

学校编码: 10384

分类号 \_\_\_\_\_ 密级 \_\_\_\_\_

学号: B200326020

UDC \_\_\_\_\_

厦 门 大 学

博 士 学 位 论 文

白氏文昌鱼 BAC 文库的构建  
及 Pax 基因家族的进化分析

Construction of BAC library for *Branchiostoma belcheri*  
and evolutionary study of Pax gene family

王蔚

指导教师姓名: 王义权 教授

专 业 名 称: 生化与分子生物学

论文提交日期: 2006 年 7 月

论文答辩时间: 2006 年 8 月 30 日

学位授予日期:

答辩委员会主席: 章晓波 研究员

评 阅 人: 王金星, 卢大儒,

陈建秀, 刘红林, 史庆华

2006 年 8 月 30 日

## 厦门大学学位论文原创性声明

兹呈交的学位论文，是本人在导师指导下独立完成的研究成果。本人在论文写作中参考的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文产生的权利和责任。

声明人（签名）：

年 月 日



## 目 录

中文摘要.....	1
英文摘要.....	3
<b>第一章 文献综述</b> .....	<b>5</b>
<b>一 基因组文库的构建和应用研究进展</b> .....	<b>5</b>
1 基因组文库的产生和发展.....	5
2 细菌人工染色体文库的构建、保存和鉴定.....	7
3 细菌人工染色体文库的应用.....	8
<b>二 文昌鱼分子生物学研究进展</b> .....	<b>10</b>
1 文昌鱼在系统发生中的地位.....	10
2 利用发育调控基因识别文昌鱼和脊椎动物之间身体结构同源性.....	12
3 文昌鱼特异的基因倍增.....	16
4 文昌鱼比较基因组学研究.....	18
5 现有的文昌鱼基因组资源.....	19
<b>三 Pax 基因家族研究进展</b> .....	<b>19</b>
1 Pax 蛋白的结构.....	22
2 Pax 基因的表达.....	21
3 Pax 基因与与肿瘤发生的关系.....	23
4 Pax 基因突变与疾病的关系.....	23
5 与 Pax 基因相关的非编码区保守序列.....	24
<b>四 论文的选题和目的</b> .....	<b>27</b>
参考文献.....	28
<b>第二章 白氏文昌鱼细菌人工染色体文库的构建</b> .....	<b>43</b>
<b>一 材料和方法</b> .....	<b>44</b>
1 高分子量 (HMW) 基因组 DNA 的抽提.....	45
2 部分酶切和大片断 DNA 的回收.....	46
3 电击感受态制备.....	47
4 连接与转化.....	47

5 重组克隆插入片断大小的鉴定.....	47
6 单克隆分离和培养.....	48
7 BAC 文库的筛选.....	48
<b>二 结果</b> .....	49
1 洗涤缓冲液的选择.....	49
2 预电泳效果.....	50
3 部分酶切条件的选择.....	52
4 片断选择.....	53
5 连接与转化.....	55
6 外源 DNA 的选择.....	55
7 文库的建立.....	57
8 文库的鉴定.....	58
9 PCR 法从 BAC 文库中筛选含 Pax 基因的阳性克隆.....	59
<b>三 讨论</b> .....	62
1 大片断 DNA 的提取方法的改进.....	62
2 连接的比例和浓度.....	63
3 文库的质量.....	64
4 文库的筛选.....	64
<b>参考文献</b> .....	65
<b>第三章 Pax 基因家族系统进化分析</b> .....	69
<b>一 材料与方法</b> .....	70
1 BAC 克隆的测序.....	70
2 文昌鱼 BAC 克隆含有基因的预测.....	71
3 Pax 基因的比较分析.....	71
<b>二 结果</b> .....	72
1 包含 Pax 基因的 BAC 克隆的测序结果.....	72
2 Pax1/9 亚基因家族的序列与结构比较.....	74
3 Pax2/5/8 亚家族基因的序列与结构比较.....	80

