

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、世界に類を見ない独創性・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）。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)。

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, Kurikiet al., BIRS Workshop, Random Fields and Stochastic Geometry, 2009, Canada) (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)。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, Horiuchi et al., BMC Genomics, 2010) や生殖隔離因子の相互作用解析と単離 (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などが進んでいるが、いずれも大型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ており、統計学や情報学はそのための補助的なツールとして使われるにすぎない。また、本計画で解析す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面的な理解を研究コミュニティに提供する。そのような計画は国内外にみられない。

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、多重ゲノムデータの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して、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、多重的データ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。
- (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、遺伝子型相関の検出も行う。
- (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、用いる戦略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あてはめを通して 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あ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ーに公開する。

サブテーマ2では、提案した並べ替え検定、多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、時系列データを含む多元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元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

〔国立情報学研究所〕 宇野毅明、井上克巳

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

〔現融合研究センター〕 辰本将司、松崎肖子、塚本ゆみ、吉田 悟、丸山多恵、Wei-Hao Shang

〔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

〔現融合研究センター〕 春島嘉章、堀内陽子、Dou Xiaoling

〔大分県立看護科学大学〕 坂口隆之

〔東京大学〕 原 尚幸

〔九州大学〕 二宮嘉行

〔政策科学大学院大〕 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、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〕 北本朝展

〔現融合研究センター〕 春島嘉章、堀内陽子、岡(木曾)彩子、杉山大樹、梅森十三、田中英希、
Dou Xiaoling

〔東京大学〕 中谷明弘

〔長浜バイオ大学〕 阿部貴志

〔首都大学〕 相垣敏郎

〔政策科学大学院大〕 土谷 隆

〔大阪大学〕 白旗慎吾

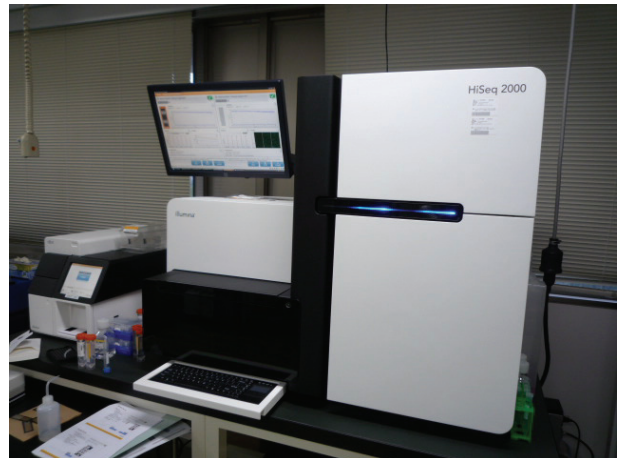
[4] 研究の進捗状況

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：大規模シークエンスの稼働

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ス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スセンターでは、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。



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)。

(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、メチレーション解析と続いている。再シーケンスの場合は、ゲノム DNA を数 μg 程度から、イルミナのプロトコー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。

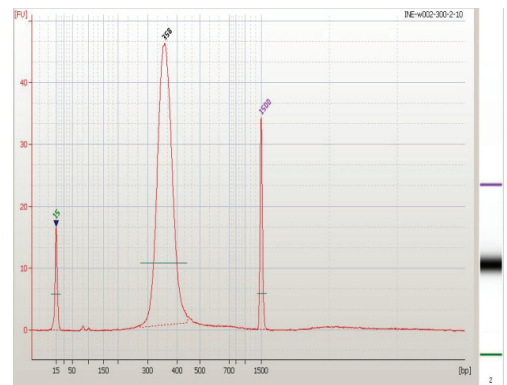
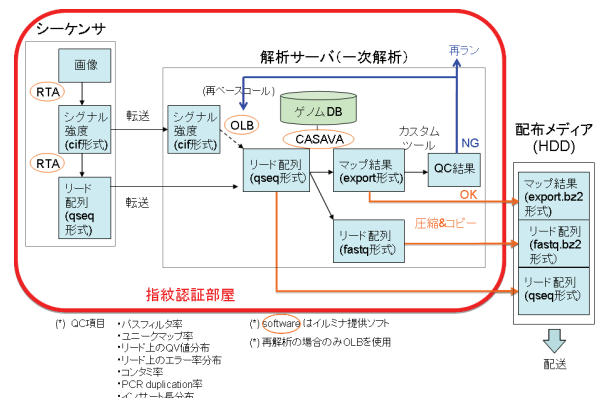


図1 Agilent バイオアナライザーによるフラグメント解析結果

(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	QV 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

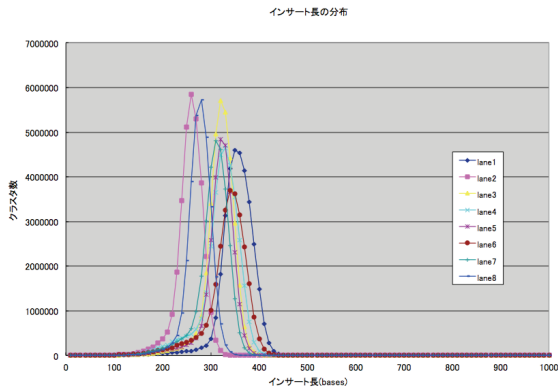


図 3 (A) 鋳型のインサート長の分布

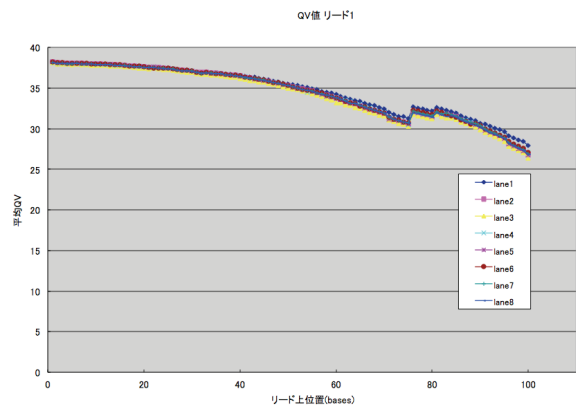
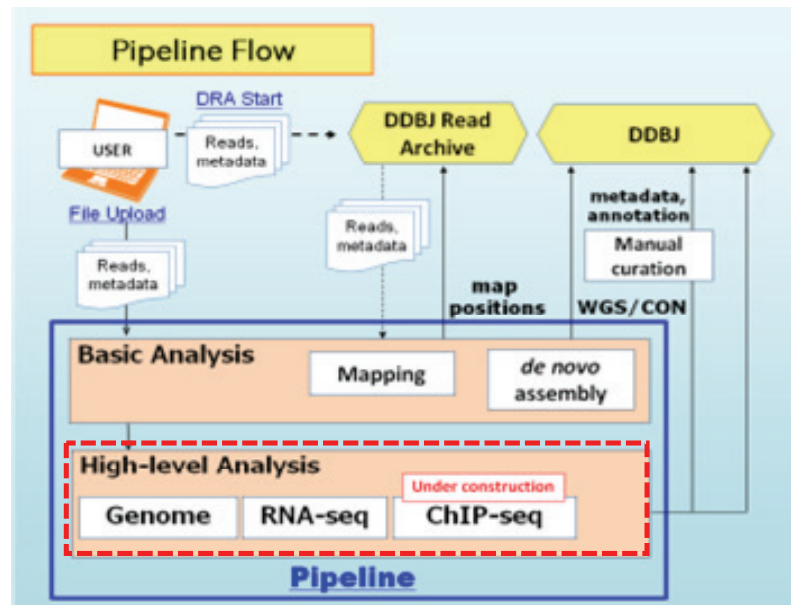


図 3 (B) リード上の各ポジションにおける平均 QV の分布

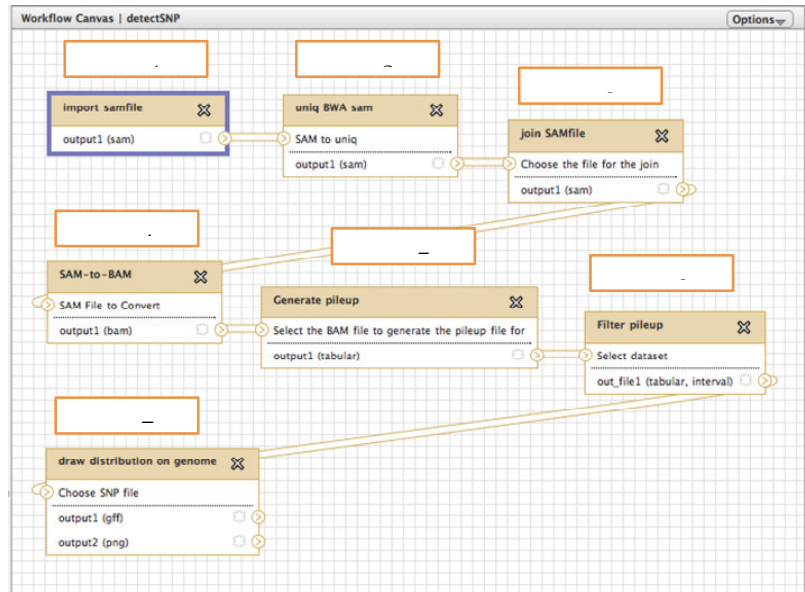
次世代シーケンシングパイプラインから出力されたすべての配列データは、DDBJ の新型シーケンサのデータアーカイブである DRA に登録される。DDBJ では、登録後のデータ処理をクラウド型で行うパイプライン開発を進めており (図 4)、マッピングと *de novo* アセンブリを行う「基礎処理部」と、ゲノム解析、RNA-seq、ChIP-seq 等の個別の高次処理を実現する「高次処理部」で構成されている。22 年度のプロジェクトでは、このうちの高次処理部を開発した。特に、高次処理部に解析ワークフロー開発基盤の Galaxy を実装し、基礎処理部と高次処理部の接続を行った。



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

図 4(上):パイプラインの処理フロー。図 5(下):Galaxy で構築したワークフロー



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年度は、まず、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から始めた。深川らは、セントロメアタンパク質を長年研究してきているので、セントロメアタンパク質の結合するゲノム領域の同定を行った。具体的には、セントロメアタンパク質 CENP-A と共免疫沈降してくる DNA を解析した。これらの DNA

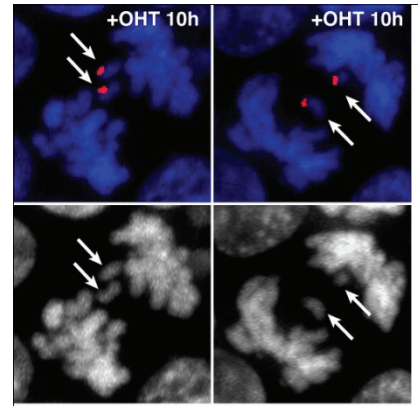
をクローニングして、そのクローンを FISH 解析してセントロメア領域に局在するかどうかを調べたところ、半数以上のクローンがセントロメアへ局在した。セントロメアへ局在する DNA に関して、シーケンス解析を行った。その結果、各染色体のセントロメアごとに、配列が異なることが判明した。

CENP-A と結合する DNA をクローニングして配列解析するのではなく、新型シーケンサーで直接解析する研究も同時におこなった。サザン解析から、反復配列の存在が予想される染色体では、反復配列と思われる位置に存在せず新たな peak をしめした。

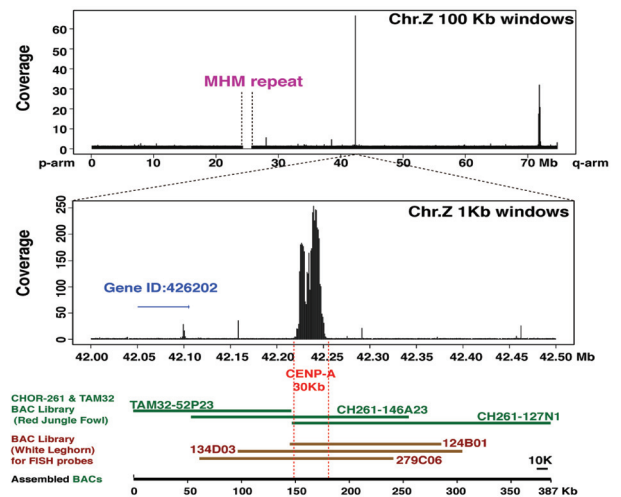
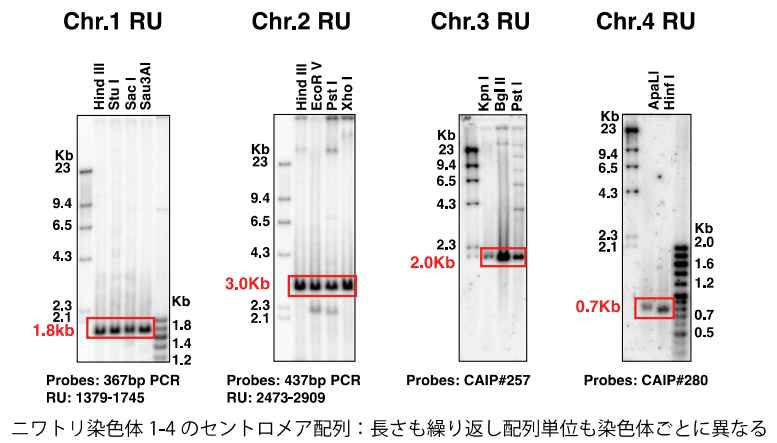
この非反復配列が、本当にセントロメアであるかどうかを確認するために FISH 解析を行い、セントロメアに存在することを確認した。今後は、この染色体領域に改変を加えた細胞を作成してこの配列の機能的意義を明らかにする。また、そのような細胞で CENP-A と結合する DNA を新型シーケンサーで解析することも計画している。ニワトリゲノム配列は、まだ不正確な情報が多く、新型シーケンサーから得られた配列のマッピングには、高度な情報学の手法が必要であり、融合研究の意義もあわせて確認できた。

1-3: 植物における高次ゲノム機能研究の超大規模化

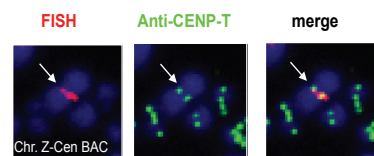
この小課題では、代表的なモデル植物であるシロイヌナズナ (*A.thaliana*) の同属近縁種セイヨウミヤ



