

Computational Biology Laboratory



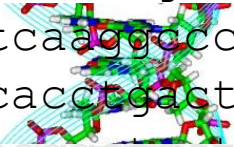
The Computational Biology Laboratory

早稲田大学 情報理工学科
情報理工・情報通信専攻
清水研究室

研究室概要

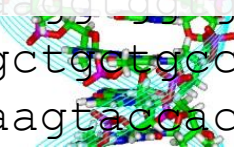
- 清水研では計算機科学のあらゆる手法を駆使して医学，生物学等の諸問題を解決する技術进行研究します。

gtacaaaaagcagaagg
gccgtcaaggccaccat
ggtgcaccgactgatgc



データ解析により
生命の秘密を
解き明かす研究

tgagagcgggtctc
gaaagcgggtctc
ccagagcgggtctc
agtggctgctgccctggc
tcacaagtaccactcagg
cctcatgggcccagcttt



究極の個人情報
と言われるゲノム
を守る研究



```
while  $i < \ell$  do  
   $x = q[i + o]$   
   $p \leftarrow \text{ExtR}(s, x)$   
  if  $\|p\|_g > 0$  then  $\triangleright$  Travers  
     $b \leftarrow x$ 
```

アルゴリズムを
極めて病気の原
因を探り当てる
研究

```
else  
   $x \leftarrow \text{ExtR}(s, x)$   
   $p \leftarrow \text{ExtR}(s, x)$   
  if  $\|p\|_r \neq 0$  then  
     $b \leftarrow x, s \leftarrow p$ 
```

現代の生命科学では情報解析が主役

ゲノム情報は主要なビッグデータです。（次ページ参照）膨大なデータを分析して病気の原因や創薬の手がかりを見つけます。

- 具体的にどんな研究ができる？

- ヒトゲノムデータと疾患データを機械学習などにより分析して、病気の原因遺伝子を探し当てる。

- どんな力が身につく？

- 文字列解析アルゴリズム，統計，機械学習等の知識が身につきます。

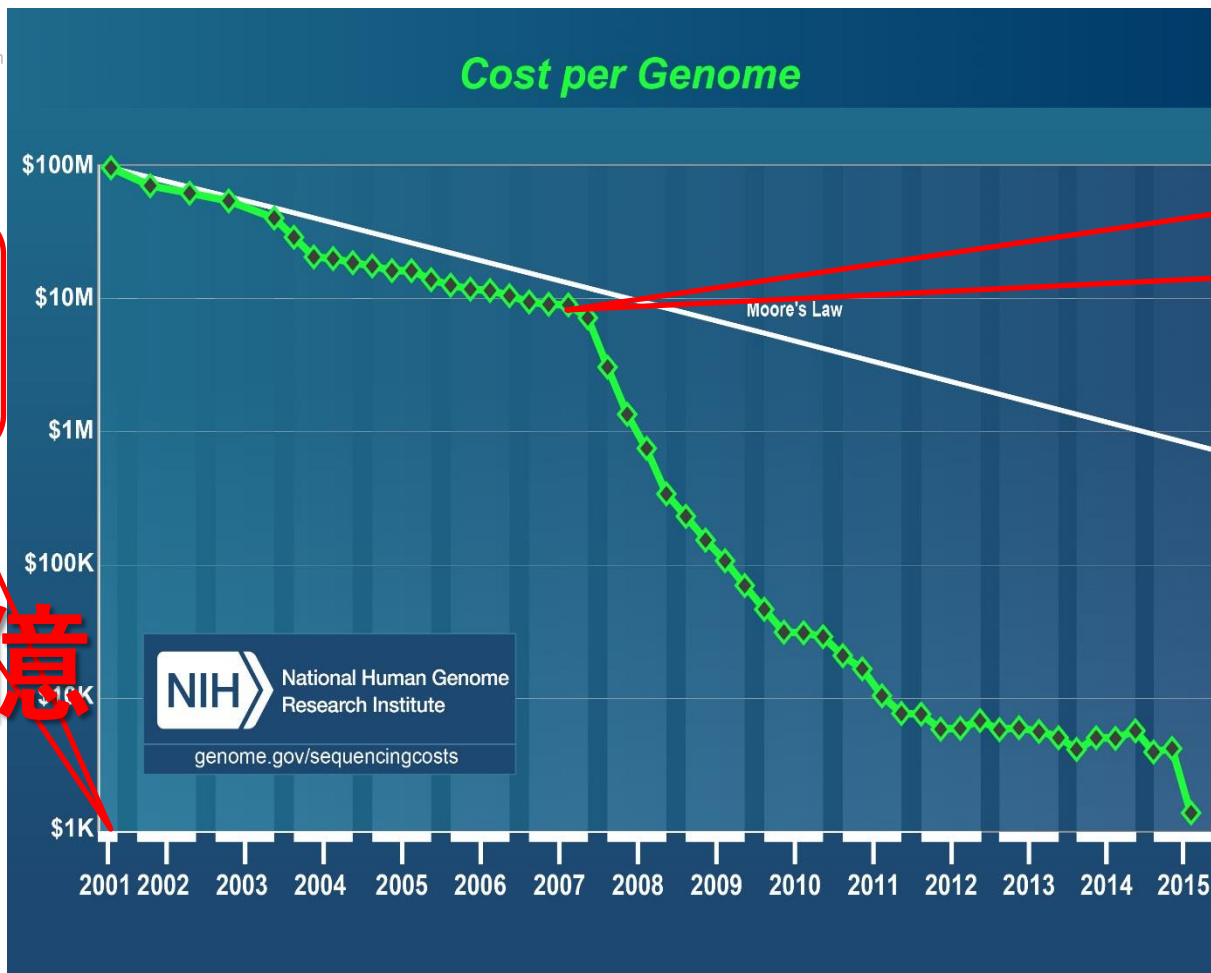
関連キーワード： 個人ゲノム，ガンゲノム解析，統計，機械学習，アルゴリズム，配列解析