4 Pax3/7 亚基因家族的序列与结构比较.....	85
5. Pax4/6 亚基因家族的序列与结构比较.....	90
6 海胆和文昌鱼中非典型 Pax 基因的发现.....	92
7 Pax 亚家族的四簇聚类定位分析.....	96
8 Pax 基因家族的系统发生关系.....	98
<b>三 讨论</b> .....	101
1 Pax4/6 亚家族倍增时间.....	101
2 Pax 亚家族基因结构的进化.....	101
3 文昌鱼比海鞘更接近于脊椎动物的直接共同祖先.....	103
4 Pax 基因家族的演化.....	105
<b>参考文献</b> .....	108
<b>第四章 Pax1/9 相关的旁系同源域和非编码区保守域的鉴定</b> .....	119
<b>一 材料与方法</b> .....	120
1 识别与 Pax 连锁的旁系同源基因.....	120
2. 脊椎动物和非脊椎动物同源基因组序列间比较.....	120
<b>二 结果</b> .....	121
1. 脊椎动物中 Pax1/9 相关的旁系同源域.....	121
2. 非脊椎动物中同源基因的连锁.....	123
3 连锁同源基因的系统进化分析.....	125
4. 非编码保守域的发现.....	131
<b>三 讨论</b> .....	134
1 含 Pax1/9 基因的同线型区域的进化.....	134
2 Pax1/9 基因座的非编码保守序列.....	137
<b>参考文献</b> .....	139
<b>总结</b> .....	144
<b>博士期间发表论文</b> .....	146
<b>致谢</b> .....	147

## Contents

<b>Abstract in Chinese</b> .....	1
<b>Abstract in English</b> .....	3
<b>Chapter 1 Introduction</b> .....	5
1. Construction and application of genomic library.....	5
2. Molecular biology study on amphioxus.....	10
3. Overview of study on Pax gene family.....	19
4. Goal of our research.....	27
<b>Reference</b> .....	28
<b>Chapter 2 BAC library construction for <i>B. belcheri</i></b> .....	43
1. Materials and methods.....	44
2. Results.....	49
3. Discussion.....	62
<b>Reference</b> .....	65
<b>Chapter 3 Evolutionary study of Pax gene family</b> .....	69
1. Materials and methods.....	70
2. Results.....	72
3. Discussion.....	101
<b>Reference</b> .....	108
<b>Chapter 4 Characterization of non-coding evolutionary conserved regions in Pax1/9 harboring Paralogon</b> .....	119
1. Materials and methods.....	120
2. Results.....	121
3. Discussion.....	134
<b>Reference</b> .....	139
<b>Conclusion</b> .....	144
<b>Publications</b> .....	146
<b>Acknowledgements</b> .....	147

## 白氏文昌鱼BAC文库的构建及Pax基因家族的进化分析

### 摘要

文昌鱼是现存生物中最近似于脊椎动物直接祖先的一个类群,在动物进化史上占据了极其重要的地位,是进化生物学、比较基因组学和发育生物学等研究的重要生物模式。白氏文昌鱼(*Branchiostoma belcheri*)分布于西太平洋沿岸的浅海沙滩,是目前世界上研究较多的几种文昌鱼之一。本研究在系统地摸索并优化了文库构建过程中几个关键技术环节的基础上,构建了国内外第一个白氏文昌鱼细菌人工染色体(Bacterial artificial chromosome, BAC)基因组文库。该文库由45,312个克隆组成,存放于118块384孔板中。通过对随机挑取318个的克隆进行脉冲场凝胶电泳检测,估算出文库克隆平均插入片断为120 kb,相当于文昌鱼基因组大小的12倍,从文库中获得任意目的片断的概率高达99.9995%。我们采用二轮PCR的方法对4个单拷贝的文昌鱼Pax基因进行筛选,共获得24个阳性克隆,平均每个基因对应6个克隆。结果显示这个文昌鱼的文库质量较高,将成为比较基因组研究的宝贵的资源,为进一步研究脊椎动物的基因组进化奠定了一定的基础,

Pax基因编码一组胚胎发育过程中起重要作用的转录因子。为了研究该基因家族的进化历史,本研究将含有Pax基因的4个阳性BAC克隆进行测序,共得到474 kb序列。分析结果表明四个白氏文昌鱼的Pax基因*BbPax1/9*、*BbPax2/5/8*、*BbPax3/7*和*BbPax6*分别跨越了15.9 kb、68.1 kb、13.1 kb和38.2 kb的基因组序列,包含外显子数目分别为5、9、5和11,编码氨基酸长度分别为363、444、470和461氨基酸。与佛罗里达文昌鱼的同源基因相比较,氨基酸水平上的同源性97.2%、98.9%、70.1%和97.0%,内含子位置完全一致。

