

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 20 日現在

機関番号：12608

研究種目：研究活動スタート支援

研究期間：2015～2016

課題番号：15H06209

研究課題名(和文)多倍体生物の全ゲノム配列構築手法の開発

研究課題名(英文)Development of an assembly method for whole genomes of polyploid organisms

研究代表者

梶谷 嶺(Kajitani, Rei)

東京工業大学・生命理工学院・助教

研究者番号：40756706

交付決定額(研究期間全体)：(直接経費) 2,100,000円

研究成果の概要(和文)：多倍体化は発酵用酵母や農業用植物に広く観察され、それら生物のゲノム配列情報解析からは産業上有用な知見が得られることが期待される。しかし、互いに類似した多数の相同染色体を持つ多倍体生物のゲノム配列を決定することは、効率的な新規アルゴリズムが要求される。本課題では、多倍体ゲノムの骨組みとなる構造(コンセンサス配列)および各々の相同染色体配列(ハプロタイプ)を決定するツールを開発した。

研究成果の概要(英文)：Polyploidization are widely observed for brewing yeasts and agricultural plants, and genome analysis of these organisms can be valuable for industries. However, multiple homologous chromosomes that are similar to each other require a novel and effective algorithm. In this research, I developed a tool to determine backbone structures of polyploid genomes (consensus sequences) and each homologous chromosome sequence (haplotype) separately.

研究分野：バイオインフォマティクス

キーワード：多倍体 ゲノム de novo アセンブリ

1. 研究開始当初の背景

ゲノム配列決定において、全ゲノムショットガン法は階層的ショットガン法と比較してコスト・時間の点で効率的であるが、真核生物の場合は相同染色体の組をまとめてシーケンシングし計算機上で配列をつなぎ合わせる(*de novo* アセンブリ)こととなる。その際、相同染色体間の差異(ヘテロ接合体)が *de novo* アセンブリの結果を悪化させることは動物・植物を問わず多くのゲノム解読計画で報告されており(*Nature* 2012:490, 49-54 など)、近交系が確立されていない非モデル生物のゲノム決定において大きな障害となっていた。対策として申請者らは高ヘテロ接合体サンプルに対応した *de novo* アセンブラ Platanus を開発し、全ゲノムショットガン法で高ヘテロ接合体サンプルのドラフトゲノムが構築可能なことを示した(*Genome Res.* 2014:24, 1384-95)。Platanus は既に多くの非モデル生物のゲノム解読計画に適用されており、ゲノム決定の効率化をもたらしている。論文発表された計画の例としては、申請者が参加したものではシーラカンス(*Genome Res.* 2013:23, 1740-8)とシロオビアゲハ(*Nat. Genet.* 2015:47, 405-9)、その他ではネムリユスリカ、クロミンククジラの解読計画が挙げられる。しかしながら、Platanus を含めた既存の *de novo* アセンブラは2倍体サンプルを想定して設計されており、倍数性が3以上のサンプルに対しては適用できないという問題点も残されている。倍数性の増加はゲノム決定時にヘテロ接合体と同様に障害となり、幾つかの解読計画ではシーケンシング前の段階で対策を講じている。例として、ジャガイモ(異質4倍体)ではホモ接合2倍体のクローンを交配実験で作成し(*Nature* 2011:475, 189-95)、コムギ(異質6倍体)では染色体腕部の置換とフローサイトメトリを用いて染色体を分離しており(*Science* 2014:345, 1251788)、それぞれの処理後にシーケンシングを行なうことで *de novo* アセンブリ時の障害を削減している。しかし、これらの手法は時間とコストがかかる上、生物種毎に実験条件を試行錯誤する必要がある、汎用的に用いられてはいない。一方、多倍体酵母ゲノムは植物と比較してサイズが小さく、リピート配列率も小さいことから全ゲノムショットガン法の適用例も存在するが(*DNA Res.* 2014:21, 299-313)、アセンブリ結果は完成ゲノムには至っていない

2. 研究の目的

ゲノムの多倍体化は特に植物・酵母において多く観察され、その中には産業上重要な生物種も含まれる。例として、主要な作物であるジャガイモは通常異質4倍体であり、2倍体の個体は食用部分のサイズが小さく栽培には適さない。多倍体ゲノム解析によって得られる知見は世界的な食料問題の解決にも大いに重要であると考えられる。酵母の多倍

体はビールやワインの発酵用の株で多く見られ、耐寒性や芳香物質など表現型への影響も調べられており、こちらも産業上重要な要素となっている。多倍体ゲノムのアセンブリ手法が確立された場合、植物・酵母のゲノム解析におけるブレークスルーとなり得るため、当該用途のソフトウェアの開発を目的とする。

