

委託プロジェクト研究
「ゲノム情報を活用した農産物の次世代生産基盤技術の開発」
平成29年度 最終年度報告書

13406471

遺伝子発現を指標にイネの生育を予測するシステムの開発

研究実施期間	平成25年度～平成29年度（5年間）
代表機関	国立研究開発法人 農業・食品産業技術総合研究機構 次世代作物開発研究センター
研究開発責任者	川原 善浩
共同研究機関	国立大学法人 東京大学大学院農学生命科学研究科
	富山県農林水産総合技術センター 農業研究所
	福岡県農林業総合試験場 農産部
	宮城県古川農業試験場 作物育種部
	国立研究開発法人 農業・食品産業技術総合研究機構（北海道農業研究センター、西日本農業研究センター）
	国立大学法人 京都大学農学研究科
普及・実用化支援組織	
研究開発責任者 連絡先	TEL : 029-838-7065 FAX : 029-838-7065 E-mail : y.kawahara@affrc.to.jp

<別紙様式3. 最終年度報告書>

I-1. 年次計画

研究課題	研究年度					担当研究機関・研究室	
	25	26	27	28	29	機関	研究室
1. 全国で活用できる遺伝子発現予測システムの開発と農業形質予測への展開							
(1) イネの栽培とモデル構築用と検証用サンプリング							
① 北海道・札幌	←				→	農研機構 北農研	水稲育種グループ
② 宮城県・古川	←				→	宮城県・古川農試	作物育種部
③ 富山県・富山	←				→	富山県・農技セ	農業バイオセンター
④ 茨城県・つくば	←				→	農研機構 次世代作物開発研究センター	フィールドオミクスユニット
⑤ 大阪府・高槻	←				→	京都大学	植物生産管理
⑥ 広島県・福山	←				→	農研機構・西日本	水稲育種グループ
⑦ 福岡県・福岡	←				→	福岡県農試	水稲育種チーム
(2) マイクロアレイ解析	←				→	農研機構 次世代作物開発研究センター	フィールドオミクスユニット
(3) Nagano et al.式モデリングの自動化	←				→	東京大学	育種学研究室
(4) 二環境因子対応モデルへの改良	←				→	東京大学	育種学研究室
(5) 遺伝子発現の分類、モデルの高度化	←				→	東京大学	育種学研究室
(6) 農業形質と対応付けしたサンプリング	←				→	東京大学	育種学研究室
(7) 農業形質と関連の深い遺伝子発現情報の探索	←				→	東京大学	育種学研究室

研究課題	研究年度					担当研究機関・研究室	
	25	26	27	28	29	機関	研究室
2. RNA-Seq 法による品種間トランスクリプトーム比較解析手法の開発							
(1) コシヒカリやコシヒカリ系地域品種のゲノム解読	←	→				農研機構 次世代作物開発研究センター	情報解析ユニット
(2) RNA-Seq法による多品種に対応した高精度な遺伝子発現解析手法の開発		←	→			農研機構 次世代作物開発研究センター	情報解析ユニット
(3) RNA-Seq解析とマイクロアレイ解析による遺伝子発現プロファイルの比較と評価	←			→		農研機構 次世代作物開発研究センター	情報解析ユニット
(4) コシヒカリとコシヒカリ系地域品種の遺伝子発現プロファイル比較解析		←			→	農研機構 次世代作物開発研究センター	情報解析ユニット
3. 遺伝子発現予測システムの高度化・汎用化のための基盤ソフトウェアシステム整備							
(1) 気象データの整備	←	→				京都大学	生態学研究センター
(2) RNA-Seqライブラリ調整の多検体・低コスト化	←	→				京都大学	生態学研究センター
(3) モデリング手法の改良、高速化	←	→				京都大学	生態学研究センター

I-2. 実施体制

課題 番号	研究項目	担当研究機関・研究室			研究担当者
		機関		研究室	
PFT 1001	全国で活用できる遺 伝子発現予測システ ムの開発と農業形質 予測への展開	農研機構 次 世代作物開発 研究センター	基盤研究領域	情報解析ユニ ット	◎ 川原 善浩
		東京大学	大学院農学生 命科学研究科 /農学部	生産・環境生 物学専攻 育 種学研究室	○ 井澤 毅 (H25～H29)
		宮城県古川農 業試験場	作物育種部		遠藤 貴司 (H25～H29)
		宮城県古川農 業試験場	作物育種部		佐伯 研一 (H26～H27)
		宮城県古川農 業試験場	作物育種部		中込 佑介 (H26～H29)
		宮城県古川農 業試験場	作物育種部	育種班	佐藤 浩子 (H26～H27、 H29)
		宮城県古川農 業試験場	作物育種部	育種班	石森 裕貴 (H28～H29)
		京都大学	大学院農学研 究科	附属農場	北島 宣 (H29)
		京都大学	農学研究科	農学専攻植物 生産管理学分 野	齊藤 大樹 (H25～H28)
		農研機構 次 世代作物開発 研究センター	基盤研究領域	フィールドオ ミクスユニッ ト	伊藤 博紀 (H27～H29)
		農研機構 次 世代作物開発 研究センター	基盤研究領域	フィールドオ ミクスユニッ ト	小郷 裕子 (H27～H29)
農研機構 西 日本農業研究 センター	水田作研究領 域	水稻育種グル ープ	中込 弘二 (H25～H29)		
農研機構 北 海道農業研究 センター	作物開発研究 領域	水稻育種グル ープ	梶 亮太 (H28～H29)		
富山県農林水 産総合センタ ー	農業研究所	農業バイオセ ンター	村田 和優 (H25～H29)		

PFT 1002	RNA-Seq法による品 種間トランスクリプ トーム比較解析手法 の開発	富山県農林水産総合技術センター	農業研究所	育種課	蛭谷 武志 (H25～H28)
		富山県農林水産総合技術センター	農業研究所	育種課	山口 琢也 (H25～H29)
		富山県農林水産総合技術センター	農業研究所	育種課	伊山 幸秀 (H25～H29)
		富山県農林水産総合技術センター	農業研究所	育種課	小島 洋一朗 (H29)
		富山県農林水産総合技術センター	農業研究所	農業バイオセンター	尾崎 秀宣 (H27～H29)
		富山県農林水産総合技術センター	農業研究所	農業バイオセンター	荘司 和明 (H28～H29)
		富山県農林水産総合技術センター	農業研究所	農業バイオセンター	藤田 健司 (H25～H26)
		富山県農林水産総合技術センター	農業研究所	農業バイオセンター・育種課	廣川 智子 (H25)
		富山県農林水産総合技術センター	農業研究所	農業バイオセンター・育種課	金田 宏 (H27～H28)
		福岡県農業総合試験場	農産部	水稻育種チーム	和田 卓也 (H25)
		福岡県農林業総合試験場	農産部	水稻育種チーム	宮原 克典 (H26～H29)
		福岡県農林業総合試験場	農産部	水稻育種チーム	山口 修 (H26～H29)
		福岡県農林業総合試験場	農産部	水稻育種チーム	石橋 正文 (H26～H29)
		北海道農業研究センター	寒地作物研究領域		藤野 賢治 (H25～H27)
農研機構 次世代作物開発研究センター	基盤研究領域	情報解析ユニット	○ 川原 善浩 (H25～H29)		

PFT 1003	遺伝子発現予測システムの高度化・汎用化のための基盤ソフトウェアシステム整備	京都大学	生態学研究センター	分子解析生態学研究部門	○ 永野 惇 (H25～H26)
		京都大学	生態学研究センター	分子解析生態学研究部門	工藤 洋 (H25～H26)

(注1) 研究開発責任者には◎、小課題責任者には○、実行課題責任者には△を付すこと。

中課題番号	13406471	研究期間	平成25～29年度
大課題名	ゲノム情報を活用した農産物の次世代生産基盤技術の開発		
中課題名	遺伝子発現を指標にイネの生育を予測するシステムの開発		
代表機関・研究開発責任者名	国立研究開発法人 農業・食品産業技術総合研究機構 次世代作物開発研究センター 川原 善浩		

I-1. 研究目的

ブランド米をはじめとして、日本の食用コメ品種の品質評価基準は厳しい。一方で、一等米比率やいもち病による被害量の年変動は大きく、同じ品種でも、栽培環境の違いによって収量や品質に大きな差が生じており、高度な栽培管理が求められている。近年、圃場で栽培するイネの生育環境（気温や日射量）と生命活動の根幹である遺伝子の働き具合（遺伝子発現）との関係性を明らかにすることで、環境情報から遺伝子発現を予測することが可能になってきた。

このため、本研究では、

1. 全国で活用できる遺伝子発現予測システムの開発と農業形質予測への展開
2. RNA-Seq法による品種間トランスクリプトーム比較解析手法の開発
3. 遺伝子発現予測システムの高度化・汎用化のための基盤ソフトウェアシステム整備により、実用的なレベルでの「環境情報から遺伝子発現を予測する技術」を開発する。また、「出穂や施肥応答などの重要な形質について、遺伝子発現との関係」を明らかにし、高度な栽培管理の実現に資する技術を開発する。また、開発する技術を様々なイネ品種に適用可能とし、より汎用的に利用できるようにするためのソフトウェア等の技術開発を目標とする。

その結果、統計モデリング手法の予測精度の向上と共に、遺伝子発現の動態から生育状況を正確に予測する技術を確立することで、それぞれの栽培環境に適した高度な栽培管理の実現に繋がることが期待される。

I-2. 研究結果

本研究では、遺伝子発現を指標にイネの生育を予測するシステムの開発を目指し、統計モデリング手法を用いた全国で活用可能な高精度な予測システムの開発（PFT1001）を中心に、その技術を様々な品種に展開していくための技術開発（PFT1002）、誰でも容易に利用できるようにするための技術開発（PFT1003）といった課題構成で研究を推進した（図1）。