结合17个物种的基因组序列的检索和GenBank中已有cDNA的同源性搜索,本研究从数据库中收集了Pax基因家族成员的序列并对其基因结构进行了注释。对4个文昌鱼Pax基因所属的4个亚家族分别进行比较分析,结果显示:1. 四个亚家族的复制都发生在脊椎动物分化早期。其中*Pax1/9*、*Pax2/5/8*和*Pax3/7*的进化历史在系统树中得到了清

晰的演绎，而*Pax4/6*亚家族的关系比较不明确且在以往研究存在争议。*Pax4*基因是*Pax*基因家族中分化程度最大的一个成员，将之纳入构树分析会导致误差的引入。对内含子位置的比较显示，在脊椎动物的*Pax4*和*Pax6*在成对域中有一个内含子位于氨基酸44和45之间，而无脊椎动物的*Pax6*基因中则位于氨基酸55和56之间。据此推断在脊椎动物形成早期位于氨基酸44/45间的内含子转移到55/56，而后发生*Pax4/6*基因的复制，由此解决了该亚家族分化时间的问题。2. 所有非脊椎动物中文昌鱼基因和脊椎动物直系同源基因间相似程度最高。在构建的进化树中，文昌鱼的4个*Pax*基因都位于脊椎动物和其他非脊椎动物基因之间；在基因结构上，文昌鱼基因中包含了脊椎动物同源基因所特有的所有保守功能域，而其他非脊椎动物基因则或多或少的缺失了一个或多个功能域；在内含子位置上，文昌鱼基因和脊椎动物基因间同源度最高，其他非脊椎动物多表现出显著的差异。这些证据都支持文昌鱼是现存的与脊椎动物亲缘关系最近的类群。在了解四个亚家族进化历史的基础上，我们进一步采用成对域的氨基酸序列对*Pax*基因家族的所有成员进行系统发生关系分析，根据聚类关系和基因结构的比较得到了*Pax*基因家族较为完整的演化过程。

比较基因组学分析显示，*Pax1/9*位于高度保守的旁系同源域中，大多数哺乳动物基因组中有10对旁系同源基因分别与*Pax1*和*Pax9*保持着连锁关系。根据这些基因的位置、转录方向和系统发生关系分析本文对这个脊椎动物旁系同源域的进化途径进行了讨论。包含*Pax1/9*在内有4个基因在脊椎动物和文昌鱼中呈紧密连锁的关系，在这个同源区域中本文发现8个非编码保守序列存在，这是第一次在非脊椎动物中发现有脊椎动物非编码保守序列的存在。文昌鱼和脊椎动物两个谱系独立进化时间累计超过1亿年，这些序列经受了强大的选择压力说明它们的存在具有重要的生物学意义，对这些序列进行功能研究将对基因调控和发育机制的进化提供重要线索。

**关键词：**白氏文昌鱼；BAC 文库；*Pax* 基因家族；进化；基因结构；非编码保守序列

## Construction of BAC library for *Branchiostoma belcheri* and evolutionary study of Pax gene family

### Abstract

As the closest living invertebrate relative of the vertebrates, amphioxus (subphylum Cephalochordata) occupy a key position in animal evolution and is regarded the best available proxy for the study of the last common ancestor of all vertebrates. In this study we reported the successful construction of BAC library of Chinese amphioxus *Branchiostoma belcheri* after optimizing a series of conditions in the library construction. The resulted library consists of 45,312 clones arrayed in one hundred and eighteen 384-well plates. The average insert fragment size was 120 kb estimated by Pulsed Field Gel Electrophoresis (PFGE) analysis of 318 randomly selected clones. The representation of the library is about 12 equivalent to the genome, allowing a 99.9995% probability of recovering any specific sequence of interest. We further formulated the library into characteristic DNA pools for two-round PCR screening for 4 single copied *Amphi-Pax* genes and identified total of 26 positive clones with average of 6.5 clones for each gene. The result indicates this library is well suited for many applications and should also serve as a valuable resource in comparative genomics study.

