

令和 5 年 4 月 14 日現在

機関番号：12601

研究種目：基盤研究(C) (一般)

研究期間：2020～2022

課題番号：20K06607

研究課題名(和文) シングルセル解析を用いて精子一細胞から高精度なゲノム構築を行う手法の開発

研究課題名(英文) Development of a method for highly accurate genome construction from single sperm cells using single-cell analysis.

研究代表者

吉武 和敏 (Yoshitake, Kazutoshi)

東京大学・大学院農学生命科学研究科(農学部)・助教

研究者番号：50646552

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：イトヨの精子シングルセルのゲノムシーケンスを行い、申請者が開発した連鎖解析ソフトウェアであるSELDLLAにさらに改良を加えることで、低カバレッジで欠損値の多いシングルセルのデータに対しても連鎖解析が可能であることを示した。(NAR Genom Bioinform. 2022 Mar 31;4)
また、連鎖解析によってゲノムを伸長した結果をグラフィカルに確認し、手動で補正する場合の補助ツールとしてSELDLA-Gを開発した。連鎖解析とは別の染色体構築手法であるHi-Cを用いたゲノム構築の際にもSELDLA-Gを利用することが可能であり、GitHubにてオープンソースソフトウェアとして公開した。

研究成果の学術的意義や社会的意義

分子生物学において生物を理解するには、その設計図となるゲノムを解読することがまず基本となるが、ゲノムを染色体の端から端まで読む解析手法はまだ発展途上である。本研究により新規にゲノムを決定する際に、染色体レベルのゲノムを構築するための新しい連鎖解析を提供することが出来た。本手法を用いることで、非モデル生物において高精度なゲノムを構築することが可能となり、その生物の特徴を理解するための基礎的な情報を入手することが出来ると考えられる。

研究成果の概要(英文)：Genome sequencing of Itoyo sperm single cells was performed and further modifications to the applicant's linkage analysis software, SELDLLA, showed that linkage analysis is possible for single sperm cell data with low coverage. (NAR Genom Bioinform. 2022 Mar 31;4)
We also developed SELDLA-G as an auxiliary tool for graphical confirmation and manual correction of genome scaffolding results obtained by linkage analysis. SELDLA-G can also be used for genome construction using Hi-C, which is a different chromosome construction method from linkage analysis, and we have released it as open source software on GitHub.

研究分野：分子生物学

キーワード：精子シングルセル 連鎖解析 ゲノム構築 SELDLA SELDLA-G Portable Pipeline イトヨ 非モデル生物

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

次世代シーケンサー(NGS)およびその周辺技術の開発・発展により、手軽にゲノム解読を行える環境が整備されつつある。特にゲノムサイズが数百万塩基程度のバクテリアゲノムであれば、1 サンプル 10 万円程度で完全にゲノムを解読することが可能である。しかしながら高等真核生物のゲノムサイズはバクテリアの 100 倍以上であることが多く、セントロメア、テロメア、rRNA のクラスターといった大きな繰り返し配列も存在するため、いまだに高等真核生物のゲノム解読は困難である。ゲノム解読の手順としては、図 1 に示すように、短い断片から段階を経て染色体の長さに伸ばしていく。しかし、スキャフォールドから染色体に伸ばすステップは未だに困難であり、NCBI のゲノムデータベースに登録されている真核生物 9,363 種のうち、染色体にまで伸長されているのはわずか 16% である。この染色体に伸長するステップで用いられる解析手法には、連鎖解析、Hi-C といった手法が存在するが、それぞれ問題点がある。連鎖解析では親の生殖細胞で生じた交叉反応を子供で検出するため、親子の DNA を準備する必要があるが、未だにゲノムが未決定の生物ではそもそも飼育系が確立していないことが多く、親子の DNA を準備するのが困難な場合が多い。また、近年開発が進んでいる細胞核内のゲノムの高次構造を利用した Hi-C という手法は、1 個体のみで解析が行えるという大きなメリットがある反面、染色体ごとにマーカーが分離される保証が無いという致命的な欠点がある。

申請者らは、連鎖解析を使用した染色体レベルのゲノム構築手法として、雌性単為発生個体(ダブルハプロイド)を用いた連鎖解析や、交雑種を用いた連鎖解析を開発してきた。これらの手法に共通する着眼点として、2 倍体のゲノムをジェノタイピングするためには、通常 30 回(30x)以上繰り返し同じ場所をシーケンスする必要があるが、1 倍体のゲノムであれば僅か 1 回(1x)だけ読めば良いという点を利用している。準備する DNA 試料として、通常の連鎖解析よりもさらに制約が加わるものであるが、その反面マーカーの分解能を著しく向上させることが可能であり、飼育系の確立している生物種であれば、非常に高精度の連鎖地図を作成可能となった。

