

## 세균종동정프로그램에 의한 몇가지 젖산균의 검색

주현성, 김동률, 리석준

오늘 분자생물학의 발전과 더불어 그 수법들이 세균동정분야에도 도입[3, 4]되고있으며 16S 리보솜리보핵산(rRNA)염기배렬분석비교에 의한 세균분류동정기술이 출현하였다.[2-4]

지금까지 분리된 근 1만종에 달하는 세균의 70여만개의 16S rRNA염기배렬자료기지가 확보[1]되었으며 미지세균의 16S rRNA염기배렬분석자료만 있으면 인터넷상에서 세계의 모든 세균들과의 상동성비교를 통해 손쉽게 동정할수 있다.

우리는 인터넷에 접속하지 않고도 누구나 손쉽게 세균검색을 할수 있도록 16S rRNA 염기배렬자료기지를 내리적재하여 세균종동정프로그램을 작성하고 그것을 젖산균의 검색에 응용하기 위한 연구를 하였다.

### 재료와 방법

젖산균배양에는 국제표준배지인 MRS배지[5](펩톤 1%, 소고기엑스 1%, 효모추출물 0.4%, 포도당 2%, 초산나트륨 0.5%, 트윈 80 0.1%, K<sub>2</sub>HPO<sub>4</sub> 0.2%, 레몬산암모늄 0.2%, 류산마그네슘 0.02%, 류산망간 0.005%, 우무 1.5%)를 리용하였다.

PCR장치(《Eppendorf》), 전기영동장치 Mupid-2plus(《Advance》), 영동겔화상입력장치(《ZF-90》), 광학현미경(《Olympus BX51》), 탁상원심분리기(《5415D》(《Eppendorf》)), 랭동원심분리기(《5417R》(《Eppendorf》)), 자동회전식현탁기(《HY-1》) 등을 리용하였다. PCR를 위한 Premix(TaqDNA폴리메라제, Mg<sup>2+</sup>, dNTP, PCR완충액이 함께 들어있음)와 상하류프라이머들은 《TaKaRa》의것을 리용하였다.

상류프라이머: 27F(GAGTTTGATCCTGGCTCAG)

하류프라이머: 1492R(TACGGTTACCTTGTACGACTT)

PCR조건은 94℃ 예비변성 10min→33회전의 94℃ 변성 1min, 53℃ 아닐링 1min, 72℃ 연장 1min→72℃ 최종연장 10min으로 설정하였다. 분석에 리용한 프로그램들은 ChromasPro.exe, MEGA 4.0.exe, Clustal-X.exe, Primer premier, Vector NTI 등이다.

### 결과 및 논의

#### 1) 16S rRNA염기배렬자료기지구축

16S rRNA염기배렬을 리용하는 세균종검색사이트에는 여러가지가 있는데 흔히 리용되고있는것은 NCBI(<http://www.ncbi.nlm.nih.gov/>), Ezbiocloud(<http://www.ezbiocloud.net/>), EzTaxon server 2.1(<http://147.47.212.35:8080/>) 등이다.

검색사이트를 리용한 세균검색에서는 인터넷에 접속한 임의의 사용자가 그 주소를 미리 알고 검색사이트에 들어간 다음 종검색하려는 세균의 16S rRNA염기배렬을 입력하면 그것은 이미 등록된 세균들의 16S rRNA염기배렬자료기지와 상동성검색을 진행하게 되며 상동성이 제일 높은 세균종이 제시되도록 되어있다. 만일 99%이상이면 같은 종으로, 97%이하이면 새로운 균일 가능성이 있다고 본다.

세균의 16S rRNA염기배열자료들은 gbbct자료기지에 등록되어있다. gbbct에는 이밖에 많은 다른 유전자들도 포함되어있으므로 gbbct자료기지에서 세균의 16S rRNA염기배열자료만을 따로 추출하여야 한다.

우리는 Genbank 230.0(2019년 5월판)의 gbbct자료기지를 리용하였는데 이것은 324개의 화일들에 나누어져있으며 매 화일들은 gz형식으로 압축되어있다.

먼저 본문형식의 원천화일로부터 MySQL형식의 gbbct자료기지를 구축하였다. 프로그램을 리용하여 gbbct자료를 하나씩 선택하고 압축을 해제한 다음 자료별로 LOCUS, DEFINITION, ACCESSION, SOURCE 등과 같은 마당이름과 그뒤에 놓이는 정보들을 추출하여 자료기지를 작성하였다. 구축된 자료기지의 마당구성을 보면 표와 같다.

