

令和 2 年 6 月 22 日現在

機関番号：82401

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K07265

研究課題名(和文) C4草本植物ソルガムの転写関連データベースの構築とその応用

研究課題名(英文) Construction of transcription related database of C4 grass plant, Sorghum bicolor and its application

研究代表者

蒔田 由布子 (Makita, Yuko)

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・研究員

研究者番号：80443026

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：ソルガムは世界五大穀物の一つであり、環境ストレスにも強いいため、バイオエタノール生産植物として注目されている。ゲノム配列以外の情報リソースに不足しており、高糖性・高バイオマス研究への応用や、環境ストレス耐性メカニズムの解明のための情報基盤整備として、ソルガムのデータベースを更新した。本課題では、糖蓄積器官の異なる2系統の発現比較を行なうことで、品種間の発現量、転写制御の比較を進めた。複数のトランスクリプトーム解析手法を組み合わせることで、効率的に糖・デンプン合成遺伝子群の転写ネットワークの解析を進め、MOROKOSHI(<http://sorghum.riken.jp>)を通じて結果を公開している。

研究成果の学術的意義や社会的意義

ソルガムは世界五大穀物の一つであり、環境ストレスにも強いいため、バイオエタノール生産植物として注目されている。ゲノム配列は公開されているものの、それ以外のリソースに著しく不足している。そこで高糖性・高バイオマス研究への応用や、環境ストレス耐性メカニズムの解明のためのリソース整備として、ソルガムのトランスクリプトーム情報基盤の整備・公開を行なう。本研究では特に、糖蓄積器官の異なる2系統の発現比較を行

研究成果の概要(英文)：C4 grass plant, sorghum ranks fifth in the global cereal production and shows strong environmental stress tolerance against drought, heat, salinity and flooding. Sorghum is also used for bioethanol production. Although the genome sequence was determined in 2009, it is still lacking the other bioinformatics resources. In this project, we updated our sorghum database to apply for the high sugar and high biomass researches and elucidation of mechanism of environmental stress tolerance. We obtained RNA-seq, CAGE and gDB-seq data from two different sugar accumulation strains, grain sorghum and sweet sorghum. Comparison of these two strains, we obtained the expression difference on the sugar transporters. Also, CAGE data suggested tissue specific transcription start sites and different transcriptional regulations. All the transcriptome data and constructed networks are accessible from MOROKOSHI database (<http://sorghum.riken.jp>).

研究分野：植物インフォマティクス

キーワード：トランスクリプトーム ソルガム RNA-seq CAGE gDB-seq C4草本植物 データベース

1. 研究開始当初の背景

ソルガム(別名モロコシ、コーリヤン)は、イネ科の C4 植物であり、世界五大穀物の一つである。サトウキビやトウモロコシの近縁で、それらと比較しても、高温、乾燥、塩害、多雨といった環境ストレスに強く、生育コストが低く抑えられる特徴がある(表1)。また食料との競合が少ないため、エタノール生産植物としても注目されている。研究分野においては、アメリカと中国が力を入れており、2009年には米中心の国際チームがゲノム配列を決定した(Paterson, et al. Nature 2009)。

(表1)エタノール生産植物の比較

	ソルガム	サトウキビ	トウモロコシ
収穫日数	4ヶ月	12ヶ月	4ヶ月
給水量(m <sup>3</sup> )	4,000	36,000	8,000
全エタノール(l/ha)	3,160	8,925	3,216
水コストを含めたエタノールコスト(US\$/kl)	75.3	111.5	89.2

(国際半乾燥熱帯作物研究所の資料を改変)

バイオマス植物として重要なソルガムではあるが、分子メカニズムの研究は進んでおらず、研究リソースも不足している。そこで当研究室では、ソルガムのバイオマス工学への応用や、環境ストレス耐性メカニズムの解明のためのリソース整備として、完全長 cDNA ライブラリ作成した(Shimada, et al. DNA Research. 2015)。その結果、約 1.1 万遺伝子(ソルガムの全遺伝子の約 1/3)をクローン化することに成功した。さらに、ソルガム全遺伝子の発現解析をするために、我々の組織特異的な RNA-Seq データに加え、NCBI SRA から公開データを利用することで、ソルガムにおいては初となるトランスクリプトームデータベース MOROKOSHI (<http://sorghum.riken.jp>)を公開した (Makita, et al. Plant Cell Physiology. 2015)。MOROKOSHI は、本データベース(DB)以前にゲノムワイドな DB が存在しなかったことから、世界中からアクセスされている。一方で、データサンプリング数やデータの種類の限定的であるために、個別研究を進める研究者からは、さらなるデータ拡充を求められていた。

