

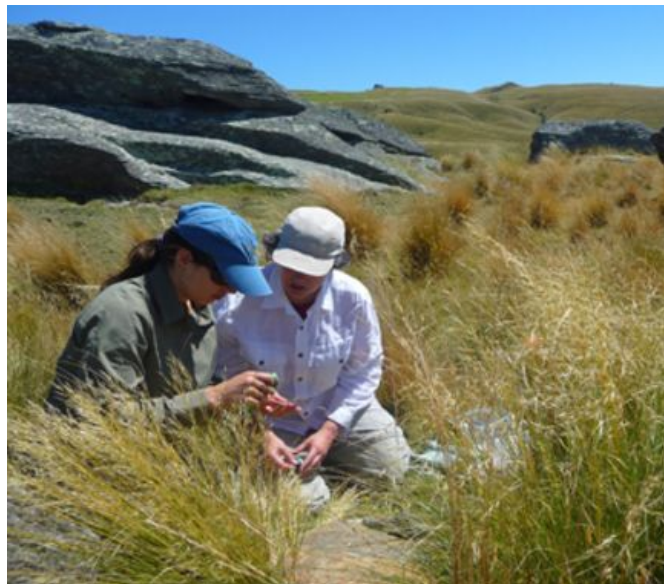
# 생물학을 위한 코딩

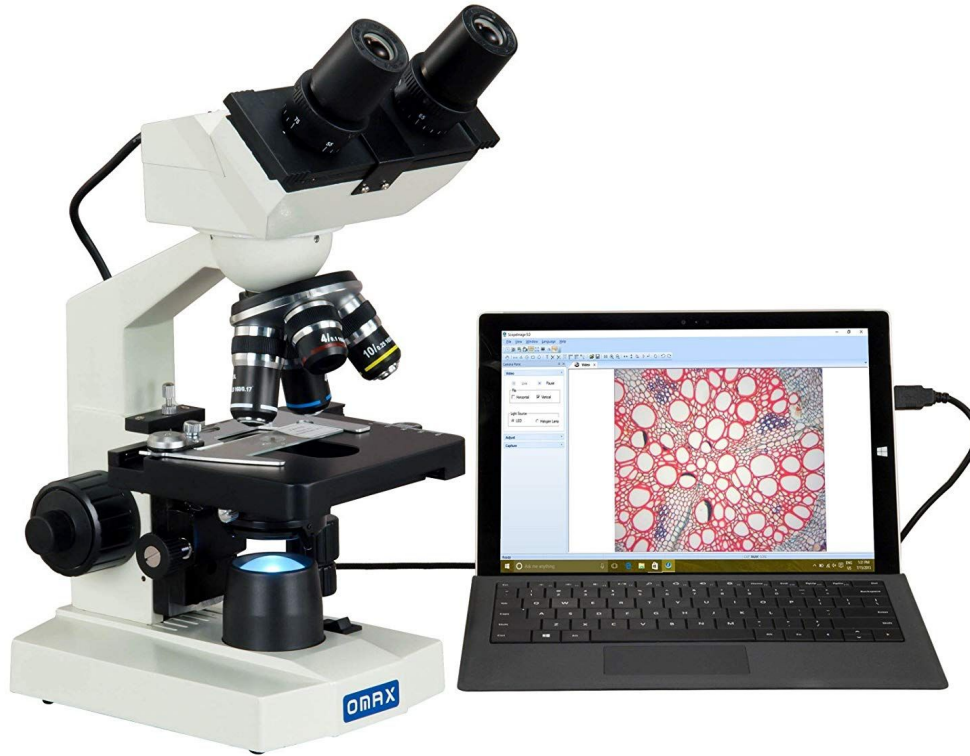


2학기

# 생물학은 어떤 학문인가?

- 관찰 위주의 학문 - 이미 존재하는 것을 해독한다.



