特集記事 ワークショップ報告

育種のための情報解析ツール使い倒し塾

神沼英里¹⁾・七夕高也²⁾・矢野健太郎³⁾・清水顕史⁴⁾・岩田洋佳⁵⁾¹⁾ 国立遺伝学研究所生命情報研究センター, 静岡県三島市, 〒411-8540²⁾ 理化学研究所環境資源科学研究センター, 茨城県つくば市, 〒305-0074³⁾ 明治大学農学部, 神奈川県川崎市, 〒214-8571⁴⁾ 滋賀県立大学環境科学部, 滋賀県彦根市, 〒522-8533⁵⁾ 東京大学大学院農学生命科学研究科, 東京都文京区, 〒113-8657

Information analysis tools for promoting the efficiency of breeding

Eli Kaminuma¹⁾, Takanari Tanabata²⁾, Kentaro Yano³⁾, Akifumi Shimizu⁴⁾ and Hiroyoshi Iwata⁵⁾¹⁾ *Genome Informatics Laboratory, National Institute of Genetics, Shizuoka 411-8540, Japan*²⁾ *Gene Discovery Research Group, RIKEN Center for Sustainable Resource Science, Ibaraki 305-0074, Japan*³⁾ *School of Agriculture, Meiji University, Kanagawa, 214-8571, Japan*⁴⁾ *School of Environmental Science, University of Shiga Prefecture, Shiga 522-8533, Japan*⁵⁾ *Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan*

キーワード

高速シーケンサー配列解析, 高効率フェノタイピング, 大規模遺伝子発現情報, QTL 解析, ゲノミックセレクション, Web ツール・データベース

はじめに

最近, 「ビッグデータ (big data)」という言葉を目にする機会が多い。ビッグデータとは情報技術 (information technology: IT) 分野の用語であり, IT の発達にともない, 様々な分野において莫大な量が蓄積されるようになった巨大なデータ群のことをいう。ビッグデータは量が膨大なだけでなく, 多くの場合, 様々な条件下で収集された様々な形式のデータを含んでおり, その解析は難しい。これは, 育種分野での事例で例えると, 実験材料や方法が異なる大量のデータをまとめて解析するようなものであり, それがどれほど困難なことなのかは容易に想像ができる。しかし, ビッグデータには, ビジネスや社会に有用な知見が未発見のまま数多く眠っていると考えられており, このことが, 様々な分野で「ビッグデータ」というキーワードを重要なものとしている。

では, 「ビッグデータ」というキーワードを念頭において, 育種や育種学の世界に視点を移してみよう。発現解析アレイや DNA アレイの登場は, 数千~数万個の遺伝子に関する発現やその DNA 多型について迅速に計測することを可能にした。また, 高速シーケンサーの登場は, 高スループット・低コストでの発現解析や DNA 多型解

析を可能にし, その勢いは先述したアレイ技術を早くも時代遅れにするが如しである。このように, ここ数年の分子遺伝学実験技術の発展は, 育種分野で蓄積されるデータ量を「ビッグデータ」とよばれる域に到達させつつある。こうした爆発的なデータ量の増加は, これまでには不可能であった新しい育種研究や育種法を実現する一方で, データ解析という手続きを従来よりもずっと複雑で困難なものにしている。一回の実験から得られるデータ量も増大している今日では, データ解析の手法や技術を知らなければ, 収集したデータを俯瞰することさえままならず, たたまた膨大な量のデータに圧倒される結果になってしまう。さらに困難なことには, 解析手法や技術が発展する速度が非常に速く, 育種や育種学に携わる者がその最新動向を把握していくことは極めて難しい。

育種の現場や研究において, バイオインフォマティクス手法を導入することによる効果は大きい。導入の利点として, 目標達成までに要する時間や労力の効率化がもたらされること, 得られたデータや知見を計算機上に蓄積し, 将来の品種育成に資する情報リソースを整備することなどが挙げられる。このような視点から, 育種分野で必要となる種々の情報解析ツールについて, その機能の概略や具体的な使用法を解説することを目的として, 本ワークショップを企画した。ワークショップでは, 5 題の講演中に, 育種分野で必要となると考えられる情報解析ツールについて, できるだけ様々なものを取り込むようにした。講演 1 および 3 では, 今後ますます増加の

