

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

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] 研究プロジェクト

(1) 目的・目標

本プロジェクトでは、超大規模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」の流れを作り、生命システム分野の展開を図る。

(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ー形成や、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。これまでの研究期間に成果も出始めており、さらに推進速度を上げ、新たな展

開を図る事が重要である。

(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して期待できる。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壤は、いずれの研究コミュニティ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 な人材育成の土壤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、我が国的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、1) これまでに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、野生イネ系統についての比較ゲノム解読、ゲノム機能領域の大規模解析等を行っている (Huang, Kurata, et al. *Nature*, 2012; Shang, et al. *Genome Res.*, 2010 等)。2) また、次世代シーケンシングシステムの開発・改良を進め、2008年7月から2013年3月までに累積で約 80×10^{12} 塩基のデータを生産するとともに、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いる。3) 超大規模データ生産とその直接的情報処理に加え、大規模ゲノムデータの情報・統計処理解析とゲノム工学等の融合を進めている。4) 染色体機能領域研究については、特に動原体領域に存在するタンパク質の同定やゲノム配列上の特徴、ゲノム改変技術を開発

してきたが (Hori et al., *Cell*, 2008; Amano et al., *J. Cell Biol.*, 2009)、さらにゲノム工学を活用して人工的に動原体を構築する実験系を確立した (Nishino et al., *EMBO J.* 32: 424-436, 2013 ; Hori et al., *J. Cell Biol.* 200: 45-60, 2013)。5) これに加え、主にシロイヌナズナ及びイネを用いて、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, Harushima, Fujisawa & Kurata, *Annals of the Institute of Statistical Mathematics*, 2014)。5) 遺伝子発現差解析に関連して、既存の外れ値の処理方法を大幅に変更してより良い結果を得ることに成功した (Fujisawa, H., Horiuchi, Y., Harushima, Y., Takada, T., Eguchi, S., Mochizuki, T., Sakaguchi, T., Shiroishi, T. and Kurata, N., *BMC Bioinformatics*, 2009)。6) 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, Kuriki, Maeno, Takada & Shiroishi, *Biometrical Journal*, 2014)。7) 新しい並べ替え技法による P 値の推定法を提案した。一般的に行われてきた並べ替え技法による P 値の推定は必ずしも妥当ではない。P 値推定が妥当になる条件を整理して、そのクラスの中で最適な検定方法を導出した (Fujisawa, H. and Sakaguchi, T., *TEST*, 2012)。8) 不均一な集団を記述するためのベルンシュタインコピュラの方法を開発し、マウスコンソミックデータに適用した (Dou, Kuriki, Lin & Richards, *Computational Statistics & Data Analysis*, 2014)。これらの成果のうち、1)、4) 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、6) については第 1 期からの継続中のテーマである。

サブテーマ 3 については、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」に基づく成果を論文化し、さらに第 II 期の研究においても、以下のような展開が進み、新たな成果が得られている。

1) マウスにおいては、第 I 期の量的形質情報の基盤整備や遺伝的解析に加えて、ゲノム配列多型 (Takada et al., *Genome Res.*, 2013[朝日新聞 2013.6.7 紹介記事掲載]) さらには、この多型情報を取り入れた複合量的形質の解析 (Oka et al., *PLoS Genetics*, 2014) に進展があった。また、表現型情報を使用した統計解析手法 (Dou X et al., *Biom. J.* 2014) をサブテーマ 2 との連携で報告した。量的形質情報の基盤整備に関しては、*Mamm. genome* 2012; *Exp. anim.* 2012 に総説を報告している。マイクロ CT 画像の情報学的解析では、情報研との共同研究による成果 (Roy S et al. *Med. Image Comput. Comput-Assist. Interv.* 2013) や遺伝解析 (Tamura et al. *Hum. Mol. Genet.*, 2013) についても論文を報告した。現在は、これまでに得られた複合量的形質データおよびゲノム配列多型情報など、大規模データを利用した各種遺伝解析を推進している。

2) マウスの社会行動を隠れマルコフモデルにより自動解析するフリーウェア (DuoMouse) を完成させ、国際誌に論文掲載された。DuoMouse は遺伝研のウェブサイトからフリーで公開し、その成果はプレスリリースされ、新聞報道された。上記 2013-2014 年以前の成果として、マウス行動解析の遺伝的基礎研究 (Takahashi et al., *Behav. Genet.*, 2006, *Mamm. Genome*, 2008, *Behav. Genet.*, 2009, Umemori et al. *BMC Genetics*, 2009) の成果を発表した。

3) イネにおいては、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 (Mizuta et al., PNAS 2010)、野生イネ多数系統の系統進化関係と栽培イネ起源地を特定 (Huang, Kurata, Wang, Wei, et al., Nature, 2012) した成果を出し、これらの成果をさらに取り込むため、多数系統の野生イネを用いた GWAS (genome-wide association study) 解析のための形質調査を行い、GWAS 解析の手法改良をサブチーム 2 と合同で進めた。また、イネ遺伝子発現制御ネットワーク解析のため、系統間の大量発現データの解析および e-QTL 用材料作成を行った。

4) ゼブラフィッシュにおいては、細胞・組織・器官特異的に Gal4 を発現するトランスジェニックフィッシュを選別する大規模スクリーニングを行い、数百に及ぶ系統を作製してきた。それらトランスジェニックフィッシュを用いて、国内共同研究者と初期発生、器官形成、神経科学に関する共同研究を行い多くの成果を得てきた (Asakawa et al., Dev Dynamics, 2012; Hirata et al., J Biol Chem, 2012; Pujiol-Marti et al., J Neurosci, 2012; Nakayama et al., Dev Growth Diff, 2012; Fukui et al., FEBS Lett, 2012; Yano et al., Development, 2012; Shimizu et al., Dev Biol, 2012; Umeda et al., Neurosci Res, 2013; Wada et al., PNAS, 2013; Kishimoto et al., Nature Neurosci, 2013; Asakawa et al., Frontiers in Neural Circuits, 2013; Muto et al., Frontiers in Neural Circuits, 2013; Banjo et al., Nature Comm, 2013; Wada et al., Curr Biol, 2013; Satou et al., Development 2013; Kwon et al., Development 2013; Nakayama et al., Mech Dev 2013; Sittaramane et al., Dev Biol, 2013)。

5) 2002 年から 2009 年までに RNAi 変異体系統 13,000 系統を作出し、首都大学東京 相垣教授と共に成虫翅の変異表現型を解析するため 20 万枚の画像を取得した。また第 2 世代の RNAi 系統を 3,000 系統作出了した。さらに CRISPR/Cas9 エンドヌクレアーゼを用いたゲノム編集技術をショウジョウバエに応用し、生殖細胞特異的発現プロモータを組み込んだ Cas9 遺伝子導入ハエ系統を用いて平均 50% という高効率でターゲット遺伝子の突然変異を誘発することができた。この手法により、ショウジョウバエのヌル変異体をハイスループットで体系的に取得する技術として発表した (Kondo & Ueda, Genetics, 2013)。

(6) 国内外における関連分野の学術研究の動向

次世代シーケンサの普及により、大規模ゲノムデータにもとづく情報処理、統計処理が個別的な研究テーマにも求められるようになり、大学等の研究室が有するゲノム解析技術レベルとの乖離が深刻化している。また、我が国では諸外国に比して特に遅れていたヒトゲノム研究材料の系統的収集を行うバイオバンクとゲノム情報解読とが一体化した東北メガバンクが稼働し、既に千人規模での日本人ゲノムデータが生産されている。シーケンス技術開発も活発に行われている。高速化、微量化、大規模化、簡便化、1 分子シーケンス等の様々なキーワードの機種開発が進んでいる。また、ナノポア技術を応用した第 4 世代シーケンサの市販が始まった。

遺伝子関連データ解析の観点からは、本プロジェクトで扱うゲノムデータ解析のための並べ替え検定手法の開発、多重性調整への応用、LASSO などの疎性・機械学習アプローチはその実用性が強く期待できる分野である。多重性に関する研究については、国際的に医学統計の観点での研究が多く見られるが、本プロジェクトが主にあつかう実験交配生物に関する多重性に関するものは多くない。

本研究で主に扱うマウス、イネ、ゼブラフィッシュ、ショウジョウバエなどのモデル生物においては、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。本プロジェクトで扱う遺伝資源と実験系系統は、世界の中でも独自のリソースであり、これまでに個々の系統レベルや遺伝子レベル、ゲノム解析などは世界中で利用され始めている。しかし、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、未だ非常に少ない。材料の優位はあるが、方法論の開発も世界中で時を追って進んできており、どのような系統やデータを用いるかの独自性を有効にするため、マッチした方法論との組み合わせの開発は緊急の問題となりつつある。動物行動研究において、社会行動や音声コミュニケーション、さらに多個体を同じケージに入れた状態での行動テストなどにより得られた時系列データの自動解析の重要性が増してきており、本プロジェクトで進めている研究は重要である。それらの行動形質に関わる多因子の解明に向けて、多数系統を交雜したヘテロジニアス集団を用いた遺伝解析法の確立は多くの研究者の注目を集めつつある。また多様な遺伝的背景を持つ実験および野生イネ集団を用いた統計的解析も、世界中で展開されつつあり、精度のよい方法論の開発や、独自の表現型の解析への適用などが待たれる。

[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、「生命システム」としての「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ーに公開・発信していく。

サブテーマ 1

新型 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、解析対象の微少化等のウエットサイドでの技術開発が、最終的には単一細胞レベルでの解析が可能になるまで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』ともいえる細胞レベルでの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国内外の研究コミュニティーと連携しながら融合の視点で研究を遂行する。一方、本プロジェクトで導入した DNA シーケンサは初期型であり、今後のアップグレードには対応できないため、第 4 世代装置の導入も含めて設備更新の検討が必要である。

サブテーマ 2

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) LASSO などの疎性、機械学習アプローチの利用
- (c) 生物・遺伝データ特有の特徴をとらえる統計データ解析

を軸として、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、イネ、ゼブラフィッシュ、ショウジョウバエを用いて、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な内容は、以下の通りである。

- (a) マウスについては、野生系統を含む多系統交雑によるヘテロジニアス集団や 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。ゲノムデータの効率的で信頼性の高い解析手法の確立。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、GWAS 解析。
- (c) ゼブラフィッシュについては、トランスポゾンを用いた遺伝学的方法論に基づく大規模遺伝子改変系統の作製とそれらからの多様な表現型の抽出および表出法の確立を行なう。
- (d) ショウジョウバエについては、RNAi knock-down 変異系統群を用いた、翅形態の変異の抽出および変異遺伝子群との相関解析を行い、また CRISPR/Cas9 を用いた変異体の作成と遺伝子変異-表現型相関解析の基盤を作る。

(2) 各年度の計画

平成 26 年度

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、データ解析を行い、相互にフィードバックしつつ「生命システム」解析に迫る。データ抽出法、解析方法論の開発を通じて、多重ゲノムデータと表現型の体系的な相互関連解析、様々な系統特性とゲノム関連情報を結びつける遺伝機能システム学として成果を挙げる事を計画している。

サブテーマ 1 では、第二世代シーケンサを用いた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時系列遺伝子発現プロファイル作成をめざし、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。とともに、時期的に市販化が予想される第 4 世代 DNA シーケンサ及び新第二世代シーケンサの導入を視野に入れ、それらに対応した情報解析手法についての検討を行う。

サブテーマ 2 では、多重性調整に関しては、H25 に開発した疑似相関を用いた多重検定手法についての論文を完成させるとともに、さらなる応用を模索する。高次元説明変数の次元圧縮を予測目的に合わせて行う方法に関しては、構築した方法がうまく働くかどうかを様々な数値実験と実データ解析を通して検証する。LASSO を用いた QTL 解析については、より多くの交互作用を含めた回帰モデルを当てはめ、その結果を遺伝学的な観点から議論する。ランキング解析については、H25 に行った 3 すくみ交互作用に基づく階層モデルの方法を理論化する。さらに、評価項目が複数ある場合に結果を統合する方法を開発する。

サブテーマ 3 では、野生系統を含む多系統交雑によるヘテロジニアス集団や 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（マウス）、一連の系統の e-QTL 情報の取得、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解析手法を構築する（イネ、マウス）。変異体の表現形質変動については、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の抽出を試みる（ゼブラフィッシュ、ショウジョウバエ）。

平成 27 年度

サブテーマ 1 では、前年度の計画を進める。特に、解析対象の微量量化技術の実現が期待できるため、発生過程における 1 細胞レベルでの精細な遺伝子発現プロファイル時系列データ解析等の当初計画の実現を図る。

サブテーマ2では、多重性調整に関しては、Graphical LASSO の情報量規準を与え、遺伝子間の関連を調べる解析のため、SCAD や MCP などといった LASSO 型の正則化法に対する情報量規準を導出し、実データに適用する。高次元説明変数の次元圧縮を予測目的に合わせて行う方法に関しては、目的変数が二値の場合や多値の場合に対応できる統計的手法を構築し、パッケージ「spcr」に組み込む。LASSO を用いた QTL 解析については、野生由来マウス系統 (MSM) と実験用マウス系統 (B6) では、マウスの攻撃行動に有意な差があることが知られていることに着目し、Kato et al. (2014) で提案した回帰分析法を発展させることにより、攻撃行動に影響する染色体領域の特定を試みる。マウス行動推定手法の研究については、精度向上を引き続き進める。また、深度センサによる情報を加味し、攻撃行動の検出も進める。また、マウス状態推定手法を他の大型動物（シバヤギ等）に適用することにより、開発アルゴリズムの汎用性検証を行う。その他、マウス行動に関するランキングデータについては、H25 に行った 3 すくみ検出に基づく階層モデルの方法を発展させ、さらに評価項目が複数ある場合に結果を統合する方法を開発する。マウス超音波データの解析については、今まで行ってきた 2 次元移動平均の方法で困難であった和声音の認識、分類に取り組む。

サブテーマ3では、野生マウス系統を含む多系統交雑によるヘテロジニアス集団を用いて、特定の行動形質（従順性）に関する選択交配をすでに多世代にわたって継続しており、それにより選択群と非選択群との間で明確な行動の違いが検出されている。また、イネの対称的形質の系統間の組換え系統間で交配自殖した集団における発現遺伝子の e-QTL 解析を行なう。これらの全ゲノム SNPs 解析を行い、遺伝子 SNP と行動（マウス）および遺伝子発現（イネ）との関連とそのネットワークを解明する。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視覚化し記述する手法を確立する。交雫固定化系統および突然変異系統におけ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、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（マウス、イネ、ゼブラフィッシュ、ショウジョウバエ）。

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、論文発表とインターネット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インターネットで公開する。新規に開発したアルゴリズム、ソフトウェアについてもインターネットで公開する。最後に研究成果を公開するためプロジェクト全体のシンポジウムを開催する。

[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、本機構以外の研究機関からも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、国立遺伝学研究所、東京工業大学、九州工科大学、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ーで共有したり、比較研究解析の基盤として利用したりする事も含め、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、シークエ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が、データ取得が多くの時間と労力を要するため、画像情報処理など先端的かつ効率的方法の体系化を進める。

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、本プロジェクト進行の間に、ゲノムプロジェクトが世界中で多様性研究、コホート研究などに向かい始め、国内外の状況も大きく変貌しつつある。加速的な連携やコミュニティー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共同研究パートナーを確保し、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

[国立遺伝学研究所・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] 豊田 敦、野口英樹、角谷徹仁、深川竜郎、堀 哲也、中村保一、
神沼英里、商 維昊

[新領域融合研究センター] 許山肖子、辰本将司、福多賢太郎、陳 薇、会津智幸、程 朝陽、
望月孝子

[東京工業大学] 黒川 顕

[九州工業大学] 矢田哲士

[慶應義塾大学] 楠原康文

[基礎生物学研究所] 長谷部光泰

[東北大学] 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、間野修平、加藤昇吾

[国立遺伝学研究所] 城石俊彦、倉田のり、高田豊行、小出 剛

[新領域融合研究センター] 木曾(岡)彩子、田邊 彰

[山形大学] 坂口隆之

[電気通信大学] 川野秀一

[九州大学] 二宮嘉行

[政策研究大学院大学] 土谷 隆

[愛知工科大学] 荒川俊也

[早稲田大学] Dou Xiaoling

[東京工業大学] 片山翔太

[筑波大学] 高橋阿貴

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]	久保貴彦、城石俊彦、高田豊行、天野孝紀、川上浩一、武藤 彩、 上田 龍、近藤 周、小出 剛
[統計数理研究所]	栗木 哲、藤澤洋徳、加藤昇吾、小山慎介
[国立情報学研究所]	北本朝展、宇野毅明
[新領域融合研究センター]	春島嘉章、堀内陽子、木曾（岡）彩子、和田浩則、田邊 彰、 近藤伸二
[ライフサイエンス統合データベースセンター]	内藤雄樹
[政策研究大学院大学]	土谷 隆
[大阪大学]	中谷 明
[新潟大学]	阿部貴志
[高知大学]	清澤秀孔
[九州大学]	二宮嘉行
[京都工芸繊維大学]	高野敏行
[首都大学東京]	相垣敏郎
[愛知工科大学]	荒川俊也
[電気通信大学]	川野秀一
[理化学研究所]	若菜茂晴、田村 勝
[東京大学]	岩田洋佳
[早稲田大学]	Dou Xiaoling
[東京工業大学]	片山翔太
[筑波大学]	高橋阿貴

