

植物 mRNA の翻訳機構に関する研究

山崎 将太郎

奈良先端科学技術大学院大学

バイオサイエンス研究科 植物代謝制御研究室

(出村 拓 教授)

平成 28 年 2 月 16 日 提出

目次

緒論.....	4
第一章 植物の成長・発達段階における翻訳状態の解析.....	7
1-1. 序論.....	7
1-2. 材料と方法.....	9
1-2-1. 使用植物条件.....	9
1-2-2. ポリソーム解析.....	9
1-2-3. ポリソーム／マイクロアレイ解析.....	9
1-2-4. ポリソーム／定量 RT-PCR 解析によるマイクロアレイデータの検証.....	13
1-3. 結果.....	17
1-3-1. 成長・発達段階における全体的な翻訳状態.....	17
1-3-2. 発芽 2、21 日目、未展開、展開葉でのゲノムワイドな翻訳状態の解析.....	19
1-3-3. 発芽 2、21 日目、未展開、展開葉での個別 mRNA の翻訳状態.....	23
1-3-4. 成長段階間、葉の発達段階間での個別 mRNA の翻訳状態の違い.....	24
1-3-5. 成長・発達を通じた mRNA 蓄積量変化と翻訳状態変化の関連性.....	27
1-3-6. 特定のタンパク質機能をコードする mRNA 種集団の翻訳状態の傾向.....	31
1-3-7. 様々な条件下での翻訳状態及びその変化の類似性と特異性.....	36
1-4. まとめ及び考察.....	40
1-4-1. 植物の成長・発達段階での各 mRNA の翻訳状態.....	40
1-4-2. 翻訳段階の生理的な役割.....	40
1-4-3. 植物 mRNA の翻訳状態の決定機構.....	42
第二章 翻訳状態の決定に関わる 5'UTR の配列的特徴の解明.....	43
2-1. 序論.....	43
2-2. 材料と方法.....	46
2-2-1. 使用植物体および培養細胞.....	46
2-2-2. Cap Analysis of Gene Expression (CAGE).....	46
2-2-3. Partial Least Squares Regression (PLS)モデルの構築.....	48
2-2-4. DNA 一過性発現実験.....	58

2-3. 結果	63
2-3-1. CAGE によるゲノムワイドな転写開始点の解析	63
2-3-2. 転写開始点の分散と変化	67
2-3-3. 転写開始点前後の配列における塩基含量の解析	71
2-3-4. 転写開始点の分散・変化と翻訳の関連性.....	75
2-3-5. 5'UTR の配列的特徴によって翻訳状態を説明できる PLS モデルの構築 ...	77
2-3-6. 未展開葉での PLS モデルから示された重要な配列的特徴	83
2-3-7. 一過性発現実験による未展開葉に関する PLS モデルの検証	85
2-3-8. 塩基置換による PLS モデルで示された重要な配列的特徴の検証.....	91
2-3-9. 様々な条件下での PLS モデルの構築とその比較.....	97
2-4. まとめ及び考察	107
2-4-1. CAGE による 5'UTR 配列の特定	107
2-4-2. 翻訳状態の決定に関わる配列的特徴.....	108
2-4-3. TSS の収束/分散と mRNA の翻訳状態	110
総括	112
謝辞	114
研究成果の公表	115
参考文献	117

緒論

遺伝子の大部分はタンパク質として発現しており、この遺伝子がタンパク質に至るまでには、転写、転写後、翻訳、翻訳後といった複数の過程が存在している。これまでに、遺伝子の発現を理解する目的で、非常に多くの転写段階での研究が行われ報告されてきた。特に近年では、マイクロレイや次世代シーケンサーといった解析を用いて、mRNA の蓄積量をゲノムワイドに解析することが容易となり、遺伝子発現ネットワークの解析や生理的な応答などを解明するために利用されている。しかしながら、mRNA とタンパク質の蓄積量の間での相関は必ずしも高くはないことが報告されており (Maier et al. 2009)、最終的なタンパク質の量を決定する翻訳と翻訳後の段階もまた非常に重要な過程であることが示されている。このような転写段階以降の重要性が示される中で、近年では実際に生体内で働くタンパク質そのものについての関心が高まっており、大規模なプロテオーム解析等が精力的に行われるようになってきている (Yamana et al. 2013, Galland et al. 2014)。こうして得られつつあるタンパク質の発現プロファイルによって、代謝等の様々な生理的な経路の理解が強く期待されるとともに、それらの研究を通して発現プロファイルを決定している各制御過程の重要性が再確認されている。一例として、動物細胞を用いて遺伝子発現における各過程の関連性を解析した研究では、mRNA の翻訳効率の差が最終的なタンパク質の量に与える影響は非常に大きく、転写効率のそれに匹敵するということが報告されている (Schwanhäusser et al. 2011)。言い換えると、単位 DNA 当たりに転写される mRNA 量が遺伝子ごとに異なるように、単位 mRNA 当たりに翻訳されるタンパク質の量も mRNA 種ごとに大きく異なっており、転写段階だけでなく翻訳段階を研究することは、最終的なタンパク質の量としての遺伝子発現を正確に理解するためには必要不可欠であると言える。加えて翻訳段階は、生体内外での様々な状況への応答にも深く関わっていることが報告されており、例えば環境ストレス下などでは、大部分の mRNA からの翻訳が速やかに抑制される (Kawaguchi et al. 2004, Matsuura et al. 2010)。翻訳反応では多量の ATP が消費されるため、その消費を抑えているのではないかと思われる。一方で、ストレス応答に関連するタンパク質をコードする mRNA は、その翻訳がストレスにおいても維持されるなど選択的な翻訳制御が行われていることが分かっている (Matsuura et al. 2010, Ueda et al. 2012)。この様に、各 mRNA の翻訳効率は状況に応じて制御されていると考えられ、遺伝子発現全体の中で翻訳過程が持つ役割や、その機構をより詳細に解析する必要がある。しかし、植物において個別 mRNA に関する報告は複数あるものの、各 mRNA の翻訳効率をゲノムワイドに解析した研究は非常に限られているのが現状であり、翻訳過程のより深い理解への大きな障害となっている。

翻訳過程は、開始、伸長、終結の3つの段階に分けることができ、他の真核生物と同様に、植物においても翻訳の開始反応がタンパク質合成の律速と考えられてお

り(Gebauer and Hentze, 2004)、mRNA の 5'末端非翻訳領域 (5'UTR) が重要な役割を担っている。この翻訳の開始は、主に 5'UTR の 5'末端に存在している Cap 構造 (m⁷GpppN) に依存しており、Cap 構造と複数の翻訳開始因子 (eukaryotic initiation factor: eIF) との相互作用を介して行われる(Gingras et al. 1999)。この Cap 構造に依存した翻訳機構のモデルでは、まず Cap 構造を認識する eIF4E と足場タンパク質である eIF4G の複合体 (eIF4F) を介して、40S リボソームサブユニットが mRNA に結合し、その後、開始前複合体を形成する。次いで開始前複合体が mRNA の 5'UTR 上をスキャニングすることによって開始コドンである AUG を認識し、60S リボソームサブユニットと結合してタンパク質合成を開始すると考えられている。この翻訳の開始を調節する機構として、主に eIF4E 等の mRNA への結合及び開始前複合体形成の阻害・促進、eIF2 α 等のいくつかの eIF 及びそれらと相互作用するタンパク質のリン酸化状態、5'UTR 内の uORF や二次構造等による本来の ORF への開始前複合体の到達の阻害などの複数の機構が存在している (Sonenberg et al. 2009)。それら複数の機構の中でも、翻訳の開始となる mRNA へのリボソームのリクルートに中心的な役割を担う eIF4E の制御は、多くの mRNA の翻訳開始に影響を与える重要な機構であると考えられている。哺乳類や酵母においては、この eIF4E に結合し、Cap 構造との相互作用を阻害する 4E 結合タンパク質 (4E-BP) の存在が知られており、この 4E-BP を介した翻訳制御モデルは非常に研究が進んでいる制御機構の一つである (Sonenberg et al. 2009, Thoreen et al. 2012)。栄養飢餓や環境ストレス、成長因子の欠如などによって 4E-BP はリン酸化され eIF4E に結合し、Cap 構造を介したリボソームのリクルートを阻害することで大部分の mRNA からの翻訳を抑制する。しかし、植物には 4E-BP のオルソログが存在せず (Browning 2004)、他の真核生物では主要なこの分子モデルを植物細胞に当てはめることはできない。また、真核生物では Cap 構造に依存せず直接的に開始前複合体を開始コドンに導く Internal Ribosome Entry Sites (IRES) という配列構造が知られており、IRES 構造を持つ mRNA の場合、上記のような翻訳開始因子が Cap 構造を認識できないような状況でも翻訳が維持されることが知られている (Schneider et al. 2001, Komar et al. 2012)。この IRES 構造は真核生物 mRNA の 5'UTR において数多く見出されているが、植物では諸説あるがその確かな存在は報告されていない。この様に、植物においても他の真核生物同様に、mRNA 種によって翻訳効率は異なり、植物が置かれた状況によってその翻訳効率が変化することは知られているが、mRNA 種によって異なる翻訳を引き起こす分子機構は、他の真核生物と比較すると大きく異なっているのではないかと考えられる。しかし、少なくとも mRNA の翻訳効率の決定には、その 5'UTR 配列が重要であることは報告されており (Roy et al. 2013, Kawaguchi et al. 2005, Branco-Price et al. 2005, Matsuura et al. 2013)、5'UTR 配列内の翻訳効率に影響を与えている配列的特徴などを特定することによって、植物 mRNA の翻訳における分子機構の解明が大きく進むと考えられる。

本研究では、植物 mRNA の翻訳機構への理解を深めることを目的とし、各 mRNA

の翻訳効率の指標値として植物体におけるリボソームローディング（翻訳状態）を解析し、それに関わる 5'UTR の配列的特徴の詳細を明らかにした。第一章では、植物体の各成長・発達段階における mRNA の翻訳状態をポリソーム/マイクロアレイによりゲノムワイドに解析し、各 mRNA の翻訳状態の違いを評価した。また、コードするタンパク質の機能別解析より、mRNA の翻訳状態と mRNA がコードするタンパク質機能との関連性について調べることで、その生理的な意義を考察した。加えて、これまでに報告されている様々な状況における各 mRNA の翻訳状態を比較することで、そこに存在すると考えられる共通する翻訳機構と条件特異的な翻訳機構についても考察を行った。第二章では、翻訳状態の決定に関わる 5'UTR の配列的特徴に焦点を当て、まずは正確な 5'UTR 配列を Cap Analysis of Gene Expression(CAGE)を用いて網羅的に決定した。そして、第一章にて評価した翻訳状態と決定した 5'UTR 配列の 2 つのゲノムワイドデータを用いた *in silico* 解析、Partial Least Squares Regression (PLS)モデルの構築を行うことで、多くの mRNA を対象として複数要素の複合的な解析を行い、翻訳状態を決定する 5'UTR の配列的特徴を明らかにした。また、構築した PLS モデルの信頼性を一過性発現実験にて評価し確認した。加えて、様々な条件下での翻訳状態についても PLS モデルを構築し、それぞれで重要な配列的特徴をモデル間で比較することで、条件間で共通する翻訳機構と条件特異的な翻訳機構について考察した。

第一章

植物の成長・発達段階における翻訳状態の解析

1-1. 序論

近年、遺伝子の発現を理解するための研究の中で、蓄積 mRNA 量と蓄積タンパク質の量の間での相関が低いことが報告され、遺伝子発現全体を理解する上で、転写だけではなく翻訳、翻訳後の過程も考慮されるようになってきた。中でも翻訳段階がタンパク質の量に与える影響は転写段階に匹敵するほど大きく、翻訳段階もまた重要であることが示されている (Schwanhäusser et al. 2011)。植物においても翻訳段階の重要性は知られており、mRNA 種によって翻訳効率は大きく異なっていること、そして様々な状況に応じて各 mRNA 種の翻訳効率は多様な変化 (違い) を示すことがこれまでに報告されている。これらの解析の多くは、ポリソーム解析というショ糖密度勾配を用いて mRNA に結合しているリボソームの数によって mRNA を分画する手法であり、リボソームローディング(翻訳状態)を翻訳効率の指標値として用いている (Davies and Abe, 1995)。この手法とゲノムワイドに mRNA の蓄積量を解析するマイクロアレイや RNA-seq 等の手法を組み合わせることでゲノムワイドな翻訳状態の解析が可能である (Melamed and Arava 2007)。加えて、近年ではリボソームフットプリント等の新たな解析手法の登場もあり (Ingolia et al. 2009)、生体内での翻訳状態の実像が少しずつ明らかとなってきた。一例として、環境ストレスによる翻訳状態(効率)の変化が知られており、高温、塩、乾燥、明暗といった幅広い刺激におけるゲノムワイドな解析結果が報告されている (Matsuura et al. 2010, Yángüez et al. 2013, Kawaguchi et al. 2004, Juntawong and Bailey-Serres 2012, Liu et al. 2012, Liu et al. 2013)。これらの環境ストレス下では、非ストレス時と比べ翻訳状態が変化しない mRNA 種から大きく抑制されるものまで、mRNA 種ごとの翻訳状態は幅広い変化を示し、大部分の mRNA 種からの翻訳は抑制される。また、個別 mRNA を対象とした研究ではあるが、いくつかの mRNA 種の翻訳状態は、植物の発達段階や細胞分化の過程によって大きく異なっていることが報告されている。例えば、花器形成時において APETALA3、PISTILATA、SUPERMAN mRNA の翻訳は、活性化されている可能性が報告されている (Tzeng et al. 2009)。またオーキシン応答因子 (ARF) は、特定のリボソームタンパク質によって翻訳状態が制御されており、その翻訳状態が成長や発達と関連することが示唆されている (Rosado et al. 2012)。加えて、発芽時のリボソームタンパク質や (Jiménez-López et al. 2011)、花粉管発芽時の NTP303 mRNA は (Hulzink et al. 2002)、その発達段階において特異的に高い翻訳状態を示すことも報告されている。この様に、植物細胞が置かれた状況や、生理的なプロセスにตอบสนองして翻訳状態は変化しており、これらの過程において、翻訳過程

での制御は生理的に重要な意味を持つことが示唆されている。さらに、花や葉、根の組織を細かく細胞種ごとに分けて、ポリソームを形成している mRNA 量を解析した研究からは、多くの遺伝子で細胞種特異的にポリソームを形成している mRNA 量が異なることが報告されている。そして、その細胞種間でのポリソーマル mRNA 量の差異の一部は、翻訳状態の違いに起因することが示唆されている (Jiao and Meyerowitz 2010)。加えて、酵母の減数分裂時の翻訳状態を、その孢子形成過程を通して詳細に解析した研究から、一部の mRNA 種の翻訳は発達段階特異的な制御を受けていることも報告されている (Brar et al. 2012)。

この様に、翻訳過程での制御が細胞内で果たす役割は決して小さくはないことが分かる。しかしながら、環境ストレス等の特殊な条件を除いて、植物体の発達段階等の過程で、翻訳状態をゲノムワイドに解析した事例はなく、各 mRNA の翻訳状態およびそれが発達等に伴ってどのように変化するかは不明であり、植物の翻訳機構を理解する上での基盤情報が不足しているのが現状であった。そこで本章では、シロイヌナズナ植物体の各成長・発達段階に着目し、発芽 2 日目と 21 日目、発芽 21 日目の未展開葉と展開葉における全 mRNA の翻訳状態をポリソーム/マイクロアレイ法によりゲノムワイドに解析し、蓄積 mRNA に対するポリソームを形成している mRNA の比率を Polysome Ratio (PR) 値として数値化した。その結果から、mRNA 種間での翻訳状態の違いを解析するとともに、各 mRNA 種の成長段階、葉の発達段階間における翻訳状態の違い (変化) についても解析を行い、植物 mRNA の翻訳機構を理解する上での基盤情報を整備した。さらに、各 mRNA の翻訳状態の違いを、コードするタンパク質機能の観点から解析し、特定のタンパク質機能を有する mRNA 種は、特徴的な翻訳状態の偏りを示す傾向があることを明らかにし、各 mRNA の翻訳状態の決定における生理的な意義について考察した。加えて、これまでに当研究室で解析されたシロイヌナズナ培養細胞を用いた通常条件(培養 3 日目)・熱ストレス下・塩ストレス下における翻訳状態も含めた、異なる条件下での各 mRNA の翻訳状態の比較を行い、様々な状況における翻訳状態を特徴づけることで、それを制御している翻訳機構についても考察した。

1-2. 材料と方法

1-2-1. 使用植物条件

本実験には、シロイヌナズナ(*Arabidopsis thaliana* Columbia-0 (Col-0))を使用した。種子は 5% 次亜塩素酸と 0.05% Triton-X の混合液で滅菌後、GM 培地に蒔き、3 日間 4°C 暗所で低温処理を行った後、22°C、16 時間明期/8 時間暗期条件で生育した。低温処理後、生育条件に移行した日を発芽 0 日目とした。

1-2-2. ポリソーム解析

1-2-2-1. サンプルング

発芽 2 日目 (2 day after germination: 2DAG)、発芽 4 日目 (4DAG)、発芽 7 日目 (7DAG)、発芽 14 日目 (14DAG)、発芽 21 日目 (21DAG)、発芽 35 日目 (35DAG) の植物を速やかに切り取り、液体窒素中で凍結させ、-80°C にて保存した。7DAG 以降の植物体については根を取り除いている。加えて、未展開葉: young leaves として 21DAG の植物より未展開葉から 3 枚を、展開葉: mature leaves として 21DAG の植物より子葉を除いた古い葉から 3 枚を切り取り、同様に保存した。

1-2-2-2. ショ糖密度勾配遠心法を用いたポリソーム解析

ショ糖密度勾配遠心を利用したポリソーム分画は、若干の改変を加えた以外は基本的に Davies らの方法に従って行った (Davies and Abe, 1995)。各サンプルを乳棒と乳鉢を用いて液体窒素中で細かく破碎した後、破碎粉末におおよそ 2 倍量の buffer U (200 mM Tris-HCl, pH8.5, 50 mM KCl, 25 mM MgCl₂, 2 mM EGTA, 100 µg/mL heparin, 100 µg/mL cycloheximide, 2% polyoxyethylene 10-tridecyl ether, and 1% sodium deoxycholate) を加え、緩やかに懸濁した。遠心 (17,500 × g, 10 min, 4°C) により細胞残さを除き、これを粗抽出液とした。この粗抽出液を RNA 濃度 250 ng/µL に調節し、buffer B (50 mM Tris-HCl, pH8.5, 25 mM KCl, and 10 mM MgCl₂) により調整した 26.25-71.25% ショ糖密度勾配液 4.85 mL 上に 300 µL 重層し、超遠心を行った (SW55Ti rotor, 55,000 rpm, 50 min, 4°C, brake-off) (Beckman Coulter, USA)。ピストン・グラジェント・フラクショネーター (BioComp, Canada) によってショ糖密度勾配の上部より約 1 mL/min の速さで吸引すると同時に、BIO-MINI UV MONITOR AC-5200 (ATTO, Japan) を用いて 254 nm の吸光度を記録した。重層した RNA 量の違いを考慮するために、吸光プロファイルの 60S リボソーム以降のピークの合計が等しくなるようにプロファイルを補正した。

1-2-3. ポリソーム/マイクロアレイ解析

1-2-3-1. サンプルング

1-2-2-1. と同様に行い、2DAG, 21DAG, young leaves, mature leaves を得た。

1-2-3-2. ショ糖密度勾配遠心法を用いたポリソーム解析

1-2-2-2. と同様に行った。

1-2-3-3. マイクロアレイ解析用 RNA の抽出

超遠心後のショ糖密度勾配液を 8 つの画分に分画した場合の、5-8 番目の画分（底側が 8 番）を混合したポリソーム画分と 1-8 番目を混合したトータル画分から、それぞれ polysome fraction RNA、total fraction RNA を抽出した。それぞれの画分には終濃度 5.5 M になるように 8 M グアニジン塩酸塩を予め加えたチューブに回収した。この時、Two-Color RNA Spike-In Kit (Agilent Technologies, USA) に含まれる spike mix A をポリソーム画分に、spike mix B をトータル画分にそれぞれ加えた。それぞれの spike mix には、*in vitro* 合成されたポリ A 配列を持つ 10 種類の転写産物が、200 倍のダイナミックレンジでかつ既知の量比で混合されている。また、それらの転写産物に対応するスポットが本研究で使用した Agilentoligoarray (Arabidopsis 4 oligo microarray 44K; Agilent Technologies) に存在する。RNA spike-in はショ糖密度勾配遠心液を回収すると同時に加えているため、その後の RNA 精製やラベリング、ハイブリダイゼーション（後述）などの過程を経ることになる。従って、RNA spike-in に対応するスポットのシグナル値を用いた補正を行うことにより、ショ糖密度勾配における実際の RNA 比率 (polysome fraction RNA vs. total fraction RNA) を試算することが可能となる (Melamed and Arava, 2007)。ショ糖溶液及びグアニジン塩酸塩の混合液に対し等量の 100%エタノールを加え、-20°Cにて一晩冷却した後、遠心操作 (15,000 × g, 90 min, 4°C) を行った。得られたペレットを 85%エタノールにて一度洗浄した後、RNeasy kit (Qiagen, Germany) に含まれる buffer RLT にてペレットを溶解し、以降は付属のプロトコールに従い RNeasy kit を用いて RNA 精製を行った。その後、更に LiCl 沈殿による精製を行った。RNA の品質は、Agilent Bioanalyzer 2100 (Agilent Technologies) を用いたオンチップ電気泳動法により検定した。

1-2-3-4. total RNA の抽出（蓄積転写産物解析用）

1-2-3-2. で得られた粗抽出液 500 μL を終濃度 5.5 M になるように 8 M グアニジン塩酸塩を予め加えておいたチューブに回収し、以降は 1-2-3-3. と同様に行った。

1-2-3-5. マイクロアレイハイブリダイゼーション

1-2-3-3. で抽出した polysome fraction RNA 及び total fraction RNA は 2 色法、1-2-3-4. で抽出した total RNA は 1 色法でのマイクロアレイハイブリダイゼーションに用いた。RNA の増幅及び蛍光標識には、Low Input Quick Amp Labeling Kit (Agilent Technologies) を使用した。まず、500 ng の RNA を鋳型に、リンカー配列として T7 プロモーター配列を含むオリゴ dT プライマーを用いた逆転写反応を

行った。合成された cDNA を鋳型に、T7 RNA polymerase *in vitro* 転写反応により、Cy3 (polysome fraction RNA, total RNA) あるいは Cy5 (total fraction RNA) で標識された CTP を取り込んだ cRNA を合成した。合成された cRNA の精製は RNeasy kit を用いて行った。2 色法での解析では polysome fraction RNA と total fraction RNA 由来の cRNA をそれぞれ 825 ng ずつ混合し、Agilent oligoarray (Arabidopsis 4 oligo microarray 44K ; Agilent Technologies) を用いた 65°C で 17 時間のハイブリダイゼーション反応に供した。1 色法での解析では total RNA 由来の cRNA 1.65 µg を、Agilent oligoarray を用いた 65°C で 17 時間のハイブリダイゼーション反応に供した。Arabidopsis 4 oligo microarray には、シロイヌナズナ由来の転写産物や前述の RNA spike-in などの塩基配列から選択された、60 mer のオリゴ DNA が 44000 スポット プリントされている。スライドを洗浄した後、Agilent Technologies Microarray Scanner (Agilent Technologies) を用いてスキャニングを行い、Cy3 及び Cy5 のシグナルを検出した。解析はそれぞれ独立した生物学的 2 反復を用いて行った。

1-2-3-6. マイクロアレイデータ解析

スキャニング画像からデータの抽出には、Feature extraction software (Agilent Technologies) を用いて行った。Feature extraction software の設定基準に従って立てられたフラグを基に、Cy3、Cy5 いずれかについてシグナル値が飽和しているスポット (glsSaturated, rlsSaturated)、スポット内のシグナルが不均一なスポット (glsFeatNonUnifOL, rlsFeatNonUnifOL)、シグナルとバックグラウンドに優位さが無いスポット (glsPosAndSignif, rlsPosAndSignif) (glsWellAboveBG, rlsWellAboveBG) を、以降の解析から除いた。正規化には、RNA spike-in に対応するスポットを基に行う方法もしくは Feature extraction software (Agilent technologies) における標準的な正規化方法である Liner&LOWESS 法 (Locally Weighted Liner Regression) を用いた。解析対象として残ったスポットに関して、以下の計算を行った。

使用した略語の説明を以下に示す。

Polysome_Fraction_[sample_name]: 各サンプルから抽出した polysome fraction RNA 由来のマイクロアレイデータにおける正規化後の Cy3 シグナル値。

Total_Fraction_[sample_name]: 各サンプルから抽出した total fraction RNA 由来のマイクロアレイデータにおける正規化後の Cy5 シグナル値。

Total_[sample_name]: 各サンプルから抽出した total RNA 由来のマイクロアレイデータにおける正規化後の Cy3 シグナル値

翻訳状態を評価するための指標として、Polysome Ratio (polysome fraction RNA の total fraction RNA に対する割合) を各スポットについて求めた。

Polysome ratio (PR)

$PR_{2DAG} = Polysome_Fraction_{2DAG} / Total_Fraction_{2DAG}$

$PR_{21DAG} = Polysome_Fraction_{21DAG} / Total_Fraction_{21DAG}$

PR_young_leaves =

Polysome_Fraction_young_leaves / Total_Fraction_young_leaves

PR_mature_leaves =

Polysome_Fraction_mature_leaves / Total_Fraction_mature_leaves

異なるサンプル間での翻訳状態の差(変化)を評価するための指標である Δ PR 値を各スポットにおいて算出した。

Δ PR_growth = $\log_{10}(\text{PR}_{21\text{DAG}} / \text{PR}_{2\text{DAG}})$

Δ PR_leaf = $\log_{10}(\text{PR}_{\text{mature_leaves}} / \text{PR}_{\text{young_leaves}})$

異なるサンプル間での蓄積転写産物量の差を示す指標である Expression score (ES) の算出を各スポットについて行った。

Expression score (ES)

ES_growth = Total_21DAG / Total_2DAG

ES_leaf = Total_mature_leaves / Total_young_leaves

Arabidopsis 4 oligo microarray の各スポットには、gene name あるいは systematic name (e.g. AGI code [The Arabidopsis Genome Initiative gene code]) が付与されている。基本的に一つの gene name (systematic name) には一つのスポットが対応しているが、複数のスポットが対応している gene name (systematic name) もいくつか存在する。各マイクロアレイデータより上記データ処理の基準を満たし、AGI code を有するものについて、複数スポットにマップされたものは平均値を計算した後、生物学的 2 反復の値を平均し、以降の解析に用いた。上記計算はすべて Microsoft Excel を使用して行った。

1-2-3-7. 遺伝子の機能分類

遺伝子がコードするタンパク質の機能分類には、MapMan (version 3.5.1) software (Thimm et al. 2004, Usadel et al. 2005, <http://mapman.gabipd.org/web/guest/mapman>) と PageMan (version 0.12) (Usadel et al. 2006, <http://mapman.gabipd.org/web/guest/pageman>) とを用いた。MapMan と PageMan のマッピングには Ath_AGI_TAIR10 (<http://mapman.gabipd.org/web/guest/mapmanstore>)を用いた。統計解析は個々の機能集団ごとに Wilcoxon tests を行い、Benjamini and Hochberg による補正を行った。有意水準は 1%とした。この計算は MapMan のアプリケーションを使用した。

1-2-3-8. Matsuura らが評価した培養細胞における翻訳状態の補正

培養細胞を用いた熱ストレス時と塩ストレス時の各 mRNA の翻訳状態をゲノムワイドに解析した結果が報告されている(Matsuura et al. 2010)。これらの翻訳状

態は本研究で用いた PR と類似した Polysome Score (PS)として計算されている。それらの PS 値を PR 値に変換した後に、本研究で得られた PR 値との比較等に用いた。計算式を以下に示す。

PR 値と PS 値は以下のように算出される。Non_Polysome_Fraction_RNA はモノソームを形成した mRNA と遊離の mRNA が含まれる画分から得られた RNA である（本研究のポリソーム解析における 1~3 番目の画分に相当）。

$$PR = \text{Polysome_Fraction_RNA} / \text{Total_Fraction_RNA}$$

$$PS = \text{Polysome_Fraction_RNA} / \text{Non_Polysome_Fraction_RNA}$$

Total_Fraction_RNA はポリソーム及びモノソームを形成した mRNA と遊離の mRNA を含む全ての画分から得られた RNA であり、以下のように定義できる。

$$\text{Total_Fraction_RNA}$$

$$= \text{Polysome_Fraction_RNA} + \text{Non_Polysome_Fraction_RNA}$$

よって PR 値は以下のように算出することができる。

$$PR = \text{Polysome_Fraction_RNA}$$

$$/ (\text{Polysome_Fraction_RNA} + \text{Non_Polysome_Fraction_RNA})$$

$$= PS / (PS + 1)$$

算出した PR 値と PS 値の対応と、環境ストレス下における翻訳状態変化を評価するための指標である ΔPR 値の計算式を以下に示す。

PR_{22°C}: PS₂₂、培養細胞 22°C コントロール条件下での翻訳状態。

PR_{37°C}: PS₃₇、培養細胞 37°C 熱ストレス条件下での翻訳状態。

PR_{0_mM}: PS₀、培養細胞 NaCl 0 mM コントロール条件下での翻訳状態。

PR_{200_mM}: PS₂₀₀、培養細胞 NaCl 200 mM 塩ストレス条件下での翻訳状態。

$$\Delta PR_{HS} = \log_{10} (PR_{37^\circ\text{C}} / PR_{22^\circ\text{C}})$$

$$\Delta PR_{SS} = \log_{10} (PR_{200\text{mM}} / PR_{0\text{mM}})$$

1-2-4. ポリソーム/定量 RT-PCR 解析によるマイクロアレイデータの検証

1-2-4-1. サンプルリング

1-2-2-1. と同様に行い、2DAG, 21DAG, young leaves, mature leaves を得た。

1-2-4-2. ショ糖密度勾配遠心法を用いたポリソーム解析

1-2-2-2. と同様に行った。

1-2-4-3. ショ糖密度勾配液からの RNA 精製

超遠心後のショ糖密度勾配液を 8 つの画分に分画し、それぞれから RNA を抽出した。加えて 5-8 番目の画分（底側が 8 番）を混合したポリソーム画分と 1-8 番目を混合したトータル画分から、それぞれ polysome fraction RNA、total fraction RNA も抽出した。それぞれの画分を、Cap 構造とポリ A 配列を有する *in vitro* 合成した *Renilla luciferase (r-luc)* mRNA 5 ng および終濃度 5.5 M になるように 8 M グアニジン塩酸塩を予め加えたチューブに回収した。合成 *r-luc* mRNA は定量 RT-PCR 法により目的 mRNA の各画分における存在比を算出する際の補正に用いた。各チューブへ混合液と等量の 100%エタノールを加え、-20°Cにて一晩冷却した後、遠心操作（15,000 × g, 90 min, 4°C）を行った。得られたペレットは 85%エタノールにて一度洗浄した後、RNeasy kit に含まれる buffer RLT にてペレットを溶解し、以降は付属のプロトコールに従い RNeasy kit を用いて RNA 精製を行った。

1-2-4-4. 定量 RT-PCR

1-2-4-3.で精製した RNA 溶液を、それぞれ等容量ずつ用いて逆転写反応を行った。逆転写反応には Transcription First Strand cDNA Synthesis Kit (Roche Applied Science, Switzerland) を付属のプロトコールに従って用いた。反応系は 13 μL とした (oligo dT プライマー使用)。PCR 反応は 5~20 倍希釈した逆転写反応溶液 2 μL を鋳型に、遺伝子特異的プライマーセット (Table 1) 及び LightCycler 480 SYBR Green I Master (Roche Applied Science) を用いて、10 μL の反応系で行った。プライマーの設計には Universal ProbeLibrary Assay Design Center/ProbeFinder (Roche Applied Science) を、SYBR Green I の蛍光強度の経時測定には LightCycler 480 System (Roche Applied Science) を、データ解析には LightCycler Data Analysis Software (Roche Applied Science) の second derivative maximum method を用いた。各画分の RNA 回収効率、RT-PCR 反応効率の違いを補正するために、各画分における目的遺伝子の結果は、ショ糖密度勾配液の回収時に加えた補正用の *r-luc* mRNA の結果で補正した。PCR 産物が単一であることは融解曲線分析もしくはアガロースゲル電気泳動により確認した。シグナルがゲノム由来でないことは逆転写反応を行っていない RNA 溶液を鋳型にした PCR 反応において、シグナルが検出されないことにより確認した。各画分に存在する mRNA 量を定量 RT-PCR 解析により定量し、それぞれ全画分に対する割合を算出した。

Target gene	Primer sequence (5' to 3')
At1g06110 (SKIP16)	CGGAGATTCCTCCCTTGTG CCGGGATAAAAACAGACGAA
At1g06760 (Histone H1)	TGCTGCAACTAAGAGGAAAGC CCTTGGCTGGTCTAGCCTTA
At1g07320 (RPL4)	CGTGACGGACCTGAATAACA CACCACCACGAACTTCACC
At1g29930 (CAB1)	GCAAGGAACCGTGAAGTAGAA TCCGAACTTGACTCCGTTTC
At2g35795	GTTGCAGCGATCGGAGAG AACCCCCGCTATAAATGGTG
At2g36400 (GRF3)	CTTCTCTGGCTCTTCTAACTCTTCA TCTTGCTTCATCTCCGAACA
At2g37450	GTTGTGGCGATTATGTGCGAG TGACCGTAGCACCAAGAGC
At2g39500	GATTCTGGATTGCTTCCTTCA GCTTCTGCAGCCACATCAT
At2g47590 (PHR2)	GGCATTAGCCCCACTCCTA CAACGCCTCTGTCTCTCCTC
At3g11400 (eIF3G-1)	CCGAAGAGATCCTTCTCGAAC CCTTTGATTCATCCGCTTTG
At3g13580 (RPL7D)	CCGGGCTAAACAGTACTCCA TTCAGCTGGATTAATTCCCTTT
At3g18780 (Actin2)	TTCCGCTCTTTCTTTCCAAG CCATTGTCACACACGATTGG
At3g26650 (CSR1)	ATCCCTGTTGCGAGTACGTT CGACAACATCATCACAAGG
At3g47610	TGCCAAGGAATATCTCGACAA CTGAACTGGCTGCTACATGG
At3g47800	CAACGAAGGCCGTAACACTC TGGACTGACCAAATCACATCA
At3g48560 (GAPA)	ACCGAAACCCGTCTCTTCTC CACAAACACTCCGGTTCCTT
At3g60240 (eIF4G)	ACGCCCAACACAGCTAAAGT CTGGAGAGGAAATGCCTGAG
At4g00040	CAATGTTGTTTCCCAGGAGAAT TCTTGACAGTTGTGCTTTTGC

At4g24690	CCAGAGTTACCTGGTCGATACA AATTCTTCAGAGATGCATCAACAT
At4g32060	GGGAGAGAAATCCTGGTGAA TCAACATCAAAGAGCATGAAGAA
At5g44572	AATCGGCTCCTTCCTCAGTT TGCTATTATTCCCCATTCG
At5g56010 (Hsp81-3)	TGAAGGTAGCAAGATGGAGGAAGTT ACCAATGTCTCAACACCCCTAATG
At5g63570 (GSA-1)	ATGTTTTGGGGACGATGTGT TCAACTCACTACACAAGAATTATTCCA
R-luc	GGATTCTTTTCCAATGCTATTGTT AAGACCTTTTACTTTGACAAATTCAGT

Table 1. 定量 RT-PCR に用いた遺伝子特異的プライマーセット

1-3. 結果

1-3-1. 成長・発達段階における全体的な翻訳状態

植物細胞内での翻訳状態(リボソームローディング)は、mRNA 種によって異なり、環境ストレス等に応じてその翻訳状態は大きく変化することが報告されている (Matsuura et al. 2010, Yángüez et al. 2013, Matsuura et al. 2010, Kawaguchi et al. 2004, Juntawong and Bailey-Serres 2012, Liu et al. 2012, Liu et al. 2013)。さらに、酵母等での研究より、その翻訳状態は細胞の分化や発達においても変化することが知られている (Brar et al. 2012)。しかし、植物に関して各成長段階や発達段階における各 mRNA 種の翻訳状態がどのようになっているのか、ゲノムワイドに解析した例は報告されていない。そこで植物の成長・発達段階に着目し、まずは全体的な翻訳状態を知るために、シロイヌナズナ野生株を用いたポリソーム解析を行った。ポリソーム解析は、ショ糖密度勾配遠心により細胞抽出液中に存在する mRNA をリボソームの結合数に応じて分画できることから、細胞内の翻訳状態を解析する手法として広く利用されている。各成長段階での翻訳状態の解析には、発芽 2 日目 (2 day after germination: 2DAG) から 35DAG までの植物を適時サンプリングして用いた (Fig. 1A)。2~4DAG については植物体全体を、7DAG 以降については根を除いた地上部を使用している。その結果、2DAG では他と比べ高いポリソームのピークが認められ、細胞全体として非常に活発な翻訳が行われていることが示唆された。それに対して特に 21~35DAG 等では、ポリソームのピークは低く、多くの mRNA は遊離の mRNA を含むノンポリソームの画分に存在しており、2DAG 等に比べあまり活発な翻訳は行われていないことが示された。このポリソームからノンポリソームへのピークの移行は、日数経過とともに徐々に顕著になっており、植物の成長に伴い細胞全体としての翻訳状態は大きく抑制されていくことが明らかとなった。一方、21DAG の未展開葉(young leaves)と展開葉(mature leaves)のプロファイルは、21DAG 地上部全体のプロファイルに似た傾向を示し、未展開葉と展開葉の間では大きな違いは認められなかった (Fig. 1B)。

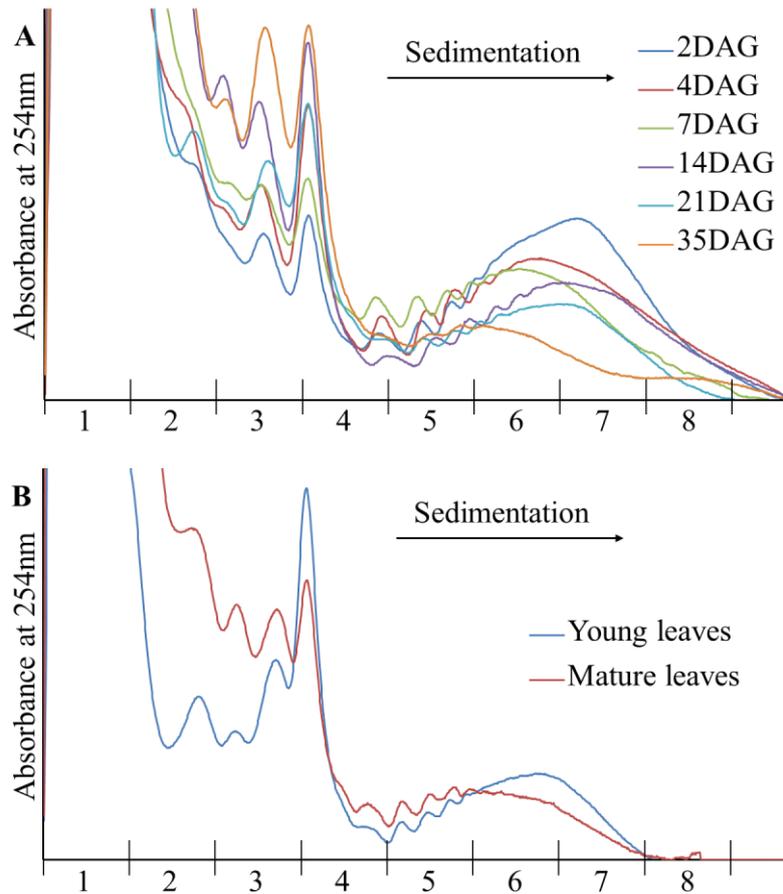


Figure 1. 各成長段階及び葉の発達段階におけるポリソーム解析

(A)2DAG から 35DAG の植物体から調製した細胞抽出液をショ糖密度勾配遠心法により分画し、254 nm の吸光を測定したプロファイルを示す。(B)21DAG の未展開葉と展開葉から測定したプロファイルを示す。沈降方向は左から右である。ポリソーム/マイクロアレイ及びポリソーム/ 定量 RT-PCR に用いた画分の番号を図中に示している。

1-3-2. 発芽 2、21 日目、未展開、展開葉でのゲノムワイドな翻訳状態の解析

1-3-1.の結果より、2DAG では全体的に活発な翻訳が行われており、21DAG 等では 2DAG と比較して、細胞全体としての翻訳状態は低いことが示された。しかし、それら全体的な挙動の中で、それぞれの mRNA 種の翻訳状態がどのように分布しているのかは不明である。加えて未展開葉と展開葉での全体的な翻訳状態に大きな違いは認められなかったが、個々の mRNA 種に着目した場合には、各 mRNA 種の翻訳状態はそれらの間で大きく異なる可能性も考えられた。そこで、個々の mRNA 種レベルでの翻訳状態を解析するために、2DAG、21DAG、21DAG の未展開葉 (young leaves)、展開葉 (mature leaves) における各 mRNA 種の翻訳状態をポリソーム/マイクロアレイ法を用いてゲノムワイドに解析した。まずショ糖密度勾配遠心により mRNA をリボソームの結合数に応じて分画し、Total fraction RNA (画分 1~8) 及び Polysome fraction RNA (画分 5~8) を回収した。その後、回収した RNA をそれぞれ Cy5 と Cy3 で蛍光標識し、cRNA を調製後、Agilent Arabidopsis 4 Oligo Microarrays 上で競合ハイブリダイゼーションさせた。それぞれのスポットから得られたシグナルの比 (Cy3: Cy5) から、個々の mRNA 種の Total fraction に対する Polysome fraction に存在する mRNA の比率 Polysome ratio (PR_2DAG、PR_21DAG、PR_young_leaves、PR_mature_leaves) を算出した。高い PR 値を示す mRNA 種は、その mRNA の多くがポリソームを形成し、活発な翻訳が行われていると考えられる。最終的に PR 値を算出することができた mRNA 数は 16917~19965 種であった。ポリソーム/マイクロアレイ解析は独立した 2 つのサンプルを用いて行ったが、対数に変換 (Log_{10}) した PR 値は、その 2 反復間で非常に高く相関していた (Fig. 2)。以降の解析では PR 値を独立した 2 つのサンプルで平均した数値を使用した。

これらのマイクロアレイデータの信頼性は、各成長・発達段階で 18 種の mRNA を用いたポリソーム/定量 RT-PCR 法によって確かめている。各条件から調整した細胞抽出液をポリソーム解析に供し、Total fraction (画分 1~8) 及び Polysome fraction (画分 5~8) に分画した後、それぞれの画分に存在する mRNA を定量 RT-PCR により定量した。それらのデータから、PR 値を算出し、ポリソーム/マイクロアレイでの結果と比較した (Fig. 3A~C)。その結果、独立した 2 つの手法間においてもすべての値は高い相関を示し ($r = 0.85 \sim 0.89$)、マイクロアレイ解析の信頼性を確認した。またいくつかの mRNA 種に関しては、より詳細にその mRNA が存在する画分を、細胞抽出液を 8 つの画分に分画して確かめている (Fig. 3D)。なお、At2g35795 の mRNA のみ主に存在する画分が他の場合と異なるが、At2g35795 mRNA の CDS 長が他に比べ短いため、mRNA の重量および結合しうるリボソームの数が少なくなるためだと考えられる。

