

表达序列标签 (EST) 分析及其在小麦研究中的应用^{*}

张延明¹, 曲 敏^{1,2}, 徐香玲¹, 李集临¹

(1. 哈尔滨师范大学生命与环境学院生物系, 哈尔滨 150080; 2. 哈尔滨商业大学 食品工程学院生物系, 哈尔滨 150076)

摘要: 简要叙述了表达序列标签 EST 技术的原理, 对小麦 EST 图谱的构建研究进展进行了综述。分别阐述了小麦七个同祖群染色体的 EST 图谱的研究进展, 同时对 EST 在小麦基因组研究中的应用前景进行了展望。

关键词: 小麦; EST; 基因组; EST-SSRs

中图分类号 S 512. 103. 53 文献标识码: A 文章编号: 1002-2767(2007)01-0082-04

Analysis of Expressed Sequence Tag (EST) and Its Application in Wheat Research

ZHANG YAN-ming¹, QU Min^{1,2}, XU Xiang-ling¹, LI Ji-lin¹

(1. Biology Department of Life Science and Environment College, Harbin Normal University, Harbin 150080; 2. Biology Department of Food Engineering College, Harbin University of Commerce, Harbin 150076)

Abstract: This article introduced the basic principles of EST analysis, summarized the advance of construction of mapped EST in wheat. The expatiation of the progress of EST mapping for wheat homologous chromosome groups 1 ~ 7. At the same time, its prospect of application on wheat genomics was also discussed.

Key words: wheat; expressed sequence tag; genomic; EST-SSRs

表达序列标签 (Expressed Sequence Tag, EST) 是美国国立卫生研究院 (National Institutes of Health, NIH) 的生物学家 Venter 于 1991 年提出的^[1]。随着人类基因组计划的开展, EST 技术首先被广泛应用于寻找人类新基因, 绘制人类基因组图谱, 识别基因组序列编码区等研究领域, 之后又被广泛应用于植物基因组研究^[2]。在植物中, 拟南芥 (*Arabidopsis thaliana* L.) 被选为第一个开展基因组计划的双子叶植物^[3]。水稻 (*Oryza sativa* L.) 是第二个被选中进行基因组研究的单子叶植物, 主要因为它的基因组小 (490 Mb), 并且是重要的粮食作物, 对谷物的比较基因组学研究表明, 水稻、小麦、玉米、高粱等的基因组成、基因顺序等存在高度共线性, 因此利用对水稻基因组研究的结果, 可从其他谷物中分离、鉴定出对应的基因。该计划于 1991 年

启动^[4], EST 计划为其中重要的组成部分。到 2003 年 6 月, 美国国家生物技术信息中心 (NCBI) 的 EST 数据库中 (dbEST) (<http://ncbi.nlm.nih.gov/dbEST/index.html>) 已录入的来自不同物种的不同组织的 EST 共有 17 291 123 条, 其中人和鼠的最多。到 2006 年 10 月 NCBI 数据库中的 EST 共有 38 953 178 条, 其中水稻 (*Oryza sativa* L.) 1 188 565 条, 玉米 (*Zea mays* L.) 1 143 728 条, 拟南芥 (*Arabidopsis thaliana* L.) 622 973 条, 大豆 (*Glycine max* L.) 359 151 条。除了模式植物和农作物以外, 近年来也开展了一些木本植物的 EST 研究, 如柑橘^[5]、杨树^[6]等等。

1 EST 技术的原理

EST 是指在来源于不同组织的 cDNA 文库中

^{*} 收稿日期: 2006-10-26

第一作者简介: 张延明 (1977-), 男, 助教, 在读博士, 主要从事细胞遗传学研究。Tel: 89180802; E-mail: blueright@163.com。

通讯作者: 李集临

随机挑选克隆、测序,得到部分 cDNA 序列,一个 EST 对应于某一种 mRNA 的 cDNA 克隆的一段序列,长度一般为 150 ~ 500 bp,只含有基因编码区域,因此,EST 可代表生物体某种组织某一时间的一个表达基因,所以被称之为“表达序列标签”;而 EST 的数目则显示出其代表的基因表达的拷贝数,一个基因的表达次数越多,其相应的 cDNA 克隆越多,所以通过对 cDNA 克隆的测序分析可以了解基因的表达丰度^[7]。目前构建 cDNA 文库一般都使用试剂盒,方法成熟,而且飞速发展的 DNA 测序技术,也使得进一步降低大规模 DNA 序列测定成本成为可能^[8]。