2. 研究の目的

ソルガムは表2に示すように大きく2系統に分けられる。分子生物学的研究でよく使われるゲノムの決定された BTx623 株は「グレイソルガム」と呼ばれ、種子にデンプンを蓄積する。一方でバイオエタノールとしてより注目されている系統は「スイートソルガム」と呼ばれ、サトウキビ同様に茎に糖を蓄積する。バイオエタノール精製過程においてスイートソルガムは、糖化のプロセスを省略することができるため、より注目度が高い。スイートソルガムは、グレイソルガムとゲノム構造がほぼ同じであるものの indel や SNP が構造遺伝子やその制御領域に多く見られ、それらが糖の蓄積部位や形態に影響を与えていると考えられる。グレイソルガムとスイートソルガムの比較は、栽培化の過程で選ばれた形質の差異であり、それがどのようなゲノム上の変異によって引き起こされたかを調べることは、基礎及び育種の観点から重要となっている。本研究では、グレイソルガムとスイートソルガムの発現量と転写開始点、転写因子の結合配列を比較することで、糖・デンプン合成遺伝子群の転写ネットワークの構築を行う。それにより高糖性ソルガム研究への貢献を目指すとともに、転写関連オミックスデータを充実させ、MOROKOSHI の更新と共に充実したデータ提供を目的とする。

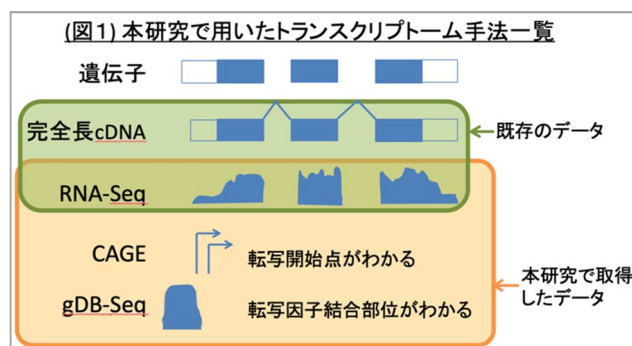
(表2)グレイソルガムとスイートソルガムの比較

	グレイソルガム	スイートソルガム
本研究で使用する系統	BTx623 (ゲノム決定株)	Keller
蓄積器官	種子	茎
主な合成物	デンプン	糖
背丈	1m	3m

3. 研究の方法

本研究では、図1に示す4つのトランスクリプトームデータを取得することで糖・デンプン合成に関する転写ネットワークの構築・解明と、それに併せてソルガムのトランスクリプトーム情報基盤の整備を進めた。

まず、グレイソルガムとスイートソルガムの糖の蓄積に関する違いを遺伝子の発現量の差異と転写開始点の違いから調べるため、RNA-Seq データと CAGE (Cap Analysis of Gene Expression) データの取得を行った。グレイソルガム



とスイートソルガムの穂、種子、茎、葉おける RNA-seq データを取得することで、系統間・組織間での比較解析を進めた。また同様のサンプルで CAGE データを得ることで、転写開始点の違いと発現量の相関解析を行い、グレイソルガムとスイートソルガムの糖・デンプン合成の違

いを明確にすることを試みた。

また、直接的に糖・デンプン合成の転写制御情報を得るために、当研究室で開発された gDB-Seq (genomic DNA binding sequence) 法を活用した。gDB-Seq 法とは、*in vitro* で転写因子が物理的に結合できるゲノム上の部位を同定するハイスループットな手法である。本課題では、ソルガムの茎部の乾汁性に関わる Dry 遺伝子と呼ばれる転写因子の DNA 結合部位をゲノムワイドに調べた。以上、図 1 の 4 手法から得られるデータは全て MOROKOSHI データベースに登録し、データの公開も行う。