[4] 研究の進捗状況

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、データ解析が進行しており、体系的な表現型-遺伝型相関解析の基盤が整い、相関解析の成果が生まれつつある。最終年度にむけて「遺伝機能システム学」の新たな展開を図っている。

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」

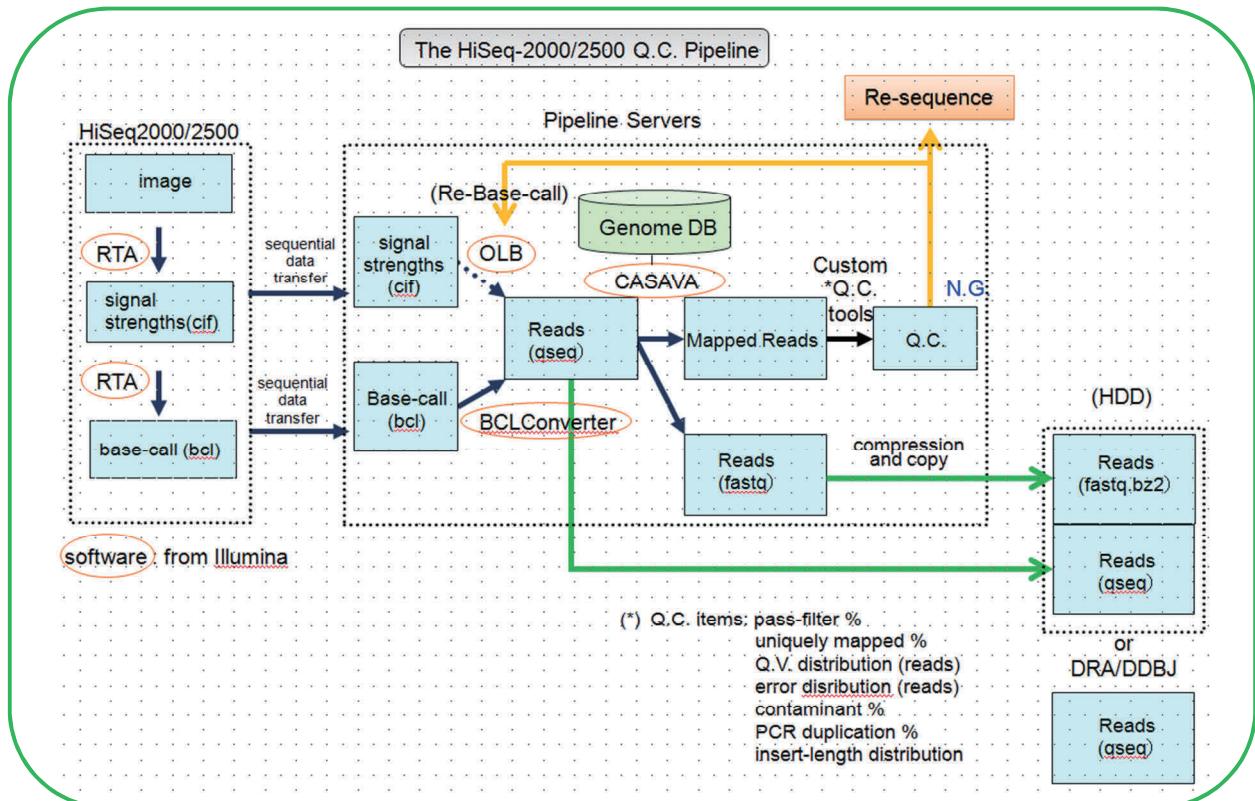
1-1 : 新型シークエンサーの稼働状況と大規模データ生産

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。本装置の導入当初の性能は、読み取り鎖長が 100 塩基、1 回の運転あたりのデータ生産量が約 200Gb であったが、その後の改良により、読み取り精度の大幅な向上と共に、1 回運転あたりのデータ



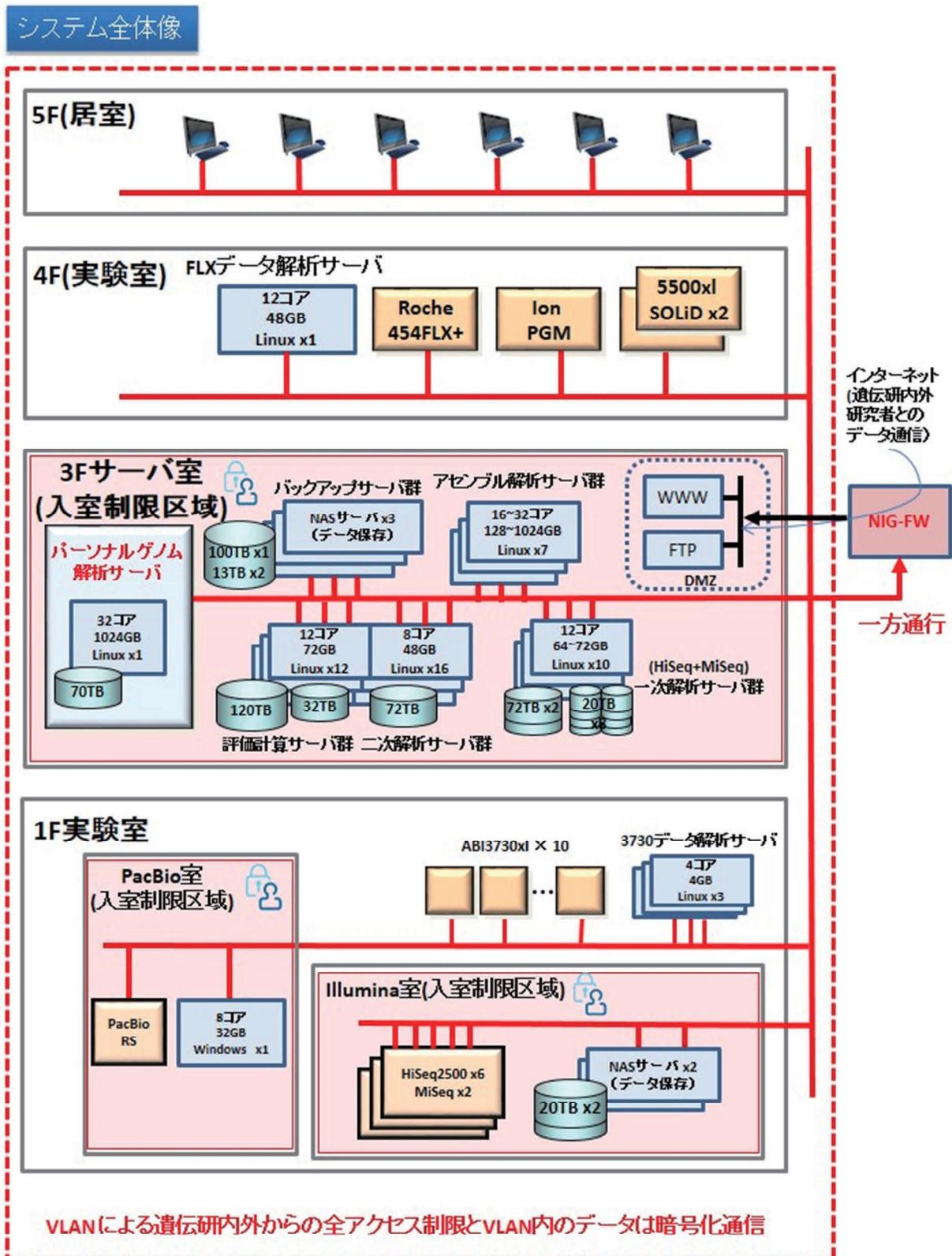
22 年度に導入した試料調製装置 c-BOT(左)と、HiSEQ2500 次世代型シーケンサ(右)。

生産量も 600Gb に増大されている。24 年度に HiSeq2500 タイプにアップグレードし、読み取り長 150bp×2, 180Gb のデータを 40 時間で産出するモードでの運転が可能となり、時間短縮、マッピングとアセンブリの効率と精度の改善が実現した。しかし、本プロジェクトで導入した DNA シーケンサは初期型であり、今後のアップグレードには対応できないため、第 4 世代装置の導入も含めて設備更新の検討が必要である。シーケンサの運用は、国立遺伝学研究所先端ゲノミクス推進センターのシーケンシング施設で行っており、世界的にも高精度のデータ生産が安定して行われるようになっている。26 年 11 月には、長年の懸案であった比較ゲノム解析研究室と先端ゲノミクス推進センターの新棟への統合移転が実現した。26 年度末には、システムの再立ち上げと新環境への慣熟もほぼ完了しており、27 年度以降の効率的運用が期待されるところである。遺伝学研究所先端ゲノミクス推進センターが運用する DNA シーケンス施設では、本プロジェクトで導入したマシン以外に、HiSeq シーケンサ 5 台と MiSeq シーケンサ 2 台、PacBio RSII 1 台、ABI3730 シーケンサ 6 台が主に使われている。これらのシーケンサーを駆使し、超大規模ゲノム関連情報の生産/解析基盤システムの構築を進めている。現在の HiSeq2500 型シーケンサによるデータ生産速度は 180Gb/40 hrs、リード長 150 bp に達しており、遺伝研スーパーコンピュータを効率よく利用するパイプライン構成と併せて国内トップクラスのゲノム解読能力を実現している。このシステムを活用し、当プロジェクト及び地球環境システム由来の試料について、25 年度末の累計で約 112Tb の塩基配列データを生産した。26 年度においては、マウス、イネ、シロイヌナズナ、ニワトリを対象に、再シーケンス、5C/HiC-seq 解析、ChIP-seq 解析、RNA-seq 解析、メチル化解析を行った。予算減のために解析シーケンス量も影響を受けており、26 年度の総レーン数は 122、総リード数 27,822,743,505、読み取り総塩基数は 4,099,863,949,650 bp であった。



上の図は、HiSeq2500 シーケンサのデータ評価/一次解析パイプラインを示している。システムのアップグレードに関わらず基本的にはこのプロセスを通過できたリードのみが以後の解析に使われる。次ページは生命情報センター西棟への統合移転後（26 年 11 月以降）のシステム配置の概略である。シ一

ケンサ室、共用サーバー室はカードキーによる施錠を行っており、さらに HiSeq 室、PacBIO 室、比較ゲノムサーバー室は生体認証による入室制限を実施し、セキュリティを高めるようになっている。また、27 年度からはシーケンサ室にセキュリティカメラを設置の予定である。





ABI3730xl キャピラリーシーケンサ



PacBIO RSII アナライザ



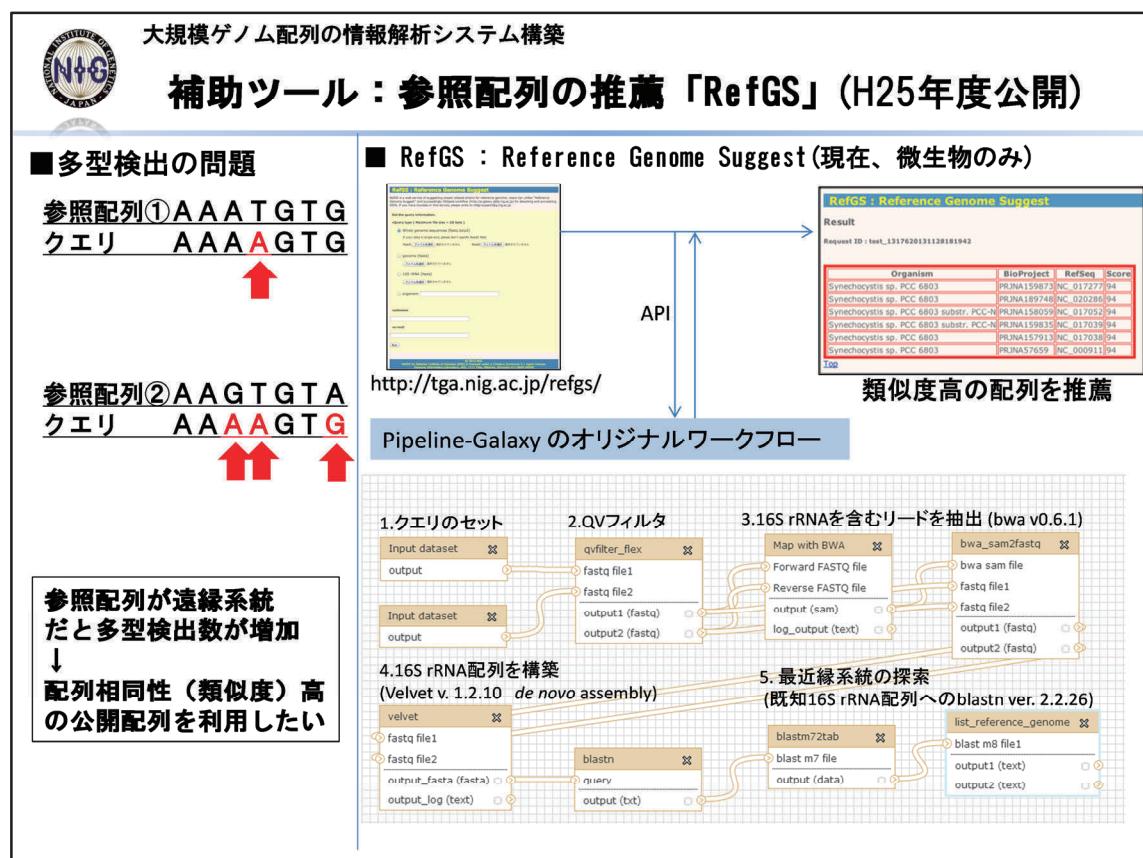
新型シーケンサ室

1-2：データ解析・管理システム：データ処理パイプラインの機能拡張

サブテーマ1 中村グループでは、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きた。基本的なデータ処理パイプラインのシステム構成は昨年度までと変更はない。平成26年度は、既存の基本的なデータ処理パイプラインのシステム構成に加えて、DNA多型注釈データベースDNAPodの解析生物種数を増やし、複数種でのオルソログ遺伝子の多型比較ワークフロー「OrthologAln」を開発した。

DNAPod (<http://tga.nig.ac.jp/dnapod/>) は、次世代シーケンサのアーカイブ配列の登録システム多型を系統単位で解析する DNA 多型注釈データベースで、本融合研究で構築された。結果を GWAS 用の遺伝子型データとしてユーザが利用できる。DNAPod には年度末までに、既発表のイネ 679 系統に加えて、新規にトウモロコシ 404 系統とソルガム 66 系統のデータを格納した。複数種でのオルソログ遺伝子の多型比較ワークフロー「OrthologAln」は、生物横断的に遺伝子領域の多型を系統単位で比較する機能を持つ。NBDC 統合植物データベース PGDBj (<http://pgdbj.jp/>) のオルソログ定義を利用して、アミノ酸配列から多型間距離を計算し、分子系統樹を構築する。実装された「OrthologAln」は、(<http://tga.nig.ac.jp/dnapod/OthologAln/>) から公開した。オルソログ遺伝子多型比較ツール「OrthologAln」の解析フローを下図に示す。Step1 で特定の生物種の参照配列遺伝子番号を選択すると、まずオルソログ遺伝子を検索する (Step2)。次にオルソログ遺伝子で多型を比較したい DNAPod 登録系統を選択する (Step3)。ユーザは DNAPod に加えて自分の解析系統のデータを追加する事が出来る (Step4)。各系統の多型はアミノ酸配列でアライメントを計算し、系統間で多型距離を計算する。NJ 法で分子系統樹を生成し、画像を出力する (Step5)。本ツールにより、ユーザはオルソログ遺伝子領域での系統単位の多型比較結果を視覚的に確認する事ができる。

今後は DNAPod の登録生物種数を多型解析の自動化により増やして、注釈データの大規模化を進めてデータマイニングに利用する。更に、ゲノム解析分野では次世代シーケンサの普及により、育種学分野や生態学分野等でフィールドでの表現型センシングや DNA 解析用試料をサンプリングする研究が進んでいる。フィールドデータは将来的に大規模データとなる為に、情報解析手法の開発が必須である。フィールド研究で重要となる地理位置情報は、DDBJ の BioSample データベース (SRA/GenBank 両方の試料情報を含む) に格納される事から、DNA 多型データの注釈情報として統合する。更にフィールドタイピング用のデータ収集基盤ツールを構築していく。



OrthologAln: オルソログ遺伝子領域多型比較機能(H26年度公開)

Step 1. 参照系統の遺伝子IDで オルソログ遺伝子検索

Orthologous Analysis

Step 1. Input TranscriptID of Rice (RAP-DB) or Maize (MaizeSequence.org) or Sorghum (JGI), example id); Rice: Os060133000-01; Maize: GRMZM2G044744_T01; Sorghum: Sb001g021190.1

Os060133000-01 Search

Step 2. 多型比較用オルソログ遺伝子を表示

Step 2. Confirm Orthologous TranscriptID.