一途をたどると思われるゲノム配列データや遺伝子発現データを解析するためのツールやデータベースに関する話題提供を行った。また、講演2では、育種分野において今後大きなボトルネックになると言われている表現型データを収集・解析するためのツールについての話題提供を行った。さらに、講演4および5では、ゲノムデータと表現型データを結びつけるアプローチとして、QTL解析とゲノミックセレクションのための手法やツールについての話題提供を行った。冒頭に述べた「ビッグデータ」という視点からみると、その到来に備えるための第一歩を踏み出すというのが本ワークショップのねらいであったと言えるかもしれない。

各講演のねらいと要旨

1. 高速シーケンサー・アーカイブ配列を用いた系統間DNA多型データベース

育種品種の系統間DNA多型は、ゲノム育種研究を加速させるための重要な基盤リソースになる。高速シーケンサーを用いてゲノムワイドに多型マーカーを構築する研究が、近年続々と発表されている。国立遺伝学研究所大量遺伝情報研究室では、高速シーケンサーの配列アーカイブ・データベース DDBJ Sequence Read Archive (DRA) (<http://trace.ddbj.nig.ac.jp/>) (Kodama *et al.* 2012) に登録されている植物・動物・微生物のゲノム配列を用いて、系統間多型の抽出と多型注釈データベースの構築を進めている。特に、DNAマーカー構築・形質予測・遺伝資源評価といった育種研究者が実践的に使えるレベルのゲノムリソースを目指して、多型データベースを整備している。第122回講演会のワークショップでは、構築中の下記2件のデータベース DNAPod と H2DB を紹介した。

① DNAPod: DNA多型注釈データベース

DNA Polymorphism Annotation Database (DNAPod; <http://tga.nig.ac.jp/dnapod/>) は、DRA 公開配列からゲノムワイドなDNA多型情報を均一な条件で解析した結果をダウンロード出来るサイトである。現在は一塩基多型 (SNP) のみを対象としており、既知遺伝子の構造注釈、同義置換/非同義置換の注釈などの情報がダウンロード出来る。2013年2月現在、イネ678系統 (内621系統は1~3xの低深度) のデータを格納している。またユーザーが独自データをDNAPodと同一条件で解析できる様に、DNAPod解析ワークフローのサービスをDDBJ Pipeline (<http://p.ddbj.nig.ac.jp/>) から公開している。DDBJ Pipelineは、シーケンサーの出力配列を解析するクラウド型パイプラインで、*de novo* アセンブリや参照配列アライメントを基盤とした解析ワークフローを提供している (Kaminuma *et al.* 2010)。

② H2DB: 形質遺伝率と関連多型データベース

Heritability Database (H2DB; <http://tga.nig.ac.jp/h2db/>) は、文献から遺伝率に関する形質とDNA多型情報を抽出

したデータベースである。DNA多型 (現在のエントリーは形質関連SNPのみ) と遺伝率推定値のデータを、PubMed abstract やオンライン雑誌から手作業で抽出している。遺伝率は育種現場で活用される重要な統計指標であるが、ヒト疾患に関する遺伝率データベースしか、これまでは存在しなかった。H2DBでは、幅広い生物種を対象に遺伝率を収集しており、2013年2月時点では、67生物種838件の遺伝率を掲載している。実験条件・形質関連SNPs・広義と狭義の区別、などの属性情報も一緒に収集しており、形質の分類はPATOオントロジー (Gkoutos *et al.* 2005) に基づいて行った。分類による遺伝率の傾向は、H2DBの論文 (Kaminuma *et al.* 2013) で報告している。

今後の発展として、DNAPodでは公開予定のカンキツ育種母本11品種の系統間多型情報を準備中で、微生物も解析を進めている。またゲノムワイドに得られる多型情報から系統群において最適マーカーを自動選択するツールや、RNA-seqの発現データの組込みを予定している。H2DBでは分子レベル形質に着目して、遺伝子発現の遺伝率情報などを増やしていく予定である。更に文献抽出だけでなく遺伝率を直接計算するツールも付加して、育種関係者が実際に利用する環境を整えて行く。

