

農林水産技術会議資料

次世代ゲノム解析技術が拓く 新しい世界

2009. 6. 16

榊 佳之
(豊橋技術科学大学)

ゲノム研究の流れ

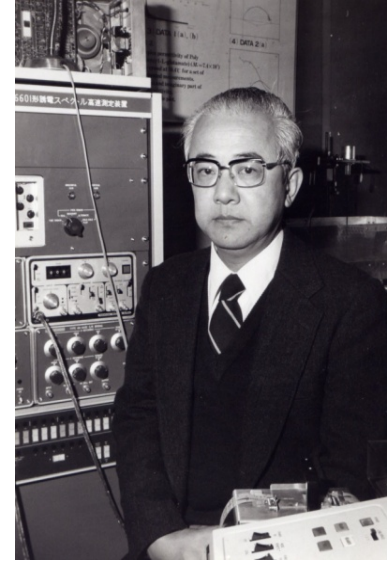
1977 DNAシーケンシング技術開発

1980s DNA自動シケンサーの開発

1953 DNA2重らせん構造



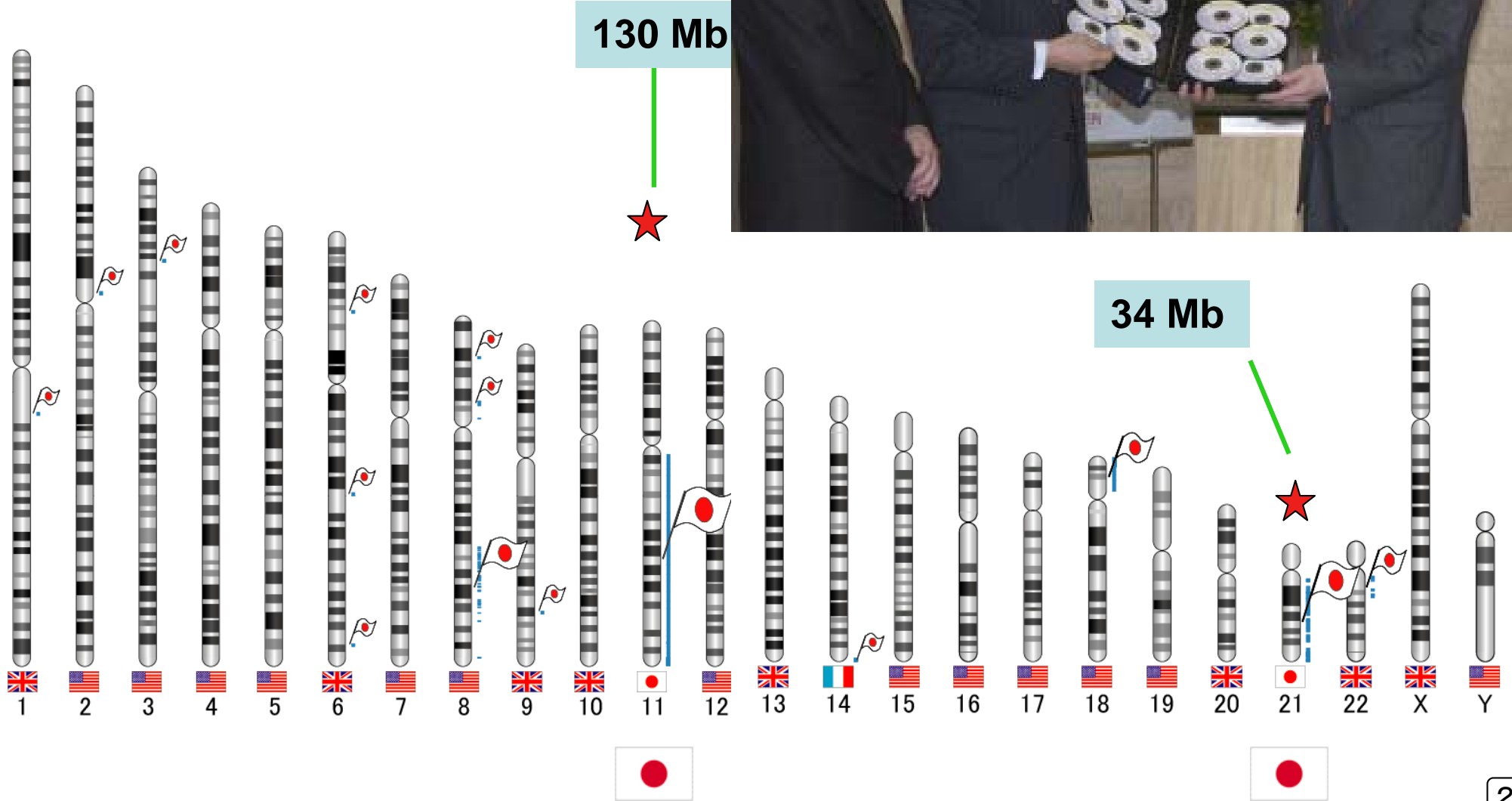
1972 遺伝子組み換え技術



1991ー ヒトゲノム計画

ヒトゲノム解読完了 (2003年4月)

