

2003年10月31日  
独立行政法人 理化学研究所

## 植物における世界最高密度の遺伝子発現地図を作製

### - 新規発現遺伝子 5800 種を発見 -

独立行政法人理化学研究所(野依良治理事長)は、機能アノテーション(機能注釈)※1 情報を付与したシロイヌナズナ※2の完全長 cDNA※3 情報や、高密度シロイヌナズナ全ゲノムアレイ※4 を用いて世界最高密度となる遺伝子発現地図※5 を作製しました。理研横浜研究所ゲノム科学総合研究センター(和田昭允センター所長)植物ゲノム機能情報研究グループの篠崎一雄プロジェクトディレクター、関原明上級研究員らの研究グループ、遺伝子構造・機能研究グループの林崎良英プロジェクトディレクターらの研究グループおよび米国ソーク生物学研究所 J. Ecker 教授らの SSP コンソーシアム※6 による共同研究の成果です。

ゲノムの塩基配列だけでは、遺伝子を完全に予測することは不可能であり、遺伝子の予測や実在を証明するためには転写産物の解析による裏づけが必要です。さらに、最終的な生物機能活性を有するタンパク質を得るためには、完全長 cDNA の解析なしでは作り出すことができません。遺伝子発現地図作製において、同定された遺伝子の配列情報、及び発現情報は、将来の有用作物の生産性向上、環境ストレス(乾燥、低温、塩、強光など)に耐性をもつ有用作物の作成、耐病性・耐害虫性有用作物の作成、食品工学、新薬の創製、有用物質の生産、さらに環境保全に役立つ植物の育成などへの応用が期待されます。また、高密度シロイヌナズナ全ゲノムマイクロアレイを使ったことで、5800 種という多量の新たな発現遺伝子を発見しました。

植物ゲノム機能情報研究グループでは、本研究で得られたシロイヌナズナ完全長 cDNA を用いて国内外の研究者との植物遺伝子の機能解析のための共同研究を進めています。また、この cDNA クローンについては、“世界標準”の研究材料として、利用を希望する研究者へ理研バイオリソースセンターから分譲しています。本研究成果は、米国の科学雑誌『Science』の10月31日号に掲載されます。また、シロイヌナズナ完全長 cDNA 情報は本研究グループのホームページ(<http://pfgweb.gsc.riken.go.jp/index.html>)でも公開しています。

## 1. 背景

2000年12月に日本の研究機関を含む国際研究チームによって、高等モデル植物の一つ「シロイヌナズナ」の全ゲノムが解析され、植物の機能解析が飛躍的に進歩しました。

植物ゲノム機能情報研究グループでは、シロイヌナズナ遺伝子(完全長 cDNA)のすべてを取り出し、その塩基配列情報などを体系的に整理した辞書となる「シロイヌナズナ遺伝子エンサイクロペディア(百科事典)」の作成を目指して、機能情報のアノテーション(注釈付け)を行っています。

これまでに種々の植物組織、発生段階の植物体、環境ストレス・植物ホルモンおよび光処理したシロイヌナズナ植物体を出発材料として、19個の完全長 cDNA ラ

イブラリーを作成し、その中から約 14,600 種（全遺伝子約 60%）の独立したシロイヌナズナ完全長 cDNA を単離し、塩基配列情報を解読（5'末端および 3'末端の端読み）し、その配列情報を公開しました。また、約 14,000 種のシロイヌナズナ遺伝子について、遺伝子の発現制御に働くプロモーター配列情報を獲得し、プロモーター・データベースを作成しました。それらの成果を「Science」（2002 年 4 月 5 日号）に発表しました。

さらに、本研究グループでは、シロイヌナズナの種々の変異体を多数作成し、個体レベルでの遺伝子の機能解析を進めています。シロイヌナズナのゲノム上にある情報を網羅的に解析を行い、約 26,000 種といわれる遺伝子の機能を明らかにしようとしています。（図 1）

