

차세대 염기서열 결정장치 나노포어를 이용한 유전체 결정

5GGCAATAACGTTTATGTTGGTTTCATGGTTTGGTCTAACTTTACC



DATA FOR ILLUSTRATIVE PURPOSES ONLY

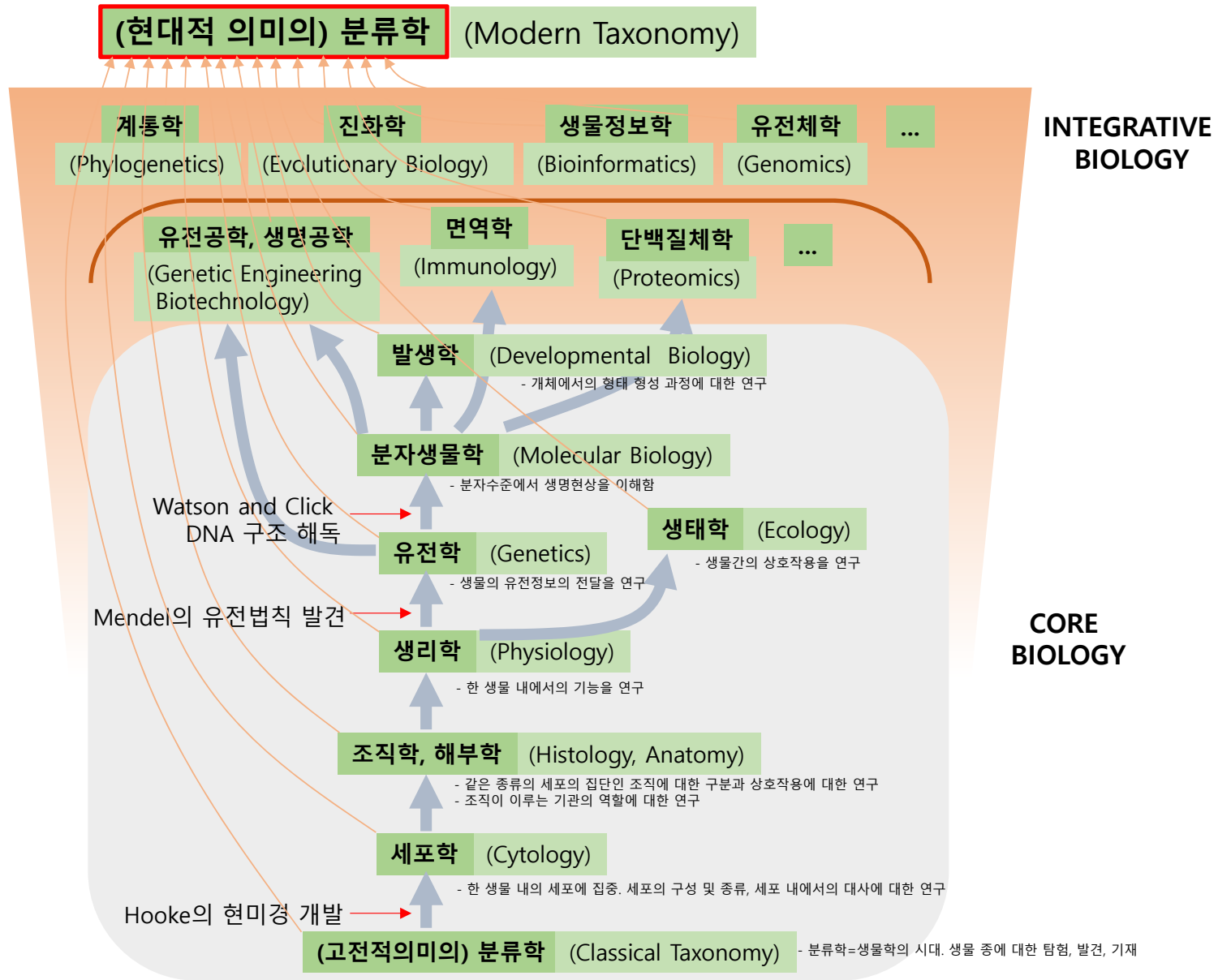
성신여자대학교 바이오생명공학과

전공기반 비교과 프로그램

**현대 생물학과 그 응용분야에 있어서
유전체 결정은
모든 다른 분야 연구와 기술개발의
기초자료가 됨!**

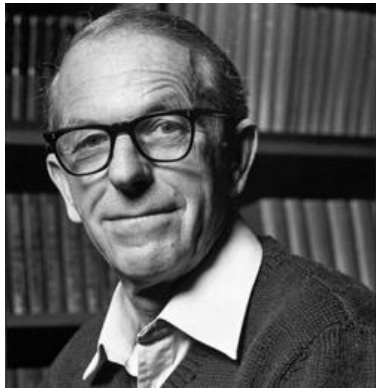
- 인간 유전체: 진단, 의약품개발, 맞춤형 의료, 단백질 기능연구...
- 식물, 동물, 미생물 유전체: 천연물의약품, 검역, 분류진화 연구...

