

基因组研究的关键平台——大片段文库

Key of genome research: large-insert genomic DNA libraries

张 洋^{1,2}, 张晓军¹, 刘 斌¹, 李富花¹, 相建海¹

(1. 中国科学院 海洋研究所, 山东 青岛 266071; 2. 中国科学院 研究生院, 北京 100039)

中图分类号: Q78

文献标识码: A

文章编号: 1000-3096(2008)04-0074-08

大片段基因组文库由于其较大的容量、遗传的稳定性和易操作性等特点而被广泛应用于人类、动植物和微生物的基因组研究中,在物理图谱构建、全基因组测序、基因图位克隆、基因结构功能和进化等研究中起到了重要作用。近年来新技术、新设备的发明和应用使得克隆载体的结构不断优化,稳定性和转化效率显著提高,大片段文库逐渐成为基因组研究的关键平台。

1 大片段基因组文库发展

1.1 概述

构建基因组文库一直是进行基因组研究的一项重要基础性工作。历经几十年的发展,目前大片段基因组文库已经成为研究真核生物基因结构与功能、基因图位克隆、比较基因组学、构建物理图谱等必不可少的工具。

大片段基因组文库的发展过程实际上就是大片段克隆载体系统的发展历程。作为第一代克隆载体的噬菌体^[1]和柯斯质粒^[2]容纳外源 DNA 片段的能力十分有限(分别在 25 kbp 和 45 kbp 左右)。而高等生物基因结构庞大复杂,调控序列往往与之相距很远,难以作为单一片段克隆于这些载体中,因此限制了它们在复杂基因组研究中的应用。

第二代克隆载体是以酵母人工染色体(Yeast Artificial Chromosome, YAC)为主要代表的。该载体是以酵母染色体为基础构建起来的,能容纳超过 1 Mbp 的外源插入片段^[3],可以用来构建复杂基因组文库,满足了高等生物基因组研究的需要。YAC 的出现带来了克隆载体和基因组文库构建的巨大飞跃,很快成为各种基因组计划中主要的克隆工具,相继构建了人^[3,4]、鼠^[5]、稻^[6,7]、拟南芥^[8,9]等一系列大片段基因组文库。尽管 YAC 载体在人类基因组等研究中发挥了不可替代的主导作用,但是由于其自身的一些缺陷,如嵌合、缺失和重排现象,转化效率

低,人工染色体与酵母的天然染色体很难分开使得插入片段不容易分离等,限制了 YAC 载体的大范围应用。

克隆载体发展的第三个时期是以细菌为宿主的大片段克隆载体系统。最初的 P1 噬菌体克隆系统是在噬菌体 P1 基因组的基础上构建起来的,其最大克隆能力是噬菌体的两倍(约 95 kbp)^[10]。

伴随着人类基因组计划的开展,许多动植物的基因组研究也相继开展起来。为了满足不同研究的需要,科研工作者相继开发了另一些以细菌为宿主的质粒载体系统,如细菌人工染色体(Bacterial Artificial Chromosome, BAC)^[11], Fosmid^[12], 噬菌体 P1 衍生的人工染色体(Bacteriophage P1 - derived Artificial Chromosome, PAC)^[13], 二元细菌人工染色体(Binary BAC, BIBAC)^[14]、传统的质粒克隆载体(Plasmid - based Cloning Vectors, PBC)^[15]和可转化的人工染色体(Transformation-competent Artificial Chromosome, TAC)^[16]等。与 YAC 相比,它们具有嵌合率低、转化率高、不易缺失和重排、外源 DNA 容易分离等特点,因此逐渐取代了 YAC 载体,得到了广泛应用。但需要强调指出的是,这些所谓的“人工染色体”是经过人工改造的质粒载体,没有细菌染色体的任何成分,它们只是以细菌作为容纳大片段 DNA 的宿主而已。

1.2 BAC 及其衍生载体系统

1992 年, Kim 等^[12]利用 BAC 载体的前体——pBAC 和 pUCcos 融合构建了 pFOS1 黏粒载体。它具有容纳 40 kbp 外源 DNA 片段的能力,拷贝数低,较柯斯

收稿日期:2006-01-04;修回日期:2006-03-07

基金项目:国家自然科学基金资助项目(30230280,30200213)

作者简介:张洋(1980-),男,山东青岛人,博士研究生,主要研究方向:分子生物学;相建海,通讯作者,电话:0532-82898568, E-mail: jhxiang@ms.qdio.ac.cn

质粒稳定等优点。随后 Kim 等利用 pFOS1 构建了人类白血病细胞系 Molt4 的 DNA 文库,并验证了插入片段的遗传稳定性。同年,两人又成功构建了另一种新型克隆载体系统——BAC 载体系统^[11]。由于它具有 YAC、P1 噬菌体载体系统无可比拟的优点而广泛应用于构建人、动植物基因组 DNA 大片段插入文库。

