

密级：

论文编号：

# 中国农业科学院 学位论文

## 茶树 cDNA 文库的构建及新梢特异表达基 因 EST 序列分析

Construction of cDNA library of tea plant (*Camellia sinensis*)  
and EST sequences analysis of leaf specific genes

研 究 生：赵丽萍

指 导 教 师：陈 亮

合作指导教师：浙江大学 高其康 研究员

申请学位类别：硕 士

专业领域名称：茶 学

研 究 方 向：分子生物学

培 养 单 位：茶叶研究所

提交日期 2004 年 6 月

# 独 创 性 声 明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知,除了文中特别加以标注和致谢的地方外,论文中不包含其他人已经发表或撰写过的研究成果,也不包含为获得中国农业科学院或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

研究生签名: 赵丽萍

时间: 2004 年 6 月 8 日

# 关于论文使用授权的声明

本人完全了解中国农业科学院有关保留、使用学位论文的规定,即:中国农业科学院有权保留送交论文的复印件和磁盘,允许论文被查阅和借阅,可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。同意中国农业科学院可以用不同方式在不同媒体上发表、传播学位论文的全部或部分内容。(保密的学位论文在解密后应遵守此协议)

研究生签名: 赵丽萍

时间: 2004 年 6 月 8 日

导师签名: 陈亮

时间: 2004 年 6 月 8 日

**本研究受**

**国家自然科学基金项目资助**

**(批准号为 30070486)**

单位代码：82101  
研究生学号：010002

## 茶树 cDNA 文库的构建及新梢特异表达基因 EST 序 列分析

论文答辩日期：2003 年 6 月 22 日

---

答辩委员会主席：杨亚军 研究员（博导）中国农科院茶叶研究所

答辩委员会成员：梁月荣 教授（博导）浙江大学

---

陈学新 教授（博导）浙江大学

---

成浩 研究员（博导）中国农科院茶叶研究所

---

江和源 副研究员（硕导）中国农科院茶叶研究所

---

论文评阅人：梁月荣 教授（博导）浙江大学

---

林智 副研究员（硕导）中国农科院茶叶研究所

---

# 目录

缩略词 .....	I
摘 要 .....	II
Abstract .....	III
第一章 文献综述与研究思路 .....	1
1.1 茶树功能基因克隆研究进展 .....	1
1.1.1 总 RNA 提取及 mRNA 分离 .....	1
1.1.2 茶树功能基因的分离与克隆 .....	2
1.2 EST 技术及其在功能基因组研究上的应用 .....	5
1.2.1 EST 及其研究概况 .....	5
1.2.2 EST 技术的原理和方法 .....	6
1.2.3 EST 的应用 .....	6
1.2.4 生物信息学在 EST 技术中的应用 .....	8
1.3 本研究的目的意义和思路 .....	9
第二章 茶树新梢 cDNA 文库的建立 .....	10
2.1 材料与方法 .....	10
2.1.1 实验材料 .....	10
2.1.2 主要试剂和常用溶液 .....	10
2.1.3 实验方法 .....	11
2.2 结果与分析 .....	15
2.2.1 RNA 质量 .....	15
2.2.2 cDNA 文库的质量 .....	16
2.3 讨论 .....	17
第三章 茶树新梢 EST 序列与功能基因分析 .....	19
3.1 材料与方法 .....	19
3.1.1 材料 .....	19
3.1.2 方法 .....	19
3.2 结果与分析 .....	23
3.2.1 新构建 <i>Sfi</i> <sup>+</sup> -pUC 载体的质量 .....	23
3.2.2 大规模质粒提取纯化及 EST 测序结果 .....	24
3.2.3 序列提交 .....	29
3.2.4 EST 序列的同一性分析 .....	29
3.2.5 与 NCBI 数据库的 BLAST 比对结果 .....	29
3.2.6 茶树 ESTs 比对上的物种分布 .....	30
3.2.7 高丰度表达的基因 .....	34
3.2.8 BlastX 比对具已知功能和具推测功能 EST 的分析 .....	34
3.3 讨论 .....	37
3.3.1 新建载体的质量 .....	37
3.3.2 EST 的 Poly(A) 特性 .....	38
3.3.3 BlastN 与 BlastX 比对结果的差异 .....	38
3.3.4 EST 测序分析的效率 .....	39
3.3.5 新梢表达基因的丰度特征 .....	39
3.3.6 新梢的主要功能基因分类 .....	39
第四章 总结与展望 .....	41
4.1 建立了 2 个高质量的 cDNA 噬菌体文库 .....	41

4.2	构建了 <i>Sfi</i> <sup>+</sup> -pUCM-T 载体 .....	41
4.3	文库测序及 EST 分析 .....	41
4.4	茶树新梢的功能基因 .....	42
4.5	今后研究方向 .....	42
参考文献 .....		43
致 谢 .....		50
附件 1 EST 序列提交格式.....		51
附件 2 GenBank 登录的基因.....		53
作者简介 .....		62

# 缩略词

## Abbreviations

英文缩写	英文全称	中文名称
Amp	Ampicillin	氨卞青霉素
ATP	Adenosine-5'-triphosphate	腺苷三磷酸
bp	Base pair	碱基对
BSA	Bovine serum albumin	牛血清白蛋白
Bt	Bacillus thuringiensis	苏云金芽胞杆菌
DEPC	Diethyl pyrocarbonate	焦碳酸二乙酯
DMSO	Dimethylsulfoxide	二甲基亚砷
EB	Ethidium bromide	溴化乙锭
EST	Expressed Sequence Tags	表达序列标签
IPTG	Isopropyl thiogalactoside	异丙基硫代半乳糖苷
LB	Luria-Bertani	一种培养基
OD	Optical density	光密度
PCR	Polymerase chain reaction	聚合酶链反应
X-Gal	Galactose	5-溴-4-氢-3-吡啶-β-D-半乳糖苷

# 摘要

茶树是一种重要的经济作物。茶叶是世界最流行的饮料之一,对人体有重要的生理保健作用。为了更好了解茶树新梢表达基因,并筛选出感兴趣的目的基因,本文构建了龙井 43 和安吉白茶等两个品种的新梢 cDNA 文库,并对龙井 43 文库进行较高通量的 EST 测序和生物信息学分析。本研究获得的主要结果如下:

1. 以 Trizol 一步法从春茶新梢中提取总 RNA,分离纯化富含 Poly A 的 mRNA,LD-PCR 反转录合成双链 cDNA,以  $\lambda$ TriplEx2 为载体,构建了龙井 43 新梢 cDNA 文库。文库的滴度分别为  $6.8 \times 10^5$  pfu/mL 和  $4 \times 10^5$  pfu/mL,符合高质量文库的要求。文库的重组率为 98.05% 和 96.15%,且插入片段的长度在 1.0-1.5 kb 左右,说明文库是完整的、有效的。两个文库总克隆数分别为  $3.5 \times 10^5$  个和  $2.1 \times 10^5$  个,高于所需标准。
2. 为了获得高质量的 EST 序列,利用原来  $\lambda$ TriplEx2 载体获得的质粒,在 pUCM-T 载体上引入 2 个 *Sfi*I 酶切位点后,构建了一个新载体 *Sfi*I-pUC。用新载体构建一个用于测序的质粒文库。
3. 从龙井 43 新梢 cDNA 质粒文库中随机选取 4320 个克隆进行测序,根据 EST 标准获得长度大于 200bp 的序列 2963 个,占总克隆的 68.60%。通过 DNATool 软件进行冗余度分析后去除的重复序列有 863 个,占总获得序列的 33.88%,获得有效 ESTs 1684 个,其平均长度为 478bp。
4. 所有有效序列与 NCBI 的核酸数据库比对,有 36.0% 序列 (607EST) 序列显示了与已知序列高度的同源性;而与蛋白质数据库进行比对时有 71.3% 的序列 (1200 EST) 显示了与已知序列的高度的同源性,并建立了茶树新梢特异表达基因数据库。
5. 有效序列 BlastX 分析表明,丰度较高的基因(表达频率=5)约有 30 个,约占总数的 10%;中等丰度基因(表达频率在 2-5 之间)约 100 个,占总数的 30%;其余低丰度的基因约占 60%。说明在茶树新梢中基因大多数呈中低丰度表达。
6. 结合拟南芥功能基因分类标准,把通过 BlastX 比对获得茶树新梢有功能描述基因(包括功能确定的及推测功能的)分为 12 类,即能量代谢相关基因(Energy),占功能基因总数的 26.39%、蛋白质合成相关基因(Protein synthesis)占 17.86%、细胞结构相关基因(Cell structure)占 10.71%、细胞抗性及其防御相关基因(Disease and defence)占 9.23%、代谢相关基因(Metabolism)占 8.33%、信号转导相关基因(Cellular signal transduction)占 5.36%、细胞内运输相关基因(Cellular transporters)占 4.66%、转录相关基因(Transcription)占 3.87%、细胞生长发育相关基因(Cell growth and division)占 3.77%、次生代谢相关基因(Secondary metabolism)占 3.47%、蛋白质降解相关基因(Protein destination and storage)占 1.69% 以及功能不明确基因及未分类的基因(Unclear classification and unclassified)占 4.66%。

关键词: 茶树, 新梢, cDNA 文库, 表达序列标签, 功能基因



# Abstract

Tea plant (*Camellia sinensis*) is an important economic crop and tea is one of the most popular beverages in the world. In order to better understand the gene expression profile of tea tender shoots and obtain the interested genes, two tender shoots cDNA libraries of Longjing 43 and Anji Baicha were constructed, and Longjing 43 library was subjected to high-throughout EST sequencing and bioinformatical analysis. The main results were as follow:

1. Total RNA was isolated from tender tea shoots using Trizol single-step method, mRNA was separated and then the double-strand cDNA was amplified by LD-PCR. After size fractionation, the ds-cDNA was ligated to  $\lambda$ TriplEx2 and recombinant bacteriophages were packaged to construct cDNA library. The titers of Longjing 43 and Anji Baicha cDNA libraries were  $6.8 \times 10^5$  and  $4 \times 10^5$  pfu/mL, respectively, matching the requirement of high quality library. The recombinant rates were 98.05% and 96.15%, respectively, and the inserts were about 1.0-1.5kb. There were  $3.5 \times 10^5$  and  $2.1 \times 10^5$  clones in total, respectively. The results indicated that the two libraries were integrated and efficacious.
2. A new vector *Sfi* $\lambda$ -pUC was constructed in order to obtain high EST sequencing quality based on the original vector pTriplEx2. Two new restriction sites of enzyme *Sfi*I were introduced into the new vector.
3. Totally 4320 clones were randomly selected from Longjing 43 cDNA library and partially sequenced by MegaBace 1000 capillary sequence machine, 2963 readable sequences that are more than 200 bp were produced from them, which was 68.60% of all the sequences. From the 2963 useful sequences, 1684 ESTs were identified by redundancy analysis with DNATool software. The redundant ratio was about 33.88%, and the average length of the 1684 ESTs was 478 bp.
4. The 1684 ESTs were blasted against the non-redundant nucleotide and protein database of NCBI and the databases of tea plant ESTs were constructed. On the basis of NCBI database searches, 36.0% of the total 1684 ESTs showed significant homology with the nucleotide database and 71.3% with the protein database.
5. BlastX analysis of significant homology with the non-redundant protein database suggested that about 10% of the total sequences were high expression abundance genes (expression frequency=5), 30% were moderate abundance genes (expression frequency scope 2-5), and 60% were low abundance genes (expression frequency<2). The results indicated that most genes in tea tender shoots were low or moderate expression abundance.
6. Genes with function annotation or putative function annotation were assigned to 12 categories according to the classification of *Arabidopsis* genes. They were genes related to energy (corresponding 26.39% of total genes), protein synthesis and storage (17.86%), cell

structure (10.71%), cell disease and defence (9.23%), metabolism (8.33%), cellular signal transduction (5.36%), cellular transporters (4.66%), transcription (3.87%), cell growth and division (3.77%), secondary metabolism (3.47%), protein destination and storage (1.69%), unclear classification and unclassified(4.66%).

**Key words:** Tea plant (*Camellia sinensis*), Tender shoots, cDNA library, Sequencing analysis, EST (expressed sequence tags), Functional genes

# 第一章 文献综述与研究思路

茶树是多年生木本经济作物,在我国栽培已有两千多年的历史,从最初的药用到现在变成流通于全球的大众化饮料已经成为一项对地方经济很可观的收入来源。目前,茶叶已成为世界公认的三大无酒精饮料之一。现代研究表明,茶叶对人体有重要的生理保健作用,如防龋齿、抗菌抗病毒抗氧化、降血压降血脂降胆固醇、抗癌抗突变等等(Yamamoto et al., 1998)。因此,从茶叶中提取有效成分加工制成医药和保健品已成为茶树综合利用的重要方向。而品种对茶叶的质量至关重要,随着茶叶市场向多元化发展,优良茶树品种供不应求。因而,采用生物技术改良或培育新一代茶树品种已成为国内外茶学研究的热点之一。

现代分子生物学技术发展十分迅速,现在已经有很多的作物建立了比较完整的分子生物学的实验体系,而茶树由于其自身特性的限制,在分子生物学尤其在功能基因组研究方面一直处于落后状态。目前为止,只有 Matsumoto、Takeuchi 等(1994) 构建了一个茶树的表达基因文库(cDNA 文库)并利用此文库获得了一些茶树上的重要基因(Takeuchi et al., 1994; Matsumoto et al., 1994; 竹内敦子等, 1994),其它已经克隆的茶树基因(表 1-1)都是通过经典基因克隆方法获得,但对茶树功能基因组研究还几乎处于空白状态。而表达序列标签(Expressed Sequence Tag, EST)技术(Adams et al., 1991)在生物功能基因组研究中发挥了巨大的作用。EST 技术已被广泛应用于新基因的克隆、组织特异性基因表达谱分析以及基因组序列的功能注释等许多研究领域。EST 测序分析成为大规模获取功能基因一种最为快捷和有效的研究手段(Delseny et al., 1997; Ablett et al., 2000),已有大量的成功经验。

因此,借鉴其它生物成功的经验,通过建立茶树表达基因文库,开展 EST 测序分析,可能在短期内获得大量的茶树功能基因表达信息,对迅速提升茶树功能基因研究水平,从基因组水平认识及从分子水平调控茶树的生长、发育、抗性和代谢规律都具有理论与实践意义。

## 1.1 茶树功能基因克隆研究进展

### 1.1.1 总 RNA 提取及 mRNA 分离

RNA 提取是一项植物分子生物学基本内容和技术。完整 RNA 的提取和纯化是进行 RNA 方面的研究工作,如 Northern 杂交、mRNA 分离、cDNA 合成及体外翻译等的前提。目前已经建立了多种 RNA 提取方法。实验室小规模提取植物总 RNA 最常用的是 Chomczynski 等(1987)提出的异硫氰酸胍-酚-氯仿一步法。此方法采用了在酸性条件下 RNA 比 DNA 更容易进入水相的原理,能较完全地去除 DNA,因为此方法使用了 RNA 的强抑制剂, RNA 被降解的可能性较小,对大多数的动植物都能得到质量很好的 RNA 样品,但对某些富含多酚、多糖等次生物质的植物却不适用。茶树幼嫩新梢中含有茶多酚,这种物质被氧化后会与 RNA 不可逆地结合,导致 RNA 活性丧失而且在用苯酚、氯仿抽提时 RNA 易丢失(Schneiderbauer et al., 1991),或形成不溶性的复合物(Graham, 1993)。目前去除酚类化合物的一般途径是在提取的初始阶段防止其被氧化,然后再将其与 RNA 分开(李宏等 1999)。防止酚类化合物被

氧化的方法主要有：还原剂法，在提取液中加入  $\beta$ -巯基乙醇、二硫苏糖醇或半胱氨酸。

螯合剂法，加入聚乙烯吡咯烷酮(PVP)，用 PVP 去除多酚时 pH 值是一个重要的影响因素，在 pH8.0 以上时 PVP 结合多酚的能力会下降。Tris-硼酸法，硼酸可以与酚类化合物结合形成复合物(Tris-硼酸的浓度不能超过 0.2 M，否则影响 RNA 的回收率)。另外还有牛血清白蛋白法(BSA)(Wang et al., 1994)及丙酮法等(Schneiderbauer et al., 1991)。Guillemant 等和 Ainsworth 认为低 pH(pH5.5)的提取液可以防止酚类化合物的离子化和进一步氧化。江昌俊等(2001)在克氏一步法分离总 RNA 的基础上，根据茶树叶片生化成分特点，做了部分修改，先使组织粉末中的 RNA 溶进变性溶液，随后不立即加入饱和苯酚和其它试剂，而是经离心收集含有 RNA 的上清液，及早除去组织碎片中影响 RNA 纯度和完整性的物质，防止 RNA 的降解。不过，Newbury 等(1977)发现 RNA 的提取的难易程度与材料中酚类物质的总量之间并无相关性，因此认为不是所有的酚类化合物都影响 RNA 的提取，但一般认为所谓的“缩合鞣质”即聚合多羟基黄酮醇类物质是影响 RNA 提取的一类化合物。

### 1.1.2 茶树功能基因的分离与克隆

茶树功能基因的克隆起步较晚，开始于 1990 年代，截止 2004 年 4 月 30 日，从 NCBI 核酸数据库获得的信息中，共有 20 多个基因被克隆(表 1-1)。主要通过三种途径。第一是利用其它作物中已知序列的克隆与 cDNA 文库直接杂交，筛选出与外来序列高度同源的 cDNA 克隆，如查耳酮合酶(Chalcone Synthase, CHS)基因(Takeuchi et al., 1994)、苯丙氨酸解氨酶(Phenylalanine Ammonia-lyase, PAL)基因(Matsumoto et al., 1994)、 $\beta$ -微管蛋白 cDNA 的克隆(竹内敦子等，1994)；第二是根据各种不同植物基因序列的同源性，利用从其它植物中获得相同基因的保守序列，以 RNA 或 cDNA 为模板，设计简并引物，进行 PCR 扩增获得目的基因，如多酚氧化酶(Polyphenol Oxidase, PPO)(赵东等，2001)，S-腺苷甲硫氨酸合成酶(S-Adenosylmethionine Synthetase, SAM)(冯艳飞和梁月荣，2001；Liang et al., 2001)；第三是利用已分离纯化酶蛋白质的氨基酸序列获得 cDNA 克隆，如咖啡碱合成酶(Caffeine Synthase, CS)(Kato, 2001)等。这些基因主要是与茶树次生代谢产物相关的酶类，如茶多酚代谢的关键酶，与咖啡碱合成有关酶以及一些影响茶叶品质的酶类。

#### 1.1.2.1 茶多酚代谢关键酶基因的克隆

茶多酚和儿茶素代谢关键酶包括苯丙氨酸解氨酶、查耳酮合酶和二氢黄酮醇 4-还原酶等。苯丙氨酸解氨酶(PAL)是微管植物普遍存在的次生代谢关键酶类，它水解苯丙氨酸产生类黄酮类合成的最初底物；查耳酮合酶(CHS)是植物类黄酮生物合成中的关键酶；而二氢黄酮醇 4-还原酶(Dihydroflavonol 4-Reductase, DFR)在儿茶素合成的后期起重要作用。PAL、CHS 和 DFR 是茶树黄酮类化合物生物合成中的重要酶类，调控着整个代谢过程，这三种酶都是利用其它作物中已知序列的克隆与茶树 cDNA 文库杂交，筛选出与外来序列高度同源的茶树 cDNA 克隆。Matsumoto、Takeuchi 等(1994)构建了薮北茶树品种(日本栽培面积最

表 1-1 在 GenBank 登录已克隆的茶树基因<sup>a</sup>

Table 1-1 Genes cloned from tea plant from Genbank

基因名称 Gene	登录号 GenBank accession	长度(bp) Length	作者 Author	年份 Year
谷氨酸盐合成酶 glutamine synthetase	AB117934	1409	Tanaka et al	2004
16S 核糖体 RNA 16S ribosomal RNA	AY563527	699	Singh et al	2004
26S 核糖体 RNA 26S ribosomal RNA	AY283368	550	Dadwal et al	2003
18S 核糖体 RNA 18S ribosomal RNA	AB120309	1800	Tanaka et al	2003
β-樱草苷酶 beta-primeverosidase	AB088027	1729	Ijima et al	2003
NADH 脱氢酶 NADH dehydrogenase (ndhF)	AF421077	1946	Anderberg et al	2002
ATP 合成酶 β 亚基 ATP synthase β subunit	AF420986	1184	Anderberg et al	2002
ATP 合成酶 α 亚基 ATP synthase α subunit	AF420952	1200	Anderberg et al	2002
紫黄素脱环氧化酶 violaxanthin de-epoxidase (vde)	AF462269	1562	Wei et al	2002
(似)β-1,3 葡萄糖苷酶 beta-1,3-glucosidase-like	AF537127	584	Jiang et al	2002
β-1,3 葡聚糖酶 beta-1,3-glucanase	AF399920	1369	Jiang et al	2001
与咖啡因合成有关 n-甲基转移酶 n-methyltransferase	AX138777	1427	Mizuno et al	2001
咖啡因合成酶 caffeine synthase	AF031280	1438	Kato et al	1999
S-腺苷蛋氨酸合成酶 S-adenosylmethionine synthase	AB041534	1303	Feng et al	2000
S-腺苷蛋氨酸合成酶 S-adenosylmethionine synthase	AJ277206	1776	Feng et al	2000
多酚氧化酶 polyphenol oxidase	AF269192	537	Xi et al	2000
5.8S 核糖体 RNA 5.8S ribosomal RNA	AF315492	653	Tang et al	2000
类 PR-1 蛋白 PR-1 like protein	AB015047	707	Tomimoto et al	1998
二氢黄酮醇 4-还原酶 dihydroflavonol 4-reductase	AB018685	1420	Takeuchi et al	1998
二氢黄酮醇 4-还原酶 dihydroflavonol 4-reductase	AB018686	1268	Takeuchi et al	1998
查耳酮合成酶 1 chalcone synthase 1	D26593	1390	Takeuchi et al	1994
查耳酮合成酶 2 chalcone synthase 2	D26594	1405	Takeuchi et al	1994
查耳酮合成酶 3 chalcone synthase 3	D26595	1425	Takeuchi et al	1994
苯丙氨酸解氨酶 phenylalanine ammonia-lyase (PAL)	D26596	2330	Matsumoto et al	1994
1,5-二磷酸核酮糖羧化酶 ribulose-1,5-biphosphate-carboxylase	X69732	900	Savolanien et al	1992

<sup>a</sup> 不包含本研究所获得的基因。Except for the genes obtained by this research.