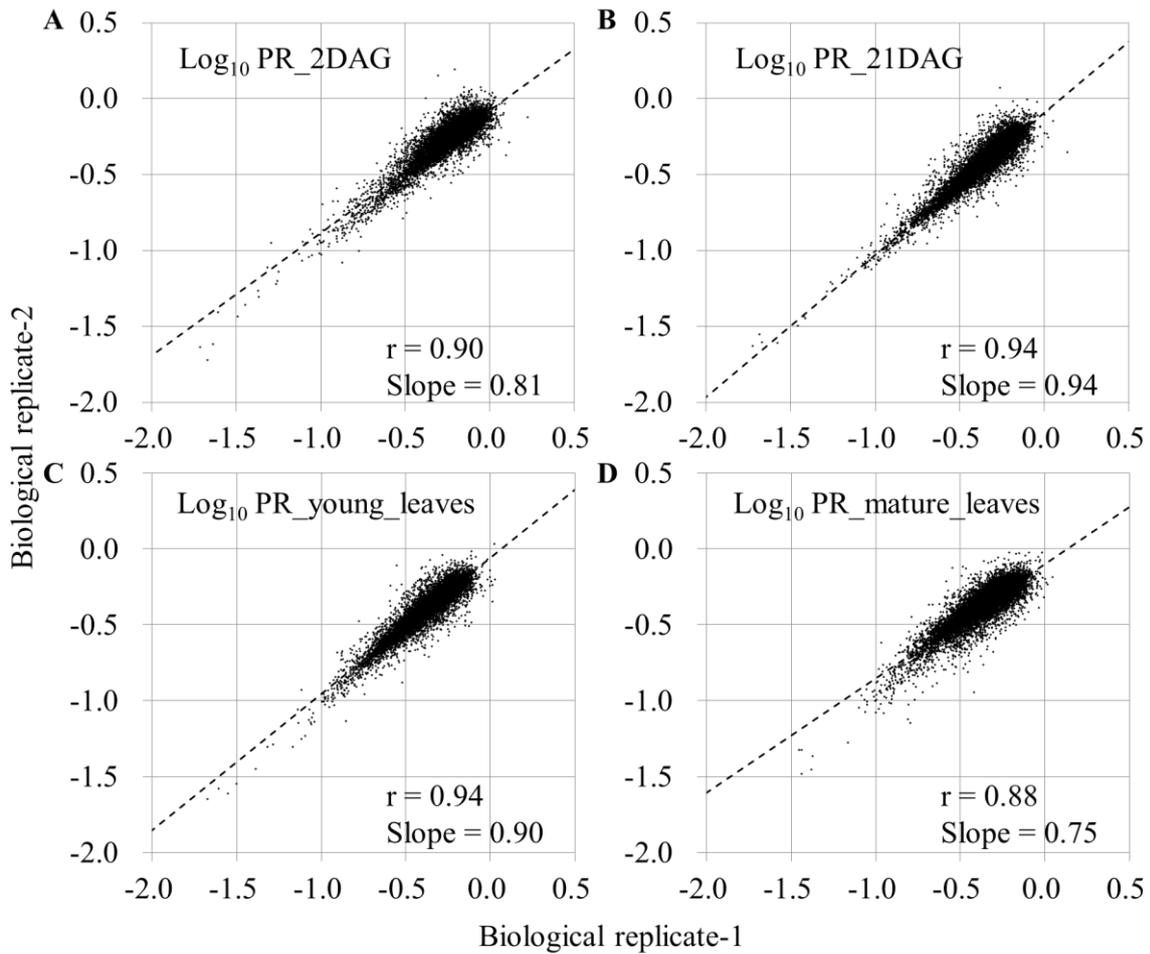


Figure 2. マイクロアレイ結果の生物学的 2 反復間での再現性

独立した生物学的 2 反復のサンプルより得られた Log_{10} に変換した各 PR 値の散布図。それぞれの散布図は(A)19965, (B)18833, (C)18122, (D)16917 の転写産物を示す。近似線 (点線) と近似式の傾き (Slope)、ピアソンの相関係数 (r) を図中に示した。

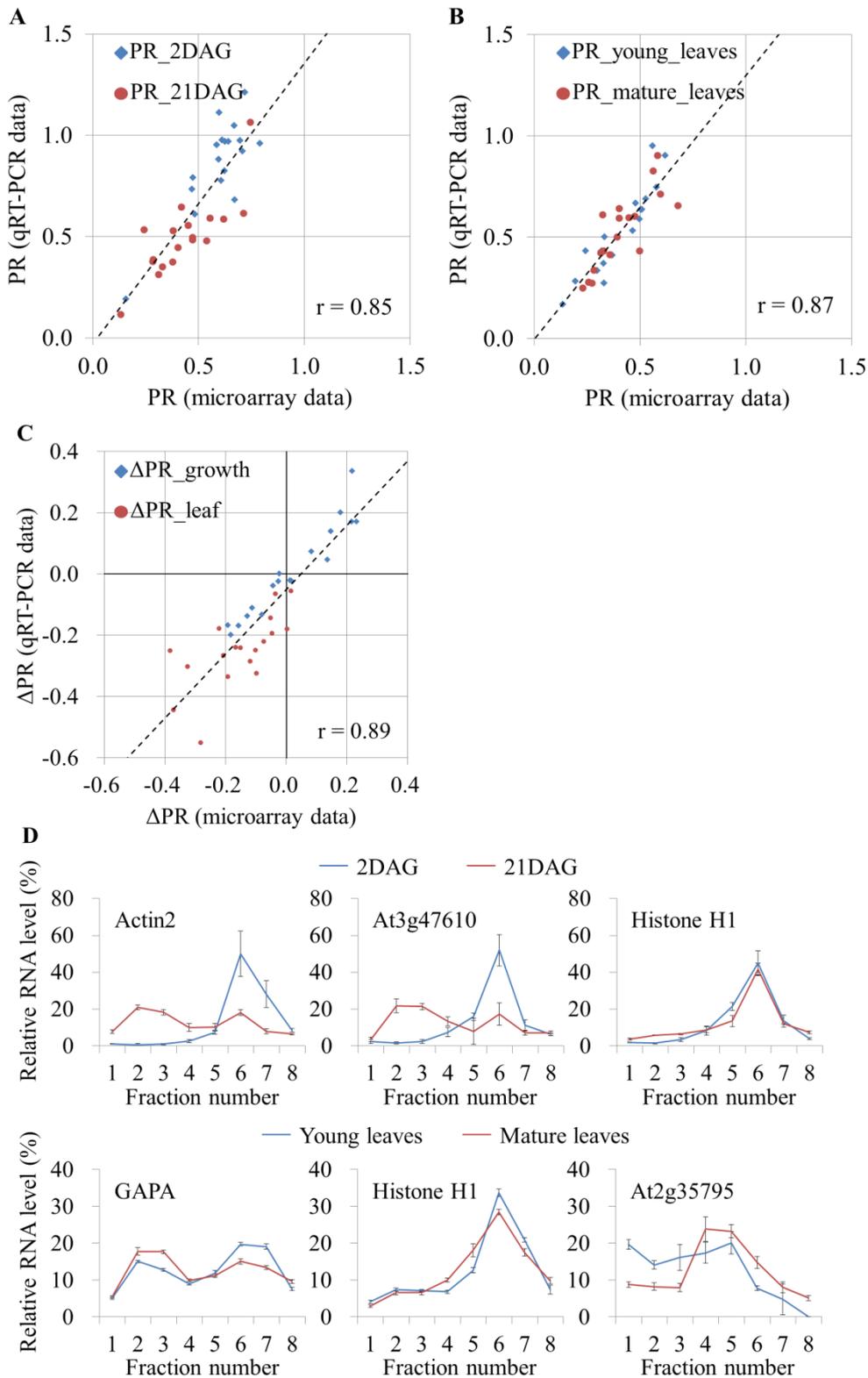


Figure 3. 定量 RT-PCR によるマイクロアレイデータの確認

(A~C) 18 遺伝子に対するポリソーム/定量 RT-PCR 解析によって、マイクロアレイデータの信頼性を確認した。ポリソーム/定量 RT-PCR 解析では、Total fraction (画分 1~8) 及び Polysome fraction (画分 5~8) のそれぞれに存在する目的の mRNA

量を定量し、各画分に等量ずつ加えた r-luc の mRNA 量で補正した。PR 値は目的 mRNA の Polysome fraction RNA と Total fraction RNA の比によって求め、 Δ PR 値(本文 1-3-4. 参照) の計算に用いた。算出した PR 値、 Δ PR 値を、マイクロアレイデータと比較した。近似線(点線)とピアソンの相関係数(r)を図中に示した。個別の結果は Table. 2 に示している。(D) いくつかの mRNA 種における 8 画分にかけて行ったポリソーム/ 定量 RT-PCR の例。縦軸は各画分に存在するそれぞれの mRNA 量を全画分に対する割合で表記している。各画分の位置は Fig. 1 に示している。

Gene name	Microarray			qRT-PCR		
	PR_2DAG	PR_21DAG	Δ PR_Growth	PR_2DAG	PR_21DAG	Δ PR_Growth
At3g13580 (RPL7D)	0.67	0.28	-0.37	0.29	1.05	-0.44
At1g29930 (CAB1)	0.59	0.24	-0.38	0.23	0.95	-0.25
At1g06110 (SKIP16)	0.61	0.29	-0.33	0.29	0.78	-0.30
At3g18780 (Actin2)	0.60	0.31	-0.28	0.31	1.11	-0.55
At3g48560 (GAPA)	0.63	0.40	-0.19	0.35	0.97	-0.34
At2g36400 (GRF3)	0.70	0.42	-0.22	0.42	0.97	-0.18
At3g26650 (CSR1)	0.61	0.38	-0.21	0.47	0.98	-0.27
At2g37450	0.48	0.33	-0.17	0.33	0.61	-0.24
At1g07320 (RPL4)	0.64	0.45	-0.15	0.45	0.97	-0.24
At4g32060	0.47	0.38	-0.10	0.35	0.79	-0.32
At3g47800	0.71	0.54	-0.12	0.54	0.92	-0.28
At3g11400 (eIF3G-1)	0.60	0.47	-0.10	0.47	0.88	-0.25
At2g39500	0.16	0.13	-0.07	0.13	0.19	-0.22
At1g06760 (Histone H1)	0.62	0.55	-0.05	0.55	0.82	-0.14
At2g47590 (PHR2)	0.79	0.71	-0.05	0.69	0.96	-0.19
At5g56010 (Hsp81-3)	0.67	0.62	-0.04	0.58	0.68	-0.07
At4g24690	0.47	0.47	0.00	0.47	0.73	-0.18
At3g60240 (eIF4G)	0.72	0.75	0.02	0.75	1.21	-0.06

Gene name	Microarray			qRT-PCR		
	PR_young leaves	PR_mature leaves	Δ PR_Leaf	PR_young leaves	PR_mature leaves	Δ PR_Leaf
At3g26650 (CSR1)	0.48	0.31	-0.18	0.67	0.42	-0.20
At3g48560 (GAPA)	0.46	0.36	-0.11	0.53	0.41	-0.11
At2g36400 (GRF3)	0.51	0.33	-0.19	0.64	0.43	-0.17
At3g13580 (RPL7D)	0.37	0.26	-0.16	0.41	0.28	-0.17
At1g07320 (RPL4)	0.53	0.39	-0.13	0.69	0.50	-0.14
At3g18780 (Actin2)	0.33	0.27	-0.08	0.37	0.27	-0.13
At3g11400 (eIF3G-1)	0.50	0.48	-0.03	0.64	0.60	-0.02
At1g06110 (SKIP16)	0.30	0.28	-0.02	0.34	0.34	0.00
At5g56010 (Hsp81-3)	0.62	0.56	-0.04	0.90	0.83	-0.04
At5g63570 (GSA-1)	0.58	0.60	0.01	0.75	0.71	-0.02
At1g06760 (Histone H1)	0.56	0.58	0.02	0.95	0.90	-0.02
At4g32060	0.33	0.40	0.08	0.50	0.59	0.07
At3g47800	0.50	0.68	0.14	0.59	0.66	0.05
At2g37450	0.32	0.45	0.15	0.43	0.60	0.14
At5g44572	0.33	0.50	0.18	0.27	0.43	0.20
At2g35795	0.24	0.40	0.22	0.43	0.64	0.17
At2g39500	0.13	0.23	0.23	0.17	0.25	0.17
At4g00040	0.20	0.32	0.22	0.28	0.61	0.34

Table 2. 定量 RT-PCR によるマイクロアレイデータの確認結果

各成長・発達段階でマイクロアレイ解析と定量 RT-PCR から得られた 18 種の mRNA の各 PR 値と Δ PR 値を示した。

1-3-3. 発芽 2、21 日目、未展開、展開葉での個別 mRNA の翻訳状態

算出した PR 値は、個々の mRNA 種のリボソームローディング（翻訳状態）を反映しており、その PR 値は幅広い分布を示し、mRNA 種によって翻訳状態は大きく異なっていた（Fig. 4）。この幅広い分布は、PR_2DAG、PR_21DAG、PR_young_leaves、PR_mature_leaves の全てで観察され（SD = 0.11~0.12）、2DAG のように細胞全体としての翻訳状態が非常に高い条件や、PR_21DAG 等のように全体的には翻訳状態が低い条件でも（Fig. 1）、個々の mRNA 種レベルで解析すると、翻訳状態が非常に高い mRNA 種から低い mRNA 種まで様々であった。この様に、全ての条件で同様に幅広い PR 値の分布が認められたが、各条件における PR 値の主要な分布の位置には違いが認められ、PR_2DAG（Median = 0.63）では、他の条件（Median = 0.46~0.48）と比べて、全体的に高い PR 値を示す傾向が認められた。このことは全体的な翻訳状態の解析結果と一致しており（Fig. 1）、植物 mRNA の翻訳状態は、mRNA 種によって大きく異なると共に、その植物の成長・発達度合いによっても大きく異なることが明らかとなった。

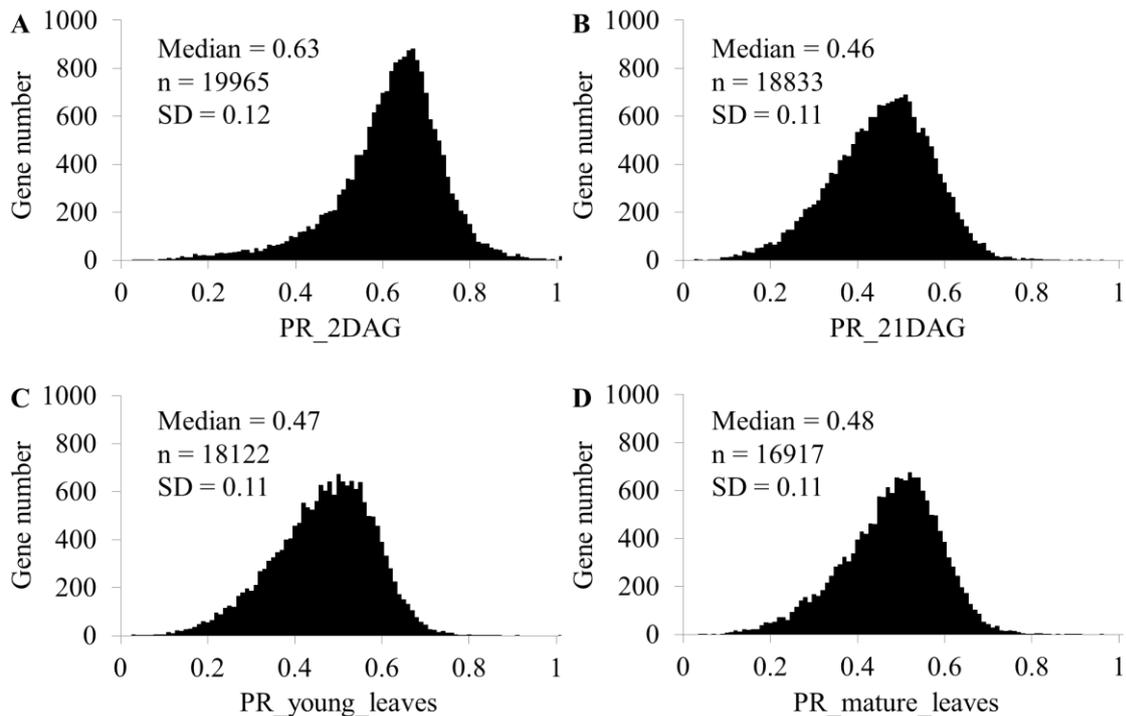


Figure 4. 翻訳状態のゲノムワイド解析より算出した PR 値の分布

(A~D) 翻訳状態の指標値である Polysome Ratio (PR) 値 (total fraction に対する polysome fraction RNA の存在比) を個々の転写産物についてマイクロアレイ解析を用いて求め、その分布をヒストグラムで示した。値は生物学的 2 反復の平均値である。図中に各条件における PR 値の中央値 (Median)、データ数 (n)、標準偏差 (SD) を示した。

1-3-4. 成長段階間、葉の発達段階間での個別 mRNA の翻訳状態の違い

これまでの結果より、全ての条件で mRNA の翻訳状態は幅広い分布を示し、その分布は、細胞全体としての翻訳状態の違いを反映し、PR_2DAG では他の条件と異なり相対的に高い PR 値を示していた。この条件間での翻訳状態の違いは、全ての mRNA 種の一様な変化なのか、それとも個々の mRNA で異なる変化を示しているのだろうか。加えて、未展開葉と展開葉では全体的な翻訳状態に大きな違いは認められなかったが、この発達段階が異なる 2 条件の間では、個々の mRNA 種レベルでも違いは認められないのだろうか。これらの疑問を解決するため、成長・発達段階間の翻訳状態の違い（変化）を ΔPR として算出し、解析を行った。解析は成長段階間 ($\Delta PR_{\text{growth}} = \log_{10}(\text{PR}_{21\text{DAG}} / \text{PR}_{2\text{DAG}})$)、葉の発達段階間 ($\Delta PR_{\text{leaf}} = \log_{10}(\text{PR}_{\text{mature leaves}} / \text{PR}_{\text{young leaves}})$) という 2 つの観点から行った。成長段階では、2 条件間の PR 値の比較を示す散布図において、PR 値が等しいことを意味する点線から大部分の mRNA は下方向に外れており、PR 値は 21DAG で減少していた (Fig. 5A)。同様の傾向は $\Delta PR_{\text{growth}}$ 値のヒストグラムからも認められ (Fig. 5C)、大部分の mRNA 種は ΔPR 値が負を示し、翻訳状態は 2DAG に比べ 21DAG で抑制されていた (Median = -0.13)。しかし、その変化は一様ではなく、翻訳状態が強く抑制されている mRNA 種から ΔPR 値が 0 付近を示し翻訳状態が変化していない mRNA 種まで幅広く存在していた (SD = 0.08)。一方、葉の発達段階間では、2 条件間の散布図において、成長段階間で観察されたような顕著な片側への偏りは認められず (Fig. 5B)、 ΔPR_{leaf} 値の分布でも多くの mRNA 種は ΔPR 値が 0 付近であり (Fig. 5D, Median = 0.01)、未展開葉と展開葉で類似した翻訳状態であった。しかし、2 条件間で翻訳状態の違いを示す mRNA 種も少なからず存在しており、その分布は ΔPR 値が負を示し、未展開葉と比較して展開葉で翻訳状態が低い（抑制された）mRNA 種から、逆に ΔPR 値が正を示し、展開葉で翻訳状態が高い（活性化された）mRNA 種まで幅広く存在していた (SD = 0.06)。

更に、詳細な葉の発達段階間の翻訳状態の違いを評価するために、翻訳状態が強く抑制された mRNA 種 (ΔPR_{leaf} 値の下位 1000 種)、強く活性化された mRNA 種 (ΔPR_{leaf} 値の上位 1000 種)、変化しなかった mRNA 種 (ΔPR_{leaf} 値が 0 付近の 1000 種) を選抜し、その PR 値の分布を解析した (Fig. 6)。Fig. 4 で示したように mRNA 種全体の PR 値の中央値は、2 点間で顕著な違いはない (Median = 0.47, 0.48)。この時、値がランダムに変化しているならば、強く抑制される（値の落ち幅が大きい）集団は、高い値から低い値への変化を示すことになり、PR_young_leaves での中央値は全体より高く、PR_mature_leaves での中央値は全体より低くなることが予想される。強く活性化される集団は、それとは逆の挙動をとり、変化しない集団は mRNA 種全体と同じ挙動をとることとなる。しかし、実際の結果を見てみると、強く抑制された mRNA 種の PR_young_leaves は、mRNA 種全体と類似した値を示し (Median = 0.46)、展開葉でのみ顕著に低い傾向であった (Median = 0.36)。一方、強く活性化された mRNA 種は、展開葉において mRNA 種全体と類似した値を示し

(Median = 0.45)、未展開葉で顕著に低い PR 値であった (Median = 0.33)。また、変化しない mRNA 種はどちらの発達段階でも mRNA 種全体よりも高い値を示した (Median = 0.51)。これらの結果は、葉の発達段階間で翻訳状態が変化しなかった mRNA 種は、翻訳状態が 2 条件でともに高い傾向がある一方で、抑制された mRNA 種は、未展開葉で平均的な翻訳状態であるが、展開葉では翻訳状態が低い傾向があり、逆に活性化された mRNA 種は、展開葉では平均的な翻訳状態であるが、未展開葉では翻訳状態が低い傾向であった。つまり、葉の発達段階を通して抑制された mRNA 種と活性化された mRNA 種は、展開葉、未展開葉それぞれで翻訳状態が全体と比較すると低い (抑制された) mRNA 種と考えられる。

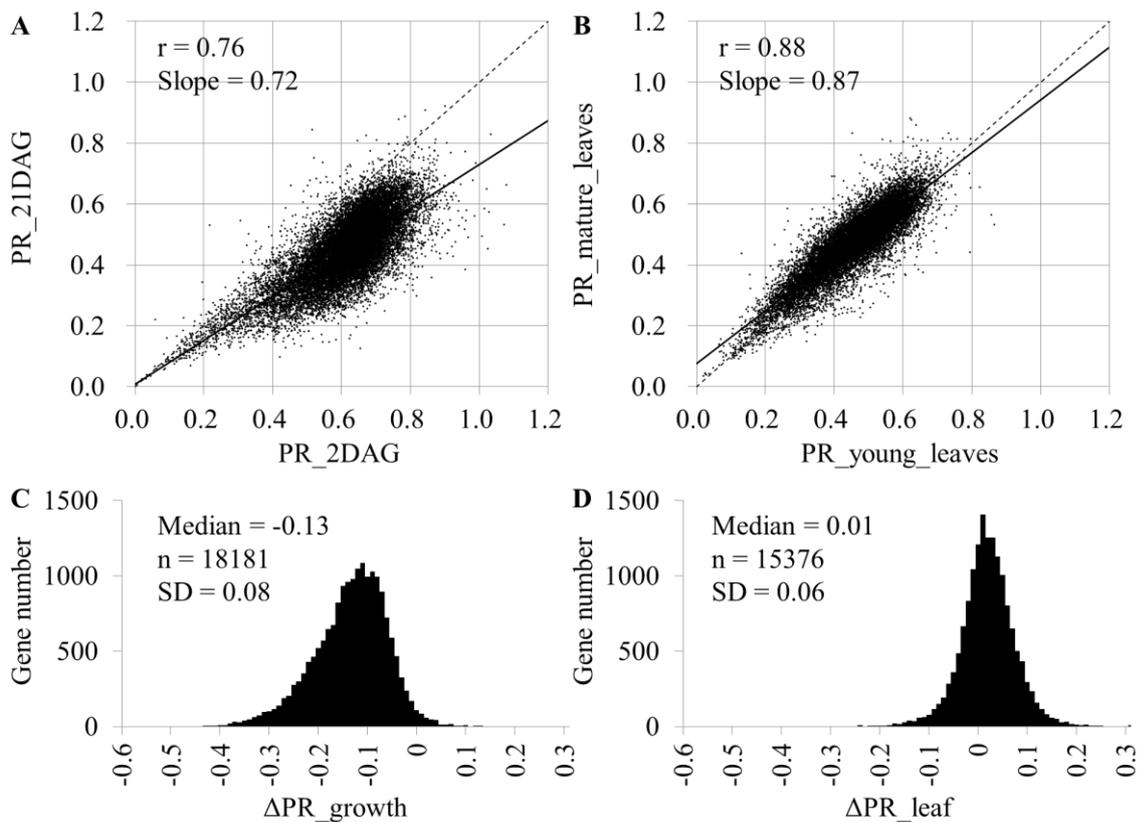


Figure 5. 成長段階、葉の発達段階間での翻訳状態の違い

(A) 成長段階間での個々の mRNA 種の PR 値の散布図。(B) 葉の発達段階間での個々の mRNA 種の PR 値の散布図。近似線 (実線) とピアソンの相関係数 (r)、近似式の傾きを図中に示す。点線は PR_21DAG と PR_2DAG、または PR_mature_leaves と PR_young_leaves の翻訳状態が等しいことを意味している。(C) 成長段階間と (D) 葉の発達段階間での翻訳状態の違い (変化) の指標値である ΔPR 値を個々の mRNA 種について求め、その分布をヒストグラムで示した。図中に各条件での PR 値の中央値 (Median)、データ数 (n)、標準偏差 (SD) を示した。

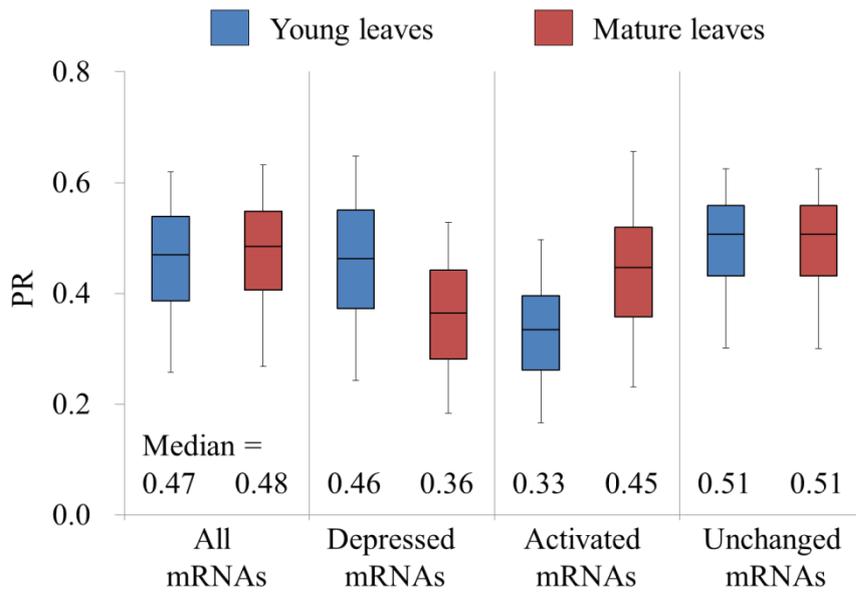


Figure 6. 葉の発達段階間での ΔPR 値別の PR 値

全 mRNA 種 (n = 15376)、葉の発達段階間の比較で展開葉での翻訳状態が抑制された mRNA 種 (n = 1000、 ΔPR 値の下位 1000 種)、活性化された mRNA 種 (n = 1000、 ΔPR 値の上位 1000 種)、変化しなかった mRNA 種 (n = 1000、 ΔPR 値が 0 付近の 1000 種) の未展開葉、展開葉での PR 値の分布を箱ひげ図で示し、各分布の中央値を図中に表記した。エラーバーの末端は分布の上位 5%位置と下位 5%位置を示し、外れ値は省略している。

1-3-5. 成長・発達を通じた mRNA 蓄積量変化と翻訳状態変化の関連性

これまでの結果によって、mRNA の翻訳状態は各成長・発達段階で大きく異なる（変化する）ことが明らかとなった。一方、遺伝子の発現を制御する非常に重要な段階である転写も、成長・発達を通して大きく変化していることが予想された。そこで、転写と翻訳の関連性を解析するために、成長段階、葉の発達段階を通じた翻訳状態変化と、蓄積 mRNA 量変化を比較することにした。2DAG、21DAG、未展開葉（young leaves）、展開葉（mature leaves）について、生物学的に独立した各 2 つのサンプルから Total RNA を抽出後、DNA マイクロアレイ解析に供した。各サンプルのシグナル比を各スポットについて求め値を平均し、蓄積 mRNA 量変化の指標である expression score (ES) 値 ($ES_{\text{growth}} = \text{Total}_{21\text{DAG}} / \text{Total}_{2\text{DAG}}$, $ES_{\text{leaf}} = \text{Total}_{\text{mature leaves}} / \text{Total}_{\text{young leaves}}$) とした。得られたシグナル値は、独立した 2 サンプル間で非常に良い再現性であった (Fig. 7, $r = 0.99 \sim 1.00$)。加えて、各条件で 18 種の mRNA について、定量 RT-PCR により Log_{10} ES 値を別に算出した結果、マイクロアレイによって得られた結果と非常によく相関した (Fig. 8, $r = 0.71$)。また、各 Log_{10} ES 値 ($\text{Log}_{10} ES_{\text{growth}}$, $\text{Log}_{10} ES_{\text{leaf}}$) の分布は、どちらも正規様の分布をしていた (Fig. 9)。今回算出した Log_{10} ES 値 ($\text{Log}_{10} ES_{\text{growth}}$, $\text{Log}_{10} ES_{\text{leaf}}$) と ΔPR 値 ($\Delta PR_{\text{growth}}$, ΔPR_{leaf}) を各 mRNA について比較したところ (Fig. 10)、それらの間には相関関係は認められず、成長段階間と葉の発達段階間の翻訳状態の制御は、全体的には蓄積 mRNA 量の制御とは独立していることが示唆された。

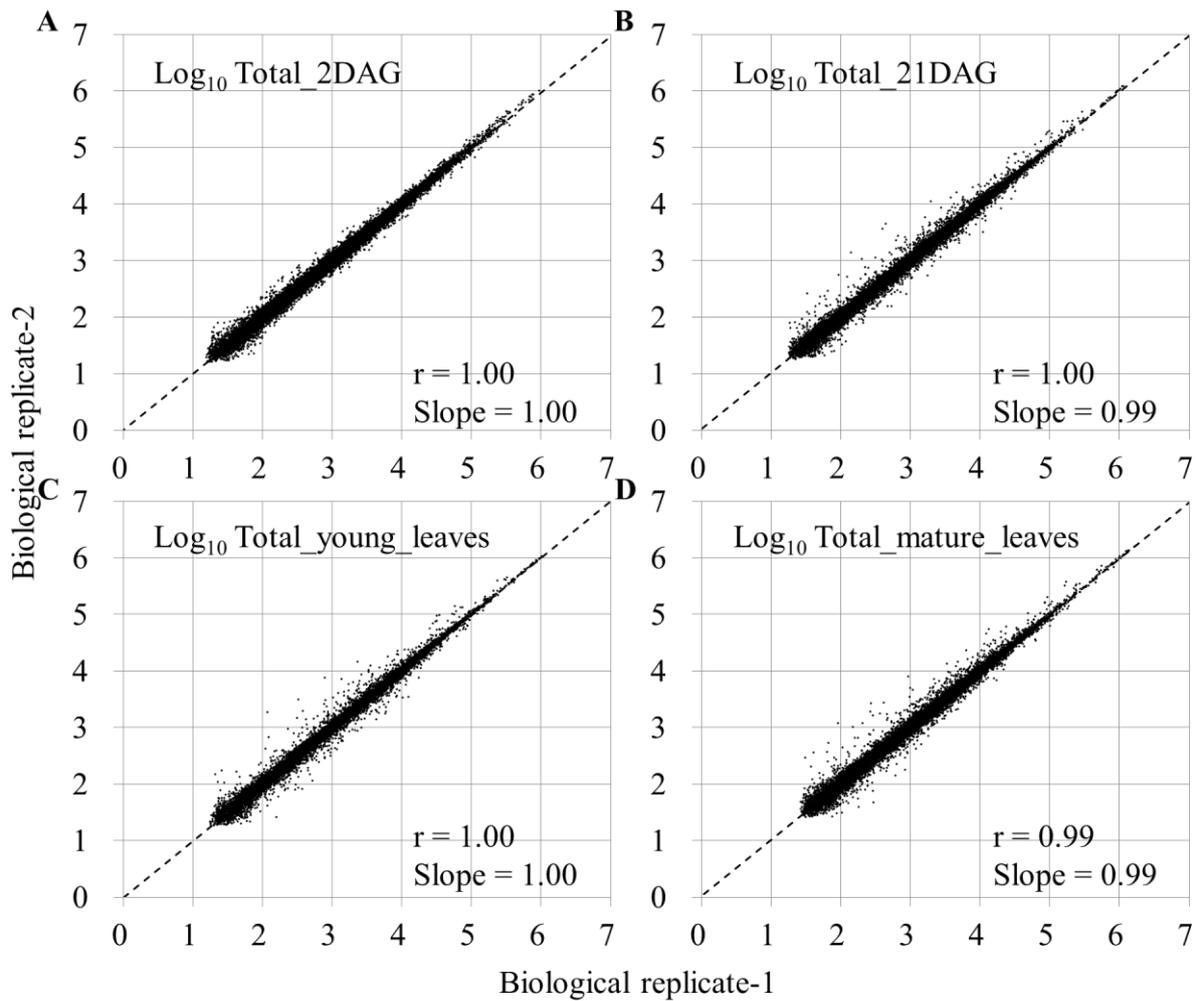


Figure 7. マイクロアレイ結果の生物学的 2 反復における再現性

独立した生物学的 2 反復のサンプルより得られた Log_{10} に変換した各シグナル強度値の散布図。それぞれの散布図は(A)20998, (B)20076, (C)19305, (D)18546 の転写産物を示す。近似線（点線）と近似式の傾き（Slope）、ピアソンの相関係数（ r ）を図中に示した。

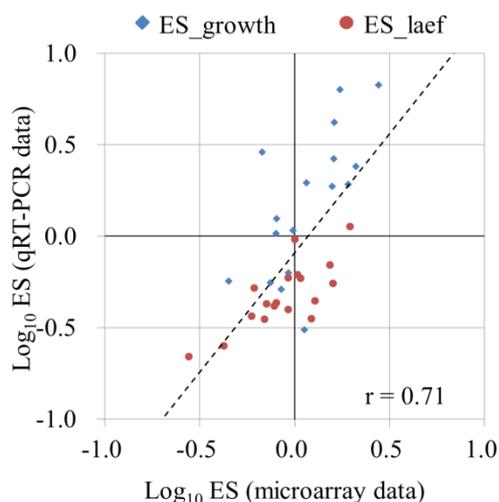


Figure 8. 定量 RT-PCR によるマイクロアレイデータの確認

18 遺伝子について Log_{10} ES 値を定量 RT-PCR によって求め、マイクロアレイデータと比較した。近似線（点線）とピアソンの相関係数（ r ）を図中に示した。個別の結果は Table 3 に示す。

Gene name	Microarray	qRT-PCR	Gene name	Microarray	qRT-PCR
	Log_{10} ES_Growth	Log_{10} ES_Growth		Log_{10} ES_Leaf	Log_{10} ES_Leaf
At3g13580 (RPL7D)	0.06	0.97	At3g26650 (CSR1)	-0.03	-0.21
At1g29930 (CAB1)	0.21	1.10	At3g48560 (GAPA)	-0.11	-0.16
At1g06110 (SKIP16)	0.32	1.05	At2g36400 (GRF3)	0.01	0.01
At3g18780 (Actin2)	-0.17	1.13	At3g13580 (RPL7D)	-0.15	-0.14
At3g48560 (GAPA)	0.24	1.47	At1g07320 (RPL4)	-0.37	-0.38
At2g36400 (GRF3)	-0.07	0.38	At3g18780 (Actin2)	0.03	0.03
At3g26650 (CSR1)	0.21	1.30	At3g11400 (eIF3G-1)	-0.23	-0.23
At2g37450	0.28	0.96	At1g06110 (SKIP16)	0.29	0.29
At1g07320 (RPL4)	-0.09	0.77	At5g56010 (Hsp81-3)	-0.21	-0.16
At4g32060	0.44	1.50	At5g63570 (GSA-1)	-0.56	-0.56
At3g47800	-0.01	0.71	At1g06760 (Histone H1)	-0.10	-0.10
At3g11400 (eIF3G-1)	-0.35	0.43	At4g32060	0.09	0.09
At2g39500	-0.13	0.42	At3g47800	0.11	0.11
At1g06760 (Histone H1)	-0.03	0.47	At2g37450	-0.16	-0.16
At2g47590 (PHR2)	-0.10	0.69	At5g44572	0.00	0.00
At5g56010 (Hsp81-3)	-0.51	-0.34	At2g35795	0.20	0.20
At4g24690	0.20	0.95	At2g39500	-0.03	-0.03
At3g60240 (eIF4G)	0.05	0.16	At4g00040	0.19	0.20

Table 3. 定量 RT-PCR によるマイクロアレイデータの確認結果

各条件でのマイクロアレイ解析と定量 RT-PCR から得られた 18 種の mRNA の各 Log_{10} ES 値を示した。

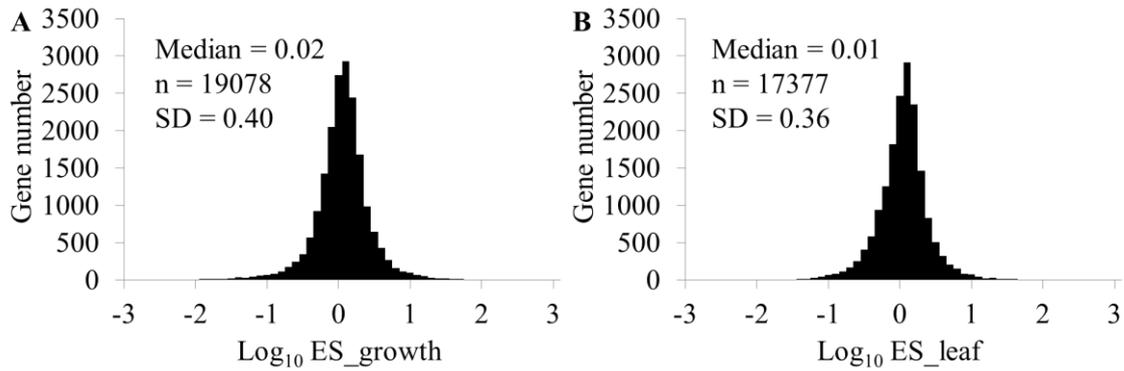


Figure 9. マイクロアレイ解析より算出した Log_{10} ES 値の分布

(A) 成長段階間、(B) 葉の発達段階間での蓄積 mRNA 変化の指標値である Expression Score (ES) 値を個々の転写産物についてマイクロアレイ解析を用いて求め、その分布をヒストグラムで示した。値は生物学的 2 反復の平均値である。図中に各条件における PR 値の中央値 (Median)、データ数 (n)、標準偏差 (SD) を示した。

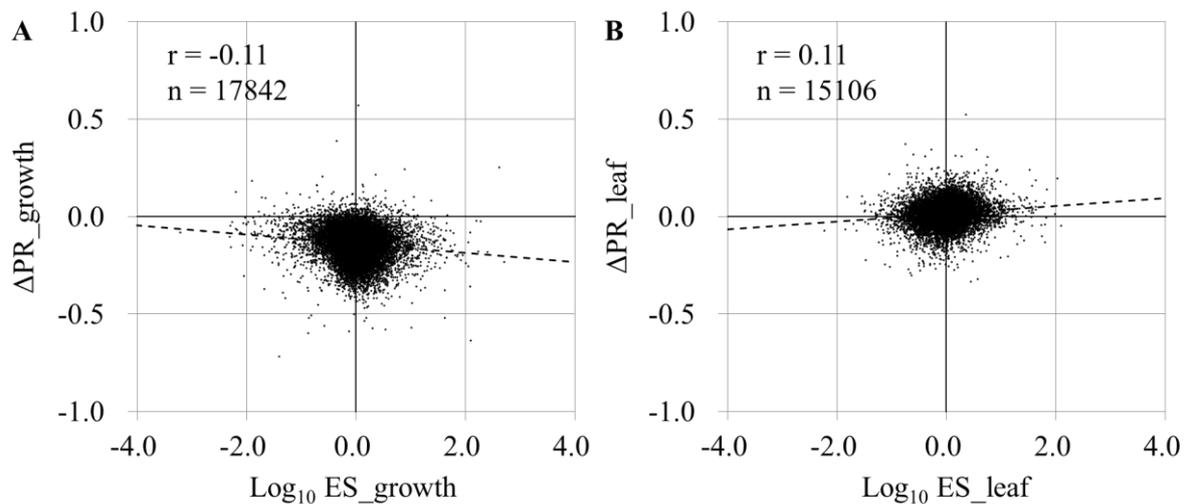


Figure 10. 成長・発達を通した蓄積 mRNA 量変化と翻訳状態変化の比較

(A) 成長段階間、(B) 葉の発達段階間での Δ PR 値と Log_{10} ES 値の散布図。ピアソンの相関係数 (r) と近似線 (実線) を図中に示す。

1-3-6. 特定のタンパク質機能をコードする mRNA 種集団の翻訳状態の傾向

翻訳状態は、mRNA 種によって大きく異なるとともに成長や発達段階を通して変化し、その制御は、転写制御とは独立していることが示唆された。このような翻訳制御の生理的な意義を考察するため、類似したタンパク質機能をコードする遺伝子集団（機能集団、全 1420 集団）に着目し、mRNA 全体の翻訳状態（PR 値）とは有意に異なる特徴的な PR 値の分布を示す機能集団が存在するのかを、解析ソフトの Map-man と Page-man (Thimm et al. 2004, Usadel et al. 2005, Usadel et al. 2006) を用いて解析した。この際、PR 値はリボソームの結合数で評価しているため、同じリボソームのリクルート効率であっても、長い mRNA の方がリボソームの結合数は多くなり PR 値は高くなる。そこで各機能集団に属する mRNA の CDS 長についても同様に評価し、PR 値が全 mRNA のそれと有意に異なる挙動をとる機能集団でも、CDS 長に有意な偏りがあり、その影響が考えられるものは除外した。解析の結果、PR_2DAG、PR_21DAG、PR_young_leaves、PR_mature_leaves のそれぞれで、上位の階層も含む 32、27、28、20 個の機能集団が mRNA 全体の PR 値の分布よりも有意に高い分布であった (Table 4)。それらの中には、TCA / org transformation、N-metabolism、amino acid metabolism、secondary metabolism、nucleotide metabolism、DNA synthesis に関する機能集団が多く含まれ、代謝に関する機能集団の翻訳状態は、高い傾向があると考えられた。加えて、PR_2DAG、PR_21DAG、PR_young_leaves、PR_mature_leaves のそれぞれで、9、8、11、11 個の機能集団が mRNA 全体の PR 値の分布よりも有意に低い分布であった (Table 4)。それらの中には hormone metabolism、RNA regulation of transcription、ubiquitin に関する機能集団が多く含まれていた。これらの機能集団の中には、各条件で類似した挙動を示すものが存在し、特に ubiquitin E3 に関連する機能集団は、全ての条件で翻訳状態は低い傾向であった。

一方、それぞれの成長・発達段階間で異なる挙動を示す機能集団も多く存在し、それぞれの成長・発達段階で特異的な翻訳状態の決定が行われている可能性が示唆された。そこで、翻訳状態の違い（変化）が生理的にどのような意義を持っているのかを考察するために、成長段階、葉の発達段階に着目し、その翻訳状態変化 (Δ PR) についても同様の解析を行った (Table 5)。その結果、 Δ PR_growth、 Δ PR_leaf のそれぞれで、20、36 個の全体の翻訳状態変化の分布より負の方向に分布が偏る、つまり成長段階や葉の発達を通して抑制されやすい機能集団を見出した。成長段階では、ribosomal protein と ubiquitin に関する機能集団が多く含まれ、葉の発達では photosynthesis、major CHO metabolism、glycolysis、cell wall、amino acid metabolism、RNA processing、ribosomal protein に関する機能集団が多く含まれていた。加えて、 Δ PR_growth、 Δ PR_leaf のそれぞれで、28、9 個の全体の翻訳状態変化の分布より正の方向に分布が偏る、つまり成長段階を通して変化しにくい、または葉の発達段階を通して活性化しやすい機能集団を見出した。成長段階では、cell wall、DNA、signaling、cell に関する機能集団が多く含まれ、葉の発達段階では ubiquitin E3 や signaling 等に関わる機能集団が含まれていた。加えて、6 個の機能集団が、成長段

階、葉の発達段階を通して共に抑制されやすい機能集団として見つかり、それらは ribosomal protein に関するものだった。これらの結果から、成長・発達に応じて特定の機能集団の翻訳状態は、より高くまたは低くなっており、遺伝子発現制御に翻訳段階が重要な意味を持っている可能性が示唆された。

BIN name (BIN code)		P-value			
		PR_2D	PR_21D	PR_YL	PR_ML
Photosynthesis	PS.lightreaction.ATP synthase (1.1.4)			3.54E-02	
	PS.lightreaction.NADH DH (1.1.6)	8.24E-03	3.88E-03	3.47E-03	8.77E-03
	PS.calvin cycle (1.3)	2.40E-02			
Major CHO metabolism	major CHO metabolism.degradation.sucrose.fructokinase (2.2.1.1)		8.14E-03	1.15E-02	4.61E-02
Fermentation	fermentation (5)	2.56E-02	1.57E-02	1.22E-02	
OPP	OPP (7)	3.79E-02			
TCA / org transformation	TCA / org transformation (8)	3.92E-06	4.72E-05	5.27E-05	3.27E-03
	TCA / org transformation.TCA (8.1)	1.57E-05	5.28E-05	3.92E-05	1.14E-03
	TCA / org transformation.TCA.pyruvate DH (8.1.1)	1.54E-02			
Cell wall	cell wall.precursor synthesis (10.1)	7.61E-03		3.79E-02	
	cell wall.pectin*esterases.PME (10.8.1)		2.59E-02		8.28E-03
N-metabolism	N-metabolism.ammonia metabolism (12.2)	4.15E-03	7.04E-03	1.22E-02	
	N-metabolism.ammonia metabolism.glutamine synthetase (12.2.2)	6.79E-03	3.53E-02	2.79E-02	
Amino acid metabolism	amino acid metabolism.synthesis.glutamate family.arginine (13.1.2.3)	2.20E-02		4.12E-02	
	amino acid metabolism.synthesis.branched chain group (13.1.4)	1.66E-02	3.05E-02	1.39E-02	
	amino acid metabolism.synthesis.serine-glycine-cysteine group (13.1.5)	1.39E-02			
	amino acid metabolism.synthesis.aromatic aa (13.1.6)	2.00E-02			
Secondary metabolism	secondary metabolism.isoprenoids (16.1)	6.94E-03			
	secondary metabolism.isoprenoids.carotenoids (16.1.4)	3.69E-02			
	secondary metabolism.phenylpropanoids (16.2)	1.67E-02	6.43E-03	1.22E-02	1.94E-02
	secondary metabolism.phenylpropanoids.lignin biosynthesis (16.2.1)		2.59E-02	1.39E-02	
	secondary metabolism.sulfur-containing.glucosinolates.synthesis (16.5.1.1)		1.40E-02	3.54E-02	
	secondary metabolism.sulfur-containing.glucosinolates.synthesis.aliphatic (16.5.1.1.1)		3.05E-02		
Hormone metabolism	hormone metabolism.ethylene.signal transduction (17.5.2)	1.84E-02	3.14E-02	2.34E-02	
Tetrapyrrole synthesis	tetrapyrrole synthesis (19)			1.39E-02	
	tetrapyrrole synthesis.magnesium chelatase (19.1)		2.73E-02	1.88E-02	
Stress	stress.abiotic (20.2)		1.57E-02	3.48E-02	5.24E-03
	stress.abiotic.heat (20.2.1)		4.86E-02		3.89E-02
Redox	redox.thioredoxin.PDIL (21.1.1)	2.19E-02	4.86E-02	4.11E-02	4.59E-02
Nucleotide metabolism	nucleotide metabolism (23)	1.81E-02			
	nucleotide metabolism.synthesis (23.1)	3.50E-04	1.23E-03	7.32E-04	
	nucleotide metabolism.synthesis.pyrimidine (23.1.1)			4.88E-02	
	nucleotide metabolism.synthesis.purine (23.1.2)	2.00E-02	8.97E-03	3.85E-03	
	nucleotide metabolism.salvage.NUDIX hydrolases (23.3.3)			4.12E-02	
Biodegradation of Xenobiotics	Biodegradation of Xenobiotics (24)				1.82E-02

