2024년 순천향대학교 과학캠프

- 생물정보학을 활용한 DNA 분석 -

지도교수 : 이용석 / 실습조교 : 정준양, 이혁

1. DNA의 구조

1953	DNA 이중나선구조 규명
1956	DNA 엮는 중합효소 발견
1960	가위 역할하는 제한효소 사용
1962	DNA 2중 나선 구조가 밝혀짐
1966	유전암호 해독법 등장
1972	DNA 재조합 생명체 등장
1977	DNA 서열 분석 방법 개발
1983	DNA 증폭 기술 개발
1984	DNA 지문분석법 발명
1990	유전자치료 실시
1994	최초의 유전자조작 식품 승인
2001	인간게놈지도 초안 작성
2003	인간게놈지도 작성 완성

DNA 구조가 밝혀진 후의 50년 역사

1953년 4월 25일, James Dewey Watson과 Francis Crick이 DNA X선 사진을 분석해 구조를 설명한 1쪽 짜리 짧은 논문을 완성해 영국 과학잡지인 '네이처'에 실었다. James Watson 과 Francis Crick, 그리고 윌킨스는 DNA 이중나선구조를 밝힌 공로로 1962년 노벨 생리의학상을 수상했다. 이렇듯 DNA의 이중나선구조가 밝혀진 지 올해로 53년이 된다. DNA는 생명체의 설계도라 할 수 있는 정보를 간직한 세포 내 물질이다. 자식이 부모를 닮는 까닭은 바로 DNA를 물려받기 때문이다. 이중나선을 이루는 사슬을 부모에게서 절반씩 물려받는 것이다. 이중나선에 담긴 DNA의 유전정보는 RNA라는 전령을 통해 단백질을 만들도록 명령을 내린다. 단백질은 생체를 구성하고, 모든 생명현상을 담당하는 실제 일꾼이다. 또한 DNA의 이중나선을 조작하면 의약품을 비롯한 유용한 물질을 생산하는 일이 가능하다. 따라서 DNA 이중나선구조의 규명은 생명공학 혁명의 출발점이 되었다. 그 이후 1966년 유전암호 해독법이 등장하고, 1972년 DNA 재조합 생명체가 등장한 후 1977년 드디어 DNA의 서열 분석 방법이 개발되었다. 염기서열 분석 즉 Nucleotide Sequencing의 원리는 1958년 1980 년 2회에 걸쳐서 노벨상을 수상한 Sanger Institute 소속의 Fredrick Sanger에 의해 밝혀졌다. 1958년 그는 insulin 단백질의 구조를 밝힌 공로를 인정 받아서 노벨상을 받았으며, 1980년 염기서열을 분석하는 원리를 밝힘으로서 노벨상을 받았다.

