

令和 3 年 4 月 20 日現在

機関番号：14603

研究種目：基盤研究(C) (一般)

研究期間：2018～2020

課題番号：18K05555

研究課題名(和文)植物コアプロモーターの配列特徴の解明

研究課題名(英文)Elucidation of sequence features of plant core promoter

研究代表者

加藤 晃 (Kato, Ko)

奈良先端科学技術大学院大学・先端科学技術研究科・教授

研究者番号：80283935

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：真核生物のプロモーター内にはコアプロモーターとして、TATA-box (TATA)、イニシエーター (Inr)、TFIIB-Recognition Element (BRE)などが存在することが知られており、RNAポリメラーゼIIによる転写開始に大きく関わっている。一方で植物コアプロモーターに関する知見は乏しいのが現状であった。本研究では、様々なコアプロモーターエレメントがmRNAの転写に与える影響を、植物を対象としてin vivoにて詳細に評価し、TATA-boxやYRルール、Tリッチ配列の転写開始点や転写効率の決定における重要性を示した。

研究成果の学術的意義や社会的意義

本研究で得られた成果は、今後の植物コアプロモーターの研究での重要な基礎となる知見を提供するものと考えられる。またこの知見を活用することで、コアプロモーターの改良による転写活性の増大が可能となり、植物で有用遺伝子を更に高発現させる上でも有用なものである。

研究成果の概要(英文)：It is known that among eukaryotic promoters, TATA-box (TATA), initiator (Inr), TFIIB-Recognition Element (BRE), etc. exist as core promoters. The core promoter is in demand for transcription initiation by RNA polymerase II. On the other hand, the current situation is that there is little knowledge about plant core promoters.

In this study, the effects of various core promoter elements on mRNA transcription were evaluated in detail in plants in vivo. As a result, it was shown that TATA-box, YR rule, and T-rich sequence are important for determining the transcription start point and transcription efficiency.

研究分野：遺伝子発現制御

キーワード：植物コアプロモーター 転写開始点 転写効率 翻訳効率

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

真核生物のプロモーター内にはコアプロモーターとして、TATA-box (TATA)、イニシエーター (Inr)、TFIIB-Recognition Element (BRE)などが存在することが知られており、RNAポリメラーゼ II による転写開始に大きく関わっている。これらコアプロモーターについて解析するためには、ゲノムワイドに転写開始点 (TSS) を決定することが必須であるが、植物ではほとんど行われておらず、植物コアプロモーターに関する知見は乏しいのが現状である。

我々はこれまでに、植物 mRNA の翻訳制御機構を解明するために、シロイヌナズナ等を対象として網羅的な転写開始点 (mRNA の 5'末端情報) の同定を行ってきた。得られた大規模データについてコアプロモーターの観点から解析したところ、多くの場合、転写開始点 (+1) は A/G、上流-1 位は T/C であり、動物と同様の YR ルールが保存されていた。また、上流-32~-25 位に TATA 様の配列が認められた。同様の解析をイネ、タバコ、バラについても行ったところ、共に YR ルールが保存されていたが、タバコとバラでは TATA の位置が 1 塩基下流側にシフトしていた。植物でよく用いられるプロモーターにカリフラワームザイクウイルス由来の 35S プロモーターがある。このプロモーターの TATA から転写開始点までの距離は、シロイヌナズナ内在プロモーターと比較すると 1 塩基短く、また、-1 位の配列は A であり、YR ルールからは外れている。そこで予備的に、この A を C に置換 (YR ルール) もしくは A の後に C を挿入 (TATA との距離) した場合の発現に及ぼす影響を一過性発現実験により調べたところ、蓄積 mRNA 量に大きな影響が認められた。一方で、転写開始点下流の配列 (5'非翻訳領域: 5'UTR 配列) が異なる場合には、変異の影響パターンが異なっていたことから、転写開始点近傍配列 (イニシエーター) 以外の下流配列もコアプロモーターとして重要であることが示唆されていた。

2. 研究の目的

本研究では、シロイヌナズナを対象として得られたゲノムワイドな転写開始点データを基に、*in silico* にて植物コアプロモーターの配列特徴を抽出するとともに、その抽出した特徴に変異や欠失を導入し、一過性発現実験による詳細な評価実験を行い、植物コアプロモーターの配列特徴を同定することを目的とする。