多的)第一个 cDNA 文库，为茶树功能基因的分离克隆奠定了基础。Matsumoto 等(1994)首先应用标记的水稻 PAL cDNA 为探针，从薮北 cDNA 文库获得茶树 PAL cDNA 的克隆；并对茶树进行 PAL-RFLP 分析，检测茶树 PAL 基因的变异；用包括 3' 非翻译区在内的 280 bp 的 PAL 探针检出了 3 个不同长度的 DNA 片段 A、B 和 D。通过对 PAL 基因的遗传规律研究，表明 3 个不同片段的遗传符合孟德尔单基因遗传定律，PAL 基因是一个复等位基因。Takeuchi 等(1994)用欧芹 CHS cDNA 为探针进行文库筛选，获得了 3 个全长 cDNA 即 CHS1、CHS2 和 CHS3，说明了 CHS 是一个多基因家族，每个 cDNA 编码 389 个氨基酸，它们之间有 93%-96% 的同源性。以 5' 非翻译区序列为基础合成寡聚核苷酸片段，其转录产物可被区分，

并且在幼叶和茎中含量丰富。Takeuchi 等(1995)还报道了避光处理遮荫条件下, PAL 和 CHS 的 mRNA 水平均明显降低, 表明 PAL 和 CHS 基因在光照下表达增加, 遮荫下表达减少。PAL、CHS 和 DFR 的克隆为应用基因工程手段调控茶树多酚类物质代谢创造了条件。

#### 1.1.2.2 多酚氧化酶基因的克隆

植物多酚氧化酶(PPO)是含铜的金属蛋白, 由核基因编码, 可能仅存在于质体中。虽然从 1895 年就开始有多酚氧化酶的研究报导(Mayer et al., 1979, 1987), 但其在植物组织中的功能并未充分了解。随着分子生物学的发展, 多酚氧化酶研究也进入了分子阶段。PPO 对茶叶的品质影响很大, 尤其是不同茶类要求不同的多酚氧化酶活性。通过传统的方法难以对酶活性进行有效的调控, 如果能通过基因工程来调控多酚氧化酶的活性, 则会有重要的理论和实际意义。赵东等(2001)根据已经发表的植物多酚氧化酶中的保守序列区, 设计简并引物, 进行 nest-PCR, 将产物与载体 pGEM-T Easy Vector 连接、转化, 得到了茶树多酚氧化酶基因的部分序列, 其长度为 1006 bp, 将核苷酸序列按三种不同的起始位置翻译成氨基酸序列, 经过比较, 最后确定正确的氨基酸序列, 其长度为 335 个氨基酸残基。并将其与已经登录于 GenBank 中植物 PPO 序列进行多重比较, 发现茶树多酚氧化酶基因与其它植物多酚氧化酶基因有较高的相似性, 尤其在铜离子结合部位, 所有的植物基本上一致。他们还进行了进化树分析, 结果表明茶树多酚氧化酶可以与大多数的木本植物聚合成一个组。

#### 1.1.2.3 与咖啡碱合成有关酶基因的克隆

目前已经克隆了 3 个与咖啡碱合成有关酶的基因, 即 S-腺苷甲硫氨酸合成酶(SAM)、咖啡碱合成酶(CS)以及与咖啡碱合成有关的 N-甲基转移酶基因。SAM 是茶树咖啡碱生物合成过程中甲基化反应的唯一甲基供体, 还是植物体内转甲基反应的甲基供体及多胺和乙烯合成的前体, 它是由 S-腺苷甲硫氨酸合成酶催化甲硫氨酸和 ATP 反应生成的, 因而对 SAM 合成酶的研究在植物逆境生理、衰老生理和茶树咖啡碱生物代谢及其调控研究上均具有重要意义。冯艳飞和梁月荣(2001)从茶叶总 RNA 反转录合成了 cDNA, 并以此为模板分别用三对 PCR 引物扩增 SAM 合成酶基因的中间主片段、3'端和 5'端片段。最后经 BLAST 比较及重叠区域拼接得到完整的 SAM 基因序列。所得序列长为 1303 bp, 编码 394 个氨基酸。并通过比较证明与其它植物的 SAM 基因有较高的同源性。Liang 等(2001)以薮北为材料提取总 RNA, 根据已知其它作物 SAM 基因的保守序列区设计上游和下游引物, 用 RT-PCR 进行 cDNA 扩增, 得到茶树 SAM 基因 cDNA 序列, 并确定此序列为 800bp, 而且实验证明 SAM 基因只在茶树的叶片中存在, 根部无此基因。

Kato(2001)以茶树嫩叶为材料, 提取总 RNA, 并获得了 cDNA, 利用 RACE (Rapid Amplification of cDNA End)技术以咖啡碱合酶的 N 端氨基酸序列设计简并特异引物, PCR 扩增获得了 CS 的 cDNA, 测得序列为 1438 bp, 编码 369 个氨基酸。但获得的 CS 基因与咖啡中此基因仅有 40%同源性。用重组的 CS 在大肠杆菌中表达, 证明 CS1 是编码咖啡碱合成酶的基因, 并得出 CS 在嫩叶中的含量很高, 但在老叶中的含量很低。

#### 1.1.2.4 其它基因的克隆

微管是构成真核细胞的骨架之一, 微管蛋白是微管的组成蛋白质, 是真核细胞内普遍存

在的蛋白质，由  $\alpha$  和  $\beta$  两个亚基组成。微管在细胞分裂、物质运输、细胞极性及生长的控制等许多生理活动中起着极其重要的作用，微管还可能参与控制细胞壁微纤丝的沉积方向，从而控制细胞的生长，并进一步控制整体植物的生长发育。有些高等植物的微管蛋白基因已经克隆出来，并测定了核酸序列。竹内敦子等(1994)首次从薮北 cDNA 文库中，用  $\beta$ -微管蛋白特异引物进行 PCR 扩增克隆到一个 505 bp 不含反义引物序列的  $\beta$ -微管蛋白 cDNA，其推导的氨基酸序列与其它高等植物有很高的同源性。 $\beta$ -1,3 葡聚糖酶、(似) $\beta$ -1,3 葡萄糖苷酶是属于水解酶类，该酶与萜烯类香气前体有密切关系，能使糖苷键合态变成游离态，还能水解野黑樱苷，释放 HCN，对植物起到一定的抗病虫害作用，它与  $\beta$  樱草苷酶都是与茶叶香气形成和释放过程相关的重要酶类。5.8S、16S、26S 核糖体 RNA 等是构成核糖体的重要组分，在研究植物起源、进化中有重要作用；1,5-二磷酸核酮糖羧化酶是卡尔文循环途径中催化二氧化碳固定的酶，是植物体内的能量代谢中的重要酶类。

所有这些茶树基因的克隆，都为从分子水平上更好地了解、研究、调控茶树的生理代谢、生长发育和品质等提供了帮助。

## 1.2 EST 技术及其在功能基因组研究上的应用

### 1.2.1 EST 及其研究概况

经典基因克隆和功能研究方法是以基因功能为基础，缺点是费时费力且一次只能克隆到一个或者少数几个与功能相关的基因。Costanzo(1983)提出表达序列标签(Expressed sequence tag, EST)概念的雏形，他们随机对肝脏 cDNA 文库测序，证实所测序列可用于研究 DNA 序列与基因功能之间的关系。Wilcox(1991)和 Okubo(1992)几乎同时描述了 EST 技术。关于 EST 技术应用的首次报道是 Adams(1991)等从三种人脑组织 cDNA 文库中随机挑取 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。Craig、Venter 等为 EST 技术的发展进行了多年不懈的努力，Craig 所在的美国国立卫生研究院(National Institutes of Health, NIH)运用了自动化测序技术，大规模工业化生产 EST 序列，并将 EST 技术从功能基因组学(Functional Genomics)扩展到结构基因组学(Structural Genomics)，直接推动了人类基因组计划(Human Genome Project, HGP)的超前完成。实际上，最初人类基因表达图谱的构建是从以 EST 测序方法为基础来获得大量 EST 开始的(Adams et al., 1991)。

人类基因组计划使生物基因组测序进入了高速发展时期。在此带动下，一些模式生物如拟南芥、大肠杆菌、酵母菌、线虫等，经济作物如玉米、水稻、小麦、棉花等以及畜禽如猪、鸡、牛等的基因组计划也相继启动或完成。1992 年开始的每年在美国的圣地亚哥召开国际基因组大会(Plant and Animal Genome Conference)，大量的研究成果都显示了基因组研究已成为国际上最前沿的领域。随着人类基因组测序和排序工作的完成，以研究基因表达和蛋白质组(Proteomics)为重点的后基因组时代(Post-Genomics Era)已经开始了。在后基因组时代，EST 作为基因表达的时空标签，依然有着重要意义。

### 1.2.2 EST 技术的原理和方法

为获知整个基因组的状况,可通过基因组 DNA 测序和源于 mRNA 的 cDNA 测序两个途径。基因组测序是一种非常有效的策略,但必须辅以计算机和软件及花费昂贵的测序费用;而 cDNA 即表达的基因只占基因组的 2%(Schuler, 1997),长度一般为 500-8 000 bp,易于克隆和测定。

EST 是指从 cDNA 文库中随机挑取克隆进行测序所获得的部分 cDNA 的 5'或 3'端短序列(约 300-500 bp)。不过,近年来越来越多的研究者把长度大于 150 bp 的短序列都作为 EST(骆蒙和贾继增, 2001)。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,它被称为 EST 计划(Adams et al., 1991)。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 2%,EST 反映的是基因组的编码部分,所以 EST 可以直接获得基因表达的信息。目前要多、快、好、省地识别和克隆新基因,EST 测序可能是一条捷径。

基因表达遵从中心法则,即遗传信息由 DNA 序列经转录、剪接后流向 mRNA,再经翻译,产生有功能的蛋白质。一个典型的真核生物 mRNA 分子由 5'-UTR(非翻译区, Untranslated regions), ORF(开放阅读框, Open reading frames), 3'-UTR 和 poly(A)四部分组成;其 cDNA 具有对应的结构。对于任何一个基因,其 5'-UTR 和 3'-UTR 都是特定的,即每条 cDNA 5'端或 3'端的有限序列即可特异性地代表生物体某种组织、某个时期的一个表达基因。来自某一组织足够数量的 EST 即可代表该组织中基因的表达情况(Mekhedov et al., 2000)。多数情况下,某组织总体 EST 中某一种 cDNA 的测量次数可以代表它在该组织中的表达丰度。

### 1.2.3 EST 的应用

随着人类、植物(拟南芥、水稻、玉米等)、动物(果蝇、家猪等)、微生物(酵母等)基因组计划的研究进展,以及近年来在荧光 DNA 大规模自动测序技术和序列信息数据处理技术上迅猛发展,已测序 EST 数量和全球各主要基因数据库中 EST 序列的信息量正以指数数量级递增。据美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI)EST 序列数据库(dbEST, 为 GenBank 中的亚库)(Boguski et al., 1993)所公布的资料,截止 2003 年 12 月 12 日,EST 数目总计已达 1938 万条以上,其中植物 EST 按数目大小排列前 10 名依次为小麦(55 万)、玉米(38.6 万)、大麦(34.8 万)、大豆(34.5 万)、水稻(26.2 万)、拟南芥(19.1 万)、高粱(16.2 万)、番茄(15.1 万)、土豆(13.2 万)、火炬松(11.1 万)等(<http://www.ncbi.nlm.nih.gov/dbEST/>)。而到 2004 年 4 月 30 日,EST 数目总计已达 2090 万条以上,仅 4 个月的时间就增长了约 160 万,可见 EST 增长速度之快。EST 测序分析已经成为一个重要的寻找、鉴定和研究基因及其表达功能的高效快速的新途径和有力工具,尤其是对于那些基因组研究相对缺乏的物种。除了在功能基因组学研究上的应用外,EST 测



序分析现已广泛运用于植物特定组织、生长发育阶段、抗性反应以及代谢过程中的基因及其表达的研究上(Sterky et al., 1998 ; Liu et al., 1995 ; Moon et al., 1998 ; Umeda et al. , 1994 ; Lange et al., 2000)。

拟南芥是 EST 计划开展最早和最系统的植物(Cooke et al., 1996) , 它仅次于人类、大鼠和线虫之后 ,居第 4 位 ,而在水稻之前(Delseny et al., 1997)。纵观已开展的植物基因组计划 ,除了拟南芥与水稻在进行 EST 测序的同时 ,还进行全基因组测序外 ,其它植物基因组计划实际都为 EST 计划。虽然 GenBank 的 dbEST 子数据库中已报道的植物有近百种 ,但真正称得上 EST 计划的 ,大多是重要的粮食作物、经济作物和几种特定的树木 ,而至今为止在 GenBank 的 dbEST 中茶树的 EST 序列只有 5 条 ,几近空白状态(<http://www.ncbi.nlm.nih.gov/dbEST/>)。

就完全基因组而言 ,EST 数据对基因识别、基因组注释和比较基因组研究是一种重要的工具。无论 EST 数据最终如何得到应用 ,利用这些数据重建非冗余性的、高度忠实的转录本数据 ,将会显著增强其中所蕴含的生物学价值。由于 EST 反映的是 mRNA 序列 ,不包括内含子 ,可以直接转换成氨基酸序列 ,在基因研究中起着标签的作用。根据 mRNA 的组织来源和生物体所处的生长阶段 ,EST 反映了其相应基因在时间和空间上的表达 ,这对于寻找在特定时期或组织表达的基因十分简便有效 ,组成同一“重叠群”(contig)的 EST 丰度也在一定程度上反映了表达量的多少。今后在应用 EST 进行研究主要可以从以下几个方面着手 :

分离与鉴定新基因 :传统的基因克隆方法是通过差减杂交、mRNA 差异显示等手段寻找出一些差异表达序列 ,结合图位克隆和转座子标签克隆等手段 ,利用分子标记或表型变化来分析鉴定基因(Rounsley et al., 1998)。近年来 ,由于 DNA 自动测序技术的快速发展 ,用 EST 取代对 cDNA 全长筛选、基因组序列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。由于在构建 cDNA 文库时要尽可能地选用全长 cDNA ,所以一旦发现有价值 EST ,可以找到对应的克隆 ,获得的全长 cDNA 可以直接用于如转基因的研究等。Van de Loo 等(1995)第一次成功地应用 EST 方法分离到蓖麻油酸(12-羟基油酸)生物合成的关键酶基因。

构建遗传学图谱(遗传图谱、物理图谱和转录图谱) :以往作为 DNA 分子遗传标记的多为 RFLP、AFLP、RAPD 和重复序列标记等。EST 片段由于其多态性高也可以作为分子标记 ,用来建立遗传连锁图谱。如 Sasaki 等(1998)用籼稻品种和粳稻品种杂交 F<sub>2</sub> 代 186 个植株作遗传图谱 ,选用了 2300 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。

基因差异表达的研究 :因为 EST 稳定性高、分析规模大 ,对 cDNA 文库随机挑取克隆进行大规模测序 ,可直接得到特定组织细胞的特定时期的表达情况 ,从而能在基因整体水平研究相关的功能及代谢。EST 的最早应用也是基于对基因表达谱的研究而开展起来的(Adams et al., 1991)。

比较基因组学研究 :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(Green et al., 1993) ,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。

应用大量的 EST 序列 ,建立基因芯片 ,进行育种早期鉴定、抗性筛选、基因突变检测等研究。

### 1.2.4 生物信息学在 EST 技术中的应用

生物信息学的兴起与 HGP(Human Genomic Project)的实施密切相关, 它的研究范围覆盖了生物信息的获取、处理、存储、共享、分析和解释等方面。生物信息学是一门交叉科学, 它综合运用数学、计算机科学和生物科学等知识来阐明各类数据的生物学意义, 数据库、计算机网络、分析软件是该学科的运作平台。随着基因组计划的实施及其它分子生物学研究的深入, 每天有大量的核酸和蛋白质序列产生。面对这些数量庞大且又高度复杂的生物数据, 靠人工完成对其分析与计算是不可想象的, 生物信息学由此而产生, 同时也奠定了其在基因组学研究中的重要地位。

生物信息学(Bioinformatics)在大规模测序中起着极为重要的作用, 它可将未知基因和已知基因进行同源性比较, 从而推测未知基因的可能功能, 并可利用电脑对 DNA 序列进行查询、组装、翻译甚至研究高级结构。

#### 1.2.4.1 EST 数据库

数据库是生物信息学的主要内容, 从数据库的种类来看, 核酸和蛋白质序列数据库是最基本的数据库。目前较为常用的核酸序列数据库有: 美国 NCBI 的 GenBank, 欧洲分子生物学实验室的 EMBL(European Molecular Biology Laboratory), 日本国家数据库 DDBJ(DNA Data Bank of Japan), 这三个数据库是收录范围最广并完全向公众开放的数据库, 在它们中均含有 EST 子数据库。在核酸序列数据库中, GenBank 中近 70% 的序列为 EST 序列(Benson et al., 2003)。我国于 1996 年在北京大学建立了生物信息中心(Centre of Bioinformatics, CBI)(罗静初等, 1998), 引进了核酸蛋白质序列、结构等近 40 个数据库, 作为 EMBL 的节点, 它建立了镜像系统, 具有多种服务功能。在各种综合数据库中, EST 的增长速度最快, 1991 年在各个公共数据库中的 EST 数目不足 2000 条, 而到 2004 年 4 月 30 日, 在最大的国际公共数据库 GenBank 中 EST 数量已增至 20,896,801 条。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外, 数据库的专门化也是一种趋势, 即建立特定物种和特定内容的数据库(Gelbart 1998), 如作物研究数据库 GrainGenes, 包括小麦、玉米、大豆基因组子数据库等。拟南芥和水稻由于其在基因组研究中的特殊作用, 也有许多独立的数据库和子数据库。



#### 1.2.4.2 EST 序列分析软件

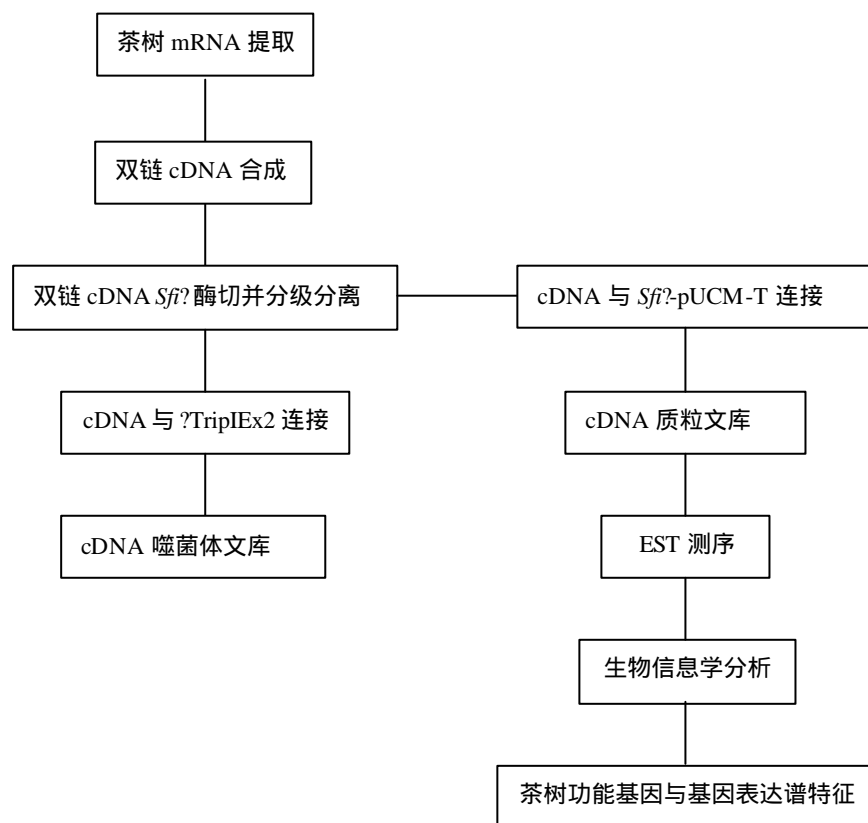
在 EST 研究中, 使用最多的方法就是序列相似性比较, 以此来确定 EST 的功能。BLAST(Basic Local Alignment Search Tool)是应用较广泛的工具软件之一, 为同源分析的软件包, 包括有 BlastN、BlastX、TblastX、BlastP 和 TblastN 等 5 个软件(夏云等, 1997)。其中 BlastN、TblastX、BlastX 是为核酸序列查询设计的, 在 EST 分析中都要用到, 其它两个是为蛋白质序列设计的。BlastN 是待查询序列及其互补序列一起对库查询, 速度快但灵敏度不高; BlastX 是将待查询的序列按 6 种可能的阅读框进行翻译, 然后对蛋白质序列进行查询, 该软件灵敏度高、对序列错误的忍耐性也大。在进行 EST 分析时, 同时使用 BlastN 和 BlastX 会得到较准确的结果。通过 BLAST 分析, 未知功能的序列与储存在序列数据库中的序列做相似性比对, 并根据序列间的吻合程度给出分 (score)值和 E (expect)值。分值越高, E 值越

低，相应的序列间相似性越强，对应功能相近的可能性就越大。核酸与核酸数据库中的序列之间比对结果中  $\text{score} > 100$  的序列认为是有确定功能的(Newman et al., 1994)。核酸与蛋白质数据库中序列之间的比对结果中  $E$  值小于 0.02 认为是有确定功能的(Pearson, 1996)。

### 1.3 本研究的目的意义和思路

借鉴其它生物成功的经验，通过建立茶树表达基因文库，开展 EST 测序分析，可能在短期内获得大量的茶树功能基因表达信息，对迅速提升茶树功能基因研究水平，从基因组水平认识及从分子水平调控茶树的生长、发育、抗性和代谢规律都具有理论与实践意义。本研究拟通过建立茶树新梢的 cDNA 文库，并对文库进行高通量 EST 测序，通过与著名核酸和蛋白质数据库的比对、分析，把 EST 作为研究手段分析茶树新梢的功能基因表达模式，为从基因组水平研究这一木本经济植物提供重要的序列信息，也为今后茶树基因克隆以及分子生物学的研究打下基础。

本实验的研究流程如下：



## 第二章 茶树新梢 cDNA 文库的建立

基因文库是指用体外克隆技术所得到的一种重组 DNA 分子群体,群体中的每个重组分子独立寄存在适当的寄主细胞(如大肠杆菌细胞)中,通过这些寄主细胞繁殖,不仅能使文库中的各种 DNA 分子得到维持和拷贝数扩增,而且还可以通过对这些寄主细胞的基因型或者表达产物的检测分析,从文库中分离出人们感兴趣的基因,开展进一步的理论和开发研究。DNA 分子贮存的遗传信息,转录成 mRNA, mRNA 再翻译成蛋白质,即从 mRNA 中我们可以获得完整的与蛋白质合成相关的基因信息。cDNA 文库在研究具体某类特定细胞中基因组的表达状态以及表达基因功能鉴定方面具有特殊的优势,从而使它在个体发育、细胞分化、细胞周期调控、细胞衰老和凋亡等生命现象的研究中具有更为广泛的应用价值,是研究工作中最常使用到的基因文库。

目前,在其它的动植物中已经成功的应用 cDNA 文库克隆了很多功能基因。茶树在分子生物学水平上的研究远远滞后于水稻、玉米、小麦等作物,在公共数据库中拥有基因序列的数量更是寥寥无几。在本研究之前也只有日本利用茶树 cDNA 文库克隆了部分重要基因(Takeuchi et al., 1994)。因此要克隆茶树上一些重要基因,最好的途径是利用茶树各种组织中特定的 mRNA,逆转录成相应的 cDNA,并以此建立可表达的 cDNA 文库,从中筛选感兴趣基因的阳性克隆。我们利用茶树的新梢,构建了两个茶树品种的 cDNA 文库,为从分子水平研究茶树生长、代谢规律及一些重要基因的克隆、筛选做准备。

### 2.1 材料与方法

#### 2.1.1 实验材料

供试品种为种植于中国农业科学院茶叶研究所内国家种质杭州茶树圃的龙井 43 和安吉白茶。取春茶一芽二叶新梢,液氮迅速冷冻后保存于-70℃ 冰箱中备用。

#### 2.1.2 主要试剂和常用溶液

RNA 提取的 Trizol 试剂购自 GibcoBRL 公司; PolyATtract mRNA Isolation Systems 购自美国 Promega 公司; SMART<sup>TM</sup> cDNA Library Construction Kit 购自 Clontech Laboratories, Inc.。

上样 buffer(1 mL): 由 45% 甲酰胺 545  $\mu$ L、37% 甲醛 196  $\mu$ L、10 $\times$  3-(N- 吗林代)丙磺酸 121  $\mu$ L、80% 甘油 76  $\mu$ L 和 10% 溴酚兰 62  $\mu$ L 配制。

1 $\times$  Lambda 稀释 buffer : 0.1 M NaCl, 10 mM MgSO<sub>4</sub>·7H<sub>2</sub>O, 35 mM Tris-HCl, 0.01% 明胶。

