

令和 5 年 6 月 13 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：2021～2022

課題番号：21K20580

研究課題名（和文）非モデル生物を用いたゲノム倍数化を介した植物の種分化機構解明の基盤構築

研究課題名（英文）Building a foundation for elucidating speciation mechanisms mediate genome polyploidization using non-model plants

研究代表者

田中 裕之（Tanaka, Hiroyuki）

東京工業大学・生命理工学院・研究員

研究者番号：60782042

交付決定額（研究期間全体）：（直接経費） 2,400,000円

研究成果の概要（和文）：本研究では、スミレ属植物でゲノム研究を展開するための基盤構築を目的として、異質4倍体であるニオイスミレ（*V. odorata*）の新規ゲノム決定を試みた。その結果、スミレ属植物で初めて染色体スケールのゲノム配列の構築に成功し、完成度の高い遺伝子セットを作成した。よって、スミレ属植物で研究を展開するための基盤構築が完了した。

研究成果の学術的意義や社会的意義

ゲノム重複による高次倍数体の形成は植物一般にみられる現象であり、染色体倍化が植物の進化・種分化において重要な役割をもつと古くから考えられてきたが、倍数体形成による遺伝子レベルの進化については未だに解明されていない点が多い。

スミレ属は日本という狭い範囲内で環境適正や形態などが多様化した数十種類の異質多倍体スミレが自生しており、ゲノム倍数化と種分化の研究において適した研究材料である。スミレ属でゲノム研究を展開するための基盤構築が実現すれば、染色体倍化がもたらす植物の進化・種分化機構の役割の解明に貢献すると考えられる。

研究成果の概要（英文）：In order to building a foundation for genomic analysis of *Viola spicose*, I performed genome sequencing project of allotetraploid *Viola odorata*. I generated a chromosome-level reference genome and high-quality predicted gene annotation dataset for *V. odorata*.

研究分野：バイオインフォマティクス

キーワード：異質倍数体 種分化 比較ゲノム解析 スミレ属

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

多くの生物種は 2 セットの染色体を持つが、セット数が増えた個体が現れ、新種として確立することがある。これをゲノム重複と呼び、ゲノム重複が生物の新規性をもたらしたという仮説が提唱されてきた (Ohno, 1970, Evolution by Gene Duplication)。近年、植物の全ゲノム重複が種分化を牽引したこと (Palfalvi et al., 2020, Current Biology) や、異質 4 倍数体植物では 2 倍体近縁種よりも遺伝子変異が高頻度におこること (Paape et al, 2018, Nat Commun.) が報告されており、ゲノム倍数化が植物の進化・種分化において重要な役割をもつことがわかってきた。しかしながら、倍数体形成による遺伝子レベルの進化については未だに解明されていない点が多い。

スミレ属は北半球の温帯域を中心に約 400 種が知られており、日本では基本種で 60 種、変種・品種・雑種を含めると約 220 種が自生している。それらの分布様式は多岐にわたり、地形的・気候的变化に富む日本列島の中で様々な環境に適応している。日本に自生するスミレ属の染色体数は報告されており、基本染色体数は $n=6$ であるが大部分は $n=10, 12, 24, 36$ などの多倍数体種である (Miyaji, 1929, CYTOLOGIA)。興味深いことに、2 倍体種の生息域は高山地帯に限られているのに対し、低山地帯に生息する種は全て多倍数体種であり、それらは、Chamaemelum 節の 2 倍体種に由来する "CHAM ゲノム" と未同定の 2 倍体種に由来する "MELVIO ゲノム" からなる異質 4 倍数体種であることが示唆されている (Marcussen et al., 2015, Syst. Biol.)。したがって、日本の低山地帯に生息するスミレ属は、Chamaemelum 節の 2 倍体種が起源であり、MELVIO ゲノムの由来となる 2 倍体種との交雑・倍数化による異質 4 倍数体形成をきっかけに低山地帯に適応し、さらに森林、平地、海浜、丘陵などのそれぞれの環境への適応型として多様化したと考えられる (図 1)。以上のことから、日本に自生するスミレ属は "ゲノム倍数化が生物に多様性をもたらす" を体現した植物であり、これらの比較ゲノム解析によって、ゲノム倍数化の環境適応の多様化機構を明らかにすることができると考えられるが、スミレ属異質 4 倍数体のゲノム解読は全く進んでいない。