世界上第一个 BAC 载体 pBAC108L 起源于 *Escherichia coli* 的 F 因子(F-factor)。天然状态下 F 因子是一个 100 kbp 左右的闭合环状 DNA 分子,其拷贝数受宿主基因组的严格控制,通常每个细胞只有 1~2 个拷贝。1992 年 Shizuya 等^[11]通过删除 F 因子上一些非必需序列并添加 *CM^r* 基因作为载体筛选标记得到了 pBAC108L,它具有稳定克隆超过 300 kbp 外源 DNA 片段的能力,但缺乏重组标记。重组克隆必须通过菌落杂交进行筛选,操作复杂费时且不准确,因此不适用于文库构建工作。为此,1996 年 Kim 等^[17]将 *lacZ* 基因引入 pBAC108L 载体作为重组筛选标记构建了 pBeloBAC11 载体。1997 年 Birren 等^[18]构建了 pIndigoBAC,该载体来源于 pBeloBAC11,由于 *lacZ* 基因的 3' 末端的一个点突变,能够使 pIndigoBAC 载体产生比单拷贝的 pBeloBAC11 更深的蓝色,从而使重组子的筛选更加迅速、精确。此后,为了方便含有 *EcoR* 大片段 DNA 的插入,Frijters 等^[19]去除了 pBeloBAC11 载体 *CM^r* 基因中的 *EcoR* 位点构建了 pECBAC1 载体。目前,pBeloBAC11 和 pECBAC1 已经被广泛用于文库构建,pECBAC1 的分子结构如图 1 所示,*oriS* 和 *repE* 基因调节 F 因子的自主复制的方向而 *parA* 和 *parB* 基因负责控制 F 因子的拷贝数。pECBAC1 的 *lacZ* 基因中含有 *Hind*、*BamH* 和 *EcoR* 三个限制性酶切位点,在克隆位点两侧除噬菌体 T7 和 SP6 启动子序列外还有两个 *NotI* 位点可以方便地将插入片段从载体上切除下来。此外,这个载体上还有噬菌体 *cosN* 位点和 P1 噬菌体 *loxP* 位点。1998 年 Mozo 等^[20]将 *Km^r* 基因插入到 pBeloBAC11 的 *CM^r* 基因的 *EcoR* 位点,构建了可克隆 *EcoR* 大片段 DNA 的 pBeloBACKan 载体。Frengen 等以 pBAC108L 和 pCYPAC2^[21]为基础构建了 pBACe 3.6 载体,其多克隆位点位于蔗糖致死基因 *sacB* 上,这样重组子可以在含有蔗糖的培养基上进行选择^[22]。2000 年,Fu 等^[23]在 pBeloBAC11 载体的基础上去除了 *lacZ* 基因两侧的 *NotI* 位点,并在 *BamH* 和 *EcoR* 中间插入了一个 *NotI* 位点,成功构建了能够克隆 *NotI* 片段的载体

pNOBAC1,*NotI* 是一个甲基化敏感的八碱基限制性内切酶,在基因组中的切点相对较少。使用这种载体构建文库可以减少筛选相关基因所需的克隆数,极大减少工作量。随后,Choi 等^[24]将 *Cre/lox*,*htp*,*nos3* 等位点引入 pBACKAN,构建了 pBAC-wich 载体。*Cre/lox* 位点特异性重组系统使外源基因可以插入到真核生物染色体特定的位点,避免了由于外源 DNA 片段的随机插入而造成的“位置效应”,使遗传互补验证试验更为精确。

由于第一代 BAC 载体 pBAC108L 缺少重组筛选标记而未能广泛应用于大片段基因组文库的构建,为了克服上述困难,Loannou 等^[13]构建了 PAC 载体 pCYPAC-1。该载体是在 P1 噬菌体载体 pAd10SacB 的基础上去除了 Ad10 片段并在其 *BamH* 克隆位点插入了 pUC19 片段构建而成的,使其兼具了 P1 噬菌体载体和 BAC 载体的优点。在

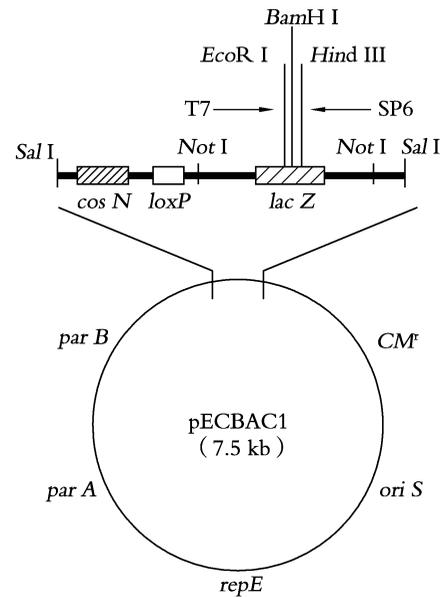


图 1 pECBAC1 载体

构建大片段基因组文库的工作中 PAC 基本上可以和 BAC 互相替代。尽管如此,它们在结构上还是有许多差异:(1) 筛选标记不同:PAC 载体是以 *sacB* 基因作为重组筛选标记,*Km^r* 基因作为抗性筛选标记。而 BAC 载体是以 *lacZ* 基因为重组筛选标记,*CM^r* 基因作为抗性筛选标记;(2) 拷贝数不同:与 P1 噬菌体和 BAC 载体类似,PAC 通常以单拷贝形式存在于宿主细胞内,但它本身有一个 P1 裂解子(P1 lytic replicon),可在 IPTG 诱导下产生多拷贝;在 pUC19 片段内含有一个质粒高拷贝数复制起点,方便了 PAC 载体的制备;(3) 载体大小不同:PAC 载体自身片段较大(约 16 kbp),构建文库没有 BAC 载体(约 7~8 kbp)效率高。

高等生物的基因组巨大,调控序列复杂,许多性状都是由在位置上紧密连锁的多个基因组成的基因簇联合控制的。这样在研究基因结构,验证基因功能的时候就需要将完整的大片段转化缺陷型品系。而传统的BAC、PAC等载体不具备将插入的大片段DNA直接转化植物的能力,而且也难以找到合适的限制性内切酶将插入片段完整地克隆到可转化生物的载体上。为了克服这个困难,Hamilton和Liu分别开发出了双元载体系统——BIBAC和TAC。这两种载体不仅能用于大片段插入文库的构建,还可以通过农杆菌介导的方法直接将插入片段导入植物体中。双元载体的成功构建加速了基因的图位克隆,而且对于实现多基因控制的品种改良也具有重要意义。