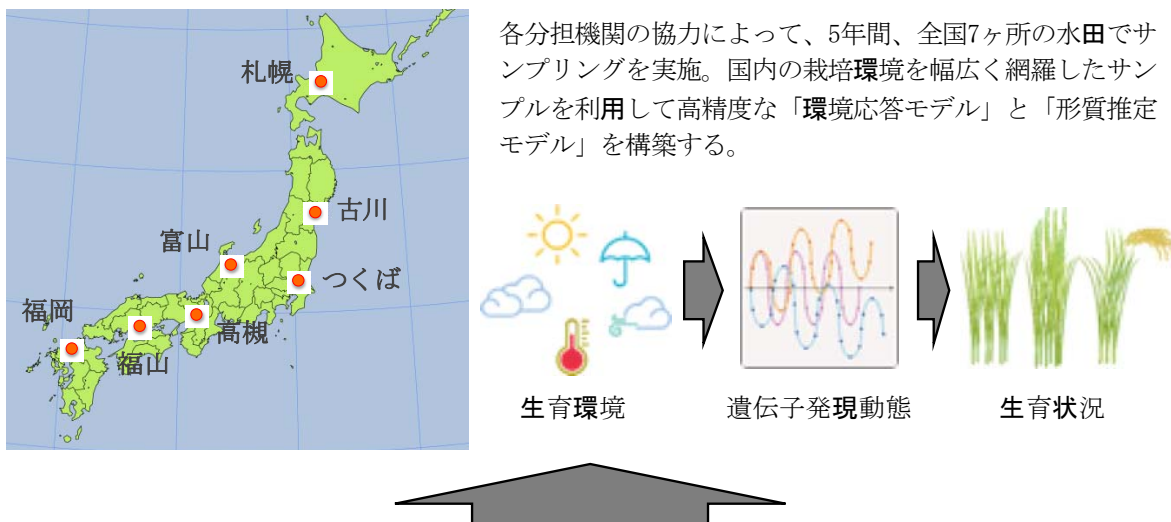
全国で活用できる遺伝子発現予測システムを開発するため、全国7か所でコシヒカリを栽培し、年に2~3回、2時間ごとの葉のサンプリングを行い、マイクロアレイによる遺伝子発現解析を行った。その結果、5年間で合計1,400以上のサンプルについて、約3万遺伝子についての遺伝子発現情報が得られた。また、より高精度に環境情報から遺伝子発現の予測を可能とするため、あらたにAPMC（adaptive population Monte Carlo ABC）法を用いた、二環境要因（日射と気温）を同時に考慮することが可能な「環境情報から遺伝子発現を予測する技術（環境応答モデル）」を開発した。2012年から2014年にかけて全国5ヶ所でサンプリングされたコシヒカリの葉のサンプルのうち、日射と気温の変動データと発現データが紐づく587サンプルを用いて遺伝子ごとにモデルを構築し、2015年と2016年の426サンプルを用いて予測精度の評価を行った。その結果、従来の手法と比べて、非常に高精度な実用レベルの環境応答モデルが構築できた。また、できるだけ少数の遺伝子発現データで重要な農業形質を予測する「形質推定モデル」の構築については、イネの体内時計を利用して、遺伝子発現から植物がもつ時刻情報を推定する技術を開発し、論文発表した（Matsuzaki et al. *Plant Cell* 2015）。さらに、イネの葉の遺伝子発現情報だけから施肥応答や葉内窒素動態、出穂日、移植日を推定できる画期的な形質推定モデルを構築した。

また、上述の環境応答モデルや形質推定モデルをコシヒカリ以外の品種にも適用可能とするため、様々な品種に対応した高精度な遺伝子発現解析を可能とするための技術開発を行った。対象品種のゲノムリシーケンスデータとRNA-Seqデータを用いて高精度な発現解析を行うワークフローを確立し、多くの遺伝子についてはマイクロアレイと非常に相関の高い遺伝子発現プロファイルが得られることを確認した。さらにその手法を用いて、国内6地点で同時に栽培、サンプリングした「コシヒカリ」と「コシヒカリ系地域品種」の葉を対象に品種間トランスクリプトーム解析を行ない、様々な遺伝子についての品種間差情報を取得し、整理した。得られたゲノム情報については、すでに公開されている300品種を超えるイネ品種のゲノムリシーケンスデータと合わせて、同じ解析ワークフローを用いて解析し、ゲノムワイドな多型情報を取得した。これらの多型情報についてはマルチゲノムブラウザ「TASUKE」を用いて、イネアノテーションデータベースRAP-DBと連携させたうえで一般公開した

（<http://rapdb.dna.affrc.go.jp>）。

統計モデリングによるゲノムワイドな遺伝子発現予測技術を様々な作物種などへ展開するため、十分な汎用性・再利用性を確保した解析ソフトウェアプラットフォームを開発した。また、より安価に質の高いトランスクリプトーム解析を実現するため、多検体RNA-Seq解析手法を開発した。

全国で活用できる遺伝子発現予測システムの開発と農業形質予測への展開 (PFT1001)



RNA-Seq法による品種間トランスクリプトーム比較解析手法の開発 (PFT1002)

高度な品種間差解析技術の開発、北海道～九州に栽培適地を持つ品種の品種間差情報の収集
様々な品種に適用可能に

遺伝子発現予測システムの高度化・汎用化のための基盤ソフトウェアシステム整備 (PFT1003)

システムの汎用化・高速化
誰でも容易に利用可能に

図1. 遺伝子発現を指標にイネの生育を予測するシステムの開発 (PFT) の課題構成と開発内容

I-3. 今後の課題

本研究による成果を新しい栽培管理技術に組み込むためには、以下に挙げるような技術開発が必要である。

1. より簡便なサンプリング手法の開発。例えば、サンプリング用の装置を取り付けたドローン等の利用が考えられる。また、技術を広く普及させるためには、サンプリングした葉内のRNAの分解を抑え、常温でのサンプリング処理を可能にする技術も必要である。
2. 遺伝子発現を定量するための手法の低コスト化。例えば、マイクロアレイ解析においては、多サンプルをこなせる超高密度なアレイの開発による低コスト化が実現可能であると考えている。また、バイオマーカー化された数十から数百の遺伝子を対象に遺伝子発現量が定量できるような、ターゲットRNA-Seq法のような手法の開発が望まれる。
3. 気象データの充実化。日射量が測定されている観測所が少ないため、日照時間等から日射量を推定する技術が必要である。
4. 統計モデリング手法の普及。将来的に統計モデリング手法による栽培管理をより一般的なものとして普及させるためには、PCスケールで計算可能なレベルにまで計算速度を高速化する必要があるだけでなく、幅広いユーザに対応できるようなドキュメントやツールのライブラリ化などの整備が必要である。

「遺伝子発現を指標にイネの生育を予測するシステムの開発」最終年度報告書

中課題番号: 13406471

研究期間: 平成 25～29 年度

中課題名: 遺伝子発現を指標にイネの生育を予測するシステムの開発 (PFT)

小課題番号: PFT1001

研究期間: 平成 25～29 年度

小課題名: 全国で活用できる遺伝子発現予測システムの開発と農業形質予測への展開

小課題代表研究機関・研究室・研究者名: 東京大学・大学院農学生命科学研究科/農学部・生産・環境生物学専攻 育種学研究室・井澤毅

1) 研究目的

明治以降に始まった作物の人工交配育種は、長く、育種家の経験に基づく達観による選抜によるところが大きかったが、ゲノムが解読され、品種間差の農業形質の QTL 解析、遺伝子単離が進み、デザインされた育種を効率的に進める技術が確立し、イネでは、その社会実装が進みつつある。

一方で、栽培現場における栽培管理技術は、篤農家の経験や農試の研究結果を元に得られた「ノウハウ」が「伝承」されているのが実情で、誰でも簡単に身に付けられるわけではなく、その伝承を効率的に進めるためのデータベース作成や、野外の環境変動が農業形質にどう影響を与えるかのデータ処理からの推定モデル（フェノロジーモデル）が開発され、一部、利用が開始されている。また、ICT 技術により、リモートセンシング等で得られたデータから、環境変動の一步先を行く栽培管理のための技術開発が開始され、今後の社会実装が期待されている。

本課題では、野外での作物の遺伝子発現データが、栽培管理や栽培シミュレーションに利用できることを示し、実用的な技術の開発を目的として、大きく、二つの目的を持って、技術開発を進めている。

一つ目は、野外の環境変動データから、全遺伝子の挙動を予測する技術の開発で、この技術開発のプロトタイプは、茨城県つくば市の水田でのサンプルのみの解析ではあったが、新農業展開ゲノムプロジェクトの成果として、論文発表されている (Nagano et al. Cell 2012)。この Cell という生物学のトップジャーナルに、植物科学・農業科学の分野で、日本の研究チームの成果が掲載されるのは 10 年来という事実からもわかるように、作物科学における日本発の大きなポテンシャルを持った新規なアプローチして評価されている。

そこで、本課題では、この技術をさらに改良し、予測精度を実用レベルに発展することを目的にする。そのために、全国 7 か所で、主要品種であるコシヒカリのサンプリングを複数年 (4～5 年) 行い、トランスクリプトーム (全遺伝子発現) 解析を進める計画である。品種間差にも対応できるように、地方代表品種も、コシヒカリと同時に栽培し、サンプリングはしておく。(予算的に、PFT1001 では、地方品種の発現解析は計画していない。)

一方で、社会実装を進めるには、野外での遺伝子発現動態と農業形質の関係を明確にしていく必要がある。そこで、二つ目の目的として、研究予算の許す限り、多くの形質に関して、野外の遺伝子発現データからの形質推定技術を開発する。一部の有用形質では、プロジェクト期間中に社会で役立つ情報を提供できる実用技術を確立することも目的にする。例えば、出穂期、窒素動態、病気にかかりやすい状態か否かといった形質を初期のターゲットにする。

将来的には、それらを組み合わせて、作物栽培管理を遺伝子発現データで行う技術を確立することを目標とする。

2) 研究成果

H24 年度に予備的に富山でサンプリングを行った。H25～H29 年度は、計画通りに、主要品種であるコシヒカリの葉をサンプリングを、札幌、古川、富山、つくば、高槻、福山、福岡の七か所で行った。各サンプリングは、2 時間ごとに、丸一日、13 サンプルを基本にしている。また、一個のサンプルは、3 個体から、一枚ずつ、完全展開直後の葉身をあつめ、その場で液体窒素で凍らせて、つくば

の生物研に送付し、解析した。H27～H29年度は予算減の影響もあり、当初の1/3の規模のマイクロアレイ解析になった。今年度までの5年間で、全国で栽培したコシヒカリの葉の遺伝子発現情報は、約1400アレイデータとなっている(図1)。今年度サンプルは現在解析中であるので、カウントしていない。それぞれに、イネの葉で発現している遺伝子、約6万プローブ、冗長性を考慮すると約3万遺伝子の発現情報が得られた。

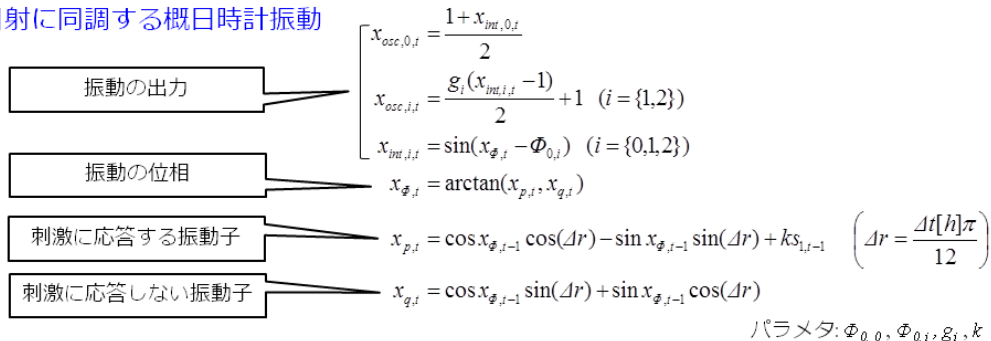
地域	2012	2013	2014	2015	2016	計
札幌	0	39	52	26	52	169
古川	0	52	52	52	0	156
富山	52	78	52	78	0	259
つくば	0	148	103	52	66	369
高槻	0	52	84	52	0	188
福山	0	52	51	0	0	103
福岡	0	48	48	52	50	204
計	52	473	442	312	168	1446

