

# 효소적으로 증폭된 DNA의 염기배열결정법과 16S like 리보솜 유전자의 증폭 및 염기배열결정에의 응용

(Sequencing of Enzymatically Amplified DNA and Its Application to 16S Like Ribosomal Gene Amplification and Sequencing)

이 재 동<sup>1</sup> · 주 우 흥<sup>2</sup>

부산대학교 자연과학대학 미생물학과<sup>1</sup> · 장원대학교 자연과학대학 생물학과<sup>2</sup>

## I. 서 론

유전자조작의 역사는 1970년대부터 시작되어, 특정유전자만 검출하는 southern hybridization법<sup>1)</sup>과 염기배열결정법<sup>2)</sup>을 비롯하여 많은 새로운 기술이 개발되었다. 이후 분자생물학에서는 필수적인 기술로서 염기배열결정법의 간편화와 신속화가 요구되어 왔다. 이에 dideoxy법<sup>3)</sup>에 의한 염기배열결정, 새로운 효소의 이용,<sup>4,5)</sup> 자동형광표식 DNA sequencer의 등장은 이러한 요구에 부응하는 기술적인 진보였다. 그러나 cloning후에 clone의 동정단계에 있어서의 간편화, 신속화의 개량은 근본적으로 진전이 없었다. 제한효소-ligase법으로 대표되는 cloning에 의한 특정DNA의 증폭법은 특정DNA와 vector를 제한효소로 절단한후 DNA ligase로 연결, 숙주(대표적으로 대장균)에 삽입시켜 증식시키는 방법이다. Vector로 사용된 plasmid는 자기복제능을 가지고 있으므로 배양후 plasmid를 추출정제, 제한효소로 절단하면 특정DNA를 회수할 수 있으나 적어도 2~3주간이 소요된다.

근년에 개발된 효소적인 DNA증폭법(polymerase chain reaction; PCR)을 이용하면 일차구조상의 단편적인 정보(부분 아미노산배열, 아미노산배열의 보존성 등)만 알면 단 수시간내 해석에 필요한 양의 DNA가 증폭되어 cDNA의 염기배열결정의 신속화, 간편화가 가능하게 되었다. 이는 획기적인 변혁으로 유전자의 구조, 기능과 발현, 유전병 및 virus감염증의 진단 및 분자진화학적인 연구에도 응용이 가능하다. 또한 자동화 DNA sequencer를 병행하여 이용하면 상상할 수 없을 정도의 신속한 해석이 가능하다. 그러므로 본문에서는 유전자증폭(gene amplification)법으로써 PCR법에 관해 기술하고자 한다. 또한 리보솜RNA는 분자시계(molecular clock)로서 생물의 계통을 논하는 데에

있어서는 최적의 조건을 갖춘 고분자화합물이다. 최근 세균(bacteria)에서는 Carl Woese<sup>6)</sup> 등에 의해 16S 리보솜 RNA의 염기배열을 바탕으로 계통이 정리되어 왔고, 균류에서도 이에 상당하는 18S 리보솜RNA의 염기배열결정이 계통진화관계의 해명에 필수적인 과정으로 주목을 받고 있다.<sup>7,8)</sup> 이에 PCR법을 이용한 16S like 리보솜DNA의 증폭법을 다루고, PCR증폭산물의 염기배열결정법에 대하여 서술한다. 최근 인위적인 leading error 등을 배제하고 신속한 자동해독과 시간적인 절약이 자동 DNA sequencer의 개발과 시판으로 가능하게 되었다. 그러므로 cDNA의 형광색표식 염기배열결정법에 대해서도 서술하고자 한다.

## II. PCR법에 의한 cDNA증폭법

PCR법은 1985년 Cetus의 R. K. Saiki 등의 연구진에 의해 고안된 방법으로 제한효소 ligase법에 의한 cloning방법을 대체할 수 있는 방법으로 주목을 받았다.<sup>9)</sup> 당초 polymerase로서 사용된 대장균 polymerase I의 Klenow단편은 열에 약해 열변성과정에서 실활되어 매 cycle마다 새로운 효소를 첨가하여야 하는 번거로움으로 일반화되지 못했다. 1987년 King에 내열성의 polymerase가 *Thermus aquaticus*로부터 개발되고, 또한 반응을 자동화하는 것도 가능하게 되어 그 이용이 분자생물학의 전반적인 영역으로 확대되었다.<sup>10)</sup>

PCR법의 개략적인 과정은 DNA의 열변성, primer의 결합, primer의 신장반응이며, 이러한 과정의 반복으로 *in vitro*에서 DNA의 증폭이 가능하다(Fig. 1). PCR법에서 통상 DNA 열변성은 92°C에서 15~30초(최초의 열변성은 좀 길게 함), primer의 결합은 45~60°C (T<sub>m</sub>-5°C)에서 15~60초, primer extension반응은 72°C에서 30~60초로 30~40 cycle을 행한다. DNA단편의 최종적인 증폭시에는 exte-

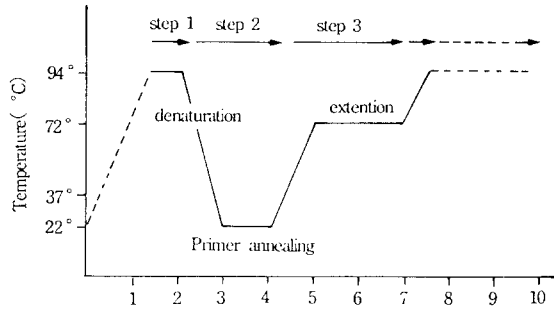


Fig. 1. Principles of polymerase chain reaction.

ension 시간을 연장하며, 증폭도를 최고로 하기 위해서는 cycle 시간을 되도록 짧게 하고 cycle의 수를 늘리는 것이 좋다.

최근 *Taq* polymerase 외에 새로운 내열성효소(*Vent*<sup>TM</sup> DNA polymerase, *Bst* DNA polymerase, *Tth* DNA polymerase)의 개발과 염기배열결정의 자동화(DNA sequencer, 형광 bio image analyzer)는 그 응용의 폭을 더욱 넓게 하고 있다. 일반적으로 2Kb 이하의 증폭은 단숨에 가능하나 2 Kb 이상의 유전자의 증폭은 중앙부분, 3'측영역, 5'측영역의

cDNA를 증폭, 염기배열결정을 하여 3개영역의 염기배열을 overlapping시켜 전체의 배열을 결정한다.

다음에는 미지의 cDNA 증폭법을 O. Ohara 등<sup>17)</sup>이 보고한 one-side polymerase chain reaction을 중심으로 소개하며, 그 과정을 Fig. 2에서 모식적으로 보여주고 있다.

