

令和 3 年 6 月 7 日現在

機関番号：15301

研究種目：挑戦的研究(萌芽)

研究期間：2019～2020

課題番号：19K22434

研究課題名(和文)植物短鎖機能性ペプチド遺伝子の探索・解析基盤の開発

研究課題名(英文)Dissection of iron deficiency response involving novel short peptide FEP1.

研究代表者

平山 隆志(Hirayama, Takashi)

岡山大学・資源植物科学研究所・教授

研究者番号：10228819

交付決定額(研究期間全体):(直接経費) 4,900,000円

研究成果の概要(和文):イネ科のモデル植物ミナトカモジグサを対象に、光環境応答処理を行った植物から、網羅的な転写物解析、ribo-RNAseq解析を行うことで、実際に翻訳がなされている短鎖ペプチド遺伝子を同定した。少なくとも百遺伝子は、新規の短鎖ペプチド遺伝子であった。また、non-coding RNA遺伝子の配列セットを定義するために、21の組織から抽出したRNAをプールし、Cap-trapping法で得た完全長cDNAライブラリの配列データを解析し約1万の非冗長なlncRNAs候補配列を同定した。タンパク質コード遺伝子とlncRNA候補配列の一部を教師データとして、判別モデルの作成を試みている。

研究成果の学術的意義や社会的意義

イネ科のモデル植物であるミナトカモジグサから、ribo-seq解析により、実際に翻訳されている新規の遺伝子を含む短鎖ペプチド遺伝子を300以上同定した。これらのデータを教師データを用いることで、植物ゲノム上の短鎖ペプチド遺伝子を探索するモデルの構築が可能となった。また、これらの遺伝子はこれまでに説明できていない現象の理解を補助する情報を提供する可能性がある。これらの情報は、植物遺伝子の機能の理解、生命現象の分子レベルでの理解、さらに人為的操作技術の開発に大きく寄与する。

研究成果の概要(英文):We identified short peptide genes that are actually translated by comprehensive transcript analysis and ribo-RNAseq analysis from the light-responsive treated plants of the grass model plant, *Brachypodium distachyon*. These included at least 100 short peptide genes that have not been reported before. To define the sequence set of non-coding RNA genes, we pooled RNA extracted from 21 tissues and analyzed the sequence data of a full-length cDNA library obtained by the cap-trapping method to identify about 10,000 non-redundant lncRNAs candidate sequences. We are attempting to create a discriminant model using some of the protein-coding genes and lncRNA candidate sequences as teacher data.

研究分野：植物分子遺伝学、植物分生物学

キーワード：短鎖ペプチド遺伝子 イネ科植物 rib-seq non-coding RNA 機械学習 *Brachypodium*

1. 研究開始当初の背景

これまでタンパク質をコードしないと考えられていた RNA 遺伝子(ncRNA 遺伝子)が、実は 100 アミノ酸残基未満の短いペプチドをコードし、生体内で重要な機能を担っている例が動物、植物を問わず報告が相次いでいる。これらは、ゲノム内には種特異的な機能性ペプチド遺伝子が数多く存在する一方で、その機能の大部分は未解明であり、それらの網羅的な同定と機能解析のための新たな研究基盤の必要性を示唆している。

2. 研究の目的

本研究は、機能性短鎖ペプチドをコードする遺伝子(機能性 sORF)の網羅的な同定とハイスループットな機能解析を可能にする基盤を構築することを目的とする。ゲノム解読が進む多様な生物種や集団から機能性 sORF を網羅的に同定し、その生理機能や遺伝子間相互作用、集団内での進化等を解明する新技術を提供することを目指す。この様な機能性短鎖ペプチドの網羅的同定にゲノムスケールの網羅性を与えられることで、ゲノム科学における遺伝子の構造についての理解が刷新されると期待される。さらに、機能性短鎖ペプチドの生理機能の解明は、環境適応性を担う遺伝子制御ネットワークに新たな遺伝子グループの統合を促し、細胞システムの理解をより深化すると考えられる。

3. 研究の方法

イネ科植物のミナトカモジグサ(*Brachypodium distachyon*)を対象に研究を行う。ゲノム進化の関係が明確で、亜科内や属内での機能性 sORF の多様性や保存性を議論できる、世界の多様な環境に適応しており機能性 sORF と環境適応性の関係を議論できる、集団レベルのゲノム解読が進んでおり機能性 sORF の種内変異を議論できる、などの目論見からである。以下の3つ研究を進める。