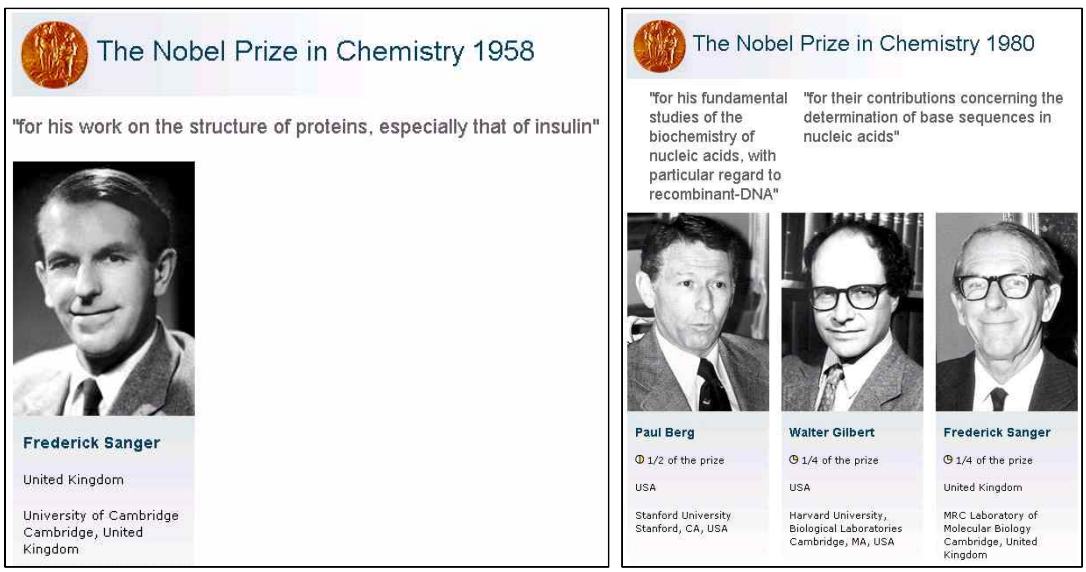


그림 1. 노벨상을 2차례 받은 Fredrick Sanger

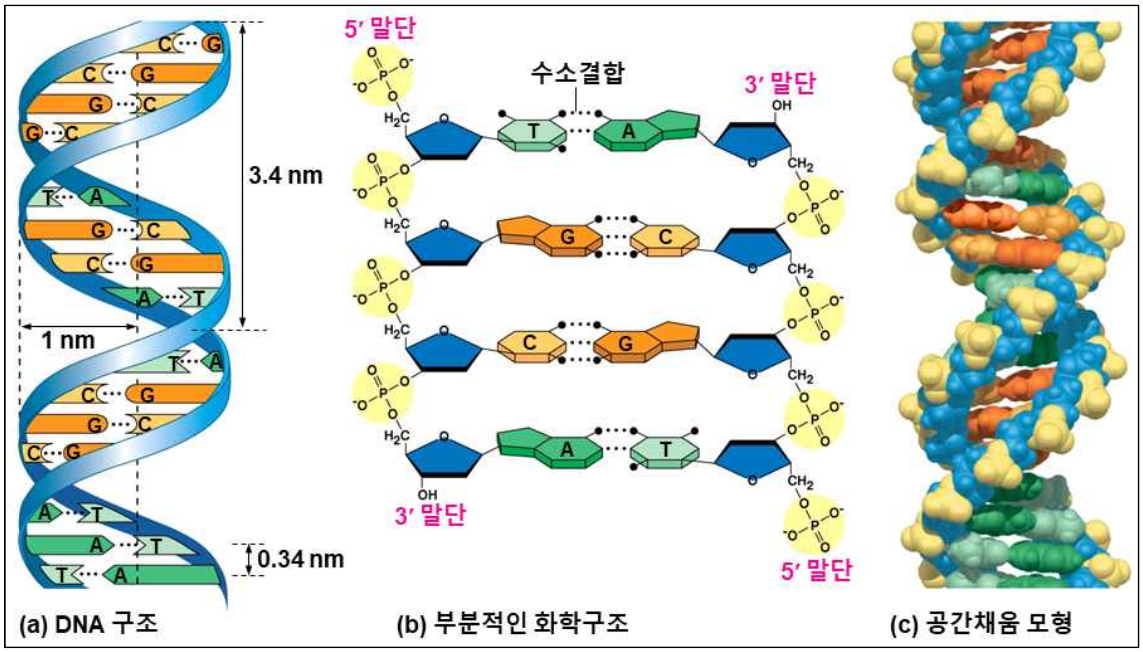


그림 2. DNA의 구조

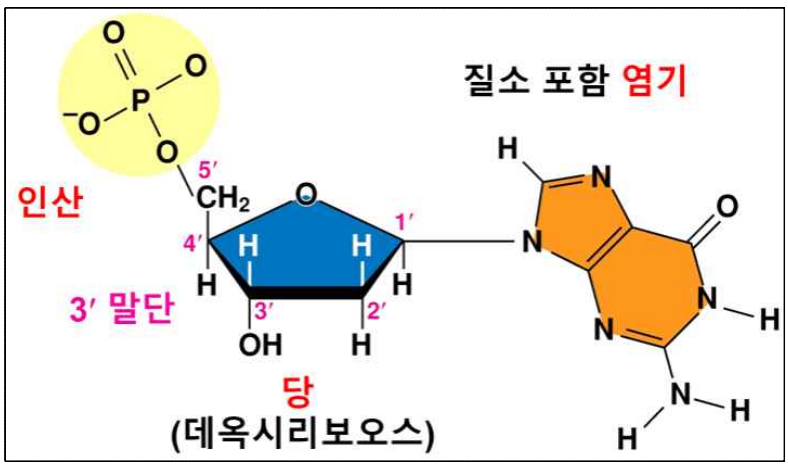


그림 3. DNA 의 뉴클레오타이드

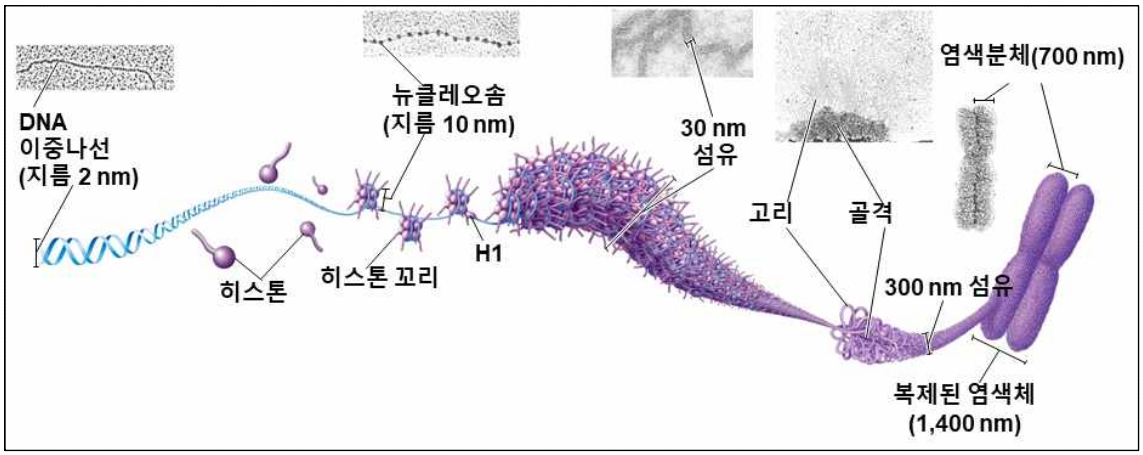


그림 4. DNA Packing

2. Central Dogma

- DNA makes RNA makes Proteins

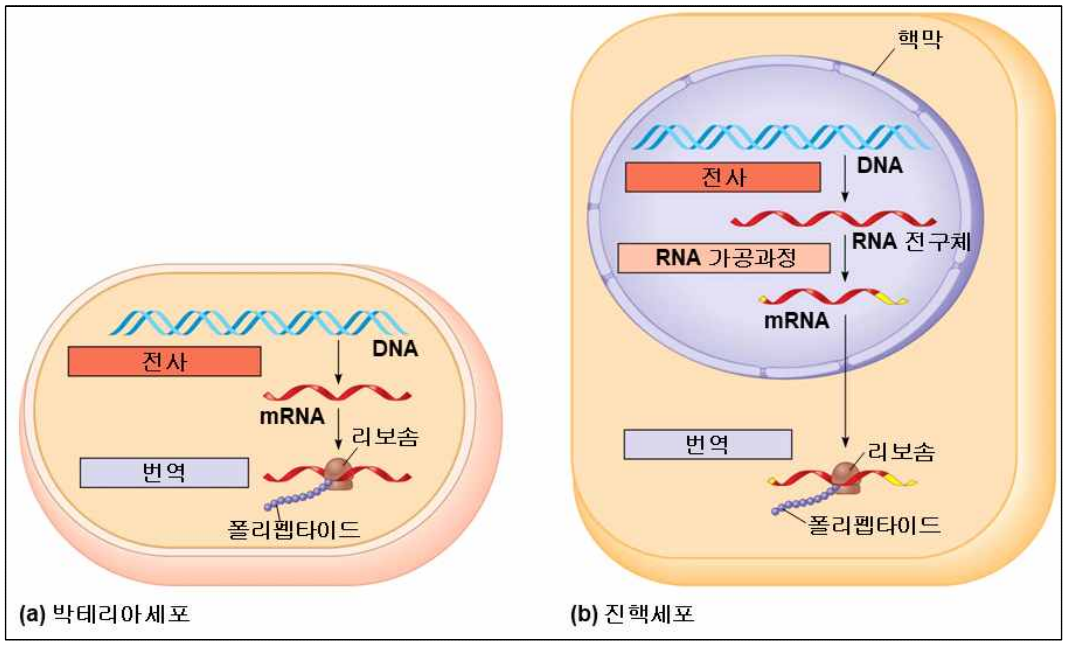


그림 5. 원핵생물과 진핵생물의 전사, 번역과정은 차이가 있다. 원핵세포에서는 mRNA의 번역이 전사가 끝나기 전에 시작할 수 있다. 진핵세포에서는 핵막에 의해 전사와 번역이 격리되어있다.

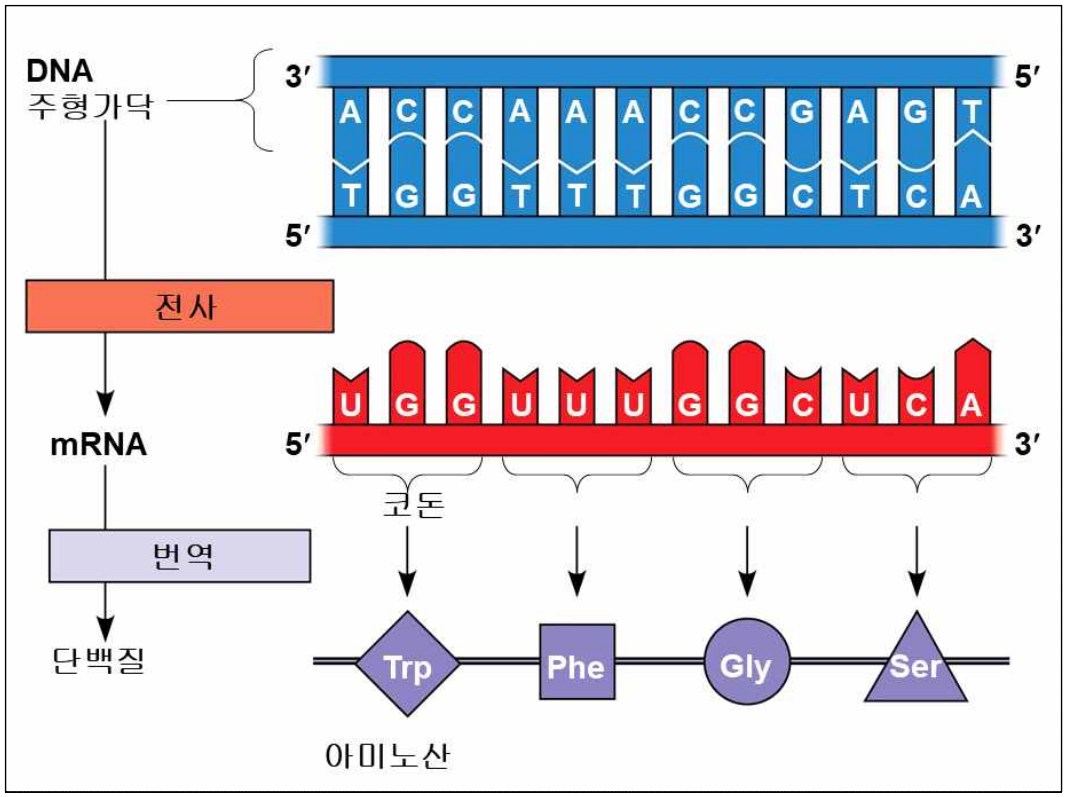


그림 6. 번역동안에 mRNA 세개염기, 즉 코돈은 5'에서 3'방향으로 읽혀진다.