2. 画像処理技術の活用による形質評価のハイスループット化

ゲノム解析技術の大きな進歩により詳細な遺伝解析が実現し、大量の系統や個体の形質を効率よく評価する手法の開発は必要不可欠となっている。近年、飛躍的に技術が進歩したデジタル画像技術は、画像から得られる形状や色彩の定量化や計測の自動化を可能とし、計測作業の効率化に有用な技術である。しかし、形質評価のハイスループット化、作業効率化に画像処理技術を活用するには、計測対象 (植物種・器官・計測する場所) や計測内容 (長さ・面積・角度・計数) に応じた画像撮影方法や解析手法、ソフトウェア開発が必要となる。本講演では、発表者が形質評価を効率化するために開発した画像解析ソフトウェアとイメージングシステムの開発事例を紹介し、技術開発における課題と解決方法について解説した。

画像解析ソフトウェアの開発について、面積計測ソフトウェア (七夕ら 2010)・粒径計測ソフトウェア (Tanabata *et al.* 2012) などのソフトウェアを紹介した。面積計測ソフトウェア、粒径計測ソフトウェアは、ドキュメントスキャナやデジタルカメラで植物器官や種子を撮影した画像ファイルを自動処理し、器官の面積や粒の形状値のハイスループット計測を実現する。こうしたソフトウェア開発を効率化・高度化を実現するツールとして、オープンソースライブラリー Open Source Computer Vision Library (OpenCV; <http://opencv.org>, <http://opencv.jp>) がある。数多くの画像処理アルゴリズムを簡単に利用でき、画像処理ソフトウェア開発の効率化に有効なツールであ

る。講演では、簡単なサンプルプログラムを使った利用方法を紹介した。

イメージングシステム開発について、イネの初期成長を自動で画像記録する装置（七夕ら 2008）の開発事例を紹介した。イメージングシステムの活用により、連続的な成長変化の解析にもついた新たな現象解明につなげることが期待できる。一方で、育種研究に有効活用するには装置開発だけでなく、実験で得られる大量の画像データを解析するソフトウェアの開発が重要となる。

大規模かつ詳細な形質評価を実現するため、画像処理技術の活用は有効な解決方法のひとつである。高度な画像処理技術を生かした解析手法やソフトウェアを開発するには、利用者と開発者が協調し研究開発を同時に進めることで研究に有用となる実用性の高い評価技術が実現できると考えている。今回のワークショップのような場を通じ、異分野の研究者がつながり新たな評価技術の開発の機会が増えると良いと思う。

本発表で紹介したソフトウェアの一部は、Web サイト (<http://phenotyping.image.coocan.jp/>) よりフリーソフトウェアとしてダウンロードして利用できる。

3. 大規模遺伝子発現情報からの遺伝子探索手法

① 大規模オミックス解析手法とソフトウェアの構築

高速シーケンサーの利用拡大にともない、公開データベース上にトランスクリプトーム情報が急増している。これらの大規模情報を精査し、遺伝子発現パターンの類似性や特異性を評価することによって、新規遺伝子の同定が促進され得る。従来、発現パターンの類似性評価では、階層的クラスタリング手法が広く用いられている。階層的クラスタリング手法では、ヒートマップとデンドログラムによって、発現パターンが類似する遺伝子群を視覚的に把握できる。ここで、階層的クラスタリング手法は、入力に用いる発現データの大規模化にともない、より多くの計算機メモリと計算時間を要求する。そのため、大規模な遺伝子発現データに対して実行できない。

明治大学や滋賀県立大学などのグループでは、多変量解析を応用し、発現パターンが類似する遺伝子群を迅速かつ簡便に抽出するための手法を開発している (Hamada *et al.* 2011, Manickavelu *et al.* 2012, Yano *et al.* 2006, 特願 2009-063273, 特許第 4302466 号)。この手法では、主成分分析と同様に、遺伝子を部分空間に投影するため、3D ビューワーによる視覚化が可能である。本手法を実行し、3D ビューワー上でのプロット描画も可能な GUI ソフトウェアが明治大学から提供されている。

発現ネットワークは、発現パターンが類似する遺伝子群 (遺伝子モジュール) を把握する上で極めて有用なツールである。発現パターンが類似する遺伝子群を網羅的に抽出することによって、発現ネットワークを直ちに構築できる。構築した発現ネットワークは、Cytoscape や Graphviz などのフリーソフトウェアを用いて容易に描画