LB 固体培养基(1L): 蛋白胨 10 g, 酵母 5 g, NaCl 5 g, 10 mM MgSO<sub>4</sub>, 琼脂 15 g, 0.2% 麦芽糖, pH 7.0, 高压灭菌 20 min。

LB 液体培养基(含氨卞青霉素)(1L): 蛋白胨 10g, 酵母 5g, NaCl 5g, 氨卞青霉素终浓度为 50  $\mu$ g/mL, 高压灭菌 20 min。

LB 固体培养基(1L): 蛋白胨 10 g, 酵母 5g, NaCl 5 g, 琼脂糖 15 g, pH7.0 高压灭菌 20 min。

LB/MgSO<sub>4</sub> 软顶层培养基(1L): 蛋白胨 10 g, 酵母 5 g, NaCl 5 g, 10 mM MgSO<sub>4</sub>, 琼脂 7.2 g, pH 7.0, 高压灭菌 20 min。

LB/MgSO<sub>4</sub> 固体培养基(1L): 蛋白胨 10 g, 酵母 5 g, NaCl 5 g, 10 mM MgSO<sub>4</sub>, 琼脂 15 g, pH 7.0, 高压灭菌 20 min。

SOC 培养基(1L): 细菌培养用胰化蛋白胨 20 g, 细菌培养用酵母提取物 5 g, NaCl 0.5 g, 2.5 Mm KCl, 10 Mm MgCl<sub>2</sub>, 20 mM 葡萄糖, pH 7.0, 高压灭菌 20 min。

## 2.1.3 实验方法

### 2.1.3.1 新梢总 RNA 的提取

取生长健壮的茶树春茶新梢 100 mg, 迅速加液氮冷冻研磨成细粉末状, 加入 1 000  $\mu$ L Trizol 试剂继续研磨后, 又再加入 1 000  $\mu$ L Trizol 试剂, 混匀分装到两个经 DEPC(焦炭酸二乙酯)处理的 1.5 mL 离心管中, 冰上放置 5 min, 各加入 200  $\mu$ L 的氯仿, 颠倒混匀, 4  $^{\circ}$ C 12 000 rpm 离心 10 min, 取 500  $\mu$ L 上清液, 加入等体积异丙醇, 上下轻轻颠倒 10 次, 室温放置 10 min, 4  $^{\circ}$ C 12 000 rpm 离心 10 min, 去上清, 沉淀中加入 1 mL 75%乙醇轻轻清洗, 倒去乙醇, 干燥后加入 20  $\mu$ L 经 DEPC 处理过的 ddH<sub>2</sub>O 溶解沉淀。甲醛凝胶变性电泳检测 RNA 质量, RNA 3.5  $\mu$ L 中加入上样 buffer 16.5  $\mu$ L, 混匀后在 95  $^{\circ}$ C 变性 2 min 后, 迅速放入冰浴中, 防止退火。电泳时, 电泳槽置于冰上, 保证低温, 电压不超过 5v/cm。RNA 可放于-70  $^{\circ}$ C 冰箱中长期保存, 备用。

### 2.1.3.2 mRNA 的分离和纯化

#### 1) 生物素标记的 Oligo(dT)探针与 mRNA 的退火

在经 DEPC 处理不含 RNAase 的 1.5 mL 离心管中加入 0.5 mg 总 RNA, 溶解于 500  $\mu$ L DEPC 处理水中, 65  $^{\circ}$ C 水浴保温 10 min, 加入 3  $\mu$ L 的 Oligo(dT)探针, 13  $\mu$ L 的 20 $\times$  SSC (1 L 20 $\times$  SSC 含: 175.3g NaCl 和 88.2g 柠檬酸钠, pH 7.0)轻轻混匀, 室温放置约 10-20 min 至冷却。

#### 2) 亲和素顺磁磁珠的冲洗

将一管亲和素顺磁磁珠(购自 Promega Corporation)轻轻悬浮后放入磁性分离架中, 使亲和素顺磁磁珠集中到管的一侧, 小心除去上清, 用 300  $\mu$ L 0.5 $\times$  SSC 清洗亲和素顺磁磁珠三次, 每次用磁性分离架集中磁珠, 去除上清液, 将清洗过的亲和素顺磁磁珠溶于 100  $\mu$ L 的 0.5 $\times$  SSC 中备用。

#### 3) 杂交体的生成和漂洗

将已退火的生物素标记探针, 加入到溶于 100  $\mu$ L 的 0.5 $\times$ SSC 的亲和素顺磁磁珠中, 轻轻混匀, 室温放置 10 min, 每隔 1-2 min 轻混匀一次, 用磁性分离架捕获磁珠亲和素顺磁磁珠, 小心去除上清, 用 300  $\mu$ L 的 0.1 $\times$ SSC 洗涤亲和素顺磁磁珠, 磁性分离架集中后去除上

清液，共重复 4 次。

#### 4) 洗脱 mRNA

将漂洗过的亲和素顺磁磁珠重新悬浮于 100  $\mu\text{L}$  的焦炭酸二乙酯水中，用磁性分离架捕获磁珠，将洗脱的水相吸至一新管中，重复清洗亲和素顺磁磁珠，共得到 250  $\mu\text{L}$  的 mRNA。

#### 5) mRNA 的沉淀和溶解

加入 0.1 $\times$ 体积的醋酸钠(pH 5.2)和 1 $\times$ 体积的异丙醇沉淀 mRNA，-20 $^{\circ}\text{C}$  过夜。12000 rpm 离心 10 min，去上清。加入 1 mL 的 75% 的乙醇悬浮后，离心片刻，去上清，干燥，复溶于 30  $\mu\text{L}$  的 DEPC 处理的水中。3.5  $\mu\text{L}$  mRNA 加入 16.5  $\mu\text{L}$  上样 buffer(按 7:33 配)混匀后在 95 $^{\circ}\text{C}$  变性 2 min，拿出后立即放入冰上，防止退火，甲醛凝胶变性电泳时，电泳槽置于冰上，保证低温。电压不超过 5v/cm。

### 2.1.3.3 cDNA 第一链合成

在 0.5 mL 的离心管中，分别加入 3  $\mu\text{L}$  的上面提取的 mRNA，1  $\mu\text{L}$  10  $\mu\text{M}$  5'PCR 第一链合成引物(5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGG CCGGG-3')，1  $\mu\text{L}$  10  $\mu\text{M}$  3' PCR 引物(5'-ATTCTAGAGGCCGAGGCGGCCGA CAT G- d(T)<sub>30</sub>N<sub>1</sub>N-3') (N=A, G, C 或者 T; N<sub>1</sub>=A, G, C)。混匀，稍离心，72 $^{\circ}\text{C}$  温育 2 min，冰上冷却 2 min，离心把组分收集到底部，每管中加入 2  $\mu\text{L}$  5 $\times$ 第一链合成 buffer(250 mM Tris, pH8.3, 30 mM MgCl<sub>2</sub>, 375 mM KCl)，1  $\mu\text{L}$  二硫苏糖醇(20 mM)，1  $\mu\text{L}$  dNTP 混合物(10 mM)，1  $\mu\text{L}$  反转录酶，至总体积为 10  $\mu\text{L}$ 。经涡旋、瞬间离心混匀，在 PTC-220 PCR 仪(MJ Research, Inc.)中 42 $^{\circ}\text{C}$  温育 1 hr 后，取出放于冰上，终止第一链的合成。放于-20 $^{\circ}\text{C}$  冰箱中可保存 3 个月，备用。

### 2.1.3.4 cDNA 第二链的合成

先把 PCR 仪加热到 95 $^{\circ}\text{C}$ 。在一个 0.5 mL 的离心管中加入 2  $\mu\text{L}$  第一链 cDNA，80  $\mu\text{L}$  无离子水，10  $\mu\text{L}$  10 $\times$  PCR buffer，2  $\mu\text{L}$  50 $\times$  dNTP mix，2  $\mu\text{L}$  10  $\mu\text{M}$  5' PCR 第二链合成引物(5'-AAGCAGTGGTATCAACGCAGAGT-3')，2  $\mu\text{L}$  3'PCR 引物(5'-ATTCT AGAGG CCGAGGCGGCCGACATG-d(T)<sub>30</sub>N<sub>1</sub>N-3') (N=A, G, C 或者 T; N<sub>1</sub>=A, G, C)，2  $\mu\text{L}$  50 $\times$  聚合酶混合物至总体积为 100  $\mu\text{L}$ ，轻轻振动混匀，稍离心，将物质收集到底部，必要时加 2 滴矿物油以覆盖液面。放进经预热到 95 $^{\circ}\text{C}$  的 PCR 仪中，PCR 反应程序如下：95 $^{\circ}\text{C}$  变性 20 s，接着 95 $^{\circ}\text{C}$  5 s，68 $^{\circ}\text{C}$  6 min 共 24 个循环，PCR 结束后，取 5  $\mu\text{L}$  用 1.1% 琼脂糖凝胶在 0.5 $\times$  TBE (45 mM Tris-硼酸，1 mM EDTA，pH8.0)中电泳。其余放于-20 $^{\circ}\text{C}$  冰箱可保存 3 个月。

### 2.1.3.5 蛋白消化及 cDNA 纯化

- 1) 取 75  $\mu\text{L}$  cDNA 加到一个 0.5 mL 的管中，加 2  $\mu\text{L}$  的蛋白酶 K(20  $\mu\text{g}/\mu\text{L}$ )
- 2) 45 $^{\circ}\text{C}$  温育 20 min
- 3) 取出后，离心一下，将组分收集到管底
- 4) 加入 23  $\mu\text{L}$  的无离子水，保证总体积为 100  $\mu\text{L}$
- 5) 加入等体积充分混匀的酚：氯仿：异戊醇(25:24:1，v/v/v)，来回混匀 1-2 min
- 6) 14000 rpm 离心 5 min
- 7) 离心后有三层，取上层到一个新的 0.5 mL 的管中

- 8) 加入 100  $\mu$ L 的氯仿：异戊醇(24:1, v/v)来回混匀 1-2 min
- 9) 14000 rpm 离心 5 min
- 10) 取上清液, 加入 10  $\mu$ L 的醋酸钠 (3M, pH4.8)、1.3  $\mu$ L glycogen (20  $\mu$ g/ $\mu$ L), 260  $\mu$ L 95% 乙醇。迅速以 14000 rpm 室温离心 20 min
- 11) 小心除去上清, 用 100  $\mu$ L 80%的乙醇洗涤沉淀
- 12) 干燥约 10 min 去除残留的乙醇
- 13) 加入 79  $\mu$ L 的无离子水, 悬浮沉淀

#### 2.1.3.6 *Sfi* I 酶切获得可连接的 cDNA

- 1) 在一个 0.5 mL 的离心管中加入, 79  $\mu$ L cDNA(经蛋白酶 K 消化), 10  $\mu$ L 10 $\times$ *Sfi* I buffer, 10  $\mu$ L *Sfi* I 内切酶(20 单位/ $\mu$ L), 1  $\mu$ L 100 $\times$ BSA(牛血清白蛋白)充分混匀
- 2) 50  $^{\circ}$ C 水浴保温 2 hr, 取出后放于-20  $^{\circ}$ C 冰箱中
- 3) 分离时再加入 2  $\mu$ L 1% 二甲苯青, 充分混匀

#### 2.1.3.7 cDNA 片段的分离

- 1) 准备 16 个 1.5 mL 离心管
- 2) 准备 CHROMA SPIN-400 柱(购自 Clontech Laboratories, Inc.)
  - a) 将 CHROMA SPIN-400 从 4  $^{\circ}$ C 中取出, 在室温下放置 1 hr, 颠倒数次彻底悬浮胶体;
  - b) 从柱中除去气泡, 用 1000  $\mu$ L 的移液器轻轻悬浮胶体, 避免产生气泡。拿走底部的盖, 让柱体自然流下。(如 3 min 之后不流了, 盖上上面的盖, 会使柱体继续流下);
  - c) 铁夹夹住柱子;
  - d) 贮存 buffer 通过柱子自然流下, 直至能看到柱体中胶的表面, 胶的高度应在 1 mL 标记处。如果少的话, 可以用其它柱中的胶来调整;
  - e) 流速大约为 1 滴/40-60 s, 1 滴约为 40  $\mu$ L。(如果流速太低, <1 滴/100 s, 或者每滴的体积<25  $\mu$ L 必须重新悬浮, 重复上述步骤)。
- 3) 贮存 buffer 流完后, 靠柱内边缘小心加入 700  $\mu$ L 的柱子 buffer, 直到流光
- 4) 当柱子 buffer 流完后(约 15-20 min), 小心将约 100  $\mu$ L 经 *Sfi* I 酶切的 cDNA 和二甲苯青的混合物加到胶顶部中心部位的表面, 不光滑的胶表面不会影响下面的分离
- 5) 进行下一步之前, 让样品充分进入胶内(在胶表面没有残留的液体)
- 6) 用 100  $\mu$ L 的柱子 buffer 清洗含有 cDNA 的管, 再轻倒入胶表面
- 7) 让 buffer 流光, 当停止时, 继续下一步, 让染料层进入胶内几毫米
- 8) 将放有 16 个管子的架子放在柱子下面, 第一个管子正在柱子下的出口
- 9) 加 600  $\mu$ L 的柱子 buffer, 立即开始收集, 每管 35  $\mu$ L(约 1 滴), 收集后的管子立即加盖, 当都收集完后, 盖上柱子的盖子
- 10) 在进行实验前, 检查每个管子的组分, 在 1.1%的琼脂糖/EB 胶上电泳, 每管用 3  $\mu$ L 连续在点样孔加样, 用 1 kb DNA Marker 作分子量标记, 150 V 电泳 10 min。收集最亮的 3-4 个组分到一个新 1.5 mL 的管中
- 11) 将 1/10 体积醋酸钠 (3M; pH4.8), 1.3  $\mu$ L glycogen(20 mg/mL), 2.5 体积 95%乙醇(-20  $^{\circ}$ C) 混匀, -20  $^{\circ}$ C 过夜

- 12) 室温 14000 rpm 离心 20 min
- 13) 小心去上清。轻离心，去剩余的上清，干燥约 10 min
- 14) 溶于 7 $\mu$ L 的无离子水，轻轻混匀。这些 *Sfi*I 酶切的 cDNA 现在已经可以与经 *Sfi*I 酶切的  $\phi$ TripIEx2 载体(购自 Clontech Laboratories, Inc.)连接

#### 2.1.3.8 cDNA 与载体连接

1) 按下列组分准备一个连接反应	对照( $\mu$ L)	样品( $\mu$ L)
cDNA	0	1.0
$\phi$ TripIEx2 载体(500ng/ $\mu$ L)	1.0	1.0
ATP(10mM)	0.5	0.5
T <sub>4</sub> DNA ligase (400 U/ $\mu$ L)	0.5	0.5
10 $\times$ ligase buffer	0.5	0.5
ddH <sub>2</sub> O	2.5	1.5
总体积	5.0	5.0

其中 10 $\times$  ligase buffer 由 500mM Tris-HCl, pH7.8, 100mM MgCl<sub>2</sub>, 100mM 二硫苏糖醇, 0.5 mg/mL BSA 组成。

- 2) 轻轻混匀，避免产生气泡，瞬间离心使组分沉到底部，16 $\times$  连接过夜
- 3) 对于每一个连接，做一个包装反应

#### 2.1.3.9 包装反应

- 1) 在每一个连接产物中加入 25  $\mu$ L 的包装蛋白
- 2) 30 $\times$  温育 90 min
- 3) 取出放于 4 $\times$  冰箱中可保存 2 周。经过扩增的文库可以在 4 $\times$  稳定保存 6-7 个月，或在 7% 二甲基亚砷(DMSO)中-70 $\times$  至少保存 1 年以上

#### 2.1.3.10 测定滴度

- 1) 原始和工作平板的制备

为复苏冰冻细胞，取 5 $\mu$ L 冰冻的大肠杆菌细胞 XL1-Blue 在含氨卞青霉素的 LB 平板上划线，37 $\times$  培养过夜后，作为原始平板，置于 4 $\times$  可保存 2 周。从原始平板中挑取单克隆在另一个含氨卞青霉素的 LB/MgSO<sub>4</sub> 培养基平板上涂板，37 $\times$  培养过夜后，作为工作平板。

- 2) 未扩增文库的滴度测定

从工作平板中挑取分离良好的单克隆接种到含 15 mL LB/MgSO<sub>4</sub>/麦芽糖培养基的 50 mL 试管中，140 rpm 震荡 37 $\times$  培养过夜直至 OD 值达到 2.0，5000 rpm 离心 5 min，去上清液，用 7.5 mL 10 mM MgSO<sub>4</sub> 重新悬浮沉淀。将包装产物用噬菌体 buffer 稀释 5-20 倍，在 200  $\mu$ L 过夜培养的 XL-1Blue 受体菌中加入 1 $\mu$ L 稀释的噬菌体，让噬菌体在 37 $\times$  吸收 10-15 min，再加入 2 mL 溶解的 LB/MgSO<sub>4</sub> 顶层琼脂。快速颠倒混匀并倒入经 37 $\times$  预热的 LB/MgSO<sub>4</sub> 平板，快速摇动平板使均匀分布，室温冷却 10 min 让顶层琼脂变硬，倒置平板在 37 $\times$  培养 6-17 hr，统计菌斑并根据公式：文库滴度 (pfu/mL)=(噬菌斑数 $\times$ 稀释倍数)/受体菌的体积，计算未扩增文库滴度。



### 2.1.3.11 文库重组率的测定

本文库可以利用 X-Gal 显色原理测定重组率,在测定滴度时加入 100 mM 的 IPTG 及 100 mM 的 X-Gal 过夜培养后,通过蓝斑(未重组克隆)、白斑(重组克隆)的数量计算重组率。

### 2.1.3.12 cDNA 插入片段的 PCR 鉴定

从 LB 平板上随机挑取 15-20 个独立的噬菌斑,分别加到含有 200  $\mu$ L 噬菌体 buffer 的离心管中,37 $^{\circ}$ C 放置 1 hr 后,保存于 4 $^{\circ}$ C,利用 TripIEx2 噬菌体 5' PCR 引物(5'-CTCCGAGATCTGGACGAGCT-3')和 3' PCR 引物(5'-GGGATATCACTC AGCATAAT-3')对构建的文库进行 PCR 鉴定。PCR 反应体积为 20  $\mu$ L,其中的组成为:含单克隆噬菌斑的稀释液 2  $\mu$ L, 25 mM  $Mg^{2+}$  1.4  $\mu$ L, 10  $\mu$ M dNTPs 0.4  $\mu$ L, 10 $\times$  PCR buffer 2  $\mu$ L, 10  $\mu$ M 5' 及 3' 引物分别 0.1  $\mu$ L, *Taq* DNA 聚合酶 0.1  $\mu$ L, ddH<sub>2</sub>O 13.9  $\mu$ L。PCR 反应条件为:95 $^{\circ}$ C 预变性 4 min, 95 $^{\circ}$ C 变性 30 s, 55 $^{\circ}$ C 退火 45 s, 72 $^{\circ}$ C 延伸 1 min, 36 个循环,最后 72 $^{\circ}$ C 延伸 10 min。反应在 MJ PTC-220 型 PCR 仪上进行。

### 2.1.3.13 cDNA 文库的扩增

从工作平板中挑取分离良好的 XL1-Blue 单克隆接种到含 15 mL 的 LB/MgSO<sub>4</sub>/麦芽糖培养基的 50 mL 三角瓶中,37 $^{\circ}$ C 140 rpm 震荡培养过夜至 OD 值达到 2.0。5000 rpm 离心 5 min,去上清液,用 7.5 mL 10 mM MgSO<sub>4</sub> 重新悬浮沉淀。在 7 mL 试管中加入 500  $\mu$ L 细菌培养液和足够稀释过的溶菌物,在 37 $^{\circ}$ C 水浴温育 15 min,加 4.5 mL 熔化的 LB/MgSO<sub>4</sub> 软顶层琼脂,快速混合并倾倒细菌/噬菌体混合物到 LB/MgSO<sub>4</sub> 平板上,使琼脂均匀铺于平板上。室温冷却 10 min,使顶层琼脂变硬,倒置平板于 37 $^{\circ}$ C 培养 6-18 hr 直至菌落长至相互接触。加入 12 mL 1 $\times$  Lambda 稀释 buffer 于每个平板,4 $^{\circ}$ C 过夜,一个平板可以成为一个扩增文库的溶菌物。平板在室温 50 rpm 震荡 1 hr 后倒入一个已灭菌的广口瓶。为去除细胞碎片和裂解剩余的完整细胞,充分混合噬菌体裂解物并倒入一个 50 mL 聚丙烯带盖的灭菌离心管中,加入 10 mL 氯仿盖紧盖子,涡旋 2 min 后,7 000 rpm 离心 10 min,上清液转移至另一个灭菌的 50 mL 离心管中,4 $^{\circ}$ C 保存。

采用 2.1.3.10 测定滴度的方法测定扩增后文库的滴度。

把经扩增的文库分装成 1 mL/管,加入终浓度 7% 的 DMSO,于 -70 $^{\circ}$ C 长期保存,尽量避免多次冻融过程。

## 2.2 结果与分析

### 2.2.1 RNA 质量

利用 Trizol 试剂及 Promega 公司的 mRNA 纯化试剂盒获得的总 RNA 和 mRNA 质量较好。总 RNA 在含有甲醛的凝胶上进行变性电泳,结果如图 2-1,可以看到 28S、18S、5S 三条带,其中 28S 的条带很亮,18S 次之,5S 的亮度非常弱,说明所提取的 RNA 基本上没有降解,完整性较好。加样孔处没有亮带,即没有 DNA 污染。龙井 43、安吉白茶总 RNA 在

260nm 和 280nm 处的光吸收值之比  $OD_{260/280}$  分别为 1.96 和 2.09，说明均一性较好，蛋白质的含量很低。分离纯化的 mRNA 电泳后呈弥散状分布，也证明了总 RNA 的完整性。用此 mRNA 合成的 cDNA 第二链的电泳结果见图 2-2，可以看到其片段主要分布在 0.5-10 kb 之间。

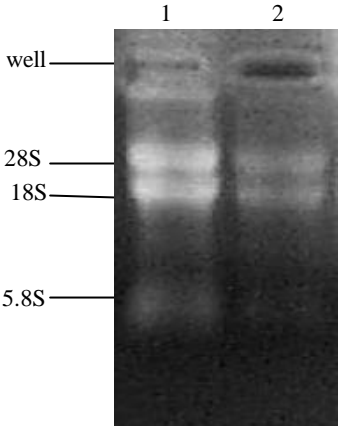


图 2-1. 总 RNA 的甲醛变性凝胶电泳  
Fig.2-1 Total RNA denatured formaldehyde gel analysis  
1:龙井 43 Longjing 43; 2:安吉白茶 Anji Baicha

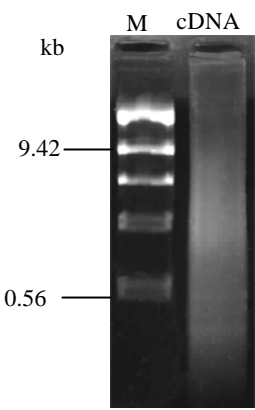


图 2-2. 龙井 43 cDNA 的 1.1%琼脂糖凝胶电泳  
M: /Hind DNA Ladder  
Fig.2-2 cDNA agarose gel electrophoresis of Longjing 43

