

# 博士論文

シロイヌナズナゲノムにおける  
隣接遺伝子の共発現データに基づいた  
オペロン様遺伝子群の推定に関する研究

和田 眞昌

2014年3月13日

奈良先端科学技術大学院大学

情報科学研究科

本論文は奈良先端科学技術大学院大学に  
博士授与の要件として提出した博士（工学）論文である

和田 眞昌

審査委員:

金谷 重彦	教授	(主指導教員)
安本 慶一	教授	(副指導教員)
Md.Altaf-UI-Amin	准教授	(副指導教員)
杉浦 忠男	准教授	(副指導教員)
小野 直亮	助教	(副指導教員)

シロイヌナズナゲノムにおける隣接遺伝子の共発現データに基づいた  
オペロン様遺伝子群の推定に関する研究

和田 眞昌

内容梗概

遺伝子のオペロン様の構造の並びは酵母や糸状菌、線虫や植物や哺乳類といった真核生物において広く見られる。植物においては、代謝経路に関係する多くのオペロン様遺伝子群の例が報告されている。例えば、シロイヌナズナにおける thalianol 代謝経路はオペロン様遺伝子群として報告されているが、この遺伝子群は共調して発現するものであり、さらに染色体上では、ごく近い領域に存在する。隣接遺伝子の発現の包括的な解析は、ゲノム内のオペロン様遺伝子群の全セットを明らかにする重要な研究である。ゲノム全体に渡るオペロン様遺伝子の予測は、機能未知遺伝子の機能アノテーションに寄与し、また進化の過程における生物学的な機能獲得についての新たな知見を得ることが期待される。本研究では隣接遺伝子のピアソン相関係数による相関と、無作為抽出した遺伝子ペアの相関の比較による、共発現遺伝子群の統計的な比較解析を行った。これはシロイヌナズナの全遺伝子について、1469 枚のマイクロアレイからなる遺伝子発現データセットを使った、疑陽性率 (FDR) を用いた統計解析による手法で行った。本研究では、シロイヌナズナのゲノムが総計で 100 個のオペロン様遺伝子群を持つことを予測した。また、FDR=0.1 という閾値において、統計的に有意な 3 個から 22 個の遺伝子を含む、34 個の遺伝子群を予測した。これらの遺伝子群の個々の遺伝子群において、機能的な関連性を持つ遺伝子には配列相同性が見られた。またこのうち 5 つの遺伝子群については、P450 遺伝子の中でもアブラナ属に特有の遺伝子を含む代謝経路に関連する遺伝子群を持ち、特に二次代謝経路において働いているものであることが予測された。

キーワード

シロイヌナズナ, オペロン, 発現情報解析, 遺伝子機能予測, False discovery rate

\*奈良先端科学技術大学院大学情報科学研究科 情報科学専攻 博士学位論文,

NAIST-IS-DD0761030 2014 年 03 月 13 日.

Prediction of operon-like gene clusters in the *Arabidopsis thaliana* genome based on co-expression analysis of neighboring genes

Masayoshi Wada

Abstract

Operon-like arrangements of genes occur in eukaryotes ranging from yeasts and filamentous fungi to nematodes, plants, and mammals. In chapter 1, we discuss about several recently characterized examples of operon-like-gene clusters in plants involved in metabolic pathways, e.g. the thalianol pathway in *Arabidopsis thaliana*. Such operon-like gene clusters are defined by their co-regulation or neighboring positions within immediate vicinity of chromosomal regions. A comprehensive analysis of the expression of neighboring genes based on 1469 microarray gene expression datasets of *A. thaliana* is presented in chapter 2 which is a crucial step to reveal the complete set of operon-like gene clusters within a genome. In chapter 3, we predicted co-expressed gene clusters by comparing the Pearson correlation coefficient of neighboring genes and randomly selected gene pairs, based on a statistical method that takes false discovery rate (FDR) into consideration. Chapter 4 contains functional analysis of the predicted operon like gene clusters aiming to contribute functional annotation efforts and provide novel insight into evolutionary aspects. We estimated that *A. thaliana* contains 100 operon-like gene clusters in total. We predicted 34 statistically significant gene clusters consisting of 3 to 22 genes each, based on a stringent FDR threshold of 0.1. Functional relationships among genes in individual clusters were estimated by sequence similarity and functional annotation of genes. Five clusters are associated with metabolism, containing P450 genes restricted to the Brassica family and predicted to be involved in secondary metabolism.

Keyword

Arabidopsis thaliana, operon, whole expression analysis, function prediction, False discovery rate  
Doctoral Dissertation, Department of Information Science, Graduate School of Information Science,  
Nara Institute of Science and Technology, NAIST-IS-DD0761030 March,13,2014.

## 目次

第1章 研究の背景 .....	1
1.1 オペロン:原核生物を中心に .....	5
1.2 真核生物におけるオペロン .....	8
1.3 植物におけるオペロン様遺伝子群 .....	11
1.4 遺伝子共発現におけるバイオインフォマティクス要素技術.....	20
1.5 本論文の構成 .....	22
第2章 オペロン様遺伝子群のゲノムからの探索法.....	23
2.1 シロイヌナズナの発現プロファイルにおける隣接遺伝子間の有意性検定.....	23
2.2 使用データ.....	26
2.3 データの補正 .....	28
2.4 隣接遺伝子の発現方向および、遺伝子間の距離の比較.....	29
第3章 FDR を用いたオペロン様遺伝子クラスターの推定 .....	37
3.1 方法 .....	37
3.2 結果と考察.....	41
第4章 遺伝子アノテーションに基づく、オペロン様遺伝子の評価 .....	48
4.1 クラスター内遺伝子の機能アノテーション方法 .....	48
4.2 結果と考察.....	52
総括 .....	61
謝辞 .....	62
業績 .....	65
参考資料 .....	66
付表 .....	72
付表 表 1.....	72
付表 表 2.....	75
付表 表 3.....	95

## 図目録

図 1	大腸菌のラクトースオペロンにおける遺伝子発現制御機構 .....	7
図 2	オペロンに注目した線虫の遺伝子発現機構 .....	10
図 3	シロイヌナズナ( <i>Arabidopsis thaliana</i> )の thalianol 生合成系と関連遺伝子群 .....	13
図 4	真核生物のオペロンと植物のオペロン様遺伝子群の比較図 .....	14
図 5	イネ( <i>Oryza sativa</i> )のモミラクトン生合成経路と関連遺伝子群 .....	15
図 6	トウモロコシ( <i>Zea mays</i> )の BX 遺群と DIBOA 生合成経路 .....	18
図 7	Avenacin 合成系の <i>Sad</i> 遺伝子群 .....	19
図 8	オペロン様遺伝子群の抽出法の概要 .....	25
図 9	発現の向きと相関解析の例 .....	31
図 10	隣接遺伝子間距離の相関解析の例 .....	31
図 11	シロイヌナズナにおける隣接遺伝子の共発現の有意性検定の結果 ..	34
図 12	隣接遺伝子および 2 遺伝子間の距離の拡大とカイ二乗値の遷移 .....	35
図 13	本研究において FDR を扱う概念図 .....	38
図 14	隣接遺伝子データから実測値、理論値を導き FDR を決定する処理の過程図 .....	40
図 15	各 window サイズにおける上位 $\alpha$ % における FDR の変化 .....	42
図 16	FDR と、FDR の値から求まる遺伝子群の数から近似曲線により推定される FDR0.5 における陽性とみなされるクラスター数 .....	43

## 表目録

表 1	原核生物および線虫におけるオペロンと植物のオペロン様遺伝子群の共通点の整理 .....	22
表 2	本研究に用いた GeneChip .....	27
表 3	ウインドウサイズごとランダムデータにより設定した閾値と、そのときの閾値以上となる推定される遺伝子群の数の関係 .....	47
表 4	本研究の解析により統計的に有意な発現をしているとして得られたシロイヌナズナ遺伝子発現群 .....	50
表 5	シロイヌナズナの遺伝子群と線虫のオペロンの差異 .....	58

# 第1章 研究の背景

ゲノム配列決定技術の進展にともない2013年現在、2725種のバクテリア、36種の植物、72種の動物ゲノムが決定され (<http://www.ncbi.nlm.nih.gov/genome/browse/>)、データベースに公開されている。ゲノムの決定により、それまで個々の遺伝子をターゲットとして研究されていた遺伝子発現の解析は、細胞あるいは組織全体の遺伝子発現量を対象とした、ゲノム全体に及ぶ遺伝子解析、つまり、トランスクリプトーム解析ができるようになった。ゲノムとは、ある生物の持つ全遺伝子、または全DNAという意味で用いられる言葉であるが、このゲノムから得られる情報は多様であり、塩基配列からはその遺伝子がどのようなアミノ酸配列に対応しタンパク質をコードしているかといった情報を得たり、配列の類似性からは他の生物との進化的な距離といった情報を得たり、塩基組成がタンパク質結合の目印となっているという場合の化学的・物理的作様に着目した配列情報を得ることが可能となっている。得られる情報には生物の遺伝子のゲノム上の位置関係も含まれており、隣接してゲノム上に配置されている、同一の染色体に存在しているといった情報も含まれている。遺伝子の位置関係からは、生物が進化するうえで、遺伝子の重複が起り、さらに重複したそれぞれの遺伝子が独自に進化し機能を獲得するという遺伝子進化の手がかりを得ることができる(Vision et al., 2000)。このことを解明するための配列解析を中心としたバイオインフォマティクスが進展

した。一方、共通もしくは一連の機能上関係のある遺伝子がゲノム上に並置して存在し、一つの転写単位として機能することが原核生物を中心に明らかとなり、このような転写単位を比較ゲノム学により解明する試みが進んでいる (Kobayashi et al., 2007)。

一方、遺伝子の発現量を測定するに当たり、多くの遺伝子を一度に測定することを目的として開発された方法として、マイクロアレイ (Schena et al., 1995)、さらに可能な限り、ゲノム上に存在するすべての遺伝子を対象とし(悉皆的に)一度にすべて(網羅的に)測定することを目的とした技術として高密度オリゴヌクレオチドアレイ、さらには新型シーケンサによる RNA-Seq 技術 (Zhong et al., 2009) など次々と新規方法が開発され、種々の生物における発現プロファイルがデータベースに格納され利用されるに至っている。転写(トランスクリプト)産物を網羅的に測定することからトランスクリプトームと名付けられ、世界中で様々な環境で種々の生物においてトランスクリプトームが測定されている。生物はゲノム上の遺伝子の全てが常に発現されているわけではなく、生物の置かれた環境により必要に応じて発現されている。つまり環境に応じた遺伝子の発現の関係が生物の有するすべての遺伝子を対象に解明されつつある。

このように悉皆的かつ網羅的に遺伝子を対象とし、ゲノムとトランスクリプトーム情報を活用することが可能となった。そこで、本研究ではゲノムとトランスクリプトーム情報を統合し、さらに分子生物学における重要な知見をバイオインフォマティクスにより明らかにすることを目標とする。ゲノム情報から遺伝子の位置関係がわかり、トランスクリプトームからは、環境・生理学的条件による遺伝子の発現量の変化の情報を得ることができる。そこで、ゲノム上に配置される遺伝子の隣接関係と共発現の有無、分子生物学における機能の類似性を検討すれば、ゲノム上の特定の領域に位置し、環境・生理学的条件に応



じた遺伝子発現を制御するシステムを有しているか否かを検討することができる。原核生物においては、複数の遺伝子が同一の mRNA に転写されることにより、環境・生理学的条件に迅速に対応するシステム、オペロンが 1960 年にジャコブとモノーにより報告された。ゲノム上に隣接し共発現する遺伝子群が、関連のある機能を有することが、バクテリアゲノムでも観察されており、その転写制御システムをオペロンと呼ぶ(Jacob et al., 1960)。そこで、このような現象がより高等な植物でも成り立っているのかを解明することを目的に、ゲノム配列が決定されかつ遺伝子発現プロファイル情報が蓄積されているモデル生物であるシロイヌナズナに注目して、ゲノム上の隣接遺伝子とその共発現の有無、さらには機能としての共通性を、バイオインフォマティクスにより検討する。

シロイヌナズナ (*Arabidopsis thaliana*) はモデル植物として、現在までの最も詳細に研究が進められた植物であり、2000 年に全ゲノムの塩基配列が決定された(Arabidopsis Genome, 2000)。染色体は 5 本から構成され、全体で 157 Mbp(百万塩基対) となっており(Bennett et al., 2003)、タンパク質をコードする遺伝子数は 27,029 遺伝子とされている(Swarbreck et al., 2008)。ゲノム情報をもとに、トランスクリプトーム、メタボロームなどのポストゲノム研究が進展し、様々な条件における組織ごとの転写プロファイルあるいはメタボロームが測定されている(Thimm et al., 2004; Fukushima et al., 2011)。このような大量情報を体系的に解析することによりさらに高次の生命現象を理解することが可能となっている。このような発現情報は膨大な数値データからさらに意味のあるデータを発掘するために、情報学や統計学を用いたバイオインフォマティクスのアプローチが必要とされる(Schilling et al., 1999; Celis et al., 2000)。

本研究ではシロイヌナズナを対象とし、(1)ゲノム上の遺伝子の位置、および(2)それぞれの遺伝子の発現プロファイルをもとにゲノム上に隣接し共発現する

遺伝子群を統計的に検出する方法を提案し、実際に、シロイヌナズナゲノム中にこのような領域がどのくらいあるかを推定する。

バクテリアゲノムに関しての研究が進んだ遺伝子の発現制御機構であるオペロンについて 1.1 節で述べる。さらに、1.2 節～1.3 節では真核生物におけるオペロン様遺伝子群について知見を整理する。真核生物では複数の遺伝子は同一の mRNA 上への転写はされないが、ゲノム上に並置され、これらの遺伝子が共発現し、分子生物学における関連がある機能を有するような遺伝子クラスターを、オペロン様遺伝子群とよぶ(Field and Osbourn, 2008)。さらに、第 1.4 節において、オペロン解析におけるバイオインフォマティクス技術を整理する。さらに、これらのオーム情報解析を考慮に入れた上で、第 1.5 節で、本論文の目的を達成するための研究戦略を述べる。

## 1.1 オペロン:原核生物を中心に

生物は、遺伝子の発現調節することにより環境に応じた恒常性の維持を達成する。これは生物が環境ならびに生理的条件に応じてタンパク質量を調整することにより、環境への適応を可能にしている。オペロンは大腸菌における発現制御機構として発見された(Jacob et al., 1960)。オペロンの特徴はまず、調節を受ける遺伝子群が、ゲノム上において一つのオペレーターの下流に存在し、一本の mRNA に複数の遺伝子をコード(ポリシストロンの mRNA)する形で転写されることである。この調節の仕組みとして正の制御、負の制御が実験により確認された。正の制御とは、転写を活性化する活性化因子により、DNA の転写量が増えるという仕組みで、オペロンの転写の頻度が増す制御である。負の制御では、DNA に結合することで転写を阻害する阻害因子として働くリプレッサーにより遺伝子発現が抑制されている。リプレッサーは立体構造を変化させる誘導物質と結合する。誘導物質と結合することで構造が非 DNA 結合型あるいは DNA 結合型になる場合がある。どちらの場合も DNA に結合することで転写を抑制するが誘導物質の濃度により抑制が切り替えられる。これらの転写産物が、同一の物質の生合成、生分解経路に関わっている。この転写調節は、例えば、細胞内の物質の濃度により、複数の遺伝子発現のオンオフを一斉に切り替えるといった調節を達成することができる(Malacinski et al., 1999)。

これらの正・負の転写制御を大腸菌で発見されたラクトースオペロンを例に説明する(図 1)。大腸菌ゲノムにおいて、*LacR*、*lacZ*、*lacY*、*lacA* がゲノム上に並置されており(図 1)、*lacZ*、*lacY*、*lacA* の三つの遺伝子は同一の mRNA に転写される。これを *lac*-オペロンと呼ぶ。これらの遺伝子はガラクトースの細胞外からの吸収と代謝に関連する。*lacR* は常に発現しており、転写、翻訳の過程を経て *LacR* リプレッサータンパク質が生合成され、大腸菌内に存在している。こ

の LacR は、*lac*-オペロンの上流のオペレーター配列に結合すると、*lac*-オペロンを形成する mRNA の合成を阻害する(Jacob et al., 1960; Malacinski et al., 1999)。一方、LacR がオペレーター配列に結合できないときには、*lac*-オペロンを形成する mRNA は合成が可能となり、最終的には三つのタンパク質 LacZ、LacY、LacA が転写・翻訳され細胞内にラクトースを取り込み代謝される。このように、同一の mRNA に複数の遺伝子が転写されることにより、遺伝子発現の制御が Lac リプレッサーのゲノム上へ結合するか否かにより、同時にこれら三つの遺伝子の発現に負の制御をかけることができる。このように、1 つの mRNA に転写される遺伝子群をオペロンと呼ぶ。

ゲノム科学の時代に入り、バクテリアゲノムにおいて DNA マイクロアレイなど多量の発現情報をもとにしたゲノム上のオペロンの構成の研究が盛んに行われ(Ermolaeva et al., 2001; Salgado et al., 2000; Wang et al., 2004; )、データベースに整理され公開されている(Huerta et al., 2001; Taboada et al., 2011)。Taboada ら(2011)はゲノム配列決定された 1200 種の原核生物について、遺伝子間距離と機能類似性により予測し公開した。一方、小林ら(2007 年)は、プロモーター配列、隣接遺伝子の発現相関を考慮してオペロンを予測した。その結果、枯草菌において 892 個のオペロンに 2183 個の遺伝子から構成されることが予測された。複数の遺伝子からなるオペロンにおいて、平均 3.71 個の遺伝子が含まれることになる。*Staphylococcus aureus* では平均 3.47 個(Wang et al.,2004)、大腸菌では 3.41 個(Salgado et al., 2000)の遺伝子が複数の遺伝子からなるオペロンに配置される(Kobayashi et al., 2007)。原核生物では、遺伝子発現を制御するために、ポリシトロニック、すなわち同一の mRNA に複数の遺伝子が同時に転写されることによる遺伝子発現制御機構において重要な役割を演じており、ゲノム上の遺伝子の配置が直接、遺伝子制御に関わっている。

大腸菌 *lac*オペロン

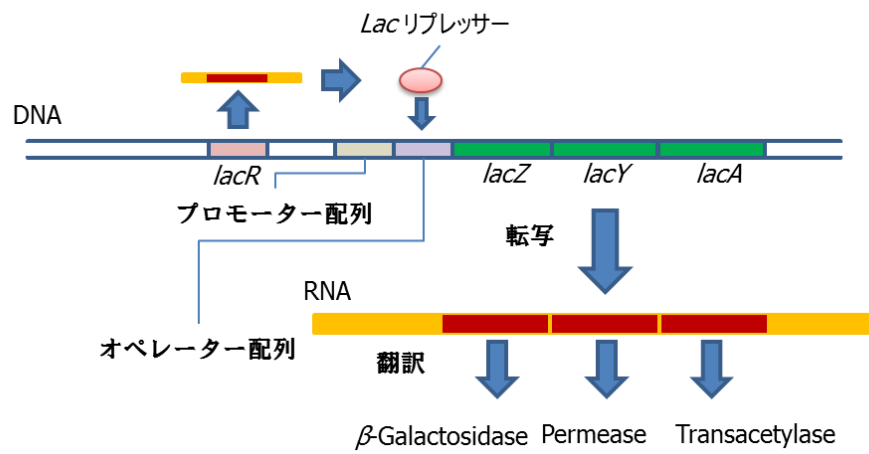


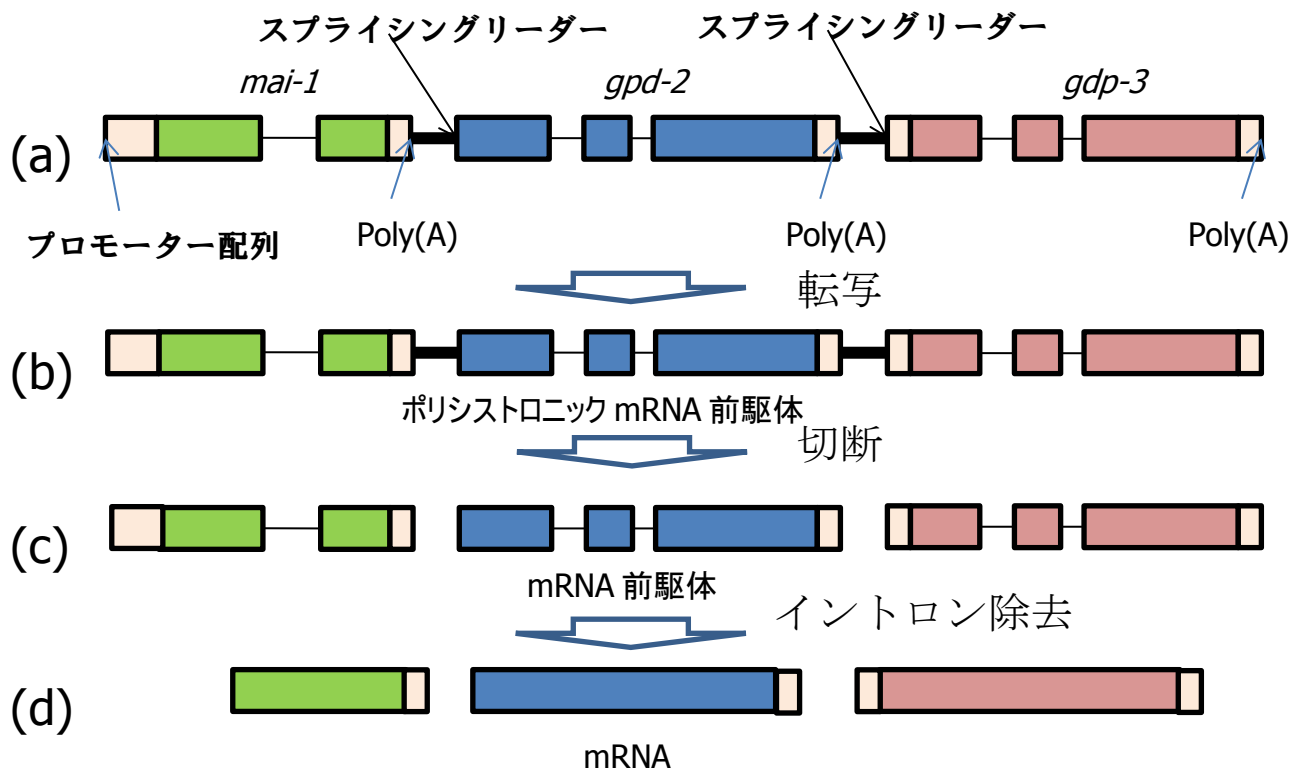
図 1 大腸菌のラクトースオペロンにおける遺伝子発現制御機構

## 1.2 真核生物におけるオペロン

真核生物において同一の mRNA に複数の遺伝子が転写される、いわゆるバクテリアにみられるオペロンが報告されている。真核生物においては線虫 (*Caenorhabditis elegans*) においてオペロンとしての遺伝子発現機構が最初に発見された (図 2)。線虫ゲノムにおいて 3 つの遺伝子、*mai-1* (ATPase inhibitor-1)、*gpd-2* (glyceraldehyde 3-phosphate dehydrogenase-2)、*gpd-3* (glyceraldehyde 3-phosphate dehydrogenase-3) は第 10 番目の染色体ゲノム上に並置する。図 2 において 3 つの遺伝子は、緑、青、桃色の四角で示されている。なお、橙色の四角は非翻訳領域を示す。3 個の遺伝子を含む mRNA が *mai-1* の上流に位置するプロモーターから転写される (図 2b)。これにより 3 遺伝子をコーディングした、ポリシストロニック mRNA 前駆体が作られる。次はポリシストロニック mRNA 前駆体が切断され、3 つの mRNA 前駆体を作られる (図 2c)。ここではまず先頭の遺伝子 *mai-1* の mRNA の 3' 末端のポリ A 配列までがひとつの mRNA 前駆体として切られる。続いて残りの遺伝子 (*gpd-2* と *gpd-3*) がトランススプライシング (trans-splicing) により mRNA が切り出される。トランススプライシングとは、シススプライシング (cis-splicing) がイントロン両端を切りイントロンを取り除くのに対し、イントロンの 3' 側を切る仕組みである。ここでは、small nuclear ribonucleoprotein (snRNP) を繋ぎかえることで、配列を置換して余分な配列を切断している。1 番目と遺伝子と 2 番目の遺伝子の間の距離は ~100bp とされ塩基 U が豊富な領域があり、下流に位置する遺伝子を守る働きがあり、下流の遺伝子の 5' 末端でトランススプライシングが起こるまでに必要である。2 番目の mRNA 前駆体の上流のスプライスリーダーでトランススプライシングが起こり、5' 末端が snRNP に置き換わることで、もとの配列は切断される。また 3' 側は一番目の遺伝子と同様にポリ A 配列で切断され、2 番目の mRNA 前駆体が切り出される。

3番目の遺伝子についても同様である。最後に、3つの mRNA 前駆体からイントロンが取り除かれ、3本の成熟した mRNA となる(図 2d)(Blumenthal and Gleason, 2003; Blumenthal, 2004; Girard et al., 2007)。このように、線虫においては、複数の遺伝子が同一の mRNA として転写されるものの、それぞれの遺伝子ごとにタンパク質に翻訳される mRNA は遺伝子ごとに存在することになる。

*Caenorhabditis elegans* のオペロン



Blumenthal and Gleason, 2003 *Nat.Rev.Genet* 4, 112-20

図 2 オペロンに注目した線虫の遺伝子発現機構



### 1.3 植物におけるオペロン様遺伝子群

植物を含む真核生物の遺伝子の発現は原核生物のように、複数の遺伝子が同一の mRNA に転写されるわけではない。基本的にはオペロンはバクテリアにおける発現制御である。また植物の場合、線虫のようなオペロンの転写は報告されていない。しかし、オペロンのように共調発現する遺伝子群が報告されており、さらに特定の代謝物の生合成経路に関わる遺伝子について、オペロン様遺伝子群(Operon-like gene clusters) として報告された(Field and Osbourn, 2008)。ここでは、植物におけるオペロン様遺伝子群の報告例を整理する。

シロイヌナズナにおけるオペロン様遺伝子群は 5 番染色体において隣接し存在し、発現相関の高い 4 つの遺伝子がシトクロム P450 遺伝子に該当する 2 つの遺伝子を含み、かつ triterpene-thalianol の生合成経路に関与するという報告がされている(Field and Osbourn, 2008) (図 3)。真核生物の線虫のオペロンと植物のオペロン様遺伝子群の比較図は図 4 のようになる。

シロイヌナズナにおいてはさら隣接する遺伝子の発現傾向を解析するアプローチとして、ゲノム上における距離、そのコードされている向きについて、といったゲノム構造からの発現解析研究も行われており、(Ren et al., 2005)、また遺伝子の距離について、クロマチン構造に注目する研究もなされている(Chen et al., 2010)が、隣接する遺伝子は共発現する傾向が高いことが確認されている。