し、遺伝子を探索できる。

② オミックス統合データベースの構築

Web データベースが提供する情報は多岐に渡る。それらには、ゲノムやトランスクリプトーム、メタボロームなどのオミックス情報、文献などから得られる知識情報、DNA や実験系統などの実験リソース情報などが含まれる。一方で、現在、多くのデータベースが開発されており、研究者が目的の情報を提供するデータベースを探すことも困難となっている。そのため、Web 上の情報を横断的に検索し得る統合データベースやポータル・サイトの構築が強く求められている。データベース Oryzabase では、イネの遺伝子や実験リソースなどの情報を統合して提供している (Kurata and Yamazaki 2006)。トマトのオミックス統合データベース TOMATOMICS (<http://bioinf.mind.meiji.ac.jp/tomatomics/>) では、トマトゲノムに対するゲノム・ブラウザーや cDNA ライブラリー、SNP などの情報を統合して提供している。Plant Genome DataBase Japan (PGDBj, <http://pgdbj.jp/>) では、オルソログや DNA マーカー、実験リソースなどの情報を格納している。バイオインフォマティクス解析の大規模化の進展によって、統合データベースや知識情報を集約した知識ベースの重要性はさらに高まるであろう。

4. QTL 解析の基礎と応用

QTL 解析は、育種学会では 20 年以上前から知られる遺伝解析手法で、マーカー遺伝子型情報を用いて、量的形質を制御する遺伝子座を検出する手法である。QTL 解析を用いると、収量や品質など様々な量的形質に関与する遺伝子座の位置や効果を定量的に評価でき、複雑な形質でも効率的な育種が可能になる。これまでに非常に多くの有用形質について QTL が報告されており、例えば Q-TARO データベース (Yonemaru *et al.* 2010) によると、イネの様々な形質について 1051 個の QTL が報告されている (2013 年 2 月現在)。

本講演は「QTL 解析の基礎と応用」と題して、当初は、QTL 解析の有効性を知らしめるきっかけとなったインターバルマッピング法の原理について解説し、大量の塩基多型情報が得られる近年の事情に適した解析方法を提示する予定で準備を進めていた。しかし実際には、大量データの取り扱いに便利な統計解析用フリーソフト R のインストールと QTL 解析の演習を行うことにした。演習形式の講演に変更した理由は、本ワークショップ主催者らの企画してきたグループ研究集会 (Breeding Informatics) において研究用ツールの使い方講習会が開催されてきたこと、またワークショップという新たな枠組みにおいては演習形式の講演にも需要があると考えたためである。今回、無線 LAN 付きの PC 持参で参加された方には、演習内で R および qtl ライブラリーのインストールを行ってもらい、Web サイト経由で配布したレジュメに従って、例データによる QTL 解析を体験していただ

いた。

QTL 解析は学会内では使用者も多くその意味で新鮮味のないテーマでもあるため、本講演は少数の参加者を見込んでいた。しかし実際は、講義室が埋まるほど多くの人に参加していただけた。また、講演後も演習内容やレジュメの内容についての質問を複数いただいたことから、本講習会の需要は一定以上あったのだろう。

演習では QTL 解析手法の中身についてまで触れることができなかった。インターバルマッピング法やその改良法による QTL 検出の成否は、解析に用いるマーカー順序やその遺伝距離の妥当性に依存する。一方、近年のシーケンス技術の進展により（本ワークショップ講演 1）、ゲノム情報の基盤がなくとも大量の塩基配列多型を得られる時代になってきた。連鎖地図の得にくい複雑な生物材料の場合は、インターバルマッピング法よりも、多型マーカーと形質値との単回帰分析により検出した有意な隣接マーカーで QTL の定量的評価を行うのがよいだろう（閾値は並べ替え検定で計算できる）。

演習に用いたレジュメは <http://www.eonet.ne.jp/~vor-dem-gesetz/QTL.htm> から利用できる。R/qtl は遺伝子型データのエラーを検出するためのコマンドが充実しており、相互作用 QTL や多重 QTL の検出などより優れた解析もできる。演習では触れられなかったそれらの内容についても、上記の Web ページに追記する予定である。