## 2. 研究成果

遺伝子発現地図の作成における、今回の研究成果は下記の 3 点になります。括弧内の研究機関名は本研究内容の正式担当機関です。

(1) タンパク質の機能解析や遺伝子組み換え作物の作成に利用可能な完全長 cDNA を単離し（植物ゲノム機能情報研究グループと遺伝子構造・機能研究グループ）、11,600 種（全遺伝子の約 40%）に関して、全長の塩基配列を決定し（SSP コンソーシアム）、機能アノテーションを行いました（植物ゲノム機能情報研究グループと SSP コンソーシアム）。（図 2）さらにデータベース解析から全遺伝子の 80% の転写産物を同定しました（植物ゲノム機能情報研究グループと SSP コンソーシアム）。

この全長の塩基配列決定したシロイヌナズナ完全長 cDNA クローンは、理研バイオリソースセンターから、これまでに世界の約 200 の研究グループに合計約 3,400 個分譲しています。このためシロイヌナズナ完全長 cDNA クローンの成果は、植物研究における世界的な標準的リソースになりつつあります。

(2) 高密度シロイヌナズナ全ゲノムマイクロアレイを世界に先駆けて作製したこと（SSP コンソーシアム）と、花や根などの植物組織より調製した RNA 由来の 4 種類のプローブとハイブリダイゼーションさせる（SSP コンソーシアム）ことにより、シロイヌナズナゲノムにおいて約 2 万種以上の発現遺伝子の存在を明らかにしました（植物ゲノム機能情報研究グループと SSP コンソーシアム）。さらにその中から本研究で 5,800 種の新たな発現遺伝子<sup>\*7</sup>があることがわかりました（植物ゲノム機能情報研究グループと SSP コンソーシアム）。

また、同定された発現遺伝子の中には、ゲノム上で互いに逆向きに転写されている遺伝子が約 7,600 種存在することがわかりました。さらに、これまでのゲノム解析だけでは発現遺伝子として認知されていないゲノム領域の中から、実際には遺伝子として転写されている領域を新たに約 2,000 個同定しました。この事は、ゲノム塩基配列を基にしたコンピュータによる遺伝子予測のみでは、正確な遺伝子領域を特定できないことを物語っています。

(3) 約 8,000 種の完全長 cDNA に関しては、部位特異的組み換えシステム<sup>\*8</sup>を用いる事によりタンパク質機能解析を迅速に行うことが可能な発現システムを構築しました（植物ゲノム機能情報研究グループと SSP コンソーシアム）。完全長

cDNA クローンは、タンパク質をコードする領域を完全にカバーしているためタンパク質の機能解析に適したリソースです。

本研究グループはシロイヌナズナ完全長 cDNA クローンの ORF(Open Reading Frame)領域<sup>\*9</sup>を PCR<sup>\*10</sup>で増幅した断片を pUNI ベクター<sup>\*11</sup>にサブクローニングしたクローン(U クローン図 3)を全長塩基配列決定するプロジェクト

(ORFeome プロジェクト)も進めてきました (植物ゲノム機能情報研究グループと SSP コンソーシアム)。作製した U クローンは、部位特異的組み換えシステムを用いて容易に種々のタンパク質発現用クローンに変換可能なため、将来の有用な遺伝子資源になると予想されます。これまでに約 8000 個の U クローン的全長塩基配列を決定しており (SSP コンソーシアム)、全長決定した U クローンは、全長配列データを DDBJ/GenBank/EMBL データベースから公開しています。

### 3. 研究成果の意義

植物科学においてもシロイヌナズナ、イネなど膨大なゲノム情報 (DNA 塩基配列) が急激に解読されつつあるなか、それらの情報をもとに、革新的な有用作物、食品、医薬品、農薬などの開発につなげるための、ゲノム機能研究の推進が重要となっています。

近年、有用作物、食品、医薬品開発に直結するゲノム機能研究の主要な課題として、遺伝子から合成されるタンパク質 (生体反応を担う主役) に関する機能・構造研究が重要視されており、激しい国際競争が展開されつつあります。

本研究により植物ゲノムにはまだ未同定の多数の有用遺伝子が埋もれていることが示唆され、ゲノム塩基配列を基にしたコンピュータによる遺伝子予測のみでは、正確な遺伝子領域を特定することが難しいことが示されました。作製したシロイヌナズナの高密度遺伝子発現地図は、植物科学における基礎研究のみならず、"イネ、コムギ、トウモロコシ、大豆"など主要穀物における有用遺伝子の探索への利用も今後期待されます。植物ばかりでなく動物においても今後本研究のように完全長 cDNA や高密度全ゲノムアレイを用いたトランスクリプトーム解析<sup>\*12</sup>をさらに推進させることにより多くの新規な有用遺伝子の単離、同定が期待されます。今回発表した論文は、トランスクリプトーム解析のマイルストーンになるものです。