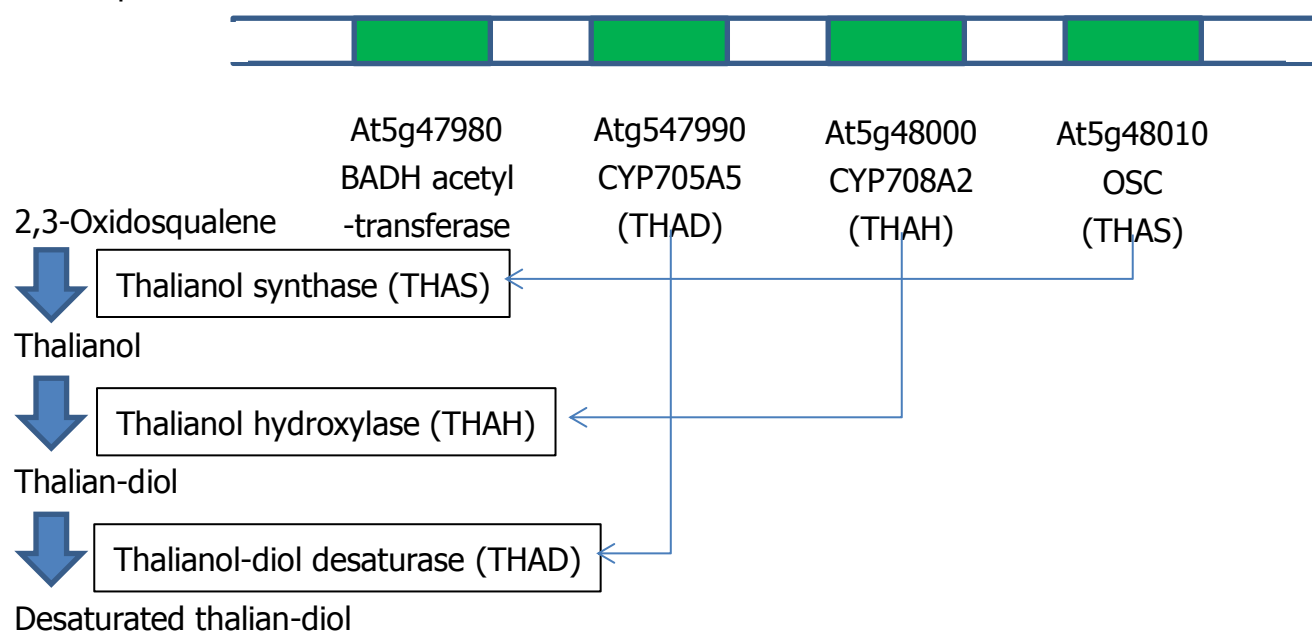
イネにおいては、キチンオリゴ糖誘導体や紫外線により生産される momilactone の合成経路に 2 つのシトクロム P450 コードが関与し、またイネの 4 番染色体上において遺伝子群となっている(Shimura et al., 2007)。

遺伝子のコードされている方向に対しての研究ではイネ(rice)やポプラ(*populous trichocarpa*)という生物間における遺伝子発現の保存性についても検討されているが、それぞれの生物ごとにおいては統計的に高い発現相関を示す 2

遺伝子の組み合わせ(*divergent* や *convergent*)であっても、他の生物においてはその遺伝子対の保存性は低いとされる(Krom and Ramakrishna, 2008)。また、イネ(*rice*)の報告ではゲノム中の共発現領域はだいたい 2~4 遺伝子からなるものであり、ゲノム中の 5%の遺伝子が含まれていると報告されている(Ren et al., 2007)。

酵母から糸状菌、線虫、植物、哺乳類といった多くの真核生物において、共発現している機能遺伝子の集中する領域を見ることができる(Hurst et al., 2004; Koonin, 2009)。これらの遺伝子群は、ゲノム上で近距離に存在し共に調節を受けているといった、オペロンのような特徴を有した構成となっている。植物、酵母、動物などの多くの真核生物における隣接遺伝子の発現は、偶然というよりは頻繁に共発現している傾向がみられる (Hurst et al., 2004)。また、真核生物のゲノムにおいては、高いレベルで共発現している機能的な隣接遺伝子が 12%含まれているという報告もある(Al-Shahrour et al., 2010)。ただし、キネトプラスト類(Clayton, 2002)や線虫(Blumenthal and Gleason, 2003; Guiliano and Blaxter, 2006; Qian and Zhang, 2008)といったオペロンを持つとして報告されている生物においても、原核生物にみられるような一本の mRNA としての転写構造をとっているわけではない。

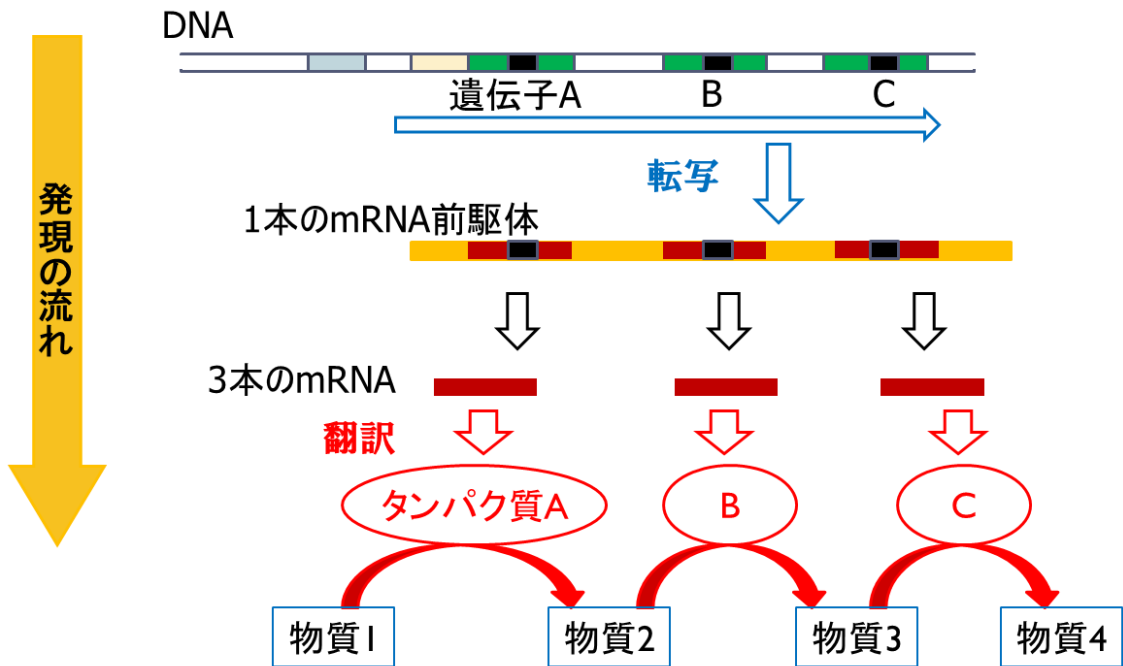
Arabidopsis Chromosome 5



Field & Osbourn, *Science* Vol. 320. no. 5875, pp. 543 – 547, 2008

図 3 シロイヌナズナ(*Arabidopsis thaliana*)の thalianol 生合成系と関連遺伝子群

真核生物におけるオペロンの発現の流れ



植物におけるオペロン様遺伝子群の発現の流れ

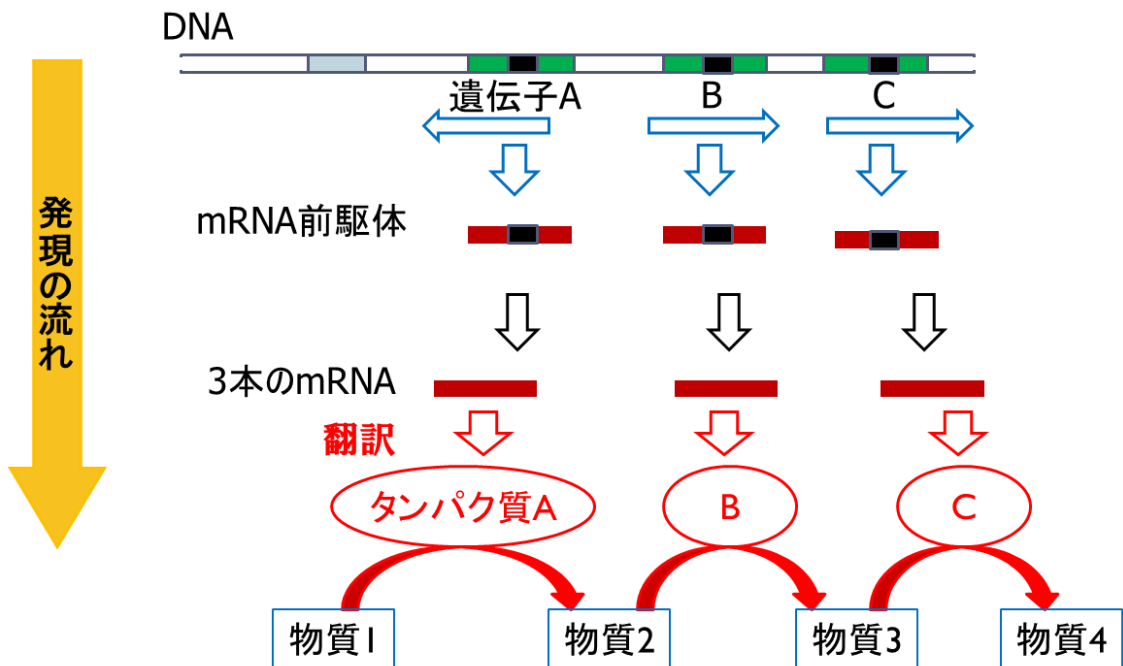


図 4 真核生物のオペロンと植物のオペロン様遺伝子群の比較図

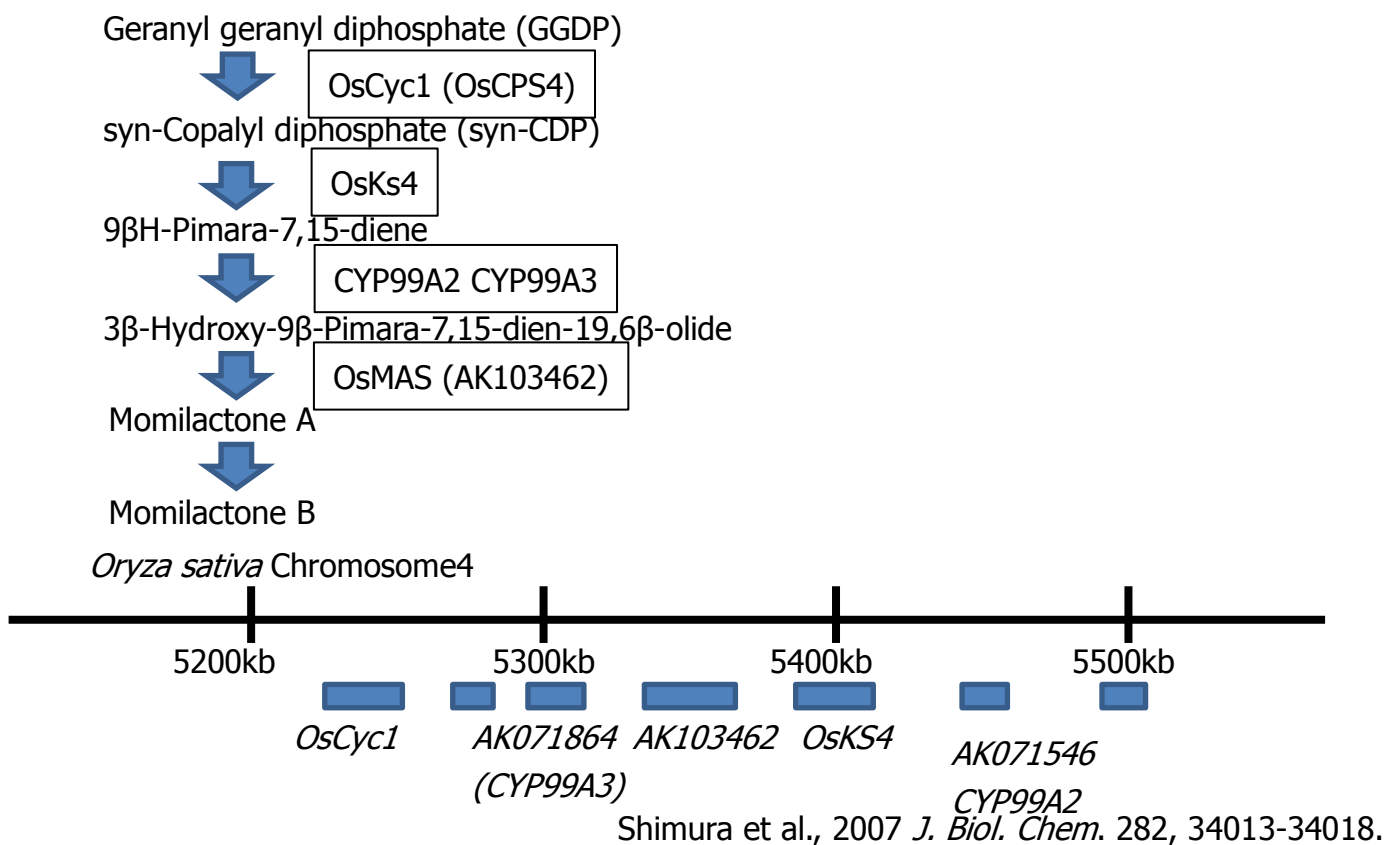


図 5 イネ(*Oryza sativa*)のモミラクトン生合成経路と関連遺伝子群

さらに、トウモロコシの benzoxazinones (Niemeyer, 1988; Bailey and Larson, 1991; Frey et al., 1995; Frey et al., 1997; Gierl and Frey, 2001)、オート麦の avenacin(Osbourn et al., 1994; Papadopoulou et al., 1999; Qi et al., 2004; Qi et al., 2006; Mylona et al., 2008)、rice の momilactone(Shimura et al., 2007)、シロイヌナズナの triterpene (thalianol) などにおいてオペロン様の遺伝子発現が報告されるに至る。トウモロコシでは 5 段階の代謝物質を経由する例も報告されており、これは indole から benzoxazinone 2,4-dihydroxy-1,4-benzoxazin-3-one (DIBOA)への合成経路となっており、トウモロコシの 4 番染色体に集まっている。このうち 4 つの遺伝子は P450 ファミリーに属する遺伝子である。この代謝経路の発現は植物の生育時に見られ、発芽後の 4 日、根と茎において benzoxazinone が高いレベルで現れている(Bailey and Larson, 1991; Frey et al., 1995; Frey et al., 1997; Gierl and Frey, 2001)。この Benzoxazinone は苗木の時期における一般的な防御機構に関係している(Niemeyer, 1988)。

また、オート麦における avenacin の代謝経路では遺伝子群となっている 7 つもの遺伝子が関わりとされている(Qi et al., 2004; Mylona et al., 2008)。そして、avenacin A-1 は 5 段階の酸化酵素反応によって合成され(図 7)、この経路に関わる酵素も Cytochrome P450(P450 遺伝子)に分類される遺伝子である(Qi et al., 2006)。またこの avenacin A-1 という物質はオート麦の茎の表皮においては広範囲における防御に使われ塩害に対する抵抗性を持ち(Papadopoulou et al., 1999)、局所的に合成されているとされる(Osbourn et al., 1994)。

このように植物において、機能的に関連性のある遺伝子群が共発現しているような事例が数多く報告されている。

これらの事象から、このように機能において関連があり、ゲノム上で隣接する遺伝子群について先行研究に倣い、オペロン様遺伝子群と呼ぶ。(Field and

Osbourn, 2008; Osbourn and Field, 2009)

植物の P450 遺伝子においては遺伝子重複により新機能獲得が起こっており (Matsuno et al., 2009)、それがオペロンのようなクラスターの形成に結果として現れていると考えられる (Field and Osbourn, 2008; Flowers and Purugganan, 2008)。以上の事からオペロンのような遺伝子発現クラスターは研究対象として生物学的にも重要な機構であると考えた。

遺伝子発現群は、アレロパシーやファイトアレキシン生成および生体異物からの防御といった機能に関連する。またシロイヌナズナにおいて同一代謝経路にある隣接遺伝子は共発現する傾向がみられると報告されている (Williams and Bowles, 2004; Ren et al., 2005)。そのため共発現遺伝子群を決定することができれば、詳細のわからない遺伝子の代謝経路上での働きを理解する助けとなる。

今回の研究では公開されているシロイヌナズナの発現データをもとに、オペロンのように共発現している遺伝子群(オペロン様遺伝子群)の総数を推定することを目的とした。

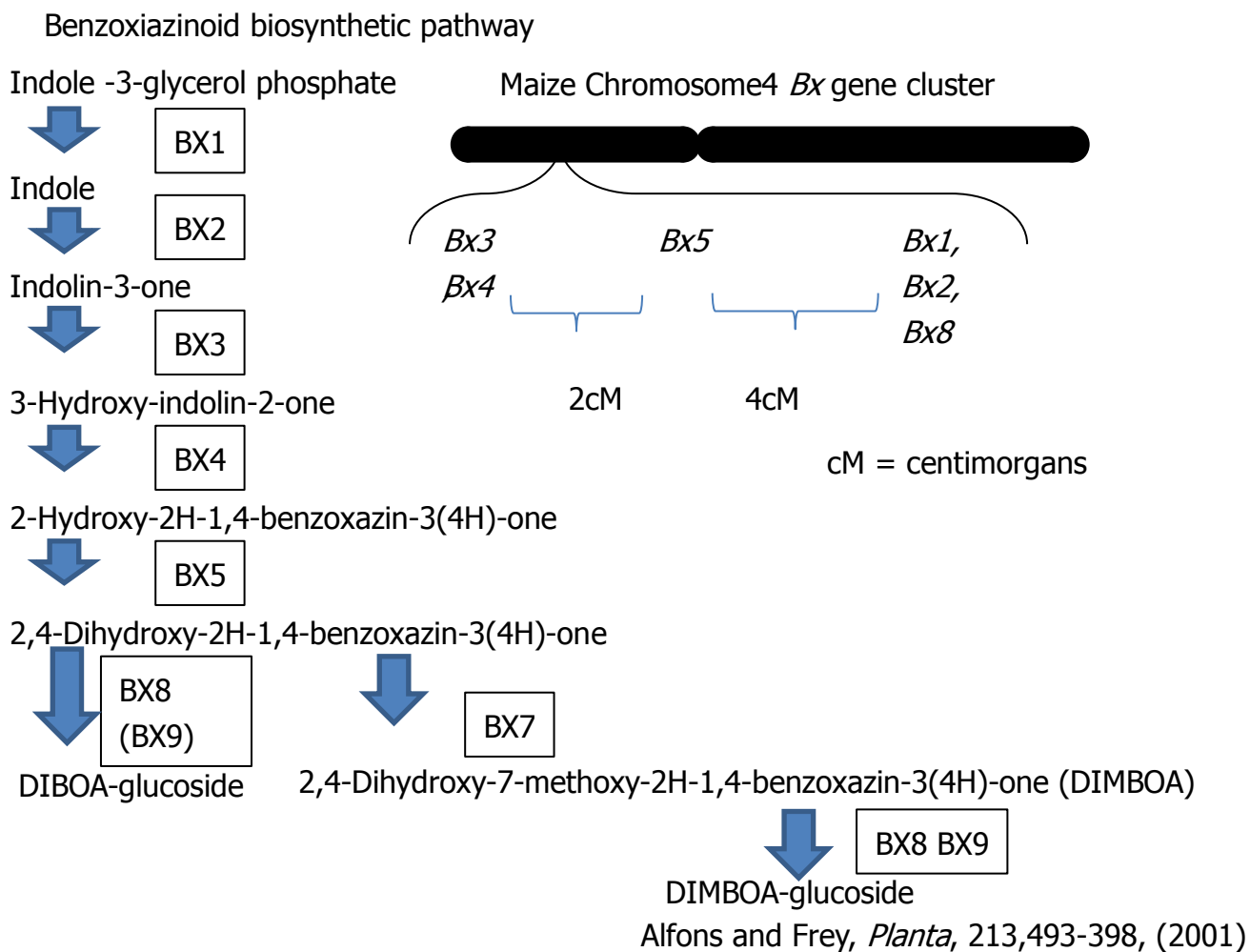


図 6 トウモロコシ(*Zea mays*)の **BX** 遺群と **DIBOA** 生合成経路



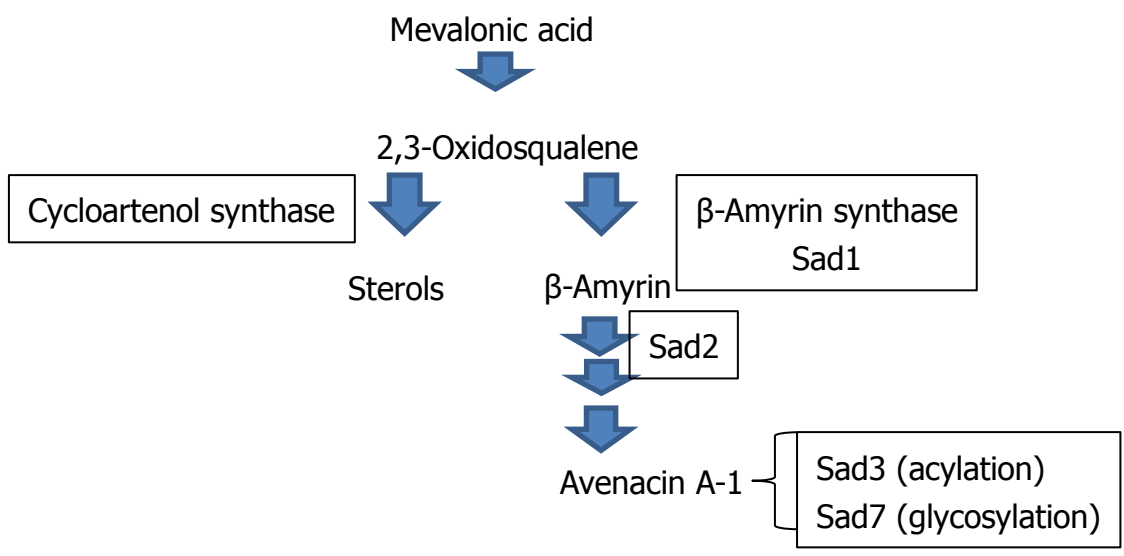


図 7 Avenacin 合成系の *Sad* 遺伝子群

## 1.4 遺伝子共発現におけるバイオインフォマティクス要素技術

ゲノムが明らかにされたモデル生物については、その発現情報を網羅的に取得し解析する手法がとられてきた。これまでの実験手法に加え、バイオインフォマティクスによる解析を行うことにより新たな知見を得る試みがなされるようになり、その解析結果を実験に生かすことにより生物研究の進展を見ることができた。このようにして生物情報を解析する分野はバイオインフォマティクスとして多くの手法による解析がなされている(Chuang et al., 2012)。そこで原核生物においてはどのような情報をもとにオペロン予測がなされているかについて整理する。

オペロン予測に関してはまず第 1 にゲノムにおける遺伝子間の距離に注目した方法が提案された。これは同じオペロンに属する遺伝子において遺伝子間が近接して配置されている傾向を統計的に把握する方法である。2 つ目は代謝経路に注目した方法である。オペロンに属する遺伝子群は規則的もしくは連続して機能し代謝経路を構成しており、遺伝子がどのような代謝経路上で働いているかという情報はオペロン決定の重要な情報となりうる。第 3 の方法としてはプロモーター、ターミネーターといった、遺伝子発現において特徴となる遺伝子上流、下流の配列情報を利用する方法である。さらに 4 つ目は、遺伝子の発現により作られるタンパク質の配列相動性に基づくもので、既知のタンパク質の配列を機能ごとにクラスタリングし、その特徴に当てはまる遺伝子を同機能と分類する、遺伝子の機能からのアプローチである。

発現を観測するための大量のデータ取得を行う上で、マイクロアレイ、なかでも高密度オリゴヌクレオチドアレイといった技術が用いられる。この発現デ

ータに、オペロン予測に関連する情報である遺伝子間距離、代謝経路、発現プロモーター、オルソログス遺伝子、遺伝子の長さやオペロン自体のサイズ、さらにはデータベースによるものなどを関連付けすることにより解析を行えるが、これらはバイオインフォマティクスとしては統計学的に、また機械学習のような方法により行われ、予測においては情報科学におけるデータマイニング技術が応用された。

このように多面的に情報が増えるにつれ、情報解析を統合することによりさらに高次の知見を高精度で得ることが可能となっている。例えば枯草菌におけるマイクロアレイによる発現データと配列解析からレギュロンと呼ばれる転写単位を予測するといったことも可能となった。これでゲノム配列から、転写因子の結合部位の予測を行い、その下流に位置する遺伝子群のマイクロアレイによる発現情報を組み合わせることにより可能となっている。また、機能相動性をもとにそれぞれの遺伝子群を決定した研究がなされている(Kobayashi et al., 2007)

## 1.5 本論文の構成

1・1～1・4における知見を要約すると表 1 のようになる。

表 1 原核生物および線虫におけるオペロンと植物のオペロン様遺伝子群の共通点の整理

	原核生物	真核生物	
		線虫	植物
隣接遺伝子	○	○	○
mRNA におけるポリシストロニック構造	○	○	
発現相関	○	○	○
遺伝子間距離	○		
機能類似性	○	○	○

表 1 によると、オペロン様遺伝子群を探索するに当たり植物においては(1)隣接遺伝子群、(2)遺伝子の発現相関、(3)遺伝子の機能類似性によりオペロン様遺伝子群を推定できると期待される。そこで第 2 章～第 4 章においては統計的評価によりオペロン様遺伝子群を検出する方法を説明する。第 2 章において、今回使用するデータセットにおいて、ターゲットとする隣接遺伝子群が過去の研究において報告されているような遺伝子の発現傾向がみられるかを確認した。そのうえで遺伝子群を探す指標とするために、どれぐらいの遺伝子数セットをオペロン様遺伝子として扱うか、閾値となる遺伝子数セットを決定した。第 3 章では隣接遺伝子群の中から、発現相関にもとづき共発現している遺伝子群をオペロン様遺伝子群として定義するための推定方法について述べた。そして第 4 章では、第 3 章において推定したオペロン様遺伝子群に含まれる遺伝子をオントロジー解析による機能推定と既知の知見にもとづいて考察し、オペロン様遺伝子群の生物学的意味を検討した。

## 第 2 章 オペロン様遺伝子群のゲノムからの探索法

### 2.1 シロイヌナズナの発現プロファイルにおける隣接遺伝子間の有意性検定

オペロンのような発現調節を受ける、また結果として同時期に働くような遺伝子を見るには、遺伝子が隣接していること、さらに隣接遺伝子が共調して働くことが確認される必要がある。今回ターゲットするオペロン様遺伝子群に含まれる遺伝子の発現は似たような発現傾向を有し、統計的に有意であることを示す必要があると考えその検定を行った。