생물학의 발달 과정



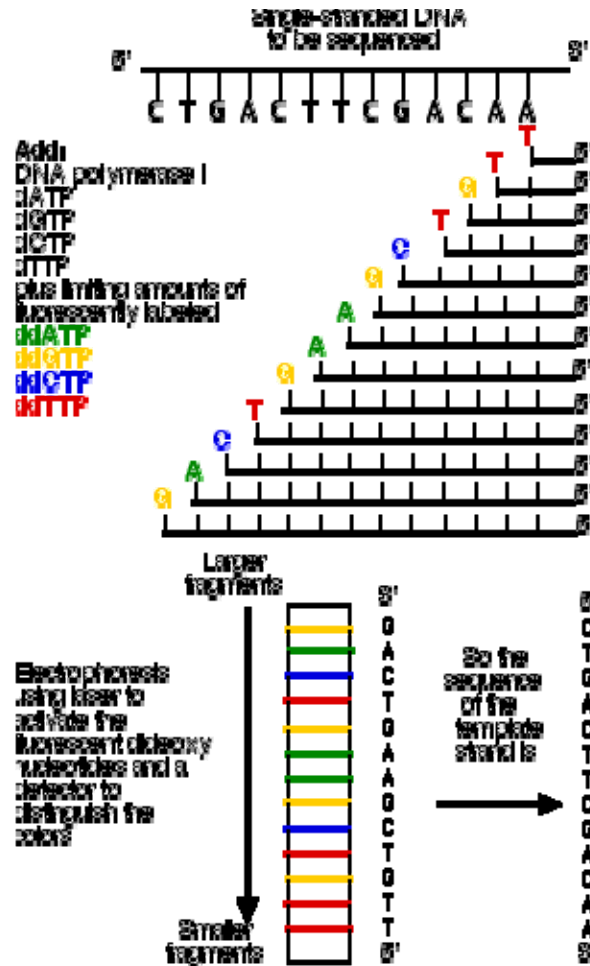
I. Introduction: 염기서열 결정의 진화

The first generation sequencing: Sanger sequencing



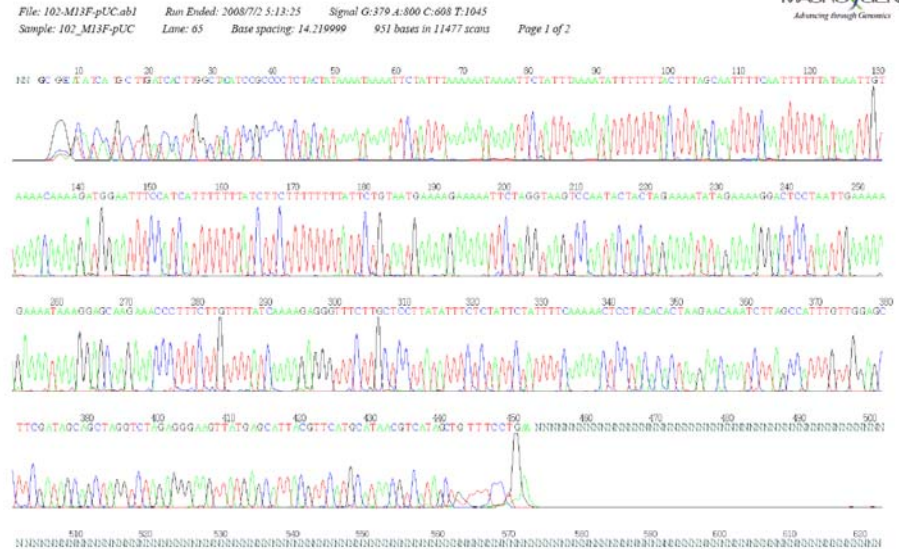
Frederick Sanger (영국)

- 두 번의 노벨상 수상자
- "termination" method
- One-dye four lane system에서 four-dye one lane system으로 발전.



현재 가장 많이 쓰는 모델: ABI 3730 (Thermo Fisher)

- Applied Biosystems Co.
→ Thermo Fisher Co.
- Ca. 900 bp/capillary
- 384 capillary/run
- 900 X 384
= ca. 350 kbp



Next Generation Sequencing (차세대 염기서열 결정)

1) The second generation sequencing:

- 1) emulsion based clonal amplification (**emPCR**)의 기술,
- 2) DNA 분자가 합성될 때 형광을 발하는 염기서열 결정기술 (**pyrosequencing**)
- 3) 광섬유들을 평행하게 붙여 만든 **pico-titer plate**
등의 신기술을 이용하여 염기서열 결정 용량은 획기적으로 증가.

대표적 기업 / 기술:

Roche (454) / 454

Solexa / Illumina → MGI (Illumina 유사 기술에 의한 중국 기업/제품)

ABI / SOLiD

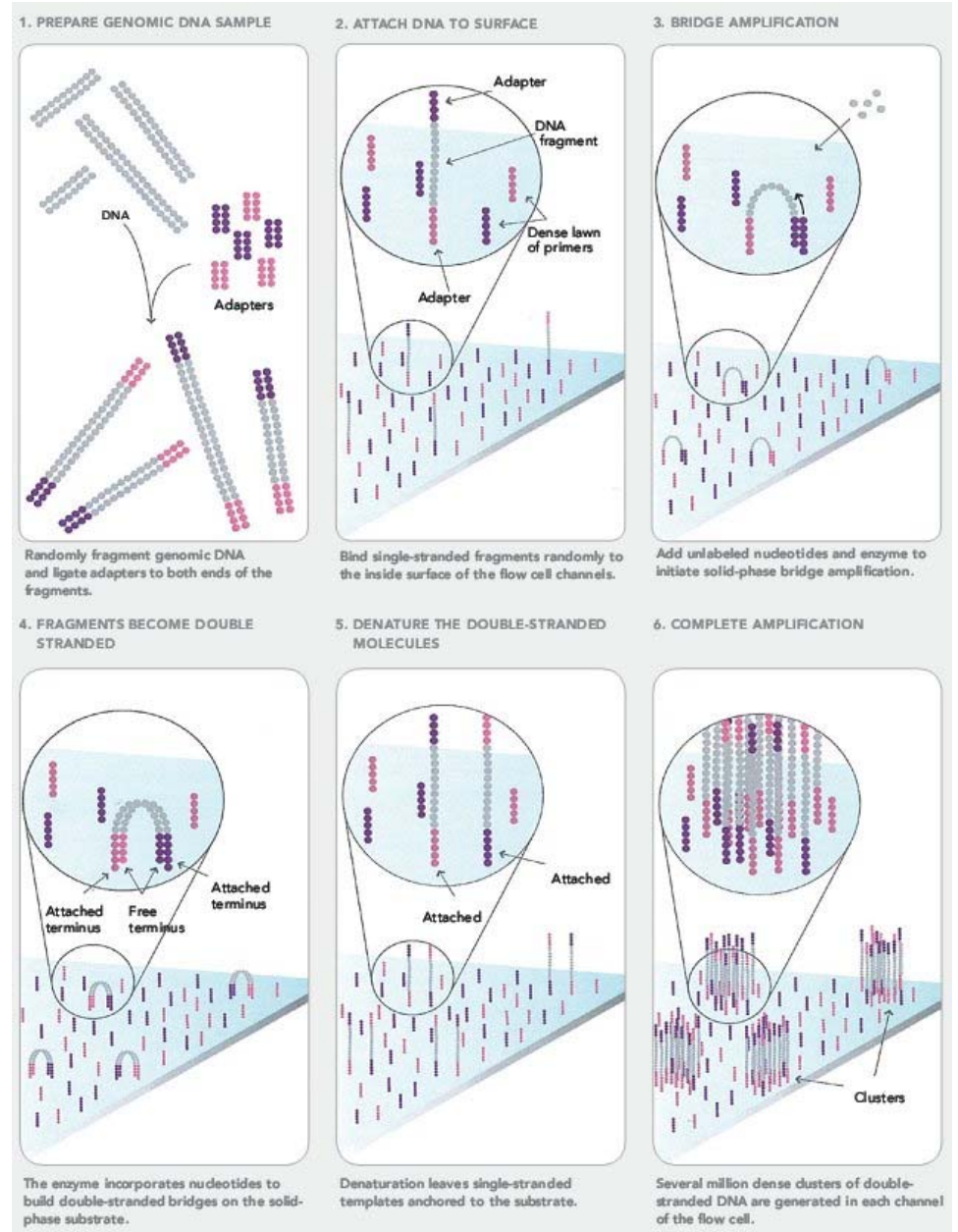
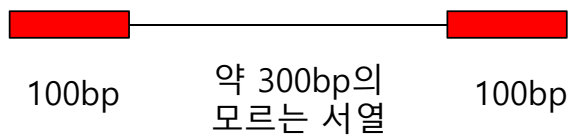
Hilicos / Ion Torrent

→ 대부분 없어지고, 현재 Illumina (and MGI)가 가장 널리 쓰임.