Misc	misc.glutathione S transferases (26.9)				3.91E-02
	misc.peroxidases (26.12)	7.39E-03			
	misc.myrosinases-lectin-jacalin (26.16)				4.28E-02
	misc.invertase/pectin methylesterase inhibitor family protein (26.18)		4.66E-03	2.34E-02	1.28E-03
RNA	RNA.processing (27.1)				1.71E-02
	RNA.processing.splicing (27.1.1)				2.17E-03
	RNA.regulation of transcription (27.3)	1.27E-02			
	RNA.regulation of transcription.MADS box transcription factor family (27.3.24)			2.19E-02	
	RNA.regulation of transcription.MYB domain transcription factor family (27.3.25)	3.56E-03			
	RNA.regulation of transcription.GeBP like (27.3.49)				3.51E-02
	RNA.regulation of transcription.SNF7 (27.3.71)				5.24E-03
	RNA.regulation of transcription.BBR/BPC (27.3.84)		4.11E-02	4.28E-02	
DNA	DNA.synthesis/chromatin structure (28.1)		1.94E-04	2.61E-03	
	DNA.synthesis/chromatin structure.histone (28.1.3)		4.94E-03		2.58E-03
	DNA.synthesis/chromatin structure.histone.core (28.1.3.2)		3.69E-03		1.14E-03
	DNA.synthesis/chromatin structure.histone.core.H3 (28.1.3.2.3)		4.29E-02		4.59E-02
Protein	protein (29)				2.45E-04
	protein.synthesis.ribosome biogenesis (29.2.2)	3.84E-03	1.75E-02	7.18E-04	
	protein.synthesis.ribosome biogenesis.export from nucleus (29.2.2.1)	3.79E-02			
	protein.synthesis.ribosome biogenesis.Pre-rRNA processing and modifications (29.2.2.3)				1.82E-02
	protein.synthesis.initiation (29.2.3)		2.09E-02	1.98E-03	
	protein.targeting (29.3)	9.04E-03			
	protein.targeting.nucleus (29.3.1)				1.39E-02
	protein.targeting.chloroplast (29.3.3)	3.98E-02	2.85E-02	1.09E-02	
	protein.degradation.serine protease (29.5.5)	9.30E-03		4.92E-02	
	protein.degradation.ubiquitin (29.5.11)	2.76E-08	9.50E-09	1.43E-08	4.76E-08
	protein.degradation.ubiquitin.E3 (29.5.11.4)	1.47E-19	2.61E-15	1.26E-15	3.82E-12
	protein.degradation.ubiquitin.E3.RING (29.5.11.4.2)	2.88E-06	1.15E-03	6.63E-04	1.82E-02
	protein.degradation.ubiquitin.E3.SCF (29.5.11.4.3)	2.55E-15	5.40E-17	7.23E-17	2.79E-14
	protein.degradation.ubiquitin.E3.SCF.FBOX (29.5.11.4.3.2)	1.61E-18	7.05E-18	3.78E-18	2.02E-15
protein.degradation.ubiquitin.proteasom (29.5.11.20)	1.40E-02				
Signalling	signalling.in sugar and nutrient physiology (30.1)				1.82E-02
	signalling.calcium (30.3)				2.88E-03
	signalling.phosphoinositides.phosphoinositide phospholipase C (30.4.4)				2.02E-02
Transport	transport.metabolite transporters at the mitochondrial membrane (34.9)				3.89E-02
	transport.Major Intrinsic Proteins (34.19)	9.79E-05			
	transport.Major Intrinsic Proteins.PIP (34.19.1)	2.40E-02			
Not assigned	transport.misc (34.99)	3.21E-06			
	not assigned.no ontology.glycine rich proteins (35.1.40)	4.34E-03			1.36E-02
	not assigned.no ontology.proline rich family (35.1.42)	4.22E-02	2.73E-02	9.21E-03	

Table 4. 全体とは有意に異なる PR 分布を示す機能集団

PR_2DAG (PR_2D)、PR_21DAG (PR_21D)、PR_young_leaves (PR_YL)、PR_mature_leaves (PR_ML) における各機能集団 (MapMan Functional Category: BIN) の wilcoxon's p-value を示す。青色は PR 値が全 mRNA の分布より有意に高いことを、赤色は有意に低いことを意味している。表には有意な機能集団のみをリストし、有意水準は $p < 0.05$ とした。

BIN name (BIN code)		P-value	
		Δ PR growth	Δ PR leaf
Photosynthesis	PS.lightreaction.photosystem II.LHC-II (1.1.1.1)		2.33.E-02
	PS.calvin cycle (1.3)		1.68.E-02
Major CHO metabolism	major CHO metabolism (2)	1.07.E-02	1.45.E-03
	major CHO metabolism.degradation (2.2)		6.09.E-03
Minor CHO metabolism	minor CHO metabolism.callose (3.6)	1.05.E-02	
Glycolysis	glycolysis (4)		9.62.E-04
	glycolysis.cytosolic branch (4.1)		2.50.E-02
Cell wall	cell wall (10)	2.20.E-03	3.90.E-02
	cell wall.precursor synthesis (10.1)		1.10.E-02
	cell wall.degradation.mannan-xylose-arabinose-fucose (10.6.2)	9.31.E-03	
Lipid metabolism	lipid metabolism.FA desaturation (11.2)	4.80.E-02	
Amino acid metabolism	amino acid metabolism (13)		3.48.E-07
	amino acid metabolism.synthesis (13.1)		7.32.E-08
	amino acid metabolism.synthesis.aspartate family (13.1.3)		1.60.E-02
	amino acid metabolism.synthesis.branched chain group (13.1.4)		9.21.E-03
	amino acid metabolism.synthesis.branched chain group.common (13.1.4.1)		4.25.E-02
Secondary metabolism	secondary metabolism.isoprenoids (16.1)		4.13.E-02
Co-factor and vitamine metabolism	Co-factor and vitamine metabolism (18)	2.10.E-03	
Stress	stress (20)	1.92.E-02	
Redox	redox.glutaredoxins (21.4)		4.43.E-02
Nucleotide metabolism	nucleotide metabolism.synthesis (23.1)		5.55.E-03
	nucleotide metabolism.synthesis.purine (23.1.2)		3.41.E-02
	nucleotide metabolism.salvage.NUDIX hydrolases (23.3.3)		3.90.E-02
	nucleotide metabolism.phosphotransfer and pyrophosphatases (23.4)	1.54.E-02	
Misc	misc.gluco-, galacto- and mannosidases (26.3)	2.74.E-02	
RNA	RNA.processing (27.1)		1.11.E-03
	RNA.processing.splicing (27.1.1)		2.75.E-03
	RNA.processing.RNA helicase (27.1.2)		1.48.E-03
	RNA.regulation of transcription.MADS box transcription factor family (27.3.24)		1.68.E-02
	RNA.regulation of transcription.Aux/IAA family (27.3.40)	4.14.E-02	
	RNA.regulation of transcription.Chromatin Remodeling Factors (27.3.44)		2.33.E-02
	RNA.regulation of transcription.General Transcription (27.3.50)	4.33.E-02	
	RNA.RNA binding (27.4)		3.02.E-02
DNA	DNA (28)	9.85.E-09	
	DNA.synthesis/chromatin structure (28.1)	1.89.E-13	
	DNA.synthesis/chromatin structure.retrotransposon/transposase (28.1.1)	1.41.E-02	
	DNA.synthesis/chromatin structure.histone (28.1.3)	2.52.E-05	
	DNA.synthesis/chromatin structure.histone.core (28.1.3.2)	3.22.E-05	
Protein	protein (29)	2.31.E-04	4.55.E-07
	protein.aa activation (29.1)		1.45.E-03
	protein.synthesis (29.2)	6.25.E-08	3.34.E-22
	protein.synthesis.ribosomal protein (29.2.1)	5.58.E-15	2.19.E-09
	protein.synthesis.ribosomal protein.prokaryotic.unknown organellar (29.2.1.1.3)	1.76.E-02	
	protein.synthesis.ribosomal protein.prokaryotic.unknown organellar.50S subunit (29.2.1.1.3.2)	4.31.E-03	
	protein.synthesis.ribosomal protein.eukaryotic (29.2.1.2)	1.15.E-16	1.49.E-13

	protein.synthesis.ribosomal protein.eukaryotic.40S subunit (29.2.1.2.1)	9.54.E-07	2.55.E-06
	protein.synthesis.ribosomal protein.eukaryotic.60S subunit (29.2.1.2.2)	1.13.E-09	4.55.E-07
	protein.synthesis.ribosome biogenesis (29.2.2)		4.71.E-06
	protein.synthesis.ribosome biogenesis.Pre-rRNA processing and modifications (29.2.2.3)		1.42.E-03
	protein.synthesis.ribosome biogenesis.Pre-rRNA processing and modifications.DExD-box helicases (29.2.2.3.5)		3.90.E-02
	protein.synthesis.initiation (29.2.3)		1.10.E-02
	protein.synthesis.elongation (29.2.4)		2.86.E-04
	protein.targeting.nucleus (29.3.1)		4.55.E-07
	protein.degradation.metalloprotease (29.5.7)	1.49.E-02	
	protein.degradation.ubiquitin (29.5.11)	2.10.E-03	
	protein.degradation.ubiquitin.E2 (29.5.11.3)	1.18.E-05	
	protein.degradation.ubiquitin.E3 (29.5.11.4)	7.11.E-03	2.49.E-02
	protein.degradation.ubiquitin.E3.SCF (29.5.11.4.3)	1.85.E-04	
	protein.degradation.ubiquitin.E3.SCF.FBOX (29.5.11.4.3.2)	2.20.E-03	
Signalling	signalling (30)	7.88.E-14	
	signalling.receptor kinases (30.2)	0.00.E+00	
	signalling.receptor kinases.leucine rich repeat VIII (30.2.8)		2.49.E-02
	signalling.receptor kinases.leucine rich repeat VIII.VIII-2 (30.2.8.2)		2.88.E-02
	signalling.receptor kinases.leucine rich repeat X (30.2.10)	4.62.E-03	
	signalling.receptor kinases.leucine rich repeat XI (30.2.11)	7.11.E-03	
	signalling.receptor kinases.Catharanthus roseus-like RLK1 (30.2.16)	1.57.E-02	
	signalling.calcium (30.3)	4.05.E-02	
	signalling.phosphoinositides (30.4)	7.93.E-03	
	signalling.G-proteins (30.5)		2.49.E-02
	signalling.light (30.11)	1.02.E-03	
Cell	cell (31)	3.54.E-02	
	cell.organisation (31.1)	4.31.E-03	
Micro RNA	micro RNA, natural antisense etc (32)		4.55.E-07
Development	development.squamosa promoter binding like (SPL) (33.3)	4.62.E-03	
Transport	transport.metabolite transporters at the mitochondrial membrane (34.9)	4.05.E-02	
	transport.ABC transporters and multidrug resistance systems (34.16)	7.11.E-03	
Not assigned	not assigned (35)	5.05.E-07	0.00.E+00
	not assigned.no ontology.pentatricopeptide (PPR) repeat-containing protein (35.1.5)	2.52.E-05	1.45.E-03
	not assigned.no ontology.pumilio/Puf RNA-binding domain-containing protein (35.1.12)	1.54.E-02	
	not assigned.no ontology.formin homology 2 domain-containing protein (35.1.20)	1.99.E-03	
	not assigned.unknown (35.2)	2.54.E-09	0.00.E+00

Table 5. 全体とは有意に異なる ΔPR 分布を示す機能集団

ΔPR_{growth} 、 ΔPR_{leaf} における各機能集団 (MapMan Functional Category: BIN) の wilcoxon's p-value を示す。青色は ΔPR 値が全 mRNA の分布より有意に高く、成長段階では翻訳状態が変化しにくいことを、葉の発達段階では活性化していることを意味する。一方、赤色は ΔPR 値が有意に低く翻訳状態が抑制されやすいことを意味している。表には有意な機能集団のみをリストし、有意水準は $p < 0.05$ とした。

1-3-7. 様々な条件下での翻訳状態及びその変化の類似性と特異性

これまでに、各成長・発達段階での各 mRNA の翻訳状態は多様であり、特定のタンパク質機能をコードする mRNA 集団の翻訳状態は、成長・発達段階間で異なる挙動を示し、それぞれの成長・発達段階に応答した翻訳状態の決定が行われていることが示唆された。この様な mRNA の翻訳状態の決定に関して、その制御の類似性と特異性に焦点を当て、更なる考察を行うために、様々な状況における個々の mRNA の翻訳状態、および翻訳状態変化を比較することにした。これまでに当研究室では、シロイヌナズナ培養細胞を用いて、熱ストレスコントロール (PR_22°C)、熱ストレス下 (PR_37°C) の翻訳状態および熱ストレスによる翻訳状態変化 (ΔPR_{HS})、塩ストレスコントロール (PR_0_mM)、塩ストレス下 (PR_200_mM) の翻訳状態および塩ストレスによる翻訳状態変化 (ΔPR_{SS}) などの環境ストレス下での翻訳状態及びその変化をポリソーム/マイクロアレイ解析によってゲノムワイドに解析している (Matsuura et al., 2010)。これらの環境ストレス下での mRNA の翻訳状態は、コントロール条件下では全体的に翻訳状態 (PR 値) は高く、環境ストレス下では各 mRNA の PR 値は全体的には低い値を示す傾向があり、その変化は、翻訳状態が変化しないもの (ΔPR 値が 0 付近) から強く抑制されるもの (ΔPR 値が負を示す) まで正規様の分布をとり、大部分の mRNA からの翻訳が抑制されることが明らかとなっている。今回の成長・発達段階における翻訳状態の解析結果に、これらの結果も合わせて比較解析を行うことにした。全条件下で PR 値が算出できた 13447 種の mRNA についてすべての条件間で PR 値を比較し、相関係数として評価したところ、mRNA 全体としてはある程度 PR 値が類似していた ($r = 0.39 \sim 0.95$)。加えて、最も高い PR 値分布の中央値を示した PR_22°C を基準に相関が高い順番に並べると (Table 6)、各条件の中央値は徐々に低くなる傾向が認められ (Table 7)、類似した中央値を示す条件間では相関が高くなっていた。なお、Matsuura らの解析では、PR 値の算出に用いたポリソーム解析時の回収した画分の位置が今回解析した場合と異なるため、本研究における PR 値の算出値よりも高く評価されている。その中でも PR_22°C、PR_0_mM、PR_2DAG の発芽初期や培養細胞での非ストレス下などの PR 値分布の中央値が高い傾向にある条件 (高 PR 条件) の間 ($r = 0.74 \sim 0.92$) と、PR_young_leaves、PR_200_mM、PR_21DAG、PR_mature_leaves、PR_37°C の十分に成長した植物体や環境ストレス下など中央値が低い傾向にある条件 (低 PR 条件) の間 ($r = 0.64 \sim 0.95$) において、特に高い相関が認められ、高 PR 条件間や低 PR 条件間では、それぞれの翻訳状態が比較的類似していると考えられる。一方で、高 PR 条件と低 PR 条件の間で比較すると相関は低くなる傾向があり ($r = 0.39 \sim 0.76$)、それらの条件間で異なる挙動を示す mRNA が多く存在していると考えられる。

低 PR 条件間で PR 値の相関は高く、これら条件での各 mRNA の翻訳状態は類似していると考えられる。しかし、それぞれの状況に応じて特異的な翻訳状態をとる mRNA 種も存在している可能性が十分に考えられた。そこで、特定の状況に応じた翻訳制御を考察するために、成長段階、熱ストレス、塩ストレスに着目した翻訳状

態変化の比較を、全条件下で ΔPR 値が計算できた 13447 種の mRNA について行った (Table 8)。その結果、成長段階、熱ストレス、塩ストレスによる翻訳状態変化の間で、ある程度は相関することが認められたが ($r = 0.56 \sim 0.71$)、成長段階と環境 (熱と塩) ストレスの間での相関は相対的に低く ($r = 0.56 \sim 0.58$)、条件間で異なる翻訳状態変化を示す mRNA 種が多く存在することが示唆された。つまり、翻訳状態が抑制される mRNA 種は常に一定ではなく、2 条件間で相関から外れる mRNA 種は成長段階もしくは環境ストレスにおいて特異的に制御されていると考えられる。このように、翻訳状態は様々な条件の間である程度の類似性を示すが、それぞれの条件間で異なる挙動を示す mRNA も多く存在することから、そこには複数の制御機構が存在していると考えられる。

	PR_22°C	PR_0_mM	PR_2DAG	PR_YL	PR_200_mM	PR_21DAG	PR_ML	PR_37°C
PR_22°C		0.92	0.74	0.63	0.62	0.61	0.54	0.39
PR_0_mM			0.75	0.72	0.75	0.70	0.61	0.51
PR_2DAG				0.76	0.61	0.74	0.65	0.42
PR_YL					0.80	0.95	0.87	0.64
PR_200_mM						0.79	0.72	0.78
PR_21DAG							0.91	0.67
PR_ML								0.66
PR_37°C								



Table 6. 様々な条件間での PR 値の比較 (各条件間での相関係数)

熱ストレスコントロール (PR_22°C)、塩ストレスコントロール (PR_0_mM)、2DAG (PR_2DAG)、未展開葉 (PR_YL; PR_young_leaves)、塩ストレス下 (PR_200_mM)、21DAG (PR_21DAG)、展開葉 (PR_ML; PR_mature_leaves)、熱ストレス下 (PR_37°C) における翻訳状態の解析結果より、全条件下で PR 値を算出できた 13447 種の mRNA について、各条件間の PR 値を比較し、相関係数をヒートマップで示した。PR 値の分布において最も高い中央値を示す PR_22°C を基準に、その PR_22°C と相関が高い順に左から並べている。PR_22°C, PR_0_mM, PR_200_mM, PR_37°C については、Matsuura ら (2010) によって報告されたマイクロアレイデータセットを改変して使用している (<http://cibex.nig.ac.jp/data/CBX69/>)。

	PR_22°C	PR_0_mM	PR_2DAG	PR_YL	PR_200_mM	PR_21DAG	PR_ML	PR_37°C
Median *	--	--	0.630	0.470	--	0.460	0.490	--
Median **	0.910	0.880	--	--	0.780	--	--	0.600

Table 7. 様々な条件での PR 値分布の中央値

熱ストレスコントロール (PR_22°C)、塩ストレスコントロール (PR_0_mM)、2DAG (PR_2DAG)、未展開葉 (PR_YL; PR_young_leaves)、塩ストレス下 (PR_200_mM)、21DAG (PR_21DAG)、展開葉 (PR_ML; PR_mature_leaves)、熱ストレス下 (PR_37°C) における翻訳状態の解析より、全条件下で PR 値を算出できた 13447 種の mRNA における PR 値分布の中央値を示している。*には本研究におけるマイクロアレイデータ、**には Matsuura ら (2010) によって報告されたマイクロアレイデータセットを改変して使用している (<http://cibex.nig.ac.jp/data/CBX69/>)。

	ΔPR_{SS}	ΔPR_{HS}	ΔPR_{growth}
ΔPR_{SS}		0.71	0.56
ΔPR_{HS}			0.58
ΔPR_{growth}			



Table 8. 様々な条件間での ΔPR 値の比較 (各条件間での相関係数)

塩ストレス (ΔPR_{SS})、熱ストレス (ΔPR_{HS})、成長段階 (ΔPR_{growth}) による翻訳状態変化の解析より、全条件下で ΔPR 値を算出できた 13447 種の mRNA について、各条件間の ΔPR 値を比較し、相関係数をヒートマップで示している。 ΔPR 値の分布において最も高い中央値を示す ΔPR_{SS} を基準に、その ΔPR_{SS} と相関が高い順に左から並べている。 ΔPR_{SS} 、 ΔPR_{HS} については、Matsuura ら (2010) によって報告されたマイクロアレイデータセットを改変して使用している (<http://cibex.nig.ac.jp/data/CBX69/>)。

	ΔPR_{SS}	ΔPR_{HS}	ΔPR_{growth}
Median	-0.052	-0.171	-0.135

Table 9. 様々な条件間での ΔPR 値分布の中央値

塩ストレス (ΔPR_{SS})、熱ストレス (ΔPR_{HS})、成長段階 (ΔPR_{growth}) による翻訳状態変化の解析より、全条件下で ΔPR 値を算出できた 13447 種の mRNA における ΔPR 値分布の中央値。 ΔPR_{SS} 、 ΔPR_{HS} については、Matsuura ら (2010) によって報告されたマイクロアレイデータセットを改変して使用している (<http://cibex.nig.ac.jp/data/CBX69/>)。

1-4. まとめ及び考察

1-4-1. 植物の成長・発達段階での各 mRNA の翻訳状態

未だ不明な点の多い植物の翻訳機構の解明に向けて、まずは植物 mRNA の翻訳状態を知ることが目的とし、各成長・発達段階での翻訳状態を解析した。細胞全体としての翻訳状態（リボソームローディング）の解析によって、2DAG 等の発芽初期では mRNA 全体としては非常に活発な翻訳が行われていること、そして 21DAG 等の成長した植物体や、21DAG の未展開葉、展開葉では mRNA の翻訳状態は全体的には悪くなっていた (Fig. 1)。続いて、個々の mRNA 種レベルでの翻訳状態をゲノムワイドに解析し、蓄積 mRNA に対するポリソームを形成している mRNA の比率を Polysome Ratio (PR) 値として数値化した。その結果、mRNA の翻訳状態は mRNA 種によって大きく異なっており、PR 値が高く活発な翻訳が行われていると考えられる mRNA 種から、PR 値が低くあまり翻訳が行われていない mRNA 種まで幅広い分布をとっていた (Fig. 4)。更に、成長段階、葉の発達段階に着目し、翻訳状態の違い（変化）を Δ PR 値として算出し解析した結果、成長段階では翻訳状態が強く抑制される mRNA 種から変化しない mRNA 種まで正規様の分布をとり、大部分の mRNA 種は 21DAG で翻訳状態が抑制されていた (Fig. 5C)。このような翻訳状態変化は、シロイヌナズナ培養細胞を短期的な環境ストレスにさらした場合でも観察されている (Matsuura et al. 2010)。一方、葉の発達段階間の翻訳状態変化では、mRNA の翻訳状態は、抑制されるものから活性化されるものまで正規様に分布していた (Fig. 5D)。更に詳細に解析した結果、それらの抑制、または活性化された mRNA 種の違いは、未展開葉と展開葉のどちらでより抑制されているかであり、活性化は未展開葉での抑制からの回復である可能性が示唆された (Fig. 6)。これらの結果から、細胞内の個々の mRNA の翻訳状態は決して一様ではなく、更にそれぞれの成長・発達段階で変化しており、植物 mRNA は条件によって多様な翻訳状態変化が起きる複雑な機構によって制御されているのではないだろうか。

1-4-2. 翻訳段階の生理的な役割

翻訳状態のゲノムワイド解析によって、成長・発達段階での各 mRNA 種の翻訳状態は非常に多様であることが示され (Fig. 4, Fig. 5)、かつ翻訳状態と蓄積 RNA 量の比較より、翻訳段階は転写段階とは独立した制御であることが示唆された (Fig. 10)。このような翻訳制御の生理的な役割の一端を解明するため、mRNA がコードするタンパク質機能の集団（機能集団）別に翻訳状態及びその変化の傾向を解析した。その結果、特定の機能集団に属する mRNA の翻訳状態が全 mRNA と比較して有意に高い、または低いことを見出した (Table 4)。特に代謝関連の機能集団に属する mRNA 種は各成長・発達段階において翻訳状態が高い傾向が認められた。一次代謝は細胞機能を維持していくために必要不可欠なものであるため、その代謝に関わるタンパ

ク質をコードしている mRNA 種の翻訳状態が高い可能性が考えられた。一方、ユビキチン（ユビキチン E3）に関連する機能集団に属する mRNA は、各成長・発達段階において翻訳状態が低い傾向が認められた。ユビキチンは主にタンパク質の分解に関わり、特に E3 は基質である特定のタンパク質を認識し分解に導く重要な役割を持つ。しかし、タンパク質の合成には非常に多くのエネルギーを消費するため、環境に適した厳密なエネルギーの管理が行われていると考えられている（Proud 2007, Branco-Price et al. 2008）。不必要なタンパク質を合成・分解するプロセスを繰り返すことは多くのエネルギーを消費するため、通常の状態においては、ユビキチン（特に E3）に関連する機能集団の翻訳状態は低い可能性が考えられた。一方で、ユビキチンはストレス応答等に重要な役割を持っていることが知られており（Sahi et al. 2006, Huang et al. 2008）、ストレス環境下において特定の転写因子を含めた不要なタンパク質を積極的に分解することに関係している。実際に、熱や塩といった環境ストレス下でのユビキチンに関連する機能集団の mRNA の翻訳は、他の多くの mRNA からの翻訳が抑制される中、維持（または若干の活性化）されることが報告されている（Matsuura et al. 2010）。

また、成長段階、葉の発達段階での翻訳状態の違い（変化）に着目した場合、mRNA 全体の翻訳状態変化とは異なる挙動を示す機能集団が存在していた（Table 5）。シグナル伝達に関わる機能集団に属する mRNA は、成長段階で翻訳状態が変化しにくい傾向であった。これは、環境ストレス等の状況の変化に迅速に応答し細胞内ホメオスタシスを維持するために、シグナル伝達経路に関わるタンパク質は非常に重要であり（Baena-González et al. 2007, Baena-González et al. 2008）、それらが常に翻訳されているためだと考えられる。また、リボソームタンパク質に関連する mRNA の翻訳状態の抑制が成長段階、葉の発達段階の両方で認められた。この様なリボソームタンパク質に関連する mRNA の翻訳抑制は、乾燥、低酸素、ショ糖飢餓、熱ストレス、塩ストレス等の応答でも報告されている（Kawaguchi et al. 2004, Branco-Price et al. 2005, Nicolai et al. 2006, Matsuura et al. 2010）。加えて、リボソームタンパク質は 2DAG や未展開葉のような器官が急速に成長する時に必要とされ、発芽初期や成長因子が添加された条件下では、非常に活発な翻訳が行われていることも報告されている（Majeran et al. 2010, Beltrán-Peña et al. 1995）。これは上述したように、タンパク質合成は最もエネルギーを消費するプロセスの一つであることから、タンパク質合成が活発な成長・発達中（2DAG や未展開葉）に比べ、成長・発達後（21DAG や展開葉）などの新規タンパク質合成があまり必要ではない状況下でのエネルギー管理に合理的と思われる。また、葉の発達段階において、光合成や糖分解に関わる機能集団の翻訳状態が抑制されることも同様に解釈できる。一方、ユビキチン E3 に関連する mRNA の翻訳状態は、葉の発達において活性化する傾向があった（未展開葉で低い翻訳状態であったものが展開葉で比較的高くなる）。タンパク質分解は器官サイズに影響を与えることが報告されており、葉の細胞内のセルサイクル関連タンパク質の分解を通して核内倍加を制御し器官サイズの決定に関わっている（Sonoda et al.

2009, Nguyen et al. 2013)。ユビキチン E3 に関わるいくつかの mRNA 種の翻訳状態が展開葉で比較的に高いということは、特定のタンパク質の分解を促進することで葉の発達の制御に関わっているのかもしれない。この様に、特定の生理的機能に関わる mRNA 種の翻訳状態は、細胞を適切に維持し、それぞれの状況に応答するために、翻訳段階で制御されているのではないだろうか。

1-4-3. 植物 mRNA の翻訳状態の決定機構

これまでに植物の各成長・発達段階での翻訳状態は複雑に制御され、その制御は生理的に重要であることを考察した。さらに植物 mRNA の翻訳機構への理解を深めるために、植物体の各成長・発達段階、環境ストレス下などの様々な条件における個々の mRNA の翻訳状態を比較した。その結果、それら様々な条件間での全ての比較から、翻訳状態はある程度相関しており、様々な条件での翻訳状態は全体的にはある程度類似していた (Table 6)。これは、様々な条件間で mRNA の翻訳状態を決定する共通の翻訳機構が存在することを示唆していると考えられた。一方で、発芽初期や培養細胞の通常条件などの PR 値分布の中央値が高い傾向にある条件 (高 PR 条件) と、十分に成長した植物体や環境ストレス下など中央値が低い傾向にある条件 (低 PR 条件) の間では、相関が低くなる傾向が認められ、これら 2 条件グループ間で翻訳機構が異なる部分があることが予想された。また、同じ低 PR 条件グループであっても相互の相関係数には差があり (Table 6)、翻訳状態変化 (ΔPR) という観点からも差が認められることから (Table 8)、各条件で特異的な翻訳機構が存在することも示唆される。例えば、未展開葉も展開葉も共に低 PR 条件であるが、 ΔPR_{leaf} が負の値を示す mRNA 種は、展開葉で翻訳状態が悪く、 ΔPR_{leaf} が正の値を示す mRNA 種は未展開葉で翻訳状態が低い。それらの mRNA 種は、特定の条件でのみそれぞれ特異的に翻訳が抑制されていることから、そこには複数の機構が存在しているのではないだろうか。

この様な共通する翻訳機構と特異的な翻訳機構という 2 つの機構によって、植物 mRNA の翻訳状態は決定されていると考察できる。翻訳状態の決定に関して、その分子機構はほとんど分かっていないが、mRNA の 5'非翻訳領域 (5'UTR) が重要であることが知られている (Roy et al. 2013, Kawaguchi et al. 2005, Branco-Price et al. 2005, Matsuura et al. 2013)。5'UTR 上には共通する翻訳機構に関わる配列的特徴と、おそらく複数種類の特異的な翻訳機構に関わるそれぞれの配列的特徴が存在し、各条件での翻訳状態の決定に関わっている可能性が考えられた。

第二章

翻訳状態の決定に関わる 5'UTR の配列的特徴の解明

2-1. 序論

第一章では、植物の成長・発達段階での各 mRNA の翻訳状態をゲノムワイドに解析し、翻訳状態は mRNA 種によって大きく違うことを示すとともに、各段階に応答した翻訳状態の決定が行われ、生理的に重要な意味を持っている可能性を示した。加えて、翻訳状態を各条件で比較することで、各 mRNA の翻訳状態を決定している機構を考察し、様々な条件における共通の翻訳機構と、特定の条件に応答した特異的な翻訳機構の存在を示唆した。しかし、このような翻訳機構の分子的な知見は、ほぼないのが現状である。そこで第二章では、各 mRNA の翻訳状態を決定している分子機構に着目した解析を行うことで、植物 mRNA の翻訳機構への理解を深めることを目的とした。

mRNA の翻訳効率を規定する非常に重要な要素として mRNA の 5'非翻訳領域 (5'UTR) が挙げられる。異なる 5'UTR 配列を連結したレポーター遺伝子の発現実験では、5'UTR の違いに起因する翻訳効率の差によってレポータータンパク質の発現量が大きく異なることが報告されている (Sato et al. 2004, Matsui et al. 2012)。更に翻訳状態が低い mRNA 由来の 5'UTR を連結した場合の翻訳効率は低く、翻訳状態が高い mRNA 由来の 5'UTR を連結した場合の翻訳効率は高いことも報告されており (Ueda et al. 2014)、mRNA の翻訳効率の決定において 5'UTR は中心的な役割を担っていると考えられている。加えて、近年の次世代シーケンサー技術の進歩によって、これまでは難しかった mRNA の末端の配列をゲノムワイドに決定することが可能になってきており、個々の mRNA の正確な 5'UTR 配列が明らかになることによって、5'UTR 配列と mRNA の翻訳状態との関連性がより詳細に解析されてきている。例えば、mRNA の転写開始点は必ずしも一つではなく、複数の転写開始点に由来する 5'UTR バリエーションを持つ mRNA 種が非常に多いことも報告されており (Yamamoto et al. 2009, Morton et al. 2014)、酵母による研究では、そのような 5'UTR バリエーション間で、それぞれの mRNA の翻訳効率に大きな違いが生じていることも明らかにされ (Rojas-Duran and Gilbert 2012, Arribere and Gilbert 2013)、5'UTR 配列の重要性が改めて示されるとともに、翻訳機構の解明において、その条件で存在する 5'UTR 配列を正確に決定することの重要性が認識されている。一方で、5'UTR のどのような配列的特徴が翻訳効率(状態)に影響を与えているのかの詳細については、植物体においてゲノムワイドに翻訳状態を解析した事例が十分ではなく、データベース等での 5'UTR の配列情報には限界があるため、未だ不明な点が多いのが現状である。しかしそうした現状の中でも、いくつかの配列的特徴が翻訳状態に影響を与

えていることが報告されている。例えば、upstream ORF (uORF) は、本来の ORF からの翻訳を妨げることが知られ、一部の遺伝子では特定の条件でのみ uORF が働いていることが報告されている (Wiese et al. 2004, Pajeroska-Mukhtar et al. 2012)。また、熱や塩、乾燥ストレス時の翻訳状態は 5'UTR 内の GC 比率と負に相関することが報告されている (Yángüez et al. 2013, Kawaguchi et al. 2005)。加えて、熱ストレス時の翻訳状態の決定には 5'UTR の 5'末端に存在している重要領域の塩基比率が関係していることも明らかとなっている (Matsuura et al. 2013)。また、5'UTR 内で強固な二次構造を形成するほど (自由エネルギーが負になるほど) 翻訳状態が悪くなる可能性が示唆されている (Kawaguchi et al. 2005)。しかし、5'UTR 上の二次構造の形成度合いと翻訳状態に相関が認められない例も存在し (Matsuura et al. 2013)、明確な結論には至っていない。同様に 5'UTR の長さ (Kawaguchi et al. 2005) や rRNA との相補な配列の長さ (Vanderhaeghen et al. 2006, Akbergenov et al. 2004) も影響を与える可能性が示されているが、5'UTR の長さがあまり影響しない事例や (Liu et al. 2012)、rRNA と相補な長さは動物では影響の正負が異なっていたり (Tranque et al. 1998) と不明確である。他にも、5'UTR 上の特定の塩基パターン (Motif) との関連性もいくつか報告されている。例えば Tract of oligo pyrimidine (TOP) 配列とも呼ばれる CU リッチな Motif を 5'UTR に持つ mRNA の翻訳は、特定の発達段階や成長因子等の存在下では活性化され、環境ストレス条件下では抑制される可能性が示唆されている (Tzeng et al. 2009, Jiménez-López et al. 2011, Miloslavski et al. 2014)。この様に翻訳状態に関わる配列的特徴についてはいくつか報告されているが、それらの解析は翻訳状態と単一の特徴との間の相関解析に留まり、GC 比率や二次構造の形成度合いといった類似した特徴が別視点から報告されているのが現状である。そのため、真に重要な特徴や各特徴の影響の大小、そしてそれらの特徴が様々な条件下で共通して翻訳状態に影響するのか、特定の条件下での特異的な翻訳状態に影響しているのみなのかといった総合的な知見はない。植物において、複数の条件下での mRNA の翻訳状態を解析するとともに、これらの条件下での個々の mRNA の正確な 5'UTR をゲノムワイドに決定し、それら情報に基づいて 5'UTR 内の様々な特徴を複合的に捉えて解析することができれば、植物翻訳機構の分子機構の解明が大きく進歩すると考えられる。

本章では、第一章でのゲノムワイド解析の結果を踏まえ、各 mRNA の翻訳状態の決定に関わる特徴について、5'UTR の配列的観点から解析を行った。まずはゲノムワイドな転写開始点の同定と各転写開始点レベルでの mRNA 蓄積量が解析可能な CAGE を、発芽 2 日目、発芽 21 日目の未展開葉・展開葉、培養細胞における通常条件・熱ストレス下・塩ストレス下について行い、それぞれの条件について個々の mRNA の正確な 5'UTR の配列を決定した。次に翻訳状態(PR 値)と 5'UTR 配列の 2 つのゲノムワイドデータを用いた Partial Least Squares Regression (PLS) モデルの構築を行い、翻訳状態の決定機構に関わる具体的な 5'UTR 上の配列的特徴の解明を行った。目的変数には、未展開葉の PR 値を、説明変数には AUGC 含量、特定の配列

パターン(Motif)、uORF、mRNA の二次構造、rRNA との相補配列長などに関連する様々な特徴を用いた。このような複数の配列的特徴を用いた、*in silico* による複合的な解析により、これまで不明瞭だった真に重要な配列的特徴を推察できるモデル式を構築した。構築したモデル式の信頼性については、塩基置換を導入した 5'UTR を用いた一過性発現実験により確認した。加えて、発芽 2 日目、展開葉、熱ストレス時の培養細胞での翻訳状態についてもモデル式の構築を行い、それぞれの条件での翻訳状態の決定に真に重要な配列的特徴を明らかにし、各モデル式で示された重要な配列的特徴を比較することで、様々な条件で共通した翻訳機構と特定の条件に応答した特異的な翻訳機構について考察した。

2-2. 材料と方法

2-2-1. 使用植物体および培養細胞

2-2-2-1. 使用植物体の条件

第一章の条件と同様である。

2-2-2-1. 使用培養細胞条件

シロイヌナズナ培養細胞 (*Arabidopsis thaliana* T87, Axelos et al. 1992) は理化学研究所ジーンバンク室植物開発銀行より分与していただいたものを使用した。培養は改変 LS 培地 (Nagata, 1992) を使用し、22°C、24 時間明期、振とう速度 120 rpm の条件で行った。1 週間ごとに定常期に達した細胞 4 mL を新しい培地 95 mL に移植し継代培養を行った。

2-2-2. Cap Analysis of Gene Expression (CAGE)

2-2-2-1. 植物体のサンプリング

発芽 2 日目 (2 day after germination: 2DAG) の植物体と、21DAG の植物より未展開葉から 3 枚を未展開葉 (young leaves) として、子葉を除いた古い葉から 3 枚を展開葉 (mature leaves) として切り取り、液体窒素中で凍結させ、-80°Cにて保存した。21DAG の植物体については根を取り除いている。

2-2-2-2. 熱および塩ストレス処理

ストレス処理には、植え継ぎ後 3 日目の培養細胞を用いた。熱ストレス処理は 37°C で 10 min の振とう培養を行った。塩ストレス処理は終濃度 200 mM となるように NaCl を細胞培養液に加え、22°C で 10 min の振とう培養を行った。各種ストレス処理後、吸引濾過により細胞を回収した。なお、塩ストレス処理した細胞に関しては、吸引濾過後に新たな改変 LS 培地 100 mL で細胞を洗浄し、再度の吸引濾過を行った。回収した細胞は、液体窒素中で凍結させ、-80°Cにて保存した。また、コントロール条件として未処理の培養細胞も同様に回収し保存した。

2-2-2-2. CAGE ライブラリーの作製

2DAG、未展開葉 (young_leaves)、展開葉 (mature_leaves)、培養 3 日目コントロール (Control)、熱ストレス下 (37°C)、塩ストレス下 (200_mM) の細胞から、Trizol reagent (Invitrogen, USA) を用いてトータル RNA を抽出し、CAGE ライブラリーの作製に供した。CAGE ライブラリーの作製は、Control、37°C、200_mM に関しては Takahashi ら (Takahashi et al. 2012) の CAGE ライブラリーの作製手法に従った。2DAG、young_leaves、mature_leaves に関しては、それらの手法を更に発展させた Murata ら (Murata et al. 2014) の nAnT-iCAGE ライブラリーの作製手

法に従った。概要を述べると、N15 ランダムプライマーを用いて相補鎖 cDNA を合成し、キャップ・トラッピング法により cDNA の 5'末端を選別した。ついで、RNaseI を用いて RNA 鎖を除去して得られた一本鎖 cDNA の 5'末端にリンカーを結合した。Takahashi ら (Takahashi et al. 2012) の手法では、このリンカーにシーケンスプライマー認識部位、アダプター配列、クラス IIs 制限酵素 (EcoP15I) の認識部位があり、EcoP15I によって 5'末端の 26 塩基を切り出した。その後、3'側にもリンカーを結合させ、CAGE ライブラリーとした。Murata ら (Murata et al. 2014) の手法では、5'末端に結合させたリンカーにはクラス IIs 制限酵素の認識部位はなく、切り出しは行わずにそのまま 3'側にもリンカーを結合させ、nAnT-iCAGE ライブラリーとした。

2-2-2-3. シークエンス解析

Control、37°C、200_mM に関しては、illumina (R) HiSeq 2000 を、2DAG、young_leaves、mature_leaves に関しては、illumina (R) HiSeq 2500 を用いてシークエンス解析を行った。シークエンス解析は、5'末端に結合させたリンカーに存在するシークエンスプライマー認識部位を用いたシングルリードにて、付属のプロトコールに従って行った。

2-2-2-4. マッピング

得られた raw データのそれぞれのタグから CAGE リンカー配列を除去し、長さを一定にトリミングした。その後、配列中に正確に読まれていないことを示す N が存在するタグを除去し、TAIR10 (<http://www.arabidopsis.org/index.jsp>) の情報を基にマッピングを行った。タグがマッピングされたゲノム上の位置を取得し、それぞれの位置におけるタグ数をカウントした。なお、複数位置にマッピングされ MapQuality が低いタグは除去した。また、独立して行った 2 つのサンプル間で、ともにゲノム上にタグが存在する位置についてのみタグ数をカウントし、Tag per million (TPM) 値に変換して使用した。

2-2-2-5. データ処理

各 Tag は、TAIR10 に登録されている染色体コードのタンパク質をコードする各遺伝子について、ストランド方向が同じで TAIR10 に登録された TSS の上流 500 nt から CDS の AUG までの範囲に存在しているものを、その遺伝子の TSS を示すものとしてアノテーションした。なお、TAIR10 では、いくつかの遺伝子についてスプライスバリエントが登録されているが、CAGE タグではバリエント間を区別することができない。そのため、スプライスバリエントが存在する場合は、TAIR10_representative_gene_models (<http://www.arabidopsis.org/index.jsp>) の情報から代表的なスプライスバリエントのみを用いている。遺伝子が存在しない領域または CDS 領域にマッピングされ、遺伝子のアノテーションをつけることができない

かったタグは除去した。最終的に得られたタグが存在する各ゲノム上の位置は、その遺伝子の転写開始点 (TSS) を示し、その位置における TPM 値をその TSS 由来の mRNA 量を示す TPM_TSS 値とした。そして各遺伝子の全 TSS における TPM 値を合計することで、遺伝子レベルでの mRNA 量 (TPM_Sum) を算出した。続いて、各 TSS 由来の mRNA の分布率 (TSS 分布率) を以下の式で算出した。

$$\text{Distribution_Ration_of_TSS}_{ij} (DR_{ij}) = \frac{\text{TPM_TSS}_{ij}}{\text{TPM_Sum}_i} \times 100$$

遺伝子 i における j 番目の TSS の TSS 分布率 (DR_{ij}) を示す。ここで、 n は TSS が存在するゲノム上の位置の数、 TPM_TSS_{ij} は遺伝子 i の j 番目の TPM_TSS 値、 TPM_Sum_i は遺伝子 i の TPM_Sum 値である。

また、各遺伝子における最大の TSS 分布率、つまり最も主要な TSS 由来の mRNA の存在比率 (Max Distribution Ratio of TSSs: MDR) を TSS の分散度合いの指標値とした。加えて、各条件間における各遺伝子の TSS 分布の変化を、変化の指標値である Δ TSS 値として以下の式から算出した。

$$\Delta\text{TSS}_i = \sum_j^n \frac{|DR_{Aij} - DR_{Bij}|}{200}$$

条件 A と条件 B 間における遺伝子 i の Δ TSS を示す。ここで、 n は TSS が存在するゲノム上の位置の数、 DR_{Aij} は条件 A の遺伝子 i の j 番目の TSS における分布率 (%)、 DR_{Bij} は条件 B の遺伝子 i の j 番目の TSS における分布率 (%) である。CAGE タグは、その 5'末端に Cap 構造に由来する G が付加されている。そのため 5'UTR 配列を決定する場合は、タグの 5'末端の 1 nt 下流からの配列を用いている。また、ある TSS から転写される mRNA の 5'UTR は、TAIR10 の情報を基に CDS の AUG までのゲノム配列を取得し、イントロンが存在する場合は、そのイントロンの配列を除去することで決定した。これら全てのデータ処理は、自らで独自に構築したプログラムを用いて行った。

2-2-3. Partial Least Squares Regression (PLS)モデルの構築

2-2-3-1. 使用したデータセット

予備的な説明変数の選抜には、ポリソーム/マイクロアレイ解析と CAGE の両方のデータが存在し、TSS が 66.7%以上 1 点に収束している遺伝子 (予備的な説明変数の選抜用データセット: 2DAG、未展開葉、展開葉、37°Cそれぞれ 857、555、643、808 個) について、それぞれの解析法に合わせて更に選抜して使用した。解析の主となる PLS モデルの構築には、分布率が 5%以上のある程度は転写されている TSS 由来の各 5'UTR バリエントの mRNA が、各遺伝子の mRNA の 80%以上

