

세균종동정프로그램에 의한 세균검색방법

강주현, 김동률, 리석준

분자생물학적수법들이 끊임없이 개발되면서 세균동정분야에도 도입[3]되어 16S 리보솜리보핵산(rRNA)염기배열분석결과의 비교에 의한 세균분류동정기술이 출현하였다.[2, 5] 지금까지 분리된 근 1만종에 달하는 세균의 40여만개의 16S rRNA염기배열자료기지가 확보[1]되었으며 미지세균의 16S rRNA염기배열분석자료만 있으면 인터넷상에서 세계의 모든 세균들과의 상동성비교를 통해 그것을 손쉽게 동정할수 있다.[4]

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

재료와 방법

세균배양에 리용한 배지는 R2A배지(《Difco》)와 TSA배지(《Difco》)이다.[2]

PCR장치(《Eppendorf》), 전기영동장치(《Mupid-2plus》), 영동겔화상입력장치(《ZF-90》), 광학현미경(《Olympus BX51》), 탁상원심분리기(《Eppendorf》), 랭동원심분리기(《Eppendorf》), 자동회전식현탁기(《HY-1》) 등을 리용하였다.

PCR를 위한 Premix(Taq폴리메라제, Mg^{2+} , dNTP, PCR완충액이 함께 들어있음.)와 상하류 프라이머(《TaKaRa》)들은 다음과 같다.

상류프라이머 27F(GAGTTTGATCCTGGCTCAG)

하류프라이머 1492R(TACGGTTACCTTGTTACGACTT)를 리용하였다.

PCR의 설정은 94℃ 10min→94℃ 1min, 53℃ 1min, 72℃ 1min; 33회전→72℃ 10min으로 하였다. 분석에 리용한 프로그램들은 ChromasPro, MEGA 4.0, Clustal-X, Primer premier, Vector NTI 등이다.

결과 및 논의

1) 16S rRNA염기배열자료기지구축에 의한 세균종동정프로그램의 개발

16S rRNA염기배열자료기지구축 16S rRNA염기배열을 리용하여 세균종을 검색하기 위한 싸이트에는 여러가지가 있는데 많은 연구자들이 흔히 리용하고있는것은 NCBI(<http://www.ncbi.nlm.nih.gov/>), Ezbiocloud(<http://www.ezbiocloud.net/>), EzTaxon server 2.1(<http://147.47.212.35:8080/>) 등이다. 검색싸이트를 리용한 세균검색에서는 인터넷에 접속한 임의의 사용자가 그 주소를 미리 알고 검색싸이트에 들어간 다음 종검색하려는 세균의 16S rRNA염기배열을 입력하면 그것은 이미 등록된 세균들의 16S rRNA염기배열자료기지와 상동성검색을 진행

하게 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 197.0(2013년 8월판)의 gbbct자료기지를 리용하였는데 이것은 106개의 화일들에 나누어져있으며 매 화일들은 gz형식으로 압축되어있다.

먼저 본문형식의 원천화일로부터 Access형식의 gbbct자료기지를 구축하였다. Access자료기지작성과정의 흐름도식을 그림에 주었다.

