

プロジェクト名： 統計・情報技術を駆使したゲノム多型と表現型多様性の連関解析システムの開発〔生物多様性解析〕

プロジェクトディレクター： 城石俊彦

1. 研究目標

複雑系としての生命システムの理解を深めるために、遺伝的分化を遂げた生物個体の多様性の比較解析が有力な方法論として浮かび上がってきている。そのためには、生物多様性を客観的に記載する数値計測技術と得られた計測値の違いをゲノム多型に結びつけるための統計手法の融合が必要である。また、研究材料となる生物遺伝資源としては、長い進化的時代を経て多数の遺伝子に塩基置換が蓄積し、多様な表現形質を示す生物系統の利用が有効である。国立遺伝学研究所は、古くから世界各地の自然集団から独自に採集し育成してきた遺伝的多様性に富んだマウスやイネ等の多数のモデル生物系統を保有している。本提案プロジェクトでは、これらの生物系統の持つ表現形質の多様性を可能な限り客観的に数値計測化し、それらのデータを対象にして統計数理研究所が培ってきた統計解析技術と国立情報学研究所の情報処理技術を活用することにより、数値計測化された表現型多様性とゲノム多型を関連づけて遺伝子（ゲノム）機能と遺伝子間ネットワークを解明することを目標とする。

2. 研究概要

生命科学における動植物の代表的なモデル生物であるマウスやイネが示す生物多様性を対象に、統計数理技術と情報処理技術を活用してゲノム多型と数値計測化された表現型を関連づけて遺伝子機能と遺伝子パスウェーを体系的に解析するシステムを開発する。また、大学等の外部機関との共同研究により、個々のゲノム機能と生物多様性を生み出した進化メカニズムの解明を行う。

3. 年度計画

テーマ	16年度 予備研究	17年度 プロジェクト初年度	18年度	19年度 中間評価	20年度	21年度
全体	←→ モデル生物の ゲノム・表現 型多様性の統 計解析システ ムの開発	←				→
表現型数値計測 システムの開発		←				→
生物多様性データの 統計的モデリングと 解析システムの開発		←				→

平成16年度（予備研究）

「モデル生物のゲノム・表現型多様性の統計的データ解析システムの開発」という研究課題で予備研究をスタートした。統計数理研究所と国立遺伝学研究所の二研究所間で以下の打合わせを行い、融合研究の実施内容について協議した。全体会議（2004年12月27日）：国立遺伝学研究所において、融合研究全体の基本方針を協議した。特に、マウス・イネの表現型、遺伝子間相互作用の数理解析手法について具体的な研究方針の検討を行った。個別会議（2005年1月26日）：統計数理研究所において、イネのマイクロアレイによる遺伝子発現解析と生殖的隔離障壁の研究について協議した。

平成17年度（プロジェクト開始）

プロジェクト全体の研究目標を参加者が共有し、研究方法などを協議するためのワークショップを開催して情報交換を行った。以下の主な研究成果があった。

- ・ マウス下顎骨の画像データについて P 型フーリエ記述子を用いた形態数値化とそれを用いた主成分分析が系統間の多様性解析に有効であることを示した。X線 CT 値による分析において、マウス内蔵脂肪と皮下脂肪を自動的に判別して各脂肪量を定量化するためのソフトウェアの開発に着手した。
- ・ マウスの社会行動と自発活動の日周期変動について数理モデルを設定した。
- ・ 遺伝的距離の大きな二つの生物系統において、一方の系統のゲノム DNA のプローブセットを用いたマイクロアレイの統計解析について検討し、SNP 由来の見せかけのシグナル強度を判定する方法論を検討した。
- ・ イネ生殖隔離障壁を引き起こす遺伝子座間相互作用検出のため、検定統計量の相関構造の特定とそれに基づいた多重性の調整法について検討した。
- ・ 新しい QTL 解析の開発とその有効性の検証を行うためのマウス F2 交配による実験データの生産を行った。

平成18年度

表現型多様性の数値化のための統計手法の開発を推進し、それらを用いた生物系統からの表現型データの生産と収集を推進した。以下の実質的な成果があった。

- ・ 複数の系統のマウス下顎骨の画像データから二値化画像を作成し、P 型フーリエ記述子を用いた数値化と数値データの多変量解析のためのソフトウェアを整備した。また、CT 画像からの脂肪組織領域自動抽出のためのデータを収集した。
- ・ マウス自発活動量データと社会行動パターンのトラッキングデータを収集し、系統による活動日周期の違いと社会行動の特徴付けについて統計解析を行った。数値化したマウス社会性の度合いを人間の観察により説明する手法を検討した。
- ・ マウス肥満関連形質とイネの穂形質と分けつ能について多数の系統のデータを収集した。また、(B6 x MSM)F2 マウス交配世代について肥満関連形質データを生産し、QTL 解析を実施し

