

# プロジェクト名： 超大容量ゲノム・多元軸表現型データの統計情報解析による遺伝機能システム学（略称：遺伝機能システム学）

プロジェクトディレクター： 倉田 のり教授（国立遺伝学研究所）

## 〔1〕 研究計画・研究内容について

### (1) 目的・目標

本プロジェクトでは、大量ゲノム配列情報や遺伝子発現情報のデータ解析手法と多元的な生物表現型多様性の統計モデリング手法を開発する。両者を統合することにより複雑な遺伝的相関構造を描出するための方法を開発し、モデル生物に適用してゲノム機能と遺伝的ネットワーク抽出を行う。これにより、多数の遺伝因子の高次連関から形成される生物多様性を、システムとして理解することを目指す。

これらの研究の効率的推進のため、本プロジェクトは以下の3つのサブテーマを設定して行う。

- (サブテーマ1) 次世代シーケンサによるゲノム関連情報の大規模生産とその情報解析手法の開発（代表、藤山秋佐夫：情報研／遺伝研）
- (サブテーマ2) 大量ゲノム関連データと多元的な生物表現型多様性データの統合による遺伝的相関構造描出のための統計手法の開発と最適化（代表、栗木 哲：統数研）
- (サブテーマ3) 大量で多元的なデータの情報・統計手法を適用したゲノム機能と遺伝的ネットワーク抽出（代表、倉田のり：遺伝研）

サブテーマ1では、最新のゲノムテクノロジーを駆使して年あたりペタバイト級の超大容量ゲノム・遺伝子関連データを系統的に生産し、統計情報解析研究と融合させることによって、生命システム原理についてのデータセントリックな理解を目指す。

サブテーマ2とサブテーマ3においては、まずマウス、イネ、ゼブラフィッシュ、ショウジョウバエなどのモデル生物のゲノム配列・遺伝子型、遺伝子発現表現型、表現型因子、背景因子、時系列情報などのゲノムおよび表現型の多型・多様性データの抽出法を確立する。抽出データの多重組み合わせや相関解析などを行う統計的解析手法の開発と、その手法によるデータ解析を行い、生命・遺伝現象に関するメカニズムを新たな手法で解明することを目的とする。とくにデータ取得技術の進展によって、時系列、e-QTL、遺伝集団構造解析の展開で生み出される新たな形での情報取得と、それによって生ずる課題に対応できる手法の開発を目指す。それらの方法論に立脚して、生物表現型多様性を多元的に様々な角度から説明できる、「遺伝機能システム学」の展開を図る。

本プロジェクトでは、サブテーマ1、2、3が相互にその情報を交換しながら、データセントリックの原理に従って生命現象の新たな方法論を見つけ、新規な原理と解釈を提案することを目指して推進する。

### (2) 必要性・重要性（緊急性）

ヒトゲノム解読計画以来、生命科学の研究スタイルは、大量のゲノム情報を基盤とするデータ駆動型へとパラダイムシフトが進行しつつある。近年の超並列大規模シーケンサの登場は世界的にもその勢いを加速させているが、我が国の大学研究機関の対応は大きく後れており、特に遺伝学の研究分野における大量のデータ処理と情報解析に不可欠な統計分野、情報分野との研究交流は、ごく一部のグループを除いては殆ど行われていない。このため、今後の情報ゲノム科学の発展に不可欠の研究コミュニティ形成や、社会にこれらの超大量データを用いた科学知の必要性を認知させる仕組みは脆弱であり、早急の基盤形成が必須である。このような基盤形成を行う場として、本プロジェクトの掲げる研究連携を、生命・遺伝研究の一つの中心拠点である遺伝研と、情報処理・解析の拠点である情報研、統計数理の拠

点である統数研を有する情報・システム研究機構で担う事は、大きな意味がある。

世界での研究分野の現状を考えると、日本でのこのような試みは遅きに失した感もあるが、緊急に進めるべき分野である。

### (3) 期待される成果等（学問的効果、社会的効果、改善効果等）

本プロジェクトの成果は、生命科学、統計学、情報学の各研究コミュニティーに対して異分野間融合研究の有効性を強くアピールするものとなると期待できる。同時に、本融合研究を通して作られる遺伝学と情報学、統計学の共同研究の土壌は、いずれの研究コミュニティーにも学問的に非常に良い影響を与えることが期待される。また、わが国の学術分野で欠落している multi-disciplinary な人材育成の土壌が参加各機関に形成されることが期待される。

多くの局面で統計的データ解析手法は生命遺伝現象データ解析に決定的に有用であるが、測定技術進歩にデータ解析が追いついていないのが現状である。また歴史的には統計学の起源のひとつは、生命・遺伝現象のデータを解析するために生み出されたものであり、現在においても生命遺伝データの解析を目的として開発される手法が普遍化されることにより統計学全体に還元されることは少なくない。それらの方法論は遺伝研究以外においても活用できる汎用的な性質を有するため、統計学への還元も期待される。

これらの新たな研究開発現場の中で学ぶ事により、大学院生、ポスドクなどが新たな学問の形成に寄与しつつ、育って行く事が望まれる。大容量、多様なデータに基づく多次元の視点からの研究が生み出す新たな研究、新たな問題が次の時代の学問の醸成につながる。いずれの視点からも、本課題が遺伝学研究、統計学、情報科学に大きな貢献をもたらすのみならず、これらの分野を超えた新しい融合領域の創成が期待される。

### (4) 独創性・新規性等

本プロジェクトで研究対象とするイネ、マウス、ゼブラフィッシュ、ショウジョウバエ等のモデル生物は、国立遺伝学研究所が独自に開発した実験系統と自然変異を豊富に包含した独創的な遺伝資源である。これらの研究資源の大容量ゲノム関連情報の生産は、世界でトップクラスのゲノム解読能力を有する国立遺伝学研究所のシーケンスセンターが担当する。また、大容量情報のデータ解析には、我が国の情報学の拠点である国立情報学研究所と統計数理解析の拠点である統計数理研究所が担当する。このように、独自の遺伝資源を軸に据え、遺伝学、情報学、統計学の統合的解析を行う「遺伝機能システム学」の構築は、世界に類を見ない新規性の高い試みであり、本プロジェクトの推進はこのような融合的研究体制を組織できる情報・システム研究機構以外ではあり得ない。さらに、本プロジェクトは、情報・システム研究機構外の大学・研究機関からも研究者が参加し、相互に密接な関係を保ちながら研究を進める点でも極めて新規性が高く、大学共同利用機関としての役割を担うものでもあり、他の研究機関で実施することは不可能である。

### (5) これまでの取り組み内容の概要及び実績

サブテーマ1については、これまでにナメクジウオゲノム、メダカゲノム、ラットゲノム、チンパンジーゲノム等の解読を大規模国際共同研究として進めてきた。また、新型シーケンサの導入とゲノム解読への利用を進め、大腸菌、線虫などの生命研究に有用な変異体のゲノム変異部位の特定や、日本産野生マウス系統についての比較ゲノム解読、ヒト個人ゲノムの解読等を行っている(Kasahara et al., Nature, 2007, Putnam et al., Nature, 2008, STAR Consosium, Nat. Genet., 2008, Rensing et al., Science, 2008 など)。また、次世代シーケンシングパイプラインの検討を進め、2008年7月から2011

年3月までに累積で  $20 \times 10^{12}$  塩基のデータを生産した。ニワトリのセントロメア領域については、そこに存在するタンパク質の同定や、培養細胞を用いてゲノム改変できる技術を開発してきた (Hori et al., Cell, 2008; Amano et al., J. Cell Biol., 2009)。また、主にシロイヌナズナを用いゲノム DNA の修飾と機能との関連解析研究を進めている。

サブテーマ2では、関連する以下の成果を得ている(1)機械学習とベータダイバージェンスの方法による形の表現型の計量化と QTL 解析のロバスト化 (Mollah et al. Neural Processing Letters, 2007 など) (2)ロバスト推定に関する基礎的な研究: 外れ値の割合が大きい場合にも潜在バイアスを小さくすることが可能な方法を提案した (Fujisawa & Eguchi, Journal of Multivariate Analysis, 2008) (3)高い相関構造を持つ多重検定の研究: 多重積分の必要なしに簡単に陽に計算ができ、近似精度が高く、保守的でもある検定方法を開発した。(Ninomiya & Fujisawa, Biometrics, 2007) (4)連鎖解析のエピスタシスの解析において、逐次解析やチューブ法を含むいろいろな方法で検定の多重性調整 p 値を与える方法を開発した (栗木哲「QTL 解析の統計モデルと検定の多重性調整」、21 世紀の統計科学、東京大学出版会、2008, Kuriki, et al., arXiv:1012.4921) (5)いくつかの古典的 QTL 解析において、影響関数を定義し、マウスのデータについて解析を行った (Dou et al., The 57th Session of the International Statistical Institute, 2009, South Africa)。(6)遺伝子数が多く実験回数が少ないデータに対して妥当な P 値を推定できる遺伝子発現差解析手法の成果をまとめた。これらの成果のうち、(1)、(4)、(5)については、サブテーマ(3)とも関連して、第1期新領域融合研究プロジェクト「生物多様性解析」を中心とする成果である。

サブテーマ3についても、サブテーマ2と同様に第1期新領域融合研究プロジェクト「生物多様性解析」に基づく成果が得られ、第II期への展開が期待される。主なものとして、ゲノム配列多型に基づいたマウス複合形質の解析基盤の整備と遺伝的解析 (Takada et al., Genome Res., 2008, Oka et al., Genetics, 2007, Amano et al., Dev. Cell, 2009)、マウス行動解析の遺伝的基礎 (Takahashi et al., Behav Genet., 2006, Mamm. Genome, 2008, Behav. Genet., 2009, Umemori et al. BMC Genetics, 2009)、イネ及びマウスの系統間ゲノム多型および発現差検定法 SNEP の開発 (Fujisawa et al., BMC Bioinfo., 2009) や生殖隔離因子の相互作用解析 (Harushima et al., Mizuta et al., PNAS, 2010) など、様々な成果を得ている。

## (6) 国内外における関連分野の学術研究の動向

新型シーケンサを用いた国際大規模ゲノム解読プロジェクトとしては、ヒト 1000Genomes Project、国際がんゲノムプロジェクト、実験系統マウスゲノム解読プロジェクト ENCODE 計画などが進んでいる。当初は、中核となる大規模ゲノムセンター間のネットワークによる計画で大規模データ生産に主眼が置かれ、統計学や情報学はそのための補助的ツールとしての存在であったが、次世代シーケンサの普及により、大規模ゲノムデータを軸とする研究スタイルが個別的な研究テーマにも使われるようになると、大規模データに対して各々の生命研究テーマや研究材料の特性に適した情報処理、統計処理を個別の研究室レベルで行うことが求められるようになってきている。本計画で実施しているセントロメアのゲノム解析は、巨大な反復配列が存在するために解析が遅れていた領域に、次世代シーケンシングと大規模統計情報処理とを導入し、既存の研究領域から新たな研究スタイルを生み出すことに成功している。本研究計画では、系統的な大規模データを基に生物学と統計情報学が融合することで得られる遺伝システム機能の多元的な理解を研究コミュニティーに提供し、従来は難しかった最先端の研究戦略を具現化するものである。

統計学の観点からは、遺伝子関連データ解析における(a)ゲノムデータ解析のための並べ替え検定手法の開発と多重性調整への応用、(b)グラフィカルモデルの開発、(c)影響分析を行うが、(a)については、

並べ替え検定の技法は広く使われているにもかかわらず、その妥当性や最適性はほとんど議論されておらず、そのような議論に基づいた多重性の調整の議論もない。また(c)については、関連する先行研究は国内外を通してほとんど見あたらない。(b)の各変数間の関連性をグラフィカルに表現する方法は、国内外で数例、統数研と東京大医科研、米国ジャクソン研究所のグループが知られている。また、既存の eQTL データ解析や他の多くの応用分野で研究があるが、それらはそれぞれ異なる考え方（統計モデル）に基づくものであり、とくに QTL 解析、eQTL 解析のデータ形式に適した決定的なものはない。

本研究で主に扱うマウス、イネ、ゼブラフィッシュ、ショウジョウバエにおいては、世界中で多様なゲノム情報の抽出や比較解析が行われているが、本プロジェクトで扱う遺伝資源と実験系統は、世界の中でも独自のリソースであり、これらの遺伝的特性、特に表現型や行動パターン、複合形質、発現遺伝子変異、時系列変化など多様な特性の抽出、およびそれらの統合的情報解析を統計学、情報学、ゲノム解析を駆使して解明しようとする試みは、未だ形を成していない。

## [2] 研究計画

### (1) 全体計画

全体としては、3つのサブテーマ間で機動的、融合的にデータ生産、方法論開発、データ解析を繰り返しつつ、「遺伝機能システム学」を創成し、抽出データ、解析方法論、多重ゲノムデータを用いた体系的な表現型解析、遺伝機能システム解析の成果をコミュニティに公開・発信して行く。

#### サブテーマ1

次世代型といわれる超並列 DNA シーケンサの利用技術開発を進め、特にサブテーマ2、3、と連携しながら遺伝学研究所が有する遺伝資源に対して豊かなゲノム情報を付加し、研究資源として高度化する。また、計画の後期については、解析微量化、効率化が進むことが期待できるため、例えば発生過程における精細な遺伝子発現プロファイル時系列データなど、従来研究では実現不可能であった、生物学者の『夢』ともいえる定量解析の実現をめざす。このためには、プロジェクト内研究者との連携に加え、「地球生命システムプロジェクト」や「データ同化プロジェクト」との連携、さらには共同利用研究機関であることの特長を生かして研究コミュニティと連携しながら計画を遂行する。

#### サブテーマ2

研究計画の概要に述べた3つの課題

- (a) ゲノムデータ解析のための並べ替え検定手法の開発と多重性調整への応用
- (b) 遺伝子情報と発現データのためのグラフィカルモデルの開発
- (c) QTL, eQTL 探索における影響分析

について、遺伝研メンバーが取得したデータから遺伝的知見を引き出すためのデータ解析を行うことを通じて、遺伝学上の発見につながる貢献をするとともに、新たなデータ解析のための方法論を開発する。

#### サブテーマ3

多様なゲノム情報、表現型情報をもつモデル生物、マウス、ゼブラフィッシュ、ショウジョウバエ、イネを用いて、遺伝機能システム学の基盤を作る。野生系統の多様性、ユニークな変異体集団など独自の系統群の持つ遺伝的変異のパワーを多面的に引出し、サブテーマ1、2と共同して多面的、多重的数据の相関解析法の開発と解析を行い、グラフィカルモデル理論や情報の階層的組立てにより、遺伝機能システム学を展開する。各生物種で取り扱う具体的内容は、以下の通りである。

- (a) マウスについては、野生系統を含む実験交配集団や2系統間染色体置換系統群等を用いた、複雑形質（生体内構造、骨格形態、行動パターンなど）の定量化、データ抽出と統計解析手法の開発。ゲノム多様性データとの相関解析。

- (b) イネについては、野生イネを中心に、ゲノム構造、発現遺伝子の質・量・変異解析。組み換え自殖系統の遺伝型評価、系統毎の表現型抽出、e-QTL 解析。野生イネ集団を用いた集団遺伝構造解析、association 解析などを行う。
- (c) ゼブラフィッシュについては、トランスポゾンタグ knock-out 変異系統群を用いた、多様な表現型の抽出とその表出法の確立、特に 2 次元から 3 次元軸への変換手法などの開発。また、遺伝子型関連の検出も行う。
- (d) ショウジョウバエについては、RNAi knock-down 変異系統群を用いて、主に多様な翅形態の変異の抽出法の開発、および変異遺伝子群との相関関係についての解析を行う。

## (2) 各年度の計画

### 平成 22 年度 (予備研究・プロジェクト開始)

今年度は第 II 期の初年度に当たり、研究体制と参加者も大幅に変更された事から、サブテーマ間、用いるストラテジー毎の相互検証を密に行い、速やかに研究が立ち上がり、効率良く融合研究を推進できる体制を構築することを第 1 の目標とする。今期プロジェクトは、効率化の点から 3 つのサブテーマを軸にそれぞれ以下の目標を掲げて研究を行う。

サブテーマ 1 では、次世代シーケンサを導入し、遺伝研が保有する生物系統を材料にゲノム配列と遺伝子発現について大量データ生産を開始する。それらの大量データをもとに情報処理パイプラインの設計を開始する。遺伝研シーケンスセンターには、新型シーケンサとしてイルミナ社製装置 2 台、ABI 社製装置 1 台、ロシュ社製装置 1 台が配置されているが、これらは文部科学省特定領域研究支援の目的で整備されたものであり、数ヶ月の待ち時間でフル稼働している状態である。このため、本プロジェクトを遂行するために、初年度において新型シーケンサ 1 式に加え情報処理に必要な記憶装置、演算装置を新たに導入する。初年度においては、パイプライン構築を主眼に研究を進めるため、参照配列ゲノムが整備されている、ヒト、マウス、イネ、トリ等を対象としたデータ生産を行う。

サブテーマ 2 では、全体計画にあげた課題 (a)~(c) について、既存統計方法についてその可能性と限界、改善する方法について第 1 期の研究結果を踏まえて検討し、研究課題、研究の方向性を検討する。課題 (a) では、いろいろな並べ替え検定の方法を考案し、計算機実験を行う。(b) では、連続変数と離散変数が混在したグラフィカルモデルにおける基本的な推測について検討する。(c) では、古典的ないくつかの QTL 解析の統計モデルについて影響分析の方法を試みる。

サブテーマ 3 では、遺伝子発現量測定や遺伝的相関構造の統計的解析手法を検討するための、実験集団のサンプル数や構成について検討を開始する。また、画像データによる 2・3 次元形態や生体内構造の計測、その他の形質計測の情報学的記述用新規アルゴリズムの検討と開発を行う。さらに、複合形質データや時系列データの統計モデリングの可能性について検討する。

### 平成 23 年度

サブテーマ 1 では、大量ゲノム配列決定パイプラインの最適化を進める。前年度に設計した情報処理パイプラインとシーケンシング生産パイプラインを融合し、サブテーマ 2、3 へのデータ提供を開始する。また、セントロメア領域、テロメア領域等従来のゲノム研究から取り残されていた重要なゲノム機能領域と、ゲノム複製、遺伝子発現、エピジェネティック制御を解明するための大規模解析を開始する。

サブテーマ 2 では、提案した並べ替え検定法や多重性調整法に対して、論理的な正当性を与える。QTL, eQTL 解析では、離散変数、連続変数を同時に扱う必要があるため、それらの混在したグラフィカルモデルのあてはめのためのアルゴリズム開発を行う。また、グラフィカルモデルの当てはめを通して eQTL 解析の結果を視覚化する方法を開発する。また、グラフィカルモデル当てはめにおける影響分析、ロバ

スト推定の方法を検討する。

サブテーマ3では、サブテーマ1で生産された遺伝子発現情報とゲノム配列情報等を多角的にデータベース化し、発現量に差のある遺伝子を検出する統計学的手法を検討する。画像データその他の形質計測データにも、新規アルゴリズム・統計モデリングによる記述手法の適用を試みる。また、隠れマルコフモデル等を適用した時系列データの自動記述・定量化を試みる。

#### 平成 24 年度（中間評価）

サブテーマ1では、大規模ゲノム配列解析パイプラインを完成させると共に、ゲノム配列データ生産を軌道に乗せる。また、データの可視化を行い、Webを通して研究コミュニティに公開する。

サブテーマ2では、提案した並べ替え検定、重性調整法の有効性を様々な擬似データによって確認する。グラフィカルモデルのモデル選択のためのアルゴリズム開発を行う。また eQTL の Lod マップから遺伝子関連性を抽出し可視化する方法を開発する。グラフィカルモデル当てはめにおける影響分析、ロバスト推定の方法を、前年度に引き続き検討する。

サブテーマ3では、生物系統毎の遺伝子発現量とゲノム多型情報を統合して、ゲノム変異に連動する発現変動要因の検出を行う。情報処理・統計モデリングによって記述した画像データやその他の形質計測データと大量ゲノム多型情報とを連関させる統計手法の最適化を行う。時系列データについてもゲノム情報との連関解析を試みる。以上のデータ群はテーマ2の遺伝的相関構造の統計的手法開発のデータソースとして利用する。

#### 平成 25 年度

サブテーマ1では、ゲノム配列データの生産と解析研究を継続する。器官形成、個体形成、環境適応など基本的な生命現象に関わる遺伝子発現調節機構について体系的な遺伝子発現プロファイル作成を行う。このため、モデル生物を対象に適切な発現データを得るために必要な試料調製手法や、個別の生命現象に特異的な大規模データを取り扱うための情報解析手法の開発を行う。

サブテーマ2では、提案した多重性調整 p 値推定法やグラフィカルモデルアルゴリズム、影響分析やロバスト推定を実装したソフトウェアを開発し、様々なゲノムデータに対する統計解析に適用する。また、それまでに得られたゲノムワイドに遺伝子発現の構造を理解する複数の統計的方法を整理し、遺伝学の具体的な問題に特化した統合手法を開発する。

サブテーマ3では、e-QTL のゲノム上の責任因子群についてのネットワークのエピスタティックな構成を取り出すための時系列データを含む多面的な表現型データを集積する。生物系統の個々の表現形質について、生産された大量ゲノム多型情報を表現型情報に結びつける統計手法の最適化を行う。サブテーマ2で開発するグラフィカルモデルのゲノム機能のネットワークへの適用を試みる。

#### 平成 26 年度

サブテーマ1では、技術開発の状況を見ながら前年度の計画をさらに推進させるとともに、解析対象とする生物種、生物系統、さらに生命現象を拡大させる。また、遺伝子発現プロファイルデータをサブテーマ2や3に提供し連携して遺伝子発現のネットワークを解明する。

サブテーマ2では、前年度に引き続き提案した多重性調整 p 値推定法やグラフィカルモデルアルゴリズム、影響分析をより様々なゲノムデータに対する統計解析に適用するとともに、生じた問題点を克服する方法を考える。

サブテーマ3では、一連の e-QTL 解析結果をとりまとめ、ゲノム上の遺伝因子群の特定とネットワークとしての因子群の相互連関を視覚化し記述する手法を確立する。画像データや時系列による多面的な

