

Computational Biology Laboratory



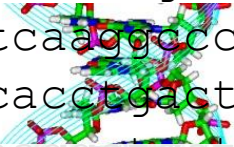
The Computational Biology Laboratory

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

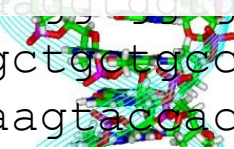
- 清水研では計算機科学のあらゆる手法を駆使して医学，生物学等の諸問題を解決する技術を研究します。

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  
     $b \leftarrow x, s \leftarrow p$ 
```

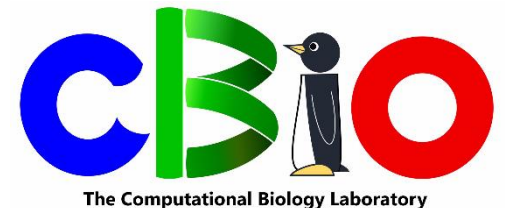
研究室の参考動画・記事

【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
- JSPS記事(ゲノムビッグデータ解析について):
https://www.jsps.go.jp/j-grantsinaid/22_letter/data/news_2018_vol1/p12.pdf



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。

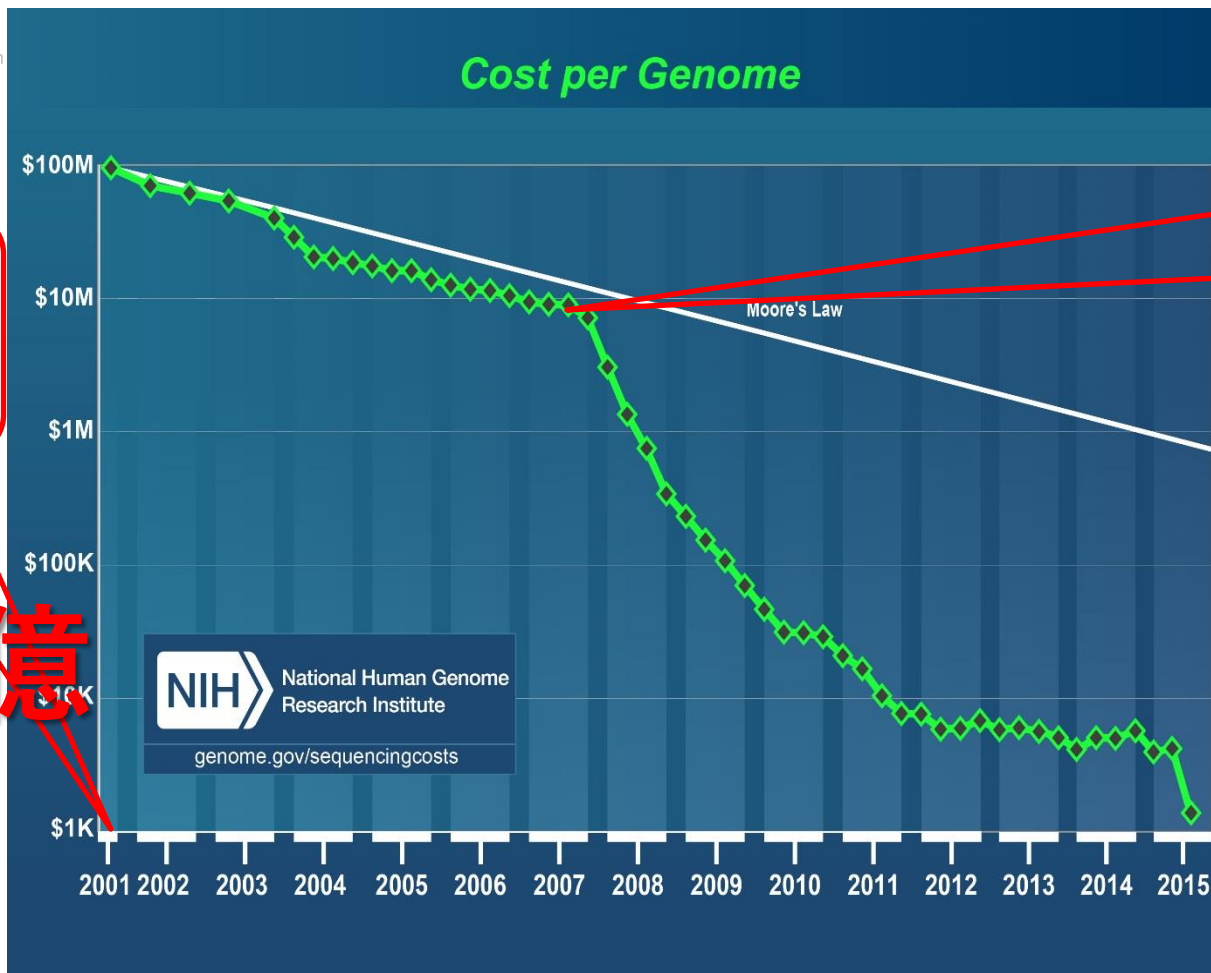
関連キーワード： 個人ゲノム，ガンゲノム解析，統計，機械学習，アルゴリズム，配列解析

Sanger
sequencer