複数の QTL を検出した。

- ・ イネおよびマウスの Affymetrics 社製の GeneChip 上のプローブセットについて、遺伝的距離の大きなイネとマウスの系統間での SNP 情報を整備した。マイクロアレイデータから SNP を検出するソフトウェアを開発し、既知 SNP のハイブリシグナル強度の実データを用いたソフトウェアの有効性の検証を行った。
- ・ イネ生殖隔離障壁の相互作用についてのカイ二乗検定統計量を配偶体及び接合体内の相互作用の各コンポーネントへ分解した遺伝モデルを作成し、再現性のある観測値がこの遺伝的モデルと一致することを確認した。また、カイ二乗確率場の最大値分布について、逐次理論による多重性調整 p 値の解析式を導出した。マウス生殖隔離の原因となる X 染色体置換系統雄の繁殖力低下についての QTL 解析を実施し、原因遺伝子が X, 1 番, 11 番染色体の複数の遺伝子間不適合であることを明らかにした。

平成 19 年度 (中間評価)

表現型多様性の数値計測手法の確立に努め、生物多様性解明のため多数の生物系統から数値計測された表現型データと実験交配個体の遺伝子型データの整備を継続した。プロジェクトの中間時点でのとりまとめを行った。

- ・ 形態多様性について新規アルゴリズムや 3 次元データの数値化についても検討した。
- ・ 系統間の実験交配個体の表現型解析と遺伝子型解析を行い、QTL 等の統計解析のための基盤情報を整備した。従来の QTL 解析手法の改良を試み、それによる量的形質の統計遺伝解析を行った。
- ・ マウスのエネルギー代謝や行動特性に関係した各形質間と多数の QTLs の相関構造 (因果関係) について SEM 等を用いて抽出する統計手法の開発を試みた。
- ・ マイクロアレイデータに基づいた SNP 検出プログラムを改良し、塩基配列未知の生物系統において SNP を考慮した遺伝子発現解析系を開発した。
- ・ 生殖隔離における遺伝子間相互作用の検出に関する多重性調整の実用性をシミュレーションにより検討した。

平成 20 年度

サブテーマ(1)「表現形質の数値計測システムの開発」では、計測された数値データを統計解析に適用するため、生物系統や遺伝的交配個体からの表現型データ収集の拡充をおこなう。表現形質の数値計測のために開発したプログラムについては、測定精度の向上と機能拡張を計り、表現型抽出力の向上を目指す。

- ・ 形態多様性についてはマウス脂肪領域判定等に使用可能なアルゴリズムを開発し、これまでに開発したものと比較して皮下脂肪と内臓脂肪の判定に関して精度向上をはかる。
- ・ エネルギー代謝関連形質の QTL 解析を行い、高い LOD 値を示す QTL を検出する。
- ・ 野生イネ 46 系統の穂形質情報を収集する。野生イネ *O. rufipogon* 由来系統の形質情報の収

集を行い、アレイ用サンプル個体を整備する。

- ・ マイクロアレイデータ解析用に開発した発現及び塩基多型の同時検出法(SNEP: Simultaneous detection of nucleotide and expression polymorphisms)を改良し、検出率や擬陽性などのパフォーマンスの改善をおこなう。

サブテーマ(2)「生物多様性データの統計的モデリングと解析システムの開発」では、各生物系統の特徴抽出を統計学的に行うための研究を一層推進する。

- ・ 形態多様性データの統計的モデリングについては新たな手法の導入を試みる。
- ・ 動物の移動行動の統計モデリングでは、ベクトル間相互作用など様々なパターン分類を応用したシミュレーションによる数理モデルの検証を行う。
- ・ 生殖的隔離に関与する遺伝子座相互作用の多重性に関しては、調整後の閾値以下の真偽判別を真の相互作用検出に応用する。
- ・ QTL解析法の最適化については、多変量の連続変量である表現型と多変量の離散変量である遺伝子型を関連づけるモデルの検討や、カーネル法の適用を検討する。

平成21年度

表現型多様性の客観的な数値計測化システムと遺伝実験データの統計解析手法を確立して、研究コミュニティに広く公開する。また、開発したシステムを利用して表現型多様性を制御する責任遺伝子(群)を探索する。生物多様性に関わる遺伝子間相互作用(ネットワーク)について、本プロジェクトで開発した統計解析手法で解明する。自然集団での生物多様性についての表現型データベースを構築して公開する。以上の確立された統計手法をマウスやイネのさまざまな形質の解析に適用し、その有効性を検証する。5年間のプロジェクトの成果を基にして国内外の研究者に広く呼びかけて公開シンポジウムを企画する。

