

# **理系基礎 生物学基礎II**

**金曜日 2限 (10:30~12:00)**

**教室 S30**

**10月26日分 井原**

# バイオインフォマティクス (Bioinformatics)

生命情報は核酸 (DNA, RNA) に何らかの形でコードされており、それが細胞というハードマシンによってタンパク質という分子に変換されて、それらが巧妙に機能することで“生物”の一見すると不思議に思える部分が作られている。これらのプロセスの全体 (あるいは一部) において、生物が採用しているルールを理解して、他の分野に利用しようとする学問が**バイオインフォマティクス**である。**数理生物学** や **生物統計学** という名称で発展してきた学問分野の発展系と捉えることもできる。

コンピュータ能力の著しい進歩によって、計算機科学が様々な分野と融合し始めた。生物学の様々な問題に対してコンピューターを使ってアプローチする学問全般に対して**バイオインフォマティクス**という総称が使われることも多い。

# 生物学の歴史

## 自然科学 → 博物学 (Natural history)

自然に存在するものについて研究する学問。生物（動物・植物）・鉱物（岩石）など、自然物について収集し、分類する学問。

## → 分類学

ヒト（あるいは動物）にまつわる興味（主に病気）から、  
解剖学、形態学、細胞学

## 19世紀以降になって

発生学、遺伝学、微生物学、生化学  
などが学問として成熟していった。

20世紀中頃になって 分子生物学（考え方は生物物理学） という領域が  
生まれ、それはすべての生物学に浸透していった。

つまり、物理化学の法則に基づいて、様々な分子が相互作用、変化すること  
で様々な生命現象の説明を試みる生物学である。

# 生物学における**ビッグデータ**

シーケンス技術の急速な向上により、生物の遺伝情報（遺伝子）を読み解くのがそれほど大きな障壁では無くなってきた。それで、様々な生物で、その遺伝子をすべて読み解くというゲノム解析が行われるようになった。同時に、個別の生命現象に関わる遺伝子（タンパク質）を対象にした従来の分子生物学から 遺伝子全体、転写物全体、タンパク質全体を総体としてすべて解析して理解しようとする分野が生まれてきた。

## オーム (ome) の時代

ゲノム (genome; gene + ome)

トランスクリプトーム (transcriptome)

プロテオーム (proteome)

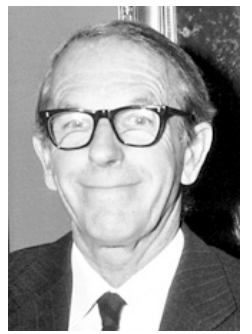
解析には、計算機が必須であり、データベースも必要になる。

# 科学技術の歴史として

## DNA塩基配列決定の歴史

DNA塩基配列法の開発 1975年

Frederick Sanger (1958, 1980 ノーベル賞)



### ddNTPの利用

$^{32}\text{P}$  ラベル (X線フィルム、現像、目視で読み取り)

蛍光ラベル (レーザー励起、自動信号読み取り)

PCR法の出現 (DNAを増やせる)

### シーケンス効率の増加

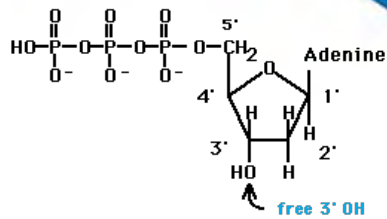
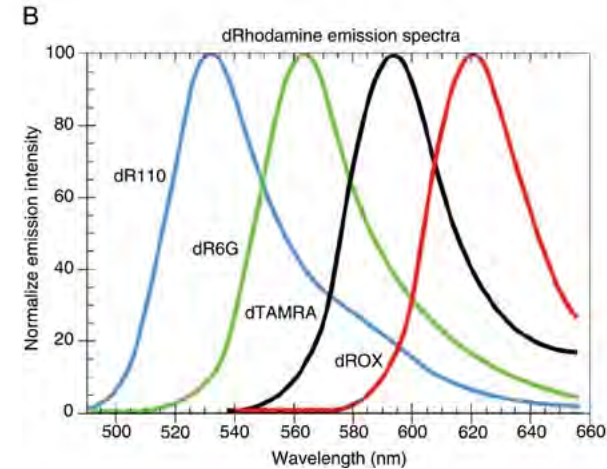
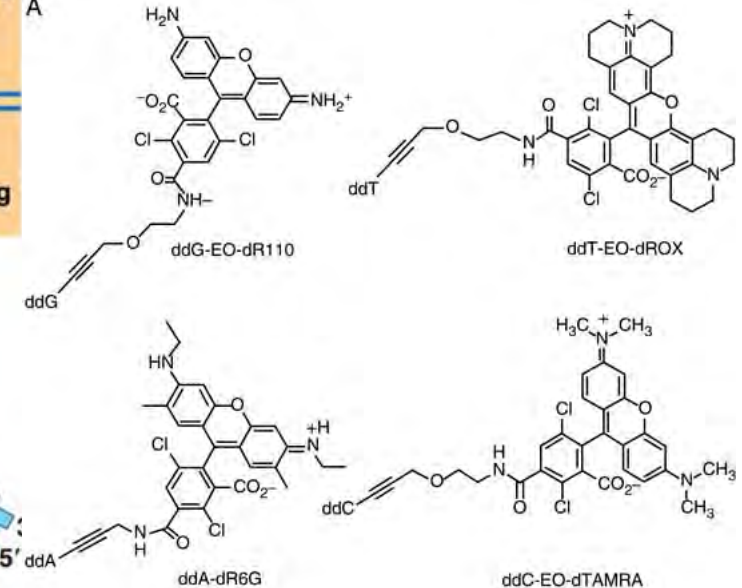
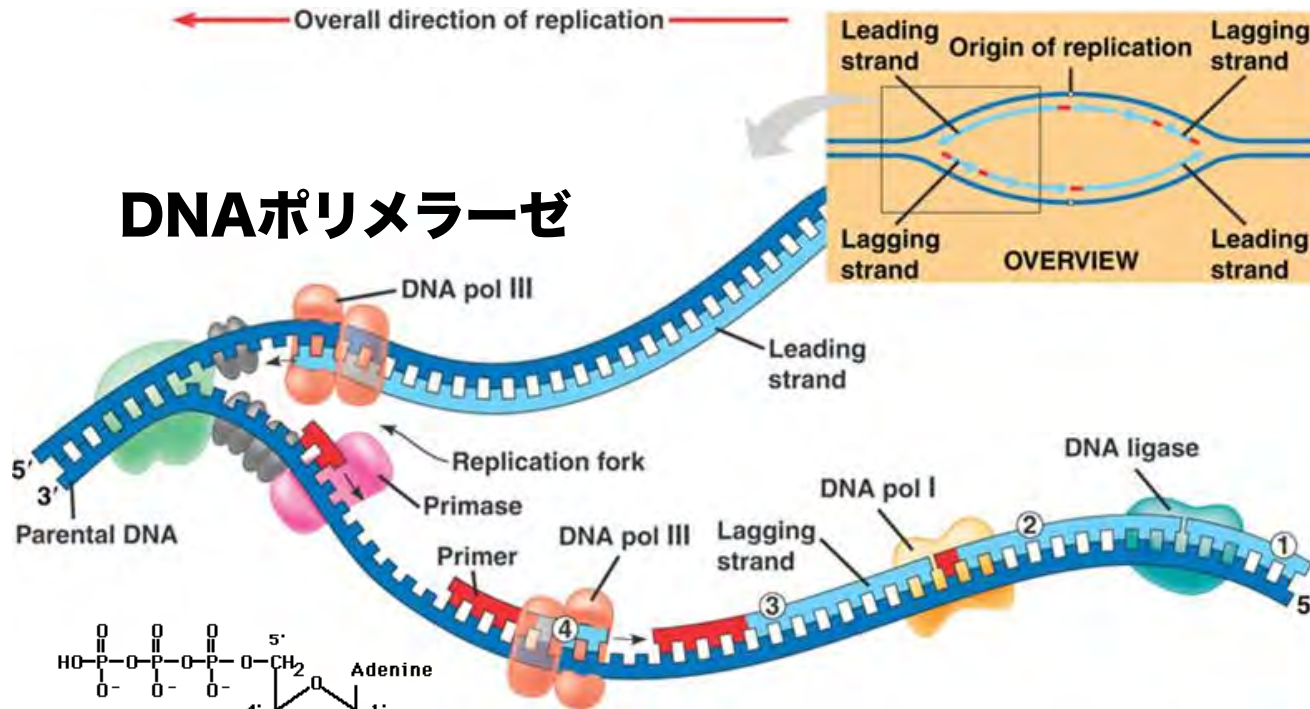
異なるシーケンス技術の模索