ゲノムビッグデータ時代の到来

<https://products.appliedbiosystems.com>



\$10000



<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。

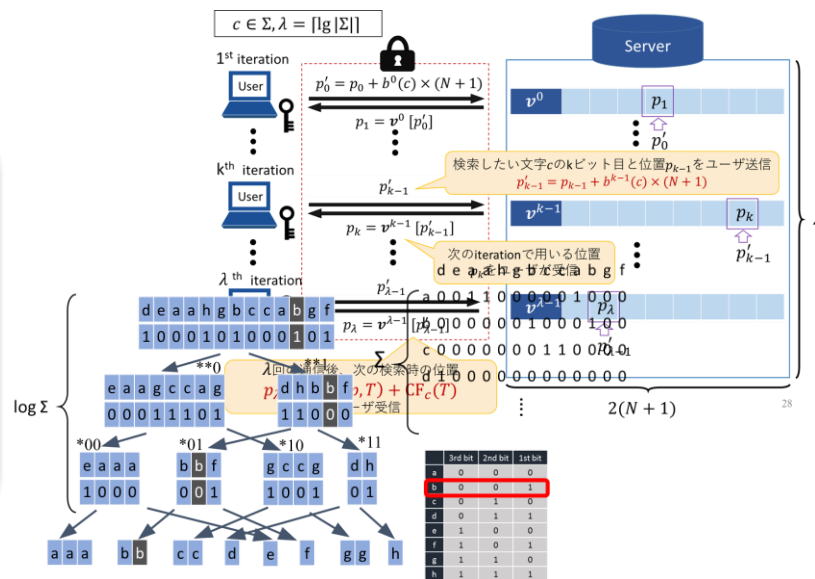
- **具体的にどんな研究ができる？**
 - 秘密計算技術（次ページ参照）を使ってゲノム情報を守りながら分析
- **どんな力が身につく？**
 - セキュリティ関連の技術を勉強できる他，ゲノム配列をはじめとした生命データの扱い方について知識を深めることができます。

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.



競技プログラミング × 生命情報科学

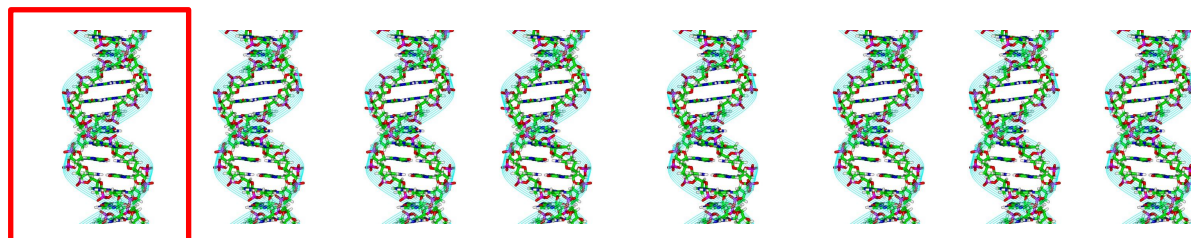
生命情報科学では競プロでおなじみの動的計画法や接尾辞配列を多用します。高度過ぎるアルゴリズムは応用先がないと思いがちかもしれませんが、生命情報科学では競プロで身に着けた力を存分に発揮し、さらに腕を磨くことができます。

- 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

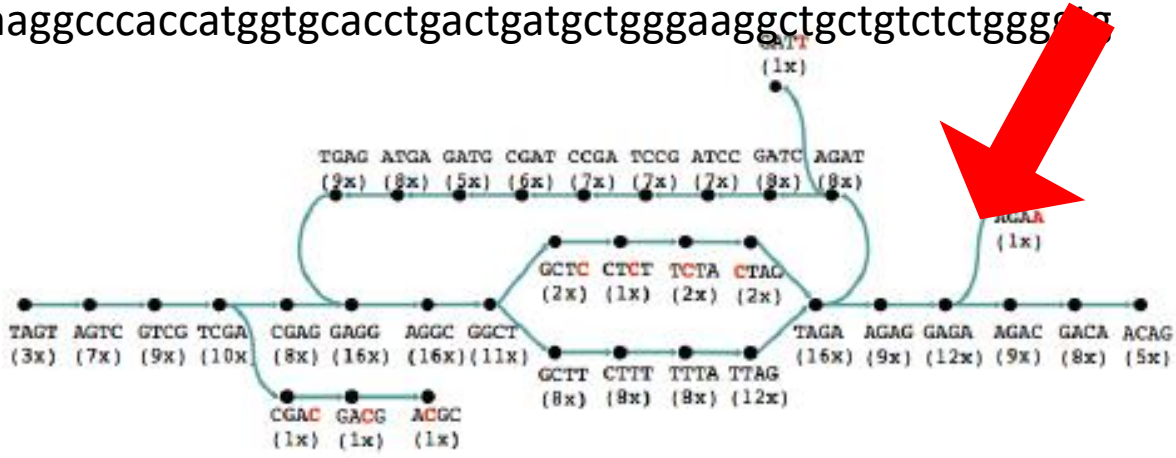
- 多様なゲノムを単一の辞書ではなく、グラフとして表現する新しい試み



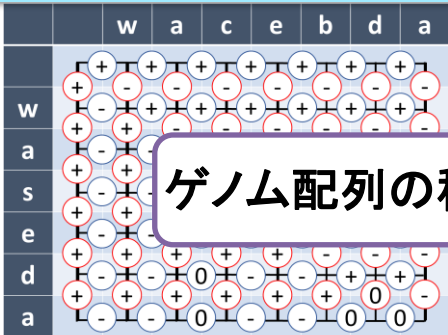
一般化圧縮接尾辞配列, PBWT, GPBWTといった最先端のデータ構造の研究が必要!