生命科学の転換点

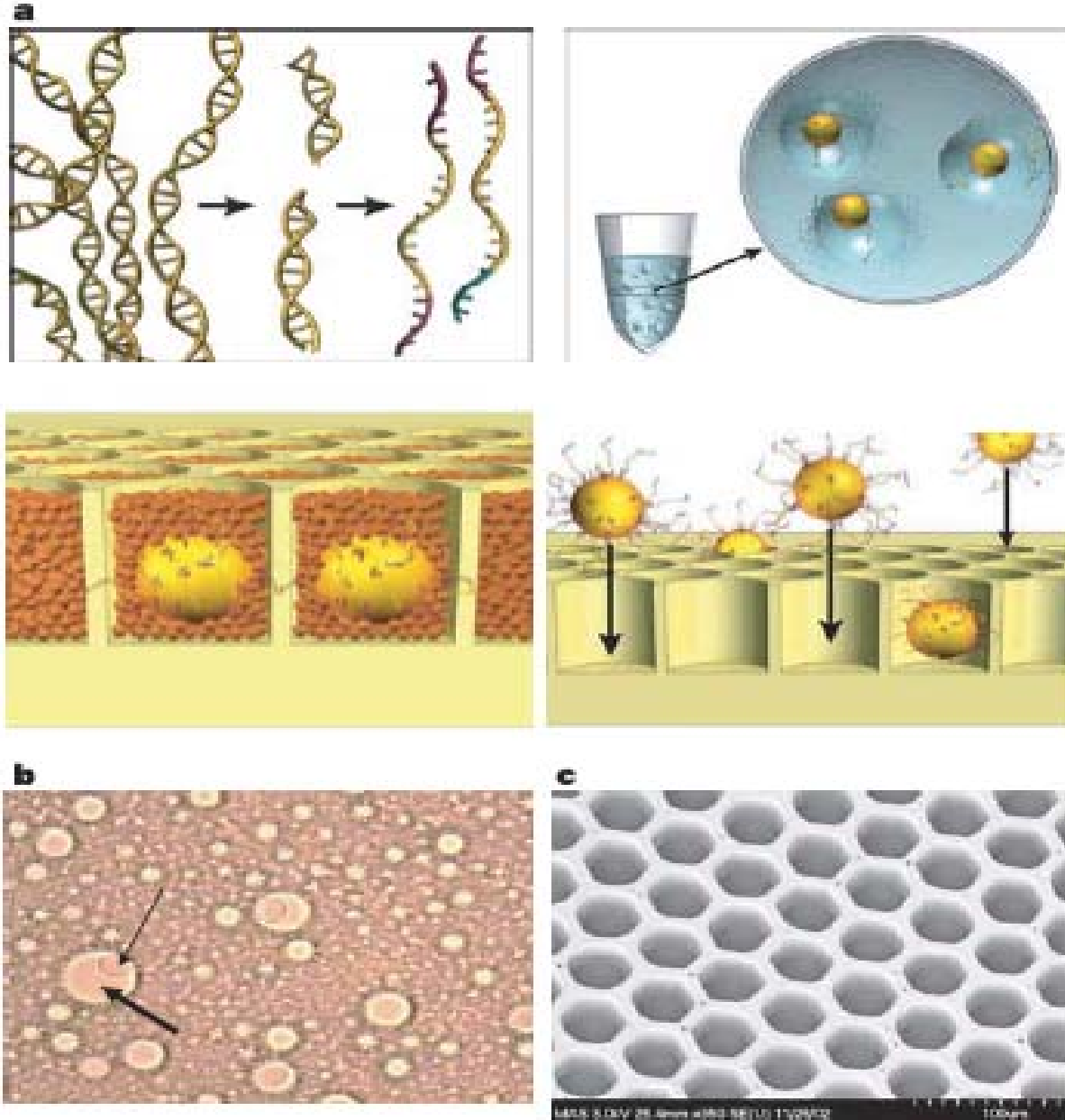


自動シーケンサーの性能の進歩 (ヒトゲノム解読時)

| 機種 | 市販年 | 解読塩基数 ／サンプル | 1度に解読でき るサンプル数 | 解読時間 (時間) | 総解読塩基数 ／台・日 |
|---------------------|------|----------------|-------------------|--------------|----------------|
| ABI373(スラブ) | 1990 | 400 | 24 | 12 | 1万 |
| ABI377(スラブ) | 1995 | 500 | 96 | 8 | 14万 |
| ABI3700(キャピラリ) | 1998 | 600 | 96 | 3 | 46万 |
| ABI3730(キャピラリ) | 2003 | 850 | 96 | 2 | 98万 |
| MegaBACE1000(キャピラリ) | 1998 | 550 | 96 | 1.5 | 84万 |
| MegaBACE4000(キャピラリ) | 2001 | 550 | 384 | 2 | 253万 |
| MegaBACE4500(キャピラリ) | 2004 | 850 | 384 | 3 | 261万 |

解読塩基数／サンプルと総解読塩基数／台・日はフル稼動したときの目安を示している。

超高速ゲノム解析システム Genome Sequencer 20 system



次世代型シーケンサーの大発展

これら次世代シーケンサーではさらなる改良が進行中である。

| | 普及年 | 配列長 (塩基数) | 解析試料数* | 解読総塩基数 | 30億塩基の収集/台 (1×分のヒトゲノム) |
|-----------|-------|--------------|-----------|------------|---------------------------|
| PacBio# | 2010? | ~ 1,500 | - | ~ 1000億/時 | 約 2分 |
| HeliScope | 2008 | 20 ~ 45 | 90,00万/時 | ~ 10 億/日 | 約 3日 |
| ABI SOLiD | 2007 | 25 ~ 35 | 30,00万/実験 | ~ 20 億/10日 | 約 15日 |
| Solexa | 2007 | 25 ~ 50 | 10,00万/実験 | ~ 10 億/5日 | 約 15日 |
| 454FLX | 2006 | ~ 250 | 40万/ 日 | ~ 1 億/日 | 約 30日 |
| 島津 DeNOVA | 2006 | ~ 1,000 | 4,608/ 日 | ~ 0.05 億/日 | 約 600日 |
| ABI3730xl | 2002 | ~ 800 | 2,304/ 日 | ~ 0.02 億/日 | 約 1,500日 |

Korf J et al.: PNAS 105, 1176-1181 (2008)

*Solexaは1回の実験に5日
*SOLiDは1回の実験に10日



島津 DeNOVA



454FLX



ABI SOLiD

The complete genome of an individual by massively parallel DNA sequencing

Nature, 2008

Dr Watson's base pairs

```
TCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTGCTGACTGCGGCATATTATATTAGCTGAT
AGCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTCG
TTAGCTGATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGCATC
CGTACTCGTACGTCGACTGCGGCATATTATATTAGCTGATCGTGAATTTCTGAATGCTAGCTGT
CCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTGCTGACTGCGGCATATTATATTAGCT
GCTAGCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACG
ATATTAGCTGATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGC
CTTCGTACGTCGACTGCGGCATATTATATTAGCTGATCGTGAATTTCTGAATGCTAGCTGTTGTG
CGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTGCTGACTGCGGCATATTATATTAGCTGATC
GCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGCATCGTGAATTTGTGAATTTAGTATG
CGAGCTGCTGCAGAGCTTCGTACGTGCTGACTGCGGCATATTATATTAGCTGATCGTGAATTTCTGA
TTTGTATGGGCCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTGCTGACTGCGGCATA
GATTTCTGAATGCTAGCTGTTGTGAATTTAGTATGGGCCCTCGTTATGATTTCTGAATGCTAGCT
GGCCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTGCTGACTGCGGCATATTATATTAG
ATGCTAGCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTA
TTATATTAGCTGATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTTAGTATGGGCCCTCGTTACG
AGCTTCGTACGTCGACTGCGGCATATTATATTAGCTGATCGTGAATTTCTGAATGCTAGCTGTTG
CTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTGCTGACTGCGGCATATTATATTAGCTGA
TAGCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTCG
ATTAGCTGATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGCAT
TCGTACGTCGACTGCGGCATATTATATTAGCTGATCGTGAATTTCTGAATGCTAGCTGTTGTGA
TTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTGCTGACTGCGGCATATTATATTAGCTGATCGT
TGTGTGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTCGACTG
GCTG TCGTGAATTTCTGACTAGCTGACTGAGCTTAAAGCTAAATAAATTTTCTGACTGCTG
TAGTATGGGCCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTGCTGATCGTGAATTT
TGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTCGACTGACTG
TCGTGATTTAGTATGGGCCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTCGACTG
CTGACTGCGGCATATTATATTAGCTGATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTTAGTAT
CGAGCTGCTGCAGAGCTTCGTACGTGCTGACTGCGGCATATTATATTAGCTGATCGTGAATTTCTGA
GATTTCTGAATGCTAGCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTGCTGCAG
CTGCGCATATTATATTAGCTGATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTTAGTATGGGCC
CTGCTGCAGAGCTTCGTACGTCGACTGCGGCATATTATATTAGCTGATCGTGAATTTCTGAATGCT
GTATGGGCCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTCGACTGCGGCATCGTACGTCG
GCTGATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGCATCGAG
ACGTGCTGACTGCGGCATATTATATTAGCTGATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTTA
GGCATCGAGCTGCTGCAGAGCTTCGTACGTGCTGACTGCGGCATATTATATTAGCTGATCGTGAAT
GTGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTCGACTGCTG
ATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTG
GCTGACTGCGGCATATTATATTAGCTGATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTTAGTAT
ATTTAGTATGGGCCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTCGACTGCGGCAT
ATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTG
AGCTGATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGCATCGA
TACGTGCTGACTGCGGCATATTATATTAGCTGATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTT
CGGCATCGAGCTGCTGCAGAGCTTCGTACGTGCTGACTGCGGCATATTATATTAGCTGATCGTGAAT
TGTGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTGCTGCAGAGCTTCGTACGTCGACTGCTG
GTGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTTCGTACGTCGACTGCGGCATATTATATTAG
ATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTTC
GCTGACTGCGGCATATTATATTAGCTGATCGTGAATTTCTGAATGCTAGCTGTTGTGAATTTAGTAT
TCGAGCTGCTGCAGAGCTTCGTACGTGCTGACTGCGGCATATTATATTAGCTGATCGTGAATTTCTG
CTAGCTGTTGTGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTTCGTACGTCGACTGCGGCAT
TATTAGCTGATCGTGAATTTAGTATGGGCCCTCGTTACGGCATCGAGCTTCGTACGTCGACTGCGGCAT
```