1996年,Hamilton等在BAC载体pBAC108L和农杆菌Ri复制子的基础上构建了BIBAC2,其能在大肠杆菌和根癌农杆菌中以单拷贝形式穿梭复制。该载体以 Km^r 基因为抗性筛选标记, $sacB$ 基因为重组筛选标记。BIBAC2载体含有两个植物转化筛选标记:一个是GUS-NPTII(葡糖醛酸糖苷酶-新霉素磷酸转移酶)。它是GUS与NPT的融合多肽,GUS用于鉴定蛋白表达活性,NPT则使载体获得卡那霉素抗性。另一个标记是潮霉素转移酶基因(*hygromycin phosphotransferase*, *hpt*),编码潮霉素转移酶,使载体获得潮霉素抗性。Hamilton等^[18]已用一段150 kbp的人类基因组DNA通过农杆菌介导成功地导入烟草和番茄的核基因组,并证明能够稳定遗传。BIBAC2的出现极大推动了高等植物的基因组研究,但随着越来越广泛的应用,BIBAC2也暴露出许多缺点:仅有 $BamH I$ 一个克隆位点;在大肠杆菌和根癌农杆菌中以单拷贝形式存在,质粒DNA的产量有限;以 $sacB$ 基因为重组筛选标记,其假阳性比例较 $lacZ$ 基因高,可达10%。

1999年,Liu等^[16]构建了可转化的人工染色体pYL TAC7,即TAC载体。该载体具有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。pYL TAC7载体在T-DNA右边界处插入了植物选择标记 hpt 基因。在TAC载体中抗性筛选基因是 Km^r 基因,重组筛选标记是 $sacB$ 基因。与BIBAC2相比,pYL TAC7含有 $Hind III$ 和 $BamH I$ 两个克隆位点,此外还有一个P1裂解子,可在IPTG诱导下产生多拷贝。与其他载体不同的是,pYL TAC7在克隆位点附近只有一个 $Not I$ 位点,当用 $Not I$ 酶切检测插入片段大小的时候应该减去23 kbp的载体长度。Liu等以pYL TAC7为载体构建了拟南芥的基因组文库,并把部分克隆通过农杆

菌介导转化拟南芥,获得了转化植株并能稳定遗传。

后来的研究发现载体拷贝数与遗传稳定性之间并不存在直接的联系,而插入片段的大小与转化效率有关:片段越大,转化效率越低。这一发现为克隆载体的发展提供了更为广阔的空间。Tao和Zhang 1998年的研究发现借助电转化技术,一些传统的载体如Plasmid、Cosmid等与BAC、PAC载体一样具有容纳外源大片DNA的能力,并在此基础上提出了PBC克隆载体的概念,其典型代表是pSLJ1711和pCLD04541。pSLJ1711是双元质粒载体,来源于质粒pKR290^[25],并在 $lacZ$ 基因中引入了dark Bluescript (dBS)多聚接头作为多克隆位点。而双元黏粒载体pCLD04541是在pSLJ1711的基础上插入 cos 位点构成的(图2)。pSLJ1711和pCLD04541可以稳定容纳高达310 kbp的外源DNA且每个细胞可以产生4~5个拷贝,以 npt (*neomycin phosphotransferase*)基因作为植物筛选标记, Tet^r 基因为抗性筛选标记,重组筛选标记是 $lacZ$ 基因^[26]。它们与其他载体相比有许多优点。首先,dBS位点的引入以及相对较高的拷贝数使它们产生比其他载体更深的蓝色,便于重组子的筛选和质粒DNA的提取;其次,有更多的克隆位点,除 $Hind III$ 、 $EcoR I$ 和 $BamH I$ 外,还有 $Xho I$ 、 $Cla I$ 、 $Xba I$ 等位点;再次,由于是双元载体,它们可以像BIBAC2、TAC一样直接用于农杆菌介导的植物转化;最后,以 $lacZ$ 基因作为重组子筛选标记使得它们优于以 $sacB$ 基因为重组子筛选标记的BIBAC2、TAC等双元载体。

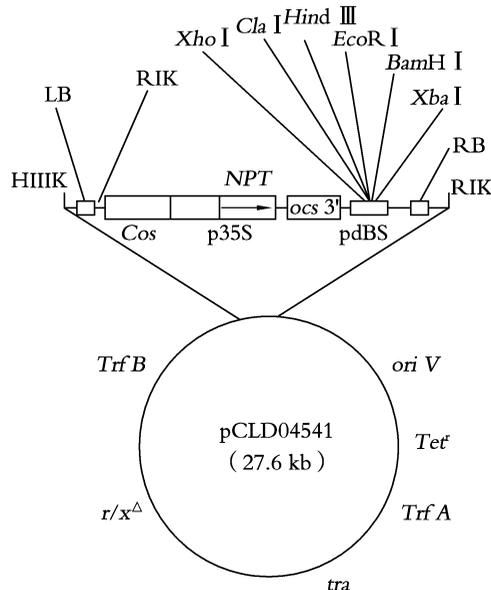


图2 双元黏粒载体pCLD04541

以细菌为宿主的大片段克隆载体系统,由于其较大的克隆能力、插入片段良好的稳定性和易操作性已经被广泛应用于真核生物基因组的文库构建。近几年来,许多大公司相继研发并推出了具有独立产权的大片段基因组文库构建试剂盒,如 Epicentre 公司的 CopyControl™ BAC Cloning Kit 和 Lucigen 公司的 CopyRight pEZ BAC Cloning Kit 等。这些

