

实验报告：增强绿色荧光蛋白基因的分子克隆

朱瑾煜（北京大学生命科学学院，学号：2000012180）

1 实验内容

1.1 增强绿色荧光蛋白（EGFP）基因的 PCR 扩增及纯化

使用 EasyPure Plasmid MiniPrep Kit 试剂盒，利用碱裂解法和吸附柱法从 2 种大肠杆菌（*Escherichia coli*）中提取 pEGFP-N3 质粒和 pET28a 质粒。然后，对 pEGFP-N3 质粒的 *EGFP* 基因片段进行 PCR 扩增，所用引物自带黏性末端，便于后续步骤中的表达载体构建。对上述 PCR 扩增产物、pEGFP-N3 质粒、pET28a 质粒进行琼脂糖凝胶电泳，在波长为 254 nm 的紫外灯下观察染色后的电泳凝胶。

1.2 *EGFP* 基因的表达载体构建

使用 *Bam* H I 和 *Xho* I 两种限制酶，对 pET28a 质粒进行快速双酶切。将酶切后的 pET28a 线性化载体和上述 PCR 产物进行电泳检测分离，根据电泳图谱估算酶切后的 pET28a 线性化载体和 *EGFP* 片段的量，然后分别切胶回收。使用 Gibson assembly 无缝连接装配载体和 *EGFP* 片段，并对 DH5 α 感受态细胞进行转化，在含卡那霉素的选择培养基上、37 °C 下过夜培养。

1.3 阳性克隆鉴定

取上述培养的菌落制成菌悬液，按 1.1 的方法提取质粒。

1.3.1 电泳检测鉴定

按 1.2 的方法对提取出的质粒进行双酶切，电泳检测双酶切片段。

1.3.2 测序鉴定及生物信息学处理

取阳性质粒送测序。收到测序结果后，利用 SnapGene 软件、BLAST 软件和 samtools 命令行工具进行序列分析，检查目的片段是否正确插入到表达载体上。所有序列数据和代码可在作者的 GitHub 仓库（<https://github.com/jyzhu-pointless/molecularLab/tree/main/molecularCloning>）中获取。

1.4 阳性克隆表达

取 1.3 中的阳性质粒，对 BL21(DE3)感受态细胞进行转化。取 2 个含卡那霉素的平板，其中一块平板加入 100 μL 20 mmol/L IPTG（诱导物），另一块平板不加诱导物，将转化细胞均分后接种到两块平板，37 $^{\circ}\text{C}$ 下过夜培养，之后置 4 $^{\circ}\text{C}$ 供拍照。

2 结果

2.1 pEGFP-N3 质粒、pET28a 质粒和 *EGFP* 片段的琼脂糖凝胶电泳结果

对 pEGFP-N3 质粒的 *EGFP* 基因片段的 PCR 扩增产物、pEGFP-N3 质粒和 pET28a 质粒进行琼脂糖凝胶电泳，在波长为 254 nm 的紫外灯下观察，结果如图 1 所示。

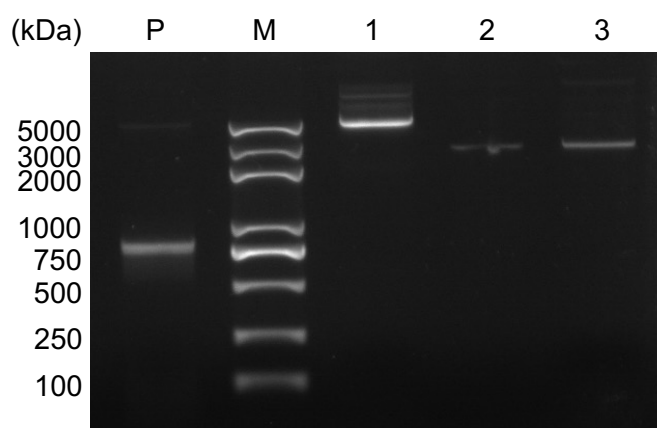


图 1. pEGFP-N3 质粒、pET28a 质粒和 *EGFP* 片段（PCR 产物）的琼脂糖凝胶电泳图谱。
P: *EGFP* 片段 PCR 产物；M: DNA marker ladder；1: pEGFP-N3 质粒；2、3: pET28a 质粒。