The Pax gene family consists of tissue-specific transcriptional regulators and plays key regulatory roles in embryonic and sensory organ development in metazoans. Four BAC clones each containing the amphioxus Pax genes were sequenced and a total of 474 kb sequence were obtained. The *BbPax1/9*, *BbPax2/5/8*, *BbPax3/7* and *BbPax6* spans 59.7 kb, 68.1 kb, 13.1 kb, 38.2 kb of genomic DNA and is composed of 5, 9, 5 and 11 exons, respectively. The deduced proteins are 363, 444, 470 and 466 aa long, comparing to 363, 445, 464 and 483 aa of the corresponding *BfPax* genes. Between the two amphioxus species the amino acid identities of the four are 97.2 %, 98.9 %, 70.1 % and 97.0 % whereas the intron position are identical.

Combining the survey of genomic assembly of 17 species and homologous search against

the cDNA dataset in GenBank, we obtained the *Pax* gene family members from public database and had their gene organization annotated. Comparison of the four *Pax* subfamilies show that 1: The four subfamilies of *Pax1/9*, *Pax2/5/8*, *Pax3/7* and *Pax4/6* were expanded at the onset of vertebrate evolution. The first three are supported by phylogenetic analyses while the fourth is less clear since *Pax4* is the most divergent member of *Pax* gene family and inclusion of sequences that show large differences would hamper tree-making methods to recover phylogenetic affinities with high confidence. Inspection of splice site locations revealed the presence of an intron between amino acid 44 and 45 of the paired domain in both vertebrate *Pax4* and *Pax6* members. However, in protochordates *Pax6* the counterpart is present between 55/56. We infer that the intron between 55/56 had transferred to 44/45 before the duplication gave rise to *Pax4* and *Pax6*, therefore the timing of duplication is solved. 2. Among all invertebrates the amphioxus *Pax* gene show the highest homology to the vertebrate orthologues as far as phylogenetic affinity, functional domain combination and intron position are concerned. Evidence support that amphioxus is the closest relative of vertebrates. Furthermore, we conducted phylogenetic analysis on the amino acid sequence of paired domain of all members belonging to *Pax* family. Based on the topology and gene organization we present the complete evolutionary scenario of *Pax* gene family.

Comparative genomics analysis revealed the *Pax1/9* genes lie in regions of extensive paralogy. In the genome of most mammalian species there are 10 pairs of paralogues show strong physical association to *Pax1* and *Pax9*. Based on the positional and phylogenetic relationship of these genes we analysed the evolution of paralogous region. A tightly linked four-gene cluster is conserved in amphioxus and vertebrate. Within this region we identified the 8 ncECR. This is the first time that the vertebrate non-coding sequence is found to be conserved in invertebrate species. These sequences that have been conserved through big phylogenetic distance promised great biological importance. Further functional study would provide insight into the evolution of gene regulation and developmental mechanism.

**Key words:** *Branchiostoma belcheri*; BAC library; *Pax* gene family; evolution; gene organization; non-coding evolutionary conserved region

## 第一章 文献综述

### 一. 基因组文库的构建和应用研究进展

随着后基因组学时代来临, 克隆、研究并利用生物的重要基因、研究它们的结构、功能和进化已经成为遗传学领域的热点之一。和原核生物相比, 真核生物基因组结构复杂, 研究起来较为困难。随着大片断 DNA 克隆技术的发展和完善, 构建真核生物大片断 DNA 的基因组文库, 并以之为平台进行基因组序列测定、物理作图、基因分离、以及基因结构核功能的分析等研究, 已经成为真核生物基因组研究中行之有效的方法<sup>[1]</sup>。

基因组文库是指含有某种生物全部基因的随机片断的重组 DNA 克隆群体。这个文库象是一个贮存有基因组全部序列的信息库, 故称为基因组文库 (genomic library), 又称为人工构建的基因“活期储蓄所”(genomic bank)。人们既可以通过基因组文库的构建、贮存和扩增特定生物基因组的全部或部分片段, 同时又能够在必需时从基因组文库中调出其中的任何 DNA 片段或目的基因<sup>[2]</sup>。