1. 중앙부분의 증폭법

아미노산의 보존영역 또는 유전자산물의 부분 아미노산 배열에 대한 지식을 바탕으로 primer를 작제한다. 이때 primer는 다소 mismatch가 허용되나 3'말단의 잔기는 정확히 작제하여야 한다. 아미노산배열에서 예상되는 전염기배열을 가진 높은 degenerate primer보다는 다소의 mismatch가 있어도 degeneracy를 줄이는 편이 특이성이 높은 cDNA 증폭에 유익하다.

이 중앙부분의 증폭에 poly A RNA를 이용하는 것이 좋으나 염색체DNA의 혼입을 가능한한 적게 하면 전세포질 RNA를 바탕으로해도 문제가 없다. DNA의 혼합방지책으로는 고염농도하의 RNA의 선택적침전,<sup>18)</sup> 산성 phenol에

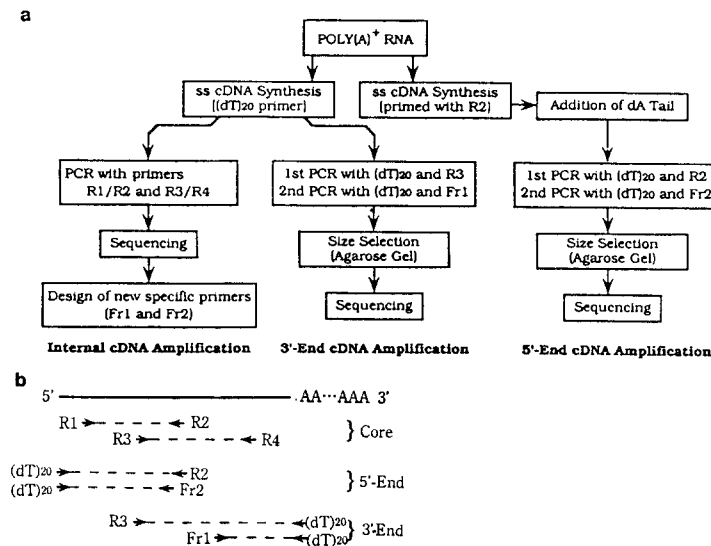


Fig. 2. Three-part strategy for cDNA amplification. (a) Steps involved in the amplification and eventual sequencing of complete cDNA transcripts. ss, Single-stranded; TM, tropomyosin. (b) Position on an idealized tropomyosin cDNA sequence of the various oligomers discussed in the text.

의한 추출<sup>19)</sup>을 들 수 있다. 전세포질 RNA를 주형으로 cDNA의 작제시는 역전사효소(reverse transcriptase ; murine moloney leukemia virus, avian myeloblast virus, Rous associated virus 2)를 사용하면 높은 수율을 얻을 수 있다. 역전사반응에 사용되는 RNA량은 보통 poly A RNA 최소 100 µg정도, 전세포질 RNA이면 최소 5~10µg을 사용한다. 역전사반응액은 멸균수에 희석비등하면 PCR에 직접 사용할 수 있으나 고염농도의 NaCl, KCl은 *Taq* polymerase를 저해하므로 최종 PCR반응액의 염농도는 0.1M이하로 유지하여야 한다.

### 2. 3'측영역의 증폭법

3'측영역의 증폭시는 대부분의 진핵생물의 mRNA가 poly A tail을 3'말단에 가지고 있으므로 이 부분에 대한 dT primer (20~30 mer)와 중앙부분의 배열정보를 이용한 특이적인 primer를 작제하여 PCR을 행한다. 이때에는 목적의 cDNA에 가끔 특이적인 primer가 결합하여 증폭된다. 그러므로 통상 PCR을 1회 더 행하며, 1회째 PCR에 사용한 primer의 결합부위의 하류의 배열을 가진 새로운 PCR용의 특이적인 primer를 사용한다. PCR반응후 특정의 목적 DNA의 증폭을 확인하기 위해 hybridization을 적당한 probe로 행한다. 특이성을 향상시키기 위해 gel에서 PCR 산물을 분리시 목적DNA를 hybridization으로 동정하여 gel로부터 회수한다.

dT primer의 hybrid형성이 불안정한 경우에는 dT primer의 5'말단에 20잔기정도의 적당한 배열을 결합시켜 anchor PCR<sup>20, 21)</sup>을 사용한다. 이 dT primer의 5'말단측의 배열로 RNA polymerase의 promoter배열 등을 이용하면, 증폭 DNA에서 RNA probe를 직접 작제가능하다. 또한 poly A부분이 상당히 길기때문에 dT primer의 결합이 일정하지 않을 수 있고, 따라서 PCR 증폭산물도 크기의 불균일성을 나타낸다. 그러나 PCR의 cycle수를 증가시키면 이를 배제할 수 있다. PCR산물의 직접 염기배열결정시 이러한 불균일성 때문에 poly A부분에서가 아니라 중앙부분에서 3'말단으로 염기배열을 결정한다.

### 3. 5'측영역의 증폭법

5'말단은 linker(adapter)의 부가, 또는 terminal deoxy-

nucleotide transferase(TdT)에 의한 homonucleotide의 부가가 cDNA의 증폭에 선행되어야 한다. 보통 dA부가가 자주 사용된다. 그러면 3'측영역의 증폭에 사용한 primer를 사용할 수 있고, back ground가 낮기 때문이다. 우선 중앙부분의 DNA에서 밝혀진 배열을 기초로하여 되도록 5'상류에 가까운 부분을 결합부위로 하는 oligomer를 준비한다. 이 primer를 이용하여 5'말단영역의 cDNA를 합성한다. 이때 primer량을 줄이고 mRNA와 primer의 hybridization시간은 길게하는 것이 좋다. cDNA합성후 phenol추출, ethanol침전에 의해 단백질 및 다량의 deoxynucleotide를 제거하며, 70% ethanol침전, 건조후 TdT반응을 행한다. 만일 sample량이 적을 경우 carrier로서 poly C를 ethanol침전시 첨가한다. TdT반응후 EDTA첨가, phenol추출, ethanol침전을 행해 cDNA를 회수하여 PCR에 사용한다. 종래에는 cloning의 경우에 5'말단부분을 해석할 수 없는 경우가 있었으나, 5'말단영역 cDNA의 증폭법은 전사개시부위의 결정에 이용된 방법과 PCR법을 병행함으로써 완전한 5'말단영역의 해석이 가능하다.