商品化试剂盒中提供数种经单酶切并去磷酸化的载体 DNA(如 *Bam*H , *Eco*R , *Hind* 等),可直接用于与大片段 DNA 的连接,从而避免了由于载体质量问题而导致的实验失败,大大缩短建库时间。同时这些载体都可被诱导产生多拷贝,方便了质粒 DNA 的提取。表 1 列出了几种在大片段克隆载体系统的发展历程中比较有代表性的载体。

表 1 用于构建大片段基因组文库的部分载体

载体名称	载体类型	载体大小 (kbp)	载体最大容量 (kbp)	载体结构	拷贝数 / 细胞	酶切位点	抗性/重组标记
pFOS1	Fosmid	9.5	40	环状	1	<i>Hind</i> , <i>Bam</i> H I	<i>CM</i> ^r , <i>Amp</i> ^r / 无
pYAC2	YAC	11.5	2 000	线状	1	<i>Sma</i>	<i>ura</i> 3 , <i>trp</i> 1 / <i>SUP</i> 4
pAd10SacBII	P1	16.0	95	线状	1~2 *	<i>Bam</i> H	<i>Km</i> ^r / <i>sac</i> B
pBAC108L	BAC	7.4	300	环状	1~2	<i>Hind</i> , <i>Bam</i> H	<i>CM</i> ^r / 无
pBeloBAC11	BAC	7.4	300	环状	1~2	<i>Bam</i> H , <i>Sph</i> , <i>Hind</i>	<i>CM</i> ^r / <i>lac</i> Z
pIndigoBAC	BAC	7.4	300	环状	1~2	<i>Bam</i> H , <i>Sph</i> , <i>Hind</i>	<i>CM</i> ^r / <i>lac</i> Z
pECBAC1	BAC	7.5	300	环状	1~2	<i>Bam</i> H , <i>Eco</i> R , <i>Hind</i>	<i>CM</i> ^r / <i>lac</i> Z
pBeloBACKan	BAC	8.5	300	环状	1~2	<i>Eco</i> R , <i>Bam</i> H , <i>Bam</i> H , <i>Sac</i> ,	<i>Km</i> ^r / <i>sac</i> B
pBACe3.6	BAC	11.5	300	环状	1~2	<i>Nsi</i> , <i>Eco</i> R , <i>Sac</i> , <i>Mlu</i>	<i>CM</i> ^r / <i>sac</i> B
pNOBAC1	BAC	7.5	300	环状	1~2	<i>Bam</i> H , <i>Not</i> , <i>Hind</i>	<i>CM</i> ^r / <i>lac</i> Z
pBACwich	BAC	11.2	300	环状	1~2	<i>Sph</i> , <i>Hind</i>	<i>CM</i> ^r / <i>lac</i> Z
pCYPAC-1	PAC	19.3	300	环状	1~2 *	<i>Bam</i> H , <i>Sca</i>	<i>Km</i> ^r / <i>sac</i> B
BIBAC2	BIBAC	23.5	300	环状	1~2	<i>Bam</i> H	<i>Km</i> ^r / <i>sac</i> B
pYL TAC7	TAC	22.9	300	环状	1~2 *	<i>Bam</i> H , <i>Eco</i> R	<i>Km</i> ^r / <i>sac</i> B
pSLJ1711	Plasmid	27.4	310	环状	4~5	<i>Xho</i> , <i>Cla</i> , <i>Eco</i> R , <i>Bam</i> H , <i>Hind</i> , <i>Xba</i>	<i>Tet</i> ^r / <i>lac</i> Z
pCLD04541	Cosmid	27.6	310	环状	4~5	<i>Xho</i> , <i>Cla</i> , <i>Eco</i> R , <i>Bam</i> H , <i>Hind</i> , <i>Xba</i>	<i>Tet</i> ^r / <i>lac</i> Z
pIndigoBAC-5 **	BAC	7.5	300	环状	1	<i>Bam</i> H , <i>Hind</i>	<i>CM</i> ^r / <i>lac</i> Z
pCCIBAC **	BAC	8.1	300	环状	1 *	<i>Bam</i> H , <i>Hind</i> , <i>Eco</i> R	<i>CM</i> ^r / <i>lac</i> Z
pEZ BAC ***	BAC	7.2	> 100	环状	1 *	<i>Bam</i> H , <i>Hind</i> , <i>Eco</i> R , <i>Sph</i> , <i>Hpa</i> , <i>Pml</i>	<i>CM</i> ^r / <i>lac</i> Z

注: *可诱导产生多拷贝, **Epicentre 公司产品, ***Lucigen 公司产品

2 大片段基因组文库的构建与鉴定

2.1 文库的构建

2.1.1 高分子质量 (HMW) DNA 的制备

高分子质量 (High Molecular Weight HMW) DNA (>600 kbp) 的成功制备是构建大片段基因组文库的关键。为了避免 HMW DNA 在制备过程中受到物理剪切,一般使用液氮将组织捣碎,然后将收集的细胞核包埋在低熔点琼脂糖中形成 plug^[17,19,27]。

2.1.2 载体的制备

构建大片段基因组文库的另一个关键步骤就是 BAC 载体 DNA 的制备。载体制备的好坏直接关系到文库的空载率和质量。为尽量避免线形载体自连接,载体的脱磷处理尤为重要,脱磷后的载体可在 -20℃ 冰箱长期保存^[19,27]。