3. DNA sequencing 의 원리

PCR 원리 (동영상 설명)를 활용한 염기서열분석 방법

1. 첫 번째 핵심원리
   Dideoxynucleotides (ddATP, ddGTP, ddCTP and ddTTP)을 이용한 Chain termination method

2. 두 번째 핵심원리
   전기영동 (electrophoresis)

3. Labeling 기술

4. 컴퓨터를 이용한 자동화 시스템

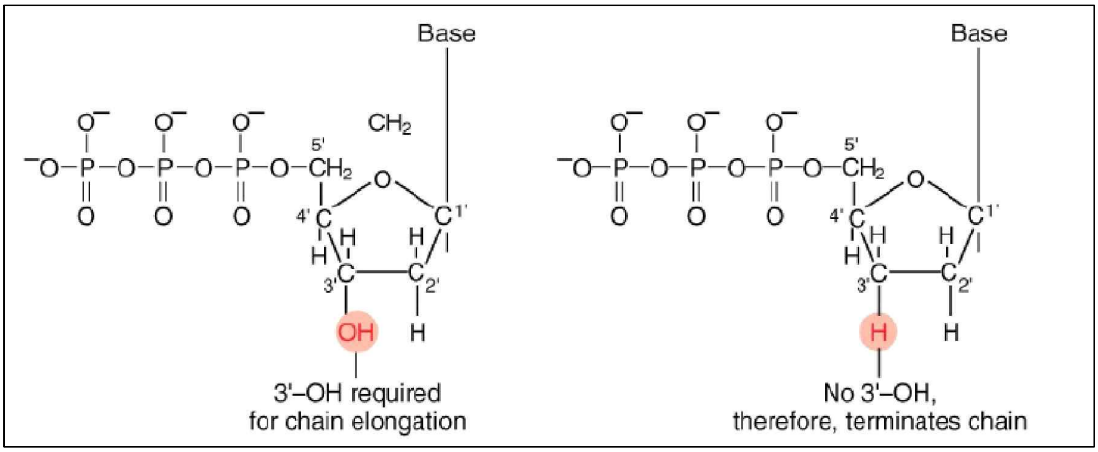


그림 7. Dideoxynucleotides 와 Deoxynucleotide 의 차이점

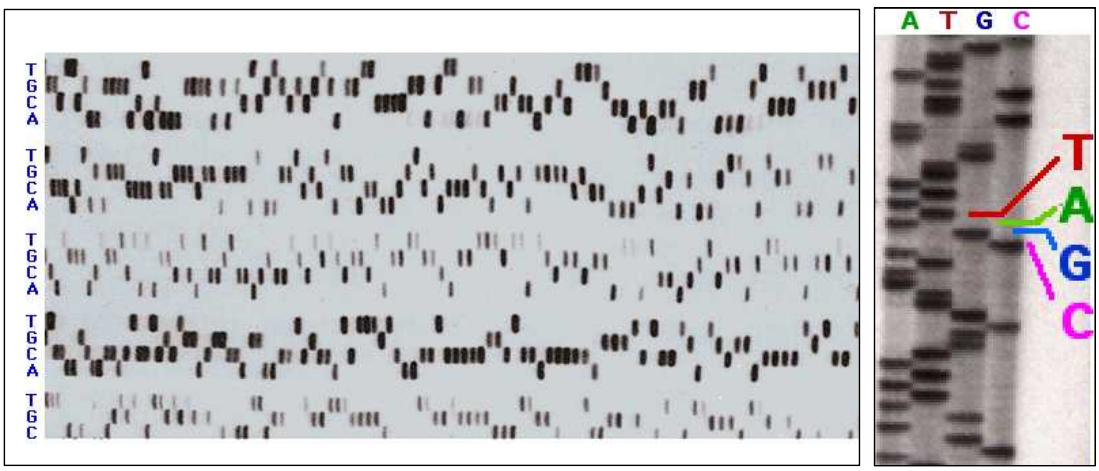


그림 8. Fredrick Sanger에 의해 개발되어진 염기서열 분석방법

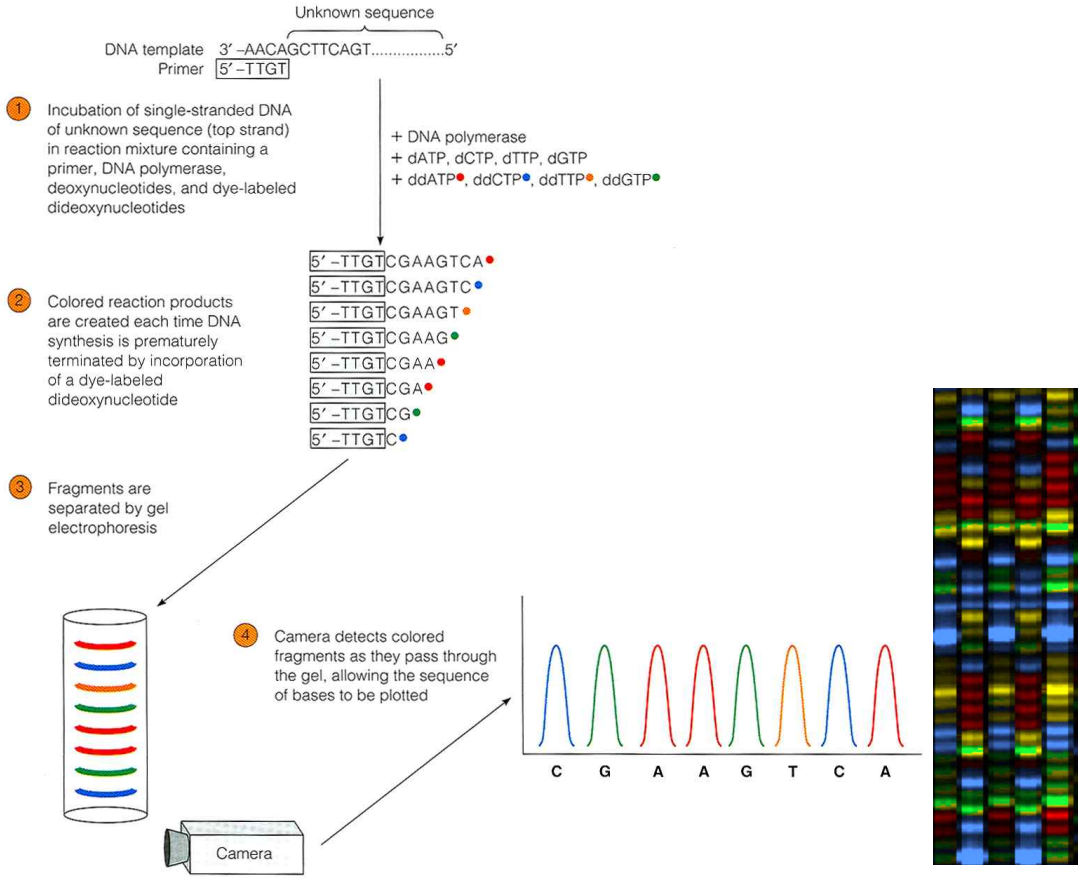


그림 9. 염기서열 자동분석기의 원리

1. PCR 을 하는데 있어서... ddNTP 를 사용하면 chain 합성이 중지되어집니다. 이것은 DNA의 ribose의 3번탄소에 OH^- 대신 H^+가 붙어있기 때문에 더이상 염기를 붙일 수 없기 때문입니다.

2~3. 이 원리를 이용 ddNTP 즉 ddATP, ddTTP, ddGTP, ddCTP 각각에 형광물질을 첨가한 후 PCR 을 한 후 전기영동을 합니다. 이때 염기서열 하나도 분리되어질 수 있도록 제작된 젤을 사용해서 합니다. 그러면 3번 그림처럼 길이에 따라 분리되어집니다.

4. 그리고 Laser를 통해 각각의 dye를 읽어서 나온 파장을 읽으면 염기서열 분석이 이루어 집니다.

5. 염기서열 분석결과는 binary file 형태의 chromatogram 파일로 나오는데 자동염기서열 분석기 중 인간게놈프로젝트에 사용되어진 Applied Biosystems (ABI) 장비에서 나오는 ab1 파일이 보편적이며 장비에 따라서 scf (standard chromatogram format), esd (megabace) 파일 등도 있다.