本研究で用いた完全長 cDNA (遺伝子の本体) は、このタンパク質を合成するための設計情報をすべて有し、タンパク質そのものを合成することができることから、ゲノム機能研究における非常に重要な「基盤ツール」となります。11,600 種という多量のシロイヌナズナ完全長 cDNA について、全長塩基配列情報、エクソンおよびイントロンの情報、ゲノム上の位置情報などを付加することは、「基盤ツール」としての価値を飛躍的に高めるとともに、今後のゲノム機能研究の効率的推進に大きく貢献するものと期待されています。また、完全長 cDNA であれば、開始コドンから終止コドンまでのタンパク質をコードする領域をカバーしているため、その遺伝子の機能を正確に個体で発現することができ、遺伝子組換え作物の作成(特に有用遺伝子の過剰発現植物を作成する場合)に用いる導入遺伝子として活用することができます。

#### 4. 今後の展開

本研究グループでは、これまでに単離したシロイヌナズナ完全長 cDNA を活用した遺伝子の発現・機能解析等に関して、国内外の研究機関とさらに積極的に共同研究を推進していきます。応用面でも、植物の乾燥や低温など環境ストレス耐性に関わる遺伝子、植物の光合成能の向上や生産性の向上に関わる遺伝子、脂質や糖質など種々の物質生産に関係する遺伝子、植物の生長制御に関係する遺伝子など有用な遺伝子の探索に役立てていきたいと考えています。

(問い合わせ先)

独立行政法人理化学研究所

横浜研究所 ゲノム科学総合研究センター

植物ゲノム機能情報研究グループ

プロジェクトディレクター

篠崎 一雄

Tel : 0298-36-4359 / Fax : 0298-36-9060

上級研究員

関 原明

Tel : 029-850-1161 / Fax : 029-856-2624

(報道担当)

独立行政法人理化学研究所 広報室

Tel : 048-467-9272 / Fax : 048-462-4715

Mail : koho@riken.jp

#### <補足説明>

##### ※1 機能アノテーション(機能注釈)

機能アノテーションとは、解析対象である遺伝子について機能の注釈付けを行うこと。例えば、既知遺伝子塩基配列情報とのホモロジー(類似性)解析、遺伝子産物であるタンパク質のモチーフ(2次構造あるいはそれらの組み合わせの構造)解析、遺伝子の属性分類等の方法で注釈付けを行う。

##### ※2 シロイヌナズナ(図4)

学名は *Arabidopsis thaliana* L. Heynh. でアラビドプシスともいう。全長約 30~40 cm のアブラナ科 1 年生草本植物。食用でも観賞用でもない、いわゆる雑草である。実験室内では約 2 か月で次世代の種子をつけることが可能なことから、シロイヌナズナはモデル高等植物として植物の遺伝子の機能解析に幅広く用いられている。さらに研究材料として非常に優れているという特徴だけでなく、シロイヌナズナを用いた研究環境が国際的に非常にオープンで種々の研究材料や情報が利用できる点にもある。具体的には、(1) 植物体が小さく、生育が容易で、生活環が 2 か月と短いなど遺伝学的研究に利用しやすい、(2) ゲノムサイズが高等植物で最小である、(3) 国際的共同研究体制が確立しており、ゲノム、遺伝子、変異体材料などのリソースセンターが存在する一などがあげられる。

### ※3 完全長 cDNA

cDNA とは、ゲノム DNA の中から不要な配列を除き、タンパク質をコードする配列のみに整理された遺伝情報物質である mRNA (メッセンジャーRNA) を鋳型にして作られた DNA のこと。完全長 cDNA は、断片 cDNA と異なり、タンパク質を合成するための設計情報をすべて有している (図 5) ため、タンパク質を合成することができる。この完全長 cDNA を効率的に合成するためには、非常に高い技術が必要とし、わが国が世界に先んじている。(図 6)