표. 자료기지의 마당구성

마당이름	의미	마당형	마당이름	의미	마당형
LOCUS	유전자자리	본문형	REFERENCE	참고문헌	본문형
DEFINITION	유전자명	본문형	COMMENT	주해	본문형
ACCESSION	등록번호	본문형	FEATURES	구성정보	본문형
VERSION	판본정보	본문형	BASE COUNT	염기수	수값형
KEYWRDS	실마리어	본문형	ORIGIN	기원생물종	본문형
SEGMENT	단편정보	본문형	SEQUENCE	배열정보	본문형
SOURCE	분리원천	본문형			

DEFINITION마당에는 유전자의 이름정보가 들어있는데 질문을 리용하여 16S rRNA염기배열들만을 모두 추출한 결과 총 1 844 429개의 배열들가운데서 340 322개의 배열이 16S rRNA염기배열이었다.

같은 세균종에 대해서도 여러개의 배열이 등록되었는데 그것은 연구자들마다 각이한 조건과 방법으로 분리결정한 배열들이 등록되어있기때문이다.

우리는 같은 세균종에서 여러개의 배열이 등록된 경우 단편배열들과 중복되는 배열들을 제거하는 방법으로 9 965종의 세균에서 총 23 543개의 16S rRNA배열로 된 자료기지를 구축하였다.

## 2) 세균종동정프로그램의 작성

세균의 검색은 해당하는 Blast프로그램을 리용하여 검색하려는 세균의 16S rRNA염기배열을 자료기지의 전체 세균 16S rRNA염기배열자료와 비교검색하고 상동성이 가장 높은 세균을 선택하는 방법으로 진행한다.

16S rRNA염기배열서고의 작성과 형식화 위의 자료기지에서부터 16S rRNA염기배열정보들을 모아 Fasta형식의 화일을 작성하였다.

Fasta형식의 화일에서 매 배열정보는 >기호로 시작하고 그뒤에 배열의 등록번호, 이름과 기원생물종정보를 주며 다음행에 배열정보가 놓인다. 배열정보는 여러개의 행으로 될 수 있다. 배열정보가운데는 A, T, G, C를 비롯한 핵산염기배열정보를 나타내는 문자외에도 수자나 공백기호와 같은 기호들도 포함될 수 있다. 아래에 그 실례를 주었다.

>gbbct000002|AB000278|Photobacterium iliopiscarium gene for 16S ribosomal RNA, p artial sequence.[Photobacterium iliopiscarium|Photobacterium iliopiscarium Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Photobacterium.

```
1 caggcctaac acatgcaagt cgaacggtaa nagattgata gcttgctatc aatgctgacg
61 ancggcgac gggtagtaaa tgccctgggaa tataccctga tgtgggggat aactattgga
121 aacgatagct aataccgcat aatctcttcg gagcaaagag ggggaccttc gggcctctcg
```

```
181 cgtcaggatt agcccagggtg ggattagcta gttggtgggg taatggctca ccaaggcgac
241 gatccctagc tggctctgaga ggatgatcag ccacactgga actgagacac ggtccagact
301 cctacggggag gcagcagtg ggaatattgc acaatggggg aaaccctgat gcagccatgc
.....
```

Blast검색을 진행하기 위해서는 배열화일을 형식화하여야 한다. 배열형식화는 Blast프로그램목록에 들어있는 Formatdb프로그램을 리용하여 진행하였다. Formatdb프로그램은 Blast검색을 고속화하기 위하여 배열정보를 비트정보로 넘기고 가공하는 프로그램으로서 배열자료기지와 배열형식을 인수로 가진다. 16S rRNA배열자료를 변환하기 위한 Formatdb프로그램의 실행방식은 다음과 같다.

Formatdb -i gbbct-16S rRNA -p F

여기서 -i 다음에 준 gbbct-16S rRNA는 Fasta형식의 16S rRNA염기배열화일이며 -p 다음에 F라고 쓴것은 입력화일이 핵산배열이라는것을 의미한다.

결과 용량이 줄어든 3개의 화일(gbbct-16S rRNA.nih, gbbct-16S rRNA.nin, gbbct-16S rRNA.nsq)이 생성되었는데 이것을 Blast검색의 검색자료기지로 리용한다.