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。

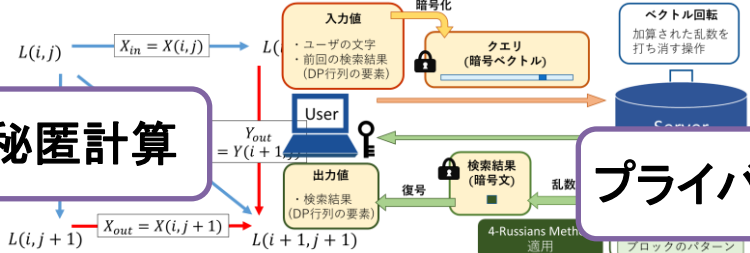


卒論テーマ例

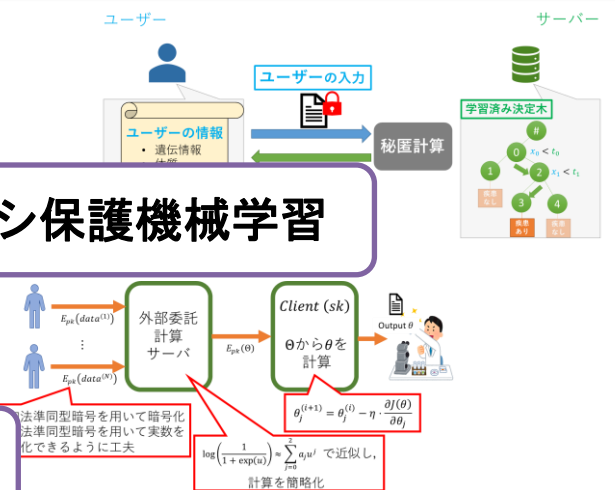


ゲノム配列の秘匿計算

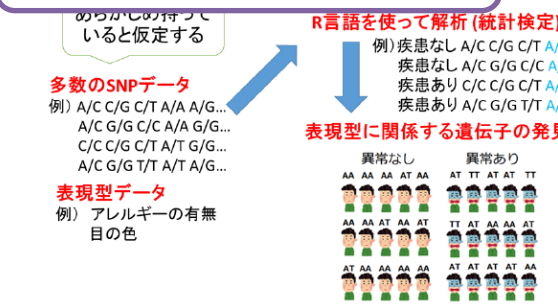
$$Enc(m1 + m2) = Enc(m1) \oplus Enc(m2)$$



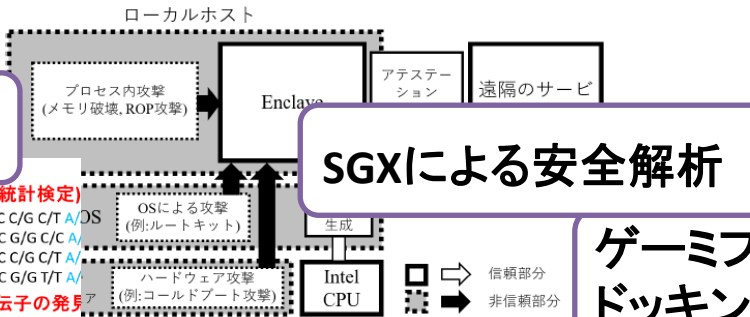
プライバシー保護機械学習



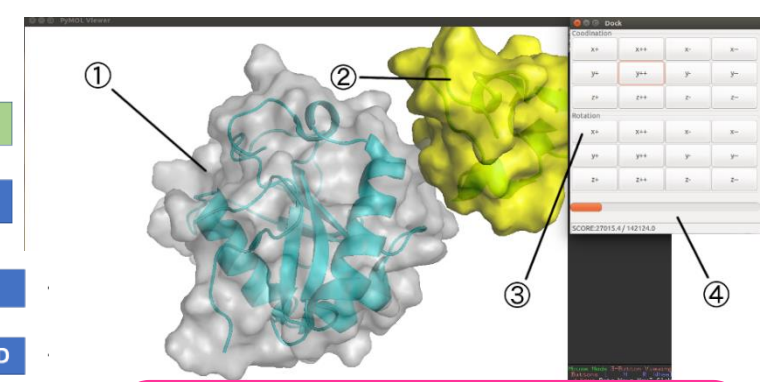
ゲノムワイド関連解析



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



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



オープンハウス・個別面談について

- 教員との個別面談期間：3月22日～3月24日
 - アポなしで対応可能な期間
3月22日(月)10:00 – 17:00, **研究室(63号館5階01室)**にお越しください。物理的に研究室内を見学したい方は、この日時でお願いします。
 - 3月22日(月)15:00 – 17:00, **バーチャル会場**にお越しください。
※アポなし対応のため、対面、オンラインを問わず、来客が重なった場合は、少しお待ち頂くことをご了承ください。
 - 3月22日10:00 – 17:00 以外の日程をご希望の場合は shimizu.kana@waseda.jpまでご連絡ください。
- 先輩が対応するバーチャルオープンハウス
 - 3月23日(火)13:00 – 15:00