Solexa / Illumina Technology

Illumina를 이용한 결과의 특징:

- 한 가닥의 DNA로 부터 염기서열을 결정할 때 순방향과 역방향으로 각각 약 100~150 bp 정도 읽게 됨.
- 그러므로 DNA를 일정크기로 잘라 만든 조각이 500bp일 경우 좌우로 100bp씩을 얻게 되고,中间的의 300bp는 모르는 서열로 연결되게 됨.

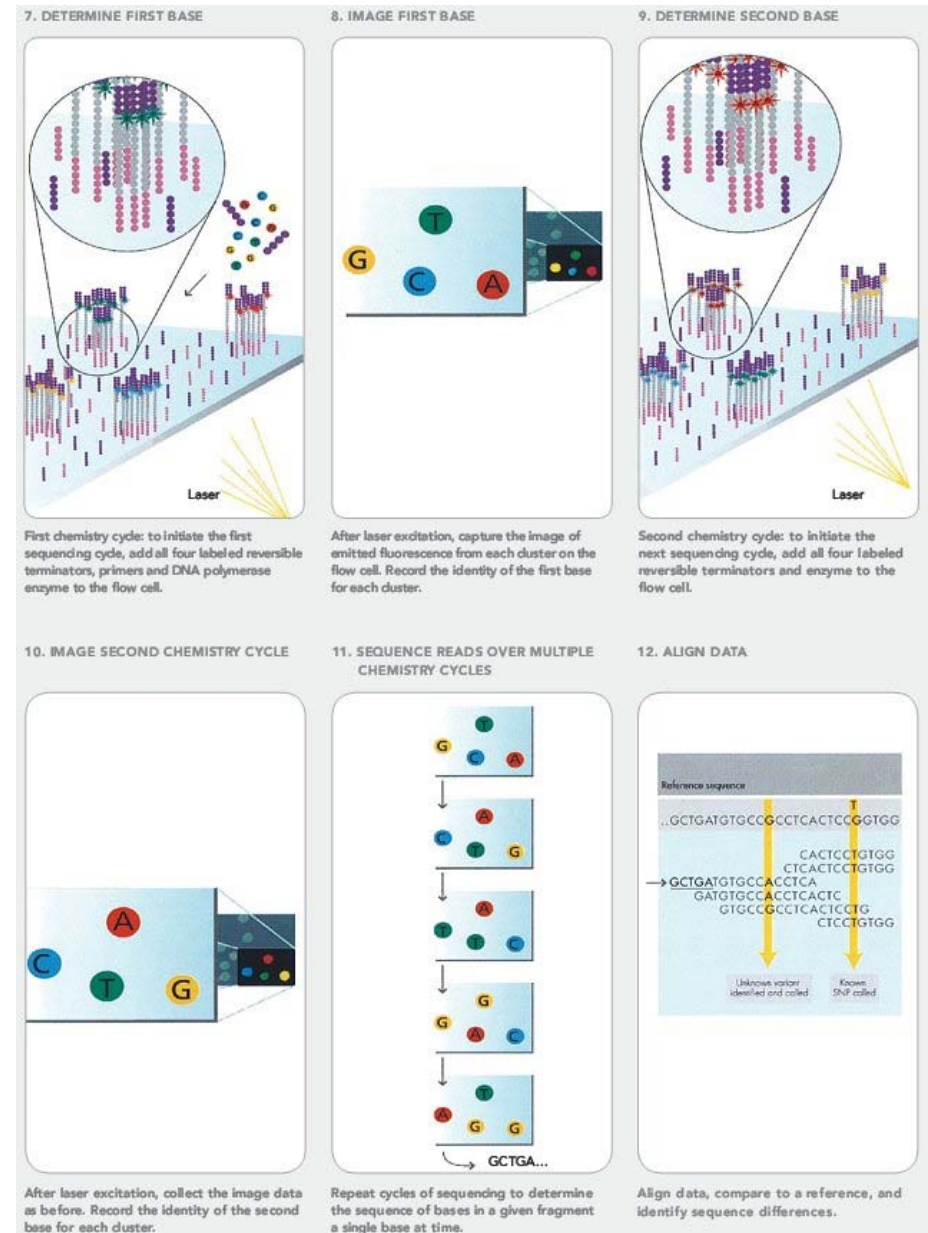


Solexa / Illumina Technology

<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

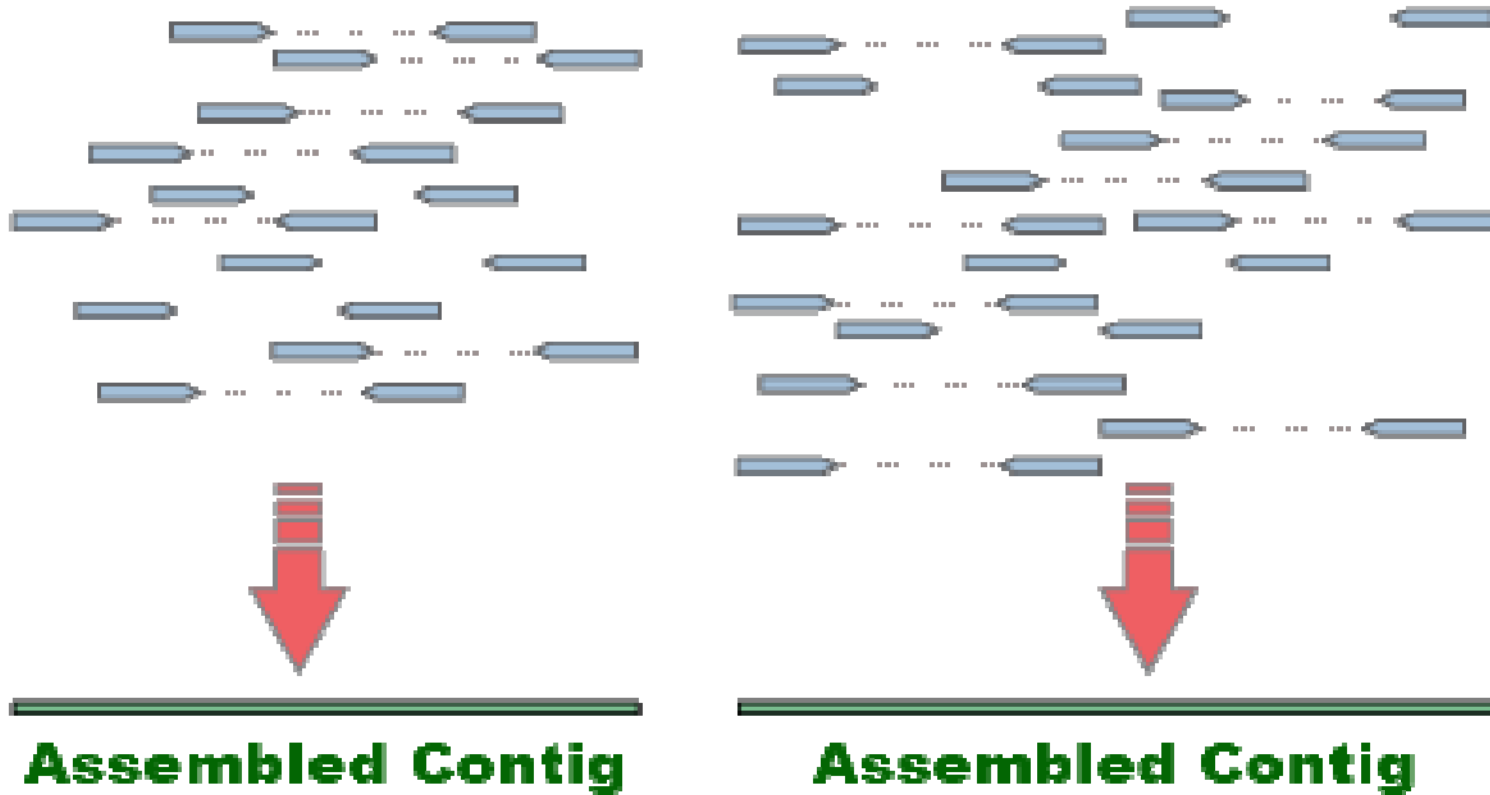
※ Illumina 기술에 의한 염기서열 결정은 “매우” 많은 염기서열을 한번에 얻을 수 있음. 그러므로 종종 많은 시료를 섞어서 염기서열을 결정하기도 한다. 이때 섞은 시료들을 구분하기 위하여 시료 각각을 구분하는 index sequence 를 각각의 시료에 붙임. 전체 염기서열 결정 후 index sequence로 시료들은 구분한 후 각각 정렬하여 결과를 얻게 됨.

※ **index sequence**: 여러 시료를 섞어 실험할 때 각각의 시료를 구분할 수 있는 짧은 염기서열. 실험의 첫 단계인 adaptor를 붙이는 과정에 index sequence를 삽입한다.



Illumina data 의 assemble 과정

Paired-End Reads



- Illumina sequence는 최근까지 유전체 결정에 있어서 주된 data로 사용되어 옴.
- Macrogen Co. service:
HiSeq X ten, HiSeq4000, NextSeq
- 진정한 인간유전체 결정 \$1,000 시대



HiSeq X System Performance Parameters	
Key Application	Large Whole-Genome Sequencing (human, plant, animal)
Output per Run	Dual flow cell: 1.6-1.8 Tb
Single Reads Passing Filter	Dual flow cell: 5.3-6 billion