图 1 显示，扩增的 *EGFP* 片段的长度在 750 bp 左右，与参考 *EGFP* 序列长度 720 bp 相近。实验中所用的 DNA marker 的 DNA 浓度为 10 ng/ μL （750 bp 加亮带为 20 ng/ μL ），而 pET28a 的条带亮度接近 marker 亮度，估计上样体系中 pET28a 的浓度为 5~10 ng/ μL 。上样体系中，pET28a 取样量为 2 μL ，总体积为 10 μL ，稀释了 5 倍。故提取的 pET28a 浓度应为 25~50 ng/ μL 。

2.2 pET28a 线性化载体和 *EGFP* 片段的琼脂糖凝胶电泳结果

取 30 μL 双酶切后的 pET28a 线性化载体样品、28 μL *EGFP* 基因片段的 PCR 扩增产物上样，进行琼脂糖凝胶电泳，在波长为 254 nm 的紫外灯下观察，结果如图 2 所示。

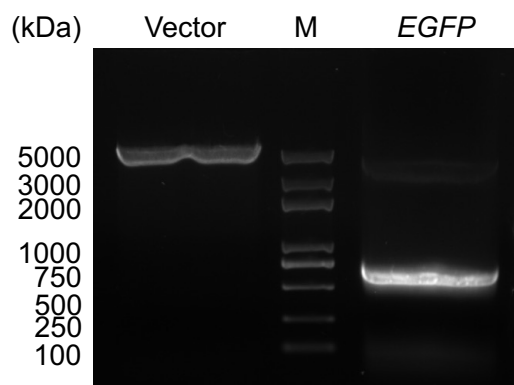


图 2. pET28a 线性化载体和 *EGFP* 片段的琼脂糖凝胶电泳图谱。Vector: 线性化载体; M: DNA marker ladder; *EGFP*: *EGFP* 基因片段。

图 2 显示，与 2.1 中实验相同；扩增的 *EGFP* 片段的长度在 750 bp 左右，与参考 *EGFP* 序列长度 720 bp 相近；线性化载体的长度约为 5000 bp。观察条带亮度，*EGFP* DNA 浓度明显高于线性化载体的 DNA 浓度（推测为 3 倍左右）。

2.3 琼脂糖凝胶电泳鉴定阳性克隆结果

对转化的菌落提取质粒，双酶切后进行琼脂糖凝胶电泳鉴定，在波长为 254 nm 的紫外灯下观察，结果如图 3 所示。

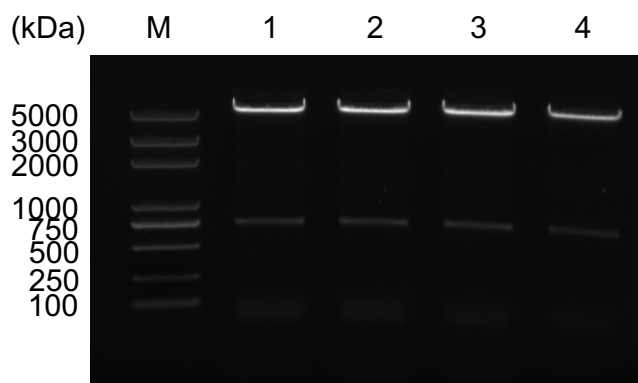


图 3. 阳性克隆的电泳鉴定结果图谱。M: DNA marker ladder; 1、2、3、4 泳道分别为所选取的 4 个转化菌落的电泳结果。

图中显示，对于 4 个转化菌落，在 5000 bp 左右和 750 bp 左右均各有一明显条带，分别对应酶切后的线性化载体和 *EGFP* 基因片段。

2.4 阳性克隆质粒测序结果

2.4.1 测序质量结果

用 SnapGene 软件打开测序数据文件 310CZ_T7-3G_Z20929219_B02.ab1，打印测序结果（seq_quality.pdf，见补充信息或 GitHub），导出 fastq（310CZ_T7-3G_Z20929219_B02.fastq，见 GitHub）和 tsv（310CZ_T7-3G_Z20929219_B02.fastq.tsv，见 GitHub）文件，可知有 916 个碱基的质量分数（quality score）都在 50 以上，占比 $916/1240=0.739$ 。由此可知，测序质量良好。

2.4.2 序列比对结果

用 SnapGene 软件打开测序序列文件 310CZ_T7-3G_Z20929219_B02.seq 和参考 *EGFP* 序列文件 EGFP_reference.fa，分别另存为 SnapGene DNA 文件 310CZ_T7-3G_Z20929219_B02.dna 和 EGFP_reference.dna。使用 Show alignment 功能，比对两个序列，如 45(310CZ_T7-3G_Z20929219_B02_Map.png) 所示。由图 4 可知，载体的第 140~859 个核苷酸的位置与 *EGFP* 序列匹配，而这两个位置刚好对应 *Bam* HI（134）和 *Xho* I（860）的酶切位点。