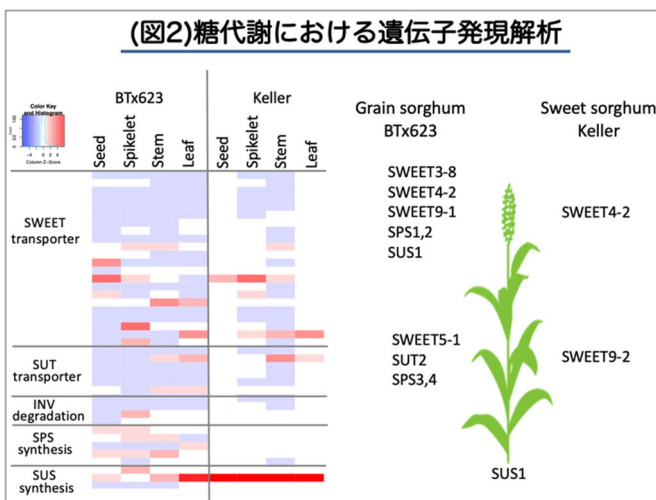
#### 4. 研究成果

##### (1) RNA-seq による系統間、組織別発現解析

種子に澱粉を蓄積するグレインソルガム (BTx623 株) と、サトウキビ同様に糖を茎に蓄積するスイートソルガムの発現比較を、穂、種子、茎、葉の 4 つの組織において行い、系統間・組織別の比較解析を行った。その結果、糖代謝において重要な遺伝子のヒートマップを図 2 に示す。

Sucrose synthase (SUS) においては、SUS1 はどちらの系統においても、全ての組織で高い発現が確認され、ソルガムの高糖性において重要な遺伝子であることが確認された。一方で、詳しく見てみると、グレインソルガムでは Sh1 の発現が、スイートソルガムでは Sus6 の発現が組織を問わず高い傾向があり、系統間での差も確認された。

また同様に、糖の輸送に関する SWEET トランスポータにおいては、糖を細胞へへ輸送する SWEET4-2 遺伝子がグレインソルガム、スイートソルガム に共通して種子と穂で高い発現を示し、SWEET4-2 遺伝子が茎よりも種子において重要な遺伝子である結果となった。一方で、SWEET4-2 以外の SWEET トランスポータにおいては、共通性が見られず、系統間での発現制御が保存されておらず、大きく変わっている様子を示した。次に、sucrose phosphate synthase (SPS) においては、グレインソルガムでの高発現に対し、スイートソルガムでは発現量が低く、系統間で差の見られる結果となった。SPS は糖代謝において重要な遺伝子であることから、スイートソルガムの発現の低さは想定外であった。この理由としては、スイートソルガム においては、sucrose のリン酸化に他の酵素 (他の経路) が主に使われていることが考えられた。

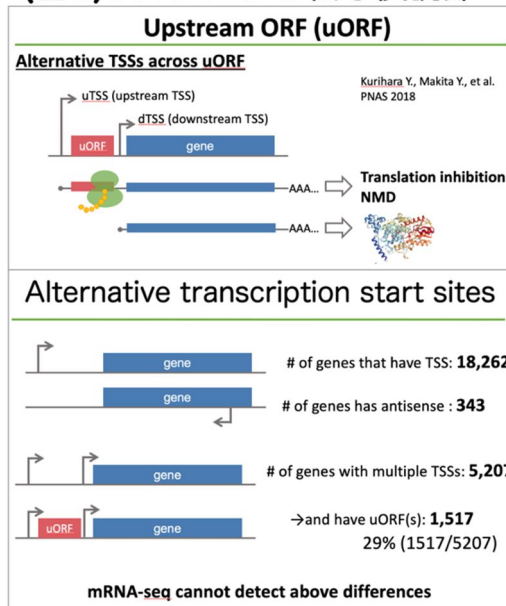


##### (2) CAGE 法による転写開始点(TSS)の同定

CAGE 法は *in vivo* でゲノムワイドに転写開始点を同定する方法であると同時に、その配列数 (リード数) から転写開始点ごとの発現量を知ることができる。本研究では同じ組織の RNA-seq と CAGE のデータを取得しているため、両者の発現量の相関を確認した。その結果、全てのサンプルにおいて 0.8 以上の高い相関性を示した。また転写開始点は、1 つのピークから遺伝子発現が行われている Steep 型と、転写開始点がゆらいで幅のある領域をもつ Broad 型に分けられる。今回、転写のゆらぎ幅と発現量の関係を確認したところ、Broad 型の方が Steep 型よりも発現量が高く、中でも TSS の幅が 50~100 塩基を超えるほどに広がるものほど、発現量がより高くなる傾向が確認された。これは、発現量が高くなるためには DNA の二重らせんが大きくほどかれる必要があり、その後の TSS の制御は、1 塩基レベルで行われていない高発現遺伝子が多いことが考えられた。

また組織別、系統別に転写開始点解析においては、まず、多くの糖代謝遺伝子の転写開始点(TSS)が系統間で保存されていた。一方で、Sucrose

##### (図3)uORFによる転写後調節



synthase 1 (SH1)においては、3箇所の TSS があり、それらを系統間、組織間で巧みに使い分けている様子が確認された。また starch synthase においては、組織別に TSS を使い分けており、その様子が系統間で保存されていた。