2 EST 在小麦研究中的应用

1998 年 8 月在第九届国际小麦遗传学会议上,由美国、澳大利亚、英国三国科学家发起,正式成立了国际小麦族 EST 协作网(International Triticeae EST Cooperative ITEC),同时也拉开了国际麦类 EST 研究的序幕。共有 12 个国家的 35 个实验室注册参加了该研究计划(<http://wheat.pw.usda.gov/genome/participants>)。注册成员与 ITEC 签署有关协议,将所获得的 EST 数据输入该组织的数据库,成员可以共享数据资源^[2]。到 2000 年 5 月,在 GenBank 中只有 9 条小麦(*Triticum aestivum* L.)ESTs,86 条大麦(*Hordeum vulgare* L.)ESTs,黑麦(*Secale cereale*)的 ESTs 一条都没有^[9]。到 2004 年 1 月,在 GenBank 中有小麦 ESTs 577 538 条,大麦 ESTs 377 074 条,黑麦 ESTs 9 194 条。到 2006 年 10 月 NCBI 的数据库中,小麦 ESTs 855 066 条,大麦 ESTs 437 321 条,黑麦 ESTs 9 195 条(http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html)。

2.1 小麦 EST 图谱的构建

2003 年 3 月 17 日负责合作绘图的十个实验室绘制出 4 485 个 ESTs (http://wheat.pw.usda.gov/NSF/progress_mapping.html)。2004 年 2 月 2 日绘制出 8 318 个 ESTs,其中 7 637 个得到证实,产生了近 40 000 个记录位点。用于分析的图谱亚群和已证实的 ESTs(4 485 个,2003 年 3 月 17 日),经过确认和去除复制位点后,产生 16 093 个位点,G. R. Lazo 等人^[10]将这些位点绘制在中国春非整倍体和缺失材料的三个基因组上,绘制在 B 基因组上的位点(5 774)多于 A 基因组(5 173)和 D 基因组(5 146)上的位点。D 基因组的 EST 的密度显著高于 A 基因组和 B 基因组。通常情况下,EST 密度的

增加同距离着丝粒的物理距离有关。EST 密度的主要区域在染色体的末端部分。绝大多数重要农艺性状基因位于 EST 密度区域。L. L. Qi, 等人^[11]对同祖群染色体的分析表明,共有 15 843 位点分布在七个同祖群染色体上,位点数量依次为 2 212、2 600、2 266、2 236、2 338、2 043 和 2 148。

2.1.1 小麦第 1 同祖群染色体 EST 图谱 J. H. Peng 等^[12]将 944 个 ESTs 产生的 2 212 EST 位点,绘制在六倍体小麦(*Triticum aestivum* L.)的第 1 同祖群染色体上。944 个 ESTs 中有 367(39%)个被绘制在第 1 同祖群中的每条染色体上,用来构建第 1 同祖群的缺失图谱。这 367 个 EST 已经存储在互联网上(<http://wheat.pw.usda.gov/pubs/2004/Genetics/>)。EST 位点在第 1 同祖群染色体上的分布是不均一的,位点数目分别是:1A 染色体为 660 个位点,1B 为 826 个位点,1D 为 726 个位点。在 A、B、D 三条染色体上,长臂上的 EST 位点数目要多于短臂,并且染色体臂上的 ESTs 的分布不是随机的,EST 簇集中在染色体短臂的末端区域和长臂的中段区域。在其他同祖群上,第 1 同祖群的 ESTs 复制位点产生的几率为 35.5%。75%的小麦第 1 同祖群染色体上的 ESTs 同水稻基因组序列($E \leq e^{-10}$)相匹配,只有 9.5%的小麦第 1 同祖群的 ESTs 同拟南芥基因组序列相匹配。

2.1.2 小麦第 2 同祖群染色体 EST 图谱 E. J. Conley 等人^[13]将 1 110 个 ESTs 产生的 2 600 个位点(http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi, 2003—3—17),通过 Southern 杂交绘制到小麦非整倍体和缺失材料的第 2 同祖群染色体上。651 个 EST 探针产生的 769 个位点绘制在 2A 染色体上,占位点总量的 29.6%,728 个 EST 探针产生的 959 个位点绘制在 2B 染色体上,占位点总量的 36.9%,725 个 EST 探针产生的 872 个位点绘制在 2D 染色体上,占位点总量的 33.5%。染色体 2B 上的 EST 位点要多于染色体 2A 和 2D。