平成22年度以降の展開

生物形態、エネルギー代謝、植物の生育特性、動物行動などの高次生命機能に係わる遺伝子群や遺伝子パスウェイの解析システムを開発することが最終目標である。この研究の延長線上には、個体レベルでの様々な表現型(ゲノム機能)を遺伝子機能のネットワークとして捉えようという Systems Genetics への発展を想定している。信頼性と再現性の高い遺伝実験データからの統計・情報解析によるデータマイニング(情報抽出)によって、実体に近いシステムズバイオロジーへの展開が期待できる。さらに、自然集団での生物多様性にも目を向けることにより、進化過程を配慮した遺伝子ネットワークのより深い理解に至ることが可能であり、その方向へと研究を展開していく計画である。

4. 研究費の推移

平成17年度実績： 157,820 千円

平成18年度実績： 135,170 千円
平成19年度実績： 118,770 千円
平成20年度実績： 137,750 千円
平成21年度見込： 128,000 千円

5. 平成20年度の研究実施体制

サブテーマ(1)：「表現型数値計測システムの開発と実験データの収集」

研究代表者：[国立遺伝学研究所] 城石俊彦

共同研究者

[国立遺伝学研究所] 倉田のり 春島嘉章 田村 勝 高田豊行 前野哲輝 岡(木曾)彩子
久保貴彦 永口 貢 堀内陽子 小出 剛

[国立情報学研究所] 北本朝展 藤山秋佐夫 佐藤真一

[統計数理研究所] 江口真透 池田思朗 藤澤洋徳 川崎能典 坂口隆之

Nurul Haque Mollah

[東京大学 大学院新領域創成科学研究科 情報生命科学専攻] 中谷明弘 廣田晃士

[長浜バイオ大学 バイオサイエンス学科 生命情報科学コース] 阿部貴志

サブテーマ(2)「生物多様性データの統計的モデリングと解析システムの開発」

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] 江口真透 奥田将己 川崎能典 坂口隆之 田中英希 種村正美
田村義保 土谷 隆 藤澤洋徳 Dou Xiaoling Nurul Haque Mollah

[国立遺伝学研究所] 小出 剛 岡 彩子 城石俊彦 杉本大樹 高田豊行 田村 勝
春島嘉章 前野哲輝 細谷正樹

[東京大学] 原 尚幸

[九州大学] 二宮嘉行

6. 平成20年度の研究進捗

サブテーマ1. 「表現型数値計測システムの開発と実験データの収集」

課題 a. 3D 画像による体脂肪計測法の開発とデータマイニング (北本, 藤山, 佐藤, 高田, 前野, 城石, 中谷, 廣田, 阿部)

[研究の目的]

X線CT装置の高性能化に伴い、現在では多数のスライス画像を短時間で取得することが可能となっている。そこで、大量に得られた画像データから対象臓器を正確に抽出し、客観的数値データに基づく測定をおこなう手法の開発は非常に重要であり、この解析手法の構築は本研究が目的

としている CT 画像を使用したマウス表現型の収集を高速に行うこと以外にも、医療分野における診断の高速化、定量化などを目的とした診断支援に応用できると考えられ、非常に有効かつ汎用性がある。本研究の目的は、マウスの全身 CT(Computed Tomography:コンピュータ断層撮影)画像から内臓および皮下脂肪、さらには脂肪組織以外の臓器を自動的に分離し、正確で客観的な数値データを算出するための手法について検討することである。この目的を達成するためには、腹筋線の自動認識が必須である。腹筋線認識については、形態テンプレートを構成するための特徴点の選定と、画像データからの特徴点の数値情報の抽出を最適に行うためのアルゴリズムの構築を進める必要がある。そこで我々はマウス全身を対象として撮影した CT 画像から脂肪組織(皮下・内臓脂肪)を自動測定し数値化するシステムを構築するため、複数個体の CT 画像を用いて各組織の CT 値の抽出と画像毎の CT 値のばらつきを想定し、ノイズに強いアルゴリズムを実装した脂肪組織の自動分離計測ソフトウェアの開発を進めている。これまでに、2種類の脂肪組織を分離するために、主に脂肪組織および腹筋線を対象として形態テンプレートを作成していたが、処理対象の臓器の範囲を広げて、抽出する情報量と精度を向上させる。複数の臓器を扱うことによって、脂肪組織のみでは判別が難しい部分での抽出精度の向上や、各臓器周辺の脂肪組織の分布情報の抽出が期待できる。

[平成20年度の進捗]

マウス個体を用いた脂肪種判定について、ピクセル間に定義したネットワークの形状の解析に基づく ActCuts 法を開発し、良好な判定結果が得られることを確認した。前年度までに開発した ActScan 法と比較して皮下脂肪と内臓脂肪の判定に関して概ね 10%以上の精度向上を確認した。また、臓器ごとの CT 値の分布を調べ脂肪領域を抽出するための CT 値の閾値の検討を行った。