ワトソン博士

DNA構造発見者



自分のゲノム公開

研究に利用へ

2007.5.30 朝日新聞

【ワシントン】上田俊英「DNAの2重らせん構造を見つけた、1962年にノーベル医学生理学賞を受けた米国の分子生物学者ジームズ・ワトソン博士（79）＝写真＝のDNAの全遺伝情報（ゲノム）が今月末に公開されることがわかった。28日発売の米誌「ニューサイエンティスト」が報じた。これのDNAはA、T、C、Gの4種類の頭文字で表される塩基が約30億（対）連続してできている。その配列に従って体のすべての形作られている。遺伝子はこのDNAのあちこちに散らばっている。DNAの塩基配列全体をゲノムと呼ぶ。

ワトソン博士は2年前に、米バイオ企業「454ライフサイエンス」が提案したゲノムの公開計画に同意。同社は博士の血液をともに87日間、約100万（約1億2千万円）かけて、塩基配列を明らかにした。結果は30日に博士に示され、31日に米国立

保健研究所（NIH）のデータベース上で公開される予定。病気や知性、性格など遺伝子との関係を調べる研究に使われることになるという。

息子2人のうち1人が精神疾患という博士は、病気を防ぐ研究への利用で一人々がより健康になれる」と遺伝子研究を進める意義を強調。「病気の原因が遺伝子とわかる（子どもが何かをできないときに、怒るのではなく、手助けしたいと思うようになるなど、人々がより慈深くなるだろう）」と述べている。

しかし、ゲノムによる「究極の遺伝子診断」が際限なく広がることには不安や異論も根強い。遺伝子治療などの進展によって、遺伝子改変時代が来ると予測する専門家もおり、博士のゲノム公開は議論を呼び起こした。

ワトソン博士と英国の故フランシス・クリック博士によるDNAの2重らせん構造発見（53年）は「20世紀最大の発見の一つ」とい

がん関連遺伝子の探索

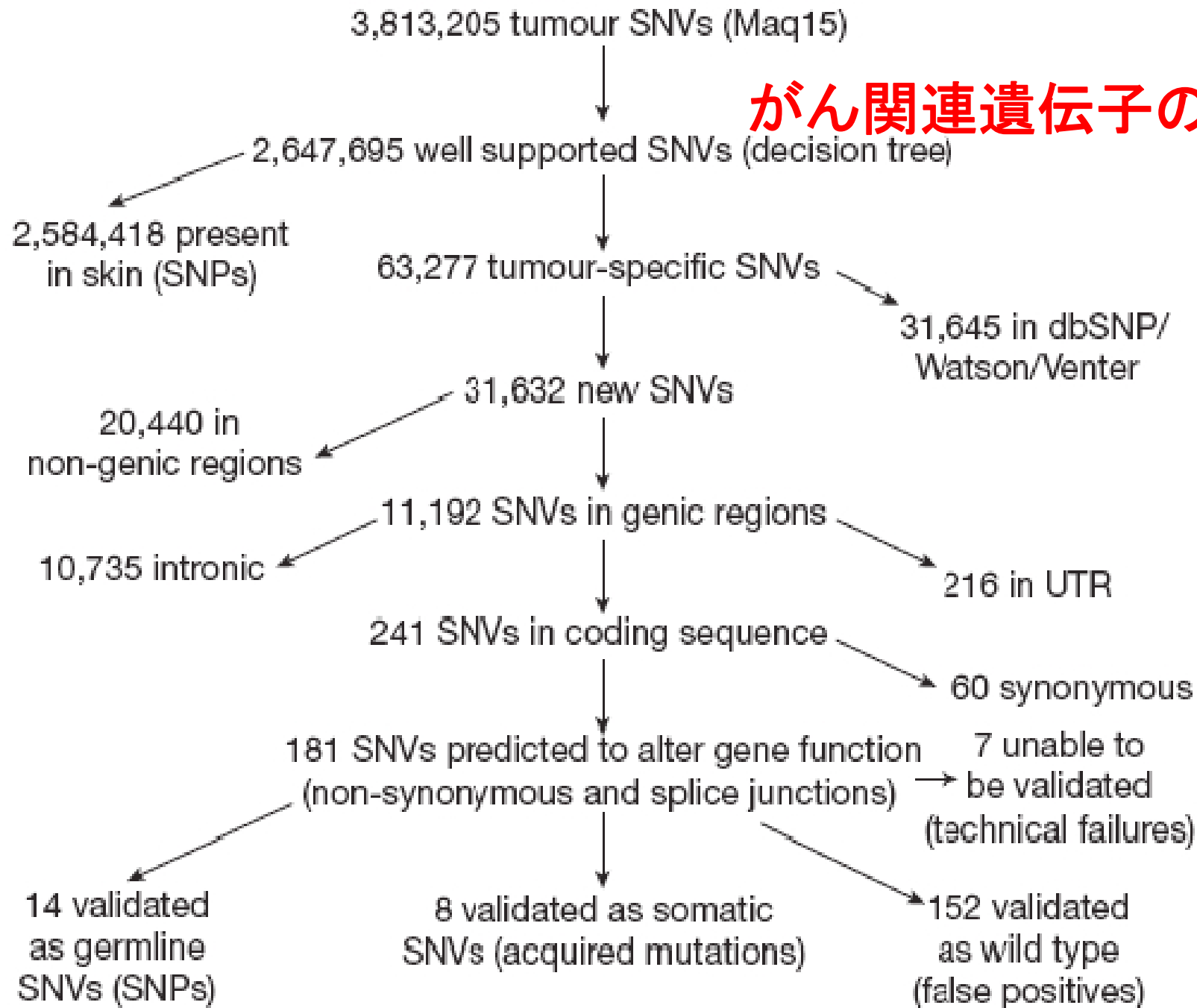
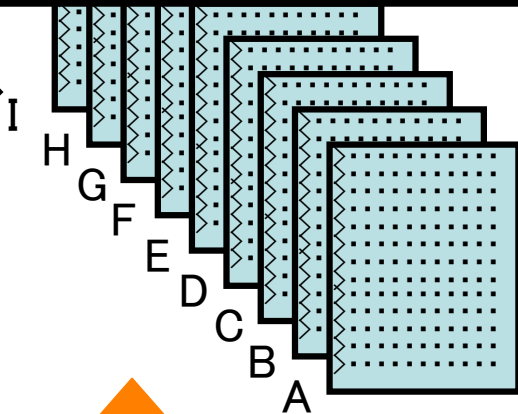


Figure 2 | Filters used to identify somatic point mutations in the tumour genome.