Sanger
sequencer



ゲノムビッグデータ時代の到来

<https://products.appliedbiosystems.com>



ヒトゲノム計画
(1990~2003)

\$30億

新型シーケンサーの
実用化

\$1000



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

Next Generation
Sequencer

ヒトゲノムの長さ ≒ 30億塩基対

ゲノム≠特別な情報. もはや“ありふれた情報”

生命情報 × セキュリティ

究極の個人情報を守るにはどうしたらよいでしょうか？ゲノム情報解析の知識とセキュリティの知識を総動員して難題に挑みます。

- **具体的にどんな研究ができる？**

- 秘密計算技術（次ページ参照）を使ってゲノム情報を守りながら分析

- **どんな力が身につく？**

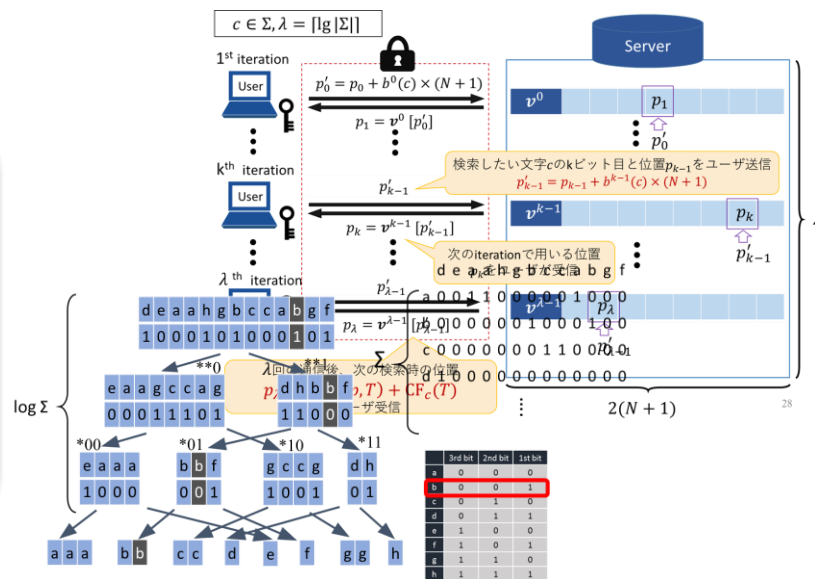
- セキュリティ関連の技術を勉強できる他，ゲノム配列をはじめとした生命データの扱い方について知識を深めることができます。

関連キーワード：個人ゲノム，個人情報保護，プライバシー保護データマイニング，暗号プロトコル，準同型暗号，秘密分散，ORAM，Trusted execution environment

秘密計算

- データを暗号化したまま目的の演算を実現する方法
 - 要素技術: 準同型暗号, 秘密分散, Garbled Circuit等
 - 応用例: 検索エンジンにキーワードを知らせずに, 検索結果だけを取得する.

学生の研究では, 準同型暗号を用いてテキストや木構造を安全に検索する手法の開発や, 秘密分散の全文検索などで成果を挙げています. また, 秘密計算とは呼びませんが, ハードウェアセキュリティの技術を使って安全にゲノム情報を解析する研究も行っています.



