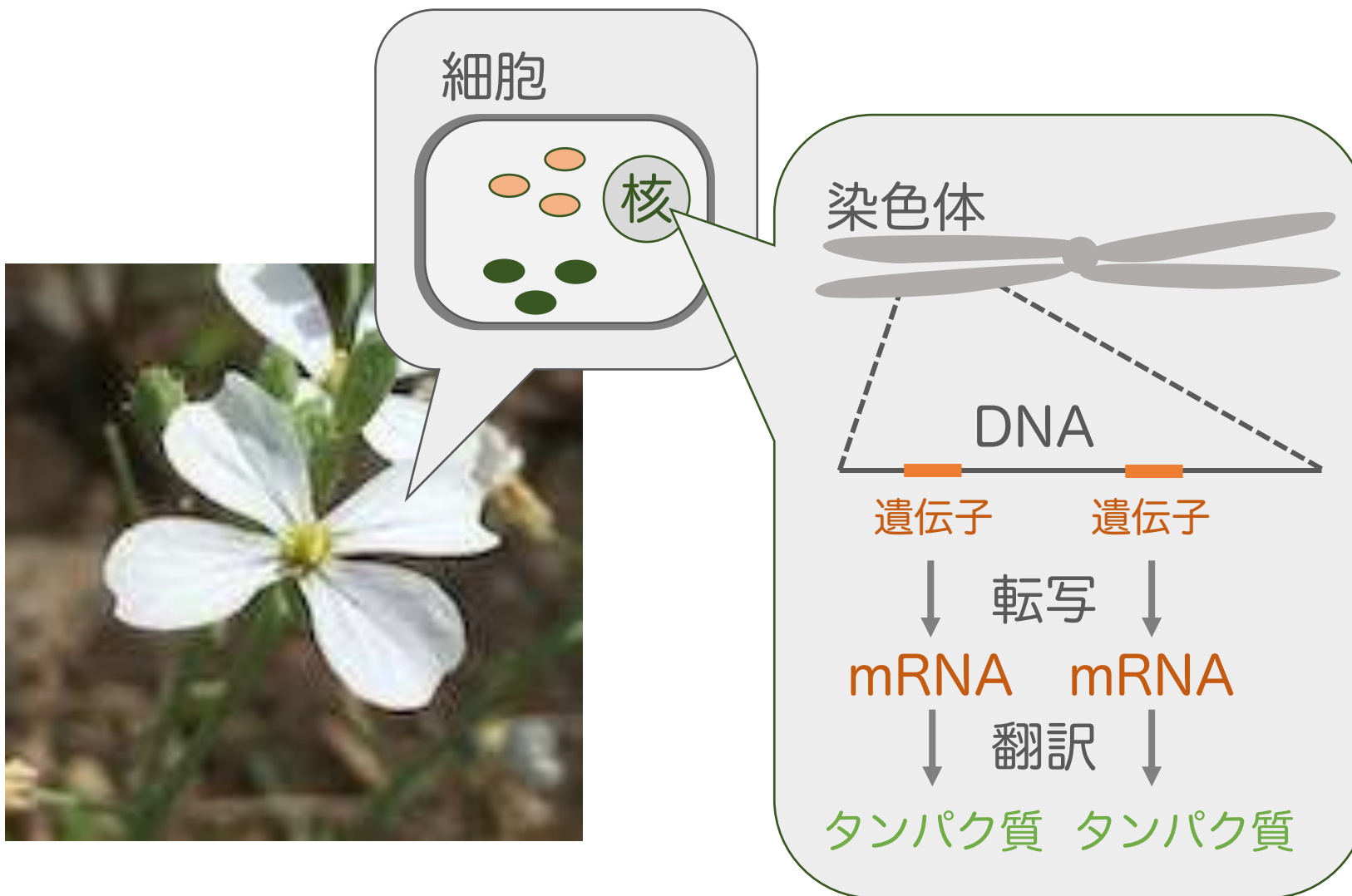


ILASセミナー2021「ゲノム生物学入門」

ゲノム解析 イントロダクション

イントロダクション

ゲノム情報（遺伝情報）



ゲノム情報は読み出せる

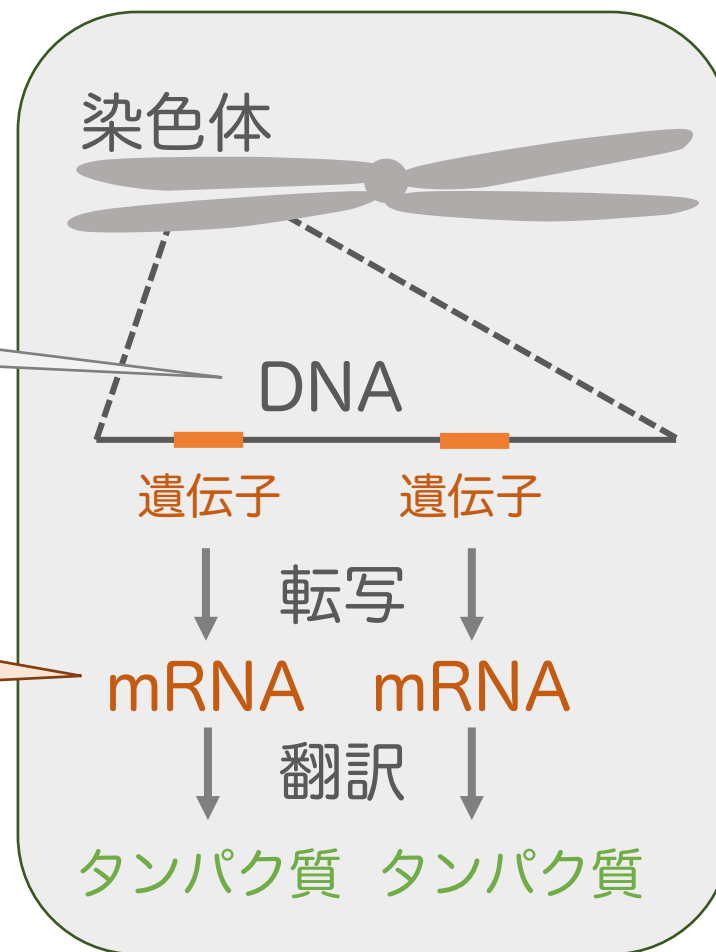
次世代シーケンサー



<https://jp.illumina.com/>

ゲノムシーケンス

RNAシーケンス
(RNA-seq)



ゲノム情報の記述



ゲノム配列 (DNA配列) は
4つの塩基で書かれる

アデニン (A)
シトシン (C)
グアニン (G)
チミン (T)

イネ第1染色体

8001番目から9000番目の塩基

```
>1 dna:chromosome chromosome:IRGSP-1.0:1:8001:9000:1
TGTTCTAATTTGCAAAACCTATTACAGTGGTTGATGTGGTTAGAACGGATAGTCATCTTGATTTCTATG
GAGAATCAAGAAATATGGCTAGGATGTCAGATATACTTGCAGTTTATGCATGGGTTGATCCTTCCACC
GGGTATTGCCAAGGTATATGCATATTTCAGTGGCATCTTTTTTTTACTTCCTTTTTTGGTTTCTTGATGTGT
CTAAATGGGATAAGTTGCTAACAGGTATGAGTGATCTACTTTCACCTTTTGTTGTTCTCTATGAGGATG
ATGCAGATGCCTTTTGGTGCTTTGAGATGCTACTGAGAAGGATGGTAATTCACATATATTTCAAATTTT
GTTGCCTATTAGTTTGGTTTCTATGTTACTCCATGCATAATATGTGCCCTGTTCCACTTCAGCGTGAAAA
TTTCCAGATGGAGGGACCAACAGGAGTTATGAAGCAGTTGCAAGCATTGTGGAAGATCATGGAAAT
AACTGATGTAGAATTATTTGAACATTTGTCAACAATTGGTGCTGAAAGCCTTCATTTTGCTTTCCGGAT