(1) 翻訳 sORF のカタログ化技術の開発：

機能性 sORF の探索では、コドン利用様式やペプチドの配列の保存性があてにならない。そこで、その発現と翻訳をそれぞれ RNA-seq と Ribo-seq(リボソームプロファイリング)で網羅的に解析し、翻訳される sORF 候補配列をカタログ化する。光条件を変化させた植物体から得た RNA サンプルを解析し、ストレス応答や生長制御に関わる機能性 sORF 候補配列を選抜する。

(2) ゲノム情報から機能性 sORF を見つける sORF 探索 AI の構築：

(1)によって得られた翻訳される sORF 候補配列カタログをサポートベクターマシン等により機械学習し、DNA 配列から翻訳される sORF を識別する AI を構築する。また、(2)で得られるデータも統合し、sORF に紐づく GRN に特徴的な機能に基づいて機能毎にグルーピングする AI を構築する。

(3) sORF がコードするペプチドの高速機能解析技術の開発：

機能性 sORF 候補配列を GFP 遺伝子とともに植物のプロトプラスト(細胞壁を取り除き、遺伝子導入を容易にした細胞)に導入し発現させる。sORF 導入細胞を、GFP 蛍光を指標にセルソーターで回収し、トランスクリプトーム解析に供する。遺伝子発現制御ネットワーク(GRN)への影響を評価し、導入した sORF の生理機能を推定する。

4. 研究成果

ミナトカモジグサ Bd21 株を、16 時間明期 8 時間暗期の条件で生育したのち、連続光下で 7 日間生育した。その後、暗所 0, 0.5, 1, 3 時間、暗所 3 時間後さらに明所 0.5, 1, 3 時間処理した植物個体 7 種類のサンプルを取得した。

1) ribo-seq 解析：これらのサンプルの組織を破砕し、全 RNA サンプル(リボソーム画分を含む)を抽出した。このうち 30ugRNA 分のサンプルを用いて、RNaseI 処理を行なった後、スクロース密度勾配遠心を行い、リボソーム画分を精製した。リボソーム画分から RNA を抽出した。RNA 変性ポリアクリルアミド電気移動で分画し 17-50bp に該当する RNA を精製した。得られた RNA にバーコード配列を含むリンカーをそれぞれのサンプルごとに付加したのち、RNA 変性ポリアクリルアミド電気移動で分画して、リンカーが結合した RNA を精製した。Ribosomal RNA を除去したのち、cDNA を合成、環状化し PCR で 8~10 回増幅しポリアクリルアミド電気泳動で分画して DNA 断片を精製し、ライブラリーとした。

2) RNA-seq 解析：項目 1) の全 RNA サンプルの 10ug 分を用いて、ribosomeRNA を除去したのち、RNAseq 用ライブラリーを作成した。

3) 上記 1) 2) で作成した配列解析用ライブラリーをそれぞれ次世代シーケンサーにより網羅的配列解析を実行した。得られた配列解析データのクオリティを整えたのち、RiboTaper により翻訳開始点を探索した。これにより、翻訳されていると推定される 30 アミノ酸残基以上 100 アミノ酸残基以下の短鎖ペプチド遺伝子を 300 個以上同定することに成功した。これらの配列解析から、シロイヌナズナやイネなどでこれまでに報告のない新規の短鎖ペプチド遺伝子かどうか評価した。その結果、少なくとも 100 以上の遺伝子が新規のものと考えられる。これらの