 - **バーチャル会場**にお越しください。

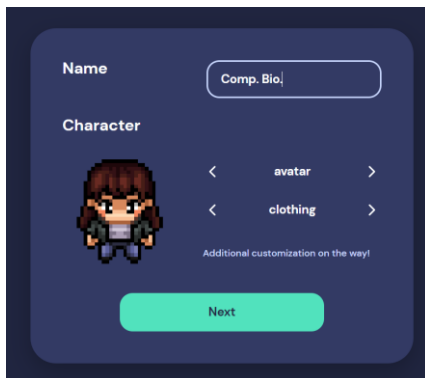
※ バーチャル会場への来場方法は次のページを参照。

バーチャル会場へのアクセス方法:

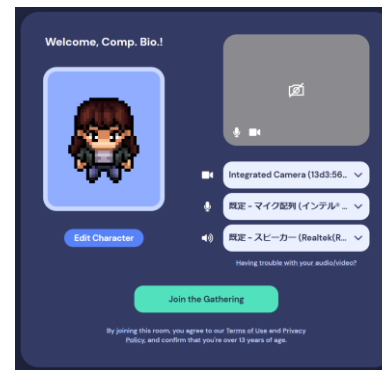
まず、<https://gather.town/app/rVqSy5VhwENaBxlc/cbiolab> にアクセスしてください。

※ 対応時間帯以外はメンバーは不在、もしくはスペースを閉じています。

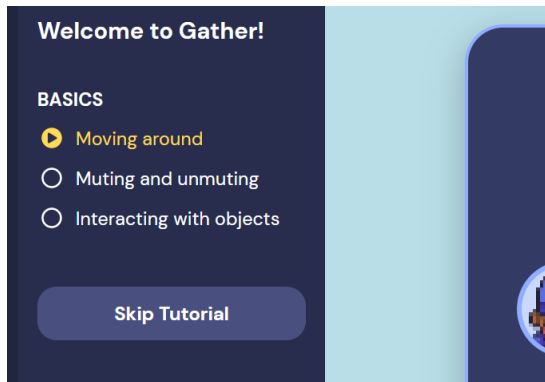
① 名前を入力



② 「Join Gathering」をクリック



③ 「Skip Tutorial」をクリックしてチュートリアルをスキップ



④ バーチャルラボに入ったら矢印のカーソルで先輩/教員のところに移動して会話開始!



2021年度のゼミについて

- 2020年度秋学期のゼミは、8割をオンライン、2割を対面により実施しました。（対面ゼミでは十分に感染対策をしたうえで、対面を希望しない学生はオンラインで参加できるようにしました。）
- 2021年度春学期もオンラインと対面の双方によりゼミを実施する予定ですが、研究室メンバーの意向や、COVID-19の状況に応じて、対面ゼミの割合を決めることとします。
- また、オンラインでのゼミ参加が快適なものとなるように、オンライン会議ツールの工夫や、貸与する機器類を充実させる予定です。

その他

- 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。

- 配属決定後の連絡方法

配属された方には、教員からメールで連絡します。

- 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>