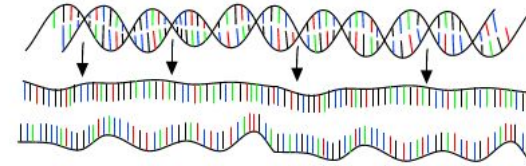


## Sequencing

30 cycles of 3 steps :

### Step 1 : denaturation

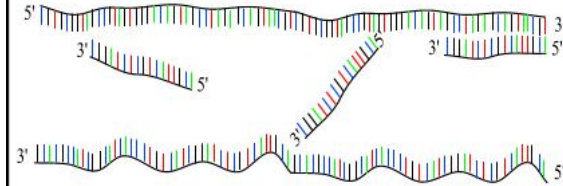
1 minut 94 °C



### Step 2 : annealing

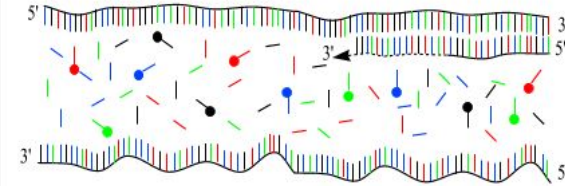
15 seconds 50 °C

1 primer !!!!

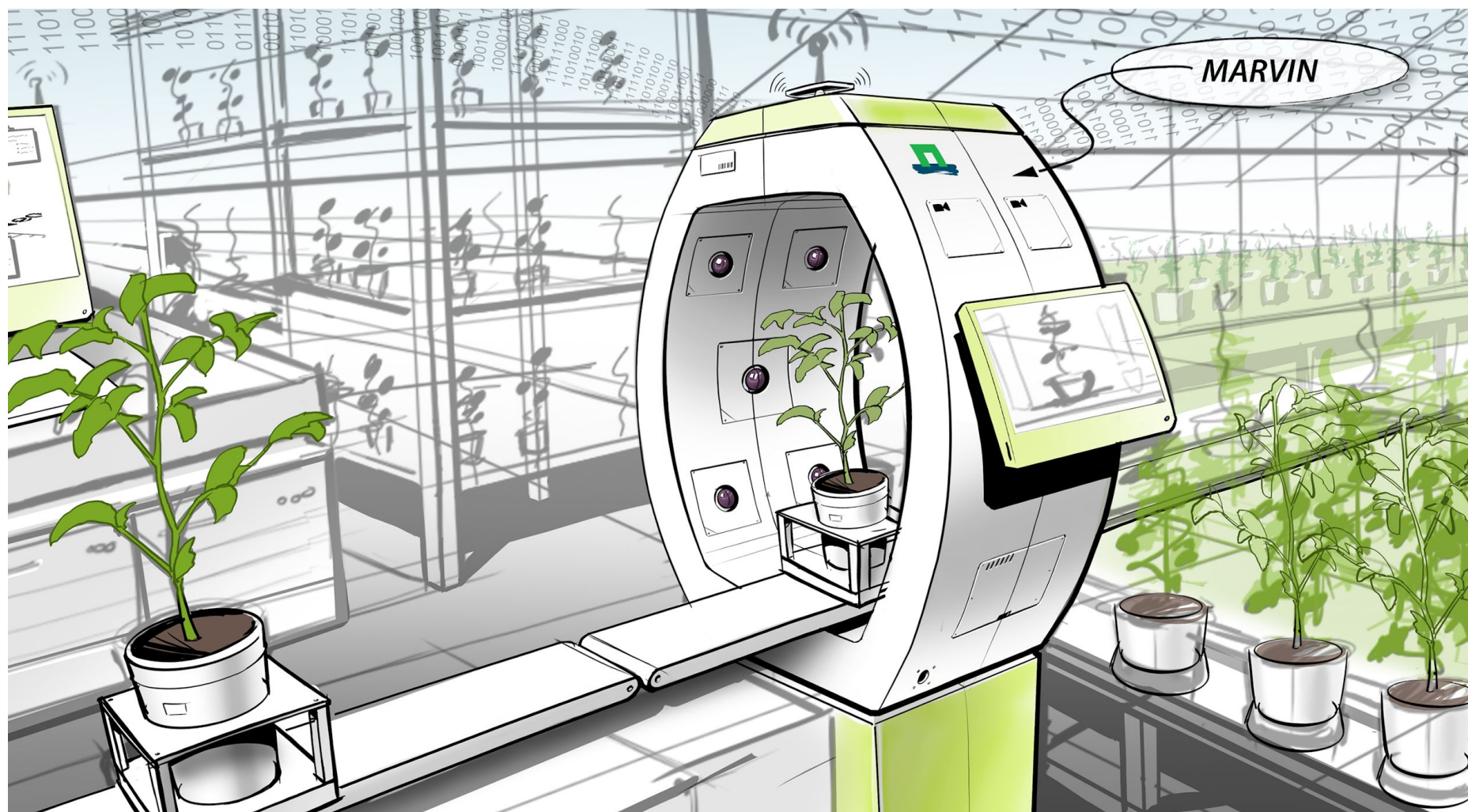


### Step 3 : extension

4 minutes 60 °C  
mixture of dNTP's  
and ddNTP's



(Andy Vierstraete 1999)

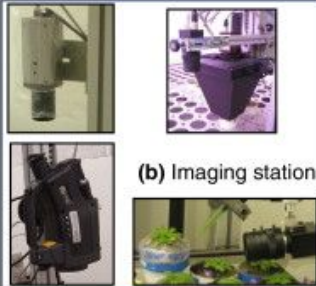








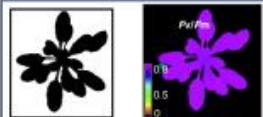