生命情報科学 × 先端アルゴリズム

ゲノム配列解析では動的計画法や接尾辞配列等を用いた高度なアルゴリズムを多用します。高度過ぎるアルゴリズムは応用先がないと思われがちかもしれませんが、生命情報科学ではアルゴリズムの工夫が疾患等のメカニズム特定に役立つのです。

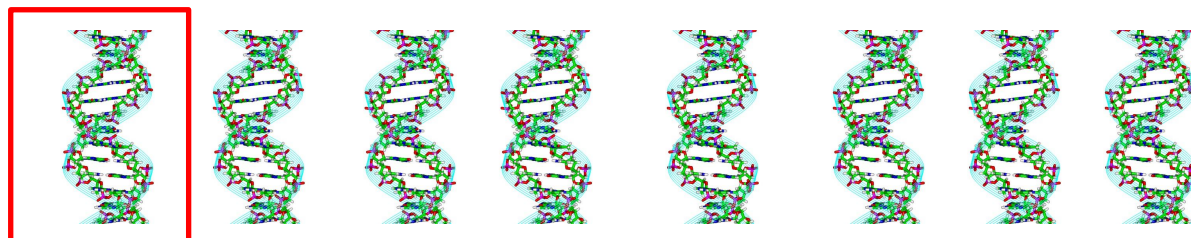
- **具体的にはどんな研究がある？**

- 数万人規模のヒトゲノム配列を高速に検索可能な索引方法の研究（次ページ参照）

関連キーワード：接尾辞木, 接尾辞配列, BWT, FM-Index, de Bruijn graph, 簡潔データ構造

ゲノム参照グラフ

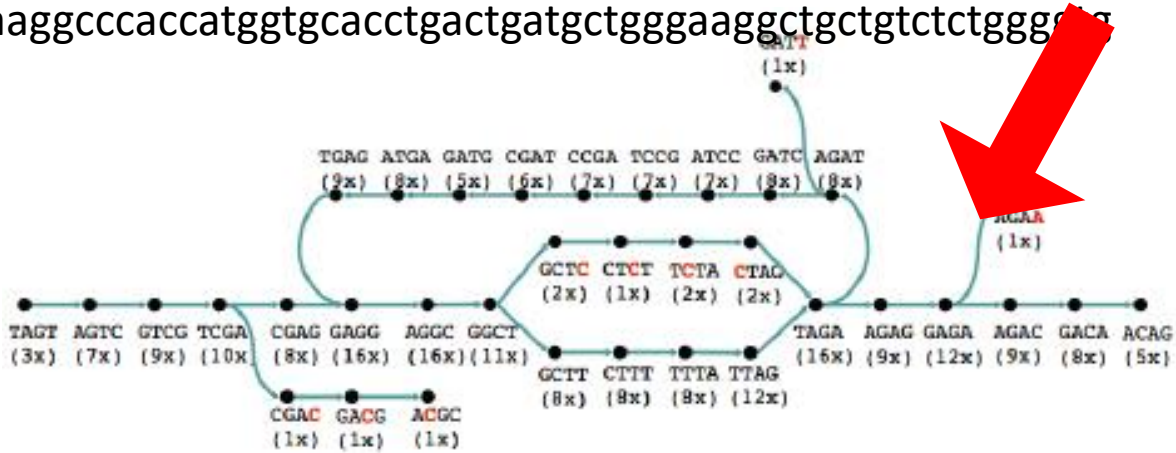
- 多様なゲノムを単一の辞書ではなく、グラフとして表現する新しい試み



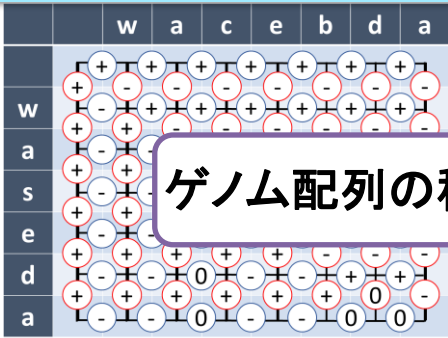
一般化圧縮接尾辞配列, PBWT, GPBWTといった最先端のデータ構造の研究が必要!

gtacaaaaagcagaagggccgtcaaggcccaccatggtgcacctgactgatgctgtgaaggctgctgtctctggcctg
 gtacaaaaagcagaaaaagccgtcaaggcccaccatggtgcacctgactgatgctgagaaggctgctgtctctggcct
 gtacaaaaagcagaagggccgtcaaggcccaccatggtgcacctgactgatgctgggaaggctgctgtctctgggga