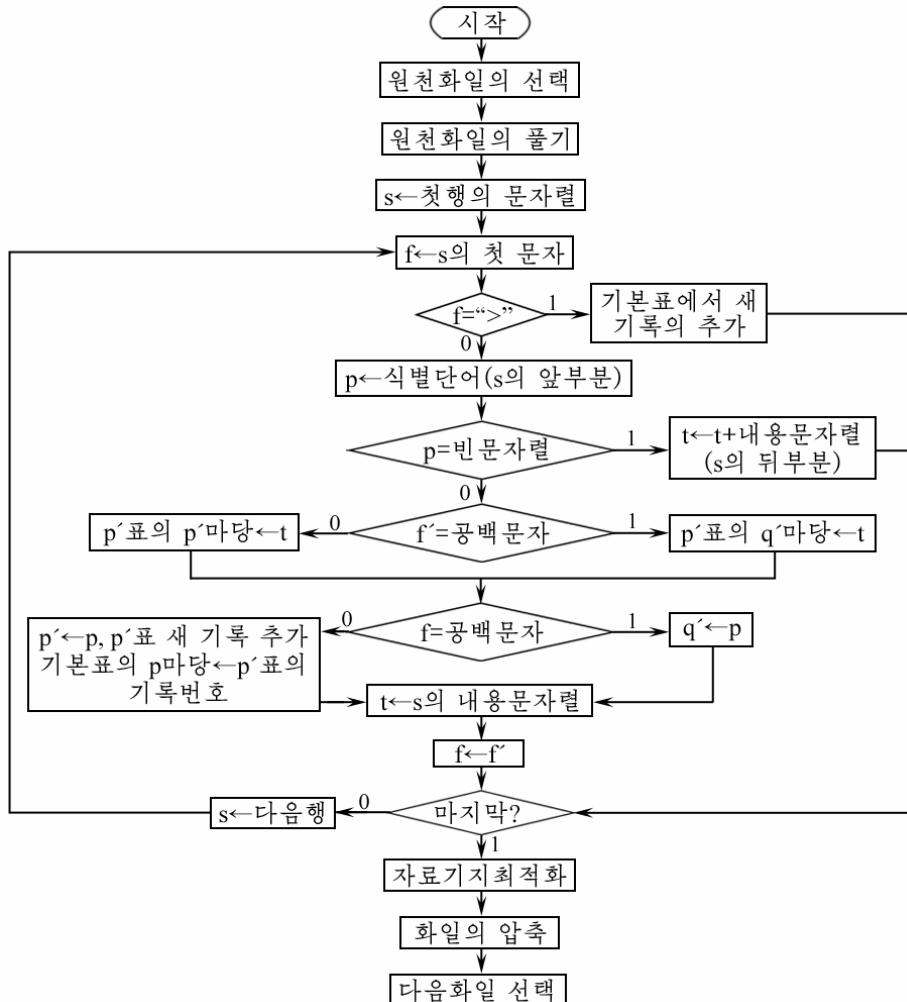


그림. 원천화일로부터 자료기지작성과정의 흐름도식

Visual C++로 작성한 프로그램을 리용하여 gbbct자료를 하나씩 선택하고 압축을 해제한 다음 자료별로 LOCUS, DEFINITION, ACCESSION, SOURCE 등과 같은 마당이름과 그 뒤에 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	열쇠어	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 1 003 316개의 배열들가운데서 440 322개의 배열이 16S rRNA염기배열이었다.

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

세균종동정프로그램의 작성 세균의 검색은 해당한 blast프로그램을 리용하여 검색하려는 세균의 16S rRNA염기배열을 자료기지의 전체 세균 16S rRNA염기배열자료와 비교검색하고 상동성이 가장 높은 세균을 선택하는 방법으로 진행한다.

16S rRNA염기배열서고의 작성과 형식화 위의 자료기지에서 16S rRNA염기배열정보들을 모아 fasta형식의 화일을 작성하였다.

fasta형식의 화일에서 매 배열정보는 >기호로 시작하고 그뒤에 배열의 등록번호, 이름과 기원생물종정보를 주며 다음 행에 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.

>gbbc000002|AB000278|Photobacterium iliopiscarium gene for 16S ribosomal RNA, partial sequence.[Photobacterium iliopiscarium|Photobacterium iliopiscarium Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Photobacterium.

```
1 caggcctaac acatgcaagt cgaacggtaa nagattgata gcttgctatc aatgctgacg
61 ancggcggac gggtagtaaa tgcctgggaa tataccctga tgtgggggat aactattgga
121 aacgatagct aataccgcat aatctcttcg gagcaaagag ggggaccttc gggcctctcg
181 cgtcaggatt agcccagggtg ggattagcta gttgggtggg taatggctca ccaaggcgac
241 gatccctagc tgggtctgaga ggatgatcag ccacactgga actgagacac ggtccagact
301 cctacgggag 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 배열자료지와 배열형식을 인수로 가진다. 16S rRNA배열자료를 변환하기 위한 formatdb프로그램의 실행방식은 다음과 같다.