(もっと安く、もっと大量に！)

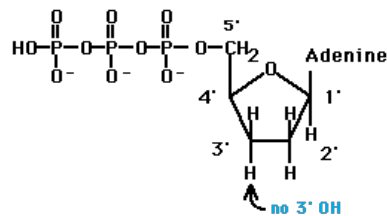
次世代シーケンサーの登場

# サンガー法による塩基配列決定方法

## DNA複製とダイデオキシヌクレオチド



dATP



ddATP

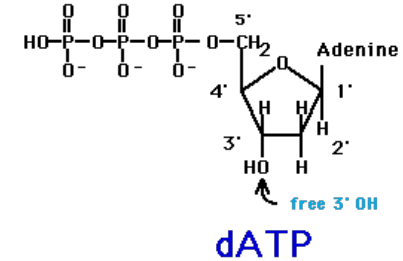
dATPの代わりにddATPを使うとddATPが取り込まれた位置でその先の反応が進まなくなる。

# サンガー法による塩基配列決定方法

## ダイデオキシヌクレオチドによるシーケンス その1 dNTPをA,T,G,Cと、ddNTPをA\*,T\*,G\*,C\*と表す

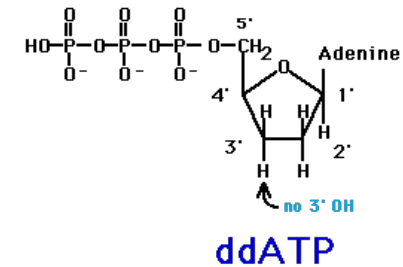
①シーケンスしたいDNA試料を、高温で2本鎖に解離させる

5'-CGATCGATGCATGCATCGATGCATCGATCGACGAAAATTTGGCGCGCGC  
3'-GCTAGCTACGTACGTAGCTACGTAGCTAGCTGCTTTTAAACCGCGCGCG



②プライマーと呼ばれる短い配列を結合させる

5'-CGATCGATGCATG  
3'-GCTAGCTACGTACGTAGCTACGTAGCTAGCTGCTTTTAAACCGCGCGCG



③DNAポリメラーゼにより伸長反応を行う

5'-CGATCGATGCATG CA  
3'-GCTAGCTACGTACGTAGCTACGTAGCTAGCTGCTTTTAAACCGCGCGCG

T C C A A C T T C A C  
 A T T C C G C A C

## RIラベルから蛍光ラベルへ

④ダイデオキシヌクレオチド (A\*, G\*, C\*, T\* で表す) が共存すると

5'-CGATCGATGCATG CATCGA  
3'-GCTAGCTACGTACGTAGCTAGCTAGCTGCTTTTAAACCGCGCGCG

T A G G C\* C T\*  