Maximum Read Length	2 x 150 bp
Run Time	< 3 days
Quality	≥75% of bases above Q30 at 2 x 150 bp

하지만... 무엇이 문제인가?

→ 100~150 bp의 너무 짧은(short-read) 서열을 제공하고 있기 때문에, 이것을 이어 붙여 유전체를 조립하면 구조적인 에러를 발생할 확률이 높아짐.

1) Mate-pair sequencing에 의한 해결

2) **Long-read data**를 생산하는 것이 궁극적 해결 방법임.

Genome assembly의 과정 (Illumina)

- 1) 추출된 DNA를 적당한 크기로 자른다.
- 2) 잘린 DNA 절편들 중 약 500 bp 의 크기를 갖는 절편들만 정제한다.
- 3) Illumina **paired-end sequencing**에서는 잘린 절편의 양쪽 끝의 각각 약 100bp 정도의 데이터를 얻을 수 있다(reads).
- 4) Assembly program을 이용하여 각각의 read들을 정렬하여 **contig**들을 생성해 낸다.
- 5) 이와는 별도로 추출된 DNA를 보다 긴 길이의 절편으로 잘라 (1K, 2K, 5K...) 이들 절편의 양쪽 끝 100bp 의 염기서열들을 결정한다 (**mate-pair library**).
- 6) 일반 paired-end sequencing을 통해 생성된 contig들과 mate-pair 결과를 합쳐 assembly한다. 이를 통해 결과적으로 전체 유전체는 부분적으로 염기서열이 완성된 contig들이 mate-pair 에 의해 연결되어 많은 gap을 포함하지만 서로의 위치관계가 명확해진 긴 염기서열을 얻게 된다. 이를 **scaffold** 또는 **super contig**라 한다

a) Multiple copies of genome



b) Sheared random fragments



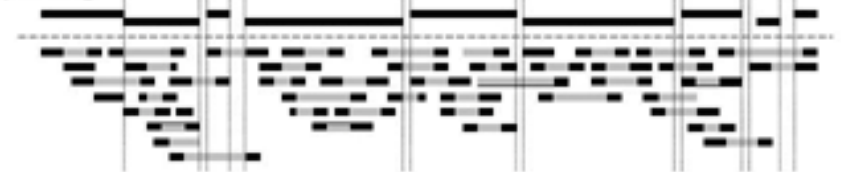
c) Size fractionated fragments



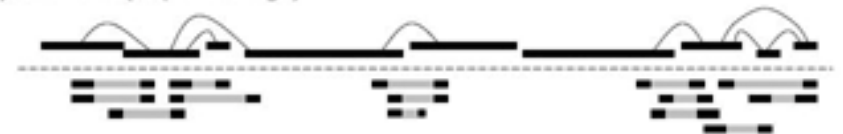
d) Reads



e) Contigs



f) Scaffolds(Super contigs)