個人のゲノム情報

ヒトゲノム配列へのマッピング



| | | | |
|---|-------|-------|-------|
| A | | | |
| B | | | |
| C | | | |
| D | | | |
| E | | | |
| F | | | |
| G | | | |

ヒトゲノム再シーケンス 1億人(日本人)

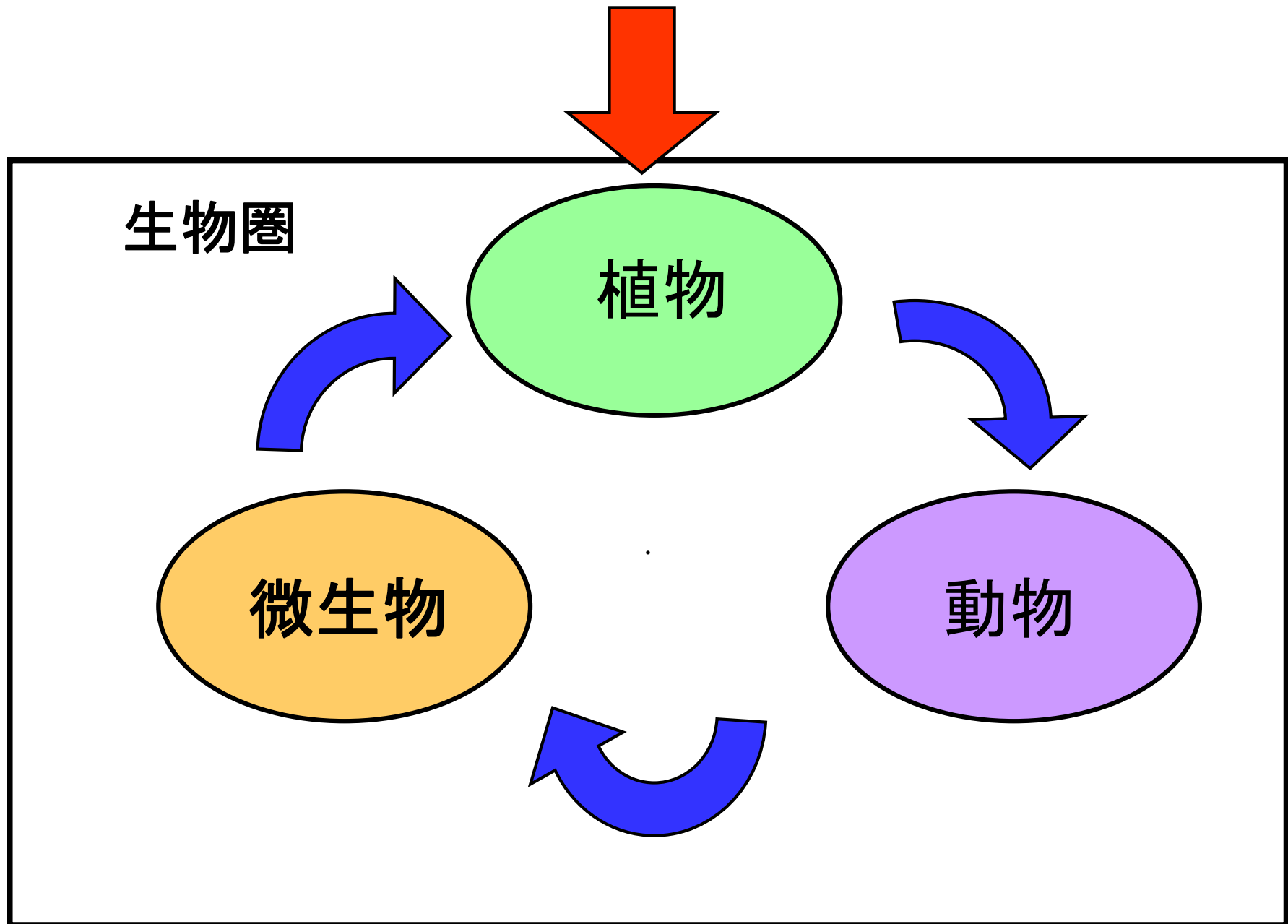


次世代シーケンサー

病気との関連解析

SNPの発見

(太陽エネルギー)