## III. 16S like 리보솜DNA의 증폭

생물의 진화는 유전자상의 중복(duplication)과 염기치환 등으로 가능하며, 그 진화의 역사는 유전자의 염기배열에 기록되어 있다. 표현형질수준에서의 단순한 계통관계의 추정과 종분화에 대한 언급은 유전자상에 존재하는 자료의 뒷받침이 없으면 주관적으로 흘러 그 성립자체가 설득력이 없게 된다. 최근 분자진화학적인 연구의 진전으로 전체genome의 해석을 바탕으로 하지 않고도 객관적인 계통진화관계에 대한 해명이 가능하게 되었다. 형태발생에 관한 유전자군의 기능, 조절 및 발현에 대한 분자생물학적인 연구가 DNA분리법의 난이성으로 진전되고 있지 않은 군류에서는 특히 분자진화적인 연구에 기초를 둔 계통진화의 관계해명이 주목을 받고 있다.

분자진화학적연구에 이용되는 고분자는 단백질, 핵산 등으로 그 정보량이 너무 크지도 너무 적지도 않으며, 통계적으로 유의성이 있는 분자이어야 할 뿐만아니라 생물의 다양성으로 주어진 시간에 많은 대상을 연구하여야 하므로 보다 간편한 연구방법의 도입이 요구된다. 그러므로 생물의 모든 종에 존재하면서 염기치환율이 높지 않은 16S like

리보솜RNA는 생물의 계통진화관계의 해명에 주요한 대상이 되고 있다.<sup>12)</sup>

16S like 리보솜RNA에 바탕을 둔 연구는 Carl Woese가 개발한 RNase T<sub>1</sub>에 의한 분해물의 oligonucleotide catalog법에서 비롯되어, Lane 등<sup>23)</sup>에 의한 부분 염기배열결정법에 의해 진행되어 왔고, 전염기배열은 리보솜DNA를 cloning하여 전결정을 하여왔다. 특히 리보솜RNA를 대상으로 한 연구에는 다량의 리보솜RNA가 필요하고, 시료 리보솜RNA가 환경에 존재하는 RNase에 의해 분해되어 버리는 경우가 많고, 염기배열결정시 부정확한 곳이 생기는 등 실험상 많은 어려움이 있었고, cloning에는 앞에서 언급한 바와 같이 clone의 확인에 많은 시간이 요구되어 data의 축적이 지연되어 왔다.

16S like 리보솜RNA의 5'측영역과 3'측영역에는 각종의 생물군에 공통적으로 보존성이 높은 영역이 존재하여 이 영역을 기초로 한 PCR용 primer를 작제할 수가 있다.<sup>24, 25)</sup> 이들 primer를 이용하면 16S like 리보솜DNA의 염기를 PCR법으로 증폭가능하다. 이 때 16S like 리보솜RNA의 5'영역과 같은 염기배열의 forward primer와 3'영역에 상보적인 염기배열의 reverse primer를 작제, PCR용 primer로 사용한다(Fig. 3).

Eubacteria 16S 리보솜DNA의 PCR에 사용되는 primer는 Table 1에 정리되어 있고, 이를 *E. coli*와 eubacteria에 공통인 16S 리보솜RNA 염기배열과 alignment한 것이 Fig. 4에 표시되어 있다. 이들 PCR primer의 조합으로 증폭된 그리고 이론적으로 증폭가능한 세균군 또는 균종은 Table 2와 Table 3에 나타내었다.

rD<sub>1</sub>과 rD<sub>2</sub> primer쌍은 다양한 eubacteria군의 증폭에 이용될 수 있다. rD<sub>1</sub> 대신 rP<sub>1</sub>으로 대체하면 더 많은 다양한 종의 증폭에 이용될 수 있으나, rD<sub>1</sub>이 3'측영역에 가까우므로 보다 많은 수의 염기증폭에는 적합하다. rP<sub>2</sub> 대신 rP<sub>1</sub>으로 대체시 길이가 적은 PCR산물이 생성되기도 한다.<sup>25)</sup>

균류에서도 forward primer로 5'-ATCTGGTTGATCCT-GCCAGT-3'의 배열을 가진 것과 reverse primer로서 5'-GATCCTTCCGACAGGTTACC-3'의 배열을 가진 것을 사용하면 16S like 리보솜DNA의 증폭이 가능하다(Fig. 4).

#### IV. PCR산물의 염기배열결정법

PCR산물의 염기배열결정은 subcloning을 하여 결정하는

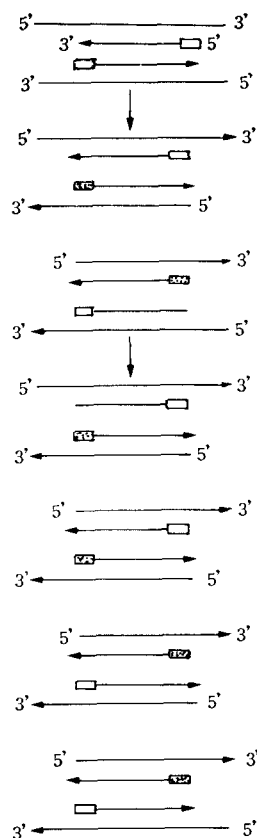


Fig. 3. Schematic illustration of 16S like ribosomal DNA amplification.

- Forward primer with a sequence corresponding to 5' region of the 16S like rDNA.
- ▨ Reverse primer complementary to 3' region of the 16S like rDNA.

방법과 직접 결정하는 방법이 있다. Subcloning시는 PCR용 primer의 5'측에 미리 polylinker를 붙여서 작제, PCR증폭하여 subcloning하거나 각각의 PCR용 primer로 blunt end인 PCR산물을 증폭하는 경우가 있다. Eubacteria 16S 리보솜DNA의 증폭시에 사용하는 primer의 polylinker는 Table 1에 표시되어 있다.<sup>26)</sup> 균류에서는 forward primer에 CCGAATTCGTCGACA(*Eco* RI와 *Sal* I site가 존재), reverse primer에 CCCGGATCCAAGCTT(*Sma* I, *Bam* HI, *Hind* III와 *Pst* I site가 존재)를 연결하여 사용할 수 있다.<sup>24)</sup> PCR산물과 vector를 제한효소로 절단하면, PCR산물중 pol-

Table 1. Summary of primers for the PCR amplification 16S of eubacterial rDNA<sup>a</sup>