### ※4 高密度シロイヌナズナ全ゲノムアレイ(図 7)

シロイヌナズナの全ゲノム配列に関して、対応する 25-mer のオリゴヌクレオチド (25 塩基からなる短いオリゴ DNA) を約 1,000 万個用意し、それらを基盤上に貼り付けたマイクロアレイ。全ゲノムの約 94% をカバーするものを世界に先駆けて作製した。

### ※5 世界最高密度の遺伝子発現地図(図 8)

本研究で得られた完全長 cDNA および全ゲノムアレイ解析で同定された発現遺伝子などの、これまでに転写されている事が確認された遺伝子をシロイヌナズナゲノム上にマッピングさせて作製した地図のことをいう。シロイヌナズナでは、2000 年 12 月に全ゲノム配列が決定され、ゲノムサイズは約 1 億 2,500 万 DNA 塩基対、全遺伝子数は約 26,000 種存在することが予想された。本研究で、シロイヌナズナの全ゲノム配列に関して、対応する 25-mer のオリゴヌクレオチド (計約 1,000 万個) を用意しそれらを基盤上に貼り付けた高密度シロイヌナズナ全ゲノムアレイ (※4 参照) を世界に先駆けて作製し、植物組織より調製した RNA 由来の 4 種類のプローブとハイブリダイゼーションさせることにより、シロイヌナズナゲノムにおいて約 2 万種の発現遺伝子領域の存在を明らかにした。その結果、高等植物では世界最高となる 6 Kb/個という密度であることがわかり、単離した約 11,600 個のシロイヌナズナ完全長 cDNA のゲノム上の位置も明らかにした。

### ※6 SSP コンソーシアム

Salk Institute (プロジェクトディレクター: J. Ecker 博士)、Stanford Genome Technology Center (プロジェクトディレクター: R. Davis 博士)、Plant Gene Expression Center (プロジェクトディレクター: A. Theologis 博士) の 3 つの研究グループからなる。植物ゲノム機能情報研究グループは、SSP コンソーシアムと約 3 年前よりシロイヌナズナ完全長 cDNA の全長塩基配列決定などに関する共同研究を行ってきた。

### ※7 発現遺伝子

ゲノム配列解析プロジェクトが種々の生物で進められており、さまざまなコンピュータプログラムを用いて遺伝子の存在が予測されている (コンピュータにより存在が予測された遺伝子を予測遺伝子という)。しかしながら、コンピュータによる予測のみでは正確な遺伝子領域を特定することが難しいことが本研究も含めて最近幾つか報告されている。コンピュータによる予測ではなく、実際に、1) cDNA とし

て単離されている、2) DNA アレイを用いた解析により遺伝子発現が確認されている、など転写されていることが実験的に確認されている遺伝子を発現遺伝子という。

#### ※8 部位特異的組み換えシステム

本研究グループが構築した迅速なタンパク質発現・機能解析システムでは、Cre-loxP の部位特異的組み換え反応（Cre リコンビナーゼにより 2 つの loxP 部位において組み換えを起こす反応）を利用している。このシステムでは、最初に目的遺伝子の ORF 断片を含む U クローンを構築しさえすれば、制限酵素やリガーゼを用いなくても ORF 断片を挿入方向と読み枠を維持させたまま種々のタンパク質発現ベクターへ移すことができる。

#### ※9 ORF(Open Reading Frame)領域

アミノ酸を指定するコドン（タンパク質のアミノ酸配列を規定するための情報をもつ 3 つの連続した核酸塩基の配列）がある程度長く連続した DNA 配列のことである。ORF の検索によりタンパク質をコードする DNA 塩基配列を見出す事ができる。

#### ※10 PCR

ポリメラーゼ連鎖反応のこと。現在の遺伝子研究に必須のひとつで、微量の DNA を短時間で 100 万倍ほどに増やすことができる。

#### ※11 pUNI ベクター

本研究グループが構築した迅速なタンパク質発現・機能解析システムにおいて、ORF 断片を挿入するためのベクター。ベクターの中に、Cre リコンビナーゼにより認識される loxP 部位が 1 個存在している。ORF 断片をこのベクターに挿入した U クローンを構築しさえすれば、部位特異的組み換え反応を用いることにより制限酵素やリガーゼを用いなくても ORF 断片を挿入方向と読み枠を維持させたまま種々のタンパク質発現ベクターへ移すことができる。（参） Liu. et al. Current Biology (1998) 8 : 1300-1309.