 T\* C G\* A\* T G A G

STOP C C T C C T\* A  
 T A\* A G G\* G C\*

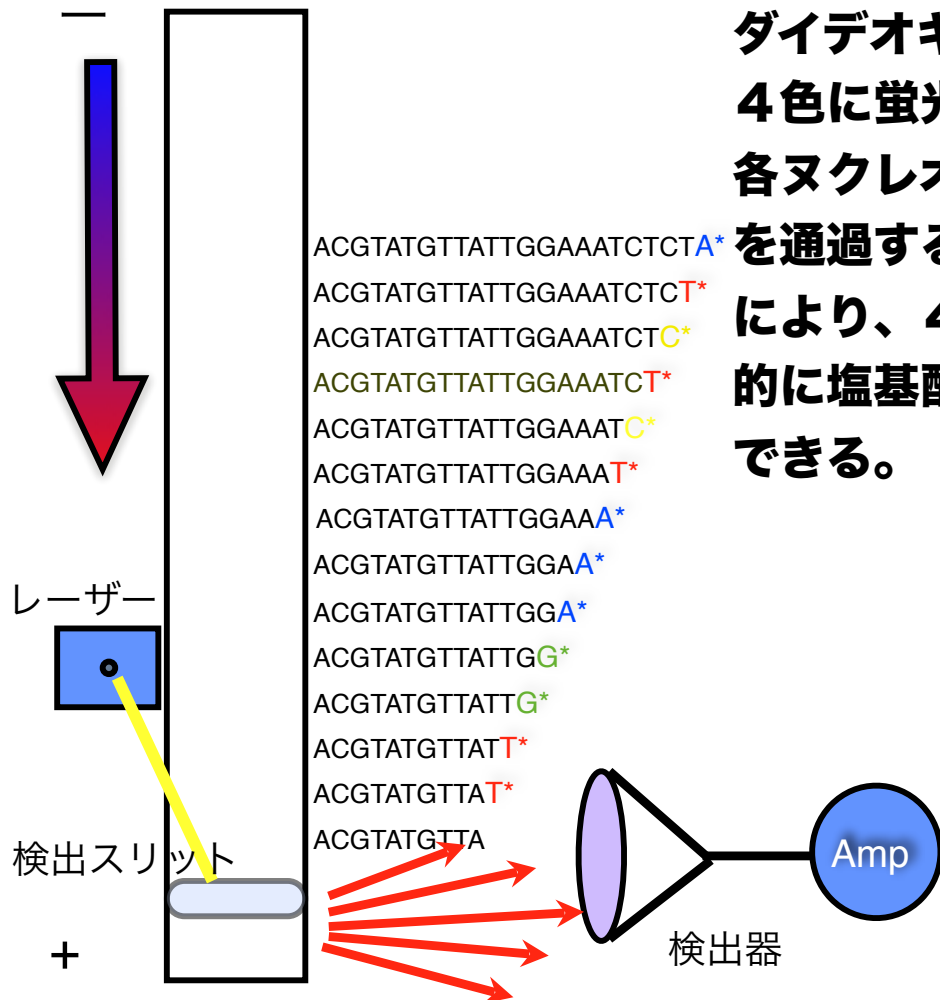


# サンガー法による塩基配列決定方法

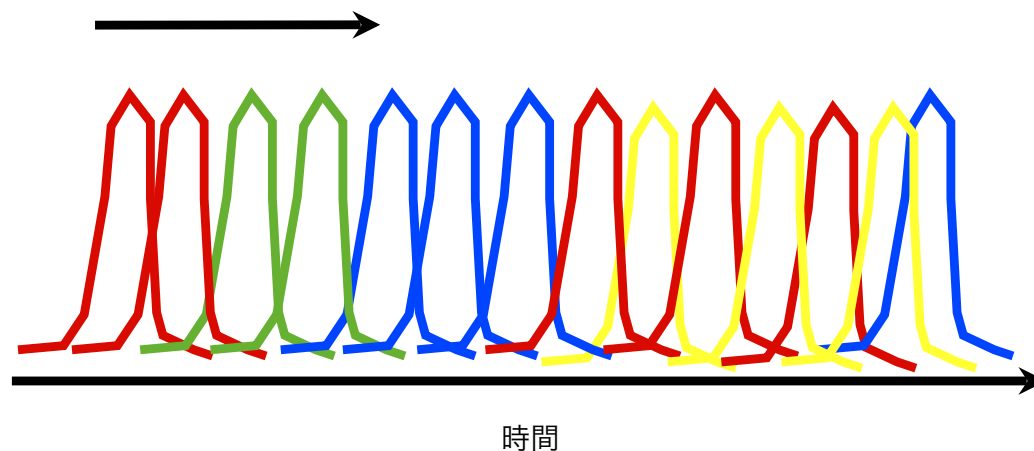
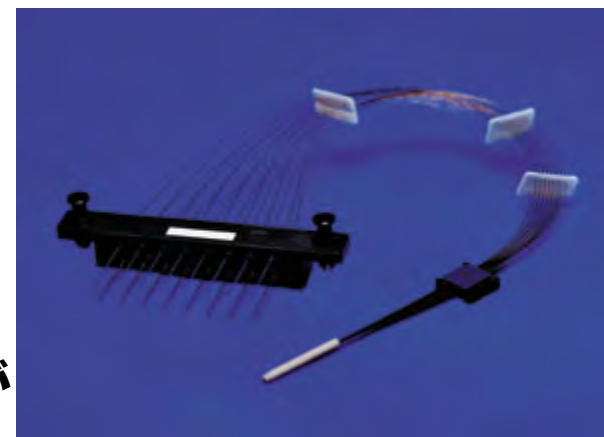
## ダイデオキシヌクレオチドによるシーケンス その2

高分子ゲルの分子ふるいを利用したDNAの電気泳動による分離と蛍光による検出

1塩基の違いで分離できる



ダイデオキシヌクレオチドを4色に蛍光ラベルすると、各ヌクレオチドは、検出スリットを通過する際に、レーザー照射により、4色に色分けされ、自動的に塩基配列情報に変換することができる。





# 実際のシーケンス例 およそ600塩基対を決定することができる。

Model 3100  
BC 1.5.1.0

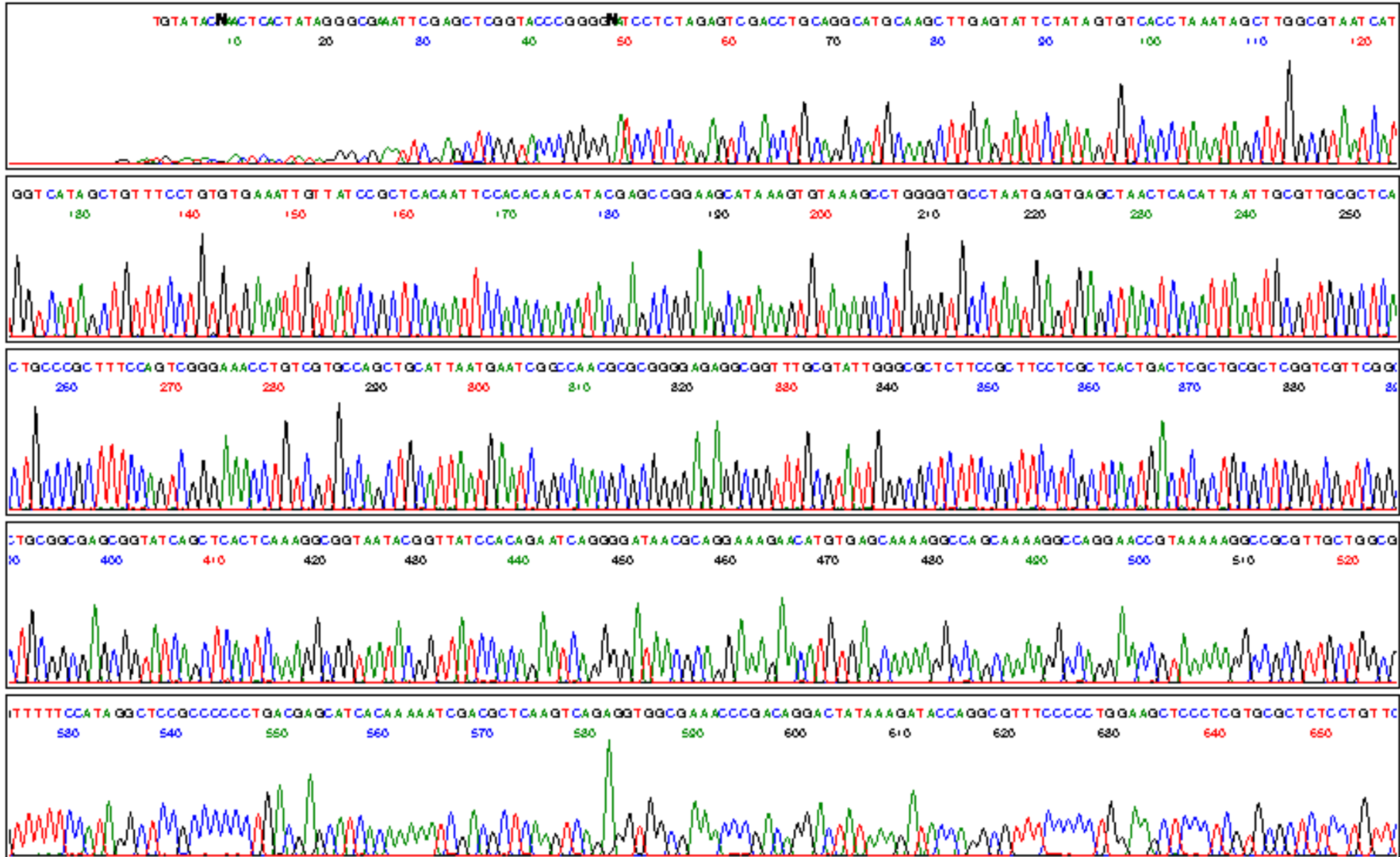
pGEM2Re\_016.ab1

pGEMCONT2  
Lane 16

Signal G:354 A:320 T:525 C:587  
KB\_3100\_POP4\_BDTv3.mob

Points 813 to 18478 Base 1: 813

Page 1 of 2  
2004年 4月 31日 (日) 9:42 PM  
2004年 4月 17日 (日) 6:09 PM  
Spacing: 20.90



# 試験管内でDNAを増やす (PCR反応)

酵素：DNAポリメラーゼ

基質：dATP, dGTP, dCTP, dTTP

プライマーと呼ばれる短いDNA

鋳型となるDNA

反応温度： 37°C

反応溶液

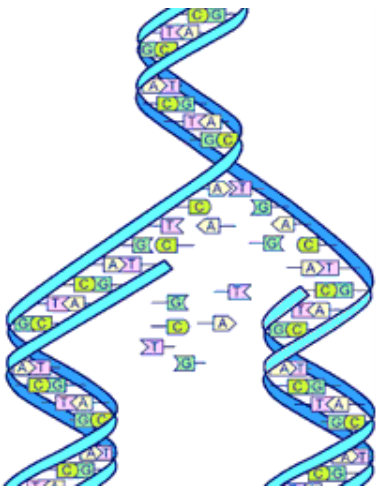
50 mM NaCl (塩)