## 2.2.2 cDNA 文库的质量

### 2.2.2.1 滴度和重组率

用包装产物感染 XL1-Blue 菌，并将感染后的菌液进行不同倍数的稀释后进行文库滴度的测定（结果见表 2-1），两个文库的感染后的菌液的体积均为 525 $\mu$ L。计算出两个文库龙井 43、安吉白茶文库滴度分别为  $6.8 \times 10^5$  pfu/mL 和  $4 \times 10^5$  pfu/mL，用 1 $\times$  Lambda 稀释 buffer 从平板上获得的菌液，其中再次涂板后进行扩增后文库的滴度的测定。分别为  $8.5 \times 10^9$  pfu/mL 和  $5.7 \times 10^9$  pfu/mL。

表 2-1 文库的滴度测定

Table2-1 Titering of the libraries			
稀释倍数 Dilution	龙井 43 Longjing 43	安吉白茶 Anji Baicha	
未扩增文库	$10^{-1}$	77	40
Original library	$10^{-2}$	6	4
平均滴度 Mean titer (pfu/mL)	$6.8 \times 10^5$	$4 \times 10^5$	
扩增文库	$10^{-4}$	940	511
Amplified library	$10^{-5}$	76	63
平均滴度 Mean titer (pfu/mL)	$8.5 \times 10^9$	$5.7 \times 10^9$	

用 IPTG 诱导并用 X-Gal 为底物培养获得蓝白斑，其中白斑为重组克隆、蓝斑为未重组的克隆，计算蓝白斑比例（表 2-2），得到龙井 43、安吉白茶文库的重组率分别为 98.05% 和 96.15%，文库总克隆数为  $3.5 \times 10^5$  个、 $2.1 \times 10^5$  个。

表 2-2 文库重组率的测定

Table 2-2 The recombinant rates of the libraries

	龙井 43 Longjing 43	安吉白茶 Anji Baicha
稀释倍数 Dilution	$10^{-4}$	$10^{-5}$
总克隆数 Total clones	940	511
蓝斑数 No. of blue spots	20	23
重组率% Recombinant rate	97.4	95.5
平均重组率% Mean rate	98.05	96.15

### 2.2.2.2 插入片段大小

对非扩增的 2 个文库随机各挑取 11 个噬菌斑，用试剂盒中提高的两端引物进行 PCR 扩增，结果如图 2-3，安吉白茶的随机克隆中有一个小于 0.5kb，一个有两条扩增条带，两个条带出现可能是挑取的时候两个距离较近噬菌斑发生污染造成的，其余大多数插入片段的长度在 1.0-1.5 kb 左右，而龙井 43 的插入片段均在 1.0-1.5kb 左右。由此说明两个文库中的插入片段的质量较好。

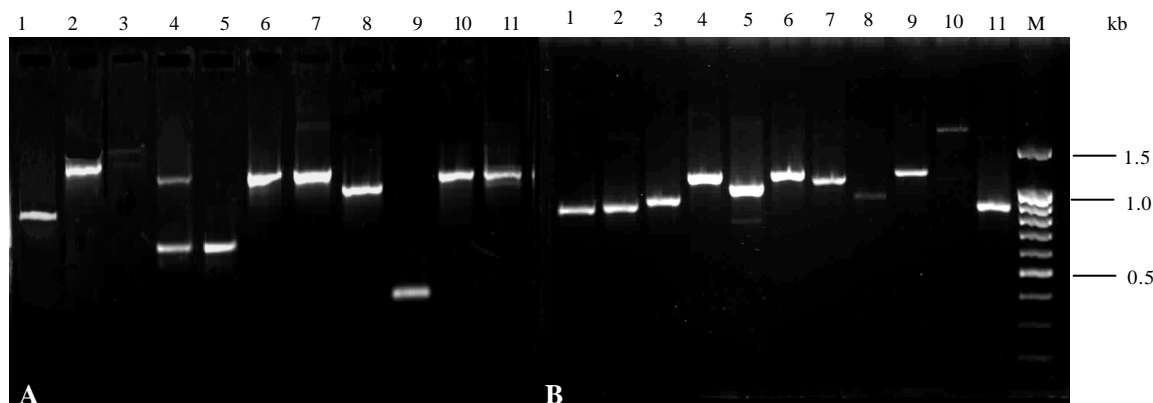


图 2-3. 龙井 43(B)与安吉白茶(A)文库插入片段的 PCR 鉴定

1-11 为随机选取的 11 个 cDNA 克隆，M 为分子量标准

Fig. 2-3 PCR analyses of 11 recombinant phages of Longjing 43 (B) and Anji Baicha (A) cDNA libraries

Lane1-11: randomly selected cDNA clones, M: 100bp +DNA ladder

## 2.3 讨论

植物组织中的 RNA 含量较低，而且离体的植物组织 RNA 降解速度很快。本研究构建的是茶树新梢 cDNA 文库，因此茶树生长状况、采样时期非常重要。一定要选择幼嫩的新梢，最好选择春茶的嫩梢。采下后迅速用液氮冷冻，然后放于-70℃中保存，保证 RNA 不被降解。

植物总 RNA 提取是分子生物学的基本技术之一，其困难首先是 RNA 非常容易降解，而 RNase 又无处不在，并且非常稳定；其次是组织中的次生物质如酚类、多糖等也会影响 RNA 的质量。茶树新梢富含茶多酚，其氧化后会与 RNA 结合。本实验方法区别于江昌俊等

(2001)对 Chomczynski 一步法分离 RNA 的改进,在于提取 RNA 试剂中含有强还原剂,充分抑制了酚类的氧化,具有快速、简便的优点。所有的试剂及所用的器皿都进行了 DEPC 处理,充分抑制了 RNase 的活性。所获得 RNA 的 28S 最亮,18S 稍弱,说明总 RNA 基本上没有降解,可以用于 mRNA 的分离。要构建 cDNA 文库,获得高质量的 mRNA 至关重要。本研究所用的试剂盒分离 mRNA 直接、快速,电泳结果表明 mRNA 的质量很好。在合成的 cDNA 中,有一些分子量小于 400bp 的片段,这就需要用 cDNA 大小分离柱将这些小片段去除,否则在连接时载体就会优先与这些小片段连接,所构建的文库主要为一些小片段克隆,其完整性就差。

文库的滴度、重组率及插入片段的大小是鉴定 cDNA 文库质量的重要指标。一般来说,文库的滴度能达到  $10^5$  pfu/mL 以上即为有效的文库(Li et al., 1998)。根据公式  $N=\ln(1-P)/\ln(1-1/n)$  ( $N$  为实际所需克隆数,  $P$  为要求的概率(一般为 99%),  $n$  为某一种类型的稀有 mRNA 在总 mRNA 中所占的比例),当要求克隆到某低丰度 mRNA 的概率为 99%时,文库的克隆数不应低于  $1.7 \times 10^5$  个(Sambrook et al., 1998)。本文所构建 cDNA 文库的滴度分别为  $6.8 \times 10^5$  pfu/mL 和  $4 \times 10^5$  pfu/mL 符合这一要求,文库的重组率分别为 98.05%和 96.15%,说明文库是完整的、有效的,并且质量较高。文库的总克隆数分别为  $3.5 \times 10^5$  个和  $2.1 \times 10^5$  个,高于所需标准,因此从所构建文库筛选低丰度的目的基因在理论上是可行的。

利用分子克隆技术,构建不同组织的 cDNA 文库,然后用适当的核酸探针、对特定抗原进行免疫学检测、通过 mRNA 的杂交体选择和翻译或者通过产生有生物学活性的分子进行同源选择是从中筛选分离目的基因、获取新基因的重要手段。尽管核酸杂交分析能够测定出基因组克隆中基因间隔序列的位置,然而更精确的结果则要通过比较基因组基因同其 mRNA 转录本在核苷酸序列结构上的差异才能确定。同样地, mRNA 的 5'末端和 3'末端精确的序列结构,也只有通过核苷酸序列的分析才能获得(吴乃虎等, 1998)。目前,由于高通量 DNA 测序设备和技术的发展, cDNA 分子不含有内含子序列和调控序列,且与蛋白质的氨基酸序列具有对应性,大规模的 cDNA 测序已成为功能基因组学研究崭新策略和途径。因此,构建 cDNA 文库是筛选新基因、分离已知基因、揭示某些因素对表达的影响的基础,具有重要的价值和意义。

## 第三章 茶树新梢 EST 序列与功能基因分析

EST 测序分析是一个重要的寻找、鉴定和研究基因及其表达功能的高效快速的新途径和有力工具，尤其是对于那些基因组研究相对缺乏的物种。除了在功能基因组学研究上的应用外，EST 测序分析现已广泛运用于植物特定组织、生长发育阶段、抗性反应以及代谢过程中的基因及其表达的研究上(Sterky et al., 1998 ;Liu et al., 1995 ;Moon et al., 1998 ;Umeda et al. , 1994 ; Lange et al., 2000)。

在已开展的植物基因组计划中，除了拟南芥与水稻在进行 EST 测序的同时，还进行全基因组测序外，其它植物基因组计划实际都为 EST 计划。虽然 GenBank 的 dbEST 子数据库中已报道的植物有近百种，但真正称得上 EST 计划的，大多是重要的粮食作物、经济作物和几种特定的树木，而至今为止在 GenBank 的 dbEST 中茶树的 EST 序列只有 5 条，几近空白状态(<http://www.ncbi.nlm.nih.gov/dbEST/>)。

为了解茶树生长发育最旺盛部位——新梢功能基因表达情况，本章构建了一个龙井 43 新梢 cDNA 质粒文库并进行高通量的测序分析，产生大量 EST 序列，并与大型基因组数据库中的巨量已知基因序列信息进行比较分析，把 ESTs 作为研究手段分析茶树新梢的基因表达模式，为从基因组水平研究这一木本经济作物提供重要的序列信息。

### 3.1 材料与方法

#### 3.1.1 材料

第二章 所获得的经过酶切过柱回收的双链 cDNA。

质粒提取溶液 :50mM Tris-HCl, pH8.0, 10mM 乙二醇四乙酸, 0.1mg/mL RNAase

质粒提取溶液 : 0.2N NaOH, 1% 十二烷基硫酸钠, 室温保存

质粒提取溶液 : 3 M 醋酸钾, 冰醋酸调节 pH 至 5.5, 4℃ 保存

#### 3.1.2 方法

##### 3.1.2.1 *Sfi*<sup>+</sup>-pUC 载体的构建

由于用于构建 cDNA 噬菌体文库试剂盒中提供用于转为质粒文库的菌株 BM25.8 获得文库的拷贝数低。而为了提高测序效率，获得高质量的序列信息，我们改用了本实验室的菌株 TG1，所以对相应的载体也进行了改变。因此在原来的文库材料的基础上，构建了一个新的质粒载体 *Sfi*<sup>+</sup>-pUC。步骤如下：

- 1) 选择一个由原噬菌体文库获得的质粒文库中的质粒 DL003E08，用双向引物进行 PCR 扩增包含 *Sfi*<sup>+</sup> 两个酶切位点在内的中间片段，反应体系如下：

Taq polymerase buffer(含 Mg<sup>2+</sup>)                      6μL

dNTP(10mM)	1.2μ L
Primer3'/5'(10mM)	0.3μ L/each
Taq polymerase	0.3μ L
ddH <sub>2</sub> O	51.4μ L
PlasmidDNA	0.5μ L
Total	60μ L

混和均匀后，收集组分，放入 PCR 仪中进行扩增，参数如下：

95 预变性 4min，95 变性 30s，54 退火 45s，72 延伸 1min，36 个循环，最后 72 延伸 10min，4μL 进行琼脂糖电泳检测。

- 2) PCR 产物纯化 :用 Vitagene DNA 纯化试剂盒(杭州维特洁生化技术有限公司)按照下面步骤进行纯化：

在 PCR 产物中加入 168μL 的 PCR-A 溶液，混和均匀后转移到 DNA 制备管中，将 DNA 制备管置于 2mL 离心管中，3600rpm 离心 1min，弃滤液，重复进行一次；

将制备管置回离心管，加入 0.5mL 的 W1 溶液，3600rpm 离心 30s，弃滤液；

将制备管置回离心管，加入 0.7mL 的 W2 溶液，3600rpm 离心 30s，弃滤液；以同样的方法再用 0.7mL 的 W2 溶液洗涤一次；

将制备管置回离心管中，最高速度离心 1min；

将制备管置于洁净的 1.5mL 离心管中，在 DNA 制备膜正中央加 30μL 水(60 预热)，室温静置 1min，最高速度离心 1min，洗脱 DNA，然后重复洗一次，获得 30μL 的纯化产物。

- 3) 将纯化的 PCR 产物与载体进行连接

PCR 纯化产物	7μ L
pUCM-T 载体	1μ L
T <sub>4</sub> -ligase	1μ L
T <sub>4</sub> ligase buffer	1μ L
Total	10μ L，16 过夜

- 4) 制备感受态细胞

- 挑取 TG1 菌单菌落于 10 mL LB 液体培养基中培养，37 培养过夜；
- 取 100μL 转接到 10 mL LB 液体培养基中，37 培养 2-3 hr，到对数生长期，取出后将菌液分装到 1.5 mL 的离心管中，冰上放置 5-10 min；
- 4 4000rpm 离心 1min，倒去上清液，用 150 μ L 0.1M CaCl<sub>2</sub> 悬浮细胞，冰上放置 3-5 min 后，4000rpm 离心 1min，去掉上清；
- 重复一次，用 50μL 的 0.1M CaCl<sub>2</sub> 悬浮细胞，获得感受态细胞，迅速放于-70 保存。

- 5) 转化

- 5μL 连接产物与 50μL 的感受态细胞混合均匀，冰上放置 30 min；
- 在 42 的恒温水浴锅热激 60 s；
- 迅速取出放置于冰上 5 min；
- 加入 700μL 的 LB 液体培养基，37 摇床 190 rpm 培养 1.5 hr；

- e) 取出,在室温下 4000rpm 离心 5 min ,然后吸去 600  $\mu$  L 上清液,用剩余的培养基悬浮细胞,获得转化产物;
- f) 将转化产物均匀涂布在已经涂好 X-gal(100mg/mL 8 $\mu$  L)和 IPTG(100mg/mL 20 $\mu$  L)的 LB/Amp 的平板上,37 培养箱过培养过夜。

#### 6) 质粒提取

提取过夜培养的转化菌的单菌落,于 LB/Amp 培养基中培养过夜,按照 (Sambrook et al., 1998)的方法提取质粒 DNA,进行序列测定。

#### 7) 质粒酶切,回收

将提取的质粒 DNA 用 *Sfi*<sup>+</sup> 酶进行切割,切去中间的插入片断,以获得具有 *Sfi*<sup>+</sup> 酶切位点的新载体,反应体系如下:

质粒 DNA	10 $\mu$ L
10xbuffer	8 $\mu$ L
<i>Sfi</i> enzyme	2 $\mu$ L
10xBSA	0.8 $\mu$ L
ddH <sub>2</sub> O	61.2 $\mu$ L
Total	80 $\mu$ L ,
	50 2hr

酶切后全部进行电泳,使用 QIAquick Gel Extraction 试剂盒回收纯化酶切片断,最后获得新载体 *Sfi*<sup>+</sup>-pUC。

### 3.1.2.2 cDNA 与新载体连接

连接体系如下:

cDNA	0.5 $\mu$ L
<i>Sfi</i> <sup>+</sup> -pUC vector	3 $\mu$ L
T <sub>4</sub> ligase	1 $\mu$ L
T <sub>4</sub> ligase buffer	1 $\mu$ L
H <sub>2</sub> O	4.5 $\mu$ L
Total	10 $\mu$ L , 16 过夜

连接产物按照前面的方法制备 TG1 菌的感受态细胞,并将连接产物进行转化,最后获得质粒文库。放于 4 保存备用。

### 3.1.2.3 cDNA 序列测定

#### 3.1.2.3.1 质粒的培养和提取

将获得的 cDNA 质粒文库稀释,进行梯度试验,选出最佳稀释浓度,在 LB 固体培养基上涂板。

- 1) 37 培养 12-16 hr 左右,直至长出菌落。
- 2) 将菌落接种在 96 孔板上,每孔加入 1mL LB(含氨卞青霉素终浓度为 50  $\mu$  g/mL)接种单细菌克隆,37 培养 16-18 hr 左右。保存菌种;3000 rpm 离心 5 min,弃去上清液,收集细胞。
- 3) 加入 200 $\mu$  L 溶液,用胶带封口,振荡混匀。

- 4) 加入 200  $\mu\text{L}$  溶液，用新胶带封口，翻转混合 10 次。此时裂解液应该透明、粘稠，不含细菌碎片。
- 5) 加入 200  $\mu\text{L}$  溶液，用新胶带封口，翻转混合 10 次。
- 6) 沸水浴 5 min，冰浴 10 min，4000 rpm 离心 20 min，上清液全部转入下面的深孔板中。
- 7) 取下深孔板，加入 300  $\mu\text{L}$  异丙醇溶液，用胶带封口，立即翻转混合 3 次。离心 15 min(4000 rpm)，沉淀 DNA。
- 8) 去掉上清，加入 300  $\mu\text{L}$  70%乙醇，4000 rpm 离心 5 min。
- 9) 用 20  $\mu\text{L}$  水重新溶解 DNA，-20 保存备用。

### 3.1.2.3.2 测序 PCR 反应和序列测定

- 1) 用水稀释 5'测序引物 (5'-CTCCGAGATCTGGACGAGC T-3')。
- 2) 在 96 孔板上，将要测序的每个模板加入下列试剂：

DYEnamicET 终止子试剂预先混合物	8 $\mu\text{L}$
5'测序引物(5 $\mu\text{M}$ )	1 $\mu\text{L}$
DNA 模板	0.2-2 $\mu\text{g}$
Total	20 $\mu\text{L}$

轻轻震荡充分混合

- 3) 按照下列 PCR 条件进行测序反应：

95	20s
50	15s
60	1min

共 32 个循环，循环完成后，简单离心收集管子底部的溶液。

- 4) 每个反应管中加 1  $\mu\text{L}$  7.5 M 乙酸铵，27.5  $\mu\text{L}$  无水乙醇，5  $\mu\text{L}$  ddH<sub>2</sub>O。
- 5) 4000 rpm 离心 40 min。
- 6) 除掉上清液，每个反应管中加 50  $\mu\text{L}$  70%的乙醇，并混合好。4000 rpm 离心 20 min。
- 7) 除去上清液，真空干燥或空气干燥沉淀约 20 min。不要过分干燥。
- 8) 将沉淀重新悬浮在 5  $\mu\text{L}$  的 loading solution 中，用力震荡 10-20 s，确保完全悬浮，简单离心收集样品。
- 9) 将样品放入 MegaBace™ 1000DNA 测序仪(Amershan Pharmacia Biotech Inc.)中，按照设定的参数从 5'端进行测序。

### 3.1.2.4 生物信息学分析

#### 3.1.2.4.1 序列编辑

以 Chromas 软件并结合手工处理，去除测序误差>5%或长度<150bp 的序列、删除序列两端的载体序列和模糊序列。将序列文件 abd 格式转化为 fasta 格式保存，其中 3'端具 Poly(A)的序列与不具有 Poly(A)尾的序列分别放于不同的文件夹中。

#### 3.1.2.4.2 序列格式转化及 GenBank 数据库提交

校正后的序列转化为通用的 FASTA 格式后，再转为 GenBank 提交格式，提交到 GenBank 网上公开。



#### 3.1.2.4.3 序列同一性分析

对所有序列进行同一性分析，也就是去除所有序列中高度重复的序列，利用 DNAtools 软件分析与人工结合的分析方法，对所有序列两两之间进行比对，经人工判断后，最终认定二者是否具有同一性、是否代表同一基因。由于是从 5'端开始测序，所以判断的标准是：1、对比的最小片断为 24bp；2、两序列存在 80%的同源性；3、一条序列几乎完全被另一条所包含。汇总同一性比对结果，可对所有序列进行 Cluster 归类(一个 Cluster 代表一个独立基因)。

#### 3.1.2.4.4 基因数据库的构建及功能基因的分析

将经过上述处理剩余的质量好的序列的 EST 逐条与 GenBank (<http://www.ncbi.nlm.nih.gov>)的核酸数据库进行 BlastN 比对分析，对 BlastN 结果分值(score 值)大于 100 的序列，给予相关的功能注释，注释内容包括序列名称、GenBank 登录号、比对上的物种的拉丁名、中文名、匹配序列的功能注释、分值、e-value、匹配碱基的数目、匹配序列等等；同样再将所有序列与蛋白质数据库进行 BlastX 比对分析，对 BlastX 结果 E 值小于 0.02 的序列(Pearson 1996)，给予相关的功能注释并同样的方法整理成数据库。对那些功能明确的基因按照其功能进行分类，分析茶树叶片中的基因表达情况。

## 3.2 结果与分析

### 3.2.1 新构建 *Sfi*?-pUC 载体的质量

以获得的在 BM25.8 菌中生成的质粒 DL003E08，经过 PCR 扩增、纯化后获得比较纯、带有 *Sfi*?的两个酶切位点的 DNA 片断，将其与载体 pUCM-T 连接后转化到 TG1 菌中，挑取阳性克隆进行培养获得质粒 DNA，纯化的质粒 DNA(图 3-1，列 1)可被 *Sfi*?比较彻底酶切(列 2，3)，只有一条带明显，切割下来的原来中间插入的小的 DNA 片段可能由于只有 200 多个碱基而并不清晰。我们把切割片段与质粒的线性形式的比较，确定切割成功。从列 3 的带型可以看出回收纯化的产物质量较好。新构建载体的长度为 2964bp(原载体 pTriplEx2 的图谱见图 3-2，序列信息见表 3-1)，新载体为原载体的两个测序引物中间的多克隆位点区域插入另外一个 pUCM-T 载体中后切掉原来的插入片断构成的。用此载体与经过酶切及分离的双链 cDNA 片段连接后转入 TG1 菌株中构成了一个新的龙井 43 的质粒文库。

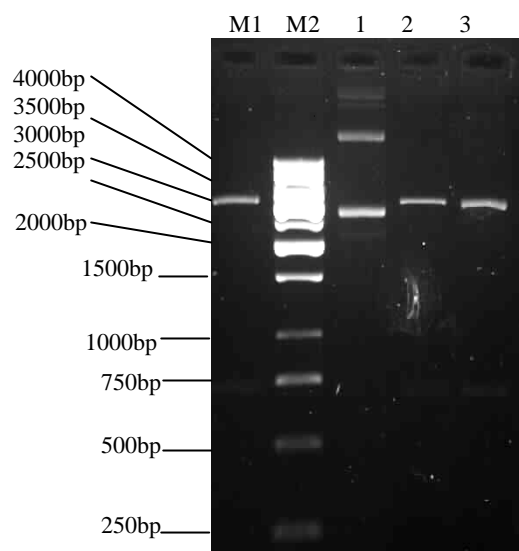
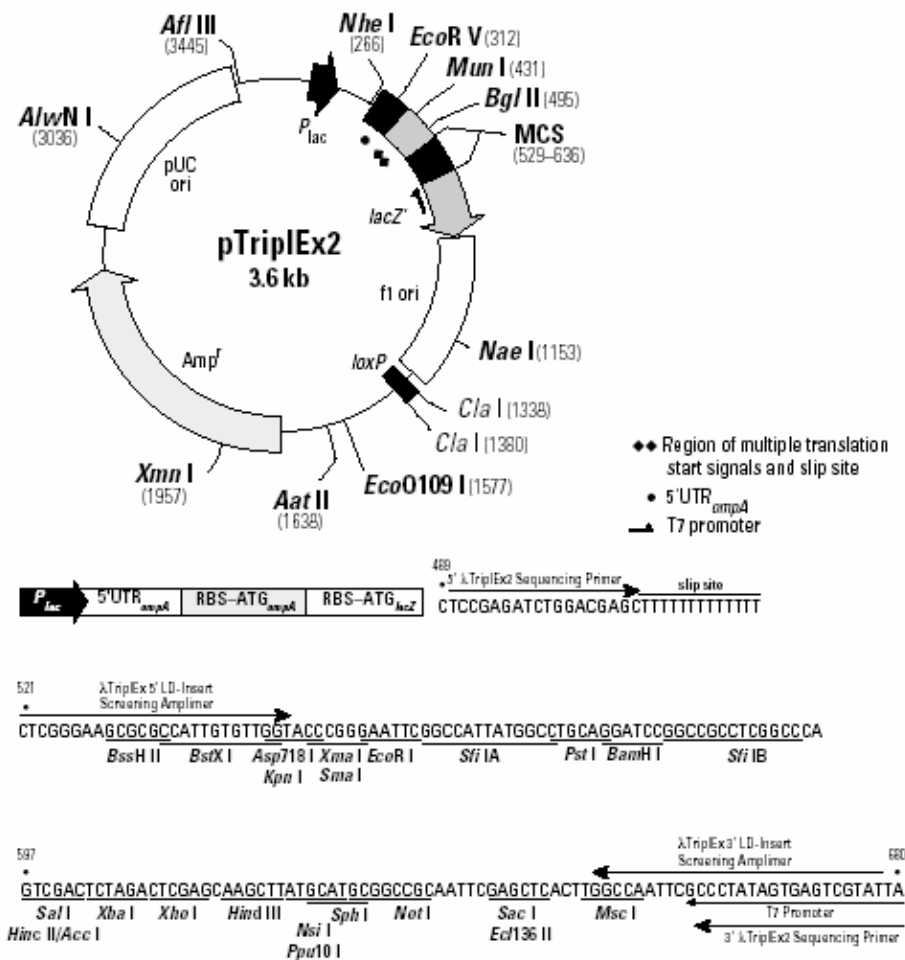


图 3-1. 重组 *Sfi*<sup>+</sup>-pUC 质粒 DNA 的 *Sfi*I 酶切鉴定  
M1, M2: 分子量标记 1. 质粒 DNA 2, 3. *Sfi*I 酶切产物  
Fig.3-1 Analysis of the *Sfi*<sup>+</sup>-pUC plasmid DNA digested by *Sfi*I in 1% agarose gel  
M1, M2: DNA maker; Lane 1: *Sfi*<sup>+</sup>-pUC DNA; Lane 2, 3: DNA digested by *Sfi*I

### 3.2.2 大规模质粒提取纯化及 EST 测序结果

选择龙井 43cDNA 文库进行进一步的测序分析。从其质粒文库中随机挑取了 4320 个单菌落进行 96 深孔板培养，共计 45 板(其中 DH12S 为 13 板，TG1 为 32 板)。经碱法抽提、纯化后的质粒溶解于 30  $\mu$ L 灭菌的 MilliQ-H<sub>2</sub>O 中。0.8% 琼脂糖凝胶电泳检测质粒 DNA(见图 3-3)的质量后，选择质量好的质粒 DNA 用 MegaBACE1000 DNA 全自动测序仪进行测序，共获得 2963 个长度大于 200 bp 并且精度大于 90% 的质量好的序列(测序结果原始图见图 3-4)，以 Chromas 软件与手工结合去除载体序列并且更正模糊的碱基，并去掉除载体后小于 150bp 的序列，获得有效序列 2547 个(表 3-2)。



Restriction map and multiple cloning site (MCS) of pTriplEx2. Unique restriction sites are bold.