```
formatdb -i gbbct-16S rRNA -p F
```

여기서 -i다음에 준 gbbct-16S rRNA는 fasta형식의 16S rRNA염기배열화일이며 -p다음에 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 blast

검색의 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`

여기서 blastall은 blast실행화일의 이름이며 -p다음에 blastn을 쓴것은 핵산배열을 핵산자료 기지에 대하여 검색한다는 의미이고 in.seq는 질문배열화일의 이름, gbbct-16S rRNA는 형식화된 자료기지화일의 이름, out.txt는 출력화일의 이름이다. 한편 -v다음의 수는 상동성 검색결과 출력되어야 할 상동배열의 개수이다.

Visual C++를 리용하여 세균검색프로그램을 작성하였다.

질문배열정보는 화일로부터 입력하거나 건반을 통하여 직접 입력할수 있게 하였다. 검색단추를 누르면 blast프로그램이 자동적으로 기동되어 검색결과가 본문형식으로 현시 되게 하였다. 결과보기단추를 누르면 상동성(%)이 제일 높은 세균이 현시되고 후보배열들의 정렬결과를 볼수 있게 하였다.

2) 세균종검색

미지세균의 분리와 검색싸이트를 리용한 종검색 토양시료로부터 계열회석법으로 회석액을 만든 다음 R2A평판배지를 리용하여 세균순수분리를 진행하였다. 2개의 미지세균을 분리하고 광학현미경상과 평판배지상에서 그 순수성을 확인한 다음 핵산추출과 PCR, 전기영동을 거쳐 16S rRNA염기배열증폭터를 조사하였다.[2] 그 16S rRNA염기배열을 전문분석소에서 분석하였는데 그 결과는 아래와 같다.

미지세균 2의 염기배열