図1 これまでにPFTプロジェクトで解析したコシヒカリの葉のマイクロアレイ解析数

2015～2017年度で予算減で解析できなかったサンプルは、凍結保存。

H28年度からH29年度、約1年間をかけて、本格的なモデル構築を行った(図2)。

日射に同調する概日時計振動



気温への蓄積的応答(成長応答)

シグモイド関数

$$x_{acc,t} = x_{acc,t-1} + \frac{\Delta t}{1 + \exp\{-a_{acc}(s_{2,t-1} - b_{acc})\}}$$

パラメタ: $x_{acc,0}, a_{acc}, b_{acc}, y_{acc}$

ゲーティングされた

環境応答(日射と気温) ($i = \{1,2\}$)

シグモイド関数

減衰

$$x_{res,i,t} = x_{res,i,t-1} + k_i \left[\frac{1}{1 + \exp\{-a_i(x_{osc,i,t-1} s_{1,t-1} - b_i)\}} - x_{res,i,t-1} \right] \Delta t$$

減衰速度

パラメタ: $x_{res,i,0}, a_i, b_i, k_i$

非負線形モデルによる統合

$$x_{obs,t} = \beta_0 + \beta_{osc,0} x_{osc,0,t} + \beta_{res,1} x_{res,1,t} + \beta_{res,2} x_{res,2,t} + \beta_{acc} x_{acc,t} \quad (\beta \geq 0)$$

図2 モデル数式

今回は、10分間隔での遺伝子発現をもとめた。訓練データとしては、2012～2014年にかけて全国5箇所からのコシヒカリの葉のサンプルで日射と気温の変動データと発現データを対応付けられる587サンプルを用い、予測精度を評価するテストデータとして、2015～2016年度の426サンプルを用い、2万7千強の遺伝子について、モデルを作成した。H28年度までに開発したAPMC法を用いたゲノムワイドな二環境を同時処理可能な環境応答モデルの構築を行った。ここで言う二環境は日射と気温変動データである。今回のモデルの斬化式は、移植日、栽培地域、年度ごとに計算する必要があり、全部で、23個の栽培グループに関して、斬化式を解いて、パラメータや初期値の評価を行った。今回の形式で、モデルを構築すると、モデル式に相当する部分は、収束が期待できるパラメータがそれぞれ1000個の数値で表現される。機会学習後の1000個のサンプルで各パラメータを表現するようにプログラムされている。これが25個集まって、1個のモデル構築になる。各遺伝子に関して得られた25個のパラメータ分布（1000個の数値で構成）は、この分布がシャープに、特定の値のみに収束している場合、そのモデルの中で、重要なパラメータとなっていることが推定できる。3年間で23個の各栽培グループで、長期間の斬化式を解く必要があり、10分間隔で、約4カ月から5カ月を想定するので、各栽培グループで2万強のステップがある。これを23個計算して、そのうえで、587サンプルの測定値との差から求めたユークリッド距離が小さくなるようにパラメータセットを適宜、変化させ、収束する条件を探した。また、モデルの評価のために、2015～2016年のサンプル、合計426サンプルをテストデータとした。その結果、非常に高予測精度で、実用レベルの遺伝子発現モデルが構築できた。具体的には、ただの平均値モデルより予測精度（テストデータに対するR2）が高い遺伝子が25573個（全体の92%）あり（図3）、予測精度が、0.6以上の遺伝子が、4492個存在する。

H24-H26の訓練データ 587サンプル
H27-H28のテストデータ 426サンプル、
APMC法で回帰

両方のサンプルで発現が確認できた
27914個の転写産物のモデルを評価

ただの平均値モデルより、
予測精度がいいものが、25573個(92%)

予測決定係数が、

0.9以上が、72個、
0.8以上が、714個、
0.7以上が、2268個
0.6以上が、4492個
0.5以上が、7243個
0.4以上が、10200個
0.3以上が、13425個

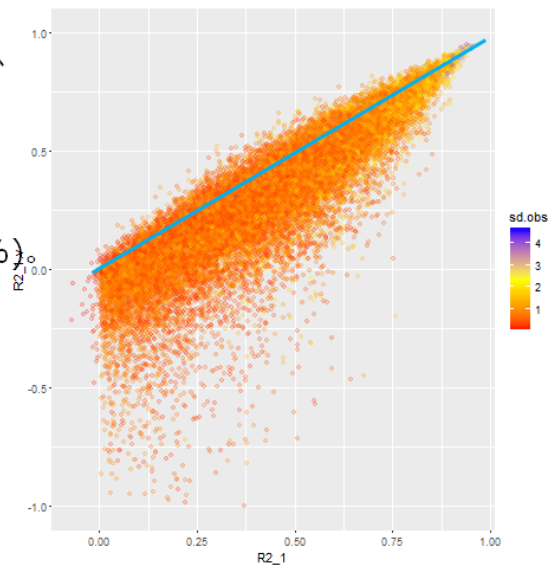


図3 新しいモデルの予測精度

一方で、予測精度が低く、発現の振れが正規分布に従うと考えられる遺伝子が、二千 ($R^2 < 0.1$) ~ 五千個 ($R^2 < 0.2$) あり、また、587 個のデータでの標準偏差が 0.5 以下の遺伝子が約三千個あることが分かった。

この結果により、イネの葉で発現する遺伝子の多くの部分が、移植日とサンプリング時刻と、気温と日射の経時変化データで、正確に推定できることがあきらかとなった (図 4)。

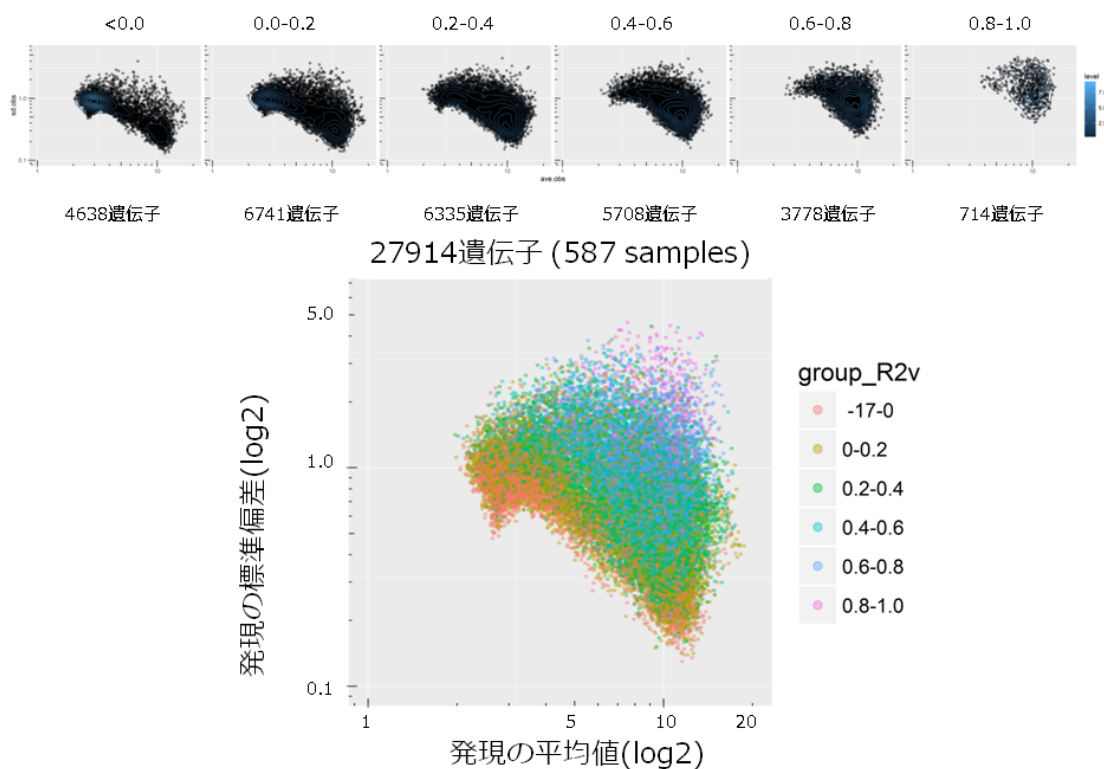


図4 新モデルの予測精度 予測R2の値での散布図。

既知の QTL として単離された有用な農業形質関連遺伝子の発現の推定・予測は、間違いなく、栽培管理上有用な情報になる。今回、予測精度が低く、一方で、発現量も高く、変動も明確な遺伝子をリストアップすると、耐病性や耐ストレス性にかかわる遺伝子が多いことがわかった。このことから、将来的に、環境情報として、これらの情報を入力することで、イネの葉の遺伝子発現は栽培環境の関数として推定可能であると期待できる。また、Hd3a/RFT1 といった花芽形成と環境を繋げる遺伝子は、それほど予測精度の高くないモデルしか構築できておらず、環境因子間の複雑な相互作用を仮定しないと、環境情報からの高精度な (1 ~ 2 日といった誤差での) 花芽形成時期の推定は依然難しいと考えられる。

また、遺伝子発現から、各種形質を推定する統計モデルに関しては、H27 年度までに、イネの体内時計の時計を利用して、植物がもつ時刻情報を遺伝子発現から推定する技術を開発し、論文発表している (Matsuzaki et al. Plant Cell 2015)。

そして、H27 年度に、施肥応答と葉内窒素動態、そして、出穂日の予測、移植日の推定を、サンプルの葉の遺伝子発現情報だけから推定できる画期的な形質推定モデルを構築した。

3) 成果活用における留意点

全国の栽培環境でも実用的な推定・予測ができるゲノムワイドな環境応答モデル (日射・気温変動

のデータと移植日情報で各遺伝子の発現を時系列的に予測するモデル)の構築は達成できた。ただし、予算減での解析サンプル数の減少による実用的な遺伝子モデルの数の目減りが起こった可能性はある。テストデータでの予測精度の評価と、どんな環境においても、一定の発現量を示す遺伝子を論理的に推定することで、コシヒカリの葉で働いている遺伝子のうち、約2万個の遺伝子が、日射変動と気温変動とサンプリング時刻情報と成長ステージ情報(移植日の情報)で、正確に予測できることがわかった。また、予測精度が高くない遺伝子は、ストレス応答性を示す遺伝子であることが示唆される結果を得ている。