を構成するものについて、それらの Log_{10} PR 値の分布が 0.005 区切りに最大でも 1 つの遺伝子しか含まれないように選抜した (PLS モデルの構築と検証用データセット: 2DAG、未展開葉、展開葉、37°Cそれぞれ 877、1023、1026、1067 遺伝子)。その後、さらにデータセットを 2 つに分け (それぞれ Log_{10} PR 値が 0.01 区切りに 1 つ以下になるように)、PLS モデルの構築用のデータセット (PLS モデルの構築用データセット: 2DAG、未展開葉、展開葉、37°Cそれぞれ 439、512、508、534 遺伝子) と検証用のデータセット (PLS モデルの検証用データセット: 2DAG、未展開葉、展開葉、37°Cそれぞれ 438、511、508、533 遺伝子) とした。

2-2-3-2. 予備的な説明変数の選抜 (Region_5', Region_AUG)

2-2-3-2-1. 5'UTR 内の特定の領域における塩基含量の評価

Region_5'および Region_AUG の選抜には、予備的な説明変数の選抜用データセットから、5'UTR 長が 50 nt 以上の遺伝子において、 Log_{10} PR 値(y)を 0.01 区切りにした中に、遺伝子が最大でも 1 個しか含まれないように N 個の遺伝子を選抜して使用した。それら N 個の遺伝子における塩基位置 k から k+L-1 の範囲 (本研究では $k+L-1 \leq 50$ とする) の長さ L の配列と Log_{10} PR 値(y)の関係を数理モデルにより表現し、この数理モデルから y 値と最も関連のある塩基配列の特徴を抽出した。まず初めに、i 番目のサンプルについて配列の k 番目から長さ L の範囲を抜き出した配列を $Seq_i = s_{ik}s_{ik+1}\cdots s_{ik+L-1}$ とした。ここで、i 番目の遺伝子における塩基配列上の位置を Region_5'の選抜時は 5'端より 3'方向に、Region_AUG の選抜時は、AUG より 5'方向に、 $s_1, s_2, \dots, s_k, \dots, s_M$ とした。また、i 番目の遺伝子の Log_{10} PR 値を y_i とした (Table 10)。そして、N 個の遺伝子における塩基位置 k から k+L-1 の範囲の長さ L の配列において、少なくとも 1 回出現する t 個の塩基からなる配列を $R_1(t), R_2(t), \dots, R_v(t), R_v(t)$ とした。本研究では $t = 1 \sim 2$ とした。また、それぞれの配列の出現頻度を $f_i^{(k, k+L-1)}(R_1(t)), \dots, f_i^{(k, k+L-1)}(R_v(t)), \dots, f_i^{(k, k+L-1)}(R_v(t))$ とした。ここで、v 番目の配列頻度を変量 $f_i^{(k, k+L-1)}(R_v(t))$ と表した (Table 11)。後述する PLS には、これらの出現頻度を用いている。

	s_1	s_2	\cdots	s_k	\cdots	s_{k+L-1}	\cdots	s_M	y
1	s_{11}	s_{12}	\cdots	s_{1k}	\cdots	$s_{1(k+L-1)}$	\cdots	s_{1M}	y_1
2	s_{21}	s_{22}	\cdots	s_{2k}	\cdots	$s_{2(k+L-1)}$	\cdots	s_{2M}	y_2
\vdots	\cdots	\cdots	\cdots	\cdots	\cdots	\cdots	\cdots	\cdots	\cdots
i	s_{i1}	s_{i2}	\cdots	s_{ik}	\cdots	$s_{i(k+L-1)}$	\cdots	s_{iM}	y_i
\vdots	\cdots	\cdots	\cdots	\cdots	\cdots	\cdots	\cdots	\cdots	\cdots
N	s_{N1}	s_{N2}	\cdots	s_{Nk}	\cdots	$s_{N(k+L-1)}$	\cdots	s_{NM}	y_N

Table 10. 配列情報と Log_{10} PR 値の定義

	$f^{(k,k+L-1)}(R_1(t))$	$f^{(k,k+L-1)}(R_2(t))$...	$f^{(k,k+L-1)}(R_v(t))$...	$f^{(k,k+L-1)}(R_V(t))$
1	$f_1^{(k,k+L-1)}(R_1(t))$	$f_1^{(k,k+L-1)}(R_2(t))$...	$f_1^{(k,k+L-1)}(R_v(t))$...	$f_1^{(k,k+L-1)}(R_V(t))$
2	$f_2^{(k,k+L-1)}(R_1(t))$	$f_2^{(k,k+L-1)}(R_2(t))$...	$f_2^{(k,k+L-1)}(R_v(t))$...	$f_2^{(k,k+L-1)}(R_V(t))$
⋮
i	$f_i^{(k,k+L-1)}(R_1(t))$	$f_i^{(k,k+L-1)}(R_2(t))$...	$f_i^{(k,k+L-1)}(R_v(t))$...	$f_i^{(k,k+L-1)}(R_V(t))$
⋮
N	$f_N^{(k,k+L-1)}(R_1(t))$	$f_N^{(k,k+L-1)}(R_2(t))$...	$f_N^{(k,k+L-1)}(R_v(t))$...	$f_N^{(k,k+L-1)}(R_V(t))$

Table 11. N個のサンプルにおける区間[k,k+L-1]の長さLの配列におけるt個の連続塩基($R_1(t), R_2(t), \dots, R_V(t)$)の頻度

2-2-3-2-2. PLSによる回帰係数(重み)とQ²値の算出

説明変数 X (N×V 行列) として 2-2-3-2-1. で評価した区間[k,k+L-1]の長さLの配列における V 個の配列頻度 $f^{(k,k+L-1)}(R_v(t))$, ($v = 1, 2, \dots, V$) と、目的変数 (N×1) として Log₁₀ PR 値(y)の線形下位形式を以下の式により表現した。

$$y = a_{R_1(t)}^{(k,k+L-1)} f^{(k,k+L-1)}(R_1(t)) + a_{R_2(t)}^{(k,k+L-1)} f^{(k,k+L-1)}(R_2(t)) + \dots + a_{R_v(t)}^{(k,k+L-1)} f^{(k,k+L-1)}(R_v(t)) + \dots + a_{R_V(t)}^{(k,k+L-1)} f^{(k,k+L-1)}(R_V(t)) + a_{0(t)}^{(k,k+L-1)}$$

ここで、 $a_{R_v(t)}^{(k,k+L-1)}$ は、v 番目の塩基頻度における回帰係数である。続いて、

Partial Least Squares (PLS)法により回帰モデル(PLS モデル)を構築し、この回帰係数を求めた。PLS 法は、説明変数 X (N×V 行列) を目的変数 y (N×1) へ線形的に関連付ける方法である。PLS は重回帰分析にみられる変数間の共線性や、応答変数よりも過剰に多い説明変数が存在する時に起きる過度の当てはめの問題を解決できるため、多変量解析の分野では広く使われている方法である。

PLS は以下の式で表した。

$$\mathbf{X} = \sum_{k=1}^D \mathbf{t}_k \mathbf{p}_k^T + \mathbf{E}$$

$$\mathbf{y} = \sum_{k=1}^D \mathbf{t}_k q_k + \mathbf{e}$$

ここで、 \mathbf{p}_k は X における k 番目の成分の重みベクトルであり、 q_k は y における k 番目の成分の係数である。また D は PLS の成分数、 \mathbf{t}_k は k 番目の潜在変数、E は X の残差であり、e は y の残差である。PLS の成分数である D は、成分数

を増加させるたびに Leave-one-out cross-validation により Q^2 値を逐次計算し、 Q^2 値が最大になるときの成分数として決定される。 Q^2 はモデルの予測精度を示す尺度であり、以下の式で表した。

$$Q^2 = 1 - \frac{\sum (y_{obs} - y_{pred})^2}{\sum y_{obs}^2}$$

ここで、 y_{obs} は実験的に得られた実測値であり、 y_{pred} は構築したモデルによる予測値である。また PLS は以下の式にまとめることができる。

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \mathbf{f}$$

このとき、 \mathbf{a} は回帰係数ベクトルであり、その要素は a_j ($j = 1, 2, \dots, N$) で表される。また \mathbf{f} は \mathbf{y} の残差である。

2-2-3-2-3. 重要領域(Region_5', Region_AUG)の選抜

Region_5'の選抜のため、5'UTR の 5'末端の塩基位置 k から $k+L-1$ の範囲の長さ L の配列において 2-2-3-2-2.で示した PLS モデルを構築し、その Q^2 値を $Q^2_{(k,k+L-1)}$ とした。それら $Q^2_{(k,k+L-1)}$ のランキングを作成し、 $Q^2_{(k,k+L-1)}$ が上位 1% に位置した Z 個の重要領域 $[k,k+L-1]$ を選抜し、新たに重要領域 Region_5' $_z$ ($z = 1, 2, \dots, Z$) とした。なお、上位 1% に位置する領域でも、より上位の領域と 50% 以上が重複している場合は除外した。Region_AUG の選抜は、AUG の塩基位置 k から 5'方向へ $k+L-1$ の範囲の長さ L の配列において Region_5'と同様に行い、 Z' 個の重要領域 $[k,k+L-1]$ を選抜し、新たに重要領域 Region_AUG $_z'$ ($z' = 1, 2, \dots, Z'$) とした。ある配列 i の Region_5' $_z$ における評価値は、塩基の出現頻度 $f_i^{Region_5'_z}(R_1(t)), \dots, f_i^{Region_5'_z}(R_V(t))$ 、Region_AUG $_z'$ における評価値は、塩基の出現頻度 $f_i^{Region_AUG'_z'}(R_1(t)), \dots, f_i^{Region_AUG'_z'}(R_V(t))$ となる。

2-2-3-3. 予備的な説明変数の選抜 (Motif)

特定の配列的パターン(Motif)の選抜には、予備的な説明変数の選抜用データセットから、5'UTR 長が 50 nt 以上の遺伝子について、 Log_{10} PR 値(y)の分布の上位 10% に存在する集団 (上位集団) と下位 10% に存在する集団 (下位集団) をさらに選抜して行った。解析は、MEME SUITE (<http://meme-suite.org/index.html>, Bailey et al. 2015) で提供されている Multiple Expectation-maximization for Motif Elicitation (MEME) を用いた。また、豊富に出現する Motif の最尤推定をする際のバックグラウンドモデルには、予備的な説明変数の選抜用データセットの全 5'UTR 配列の AUGC 含量から算出した 0~1 order Markov モデルを使用した。解析は、”Discriminative mode”を用い、上位集団に豊富に含まれている Motif を選抜する際

は下位集団を対照に、下位集団では上位集団を対照にした。相補鎖は解析対象から外し、 $E\text{-value} < 0.5$ かつ、解析対象とした遺伝子集団の 25%以上の 5'UTR に出現する Z'' 個の Motif $_{z''}$ ($z'' = 1, 2, \dots, Z''$) を選抜した。後の解析では、対象となる遺伝子の Motif $_{z''}$ を、MEME SUITE にて提供されている Find Individual Motif Occurrences (FIMO) を用いて、評価して使用している。具体的には、FIMO により、5'UTR 配列から $p\text{-value} < 0.1$ でその Motif $_{z''}$ に類似している配列を探索し、各遺伝子 i で最も $p\text{-value}$ が低い配列について、Motif $_{z''}$ との類似度を $f_i^{\text{Motif}_{z''}} = -\text{Log}_{10} p$ として評価した。評価は、各遺伝子の 5'UTR 全体、Cap または AUG から 25 nt, 50 nt, 75 nt, 100 nt の 9 個の領域 ($R'_{v'}$, $v' = 1, 2, \dots, 9$) について行ない、ある配列 i の Motif $_{z''}$ の各領域 $R'_{v'}$ における評価値は、類似度 $f_i^{\text{Motif}_{z''}}(R'_{v'})$, \dots , $f_i^{\text{Motif}_{z''}}(R'_9)$ となる。

2-2-3-4. 予備的な説明変数の選抜 (Other)

2-2-3-4-1. 想定したその他の特徴(Other)の評価

Other の選抜には、予備的な説明変数の選抜用データセットから、 Log_{10} PR 値(y)を 0.01 区切りにした中に、遺伝子が最大でも 1 個しか含まれないように N 個の遺伝子を選抜して使用した。それら N 個の遺伝子における(1) 5'UTR 全体の塩基含量、(2) uORF、(3)二次構造、(4) rRNA との相補長、(5) mRNA 長、(6) 転写量に関わる様々な特徴を評価した。以下で各項目を具体的に述べる。

(1) 5'UTR 全体の塩基含量は、遺伝子 i の 5'UTR 全体の配列における 1 塩基パターン (A, U, G, C) および 2 塩基パターン (AA, AU, \dots , CC) の合計 20 種の塩基パターンの出現頻度を $f_i^{\text{TempOther}_1}$, \dots , $f_i^{\text{TempOther}_{20}}$ として評価した。

(2) uORF は、AUG と過去に非 AUG 開始コドンとして報告されている UUG, CUG, ACG, AUU, GUG, AUA, AUC, AAG, AGG について (Brar et al. 2012, Ohta et al. 2010, Depeiges et al. 2006)、遺伝子 i の 5'UTR 上に存在する uAUG, uUUG, \dots , uAGG と、それと同一フレームで終止コドンが存在している uORF (領域重複あり), uORF_uug, \dots , uORF_agg、領域が重複していない uORF', uORF_uug', \dots , uORF_agg' の合計 30 個の配列的特徴の頻度を $f_i^{\text{TempOther}_{21}}$, \dots , $f_i^{\text{TempOther}_{50}}$ として評価した。

(3) 二次構造は、ViennaRNA Package (Lorenz et al. 2011, <http://www.tbi.univie.ac.at/RNA/>) の RNAfold を用いて、遺伝子 i の 5'UTR 全体、CDS 全体、3'UTR 全体、mRNA 全体と Fig. 11 で示す 49 領域の合計 53 領域の二次構造の形成度合い ($-\Delta G$) を、 $f_i^{\text{TempOther}_{51}}$, \dots , $f_i^{\text{TempOther}_{103}}$ として評価した。 $-\Delta G$ の計算には Turner ら (Turner et al. 2004) のパラメータを使用した。

(4) rRNA との相補長は、ViennaRNA Package の RNAduplex を用いて、遺伝子 i の 5'UTR 全体、CDS 全体、3'UTR 全体、mRNA 全体と Fig. 11 で示す 49 領域の合計 53 領域の、5'UTR が 18S rRNA と相補する最大配列長 (nt) を、

$f_i^{TempOther_{104}}, \dots, f_i^{TempOther_{156}}$ として、25S rRNA と相補する最大配列長 (nt) を、 $f_i^{TempOther_{157}}, \dots, f_i^{TempOther_{209}}$ として評価した。計算時のパラメータには Turner ら (Turner et al. 2004) のものを、18S rRNA と 25S rRNA の配列は TAIR10 のものを使用した。

(5) mRNA 長は、遺伝子 i の 5'UTR、CDS、3'UTR、mRNA 長とその対数值(Log₁₀)、合計 8 個の配列的特徴を $f_i^{TempOther_{210}}, \dots, f_i^{TempOther_{217}}$ として評価した。

(6) 転写量に関わる特徴は、遺伝子 i の TPM_Sum 値とその対数值(Log₁₀)、Max Distribution Ratio of TSS (MDR) を $f_i^{TempOther_{218}}, \dots, f_i^{TempOther_{220}}$ として評価した。

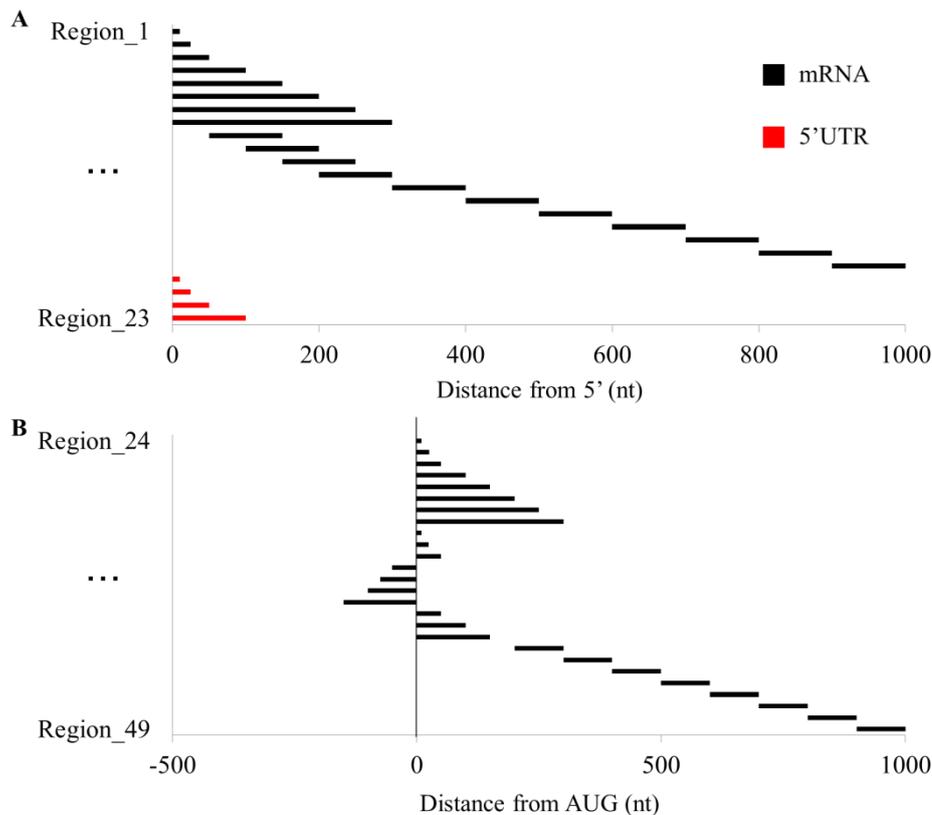


Figure 11. 二次構造および rRNA との相補長を計算した mRNA 上の領域

(A) mRNA の 5'末端からの領域。0 nt の位置は Cap の位置を示す。黒で示す領域は mRNA 配列を対象に、赤で示す領域は 5'UTR 配列を対象に解析している。(B) mRNA の AUG からの領域。0 nt の位置は AUG を示す

2-2-3-4-2. Other を用いた PLS モデルの構築

評価した TempOther_v, ($v = 1, 2, \dots, 220$) を以下の式で標準化し、説明変数 X ($N \times 220$ 行列) とした。

$$f_i^{vTempOther_v} = \frac{(f_i^{TempOther_v} - \overline{f^{TempOther_v}})}{\sigma}$$

この時、 $\overline{f_i^{TempOther_v}}$ と σ はデータセット中の全遺伝子の $f_i^{TempOther_v}$ の平均値と標準偏差を示す。続いて、Log₁₀ PR 値(y)を目的変数 (N×1) に用いて 2-2-3-2-2. と同様に PLS モデルを構築し Q² 値を算出した。220 個の標準化した評価値 $f_i^{TempOther_v}$ (v = 1, 2, ..., 220) と目的変数(y)の線形下位形式を以下の式により表現した。

$$y = a_1 f_1^{TempOther} + a_2 f_2^{TempOther} + \dots + a_v f_v^{TempOther} + \dots + a_V f_{220}^{TempOther} + a_0$$

2-2-3-4-3. Q² 値による Other の選抜

2-2-3-4-2.にて構築した PLS モデルの Q² 値を $Q_{OtherTemp_0}^2$ とし、構築時の説明変数行列から各特徴 v の評価値 $f_i^{TempOther_v}$, (v = 1, 2, ..., 220) を 1 つずつ除いた PLS モデルの $Q_{OtherTemp_v}^2$ を計算し、 $\Delta Q_{OtherTemp_v}^2$ を以下の式で算出した (Table 12)。この $\Delta Q_{OtherTemp_v}^2$ が正を示す特徴 TempOther_v は、PLS モデルの精度に負の影響を与えているため、PLS モデルから削除した。削除後の全特徴 V 個を用いて、また PLS モデルを作成し、この PLS モデルを $PLSModel_{other_0}$ 、その Q² 値を $Q_{Other_0}^2$ とした。この際に、 $Q_{Other_0}^2$ が $Q_{OtherTemp_0}^2$ よりも低くなる場合は、最も $\Delta Q_{OtherTemp_v}^2$ が大きい特徴 v のみを削除し、その PLS モデルを $PLSModel_{other_0}$ 、Q² 値を $Q_{Other_0}^2$ とした。この PLS モデル $PLSModel_{other_0}$ から特徴 TempOther_v の $\Delta Q_{OtherTemp_v}^2$ を計算し、PLS モデルの精度に負の影響を与える特徴の削除を繰り返した。この特徴の削除を、 $Q_{Other_c}^2$ が 5 回連続で低下するまで C 回繰り返し、最も高い $Q_{Other_c}^2$ を示す $PLSModel_{other_c}$ に含まれる Z''' 個の特徴を Other_{Z'''} (Z''' = 1, 2, ..., Z''') として選抜した。ある配列 i の Other_{Z'''} の評価値は、特徴量 $f_i^{Other_{Z'''}}$ となる。解析はオープンソースされている ActiveSeqAnalyzerTM (<http://kanaya.naist.jp/ActiveSeqAnalyzerTN/intro.html>) に独自のプログラムを追加して使用した。

$$\Delta Q_{OtherTemp_v}^2 = Q_{OtherTemp_0}^2 - Q_{OtherTemp_v}^2$$

PLSModel	TempOther ₁	...	TempOther _v	...	TempOther _v	Q ²
$PLSModel_{other_0}$	$f_i^{TempOther_1}$...	$f_i^{TempOther_v}$...	$f_i^{TempOther_v}$	$Q_{OtherTemp_0}^2$
$PLSModel_{other_1}$	-----	...	$f_i^{TempOther_v}$...	$f_i^{TempOther_v}$	$Q_{OtherTemp_1}^2$
...
$PLSModel_{other_v}$	$f_i^{TempOther_1}$...	-----	...	$f_i^{TempOther_v}$	$Q_{OtherTemp_v}^2$
...
$PLSModel_{other_V}$	$f_i^{TempOther_1}$...	$f_i^{TempOther_v}$...	-----	$Q_{OtherTemp_V}^2$

Table 12. PLS モデルの精度に負の影響を与える説明変数の削除

2-2-3-5. 解析の主となる PLS モデルの構築と説明変数の選抜

2-2-3-5-1. 選抜した予備的な説明変数を用いた PLS モデルの構築

PLS モデルの構築と検証用データセットの N 個の遺伝子の各 U 個の 5'UTR バリエント mRNA における、2-2-3-2-3.にて選抜した Z 個の Region_5' $_z$ ($z = 1, 2, \dots, Z$)、 Z' 個の Region_AUG $_{z'}$ ($z' = 1, 2, \dots, Z'$) の塩基の出現頻度、2-2-3-3.にて選抜した Z'' 個の Motif $_{z''}$ ($z'' = 1, 2, \dots, Z''$) の領域 ($R'_{v'}$, $v' = 1, 2, \dots, 9$) の類似度、2-2-3-4-3.にて選抜した Z''' 個の Other $_{z'''}$ ($z''' = 1, 2, \dots, Z'''$) について、合計 V ($Z + Z' + Z'' + Z'''$) 個を f_{iu}^v ($v = 1, 2, \dots, V$) として評価した。 U 個の 5'UTR バリエント mRNA の評価値を遺伝子単位で合成し、遺伝子 i の評価値 f_i^v ($v = 1, 2, \dots, V$) を算出後、標準化し説明変数 x_i^v ($v = 1, 2, \dots, V$) とした。計算式を以下に示す。

$$f_i^v = \sum_{u=1}^U (f_{iu}^v \times \frac{DR_u}{\sum_{u=1}^U DR_u})$$

$$x_i^v = \frac{(f_i^v - \bar{f}^v)}{\sigma}$$

この時、 DR はその 5'UTR バリエント mRNA の存在比率、 \bar{f}^v と σ はデータセット中の全遺伝子の f_i^v の平均値と標準偏差を示す。算出した説明変数 X' ($N' \times V$ 行列) から、PLS モデルの構築用データセットの N 個のデータを抽出し、 x_i^v ($i = 1, 2, \dots, N$), ($v = 1, 2, \dots, V$) を説明変数 X ($N \times V$ 行列)、 Log_{10} PR 値 (y) を目的変数 ($N \times 1$) に用いて 2-2-3-2-2.と同様に PLS モデルを構築し Q^2 値を算出した。 V 個の説明変数と目的変数 (y) の線形下位形式を以下の式により表現した。なお、説明変数 x_i^v ($v = 1, 2, \dots, V$) と、 $x_i^{\text{Region}_5'1}$ ($R_1(t)$), \dots , $x_i^{\text{Region}_5'Z}$ ($R_V(t)$), $x_i^{\text{Region}_AUG1}$ ($R_1(t)$), \dots , $x_i^{\text{Region}_AUGZ'}$ ($R_{V'}(t)$), $x_i^{\text{Motif}1}$ (R'_1), \dots , $x_i^{\text{Motif}Z''}$ (R'_9), $x_i^{\text{Other}1}$, \dots , $x_i^{\text{Other}Z'''}$ は同じ意味を表している。

$$y = a_1 x_1 + a_2 x_2 + \dots + a_v x_v + \dots + a_V x_V + a_0$$

2-2-3-5-2. Q^2 値による重要な配列的特徴の選抜

2-2-3-5-1.にて構築した PLS モデルの Q^2 値を $Q^2_{Temp_0}$ とし、構築時の説明変数行列から各特徴 v の評価値 $x_i^{\text{Other}v}$, ($v = 1, 2, \dots, V$) を 1 つずつ除いた PLS モデルの $Q^2_{Temp_v}$ を計算し、 $\Delta Q^2_{Temp_v}$ を算出した。以降は 2-2-3-2-3.と同様に、PLS モデルの精度に負の影響を与える特徴の削除を繰り返した。最も高い Q^2 を示した PLS モデル ($PLSModel_c$) に含まれる V' 個の説明変数 $x^{v'}$ ($v' = 1, 2, \dots, V'$) を重要な配列的特徴として選抜した。選抜した重要な配列的特徴 v' が PR 値に与える影響 (重み) として v' の回帰係数 $a^{v'}$ を使用した。

2-2-3-5-2. 構築した PLS モデルによる予測値と残差の算出

PLS モデルの検証用データセットや予備的な説明変数の選抜用データセットにおける遺伝子 i について、構築した PLS モデルの V 個の説明変数 x_i^v ($v = 1, 2, \dots, V$) を評価・算出し、算出した x_i^v と a^v から y_i の予測値 (Predicted_ y_i) と y_i の残差 (Residual $_i$) を計算した。なお x_i^v の標準化の際の平均値と標準偏差には、PLS モデルの構築と検証用データセットのものを使用している。計算式を以下に示す。

$$\text{Predicted_}y_i = \sum_{v=1}^V a^v \times x_i^v$$

$$\text{Residual}_i = y_i - \text{Predicted_}y_i$$

2-2-3-6. 予備的な説明変数の再選抜

予備的な説明変数の選抜用データセットの遺伝子 N 個について、 $PLSModel_c$ における、遺伝子 i の残差 Residual $_i$ を算出した。この時、この残差 (y 値、つまり Log $_{10}$ PR 値を構築した PLS モデルが説明できていない大きさ) が予備的な説明変数の選抜時の Region_5', Region_AUG, Motif, Other に生じている (計算できていない影響がある) と仮定するならば、Region_5', Region_AUG, Motif, Other の遺伝子 i の Log $_{10}$ PR 値への真の影響 ($\sum True_a^{Region_5'v} \times True_x_i^{Region_5'v}$ など) は以下の式で表すことができる。

$$\text{Residual}_i = \text{Residual}_i^{\text{Region_}5'} + \text{Residual}_i^{\text{Region_}AUG} + \text{Residual}_i^{\text{Motif}} + \text{Residual}_i^{\text{Other}}$$

$$\sum_{v'=1}^{V'} (True_a^{\text{Region_}5'v'} \times True_x_i^{\text{Region_}5'v'})$$

$$= \sum_{v'=1}^{V'} (a^{\text{Region_}5'v'} \times x_i^{\text{Region_}5'v'}) + \text{Residual}_i^{\text{Region_}5'}$$

$$\sum_{v'=1}^{V'} (True_a^{\text{Region_}AUGv'} \times True_x_i^{\text{Region_}AUGv'})$$

$$= \sum_{v'=1}^{V'} (a^{\text{Region_}AUGv'} \times x_i^{\text{Region_}AUGv'}) + \text{Residual}_i^{\text{Region_}AUG}$$

$$\sum_{v'=1}^{V'} (True_a^{\text{Motif}v'} \times True_x_i^{\text{Motif}v'}) = \sum_{v'=1}^{V'} (a^{\text{Motif}v'} \times x_i^{\text{Motif}v'}) + \text{Residual}_i^{\text{Motif}}$$

$$\sum_{v'=1}^{V'} (True_a^{Other_{v'}} \times True_x_i^{Other_{v'}}) = \sum_{v'=1}^{V'} (a^{Other_{v'}} \times x_i^{Other_{v'}}) + Residual_i^{Other}$$

この時、 v と v' はそれぞれの PLS モデルに重要な説明変数の数、 $True_x_i$ は遺伝子 i の真の説明変数、 $True_a$ はその重み、 x_i は遺伝子 i の既存の PLS モデルにおける説明変数、 a はその重みを示す。この真の説明変数はあくまで仮定であり、真の説明変数を選抜することは事実上不可能であるが、少しでも尤もらしい説明変数を用いた PLS モデルを構築するため、予備的な説明変数の選抜時の y 値を Log_{10} PR 値ではなく以下で計算する値とし、予備的な説明変数 (Region_5', Region_AUG, Motif, Other) の再選抜を行った。

$$y(\text{Region_5}') = \sum_{v'=1}^{V'} (a^{\text{Region_5}'_{v'}} \times x_i^{\text{Region_5}'_{v'}}) + Residual_i$$

$$y(\text{Region_AUG}) = \sum_{v'=1}^{V'} (a^{\text{Region_AUG}_{v'}} \times x_i^{\text{Region_AUG}_{v'}}) + Residual_i$$

$$y(\text{Motif}) = \sum_{v'=1}^{V'} (a^{\text{Motif}_{v'}} \times x_i^{\text{Motif}_{v'}}) + Residual_i$$

$$y(\text{Other}) = \sum_{v'=1}^{V'} (a^{\text{Other}_{v'}} \times x_i^{\text{Other}_{v'}}) + Residual_i$$

再選抜後の説明変数を用いて同様に PLS モデルの構築を 2-2-3-5-2 の操作まで行った。それを用いて PLS モデルの検証用データセットにおける予測値を算出した。算出した予測 Log_{10} PR 値と実際の Log_{10} PR 値の間におけるピアソンの相関係数を計算し、PLS モデルの予測精度とした。再選抜の結果による予測精度の低下が 5 回連続して確認されるまで再選抜を繰り返し、再選抜の中で最も予測精度の高い PLS モデル ($PLSModel_{Best}$) を使用した。再選抜の評価に予測精度を使用した理由は、再選抜を繰り返すことによる説明変数の y 値への過度の当てはめを避けるためである。

2-2-3-7. 最終的な PLS モデルの確定

2-2-3-6. で構築した $PLSModel_{Best}$ で示された重要な説明変数から、さらに偽陽性が疑われる説明変数を除去するために、PLS モデルの検証用データセットにおける $PLSModel_{Best}$ の予測値と予測精度を算出し、除外することで予測精度が上がる、つまりその予測精度に負の影響を与えている説明変数を除去した。その後、PLS モデルの構築用データセットの N 個の遺伝子における除去後の V 個の説明変数 X

($N \times V$ 行列) と Log_{10} PR 値を用いて、再度 $PLSModel_{Final}$ を構築した。この $PLSModel_{Final}$ を最終的な PLS モデルとした。

2-2-3-8. 重要な配列的特徴のクラスタリング

未展開葉の PLS モデルの構築と検証用データセット 1023 遺伝子を使用し、構築した 2DAG、未展開葉、展開葉、 37°C におけるそれぞれの $PLSModel_{Final}$ の説明変数、合計 45 個の x_i^v ($v = 1, 2, \dots, 45$) を評価・算出した (1023×45 行列)。このデータを用いて各説明変数間のユークリッド距離を計算し、ウォード法による階層的クラスタ分析を行った。

2-2-4. DNA 一過性発現実験

2-2-4-1. プラスミドの構築

異なる 5'UTR の 5'端に Cla I サイト、3'端には Aat II サイトを持つように設計したプライマーを用い (Table 13)、cDNA を鋳型とした PCR 反応にて目的 DNA 断片の増幅を行った。その後、Cla I、Aat II で切り出し、35S プロモーターの支配下に *F-luc* 遺伝子と HSP ターミネーターを持つプラスミド pBluescript II KS+ の Cla I /Aat II サイトに挿入した。5'UTR の塩基置換系列の導入には、末端の配列が Cla I、Aat II による切断部位と同じになるように設計し、相当する変異を含むプライマーを用い (Table 14)、アニーリング後、同様に pBluescript II KS+ の Cla I /Aat II サイトに挿入した。補正用の *R-luc* 遺伝子については、35S プロモーター、*R-luc* 遺伝子、HSP ターミネーターからなる発現カセットを持つ pBluescript II KS+ を用いた。

Name	Primer sequence (5' to 3')
At1g06760-1	TATCGATGGAGAGAACAATCCTCATAATCACTTTC TGACGTCCTCCATCGTCTTCTGAACTTAAGATC
At1g34000-1	TATCGATGGAGAGAAAGACAATTCAACTAACAAAA TGACGTCCTCCATTTTTTTTGTAGTTGAATTGT
At4g09650-1 TSS1	TATCGATGGAGAGAAGAGAAACCACAAATCTCTCT TGACGTCCTCCATTGTTGTTGTTGAGAGAGTTT
At4g09650-2 TSS2	TATCGATGGAGAGAACAATCTCTCTTTCTCTCAA TGACGTCCTCCATTGTTGTTGTTGAGAGAGTTT
At1g20440-1 TSS1	TATCGATGGAGAGAAAACATTACTCATTACAAA TGACGTCCTCCATAGTTAATCTTGATTTGATTA
At1g20440-2 TSS2	TATCGATGGAGAGAACAACATTACTCATTACAAA TGACGTCCTCCATAGTTAATCTTGATTTGATTA
At1g20440-3 TSS3	TATCGATGGAGAGAAACAACATTACTCATTACAAA TGACGTCCTCCATAGTTAATCTTGATTTGATTA

SS4	CGATGGAGAGAAGTGGATAAAGCTTCTCACTTTCAGTTCAATTGGATAG CCGGAAAGGCTCTCCTCCAAGTGGCGTTTATATGGAAAGACGT CTTCCATATAAACGCCACTTGGAGGAGAGCCTTCCGGCTATCCAATTG AACTGAAAGTGAGAAGCTTATCCACTTCTCTCCAT
SS4 _NC	CGATGGAGAGAAGCCCATAAAGCTTCTCACTTTCAGTTCAATTGGATAG CCGGAAAGGCTCTCCTCCAAGTGGCGTTTATATGGAAAGACGT CTTCCATATAAACGCCACTTGGAGGAGAGCCTTCCGGCTATCCAATTG AACTGAAAGTGAGAAGCTTATGGGCTTCTCTCCAT
SS4 _PA	CGATGGAGAGAAGAAAATAAAGCTTCTCACTTTCAGTTCAATTGGATAG CCGGAAAGGCTCTCCTCCAAGTGGCGTTTATATGGAAAGACGT CTTCCATATAAACGCCACTTGGAGGAGAGCCTTCCGGCTATCCAATTG AACTGAAAGTGAGAAGCTTATTTTCTTCTCTCCAT
SS4_P _Multi	CGATGGAGAGAAGAAAATAAATCCCTCATCCCCAACCCAACCAAGTAC CGAAGAACTAACCAAAAAATACAACATTCAAATGGAAAGACGT CTTCCATTTGAATGTTGTATTTTTTGGTTAGTTCTTCGGTACTTGGTTGG GTTGGGGATGAGGGATTTATTTTCTTCTCTCCAT

Table 14. 一過性発現実験用のプラスミド構築に用いたプライマーセット
Cla I 切断部位を形成する位置を赤色、Aat II 切断部位を形成する位置を青色で示す。

2-2-4-2. シロイヌナズ T-87 培養細胞からのプロトプラストの調製

シロイヌナズナ T87 培養細胞からのプロトプラスト調製は、Satoh ら (Satoh et al. 2004) の方法に若干の変更を加えて行った。培養細胞を 0.4 M マンニトールで洗浄した後、酵素液 (0.4 M Mannitol, 1% Cellulase RS (Yakult, Japan), 0.1% Pectolyase (Kikkoman, Japan), pH 5.5) を加え、25°C にて 1~2 時間穏やかに攪拌した。40 μ m ナイロンメッシュ (Cell Strainer; BD Falcon, USA) でろ過した後、400 rpm で 5 分間遠心し、上清を回収した。回収した上清に 0.4 M マンニトールを加え洗浄し、800 rpm で 5 分間遠心することによりプロトプラストを得た。0.4 M マンニトールによる洗浄は 2 回行っている。プロトプラストを W5 溶液 (154 mM NaCl, 125 mM CaCl₂, 5 mM KCl, 2 mM Mes-KOH, pH 5.6) に再懸濁し、氷中に 30 分静置した。細胞数の計測は血球計算板を用いて行い、W5 溶液と等量の MMg 溶液 (0.4 M mannitol, 15 mM MgCl₂, 4 mM Mes-KOH, pH 5.7) を加え、400 rpm で 5 分間遠心することによりプロトプラストを回収し、プロトプラスト濃度が 1×10^4 cell/ μ l になるように MMg 溶液に懸濁した。

2-2-4-3. プロトプラストへの DNA の導入

DNA のプロトプラストへの導入は、基本的に Kovtun ら (Kovtun et al. 2000) の polyethlen glycol (PEG) を用いた方法に従った。DNA (*F-luc* 0.4 μ g, *R-luc*

0.04 μ g, total volume 5 μ l) に 1×10^4 cell/ μ l のプロトプラストを 50 μ l 加えた後、混合液と等量の PEG 溶液 (40% PEG 4000, 0.2 M Mannitol, 0.1 M $\text{Ca}(\text{NO}_3)_2$) (Sheen, 2001) を加えてゆっくりと混和し、細胞を 20 分間室温にて静置した。その後、protoplast-medium (Dansako et al., 2003) を 500 μ l 加え、22 $^{\circ}$ C で 6 時間静置し、遠心操作によって上清を取り除き、液体窒素で凍結して -80 $^{\circ}$ C にて保存した。

2-2-4-4. ルシフェラーゼ活性測定

細胞の溶解は、 5×10^5 個のプロトプラストあたり 50 μ l の passive lysis buffer (Promega, USA) を用い、ミキサー処理によって溶解し、氷上で 15 分間静置した。その後、4 $^{\circ}$ C、14000 rpm で 10 分間遠心し、上清中の F-luc と R-luc の活性値を測定した。測定には Dual-luciferase reporter assay system (Promega) とルミノメータ (Lumat LB 9501; Berthold, Northern Black Forest, Germany) を付属のプロトコールに従って使用した。

2-3. 結果

2-3-1. CAGE によるゲノムワイドな転写開始点の解析

植物においても、翻訳状態は mRNA 種によって異なっており、加えて成長・発達、環境ストレス下でその翻訳状態は異なっている（変化している）ことが明らかとなった。一般的に、mRNA の翻訳状態（効率）を規定する非常に重要な要素として mRNA の 5'非翻訳領域(5'UTR)が挙げられる (Roy et al. 2013, Kawaguchi et al. 2005, Branco-Price et al. 2005, Matsuura et al. 2013)。この 5'UTR 配列は、決して同じタンパク質をコードする mRNA でも同様ではなく、転写開始点(Transcription Start Site: TSS)の違いや 5'UTR 内のスプライシングの有無によって様々なバリエーションが存在することが知られている。特に、近年の次世代シーケンサー技術の発達によって可能になったゲノムワイドな TSS の特定によって、複数の TSS に由来する 5'UTR バリエーションを持つ mRNA 種は非常に多いことが知られるようになってきている (Yamamoto et al. 2009, Morton et al. 2014)。酵母による研究では、このような TSS の違いに起因する 5'UTR バリエーション間で翻訳効率に大きな違いが認められ (Rojas-Duran and Gilbert 2012, Arribere and Gilbert 2013)、また、植物においても 5'UTR 配列の 5'末端のわずかな塩基の違いによって、熱ストレスに応答した翻訳状態変化の挙動が大きく変化することが報告されている (Matsuura et al. 2013)。これらの報告を踏まえると、5'UTR バリエーションも含めた正確な 5'UTR 配列の決定が翻訳機構を解析するために必要不可欠であると考えられる。TSS すなわち 5'UTR 配列をゲノムワイドに決定する手法の一つに Cap Analysis of Gene Expression (CAGE) がある (Takahashi et al. 2012, Murata et al. 2014)。本研究でも CAGE によって各 5'UTR 配列の決定及びその存在比率の解析を、mRNA の翻訳状態に関するゲノムワイドデータを取得している各条件について行った。

CAGE 解析には、シロイヌナズナ植物体の発芽 2 日目 (2day after germination: 2DAG)、21DAG の未展開葉 (young leaves)、展開葉 (mature leaves)、シロイヌナズナ培養細胞 T-87 の培養 3 日目コントロール (Control)、37°C 熱ストレス条件下 (37°C, 10 min)、NaCl 200 mM 塩ストレス条件下 (200 mM, 10 min) の細胞を用いた。各サンプルから抽出した Total RNA から CAGE ライブラリーを作製し、シーケンズ解析を行った。すべてのサンプルで 1000 万以上の Tag 配列が取得でき、良好なシーケンズ結果が得られた (Fig. 12A)。それらの取得した Tag 配列を、シロイヌナズナのゲノム情報として公開されている TAIR10 の情報に従ってマッピングを行った。各サンプルは独立した 2 つのサンプルを用いて行っており、2 つのサンプル間で共にゲノム上に Tag が存在する位置についてのみ Tag 数をカウントし、各ゲノム上の Tag 数を Tag per million (TPM) 値に変換した。各 Tag は TAIR10 に登録されている染色体コードのタンパク質をコードする各遺伝子について、ストランド方向が同じで TAIR10 に登録された TSS の上流 500 nt から CDS の AUG までの範囲に存在して

いるものを、その遺伝子の TSS を示すものとしてアノテーションした。なお、TAIR10 では、いくつかの遺伝子についてスプライスバリエントが登録されているが、TAIR10_representative_gene_models の情報から代表的なスプライスバリエントのみを用いている。マッピングの結果、94%以上の Tag がゲノム上にマッピングされていた (Fig. 12B)。ゲノム上に非特異的にマッピングされ除外されたことを示す Low MapQuality の Tag が 2DAG、young leaves、mature leaves で非常に多くなっていたが、Low MapQuality の Tag は、ほぼ全て rRNA の領域にマッピングされており、これらのサンプルでは Cap trapping 等での精製時に残留した rRNA が多かったと思われる。しかし、これらを除いた Tag の 75%以上はアノテーションが付けられており (Fig. 12C)、シーケンス及びマッピング結果に問題はないと考えられ、最終的に全てのサンプルで 400 万以上の Tag を解析に用いることができた。これらのデータ処理によって、各遺伝子のアノテーションがつけられた Tag が存在する各ゲノム上の位置は、その遺伝子の TSS を示し、その位置における TPM 値 (TPM_TSS) はその TSS 由来の mRNA 量を表している。そして各遺伝子の全 TSS における TPM 値を合計することで遺伝子レベルでの mRNA 量 (TPM_Sum) を算出した。独立した 2 つのサンプル間での TPM_Sum 値には高い相関 ($r = 0.98 \sim 0.99$) があり (Fig. 13)、再現性が確認されている。より信頼性の高いデータを使用するため、以降の解析では発現量が低い遺伝子と 2 反復間で発現量が大きく異なる遺伝子を除去したデータセット (Fig. 13 の緑色で示す) を用い、TPM_TSS 値と TPM_Sum 値は 2 反復間での平均値とした。