表現形質変動についても、実データを用いたゲノム機能との連関解析を行い、ゲノム機能ネットワーク全体の描出を試みる。

#### 平成 27 年度

全プロジェクトを通して得られた研究成果をとりまとめ、論文発表と Web を通じた公開を行う。特に、次世代シーケンサで生産したゲノム関連情報については適宜アノテーションをつけてデータベース化して web 公開する。新規に開発したアルゴリズム、ソフトウェアについても Web 発信する。また e-QTL データの統合のための国際的な共同体制の構築をめざす。最後に研究成果を公開するための国際シンポジウムを開催する。

#### 平成 28 年度以降の展開

サブテーマ 1 については本プロジェクト発足時には既にゲノムデータの生産量の増大速度が CPU におけるムーアの法則を凌駕していたことが示すように、量的側面からだけ見ても本プロジェクトの中心テーマでもある超大容量データを基軸とする現代生物学の流れは一層の広がりを見せることが予想される。したがって、こうした状況に対処するためには電子ハコモノ整備的な計算機ハードウェアの進歩と従来型の情報工学に依存することは不可能であり、新たな発想のもとに情報学、統計学と融合した新しい学術分野の創造と展開が必要である。本プロジェクトを先駆けとする学術分野での具体的な展開としては、個体を形成する細胞から集団を対象とし、さらに時間軸、空間軸を加味した統合的なゲノム・遺伝機能システムが展開されるであろう。本計画で得られたセントロメアやテロメアのゲノム領域解析からの知識をもとに、制御可能な人工染色体の作成も期待される。

サブテーマ 2 とサブテーマ 3 では、両者を組み合わせる事により、以下のような展開が可能になると考えられる。マウス行動データに関しては、ここまでに確立してきた表現形質の抽出法をもとに、多様性を持つ生物集団における表現形質の時系列変化を解析し、更にゲノム遺伝子座マーカーの浮動および選択について、実際の実験結果とシミュレーションによる結果とを比較する段階に入ってゆき、多因子形質の時系列変化におけるゲノムの効果を明らかにしてゆく。またショウジョウバエは翅の形成をモデルとした形態形成に機能する全遺伝子のネットワークを元に、ネットワークを構成するそれぞれのシグナル伝達系が初期胚の各組織の形成や寿命など他の生命現象にどのように関与するかを体系的に検証していく。イネにおいては表現形質を形成する遺伝子発現ネットワークからその遺伝子群の作用・機能メカニズムを解明する。また多様な集団を用いてイネゲノム構造の進化とイネの起源を追うことにより遺伝現象の仕組みや表現型獲得の仕組みを明らかにしていくことが可能になる。

### [3] 研究推進・実施体制

サブテーマ 1 では、国立遺伝学研究所シーケンスセンターが直接的なデータ生産に関わるが、従来のゲノム解読では対象になっていなかったエピゲノム修飾、機能的ヘテロクロマチン領域の研究グループや、メタゲノム、個人ゲノム、単一細胞ゲノム等の国立遺伝学研究所の先端的ゲノム研究グループが参加する。また、基礎生物学研究所、理化学研究所、霊長類研究所からゲノム研究グループが参加する予定である。情報解析手法の研究開発は、国立情報学研究所、国立遺伝学研究所、東京工業大学、京都大学、慶應義塾大学から、ゲノム情報解読やシステム生物学の研究グループが参加する。また、新領域研究「地球生命システムプロジェクト」からの試料解析も実施する。さらにサブテーマ 2 および 3 関連のデータ取得では、多様な野生および実験系統からのゲノムおよび発現遺伝子解析等も行うため、これらのデータを各生物のコミュニティで共有したり、比較研究解析の基盤として利用する事も含め、国内外の研究グループと協力、連携する体制を促進する。すでにマウス、イネなどにおける国際連携体制は

整っており、具体的データにより今後の展開を図る。そのためには、シーケンサーのフル稼働とデータ解析部隊の充実は必須である。

サブテーマ2および3の実施に当たっては、データ取得のために遺伝研に研究スタッフを配置する。多次元の遺伝情報、発現変異情報、表現型情報、時系列情報の大量データ取得と解析が鍵になる。しかし、本プロジェクトの目指す遺伝的多様性を軸に据えた同様な研究は未だ国内外で部分的にしか取り組みはなく、今後国内外の状況を見ながら順次連携やコミュニティ形成を推進して行く。統計解析の体制としては、統数研、東京大、九州大などから、統計推測（ロバスト、影響分析、グラフィカルモデル、多重性調整など）と知識を有し、かつデータ解析についても経験と興味を有する研究スタッフをおく。全メンバーが協力してデータ解析の方法論の開発にあたる。

(サブテーマ1) 次世代シーケンサによるゲノム関連情報の大規模生産とその情報解析手法の開発

・研究代表者

〔国立情報学研究所〕 藤山秋佐夫

・共同研究者

〔国立遺伝学研究所〕 角谷徹仁、深川竜郎、堀 哲也、中村保一、神沼英理、豊田 敦

〔融合研究センター〕 辰本将司、松崎肖子、丸山多恵子、Wei-Hao Shang、小林啓恵、商 維昊、望月孝子

〔東京工業大学〕 黒川 顕

〔京都大学〕 矢田哲士

〔慶應義塾大学〕 榊原康文

〔基礎生物学研究所〕 長谷部光泰

〔理化学研究所〕 黒木陽子

(サブテーマ2) 大量ゲノム関連データと多元的な生物表現型多様性データの統合による遺伝的相関構造抽出のための統計手法の開発と最適化

・研究代表者

〔統計数理研究所〕 栗木 哲

・共同研究者

〔統計数理研究所〕 藤澤洋徳、田村義保、川崎能典

〔国立遺伝学研究所〕 城石俊彦、高田豊行、岡(木曾)彩子、小出 剛、倉田のり

〔融合研究センター〕 春島嘉章、堀内陽子、Dou Xiaoling

〔山形大学〕 坂口隆之

〔新潟大学〕 原 尚幸

〔九州大学〕 二宮嘉行

〔政策研究大学院大〕 土谷 隆

(サブテーマ3) 大量で多元的なデータの情報・統計手法を適用したゲノム機能と遺伝的ネットワーク抽出

・研究代表者

〔国立遺伝学研究所〕 倉田のり

・共同研究者

〔国立遺伝学研究所〕 城石俊彦、川上浩一、武藤 彩、上田 龍、小出 剛、高田豊行、久保貴彦



〔統計数理研究所〕	田村義保、土谷 隆、栗木 哲、藤澤洋徳、川崎能典
〔国立情報学研究所〕	井上克巳、北本朝展
〔融合研究センター〕	春島嘉章、堀内陽子、岡(木曾)彩子、梅森十三、田中英希、 Dou Xiaoling、阿部玄武、後藤達彦、宋 剛秀
〔新潟大学〕	中谷明弘、阿部貴志
〔首都大学〕	相垣敏郎
〔政策研究大学院大〕	土谷 隆
〔大阪大学〕	白旗慎吾

## 〔4〕 研究の進捗状況

### サブテーマ1

「次世代シーケンサによるゲノム関連情報の大規模生産とその情報解析手法の開発」

#### 1-1：大規模シーケンスの稼働

22 年度に、最新の次世代型シーケンサとして Illumina 社製 HiSeq2000 型及び試料調製装置 c-BOT を導入した。同装置は、アジア地域で中国 BGI に次ぐ国内第 1 号機であり、当初計画よりも数ヶ月遅れて稼働を開始した。本装置は、読み取り鎖長が 100 塩基、1 回の運転あたりのセータ生産量が約 200Gb で、これまでに遺伝学研究所シーケンサセンターで運用していた Illumina 社製 GenoemAnalyzerGAIIX に較べて約 5 倍の効率向上を果たしている。一方で、読み取り精度が納入時には GAIIX よりも低いなど、情報処理段階での負担が大幅に増加した。さらに 23 年度には、読み取り精度の大幅な向上とデータ生産能力がさら



22 年度に導入した試料調製装置 c-BOT(左)と、HiSEQ2000 次世代型シーケンサ(右)。

さらに 3 倍に増大する予定であり、サブテーマ 2、3 はもとより、「地球生命システム」プロジェクトへの一層の寄与が期待されるが、同時に情報統計処理段階への負荷が大幅に増大することが予想される。国立遺伝学研究所シーケンサセンターでは、本プロジェクトで導入した装置を含む 7 台の次世代型シーケンサー (HiSeq2000 (Illumina), 5500xl SOLiD (ABI), GS FLX (Roche)) からなる大規模配列決定システムを保有している。このため、これらを有機的に運用するためのシステム構築を行うとともに、配列データの処理を迅速に実行するための専用解析サーバーからなる情報処理パイプラインを整備した。具体的には、下記の通りである。

#### (1) 大規模配列決定システム

シーケンス用の鋳型 DNA の調製と品質の確認：サンプルの生物種は、参照ゲノム配列が整備されているイネ (39 サンプル)、ニワトリ (11 サンプル)、シロイヌナズナ (2 サンプル) を中心に進めている。解析の目的は、おもに再シーケンスによる多型情報の抽出が一番多く、次に RNA-seq 解析、メチレーション解

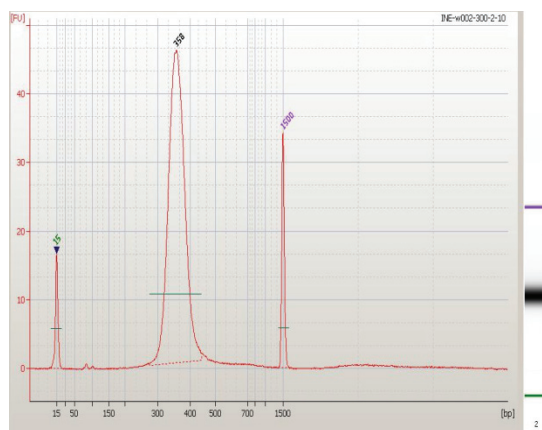
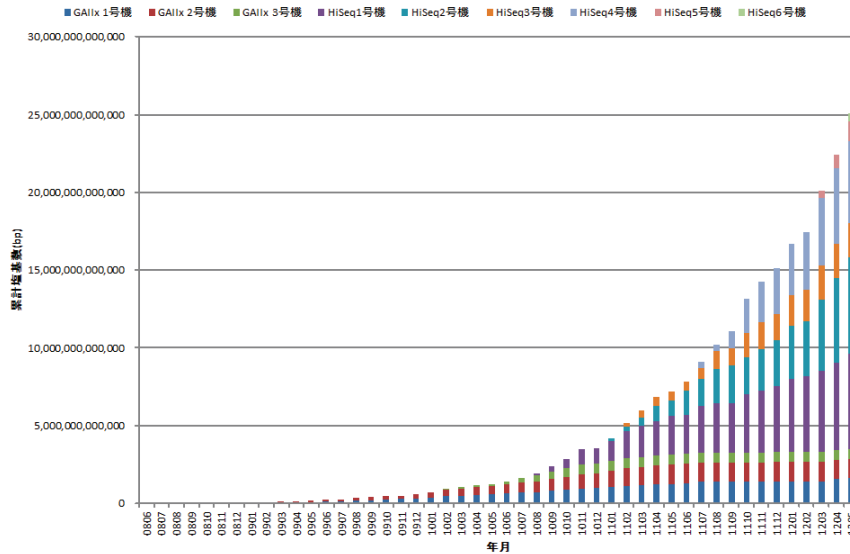


図1 Agilent バイオアナライザーによるフラグメント解析結果

析と続いている。再シーケンスの場合は、ゲノム DNA を数  $\mu\text{g}$  程度から、イルミナのプロトコルを少し改変した方法によりシーケンス用の鋳型を調製した。また、サンプルの品質を調べるために、バイオアナライザーによる解析を行っている（図 1）。その他のサンプルについて支援依頼者と協力して、鋳型調製条件等の設定に努めた。

本報告書執筆時のデータ生産実績を下図に示す。H23 年度末までの累積塩基数は  $20 \times 10^{12}$  を越えている。

H23 年度の融合研究では、本プロジェクト由来の試料に加え、シーズ探索プロジェクトから 12 検体、地球生命システム学プロジェクトから 12 検体、総計 157 検体について次世代シーケンシングを実施し、一次解析を行った後に各グループにデータと関連情報を還元した。



## (2) データ解析・管理システム

ラン終了後のデータ処理と管理についての解析パイプラインの流れを図 2 に示した。全体のシステムは、

### (i) パイプライン解析

- A. ベースコールファイルの作成 (qseq 形式)
- B. マッピング結果ファイルの作成 (export 形式)
- C. qseq 形式を fastq 形式 (リード配列ファイル) へ変換

### (ii) ラン結果管理機能：

リード配列の統計情報を WWW インターフェースで表示

### (iii) 配列精度の確認：マッピング結果をもとにリ

ードの精度に関する統計情報をラン単位、レーン単位に算出するとともに、各リードの QV 値分布、ミスマッチサイト、重複リードの割合、コンタミ率などの一覧を作成（例：表 1、図 3）

### (iv) DRA への登録：パイプライン解析結果をもとに必要となるデータセットを作成

### (v) データ自動転送：二次解析で必要となるデータおよび統計情報を計算機へ自動転送

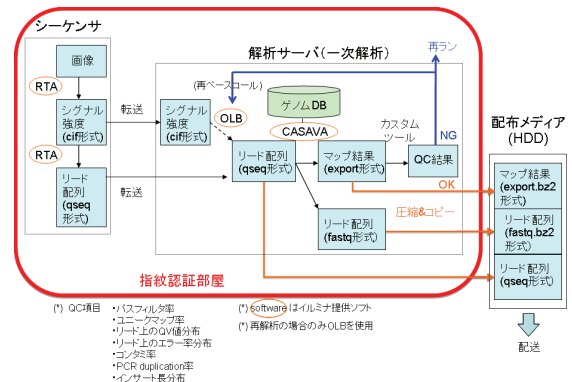


表 1 マップ結果のサマリー

サンプル	レーン	リード	総リード数	QFリード数	(%)	ユニークマップ	(%)	マルチマップ	(%)
イネゲノム W0593	1	1	68,810,188	61,548,183	89	37,407,220	61	21,055,051	34
	1	2	68,810,188	61,548,183	89	36,937,405	60	20,958,766	34
イネゲノム W1236	2	1	67,513,999	60,179,809	89	35,132,331	58	22,043,912	37
	2	2	67,513,999	60,179,809	89	34,759,448	58	21,940,305	36
イネゲノム W0630	3	1	70,669,216	62,069,990	88	38,205,390	62	20,964,890	34
	3	2	70,669,216	62,069,990	88	37,702,431	61	20,837,205	34
イネゲノム W0120	4	1	71,698,788	63,647,820	89	38,176,760	60	22,309,174	35
	4	2	71,698,788	63,647,820	89	37,693,398	59	22,198,659	35
イネゲノム W1715	5	1	57,700,292	51,821,628	90	31,001,200	60	18,293,006	35
	5	2	57,700,292	51,821,628	90	30,628,863	59	18,258,413	35
イネゲノム W0593	6	1	60,269,478	53,953,209	90	31,360,117	58	19,721,333	37
	6	2	60,269,478	53,953,209	90	30,871,873	57	19,641,414	36
イネゲノム W1230	7	1	63,878,392	57,524,502	90	33,834,355	59	20,938,058	36
	7	2	63,878,392	57,524,502	90	33,487,797	58	20,857,399	36
イネゲノム W1981	8	1	66,961,096	58,848,690	88	33,836,438	57	22,008,415	37
	8	2	66,961,096	58,848,690	88	33,455,884	57	21,903,085	37

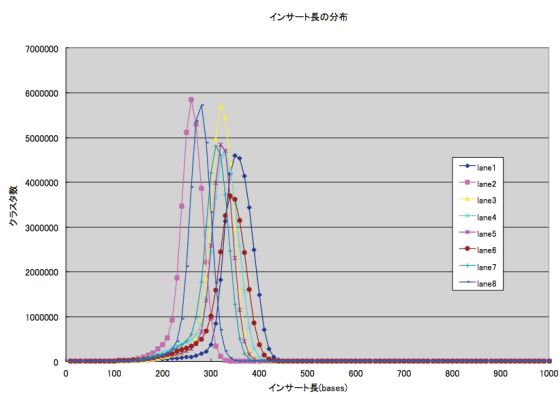


図 3 (A) 鋳型のインサート長の分布

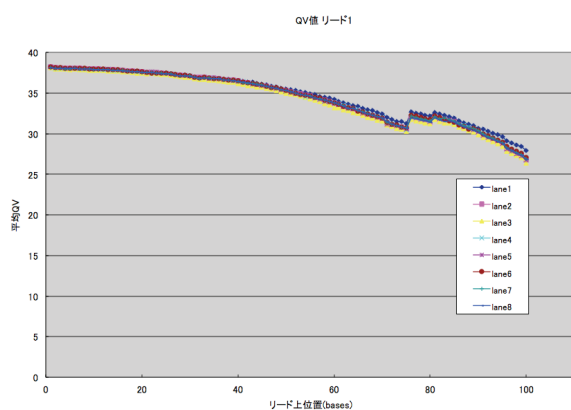


図 3 (B) リード上の各ポジションにおける平均 QV の分布

次世代シーケンシングパイプラインから出力されたすべての配列データは、DDBJ の新型シーケンサのデータアーカイブである DRA に登録される。DDBJ では、登録後のデータ処理をクラウド型で行うパイプライン開発を進めており (図 4)、マッピングと *de novo* アセンブリを行う「基礎処理部」と、ゲノム解析、RNA-seq, ChIP-seq 等の個別の高次処理を実現する「高次処理部」で構成されている。22 年度のプロジェクトでは、このうちの高次処理部を開発した。特に、高次処理部に解析ワークフロー開発基盤の Galaxy を実装し、基礎処理部と高次処理部の接続を行った。

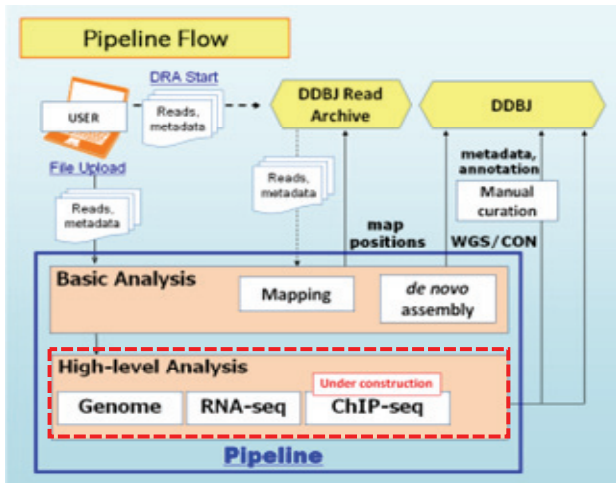


図 4: パイプラインの処理フロー

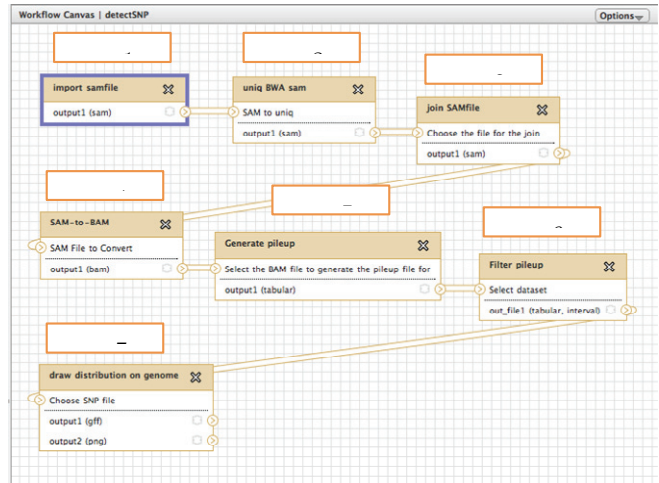
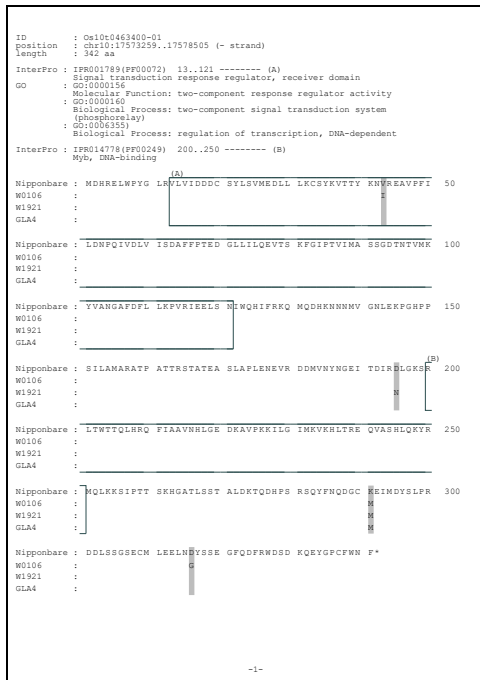


図 5: Galaxy で構築したワークフロー

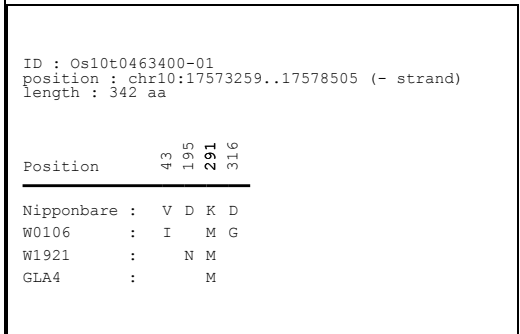
23 年度には、遺伝研植物遺伝研究室との共同研究である「イネ野生種のゲノム解析」にて構築したゲノム解析のワークフローを、種々の動植物や微生物を用いた解析にも適用できるよう増強し、複数系統のゲノムワイドな DNA 多型の検出、注釈付けを行うワークフローの開発とデータベースの構築を進めている。このワークフローでは、DDBJ Sequence Read Archive (DRA) に登録されたゲノムリシーケンスデータをクエリとしリファレンス配列にマッピングして、DNA 多型を検出する。次に、検出した DNA 多型に構造注釈を行い、視覚化して表示を行う。大量研で開発している「DDBJ Pipeline」では昨年度までに、新型シーケンサの出力配列から、ゲノムワイドに DNA 検出するサービスの構築は完了していたため、その後の DNA 多型に対する注釈付けを行う部分の開発が必要となった。そこで、H23 年度は、イネ、シロイヌナズナ、トウモロコシ、ソルガムの SNPs に対する構造注釈付けを行うワークフローを実装した。このワークフローでは、以下の 2 つの構造注釈ができる。(1)既知の注釈情報を使って、遺伝子領域の SNPs を検出する。(2)遺伝子領域の SNPs が同義置換/非同義置換のどちらを起すかを検出する。構造注釈を行った結果は、遺伝子領域のアライメントを塩基配列もしくはアミノ酸配列として視覚化することができる。