セントロメアを除くと染色体分配に失敗する

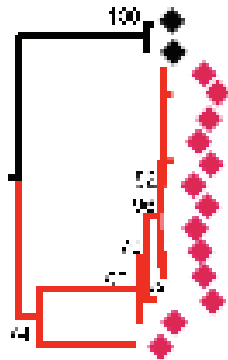


Z 染色体では、CENP-A 結合配列が非反復配列領域に存在した



非反復配列が Z 染色体ではセントロメアにある

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。予備的結果では動原体特異的に挿入が起きていた。今後は、*Tal1* の転移が起きた系統のゲノム配列を読むことで、*Tal1* 挿入部位の特異性を定量化する予定である。



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。

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みだしている。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。

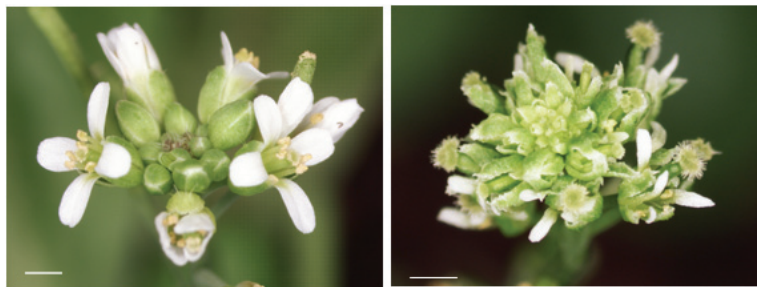


図: 低メチル化によって増殖するトランスポゾンが湯発したシロイヌナズナの発生異常(右)。左は正常なシロイヌナズナの花。

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。今後は、この実験系を用いて、シロイヌナズナの突然変異体やヒメツリガネゴケ(基礎生物学研究所、長谷部グループ研との共同研究)の生活環におけるゲノムDNAメチル化状態の変動を調べる予定である。

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。

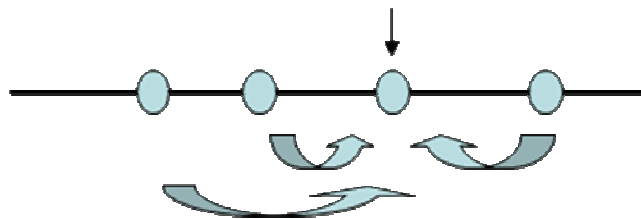
しかしながら、マイクロアレイデータは、たくさんの遺伝子からデータを観測している。そのため、

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



本年度は、もっとも標準的に利用されている平均の同等性検定を扱った。平均の同等性検定については、多くの議論がなされているが、どの手法もアドホックな提案である。おのおのの遺伝子に対して、遺伝子発現差があるかないかは、事前に分からない。そのデータを区別せずにアドホックに使ってしまうのは、データが濁ってしまい、妥当な結果がもたらされないことは容易に想像できる。

そのような状況で、Efron が提案したバランス型並べ替え手法の利用と、Pan が提案した検定統計量の工夫は、P 値を推定する際に、数理的な妥当性を持っている。特に、前者の手法は、実験回数が偶数回に限るのに対して、後者の手法は、実験回数が奇数回でも利用でき、Efron の手法の拡張になっている。しかし、そのような方法が、何らかの意味で最も優れているかどうかについては、全く検討されていない。

まずは、数理的な妥当性を持つためには、検定統計量 T がどのような性質を持てばよいかを考えた。その条件は下記で整理できた。

$$p = \Pr(T(\mathbf{Z}) \geq t | H_0) = \Pr(T^{\text{null}}(\mathbf{Z}) \geq t)$$

ここで、 \mathbf{Z} は、データを象徴して表現している。最初の等号は P 値の定義である。P 値は帰無仮説が正しい下で定義される。その P 値が、かりに、ある適当な帰無統計量 T (NULL) によって、二つ目の等号が成り立つとしよう。二つ目の等号のポイントは、その等号が、帰無仮説が正しいか正しくないかに依存せずに成立するという点である。このような等式が成り立てば、他の遺伝子のデータを借りることは容易である。P 値は、次のような数式で、簡単に経験推定をすることが可能になる。

$$\hat{p}(T(\mathbf{Z}^\dagger); t_g) = \frac{1}{BG} \sum_{b=1}^B \sum_{g'=1}^G I(T^{\text{null}}(\mathbf{Z}_{bg'}^\dagger) \geq t_g)$$

これは、遺伝子番号 g に対して経験推定された P 値である。並べ替え標本の番号が b で表現されている。 \mathbf{Z}^\dagger はある意味での並べ替え標本を象徴的に表している。ここでポイントとなるのは、経験推定が BG 回の標本平均となっていることである。もしも遺伝子一つに対して普通に経験推定をすると B 回の

標本平均となる。実験回数が少ないときは B は小さくなるので経験推定が安定しない。しかし、マイクロアレイデータでは、遺伝子数 G は大きいので、結果的に BG は大きくなり、経験推定は安定する。

さて、上述の条件が成り立つような検定統計量の中で、検出力を最大にするものは、どのようなものであるか。本研究では、条件の数式で、一番右側の値がパラメータに依存しないことに着目して、そのような不変性を導入することで、条件を満たす検定統計量のクラスを作り上げた。さらに、正規性を仮定することで、そのようなクラスの中で最強力な検定を導出した。

最初は、データの背後にある分布が、位置パラメータに対して左右対称であることを想定した。Pan の方法が妥当性を持つために必要とされる条件である。その条件の下では、最強力な検定をもたらす検定統計量は、Pan の検定統計量とほぼ同じものとなった。この結果として、Pan が提案した方法は、ある意味で最強力であることが証明された。

次に、データの背後にある分布が、検定される二つの集団において、位置パラメータを除いて同じであると想定した。この場合には、先ほど得られた検定統計量と、ほんの少し違う検定統計量を得られた。それが以下である。

$$T_p(Z) = \frac{|Q_p|}{\sqrt{\{(N-4)S^2 + Q_0^2\}/(N-3)}},$$

$$Q_p = \left\{ \frac{n_1 m_1}{n} (\bar{X}_{(1)} - \bar{Y}_{(1)}) + \frac{n_2 m_2}{m} (\bar{X}_{(2)} - \bar{Y}_{(2)}) \right\} / \sqrt{\frac{n_1 m_1}{n} + \frac{n_2 m_2}{m}}$$

$$Q_0 = \sqrt{\frac{n_1 m_1 n_2 m_2 / nm}{n_1 m_1 / n + n_2 m_2 / m}} \{ (\bar{X}_{(1)} - \bar{Y}_{(1)}) - (\bar{X}_{(2)} - \bar{Y}_{(2)}) \}$$

この検定統計量の最大の特徴は Q_0 である。この Q_0 が Pan タイプの検定統計量との差異である。この統計量のおかげで、Pan タイプの検定統計量と比べると自由度が一つ大きい。そのため、Pan タイプの統計量と比べて、検出力も高く、P 値推定も安定すると期待できる。通常であれば、自由度 1 の差は大きくないが、マイクロアレイデータの場合には、実験回数が少なく、結果として、自由度 1 の違いは大きな差を提示すると予想できる。

また、問題を数理的に整理したことによるメリットが、この数式には表れている。Pan の手法は、標本を二分割するというアイデアにたどりつけば、アドホックに提案することは容易であろう。上記の検定統計量も、結果的には、標本を二分割した検定統計量に基づいているが、統計量を Q_0 導入するということは、なかなか、アドホックには気づきにくいだろう。問題を数理的な最適性に基づいてアプローチしたことによって、その統計量が自然に表面にあらわれている。このように、アイデアを必要以上に凝らさずとも、基本的なアイデアがしっかりとしていれば、後は自動的に良い手法を構成できるのが、数理的アプローチの利点である。

次年度は、分散の同等性検定についても検討し、さらに、提案した手法のパフォーマンスを数値的に調べたいと考えている。可能であれば、近年になって盛んに研究されている FDR についても、考察を進めて行きたいと考えている。

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。この多重検定問題に対し、多重性調整した p 値

の基本的な評価方法を開発し、それに基づき遺伝的不適合を引き起こす部位の候補を与えたことが現段階までの研究成果である。

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_2 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 $X^{(s,t+1)}$ 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。

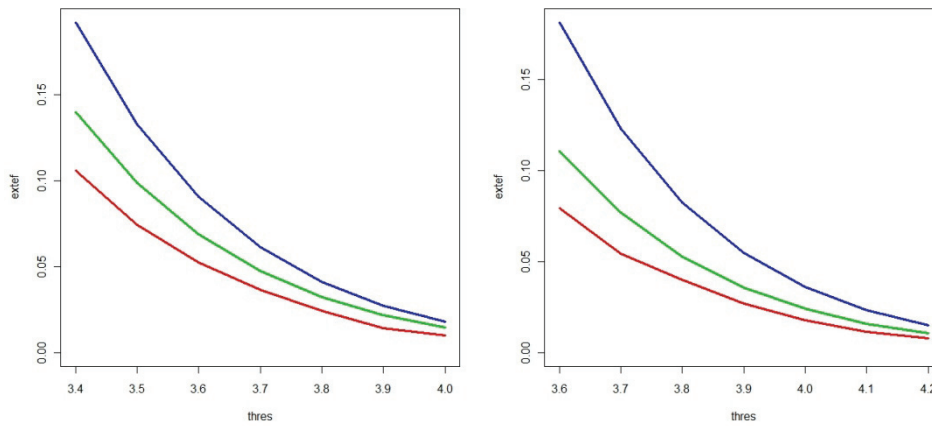


図: マーカーが 10cM(左)あるいは 5cM(右)間隔で存在した場合の確率評価。
赤線、緑線、青線はそれぞれ真値(モンテカルロ法による近似)、提案方法、Bonferroni 法に対応する

現在取り組んでいるのは、より直接的に遺伝的不適合をモデル化し、それに基づき検定をおこなうことである。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ることである。上述の多重検定に比べ、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ため、多重性調整は困難になる。しかし、もし実現できれば、遺伝的不適合を引き起こす遺伝子座についてより詳細な情報を得られるはずである。

また、解析の精度を高めるために、正常なマウスのデータも組み入れて検定をおこなうことが考えられる。例えば、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を検出できれば、それが遺伝的不適合に関わっていると判断することができる。このことは各検定において独立性の検定を考慮することで実現でき、その検定統計量の相関構造も、上述の提案方法のものよりは複雑な計算になるものの導出可能である。しかし、正常なマウスの群に異常なマウスが含まれている可能性が高い、という問題が実はあり、これは検定の結果を大きく左右する。そこで、その可能性を考慮した仮定のもと、具体的には潜在変数を用いたモデリングをすることにより、検定を定式化することを現在検討中である。

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関連解析において、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。

そこで本研究では、ガウスグラフィカルモデルにおける共分散行列の最尤推定量の標準的な計算アルゴリズムである比例反復法の局所計算アルゴリズムの開発を行った。変数の次元を n とする。通常のプロポーション反復法の場合、一回の更新則には、 $(n-1) \times (n-1)$ の逆行列の計算を含むために、計算オーダーは n^3 となり、遺伝子データのように n が大きい場合には計算コストが高くなる。しかしながら、遺伝子データの場合には、遺伝子間の条件付独立関係を表わすグラフは疎であることが多く、その場合にグラフの分解に対応した計算アルゴリズムの局所化により、計算コストを劇的に減少させることが可能となる。

グラフに長さが 4 以上の弦のない閉路が存在した場合には、下図の(a)→(b)のように、弦を加えるという操作を繰り返すと、最終的に長さが 4 以上の弦のない閉路が存在しないグラフができる。この操作のことをグラフの三角化と言い、結果得られる下図(b)のようなグラフのことをコーダルグラフと言う。

Hara and Takemura (2010)では、コードグラフの分解に基づいて、アルゴリズムが分解され、例えばサイクルモデルと呼ばれる多角形の疎グラフに対応するモデルの場合には、線形のオーダーまで改良できることを示した。また数値実験によって、特に次元が 100 を超えるような場合には、従来法に比べても有用なアルゴリズムであることを示した（表を参照）。

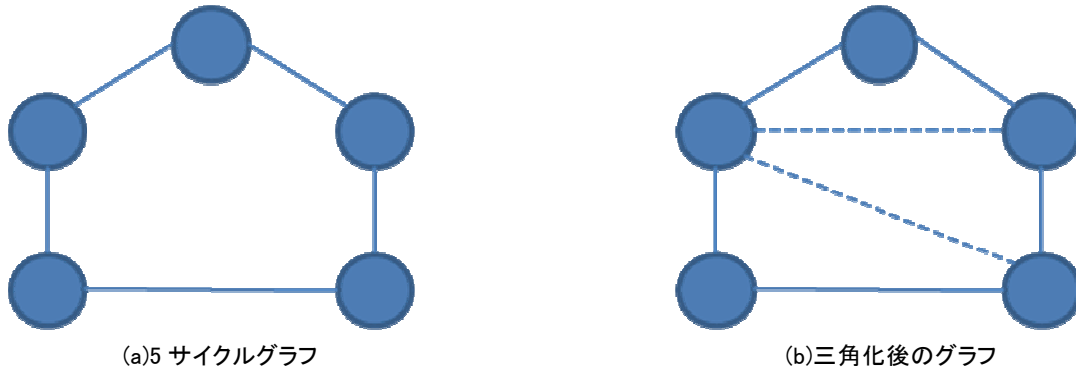


表 サイクルモデルにおける従来手法と提案手法の CPU 時間

変数の次元	提案法	従来法
5	1.590	2.261
10	3.442	2.523
50	18.63	5.482
100	38.11	17.48
200	78.87	104.91
300	120.03	361.01
500	225.94	1292.1
1000	511.60	6625.8

このような比例反復法の局所化アルゴリズムは、離散グラフィカルモデルの推定については、Malvestuto and Moscarini (2001)などで過去にも議論されてきた。一般に離散グラフィカルモデルと呼ばれるモデルでは、グラフの辺やクリークに対応した交互作用項に、構造のない自由パラメータが割り当てられる。しかし交互作用項のモデル化が有用である場面も現実には多い。そこで、Hara, Sei and Takemura (2010)では、より一般の対数線形モデルと呼ばれるクラスのモデルにおける、推測アルゴリズムの局所化について考察をおこなった。

離散のグラフィカルモデルにおけるアルゴリズム局所化も、ガウスの場合と同様にグラフの分解に基づいている。そしてその議論はグラフの単体的頂点に対応する変数の周辺和に対してモデルが閉じているという性質に依存している。しかし交互作用項に構造を入れた場合この性質は必ずしも満たされない。一方、変数の条件付き独立関係は、交互作用項の構造の有無に依存せず、元のグラフに対応する。これは交互作用項に構造の入れた場合、グラフの分解とモデルの分解が必ずしも対応しないということを意味している。

Hara, Sei and Takemura (2010)では、すべての対数線形モデルに対し、そのモデルの分解可能性に対応したグラフが一意的に定まること、推測アルゴリズムもそのグラフに基づいて自然に局所化されることを示した。その結果、遺伝子型の数を I 、分解可能性に対応するグラフの極大クリークの頂点数の最大値を c としたときに、従来法だと一回の更新で I^n 個の正規確率を更新する必要があったのが、 I^c 個

の更新で済むことになる。遺伝子のモデルのように疎なモデルの場合には、 $c \ll n$ であることから、効率的なアルゴリズムであるということが言える。

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。その中で最も基本的なモデルは、たとえば以下の式(1)で示される単一マーカー分析 (single gene model) であろう。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もあるため、現実によく用いられている (F. A. Wright and A. Kong (1997), Linkage mapping in experimental crosses: The robustness of single-gene models, *Genetics* **146**, 417-425).