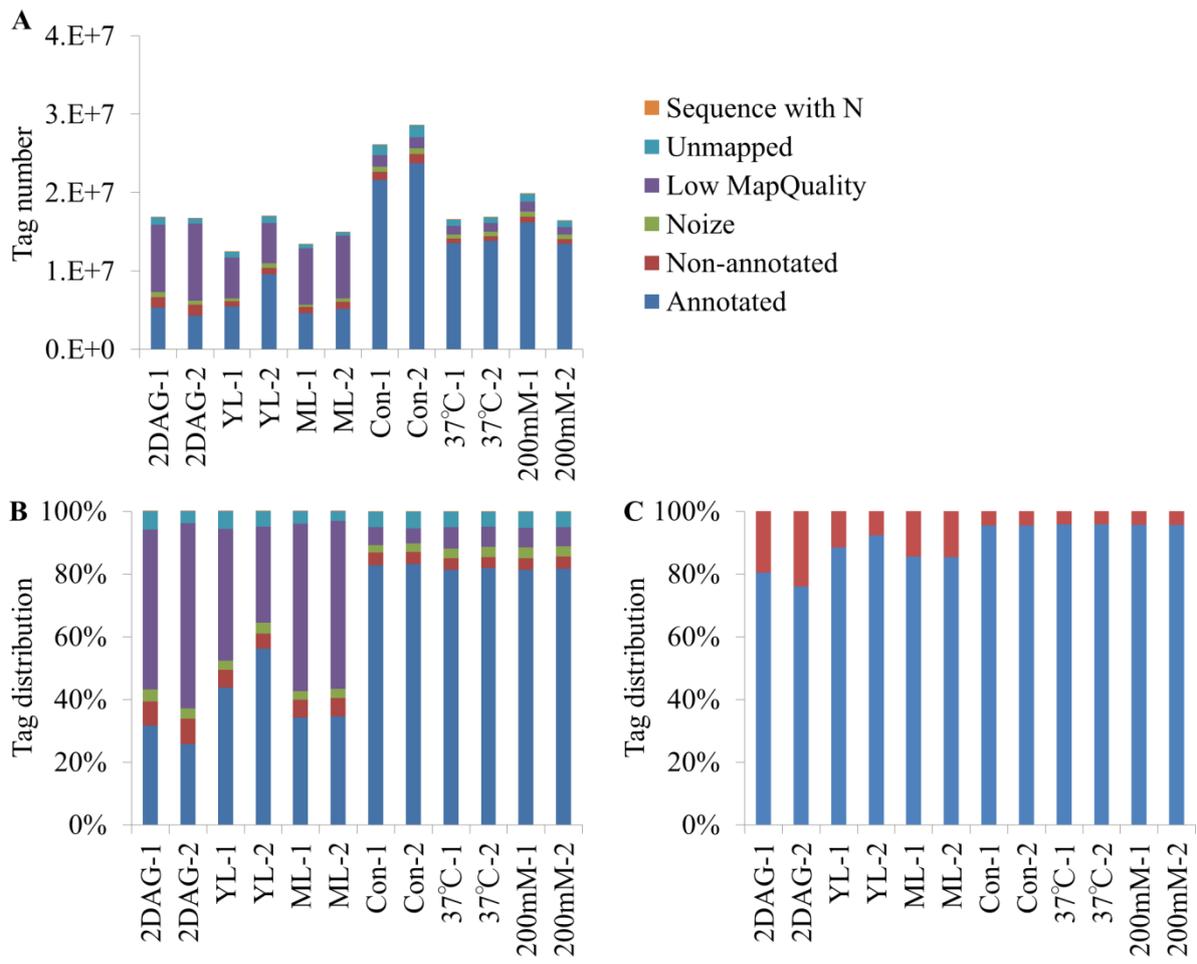


Figure 12. CAGE ライブラリーのシーケンス解析及びマッピングの結果

得られた全ての結果を (A) Tag 数、(B) Tag の比率で示している。"Sequence with N"には Tag 配列中に正確に読まれていないことを示す N が存在し、除外された Tag を示している。"Unmapped"は TAIR10 のゲノム上のどこにもマッピングされずに除外された Tag を示している。"Low MapQuality"はゲノム上の 2 か所以上に非特異的にマッピングされ除外された Tag を示している。"Noize"はゲノム上にマッピングはされたが、独立の 2 反復間のどちらかでは検出されずノイズとして除去された Tag を示している。"Non-annotated"は遺伝子のコード領域の上流以外にマッピングされた Tag を示し、本研究では使用していない。"Annotated"は遺伝子のコード領域の上流にマッピングされ、本研究で用いた Tag を示す。(C) ゲノム上に正しくマッピングされた"Non-annotated"と"Annotated"の Tag のみを比率で示している。解析は 2DAG (2DAG-1, 2DAG-2)、21DAG の未展開葉 (YL-1, YL-2)、展開葉 (ML-1, ML-2)、培養細胞 3 日目コントロール (Con-1, Con-2)、37°C 熱ストレス条件下 (37°C-1, 37°C-2)、NaCl 200 mM 塩ストレス条件下 (200_mM-1, 200_mM-2) について独立した 2 つのサンプルを用いている。

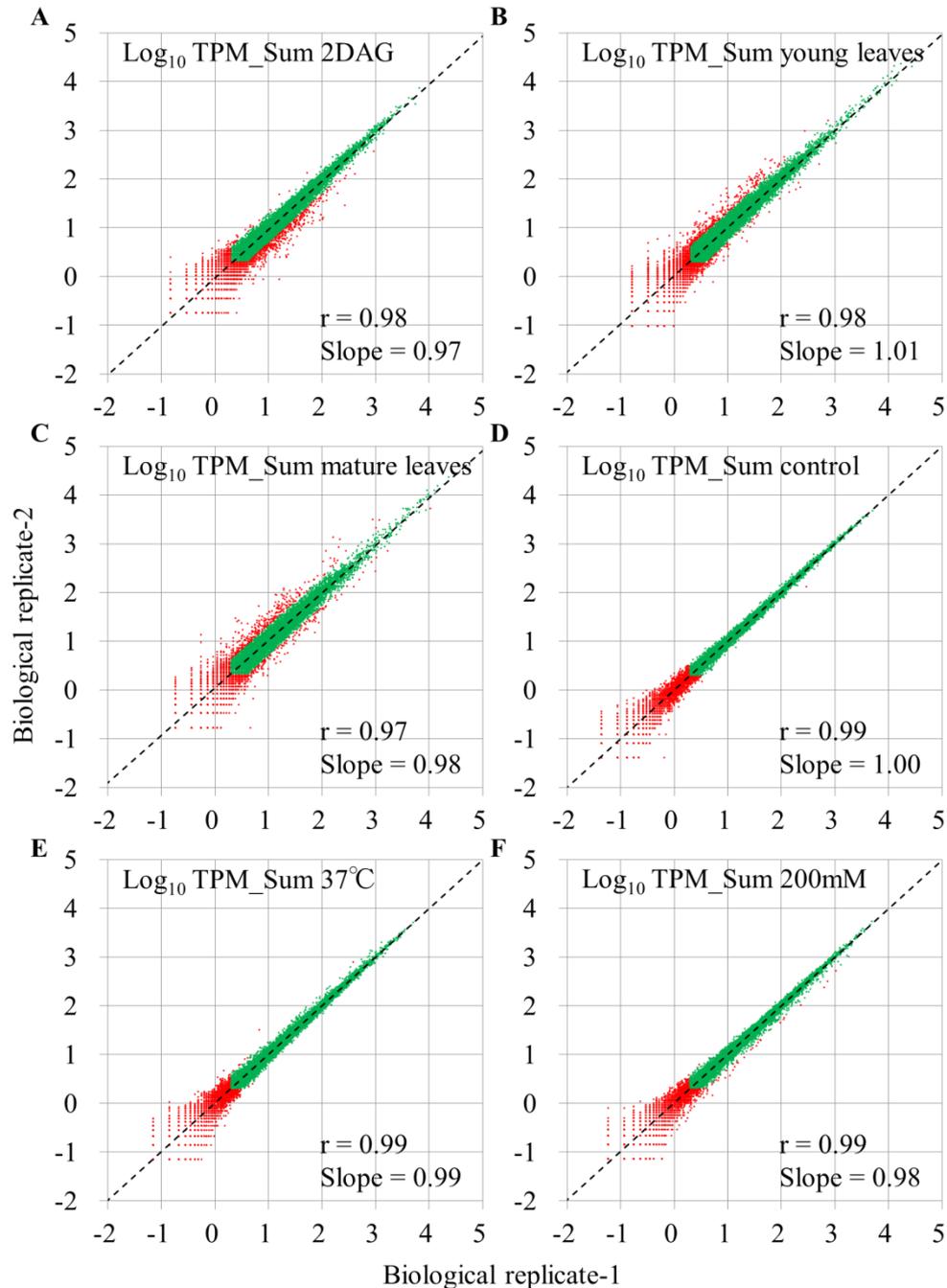


Figure 13. CAGE 結果の生物学的 2 反復における再現性

独立した生物学的 2 反復のサンプルより得られた Log_{10} -transformed TPM_Sum 値の散布図。それぞれの散布図は(A) 2DAG ($n = 18090$)、(B)未展開葉 ($n = 17083$)、(C)展開葉 ($n = 16300$)、(D) 培養細胞の通常条件 ($n = 17895$)、(E)熱ストレス下 ($n = 17208$)、(F) 塩ストレス下 ($n = 17509$) における遺伝子あたりの Tag 数を示す。近似線(点線)と近似式の傾き (Slope)、ピアソンの相関係数 (r) を図中に示した。また赤色で示した遺伝子は発現量が低いか、2 反復間で値が大きく異なるという理由で解析からは除外した。以降の解析では緑色で示した 13426 (A)、12746 (B)、11903 (C)、14391 (D)、13938 (E)、14309 (F)の遺伝子を対象としている。

2-3-2. 転写開始点の分散と変化

算出した TPM_TSS 値と TPM_Sum 値は、それぞれ各 TSS 由来の mRNA 量と、各遺伝子レベルでの mRNA 量を表している。各条件における発現量が低い遺伝子と 2 反復間で発現量が大きく異なる遺伝子を除去したデータの TPM_Sum 値は、正規様の分布を示した (Fig. 14A)。続いて、それぞれの遺伝子における TSS の分散を評価するため、各 TSS 由来の mRNA の分布率 (TSS 分布率) を算出し、各遺伝子における最大の TSS 分布率を、最も主要な TSS 由来の mRNA の存在比率 (Max Distribution Ratio of TSSs: MDR) とした (Fig. 14B)。MDR は 1 つの 5'UTR 配列で、その mRNA を表現できる比率を意味し、100 に近いほど TSS が 1 点に収束し他の TSS 由来のバリエーションが存在しないことを、0 に近いほど TSS が複数に分散し多くの TSS 由来のバリエーションが存在していることを示している。代表的な TSS 分布について未展開葉のデータを例に Fig. 15 に示した。その分布は、eukaryotic initiation factor 4G (eIF4G) のように非常に TSS が分散している遺伝子 (Fig. 15A) から Chlorophyll A/B binding protein (CAB1) のように TSS がほぼ 1 点に収束している遺伝子 (Fig. 15D) まで幅広く、その全条件における MDR の中央値は 30.2~32.0 であり (Fig. 14B)、多くの遺伝子の TSS は Ribosomal protein L27 (RPL27) のように分散し複数の 5'UTR バリエーションを持っていた (Fig. 15B)。また、Arabinogalactan protein 21 (AGP21) のように 50%以上の TSS が 1 点に収束している遺伝子 (Fig. 15C) は全体の 16~19%、CAB1 のように 75%以上が収束している遺伝子は全体の 2~4%しか存在していなかった。次に最も主要な TSS から AUG までの配列を TAIR10 のゲノム情報から抽出し、イントロンが存在する場合はイントロン部分を除去することで、各遺伝子由来の mRNA において最も主要な 5'UTR 配列を取得した。その 5'UTR 長の中央値は 83~93 nt であり、80%以上の 5'UTR は 200 nt 以下であった (Fig. 14C)。加えて、植物体である 2DAG、未展開葉、展開葉間と、培養細胞の Control、37°C、200_mM 間のそれぞれで TSS の分布を比較し、TSS 分布の変化を Δ TSS 値として評価した。この Δ TSS 値が 0 を示すと TSS の分布が 2 条件間で完全に一致し、1 を示すと TSS の分布が完全に異なることを意味している (Fig. 14D)。 Δ TSS 値の分布より、10 min という短期的なストレスである 37°C や 200_mM では、ほぼ全ての遺伝子で大きな変化は認められなかったが (Median = 0.12~0.13)、発達段階が異なる植物体間での比較では、TSS の変化が相対的に大きい傾向があった (Median = 0.21~0.28)。未展開葉と展開葉間の TSS 変化を例にすると、大部分の遺伝子における TSS は 3-ketoacyl-CoA synthase 10 (KCS10) のように顕著な変化は示さなかったが (Fig. 16A)、一部の遺伝子では Partner of SLD five 1 (PSF1) のように TSS が顕著に変化していた (Fig. 16B)。PSF1 の例では、未展開葉では主に AUG から -320~-260nt の位置に存在していた TSS が、展開葉では -410~-350nt の位置に変化している。このようにシロイヌナズナにおいても、多くの遺伝子には複数の TSS に由来する 5'UTR バリエーションが存在し、いくつかの遺伝子では条件によって TSS が変化していることが示された。

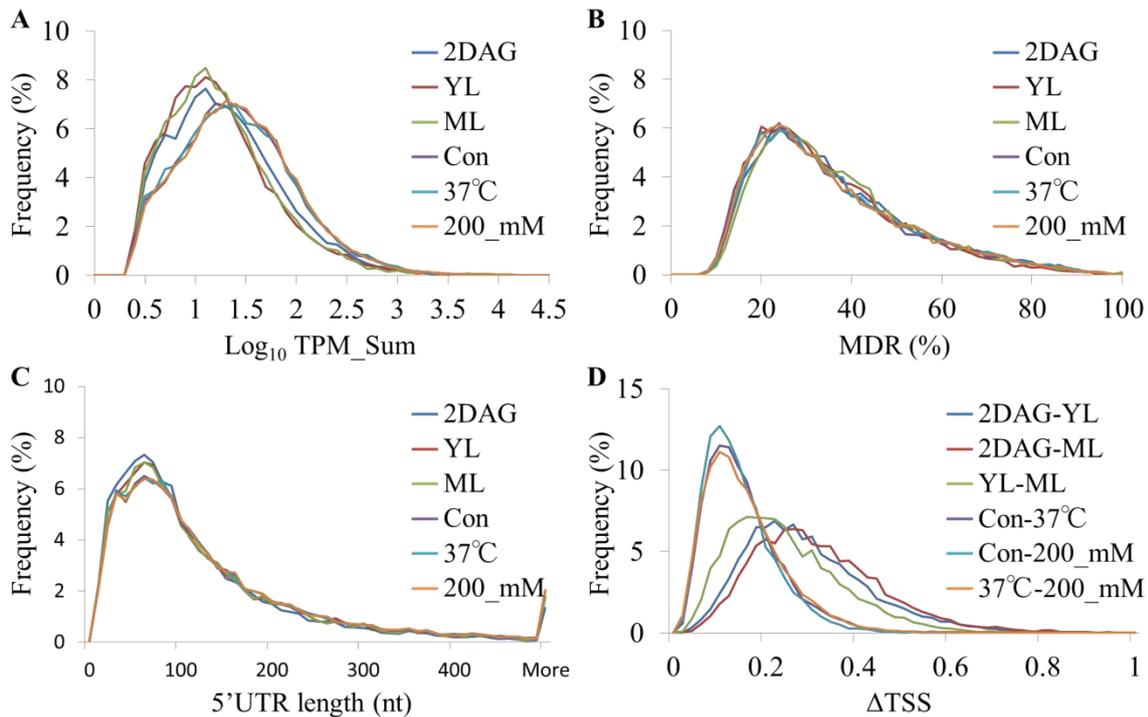


Figure 14. CAGE による TSS の解析結果の分布

(A) 2DAG、未展開葉 (YL)、展開葉 (ML)、Control (Con)、37°C、200_mM での遺伝子レベルの mRNA 量を示す TPM_Sum 値の分布、(B) 最も主要な TSS 由来の mRNA の存在比率を示す Max Distribution Ratio of TSSs (MDR) 値の分布、(C) 最も主要な TSS に由来する 5'UTR 長の分布、(D) 2 条件間の TSS の変化を示す Δ TSS の分布。(A-C)では Fig. 13 の緑色で示される 13426 (2DAG)、12746 (young leaves: YL)、11903 (mature leaves: ML)、14391 (Control: Con)、13938 (37°C)、14309 (200_mM) の遺伝子を、(D) では 2 条件間で共にデータが存在した 10986 (2DAG-YL)、10163 (2DAG-ML)、11089 (YL-ML)、13830 (Con-37°C)、14144 (Con-200_mM)、13783 (37°C-200_mM)の遺伝子を対象としている。

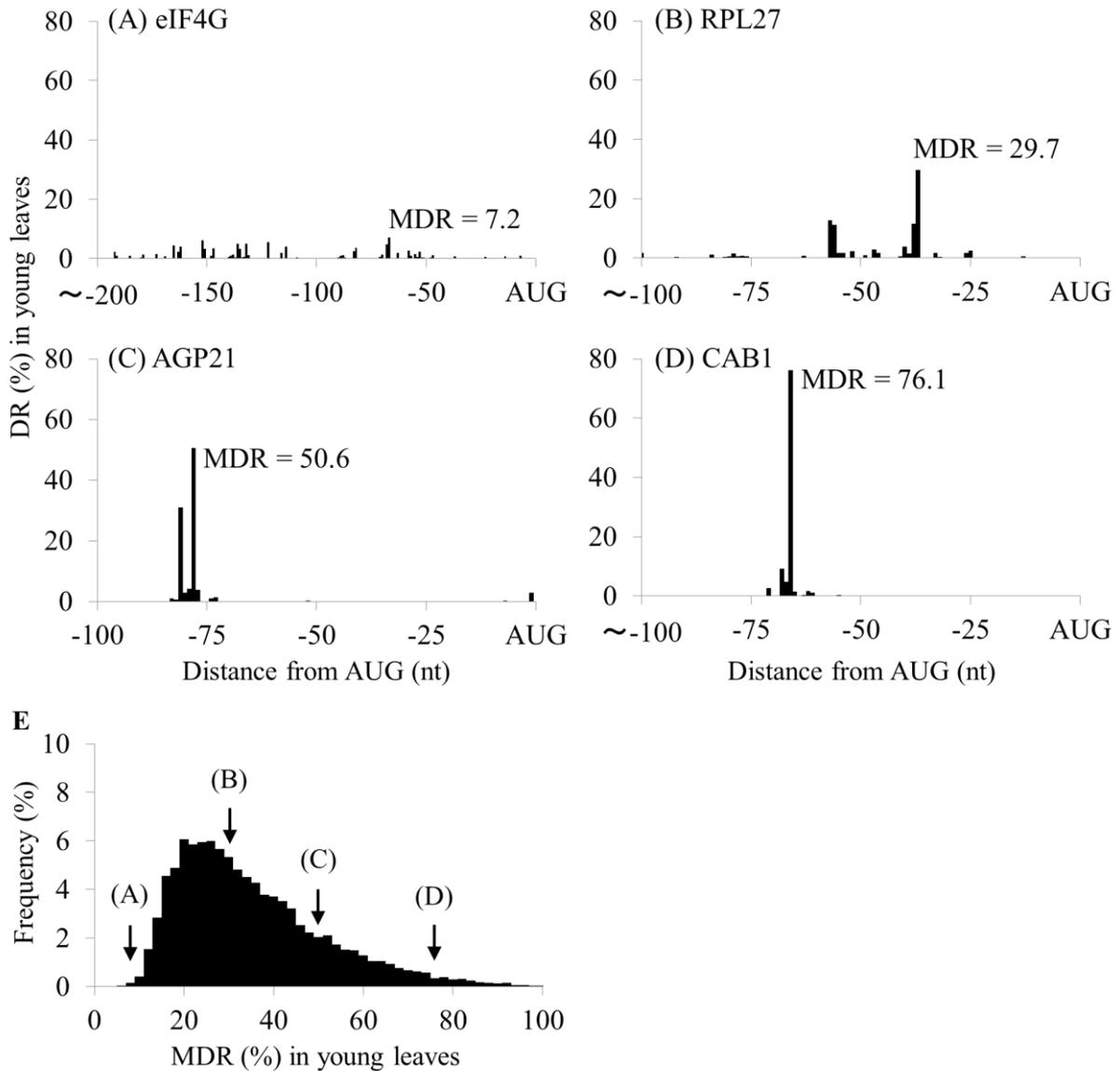


Figure 15. 未展開葉での TSS 分布の例

(A)eIF4G、(B)RPL27、(C)AGP21、(D)CAB1 遺伝子の未展開葉での TSS の分布。横軸は AUG からの距離を縦軸は Distribution ration of TSSs (DR) 値を示している。

(E) 未展開葉で最も主要な TSS 由来の mRNA の存在比率を示す Max Distribution Ratio of TSSs (MDR) 値の分布と、その分布における (A-D) の遺伝子の位置を矢印で示している。

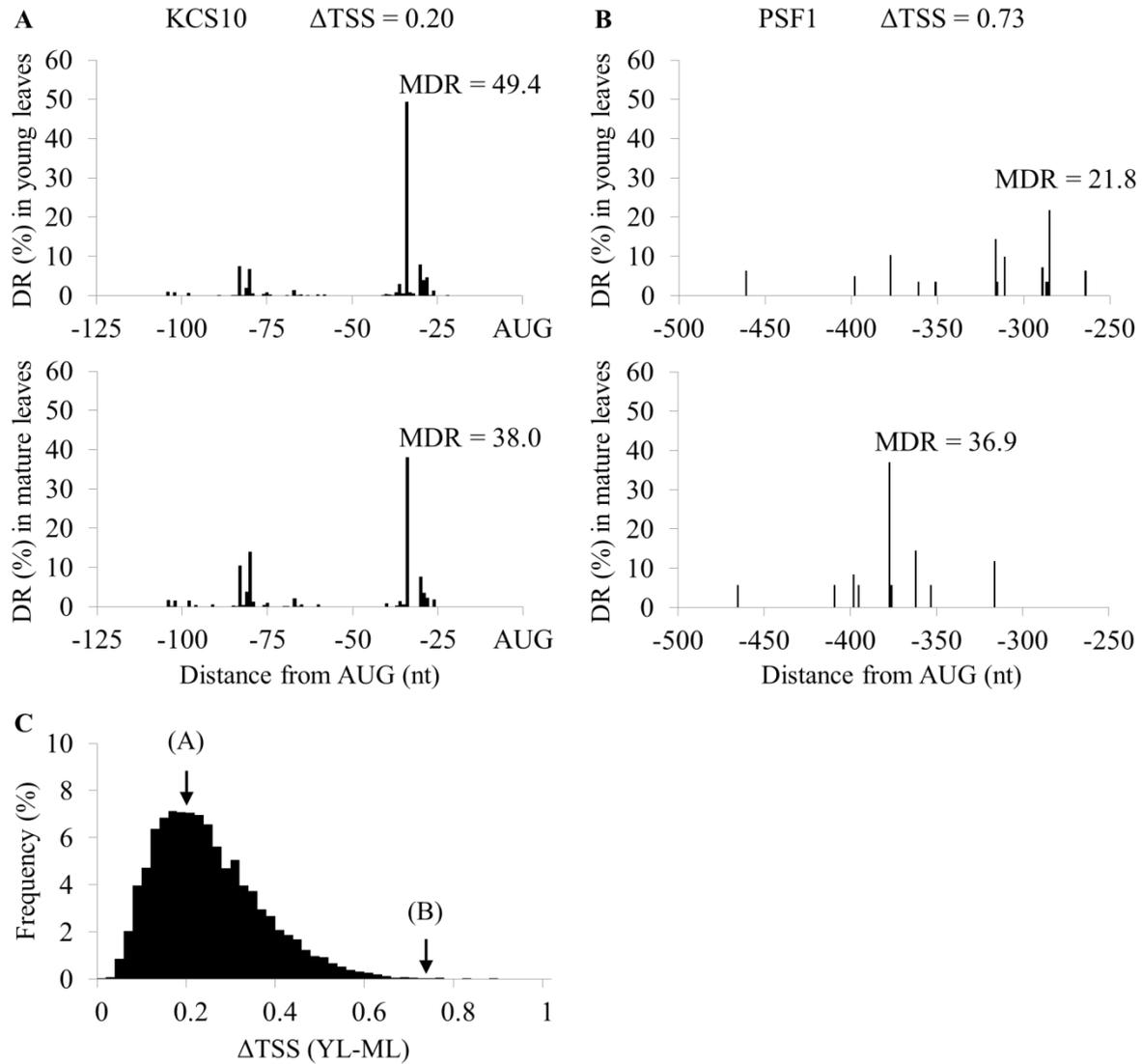


Figure 16. 未展開葉と展開葉の間での TSS 変化の例

(A)KCS10、(B)PSF1 遺伝子の未展開葉（上段）と展開葉（下段）での TSS の分布。横軸は AUG からの距離を縦軸は Distribution ration of TSSs（DR）値を示している。

(C) 未展開葉と展開葉の間での TSS 変化を示す ΔTSS 値の分布と、その分布における (A、B) の遺伝子の位置を矢印で示した。

2-3-3. 転写開始点前後の配列における塩基含量の解析

遺伝子によって、TSS は非常に分散しているものから 1 点に収束しているものまで様々であり、いくつかの遺伝子では TSS が成長・発達段階間で変化していた。どの位置から転写されるかは、その分散度合いも含めて、プロモーターの性質だけでなく転写された mRNA の質 (5'UTR の配列) に関わる重要な情報である。そこで、TSS 周辺の配列の塩基含量からその決定に関わる機構についても考察するために、各遺伝子の最も主要な TSS の上流 50 nt、下流 50 nt の ATGC の比率を算出した (Fig. 17)。なお、開始 AUG の出現による塩基の偏りを無視するために、対象とした遺伝子は最も主要な TSS 由来の 5'UTR の長さが 50 nt 以上のものとした。解析の結果、全ての場合において TSS から約-30 nt (0 nt は 5'UTR の 5'末端の塩基となる) の位置に TATA box と思われる A と T の集中が認められた。加えて TSS 前後数 nt にも顕著な塩基含量の偏りが認められ、-2 nt では T、-1 nt では T と C、0 nt では A と G、3 nt では C の集中が見られた。これらの配列は転写装置が転写を開始する上で重要な要素である可能性がある。加えて、培養細胞である Control、37°C、200_mM では 24~25 nt の位置に T の比率が高い傾向が認められた。

続いて、TSS 前後の塩基の偏りが TSS の分散に与える影響を評価するために、未展開葉を例にして TSS が収束している遺伝子 (MDR の上位 1000 遺伝子) と分散している遺伝子 (MDR の下位 1000 遺伝子) の塩基含量を比較した (Fig. 18)。その結果、TSS が収束している遺伝子では TSS から-30 nt 付近に位置する TATA box の存在がより明瞭であったのに対し、分散している遺伝子ではあまり顕著な偏りは認められなかった。また、収束している遺伝子では、TSS 前後の塩基の偏りも全 mRNA の場合と比較すると僅かながら増加していた。TATA box を有するプロモーターの場合は TSS が収束しており、TATA-less プロモーターでは TSS が分散する傾向があると考えられる。

また、TSS の変化に関しても未展開葉と展開葉を例に、TSS の変化が小さい遺伝子 (Δ TSS の下位 1000 遺伝子)、大きい遺伝子 (Δ TSS の上位 1000 遺伝子) について、それぞれでの未展開葉と展開葉での塩基含量の偏りを解析した (Fig. 19)。結果として TSS が変化しない遺伝子と大きく変化する遺伝子では TATA box の存在にはっきりとした違いがあり、TSS が変化する遺伝子では、TATA box の存在は不明瞭だった。TATA-less プロモーターの方が、より TSS が変化しやすい傾向と思われる。また、前後数 nt での塩基含量には大きな差異は認められなかった。

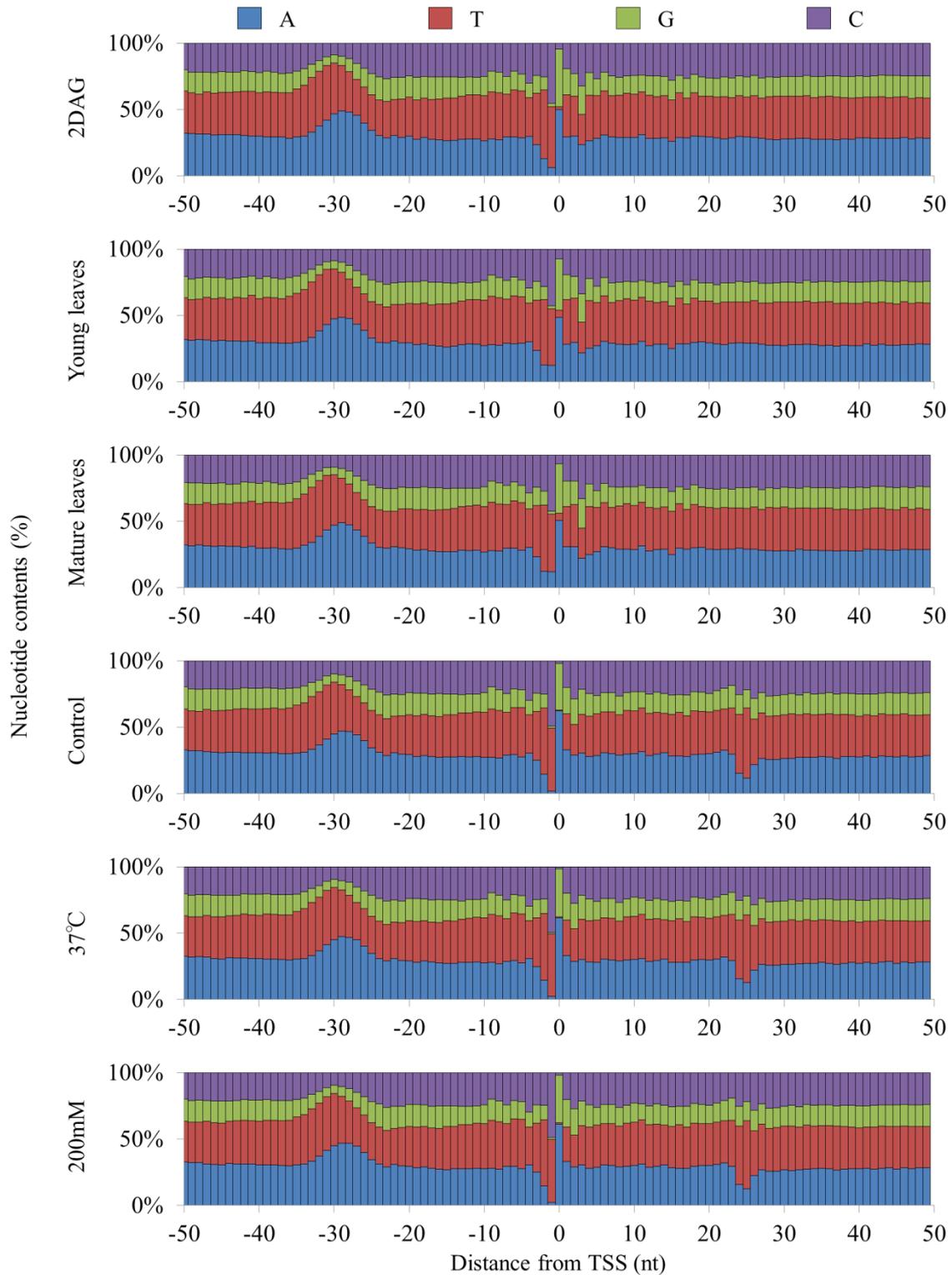


Figure 17. TSS 周辺の塩基含量

2DAG (n = 9804)、未展開葉 (n = 9533)、展開葉 (n = 8918)、Control (n = 10959)、37°C (n = 10959)、NaCl 200 mM (n = 10923) で最も主要な TSS の前後それぞれ 50 nt における ATGC 含量を示している。横軸は TSS からの距離を示し、0 は 5'UTR の 5' 末端に位置する。

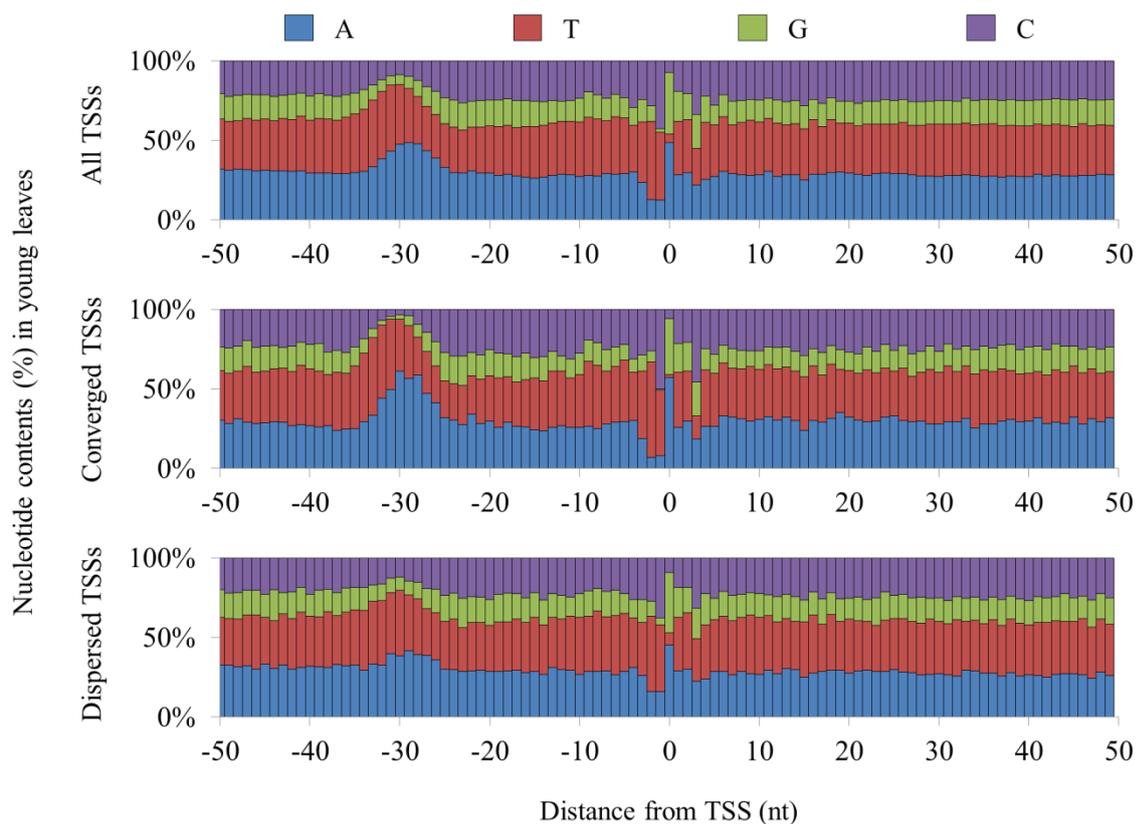


Figure 18. 収束または分散している遺伝子の TSS 周辺の塩基含量

未展開葉のデータセットの全遺伝子 (n = 9533)、TSS が収束している遺伝子 (n = 1000; MDR 値の上位)、分散している遺伝子 (n = 1000; MDR 値の下位) の最も主要な TSS の前後それぞれ 50 nt における ATGC 含量を示している。横軸は TSS からの距離を示し、0 は 5'UTR の 5'末端に位置する。

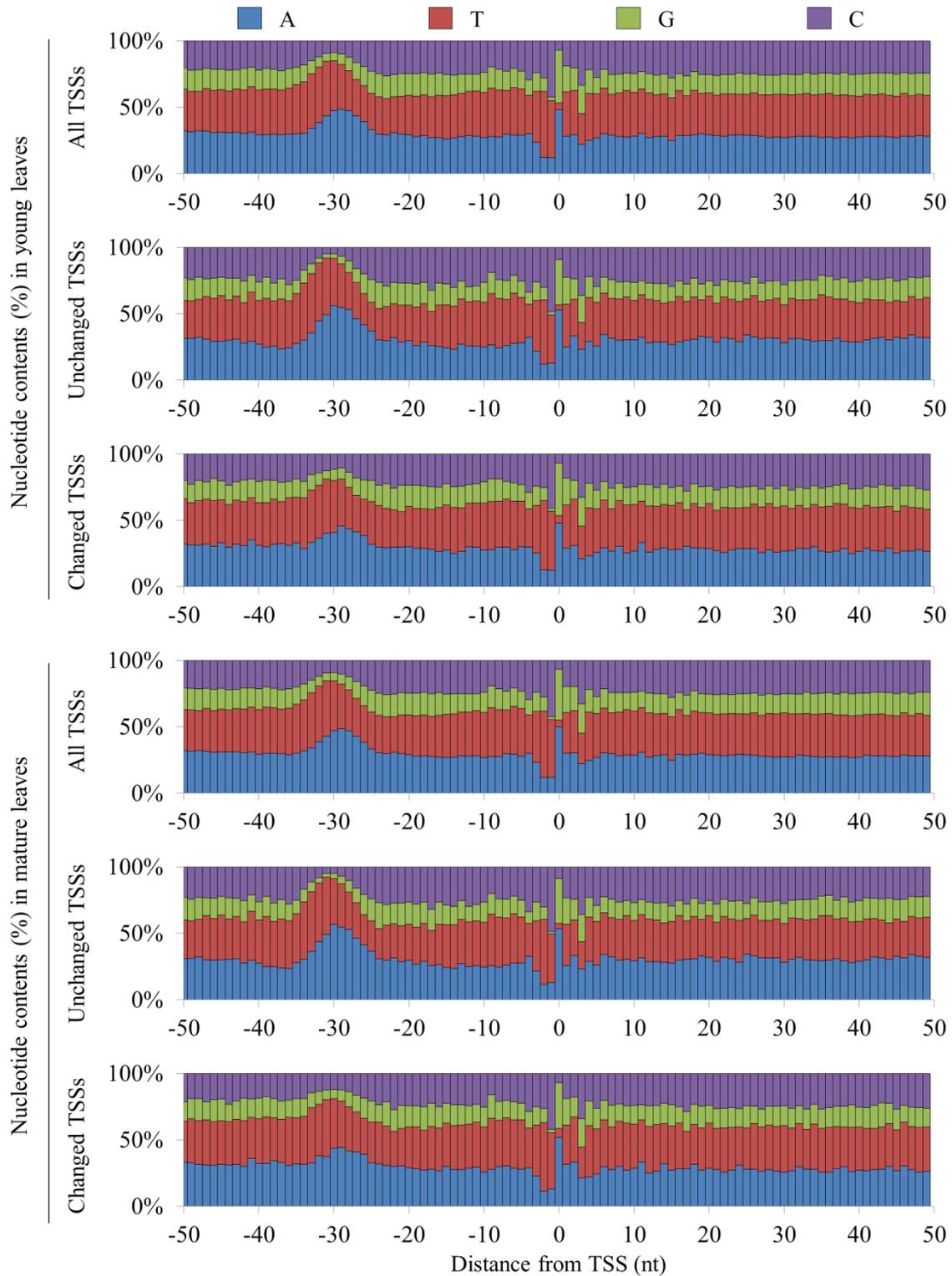


Figure 19. TSS が変化する遺伝子と変化しない遺伝子の TSS 周辺の塩基含量

未展開葉と展開葉で共にデータが存在する全遺伝子 (n = 8134)、TSS が変化しない遺伝子 (n = 1000; Δ TSS 値の下位)、変化する遺伝子 (n = 1000; Δ TSS 値の上位) の未展開葉 (上段) と展開葉 (下段) で最も主要な TSS の前後それぞれ 50 nt における各位置の ATGC 含量を示している。位置の 0 は 5'UTR の 5'末端に位置する。

2-3-4. 転写開始点の分散・変化と翻訳の関連性

CAGEによって多くの遺伝子の TSS は分散し、条件によって TSS が変化することが明らかとなり、5'UTR の多様性が明らかとなった。5'UTR 配列は翻訳状態を規定している非常に重要な要素として知られており、今回の解析によって示された TSS の分散や変化が翻訳状態の規定に影響を与えている可能性が考えられた。そこでまずは、未展開葉を例に第一章で算出した PR 値より、翻訳状態が低い mRNA 種 (PR 値の下位 500 種)、中間的な翻訳状態を示す mRNA 種 (PR 値の中央値付近の 500 種)、翻訳状態が高い mRNA 種 (PR 値の上位 500 種) に分け、TSS の分散度合いの指標値である MDR 値に偏りが見られるかを解析した (Fig. 20A)。その結果、mRNA 全体と比較して、それらの分布に顕著な偏りは認められず、全体的には TSS の分散度合いと翻訳状態は相関していなかった。しかし、僅かながら翻訳状態が低い mRNA 種と高い mRNA 種は、全 mRNA と比較して MDR が高い傾向があり、それらの mRNA 種は TSS が 1 点に収束している傾向がある可能性が示唆された。TSS が分散している場合、複数存在する 5'UTR バリエントが示す様々な翻訳状態が平均化されることで、顕著な挙動をとりにくい一方で、TSS が 1 点に収束している場合は、平均化されないためにその mRNA の翻訳状態が反映されやすい可能性がある。また、TSS の変化 (Δ TSS) と翻訳状態の変化 (Δ PR) を、未展開葉と展開葉を例に、翻訳状態が抑制された (展開葉で翻訳状態が低い) mRNA 種 (Δ PR_leaf の下位 500 種)、変化しなかった mRNA 種 (Δ PR_leaf が 0 付近の 500 種)、活性化された (未展開葉で翻訳状態が低い) mRNA 種 (Δ PR_leaf の上位 500 種) に分けて解析した (Fig. 20B)。その結果、翻訳状態が条件間で異なる (抑制または活性化された) mRNA 種は Δ TSS が低い傾向にあった。この結果は、未展開葉と展開葉との間での翻訳状態の変化は、TSS 変化によって生じた異なる 5'UTR 配列に起因したものではなく、未展開葉および展開葉それぞれで異なる翻訳機構が存在しているためと考えられる。

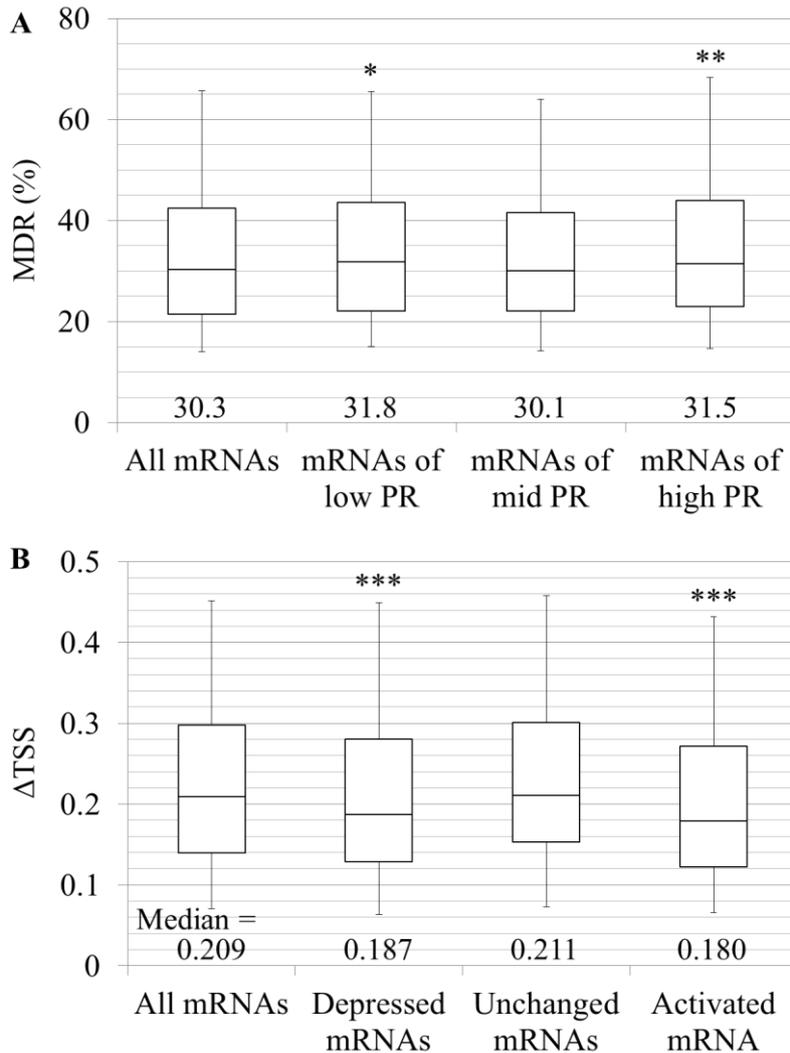


Figure 20. TSS の分散・変化と翻訳状態の関連性

(A) 未展開葉のデータが存在する全 mRNA 種 (n = 12258)、翻訳状態が低い mRNA 種 (n = 500; PR 値の下位)、中間的な翻訳状態を示す mRNA 種 (n = 500; PR 値の中央値付近)、翻訳状態が高い mRNA 種 (n = 500; PR 値の上位) での MDR 値の分布を箱ひげ図で示し、各分布の中央値を図中に記した。(B) 未展開葉と展開葉でデータがともに存在する全 mRNA 種 (n = 10299)、翻訳状態が抑制された mRNA 種 (n = 500; ΔPR_{leaf} 値の下位)、変化しなかった mRNA 種 (n = 500; ΔPR_{leaf} が 0 付近)、活性化された mRNA 種 (n = 500; ΔPR_{leaf} 値の上位) での ΔTSS 値の分布を箱ひげ図で示し、各分布の中央値を図中に表記した。エラーバーの末端は分布の上位 5% 位置と下位 5% 位置を示し、外れ値は省略している。加えて、全 mRNA 種の分布との差を Welch's t test によって評価した。* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2-3-5. 5'UTR の配列的特徴によって翻訳状態を説明できる PLS モデルの構築

第一章において、ゲノムワイドな翻訳状態の決定を行い、PR 値として翻訳状態を数値化した。その結果、翻訳状態は mRNA 種によって様々であり、加えて植物が置かれた状況に応じて翻訳状態は変化していた。これまでに mRNA の翻訳状態に影響を与えていると考えられる 5'UTR の配列的特徴がいくつか報告されている。しかし、それらの多くは翻訳状態と単一の特徴との間の相関解析に留まり、翻訳状態との間に緩やかな相関が確認されたに過ぎず、真に重要な特徴や各特徴の影響の大小は不明瞭なままだった。それらの問題を解決するために、数値化した翻訳状態と特定した 5'UTR 配列を基に *in silico* で複数の配列的特徴を用いた複合的な解析を行った。