2. 研究の目的

本研究では、ダブルハプロイド、交雑種の系をさらに発展させ、近年開発が進んでいるシングルセルの技術を取り入れ、精子 1 細胞を直接ジェノタイピングすることを目指す。精子をジェノタイピング出来れば、親の生殖細胞で生じた交叉反応を直接検出することが可能であり、その子孫を準備する必要が無い。さらに精子であれば 1 倍体であるため、ダブルハプロイドや交雑種で培った 1x と少ないデータ量から効率よくジェノタイピングを行い、高密度に連鎖地図を作成することも可能である。

これまでに精子 1 細胞からのジェノタイピングは、ヒト、ミジンコ等で報告があるが、解析した細胞数は高々 100 個程度と少ない。本研究では 1x のデプスで精子 1,000 細胞を読み取り、超高密度の連鎖地図を作成し、高精度のゲノムを構築することを目標とする。また、連鎖解析におけるデータ解析では、未だにマーカー数が千個程度の時代に作られた解析ツールが頻りに用いられているが、本研究のようにサンプル数 1 万、マーカー数 100 万個の規模になると、それらのツールは全く機能しない。申請者らはダブルハプロイドや交雑種での超高密度連鎖解析ツールとして独自に SELDLA というソフトウェアを開発しており、それを精子一細胞系に導入することで、既存のツールでは解析不可能な規模でのデータ解析を行う。

3. 研究の方法

本研究では、商用レベルで利用可能となったシングルセルゲノム解析のプラットフォームである 10x Genomics 社のライブラリ調整装置を用い、イトヨの精子を対象にシングルセル全ゲノム増幅を行った。10x Genomics 社のライブラリ調整装置では一細胞ごとに異なるバーコード DNA が付与され、それを Illumina 社のシーケンサーで一度に読み取って、データ解析時にデータを細胞ごとに分離していく。そのため、シーケンサーの読み取る量を増やすことで、容易に 1,000 細胞のシーケンスも行うことが可能である。イトヨは脊椎動物の中ではゲノムサイズが 400 Mb と小さく、シーケンス時のデータ量を低く設定することが可能であった。

データ解析には申請者らの開発した SELDLA というソフトウェアを改良し、欠損値の多い精子 1 細胞での解析に適したアルゴリズムを実装した。また一連の解析を誰でも簡単に行えるように SNP 抽出から連鎖解析によるゲノム構築までの解析フローを Portable Pipeline に実装し、公開した。

4 . 研究成果

本研究では 10x Genomics 社のシングルセルライブラリ調整装置を用い、Illumina 社の次世代シーケンサーを使って大量のシーケンスを行ったデータを解析して、精子を一細胞ずつジェノタイプングし、連鎖解析を行い、高精度のゲノムを構築することが目標である。イトヨの精子シングルセルのゲノムシーケンスを行い、申請者が開発した連鎖解析ソフトウェアである SELDLA にさらに改良を加えることで、低カバレッジで欠損値の多いシングルセルのデータに対しても連鎖解析が可能であることを示した。(Yoshitake, et. al., NAR Genom Bioinform. 2022 Mar 31;4(2))

さらに、連鎖解析によってゲノムを伸長した結果をグラフィカルに確認し、手動で補正する場合の補助ツールとして SELDLA-G を開発した。SELDLA-G は Windows や Mac で動作するソフトウェアであり、GPU を用いて高速に描画したり、マーカー間のフェーズ一致度を高速に計算したりするなど、操作性が高い。また連鎖解析とは別の染色体構築手法である Hi-C を用いたゲノム構築の際にも SELDLA-G を利用することが可能であり、SALSA という Hi-C 解析ツールの出力を入力としてコンタクトマップを編集することが出来る。開発したソフトウェアは GitHub にてオープンソースソフトウェアとして公開した。(https://github.com/c2997108/SELDLA-G)

また、シーケンスデータを解析し、SNP コールから連鎖解析までを行う一連の解析フローを Windows でもグラフィカルに行えるように Portable Pipeline を開発し、こちらも GitHub にて公開した。(https://github.com/c2997108/OpenPortablePipeline)

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 Yoshitake Kazutoshi, Ishikawa Asano, Yonezawa Ryo, Kinoshita Shigeharu, Kitano Jun, Asakawa Shuichi	4. 巻 4
2. 論文標題 Construction of a chromosome-level Japanese stickleback species genome using ultra-dense linkage analysis with single-cell sperm sequencing	5. 発行年 2022年
3. 雑誌名 NAR Genomics and Bioinformatics	6. 最初と最後の頁 online
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/nargab/lqac026	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------