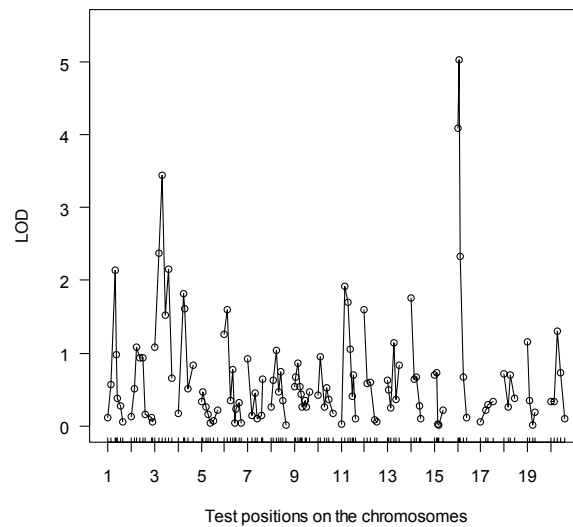


図 1: ロッドスコア

図 1 は、 F_2 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 ($\log_{10} [\text{ng/ml}]$) 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

$$w_{ij} = \begin{cases} 1 & (z_{ij} = \pm 1), \\ -1 & (z_{ij} = 0) \end{cases}$$

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、昨年度までの融合研究において、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

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

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

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

また c_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。

以下では、 $\text{EIF}_i(c)$ 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 100α パーセント点 (例えば $\alpha = 0.05$) をその閾値とする。ただしこの分布は、QTL の真のモデルに依存するので簡単に推定することはできない。ここでは QTL が存在しないというモデルの下で、この分布の近似を考える。

モデル(1)を行列表示して、

$$y = X_j \theta + \varepsilon, \quad X_j = \begin{pmatrix} z_{1j} & w_{1j} & 1 & u_1 \\ \vdots & \vdots & \vdots & \vdots \\ z_{Nj} & w_{Nj} & 1 & u_N \end{pmatrix}_{N \times 4}$$

と書く。 $\hat{\sigma}_j = \sigma^2(1 + O_p(n^{-1/2}))$ に注意すると、 $e_i = (0, \dots, 0, 1, 0, \dots, 1)'$ (i 番目は 1、他は 0) を用いて

$$\hat{\ell}_{ij} = -\frac{1}{2\hat{\sigma}_j^2}(e_i' H_j y) - \frac{1}{2} \log \hat{\sigma}_j^2 = -\frac{1}{2\sigma^2}(e_i' H_j y) - \frac{1}{2} \log \sigma^2 + O_p(n^{-1/2}),$$

$$\tilde{\ell}_i = -\frac{1}{2\tilde{\sigma}^2}(e_i' H_0 y) - \frac{1}{2} \log \tilde{\sigma}^2 = -\frac{1}{2\sigma^2}(e_i' H_0 y) - \frac{1}{2} \log \sigma^2 + O_p(n^{-1/2}),$$

ただし

$$H_j = I - X_j(X_j' X_j)^{-1} X_j', \quad H_0 = I - X_0(X_0' X_0)^{-1} X_0', \quad X_0 = \begin{pmatrix} 1 & u_1 \\ \vdots & \vdots \\ 1 & u_N \end{pmatrix}$$

と書ける。帰無仮説の下では、

$$y = X_0 \begin{pmatrix} \mu \\ \nu \end{pmatrix} + \varepsilon, \quad \varepsilon \sim N_n(0, \sigma^2 I)$$

であるが、 $H_j X_0 = 0$, $H_0 X_0 = 0$ に注意し、また ε/σ を改めて ε とおくと、 $\varepsilon \sim N(0, I_N)$ 、

$$\begin{aligned} \text{EIF}_i(j) &\approx -\frac{N}{2 \log 10} \left\{ (e_i' H_j \varepsilon)^2 - (e_i' H_0 \varepsilon)^2 - \frac{1}{N} \sum_{i=1}^N (e_i' H_j \varepsilon)^2 + \frac{1}{N} \sum_{i=1}^N (e_i' H_0 \varepsilon)^2 \right\} \\ &= -\frac{N}{2 \log 10} \varepsilon' \left\{ H_j e_i e_i' H_j - H_0 e_i e_i' H_0 - H_j \left(\frac{1}{N} \sum_{i=1}^N e_i e_i' \right) H_j + H_0 \left(\frac{1}{N} \sum_{i=1}^N e_i e_i' \right) H_0 \right\} \varepsilon \\ &= -\frac{N}{2 \log 10} \varepsilon' \left\{ H_j e_i e_i' H_j - H_0 e_i e_i' H_0 - H_j + H_0 \right\} \varepsilon. \end{aligned}$$

これより

$$\text{EIF}_i(c) \approx -\frac{N}{2 \log 10} \varepsilon' \left\{ \sum_{j \in J} c_j (H_j e_i e_i' H_j - H_0 e_i e_i' H_0 - H_j + H_0) \right\} \varepsilon.$$

これを用いて、シミュレーションによって $\max_{i=1, \dots, N} |\text{SEIF}_i(c)|$ の分布を推定し、閾値を設定することができる。

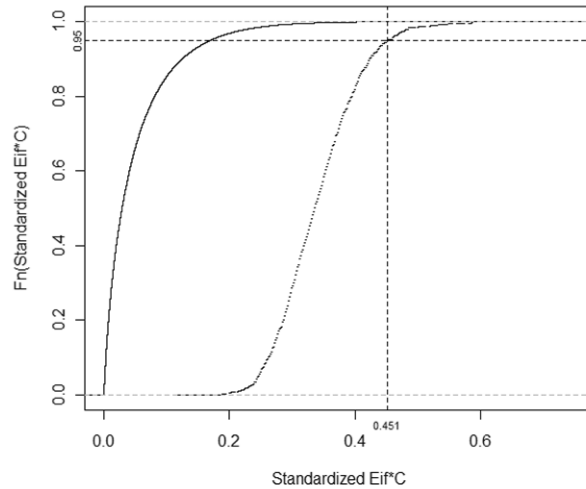


図 2: 基準化経験影響関数とその最大値の経験分布

図 2 は、アディポネクチンのデータについて、スコア $c = (1.04, -2.36, 1.31)$ を用いて計算した $\max_{i=1, \dots, N} |\text{SEIF}_i(c)|$ の分布の推定値（経験分布）である。95%点は 0.451 である。

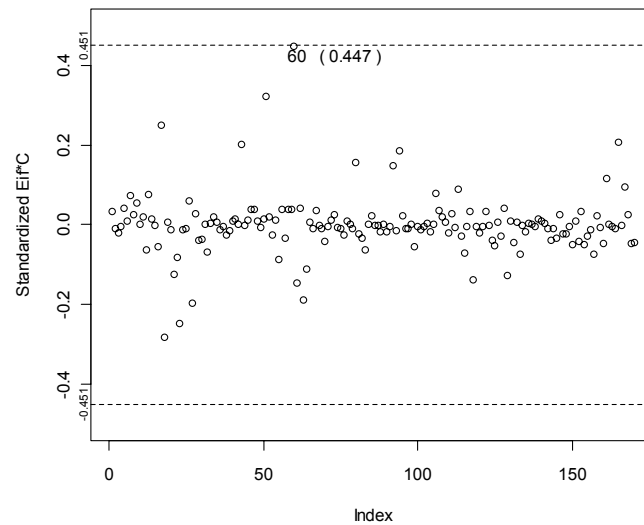


図 3: 凸性の基準化経験影響関数

図 3 は、 $\text{SEIF}_i(c) (i = 1, \dots, N = 170)$ のプロットである。絶対値が閾値に近い個体として、個体 60 が検出されている。この個体 60 ($\text{SEIF} = 0.447$) は、ロッドスコアの凸性の形状に影響を与える個体であり、注意すべき個体であることが分かる。

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」
 マウス複雑形質（生体内構造、骨格形態、行動パターンなど）の表現型数値データの統計解析手法の開発を行った。また、開発した手法を用いて野生系統を含む実験交配集団や 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収集を実施した。

3-1 : 3D 画像による体脂肪計測法の定量化とデータマイニング（中谷、北本、阿部、高田、城石）

本研究の目的は、第 1 に、融合研究プロジェクト研究第 I 期に行った生体イメージングからの正確な表現型測定の手法を一段と改善することである。特にマウスの全身 CT（Computed Tomography: コンピュータ断層撮影）画像から自動的に分離した内臓および皮下脂肪、さらには脂肪組織以外の臓器の数値データ計測アルゴリズムを完成させる。第 2 に、この手法を用いて得られたマウスの数値データの空間的な分布を量的形質として定量・測定を行い遺伝解析（QTL 解析）に繋げることである。

平成 22 年度は、基準系統として C57BL/6J(B6) 系統をサンプルにアルゴリズムの開発を進めた。また、10 および 25 週齢の個体より得られた CT 画像セットを使用し、開発中の自動体脂肪計測アルゴリズムにより得られた結果と同一個体の解剖による臓器重量データと比較した。

B6 系統の 25 週齢雌のある個体での各スライスでの測定値をプロファイル化した結果を図 1A に示す。

測定には、全身に渡る体腔境界の位置情報を用いた。これにより任意の断面での内臓脂肪組織と皮下脂肪組織および筋肉の空間的な分布を可視化することが可能になった。

図 1B は、脊椎付近での矢状断面での分布状態、図 1C は、脊椎腹側での冠状断面での分布状態を図示したものである。この俯瞰図により任意の領域内での分布状態を得ることができるため、特定の臓器

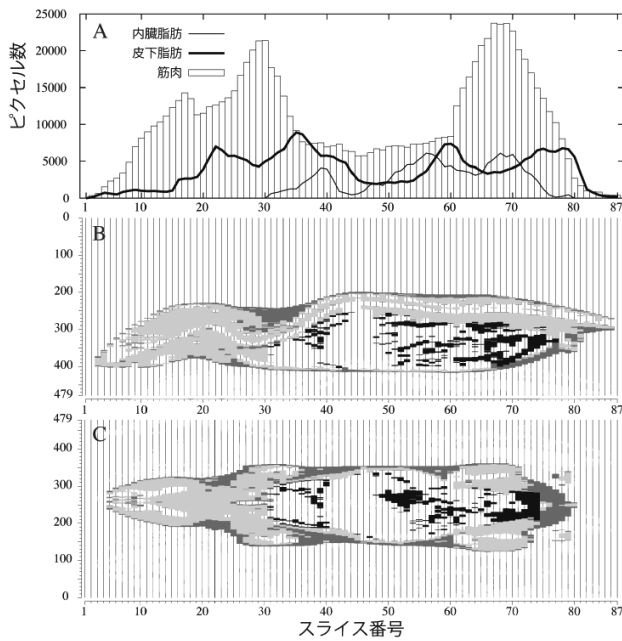


図 1A 脂肪組織量と筋肉量のプロファイル。A) 各スライスでの内臓脂肪組織量(細線グラフ)と皮下脂肪組織量(太線グラフ)および筋肉量(棒グラフ)。B) 矢状断面で内臓脂肪組織(黒色)と皮下脂肪組織(濃灰色)および筋肉(薄灰色)の空間的分布。薄灰色部分には、脳や脊髄も含む。C) での縦軸 250 の位置での断面。C) 同様に冠状断面での空間的分布。B) での縦軸 350 の位置での断面。

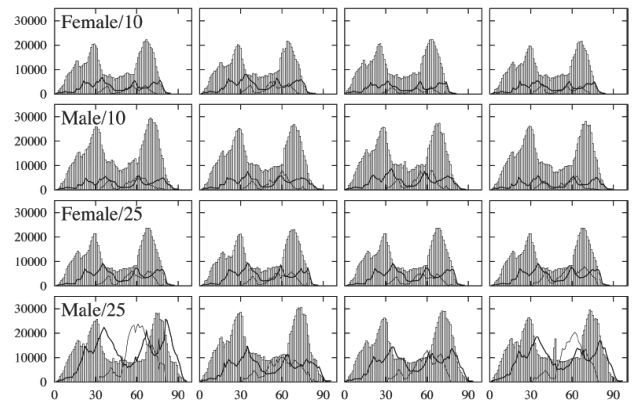


図 2 週齢と性別ごとの脂肪組織量と筋肉量のプロファイル。各行は、性別 (Female: 雌 / Male: 雄) と週齢 (10: 10 週齢 / 25: 25 週齢) ごとの 4 サンプルに相当。縦軸は各組織に相当するピクセル数。

の周囲に存在する脂肪組織の組成を定量するための基礎データとして使用することができた。評価用サンプル 16 個体について図 1A に相当するプロファイルを生成した結果を図 2 に示す。