図 8 にこの章の解析の概略を示す。まず利用するデータは様々な実験において、Affymetrix 社の ATH1 genome array によって得られた発現のデータである(図 8-1)。ただし、実験ごとのバックグラウンドのノイズがあるのでこれを補正する(図 8-2)。さらにそれぞれの実験においても、実験誤差をなるべくなくすために、複数回の実験が存在するものを用いて平均値を出すことを考慮し、繰り返し実験があるものを対象としてしぼりこんだ(図 8-3)。このデータを元に、実験区間 599 件について、22591 遺伝子の発現プロファイルを作成した(図 8-4)。発現プロファイルを用いて、全遺伝子について 1 対 1 対応のピアソン相関係数を求めた。これを遺伝子同士の相関として用いた(図 8-5)。このデータのうち、染色体

上での位置情報をもとに、隣に存在する遺伝子とペアを取れるものを選び、隣接遺伝子対をターゲットとしたデータ 16972 件に絞り込んだ(図 8-6)。全ての遺伝子発現ペアの中から、ランダムに 169720 件のデータセットを取りだし、隣接遺伝子のデータセット比較を行った(図 8-7)。ここから統計的に有意となる遺伝子クラスター数を確認した(図 8-8)。以上の作業を経て、今回のターゲットの遺伝子セットのサイズを決定した(図 8-9)。この決定したサイズをもとに、次章で、FDR をもとにオペロン様な遺伝子領域の決定を行った。

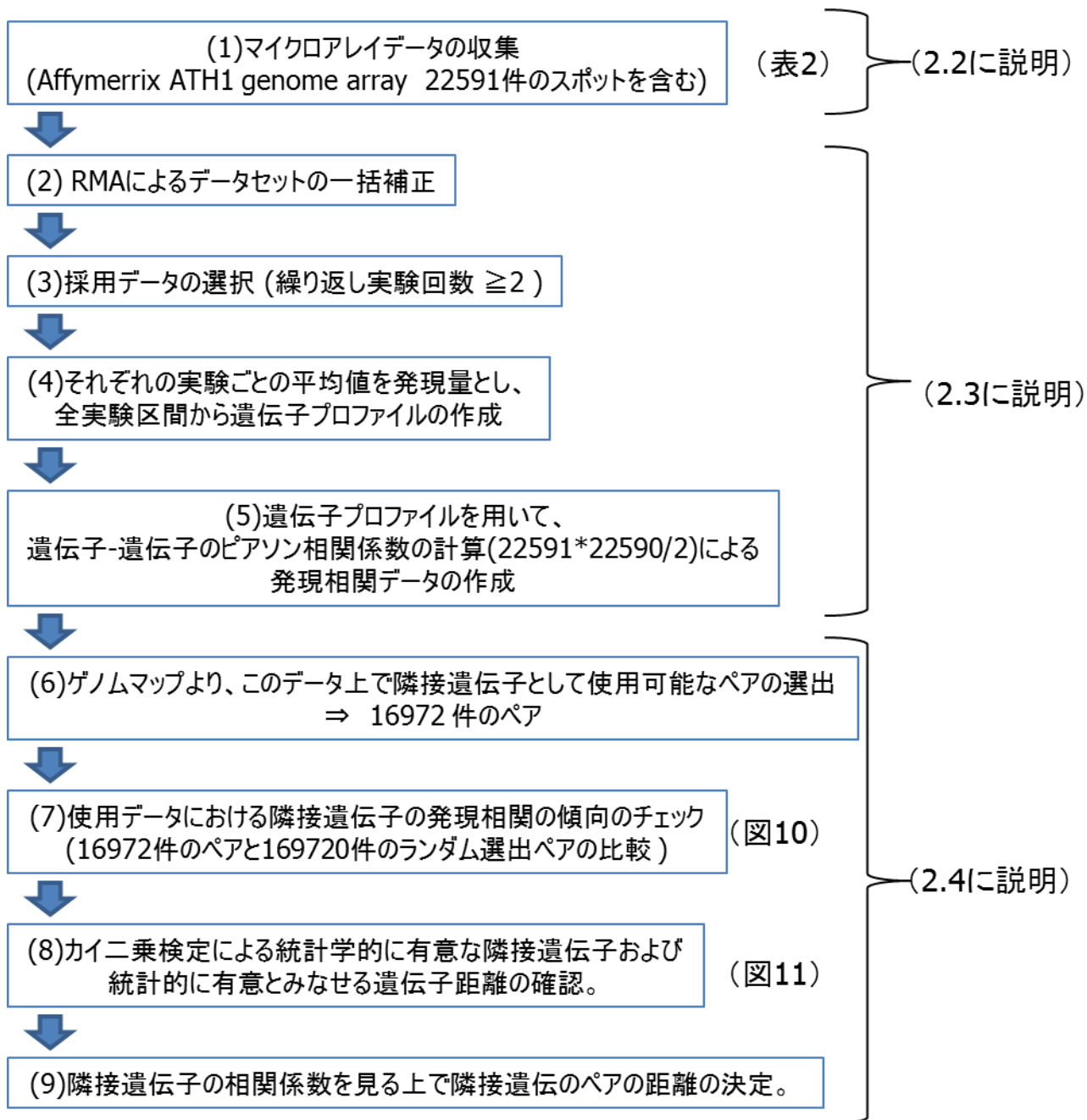


図 8 オペロン様遺伝子群の抽出法の概要

## 2.2 使用データ

本研究では、各々の遺伝子の発現相関の有無とゲノム上の隣接性を考慮して、オペロン様遺伝子群を統計的に推定することを目標とする。そのため、できる限り多くの実験条件で測定された発現プロファイルデータを集めて解析に用いることにした(Fukushima et al., 2008)。そこで、公開されているデータベースから、ATH1 Genome Arrays の GeneChip による測定データ 1469 種を集めた。ATH1 Genome Array はシロイヌナズナの 22591 の遺伝子の発現を一度に計測できる GeneChip®である。表 2 に使用した GeneChip の測定条件の概要を示すように、17 種の植物の組織において発現プロファイル測定を行っており、本研究における目的を達成するための発現プロファイル測定の多様性が反映されると期待される。なお、1469 枚のアレイは以下の公開データベースから無償で得ることができる。

The Arabidopsis Information Resource (TAIR) (Huala et al., 2001) (977 Arrays; <http://www.arabidopsis.org/portals/expression/microarray/ATGenExpress.jsp>),

European Arabidopsis Stock Centre (NASC) (474 Arrays; <http://affymetrix.arabidopsis.info/narrays/help/helpindex.html>),

Gene Expression Omnibus (GEO) in NCBI (18 Arrays; <http://www.ncbi.nlm.nih.gov/geo/>)



表 2 本研究に用いた GeneChip

Tissue	the number of experimental conditions	total number of experiments
Aerial part	16	35
Cell suspension	6	18
Cotyledons	4	9
Flower	27	75
Hypocotyl	5	14
Leaf	118	327
Petiole	3	9
Pollen	4	9
Root	85	181
Rosette leaves	22	66
Seedling	179	430
Shoot	74	154
Shoot apex	24	52
Stamen	2	6
Stem	7	17
Tumor	2	4
Whole plant	21	63
Total	599	1469

## 2.3 データの補正

GeneChip の蛍光強度から発現量を求める方法として Robust Multi-array Average (RMA) (Bolstad et al., 2003; Irizarry et al., 2003)を使用して補正を行った。この処理に当たっては、統計処理ソフトウェア R の生物統計用のパッケージ、Bioconductor を用いており、そのマイクロアレイを処理する Affy パッケージを使用した。Robust Multi-array Average (RMA)法により、ミスマッチプローブを使わない、パーフェクトマッチのみを用いた補正を行った。RMA は、バックグラウンド補正を行う RMA 折畳み補正(RMA convolution)、分位数正規化(quantile normalization)、中央値分散分析(Median polish)の 3 ステップからなる(Bolstad, 2004)。

1469 枚のアレイは、それぞれ同一の繰り返し実験による発現データを含んでいる。そこで、このような繰り返し実験については平均値を代表値に用いた。この過程を経て、一つの遺伝子について 599 種の発現データを得た。これにより 22591 遺伝子について、599 の実験区間からなる発現プロファイルデータを構築した。この発現プロファイルを使い、遺伝子間のピアソン相関係数を得ることで、共発現の指標として用いた。

## 2.4 隣接遺伝子の発現方向および、遺伝子間の距離の比較

共発現のうち、今回ターゲットとなる隣接遺伝子のペアは16972ペアである。これは現在、公開されている GeneChip で使用されているデータのうち、染色体上の位置情報をもとに隣接関係にあると判別できる物の数である。

この研究においてターゲットとなる隣接遺伝子間の共発現関係について統計的に有意な差が得られるかを見るためにカイ二乗検定( $\chi^2$ -statistics)による、ランダムに抽出したペアによるデータ分布と、隣接遺伝子によるデータ分布の比較を行った。

このランダムなデータセット  $S_{rand}$  は全遺伝子の遺伝子遺伝子発現相関から統計的に有意とみなせるだけの量として全部で 169720 個のペア(隣接する遺伝子ペアの 10 倍をもとに決定)をランダムに抽出したものである。この全データセット数は  $T$  とする。

$S_{rand}$  はピアソン相関の-1 から 1 にあたる区間におけるデータのヒストグラムである。この分布に区間を設定する。それぞれの区間におけるデータ数が全て同じになるように相関係数で区切り、その数を  $M$  とすると、1 区間あたりに存在するデータ数は  $|S_{rand}|/M$  と表すことができる。今回、 $M = 20$  とし、区間あたりのデータ数は 8436 個とした。

ランダムに抽出した相関係数データをマイナス側から数え上げ、8436 個ごとに相関係数の区切りを設けて、実際の隣接遺伝子のペアのデータがそれぞれの区間にいくつ含まれるかを数える。

この実際のデータをそれぞれの区間は  $c(i = 1, \dots, M)$  とし、存在するデータ数は  $r_{i-1} \leq r < r_i$  に含まれる  $n_i$  とする。以上より次式により検定を行った。

$$\chi^2 = \sum_{s=i}^M \frac{(n_i - T/M)^2}{T/M}$$

ここで、 $n_i$ は相関係数において $r_{i-1} \leq r < r_i$ の区間に存在する実データにおけるペアの数である。また、 $T$ はターゲット全体のペアの数である。この研究においては、先に述べたように $M=20$ とする。このカイ二乗値における  $p$ -values はカイ二乗分布において自由度 $\phi = M - 1$ を用いた。

以上の方法により、隣接遺伝子、隣接する遺伝子の発現の向き、さらに隣接する遺伝子の間にいくつかの遺伝子が入った場合にその発現傾向がどのように変わっていくかを解析した。

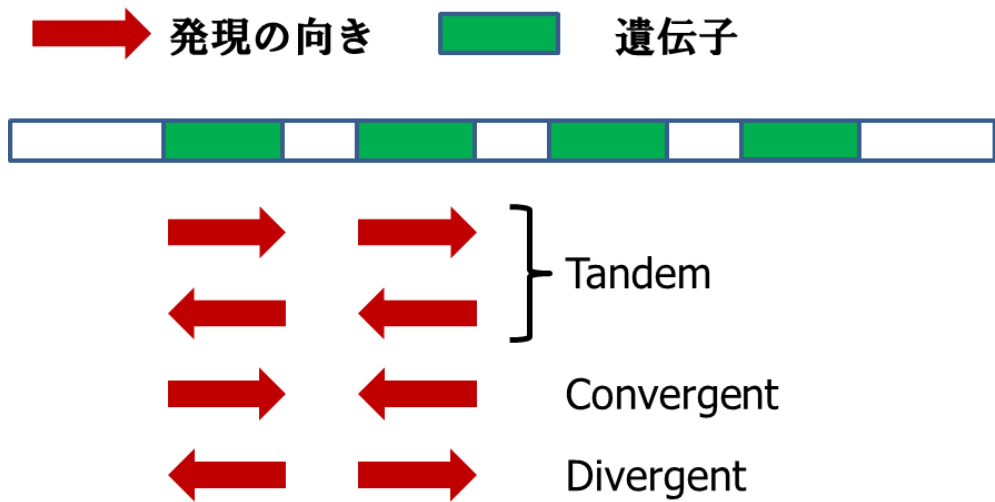


図 9 発現の向きと相関解析の例

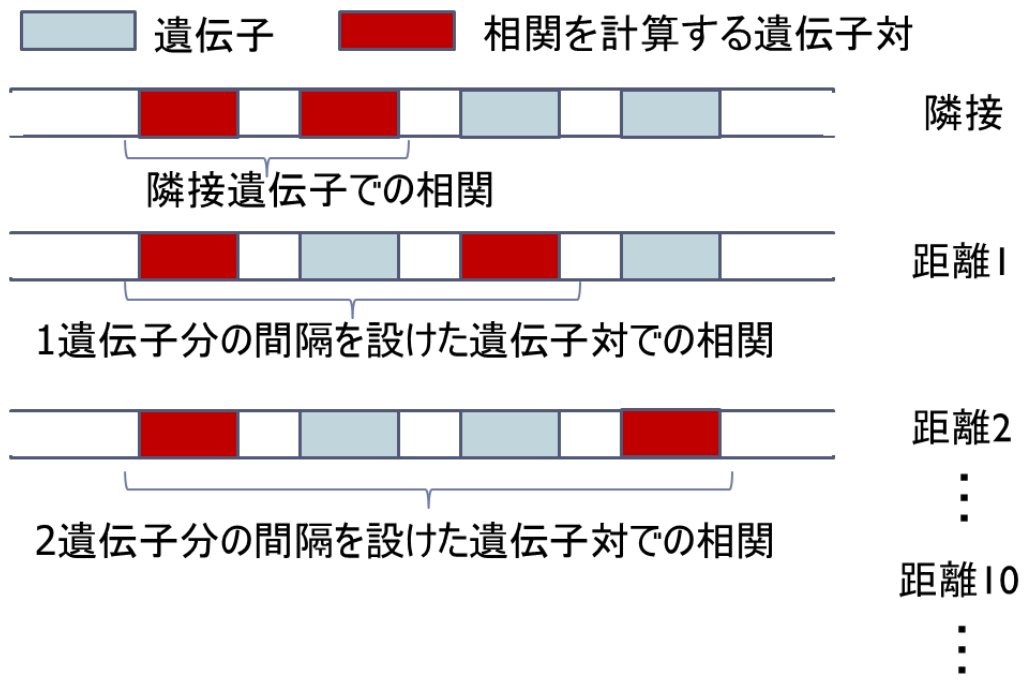


図 10 隣接遺伝子間距離の相関解析の例

まず使用するデータセットの隣接遺伝子について、いくつかの比較検定を行った。

隣接遺伝子についての検定のほか、その向きごとに tandem ( $\rightarrow\rightarrow$ )、convergent ( $\rightarrow\leftarrow$ )、divergent ( $\leftarrow\rightarrow$ )といった関係においても検定を行った。そして、隣接遺伝子の場合、間の距離を 0 としてこれを N0、さらに間に一つ遺伝子が入る距離 1 を N1 とし、N1 から N10 までについて検定を行った。

図 11 はまず、N0 から N4、そして tandem ( $\rightarrow\rightarrow$ )、convergent ( $\rightarrow\leftarrow$ )、divergent ( $\leftarrow\rightarrow$ )の分布とランダムデータの分布を比較した結果である。このグラフは相関係数-1.0 から 1.0 までを  $r_i$  で表された 20 区間に分けて比較したものである。比較方法はそれぞれの区間においてデータ個数  $n_i$  の全量に占める割合から、ランダムの一律 5%を引くことで、ランダムデータと比較した場合の偏りを示したものである。この三つの分布において tandem のペアにおけるカイ二乗値( $\chi^2$ -value)とその  $p$  値は  $\chi^2 = 1316.15$ ;  $p = 1.2 \times 10^{-267}$  となり、convergent ( $\chi^2 = 104.5$ ;  $p = 8.16 \times 10^{-14}$ ) および divergent ( $\chi^2 = 403.57$ ;  $p = 7.95 \times 10^{-74}$ )より、大きい値となった。このように tandem 方向に遺伝子発現をする隣接遺伝子ペアの相関係数は、他の二つのタイプの方向のペアに比べより共発現している傾向が見られた。また divergent タイプの遺伝子についても tandem タイプほどでないにしろ、有意に強く表れている。ただ、ヒトのゲノムにおける報告では、divergent タイプの転写がされる隣接遺伝子ペアにおいて、特に強い共発現が見られるとある (Trinklein et al., 2004)が、シロイヌナズナのゲノムの発現プロファイルにおいては、どの方向性のタイプにおいても隣接遺伝子は有意な相関を示す傾向があることが見られる。

図 11 の結果に示す N1~N4 におけるカイ二乗値と  $p$ -value はそれぞれ、 $\chi^2 =$

442.48,  $p = 6.14 \times 10^{-82}$ ; N1,  $\chi^2 = 253.17, p = 7.04 \times 10^{-43}$ ; N2,  $\chi^2 = 123.71, p = 2.23 \times 10^{-17}$ ; N3,  $\chi^2 = 125.75, p = 9.26 \times 10^{-18}$ ; N4 となり、カイ二乗値は、隣接遺伝子における値に比べ、遺伝子間に一つとばし、二つとばしと間に遺伝子が増える、つまり遺伝子間の距離が広がるにつれて多くの場合で減っていくことを指し示した。どの値においても、これらの数値は p-value から統計的に有意であるといえる。しかしながら、これらの結果は、もっとも近い隣接遺伝子 N0 にくらべ、どれも小さい値を取った。図 12 は 同様に N10 までターゲットとする遺伝子ペアの間の遺伝子数を増やしたものである。このデータの数値は過去の報告にある、シロイヌナズナの遺伝子において隣接遺伝子からなる 2~4 遺伝子の共発現遺伝子の発現クラスターが 5~10% 存在するという報告と一致する部分を見ることができた (Ren et al., 2005; Zhan et al., 2006)。今回使用するデータは、過去に研究されたデータと同じような性質を持つデータセットを用いたものであることが証明できたと考える。

N11 においてはカイ二乗値が  $\chi^2 = 48.86, p = 1.92 \times 10^{-4}$  となり、統計的に有意であったが、この p-value はかなり大きい。さらに線虫における 1 オペロン中の遺伝子数は 8 個 (Blumenthal et al., 2002)、と報告されていることから、今回、遺伝子群を推定するに当たっては、この数字以上の大きさも考慮したいと考えた。p-value を見ると、N5~N8 では p-value が  $10^{-11} \sim 10^{-12}$  の範囲で推移しており、N9、N10 と低くなる。N8 なら、間にある遺伝子は 8、全体に含まれる遺伝子数は 10 であり、前述の線虫の例もカバーできさらに統計的に有意でもある。このことから、今回、推定する遺伝子数を 10 個までとした。

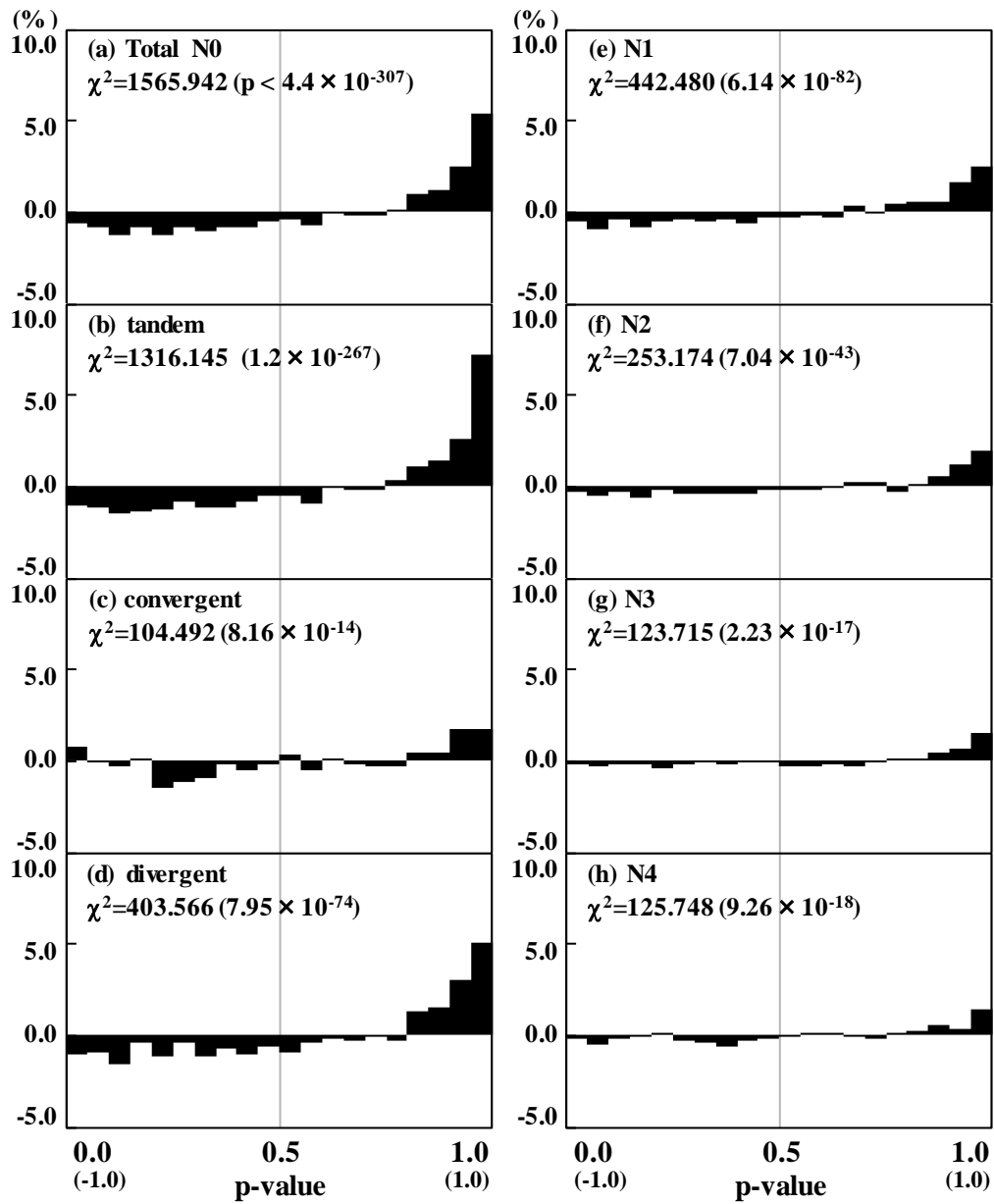


図 11 シロイヌナズナにおける隣接遺伝子の共発現の有意性検定の結果



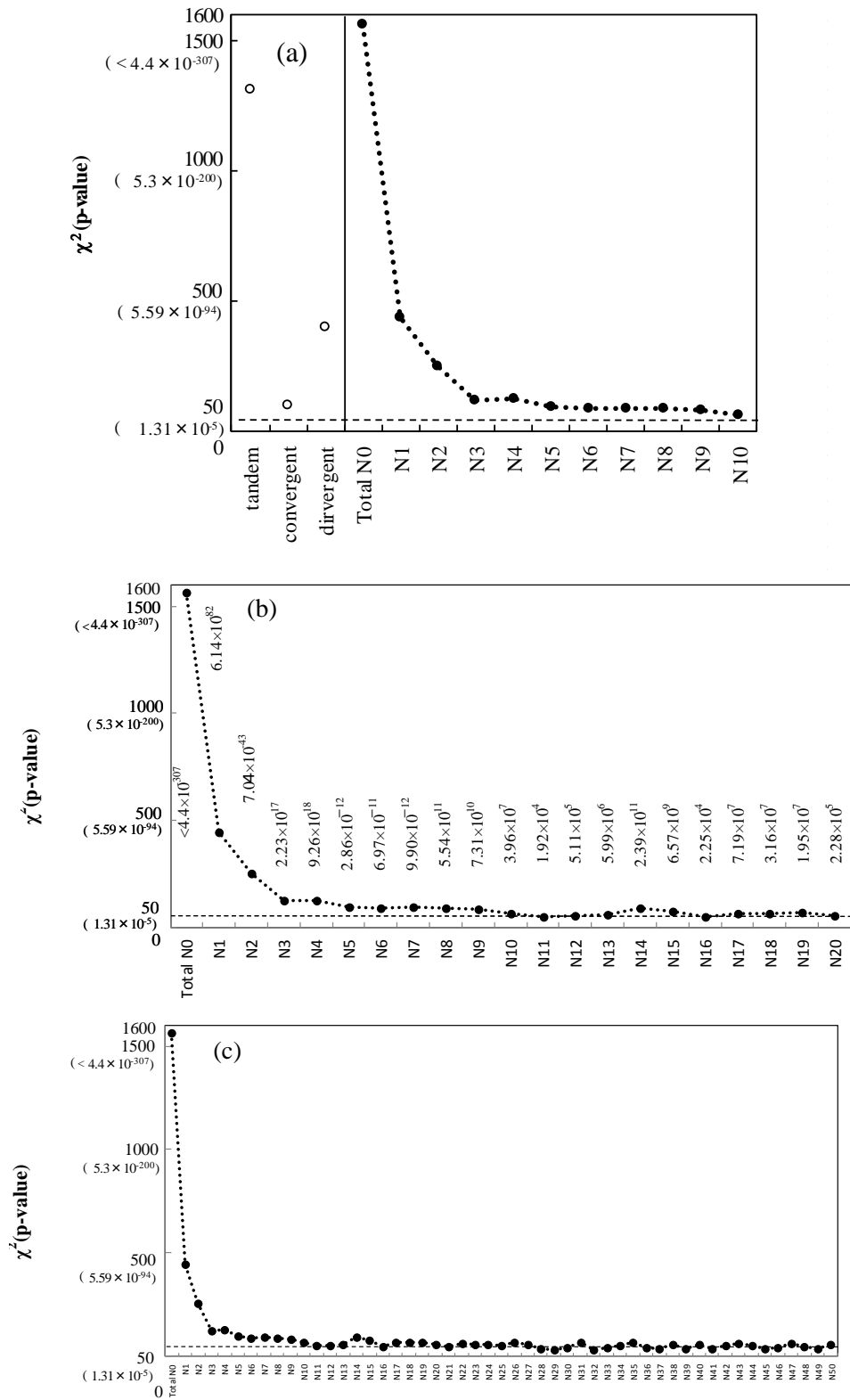


図 12 隣接遺伝子および 2 遺伝子間の距離の拡大とカイ二乗値の遷移

真核生物のゲノムにおける共発現遺伝子群の発現機構は、ゲノム上の局所においては双方向への発現が可能なプロモーター領域や共通のエンハンサーの利用によるものや、ゲノム全体においては、染色体上においての類似の発現パターンを示すようなクロマチン構造によるものがある(Hurst et al., 2004)。クロマチンを考慮した調節機構の存在はしばしば重要なものとみなされているが(Cremer and Cremer, 2001; Sproul et al., 2005)、小サイズの遺伝子群における共発現遺伝子の大部分は、発現調節による遺伝子の共調によるものだと考えられる(Semon and Duret, 2006)。

今回使用する発現データにおいて、16972 個の近接する遺伝子ペアは、ランダムに選択した遺伝子ペアと比較して共発現する傾向を持つことが示された。従って、シロイヌナズナにおいて見られる隣接する共発現遺伝子は、ゲノムにおいて統計的に有意な共発現をする遺伝子群を見つけ出す方法を用いるのに、現実的な適切なデータであると考えられた。