Rice (6 Results)	TranscriptID	RefSeq ProteinID	RefSeq Definition
<input checked="" type="checkbox"/>	Os010142300-01	NP_001041981.1	Os01g0142300
<input type="checkbox"/>	GRMZM2G044744_T01	NP_001123590.1	
<input checked="" type="checkbox"/>	Os020807100-01	NP_001048453.1	Os02g007100, partial
<input type="checkbox"/>	Os020807100-02	NP_001048454.1	Os02g007100, partial
<input checked="" type="checkbox"/>	Os05033600-01	NP_001056142.1	Os05g033600
<input type="checkbox"/>	Os060133000-01	NP_001056704.1	Os06g0133000

Maize (5 Results)	TranscriptID	RefSeq ProteinID	RefSeq Definition
<input checked="" type="checkbox"/>	GRMZM2G1004744_T01	NP_001123590.1	hypothetical protein
<input type="checkbox"/>	GRMZM2G100652_T02	NP_001130556.1	starch synthase IV
<input type="checkbox"/>	GRMZM2G117153_T01	NP_001150015.1	uncharacterized protein LOC100192061 precursor
<input type="checkbox"/>	GRMZM2G117153_T01	NP_001150015.1	glycosyl transferase, group 1 family protein

Sorghum (7 Results)	TranscriptID	RefSeq ProteinID	RefSeq Definition
<input type="checkbox"/>	Sb02g008800.1	XP_002461829.X	hypothetical protein
<input checked="" type="checkbox"/>	Sb02g009870.1	XP_002461889.X	hypothetical protein
<input checked="" type="checkbox"/>	Sb03g006480.1	XP_002457385.1	hypothetical protein
<input type="checkbox"/>	Sb03g006480.1	XP_002457385.1	SORBIRAFT_02g008800
<input type="checkbox"/>	Sb03g006480.1	XP_002457385.1	SORBIRAFT_02g009870
<input type="checkbox"/>	Sb03g006480.1	XP_002457385.1	SORBIRAFT_02g009870
<input type="checkbox"/>	Sb03g006480.1	XP_002457385.1	SORBIRAFT_02g009870

Step 3. 解析したいDNAポッド 登録系統を選択

Step 3. Select Strain.

Total limit : 50 strains

Rice				
Any Species	Any Subspecies	Context Strain Name	Context ID	ID
<input type="checkbox"/>	SRA sample id	species	subspecies	strain
<input type="checkbox"/>	SR5086334	Oryza sativa	Javaica	Haginomae Mochi
<input type="checkbox"/>	SR5086335	Oryza sativa	Indica	Mirai
<input type="checkbox"/>	SR5086336	Oryza sativa	Indica	Shokuhwan
<input type="checkbox"/>	SR5086337	Oryza sativa	Javaica	Darmali
<input type="checkbox"/>	SR5086338	Oryza sativa	Indica	Laung Pratet
<input type="checkbox"/>	SR5086339	Oryza sativa	Indica	rice

Maize				
Any Species	Any Subspecies	Context Strain Name	Context ID	ID
<input type="checkbox"/>	SRA sample id	Zea species	Subspecies	strain
<input checked="" type="checkbox"/>	SR5072476	Zea mays	marvelous	TILK
<input type="checkbox"/>	SR5072477	Zea mays	marvelous	TH15
<input type="checkbox"/>	SR5072480	Zea mays	mays	TX303
<input type="checkbox"/>	SR5072482	Zea mays	mays	V1051247
<input type="checkbox"/>	SR5072483	Zea mays	mays	V1051248
<input type="checkbox"/>	SR5072484	Zea mays	mays	V1051249

Sorghum				
Any Species	Any Subspecies	Context Strain Name	Context ID	ID
<input type="checkbox"/>	SRA sample id	species	subspecies	strain
<input checked="" type="checkbox"/>	SR5266351	Sorghum bicolor	bicolor	Rio
<input type="checkbox"/>	SR5266352	Sorghum bicolor	bicolor	E-Tian
<input type="checkbox"/>	SR5266353	Sorghum bicolor	bicolor	Keller
<input type="checkbox"/>	SR5308944	Sorghum bicolor	bicolor	ZM/A 5298
<input type="checkbox"/>	SR5309195	Sorghum bicolor	bicolor	PI 505722
<input type="checkbox"/>	SR5309196	Sorghum bicolor	bicolor	PI 508366

Step 4. ユーザの 解析データUpload

Upload your variant call file (.option)

Select Organism: Rice

Select a variant call format (VCF) file:

Upload file includes only homozygous SNPs data detected with GT fields in VCF.

File name (excluding extension) is the strain name.

A prohibited character... (underscore)

example) Koshibakari.vcf → Koshibakari

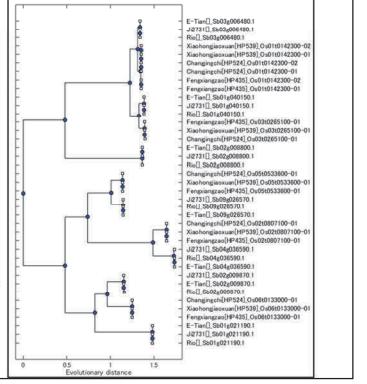
Sample upload file (Rice): Koshibakari.vcf

ファイル選択: Koshibakari.vcf

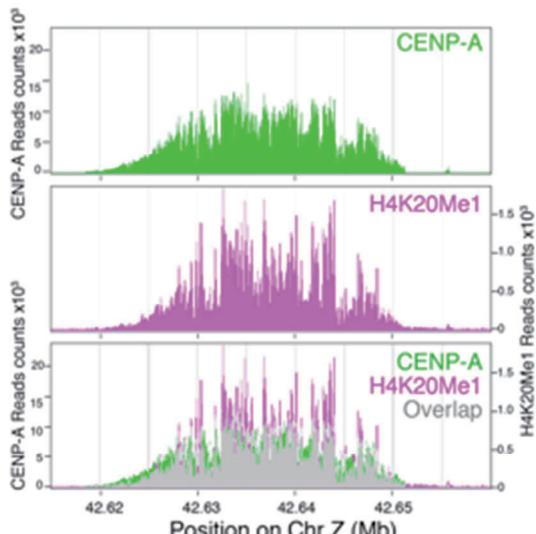
E-Mail (option): smi@q.ac.jp (recommended) Email notification when your job is completed.

Submit your job

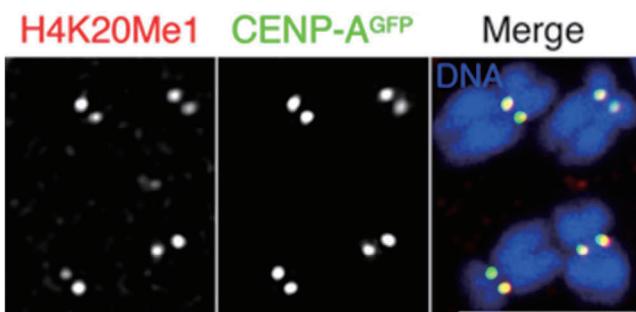
Step 5. 分子系統樹による 系統間多型比較



1-3 : 高次ゲノム機能領域の解析



ネオセントロメアにおける CENP-A 抗体と H4K20me1 抗体による ChIP-seq プロファイル



ヒト細胞における H4K20me1 のセントロメア局在

サブテーマ1のチーム内では、次世代シーケンサを活用した大規模シーケンシングを基盤とする融合的研究が進み、昨年度からの研究を継続発展させている。深川グループによる染色体高次構造に関する研究では、これまでにネオセントロメアを培養細胞で構築できる独自に開発してきた。本年度は、その実験系を活用してセントロメア形成に必須なエピジェネティックマーカーの探索をヒストンに対する特

異抗体を利用した ChIP および次世代シーケンサ解析によって行った。その結果、セントロメアのクロマチンには、ヒストン H4 の 20 番目のリジンが特異的にモノメチル化されることが明らかになった（上図）。さらに、そのメチル化が起きないようすると、動原体構築に異常がおこることが判明した。解析には、深川グループの研究に加え、ゲノム解析を専門とする藤山のグループと、情報学専門とする榎原のグループが融合することで行なわれた。これらの成果に関しては、Dev Cell, 29: 740-749, 2014; Dev Cell, 30: 496-508, 2014; Curr Biol. 24: R946-947, 2014 へ論文報告した。

1-4 : 高次ゲノム機能領域の解析-2

サブテーマ 1 の角谷グループは、ゲノム機能調節の重要なメカニズムの一つであるメチル化とトランスポゾン転移について、シロイヌナズナやイネをモデルとした研究を進めている。2014 年度は、クロマチン再構成因子 DDM1 (DECREASE IN DNA METHYLATION 1) 遺伝子の変異体を用いた解析を行った。この変異体では、ゲノムワイドに反復配列領域の DNA メチル化が低下する。一方、不思議なことに少数の遺伝子で DNA メチル化が上昇することが知られていた (Saze & Kakutani 2007 EMBO J; Sasaki et al 2012 Plant J)。ゲノムワイドの DNA メチル化解析を行ったところ、遺伝子における DNA メチル化は変異がホモになった 1 世代目ではおこらず、数世代を経た個体のみで見いだされたことがわかった。さらに興味深いことに、*ddm1* 変異体由来の染色体を持つ個体では、他の座で異所的な DNA メチル化の上昇が観察された（図）。これらの結果から、ゲノム全体での抑制クロマチンの量を補正する負のフィードバック機構の存在が示唆された (Ito et al PLoS Genetics, in press)。

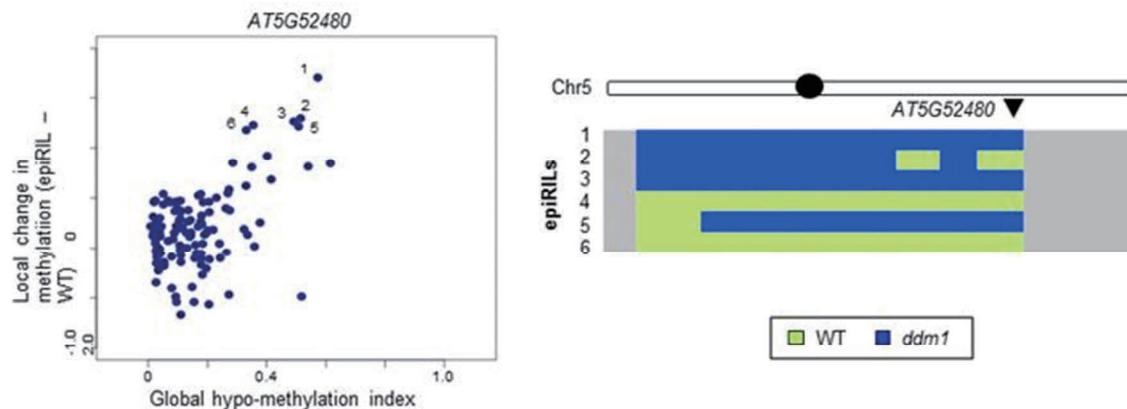


図 (左) *ddm1* 由来する染色体によるゲノム全体の DNA 低メチル化と局所的 DNA 高メチル化の関係。両者に正の相関があった。(右) 局所的な高メチル化のおきた座は、*ddm1* 由来のものとそうでないものの両方が含まれた。これらの結果から、ゲノム全体での抑制クロマチンの量を補正する負のフィードバック機構の存在が示唆される。

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」

2-1 : 疑似相関を利用した新しい多重性調整手法の開発 [二宮、栗木、高田、城石]

前年度までにおいて、未知の相関をもつ複数の t 検定からなる多重検定問題に対し、ある種の疑似相関を未知の相関の代わりに用いる新しい多重性調整法を開発した。そして、純系のマウス MSM とそのコンソミックマウスとの特徴量を比較する多重検定問題に適用した。結果、元々の解析 (Takada et al., 2008) で用いられている Bonferroni 法に比べ、56 個（全部で 2059 個）の仮説をさらに棄却することができた。

本年度は、その手法を理論的に整備するとともに、数値実験による有効性の確認を行った。提案手法

は疑似相関を用いているが、どのような疑似相関でも用いてよいわけではない。理論的な整備においては、まずFWER (familywise error rate) を漸近コントロールするような疑似相関を与えた。また、その疑似相関の使い方は必ずしも一通りではないため、ある使い方のクラスの中で、ある種の最適性を満たすような使い方を提案した。さらに、疑似相関行列が非正定値となってしまう可能性がある問題に対し、FWER の漸近コントロールを保証しつつ疑似相関行列を正定値化する二つの方法（下表では Proposal 1、Proposal 2 と記載）を考えた。

数値実験においては、ある特殊な相関構造をもった多重検定問題を考え、提案手法と Bonferroni 法、通常の相関を用いる MaxT 法 (Dudoit and van der Laan, 2008, 2.6 節)、MaxT 法をステップダウン方式で用いる StepMaxT 法 (Dudoit and van der Laan, 2008, 5.2 節) を比較した。表 1 では FWER を評価している (ρ : 検定間の相関、 n : 標本数、 p : 検定の数)。提案手法は FWER を MaxT 法と同程度にコントロールしていることが確かめられる。一方、表 2 では検出力を評価している (μ : 二群の母平均の差)。この結果により、提案手法は単に検出力の向上を保証するだけでなく、通常の相関の利用やステップダウン法の利用による向上よりずっと大きな向上が場合によっては見込めることがわかった。

表 1 提案手法と既存手法の FWER 比較。数値はサイズを 5% と設定したときの FWER の平均(%)、丸カッコ内の数値はその標準偏差(%)、鍵カッコ内の数値は最小 p 値の平均である。

ρ	n	P	Bonferroni	(Step)MaxT	Proposal 1	Proposal 2
0	12	50	4.53 [52.5]	4.64 (0.54) [51.3]	4.59 (0.59) [51.1]	4.64 (0.56) [51.2]
0.6	12	50	3.77 [64.2]	5.07 (0.82) [48.5]	5.15 (1.07) [48.8]	5.26 (1.20) [48.7]
0.3	6	50	3.53 [57.1]	4.38 (0.75) [48.6]	4.24 (0.72) [49.3]	4.21 (0.82) [48.9]
0.3	18	50	4.44 [56.7]	4.90 (0.72) [50.5]	4.86 (0.64) [50.5]	4.84 (0.71) [50.6]
0.3	12	20	4.26 [58.5]	4.92 (0.53) [49.0]	4.99 (0.59) [48.9]	4.96 (0.55) [49.0]
0.3	12	80	4.59 [54.0]	5.04 (0.82) [48.5]	5.06 (0.87) [48.6]	4.99 (0.83) [48.6]

表 2 提案手法と既存手法の検出力比較。数値は検出力(%)、鍵カッコ内の数値は平均 p 値(%)である FWER 比較。数値はサイズを 5% と設定したときの FWER の平均(%)、カッコ内の数値はその標準偏差(%)である。

P	n	p	μ	Bonferroni	MaxT	StepMaxT	Proposal 1	Proposal 2
0	12	50	1.2	30.4 [34.4]	30.9 [33.8]	34.7 [28.8]	39.0 [23.8]	38.8 [23.7]
0.6	12	50	1.2	30.5 [34.1]	36.3 [25.1]	41.3 [23.3]	54.2 [17.0]	54.6 [16.5]
0.3	6	50	1.2	6.9 [65.9]	7.7 [59.6]	8.2 [58.6]	14.7 [48.0]	12.4 [50.3]
0.3	18	50	1.2	59.8 [14.1]	61.6 [12.5]	70.2 [8.7]	74.5 [7.0]	74.3 [7.0]
0.3	12	20	1.2	46.8 [20.4]	48.6 [17.1]	57.1 [13.4]	63.8 [10.5]	64.0 [10.3]
0.3	12	80	1.2	26.3 [39.9]	27.3 [35.6]	31.8 [32.2]	41.2 [25.3]	41.6 [24.7]
0.3	12	50	0.9	14.3 [55.0]	14.9 [50.6]	16.9 [48.5]	21.4 [42.9]	21.8 [42.3]
0.3	12	50	1.5	55.8 [14.9]	57.8 [12.9]	67.8 [8.8]	77.2 [5.7]	77.1 [5.6]

2-2：高次元説明変数の次元圧縮を予測目的に合わせて行う方法 [川野、藤澤、高田、城石]

サブテーマ 3 (城石プロジェクト) の研究成果として、コンソミック系統および両親系統（計 33 種類）に関する 36 形質の測定データが得られている。そのデータに基づいて、コンソミック系統を形質によって特徴づけられないか、というテーマを考えることができる。この問題は統計学の言葉では、形質を説明変数、系統を目的変数とする判別分析である。

判別分析では離散データをターゲットにする。しかし、最初から離散データをターゲットにするのは大変なので、まずは連続データをターゲットとして関連手法を提案することを考えた。その具体的な手法は昨年度の報告書で触れた通りである。

本年度はその成果を論文としてまとめて投稿した。本論文はその後、多くのデータ解析とシミュレーションを追加し、Computational Statistics and Data Analysis に受理されたところである。また、その統計手法を実行できる統計ソフトウェアを R の上で実装し、パッケージ「spcr」を配布している。

spcr: Sparse principal component regression

This package computes the sparse principal component regression. The regularization parameters are optimized by cross-validation.