**(a)** Phenotyping platform with environmental sensors



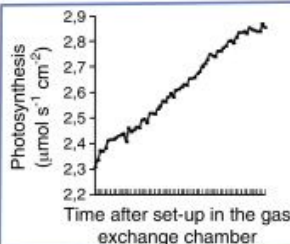
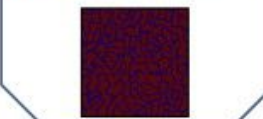
**(b)** Imaging station



**(d)** Other phenotyping technologies

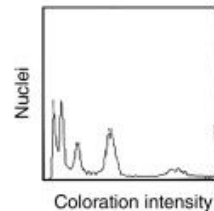
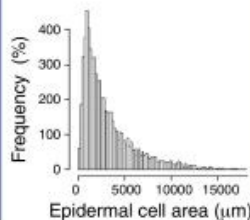


**(c)** Image analysis

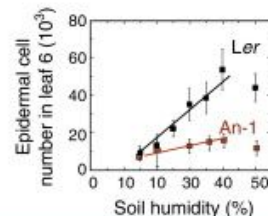
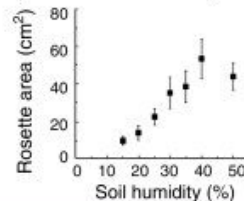
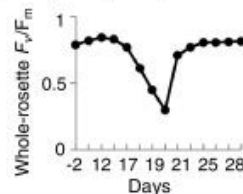
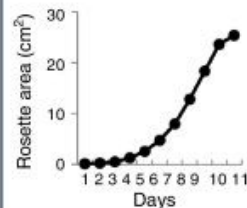


