisada 等^[30]还运用来自 192 个 cDNA 克隆的 EST 对上岛脐橙(*Citrus sinensis*)的未成熟种子

进行分析。Lin 等^[31]曾对白菜花芽 cDNA 文库的 965 个 EST 序列进行了同源性分析。由于用 EST 研究基因表达的稳定性高且分析规模大,使其成了研究基因差异表达的热点。在植物上诸如水稻胚乳发育中的基因表达^[32],拟南芥防御系统基因表达^[33],杨树^[34]和松树^[35]木质形成中的基因表达等,都是通过 EST 分析来进行的。用 EST 研究植物基因表达所涉及的植物种类多,研究内容广泛。

利用来自特定组织的 EST 还可对未知功能的基因进行功能分析。潘巍等^[36]用肝细胞癌组织的 mRNA 反转录合成 cDNA 探针,杂交筛选获得了在肝癌和正常肝中表达量有明显差异的 EST 片段,根据 GenBank 的相关数据合成引物,从人肝 cDNA 文库中扩增出了一个新基因,经查该基因已登录 GenBank 但功能未知。他们进一步分析了该基因的转录结合位点、同源性、氨基酸组成,并预测了蛋白二级结构及蛋白定位,用 Northern 杂交分析了该基因在肝、胰脏、肾、脑和肺等不同组织的分布,脂质体转染 NIH3T3 细胞检测表达,推测其可能与肝癌相关,暂命名为 *fup1*。

3 EST 应用中应注意解决的问题

3.1 EST 基因克隆和基因表达分析的精确性

基于 EST 电子杂交的基因克隆为预测基因序列和功能提供了经济、快捷的手段和大量信息。但由于 mRNA 在不同时空的丰度差异、反转录酶的误配、同功密码子的存在、物种间的特异性、电子杂交对错误碱基的敏感性等,常造成预测的精度较低。为了克服这些问题,在序列分析时应用多种综合软件如 GeneScan、Phred、Polyphred 等,使候选序列同见于 2 个或 2 个以上的文库,以有效排除假阳性结果^[37]。还有一个值得注意的问题是:在对 cDNA 克隆随机测序时,如何避免重复测序或漏测,因为高丰度的基因会被反复测序,而低丰度的基因有可能漏测^[38,39]。要解决这一问题除了建立标准化的 cDNA 文库,使 cDNA 文库中基因出现的频率基本一致外,还可采用以下 4 项措施:建立差减文库,扣除管家基因,使所研究的差异基因富集;用出现频率高的基因作探针,文库杂交筛选后再测序;用 mRNA 差别显示技术获得特异性片段,测序分析;

在 EST 数据库中剔除冗余重复序列,DBEST 和 Nr 都是这样的数据库。

3.2 重视基础实验的描述标准化

应用 EST 数据结合生物信息学分析,可为生物

学研究提供极有价值的参考和新的研究思路, 但一个基因的原初确定及其表达分析还要结合实验的方法来解决。这就要求要注意实验过程和描述标准化, 以建立标准注释的公共数据库, 使实验数据具有更准确的可比性。具体而言, 标准化的基因表达数据库要有以下 5 类必要的信息: 联系信息, 提交数据的实验室或研究人员的信息; 杂交靶探针信息, 对阵列上的每个“点”, 应有相应的 DNA 序列在公共数据库中的编号; 杂交样本、细胞类型和组织来源要用标准语言描述; mRNA 转录定量; 统计学意义, 应经济合理且有足够重复次数地重复一个实验, 以便给出基因表达测定的变异情况。研究者应注意实验描述的标准化, 使研究结果快速登陆数据库。

4 展 望

虽然 EST 有所获基因组信息不全, 一些调控序列、内元序列等在基因表达调控中起重要作用的信息不能体现出来等缺点, 但它在基因组学研究中的优势是显而易见的^[1]。随着人类基因组计划和一