微生物の世界が見えだした

$\leq 1\%$



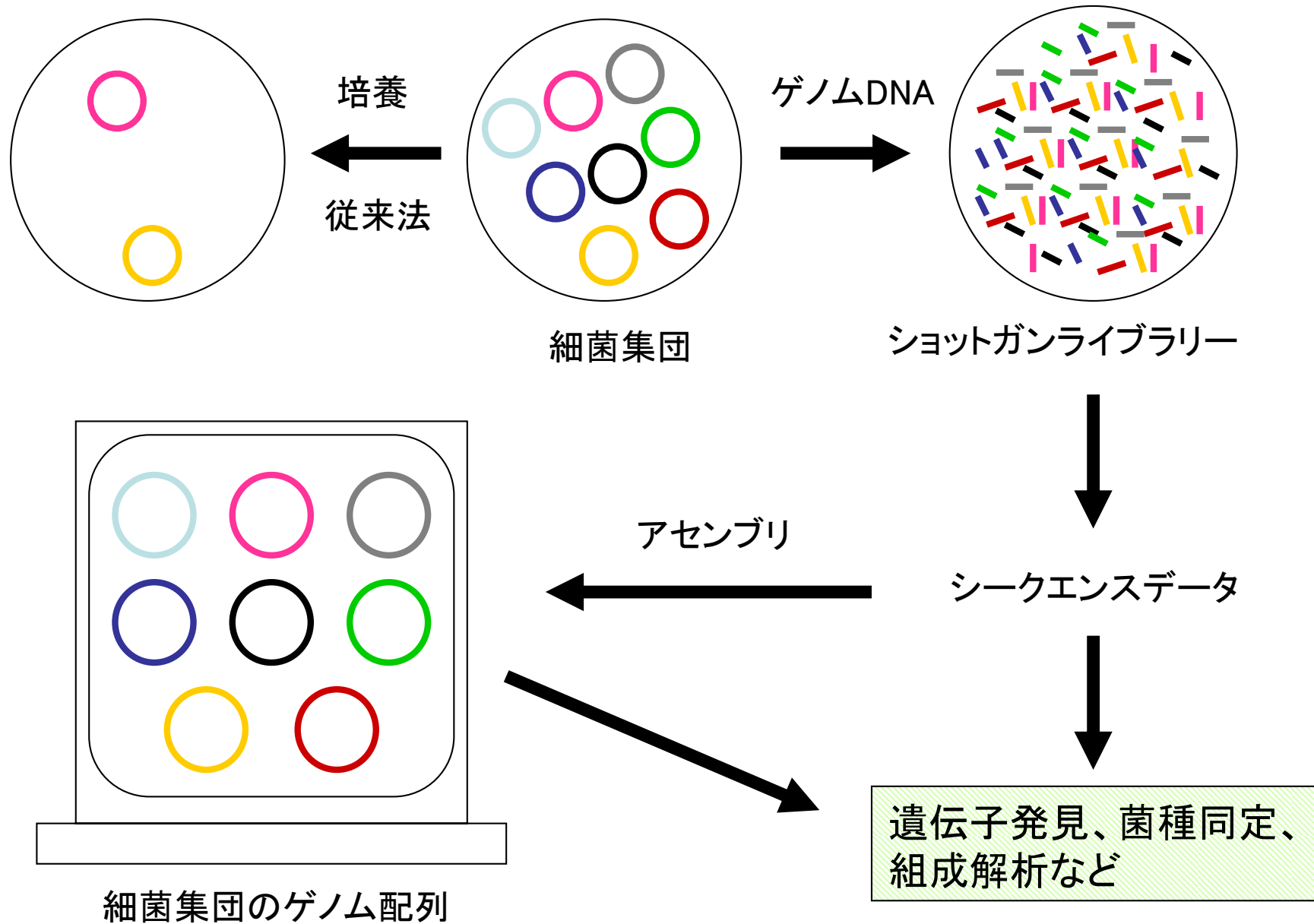
難培養性

$99\% \leq$

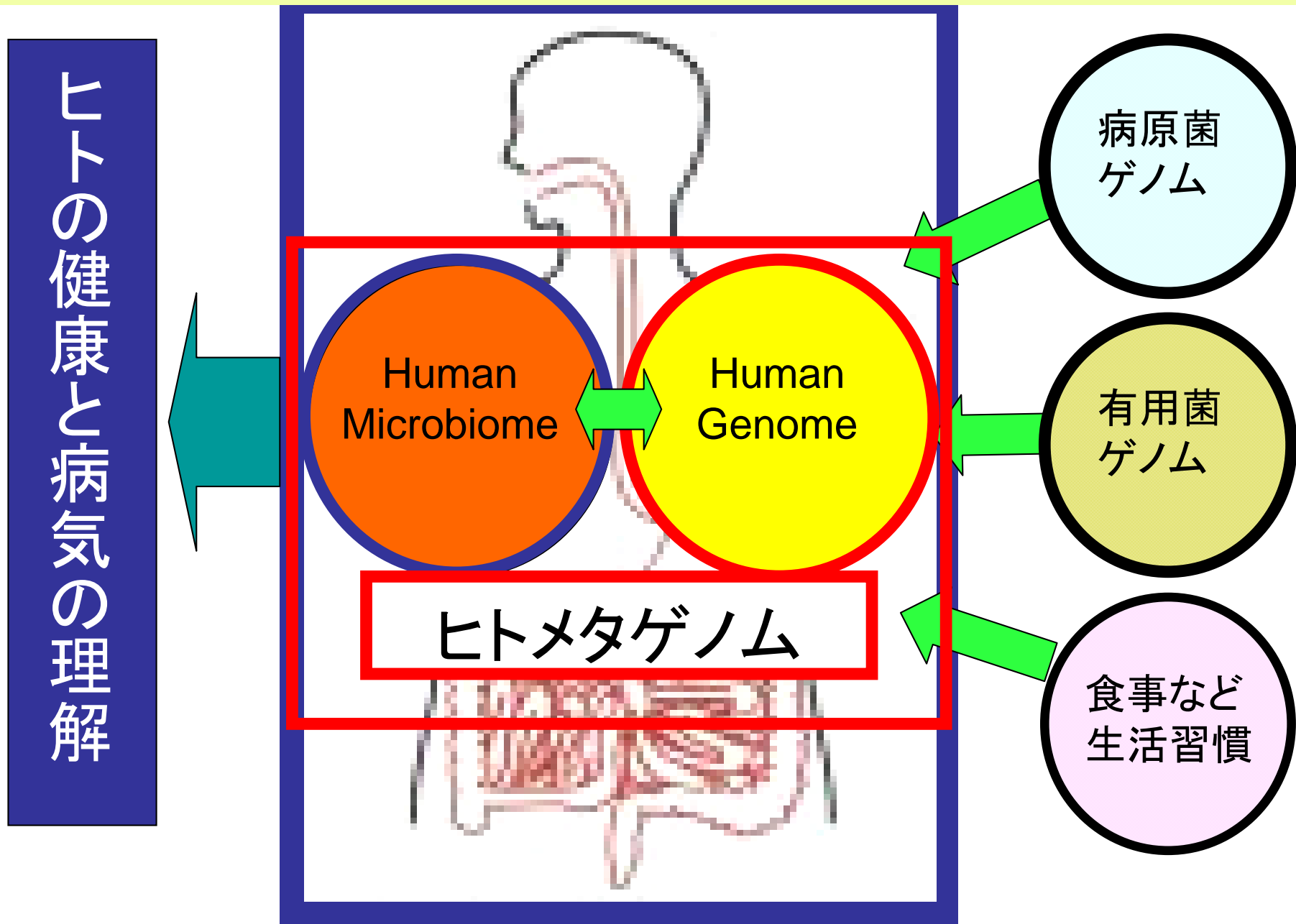
健康
環境
エネルギー

環境生息微生物
群など細菌集団の
ゲノム解析

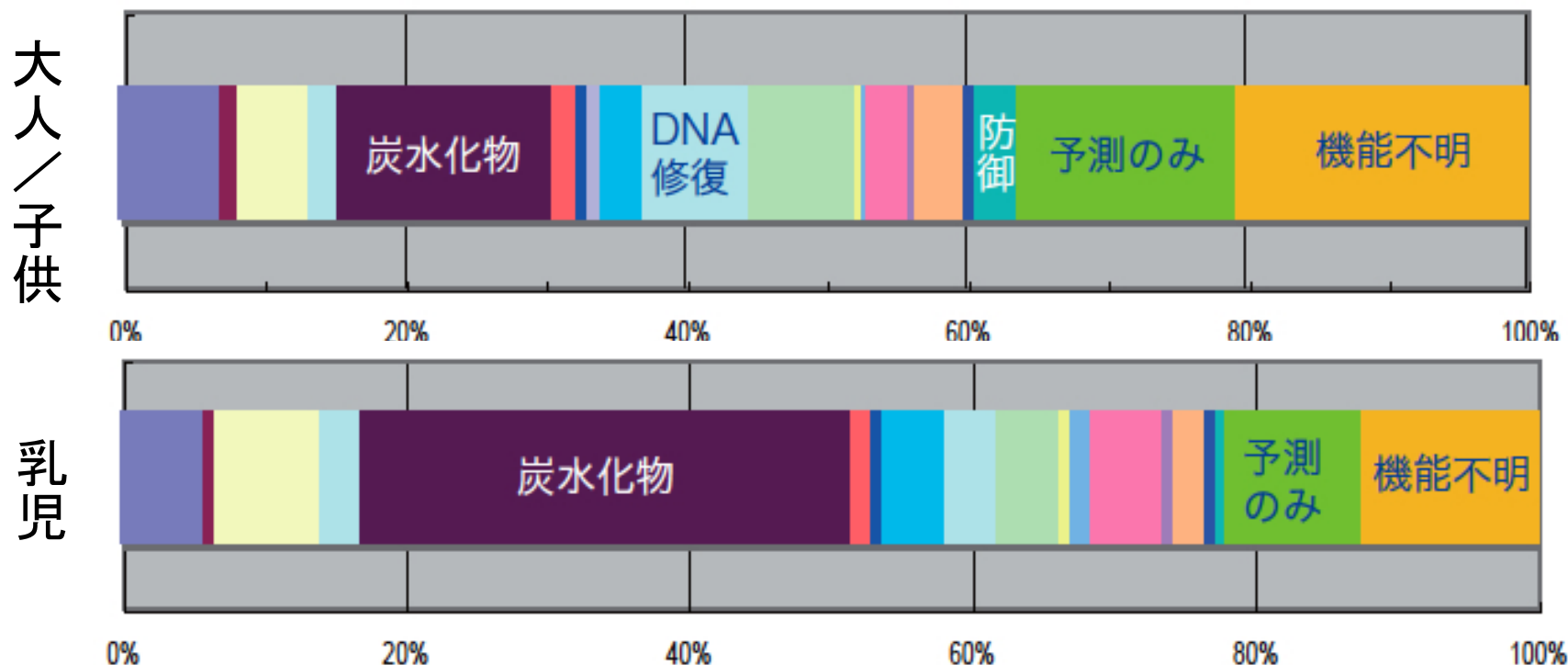
細菌のメタゲノム解析



ヒトはヒトゲノムとヒト常在菌ゲノムからなる superorganism (超有機体)



