

平成23年 4 月 20 日現在

機関番号： 82401  
 研究種目： 若手研究(B)  
 研究期間： 2009～2010  
 課題番号： 21780011  
 研究課題名（和文）  
 比較ゲノムリンクによる穀物ゲノム情報の育種活用基盤の構築  
 研究課題名（英文） Construction of genome information resources for cereal breeding based on comparative "genome-link"

## 研究代表者

持田 恵一 (MOCHIDA KEIICHI)  
 独立行政法人理化学研究所・バイオマス研究基盤チーム・上級研究員  
 研究者番号： 90387960

## 研究成果の概要（和文）：

本研究では、穀物の遺伝子発現ネットワークの情報学的な解析と関連データベース群の体系的な統合を行い、穀物での有用遺伝子探索とその育種利用に向けたゲノム知識基盤の構築を行った。特に、オオムギ遺伝子共発現解析を行い、ムギ類での有用遺伝子探索のための知識基盤を構築した(Mochida et al. 2011)。また、比較ゲノム情報基盤に基づいた、イネ科6植物の転写因子データベースを構築し、イネ科穀物での遺伝子制御ネットワークを理解するための知識基盤を整備した(Mochida et al. provisionally accepted)。

## 研究成果の概要（英文）：

In this study, we analyzed regulatory networks of cereal transcriptomes and performed integration of related databases to establish genomic knowledge infrastructure for gene discovery and those application to crop improvement. In particular, we carried out co-expression analysis of barley transcriptome to gain comprehensive landscape of a barley transcriptome, which was to develop a knowledge infrastructure for gene discovery in Triticeae crops (Mochida et al. 2011). Furthermore, we developed a database on transcription factors found in 6 Poaceae species based on the comparative genome informatics infrastructure, which should be beneficial database to elucidate regulatory networks and those comparison among Poaceae cereals (Mochida et al. accepted after minor revision).

## 交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,400,000	420,000	1,820,000
2010年度	1,300,000	390,000	1,690,000
年度			
年度			
年度			
総計	2,700,000	810,000	3,510,000

研究分野：農学

科研費の分科・細目：農学・育種学

キーワード：イネ科植物、遺伝子共発現、比較ゲノム、ゲノムデータベース

### 1. 研究開始当初の背景

作物のゲノム情報とモデル植物の知見を横断的に観察することが、作物研究の加速に不可欠である。各々の作物で EST や全長 cDNA の大量解析など、網羅的な知識の収集が進んでいる。また、全ゲノムの概要解読がいくつかの作物あるいはその近縁モデル植物種でなされ、近縁モデル植物(ゲノム)⇔作物(発現遺伝子)の「比較ゲノムリンク」を構成することができるようになってきた。

イネ科作物では、イネゲノム配列の完全解読(IRGSP 2005)に加えて、トウモロコシの近縁モデルとしてモロコシが、ムギ類の近縁モデルとして *Brachypodium* のゲノム概要配列情報が公開された (<http://www.brachypodium.org/>)。一方で、コムギ、オオムギ、トウモロコシで cDNA 配列クローンの収集と配列解読が大規模に進められた。

申請者は、6 倍性コムギの EST および完全長 cDNA の網羅的な解析を推進し、コムギ遺伝子の発現マップを作成し、倍数性コムギの遺伝子発現パターンを初めて体系的に明らかにした。(Ogihara & Mochida 2003 et al, Mochida et al.2004)。また、クラスタリングしたコムギの全発現遺伝子をイネゲノム情報と比較し、二つのイネ科作物の遺伝子の特徴を明らかにした(Mochida et al, 2006)。また、理研オミックス基盤研究領域、横浜市立大学と共同でコムギ完全長 cDNA クローンの収集と完全解読、機能注釈づけを進めている。一方で、国際的には、2 倍性植物であるオオムギで cDNA 解析がすすめられ、とくに cDNA 配列をマーカーとした連鎖地図の作成が複数の研究機関で個別に進められた(Rostoks et al. 2005, Stain et al. 2007)。

しかし、一部を除いてゲノム構造が極めてよく保存されているコムギとオオムギ間でそのゲノム情報の比較や統合的な活用はほとんど実践されていない。申請者は、世界の研究機関から公開されている別個のオオムギおよび 2 倍体コムギ cDNA 連鎖地図情報をひとつのデータベースとして統合した **Triticeae Mapped EST Database(TriMEDB)** を構築し、公開した(Mochida et al.2008)。TriMEDB は遺伝学的に染色体上の位置が判明する全てのオオムギ cDNA と、それと対応づけられる全てのコムギ発現遺伝子配列情報を検索、閲覧可能にした初めてのデータベースで、独自に EST マーカー配列をクラスタリングし、2700 以