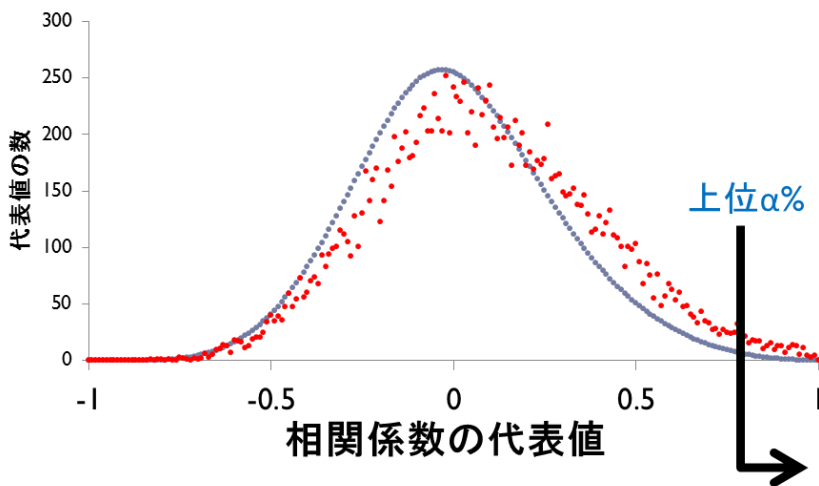
## 第 3 章 FDR を用いたオペロン様遺伝子クラスターの推定

### 3.1 方法

シロイヌナズナのゲノムはおよそ 27000 遺伝子とされているが、そのうち、ATH1 Genome array により発現強度を測定できるのは 22591 遺伝子となる。現状では全遺伝子をカバーするには至らない。そのため、隣接遺伝子の相関係数についてもすべての遺伝子について計算できるわけではない。使用可能な隣接遺伝子ペアはシロイヌナズナの 5 本の染色体それぞれで隣接であるとされているものの位置を考慮して決定する。よって、本研究で、隣接遺伝子として相関係数を計算することができたのは 16972 個のデータ隣接ペアということになった。この 16972 個の相関係数は  $N_{real}$  と表すことにした。そして、単純に連続する  $n$  個の遺伝子からは  $n - 1$  個の相関係数が存在することになる。

ここで、本研究の目的は False discovery rate(FDR)(Benjamini and Hochberg, 1995)をもとに統計的に有意な相関係数の値による隣接遺伝子のグループを選択することである。遺伝子ペア  $x$  での有意なレベル  $\alpha$  は以下のようにして計算した。

隣接遺伝子群とランダム選出の遺伝子群の  
相関係数の分布比較の概念図



隣接遺伝子群のデータセットの  
上位α%  
⇒理論値(期待値)

ランダムにN個の隣接遺伝子対の  
相関を選択し上位α%の下限となる  
代表値を決める

連続する隣接遺伝子対N個の代表値  
>閾値  
(ランダムな隣接遺伝子N個の  
代表値による分布のα%)  
⇒観測値

$$FDR = \frac{\text{理論値}}{\text{観測値}}$$

図 13 本研究において FDR を扱う概念図

**Step 1:** 比較する遺伝子ペア数と同じ数という考えに基づき、ランダムに  $x$  ペアを  $N_{real}$  から選びその平均値を出す。

**Step 2:** Step 1 を 100000 回繰り返してすべての操作における平均値を得る。

$$R(x) = \{ \bar{r}_i(x), i = 1, 2, \dots, 100,000 \}$$

**Step 3:**  $R(x)$  の要素 100000 件を数値の降順で並べ替え、 $R(x)$  の上位  $\alpha\%$  に対応する相関係数の値を閾値として  $Thr(\alpha)$  とする。このとき  $R(x)$  で用いた  $\alpha\%$  に対応するだけの実際の相関係数のデータ個数を  $N_{rand}(\alpha)$  とする。

**Step 4:** 実データにおける連続する  $x$  個の相関係数における平均値が  $Thr(\alpha)$  より大きい箇所を  $N_{real}(\alpha)$  とする。ただし、連続する相関の性質上、負の相関を含むものは除く。

**Step 5:** FDR の計算は  $FDR(\alpha) = N_{rand}(\alpha)/N_{real}(\alpha)$  から導く。

この研究では  $x = 1, 2, 3, 4, 5, 6, 7, 8, 9$  を使い、 $\alpha = 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1.0$  を  $x$  に相当する値として用いた。

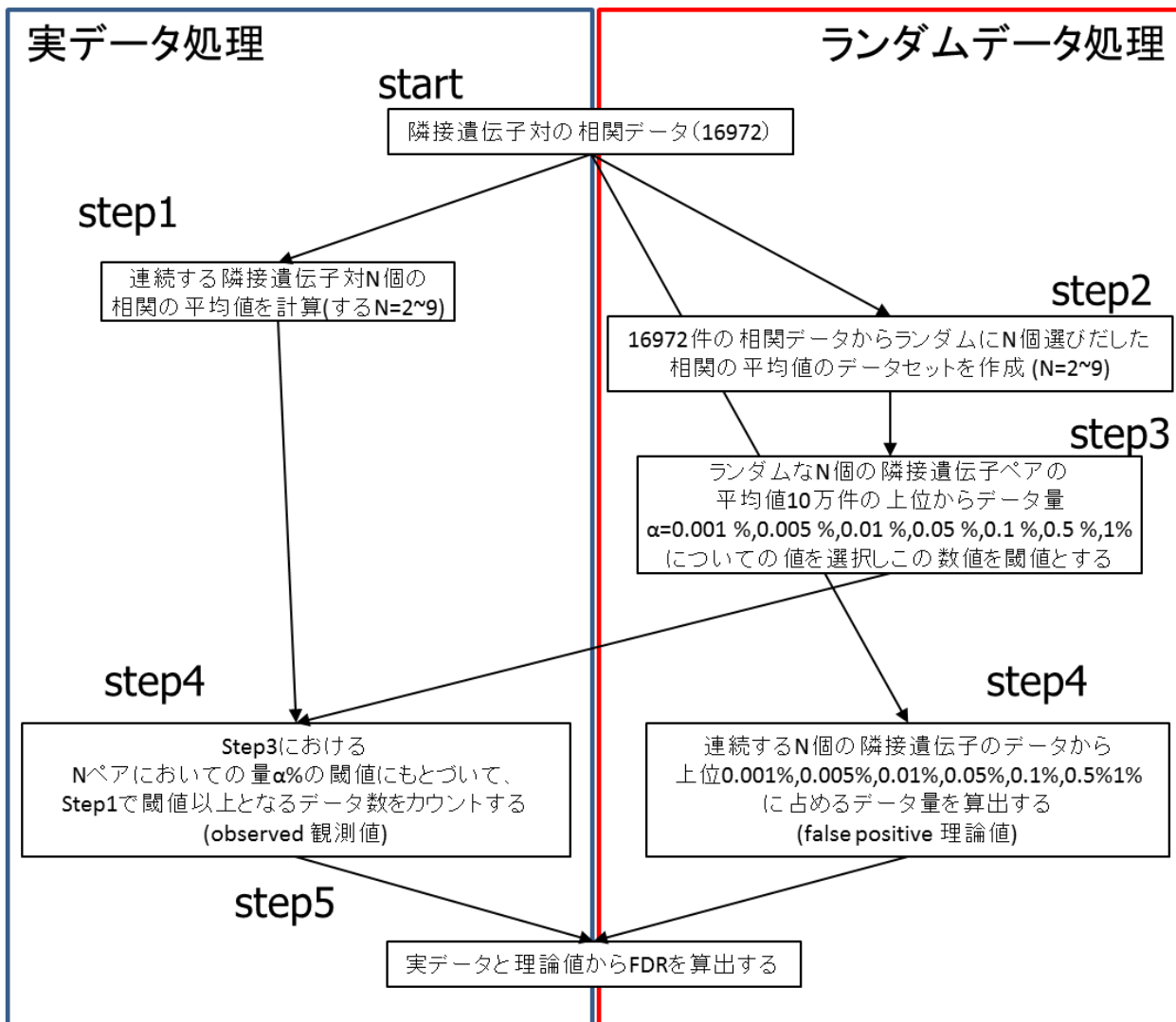


図 14 隣接遺伝子データから実測値、理論値を導き FDR を決定する処理の過程図

## 3.2 結果と考察

本研究の目的は、統計的に有意に、16792 件のデータセットから遺伝子グループを選択することであった。この目的において FDR(Benjamini and Hochberg, 1995)を用いることで、伝統的な多重検定を用いた遺伝子グループ決定を行った場合において発生する問題となる、相関は高いが偽陽性であるような結果の場合における、さらに種々の検定を考慮しなければならないという問題に対応できたと考える。この方法では設定したサイズの隣接する遺伝子の相関係数の値の平均値を、同サイズで設定したランダムに選択した遺伝子の相関係数の平均値と比較することをベースとしている。この比較検定ではランダムな選択による平均値の決定を 100000 回繰り返しその分布を推定した。この方法においては設定としては 2~10 遺伝子を選択することにより統計的有意となる閾値を  $\alpha$  ごとに求め、それぞれの  $\alpha$  の場合における FDR、 $FDR(\alpha)$  を決定するという考えを持って進めた。また連続する  $n$  個の遺伝子においては  $n-1$  の相関係数があると考えで行っている。ただし、相関係数をベースとしているので、連続を取ることにより遺伝子数と相関係数の数はこのルールに一致しない場合もある。最終的に候補とするオペロン様な遺伝子群を推定するに当たり、重複する箇所は結合を行った。

今回の研究では単純な 2 遺伝子だけの遺伝子群の相関係数、これは隣接遺伝子の相関係数そのものであるが、この相関が 1 個のデータからなるデータセットは FDR が高くなりすぎる(FDR1 を超える)ため、対象データから除いた。

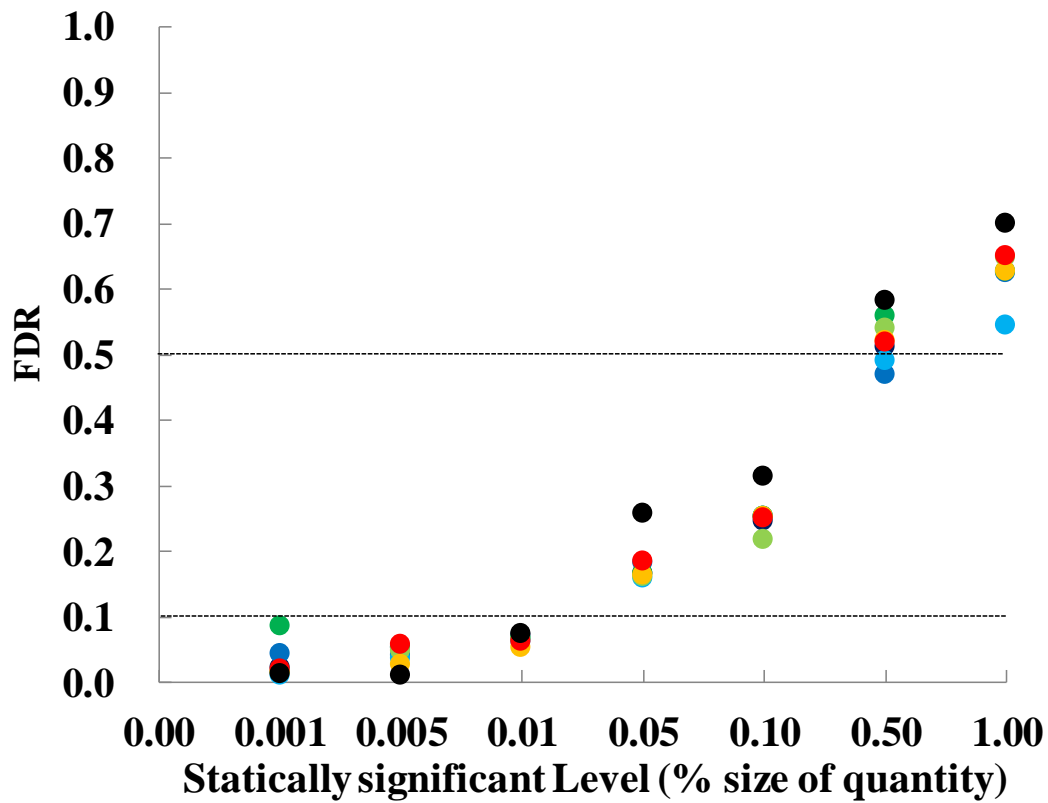


図 15 各 window サイズにおける上位  $\alpha$  % における FDR の変化  
 FDR と  $\alpha$  (0.005%~1%) の関係 遺伝子群のサイズの代表的なサイズを基準として、黒 (●) は 3、赤 (●) は 4、橙 (●) が 5、黄緑 (●) が 6、緑 (●) が 7、水色 (●) が 8、青 (●) が 9、藍色 (●) が 10 で表す。



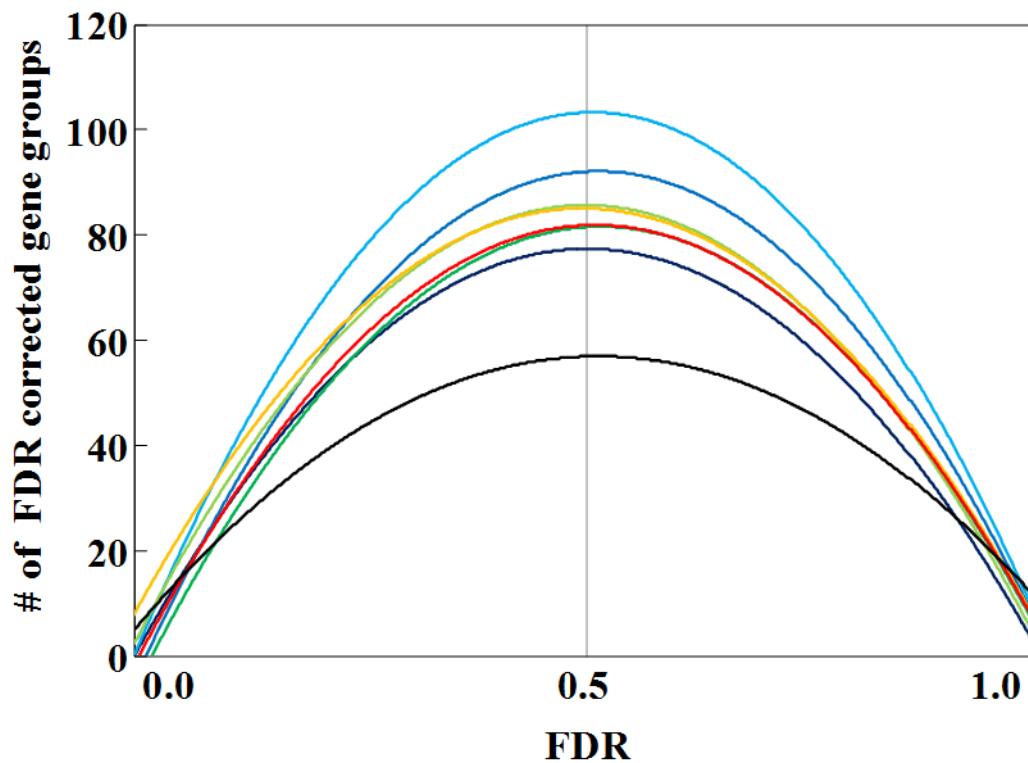


図 16 FDR と、FDR の値から求まる遺伝子群の数から近似曲線により推定される FDR0.5 における陽性とみなされるクラスター数

遺伝子群のサイズと示す線の色、その頂点の値と近似曲線の相関は以下のようになる。3 ; 黒 (-) ; 57(0.911)、4 ; 赤 (-) ; 82(0.969)、5 ; 橙 (-) ; 85(0.933)、6 ; 黄緑 (-) ; 86(0.977)、7 ; 緑 (-) ; 90(0.929)、8 ; 水色 (-) ; 103(0.914)、9 ; 青 (-) ; 92(0.960)、10 ; 藍色 (-) ; 77(0.996)

図 15 のプロットグラフは今回の研究で用いた $\alpha$ において、遺伝子グループごとに FDR がどの値を取るかを表したものである。 $\alpha$ が小さくなると、使用するデータ量が少なくなり、FDR が小さくなる傾向がみられる。つまり FDR が小さくなる条件とは、 $\alpha$ が小さい状態であることがわかる。 $\alpha$ が小さく設定される場合、取得した遺伝子群においては、その大部分が真の陽性値をとる遺伝子群であるという事がわかる。またここで、今回使った $\alpha$ とその対応する遺伝子数グループの数を  $N$  とすると、偽陽性を除いたデータを推定することもできると考えた。このようにして FDR をもとに陽性と考えらえる遺伝子群数を  $(1 - \text{FDR}(\alpha))$  を推定した。図 16 はその推定されたデータをもとに FDR と FDR より導いた陽性データ数の関係性をもとに描いた二次回帰曲線である。なお、この回帰曲線は信頼に値するものであり ( $r > 0.911$ )、推測された遺伝群の数との間に高い相関があると考えられる。そこで、ここでの数字をもとにシロイヌナズナにおけるオペロン様な遺伝子群について最大でいくらとりうるかを推定した。図 16 において曲線の最大値となるのは、おおよそではあるが、 $\text{FDR}=0.5$  のときであると仮定する。これは FDR を指標とするため、データ総量の半分が  $\text{FDR}=0.5$  において偽陽性となると考えられるからである。そしてこの  $\text{FDR}=0.5$  という値は図 15 においては  $\alpha = 0.5\%$  が相当する。 $\alpha = 0.5\%$  における、遺伝子群のサイズ 3~10 のデータを整理し、重なるところは結合すると、 $\alpha = 0.5\%$ 、 $\text{FDR}=0.5$  においては 201 遺伝子群が推定される。つまり今回用いたデータからシロイヌナズナは、おおよそ  $(1 - 0.5) \times 201 = 100.5$ 、おおよそ 100 個のオペロン様な遺伝子群を持っていると推測できることになる。しかし、この閾値で予測される遺伝子群は、そもそも半分が間違いのものとなる。ここで FDR が 0.1 となる閾値に注目すると、この場合に予測される遺伝子群においては、間違っって予測されたものが含まれ

る可能性は、10 件あたり 1 件ということになる。さらに  $\alpha$  下げれば、予測される遺伝子数はさらに減るが、謝って推測される遺伝子群自体も減らすことができるとも考えられる。今回、利用可能な最小の  $\alpha=0.001\%$  においては、63 個の遺伝子群が予測され、重複される部分を統合すると、13 個の遺伝子群が推定されたものとして取得できた。ただし、この場合の FDR は、遺伝子サイズごとに、10 : 0.02、9 : 0.04、8 : 0.01、7 : 0.08、6 : 0.02、5 : 0.01、4 : 0.01、3 : 0.01 となっている。最も高い FDR である 0.08 を基準に考えると、13 個の予測される遺伝子群の中に、まだ 1 個の間違った遺伝子が含まれることになるが、0.001% における FDR の平均を見ると、0.02 であり、この場合なら含まれる誤判定されるデータは 1 件以下とみなすことも可能である。また、ひとつだけ高い FDR を持つデータについては、実際にターゲットとして実験を進める上では、このデータを使わないと言う選択も可能である。今回扱った遺伝子グループサイズすべてにおいては、 $\alpha$  が 0.01% の時にはすべての場合において、FDR が 0.1 以下となっていることから、この閾値なら FDR の値自体を考慮して、統計的に有意な遺伝子群を選択できていると考えた。この  $\alpha = 0.01\%$  においては全部で 212 個の遺伝子群を推定した。この 212 個の遺伝子群から重複を排除し最終的に 34 個の遺伝子群を得た。たとえば、この 34 個の遺伝子群について、FDR をもとに評価すると  $\alpha = 0.01\%$  での FDR は図 14 では最大で 0.07 相当に見ることができるので、 $34 \times (1 - 0.07) = 32$  の遺伝子群が統計的に有意なものと考えることができる。とはいえ、隣接遺伝子については、前章で示したように相関が高い傾向もあり、またシロイヌナズナについても隣接遺伝子は共発現する傾向が高いと言われており (Williams and Bowles, 2004)、*thalianol* 合成経路として報告された一つの遺伝子群にある 4 遺伝子が酵素として働くオペロンのような形となることが説明できると考えた (Field and Osbourn, 2008)。つまり、共発現の調節や、機能的に関

連する遺伝子群であることが見られるのがよい証明と考える。よって、この研究では統計的にもやや厳しめの数値で有意であると考えた $\alpha = 0.01\%$ の閾値での遺伝子群に焦点をあて、その遺伝子の特徴を見ることによって、遺伝子群がオペロン様と呼べるにふさわしいものを含んでおり、FDR をもとにした遺伝子群推定が有用であることを確認するという方向性を考えた。

表 3 ウィンドウサイズごとランダムデータにより設定した閾値と、そのときの閾値以上となる推定される遺伝子群の数の関係

rate of random 100000	windows size								
		9	8	7	6	5	4	3	2
0.001%	Ave. Cor	0.5479	0.6043	0.5771	0.7118	0.6797	0.7314	0.8454	0.9651
	Ap. Data	8	4	16	2	11	12	9	13
0.005%	Ave. Cor	0.4968	0.4916	0.5259	0.5915	0.6378	0.66	0.7763	0.929
	Ap. Data	16	21	23	16	18	32	15	19
0.010%	Ave. Cor	0.4652	0.4728	0.5051	0.5583	0.6022	0.6507	0.7473	0.8907
	Ap. Data	24	27	26	26	26	33	27	23
0.050%	Ave. Cor	0.4041	0.4314	0.4544	0.4954	0.5353	0.5976	0.6815	0.7938
	Ap. Data	51	47	54	51	53	52	46	33
0.100%	Ave. Cor	0.3715	0.3969	0.4304	0.4656	0.5006	0.5645	0.642	0.747
	Ap. Data	69	67	67	67	78	67	68	54
0.500%	Ave. Cor	0.2661	0.3079	0.3408	0.3825	0.4248	0.4798	0.5534	0.6549
	Ap. Data	166	181	173	152	157	163	164	146
1.000%	Ave. Cor		0.232	0.2864	0.3391	0.3814	0.4373	0.5036	0.6017
	Ap. Data		272	312	271	262	271	261	243

「Ave. Cor.」はクラスター内の相関に適応する閾値となる値、「Ap. Data」は実際に閾値を満たしたクラスターの数、また赤色は今回の研究で用いた FDR0.1 以下となる閾値でのクラスター数である。

## 第 4 章 遺伝子アノテーションに基づく、オペロン様遺伝子の評価

FDR により、統計的に有意な方法を用いて、相関係数をもとにした共発現領域を定義することはできた。しかしながら、オペロンの定義にあるように、機能についての知見がないままで、オペロン様な遺伝子を定義したとは言えないと考える。よって、FDR によって抽出した遺伝子群について、そこに含まれる遺伝子について、種々の遺伝子アノテーション情報を使い情報を付加することで、それぞれの遺伝子群がオペロン様として予測に値するものであるかを解析した。

### 4.1 クラスター内遺伝子の機能アノテーション方法

シロイヌナズナの遺伝子の機能アノテーションについて、最初に“Arabidopsis Gene Classifier” (Takahashi et al., 2009) を用いた。このデータベースは、公開データベースである複数の研究結果をもとに、さらに他のリファレンスデータ、主に発表された論文から読み取れる情報をもとに構築されており、複数の遺伝子のアノテーションを見ることができるよう設計されている。特にもととなったデータセットとしては MIPS (Mewes et al., 2002) や RARTF (Iida et al., 2005) という大規模なものや、ほかのデータベース (Shiu and Bleecker, 2001; Wintz et al., 2003) の内容を含めたものとなっている。オペロン様遺伝子群については次の式

により、カイ二乗検定を用いて検定を行った。

$$\chi^2 = \sum_{i=1}^r \frac{\left(n_i - N \frac{t_i}{T}\right)^2}{N \frac{t_i}{T}} \quad (2)$$

ここで  $n_i$  と  $t_i$  はそれぞれクラスター内に含まれる遺伝子とデータベース上に登録されている遺伝子の数であり、 $i$  番目の機能に属する 遺伝子の数となり ( $i = 1, 2, \dots, r$ ) それぞれ Arabidopsis Gene Classifier による。  $N$  はあるクラスターに含まれた遺伝子数であり  $T$  はデータベース上での遺伝子の総数である。この検定においては “unclassified protein” と表記される機能未知は含まない。機能の数は “unclassified protein” の機能クラスを除いて、 $r$  は 22 としている。ここでの  $p$ -value はカイ二乗分布より求め、このとき自由度は  $\phi = r - 3$  とした。また最後にボンフェローニ法によりカイ二乗値を調整した (Holm, 1979)。

さらに、アノテーションについては遺伝子クラスター内の遺伝子を Gene Ontology によるアノテーションの付加も調べた。これは TAIR による web アプリケーションである GO slim terms を用いた (<http://www.arabidopsis.org/tools/bulk/go/index.jsp>) (付表 表 1)。さらに超幾何分布 (Hypergeometric distribution) をもとに Gene Ontology の  $p$ -value を計算する方法を採用している GOSTATS (Falcon and Gentleman, 2007) を使ったデータも参考にした (付表 表 2)。

また、クラスター内の遺伝子において BLAST により配列相動性を確認することにより遺伝子の重複による影響を考えた。これはシロイヌナズナの全遺伝子をデータベースとして、相動性検索を行い、E-value を閾値として  $10^{-5}$  以下となるものがクラスター内に含まれる場合、その遺伝子を重複遺伝子と考えた。

表 4 本研究の解析により統計的に有意な発現をしているとして得られたシロイヌナズナ遺伝子発現群 (3章における $\alpha=0.01\%$ 。さらに CID に下線があるものは $\alpha=0.001\%$ において確認された遺伝子群)。表上半分の遺伝子名についている色は、黒以外の色については、BLASTにより、遺伝子群内で同色の遺伝子と有意な配列相動性を見られた。表の株は特徴的な機能であり、遺伝子群内で多くを占めた機能について太文字で表す。また最下部、“\*”は5%有意水準、“\*\*”は1%有意水準を満たすものであったことを示す。(次ページに続く)