自動解析システム ActCuts法の開発：

CT 画像のピクセル間に定義したネットワーク形状の解析に基づく ActCuts 法と呼ばれる手法を開発した。ActCuts 法は、前年度までに開発した ActScan 法を補完および拡張したものである。ActCuts 法は、まず CT 画像から脂肪組織に相当する CT 値をもつピクセルを抽出し、画像上で隣接したピクセル同士を辺で結ぶことによって、CT 画像全体を覆うグリッド状のピクセル間ネットワークを生成する。この際に、CT 値が類似し合ったピクセル群がセグメント化される性質に注目して内臓脂肪と皮下脂肪の分離を行う。両脂肪が腹膜によって完全に分離される理想的な状況では、連結した部分ネットワークを抽出することによって、腹膜の内外の脂肪を抽出できるが、実際のデータでは腹膜の位置自体も予測する必要がある。また、腹膜が画像上で断続的になっており、両脂肪が腹膜によって完全に分離されていないこともある。そこで、ActCuts 法では、明らかに腹膜の中と外に位置するピクセルペア群を初期解から抽出し、このピクセルペア群間でネットワークが最も「細く」なっている位置(隘路・くびれ)を腹膜として抽出する。初期解として、前年度までに開発した ActScan 法による解、或いは、新規に開発した GUI によって人間が対話的に与える解を用いることができる。

図 1 は、ピクセル間ネットワークに基づいた脂肪種判定アルゴリズムの概略を示したものであ

る。まず、隣接した脂肪ピクセル同士をグリッド状に接続する。次に、2つの仮想的なノード S と T を用意し、各脂肪ピクセルを、初期解で内臓脂肪であれば S に、皮下脂肪であれば T に接続する。その上で、ノード S から T に向かうネットワークフロー（水色矢印）を計算し、流れの幅が最も狭くなる部分を2つの脂肪の境界線として抽出する（青線）。境界線のどちら側にあるかによって、脂肪種を判定する（内臓脂肪を赤色、皮下脂肪を緑色で表示）。ピクセル間ネットワークを隣接したスライス間にもまたがって3次的に定義することで、単独のスライスでは判定できない領域も体軸に沿った前後関係から情報を補完して判定精度を改善することが可能になっている。

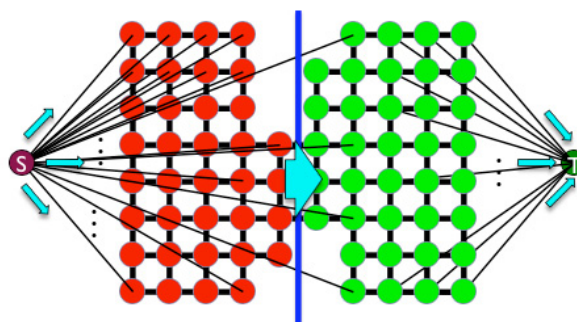


図1 ピクセル間ネットワークに基づいた脂肪種判定アルゴリズムの概略

図2は、ある個体の連続する3スライス ($k-1$, k , $k+1$) での判別結果を示したものである（実際の判別処理は1個体分の全スライスに対して行われる）。ここでは、内臓脂肪を赤色で、皮下脂肪を緑色で示している。上段は、前年度までに開発した ActScan 法による判別結果である。自由に変形する曲線をフィッティングすることによって腹膜の認識を行うが、この例では、腹膜以外の境界線を誤抽出している。中段は、本年度に開発した ActCuts 法による判別結果である。

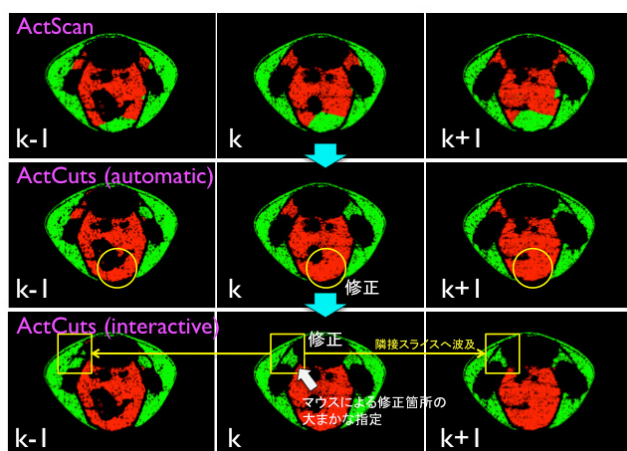


図2 各種アルゴリズムによるCTスライスの領域判別結果

上段に示した ActScan 法によって得られた結果を初期解として用いている。丸印で示した領域の判別結果が修正されている。下段は、中段に示した結果から、対話的な初期解の追加と ActCuts 法の適用を繰り返して得た判別結果である。特定のスライス上での修正は、ピクセル間ネットワークを介して隣接するスライスにも波及するようになっている。