3. 研究の方法

(1) これまでに、網羅的な転写開始点に関する予備的なデータ処理は完了していた。一方で遺伝子 (プロモーター) の種類によっては、TATA の有無や転写開始点の収束 / 分散の度合いが異なっているため、それらに関して詳細なデータ処理を行い、遺伝子種 (プロモーター種) によってグループ分けを行った。

(2) プロモーター種によっては、コアプロモーター配列が異なることが予想されるため、上記グループ分けに従って、転写開始点の位置情報に基づいた周辺塩基の出現頻度を *in silico* にて解析し、各グループで共通するもしくは特徴的なモチーフを抽出した。

(3) 変異の導入等の予備実験では、レポーター活性を指標とした一過性発現実験を行ったが、レポーター活性は、転写および翻訳過程双方の影響を受ける。そのため、本研究では、代表的なプロモーター種に対して、それぞれ予想されるモチーフに変異を導入し、一過的にプラスミド DNA をシロイヌナズナプロトプラストに導入後、RNA を精製 (DNase によるプラスミド DNA の除去) し、定量 PCR によって転写された mRNA を定量した。また、変異等の導入による転写開始点の変化の有無を 5'RACE 法により確認した。

4. 研究成果

(1) まず、植物プロモーターでの TATA-box と TSS の収束の関係性を調べるために、データを解析した。確かな発現が確認できた 7164 の遺伝子を、最も主要な TSS の使用頻度、つまり 1 点への収束度合いを表す rTSS-usage の数値と、TATA-box が存在する TSS の上流 35 塩基から 26 塩基の 10 塩基内の T または A の比率に基づいて 25 のクラスに分割した (表 1)。その結果、rTSS-usage が低い分散型の遺伝子の行 (表 1) を見ると、TATA-less 型に属する遺伝子が多く、反対に rTSS-usage が高い収束型の遺伝子を見ると TATA 型に属する遺伝子が多いことが認められた。このことから TATA-box と TSS の収束には確かな関係性があることが明らかとなった。

表 1. TATA-box 配列領域の T または A の塩基比率と TSS の収束度合いを指標に分類された遺伝子の数

class	TA-ratio-1	TA-ratio-2	TA-ratio-3	TA-ratio-4	TA-ratio-5	total
rTSS-usage-1	368	312	274	245	234	1433
rTSS-usage-2	343	306	270	262	252	1433
rTSS-usage-3	294	312	293	269	264	1432
rTSS-usage-4	268	268	296	308	293	1433
rTSS-usage-5	160	235	299	349	390	1433
total	1433	1433	1432	1433	1433	7164

(2) 続いて、TATA-box を持ち TSS が 1 点に収束している遺伝子を「TATA 収束型」、TATA-box を持ちながら TSS が分散している遺伝子を「TATA 分散型」、TATA-box を持たないが TSS が収束している遺伝子を「TATA-less 収束型」、TATA-box を持たず TSS が分散している遺伝子を「TATA-less 分散型」と定義し、それらの遺伝子における、TSS 前後の塩基の出現頻度を解析した (図 1)。すべての型で、TSS 前後のコアプロモーター配列 Initiator (Inr) に由来すると考えられる塩基の偏りが顕著に検出された。特に偏りが強い位置は、TSS の前後 1 塩基であり、これは Inr の YR ルールを反映していると考えられる。一方で、この塩基の偏りは、TATA-less 型に着目すると分散型よりも収束型で顕著であり、また TATA 型に着目すると -1 位における C の比率が収束型より分散型で低いなど、収束型と分散型の間で YR ルールを有する遺伝子の比率や YR ルール内での塩基の組み合わせの出現頻度に違いが見受けられた。