2011 年 12 月の日本分子生物学会大会と 2012 年 3 月の日本植物生理学会年会にて、遺伝研比較ゲノム解析研究室が塩基配列決定を行い、DRA に登録された植物遺伝研究室の解析データである、野生イネ *O. rufipogon* の複数系統の開花の促進に働く遺伝子 *Ehd1* の機能の有無を、形質が注釈付けされた SNPs の系統間比較の例として行った解析例を中心に報告を行った。

図 1: 野生イネ *O. rufipogon* (W0106, W1921) と DRA 登録データ *O. sativa* Indica cv. Guangluai-4 (GLA4) の系統間の SNPs 比較。開花の促進に働く遺伝子 (*Ehd1*) は、*O. sativa* cv. Taichung 65 では、機能していない。原因として、219 番目の残基が日本晴はグリシン(G)に対して、アルギニン(R)に変異していることが知られている。*O. rufipogon* 2 系統、*O. sativa indica* cv. guangluai-4 は、日本晴と同じグリシン(G)を示した。



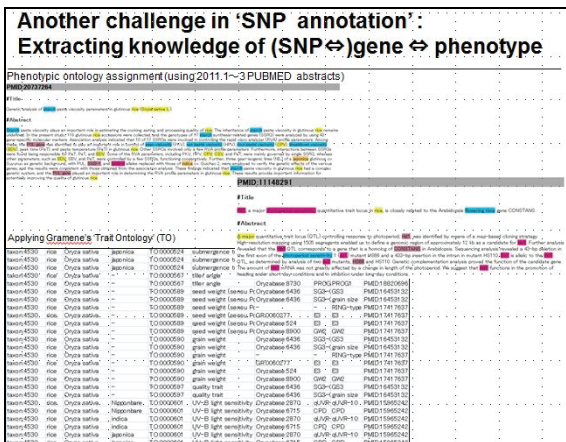
(1) タンパク質全長のアライメント



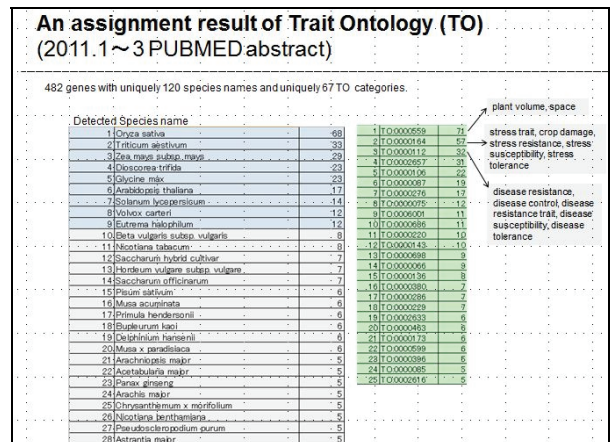
(2) アミノ酸置換部位のアライメント

DNA 多型データベースでは、形質を表現する注釈データとの関係性の蓄積が重要である。このため多形(SNP)が報告された遺伝子と形質間の共起関連を NCBI PubMed 形式の要旨から抽出するテキストマイニングプログラムを開発した。2011年1月から3月までの約20kのPubMedの文献アブストラクトを開発したプログラムでの解析を試行した結果、多くはヒトとマウスのデータであったが、植物については、Trait Ontology (TO) に登録された語句を割り当てる term mapping 処理を実施したところ、120生物種に482の共起関連性を抽出することができた。また67のTOカテゴリを割り当てることができた(図2)。今後、DNA多型データベースの検索に用いるため必須となる注釈データとして応用するために、これら自動抽出結果を手動でキュレーションし改善するシステムの構築を行って、利用可能な精度の形質注釈データへと発展させる予定である。

図2 遺伝子 (SNP) -形質間の共起関連をテキストマイニングプログラムにて抽出



(1) 自動抽出した term の表示



(2) 抽出した TO の分類結果

また、動植物や微生物の複数系統のゲノムワイドな DNA 多型の検出、注釈付けを行うワークフローの開発とデータベースの構築も進めている。このワークフローでは、DDBJ Sequence Read Archive (DRA)に登録されたゲノムリシーケンスデータをクエリとしリファレンス配列にマッピングして、DNA 多型を検出する。近縁種を含む複数系統間での genome diversity を比較するデータは DRA に登録されているが、ゲノム育種を目的としたデータは未だ公開されていない。この為、独) 農業・食品産業技術総合研究所の果樹ゲノム研究チーム(代表：清水徳朗博士)と共同研究体制を組んでカンキツ育種で重要な 11 品種のゲノムをシーケンスした。

「次世代シーケンサを用いたカンキツ品種の高密度DNA多型解析」  
融合共同研究、遺伝研中村研

※高密度多型解析+データベース構築  
※交雑品種の遺伝関係の検証  
有用品種でヘテロ接合アリルが減少してホモ化する問題が顕在化。→高密度多型の確保へ

#	品種名	特徴	
#1	温州ミカン	育種母本, 日本有用品種	国内の主要カンキツ品種 ・清見など多くの後継品種の母本 ・多くの重要な特性(高い単為結果性等)
#2	ダンシータンゼリン	育種母本	
#3	地中海マンダリン	育種母本	
#4	キングマンダリン	育種母本	
#5	紀州ミカン	育種母本	・育種上の重要な特性を持つ
#6	ポンカン	育種母本	
#7	クレメンチン	育種母本(REFERENCE)	
#8	スイートオレンジ	育種母本(REFERENCE)	
#9	晩白柚	育種母本	
#10	アンコール(#3 x #4)	育種母本	・主要なカンキツの母本 ・交雑履歴があるので、多型と連鎖ブロックの遺伝性の検証に利用
#11	清見(#1 x #8)	育種母本, 果樹研有用品種	

所属担当	分担内容	参考
果樹研カンキツ研究 興津拠点：清水	試料, DNA抽出	<a href="http://fruit.naro.affrc.go.jp/seika/seikajyoho.html">http://fruit.naro.affrc.go.jp/seika/seikajyoho.html</a>
遺伝研：藤山研	シーケンシング	<a href="http://www.nig.ac.jp/section/fujiyama/fujiyama-j.html">http://www.nig.ac.jp/section/fujiyama/fujiyama-j.html</a>
遺伝研：中村研	バイオインフォマティクス, DB登録	<a href="http://charles.genes.nig.ac.jp/about/research2/">http://charles.genes.nig.ac.jp/about/research2/</a>

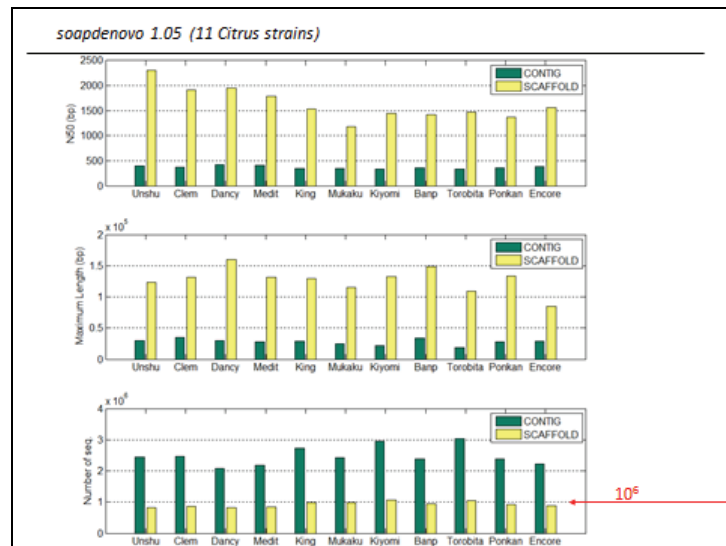
11 品種のリストを図 1 に示す。シーケンスリードは、参照配列(JGI *Citrus clementine*, scaffolds version 0.9, 半数体) にアライメントして SNPs を抽出した。参照配列の全長は 296 Mb で、Scaffold 数は(N=1128) である。Scaffold 数が多いので、CDS を含む配列(N=235)を抽出して BWA (version 0.5.9) でアライメントし、samtools で SNPs を抽出した(表 1)。SAM ファイル中で paired-end reads のうち一方でもユニークにマッチする条件(XT:A:U)を取った。参照配列の品種 *C.clementine* を除くと 25~68 万の SNPs を検出した。SNP の総数は、半数体参照配列と同一品種の 2 倍体の品種(*C.clementine*) で一番少なく、遺伝的に最も遠縁の Banpeiyu が最大になった。ホモ型 SNPs とヘテロ型 SNPs の割合は、品種により異なり、カンキツ属の中で比較的ホモ化が進んでいる Banpeiyu ではホモ型 SNPs の割合が高い事が示された。

表 1:カンキツの育種用品種 11 系統の SNPs 抽出結果。*C.clementine* 半数体配列のうち CDS を有する contigs を参照配列として抽出した。

ID	Strain Name	homo SNPs## **	hetero SNPs#	Total SNPs#
A1_s12	C.unshu Miyagawa-wase	221564	173948	395512
A1_s34	C.clementine	1245	172868	174113
A1_s56	Dancy Tangerine	163084	124301	287385
A1_s78	Medit. Mandarin	130866	127171	258037
A2_s12	King Mandarin	175177	203814	378991
A2_s34	Mukaku Kishu	177934	153737	331671
A2_s56	Kiyomi	208876	217144	426020

A2_s78	Banpeiyu	587978	95780	683758
A3_s12	Torobita Orange	115061	267902	382963
A3_s34	Ponkan	136739	162884	299623
A3_s56	Enkore	92017	153148	245165

また遺伝研のスーパーコンピュータに大容量メモリの計算機が導入されたので、試験的にゲノムのデータから *de novo assembly* を試みた (図 2)。



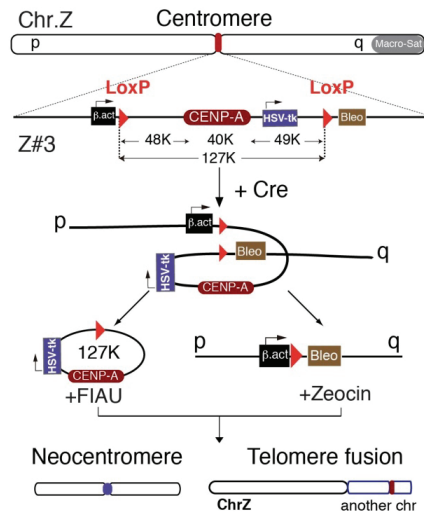
今後は、11 品種間での品種間多型を抽出していく。特に、今年度融合研究で開発した DNA 多型注釈ワークフローを用いて、タンパク質コーディング領域に含まれるか等の品種間多型の構造注釈付を行う。最終的には、育種研究者が利用可能な有用 SNPs マーカの絞りこみツールを開発して、次世代シーケンサの解析ツールへワークフローとして組み込む予定である。

## 1-2：高次ゲノム機能領域の解析

この課題については、セントロメアのゲノム領域の実体を知ることが目的に研究を推進する。細胞が分裂される際には、全ゲノム情報を持つ染色体が娘細胞へと分配される。染色体分配は、両極から伸びた紡錘糸が、染色体の特殊領域を捉えることで遂行される。この紡錘糸が捉える特殊領域はセントロメアと呼ばれ、多くの生物種では、セントロメア領域は巨大な反復配列から構成されている。しかし、一般にゲノム解読解析を行うときには、領域が巨大で複雑な反復構造を構成しているためセントロメア領域は対象から除外されており、また、解読後にこの配列の機能解析を行うことは、きわめて困難であると考えられてきた。本研究では、ニワトリのセントロメアタンパク質の研究を長年行ってきた深川のグループと、ゲノム解析を専門とする藤山のグループと、情報学専門とする榊原のグループが融合することで、ニワトリゲノムを対象として、セントロメアのゲノム配列の実体解明と、その配列の機能的意義を明らかにできると考えている。

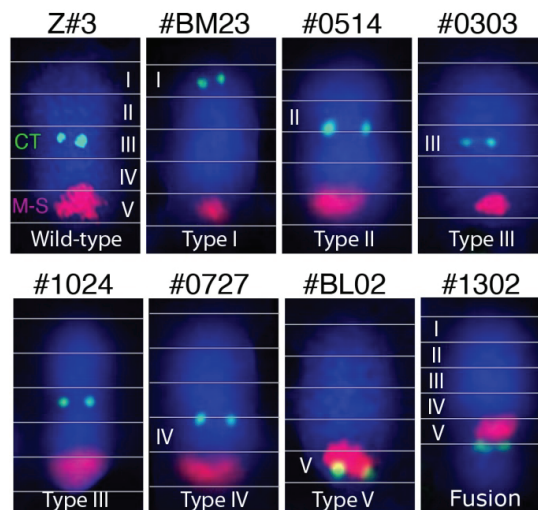
22 年度に、ニワトリゲノムからセントロメアを構成するゲノム配列の同定を行ない、いくつかの染色体において、反復配列を持たない短いセントロメアを持つことを明らかにした。23 年度は、この染色体領域に改変を加えた細胞を作成してこの配列の機能的意義を明らかにすることを目指した。具体的には、セントロメアの両脇に LoXP 配列を導入した細胞を作成して、Cre-リコンビナーゼを条件的に発現

させ、内在性のセントロメアを取り除くシステムを構築した。その結果、ほとんどの細胞では、内在性のセントロメアを取り除くと死滅した。この系で、内在性のセントロメアがないにも関わらず生育してくる細胞を選択に分離した。生育細胞において、本来の場所と異なる場所にセントロメアができることを期待した実験系である。大量のスクリーニングを行なった結果、本来のセントロメアと異なる場所に新しくセントロメアを獲得した染色体をもつ細胞株を 100 種類以上同定できた。



本来のセントロメアを除いて新しいセントロメア (ネオセントロメア) を作成する実験系

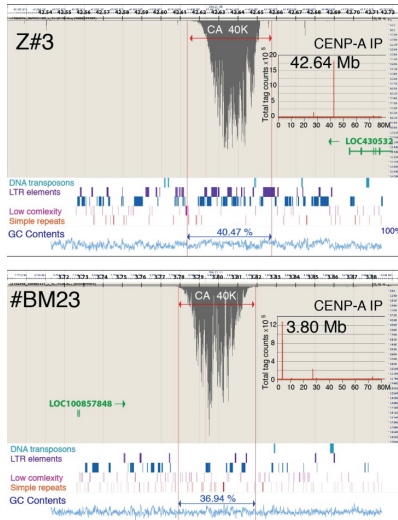
FISH 法により、新しくできたセントロメア (ネオセントロメア) をマッピングした結果、



新しく作成したネオセントロメアを持つ各種染色体

多様な種類が存在していた。そのうちのいくつかについては、セントロメア蛋白質に結合する DNA 領域を新型シーケンサーで解析する ChIP-seq 法により、解析した。その結果、いくつかのネオセントロメアを高精度でマッピングすることに成功した。

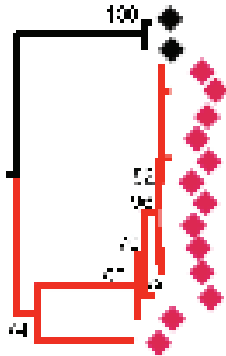




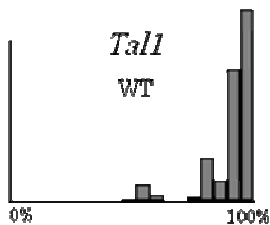
新型シーケンサーによる ChIP-seq 解析令 Z#3 は、本来のセントロメア #BM23 は、本来の場所と全く異なる領域にセントロメアができています

### 1-3 : 植物における高次ゲノム機能研究の超大規模化

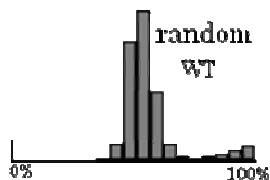
この小課題では、代表的なモデル植物であるシロイヌナズナ(*A.thaliana*)の同属近縁種セイヨウミヤマハタザオ(*A. lyrata*)の内在レトロトランスポゾン *Tal1* の全染色体動態解析を行った。*Tal1* は、この植物の動原体に分布する。この偏った分布が挿入部位特異性によるのか自然選択によるのかを知るため、このトランスポゾンをシロイヌナズナに形質転換し、転移を誘発した。予備的結果は動原体特異的に挿入が起きたことを示唆するものであったが、23 年度に、形質転換体においてゲノム配列解析を行った結果、*Tal1* はシロイヌナズナにおいても、ほぼ全て動原体配列に転移することが明らかになった (Genes Dev 26, 705-713)。



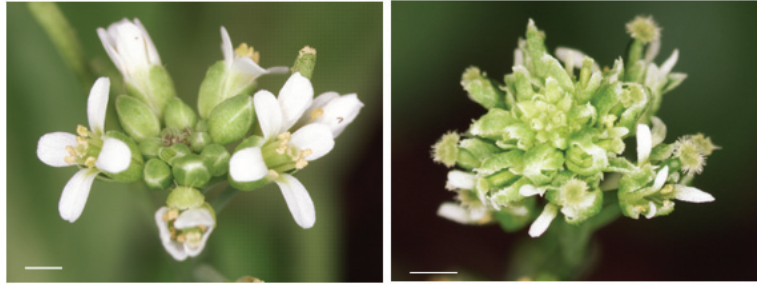
左図:シロイヌナズナの *COPIA93*(黒菱形)とセイヨウミヤマハタザオの *Tal1* および類似配列(赤菱形)の配列比較による系統樹。後者は比較的最近増殖したことがわかる。この *Tal1* の増殖をシロイヌナズナで再現しようとしている。



左図:シロイヌナズナに導入した *Tal1* は動原体に特異的に挿入した。個々の挿入位置に隣接する配列が動原体配列とどれだけ一致するかを示した(上パネル)。同じゲノム配列データでランダムにとった配列での解析をコントロールとして示した(下パネル)。(Genes Dev 2012 Tsukahara et al より)



また、シロイヌナズナのゲノム DNA メチル化が低下した条件下で、多くのトランスポゾンが転移することを見いだしている。今後はこれらの系統のゲノム塩基配列を読むことで、ゲノム構造の変化（トランスポゾンの転移やその他の再編成）を網羅的に調べる予定である。



図： 低メチル化によって増殖するトランスポゾンが湯発したシロイヌナズナの発生異常(右)。左は正常なシロイヌナズナの花。

シロイヌナズナを用いた転移因子の動態解析のほかに、22年度にシロイヌナズナにおいて全ゲノム1塩基レベルでの DNA メチル化プロファイリングの実験系を動かした。23年度は、この実験系を用い、シロイヌナズナやイネ、ヒメツリガネゴケ（基礎生物学研究所、長谷部グループ研との共同研究）を用いて、クロマチンの制御を行う遺伝子の変異体、発生や環境条件の異なるサンプルを用いた解析を進めている。

## サブテーマ2

「大量ゲノム関連データと多様な生物表現型多様性データの統合による遺伝的相関構造描出のための統計手法の開発と最適化」

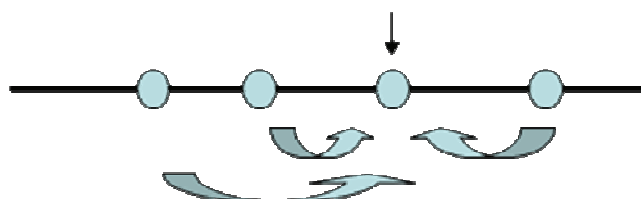
### 2-1：ゲノムデータ解析のための並べ替え検定手法の開発 [藤澤、坂口]

ゲノムデータの一つであるマイクロアレイデータは、しばしば、実験回数が少ない。そのため、多くの統計的手法は、安定的な結果を提示しにくいという問題を抱えている。マイクロアレイデータに対して典型的に行われる遺伝子発現差解析も同じ問題を内在している。検定の代表的な指標であるP値を並べ替え技法によって推定するのに、十分な実験回数を確保できない。

しかしながら、マイクロアレイデータは、たくさんの遺伝子からデータを観測している。そのため、全体の標本数は非常に多い。そこで、ある遺伝子における遺伝子発現差を検出しようとしてP値を推定するとき、他の遺伝子のデータを援用できないかと工夫がされ続けている。そして、実験回数の不足を代用するのに、しばしば、理論的な保証もなしに、アドホックな手法が提案されている。

本研究では、アドホックな手法に、統一的な観点を与えることで、問題の論点を数理的に整理し、可能であれば、数理的なアプローチに基づいて、よりよい手法を提案することを目的とする。まずは、遺伝子発現差解析において、他の遺伝子のデータを利用することが妥当になる条件を数理的に整理して、その条件を満たす検定の中でよりよい手法を提案することを考える。

### 遺伝子発現差解析：P値推定



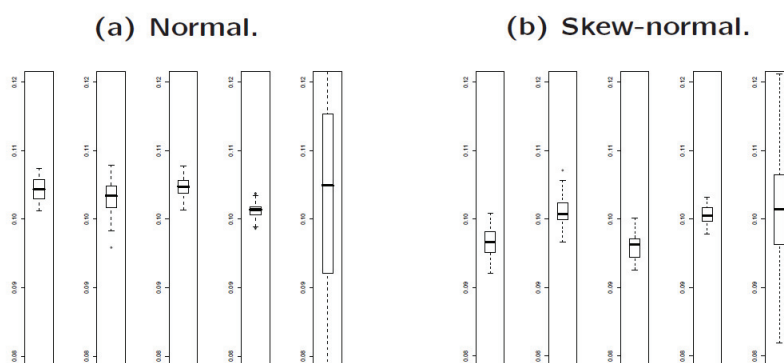
平均差の検定に関しては、上述の目標を達成できるような手法が得られた。以下は、その手法がどのくらいのパフォーマンスを示すかを、実際に検証した結果である。数理的最適性から得られた検定統計量は  $T_s$  と  $T_p$  であった。特に、 $T_s$  は過去に  $Pan$  によって提案された検定統計量に近く、 $T_p$  は状況に応じてより良い手法になることが理論的に期待できる。

導出した方法を有名な Golden Spike Data に応用した結果を下図で示している。 $\alpha$  は有意水準である。Bon は Bonferroni 補正の場合である。 $T_p$  の方が検出力は高いことが、理論的に期待されていたが、実際に、同じ有意水準において、 $T_p$  の方が遥かに多くの遺伝子を発見している。また、偽陽性も非常に低いことが見て取れる。

$\alpha$	$10^{-2}$	$10^{-3}$	$10^{-4}$	$10^{-5}$	Bon*
<b># of detected genes</b>					
$T_s$	340	9	0	0	0
$T_p$	1013	525	161	16	16
<b># of truly detected genes</b>					
$T_s$	317	8	0	0	0
$T_p$	890	508	158	16	16

P 値推定としてのパフォーマンスも検証した。下図はその結果である。詳細なシミュレーションデザインは省略するが、50 回の結果をボックスプロットしている。左側から、 $Pan$  の検定統計量、 $T_s$  と  $T_p$  は既述した通り、 $T_{s2}$  は  $T_s$  を少し修正したもの、 $Perm$  は通常用いられている単純な並べ替え技法による方法である。通常用いられている方法はばらつきが大きく使えそうもない。検出力が最も高い検定統計量  $T_p$  が P 値推定としても最もよいパフォーマンスを示した。これは、検出力を最大とする機構の中に、統計量の分散を小さくする機構が含まれており、その結果と考えられる。

$(n, m) = (10, 5)$ :  $T_{Pan}$ .  $T_s$ .  $T_{s2}$ .  $T_p$ .  $Perm$ .