**(e)** Raw data

Area ( $\text{cm}^2$ )	Leaves	Bolted
32.5	16	No



**(f)** Integrating time, environment and genetics in data analyses



# 생물학은 어떤 학문인가?

- 관찰이후의 추론이 중요한 학문

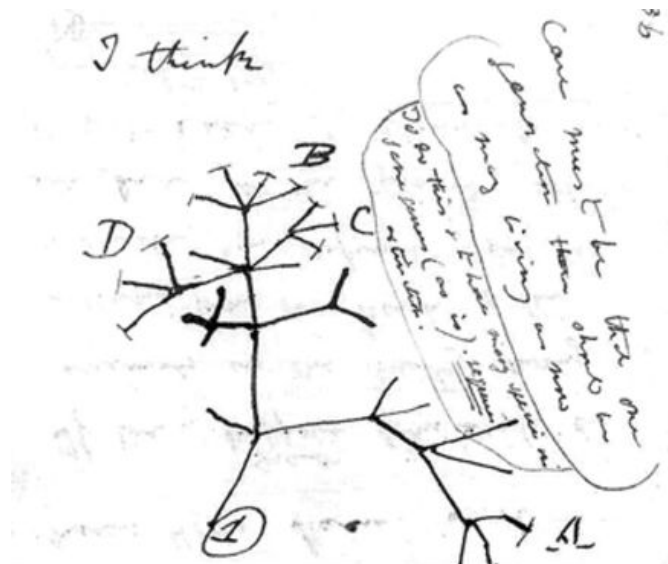
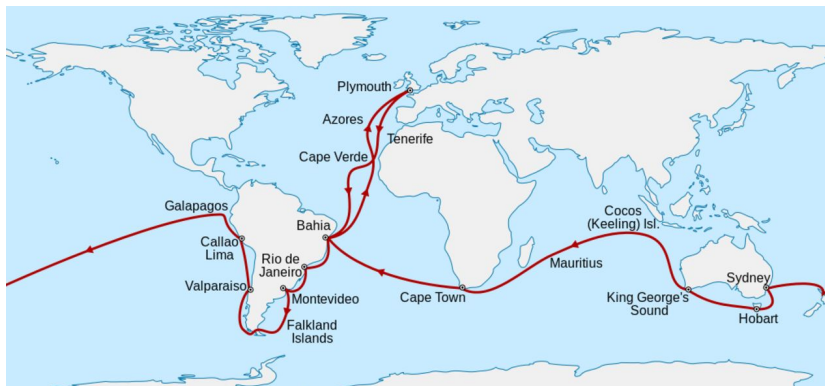






# 생물학은 어떤 학문인가?

- 관찰이후의 추론이 중요한 학문



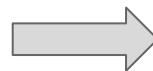
찰스 다윈이 '생명의 역사는 기존의 종으로부터 새로운 종이 가지를 쳐온 과정'이라는 개념을 설명하기 위해 그린 '생명의 나무' 그림, 동아일보 자료 사진



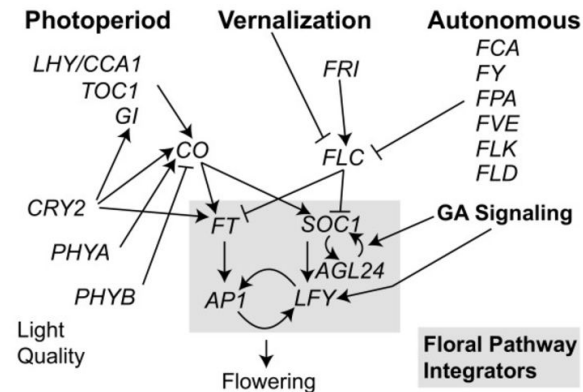
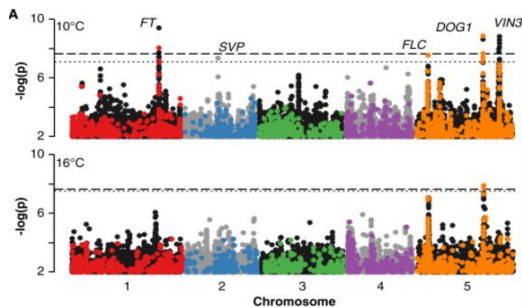
다양한  
관찰