	1세대 염기서열분석기	2세대 염기서열 분석기 (NGS 1세대)			3세대 염기서열 분석기 (NGS 2세대)		
장비사진							
방식	Sanger sequencing	454 Roche GS-FLX+ system	Illumina Hiseq X System	ABI Solid 4hq system	Ion Torrent S5/S5 XL system	PacBio RS Ⅱ	Oxford Nanopore MinION
	체인 종결자 방식	Pyrosequencing	합성에 의한 서열분석법	Ligation 방식의 서열분석	Ion semiconductor 방식	단일분자 실시간 서열분석	Single Molecule sequencers
Read length	400~900 bp	700 bp (최대 1,000 bp 까지)	2x150 bp	2 x 75 또는 75 x 35 bp	200 bp 이상	250 bp – 40 kb	100 kb 이상
Accuracy	99.90%	99.997%	97%	99.99%	98%	99.999%	96%
Reads per run	384	1 million	1.6 to 1.8 Tb	Up to 300 Gb	60 to 80 million	35,000~75,000	10GB~20GB (standard)
Time per run	20분~3시간	23 시간	3일 이상	3일~2주	2~4 시간	0.5~6시간	6~48시간
COST (US$)	$500/million bases	$8.57/million bases	$1,000/genome†	$1,000/genome†	$11.43/million bases	$0.4-$0.8/million bases	$6.44-$17.90/ million bases
장점	분석서열 길이가 길고 많은 응용프로그램 활용가능	속도가 빠르고 분석서열 길이가 길다.	분석된 서열 데이터 량이 많다	분석 비용이 싸다	비교적 장비가격이 싸고, 분석속도가 빠름	길게 읽는 장비이며 분석속도가 빠름	장비를 사지않아도 되는 KIT 형식
단점	대량서열 분석시 분석비용이 비쌈	분석비용이 타 장비에 비해 비싼 편이며, Homo polymer sequencing error 有	장비가격이 매우 비싸다	다른 방식에 비해 분석속도가 매우 느리며 분석된 서열 길이가 짧다	Homo polymer sequencing error 有	분석비용은 적당한 편이나 장비가격이 매우 비쌈	현재까지는 정확도가 매우 떨어짐

† genome : genome represents sequencing of the human genome at 30X coverage.

그림 10. 자동염기서열분석기의 발달사

4. NCBI 소개 및 BLAST 프로그램 설명

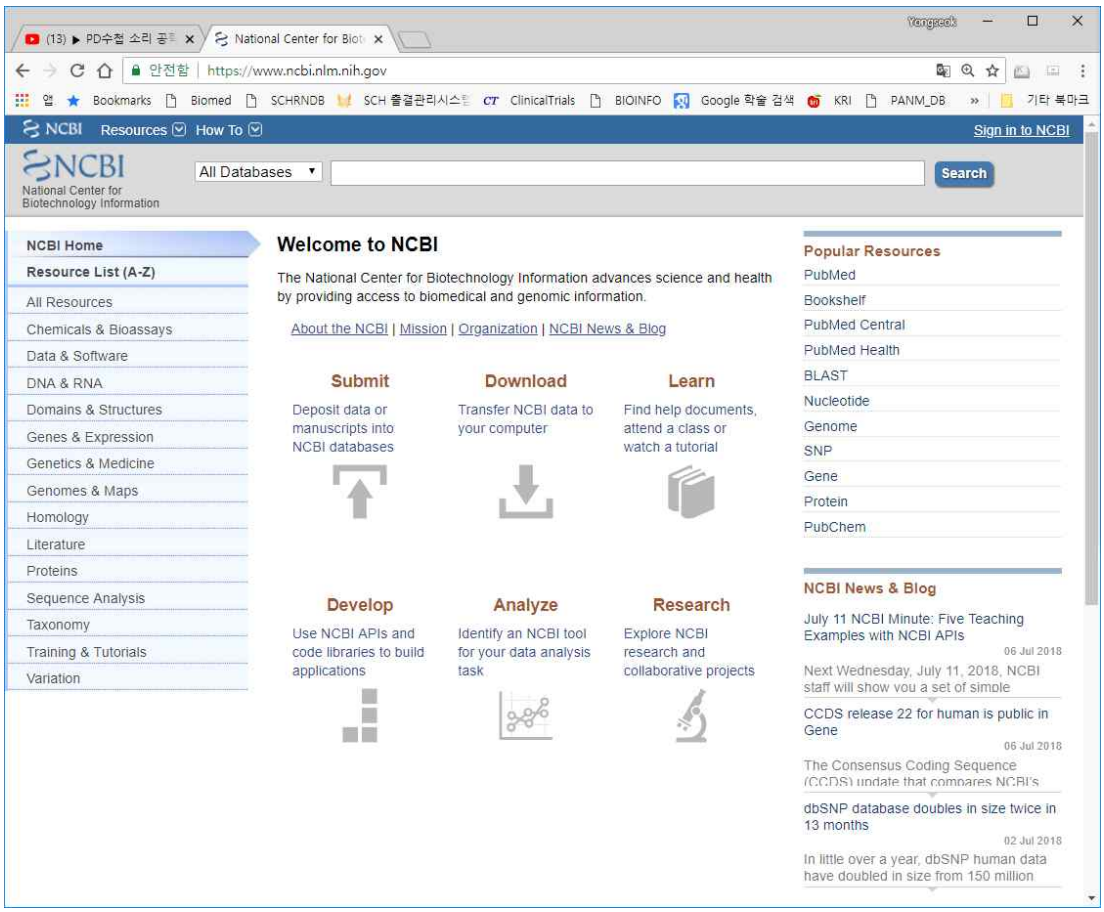


그림 11. NCBI (The National Center for Biotechnology Information) 웹페이지. https://www.ncbi.nlm.nih.gov

PROGRAM	QUERY	DB
blastp	protein	protein
blastn	nucleotide	nucleotide
blastx	NT→AA	protein
tblastn	protein	NT→AA
tblastx	NT→AA	NT→AA

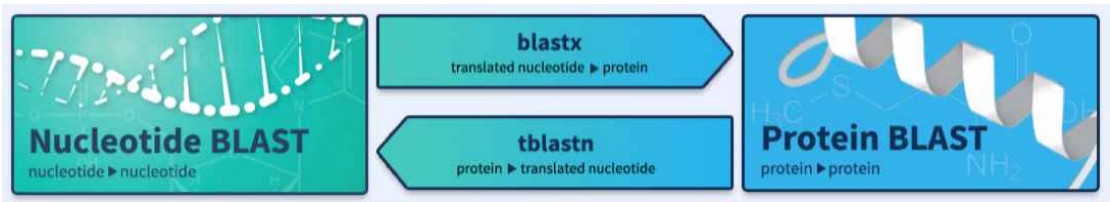


그림 12. BLAST SW 종류. https://blast.ncbi.nlm.nih.gov/Blast.cgi

5. DNA 추출 및 실험방법 간단 소개

<참고자료 : 실재 실험은 진행하지 않습니다>

실험 목적

DNA 추출 원리를 이해하고, 유전체 DNA를 직접 추출하여 눈으로 확인한다.

실험 개요

1. 유전체 DNA(Genomic DNA)란?

일반적으로 세포내에 있는 전체 DNA를 유전체(genomic) DNA라고 한다. 즉 그 생명체가 살아가는데 필요한 모든 유전정보를 담고 있는 물질을 말한다. DNA를 추출하는 방법에는 여러 가지 방법이 있고 세포조직 및 DNA구조에 따라 추출방법이 다르다. 본 실험에서는 식물조직으로부터 유전체 DNA를 추출하는 방법과 원리에 대해 알아본다.

2. DNA 추출의 원리

세포로부터 DNA를 분리하기 위해서는 먼저 DNA가 세포 밖으로 유출될 수 있도록 세포벽 및 세포막을 분해시켜야 하며, 이어서 세포막을 파괴할 때 유리되는 DNA를 분리하여야 하고, 단백질이나 RNA의 혼입을 최소로 줄이면서 DNA를 분리해야 한다. 세포벽을 가지는 생물체의 경우 라이소자임(lysozyme) 효소를 처리하여 세포벽을 분해시키고 세정제의 일종인 Sodium Dodecyl Sulfate (SDS)를 처리하여 세포막을 분해한다. 이렇게 되면 DNA가 세포 밖으로 유출된다. SDS는 세포막을 분해시킬 뿐 아니라 단백질을 변성시키기도 하므로 핵산가수분해효소들의 작용에 필요한 금속양이온과 착염을 형성함으로써 효소의 활성을 저해시킨다.

한편, 세포 밖으로 유출된 DNA 분획은 페놀, 클로로포름, 아이소아밀알콜 등의 유기용매를 이용하여 불순물을 제거한 후, 에탄올을 이용하여 침전시켜서 분리할 수 있다. 또 다른 방법으로는 실리카 등을 이용한 고체로 된 층에 정전기적 인력에 의해 DNA가 결합하는 원리를 이용하여 분리할 수 있다. 적정한 농도의 염(Na^+)과 높은 pH를 갖는 완충용액에 DNA를 넣고 통과시키면 실리카로 구성된 막에 DNA가 결합하게 된다. 결합된 DNA는 물 또는 *TE 완충용액을 이용해서 녹일 수 있다.