2.1.3 小麦第 3 同祖群染色体 EST 图谱 J. D. Munkvold 等人^[14]将 996 个 ESTs 产生的 2 266 个限制性片段(位点)绘制在小麦第 3 同祖群染色体上。这些位点在染色体 3A、3B、3D 上的位点数目依次为 634、884 和 748。除了 3DS 和 3DL 以外,高密度 ESTs 区域存在于染色体末端,低密度 ESTs 区域存在于染色体臂的中部。染色体 3B 的 ESTs 位点最多,其次为 3D 和 3A,3D 的位点密度最大

(1.1), 然后是 3B (1.02), 最低的是 3A (0.88)。在第 3 同祖群染色体公认的图谱中大约有 232 (44%) 个序列同水稻染色体 1 相匹配。第 3 同祖群染色体中的 ESTs 单基因图谱大约有 21 个同拟南芥编码区相匹配。

2.1.4 小麦第 4 同祖群染色体 EST 图谱 Miftahudin, K. 等人^[15] 经过 938 个 ESTs 杂交检测, 将 1 918 个位点绘制在小麦第 4 同祖群染色体上, 这 938 个 ESTs 来源于 26 个 cDNA 文库。研究发现这 1 918 个位点在第 4 同祖群的三条染色体上的分布是不均一的, 在 4A、4B、4D 染色体上绘制的比率分别为 41%、28% 和 31%。第 4 同祖群染色体特有的 119 个 ESTs, 被用来绘制公认的第 4 同祖群染色体图谱。在这些 ESTs 中, 有 49% 同水稻染色体 3 有同源序列, 12% 同水稻其他染色体有匹配序列, 39% 同水稻没有匹配序列。使用 BLAST X (Basic Local Alignment Search Tool, BLAST) 搜寻全部蛋白质数据库, 发现 42% 的第 4 同祖群染色体的 ESTs 可划分到功能类别。

2.1.5 小麦第 5 同祖群染色体 EST 图谱 A. M. Linkiewicz 等人^[16] 构建了小麦染色体 5A、5B、5D 的高密度缺失图谱 (High-density deletion bin maps), 其中包括用 1 052 个 EST 探针绘制的 2 338 个位点, 和 217 个以前已经绘制的位点 (位点总数为 2 555 个)。研究结果表明, 在 5B 染色体上绘制出相当高的位点数 (38%), 5A 和 5D 染色体分别为 34% 和 28%。在三条染色体上产生多态水平的差异, 一部分原因可能是因为这些位点差异造成的。在 5B 染色体上发现复制位点的数量也很多 (42%)。除了已知的 5AL/4AL 易位和染色体 5A 的部分区域以外, 在第 5 同祖群染色体中观察到了良好的完整的共线性。在小麦第 5 同祖群的低拷贝 ESTs 数和水稻染色体 12 (88 个 ESTs)、9 (72 个 ESTs) 和 3 (84 个 ESTs) 的低拷贝 ESTs 数之间, 可以观察到统计学意义显著的共线性。

2.1.6 小麦第 6 同祖群染色体 EST 图谱 H. S. Randhawa 等人^[17] 用来源于 146 条染色体、染色体臂、非整倍体亚臂和缺失的材料的 37 个 cDNA 文库的 7 965 个独立成分, 进行小麦 ESTs 染色体定位, 有 882 个 ESTs 被确认在小麦第 6 同祖群染色体上。这 882 个 ESTs 被绘制成 25 个区域, 这些区域在 23 个断裂位点的侧面。用 882 个 ESTs 检测了 5 154 个限制性片段, 其中有 2 043 个 (位点) 定位在第 6 同祖群的染色体上, 806 个绘制在其他同祖

群染色体上。结果表明, 6B 染色体上绘制的位点最多, 而 6D 染色体上绘制的位点最少。同水稻基因组序列进行同源比较, 小麦第 6 同祖群有 43% 的 ESTs 得到确认。在这些 ESTs 中, 58% 存在于水稻染色体 2 上, 其余的存在于其他水稻染色体上。

2.1.7 小麦第 7 同祖群染色体 EST 图谱 K. G. Hossain 等人^[18] 将来源于 919 个 ESTs 克隆的 2 148 个位点绘制在小麦第 7 同祖群上。其中 528 个 ESTs 绘制在染色体 7A 上, 确认了 661 个位点; 549 个 ESTs 绘制在染色体 7B 上, 确认了 719 个位点; 613 个 ESTs 绘制在 7D 染色体上, 确认了 768 个位点。在大多数情况下, 位点数量在着丝粒区域很低, 在末端区域逐渐增加。在 919 个 ESTs 中, 有 44 个绘制在 7A 和 7D 染色体上。在第 7 同祖群中, 复制位点的水平为 24%, 绝大多数位点定位在染色体末端区域。119 个 ESTs 探针杂交三个片段, 绘制在第 7 同祖群的 3 条染色体上, 这些 ESTs 被定为路标探针, 用来构建公认的第 7 同祖群染色体图谱。小麦、水稻和大麦的路标探针的顺序和比较图谱的产生是依据水稻 BAC/PAC 文库和定位在水稻第 6、8 条染色体上的遗传标记。