16 個体それぞれに関して、内臓脂肪組織と皮下脂肪組織に相当するピクセル数の総和と実際に解剖して得られた組織重量の関係を図 3 に示す。

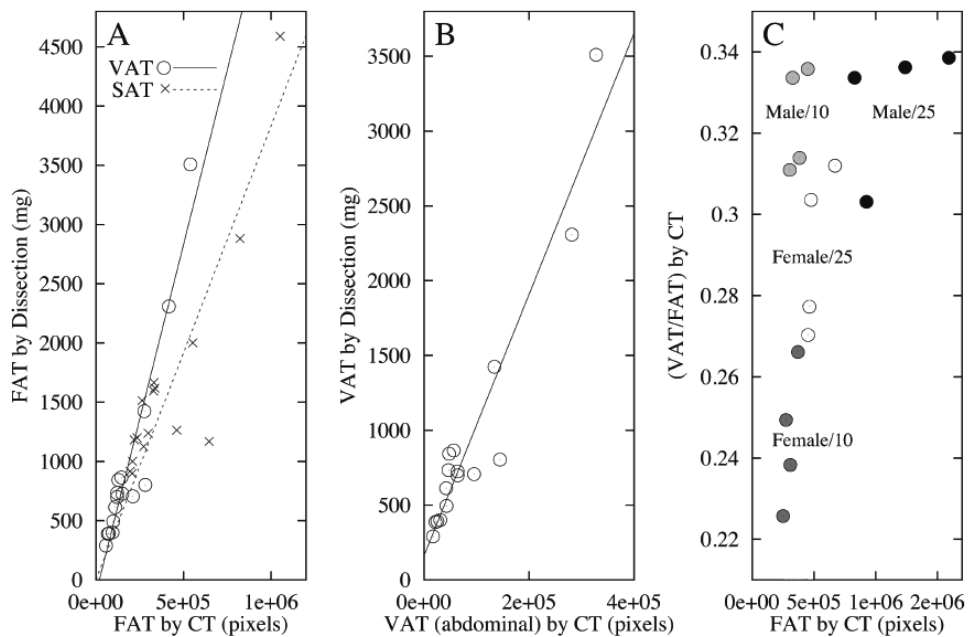


図 3 計測値間の相関関係。A) X 線 CT による脂肪組織量(ピクセル数)と解剖による脂肪組織重量 (mg) の関係。○は内臓脂肪組織での関係(相関係数=0.95)、×は皮下脂肪組織での関係(相関係数=0.88)を示す。実線 ($y = 0.0059x - 115.13$) と破線 ($y = 0.0038x + 1.01$) は両脂肪組織での回帰直線を示す。B) X 線 CT による腹部内臓脂肪組織量(ピクセル数)と解剖による内臓脂肪組織量 (mg) の関係(相関係数=0.95)。実線 ($y = 0.0087x + 162.63$) は回帰直線を示す。C) X 線 CT による総脂肪組織量(ピクセル数)と X 線 CT による内臓脂肪組織率の関係。週齢と性別の情報を付記している。VAT:内臓脂肪組織、SAT:皮下脂肪組織。

いずれも X 線 CT での測定結果はピクセル数によるものである。図 3A は、X 線 CT による脂肪組織量（横軸）と解剖による脂肪組織重量（縦軸）の関係を内臓脂肪組織と皮下脂肪組織について示したものであり、両者の間には高い相関関係が認められた（それぞれの相関係数は 0.95 と 0.88）。図中に回帰直線を示す通り、X 線 CT による評価結果を用いて解剖による結果を推定することが可能であることを示唆する。ただし、両回帰直線の傾きは異なっており（0.0059 と 0.0087）、ピクセルあたりの脂肪組織重量がこれらの脂肪組織で異なることを示す結果となっている。

これは、脂肪量の多い個体の値に影響を受けたものであり、皮下脂肪組織量に関して、X 線 CT での過大評価、もしくは、解剖での過小評価のいずれかが推測される。また、図 3B は、X 線 CT による腹部領域の内臓脂肪組織量（横軸）と解剖による全身の内臓脂肪組織量（縦軸）の関係を示している（相関係数は 0.95）。以上の散布図中で脂肪量が多い 4 個体は、いずれも 25 週齢の雄である。これらの 4 個体は、他の性別と週齢と比較して、個体間の変異が大きくなっているが、概ね良好な結果を得ていると言える。また、図 3C は、X 線 CT による総脂肪組織量（横軸）と X 線 CT による内臓脂肪組織率（縦軸）の関係を示している。内臓脂肪組織率は、内臓脂肪組織に相当するピクセル数の全脂肪組織に相当するピクセル数に対する比率である。散布図上で、週齢と性別のクラスごとに個体がクラスタ化しており、仮にそれぞれのクラスで共通の性質をもっているとするならば、本手法がそれらの性質を分離するに足る精度を実現していることを示唆している。

最後に、画像の生成など、データ収集の自動化については大体 1CPU で数時間のオーダーで（全身のスライス画像について）完了することができた。今後は、F2 集団を始めとする多様な遺伝的背景をもつ個体群での定量を行うことにより、表現型に関連した遺伝子座を探索する。

3-2：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、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、哺乳類ではじめて、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。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も明らかな日本産野生由来 MSM/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。

平成 22 年度は、(B6xMSM) F1 および (MSMxB6) F1 個体の交配・生産を開始した。2011 年 3 月の時点で、(B6xMSM)F1: 87 匹（雄 41 および雌 46 匹）、(MSMxB6)F1: 46 匹（雄 24 および雌 22 匹）、計 133 匹について試料の採取ならびに解剖時に臓器重量などの基本的な表現型データの収集を行うことができた。具体的に F1 個体は通常餌（CE-2, 346.8kcal/100g, 4.6% from fat）で 10 週間飼育しそ

解剖学的観察	血液生化学	血液成分-4	
	体躯	血液成分-1	（肝臓機能）
	体重	（脂質）	総ビリルビン
体長	総コレステロール	ALT	
尾長	HDL-コレステロール	AST	
身体の構成	Non-HDL-コレステロール	乳酸デヒドロゲナーゼ	
	中性脂肪	アミラーゼ	
	体脂肪重量	血液成分-2	血液成分-5
	除脂肪体重	（ミネラル）	総タンパク質
体脂肪率	無機リン	アルブミン	
肥満度	カルシウム	グロブリン	
カリウム			
臓器重量	心臓	血液成分-3	
	肝臓	（腎臓機能）	
	脾臓	クレアチニン	
	腎臓	尿素窒素	
	精巣	尿酸値	

図1 本解析で観察している表現型パラメータ。各項目の測定は 10 週齢を対象にしている。

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、収集を行った。血液はドライケム（フジフィルムメディカル）およびベトスキャン（Abaxis社）による各種血中パラメータの測定を開始した。遺伝子発現解析に使用する資料は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。なおデータ収集と資料保存は平成23年度も継続して行っており、最終的なデータがでそろった時点で各組み合わせが表現型にどのように影響を与えるか統計解析を行い、遺伝子発現解析との相関解析を解析することを目指す。

我々は、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表現型データおよび、RNA 採取用組織の収集・保存を開始しており、次年度以降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考えます。

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。

平成22年度は、マウスの社会行動を調べるテストの一つとして、ソーシャルインタラクションテストを行った（図1）。このテストでは、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。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。一方、このテストを自動解析するソフトも存在し、そこでは2個体の位置をトラッキングし、2個体がある任意の距離以内に近づいた際に社会的接触があったと抽出する。しかし、この設定された任意の距離が、実際に生物学的な意味を伴うかは疑問である。本研究では、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



図1. Social interaction テスト

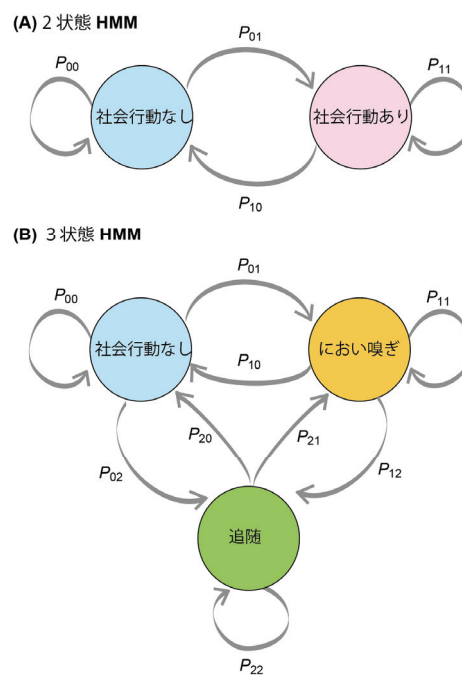


図2. 隠れマルコフモデル(HMM)

みた。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、追従、攻撃行動）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「追従」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、追従で 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。本研究ではまた、攻撃行動も含めた 5 状態の推測も試みたが、攻撃行動の出現がかなり低かったことにより、正しく学習させることができなかった。

続いて、2 状態 HMM による状態推定を用いて、マルコフ推移確率におけるマウス社会行動の系統比較を行った。社会行動の有無の状態間の推移確率 P01（無 → 有）と P10（有 → 無）の 2 次元空間において、各系統の特徴を記述した。この解析では、C57BL/6J（B6）系統と MSM 系統から作出された一連のコンソミックマウス系統を用いた。これまで、MSM 系統は B6 系統よりも 3 倍近く長い社会的接触を示すことが示されていた。マルコフ推移確率空間において、B6 系統は左上（P01 が低く、P10 が高い）に位置した（図 3）。これはつまり、この系統が社会行動を開始しにくく、社会行動が出現した場合それがすぐに終了することを意味する。一方 MSM 系統は右下（P01 が高く、P10 が低い）に位置しており、この系統は社会行動を開始しやすく、その状態が比較的長く継続することを示す。一連のコンソミック系統をこのマルコフ空間上にプロットした結果、多くの系統が B6 と類似した位置か、MSM と B6 の間に存在するということが明らかになった。興味深いのが、6 番染色体（6C）と 17 番染色体のコンソミック系統で、これらは自動解析によって社会的接触が同様に長い系統であったが、マルコフ空間上では異なる特徴を示した。17 番染色体コンソミックは MSM とほぼ同様の位置にプロットされた一方、6 番染色体コンソミックは左下（P01、P10 ともに低い）に位置した。この結果は 6 番染色体コンソミック系統は社会行動を開始させにくい、いったん出現するとそれが長く持続するという特徴を持つことを示す。また、性差についても注目すべき特徴が観察された。マルコフ空間上において、13 番染色体と 15 番染色体のコンソミック系統では有意な性差が観察され、雌は B6 寄り、雄は MSM 寄りにプロットされた（図 4）。実験者の観察によりこれらの系統ではオスにおいてのみ攻撃行動が出現することが明らかになっており、このマルコフ空間における性差は、社会行動の質の違いを抽出することができたといえる。

本研究では、HMM を用いてマウスの社会行動状態の推測を人間の観察と高い一致率で行うことに成功した。また、

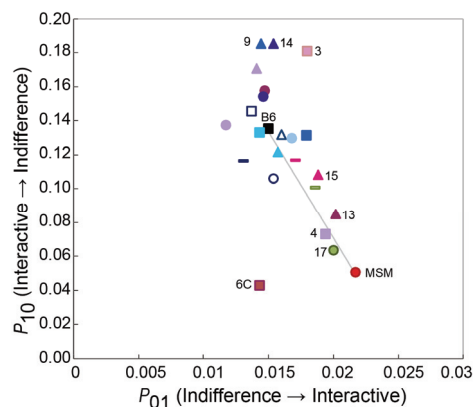


図 3. マルコフ推移確率空間における各系統の平均値プロット。B6-MSM 間を灰色の線でつないでいる。

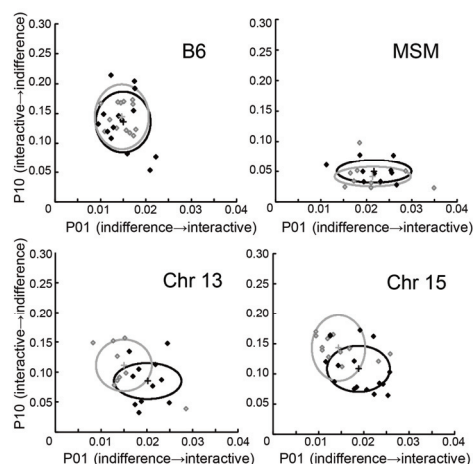


図 4. マルコフ推移確率空間における性差。各点は各個体の値、○は誤差楕円体を示す。雄データは黒、雌データは灰色で示す。

マルコフ推移確率を用いて系統比較を行った結果、自動解析による2個体間の距離データでは検出できなかった系統差が明らかになった。それらの系統差は、実際に人間の観察では抽出されていることから（例えば攻撃行動など）、より正しく社会行動の質の違いを反映している可能性が高い。最後に、本研究ではこれらの解析を行うためのソフトウェアを作出した。このソフトウェアは、本共同研究の成果が論文として発表後、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。

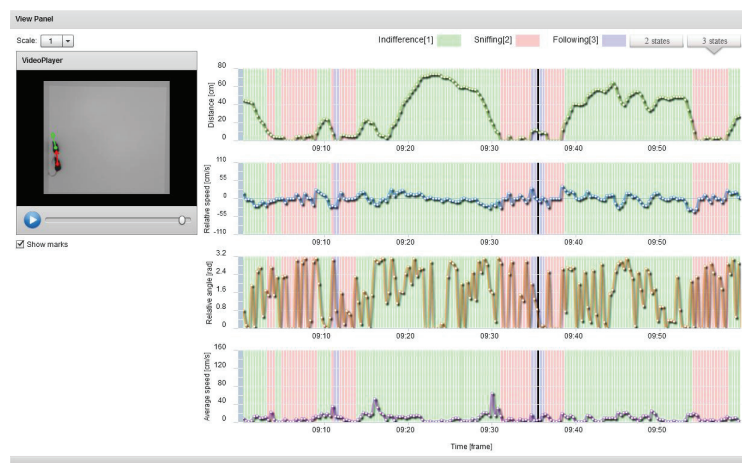


図 5. Social Interaction 行動解析ソフトウェア

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、周波数と振幅の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、周波数と振幅の3次元で観測されている。そのうち、振幅が声の大きさを表すため、決めた振幅の以上のもので、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。さらに、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予定である。