メタゲノム解析からわかった腸内細菌叢のはたらき

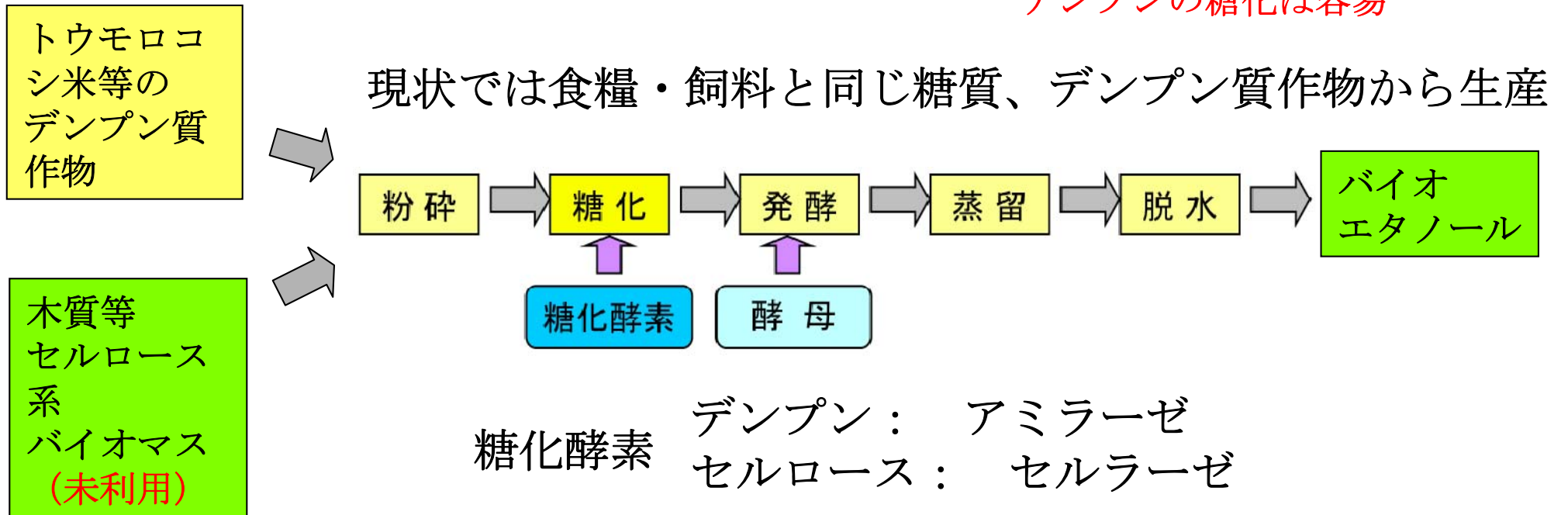


- エネルギー生産、変換
- 細胞分裂
- アミノ酸代謝と輸送
- 核酸代謝と輸送
- 炭水化物代謝と輸送
- 補酵素代謝と輸送
- 脂質代謝と輸送
- 転写
- 翻訳
- DNA複製、修復、組み換え
- 細胞壁・膜合成
- 細胞運動
- 転写後修飾、シャペロン
- 無機イオン代謝と輸送
- 二次代謝物質生合成、異化、輸送
- シグナル伝達
- 細胞内輸送と分泌
- 防御機構
- 一般的な機能予測のみ
- 機能不明

セルロース系バイオマスからのバイオエタノール生産の問題点

バイオエタノール製造プロセス

デンプンの糖化は容易

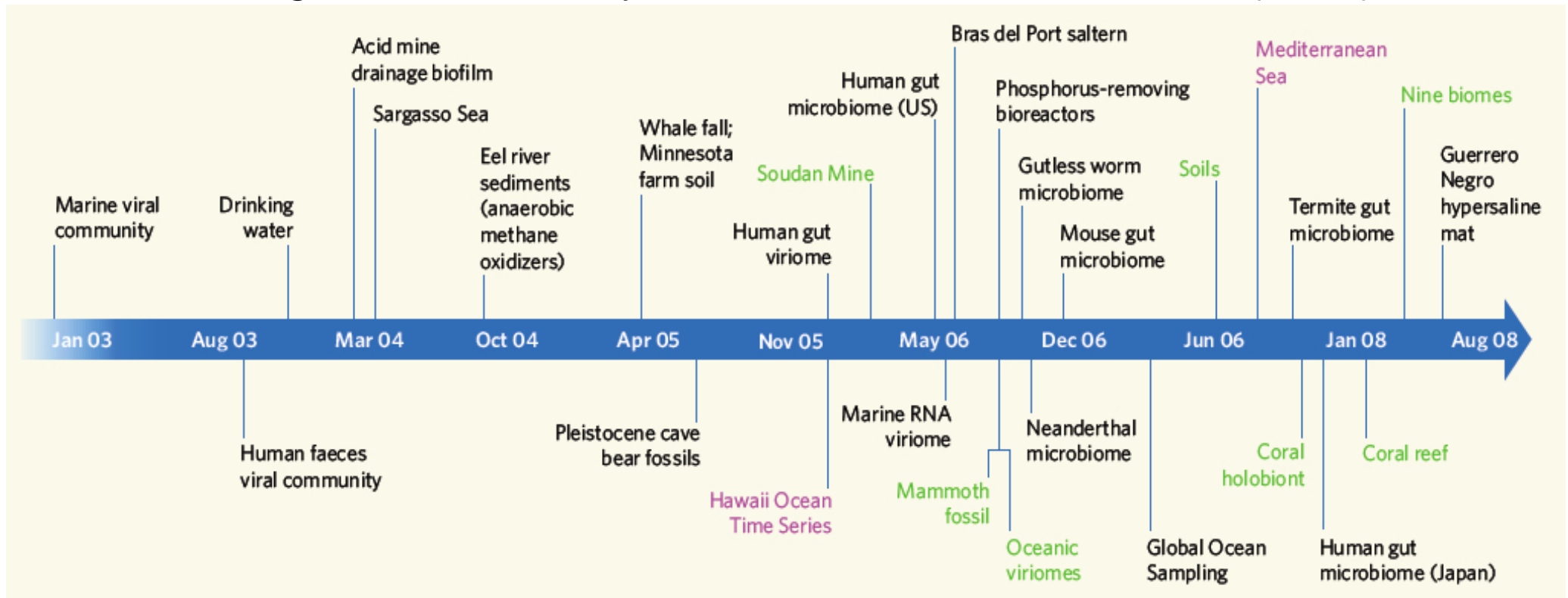


難分解性のセルロースの糖化には強力な（効率的な）セルラーゼが必要

自然界で効率良くセルロース資源を利用しているシロアリなどの共生微生物群の酵素の利用に期待

Timeline of sequence-based metagenome projects since 2003

Hugenholtz P and Tyson GW: Nature 455, 481-483 (2008)



3730 dye-terminator shotgun sequencing (black)