Primer	Sequence (5' to 3')	Designed for
fD1	ccgaattcgtcgacaacAGAGTTTGATCCTGGCTCAG	Most eubacteria
fD2	ccgaattcgtcgacaacAGAGTTTGATCATGGCTCAG	Enterics and relatives
fD3	ccgaattcgtcgacaacAGAGTTTGATCCTGGCTTAG	<i>Borrelia</i> spirochetes
fD4	ccgaattcgtcgacaacAGAATTTGATCTTGGTTCAG	Chlamydiae
rD1	cccgggatccaagcttAAGGAGGTGATCCAGCC	Many eubacteria
rP1	cccgggatccaagcttACGGTTACCTTGTTACGACTT	Enterics (and most eubacteria)
rP2	cccgggatccaagcttACGGTACCTTGTTACGACTT	Most eubacteria
rP3	cccgggatccaagcttACGGATACCTTGTTACGACTT	Fusobacteria (and most eubacteria)

<sup>a</sup>Primer abbreviations : f, forward ; r, reverse ; D, distal ; P, proximal. All primer sequences are presented in 5' to 3' orientation. Linker sequences containing restriction sites for cloning are designated in lowercase letters. The "f" series of linkers all contain *EcoRI* and *SalI* sites, and the "r" series all contain *HindIII*, *BamHI*, and *XmaI* recognition sequences. Reverse primers produces sequences complimentary to the rRNA. Primers rP1, rP2, and rP3 are identical except for the 17th base from the 3' end. Under most amplification conditions, they should be functionally equivalent. Primer rP2 has the sequence corresponding to the greatest diversity of bacteria.

Table 2. Primer combinations that have been proven to produce an approximately 1,500-bp fragment

Species	Primer pair
<i>Neisseria gonorrhoeae</i>	fD1 + rD1
<i>Coxiella burnetii</i>	fD1 + rD1
<i>Anaplasma marginale</i>	fD1 + rD1
<i>Neisseria meningitidis</i>	fD1 + rD1
<i>Bacteroides fragilis</i>	fD1 + rP2
<i>Borrelia burgdorferi</i>	fD3 + rD1
<i>Borrelia hermsii</i>	fD3 + rD1
<i>Clostridium perfringens</i>	fD1 + rD1
<i>Mycoplasma pneumoniae</i>	fD1 + rP1
<i>Mycoplasma hominis</i>	fD1 + rP1
<i>Mycoplasma genitalium</i>	fD1 + rP1
<i>Ureaplasma urealyticum</i>	fD1 + rP1
<i>Campylobacter jejuni</i>	fD1 + rP1
<i>Shigella flexneri</i>	fD2 + rP1
<i>Shigella sonnei</i>	fD2 + rP1
<i>Chlamydia psittaci</i>	fD4 + rD1
<i>Chlamydia trachomatis</i>	fD4 + rD1
<i>Chlamydia pneumoniae</i>	fD4 + rD1
<i>Mycobacterium bovis</i>	fD1 + rD1
<i>Legionella pneumophila</i>	fD1 + rD1

Table 3. Theoretical specificity of amplification primers for 16S rDNA

Primer	Phylogenetic grouping and genera which should amplify with indicated primer <sup>a</sup>
fD1	Gram-positive bacteria and relatives <i>Bacillus, Clostridium, Staphylococcus, Listeria, Lactobacillus, Streptococcus, Mycoplasma, Spiroplasma, Ureaplasma, Acholeplasma, Erysipelothrix, Fusobacterium, Arthrobacter, Mycobacterium, Streptomyces</i> Purple bacteria and relatives (proteobacteria) <i>Rochalimaea, Brucella, Rhodopseudomonas, Agrobacterium, Rhodospirillum, Pseudomonas, Neisseria, Caulobacter, Myxococcus, Campylobacter, Rickettsia, Ehrlichia</i> Cyanobacteria <i>Anacysits (Synechococcus)</i> Bacteroides/flavobacteria <i>Bacteroides, Flavobacterium</i> Deinococcus and relatives <i>Deinococcus, Thermus</i> Spirochetes <i>Treponema, Spirochaeta</i> Planctomyces and relatives <i>Planctomyces</i> Chlorobium-green sulfur bacteria <i>Chlorobium</i> <i>Thermotoga</i> <i>Thermotoga</i>
fD2	Enteric members of gamma subdivision of proteobacteria <i>Escherichia, Shigella, Salmonella, Serratia, Erwinia, and Citrobacter, etc.(all the enterics) ; Oceanopirillum, Haemophilus, Acinobacillus, Vibrio, Pasteurella</i>
fD3	Spirochetes of the genus <i>Borrelia</i>
fD4	Genus <i>Chlamydia</i>
rD1	Purple bacteria and relatives (proteobacteria) <i>Pseudomonas, Neisseria, Rochalimaea, Agrobacterium, Myxococcus, Desulfovibrio</i> Gram-positive bacteria and relatives <i>Bacillus, Staphylococcus, Arthrobacter, Streptomyces, Mycobacterium, Heliobacterium</i> Cyanobacteria <i>Anacysits (Synechococcus)</i> Spirochetes <i>Treponema, Leptospira</i> Planctomyces <i>Planctomyces</i> <i>Chlorobium</i> <i>Chlorobium</i> <i>Thermotoga</i> <i>Thermotoga</i> (plus selected archaebacteria)
rP1, rP2, or rP3 (probably all functionally equivalent	Should prime all bacteria, plus plant mitochondria, chloroplasts, archaebacteria, and <i>Dictyostelium</i> , but not yeasts or vertebrates