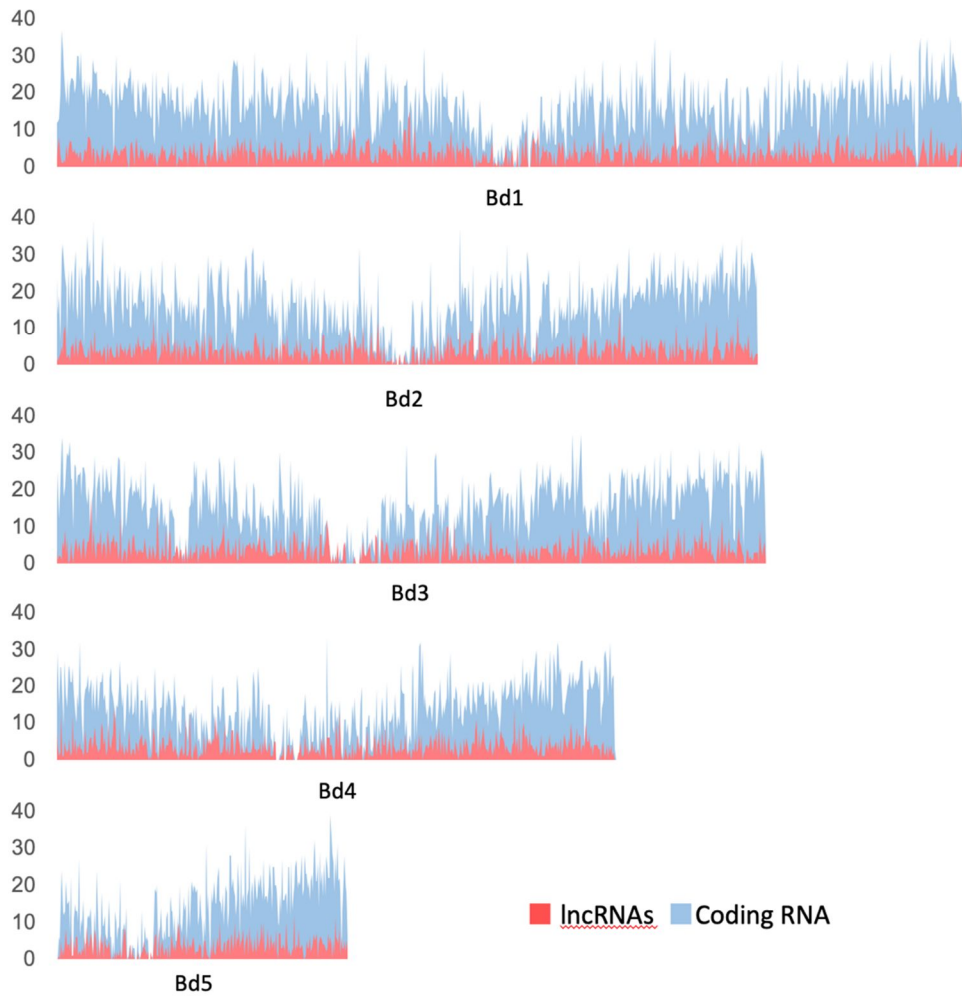
機能を明らかにすることで、今まで説明ができなかった生命現象について理解する糸口が見出せる可能性がある。

4) 短鎖ペプチド遺伝子の分子機能を明らかにするための、プロトラストを用いたハイスループット機能解析プラットフォームの構築およびその運用試験を行うために、新規の短鎖ペプチド遺伝子と植物種間で保存されているが機能不明のペプチド遺伝子、さらにレファレンスとして機能既知のペプチド遺伝子類似遺伝子、計 30 遺伝子を選抜し、それらの cDNA をクローニング、植物細胞発現ベクターに挿入したコンストラクトをそれぞれ構築し、トランスフェクション用高品質プラスミド DNA を得た。

transcript_id	ORF_pept
KQK12381	MASVGFGTSMVAAPVAAAAGMGSGRRLRRTQVVAATRGOAAPPKKQEK
KQK03350	MEGSSSRTPEITVPAPRPAAGGGAVDAVKAATKEPVSPGSPSPASRK
KQK08916	MSTAAANWCYATVAPRAKSVVVASLGTAPSSSSGFRPRLIRNAPVQ
PNT76909	MASRILVVLVIMASLMFIAQDVLAAARVLAQTSYIGDFPPAASYGSA
KQK14585	MALASNRDLFTYGGDKPTLVPIAKAIKMEAAAAGAGGPWLPQAAA
KQK11135	MAHFQEVYDCSEEVAMGNPAGGFRHGHGSGGGIQQHVAKETFVQE
KQK11732	MRAARKAGRAEVEVDDPRIGDWAGPQLELGFPGGNTNHIQTQKQHVK
KQJ84283	MASPATPEPTGTQEEADARQSLMGISRSVPEAGAALSGKLPHNGRV
KQK08966	MAPASSKVMGHLAQNGGIASVAVYAAAPCDACCGGSRHRKAETDGD
KQK10926	MPRLHADSSCSEQAAVLSPDQFPGGGEVQSQPKISEWERSFSTTTAD
KQK06152	MKSSTLTAILVLAQILVMGLAEVNAVGYWPKCCDNCRTFSGVMVCCD
KQJ94588	MDDDDSSSSPVEKISEQNSNTIASSSSSTGSTSDTESDKSDLQISDKS
KQJ93551	MSKMSFRALFLAVLMIICALQTVSVQGGHGSVSSGIRNAPPSRPSGRV
KQK01658	MAQPIDGSPVQGLWGPVQSFAGPHGPADMVPMRPAAGAQQV
KQK00372	MDARAPRHQRLRTPRGPPTPNHPRNPSTPGQNPSPAPAAAPRSVAT
KQJ84358	MAAVANSKGRAIAGNFVARILAGKAASPRAVHASAYDKNVEEQVRPAF
KQJ82670	MAASLNRLRSIGIHLLATGVEASKPASRGFHATGVKRMGGHGHDEPY
KQJ93597	MALRALYNEIRAMKVREVPAYLKPRLTWGNVKKSTDQAVDRYIEKYIE
KQK21156	MARERNLEKLGKGGKGSQLEANKKAMNIQKICMQTFICTTSEAKCKEH
KQK13912	MAGIHKIEEKLHMGGSDHKDEHKKEEHEKHKDGEHKKDEGHEKGM
KQK04141	MERVASRCASLLAQRGYSVSAALAKGAGRSAEERTAVAVKRTMAAKK
KQK10364	MVLAATVLAPPAMVRAAISCSAVYSTLMPCLQFVQGGGAPSRGCCSGI
KQJ97366	MAKAAATCVVAMALALAVLMALASGATAAQCNCAGQLIVCAPAIIGGAAP
KQJ94500	MLEETKHGRNEETIKYQEEIATLNEEMKKLQVQEKTEEVHVAEDKAL
KQK18573	MSEESGRPLPKFGEWDVNDPASADGFTVIFNKARDEKAGNGQDTE
KQJ94665	MSGVWVFKDGVRRVEKDNPGSSSSNSGSMGRPKVLVHVPSGEVVL
KQK07642	MAGTANCIDIIAILPPLGVFLKFGCGHEFWICLLTFLGYIPGIYAIYAITK
KQK14336	MWQEPLNPGNWKEEHFVLSLAMWGGIFYGVAKLFGGKEDKTEVAF
KQK15163	MAIRGVPSAREMTVEDFKLWLKQFVDGDGRISRGEALRRRRGGW
KQJ99669	MASNVDTTGEEGTFRSTTPTGAAASSPRMMRRSFSSASSSGSHGGG