$T_p$  が適当な条件のもとで検出力を最大にすることは理論的に証明された。実際にどの程度の検出力の差があるかも調べた。その結果が下図である。(詳細なシミュレーションデザインは省略する。) 構築された検定統計量  $T_p$  を使った方法がはっきりと検出力が高いことが見て取れる。特に、過去にアドホックに提案されていた  $T_{pan}$  に比べると検出力の差は顕著である。

Power of test when  $\sigma_Y^2 = 1$ .

	$T_{Pan}$	$T_s$	$T_{s2}$	$T_p$
<b>(b-ii) <math>n = 10, m = 5, n_1 = 7, m_1 = 3</math></b>				
$\alpha = 0.01$				
$\mu_Y = 1$	0.108	0.138	0.123	0.158
$\mu_Y = 2$	0.522	0.644	0.580	0.710
$\mu_Y = 3$	0.886	0.966	0.933	0.981
$\alpha = 0.001$				
$\mu_Y = 1$	0.023	0.029	0.025	0.034
$\mu_Y = 2$	0.182	0.301	0.237	0.329
$\mu_Y = 3$	0.537	0.765	0.638	0.833

2-2 : 遺伝的不適合に関連する遺伝子部位の検出問題のための多重性調整 [二宮、梅森、小出]

マウス KJR とマウス BLG2 をかけ合わせて作られたマウスには、運動失調や発達遅延など多面的な欠陥が観測される。つまり、遺伝的不適合により神経発達異常がもたらされていると考えられる。では遺伝子のどのあたりの部位が遺伝的不適合を引き起こしているか、といったことが課題として生じるわけだが、この課題は多重検定問題として定式化される。どのようなデータを用いるか、そしてどのようなモデルを仮定するかといった設定により、この定式化は異なる。より正確な情報を得るためには設定をより複雑にする必要があるが、一方で、複雑にすると定式化された検定問題の多重性調整は困難になる傾向がある。そこで、まずは最も単純な設定において多重性調整した p 値を評価する方法を開発した。そして現在は、より正確な情報を得るための方法の拡張を遂行しているところである。その結果、遺伝的不適合を引き起こす部位のいくつかの候補を与えることに成功している。

最も単純な設定における方法の詳細を述べると、まず使っているデータは、神経発達異常を示した 162 個体の、平均 10cM 間隔に設定してある 120 箇所のマイクロサテライトマーカーにおける遺伝型である。各個体は、純系の KJR と純系の BLG2 の交配後の自殖 2 代目、つまり F<sub>2</sub> である。したがって、神経発達異常をもたらす遺伝的不適合を引き起こしている遺伝子の部位から十分にはなれたマーカーにおいては、遺伝型として K/K (KJR ホモ)、K/L (ヘテロ)、L/L (BLG2 ホモ) が 1:2:1 といったメンデル比の割合で現れる。一方、マーカーがその遺伝子部位に近ければ近いほど、遺伝型の出現頻度はメンデル比から乖離した傾向をもつようになる。そこで、メンデル比から有意に乖離している出現頻度をもつマーカーを検出することにより、遺伝的不適合を引き起こす遺伝子部位の探索を行った。

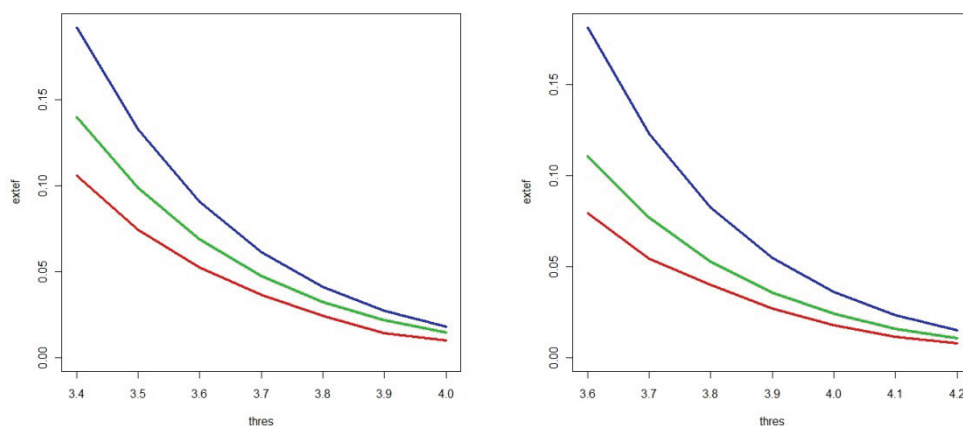
発現の仕方については、KJR の完全優性、優性効果なし、BLG2 の完全優性の三種を考えた。これらをそれぞれタイプ 1, 2, 3 の発現と呼ぶことにする。また「s 番目のマーカーではメンデル比が保たれている」という仮説を  $H_0^{(s)}$ 、「s 番目のマーカーではメンデル比がタイプ t の発現により崩れている」という仮説を  $H_1^{(s,t)}$  とし、 $H_0^{(s)}$  と  $H_1^{(s,t)}$  を比較する検定をすべての s, t ( $1 \leq s \leq 120, 1 \leq t \leq 3$ ) で考えること、つまり 360 個の検定からなる多重検定として問題を捉えた。これにより、例えば  $H_0^{(s')}$  と  $H_1^{(s',t')}$  を比較する検定の多重性調整した p 値が、まわりの検定の p 値より小さい上に有意水準を下回っていれば、s' 番目のマーカー付近にタイプ t' に近い発現の仕方でも遺伝的不適合を引き起こす遺伝子座が存在する、と主張することができる。

検定統計量としては、スコア検定統計量を用いた。 $H_0^{(s)}$  と  $H_1^{(s,t)}$  を比較するスコア検定統計量を  $T^{(s,t)}$ 、その実現値を  $c^{(s,t)}$  としたとき、各検定の多重性調整した p 値は帰無仮説のもとでの確率  $P(\max_{1 \leq s \leq 120, 1 \leq t \leq 3} T^{(s,t)} \geq c^{(s,t)})$  で与えられる。この確率の厳密な値を容易な計算で求めることは一般にできないため、

ここでは保守的に検定をおこなうことを考え、確率の漸近的な上界を容易な計算で求めることを目指した。つまり、 $X^{(s,t)}$ を  $T^{(s,t)}$ の収束先とし、 $P(\max_{1 \leq s \leq 120, 1 \leq t \leq 3} X^{(s,t)} \geq c^{(s,t)})$ の上界をなるべくタイトに評価する問題を扱った。

最も容易な上界は **Bonferroni** の不等式によるものである。すべての  $X^{(s,t)}$ は帰無仮説のもとでカイ二乗分布にしたがうため、その分布関数を  $G_0$ とすれば、確率の評価式は  $360\{1-G(c^{(s,t)})\}$ で与えられる。しかし、これは  $X^{(s,t)}$ が独立に近いときに有効な方法であり、本問題では実際の確率とかけはなれた上界を与えてしまうことになる。なぜなら、マーカー間は平均  $10cM$  しかはなれておらず、隣接するマーカーに対する遺伝型は高確率で同じものとなるため、 $X^{(s,t)}$ と  $X^{(s+1,t)}$ は高い相関をもつからである。また、例えば  $s$  番目のマーカーにおいてメンデル比が崩れているならば、どの発現のタイプでみてもメンデル比は崩れているわけであり、つまり  $X^{(s,t)}$ と  $X^{(s+1,t)}$ は高い相関をもつからである。

この **Bonferroni** の方法を改良するものとして様々な方法が提案されているが、 $\{X^{(s,t)}\}$ の相関構造は複雑であり、有効な改良はほとんどない。本研究においては、このような問題に対応するために提案された **Ninomiya and Fujisawa (2008)**の方法を適用し、裾確率のタイトな上界を与えた。また、その上界を与えるために必要となる  $\{X^{(s,t)}\}$ の相関構造を計算した。下図は、人工的な設定において、**Bonferroni** の方法と提案方法を比較する数値実験の結果を示したものである。提案方法の有用性が確認できる。そして提案方法の適用により、メンデル比を有意に崩れていると判断できる遺伝子部位を二箇所検出することができた。つまり、この二部位が遺伝的不適合に関連していることが予想される。



図：マーカーが  $10cM$ (左)あるいは  $5cM$ (右)間隔で存在した場合の確率評価。  
赤線、緑線、青線はそれぞれ真値(モンテカルロ法による近似)、提案方法、Bonferroni 法に対応する

より正確な情報を抽出するために次に取り組んだのは、直接的な遺伝的不適合のモデル化と、その設定下での多重検定の構成である。具体的には、二つ以上のマーカーに対する遺伝型を同時に考え、どの組み合わせがメンデル比を崩すかを調べた。最も単純な設定下での多重検定に比べ、検定数が大きく増加する上にそれらはより複雑な相関構造をもつことになるが、**Ninomiya and Fujisawa (2007)**の方法と **Taylor et al. (2007)**の方法を組み合わせることで拡張することにより、多重性調整法を開発した。その結果、前述の二箇所の遺伝子部位とは異なる遺伝子部位を検出することに成功している。

さらに解析の精度を高めるため、正常なマウスのデータも組み入れた設定下で多重検定を行った。異常なマウスと正常なマウスの遺伝型が有意に異なる遺伝子座が遺伝的不適合に関わっていると考えられることから、各検定は独立性の検定として問題を構成した。そして検定統計量の相関構造を導出し、多重性調整法を開発した。ただし、正常なマウスと異常なマウスとでは使っているマーカー遺伝子座が若干異なっていたため、共通のマーカー遺伝子座における遺伝型情報のみを用いた。しかし、それが理由で情報を損失し、遺伝子部位の検出に失敗した。現在、全マーカー遺伝子座における遺伝型情報を用い

る方法への修正を行っているところである。

例えば遺伝的不適合を表す指標が複数種ある場合、各指標で検定を構成できることからやはり関連遺伝子検出は多重検定問題となる。そしてこの場合に困難なのは、検定間の相関構造が未知であるためにそれを推定しなければならないことである。現在このケースにおける多重性調整法の開発を目指しており、どのような推定が検出力を上げるかを理論的に考察している段階である。

### 2-3 : 遺伝子情報と発現データのためのグラフィカルモデルの開発 [原、栗木]

グラフィカルモデルは、変数間の条件付独立関係、因果関係をグラフで表した統計的モデルである。グラフィカルモデルは遺伝連関解析において、遺伝子間の相互作用や、遺伝子発現の因果関係のモデリングに有用であると考えられている。遺伝子データの場合、モデルの次元が非常に大きくなること、分解可能モデルのような特殊なモデルを除き、最尤推定量などの統計量が明示的に得られないことなどから、効率的な推測アルゴリズムの開発は重要な課題であると言える。

本研究ではマウスの 2 系統間染色体置換（コンソミック）系統群を用いた遺伝子間相互作用ネットワーク推定のための、統計的モデルの構築、および LASSO 型推定手法を用いた解析の可能性について検討を行った。C57BL/6J (B6) 系統と MSM 系統から作出されたコンソミックマウス系統と B6 系統の肝臓の遺伝子発現量データが、(遺伝子) × (コンソミック系統) の配列データで与えられている。これを用いて遺伝子の制御関係のネットワークを推定することを考える。モデルとしては例えば一元配置モデル

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ij}$$

を用いることが考えられる。ここで  $i=1, \dots, N$  は各遺伝子を、また  $j=1, \dots, M$  は各染色体に対するコンソミック系統を表す添字で、 $y_{ij}$  は B6 系統とコンソミック系統の発現量の比を表すものとする。そのときに、 $\gamma_{ij} \neq 0$  のときに  $j \rightarrow i$  制御関係が存在するとして、 $\gamma_{ij}$  を推定することを考える。しかし一般にこうした制御ネットワークは疎であることが想定できる。そこでこのモデルを LASSO 型の推定方式で推定することが考えられる。上のモデルも線形モデルであるから、LASSO 型の推定は原理的には可能であるが、パラメータに階層関係があるため、その推定アルゴリズムの導出は必ずしも自明ではない。推定アルゴリズムの導出と、その最適性の理論的評価などは現在検討中である。また上のモデルはつの遺伝子間の相互作用のモデルであるが、ここで考えている手法の 3 因子以上の交互作用を含むモデルへの一般化、実際のデータへの適用なども今後の課題である。

### 2-4 : QTL, eQTL 探索における影響分析 [栗木、ドウ、城石、高田]

QTL 解析のための統計モデルには、さまざまなものが提案されているが、その中で最も基本的なモデルは以下の式 (1) で示される単一マーカー分析 (single gene model) である。QTL が 1 つしかないという仮定は現実には正しくない場合もあると思われるが、モデルが簡潔でありスキャン統計量としてのロバスト性があるため、現実に広く用いられている (F. A. Wright and A. Kong (1997))。

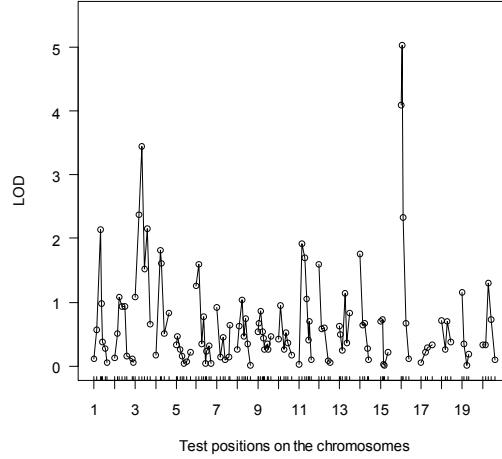


図 1: ロッドスコア

図 1 は、F<sub>2</sub> 集団の雄雌マウス 170 個体について、119 座の遺伝子型と、表現型として血中アディポネクチン (log<sub>10</sub> [ng/ml]) を計測し、単一マーカー分析によって得られた LOD スコアである。ここで用いている統計モデルは、以下のようなものである：

$$y_i = \mu + \alpha z_{ij} + \beta w_{ij} + \nu u_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \quad (1)$$

ここで  $i = 1, \dots, N$  は個体を、また  $j = 1, \dots, M$  は遺伝子座個体を表す添字である。データとしては第  $i$  個体について、表現型  $y_i$ 、遺伝子型  $z_i = (z_{i1}, \dots, z_{iM})$  ( $z_{ij} = -1, 0, 1$ )、また共変量 (性別)  $u_i = 0, 1$  が得られている。また  $w_{ij} = 2|z_{ij}| - 1$  は、優性効果を表すダミー変数である。第  $j$  座のロッドスコアは

$$\text{LOD}(j) = \frac{N}{2 \log_{10}} \sum_{i=1}^N \log \frac{\tilde{\sigma}^2}{\hat{\sigma}_j^2},$$

ただし本報告を通して、モデル (1) の下での  $\theta = (\alpha, \beta, \mu, \nu, \sigma^2)'$  の推定量を  $(\hat{\alpha}_j, \hat{\beta}_j, \hat{\mu}_j, \hat{\nu}_j, \hat{\sigma}_j^2)'$ 、QTL が存在しないという帰無仮説 ( $\alpha = 0, \beta = 0$ ) の下での推定量を  $(0, 0, \tilde{\mu}, \tilde{\nu}, \tilde{\sigma}^2)$  とおく。

第 16 染色体上に見られるピークは既知の遺伝子である。2 番目に高いピークとして第 3 染色体に 2 峰のピークが見られる。この LOD の形が真に 2 つの QTL によるものか、あるいは単に確率的な揺らぎによるものかの違いは、遺伝学的には非常に重要である。そのために遺伝学研究においては、このような LOD スコアの形が認められた場合には、LOD スコアの形状に大きな影響を持つ個体を洗い出し、そのデータを再吟味することが行われている。しかしそのようなことを系統的に行う方法論が知られていなかった。

そのため、いままでの融合研究では LOD スコアの形に影響を与える個体をスクリーニングするための経験影響関数を定義し、その値をもって個体の影響度とする方法を提案した。提案した経験影響関数は次のようなものである：

$$\text{EIF}_i(c) = \sum_{j \in J} c_j \text{EIF}_i(j)$$

ただし

$$\text{EIF}_i(j) = \frac{N}{\log_{10}} \left\{ (\hat{\ell}_{ij} - \tilde{\ell}_j) - \frac{1}{N} \sum_{i=1}^N (\hat{\ell}_{ij} - \tilde{\ell}_j) \right\},$$

$$\hat{\ell}_{ij} = -\frac{1}{2\hat{\sigma}_j^2}(y_i - \hat{\mu} - \hat{\alpha}_j z_{ij} - \hat{\beta}_j w_{ij} - \hat{v}_j u_i)^2 - \frac{1}{2} \log \hat{\sigma}_j^2,$$

$$\tilde{\ell}_i = -\frac{1}{2\tilde{\sigma}^2}(y_i - \tilde{\mu} - \tilde{v}u_i)^2 - \frac{1}{2} \log \tilde{\sigma}^2,$$

また  $\alpha_i$  は想定する LOD スコアの形によって決める定数である。例えば、ここで考える 2 峰性の場合には、 $J$  は 3 座の遺伝子座、また  $c = (1, -2, 1)$  のように選ぶ。

本年度研究では昨年度に引き続き、 $\text{EIF}_i(c)$  がどのくらいの大きさのときに、それが影響度を持つ個体であるかを判定するための合理的な閾値の設定を考える。ここでは、基準化経験関数

$$\text{SEIF}_i(c) = \text{EIF}_i(c) / \sqrt{\sum_{i=1}^N \text{EIF}_i(c)^2}$$

の最大値  $\max_{i=1, \dots, N} \text{SEIF}_i(c)$  の上側 100 $\alpha$  パーセント点 (例えば  $\alpha = 0.05$ ) をその閾値とする。

Step1. 与えられたデータから飽和回帰モデル

$$y_i = \mu + \sum_{j \in J} (\alpha_j z_i^{(j)} + \beta_j w(z_i^{(j)})) + v u_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2).$$

のパラメータを推定する。ここで、 $J$  は QTL が存在する可能性のあるすべての範囲である。

Step2. この回帰モデルに基づいて観測データと同じサイズの新しいデータセット生成し、最大の  $|\text{EIF}(c)|$  を計算する。

Step3. Step2 を十分に繰り返し、最大 EIF の経験分布関数を求め、各個体の  $|\text{EIF}|$  がこの分布関数の上側確率が p 値として計算できる。小さい p 値を持つ個体は影響度の大きい個体として判断される。

この方法を評価するために、以下のシミュレーションを行った。

第 3 染色体のデータから、 $J = \emptyset$ ,  $J = \{3\}$ ,  $J = \{3, 5\}$ ,  $J = \{3, 4, 5\}$  の 4 つのモデルを仮定する。各モデルでは影響度の大きい個体があるかないかの 2 通りのデータセット  $D$  を生成する。前述した方法のように飽和モデルでは、新たに生成したデータセット  $\tilde{D}$  の  $\tilde{T} = \max |\text{SEIF}|$  の分布とデータセット  $D$  の  $T = \max |\text{SEIF}|$  の分布を比べた。

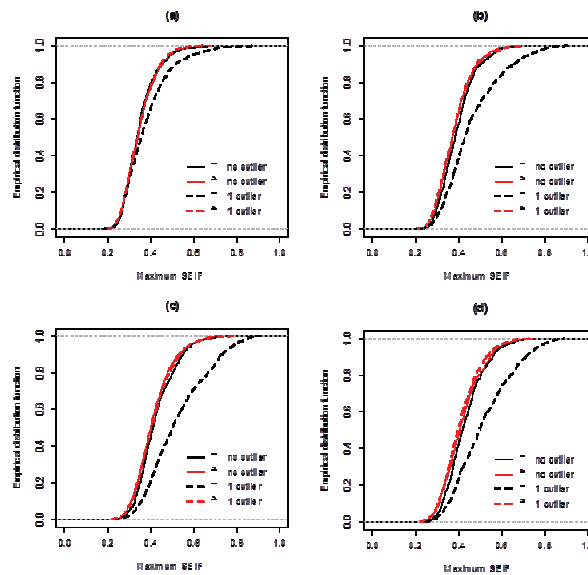


図 2: との分布



結果の図 2 (a)-(d)では、外れ値の有無に関わらず、2つの $\bar{T}$ の分布が近いので、提案する手法ははずれ値に頑健であることが分かる。また、 $\bar{T}$ の分布ははずれ値がないときの $T$ の分布に近いことから、偽陽率がコントロールされ、はずれ値があるときに検出力があることが分かる。さらに、4つのモデルの結果は同じような傾向を示しているため、提案手法は真のモデルに依存することが少ない安定した手法であるといえる。