<sup>a</sup>Primers are considered applicable if there is a perfect match for approximately 15 bases at the 3' end of the primer. The list is definitive only in the sense that the taxa mentioned represent the sequences available to the authors. The absence of a genus from the list does not imply that the primer will not work. Because the majority of the available rRNA sequences are derived from direct sequencing of rRNAs with reverse transcriptase, there is far less information available about the 3' end of the 16S. In some cases, the indicated genus is represented by numerous species; in other cases the indicated genus is represented by only one.

```

CONS=90%      AGAGUUUGAUC UGGCUCAG  GAACCGUGGCGG - CC U A ACAUGCAAGUCG CG
E. coli       AAAUUUAGAGUUUGAUCUAGGUCUCAGAUUAGAACCGUGGCGGCA-GGCCUAACACAUAGCAAGCGUAACAGGAAGAAGCUUGC

fD1          5'ccqaattcgtcgacaacAGAGTTTCATCCTGGCTCAG-3' (extend) ----)
fD2          5'ccqaattcgtcgacaacAGAGTTTCATCCTGGCTCAG-3' (extend) ----)
fD3          5'ccqaattcgtcgacaacAGAGTTTCATCCTGGCTTAG-3' (extend) ----)
fD4          5'ccqaattcgtcgacaacAGAATTTCATCCTGGTTTCAG-3' (extend) ----)

CONS=90%      G GGC  ACGGGUG GUAA  U  A  U  CC      GG A A C      GAAA      UAAUACC  AU
E. coli       UUCUUUGCUGACCGAGUGGCGGACGGGUGAGUAAUGUCUGGGAAACUGCTUGAUGGAGGGGUAACUACUGGAAACGGUAAUCGUAUACGCCAUAAACGCCAUAAACGUCC

CONS=90%      AAAG      GA      AU AG U GUUGG  GGUAA GGC  ACCAAG C A CA  U
E. coli       CAAGACCAAAGAGGGGACCUCGGGGCCUCUUGCCAUCCG AUGUGCCAGAUUGGAAUAGCUAGUAGGUGGGUAACUCCUCACCUUAGGGCAGCAUCCUCU

CONS=90%      A C G  CUCAGAGG  GA C G  CACA  UGG ACUCAGACACCG  CEA ACUCCUACCGGAGCCAGCAGU  GGAU U  CAAUGG  G AA  CU
E. coli       AGCUGGUCUGAGAGG AUGACCCACACUGGAACUGAGACACGGUCCAGACTUCUACGGGAGGCAGCAGUGGGAAU AUGGCACAAUUGGCCGCAAGCCU

CONS=90%      GA  AGC A GCGCCUG  GA GA G  U CG  GUAAA  CU U      GA G      UGAC UA
E. coli       GAUGCAGCCAGCCCGGUAUGAAGAAGCCUUCGGUUGUAAAGUAUUUACCGGGGAGGAGGGAGUAAGCUUAAUACCUUUGCUCUUGCAGUUU

CONS=90%      A  AACC  CCGCUAACU  CUGCCAGCAGCCCGGUAAUAC  AGG  GC  AGCGUU  CCGA U A  UGGCCGUAAG  G  G  AGG  G
E. coli       CCCGCAAGAAGACCCCGCUAACUCCGUGCCAGCAGCCCGGUAAUACGGAGCGUGCAAGCGUAAUCCGAAUACUUGCCGUAAGCCGACCCAGCGG

CONS=90%      G      AGU  G  GU AAA      GCU AAC      A AC      CU GA      AG GG      GAAUU  GUGUA
E. coli       GUUUGUUAAGUCAGAUUGAUAUCCCGGCUAACCUUGGAAACUGCAUCUGAUACUGGCAAGCUUGAGUCUCUGAGGCGGGGAGAAUCCAGGUGUA

CONS=90%      G  GGUGAAAU  CCUAGA AU  A  GAA ACC  U  GCGAAGCC      CUGG      A  UCAC  CU  A  G  CGAAAGCGUGGGGAGC  AACAGGA
E. coli       CGGUGAAAUAGCUAGAGAUUCUGGAGAAUACCGUGGCGAAGGGGGCCCGCUGGACGAAGACUGACGCUACGUGCGAAAGCGUGGGGAGCAACAGGA

CONS=90%      UURGAUACCCUGGAGUCCACGC  UAAACGAUG  U  GU  G      G  AG  AACGC  UAA  CCGCCUGG
E. coli       UUAGAUAACCCUGGAGUCCACGCAGCCGUAACGAUGUCGACUUGGAGGUGUGCCUUGAGGCGUGGCUUCCGGAGCUAACCGGUUAGUCCAGCCCGUGG

CONS=90%      AGUACG  CGCAAG  U  AAACUAAA  GAAUUGACGGGG  CCCGCACAAGCGG  GGAG  AUGUGGUUUAAUUGCA  G  ACGCG  GAACCUUACC
E. coli       GAGUACGGCCGCAAGGUAAAACUCAAUGAAUUGACGGGGCCCGCACAAAGCGGUGGAGCAUGUGGUUUAAUUGCAUGCAACCGCAAGAACCUUACCU

CONS=90%      UUGACAU      -GA A      -      -      ACAGGUG  UGCAUGG  UGUCGUCAGCUCGUG  CCUGAG  U
E. coli       GUCUUGACAUCCACCGAA-GUUUUA-GAGAUAGAAU-GUGCCUUCG-GGAACCGAGACAGGUGCUGCAUGGCUUGCUGCAGCUCGUGUUGUUGAAU

CONS=90%      GUUGGGUUAAGUCCCGCAACGAGCGCAACCC      GUU C A C      G G ACUC      ACUGCC  G  AA  GGAGGAAG
E. coli       GUUGGGUUAAGUCCCGCAACGAGCGCAACCCUUAUCCUUGUUGGCAACCGGUC-CGGCCGGAAACUCAAAGGAGACUGCCAGUGAAACUUGGAGGAAG

CONS=90%      G  GGA  GACGCUA  UC  UCAUG  CCCUUA  G  GGGCUACACAGU  CUACAUGG  ACA  G  G  GC  A  G  GA  AGC  AA  C
E. coli       UGGGGAUGACCGUAAGUCAUAUGGCCUUAACGACAGGGCUACACAGGUCUACAUUGGCGCAUACAAGAGAAGCGACTUCGCGAGAGCAAGCGGAC

CONS=90%      AAA  UC  AGU  CCGAU  G  CUGCAACUCG  C  UGAAG  GGA  U  GCUAGUAAUCG  AUCAG  A  CCGGUAUACGUUC
E. coli       UCAUAAAGUGGCGUAGUCCCGAUUGGAGUCUGCAACUCGACUCCAUGAAGUCGGAUCCGUAUUCGUGGAUCAGAAUGCCACGGUAAUACGUUC

CONS=90%      CCGG  CUUGUACACACCCCGCUA  CA  G  AG      AAG      AACC  GGA      C  A  GU  G
E. coli       CCGGGCCUUGUACACACCCCGCUACACCAUGGGAGUGGGUUGCAAAGAAGUAGCUUAAACCUCCGGAGGGCCUUAACCACUUGUGAUUCAUG

CONS=90%      A  UGGG  AAGUCGUAACAAGGUA  CC  UA  GAA  UG  GG  UGGAU  ACCUCCUUU
E. coli       ACUGGGUGAAGUCGUAACAAGGUAACCGUAGGGAAACCUUGCGGUUGGAUACCUCCUUA

fD1          (<---- (extend) 3'-CCGACCTAGTGGAGGAAAttcgaacctagggccc5'
fP1          (<---- 3'-TTCAGCATTCCTCCATTCGGCAAttcgaacctagggccc-5'
fP2          (<---- 3'-TTCAGCATTCCTCCATTCGGCAAttcgaacctagggccc-5'
fP3          (<---- 3'-TTCAGCATTCCTCCATAGGCAAttcgaacctagggccc-5'
    
```

Fig. 4. Sequence alignment of the amplification primers with 16S rRNAs of *E. coli* and a eubacterial consensus sequence. The consensus shows positions that greater than 90% conserved for a phylogenetically diverse collection of approximately 85 bacterial sequence.

ylinker내와 vector는 같은 sticky end가 생기며, 이를 ligation하여 subcloning한다.