* TE 완충용액 : Tris, EDTA로 구성되어 있어 DNA와 RNA의 변성 및 분해를 방지

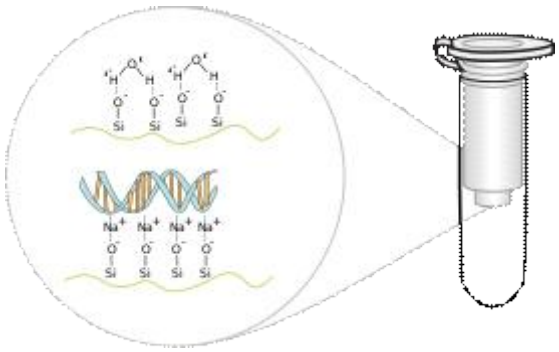


그림 2-1. 실리카 막에 물 또는 DNA와 결합

3. 간단한 DNA 추출

위에서 설명한 일반적인 DNA추출법은 잘 정제된 DNA를 얻을 수 있으나 실험절차가 복잡하고 페놀, 클로로포름 등의 몸에 유해한 시약을 사용하며 고가의 kit와 장비를 필요로 한다. 그러나 실생활에서 쉽게 구할 수 있는 세정제와 소금과 에탄올을 이용해서도 간단히 DNA추출 실험을 해볼 수 있다. 세정제는 SDS용액처럼 세포막을 분해시키고 유리되어 나온 DNA는 소금의 Na+와 결합하고 후에 에탄올을 처리하면 DNA가 에탄올 안에서 엉켜 눈으로 확인을 할 수 있게 되는 원리다. 하지만 이렇게 얻은 DNA는 단백질이나 기타 물질들이 정확하게 분리되어 있지 않으므로 다시 정제 과정을 거쳐야 PCR실험 등에 이를 적용할 수 있다.

실험 목표

DNA 추출의 원리를 이해하고, 동물 조직에서 DNA를 추출하여 전기영동을 통해 DNA가 추출되었음을 확인한다.

실험 재료

Animal tissue, DNA extraction buffer (10 mM NaCl, 20 mM Tris-HCl, pH 8.0, 1 mM EDTA, RNAase), 10% sodium dodecyl sulfate (SDS), proteinase K, Phenol equilibrated with 0.1 M Tris-HCl(pH 8.0), Chlorform: isoamyl alcohol(24:1), 3 M sodium acetate (pH 5.2), 100% Ethanol, 70% Ethanol, TE buffer(10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA), 1X TAE buffer(40 mM Tris-acetate, 1 mM EDTA), D.W., 막자사발 1개, 비커, microtube, pipette

실험 방법

1. DNA 추출

1) 동물조직을 잘게 자른 후 막자사발에 넣고서 액체질소를 부어서 조직을 가루가 되도록 갈아서 분쇄한다.
2) 막자사발 안의 분쇄된 조직에 DNA extraction buffer를 950㎕으로 넣어준 후 잘 섞어준 후 튜브에 옮겨 담는다.
3) 이 혼합물에 10% SDS 용액 50㎕를 넣고 잘 섞는다.
4) proteinase K solution 용액 30㎕를 첨가한 후, 뚜껑을 단단히 닫고 위아래로 힘차게 힘들어서 섞어준다.
5) 37℃의 항온수조 안에 적어도 6시간에서 최대 2일정도 배양시키거나 또는 55℃에서 3시간동안 둔다. 배양하는 동안 튜브를 몇 번 천천히 뒤집어 놓는다. 완전히 용해된 샘플은 점성이 있고 거의 투명하다.
6) 같은 농도의 페놀 일정한 양을 추가하고 뚜껑을 단단히 닫은 후 1분 동안 뒤집어가

면서 천천히 섞는다.

7) 상온에서 원심분리기에서 1700 g로 튜브를 돌린다.

8) 플라스틱 피펫으로 위의 액체층을 빨아드려 다른 새 튜브에 옮겨 라벨링한다. 이것은 페놀을 옮기거나 두 층 사이로부터의 흰 단백질같은 물질을 옮기지 않도록 주의한다.

9) 페놀 추출(단계6)을 다시 한번 반복한다.

10) 페놀대신 같은 농도의 클로로포름 일정한 양을 넣어 추출하고 원심분리 후 상층액만 모아서 다른 새 튜브에 옮겨 담는다.

11) 클로로포름 추출을 반복한다.

12) 액체상태의 DNA 용액에 0.1배 용량의 3M sodium acetate (pH 5.2)을 넣고, 돌려가며 잘 섞는다.

13) 튜브에 2배 용량의 chilled 100% Ethanol을 넣는다. 뚜껑을 단단히 닫은 후 1분 동안 뒤집어 가며 천천히 섞는다. 그러면 하얀실 같은 모양이 만들어진다.

14) 소독한 막대를 이용하여 석출된 DNA를 감아낸다. (만약 석출된 것이 보이지 않는다면, 마이크로퍼지로 최고 속도로 10분 동안하고, 알갱이를 70% 에탄올에 씻어낸다.)

15) 1 mL 의 차거운 70% 에탄올로 감아낸 DNA 전체를 찍어가면서 씻어낸다.

16) 석출물이 투명해질 때까지 건조시킨다. 대부분 약 5~10분 정도이다.

17) TE 버퍼의 적당한 볼륨에서 석출물을 녹인다. (대개 약 100~500 μL; 1 μg/μL 의 농도의 DNA을 기준)

18) TE 버퍼에서 DNA가 녹게 하기 위해서 50°C에서 4시간동안 놓고, 배양하는 동안 규칙적으로 천천히 흔들어준다. DNA가 적당히 녹을 수 있게 만들지 못하면 용액 안의 DNA 가 불규칙한 분포하게 한다.

2. 추출한 DNA 확인

1) 추출된 DNA의 확인은 분광광도계 등 흡광도를 통해서 계산한다.

2) 260nm에서 측정하여 결정.

3) 흡광도를 이용한 DNA 농도 계산법

① 이중가닥 DNA (Doubled-stranded DNA, dsDNA)

: 260nm에서 흡광도(A) 값이 1일 경우, 50 μg/ml H_2O

② 단일가닥 DNA (single-stranded DNA, ssDNA)

: 260nm에서 흡광도(A) 값이 1일 경우, 33 μg/ml H_2O

참고문헌

1. 생명과학교재편찬위원회. 생명과학개론, 2009. 지코사이언스.
2. 일반생물학실험(2009), 한양대학교 분자생명과학부, 한양대학교 출판부.
3. QIAamp DNA Mini and Blood Mini Handbook.
4. 일반생물학실험(2008), 권혁빈 외, 지코사이언스.

6. DNA Sequencing 및 DNA Sequencing 결과 확인

1) Chromas 프로그램을 이용하는 법 => 분석할 자료가 소량일 경우

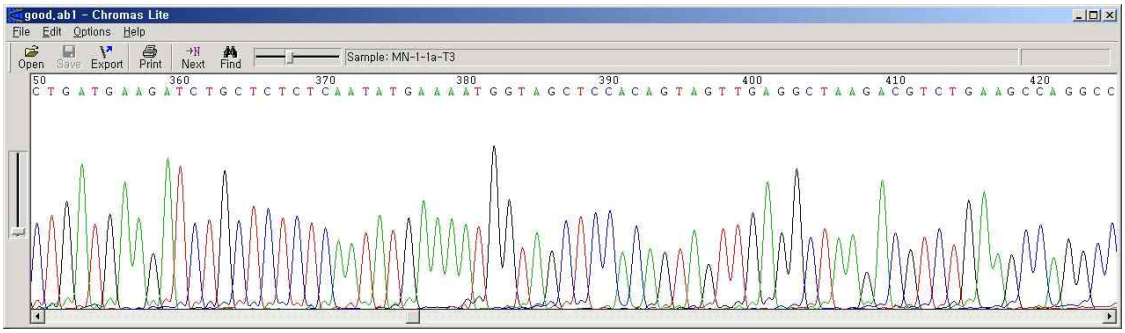


그림 13. Chromas 를 통해 ab1 파일을 열어본 모습

Chromas 프로그램을 이용하여 sequence file 을 확인 및 수정할 수 있으며, FASTA 형식으로 만들어 일반 editor에서 볼 수 있다. 또한 출력을 통해 chromatogram 을 확인 할 수도 있다.