3 展望

EST 技术已经在小麦基因组研究中发挥了重要作用, 为评定基因组中基因密度区域、分布情况以及研究基因组进化等方面提供了一种有力手段。同时, 小麦丰富的可用资源也为研究者提供了用武之地, 如小麦缺体—四体系统 (nullisomic—tetrasomic, NT)、双端体系统 (ditelosomic)、缺失系 (deletion lines) 等。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也将发挥更大的作用^[19], 这必将推动小麦高分辨率基因组图谱中基因精确定位的研究进程, 为深入研究多倍体基因组同相关物种的共线性分析及比较基因组研究提供帮助。

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] 骆蒙, 贾继增. 国际麦类基因组 EST 计划研究进展[J]. 中国农业科学, 2000, 33(6): 110-112.

[3] Rounsley S, Linx K K. Large scale sequencing of plant genome[J]. Curr Opin Plant Biol, 1998, 1(2): 136-141.

[4] Sasaki T. The rice genome project in Japan[J]. Proc Natl Acad Sci USA, 1998, 95(5): 2027-2028.

[5] Hisada S, Akinama T, Edo T, et al. Expressed sequence tags of citrus fruit during rapid cell development phase[J]. Amer

Soc HortSci, 1997, 122 (6): 808-812.

[6] Desprez T, Amsalem J, Caboche M, et al. The Arabidopsis thaliana cDNA sequencing project[J]. Plant J, 1998, 14 (5): 643-652.

[7] 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.

[8] 李红, 卢孟柱, 蒋湘宁. 表达序列标签(EST)分析及其在林木研究中的应用[J]. 林业科学研究, 2004, 17(6): 804-809.

[9] Bennett M. D., I. J. Leitch. 2003 Plant DNA C—values database [DB/OL]. [http://www. rbgkew. org. uk/ cval/ home. page. html](http://www.rbgkew.org.uk/cval/home.page.html). 2003-01-10.

[10] G. R. Lazo, S. Chao, D. D. Hummel et al. Development of an Expressed Sequence Tag (EST) Resource for Wheat (Triticum aestivum L.); EST Generation, Unigene Analysis, Probe Selection and Bioinformatics for a 16, 000—Locus Bin—Delineated Map[J]. Genetics, 2004, 168: 585-593.

[11] L. L. Qi, B. Echaliér, S. Chao, et al. A Chromosome Bin Map of 16, 000 Expressed Sequence Tag Loci and Distribution of Genes Among the Three Genomes of Polyploid Wheat[J]. Genetics, 2004, 168: 701-712.

[12] J. H. Peng, H. Zadeh, G. R. Lazo, et al. Chromosome Bin Map of Expressed Sequence Tags in Homoeologous Group 1 of Hexaploid Wheat and Homoeology With Rice and Arabidopsis [J]. Genetics, 2004, 168: 609-623.

[13] E. J. Conley, V. Nduati, J. L. Gonzalez—Hernandez et al. A 2600—Locus Chromosome Bin Map of Wheat Homoeologous Group 2 Reveals Interstitial Gene—Rich Islands and Colinearity With Rice[J]. Genetics, 2004, 168: 625-637.

[14] J. D. Munkvold, R. A. Greene, C. E. Bermudez—Kandianis et al. Group 3 Chromosome Bin Maps of Wheat and Their Relationship to Rice Chromosome 1[J]. Genetics, 2004, 168: 639-650.

[15] Miftahudin, K. Ross, X. —F. Ma et al. Analysis of Expressed Sequence Tag Loci on Wheat Chromosome Group 4 [J]. Genetics, 2004, 168: 651-663.

[16] A. M. Linkiewicz, L. L. Qi, B. S. Gill et al. A 2500—Locus Bin Map of Wheat Homoeologous Group 5 Provides Insights on Gene Distribution and Colinearity With Rice[J]. Genetics, 2004, 168: 665-676.

[17] H. S. Randhawa, I. M. Dilbirligi, I. D. Sidhu, et al. Deletion Mapping of Homoeologous Group 6—Specific Wheat Expressed Sequence Tags[J]. Genetics, 2004, 168: 677-686.