#### ※12 トランスクリプトーム解析

1 つのゲノムまたは特定の細胞・組織・器官の中で生産される転写産物全体をトランスクリプトームという。近年、DNA マイクロアレイなどを用いた遺伝子発現解析システムが開発され、トランスクリプトームの解析が可能になった。

# ゲノム科学的手法によるシロイヌナズナ遺伝子の機能解析

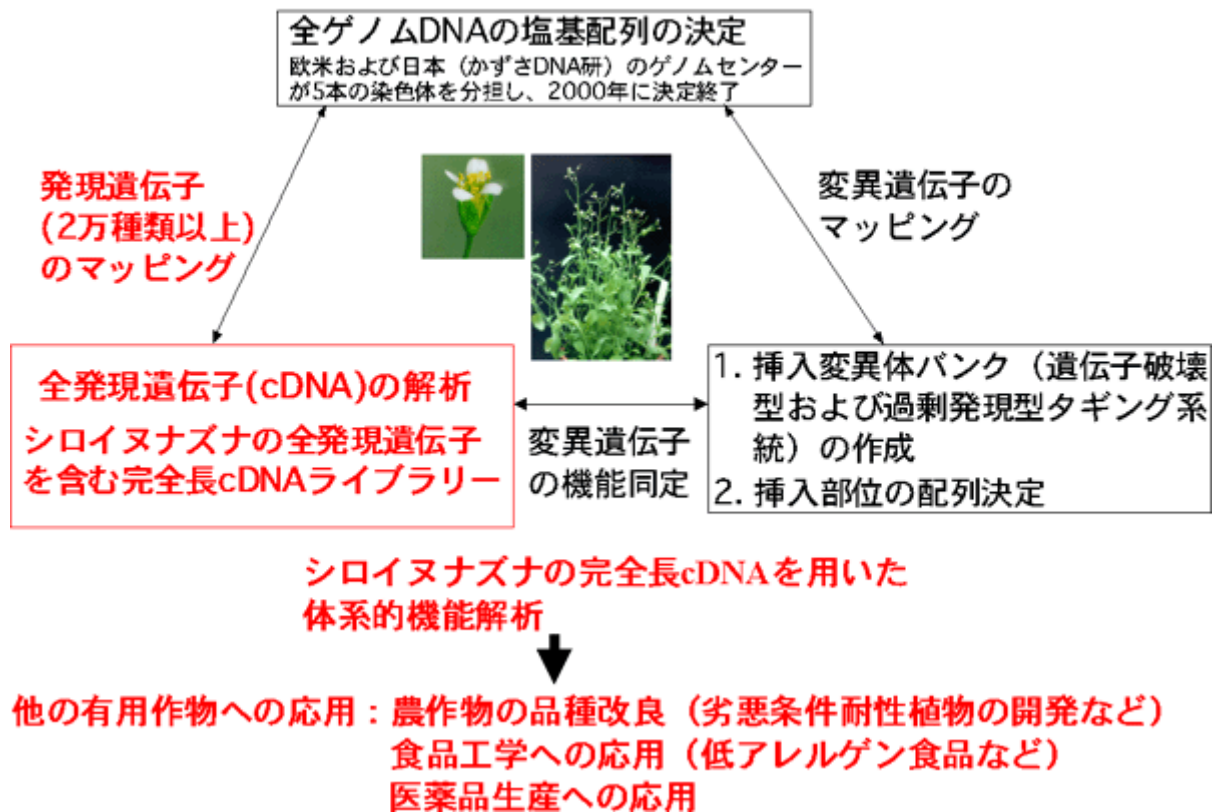


図 1

## 11,600種のシロイヌナズナ完全長cDNAのデータベース：

1. 全長塩基配列情報、ゲノム上の位置情報
2. 機能情報（既知の遺伝子なのか、新規の遺伝子なのか、既知の遺伝子とどのような類似性があるのか、どのような性質の遺伝子なのかなど）の注釈付け

Yamada et al. (2003) Science (in press).