장단점 : 눈으로 직접 모든 데이터를 확인하므로 신뢰성 면에서는 매우 좋다. 하지만 대량의 데이터를 처리하기에는 부적합 하여 자동화할 수 없다.

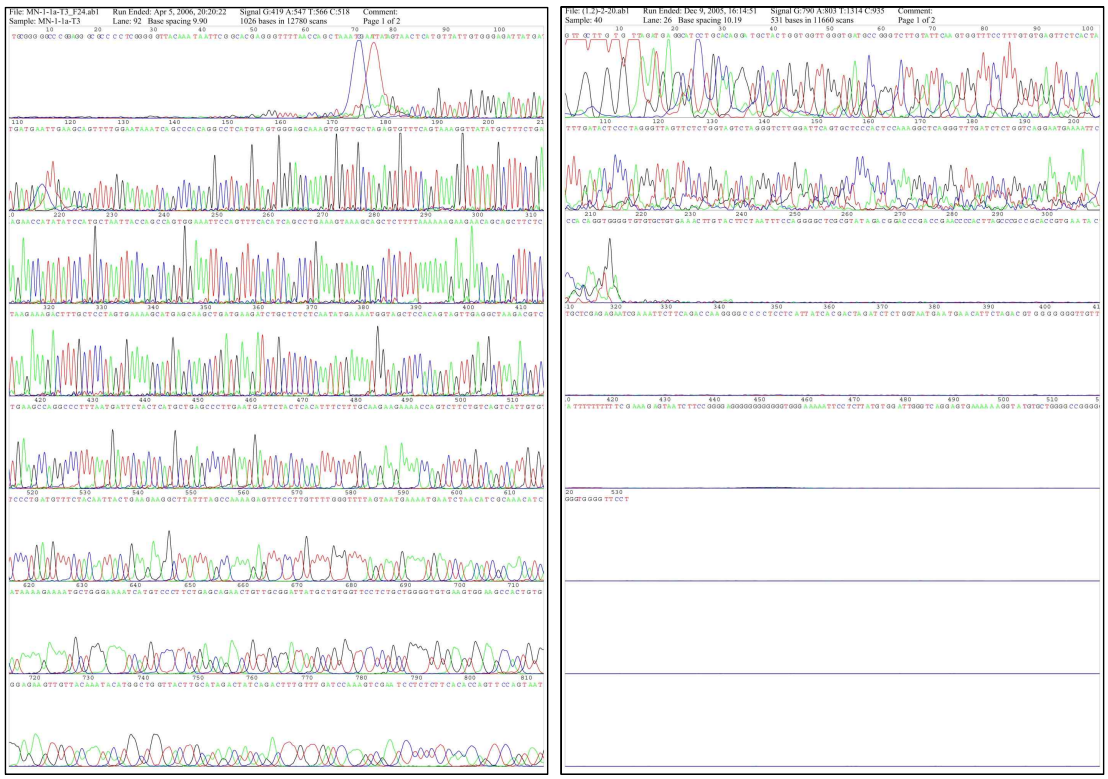


그림 14. Chromas 를 통해 출력을 한 모습 (좌 ; 좋은샘플결과, 우;나쁜샘플결과)

2) Phred 프로그램을 이용하는 법 => 분석할 자료가 대량일 경우

```
>good.ab1   1054      0   1054  ABI
TCGCGGCCCCCGCGGGCGCCCCCTGGGGGATCCAATAATTCGGCACGAGG
GTTTTAACCAGCTAAACCTTTATGTAACTCATGTTATTGTGGGAGATTAT
GATGATGACTTGAAGCAGTTTTGGAATAAATCAGCCCACAGGCCTCATGT
AGTGGGAGCAAAGTGGTTGCTAGAGTGTTTCAGTAAAGGTTATATGCTTT
CTGAAGAACCATATATCCATGCTAATTACCAGCCAGTGGAAATTCCAGTT
TCACATCAGCCTGAAAGTAAAGCAGCTCTTTTAAAAAAGAAGAACAGCAG
CTTCTCTAAGAAAGACTTTGCTCCTAGTGAAAAGCATGAGCAAGCTGATG
AAGATCTGCTCTCTAATATGAAAATGGTAGCTCCACAGTAGTTGAGGCT
AAGACGTCTGAAGCCAGGCCCTTTAATGATTCTACTCATGCTGAGCCCTT
GAATGATTCTACTCACATTTCTTTGCAAGAAGAAAACCAGTCTTCTGTCA
GTCATTGTGTCCCTGATGTTTCTACAATTACTGAAGAAGGCTTATTTAGC
CAAAAGAGTTTCCTTGTTTTGGGTTTTAGTAATGAAAATGAATCTAACAT
CGCANACATCATAAAAGAAAATGCTGGGAAAATCATGTCCCTTCTGAGCA
GAACTGTTGCGGATTATGCTGTGGTTCCTCTGCTGGGGTGTGAAGTGGAA
GCCACTGTGGGGAGAAGTTGTTACAAATACATGGCTGGTTACTTGCATAG
ACTATCAGACTTTGTTTGATCCAAAGTCGAATCCTCTCTTCACACCAGTT
CCAGTAATGACAGGAATGACTCCTTTTAGAGGATTGTGTTATTTCATTTA
GCCAGTGTGCTGGAGCAGAAAAAGAGTCTTTAACATCCCTAGCAAACCCT
CCTTGGAACAAGTGTTCAAGAAATACTTTTGGTCCCCAAATCCAATGGCA
AAAAAAAGGGCAGGGTTTGCCCAGTAACCTCCCTCTTAAACACTGAAAAA
AAACACTGGGGGGGGGCCCCCAAAAATATTAAAAACCTGGCCAAAAAAAA
AAAG
```

```
>bad.ab1    969      0    969  ABI
TTTTGTTTGTTGTTTAGATGAGGCAGTCTGCACAGGGTGCTACTGGTGGT
TGGGTGATGCCGGGTCTTGTATTCAAGTGGTTTCCTTTGTGTGAGTTCTC
ACTATTTGATACTCCCTAGGGTTAGTTCTCTGGTAGTCTAGGGTCTTGGA
GTCAGTGCTCCCACTCCAAAGGCTCAGGGTTTGATCTCTGGTCAGGAATG
AAAATTCGAGAAGTGAGNTNNGTNTGTGNNNNTTNTAANNNTNTTTTNN
NNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTNTTTT
TTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTNTTTTTTTTTTTTTTT
TTTTTTTTNTTTTTTTTTTTTTTTNTTTTTTNTTTTTTNTNTTTTNT
TTTTTTTTTTTTTTTTTTNNTTTTTNTTTTTTNTTTTTTTTNTTTTTT
TTTTTTTTTTTTTTTTTTTTTTTTNNTTNTTTNNNTNNTTTTTTTTTT
TTTTNTTTTTTTTTTTTTTTTTTTNTNNTTTTTTTTTTTTTTTTTTTT
TTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTNTNNTTTTNTTTTT
TTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTNTTTTTTTTT
NNNNTTTTTNNTNTNTTNTTTTTNTTTTTTNTNTTNTNTNTTNNNTTT
NTTNNNNNTTTTNTNTTTTTTTTTTTNTTNTNTNTNNTTTNNTTTTN
TTTTTTTTTTTTTTTNTNTTTNNNTTTTTTNTTTTNNNTTNTNNTT
TTTTTTNTTTNTTTNTTTNNTTTTTTNTTTTTTTTTTTTTNTNTNTNNTT
TTTTTTNNTTTTNTNTTNTNNTNNTTNNNNNNNNNNNNNNNNNNNN
NTTNNNNTNNTNTTTTTTTTTTNTNNTNNNTTNNNTNNNTTTNN
NNTTTTTTTTTTNTTTTTT
```

그림 15. Phred 를 통해 base celling을 한 모습 (좌 ; 좋은샘플결과, 우;나쁜샘플결과)

장단점 : 분석하여야 할 파일이 많은 경우 파일 하나하나 열어 보기가 매우 어려우므로, Phred 라는 소프트웨어를 이용하여 대량으로 염기서열을 분석 할 수가 있다. 이 경우 각 염기서열의 peak 가 점수화 되어 나타나며 점수에 따라 자동으로 trimmimg 을 할 수가 있다. 단 이 프로그램은 LINUX 에서만 돌아간다. 하지만 완전 자동화 할 수 있어 데이터가 대량화 될수록 유용하다.

