

情報・システム研究機構シンポジウム

「分野を超えたデータサイエンスの広がり～自然科学から人文社会科学まで～」



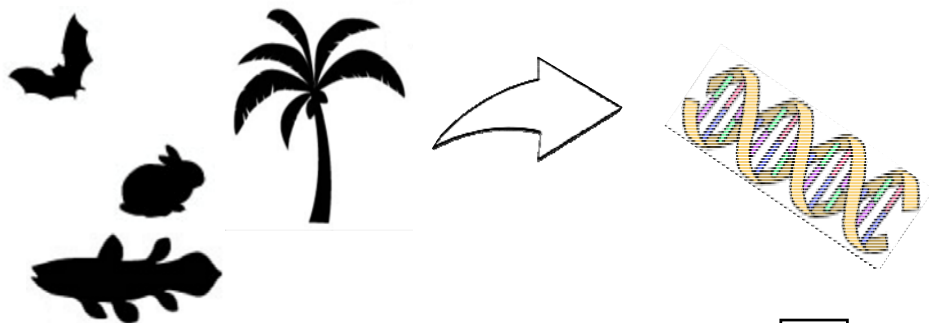
野口 英樹

近藤 伸二

福多 賢太郎

江島 史緒

# ゲノムデータ



◆ ゲノムDNA

◆ 転写産物 (RNA)