```
GCTCCCTCCCACAAGGGGTTAGGCCACCGCTTCGGGTGTTACCAACTTTTCGTGACTTGACGGGC
GGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGCGTTGCTGATCTGCGATTACTAGCGACTCCGAC
TTCATGGGGTCGAGTTGCAGACCCCAATCCGAAGTACGACCGGCTTTTTGGGATTAGCTCCACCTCAC
AGTATCGCAACCCCTTTGTACCGGCCATTGTAGCATGCGTGAAGCCCAAGACATAAGGGGCATGATGAT
TTGACGTCGTCCCCACCTTCTCCGAGTTGACCCCGGCAGTCTCCCATGAGTCCCCGCCATTACGCGC
TGGCAACATGGAACGAGGGTTGCGCTCGTTGCGGGACTTAACCCAAACATCTCACGACACGAGCTGACG
ACAACCATGCACCACCTGTGAACCGGCCCCAAAGGGGAAACCACATTTCTGCGGGCGGTCCGGTCCATG
TCAAGCCTTGGAAGGTTCTTCGCGTTGCATCGAATTAATCCGCATGCTCCGCCGCTTGTGCGGGCCC
CCGTCAATTTCCTTTGAGTTTTAGCCTTGCGGCCGTACTCCCCAGGCGGGGCACTTAATGCGTTAGCTA
CGGCGCGGAAAACGTGGAATGTCCCCACACCTAGTGCCCAACGTTTACGGCATGGAATACCAGGGTA
TCTAATCCTGTTGCTTCCCATGCTTTGCTCCTCAGCGTCAGTTAATGCCAGAGACCTGCCTTCGC
CATCGGTGTTCTCCTGATATCTGCGCATTTACCGCTACACCAGGAATCCAGTCTCCCCCTACATCA
CTCTAGTCTGCCGTACCCACCGCAGATCCGGAGTTGAGCCCCGGACTTTACGGCAGACGCGACAAA
CCGCCTACGAGCTCTTTACGCCCAATAATTCGGGATAACGCTTGCGCCCTACGTATTACCGCGGCTGC
TGGCACGTAGTTAGCCGGCGCTTCTTCTGCAGGTACCGTCACTTTTCGCTTCTTCCCTACTGAAAGAGG
TTTACAACCCGAAGGCCGTCATCCCTCACGCGGCGTCGCTGCATCAGGCTTGCGCCCATTGTGCAATA
TTCCCCACTGCTGCCTCCCGTAGGAGTCTGGGCCGTGTCTCAGTCCAGTGTGGCCGGTCACCCCTCTC
AGGCCGGCTACCCGTCGTCGCTTGGTGAGCCATTACCTCACCAACAAGCTGATAGGCCGCGAGTCCA
TCCAAAACCACAAAAGCTTTCCACCCCCCACCATGCGATGAGGAGTCATATCCGGTATTAGACCCAGT
TTCCAGGCTTATCCAGAGTTAAGGGCAGGTTACTCACGTGTTACTACCCGTTGCCACTAATCCC
CCCAGCAAGCTGGAGATCATCGTTGCACTTGCA
```

미지세균 6의 염기배열

TCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAACGCGTGGAATCTACCCCTTGACTACG
 GAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCTCTCGGGAGAAAGATTTATCGGTCAAGGA
 TGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCATAGCTGGTCT
 GAGAGGATGATCAGCCACATTGGGACTGAGACACGGCCCAAACCTCTACGGGAGGCAGCAGTGGGGAA
 TATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAA
 GCTCTTTACCGGAGAAGATAATGACGGTATCCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
 CGCGGTAATACGAAGGGGGCTAGCGTTGTTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGA
 TCAGTCAGGGGTGAAATCCCAGGGCTCAACCCTGGAAGTGCCTTTGATACTGTGATCTGGAGTATGG
 AAGAGGTGAGTGGAATTCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAA
 GGCGGCTCACTGGTCCATTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCC
 TGGTAGTCCACGCCGTAAACGATGAATGTTAGCCGTCGGGCAGTATACTGTTTCGGTGGCGCAGCTAAC
 GCATTAAACATTCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCCGC
 ACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATGCCCCG
 GCTACTTGCAGAGATGCAAGGTTCCCTTCGGGGACCGGGACACAGGTGCTGCATGGCTGTCGTCAGCT
 CGTGTGCTGAGATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTCGCCCTTAGTTGCCAGCATTACAG
 TTGGGCACTCTAAGGGGACTGCCGGTGATAAGCCGAGAGGAAGGTGGGGATGACGTCAAGTCCTCATG
 GCCCTTACGGGCTGGGCTACACACGTGCTACAATGGTGGTGACAGTGGGCAGCGAGCACGCGAGTGTG
 AGCTAATCTCCAAAAGCCATCTCAGTTCGGATTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCG
 CTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGACACACCGCCCCGTCACAC
 CATGGGAGTTGGTTTTACCCGAAGGTAGTGCGCTAACCGCAAGGAGGCAGC

Chromas 2.0 프로그램을 리용하여 염기배열분석결과와 정확성을 확인하였으며 상류와 하류의 배열분석이 잘된 부분들을 따내어 Vector NTI에서 연결(Contig)하여 전체 배열을 얻어내었다.

이 16S rRNA염기배열자료들을 인터넷상에서 세계 모든 세균들의 16S rRNA염기배열자료들과 비교하여 종검색을 진행하였다.

검색결과에 의하면 미지세균 2와 6은 비교되는 종과 16S rRNA염기배열의 상동률이 각각 99.928, 99.776%이므로 그 상동률이 99%이상이면 비교되는 종으로 본다는 선행연구 결과[3, 5]에 기초하여 각각 *Arthrobacter pascens*, *Rhizobium leguminosarum*으로 그 분류학적 위치가 밝혀졌다.

세균종동정프로그램을 리용한 종검색 세균종을 검색하기 위한 대면부를 작성하였다.

검색배열창에 미지세균 2와 6의 염기배열자료를 복사해넣든가 아니면 《화일로부터 입력》단추를 클릭하여 해당 16S rRNA염기배열자료를 불러들인다. 《검색》단추를 클릭하고 잠깐 기다리면 결과가 현시된다. 《결과보기》단추를 클릭하면 구체적인 자료를 열람할수 있다.

세균종동정프로그램을 리용한 검색에서도 미지세균 2와 6은 각각 *Arthrobacter pascens*, *Rhizobium leguminosarum*으로 동정되었다.

세균검색프로그램을 리용한 이러한 검색결과는 검색싸이트를 리용하여 동정한 검색 결과와도 일치한다. 이것은 인터넷에 접속하지 않고도 임의의 시간에 손쉽게 세균종검색을 진행할수 있다는것을 보여준다.

맺 는 말

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

2) 개발한 세균종동정프로그램을 리용하여 미지세균 2와 6을 종검색한 결과 각각 *Arthrobacter pascens*, *Rhizobium leguminosarum*으로서 검색싸이트를 리용한 검색결과와 일치하였다.

참 고 문 헌

- [1] 김일성종합대학학보(자연과학), 54, 12, 124, 주체97(2008).
- [2] 김동률 등; 조선민주주의인민공화국 과학원통보, 2, 59, 주체103(2014).
- [3] N. R. Krieg et al.; Bergey's Manual of Systematic Bacteriology 4, Springer, 1~56, 2010.
- [4] J. P. Harley; Laboratory Exercises in Microbiology, Springer, 37~186, 2005.
- [5] 陶天申 等; 原核生物系统学, 化学工业出版社, 15~116, 2007.

주체105(2016)년 1월 5일 원고접수

Identification of Bacteria by using Program

Kang Ju Hyon, Kim Tong Ryul and Ri Sok Jun

By using this bacteria identification-program, we can identify some bacteria with its 16S rRNA gene database without Internet service.

Key words: 16S rRNA, bacteria, identification