上述の3手法の判別結果を比較したものを図3に示す。ActScan 単独では良好な判別精度（別途生成した「手作業」による正解との比較）が得られていなかった28個体を選択して、全体の判別精度を対話的な ActCuts 法で得られた結果と比較したものである。横軸は個体 ($1 \sim 28 \times 3$) を示し、縦軸は判別精度（対話的な ActCuts 法との相対評価）を示している。自動的な ActCuts 法は、対話的な ActCuts 法の概ね 95~99% 程度の精度を達成していること、また、ActScan 法に比べて 10% 以上向上していることが確認できた。

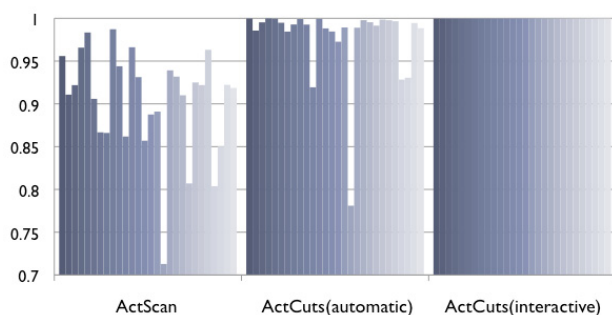


図3 各種アルゴリズムによるCTスライスの領域判別結果比較

今後の課題

これまでに開発した ActScan 法と ActCuts 法は互いに補完関係にあり、対話的な ActCuts 法も含めて良好な結果を達成している。但し、当初に予定した、脂肪以外の臓器の情報の加味は未完成なままになっている。一つ目の原因として、臓器間での CT 値の差異が予想以上に少なかったことが挙げられる。図 3 は、下段に示す臓器の CT 値の分布を示したものである。中段は、特定の断面における CT 値の分布を色で示している（赤い領域ほど CT 値が高い）。上段は、各臓器での CT 値の頻度分布（割合）を示している。上段の縦軸は CT 値、横軸は中下段での位置、色によって CT 値の頻度（赤いほど頻度が高い）を示している。例えば、内臓脂肪は-250 以下の CT 値となっていることが分かる（上段）。一方、皮下脂肪は内臓脂肪に比べて均一ではなく（中段）、-125 以下の CT 値となっていることが分かる（上段）。また、褐色脂肪組織は、特異的な CT 値となっているが、その他の臓器に関しては、殆ど差異がないため、CT 値の分布のみでは抽出できないことが分かる（上段）。

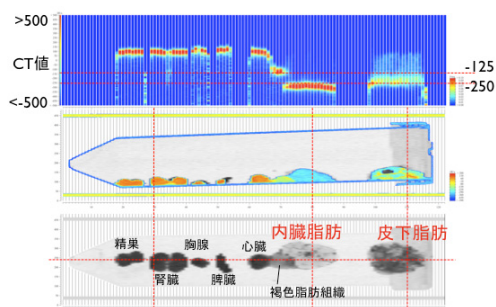


図 4 各種臓器の CT 値の分類

このため、脂肪以外のより多くの臓器の情報を加味するためには、各臓器の位置的な情報を用いた抽出処理が必要になる。図 5 は、1 個体分のスライスを統合することによって得たマウス全体像を特定の 2 平面で仮想的に切断したものである。横軸はスライス番号を示しており、実際のスライス位置を縦線を表示している。各ピクセルは、CT 値に従って着色しており（赤い領域ほど CT 値が高い）、また、隣り合うスライス CT 値を用いてスライス間領域の CT 値を予測および補完して、擬似的に解像度を高めて表示している。これまでは、主に体軸に垂直な断面による解析を行ってきたが、体軸に平行な断面を用いた解析の可能性を示しており、多様な角度からの解析の必要性を示唆しているものと考えられる。

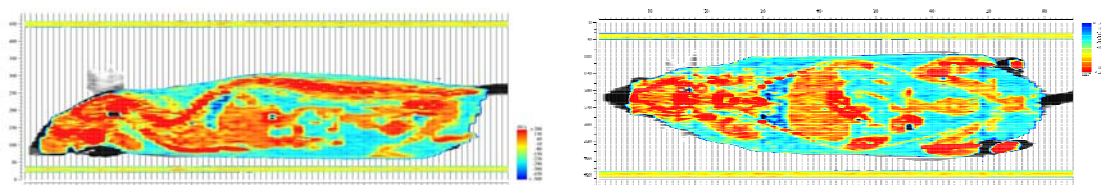


図5 スライスを統合することによって得たマウス全体像

課題 b. マウスエネルギー代謝関連表現型計測法の開発と実験データの収集 (高田, 前野, 城石)

[研究目的]