また、形質予測モデル構築には、各サンプルに関連性を持たせての形質調査が必要であるが、その部分で、予算減により、病害応答の遺伝子発現バイオマーカーの開発は中断せざるを得なかった。そして、LCTとの共同研究を介して、各種元素動態の遺伝子発現バイオマーカーの開発では、有望な予備的な結果は得られているが、この3月の農水省の科学計算機システムの更新に伴うスクリプト動作不良への対応から、最終年度まで、先に進めることができなかった状況である。

プロジェクト開始当初の目標であるコシヒカリの日射や気温等の栽培環境変化の影響は、出穂期で予測が可能になるが、それを超えて、遺伝背景が異なる新品種等の推定には、大規模な解析は不要ではあるが、解析の継続が必要である。

一方で、施肥応答性の分けつ制御遺伝子の発見等の遺伝子発現バイオマーカーの開発は、当初の予定になかった画期的な成果と言える。

4) 今後の課題

今回の成果を、新しい栽培管理技術に組み込むための必要な技術開発は明確である。ひとつは、野外での葉のサンプリングの自動化で、これには、サンプリング用の装置を取り付けたドローン等の利用開発が必要である。また、サンプリングをした葉内のRNAを常温で壊さない前処理用の試薬の開発が必須である。さらに、発現量の測定の低コスト化も必要である。このためには、マイクロアレイ解析等の一回のハイブリを細区画化したり、RNA-Seqのライブラリを作成するための特異的な遺伝子の発現に注目したcDNAの選抜技術の開発が必要となる。

「遺伝子発現を指標にイネの生育を予測するシステムの開発」最終年度報告書

中課題番号: 13406471

研究期間: 平成 25～29 年度

中課題名: 遺伝子発現を指標にイネの生育を予測するシステムの開発 (PFT)

小課題番号: PFT1002

研究期間: 平成 25～29 年度

小課題名: RNA-Seq 法による品種間トランスクリプトーム比較解析手法の開発

小課題代表研究機関・研究室・研究者名: 農研機構 次世代作物開発研究センター・基盤研究領域・情報解析ユニット・川原善浩

1) 研究目的

日本各地で様々な地域環境に適応したイネ品種が栽培されており、それらの遺伝子発現プロファイル比較により、品種間の形質の違いに関わる遺伝子候補の探索が可能である。しかしながら、従来のマイクロアレイによる発現解析の精度はプローブ設計に強く依存しており、様々な品種に対応するためには、品種ごとに遺伝子配列やアノテーション情報が必要になるなどの問題がある。本課題では、超並列シーケンシング技術を活用した多品種に対応したイネの網羅的な遺伝子発現プロファイル解析手法を開発し、圃場環境における高精度な品種間トランスクリプトーム比較解析の実現を目的とする。さらに、品種間の様々な形質の違いに関わる遺伝子候補を効率よく絞り込むことを目的とし、複数のイネ品種間のゲノム配列上の差異、遺伝子構造、遺伝子発現プロファイル、様々なアノテーション情報を同時に俯瞰、比較可能なシステムを開発する。また、様々な品種に対応したマイクロアレイプローブのデザインに活用できるよう、RNA-Seq 解析によって得られる各品種の葉で発現している全転写産物の配列情報や多型情報を整備する。

2) 研究成果

1) コシヒカリやコシヒカリ系地域品種のゲノム解読

遺伝子発現解析での活用を目的として、コシヒカリとコシヒカリ系地域品種7種(ほしのゆめ、ひとめぼれ、てんたかく、てんこもり、ヒノヒカリ、元気つくし、コシヒカリ近中四 SBL1 号)を対象として、イルミナ社の超並列シーケンサーを用いてゲノムリシーケンシング解析を行った。得られたペアエンドリード配列をイネ(日本晴)のリファレンスゲノム配列(IRGSP-1.0)にアラインメントし、各品種について日本晴ゲノムに対する多型サイトを検出した。そして、得られた多型情報を整理し、地域品種のゲノム構造の概要を明らかにした(図1)。

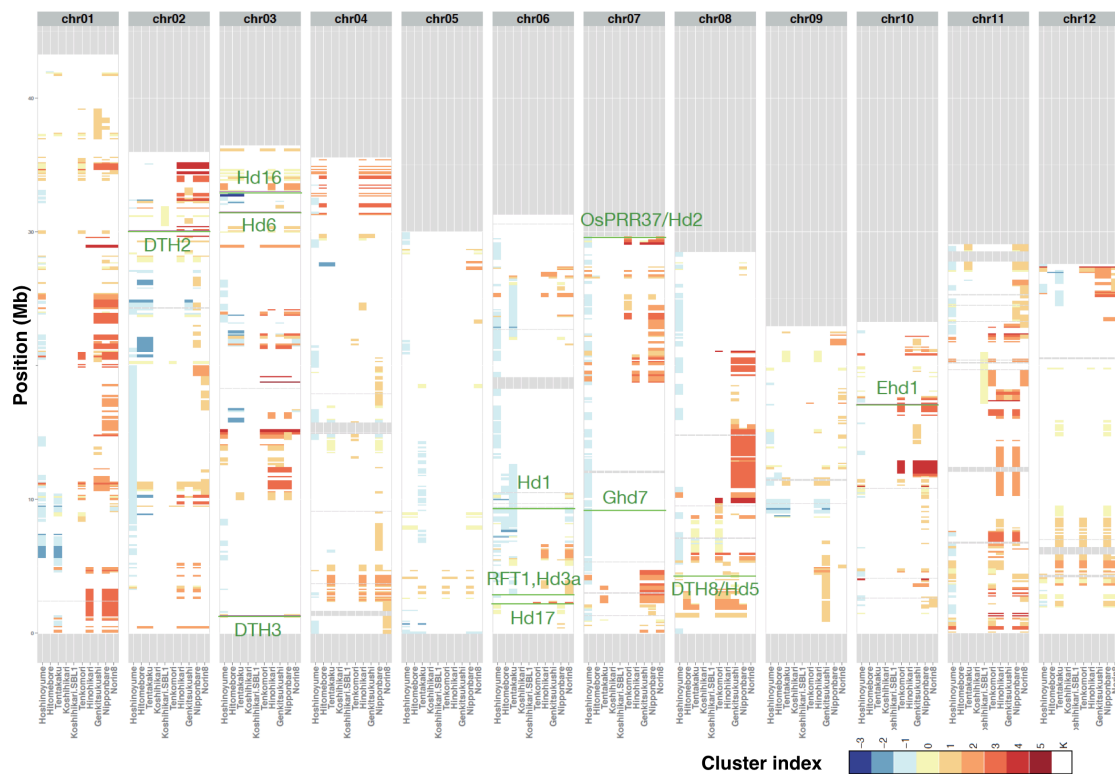


図1 コシヒカリと異なる遺伝子型をもつ領域を早晩性にもとづいて分類した染色体地図

100kbの領域ごとに遺伝子型によるクラスタリングをおこない、コシヒカリとコシヒカリSBL1を基準の0、それよりも早生の品種（ほしのゆめ、ひとめぼれ、てんたかく）に-1、晩生の品種（てんこもり、ヒノヒカリ、元気つくし、農林8号、日本晴）に1のスコアを与え、クラスター内のメンバーの合計値をそのクラスターのインデックスとして分類した。その結果を染色体ごとに左から、ほしのゆめ、ひとめぼれ、てんたかく、てんこもり、ヒノヒカリ、元気つくし、日本晴、農林8号の順に並べて色分けした。早生品種に特徴的なゲノム領域は青く、晩生品種に特徴的な領域は赤く示されている。

本課題で得られたリード配列データはDDBJのDDBJ Read Archive (DRA) から取得可能である（アクセス番号：DRA006061）。また、本課題で対象としたコシヒカリやコシヒカリ系地域品種だけでなく、すでに公開されている300種を超えるイネ品種のゲノムリシーケンスデータを合わせ、同じ解析ワークフローを用いて解析を行った。そして、得られたリファレンスゲノム（日本晴）に対する多型情報をマルチゲノムブラウザ「TASUKE」に載せ、一般公開した（図2および、RAP-DB[<http://rapdb.dna.affrc.go.jp>]のトップページのメニュー Browser-> TASUKEより閲覧可能）。イネアノテーションデータベースRAP-DBと密な連携をとることで、遺伝子アノテーション情報と相互に閲覧が可能となっており、ユーザは任意の遺伝子やその近傍にある多型情報を容易に調べることができる。また、TASUKEに実装された機能によって、マーカー候補となる任意の多型を対象としたPCRプライマーの設計等も可能となっている。このように本課題で得られたコシヒカリやコシヒカリ系地域品種のゲノム情報については新たなマーカー開発や各品種用にマイクロアレイプローブをデザインする際に活用できるように情報を整備、公開した。