图 3-2. pTriplEx2 载体部分图谱  
Fig 3-2: The partial map of vector pTriplEx2



图 3-3. 质粒 DNA 的 0.8% 琼脂糖电泳分析  
Fig.3-3. Analysis of plasmid DNA in 0.8% agarose gel

表 3-1 *Sfi* Ⅱ-pUC 载体序列

Table 3-1 The DNA sequence of *Sfi* Ⅱ-pUC vector

*Sfi* Ⅱ + pUCm Vector 2964bp

tcgcgcgtttcgggtgatgacgggtgaaacacctgacacatgcagctcccgagacggtcacagctgtctgtaagcggatgccgggagcaga  
 caagcccgatcagggcgctcagcgggtgttgccgggtgtcggggctggcttaactatcgggcatcagagcagattgtactgagagtgcacc  
 atatcggtgtgaaataccgcac agatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgaactgttggaagg  
 gcgatcgggtgcgggcctcttcgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacgccagggtttccc  
 agtcacgac **GTTGTAAAACGACGGCCAG** TGAATTCGAGCTCGGTACC **GTAATACGACTCAC** M13/pUC Sequencing Primer  
 T7 Promoter(pUCm-T)  
**TATAGGGCG** ACATATGATCGATGATATCCCATGGGCGGCCGCTGCAGACCAGGTCT **CT** pUCm-T vector  
 5' pTriplEx2 Sequencing Primer  
**CCGAGATCTGGACGAGC** TTTTTTTTTTTTCTCGGGAAGCGGCCATTGTGTTGGTACCC  
 GGGAATTC **GGCCATTATGGCC** TGCAGGATCC **GGCCGCCTCGGCC** CAGTCGACTCTAGA  
 SfiIA Insert seq SfiI B  
 CTCGAGCAAGCTTATGCATGCGGCCGCAATTCGAGCTCACTTGGCCAATTCC **CCCTATA**  
 3' pTriplEx2 Sequencing Primer pUCm-T vector  
**GTGAGTCGTATTA** AGACTGGAGATCTGGATCCCTCGAGTCTAGAGTCGACCTGCAGGC  
 M13/pUC Reverse Primer  
 ATGCAAGCTTGGCGTAATCAT **GGTCATAGCTGTTTCCTG** tgtgaaattgtatccgctcacaattccacaaa  
 catacagaccggaagcataaagtgtaaagcctggggtgcctaatgagtgaactacattaattgcgttgcgctcactgcccgctttccag  
 tcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttccgctc  
 actgactcgtcgcgtcggctgttcggctcggcgagcgggtatcagctcactcaaaaggcggtatacggttatccacagaatcaggggataa  
 cgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccggtgtcggcggttttccataggtccgccc  
 ccctgacgagcatcacaaaaatcgacgtcaagtgcagaggtggcgaaaccgacagactataaagataaccaggcggtttccccctggaagc  
 tccctcgtcgcgtctcgttccgacctgccgttaccggataacctgtccgcttttcccttcgggaagcgtggcggttttcaatgctcacgc  
 tgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccggtcagcccgaccgctgcgccttatccggtg  
 actatcgtcttgagtccaacccggtgaagacacgactatcgcc actggcagcagccactggtaacaggattagcagagcaggtatgtaggc  
 ggtgctacagagttctgaagtgggtggcctaactacggctacactagaaggacagatattgggtatctgcgctctgctgaagccagttacctcgg  
 aaaaagagttggtagctcttgatccggcaaaacaaccaccgctggtagcgggtggtttttgttgcaagcagcagattacgcgcagaaaaaa  
 ggatctcaagaagatcctttgatcttttctacggggtcagctcagtggaacgaaactcacgttaagggtatttgggtcatgagattatcaaaa  
 aggatcttcacatagatccttttaaaataaaatgaagttttaaatcaatcaaaagtatatatgagtaaaactggctgacagttaccaatgcttaatca  
 gtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggttaccat  
 ctggccccagtgctgcaatgataccgcgagaccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagc

gcagaagtggctcgtcaactttatccgcctccatccagctatttaattgttccgggaagctagagtaagtagttgccagttaatagtttgcga  
acgttgttgcattgctacaggcatcgtggtgtcacgctcgtcgttggatggcttcattcagctccggtccaacgatcaaggcgagttacat  
gatccccatgttgtgcaaaaaagcggttagctccttcggctcctccgatcgttgtcagaagtaagttggccgagtggtatcactcatggttatgg  
cagcactgcataattcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatcggg  
cgaccgagttgcttgc ccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgtcatcattggaaaacgttcttcgg  
ggcgaaaactctcaaggatcttaccgctgttgagatccagttcgtatgaaccactcgtgcaccaactgatcttcagcatcttttactttaccag  
cgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaaggggaataagggcgacacggaaatgttgaatactatactcttctctttt  
caatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataacaaataggggtccgcgcacattt  
ccccgaaaagtgccacgtgacgtctaagaaaccattattatcatgacattaacctataaaaaataggcgatatcacgaggcccttcgtc

表 3-2 龙井 43EST 序列分析结果

Table 3-2 Analysis results of longjing 43 EST sequences

序列类别 Sequence catalogs	克隆数 Clones
提取出的有用序列 Useful sequences	2963
空载和去载体后<150bp Shorter than 150bp	416
有效序列 Sequences greater than 150bp	2547
重复 Repeats	863
获得的有效 ESTs Valid ESTs	1684
重复率(重复数/有效序列) Repeat ratio	33.88%

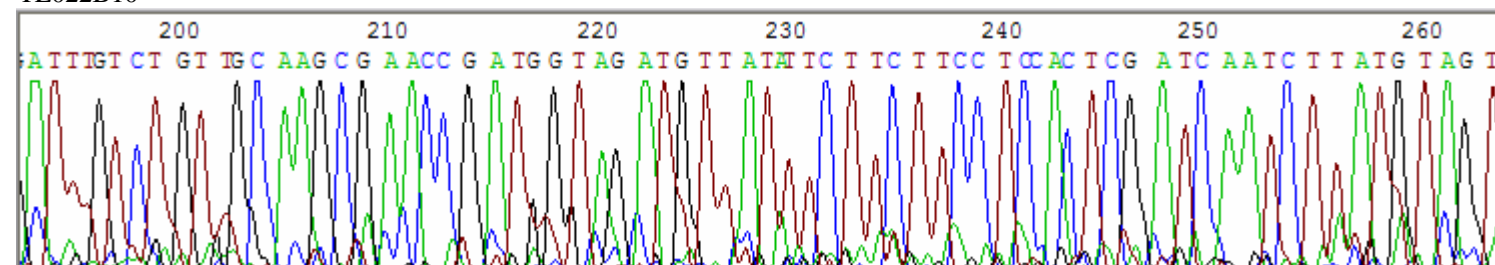
有效序列中,最长为 809 bp,最短为 150 bp,平均长度为 478 bp。其中 3'端具有 Poly(A)的序列有 511 条,占有效序列的 30.3%(表 3-3)。有 250 个序列具有左右两侧的载体序列,说明这些序列已经测通,这部分序列占总序列的 15%左右,占 3'端具有 Poly(A)序列的 50%左右。序列长度的分布图 3-5,可以看出,整体测序质量较好,只有 4%的序列小于 200bp;序列长度在 600-700 bp 之间约占 24.3%;67%以上的序列大于 400bp;完全符合进一步研究的要求。

表 3-3 EST 序列末端信息

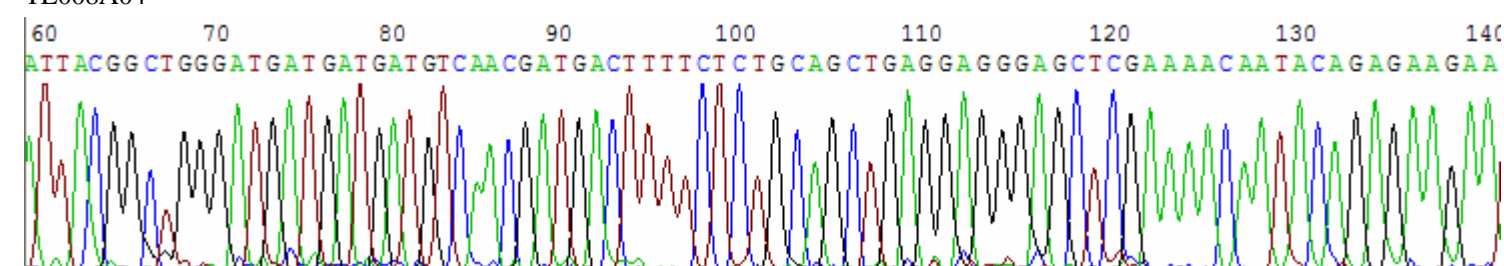
Table3-3 The poly(A) information of EST sequences

	EST个数 Number	百分比 Percent(%)
Poly(A)	511	30.3
No poly(A)	1173	69.7
Total	1684	100

TL022B10



TL008A04



TL016A01

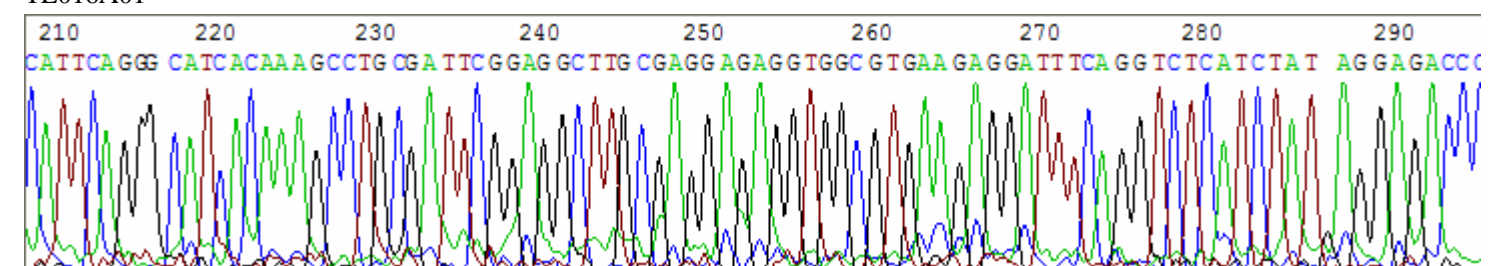


图 3-4. 测序结果示意图

Fig. 3-4 Sequencing sketch map of the EST

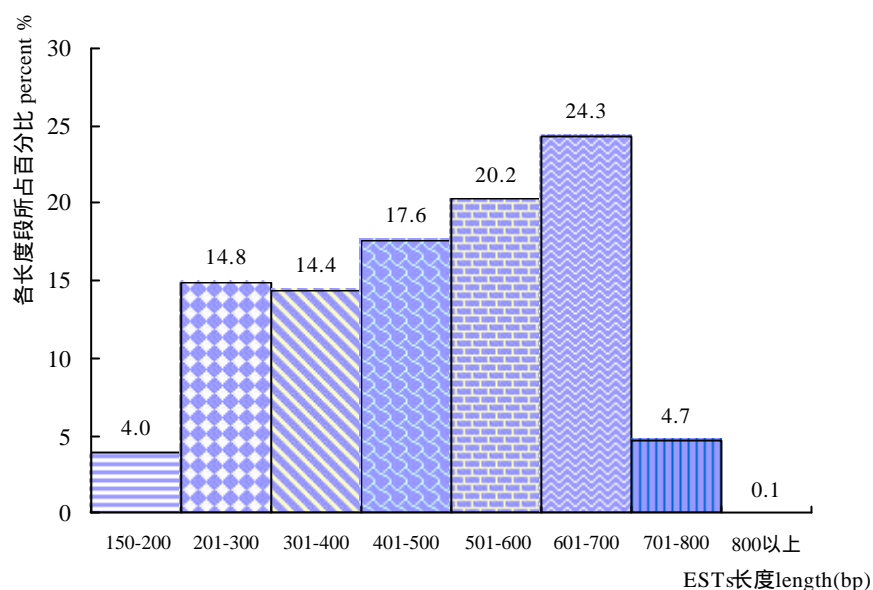


图 3-5 EST 长度的分布图  
Fig. 3-5 The length distribution of tea plant ESTs

### 3.2.3 序列提交

将 1643 个有效的序列转化成 FASTA 格式的文本文件，进而转为 GenBank 要求的 EST 序列的提交格式后（见附件 1）分批提交到 GenBank 数据库。目前正在受理中。

### 3.2.4 EST 序列的同一性分析

用 DNAtools 软件对所有有效序列进行同一性分析，参数设定为 24bp，结合手工操作，发现有 863 个序列为与其它序列重复的序列，即与其它的序列基本相同或者属于其它序列的一部分。将这些冗余序列去除后，最后获得 1684 个高质量的非冗余序列进行 BlastN 和 BlastX 比对分析。

### 3.2.5 与 NCBI 数据库的 BLAST 比对结果

将 1684 个 EST 序列逐条与 GenBank(<http://www.ncbi.nlm.nih.gov>)的核酸数据库进行 BlastN、BlastX 比对分析，对 BlastN 结果分值(score 值)大于 100 的序列及 BlastX 中 E 值<0.02 的序列，给予相关的功能注释，最终将所有序列分为 5 种类型：(1)功能已知基因(known genes)；(2)具有推测功能的基因(putative genes)；(3)已知定位信息的 ESTs(locational genes)；(4)未知 mRNA(unknown mRNA)；(5)新基因(Novel genes)，见表 3-4。

表 3-4 ESTs 与 NCBI数据库 BLAST 比对结果

Table 3-4 The BLAST results of 1684 ESTs against NCBI database

基因类别 Gene catelogs	基因数目 Gene numbers		占总克隆数目的百分比(%) Percent of total clones	
	BlastN	BlastX	BlastN	BlastX
已知功能基因 Known genes	481	898	28.7	53.3
具有推测功能的基因 Putative genes	89	110	5.3	6.5
已知染色体定位基因 Locational genes	12	-	0.7	-
未知 mRNA Unknown mRNA	25	192	1.5	11.4
新基因 Novel genes	1077	484	63.8	28.7
合计 Total	1684	1684	100	100

从表 3-4 中可以看出,在本研究所测到的 1684 个基因 EST 序列中,BlastN 比对结果中,与网上数据库同源性很低的序列为 1077 个,占总序列的 63.8%,大多数基因为新基因;只有 12 个为染色体定位的序列,占总序列的 0.7%;而比对注释为未知 mRNA 的序列还有 25 个,以上 3 个类型的序列合计约有 66%,也即龙井 43 中有 66%的表达基因为功能未知基因。而 BlastX 比对结果中,新基因的数目是 484 个,只占总数的 28.7%。

根据 BLAST 比对结果,分别构建了龙井 43 核酸、蛋白质比对的两个独立基因数据库,所包含的内容及格式见表 3-5(1)(2)。两个数据库都是根据 BLAST 获得的结果,自己编制程序提取有用信息构成的,直观地反应任一 EST 的信息。

### 3.2.6 茶树 ESTs 比对上的物种分布

具已知或推测功能的607条ESTs,通过与NCBI的非冗余蛋白质和核酸数据库BLASTX与BLASTN比对,共得到12 755条比对结果,其中茶树只有80条(表3-6);比对涉及的物种(变种)为344 个,主要为下面物种,如拟南芥、水稻(籼稻)等。

在BlastN的比对结果中,已知功能序列中有部分与GenBank中茶树已克隆基因序列高度相似,这些序列见表3-7。



表 3-5(1) 龙井 43 新梢 EST BlastN 结果统计数据库

Table 3-5(1) Database of blastN results of tea plant

克隆 Clones	最高同源序列基因登录号 GenBank accessions No.	物种 Species	功能注释 Functional annotations	分值 Score	E 值 e-value
TL024D03	gi 3204128 emb AJ006768.1 CAR676	鹰嘴豆 <i>Cicer arietinum</i>	组蛋白 H2A mRNA for histone H2A Length = 685	206.00	3e-50
TL024D05	gi 15149874 emb AJ277599.2 QSU27759	软木橡树 <i>Quercus suber</i>	金属硫蛋白 mRNA for metallothionein-like protein (mt gene)Length = 309	234.00	1e-58
TL024E02	gi 21490 emb X04401.1 STLS1	马铃薯 <i>Solanum tuberosum</i>	光诱导组织特异性 ST-LS1 基因 For light inducible tissue-specific ST-LS1 gene Length = 560	212.00	6e-52
TL024E10	gi 2565418 gb AF026803.1 AF02680	红豆草 <i>Onobrychis viciifolia</i>	组蛋白 H3 Histone H3 mRNA, complete cds Length = 590	266.00	4e-68
TL024F03	gi 9967142 dbj AB031280.1	茶树 <i>Camellia sinensis</i>	咖啡因合成酶 1 TCS1 mRNA for caffeine synthase, complete cds Length = 1438	1088.00	0.0
TL024F05	gi 2746720 gb AF038387.1 AF03838	甜椒 <i>Capsicum annuum</i>	组蛋白 H4 Histone H4 mRNA, complete cds Length = 483	115.00	8e-23
TL024F09	gi 32984510 dbj AK099301.1	粳稻 <i>Oryza sativa</i> cv. <i>japonica</i>	(japonica cultivar-group) cDNA clone:J023136B12, full insert sequence Length = 381	107.00	2e-20
TL024G02	gi 882241 gb U19738.1 FLU1973	茄子 <i>S. melongena</i>	碳酸酐酶 1 Carbonic anhydrase 1 mRNA, complete cds Length = 1152	101.00	2e-18
TL024G07	gi 5706703 gb AF112440.1 AF11244	一枝黄花 <i>Flaveria linearis</i>	核糖体蛋白 S26 Ribosomal protein S26 (RPS26) mRNA, complete cds Length = 610	218.00	9e-54
TL024G11	gi 2677823 gb U93164.1 PAU9316	豌豆(pea) <i>Pisum sativum</i>	脱落胁迫成熟蛋白 Absciscic stress ripening protein homolog mRNA, complete cds Length = 976	101.00	1e-18

表 3-5(2) 龙井 43 新梢 EST BlastX 结果统计基因数据库

Table3-5(2) Database of blastX results of tea plant

克隆 Clones	长度 bp length	最高同源序列基因登录号 GenBank accession No.	物种 Species	功能注释 Functional annotation	分值 score	E 值 e-value
TL003E05	584	gi 7387849 sp O04974 LU1B_LYCPN	<i>Lycopersicon pennellii</i> 野生番茄	2-isopropylmalate synthase B (Alpha-isopropylmalate synthase B) 2 - 异丙基苹果酸合酶	261	7.00E-69
TL025F07	602	gi 2344781 dbj BAA21897.1	<i>Ipomoea nil</i> 裂叶牵牛	2-oxoglutarate 3-dioxygenase; flavanone 3-hydroxylase; naringenin 2 - 酮戊二酸 3 - 双加氧酶	167	8.00E-41
DL008A07	367	gi 18417346 ref NP_568299.1	<i>Arabidopsis thaliana</i> 拟南芥	30S ribosomal protein S13, chloroplast precursor (CS13) 30S 核糖体蛋白 S13	61.6	3.00E-09
TL025H03	637	gi 30841448 gb AAP34362.1	<i>Gossypium barbadense</i> 海岛棉	Fiber protein Fb15 光纤蛋白 15	141	9.00E-33
TL022A07	559	gi 38154485 gb AAR12194.1	<i>Nicotiana benthamiana</i> 烟草	molecular chaperone Hsp90-2 分子伴侣 Hsp90-2	160	1.00E-57
TL028C11	689	gi 9367321 emb CAB97360.1	<i>Juglans nigra</i> 黑核桃	Flavanone 3-hydroxylase 黄烷酮 - 3 - 羟化酶	258	1.00E-79
TL012E02	299	gi 29123528 gb AAO63022.1	<i>Allium cepa</i> 洋葱	Flavanone 3-hydroxylase 黄烷酮 - 3 - 羟化酶	54.3	5.00E-07
TL011B09	454	gi 29123532 gb AAO63023.1	<i>Allium cepa</i> 洋葱	Flavonol synthase 黄酮醇合成酶	55.8	4.00E-13
TL003D05	687	gi 14916566 sp Q9ZWQ9 FLS_CITUN	<i>Citrus unshiu</i> 温州蜜柑	Flavonol synthase (FLS) (CitFLS) 黄酮醇合成酶 FLS	127	2.00E-28
TL007G04	699	gi 15240834 ref NP_196382.1	<i>Arabidopsis thaliana</i> 拟南芥	Formin homology 2 (FH2) domain-containing protein 形成素同族体 2 包含蛋白	177	1.00E-43
TL004F10	315	gi 21592495 gb AAM64445.1	<i>Arabidopsis thaliana</i> 拟南芥	Fructokinase-like protein 果糖激酶类似蛋白	62	2.00E-09
TL021F04	453	gi 27728731 gb AAO18639.1	<i>Actinidia deliciosa</i> 猕猴桃	Galactose dehydrogenase 半乳糖脱氢酶	132	6.00E-31