本研究は遺伝的に離れた複数のマウス系統を基盤にして, 肥満ならびにエネルギー代謝に関連する様々な表現型を計測し実験データを収集することで, これらを制御する遺伝子群を統計遺伝学的に検出するための実験データの整備を目的としている。具体的には, 西ヨーロッパ産汎用実験系統である C57BL/6J(B6: *Mus musculus domesticus*), これとは遺伝的距離が離れた関係にある日本産野生由来系統の MSM/Ms (MSM : *Mus musculus molossinus*), および MSM/Ms と同亜種に属する JF1/Ms(JF1)を使用し, 解剖学的, 血液生化学的パラメータをはじめとしたエネルギー代謝関連表現型を経時的に計測する。最終的には, 得られた数値データを使用し, 両者のゲノム多様性を指標とした多段階の全ゲノムスキャンを行い, 各種表現型の責任遺伝子領域とそれらが関与するエピスタシスの検出を行うことを目的としている。

[平成 20 年度の進捗]

平成 20 年度は, これまで JF1 系統と B6 系統の F2 世代個体を用いたエネルギー代謝関連形質の QTL 解析 ((B6xJF1) F2 解析系, 雄 85 匹, 雌 86 匹使用) を詳細におこない, エネルギー代謝関連表現型遺伝子座について, ゲノムの 30 か所以上高い LOD 値 (2.5 以上) を示す QTL を検出した (図 1)。

Trait	LOD					
	2.5-3	3-3.5	3.5-	2.5-3	3-3.5	3.5-
BW_10W	0	0	0	1	0	0
BW_15W	0	0	0	2	0	0
BW_20W	1	0	0	1	1	0
BW_25W	1	0	0	0	2	0
TTL_10W	1	0	1	0	0	0
TTL_25W	0	0	1	0	0	1
TL_10W	1	0	1	0	0	1
TL_25W	1	0	1	0	0	1
BL_10W	1	0	0	0	0	0
BL_25W	0	0	0	0	1	0
BMI_10W	0	0	0	1	0	0
BMI_25W	1	0	0	0	0	1
CC	0	0	1	0	0	1
LW	1	0	0	1	0	0
TG	0	0	0	0	0	0
TTLC	1	0	0	0	0	0
HDLC	0	0	0	0	0	0
NHDLC	0	0	0	0	0	0
INS	0	0	0	0	0	0
LEP	0	2	0	0	0	0
ADP	0	0	1	0	0	1
Total	9	2	6	6	4	6

Trait	
BW_10W	: 10週令体重
BW_15W	: 15週令体重
BW_20W	: 20週令体重
BW_25W	: 25週令体重
TTL_10W	: 10週令全長
TTL_25W	: 25週令全長
TL_10W	: 10週令尾長
TL_25W	: 25週令尾長
BL_10W	: 10週令胴長
BL_25W	: 25週令胴長
BMI_10W	: 10週令BMI
BMI_25W	: 25週令BMI
CC	: 毛色
LW	: 肝臓重量
TG	: トリグリセリド
TTLC	: 総コレステロール
HDLC	: HDLコレステロール
NHDLC	: nonHDLコレステロール
INS	: Insulin
LEP	: Leptin
ADP	: Adipone ctin

図 1 (B6xJF1) F2 解析系を使用した QTL 解析のまとめ。雄 (M), 雌 (F) の結果を示す。本解析によりエネルギー代謝に関係すると考えられる QTL を複数確認することができた。図では雌雄の QTL は別に表示した。雌雄合わせて計 33 の QTL を確認したが雌雄で重複するものや測定項目で重複する QTL が存在すると示唆される。

特に肥満関連のパラメータとして, レプチン, アディポネクチンについてはその濃度を制御していると考えられる QTL を複数検出することができた。次に, 複数の QTL が関与する表現型についても解析するため, r/qtl の Two QTL genome scan についても行った (図 2)。左上にエピスタシス LOD, 右下に Additive LOD を示す。前年度に行った体重以外のすべての項目について解析をおこなった結果, 各形質には多因子により制御されていると考えられる遺伝子座が複数確認できる。これと比較して, 例として示した毛色は一点の強い効果を示す遺伝子座が相加効果を示すエリアに雌雄に共通して確認できる。これらの QTL が相加的効果以外にも相乗的効果の存在が示唆されることから, 今後はこれらの効果の真偽を統計学的に検証する必要がある。現在進行中の課題である B6 と MSM ならびに B6 と JF1 系統の F2 世代の解析に関しては, 高密度の SNP 検出システムが整備でき次第, 遺伝マーカー数を増やし, より密度の高いマッピングパネルを作成して QTL の候補領域を絞り込む。今年度までに実施した B6-MSM, B6-MSM のデー