Version: 1.2.1
Published: 2014-12-04
Author: Shuichi Kawano
Maintainer: Shuichi Kawano <skawano at ai.is.uoc.ac.jp>
License: [GPL-2](#) | [GPL-3](#) [expanded from: GPL (≥ 2)]
URL: <https://sites.google.com/site/shuichikawanoen/>
NeedsCompilation: yes
CRAN checks: [spcr results](#)

Downloads:

Reference manual: [spcr.pdf](#)
Package source: [spcr_1.2.1.tar.gz](#)
Windows binaries: r-devel: [spcr_1.2.1.zip](#), r-release: [spcr_1.2.1.zip](#), r-oldrel: [spcr_1.2.1.zip](#)
OS X Snow Leopard binaries: r-release: [spcr_1.2.1.tgz](#), r-oldrel: [spcr_1.2.1.tgz](#)
OS X Mavericks binaries: r-release: [spcr_1.2.1.tgz](#)
Old sources: [spcr archive](#)

2-3：回帰分析を用いたマウス活動量データの遺伝解析〔加藤、栗木、小出〕

今年度は、昨年度までの研究結果をまとめた論文を学術誌「Heredity」に出版した。出版された論文の概要は以下の通りである。

Nishi *et al.* (2010) は、野生由来マウス系統 (MSM) と実験用マウス系統 (B6) の行動形質の比較を行い、2 系統のマウスの活動量 (home-cage activity) に有意な差があることを明らかにした。さらに、第 6 番染色体の遺伝子型の違いが、活動量に大きく影響していることも示した。

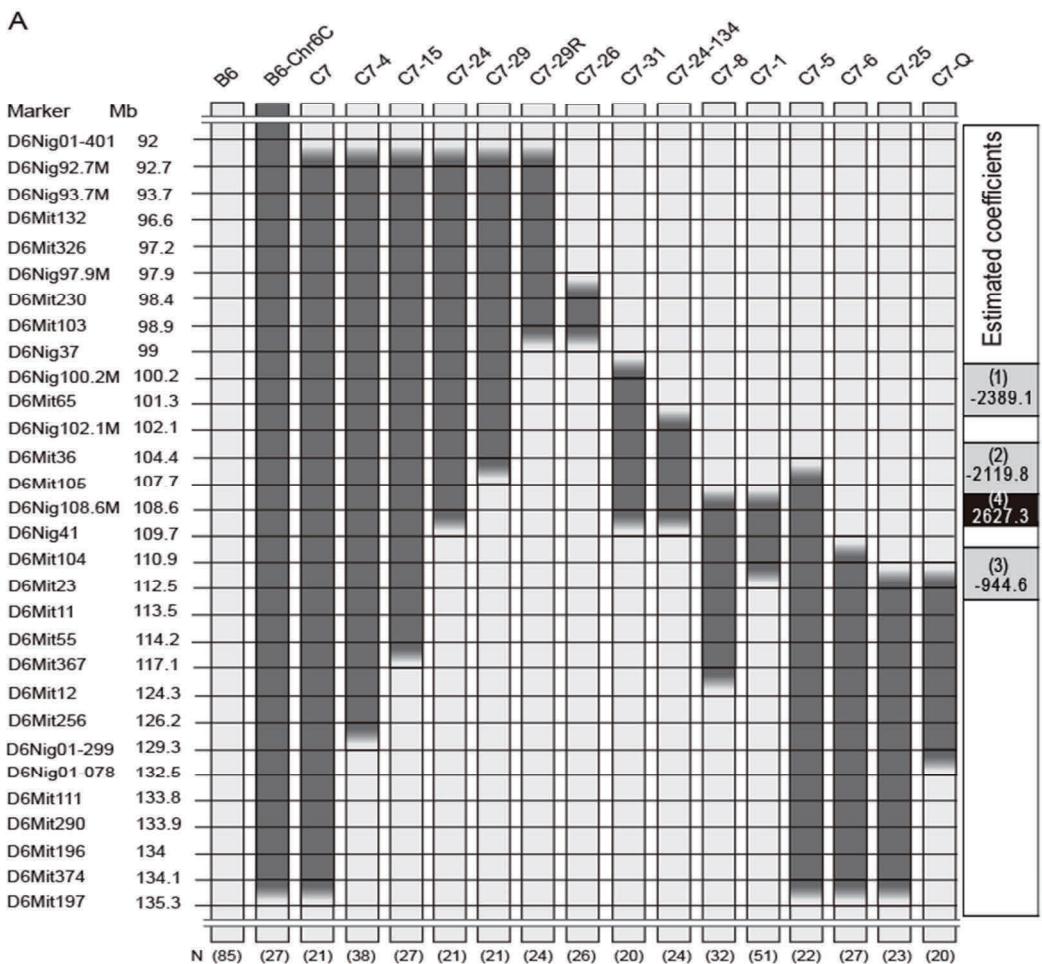
そこで論文では、第 6 番染色体のどの領域がマウスの活動量に影響するのかを調べるため、第 6 番染色体の一部を MSM 由来のもので置換した B6 を親系統とするコンジェニック系統を 15 系統樹立し、それらの系統とその親系統の活動量データから回帰分析の手法を用いて活動量に影響する領域の特定を試みた。

回帰分析においては、マウス系統の活動量を被説明変数、microsatellite marker における遺伝子型を説明変数とした。しかし今回のデータに対して、通常の最小二乗法に基づく回帰モデルを当てはめることに大きな問題点があった。それは、microsatellite marker (説明変数) の数がマウスの系統数 (標本数) よりも大きくなるという問題、つまり $p >> n$ 問題である。このようなデータに対しては、通常の最小二乗誤差に基づく回帰モデルでは解が一意に求まらないことが統計学において知られている。

この問題を克服するため、論文では以下のように解析を行った。まず初めに、lasso (Tibshirani, 1996) を用いて変数選択を行い、説明変数が標本数よりも小さくなるまで変数を削減した。それから、通常の最小二乗法による回帰モデルを当てはめ、all subsets を用いてさらなる変数選択を行った。

その結果、活動量に有意に影響を与える染色体領域を推定することに成功した。また、以下の図に示されているように、回帰係数の推定値から、推定された染色体領域において活動量がどの程度増加または減少するのかもあわせて推測することができた。今回の解析により活動量を大きくすると推定された

領域は、追加して行った実験の結果とも一致した。



2-4 : B スプラインコピュラの提案とそのマウス・コンソミック系統データ解析への応用

[Dou、栗木、城石、高田]

H24 年度の研究成果として、ベルンシュタインコピュラを提案し、それをマウス・コンソミックデータ解析に適用した。本年度は、ベルンシュタインコピュラを特別の場合として含む B スプラインコピュラを提案し、同種の解析を行った。

次数 d で節点を

$$\underbrace{t_{-d} = \dots = t_0}_{d+1} = 0 \leq \underbrace{t_1 \leq \dots \leq t_{m-1}}_{m-1} \leq 1 = \underbrace{t_m = \dots = t_{m+d}}_{d+1}.$$

(総数 $m + 2d + 1$) とする区間 $[0,1]$ 上の B スプライン関数を $N_{k-d-1}^d(x)$ とおく (de Boor, 2001)。このとき、2 变量 B スプラインコピュラは

$$c(u, v; R) = \sum_{k=1}^n \sum_{l=1}^n r_{kl} \phi_{k,n}(u) \phi_{l,n}(v)$$

で定義される。ここで R は $n \times n$ はパラメータ行列で $r_{kl} \geq 0$, $\sum_{l=1}^n r_{kl} = q_{k,n}$, $\sum_{k=1}^n r_{kl} = q_{l,n}$ をみたすもの。また

$$\phi_{k,n}(x) = \frac{1}{q_{k,n}} N_{k-d-1}^d(x), \quad q_{k,n} = \int_0^1 N_{k-d-1}^d(x) dx$$

である。

特殊な場合として、 $m=1$ 、すなわち 0、1 以外の節点がない場合を考えると、 $N_{k-d-1}^d(x)$ はベルンスタイン多項式となり、また $q_{k,n} = 1/n$ である。

$c(u;v;R)$ が実際にコピュラ密度関数であることは、性質

$$1 = \sum_{k=1}^n N_{k-d-1}^d(x) \text{ for all } x \ (n = k + m)$$

より従う。これはいわゆる 1 の分割である。

B スプラインコピュラについて、いくつかの理論的な性質を導くことができる。まず最大の相関は、パラメータ行列 R が対角行列

$$R = (r_{kl}) = \text{diag}(q_{k,n})_{1 \leq k \leq n}.$$

のとき達成される。またその最大値は、節点が等間隔 $t_i = i/m$ ($i = 1, \dots, m - 1$) のとき、

$$1 - \frac{d+1}{(n-d)^2} + \frac{d(d+3)(2d+3)}{5(d+2)(n-d)^3}, \quad n \geq 2d$$

であったえられる。これより、B スプラインコピュラはベルンシュタインコピュラよりも広いレンジの相関を表現できることが分かる。

モデル当てはめの手順は、ベルンシュタインコピュラの場合とほとんど同様である。データ (x_t, y_t) , $t = 1, \dots, N$, をランク統計量

$$u_t = \frac{N}{N+1} F_N(x_t), \quad v_t = \frac{N}{N+1} G_N(y_t)$$

に変換し (F_N , G_N は経験分布)、EM アルゴリズム (Dou, et al., CSDA, 印刷中) によって、擬似 MLE

$$\hat{R} := \underset{R}{\operatorname{argmax}} \prod_{t=1}^N c(u_t, v_t; R)$$

を求める。密度関数の推定量は $c(F_N(x), G_N(y); \hat{R}) \hat{f}_N(x) \hat{g}_N(y)$ で得られる。

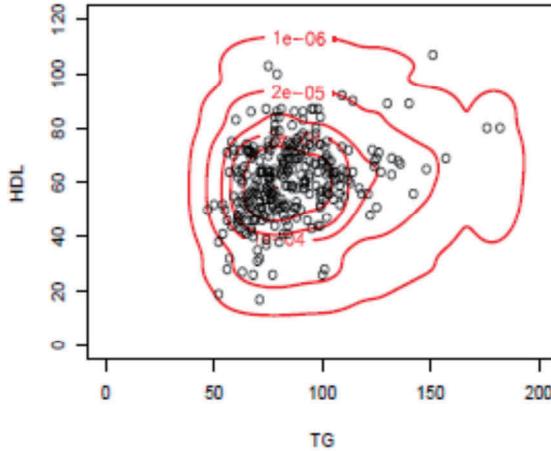


図 1 コンソミックマウスデータと B スプラインコピュラの当てはめ

図 1 はコンソミックマウスデータに対して本モデルを当てはめた例である ($d = 2$; $n = 3$)。Triglycerides (TG) と plasma high-density lipoprotein cholesterol (HDL) は代謝症候群に関する重要な指標であり、人間の心臓血管発病と関連性がある。本データは、肥満に関する遺伝子を特定するために、TG と HDL データが 314 匹の生後 10 週間のメスのコンソミックマウスから観測した (Takada

and Shiroishi, 2012) ものである。本データセットは C57BL/6 (B6) と MSM/Ms (MSM) の純粋の血統で作られた 30 種類のコンソミック血統を含んでいる。このような不均一性の高いデータセットに対して、B スプラインコピュラのアプローチは柔軟なモデルを提供することが分かる。

2-5：マウス並走状態除去による自動推定アルゴリズム一致率向上の試み

[土谷、荒川、柿原、高橋、杉本、田邊、小出]

人間観察により動画像のフレーム単位で並走状態と判定したデータと、その際のマウス行動のデータを教師データとし、サポートベクタマシン (SVM) を用いて並走状態を学習させた。その結果に基づき、SVM を適用しない場合と一致率 (Percent concordance) を比較した。その結果、SVM を適用した場合、若干の一致率向上が見られた。しかし大幅な改善には至っていないため、今後更に並走除去の手法を検討し、一致率の向上を図る。

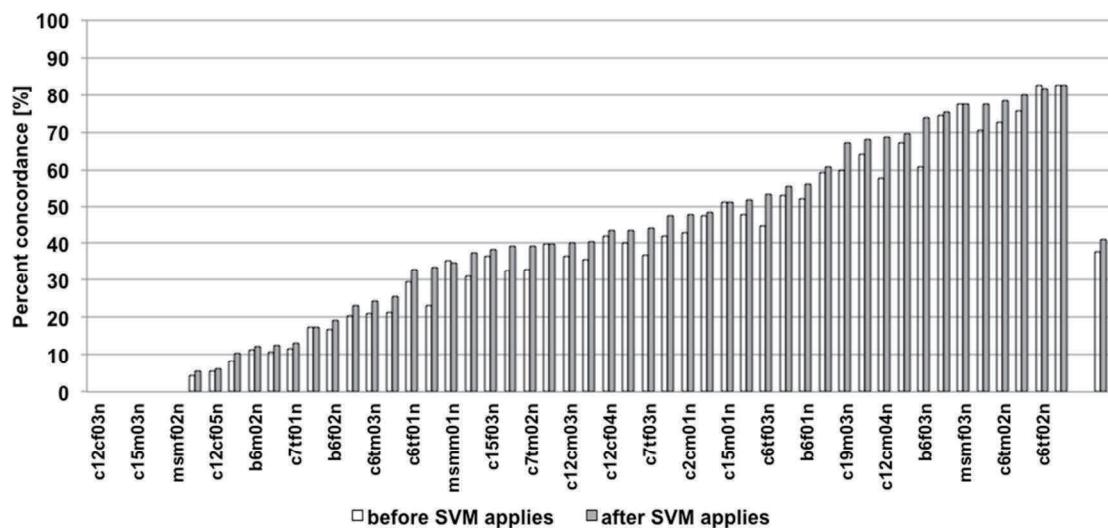


図 1 SVM 適用前後の一致率の比較

攻撃行動の検出については、深度センサを用いた検証を進めているが、画像処理に不十分な点があるため、マウスの 3 次元物理量が上手く抽出できていない。画像処理アルゴリズムを改善し、3 次元物理量が抽出できるように進める。

また、従来の推定に関して更に検討を行うため、隠れマルコフモデルの物理量について、マウスの物理量の分布を実データに基づいたヒストグラムを用いた場合と、混合ガウス分布を用いた場合を比較した。目視観察による状態と隠れマルコフモデルによる状態推定結果を比べて「一致率」の観点から検証した場合、Sniffingにおいては混合ガウス分布 (80.3%) の方がヒストグラム (78.6%) に比べて若干良い結果であった。しかし、Following の場合はヒストグラム (37.7%) の方が混合ガウス分布 (30.0%) に比べて一致率が高い傾向が見られた。