질문배열의 입력과 Blast의 실행 Blast프로그램은 여러가지 인수를 가지는데 그가운데서 중요한것은 실행프로그램의 종류, 질문배열화일의 이름, 자료기지화일의 이름, 출력화일의 이름, 출력화일의 형식, 검색배열의 개수 등이다.

질문배열을 입력하면 다음의 인수들을 가지고 Blast를 실행한다.

Blastall -p blastn -i in.seq -d gbbct-16S rRNA -o out.txt -v 500

여기에서 Blastall은 Blast실행화일의 이름이며 -p다음에 blastn을 쓴것은 핵산배열을 핵산자료기지에 대하여 검색한다는 의미이고 in.seq는 질문배열화일의 이름, gbbct-16S rRNA는 형식화된 자료기지화일의 이름, out.txt는 출력화일의 이름이다. -v 다음의 수는 상동성검색결과 출력되어야 할 상동배열의 개수이다.

컴퓨터망을 통한 세균검색프로그램은 PHP언어로 작성하였다.

질문배열정보는 화일로부터 입력하거나 건반을 통하여 직접 입력할수 있게 하였다. 검색단추를 클릭하면 봉사기에서 Blast프로그램이 자동적으로 기동되어 검색결과가 의뢰기컴퓨터에 현시되게 하였다. 결과보기단추를 클릭하면 상동성(%)이 제일 높은 세균이 현시되고 후보배열들과의 정렬결과를 볼수 있게 하였다.

3) 미지세균의 분리와 세균중동정프로그램을 리용한 종검색  
공산유제품으로부터 계열희석법으로 희석액을 만든 다음 MRS평판배지를 리용하여 젖산균순수분리를 하였다.

2개의 미지세균을 분리하고 다시 광학현미경상과 평판배지상에서 그 순수성을 확인한 다음 핵산추출과 PCR, 전기영동을 거쳐 16S rRNA염기배열증폭띠를 조사[2]하였다.(그림 1)

2개의 미지세균의 16S rRNA염기배열증폭띠를 확인한데 기초하여 그것의 16S rRNA염기배열분석을 전문분석기관에 의뢰하였다.

이 균그룹들의 16S rRNA염기배열분석결과는 다음과 같다.

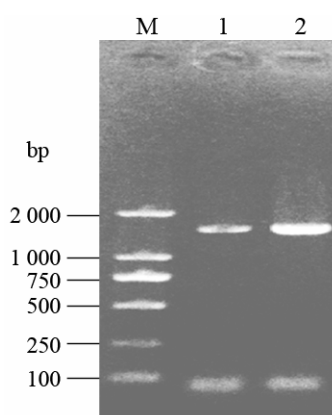


그림 1. 분리균그룹들의 16S rRNA염기배열증폭띠 전기영동상  
M은 분자크기표식자(DL2000 DNA Marker), 1, 2는 분리균그룹

## 미지세균 1의 16S rRNA염기배열(1 418bp)

accaacagattcacttcggtgatgacgttgggaacgcgagcggcggatgggtgagtaaacacgtggg  
 gaacctgccccatagtcctgggataccacttggaaacaggtgctaataccggataagaaagcagatc  
 gcatgatcagcttataaaaggccgcgtaagctgtcgctatgggatggccccgcggtgcattagcta  
 gttggtagggtaacggcctaccaaggcaatgatgcatagccgagttgagagactgatcggccacat  
 tgggactgagacacggcccaaactcctacgggaggcagcagtaggggaatcttcacaatggacgaa  
 agtctgatggagcaacgccgcgtgagtgaaagaggttttcggatcgtaaagctctgttgttgggtga  
 agaaggatagaggtagtaactggcctttatattgacggtaatcaaccagaaagtcacggctaactac  
 gtgccagcagccgcggttaatacgttaggtggcaagcgttgtccggatttatgggctaaagcgagc  
 gcaggcgaagaataagtcctgatgtgaaagccctcggcttaaccgaggaactgcatcggaaactgt  
 tttcttgagtgcagaagaggagagtggaaactccatgtgtagcgggtggaatgctagatatatgga  
 agaacaccagtggcgaaggcggctctctgggtctgcaactgacgctgaggctcgaaagcatgggtag  
 cgaacaggattagataccctggtagtccatgccgtaaacgatgagtgctaagtgttgggaggtttc  
 cgctctcagtgctgcagctaacgcattaaagcactccgcctggggagtagcaccgcaaggttgaaa  
 ctcaaaggaattgacggggggccgcacaaagcgtggagcatgtggtttaattcgaagcaacgcgaa  
 gaaccttaccaggtcttgacatctagtgcactccgtagagatacggagttcccttcggggacacta  
 agacaggtggtgcatggctgtcgctcagctcgtgtcgtagatgttgggttaagtcccgaacgagc  
 gcaaccttgtcatttagttgccagcattaaagtgggcactctaatagagactgccggtgacaaaccg  
 gaggaaggtggggatgacgtcaagtcatcatgccccttatgacctgggctacacacgtgctacaat  
 ggacagtacaacgaggagcaagcctgcgaaggcaagcgaatctcttaaagctgttctcagttcgga  
 atgcagctctgcaactcgaactgcacgaagctggaatcgctagtaatcgcggtatcagcacgcgcggt  
 gaatacgttcccgggccttgtagacacaccgcccgtcacaccatgggagtcgtgcaatgcccaaagccg  
 gtggcctaaccttcgggaaggagccgtctaag