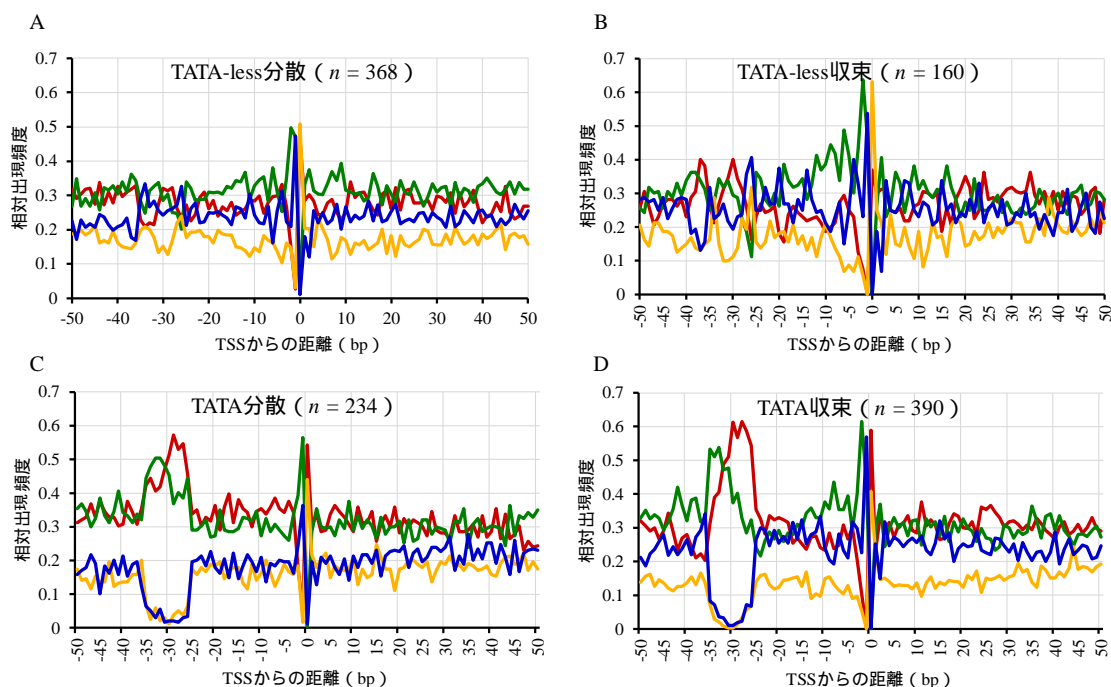


図 1 TATA-box のある領域の T または A の比率に基づいてクラス分けされた各遺伝子の TSS 周辺領域 (-50/+50) の相対的な塩基出現頻度

次に、YR ルールに着目し、その出現頻度の解析を行った。その結果、TSS が分散している遺伝子の中には、YR ルールから外れたものが存在した一方で、収束している遺伝子では、ほぼすべての TSS が YR ルールに従っていたことが明らかになった。加えて、TATA-less 型に着目すると、YR に従う配列パターンのうち「TG」や「CG」の比率は、分散型では相対的に低く、反対に「TA」は収束型で低い傾向があった。このことから、TATA-less 型における TSS の収束や分散に YR のパターンが関わっている可能性が示唆された。加えて、TATA 型に着目すると、収束型では 4 割を超える遺伝子が、YR に従う配列パターンのうち CA であったが、分散型では TA や TG が多いことが分かり、YR ルールから逸脱した配列や TA や TG を有する遺伝子は TATA-box を有していても TSS が分散しやすい可能性が示唆された。また、TATA-box を持つが否かで収束や分散に関わる YR パターンが異なると考えられた。TATA-box や YR ルール以外で、TSS の収束・分散に関わると考えられる特徴として、TSS 上流の T リッチ配列が観察された。TATA-less 収束型および TATA 収束型の遺伝子の TSS 上流 10 塩基前後の領域内には T が豊富に含まれている。これは、TATA-less 型に着目するとより顕著な傾向が見えており、分散型と比較して収束型では、T の比率が TSS の上流領域で明らかに高くなっている。このように、T リッチな配列は、TSS の収束に関わり、TATA-less 型であっても TSS を収束に向かわせると考え

られる。また、他の特徴としては、TSS 周辺の全体的な C の比率が挙げられる。特に、TATA 分散型の遺伝子では、全体的に C の比率が低い傾向が認められ、TSS 周辺に C が少ない場合、TATA-box を有していても TSS が分散する可能性が示唆された。このように、プロモーターの型によって YR ルールや T リッチ、全体的な C の比率などの出現頻度が異なっており、TSS の収束・分散の決定に関与するコアプロモーターは、TATA-box の他にも複数存在することがうかがえた。