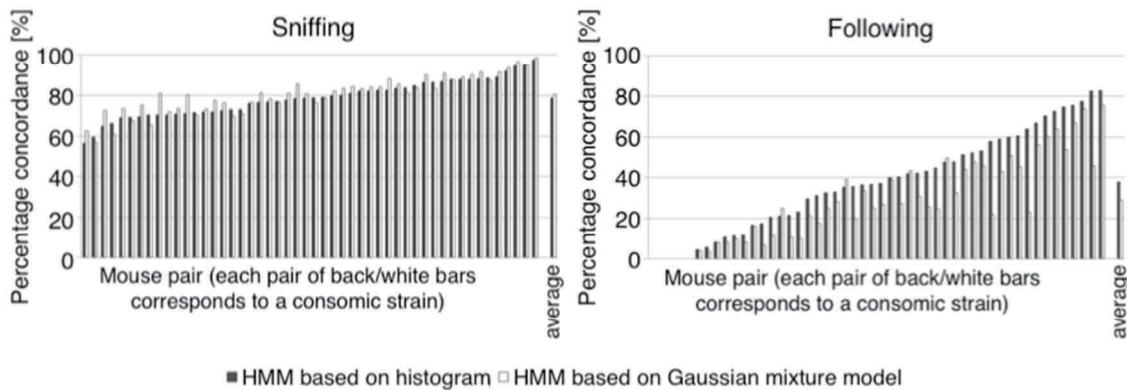


図 2 ヒストグラムと混合ガウス分布による一致率の比較

また、これまでの研究成果のまとめとして、Springer より論文招待され、投稿した。

上記に加えて、マウス状態推定手法の汎用化を検証するため、シバヤギを例とした大型動物の状態推定を試みた。マウス状態推定と同様に、隠れマルコフモデルを用い、シバヤギの座標と速度を物理量とし、雌シバヤギの「発情 (Estrus)」「非発情 (Not estrus)」および雌シバヤギが「雄に接近する行動 (Approach)」「雄の近くに滞在する行動 (Standing)」の推定を実施した。その結果、マウス状態推定において構築した手法が大型動物でも適用できる可能性が示唆された。マウス状態推定と同じアルゴリズムを組み込むことにより、日常の繁殖管理システム構築ができることが期待される。

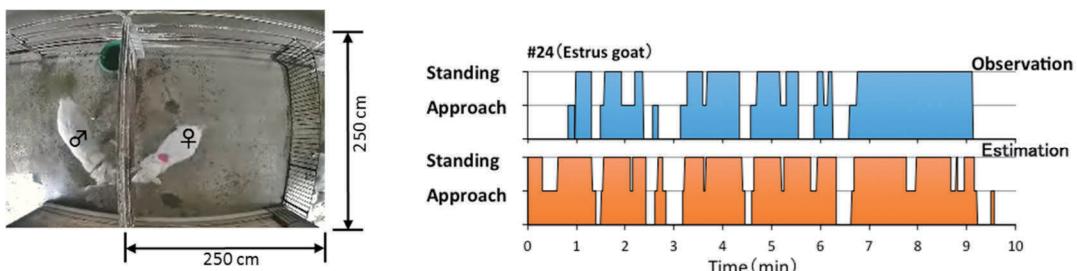


図 3 シバヤギによる状態推定実験(左)と、目視観察と状態推定結果の比較(右)

2-6：マウス超音波発声データの解析 [Dou、栗木、杉本、高橋、小出]

マウス超音波発声データを自動的に解析するために、本研究ではマウスの超音波発声 (USV) データのパターン認識を行うための効率的なデータ解析手法を提案し、フリーソフトウェアで解析を行う事を目指している。本研究は大阪大学白旗慎吾名誉教授の協力により進めている。

マウスの超音波発声の長さは様々である。今まででは、発声パターンの形だけに注目して分類を行ってきたが、今年度は最初に長さで分類した後に、形でクラスタリングを行った。このことによって、分類の精度が良くなかった。この方法によって、長さの異なる USV の機能についても調べる事ができるようになった。

以下は BALB/cAnN マウスの発声データの解析結果の一例である。図 1 は USV の長さのヒストグラムである。図 2 は USV の長さについて 4 つのグループに分けたものである。短い USV 曲線をさらに形で分類した結果は図 3 と図 4 である。

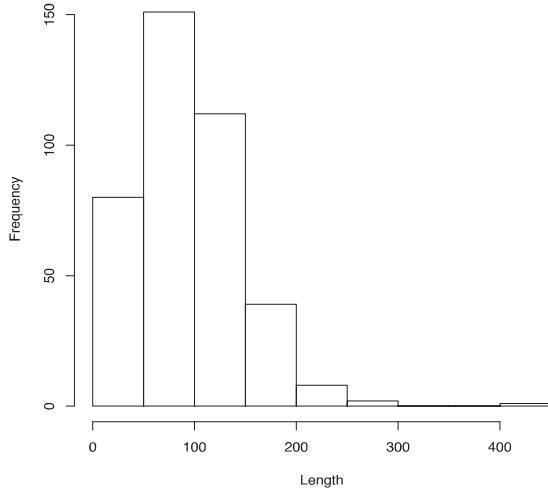
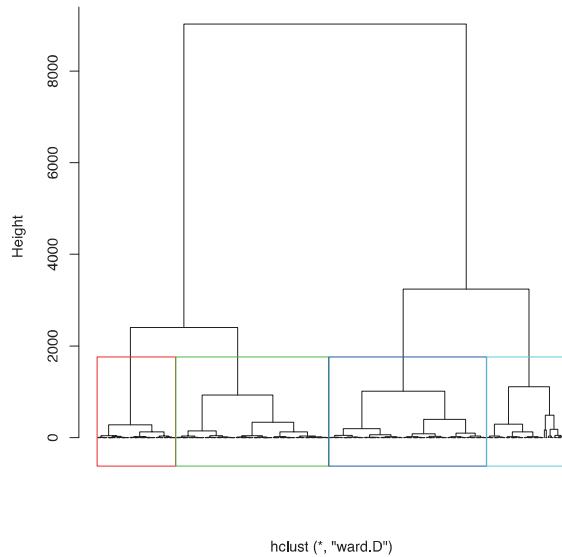


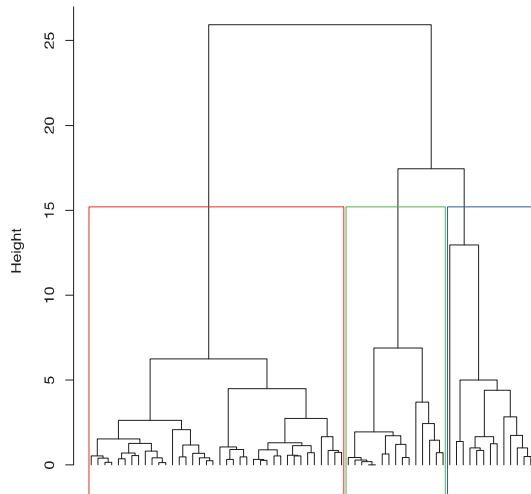
図 1



`hclust (*, "ward.D")`

図 2

Group I



`hclust (*, "ward.D")`

図 3

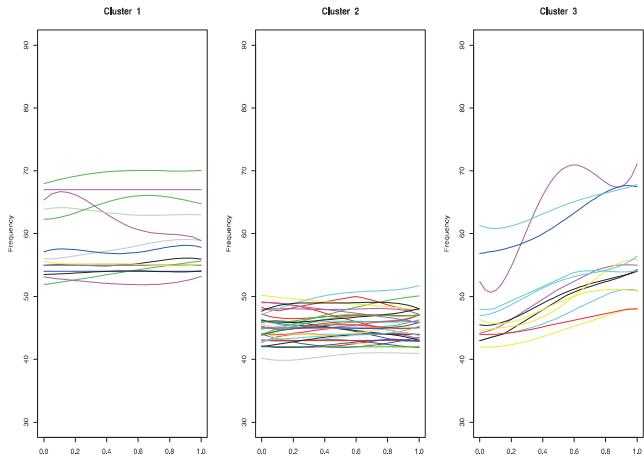


図 4

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」

3-1：次世代シーケンサによるアレル発現量バイアス解析およびエピゲノム解析

〔豊田、藤山、阿部、中谷、近藤、清澤、藤澤、栗木、二宮、原、天野、岡、高田、城石〕

本テーマの目的は、リファレンス配列が整備されている複数マウス系統間のゲノム多型情報と、これら系統間の交配系から得られた試料を用いた各種表現型収集、遺伝子発現解析およびエピゲノム解析を

を行うことにより、対立遺伝子間の発現量差を制御するゲノム機能の解明に繋がる研究を行うことである。主に使用するマウスは、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 C57BL/6J (以降 B6 と表記)、および国立遺伝学研究所が樹立し、我々がその塩基配列を明らかにした日本産野生由来の MSM 系統、さらにはこれらの正逆交配により得られた F1 個体である。両系統間には、亜種に相当する 1,400 万カ所以上の莫大な SNP 情報があるため、特に F1 個体では、このゲノム多様性を利用した遺伝子発現情報を次世代シーケンサによる RNA-Seq 法で解析し、両親由来の対立遺伝子発現バイアスのプロファイリングを行うことが可能である。我々はこれまでに、B6 および MSM の SNP 情報を取り入れ、B6-MSM 両アレルを区別した遺伝子発現解析を行うためのプラットフォームを DDBJ スパコン内に構築し、解析を行っている。

本年度は、昨年度までにサブテーマ 1 との連携により得られた、肝臓および精巣由来サンプルの解析 (フェーズ 1) について、性別および組織ごと、各 6 サンプルのデータ解析について、発現遺伝子の検出に関わるパラメータの最適化を行い、アレル発現量バイアスを示す遺伝子群の抽出を行った。図 1 に結果を示す。雌雄の肝臓および、精巣について、緑円は MSM アレルの発現量が B6 アレルよりも高い遺伝子群、青円は B6 アレルの発現量が MSM アレルよりも高い遺伝子群、赤円は、BM および MB どちらの組み合わせでも母方のアレルの発現量が高い遺伝子群、黒円は同様に父方のアレルの発現量が高い遺伝子群を示す。

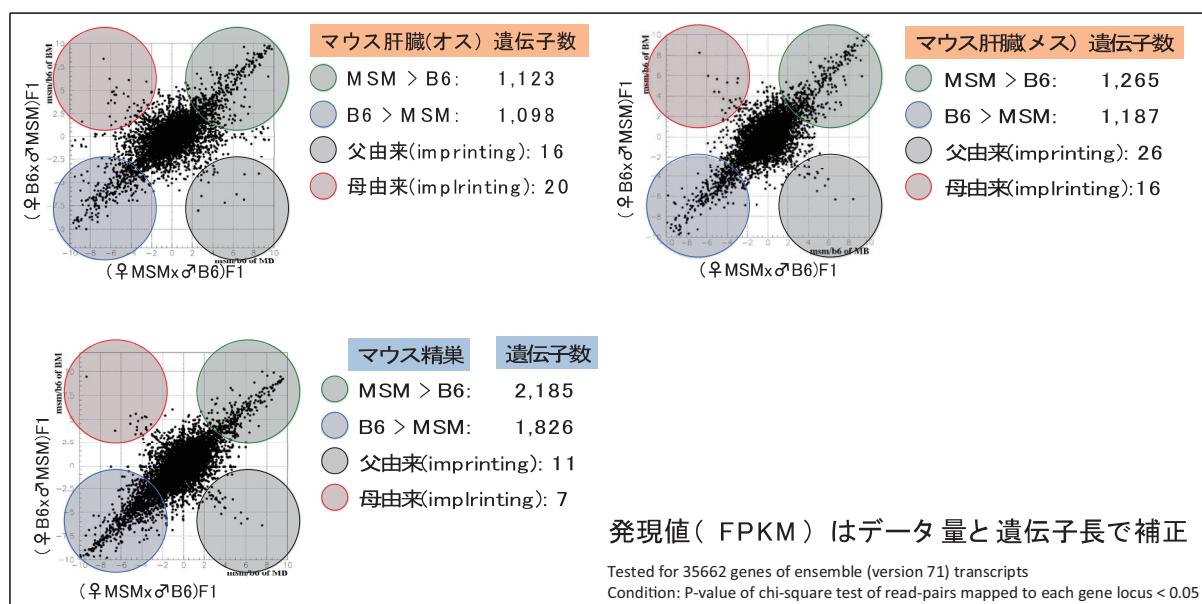


図 1 F1 個体の肝臓および精巣での遺伝子発現のアレルバイアス。non-coding 遺伝子を含む転写産物 35,662 種 (ensemble version 71)を対象にしたプロット図を示す。各プロット図について、左側に F1(B6 ♀ x MSM ♂)について、上部に MSM アレル高発現、下部に B6 アレル高発現を示す遺伝子群がプロットされる。一方、下部に F1(MSM ♀ x B6 ♂)について、左側に B6 アレル高発現、右側に MSM アレル高発現がプロットされる。各図の中心部分は F1 個体で発現差のない遺伝子群の集合である。

次に、検出した遺伝子の他の解析系による検証を行うため、アレル発現バイアスが確認された遺伝子について、TaKaRa 社製、サイクリングプローブを利用した定量 PCR による検出系 (Cyclease®PCR 法) を使用して、アレルごとの遺伝子発現を確認した。この手法では、SNP を認識した標的遺伝子の発現解析を正確かつ簡便に行うことができる。図 2 に本手法により得られた結果を示す。F1 個体の肝臓および脳での解析結果について示している。例として F1 (B6 ♀ x MSM ♂) 個体由来の組織から精製した RNA を逆転写した cDNA をサンプルとして使用して行った解析結果を示す。例として示した遺伝子

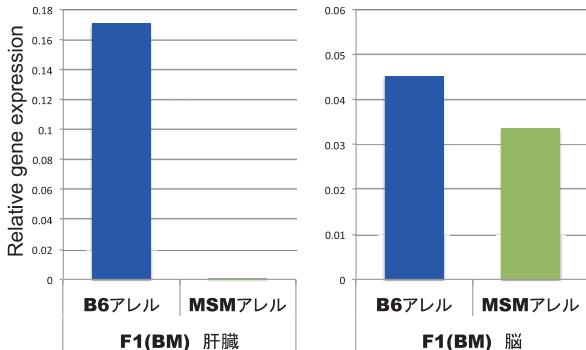


図 2 RNA-Seq による遺伝子発現量解析により、F1 の肝臓で B6-MSM アレル間の発現差が観察された遺伝子について TaKaRa 社製、サイクリングプローブを利用した肝臓由来サンプルにおける定量 PCR による解析により得られた結果を示す(左パネル)。コントロールである脳由来サンプルの解析では、B6 および MSM の両方のアレルからの発現が確認できる(右パネル)。

は、RNA-Seq 解析の結果と同様に、肝臓では F1 個体で B6 アレルの発現しか観察されない。一方、脳では B6 アレル、MSM アレル両方からの発現が確認できる。今後は、費用対効果を考えた上で、本手法を RNA-Seq による遺伝子発現量バイアス解析において抽出された複数の遺伝子群の検証に利用する。

これまでに我々が構築した、亜種間ゲノム多型を利用したプラットフォームを使用することで、B6 および MSM のアレルを区別した遺伝子発現解析を全ゲノムに渡って効率的に行なうことが可能になったが、各アレル由来の遺伝子発現量と、遺伝子発現調節に関わるヒストン修飾 (H3K4me1 など) との関係、すなわち「マウスの系統間差を生み出す遺伝子発現制御領域の多型」を調査するため、予備的な ChIP-seq 解析を開始した。まず、

MSM、B6、F1 (B6/MSM)、F1(MSM/B6) の 10 週齢の雄の肝臓について解析を開始した。組織サンプルはホルマリン等で固定してから核を回収し、−80°C にて保存した。各種抗体を使用して免疫沈降した DNA サンプルは、品質および濃度を測定後、イルミナ社製、TruSeq ChIP Sample Prep Kit より解析試料を作製した。塩基配列解読用のサンプルは、バイオアナライザーにより再度品質検査を行い、サブテーマ 1 によるデータ解析に供した。イルミナ社製、Hiseq2000 あるいは Hiseq2500 により生産された 1 次データは、サブテーマ 1 のサーバーよりデータを受け取り、国立遺伝学研究所 DDBJ スーパーコンピュータの「tricmouse」アカウントで 2 次・3 次解析を行った。結果のマッピングには、マウスの基準配列である B6 系統の mm9、mm10、我々がゲノムを公開している MSM の多型情報を使用した。なお、2 次・3 次解析を行う過程で、データの閲覧および解析手法の検討を「tricmouse」アカウントを利用してメンバー間で行った。図 3 にこれまでに得られた結果の一例 [F1(B6♀ × MSM♂) 肝臓、オス個体] を示す。