2.1.3 HMW DNA 的部分酶切和片段选择

通过特定的限制性内切酶把包埋于 plug 中的基因组 DNA 消化成不同长度的可克隆片段。部分酶切 HMW DNA 时通常要考虑以下几个因素:所用限制性内切酶的种类和浓度;酶切消化的时间;影响酶活力的因子的浓度,如 Mg^{2+} 等。部分酶切后通过两次脉冲场电泳 (Pulse Field Gel Electrophoresis, PFGE) 分别回收大小比较集中的片段 (100 ~ 400 kbp)^[11,19,27]。其目的是筛除过小和过大的片段,因为小片段在与载体连接时比大片段有更大的竞争优势,而过大的片段即使能够连入载体也很难成功转入受体细胞 DH10B。由于每种生物基因组的碱基组成存在差异,所以在实际的建库工作中如果只用一种限制性内切酶的大片段构建基因组文库就很有可能由于酶切位点分布不均匀产生过小或过大片段,最终导致文库的代表性不强。在这种情况下,一味地增加克隆数目是不能增加基因组覆盖率的。除此之外,一些生物基因组中某些区域对于某种载体来说是不可克隆的,或是克隆的片段不能稳定地存在于宿主细胞中,这就需要利用其他载体系统和宿主细胞来克隆这些基因组区域。

研究表明,采用两种或多种碱基偏好不同的限制性内切酶及不同的载体系统构建的大片段基因组文库,具有较好的互补性和代表性,在此基础上可以构建出高质量的物理图谱。例如 2001 年发表的人类基因组物理图谱就是用 *Hind* III, *Eco*R I 等多种限制性内切酶, BAC, PAC 和 YAC 三种载体系统构建起来的^[28]。

2.1.4 连接和转化

大片段 DNA 是否能有效连接到载体上主要取决于载体与 DNA 片段的比例,对于不同生物采用的比例不同,大多数采用摩尔比 $1 : 4 \sim 10$ ^[11,19,27]。

电转化技术的发展和 *E. coli* 菌株 DH10B 细胞

系的发现^[28],使得大片段基因组文库得以迅速发展起来。电转化技术比已往的化学转化效率更高 (> 1×10^{10} cfu (细菌群落总数)/ μ g pUC19 DNA),通过电脉冲可以将插入大片段的载体转入细菌宿主细胞^[11]。DH10B 及其派生菌株由于可以稳定容纳含有大片段的载体而被广泛应用于大片段基因组文库的构建工作^[14]。电转化效率与插入片段的大小有关,插入片段越大,转化效率越低;反之,越高。此外,转化条件如电压、电容、电阻、温度等物理参数也直接影响转化效率、插入片段的大小及假阳性克隆的比例^[30]。

2.1.5 克隆的挑取与保存

克隆的分捡有手工和机器分捡两种。利用蓝白斑筛选挑取阳性克隆于含有相应抗生素的冻存培养基中,37℃ 过夜培养后保存于超低温冰箱中。为了方便筛选,一般将保存于保菌板中的有序克隆制成混合池或点在高密度尼龙膜上。这样就可以通过 PCR 或是杂交快速定位目的克隆。

2.1.6 目的克隆的筛选

高等生物的大片段基因组文库由几万到几十万个克隆组成,因此建立高效的管理和筛选系统是十分必要的。文库的筛选系统有 PCR 筛选系统和尼龙膜杂交筛选系统两种。PCR 筛选系统是将有序的克隆混合成行池、列池、板池以及超级池等一系列不同容量的混合池,利用对目的序列已知的信息设计特异引物,对各级混合池进行 PCR 筛选,从而确定目的克隆的位置。高密度尼龙膜杂交筛选系统是近年来发展起来的新型筛选系统,它具有易保存,操作方便、可反复使用等优点。使用机械手按照一定的矩阵形式将有序克隆按一定的顺序点在膜上,然后放在固体培养基上培养,待菌落长出之后将其裂解,使 DNA 结合在尼龙膜上,就可以用 Southern 杂交的方法筛选阳性克隆。

2.2 文库的鉴定

一个高质量的大片段基因组文库是构建物理图谱、全基因组测序、基因图位克隆等研究的基础。因此,文库的鉴定工作就显得尤为重要,鉴定内容包括插入片的平均大小、基因组的覆盖率、空载率、嵌合率、稳定性和细胞器 DNA 的污染等。

理论上对于一个确定的基因组,平均插入片段越大,覆盖整个基因组所需的克隆数就越少,在构建物理图谱时所需组装的 contig 和所产生的 gap 就减少,图谱的质量也就越高。2000 年, Soderlund 等^[31]以人类基因组一段 110 Mbp 长的序列为材料,利用 FPC (Fingerprinted Contigs) 软件研究了不同大小的插入片段对所构建的物理图谱质量的影响。结果显示,当文库的平均插入片段从 100 kbp 增加到 165 kbp 时, contig 的数目由 212 个减少到 38 个,减

少 82%。而平均插入片段继续由 165 kbp 增加到 200 kbp 时, contig 的数目仅仅缩小到 26 个, 减少 88%。因此, 一个平均插入片段在 150~160 kbp 的基因组文库是构建物理图谱的最佳选择。

