

令和 2 年 7 月 13 日現在

機関番号：82508

研究種目：基盤研究(B) (一般)

研究期間：2017～2019

課題番号：17H03769

研究課題名(和文)花の模様形成を決める細胞の位置別のエピジェネティックスの解明

研究課題名(英文)Epigenetics analysis for determination of floral patterns in each cell

研究代表者

平川 英樹(Hirakawa, Hideki)

公益財団法人かずさDNA研究所・ゲノム情報解析施設・施設長

研究者番号：80372746

交付決定額(研究期間全体)：(直接経費) 14,400,000円

研究成果の概要(和文)：セントポーリアは組織培養における不定芽の誘導によって多様な花の模様変異体が出現する。その原因が遺伝子発現の違いによるものがエピジェネティックな制御によるものを調べ、花の模様形成のメカニズムを明らかにすることを目的とした。品種'キラウエア'の全ゲノム配列をSequelによるロングリードを用いてアセンブルし、2,200本(総延長：708.7 Mb、N50長：1.3 Mb)のドラフト配列を構築した。また、フェージングしたゲノム配列も構築した。セントポーリアのゲノム解読や遺伝子予測、発現量解析などを行ったが、対立遺伝子間の比較や共発現ネットワークの構築といった解析まで実施することができなかった。

研究成果の学術的意義や社会的意義

セントポーリアは特徴的な生理・生態的特徴を持つことから花の模様形成の研究材料として用いている研究者が多い。本研究で得られた全ゲノム配列は多くの研究に貢献でき、セントポーリアが花の模様形成のモデル植物になりうると思われる。また、本研究によって、花の模様形成を解明するだけでなく、花の模様といった不安定な形質に関わる原因遺伝子の特定を行う多くの研究者に共通の方法論を与えることができる。セントポーリアは園芸作物として広く流通しており、花の模様形成のメカニズムが明らかになればより多様な模様の品種を育成できるようになることが期待される。

研究成果の概要(英文)：In saintpaulia, various flower mutants could be generated by the adventitious shoot in tissue culture. The purpose of this study is to investigate that the mutants were generated by changes of gene expression patterns or epigenetic controls and to clarify the mechanisms of forming the flower patterns. The genome sequence of saintpaulia cultivar 'Kilauea' has been determined by using the long-reads sequenced by Sequel. The 2,200 contigs (total length: 708.7 Mb, N50 length: 1.3 Mb) were determined as draft genome sequence. The genes were predicted by using RNA-Seq data sampled from several kinds of tissues. The phased genome sequence of the cultivar 'Kilauea' was also constructed by using 10x Genomic linked-reads, but currently the analyses for comparisons of alleles and construction of co-expression networks were not performed.

研究分野：バイオインフォマティクス

キーワード：花の模様形成 ゲノム解読 RNA-Seq解析 フェージング エピジェネティックス データベース

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

セントポーリア (*Saintpaulia* spp.) は、花色の多様性に加えて、複雑な模様を持つ品種が多数存在する。組織培養によって不定芽 (葉脇など特定の箇所以外から出る芽) を誘導すると花の縞のパターンが逆転した個体やまだら模様の個体、単色の個体など多様な変異体が発現する。セントポーリアはゲノムサイズが小さいこと、二倍体であり他の品種との交雑が可能であることなどから、花の模様研究のモデル植物として最適であると考えられている。縞模様品種の中には不定芽由来植物において縞模様が発現する品種群があり、このような品種群から植物体を再生させると、親植物体と同じ縞模様が出現するとともに、親と縞のパターンが逆転する個体、まだら模様の個体、単色の個体など多くの変異体が発現するタイプ (非キメラ) が存在する。一方、出現個体の割合では縞模様の表皮の表現型を持つ単色個体が多く、また、表皮を除去した外植体を用いた場合においても、親個体の L2 層の表現型の単色花個体が多くなるタイプ (キメラ) も存在する。縞模様品種は全て周縁キメラとして扱われ、非キメラタイプの縞模様に関しては組織培養個体の分離がはっきりとしないため長年の間考察が混乱しており、エピジェネティックな制御を受けている可能性が考えられている。花の模様は遺伝子不安定性のみならず遺伝子間の相互作用などが関わる複雑な現象であり、これらの現象が生じる原因を明らかにすることが重要である。