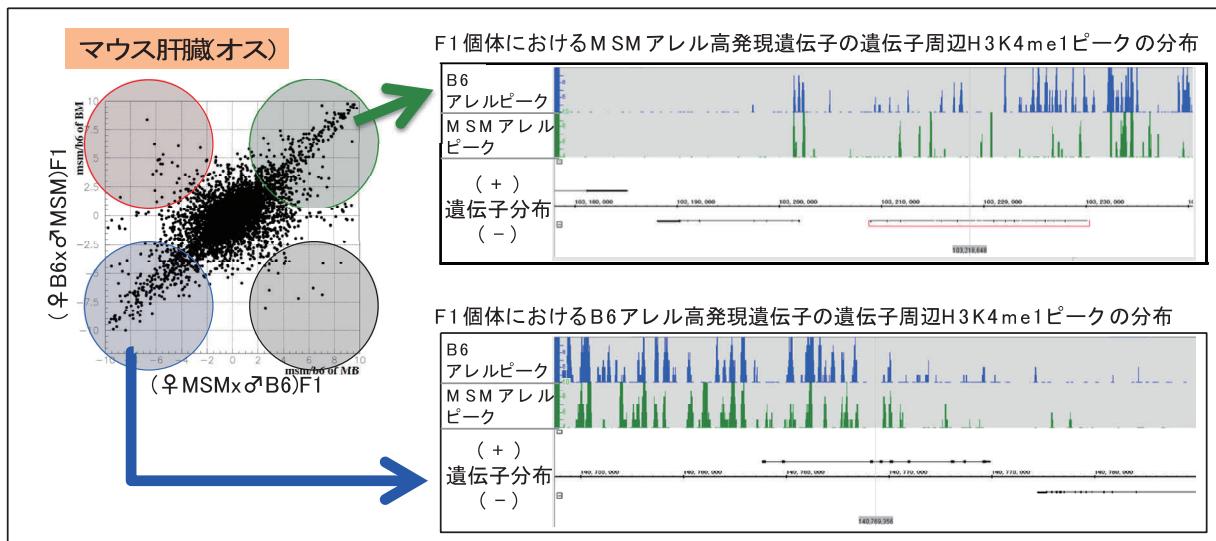


図 3 発現バイアスのある遺伝子周辺におけるメチル化ピークの分布。図 1 に示したアレル発現量バイアスを示す遺伝子から、B6 アレル高発現(右側上部パネル)、および MSM アレル高発現(右側下部パネル)を示す遺伝子(Ensembl gene version 71)を各 1 種類選択し、遺伝子領域の前後の領域を示した。左から染色体のセントロメア側、右側はテロメア側を示す。B6 アレルピークを青色、MSM アレルピークを緑色で示している。シグナル強度はピークの高さで示される。共通の座位に位置するシグナル強度の違うピークや異なる座位のピークに存在するゲノム多型が、遺伝子発現差を生み出す制御配列の候補となる。

ChIP-seq により得られたデータは B6-MSM 間の SNP を指標にしたマッピングを行い、B6、および MSM 由来アレルを効果的にマップすることができた。H3K4me1 はエンハンサーのマークとして使用されているが、図 1 に示したアレル発現量バイアスを示す遺伝子周囲の H3K4me1 にも B6 および MSM それぞれに特異的なピークが観察された。今後は、遺伝子発現データやこれらピークに関連した統計解析を行い、アレル発現量バイアスに影響をおよぼすピークの抽出を行う。ChIP-seq 解析については、他のメチル化ヒストン抗体を使用して免疫沈降したサンプルや、染色体構造の変化を検出するための各種手法により得られたサンプルへと拡張する。

今後は、これまで複数のマウス系統を用いて収集、解析してきた複合形質のデータをカタログ化し、次世代シーケンサによるデータ解析から、GO 解析やパスウェイ解析を行い、表現型差に関連した遺伝子群の発現差と、その発現調節エレメントの機能を解明する。

3-2：大規模実験交配集団を対象とした遺伝的構造の統計解析 [栗木、岡、高田、城石]

本テーマは、マウス汎用近交系統および亜種由来系統の大規模交配により F2 個体を作製し、複数の形質データを観測するとともに、シーケンス社製マスアレイを利用した高速 SNP タイピングによる遺伝子型情報を組み合わせ、形質データに影響をおよぼす遺伝的構造（主にエピスタシス）をサブテーマ 2 と連携した解析により、統計的に検出することを目的としている。これまでに、B6 および MSM を基盤とした F2 個体 3000 頭超の形質データと全ゲノムに渡る 200 カ所程度の SNP によるジェノタイピングによる遺伝子型情報量の収集が終了した。本年度は、これまでに本テーマで解析した西ヨーロッパ産由来 B6、および日本産野生由来の MSM 系統とは別亜種の台湾産 (*Mus m castaneus*) を基盤とした F2 個体の全ゲノムに渡る 200 カ所程度の SNP によるジェノタイピングを開始した。これにより、B6 および MSM 双方と遺伝距離が遠い系統も導入することができ、実験動物としての「マウス」が保持するゲノム多型を全方位から網羅したエピスタシスの検出をサブテーマ 2 と連携し試みる。

3-3：野生由来ヘテロジニアスストックによる高精度 QTL 遺伝子マッピング法の確立

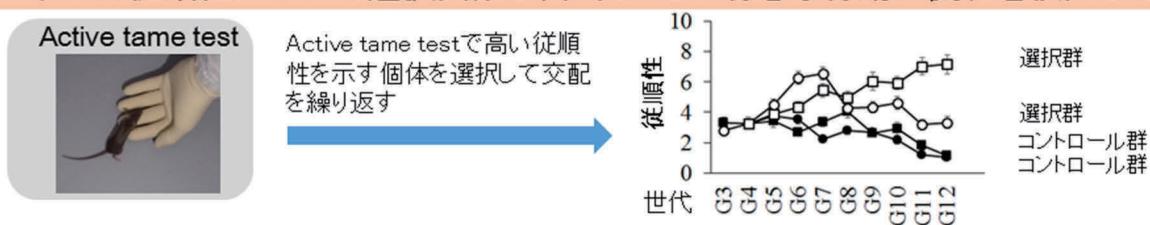
[松本、田邊、高橋、栗木、加藤、二宮、高野、小出]

性格・行動や体質など、これらわれわれの身近な形質は多因子形質であり、これまで既存の遺伝解析の方法では遺伝子同定に必要な十分な遺伝的マッピング精度や、それら多因子間の相互作用なども含めたネットワーク全体の基盤情報を得ることは極めて難しいことが知られている。その原因として、小さな効果を持った多数の遺伝子座の存在や遺伝子間相互作用などが考えられているが、その実態はまだ明らかになっていないのが実情である。近年、多数の遺伝的に異なるマウス系統を交雑したヘテロジニアス集団が、高い遺伝子座のマッピング精度を示すことが示されている。代表的なヘテロジニアス集団は、8 つの近交系統を相互に交配し、その後近交化を避けるためにランダム交配を続けることで作出されるものである。我々は、野生系統の遺伝的多様性に着目し、*Domesticus* 亜種由来の 2 系統 (PGN2, BFM/2)、*Castaneus* 由来の 1 系統 (HMI)、それから *Musculus* 亜種由来の 5 系統 (BLG2, NJL, MSM, CHD, KJR) の合計 8 種類の野生由来系統をもとに 16 ペアで一つの集団を維持することで野生由来ヘテロジニアス集団の作製を進めてきた。これら野生由来の 8 系統は、系統間で遺伝的な差が大きく、その遺伝的マーカーおよび形質の多様性を持つことが大きな特徴である。そのため、これらの野生系統から作出されるヘテロジニアス集団内では膨大な多様性が期待できる。私たちは、この集団に対して従順性にかかる形質の選択交配を行うことで、この形質に関して顕著な違いを示す集団を作製することに成功した（図）。そこで、これらの個体に関して、全ゲノムの SNPs タイピングを行い、選択集団と非選択集団を比較することによる精度の高い遺伝解析を行うことを試みている。また、この野生由来ヘテロジニアス集団は、ほかにもさまざまな多因子形質の遺伝解析に利用できる可能性を持っており、F2 集団などを用いた

QTL 解析では十分な精度の遺伝子座情報が得られない場合に、このヘテロジニアス集団を利用した解析を行うことで精度の高い遺伝解析が可能になると期待できる。現在、13 世代目を交配中である。

これまでの研究で、8 つの親系統から最初から計画的に世代をおって 16 ペアの集団として作出された本ストックが、集団内のアレル頻度の推定において非常に正確な実測値を示すことを明らかにした。昨年度は、これらの集団の部分的な個体群について全ゲノムの SNPs タイピングを行い、そのデータとともに選択群で有意にアレル頻度の増加とヘテロ接合度の減少を示す領域があることが分かった。

・マウスの従順性について選択交配し、世代ごとの有意な行動の変化を検出した



・選択群において特定のアレル頻度が増加し、ヘテロ接合度が減少した領域を同定

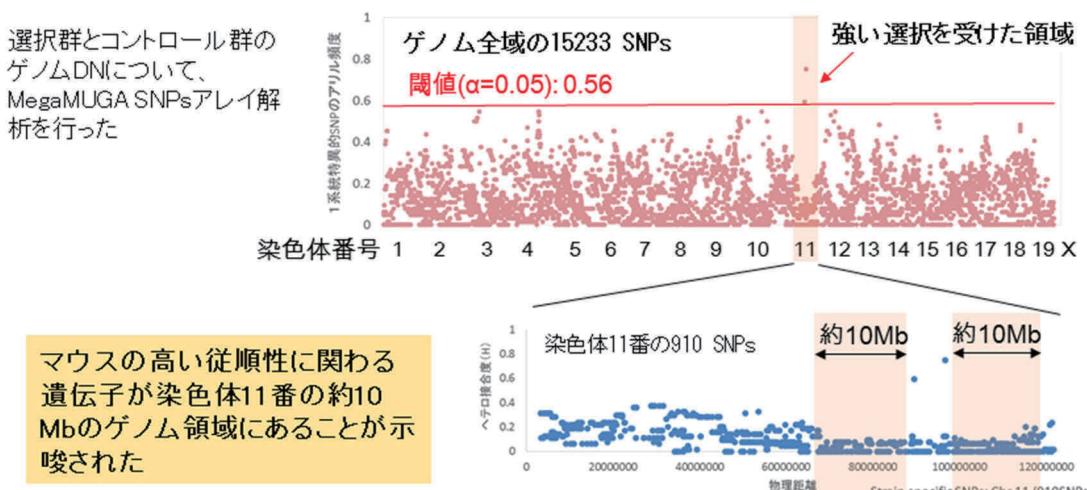


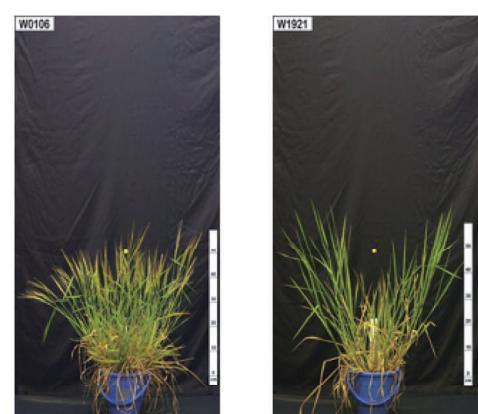
図 16 ペア(総アレル数 64)のヘテロジニアス集団における選択交配の結果と、その全ゲノム SNPs データ解析結果

3-4 : 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、栗木、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。

このように野生イネ *O. rufipogon* は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

図1 *Oryza sativa*の祖先種*O. rufipogon*
(左)一年生型のW0106 (右)多年生型のW1921。



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 の交雑後代による実験集団を育成し、これを用いた量的な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) による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。H26 年度までに、異なる F_2 個体どうしを交配した F_3 他殖系統群 333 個体から自殖系統 (RILs; recombinant inbred line) の育成を進め F_7 世代種子 200 系統を獲得した。この F_7 世代の個体のヘテロ率は 3.125%、1 モルガンあたりの観測可能な組み換え箇所は 2,775 箇所と推定される。またイネゲノムの総物理距離は 400Mb、総遺伝地図距離は約 16 モルガンとされているためこの F_7 の実験集団約 200 系統を用いた場合、総組み換え箇所 8,880 箇所で、組換え箇所間の平均物理距離は 45kbp で平均 4.5 個の遺伝子があると予想される。H27 年度はこの実験集団 200 系統を展開し、開花 1 ヶ月後の節部発根能力の形質調査と遺伝子発現量の相関解析を遂行する計画である。また課題 3－5 の研究成果より、遺伝子の発現多型は、異なる特徴を持つ 2 タイプが有る事が示された。両者を区別するには複数組織の発現解析を行う必要がある。そこで親系統 W0106 と W1921 間の複数組織の次世代シーケンサーによる発現解析準備に取り組んでいる。また同時に一年生、多年生の指標となる節部からの発根について親系統で複数のステージで発現解析を行い、eQTL 解析に適したステージを決定する。また次世代シーケンサーによって解読した親系統 W0106 と W1921、 F_1 のゲノム塩基配列多型情報の解析により平均 625bp に一箇所の割合で F_1 のヘテロ SNP サイト 593,782 箇所を抽出した。このうち遺伝子領域に属するのは 135,465 箇所であり、eQTL 解析において 8,880 箇所の組み換え箇所の判別に十分な数である。

3-5：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。しかしながらこれまでに遺伝子分子進化の観点から生物多様性研究はなされてきたものの発現進化との関連性の報告は少ない。そこでイネ亜種間の多様性を発現進化の観点からも理解するために、公的データベース GEO より Affymetrix 社 Rice Genome Array の発現データを多量に取得しイネ亜種間の発現多型の検出及びその特徴解析を試みた。

GEO より取得したジャポニカ (Nipponbare)、インディカ 3 品種 (Minghui63, Zhenshan97, 9311) 合わせて 321 アレイデータ (121 組織分) から類似アレイデータを推定するため、予めジャポニカ・インディカ間で発現多型を示さない遺伝子群を用い全 321 アレイデータのクラスタリング解析を遂行し、Nipponbare とインディカ 2 品種 (Minghui63, Zhenshan97) のアレイデータが同一クラスターに属する 5 組織 (胚乳、成熟茎、幼穂、幼葉、根) を取得した。5 組織それぞれについて Nipponbare とインディカ 2 品種 (Minghui63, Zhenshan97) 間、およびインディカ亜種内 (Minghui63 と Zhenshan97) の発現多型を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 (SNEP 法) を用い推定した。加えて Nipponbare の 44 組織、Minghui63 の 36 組織、Zhenshan97 の 39 組織の発現データを用い発現多型遺伝子群を、一方の亜種ではいずれかの組織で高発現を確認できるがもう一方の亜種ではすべての組織で発現が見られない “サイレント型” と、いずれの亜種においても様々な組織で発現が見られる “非サイレント型” に分類した。サイレント型発現多型遺伝子と非サイレント型発現多型遺伝子については次の特性：組織ごとの検出数、遺伝子発現の組織特異性指標、遺伝子の構造差、分子進化速度、ゲノム上の分布、遺伝子の機能、遺伝子制御機構を調査し、それぞれの異なる特徴を明らかにした。(表 1)

表 1 サイレント型 /DE 遺伝子とノンサイレント型遺伝子の特性

	サイレント型発現多型遺伝子	非サイレント型発現多型遺伝子
検出数	<ul style="list-style-type: none"> 各組織でほぼ一定。 ・ジャポニカでサイレンシングされている遺伝子数はインディカより少ない。 	<ul style="list-style-type: none"> 組織ごとに異なる。 ・ジャポニカとインディカの同一組織で高発現する非サイレント型遺伝子の数は同程度。
組織発現特異性との関係	・無関係。	・組織特異的に発現する遺伝子で高頻度。
遺伝子構造差	・大きく構造が変異。	・発現多型のない遺伝子と同程度。
遺伝子分子進化	・制約から解放。	・発現多型のない遺伝子と同程度。
ゲノム上の分布	・クラスター化して分布するものもある。	・ほぼ一様に分布。
トランスポゾン関連遺伝子	・濃縮されている。	・発現多型のない遺伝子と同程度。
遺伝子オントロジー(GO)エンリッチメント解析	<ul style="list-style-type: none"> ・"アボトーシス", "タンパク質リン酸化", "防御反応", など。 ・ジャポニカでサイレンスされる遺伝子群とインディカでサイレンスされる遺伝子群のGO termは同じ。 	<ul style="list-style-type: none"> ・"ストレス応答", "低温応答", "過酸化水素に対する応答", など。 ・ジャポニカでより高発現する遺伝子群とインディカでより高発現する遺伝子群ではエンリッチされるGO termは異なる。
eQTLによる制御	主に <i>cis</i> -eQTL	<i>cis</i> -と <i>trans</i> -eQTLの両者

indica 2 亜種 (Minghui63 と Zhenshan97) についてはこれらの交雑後代 RIL を用いた eQTL 解析結果が報告されている。親品種間の発現多型遺伝子をサイレント型と非サイレント型に分け、eQTL 解析の結果と照合すると、サイレント型が主に *cis*-eQTL 制御を受けているのに対し非サイレント型は *cis*、*trans*-eQTL の両者による制御を受けていることが明らかとなった (表 1)。