基因组文库的覆盖率也是构建物理图谱不可忽略的重要因素。一直以来, 科学家对“用多大覆盖率的大片段基因组文库来构建物理图谱是最有效、最经济的”这个问题存在很大争议——覆盖率太小会出现很多 gap 从而影响物理图谱的质量, 覆盖率太大不但工作量会成倍增加而且是一种不必要的浪费。为了解决这一争论, Xu 等^[32] 2004 年利用计算机模拟了拟南芥第 2 和第 4 号染色体以及人类的第 22 号染色体的 contig 组装回答了这个问题。研究发现, 当覆盖率从 5 × 增加到 10 × 时, 覆盖 3 条染色体的 contig 数目都显著的减少, 而当覆盖率继续从 10 × 增至 15 × 时, contig 的数目几乎没有变化, 甚至在某些情况下还可能增加。因此, 10 × 覆盖率的文库就完全可以满足高质量物理图谱的构建和全基因组测序的需要。

通常一个合格的基因组文库, 克隆的空载率要在 5% 以下。

所谓嵌合(chimerism), 是指连接在一个载体上的插入片段包含有两个或两个以上来源于不同染色体位置的片段。基因组文库的嵌合体率越低, 文库的质量就越高。嵌合率过高会造成基因测序和基因作图的严重错误。

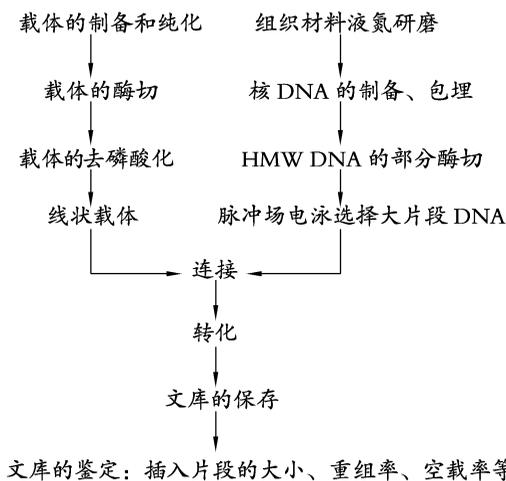


图 3 大片段基因组文库的构建流程

BAC 克隆的稳定性鉴定是指, 随机挑选一定数量的克隆接种培养基连续继代培养, 分别提取第 0 代和第 100 代克隆的质粒 DNA, 酶切、电泳检测 BAC 克隆中外源 DNA 在继代培养中是否丢失或发生重组等。

核基因组文库如果混杂线粒体等细胞器 DNA 不但影响文库质量, 还会把图位克隆引向歧途。在文库构建过程中, 采用脉冲场电泳技术制备的高纯

度 HMW DNA 可以从根本上避免细胞器 DNA 的污染。同时, 可以利用线粒体等细胞特异性探针对大片段基因组文库进行随机的菌落原位杂交, 以检测是否有细胞器 DNA 的污染。

3 大片段基因组文库展望及其在海洋生物研究中的应用

从 1987 年 Burke 等成功将大片段 DNA 克隆到 YAC 载体开始, 大片段基因组文库技术体系不断发展完善并日益显现出其在基因组学研究中不可替代的重要地位。然而, 该体系由于核心技术很难在短时间内有较大的突破, 所以仍存在一些不足之处。首先是对宿主细胞承载能力和转化效率方面的研究: 例如 DH10B 宿主细胞最高只能容纳 310 kbp 的外源 DNA 分子。究其原因, 一方面以方波和指数衰减为基础的电转化技术制约了更大片段 DNA 的进入; 另一方面, 宿主细胞本身的属性也是影响承载能力和转化效率的关键因素, 例如最近 Stratagen 公司的研究表明, 该公司新研发的 XL10-Gold Cells 感受态细胞较传统 DH10B 感受态细胞转化效率高 25 倍, 可以大大缩短建库时间, 节约经费。

其次, 作为绘制物理图谱的关键技术——DNA 测序技术发展缓慢, 尤其是 BAC DNA 末端测序由于分子质量大, 容易产生二级结构等因素导致测序成本仍然比较昂贵。目前, 美国科学家已经对纽约曼哈顿区的空气进行了采样, 准备用新研制的 454 测序平台对所采集空气中的细菌进行全基因组测序。每个 454 测序平台可在 4.5 h 内完成 2 500 万碱基的测序工作, 是传统的 Sanger 测序效率的近百倍, 且成本较低得多^[33], 其有望成为下一代 DNA 测序技术核心。

在人类基因组计划的带动下, 越来越多的动植物基因组计划相继启动。以 BAC, BIBAC 及 PBC 载体等为代表的新型克隆载体系统在动植物和人类基因组研究中发挥了巨大作用。大片段文库已经成为物理作图、基因组测序、图位克隆等的重要工具。在未来, 基因表达调控、基因互作、BAC 转基因技术、品种改良等工作将成为大片段基因组文库利用和发展方向。

21 世纪是海洋世纪, 海洋容纳了地球上生物种类的 80%, 是人类最大也是最后一块尚未开垦的基因宝库。与陆地生物相比, 海洋生物基因组计划的实施起步稍晚, 但海洋生物以其独特的发育和进化地位以及重要的经济价值, 很快成为生物学研究的热点。研究发现, 海洋生物基因组一般都比较庞大, 冗余序列较多, 仅仅利用小片段基因组文库, 尤其是 cDNA 文库往往很难克隆基因全长及其上下游的调控序列, 从而阻碍了功能基因的研究。近年来, 许多