些模式生物基因组计划的完成, 有关基因组研究的重点将转向功能基因组和后基因组的研究, EST 作为基因组学与功能基因组学及后基因组学之间的桥梁, 将会起到更加显著的作用。模式生物基因组计划完成后, 其他生物的基因组测序有无必要正在争论之中, 但 EST 应用却方兴未艾^[5, 9]。在未来的几年中, 基因定点突变技术、定向重组技术、基因表达 DNA 芯片技术和蛋白质组计划都会有长足发展, 更多的全长序列 cDNA 文库将被建立, EST 将用于在组织、细胞和亚细胞水平上, 研究不同环境状态下的基因表达系统以及生物的系统发育^[40]。比较不同的 EST 数据, 结合实验方法还可获得单核苷酸多态性 (single nucleotide polymorphism, SNP), 可以构建更为个性化的基因图^[41, 42]。这些研究提供的资料将丰富和发展生物信息学, 人们将开发模型植物 (virtual plant)^[40], 甚至在硅片上进行“杂交”作图, 定位与复杂疾病相关的基因^[43], 这将使基因组研究进入一个全新的阶段。

[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] Collins F S, Patrino A, Jordan E, et al New Goals for the U. S Human Genome Project: 1998- 2003[J]. Science, 1998, 282(5389): 682- 689
- [3] Hattori M, Tsukahara F, Furuhashi Y, et al A novel method for making nested deletions and its application for sequencing of a 300 kb region of human App Locus[J]. Nucleic Acids Res, 1997, 25: 1802- 1808
- [4] Harte F, Tosser-klopp G, Cloucard-martinato C. Expressed sequenced tags for genes: review[J]. Genet Sel Evol, 1998, 30 (5): 521- 541
- [5] 骆 蒙, 贾继增. 植物基因组表达序列标签(EST)计划研究进展[J]. 生物化学与生物物理进展, 2001, 28(4): 494- 497
- [6] 陆佳韵, 王秀琴, 吴 旻. EST 及其应用[J]. 生命科学, 1999, 11(4): 186- 189
- [7] 王华春, 陈清轩. 更好地利用 EST 数据库[J]. 生物化学与生物物理进展, 2000, 27(4): 442- 444
- [8] The Arabidopsis Initiative Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*[J]. Nature, 2000, 408(14): 796- 815
- [9] Adams D. Now for the hard ones[J]. Nature, 2000, 408(14): 792- 793
- [10] 薛勇彪. 国家重点基础研究发展规划项目: 水稻重要性状的功能基因组学研究项目简介[J]. 植物学报, 2001, 43(7): 769- 770
- [11] Venter J C, Adams M D, Myers E W, et al The sequence of the human genome [J]. Science, 2001, 291: 1304- 1351
- [12] International Human Genome Sequencing Consortium. Initial Sequencing and analysis of the human gene[J]. Nature, 2001, 409: 860- 921
- [13] The *C. elegans* Sequencing Consortium. Sequence and analysis of the genome of *C. elegans*[J]. Science, 1998, 282: 2012- 2018
- [14] Adams M D, Celniker S E, Holt R A, et al The Genome sequence of *Drosophila melanogaster*[J]. Science, 2000, 287: 2185- 2195
- [15] Gelfand M S. Prediction of function in DNA Sequence analysis[J]. J Comput Biol, 1995, 2: 87- 115
- [16] 梁国栋. 最新分子生物学实验技术[M]. 北京: 科学出版社, 2001. 366- 370
- [17] Calvet J P. SOFTWARE: Comprehensive sequence analysis[J]. Science, 1998, 282(5391): 1057- 1058
- [18] Altschul S F, Madden T L, Schaffer A A, et al Capped BLAST and PSI-BLAST: a new generation of protein database search program[J]. Nucleic Acids Res, 1997, 25(17): 3389- 3402
- [19] 罗静初, 江 涛, 李 兵, 等. 分子生物信息检索系统和数据库[J]. 高技术通讯, 1998, 10(10): 61- 63