複数のパラメータと目的の値との関係性を解析する（複数のパラメータで目的の値を説明する）にはいくつかの手法があるが、最も基本的な手法の一つに重回帰分析法があり、複数のパラメータを説明変数 (x)、目的の値を目的変数 (y) として計算的に説明変数の重み（目的変数に与える影響の大きさ）を評価するものである。本研究では、重回帰分析法に存在する多重共線性等の欠点を改良した Partial Least Squares Regression (PLS)法を用いた。目的変数には翻訳状態の指標値である PR 値の対数値を、説明変数には CAGE で特定した 5'UTR 配列を基に後述する様々な配列的特徴を用いた。PLS 法によって、配列的特徴（説明変数）で翻訳状態（目的変数）を説明する式（PLS モデル）の構築を行い、構築した PLS モデルからモデルの精度（ Q^2 値）に負の影響を与える説明変数を除いていくことで、真に重要な配列的特徴を選抜した。なお PLS モデルの構築を行う上で目的変数である PR 値と説明変数の基となる 5'UTR 配列は 1:1 で対合していることが望ましいが、多くの遺伝子は複数の 5'UTR バリエントを有する一方で、PR 値は各遺伝子でそれぞれ 1 個の値しか存在しない。しかし、CAGE の結果より、TSS が 1 点に収束し、5'UTR 配列が 1 種しか存在しない遺伝子は非常に限られるため正確な結果を導くことができない。この問題を解決するために、1:1 対合が必要な解析が含まれる予備的な説明変数の選抜のみを、TSS がほぼ 1 点に収束し 5'UTR 配列が 1 つしか存在しない遺伝子集団で行った。その後、予備的に選抜した説明変数を複数の 5'UTR バリエントに対して評価し、バリエント間で評価値をまとめ、複合 5'UTR 配列における説明変数を計算することで、TSS が 5'UTR バリエントをある程度有する遺伝子についても、PR 値と複合 5'UTR の説明変数を 1:1 対合させ、PLS モデルの構築と重要な説明変数（配列的特徴）の選抜を行った。モデルの構築は、一般的に目的変数の値幅が大きく、適度に分散し、必要な説明変数が少ないほど容易になる。未展開葉の PR 値は幅広い分布をとり、かつ他の条件での PR 値と比較的に良く相関し最も中間的もしくは一般的な翻訳状態を示している。そのため、他の条件と比べ特異的な制御を受ける遺伝子が少なく、必要な配列的特徴も単純化できると考えられた。そこで、まずは未展開葉での PLS モデルを構築した。以下でも未展開葉を例に説明をする。

予備的な説明変数の選抜の流れを Fig. 21 にまとめている。PR_young_leaves 値と 5'UTR 配列の両方の情報が存在している遺伝子は 12258 個あり、そのうち、TSS が

66.7%以上 1 点に収束している遺伝子は 555 個であった。この 555 遺伝子を、それぞれの解析に応じて更に選抜し、予備的な説明変数の選抜に用いた。予備的な説明変数の選抜は、以下の 3 つの独立した解析によって行った。5'UTR の 5'末端または AUG からの位置依存的な領域内での塩基含量に関する説明変数 (Region_5'、Region_AUG) の選抜、特定の塩基パターンに関する説明変数 (Motif) の選抜、その他の配列的特徴等に関する説明変数 (Other) の選抜である。Region_5'または Region_AUG の選抜では、PR 値が高い遺伝子から低い遺伝子まで適度に分散している 5'UTR 長が 50 nt 以上の遺伝子集団を用い、5'UTR の 5'末端または AUG から 50 nt の範囲内における全ての位置と長さの領域 (1~1、1~2・・・1~50、2~2、2~3・・・49~50、50~50) で塩基含量 (A、U、G、C または AA、AU・・・CC) を数え、その塩基含量から PR 値を説明する簡易的な PLS モデルを構築した。構築したモデルの精度を表す Q^2 値で領域を評価し、 Q^2 値が他に比べて高い領域を選抜した。 Q^2 値が高い領域は、その塩基含量で PR 値を説明できることを意味し、翻訳状態を規定している特徴の一つの可能性であると考えられる。Motif の選抜では、5'UTR 長が 50 nt 以上で PR 値が特に高い遺伝子、または特に低い遺伝子を用い、それらに豊富に含まれている塩基パターンを Multiple Expectation-maximization for Motif Elicitation (MEME) を用いて選抜した。Other の選抜では、PR 値が適度に分散している遺伝子集団において、5'UTR 全体の塩基含量に関する特徴、uORF に関連する特徴、二次構造に関連する特徴として mRNA の様々な領域内での $-\Delta G$ 、rRNA との相互作用に関連する特徴として mRNA の様々な領域内での 18S または 25S rRNA との相補な連続配列の最大長、5'UTR や CDS 等の mRNA 長に関連する特徴、転写に関連する特徴として TPM_Sum 値や MDR 値、合計 220 種の配列的特徴を評価し、それらを説明変数、PR 値を目的変数として予備的な PLS モデルを構築し、モデルの Q^2 値に負の影響を与えない配列的特徴を選抜した。

こうして予備的に選抜した説明変数を用いて解析の主となる PLS モデルの構築を行った。PLS モデルの構築の流れを Fig. 22 にまとめている。データセットには、分布率が 5% 以上のある程度は転写されている TSS 由来の各 5'UTR バリエントの mRNA の合計が、各遺伝子の mRNA の 80% 以上を構成する 5744 遺伝子を選び、PR 値が分散するように 1023 遺伝子を更に選抜し使用した。これらの遺伝子は平均して 5.2 個のある程度は転写されている TSS 由来の 5'UTR バリエントを持つ。これらの遺伝子の PR 値を目的変数、予備的な説明変数の選抜によって示された配列的特徴の特徴量を説明変数として PLS モデルの構築を行い、モデルの Q^2 値に負の影響を与えない説明変数を重要な配列的特徴として選抜した。この時、構築には 1023 遺伝子の半数である 512 遺伝子 (データセット A) のみを用いており、残りの 511 遺伝子 (データセット B) を用いて、構築した PLS モデルから予測値を算出し、その予測精度から信頼性を評価した。その後、構築したモデルにおける残差 (説明しきれていない PR 値) を計算し、残差が小さくなるように Region_5'、Region_AUG、Motif、Other の再選抜を行った。予備的な説明変数の再選抜の流れを Fig. 23 にまとめている。

る。再選抜は、5 回連続で再選抜の結果による信頼性の低下を確認した時点で終了し、それまでの再選抜の中で最も信頼性が高いモデルを採用した。こうして構築したモデルには、計算上はモデルの精度に正の影響を与える重要な配列的特徴のみが説明変数として用いられている。しかし、そこに偽陽性が存在する可能性も十分に考えられる。そこで、採用したモデルに関しては、データセット B を用いた予測精度による信頼性の評価の際に、その予測精度に負の影響を与えている説明変数を除去した。除去後に各説明変数の重みを再計算し、最終的なモデルとした。

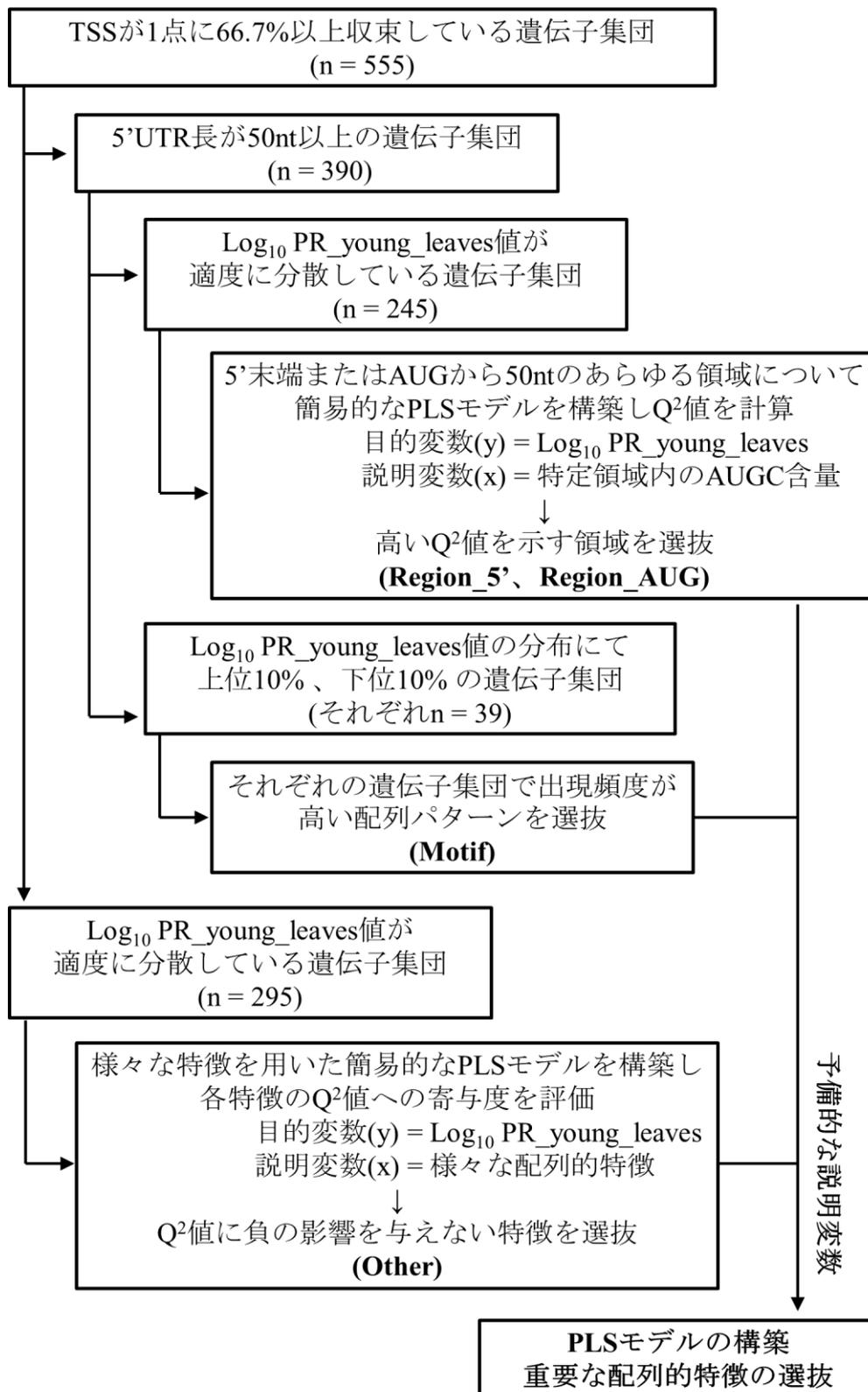


Figure 21. 予備的な説明変数の選抜の流れを示した模式図

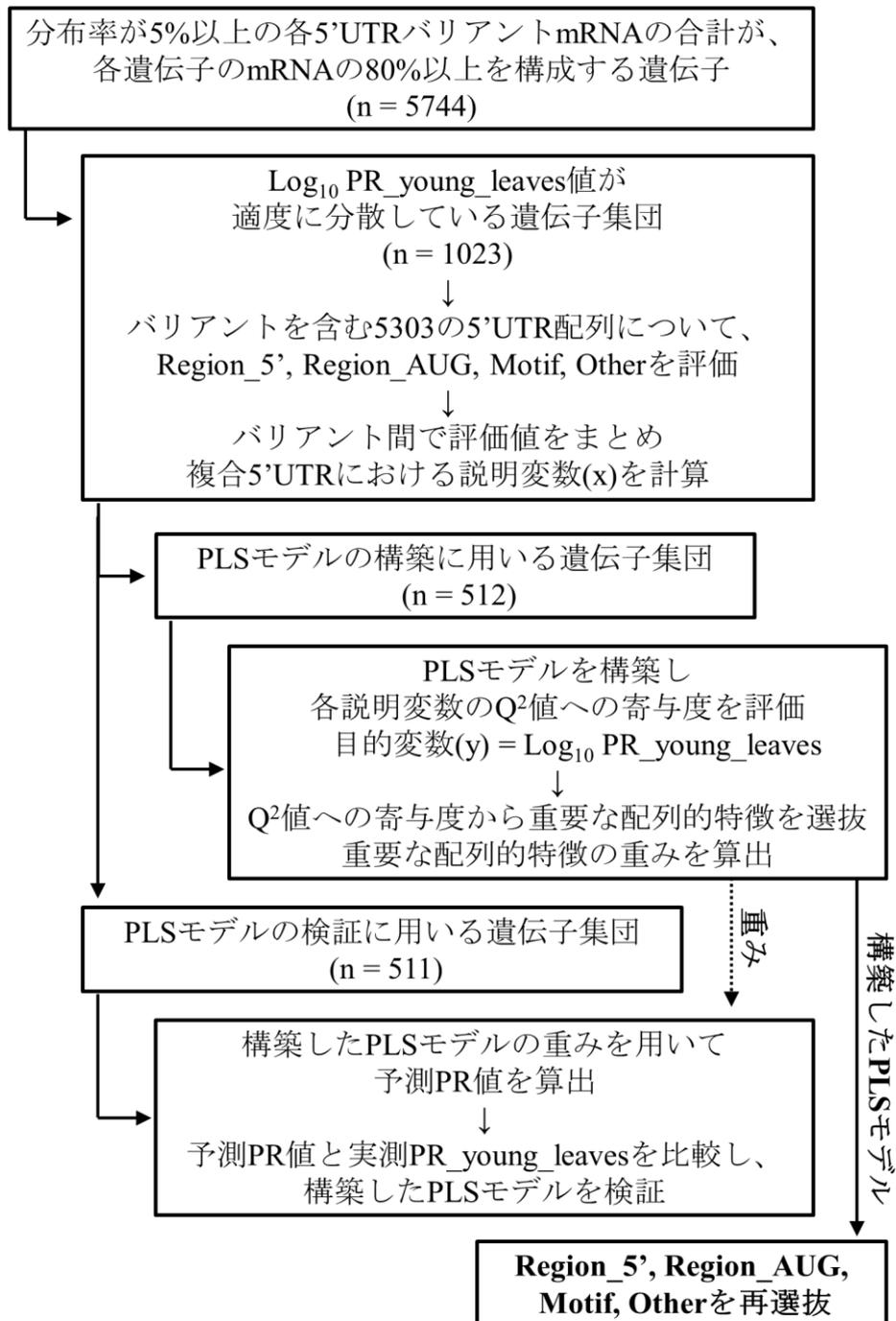


Figure 22. PLS モデルの構築と重要な配列的特徴の選抜の流れを示した模式図

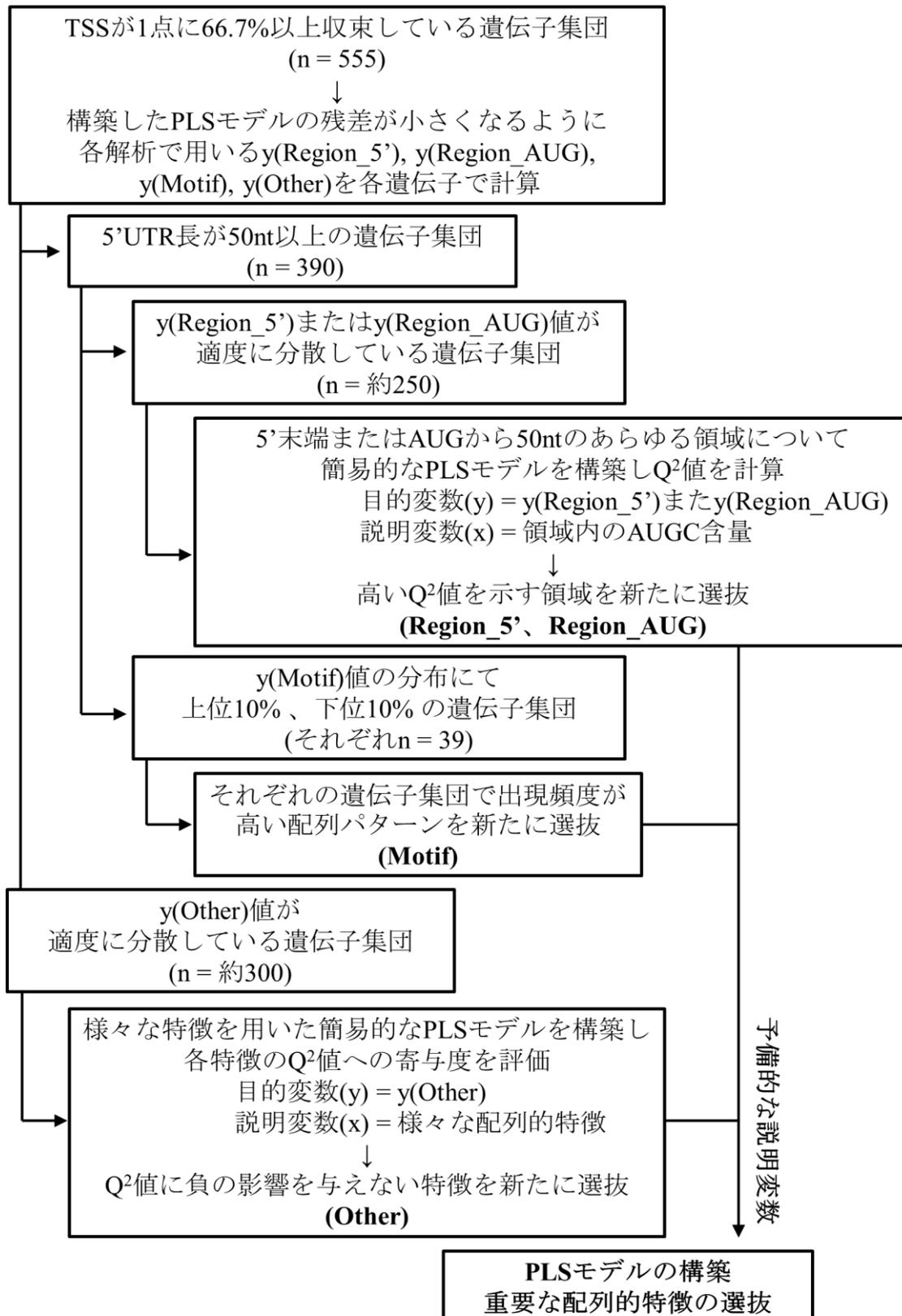


Figure 23. 予備的な説明変数の再選抜の流れを示した模式図

2-3-6. 未展開葉での PLS モデルから示された重要な配列的特徴

未展開葉の Log_{10} PR_young_leaves (目的変数) を配列的特徴 (説明変数) によって説明できる PLS モデルの構築では、最も翻訳状態をよく説明できた予備的な説明変数の再選抜 5 回目のものを最終的なモデルとした。構築の結果、PLS モデルの構築に用いた 512 遺伝子 (データセット A) において高い Q^2 値が示され、実測 Log_{10} PR_young_leaves とモデルによる予測 Log_{10} PR_young_leaves はよく相関していた (Fig. 24A)。モデルの検証に用いた 511 遺伝子 (データセット B) でも実測値と予測値の間で高い相関が見られ (Fig. 24B)、構築したモデルは PR 値を十分に説明できるものであった。このモデルでは、9 個の重要な配列的特徴が示され (Table 15)、それらは 5'UTR 上の二次構造 (3 個) と、uORF (1 個)、5'UTR 上の特定領域における塩基含量 (3 個)、25S rRNA との相互作用 (1 個)、CDS 長 (1 個) に関連するものだった。特に翻訳状態に負の影響を与えている配列的特徴としては、mRNA の Cap 側 100 nt 内の二次構造の形成度合い ($-\Delta G_{1\sim 100_from_Cap}$) と uORF の数 (uORF)、5'UTR の AUG から -150~-50 の領域内の二次構造の形成度合い ($-\Delta G_{-150\sim -51_from_AUG}$) だった。逆に特に翻訳状態に正の影響を与えている配列的特徴としては、CDS 長 ($\text{Log}_{10_CDS_Length}$) と 5'UTR の Cap 側 4 nt 内の A の含量 ($A_{1\sim 4_from_Cap}$) だった。CDS 長が非常に強い正の影響を与えているが、PR 値はリボソームの結合数で評価しているため、同じリボソームのリクルート効率でも長い mRNA の方がリボソームの結合数は多くなり、PR 値は高い傾向となるためだと考えられる。また 80% 以上の 5'UTR の長さは、200 nt 以下であるため (Fig. 14C)、 $-\Delta G$ が強い負の影響を与えている mRNA の Cap 側 100 nt 内の領域と、AUG から -150~-50 の領域は一部の mRNA では重複していると考えられる。また、 $-\Delta G$ が強い負の影響を与えている Cap 側 300 nt の領域は多くの遺伝子で AUG 以降も含まれている。

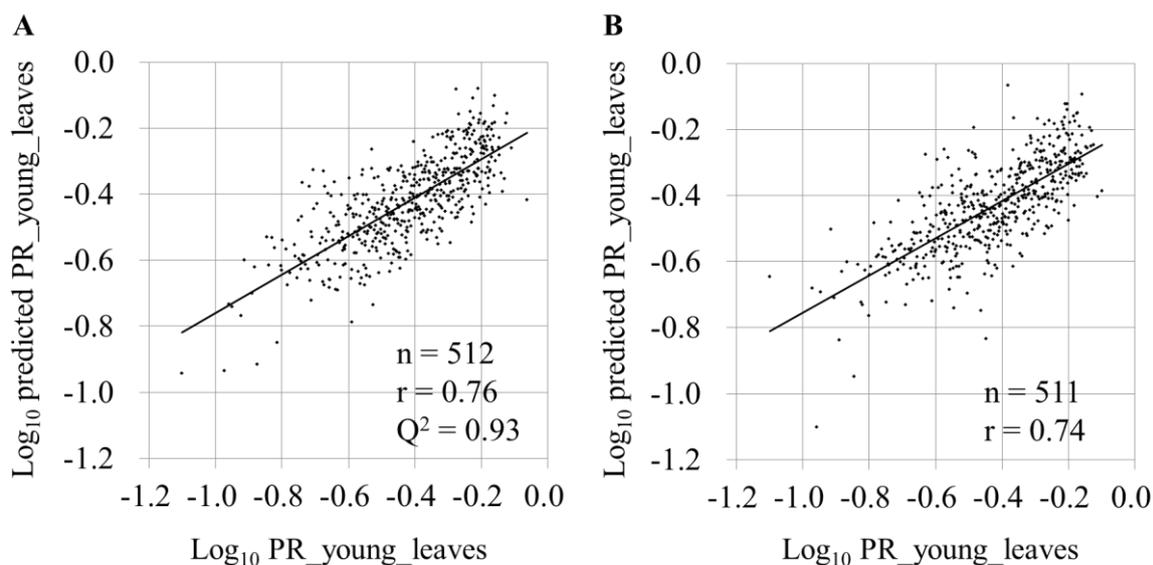


Figure 24. 未展開葉の PLS モデルによる予測 PR 値と実測 PR 値の相関

(A) PLS モデルの構築に用いた 512 遺伝子での実測 Log_{10} PR_young_leaves と予測 Log_{10} PR_young_leaves の散布図。データ数 (n) と近似線 (実線)、相関係数 (r)、モデルの精度 (Q^2) を図中に示す。(B) PLS モデルの構築に用いた 512 遺伝子とは異なる 511 遺伝子での実測 Log_{10} PR_young_leaves と予測 Log_{10} PR_young_leaves の散布図。データ数 (n) と近似線 (実線)、相関係数 (r) を図中に示す。

Explanatory variable	Weight
$-\Delta G_{1\sim 100_from_Cap}$	-0.050
uORF	-0.033
$-\Delta G_{-150\sim -51_from_AUG}$	-0.020
$-\Delta G_{1\sim 300_from_Cap}$	-0.015
C _{1~4} _from_Cap	-0.012
GC _{-23~-5} _from_AUG	-0.008
25SrRNA_Duplex_Length _{0~25} _from_AUG	0.012
A _{1~4} _from_Cap	0.016
$\text{Log}_{10_CDS_Length}$	0.111

Table 15. 未展開葉の PLS モデルで示された重要な配列的特徴とその重み

最終的な PLS モデルに含まれている説明変数 (重要な配列的特徴) と重み (PLS モデルへ与える影響) の一覧。重みが負の値を取るほど Log_{10} PR_young_leaves に負の影響を与えていることを、正の値を取るほど正の影響を与えていることを意味している。“ $-\Delta G$ ”は二次構造の形成度合いを、“uORF”は uORF の数を、“C, GC, A”は塩基の比率を意味し、特定の領域内でのみ評価を行った場合は、基準となった Cap または AUG からの位置を合わせて表記している。

2-3-7. 一過性発現実験による未展開葉に関する PLS モデルの検証

構築した未展開葉の PLS モデルは、実際の PR 値を十分に予測することができ、そのモデルから未展開葉の翻訳状態の決定に重要である配列的特徴が示された。このモデルの信頼性を更に確かなものとし、加えて翻訳状態（リボソームローディング）だけではなく、翻訳効率（実際にタンパク質ができる効率）も説明することが可能であることを確認するために、一過性発現実験から得られたデータを用いた検証を行った。また、一過性発現実験では同じレポーター遺伝子を用いることで、PR 値に大きな影響を及ぼす CDS 長に関する説明変数以外を重点的に評価できる。当研究室ではこれまでに、39 遺伝子の 5'UTR (Table 16) をレポーター遺伝子 (*F-luc*) に連結し、*in vitro* 合成した RNA を用いた RNA 一過性発現実験を行っており、様々な 5'UTR 配列による翻訳効率の違いが評価されている (Fig. 25A)。この一過性発現実験では、シロイヌナズナ T-87 培養細胞の培養 3 日目（ポリソーム/マイクロアレイ結果の 22°C、CAGE の Control の細胞とほぼ同一）を用いている。今回 PLS モデルを構築した未展開葉とは条件が異なるが、培養 3 日目の翻訳状態は、未展開葉の翻訳状態と良く相関し、その相関係数は 0.63 である (Table 6)。また、各 5'UTR に依存した翻訳効率の違いは、レポーター遺伝子である *F-luc* の活性値を供導入した *R-luc* の活性値によって導入効率を補正した F/R 値として評価されている。なお、PR 値とリボソームの結合数は指数近似するが、リボソームの結合数と翻訳能力は線形近似するため、PR 値に合わせ 39 種の 5'UTR に関して得られた F/R 値を対数変換して示した (Fig. 12A)。39 種の 5'UTR の翻訳効率を PLS モデルから予測した結果 (Fig. 25B)、実測された翻訳効率 (\log_{10} F/R 値) と予測 PR 値はよく相関しており (Fig. 12C, $r = 0.77$)、今回構築した PLS モデルの信頼性を確認するとともに、実際にタンパク質を合成する効率である翻訳効率も十分予測できるものであった。一方で、RNA 一過性発現実験では、転写の影響を無視して翻訳効率を評価することができるが、RNA を *in vitro* 合成する実験上 5'末端に必ず GG 配列が付加されてしまう。今回の未展開葉の PLS モデルでは 5'末端 4 nt 中の A と C の含量といった末端の配列も重要な配列的特徴として示されており、それらの影響を評価するには不十分な点があった。

そこで、DNA 一過性発現実験により更なる検証を行うことにした。DNA 一過性発現実験では恒常的に発現し、TSS が 1 点に収束することが知られている CaMV35S プロモーターを使用し、その TSS とレポーターである *F-luc* 遺伝子の間に CAGE データより選抜した 5'UTR を挿入し、導入した 5'UTR の翻訳効率を F/R 値として評価した。7 種の遺伝子由来の、異なる 5'UTR バリエーションを含む 10 種の 5'UTR (Table 17) の翻訳効率を評価し (Fig. 26A)、PLS モデルによる予測値と実測した F/R 値を比較した結果 (Fig. 26B)、非常によく相関した (Fig. 26C, $r = 0.84$)。これらの結果によって、今回構築した PLS モデルが実際の翻訳効率を十分に説明することが可能であるということが確認された。

Gene	5'UTR sequence
At1g01430	GGGCTGTGTCTTTTTTATAATGATGAAGTTCCTCGCTTCAACGCCAACCAAAAACCTCCAGAAAGCT AATTAGGGAACATTTGAAAAAACTCAGTGATCTGCTCTTCTCTAATTGCA
At1g06760	GGCTTCACAATCCTCATAATCACTTTCGAAAATTACATTTACGCTTCTTGCAATCAAATTTCCGATCTT AAGTTCAGAAGACG
At1g07230	GGAGATCCAAAGAGGACTTTTTTCAGCTTCATTGCTCTCTTAAAATCTGTGGTAATCTCTCTCCATTG TTATTCTTCAAGTAAATAATGCCCTTGTCTCTCTCAACACCATTATCTTATAAACCTAGACACTC CGCCGAATTTTACCCTTCTCCACTGGTCATCCGCGTTTTTTTTCCGGGTCTAAGGTCACCGGAAAA
At1g09070	GGATCATAAAAAACATTCTCAGAGCAAATTTCAAGTTTATTTTCATTCAATTCGATCCGAGTAACC
At1g20110	GGATCTTCGTCATCGTTTCCAATTTCCGATATAATCTGATCAAATCATCGCCGACGATCACCGATTGAG CG
At1g27730	GGAGCAATTAACAATTTCTTCACTGCAATTCACAAGCAACCTTCAAACCTAAAACCTCGAGAGACAAGA AATCCTCAGAATCTTTAACTTA
At1g47330	GGAAAACATCAAAGAGTCACTCTAAACTCATCTCTCGCCGAATTCTCCCAACAATTTCCGCCGGA AACAAATTTCTCAGATTCCGGTAACTCTAAAACCT
At1g55160	GGTTGAAGAAGAAGAGTAAAAGAAGAAACCTGGGAAGAAGTGAGAACTCATCTGATCGCC
At1g55330	GGATCATCACAAACAAAATCAAACAAGAATAACAAAATCTTTCTTATAAATTTTATTTCAGAGAC ATCAAAGGAGAATTA
At1g56580	GGTAAGAGACAGAGAGATCTTAACACAAAACAAAGCAAACACCAAAAAAACAGAG
At1g69490	GGATTTCAAGAACCTTGAGAACATCAAAAACTAACACAGAAAGAAAAAAAACAGTTCCTGTTCTATT AGATTGTTTTCTAAATTTGTCTGAAAATC
At1g77120	GGTACATCACAATCACACAAAACCTAACAAAAGATCAAAGCAAGTTCTTCACTGTTGATA
At1g77760	GGACCAAAAAATTTCAAAAACCAAAACCAAAAAAATAGCTTATCTCTCTTATCACCGGTTCT AATCACTAAAAC
At1g78080	GGATCGGTGAGGTTGAGAGTAATTCACTACACACACAAAAAATAAATGAGTGCCTCCCCAAAA ACAAAATTTGGTAGATAACGAGCAATTTGTTTTTTTCAGATTTGATCCTGAATTTTACATTTTTTTTTT GCAATCTCCCCCTAATCTGTTGTTTCTCGCTTCTTCTGTTAATCATCTGTCTTTCAAAAAGAAAGAA AAAAGAAAAATTCGATTTCTGGGTTGTTTTGTGCATACAGAAAAAAA
At2g15290	GGAAAATAAACTCTCCACTAAAAAAT
At2g21790	GGCTAGAGATTTTGAAACCGAAACCCAAAAACCCCTTTGACGCCTCCTTCTTATCTCTTTATAAA AAACCATTTCTTTCTGCAACATCGTTGCTTATCATCAGACGCACATCACCTGTCGATAAAAATTCCTC TGAGAGTGTTTTTTTGTTTTCTTCTGACAAAGAAAT
At2g30250	GGAGTTTTATCCTCTTTTTTCAATTCAAGATCAATTTCAAAAAAGTTTCTTTTTTAGCGAATCCTCTG TTTTAGAAATCTTAAAGTTGTCTCCTTTATTAAGACCATC
At2g34050	GGAGGGTTCTAAATCTGAGATTTCCAGAGAACTGTGCATATCTTCATAGTTTCTTTGAATTTACCGTA TTCATACCATAAAA

At2g39700	GGACCATCCTCATCTCTCAATCAACAAACCTAACTCTTTCTCTATACAAGAGCGCAAGCTCGAAGGCG CTCTTGTTCTTTCTTCATTTTCTCCTTAACAATCACCCATAGTAACTTTAAAACCTCTCTGTTTCTAGT AATTCACACACAAA
At2g41630	GGCCTTCTCCACATAAGTTACATCTCTCGTGTGTTTGTGTTTCTTTGTCTCCGATTTTTTCGCGACGAAGA AGAAGACGAGAGATAGAGAGAGAAGTAGAGAAATCGAAGGAATCTGTAACCGATTTTAAGATCTCAA TTTTTAGGGTTGTTGATTTTTCAATTTCTGGGTTAATTTTTTTAGGGTTTTTCATTGGAATC
At2g46390	GGGAAAACGAAAAGTCCGAAGAAGAAGAAGAAGAAGGAGAAAAG
At3g11120	GGGAATTGGCGCCGCTCATTTCCTGAAATTTGCATAGAATCAGAGGAGTGAAGAGTTTATCTTCTG TGAATTCACC
At3g15450	GGATAACACATTTCAAGCATTGGATTAATCAAAGACAAAGAAAACGAAA
At3g46620	GGAACTCAGATCCTTACCAGCAACACAAGTTATTTCAAAAAG
At3g47610	GGCCTTCTTGTCGCTGTTTCGAAGAGACTAAAGGCGACGGAGAGAATCGGAGAAGAAG
At3g51860	GGCAGATGTAGTAGAATCAAAAACGTCTTAAAAC
At4g02820	GGAAACTAAACAAA
At4g12000	GGATTAACAAACAAACCGAAAAAAGAAAAAACTCATCTTTCTCCAAAATCACACAAATCTTCTTTAT TTGTTATTCTCAATCCTTCCTTCATCCCCAGGTTCTTTTCGATTGTTGAGTCATTCAATTTTTCCATCAC TGGGTTTTTCTCTCTGAATCCGATCGGAGAATCCAGTCGATTACTAATCTAGCGCTCTCTTTTTTTCTAC TCG
At4g14560	GGACACAAGCATTTCAGGATATCAAATCACAATCCCAAGAAGAGCAATAACAAGAGAAGAAGAA GTAGTTCAAGAATTAAGGAAGAGAGCTTCTCCGTTAAAGTATAGTGAGAGAAT
At4g15000	GGACTTAGGGTTCATAGCAGCCAGAGAGAGAGACAAGTGAGAGGGATCTACCAAACGAAGCAACA
At4g18430	GGCAACATCTCTGAAATCTTCTTTACGATTTTGTAACCTTCTATCAAATACCTTAATA
At4g32060	GGGTCAAAAAGAAATCTGCGAAAAATAAGAGAGAGAGAGAGAGAAAACAGGATCATTCAAT
At5g03230	GGAACTCTCAAATTCACAAAAACCTCTCAGATTTTTTCTACAAGCTGATCAAAACCCAGAATCCAGT TTTCATAAAAACATCATCTTTCTTTATTCTTCTTTACACGCTCTGTTTTTCATCA
At5g08650	GGAGTGCC
At5g09880	GGATCTCATTAGCTTCTTCCAATTCGAAACCCTAACAAAAGCCCTAATTTACGAAATACGCCGAATCG TTGACGAGA
At5g11670	GGATCATCATTCCACATAAACAAAATATCTTCTGCTCATTATTCATCATCAATCTACTACTCCGTAAA CTTTGCTTCCTCTGTTTGTATTAGCTCAGAAGAGAT
At5g39740	GGGCGCCTTTGCCATCTTTGTCATCATCTTCACAGGAAACA
At5g54510	GGACACCCTTAAAGCTTCAACAAAACAGATCAAGCTTCTTTCACCATTTTCACTCTCTTTAAGCTTT CTTTCTTAATTTCTCTCATTTCGAATTTTAAACACAAAACCTAAACG
At5g57440	GGGGTTCGTCGTTATCGGTATCGAATCAACA

Table 16. RNA 一過性発現実験に用いられた遺伝子の 5'UTR 配列
解析の都合上、挿入された 2 塩基の G を赤色で示している。

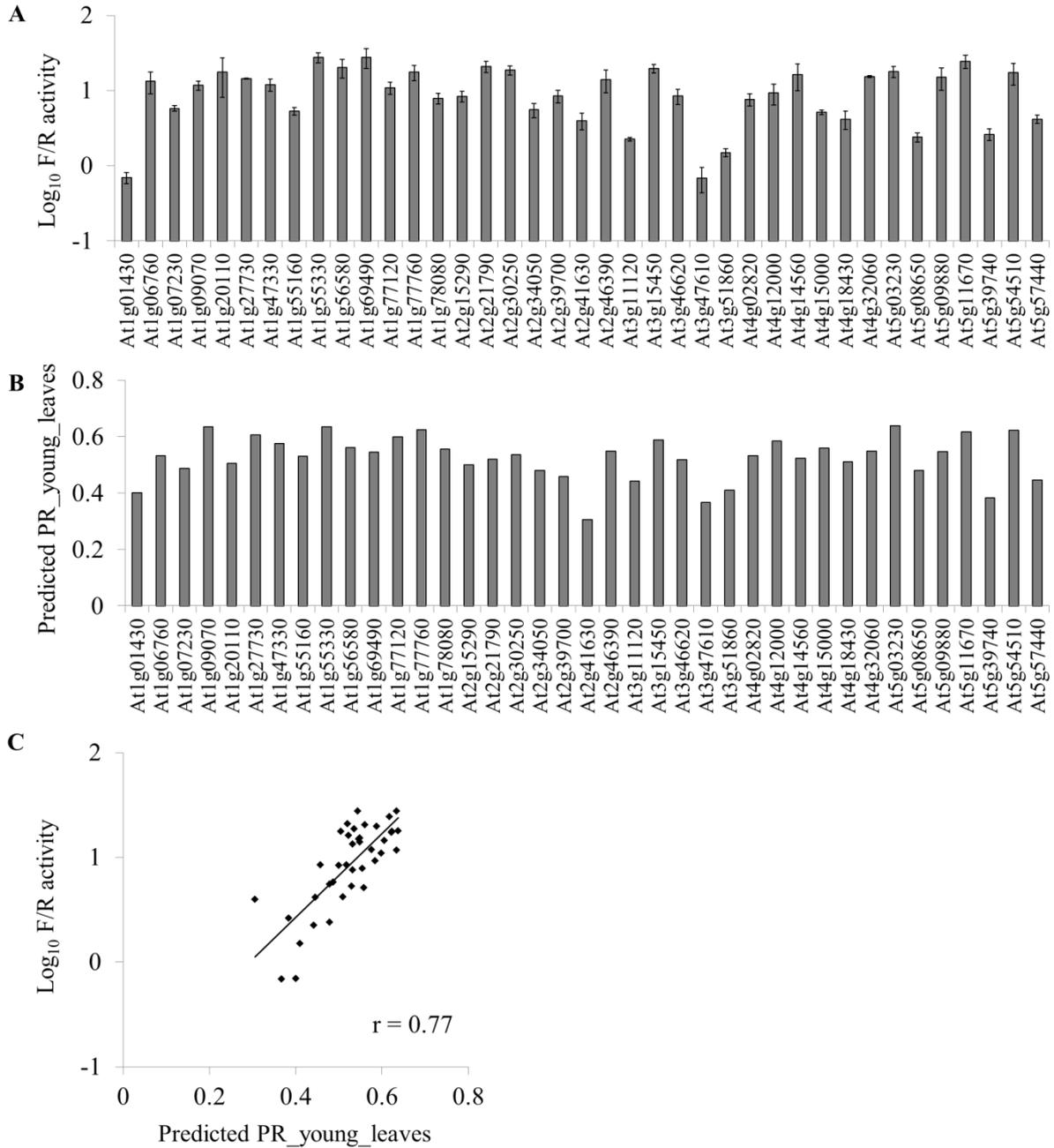


Figure 25. RNA 一過性発現実験による PLS モデルの検証

(A) Matsuura ら (Matsuura et al. 2013) によって報告された 39 遺伝子由来の 5'UTR をレポーター遺伝子 (F-luc) に連結した各構築での F-luc 活性値を示す。F-luc 活性値は供導入した R-luc 活性値によって F/R 活性値として補正し対数値に変換している。(B) 39 遺伝子由来の 5'UTR についての予測 PR_young_leaves 値を示している。(C) 39 遺伝子由来の 5'UTR についての予測 PR_young_leaves 値と実測された Log₁₀ F/R 活性値の散布図。近似線 (実線) とピアソンの相関係数 (r) を図中に示す。

Gene	5'UTR sequence
At1g77120.1	TACATCACAATCACACAAAACAAACAAAAGATCAAAGCA AGTTCTTCACTGTTGATA
At1g06760.1	CAATCCTCATAATCACTTTCGAAATTACATTTACGCTTTCTT GCAATCAAATTTTCCGATCTTAAGTTCAGAAGACG
At1g34000.1	AGACAATTCAACTAACAAAAA
At4g09650.1 TSS1	GAGAAACCACAAATCTCTCTTTCTCTCAAACCTCTCTCAACAA CAACA
At4g09650.1 TSS2	CAAATCTCTCTTTCTCTCAAACCTCTCTCAACAACAACA
At1g20440.1 TSS1	AAACATTACTCATTACAAAACCATCTTAAAGCAACTACAC AAGTCTTGAAATTTTCTCATATTTTCTATTTACTATATAAAC TTTTAATCAAATCAAGATTAAC
At1g20440.1 TSS2	CAAACATTACTCATTACAAAACCATCTTAAAGCAACTACA CAAGTCTTGAAATTTTCTCATATTTTCTATTTACTATATAAA CTTTTAATCAAATCAAGATTAAC
At1g20440.1 TSS3	ACAAACATTACTCATTACAAAACCATCTTAAAGCAACTAC ACAAGTCTTGAAATTTTCTCATATTTTCTATTTACTATATAA ACTTTTAATCAAATCAAGATTAAC
At5g13420.1	GATCGATCAAACCAAGAAAAAACACTTTCGTATTTCCCTCG ACGAAAAA
At2g22230.1	ACCTGCAAAAACAACATCTCTCACATTCTCTCTAAACTCTCT TCACTGTTACTCGACA

Table 17. DNA 一過性発現実験に用いた遺伝子の 5'UTR 配列

At4g09650.1 と At1g20440.1 は異なる TSS に由来する 5'UTR バリエントも試験している。

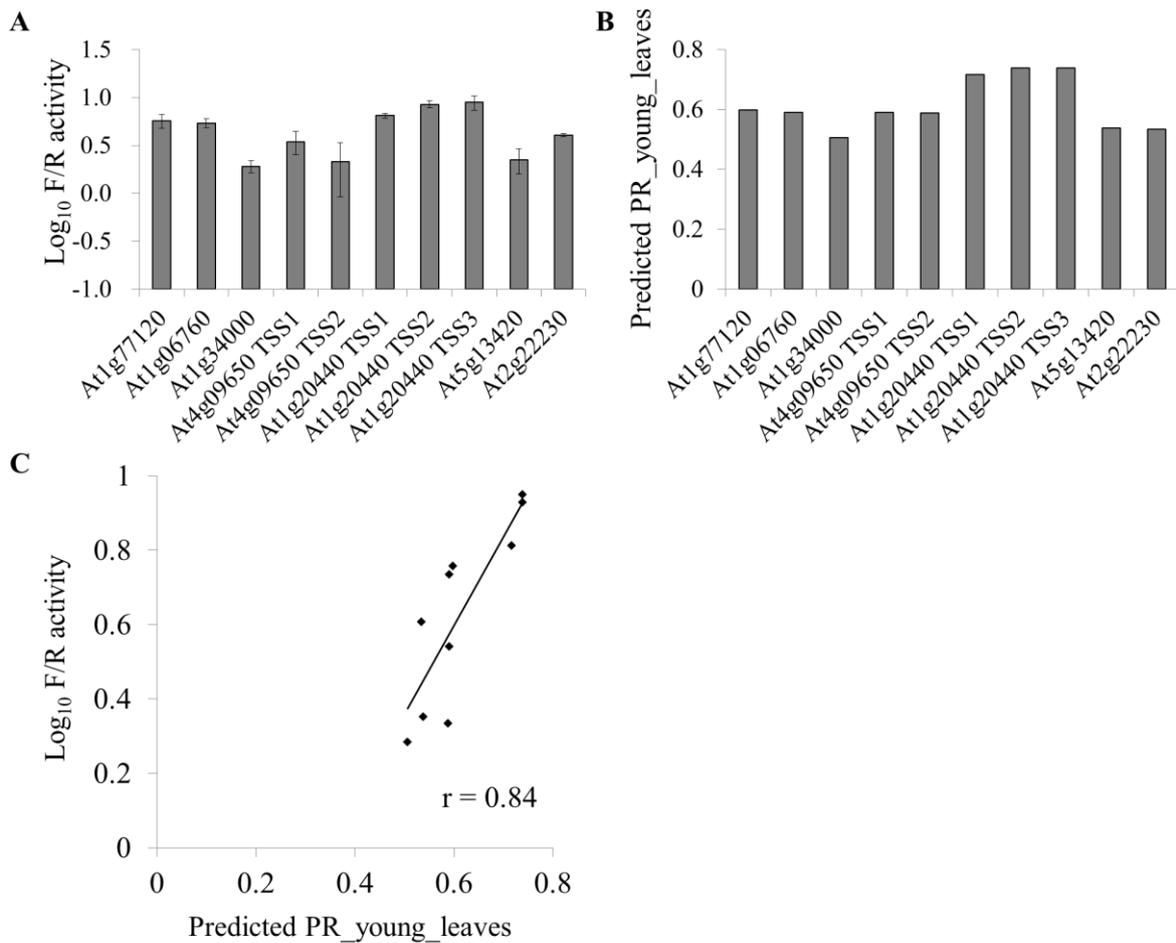


Figure 26. DNA 一過性発現実験による PLS モデルの検証

(A) 7種の遺伝子由来の5'UTRバリエントを含む10種の5'UTRをレポーター遺伝子(F-luc)に連結した各構築でのF-luc活性値を示す。F-luc活性値は供導入したR-luc活性値によってF/R活性値として補正し対数値に変換している。(B) 10種の5'UTRについての予測PR_young_leaves値を示している。(C) 10種の5'UTRについての予測PR_young_leaves値と実測したLog₁₀F/R活性値の散布図。近似線とピアソンの相関係数(r)を図中に示す。