(3) TATA-box をはじめとする各コアプロモーターの重要性を検証するため、データ解析で示された重要な配列的特徴が遺伝子の発現に与える影響を *in vivo* にて評価することとした。そのためにまずは、改変を加えるプロモーターの選抜および改変する配列の決定を行った。得られたデータ解析の結果の全てを一度に検証することは難しいため、本研究では典型的なプロモーターの型である TATA 収束型と TATA-less 分散型について、改変と検証が容易な TATA-box と YR ルール、T リッチな配列についての検証を行うこととした。まず、mRNA の定性・定量を容易とするため、CAGE 解析での mRNA 蓄積量の評価結果より、mRNA100 万本あたりの蓄積量が 1000 本以上である発現量の高い遺伝子のみを選抜した。TATA 収束型プロモーターとして、主要な TSS の上流-35 から-26 の 10 塩基内に T または A を 7 塩基以上有し、TSS 上の YR ルールに従う遺伝子 (96 個) の中から TSS の収束度が高い EF1 (AT5G60390) と HOG1 (AT4G13940) を選抜した。また、TATA-less 分散型のプロモーターとして、主要な TSS の上流-35 から-26 の 10 塩基内に T または A を 5 塩基以下しか有しておらず YR ルールに従う遺伝子 (7 個) の中から収束度が低い CytOx (AT4G21105) と RPL29 (AT3G09500) を選抜した。

選抜されたプロモーターについて、コアプロモーターの重要性を解析するためには、各プロモーター配列の改変が転写効率や TSS の収束・分散に与える影響を、レポーター遺伝子を用いた一過性発現実験にて評価する必要がある。この際に、TSS の位置と分布を詳細に評価するためには、少なくとも 100 を超える TSS の位置と出現頻度を解析する必要がある。しかし、複数のプロモーターについて、通常の 5'RACE 法を用いて各プロモーターについて cDNA を 1 本ずつクローニングし、サンガー法にて TSS を同定する手法では、多大な時間と労力が必要である。また、TSS の網羅的な解析手法である CAGE などの次世代シーケンサーを用いる場合、技術の専門性が高く、流通していない試薬を使用することから、一般的に専門業者への外注にて解析することとなり、費用と結果が得られるまでの時間が問題となる。そこで本実験では、環状 RACE 法とロングリードシーケンサーを組み合わせたハイスループットかつ、安価で簡易的な RACE 法の構築を行った。

まず、TATA-box の重要性解析のために、TATA 収束型の TSS の-33 から-30 塩基の“TATA”を“CGCG”へと置換し、一過性発現実験で評価した。TSS の位置と分布を評価した結果、CAGE データの解析結果と同様に TATA 収束型プロモーターからの転写産物は、多くが 1 点に収束していた。一方で、それらのプロモーターから TATA-box を欠損させると、本来の TSS から数十塩基上流へ移動した上に、複数の TSS から分散して転写されることが明らかとなった。EF1 α プロモーターの場合、TATA-box 欠損系列では本来の TSS はほぼ使われず、主要な TSS が大きく上流へ変動した。HOG1 プロモーターで TATA-box を欠損させた場合でも、本来の TSS からの転写は確認できるもののその割合は減少し、上流からの転写が新たに確認された。また、mRNA 蓄積量およびタンパク質発現量を一過性発現実験で評価した結果、TATA-box を欠損させると、mRNA 蓄積量ならびにタンパク質発現量は、 $p < 0.05$ 水準で有意に低下していた。また、mRNA 蓄積量とタンパク質の活性値より算出した翻訳効率についても、統計的な有意ではないものの最大で 60% に低下していた。これらの結果から、TATA-box の有無が、TSS の分布や収束・分散、転写効率に大きな影響を与え、その結果、TSS の変化に伴う 5'UTR の違いにより翻訳効率にも影響し得ることが示唆された。