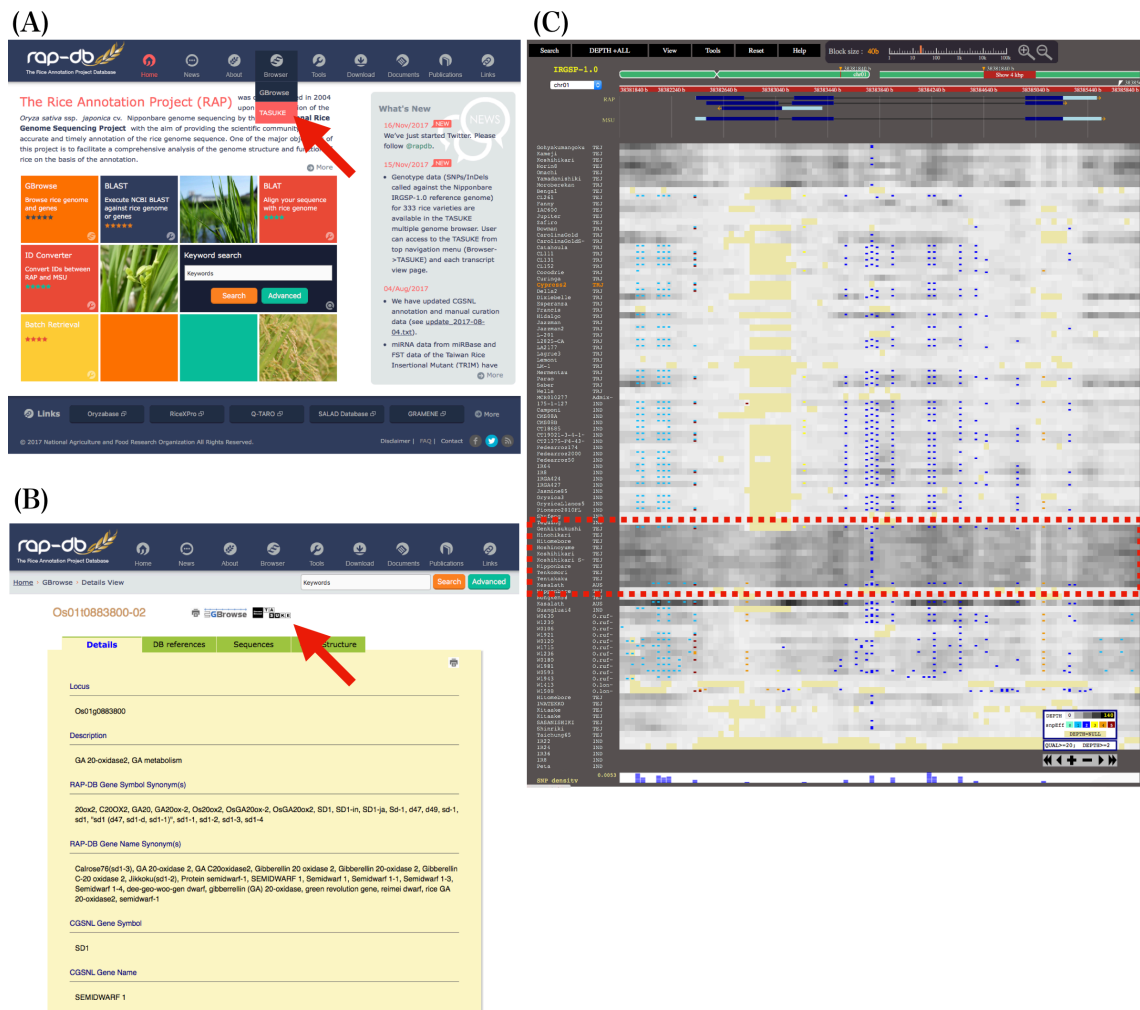


図2 RAP-DBとコシヒカリ系品種のゲノム情報を公開するTASUKEの連携

(A) RAP-DBのトップページに配置されたTASUKEへのリンク (矢印)、(B) 各転写産物の機能アノテーションページ (例: SEMIDWARF 1, SD1) に配置されたTASUKEへのリンク (矢印)、(C) TASUKEによってSD1遺伝子座周辺の多型情報を表示している。縦軸に各品種、横軸が染色体に沿ったゲノム領域を表しており、各ブロックが40bpの領域に相当し、背景色の濃淡でread depth、ブロック中に表示される四角形やその色の濃淡が多型の存在やその頻度を表している。PFT1002で解読したコシヒカリとコシヒカリ系地域品種の多型情報は点線枠部分に表示されている。

2) RNA-Seq 法による多品種に対応した高精度な遺伝子発現解析手法の開発

非リファレンス (日本晴以外の) 品種を対象に RNA-Seq 法による遺伝子発現解析を行う際、解析精度を落とす原因の一つとして、品種間のゲノム配列の差異が挙げられる。そのようなゲノム配列の差異を解消するため、ゲノムリシーケンス解析から得た多型情報を元に、リファレンスゲノム上の多型サイトを対象品種タイプに置換したゲノム配列「品種タイプゲノム」を作成し、それを RNA-Seq 解析に用いる手法を考案した (図3)。また、リファレンス品種から遠縁の品種にも対応できるように、リファレンスゲノムにアラインメントできなかった各品種のゲノムリシーケンスデータ (アンマッピングリード) を元に「品種特異的なゲノム配列」を de novo アセンブルによって作成し、品種タイプのゲノム配列と合わせて発現解析に用いることで、発現解析のさらなる高度化を図った。一連の解析パイプラインは農研機構 高度解析センターの情報

解析サーバ内に構築し、クラスタ計算機を用いることで複数の品種やサンプルを対象とした並列計算が可能である。「品種タイプゲノム」と「品種特異的配列」の作成については、収束するまで変異の検出と置換を繰り返すため、計算時間はリファレンス品種と対象品種の近縁度に依り、概ね数日から数週間程度で完了する。RNA-Seq 解析については1 サンプル当たり数時間で完了し、数十～百サンプルを同時に解析可能である。

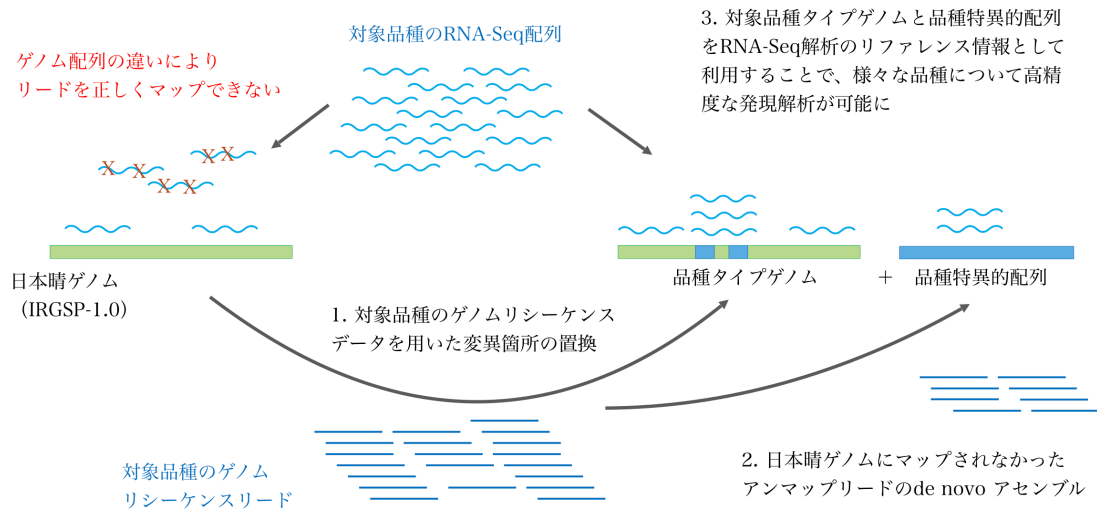


図3 ゲノムリシーケンスデータを用いて作成した「対象品種タイプゲノム配列」と「リファレンス品種には無い対象品種特異的配列」を活用した、様々なイネ品種に対応した高精度なRNA-Seq解析のワークフロー

3) RNA-Seq 解析とマイクロアレイ解析による遺伝子発現プロファイルの比較と評価

超並列シーケンサーを用いた RNA-Seq 法による遺伝子発現解析の高精度化を目的とし、圃場環境で栽培されたコシヒカリの葉のサンプルの一部（47 タイムポイントの時系列サンプル）を用いて、RNA-Seq 解析とマイクロアレイ解析を行い、両技術によって得られる遺伝子発現量や発現プロファイルを比較した。イネの葉で確実に発現していると思われる（どちらかの技術での平均発現量（log2）が5以上）かつタイムポイント間で大きな変動を示す（全タイムポイントの標準偏差が1以上）3,777 遺伝子に対象を絞り、それぞれの技術によって得られる時系列遺伝子発現プロファイルの差を調べた。その結果、遺伝子発現プロファイルは概ね2つの技術で一致していることが分かったが、一部の遺伝子では大きな違いも見られた（図4）2つの技術によって得られる発現プロファイルに違いが生まれる理由を調べるため、差異の大きな遺伝子の特徴や設計されたプローブ位置などを詳細に調べた。その結果、マイクロアレイ側の問題として、不正確な遺伝子構造アノテーションに起因するプローブ設計ミスが多く見られた。このような情報は、今後新たにマイクロアレイプローブをデザインする際に役立つと考えられる。また、RNA-Seq 側の問題として、ゲノム中に多コピー存在する重複遺伝子や転写配列長が短い遺伝子（200-300bp 以内）について、現在のライブラリ調製法やデータ解析手法では発現の検出が困難であることが分かった。しかしながら、重複遺伝子についてはマイクロアレイにおいてもクロスハイブリの問題があるため、正しい発現量が得られているかは検証の必要がある。

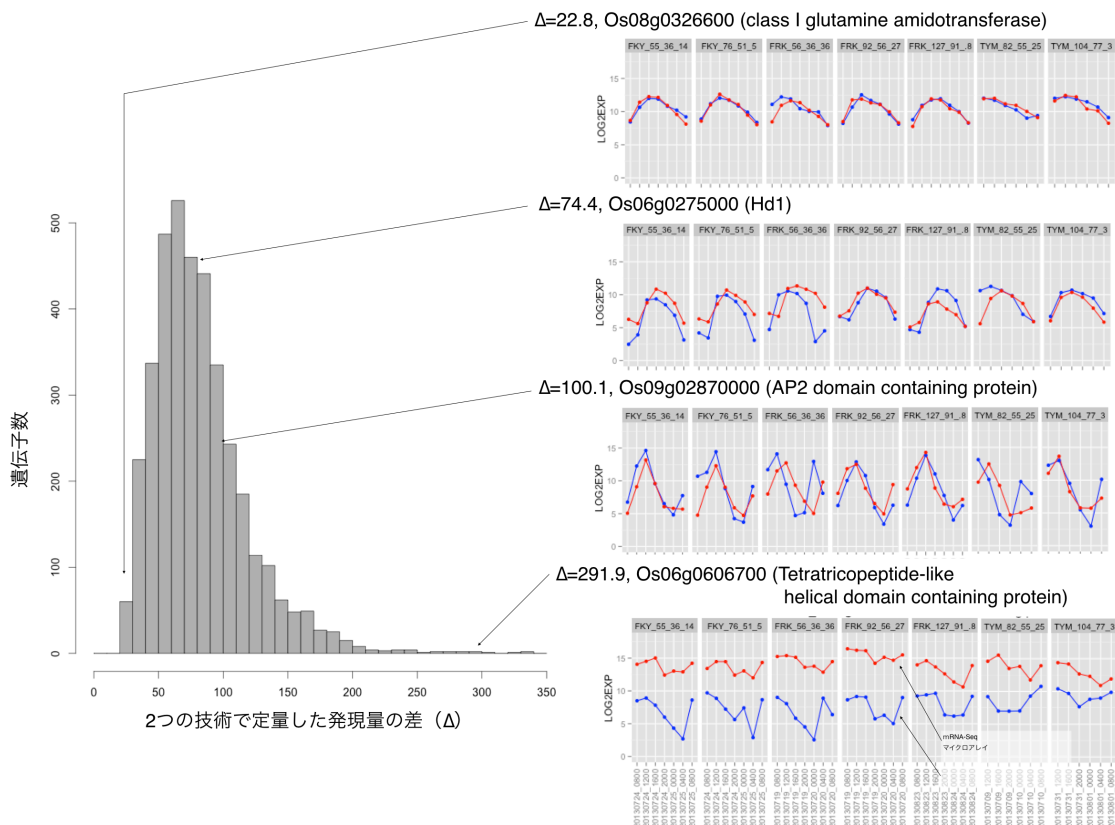


図4. マイクロアレイとRNA-Seq解析による発現プロファイルの差
 どちらかの技術での平均発現量 (log2) >=5、かつ全タイムポイントの標準偏差>=1の条件を満たす、
 3,777遺伝子を対象とし、各遺伝子の47タイムポイントについての2技術間の発現量差の絶対値の和の
 ヒストグラムと代表的な発現プロファイル