2-3-8. 塩基置換による PLS モデルで示された重要な配列的特徴の検証

これまでの結果から、未展開葉に関する PLS モデルによって、翻訳状態の決定に関わる重要な配列的特徴が示され、その PLS モデルの信頼性を一過性発現実験により確認した。更なる検証として、未展開葉に関する PLS モデルによって示された重要な配列的特徴について個々にその重要性を確かめるため、いくつかの 5'UTR 上の配列的特徴量を変化させた配列置換系列を作製し、レポーター遺伝子である F-luc に連結し、シロイヌナズナ T-87 培養細胞を用いた DNA 一過性発現実験によって、レポーター活性を指標とした評価を行った。今回、検証した配列的特徴は、翻訳状態に強い負の影響を与える mRNA の Cap 側の二次構造 ($-\Delta G_{1\sim 100_from_Cap}$) と Cap 側 4 nt での C の含量 ($C_{1\sim 4_from_Cap}$)、強い正の影響を与える Cap 側 4 nt での A の含量 ($A_{1\sim 4_from_Cap}$) である (Table 15)。他にモデルによって示された影響力が大きい配列的特徴として、強い負の影響を与える uORF があるが、uORF が本来の AUG からの翻訳を抑制することは既によく知られているため本研究では改めての検証は行っていない。置換系列 (Substitute series: SS1, SS2) には転写開始点が 1 点に 66.7% 以上収束し、今回検証する配列的特徴を本来の 5'UTR 内にあまり持たない At5g08680.1 と At5g55250.1 の 5'UTR を選抜した。そこに負の配列的特徴である Cap 側の二次構造の形成度合いを増加させた配列 (強く形成する N- ΔG 、弱く形成する N- ΔG_{mild})、Cap 側 4 nt の C を増加させた配列 (NC)、正の配列的特徴である Cap 側 4 nt の A を増加させた配列 (PA)、正と負の配列的特徴を複合的に配列置換し、モデル上は非常に高い PR 値が期待できる配列 (P-Multi) を作製し検証した (Table 18, Table 19)。この時、他の配列的特徴の量はできる限り変化させない様に配列を置換した。なお、P-Multi の Cap 側 4 nt はモデル上では AAAA が望ましいが、35S プロモーターの TSS 下流に A が長く連続する 5'UTR を連結した場合、本来の TSS から転写されない事例があったため、今回は GAAA とした。これらの配列を用いて一過性発現実験を行い各置換系列の翻訳効率を調べたところ (Fig. 27A)、概ねモデルから予測した結果と一致し (Fig. 27B)、実測した翻訳効率と予測 PR 値の間には高い相関 ($r = 0.71$) が認められた (Fig. 27C)。特にモデルによって示された負の配列的特徴である $-\Delta G$ を増加させた場合では、顕著に翻訳効率が低下しており、加えて $-\Delta G_{mild}$ 等の結果から $-\Delta G$ が増加するにつれ翻訳効率も徐々に低下していくことも示された。また、モデルによって示された負の配列的特徴である Cap 側 4 nt の C を増加させると、SS1 では翻訳効率の減少が確認された一方で、SS2 では大きな変化は認められなかった。正の配列的特徴である Cap 側 4 nt の A を増加させた場合では、SS1 では逆に少し翻訳効率が下がったものの、SS2 では翻訳効率の増加が確認された。しかし、モデル上では非常に高い翻訳効率が期待される P-Multi は、SS1、SS2 でともに翻訳効率が落ちる結果となった。この結果は、PLS モデルの構築に用いた未展開葉での翻訳状態と一過性発現実験に用いているプロトプラスト化した培養細胞での翻訳状態の違いか、この PLS モデルでは評価されていない他の重要な配列パターンが出現したことに起因している可能性が考えられる。

予測と大きく異なるこの P-Multi を除外した場合、翻訳効率と予測 PR 値の相関係数は 0.88 であった。

Fig. 27 の結果には不明瞭な点があったため、はっきりとした影響が確認できなかった Cap 側 4 nt での A と C 含量の影響と複合的な置換である P-Multi について再検証を行った。新たに置換系列として At5g24840.1 と At4g38160.3 の 5'UTR を SS3、SS4 として選抜し、SS1 や SS2 と同様に置換系列を作製し (Table 20, Table 21)、一過性発現実験にて翻訳効率を測定した。その結果、実測した翻訳効率 (Fig. 28A) と予測した PR 値 (Fig. 28B) の間で非常に高い相関 ($r = 0.93$) が認められた (Fig. 28C)。Cap 側 4 nt の A と C 含量に関しては、SS3 では顕著な変化ではなかったが、SS4 では予測通りの結果を示した。加えて、SS3 と SS4 の両方で P-Multi は、本来の配列よりも高い翻訳効率を示しており、PLS モデルによる予測と類似した結果となった。一部では配列置換による影響が少なかったものの全体的には、モデルによって示された配列的特徴は、実際の翻訳効率にも影響を与えていることが示され、構築した PLS モデルの精度と、そのモデルによって示された配列的特徴の重要性を確認できた。しかし、前述のように未展開葉でのデータから構築した PLS モデルと、評価に用いたプロトプラスト化した細胞での翻訳状態の差異について考慮する必要がある。

Name	5'UTR sequence
SS1	ATTCATTTGATAACCCTAGTAGCTCCTTTTTTCTCTATCTCGCTTTCTCTGTTTCACCACAGTC
SS1_NC	ACC CATTTGATAACCCTAGTAGCTCCTTTTTTCTCTATCTCGCTTTCTCTGTTTCACCACAGTC
SS1_N-ΔG	ATTCATTTGATAACCCTAGTAG GCTAGGGTTATCAAGTGAGC TTTCTCTGTTTCACCACAGTC
SS1_N-ΔG_mild	ATTCATTTGATA CCAAGTGAAT TCCTTTTTTCTCTATCTCGCTTTCTCTGTTTCACCACAGTC
SS1_PA	AA CATTTGATAACCCTAGTAGCTCCTTTTTTCTCTATCTCGCTTTCTCTGTTTCACCACAGTC
SS1_P_Multi	GAAA ATTTGATAACCCTAG TAATCTTCCCTC CCTATT TCACCCCTCTCA TT ATAGATCACAC
SS2	ATTCACAAAAAGAAGGTGTAGAGAAAGTTAAAGAGAGAAGGAGAGATCCATAGAGAAAGAGAAAGAGA
SS2_NC	ACC CACAAAAAGAAGGTGTAGAGAAAGTTAAAGAGAGAAGGAGAGATCCATAGAGAAAGAGAAAGAGA
SS2_N-ΔG	ATTCACAAAAAGAAGGTGTAGAG TGCGCCTTCTTTTTGTGAAT AGATCCATAGAGAAAGAGAAAGAGA
SS2_N-ΔG_mild	ATTCACAAAAAG ATTTTGTGAAT AAAGTTAAAGAGAGAAGGAGAGATCCATAGAGAAAGAGAAAGAGA
SS2_PA	AA CACAAAAAGAAGGTGTAGAGAAAGTTAAAGAGAGAAGGAGAGATCCATAGAGAAAGAGAAAGAGA
SS2_P_Multi	GAAA ACAAAAAGAAGGTGTAGAGAAAGTTAAAGAGAGAAGGAGAGA AGG ATAGAGAAAGAGAAAGAGA

Table 18. DNA 一過性発現実験に用いた置換系列 SS1 と SS2 の 5'UTR 配列

置換系列に用いた SS1、SS2 の配列とその置換系列の配列を示している。塩基置換した配列を赤で示した。

Name	Predicted PR	Region			Secondary_Structure		
		From Cap 1~4		From AUG -23~-8	From_Cap		From_AUG
		A	C	GC	1-100 -ΔG	1-300 -ΔG	-150~-51 -ΔG
SS1_NC	0.51→0.46	1	1→3	1	13.1	67.9→69.2	0
SS1_N-ΔG	0.51→0.33	1	1	1	13.1→36.8	67.9→87.9	0
SS1_N-ΔG_mild	0.51→0.47	1	1	1	13.1→18.3	67.9→69.2	0
SS1_PA	0.51→0.57	1→3	1	1	13.1	67.9→66.7	0
SS1_P_Multi	0.51→0.73	1→3	1→0	1→0	13.1→3.8	67.9→57.8	0
SS2_NC	0.62→0.57	1	1→3	0	4.2	59.9→58.8	0→0.2
SS2_N-ΔG	0.62→0.38	1	1	0	4.2→29.7	59.9→83.7	0
SS2_N-ΔG_mild	0.62→0.55	1	1	0	4.2→10.7	59.9→65.1	0
SS2_PA	0.62→0.69	1→3	1	0	4.2→ 4.7	59.9→59.1	0
SS2_P_Multi	0.62→0.78	1→3	1→0	0	4.2→ 0.1	59.9→56.2	0

Table 19. SS1 と SS2 の各置換系列の配列的特徴の変化

それぞれの置換系列での配列的特徴の変化とそれに伴う予測 PR 値の変化を示している。

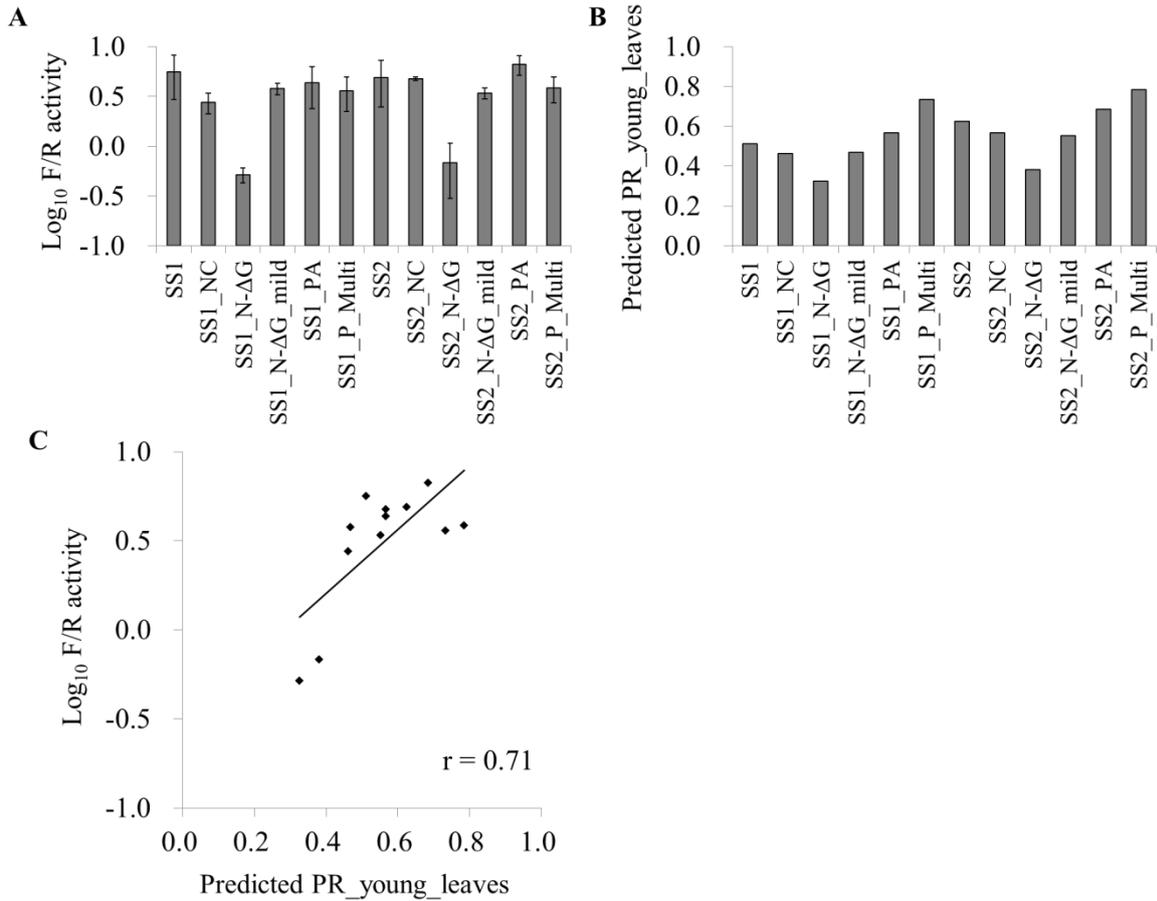


Figure 27. DNA 一過性発現実験による置換系列 SS1 と SS2 の評価

(A) 5'UTR 配列の置換系列をレポーター遺伝子 (F-luc) に連結した各構築での F-luc 活性値を示す。F-luc 活性値は供導入した R-luc 活性値によって F/R 活性値として補正し対数値に変換している。(B) 置換系列についての予測 PR_young_leaves 値を示している。(C) 置換系列についての予測 PR_young_leaves 値と実測した翻訳効率を表す Log₁₀ F/R 活性値の散布図。近似線 (実線) とピアソンの相関係数 (r) を図中に示す。

Name	5'UTR sequence
SS3	GAGGGTTAGTTTTCGAAGAACGACGAGCTCGAGATTATTGATTTTCAAGTGGAGATTGAATCAGTGATA
SS3_NC	GCCC GTTAGTTTTCGAAGAACGACGAGCTCGAGATTATTGATTTTCAAGTGGAGATTGAATCAGTGATA
SS3_PA	GAAAGTTAGTTTTCGAAGAACGACGAGCTCGAGATTATTGATTTTCAAGTGGAGATTGAATCAGTGATA
SS3_P_Multi	GAAACAAACAAAAGGAAGAGCAACGATCGGGAGAAATAATTAAGAATAAGGAGATAGAAAGAGTGAAA
SS4	GTGGATAAGCTTCTCACTTTGAGTTCAATTGGATAGCCGGAAAGGCTCTCCTCCAAGTGGCGTTTAT
SS4_NC	GCCATAAGCTTCTCACTTTGAGTTCAATTGGATAGCCGGAAAGGCTCTCCTCCAAGTGGCGTTTAT
SS4_PA	GAAAATAAGCTTCTCACTTTGAGTTCAATTGGATAGCCGGAAAGGCTCTCCTCCAAGTGGCGTTTAT
SS4_P_Multi	GAAAATAATCCCTCATCCCAACCCAAGTACCGAAGAACTAACCAAAAATACAACATTCAA

Table 20. DNA 一過性発現実験に用いた置換系列 SS3 と SS4 の 5'UTR 配列

置換系列に用いた SS1、SS2 の配列とその置換系列の配列を示している。塩基置換した配列を赤で示した。

Name	Predicted PR	Region			Secondary_Structure		
		From Cap 1~4		From AUG -23~-8	From_Cap		From_AUG
		A	C	GC	1-100 -ΔG	1-300 -ΔG	-150~-51 -ΔG
SS3_NC	0.52→0.42	1→0	0→3	0	16.1→17.2	70.8→73.9	0.3→0
SS3_PA	0.52→0.58	1→3	0	0	16.1→15.4	70.8→70.4	0.3→1
SS3_P-Multi	0.52→0.78	1→3	0	0	16.1→ 0.1	70.8→56.5	0.3→0
SS4_NC	0.43→0.37	0	0→3	1	20.8→20.5	79.5→84.2	2→0.2
SS4_PA	0.43→0.51	0→3	0	1	20.8→20.5	79.5→76.3	2→ 0
SS4_P-Multi	0.43→0.78	0→3	0	1→0	20.8→ 0.1	79.5→56.4	2→ 0

Table 21. SS3 と SS4 の各置換系列の配列的特徴の変化

それぞれの置換系列での配列的特徴の変化とそれに伴う予測 PR 値の変化を示している。

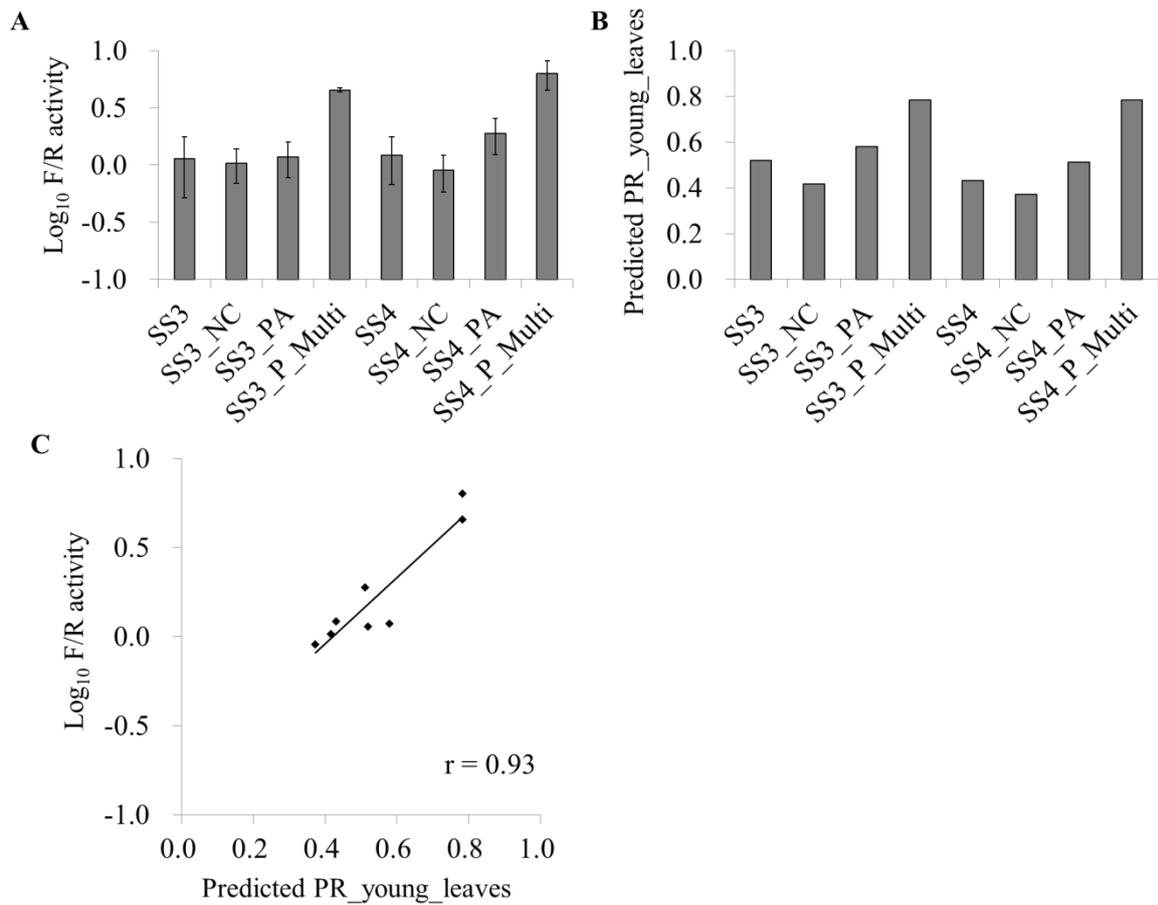


Figure 28. DNA 一過性発現実験による置換系列 SS3 と SS4 の評価

(A) 5'UTR 配列の置換系列をレポーター遺伝子 (F-luc) に連結した各構築での F-luc 活性値を示す。F-luc 活性値は供導入した R-luc 活性値によって F/R 活性値として補正し対数値に変換している。(B) 置換系列についての予測 PR_young_leaves 値を示している。(C) 置換系列についての予測 PR_young_leaves 値と実測した翻訳効率を表す Log₁₀ F/R 活性値の散布図。近似線とピアソンの相関係数 (r) を図中に示す。

2-3-9. 様々な条件下での PLS モデルの構築とその比較

ここまでに、未展開葉において、5'UTR の配列的特徴で mRNA の翻訳状態を説明できる PLS モデルの構築を行い、構築した PLS モデルが実際のタンパク質の生産効率である翻訳効率を十分に説明できることを確認し、構築した PLS モデルの信頼性を確かめた。この未展開葉でのモデルによって、翻訳状態に影響を与える重要な配列的特徴が示されているが、翻訳状態は様々な条件で異なっていることから、それぞれの翻訳状態には異なる配列的特徴が関係している可能性が考えられた。そこで、翻訳状態 (PR 値) と 5'UTR の配列情報をともに取得している 2DAG、展開葉 (mature leaves)、培養 3 日目コントロール (Control)、熱ストレス下 (37°C)、塩ストレス下 (200 mM) についても未展開葉と同様に PLS モデルの構築を行い、その条件における翻訳状態の決定に重要な配列的特徴を解析した。

PR_2DAG を配列的特徴で説明する PLS モデルの構築では、最も翻訳状態をよく説明できた予備的な説明変数の再選抜 5 回目のモデルを使用した。この PLS モデルは翻訳状態を十分に説明でき (Fig. 29)、この PLS モデルより 8 個の重要な配列的特徴が示された (Table 22)。特に翻訳状態に影響を与えていた配列的特徴には、二次構造の形成度合いや A リッチな配列パターン (Motif) などがあつた (Fig. 30)。

展開葉での PLS モデルの構築では、最も翻訳状態をよく説明できた予備的な説明変数の再選抜 6 回目のモデルを使用した。この PLS モデルも翻訳状態を十分に説明でき (Fig. 31)、この PLS モデルより 16 個の重要な配列的特徴が示された (Table 23)。特に翻訳状態に影響を与えていた配列的特徴には、二次構造の形成度合いや CT リッチな Motif (Fig. 32)、A 含量などがあつた。なお、この PLS モデルで示されている uCUG は過去に非 AUG 翻訳開始コドンとして報告がある配列であり (Brar et al. 2012, Ohta et al. 2010, Depeiges et al. 2006)、uORF に関連した配列的特徴である。

37°Cでの PLS モデルの構築では、最も翻訳状態をよく説明できた予備的な説明変数の再選抜なしのモデルを使用した。この PLS モデルも翻訳状態を十分に説明でき (Fig. 33)、この PLS モデルより 12 個の重要な配列的特徴が示された (Table 24)。特に翻訳状態に影響を与えていた配列的特徴には、二次構造の形成度合いや C 含量、A 含量、A リッチな Motif (Fig. 34) などがあつた。

残りの Control と 200 mM に関しては、十分な精度の PLS モデルの構築には至らなかった。これは、精度の高い PLS モデルの構築には、目的変数 (PR 値) がある程度幅広く分布している必要があるが、Control と 200 mM では他の条件に比べ PR 値の分布が狭かったのが原因であると考えられる。

翻訳状態を十分に説明できた 2DAG、未展開葉、展開葉、37°Cの PLS モデルから合計 45 個 (モデル間の重複を含む) の様々な配列的特徴が示され、それらの中には二次構造の形成度合いなどの全てのモデルに含まれていた配列的特徴や、逆に展開葉のモデルでのみ示されている CT リッチな Motif 等の条件特異的な配列的特徴が存在していた。各条件での翻訳状態の決定に関わる配列的特徴の違いを更に解析するため、各モデルで示された 45 個の配列的特徴について、ワード法による階層的

クラスター分析を行った (Fig. 35)。データセットには、TSS が分散しすぎておらず、幅広い PR 値分布を示す 1023 遺伝子の配列情報を使用し、各遺伝子における 45 個の配列的特徴の量を評価した。類似する配列的特徴ならば各遺伝子の評価値も近い値となり 1 個のクラスターを形成するが、大きく異なる配列的特徴ならば各遺伝子での評価値も大きく異なる値を示すこととなり異なるクラスターを形成することとなる。その結果、それぞれの配列的特徴を、類似した配列的特徴ごとに分けることができ、それらを 8 個のクラスターとした (Fig. 22、赤枠で示す)。クラスター1は、uORF や 5'UTR の長さに関連するものが多く (- $\Delta G_{-150 \sim -51_from_AUG}$ という配列的特徴が含まれるが、長い 5'UTR を持つ遺伝子での特徴である)、それらに関連したクラスターと考えられる。クラスター2は、C または T に関わるものが多く、C 及び T の含量に関連したクラスターと考えられる。クラスター3 に属する特徴は、CDS 内の 25S rRNA との相補配列長や Cap 側 4 nt の C 含量など大きく異なる配列的特徴であり解釈はできていない。クラスター4 は、Cap 側の A 含量、クラスター5 は、CDS 長を表していた。クラスター6 は、A 含量または A (A 及び G) リッチな Motif を示し、特に影響が大きいものには Cap から 50 nt 内を示すものが多く、Cap 側の A 含量に関連したクラスターと考えられる。クラスター7 は、Cap 側 100 nt の二次構造または G 含量に関わるものが多く Cap 側の二次構造に関連したクラスターと考えられる。クラスター8 は、Cap から広い範囲内での二次構造に関連していると考えられる。

それぞれのクラスターが各条件での翻訳状態に与える影響を評価するため、各条件での重みをクラスター単位で合計した (Table 25)。全ての条件で翻訳状態に影響を与えるクラスターは、5 と 7 であった。CDS 長を示すクラスター5 は、PR 値 (リボソームローディング) に CDS 長が直接的に関わるため全ての条件で強い正の影響を与えていた。Cap 側の二次構造に関連しているクラスター7 は、全ての条件で負の影響を与えており、未展開葉では他より影響が強かった。クラスター1、2、6、8 は、条件間で与える影響の有無が異なっていた。uORF や 5'UTR の長さに関連するクラスター1 は、37°C 以外では負の影響を与えており、特に 2DAG と未展開葉での影響が大きかった。C と T の含量に関連するクラスター2 は、展開葉と 37°C でのみ負の影響を与えていた。Cap 側の A 含量に関連するクラスター6 は、37°C でのみ強い正の影響を与えていた。Cap から広い範囲での二次構造に関連するクラスター8 は、2DAG 以外では負の影響を与えており、特に展開葉と 37°C で影響が大きかった。これらの結果によって、全ての条件で翻訳状態に影響を与える共通する配列的特徴の存在と、特定の条件でのみ翻訳状態に影響を与える特異的な配列的特徴が存在することが明らかとなった。このことは、第一章において考察した様々な条件で共通する翻訳機構と、特定の条件に特異的な翻訳機構にそれぞれ関わっていると考えられる。

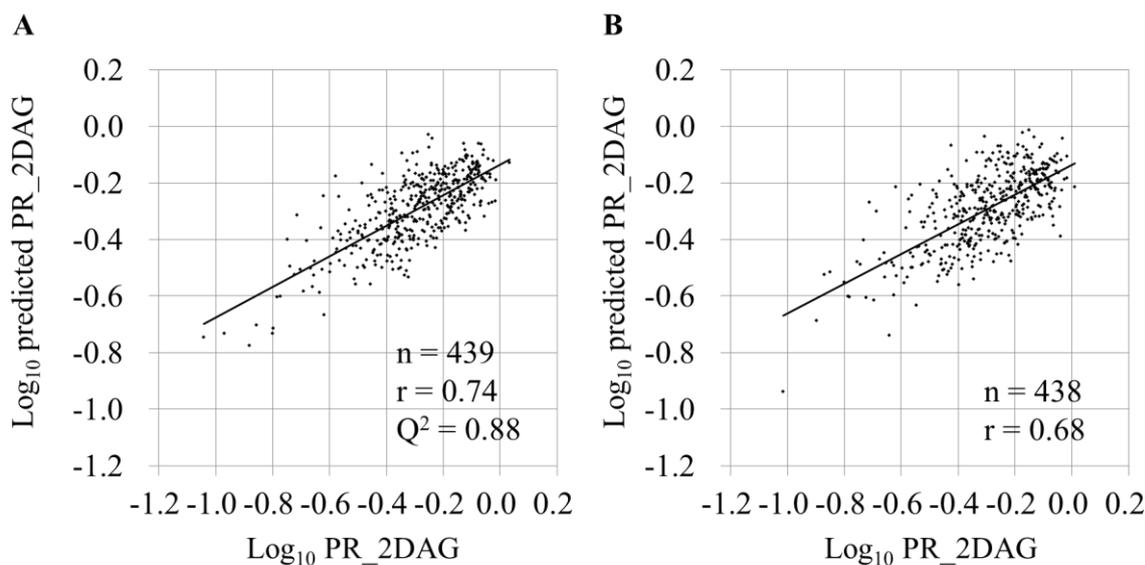


Figure 29. 2DAG の PLS モデルによる予測 PR 値と実測 PR 値の相関

(A) PLS モデルの構築に用いた 439 遺伝子についての実測 Log_{10} PR_2DAG と予測 Log_{10} PR_2DAG の散布図。データ数 (n) と近似線 (実線)、相関係数 (r)、モデルの精度 (Q^2) を図中に示す。(B) PLS モデルの構築に用いた 439 遺伝子とは異なる 438 遺伝子についての実測 Log_{10} PR_2DAG と予測 Log_{10} PR_2DAG の散布図。データ数 (n) と近似線 (実線)、相関係数 (r) を図中に示す。

Explanatory variable	Weight
$-\Delta G_{-150 \sim -51_from_AUG}$	-0.047
$-\Delta G_{1 \sim 100_from_Cap}$	-0.026
Motif_2DAG-1	-0.024
25SrRNA_Duplex_Length_301~400_from_Cap	-0.013
T_1~6_from_Cap	0.011
25SrRNA_Duplex_Length_501~600_from_Cap	0.015
Motif_2DAG-1_1~50_from_Cap	0.035
$\text{Log}_{10_CDS_Length}$	0.113

Table 22. 2DAG の PLS モデルで示された重要な配列的特徴とその重み

最終的な PLS モデルに含まれている説明変数 (重要な配列的特徴) と重み (PLS モデルへ与える影響) の一覧。重みが負の値をとるほど Log_{10} PR_2DAG に負の影響を与えていることを、正の値をとるほど正の影響を与えていることを意味している。説明変数に含まれている Motif_2DAG-1 の配列パターンは Fig. 30 に示してい

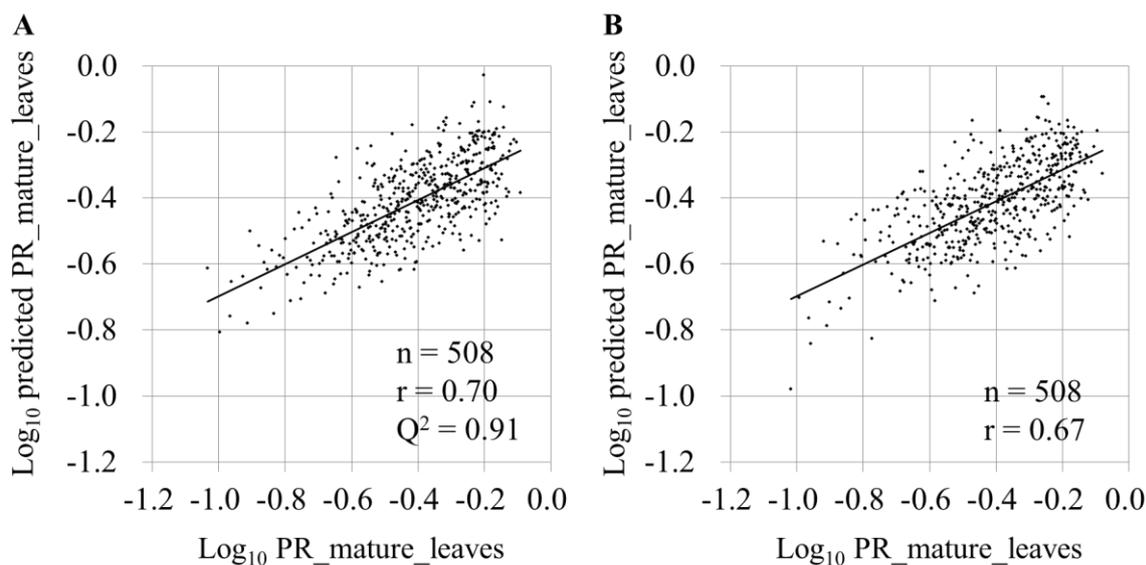


Figure 31. 展開葉の PLS モデルによる予測 PR 値と実測 PR 値の相関

(A) PLS モデルの構築に用いた 508 遺伝子についての実測 Log_{10} PR_mature_leaves と予測 Log_{10} PR_mature_leaves の散布図。データ数 (n) と近似線 (実線)、相関係数 (r)、モデルの精度 (Q^2) を図中に示す。(B) PLS モデルの構築に用いた 508 遺伝子とは異なる別の 508 遺伝子についての実測 Log_{10} PR_mature_leaves と予測 Log_{10} PR_mature_leaves の散布図。データ数 (n) と近似線 (実線)、相関係数 (r) を図中に示す。

Explanatory variable	Weight
$-\Delta G_{1\sim 300_from_Cap}$	-0.042
$-\Delta G_{1\sim 50_from_Cap}$	-0.030
Motif_ML-1_1 \sim 25_from_Cap	-0.015
25SrRNA_Duplex_Length_-150 \sim -51_from_AUG	-0.012
G_1 \sim 13_from_Cap	-0.012
Motif_ML-1	-0.010
uCUG	-0.008
5'UTR_Length	-0.007
Motif_ML-2_-100 \sim -1_from_AUG	-0.006
Motif_ML-2	-0.004
$-\Delta G_{-5\sim 4_from_AUG}$	0.006
Motif_ML-2_-10 \sim -1_from_AUG	0.010
T_-25 \sim -5_from_AUG	0.010
A_1 \sim 13_from_Cap	0.013
A_-50 \sim -8_from_AUG	0.015
Log_{10} _CDS_Length	0.100

Table 23. 展開葉の PLS モデルで示された重要な配列的特徴とその重み

最終的な PLS モデルに含まれている説明変数（重要な配列的特徴）と重み（PLS モデルへ与える影響）の一覧。重みが負の値をとるほど Log_{10} PR_mature_leaves に負の影響を与えていることを、正の値をとるほど正の影響を与えていることを意味している。説明変数に含まれている Motif_ML-1 と Motif_ML-2 の配列パターンは Fig. 32 に示している。"-ΔG"は二次構造の形成度合いを、"uCUG"は CUG の配列パターンの数を、"G, T, A"は塩基の比率を意味し、特定の領域内でのみ評価を行った場合は、基準となった Cap または AUG からの位置を合わせて表記している。

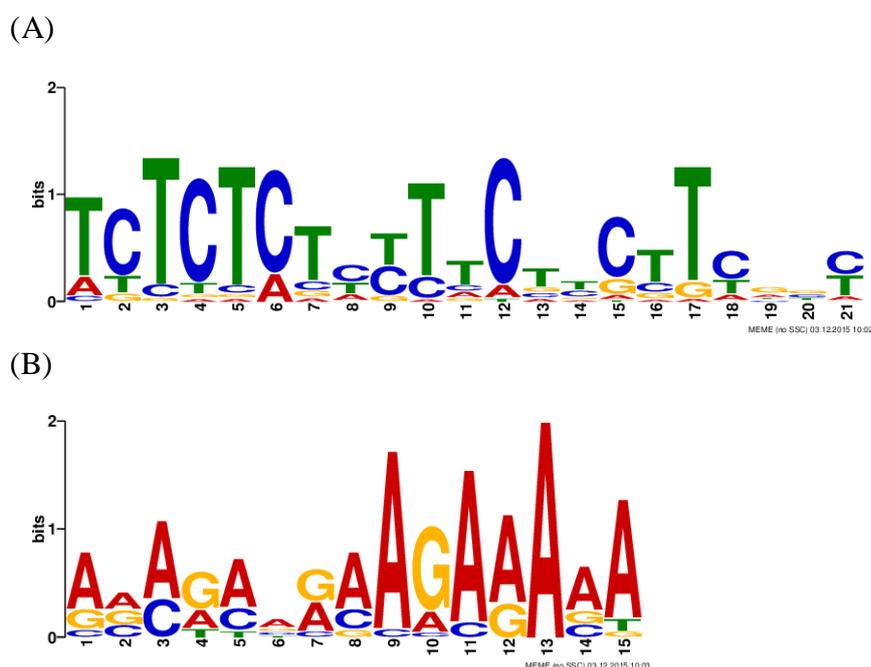


Figure 32. 展開葉の PLS モデルで示された Motif の配列パターン

展開葉での PLS モデルで示された Motif_ML-1 (A) と Motif_ML-2 (B) の配列パターン。文字の高さは、その位置での塩基の偏りの大きさを示している。

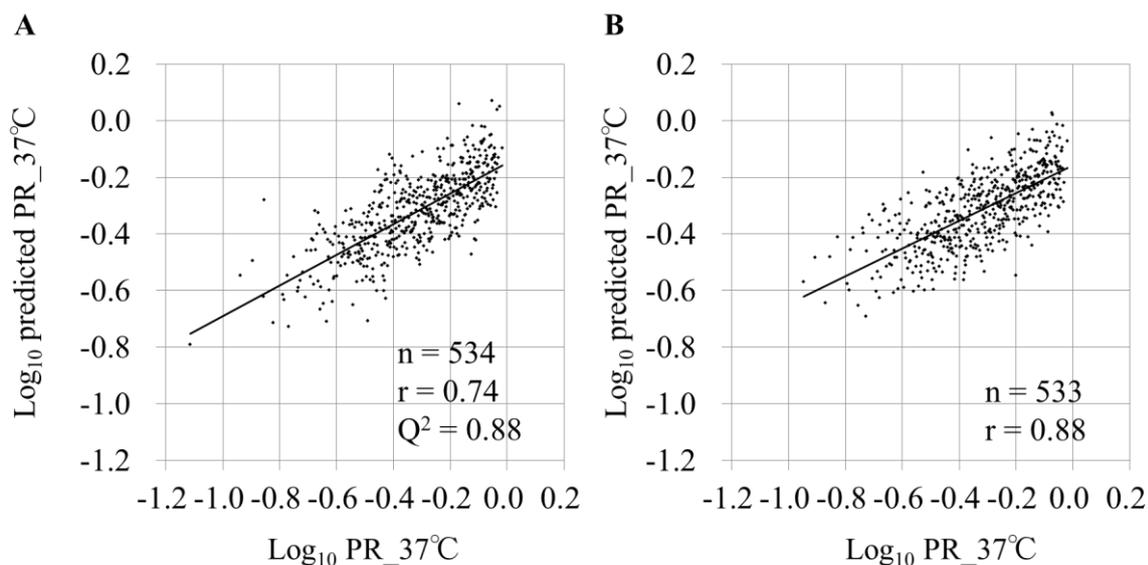


Figure 33. 37°Cの PLS モデルによる予測 PR 値と実測 PR 値の相関

(A) PLS モデルの構築に用いた 534 遺伝子についての実測 $\text{Log}_{10} \text{PR}_{37^\circ\text{C}}$ と予測 $\text{Log}_{10} \text{PR}_{37^\circ\text{C}}$ の散布図。データ数 (n) と近似線 (実線)、相関係数 (r)、モデルの精度 (Q^2) を図中に示す。(B) PLS モデルの構築に用いた 534 遺伝子とは異なる別の 533 遺伝子についての実測 $\text{Log}_{10} \text{PR}_{37^\circ\text{C}}$ と予測 $\text{Log}_{10} \text{PR}_{37^\circ\text{C}}$ の散布図。データ数 (n) と近似線 (実線)、相関係数 (r) を図中に示す。

Explanatory variable	Weight
$-\Delta G_{1\sim 200_from_Cap}$	-0.064
$C_{1\sim 45_from_Cap}$	-0.056
$-\Delta G_{1\sim 250_from_Cap}$	-0.045
Motif $_{37^\circ\text{C}-1_100\sim-1_from_AUG}$	-0.031
GG	-0.023
Motif $_{37^\circ\text{C}-1_25\sim-1_from_AUG}$	-0.012
Motif $_{37^\circ\text{C}-2_10\sim-1_from_AUG}$	0.018
$T_{1\sim 45_from_Cap}$	0.022
Motif $_{37^\circ\text{C}-1_50\sim-1_from_AUG}$	0.023
$A_{1\sim 45_from_Cap}$	0.049
$-\Delta G_{101\sim 200_from_Cap}$	0.065
$\text{Log}_{10_CDS_Length}$	0.066

Table 24. 37°Cの PLS モデルで示された重要な配列的特徴とその重み

最終的な PLS モデルに含まれている説明変数 (重要な配列的特徴) と重み (PLS

モデルへ与える影響)の一覧。重みが負の値をとるほど $\text{Log}_{10} \text{PR}_{37^\circ\text{C}}$ に負の影響を与えていることを、正の値をとるほど正の影響を与えていることを意味している。説明変数に含まれている Motif_{37°C-1} と Motif_{37°C-2} の配列パターンは Fig. 34 に示している。”- ΔG ”は二次構造の形成度合いを、”C, GG, T, A”は塩基の比率を意味し、特定の領域内でのみ評価を行った場合は、基準となった Cap または AUG からの位置を合わせて表記している。

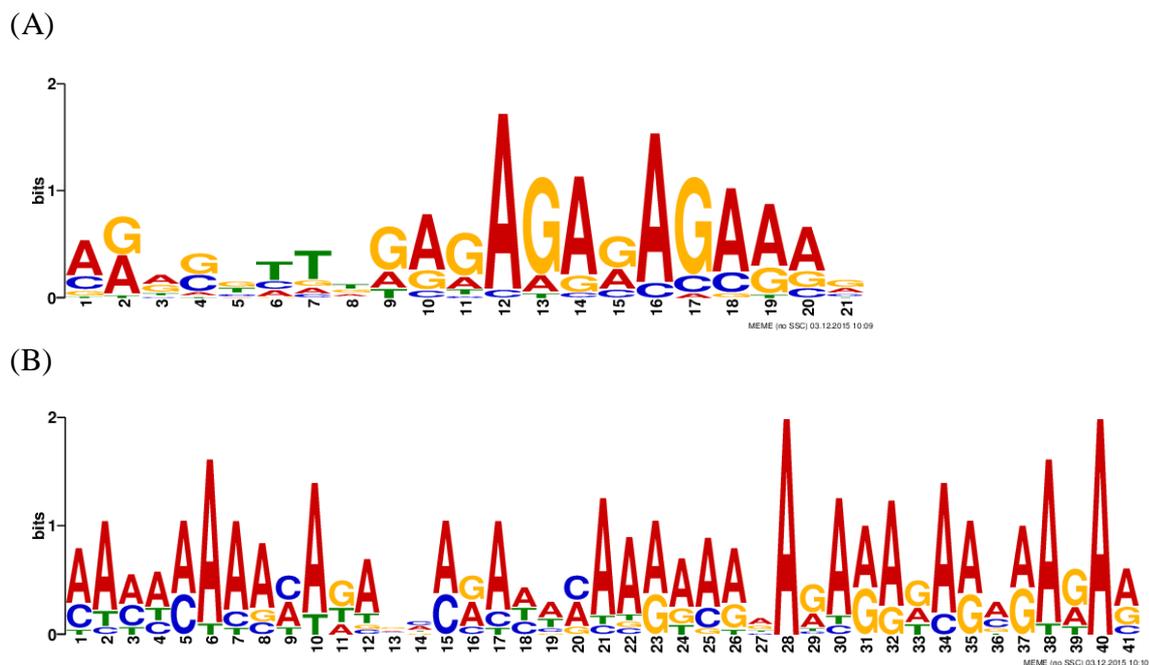


Figure 34. 37°Cの PLS モデルで示された Motif の配列パターン

37°Cでの PLS モデルで示された Motif_{37°C-1} (A) と Motif_{37°C-2} (B) の配列パターン。文字の高さは、その位置での塩基の偏りの大きさを示している。