上の非冗長な cDNA マーカーの注釈情報とイネゲノム情報との比較結果も提供する。**GrainGene** や **Gramene** といった既存のデータベースに比べて、塩基配列相同性に基づくイネ-コムギ-オオムギ間の比較が可能な点が特徴的で、大麦と小麦のゲノム情報の統合的な育種活用に先鞭をつけた。さらに申請者は、理化学研究所で配列解読した 6000 以上のコムギ全長 cDNA 配列を含む、オオムギ、コムギのすべての公開済み全長 **Coding Sequence** の配列情報とその詳細な注釈情報を提供する **Triticeae Full-length CDS Database(TriFLDB)** を構築し、公開した(Mochida et al.2009)。

一方で、イネ科作物で網羅的な遺伝子発現情報が充実してきた。イネ、オオムギで網羅的な発現解析情報が公開されている。申請者はコムギ cDNA 配列情報から高密度オリゴアレイを設計し、特に非生物学的ストレス耐性の解析を進めている(Kawaura & Mochida et al. 2005, Kawaura and Mochida et al. 2008)。これらの網羅的遺伝子発現情報で相関解析を行い共発現遺伝子群を絞り込むことが遺伝子の機能解析や候補遺伝子の探索に有効であることが示された(Obayashi et al.2007)。

### 2. 研究の目的

本研究は、モデル植物で加速的に収集の進むゲノム情報学的知見を、作物の育種研究に応用するため、理化学研究所植物科学研究センター(理研PSC)で開発、公開されてきた知識データベースを、比較ゲノム科学の視点で有機的に統合化、発展的に機能を強化した知識データベースを開発する。特に、発現遺伝子の連鎖地図⇔全長遺伝子配列の構造と機能注釈⇔遺伝子の発現プロファイルの 3 種類のゲノム情報の関連づけを行う。イネ科作物をとりあげ、特にイネ、オオムギ、コムギ、*Brachypodium* の cDNA とゲノム情報をもとに、イネ科作物「比較ゲノムリンク」を作成し、麦類ゲノム育種を加速するための情報基盤を確立する。これにより、麦類のゲノム育種に活用するゲノム情報基盤の機能を強化するために、イネ、オオムギ、コムギの遺伝子発現情報を統合し、発現遺伝子の連鎖地図・全長遺伝子配列の構造と機能注釈・遺伝子の発現プロファイルの 3 情報を統合的に活用したジーンディスカバリーを可能にする。

### 3. 研究の方法

本研究は、2 年間で、(1)コムギ、オオムギ、イネの網羅的遺伝子発現解析情報を統合的に比較、閲覧可能なデータベースの構築、(2)ムギ類データベース TriMEDB、TriFLDB と統合と機能の高度化、および(3)イネ、モロコシ、*Brachypodium* のゲノム情報との統合を、検索システムの構築も含めて

行った。

#### 網羅的遺伝子発現解析情報の収集と共発現データの作成

主要な遺伝子発現データレポジトリである PLEXDB (Wise et al. 2007)、NCBI GEO (Barrett et al. 2006) からオオムギおよびイネの網羅的遺伝子発現データを収集した。また、理研植物センターで申請者等が収集したコムギマイクロアレイデータを整理した。これらの情報を統合的に扱えるデータ構造に変換しデータベースに格納した。それぞれのプラットフォームに応じた正規化を行い、相関解析により共発現遺伝子グループ情報を構築する。相関係数には、Pearson の積率相関係数を用いた。相関係数の計算に用いるハイブリデータの選定は、シロイヌナズナの共発現遺伝子データベース ATTED (Obayashi et al. 2007) での取り扱い方法を参考にして吟味した。

#### それぞれの作物マイクロアレイプローブと遺伝子代表配列の対応付け

マイクロアレイに搭載されたプローブを既存の cDNA クラスターの代表配列および、イネ遺伝子に対応づけた。コムギ、オオムギについては、主要な cDNA 代表配列を提供するデータベースである、NCBI UniGene, TIGR GeneIndex, PlantGDB, HarvEST の当該植物のデータとの対応付けを行った。イネについては RAP-DB と TIGR Pseudomolecule を用いた。それぞれのデータを同一のデータベースに格納し、異なるマイクロアレイと異なる cDNA クラスター配列を統一して扱えるようにした。