次に、TATA-less 分散型プロモーターを持つ遺伝子 CytOx、RPL29 を用いて、TATA-box 配列の置換による影響を解析した。各プロモーターの主要な TSS の上流 30 塩基前後の配列を“TATATA”に置換することで、TATA-box の有無が異なるコンストラクトを構築し、その影響を、一過性発現実験で評価した。TSS の位置と分布を評価した結果、未改変 TATA-less 分散型プロモーターからの転写産物は、分散して転写される傾向がみられた一方で、TATATA 置換系列では、主に 2 点に収束した。CytOx の場合、未改変系列では広い範囲に分散していたのに対し、“TATATA”を置換した CytOx-TATATA の場合、狭い範囲の分散に変化した。また、mRNA 蓄積量およびタンパク質発現量を一過性発現実験で評価した結果、TATATA へと配列を置換すると、mRNA 蓄積量ならびにタンパク質の蓄積量が向上する傾向が見られた。特にタンパク質の蓄積量に関しては $p < 0.05$ 水準で有意に増加していた。このことから CytOx に TATA-box を加えた場合、TSS は収束し、mRNA およびタンパク質蓄積量が向上する傾向であった。RPL29 では、TATATA への置換によって特に 2 番目以降の TSS の位置と出現頻度に大きな違いが見られ、TSS が大きく変動していた。また、TATATA への置換によって転写される頻度が増加した TSS は、その上流 30 塩基ほどの位置に TATA-box 様の配列を有しており、TATATA の改変がより上流の TATA-box を活性化している可能性があった。一方で、mRNA 蓄積量およびタンパク質発現量を一過性発現実験で評価した結果、mRNA 蓄積量、タンパク質発現量および翻訳活性は大きな変化はなく、若干の低下する傾向があった。これらの結果から、TATA-less 分散型プロモーターに

TATATA を創出しても、必ずしも TATA 型のプロモーターのように TSS が収束するわけではなく、転写効率に与える影響もプロモーターによって異なっていることが明らかとなった。加えて、TSS や蓄積量の変化は、TATA 型プロモーターから TATA-box を欠損させたときに比べると小さく、TATA-less 型に TATA-box の配列を加えても、TATA 型での TATA-box と同等の効果は有していないことが示唆された。

次に、TATA 収束型 EF1 α プロモーターの主要な TSS 前後の配列 (-1/0) を本来の TG から様々な配列に置換したプラスミドを用いて、一過性発現実験で評価した。YR ルールに従う別の配列である CG へと置換した場合は、TSS はほぼ変化していなかった。同様に、YR ルールに従う TA へと置換した場合は、若干の変化が確認された。続いて、YR ルールではない別の配列に置換した場合は、YR ルールに従う配列に置換した場合に比べ、TSS への影響が大きく、より分散する傾向が認められた。特に、TT や TC では、分散の傾向が顕著であった。このように、YR 部分の配列は、TSS の収束や分散に影響を与えており、収束度合いはルールに従う配列の方が大きいことが示された。また mRNA 蓄積量およびタンパク質発現量を一過性発現実験で評価した結果、mRNA 蓄積量は全体的に低下している傾向があり、YR の改変は転写効率に若干の影響を及ぼしている可能性があるが、TT に置換した場合のタンパク質の蓄積量以外では、 $p < 0.05$ 水準で有意な変化は認められなかった。翻訳の効率には 5'UTR 配列が重要であることが知られているため、TT への置換で観察された本来の TSS よりも下流側の TSS から転写された 5'UTR の翻訳効率は、他よりも悪い可能性が考えられた。この結果より、Inr の YR ルールの配列は、TSS の収束や分散に重要であり、YR ルールに従う場合では TSS が収束しやすくなる一方で、mRNA の転写効率にはあまり強い影響を与えないことが示唆された。

最後に TATA-less 分散型プロモーター CytOx を用いて、T リッチの重要性を解析した。CytOx プロモーターの主要な TSS および 2 番目に主要な TSS の -1 位から上流 10 塩基を、2 種類の T リッチ配列にそれぞれ置換したコンストラクトを構築した。その影響を、一過性発現実験で評価した。2 種の T リッチを 2 か所で置換した結果、全てのプロモーターで T リッチな配列は TSS の分布に影響を与えていることが示唆された。しかし、移動した TSS の位置は、置換系列間で様々であった。同じ配列の異なる位置への置換系列でさえも差が認められており、置換する位置、つまり周辺の塩基との関係性も重要であることが示唆された。また、mRNA 蓄積量およびタンパク質発現量を一過性発現実験で評価した結果、mRNA 蓄積量には、有意な差は認められなかった。また、タンパク質の発現量にも有意な差は認められなかった。今回の結果から、T リッチな配列は、転写効率に若干の負の影響を与えている可能性が示唆された。また、T リッチな配列への置換によって生じた TSS の変動によって 5'UTR 配列が変わり、翻訳の効率が変化している可能性がある。しかし、この結果だけでは、評価しなかった「収束型プロモーターに存在する T リッチな配列」が遺伝子の発現量に及ぼす影響を結論付けることはできない。そのため、T リッチな配列が発現量へ与える影響に関しては更なる解析が必要であると考えられる。これらの結果から、収束型プロモーターで見られた T リッチな配列は、転写効率や TSS の位置に影響を与えることもあるが、周辺の塩基との関係性が強く、その配列のみでは必ずしも TSS を収束させるわけではないと考えられた。