[18] K. G. Hossain, V. Kalavacharla, G. R. Lazo, et al. A Chromosome Bin Map of 2148 Expressed Sequence Tag Loci of Wheat Homoeologous Group 7[J]. Genetics, 2004, 168: 687-699.

[19] 杨克强, 王跃进, 张今今, 等. 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[J]. 西北农林科技大学学报(自然科学版), 2002, 30(4): 142

(上接 81 页)

[7] 陈学留, 张建华. 玉米根系生长与叶片衰老的相关观察. 莱阳农学院学报[J]. 1994, 11(1): 17-20.

[8] 王泽立. 玉米耐旱性性状遗传及对干旱的反应[A]. 山东省节水农业研究汇编[C], 山东: 山东农业出版社, 1995. 94-99.

[9] 朱志华. 不同抗旱性冬小麦幼苗根系对水分胁迫的反应[J]. 植物生理学通讯, 1996, 32 (6): 410-413.

[10] Passioura J B. Roots and drought resistance[J]. Agric water management, 1983, 7: 265-280.

[11] 李运朝. 玉米抗旱性研究进展[J]. 玉米科学, 2004, 12(1): 63-68.

[12] 刘玉涛, 邱振英, 王宇先, 等. 春玉米抗旱性鉴定指标比较研究[J]. 玉米科学, 2006, 14(4): 117-120.

[13] Edmeades G O. Causes for silk delay in a lowland tropical maize population[J]. Crop sci., 1993, 33(5): 1029-1035.

[14] 于海秋, 徐克章, 陈学求, 等. 玉米主要抗旱性状的配合力及遗传分析Ⅲ 生育时期[J]. 玉米科学, 2003, 11(3): 18-22.

[15] Hall A. J, Lemloff J. H, Jrapani N. Water stress before and during flowering in maize and its effects on yield, its components and their altermimants[J]. Mayolica 1981, 26: 19-38.

[16] DNN, E. W Daynard T, B Muldoon J, F, et, al. Resistance to drought and density stress in Canadian and European maize hybrids[J]. Plant Sci, 1984, 64: 575-5850.

[17] Kramer, P. J. Water relations of plants[M]. New York: Academic Press, 1983. 489.

[18] 罗淑平. 玉米抗旱性及鉴定指标的相关性分析[J]. 干旱地区农业研究, 1990, 8(3): 72-78.

[19] 张宝石, 徐世昌, 宋风斌, 等. 玉米抗旱基因型鉴定方法和指标的探讨[J]. 玉米科学, 1996, 4(3): 19.

[20] 黎裕. 作物抗旱鉴定的方法与指标[J]. 干旱地区农业研究, 1993, 3(1): 26-29.

[21] 林秋萍, 贡冬花, 李普安, 等. 夏玉米干旱适应性及其生理机制的研究[J]. 华北农学报, 1990, 5(4): 15-19.

[22] 王金胜. 水分胁迫对玉米幼苗几种生理生化指标的影响及其与抗旱性的关系[J]. 山西农业学报, 1992, 12(2): 137-140.

[23] 鲍巨松, 杨成书, 薛吉全, 等. 不同生育时期水分胁迫对玉米生理特性的影响[J]. 作物学报, 1991, 17 (4): 261-265.

[24] 王邦锡, 黄文常. 不同植物在水分胁迫条件下脯氨酸积累与抗旱性的关系[J]. 植物生理学报, 1989, 15(1): 46-51.

[25] 关义新, 戴俊英, 陈军, 等. 土壤干旱下玉米叶片游离脯氨酸积累及其与抗旱性的关系[J]. 玉米科学, 1996, 4(1): 43-45.

[26] Sincer, T. N. Changes in praline contraction exit plant tissues [J]. Austra Biol Sci., 1973, 26(1): 78-90.

[27] 张海明, 王茅雁, 侯建华, 等. 干旱对玉米过氧化氢酶、MDA 含量及 SOD、CAT 活性的影响[J]. 内蒙古农牧学院学报, 1993, 14(4): 92-95.

[28] 侯建华, 吕凤山. 玉米苗期抗旱性鉴定的研究[J]. 华北农学报 1995, 10(3): 89-93.

[29] 马瑞昆. 综述麦类作物和干旱的农学及生理研究[J]. 农作物研究资料, 1986, 4(1): 73-81.

[30] HUJMACHER B. 气孔和非气孔因素对棉花光合速率的控制作用[J]. 国外农学—棉花, 1986, (1): 24-28.