Polylinker가 없는 primer를 사용하면 blunt end를 가진 PCR산물이 생성되므로 증폭시에는 이를 T<sub>4</sub> DNA ligase로

vector와 연결한다. 이때 제한효소로 vector를 절단하여 sticky end가 생기며 이를 DNA polymerase 또는 S<sub>1</sub> nuclease처리로 sticky end를 blunt end로 바꾸어 T<sub>4</sub> DNA ligase로 결합시킬 수 있다. 또는 합성 linker를 PCR산물에





결합시켜 새로운 말단(sticky end)을 만들어 같은 sticky end의 vector에 subcloning하기도 한다.

직접 염기배열결정법의 큰 이점은 시간을 단축할 수 있다는 점이다. 또한 잘못된 염기가 들어갈 확률이 Klenow fragment를 사용한 경우  $8 \times 10^{-6}$ /nucleotide/PCR cycle인데 비해, *Taq* polymerase는  $2 \times 10^{-4}$ /nucleotide/PCR cycle (*Taq* polymerase는 proof-reading기능을 가지지 않음)로 잘못 중합된 염기가 존재할 수 있는 가능성 때문에 subcloning시는 복수의 clone에 대해 염기배열을 확인하여야 하나, 직접 PCR산물의 염기배열을 결정하면 이러한 artifact는 검출되지 않는다. 직접 염기배열결정시 세균의 16S 리보솜DNA의 경우에는 Lane 등<sup>23)</sup>의 universal primer A, B, reverse B, C가 sequencing primer로 사용될 수 있다. 이때도 보존성이 높은 영역에서 sequencing primer를 작제한다. 현재 세균에서는 forward sequencing primer로 339~357,

785~805, 907~926, 1391~1406과 같은 위치(*E. coli* 리보솜RNA의 염기위치)가 사용되고, reverse sequencing primer는 357~342, 536~519, 802~785, 926~907, 1115~1100, 1406~1392, 그리고 1513~1494 위치에 상보적인 염기로 된 primer가 사용된다.<sup>25)</sup> 균류에서 사용될 수 있는 sequencing primer는 역시 보존성이 높은 영역과 같은 염기배열을 가진 forward primer와 상보성을 가진 reverse primer를 사용할 수 있는데 이는 Table 4에 나타내어져 있다.<sup>14)</sup>

염기배열결정은 Sanger의 ddNTP를 chain terminator로 사용한 dideoxy법<sup>3)</sup>이 이용된다. 이때 보통 방사성동위원소 ( $P^{32}$ ,  $S^{35}$ )로 표식된 deoxy nucleotide( $^{32}P$ -dATP,  $^{35}S$ -dATP)를 반응액에 넣어 합성된 DNA가 표식되게 한다 (Fig. 6, A). 방사성동위원소를 사용하여 염기배열결정을 수동으로 행하면, 피폭의 가능성도 있고, 폐기물의 보존,

Table 4. List of oligonucleotides, complementary to conserved 18S rDNA regions, which are used as primers for sequence determination of the 18S rRNA gene of *Candida albicans*

Number,	Sequence(a)	Strand (b)	Corresponding position in the 18S rDNA sequence of <i>C. albicans</i>
1	CTGGTTGATYCTGCCAGT	f	4- 21
2	GAAACTGCGAATGGCTCATT	f	83- 102
3	AATGAGCCATTCGCATTTTC	r	102- 83
4	AGGGYTCGAYYCCGGAGA	f	368- 385
5	TCTCCGRRTCGARCCCT	r	385- 368
6	ATTACCGCGGCTGCTGGC	r	580- 563
7	CGCGGTAATTCAGCTCCA	f	572- 590
8	TTGGYRAATGCTTTTCGC	r	398- 922
9	TTRATCAAGAACGAAAAGT	f	950- 967
10	CCGTCAATTYTTTRA	r	1136- 1117
11	AATTTGACTCAACACGGG	f	1169- 1186
12	ATAACAGGTCTGTGAGCCC	f	1409- 1428
13	GGGCATCACAGACCTGTTAT	r	1428- 1409
14	TTTGYACACACCGCCCGTTCG	f	1614- 1633
15	GACGGGCGGTGTGTRC	r	1632- 1617
16	CYGCAGGTTACCTACRG	r	1777- 1760

(a) The simultaneous presence of two bases at a single position is indicated by R(=A+G) or by Y(=C+T).

(b) Oligonucleotides with a sequence corresponding to the RNA-like strand are indicated by "f", oligonucleotides complementary to the RNA-like strand are indicated by "r".

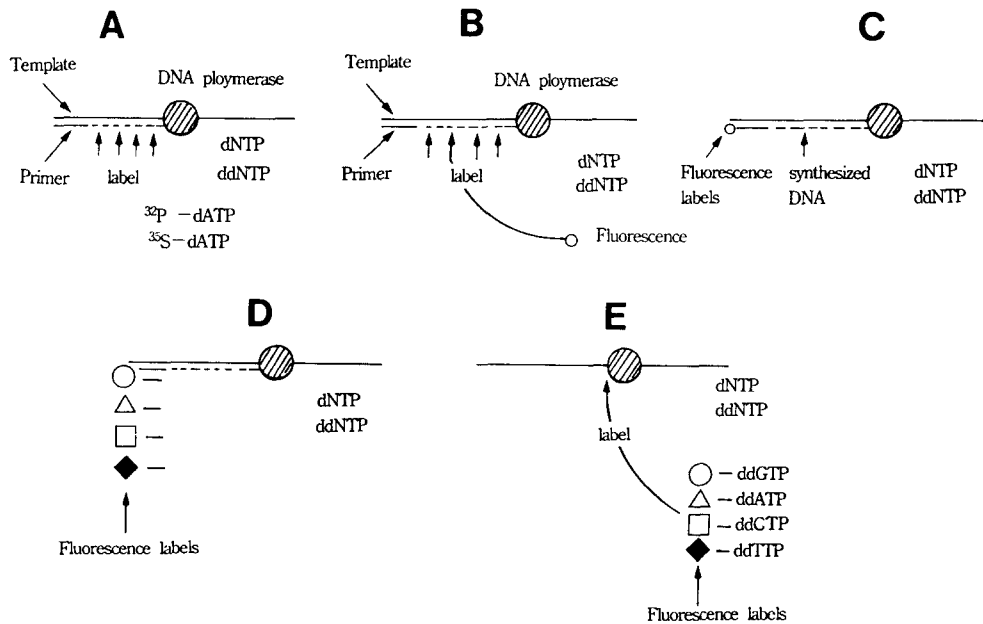


Fig. 6. Schematic diagram for DNA labeling.