2. 研究の目的

そこで本申請課題では、スミレ属植物を "ゲノム倍数化の環境適応の多様化" のモデル植物として研究を展開するための基盤構築を目的として、異質 4 倍数体種ニオイスミレ (*Viola odorata*) の新規ゲノム決定と、遺伝子予測を実施した。

3. 研究の方法

【DNA シーケンスデータの取得】

ニオイスミレの葉から塩化ベンジル法によって DNA を抽出し、PacBio ライブラリを作成した。さらに、ニオイスミレの葉を試料として Dovetail Omni-C® Kit によって Hi-C ライブラリを作成した。PacBio ライブラリは PacBio Sequel II の CCS (Circular Consensus Sequencing) mode によって、Hi-C ライブラリは Illumina novaseq 6000 によってシーケンスを実施した。これらのシーケンスは共同研究者である遺伝学研究所の豊田 敦博士によって実施していただいた。

【RNA シーケンスデータの取得】

ニオイスミレの組織 (葉、茎、根、花) を採取し、RNeasy Protect Kits (QIAGEN) を用いて標準プロトコールに従って RNA 抽出を実施した。RNA-seq ライブラリの作成およびシーケンシングは外部受託サービスを利用した。

【ゲノム配列の構築】

既に申請者が取得した Illumina pair-end と Illumina mate-pair (短い DNA 断片の両端配列を取得) に加えて今回取得した Pacbio CCS リードの 3 種類のデータを組み合わせてコンティグ配列の de novo アセンブリの実施と、Hi-C リードを使用したコンティグ配列の Hi-C スキャフォールドリングを実施した。コンティグ配列の de novo アセンブリと Hi-C スキャフォールドリングはそれぞれ種々のツールを検討し、配列の連続性や双子葉植物に共通する Single copy orthologs (BUSCO: eudicots_odb10) の検出率など評価指標によって、最も完成度が高いと考えられる構築結果を選定した。

【ニオイスミレゲノムの遺伝子予測】

まず、RNA-seq データを使用した方法を 2 種類、キントラノオ目の近縁種 (*V. pubescens*, *Populus trichocarpa*, *Manihot esculenta*, *Ricinus communis*) 遺伝子配列のホモロジー情報を使用した方法、プログラムによる機械的予測 (ab initio) を 2 種類の合計 5 種類の方法によってニオイスミレゲノム内の遺伝子領域を予測した。最終的に、それぞれの予測結果を精査・統合し、ニオイスミレゲノムの遺伝子領域を特定した。予測遺伝子の機能アノテーションは、Blast を用いたタンパク質データベース (NCBI nr, UniProtKB/Swiss-Prot) との相同性検索と、InterProScan

によるド配列中のモチーフおよびドメインの探索を実施した。

4. 研究成果

【DNA・RNA シーケンスデータの取得】

植物からの DNA 抽出法としては C-TAB 法が一般的に知られているが、スミレ属植物は多糖類やポリフェノール類などを多く含んでいることが原因で、他の植物と比較して抽出効率が悪かった。そこで他の抽出方法を検討した結果、塩化ベンジル法をベースとした抽出法によって高分子 DNA を高効率で抽出可能であることがわかった。本方法によって、異質倍数体スミレの葉から約 35 µg の DNA を抽出することができた。この DNA を用いて PacBio ライブラリを作成し、PacBio Sequel II の CCS (Circular Consensus Sequencing) mode によってゲノムシーケンスを実施することで、21.7 GB の CCS リードデータを取得した。さらに、ニオイスミレ 300 mg の葉を資料として、Dovetail Omni-C® Kit によって Hi-C ライブラリを作成し、Illumina novaseq 6000 によって 74.3 Gb のリードデータを取得した。