今後の具体的な研究手法と解析計画は以下ようになる：

Step 1. 3次元データ（周波数、時間にたいする信号強度データ）について、ノイズ除去のための平滑化を行う。（移動平均、対数や平方根を取るなどの手法を用いる）

Step 2. 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、直交多項式、ウェーブレット法など）

Step 3. Step 2 で得られた関数データ（音声波形データ）をクラスタリングし、声のパターンの類型化を行う。

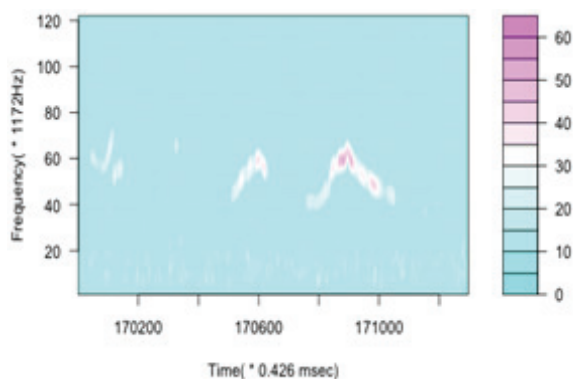


図 1 移動平均と平方根処理を行った3次元の超音波データ

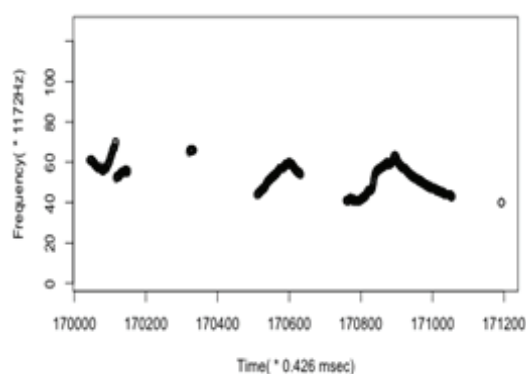


図 2 ノイズの除去とピーク周波数の推定を行い得られた発声の波形

Step 4. それを元に、今後どのようなデータを取得するか、検討する。例えば、外部刺激を与えて、データを取得するなどが考えられる。

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、解析集団の作製方法を検討した。

3-5 : 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、久保、倉田)

この研究は、栽培イネ *Oryza sativa* の直接の祖先種である *O. rufipogon* の一年生、多年生グループの代表的な系統、W0106 と W1921 を用い、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) を同じ解析集団にて行うことにより、表現形質を形成する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抽出し、表現形質とその遺伝子発現のネットワークを解明することである。

栽培イネ *O. sativa* は、形質やゲノム塩基配列などによりジャポニカ、インディカなどの亜種に分類されている。一方栽培イネ祖先種である野生イネ *O. rufipogon* は、分類学上の亜種は定義されていないが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亜種に相当するような集団構造が存在することが示唆されている。融合研究プロジェクト第 I 期には *O. rufipogon* について、量的形質についてデータを取得した。柱頭長、種子稔性、穂の形 (着粒数と相関あり)、発根・再生力など多くの形質が、大きく多年生系統群、一年生系統群の 2 つのグループに分かれた (図 1)。このように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。そのため表現型の離れた代表的な 2 品種を用いた交雑後代の実験集団による遺伝解析も平行して行うこととし、一年生の代表品種として W0106、多年生の代表品種として W1921 を選択した。両品種の形態を図 2 に示す。

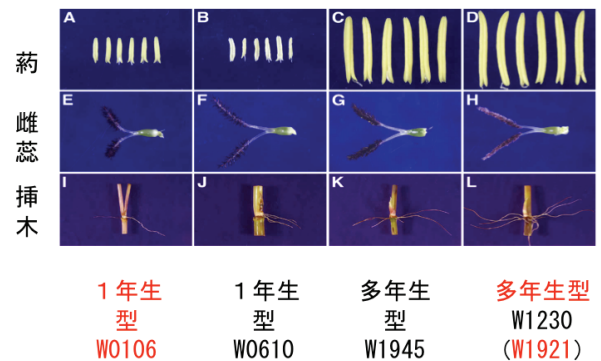


図 1、*O. rufipogon* の葯 (A-D)、雌蕊 (E-H)、挿木 (I-L)。1 年生型の W0106 (A,E,I)、W0610 (B,F,J)。多年生型の W1945 (C,G,K)、W1230 (D,H,L)。

22 年度は QTL の検出能、組換え分解能、RNA-seq の費用労力等を考慮した解析集団の検討をし、育成を進めた。解析する実験集団は、形質調査や遺伝子発現量測定の実現性を得るため遺伝的に固定化された組換え固定化集団が望ましい。一般に QTL や eQTL の検出能や分解能は用いる集団の大きさに依存し、集団のサイズは多いほど良い、しかし解析集団の各個体の全発現遺伝子の発現量及び全ゲノムの遺伝型の測定にかかる費用や労力を考慮すると集団の大きさは 200 程度とした。次に、解析集団の解像度を考えた。集団のサイズが一定のとき検出される QTL の分解能を高めるためにはより遺伝型の組換えが多い個体の育成が必要となる。組換え固定化集団の育成手法としては、単粒系統 (single-seed descent) 法による組換え自殖系統 (recombinant inbreed line) の育成、または葯培養による倍加半数体系統 (doubled haploid line)

1 年生型 W0106 多年生型 W1921

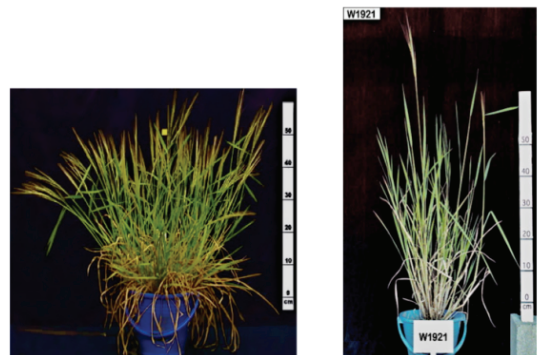


図 2、*O. rufipogon* 1 年生型の W0106、多年生型の W1921 の草型。

が用いられる。イネは自殖の傾向が強いため RILs の育成は比較的労力はかからないが、遺伝型固定のための 10 世代以上の時間がかかる。さらに個体当たりの観測可能な組換え数は、世代経ることによって生じる新たな組換え箇所と遺伝型ホモ型への固定化のため減る組換え箇所が釣り合うため F₂ 個体と同じである。200 個体を用いた時の最大の分解能は 0.25cM でイネの場合の平均物理距離は 63kb でそこには約 6 個の遺伝子が含まれることが予想される。F₁ 個体の葯培養による DHL では、労力はかかるが固定化集団の育成時間は短い。しかし分解能は F₂ 集団の半分で 0.5cM (平均物理距離 125kb) になってしまう欠点を持つ。自殖を進めることによりヘテロ型半数の遺伝型は次世代で固定化されるが、自殖後代で新たに生じる観測可能な組換えはその親がヘテロ型であった箇所のみである。異なる系統間の交雑を行えばヘテロ型の割合を維持出来るので観測可能な組換えは増えるが、固定化が遅れるだけでなくその交雑両親として必要な系統数は直接の子孫系統数の 2 倍必要となる。以上の考察をもとに、

図 3 に示す他殖(F₃)と自殖(F₄)を組み合わせた計画による解析集団の育成を開始した。

22 年度は 500 以上の F₁ の自殖種子 (F₂)を確保したので、23 年度以降は異なる F₂ 個体を用いた F₃ 他殖系統群を育成し、自殖 F₄ 系統の葯培養による DHLs の育成を行う予定である。この DHLs 200 系統を用いた集団の分解能は

0.22cM (平均物理距離 56kb) である。葯培養による DHL 化が全ての系統で成功する保証はないので、人工気象器や熱帯での世代促進による RILs の育成も併せて進めて行く予定である。この RILs 200 系統を用いた分解能は 0.2cM (平均物理距離 50kb) となる。

次に次世代シーケンサーにより取得される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が課題となる。22 年度は親品種 W106,W1921 のゲノム配列情報を取得するため、次世代シーケンサーよりゲノムデータを取得した。Nipponbare のゲノム配列をマップして各品種のゲノム配列をリシーケンスして塩基配列多型、挿入欠失配列等を調査した。今年度以降、2 ゲノムのドラフト配列をもとに発現解析および配列変異解析を行う予定である。

3-6 : 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、久保、倉田)

また、融合研究プロジェクト第 I 期により開発した構造多型を考慮したマイクロアレイによる発現解析手法をもとに、ゲノム配列が既知の異なるイネ 2 品種 (ジャポニカ日本晴とインディカ 9311) の苗条 (shoot)、幼穂 (young panicle) の遺伝子発現量の差の解析を行った。マイクロアレイによる発現解析はプローブのハイブリダイゼーションシステムによるため、Affymetrix rice genome array の約 63 万のプローブのうち日本晴で想定されている全ての遺伝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 万プローブのみを用い 44,447 遺伝子の発現量比較を行った。

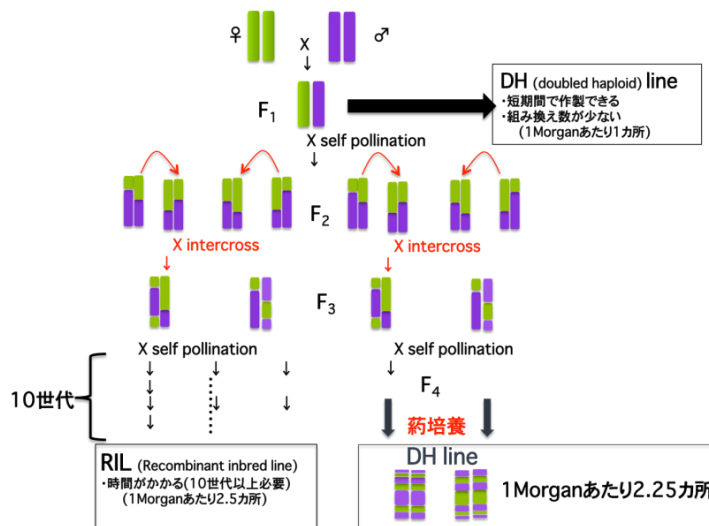


図3. 様々な組換え固定化集団による解像度の比較

表1に shoot と young panicle で高発現している遺伝子の数と発現量が異なる遺伝子の数をまとめた。それぞれの組織で高発現する遺伝子数は品種組織間で大きな差はなく今回解析対象とした遺伝子の 30%である。発現量比較の対称とした遺伝子のうち 93-11 のゲノム配列をもとにアノテーションされている遺伝子数は 28,052 で対象

表 1 Affymetrix Rice Genome Array による 44,447 遺伝子の品種間発現差

高発現遺伝子数 (>log ₂ 7)	日本晴	93-11 (Annotated)	発現量が異なる
Young Panicle	13,475	13,442 (12,207)	953
Shoot	13,535	13,472 (12,332)	972
共通			487

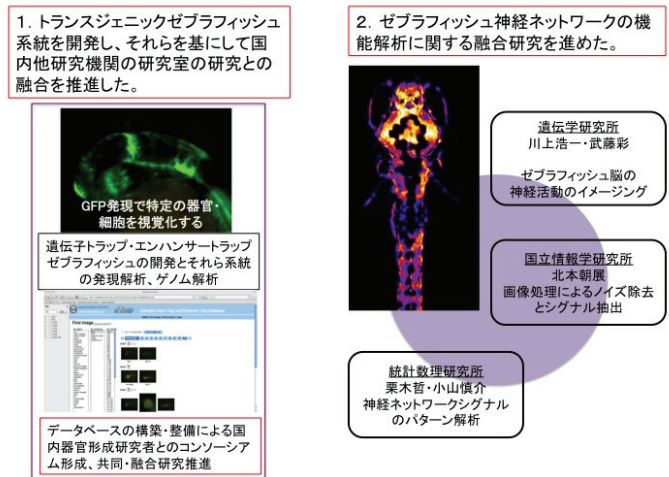
遺伝子の 63%であるが高発現遺伝子の約 90%は 93-11 でもアノテーションされている遺伝子であった。日本晴または 93-11 で高発現する遺伝子のうち品種間で発現量の差が見られるものは約 7%程度であった。この品種間で発現量に差のある遺伝子の約半数の 487 個は shoot, young panicle でともに発現量差が確認された遺伝子であった。この 487 個の遺伝子について公的な遺伝子発現データベースのデータをもとに解析すると、353 遺伝子は組織を問わずジャポニカ品種で高発現量し、インディカ品種ではいずれの組織においても発現は観測されなれない。また、71 遺伝子は逆にインディカ品種で高発現しているが、ジャポニカ品種では発現していなかった。これらの 424 遺伝子は一方の亜種では遺伝子そのものがないか、恒常的に発現が抑制されている遺伝子と推定された。現在この内容を論文に作成している。

3-7: 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

トランスポゾンのランダムな挿入に基づくトランスジェニック系統、ノックアウト変異系統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を行う。さらに発現パターンの数理解析などを行うため、H22 度は以下の研究を進めた。

(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行い、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、合計 800 系統作製した。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。研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。脊椎の運動神経、脳内の神経活動の時空間パターンを情報研の北本博士と共同でデータ取得するためのプログラムの開発を開始し、脊椎の運動神経の発火パターンのデータを取得するプログラムを作製しつつある。またそのような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開始した。



3-8: 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助的役割を担い、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

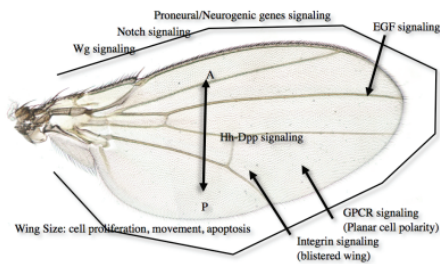


図 1. ショウジョウバエの翅形成に働くシグナル伝達系

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。

H22年度は遺伝研の RNAi（遺伝子をノックダウンする）システム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。新たに 3,100 系統の変異翅の画像を取得した。これまでに 12,000 の RNAi 系統を処理し、画像取得が可能であったものは 9,000 系統であった。現在進行中のサンプルを除き、大きく形態が変化するために平面としての画像取得が不可能なもの、あるいは個体が致死となって翅の形成が観察されなかったものが 1,500 系統存在することも明らかとなった。

これらの画像に対して、微細な形態異常を検出するプログラムを作出することが本研究テーマの主要な計画の一つであるが、今年度は翅の面積を計算するためのプログラムの開発を行った。図 2 には、首都大に維持されている GS（遺伝子の強制発現が可能）系統の例を掲載している。このプログラムを用いたデータの検証はこれからである。

実際の翅形態の異常は、翅脈の有無やその幅広さ、また辺縁部の切れ込みなど、シグナル伝達系の種類によって特徴が異なっている。これらの特徴をどのように取得していくかがこれからの問題である。さらにこれらの画像データは各系統 20 サンプル、総計 20 万サンプル以上にのぼる。このような膨大なサンプルを処理する新たなプログラムの開発については統計学的なアプローチが必須であると言える。