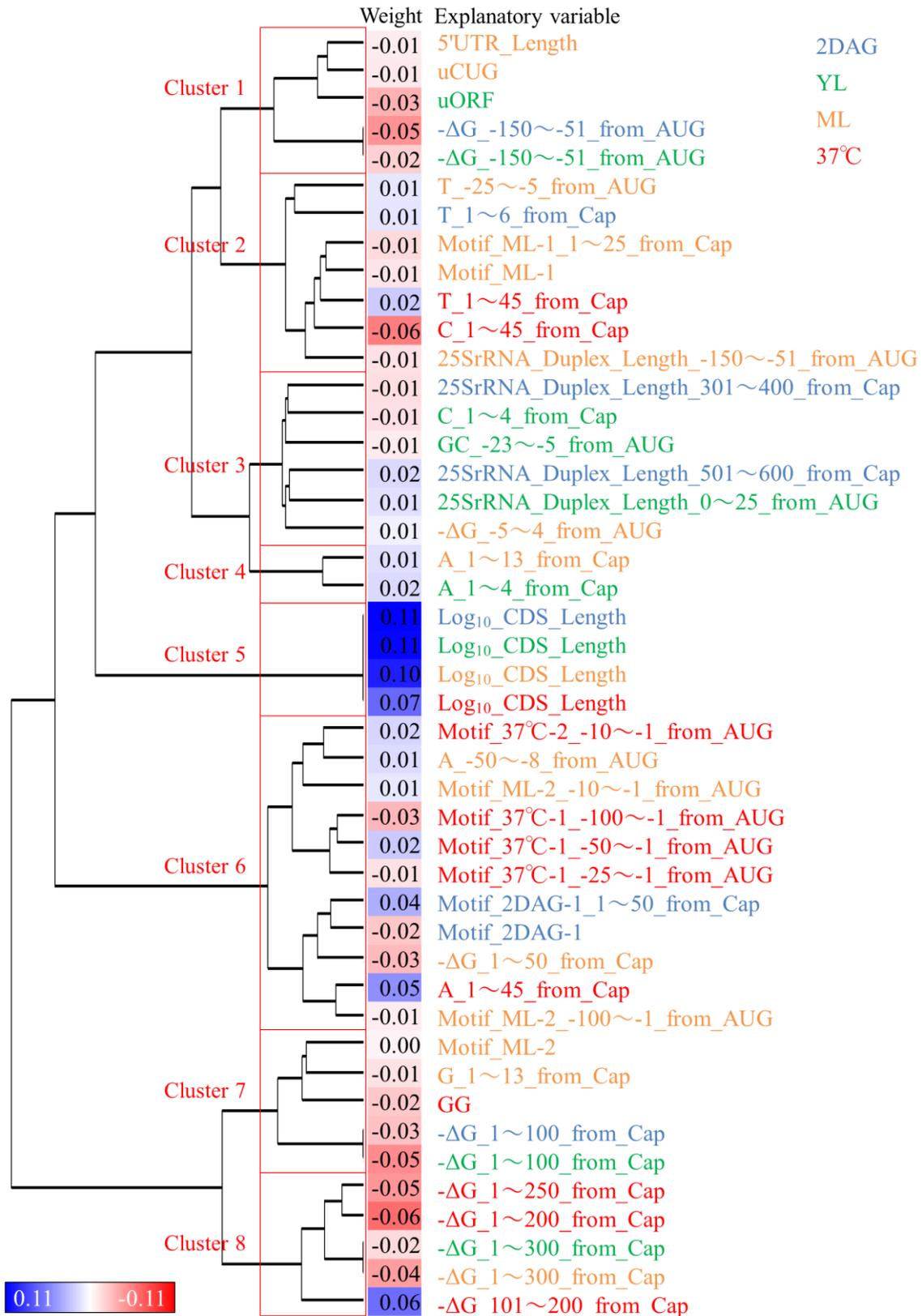


Figure 35. 各 PLS モデルで重要な配列的特徴のクラスタリング

2DAG、未展開葉 (young leaves: YL)、展開葉 (mature leaves: ML)、37°Cの PLS モデルでの合計 45 個の説明変数 (重要な配列的特徴) について、1023 遺伝子の配

列データを基に特徴量の評価を行い、それぞれの遺伝子について、各特徴量の類似度からクラスタリングを行った。各 PLS モデルで示された重要な配列的特徴を色別で示し、それぞれの重みをヒートマップで表している。加えて、クラスタリング結果を基に配列的特徴を 8 個のクラスターに分け、それぞれのクラスター (Cluster1~8) の位置を示している。

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
2DAG	-0.05	0.01	0.00	0.00	0.11	0.01	-0.03	0.00
YL	-0.05	0.00	-0.01	0.02	0.11	0.00	-0.05	-0.02
ML	-0.02	-0.03	0.01	0.01	0.10	-0.01	-0.02	-0.04
37°C	0.00	-0.03	0.00	0.00	0.07	0.05	-0.02	-0.04



Table 25. 配列的特徴のクラスター別の各 PLS モデルへの影響

Fig. 35 において分けられた説明変数 (配列的特徴) の各クラスターについて、各条件 (Fig. 35 の色別) で配列的特徴の重みを合計し、ヒートマップで示している。

2-4. まとめ及び考察

2-4-1. CAGE による 5'UTR 配列の特定

5'UTR 配列は、各 mRNA の翻訳状態の決定において中心的な役割を果たしていることが知られている。そこで第一章で考察した翻訳機構について、5'UTR 配列の違いという観点から解析するために、まずは CAGE を用いて TSS を調べることで、正確な 5'UTR 配列のゲノムワイドな特定を行った。その結果、1 万を超える遺伝子について、様々な 5'UTR バリエーションとその存在比を含む詳細な 5'UTR 配列の情報を取得することができた。CAGE によって得られた個々の遺伝子の詳細な 5'UTR 配列の情報は、翻訳状態を規定している 5'UTR の配列的特徴に関する解析のために非常に重要な知見となった。

加えて、今回の CAGE 結果を用いて、TSS の決定に関する基本的な解析も行った。まず各遺伝子の TSS の分布からは、TSS が 1 点に収束している遺伝子から非常に分散し多様な 5'UTR バリエーションを持つ遺伝子まで幅広く存在し、多くの遺伝子の TSS は分散していることが明らかとなった (Fig. 14B, 15)。また TSS が 1 点に収束している遺伝子では、TSS 上流に存在する TATA box の存在が明瞭であったが、TSS が分散している遺伝子では、TATA box の存在は曖昧だった (Fig. 18)。加えて、TSS の前後数 nt の領域にもその塩基含量に顕著な偏りが認められ (Fig. 17)、特定の塩基パターンがある可能性が示唆された。この領域は一般的に、Initiator (Inr) として知られており、その配列としてショウジョウバエでは TCAKTY が、ヒトでは YYANWYY (下線部が mRNA の 5'末端となる位置を、K は G, T、Y は C, T、W は T, A を意味している) が提唱されている (Kadonaga 2012)。一方で、5'末端となる位置の A はショウジョウバエやヒトでは非常に高度に保存されているが、本研究の結果では、A または G であるなど、一部異なる部分もあり、これらの塩基パターンは、植物での転写装置が転写を開始するために重要な配列であると考えられる。また、いくつかの遺伝子では成長・発達や環境ストレスを通して TSS が変化していた (Fig. 14D, 16)。このような TSS が大きく変化する遺伝子では、TSS が変化しない遺伝子に比べ、その TSS 上流に位置する TATA box の存在が曖昧であり、TATA-less プロモーターの方が TSS の変化が起きやすい傾向があると思われる (Fig. 19)。植物においてコアプロモーター因子として TATA box や Y patch、Inr 等が報告されており (Yamamoto et al. 2009)、3 割ほどの遺伝子が TATA box を有することが知られている。またその報告の中で明確な TATA box や Y patch、Inr を有する遺伝子の TSS は収束している傾向があり、特に TATA box の有無は TSS の収束度合いに大きな影響があることが示されている。本研究の結果も踏まえて考察すると TATA box を有する遺伝子は TSS が収束し、条件によっても TSS が変化しにくく、一様な 5'UTR 配列を有する傾向があるのではないだろうか。

2-4-2. 翻訳状態の決定に関わる配列的特徴

翻訳状態の決定に関わる配列的特徴を解析するために、主に 5'UTR の配列的特徴によって翻訳状態が説明できる PLS モデルを作成し、重要な配列的特徴を選抜するとともに、示された配列的特徴が与える影響を評価した。PLS モデルの構築は 2DAG、未展開葉、展開葉、37°C など複数の翻訳状態について行い (Fig. 24, 29, 31, 33)、それらの翻訳状態の決定に関わる重要な配列的特徴を示した (Table 15, 23, 24, 25)。また、未展開葉での PLS モデルに関しては、一過性発現実験によって PLS モデルが翻訳効率、つまり実際のタンパク質を生産する効率を十分に説明できることを検証するとともに、5'UTR 上の配列的特徴を変化させた配列置換系列によって、その重要性を確認した (Fig. 25, 26, 27, 28)。複合的に配列を置換し、PLS モデル上は非常に高い翻訳効率を示す SS1 と SS2 の P-Multi に関してのみ、予測した翻訳状態と実測の翻訳効率間で大きな違いが認められたが (Fig. 27)、これは構築に用いた未展開葉での翻訳状態と、一過性発現実験に用いたプロトプラスト化した培養細胞での翻訳状態の違いに起因している可能性が考えられた。他の PLS モデル例えば展開葉での PLS モデルで SS1 と SS2 の置換系列を予測した結果では、予測した翻訳状態と実測した翻訳効率間の相関 ($r = 0.82$) は未展開葉 ($r = 0.71$) に比べると高く、P-Multi の予測値も基の配列の予測値とほぼ同じ値となり実測値に近かった。プロトプラスト化した培養細胞の翻訳状態は展開葉の翻訳状態に近いのかもしれない。しかし、一過性発現実験による検証全体では、翻訳状態の予測値と実際の翻訳効率は高い相関を示したことから、今回 PLS モデルを構築した手法は、これまでの特徴単体での解析では難しかった複数の特徴が影響を与えるような複雑な機構を十分に説明することが可能であり、重要な特徴とその影響を客観的に表す有用な手法であると考えられる。

第一章にて、翻訳状態の決定機構には、様々な状況で共通した翻訳機構とそれぞれの状況に応答した特異的な翻訳機構が存在する可能性を考察した。これらの翻訳機構の一端を明らかとするため、構築した 2DAG、未展開葉、展開葉、37°C の PLS モデルによって示された重要な配列的特徴を比較し、示された配列的特徴の差異を考察した。考察には配列的特徴をいくつかのクラスターに分けて行い (Fig. 35, Table 25)、それぞれのクラスターが各条件の翻訳状態に与える影響は、全ての条件で共通した影響から特定の条件だけの特異的な影響まで様々であることが示された。Cap 構造側の二次構造に関連しているクラスターは、全ての条件において負の影響が認められた。この様々な条件で共通して翻訳状態に影響を与えている Cap 構造側の二次構造に関する配列的特徴は、第一章で考察した様々な条件下で共通した翻訳機構に関係している可能性が考えられる。この Cap 構造側の二次構造が翻訳へ及ぼす影響は、動物細胞での研究でも知られており、Cap 構造に非常に近い位置 (Cap から 45 nt) に存在するヘアピン構造 (強固な二次構造) が翻訳効率を大きく低下させることが報告されている (Babendure et al. 2006)。この Cap 構造近傍でのヘアピン構造は、Cap 構造を認識する eIF4G を含む開始前複合体の mRNA への結合を物理的に阻害す

ることで翻訳を抑制していると考えられており、様々な条件下で共通した翻訳機構の分子機構の1つなのではないだろうか。興味深いことに、この Cap 構造近傍の二次構造による翻訳の阻害は、酵母では確認されておらず、開始 AUG 近くの領域で形成される二次構造による翻訳の阻害が観察されている (Vega-Laso et al. 1993, Koloteva et al. 1997)。このことから、動物等の高等真核生物では、Cap 構造と開始前複合体の結合が重要なものに対し、酵母等の下等な真核生物では、リボソームのスキャニングが重要なのではないかと考えられている (Koloteva et al. 1997)。本研究の結果では、動物細胞と同様に Cap 構造近傍の二次構造の重要性が示されているが、2DAG 以外の PR 値の分布が全体的に低い傾向がある条件、特に展開葉や 37°C において、Cap 構造から 200~300 nt の広い範囲での二次構造の形成度合いが翻訳状態の決定に負の影響を及ぼすことが示されている。植物では 5'UTR 長が比較的短いため (Fig. 14C)、この範囲には開始 AUG が含まれる可能性が高く、この領域内の二次構造は、酵母と同様にリボソームのスキャニングまたは翻訳の伸長に影響を及ぼしている可能性が考えられる。リボソームが mRNA 上をスキャニングする際には、RNA ヘリカーゼである eIF4A (eIF4A-1, eIF4A-2) によって二次構造を形成している 2 本鎖を解きながら進んでいくことが知られている。この eIF4A-1 と eIF4A-2 は、分裂組織等の活発な成長が行われている組織では、発現が誘導され、細胞周期の進行や細胞サイズ決定に重要であることが示唆されている (Le et al. 1998, Bush et al. 2015)。加えて低酸素ストレス時には eIF4A-1 タンパク質と高い相同性を示す eIF4A-3 の細胞内局在が核質から核小体とスプライシング小斑点に速やかに変化することも報告されている (Koroleva et al. 2009)。eIF4A の発現が誘導される発芽初期ではヘリカーゼ活性が高く、展開葉や環境ストレス時では発現量の低下や局在の変化によってヘリカーゼ活性が低下し、mRNA のスキャニングに影響を及ぼしている可能性が考えられる。この Cap 構造から広い範囲での二次構造は、スキャニングへの影響によって成長した植物や環境ストレス下での特異的な翻訳機構に関わっているのではないだろうか。

加えて、他にも限られた条件下でのみ特異的に翻訳状態に影響を与える配列的特徴のクラスターも存在し、C 及び U の含量に関連するクラスターは、展開葉と 37°C でのみ負の影響を与えていた。C 及び U リッチな配列は Tract of OligoPyrimidine (TOP) 配列として知られ、詳細な分子機構は分かっていないが mTOR 経路による翻訳制御に関わることが示唆されている (Iadevaia et al. 2008, Xiong and Sheen 2014)。実際に TOP 配列を持つ mRNA の一部からの翻訳は、花器の形成過程 (Tzeng et al. 2009) や成長因子 (Jiménez-López et al. 2011)、オーキシンの存在下 (Schepetilnikov et al. 2013) では活性化され、動物細胞での低酸素ストレス条件下 (Miloslavski et al. 2014) では抑制されることが報告されている。C 及び U リッチな配列は、mTOR 経路などを通して、十分に発達した器官である展開葉や熱ストレス下における特異的な翻訳機構に関わっているのかもしれない。また、Cap 構造側の A 含量に関連するクラスターは、熱ストレス時にのみ特異的に翻訳状態に正の影響を与えていた。植

物培養細胞における熱ストレスによる翻訳状態の変化を解析した研究では、熱ストレスによる翻訳の抑制に関わる重要な特徴として、Cap 構造から主に 34 nt 内の重要領域が示されている (Matsuura et al. 2013)。そして、その領域で熱ストレス下でも高い翻訳を維持する最適な配列は A リッチであることが示されており、今回の結果と一致している。この熱ストレス下での翻訳の抑制に関わる機構は明らかになっていないが、熱ストレス下では何らかの翻訳開始因子や RNA 結合因子といったトランス因子が Cap 構造側の A リッチな配列を認識し、高い翻訳状態を保つ機構が存在していると考えられる。また uORF や 5'UTR の長さに関連したクラスターは、37°C 以外では負の影響を与えており特に 2DAG と未展開葉での影響が大きかった。uORF は下流に存在する本来の ORF からの翻訳を抑制することが知られており、eIF2 α のリン酸化状態がその制御に影響を与えることが報告されている。例えば、TBF1 遺伝子は成長している植物体では uORF によって本来の AUG からの翻訳が抑えられているが、病原体の進入時には eIF2 α のリン酸化を介して本来の AUG からの翻訳が行われる (Pajerowska-Mukhtar et al. 2012)。また酵母の GCN4 遺伝子の翻訳も eIF2 α のリン酸化と uORF によって制御され、通常時は翻訳が抑えられるがアミノ酸の飢餓状態では翻訳が活性化される (Hinnebusch 2005)。通常の植物体、特に 2DAG と未展開葉でも uORF を持つ mRNA の翻訳状態は、eIF2 α のリン酸化を介して uORF による特異的な翻訳機構で制御されているのかもしれない。これらの結果は、植物 mRNA の翻訳状態は、今回示された各 mRNA の配列的特徴に基づいて、共通する翻訳機構と、主に条件特異的に変化するトランス因子の量や修飾状態によって制御される複数の特異的な翻訳機構によって複雑に制御され決定されていることを示唆している。本研究は、これまで報告および推察されてきた様々な配列的特徴について、各条件での PLS モデルを構築することで複合的に解析し、これまでは不明瞭だった、それぞれの条件での翻訳状態の決定に真に重要な配列的特徴を示すとともに、条件間でのそれら配列的特徴の寄与度の違いを明らかにした。これらの知見は、各 mRNA の翻訳状態を決定しているその分子機構の解明に非常に重要な情報になる。

2-4-3. TSS の収束/分散と mRNA の翻訳状態

CAGE によって明らかになった TSS の分散による 5'UTR 配列バリエーションの存在や、その TSS の変化による 5'UTR 配列の変化は、今回の PLS モデルによって示された重要な配列的特徴の有無や特徴量に影響を与え、翻訳状態に影響を及ぼしている可能性が考えられる。5'UTR 配列バリエーションが多く存在する場合、一般的には 5'UTR 配列の重要な配列的特徴の量がバリエーション間でばらけることによって、それに依存した様々な翻訳状態を示す mRNA が存在し、その遺伝子単位での翻訳状態は平均化され常に一般的な翻訳状態をとるとも考えられる。実際に、非常にわずかながら翻訳状態が非常に高い、または低いといった極端な挙動を示す遺伝子は、他の遺伝子と比べ TSS が収束している傾向があった (Fig. 20A)。また過去の研究より、条件に

よって TSS が変化し 5'UTR の配列的特徴の有無や特徴量が変わることによって、結果として特異的に翻訳状態が制御されている遺伝子が哺乳動物や酵母において報告されている (Law et al. 2005, Brar et al. 2012, Badhai et al. 2011)。しかし興味深いことに、葉の発達段階間での TSS の変化と翻訳状態変化の関係を解析した結果からは、葉の発達段階間で翻訳状態が大きく異なる遺伝子の TSS は、他の遺伝子に比べ変化していないことが示された (Fig. 20B)。このような結果となった理由には、植物の場合、TSS が変化する遺伝子は TATA-less プロモーターが多く TSS が分散するものが多いため、翻訳状態が平均化されることで明確な変化を示しにくいからではないかと考えられる。この結果は、一般的に未展開葉と展開葉の翻訳状態の違いは、TSS の変化によって生じる 5'UTR の配列的特徴の変化ではなく、5'UTR 上に存在している特異的な翻訳機構に関わる配列的特徴に依存していると考えられる。

総括

本研究は、遺伝子の発現制御に重要であることが知られる「翻訳」について、その制御に非常に重要な 5'UTR の配列に着目し、植物 mRNA の翻訳機構への理解を深める目的で行ったものである。

まず第一章では、植物 mRNA の翻訳機構を理解するための基盤情報となる翻訳状態を解析し、遺伝子発現における翻訳段階の重要性を示すとともに、翻訳状態を決定している翻訳機構を考察した。具体的には、シロイヌナズナの各成長・発達段階での翻訳状態をゲノムワイドに解析し、翻訳状態は mRNA 種によって大きく異なり、活発な翻訳が行われていると考えられる mRNA から、あまり翻訳が行われていないものまで幅広く存在していることを明らかにした。また、各 mRNA 種の翻訳状態は、成長・発達段階間で変わらないものから大きく変化するものまで様々であり、その発達段階に応じて各 mRNA の翻訳状態が決定されている可能性を示した。この様な各 mRNA の翻訳状態の違いを、タンパク質機能の観点から解析し、翻訳段階での制御の生理的な重要性の一端を示した。加えて、これまでに当研究室で解析された培養細胞の環境ストレス下等も含めた様々な条件間での翻訳状態の比較から、各条件での翻訳状態を制御している機構について考察し、様々な条件で共通する翻訳機構と、特定の状況に応答した特異的な翻訳機構の存在を示唆した。

続く第二章では、mRNA の翻訳状態の決定に中心的な役割を担っていると考えられている 5'UTR 配列に焦点を当て、植物 mRNA の翻訳状態を決定している分子機構の理解を目指した。まずは CAGE を行い、様々な条件での正確な 5'UTR の配列を決定し、その多様性と分布を調べた。次に翻訳状態と 5'UTR 配列の 2 つのゲノムワイドデータを用いた PLS モデルの構築を行い、翻訳状態の決定機構に関わる具体的な 5'UTR 上の配列的特徴の解明を行った。構築した PLS モデルからいくつかの重要な配列的特徴が明らかとなり、構築したモデルは、翻訳状態を十分に説明可能であることを示した。また、構築した PLS モデルは、実際のタンパク質の生産効率も十分に説明できることを一過性発現実験によって検証し、その高い信頼性を確認した。加えて、様々な条件の翻訳状態に関する PLS モデルを構築し、それぞれの条件で重要な配列的特徴を示すとともに、その配列的特徴を比較することで、各条件下で翻訳状態を決定している機構について考察し、5'末端の二次構造が共通の翻訳機構に関わること、5'UTR の CU リッチや A リッチな配列などが特定の条件に応答した特異的な翻訳機構に関わることを示唆した。

本研究の結果は、様々な条件での植物 mRNA の翻訳状態について、それを決定している配列的特徴を明らかにし、各条件でのそれら配列的特徴の寄与度の違いを示すとともに、様々な条件での共通または特異的な翻訳機構に特徴づけた。本研究の結果は、植物 mRNA の翻訳機構への理解を大きく進歩させ、将来の翻訳制御の分子機構の解明に非常に重要な情報となることが期待される。

今後の発展としては、様々な条件や発達段階、器官、細胞型での植物について、リボソームフットプリント法のような mRNA の翻訳状態の高解像度解析や、ポリソーム/CAGE 法のような各 5'UTR バリエーションについての翻訳状態解析による大規模かつ詳細な情報の蓄積が望まれる。より大規模かつ詳細な情報があれば、翻訳状態に影響を与えている配列的特徴をより詳細に解析することも可能になり、本研究で明らかにした配列的特徴のより具体的な配列パターンや重要な位置を示すことや、翻訳制御に関わる新たな配列的特徴を見出すことも可能になるだろう。

最後に、本研究を別の視点から捉えてみると、外来遺伝子の発現を制御する上でも非常に有用であると言える。本研究によって構築した PLS モデルは、配列から翻訳状態を予測できるだけでなく、目的とする翻訳状態に適した 5'UTR の配列を設計し、タンパク質の生産量を操作できる可能性を秘めている。例えば、有用物質（生分解性プラスチック、植物性油脂、医療用タンパク質など）の大量生産のために、目的の外来遺伝子を効率的に翻訳できる 5'UTR 配列を、モデルによる翻訳状態の予測値から選抜でき、更に 5'UTR 配列の最適化を行うことで、目的の遺伝子に適した高発現系の開発が行える可能性がある。この可能性の一端は、第二章の一過性発現実験による PLS モデルの検証において、配列置換によって基となった配列よりも翻訳効率を大きく上昇させることに成功した置換配列 SS3_ P-Multi と SS4_ P-Multi の結果によって示されている (Fig. 28)。

謝辞

本研究を遂行するにあたり、御指導、御鞭撻を賜りました出村拓教授に厚く御礼申しあげます。加藤晃助教には大変お忙しい中、直接の懇切なる御指導ならびに格別なる御高配を賜り深く御礼申し上げます。また、久保稔特任准教授、大谷美沙都助教、米田新助教、情報科学研究科の金谷重彦教授、大阪大学の松浦秀幸助教、埼玉大学の山口雅利准教授には折々に御指導ならびに適切且つ有益な御助言を頂き深く感謝いたします。また、梅田正明教授、高山誠司教授には、本研究のアドバイザーとして適切且つ有益な御助言を頂き心から感謝いたします。

植物代謝調節学講座の皆様にも本当にお世話になりました。特に上田清貴博士には研究における御指導のみならず、多岐に渡ってお世話になりました。厚く御礼申し上げます。田中雅恵氏、大平菜緒氏、原田麻記氏には、事務的な面で何かとお世話になりました。津田貴子氏、金城聖子氏には試薬類の作製や培養細胞の継代など、大変お世話になりました。厚く御礼申し上げます。ここにすべての方の名前を挙げることはできませんが、佐野亮輔博士、大河原錬也氏、矢村寿啓氏、西垣直哉氏、畑健介氏、横河由樹子氏、今瀬諒司氏、佐藤伊純氏には私の至らないところもあり、迷惑をかけることもありましたが、研究の遂行にあたって様々な面でお世話になりました。植物代謝調節学講座の皆様の御指導、御助言、御協力等に対し、心から御礼申し上げます。皆様のお力添えや励まし無くして本論文の完成はあり得ませんでした。

最後に、家族にはいつも私の意思を尊重していただき、暖かく応援してもらいました。この場を借りて深く御礼申し上げます。ありがとうございました。

研究成果の公表

本研究の成果は、筆頭著者として執筆した論文と、複数の学会にて発表している。また、関連する研究を、共著論文として発表するとともに著書も書いている。以下に内容を記す。

論文（筆頭著者）

Yamasaki, S., Matsuura, H., Demura, T. and Kato, K. (2015) Changes in Polysome Association of mRNA Throughout Growth and Development in *Arabidopsis thaliana*. *Plant Cell Physiol.* 56:2169-2180

論文（共同著者）

Matsui, T., Matsuura, H., Sawada, K., Takita, E., Kinjo, S., Takenami, S., Ueda, K., Nishigaki, N., Yamasaki, S., Hat, a K., Yamaguchi, M., Demura, T. and Kato, K. (2012) High level expression of transgenes by use of 5'-untranslated region of the *Arabidopsis thaliana arabinogalactan-protein 21* gene in dicotyledons. *Plant Biotechnol.* 29: 319-322

Ueda, K., Okawara, R., Yamasaki, S., Sanada, Y., Kinoshita, E., Yoneda, A., Demura, T. and Kato K. (2014) Efficient transgene expression by alleviation of translational repression in plant cells. *J. Biosci. Bioeng.* 118:434-440

学会発表

シロイヌナズナ植物体の成長に伴う翻訳状態変化の解析
山崎 将太郎, 上田 清貴, 米田 新, 出村 拓, 加藤 晃
第 30 回日本植物細胞分子生物学会大会 2012 年 8 月 5 日

成長と発達に伴う翻訳状態変化のゲノムワイドな解析
山崎 将太郎, 上田 清貴, 米田 新, 出村 拓, 加藤 晃
第 476 回日本農芸化学会関西支部大会 2012 年 9 月 29 日

植物体の成長／発達に伴う翻訳状態変化の解析
山崎 将太郎, 上田 清貴, 米田 新, 出村 拓, 加藤 晃
第 54 回日本植物生理学会年会 2013 年 3 月 21 日

植物体の成長／発達に伴う翻訳レベルでの遺伝子発現変化
山崎 将太郎, 上田 清貴, 米田 新, 出村 拓, 加藤 晃

第 31 回日本植物細胞分子生物学会大会 2013 年 9 月 19 日

植物の成長や発達に伴う翻訳制御の解析

山崎 将太郎, 上田 清貴, 出村 拓, 加藤 晃

第 32 回日本植物細胞分子生物学会大会 2014 年 8 月 21 日

植物の成長・発達に伴う翻訳状態変化に関わる 5' UTR の配列的特徴の解析

山崎 将太郎, 上田 清貴, 出村 拓, 加藤 晃

生命情報科学若手の会 第 6 回研究会 2014 年 10 月 29 日

5'UTR を介した植物翻訳制御機構の解明 (CAGE データの活用)

山崎 将太郎, 出村 拓, 加藤 晃

理研公開シンポジウム

CAGE 法を用いた新たなトランスクリプトーム解析事例と実際 2014 年 11 月 19 日

in silico による植物 mRNA の翻訳状態を規定する 5'UTR の配列的特徴の解析

山崎 将太郎, 出村 拓, 加藤 晃

NGS 現場の会 第 4 回研究会 2015 年 7 月 2 日~3 日

mRNA からの翻訳効率を改良した導入遺伝子高発現系

山崎 将太郎, 出村 拓, 加藤 晃

第 67 回生物工学会大会 2015 年 10 月 28 日

翻訳状態を決定する 5'UTR の配列的特徴

山崎 将太郎, 出村 拓, 加藤 晃

第 57 回日本植物生理学会年会 2016 年 3 月 (予定)

著書

山崎 将太郎, 上田 清貴, 加藤 晃 (2013) 環境ストレスの影響を考慮した植物発現ベクターの開発. *生物工学会誌* 91: 356-360

山崎 将太郎, 加藤 晃 (2014) 植物への有用遺伝子導入. *生物工学会誌* 92: 617-621

参考文献

Akbergenov, R.Zh., Zhanybekova, S.Sh., Kryldakov, R.V., Zhigailov, A., Polimbetova, N.S., Hohn, T. et al. (2004) ARC-1, a sequence element complementary to an internal 18S rRNA segment, enhances translation efficiency in plants when present in the leader or intercistronic region of mRNAs. *Nucleic Acids Res.* 32:239-47.

Arribere, J.A. and Gilbert, W.V. (2013) Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Res.* 23: 977-987.

Axelos, M., Curie, C., Mazzolini, L., Bardet, C. and Lescure, B. (1992) A protocol for transient gene expression in *Arabidopsis thaliana* protoplasts isolated from cell suspension cultures. *Plant Physiol. Biochem.* 30: 123-128.

Babendure, J.R., Babendure, J.L., Ding, J.H. and Tsien, R.Y. (2006) Control of mammalian translation by mRNA structure near caps. *RNA.* 12: 851-861.

Badhai, J., Schuster, J., Gidlöf, O. and Dahl, N. (2011) 5'UTR variants of ribosomal protein S19 transcript determine translational efficiency: implications for Diamond-Blackfan anemia and tissue variability. *PLoS One* 6: e17672.

Baena-González, E., Rolland, F., Thevelein, J.M. and Sheen, J. (2007) A central integrator of transcription networks in plant stress and energy signalling. *Nature* 448: 938–942.

Baena-González, E. and Sheen, J. (2008) Convergent energy and stress signaling. *Trends Plant Sci.* 13: 474–482.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L. et al. (2015) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37: W202-W208.

Beltrán-Peña, E., Ortíz-López, A. and Sánchez de Jiménez, E. (1995) Synthesis of ribosomal proteins from stored mRNAs early in seed germination. *Plant Mol. Biol.* 28: 327–336 .

Branco-Price, C., Kawaguchi, R., Ferreira, R.B. and Bailey-Serres, J. (2005) Genome-wide analysis of transcript abundance and translation in *Arabidopsis* seedlings subjected to oxygen deprivation. *Ann. Bot.* 96: 647–660.

- Branco-Price, C., Kaiser, K.A., Jang, C.J., Larive, C.K. and Bailey-Serres, J. (2008) Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in *Arabidopsis thaliana*. *Plant J.* 56: 743–755.
- Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T. and Weissman, J.S. (2012) High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. *Science* 335: 552-557.
- Browning, K.S. (2004) Plant translation initiation factors: it is not easy to be green. *Biochem. Soc. Trans.* 32: 589-591.
- Bush, M.S., Crowe, N., Zheng, T. and Doonan, J.H. (2015) The RNA helicase, eIF4A-1, is required for ovule development and cell size homeostasis in *Arabidopsis*. *Plant J.* 84: 989-1004.
- Dansako, T., Kato, K., Satoh, J., Sekine, M., Yoshida, K. and Shinmyo, A. (2003) 5' Untranslated region of the *HSP 18.2* gene contributes to efficient translation in plant cells. *J. Biosci. Bioeng.* 95: 52-58.
- Davies, E., and Abe, S. (1995) Methods for isolation and analysis of polyribosomes. *Methods Cell Biol.* 50: 209-222.
- Depeiges, A., Degroote, F., Espagnol, M.C. and Picard, G. (2006) Translation initiation by non-AUG codons in *Arabidopsis thaliana* transgenic plants. *Plant Cell Rep.* 25:55-61.
- Galland, M., Huguet, R., Arc, E., Cueff, G., Job, D. and Rajjou, L. (2014) Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during *Arabidopsis* seed germination. *Mol. Cell. Proteomics.* 13: 252-268.
- Gebauer, F. and Hentze, M.W. (2004) Molecular mechanisms of translational control. *Nat. Rev. Mol. Cell Biol.* 5: 827-835.
- Gingras, A.C., Raught, B., and Sonenberg, N. (1999) eIF4F initiation factors: Effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem.* 68: 913-963.
- Hinnebusch, A.G. (2005) Translational regulation of GCN4 and the general amino acid control of yeast. *Annu. Rev. Microbiol.* 59: 407-450.

- Huang, B. and Xu, C. (2008) Identification and characterization of proteins associated with plant tolerance to heat stress. *J. Integr. Plant Biol.* 50: 1230–1237.
- Hulzink, R.J., de Groot, P.F., Croes, A.F., Quaedvlieg, W., Twell, D., Wullems, G.J., et al. (2002) The 5'-untranslated region of the *ntp303* gene strongly enhances translation during pollen tube growth, but not during pollen maturation. *Plant Physiol.* 129: 342-353.
- Iadevaia, V., Caldarola, S., Tino, E., Amaldi, F. and Loreni, F. (2008) All translation elongation factors and the e, f, and h subunits of translation initiation factor 3 are encoded by 5'-terminal oligopyrimidine (TOP) mRNAs. *RNA.* 14: 1730-1736.
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R. and Weissman, J.S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324: 218-223.
- Jiao, Y. and Meyerowitz, E.M. (2010) Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Mol. Syst. Biol.* 6: 419.
- Jiménez-López, S., Mancera-Martínez, E., Donayre-Torres, A., Rangel, C., Uribe, L., March, S., et al. (2011) Expression Profile of Maize (*Zea mays L.*) Embryonic Axes During Germination: Translational Regulation of Ribosomal Protein mRNAs. *Plant Cell Physiol.* 52: 1719–1733.
- Juntawong, P. and Bailey-Serres, J. (2012) Dynamic light regulation of translation status in *Arabidopsis thaliana*. *Front Plant Sci.* 3: 66.
- Kadonaga, J.T. (2012) Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip. Rev. Dev. Biol.* 1: 40-51.
- Kawaguchi, R., Girke, T., Bray, E.A. and Bailey-Serres, J. (2004) Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in *Arabidopsis thaliana*. *Plant J.* 38: 823-839.
- Kawaguchi, R. and Bailey-Serres, J. (2005) mRNA sequence features that contribute to translational regulation in *Arabidopsis*. *Nucleic Acids Res.* 33: 955-965.
- Koloteva, N., Müller, P.P., McCarthy, J.E. (1997) The position dependence of translational regulation via RNA-RNA and RNA-protein interactions in the 5'-untranslated region of

eukaryotic mRNA is a function of the thermodynamic competence of 40 S ribosomes in translational initiation. *J. Biol. Chem.* 272: 16531-16539.

Komar, A.A., Mazumder, B. and Merrick, W.C. (2012) A new framework for understanding IRES-mediated translation. *Gene.* 502 :75-86.

Koroleva, O.A., Calder, G., Pendle, A.F., Kim, S.H., Lewandowska, D., Simpson, C.G. et al (2009) Dynamic behavior of Arabidopsis eIF4A-III, putative core protein of exon junction complex: fast relocation to nucleolus and splicing speckles under hypoxia. *Plant Cell.* 21: 1592-606.

Kovtun, Y., Chiu, W.L., Tena, G. and Sheen, J. (2000) Functional analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. *Proc. Natl. Acad. Sci. USA* 97: 2940-2945.

Law, G.L., Bickel, K.S., MacKay, V.L. and Morris, D.R. (2005) The undertranslated transcriptome reveals widespread translational silencing by alternative 5' transcript leaders. *Genome Biol.* 6: R111.

Le, H., Browning, K.S., Gallie, D.R. (1998) The phosphorylation state of the wheat translation initiation factors eIF4B, eIF4A, and eIF2 is differentially regulated during seed development and germination. *J. Biol. Chem.* 273: 20084-20089.

Liu, M.J., Wu, S.H., Chen, H.M. and Wu, S.H. (2012) Widespread translational control contributes to the regulation of Arabidopsis photomorphogenesis. *Mol. Syst. Biol.* 8: 566.

Liu, M.J., Wu, S.H., Wu, J.F., Lin, W.D., Wu, Y.C., Tsai, T.Y., et al. (2013) Translational Landscape of Photomorphogenic Arabidopsis. *Plant Cell* 25: 3699-3710.

Lorenz, R., Bernhart, S.H., Höner-Zu-Siederdisen, C., Tafer, H., Flamm, C., Stadler, P.F. et al. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6:26.

Maier, T., Güell, M. and Serrano, L. (2009) Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* 583: 3966-3973.

Majeran, W., Friso, G., Ponnala, L., Connolly, B., Huang, M., Reidel, E., et al. (2010) Structural and metabolic transitions of C4 leaf development and differentiation defined by microscopy and quantitative proteomics in maize. *Plant Cell* 22: 3509-3542.

Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M. and Turner, D. H.

(2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA* 101: 7287-7292.

Matsui, T., Matsuura, H., Sawada, K., Takita, E., Kinjo, S., Takenami, S. et al. (2012) High level expression of transgenes by use of 5'-untranslated region of the *Arabidopsis thaliana arabinogalactan-protein 21* gene in dicotyledons. *Plant Biotechnol.* 29: 319-322.

Matsuura, H., Shinmyo, A. and Kato, K. (2008) Preferential translation mediated by *Hsp81-3* 5'-UTR during heat shock involves ribosome entry at the 5'-end rather than an internal site in *Arabidopsis* suspension cells. *J. Biosci. Bioeng.* 105: 39-47.

Matsuura, H., Ishibashi, Y., Shinmyo, A., Kanaya, S. and Kato, K. (2010) Genome-Wide Analyses of Early Translational Responses to Elevated Temperature and High Salinity in *Arabidopsis thaliana*. *Plant Cell Physiol.* 51: 448-462.

Matsuura, H., Takenami, S., Kubo, Y., Ueda, K., Ueda, A., Yamaguchi, M., et al. (2013) A computational and experimental approach reveals that the 5'-proximal region of the 5'-UTR has a Cis-regulatory signature responsible for heat stress-regulated mRNA translation in *Arabidopsis*. *Plant Cell Physiol.* 54: 474-483.

Melamed, D. and Arava, Y. (2007) Genome - Wide Analysis of mRNA Polysomal Profiles with Spotted DNA Microarrays. *Methods Enzymol.* 431: 177-201.

Miloslavski, R., Cohen, E., Avraham, A., Iluz, Y., Hayouka, Z., Kasir, J. et al. (2014) Oxygen sufficiency controls TOP mRNA translation via the TSC-Rheb-mTOR pathway in a 4E-BP-independent manner. *J. Mol. Cell Biol.* 6: 255-266.

Morton, T., Petricka, J., Corcoran, D.L., Li, S., Winter, C.M., Carda, A., et al. (2014) Paired-end analysis of transcription start sites in *Arabidopsis* reveals plant-specific promoter signatures. *Plant Cell* 26: 2746-2760.

Murata, M., Nishiyori-Sueki, H., Kojima-Ishiyama, M., Carninci, P., Hayashizaki, Y. and Itoh, M. (2014) Detecting expressed genes using CAGE. *Methods Mol. Biol.* 1164:67-85.

Nagata, T., Nemoto, Y., and Hasezawa, S. (1992) Tobacco BY-2 cell line as the "HeLa" cell in the cell biology of higher plants. *Int. Rev. Cytol.* 132: 1-30.

Nguyen, H.M., Schippers, J.H., Gõni-Ramos, O., Christoph, M.P., Dortay, H., van der Hoorn, R.A., et al. (2013) An upstream regulator of the 26S proteasome modulates organ

size in *Arabidopsis thaliana*. *Plant J.* 74: 25-36.

Nicolai, M., Roncato, M.A., Canoy, A.S., Rouquié, D., Sarda, X., Freyssinet, G., et al. (2006) Large-scale analysis of mRNA translation states during sucrose starvation in *Arabidopsis* cells identifies cell proliferation and chromatin structure as targets of translational control. *Plant Physiol.* 141: 663–673.

Ohta, T., Matsuoka, H., Nomura, Y. and Tozawa, Y. (2010) Control of translational initiation in the wheat-embryo cell-free protein expression system for producing homogenous products. *Protein Expr Purif.* 73:15-22.

Pajeroska-Mukhtar, K.M., Wang, W., Tada, Y., Oka, N., Tucker, C.L., Fonseca, J.P., et al. (2012) The HSF-like transcription factor TBF1 is a major molecular switch for plant growth-to-defense transition. *Curr. Biol.* 22:103-112.

Proud, C.G. (2007) Signalling to translation: how signal transduction pathways control the protein synthetic machinery. *Biochem. J.* 403: 217-234.

Rojas-Duran, M.F. and Gilbert, W.V. (2012) Alternative transcription start site selection leads to large differences in translation activity in yeast. *RNA* 18: 2299-2305.

Rosado, A., Li, R., van de Ven, W., Hsu, E. and Raikhel, N.V. (2012) *Arabidopsis* ribosomal proteins control developmental programs through translational regulation of auxin response factors. *Proc. Natl. Acad. Sci. USA* 109: 19537-19544.

Roy, B. and von Arnim, A.G. (2013) Translational Regulation of Cytoplasmic mRNAs. *Arabidopsis Book* 11: e0165.

Sahi, C., Singh, A., Blumwald, E. and Grover, A. (2006) Beyond osmolytes and transporters: novel plant salt-stress-tolerance-related genes from transcriptomic profiling data. *Physiol. Plant.* 127: 1–9.

Satoh, J., Kato, K. and Shinmyo, A. (2004) The 5'-untranslated region of the tobacco *alcohol dehydrogenase* gene functions as an effective translational enhancer in plant. *J. Biosci. Bioeng.* 98:1-8.

Schepetilnikov, M., Dimitrova, M., Mancera-Martínez, E., Geldreich, A., Keller, M. and Ryabova, L.A. (2013) TOR and S6K1 promote translation reinitiation of uORF-containing mRNAs via phosphorylation of eIF3h. *EMBO J.* 32: 1087-1102.

Schneider, R., Agol, V.I., Andino, R., Bayard, F., Cavener, D.R., Chappell, S.A., Chen, J.J. et al. (2001) New ways of initiating translation in eukaryotes. *Mol. Cell. Biol.* 21: 8238-8246.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011) Global quantification of mammalian gene expression control. *Nature* 473: 337–342.

Sonenberg, N. and Hinnebusch, A.G. (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136: 731-745.

Sonoda, Y., Sako, K., Maki, Y., Yamazaki, N., Yamamoto, H., Ikeda, A., et al. (2009) Regulation of leaf organ size by the Arabidopsis RPT2a 19S proteasome subunit. *Plant J.* 60: 68-78.

Sugio, T., Matsuura, H., Matsui, T., Matsunaga, M., Noshio, T., Kanaya et al. (2010) Effect of the sequence context of the AUG initiation codon on the rate of translation in dicotyledonous and monocotyledonous plant cells. *J. Biosci. Bioeng.* 109: 170-3.

Takahashi, H., Lassmann, T., Murata, M. and Carninci, P. (2012) 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nat. Protoc.* 7: 542-561.

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., et al. (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* 37: 914-939.

Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, NS. and Sabatini D.M. (2012) A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* 485: 109-113.

Tranque, P., Hu, M.C., Edelman, G.M. and Mauro, V.P. (1998) rRNA complementarity within mRNAs: a possible basis for mRNA-ribosome interactions and translational control. *Proc. Natl. Acad. Sci. USA* 95:12238-12243.

Tzeng, T.Y., Kong, L.R., Chen, C.H., Shaw, C.C. and Yang, C.H. (2009) Overexpression of the Lily p70^{s6k} Gene in Arabidopsis Affects Elongation of Flower Organs and Indicates TOR-Dependent Regulation of AP3, PI and SUP Translation. *Plant Cell Physiol.* 50: 1695–1709.

Ueda, K., Okawara, R., Yamasaki, S., Sanada, Y., Kinoshita, E., Yoneda, A. et al. (2014) Efficient transgene expression by alleviation of translational repression in plant cells. *J. Biosci. Bioeng.* 118: 434-440.

Ueda, K., Matsuura, H., Yamaguchi, M., Demura, T. and Kato, K.. (2012) Genome-wide analyses of changes in translation state caused by elevated temperature in *Oryza sativa*. *Plant Cell Physiol.* 53:1481-1491.

Usadel, B., Nagel, A., Thimm, O., Redestig, H., Blaesing, O.E., Palacios-Rojas, N., et al. (2005) Extension of the Visualization Tool MapMan to Allow Statistical Analysis of Arrays, Display of Corresponding Genes, and Comparison with Known Responses. *Plant Physiol.* 138: 1195-1204.

Usadel, B., Nagel, A., Steinhauser, D., Gibon, Y., Bläsing, O.E., Redestig, H., et al. (2006) PageMan: An interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. *BMC Bioinformatics* 7: 535.

Vanderhaeghen, R., De-Clercq, R., Karimi, M., Van-Montagu, M., Hilson, P. and Van-Lijsebettens, M. (2006) Leader sequence of a plant ribosomal protein gene with complementarity to the 18S rRNA triggers in vitro cap-independent translation. *FEBS Lett.* 580: 2630-2636.

Vega-Laso, M.R., Zhu, D., Sagliocco, F., Brown, A.J., Tuite, M.F. and McCarthy, J.E. (1993) Inhibition of translational initiation in the yeast *Saccharomyces cerevisiae* as a function of the stability and position of hairpin structures in the mRNA leader. *J. Biol. Chem.* 268: 6453-62.

Wiese, A., Elzinga, N., Wobbes, B. and Smeekens, S.. (2004) A conserved upstream open reading frame mediates sucrose-induced repression of translation. *Plant Cell* 16: 1717-1729.

Xiong, Y. and Sheen, J. (2014) The role of target of rapamycin signaling networks in plant growth and metabolism. *Plant Physiol.* 164:499-512.

Yamamoto, Y.Y., Yoshitsugu, T., Sakurai, T., Seki, M., Shinozaki, K. and Obokata, J. (2009) Heterogeneity of Arabidopsis core promoters revealed by high-density TSS analysis. *Plant J.* 60: 350-362.

Yamana, R., Iwasaki, M., Wakabayashi, M., Nakagawa, M., Yamanaka, S. and Ishihama, Y. (2013) Rapid and Deep Profiling of Human Induced Pluripotent Stem Cell Proteome by One-shot NanoLC–MS/MS Analysis with Meter-scale Monolithic Silica Columns. *J. Proteome Res.* 12: 214–221.

Yángüez, E., Castro-Sanz, A.B., Fernández-Bautista, N., Oliveros, J.C. and Castellano, M.M. (2013) Analysis of Genome-Wide Changes in the Translatome of Arabidopsis Seedlings Subjected to Heat Stress. *PLoS One* 8: e71425.