タを使用して各 QTL 解析について同様の解析を行うとともに各形質間の相関解析を実施する。

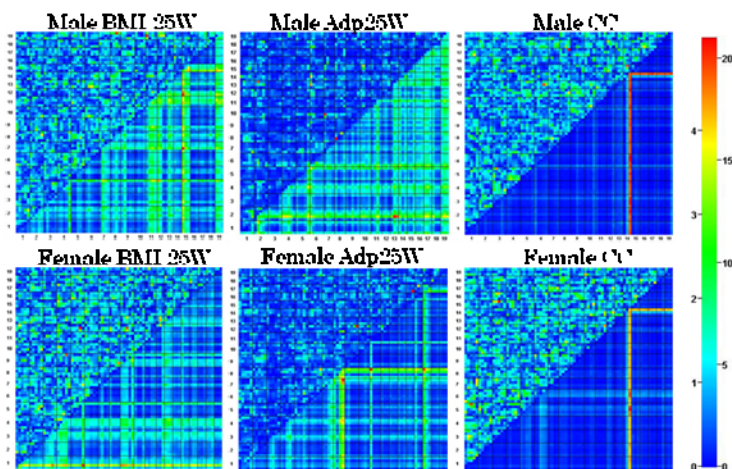


図 2 B6xJF1 F2 解析により示唆されるエネルギー代謝関連表現型の QTL。Two QTL genome scan の結果を示す。雌雄の 25 週令の BMI, アディポネクチン(Adp), コントロールとして毛色(CC)の結果を例として示す。各項目の左上は epistasis に関する LOD, 右下は相加効果に関する LOD を示す。各図の下および左から染色体 1 番から順に 19 番までを示している。各形質には多因子により制御されていると考えられる遺伝子座が確認される。このほかにもエネルギー代謝に関する QTL 解析では複数の遺伝子座による効果と考えられる形質が複数観察された。毛色に関する解析結果は一点の強い効果を示す遺伝子座が相加効果を示すエリアに雌雄に共通して確認できる。LOD と色の関係は右端に記載。解析は r/qt1 によりおこなった。

今後の課題

これまでに得られた (B6xJF1) F2 解析系のエネルギー代謝に関連した表現型データによる QTL 解析に関しては、F2 個体の全ゲノムを対象とした粗いジェノタイピングによる結果のみである。今後は、解析に使用した (B6xJF1) F2 解析系の個体数が少ないのでこの問題を解消する必要がある。さらに一層高詳細に染色体領域の解析を行うためにより多くの SNP 座位の選定を行う必要がある。高詳細に全ゲノムスキャンをおこなうことでより密度の高いマッピングパネルを作成して QTL の候補領域を絞り込むことが可能になると期待できる。また、今年度までに実施した B6-MSM, B6-MSM のデータも使用して各 QTL 解析について同様の解析を行うとともに、これまでに収集した各形質間の相関解析を実施する。その他、今年度から新たなエネルギー代謝関連表現型として呼気測定を開始した。これに関しては現在計測データを蓄積中である。

課題 c. イネストレス耐性・穂形質の系統データ収集と遺伝子発現データとの相関解析法の開発(藤澤, 江口, 倉田, 春島, 堀内, 久保)

【研究目的】

本課題では、多様な生育特性を示すイネ栽培系統や近縁野生イネの形質多様性と、塩基配列レベルの遺伝子構造多型および遺伝子発現多様性との相関を抽出し、個々の形質に関与する遺伝的要因を発見するため、統計的な処理法や解析集団、解析形質の適正な選択法の確立を目指している。形質として穂の器官の形状および生育特性等を中心に系統毎の形質を計測し、もう一方では全ゲノムにおける系統間での SNP データの抽出と、遺伝子発現データの取得を行う。今日では

次世代シーケンサーの発展により遺伝子構造多型と発現多型を同時に見積もる有効な情報収集が可能となりつつあるが、大量のデータから有効な情報を抽出する有効な手法については未だ確立していない。これらの構造、発現多型の大量データと形質多様性のアソシエーション解析においては、特に融合研究としての特性を生かし取り組みたい。

[平成20年度の進捗]

19年度までの、栽培イネ祖先種である *Oryza rufipogon* に絞った穂形質と分けつ能についての調査により、分類学上で言われている多年生と一年生の分類特性が種子稔性、葯長などと高い相関があるという結果を得ている。一方で一次枝梗の数、一次枝梗あたりの穎花数などの調査されたイネの種子生産効率に影響を与える形質の多くは一年生多年生の分類特性との相関は低かった。これらの形質がどのような遺伝的制御を受け多様性を示すのか大変興味深い問題である。

20年度は特に穂形質(穂型, 穎花長, 葯長, 粒重, 1穂穎花数, 稔性)データを蓄積し整備した。さらに遺伝子構造または遺伝子発現の差がマイクロアレイにより測定可能であるか検討した。Affymetrix社のGeneChip Rice Genome arrayは主に栽培イネ *O.sativa*の *japonica*由来の転写産物を基に設計されている。アレイを用い発現多型や塩基多型を検出する手法は課題dにて、ゲノム配列が既知の栽培イネ