表3-6 茶树ESTs比对上的物种信息统计

Table 3-6 The species distribution of tea plant ESTs BLAST result

物种 species	核酸序列数 nucleotide sequences	占总物种百分比 Percent%
拟南芥 <i>Arabidopsis thaliana</i>	4 294	33.7
籼稻 <i>Oryza sativa</i> ssp. <i>Indica</i>	1 004	7.9
烟草 <i>Nicotiana tabacum</i>	443	3.5
木烟草 <i>Nicotiana sylvestris</i>	401	3.1
大豆 <i>Glycine max</i>	368	2.9
马铃薯 <i>Solanum tuberosum</i>	367	2.9
甜玉米 <i>Zea mays</i>	365	2.9
番茄 <i>Lycopersicon esculentum</i>	345	2.7
矮牵牛 <i>Petunia hybrida</i>	212	1.7
小鼠 <i>Mus musculus</i>	206	1.6
豌豆 <i>Pisum sativum</i>	162	1.3
松叶菊属 <i>Mesembryanthemum crystallinum</i>	148	1.2
陆地棉 <i>Gossypium hirsutum</i>	129	1.0
小麦 <i>Triticum aestivum</i>	104	0.8
茶树 <i>Camellia sinensis</i>	80	0.6
其它(单个物种少于 100 条)	4 127	32.4
Others (less than 100 entries for individual species)		
合计 Total	12 755	100

表3-7. 比对上的茶树已克隆基因

Table 3-7 BLAST matched cloned genes of tea plant in the GenBank

克隆号 Clone ID	功能描述 Functional annotation
TL018D10	ITS1, 5.8SrRNA, ITS2, 28S rRNA
TL004C05	5S 核糖体 RNA(5S rRNA)
TL011G03	叶绿体上转移 Leu 的 tRNA(chloroplast tRNA-Leu gene for transfer RNA-Leu)
TL022H10	$\beta$ -1,3-葡萄糖苷酶( $\beta$ -1,3-glucosidase)
TL007A08	$\beta$ -樱草糖苷酶( $\beta$ -primeverosidase)
TL016G12	成熟酶(maturase)
TL012A03	查尔酮合成酶 1(chalcone synthase 1)
TL025D10	查尔酮合成酶 2(chalcone synthase 2)
DL007F02	查尔酮合成酶 3(chalcone synthase3)
TL027F02	S-腺苷甲硫氨酸合成酶(S-adenosylmethionine sythetase)
TL014D02	咖啡因合成酶(caffeine synthase)

### 3.2.7 高丰度表达的基因

植物表达的基因因组织、器官和发育阶段而异，不同cDNA文库的基因具有不同的时空和组织特异性。在茶树新梢cDNA文库中与细胞生长/分化、能量有关的ESTs非常丰富，如与光合作用直接有关4个基因的ESTs就达到115个，占总数的6.8%(表3-8)。

表3-8 茶树新梢cDNA文库的一些高丰度基因

Table3-8 Some abundant genes in the tea young shoots cDNA library

高丰度基因Abundant Genes	EST 数量numbers
组蛋白 Histone	59
类金属硫因蛋白 Metallothionein-like protein (mt gene)	42
核酮糖1,5-二磷酸羧化/加氧酶 Ribulose biphosphatecarboxylase/oxygenase (Rubisco)	42
叶绿素a/b结合蛋白 Chlorophyll a/b-binding protein(CAB)	37
捕光叶绿素a/b结合蛋白 light harvesting chlorophyll A/B binding protein(Lhcb)	19
光诱导组织特异性ST-LS1蛋白 Light inducible tissue-specific ST-LS1	17

### 3.2.8 BlastX 比对具已知功能和具推测功能 EST 的分析

为对茶树的基因表达谱有一个更加深入的了解，对已有功能注释的基因进行分类。按照拟南芥功能分类标准(The EU Arabidopsis Genome Project., 1998)(表 3-9)，把 BlastX 结果中具有已知功能或推测功能的 1008 个 EST 根据所参与的细胞过程将其分为 12 个大类(表 3-10)：能量代谢相关基因(Energy)、蛋白质合成相关基因(Protein synthesis)、细胞结构相关基因(Cell structure)、细胞抗性及防御相关基因(Disease and defence)、代谢相关基因(Metabolism)、信号转导相关基因(Signal transduction)、细胞内运输相关基因(Transporters)、转录相关基因(Transcription)、次生代谢相关基因(Secondary metabolism)、细胞生长发育相关基因(Cell growth and division)、蛋白质降解相关基因(Protein destination and storage)和功能不明确基因(Unclear classification and unclassified)。

表 3-9 植物基因的功能分类

Table 3-9 Functional catalogues of plant genes

01 Metabolism	02 Energy	03 Cell growth/division	04 Transcription
01.01 Amino acid	02.01 Glycoslysis	03.01 Cell growth	04.01 rRNA synthesis
01.02 Nitrogen and sulphur	02.02 Gluconeogenesis	03.13 Meiosis	04.10 tRNA synthesis
01.03 Nucleotides	02.07 Pentose phosphate	03.16 DNA synth/replication	04.19 mRNA synthesis
01.04 Phosphate	2.10 TCA pathway	03.19 Recombination/repair	04.1901 General TFs
01.05 Sugars and polysaccharides	2.13 Respiration	03.22 Cell cycle	04.19.4 Specific TFs
01.06 Lipid and sterol	2.16 Fermentation	03.25 Cytokinesis	04.1907 Chromatin modificarion
01.07 cofactors	2.20 E-transport	03.26 Growth regulators	04.22 mRNA processing
	2.30 Photosynthesis	03.99 other	04.31 RNA transport
05 Protein synthesis	06 Protein destination and storage	07 Transporters	08 Intracellular traffic
05.01 Ribosomal proteins	06.01 Folding and stability	07.01 Ions	08.01 Nuclear
05.04 Translation factors	06.04 Targetting	07.07 Sugars	08.02 Chloroplast
05.07 Translation control	06.07 Modification	07.10 Amino acids	08.04 Mitochondrial
05.10 tRNA synthases	06.10 Complex assembly	07.13 Lipids	08.07 Vesicular
05.99 Others	06.13 Proteolysis	07.16 Purins/pyrimidines	08.10 Peroxisomal
	06.20 Storage proteins	07.22 Transport ATPases	08.13 Vacuolar
		07.25 ABC-type	08.16 Extracellular
		07.99 Others	08.19 Import
			08.99 Others
09 Cell structure	10 Signal transduction	11 Disease/defence	20 Secondary metabolism
09.01 Cell wall	10.01 Receptors	11.01 Resistance genes	20.1 Phenylpropanoids/phenolics
09.04 Cytoskeleton	10.04 Mediators	11.02 Defence-reglated	20.2 Terpenoids
09.07 ER/Golgi	10.0404 Kinases	11.03 Cell death	20.3 Alkaloids
09.10 Nucleus	10.0407 P hosphatases	11.04 Cell rescue	20.4 Non-protein amino acids
09.13 Chromosomes	10.0410 G proteins	11.05 Stress responses	20.5 Amines
09.16 Mitochondria	10.99 Others	11.06 Detoxification	20.6 Glucosinolates
09.19 Peroxisome		11.99 Others	20.99 Others
09.25 Vacuole			
09.26 Chloroplast			
09.99 Others			
	12 Unclear classification	13 Unclassified	14 Transposons
			14.01 LTR retroelements
			14.02 Non-LTR retroelements
			14.99 Other
Fifteen putative cellular roles for genes in plants are shown that are broadly based on the yeast functional catalogue, New categories, adapted for plant-specific roles, have been made for secondary product metabolism and disease, defence and stress responses, and several new subcategories of genes involved in photosynthesis, chloroplast structure, storage processes and polysaccharide metabolism have been established. Categories between 14-19 have been reserved for gene categories in other organisms.			

表 3-10 已知功能或具有推测功能的基因分类

Table3-10 The classification of genes with function annotation or putative function annotation

分类信息	所测基因数目	占功能基因百分比	占有效 EST 百分比
Functional category	genes	% of functional genes	% of total ESTs
能量 Energy	266	26.39	15.80
蛋白质合成 Protein synthesis	180	17.86	10.69
细胞结构 Cell structure	108	10.71	6.41
抗病及防御 Disease and defence	93	9.23	5.52
代谢 Metabolism	84	8.33	4.99

信号传导 Signal transduction	54	5.36	3.21
胞内运输 Transporters	47	4.66	2.80
转录 Transcription	39	3.87	2.32
次生代谢 Secondary metabolism	35	3.47	2.08
细胞生长分裂 Cell growth and division	38	3.77	2.26
蛋白质降解 Protein destination and storage	17	1.69	1.01
分类不清 Unclear classification	47	4.66	2.79
合计 Total	1008	100	59.86

从表 3-10 中可以看出，编码能量代谢相关基因的 EST 数量最多，共有 266 个，占有功能描述基因的 26.39%。在此类基因中，编码与光合作用相关基因的 EST 序列最多，其中主要是叶绿素 a/b 结合蛋白和 1,5-二磷酸核酮糖羧化酶(表 3-8)，因为本研究取材是春天新长出的幼嫩新梢，其细胞生长旺盛，细胞分裂和伸长加速，为满足自身生命活动的需要，光合作用加强，所以与光合作用相关的基因表达丰度较高。另外还有与呼吸作用、三羧酸循环等相关的基因。

与蛋白质合成相关的 EST 数量为 180 个，仅低于能量相关的基因。其中表达丰度最高的是核糖体蛋白。核糖体是蛋白质合成的场所，所有蛋白质的合成都需要在核糖体上进行。真核生物的核糖体的沉降系数为 80S，由 60S 的大亚基和 40S 的小亚基组成。40S 的小亚基由 18S 的 rRNA 分子和 30 种不同的核糖体蛋白质组成，60S 大亚基由各一分子的 5SrRNA 分子、5.8SrRNA、28SrRNA 和约 45 种蛋白质组成。蛋白质分子主要分布在核糖体表面，而 rRNA 位于内部，二者靠非共价键结合在一起，核糖体中蛋白质约占 40%，RNA 约占 60%。本研究中检测到大量的核糖体蛋白。核糖体常常分布在细胞蛋白质合成旺盛的区域，其数量和蛋白质合成强度有关。在茶树新梢中，储藏蛋白和其它蛋白的合成反应的强度应该是比较高的，所以需要合成大量的核糖体蛋白。

本研究中检测到与细胞结构相关的 EST 有 108 个，其中主要包括组蛋白、微管蛋白、肌动素合成相关蛋白等。组蛋白是真核生物与 DNA 一起组成染色质的主要组分，编码组蛋白的基因普遍存在与低等植物和高等植物中，组蛋白的功能是在染色质中形成核小体结构，同时也参与基因的表达调控过程。组蛋白有 5 类：H1、H2A、H2B、H3、H4，其中 H2A、H2B、H3、H4 组蛋白组成核小体的八聚体核心部分。H3、H4 为多基因家族，成员较多，本研究中 5 种组蛋白都被检测到，并且 H3、H4 的丰度相对较高。

从表中可以看出参与抗病、抗逆过程的防御性蛋白基因的 EST 达到 93 个，占有功能描述基因的 9.23%，说明茶树的新梢具有一套抵御外界不良环境及病虫害的防御体系。

编码代谢基因的 EST 有 84 个，占有功能描述基因的 8.33%，它们大多数是与代谢相关的酶类。其中酶的种类很多且参与不同的功能，不完全属于一条代谢途径。没有出现相对丰度较高的基因。

植物抗病、抗逆反应涉及到复杂的信号转导，本研究中与信号转导相关的基因的 EST 有 54 条，约占有功能描述基因的 5.36%。如此高比例的信号转导基因的表达暗示了植物与

动物相似,存在复杂的信号转导机制,可以对内外环境的变化作出应答,以调整机体状态来适应环境的变化。其中泛在蛋白和多聚泛在蛋白的表达丰度最高,在真核细胞中,泛在蛋白最主要的功能是作为识别信号与目的蛋白结合并调节蛋白质降解。泛在蛋白是一由 76 个氨基酸残基组成的高度保守蛋白质,普遍存在于所有真核生物的细胞内。泛在蛋白伸展/融合蛋白基因在分生组织中表达丰度较高,而在成熟组织中明显降低;而聚泛在蛋白基因在所有植物的器官中组成性表达,在幼嫩组织中表达水平较高,这与本文得到的结果完全相符。

与细胞内运输机制相关的基因有 47 个,其中主要水分通道蛋白、质膜 ATP 酶( $H^+$ -ATPase)以及跨膜蛋白等。

与转录相关的 EST 有 39 个,其中表达丰度较高的有 Myb 蛋白及 zinc-finger 蛋白等。另外与细胞生长相关的 EST 种的蛋白主要是硫氧还蛋白和细胞周期蛋白等。转录因子的研究是目前分子生物学研究的热点,植物转录因子在植物的生长发育及其对外界环境的反应中起着重要的作用(Zhang et al., 1999; Després et al., 2000)。但迄今为止大部分高等植物的转录因子功能尚未知。研究发现,真核生物基因组中转录因子占相当大比例,根据基因组计划所得数据,估计果蝇(*Drosophila*)、线虫(*Caenorhabditis elegans*)和拟南芥(*Arabidopsis thaliana*)的基因组中的转录因子分别约占其基因总量的 5%、2.7%和 5%。本研究中茶树新梢组织中与转录相关的 EST 所占比例为 3.87%,与前面几个物种的估计值基本相符。

茶树叶片中有很多次生代谢产物,如茶多酚、儿茶素、咖啡碱、多糖等等(Hertog et al., 1993),其中仅茶多酚类就占茶叶干重的 20%-35%,它们在防御紫外线、作为植物-微生物相互作用的信号分子和植物防御反应中起着重要的作用。本实验检测到的与次生代谢相关的基因有 35 个,占 3.47%,其中主要有黄烷酮合成酶、查尔酮合成酶、咖啡因合成酶等。

## 3.3 讨论

### 3.3.1 新建载体的质量

在 T 载体的基础上引入原载体序列的 *Sfi*I 两个酶切位点,构建了一个新的载体 *Sfi*I-pUC 载体。因为 T 载体具有不需对引物进行特殊酶切位点的修饰,即能直接进行 PCR 产物的克隆,并具有克隆效率高的特点。在 pUCM-T 载体中,外源片断的克隆将打破载体中的  $\beta$  半乳糖苷酶的编码序列,因此重组子可以在指示平板上用颜色反应加以筛选。将 PCR 产物克隆到 pUCM-T 载体后转入大肠杆菌 TG1 感受态细胞中后,筛选了有外源片断插入的白斑进行培养然后酶切获得需要的载体。对于新建的载体,其用颜色反应筛选重组子的功能不能确定,因为在 T 载体上加入的片断的长度不是 3 核苷酸的倍数,且插入位点在  $\beta$  半乳糖苷酶基因的阅读框内,已经破坏了此基因的阅读框,因此新构建的载体不能用来进行颜色筛选,即使出现白斑现象也不能确定其不是重组子。载体上有两个序列相同的但是方向不同的引物位点,即用单引物就可以扩增出插入片断,这样就造成了此引物不能用来进行测序,这是新载体需要继续改进的地方。构建的新的龙井 43 的质粒文库所包含的克隆数较多(数据未统计)。用此文库进行测序可以一定程度上避免产生过多的影响 EST 的分析及组织中的基因表达情况的冗余序列。

### 3.3.2 EST 的 Poly(A)特性

EST 测序是目前应用最广泛的寻找新基因的手段。从统计学分析,为获得某组织中主要表达基因,至少需要 1000 条 ESTs(Audic et al., 1997)。本研究共获得了 1684 个高质量的茶树新梢 EST,故基本可以反应出新梢的基因表达情况。

3'端非翻译区在调节基因表达中起重要作用,而且可以特异性地代表生物体某些组织某个时期的一个表达基因(Coulson et al., 1997; Wu et al., 2000),而 5'端非翻译区的长度和碱基顺序在不同的生物和基因中变化很大。最初选择对特异性强的 3'端进行测序,得到的大部分序列具有 poly(A)尾结构(数据未详细统计)。改变载体后由于 3'端的引物序列距离插入片段较远,影响测序质量,因此选择进行 5'端测序,仍有一部分序列(30.3%)获得了 poly(A)尾结构,这部分序列即为测通序列。对龙井 43 新梢 cDNA 文库随机选取 4320 个克隆的序列进行测定,可以代表文库中基因的分布。测序获得有用序列为 2963 个,测序成功率为 68.6%,其平均长度为 478bp。为了获得高质量的序列,剔除了测序误差大于 5%和除去载体序列后长度小于 150bp 的序列,并去除了核糖体 RNA、叶绿体和线粒体以及载体序列的污染。

### 3.3.3 BlastN 与 BlastX 比对结果的差异

所有 1684 EST 与 NCBI 的核酸数据库比对时,有 36.2%的 EST 序列显示了与已知序列很大的同源性;而与其蛋白质数据库进行比对时,有 71.3%的序列显示了与已知序列很大的同源性。很显然,这些值都是暂时的,会随着数据库中数据信息的增加而得到提高。核酸比对的其余 63.8%为未知功能新基因,这与 1997 年 1.5 万个拟南芥不重复 ESTs 中有 60%为未知新基因(Delseny et al., 1997),水稻 ESTs 研究中,1996 年和 1997 年的 EST 中均有 75%为新基因(Yamamoto et al., 1997)基本相似。由于核酸与蛋白质两个数据库的范围不完全一致,BlastN 的范围是 All GenBank+EMBL+DDBJ+PDB sequences (but no EST ,STS ,GSS ,or phase 0 , 1 or 2 HTGS sequences),而 BlastX 的范围是 All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF,所以其获得新基因的数量也有很大的差异。一些 EST 分析者采用的是 3'端测序并进行分析,他们在进行与数据库比对时,两个数据库比对的结果中,从核酸数据库比对确定新基因的数量少于蛋白质比对获得的数量(Carson et al., 2002; Michael et al., 2003),他们认为那些与蛋白质数据库没有明显的同源序列而在核酸数据库中有明显的同源序列是因为这些序列长度太短,可能不包括蛋白质编码序列造成结果的差异。本文的比对结果中只有 16 个是与蛋白质没有比对上而与核酸比对上的 EST 序列,即仍有很大一部分 EST 序列是与蛋白质序列有明显的同源性而与核酸序列没有明显的同源性,这可能是由于 3'端测序获得的序列含有较长的非编码区,一般在 250bp 左右。Michael 等(2003)的研究表明,序列长度越短,与蛋白质数据能比上的有明显功能的可能性就越低,而本研究大部分是进行的 5'端测序,并且序列的平均长度为 478bp,这些 5'端序列所包含的非编码区



序列相对较短,大部分会包括蛋白质的编码序列,因此与蛋白质数据库有明显同源性的序列要多一些。具体结果还有待于进一步的分析。

### 3.3.4 EST 测序分析的效率

从 BlastN 比对结果中可以看出,与本文 EST 比对上的物种最多的是拟南芥(4294 条),其次是水稻(1004 条),而茶树只有 80 条,这个结果反应了在 GenBank 的核酸数据库中不同植物序列的丰富程度,也可以看出数据库中茶树的序列很少。还有一些是非植物序列,包括酵母、细菌和人类等的序列有一定同源性,表明这些序列可能含有高度保守序列。

经 BlastN 比对所获得具有明确或推测功能的 607 个 EST,可以归类为 130 多个茶树功能基因(数据未详细给出)。其中已经获得了茶树中大部分(80%以上)已克隆的基因的 ESTs。茶树功能基因的克隆始于 1990 年代,到目前为止国内外在 GenBank 登录的基因才 20 多个(不包含我们的序列)。显而易见,EST 测序分析在发现、鉴定茶树功能基因中的效率。

### 3.3.5 新梢表达基因的丰度特征

通过随机测序所获得同一基因的 EST 数目在一定程度上可代表该基因在该组织中的表达丰度。一般而言,高丰度表达的基因常常是细胞特异性的基因,而大多数看家基因呈中低表达丰度。以往的大量经典的生化研究表明(Bisshop et al., 1974),大多数已分化的哺乳动物细胞通常有 10000-30000 个基因表达,其中高丰度基因约占总 mRNA 的 20%以内,中等丰度的基因约有几百种,约占总 mRNA 的 30-40%,剩下低丰度的基因约占 50%。所获得序列中通过 DNATool 软件去除的重复序列有 863 个,占总获得序列的 33.88%。当去除重复序列后进行 BLAST 分析时,丰度较高的基因(表达频率=5)约有 30 个,约占总数的 10%,中等丰度基因(表达频率在 2-5 之间)约 100 个,占总数的 30%,其余低丰度的基因约占 60%,这与 Bisshop 等的结果基本相符,即大多数基因呈中低丰度表达。

### 3.3.6 新梢的主要功能基因分类

由于目前对茶树基因的研究较少,GenBank 数据库中的茶树功能基因很少,所以本研究直接与 GenBank 上的总核酸数据库和蛋白质数据库进行查询比较,结果获得的新基因就意味着可能是在茶树新梢组织中特异表达的基因。结合拟南芥基因功能分类标准(The EU Arabidopsis Genome Project., 1998),把茶树新梢组织中有功能描述的基因分为 12 类。通过对所有已知功能基因的分析发现:与编码能量代谢相关基因的 EST 数量最多,共有 266 个,占有功能描述基因的 26.39%,其次是与蛋白质合成有关的 EST 占功能描述基因的 17.86%,与代谢相关的 EST 占 8.33%。在能量代谢相关基因中,编码与光合作用相关的基因的 EST 序列最多,另外还有与呼吸作用、三羧酸循环等相关的基因,这可能主要是因为的研究材料

是新生的茶树新梢，其细胞生长和分裂活跃，需要大量的物质合成和代谢以及较强的光合作用以满足其生长发育的需要。另外，与细胞结构、抗性和信号转导相关的 EST 含量也较高，分别占 10.71%、9.23%和 5.36%。与次生代谢相关的基因有 35 个，占 3.47%，其中主要有黄烷酮合成酶、查尔酮合成酶、咖啡因合成酶等。这些都是影响茶叶品质的重要酶类，为我们今后对茶树次生代谢研究提供了一些序列信息。

## 第四章 总结与展望

随着分子生物学技术迅速发展,很多种作物已经建立了比较完整的分子生物学的实验体系,而茶树在分子生物学尤其在功能基因组学研究方面一直处于滞后状态。日本的 Matsumoto、Takeuchi 等(1994) 构建了一个茶树的表达基因文库(cDNA 文库)并利用此文库获得了一些茶树上的重要基因(Takeuchi et al., 1994; Matsumoto et al., 1994; 竹内敦子等, 1994),而其它已经克隆的茶树基因都是通过经典基因克隆方法获得,但对茶树功能基因组研究还几乎处于空白状态。本研究主要在以下几个方面取得了重要的进展。

### 4.1 建立了 2 个高质量的 cDNA 噬菌体文库

我们以 Trizol 一步法从春茶新梢中提取总 RNA,分离纯化富含 Poly A 的 mRNA,LD-PCR 反转录合成双链 cDNA,以  $\lambda$ Triplex2 为载体,构建了龙井 43、安吉白茶等两个茶树品种的 cDNA 文库,滴度分别为  $6.8 \times 10^5$  pfu/mL 和  $4 \times 10^5$  pfu/mL,符合高质量文库的要求,文库的重组率为 98.05%和 96.15%,说明文库是完整的、有效的。两个文库的总克隆数分别为  $3.5 \times 10^5$  个和  $2.1 \times 10^5$  个,高于所需标准,因此从所构建文库筛选低丰度的目的基因在理论上是可行的。我们通过下一步的测序分析,获知茶树叶片的基因表达情况后,将以此文库为基础,筛选我们感兴趣的的目的基因的克隆。