10 mM Tris-HCl (緩衝剤)

10 mM MgCl<sub>2</sub>

1 mM DTT (還元剤)

pH 7.9 (@ 25°C)



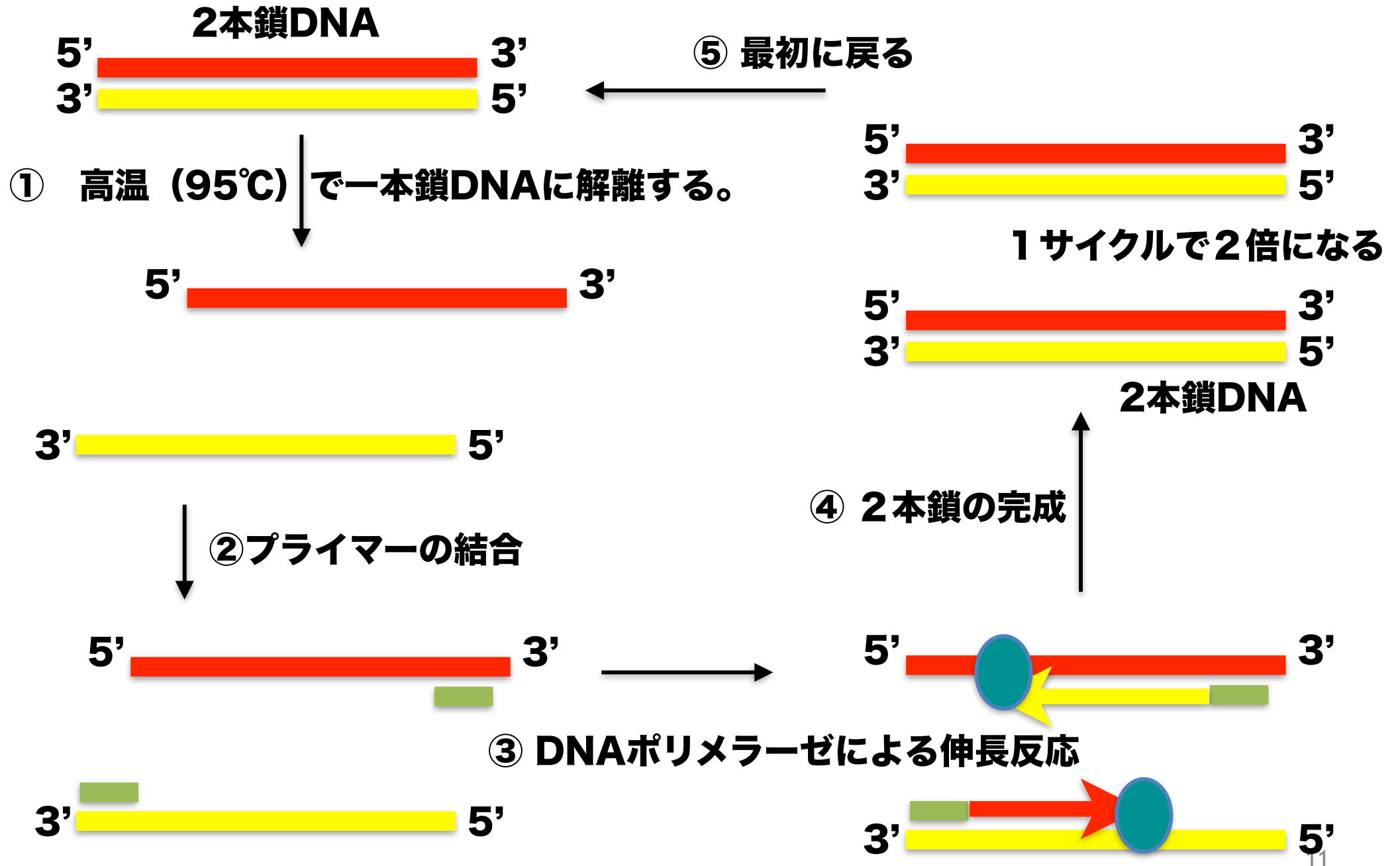
5'-AGTTGCGATCGATCGATCGAT-OH 3'

DNAは **5'->3'** の方向にのびる。

試験管の中で必要な材料をそろえれば、DNAを増やすことができる。ただし、配列がわかっている (プライマーの設計) 必要がある。

# PCR (Polymerase chain reaction) 法

## ポリメラーゼ連鎖反応



# 次世代シーケンサー技術

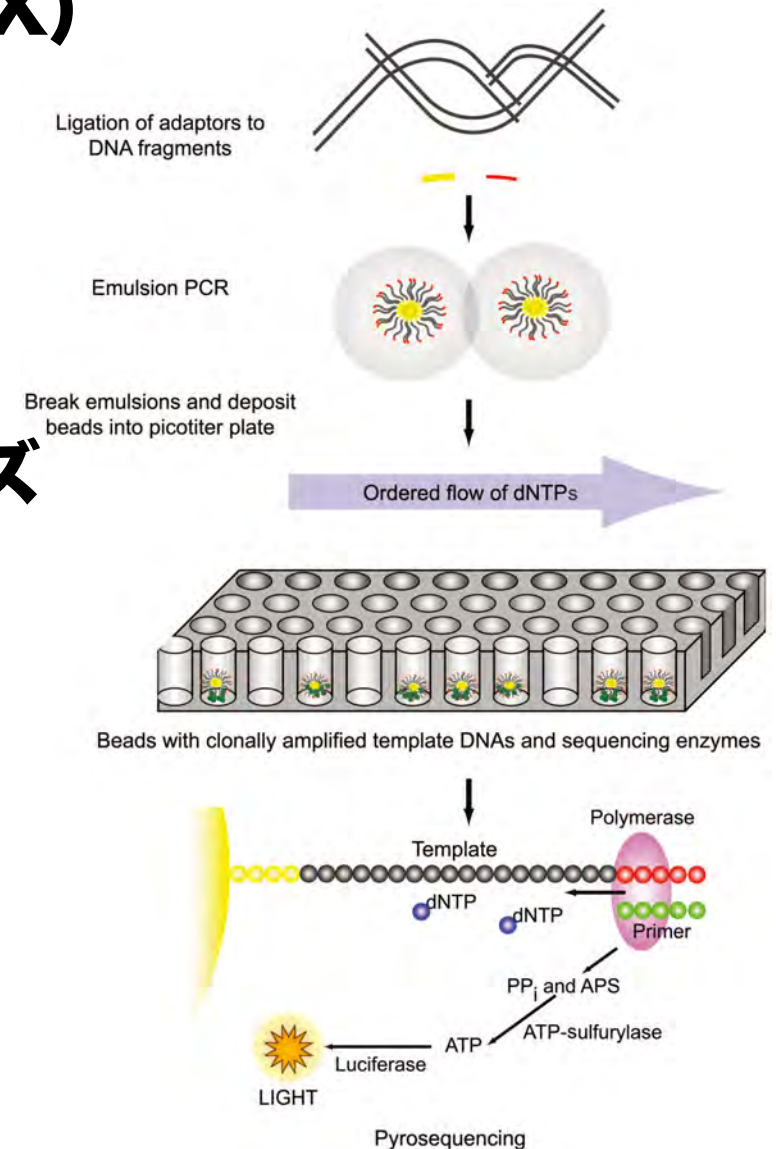
## (Next Generation Sequencing technology)

