

表达序列标签(EST)及其在抗孢囊线虫大豆研究中的应用^{*}

毕影东¹, 郭东林¹, 杨冬鹤¹, 徐香玲¹, 李集临¹, 王晓萍¹, 柏永军²

(1 哈尔滨师范大学生物系, 哈尔滨 150080; 2 黑龙江省克山县北联镇农业综合服务中心, 克山 161633)

摘要: 表达序列标签(expressed sequence tags, EST)是一种快捷、高效地揭示基因组信息的方法。本文对EST概念、技术原理及其在基因组研究中的应用和进展等方面内容进行了综述,并对EST在大豆基因组研究和抗孢囊线虫大豆研究中的应用进展作一介绍。

关键词: 表达序列标签; 大豆; 抗孢囊线虫

中图分类号: S 565.103.53 文献标识码: A 文章编号: 1002-2767(2006)03-0090-04

EST and its Application on Soybean with Resistance to the Soybean Cyst Nematode

BI Ying-dong¹, GUO Chang-hong¹, YANG Dong-he¹, XU Xiang-ling¹,
LI Ji-lin¹, WANG Xiao-ping¹*, BAI Yong-jun²

(1 Biology Department of Harbin Normal University, Harbin 150080; 2 Beilian Agricultural Service Center, Keshan, Keshan 161633)

Abstract: Expressed Sequence Tags (EST) has been proved to be a rapid and efficient approach to disclose the information of genome. The concept, technique principle and its application on soybean with resistance to the soybean cyst nematode (SCN) were summarized in this paper.

Key words: expressed sequence tags (EST); soybean; soybean cyst nematode (SCN)

表达序列标签(expressed sequence tags, EST)是cDNA的部分序列,长度一般为300~500 bp,由大规模cDNA克隆一次性测序得到。由于cDNA的5'端或3'端的有限序列可特异性地代表一个基因,故被称为“表达序列标签”^[1]。EST技术是一种快速有效揭示基因组容量的方法,随着分子生物学和生物信息学的不断发展,人们越来越认识到EST在基因的发现、基因作图和基因组序列中编码区域的确定等方面起到重要作用,受到全世界许多实验室的重视,已成为基因发现、基因表达及重组蛋白表达等研究的强有力的分子生物学工具。它在新基因资源中扩展最为迅速,到2003年4月dbEST库中有1600多万条EST序列,来自于约500种不同物

种的组织和细胞,NCBI现在可以提供大量EST数据的查询^[2]。

1 EST技术的原理和方法

EST是指从cDNA文库中随机挑取克隆,对其进行一轮大规模测序所获得的部分cDNA的5'或3'端序列,长度一般为100~500 bp。5'UTR和3'UTR都是特定的,即每条cDNA的5'端或3'端的有限序列即可特异性代表生物体某种组织某个时期的一个表达基因。一个基因的表达次数越多,能够测到的相应EST也越多,所以通过EST分析即可了解基因表达情况和表达丰度。将得到的ESTs与

* 收稿日期: 2005-12-31

基金项目: 黑龙江省教育厅项目(10541088)

第一作者简介: 毕影东(1974-),男,黑龙江省望奎县人,硕士,从事分子遗传学研究。Tel: 0451-59788410; E-mail: ydbi308@163.com.

通讯作者: 王晓萍

GenBank 等数据库中的数据进行相似性分析、核酸或蛋白质序列的同源性比较等生物信息学分析是近几年来分离与克隆新基因及基因功能研究的一个行之有效的手段。

2 表达序列标签研究内容 and 应用进展

用 EST 技术来进行基因组研究的思想, 是由美国科学家 Venter 等在人类基因组计划开始时提出的, 称为 EST 计划。这一基因组研究策略的优越性显而易见, 因为表达基因只占整个基因组的 3% ~ 5%, EST 反映的是基因的编码部分, 所以 EST 计划可以直接获得基因表达的信息。另一方面, 用 EST 代替基因组测序, 可使研究费用大大降低, 效率大大提高, 具有多、快、好、省之特点。Adams (1991) 等从 3 种人脑组织 cDNA 文库随机挑取 609 个克隆进行测序, 得到一组人脑组织的 EST, 分析结果表明其中 36 个代表已知基因, 337 个代表未知基因, 这是关于 EST 技术应用的首次报道^[1]。EST 现在已经得到了广泛应用, 特别是在功能基因组研究中发挥作用巨大, 是研究基因最有效的捷径^[3]。应用在构建遗传学图谱、基因定位、分离与鉴定新基因、基因表达的研究、比较基因组学研究、生物信息学分析等方面。