## 미지세균 2의 16S rRNA염기배열(1 369bp)

ggatgggtgagtaaacacgtgggcaatctgccgtaaagactgggataccacttggaaacaggtgcta  
 ataccggataacaacatgaatcgcatgattcaagtttgaaaggcggcgtaagctgtcacttttagga  
 tgagcccgcggcgcattagctagttgggtgggggtgaaggcctaccaaggcaatgatgcgtagccgag  
 ttgagagactgatcggccacattgggactgagacacggcccaaactcctacgggaggcagcagtag  
 gaaatcttcacaatggacgcaagtctgatggagcaacgccgcgtgagtgaaagaggttttcggat  
 cgtaaagctctgttgttgggtgaagaaggatagaggcagtaactgggtctttatattgacggtaatcaa  
 ccagaaagtcacggctaactacgtgccagcagccgcggtaatacgttaggtggcaagcgttgtccgg  
 atttattgggcgtaaagcgagcgcaggcgggaatgataagtctgatgtgaaagcccacggctcaacc  
 gtggaactgcatcggaaactgtcattcttgagtgcagaagaggagagtggaattccatgtgtagcg  
 gtggaatgcgtagatatatggaagaacaccagtggcgaaggcggctctctgggtctgcaactgacgc  
 tgaggctcgaaagcatgggtagcgaacaggattagataccctggtagtccatgccgtaaacgatga  
 gcgctaggtgttggggactttccggctcctcagtgccgcagcaaacgcattaaagcgtccgcctggg  
 gactacgaccgcaaggttgaaactcaaaggaattgacggggggccgcacaaagcgggtggagcatgtg  
 gtttaattcgaagcaacgcgaagaaccttaccaggtcttgacatcctgtgctacacctagagatag  
 gtggttcccttcggggacgcagagacaggtgggtgcatggctgtcgctcagctcgtgtcgtagatgt  
 tgggttaagtcccgaacgagcgcgaaccttgtctttagttgccatcattaagtgggcactctaa  
 agagactgccggtgacaaaccggaggaaggtggggatgacgtcaagtcatcatgccccttatgacc  
 tgggctacacacgtgctacaatgggcagtagcagcgaagcgaacccgcgagggtaagcggatctc  
 ttaaagctgttctcagttcggactgcaggctgcaactcgctgcacgaagctggaatcgctagtaa  
 tcgcggtatcagcacgcgcggtgaatacgttcccagccttgtagacacaccgcccgtcacaccatgg  
 aagtcgtgcaatgccccaaagtcgggtgggataacctttataggagtcagcc

Chromas 2.0 프로그램을 리용하여 염기배열 분석 결과의 정확성을 확인하였으며 상류와 하류의 배열 분석이 잘된 부분들을 따내어 Vector NTI에서 연결(Contig)하여 전체 배열을 얻어내었다.