#### 1 基因组文库的产生和发展

克隆载体作为基因组文库构建的重要媒介, 提高其容纳量一直是重要发展方向之一。自 1973 年 Cohen 构建了第一个质粒载体 pS101 以来, 越来越多的克隆载体相继出现, 使得克隆载体的整体结构和克隆效率有了很大改善。载体的发展大致经过三个阶段: 第一代是以  $\lambda$  噬菌体和粘粒为代表的载体; 第二代的代表为 YAC (yeast artificial chromosome)、BAC (bacterial artificial chromosome) 及 PAC (P1-derived artificial chromosome); 第三代为 BIBAC (binary BAC) 及 TAC (transformation-competent artificial chromosome) 为代表的新型文库载体, 不仅具有第二代载体的所有特征, 而且具备了植物的转化功能<sup>[3]</sup>。

##### 1.1 第一代载体

$\lambda$  噬菌体是最早使用的克隆载体, 其插入片断仅 20 kb 左右<sup>[4]</sup>。1979 年 Royal 等确定了在粘粒 (Cosmid) 载体中克隆大片断的可能性<sup>[5]</sup>。此后粘粒载体用于构建果蝇、小鼠、人等基因组文库, 并从这些文库中成功分离得到基因。肺炎支原体基因组全序列的测定就是粘粒文库的基础上完成的<sup>[6-8]</sup>。粘粒是用包含  $\lambda$  噬菌体 cos 位点的质粒。载体

长 4~6 kb, 平均插入一般为 40 kb 左右。重组的粘粒可承载外源 DNA 片段, 并象  $\lambda$  噬菌体一样感染大肠杆菌, 在细菌细胞中复制和增殖。粘粒载体优点在于稳定性好, 嵌合体少。缺点是插入片断过小, 若用于染色体步行 (chromosome walking) 则所需步行次数太多, 很大程度限制了它的应用。

## 1.2 第二代载体

第二阶段的克隆载体与第一代载体相比, 显著特点是载体的容载能力扩大。酵母人工染色体 YAC 是最早发展的真正意义上的人工染色体, 具备在寄主细胞中稳定地复制和遗传所必须的三个元件: 复制起始区 (origin of replication), 参与染色体 DNA 复制起始结构的形成; 着丝粒 (centromere), 负责染色体向子细胞的传递; 端粒 (telomere), 对染色体 DNA 两个末端起封口和保护作用<sup>[9]</sup>。酵母人工染色体的平均插入片段可达到 1 Mb 左右, 因此构建基因组文库时, 只需要较少的克隆数便可以覆盖整个基因组。因此 YAC 很快成为构建复杂基因组 DNA 文库的主要载体系统。在人类基因组计划的早期, YAC 文库被用于基因组框架的构建<sup>[10]</sup>。

虽然 YAC 能容载较大的片段, 但同时也有一些自身难以克服的缺点: 一是嵌合现象的存在。即一个 YAC 克隆中的插入子可能来自两个或多个不同的片段, 嵌合体的比例达到 40%~60%, 这种现象使染色体步行和基因分离难度加大; 二是 YAC 克隆的稳定性差, 插入片段存在重排 (sequence rearrangement) 和丢失 (insert deletion) 现象, 这对染色体物理图谱的构建和基因分离都十分不利; 三是插入片段的分离和纯化困难; 四是不能采用电转化, 转化效率低。这些缺点限制了 YAC 的应用<sup>[11, 12]</sup>。因此科学家开始把眼光转向寻找更好的载体系统。1992 年 BAC 载体应时而生<sup>[13]</sup>。

细菌人工染色体 (BAC) 是以大肠杆菌致育因子 (F 因子) 为基础的合成载体。F 因子具有稳定遗传的特点, 其复制是严谨型的, 在每个细胞中仅有 1~2 个拷贝, 减小了插入片段发生重组的机率, 理论上 F 因子能够携带 1 Mb 的外源片段, 在实际应用中克隆的平均大小约为 120 kb, 个别 BAC 重组子中含有的基因组 DNA 最大可达 300 kb。虽然容量没有 YAC 载体大, 但 BAC 载体拥有 YAC 载体所缺乏而科学家所需的多个优点: 一是以大肠杆菌为寄主, 采用电转, 转化率高, 因此构建 BAC 文库比 YAC 文库更容易; 二是 BAC 载体以环型超螺旋状态存在, 从大肠杆菌中提取质粒较方便, 载体