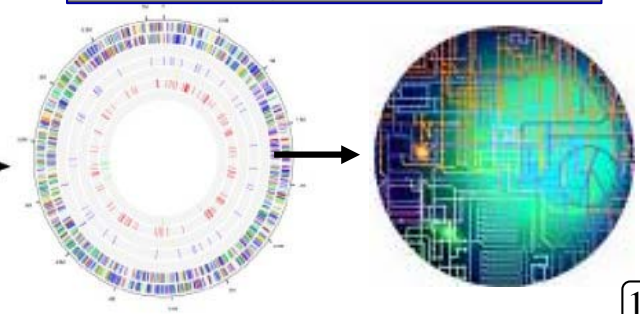
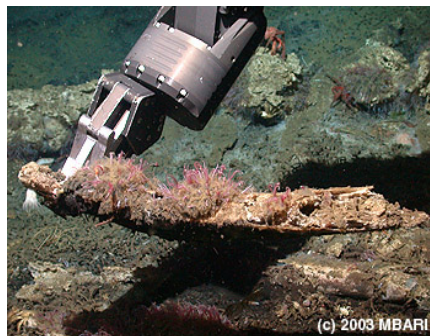
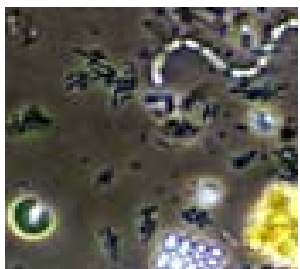
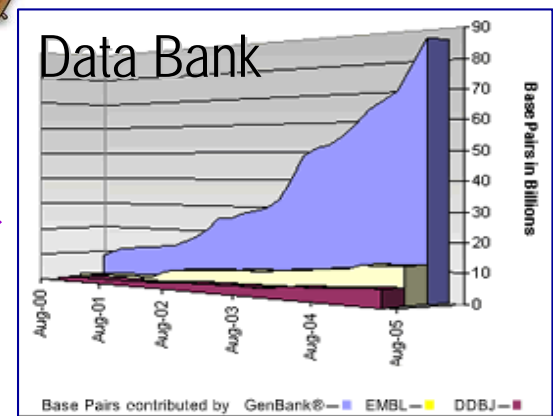
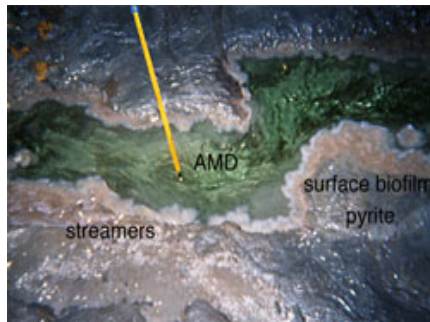
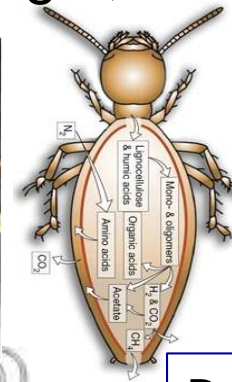
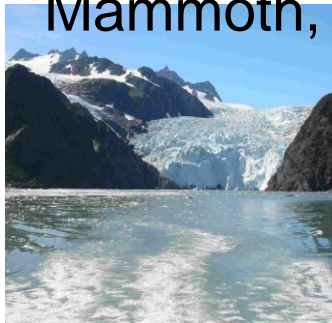
Fosmid library sequencing (pink)

454 Pyrosequencing (green)

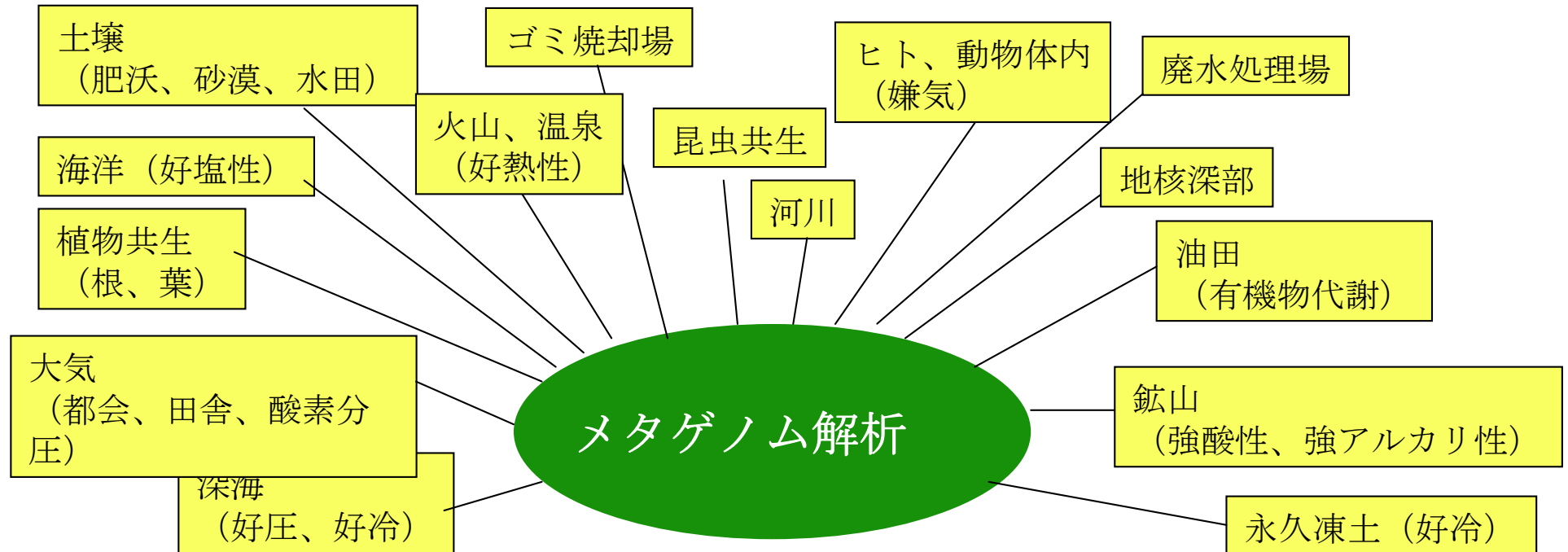
What kinds of projects are being conducted?

> 130 large and small metagenomic projects are going on around the world, including:

Soil, Air, Ocean, Deep sea, Whale falls, Volcanic, Extreme environments, Hot springs, Human gut, Oral cavity, Termite gut, Bovine, Water bear, Mammoth, Fungal,



自然環境中の細菌メタゲノム研究



これまでの10万倍の新規細菌、遺伝子、代謝物の発見

- 各環境細菌種、代謝、遺伝子組成の比較解析⇒環境の特徴、細菌-細菌間、細菌-環境間における相互作用及び生存戦略の総合的解明 (環境Genomics)。
- 有用細菌、遺伝子、代謝物の発掘と医療及び産業有用物質の開発。
- 石油代換え人工細菌などの作成 (合成 Genomics)。
- 海、土壌、河川などの環境汚染の浄化及び診断システムの開発。
- ヒトの健康維持への細菌の活用 (プロバイオティクス、生活改善薬)

J. Craig Venter

I N S T I T U T E

Synthetic Biology

Group Leader: Hamilton O. Smith, M.D.

Biological Energy

Group Leader: Hamilton O. Smith, M.D.

Nature (2002) 420:350

**Hydrogen production from water,
Ethanol and Butanol from cellulose**

Synthesis of new genomes and bacteria carrying the best gene sets for specific targets.

Venter aims for maximum impact with minimal genome

Erika Check, Washington

Not for the first time, geneticist Craig Venter's latest wheeze has set US biology abuzz. This time he has reignited the debate over open publication of research results by declaring that he may not release all of the details of his new project.