このように、転写開始点を使い分けることは、その後のタンパク質の局在と発現調節に大きく影響することがある。転写開始点が下流に移動すると、アミノ酸配列の N 末端側が数塩基から数十塩基短くなる。それにより N 末端にあった細胞内局在のシグナル配列が除去され、タンパク質の輸送先が変わることが知られている。今回の解析でもミトコンドリア移行シグナルが除去される例などが確認された。また細胞内局在以外にも、転写開始点をシフトさせる転写後調節として、uORF を介した制御がある(図 3)。uORF(upstream ORF)とは CDS の上流のプロモータ領域に存在する短い ORF 構造で、uORF が転写されると mRNA-seq 上は発現が確認されるものの、実際にはリボソームが uORF にコードされた短いペプチドのみを合成し、その後解離してしまうため本来の CDS では翻訳が行われなくなる。また NMD により mRNA 自体も分解が促進される。ゲノムワイドに uORF を調べる方法としては、CAGE, mRNA-seq に加え、リボソーム フットプリント法などを組み合わせることがある(Kurihara, Makita *et al.* PNAS. 2018)。今回は CAGE データから複数 TSS を持つ遺伝子を検出することで、uORF による調節下にあると考えられる候補遺伝子の探索を行った。その結果、組織別で 2 つ以上の TSS を持つ遺伝子が 5207 個みつかった。その内、それら 2 つの TSS 間に uORF の候補配列となるペプチド配列が見つかるものが 1517 遺伝子あった。これらはソルガムが組織別に uORF を使い分けることによる転写後調節を行っていることを示していた。

### (3) gDB-seq による D gene のゲノムワイドな結合領域

gDB-seq 法とは、転写開始点をゲノムワイドに同定する ChIP-seq 法の in vitro 版であり、形質転換を必要としないメリットがある。ソルガムの形質転換法は当研究室においても確立しているもの、安定した系統を作成し、選抜を終えるには 2 年かかる。そこで今回は、gDB-seq 法により、ソルガムの農業的な形質として非常に重要な、茎の部分の乾汁をコントロールする D 遺伝子 (Dry gene, 転写因子) の結合をゲノムワイドに調べた。その結果、78 遺伝子のプロモータ付近に D 遺伝子の結合を確認することができた。それら D 遺伝子の被制御遺伝子群を GO 解析により機能を確認すると、光合成関連やヌクレオチド代謝関連の遺伝子が確認され、今後の詳細な機能解析の一助となるデータを得ることができた。現在、本データにおいては論文を執筆中である。

### (4) ソルガムトランスクリプトーム DB、MOROKOSHI の更新

MOROKOSHI DB (<http://sorghum.riken.jp>) は、ソルガムの cDNA 情報と遺伝子機能アノテーションを提供するデータベースとして、2015 年より公開している (Makita Y., *et al.* Plant Cell Physiology. 2015)。本プロジェクトにより mRNA-seq に加え、CAGE, gDB-seq といった多角的なトランスクリプトームデータを統合することができた。また独自のデータに加え、12 論文から抽出した 48 サンプルの mRNA-seq データやそれらを用いた共発現解析による転写ネットワークの構築を行った。また公開データより PacBio シーケンサによる完全長 cDNA; iso-seq のデータ、クロマチンの開かれた領域を調べる ATAC-seq, in vivo でゲノムワイドなタンパク質の結合領域を調べる ChIP-seq のデータ統合も進めた。さらに、毎年開かれている日本国内のソルガム研究会においては、ソルガム、アノテーションを充実させる会を共同で主催し、日本中のソルガム研究者が集まるタイミングで、その貴重な知識を MOROKOSHI DB より公開すべく、マニュアル アノテーションの会を開催している。すでに 3 年連続で集まることで、多数の論文情報やシノニム情報、貴重な生物学的な機能に関する情報を集めることに成功し、本プロジェクトの最終年度に MOROKOSHI DB より公開を開始することができた。

また、ウェブアプリケーションの拡充としては、オンライン上で Heatmap を描くツールやサンプルごとの発現量を可視化するツールを R の Shiny を用いて実装した(図 4)。これにより、研究者が生データをダウンロードし解析する必要なく、自分の注目してい