### 4.2 构建了 $Sfi$ -pUCM-T 载体

用 BM25.8 获得的 cDNA 质粒文库的拷贝数低。而为了提高测序效率,获得高质量的序列信息,我们改用了本实验室的菌株 TG1,所以对相应的载体也进行了改变。因此我们在原来的文库材料的基础上,构建了一个新的质粒载体  $Sfi$ -pUC。它是利用由原载体  $\lambda$ Triplex2 获得的质粒,在 pUCM-T 载体上引入 2 个  $Sfi$  酶切位点后获得的。并用新的载体与 cDNA 连接构建了一个与噬菌体文库含有相同信息的龙井 43 的质粒文库。

### 4.3 文库测序及 EST 分析

对龙井 43 新梢 cDNA 文库中随机选取的 4320 个克隆的序列进行测定,可以代表文库中基因的分布。根据 EST 标准获得长度大于 200bp 的序列共 2963 个,占总测序克隆数的 68.60%,其平均长度为 478bp。为了获得高质量的序列,剔除了测序误差大于 5%和除去载体序列后长度小于 150bp 的序列,并去除了核糖体 RNA、叶绿体和线粒体以及载体序列的污染。最后共获得了高质量的序列 1684 个。所有高质量的基因与 NCBI 的核酸数据库比对,有 36.2%的 EST 序列显示了与已知序列的很大的同源性,而与其蛋白质数据库进行比对时有 71.3%的序列显示了与已知序列的很大的同源性。

对获得的序列通过 DNATool 软件去除的重复序列有 863 个,占总获得序列的 33.88%。

当去除重复序列后进行 BLAST 分析时,丰度较高的基因(表达频率=5)有约 30 个,约占总数的 10%,中等丰度基因(表达频率在 2-5 之间)约 100 个,占总数的 30%,其余低丰度的基因约占 60%。这与 Bisshop 等的结果基本相符,即大多数基因呈中低丰度表达。

#### 4.4 茶树新梢的功能基因

结合拟南芥基因功能分类标准,把茶树新梢组织中有功能描述的基因分为 12 类。通过对所有已知功能基因的分析发现:与编码能量代谢相关基因(Energy)的 EST 数量最多,共有 266 个,占有功能描述基因的 26.39%、蛋白质合成相关基因(Protein synthesis)的 EST 占有功能描述基因的 17.86%、细胞结构相关基因(Cell structure)占 10.71%、细胞抗性及其防御相关基因(Disease and defence)占 9.23%、代谢相关基因(Metabolism)占 8.33%、信号转导相关基因(Cellular signal transduction)占 5.36%、细胞内运输相关基因(Transporters)占 4.66%、转录相关基因(Transcription)占 3.87%、次生代谢相关基因(Secondary metabolism)占 3.47%、细胞生长发育相关基因(Cell growth and division)占 3.77%、蛋白质降解相关基因(Protein destination and storage)占 1.69%和功能不明确基因(Unclear classification and unclassified)4.66%。

部分功能明确的 EST 序列申请了专利保护,目前正在受理中。部分功能明确的基因序列在 GenBank 上登录,详细信息见附件 2。

#### 4.5 今后研究方向

随着分子生物学研究的深入,不断有新的序列产生,也有新的功能被鉴定出来。因此对 EST 的分析是一个持续的工作过程,需要随时跟踪最新的进展,并且由此做出功能上的预期。在本研究的基础上,今后的研究主要有以下几个方向:

- 1、凭借 EST 分析获得的数据信息,对感兴趣的 EST 进行功能验证。即根据 EST 序列设计引物,从 cDNA 文库或基因组中扩增上述序列,以期获得序列全长。
- 2、根据 EST 分析获得的信息,制作由茶树新梢组织基因构成的基因芯片。可以用于检测不同环境条件、不同品种茶树新梢基因表达谱差异。
- 3、构建茶树根系、幼果等,或不同环境条件如胁迫、病虫害危害前后的 cDNA 文库并进行测序,与此结果比较,研究茶树基因表达的差异。

## 参考文献

1. 丁克越, 沈岩. ESTs 数据分析及 ESTs 数据系统. *国外医学分子生物学分册*, 2002, 24(2): 113-116
2. 冯艳飞, 梁月荣. 茶树 S-腺苷甲硫氨酸合成酶基因的克隆和序列分析. *茶叶科学*, 2001, 21(1): 21-25
3. 高其康, 野蚕黑卵蜂寄主识别利它素的分子生物学研究. *博士学位论文*, 2000
4. 关敏, 孙蕾, 常雅萍, 于永利. 人胎脑未知基因 EST 的发现. *白求恩大学学报*, 2001, 27(6): 589-561
5. 贺林主编. 解码生命—人类基因组计划及后基因组计划, 北京: 科学出版社, 2000: 173
6. 江昌俊, 王朝霞, 李叶云. 茶树中提纯总 RNA 的研究. *茶叶科学*, 2000, 20(1): 27-29
7. 李大力. 一种从富含次生物质的植物中提取 RNA 的方法. *南京理工大学学报*, 2001, 25(5): 547-549
8. 李宏, 王新力. 植物组织 RNA 提取的难点及对策. *生物技术通报*, 1999, 1: 36-39
9. 陆佳韵, 王秀琴, 吴昊, EST 及其应用. *生命科学*, 1999, 11(4): 186-18
10. 罗静初, 江涛, 李兵. 分子生物信息镜像系统和数据库. *高技术通讯*, 1998, 10(10): 61-63, 53
11. 骆蒙, 孔秀英, 霍纳新, 周荣华, 贾继增, 小麦抗白粉病侵染初期的表达序列标签分析. *遗传学报*, 2002, 29(6): 525-530
12. 骆蒙, 贾继增. 植物基因组表达序列标签(EST)计划研究进展. *生物化学与生物物理进展*, 2001; 28(4): 494-497
13. 沈倍奋主编, 《分子文库》, 北京: 科学出版社, 2001
14. 魏英杰, 丁金凤, 许有元, 钱杰, 高润霖. 血管成形术后血管平滑肌细胞增殖相关 ESTs 的筛选和克隆. *高血压杂志*, 1999, 7(4): 352-356
15. 吴乃虎编著, 《基因工程原理》, 北京: 科学出版社, 2000
16. 杨晓明, 贺福初, 邢桂春. 人胎肝 cDNA 文库的免疫学筛选. *免疫学杂志*, 1996, 12(2): 78-81
17. 赵东, 刘祖生, 奚彪. 茶树多酚氧化酶基因的克隆及其序列比较. *茶叶科学*, 2001, 21(2): 94-98
18. 赵丽萍, 陈亮, 高其康. 茶树基因克隆、遗传转化研究进展及 EST 在功能基因组研究上得应用. 第三届海峡两岸茶业学术研讨会论文集, 2003, pp: 292-300
19. 周国岭, 杨光圣, 傅延栋. 基因克隆技术. *华中农业大学学报*, 2001, 20(6): 584-592
20. Ablett E, Seaton G, Scott K, Shelton D, Graham MW, Baverstock P, Lee LS and Henry R. Analysis of grape ESTs: global gene expression patterns in leaf and berry, *Plant Science* 2000, 159: 87-95

21. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR and Venter JC. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 1991,252(5013): 1651-1656
22. Ainsworth C. Isolation of RNA from floral tissue of *Rumex acetosa* (sorrel). *Plant Mol Biol Reprtr*. 1994, 12:198-203
23. Audic S and Claveris JM. The significance of digital gene expression profile. *Genome Res*. 1997,7:986-995
24. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J and Wheeler DL. GenBank, *Nucl. Acids Res*. 2003, 31(1): 23-27
25. Bisshop JO, Morton JG, Rosbash M. Three abundance classes in Hela Cell messenger RNA. *Nature*, 1974, 250:199-204
26. Boguski MS, Lowe TM, Tolstoshev CM. dbEST-database for expressed sequence tags. *Nat. Genet*. 1993, 4:332-333
27. Borges MF, Spohn PK, Coulson AS. Using 3' untranslated sequences to identify differentially expressed genes in Leishmania. *Gene*, 1997, 196:159-164
28. Carson DL, Botha EC, Genes expressed in sugarcane maturing internodal tissue. *Plant Cell Rep*. 2002, 20:1075-1081
29. Carulli JP, Artinger M, Swain PM. High throughput analysis of differential gene expression. *Cell Biochem Suppl*. 1998, 30-31:286-296
30. Chomczynski P, Sacchi N. Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem*. 1987,162:156-159
31. Cleveland DW and Sullivan KF. Molecular biology and genetics of tubulin. *Annu. Rev. Biochem*. 1985,54:331-365
32. Cooke R, Raynal M, Laudie M, Grellet F, Delseny M, Morris PC, Guerrier D, Giraudat J, Quigley F, Clabault G, Li YF, Mache R, Krivitzky M, Gy IJ, Kreis M, Lecharny A, Parmentier Y, Marbach J, Fleck J, Clement B, Philipps G, Herve C, Bardet C, Tremousaygue D, Hofte H. Further progress towards a catalogue of all *Arabidopsis* genes: analysis of a set of 5000 non-redundant ESTs. *Plant*, 1996,9:101-124
33. Costanzo F, Castagnoli L, Dente L, Arcari P, Smith M, Costanzo P, Raugei G, Izzo P, Pietropaolo TC, Bougueleret L, Cimino F, Salvatore F, Cortese R. Cloning of several cDNA segments coding rot human liver proteins. *EMBO* 1983, 2:57-60
34. Csaikl UM, Bastian H, Brettschneider R, Gaush S, Meir A, Schauerte M, Scholz F, Sperisen C, Vornam B, Ziegenhagen B. Phenylalanine ammonia-lyase gene organization and structure. *Plant Mol. Biol*. 1989,12:367-383
35. Delseny M, Coole K, Raynaol M, Gellet F. The *Arabidopsis thaliana* cDNA sequencing project. *FEBS letter*, 1997, 403(3):221-224
36. Després C, Delong C, Glaze S, et al. The Arabidopsis NPR1/NIM1 protein enhances the

- DNA binding activity of a subgroup of the TGA family of bZIP transcription factors. *Plant Cell* 2000,12:279-290
37. Ewing RM, Kahla AB, Poirot O, Lopez F, Audic S and Claverie JM. Large scale statistical analysis of rice ESTs reveals correlated patterns of gene expression. *Genome Res.* 1999,9(10): 950
  38. Frohman MA, Dush MK, Martin GR. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci., USA* 1988, 85(23): 8998-9002
  39. Gao QK, Hu C. Construction of a cDNA library of host recognition kairomone for *telenomus theophila*. *Entomologia Sinica* 2001,9(1): 35–39
  40. Gill RW, Sanseau P. Rapid in silico cloning of genes using expressed sequence tags (ESTs). *Biotechnol. Annu. Rev.* 2000, 5:25-44
  41. Graham G C. A method for extraction of total RNA from *Pinus radiata* and other conifers. *Plant Mol. Biol. Repr.* 1993,11:32-37
  42. Guillemaut P, Marchal-Drouard L. Isolation of plant DNA: a fast inexpensive and reliable method. *Plant Mol. Biol. Repr.* 1992,10:69-74
  43. Guo W, Yamaguchi K, Watanabe N, Usui T, Luo SJ, Sakata K.. A primeverosidase as a main glycosidase concerned with the alcoholic aroma formation in tea leaves. *Biosci. Biotech. Biochem.* 1995,59(5): 962-964
  44. Gutierrez RA, Ewing RM, Cherry J M and Green PJ. Identification of unstable transcripts in *Arabidopsis* by cDNA microarray analysis: Rapid decay is associated with a group of touch- and specific clock-controlled genes. *Proc. Natl. Acad. Sci. USA* 2002, 99(17): 11513-11518
  45. Hackett CA, Wachira FN, Paul S, Powell W, Waugh R. Construction of a genetic linkage map for *Camellia sinensis* (tea). *Heredity* 2000, 85:346-55
  46. Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T. A high-density rice genetic linkage map with 2275 markers using a single F<sub>2</sub> population. *Genetic* 1998, 148(1): 1-16
  47. Hately F, Tosser-Klopp G, Cloucard-martinato C. Expressed sequenced tags for gene: a review. *Genet. Sel. Evol.* 1998,30(5): 521-541
  48. Hertog MGL, Hollman PCH, Katan MB, Kromhout D, Intake of potentially anticarcinogenic flavonoids and their determinations in adults in the Netherlands, *Nutr. Cancer*, 1993, 20:21-29
  49. [http://www.ncbi.nlm.nih.gov/dbEST/dbEST\\_summary.html](http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html)
  50. Iwasa K, Biosynthesis of catechins in tea plant. *Bull Natl Res Inst Tea* 1977,13:101-126
  51. Jan, VD. cDNA Encoding S-adenosyl-L-methionine synthetase from poplar. *Plant physiol.* 1993,102:1365-1366
  52. Jefferson RA, Kavanagh TA, Bevan MW. GUS fusions;  $\beta$ -glucuronidase as a sensitive and

- versatile gene. *EMBO* 1987, 6, 3901-3907
53. John ME. An efficient method for isolation of RNA and DNA from plants containing polyphenolics. *Nucl. Acids Res.* 1992,20:2381
  54. Kato M. Biochemistry and molecular biology in caffeine biosynthesis--molecular cloning and gene expression of caffeine synthase. In: *Proceedings of 2001 International Conference on O-Cha (Tea) Science* (Session II), 21-24, Shizuoka, Japan
  55. Lange BM, Wildung MR, Stauber EJ, Sanchez C, Pouchnik D and Croteau R. Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes. *Proc. Natl. Acad. Sci., USA* 2000, 97(6): 2934-2939
  56. Lee CM, Lee YJ, Lee MH, Nam HG, Cho TJ, Hahn TR, Cho MJ, Sohn U, Large-scale analysis of expressed genes from the leaf of oilseed rape (*Brassica napus* L.). *Plant Cell Rep.* 1998,17:930-936
  57. Li TW, Xiang JM and Liu RY. Construction of cDNA library of shrimp *Penaeus chinensis* (Crustacea, Decapoda). *Acta Zoological Sinica*, 1998, 44(2): 237-238
  58. Liang YR, Tanaka J, Takeda Y. Molecular cloning and sequencing of a *Camellia sinensis* cDNA encoding S-adenosylmethionine synthetase. In: *Proceedings of 2001 International Conference on O-Cha (Tea) Science* (Session II), 58-61, Shizuoka, Japan
  59. Luo M, Kong XY Jiang T. Analysis of resistance to Powdery-Mildew in wheat based on expressed sequence tags technique. *Acta Botanica Sinica*, 2002, 44 (5): 567-572
  60. Ma SJ, Mizutani M, Hiratake J. Substrate specificity of  $\beta$ -primeverosidase, a key enzyme in aroma formation during oolong tea and black tea manufacturing. In: *Proceedings of 2001 International Conference on O-Cha (Tea) Science* (Session II), 15-21, Shizuoka, Japan
  61. Matsumoto S, Takeuchi A, Hayatsu M. Molecular cloning of phenylalanine ammonia-lyase cDNA and classification of varieties and cultivars of tea plants (*Camellia sinensis*) using the tea PAL cDNA probe. *Theor. Appl. Genet.* 1994,89:671-675
  62. Mayer AM, Harel E. Polyphenol oxidases in plants. *Phytochemistry* 1979,18:193-215
  63. Mekhedov S, Martinez O and Ohlrogge J. Towards a functional catalog of the plant genome: a survey of genes for lipid biosynthesis, *Plant Physiol.* 2000, 122:389-401
  64. Michael B, Robert S, Jose C, Phat D, Wayne H, Randall N. An expressed sequence tag(EST) set from *Citrus sinensis* L. Osbeck whole seedlings and the implications of further perennial source investigations. *Plant Science*, 2003, 165:415-422
  65. Nelson PS, Ng WL, Schummer M, True, L D, Liu, A Y, Bumgarner, R E, Ferguson C, Dimak A, Hood L. An expressed sequence tag database of the human prostate: sequence analysis of 1168 cDNA clones. *Genomics*, 1998, 47(1): 12-25
  66. Newbury HJ, Possingham JV. Factors affecting the extraction of intact ribonucleic acid from plant tissues containing interfering phenolic compounds. *Plant Physiol.* 1997, 60,543-547
  67. Newman,T., deBruijn,FJ, Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N.,Somerville,S., Thomashow,M., Retzel,E. and Somerville,C Genes galore: a



- summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones. *Plant Physiol.* 1994, 106:1241-1255
68. Ogawa K, Moon JH, Guo W, Yagi A, Watanabe N, Sakata K, Z. Naturforsch., A study on tea aroma formation mechanism: alcoholic aroma precursor amounts and glycosidase activity in parts of the tea plant. *Zeitschrift für Naturforschung. Section C. Biosciences* 1995,50(7/8): 493-498
  69. Okubo K, Hori N, Matoba R, Niiyama T, Fukushima A, Kojima Y, Matsubara K. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nat. Genet.* 1992, 2:173-179
  70. Park YS, Kwak JM, Kwon OY, Kim YS, Lee DS, Cho MJ, Lee HH, Wam HG. Generation of expressed sequence tags of random root cDNA clones of *Brassica rapus* by single-run partial sequencing. *Plant Physiol* 1993,103:359-370
  71. Pearson WR, Effective protein sequence comparison. *Methods Enzymol.* 1996,266:227-258
  72. Rounsley S, Linx K. Large scale sequencing of plant genome. *Curr. Opin. Plant Biol.* 1998,1(2): 136-141
  73. Saijo R, Effect on shade treatment on biosynthesis of catechins in tea plants. *Plant Cell Physiol* 1980,21:989-998
  74. Sakata K, Watanabe N, Usui T. Molecular basis of alcoholic aroma formation during tea processing. In “*Food for Health in the Pacific Rim*” *Proceedings of the 3<sup>rd</sup> International Conference of Food Science and Technology*, ed. By JR Whitaker, et al., Davis, California, Foods Nutrition Press, Inc., Trumbull, Connecticut, 1999,p. 93-105
  75. Sambrook J, Fritsch EF and Maniatis T. *Molecular Cloning: A Laboratory Manual* (2<sup>nd</sup> ed), Cold Spring Harbor Laboratory Press, 1998
  76. Sasaki T, Benjamin Burr. International Rice Genome Sequencing Project: the effect to completely sequence the rice genome. *Curr. Opin. Plant Biol.* 2000,3:138-141
  77. Sasaki T. The rice genome project in Japan. *Proc. Natl. Acad. Sci., USA* 1998,95(5): 2027-2028
  78. Schneiderbauer A, Sandermann HJ, and Ernst D. Isolation of functional RNA from plant tissues rich in phenolic compounds. *Anal. Biochem.* 1991, 197,91-95
  79. Schuler GD. Pieces of the puzzle: expressed sequence tags and the catalog of human genes. *Journal of Molecular Medicine*, 1997, 75:694-698
  80. Sterky F, Regan S, Karlsson J, Hertzberg M, Rohde A, Holmberg A, Amini B, Bhalerao R, Larsson M, Villarroel R, Montagu MV, Sandberg G, Olsson O, Teeri TT, Boerjan W, Gustafsson P, Uhlén M, Sundberg B and Lundeberg J. Gene discovery in the wood-forming tissues of poplar: analysis of 5,692 expressed sequence tags. *Proc. Natl. Acad. Sci., USA* 1998, 95(22): 13330-13335
  81. Sugui JA, Deising HB. Isolation of infection-specific sequence tags expressed during early stages of maize anthracnose disease development. *Molecular Plant Pathology*,

2002, 3(4) , 197–203

82. Suzuki T, Takahashi E. Biosynthesis of caffeine by tea-leaf extracts—Enzymic formation of theobromine from 7-methylxanthine and of caffeine from theobromine. *Biochem.* 1975,146:87-96
83. Takeuchi A, Matsumoto A, Hayatsu M. Chalcone synthase from *Camellia sinensis* : Isolation of the cDNAs and the Organ-specific and Sugar-responsive expression of the Genes. *Plant Cell Physiol.* 1994,35(7): 1011-1018
84. Takeuchi A, Matsumoto S, Hayatsu M. Effects of shading treatment on the expression of the genes for chalcone synthase and phenylalanine ammonia-lyase in tea plant (*Camellia sinensis*). *Bull. Natl. Res. Inst. Veg.Ornam. Plants & Tea, Japan (B)* 1995, 8: 1 -9
85. The EU Arabidopsis Genome Project. Analysis of 1.9Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature* 1998, 391(29):485-492
86. Umeda M, Hara C, Matsubayashi Y, Li HH, Liu Q, Tadokoro F, Aotsuka S, Uchimiya H. Expressed sequence tags from cultured cells of rice (*Oryza sativa* L.) under stressed conditions: analysis of transcripts of genes engaged in ATP-generating pathways. *Plant Mol. Biol.* 1994,25(3): 469-478
87. van de Loo FJ, Broun P, Turner S, Somerville C. An oleate 12-hydroxylase from *Ricinus communis* L. is a fatty acyl desaturase homolog. *Proc. Natl. Acad. Sci., USA* 1995, 92:6743-6747
88. Vaughn KC, Duke SO. Function of polyphenol oxidase in higher plants. *Physiol. Plant,* 1984,60:106-112
89. Venter JC, Adams MD, Myers EW, *et al.* The sequence of the human genome. *Science* 2001, 291(5507): 1304-51
90. Wang CS, Vodkin LO. Extraction of RNA from tissues containing high levels of procyanidins that bind RNA. *Plant Mol. Biol. Reprtr.* 1994, 12:132-145
91. Woltering, EJ, Ten Have, A., Larsen, PB, Woodson, WR. Expression of ethylene biosynthetic pathway transcripts in senescing carnation flowers. *Plant Physiol.* 1992, 99:526-532.
92. Wu J, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yazaki J, Fujii F, Shomura A, Ando T, Kono I, Waki K, Yamamoto K, Yano M, Matsumoto T and Sasaki T. A comprehensive rice transcript map containing 6591 Expressed Sequence Tag sites. *The Plant Cell*, 2002, 14: 525-535
93. Yamamoto K and Sasaki T. Large-scale EST sequencing in rice. *Plant Mol. Biol.* 1997, 35(1): 135-144
94. Yamamoto T, Juneja LJ, Chu DC, Kim M .Chemistry and application of green tea. *New York: CRC Press*, 1998, pp1-158
95. Yang, S.F., Hoffman. Ethylene biosynthesis and its regulation in higher plant. *Annu Rev Plant Physiol.* 1984, 35:155-189.
96. Young R.A., Davis, Efficient isolation of gene by using antibody probes. *Proc. Natl. Acad.*

*Sci., USA* 1993, 80:1194

97. Zhang Y, Fan W, Kinkema M, Li X, Dong X. Interaction of NPR1 with basic leucine zipper protein transcription factors that bind sequences required for salicylic acid induction of the PR-1 gene. *Proc. Natl. Acad. Sci., USA* 1999,96:6523-6528
98. 竹内敦子・松元、哲・早津雅仁. Amplification of  $\beta$ -Tubulin cDNA from *Cammellia sinensis* by PCR. *Bull. Natl. Res. Inst. Veg. Ornam. Plants & Tea, Japan (B)* 1994,7 : 13-20

# 致 谢

本论文是在中国农业科学院茶叶研究所陈亮副研究员和浙江大学高其康研究员两位导师共同悉心指导下完成的。三年来，每一滴的长进和提高无不浸透着两位导师的心血。两位导师学识之渊博、思维之敏锐以及治学之严谨都给我留下了深刻的印象并且使我终身受益。导师们求实的精神、高尚的人格和呕心沥血的敬业精神为我树立了楷模。在此论文完成之际，谨向两位导师表示衷心的感谢并致以崇高的敬意。

在论文完成的过程中，我非常荣幸地得到了中国农业科学院茶叶研究所领导、老师们的热心帮助，是他们抽出宝贵的时间给我补充茶学方面的知识，在此向茶叶所领导和老师们表示衷心的感谢，特别是课题组的虞富莲研究员以及姚明哲和王新超两位师兄，在生活上和实验上都给予我无微不至的关心和帮助，在此一并表示衷心的感谢。

论文的大部分内容是在浙江大学华家池校区分析测试中心大分子实验室完成的。实验室的毛伟华老师和徐幼萍老师在实验上给了我非常大的关心和帮助，还有孙守慧博士、刘贤德、丁矛、韦国栋、李红亮、赵敏、牛黎明和聂文敏硕士以及李柱同学等都给了非常大的帮助。我的学友王颖在三年的学习生活中给了我许多的关心和帮助。

本实验的顺利完成也离不开中国农业科学院研究生院的领导和各位老师的关心和支持，在此一并表示感谢。

值此论文完成之际，特别感谢我的父母和家人在我求学路上给予我精神和物质上的支持。

最后，谨向所有关心、支持和帮助我的亲友、老师与同学表示诚挚的谢意！

赵丽萍

2004年5月于杭州

# 附件 1 EST 序列提交格式

## PUBLICATION FILES

TYPE: Pub

TITLE:

Generation and large-scale analysis of expressed sequence tags (EST) from young leaf cDNA library of tea plant [Camellia sinensis (L.) O. Kuntze]

AUTHORS: Chen, L.; Zhao, L.P.; Gao, Q.K.