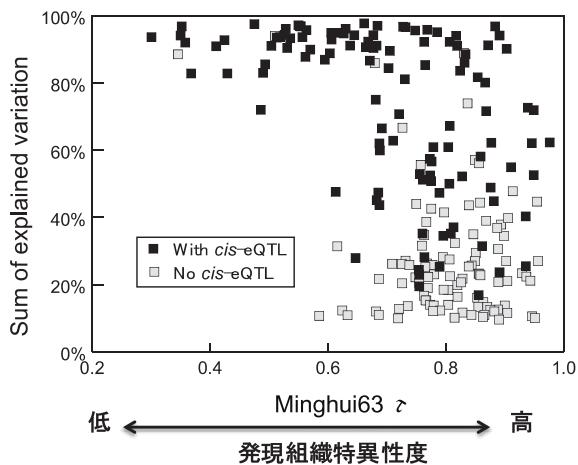


図1 非サイレント発現多型遺伝子の発現組織特異性度と発現制御eQTLの関係 芽生え初期におけるインディカ亜種間非サイレント型遺伝子多型遺伝子のeQTL説明変数の合計(Wang et al., 2010)を Minghui63における発現組織特異性度に対しプロットした。

また 211 個の非サイレント型発現多型遺伝子については、遺伝子発現の組織特異性度と、検出された eQTL で集団内の遺伝子変動がどの程度説明出来るかについて、*cis*-eQTL が検出された遺伝子とされなかった遺伝子を分け図 1 で比較した。組織特異性度の高い遺伝子の多くは *cis*-eQTL による制御を受ける傾向が少なく、また検出された eQTL ではその集団内の遺伝子変動を説明しきれていない。つまり、発現組織特異性度の高い非サイレント型発現多型は遺伝子自身または近傍の配列の差によって発現量に違いが生じるのではなく、複数の *trans*-eQTL による複雑な機構により発現量の差を生じた。また、発現組織特異性度の高い遺伝子の発現進化が急であることから、遺伝子発現に正の自然淘汰が働いていることが示唆される。本結果は論文投稿中である。

3-6 : 多系統野生イネゲノム構造と形質相関解析 (GWAS 解析) による特性寄与遺伝子の検索

[倉田、久保、岩田、藤澤、栗木、上海植物科学研究所]

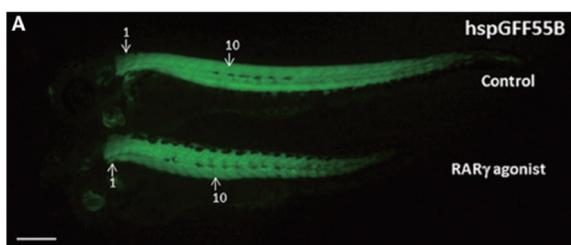
2012 年度に、アジア各地から収集した野生イネ *O. rufipogon* 446 系統と栽培イネ *O. sativa* 1083 系統 (*japonica*, *indica* を含む) をゲノム解析し、ゲノム全長の構造変異のパターンを用いて、1529 系統間の相互関係を明らかにした。この野生イネ集団を用いて、遺伝型の多型と形質変異の相関の有無を全ゲノムに渡り検索する、GWAS (genome wide association study) 解析を行うため、2012 年から 2014

年度までの3年間をかけて、各年度15種類以上の形態形質や生態型の観測および大規模データの収集を行い、特性を支配する遺伝子の検出を試みている。2014年度は、2013年度の形質データの整理が終了し、2年度分のデータの利用が可能になった。これらのデータをさまざまな角度から検討し、2年分のデータの比較検討を行っている。これまでの解析では、446系統はゲノム構造から大きく3つの遺伝的集団に分かれるが、集団毎の予備的GWAS解析、全系統で行なった結果、さらに年次間の解析結果で、ゲノム上の形質と相関する遺伝子領域のピークが異なる場合が多く見られた。データのバラツキや、形質毎の特性の精査が必要と考えられ、順次進めてデータの再整備を行なった。平行して2014年度には、3年度目の形質情報の大量収集(20形質×約1200個体)を行なっており、今後3年間の調査結果を検討しつつ、GWASの方法改良と合わせて、原因遺伝子の特定を進める。GWAS解析法に関する新たな改良法の提案、実際のデータ適用場面でのさらなるアイデアについて、統計数理研、東大グループと検討を進めている。

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 [川上、武藤、北本、栗木、小山]

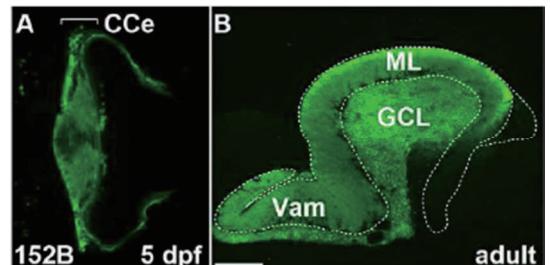
トランスポゾンのランダムな挿入に基づくトランスジェニック系統、ノックアウト変異系統の網羅的な開発とそれらに基づく多様な表現型の抽出とその表出法の確立のための研究を行った。神経活動パターンの画像解析、数理解析等の研究を進めた。下図に概略を示す。

- (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1300系統以上とした。これらを基にして、国内外の器官形成研究者と共同研究を展開することにより研究の融合をはかった。千葉大学伊藤博士との共同研究では、脊髄の神経細胞分化におけるNotchシグナルの役割を明らかにした(Okigawa et al., Dev Biol 2014)。国立遺伝研融合研究員和田博士は、ゼブラフィッシュ側線鱗の骨リモデリングによる形成メカニズムを明らかにした(Wada et al., Dev Biol 2014)。



英国アストン大学 Johnson 博士との共同研究でレチノイン酸受容体 γ の初期発生、ヒレ再生における役割を明らかにした(Wai et al., Stem Cells Dev 2014) (左図:レチノイン酸受容体 γ アゴニストの体節形成における役割)。名古屋大学日比博士との共同研究でトランスジェニックフィッシュの小脳におけるGal4発現の詳細な解析

を行なった(Takeuchi et al., 2014) (右図:小脳の顆粒細胞の可視化)。国立遺伝学研究所平田博士との共同研究でナトリウムチャンネルの品質管理が神経細胞の活動電位発生のために重要であることを明らかにした(Ogino et al., 2014)。



- (2) これまで新しい表現型抽出法としてゼブラフィッシュの神経細胞の活動を、カルシウムインディケーター-GCaMPを用いてイメージングするシステムの開発に成功してきた(Muto et al., PNAS, 2011、Muto et al. Curr. Biol. 2013)。この方法を用いて、脳神経活動の長時間イメージングを行い、得られた神経細胞の発火パターンの数理学的解析を国立情報研北本研究室と進めつつある。

3-8：ショウジョウバエ翅形態異常のゲノム相関解析 [上田、近藤、相垣、内藤]

これまで RNAi 変異体等を用いて、8,070 遺伝子に対応する翅異常表現型 26 万枚を記録した。しかしながら RNAi 変異体ではターゲット遺伝子のノックダウンが効率よく行われるという保証はない。私達の用いている変異体系統では RNAi ベクターがトランスポゾンによって染色体にランダムに挿入されており、その挿入位置によって RNAi 効果を産み出す二重鎖 RNA の発現量が変化すること、また当該遺伝子の発現する細胞によって GAL4-UAS 発現誘導システムの効率に差があることなど、原理的な問題である。これを克服するために、最近開発された高効率 RNAi ベクター pValium20 を用いて新たに第二世代ライブラリを作成することにした。大規模なトランスジェニック・ライブラリを短期間で作出する方法を考案し、2 年間で約 3,000 の新しい系統を作出することに成功した。これらの系統は平成 27 年 1 月より遺伝研・ショウジョウバエストックセンター (NIG-Fly) より送料等実費で国内外に提供されている。現在ハーバード大を中心とした国際的な共同研究により、全ゲノムをカバーする RNAi 表現形データベースの構築が進行中である。

ショウジョウバエにおける RNAi 実験は簡便ではあるが、上記の偽陰性の可能性と、さらにオフターゲット効果による偽陽性が大きな問題である。したがって、最終的に遺伝子の機能を確定するためには、ヌル変異体の観察が必須である。昨年度は生殖細胞特異的に人工スクレアーゼ CRISPR/Cas9 を発現するトランスジェニック系統を開発し、これを用いて高効率で変異体を得る方法を確立した。平成 26 年度は約 200 の遺伝子についてヌル変異体系統を作出した。また全遺伝子の変異体作成を目標として、DBCLS 内藤の協力を得、開発された最適 guide RNA の自動設計法のもとに大規模な変異体作成を行っている。さらに今年度は GFP や GAL4 などのマーカー遺伝子をハイスループットでノックインする系も構築した。この系を用いて、特定の組織・細胞をラベルする GAL4 系統を 100 系統以上作出した。現在、同技術の提供を軸とした共同研究を、所外の研究者 16 名（国外 11、国内 5）と行っている。

[5] 研究成果物

サブテーマ 1

① 知見・成果物・知的財産権等

1. オルソログ遺伝子のアミノ酸配列比較プログラム 「OrthologAln」

<http://tga.nig.ac.jp/dnapod/OrthologAln>

② 成果発表等

<論文発表>

[学術論文]

1. Mochizuki, T., Tanizawa, Y., Fujisawa, T., Nikoh, N., Shimizu, T., Toyoda, A., Fujiyama, A., Kurata, N., Nagasaki, H., Kaminuma, E., Nakamura, Y. DNApod: DNA polymorphism annotation database from next-generation sequence read archives. *BMC Genomics*, in preparation.
2. Shenton, M.R., Ohyanagi, H., Wang, Z.X., Toyoda, A., Fujiyama, A., Nagata, T., Feng, Q., Han, B., Kurata, N. Rapid turnover of antimicrobial-type cysteine-rich protein genes in closely related *Oryza* genomes. *Mol Genet Genomics*. 2015 Apr 5. [Epub ahead of print] PMID:25842177
3. Arimura, Y., Shirayama, K., Horikoshi, N., Fujita, R., Taguchi, H., Kagawa, W., Fukagawa, T., Almouzni, G., Kurumizaka, H. Crystal structure and stable property of the cancer-associated heterotypic nucleosome containing CENP-A and H3.3. *Sci Rep* 4: 7115, 2014.
4. Fukagawa, T., Earnshaw, W.C. Neocentromeres.. *Curr. Biol.* 24:R946-947, 2014.
5. Kagawa, N., Hori, T., Hoki, Y., Hosoya, O., Tsutsui, K., Saga, Y., Sado, T., Fukagawa, T. The

- CENP-O complex requirement varies among different cell types. *Chr.Res* 22: 293-303, 2014.
6. Fukagawa, T., Earnshaw, WC. The centromere: chromatin foundation for the kinetochore machinery. *Dev. Cell* 30: 496-508, 2014.
 7. Hori, T., Shang, WH., Toyoda, A., Misu, S., Monma, N., Ikeo, K., Molina, O., Vargiu, G., Fujiyama, A., Kimura, H., Earnshaw, WC., Fukagawa, T. Histone H4 Lys 20 monomethylation of the CENP-A nucleosome is essential for kinetochore assembly. *Dev. Cell* 29: 740-749, 2014.
 8. Ito, T., Tarutani, Y., To, TK., Kassam, M., Duvernois-Berthet, E., Cortijo, S., Takashima, K., Saze, H., Toyoda, A., Fujiyama, A., Colot, V., Kakutani, T. Genome-wide negative feedback drives transgenerational DNA methylation dynamics in *Arabidopsis*. *PLOS Genetics*, in press.

[データベース]

1. DNApod: DNA polymorphism annotation database (DB 更新)

[著書等]

1. 長崎英樹, 望月孝子, 谷沢靖洋, 神沼英里, 中村保一: VI-4 DDBJ Read Annotation Pipeline 解析パインによる RNA-Seq de novo assembly とイネ多型解析. 羊土社 pp. 352-360, 2014.

<会議発表等>

[招待講演]

国際

1. Fukagawa, T. Centromere specific chromatin structure. IIAS symposium. 2014 May 12, Kyoto, Japan.
2. Fukagawa, T. Centromere specific chromatin structure. Gordon Conference. 2014 July 27, Boston, USA.
3. Fukagawa, T. Centromere specification and kinetochore assembly. 3R Symposium. 2014 Nov 17, Shizuoka, Japan.

国内

1. 神沼英里、望月孝子、長崎英樹、谷沢靖洋、小笠原理、大久保公策、高木利久、中村保一: 遺伝研スパン概要と、NGS アーカイブデータを用いた SNP 解析. ワークショップ「NGS 使い倒し講座 - Breeding Informatics 研究 XIII」日本育種学会第 126 回講演会. 2014.9.26, 宮崎.
2. 望月 孝子: NGS 公共データから構築した DNA 多型注釈データベースと解析ワークフローの紹介. データベース講習会. 第 56 回日本植物生理学会年会. 2015.3.16, 東京.
3. 深川竜郎 : ゲノム安定性を保障するセントロメアの構造基盤. 日本癌学会. 2014.9.25-27, 横浜.

[一般講演]

国際

1. Mochizuki, T., Tanizawa, Y., Fujisawa, T., Nikoh, N., Toyoda, A., Fujiyama, A., Kurata, N., Nagasaki, H., Shimizu, T., Kaminuma, E., Nakamura, Y. DNA Polymorphism Database from New-Generation Sequence Read Archive and Analytical Workflow. Plant & Animal Genome XXIII. 2015 January 14, San Diego, USA.
2. 角谷徹仁. DNA methylsation and TE dynamics. Epigenetic regulation of organismal function and

response to the environment. 2015.May 20, Philadelphia, USA.

国内

1. 角谷徹仁：エピジェネティクスと遺伝学. 第 8 回日本エピジェネティクス研究会年会. 2014.5.27, 東京.
2. 角谷徹仁：Epigenetic silencing and anti-silencing of TEs in Arabidopsis. 日本進化学会第 16 回大阪大会. 2014.8.22, 大阪.
3. 角谷徹仁：クロマチン長期記憶. 日本植物学会第 78 回大会シンポジウム「環境刺激に応答した植物の情報伝達と成長制御」. 2014.9.12, 川崎.
4. 角谷徹仁：DNA メチル化とトランスポゾン動態. 日本植物学会第 78 回大会シンポジウム「細胞遺伝学の伝統と新たな地平」. 2014.9.13, 川崎.
5. 保坂碧、稻垣宗一、佐瀬英俊、角谷徹仁：シロイヌナズナのヒストン H3K9 脱メチル化酵素 IBM1(increase in Bonsai Methylation1)と転写機構の相互作用. 日本遺伝学会第 86 回大会. 2014.9.17, 長浜.
6. 伊藤佑、樽谷芳明、佐瀬英俊、角谷徹仁：シロイヌナズナの ddm1 (decrease in DNA methylation1) 変異は自殖の過程において双方向の DNA メチル化変化を引き起こす. 日本遺伝学会第 86 回大会. 2014.9.17, 長浜.
7. 角谷徹仁: Epigenetic silencing and anti-silencing of TEs in Arabidopsis. 第 38 回内藤コンファレンス. 2014.10.8, 札幌.
8. 角谷徹仁：トランスポゾンの抑制と抗抑制. 第 37 回日本分子生物学会. 2014.11.26, 横浜.

[ポスター]

国内

1. 望月孝子, 谷澤靖洋, 藤澤貴智, 長崎英樹, 神沼英里, 清水徳朗, 豊田敦, 藤山秋佐夫, 倉田のり, 二河成男, 中村保一. DNApod : NGS アーカイブ配列からの統合 DNA 多型注釈データベース 生物横断的解析への活用. 第 37 回日本分子生物学会年会. 2014.11.25, 横浜.