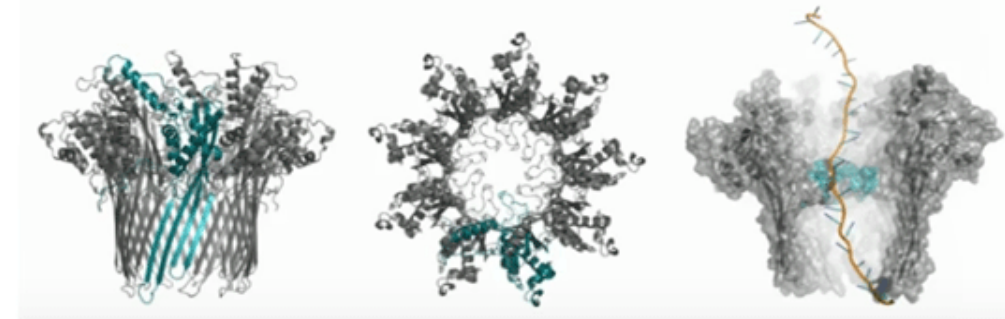
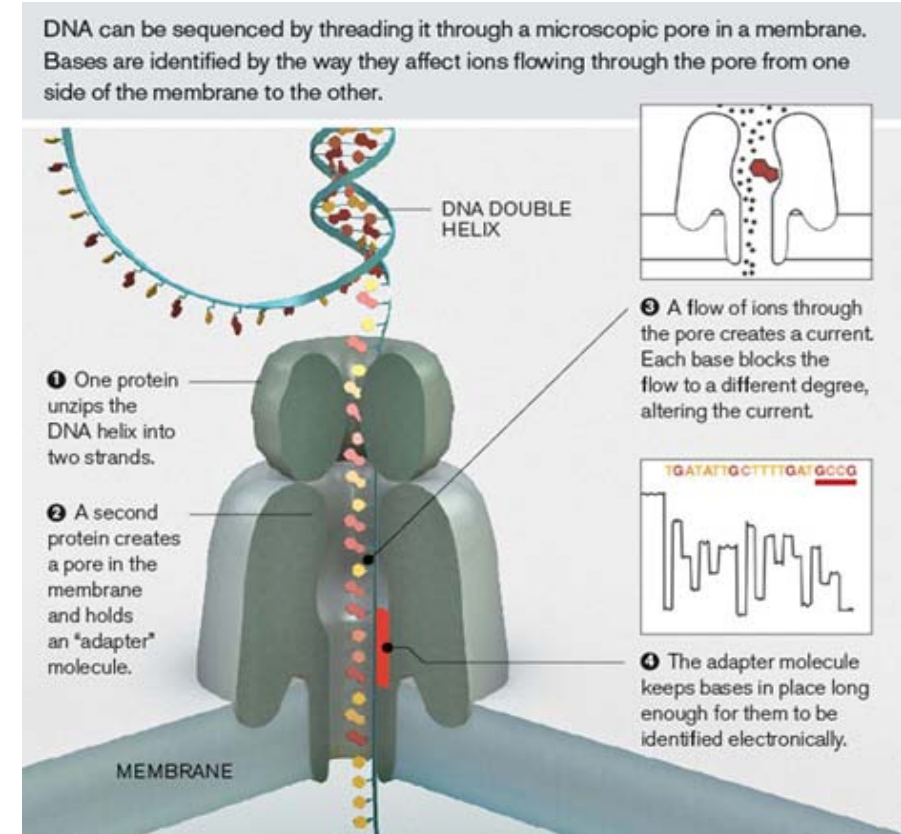
Next Generation Sequencing (차세대 염기서열 결정)

2) The Third generation sequencing:

- 100-150 bp 길이의 short-read를 생성해 내는 2세대 염기서열결정장치와는 달리 수십 kbp 에 달하는 long-read를 생성해 내는 기술
- 대표적 기업 / 기술:
 - 1) PacBio / PacBioHiFi
 - 2) Oxford Nanopore Technology (ONT) / Nanopore
- **PacBioHiFi** 주로 10~15 kbp (up to 25 kbp)의 제한된 길이의 대용량 데이터 생산하여 sequencing servic업체에 의해 운영되는 것에 반해, **Nanopore**는 연구실 단위의 실험이 가능하고, 추출된 DNA의 길이만큼 훨씬 긴 서열의 생성이 가능함.

ONT / Nanopore technology

- **Oxford Nanopore Technologies** (ONT): 2005년에 University of 로부터 설립.
- “Nanopore (나노포어)”라는 막단백질을 이용한 획기적인 NGS.
- 막단백질의 직경이 수 나노미터를 이루기 때문에 이런 이름이 붙여짐.
- 세포에서는 막에 위치하는 나노포어를 통해 이온들이 막 내외로 이동하는데, 이온이 지나가면서 막에 전류를 발생시키게 됨. 그런데 나노포어로 DNA나 RNA를 통과시키면, 이들이 이온의 흐름을 방해하여 여기서 발생하는 전류에도 변화가 생기게 되어 이런 막 단백질의 전류의 변화를 측정하는 것임.
- 즉, A, C, G, T 각 염기서열마다 이온의 흐름을 방해시키는 정도가 다르고 이에 의해 전류가 변화하는 정도도 달라지게 되므로, 이를 분석하여 서열을 파악할 수 있음.

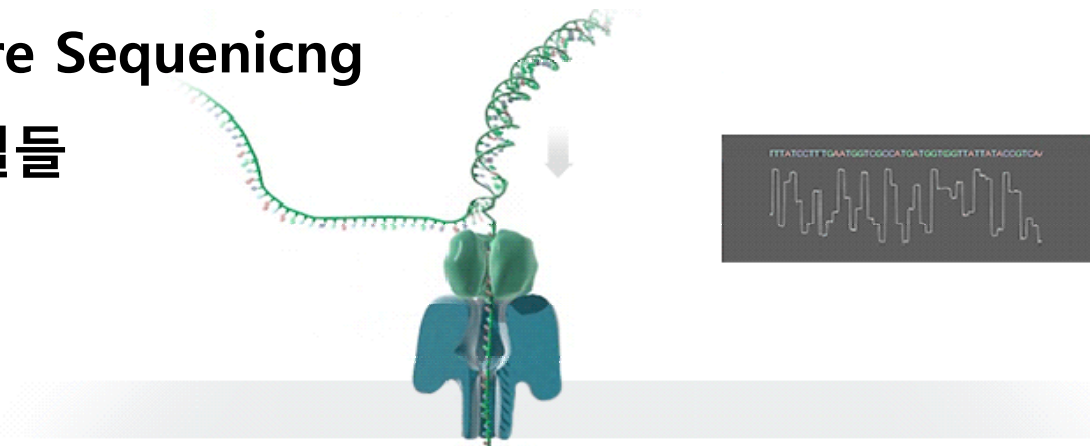


ONT 시퀀싱에 사용되는 나노포어인 CsgG의 구조

<https://www.youtube.com/watch?v=E9-Rm5AoZGw>

Nanopore Sequencing

기기 모델들



가장 작은 model인 MinION은 노트북에 연결하여 현장사용도 가능하다.