RNA-seq データは、ニオイスミレの各組織由来の RNA-seq データを、合計 53.4 Gb 取得した。Trinity による RNA-seq リードの de novo アセンブルによって遺伝子セットを作成し、BUSCO を用いた Single copy orthologs の検出率を評価した。その結果、双子葉植物に共通する Single copy orthologs (eudicots_odb10) の検出率は 96.6%であった。以上の結果から、今回取得した RNA-seq データは遺伝子アノテーションに使用するうえで十分なクオリティであると考えられた。

【染色体スケールのゲノム配列の構築】

Pacbio CCS リードを入力データとした hifiasm v1.16.1 によるアセンブルを実施した結果、コンティグ配列の合計長は 691.9 Mb のコンティグ配列が得られた。連続性の指標となる N50(配列を長い方から順に足していき、全体の半分の長さまで達した時の配列の長さ)は 29.6 Mbp であり、スミレ属において染色体配列が唯一報告されている 2 倍体スミレ *V. pubescens* の N50 (3.5 kbp) を大幅に上回っていた。続いて、コンティグ配列に含まれるオルガネラゲノム由来のコンティグや冗長な配列を除去し、Hi-C リードデータを活用したコンティグ配列のスキファオールディングを実施した。Hi-C リード (73 Gb) を入力データとして Yaha によって Hi-C スキファオールディングを実行し、その後 Juicebox によってリードマッピング情報を可視化し、ミスアセンブル、ミススキファオールディングのマニュアル修正を実施した。最終的に単相(n)の染色体数と同一である 10 本の配列を構築した。合計長は 650 Mbp であり、推定ゲノムサイズと同程度の値を示した。全コンティグの 98.8%は染色体スキファオールドに帰属しており、スミレ属植物において初めて染色体スケールのゲノム配列の構築に成功した。

【DNA シーケンスデータの取得】

まず、RNA-seq データを使用した 2 種類の方法によって 87,659 遺伝子と 129,796 遺伝子、ホモロジー情報を使用した方法によって 235,919 遺伝子、プログラムによる機械的予測(ab initio)を活用した 2 種類の方法によって 33,981 遺伝子と 46,773 遺伝子が予測された。そしてそれら 5 種類の遺伝子予測結果を統合することで、最終予測結果として 37,784 遺伝子が予測された。BUSCO を用いた Single copy orthologs の検出率から、今回予測された遺伝子セットの完成度評価を実施した。その結果、ニオイスミレの予測遺伝子セットからは、双子葉植物に共通する Single copy orthologs の 98.2%が検出されており、2 倍体スミレ *V. pubescens* の遺伝子セットの Single copy orthologs 検出率(76.9%)を大幅に上回っており、高精度な遺伝子予測(NCBI RefSeq)が実施されているキントラノオ目の近縁種の Single copy orthologs 検出率 (*M. esculenta*, 99.5%; *P. trichocarpa*, 98.8%; *R. communis*, 99.6%) に匹敵する値であった。さらに、予測遺伝子の機能アノテーションを実施した結果、37,227 遺伝子(全遺伝子中の 98.5%)がアノテーションされた。

本研究によって、スミレ属植物において初めて染色体スケールのゲノム配列の構築に成功した。今回作成したゲノム配列および予測遺伝子セットは、先行研究のスミレゲノム(*V. pubescens*)よりも高い完成度を示した。予測遺伝子の大部分が機能アノテーションされており、本研究の目的であったスミレ属植物でゲノム研究を展開するための基盤構築が完了した。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 田中裕之、堀立樹、山本 祥平、豊田敦、矢野 健太郎、山根 京子、伊藤武彦
2. 発表標題 ワサビゲノム解読の現状
3. 学会等名 日本育種学会第143回講演会
4. 発表年 2023年

〔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
---------	---------