4) コシヒカリとコシヒカリ系地域品種の遺伝子発現プロファイル比較解析

本課題で開発した、ゲノムリシーケンシングデータを活用した、RNA-Seq法による多品種に対応した高精度な遺伝子発現解析手法を用いて、国内6地点の同環境で栽培したコシヒカリとコシヒカリ系地域品種7種(1地点につき1~2品種)の葉のサンプル(1地点の1品種につき12サンプル)を対象に品種間トランスクリプトーム比較解析を行った(図5A)。全156サンプルについてTotal RNAを抽出した後、ライブラリ調製(stranded mRNA-Seq)やシーケンシング解析(100bp x 2のペアエンド)については外注にて行った。各品種について得られたRNA-Seqデータは、上述の手法によって作成した「品種タイプゲノム配列」と「品種特異的配列」に対してアラインメントし、遺伝子発現プロファイルを取得した。その際、RAP-DBから提供されている遺伝子アノテーション(IRGSP-1.0)だけでなく、米国ミシガン州立大学のグループが公開するイネゲノムアノテーション(RGAP 7)やPacBio社のロングリードタイプのシーケンサーによって解読した葉由来の完全長cDNA配列も合わせた「統合イネ遺伝子アノテーション(67,550遺伝子座)」を作成し、解析に用いた(図5B)。そして、本課題のコシヒカリとコシヒカリ系地域品種7種のゲノム、トランスクリプトーム解析によって得られたゲノム配列や遺伝子発現プロファイル情報については、ゲノムブラウザ「JBrowse」や、独自に開発した遺伝子発現比較ビューア等を通して閲覧できるように整備し、複数のイネ品種間のゲノム配列上の差異、遺伝子構造、遺伝子発現プロファイル、様々なアノテーション情報を同時に

俯瞰、比較可能なシステムとして開発を行った（図5C）。

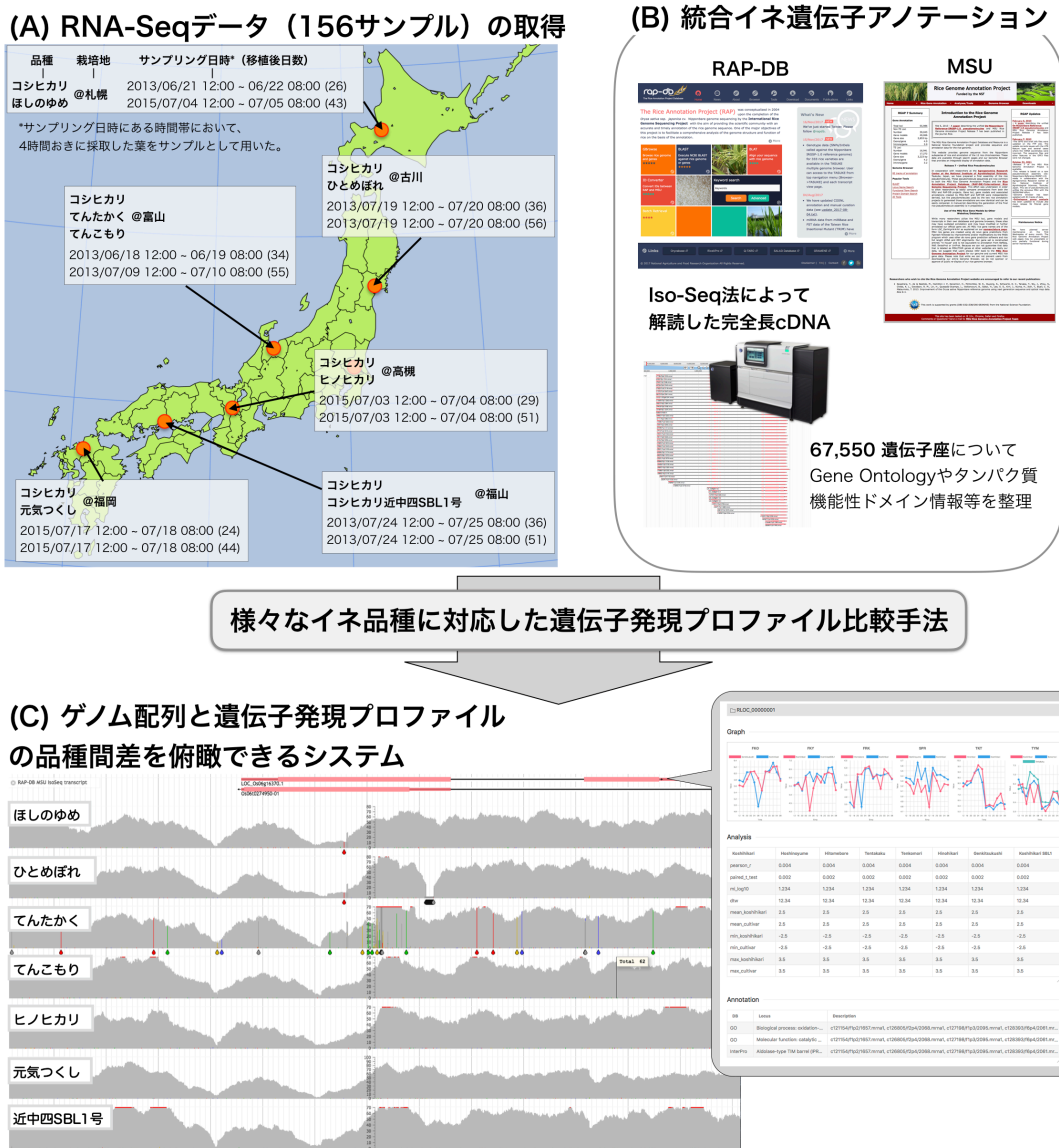


図5 複数品種のゲノム配列と遺伝子発現プロファイルの品種間差を俯瞰できるシステムの構築

(A) コシヒカリとコシヒカリ系地域品種7種のサンプル情報。(B) 発現解析の対象とした統合イネ遺伝子アノテーション情報 (RAP-DB、MSU、Iso-Seq法による完全長cDNA解読に基づく遺伝子構造)。(C) コシヒカリと各コシヒカリ系地域品種のゲノム配列間の差異 (SNPや比較的小さいInDel、ゲノム領域の大きな欠失等) を閲覧可能なゲノムブラウザ (JBrowse) と各サンプリング地点や品種についての遺伝子発現プロファイルや品種間差を表す様々な統計量、遺伝子機能アノテーションなどを閲覧可能な新たに開発した遺伝子発現比較プロファイルビューア (表示しているデータは開発に用いているテストデータ)。

3) 成果活用における留意点

本課題で開発した手法により、ゲノム配列の異なる品種間であっても高精度に遺伝子発現プロファイル比較が可能となり、遺伝子発現プロファイルの品種間差から、品種間の形質の違いの要因となる遺伝子候補の絞り込みが可能となった。しかしながら、同じ環境で栽培したイネを同じタイミングで

サンプリングした場合でも、出穂期が大きく異なる品種間では異なる生育ステージの遺伝子発現プロファイルと比較してしまうことになる。そのため、なるべく出穂期が近い材料を比較対象とする、または解析結果を解釈する際に、生育ステージによって遺伝子発現が変化する遺伝子を区別するといったことが必要になる。

また、同一 RNA サンプルを用いた比較解析によって、マイクロアレイと RNA-Seq の両技術から得られる発現プロファイルは非常に相関が高いことが明らかになったが、遺伝子によっては技術間で発現プロファイルに大きな差がみられるため、両技術によって得られたデータを混ぜて解析する場合には注意が必要である。

4) 今後の課題

- 1) 今後、ロングリードタイプのシーケンシング技術の性能（精度や解読量）がさらに向上し、様々な品種についてリファレンスと呼べるクオリティのゲノム配列が比較的容易に得られるようになる。そのような状況になると、得られたゲノム配列に遺伝子アノテーションを付与し、品種間で遺伝子の対応付けをした上で RNA-Seq 法によって品種間トランスクリプトーム比較解析を行うのが、もっとも効率がよく、精度も高い手法となるはずである。そのためには、リファレンスゲノム配列のアセンブル手法だけでなく、ゲノムワイドに迅速に遺伝子アノテーションを付与するための手法、品種間ゲノムアラインメントやそれに基づく遺伝子やゲノム領域の対応付けをするための手法の確立が重要になるだろう。
- 2) Illumina 社の試薬やプロトコルに基づいた RNA-Seq 法による網羅的な遺伝子発現解析は、ライブラリ調製やシーケンシングを合わせて 1 サンプル当たり 4-6 万円と決して安価とは言えない。1 つの解決策としては、PFT1003 で開発する多検体のライブラリ調製を低コストで行う手法の確立である。また、PFT1001 で開発された遺伝子動態からイネの形質を予測する技術については、最終的にはいくつかのマーカー遺伝子の発現情報が得られれば十分であるため、今後は多数のサンプルについて、数十から数百のマーカー遺伝子の発現量を一度にかつ安価に定量する手法（Targeted RNA-Seq 法）の確立が重要になると考えられる。

「遺伝子発現を指標にイネの生育を予測するシステムの開発」最終年度報告書

中課題番号：13406471

研究期間：平成25～29年度

中課題名：遺伝子発現を指標にイネの生育を予測するシステムの開発(PFT)

小課題番号：PFT1003

研究期間：平成25～26年度

小課題名：遺伝子発現予測システムの高度化・汎用化のための基盤ソフトウェアシステム整備

小課題代表研究機関・研究室・研究者名：京都大学・生態学研究センター・永野惇

1) 研究目的

野外における遺伝子発現のモデリングには、現在、特定のデータセット・モデルに特化して書かれたプログラムが用いられている。その実行や結果を用いたデータマイニング・予測計算に際しても、ad hocに行われているため、モデルの改善や環境データの追加、他の系統、作物種などへの展開に際して、毎回大きな労力がかかる。この問題を解消し、野外遺伝子発現のモデリングをイネの生育予測をはじめとする様々な現実の課題に応用可能とするため、十分な汎用性・再利用性を確保した解析ソフトウェアプラットフォームを開発することを目的とした。また、質の高いモデリングには、質・量ともに高い水準のトランスクリプトームデータが不可欠である。しかしながら、現行のトランスクリプトーム解析の手法は、コストが非常に高く、手間もかかる。このコスト・手間の問題が幅広い作物種などへの、トランスクリプトームのモデリング手法の適用の障害となっている。そこでこの障害を解消するため、低コスト・高効率・高汎用性の多検体RNA-Seq解析手法の開発を目的とした。