SSP Consortium Full-length cDNA/ORF Clone Search Result

You search for **CLONE ALL** at the SSP cDNA database (last update: Oct 14 18:30).

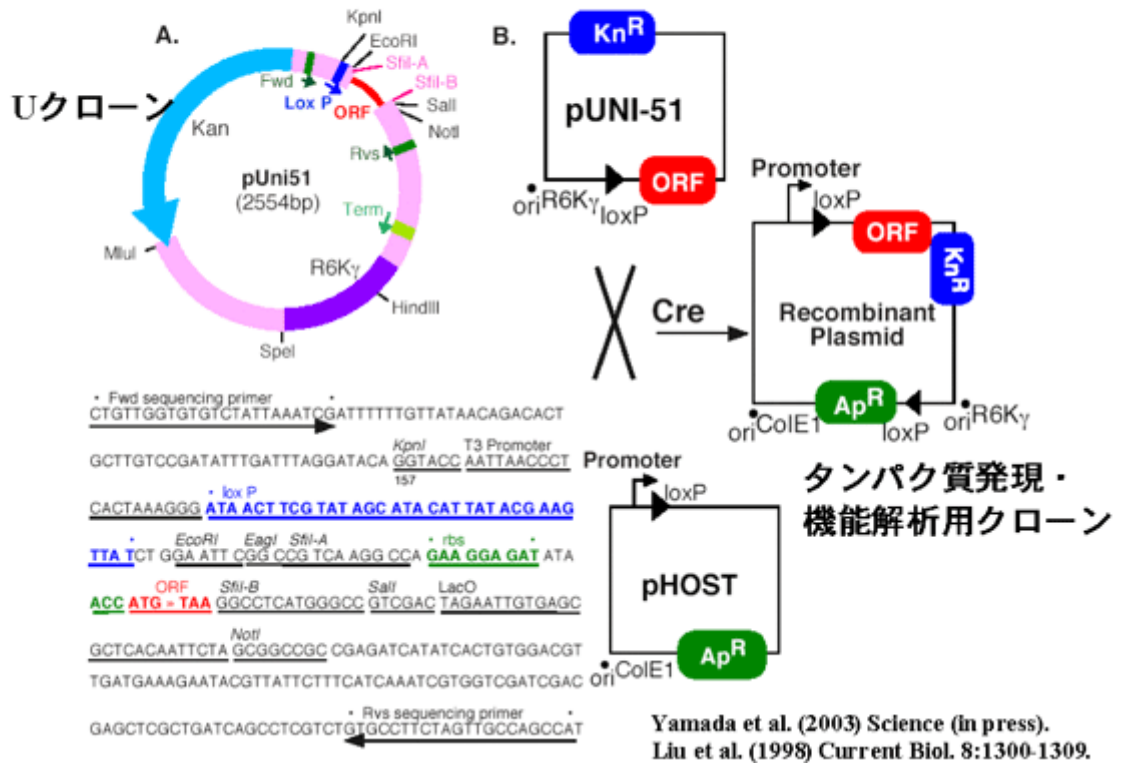
Display All Search: GenBank Clone SSP Submit

#	Clone	Source	Size (bp)	Gene Name			Chr	Putative Function	Group	GenBank		TDNA Link
				New	Old	Coding				Accession	Submitted	
1	S1	015C31A03	1714	AT3g11890	F26K24_18	66.1568	3	hypothetical protein	Salk	AF324982	16-MAR-2001	<a href="#">tdnaexpress</a>
2	S2	000C07C07	1395	AT5g54900	MBG8_17	30.1193	5	polyadenylate-binding protein	Salk	AF324983	16-MAR-2001	<a href="#">tdnaexpress</a>
3	S3	000C07B02	1261	At2g27680	F15K20_22	196.1176	2	unknown protein	Salk	AF324984	16-MAR-2001	<a href="#">tdnaexpress</a>
4	S4	000C10D06	1234	At1g24360	F21I9_300	85.1044	1	putative 3-oxoacyl [acyl-carrier protein] reductase	Salk	AF324985	16-MAR-2001	<a href="#">tdnaexpress</a>
5	S5	000C08A06	431	At2g17660		78.287	2	hypothetical protein	Salk	AF324986	16-MAR-2001	<a href="#">tdnaexpress</a>
6	S6	000C09E05	1207	At1g11210	T28P6_5	82.1008	1	hypothetical protein	Salk	AF324987	16-MAR-2001	<a href="#">tdnaexpress</a>