5. ゲノミックセレクションのための統計手法

現在、ゲノムワイド DNA 多型に基づき個体の遺伝的能力を予測して選抜を行うゲノミックセレクション（genomic selection, Meuwissen *et al.* 2001）という手法が、新しい選抜法として注目を集めている（Heffner *et al.* 2009, Jannink *et al.* 2010, 岩田 2012）。ゲノミックセレクションでは、選抜を行う際に対象とする環境での栽培試験を行う必要が無い。その代わりに、DNA 多型に基づいて目的形質を予測し、予測値の優劣に従って選抜を行う。ゲノミックセレクションを世代促進と組み合わせれば、年に複数回選抜を行うことも可能となり、育種的大幅な高速化につながると期待されている。このような手法が実用技術として注目される背景には、昨今のゲノム解析技術の進展がある。例えば、高速シーケンサーを用いたジェノタイピングにより、数百個体について数千～数万箇所の DNA 多型を一度に計測することも可能となっている（Davey *et al.* 2011）。こうしたゲノム解析技術の進展により、表現型選抜よりもゲノミックセレクションのほうが、コスト面から考えても有利になりつつある。そこで、本講演では、今後ますます重要な選抜法となると考えられるゲノミックセレクションについて、そこで必要となる統計モデルおよび手法の解説を行った。

ゲノミックセレクションでは、トレーニング集団とよばれる系統群を用いて、予測モデルが構築される。具体的には、トレーニング集団で収集された目的形質の表現

型データとゲノムワイドマーカーの遺伝子型データに関連づける予測モデルを構築する。ゲノムワイドアソシエーション研究（genome-wide association study, GWAS）も同様のデータから関連遺伝子の検出を試みるが、ゲノミックセレクションと GWAS ではモデル化の目的が大きく異なる。GWAS では表現型変異を司る遺伝機構をモデル化することに興味があるのに対し、ゲノミックセレクションではマーカー遺伝子型から表現型変異を精度良く予測することがモデル化の目的である。したがって、その予測モデル構築においては、量的遺伝学に基づくモデルだけでなく、サポートベクターマシン（support vector machine）やニューラルネットワーク（neural network）等に代表される機械学習（machine learning）法まで、様々な方法が用いられる（Lorenz *et al.* 2011, Heslot *et al.* 2012）。講習会では、統計解析用フリーソフト R を用いて計算できる幾つかの方法について、その特徴と具体的な R コードについて解説した。また、様々な方法の中から最適なものを決定するための交差検証（cross-validation）法についても解説した。

講習会の冒頭でも述べたように、予測モデルそのものは R が使えれば直ぐに計算できるものも少なくない。また、各方法についてその仕組みを詳細に理解しておく必要は必ずしも無い。“ブラックボックス”であっても予測精度が高ければ良いというのが予測モデルに求められる特性だからである。ゲノミックセレクションに用いられるモデル化手法について学ぶには、「習うより慣れよ」で、まずは講習会で紹介した R スクリプトを使って、実際に計算してみることを勧める。講習会で用いた R スクリプト、サンプルデータ、プレゼンファイルは（<http://papaya.ab.a.u-tokyo.ac.jp/groups/iwata/>）からダウンロードできる。なお、ゲノミックセレクションのサンプルデータには一般公開されているものがいくつかある。例えば、Endelman and Jannink（2012）には、トウモロコシ、イネ、オオムギ、ブタのデータのダウンロードサイトが記述されている。

おわりに

本ワークショップには、非常に多くの人にご参加いただいた。これはワークショップを企画した著者らにとっては大変ありがたいことであった。一方、参加者の多さは、本ワークショップのような情報交換の場が少ないことを反映したものとも考えられた。通常の学会発表では、情報解析手法やツールに関してその詳細を情報交換することは難しい。今回のワークショップは、同様のワークショップを今後も継続的に企画することの必要性を再認識する機会ともなった。