(図4)MOROKOSHI データベース



る遺伝子の発現量の確認を進めることができる。またユーザからの要望により、データのダウンロード機能も充実させ、今までの全データを一括してダウンロードする方法に加え、Web サイト上で指定した複数遺伝子だけのデータダウンロードが可能となった。このように、ゲノムワイドなデータに加え、マニュアルキュレーションデータ、ウェブアプリケーションを拡充することができた。本データベースはソルガムのみならず、近縁種であるトウモロコシ・サトウキビのモデルデータベースとしても重要である。

5. 主な発表論文等

〔雑誌論文〕 計2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 Yukio Kurihara, Yuko Makita, Mika Kawashima, Tomoya Fujita, Shintaro Iwasaki, Minami Matsui	4. 巻 115
2. 論文標題 Transcripts From Downstream Alternative Transcription Start Sites Evade uORF-mediated Inhibition of Gene Expression in Arabidopsis	5. 発行年 2018年
3. 雑誌名 Proceedings of the National Academy of Sciences	6. 最初と最後の頁 7831-7836
掲載論文のDOI（デジタルオブジェクト識別子） 10.1073/pnas.1804971115.	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 -

1. 著者名 Tomoko Kuriyama, Setsuko Shimada, Minami Matsui	4. 巻 36
2. 論文標題 Improvement of Agrobacterium-mediated Transformation for Tannin-Producing Sorghum	5. 発行年 2019年
3. 雑誌名 Plant Biotechnol (Tokyo)	6. 最初と最後の頁 43-48
掲載論文のDOI（デジタルオブジェクト識別子） 10.5511/plantbiotechnology.19.0131a.	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計9件（うち招待講演 0件/うち国際学会 1件）

1. 発表者名 Minami Matsui, Tomoko Kuriyama, Mika Kawashima, Haruka Shimohira, Yuko Makita
2. 発表標題 Comparison of Expression profile of grain and sweet sorghum by using CAGE methods
3. 学会等名 第397回生存圏シンポジウム クリーンエネルギー生産に向けたリグノセルロース分子育種の現状と展望
4. 発表年 2019年

1. 発表者名 Tomoko Kuriyama, Yuko Makita, Mika Kawashima, Setsuko Shimada and Minami Matsui
2. 発表標題 ソルガムのマルチトランスクリプトーム解析と遺伝子導入技術の応用
3. 学会等名 第9回ソルガム研究会
4. 発表年 2018年

1. 発表者名 Yuko Makita, Mika Kawashima, Tomoko Kuriyama, Setsuko Shimada, Minami Matsui
2. 発表標題 MOROKOSHI: Transcriptome database in Sorghum bicolor and its updates
3. 学会等名 The 16th International Conference on Bioinformatics (国際学会)
4. 発表年 2017年

1. 発表者名 蒔田由布子、川島美香、栗山朋子、嶋田勢津子、松井南
2. 発表標題 C4草本植物ソルガムにおけるマルチトランスクリプトーム解析と MOROKOSHI データベースの紹介
3. 学会等名 第35回日本植物細胞分子生物学会大会
4. 発表年 2017年

1. 発表者名 蒔田由布子、川島美香、栗山(近藤)朋子、嶋田勢津子、松井南
2. 発表標題 バイオマス植物ソルガムにおけるマルチトランスクリプトーム解析
3. 学会等名 第40回日本分子生物学会年会
4. 発表年 2017年

1. 発表者名 蒔田由布子、川島美香、栗山朋子、嶋田勢津子、松井南
2. 発表標題 組織別RNA-seqによるグレイソルガム(BTx623)とスイートソルガム(Keller)の比較
3. 学会等名 第8回ソルガム研究会
4. 発表年 2017年

1. 発表者名 Minami Matsui, Tomoko Kuriyama, Mika Kawashima, Haruka Shimohira, Yuko Makita
2. 発表標題 Comparison of Expression profile of grain and sweet sorghum by using CAGE methods
3. 学会等名 第397回生存圏シンポジウム クリーンエネルギー生産に向けたリグノセルロース分子育種の現状と展望
4. 発表年 2019年

1. 発表者名 蒔田由布子, 栗山朋子, 下平春香, 嶋田勢津子, 松井南
2. 発表標題 ソルガムの転写制御解析とMOROKOSHI データベースの更新
3. 学会等名 第10回ソルガム研究会
4. 発表年 2019年

1. 発表者名 Tomoko Kuriyama, Yuko Makita, Mika Kawashima, Setsuko Shimada and Minami Matsui
2. 発表標題 ソルガムのマルチトランスクリプトーム解析と遺伝子導入技術の応用
3. 学会等名 第10回ソルガム研究会
4. 発表年 2019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
研究協力者	栗山 朋子  (Kuriyama Tomoko)		