MinION
Mk1C

Available to pre-order

Complete sequencing, analysis, and viewing device

Up to 30 Gb data / flow cell 512 channels*




MinION
Mk1B

Commercially available

Portable, USB powered biological analysis

Up to 30 Gb data / flow cell 512 channels*




GridION
Mk1

Commercially available

Five flow cell capacity and integrated computing

Up to 150 Gb (5 x 30 Gb) data / device 5 x 512 channels*



PromethION
P24 P48


Commercially available

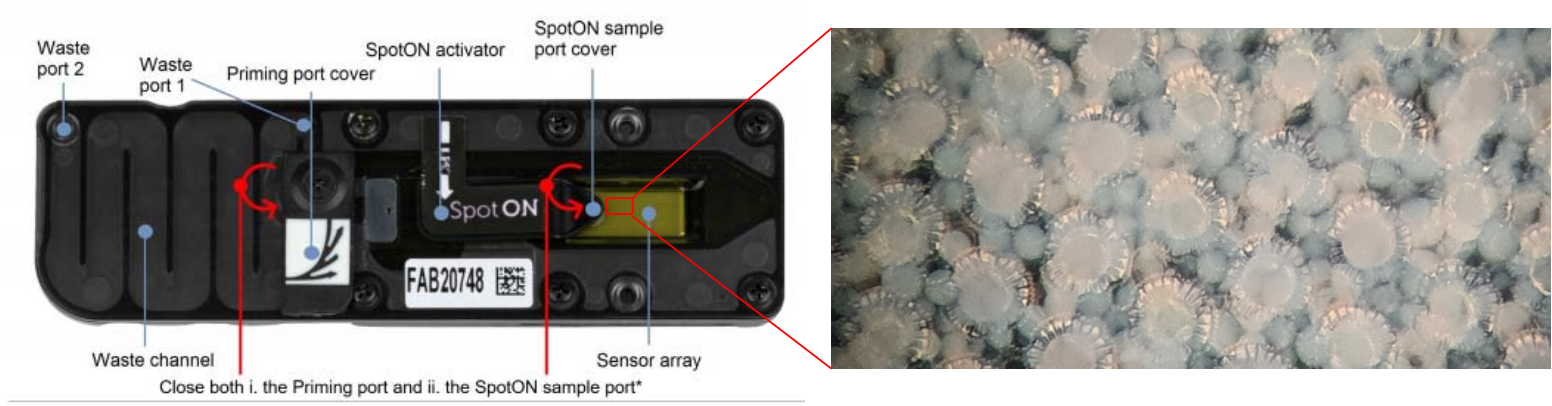
High-throughput, versatile benchtop system (P24 or P48)

P24	P48
>3.5 Tb data / device	>7 Tb data / device
24 x 3,000 channels*	48 x 3,000 channels*

Flongle

Adapter for MinION/GridION, supports smaller single-use flow cells. Up to 1.8 Gb currently; towards 3 Gb.





Molecular weight
T: 126.11 g/mol
A: 135.13 g/mol
G: 151.13 g/mol
C: 190.24 g/mol

Summary:

Comparison Sequencing Platforms

	Platform	Amplification Method	Sequencing Method	Detection Method	Average read length	Errors
○	Illumina (HiSeq, MiSeq etc)	bridge PCR	sequencing by synthesis	Light	100-200 bp	~ 0.1 %
×	Life Tech Ion Torrent / Proton	emulsion PCR	Ion semiconductor sequencing	pH	200-400 bp	~ 1 %
×	Roche 454	emulsion PCR	Pyrosequencing, cleavage of released pyrophosphate	light	700 bp	~1 % High error rate in homopolymer
×	Life Tech SOLiD	emulsion PCR	sequencing by ligation of hybridizing labeled oligos	light	100 bp	~ 0.1 %
○	Pacific Biosciences PacBio	No amplification, single-molecule sequencing	polymerase incorporating colored NTPs	light	4.5 – 8 kb	>10 %
○	Oxford Nanopore MinION	No amplification, single molecule nanopore sequencing	DNA molecule traverses pore	current	> 5.4 kb	< 5 %

Further reading, great lecture: Sequencing technology - Past, Present and Future, http://www.molgen.mpg.de/899148/OWS2013_NGS.pdf

× 현재 사장되어가는 기술들

II. 실험 스케줄

<u>주차별</u> 주제	강사	날자	시간	내용	<u>참여예상</u> <u>학생수</u>
High-molecular weight DNA 추출	김상태	7월 6일 (목)	9:00~12:00	- DNA 추출 이론 강의 - DNA 추출 1단계	10
			13:00~16:00	- DNA 추출 2단계	10
		7월 7일 (금)	9:00~12:00	- 전기영동 및 Qubit 정량 - <u>Nanopore</u> 이론 강의 - 길이측정 분석의뢰	10
Nanopore running	<u>JCBIO</u>	8월 3일 (목)	9:00~12:00	- 라이브러리 작성 실습	10
			13:00~16:00	- <u>Nanopore</u> running	10
		8월 17일 (목)	9:00~12:00	- 결과 생성 및 결과 분석	10

(주) JCBIO: Oxford Nanopore의 Korean distributor

- 조교: 식물분자계통학실 서정우 (석박통합과정), 이승연 (석사과정)

- 실험: 3개 조로 진행.

1조	4학년	손영비
담당:	4학년	정의정
서정우	4학년	고채빈
2조	4학년	정다영
담당:	4학년	조소희
이승연	4학년	조유연
3조	4학년	김예슬
담당:	4학년	유승진
김상태	3학년	서지예
	3학년	한세희

- 실험 대상: 목련 (*Magnolia kobus* DC.)