本ワークショップのテーマは、情報解析ツールを使いこなすためのノウハウを紹介することであった。「ツール」は便利に使えることが重要であり、使うことが利用

者にとって大きな負担になるようでは、本当の意味での「ツール」とは言えない。情報解析の手法やソフトウェアをいかに「ツール」とするかは、ツール開発にも携わる著者らの重要な使命の一つでもある。しかし、それは言うほどたやすいことではない。情報解析の分野も、他の分野と同様に細分化が進んでおり、著者らでさえも自分の研究に関連する分野を少しでも離れると最新動向が全く分からないのが実情である。また、問題をさらに難しくしているのは、解析対象となる情報そのものが非常に速い速度で変化していることである。例えば、数年前までは、遺伝子発現データと言えばマイクロアレイの画像解析から得られたデータだったわけだが、最近では、高速シーケンサーで網羅的に配列解読された RNA のリードデータに代わりつつある。このような状況下では、限られた数の研究者だけで様々な情報解析ツールについてその最新動向を把握し続けるのは容易ではない。読者の皆様には、今後、ワークショップの参加者としてだけでなく、是非、講師としてご講演いただき、ご自分の研究が関連する分野の情報解析ツールについて、その使い方に関するノウハウをご紹介いただきたい。

大規模なオミックス解析を処理するために膨大な計算機リソースを投入し続ける従来のアプローチでは、研究を発展し続けることはできない。この要因として、解析サーバーの導入・運営に要する費用や外部委託費が研究費を圧迫し、本来の研究目的を十分に実施できないこと、また、解析を実行する上で解析サーバーの操作を要するために、データ生産者であるフィールド研究者やウェット研究者が自ら解析を実施し難いことなどが挙げられる。特に、解析をバイオインフォマティクスや外部委託に依存する場合、きめ細やかな対応が困難なばかりか、意思疎通のずれによって間違った解析結果が送られてくることもあり得る。誤りに気づかないまま、研究を展開してしまうなど、問題点は大きい。フィールド・ウェット研究者が、計算機の操作法を少しでも習得すれば、独自に解析を実行できるはずである。また、バイオインフォマティクスは、新たな手法を開発する際には、GUI ソフトウェアや Web ツールを構築すべきである。

冒頭でも述べたように育種分野で蓄積されるデータは「ビッグデータ」の域に達しつつある。今後も爆発的に増加していくデータをどのように解析し、そこからどのようにして知見を引き出していくかということは、今後、非常に重要な課題となっていくと考えられる。さらには、育種学研究を含む植物分野では、国内においてバイオインフォマティクス手法に精通した人材が少ないという問題も抱えている。我が国が育種学研究を牽引し続けるためにも、若手研究者に対してバイオインフォマティクス手法を習得するための機会をできる限り提供し続けることが重要である。今後、本ワークショップに引き続き開催されるであろうワークショップが、そのような課題に関する議論の場を提供することができれば幸いである。

謝 辞

本ワークショップは、国内のバイオインフォマティクス研究をリードしている研究者の方々に話題提供をしていただくことができたことに感謝の意を表する。また、電源や無線 LAN などの設営をしてくださった大会運営の方々にもたいへんお世話になった。また、ワークショップの運営にご協力いただいた明治大学のスタッフにも深く感謝したい(岩田・矢野 記す)。

引用文献

高速シーケンサー・アーカイブ配列を用いた系統間 DNA 多型データベース

- Kodama, Y., M. Shumway and R. Leinonen (2012) The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res.* 40: D54–D56.
- Kaminuma, E., T. Fujisawa, N. Sakamoto, Y. Tanizawa, N. Kurata, T. Shimizu and Y. Nakamura (2013) H2DB: a heritability database across multiple species by annotating trait-associated genomic loci. *Nucleic Acids Res.* 41: D880–D884.
- Kaminuma, E., J. Mashima, Y. Kodama, T. Gojobori, O. Ogasawara, K. Okubo, T. Takagi and Y. Nakamura (2010) DDBJ launches a new archive database with analytical tools for next-generation sequence data. *Nucleic Acids Res.* 38: D33–D38.
- Gkoutos, G.V., E.C. Green, A.M. Mallon, J.M. Hancock and D. Davidson (2005) Using ontologies to describe mouse phenotype. *Genome Biol.* 6: R8.