Chromosome	1									2				3			
Cluster ID (CID)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	At1g14930	At1g23530	At1g29430	At1g39270	At1g51830	At1g52050	At1g63450	At1g72820	At1g75850	At2g07190	At2g07620	At2g11210	At2g19720	At2g21320	At3g16430	At3g21830	At3g22240
	At1g14940	At1g23540	At1g29440	At1g39350	At1g51840	At1g52060	At1g63460	At1g72830	At1g75860	At2g07200	At2g07630	At2g11220	At2g19730	At2g21330	At3g16440	At3g21840	At3g22250
	At1g14950	At1g23550	At1g29450	At1g39430	At1g51850	At1g52070	At1g63470	At1g72840	At1g75870	At2g07230	At2g07672	At2g11230	At2g19740	At2g21340	At3g16450	At3g21850	At3g22260
		At1g23560	At1g29460	At1g40150	At1g51860	At1g52080	At1g63480	At1g72850	At1g75880	At2g07240	At2g07673	At2g11240	At2g19750	At2g21370	At3g16460	At3g21860	At3g22290
		At1g23570	At1g29470	At1g40230			At1g63490	At1g72860	At1g75890	At2g07280	At2g07680	At2g11270		At2g21380			At3g22300
		At1g23580	At1g29480	At1g41795			At1g63500	At1g72890	At1g75900	At2g07290	At2g07690	At2g11360		At2g21385			At3g22310
		At1g23590	At1g29490	At1g41797			At1g63540	At1g72900	At1g75910	At2g07300	At2g07691	At2g11370					At3g22320
		At1g23600	At1g29500	At1g41810			At1g63550	At1g72920	At1g75920	At2g07320	At2g07692	At2g11680					At3g22340
		At1g23610	At1g29510	At1g41820			At1g63560	At1g72930	At1g75930	At2g07330	At2g07693	At2g11690					At3g22350
		At1g23630		At1g41825			At1g63580	At1g72940	At1g75940		At2g07701	At2g11890					At3g22360
		At1g23640		At1g41860			At1g63590	At1g72950			At2g07702	At2g11910					At3g22370
		At1g23650		At1g41870			At1g63600				At2g07705	At2g12170					At3g22380
		At1g23670									At2g07706	At2g12190					At3g22400
		At1g23680									At2g07713	At2g12210					At3g22430
		At1g23690									At2g07714	At2g12230					At3g22435
		At1g23700									At2g07715						At3g22440
											At2g07718						
											At2g07719						
											At2g07721						
											At2g07722						
											At2g07724						
											At2g07725						
Number of Genes	3	16	9	12	4	4	12	11	10	9	22	15	4	6	4	4	16
Cell cycle and DNA processing					1						1						
Cell fate																	
Cell rescue, defense and virulence			1		3		8						1		4		
Cellular communication / Signal transduction		2	6		4		6	1									
Metabolism							1		7			2		2			2
Protein fate(folding, modification, destination)										1		1				4	1
Protein synthesis													4				1
Storage protein																	
Transcription							3	1			1	1		1			2
Transport facilitation								1	1		1			1			
Transposable elements, viral and plasmid proteins				1						3		4					1
Unclassified proteins	3	14	2	11			2		2	5	19	7		1			9
Significance level	N/A	-	**	-	-	**	-	**	-	**	*	**	**	**	**	**	**
Significance level (Bonferroni correction)	N/A	-	-	-	-	-	-	**	-	**	-	*	**	**	**	-	-



表 4(続き)

Chromosome	3			4							5						
	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Cluster ID (CID)	At3g26210	At3g28770	At3g30396	At4g05190	At4g12470	At4g15330	At4g16860	At4g22030	At4g23130	At4g27120	At5g07440	At5g28550	At5g38120	At5g45420	At5g47980	At5g52670	At5g61660
	At3g26220	At3g28780	At3g30400	At4g05200	At4g12480	At4g15340	At4g16870	At4g22040	At4g23140	At4g27130	At5g07450	At5g28560	At5g38130	At5g45430	At5g47990	At5g52680	At5g61670
	At3g26230	At3g28790	At3g30440	At4g05210	At4g12490	At4g15350	At4g16880	At4g22050	At4g23150	At4g27140	At5g07460	At5g28570	At5g38140	At5g45480	At5g48000	At5g52690	At5g61700
	At3g26270	At3g28830	At3g30450	At4g05230	At4g12500	At4g15360	At4g16890	At4g22060	At4g23180	At4g27150	At5g07470	At5g28580	At5g38150	At5g45490	At5g48010	At5g52710	At5g61710
	At3g26280	At3g28840	At3g30480	At4g05240	At4g12510	At4g15370		At4g22070	At4g23190	At4g27160	At5g07510	At5g28590	At5g38160	At5g45500	At5g48020	At5g52720	At5g61720
	At3g26290		At3g30490	At4g05250		At4g15380		At4g22080	At4g23200	At4g27170	At5g07520		At5g38170	At5g45510	At5g48030	At5g52740	At5g61730
	At3g26300		At3g30500	At4g05260		At4g15390		At4g22120	At4g23210		At5g07530		At5g38180	At5g45530		At5g52750	At5g61740
	At3g26310			At4g05270		At4g15400		At4g22130	At4g23220		At5g07540		At5g38210	At5g45540		At5g52760	
	At3g26320			At4g05280		At4g15410		At4g22140	At4g23230		At5g07550		At5g38220			At5g52770	
						At4g15415		At4g22150	At4g23240		At5g07560		At5g38290			At5g52780	
								At4g22160	At4g23250				At5g38300				
								At4g22190	At4g23260								
								At4g22200	At4g23270								
									At4g23280								
									At4g23290								
									At4g23300								
									At4g23310								
									At4g23320								
									At4g23330								
									At4g23400								
									At4g23410								
									At4g23420								
Number of Genes	9	5	7	9	5	10	4	13	22	6	10	5	11	8	6	10	7
Cell cycle and DNA processing				1							1						
Cell fate									1								
Cell rescue, defense and virulence			1			3							3			1	
Cellular communication / Signal transduction				1		2		1	18				1	1			
Metabolism	9					6		1	1		7		2		3		
Protein fate(folding, modification, destination)			1	7			1	2			2	1					
Protein synthesis										1	1						
Storage protein										4							
Transcription								2				1	2	1			
Transport facilitation					5			1	1				3				3
Transposable elements, viral and plasmid proteins			1														
Unclassified proteins		5	4			2		6		1		3	3	3	3	9	4
Significance level	-	N/A	-	**	**	-	*	**	**	**	-	**	*	-	*	-	-
Significance level (Bonferroni correction)	-	N/A	-	**	*	-	-	*	**	**	-	*	-	-	-	-	-

## 4.2 結果と考察

本節では、それぞれの遺伝子群を説明するに当たり、遺伝子群にナンバーをつけ、CID(Cluster ID)という略称を用いて説明を行う。推定された遺伝子群の結果について TAIR によるオントロジーをもとにした機能クラスタリングではまず多くの遺伝子群に、関連機能が集まっている遺伝子群を見ることができた。しかし、詳細を説明するには情報不足のため、さらに Gene Classifier や GOSTATS によるデータの分析を行った。その結果を表 4 に示す。遺伝子群内で相動性検索の結果が  $E < 10^{-5}$  となるものについては、その遺伝子名の色を変えてその対応を表している。CID = 9, 18, 23, 28, 32 の 5 つの遺伝子群は “metabolism” の機能クラスを持っていた。CID = 6, 8, 15, 24, 31 の 5 つの遺伝子群は “cell rescue, defense and virulence” を持っていた。CID = 3, 7, 22 の 3 つの遺伝子群は “cellular communication/signal transduction” を持っていた。CID = 27 は “storage protein や “protein synthesis” を持つ遺伝子群である。GOSTATS による GO によるクラスタリング結果は付表 表 2 に示したようになった。CID1 の 3 遺伝子は “response to biotic stimulus” は生体刺激に反応する遺伝子群であり、これは生体防御に働くと考えられる。CID3 では 9 遺伝子中、7 つの遺伝子が “response to stimulus” に分類され植物ホルモン刺激に応答する遺伝子であった。CID5 はタンパク質のリン酸化にかかわる遺伝子群であった。CID14 におけるほとんどの遺伝子は、機能的な詳細ではクラスタリングされないが、細胞の構成要素の項目より、クロロプラストの構成要素に関わる遺伝子群だとわかる。CID30 においては多くの遺伝子は “lipid transport” に属する。CID33 の遺伝子群は、金属イオン結合と輸送に関わる遺伝子群だとされている。CID34 で特徴的なのは分子機能として “ATPase activity” が固まっていることだ。このように遺伝子群の大部分は豊富

なカテゴリーのどれかに分類することができた。続いて、既知または関連性のある分子および生物学的知識を用いて考察を行った。

植物に関する先行研究によると、多くの代謝経路に関する遺伝子はクラスターとなっていないともいう報告もあるが、クラスター化は遺伝子の有益な組み合わせを継承するのに有益である(Gierl and Frey, 2001; Qi et al., 2004; Wong and Wolfe, 2005)。代謝経路の遺伝子群の途絶は有害な中間体の蓄積を引き起こす(Myrona et al., 2008)。“metabolism”と関連づけられる CID = 9, 18, 23, 28, 32 だが、このうち3つの CID = 18, 23, 32 は P450 を含む遺伝子群であり、種特異的な代謝経路を含むと考えられる(Mizutani and Ohta, 2010)。4つの酸化スクアレン環化酵素遺伝子 (OSC; At4g15340, At4g15370, At5g48010, At5g36150)はシロイヌナズナのゲノム上で P450 遺伝子群として存在し、ブラシカ属にだけ存在するものであり、二次代謝産物に関わると予測されている(Field and Osbourn, 2008)。また二つの遺伝子が CID23 に含まれているものである。4つの P450 遺伝子(At5g48010, At5g48000, At5g47990, At5g47980)を含む CID32 はすでにオペロン様として定義されていた遺伝子で、thalianol 代謝の反応経路で働く酵素(OSC, CYP708, CYP705, アクチルトランスフェラーゼ)をコードするものとして報告されている(Field and Osbourn, 2008)。3つの遺伝子、At5g48010、At5g48000、At5g47990 のゲノム上での並びは近縁種である *A. lyrata* でも確認される(Amoutzias and Van de Peer, 2008)。この研究で簡単にこの価値を述べるとすると、多くのアレイにおいてこれらは共発現しているということである。GOSTATS による *p*-value によるアノテーションにおけるクラスタリングにおいても CID23 と CID32 は有意な結果が見られ、GOterm において “triterpenoid biosynthetic process” を示しており、二次代謝系に関わるとものであることがわかる。またこの *p*-value をもとにしたアノテーションにおいては、CID23 では GO terms として “oxygen binding”、“heme

binding”そして“electron career activity”といった性質が見られた。

CID23 においては二つの酸化スクアレン環化酵素 (OSC)が P450 遺伝子群として示されている(Mizutani and Ohta, 2010)。一つ目の OSC 遺伝子群は baruol 生合成経路に関わる At4g15370 (*BARSI*; OSC)、At4g15360 (*CYP705A3*)、At4g15350 (*CYP705A2*)の三つからなる遺伝子群である。もう一つは arabidiol 生合成経路に関わる At4g15340 (*ATPEN1*; OSC) と At4g15330 を含んでいる。CID23 はこのすべてを含むことから、これらの遺伝子群は baruol と arabidiol の合成系において同時に働いていると考えられる。酸化スクアレン環化酵素については、CID32 に含まれる At5g48010 も該当している。この遺伝子の属する遺伝子群は thalianol 合成系の遺伝子群となっている。これら baruol や arabidiol、thalianol の明確な機能は報告されていない。しかし、thalianol 合成系遺伝子群のノックアウト実験、および、過剰発現実験においては、シロイヌナズナの成長における根の長さの変化がみられるという、シロイヌナズナの成長を調節する化合物であるという結果が示唆されている(Field and Osbourn, 2008)。この一例から、他の化合物においても同様の効果があると明らかには言えないが、機能未知の遺伝子をターゲットとする上での一助になると考えられる。

CID18 では 9 つの P450 遺伝子が遺伝子群を形成していると考えられる (*CYP71B23*; At3g26210, *CYP71B3*; At3g26220, *CYP71B24*; At3g26230, *CYP72B25*; At3g26270, *CYP72B4*; At3g26280, *CYP71B26*; At3g26290, *CYP71B34*; At3g26300, *CYP71B35*; At3g26310, *CYP71B36*; At3g26320)。この CYP71B サブファミリーに属する P450 遺伝子は non-A タイプと呼ばれ、種特異的な二次代謝産物の生合成経路の多様性に寄与していると考えられている。この CID18 での 9 遺伝子はすべて、GO term による分類では“oxidation-reduction process”、“oxygen binding, heme binding”、“electron career activity”の分類を持つ。これら CYP71B サブファミリ

一に属する遺伝子 17 個中 9 個が、この 3 番染色体の小さな領域に集中しているということになる。現在の結果が指し示すのは、これら 9 つの P450 遺伝子は共発現しているが、その代謝における作用はまだ明らかとなっていないということである。

さらに推定された遺伝子群について、共発現遺伝子群が機能的な関連性を持つ例を示す。CID13(*RPS15AB*, *RPL28A*, *RPL31A*, *RPS30A*)は細胞内気質のリボソームに関わる遺伝子群である。CID = 16, 21 はユビキチン修飾によるタンパク質分解系に関わる。“lipid and fatty-acid”のアノテーションがついた CID9 (*EXL1*, *2*, *3*, *4*, *5*, *6*, *ATA27*)と CID28 (*GRP14*, *18*, *17*, *16*, *19*, *20*)は脂質や脂肪酸の合成に関わる。CID22 および CID30 は脂質輸送に関わる。さらにこの CID22 は炭水化物代謝系に関わる。これらの所見はシロイヌナズナのゲノムにおいて、種々の生合成に関与する隣接遺伝子が、遺伝子量を調整するためには共調するよう発現を調整されていると示唆される。言い換えると、これらの遺伝子は、その遺伝子がコードするタンパク質はバイオマシナリーシステムを構成するために共調させられている(Williams and Bowles, 2004)。さらに結果を見ていくと、推定された遺伝子群について、未知の機能と思わせる機構が示唆された。それは独立するオペロン様遺伝子群がお互いの調節に関わっているのではないかというものである。顕花植物において、自他を区別するシステムは配偶者を認識する分子によるものが代表的である。CID9 に含まれる “extracellular lipase family II” のアノテーションを与えられる遺伝子群(*At1g75880*, *EXL1*; *At1g75890*, *EXL2*; *At1g75900*, *EXL3*; *At1g759*, *EXL4*; *At1g75920*, *EXL5*; *At1g75930*, *EXL6*)と CID28 に含まれる “glycine-rich protein/oleosins” の遺伝子群 (*At5g07510*, *GRP14*; *At5g07520*, *GRP18*; *At5g07530*, *GRP17*; *At5g07540*, *GRP16*; *At5g07550*, *GRP19*; *At5g07560*, *GRP20*)は植物の生殖において発現する。これら、リパーゼやオレオシンは花粉

を包むタンパク質の 90%以上を構成しており、柱頭における花粉の水和反応の開始に関与する(Mayfield et al., 2001)。1 番染色体にある CID9 と 5 番染色体にある CID28 は互いに独立しているが、受粉時において両者は機能的に共発現するだろう。これらの遺伝子群は、植物ゲノムが、機能遺伝子をクラスター化して持つことにより、利点をもたらすことを暗示する。異なる環境条件に適応する必要性があれば、代謝系の遺伝子群の形成は激しく働かせることになる。この所見はゲノムの柔軟性に言及する注目すべき点である。

この研究において、GO アノテーションの解析により、“defense response and immune response” のアノテーションがついた遺伝子群は CID=8 および CID=24 である。CID8 の 11 遺伝子中の 9 遺伝子(At1g72850, At1g72860, At1g72890, At1g72900, At1g72920, At1g72930, At1g72940, At1g72950)は “defense response” であるが、これはすでに共発現する遺伝子として報告されている (Meyers et al., 2002)。CID24 に含まれる “immune response” (免疫応答系) の 4 遺伝子(*RPP4*, At4g16860; *Copia4*, At4g16870; At4g16880; *SNCI*, At4g16890)は先行研究によると、これも強い共発現をする遺伝子群として報告された 226 遺伝子群に含まれるものの 1 遺伝子群である。(Zhan et al., 2006; Yi and Richards, 2007)。

この研究では FDR を 0.1 と厳しくとることにより偽陽性を小さく押さえることにした。しかしながら、さらに多くのオペロン様な遺伝子群をさがすならば FDR をゆるめに設定することで、たとえば、FDR を 0.25 とすることでクラスターの数は 57 に増える。この 57 に増えた中には、マルネラール代謝経路として関連付けられる遺伝子群として報告されたものが含まれてもいる(Field et al., 2011) (付表 表 3)。なお、さらに条件を厳しく、つまり 3 章における  $\alpha = 0.001\%$  において推定される FDR=0.02 では、13 個の遺伝子群 (CID = 2,3,9,11,22,23,26,27,28,29,30,32,33 ; 表 4 CID に下線) が対応する。このような自

由な閾値設定は恣意的なデータ選択が可能ではないかという問題点も考えられるが、取得できる予測される遺伝子群には FDR による信頼性を与えられていることから、データに含まれる誤りを考慮したデータ解釈が可能という点で有利であると考えられる。

オペロンは主に、原核生物で見られるものとされてきた。しかし、線虫 (*Caenorhabditis elegans*) においてはオペロンを定義した例があることを 1 章で紹介したとおりである。この線虫はおよそ 900 個に及ぶオペロンが決定されている (Blumenthal and Gleason, 2003)。線虫のオペロンはユニークな、トランススプライシングの過程を得て一本の RNA になるが、それは、連続して複数の遺伝子をコードしている pre-mRNA から一つの遺伝子をコードする mRNA が形成される現象である (Spieth et al., 1993)。線虫のオペロンの機能について、報告されている情報を参照すると、機能クラスごとにどの程度の遺伝子がオペロンに含まれているかが確認されている。多くのオペロンに含まれる遺伝子が、転写、スプライシング、翻訳といった発現制御のステップに関与しているとされている。また、組織特異的に、もしくは細胞特異的に働く遺伝子、コラーゲン、雌性ホルモンにより肝臓で作られるビテログニンへ関連する遺伝子、主だった精子関連のタンパク質生産に関わる遺伝子といったものは、ほとんどオペロンに含まれないとされている。そして、線虫のオペロンは餓死やストレスに対応するための、素早い遺伝子発現のレベルの変化といった必要性に応じたシステムであると考えられている。今回の研究において、シロイヌナズナについては 100 個のオペロン様な遺伝子クラスターが含まれると予測したが、この数字は線虫のものに比べ非常に少ない。

表 5 シロイヌナズナの遺伝子群と線虫のオペロンの差異

	シロイヌナズナ	線虫(Blumenthal et al., 2003)
発現調節系の遺伝子群のオペロンはあるか？	機能的に多くは見つからなかった (本研究)	ほとんどがこのタイプに属する遺伝子群とされる
・特異的な機能 ・特異的な組織での働き	花粉 リパーゼの生成	見つかっていない
	細胞の防御など特定の機能	
	根(Thaianol 合成系遺伝子群)	

これは、線虫のオペロンで特徴的であった、発現制御系の遺伝子を共発現遺伝子群として保存している数が少ないからではないかと考える。移動しない植物の場合と線虫のような動物での違いが、オペロン様遺伝子群の構成の違いとなったと考えられる。植物の場合、二次代謝系において多用される外敵からの生体防御に適応するような遺伝子群が共発現遺伝子群として働くようにできているのではないかと考えた。両者の差異を表 5 に示す。しかしながら、今回の研究により推定された遺伝子群を制御する機構については詳しい働きはわからない。環境の変動などにおいては、本研究では取り上げなかったエピジェネティックな制御をしているという可能性も考えられる。遺伝子間距離が 400bp 以内の場合、遺伝子がコードされている向きが *divergent* な二つの遺伝子における共発現はクロマチン構造の影響によるものという報告がある(Chen et al., 2010)。この場合、400bp 以上では、*convergent* な 2 遺伝子との間に差は見られないとされる。今回のオペロン様遺伝子群の推定においては、隣接する遺伝子という条件に着目し、遺伝子間の塩基距離については考慮していないが、予測された遺伝子群においては *divergent* な関係にある 2 遺伝子も含まれている。このな



かで、前述の 400bp 以内に近接している 2 遺伝子を含み、かつ相関係数が高いものとして  $r > 0.5$  となる隣接遺伝子が含まれているかを確認した結果、CID=8,12,13,31 において近接する divergent な 2 遺伝子を見ることができた。これらの遺伝子群の共発現はクロマチン構造の変化による遺伝子発現の影響を持つ可能性が考えられる。

多くの二次代謝物の遺伝子クラスターは真菌のゲノムからも見つかったが、その多くの生産物は植物や動物にとって毒として働くものである (Yu and Keller, 2005)。これは、シロイヌナズナにおいて予測した、オペロン様遺伝子群についても、表 4 の “metabolism” や “cell rescue, defense and virulence” といった項目を見る限り、同様に対応するものではないかと考えられる。また植物のオペロン様遺伝子群の遺伝子が、害虫や病気といったものに対抗するものが集中している事実からもそのように考えられる。

本研究においては、予測した遺伝子群内において遺伝子に相動性が見られる、遺伝子の重複箇所を示した。このような状況は、植物が生合成を行う上で有利に働くと考えられる。たとえば、P450 として機能推定された遺伝子群 CYP タンパク質は水酸化酵素ファミリーであるが、多くの遺伝子について明確な働きが分かってはいない。しかし判明している生合成経路を見てみると、植物ホルモンであるジベレリンを生合成する経路においては、複数の CYP タンパク質が関わっている (Mizutani and Ohta, 2010)。この例では、同じファミリーに属する複数の遺伝子が、生合成経路のどこで働くかは厳密に定められており、CYP 遺伝子を使い分けることで複数の生成物生産が可能としている。また、それぞれの経路に置いて複数の CYP タンパク質が働いている。本研究でのオペロン様遺伝子群においてこのような条件を考えると、生合成経路において重複した同様の機能を持ち、かつその機能する経路が順にきまっている遺伝子産物ならば、共発

現することにより、効率的な物質生産に寄与することが考えられる。

以上のように本研究では生物実験で得られた知見をベースにして、オペロン様な遺伝子に注目したこと、それを生物の大量データ解析により共発現領域を予測し、解析を行うことにより得られたが、この結果は続いて生物実験へのフィードバックを期待される。共発現に基づいてゲノムのデータマイニングを行うという観点は、質量解析のデータなどと合わせて、未知の代謝経路の解明につなげる戦略を期待される分野である(Castillo et al., 2013)。また、共発現とそれに共調して働く代謝経路の関係は、いまだ機能未知である遺伝子についての研究を深めるための重要なツールとして期待されている(Tohge and Fernie, 2012)。

## 総括

本研究は、シロイヌナズナにおいて、共発現のデータをもとに FDR を用いて 100 個のオペロン様な遺伝子クラスターが存在することを予測し、そのうちの 34 個のクラスターはそれぞれが含む 3~22 遺伝子を含み、FDR が 0.1 以下という条件で統計的に有意だと決定した遺伝子群を予測した。27 クラスターにおいては BLAST により  $E < 10^{-5}$  という条件により、遺伝子重複と考えられるものが含まれていた。5 つのクラスターは P450 の遺伝子を含み、それらはブラシカ属に限定される遺伝子群であり、また二次代謝系に関わる遺伝子だと予測されるものであった。さらに、推測された遺伝子群は、リボソームタンパクや、ユビキチンプロテアソーム系、脂質代謝系、脂質運搬、炭水化物代謝、またゲノム上では別の染色体に乗っている遺伝子の共発現などを見ることができた。

推定された遺伝子群の数は、0.1 という厳しい FDR によるデータのカットの影響もあるが、バクテリアに見られるオペロンに比べても、とても少ないものであった。にもかかわらず、今回、シロイヌナズナにおいて予測したオペロン様遺伝子群に含まれる遺伝子は、類似の機能を持つものが含まれるものが選別されており、今後の機能的な、生合成経路の解明に寄与することができると考える。

## 謝辞

本研究を遂行するに当たり、奈良先端科学技術大学院大学情報科学研究科の金谷 重彦 教授から非常に長期に渡りご指導いただきました。本学への進路を決めるに当たり、入学前からこの分野についての説明をしていただき多くの助言をいただきました。研究室に配属後は、研究室をあげての学生のサポートをしていただき、時には厳しい言葉をいただきながら、時には深夜まで自ら発表指導に当たっていただきました。博士後期課程においては、非常に長きに渡り、研究内容についての取り組みを指導していただきました。面白い内容があれば、やや研究フィールドから離れているようなものでも導入し、その中から活かせるものを探す姿勢など、非常に研究に真摯な姿勢をことあるたびに教えて頂きました。情報生命学という比較的新しい分野であった研究フィールドの楽しみ方のようなものを様々な視点を通して学ぶ機会を与えていただきました。解析においては多くの助言をいただき、さらに教授の縁から多くの研究者を通してこの研究に助言を受けることもできました。ユニークな人材が集う研究室運営から、面白い人々に会える機会も頂きました。ここに厚く御礼申し上げます。

今回の博士論文の審査について、副指導教員を担当していただきました、奈良先端科学技術大学院大学情報科学研究科 安本 慶一 教授に、この場を借りて御礼申し上げます。異分野からの審査を担当していただくことは大変なことにもかかわらず、しかしながら異なった視点での審査を担当頂けることは有り難いことであると感謝します。

奈良先端科学技術大学院大学情報科学研究科の Md.Altaf-Ul-Amin 准教授からは、様々な場面で研究のご指導をいただきました。情報解析ツールの提供や手法の解説といったことにとどまらず、手法やアルゴリズムの確認において多くの時間を割いていただきました。また、発表の資料のチェック、論文のチェ

ックなども含めてのご助力をいただきました。論文執筆にあたりましては英語指導など、非常に細かなことまでしていただきました。時に、ご家族そろって研究室に楽しい時間を提供していただくこともありました。御礼申し上げます。

奈良先端科学技術大学院大学情報科学研究科の 杉浦 忠男 准教授には、本学の博士前期課程のころから、別の研究室でありながらも学内で情報生命科学という分野ということもありまして、たびたびご指導いただく機会がありました。同研究室となってからも、つたない発表を見守っていただき、また声をかけていただきました。ここに御礼申し上げます。

奈良先端科学技術大学院大学情報科学研究科の小野 直亮 助教からはこの 2 年間、様々な分野に渡り、面白い話題を提供していただきました。また研究の話題においても、広範な視点を持ってご指導していただき、ソフトウェアの指導などもしていただきました。また研究発表の内容についても改善点を示していただきました。ここに御礼申し上げます。

奈良先端科学技術大学院大学情報科学研究科の 佐藤 哲大助教も他の研究室でありながら、大学院での課程における発表の場ではその都度、発表を通しての鋭い質問などを頂きました。同研究室になってからも温かく見守っていただきましてありがとうございます。ここに御礼申し上げます。

千葉大学の 高橋 弘喜 准教授は当大学院での学生時代の先輩として研究に意見していただき、後に助教として投稿論文の指導や発表の指導していただき、また作成されたデータベースが当研究においても大きな助力となりました。御礼申し上げます。

前橋工科大学の 中村 建介准教授は当研究室に所属されていたとき、研究発表において助言を頂き、また投稿論文の内容についても細かく指導していただきました。ここに御礼申し上げます。