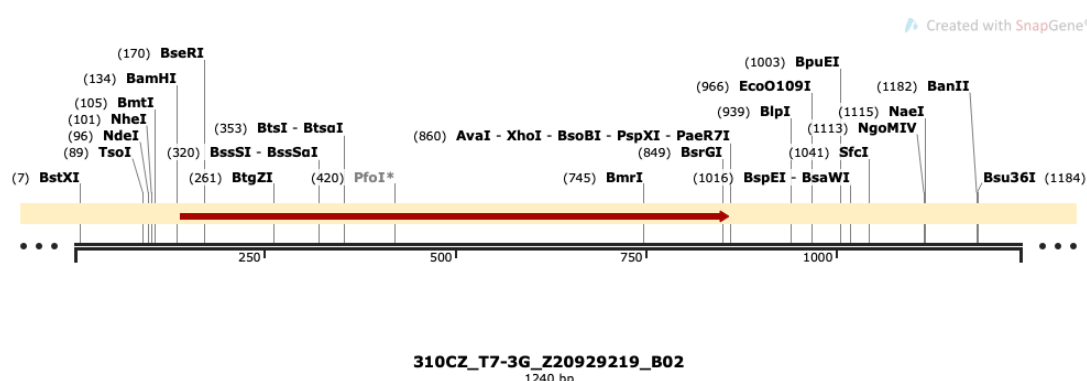


图 4. SnapGene 序列比对结果。黑色线条表示测序的载体序列，红色粗箭头表示参考 *EGFP* 基因序列。图中标出了测序序列上的各种常见限制酶的酶切位点。

用 NCBI BLAST 的 `blastn` 功能对测序序列和参考 *EGFP* 序列进行序列比对（比对结果见补充信息），并从 SnapGene 导出 alignment 的 `sam` 文件（310CZ_T7-3G_Z20929219_B02.sam），在 Linux 命令行中使用 `samtools` 软件对 `sam` 文件进行分析（见 GitHub），得到的匹配程度均为 100%，无错配。这说明扩增的 *EGFP* 基因不仅成功装配到了载体上，而且没有发生突变。

2.5 阳性克隆表达结果

阳性质粒转化 BL21(DE3) 细胞，分别培养在含 IPTG 诱导物和不含 IPTG 诱导物的平板上（方法详见 1.4），置于暗室内拍照观察绿色荧光强度，如图 5 所示。可以看出，含 IPTG 诱导物的平板中菌落的绿色荧光强度明显高于不含 IPTG 诱导物的平板中菌落的绿色荧光强度。

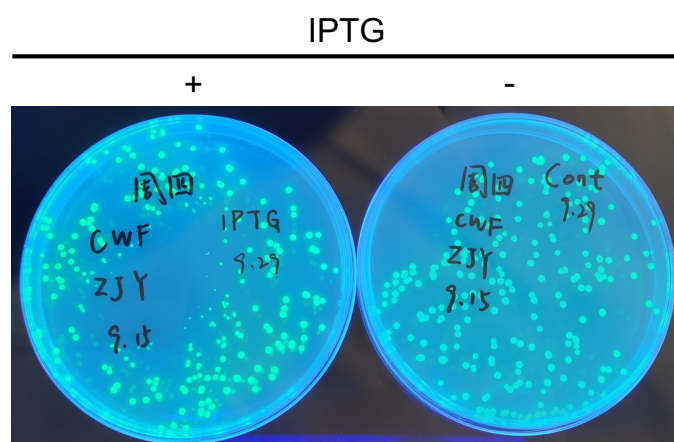


图 5. 阳性克隆表达菌株的荧光强度与诱导物 IPTG 是否存在的关系。+表示平板中含有 100 μ L 20 mmol/L IPTG，-表示平板中不含 IPTG。

3 讨论

本实验是对 *EGFP* 的分子克隆实验，具体而言是将 *EGFP* 基因扩增后，通过双酶切法构建表达载体，将表达载体转化到感受态细胞中，然后用酶切电泳和序列测定的方式检测表达载体是否被成功转化，用观测表型（荧光强度）的方式检测表达载体能否在细胞中正常表达。检测结果均完全符合预期，达到了实验目的。