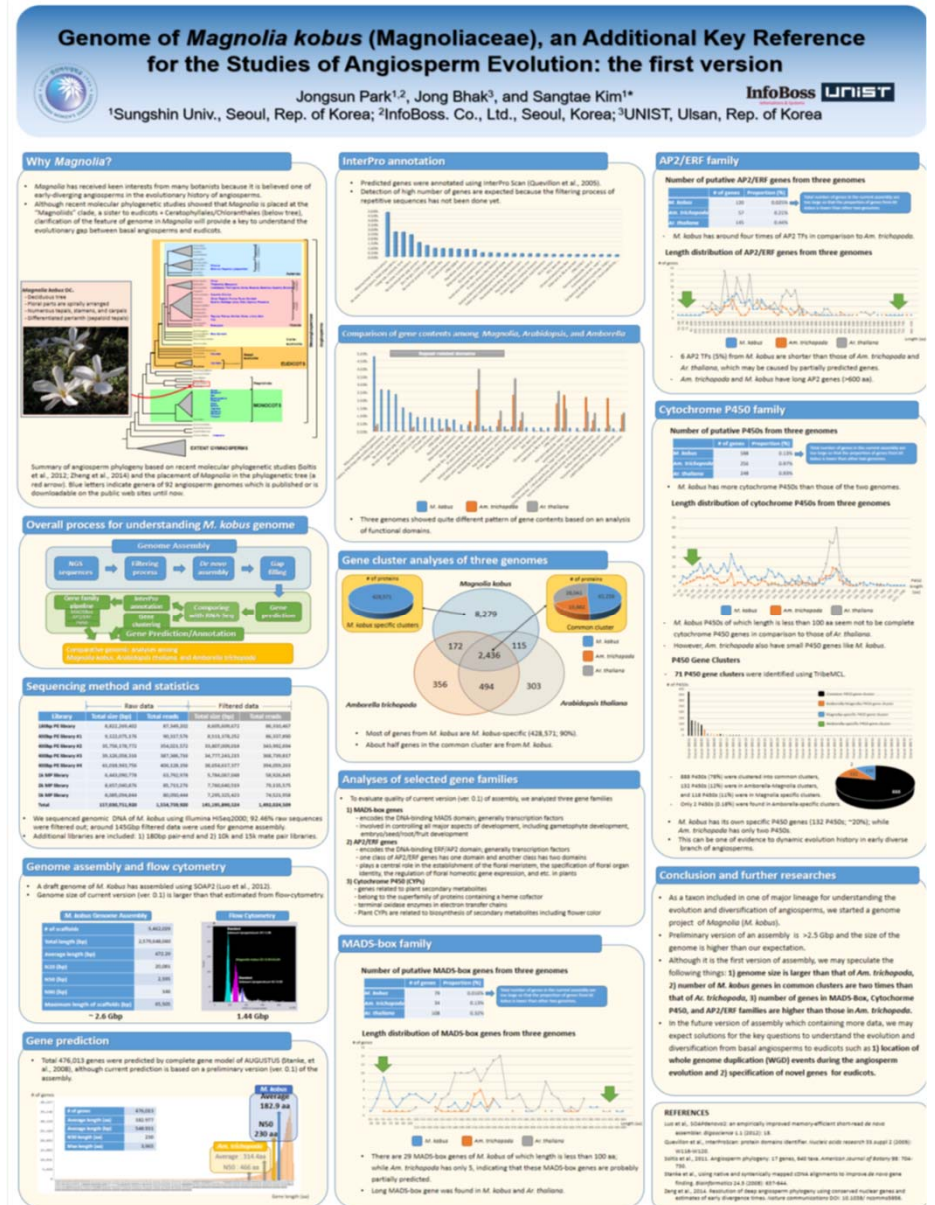
- Basal angiosperm

- 한국, 일본에 분포하지만, 한국에서는 제주도에 약 1,000여 개체 만이 존재하는 멸종위기종임.

- 성신여대 식물분자계통학실에서 지속적으로 분류, 계통, 진화, Evo-Devo에 대한 연구를 진행하고 있는 중임.
- 2016년 Illumina sequencin에 의해서 ver. 1.0 genome이 완성되었지만, 완성도가 너무 낮아 논문으로 발표되지 못한 상태임.
- 예상 유전체 크기: 1.44 Gbp (flowcytometry)
- Ver. 1.0의 assembly: 2.6 Gbp

# of scaffolds	5,462,029
Total length (bp)	2,579,648,040
Average length (bp)	472.29
N20 (bp)	20,081
N50 (bp)	2,595
N90 (bp)	146
Maximum length of scaffolds (bp)	65,505

→ long-read data의 추가가 필요함!



N50

- 일반적으로 유전체 수준에서 assembly가 잘 되었는지 안되었는지는 **N50** 값으로 제시함.
- 전체 contig 들을 크기순으로 배열하여 큰 것으로부터 크기를 차례로 더하여 더한 값이 전체 유전체 크기의 50%를 넘는 순간의 contig의 크기를 N50라 함.
- N50 는 평균값 또는 중간값과는 다른 의미의 수치임!!!
- 3 3 4 6 7 8 8 9 9 9 10 11 13 25 의 길이의 contig들이 있을 때

$$\text{Mean} = 125/14 = 8.93 \quad (\text{sum}=125)$$

$$\text{Median} = (3+25)/2 = 14$$

$125/2 = 62.5$ sum of contig lengths reach to 62.5 (from the largest to the smallest)

$$25+13+11+10=59$$

$$25+13+11+10+9=68$$

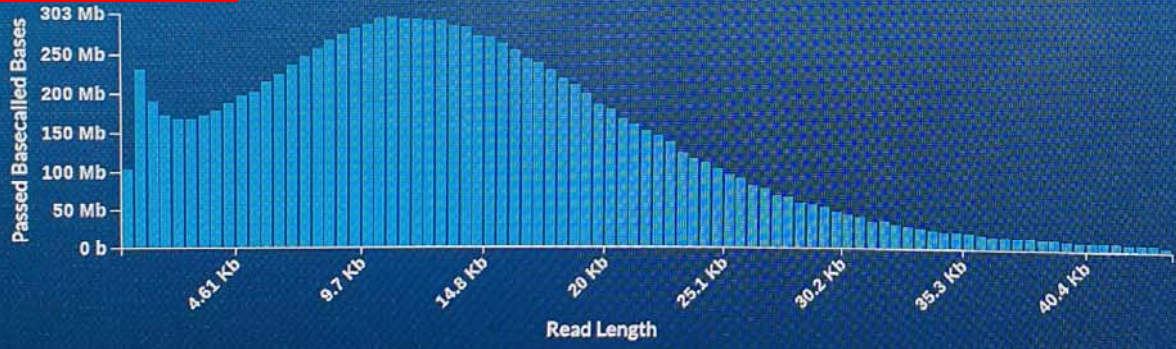
Therefore, **N50** = 9

Position	Flow cell ID	Sample ID	Health	Run time	Run state	Reads	Bases	Baseca
<input type="checkbox"/>	MN35716	FAO37153	no_sample	44.8 H / 7.2 H	Active	2.52 M	14.9 Gb basecalled 16.32 Gb estimated	100%

Scroll right >

Read length histogram

Estimated N50: 13.15 Kb



Read Bases

Lengths Counts Estimated Basecalled

Hide outliers Split by read end reason

- 사용할 Oxford Nanopore의 flowcell 및 시약: **MinION Starter Pack**

store.nanoporetech.com/minion-basic.html

PRODUCTS APPLICATIONS STORE RESOURCES SUPPORT ABOUT


HOME / STORE / DEVICES

< Configure your package

SELECT PACKAGE FLOW CELLS SEQUENCING KITS TRAINING CONFIRM


Select your flow cell type [Continue >](#)

1x Flow Cell (R10.4.1)

 R10 is our nanopore chemistry designed to deliver highest consensus accuracy. Paired with the Kit 14 chemistry, R10.4.1 generates data at a modal accuracy above 99%.
Note: R10.4.1 flow cells currently require Kit 14 chemistry.
Product lead time: 1 week

[Select](#)


1x Flow Cell (R9.4.1)

 The MinION and GridION Flow Cell contains up to 512 nanopore channels for sequencing DNA or RNA in real-time.