図 2



完全長cDNAクローンは、タンパク質をコードする領域を完全にカバーしているため、タンパク質の機能解析に適したリソースである。Uクローン（シロイヌナズナ完全長cDNAのORF領域をPCRで増幅した断片をpUNIベクターにサブクローニングしたもの）を用いた迅速なタンパク質発現・機能解析システムを構築した。



## モデル植物としてのシロイヌナズナの特徴



図 4

完全長cDNA（遺伝子の本体）には、タンパク質の製造や働きに関するすべての情報が整理された形で備わっている。

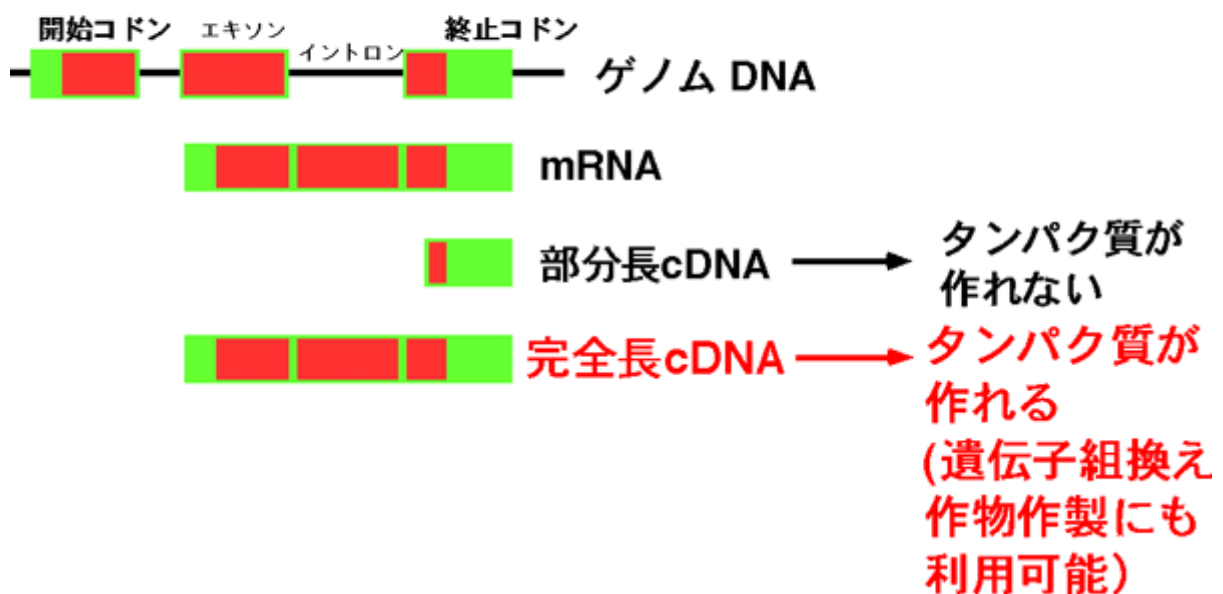
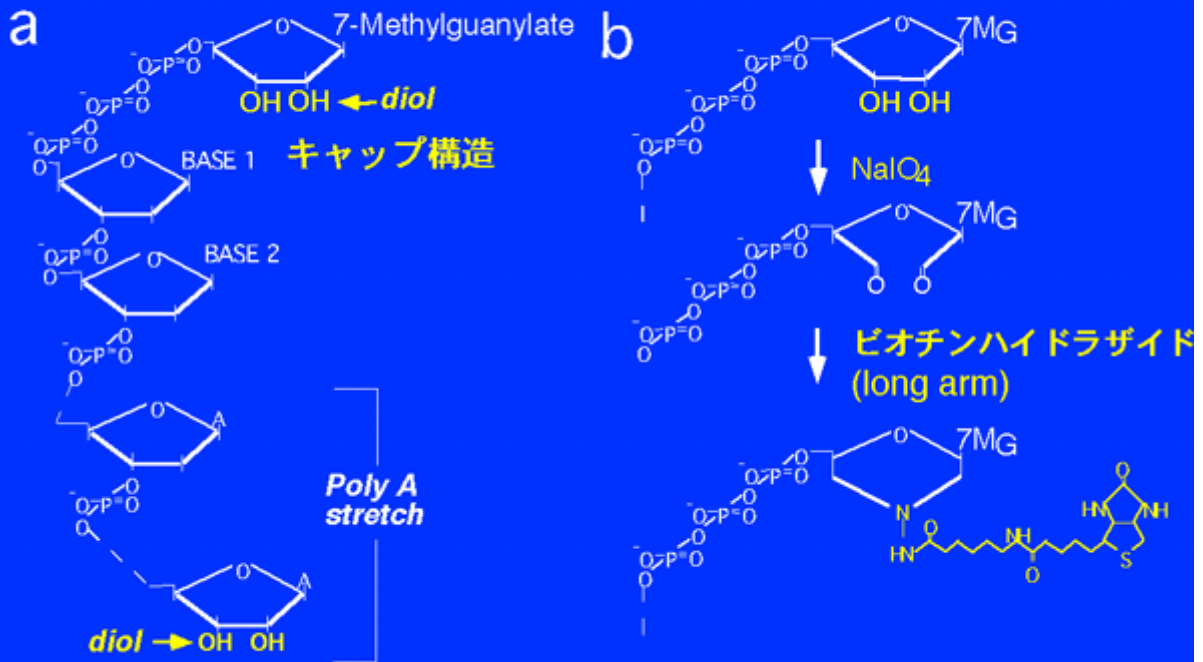