2.1 构建遗传学图谱

遗传图谱、物理图谱和转录图谱是基因组计划要获得的三张遗传学图谱。构建染色体物理图谱需要大量的单拷贝短序列 (Sequence Tags Site, STS) 作为界标, 由于大多数基因是单拷贝的, 所以 EST 可以用来充当 STS^[4]。同样 EST 片段由于其多态性高可以作为分子标记, 用来建立遗传连锁图谱^[3, 5]。如有人用日本的水稻品种 Nippon bare 和印度品种 Kasalath 杂交 F₂ 代 186 个植株作遗传图谱, 选用了 2 300 个 DNA 标记, 12 个连锁群, 总的遗传距离为 1550cM, 其中 70% 的标记来源于 EST^[6]。基因图谱又称转录图谱, 为染色体 DNA 的某一段内所有可转录序列的分布图, EST 为转录基因的产物, 可直接用于图谱的构建。由于 RNA 的 3' UTR 是代表每个基因的特异序列, 将对应于 3' UTR 的 EST 序列进行 RH (Radition Hybrid) 定位, 即可构成由基因组成的图谱^[7]。

2.2 基因定位

EST 处于基因的编码区, 因而 EST 是与目标基因共分离的。以此为基础, 综合分析已有的遗传图谱、物理图谱和基因图谱, 可将基因在染色体上准确定位。先参考遗传图谱与物理图谱, 根据孟德尔

遗传特征, 将相关基因定位在更狭小范围的染色体基因组区段, 再通过查询以 ESTs 为基础的基因图谱, 即可获得这一目的基因。

2.3 分离与鉴定新基因

分离基因的经典方法为图位克隆和转座子标签克隆, 是利用分子标记或表型变化来分析鉴定基因。以 EST 为重要来源的染色体物理图谱进一步方便了对候选基因的连锁分析, 它可将基因确定在更狭小的染色体区段内, 缩小了基因的筛选范围。用 EST 取代对 cDNA 全长的筛选、基因组序列的鉴定等繁琐的实验操作, 可大大地提高分离基因的效率。将所获 EST 用生物信息学方法与各公共数据库中已知序列进行比较, 可迅速而准确地确定基因功能^[3]。李红等利用差异显示 PCR 技术获得的一条在胃癌和正常组织有差异表达的 EST 序列 (W 123 (GenBank 登录号为 AF150631), 通过与 GenBank 的 dbEST 库进行电子杂交, 采用 RNA 印迹对目的序列进行初步鉴定, 并进行组织分布分析。^[8]

2.4 基因表达的研究

由于用 EST 技术研究基因的表达稳定性高且分析规模大, 使其成为研究基因表达的热点。在获得大量 EST 的基础上进行数据库相似性检索, 获得对基因组结构、组织、表达等认识。高瑞娟等以结球白菜 (*Brassica rapa* L. ssp. *pekinensis*) 结球前期心叶为材料, 构建了结球前期球叶 cDNA 文库, 经分析约 77% 的功能已知蛋白质来自拟南芥 (*Arabidopsis thaliana*), 另外还有 60 个 EST 与已发表的植物基因没有同源性, 这一部分 EST 对于研究结球白菜独特的生理和形态发育途径以及开展基因组分子作图具有重要的意义。大量 EST 的获得为进一步了解结球白菜结球机制及获取相关基因提供了重要的序列信息^[9]。基因表达谱的研究更是 EST 研究中的热点, 如水稻胚乳发育中的基因表达^[10]、拟南芥防御系统基因表达^[11]、油菜保卫细胞的代谢研究^[12] 等等, 涉及的植物种类多, 研究内容广泛。