学生の研究では、グラフ化したゲノム情報を効率的に比較して、ガンの原因遺伝子を発見する研究で成果を挙げています。

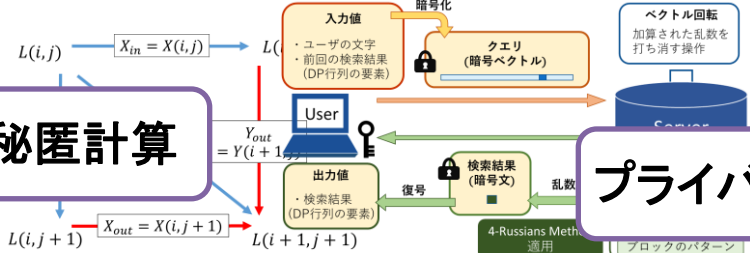


卒論テーマ例



ゲノム配列の秘匿計算

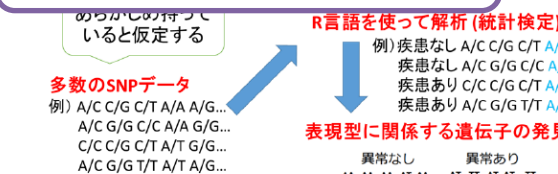
$$Enc(m1 + m2) = Enc(m1) \oplus Enc(m2)$$



プライバシー保護機械学習



ゲノムワイド関連解析

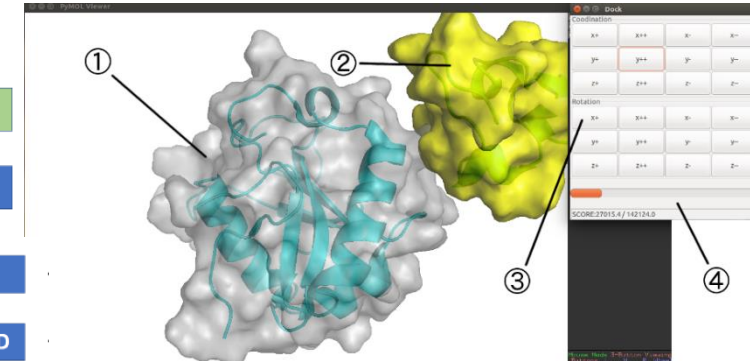


SGXによる安全解析

ゲーミフィケーションによる生体分子ドッキングシミュレーション

DNAストレージ

DTCゲノミクスの安全性
Spaced Seed検索のデータ構造
ゲノムグラフにおけるアラインメントアルゴリズム
など



色々な課題に挑戦

Huffman code	12011	02110	02101	222111	01112	222021
DNA nucleotides	GCGAG	TGAGT	ATCGA	TGCTCT	AGACC	ATGTGA

研究室の様子



国内外での研究発表や学会での受賞など、多くの学生が大学外からも認められる活躍をしています。



壁一面のホワイトボードや大きなビーズクッションがあります。キッチンのある部屋ではホームパーティーのような宴会を開きます。



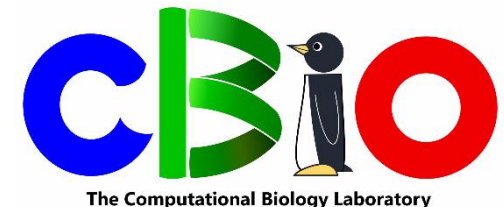
研究室の参考動画・記事

【研究室紹介動画・パンフレット】 ※ 研究室の様子, 学生インタビュー等.

- パンフレット: http://www.waseda.jp/nyusi/ebro/ug/se_jp_2019/html5.html#page=15
- 動画: <https://www.youtube.com/watch?v=VmsU9H3eGPU>

【研究内容紹介動画・記事】

- IPSJONE(<https://ipsj-one.org/2016/>)講演:
https://ipsj-one.org/2016/videos/24_shimizu_fs.mp4
- 河合塾・みらいぶプラス掲載記事:
生命情報科学について: <https://www.milive-plus.net/gakumon161202/02/>
ゲノム情報保護について: <https://www.milive-plus.net/gakumon161202/>
- 学部web(生命情報科学について):
<http://www.fse.sci.waseda.ac.jp/june-1-2016/>
- 早稲田理工 by AERA(秘匿ゲノム検索について):
https://www.cbio.cs.waseda.ac.jp/assets/materials/AERA_KS.pdf
- CAMPUS NOW 10月号(秘匿ゲノム検索について):
<https://waseda.box.com/s/h3n2ogcy1k8ivxz5oqnbpm8s07dakga>
- JSPS記事(ゲノムビッグデータ解析について):
https://www.jps.go.jp/j-grantsinaid/22_letter/data/news_2018_vol1/p12.pdf



その他

- オープンハウス

オープンハウスは、対面、オンラインの双方で実施します。

詳しくは、web (<https://www.cbio.cs.waseda.ac.jp/docs/posts/>)をご参照ください。

- 研究環境等

全員に対して個人用の席とPCが割り当てられます。テーマや研究の進捗にもよりますが、学外の一流研究者との議論も活発に行います。

- 学生の活躍

国内外での研究発表や学会での受賞など、多くの学生が大学外からも認められる活躍をしています。

- 研究室の所在, 連絡先

63号館5階01, 21室

E-mail: shimizu.kana@waseda.jp

URL: <https://www.cbio.cs.waseda.ac.jp>