<受賞>

なし

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

1. RefGS プログラム : <http://cran.r-project.org/web/packages/spcr/index.html>

② 成果発表等

<論文発表>

[学術論文]

1. Arakawa, T., Tanave, A., Ikeuchi, S., Takahashi, A., Kakihara, S., Kimura, S., Sugimoto, H., Asada, N., Shiroishi, T., Tomihara, K., Tsuchiya, T., Koide, T. A male-specific QTL for social interaction behavior in mice mapped with automated pattern detection by a hidden Markov model incorporated into newly developed freeware. *J Neurosci Methods* 234: 127-134, 2014.
2. Kato, S., Ishii, A., Nishi, A., Kuriki, S., Koide, T. Segregation of a QTL cluster for home-cage activity

- using a new mapping method based on regression analysis of congenic mouse strains. *Heredity* 113: 416-423, 2014.
3. Kawano, S., Fujisawa, H., Takada, T., Shroishi, T. Sparse principal component regression with adaptive loading. *Computational Statistics & Data Analysis*, in press.
 4. Dou, X., Kuriki, S., Lin, G.D., Richards, D. EM algorithms for estimating the Bernstein copula. *Computational Statistics & Data Analysis*, in press.
 5. Dou, X., Shirahata, S., Sugimoto, H., Koide, T. Noise reduction and classification of mouse ultrasonic vocalization data. ASTE Special Issue on the “Financial & Pension Mathematical Science”, 12: 71-79 (March 2015).

<会議発表等>

[招待講演]

国際

1. Dou, X., Shirahata, S., Sugimoto, H., Koide, T. Functional Clustering of Mouse Ultrasonic Vocalization Data. Workshop on Statistical Methods for Large Complex Data, National Sun Yat-sen University, Kaohsiung, Taiwan, March, 2015.

国内

1. 二宮嘉行：疑似相関を用いた多重性調整およびその応用. 応用統計学会. 2014.5.22, 東京.

[一般講演]

国際

1. Kawano, S., Fujisawa, H. Sparse principal component regression for simultaneous dimension reduction and variable selection. IMS-APRM. June 2014, Taiwan.
2. Dou, X., Kuriki, S., Lin, G., Richards, D. A class of B-spline copulas: dependence structure and estimation. 7th International Conference of the ERCIM WG on Computational and Methodological Statistics (ERCIM 2014). 2014 December 6, University of Pisa, Italy.

国内

1. 荒川俊也, 遠藤なつ美, 田中知己：隠れマルコフモデルを用いたヤギの行動自動推定の試み. 研究集会「最適化：モデリングとアルゴリズム 28」. 2015.3.19-20, 立川.
2. 遠藤なつ美, 荒川俊也, 田中知己：ヤギの行動解析に基づく発情の自動推定法の検討. 日本畜産学会第119回大会. 2015.3.26-30, 宇都宮.
3. 川野秀一, 藤澤洋徳, 高田豊行, 城石俊彦：スパース学習による主成分回帰モデル. 情報論的学習理論ワークショップ. 2014.11, 名古屋.
4. 川野秀一, 藤澤洋徳, 高田豊行, 城石俊彦：スパース学習による1段階主成分回帰モデル. 統計的推測法における最近の展開. 2014.12.14, 山口.
5. 川野秀一, 藤澤洋徳, 高田豊行, 城石俊彦：スパース正則化に基づく主成分回帰モデリング. 統計関連学会連合大会. 2014.9, 東京.
6. 豊田淳, 後藤達彦, 荒川俊也, 小針大助, 岡山毅：赤外線深度センサを利用した動物、植物の動きの解析. 第87回日本生化学会大会. 2014.10.15-18, 京都.

<受賞>
なし

③ その他の成果発表

1. 新聞記事「マウスの行動自動解析 三島の遺伝研・小出准教授らソフト開発」静岡新聞, 2014年5月23日
2. 新聞記事「マウス観察 早・安・楽に 愛知工科大など行動解析ソフト開発」中日新聞, 2014年5月28日
3. 新聞記事「デュオマウス開発話題に 実験用マウスの行動を自動解析／愛知工科大 荒川氏らが開発」東日新聞, 2014年6月1日
4. 第6回幸田プレステージレクチャーズにおける Duomouse の展示, 2015年1月17日, 愛知県幸田町民会館
5. 未来展2015における Duomouse の展示, 2015年3月25-26日, 愛知県名古屋市吹上ホール

サブテーマ3

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. Kawano, S., Fujisawa, H., Takada, T., Shiroishi, T. Sparse principal component regression with adaptive loading. *Computational Statistics & Data Analysis*, 2015, doi: 10.1016/j.csda.2015.03.016.
2. Oka, A., Shiroishi, T. Regulatory divergence of X-linked genes and hybrid male sterility in mice. *Genes Genet Syst* 89: 99-108, 2014.
3. Tamura, M., Amano, T., Shiroishi, T. The Hand2 gene dosage effect in developmental defects and human congenital disorders: *Curr Top Dev Biol* 110: 129-152, 2014.
4. Oka, A., Takada, T., Fujisawa, H., Shiroishi, T. Evolutionarily diverged regulation of X-chromosomal genes as a primal event in mouse reproductive isolation. *PLoS Genet* 10:e1004301, 2014.
5. Dou, X., Kuriki, S., Maeno, A., Takada, T., Shiroishi, T. Influence analysis in quantitative trait loci detection. *Biom J* 56: 697-719, 2014.
6. Takahashi, A., Shiroishi, T., Koide, T. Genetic mapping of escalated aggression in wild-derived mouse strain MSM/Ms: association with serotonin-related genes. *Frontiers in Neuroscience* Volume8, Article156, 2014.
7. Kato, S., Ishii, A., Nishi, A., Kuriki, S., Koide, T. Segregation of a QTL cluster for home-cage activity using a new mapping method based on regression analysis of congenic mouse strains. *Heredity* 113: 416-423, 2014.
8. Arakawa, T., Tanave, A., Ikeuchi, S., Takahashi, A., Kakihara, S., Kimura, S., Sugimoto, H., Asada, N., Shiroishi, T., Tomihara, K., Tsuchiya, T., Koide, T. A male-specific QTL for social interaction behavior in mice mapped with automated pattern detection by a hidden Markov model incorporated into newly developed freeware. *Journal of Neuroscience Methods* 234: 127-134, 2014.
9. Takahashi, A., Nagayasu, K., Nishitani, N., Kaneko, S., Koide, T. Control of intermale aggression by the medial prefrontal cortex activation in the mouse. *PLoS ONE* 9(4):e94657, 2014.

10. Kanno, K., Kokubo, H., Takahashi, A., Koide, T. Ishiura S. Enhanced prepulse inhibition and low sensitivity to a dopamine agonist in Hesr1-knockout mice. *Journal of Neuroscience Research* 92: 287-297, 2014.
11. Kuriki, S., Harushima, Y., Fujisawa, H., Kurata, N. Approximate tail probabilities of the maximum of a chi-square field on multi-dimensional lattice points and their applications to detection of loci interactions. *Annals of the Institute of Statistical Mathematics*. 66: 725-757. 2014.
12. Tsuda, K., Kurata, N., Ohyanagi, H., Hake, S. Genome-wide study of KNOX regulatory network reveals brassinosteroid catabolic genes important for shoot meristem function in rice. *Plant Cell*, 26: 3488-3500, 2014. doi: 10.1105/tpc.114.129122.
13. Yamamoto, A., Yoshii, M., Murase, S., Fujita, M., Kurata, N., Hobo, T., Kagaya, Y., Takeda, S., Hattori, T. Cell-by-cell developmental transition from embryo to postgermination phase revealed by heterochronic gene expression and ER-body formation in *Arabidopsis leafy cotyledon* mutants. *Plant Cell Physiol*. 55: 2112-2125, 2014
14. Shenton, M., Ohyanagi, H., Wang, Z-X., Toyoda, A., Fujiyama, A., Nagata, T., Feng, Q., Han, B., Kurata, N. Rapid turnover of antimicrobial-type cysteine-rich protein genes in closely related *Oryza* genomes. *Mol Genet Genomics*. Published on line 05 April. 2015. DOI 10.1007/s00438-015-1028-4
15. Okigawa, S., Mizoguchi, T., Okano, M., Tanaka, H., Isoda, M., Jiang, Y.J., Suster, M., Higashijima, S.-I., Kawakami, K., Itoh, M. Different combinations of Notch ligands and receptors regulate V2 interneuron progenitor proliferation and V2a/V2b cell fate determination. *Developmental Biology* 391: 196-206, 2014.
16. Wada, H., Iwasaki, M., Kawakami, K. Development of the lateral line canal system through a bone remodeling process in zebrafish. *Developmental Biology* 392: 1-14, 2014.
17. Wai, H.A., Wada, H., Kawakami, K., Müller, F., Vernallis, A.B., Brown, G., Johnson, W.E. The development and growth of tissues derived from cranial neural crest and primitive mesoderm is dependent on the ligation status of retinoic acid receptor (RAR) γ : evidence that RAR γ functions to maintain stem/progenitor cells in the absence of retinoic acid. *Stem Cells and Development* 24: 507-519, 2015.
18. Takeuchi, M., Matsuda, K., Yamaguchi, S., Asakawa, K., Miyasaka, N., Lal, P., Yoshihara, Y., Koga, A., Kawakami, K., Shimizu, T., Hibi, M. Establishment of Gal4 transgenic zebrafish lines for analysis of development of cerebellar neural circuitry. *Developmental Biology* 397: 1-17, 2015.
19. Ogino, K., Low, S.E., Yamada, K., Saint-Amant, L., Zhou, W., Muto, A., Asakawa, K., Nakai, J., Kawakami, K., Kuwada, J.Y., Hirata, H. RING finger protein 121 facilitates the degradation and membrane localization of voltage-gated sodium channels. *Proc. Natl. Acad. Sci. USA* 112: 2859-2864, 2015.
20. Kondo, S. New horizons in genome engineering of *Drosophila melanogaster*. *Genes Genet Syst*. 89: 3-8, 2014.

[解説・総説]

1. 高田豊行、城石俊彦：日本産マウス系統 MSM と JF1 のゲノム解読からみえてきた実験用マウスの起源。医学のあゆみ 248:171-173, 2014.
2. 高田豊行：マウスのゲノムに遺る江戸時代の愛玩ねずみ。遺伝 68:7-10, 2014.

<会議発表等>

[招待講演]

国際

1. Kawakami, K. The Gal4 driver transgenic resource for developmental biology and neuroscience. ZDM7: Zebrafish Disease Models Conference. 2014.6.28-7.1, Madison, Wisconsin.
2. Kawakami, K. The *Tol2* transposon technology in vertebrates. 12th Transgenic Technology Meeting by International Society for Transgenic Technologies. 2014.10.6-8, Edinburgh.
3. Kawakami, K. The *Tol2* transposon technology and its applications in zebrafish. 12th Transgenic Technology Meeting: Zebrafish workshop 2014. 2014.10.9, Edinburgh.
4. Kawakami, K. Genetic dissection of the zebrafish brain by the Gal4-UAS approach. Cold Spring Harbor Asia Meeting: Genetics, Genomics & Phenomics of Fish. 2014.10.20-24, Suzhou.
5. Kawakami, K. Genetic dissection of the zebrafish brain by the Gal4-UAS approach. The 3rd Imaging Structure & Function in the Zebrafish Brain Symposium. 2014.12.4-17, Paris.

国内

1. 岡彩子：遺伝子発現制御の進化的分化と生殖隔離. 国立遺伝学研究所研究会 「表現型と遺伝子型のミックシングリンクをつなぐ - ゲノムダイナミクスによる遺伝子発現の質的・量的制御 -」 . 2014.6.27, 三島.

[一般講演]

国際

1. Mochizuki, T., Tanizawa, Y., Fujisawa, T., Nikoh, N., Toyoda, A., Fujiyama, A., Kurata, N., Nagasaki, H., Shimizu, T., Kaminuma, E., Nakamura, Y. DNA Polymorphism Database from New-Generation Sequence Read Archive and Analytical Workflow. Plant & Animal Genome XXIII. 2015 January 14, San Diego, USA.

国内

1. 高田豊行、近藤伸二、阿部貴志、清澤秀孔、豊田敦、藤山秋佐夫、城石俊彦：マウス野生由来系統のゲノム情報を用いたアレル発現量解析. 日本遺伝学会第 86 回大会. 2014.9.17-19, 長浜.
2. 殿崎薰、関根大輔、大西孝幸、吉海弘康、倉田のり、木下哲：倍数性操作によるイネ胚乳の生殖的隔離の打破. 日本育種学会第 127 回講演会. 2015.3.22, 町田.

[ポスター]

国際

1. Horiuchi, Y., Harushima, Y., Fujisawa, H., Mochizuki, T., Fujita, M., Ohyanagi, H., Kurata, N. Two contrasting types of differentially expressed genes between japonica and indica cultivars. 12th International Symposium on Rice Functional Genomics. 2014.11.16-19, Tucson, USA.
2. Ohyanagi, H., Kobayashi, M., Fujita, M., Shenton, M., Kubo, T., Toyoda, A., Fujiyama, A., Yamazaki, Y., Goicoechea, JL., Wang, RA., Yano, K., Kurata, N. Deciphering the *Oryza officinalis* genome sequence and construction of genomic information infrastructure for *Oryza* wild accessions. 12th International Symposium on Rice Functional Genomics. 2014.11.16-19, Tucson, USA.
3. Kitazumi, A., Ohyanagi, H., Fujita, M., Toyoda, A., Fujiyama, A., Kurata, N., De los Reyes BG. Comparing the compositional complexities of the low temperature stress transcriptomes of AA and

CC genomes of Oryza. 12th International Symposium on Rice Functional Genomics. 2014.11.16-19, Tucson, USA.

4. Ohyanagi, H., Ebata, T., Huang X., Gong, H., Fujita, M., Toyoda, A., Fujiyama, A., Wang ZX., Han, B., Yamazaki, Y., Kurata, N. Consideration of O. Ruffipogon Diversity with a SNP Database. Plant & Animal Genome XXIII. 2015.1.12, San Diego, USA.

国内

1. 近藤伸二、加藤英政、鈴木穣、高田豊行、城石俊彦、菅沼成文、清澤秀孔 : in vitro 神経細胞分化系を用いた網羅的な組織特異的ゲノム刷り込み領域の解析. 第 37 回日本分子生物学会年会. 2014.11.25-27, 横浜.
2. Toyoyuki Takada: NIG Mouse Functional Genomics Database. モロシヌス研究会. 2014. 6.27-28, 修善寺.
3. Takanori Amano: A novel regulatory element for Shh expression in the lung and gut of mouse embryos. モロシヌス研究会. 2014.6.27-28, 修善寺.
4. 松本悠貴、後藤達彦、中岡博史、田邊彰、井ノ上逸朗、Richard Mott、小出剛 : 野生由来ヘテロジニアスストックマウスを用いた従順性行動に関する遺伝子座の探索. 第 61 回日本実験動物学会総会. 2014.5.15-17, 札幌.
5. 今井悠二、松本悠貴、田邊彰、小出剛、半澤直人 : 野生由来ヘテロジニアスストックマウスを用いて明らかにするジアゼパム抗不安効果の遺伝的多様性. 第 61 回日本実験動物学会総会. 2014.5.15-17, 札幌.
6. 田邊彰、石井亜矢子、高橋阿貴、小出剛 : 日本産野生由来マウス系統 MSM の行動遺伝学が明らかにする実験用マウス PACAP 遺伝子の進化. 第 61 回日本実験動物学会総会. 2014.5.15-17, 札幌.
7. 平田晴菜、高橋阿貴、渡邊和忠、小出剛、霜田靖 : 発達過程の大脳基底核に発現する Caspr3 欠損が行動に及ぼす影響. 第 37 回日本神経科学大会. 2014.9.11-13, 横浜.
8. 今井悠二、松本悠貴、田邊彰、半澤直人、小出剛 : ジアゼパム抗不安効果の多様性解明に向けた野生由来ヘテロジニアスストックマウスを用いた解析. 第 37 回日本神経科学大会. 2014.9.11-13, 横浜.
9. 高橋阿貴、小出剛 : 背側縫線核の興奮性入力が雄マウスの攻撃行動の強度を決定する. 第 37 回日本神経科学大会. 2014.9.11-13, 横浜.
10. 田邊彰、高橋阿貴、小出剛 : 野生由来マウスの行動のストレス応答に関するゲノムマッピングと遺伝子解析. 第 37 回日本神経科学大会. 2014.9.11-13, 横浜.
11. 小出剛、石井亜矢子、細川暢子、西明紀 : 高自発活動マウスの分子遺伝学的解析により Edem1 は候補遺伝子より除外される. 第 37 回日本神経科学大会. 2014.9.11-13, 横浜.
12. 望月孝子、谷澤靖洋、藤澤貴智、長崎英樹、神沼英里、清水徳朗、豊田敦、藤山秋佐夫、倉田のり、二河成男、中村保一. DNApod : NGS アーカイブ配列からの統合 DNA 多型注釈データベース 生物横断的解析への活用. 第 37 回日本分子生物学会年会. 2014.11.25, 横浜.

<受賞>
なし

③ その他の成果発表：なし