다음 세균종동정 프로그램의 검색대면부에서 검색배열창에 미지세균 1과 2의 16S rRNA 염기배열자료를 복사하여 검색을 진행하였다.(그림 2, 3)



그림 2. 미지세균 1의 16S rRNA염기배열을 리용한 검색결과

그림 2에서 보는바와 같이 미지세균 1의 16S rRNA염기배열자료를 지금까지 세계적으로 등록된 모든 16S rRNA염기배열자료들과 상동성검색을 진행한 결과 *Lactobacillus acidophilus* B4 군그루의 16S rRNA염기배열과 상동성이 가장 높았다.(99.859%)

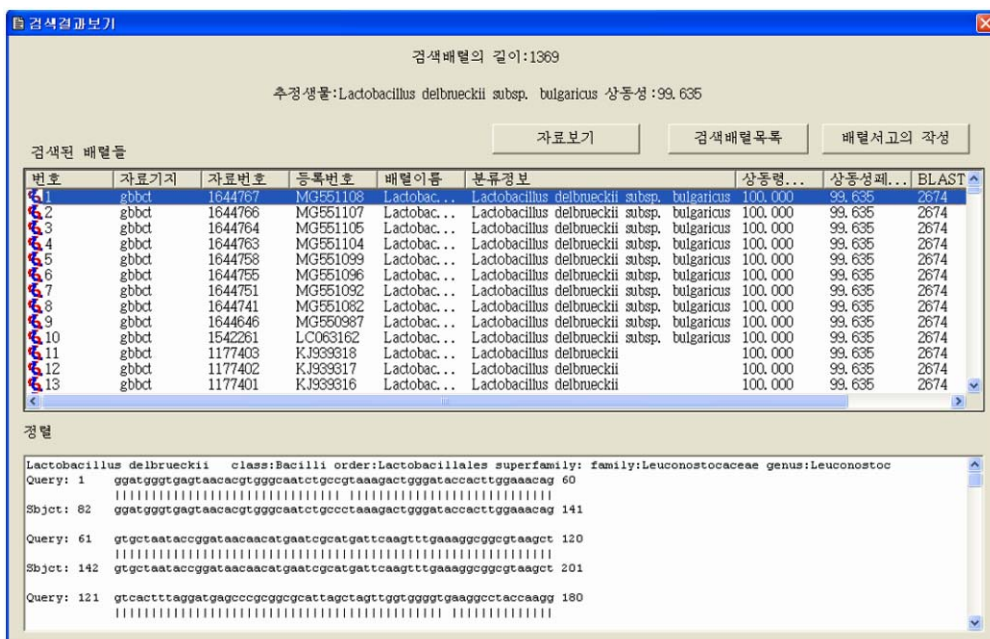


그림 3. 미지세균 2의 16S rRNA염기배열을 리용한 검색결과

그림 3에서 보는바와 같이 미지세균 2의 16S rRNA염기배열자료를 가지고 우와 같이 상동성검색을 진행한 결과 *Lactobacillus delbrueckii* subsp. *bulgaricus* 균그룹의 16S rRNA염기배열과 상동성이 가장 높았다.(99.635%)

이 자료들과 미지세균 1과 2의 형태학적특성, 생리생화학적특성자료들에 기초하여 이 균그룹들을 각각 *Lactobacillus acidophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus*로 동정할수 있었다.

## 맺 는 말

1) 인터넷에 접속하지 않고도 16S rRNA염기배열자료만 있으면 손쉽게 미지세균의 종검색을 할수 있는 세균종동정프로그램을 개발하였다.

2) 이 세균종동정프로그램을 리용하여 산유시료에서 분리한 미지세균 1과 2의 분류학적위치는 각각 *Lactobacillus acidophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus*라는것을 밝혔다.

## 참 고 문 헌

- [1] 김일성종합대학학보(자연과학), 54, 12, 124, 주체97(2008).
- [2] 김동률 등; 조선민주주의인민공화국 과학원통보, 2, 59, 주체103(2014).
- [3] Paul De Vos et al.; Bergey's Manual of Systematic Bacteriology 3, Springer, 464~735, 2009.
- [4] P. T. Kathleen et al.; Foundations in Microbiology, McGraw-Hill, 59~129, 2018.
- [5] 孟祥晨 等; 乳酸菌与乳品发酵剂, 科学出版社, 14~84, 2009.

주체109(2020)년 1월 5일 원고접수

## Identification of Some Lactic Acid Bacteria by Using Identification-Program

*Ju Hyon Song, Kim Tong Ryul and Ri Sok Jun*

Using this bacteria identification-program, with not connection of the internet service, enables us to identify lactic acid bacteria with its 16S rRNA gene data.

Keywords: 16S rRNA, bacteria, identification