骆蒙等以抗白粉病品系“百农 3217 × Mardler” BC5F₄ 为材料, 构建了一个白粉病菌接种初期的抑制消减杂交 cDNA 文库, 测序获得 760 条 ESTs。与 GenBank 序列进行 BLASTX 分析, 获功能已知 ESTs 271 条。通过分析抗病相关基因, 推测 G 蛋白介导的信号传导途径、SA 信号传递系统、MAP 相关信号传递系统等参与了小麦抗白粉病过程。

SAR 基因在抗相关 ESTs 中的种类与数量最多。数据显示苯丙烷代谢途径、细胞壁结构修饰作用、细胞保卫机制参与了抗病过程。未知功能 ESTs 与 GenBank 序列进行 BLASTN 分析, 其中许多与病原菌、非病原菌诱导 cDNA 文库来源的 ESTs 同源; 新 ESTs 占全部 ESTs 的 16.6%^[13]。

2.5 比较基因组学研究

植物比较基因组学研究是从比较不同植物的遗传图谱开始的, 如禾本科植物遗传作图的共线性研究。利用 EST 中古老保守序列(ancient conserved region, ACR)来研究物种之间的进化关系, 可以避免选择家系、构建群体、大量的统计分析等周期长而繁琐的遗传图谱制作过程, 并且使结果更加准确。如对拟南芥的研究表明, 约 2/3 以上的 EST 中有 ACR 序列, 据此可以判定进化关系^[14]。

2.6 生物信息学分析

在各综合数据库中, EST 的增长速度最快。数据库的发展除了具有增长与更新速度快、复杂度与网络化程度高的特点, 根据 EMBL 统计截止 2003 年 4 月 14 日有约 1 610 万个 EST 供公共使用, 含盖约 200 个物种。数据库的专门化也是一趋势, 即建立特定物种或特定研究内容的数据库如拟南芥等模式生物的数据库和水稻、大豆等农业作物的数据库^[3]。随着研究的深化, 各种组织、器官的数据库也不断涌现。1998 年 8 月在 Saskatoon 大学举行的第九届国际小麦遗传大会, 以美国、英国、加拿大等国家为首成立小麦 EST 研究国际合作组织, 提出了建立小麦族 EST 公共数据库的行动计划。并成立了国际小麦表达序列标签联盟(TTEC)长期的目标是确定和破译小麦基因组所有基因在染色体上的定位和功能。目前在根、茎、叶、小穗、胚、盾片、幼苗和种皮等不同组织器官中, 开展了小麦抗病性、抗逆性等重要农艺性状的 EST 测序^[15]。

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。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)。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]。

3 大豆 EST 及大豆抗孢囊线虫的研究进展

3.1 大豆 EST 研究进展

目前, 转基因大豆主要是抗除草剂品种。今后, 抗虫、改善营养成分(如脂肪酸组成)将是转基因大豆的新重点。过去的十几年中, 大豆遗传图谱和生物技术取得了长足进展, 现已建立了涵盖 20 个连锁群, 包括 1 400 个标记的遗传图谱。中国科学院遗传与发育生物学研究所 A. -G. Tian, J. Wang 等构建了科丰 1 号在 SA 胁迫下的 cDNA 文库, 滴度达 4.2×10^7 , 对 3 万多个大豆 EST 进行了测序和聚类分析, 结果表明其中包括 7 536 个仅由一个 EST 组成(singletons)和 2 937 个片段重叠群(contigs)。有 26 个和抗性基因、依赖抗性基因的植物防卫基因以及耐逆相关基因有同源性。根据序列分析结果, 发展了 600 多个大豆新的 SSR 标记。Jiw 等构建了野生大豆在高盐状态下 cDNA 文库测序分析获得了 696 个特异 EST 序列有助于大豆抗胁迫机制的研究^[17]。根据拟南芥基因编码产物的分类将这些 unigene 进行了注释。认为和拟南芥不同源的基因是大豆特异的基因, 主要是编码和调节发育以及种子储藏蛋白合成有关的基因。SSR 分析结果说明大豆基因组比拟南芥和苜蓿均复杂。大豆 unigene 基因序列中的 GC 含量与拟南芥和苜蓿的相似。综合分析 EST 数据和 BAC-contig 序列推算出大豆基因组中的总基因数约为 63 000^[17]。美国公立大豆表达序列标记(EST)项目, 旨在通过高产率的序列分析技术对成千上万个基因表达的 mRNA 完成序列分析, 从而把研究目标定位于特定的组织或器官上, 以获取大豆基因表达的多样性信息。由美国 BELTSVILLE 农业科研中心建立的大豆基因及微阵列数据库(SGMD), 该数据库包括与大豆抗孢囊线虫相关的基因组成, EST 及微阵列数据以及利用大豆 EST 进行表达分析的工具。目前已储存 5 000 多万个微阵列数据和 20 000 个 ESTs。最近该数据库也收录了大豆抗孢囊线虫 ESTs。大豆基因及微阵列数据库(SGMD)已成为通过表达序列标签分析发现新基因的有利工具之一^[18, 19]。