### 1.パイロシーケンシング (454 GS FLX)

2005年に登場

Emulsion PCRという発想

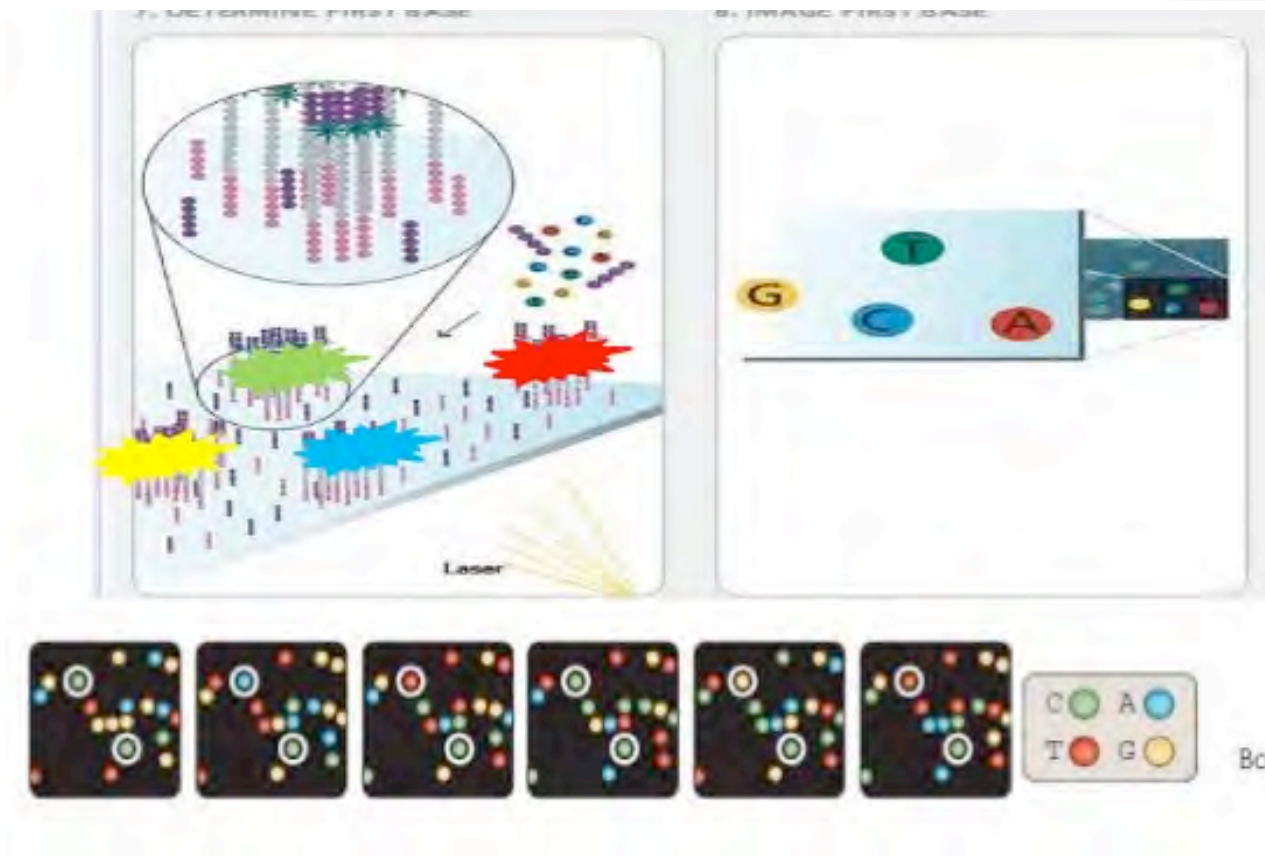
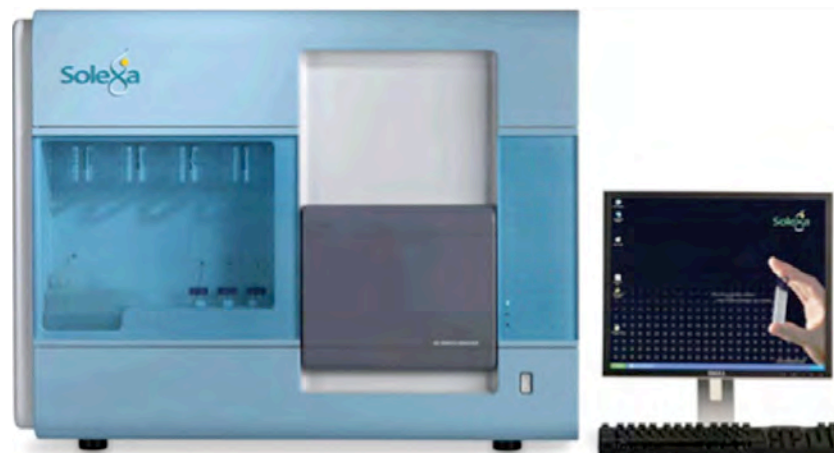
小さい整列された穴にビーズ  
を落とすイメージ  
光をダイオードで検出して