### サブテーマ 3

「大量で多角的なデータの情報・統計手法を適用したゲノム機能と遺伝的ネットワーク抽出」

複雑形質（生体内構造、生化学データ、骨格形態、行動パターンなど）の表現型計測を行いデータを拡充した。汎用実験系統、野生由来系統、さらにはそれらの交配集団や2系統間染色体置換（コンソミック）系統群等を対象としたデータ解析をおこなった。特に、マウスゲノム多型性と遺伝子発現情報を利用して、対立遺伝子による遺伝子発現量の差を制御する遺伝機構を解析するため、各種交配系統の整備および次世代シーケンサーによる対立遺伝子発現データの収集を開始した。

#### 3-1：次世代シーケンサーによる対立遺伝子発現バイアスの遺伝的制御機構の解明（豊田、藤山、阿部、中谷、藤澤、栗木、二宮、原、井上、宋、岡、高田、城石）

本研究はゲノムのリファレンス配列が整備されたマウス系統の遺伝子多型情報と、計画的交配による個体群から得られた組織の遺伝子発現情報を利用して対立遺伝子による遺伝子発現量の差を制御する遺伝機構を解析することである。我々が使用するマウス C57BL/6J(B6)および MSM 系統間には膨大な cSNP (26,000 の non-synonymous 塩基置換を含む) があるため、哺乳類でゲノム上の全遺伝子について次世代シーケンサーを用いた RNA-Seq 法で両親由来のアリル発現のバイアス（極端な場合にはインプリンティング現象）のプロファイリングが可能となる。世界的に見てもこのような研究は黎明期にあり、しかも使用するマウス系統は国立遺伝学研究所が樹立し塩基配列が明らかな日本産野生由来 MSM/Ms 系統や JF1/Ms 系統、およびマウスで最初にゲノム配列が明らかにされた汎用実験用系統である西ヨーロッパ産由来の B6 系統の交配により得られる F1 個体を当面の対象としている（図）。

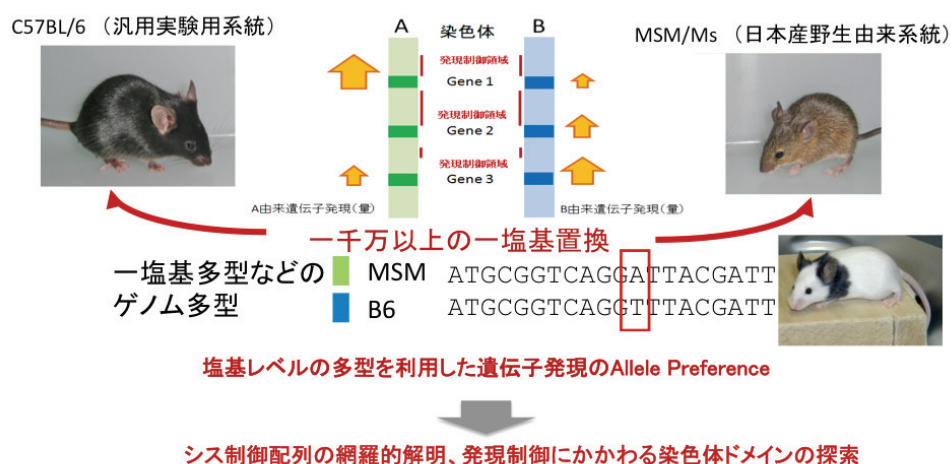


図 C57BL/6、MSM/MsおよびJF1/Msのゲノム情報と表現型情報を利用した次世代シーケンサーによる対立遺伝子発現バイアスの遺伝的制御機構の解明

本年度は平成 22 年度より開始している (B6xMSM) F1 および(MSMxB6)F1 個体の交配・生産を継続するとともに、新たに MSM 系統と同じ *molossinus* 亜種由来の愛玩用系統 JF1 系統を使用した (B6xJF1) F1 および(JF1xB6)F1 個体についても同様に交配・生産と表現型データの収集を開始した。これらより得られる発現解析の結果も利用することで、*molossinus* 亜種の持つ B6 系統に対する広範な遺伝子発現制御の解析が行えるものとする。

これまでの B6、MSM および JF1 各系統による交配により、(B6xMSM) F1: 140 匹(雄 72 および雌 68 匹)、(MSMxB6)F1 : 67 匹 (雄 35 および雌 32 匹)、(B6xJF1) F1: 146 匹(雄 67 および雌 79 匹)、(JF1xB6)F1 : 60 匹 (雄 28 および雌 32 匹) の 400 匹以上の個体から表現型データおよび RNA 収集用の臓器サンプルを収集することができた。対象とする項目は前年度の報告書に記載済みである。なお、表現型データに関しては、これまでと同様に通常餌 (CE-2, 346.8kcal/100g, 4.6% from fat)、10 週間飼育しその後解剖した。個体は離乳より雄は単飼、雌は最大 5 匹までの群飼により維持した。測定週令に達した個体は、体重、体長、対象臓器を計測しサンプルの収集を行った。遺伝子発現解析に使用する試料も同様に 10 種類 (大脳、小脳、肝臓、腎臓、心臓、胸線、膵臓、脾臓、白色脂肪、褐色脂肪、精巣) を対象とし保存を行った。今年度を持って B6、MSM および JF1 による交配系統の表現型収集は無事に終了することができた。今後は、これらデータを使用して、各系統の組み合わせが表現型にどのように影響を与えるかについて、サブテーマ 2 と連携して統計解析を行い、遺伝子発現解析との相関解析を解析することを目指している。さらに一部の個体より得られた RNA については、サブテーマ 1 との連携により、HiSeq2000 をもちいた直接塩基配列読取により遺伝子発現のデータ取得を開始した。次年度はこれらのデータを使用した遺伝子発現解析に着手する。

(B6 x MSM)

### 3-2 : 大規模実験交配集団を対象とした遺伝的構造 (主にエピスタシス) の統計解析 (栗木、岡、高田)

これまで行われてきた動物モデルの F2 解析では、せいぜい数百匹の組み替え体による連鎖解析が限界であったので、エピスタシスの検出が困難であった。本テーマは、B6 および MSM を基盤として、3000 匹を目的とした F2 組み替え体と、廉価で高速な SNP タイピングを組み合わせ、エピスタシスの検出と同様に挑戦することを目的としている。今年度までに 2,500 個体 (雄) 以上の形態表現型データを収集し、一部のデータについては解析を開始した (図 1)。全てのサンプルについて、ゲノム DNA のアーカイブ化が進行中であり、アーカイブの約 600 の DNA については、本プロジェクトにより構築した B6-MSM 間の MassARRAY による 200 座位近い高速 SNP タイピングにより SNP タイピングが終了している。パイロット解析として約 500 個体分の Genotyping を行い、既存の QTL 解析ソフトによりエピスタシスの検出を試み、複数の候補座位の存在を明らかにした (図 2)。今後これらの結果を使用し、サブテーマ 2 と連携を取りながらエピスタシスについて検出および検証可能な統計的手法を試行し、本テーマの目的とする 3,000 超の各種データに適応させる。

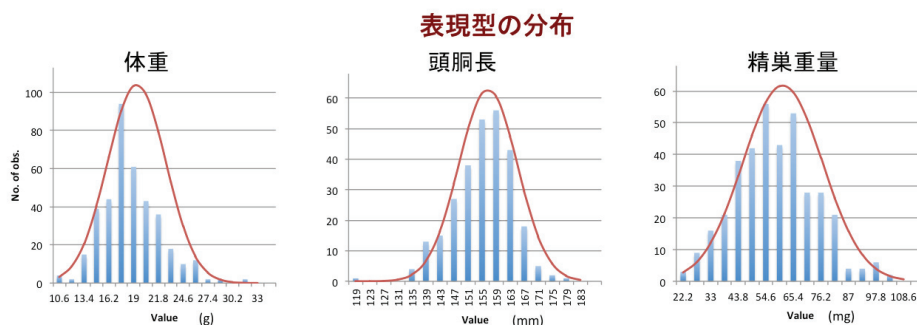


図1 エピスタシス検出のためのF2 ((♀B6x♂MSM)F1 x (♀B6x♂MSM)F1)解析群の表現型分布(2,000 超の体重、頭胴長および精巣重量のデータ)

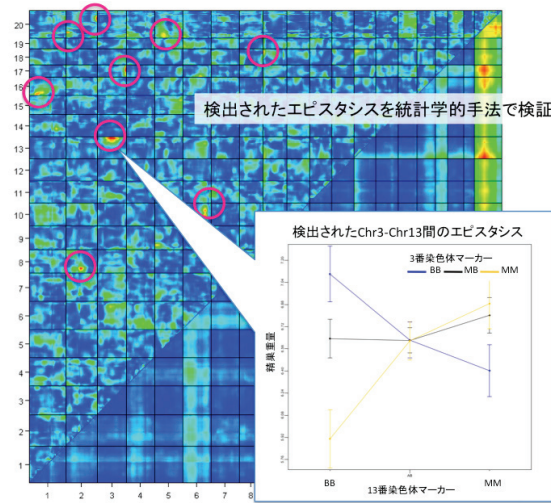


図2 既存のソフトウェアによる遺伝子座間のQTL解析を行った結果、複数のエピスタシス制御の存在が示唆された(マゼンダ色丸囲み)。縦軸および横軸に染色体およびマーカーが配置されている。ロッドスコアは青(低)一赤(高)で示される。これらの効果について、False positiveおよびFalse negativeを検出できるか検討する。

### 3-3：隠れマルコフによるマウスの社会行動状態の自動抽出と、マルコフ空間による系統比較（土谷、荒川、柿原、高橋阿貴、梅森、杉本、田邊、小出）

本研究の目的は、これまでは人による観察法が中心であった社会行動解析に関して、隠れマルコフモデルによるコンピュータ学習により自動推定を可能とし、大量社会行動テスト記録画像から、有用な行動データを自動抽出することを目的とする。

社会行動は多くの哺乳類において観察され、その動物の適応に大きく関わっている。人間社会においても、社会的な相互作用がよりよい生活を行う上で重要な意味を持っている。社会行動の疾患である自閉症は、比較的高い発症率（約 1%）が報告されており、その治療のためにも社会行動の生物学的基盤の理解が求められている。社会行動の遺伝学的・生物学的基盤を調べる上で、マウスがモデル動物としてよく用いられている。

これまでに、国立遺伝学研究所で行ったソーシャルインタラクションテスト（図 1）の結果をもとに、2 個体の動きを隠れマルコフモデルにより自動推定する手法の確立を進めてきた。ソーシャルインタラクションテストとは、2 個体のマウス（同系統・同性）を新奇なオープンフィールドに入れ、そこで生ずる社会的な相互作用を 10 分間観察するものである。このテスト場面では、様々な社会行動が出現し（例えば相互におい嗅ぎ、追隨行動、攻撃行動など）、それらを詳細に解析するうえでは、実験者が各行動の出現を各時間において記録するという観察法が用いられる。しかし、この方法は詳細であるというメリットの一方でかなり時間がかかり、ノックアウトマウスのスクリーニングや、量的遺伝子座の同定（QTL 解析）など大量解析においては実用的ではない。そこで、隠れマルコフモデル（HMM）を用いて、実験者による観察法から得られる結果を自動解析による位置データのみから推測することを試みた。



図 1. ソーシャルインタラクションテスト

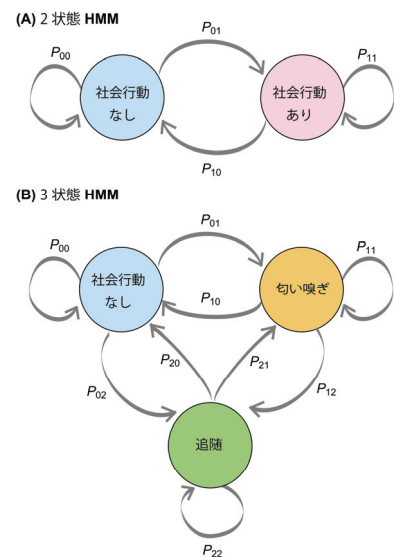


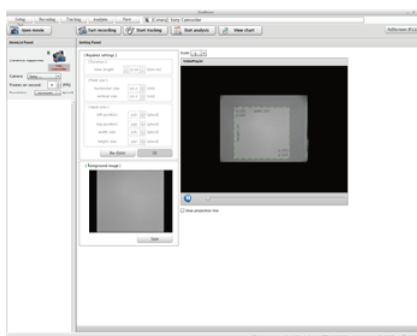
図 2. 隠れマルコフモデル(HMM)

様々な系統を含む 50 ペアのマウスの 10 分間の相互社会行動について、人間による観察（社会的におい嗅ぎ、性器嗅ぎ、追隨）と、自動解析ソフトによるマウスの位置のトラッキング（1 秒間に 3 フレームの位置情報を取得、合計 1800 フレーム）を行った。マウスの位置データから、4つの指標（2 個体間の距離、相対速度、相対角度、平均速度）を算出し、それらのヒストグラム分布と人間の観察による行動状態を HMM に学習させた。本研究では 2 つの HMM（2 状態 HMM と 3 状態 HMM）を用いて行動状態を推測した（図 2）。2 状態 HMM では、社会行動の出現の有無の 2 状態を抽出した。3 状態 HMM では、社会行動の中でも「におい嗅ぎ」と「追隨」を分けて抽出した。いずれのモデル推定値も人間による観察結果とかなり高い一致率を示し、2 状態 HMM では 0.983 の相関、3 状態 HMM ではにおい嗅ぎで 0.976、追隨で 0.750 の相関が認められたことから、ある程度信頼性の高い推測が行うことができたといえる。

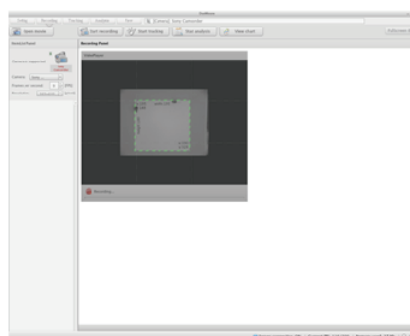
これらの一連の解析をより効率よく行うために、データ取得からトラッキング解析、行動の自動推定、さらには解析結果の表示まですべてを行うことが可能なソフトウェア(DuoMouse)を開発した。このソフトウェアは、Adobe 社の Flash action script と呼ばれるマルチプラットフォーム言語を用いて開発しており、Windows、Mac、Linux のどの OS でも実行することが可能である。

このソフトウェアは、本共同研究の成果を論文として発表後、ホームページより自由にダウンロードできるものとして公開する予定である。

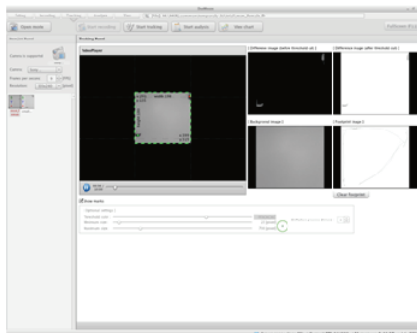
セッティング画面



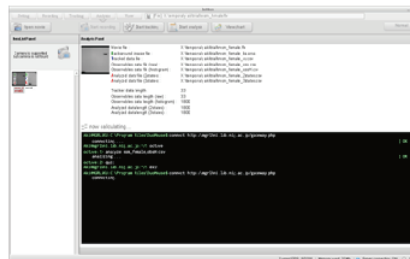
レコーディング画面



トラッキング画面



自動解析画面



解析結果表示画面

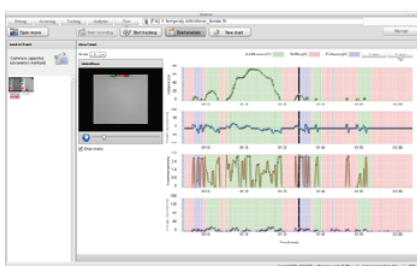


図3 社会行動自動解析ソフトウェア(DuoMouse)の各解析ステップの画面

ここまでの解析において、動物の社会行動は、人間が「動物の社会的行動」を定義し、動画像を見てどのような社会的行動を行っているかを観察して解析している。

一方、マウス行動を「局所的に観察して一時点での行動を社会的行動として判定して捉える」と、  
「マウス行動をより長い期間経時的に観察した時の全体的な印象としての社会性の判定」、は別である。  
両者の関係は、認知科学的に興味深い問題であると考えられる。本研究においてはこの問題に取り組み以下の結果を得た。社会性は概して MSM マウスが高いとされているので、動画像を始めから終わりまで通して観察して感じた社会性の度合いを簡単に MSM 度と呼ぶ。ここでは、21 名の被験者に 59 種類のマウスペアの画像を見せて、MSM 度を 0 点から 100 点の間で得点化してもらい、各画像について、21 名の平均点を、その画像（マウスペア）の MSM 度と定義した。そして、MSM 度を、動画像の特徴量を用いて説明することを試みた結果、

MSM 度 = 1235.74 × (無関心から社会的行動への遷移確率)

− 395.84 × (社会的行動から無関心への遷移確率) + 71.02

という簡単な線形式で MSM 度が良く説明できることが分かった (図 4 左側)。ここで、遷移確率は、隠れマルコフモデルを用いて計算されたものである。このモデルは、MSM 度を、動画像に関する典型的な特徴量である「2 匹のマウスが社会的状態にある確率 (比率)」を説明変数として説明する (図 4 右側) よりもあてはまりがよく、人間がマウスの社会性に関する印象を形成する際には、大域的に捉えるよりも、局所的な認識を繰り返し脳内に刷り込んでいることを示唆している。

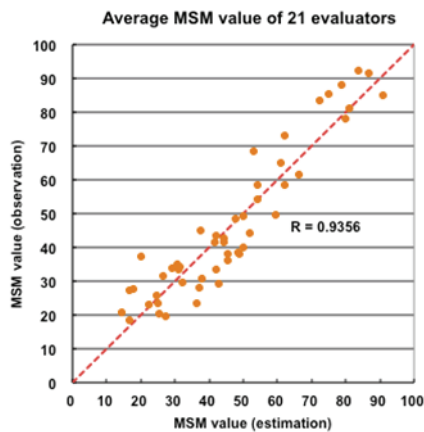
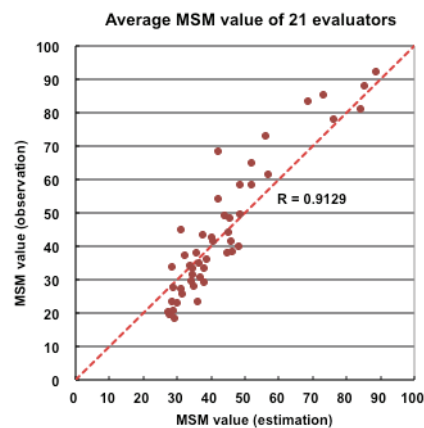


図 4 推定 MSM 度と観測 MSM 度の散布図  
(遷移確率による推定)



推定 MSM 度と観測 MSM 度の散布図  
(定常確率による推定)

### 3-4 : マウスの超音波発声データの解析 (Dou Xiaoling、白旗、栗木、杉本、高橋、小出)

音声データは、時間、周波数と信号強度の 3 次元で観測されるものであり複雑な情報を含む。更に、周囲のノイズも含まれるため、実際にその音声パターンの抽出・解析を行なうには困難が伴う。本研究では、マウスの超音波発声パターンの抽出を行うための、適切な分析手法の確立を目的とした。

マウスの超音波発声のデータは時間、周波数と強度の 3 次元で観測されている。そのうち、強度が声の大きさを表すため、決めた強度の以上のもとで、超音波データの変化パターンを時間と周波数の 2 次元に縮約し、ノイズを取り除き、曲線として解析するための予備的な検討を進めた。また、得られた曲線の断片を関数化し、関数のクラスター分析を行う方法も検討した。

具体的な研究手法と解析計画は以下のようなになる：

Step1. 3 次元データ (周波数、時間に対する信号強度データ) について、移動平均の方法でノイズを除去する。

Step2. 信号強度が強い時間を切り出し（音声波形切り出し）、さらに各時点について最も信号強度が大きい周波数（ピーク周波数）を推定し、データを関数データ（時間に対するピーク周波数）の形に縮約する。（検討している手法は B-spline 基底関数法）

Step3. Step2 で得られた関数データ（音声波形データ）をクラスタリングする。

特に、Step2 では最小信号強度、音声波形の間隔、音声波形の長さの最小値および B-spline 基底関数の次数などをパラメータとして推定する必要がある。Step 3 ではクラスターの数を決める方法も考えなければならない。今後主にこれらのパラメータを推定する方法の確立を目指す。