3.2 抗孢囊线虫大豆 EST 的研究进展

据美国乔治马森大学 Alkharouf 和马里兰大学 Khand 等研究, 大豆抗孢囊线虫的机制十分复杂, 涉及到很多基因且大多不明。他们利用 3 号小种孢囊线虫接种大豆 12 h、2~4 d、6~8 d 后, 分别从根部提取 mRNA 构建 cDNA 文库通过表达序列标签分析可加快对参与抗病过程基因的了解, 他们从三个文库中分析了 3 454 个 cDNA 克隆, 其中有 25 个

cDNA 克隆源于孢囊线虫 RNA。在接种孢囊线虫的大豆根中受逆境诱导基因如 SAM22 和谷胱甘肽转移酶基因表达量较未接种的大豆根有大幅度提高, 早期防御基因如氧化酶基因在接种 12 h 表达量最多, 而在 6~8 d 的文库中这些基因表达量不大, 但编码未知蛋白质和逆境诱导蛋白持续大量表达, 这些 EST 和相关信息有助发现新基因和了解抗虫相关的蛋白质^[20]。Matthews 等在题为“运用生物信息学研究大豆抗孢囊线虫的机制和抗性基因”一文中指出: 在大豆基因组测序还未完成, 并且很少对大豆与大豆孢囊线虫的相互作用机制的基因功能研究的情况下, 我们应用生物信息学分析大豆的 ESTs 了解大豆抗孢囊线虫的相关基因和抗性机制。我们发现了很多抗性候选基因与大豆抗孢囊线虫相关的信号传导途径。我们建立了世界上第一个大豆 ESTs 的公用微阵列数据库, 该数据库的在线分析工具已成为研究分析大豆抗虫、抗病相关基因表达的快速、高效的方法之一。另外, 德国 KIEL 大学的 Daguanglain 等克隆了一个甜菜抗孢囊线虫基因 HS1pro-1, 该基因编码 282 个氨基酸的蛋白质, 同其他高等植物抗病基因一样富含亮氨酸并有跨膜成分。该基因的成功克隆有助于对大豆抗孢囊线虫机制的了解^[21]。

4 EST 尚存在的问题

虽然 EST 技术功能强大, 广泛应用于基因组研究的各个领域, 但就目前的技术而言, 仍然存在一定的局限性^[7]。①由于 EST 是经一次测序所得到的会包含测序错误, Arronson^[22]认为, 约 0.5% 的 EST 的克隆是错误或是相同的克隆, 约 5% 克隆的插入片段是反向的, 约 2.5% 的克隆是引物结合在 cDNA 内部, 所以所得序列并非从末端开始。② EST 的测序方法存在误差及许多应用程序也存在一定的局限, 使独立基因 (Unigene) 在聚类拼接过程中可能将同源性较高的不同基因混同在一起。③由于 EST 是一个基因的部分序列, 所以它所揭示的基因组信息不全, 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。④对 cDNA 文库质量有一定的要求。高、中表达丰度基因的 EST 存在冗余性, 重复测序造成人力物力的巨大浪费, 且表达丰度较低的稀有基因容易疏漏。

5 展望

人类基因组及一些模式生物基因组的测序工作陆续完成, 预示着以测序为基础的基因组时代的结

束, 后基因组时代的到来, EST 数据已广泛应用于微阵列分析、蛋白质组及基因组各项生物学研究^[23], 进而发现抗逆相关基因或者利用有利基因进行品种改良等, 是新时代里生物科学工作者所面临的新的巨大挑战。因此 EST 技术必将成为一种强有力的工具, 帮助我们揭示基因组所包含的信息, 使基因组研究进入一个全新的阶段。