海洋生物的大片段基因组文库已经构建完成,这为深入研究海洋生物基因组、开发利用海洋生物资源奠定了基础。1997年美国农业部正式确定了以牡蛎、对虾、罗非鱼、鲟鱼和鲑鱼5种水产动物的基因组计划;2005年美国水产动物基因组协调委员会提交的《水产经济动物基因组研究》白皮书中再次强调要加强上述5种水产经济动物的基因组研究。经过近10年的努力,科学家利用微卫星、AFLP等标记绘制了它们的遗传连锁图^[34~43];构建了除对虾以外的高质量BAC文库^[44~48];Sieson^[49]和Katagiri^[50]于2005年1月和3月份分别宣布鲑鱼与罗非鱼物理图谱构建完成,牡蛎和鲟鱼的物理图谱也将在最近建成。海洋生物的基因组时代已经来临,而中国海洋生物的基因组研究目前刚刚起步,基础薄弱,总体水平落后于美国等发达国家,也落后于陆地动植物,这就需要积极开展与国内外先进实验室的合作与交流,引进技术和设备,促进中国海洋生物基因组研究的快速发展。

参考文献:

- [1] Murray N E, Murray K. Manipulation of restriction targets in phage lambda to form receptor chromosomes for DNA fragments [J]. *Nature*, 1974, **251** (5 475): 476-481.
- [2] Collins J, Hohn B. Cosmids: a type of plasmid gene-cloning vector that is packageable *in vitro* in bacteriophage lambda heads[J]. *Proc Natl Acad Sci USA*, 1978, **75** (9): 4 242-4 246.
- [3] Burke D T, Carle G F, Olson M. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors[J]. *Science*, 1987, **236**: 806-812.
- [4] Chumakov I M, Rigault P, Le Gall I, *et al.* A YAC contig map of the human genome[J]. *Nature*, 1995, **377**(6 547 Suppl.): 175-297.
- [5] Nusbaum C, Slonim D K, Harris K L, *et al.* A YAC-based physical map of the mouse genome[J]. *Nature Genet*, 1999, **229** (4): 388-393.
- [6] Kurata N, Umehara Y, Tanoue H, *et al.* Physical mapping of the rice genome with YAC clones[J]. *Plant Mol Biol*, 1997, **35**(1-2): 101-113.
- [7] Saji S, Umehara Y, Antonio B A, *et al.* A physical map with yeast artificial chromosome (YAC) clones covering 63% of the 12 rice chromosomes [J]. *Genome*, 2001, **44**: 32-37.
- [8] Canillieri C, Lafleuriel J, Macadre C, *et al.* A YAC contig map of *Arabidopsis thaliana* chromosome 3[J]. *Plant J*, 1998, **14**: 633-642.
- [9] Schmidt R, West J, Love K, *et al.* Physical map and organization of *Arabidopsis thaliana* chromosome 4 [J]. *Science*, 1995, **270**: 480-483.
- [10] Sternberg N. Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs [J]. *Proc Natl Acad Sci USA*, 1990, **87**: 103-107.
- [11] Shizuya H, Birren B, Kim U-J, *et al.* Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector [J]. *Proc Natl Acad Sci USA*, 1992, **89**: 8 794-8 797.
- [12] Kim U J, Shizuya H, de Jong P J, *et al.* Stable propagation of cosmid size human DNA insert in an F factor based vector [J]. *Nucleic Acids Research*, 1992, **20** (5): 1 083-1 085.
- [13] Ioannou P A, Amemiya C T, Garnes J, *et al.* A new bacteriophage P1-derived vector for the propagation of large human DNA fragments [J]. *Nature Genet*, 1994, **6**: 84-89.
- [14] Hamilton C M. A binary-BAC system for plant transformation with high-molecular-weight DNA [J]. *Gene*, 1997, **200**: 107-116.
- [15] Tao Q, Zhang H B. Cloning and stable maintenance of DNA fragments over 300 kb in *Escherichia coli* with conventional plasmid - based vectors [J]. *Nucleic Acids Research*, 1998, **26**: 4 901-4 909.
- [16] Liu Y G, Shirano Y, Fukaki H, *et al.* Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning [J]. *Proc Natl Acad Sci USA*, 1999, **96**: 6 535-6 540.
- [17] Kim U J, Birren B, Slepak T, *et al.* Construction and characterization of a human bacterial artificial chromosome library [J]. *Genomics*, 1996, **34**: 213-218.
- [18] Birren B, Green E D, Klapholz S, *et al.* Bacterial Artificial Chromosomes [A]. Birren B. Cloning Systems (3), In *Genome Analysis: A laboratory manual* [C]. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1997. 242-295.
- [19] Frijters A C, Wang Z, Van Damme M, *et al.* Construction of a BAC library containing large *EcoR* I and *Hind* genomic fragments of lettuce [J]. *Theor Appl Genet*, 1997, **94**: 390-399.
- [20] Mozo T, Fischer S, Shizuya H, *et al.* Construction and characterization of the IGF *Arabidopsis* BAC library [J]. *Mol Gen Genet*, 1998, **258**: 562-570.
- [21] Ioannou P A, de Jong P J. Construction of bacterial artificial chromosome libraries using the modified P1 (PAC) system [A]. Dracopoli N C, Haines J N, Korf B R, *et al.* Current Protocols in Human Genetics (5) [C]. New York: John Wiley & Sons, 1996. 5, 15.
- [22] Frengen E, Weichenhan D, Zhao B, *et al.* A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites [J]. *Genomics*, 1999, **58** (3): 250-253.
- [23] Fu H, Dooner H K. A gene-enriched BAC library for cloning large allele-specific fragments from maize: isolation of a 240-kb contig of the bronze region [J].