#### データセットの関連付けと統合化

イネ、モロコシ、Brachypodium のゲノム情報をゲノムブラウザ Gbrowse (Stein et al. 2002) 上に実装し、ゲノム配列とともに予測遺伝子とその構造、機能アノテーションを統合した。イネ科作物ゲノム情報に、アレイプローブ配列を対応付け、イネ、コムギ、オオムギの遺伝子発現情報を参照できるように実装した。同時に、このゲノム情報は TriMEDB、TriFLDB ともデータ連携をもち、それらの情報をゲノム配列上で一元的に検索できるようにデータの関連づけを行った。

## 4. 研究成果

### (1) オオムギトランスクリプトームの共発現解析とムギ類遺伝子探索への応用

本研究の推進により、イネ科作物ゲノム情報の統合化とムギ類研究を加速するための情報基盤が構築できた。そこで、本情報基盤を活用し、オオムギの遺伝子共発現ネットワークの解析と、ムギ類遺伝子探索への活用を実践した。

オオムギの遺伝子共発現ネットワークを描出するために、45 種類のマイクロアレイ実験シリーズから成る 1347 の GeneChip データセットを用いた。遺伝子-遺伝子間の総当たり組み合わせについて、類似組織の偏りを補正するための係数を乗じた相関係数を計算し、オオムギトランスクリプトームの全体的な遺伝子共発現ネットワークを構築した。次に、機能単位としてのネットワークモジュールを探索するために、遺伝子共発現ネットワーク中からサブネットワークモジュールへの分類と、それぞれのモジュール内遺伝子の詳細なアノテーションを行った、それぞれのネットワークモジュールの特徴づけを行うために、シロイヌナズナおよびブラキボディウムの遺伝子との比較を行った。オオムギ遺伝子と、これら 2 つのモデル植物遺伝子との比較解析に基づいて、Gene Ontology の各種オントロジー名の共在関係を調査した。その結果、乾燥ストレスへの応答と、セルロース生合成に関わることが示唆される機能モジュールを見いだすことに成功した。これらの機能モジュールのそれぞれに、その遺伝子発現制御ネットワークに関わる遺伝子群を見いだした。また、イネ科植物ゲノム情報をもちいた進化階層的な比較解析から、ムギ類に特異的なネットワークモジュールの探索を試み、ムギ類に特異的な遺伝子群で構成されるモジュール群をその候補として見いだした。これらの結果は、コムギ遺伝子との比較ゲノム解析を通じて、ムギ類での遺伝子探索に利用可能であることを示した。本研究でのオオムギ共発現解析結果について、解析結果の統合と種々の検索インタフェースを提供するウェブページを公開した (<http://coexpression.psc.riken.jp/barley/>)。

本研究は、初めて網羅的にオオムギの遺伝子共発現解析を実施したものであり、本研究課題である比較ゲノムリンクに基づくゲノム情報の統合により、穀物ゲノム情報の育種活用基盤の構築とその利用を行った成果として、論文公表を行った (Mochida et al. Plant and Cell Physiol. 2011)。

### (2) イネ科植物の転写因子統合データベースの構築と統合化

転写因子とプロモーター上のシス制御因子との相互作用は、分子スイッチとして、遺伝子発現制御ネットワークの機能発現に不可欠である。そこで、比較ゲノムリンクに基づくゲノム情報をイネ科植物の転写因子に適用を拡張した。ブラキボディウム、トウモロコシ、イネ、ソルガム、オオムギ、コムギの 6 種のイネ科草本植物について、ゲノムあるいは cDNA 配列情報から転写因子の計算機的な探索を行った。また、探索された転写因子について、関連する種々の情報を統合したデータベース GramineaeTFDB を構築した。それぞれの転写因子をコードするとされる遺伝子配列に関して、ゲノム配列構造の特徴、タンパク質機能ドメイン、プロモーター領域の構造、Gene Ontology アノテーション、完全長 cDNA の情報などを統合し、ウェブ上でアクセスできるデータベースとして公開した (<http://gramineaeTFDB.psc.riken.jp/>)。また、外部の