## 2. SBS (Sequence by synthesis)ケミストリー

### Bridge PCRという発想

ヌクレオチドの片方をマスクして、  
1回の反応で止める。

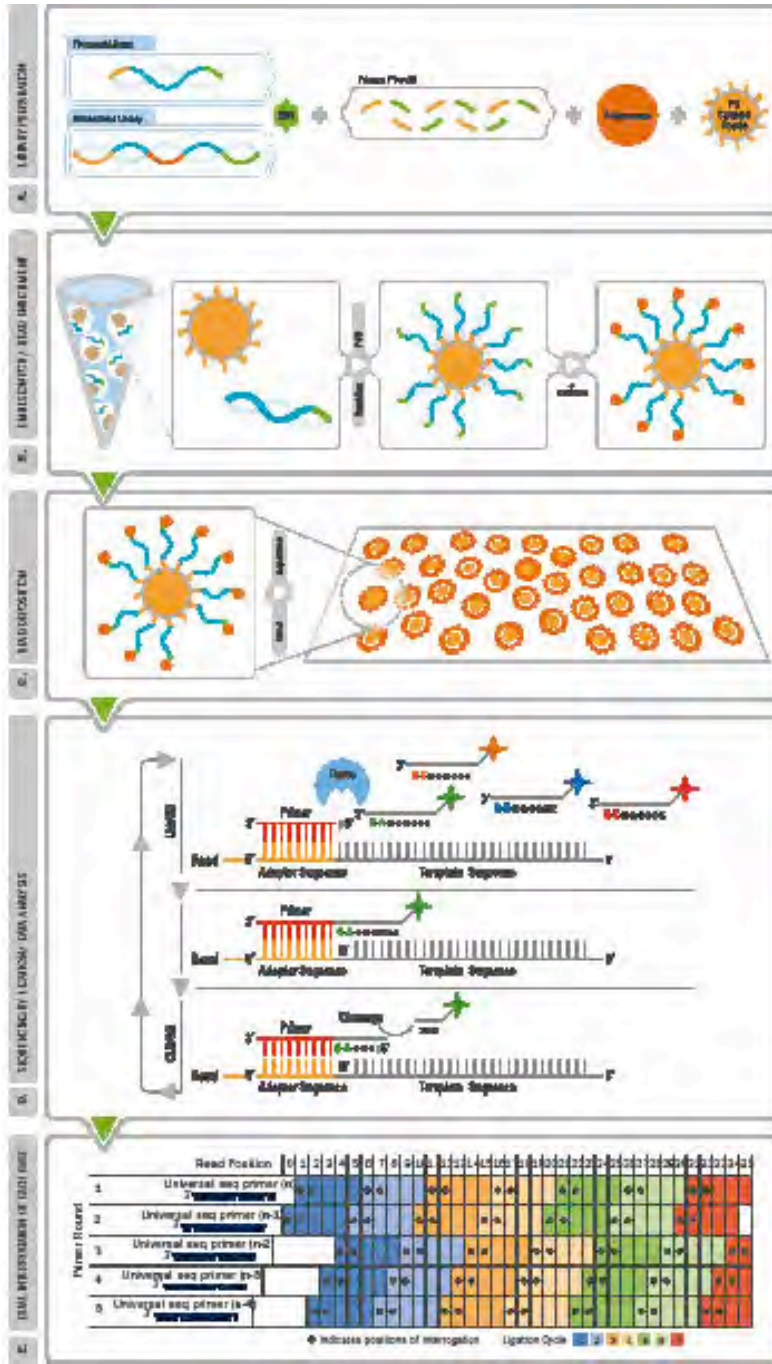


顕微鏡で拡大して、CCD  
カメラで画像解析

大量データを並列に処理  
できる。

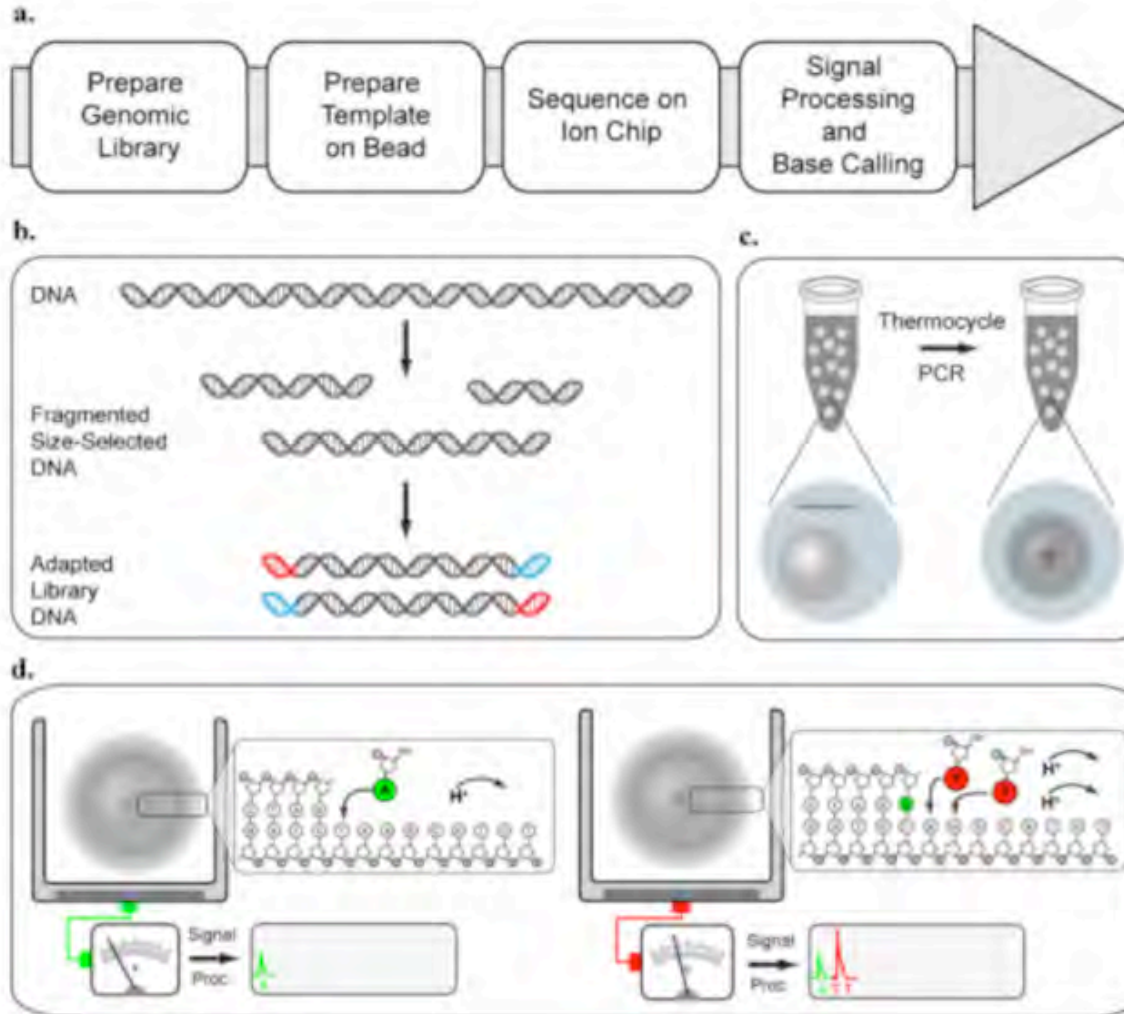


# 3. SOLiDシステム (Sequence of Ligation)

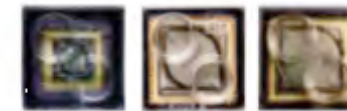


顕微鏡(オリンパス)で拡大して、CCDカメラ(浜松フォトニクス)で画像解析  
 Emulsion PCR (454で用いられた技術)  
 スライドガラスにビーズを結合  
 ビーズ上のDNAに合成オリゴをLigationで  
 つなげてシーケンスを行う