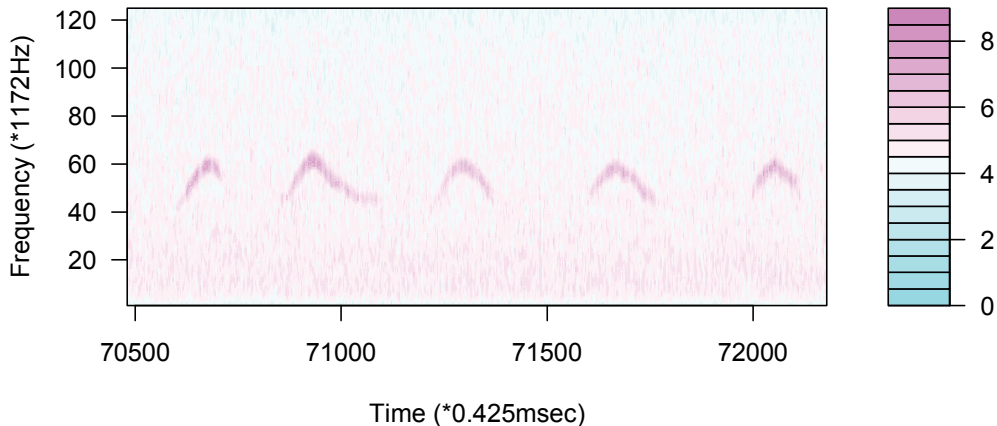


図1. FFTして得られたマウス(BALB/cAnN)超音波のデータの一部

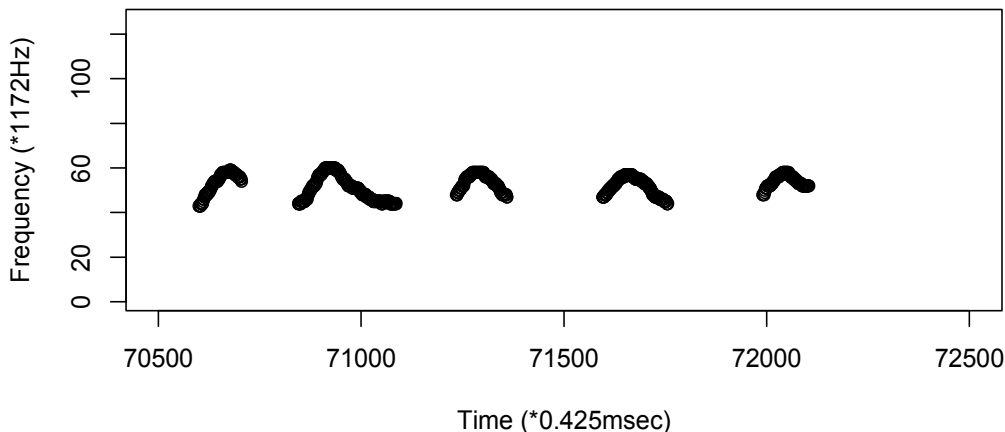


図2. 移動平均を取り、ピーク周波数を推定し、得られた超音波発声の波形

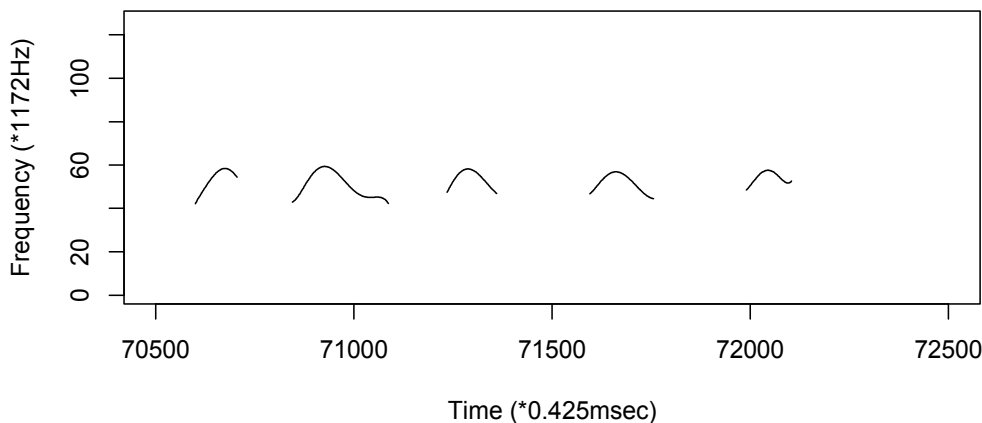


図3. 関数として得られた音声の波形

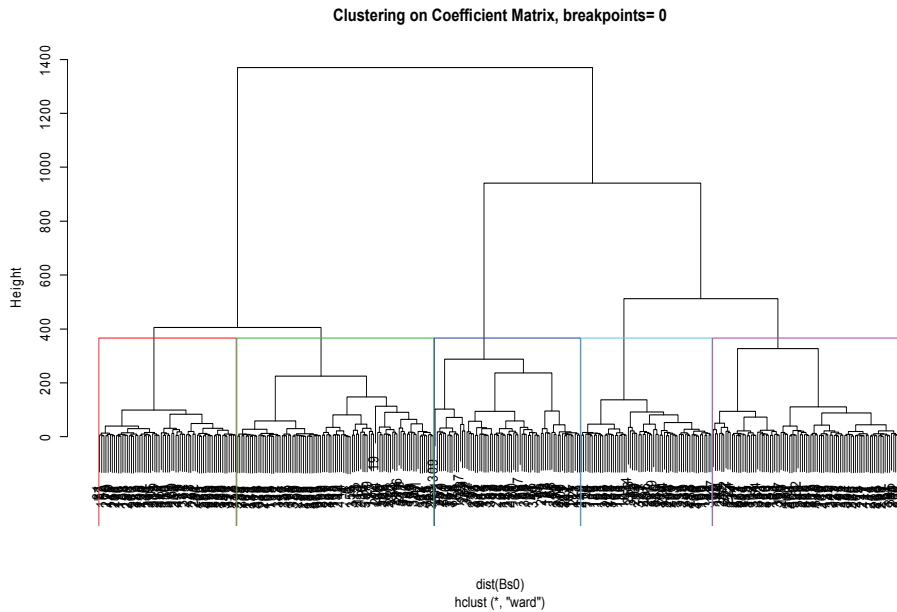


図 4. 関数データの階層的クラスタリング

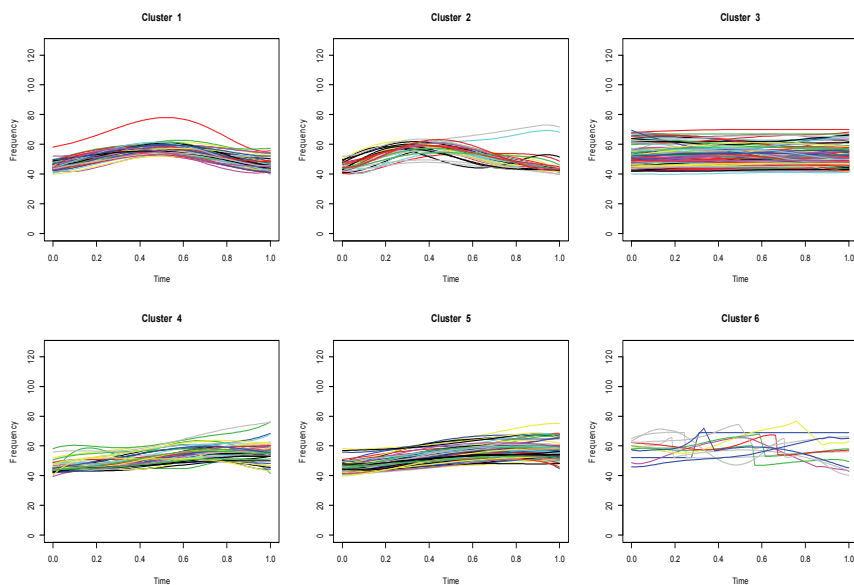


図 5. K-means 法による BALB/cAnN の超音波発声データの分類

イネについては、多様な遺伝形質を持つ野生イネのゲノム構造、発現遺伝子の構造及び発現量差の解析（eQTL 解析を含む）を行うのに必要な基本情報を取得し、順次マッピング、アセンブル、比較解析などの解析集団の作製方法を検討した。

3-5 : 遺伝子型-表現型相関解析のための野生イネ遺伝型構造解析と表現型解析の基盤 (久保、藤山、豊田、中村、神沼、望月、倉田、海外共同)

野生イネ系統群を用いたゲノムと遺伝型解析および表現型解析を基盤として、遺伝型と表現型のアソシエーション解析を行うため、H23年度は以下の内容のサンプリング、次世代シーケンスによる解読と基本系統のゲノム配列解析、系統間比較による遺伝型推定、系統間進化関係の解析を行った。

図3-5. 次世代シーケンサーのW0106, W1921, W1943, Guangluai-4, Nongken-58のnipponbareゲノムへのマッピング

Chromosome	Length of IRGSPb5 (bp)	Map rate ( MappedReads / All Reads % ) , Number of Mapped Reads					Coverage (%)					Depth average ( Depth / Length )				
		W0106	W1921	W1943	Guangluai-4	Nongken-58	W0106	W1921	W1943	Guangluai-4	Nongken-58	W0106	W1921	W1943	Guangluai-4	Nongken-58
		1	45038604	9.0%	8.8%	11.0%	10.3%	11.1%	87.7%	87.4%	91.3%	88.4%	93.1%	12.1	11.0	14.5
2	36792247	7.369561	7.3%	9.0%	8.6%	9.3%	89.3%	89.6%	93.6%	90.7%	94.8%	12.4	11.3	14.7	14.7	12.1
3	37312367	6.154260	5.823827	5.503489	8.577217	11.313320	8.0%	7.8%	8.5%	9.0%	8.8%	92.6%	92.3%	92.4%	93.1%	94.9%
4	36060865	6.522752	6.227849	5.226117	8.978284	10.767674	8.4%	8.3%	9.7%	10.5%	11.0%	84.0%	84.3%	90.3%	88.4%	95.1%
5	30073438	6.863760	6.583178	5.939467	10.495248	13.370349	6.6%	6.9%	8.3%	8.2%	7.3%	91.8%	91.8%	95.3%	94.4%	96.6%
6	32124789	5.386468	5.463114	5.079882	8.167914	8.880676	6.4%	6.7%	8.3%	7.6%	8.0%	86.4%	86.5%	91.1%	89.2%	94.2%
7	30357780	5.5273797	5.329191	5.084895	7.622587	9.737444	5.9%	5.9%	6.8%	7.0%	7.2%	85.2%	85.8%	87.8%	86.8%	95.1%
8	28530027	6.4843303	4.712181	4.141353	7.024546	8.807635	6.2%	6.1%	7.8%	7.2%	7.3%	88.6%	88.4%	95.7%	90.3%	97.0%
9	23895721	5.5097287	4.852871	4.764902	7.176549	8.900338	5.7%	5.8%	6.1%	6.1%	6.3%	83.9%	83.6%	88.0%	86.3%	92.7%
10	23703430	4.656704	4.571614	3.723909	6.099332	7.621019	6.5%	6.7%	6.5%	7.5%	7.9%	84.8%	86.0%	88.9%	87.6%	93.4%
11	31219694	5.5290521	5.339186	3.958432	7.495916	9.646737	5.2%	5.3%	6.5%	6.3%	6.8%	76.7%	77.5%	82.7%	78.6%	87.1%
12	27679166	4.4271928	4.217471	3.976976	6.305766	8.262777	5.7%	5.7%	7.2%	7.0%	8.0%	83.0%	83.3%	90.0%	85.0%	94.6%
Total or Average	382788128	81.21%	81.38%	95.70%	95.10%	99.05%	86.4%	86.6%	90.7%	88.4%	94.1%	12.6	11.7	14.8	15.3	12.3
Num of hit Reads		66407676	64628606	58577890	95243684	120639210										
Num of all Reads		81773774	79418458	61209112	100150796	121790258										

まず栽培イネの直接の祖先種である *O. rufipogon* の3種類の系統 (W0106, W192, W1943)を次世代シーケンサーでゲノムの x11~x15 倍の厚さで読み、レファレンスゲノムである *O. sativa japonica*, *Nipponbare* ゲノムにマッピングした。また栽培イネの *japonica* (Nongken-58) 及び *indica* (Guangluai-4)のゲノムリードデータを入手し、これら2系統もマッピングし、5系統は各々86-94%のゲノムをカバーでき、ゲノムの大半を再構成できた (表 3-5)。

*Nipponbare* と合わせ、栽培イネ3系統、野生イネ3系統のゲノム間で SNP と Indel の比較解析を行った。これらの過程で、種々のシステムの構築や、ソフトウェアの利用、解析方法等を精査し、現時点での最適なゲノム再構築法と問題点を抽出した。また、SNP を用いた構造比較解析によって、6系統間の Hapmap を作成し、詳細な構造比較と6系統の類似性、ゲノム断片の持ち込みなどの状態が明らかとなった (図 3-5 に一部を表示)。この結果は、現在論文作成中である (Ohyanagi et al.)。

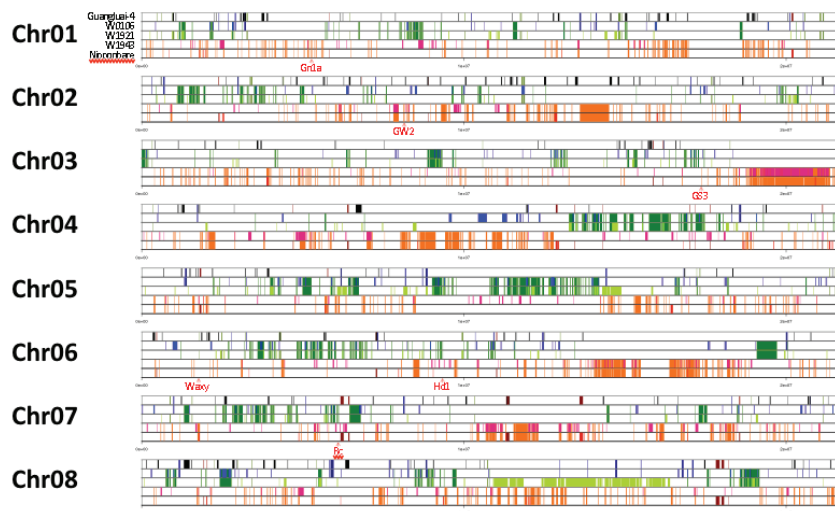


図 3-5. 染色体 1-12 の Hapmap の一部を表示

さらに、中国との共同研究により、400 系統以上の *O. rufipogon* を系統毎に薄く読み (x2-x3)、予測 Hapmap を構築するアルゴリズムを用いて、全系統のゲノム再構築を試みた。再構築ゲノム配列の比較を行い、系統間の類縁関係、進化関係を予測した。各系統のアジアに置ける原産地域との関連から、栽培イネの起源までをたどる事が出来た。さらに、H23 年度に計測できた草型 (開張度) の表現型に関する



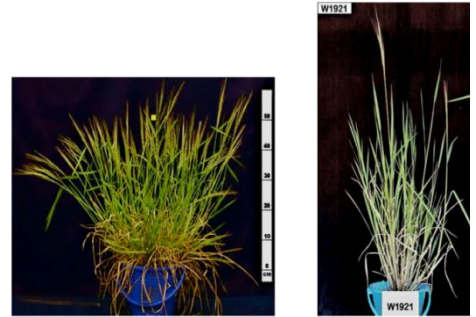
るアソシエーション解析の結果も解析が進めたが、現在論文準備中のため詳細は H24 年度の成果での報告とする。

3-6 : eQTL 解析による表現形質と遺伝子発現のネットワークの解明 (堀内、春島、藤澤、久保、倉田)

栽培イネ *Oryza sativa* の祖先種である野生イネ *O. rufipogon* は、栽培イネより多様な遺伝的形質を示し、生態、形質調査から数種類の集団構造が存在することが示唆されている。第 I 期融合研究プロジェクトで推進した。 *O. rufipogon* の量的形質調査からも、柱頭長、種子稔性、穂の形 (着粒数と相関あり)、発根・再生力など多く形質が多年生系統群、一年生系統群で大きく2つに分かれることが示唆されている。

1 年生型 W0106

多年生型 W1



栽培イネ *Oryza sativa* の祖先種である野生イネ *O. rufipogon* は、栽培イネより多様な遺伝的形質を示し、生態、形質調査から数種類の集団構造が存在することが示唆されている。第 I 期融合研究プロジェクトで推進した。 *O. rufipogon* の量的形質調査からも、柱頭長、種子稔性、穂の形 (着粒数と相関あり)、発根・再生力など多く形質が多年生系統群、一年生系統群で大きく2つに分かれることが示唆されている。

図 1. *O. rufipogon* の葯(A-D)、雌蕊(E-H)、挿木(I-L)。1年生型の W0106(A,E,I)、W0610(B,F,J)。多年生型の W1945(C,G,K)、W1230(D,H,L)

野生イネにおいても集団構造を持つため、様々な品種を用いたアソシエーション解析による表現型及び遺伝子発現量の遺伝解析は困難を伴う。そこで本研究は表現型の離れた一年生、多年生グループから代表的な2系統 W0106 と W1921 を選抜し、この交雑後代の実験集団を量的な表現型の遺伝形質の遺伝解析 (QTL 解析) と遺伝子発現量の遺伝解析 (eQTL 解析) に用いることとし、表現形質を形成するゲノム上の遺伝因子群の特定とその遺伝子発現のネットワークの解明を目指している。

22 年度には、解像度の高い解析集団として組換え固定化集団を育成するため、異なる F<sub>2</sub> 個体どうしを交配して F<sub>3</sub> 他殖系統群を得るため、多数の F<sub>2</sub> 種子を作成した。交配した F<sub>3</sub> 他殖系統を自殖させ、自殖 F<sub>4</sub> 系統から葯培養による倍加半数体系統(DHLs : doubled haploid line)の育成を行うため DHLs 200 系統を用いると集団の分解能は 0.22cM(平均物理距離 56kb)と計算される (図 1)。

H23 年度は実際に異なる F<sub>2</sub> 314 個体を交雑し 157 の F<sub>3</sub> 他殖系統を取得し、冬期にインキュベーターを用いて 72 系統分の自殖 F<sub>4</sub> 種子を獲得した。H24 年度も引き続き自殖 F<sub>4</sub> 種子取得に向けた交雑、育成を行うと同時に、葯培養による DH 化検討実験と DHLs 育成を行う。

これらの系統を用いた e-QTL 解析のため、ゲノム塩基配列も未知で、遺伝子のアノテーションもない野生イネ品種の大量の RNA-seq のデータをどのように処理し、遺伝子発現量を推定するかを検討している。3-5 で示したように、H22 年度から H23 年度に W106, W1921 のゲノム配列情報を再構築するため、次世代シーケンサーよりゲノムデータを取得、マッピングし、Nipponbare ゲノム配列にマップして各品種のゲノム配列の塩基配列多型、挿入欠失配列等を調査した。また H23 年度は W106, w1921 の RNA-seq のデータを取得し、Nipponbare ゲノム配列にマップし日本晴の RNA-seq のデータをマップしたものとの比較を進めた。H24 年度以降、これらの配列変異情報を基盤に両ゲノム間の発現比較等を行う予定である。

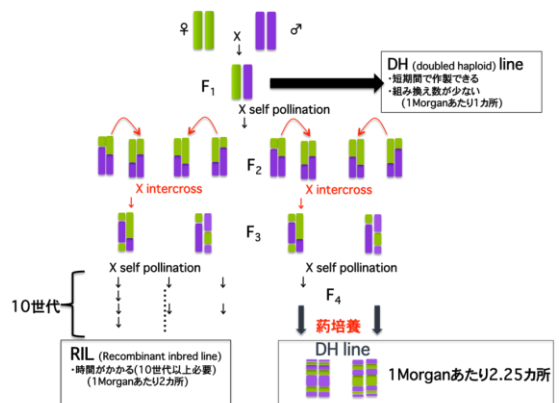


図 1 様々な組換え固定化集団による解像度の比較

### 3-7：構造多型を考慮した発現解析手法を用いた遺伝子発現量差の解析（堀内、春島、藤澤、久保、倉田）

生物の形質や形態の多様性は発現する遺伝子の塩基配列構造や遺伝子の発現量の差によりもたらされる。本研究は栽培イネのジャポニカとインディカの亜種間において各組織でどの程度の遺伝子が高発現し、どの程度の遺伝子の発現量が異なるのかを Affymetrix 社 Rice Genome Array を用いて解析を行った。また発現量が異なる遺伝子についてその組織間における発現変動の特徴や遺伝子進化との関連性を調査した。Rice Genome Array はイネのゲノム配列が決定される以前に設計されたものであったので、array の約 63 万のプロープのうちジャポニカ Nipponbare で想定されている全ての遺伝子が発現しても特定の遺伝子にのみハイブリダイゼーションする 51 万プロープのみを用い、計 44,447 遺伝子を発現差解析の解析対象とした。またイネ 2 亜種の発現データとして公的データベース GEO よりジャポニカ (Nipponbare)、インディカ 3 品種 (Zhenshan97, Minghui63, 9311) 合わせて 321 アレイデータ (121 組織) を取得し、融合研究プロジェクト第 I 期により開発した遺伝子の塩基多型に影響されない発現量推定法を用い解析した。まず、第 I 期プロジェクトで取得した Nipponbare と 93-11 の葉と幼穂の array データよりいずれの組織でも高発現しかつ 2 品種間で遺伝子発現量に差のない 487 遺伝子を用い全 321 アレイデータのクラスタリング解析を行った。アレイデータは類似組織毎にクラスタリングされ、胚乳、成熟蒴、幼穂、幼葉、根の計 5 組織の類似アレイのクラスターにはジャポニカとインディカ由来の品種がともに含まれていた。この同一クラスターに含まれるアレイデータは同じ組織のものとしてジャポニカとインディカ間の発現差解析に用いた。5 組織それぞれについて、Nipponbare とインディカ 2 品種間の発現解析を行った。ジャポニカ 5 組織で高い発現がみられたのは全解析遺伝子 44447 個のうち 13433 ~ 15173 個 (30~34%) で、発現遺伝子数は組織間で類似していた。一方、ジャポニカ、インディカ間で有意に発現差がある遺伝子を 5 組織から抽出すると、発現差がある遺伝子数は 169~967 個と用いる組織によりばらついていた。次に、各組織で有意に発現差を示す遺伝子の発現特性を 121 組織データのクラスタリング解析により調べた。ジャポニカとインディカの遺伝子発現差を調べたどの組織データでも、ジャポニカとインディカで発現差のあった遺伝子で 321 のアレイデータをクラスタリングすると、アレイデータは組織毎ではなく亜種毎に大きなクラスターを作る。また発現差のあった遺伝子群は、その遺伝子が特定の組織でのみ発現し、発現差も組織特異性がある遺伝子群 (A) と、一方の亜種では全組織で高発現するが、もう一方の亜種では発現がない遺伝子群 (B: 亜種特異的に発現する遺伝子) が観測された (図 1)。

また各組織データで確認された (A) と (B) の遺伝子群を 5 組織間で比較したところ、(A) は検出された組織でのみ特異的に発現差がみられていた。また、生殖関連組織ほど (B) の品種特異的に発現する遺伝子の割合が多かった (図 2)。

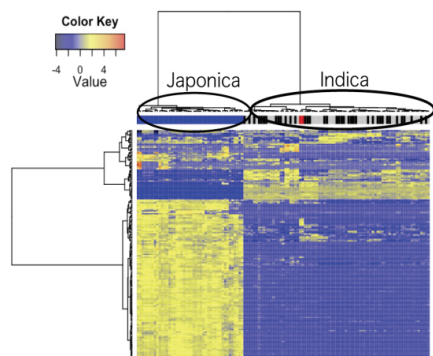


図1 受精後5-10dの胚乳においてジャポニカ、インディカ間で有意に発現差を示した169遺伝子の121組織発現ヒートマップ  
両品種共にいずれかの組織で発現がある遺伝子群(A)と、一方の品種の全組織で高発現するが、もう一方の品種の全組織で発現がない遺伝子群(B)が観測された。  
Bの発現パターンに従い121アレイデータはジャポニカとインディカの2つのクラスターに分類された。