参考文献:

- [1] AdamsMD, KelleyJM, GocayneJD, et al.. Complementary DNA sequencing: expressed sequence tags and human genome project[J]. Science, 1991, 252: 1651-11656.
- [2] Stephen Rudd. Expressed sequence tags: alternative or complement to whole genome sequence[J]. TRENDS in Plant Science 2003, 8(7): 321; 329
- [3] LuoMeng, JiajiZeng. Progress in Expressed Sequence Tags (EST) Project of Plant genome[J]. Prog. Biochem. Biophys. 2001, 28(4): 494-497
- [4] GilpinBJ, McallumTA, FrewTJ. A linkage map of the pea (Pisum sativum L.) genome containing cloned sequences of known function and expressed sequence tags (EST)[J]. Theor Appl Genet, 1997, 95(8): 1289-1299
- [5] HanushimaY, YanoM, ShomuraA, et al.. A high density rice genetic linkage map with 2275 markers using a single F₂ population[J]. Genetics 1998, 148(1): 1-16
- [6] SasakitT. The rice genome project in japan[J]. Proc Natl Acad Sci USA, 1998, 95(5): 2027-2028.
- [7] 万海伟, 杜立新. 表达序列标签(EST)在基因组学研究中的应用[J]. 生物技术通报, 2004, (1): 35-38
- [8] 李红. 胃癌相关 cDNA 片段的快速克隆和表达分析[J]. 生物化学与生物物理进展, 2002, 29(4): 604-609
- [9] 高瑞娟, 戴大鹏, 马荣才等. 结球白菜结球前期基因表达序列标签(EST)分析[J]. 农业生物技术学报, 2004; 12(1): 24-29
- [10] LiuJY, HaraC, UmedaM, et al.. Analysis of randomly isolated cDNAs from developing endosperm of rice: evaluation of expressed sequence tags and expression levels of mRNAs[J]. Plant Molecular Biology, 1995, 29(4): 685-689
- [11] EppleP, ApelK, Bohlmann H. ESTs reveal multigene family for plant defensins in Arabidopsis thaliana[J]. FEBS Letters, 1997, 400(2): 168-192
- [12] SterkyF, RegagnS, KarlssonJ, et al.. Gene discovery in the wood forming tissue of poplar: Analysis of 5692 expressed sequence tags[J]. Proc Natl Acad Sci USA, 1998, 95(22): 1333-1335
- [13] 骆蒙, 孔秀英, 霍纳新等. 小麦抗白粉病侵染初期的表达序列标签分析[J]. 遗传学报, 2002, 29(6): 525-530
- [14] GreenP, LipmanD, HittierL, et al.. Ancient conserved region in new gene sequence and protein database[J]. Science 1996, 259(5102): 1711-1716
- [15] 张正斌, 徐萍. 小麦基因组研究[J]. 遗传 HEREDITAS(Beijing), 2002, 24(3): 389-394
- [16] 梁国栋. 最新分子生物学实验技术[M]. 北京: 科学出版社,

大棚豆角延后丰产栽培技术要点

黄 英

(黑龙江省庆安县农业技术推广中心, 庆安 152400)

利用大棚秋季栽培蔬菜, 可大幅度延长蔬菜供应期, 增加淡季蔬菜品种, 既解决了淡季供应问题, 也提高了蔬菜单位面积产量, 增加了农民的经济效益。在这里介绍一下大棚秋季栽培豆角的技术。该项技术经过几年的试验表明, 产量可达 3 000 kg/667m², 销售价格平均 1.6 元/kg, 收益达 4 800 元/667m², 去掉人工费、病虫害防治费等各项费用, 纯收入 4 300 元/667m²。现将其栽培技术介绍如下:

1 品种选择

选择蔓生、产量高、品质好的早熟、中早熟品种。我县采用的品种为紫花油豆。

2 整地施肥

前茬作物拉秧后, 立即拔除植株, 将土壤深翻 15~20 cm, 晒 3~5 d。施农家肥 4 000~5 000 kg/667m², 磷酸二铵 10 kg/667m², 均匀地撒到地表, 然后起垄。小垄 60~70 cm, 大垄 90~120 cm, 打碎坷垃以待播种。