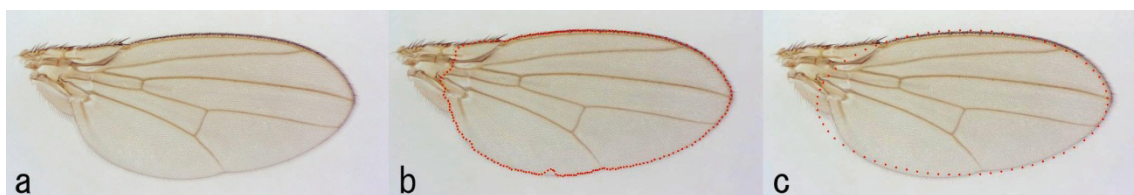


図 2. a:GS80849 変異体の翅、 b:輪郭取得、 c:面積を計算するための近似楕円

[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. Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, Kawashima K, Minami C, Muraki A, Nakazaki N, Takahashi C, Nakayama S, Kishida Y, Kohara M, Yamada M, Tsuruoka H, Sasamoto S, Tabata S, Aizu T, Toyoda A, Shin-I T, Minakuchi Y, Kohara Y, Fujiyama A, Tsuchimoto S, Kajiyama S, Makigano E, Ohmido N, Shibagaki N, Cartagena JA, Wada N, Kohinata T, Atefeh A,

- Yuasa S, Matsunaga S, Fukui K. Sequence Analysis of the Genome of an Oil-Bearing Tree, *Jatropha curcas* L. *DNA Res.* 2010 Dec 13. [Epub ahead of print] PubMed PMID: 21149391.
2. 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.* 2010 Nov 18. [Epub ahead of print] PubMed PMID:21085194.
 3. Taji T, Komatsu K, Katori T, Kawasaki Y, Sakata Y, Tanaka S, Kobayashi M, Toyoda A, Seki M, Shinozaki K. Comparative genomic analysis of 1047 completely sequenced cDNAs from an Arabidopsis-related model halophyte, *Thellungiella halophila*. *BMC Plant Biol.* 2010 Nov 24;10:261. PubMed PMID: 21106055; PubMed Central PMCID: PMC3017837.
 4. Shang WH, Hori T, Toyoda A, Kato J, Pependorf K, Sakakibara Y, Fujiyama A, Fukagawa T. Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Res.* 2010 Sep;20(9):1219-28. Epub 2010 Jun 9. PubMed PMID: 20534883; PubMed Central PMCID: PMC2928500.
 5. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Takamatsu K, Chuma S, Kojima-Kita K, Shiromoto Y, Asada N, Toyoda A, Fujiyama A, Totoki Y, Shibata T, Kimura T, Nakatsuji N, Noce T, Sasaki H, Nakano T. MVH in piRNA processing and gene silencing of retrotransposons. *Genes Dev.* 2010 May;24(9):887-92. PubMed PMID:20439430; PubMed Central PMCID: PMC2861188.
 6. Nishito Y, Osana Y, Hachiya T, Pependorf K, Toyoda A, Fujiyama A, Itaya M, Sakakibara Y. Whole genome assembly of a natto production strain *Bacillus subtilis* natto from very short read data. *BMC Genomics.* 2010 Apr 16;11:243. PubMed PMID: 20398357; PubMed Central PMCID: PMC2867830.
 7. Ohnishi Y, Totoki Y, Toyoda A, Watanabe T, Yamamoto Y, Tokunaga K, Sakaki Y, Sasaki H, Hohjoh H. Small RNA class transition from siRNA/piRNA to miRNA during pre-implantation mouse development. *Nucleic Acids Res.* 2010 Aug;38(15):5141-51. Epub 2010 Apr 12. PubMed PMID: 20385573; PubMed Central PMCID: PMC2926599.
 8. Takahashi S, Takagi H, Toyoda A, Uramoto M, Nogawa T, Ueki M, Sakaki Y, Osada H. Biochemical characterization of a novel indole prenyltransferase from *Streptomyces* sp. SN-593. *J Bacteriol.* 2010 Jun;192(11):2839-51. Epub 2010 Mar 26. PubMed PMID: 20348259; PubMed Central PMCID: PMC2876496.
 9. Toh H, Oshima K, Toyoda A, Ogura Y, Ooka T, Sasamoto H, Park SH, Iyoda S, Kurokawa K, Morita H, Itoh K, Taylor TD, Hayashi T, Hattori M. Complete genome sequence of the wild-type commensal *Escherichia coli* strain SE15, belonging to phylogenetic group B2. *J Bacteriol.* 2010 Feb;192(4):1165-6. Epub 2009 Dec 11. PubMed PMID: 20008064; PubMed Central PMCID: PMC2812965.
 10. Matsui H, Taniguchi Y, Inoue H, Kobayashi Y, Sakaki Y, Toyoda A, Uemura K, Kobayashi D, Takeda S, Takahashi R. Loss of PINK1 in medaka fish (*Oryzias latipes*) causes late-onset decrease in spontaneous movement. *Neurosci Res.* 2010 Feb;66(2):151-61. Epub 2009 Nov 4. PubMed PMID: 19895857.
 11. Sakudoh T, Nakashima T, Kuroki Y, Fujiyama A, Kohara Y, Honda N, Fujimoto H, Shimada T, Nakagaki M, Banno Y, Tsuchida K.: Diversity in Copy Number and Structure of a Silkworm

Morphogenetic Gene as a Result of Domestication. *Genetics*. (2011) Jan 17. [Epub ahead of print]
PMID: 21242537

12. Shinya Ohta, Jimi-Carlo Bukowski-Wills, Luis Sanchez-Pulido, Flavia de Lima Alves, Laura Wood, Zhuo Chen, Melpi Platani, Lutz Fischer, Damien F. Hudson, Chris P. Ponting, Tatsuo Fukagawa, William C. Earnshaw, and Juri Rappsilber. "The protein composition of mitotic chromosomes determined using multi-classifier combinatorial proteomics." *Cell* Vol. 142, 810-821 (2010).
13. Susana Abreu Ribeiro, Paola Vagnarell, Yimin Dong, Tetsuya Hori, Bruce F. McEwen, Tatsuo Fukagawa, Cristina Flors, and William C. Earnshaw. "A super-resolution map of the vertebrate kinetochore." *Proceeding of National Academy of Science USA* Vol. 107, 10484-10489 (2010).
14. Katherine Johnston, Ajit Joglekar, Tetsuya Hori, Aussie Suzuki, Tatsuo Fukagawa, and E. D. Salmon. "Vertebrate Kinetochore Protein Architecture: Protein Copy Number." *Journal of Cell Biology* Vol. 189, 937-943. (2010).
15. Inagaki, S., Miura-Kamio, A., Nakamura, Y., Lu F., Cui X., Cao X., Kimura, H., Saze, H., Kakutani, T., Autocatalytic differentiation of epigenetic modifications in Arabidopsis genome, *EMBO J* 29 3496-3509, 2010
16. Fujimoto, R., Sasaki, T., Kudoh, H., Taylor, JM., Kakutani, T., Dennis, ES., Epigenetic variation in the FWA gene within the genus Arabidopsis, *Plant J*. in press
17. Toyama, M., Matusda, K., Kakutani, T., Terao-Morita, M., Tasaka, M., Developmental changes in crossover frequency in Arabidopsis, *Plant J*. 14 81-87, 2011
18. 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. 薦田多恵子, 藤山秋佐夫 (分担執筆): 生物学事典 (石川統, 黒岩常祥, 塩見正衛, 松本忠夫, 守隆夫, 八杉貞男, 山本正幸編集), 2010, 東京化学同人
2. 藤山秋佐夫(分担執筆): 第1章 ゲノムの進化とヒトへの道「現代生物科学入門第2巻, ゲノム科学の展開」, pp1-24, 2011, 岩波書店

[解説・総説]

1. Inagaki, S., Kakutani, T., Control of genic DNA methylation in Arabidopsis, *J. Plant Research* 123 299-302, 2011
2. Saze, H., Kakutani, T., Differentiation of epigenetic modifications between transposons and genes. *Curr Opin in Plant Biol* 14, 81-87, 2011

[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.bioportal.jp/>

<会議発表等>

[招待講演]

国 際

1. 藤山秋佐夫: ヒトとチンパンジーの個人ゲノム解読, 生物多様性国際会議「霊長類のゲノム多様性研究」, 京都大学霊長類研究所主催(2010年3月4~6日, 犬山国際観光センター, 犬山市)
2. Asao Fujiyama: Realistic Approach to the New-Generation Sequencing Pipeline, 3rd Annual Illumina Asai Pacific User Symposium, 12-14 Apr, 2010, Phuket, Thailand
3. Yasukazu Nakamura, Eli Kaminuma, Takako Mochizuki, Yuichi Kodama, Satoshi Saruhashi, Hideki Nagasaki, Hideaki Sugawara, Toshihisa Takagi, Kousaku Okubo: DDBJ Read Annotation Pipeline: A cloud computing based analytical tool for next-generation sequencing data NIAS Symposium RAP7 meeting, 14 October, 2010, Tokyo, Japan
4. Tatsuo Fukagawa: Structural dynamics of inner-kinetochore for faithful chromosome segregation, EMBO workshop on chromosome segregation and aneuploidy, 19-23 June, 2010, Edinburgh, UK

国 内

1. 中村保一: 「DDBJ / INSDC の新型シーケンサデータへのとりくみ - DDBJ Sequence Read Archive (DRA) と解析支援系 DRA pipeline -」, 第 33 回日本分子生物学会年会 ランチョンセミナー「次世代シーケンサデータから見えてくる生命像」, 神戸, 日本, 2010年12月7日
2. 神沼英里, 児玉悠一, 望月孝子, 猿橋智, 長崎英樹, 菅原秀明, 大久保公策, 高木利久, 中村保一: 「次世代シーケンサーのアーカイブ DB とクラウド型データ解析システムの紹介」, 第 148 回農林交流センターワークショップ「次世代シーケンサーを利用したゲノム解析の実際」, つくば, 日本, 2010年9月9-10日
3. 深川竜郎: 動物細胞のキネトコア構造, 構造エピゲノム研究会 発足記念シンポジウム, 横浜, 日本, 2010年4月30日
4. 深川竜郎: 染色体分配におけるインナーキネトコアの構造変化, 第 62 回日本細胞生物学会, 大阪, 日本, 2010年5月19日
5. 深川竜郎: Structural dynamics of inner-kinetochore for faithful chromosome segregation, BMB2010 第 33 回日本分子生物学会年会・第 83 回日本生化学会大会合同大会, 神戸, 日本, 2010年12月9日

[一般講演]

国 際

1. Kakutani, T., Genetics of DNA methylation in genes and transposons in Arabidopsis, ASM Conference “Mobile DNA”, Montreal, Canada, 2010年4月27日
2. Kakutani, T., Genetics of DNA methylation in genes and transposons in Arabidopsis, International Conference of Arabidopsis Research, Yokohama, Japan, 2010年6月8日
3. Kakutani, T., Genetics of DNA methylation in genes and transposons in Arabidopsis, Cold Spring Harbor Asia meeting “From Plant Biology to Crop Biotechnology”, Suzhou, China
4. Kakutani, T., Genetics of DNA methylation in genes and transposons in Arabidopsis. Plant & Animal Genome XIX, EPIC: Plant Epigenome Project, San Diego, U.S.A. 2011年1月15日
5. Kakutani, T., Genetics of DNA methylation in genes and transposons in Arabidopsis, CDB Symposium 2011 Epigenetic Landscape in Development and Disease, 神戸, 日本 2011年3月14日
6. Yuichi Kodama, Saruhashi Satoshi, Eli Kaminuma, Hideaki Sugawara, Toshihisa Takagi, Kousaku Okubo, Yasukazu Nakamura, DDBJ Sequence Read Archive / DDBJ Omics Archive, Fourth Biocuration Conference, Tokyo, Japan, 2010年10月12-13日
7. Eli Kaminuma, Takako Mochizuki, Yuichi Kodama, Satoshi Saruhashi, Hideaki Sugawara,

Kousaku Okubo, Toshihisa Takagi and Yasukazu Nakamura, DDBJ Sequence(発表時追加) Read Annotation Pipeline —A cloud computing-based analytical tool for next-generation sequencing data, The tenth Cold Spring Harbor Laboratory/Wellcome Trust conference on Genome Informatics, Hinxton, UK, 2010年9月15-19日

8. Yuichi Kodama, Eli Kaminuma, Takako Mochizuki, Hidemasa Bono, Hideaki Sugawara, Toshihisa Takagi, Kousaku Okubo, Yasukazu Nakamura, DDBJ Omics Archive and DDBJ Read Annotation Pipeline: archive and analyze quantitative data from next generation platforms, 13th International MGED Meeting (MGED13), Boston, MA, USA, 2010年7月13-15日
9. Eli Kaminuma, Takako Mochizuki, Yuichi Kodama, Satoshi Saruhashi, Hideaki Sugawara, Kousaku Okubo, Toshihisa Takagi, Yasukazu Nakamura, DDBJ Read Annotation Pipeline: A web-based analytical tool for next-generation sequencing data, 18th Annual International Conference on Intelligent Systems for Molecular Biology(ISMB2010), Boston, MA, USA 2010年7月9日
10. Hori, T., Shang, W., Takeuchi, K., Toyoda, A., Sakakibara, Y., Fujiyama, A., Fukagawa, T., Unique Feature of Chicken Centromere DNA and Engineering of Centromere Using DT40 Cells, 50th ASCB Annual Meeting, Philadelphia, USA, 2010年12月12日
11. Suzuki, A., Hori, T., Nishino, T., Miyagi, A., Morikawa, K., Fukagawa, T., Stretched Distribution of Inner Kinetochores Occurs Dependent on Tension from Spindle Microtubules, 50th ASCB Annual Meeting, Philadelphia, USA, 2010年12月12日