公開データベース上の有用な情報との間で相互参照を可能にすることで、情報の統合性と利便性を向上させた。また、それぞれの転写因子をコードする遺伝子のゲノム配列の上流に存在するシスエレメントについて、検索、閲覧できる機能を GramineaeTFDB に実装した。これにより、利用者は、それぞれの転写因子の機能をプロモーターの構造からも予測し、絞り込むことが可能となった。また、トウモロコシ、イネ、オオムギに関しては、トランスクリプトーム情報へのリンクを提供した。また、イネ-シロイヌナズナ FOX 変異体情報と、トウモロコシの Ds タグライン情報、イネの T-DNA タグライン情報へのハイパーリンクを提供した。このデータベースは、本研究課題の比較ゲノムリンクに基づくゲノム情報の統合により、6種のイネ科植物の転写因子の機能に関わる情報の統合化を初めて体系的に進めたものとして、論文を投稿した(Mochida et al. accepted after minor revision, DNA Res.)。

### (3)比較ゲノムリンクの他生物種への応用

イネ科植物で構築した比較ゲノムリンクを背景としたゲノム情報の統合思想とその解析手法を、イネ科植物同様にゲノム情報の蓄積が進むマメ科植物への適用を試みた。特に、環境ストレス応答など、種々の重要な細胞システムに関わると考えられている Two-component system (TCS)に関わる遺伝子探索に活用した。シロイヌナズナおよびイネのゲノム情報をもとに、比較ゲノム科学的な手法により、ダイズゲノム情報から、TCSに関わる遺伝子群を探索した。また、マメ科植物の比較ゲノムリンクに基づく情報基盤を用いて、その遺伝子発現やプロモーターモチーフ、進化的解析による遺伝子サブファミリーなどを調査した。その結果、ストレス応答に関わると示唆される TCS 関連遺伝子を予測することに成功した。このダイズでの解析は、本課題の推進による比較ゲノムリンクによりゲノム情報を統合する思想に基づく、ゲノム情報からの知識抽出事例として、ダイズで初めて TCS 関連遺伝子の探索と機能予測を行ったものであり論文を公表した(Mochida et al. DNA Res. 2011)。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 3 件)

① Mochida K., Uehara-Yamaguchi Y, Yoshida T, Sakurai T, Shinozaki K. Global landscape of a co-expressed gene network in barley and its application to gene discovery in triticeae crops. *Plant Cell Physiol.* 2011 May;52(5):785-803. 2011 Mar 24.

② Mochida K., Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K, Tran LS. Genome-wide analysis of two-component systems and prediction of stress-responsive two-component system members in soybean. *DNA Res.* 2010;17(5):303-24. Epub 2010 Sep 3.

③ Mochida K., Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K, Tran LS. *In silico* analysis of transcription factor repertoires and prediction of stress-responsive transcription factors from six major Gramineae plants. *DNA Res.* (accepted after minor revision)

[学会発表] (計 2 件)

① 第 52 回日本植物生理学会年会  
イネ科植物転写因子の統合データベース GramineaeTFDB 持田恵一, 吉田拓広, 櫻井哲也, 篠崎和子, 篠崎一雄, チャン ファンラム ソン  
2011 年 3 月 20 日~22 日(東北大学)

② 第 52 回日本植物生理学会年会  
オオムギの遺伝子共発現ネットワーク解析 持田恵一, 上原由紀子, 吉田拓広, 櫻井哲也, 篠崎一雄  
2011 年 3 月 20 日~22 日(東北大学)

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

○取得状況 (計 0 件)

[その他]

ホームページ等

<http://gramineatfdb.psc.riken.jp/>

<http://coexpression.psc.riken.jp/barley/>

## 6. 研究組織

### (1) 研究代表者

持田 恵一 (MOCHIDA KEIICHI)  
独立行政法人理化学研究所・バイオマス研究基盤  
チーム・上級研究員  
90387960

### (2) 研究分担者

### (3) 連携研究者

### (4) 研究協力者

櫻井 哲也 (SAKURAI TETSUYA)  
独立行政法人理化学研究所・ゲノム情報統合化研究ユニット・ユニットリーダー  
90415167  
吉田 拓広 (YOSHIDA TAKUHIRO)  
独立行政法人理化学研究所・ゲノム情報統合化研究ユニット・テクニカルスタッフ  
上原由紀子 (UEHARA YUKIKO)  
独立行政法人理化学研究所・バイオマス研究基盤  
チーム・テクニカルスタッフ