中的 lacZ 元件使人们能够利用颜色差异区分重组子，限制性位点能够用于大片段基因组 DNA 的克隆。三是 BAC 的复制子来源于 F 因子，可稳定遗传，嵌合及重组现象少；四是可以通过菌落原位杂交来筛选目的基因，方便快捷；五是 BAC 载体在克隆位点的两侧具有 T7 和 Sp6 聚合酶启动子，可以用于转录获得 RNA 探针或直接用于插入片段的末端测序<sup>[14-16]</sup>。

1994 年，Ioannou 等发展了源于 P1 噬菌体的人工染色体（P1-derived artificial chromosome, PAC）<sup>[17]</sup>。这种载体结合了 P1 和 F 因子的特性，通过电击转化大肠杆菌细胞，以单拷贝形式稳定传代，无嵌合体。基于 PAC 的人类基因组文库插入片段的大小在 60~150 kb 之间<sup>[17,18]</sup>。PAC 载体的一个不足之处是自身片段较大（约 16 kb），构建文库没有 BAC 载体（约 7~8kb）效率高，鸟枪法测序费用较大。另一个缺点在于现有 PAC 载体尚缺乏选择标记<sup>[19,20]</sup>。PAC 载体在基因分离和序列分析当中，可作为 YAC 连续克隆群的重要补充，日本水稻基因组计划（RGP）已把水稻的 PAC 文库用于物理图谱的构建<sup>[21]</sup>。

### 1.3 第三代载体

人工染色体系统虽然有许多优势，但都无法直接进行植物转化，在候选基因的功能研究中需要将外源片段进行亚克隆，因而工作量大，同时也有漏失目的 DNA 片段的可能。近年来，通过对一些已有的基因组文库载体进行改造后发展了既可以用于克隆，又能够直接用于植物转化的大片段双元细菌人工染色体（BIBAC）。1997 年，Hamilton 等结合 BAC 载体和根癌农杆菌（*Agrobacterium tumefaciens*）双元载体的特点，构建了双元 BAC 载体 BIBAC2，这种载体在结构上具有 BAC 载体的复制系统，又加入作用于农杆菌的 Ri 复制子和抗卡那霉素筛选标记，因此 BIBAC 能在大肠杆菌和根癌农杆菌中穿梭复制。Hamilton 等已将一 30 kb 的酵母 DNA 和一 152 kb 的人类基因组 DNA 通过农杆菌介导成功地导入烟草核基因组，并证明能够稳定遗传<sup>[22]</sup>。

1999 年 Liu 等结合 PAC 和双元载体的特点，构建了植物可转化基因人工染色体 TAC（transformation-competent artificial chromosome）载体 pYL TAC7，TAC 载体具有 P1 复制子和 Ri 质粒 pRiA4 复制子，能在大肠杆菌和农杆菌中穿梭复制<sup>[23]</sup>。1998 年，Noutoshi 等以小球藻最小的染色体为模型，设计了植物人工染色体（plant artificial chromosome），

可插入超过 10 Mb 的片断。植物人工染色体可在植物细胞中自主复制,使植物超长 DNA 克隆和表达成为可能<sup>[24]</sup>。

## 2 细菌人工染色体文库的构建、保存和鉴定

近年来在基因组文库中占主导地位的载体主要是人工染色体载体系列: YAC、BAC 和 PAC。三种载体中 BAC 载体具明显优势:它嵌合与重排频率低、易于分离插入 DNA、操作简单和可采用高效的电转化体系等优点弥补了 YAC 的不足;在另一方面, BAC 在转化前无需对重组子进行包装的优点是 PAC 所不及的。所以近年来 BAC 载体得到了迅速的发展并在构建基因组文库方面得到了广泛的应用。

BAC文库构建技术已日渐成熟。其过程主要包括载体(vector)及栓塞(plug)的制备、部分酶切(partial digestion)、大片段的选择(large-size DNA selection)、连接(ligation)、转化(transformation)及克隆的挑取(picking)。阳性克隆的挑选有人工和全自动分子生物学工作台技术(robotics)两种<sup>[25]</sup>。一般通过lacZ的插入失活,显示蓝白斑的负向选择来筛选重组子,但是在基因组文库的构建中常发现有1~10%左右的假阳性,当用工作台操作时,假阳性率更高<sup>[13, 25]</sup>。现在又产生了正向选择的BAC载体pBACe3.6<sup>[26]</sup>,其多克隆位点位于蔗糖致死基因sacB上,重组子可以在含有蔗糖的培养基上进行选择,产生的零背景可使机器人进行挑选重组子时不必进行菌落之间的分辨,便于自动挑取收集。