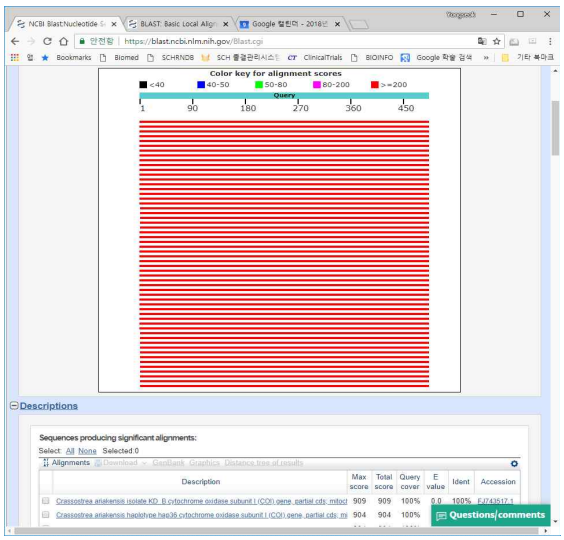
```
>good.ab1   1054      0   1054  ABI
7 9 9 7 7 9 9 8 8 13 12 9 7 6 11 10 9 9 9 10 7 7 8
8 9 10 13 10 13 13 11 6 6 9 9 9 10 14 14 23 17 20
13 13 13 19 27 30 47 56 56 56 46 44 40 40 29 29 18
18 21 23 21 17 10 7 7 15 15 15 8 8 8 10 10 10 19 24
32 40 40 40 40 46 40 40 40 40 51 51 51 51 51 51 40
40 46 46 46 40 40 40 48 48 48 21 19 12 12 18 18 15
18 25 21 34 34 40 44 56 56 44 51 51 51 51 56 56 56
51 45 45 40 40 40 45 46 51 51 45 45 45 56 43 51 56
56 56 56 48 44 44 44 47 44 56 56 56 56 56 56 56 56
56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 51 46
46 42 42 42 56 56 56 56 56 56 56 56 56 56 56 56 56 56
56 56 56 56 56 44 44 44 47 47 46 46 51 51 51 51 56
56 51 46 43 42 42 44 43 51 51 51 51 51 51 43 44 56
56 56 56 56 56 56 50 50 50 56 56 56 56 56 56 56 56
56 56 56 56 56 56 56 56 44 44 44 50 50 56 56 56
56 56 56 56 56 56 56 50 50 47 47 50 50 56 56 56 56
56 56 56 50 50 44 44 44 44 44 44 50 50 56 56 56 50
47 47 56 56 56 56 56 50 50 44 44 44 50 42 42 42 43
43 51 56 56 56 56 56 56 56 56 56 50 56 56 56 56 56
56 56 44 43 43 42 42 42 56 56 56 56 42 42 42 42 42
42 50 50 50 56 56 56 56 56 56 56 56 51 51 51 51 51
43 56 56 56 56 56 56 56 56 56 56 56 50 50 50 56 56
56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 50 50
44 44 44 56 56 56 56 50 50 50 50 50 50 56 56 50 50
50 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56
56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56
56 51 51 51 51 51 51 56 56 56 56 56 56 56 56 56 56
56 56 56 56 56 56 44 44 44 44 44 44 44 44 48 48 56
56 56 56 56 56 56 56 56 56 56 56 56 56 56 42 42 56 46
46 42 43 42 42 42 42 42 56 56 51 51 51 51 51 56 56
56 56 43 43 42 40 40 40 56 48 40 40 40 40 33 28 28
37 25 34 37 40 40 40 40 48 48 40 40 40 37 25 25 48
31 40 40 40 40 40 40 47 56 56 56 47 47 56 56 56 42
40 40 40 56 48 48 40 40 40 40 48 40 37 34 34 34 40
40 40 40 40 40 40 37 29 29 29 29 29 26 20 17 4 0 4
21 24 40 40 40 40 40 32 32 26 24 16 22 48 40 40 40
31 31 34 25 29 29 26 25 25 29 25 29 40 40 29 34 34
37 34 34 29 31 31 24 18 18 15 12 12 17 25 40 40 40
40 40 36 34 37 34 37 37 37 32 40 31 31 26 21 17 18
18 19 22 21 27 27 31 31 40 40 40 37 32 28 28 27 27
21 20 19 21 18 15 15 15 11 15 12 14 22 25 31 24 17
12 15 11 9 8 8 8 8 8 10 10 13 16 16 21 23 15 20 16
19 20 18 18 24 22 20 27 27 25 25 25 22 27 27 25 25
25 28 14 14 8 8 8 14 16 19 11 11 8 8 7 11 9 18 10
```

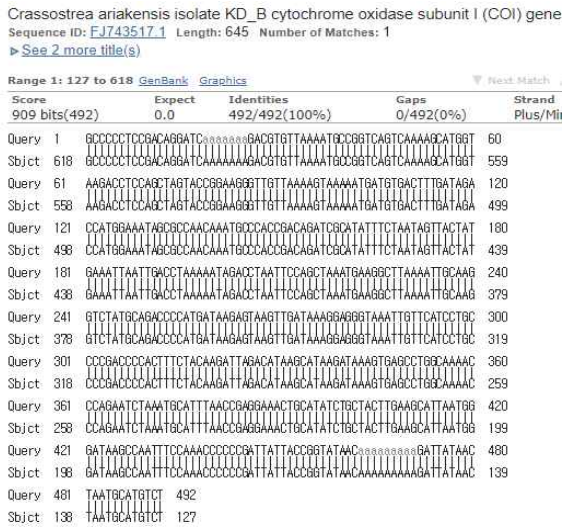
```
>bad.ab1    969      0    969  ABI
8 8 9 9 8 8 6 7 8 8 15 14 14 12 8 7 7 13 19 21 16
10 7 7 7 6 6 13 14 19 11 11 11 9 8 8 12 17 26 40 32
32 29 29 29 31 35 35 35 34 35 35 29 25 25 29 29 25
20 18 18 13 22 21 24 25 31 25 28 28 30 30 31 29 24
24 21 28 28 30 30 38 42 42 42 42 44 34 30 30 33 33
32 32 34 30 33 28 28 24 24 24 24 30 27 33 22 22 22
27 27 24 24 24 33 33 35 35 38 35 35 33 33 24 22 22
30 25 29 33 30 31 30 33 29 28 28 33 30 35 33 30 34
28 28 30 30 24 24 24 17 22 17 19 19 13 15 14 14 14
33 14 15 15 20 19 25 25 29 29 33 28 22 22 22 24 20
14 12 18 20 23 18 18 14 16 14 15 15 16 20 20 22 20
17 18 20 23 20 21 24 24 16 14 7 7 7 7 8 11 11 16 10
8 8 8 4 0 4 0 0 4 4 0 4 15 17 4 0 0 0 0 4 4 0 4 14
4 0 0 0 4 0 4 10 19 10 4 0 0 0 0 0 0 4 8 8 9 4 4 4
6 6 6 6 4 4 4 4 4 4 6 4 4 4 4 4 6 6 4 0 4 4 4 6 9
7 4 4 4 4 4 7 10 4 0 4 4 4 7 4 4 4 9 4 6 6 6 6 6 6
6 6 4 4 4 4 4 6 4 0 4 6 6 6 6 6 6 6 4 4 4 4 4 0 4
6 6 6 6 6 4 4 4 4 6 6 6 9 9 9 9 9 8 8 6 4 4 4 0 4
8 8 8 8 9 9 7 7 7 8 8 9 4 0 4 9 9 9 4 0 4 8 8 6 6
7 6 6 4 0 4 4 0 4 6 6 4 0 4 4 4 4 9 7 8 11 11 9 9
9 15 8 10 7 6 8 4 0 0 4 4 6 9 9 4 0 4 4 4 4 8 4 0
4 9 9 9 4 4 4 4 4 0 4 4 4 4 4 4 4 4 4 4 9 9 8 6
6 9 7 6 6 6 10 11 6 6 6 8 8 4 6 4 0 0 4 4 0 4 15 9
4 0 0 0 4 0 0 4 9 9 9 15 14 8 6 6 7 7 12 9 9 4 0 4
7 6 9 16 10 9 16 11 21 21 11 11 10 19 8 6 4 0 4 4
0 0 4 9 15 6 6 8 6 6 6 7 6 6 8 19 9 9 9 9 9 9 9 11
11 13 18 10 21 8 6 6 6 6 8 9 6 6 12 13 9 9 6 4 0 4
9 6 8 8 9 9 9 9 9 9 9 9 4 0 4 0 0 4 6 9 4 0 4 6 6
16 9 6 6 6 6 8 10 8 8 6 6 9 9 9 9 8 9 6 4 0 4 9 11
21 14 12 12 9 9 9 19 19 13 13 10 8 6 6 6 6 4 0 4 6
9 16 12 14 11 11 4 0 0 0 0 4 19 10 10 9 4 0 0 4 0
4 0 4 4 0 4 9 9 7 4 0 4 8 6 9 6 6 4 0 4 0 4 4 4 0
4 0 4 6 4 0 0 0 4 8 4 0 4 6 4 0 0 0 0 0 4 9 15 9 4
0 4 6 6 6 7 6 6 6 6 6 6 4 0 4 10 4 0 4 4 0 4 0 0 4
6 9 4 0 0 4 8 6 6 4 0 4 19 13 6 6 6 6 6 6 13 7 7 8
8 4 0 4 4 0 4 9 9 4 0 0 0 4 8 8 9 9 8 4 0 4 6 6 6
4 0 0 0 4 4 0 4 0 0 4 9 9 9 9 10 10 9 9 4 0 4 9 9
4 0 4 6 9 4 0 0 4 6 6 9 6 9 4 0 4 4 4 6 6 6 6 6 6
8 6 7 4 0 4 0 4 4 0 0 4 6 6 8 6 6 6 4 0 0 0 4 8 6
6 4 0 4 6 4 0 4 4 0 0 4 4 0 0 4 6 4 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 4 6 4 0 0 0 0 4 0 0 4
4 0 4 6 9 6 6 6 6 6 9 6 4 0 4 0 0 4 4 0 0 0 4 9 4
0 0 0 4 0 0 0 4 11 11 4 0 0 0 0 4 6 9 9 8 9 9 7 7
4 0 4 8 9 9 11 8
```