- [20] 孙焕东,展成岗,贺福初. 一个基于Blast程序的多重序列对齐程序——Mblast[J]. 生物化学与生物物理进展, 2001, 28(3): 419- 422
- [21] 李永青,曾伟奇,朱传炳,等. 人类新基因ZNF322的初步研究[J]. 生命科学研究, 2001, 5(2): 141- 145
- [22] 李刚,胡迎春,张开泰,等. “电子”cDNA文库筛选指导基因的全长cDNA克隆[J]. 生物技术通讯, 2000, 11(1): 1- 4
- [23] Borgen N D, Sato W L, Drewes L R. A Lock-clocking oligo (dT) primer for 5' and 3' RACE PCR [J]. PCR Methods Applic, 1992, 2: 144- 148
- [24] 邢桂春,张成岗,魏汉东,等. 采用RACE技术获得全长人新基因MAGE-D1[J]. 中国生物化学与分子生物学报, 2001, 17(2): 203- 208
- [25] 罗瑛,隋建丽,铁轶,等. 生物信息学辅助定位及延伸辐射诱导未知表达序列标签[J]. 生物化学与生物物理进展, 2001, 28(2): 188- 191
- [26] 韩为东,于力,楼方定,等. 一个新的白血病相关基因LRD16全长cDNA克隆序列分析及表达特征[J]. 中国生物化学与分子生物学报, 2001, 17(2): 209- 214
- [27] QIAN Jun, ZHANG Xiao-Hui, YANG Jian-Bo, et al. Cloning and expression analysis of a novel gene UBA P1, possibly involved in ubiquitin pathway[J]. Acta Biochim Biophys Sinica, 2001, 33(2): 147- 152
- [28] Rubin G M, Hong L, Brokstein P, et al. A *Drosophila* Complementary DNA Resource [J]. Science, 2000, 287: 2222- 2224
- [29] Hisada S, Akihama T, Endo T, et al. Expressed sequence tags of *Citrus* fruit during rapid cell development phase [J]. J Amer Soc Hort Sci, 1997, 122(6): 808- 812
- [30] Hisada S, Moriguchi T, Hidaka T, et al. Random sequencing of Sweet Orange (*Citrus sinensis* Osbeck) cDNA library derived from young seeds [J]. J Japan Hort Sci, 1996, 65(3): 487- 495
- [31] Lin C D, Kim H Y, Kim M G, et al. Expressed sequence tags of Chinese cabbage flower bud [J]. Plant Physiol, 1996, 111: 577- 588
- [32] Liu J Y, Hara C, Umeda M, et al. Analysis of randomly isolated cDNAs from developing edospem of rice (*Oryza sativa*): evaluation of expressed sequence tags and expression levels of mRNA s [J]. Plant Molecular Biology, 1995, 29(4): 685- 689
- [33] Epple P, Apel K, Bohmann H. ESTs reveal multigene family for plant defensins in *Arabidopsis thaliana* [J]. FEBS Letters, 1997, 400(2): 168- 192
- [34] Sterky F, Regan S, Karlsson J, et al. Gene discovery in the woodforming tissue of poplar: Analysis of 5692 expressed sequence tags [J]. Proc Natl Acad Sci USA, 1998, 95(22): 13330- 13335
- [35] Allona I, Quinn M, Shoop E, et al. Analysis of xylem formation in pine by cDNA sequencing [J]. Proc Natl Acad Sci USA, 1998, 95(16): 9689- 9693
- [36] 潘巍,席全胜,夏双络,等. 肝癌中高表达基因fup1的克隆及其功能[J]. 生物化学与生物物理学报, 2001, 33(2): 177- 178
- [37] 陈炜,张戈,张恩仲. 基于生物信息学的SNP候选位点搜寻方法[J]. 遗传, 2001, 23(2): 153- 156
- [38] Yamamoto K, Sasaki T. Large-scale EST sequencing in rice [J]. Plant Molecular Biology, 1997, 35(1): 135- 144
- [39] Hoog C. Isolation of a large number of novel screening strategy [J]. Nucleic Acids Res, 1991, 19(22): 6123- 6127
- [40] Somerville C, Dangl J. Plant Biology in 2010 [J]. Science, 2000, 290: 2077- 2078
- [41] Wang D G, Fan J B, Siao C T, et al. Large scale identification, mapping and genotyping of single-nucleotide-polymorphism in the human genome [J]. Science, 1998, 280: 1048- 1077
- [42] The International SNP Map working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms [J]. Nature, 2001, 409(15): 928- 933
- [43] Grup A, Gemer S, Ustuka J, et al. In silico Mapping of complex Disease-Related Traits in mouse [J]. Science, 2001, 292: 1915- 1918

Gene cloning and expression analysis based on expressed sequence tag

YANG Ke-qiang, WANG Yue-jing, ZHANG Jin-jin, WANG Xi-ping, ZHANG Jian-xia, WAN Yi-zhen

(College of Horticulture, Northwest Sci-Tech University of Agriculture and Forestry, Yangling, Shaanxi 712100, China)

Abstract: Expressed sequence tag (EST) is a partial cDNA sequence of the gene which expressed in specific tissue and specific stage. With the development of bioinformatics, EST represents a useful tool of gene identification, gene location and gene expression. The paper summarized the advance of gene identification and gene expression by ESTs and discussed some problems and tendencies in these procedures as well.

Key words: expressed sequence tag; gene cloning; gene expression