2. 研究の目的

花の模様形成がエピジェネティックな制御を受けているのか、もしくは、遺伝子発現による制御を受けているのかを網羅的に調べるために、まずは、非キメラであるセントポーリア (*Saintpaulia ionantha*) の品種 'キラウエア' について、精度の高いセントポーリアのゲノム配列を構築する。'キラウエア' とキメラである品種 'チョコ' の縞模様が異なる各変異系統について、花弁部位別および茎頂分裂組織茎、花弁などにおける RNA 転写発現の解析を行い、両者間の発現状態を比較する。次に、全ゲノム DNA メチル化解析を行い、両者間のメチル化状態を比較する。さらには、ゲノムワイドなメチル化および遺伝子発現に基づき共発現解析を行うことで共発現ネットワークを構築し、代謝経路の解析、アソシエーション解析を実施することで、模様形成の原因となる遺伝子を特定する。

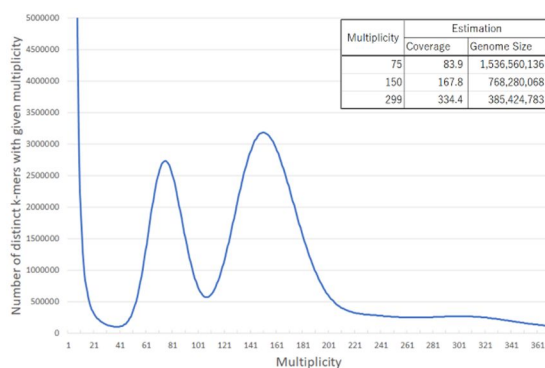
3. 研究の方法

Pacific Biosciences 社製の Sequel を用いて 'キラウエア' について 12 セル分のロングリードデータを取得し、FALCON により *de novo* アセンブルを行い、FALCON-Unzip によりフェージングを行う。'キラウエア' の縞花の花弁中心部の白部と青部、白花と青花の蕾、'チョコ' の青花、白花の花弁からそれぞれ得られた RNA-Seq リードに対して、Trinity により *de novo* アセンブルを行い、unigene を構築し、その後、発現量解析を行う。Bisulfite-Seq を行い、ゲノムワイドなエピジェネティックの状況を調べる。

4. 研究成果

まずは、ゲノムワイドな解析の基盤として、セントポーリア 'キラウエア' の全ゲノム配列の解読を行った。セントポーリアは他殖性であるため、薬培養由来の半数体を用いた。Sequel を用いて得られた 12 セル分のロングリードに対して FALCON を用いて *de novo* アセンブルを行い、FALCON-Unzip を用いてフェージングを行った。この結果、3,221 本の primary コンティグ (総延長: 704 Mb、N50 長: 1.25 Mb) と 1,508 本の haplotig (総延長: 121 Mb、N50 長: 131 kb) を合わせることで 4,692 本からなるドラフト配列 SIO_r1.0 (総延長: 825 Mb、N50 長: 958 kb) を構築した。primary コンティグの BUSCO の Complete は 97.1% であった。その後、イルミナ社製 HiSeq によりゲノム由来のペアエンドリードを取得し、kmer 頻度分布を調べたところ、明瞭な 2 つのピークが見られたことから対象としたゲノムはヘテロ性が高いことが分かった。kmer 頻度分布のプロットを用いてゲノムサイズを推定した結果、768 Mb となった (図)。

そこで、最新版の FALCON と FALCON-Unzip を用いヘテロ性を考慮して *de novo* アセンブルとフェージングを再実行し、Arrow と Pilon で補正した結果、2,200 本の primary コンテ



ィグ(総延長:708.7 Mb、N50 長:1.3 Mb)と7,658本のhaplotig(総延長:308.7 Mb、N50 長:141.0 kb)が得られた。primary コンティグの BUSCO の Complete は 97.3%(Complete and single-copy:64.1%、Complete and duplicated:33.2%)、Fragmented は 0.6%、Missing は 2.1% となった。SIO_r1.0 よりも良好な結果が得られたため、本結果をドラフト配列 SIO_r2.0 として決定した。

‘キラウエア’の縞花の花弁の白部(中心部)と青部(周縁部)、白花と青花の蕾、‘チコ’の白花、青花の花弁に対してイルミナ社製 HiSeq により得られた RNA-Seq リードをすべて合わせて *de novo* アセンブルを行うことで 161,685 本のコンティグ(BUSCO Complete:95.9%)を得た。その後、バリエーションの除去を行うことで 36,670 本の unigene(BUSCO Complete:85.4%)を構築した。その unigene に対して‘キラウエア’と‘チコ’の RNA-Seq リードをそれぞれマッピングすることで発現量の差を調べたが、有力な候補となる遺伝子は見当たらなかった。