3. 研究の方法

入力データは現時点で塩基あたりのコストが最小の Illumina 社製シーケンサーのものであり、開発段階での対象生物種は多倍体の酵母を用いる。*de novo* アセンブリの方法としては次の2種類の方針に従ったツールを実装する。

(1) 相同染色体のコンセンサス配列構築

倍数性が3以上のサンプルに対応した *de novo* アセンブラを、申請者が過去に開発した2倍体用 *de novo* アセンブラ Platanus(*Genome Res.* 2014:24, 1384-95)を拡張することで開発する。Platanus は最初にシーケンスリードから長さ k の部分文字列(k -mer)を抽出し、それらを節点とする de Bruijn グラフと呼ばれるデータ構造を作り、その中の経路をアセンブリ結果(contig)として出力する。次に、ペアードエンド情報を用いて配列の伸長を行う scaffolding という工程に移るが、その際は contig を節点とした scaffold グラフという構造を構築する。ここでも同様にグラフ上の経路を結果配列(scaffold)とする。de Bruijn グラフと scaffold グラフの両方において、相同染色体間で差異がある場合(ヘテロ領域)、バブルと呼ばれる特徴的な構造が出現する。これを統合することで長い配列を構築することができるが、アセンブルされた配列は両方のハプロタイプをモザイク状に含んでおりコンセンサス配列となる。Platanus の特徴としては、構造変異を含む領域に対しても、ヘテロ領域であるかどうかを精度良く判定し統合を行えるという点が挙げられる。その時に主に用いられる情報は coverage depth(ある領域がシーケンスされた回数)であり、ヘテロ領域は片方の相同染色体由来のため値が低くなる性質を利用する。倍数性が3以上のサンプルでは、バブルの中にさらに分岐が含まれることでグラフが複雑化してしまうため、始点と終点を共有する経路の集合を「スーパーバブル」として検出し、そこから1つの経路を選んでコンセンサス配列とする解決法が考えられる。ここで、単に相同性の高い配列の組を統合してしまうとパラログ遺伝子などを誤ってまとめてしまうことになるため、各経路で coverage depth が低くなっているかどうかを確認した上で統合の可否を判定する機能を組み込む予定である。Platanus には単純なバブルの統合など基本的な機能は既に実装されている上、開発は申請者自身に

よって行われたため、比較的低い開発コストで新機能を実装することができる。

(2) ハプロタイプ配列構築

前年度に開発した *de novo* アセンブラとは異なり、相同染色体を統合せずにハプロタイプ毎に配列を出力するツールを開発する。相同染色体間の差異が密かつ均等に分布しているならば、バブル統合機能を無効にするのみで長いハプロタイプ配列を構築可能であると予想されるが、実際のゲノム内では変異の密度に大きな偏りが存在している。そのため変異が存在せず連鎖が不明な領域が現れ、ハプロタイプ配列を伸長することができなくなる。本研究ではインサートサイズが大きいメイトペアライブラリを用いて高精度で連鎖関係を決定し、長いハプロタイプ配列を構築するアルゴリズムを開発する。結果のハプロタイプ配列を前年度に開発されたツールが作るコンセンサス配列上に並べ、各染色体と対応付けられた形で出力される機能も実装する。

4. 研究成果

平成 27 年度終了時までには共同研究の体制を整え、ラガービール酵母 *Saccharomyces pastorianus* の複数株のシークエンシング結果を入手している。これらのサンプルは異種交配株であり、倍数性は 3 または 4 であるが、特定のゲノム領域で倍数性が異なる性質（異数性）を持つ。Illumina 社製 DNA シークエンサのデータが得られた 4 株については、ゲノム領域間の倍数性の差異による coverage depth のばらつきに対応するため、エラー由来配列の除去やヘテロ領域の統合に関するパラメータを調整し、効果的に相同染色体のコンセンサス配列を構築する方法を実装した。その結果、400 kbp 以上の N50 長（アセンブリ結果の長さの評価指標）を示すドラフトゲノムを構築することに成功した（表 1）。それらの配列について、祖先種ゲノムとの配列比較や mate-pair ライブラリのマッピングによる精度評価も行なっているが、構造的なエラーは検出されていない。ドラフトゲノム群はラガービール酵母の比較ゲノム解析の論文 (Okuno et al. *DNA Res.* 2016:23, 67-80) と関連付けて公共データベースへの登録および公開もなされており、代表者も当該論文の第 2 著者となっている。また、Illumina シークエンサのデータに加えて、多倍体酵母の 1 株については Pacific Biosciences 社製の 1 分子 DNA シークエンサのデータも入手することができ、そのデータに対応したアルゴリズムの開発も開始することができた。