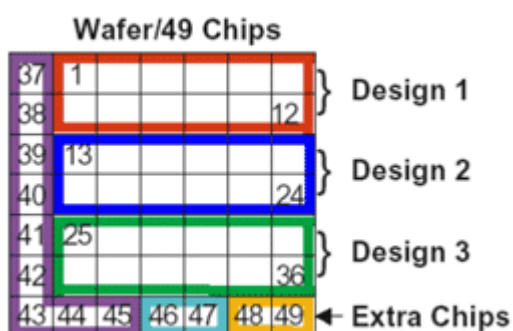
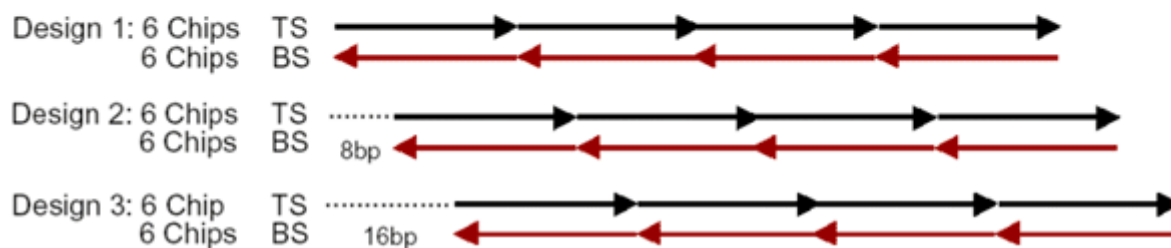
図 5

完全長cDNAを効率的に合成するには、非常に高い技術が必要とし、わが国が世界に先んじている。



ビオチン化キャップトラッパー法（真核生物のmRNAに特有なキャップ構造をビオチン化ラベルする事により完全長cDNAを効率よく選択する方法）（Carninci et al. (1996) Genomics 37, 327；理研林崎グループ）

高密度シロイヌナズナ全ゲノムアレイ（シロイヌナズナ全ゲノム配列に関して、対応する25-merのオリゴヌクレオチドを合計約1,000万個用意し基盤上に高密度に整列させたもの、ゲノム全体の約94%をカバー）を世界で初めて作製した。



37-45: Expression Chips  
46-47: Promoter Mapping Chips  
48-49: Splice Junction Chips

Yamada et al. (2003) Science (in press).

高密度シロイヌナズナ全ゲノムアレイを用いて、  
葉、根などの植物組織より調製したRNA由来の  
プローブ（計4種類）とハイブリさせ解析した結果  
得られた3番染色体の高密度遺伝子発現地図



Yamada et al. (2003) Science (in press).