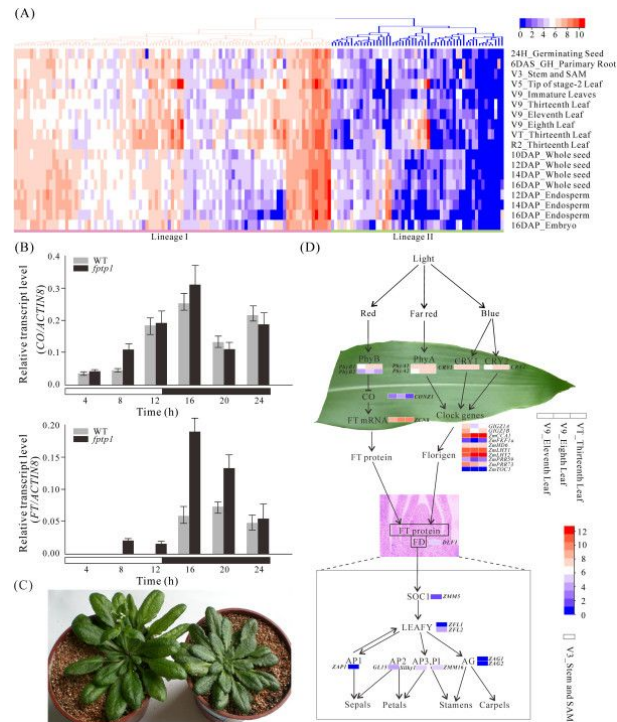
추론한 모델  
(가설)



뒷받침하는 증거



10.1186/2041-9139-2-4



# 표현형 관찰



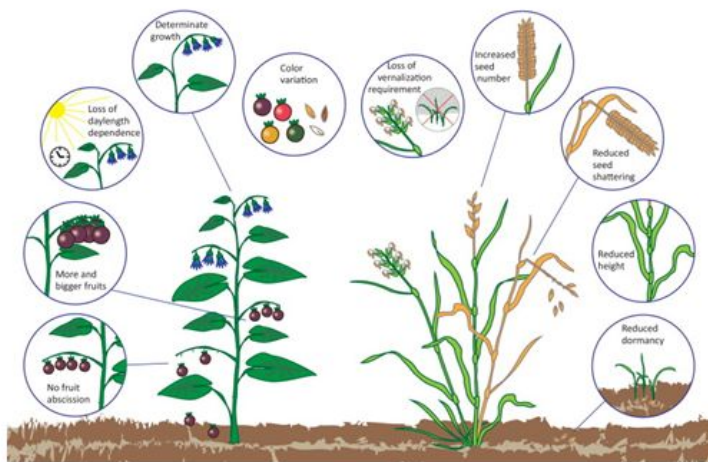
채집

농경의 시작  
(~ 5,000 BP)



멘델의 법칙 (1865)

- 과실이 크다.
- 떨어지지 않는다.
- 예쁘다.
- 다루기 쉬운 정도의 크기이다.
- 다시 심으면 잘 난다.
- 생산량이 많다.





# 유전형 관찰



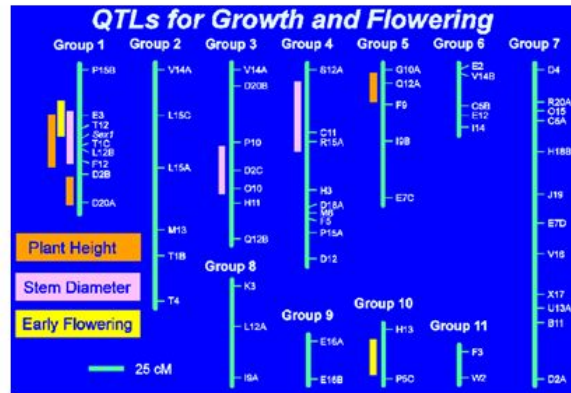
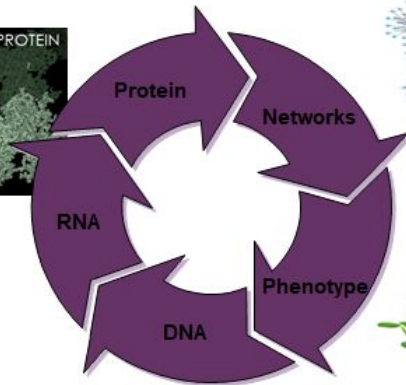
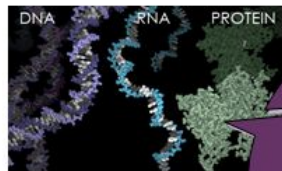
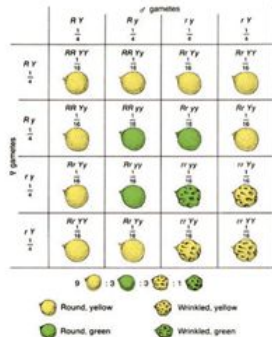
멘델의 법칙 (1865)