次に、‘キラウエア’の RNA-Seq リードを SIO_r2.0 の primer 配列にマッピングし、BRAKER2 により遺伝子予測を行ったところ、143,914 個の遺伝子が得られた。次に、EMBL-EBI の TrEMBL(UniProt、Swiss-Prot) NCBI の NR、*Erythranthe guttata* の pep 配列に対するホモロジー検索、Pfam に対するドメイン検索、RNA-Seq のマッピングを行った。ホモロジー検索とドメイン検索でヒットし、RNA-Seq のマッピングの結果 TPM>1.0 の条件を満たすものを信頼度が高い遺伝子セット HC(High-Confidence)と定義し、それ以外を信頼度が低い遺伝子セット LC(Low-Confidence) 転移因子 TE(Transposable Element)と定義した。その結果、HC(配列数:35,648)、LC(配列数:72,905) TE(配列数:26,557本)となった。HCの BUSCO の結果は Complete が 93.3%(Complete and single-copy:62.4%、Complete and duplicated:30.9%)、Fragmented が 3.0%、Missing が 3.7%となった。一方、LCの BUSCO の結果は Complete が 8.1%(Complete and single-copy:6.3%、Complete and duplicated:1.8%)、Fragmented が 4.7%、Missing が 87.2%、TE については、Complete が 0.7%(Complete and single-copy:0.6%、Complete and duplicated:0.1%)、Fragmented が 0.3%、Missing が 99.0%となった。上記の結果から、信頼度の高い 35,648 個の遺伝子を得ることができた。SIO_r2.0 のゲノム配列と遺伝子配列(HC)を表に示した。

葉や葉柄、カルスなど 19 サンプルについて HiSeq を用いてシングルエンド RNA-Seq を取得し、白色およびピンク色の花など 6 サンプルについて HiSeq を用いてペアエンド RNA-Seq を取得した。また、Sequel により Iso-Seq データを取得した。Iso-Seq については 148,645 リード(総延長:241 Mb、BUSCO Complete:91.3%)が得られ、それに対してクラスタリングを行い、重複を除いた結果、61,750 本の Non-redundant transcripts 配列(総延長:109 Mb、BUSCO Complete:88.4%)が得られた。その配列に対して ANGEL により ORF を予測した結果、27,106 本の完全長 ORF(総延長:31 Mb、BUSCO Complete:55.0%)が得られた。現在、その配列を遺伝子予測に用いることで、予測精度の向上を図っている。

RNA-Seq 以外にも‘キラウエア’の縞花の花弁の白部と青部から small RNA-Seq をイルミナ社製 MiSeq により各サンプルあたり 36 塩基で 5M リードを取得し、‘チコ’についても白花と青花の花弁から small RNA-Seq を取得した。今後、ゲノム配列に対して small RNA をマッピングし、マップされた領域を調べる。

また、‘キラウエア’と‘チコ’についてフェージングされた配列間で比較するため、10x Genomics の linked-reads を得た。Linked-reads に対して supernova により *de novo* アセンブルしたところ、‘キラウエア’については 28,863 本の pseudohap 配列(総延長:637.5 Mb、N50 長:669.8 kb、BUSCO Complete:96.0%)、‘チコ’については 27,711 本の pseudohap 配列(総延長:638.0 Mb、N50 長:1.1 Mb、BUSCO Complete:96.5%)を得た。さらに、‘キラウエア’と‘チコ’についてフェージングされた pseudohap2.1 と pseudohap2.2 をそれぞれ得た。

フェージングされたハプロイド単位でのスキャフォールドを得ることができたため、今後、対立遺伝子間の配列の違いも考慮して RNA 解析などを行う。RNA 解析を行うことで花の様形成が遺伝子による制御でないと考えられた場合、Bisulfite-Seq 解析を実施する。本研究によりセントポーリアのゲノム配列の解読を進めることは出来たが、ハプロタイプを考慮した高精度な共発現ネットワーク解析、代謝経路の解析、アソシエーション解析を実施することが出来なかった。引き続き解析を実施することで花の様形成のメカニズムについての研究を進める。

表 セントポーリア‘キラウエア’のゲノム配列および遺伝子

	primary	haplotig	cds(HC)
配列数	2,200	7,658	35,648
総延長(bases)	708,683,047	308,684,504	42,852,762
平均長(bases)	322,129	40,309	1,202
最大長(bases)	8,369,882	1,168,399	17,301
最小長(bases)	16,884	104	87
N50長(bases)	1,262,744	141,025	1,638
G+C%	35.1	35.1	45.0
Complete	97.3	64.5	93.3
(Single)	64.1	57.2	62.4
(Duplicated)	33.2	7.3	30.9
Fragmented	0.6	4.4	3.0
Missing	2.1	31.1	3.7

5. 主な発表論文等

〔雑誌論文〕 計0件

〔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
研究分担者	細川 宗孝 (Hosokawa Munetaka) (40301246)	近畿大学・農学部・教授 (34419)	
研究分担者	白澤 健太 (Shirasawa Kenta) (60527026)	公益財団法人かずさDNA研究所・先端研究開発部・主任研究員 (82508)	