3 适时播种, 合理密植

3.1 选种

选豆粒大小整齐饱满, 颜色一致而有光泽, 机械损伤, 无虫孔, 无病斑的种子。

3.2 浸种催芽

将选好的干种子放到 20℃左右温水中浸 3~4 h, 然后捞出放在干净的盆等容器中, 上面盖上一层湿布, 放在 26℃左右条件下催芽, 每天用 20℃左右温水投洗一次, 2 d 内即可出芽。

3.3 播种时间

直播 7 月 10 日左右。如果前茬倒不出来, 可进行营养钵育苗, 每钵点种 3~4 粒, 2 片真叶定苗, 每垄 2 株壮苗, 育苗播种时间为 7 月 5 日左右, 7 月末移栽。

3.4 合理密植

60~70 cm 的小垄垅距 30 cm, 保苗株数为 3 600 株/667m²; 90~120 cm 的大垄垅距 35 cm, 垄上双行, 保苗株数为 3 800~4 000 株/667m²。

4 棚内管理

总的管理原则是生育前期温度较高, 条件适宜, 植株生长较快, 防止徒长, 控制浇水; 后期温度偏低, 停止浇水施肥, 注意防寒保温, 尽量延长收获期。

4.1 温度管理

从出苗到开花, 白天温度保持在 22~25℃, 夜间保持在 10~15℃。生育后期随气温变化逐渐减少通风或不通风, 提高棚内温度。为了延长生长期四周围上草帘。

4.2 肥水管理

由于是座水淹种, 所以缓苗后到开花结荚前不用浇水, 要严格控制水分, 否则会引起徒长而落花。当幼荚座住后浇一次水并结合浇水追施一次腐熟好的有机肥。如果没有有机肥, 可追尿素 5 kg/667m², 磷酸二氢钾 0.5 kg/667m², 到蔓爬满架的盛产期, 可用 0.01%~0.03% 的铜酸铵或硫酸铜进行根外追肥, 促使早熟和提高产量。

4.3 适时支架

当蔓抽出 30 cm 时, 应进行支架或吊蔓。双行密植的将两行并拢成“人”字架, 在吊蔓时要顺着植株长势进行缠绕。

5 病害防治

5.1 锈病

发病初期及时喷药防治。常用药剂有: 50% 萎锈灵可湿性粉剂 1 000 倍液, 50% 多菌灵可湿性粉 800~1 000 倍液, 每隔 7~10 d 喷药一次, 喷 2 次即可。

5.2 害虫

主要有蚜虫, 蚜虫可用 10% 的吡虫啉可湿性粉剂 1 000~1 500 倍液喷雾防治。

* 收稿日期: 2006-04-03

作者简介: 黄英(1972-), 女, 黑龙江省庆安县人, 农艺师, 从事农业技术推广工作 0455-2869541。

2001.366-370.

[17] Ji W, Li Y, Li J, et al. Generation and analysis of expressed sequence tags from NaCl-treated Glycine soja[J]. BMC Plant Biol, 2006, 22(6): 4.

[18] A. -G. Tian, J. Wang. Characterization of soybean genomic features by analysis of its expressed sequence tags[J]. Theor Appl Genet, 2004, 108(5): 903-913

[19] Nadim W. Alkharof, Benjamin F. Mtthew s SGMD the Soybean Genomics and Microarray Database[J]. Nucleic Acids Research, 2004, 32: 398-400

[20] Alkharouf N, Khan R, Matthews B. Analysis of expressed

sequence tags from roots of resistant soybean infected by the soybean cyst nematode[J]. Pest Manag Sci, 2003, 59(6-7): 748-53.

[21] Daguang Cai, Michael Kleine, et al. Positional Cloning of a Gene for Nematode Resistance in Sugar Beet[J]. science, 1997, 275(7): 832-834

[22] GeyerH, SchmittS, WuhrerM, GeyerR. Structural analysis of glycoconjugates by on-target enzymatic digestion and MALDI-TOF-MS[J]. Anal Chem, 1999, 71: 476-482.

[23] Shafer P, Lin DM, Yona G. Mapping EST sequences to proteins[J]. BMC Genomics, 2006, 7(1): 41