그림 16. Phred 를 통해 base calling을 하여 quality를 점수화 한 모습
(좌 ; 좋은샘플결과, 우;나쁜샘플결과)

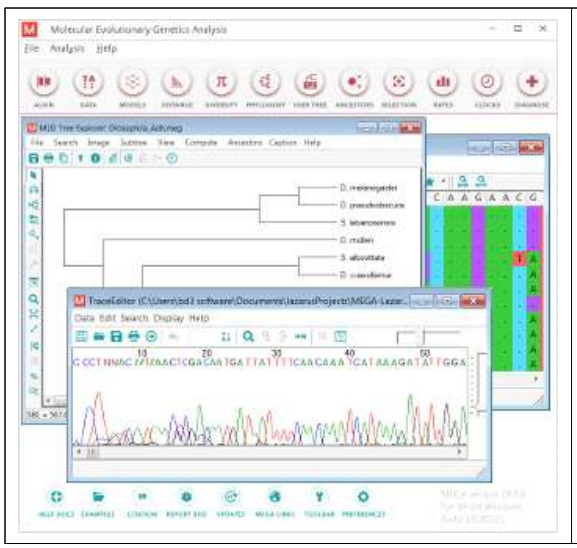
7. BLAST 수행 결과에 의한 유사서열의 확보

```
>1_COI-HCO2198_D01.ab1      492        0      492   ABI trimmed
GCCCCCTCCGACAGGATCAAAAAAAGACGTGTTAAAATGCCGGTCAGTCA
AAAGCATGGTAAGACCTCCAGCTAGTACCGGAAGGGTTGTTAAAAGTAAA
AATGATGTGACTTTGATAGACCATGGAAATAGCGCCAACAAATGCCCACC
GACAGATCGCATATTTCTAATAGTTACTATGAAATTAATTGACCTAAAAA
TAGACCTAATTCCAGCTAAATGAAGGCTTAAAATTGCAAGGTCTATGCAG
ACCCCATGATAAGAGTAAGTTGATAAAGGAGGGTAAATTGTTCATCCTGC
CCCGACCCCACTTTCTACAAGATTAGACATAAGCATAAGATAAAGTGAGC
CTGGCAAAACCCAGAATCTAAATGCATTTAACCGAGGAAACTGCATATCT
GCTACTTGAAGCATTAATGGGATAAGCCAATTTCCAAACCCCCCGATTAT
TACCGGTATAACAAAAAAAAAGATTATAACTAATGCATGTCT
```





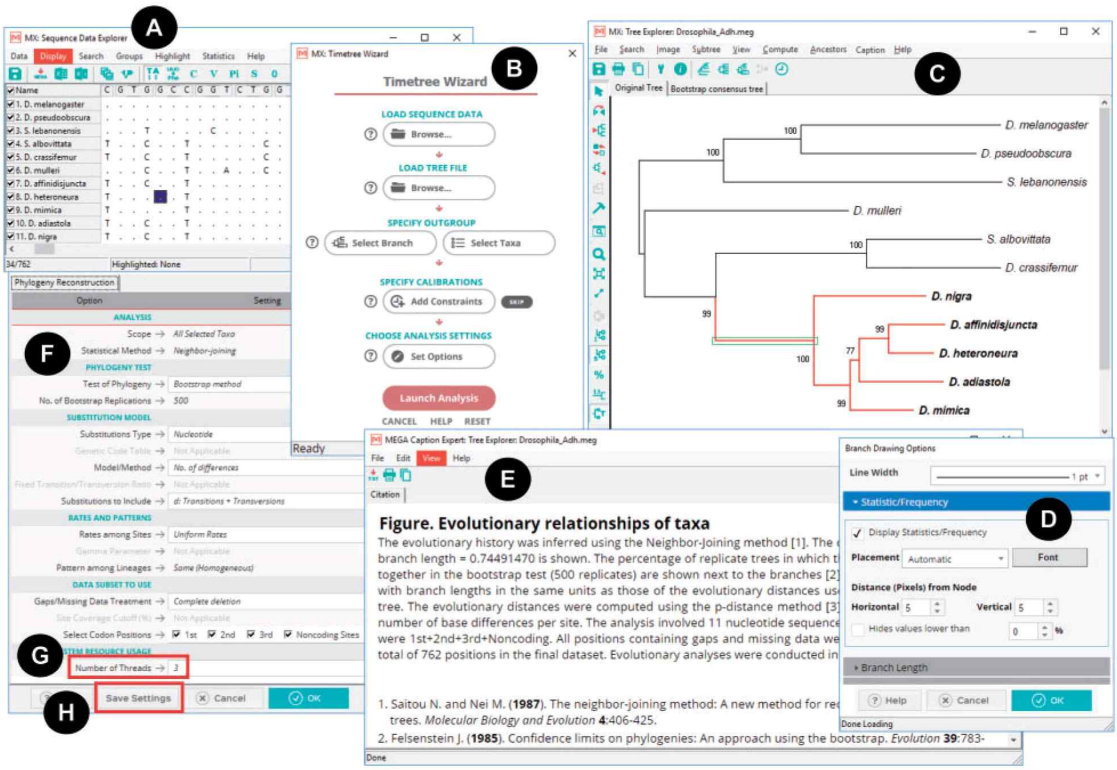
8. MEGA Software 사용법



https://www.megasoftware.net 웹싸이트에서 MEGA-X 다운로드 및 설치

Windows version 으로 본인 사용컴퓨터 OS가 32 비트인지 64 비트인지 확인 필요

- Multiple sequence Alignment 수행 후 Molecular Phylogenetic Tree 작성



(A) Sequence Data Explorer
(B) Timetree Wizard
(C) Tree Explorer
(D) Options dialogs
(E) Caption Expert
(F) The redesigned “Analysis Preferences” dialog box
(G) the number of threads to employ
(H) a“Save Settings” button for generating an “.mao” file for use with MEGA-CC for high throughput analysis.