計算システムズ生物学研究室の多くの方々からも多大なご助力を頂きました。森田 晶 研究員には、当研究室で処々のきめ細やかな業務のほか、使用するデータベースの管理などの面で、たびたびご助力を頂きました。また研究内容をデータベースとして外部に公開するに当たり多くの要望を汲んだ公開状況を作ってくださいました。ここに御礼申し上げます。中村 由紀子 研究員からは、この研究に所属されている期間や外部で研究員をされていた期間も通して、植物研究における助言を頂く機会もあり、また研究を進めるにあたっての心構えをご助言いただくこともありました。ここに御礼申し上げます。学生の方々からも多大なご協力を頂きました。池田 俊さん、桂樹 哲雄さんには研究についての助言、また違うバックグラウンドから見たときの本研究に対する視点などを頂く機会もありました。また研究室内での計算機の扱いなどでも助力いただき、またときに様々な面白い話題を提供していただきました。Nelson Kibinge さんや Li Donghan さんといった海外からの留学生の方々も増えて、面白い話を聞く機会も頂きました。また名前を挙げられなかったほかの学生の皆様にも研究や研究室での生活を楽しませて頂きました。ここに御礼申し上げます。

長きに渡り、比較ゲノム学講座および計算システムズ生物学講座で過ごす中で、多種多様な人材が集まっていた研究環境であったため、多くのバックグラウンドの違う人々が集まっており、先輩や後輩といった関係なく様々なことを学ぶ機会があったことに感謝し、また、そのような土壌を育て下さった教官の皆様に改めて感謝します。

最後に、ここに名前を挙げた皆様には今一度、また名前をおのせできなかった皆様におきまして今ここで感謝させていただきます。多くの皆様から様々な場面でご助力いただき、また、幾多のご心配おかけしたことと思いますが、気長に見守って頂きました。心から御礼申し上げます。

# 業績

## 査読付学術論文

1. Masayoshi Wada, Hiroki Takahashi, Md. Altaf-Ul-Amin, Kensuke Nakamura, Masami Y. Hirai, Daisaku Ohta, Shigehiko Kanaya. Prediction of operon-like gene clusters in the *Arabidopsis thaliana* genome based on co-expression analysis of neighboring genes. *Gene*, 503(1), 56–64. 2012 第2章～第4章に対応
2. Atsushi Fukushima, Masayoshi Wada, Shigehiko Kanaya, and Masanori Arita. SVD-based Anatomy of Gene Expressions for Correlation Analysis in *Arabidopsis thaliana*. *DNA Research*. 15(6): 367–374 2008 第2章に対応
3. Hiroki Takahashi, Mai Kawazoe, Masayoshi Wada, Aki Hirai, Kensuke Nakamura, Md. Altaf-Ul-Amin, Yuji Sawada, Masami Yokota Hirai, and Shigehiko Kanaya. KNApSACk gene classification system for *Arabidopsis thaliana*: Comparative genomic analysis of unicellular to seed plants. *Plant biotechnology* 26(5), 509-516, 2009 December 第4章に対応

## 査読付国際会議発表

1. Masayoshi Wada, Md. Altaf-Ul-Amin, Aki Hirai, Yukiko Nakamura, Masami Yokota Hirai, Kazuki Saito, Atsushi Fukushima, Masanori Arita, Daisuke Shibata, Shigehiko Kanaya. Analysis on functional relations of co-expressed gene pairs of *Arabidopsis thaliana*. 5th International Conference on Plant Metabolomics. 2008 July. 第2章に対応
2. Masayoshi Wada, Kensaku Nishikata, Md. Altaf-Ul-Amin, Hiroko Asahi, Yoko Shinbo, Ken Kurokawa, Shigehiko Kanaya. Functional units determined by integrated analysis based on protein-protein interaction and co-expression data in *Arabidopsis thaliana*. The Sixth Asia Pacific Bioinformatics Conference, 2008 January. 第2章に対応

## 參考資料

- Al-Shahrour, F., Minguez, P., Marques-Bonet, T., Gazave, E., Navarro, A. and Dopazo, J., 2010. Selection upon Genome Architecture: Conservation of Functional Neighborhoods with Changing Genes. *PLoS Comput Biol* 6.
- Amoutzias, G. and Van de Peer, Y., 2008. Together we stand: Genes cluster to coordinate regulation. *Dev. Cell* 14, 640-642.
- Arabidopsis Genome, I., 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796-815.
- Bailey, B.A. and Larson, R.L., 1991. Maize microsomal benzoxazinone N-monooxygenase. *Plant Physiol.* 95, 792-796.
- Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57, 289-300.
- Bennett, M.D., Leitch, I.J., Price, H.J. and Johnston, J.S., 2003. Comparisons with *Caenorhabditis* (approximately 100 Mb) and *Drosophila* (approximately 175 Mb) using flow cytometry show genome size in *Arabidopsis* to be approximately 157 Mb and thus approximately 25% larger than the *Arabidopsis* genome initiative estimate of approximately 125 Mb. *Ann Bot* 91, 547-57.
- Blumenthal, T., 2004. Operons in eukaryotes. *Briefings in Functional Genomics & Proteomics* 3, 199-211.
- Blumenthal, T., Evans, D., Link, C.D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W.L., Duke, K., Kiraly, M. and Kim, S.K., 2002. A global analysis of *Caenorhabditis elegans* operons. *Nature* 417, 851-854.
- Blumenthal, T. and Gleason, K.S., 2003. *Caenorhabditis elegans* operons: form and function. *Nat Rev Genet* 4, 112-20.
- Bolstad, B.M., 2004. Low-level Analysis of High-density Oligonucleotide Array Data: Background, Normalization and Summarization, *Biostatistics*. UNIVERSITY OF CALIFORNIA.
- Bolstad, B.M., Irizarry, R.A., Astrand, M. and Speed, T.P., 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185-193.
- Castillo, D.A., Kolesnikova, M.D. and Matsuda, S.P.T., 2013. An Effective Strategy for Exploring Unknown Metabolic Pathways by Genome Mining. *Journal of the American Chemical Society* 135, 5885-5894.
- Celis, J.E., Kruhoffer, M., Gromova, I., Frederiksen, C., Ostergaard, M., Thykjaer, T.,



- Gromov, P., Yu, J.S., Palsdottir, H., Magnusson, N. and Orntoft, T.F., 2000. Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *Febs Letters* 480, 2-16.
- Chen, W.H., de Meaux, J. and Lercher, M.J., 2010. Co-expression of neighbouring genes in *Arabidopsis*: separating chromatin effects from direct interactions. *Bmc Genomics* 11, 8.
- Chuang, L.-Y., Chang, H.-W., Tsai, J.-H. and Yang, C.-H., 2012. Features for computational operon prediction in prokaryotes. *Briefings in Functional Genomics* 11, 291-299.
- Cremer, T. and Cremer, C., 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* 2, 292-301.
- Falcon, S. and Gentleman, R., 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics* 23, 257-258.
- Field, B., Fiston-Lavier, A.S., Kemen, A., Geisler, K., Quesneville, H. and Osbourn, A.E., 2011. Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proceedings of the National Academy of Sciences of the United States of America* 108, 16116-16121.
- Field, B. and Osbourn, A.E., 2008. Metabolic diversification--independent assembly of operon-like gene clusters in different plants. *Science* 320, 543-7.
- Flowers, J.M. and Purugganan, M.D., 2008. The evolution of plant genomes - scaling up from a population perspective. *Current Opinion in Genetics & Development* 18, 565-570.
- Frey, M., Chomet, P., Glawischnig, E., Stettner, C., Grun, S., Winklmaier, A., Eisenreich, W., Bacher, A., Meeley, R.B., Briggs, S.P., Simcox, K. and Gierl, A., 1997. Analysis of a chemical plant defense mechanism in grasses. *Science* 277, 696-699.
- Frey, M., Kliem, R., Saedler, H. and Gierl, A., 1995. Expression of a cytochrome P450 gene family in maize. *Mol. Gen. Genet.* 246, 100-109.
- Fukushima, A., Kusano, M., Redestig, H., Arita, M. and Saito, K., 2011. Metabolomic correlation-network modules in *Arabidopsis* based on a graph-clustering approach. *BMC Syst Biol* 5, 1.
- Fukushima, A., Wada, M., Kanaya, S. and Arita, M., 2008. SVD-based Anatomy of Gene Expressions for Correlation Analysis in *Arabidopsis thaliana*. *DNA Research* 15, 367-374.
- Gierl, A. and Frey, M., 2001. Evolution of benzoxazinone biosynthesis and indole production in maize. *Planta* 213, 493-498.
- Girard, L.R., Fiedler, T.J., Harris, T.W., Carvalho, F., Antoshechkin, I., Han, M., Sternberg, P.W., Stein, L.D. and Chalfie, M., 2007. *WormBook: the online review of*

- Caenorhabditis elegans biology. Nucleic Acids Research 35, D472-D475.
- Holm, S., 1979. A SIMPLE SEQUENTIALLY REJECTIVE MULTIPLE TEST PROCEDURE. Scand. J. Stat. 6, 65-70.
- Huala, E., Dickerman, A.W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., Hanley, D., Kiphart, D., Zhuang, M.Z., Huang, W., Mueller, L.A., Bhattacharyya, D., Bhaya, D., Sobral, B.W., Beavis, W., Meinke, D.W., Town, C.D., Somerville, C. and Rhee, S.Y., 2001. The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucl. Acids Res. 29, 102-105.
- Hurst, L.D., Pal, C. and Lercher, M.J., 2004. The evolutionary dynamics of eukaryotic gene order. Nat. Rev. Genet. 5, 299-310.
- Iida, K., Seki, M., Sakurai, T., Satou, M., Akiyama, K., Toyoda, T., Konagaya, A. and Shinozaki, K., 2005. RARTF: Database and tools for complete sets of Arabidopsis transcription factors. DNA Res 12, 247-256.
- Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B. and Speed, T.P., 2003. Summaries of affymetrix GeneChip probe level data. Nucl Acids Res. 31, e15.
- Jacob, F., Perrin, D., Sanchez, C. and Monod, J., 1960. Operon: a group of genes with the expression coordinated by an operator. C R Hebd Seances Acad Sci 250, 1727-9.
- Kobayashi, H., Akitomi, J., Fujii, N., Kobayashi, K., Altaf-Ul-Amin, M., Kurokawa, K., Ogasawara, N. and Kanaya, S., 2007. The entire organization of transcription units on the Bacillus subtilis genome. BMC Genomics 8, 11.
- Koonin, E.V., 2009. Evolution of genome architecture. Int. J. Biochem. Cell Biol. 41, 298-306.
- Krom, N. and Ramakrishna, W., 2008. Comparative analysis of divergent and convergent gene pairs and their expression patterns in rice, Arabidopsis, and Populus. Plant Physiol 147, 1763-1773.
- Malacinski, M.G., Freifelder, D. and 川喜田正夫, 1999. 分子生物学の基礎 第3版, 3 ed. 東京化学同人.
- Matsuno, M., Compagnon, V., Schoch, G.A., Schmitt, M., Debayle, D., Bassard, J.E., Pollet, B., Hehn, A., Heintz, D., Ullmann, P., Lapierre, C., Bernier, F., Ehlting, J. and Werck-Reichhart, D., 2009. Evolution of a Novel Phenolic Pathway for Pollen Development. Science 325, 1688-1692.
- Mayfield, J.A., Fiebig, A., Johnstone, S.E. and Preuss, D., 2001. Gene families from the Arabidopsis thaliana pollen coat proteome. Science 292, 2482-2485.
- Mewes, H.W., Frishman, D., Guldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Morgenstern, B., Munsterkotter, M., Rudd, S. and Weil, B., 2002. MIPS: a database for genomes and protein sequences. Nucl Acids Res 30, 31-34.

- Meyers, B.C., Morgante, M. and Michelmore, R.W., 2002. TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes. *Plant J* 32, 77-92.
- Mizutani, M. and Ohta, D., 2010. Diversification of P450 Genes During Land Plant Evolution. *Annu. Rev. Plant Biol.* 61, 291-315.
- Mylona, P., Owatworakit, A., Papadopoulou, K., Jenner, H., Qin, B., Findlay, K., Hill, L., Qi, X., Bakht, S., Melton, R. and Osbourn, A., 2008. *Sad3* and *Sad4* are required for saponin biosynthesis and root development in oat. *Plant Cell* 20, 201-212.
- Niemeyer, H.M., 1988. Hydroxamic acids (4-hydroxy-1,4-benzoxazin-3-ones), defence chemicals in the gramineae. *Phytochemistry* 27, 3349-3358.
- Osbourn, A.E., Clarke, B.R., Lunness, P., Scott, P.R. and Daniels, M.J., 1994. An oat species lacking avenacin is susceptible to infection by *Gaeumannomyces graminis var. tritici*. *Physiol. Mol. Plant Pathol.* 45, 457-467.
- Osbourn, A.E. and Field, B., 2009. Operons. *Cell. Mol. Life Sci.* 66, 3755-3775.
- Papadopoulou, K., Melton, R.E., Leggett, M., Daniels, M.J. and Osbourn, A.E., 1999. Compromised disease resistance in saponin-deficient plants. *Proc. Natl. Acad. Sci. U. S. A.* 96, 12923-12928.
- Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R. and Osbourn, A., 2004. A gene cluster for secondary metabolism in oat: Implications for the evolution of metabolic diversity in plants. *Proc. Natl. Acad. Sci. U. S. A.* 101, 8233-8238.
- Qi, X., Bakht, S., Qin, B., Leggett, M., Hemmings, A., Mellon, F., Eagles, J., Werck-Reichhart, D., Schaller, H., Lesot, A., Melton, R. and Osbourn, A., 2006. A different function for a member of an ancient and highly conserved cytochrome P450 family: From essential sterols to plant defense. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18848-18853.
- Ren, X.Y., Fiers, M., Stiekema, W.J. and Nap, J.P., 2005. Local coexpression domains of two to four genes in the genome of *Arabidopsis*. *Plant Physiol.* 138, 923-934.
- Ren, X.Y., Stiekema, W.J. and Nap, J.P., 2007. Local coexpression domains in the genome of rice show no microsynteny with *Arabidopsis* domains. *Plant Mol. Biol.* 65, 205-217.
- Schilling, C.H., Schuster, S., Palsson, B.O. and Heinrich, R., 1999. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol Prog* 15, 296-303.
- Semon, M. and Duret, L., 2006. Evolutionary origin and maintenance of coexpressed gene clusters in mammals. *Mol Biol Evol* 23, 1715-1723.
- Shimura, K., Okada, A., Okada, K., Jikumaru, Y., Ko, K.W., Toyomasu, T., Sassa, T., Hasegawa, M., Kodama, O., Shibuya, N., Koga, J., Nojiri, H. and Yamane, H., 2007. Identification of a biosynthetic gene cluster in rice for momilactones. *J. Biol. Chem.*

282, 34013-34018.

- Shiu, S.H. and Bleecker, A.B., 2001. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci. USA* 98, 10763-10768.
- Spieth, J., Brooke, G., Kuersten, S., Lea, K. and Blumenthal, T., 1993. Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. *Cell* 73, 521-532.
- Sproul, D., Gilbert, N. and Bickmore, W.A., 2005. The role of chromatin structure in regulating the expression of clustered genes. *Nat. Rev. Genet.* 6, 775-781.
- Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., Radenbaugh, A., Singh, S., Swing, V., Tissier, C., Zhang, P. and Huala, E., 2008. The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Research* 36, D1009-D1014.
- Takahashi, H., Kawazoe, M., Wada, M., Hirai, A., Nakamura, K., Altaf-Ul-Amin, M., Sawada, Y., Hirai, M.Y. and Kanaya, S., 2009. KNApSAcK gene classification system for *Arabidopsis thaliana*: Comparative genomic analysis of unicellular to seed plants. *Plant Biotechnol.* 26, 509-516.
- Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L.A., Rhee, S.Y. and Stitt, M., 2004. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant Journal* 37, 914-939.
- Tohge, T. and Fernie, A.R., 2012. Co-expression and co-responses: within and beyond transcription. *Frontiers in plant science* 3, 248.
- Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otiillar, R.P. and Myers, R.M., 2004. An abundance of bidirectional promoters in the human genome. *Genome Res.* 14, 62-66.
- Vision, T.J., Brown, D.G. and Tanksley, S.D., 2000. The origins of genomic duplications in *Arabidopsis*. *Science* 290, 2114-2117.
- Williams, E.J.B. and Bowles, D.J., 2004. Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. *Genome Res* 14, 1060-7.
- Wintz, H., Fox, T., Wu, Y.Y., Feng, V., Chen, W.Q., Chang, H.S., Zhu, T. and Vulpe, C., 2003. Expression profiles of *Arabidopsis thaliana* in mineral deficiencies reveal novel transporters involved in metal homeostasis. *J Biol Chem* 278, 47644-47653.
- Wong, S. and Wolfe, K.H., 2005. Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nat Genet* 37, 777-82.
- Yi, H. and Richards, E.J., 2007. A cluster of disease resistance genes in *Arabidopsis* is

coordinately regulated by transcriptional activation and RNA silencing. *Plant Cell* 19, 2929-2939.

Yu, J.H. and Keller, N., 2005. Regulation of secondary metabolism in filamentous fungi. *Annu. Rev. Phytopathol.* 43, 437-458.

Zhan, S., Horrocks, J. and Lukens, L.N., 2006. Islands of co-expressed neighbouring genes in *Arabidopsis thaliana* suggest higher-order chromosome domains. *Plant J* 45, 347-357.

# 付表

## 付表 表 1.

推定した遺伝子群に含まれる遺伝子機能の数。TAIR の GOslim 検索による。

Keyword Category	Functional Category	Gene Count												
		1	2	3	4	5	6	7	8	9	10	11	12	
GO Cellular Component	cell wall													
	chloroplast		1									1	1	
	cytosol													
	ER									1				
	extracellular									6				
	Golgi apparatus		1											
	mitochondria		1						1			1		
	nucleus		1						1			1		
	other cellular components				3	4			1			2		
	other cytoplasmic components								1	1		2	1	
	other intracellular components		1						2	1		4	1	
	other membranes								9	1		2		
	plasma membrane				1			2						
	plastid								3				1	
ribosome								3			2			
unknown cellular components		3	11	1	1	1		4		2	5	1	2	
GO Molecular Function	DNA or RNA binding							2					1	
	hydrolase activity									7	2		2	
	kinase activity		2			3		1						
	nucleic acid binding													
	nucleotide binding		2			1		1	7				1	
	other binding							1	1	1			1	
	other enzyme activity							2		1		1	2	
	other molecular functions												1	
	protein binding						1							
	receptor binding or activity								9					
	structural molecule activity											2		
	transcription factor activity								1	1				
	transferase activity		3				1		1		6		1	
	transporter activity												1	
unknown molecular functions		3	1	8	1	1	3	4		2	2	14	2	
GO Biological Process	cell organization and biogenesis													
	developmental processes													
	DNA or RNA metabolism												1	
	electron transport or energy pathways												1	
	other biological processes		1	7							6		1	
	other cellular processes		2			3			2	2	1		4	2
	other metabolic processes		3			3			4	1	7		5	3
	protein metabolism		2			3			1			2	2	
	response to abiotic or biotic stimulus		3											
	response to stress		3						1	9				
	signal transduction						1			9				
	transport										1	1		1
	unknown biological processes		12	2	1	1	3	7			2	2	14	2

Keyword Category	Functional Category	Gene Count											
		13	14	15	16	17	18	19	20	21	22	23	24
GO Cellular Component	cell wall	1											
	chloroplast	2	5			1							
	cytosol	4	1	1									
	ER												
	extracellular		1										
	Golgi apparatus												
	mitochondria					3							
	nucleus			2		2				1			1
	other cellular components	1	1	2		1	5	1		1	3	4	1
	other cytoplasmic components	4	4			1							1
	other intracellular components	4	4		3	2				1		1	
	other membranes			1									
	plasma membrane	2				1							
	plastid		4										
ribosome	4	1			1								
unknown cellular components			2	1	2	1			6		2	1	
GO Molecular Function	DNA or RNA binding					2						1	
	hydrolase activity					2				1			
	kinase activity		1							1			
	nucleic acid binding												
	nucleotide binding		1										1
	other binding	2	2		1	9					5	4	
	other enzyme activity	1		4	3	9							6
	other molecular functions					9							5
	protein binding				2	1						1	2
	receptor binding or activity												
	structural molecule activity	4				1							
	transcription factor activity		1										
	transferase activity		1			2				1			2
	transporter activity		1										
unknown molecular functions		1	2		6		3		5			1	
GO Biological Process	cell organization and biogenesis								1	1		1	
	developmental processes		1			1			1			1	
	DNA or RNA metabolism												
	electron transport or energy pathways					1							
	other biological processes		1	1		1					1		1
	other cellular processes	4	5		4	3		1		2		3	1
	other metabolic processes	4	4		4	8	9			3			9
	protein metabolism	4			4	1				1			
	response to abiotic or biotic stimulus			2		2					3		2
	response to stress			2		2					3	1	2
	signal transduction					1						1	1
	transport		1									5	
	unknown biological processes		1	1		6		3		1		2	1

Keyword Category	Functional Category	Gene Count												
		25	26	27	28	29	30	31	32	33	34			
GO Cellular Component	cell wall													
	chloroplast		2				1			1				
	cytosol				2		1	1						
	ER							1	2					
	extracellular				5									
	Golgi apparatus													
	mitochondria		1		1					1				
	nucleus							1						
	other cellular components	3	8	4			3	1						2
	other cytoplasmic components	1		2	1									1
	other intracellular components	1		2	1	1								1
	other membranes		3		7									1
	plasma membrane	2	5						2					2
	plastid												1	1
ribosome														
<b>unknown cellular components</b>	<b>2</b>	<b>2</b>		<b>1</b>	<b>2</b>	<b>3</b>	<b>3</b>	<b>3</b>	<b>3</b>	<b>8</b>				
GO Molecular Function	DNA or RNA binding	1		1			1	1						
	hydrolase activity	1					1						3	
	kinase activity	1	18		1		1	1						
	nucleic acid binding													
	nucleotide binding	2	6	1				2						
	other binding	1	1	4	7		3		2	9				
	other enzyme activity	1	1		3		1		3					
	other molecular functions			4	4									
	protein binding	1	3	1						1				
	receptor binding or activity													
	structural molecule activity													
	transcription factor activity	1					1	1						
	transferase activity	1	8		1		2	1	1					
	transporter activity	1	1										3	
<b>unknown molecular functions</b>	<b>3</b>	<b>1</b>			<b>3</b>	<b>1</b>	<b>4</b>	<b>1</b>	<b>1</b>	<b>3</b>				
GO Biological Process	cell organization and biogenesis						1		1				1	
	developmental processes	1	1						1	4				
	DNA or RNA metabolism													
	electron transport or energy pathways													
	other biological processes	1	5		7									
	other cellular processes	4	19	1	4		4	2	4				4	
	other metabolic processes	4	19	1	4		5	2	5				4	
	protein metabolism	2	15	1	3		2	1	1					
	response to abiotic or biotic stimulus		6		3		1							
	response to stress		7		3				1					
	signal transduction													
	transport	1	1	4	1		3				4		3	
	<b>unknown biological processes</b>	<b>5</b>	<b>1</b>				<b>3</b>	<b>2</b>	<b>4</b>	<b>1</b>	<b>1</b>	<b>3</b>		



## 付表 表 2.

GOSTATS を用いた、推定した遺伝子群についての Gene ontology による遺伝子  
遺伝子のアノテーションの機能分類の一覧。

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
1 (3)	BP	GO:0009607	1.60E-05	3	625	response to biotic stimulus
		GO:0006952	3.19E-05	3	786	defense response
		GO:0006950	0.00066447	3	2161	response to stress
		GO:0050896	0.00556698	3	4388	response to stimulus
2 (16)	MF	GO:0003950	0.00420575	1	8	NAD+ ADP-ribosyltransferase activity
		GO:0016763	0.02396102	1	46	transferase activity, transferring pentosyl groups
		GO:0004674	0.04821547	2	679	protein serine/threonine kinase activity
3 (9)	BP	GO:0009733	9.02E-13	7	286	response to auxin stimulus
		GO:0009725	2.47E-09	7	885	response to hormone stimulus
		GO:0009719	4.31E-09	7	959	response to endogenous stimulus
		GO:0010033	2.33E-08	7	1223	response to organic substance
		GO:0042221	7.85E-07	7	2037	response to chemical stimulus
		GO:0050896	0.00014092	7	4388	response to stimulus
	CC	GO:0005794	0.04120451	1	228	Golgi apparatus
5 (4)	BP	GO:0006468	0.00020355	3	927	protein phosphorylation
		GO:0016310	0.00023869	3	978	phosphorylation
		GO:0006796	0.00030575	3	1063	phosphate metabolic process
		GO:0006793	0.00030661	3	1064	phosphorus metabolic process
		GO:0006464	0.00127434	3	1722	protein modification process
		GO:0043412	0.00167871	3	1891	macromolecule modification
		GO:0044267	0.00997668	3	3485	cellular protein metabolic process
		GO:0019538	0.01353488	3	3875	protein metabolic process
		GO:0007167	0.02084306	1	130	enzyme linked receptor protein signaling pathway
		GO:0007169	0.02084306	1	130	transmembrane receptor protein tyrosine kinase signaling pathway
	GO:0007166	0.02402052	1	150	cell surface receptor linked signaling pathway	
	GO:0044260	0.04680844	3	6010	cellular macromolecule metabolic process	
	MF	GO:0016301	0.00060708	3	1336	kinase activity
GO:0016772		0.00092053	3	1538	transferase activity, transferring phosphorus-containing groups	
GO:0016740		0.00507937	3	2753	transferase activity	
6 (4)	CC	GO:0012505	0.00059678	4	3407	endomembrane system
	MF	GO:0003779	0.01081289	1	67	actin binding
		GO:0008092	0.01754626	1	109	cytoskeletal protein binding
7 (12)	CC	GO:0031225	0.00542363	2	247	anchored to membrane
		GO:0031224	0.04903727	2	793	intrinsic to membrane
	MF	GO:0004602	0.00364048	1	9	glutathione peroxidase activity
		GO:0016684	0.04095663	1	103	oxidoreductase activity, acting on peroxide as acceptor
		GO:0004601	0.04095663	1	103	peroxidase activity

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
8 (11)	BP	GO:0045087	6.42E-14	8	298	innate immune response
		GO:0006955	7.15E-14	8	302	immune response
		GO:0006915	8.42E-14	7	150	apoptosis
		GO:0002376	1.14E-13	8	320	immune system process
		GO:0012501	1.59E-12	7	227	programmed cell death
		GO:0006952	1.63E-12	9	786	defense response
		GO:0016265	5.13E-12	7	268	death
		GO:0008219	5.13E-12	7	268	cell death
		GO:0007165	2.42E-09	8	1115	signal transduction
		GO:0023052	3.24E-09	8	1157	signaling
		GO:0006950	1.36E-08	9	2161	response to stress
		GO:0051716	3.61E-08	8	1572	cellular response to stimulus
		GO:0050794	3.50E-07	9	3126	regulation of cellular process
		GO:0050789	9.82E-07	9	3518	regulation of biological process
		GO:0065007	2.57E-06	9	3929	biological regulation
		GO:0050896	6.68E-06	9	4388	response to stimulus
		GO:0009987	0.00195969	10	10997	cellular process
		GO:0006839	0.03110805	1	71	mitochondrial transport
	CC	GO:0031224	5.52E-12	9	793	intrinsic to membrane
		GO:0044425	3.24E-10	9	1250	membrane part
		GO:0016020	1.24E-06	10	4495	membrane
		GO:0016602	0.0065446	1	13	CCAAT-binding factor complex
		GO:0005623	0.00925107	11	14237	cell
		GO:0044464	0.00925107	11	14237	cell part
		GO:0005667	0.01603976	1	32	transcription factor complex
	MF	GO:0004888	8.25E-19	9	159	transmembrane receptor activity
		GO:0004872	5.15E-18	9	194	receptor activity
		GO:0004871	2.05E-15	9	374	signal transducer activity
		GO:0060089	2.05E-15	9	374	molecular transducer activity
		GO:0005524	5.68E-07	7	1426	ATP binding
		GO:0032559	5.85E-07	7	1432	adenyl ribonucleotide binding
		GO:0030554	5.85E-07	7	1432	adenyl nucleotide binding
		GO:0035639	1.52E-06	7	1648	purine ribonucleoside triphosphate binding
GO:0032553		1.55E-06	7	1654	ribonucleotide binding	
GO:0032555		1.55E-06	7	1654	purine ribonucleotide binding	
GO:0017076		1.58E-06	7	1658	purine nucleotide binding	
GO:0000166		9.24E-06	7	2156	nucleotide binding	
GO:0005488	0.01749678	8	9121	binding		