On 21 November, Venter said he intends to synthesize a bacterial genome from scratch.

Synthetic Genomics

the methods for this work, in case they were used to make biological weapons. "Depending on what is happening in this field, we may not disclose all these details," he says.

The planned work will build on an earlier project in bacterial genomics that he started at

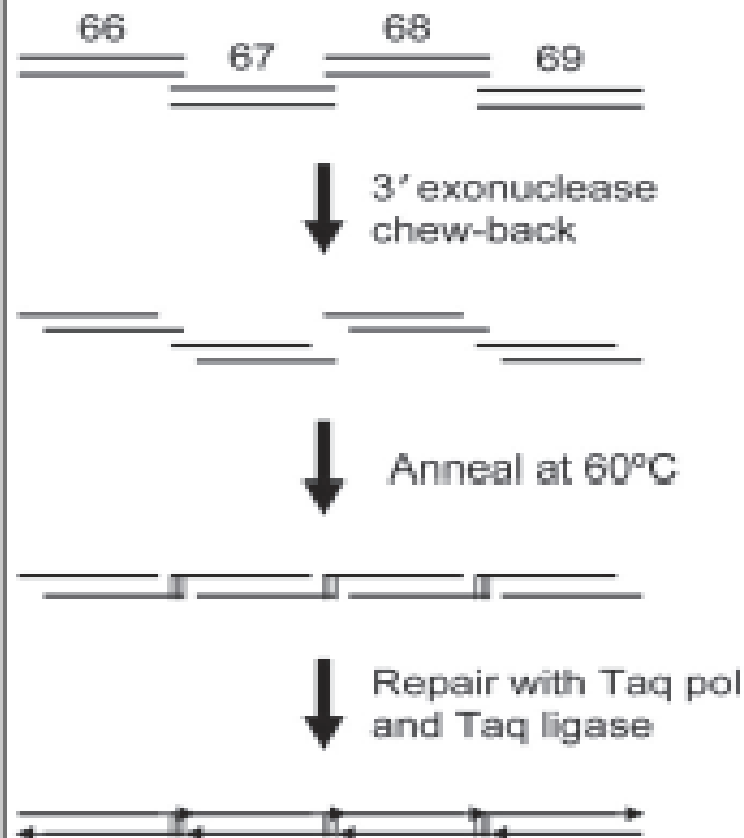


Good news? Critics of Craig Venter's project claim that he is planning to create a new form of life.

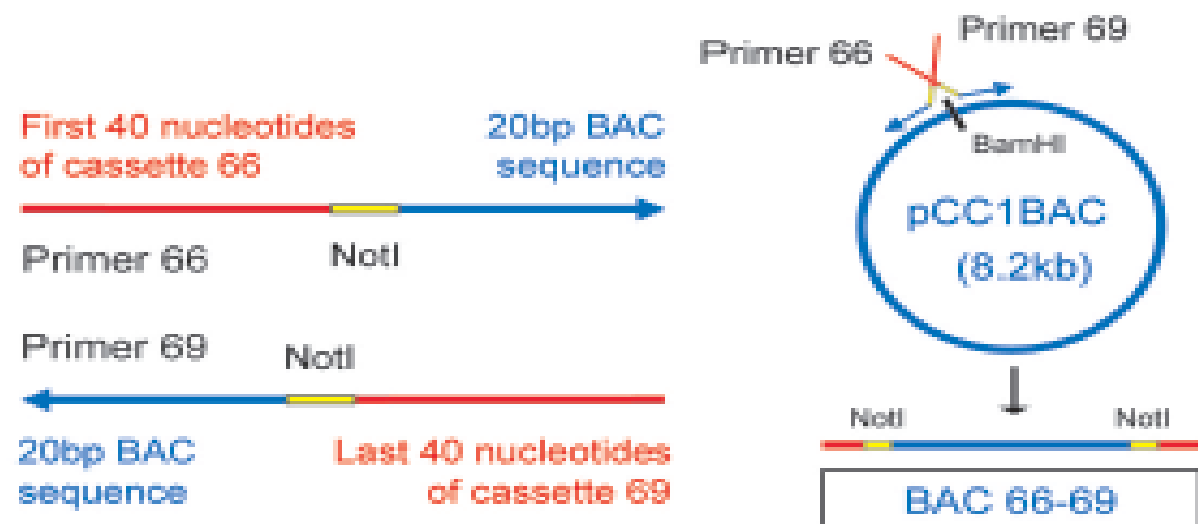
Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome

Daniel G. Gibson, Gwynedd A. Benders, Cynthia Andrews-Pfannkoch, Evgeniya A. Denisova, Holly Baden-Tillson, Jayshree Zaveri, Timothy B. Stockwell, Anushka Brownley, David W. Thomas, Mikkel A. Algire, Chuck Merryman, Lei Young, Vladimir N. Noskov, John I. Glass, J. Craig Venter, Clyde A. Hutchison III, Hamilton O. Smith*

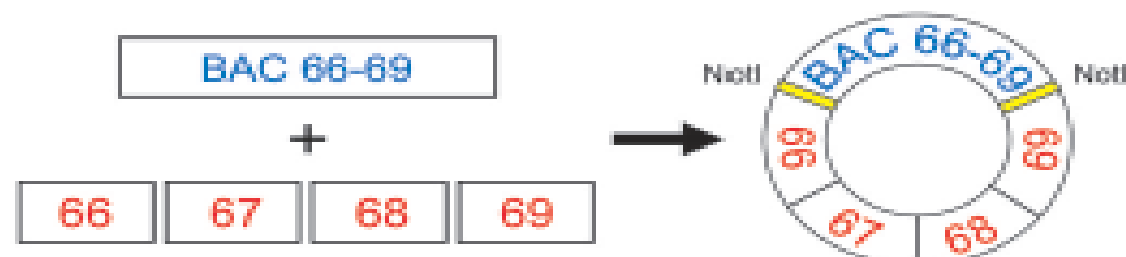
A. Assembly of cassettes 66 to 69 by *in vitro* recombination



B. PCR amplification of pCC1BAC vector using primers with Not I sites flanking the BamH I cloning site



C. Simultaneous assembly with PCR-amplified BAC



今後のバージョンアップと第3世代シーケンサー

| 454FLXTi | 2009 | 2010 | 2012 |
|--------------------------|-----------|-------------|-----------|
| Av. read length | 450 bases | 1,000 bases | 350 bases |
| Read# | 1.2 M | 2 M | 300 M |
| Total bases / run | 0.5 Gb | 2 Gb | 100 Gb |
| Run time | 10 hrs | 20 hrs | ? |
| Total bases /day (24 hr) | 1 Gb | 2 Gb | ? |
| Cost / run | \$6,000 | \$1,000 | \$2,000 |

emPCRなどの前処理の自動化(2010～)

De novo sequencing, finishing(2010～)

| SOLiD | 2009 SOLiD3 | 2009秋 SOLiD3Plus | 2010 SOLiD4 |
|--------------------------|-------------|------------------|-------------|
| Av. read length | 50 bases | 75 bases | 100 bases |
| Read# | 400 M | 1,000 M | 1,400 M |
| Total bases / run | 20 Gb | 80 Gb | 120 Gb |
| Run time | 10 days | ? | ? |
| Total bases /day (24 hr) | 2 Gb | ? | ? |
| Cost / run | \$23,000 | ? | ? |

2011?

The 3rd generation

- ABI+VisiGen
- Pacific Biosciences
- Oxford Nanopore Tech.