DNAシーケンサー

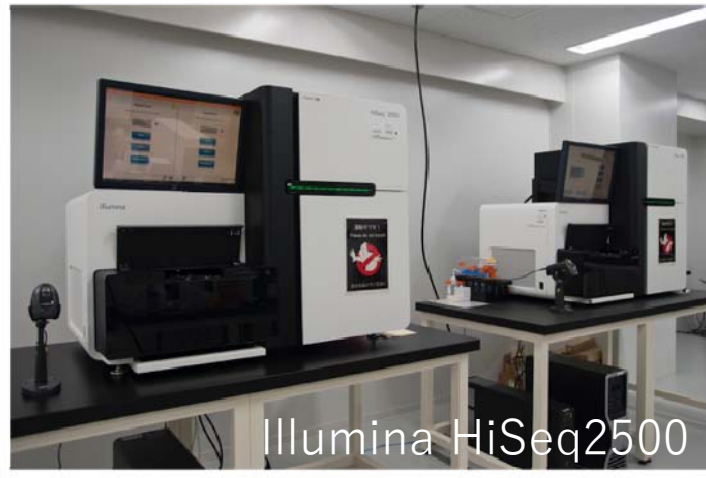


塩基配列データ

**多様・大量**

# 次世代シーケンサー (遺伝研先端ゲノミクス推進センター)

リード長      リード数      データ量



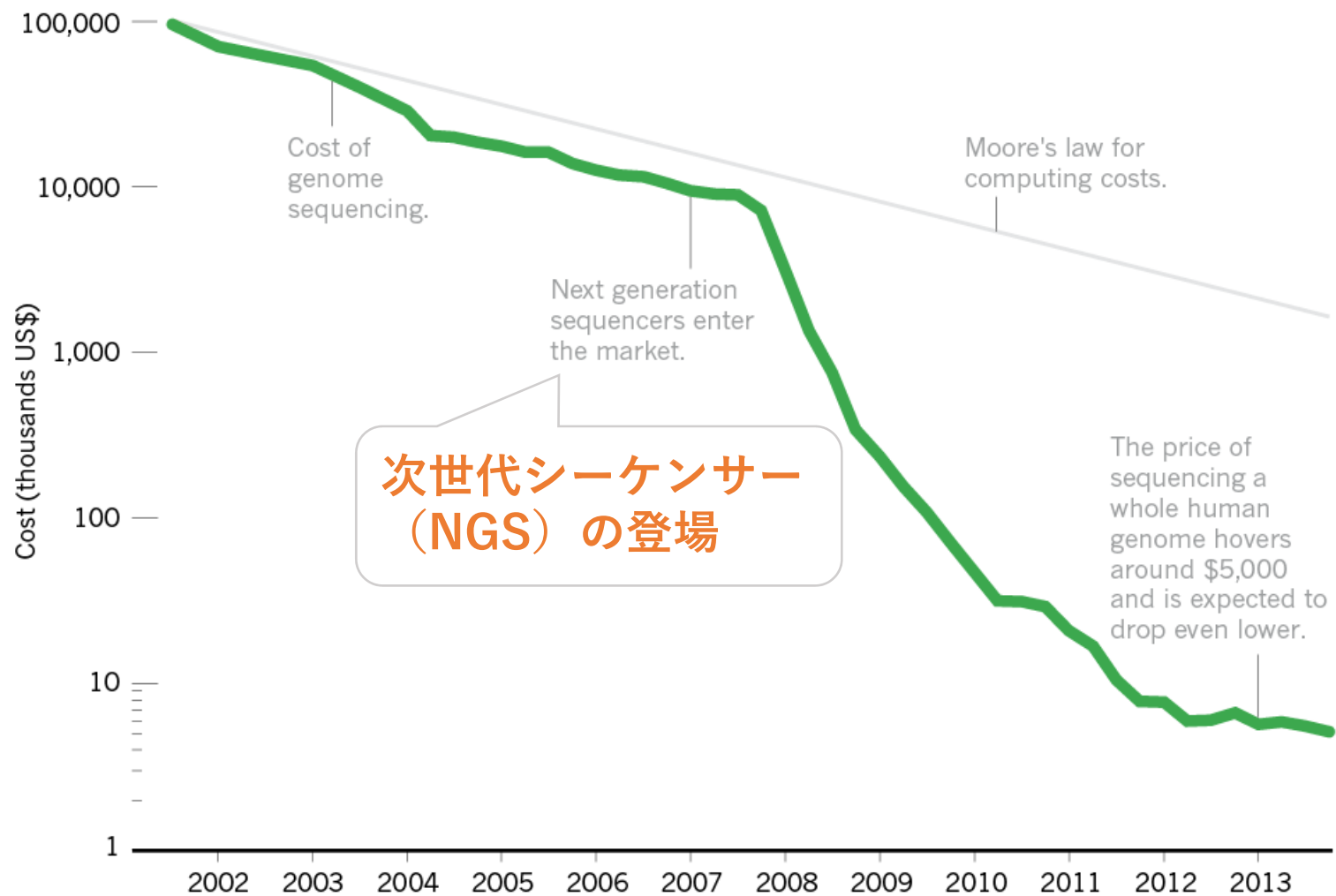
~250 bp      約2.5億      約60 Gb



平均15 Kbp      5-7万      ~1 Gb

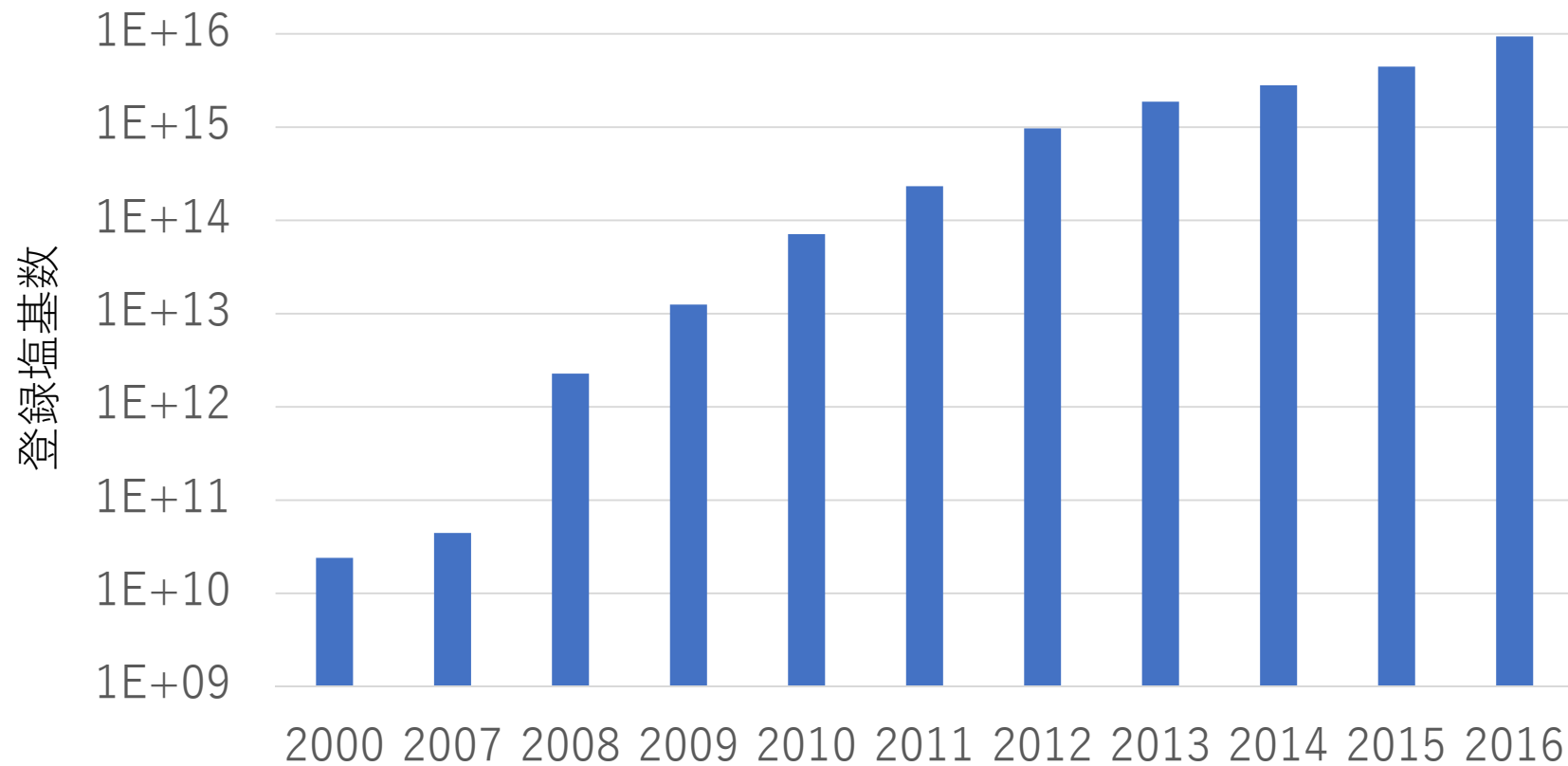
# Falling fast

In the first few years after the end of the Human Genome Project, the cost of genome sequencing roughly followed Moore's law, which predicts exponential declines in computing costs. After 2007, sequencing costs dropped precipitously.