2) 研究成果

(1) 気象データの整備

全国の気象データは主として気象庁が観測したものが利用可能だが、観測形態、年代、観測対象要素などによって複数のデータソースに分散しており、利用には整理が必要である。例えば、アメダス観測点と気象官署では観測対象要素、データソース、データ形式が異なっている。また、同じアメダス観測点の時別データでも、年代によってデータソース、データ形式が異なっている。さらに、一定割合で欠測が見られることから、これらの欠測にも補間など何らかの対応が必要となる。そこで基盤情報の整備として、全国約1,300地点の観測点のデータの整理を行った。まず、複数のデータソースからのデータを整理・統合し、MySQLのデータベースを構築した。さらに、欠測の補間処理として、1時間未満の短期間の欠測に関してはスプライン関数による補間を、1日未満の中期間の欠測に関しては前後3日間のデータから時間ごとに算出した平均値による補間を、1日以上長期間の欠測に関しては平年値による補間を、それぞれ行った。これらの処理の結果、最も古い地点では1961年からの気象データを遺伝子発現の予測計算に利用可能な形態に整理することが出来た。

(2) RNA-Seqライブラリ調整の多検体・低コスト化

多検体・低コストRNA-Seqライブラリ調製法をさらに改良するために、Selective depression法の導入を行った。Selective depression法はrRNAなどtotal RNA中に大量に存在するRNAを選択的に分解する方法である。通常のRNA-Seqライブラリ調製ではオリゴdTビーズを用いたmRNAの濃縮が良く用いられるため、ポリAを持たないRNAは除去される。一方、Selective depression法ではrRNAなどのターゲットを分解するので、残ったシーケンス対象RNAにはポリAを持たないものも含まれ、未知の転写産物の発見につながる利点がある。また、rRNA以外に大量に存在する転写産物があった場合、それらも分解対象として実験することが出来るため、定量効率が高まる利点もある。実際にイネのrRNA配列を対象としてSelective depression法による分解を施し、イネ葉サンプルのRNA-Seqを行ったところrRNAを全リードの1%以下まで抑制できた。しかしながら、photosystem D1などいくつかの葉緑体にコードされる遺伝子が多量に発現していることが分かった。上位10遺伝子で残りのRNAの70~90%程度を占めていた(図1)。そこで分解対象にこれら10遺伝子も含めて再度Selective depression法による分解、RNA-Seqを行ったところ、良好な定量結果が得られた。

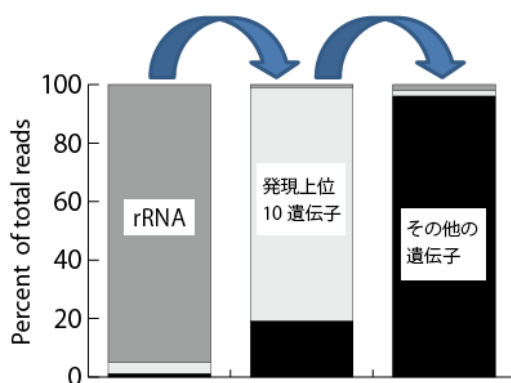


図1 selective depression法によって効率的な定量が可能になった。

(3) モデリング手法の改良、高速化

現行のプログラムは、メンテナンス性、拡張性、汎用性に著しく乏しい。そこでまず、モデリング手順の全体を整理し、今後の開発をスムーズに進めるためにモジュール化する単位を分析した。分析の結果、大きく分けて、最適化アルゴリズム部分、モデルの実装表現部分、データの入出力部分に切り分けることとした。これをもとに、オブジェクト指向による実装を行い、フィールドオミクスモデリング解析ソフトウェアプラットフォーム (FOMP) のフレームワークとした。

先行研究においては100ノードの大型計算機で3カ月程度の計算時間を要していた。そのため、フィールドオミクスの研究手法を他の幅広い作物種に適用することは困難であった。そこで、解析の高速化のために、モデリング手法の改良に取り組んだ。まず、ゲート効果(特定の時刻帯にのみ環境応答を示す現象)を表す関数の実装を見直した。こ

れまででは、「ゲート効果なし」「サイン型ゲート」「矩形ゲート」の3種類に関して、それぞれ別々のプログラムとして実装・推定・モデル選択を行っていた。これをハイボリックタンジェント関数によって表現することで、1つの関数で統一して扱うことを可能とした(図2)。この改良によって、プログラムがシンプルになること以外に、「ゲート効果なし」「サイン型ゲート」「矩形ゲート」のそれぞれの中間的な形も含めて推定が可能となることや、これまで行っていた複雑なモデル選択の手順の簡略化、などいくつかの良い性質が得られた。

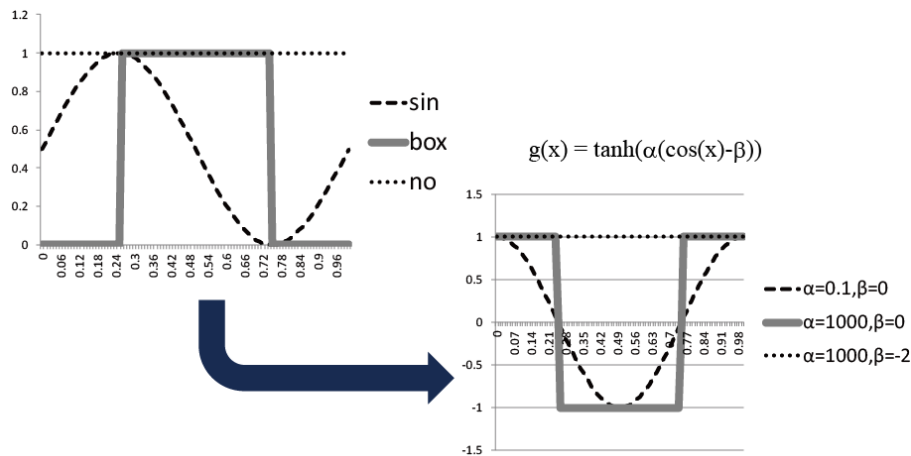


図2 ハイボリックタンジェント関数によってゲート効果を実装した。

さらに、ステップワイズに複雑な手続きで行っていた線形モデル部分の変数選択の改良に取り組んだ。スパース推定の手法であるLASSO法を適用することで、これまで複雑な処理をしていた部分を簡潔かつ高速に行うことが可能となった。

以上のような改良の効果を迅速に調べるため、FOMPを活用してテストのための環境を整備した。加えて、テストを迅速かつ十分に行うため約27,000遺伝子の中から、全体の発現パターンのばらつきを反映する100遺伝子を選定した。この100遺伝子を用いたテストの結果、ゲート関数の見直し、LASSO法の適用などの改良によって、およそ40倍の高速化が達成出来ていることが分かった。また、LASSO法のペナルティの値をクロスバリデーションによって遺伝子ごとに自動的に最適化することによって、過学習を防ぎ予測精度を維持した結果を得ることが可能となった。

さらなる高速化・使い勝手の向上を達成するため、パラメータの最適化アルゴリズムの改良を行った。これまでの方法では全ての遺伝子を同等にひとつひとつパラメータの最適化を行ってきた。ここで、イネゲノム中に3万程度の遺伝子があるが、それらが互いに互いの発現を制御している関係があるため、完全にばらばらの3万通りの発現パターンがあるわけではなく、互いによく似た発現パターンを示す遺伝子が多くあることが期待される。実際、簡単な相関解析などによっても互いによく似た発現を示す遺伝子のクラスタを見出すことが出来る。そこで、あらかじめ全遺伝子をよく似た発現パターンを示す遺伝子クラスタに分類しておき、それぞれの遺伝子クラスタ内でモデリング結果を共有しながらパラメータの最適化を行うと精度を保ちながら計算時間を大幅に節約できると予想される。そこで、このアイデアを実装した最適化プログラムを作成した。ま

ず、全遺伝子を100～500程度のクラスタに振り分けただけで、それぞれのクラスタの代表になる遺伝子を従来法でモデリングする。さらに、クラスタ内の他の遺伝子の最適化に際しては、クラスタの代表遺伝子で選ばれたモデル形式・パラメータを再利用し、そこから最適化をスタートする。この処理によってある程度精度を保ちつつ、大幅に計算時間を短縮することが可能となった。

以上のような改良から計算速度に関しては、現行のモデリングに関しては汎用のPCスケールで計算可能なレベルまで高速化することができた。しかしながら、未だトランスクリプトーム全体を計算するには数日間を要する状態であり、また、さらに多くのデータ、複雑なモデルに対応するためには、大型計算機を用いても長い計算時間を要すると考えられた。そこで、システム全体をより汎用的な環境での実行を可能とするために、クラウド型計算資源での実行に向けてテストを行った。これまでに開発したプログラムを用いてAmazon EC2環境での実行をテストしたところ、問題なく実行可能であることを確認した。

3) 成果活用における留意点

「(3) モデリング手法の改良、高速化」の成果における「よく似た発現パターンを示す遺伝子クラスタに分類しておき、それぞれの遺伝子クラスタ内でモデリング結果を共有しながらパラメータの最適化を行う方法」を用いた場合、従来の「全ての遺伝子をひとつひとつパラメータの最適化を行う方法」と比較して、やや予測精度が低下することがある。なるべく高い精度を求める場合で、計算資源が許す場合は後者の方法を用いることが推奨される。

4) 今後の課題

(1) 気象データの整備

気象データに関して、基本的な部分の整備は完了したが、実際の遺伝子発現のモデリングに用いるためにはさらなる整備が必要である。例えば、約1,300地点のうち、日射量が測定されているのは67地点のみであり、他の多くの地点では日照時間のデータが取られている。そこで、日照時間からの日射量の推定などが必要となる。

(2) RNA-Seqライブラリ調整の多検体・低コスト化

多検体のトランスクリプトームを効率よく取得するシステムを確立できているが、実験のバッチによっては成績が良くないことがあるなど不安定性が残っている。今後はより安定した結果を得られるようにプロトコルの簡略化や機械化などが必要と考えられる。

(3) モデリング手法の改良、高速化

計算速度に関しては、平成26年度までの成果によって、現行のモデリングに関しては汎用のPCスケールで計算可能なレベルまで高速化することができた。さらにAmazon EC2での実行によって潜在的に幅広いユーザーに対応できる状況に近づいたと考えられる。しかし