TATA 収束型プロモーターおよび TATA-less 分散型プロモーターに TATA-box の有無の異なる改変を加えた結果、TATA-box 欠損系列では TSS、転写効率ともに大きな変化があった一方で、TATA 配列挿入系列では、TATA-box を欠損させたプロモーターでの TSS、転写効率への変化ほど大きく影響しなかった。このことから、TATA 型と TATA-less 型のプロモーターでの転写制御の違いは、TATA-box だけが影響しているのではなく、他のコアプロモーターの影響も考えられた。TATA 収束型プロモーターにおいて、Inr 内の YR ルールに関する改変の結果、YR ルール内の改変系列では、TSS は依然として 1 点に収束したものの、転写効率については各改変系列で若干異なる結果が得られた。一方で、YR ルール外の改変は、TSS が収束から分散に転じ、転写効率がそれほど変わらないが、低下した。このことから、YR ルールは転写効率と TSS の 1 点への収束に関わることが示された。さらに、YR ルール内の改変系列において、TSS 決定に及ぼす影響の違いを比較した結果、YR 配列パターンによって収束度合いが異なった。特に「CA」および「TG」が重要なパターンであると考えられた。TATA-less 分散型プロモーターにおいて、TSS 上流を T リッチ配列に改変した結果、T リッチな配列は、TSS や転写効率に影響を及ぼすものの、想定される TSS の収束は必ずしも認められるわけではなかった。さらに、T リッチの配列パターンだけでなく、その周辺の他の配列との組み合わせによって影響が異なっていることが示唆された。本研究は、様々なコアプロモーターが mRNA の転写に与える影響を *in vivo* にて詳細に評価し、TATA-box や YR ルール、T リッチ配列の TSS や転写効率の決定における重要性を植物で示した。これらの結果は、今後の植物コアプロモーターの研究での重要な基礎となる知見を提供するものと考えられる。

5. 主な発表論文等

〔雑誌論文〕 計1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 0件）

1. 著者名 Yamasaki Shotaro, Suzuki Atsunobu, Yamano Yasuaki, Kawabe Harunori, Ueno Daishin, Demura Taku, Kato Ko	4. 巻 35
2. 論文標題 Identification of 5'-untranslated regions that function as effective translational enhancers in monocotyledonous plant cells using a novel method of genome-wide analysis	5. 発行年 2018年
3. 雑誌名 Plant Biotechnology	6. 最初と最後の頁 365 ~ 373
掲載論文のDOI（デジタルオブジェクト識別子） 10.5511/plantbiotechnology.18.0903a	査読の有無 有
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5件（うち招待講演 1件/うち国際学会 1件）

1. 発表者名 田中彩, 山崎将太郎, 川邊陽文, 出村拓, 加藤晃
2. 発表標題 転写開始点の決定に重要なコアプロモーター因子の解析
3. 学会等名 第8回植物RNA研究ネットワークシンポジウム
4. 発表年 2019年

1. 発表者名 山崎将太郎, 上野大心, 加藤晃
2. 発表標題 mRNA配列から翻訳状態を説明できる数理モデル～基礎と応用～
3. 学会等名 第1回植物インフォマティクス研究会
4. 発表年 2018年

1. 発表者名 Ko Kato
2. 発表標題 Transgene expression system in plants optimizing translation process using informatics.
3. 学会等名 The 30th Annual Meeting of the Thai Society for Biotechnology (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 川邊陽文、山崎将太郎、加藤晃
2. 発表標題 有用遺伝子高発現系の確立へ向けたコアプロモーター配列の解析
3. 学会等名 第37回日本植物細胞分子生物学会
4. 発表年 2019年

1. 発表者名 田中彩、山崎将太郎、川邊陽文、西村侑美、出村拓、加藤晃
2. 発表標題 植物のコアプロモーター配列の改変が遺伝子発現に及ぼす影響
3. 学会等名 2020年度日本生物工学会関西支部学生オンライン発表会
4. 発表年 2020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