文库的保存,已经发展了单克隆保存法<sup>[27]</sup>和混合池存放法<sup>[28]</sup>两种方法。后者对较大基因组的生物文库保存来说是一个好的选择。

文库的鉴定内容包括插入片段长度、空载率、克隆的多少及细胞器 DNA 的污染等。一个好的文库应该具有较大的平均插入片段及覆盖率,较小的空载率,细胞器 DNA 的污染率很低<sup>[29]</sup>。不同用途对文库的要求有很大的差别。如用来大规模测序,则要求插入片段越长越好;如克隆一个不大的基因,则片段较小有利于转化工作。文库的覆盖度也可以根据研究的需要有所差异,如用于克隆基因和测序,则覆盖率要求高;如用于克隆着丝粒的重复序列,那库容量则可大大缩小,因为阳性克隆的筛选比例最高可以达到10%以上<sup>[30]</sup>。

## 3 细菌人工染色体文库的应用

高质量的基因组文库的建成极大的方便了基因、基因组的研究工作。在大片段基因组文库基础上可进行大规模的物理作图 (physical mapping)<sup>[31]</sup>、测序 (sequencing)<sup>[32]</sup>、基因定位克隆 (position cloning)<sup>[33]</sup>及基因转化、分子标记 (molecular marker) 的发掘<sup>[34]</sup>、着丝粒 (centromere)<sup>[30, 35]</sup>及比较基因组 (Comparative genomics) 的研究工作<sup>[36, 37]</sup>。

### 3.1 图位克隆

基因的图位克隆 (map-based cloning) 又称定位克隆 (positional cloning), 用该方法分离基因是根据目的基因在染色体上的位置进行的, 无需预先知道基因的DNA顺序, 也无需预先知道其表达产物的有关信息, 但首先要有一个根据目的基因建立起来的遗传分离群体, 然后开展以下工作: 首先找到与目的基因紧密连锁的分子标记, 然后用遗传作图和物理作图将目的基因定位在染色体的特定位置, 构建含有大插入片段的基因组文库 (BAC或YAC)。以与目的基因连锁的分子标记为探针筛选基因组文库, 用获得的阳性克隆构建目的基因区域的重叠群; 找到与目的基因紧密连锁的分子标记和鉴定出分子标记所在的大片段克隆以后, 接着是以该克隆为起点进行染色体步行, 逐渐靠近目的基因, 以该克隆的末端作为探针筛选基因组文库; 鉴定和分离出邻近基因组片段的克隆, 再将这个克隆的远端末端作为探针重新筛选基因组文库; 继续这一过程, 直到获得具有目的基因两侧分子标记的大片段克隆或重叠群; 随后通过亚克隆获得含有目的基因的小片段克隆。最后通过遗传转化和功能互补验证最终确定目的基因的碱基序列<sup>[2]</sup>。

### 3.2 物理图谱的构建与全基因组测序

基因组物理图谱是指一系列DNA的限制性酶切片段, 沿染色体的有序排列形式。它与传统的遗传图谱的主要区别在于, 能反映基因组中基因或标记间的实际距离, 是基因组组织结构真实反映; 它所容纳的信息量大, 不仅包含了基因的编码区, 也包含了内含子和基因的调控区。物理图谱的构建是通过菌落杂交, PCR和DNA指纹等方法来确定克隆之间的重叠关系, 将顺序重叠的克隆片段排列在一起<sup>[31]</sup>。

基因组测序需要把已确定物理位置的克隆大片段进行亚克隆, BAC文库性质稳定, 保真度高, 而且与普通的自动化DNA制备纯化程序相容, 所以成为大规模测序中优先选择的克隆系统<sup>[32]</sup>。BAC既可以亚克隆到质粒上进行测序, 也可以直接作为原始底物

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to [etd@xmu.edu.cn](mailto:etd@xmu.edu.cn) for delivery details.

厦门大学博硕士论文摘要库