ながら、実際の利用に際しては多くの予備知識が要求され、かなり高いハードルが残っていると云わざるを得ない。そこで、このハードルの解消のために、セットアップ・使用方法に関するドキュメントの整備、インストールを簡便に行えるようにライブラリとしての整備、セットアップ済み環境のAmazon EC2インスタンスの整備などが必要であろう。

成果等の集計数

課題番号	学術論文		学会等発表(口頭またはポスター)		出版図書	国内特許権等		国際特許権等		報道件数	普及しうる成果	発表会の主催(シンポジウム・セミナー)	アウトリーチ活動
	和文	欧文	国内	国際		出願	取得	出願	取得				
13406471	0	6	28	5	0	1	0	0	0	1	0	0	0

(1)学術論文

区分: ①原著論文、②その他論文

整理番号	区分	機関名	タイトル	著者	掲載誌	巻(号)	掲載ページ	発行年	発行月
1	②	農業生物資源研究所	Critical gates in day-length recognition to control the photoperiodic flowering	Osugi A, Izawa T	Adv Bot Res	72(4)	103-130	2014	1
2	①	農業生物資源研究所	Punctual transcriptional regulation by the circadian clock under fluctuating field conditions	Jun Matsuzaki, Yoshihiro Kawahara, Takeshi Izawa	Plant Cell	27(3)	633-648	2015	3
3	②	農業生物資源研究所	Deciphering and prediction of plant dynamics under field conditions	Izawa T	Curr Opin Plant Biol.	24	87-92	2015	4
4	①	東京大学、農業生物資源研究所	Hd1,a CONSTANS ortholog in rice, functions as an Ehd1 repressor through interaction with monocot-specific CCT-domain protein Ghd7	Y. Nemoto, Y. Nonoue, M. Yano, T. Izawa	Plant J	86(3)	221-233	2016	5
5	①	旧生物研 東京大学	Synthetic control of flowering in rice independent of the cultivation environment	Okada R, Nemoto Y, Endo-Higashi N, Izawa T	Nature Plants	3	17039	####	3
6	①	旧生物研 東京大学	Fine-tuning of setting of critical day length by two casein kinases in rice photoperiodic flowering	Nemoto Y, Hori K, Izawa T	Journal of Experimental Botany		in press	####	12

(2)学会等発表(口頭またはポスター)

整理番号	タイトル	発表者名	機関名	学会等名	発行年	発行月
1	Time-keeping in rice under fluctuating field conditions	Takeshi Izawa	農業生物資源研究所	Gordon Research Conference 2013: Chronobiology	2013	7
2	野外環境下におけるトランスクリプトームダイナミクスの解明と予測	永野 惇	京都大学	第3回NGS現場の会	2013	9
3	2つの大規模データ: 気象とトランスクリプトームをあわせて考える	永野 惇	京都大学	岡山大学資源植物科学研究所ワークショップ「大規模データと情報科学による植物科学研究の未来」	2013	10
4	栽培環境変動データからイネの遺伝子発現を予測する!	井澤 毅	農業生物資源研究所	生物研創立30周年記念シンポジウム	2013	10
5	Modeling of field transcriptomes: from prediction toward designing	永野 惇	京都大学	International Symposium on Toward the use of atmospheric CO ₂ - from photosynthesis to biorefinery	2013	11
6	気象-オミクスモデル: イネにおけるフィールドトランスクリプトミクス	永野 惇	京都大学	第20回時間生物学会総会シンポジウム	2013	11
7	気象データと網羅的遺伝子発現データの統合	永野 惇	京都大学	環境・生態データと統計解析(H25統計数理研究所共同研究集会)	2013	11
8	野外環境下におけるトランスクリプトームダイナミクスの解明と予測	永野 惇	京都大学	第6回定量生物学の会年會	2013	11

9	栽培環境下での作物の遺伝子発現解析のすすめ	井澤 毅	農業生物資源研究所	植物科学シンポジウム「持続可能な資源の開発に向けた植物科学」	2013	12
10	Punctual transcriptional regulation of circadian clocks under fluctuating field conditions	Takeshi Izawa	農業生物資源研究所	Cold Spring Harbor Meeting in Asia 2014 :Genome Assisted Biology of Crops and Model Plant Systems	2014	5
11	水田で育つイネのフィールドトランスクリプトームデータから有用な情報を抽出する！	井澤 毅	農業生物資源研究所	BIOtech 2014	2014	5
12	イネの概日時計は環境が変動する野外でどれだけ正確に時を刻むか？」	松崎潤, 井澤毅	農業生物資源研究所	イネ遺伝学・分子生物学ワークショップ2014	2014	7
13	Deciphering and Prediction of Transcriptome Dynamics in a Real World	永野 惇	京都大学	国立環境研究所・生物センターシンポジウム	2014	9
14	フィールド・トランスクリプトミクスによる環境応答の解明と予測	永野 惇	京都大学	日本育種学会シンポジウム	2014	9
15	北海道から九州で栽培されるコシヒカリ系地域品種のゲノム比較解析	川原善浩, 田部典子, 伊藤剛, 井澤毅	農業生物資源研究所	日本育種学会第126会講演会	2014	9
16	Deciphering and reproduction of transcriptome dynamics under fluctuating field environments	永野 惇	京都大学	内藤コンファレンス 生物システムの物質的基盤	2014	10

17	Punctual transcriptional regulation of circadian clocks under fluctuating field conditions	Takeshi Izawa	農業生物資源研究所	第12回日仏植物科学ワークショップ	2014	10
18	ノイズの海からシグナルをすくい上げるには？:気象と遺伝子発現の解析から考える	永野 惇	京都大学	EvoDevo青年の会	2014	10
19	Deciphering and Prediction of Transcriptome Dynamics in a Real World	永野 惇	京都大学	分子生物学会シンポジウム	2014	11
20	水田でイネはいかに刻を刻んでいるのか？	井澤 毅	農業生物資源研究所	第21回日本時間生物学会シンポジウム6	2014	11
21	Iso-Seq法によるイネの葉のトランスクリプトーム解析から見えること	川原善浩	農業生物資源研究所	NGS現場の会 第四回研究会	2015	7
22	コシヒカリ系品種の比較ゲノム解析と病害抵抗性をもつコシヒカリNILを例とした比較トランスクリプトーム解析	川原善浩, 田部典子, 井澤毅	農業生物資源研究所	日本育種学会第128回講演会 講演要旨集 育種学研究	2015	9
23	「植物栄養の多面的解析と応用に向けて」(フィールドで得られたトランスクリプトームデータで、作物の健康状態を知ることはできるか?)	井澤 毅	農業生物資源研究所	日本土壌肥料学会2015年度京都大会	2015	9
24	遺伝子の働きで診る「イネの健康診断」実現への取り組み	井澤 毅	農業生物資源研究所	アグリビジネスフェア	2015	11
25	遺伝子発現情報を利用したイネの細密健康診断の可能性	井澤 毅	農業生物資源研究所	ゲノム情報活用時代における作物育種の進展と将来展望	2015	11

26	「ラボとフィールドをつなぐ植物科学」遺伝子発現データを基にした「作物の健康診断」を目指して	井澤 毅	農業生物資源研究所	植物科学シンポジウム2015	2015	12
27	Photoperiodic control of flowering in rice	Takeshi Izawa	東京大学	The 64th NIBB Conference	2016	4
28	田んぼで育っているイネの遺伝子の動きをよく見てみよう！	井澤 毅	東京大学	第48回名古屋大学IGERグリーン自然科学レクチャー	2016	7
29	一分子リアルタイムDNAシーケンサー(Iso-Seq法)によるイネの葉のトランスクリプトーム解析	川原善浩, 田部典子, 井澤毅	農研機構 次世代作物開発研究センター	日本育種学会第130回講演会 講演要旨集 育種学研究	2016	9
30	Photoperiodic flowering under fluctuating natural conditions - Time-keeping dynamics in rice transcriptome in the fields - 自然環境下での光周性花芽形成 —野外におけるトランスクリプトームの経時動態—	井澤 毅	東京大学	第23回日本時間生物学会学術大会 特別企画シンポジウム「基礎と応用の融合」	2016	11
31	野外で育つ植物のありのままのすがたを遺伝子の動きから見てみよう	井澤 毅	農業生物資源研究所	シンポジウム「植物の繁殖戦略を考える」	2104	11
32	オミックス比較解析からみえるコシヒカリ系イネ品種の違い	川原 善浩	農研機構 次世代作物開発研究センター	NGS現場の会第五回研究会	2017	5
33	ゲノム、トランスクリプトーム比較解析によって見えてきたコシヒカリ系イネ品種間の違い	川原 善浩	農研機構 次世代作物開発研究センター	日本育種学会第133回講演会	2018	3

(3) 出版図書

区分: ①出版著書、②雑誌(注)(1)学術論文に記載したものを除く、重複記載をしない。)、③年報、④広報誌、⑤その他

整理番号	区分	著書名(タイトル)	著者名	機関名	出版社	発行年	発行月
1		「該当無し」					

(4) 国内特許権等

整理番号	特許権等の名称	発明者	権利者 (出願人等)	機関名	特許権 等の種 類	番号	出願年月日	取得年月 日
1	特性推定モデル生成装置および方法、解析対象の特性推定装置および方法	井澤 毅, 深田 麻衣子	生物研	生物研	特許権	2015-176416	2015年9月8日	

(5) 国際特許権等

整理番号	特許権等の名称	発明者	権利者 (出願人等)	機関名	特許権 等の種 類	番号	出願年月日	取得年月 日	出願国
	「該当無し」								

(6) 報道等

区分: ①プレスリリース、②新聞記事、③テレビ放映、④その他

区分	記事等の名称	掲載紙・放送社名等	掲載年	掲載月	掲載日	機関名	備考
②	ゲノム情報で進むイネ改良	毎日新聞	2014	8	7	農業生物資源研究所	

(7) 普及に移しうる成果

区分: ①普及に移されたもの、製品化して普及できるもの、②普及のめどがたったもの、製品化して普及のめどがたったもの、③主要成果として外部評価を受けたもの

区分	成果の名称	機関名	普及(製品化) 年月	主な利用場面	普及状況
	「該当無し」				

(8) 発表会の主催の状況
 (シンポジウム・セミナー等を記載する。)

整理番号	発表会の名称	年月日			開催場所	参加者数	機関名	備考
1	「該当無し」							

(9) アウトリーチ活動の状況

当事業の研究課題におけるアウトリーチ活動の内容は以下のとおり。

区分：①一般市民向けのシンポジウム、講演会及び公開講座、サイエンスカフェ等、②展示会及びフェアへの出展、大学及び研究所等の一般公開への参画、

③その他(子供向け出前授業等)

整理番号	区分	アウトリーチ活動	年月日			開催場所	参加者数	主な参加者	機関名	備考
1		「該当無し」								