



Tokyo Tech

Press Release

2023 年 8 月 7 日

東京工業大学

## 両親由来のゲノム配列を染色体スケールで決定する新手法

－両親間の大規模な変異の解析を可能に－

### 【要点】

- 染色体スケールでつながった両親由来のゲノム配列を区別する情報解析手法を開発。
- 哺乳類、鳥類、魚類などを対象にしたテストで高精度での配列決定を実証。
- 従来は解析が困難だった両親間の大規模な変異の解析を実現。

### 【概要】

東京工業大学 生命理工学院 生命理工学系の大内俊大学院生(博士後期課程 3 年)、伊藤武彦教授、梶谷嶺助教(研究当時)の研究チームは、真核生物のゲノム配列決定において、両親由来の配列を区別し、**染色体スケール**(用語 1)でつながった配列をそれぞれ決定する、新しい情報解析手法の開発に成功した。

DNA 配列決定の技術革新により、ゲノム中の染色体をほぼ全長で決定できるようになってきている。しかしゲノム配列決定では、染色体スケールの配列決定ツールの結果に対して手動による修正が必要なことや、真核生物では両親からそれぞれ受け継いだ 2 本の相同染色体のうち 1 本分しかゲノム配列決定できず、差異を考慮できないことが課題として残されている。そのため、配列データのみから両親由来の配列を区別し、染色体スケールでつながった配列をそれぞれ高精度で決定することが可能な解析手法が求められていた。

そこで研究チームは、「GreenHill (グリーンヒル)」と呼ばれる新しいプログラムを開発し、**次世代シーケンサー**(用語 2)の大規模な断片配列データと **Hi-C**(用語 3)と呼ばれる染色体立体配座捕捉法を用いて、両親由来の配列を染色体スケールで再構築することを可能にした。この新たな情報解析手法により、両親由来の配列を染色体スケールで簡便に決定できるようになり、両親間の大規模な変異の解析など、さまざまな下流解析への貢献が期待される。

本研究成果は、2023 年 7 月 11 日付の「*Genome Biology*」に掲載された。

## ●背景

ロングリードシーケンス技術や Hi-C 法など DNA 配列決定の技術革新により、少人数の研究グループでも、ゲノム中の染色体のほぼ全長の配列を決定できる時代が到来してきている。しかしながら、ヒトを含む 2 倍体生物のゲノム配列決定には依然として、以下のような課題がある。

1 つ目の課題は、既存の Hi-C 法を用いた染色体スケールの配列決定ツールの結果にはエラーが含まれており、手動の修正が必要なことである。手動修正には大幅な時間がかかるうえに、作業担当者により結果が異なる再現性の問題もあり、自動化が望まれていた。

2 つ目の課題は、2 倍体生物のゲノムは、母親と父親からそれぞれ受け継いだ 2 本の相同染色体を持っているが、そのうち 1 本分しか決定できないことである。そこで、2 本の相同染色体の違いを無視し、モザイク状につなが合わせた擬似的な 1 本のゲノム配列を決定する手法が長く用いられてきた。しかし近年、両親由来のゲノム配列(相同染色体間)で差異が大きい領域が表現型と関連する事例(ヒトの免疫型決定など)が報告されており、両親由来のゲノム配列を分けて解析することが重要であることがわかってきている。しかしながら、両親由来のゲノム配列を区別し、染色体スケールでつながった配列をそれぞれ決定することは、技術やコストの面から多くの問題が存在していた。

## ●研究成果

本研究で開発された解析手法では入力として、次世代シーケンサーの断片配列データを既存ツールでつながれた配列 (contig) を用いる (図 a)。解析ではまず、この contig 内の両親由来の配列から対応する配列を検出し、一つにまとめる (図 b)。そのうえで、まとめた配列をロングリードや Hi-C のデータを用いて染色体スケールまでつなげる (図 c)。最後に、最初にまとめた配列を二つに分け、ロングリードや Hi-C のデータを用いて、母親由来と父親由来の配列に分けることで、染色体スケールの両親由来のゲノム配列を構築する (図 d)。

研究チームは、この手法を「GreenHill」というプログラムに実装し、公開した (<https://github.com/ShunOuchi/GreenHill>)。さらにこの手法を実際に、線虫、牛、キンカチョウ、インコ、クロサイ、コチョウザメなどの各種生物に適用し、長くつながった両親由来の配列をそれぞれ既存の手法よりも高精度で決定できることを実証した。