왓슨 & 크릭  
DNA의 발견 (1950년대)



PCR의 발명 (1983),  
각종 다형성 (Polymorphism) 관찰



# 유전형관찰 - 생물 빅데이터 시대



애기장대  
유전체  
해독 (2000)



인간유전체  
해독 (2001)

표준유전체를 작성하기 위해  
전 세계 연구그룹의 참여  
천문학적인 금액 소요



Illumina solexa GA의  
개발 (2006) - Next Generation Sequencer (NGS)



## ARTICLE

Received 3 Feb 2014 | Accepted 30 Sep 2014 | Published 11 Nov 2014

[Link to this Article](#)

[OPEN](#)

## Genome sequence of mungbean and insights into evolution within *Vigna* species

Yang Jae Kang<sup>1</sup>, Sue K. Kim<sup>2</sup>, Moon Young Kim<sup>3</sup>, Puji Lestari<sup>1,2</sup>, Kil Hyun Kim<sup>3</sup>, Bo-Kaun Ha<sup>4</sup>, Tae Hwan Jun<sup>5</sup>, Won Joo Hwang<sup>6</sup>, Taeyoung Lee<sup>7</sup>, Jaeyun Lee<sup>8</sup>, Sangra Shin<sup>1</sup>, Min Young Yoon<sup>9</sup>, Young Eun Jang<sup>1</sup>, Kwang Soo Han<sup>1</sup>, Puritane Tepraporn<sup>10</sup>, Na Yoon<sup>1</sup>, Prakit Sontak<sup>1</sup>, Patcharin Tanya<sup>10</sup>, Kwang Soo Kim<sup>1</sup>, Jae-Gyun Gwak<sup>1</sup>, Jung-Kyung Moon<sup>1</sup>, Yeong-Ho Lee<sup>1</sup>, Beom-Seock Park<sup>1</sup>, Aureliano Bombarely<sup>11</sup>, Jeffrey J. Doyle<sup>9</sup>, Scott A. Jackson<sup>12</sup>, Roland Schaffner<sup>13</sup>, Peerasak Srinives<sup>14</sup>, Rajeev K. Vanshrey<sup>12</sup> & Suk-Ho Lee<sup>1,13</sup>

녹두 유전체 해독 (2014)

단일 연구그룹에서  
표준유전체 완성 가능

## Before - NGS 의 출현 - After



ABI3730xl

1번 돌렸을 경우

출력하는 각 염기서열 길이 : ~ 700 base  
총 생산하는 염기서열 길이 : 60,000 base  
가격 : 약 50만원



Illumina HiSeq 2000

1번 돌렸을 경우

출력하는 각 염기서열 길이 : ~ 150 base  
총 생산하는 염기서열 길이 : 30,000,000,000 base  
가격 : 약 400 만원