- Genome Research**, 2000, 10:866-873.
- [24] Choi S, Begum D, Koshinsky H, *et al.* A new approach for the identification and cloning of genes: the pBACwch system using Cre/*lox* site-specific recombination[J]. **Nucleic Acids Res**, 2000, 28 (7) :E19.
- [25] Ditta G, Stanfield S, Corbin D, *et al.* Broad host range DNA cloning system for Gram negative bacteria: Construction of a gene bank of *Rhizobium melilot* [J]. **Proc Natl Acad Sci USA**, 1980, 77:7 347-7 351.
- [26] Jones J D G, Shlumukov L, Carland F, *et al.* Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants[J]. **Transgenic Research**, 1992, 1:285-297.
- [27] Zhang H B. Manual for construction and manipulation of large-insert bacterial clone libraries [M]. Texas, USA: Texas A & M University, 2000.
- [28] McPherson J D, Marra M, Hillier L, *et al.* A physical map of the human genome[J]. **Nature**, 2001, 409 (6822) :934-941.
- [29] Hanahan D, Jersee J, Bloom F R. Plasmid transformation of *Escherichia coli* and other bacteria [J]. **Methods Enzymol**, 1991, 204:63-113.
- [30] Sheng Y, Mancino V, Birren B. Transformation of *Escherichia coli* with large DNA molecules by electroporation[J]. **Nucleic Acids Research**, 1995, 23 (11) :1 990-1 996.
- [31] Soderlund C, Humphray S, Dunham A, *et al.* Contigs built with fingerprints, markers, and FPC V4. 7[J]. **Genome Research**, 2000, 10(11) :1 772-1 787.
- [32] Xu Z, Sun S, Covaleda L, *et al.* Genome physical mapping with large-insert bacterial clones by fingerprint analysis: methodologies, source clone genome coverage, and contig map quality[J]. **Genomics**, 2004, 84(6) :941-951.
- [33] Margulies M, Egholm M, Altman W E, *et al.* Genome sequencing in microfabricated high-density picolitre reactors [J]. **Nature**, 2005, 437:376-380.
- [34] Waldbieser G C, Bosworth B G, Nonneman D J, *et al.* A microsatellite-based genetic linkage map of the channel catfish, *Lctalurus punctantus*[J]. **Genetics**, 2001, 158: 727-734.
- [35] Liu Z, Karsi A, Li P, *et al.* An AFLP-based genetic linkage map of channel catfish (*Lctalurus punctantus*) constructed by using an interspecific hybrid resource family[J]. **Genetics**, 2003, 165 :687-694.
- [36] Sakamoto T, Danzmann R G, Gharbi K, *et al.* A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates [J]. **Genetics**, 2000, 155 :1 331-1 345.
- [37] Nichols K M, Young W P, Danzmann R G, *et al.* A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*) [J]. **Animal Genetics**, 2003, 34: 102-115.
- [38] Lee B Y, Lee W J, Streeleman J T, *et al.* A second-generation genetic linkage map of tilapia (*Oreochromis spp.*) [J]. **Genetics**, 2005, 170(1) :237-244.
- [39] Albertson R C, Streeleman J T, Kocher T D. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes [J]. **Proc Natl Acad Sci USA**, 2003, 100:5 252-5 257.
- [40] Hubert S, Hedgecook D. Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*[J]. **Genetics**, 2004, 168 :351-362.
- [41] Yu Z, Guo X. Genetic linkage map of the eastern oyster *Crassostrea virginica* Gmelin [J]. **Biol Bull**, 2003, 204 :327-338.
- [42] Li Y, Byrne K, Miggiano E, *et al.* Genetic mapping of the kuruma prawn *Penaeus japonicus* using AFLP markers [J]. **Aquaculture**, 2003, 219:143-156.
- [43] P é F, Erazo C, Zhinaula M, *et al.* A sex-specific linkage map of the white shrimp *Penaeus (Litopenaeus) vannamei* based on AFLP markers [J]. **Aquaculture**, 2005, 242: 105-118.
- [44] Quiniou S M A, Katagiri T, Clem W, *et al.* Construction and characterization of a BAC library from a gynogenetic channel catfish, *Ictalurus punctatus* [J]. **Genet Sel Evol**, 2003, 35 :673-684.
- [45] Phillips R. Physical and genetic mapping of the rainbow trout major histocompatibility regions: evidence for duplication of the class I region [J]. **Immunogenetics**, 2003, 55 :561-569.
- [46] Gahr S A, Palti Y, Rexroad C E . Genomic characterization of a novel pair of Id genes in the rainbow trout (*Oncorhynchus mykiss*) [J]. **Animal Genetics**, 2004, 35:317-320.
- [47] Katagiri T, Asakawa S, Minagawa S, *et al.* Construction and characterization of BAC libraries for three fish species: rainbow trout, carp and tilapia [J]. **Animal Genetics**, 2001, 32:200-204.
- [48] Watanabe M, Kobayashi N, Fujiyama A, *et al.* Construction of a BAC library for *Haplochromis chilotes*, a cichlid fish from Lake Victoria [J]. **Genes Genet Syst**, 2003, 78:103-105.
- [49] Siemon H S N, Artieri C G, Bosdet I E, *et al.* A physical map of the genome of Atlantic salmon, *Salmo salar* [J]. **Genomics**, 2005, 86(4) :396-404.
- [50] Katagiri T, Kidd C, Tomasino E, *et al.* BAC-based physical map of the Nile tilapia genome [J]. **BMC Genomics**, 2005, 6:89-94.

(本文编辑:刘珊珊)