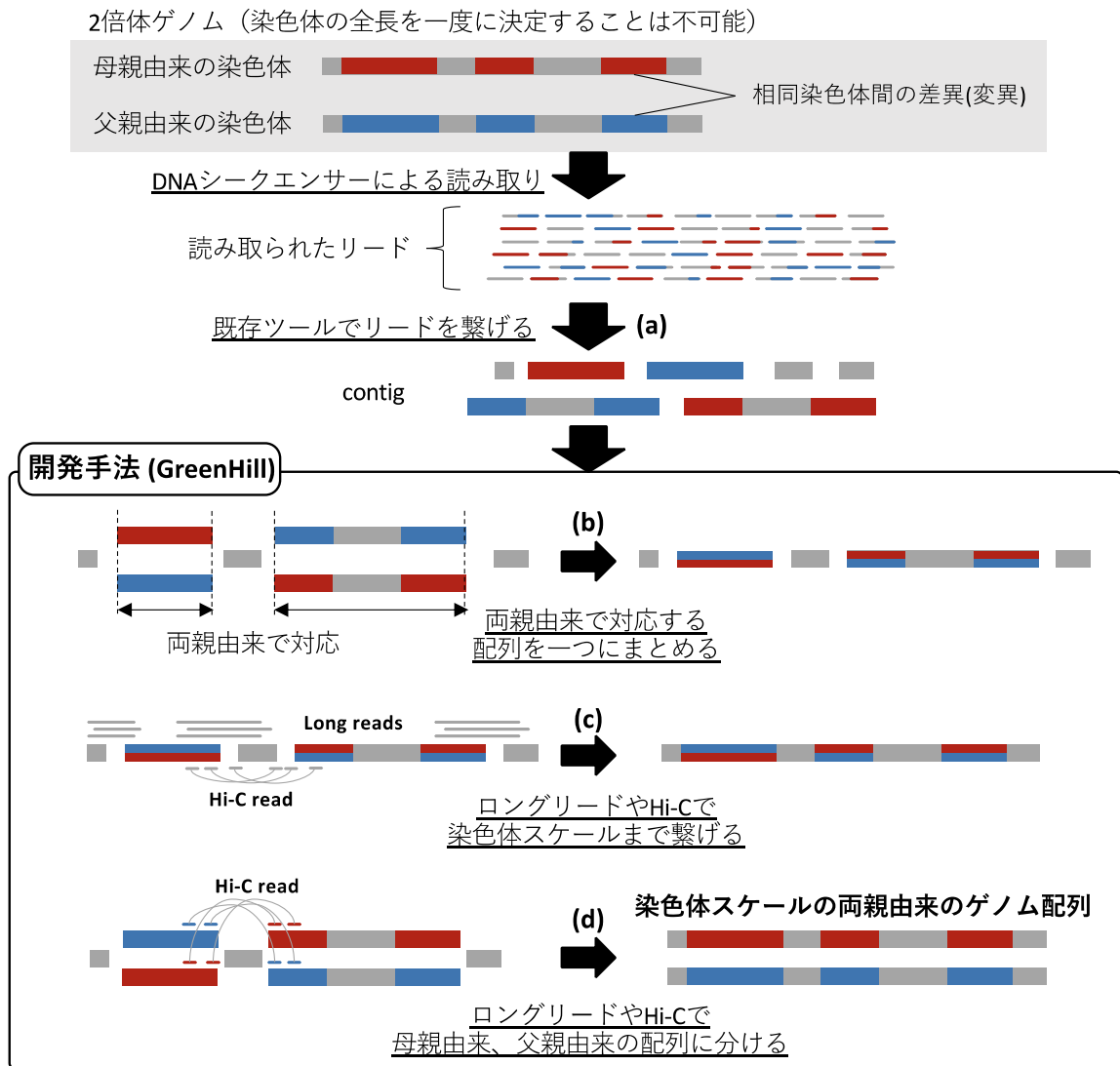


図 新たな情報解析プログラム「GreenHill」のアルゴリズムの模式図

## ●社会的インパクト

今回の研究で開発した手法は、2倍体生物のゲノム配列決定の真のゴールである「両親から引き継いだすべての染色体の全長の配列を決定すること」の達成に大きく貢献すると期待される。また、この開発手法は多くの生物種に対して高い性能を発揮しており、ゲノム配列決定プロジェクトを幅広く推進することも見込まれる。両親由来のゲノム配列を区別し、染色体スケールで決定することは、相同染色体間の大規模な変異の解析などの下流解析をする上で重要であり、がんゲノムの変異解析や植物の育種のためのゲノム解析など、幅広い分野への応用が期待される。

## ●今後の展開

本研究により、配列データのみから両親由来の配列を染色体スケールで決定する方法の開発に成功した。今後は、倍数性が高いなどの特徴を持つ複雑なゲノムの配列決定にも対応することで、より汎用性の高いゲノム配列決定手法の開発を目指す。

## ●付記

本研究は、科学研究費助成事業・学術変革領域研究（学術研究支援基盤形成）「先進ゲノム解析研究推進プラットフォーム」(JP22H04925(PAGS))、基礎研究 B（JP22H02598）などの支援を受けて行われた。

## 【用語説明】

- (1) **染色体スケール**：(今回用いている意味としては) ゲノム中の全染色体のほぼ全長配列を決定すること。染色体のほぼ全長の配列を決定することで、ゲノムの詳細な情報を得ることができる。
- (2) **次世代シーケンサー**：2004 年頃から登場した、新しいタイプの塩基配列解読装置（シーケンサー）。産出するデータ量が多いのが特徴で、近年では読み取れるゲノム断片の長さが従来のものより長いロングリードと呼ばれる技術が登場している。
- (3) **Hi-C**：染色体立体配座捕捉法。ゲノムの 3 次元構造の解析法の 1 つで、近年はゲノム配列決定にも応用されている。

## 【論文情報】

掲載誌：*Genome Biology*

論文タイトル：GreenHill: a de novo chromosome-level scaffolding and phasing tool using Hi-C

著者：Shun Ouchi, Rei Kajitani, and Takehiko Itoh

DOI：10.1186/s13059-023-03006-8

## 【問い合わせ先】

東京工業大学 生命理工学院 生命理工学系 教授

伊藤 武彦（いとう たけひこ）

Email: takehiko@bio.titech.ac.jp

TEL: 03-5734-3430 FAX: 03-5734-3630

## 【取材申し込み先】

東京工業大学 総務部 広報課

Email: media@jim.titech.ac.jp

TEL: 03-5734-2975 FAX: 03-5734-3661