# 4. Ion Torrent / Ion Proton 半導体シーケンサー



Ion PGM™ Sequencer



Ion Proton™ Sequencer



## Emulsion PCR

**DNA合成反応では、 $H^+$ がリリースするので、酸性化が起こる。この変化を半導体で検出する。**

**――> 蛍光試薬は不要**

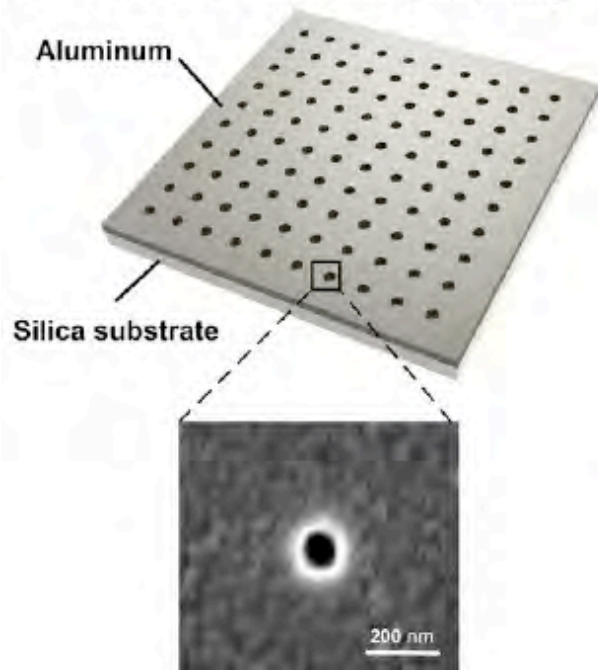
**pH変化だけなのでコストは下がる。**



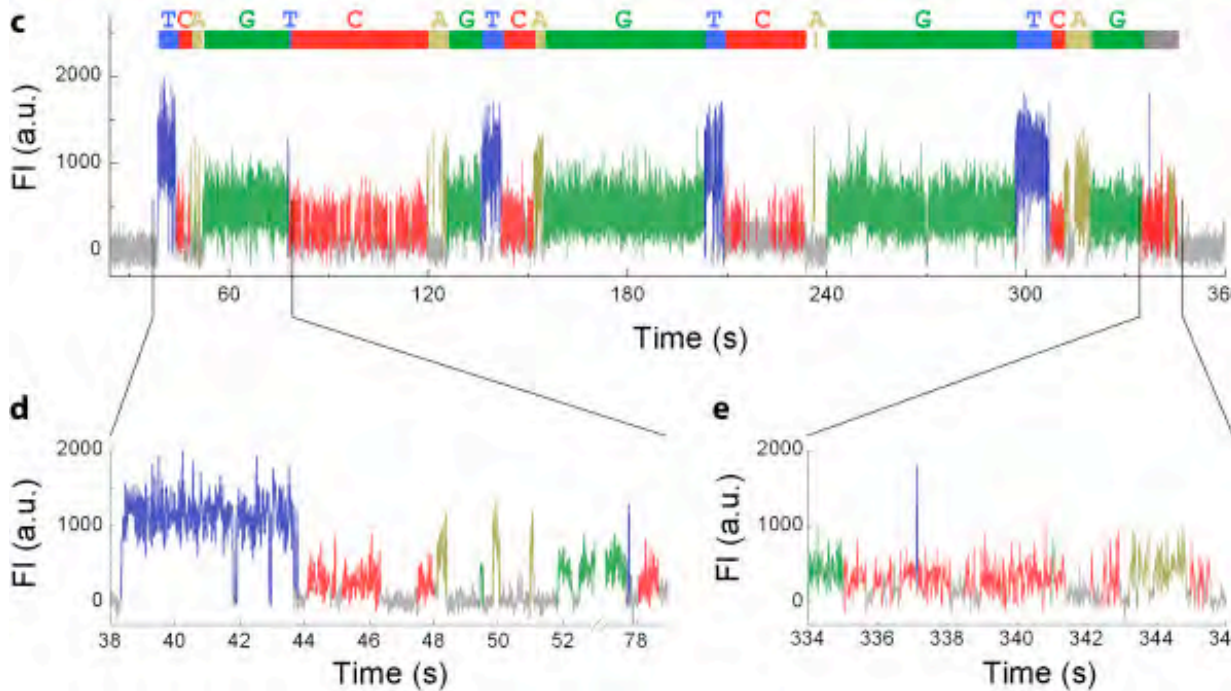
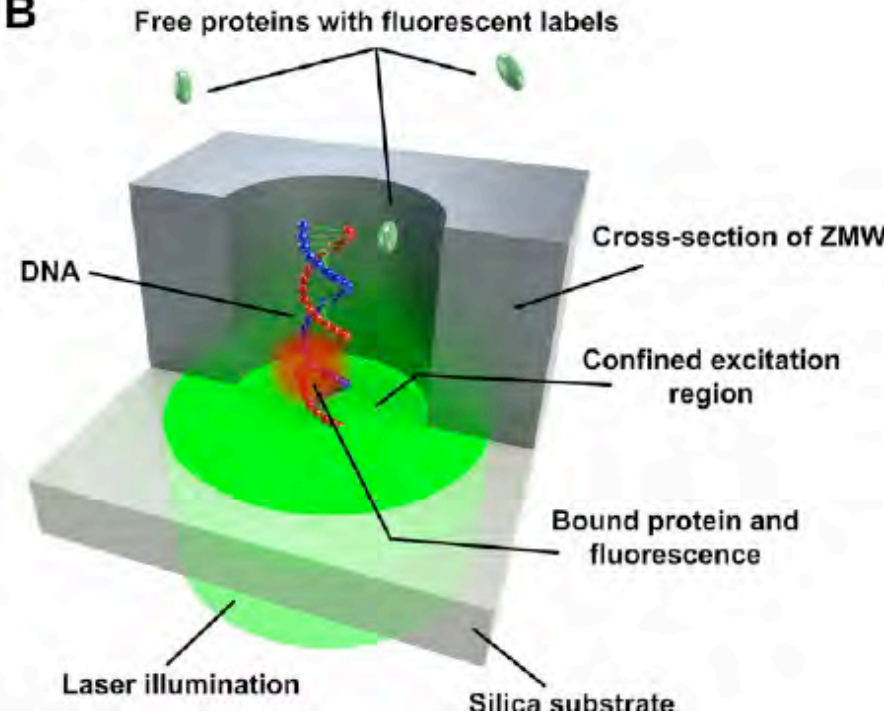
# 5. 一分子リアルタイムDNAシーケンサー

PacBio

**A** Zero Mode Waveguide (ZMW) Array



**B**



**Zero-mode waveguide**  
**連続信号をビデオモニターで記録する。**  
**1分子のシーケンス!**

# 6. ナノポアシーケンサー

大きな機器 (1億円) は不要

USBに直接差し込んで使える、使い捨て  
サイズ (10万円)

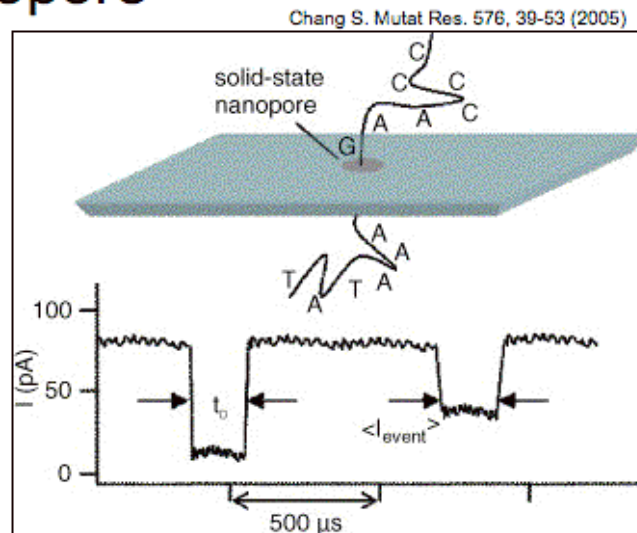
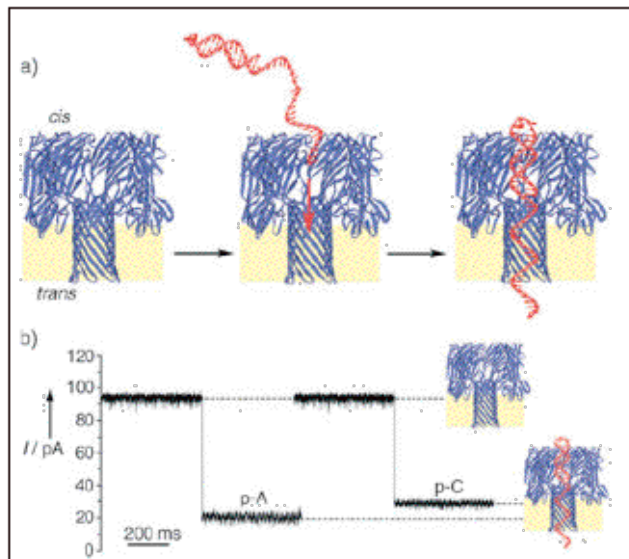
小さいので、オンサイトでの利用が可能