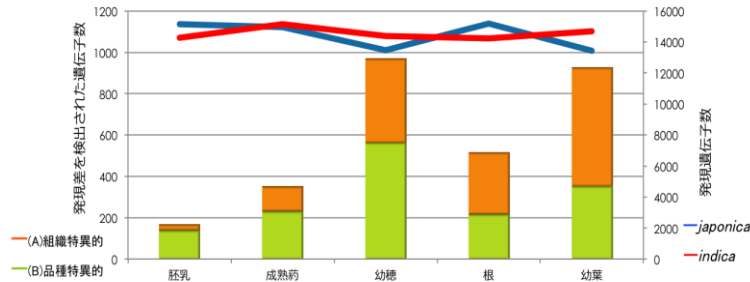


図2 ジャポニカ-インディカ間で発現差示す遺伝子と発現遺伝子数

5 組織からは計 1404 個のジャポニカがインディカより有意に発現差が高い遺伝子が抽出された。この内、40%は(B)ジャポニカの全組織で高発現を示すがインディカで発現がみられない亜種特異的に発現する遺伝子群であり、残り 60%は(A)特定の組織でのみ発現差が観測された遺伝子群であった。次に、ジャポニカ-インディカ間の発現多型と遺伝子進化速度の相関性を調べるため、インディカ 93-11 ゲノムの遺伝子アノテーションより相同遺伝子を検索した。亜種間で発現差がありジャポニカで高発現する遺伝子のうち(B)に属する遺伝子群はインディカで相同遺伝子が見つかる割合が低く、相同遺伝子が検出されても、Ka/Ks 値が高く進化速度の速い遺伝子の割合が遺伝子発現に亜種間で差がないもの、または(A)群に属する遺伝子群と比較し有意に多かった。遺伝子発現量差に組織特異的に現れる遺伝子群(A)は相同性遺伝子がインディカゲノムで検出される割合も、遺伝子進化速度が早い遺伝子の数の割合もジャポニカ-インディカ間で遺伝子発現差の変わらないものと差はなかった。gene ontology 解析でも(A)群と(B)群の遺伝子は差があり、(B)には病害抵抗性関連遺伝子が多く確認され、一方(A)群には転写因子関連が多く確認されていた。以上より、ジャポニカ-インディカ間において発現差を示す遺伝子は A: 特定の組織でのみ発現差を示す遺伝子群と B: 亜種特異的に発現する遺伝子の 2 種類に大きく分類され、A)と B)では遺伝子進化速度や遺伝子発現制御機構、さらに遺伝子機能自体にも違いがあることが示唆された。この結果は H24 年度に論文投稿を行う予定である。

### 3-8: ゼブラフィッシュの多様な表現型の抽出とその表出法の確立

トランスポゾンランダムな挿入に基づくトランスジェニックシステム、ノックアウト変異システムの網羅的な開発とそれらに基づく多様な表現型の抽出とその表出法の確立のための研究を行った。また発現パターンの画像解析、数理解析等の研究を進めた。図1に概略を示す。それぞれについて以下に述べる。

(1) モデル脊椎動物ゼブラフィッシュにおいて、トランスポゾンを用いた遺伝子トラップ・エンハンサートラップスクリーンを実施し、Gal4を細胞・組織・器官特異的に発現するトランスジェニックゼブラフィッシュを新たに 100 系統作製し、合計 1000 系統以上とした。これらを用いて、国内外の器官形成研究者と研究の融合をはかった。名古屋市立大学のトランスジェニックフィッシュシステムを用いて、新生ニューロンの嗅球への移動の研究を行い、その経路を明らかにした (Kishimoto et al. 2011)。名古屋大学伊藤博士とは、トランスポゾン挿入によって作製された deltaA 遺伝子変異を利用して側線神経細胞分化研究を展開した (Mizoguchi et al. 2011)。このように融合研

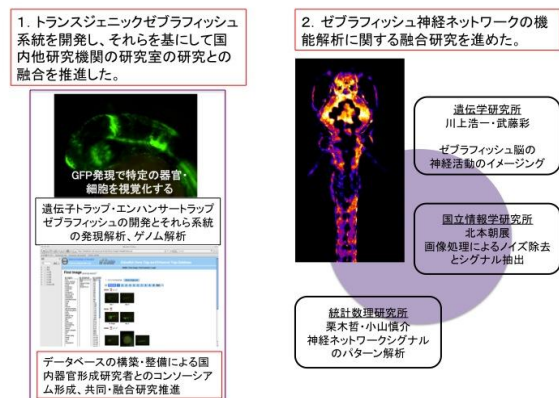


図1. ゼブラフィッシュの多様な表現型の抽出とその表出法の確立のための研究の概略

究は順調に進捗している。

(2) 新しい表現型抽出法としてゼブラフィッシュの神経活動を、カルシウムインディケータ GCaMP を用いてイメージングするシステムの開発に成功した (Muto et al. 2011)。脊椎の運動神経、脳内の神経活動の時空間パターンをイメージ動画から自動的にデータ取得するための画像解析プログラムの開発を情報研の北本博士と共同で行い、脊椎の運動神経の発火パターンをデータ取得することに成功した (図 2)。

さらに、このようにして得られた時空間パターンを数理的に解析するために、統数研栗木博士との共同研究を進めている。

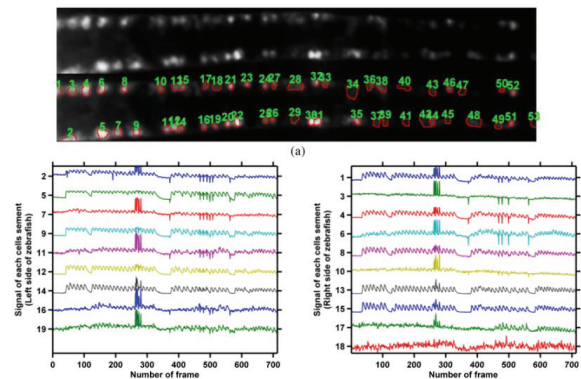


図2. 神経活動をカルシウムイメージングした動画からのデータの取得プログラムの開発。神経活動の発火のイメージングおよび解析すべき領域の自動設定(上)。各々の解析領域におけるシグナルの経時変化(下)。

### 3-9 : ショウジョウバエ翅形態異常のゲノム相関解析

ショウジョウバエのゲノムは 15,000 個の遺伝子をコードしている。これらの 3 割程度は個体の発生に重要な役割を果たしており、突然変異で致死となることが知られている。残りの遺伝子はこれらの補佐的役割を担い、システムの堅固さを保証しているのではないかと考えられるが、必ずしも明確な結論は得られていない。そこで本研究では、遺伝研に維持されている 7,100 遺伝子の突然変異体 (RNAi 変異体) の表現型を詳細に解析することによりこの問題にアプローチする。

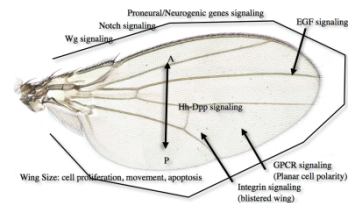


図1. ショウジョウバエの翅形成に働くシグナル伝達系

ショウジョウバエの翅は薄いシート状の構造物である。この翅は体の他の器官と同様に、それぞれ多くの遺伝子が関与する幾つかのシグナル伝達系が互いに協調して働くことによって作られる (図 1)。それぞれのシグナル伝達系は翅の前後軸方向あるいは端部一基部方向の増殖や、翅脈の形成等を支配するため、それらの変異体では特徴的な形態異常が観察される。従って、微かな形態異常を検出することができれば上記の補佐的な遺伝子の機能も同定できるかもしれない。

H23 年度は RNAi 系統を使って翅に遺伝子変異を誘導し、その画像を取得する作業を継続した。これまでに首都大学の GS 変異体系統も使って取得した画像は合計で 258,000 枚を越える。必ずしも明確な輪郭がトレースできないものもあるが、RNAi 系統で複数サンプルが取得できたものは 5,214 遺伝子分、9,216 系統に達した。現在、これらの画像に対して基本的な統計解析を行い、また昨年度開発した翅面積を計算するためのプログラムを適用して、各遺伝子の翅の形成に及ぼす機能を解析している。

## [5] 研究成果物

### サブテーマ 1

① 知見・成果物・知的財産権等：なし

② 成果発表等

<論文発表>

[学術論文]

1. Izutsu M, Zhou J, Sugiyama Y, Nishimura O, Aizu T, Toyoda A, Fujiyama A, Agata K, Fuse N.: Genome features of "Dark-fly", a *Drosophila* line reared long-term in a dark environment. *PLoS*

- One*. 2012;7(3):e33288. Epub 2012 Mar 14.
2. Izutsu M, Zhou J, Sugiyama Y, Nishimura O, Aizu T, Toyoda A, Fujiyama A, Agata K, Fuse N.: Genome features of "Dark-fly", a *Drosophila* line reared long-term in a dark environment. *PLoS One*. 2012;7(3):e33288. Epub 2012 Mar 14.
  3. Murtagh V, O'meally D, Sankovic N, Delbridge ML, Kuroki Y, Boore JL, Toyoda A, Jordan KS, Pask AJ, Renfree MB, Fujiyama A, Graves JA, Waters PD: Evolutionary history of novel genes on the tammar wallaby Y chromosome: Implications for sex chromosome evolution. *Genome Res*. 2011 Nov 29. [Epub ahead of print] PMID: 22128133.
  4. Takehana Y, Naruse K, Asada Y, Matsuda Y, Shin-I T, Kohara Y, Fujiyama A, Hamaguchi S, Sakaizumi M: Molecular cloning and characterization of the repetitive DNA sequences that comprise the constitutive heterochromatin of the W chromosomes of medaka fishes. *Chromosome Res*. 2011 Nov 29. [Epub ahead of print] PMID: 22124859
  5. Renfree MB, Papenfuss AT, Deakin JE, Lindsay J, Heider T, Belov K, Rens W, Waters PD, Pharo EA, Shaw G, Wong ES, Lefèvre CM, Nicholas KR, Kuroki Y, Wakefield MJ, Zenger KR, Wang C, Ferguson-Smith M, Nicholas FW, Hickford D, Yu H, Short KR, Siddle HV, Frankenberg SR, Chew KY, Menzies BR, Stringer JM, Suzuki S, Hore TA, Delbridge ML, Mohammadi A, Schneider NY, Hu Y, O'Hara W, Al Nadaf S, Wu C, Feng ZP, Cocks BG, Wang J, Flicek P, Searle SM, Fairley S, Beal K, Herrero J, Carone DM, Suzuki Y, Sugano S, Toyoda A, Sakaki Y, Kondo S, Nishida Y, Tatsumoto S, Mandiou I, Hsu A, McColl KA, Lansdell B, Weinstock G, Kuczek E, McGrath A, Wilson P, Men A, Hazar-Rethinam M, Hall A, Davis J, Wood D, Williams S, Sundaravadanam Y, Muzny DM, Jhangiani SN, Lewis LR, Morgan MB, Okwuonu GO, Ruiz SJ, Santibanez J, Nazareth L, Cree A, Fowler G, Kovar CL, Dinh HH, Joshi V, Jing C, Lara F, Thornton R, Chen L, Deng J, Liu Y, Shen JY, Song XZ, Edson J, Troon C, Thomas D, Stephens A, Yapa L, Levchenko T, Gibbs RA, Cooper DW, Speed TP, Fujiyama A, M Graves JA, O'Neill RJ, Pask AJ, Forrest SM, Worley KC: Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biol*. 2011 Aug 29;12(8):R81. [Epub ahead of print] PMID: 21854559.
  6. Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, Tanaka M, Fujie M, Fujiwara M, Koyanagi R, Ikuta T, Fujiyama A, Miller DJ, Satoh N: Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature*. 2011 Jul 24;476(7360):320-3. doi: 10.1038/nature10249. PMID: 21785439
  7. Liu W, Morito D, Takashima S, Mineharu Y, Kobayashi H, Hitomi T, Hashikata H, Matsuura N, Yamazaki S, Toyoda A, Kikuta K, Takagi Y, Harada KH, Fujiyama A, Herzig R, Kricshek B, Zou L, Kim JE, Kitakaze M, Miyamoto S, Nagata K, Hashimoto N, Koizumi A: Identification of RNF213 as a susceptibility gene for moyamoya disease and its possible role in vascular development. *PLoS One*. 2011;6(7):e22542. Epub 2011 Jul 20.
  8. Takahashi S, Toyoda A, Sekiyama Y, Takagi H, Nogawa T, Uramoto M, Suzuki R, Koshino H, Kumano T, Panthee S, Dairi T, Ishikawa J, Ikeda H, Sakaki Y, Osada H: Reveromycin A biosynthesis uses RevG and RevJ for stereospecific spiroacetal formation. *Nat Chem Biol*. 2011 Jun 5;7(7):461-8. doi: 10.1038/nchembio.583. PMID: 21642985.
  9. Ono R, Kuroki Y, Naruse M, Ishii M, Iwasaki S, Toyoda A, Fujiyama A, Shaw G, Renfree MB, Kaneko-Ishino T, Ishino F: Identification of tammar wallaby SIRH12, derived from a

- marsupial-specific retrotransposition event. *DNA Res.* 2011;18(4):211-9.
10. Ohnishi Y, Totoki Y, Toyoda A, Watanabe T, Yamamoto Y, Tokunaga K, Sakaki Y, Sasaki H, Hohjoh H: Active role of small non-coding RNAs derived from SINE/B1 retrotransposon during early mouse development. *Mol Biol Rep.* 2011 May 13.
  11. Watanabe T, Tomizawa S, Mitsuya K, Totoki Y, Yamamoto Y, Kuramochi-Miyagawa S, Iida N, Hoki Y, Murphy PJ, Toyoda A, Gotoh K, Hiura H, Arima T, Fujiyama A, Sado T, Shibata T, Nakano T, Lin H, Ichihyanagi K, Soloway PD, Sasaki H: Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse *Rasgrf1* locus. *Science.* 2011 May 13;332(6031):848-52. PMID: 21566194.
  12. Kazuki Y, Hoshiya H, Takiguchi M, Abe S, Iida Y, Osaki M, Katoh M, Hiratsuka M, Shirayoshi Y, Hiramatsu K, Ueno E, Kajitani N, Yoshino T, Kazuki K, Ishihara C, Takehara S, Tsuji S, Ejima F, Toyoda A, Sakaki Y, Larionov V, Kouprina N, Oshimura M: Refined human artificial chromosome vectors for gene therapy and animal transgenesis. *Gene Ther.* 2011 Apr;18(4):384-93. Epub 2010 Nov 18. PMID: 21085194; Central PMCID: PMC3125098.
  13. Nishino T, Takeuchi K, Gascoigne K, Suzuki A, Hori T, Oyama T, Morikawa K, Cheeseman I M, and Fukagawa T: CENP-T-W-S-X Forms a Unique Centromeric Chromatin Structure with a Histone-like Fold. *Cell* (2012) 148, 487-501.
  14. Perpelescu M, and Fukagawa T: The ABCs of CENPs. *Chromosoma* (2011). 120, 425-446.
  15. Gascoigne KE, Takeuchi K, Suzuki A, Hori T, Fukagawa T, and Cheeseman I.M.: Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. *Cell* (2011)145, 410-422.
  16. Toyama, M., Matusda, K., Kakutani, T., Terao-Morita, M., Tasaka, M., Developmental changes in crossover frequency in Arabidopsis, *Plant J.* 14 81-87, 2011
  17. To TK, Kim JM, Matsui A, Kurihara Y, Morosawa T, Ishida J, Tanaka M, Endo T, Kakutani T, Toyoda T, Kimura H, Yokoyama S, Shinozaki K, Seki M. Arabidopsis HDA6 regulates locus-directed heterochromatin silencing in cooperation with MET1 *PLoS Genet* 7, e1002055, 2011.

[データベース] なし

[著書等]

1. 水口洋平, 豊田敦 「脊椎動物の *de novo* assembly, メダカ近交系全ゲノム解析」 細胞工学 2011, 30, 801-805.
2. 水口洋平, 豊田敦 「1000 ゲノムプロジェクトの目標と展開」 医学のあゆみ 2011, **236**, 616-622.

[解説・総説]

[その他]

1. 丸山多恵子, 藤山秋佐夫 : 日本語バイオポータル : <http://www.biportal.jp/>

<会議発表等>

[招待講演]

国際

1. Fukagawa, T., Genetic engineering of chicken centromeres, エジンバラ (イギリス) , 2011/9/17~20, 6th International Chick Meeting

2. Fukagawa, T., Genetic engineering of vertebrate centromeres, ボストン(米国), 2011/11/16~17, The Boston Area Mitosis and Meiosis (BAMM) meeting

国内

1. 藤山秋佐夫：最先端ゲノミクスが拓く生命科学の新展開と薬学への応用の可能性について北海道大学薬学部, 平成 23 年 11 月 16 日(水), 18:30-20:00
2. 深川竜郎, 高等動物におけるキネトコア形成機構, 京都市, 2011/9/21~24, 第 84 回日本生化学会大会
3. 深川竜郎, キネトコア構造を決定するエピジェネティックス機構, 横浜市, 2011/4/25, 構造エピゲノム研究会第 3 回ワークショップ
4. Fukagawa, T., A unique centromeric chromatin structure in vertebrate cells, 横浜市, 2011/12/13~16, 第 34 回日本分子生物学会年会

[一般講演]

国際

1. Nishino, T. et al., Structural cell biochemistry of a novel histone fold vertebrate kinetochore complex: CENP-TW and CENP-SX form a heterotetramer, デンバー(米国), 2011/ 12/3~6, 51th ASCB Annual Meeting
2. Nishino, T. et al., Structural cell biochemistry of a novel histone fold vertebrate kinetochore complex, ウェストダーバー(米国), 2011/ 7/11~15, Gordon Research Conferences : Chromosome Dynamics

国内

1. Osakabe, A. et al., Histone chaperone activity of a novel histone interacting factor SPT2, 横浜市, 2011/12/13~16, 第 34 回日本分子生物学会年会
2. Hori, T. et al., Ectopic localization of CCAN proteins induces centromere formation in vertebrate cells, 横浜市, 2011/12/13~16, 第 34 回日本分子生物学会年会
3. Nishimura, K. et al., Mcm8 and Mcm9 form a novel complex involved in resistance to DNA crosslinking reagents, 横浜市, 2011/12/13~16, 第 34 回日本分子生物学会年会
4. Kitamura, H. et al., Interaction between the actin-related protein Arp6 and nuclear myosin I and their contribution to the nuclear, 横浜市, 2011/12/13~16, 第 34 回日本分子生物学会年会
5. 竹内康造他, CENP-T-W-S-X 複合体は, 動原体タンパク質群の集合機構に重要な役割を担う, 横浜市, 2011/12/13~16, 第 34 回日本分子生物学会年会
6. 橋本瑞代他, ニワトリ DT40 細胞を用いた MCM-BP(MCM-binding protein)の機能解析, 横浜市, 2011/12/13~16, 第 34 回日本分子生物学会年会
7. 西淵いくの他, DNA 損傷応答におけるヒストンバリエント H2A.Z isoform の関与, 横浜市, 2011/12/13~16, 第 34 回日本分子生物学会年会

③ その他の成果発表：なし

サブテーマ 2

① 知見・成果物・知的財産権等：なし

② 成果発表等

<論文発表>

[学術論文]

1. 1. Fujisawa, H. (2011). Normalized estimating equation for robust parameter estimation. *ISM Research Memorandum*. No. 1146.
2. Fujisawa, H. and Abe, T. (2012). A family of unimodal skew-symmetric distributions with mode-invariance. *ISM Research Memorandum*. No. 1151.
3. Kato, N. and Kuriki S. (2011). Likelihood ratio tests for positivity in polynomial regressions. arXiv:1108.1033.
4. 加藤直広, 栗木哲 (2012). 2次多項式回帰曲線の正值性検定, 応用統計学, 印刷中.
5. Kumasaka, N., Fujisawa, H., Hosono, N., Okada, Y., Takahashi, A., Nakamura, Y. Kubo, M. and Kamatani, N. (2011). PlatinumCNV: a Bayesian Gaussian Mixture Model for Genotyping Copy Number Polymorphisms Using SNP Array Signal Intensity Data. *Genetic Epidemiology*, 35, 831-844.
6. Hara, H., Sei, T. and Takemura, A. (2012). Hierarchical subspace models for contingency tables. *Journal of Multivariate Analysis*. 103, 19-34.
7. Hara, H. and Takemura, A. (2011). A Markov basis for two-state toric homogeneous Markov chain model without initial parameters. *Journal of the Japan Statistical Society*, 41, 33-49.
8. Huang, J., Dou, X., Kuriki, S. and Lin, G. (2012). Dependence structure of bivariate order statistics from bivariate distributions with applications. *ISM Research Memorandum* No. 1156.
9. Takemura, A. and Hara, H. (2011). Markov chain Monte Carlo test of homogeneity of Markov chains. *Statistical Methodology*, 9, 392-406.
10. Yanagihara, H. and Fujisawa, H. (2012). Iterative bias correction of the cross-validation criterion. *Scandinavian Journal of Statistics*, 39, 116-130.

<会議発表等>

[招待講演]

国 際

1. Fujisawa, H., Optimal significance analysis of microarray data in a class of tests whose null statistic can be constructed, 7th IASC-ARS Joint 2011 Taipei Symposium, Taipei, Taiwan, 2011.12.17.
2. Fujisawa, H., Gamma-divergence and robust parameter estimation, International Workshop on Anomalous Statistics, Generalized Entropies, and Information Geometry, Nara, Japan, 2012.03.09.
3. Kuriki, S., Volume of tubes and the largest eigenvalue of a Wishart matrix, CIMPA-MICINN-UNESCO RESEARCH SCHOOL "ANALYTICAL AND ALGEBRAIC TOOLS IN STATISTICS AND GRAPHICAL MODELS", Hammamet, Tunisia, 2011.09.07.
4. Kuriki, S., Integral Geometric Approach to Statistical Distribution Theory with Application to Testing Positivity of Polynomial, ISM Symposium, Tokyo, Japan, 2012.03.24.