画像処理技術の活用による形質評価のハイスループット化

- 七タ高也・山田哲也・清水悠介・篠崎良仁・金勝一樹・高野 誠 (2010) 植物器官のデジタル画像面積を効率的に計測できる領域抽出ソフトウェアの開発. *園芸学研究* 9: 501–506.
- Tanabata, T., T. Shibaya, K. Hori, K. Ebana and M. Yano (2012) *SmartGrain*: High-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiol.* 160: 1871–1880.
- 七タ高也・清水久代・篠村知子・高野 誠・宮村(中村)浩子・斎藤隆文 (2008) イネ初期生育期における表現型解析のための画像計測システム. *電気学会論文誌 C 部門* 128-C: 962–969.

大規模遺伝子発現情報からの遺伝子探索手法

- Hamada, K., K. Hongo, K. Suwabe, A. Shimizu, T. Nagayama, R. Abe, S. Kikuchi, N. Yamamoto, T. Fujii, K. Yokohama *et al.* (2011) OryzaExpress: An integrated database of gene expression networks and omics annotations in rice. *Plant Cell Physiol.* 52: 220–229.
- Kurata, N. and Y. Yamazaki (2006) Oryzabase. An integrated biological and genome information database for rice. *Plant Physiol.* 140: 12–17.
- Manickavelu, A., K. Kawaura, K. Oishi, T. Shin-I, Y. Kohara, N. Yahiaoui, B. Keller, R. Abe, A. Suzuki, T. Nagayama *et al.* (2012) Comprehensive functional analyses of expressed sequence tags in common wheat (*Triticum aestivum*). *DNA Res.* 19: 165–177.

Yano, K., K. Imai, A. Shimizu and T. Hanashita (2006) A new method for gene discovery in large-scale microarray data. *Nucleic Acids Res.* 34: 1532–1539.

QTL 解析の基礎と応用

Yonemaru, J., T. Yamamoto, S. Fukuoka, Y. Uga, K. Hori and M. Yano (2010) Q-TARO: QTL annotation rice online database. *Rice* 3: 194–203. (<http://qtaro.abr.affrc.go.jp/>)

ゲノミックセレクションのための統計手法

Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen and M.L. Blaxter (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12: 499–510.

Endelman, J.B. and J.L. Jannink (2012) Shrinkage estimation of the realized relationship matrix. *G3 (Bethesda)* 2: 1405–1413.

Heffner, E.L., M.E. Sorrells and J.L. Jannink (2009) Genomic selec-

tion for crop improvement. *Crop Sci.* 49: 1–12.

Heslot, N., H.P. Yang, M.E. Sorrells and J.L. Jannink (2012) Genomic selection in plant breeding: a comparison of models. *Crop Sci.* 52: 146–160.

岩田洋佳 (2012) 「ゲノム育種」再び。次世代シーケンサーは新しい育種の扉を開くのか? *作物研究* 57: 77–82.

Jannink, J.L., A.J. Lorenz and H. Iwata (2010) Genomic selection in plant breeding: from theory to practice. *Brief. Func. Genomic Progeomic.* 9: 166–177.

Lorenz, A.J., S. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, K.P. Smith, M.E. Sorrells and J.L. Jannink (2011) Genomic selection in plant breeding: knowledge and prospects. *Adv. Agron.* 110: 77–123.

Meuwissen, T.H., B.J. Hayes and M.E. Goddard (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.

当日 演題

講演

1. 「新型シーケンサのアーカイブ配列と解析パイプライン：系統間 SNP 解析を事例として」
神沼英里¹・望月孝子¹・長崎英樹¹・児玉悠一¹・猿橋智¹・大久保公策¹・高木利久¹・大柳一¹・倉田のり¹・清水徳朗²・中村保一¹ (1. 国立遺伝学研究所, 2. 農業・食品産業技術総合研究機構 果樹研究所)
2. 「画像処理技術の活用による形質評価のハイスループット化」
七夕高也 (農業生物資源研究所)
3. 「大規模遺伝子発現情報からの遺伝子探索手法」
矢野健太郎・土田博子・五十嵐香理 (明治大学農学部生命科学科)
4. 「QTL 解析の基礎と応用」
清水顕史 (滋賀県立大学環境科学部)
5. 「ゲノミックセレクションのための統計手法」
岩田洋佳 (東京大学大学院農学生命科学研究科)