表 1. 公開済みの多倍体酵母ドラフトゲノム

Strain	CBS1503	CBS1513
Genbank accession	GCA_001515 425.1	GCA_001515 445.1
BioSample ID	SAMD00035 484	SAMD00035 486
Total length (bp)	17,195,167	19,248,212
# sequences	631	178
N50 length (bp)	484,478	644,406
Max length (bp)	1,060,739	1,050,489

Strain	CBS1538	W34/70
Genbank accession	GCA_001515 465.1	GCA_001515 485.1
BioSample ID	SAMD00035 486	SAMD00035 489
Total length (bp)	14,404,124	22,500,926
# sequences	277	495
N50 length (bp)	428,791	723,289
Max length (bp)	760,567	1,455,873

平成 28 年度からは、コンセンサス配列ではなく各ハプロタイプ配列を構築するアルゴリズムの開発を重点的に行なった。野生の生物集団ではハプロタイプ配列間の差異が特に大きくなる (>5%) ゲノム領域が存在している例が当該年度までに多く報告され、本課題で対象に入る異質倍数体では更に差異は増大していると想定される。このようなゲノム領域が存在すると、単純に類似度の高い配列を相同領域として統合することは困難となるため、各ハプロタイプ配列をできるだけ別々に構築し、最後に各相同領域で 1 つのハプロタイプを選択していくことでコンセンサス配列を構築するソフトウェアを開発した。この方針の利点としては、コンセンサス配列を作る前の、個々のハプロタイプ配列もアセンブリ後の解析に活用することができる点である。

開発の前段階としては、倍数性が 3 以上のサンプルより *de novo* アセンブリが容易であると考えられる 2 倍体生物を対象としたアルゴリズムを開発し、それを発展させる形で本課題の対象の多倍体サンプル向けのソフトウェアを実装した。具体的には、*de novo* アセンブリ時に用いられるグラフ構造である de Bruijn グラフと scaffold グラフの両方に対して、Illumina シークエンサの paired-end (mate-pair) あるいは 1 分子シークエンサーのロングリードをマップし、分岐構造を解決することを繰り返す。多倍体用には、グラフ中の制限付き経路探索を実装することで、計算時間を抑えつつ分岐の多い構造に対応した（図 1）。成果のツールの性能は、多数のハプロタイプを含むメタゲノムデータへの応用という形で、第 11 回日本ゲノム微生物学会年会にて発表を行なった。

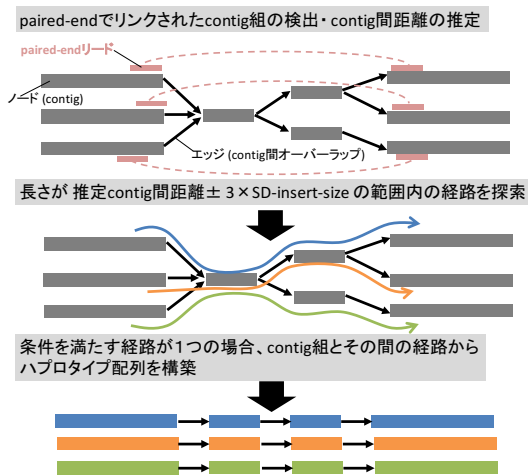


図 1. 多ハプロタイプ構築アルゴリズムの模式図

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 1 件)

① Miki Okuno, Rei Kajitani, Rie Ryusui, Hiroya Morimoto, Yukiko Kodama, Takehiko Itoh. Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization. *DNA Research*, 査読有, Vol. 23, Issue 1, pp. 67-80, DOI:<https://doi.org/10.1093/dnares/dsv037>

[学会発表] (計 1 件)

① 梶谷嶺、小椋義俊、後藤恭宏、吉村大、奥野未来、林哲也、伊藤武彦、メタゲノムショットガンデータからの株ハプロタイプ配列構築手法の開発、第 11 回日本ゲノム微生物学会年会、2017 年 3 月 4 日、慶應義塾大学 湘南藤沢キャンパス (神奈川県 藤沢市)

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

名称：
 発明者：
 権利者：
 種類：
 番号：
 出願年月日：
 国内外の別：

○取得状況 (計 0 件)

名称：
 発明者：
 権利者：
 種類：
 番号：
 取得年月日：
 国内外の別：

[その他]
 ホームページ等

6. 研究組織

(1) 研究代表者

梶谷 嶺 (KAJITANI, Rei)
 東京工業大学・生命理工学院・助教
 研究者番号：40756706

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：

(4) 研究協力者

()