国内

1. 角谷徹仁, DNAメチル化とトランスポゾンとシロイヌナズナのエピ遺伝学, 日本遺伝学会第82回大会シンポジウム, 札幌, 日本, 2010年9月20日
2. 河邊昭, 塚原小百合, 角谷徹仁, Arabidopsis属におけるCOPIA93/COPIA20ファミリーの分布と挿入位置特異性, 日本遺伝学会第82回大会, 札幌, 日本, 2010年9月21日
3. 塚原小百合, 小林啓恵, 河邊昭, 角谷徹仁, Arabidopsis lyrataにおけるCOPIA93レトロトランスポソンのセントロメア特異的分布の形成機構, 日本遺伝学会第82回大会, 札幌, 日本 2010年9月21日
4. 佐々木卓, 三浦明日香, 小林啓恵, 伊藤佑, 角谷徹仁, RNAi経路とクロマチン再構成による植物発生とDNAメチル化の制御, 日本遺伝学会第82回大会, 札幌, 日本, 2010年9月21日
5. 佐瀬英俊, 高嶋和哉, 北山淳子, 小林啓恵, 角谷徹仁, シロイヌナズナの遺伝子領域におけるヘテロクロマチン修飾除去メカニズムの解析, 日本遺伝学会第82回大会, 札幌, 日本 2010年9月21日
6. 角谷徹仁, シロイヌナズナにおけるDNAメチル化の遺伝学, 大阪大学蛋白質研究所セミナー「ゲノム記憶の成立機序とその制御」大阪, 日本, 2010年11月19日
7. 中村保一, 「新型シーケンサーから得られるデータをどう解釈し活用するか: 統合データベースプロジェクトからの提案」, 第33回日本分子生物学会年会 ワークショップ 1W20-p, 神戸, 日本 2010年12月7日
8. 児玉悠一, 猿橋智, 坂井勝呂, 神沼英里, 菅原秀明, 高木利久, 大久保公策, 中村保一, 「DDBJ Sequence Read Archive / DDBJ Omics Archive」, 第33回日本分子生物学会年会 ポスター発表 2P-1203, 神戸, 日本 2010年12月7日
9. 神沼英里, 望月孝子, 児玉悠一, 猿橋智, 菅原秀明, 大久保公策, 高木利久 1, 中村保一, 「次世代シーケンサーのアーカイブDBとクラウド型解析パイプライン」, NIASシンポジウムイネ遺伝学・分子生物学ワークショップ2010プログラム, つくば, 日本 2010年7月2日
10. 越阪部晃永, 立和名博昭, 堀哲也, 小布施力史, 木村宏, 深川竜郎, 胡桃阪仁志, 新規ヒストン結合因

子ヒト SPT2 の DNA 損傷修復における機能解析, BMB2010 第 33 回日本分子生物学会年会・第 83 回日本生化学会大会合同大会, 神戸, 日本, 2010 年 12 月 9 日

11. 竹内康造, 堀哲也, 西野達哉, 立和名博昭, 越阪部晃永, 胡桃阪仁志, 深川竜郎, DNA 結合活性を有する CCAN タンパク質群の機能解析, BMB2010 第 33 回日本分子生物学会年会・第 83 回日本生化学会大会合同大会, 神戸, 日本, 2010 年 12 月 9 日
12. 香川尚子, 堀哲也, 保木裕子, 佐渡 敬, 山縣一夫, 小久保博樹, 佐賀裕美子, 深川竜郎, BMB2010 第 33 回日本分子生物学会年会・第 83 回日本生化学会大会合同大会, 神戸, 日本, 2010 年 12 月 9 日
13. 北村大志, 大淵恵理, 田辺秀之, 小布施力史, 堀哲也, 深川竜郎, 原田昌彦, 細胞核内のクロマチン空間配置におけるアクチン関連タンパク質 Arp6 の機能解析, BMB2010 第 33 回日本分子生物学会年会・第 83 回日本生化学会大会合同大会, 神戸, 日本, 2010 年 12 月 9 日
14. 松田涼, 堀哲也, 竹内康造, 北村大志, 深川竜郎, 原田昌彦, 遺伝子破壊細胞によるヒストンバリエント H2A.Z アイソフォームの機能解析, BMB2010 第 33 回日本分子生物学会年会・第 83 回日本生化学会大会合同大会, 神戸, 日本, 2010 年 12 月 9 日
15. Wei-Hao Shang, Tetsuya Hori, Atsushi Toyoda, Kris Pendorf, Yasubumi Sakakibara, Asao Fujiyama, Tatsuo Fukagawa, Creation of chromosomes containing neocentromere in chicken DT40 cells, BMB2010 第 33 回日本分子生物学会年会・第 83 回日本生化学会大会合同大会, 神戸, 日本, 2010 年 12 月 9 日

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. Hara, H. and Takemura, A. (2010). A localization approach to improve iterative proportional scaling in Gaussian graphical models. *Communications in Statistics - Theory and Methods*, 39, 1643-1654.
2. Fujisawa, H. and Sakaguchi, T. (2010). Optimal significance analysis of microarray data in a class of tests whose null statistic can be constructed. *ISM Research Memorandum*, No.1127.
3. Fujisawa, H. (2010). Robust model selection criteria based on gamma-cross entropy. *ISM Research Memorandum*, No.1128.
4. Dou, X., Kuriki, S., Maeno, A., Takada, T., and Shiroishi, T. (2010). Influence Analysis in Quantitative Trait Loci Detection, *ISM Research Memorandum*, No.1136.
5. Dou, X., Kuriki, S., and Lin, G. D. (2010). On Baker's bivariate distributions with fixed marginals, *ISM Research Memorandum*, No.1135.

<会議発表等>

[招待講演]

国際

1. Dou, X., Statistical inference for Baker's distribution. 中央研究院統計科學研究所演講, 台北, 2010 年 12 月 6 日

[一般講演]

国 際

1. Hara, H., Sei, T., and Takemura, A., Hierarchical subspace models for contingency tables. IMS annual Meeting, Chalmers University of Technology, Gothenburg, Sweden, 2010年8月12日

国 内

1. 藤澤洋徳, 坂口隆之: 遺伝子発現差解析における並べ替え技法による P 値の推定, 統計関連学会連合大会, 東京, 2010年9月7日
2. 藤澤洋徳: ロバスト推測の基礎とダイバージェンス型メソッドへの発展, 第13回情報論的学習理論ワークショップ, 東京, 2010年11月5日
3. 藤澤洋徳, 坂口隆之: 遺伝子発現差解析における並べ替え技法による P 値推定と最適な有意性検定, 科研費研究集会「生物情報解析の理論的基礎とその応用」, 東京, 2010年12月7日
4. 藤澤洋徳: 外れ値の割合が大きい場合にもバイアスが小さいロバスト推定, 広島大学理学部数学教室談話会, 広島, 2011年1月25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

1. ソフトウェア: **Social Interaction** 行動自動解析ソフトウェア (土谷, 小出等)
2. トランスジェニックフィッシュ系統を基にして国内外の研究者と共同研究を展開し, 側線形成, 脊索細胞分化における Notch シグナル, リンパ管形成, Fanconi 貧血症原因遺伝子, 水晶体形成, 恐怖行動メカニズム, 神経活動のカルシウムイメージングにおいて新しい知見を得ることができた。この中でトランスジェニックフィッシュをまとめたデータベースを整備し, 論文発表にいたった。

② 成果発表等

<論文発表>

[学術論文]

1. Ishii, A., Koide, T., Takahashi, A., Shiroishi, T., Hettlinger, TP., Frank, ME., Savoy, LD., Formaker, BK., Yertutanol, S., Lionikas, A., and Blizard, DA. B6-MSM Consomic Mouse Strains Reveal Multiple Loci for Genetic Variation in Sucrose Octaacetate Aversion. *Behav Genet.* In-press (2011).
2. Nishi A, Ishii A, Takahashi A, Shiroishi T, Koide T. QTL analysis of measures of mouse home-cage activity using B6/MSM consomic strains. *Mamm Genome.* (2010).
3. Oka A, Mita A, Takada Y, Koseki H, Shiroishi T. Reproductive isolation in hybrid mice due to spermatogenesis defects at three meiotic stages. *Genetics.* 186: 339-51 (2010).
4. Takahashi A, Tomihara K, Shiroishi T, Koide T. Genetic mapping of social interaction behavior in B6/MSM consomic mouse strains. *Behav Genet.* 40: 366-376 (2010).
5. Dowse, H., Umemori, J., Koide, T. Ultradian Components in the Locomotor Activity Rhythms of the Genetically Normal Mouse, *Mus musculus.* *J Exp. Biol.* 213:1788-1795, 2010.
6. Horiuchi, Y., Harushima, Y., Fujisawa, H., Mochizuki, T., Kawakita, M., Sakaguchi, T., and Kurata N. A simple optimization can improve the performance of single feature polymorphism detection by Affymetrix expression arrays. *BMC Genomics* 11(1):315, 2010.
7. Fujii, S., Yamada, M., Fujita, M., Itabashi, E., Hamada, K., Yano, K., Kurata, N., and Toriyama, K.

- Cytoplasmic-nuclear genomic barriers in rice pollen development revealed by comparison of global gene expression profiles among five independent cytoplasmic male sterile lines. *Plant Cell Physiol.* 51(4):610-620, 2010.
8. Mizuta, Y., Harushima, Y., Kurata, N. Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. *Proc. Natl. Acad. Sci. USA* 107(47):20417-20422, 2010.
 9. Fujita, M., Horiuchi, Y., Ueda, Y., Mizuta, Y., Kubo, T., Yano, K., Yamaki, S., Tsuda, K., Nagata, T., Niihama, M., Kato, H., Kikuchi, S., Hamada, K., Mochizuki, T., Ishimizu, T., Iwai, H., Tsutsumi, N., Kurata, N. Rice expression atlas in reproductive development. *Plant Cell Physiol.* 51(12):2060-2081, 2010.
 10. Kurata, N., Satoh, H., Kitano, H., Nagato, Y., Endo, T., Sato, K., Akashi, R., Ezura, H., Kusaba, M., Kobayashi, M., Nitasaka, E., Kasai, F., Yamazaki, Y., and Yoshimura, Y. NBRP, National Bioresource Project of Japan and plant bioresource management. *Breeding Science* 60(5):461-468, 2010.
 11. Nonomura, K. I., Morishima, H., Miyabayashi, T., Yamaki, S., Eiguchi, M., Kubo, T., and Kurata, N. The wild *Oryza* collection in National BioResource Project (NBRP) of Japan: History, biodiversity and utility. *Breeding Science* 60(5):502-508, 2010.
 12. Yamazaki, Y., Sakaniwa, S., Tsuchiya, R., Nonomura, K. I., and Kurata, N. Oryzabase: an integrated information resource for rice science. *Breeding Science* 60(5):544-578, 2010.
 13. Yamaki, S., Miyabayashi, T., Eiguchi, M., Kitano, H., Nonomura, K. I., and Kurata, N. Diversity of panicle branching patterns in wild relatives of rice. *Breeding Science* 60(5):586-596, 2010.
 14. Hamada, K., Hongo, K., Suwabe, K., Shimizu, A., Nagayama, T., Abe, R., Kikuchi, S., Yamamoto, N., Fujii, T., Yokoyama, K., Tsuchida, H., Sano, K., Mochizuki, T., Oki, N., Horiuchi, Y., Fujita, M., Watanabe, M., Matsuoka, M., Kurata, N., Yano, K. OryzaExpress: An Integrated Database of Gene Expression Networks and Omics Annotations in Rice. *Plant Cell Physiol.* 52(2):220-229, 2011.
 15. Ishikawa, R., Ohnishi, T., Kinoshita, Y., Eiguchi, M., Kurata, N., Kinoshita, T. Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncytial nuclear division. *Plant J.* 65(5):798-806, 2011.
 16. Nonomura, K-I, Eiguchi M., Nakano, M., Takashima, K., Komeda, N., Fukuchi, S., Miyazaki, S., Miyao, A., Hirochika, H., Kurata, N. A novel RNA-recognition-motif protein is required for premeiotic G1/S-phase transition in rice (*Oryza sativa* L.). *PLoS Genetics* 7(1):e1001265, 2011.
 17. Yamaki, S., Nagato, Y., Kurata, N., Nonomura, K.I. Ovule is a lateral organ finally differentiated from the terminating floral meristem in rice. *Developmental Biology* 351(1):208-216, 2011.
 18. Ohnishi, T., Takanashi, H., Mogi, M., Takahashi, H., Kikuchi, S., Yano, K., Okamoto, T., Fujita, M., Kurata, N., Tsutsumi, N. Distinct Gene Expression Profiles in Egg and Synergid Cells of Rice as Revealed by Cell Type-specific Microarrays. *Plant Physiol.* 155(2):881-891, 2011.
 19. Pujol-Martí, J., Baudoin, J.P., Faucherre, A., Kawakami, K., and López-Schier, H. Progressive neurogenesis defines lateralis somatotopy. *Developmental Dynamics* 239, 1919-1930 (2010).
 20. Yamamoto, M., Morita, R., Mizoguchi, T., Matsuo, H., Isoda, M., Ishitani, T., Chitnis, A.B., Matsumoto, K., Crump, J.G., Hozumi, K., Yonemura, S., Kawakami, K., and Itoh, M. Mib-Jag1-Notch signalling regulates patterning and structural roles of the notochord by controlling cell-fate decisions. *Development* 137, 2527-2537 (2010).
 21. Bussmann, J., Bos, F.L., Urasaki, A., Kawakami, K., Duckers, H.J., and Schulte-Merker, S. Arteries

provide essential guidance cues for lymphatic endothelial cells in the zebrafish trunk. *Development* 137, 2653-2657 (2010).

22. Rodríguez-Marí, A., Cañestro, C., Bremiller, R.A., Nguyen-Johnson, A., Asakawa, K., Kawakami, K., and Postlethwait, J.H. Sex reversal in zebrafish *fancl* mutants is caused by Tp53-mediated germ cell apoptosis. *PLoS Genetics* 6:e1001034 (2010).
23. Imai, F., Yoshizawa, A., Fujimori-Tonou, N., Kawakami, K., and Masai, I. The ubiquitin proteasome system is required for cell proliferation of the lens epithelium and for differentiation of lens fiber cells in zebrafish. *Development* 137, 3257-3268 (2010).
24. Asakawa, K., and Kawakami, K. A transgenic zebrafish for monitoring in vivo microtubule structures. *Developmental Dynamics* 239, 2695-2699 (2010).
25. Kawakami, K., Abe, G., Asada, T., Asakawa, K., Fukuda, R., Ito, A., Lal, P., Mouri, N., Muto, A., Suster, M.L., Takakubo, H., Urasaki, A., Wada, H., and Yoshida, M. zTrap: zebrafish gene trap and enhancer trap database. *BMC Developmental Biology* 10, 105 (2010).
26. Agetsuma, M., Aizawa, H., Aoki, T., Nakayama, R., Takahoko, M., Goto, M., Sassa, T., Amo, R., Shiraki, T., Kawakami, K., Hosoya, T., Higashijima, S., and Okamoto, H. The habenula is crucial for experience-dependent modification of fear responses in zebrafish. *Nature Neuroscience* 13, 1354-1356 (2010).
27. Muto, A., Ohkura, M., Kotani, T., Higashijima, S.-i., Nakai, J., and Kawakami, K. Genetic visualization with an improved GCaMP calcium indicator reveals spatiotemporal activation of the spinal motor neurons in zebrafish. *Proc. Natl. Acad. Sci. USA* 108, 5425-5430 (2011).