国 内

1. 藤澤洋徳 : ロバスト推測～外れ値への対処方法～, 応用統計学会, 大阪, 2011年06月03日
2. 藤澤洋徳 : Model selection criteria in multivariate regression, New developments in multivariate analysis, 沖縄, 2012年01月21日

[一般講演]

国 際

1. 1. Dou, X., Shirahata, S., Sugimoto, H. and Koide, T., Clustering on Mouse Ultrasonic Vocalization



- Data. Joint Meeting of the Korea-Japan Conference of Computational Statistics and the 25th Symposium of Japanese Society of Computational Statistics, Busan, Korea, 2011.11.11-12.
2. Dou, X., Shirahata, S., Sugimoto, H., and Koide, T., Cluster Analysis of Mouse Ultrasonic Vocalization Data. Joint2011 Taipei Symposium, Taipei, Taiwan, 2011.12.18.
  3. Fujisawa, H., Optimal significance analysis of microarray data in a class of tests whose null statistic can be constructed, International Conference on Advances in Probability and Statistics, Hong Kong, China, 2011.12.30.
  4. Hara, H., Bayes admissible estimation of means in the Poisson graphical models. 2011 Joint Statistical Meeting, Miami Beach, USA, 2011.08.02.
  5. Kato, N. and Kuriki, S., Likelihood ratio test for the positive polynomial cone hypotheses and associated confidence bands, ISI2011, Dublin, Ireland, 2011.08.24.

国内

1. Dou Xiaoling, 栗木 哲, 前野哲輝, 高田豊行, 城石俊彦: QTL解析における影響分析, 白旗慎吾教授定年退職記念シンポジウム, 大阪, 2012年3月17日
2. 小川光紀, 原 尚幸, 竹村彰通: 無向グラフに対する Graver 基底とそのランダムグラフの検定への応用, 応用統計学会, 大阪, 2011年06月3日
3. 小川光紀, 原 尚幸, 竹村彰通: 無向グラフに対するマルコフ基底の構造とその応用, 統計関連学会連合大会, 福岡, 2011年09月05日
4. 熊坂夏彦, 藤澤洋徳, 細野直哉, 高橋 篤, 久保充明, 鎌谷直之, Bayesian Gaussian Mixture Model によるアレル特異的なコピー数多型のタイピング手法の開発, 統計関連学会連合大会, 福岡, 2011年09月06日
5. 二宮嘉行: スキャン統計量のための裾確率評価, 計量生物学会, 大阪, 2011年06月02日
6. 二宮嘉行: A p-value evaluation for multiple testing problem based on highly correlated test statistics. Meetings of Hiroshima Statistics Study Group, 広島, 2011年11月4日
7. 原尚幸: 最近の計算代数統計学の展開, 統計関連学会連合大会, 福岡, 2011年09月05日
8. 原尚幸, 青木 敏, 竹村彰通: Running Markov chain without Markov basis, 統計関連学会連合大会, 福岡, 2011年09月06日
9. 藤澤洋徳: マイクロアレイデータを利用した遺伝子発現差解析において帰無統計量が構成できる検定の中で最適な有意性検定, 名古屋統計セミナー, 名古屋, 2011年05月20日
10. 藤澤洋徳:  $\gamma$  相互エントロピーに基づいたロバストなモデル選択規準, 統計関連学会連合大会, 福岡, 2011年09月06日
11. 藤澤洋徳: 遺伝子発現差解析におけるP値推定と最適な有意性検定, 統計科学セミナー, 福岡, 2012年01月13日

③ その他の成果発表: なし

サブテーマ3

① 知見・成果物・知的財産権等

② 成果発表等

<論文発表>

[学術論文]

1. Oka, A., Mita, A., Takada, Y., Koseki, H., and Shiroishi, T. (2010). Reproductive isolation in hybrid mice due to spermatogenesis defects at three meiotic stages. *Genetics* 186, 339-51.
2. Kondrashov, N., Pusic, A., Stumpf, CR., Shimizu, K., Hsieh, AC., Xue S., Ishijima, J., Shiroishi, T., and Barna, M.: Ribosome-mediated specificity in hox mRNA translation and vertebrate tissue patterning. *Cell* 145: 383 – 397, 2011.
3. Yonekawa H., Takada T., Shitara, H., Taya C., Matsushima Y., Matsuoka K., Kikkawa Y.: Mouse models for atopic dermatitis developed in Japan. *Atopic Dermatitis - Disease Etiology and Clinical Management*, ed, Jorge Esparza-Gordillo, 21-38, 2012, InTech, ISBN 978-953-308-67-5.
4. Takada T and Shiroishi T. Complex quantitative traits cracked by the mouse inter-subspecific consomic strains. *Experimental Animals*, (2012) In press.
5. Sugimoto H., Okabe S., Kato M., Koshida N., Shiroishi T., Mogi K., Kikusui T., Koide T. A role for strain differences in waveforms of ultrasonic vocalizations during male–female interaction. *PLoS ONE* 6: e22093, (2011).
6. Ishii A., Koide T., Takahashi A., Shiroishi T., Hettinger T.P., Frank M.E., Savoy L.D., Formaker B.K., Yertutanol S., Lionikas A., Blizard D.A. B6-MSM consomic mouse strains reveal multiple loci for genetic variation in sucrose octaacetate aversion. *Behav. Genet.* 41: 716-723, (2011).
7. Koide T., Ikeda K., Ogasawara M., Shiroishi T., Moriwaki K., Takahashi A. A new twist on behavioral genetics by incorporating wild-derived mouse strains. *Exp. Anim.* 60: 347-354, (2011).
8. De Graeve, F.M., Van de Bor, V., Ghiglione, C., Cerezo, D., Jouandin, P., Ueda, R., Shashidhara, L.S., Noselli, S. *Drosophila* apc regulates delamination of invasive epithelial clusters. *Developmental Biology* 368, 76-85 (2012).
9. Ito, Y., Kimura, F., Hirakata, K., Tsuda, K., Takasugi, T., Eiguchi, M., Nakagawa, K., Kurata, N. Fatty acid elongase is required for shoot development in rice. *Plant J.* 66: 680-688, 2011.
10. Ito, Y., Thirumurugan, T., Serizawa, A., Hiratsu, K., Ohme-Takagi, M., Kurata, N. Aberrant vegetative and reproductive development by overexpression and lethality by silencing of OsHAP3E in rice. *Plant Science* 181: 105-110, 2011.
11. Kubo, T., Yoshimura, A., Kurata, N. Hybrid Male Sterility in Rice Is Due to Epistatic Interactions with a Pollen Killer Locus. *Genetics* 189: 1083-1092, 2011.
12. Tsuda, K., Ito, Y., Sato, Y., Kurata, N. Positive autoregulation of a KNOX gene is essential for shoot apical meristem maintenance in rice. *Plant Cell* 23: 4368-4381, 2011.
13. 宋 剛秀, 井上克巳. モデル生成を用いた代謝ネットワークにおける極小活性パスウェイの列挙. 人工知能学会誌, 27(3):204-212, 2012.

[データベース]

<会議発表等>

[招待講演]

国際

1. Takada T.: Functional genomics powered by mouse inter-subspecific genome diversity. Mouse Genetics Symposium, Seoul National University, Oct. 27, 2011, Seoul, Korea.
2. Kurata, N., Mizuta, Y., Kubo, T., Harushima Y. Reproductive barriers in rice diversification. SMBE 2011 Kyoto(Symposium6: Molecular bases of speciation), 2011年7月26-30日, 京都.

3. Kurata, N. ASSEMBLING THE O. OFFICINALIS, CC GENOME, SHORT READS. 9th International Symposium on Rice Functional Genomic (Workshop1: International Oryza Map Alignment Project), 2011年11月7-9日, Taipei, Taiwan.

国内

1. 城石俊彦. 日本産愛玩用マウスゲノムの遙かな旅. 理研バイオリソースセンター10周年記念公演シンポジウム, つくば国際会議場エポカルつくば大ホール, 2011, 7.1, 筑波.

[一般講演]

国際

1. Arakawa, T. An automated markov analysis of social interaction behavior of mouse and its applications to genetic mapping. ISM - ISI - ISSAS Joint Conference 2012, 2012年2月2日-3日, The Institute of Statistical Mathematics, Tachikawa, Japan.
2. Soh, T., Inoue, K., Baba, T., Takada, T., Shiroishi, T. Predicting Gene Knockout Effects by MinimalPathway Enumeration. In: Petre Dini and Pascal Lorenz (eds.), Proc. 4th Intl Conference on Bioinfo.Biocomput. Systems and Biotech.(BIOTECHNO 2012), pp.11-19, ThinkMind, (Selected for one of the two papers invited for IARIA Journals.) St. Maarten, Netherlands Antilles, March 25-30, 2012.
3. Sugimoto, H., Okabe, S., Kikusui, T., Koide, T. A role of strain difference in waveforms of male ultrasonic vocalization for social behavior. 13th Annual Meeting for International Behavioural and Neural Genetics Society. Rome, Italy May 10-14, 2011.
4. Koide, T., Tanave, A., Sugimoto, H., Takahashi, A. Wild-derived stock of mice, a useful resource for studying genetic basis of anxiety-like behavior. 13th Annual Meeting for International Behavioural and Neural Genetics Society. Rome, Italy May 10-14, 2011.
5. Ohyanagi, H., Kurata, N. DIVERSIFICATION AND EVOLUTION OF TWO SUBSPECIES OF ORYZA SATIVA AND THREE WILD ACCESSIONS OF ORYZA RUFIPOGON. 9th International Symposium on Rice Functional Genomic, 2011年11月7-9日, Taipei, Taiwan.
6. Takasugi, T., Fujita, M., Kurata, N., Ito, Y. A SEARCH FOR A GENE CONNECTING CYTOKININ AND A KNOX GENE. 9th International Symposium on Rice Functional Genomic(ポスター), 2011年11月7-9日, Taipei, Taiwan.
7. Ohyanagi, H., Nagasaki, H., Nagata T., Mochizuki, T., Kaminuma, E., Nakamura, Y., Takeshita, S., Aizu, T., Toyoda, A., Fujiyama, A., Zhao, Q., Han, B., Kurata, N. An evolutionary aspect for two subspecies of Oryza sativa and three wild accessions of Oryza rufipogon. Plant and Animal Genome XX(ポスター), 2012年1月14-18日, San Diego, U.S.A

国内

1. 高田豊行, 三田晃彦, 森脇和郎, 米川博通, 城石俊彦. 日本産野生由来マウス系統 MSM/Ms を用いた加齢性エネルギー代謝の機能ゲノム解析. 日本遺伝学会, 2011, 9.20-22, 京都.
2. 城石俊彦. 野生マウスが拓く新しい実験動物の地平. 第58回日本実験動物学会 シンポジウム タワーホール船堀, 2011, 5.25, 東京.
3. 高田豊行. MSM/Ms 系統を基盤にしたマウス機能ゲノム学データベースの構築, 日本実験動物学会総会・ワークショップ「国内実験動物データベースの現状と展望」第58回日本実験動物学会総会, 2011, 5月25日~27日, 東京.
4. 岡彩子, 城石俊彦. マウス亜種間における生殖隔離の遺伝的メカニズム. 日本進化学会第13回大会,

2011, 7.29~31, 京都.

5. 岡彩子, 城石俊彦. マウス亜種間の生殖隔離の遺伝的メカニズム, 日本遺伝学会第 83 回大会, 2011, 9.20~22, 京都.
6. 宋 剛秀, 井上克巳, 馬場知哉, 高田豊行, 城石俊彦. 極小活性パスウェイの列挙を用いた大腸菌における遺伝子欠損の影響予測. 情報処理学会研究報告, Vol.2012-ICS-165, No.7, pp.1-6, 情報処理学会, 2012. 第 165 回情報処理学会知能システム研究会, 2012 年 1 月 12 日, 国立情報学研究所, 東京.
7. 宋 剛秀, 井上克巳, 馬場知哉, 高田豊行, 城石俊彦. 極小活性パスウェイの列挙を用いた遺伝子ノックアウトの影響予測. 第 47 回人工知能学会分子生物情報研究会, 2011 年 12 月 15 日, 慶応義塾大学日吉キャンパス, 横浜.
8. 荒川俊也, 高橋阿貴, 田邊彰, 柿原聡, 木村真吾, 杉本大樹, 城石俊彦, 富原一哉, 小出剛, 土谷隆. 隠れマルコフモデルによるマウス状態の自動推定とコンソミックマウスの特徴付け, 統計数理研究所共同研究集会「最適化: モデリングとアルゴリズム」, 2012 年 3 月 16 日, 政策研究大学院大学, 東京, 日本
9. 菊水健史, 小出剛: マウス求愛歌の行動学的解析 (シンポジウム) 第 34 回日本神経科学大会, 2011 年 9 月 14-17 日, 横浜
10. 菅野康太, 小久保博樹, 相賀裕美子, 高橋阿貴, 小出剛, 石浦章一: ドーパミントランスporter 発現制御候補因子 *Hesr1/2* ノックアウトマウスの行動解析 (ポスター), 第 34 回日本神経科学大会, 2011 年 9 月 14-17 日, 横浜
11. 後藤達彦, 石井亜矢子, 西明紀, 高橋阿貴, 城石俊彦, 小出剛: 時間依存的な自発活動量のマウス系統差に関わる多因子の遺伝学的解析 (ポスター), 第 34 回日本神経科学大会, 2011 年 9 月 14-17 日, 横浜
12. 小出剛, 高橋阿貴, 城石俊彦, 田邊彰: MSM-B6 コンソミック系統を用いた不安様行動の分子遺伝学的解析 (ポスター), 第 34 回日本神経科学大会, 2011 年 9 月 14-17 日, 横浜
13. 杉本大樹, 高橋阿貴, 小出剛: アセチルコリン受容体調節因子 *Lynx1* の社会行動への関与 (ポスター), 第 34 回日本神経科学大会, 2011 年 9 月 14-17 日, 横浜
14. 高橋阿貴, Klaus Miczek, 小出剛: 背側縫線核 GABA<sub>B</sub> 受容体の活性化による過剰な攻撃行動の詳細な行動学的・神経学的解析 (ポスター), 第 34 回日本神経科学大会, 2011 年 9 月 14-17 日, 横浜
15. 田邊彰, 高橋阿貴, 荒川俊也, 柿原聡, 木村真吾, 杉本大樹, 城石俊彦, 富原一哉, 土谷隆, 小出剛: 隠れマルコフモデルを用いたマウス社会行動解析ソフトウェアの開発 (ポスター), 第 34 回日本神経科学大会, 2011 年 9 月 14-17 日, 横浜
16. 小出剛: 野生由来マウス系統を用いた行動遺伝学 (シンポジウム) 第 58 回日本実験動物学会総会, 2011 年 5 月 25-27 日, 東京
17. 杉本大樹, 菊水健史, 城石俊彦, 小出剛: マウス交配行動時の超音波発声の役割 (口頭発表), 第 58 回日本実験動物学会総会, 2011 年 5 月 25-27 日, 東京
18. 高橋阿貴, 城石俊彦, 小出剛: 日本産野生由来マウス系統 MSM の過剰な攻撃行動とセロトニン受容体 mRNA の系統差 (口頭発表), 第 58 回日本実験動物学会総会, 2011 年 5 月 25-27 日, 東京
19. 大柳一, 長崎英樹, 永田俊文, 望月孝子, 神沼英里, 中村保一, 竹下紗由美, 会津智幸, 豊田敦, 藤山秋佐夫. Q. Zhao, B. Han, 倉田のり. 高速 DNA シーケンサーで探る栽培種近縁野生イネ *Oryza rufipogon* 3 系統の多様性と進化. 日本育種学会第 120 回講演会, 2011 年 7 月 22-24 日, 福井
20. 高梨秀樹, 平田悠人, 大柳一, 瓦間淳子, 永田俊文, 豊田敦, 藤山秋佐夫, 倉田のり, 堤伸浩. 次世代シーケンサーを用いたイネ雌性配偶体構成細胞のトランスクリプトーム解析. 日本育種学会第 120 回講演会, 2011 年 7 月 22-24 日, 福井

21. 山木辰一郎, 長戸康郎, 野々村賢一, 倉田のり. イネ GYPSY EMBRYO 遺伝子は穂の枝分かれと花分裂組織の有限性を制御する. 日本育種学会第 120 回講演会, 2011 年 7 月 22-24 日, 福井
22. 久保貴彦, 藤田雅丈, 高橋宏和, 中園幹生, 堤伸浩, 倉田のり. イネ雌性配偶子形成過程のトランスクリプトーム解析. 日本育種学会第 120 回講演会, 2011 年 7 月 22-24 日, 福井
23. 濱田和輝, 深澤開, 長山大志, 横山幸治, 土田博子, 五十嵐香理, 倉田のり, 矢野健太郎. イネの遺伝子発現ネットワーク解析とデータベース. 日本育種学会第 120 回講演会(ポスター), 2011 年 7 月 22-24 日, 福井
24. 小宮怜奈, 大柳一, 新濱充, 渡部聡朗, 筒井康博, 望月孝子, 神沼英里, 中村保一, 倉田のり, 野々村賢一. イネ生殖細胞特異的 Argonaute タンパク質 MEL1 と結合する small RNAs の同定. 第 34 回日本分子生物学会年会, 2011 年 12 月 13-16 日, 横浜
25. 望月孝子, 長崎英樹, 神沼英里, 大柳一, 倉田のり, 二河成男, 中村保一. 新型シーケンサ・アーカイブ配列からの DNA 多型注釈データベース構築. 第 34 回日本分子生物学会年会(ポスター), 2011 年 12 月 13-16 日, 横浜
26. Takasugi, T., Fujita, M., Kurata, N., Ito, Y. Cytokinin-induced KNOX gene expression in rice - a search for a gene connecting cytokinin and KONX gene. 第 34 回日本分子生物学会年会(ポスター), 2011 年 12 月 13-16 日, 横浜
27. Hamada, K., Fukazawa K., Nagayama, T., Yokoyama, K., Tsuchida, H., Igarashi, K., Kurata, N., Yano, K. OryzaExpress : An integrated Database for Gene Expression Networks in Rice. 第 34 回日本分子生物学会年会(ポスター), 2011 年 12 月 13-16 日, 横浜
28. 津田勝利, 伊藤幸博, 佐藤豊, 倉田のり. 茎頂分裂組織の維持に必須なイネ KNOX 遺伝子の正の自己制御. 第 53 回日本植物生理学会年会, 2012 年 3 月 16-18 日, 京都
29. 小宮怜奈, 大柳一, 新濱充, 渡部聡朗, 筒井康博, 米田弥生, 望月孝子, 神沼英里, 倉田のり, 野々村賢一. イネ生殖細胞特異的 Argonaute タンパク質 MEL1 と結合する small RNAs の同定. 第 53 回日本植物生理学会年会, 2012 年 3 月 16-18 日, 京都
30. 高橋秀樹, 平田悠人, 大柳一, 瓦間淳子, 永田俊文, 太田垣俊吾, 豊田敦, 藤山秋佐夫, 倉田のり, 堤伸浩. 次世代シーケンサーを用いたイネ雌性配偶体構成細胞のトランスクリプトーム解析. 第 53 回日本植物生理学会年会, 2012 年 3 月 16-18 日, 京都
31. 関根大輔, 大西孝幸, 古海弘康, 吉野みほ子, 倉田のり, 木下哲. イネの倍数体間交雑における胚乳発生異常の解析. 第 53 回日本植物生理学会年会, 2012 年 3 月 16-18 日, 京都
32. 望月孝子, 長崎英樹, 藤澤貴智, 神沼英里, 大柳一, 倉田のり, 二河成男, 中村保一. 新型シーケンサ・アーカイブ配列からの植物 DNA 多型注釈データベース構築. 第 53 回日本植物生理学会年会(ポスター), 2012 年 3 月 16-18 日, 京都
33. 濱田和輝, 深澤開, 長山大志, 横山幸治, 土田博子, 五十嵐香里, 倉田のり, 矢野健太郎. OryzaExpress: イネの遺伝子発現ネットワークとオミックス情報統合データベース. 第 53 回日本植物生理学会年会(ポスター), 2012 年 3 月 16-18 日, 京都
34. 堀内陽子, 春島嘉章, 藤澤洋徳, 大柳一, 藤田雅丈, 望月孝子, 倉田のり. イネジャポニカ, インディカ間で有意に発現量に差がみられる遺伝子の遺伝子及び周辺塩基配列構造の大規模解析. 日本育種学会第 121 回講演会. 2012 年 3 月 28-30 日, 宇都宮
35. 大柳一, 長崎英樹, 永田俊文, 望月孝子, 神沼英里, 中村保一, 竹下紗由美, 会津智幸, 豊田敦, 藤山秋佐夫, Zhao, Q., Han, B., 倉田のり. 栽培イネ *Oryza sativa* 2 亜種と野生イネ *Oryza rufipogon* 3 アクセッションにおける進化的関係. 日本育種学会第 121 回講演会. 2012 年 3 月 28-30 日, 宇都宮
36. 関根大輔, 大西孝幸, 古海弘康, 倉田のり, 木下哲. イネの胚乳発生サイズを制御する父母ゲノムの効

- 果. 日本育種学会第 121 回講演会. 2012 年 3 月 28-30 日, 宇都宮
37. 武田泰実, 池田真由子, 山口秀和, 村田強, 山木辰一郎, 加藤淳太郎, 野々村賢一, 倉田のり, 北野英己. ゲノム構成と形態的特徴から見た野生イネ遺伝資源の再評価. 日本育種学会第 121 回講演会. 2012 年 3 月 28-30 日, 宇都宮
38. 濱田和輝, 深澤開, 長山大志, 横山幸治, 土田博子, 五十嵐香里, 倉田のり, 矢野健太郎. 大規模遺伝子発現情報に基づく遺伝子発見手法の開発. 日本育種学会第 121 回講演会. 2012 年 3 月 28-30 日, 宇都宮
39. Shenton, M., 倉田のり. *Oryza* 属の種間交雑における花粉管伸長障害の特性解析. 日本育種学会第 121 回講演会(ポスター). 2012 年 3 月 28-30 日, 宇都宮
40. 太田垣駿吾, 大柳一, 豊田敦, 藤山秋佐夫, 倉田のり. 次世代シーケンサーを用いたイネ属での雑種強勢機構の解析. 日本育種学会第 121 回講演会(ポスター). 2012 年 3 月 28-30 日, 宇都宮
39. 久保貴彦, 藤田雅丈, 倉田のり. イネ雌性不稔に関わる遺伝子発現プロファイリング. 日本育種学会第 121 回講演会(ポスター). 2012 年 3 月 28-30 日, 宇都宮

<受賞>

- ③ その他の成果発表: なし