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
9 (10)	BP	GO:0019953	2.46E-13	6	83	sexual reproduction
		GO:0000003	1.34E-06	6	1096	reproduction
		GO:0042147	0.00282474	1	7	retrograde transport, endosome to Golgi
		GO:0016197	0.006446	1	16	endosome transport
	CC	GO:0005576	5.02E-09	6	437	extracellular region
		GO:0005788	0.00412352	1	10	endoplasmic reticulum lumen
		GO:0044432	0.02854877	1	70	endoplasmic reticulum part
	MF	GO:0016298	2.58E-12	6	141	lipase activity
		GO:0008415	1.03E-11	6	177	acyltransferase activity
		GO:0016747	6.67E-11	6	241	transferase activity, transferring acyl groups other than amino-acyl groups
		GO:0016746	1.27E-10	6	268	transferase activity, transferring acyl groups
		GO:0004091	5.39E-10	6	341	carboxylesterase activity
		GO:0016788	2.99E-07	6	984	hydrolase activity, acting on ester bonds
		GO:0016787	9.13E-06	7	2911	hydrolase activity
GO:0016740		0.00011912	6	2753	transferase activity	
GO:0003824	0.00791253	7	8170	catalytic activity		
10 (9)	BP	GO:0006508	0.00403643	2	654	proteolysis
		GO:0008234	0.00011313	2	108	cysteine-type peptidase activity
	MF	GO:0070011	0.00239104	2	500	peptidase activity, acting on L-amino acid peptides
		GO:0008233	0.00266253	2	528	peptidase activity
11 (22)	BP	GO:0006268	0.00507941	1	9	DNA unwinding involved in replication
		GO:0006270	0.00676721	1	12	DNA-dependent DNA replication initiation
		GO:0032392	0.00732922	1	13	DNA geometric change
		GO:0032508	0.00732922	1	13	DNA duplex unwinding
		GO:0022904	0.01906344	1	34	respiratory electron transport chain
		GO:0006261	0.029569	1	53	DNA-dependent DNA replication
		GO:0008283	0.03997001	1	72	cell proliferation
		GO:0022900	0.04594423	1	83	electron transport chain
	GO:0071103	0.04918833	1	89	DNA conformation change	
	CC	GO:0000790	0.01023406	1	16	nuclear chromatin
		GO:0044454	0.01847754	1	29	nuclear chromosome part
		GO:0000228	0.02288988	1	36	nuclear chromosome
		GO:0000785	0.03415147	1	54	chromatin
	MF	GO:0042623	0.01375334	2	320	ATPase activity, coupled
GO:0008094		0.01855815	1	33	DNA-dependent ATPase activity	
	GO:0016887	0.02398688	2	430	ATPase activity	

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
12 (15)	BP	GO:0009190	0.00040394	1	2	cyclic nucleotide biosynthetic process
		GO:0009187	0.00040394	1	2	cyclic nucleotide metabolic process
		GO:0046058	0.00040394	1	2	cAMP metabolic process
		GO:0006171	0.00040394	1	2	cAMP biosynthetic process
		GO:0009124	0.00322789	1	16	nucleoside monophosphate biosynthetic process
		GO:0009123	0.00342935	1	17	nucleoside monophosphate metabolic process
		GO:0006164	0.0194414	1	97	purine nucleotide biosynthetic process
		GO:0072522	0.02162687	1	108	purine-containing compound biosynthetic process
		GO:0009165	0.02499673	1	125	nucleotide biosynthetic process
		GO:0034404	0.02835731	1	142	nucleobase, nucleoside and nucleotide biosynthetic process
		GO:0034654	0.02835731	1	142	nucleobase, nucleoside, nucleotide and nucleic acid biosynthetic process
		GO:0006163	0.02875206	1	144	purine nucleotide metabolic process
		GO:0072521	0.03465796	1	174	purine-containing compound metabolic process
		GO:0009117	0.04482678	1	226	nucleotide metabolic process
	GO:0006753	0.04482678	1	226	nucleoside phosphate metabolic process	
MF	GO:0004016	0.00060753	1	3	adenylate cyclase activity	
	GO:0016849	0.00080998	1	4	phosphorus-oxygen lyase activity	
	GO:0009975	0.00303471	1	15	cyclase activity	
	GO:0046912	0.00343879	1	17	transferase activity, transferring acyl groups, acyl groups converted into alkyl on transfer	
	GO:0019825	0.04592383	1	231	oxygen binding	

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
13 (4)	BP	GO:0006412	7.53E-06	4	1298	translation
		GO:0044267	0.00039227	4	3485	cellular protein metabolic process
		GO:0010467	0.00040967	4	3523	gene expression
		GO:0042254	0.00043716	2	213	ribosome biogenesis
		GO:0022613	0.0004705	2	221	ribonucleoprotein complex biogenesis
		GO:0034645	0.00050116	4	3705	cellular macromolecule biosynthetic process
		GO:0009059	0.00050768	4	3717	macromolecule biosynthetic process
		GO:0019538	0.00059971	4	3875	protein metabolic process
		GO:0071843	0.0009166	2	309	cellular component biogenesis at cellular level
		GO:0044249	0.0014758	4	4853	cellular biosynthetic process
		GO:0009058	0.00169109	4	5021	biosynthetic process
		GO:0044085	0.00300546	2	563	cellular component biogenesis
		GO:0044260	0.00347207	4	6010	cellular macromolecule metabolic process
		GO:0043170	0.00485408	4	6535	macromolecule metabolic process
		GO:0044237	0.01157819	4	8121	cellular metabolic process
		GO:0071841	0.01174641	2	1130	cellular component organization or biogenesis at cellular level
		GO:0044238	0.01231379	4	8247	primary metabolic process
		GO:0071840	0.02122406	2	1536	cellular component organization or biogenesis
		GO:0008152	0.02807955	4	10134	metabolic process
		GO:0009987	0.0389389	4	10997	cellular process
		GO:0022626	3.86E-08	4	307	cytosolic ribosome
	GO:0044445	5.17E-08	4	330	cytosolic part	
	GO:0005840	2.07E-07	4	466	ribosome	
	GO:0030529	6.41E-07	4	618	ribonucleoprotein complex	
	GO:0005829	1.72E-06	4	790	cytosol	
	GO:0043228	3.95E-06	4	973	non-membrane-bounded organelle	
	GO:0043232	3.95E-06	4	973	intracellular non-membrane-bounded organelle	
	GO:0032991	3.95E-05	4	1729	macromolecular complex	
	GO:0022627	0.00013713	2	105	cytosolic small ribosomal subunit	
	GO:0022625	0.00018824	2	123	cytosolic large ribosomal subunit	
	GO:0044446	0.00023598	4	2702	intracellular organelle part	
	GO:0044422	0.00023668	4	2704	organelle part	
	GO:0015935	0.00023694	2	138	small ribosomal subunit	
	GO:0015934	0.0003851	2	176	large ribosomal subunit	
	GO:0044444	0.00521611	4	5857	cytoplasmic part	
	GO:0005737	0.00694754	4	6292	cytoplasm	
	GO:0043229	0.01448215	4	7560	intracellular organelle	
	GO:0043226	0.01451283	4	7564	organelle	
	GO:0044424	0.02233886	4	8425	intracellular part	
	GO:0005622	0.02562468	4	8719	intracellular	
	GO:0005886	0.03721225	2	1817	plasma membrane	
	MF	GO:0003735	6.59E-08	4	397	structural constituent of ribosome
	GO:0005198	2.13E-07	4	532	structural molecule activity	

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
14 (6)	BP	GO:0046365	0.0001612	2	82	monosaccharide catabolic process
		GO:0046164	0.00018994	2	89	alcohol catabolic process
		GO:0005998	0.00024239	1	1	xylulose catabolic process
		GO:0005997	0.00048472	1	2	xylulose metabolic process
		GO:0005996	0.00051026	2	146	monosaccharide metabolic process
		GO:0044275	0.00056743	2	154	cellular carbohydrate catabolic process
		GO:0016052	0.00062757	2	162	carbohydrate catabolic process
		GO:0019323	0.00121144	1	5	pentose catabolic process
		GO:0044282	0.00135732	2	239	small molecule catabolic process
		GO:0006066	0.00160322	2	260	alcohol metabolic process
		GO:0006098	0.00604496	1	25	pentose-phosphate shunt
		GO:0006740	0.00628612	1	26	NADPH regeneration
		GO:0044262	0.00629345	2	522	cellular carbohydrate metabolic process
		GO:0006739	0.00700932	1	29	NADP metabolic process
		GO:0019321	0.00821367	1	34	pentose metabolic process
		GO:0046496	0.0084544	1	35	nicotinamide nucleotide metabolic process
		GO:0072524	0.00941681	1	39	pyridine-containing compound metabolic process
		GO:0019362	0.00941681	1	39	pyridine nucleotide metabolic process
		GO:0006733	0.0113393	1	47	oxidoreduction coenzyme metabolic process
		GO:0007018	0.01157939	1	48	microtubule-based movement
		GO:0009056	0.01438482	2	801	catabolic process
		GO:0006855	0.01493559	1	62	drug transmembrane transport
		GO:0015893	0.01541427	1	64	drug transport
		GO:0042493	0.01565354	1	65	response to drug
		GO:0005975	0.01736428	2	884	carbohydrate metabolic process
		GO:0006007	0.01852097	1	77	glucose catabolic process
		GO:0019320	0.01875961	1	78	hexose catabolic process
		GO:0006006	0.01971368	1	82	glucose metabolic process
		GO:0007017	0.02637054	1	110	microtubule-based process
		GO:0019318	0.02684459	1	112	hexose metabolic process
		GO:0006732	0.03487397	1	146	coenzyme metabolic process
		GO:0044281	0.03896937	2	1359	small molecule metabolic process

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
14 (6)	CC	GO:0009507	0.00022702	5	2913	chloroplast
		GO:0009536	0.00026077	5	2997	plastid
		GO:0005622	0.00410023	6	8719	intracellular
		GO:0009941	0.00580837	2	441	chloroplast envelope
		GO:0009526	0.00643943	2	465	plastid envelope
		GO:0044444	0.00652593	5	5857	cytoplasmic part
		GO:0005737	0.00913795	5	6292	cytoplasm
		GO:0031967	0.01549117	2	733	organelle envelope
		GO:0031975	0.01549117	2	733	envelope
		GO:0043231	0.01747836	5	7233	intracellular membrane-bounded organelle
		GO:0043227	0.01751187	5	7236	membrane-bounded organelle
		GO:0010287	0.01830968	1	67	plastoglobule
		GO:0043229	0.02142739	5	7560	intracellular organelle
		GO:0043226	0.02147953	5	7564	organelle
		GO:0031977	0.02345082	1	86	thylakoid lumen
		GO:0044434	0.02604332	2	964	chloroplast part
		GO:0044435	0.02890258	2	1019	plastid part
		GO:0044424	0.03512144	5	8425	intracellular part
	MF	GO:0004856	0.00048606	1	2	xylokinaase activity
		GO:0004332	0.00170034	1	7	fructose-bisphosphate aldolase activity
		GO:0016832	0.00460959	1	19	aldehyde-lyase activity
		GO:0019200	0.00944265	1	39	carbohydrate kinase activity
		GO:0015238	0.01569638	1	65	drug transmembrane transporter activity
		GO:0003777	0.01593625	1	66	microtubule motor activity
		GO:0003774	0.02072341	1	86	motor activity
		GO:0016830	0.02549116	1	106	carbon-carbon lyase activity
		GO:0008270	0.03459259	2	1271	zinc ion binding
		GO:0015297	0.03496863	1	146	antiporter activity

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
15 (4)	BP	GO:0009409	0.00072178	2	274	response to cold
		GO:0009266	0.00155233	2	403	response to temperature stimulus
		GO:0010043	0.00628751	1	39	response to zinc ion
		GO:0009628	0.01778957	2	1401	response to abiotic stimulus
		GO:0006950	0.04056542	2	2161	response to stress
	MF	GO:0005507	0.00046882	2	220	copper ion binding
		GO:0046914	0.03192501	2	1898	transition metal ion binding
		GO:0046872	0.04416371	2	2255	metal ion binding
		GO:0043169	0.04915333	2	2388	cation binding
		GO:0043167	0.0491917	2	2389	ion binding
16 (4)	BP	GO:0006511	4.04E-06	3	250	ubiquitin-dependent protein catabolic process
		GO:0019941	4.09E-06	3	251	modification-dependent protein catabolic process
		GO:0043632	4.09E-06	3	251	modification-dependent macromolecule catabolic process
		GO:0051603	4.29E-06	3	255	proteolysis involved in cellular protein catabolic process
		GO:0044257	4.71E-06	3	263	cellular protein catabolic process
		GO:0030163	5.44E-06	3	276	protein catabolic process
		GO:0044265	6.12E-06	3	287	cellular macromolecule catabolic process
		GO:0009057	9.38E-06	3	331	macromolecule catabolic process
		GO:0044248	5.73E-05	3	606	cellular catabolic process
		GO:0006508	7.20E-05	3	654	proteolysis
		GO:0009056	0.00013177	3	801	catabolic process
		GO:0007346	0.00258309	1	16	regulation of mitotic cell cycle
		GO:0000278	0.00805554	1	50	mitotic cell cycle
		GO:0044267	0.00997668	3	3485	cellular protein metabolic process
		GO:0019538	0.01353488	3	3875	protein metabolic process
		GO:0051726	0.01877355	1	117	regulation of cell cycle
		GO:0009987	0.0389389	4	10997	cellular process
	GO:0007049	0.04198467	1	264	cell cycle	
	GO:0044260	0.04680844	3	6010	cellular macromolecule metabolic process	
	CC	GO:0019005	9.41E-09	3	30	SCF ubiquitin ligase complex
		GO:0031461	1.46E-06	3	157	cullin-RING ubiquitin ligase complex
		GO:0000151	3.92E-06	3	218	ubiquitin ligase complex
		GO:0043234	0.00049388	3	1100	protein complex
		GO:0032991	0.00187662	3	1729	macromolecular complex
	MF	GO:0004842	1.40E-08	4	270	ubiquitin-protein ligase activity
		GO:0019787	1.55E-08	4	277	small conjugating protein ligase activity
		GO:0016881	2.08E-08	4	298	acid-amino acid ligase activity
GO:0016879		3.62E-08	4	342	ligase activity, forming carbon-nitrogen bonds	
GO:0016874		1.15E-07	4	456	ligase activity	
GO:0003824		0.01199146	4	8170	catalytic activity	



Class	Ontology	GOMFID	Pvalue	Count	Size	Term	
17 (16)	BP	GO:0031930	0.00226048	1	4	mitochondria-nucleus signaling pathway	
		GO:0009409	0.01017539	2	274	response to cold	
		GO:0042752	0.01125503	1	20	regulation of circadian rhythm	
		GO:0009266	0.02114107	2	403	response to temperature stimulus	
		GO:0048511	0.03615016	1	65	rhythmic process	
			GO:0007623	0.03615016	1	65	circadian rhythm
		CC	GO:0005763	0.00412352	1	9	mitochondrial small ribosomal subunit
			GO:0000418	0.00549462	1	12	DNA-directed RNA polymerase IV complex
			GO:0005761	0.00732011	1	16	mitochondrial ribosome
			GO:0005665	0.00777601	1	17	DNA-directed RNA polymerase II, core complex
			GO:0005739	0.01103973	3	1074	mitochondrion
			GO:0000428	0.01232467	1	27	DNA-directed RNA polymerase complex
			GO:0055029	0.01232467	1	27	nuclear DNA-directed RNA polymerase complex
			GO:0030880	0.01232467	1	27	RNA polymerase complex
			GO:0000314	0.0127785	1	28	organellar small ribosomal subunit
			GO:0016591	0.01413888	1	31	DNA-directed RNA polymerase II, holoenzyme
			GO:0070013	0.01999656	2	488	intracellular organelle lumen
			GO:0043233	0.02007376	2	489	organelle lumen
			GO:0031974	0.02053967	2	495	membrane-enclosed lumen
			GO:0005759	0.02136574	1	47	mitochondrial matrix
		GO:0000313	0.02630651	1	58	organellar ribosome	
		GO:0032991	0.03920508	3	1729	macromolecular complex	
		MF	GO:0009916	4.47E-06	2	6	alternative oxidase activity
			GO:0016682	0.00011195	2	28	oxidoreductase activity, acting on diphenols and related substances as donors, oxygen as acceptor
			GO:0016679	0.00019667	2	37	oxidoreductase activity, acting on diphenols and related substances as donors
			GO:0016165	0.00452863	1	8	lipoxygenase activity
			GO:0016702	0.01576702	1	28	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen
			GO:0051213	0.01688436	1	30	dioxygenase activity
			GO:0016701	0.02078576	1	37	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen
	GO:0003899		0.03569977	1	64	DNA-directed RNA polymerase activity	
	GO:0034062		0.03898522	1	70	RNA polymerase activity	
	GO:0016491	0.04533059	3	1451	oxidoreductase activity		

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
18 (9)	BP	GO:0055114	7.54E-14	9	866	oxidation-reduction process
		GO:0008152	0.00032234	9	10134	metabolic process
	CC	GO:0012505	0.00048652	5	3407	endomembrane system
	MF	GO:0019825	4.70E-19	9	231	oxygen binding
		GO:0020037	3.53E-18	9	288	heme binding
		GO:0004497	5.62E-18	9	303	monoxygenase activity
		GO:0046906	9.51E-18	9	321	tetrapyrrole binding
		GO:0005506	2.20E-17	9	352	iron ion binding
		GO:0009055	4.31E-17	9	379	electron carrier activity
		GO:0016491	8.18E-12	9	1451	oxidoreductase activity
		GO:0046914	9.23E-11	9	1898	transition metal ion binding
		GO:0046872	4.36E-10	9	2255	metal ion binding
		GO:0043169	7.32E-10	9	2388	cation binding
		GO:0043167	7.34E-10	9	2389	ion binding
		GO:0003824	4.75E-05	9	8170	catalytic activity
GO:0005488		0.00012801	9	9121	binding	
19 (5)	BP	GO:0010584	0.00306686	1	19	pollen exine formation
		GO:0010208	0.00371162	1	23	pollen wall assembly
		GO:0010927	0.00371162	1	23	cellular component assembly involved in morphogenesis
		GO:0045229	0.01717939	1	107	external encapsulating structure organization
		GO:0048646	0.01813612	1	113	anatomical structure formation involved in morphogenesis
		GO:0009555	0.02893031	1	181	pollen development
		GO:0048229	0.04073218	1	256	gametophyte development
		GO:0071844	0.04151513	1	261	cellular component assembly at cellular level
		GO:0022607	0.0438611	1	276	cellular component assembly
GO:0032989	0.04526661	1	285	cellular component morphogenesis		
21 (9)	BP	GO:0051225	0.00048471	1	4	spindle assembly
		GO:0007051	0.00072701	1	6	spindle organization
		GO:0070925	0.00084814	1	7	organelle assembly
		GO:0000226	0.00556473	1	46	microtubule cytoskeleton organization
		GO:0000279	0.01026645	1	85	M phase
		GO:0022403	0.01219106	1	101	cell cycle phase
		GO:0007017	0.01327256	1	110	microtubule-based process
		GO:0007010	0.01543319	1	128	cytoskeleton organization
		GO:0022402	0.01890756	1	157	cell cycle process
		GO:0071844	0.03130016	1	261	cellular component assembly at cellular level
		GO:0007049	0.03165608	1	264	cell cycle
	GO:0022607	0.0330789	1	276	cellular component assembly	
	CC	GO:0005876	0.00183453	1	5	spindle microtubule
		GO:0005874	0.01241682	1	34	microtubule
		GO:0005819	0.01241682	1	34	spindle
		GO:0015630	0.02721044	1	75	microtubule cytoskeleton
		GO:0044430	0.04039397	1	112	cytoskeletal part
	MF	GO:0003777	0.02119254	1	66	microtubule motor activity
GO:0003774		0.02753644	1	86	motor activity	

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
22 (5)	BP	GO:0006869	6.90E-12	5	147	lipid transport
		GO:0010876	1.24E-11	5	165	lipid localization
		GO:0071702	1.43E-10	5	268	organic substance transport
		GO:0033036	8.03E-09	5	597	macromolecule localization
		GO:0050832	1.37E-06	3	129	defense response to fungus
		GO:0006810	1.84E-06	5	1766	transport
		GO:0051234	1.88E-06	5	1774	establishment of localization
		GO:0051179	2.28E-06	5	1843	localization
		GO:0009620	3.09E-06	3	169	response to fungus
		GO:0051707	0.00012418	3	581	response to other organism
		GO:0009607	0.00015422	3	625	response to biotic stimulus
		GO:0051704	0.0002916	3	775	multi-organism process
		GO:0006952	0.00030401	3	786	defense response
		GO:0006950	0.00580601	3	2161	response to stress
		GO:0009627	0.00785336	1	39	systemic acquired resistance
		GO:0009814	0.02361027	1	118	defense response, incompatible interaction
		GO:0050896	0.04192389	3	4388	response to stimulus
	CC	GO:0012505	0.00381964	3	3407	endomembrane system
	MF	GO:0008289	6.58E-11	5	229	lipid binding
		GO:0005488	0.00688111	5	9121	binding

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
23 (10)	BP	GO:0016104	7.74E-06	2	12	triterpenoid biosynthetic process
		GO:0006722	1.07E-05	2	14	triterpenoid metabolic process
		GO:0055114	0.00016283	4	866	oxidation-reduction process
		GO:0010685	0.00036358	1	1	tetracyclic triterpenoid metabolic process
		GO:0010686	0.00036358	1	1	tetracyclic triterpenoid biosynthetic process
		GO:0016114	0.0003659	2	80	terpenoid biosynthetic process
		GO:0006721	0.00060561	2	103	terpenoid metabolic process
		GO:0010263	0.00072704	1	2	tricyclic triterpenoid biosynthetic process
		GO:0008299	0.000807	2	119	isoprenoid biosynthetic process
		GO:0006720	0.00116172	2	143	isoprenoid metabolic process
		GO:0010683	0.00145361	1	4	tricyclic triterpenoid metabolic process
		GO:0019742	0.00363049	1	10	pentacyclic triterpenoid metabolic process
		GO:0019745	0.00363049	1	10	pentacyclic triterpenoid biosynthetic process
		GO:0008610	0.00783883	2	379	lipid biosynthetic process
		GO:0019748	0.00874014	2	401	secondary metabolic process
		GO:0044255	0.01185549	2	470	cellular lipid metabolic process
	GO:0006629	0.02851173	2	748	lipid metabolic process	
	CC	GO:0000159	0.00480944	1	15	protein phosphatase type 2A complex
		GO:0008287	0.01309886	1	41	protein serine/threonine phosphatase complex
	MF	GO:0012505	0.01403055	4	3407	endomembrane system
		GO:0019825	1.50E-06	4	231	oxygen binding
		GO:0020037	3.61E-06	4	288	heme binding
		GO:0004497	4.41E-06	4	303	monooxygenase activity
		GO:0046906	5.54E-06	4	321	tetrapyrrole binding
		GO:0005506	7.98E-06	4	352	iron ion binding
		GO:0009055	1.07E-05	4	379	electron carrier activity
		GO:0031559	1.15E-05	2	13	oxidosqualene cyclase activity
		GO:0016866	0.0002412	2	58	intramolecular transferase activity
		GO:0080011	0.00040509	1	1	baruol synthase activity
		GO:0034075	0.00040509	1	1	arabidiol synthase activity
		GO:0016491	0.00187447	4	1451	oxidoreductase activity
		GO:0003824	0.00322806	8	8170	catalytic activity
		GO:0016853	0.00357759	2	226	isomerase activity
		GO:0016747	0.0040562	2	241	transferase activity, transferring acyl groups other than amino-acyl groups
		GO:0016746	0.00498886	2	268	transferase activity, transferring acyl groups
		GO:0046914	0.00501667	4	1898	transition metal ion binding
		GO:0008601	0.00646371	1	16	protein phosphatase type 2A regulator activity
		GO:0019888	0.00726902	1	18	protein phosphatase regulator activity
		GO:0019208	0.00767146	1	19	phosphatase regulator activity
		GO:0008187	0.00767146	1	19	poly-pyrimidine tract binding
		GO:0008266	0.00767146	1	19	poly(U) RNA binding
		GO:0046872	0.00929487	4	2255	metal ion binding
		GO:0043169	0.01137476	4	2388	cation binding
GO:0043167		0.01139148	4	2389	ion binding	
GO:0003727		0.01208858	1	30	single-stranded RNA binding	

Class	Ontology	GOMFID	Pvalue	Count	Size	Term	
24 (4)	BP	GO:0009814	6.74E-05	2	118	defense response, incompatible interaction	
		GO:0045087	0.00042988	2	298	innate immune response	
		GO:0006955	0.00044147	2	302	immune response	
		GO:0002376	0.00049551	2	320	immune system process	
		GO:0009862	0.00145366	1	12	systemic acquired resistance, salicylic acid mediated signaling pathway	
		GO:0051707	0.00162415	2	581	response to other organism	
		GO:0009607	0.00187743	2	625	response to biotic stimulus	
		GO:0051704	0.00287576	2	775	multi-organism process	
		GO:0006952	0.00295714	2	786	defense response	
		GO:0009816	0.00326876	1	27	defense response to bacterium, incompatible interaction	
		GO:0009863	0.00363152	1	30	salicylic acid mediated signaling pathway	
		GO:0071446	0.00363152	1	30	cellular response to salicylic acid stimulus	
		GO:0009627	0.00471926	1	39	systemic acquired resistance	
		GO:0009817	0.00496087	1	41	defense response to fungus, incompatible interaction	
		GO:0050832	0.01555313	1	129	defense response to fungus	
		GO:0009751	0.0173511	1	144	response to salicylic acid stimulus	
		GO:0009620	0.02034284	1	169	response to fungus	
		GO:0006950	0.02152475	2	2161	response to stress	
		GO:0042742	0.02499783	1	208	defense response to bacterium	
		GO:0009617	0.03201192	1	267	response to bacterium	
		GO:0009733	0.03426351	1	286	response to auxin stimulus	
		GO:0071310	0.04676098	1	392	cellular response to organic substance	
		CC	GO:0042598	0.00266006	1	29	vesicular fraction
			GO:0005792	0.00266006	1	29	microsome
	GO:0005624		0.00311833	1	34	membrane fraction	
	GO:0005626		0.00311833	1	34	insoluble fraction	
	GO:0000267		0.0035765	1	39	cell fraction	
	MF	GO:0030275	0.00024304	1	2	LRR domain binding	
		GO:0019904	0.00048605	1	4	protein domain specific binding	
		GO:0005515	0.00835267	2	1327	protein binding	