## Nanopore

### Synthetic nanopore

Collins (U of Maine)  
Davis (Stanford U)  
Ramsey (U of North Carolina)  
Golovchenko/Branton/Deamer (Harvard)  
Timp (UIUC)



### Biological nanopore

Ghadiri/Bayley (Scripps/Oxford)

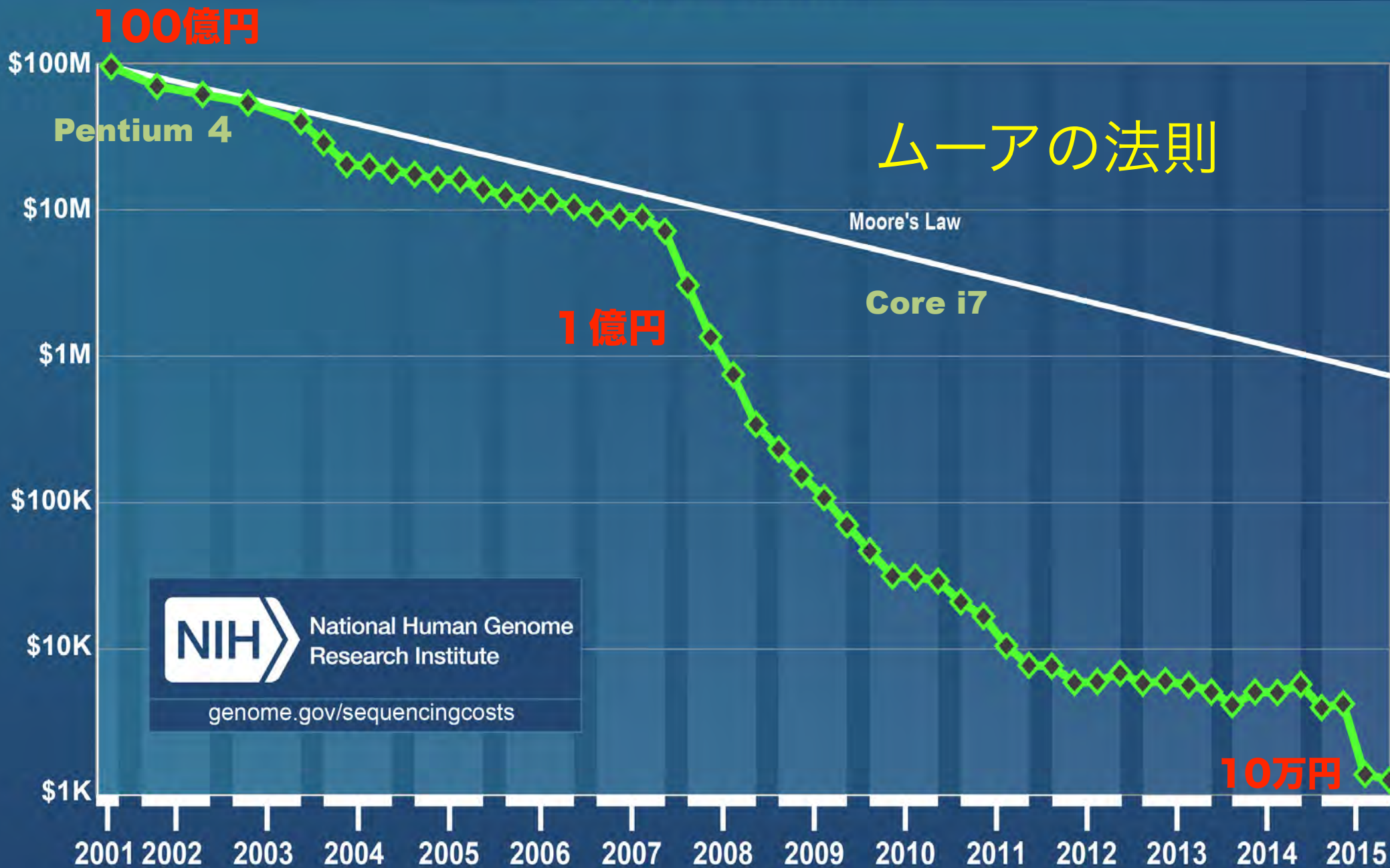
### Nanopore with modified analyte

Benner (U of Florida)  
Marziali (U of British Columbia)  
Meller (Rowland Inst)  
Ju (Columbia)

1分子をシーケンスする  
DNA輸送タンパク質を利用  
電気信号を検出する



# Cost per Genome



# ムーアの法則(Moore's law)とは

インテル創業者の1人であるゴードン・ムーアが、1965年に自らの論文上で唱えた

**「半導体の集積率は18か月で2倍になる」**

という半導体業界の経験則です。

つまり、半導体の微細化技術の向上により、半導体の最小単位であるトランジスタの数が、同じ面積でも18ヶ月で2倍になるということです。

つまり、CPUの性能も18ヶ月ごとに2倍になる。

性能あたりのコストは半分に下がる？

次世代シーケンサーの進化によるコストダウンは、このムーアの法則をはるかに超えていくものであった。

でも、そろそろ、頭打ちかな…

**2003年 ヒトゲノム塩基配列（28億6千万基）99%解読**

**解読期間：13年      費用：約30億\$**

**2008年 次世代シーケンサーの登場**

**2014年 米国「1000ドルゲノムプロジェクト」1人分のDNA解析を達成**

**解読期間：2日      費用：1000\$**

**（ただし、機械は50人分を1回で解析するので、1000 X 50 = 50000\$ は必要）**

**イルミナ社は、100\$ゲノム解析を目指す。**

**近い将来、1万円でヒトゲノムが決まる時代が来る！**

# ゲノムから何が分かるか？

アンジェリーナ・ジョリー



乳房、卵巣、卵管を摘出  
2013年

**統計学**をしっかりと学  
びましょう。

Rの学習

BRCA1 and BRCA2  
Hereditary Breast and Ovarian  
Cancer

**breast cancer susceptibility gene 1**  
がん抑制遺伝子  
核内で遺伝子組換え修復に関わるタンパク質をコードす  
る遺伝子。

損傷 → 修復の繰り返し

単に恐れるのではなく、確かな知識を身につけて  
状況判断し、その時の最善と判断する（これは各  
人で異なる）行動をとる。それが、本当の意味で  
“知恵”と“勇気”もった文化人だと思います。

人生に正解、不正解はありません。

統計データを読み解いて全体の傾向を見るだけの  
時代から、データを個別にとらえて、対応する新  
しい時代がそこまで来ています。



# ビッグデータから何が分かるか？

例えば、日本人全員のゲノム配列が分かれば？

地球上の生物すべてのゲノム配列が分かれば？

問題が与えられて、それを解く能力も大事ですが、何が面白い問題なのか？を見つける能力はもっと大切です。色々なことを想像して、何が足りないのか？どんな視点で見れば面白いか？あれこれ考え（時間制限はありません）で、自分の問題を見つける。後は、残りの時間でそれを解くためにひたすら走る（より細かい問題を解く連続になる）。

ゲノム情報には、その生物のすべての情報が入ってるはずですが。面白いと思いませんか？