# 생물정보학의 도래

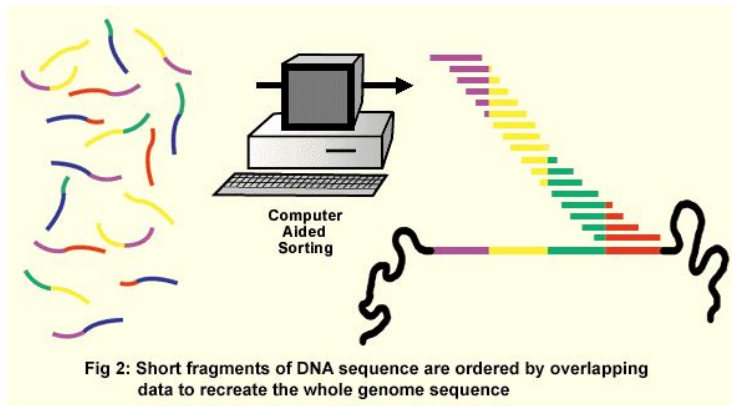
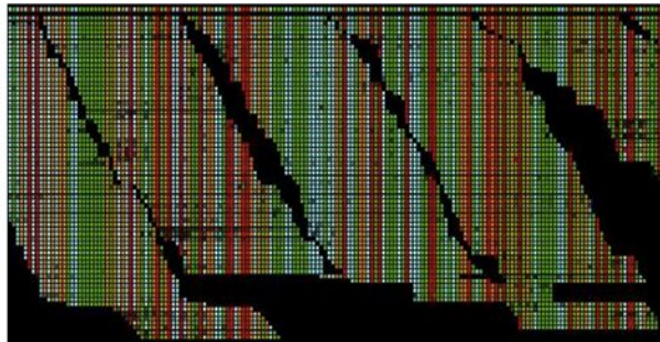


Fig 2: Short fragments of DNA sequence are ordered by overlapping data to recreate the whole genome sequence

“생물 현상을 관찰하는데 쓰는 도구에 컴퓨터가 들어오다!”

> 생물 정보학의 출현



# Python

[GO](#)[Socialize](#)[About](#)[Downloads](#)[Documentation](#)[Community](#)[Success Stories](#)[News](#)[Events](#)

```
# Python 3: Fibonacci series up to n
```

```
>>> def fib(n):
```

```
>>>     a, b = 0, 1
```

```
>>>     while a < n:
```

```
>>>         print(a, end=' ')
```

```
>>>         a, b = b, a+b
```

```
>>>     print()
```

```
>>> fib(1000)
```

```
0 1 1 2 3 5 8 13 21 34 55 89 144 233 377 610 987
```

## Functions Defined

The core of extensible programming is defining functions. Python allows mandatory and optional arguments, keyword arguments, and even arbitrary argument lists. [More about defining functions in Python 3](#)

[1](#)[2](#)[3](#)[4](#)[5](#)

# 왜 파이썬인가?

- <http://www.numpy.org/> - 연산특화 라이브러리
- <https://www.scipy.org/> - 각종 과학관련 수식 라이브러리
- <http://scikit-learn.org/> - 기계학습 라이브러리
- <http://scikit-image.org/> - 이미지 인식 라이브러리
- <https://matplotlib.org/> - 시각화 라이브러리
- <https://www.tensorflow.org/> - 딥러닝 라이브러리

# 생물학의 미래

- 생물정보학이란 학문이 따로 존재하지 않고, 생물학에 흡수된다.
- 모든 생물학자들은 컴퓨터 기반 분석을 할 줄 안다.
- 컴퓨터를 통한 “관찰”을 수행한다.
- 생물현상을 컴퓨터를 통해 다양하게 생성된 가설에 비추어 분석하고 최적의 가설(모델)을 찾는다.
- 단일 유전자 관찰을 통해 만들어내는 가설들 역시 생물 빅데이터에 의해서 재검증되어야 한다.
- 4차산업혁명이라고 불리우는 인공지능 기술의 발전 역시 생물학에 큰 지형변화를 가져올 것이다.



# 그래서

- 이번 학기에 python을 배워본다.
- 재미없는 컴퓨터 이야기 보다는 생물 데이터 위주의 강의를 할 예정.
- 수업자료들은 전부 웹페이지에 올려둘 예정.
- 성적? 40:40:20
- 현재 위치 : 27동 419호
- Email : kangyangjae@gnu.ac.kr
- 질문 항상 받음