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
25 (13)	BP	GO:0042391	0.00040398	1	1	regulation of membrane potential
		GO:0009911	0.01405204	1	35	positive regulation of flower development
		GO:2000243	0.01405204	1	35	positive regulation of reproductive process
		GO:0048582	0.01803414	1	45	positive regulation of post-embryonic development
		GO:0051094	0.02081299	1	52	positive regulation of developmental process
		GO:0006873	0.03107256	1	78	cellular ion homeostasis
		GO:0055082	0.03342638	1	84	cellular chemical homeostasis
		GO:0009845	0.03538395	1	89	seed germination
		GO:0090351	0.03733796	1	94	seedling development
		GO:0050801	0.03928841	1	99	ion homeostasis
		GO:0009909	0.04395497	1	111	regulation of flower development
		GO:0006355	0.04700605	2	880	regulation of transcription, DNA-dependent
		GO:0051252	0.04797738	2	890	regulation of RNA metabolic process
	MF	GO:0005242	0.00327696	1	9	inward rectifier potassium channel activity
		GO:0005249	0.0079935	1	22	voltage-gated potassium channel activity
		GO:0022843	0.00835549	1	23	voltage-gated cation channel activity
		GO:0005267	0.00871736	1	24	potassium channel activity
		GO:0030570	0.00944074	1	26	pectate lyase activity
		GO:0016837	0.00944074	1	26	carbon-oxygen lyase activity, acting on polysaccharides
		GO:0015276	0.01052495	1	29	ligand-gated ion channel activity
		GO:0030551	0.01052495	1	29	cyclic nucleotide binding
		GO:0022834	0.01052495	1	29	ligand-gated channel activity
		GO:0005261	0.01196891	1	33	cation channel activity
		GO:0005244	0.01377123	1	38	voltage-gated ion channel activity
		GO:0022832	0.01377123	1	38	voltage-gated channel activity
		GO:0022836	0.02166676	1	60	gated channel activity
		GO:0004190	0.0287957	1	80	aspartic-type endopeptidase activity
		GO:0070001	0.0287957	1	80	aspartic-type peptidase activity
		GO:0005216	0.03057071	1	85	ion channel activity
GO:0022838	0.04361638	1	122	substrate-specific channel activity		
GO:0015267	0.04396679	1	123	channel activity		
GO:0022803	0.04396679	1	123	passive transmembrane transporter activity		
GO:0016835	0.04431708	1	124	carbon-oxygen lyase activity		

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
26 (22)	BP	GO:0006468	1.93E-12	11	927	protein phosphorylation
		GO:0016310	3.45E-12	11	978	phosphorylation
		GO:0006796	8.50E-12	11	1063	phosphate metabolic process
		GO:0006793	8.58E-12	11	1064	phosphorus metabolic process
		GO:0006464	1.50E-09	11	1722	protein modification process
		GO:0043412	4.04E-09	11	1891	macromolecule modification
		GO:0042742	2.28E-07	5	208	defense response to bacterium
		GO:0009617	7.83E-07	5	267	response to bacterium
		GO:0044267	2.31E-06	11	3485	cellular protein metabolic process
		GO:0019538	6.71E-06	11	3875	protein metabolic process
		GO:0006952	9.22E-06	6	786	defense response
		GO:0051707	3.43E-05	5	581	response to other organism
		GO:0009607	4.86E-05	5	625	response to biotic stimulus
		GO:0009751	0.00012349	3	144	response to salicylic acid stimulus
		GO:0051704	0.00013429	5	775	multi-organism process
		GO:0012501	0.00047059	3	227	programmed cell death
		GO:0044260	0.00047421	11	6010	cellular macromolecule metabolic process
		GO:0016265	0.0007626	3	268	death
		GO:0008219	0.0007626	3	268	cell death
		GO:0043170	0.00102647	11	6535	macromolecule metabolic process
		GO:0010033	0.00109732	5	1223	response to organic substance
		GO:0009987	0.00159658	14	10997	cellular process
		GO:0042221	0.00172255	6	2037	response to chemical stimulus
		GO:0006950	0.00233683	6	2161	response to stress
		GO:0044237	0.00698244	11	8121	cellular metabolic process
		GO:0044238	0.00795381	11	8247	primary metabolic process
		GO:0002237	0.00821185	1	12	response to molecule of bacterial origin
		GO:0046777	0.01500654	1	22	protein autophosphorylation
		GO:0045087	0.0174363	2	298	innate immune response
		GO:0006955	0.01787959	2	302	immune response
	GO:0009816	0.01838743	1	27	defense response to bacterium, incompatible interaction	
	GO:0002376	0.01993367	2	320	immune system process	
	GO:0050896	0.02031189	7	4388	response to stimulus	
	GO:0009626	0.02846469	1	42	plant-type hypersensitive response	
	GO:0034050	0.02913303	1	43	host programmed cell death induced by symbiont	
	GO:0008152	0.04146591	11	10134	metabolic process	
	GO:0005623	0.01509692	14	14237	cell	
	GO:0044464	0.01509692	14	14237	cell part	
	GO:0012505	0.02037709	6	3407	endomembrane system	
	GO:0005886	0.03121113	4	1817	plasma membrane	
	GO:0016301	2.66E-22	17	1336	kinase activity	
	GO:0016772	2.95E-21	17	1538	transferase activity, transferring phosphorus-containing groups	
GO:0016740	6.11E-17	17	2753	transferase activity		
GO:0003824	6.78E-09	17	8170	catalytic activity		
GO:0004672	4.66E-08	8	881	protein kinase activity		
GO:0016773	1.56E-07	8	1031	phosphotransferase activity, alcohol group as acceptor		
GO:0004674	4.05E-06	6	679	protein serine/threonine kinase activity		
GO:0005524	0.00026182	6	1426	ATP binding		
GO:0032559	0.00026788	6	1432	adenyl ribonucleotide binding		
GO:0030554	0.00026788	6	1432	adenyl nucleotide binding		
GO:0035639	0.00057181	6	1648	purine ribonucleoside triphosphate binding		
GO:0032553	0.00058303	6	1654	ribonucleotide binding		
GO:0032555	0.00058303	6	1654	purine ribonucleotide binding		
GO:0017076	0.00059061	6	1658	purine nucleotide binding		
GO:0000166	0.00234202	6	2156	nucleotide binding		
	CC					
	MF					

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
27 (6)	BP	GO:0006869	5.94E-09	4	147	lipid transport
		GO:0010876	9.47E-09	4	165	lipid localization
		GO:0071702	6.66E-08	4	268	organic substance transport
		GO:0033036	1.64E-06	4	597	macromolecule localization
		GO:0006810	0.00012176	4	1766	transport
		GO:0051234	0.00012395	4	1774	establishment of localization
		GO:0051179	0.00014407	4	1843	localization
	GO:0006413	0.01485917	1	74	translational initiation	
	CC	GO:0012505	0.00059678	4	3407	endomembrane system
	MF	GO:0045735	1.37E-10	4	58	nutrient reservoir activity
		GO:0008289	3.58E-08	4	229	lipid binding
		GO:0005488	0.00688111	5	9121	binding
		GO:0003743	0.01789834	1	89	translation initiation factor activity
GO:0008135		0.02684999	1	134	translation factor activity, nucleic acid binding	
28 (10)	BP	GO:0019915	1.22E-17	6	18	lipid storage
		GO:0019953	2.46E-13	6	83	sexual reproduction
		GO:0010876	1.64E-11	6	165	lipid localization
		GO:0033036	3.71E-08	6	597	macromolecule localization
		GO:0000003	1.34E-06	6	1096	reproduction
		GO:0051179	2.73E-05	6	1843	localization
		GO:0009859	0.00040398	1	1	pollen hydration
		GO:0006833	0.00403315	1	10	water transport
		GO:0042044	0.00403315	1	10	fluid transport
		GO:0006979	0.00532872	2	278	response to oxidative stress
		GO:0009875	0.01564662	1	39	pollen-pistil interaction
		GO:0042221	0.0430429	3	2037	response to chemical stimulus
		GO:0051726	0.0462806	1	117	regulation of cell cycle
	CC	GO:0005576	7.36E-07	5	437	extracellular region
		GO:0016020	0.00104925	7	4495	membrane
		GO:0012511	0.00595127	1	13	monolayer-surrounded lipid storage body
		GO:0005811	0.00640774	1	14	lipid particle
		GO:0031012	0.00777601	1	17	extracellular matrix
		GO:0044421	0.01368561	1	30	extracellular region part
		GO:0005829	0.04870253	2	790	cytosol
	MF	GO:0008289	1.22E-10	6	229	lipid binding
		GO:0045735	5.70E-09	4	58	nutrient reservoir activity
		GO:0008113	1.34E-05	2	14	peptide-methionine-(S)-S-oxide reductase activity
		GO:0016671	5.16E-05	2	27	oxidoreductase activity, acting on a sulfur group of donors, disulfide as acceptor
		GO:0004352	0.00040509	1	1	glutamate dehydrogenase activity
		GO:0016667	0.0007445	2	102	oxidoreductase activity, acting on a sulfur group of donors
		GO:0004353	0.00081003	1	2	glutamate dehydrogenase [NAD(P)+] activity
		GO:0016639	0.00161947	1	4	oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor
		GO:0004693	0.01288979	1	32	cyclin-dependent protein kinase activity
		GO:0016491	0.01780301	3	1451	oxidoreductase activity
		GO:0050897	0.01808342	1	45	cobalt ion binding
		GO:0016638	0.01927845	1	48	oxidoreductase activity, acting on the CH-NH2 group of donors
GO:0005488	0.03531717	7	9121	binding		



Class	Ontology	GOMFID	Pvalue	Count	Size	Term
30 (11)	BP	GO:0006869	2.39E-05	3	147	lipid transport
		GO:0010876	3.37E-05	3	165	lipid localization
		GO:0071702	0.00014235	3	268	organic substance transport
		GO:0033036	0.00147577	3	597	macromolecule localization
		GO:0009903	0.00242165	1	6	chloroplast avoidance movement
		GO:0051234	0.00388665	4	1774	establishment of localization
		GO:0009902	0.00403315	1	10	chloroplast relocation
		GO:0051667	0.00403315	1	10	establishment of plastid localization
		GO:0051644	0.00443566	1	11	plastid localization
		GO:0051179	0.0044651	4	1843	localization
		GO:0051656	0.00604423	1	15	establishment of organelle localization
		GO:0051640	0.0084527	1	21	organelle localization
		GO:0009637	0.02041644	1	51	response to blue light
		GO:0006810	0.02975673	3	1766	transport
		GO:0009658	0.02989372	1	75	chloroplast organization
	GO:0009657	0.04511842	1	114	plastid organization	
	MF	GO:0008289	6.35E-05	3	229	lipid binding
		GO:0004045	0.00291333	1	8	aminoacyl-tRNA hydrolase activity
		GO:0016207	0.00473032	1	13	4-coumarate-CoA ligase activity
		GO:0016405	0.00726917	1	20	CoA-ligase activity
GO:0016878		0.00763139	1	21	acid-thiol ligase activity	
		GO:0016877	0.00980226	1	27	ligase activity, forming carbon-sulfur bonds
31(8)	BP	GO:0006915	0.04165898	1	150	apoptosis

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
32 (6)	BP	GO:0080003	3.96E-12	3	3	thalianol metabolic process
		GO:0010683	1.58E-11	3	4	tricyclic triterpenoid metabolic process
		GO:0006722	1.44E-09	3	14	triterpenoid metabolic process
		GO:0006721	6.95E-07	3	103	terpenoid metabolic process
		GO:0006720	1.87E-06	3	143	isoprenoid metabolic process
		GO:0048364	1.26E-05	3	270	root development
		GO:0022622	1.26E-05	3	270	root system development
		GO:0019748	4.12E-05	3	401	secondary metabolic process
		GO:0044255	6.61E-05	3	470	cellular lipid metabolic process
		GO:0048856	8.29E-05	4	1602	anatomical structure development
		GO:0007275	0.00017171	4	1927	multicellular organismal development
		GO:0000740	0.00020199	1	1	nuclear membrane fusion
		GO:0010198	0.00020199	1	1	synergid death
		GO:0090174	0.00020199	1	1	organelle membrane fusion
		GO:0032501	0.0002047	4	2015	multicellular organismal process
		GO:0032502	0.00024397	4	2107	developmental process
		GO:0006629	0.00026258	3	748	lipid metabolic process
		GO:0048513	0.00033347	3	811	organ development
		GO:0048731	0.00033469	3	812	system development
		GO:0051084	0.00040394	1	2	'de novo' posttranslational protein folding
		GO:0051085	0.00040394	1	2	chaperone mediated protein folding requiring cofactor
		GO:0010263	0.00040394	1	2	tricyclic triterpenoid biosynthetic process
		GO:0006458	0.00040394	1	2	'de novo' protein folding
		GO:0061077	0.00040394	1	2	chaperone-mediated protein folding
		GO:0010623	0.00060586	1	3	developmental programmed cell death
		GO:0009558	0.00100961	1	5	embryo sac cellularization
		GO:0007349	0.00141323	1	7	cellularization
		GO:0019742	0.00201841	1	10	pentacyclic triterpenoid metabolic process
		GO:0019745	0.00201841	1	10	pentacyclic triterpenoid biosynthetic process
		GO:0016104	0.0024217	1	12	triterpenoid biosynthetic process
		GO:0010197	0.00423489	1	21	polar nucleus fusion
		GO:0000741	0.00443619	1	22	karyogamy
GO:0009559	0.00443619	1	22	embryo sac central cell differentiation		
GO:0048284	0.00443619	1	22	organelle fusion		

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
32 (6)	BP	GO:0006997	0.00564333	1	28	nucleus organization
		GO:0006944	0.00644743	1	32	cellular membrane fusion
		GO:0061025	0.00644743	1	32	membrane fusion
		GO:0009561	0.01025983	1	51	megagametogenesis
		GO:0016114	0.01605619	1	80	terpenoid biosynthetic process
		GO:0009553	0.01924253	1	96	embryo sac development
		GO:0061024	0.02103122	1	105	membrane organization
		GO:0016044	0.02103122	1	105	cellular membrane organization
		GO:0048646	0.02261897	1	113	anatomical structure formation involved in morphogenesis
		GO:0008299	0.02380843	1	119	isoprenoid biosynthetic process
		GO:0009856	0.0332825	1	167	pollination
		GO:0048610	0.03367565	1	169	cellular process involved in reproduction
		GO:0044237	0.04270225	4	8121	cellular metabolic process
		GO:0048468	0.04346291	1	219	cell development
		GO:0012501	0.04502149	1	227	programmed cell death
		GO:0044238	0.04516449	4	8247	primary metabolic process
		GO:0006457	0.04677246	1	236	protein folding
		CC	GO:0005783	0.00477826	2	399
	MF	GO:0051746	0.00024305	1	1	thalianol synthase activity
		GO:0080014	0.00024305	1	1	thalianol hydroxylase activity
		GO:0080004	0.00024305	1	1	thalian-diol desaturase activity
		GO:0019825	0.0012758	2	231	oxygen binding
		GO:0031559	0.00315585	1	13	oxidosqualene cyclase activity
		GO:0016866	0.01401593	1	58	intramolecular transferase activity
		GO:0016627	0.01689524	1	70	oxidoreductase activity, acting on the CH-CH group of donors
		GO:0051082	0.02549116	1	106	unfolded protein binding
		GO:0031072	0.03023954	1	126	heat shock protein binding
		GO:0008415	0.04226073	1	177	acyltransferase activity
	GO:0016491	0.04420107	2	1451	oxidoreductase activity	

Class	Ontology	GOMFID	Pvalue	Count	Size	Term
33 (10)	BP	GO:0030001	2.18E-08	4	203	metal ion transport
		GO:0006812	1.44E-07	4	325	cation transport
		GO:0006811	3.81E-07	4	414	ion transport
		GO:0006810	0.00012176	4	1766	transport
		GO:0051234	0.00012395	4	1774	establishment of localization
	MF	GO:0051179	0.00014407	4	1843	localization
		GO:0046872	4.01E-09	9	2255	metal ion binding
		GO:0043169	6.68E-09	9	2388	cation binding
		GO:0043167	6.71E-09	9	2389	ion binding
		GO:0005488	0.00085469	9	9121	binding
34 (7)	BP	GO:0009661	0.00016159	1	1	chromoplast organization
		GO:0016120	0.00113072	1	7	carotene biosynthetic process
		GO:0016119	0.00193779	1	12	carotene metabolic process
		GO:0016117	0.00355046	1	22	carotenoid biosynthetic process
		GO:0016109	0.00355046	1	22	tetraterpenoid biosynthetic process
		GO:0016116	0.00467817	1	29	carotenoid metabolic process
		GO:0016108	0.00467817	1	29	tetraterpenoid metabolic process
		GO:0046246	0.00483919	1	30	terpene biosynthetic process
		GO:0042214	0.00693071	1	43	terpene metabolic process
		GO:0016114	0.01286545	1	80	terpenoid biosynthetic process
		GO:0046148	0.01606232	1	100	pigment biosynthetic process
		GO:0006721	0.01654118	1	103	terpenoid metabolic process
		GO:0009657	0.0182955	1	114	plastid organization
		GO:0008299	0.01909215	1	119	isoprenoid biosynthetic process
		GO:0042440	0.01941067	1	121	pigment metabolic process
	GO:0006720	0.02290929	1	143	isoprenoid metabolic process	
	CC	GO:0005886	0.03721225	2	1817	plasma membrane
	MF	GO:0043492	5.57E-06	3	163	ATPase activity, coupled to movement of substances
		GO:0042626	5.57E-06	3	163	ATPase activity, coupled to transmembrane movement of substances
		GO:0016820	5.78E-06	3	165	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances
		GO:0015405	1.30E-05	3	216	P-P-bond-hydrolysis-driven transmembrane transporter activity
		GO:0015399	1.31E-05	3	217	primary active transmembrane transporter activity
		GO:0042623	4.19E-05	3	320	ATPase activity, coupled
GO:0016887		0.00010095	3	430	ATPase activity	
GO:0022804		0.00022536	3	564	active transmembrane transporter activity	
GO:0017111		0.00047792	3	728	nucleoside-triphosphatase activity	
GO:0016462		0.0005339	3	756	pyrophosphatase activity	
GO:0016818		0.00054014	3	759	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	
GO:0016817		0.00055276	3	765	hydrolase activity, acting on acid anhydrides	
GO:0022857	0.00106703	3	958	transmembrane transporter activity		
GO:0005215	0.0022867	3	1246	transporter activity		
GO:0016787	0.02487117	3	2911	hydrolase activity		

## 付表 表 3.

$\alpha=0.05\%$ において推定される遺伝子群の一覧 (FDR はおよそ 0.25 に対応する)

CID	Gene
1	At1g14930, At1g14940, At1g14950
2	At1g23530, At1g23540, At1g23550, At1g23560, At1g23570, At1g23580, At1g23590, At1g23600, At1g23610, At1g23630, At1g23640, At1g23650, At1g23670, At1g23680, At1g23690, At1g23700
3	At1g26380, At1g26390, At1g26400, At1g26410, At1g26420, At1g26440, At1g26450, At1g26460, At1g26470, At1g26480, At1g26520, At1g26530, At1g26540
4	At1g29430, At1g29440, At1g29450, At1g29460, At1g29470, At1g29480, At1g29490, At1g29500, At1g29510
5	At1g32060, At1g32070, At1g32080
6	At1g37160, At1g37170, At1g39190, At1g39270, At1g39350, At1g39430, At1g40150, At1g40230, At1g41795, At1g41797, At1g41810, At1g41820, At1g41825, At1g41860, At1g41870
7	At1g51830, At1g51840, At1g51850, At1g51860, At1g51880, At1g51890, At1g51900, At1g51910, At1g51915, At1g51920, At1g51940, At1g51950
8	At1g52050, At1g52060, At1g52070, At1g52080
9	At1g55820, At1g55830, At1g55840, At1g55850, At1g55870, At1g55880, At1g55890, At1g55900
10	At1g63440, At1g63450, At1g63460, At1g63470, At1g63480, At1g63490, At1g63500, At1g63540, At1g63550, At1g63560, At1g63580, At1g63590, At1g63600
11	At1g67860, At1g67865, At1g67870
12	At1g72810, At1g72820, At1g72830, At1g72840, At1g72850, At1g72860, At1g72890, At1g72900, At1g72920, At1g72930, At1g72940, At1g72950, At1g72960, At1g72970
13	At1g75850, At1g75860, At1g75870, At1g75880, At1g75890, At1g75900, At1g75910, At1g75920, At1g75930, At1g75940
14	At2g07190, At2g07200, At2g07230, At2g07240, At2g07280, At2g07290, At2g07300, At2g07320, At2g07330
15	At2g07620, At2g07630, At2g07672, At2g07673, At2g07691, At2g07692, At2g07693, At2g07701, At2g07702, At2g07705, At2g07706, At2g07713, At2g07714, At2g07715, At2g07718, At2g07719, At2g07721, At2g07722, At2g07724, At2g07725, At2g07680, At2g07690
16	At2g11120, At2g11130, At2g11140, At2g11150, At2g11210, At2g11220, At2g11230, At2g11240, At2g11270, At2g11360, At2g11370, At2g11680, At2g11690, At2g11890, At2g11910, At2g12170, At2g12190, At2g12210, At2g12230, At2g12240, At2g12250, At2g12290, At2g12300, At2g12910, At2g12920
17	At2g17060, At2g17070, At2g17080
18	At2g19720, At2g19730, At2g19740, At2g19750
19	At2g21310, At2g21320, At2g21330, At2g21340, At2g21370, At2g21380, At2g21385
20	At2g29530, At2g29540, At2g29550, At2g29560, At2g29570
21	At3g03300, At3g03305, At3g03310, At3g03320, At3g03330, At3g03340, At3g03360, At3g03370
22	At3g07820, At3g07830, At3g07840, At3g07850
23	At3g16430, At3g16440, At3g16450, At3g16460, At3g16470
24	At3g21780, At3g21790, At3g21800, At3g21830, At3g21840, At3g21850, At3g21860

CID	Gene
25	At3g22240, At3g22250, At3g22260, At3g22290, At3g22300, At3g22310, At3g22320, At3g22340, At3g22350, At3g22360, At3g22370, At3g22380, At3g22400, At3g22430, At3g22435, At3g22440, At3g22450
26	At3g26210, At3g26220, At3g26230, At3g26270, At3g26280, At3g26290, At3g26300, At3g26310, At3g26320
27	At3g28770, At3g28780, At3g28790, At3g28830, At3g28840, At3g28850
28	At3g30380, At3g30390, At3g30396, At3g30400, At3g30440, At3g30450, At3g30480, At3g30490, At3g30500
29	At3g31910, At3g31915, At3g31920, At3g31930, At3g32050, At3g32060, At3g32080, At3g32090, At3g32100, At3g32110, At3g32120, At3g32130, At3g32140, At3g32150, At3g32160, At3g32180, At3g32190, At3g32250, At3g32260, At3g32270, At3g32280, At3g32290, At3g32900, At3g32910
30	At3g42570, At3g42580, At3g42590, At3g42600, At3g42630, At3g42640, At3g42660, At3g42670
31	At4g04070, At4g04080, At4g04110, At4g04120, At4g04130, At4g04180, At4g04190, At4g04210, At4g04220, At4g04260, At4g04270, At4g04280, At4g04290, At4g04320, At4g04330, At4g04340, At4g04350, At4g04410, At4g04420
32	At4g05190, At4g05200, At4g05210, At4g05230, At4g05240, At4g05250, At4g05260, At4g05270, At4g05280
33	At4g12470, At4g12480, At4g12490, At4g12500, At4g12510, At4g12530, At4g12540
34	At4g15330, At4g15340, At4g15350, At4g15360, At4g15370, At4g15380, At4g15390, At4g15400, At4g15410, At4g15415
35	At4g15680, At4g15690, At4g15700
36	At4g16870, At4g16860, At4g16880, At4g16890
37	At4g19490, At4g19500, At4g19510, At4g19520, At4g19530, At4g19540, At4g19550, At4g19570, At4g19580, At4g19600, At4g19610
38	At4g22030, At4g22040, At4g22050, At4g22060, At4g22070, At4g22080, At4g22120, At4g22130, At4g22140, At4g22150, At4g22160, At4g22190, At4g22200
39	At4g23130, At4g23140, At4g23150, At4g23180, At4g23190, At4g23200, At4g23210, At4g23220, At4g23230, At4g23240, At4g23250, At4g23260, At4g23270, At4g23280, At4g23290, At4g23300, At4g23310, At4g23320, At4g23330, At4g23400, At4g23410, At4g23420
40	At4g26500, At4g26510, At4g26520, At4g26530
41	At4g27120, At4g27130, At4g27140, At4g27150, At4g27160, At4g27170
42	At4g33470, At4g33480, At4g33490, At4g33500, At4g33510, At4g33520, At4g33530
43	At4g38840, At4g38850, At4g38860, At4g38870, At4g38880
44	At5g01530, At5g01540, At5g01550, At5g01560, At5g01600, At5g01610, At5g01620, At5g01650, At5g01660, At5g01670, At5g01680, At5g01690, At5g01700
45	At5g06720, At5g06730, At5g06740, At5g06750
46	At5g07410, At5g07420, At5g07430
47	At5g07440, At5g07450, At5g07460, At5g07470, At5g07510, At5g07520, At5g07530, At5g07540, At5g07550, At5g07560
48	At5g28550, At5g28560, At5g28570, At5g28580, At5g28590
49	At5g38120, At5g38130, At5g38140, At5g38150, At5g38160, At5g38170, At5g38180, At5g38210, At5g38220, At5g38290, At5g38300

CID	Gene
50	At5g38895, At5g38900, At5g38950, At5g38960, At5g38970, At5g38980, At5g38990, At5g39020, At5g39030
51	At5g42580, At5g42590, At5g42600
52	At5g45410, At5g45420, At5g45430, At5g45480, At5g45490, At5g45500, At5g45510, At5g45530, At5g45540
53	At5g47980, At5g47990, At5g48000, At5g48010, At5g48020, At5g48030
54	At5g49240, At5g49250, At5g49260
55	At5g52670, At5g52680, At5g52690, At5g52710, At5g52720, At5g52740, At5g52750, At5g52760, At5g52770, At5g52780
56	At5g61660, At5g61670, At5g61700, At5g61710, At5g61720, At5g61730, At5g61740, At5g61750
57	At5g63930, At5g63940, At5g63950, At5g63960