처리과정에서 환경오염의 문제가 심각하게 된다. 최근 몇 개의 회사에서 자동화 DNA sequencer가 개발, 실용화되어 점차 그 사용빈도를 늘고 있다. 이들 자동화 DNA sequencer는 형광색소를 사용하는데 그 표식의 안정성이 높고 (6개월 이상), loading 후 8~10시간내에 결과를 얻을 수 있고, 직접 computer에 결과가 입력되므로 data처리가 신속하게 행하여 질 수 있다는 이점이 있다.

시판되고 있는 자동화 DNA sequencer는 형광발색단을 부가하는 방식에 따라 대별되는데, primer에 부가하는 방식 (Fig. 6C, D)과 dideoxynucleotide에 부가하여 표식하는 방식 (Fig. 6, E)이 있다. Primer에 부가하는 방식은 Hitachi와 Applied Biosystem의 DNA sequencer가 채용하고 있다. Hitachi HQ-3000은 한 primer를 형광색소로 표식하여 각각의 염기에 대하여 dideoxy반응을 행하여 각각 별개의 lane에서 전기영동을 행한다 (Fig. 6, C). 미국 Applied Biosystems DNA Sequencer는 4종의 다른 형광발색단을 primer에 dideoxy법으로 염기배열을 결정한다 (Fig. 6, D). 각각의 형광발색단을 ddNTP와 대응시켜 각각의 염기별로 반응을 행하며 반응후 이 4종의 반응액을 혼합하여 한 lane으로 전기영동한다. 이들 형광발색단은 물리화학적 성질이 상이하야 전기이동도가 상이하야므로 형광색소별

로 보정을 하여야 하고, 각 발색단의 여기광파장이 상이하야므로 이를 detector하는 장치가 필요하다. Dupont의 Genesis 2000 DNA Sequencer는 4종의 형광발색단이 부가된 dideoxynucleotide를 사용하며 (Fig. 6, E), 이들 형광발색단은 물리적성질이 거의 비슷하다. 반응을 한개의 실험관 내에서 행할 수 있고 primer는 보통의 oligonucleotide이므로 경제적이다. 또한 형광발색단이 dideoxynucleotide에 부가되어 있으므로, dideoxy법에서 자주 접하는 DNA 2차구조의 영향으로 생기는 stop현상이 일어나기 어렵기 때문에 정확한 해석이 가능하다. 이들 자동 DNA sequencer는 DNA 해석용 software (DNAsis, GENETYX, MacMolly 등)의 개발과 실용화와 더불어 그 진가를 발휘할 것으로 생각된다. 또한 형광물질해상장치 (형광 Bio Image Analyzer)는 형광색소로 표식된 deoxynucleotide를 사용하여 dideoxy반응시에 표식하며, 반응생성물을 전기영동하여 glass판, 건조 gel, transfer한 membrane filter를 직접 set하여 단시간 (30-1시간)에 computer로 자동해상하는 방식으로 그 이용이 증대될 것으로 예상된다. 특히 여과지에의 전사, 건조가 필요없으므로 대폭적으로 시간단축이 가능하다는 데에 그 이점이 있다.

지금까지 기술한 자동화 DNA sequencer는 설치비용이

많이 든다. 그래서 형광색소를 사용한 수동적인 DNA 염기배열결정법으로서 형광표식부위를 효소적으로 발색시키는 방법의 연구가 진행되고 있다. 그 방법을 개략적으로 설명하면, primer 또는 deoxynucleotide를 photobiotin으로 표식하여 dideoxy법으로 sequencing반응후, gel상의 DNA 단편을 blotting하여 여과지에 전사시키며, 전사된 DNA 단편을 UV조사(파장 254nm, 출력 100 $\mu$  Watts/cm이상)로 고정한다(cross-link). 최종적으로 blocking반응 후 biotin-streptavidin-conjugated alkaline phosphatase 결합을 시키며, DNA검출은 X선 film에 DNA면을 밀착시켜 감광함으로써 행한다(Fig. 7). 이 때 화학발광반응은 2시간에서 최고에 달하여 1주간정도 지속되어 염기배열결정이 가능하다. 형광색소의 화학발광반응이 일반화되면 자동 DNA sequencer의 설치없이도 염기배열결정에 형광색소를 사용 가능하고 특히 별도의 설비없이 일반적인 실험실에서도 염기배열결정이 가능하게 될 것이다.

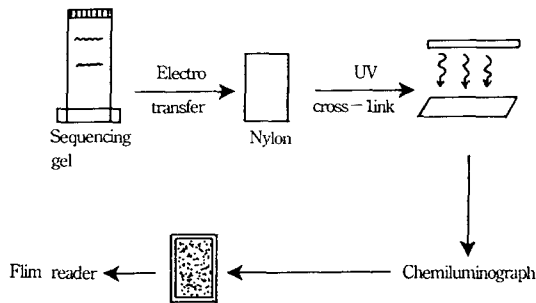


Fig. 7. The steps of chemiluminographic sequencing.

### V. 최근의 16S like 리보솜유전자의 증폭에 의한 분자진화학적인 연구와 그 이용

균류의 계통진화학적인 연구에서는 PCR을 사용한 연구가 미진한 상태이다. L. Hendriks 등의 리보솜DNA의 분자진화적인 연구에서 여성에게 성가신 병인 *Candida*종의 원인균으로 이제까지 담자균계로 알려져 있던 *Candida albicans*가 자낭균계에 근연임이 밝혀졌다<sup>14)</sup> (Fig. 8). 고등식물에 기생하는 생활환, 생리생화학적인 성질이 불분명한 *Taphrina*속과 히말라야의 토양에서 분리된 *Saitoella*속은 미세형태학적, 화학분류학적으로 유사성이 시사되어져 왔

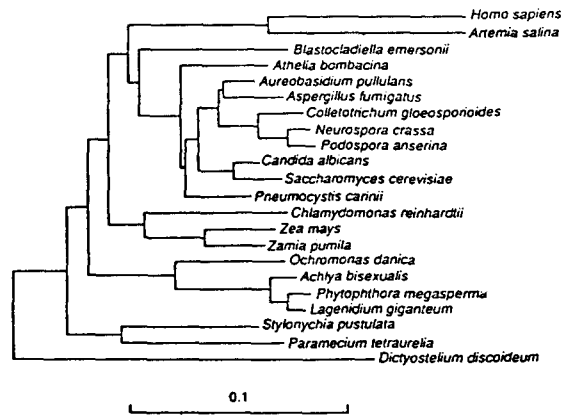


Fig. 8. Phylogenetic tree based on small-subunit rRNA sequences showing positions of *Athelia bombacina*, *Aureobasidium pullulans*, and *Colletotrichum gloeosporioides* among those of 19 other eukaryotes. Lineage were deduced from structural distance data as described by Elwood *et al.* (1985). Evolutionary distance between nodes of the tree are represented by the horizontal component of their separation. Distance corresponding to 10 changes per 100 nucleotide positions is indicated by the bar.