The \$1,000 genome, *Nature* 507:294-295, 2014

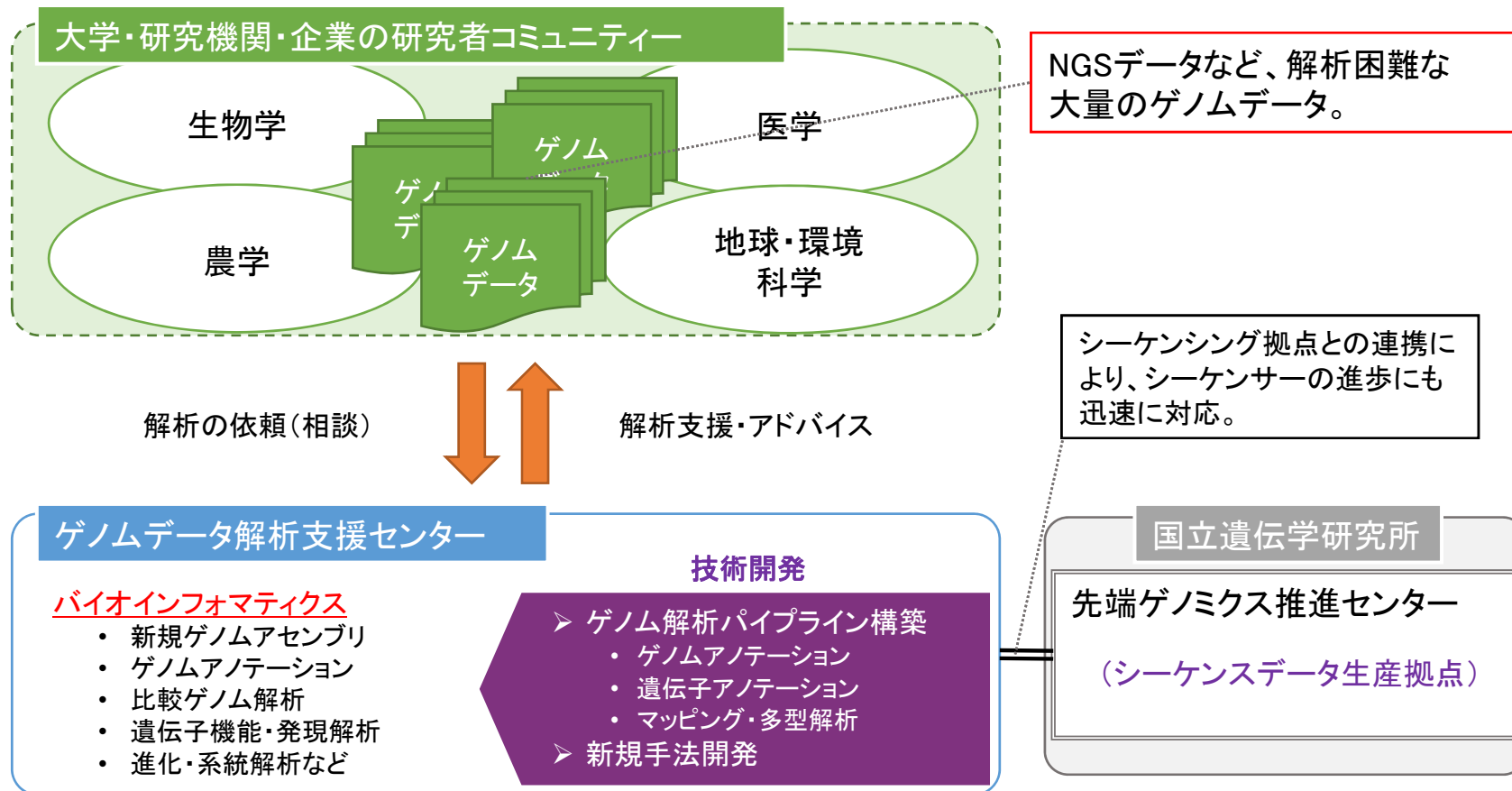
# 公共データベースの登録塩基数



**NCBI SRA** (<https://www.ncbi.nlm.nih.gov/sra>)

# ゲノムデータ解析支援センター

最先端のバイオインフォマティクス技術を用いて、多様なゲノムデータの情報解析支援を実施する。それによりゲノムデータ駆動型の研究の推進を図り、生物学のみならず医学や環境学などのさまざまな研究分野に貢献、その発展に寄与する。



# ゲノムデータ解析

- ***De novo*ゲノムシーケンス（新規ゲノム決定）**

- *De novo*ゲノムアセンブリ
- ゲノムアノテーション

- **ゲノムリシーケンス（変異・多型解析）**

- 全ゲノムリシーケンス
- ターゲットリシーケンス・エピジェネティクス解析

- **トランスクリプトーム解析**

- 遺伝子構造・発現解析
- Non-coding RNA解析

- **メタゲノム解析**

- メタゲノムアセンブリ
- 種分類、遺伝子予測

## 2016年度 支援課題

年度	生物種	解析カテゴリー
2016上期	ルーセットオオコウモリ	<i>De novo</i> ゲノム
	ゲンジ・ヘイケボタル	<i>De novo</i> ゲノム
	デメキン	リシーケンス
	ココナッツ	<i>De novo</i> ゲノム
	タマーワラビー	<i>De novo</i> ゲノム
2016下期	日本産マウス (MSM)	リシーケンス
	ゼブラフィッシュ	発現解析 (RNA-seq)
	アクネ菌	<i>De novo</i> ゲノム
	ウサギ	リシーケンス
	マウス (B6/MSMコンソミック)	発現解析 (アレイ)
	南極菌類	<i>De novo</i> ゲノム



## 今後の計画

- 高度なゲノムデータ解析支援の実施
- ゲノム解析パイプラインの構築
- 高速・省メモリな解析アルゴリズムの開発
- 人材育成（解析講習会等）



<https://genome-info.nig.ac.jp>