機能解析ように選抜された短鎖ペプチド遺伝子 (30) のリスト (配列は一部)。上半分は他の報告がない新規植物遺伝子。下半分は、機能既知を含む報告のある遺伝子。

5) 配列情報から短鎖ペプチド遺伝子をコードするかどうかを判別するモデルの作成を行うために、ミナトカモジグサにおいて、non-coding RNA 遺伝子の配列セットを定義することを目的として、完全長 cDNA ライブラリーの配列データを解析し lncRNAs 候補配列を同定した。具体的には、ミナトカモジグサの様々な組織(21 組織)から抽出した RNA をプールし、Cap-trapping 法で得た cDNA ライブラリーを Illumina の次世代シーケンサーにより、鎖特異的なシーケンシングを行った。これまでに収集したミナトカモジグサの RNA-seq データと合わせてリファレンスゲノムにマッピングし、5' および 3' 端を完全長 cDNA 配列データから同定するとともに、複数のアルゴリズムに基づくコーティングポテンシャルが閾値未満となる約 1 万種類の lncRNAs 候補配列を同定した。これらの lncRNAs 候補配列セットは、sORF かどうかを判別する AI 学習をするうえでのネガティブデータセットとして用いることができる。他方、上記の ribo-seq 解析で得た短鎖ペプチド遺伝子配列は、ポジティブデータセットとして用いることができるため、sORF 判別モデルを機械学習によって構築するための基盤となるデータセットが整備できた。また、転写制御ネットワークの特徴的な機能に基づいて、機能毎に各遺伝子をグルーピングするために、根・葉・穂などの代表的な器官と、塩・乾燥・高温などの環境ストレスを処理した組織など、32 組織から得られた RNA-seq データを得た。上記、RNA-seq 解析で得た遺伝発現データと合わせて、各遺伝子の発現パターンと転写因子から GRN を推定し、各 GRN のモジュールにエンリッチする GO Terms や転写因子の機能等から各ネットワークに特徴的な機能を推定するための基盤を構築した。



21 の組織を対象にした Cap-trapping 法による RNA 解析結果から抽出したタンパク質コード遺伝子と non-coding RNA 遺伝子のゲノム上の分布図。これらのリストと、ribo-seq 解析のデータを教師データとして sORF 判別モデルの構築が可能となった。

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 1件）

1. 発表者名 Takashi Hirayama, Satoshi Okada, Gui Jie Lei, Naoki Yamaji, Keiichi Mochida, Jian Feng Ma
2. 発表標題 FEP1, a novel short peptide involved in iron homeostasis of Arabidopsis
3. 学会等名 Chemistry and Plant Biology, ERATO International Symposium (国際学会)
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

コロナ感染症蔓延の影響のため、計画していた国際共同研究によるribo-seq解析は実施できず、研究期間および配分予算内での研究計画全ての実施は困難となった。研究を継続し当初の目的を達成する予定である。

研究内容
<http://www.rib.okayama-u.ac.jp/ers/research2.html>

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	持田 恵一 (Mochida Keiichi) (90387960)	国立研究開発法人理化学研究所・環境資源科学研究センター・チームリーダー (82401)	

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
---------	---------