GCTGCTAGTGCTTTTTTCGTCGAGAGCTGTCCTTTGAGGAGTCACTTAGCATGTGGGAGGTGTGGCAT
TTGAAGACACATGTATTATCTTTTTTTCAGTTGTTGTTCTTCGTAGGAAAGCTCTAATGGATGTTTGTTT
GTCAGCTGGTGTAGGTCTAGGGGATGTTGGTTAAGCCTTTTTTATTTCAAAAATTGTATTACTAGTTAA
ATATTTCAAATAGTAAAATAGGGTACATCTTCGATTTAGGTTCAAATCATCAGAGGTGTAGGTAGATG
GTTCAATTACGATTGTAGTTAATGTTTGTGATATCATGTGGCATGTGAACTTTATACTCCCTCGGGTCACA
AATATTTGACGTTTAGGACAAGATGTGGTCAAACCTTTTAAAATTTTGACCGTCAATAACAGCTCAAAT
GCGTAGCTTATAAAAAACACAAAAAAAGATATGCGGAGA
```

ゲノムの大きさ

生物種	ゲノムの大きさ
ヒト	3,000 Mbase (3億塩基)
シロイヌナズナ	130 Mbase (1,300万塩基)
イネ	390 Mbase (3,900万塩基)
コムギ	17,000 Mbase (17億塩基)

ゲノム中の遺伝子の個数

生物種	ゲノムの大きさ	遺伝子数
ヒト	3,000 Mbase (3億塩基)	26,000 遺伝子
シロイヌナズナ	130 Mbase (1,300万塩基)	27,000 遺伝子
イネ	390 Mbase (3,900万塩基)	37,000 遺伝子
コムギ	17,000 Mbase (17億塩基)	120,000 遺伝子

ゲノム解析は「ビッグデータ解析」

1塩基ずつ（1遺伝子ずつ）見て調べるのはほぼ不可能



コンピュータの力を借りよう！

プログラミング言語「Python」を勉強する

事前準備

事前準備

- インターネットに接続されている
- ブラウザがインストールされている

利用可能なブラウザ

- Google Chrome
- Firefox
- Safari (Mac)

- Googleアカウントを持っている

解析環境: Google Colaboratory