YEAR: 2004

STATUS: 1

||

## LIBRARY FILES

TYPE: Lib

NAME: Young leaf cDNA library of tea plant (Camellia sinensis)

ORGANISM: Camellia sinensis

CULTIVAR: Longjing 43

TISSUE: Young leaf

STAGE: Growing stage

HOST: E. coli TG1

VECTOR: lambda TriplEx2

V\_TYPE: Phage

RE\_1: SfiIA and B

||

## CONTACT FILES

TYPE: Cont

NAME: Chen, Liang

TEL: 86 571 8665 2835

EMAIL: tbtri@mail.hz.zj.cn

LAB: Laboratory for Germplasm, Breeding and Molecular Biology

INST: Tea Research Institute, Chinese Academy of Agricultural Sciences

ADDR: 1 Yunqi Road, Hangzhou, Zhejiang 310008, P.R. China

||

## EST FILES

TYPE: EST

STATUS: New

CONT\_NAME: Chen, Liang

Generation and large-scale analysis of expressed sequence tags (EST) from young leaf cDNA library of tea plant [*Camellia sinensis* (L.) O. Kuntze]

CLONE: TL001A01

P\_END: 5'

PUBLIC:

POLYA:unknown

ACGGCGTGGACCATCCAGTATTCTCTCTCTCTCTCTCTCTCTCTCAAGGCCA  
GTTTTTTTTTCTCTCTCTAAGAACCTAGTCTCAAATGGCTCGTACAAAGCAGACAGCTC  
GGAAGTCCACCGGAGGCAAGGCACCAAGGAAGCAACTCGCCACCAAGGCCGTCGCA  
AGTCCGTCCAGCCACCGGTGGTGTGAAGAAGCCCCACCGTTCGCGCCCGGCACGGTGG  
CGCTCCGGGAGATCCGAAAGTACCAGAAATCCACAGAGCTTCTGATCCGAAAGCTTCC  
ATTCCAGAGACTTGCTCGTGAGATCGCTCAGGACTTCAAGACCGATCTCCGATTCCAG  
AGCTCCGCTGTTGCGGCCCTCCAGGAGGCTGCTGAGTCTTACCTTGTTGGACTGTATG  
AAGACACCAATCTTTGTGCTATTACGCTAAGCGTGTACAGATTATGCCTAAGGATATG  
CAGCTTGCTCGCCGCATCAGAGGCGAACGTGCTTAGGGTTTGAGATTGTTGCCTAATGT  
GATTGTGAATCAATGGTTAGTTGTTTGTATAGGTGGTTAGGGTTAGTGTGTCTGGTTTCT  
AGTGCTAGTCGGTTCTGTGTTTTAATTGGTGTCTATCTACTGTTGAAGCTTTATAAAGTT  
ATTTCAATGAAATTATGCTATTATCAAAAAAAAAAAAAAAAAAAAAAAAAA

11

## 附件 2 GenBank 登录的基因

1: [AY570925](#). *Camellia sinensis*...[gi:46242818] [Links](#)

LOCUS AY570925 658 bp mRNA linear PLN 11-APR-2004

DEFINITION *Camellia sinensis* var. *sinensis* S-adenosylmethionine synthase 2 mRNA, partial cds.

ACCESSION AY570925

VERSION AY570925.1 GI:46242818

KEYWORDS .

SOURCE *Camellia sinensis* var. *sinensis* (green tea)

ORGANISM [Camellia sinensis var. sinensis](#)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta;  
Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
Ericales; Theaceae; *Camellia*.

REFERENCE 1 (bases 1 to 658)  
AUTHORS Chen, L., Zhao, L.P. and Gao, Q.K.  
TITLE S-adenosylmethionine synthetase 2 gene isolated from tea plant (*Camellia sinensis* var. *sinensis*)  
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 658)  
AUTHORS Chen, L., Zhao, L.P. and Gao, Q.K.  
TITLE Direct Submission  
JOURNAL Submitted (01-MAR-2004) Key Laboratory of Tea Chemical Engineering, Ministry of Agriculture, Tea Research Institute, Chinese Academy of Agricultural Sciences, 1 Yunqi Road, Hangzhou, Zhejiang 310008, China

FEATURES Location/Qualifiers  
source 1..658  
/organism="Camellia sinensis var. sinensis"  
/mol\_type="mRNA"  
/variety="sinensis"  
/cultivar="Longjing 43"  
/db\_xref="taxon:238974"  
/note="common: green tea"  
[CDS](#) 124..>658  
/codon\_start=1

/product="S-adenosylmethionine synthase 2"  
 /protein\_id="[AAS83521.1](#)"  
 /db\_xref="GI:46242819"  
 /translation="METFLFTSESVNEGHPDKLCDQISDAVLDACLEQDQDSKVACET  
 CTKTNMVMVFGEITTKAAVDYEKIVRDTCTMGLVSDDVGLDADNCKVLVNI EQQSPD  
 IAQGVHGHLTRPEEMGAGDQGHMLGYATDETSELMPLSHVLATKLGARLTEVRKNGTCPWLRPDGKTQVTVEYYH"  
 ORIGIN

```

      1 ggggcaaaag ccctaggaag cagctaaggg gtctctcttc tctctctctg tctctcttcg
     61 catccttttt ttgggggttg atatctcttc gcatccttgt tttatcatct tcggtttgag
    121 agaatgaaa ctttcctatt cacttccgaa tctgtgaatg agggtcaccc agacaagctc
    181 tgtgatcaga tatctgatgc agttctcgat gcctgccttg aacaagatca agacagcaaa
    241 gttgcctgtg agacttgac caagactaac atggatcatg tcttcggtga aatcacaacc
    301 aaagccgagg tagactatga gaagattgtg cgtgacacat gccgcacat gggattgggt
    361 tccgatgatg ttggtcttga tgctgataat tgcaagggtc tagtcaacat cgagcaacag
    421 agccctgata ttgcccaggg tgtccacggt catctcacta agcgccccga ggagatgggt
    481 gctggtgacc agggtcacat gttgggctat gccacagatg aaacttccga gctgatgccc
    541 ctgagccatg tccttgcaac caagcttggg gccgcctca ctgaagtgcg aaagaatgga
    601 acttgcccct ggttgagacc tgatggcaaa acccaagtca ctgttgagta ttaccatg
  
```

//

---

1: [AY570926](#). *Camellia sinensis*...[gi:46242820] [Links](#)

LOCUS AY570926 600 bp mRNA linear PLN 11-APR-2004

DEFINITION *Camellia sinensis* var. *sinensis* chalcone synthase 1 mRNA, partial cds.

ACCESSION AY570926

VERSION AY570926.1 GI:46242820

KEYWORDS .

SOURCE *Camellia sinensis* var. *sinensis* (green tea)

ORGANISM [Camellia sinensis var. sinensis](#)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta;  
 Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 Ericales; Theaceae; *Camellia*.

REFERENCE 1 (bases 1 to 600)  
 AUTHORS Chen, L., Zhao, L.P. and Gao, Q.K.  
 TITLE Chalcone synthase 1 gene isolated from tea plant (*Camellia sinensis* var. *sinensis*)  
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 600)



AUTHORS Chen,L., Zhao,L.P. and Gao,Q.K.  
TITLE Direct Submission  
JOURNAL Submitted (01-MAR-2004) Key Laboratory of Tea Chemical  
Engineering,Ministry of Agriculture, Tea Research Institute, Chinese Academy of  
Agricultural Sciences, 1 Yunqi Road, Hangzhou, Zhejiang 310008,China

FEATURES Location/Qualifiers

source 1..600  
/organism="Camellia sinensis var. sinensis"  
/mol\_type="mRNA"  
/variety="sinensis"  
/cultivar="Longjing 43"  
/db\_xref="taxon:238974"  
/note="common: green tea"  
[CDS](#) <1..429  
/note="CHS1"  
/codon\_start=1  
/product="chalcone synthase 1"  
/protein\_id="[AAS83522.1](#)"  
/db\_xref="GI:46242821"

/translation="PDSGATDGHLEVALTFHLLKDVPLISKNIKSLAEAFQPLGISDWNSLFWIARPGGPAILDQ  
VELKSGLKEEKLRAHRHVLSEYGNMSSACVLFILDEMRRKSAADGLKTTGEGLEWGVLFGGPGLTVETVVLHSLST"

ORIGIN

```
1 cccgatagt acggtgccac tgacggacac ctccgtgaag tggcccttac atttcacctc
61 ctcaaggatg ttcttgggct catctccaag aacattgaaa agagtttggc ggaagcattc
121 caacccttgg gcatctctga ttggaactct ttgttttggga ttgcacgtcc tgggtgggcct
181 gctattttgg atcaggtgga attaaagtcg ggccttaagg aagagaagct acggggccacg
241 aggcacgtgc taagttagta cggtaatatg tcaagtgcgt gcgtgttgtt tatattggat
301 gagatgagga agaagtccgc ggcggatggg ctcaagacca ctggtgaagg gctcgagtgg
361 ggagtgctat tcgggttcgg accagggtc actgttgaga ctgtggtgct ccatagcttg
421 tctacataaa atggctgcta gcctattgag tattgtttcg gggttttctt caacttgtga
481 ttgtgttatg ggttggcttt gtttgggtct actttatttc ctgcatttaa agtaagaatg
541 aaatttatgt aatgagatta taatggaaaa taacatgttc ttgcaatat gtatgcaaaa
```

//

1: [AY570927](#). Camellia sinensis...[gi:46242822]

[Links](#)

LOCUS AY570927 677 bp mRNA linear PLN 11-APR-2004

DEFINITION *Camellia sinensis* var. *sinensis* chalcone synthase 3 mRNA, partial cds.  
 ACCESSION AY570927  
 VERSION AY570927.1 GI:46242822  
 KEYWORDS .  
 SOURCE *Camellia sinensis* var. *sinensis* (green tea)  
 ORGANISM [Camellia sinensis var. sinensis](#)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta;  
 Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 Ericales; Theaceae; *Camellia*.  
 REFERENCE 1 (bases 1 to 677)  
 AUTHORS Chen, L., Zhao, L.P. and Gao, Q.K.  
 TITLE Chalcone synthase 3 gene isolated from tea plant (*Camellia sinensis* var. *sinensis*)  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 677)  
 AUTHORS Chen, L., Zhao, L.P. and Gao, Q.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-MAR-2004) Key Laboratory of Tea Chemical Engineering, Ministry of Agriculture, Tea Research Institute, Chinese Academy of Agricultural Sciences, 1 Yunqi Road, Hangzhou, Zhejiang 310008, China  
 FEATURES Location/Qualifiers  
 source 1..677  
 /organism="Camellia sinensis var. sinensis"  
 /mol\_type="mRNA"  
 /variety="sinensis"  
 /cultivar="Longjing 43"  
 /db\_xref="taxon:238974"  
 /note="common: green tea"  
[CDS](#) <1..502  
 /note="CHS3"  
 /codon\_start=2  
 /product="chalcone synthase 3"  
 /protein\_id="[AAS83523.1](#)"  
 /db\_xref="GI:46242823"  
 /translation="RPFQEEVEEPTSVGSLVSAAPTILPSSDGAIDGHLREVACTVHLLKDVPRILSMNVEKSLVEAFQ  
 PLGISDWNLSLFWIAHPGGPAILDQVELKLGLKEEKLRAHRLSEYGNMSSACVLFILDEMRKKSAAEGLKTTGEGLEW  
 GVLFGGPGLTVETVVLHSLCT"

## ORIGIN

```
1  cagacccaaa ttccaagagg ttgaggaacc cacctctgtt ggaagcttgg tttcagcggc
61  cccaaccatt cttcctagta gtgacggtgc catcgacgga cacctccgtg aagtggcctg
121 tacagttcat ctctcaagg atgttccaag gctcatctct atgaacgttg aaaagagcct
181 ggtggaagca ttccagcctt tgggcatttc tgattggaat tccctttttt ggattgcaca
241 ccctgggtggg cctgctattt tggaccaggt agagttaaaa ttgggcctta aggaagagaa
301 gctacgggcc acaaggcacg tgctaagcga gtatggtaac atgtcaagcg cgtgcgtgct
361 attcatacta gatgagatga ggaagaaatc agctgaagaa gggctcaaga ccactggcga
421 agggcttgag tggggtgtgc tatttgggtt tggacctggg ctaccggtg agactgtggt
481 gctccacagt ttgtgcactt aaaacttaaa agggctactt aaagtataag cattgtttct
541 tgtttttctt ttaatttgtg gttgtgttgt tgaatggctt tgtttggtca tactctgtgt
601 gcattttaaa gtaagaaatg cgaatgaaag ttatgtagtg acttttaaaa gaaataaaat
661 tttctttatc gaaaaaa
```

//

---

1: [AY570928](#). *Camellia sinensis*...[gi:46242824]

[Links](#)

LOCUS AY570928 147 bp DNA linear PLN 11-APR-2004

DEFINITION *Camellia sinensis* var. *sinensis* tRNA-Leu (trnL) gene, partial sequence; chloroplast.

ACCESSION AY570928

VERSION AY570928.1 GI:46242824

KEYWORDS .

SOURCE chloroplast *Camellia sinensis* var. *sinensis* (green tea)

ORGANISM [Camellia sinensis var. sinensis](#)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; Ericales; Theaceae; *Camellia*.

REFERENCE 1 (bases 1 to 147)

AUTHORS Chen,L., Zhao,L.P. and Gao,Q.K.

TITLE Chloroplast transfer RNA-Leu gene isolated from tea plant (*Camellia sinensis* var. *sinensis*)

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 147)

AUTHORS Chen,L., Zhao,L.P. and Gao,Q.K.

TITLE Direct Submission

JOURNAL Submitted (01-MAR-2004) Key Laboratory of Tea Chemical Engineering, Ministry of Agriculture, Tea Research Institute, Chinese Academy of Agricultural Sciences, 1 Yunqi Road, Hangzhou, Zhejiang 310008, China

FEATURES	Location/Qualifiers
source	1..147 /organism="Camellia sinensis var. sinensis" /organelle="plastid:chloroplast" /mol_type="genomic DNA" /variety="sinensis" /cultivar="Longjing 43" /db_xref="taxon:238974" /note="common: green tea"
<a href="#">gene</a>	<1..>147 /gene="trnL"
<a href="#">tRNA</a>	<1..>35 /gene="trnL" /product="tRNA-Leu"

ORIGIN

```

1 ggggggatat ggcgaaattg gtagacgcta cggacttaat ggattgagcc tggatggaa
61 acctactaag tgataacttt caaattcaga gaaaccctgg aattaataaa aatgggcaat
121 cctgagccaa atcctgtttt tctaaag
//

```

---

1: [AY570929](#). Camellia sinensis...[gi:46242825] [Links](#)

LOCUS AY570929 716 bp mRNA linear PLN 11-APR-2004

DEFINITION Camellia sinensis var. sinensis caffeine synthase 1 mRNA, partial cds.

ACCESSION AY570929

VERSION AY570929.1 GI:46242825

KEYWORDS .

SOURCE Camellia sinensis var. sinensis (green tea)

ORGANISM [Camellia sinensis var. sinensis](#)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta;  
Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
Ericales; Theaceae; Camellia.

REFERENCE 1 (bases 1 to 716)  
AUTHORS Chen,L., Zhao,L.P. and Gao,Q.K.  
TITLE Caffeine synthase gene isolated from tea plant (Camellia sinensis var. sinensis)  
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 716)  
AUTHORS Chen,L., Zhao,L.P. and Gao,Q.K.

TITLE Direct Submission  
JOURNAL Submitted (01-MAR-2004) Key Laboratory of Tea Chemical  
Engineering, Ministry of Agriculture, Tea Research Institute, Chinese Academy of  
Agricultural Sciences, 1 Yunqi Road, Hangzhou, Zhejiang 310008, China

FEATURES Location/Qualifiers  
source 1..716  
/organism="Camellia sinensis var. sinensis"  
/mol\_type="mRNA"  
/variety="sinensis"  
/cultivar="Longjing 43"  
/db\_xref="taxon:238974"  
/note="common: green tea"  
[CDS](#) 66..>716  
/codon\_start=1  
/product="caffeine synthase 1"  
/protein\_id="[AAS83524.1](#)"  
/db\_xref="GI:46242826"  
/translation="MELATAGKVNVEVLFMNRGEGESSYAQNSSFTQQVASMAQPALEN  
AVETLFSRDFHLQALNAADLGCAAGPNTFAVISTIKRMMEKKCRELNCQTLELQVYLN  
DLFGNDFNTLFGKLSSEVIGNKCEEVSCYVMGVPGSFHGRFLFPRNSLHLVHSSYSVHW  
LTQAPKGLTSREGLALNKGKIYISKTSPPVVRESYLSQFHEDSQCFSM LDPGRGGSQW"

#### ORIGIN

```
1 ggggattcat atcactgctg tggcagctgg cctctttgct ataaaaatta cttttccgac
61 gaggcatgga gctagctact gcggggaagg tgaacgaagt gttgttcattg aacagaggag
121 aaggagaaag tagttatgca caaaactctt ctttcacgca acaagtggcc tcaatggcac
181 agccagcgct agaaaatgca gttgaaactc tcttctccag agatttcac cttcaagctc
241 ttaacgcagc ggacttgggt tgtgcagcgg gtccaaacac attcgagtg atttctacga
301 tcaagagaat gatggaaaag aaatgcaggg aattgaattg ccaaactg gaacttcagg
361 ttacttgaa tgatctttt ggaaatgatt tcaataccct cttcaaaggc ctgtcgcttg
421 aggttattgg taacaaatgt gaggaagttt cgtgttatgt gatgggagta ccgggggtctt
481 tccatggtcg gctttttcct cgtaacagct tacatttagt tcattcctct tacagtgttc
541 attggcttac tcaggcacca aaaggactca caagcagaga aggcttggca ttaaacaagg
601 ggaagattta catatcaaag acaagccctc ctgttgtaag agagagctac ttatctcaat
661 ttcatagaaga ttcacaatgt ttctcaatgc tagatccaag aggtgggtcc caatgg
```

//

1: [AY570930](#). Camellia sinensis...[gi:46242831]

[Links](#)

LOCUS AY570930 506 bp mRNA linear PLN 11-APR-2004

DEFINITION *Camellia sinensis* var. *sinensis* chalcone synthase 2 mRNA, partial cds.  
 ACCESSION AY570930  
 VERSION AY570930.1 GI:46242831  
 KEYWORDS .  
 SOURCE *Camellia sinensis* var. *sinensis* (green tea)  
 ORGANISM [Camellia sinensis var. sinensis](#)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta;  
 Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 Ericales; Theaceae; *Camellia*.  
 REFERENCE 1 (bases 1 to 506)  
 AUTHORS Chen, L., Zhao, L.P. and Gao, Q.K.  
 TITLE Chalcone synthase 2 gene isolated from tea plant (*Camellia sinensis* var. *sinensis*)  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 506)  
 AUTHORS Chen, L., Zhao, L.P. and Gao, Q.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-MAR-2004) Key Laboratory of Tea Chemical Engineering, Ministry of Agriculture, Tea Research Institute, Chinese Academy of Agricultural Sciences, 1 Yunqi Road, Hangzhou, Zhejiang 310008, China  
 FEATURES Location/Qualifiers  
 source 1..506  
 /organism="Camellia sinensis var. sinensis"  
 /mol\_type="mRNA"  
 /variety="sinensis"  
 /cultivar="Longjing 43"  
 /db\_xref="taxon:238974"  
 /note="common: green tea"  
[CDS](#) <1..336  
 /note="CHS2"  
 /codon\_start=1  
 /product="chalcone synthase 2"  
 /protein\_id="[AAS83525.1](#)"  
 /db\_xref="GI:46242832"  
 /translation="GRKSLNEAFQPLNITDGNSLFWIAHPGGPAIWDQVELKLALKPE  
 RLRATRQVLSEYGNMSSACVWFILDEMRSKSAKGLKTTGEGLDGGVLLGFGPGLTVE  
 TVVLHSVST"

ORIGIN

1 gggagaaaga gtctaaatga ggcattccaa cccttgaaca tctatgatgg gaattccctt  
61 ttctggatag cccatcctgg tggccctgcc atttgggacc aagtggaatt gaagttggcc  
121 cttagcccg agaggcttcg ggcacgaga caggtgctaa gtgagtacgg caacatgtcc  
181 agcgcttgtg tgtggtttat cctagatgag atgagaaaga gttcagcaaa gaaaggactc  
241 aagaccacgg gtgaggggct cgacgggggt gtgctcttgg gggtcggacc agggctcacc  
301 gttgagactg tggtgctaca tagcgtgtct acttaatttt aaaatgttgc tctatcattg  
361 aacttggttg tagtgaggtc ttttttttt ttaatgtgtg attgtgttgg cttttattgg  
421 gttccacttt gcttgggtgg ttggcatttg gtaatgatac tcatgccgtg aattaaaatt  
481 taaataatat atgtttctgt gacaaa

//

# 作者简介

姓名：赵丽萍

性别：女

出生日期：1979 年 5 月 25 日

出生籍贯：黑龙江省肇东市

学历：硕士研究生

工作经历：无

曾获何种荣誉：无

发表论著情况等：

赵丽萍，陈亮，高其康. 茶树基因克隆、遗传转化研究进展及 EST 在功能基因组研究上得应用. *第三届海峡两岸茶业学术研讨会论文集*，2003, pp :292-300

陈亮，赵丽萍，高其康. 茶树新梢 cDNA 文库的构建和 EST 测序成功率的初步分析. *茶叶科学*，2004，24(1)：18-22

陈亮，赵丽萍，高其康. cDNA 文库的建立和 EST 测序分析：一个发现和鉴定茶树功能基因的高效快速新途径. *第三届全国青年茶学学术讨论会论文集*，2003，pp:1-7

陈亮，赵丽萍，高其康. 茶树新梢 cDNA 克隆测序和表达序列标签(ESTs)特性分析. *农业生物技术*. 2004, 12(已接受)