在图 1 的凝胶电泳中，除 PCR 产物（*EGFP* 基因片段）外，两种质粒也分别上样，它们最明显的条带分别位于图中的 5000 bp 和 3000 bp 的位置。然而，直

接认为这两种质粒的长度就是 5000 bp 和 3000 bp 是不合适的，因为质粒主要以共价闭环 DNA (covalent closed circular DNA, cccDNA) 的形式存在，而所用的 DNA marker 是线性 DNA (linear DNA, lDNA)。在电泳中，对于同等长度的 DNA，超螺旋的共价闭环 DNA 的迁移率高于线性 DNA 的迁移率，后者又高于开环 DNA (open circular DNA, ocDNA) 的迁移率。由于多种理化因素的影响，一部分质粒会转化为线性或开环 DNA 的形式，从而在电泳中表现为一条明显的主带 (cccDNA) 和 1~2 条不明显的副带 (lDNA 或 ocDNA)。这在图 1 的 1 号泳道 (pEGFP-N3 质粒) 中很明显。

实验中使用的 PCR 引物自带粘性末端，这使得 PCR 产物可以直接与线性化载体进行 Gibson assembly，而无需再酶切 PCR 产物。图 1 和图 2 中可以看到，PCR 产物中有不明显的 5000 bp 左右的非特异性条带，推测为 PCR 反应中的模板 DNA 及其他非特异性扩增产物。由于使用的引物较特殊，在设计引物时应特别注意，尽量减少非特异性扩增的出现。

在图 3 的阳性克隆鉴定电泳中，5000 bp 左右的条带较 750 bp 左右的目标条带更明显，这可能是因为有的载体没有被插入 *EGFP* 基因片段 (即空载体)。为了避免出现大量细菌中了空载体的情况，可考虑使用正负双向选择系统，即使用带有两个抗性基因的载体，将目的基因插入其中一个抗性基因，在培养过程中，选择对被插入目的基因破坏的抗性基因所对应的抗生素无抗性、而对另一种抗生素无抗性的菌株。除了合理筛选以外，控制好酶切时间、使用双酶切防止空载体自身环化也是十分有必要的。

对测序质量和序列信息的定量分析表明，*EGFP* 基因被完整无误地插入了载体并得以转化。其中运用了与序列比对相关的大量生物信息学方法和工具，这表明分子生物学的实验和研究中是离不开生物信息学的。对阳性克隆的表型检测表明，*EGFP* 基因不但被插入载体得以转化，还能在宿主细胞中成功表达。需要指出的是，不含 IPTG 诱导物的平板中也显示出绿色荧光，这可能是因为 *EGFP* 基因除了诱导表达外，本身也有一定程度的基本表达。如果能加上未转化的菌株平板作为空白对照组，结论可能会更有说服力。

参考文献:

1. 分子生物学实验授课 PPT
2. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." *J. Mol. Biol.* 215:403-410.
3. SnapGene software (www.snapgene.com).

补充信息（Supplementary Information）

以下列出 BLAST 序列比对结果和测序质量分析图谱（seq_quality.pdf）。其他数据可前往 GitHub 或作业附件查看。

BLAST 序列比对结果

Score	Expect	Identities	Gaps	Strand
1330 bits(720)	0.0	720/720(100%)	0/720(0%)	Plus/Plus

```
Query 1  ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 140 ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGAC 199

Query 61  GGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 200  GGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAC 259

Query 121  GGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 260  GGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACC 319

Query 181  CTCGTGACCACCCTGACCTACGGCGTGCAAGTTCAGCCGCTACCCCGACCACATGAAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 320  CTCGTGACCACCCTGACCTACGGCGTGCAAGTTCAGCCGCTACCCCGACCACATGAAG 379

Query 241  CAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGACCATCTTC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 380  CAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGACCATCTTC 439

Query 301  TTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 440  TTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCTG 499

Query 361  GTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 500  GTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC 559

Query 421  AAGCTGGAGTACAACCTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAAC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 560  AAGCTGGAGTACAACCTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAAC 619

Query 481  GGCATCAAGGTGAACCTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCC 540
```


Sbjct	620	GGCATCAAGGTGAAC TTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCC	679
Query	541	GACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCACAACCAC	600
Sbjct	680	GACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCACAACCAC	739
Query	601	TACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCAACGAGAAGCGCGATCACATGGTC	660
Sbjct	740	TACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCAACGAGAAGCGCGATCACATGGTC	799
Query	661	CTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA	720
Sbjct	800	CTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA	859

测序质量分析图谱（见下一页）