[Select](#)

Selected package

MinION Starter Pack [StarterPack](#)



Total: \$1,000.00

[Chat](#)

그런데...

재료가 좋아야 좋은 결과가 나옴.

→ 일반 DNA 추출과정으로는 조각난 DNA가 추출됨.

→ 그러므로 되도록이면 “긴” DNA를 추출하는 것이 성공적인
Nanopore sequencing의 1차적 관건임!

※ Long DNA = **High-Molecular Weight (HMW) DNA**

II. 식물체로부터의 HMW DNA 추출

최근 성신여자대학교 식물
분자계통학실에서 개발하
여 출판된 식물 HMW
DNA 추출 방법
(2023년 6월 공식 출판)

본 프로그램의 첫 part는
이 논문의 protocol에 준
하여 수행 됨.

Applied in Plant Sciences
의 DNA추출을 위한 특별
호에 게재됨.







<https://bsapubs.onlinelibrary.wiley.com/toc/21680450/2023/11/3>

Received: 26 September 2022 | Accepted: 4 May 2023

DOI: 10.1002/aps3.11528

PROTOCOL NOTE

High-molecular-weight DNA extraction for long-read sequencing of plant genomes: An optimization of standard methods

Myoungbo Kang¹  | Andre Chanderbali²  | Seungyeon Lee¹  |
Douglas E. Soltis^{2,3}  | Pamela S. Soltis²  | Sangtae Kim¹ 

¹Department of Biotechnology, Sungshin Women's University, Seoul 01133, Republic of Korea

²Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611, USA

³Department of Biology, University of Florida, Gainesville, Florida 32611, USA

Correspondence

Sangtae Kim, Department of Biotechnology, Sungshin Women's University, Seoul 01133, Republic of Korea.
Email: amborella@sungshin.ac.kr

This article is part of the special issue "Emerging Methods in Botanical DNA/RNA Extraction."

Abstract

Premise: Developing an effective and easy-to-use high-molecular-weight (HMW) DNA extraction method is essential for genomic research, especially in the era of third-generation sequencing. To efficiently use technologies capable of generating long-read sequences, it is important to maximize both the length and purity of the extracted DNA; however, this is frequently difficult to achieve with plant samples.

Methods and Results: We present a HMW DNA extraction method that combines (1) a nuclei extraction method followed by (2) a traditional cetyltrimethylammonium bromide (CTAB) DNA extraction method for plants with optimized extraction conditions that influence HMW DNA recovery. Our protocol produced DNA fragments (percentage of fragments >20 kbp) that were, on average, ca. five times longer than those obtained using a commercial kit, and contaminants were removed more effectively.

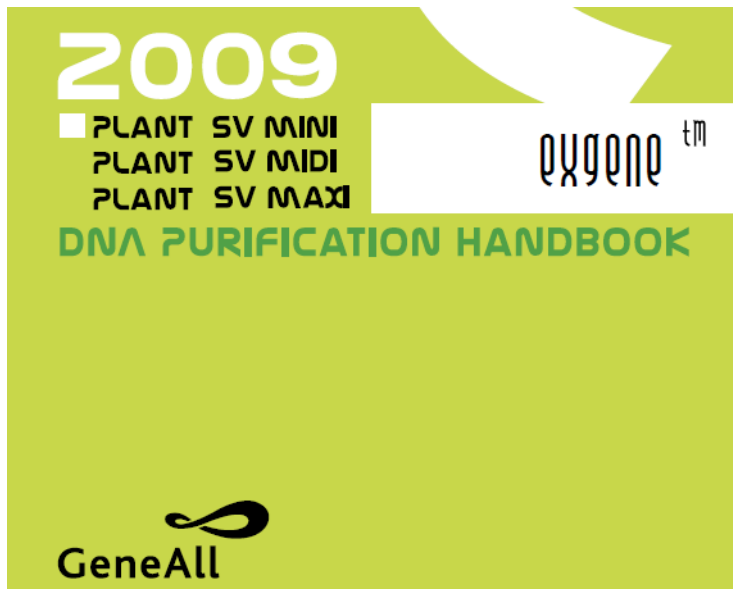
Conclusions: This effective HMW DNA extraction protocol can be used as a standard protocol for a diverse array of taxa, which will enhance plant genomic research.

KEYWORDS

CTAB, DNA extraction, Femto Pulse system, high-molecular-weight DNA, nuclei extraction

기초지식: DNA의 추출을 위한 버퍼와 DNA의 성질(일반생물학실험 자료):

- 파쇄된 조직을 extraction buffer (CTAB buffer) 와 섞으면 조직으로부터 DNA가 분리된다 (버퍼속의 EDTA가 킬레이트(chelate) 작용을 함)
 - ※ 킬레이트: 한 개의 리간드가 금속 이온과 두 자리 이상에서 배위결합을 하여 생긴 착이온을 뜻한다.
- 조직이 파쇄되면 DNase가 세포내에서 빠져나와서 DNA를 파괴하게 된다. 그러므로 chloroform등의 단백질 비활성화 물질을 처리함으로써 모든 효소작용을 정지시킨다.
- DNA는 염(salt)의 존재 하에서 70%정도의 EtOH에서 엉기는 (pellet을 형성) 성질을 갖고 있다. 이 때 원심분리를 하면 엉긴 DNA는 가라앉고 다른 이물질은 용액속에 남아있게 된다. 이러한 성질을 이용하여 순수한 DNA를 추출할 수 있다.
- DNA는 TE (Tris-EDTA) buffer에서 매우 잘 녹는다.
- 전통적인 식물 DNA 추출방법: CTAB method (일반생물학실험에서 다룸) link:



상업용 키트에 의한 일반적인 DNA 추출법:
 Filter-binding method에 의한 방법
 그러나 매우 단편화된 DNA가 추출됨.
 → PCR 등 일반 실험을 위해서는 문제 없지만,
 Nanopore로는 적합하지 않음.