다. 최근 *Taphrina wiesneri*(벚꽃의 witches' broom 병원균)과 *Saitoella complicata*의 18S 리보솜DNA를 PCR로 증폭한 연구<sup>27)</sup>에서 이들은 담자균류와 자낭균류가 아닌 제3의 계통군을 형성하여, 담자균류와 자낭균류가 계통진화상으로 출현하기 이전에 분기되었음이 시사되었다.

PCR법에 의해 16S 리보솜RNA 유전자(약 1500염기)를 해석한 결과, 호알카리성 *Bacillus*속 세균중에 계통이 상이한 균군이 혼재하고 있을 것이라는 지금까지의 생각과는 상이하게 호알카리성 *Bacillus*속은 상호간에 약 95%의 유사도를 나타내어 계통적으로 단일한 것으로 밝혀졌다.<sup>28)</sup>

16S like 리보솜DNA의 증폭과 그 염기배열결정법으로 계통진화관계가 분명하게 되면, 이들 data의 축적을 토대로 하여 균종의 판별에도 이용할 수 있다. 또한 인공배양할 수 없는 미생물의 해석에 유용하게 사용할 수 있을 것이다. 이런 방법은 고등생물에 기생 또는 공생하는 미생물에도 적용할 수 있어서, 미생물생태학의 연구에 광범위하게 사용될 수 있을 것이다. 또한 형태를 제어하는 유전자군의

구조, 조절 및 발현에 대한 구조분자생물학적인 연구를 병행하면, 분자진화와 형태진화를 접목한 통일된 시각의 진화과정의 규명에도 이바지할 것이다.

### 참 고 문 헌

1. Southern, E. M., *J. Mol. Biol.*, **98**, 503-517(1975).
2. Maxam, A. M., and Gilbert, W., *Proc. Natl. Acad. Sci. USA*, **74**, 560-564(1977).
3. Sanger, F., Nicklen, S., and Coulson, A. R., *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467(1977).
4. Zagursky, R. J., Baumeister, K., Lomax, N., and Ber-  
man, M. L., *Gene Anal. Tech.*, **2**, 89-94(1985).
5. Tabor, S., and Richardson, C. C., *J. Biol. Chem.*, **262**,  
15330-15333(1987).
6. Tabor, S., and Richardson, C. C., *Proc. Natl. Acad. Sci.  
USA*, **84**, 4767-4771(1987).
7. Tabor, S., and Richardson, C. C., *J. Biol. Chem.*, **264**,  
6447-6458(1989).
8. Tabor, S., and Richardson, C. C., *Proc. Natl. Acad. Sci.  
USA*, **86**, 4076-4080(1989).
9. Innis, M. A., Myambo, K. B., Gelfand, D. H., and  
Brow, M. A. D., *Proc. Natl. Acad. Sci. USA*, **85**, 9436-  
9440(1988).
10. Chien, A., Edgar, D. B., and Trela, J. M., *J. Bacteriol.*,  
**127**, 1550-1557(1976).
11. Lawyer, F. C., Stoffel, S., Saiki, R. K., and Myambo,  
K., *J. Biol. Chem.*, **264**, 6427-6437(1989).
12. Woose, C. R., *Microbiol. Rev.*, **51**, 221-271(1987).
13. Edman, J. C., Kovacs, J. A., Masur, H., Santi, D. V.,  
Elwood, H. J., and Sogin, M. L., *Nature*, **334**, 519-  
522(1988).
14. Hendriks, L., Goris, A., Neffs, J. M., van de Peer, Y.,  
Hennebert, G., and de Wachter, R., *Syst. Appl. Micro-  
biol.*, **12**, 223-229(1989).
15. Mullis, K. B., and Faloona, F. A., *Methods Enzymol.*,  
**155**, 335-350(1987).
16. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J.,  
Higuchi, R., Harn, G. T., Mullis, K. B., and Erlichm  
H. A., *Science*, **239**, 487-491(1988).
17. Ohara, O., Dorit, R. L., and Gilbert, W., *Proc. Natl.  
Acad. Sci. USA*, **86**, 5673-5677(1989).
18. Wallace, D. H., *Methods Enzymol.*, **152**, 41-48(19  
87).
19. Chomczynsky, P., and Sacchi, N., *Anal. Biochem.*, **162**,  
156-159(1987).
20. Frohman, M. A., Dush, M. K., and Martin, G. R., *Proc.  
Natl. Acad. Sci. USA*, **85**, 8988-9002(1988).
21. Loh, E. Y., Elliot, J. F., Cwirla, S., Lanier, L. L., and  
Davis, M. M., *Science*, **243**, 217-220(1989).
22. Fox, G. E., Stackebranddt, E., Hespell, R. B., Gibson,  
J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W.  
E., Tanner, R., Magrum, L., Zablen, L. B., Blakemore,  
R., Gupta, R., Bonen, L., Lewis, B. N., Stahl, D. A.,  
Luehrsen, K. R., Chen, K. N., and Woese, D. R., *Science*,  
**209**, 457-463(1980).
23. Lane, D., Pace, B., Olsen G. J., Stahl, D. A., Sogin,  
M. L., and Pace, N. R., *Proc. Natl. Acad. Sci. USA*,  
**82**, 6955-6959(1985).
24. Illingworth, C. A., Andrews, J. H., Bibeau, C., and  
Sogin, M. L., *Exp. Mycol.*, **15**, 65-75(1991).
25. Weisburg, W. G., Barns, S. M., Pelletier, D. A., and  
Lane, D. J., *J. Bacteriol.*, **173**, 697-703(1991).
26. Elwood, H. J., Olsen, G. J., and Sogin, M. L., *Mol.  
Biol. Evol.*, **2**, 399-410(1985).
27. 西田洋巳, 杉山純多, 日本農藝化學會 大會講演要指集,  
p. 531(1992).
28. 玉岡迅, 秋葉眺彦, 堀越弘毅, 日本農藝化學會 大會講演  
要指集, p. 527(1992).