[データベース]

1. zTrap: zebrafish gene trap and enhancer trap database. <http://kawakami.lab.nig.ac.jp/ztrap/> (論文 25 ; 浅川, 武藤, 川上)

<会議発表等>

[招待講演]

国 際

1. 小出剛: “Attempts toward understanding genetic basis of non-Mendelian inheritance: behavior and genetic incompatibility” セミナー, 2010年10月22日, Institute of Molecular Genetics of Montpellier, Montpellier, France
2. Kurata N. Various Reproductive Barriers in the Hybrids and Its Relation to Evolution. 8th International Symposium on Rice Function Genomics, 2010年10月18日-20日, Bento Gonzalves, Brazil
3. Kawakami, K. Transgenesis and genome analysis with the Tol2 transposable element in mice and zebrafish. “FASEB Summer Research Conferences, Genome Engineering: Research and Therapeutic Applications” Steamboat, Colorado. June 6-11 (2010).
4. Kawakami, K., Abe, G., Asakawa, K., Fukuda, R., Lal, P., Muto, A., Suster, M.L., Takakubo, H., Urasaki, A. zTrap and NIGKOF: resources for gene trap and enhancer trap lines and knockout fish. “9th International Conference on Zebrafish” Madison, USA. June 16-20 (2010).
5. Kawakami, K. The *Tol2*-mediated Gal4-UAS system and its application to the study of functional neural circuits in zebrafish. “Imaging Structure and Function in the Zebrafish Brain” Lisbon,

Portugal Dec 13-15 (2010).

国内

1. 城石俊彦：「マウス亜種間コンソミック系統を用いた多因子形質の遺伝解剖」，第 24 回日本糖尿病・肥満動物学会学術集会特別講演，2010 年 1 月 23 日，大阪
2. 城石俊彦：「マウス亜種間コンソミック系統による複合形質の遺伝解剖」，臨床研セミナー特別企画：「マウス遺伝学の今」2010 年 3 月 24 日，東京
3. 高田豊行：「多因子形質解析とその実験モデルマウスの開発」第 57 回日本実験動物学会総会，2010 年 5 月 12 日～14 日，京都
4. 城石俊彦：「コンソミックマウスのすべて」，第 4 回ターゲットバンセミナー 2010 年 7 月 30 日，東京
5. 小出剛：「野生由来マウス系統間にみられる行動多様性の遺伝的基盤」放射線医学総合研究所セミナー，2010 年 11 月 11 日，千葉
6. 小出剛：「時間依存的な活動量に関わる遺伝的要因の解析」第 17 回日本時間生物学会学術大会 2010 年 11 月 20・21 日，東京
7. 水多陽子，春島嘉章，倉田のり，イネ亜種間交雑で生殖的隔離を引き起こす重複遺伝子 DPL1, 2 の解析，日本遺伝学会第 82 回大会，2010 年 9 月 20 日-23 日，北海道大学高等教育機能開発総合センター，日本

[一般講演]

国際

1. Toyoyuki Takada, Akihiko Mita, Shigeharu Wakana, Kazuo Moriwaki, Hiromichi Yonekawa, Toshihiko Shiroishi. Age-associated change of energy metabolism are genetically dissected by mouse inter-subspecific consomic strains. 23th International Mammalian Genome Conference, 2010. 10.17-21, Crete, Greece.
2. Ishii, A., Nishi, A., Shiroishi, T., Takahashi, A., Koide, T.: Consomic analysis of genetic factors related to temporal difference of home-cage activity between B6 and MSMM. (ポスター発表)24th International Mammalian Genome Conference, October 17-21, 2010, Crete, Greece.
3. Ohyanagi, H., Kaminuma, E., Nagata, T., Mochizuki, T., Nagasaki, H., Nakamura, Y., Aizu, T., Toyoda, A., Fujiyama, A., Kurata, N. TOWARDS THE DRAFT GENOME SEQUENCE OF ORYZA OFFICINALIS: THE INITIAL LANDMARK IN ORYZA CC GENOMES. 8th International Symposium on Rice Function Genomics, 2010 年 10 月 18 日-20 日，Bento Gonçalves, Brazil
4. Tsuda, K., Sato, Y., Ito, Y., Kurata, N. Positive Autoregulation of KNOX genes in rice. (ポスター発表) 第 2 回 NIBB-MPIZ 合同シンポジウム「Plant Science Communications 2010」，2010 年 11 月 16 日-18 日，岡崎コンファレンスセンター，日本

国内

1. 高田豊行. マウス亜種間の表現型多様性とゲノム多型性に基づいた機能ゲノム学のためのデータベース，日本遺伝学会・ワークショップ「モデル動物における表現型解析の標準化と情報統合」2010, 9.20-22, 札幌.
2. 高田豊行. マウス MSM/Ms 系統のゲノム解析. 第 24 回モロシヌス研究会，2010, 9.17-18, 阿蘇.
3. 高田豊行，三田晃彦，森脇和郎，米川博通，城石俊彦：マウス亜種間コンソミック系統群による加齢性エネルギー代謝表現型の遺伝解剖. 第 57 回日本実験動物学会総会，2010, 5.12-14, 京都.
4. 小出剛，石井亜矢子，西明紀，田邊彰，城石俊彦，高橋阿貴：マウス行動の多様性を生み出す遺伝-環境要因 (ワークショップ口頭発表) 第 82 回日本遺伝学会大会ワークショップ 2010 年 9 月 20～23 日

札幌

5. 梅森十三, 宇野毅明, 湯浅茂樹, 小出剛: 遺伝的不適合マウスにおけるエピジェネティック異常 (口頭, 日本語) 第 82 回日本遺伝学会, 2010 年 9 月 20 日, 札幌
6. 梅森十三, 宇野毅明, 湯浅茂樹, 小出剛: Epigenetic abnormality of genetic incompatibility mouse, Genic mice (口頭, 英語) 第 33 回日本分子生物学会年会, 2010 年 9 月 3 日, 神戸
7. Takahashi, A., Shimamoto, A., Boyson, C.O., Koide, T., DeBold, J.F., Miczek, K.A. GABAB receptor modulation of serotonin neurons in the dorsal raphe nucleus escalates aggression in mice. (口頭, 英語) Neuro2010 2010.9.3 神戸
8. 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 (ポスター発表) 第 57 回日本実験動物学会 2010 年 5 月 12 日
9. 杉本大樹, 高橋阿貴, 城石俊彦, 小出剛 アセチルコリン受容体モジュレーターLynx1 のマウス攻撃行動への関与 (ポスター発表) 第 33 回日本神経科学会 2010 年 9 月 2 日, 神戸
10. 杉本大樹, 菊水健史, 城石俊彦, 小出剛: マウス交配時の超音波発声の役割 (ポスター発表) 第 57 回日本実験動物学会, 2010 年 5 月 12-14 日, 京都
11. 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系統差に関わる多因子の遺伝学的解析 (ポスター発表) 第 33 回日本分子生物学会年会 2010 年 12 月 7-10 日 神戸
12. 高梨秀樹, 大西孝幸, 茂木美来, 菊地俊介, 矢野健太郎, 岡本龍史, 藤田雅丈, 倉田のり, 堤伸浩, イネ雌性配偶体構成細胞における遺伝子発現プロファイリング, イネ遺伝学・分子生物学ワークショップ 2010, 2010 年 7 月 2 日-3 日, つくば国際会議場, 日本
13. 水多陽子, イネ亜種間雑種で花粉不発芽を引き起こす重複遺伝子 DPL1, 2 の解析. 科学研究費特定領域研究「植物ゲノム障壁」若手ワークショップ, 2010 年 7 月 12 日-13 日, 名古屋大学, 日本
14. 木下哲, 大西孝幸, 石川亮, 木下由紀, 永口貢, 倉田のり, 高等植物は胚乳における種の障壁成立のルールとメカニズム, 日本遺伝学会第 82 回大会, 2010 年 9 月 20 日-23 日, 北海道大学高等教育機能開発総合センター, 日本
15. Ohnishi, T., Ishikawa, R., Kinoshita, Y., Eiguchi, M., Kurata N., and Kinoshita T. Interspecific hybridization in *Oryza* is associated with altered endosperm development and deregulation of OsMADS87 imprinting. BMB2010 (第 33 回日本分子生物学会年会 第 83 回日本生化学会大会 合同大会), 2010 年 12 月 7 日-10 日, 神戸ポートアイランド, 日本
16. 野々村賢一, 永口貢, 宮崎さおり, 米田典央, 宮尾安藝雄, 廣近洋彦, 倉田のり, 植物の減数分裂前 G1/S 移行を制御する新規 RRM 蛋白質の解析, BMB2010 (第 33 回日本分子生物学会年会 第 83 回日本生化学会大会 合同大会), 2010 年 12 月 7 日-10 日, 神戸ポートアイランド, 日本
17. 須崎大地, 永田俊文, 植田美那子, 倉田のり, 東山哲也, 顕微細胞操作と大規模発現解析を駆使して雌性配偶体が機能を獲得する仕組みに迫る, 第 52 回日本植物生理学会年会, 2010 年 3 月 11 日, 植物生理学会ホームページ上
18. 水多陽子, 春島嘉章, 倉田のり, イネ亜種間交雑で生殖的隔離を引き起こす重複遺伝子 DPL1, DPL2 の解析, 第 52 回日本植物生理学会年会, 2010 年 3 月 11 日, 植物生理学会ホームページ上
19. 津田勝利, 伊藤幸博, 佐藤豊, 倉田のり, イネにおける KNOX 遺伝子の自己制御, 第 52 回日本植物生理学会年会, 2010 年 3 月 11 日, 植物生理学会ホームページ上
20. 濱田和輝, 本郷耕平, 山本直樹, 藤井貴朗, 望月孝子, 諏訪部圭太, 倉田のり, 矢野健太郎, *OryzaExpress*: イネの遺伝子発現ネットワーク解析とデータベース構築(ポスター発表), 日本育種学会 第 118 回講演会, 2010 年 9 月 23 日-26 日, 秋田県立大学, 日本
21. M. Shenton, 倉田のり, C ゲノムとの種間交配における *O. sativa* 花粉管伸長停止の特性解析 (ポスタ

- 一発表), 日本育種学会 第 118 回講演会, 2010 年 9 月 23 日-26 日, 秋田県立大学, 日本
22. Hamada, H., Suwabe, K., Hongo, K., Yamamoto, N., Fujii, T., Mochizuki, T., Kurata, N. and Yano, K. Gene expression network analysis from large-scale microarray data in rice. (ポスター発表) BMB2010 (第 33 回日本分子生物学会年会 第 83 回日本生化学会大会 合同大会), 2010 年 12 月 7 日-10 日, 神戸ポートアイランド, 日本
 23. 野々村賢一, 永口貢, 宮崎さおり, 米田典央, 宮尾安藝雄, 廣近洋彦, 倉田のり, 植物の減数分裂前 G1/S 移行を制御する新規 RRM 蛋白質の解析(ポスター発表), BMB2010 (第 33 回日本分子生物学会年会 第 83 回日本生化学会大会 合同大会), 2010 年 12 月 7 日-10 日, 神戸ポートアイランド, 日本
 24. 永田俊文, 望月孝子, 大柳一, 長崎英樹, 神沼英里, 中村保一, 会津智幸, 豊田敦, 藤山秋佐夫, 倉田のり, 高速シークエンサーを用いたイネ近縁種の比較ゲノム解析(ポスター発表), BMB2010 (第 33 回日本分子生物学会年会 第 83 回日本生化学会大会 合同大会), 2010 年 12 月 7 日-10 日, 神戸ポートアイランド, 日本
 25. 須崎大地, 永田俊文, 上田美那子, 佐々木成江, 倉田のり, 東山哲也, 顕微鏡操作と大規模発現解析により被子植物雌性配偶体の発生と機能獲得機構に迫る (ポスター発表), BMB2010 (第 33 回日本分子生物学会年会 第 83 回日本生化学会大会 合同大会), 2010 年 12 月 7 日-10 日, 神戸ポートアイランド, 日本
 26. 坂庭真悟, 土屋里枝, 高橋由佳, 野々村賢一, 宮林登志江, 久保貴彦, 倉田のり, 山崎由紀子, Improvement to Oryzabase: Integrated Rice Science Database(ポスター発表), BMB2010 (第 33 回日本分子生物学会年会 第 83 回日本生化学会大会 合同大会), 2010 年 12 月 7 日-10 日, 神戸ポートアイランド, 日本
 27. 木本剛彰, 柴博史, 岩野恵, 高橋宏和, 中園幹生, 藤田雅丈, 倉田のり, 磯貝彰 他, レーザーマイクロダイセクションを利用したシロイヌナズナ薬タペート組織における遺伝子発現様態の網羅的解析 (ポスター発表), 第 52 回日本植物生理学会年会, 2010 年 3 月 11 日, 植物生理学会ホームページ上
 28. 永田俊文, 大柳一, 長崎英樹, 望月孝子, 神沼英里, 中村保一, 会津智幸, 豊田敦 他, 高速シークエンサーを用いたイネ近縁種の比較ゲノム解析(ポスター発表), 第 52 回日本植物生理学会年会, 2010 年 3 月 11 日, 植物生理学会ホームページ上
 29. 濱田和輝, 本郷耕平, 諏訪部圭太, 清水顕史, 長山大志, 阿部伶奈, 菊地俊介, 山本直樹 他, OryzaExpress: イネの機能アノテーションと遺伝子発現ネットワークの統合データベース(ポスター発表), 第 52 回日本植物生理学会年会, 2010 年 3 月 11 日, 植物生理学会ホームページ上

<受賞>

1. 高田豊行 日本実験動物学会奨励賞 第 57 回日本実験動物学会総会「多因子形質解析とその実験モデルマウスの開発」
2. 濱田和輝, 山本直樹, 望月孝子, 諏訪部圭太, 倉田のり, 矢野健太郎, 第 117 回講演会日本育種学会優秀発表賞
3. 水多陽子, 春島嘉章, 倉田のり, 第 117 回講演会日本育種学会優秀発表賞

③ その他の成果発表: なし