の日本晴(*japonica*)と93-11(*indica*)間で開発してきた。栽培イネ *indica*の祖先種である野生イネ *O. rufipogon*; W106と、栽培イネとはゲノム種も異なる *O. punctata*のW1514の2種、いずれも塩基配列が未知の系統を用いて、アレイで発現量多型や塩基配列多型が別々に検出できるか検討した。ゲノムDNAのハイブリダイゼーションにより系統間の塩基配列の差がどの程度シグナル強度に影響を与えるか、またmRNAハイブリダイゼーションにより発現データにその塩基配列構造差が影響するのか比較した(図1)。対角線より右下は日本晴のシグナルが高い事を示す。W106対日本晴のシグナル強度比較は、93-11対日本晴のデータとゲノムハイブリダイゼーション、mRNAハイブリダイゼーション共に良く似ていた。アレイ設計の基データとして使われた系統と遺伝的に近い *O. sativa*内や *O. rufipogon*の系統では下記課題「マイクロアレイ・統計解析システムの開発」で開発した手法を使い発現多型や塩基多型を検出する事が可能と考えられる。一方BBゲノム種W1514は、塩基配列が大きく異なることがゲノムハイブリダイゼーションのシグナル強度より明らかである。アレイが日本晴を基に設計されているので、日本晴よりもシグ

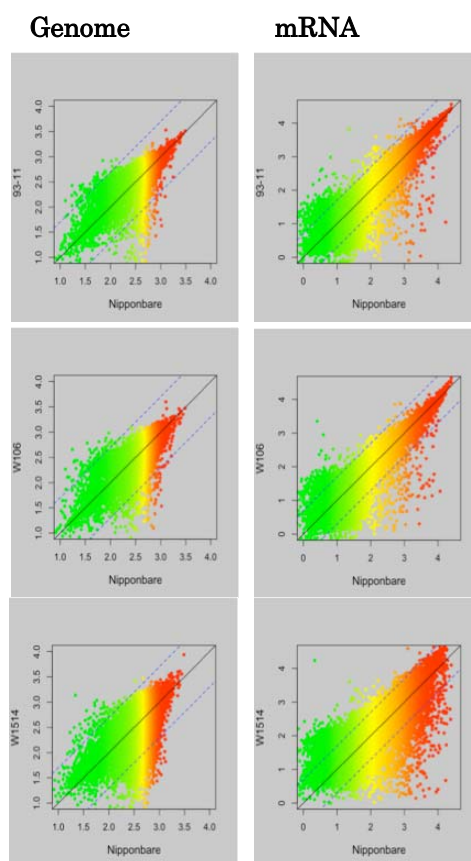


図1

ナル強度がかなり低いものが多いのは予想通りであるが、日本晴よりもシグナル強度が高いものも多数ある。これは、プローブを設定した配列が W1514 では重複しているためと考えられる。さらに、W1514 対日本晴の mRNA ハイブリダイゼーションにおけるシグナル強度比較も、他の 2 つとはかなり異なる。課題 d で開発した手法に加え、遺伝子の重複の程度を調べ重複遺伝子の発現量を推定する方法や、アレイのプローブ設定領域に多数の塩基置換が含まれる場合の遺伝子発現量を推定する改良等の課題が残る。

発現遺伝子の塩基配列が良く調べられている系統では、次世代シーケンサーを用いて遺伝子の数塩基の差や、発現の差を直接観測する試みが行われている。ゲノムの塩基配列が既知の系統と遺伝的に離れた系統で、次世代シーケンサーでの発現遺伝子の解析を行う可能性を検討する必要がある。発現遺伝子の塩基配列多型と発現量多型を同時に調べることができるか、同時にゲノム塩基配列も決定する必要があるのか、アレイ解析を組み合わせた方が良いのかなど、課題 d と合わせて今後実際にデータを測定しつつ検討する予定である。

課題 d. マイクロアレイによるイネ・マウス遺伝子発現データの統計解析システムの開発 (藤澤, 江口, 坂口, 池田, 倉田, 春島, 堀内, 城石, 高田)

[研究目的]

本課題では、多様な生物特性を理解するため、塩基配列レベルの多型および遺伝子発現量の多型を検出する方法の確立を目的とする。Affymetrix 社の GeneChip expression array はゲノム全体の遺伝子の発現量を比較する事を目的に作られたものだが、ゲノム全体に渡る塩基配列の差も見積もる事が出来、生物の多様性を研究する上で有用性が期待される実験手法である。しかしながら、アレイ設計と異なる系統の遺伝子発現と塩基多型を区別して検出することは容易ではない。GeneChip は遺伝子発現量の推定を PM プローブと MM プローブと呼ばれる 25mer のオリゴプローブの 11 組を 1 つのセットとして用いている。PM プローブは標的とする遺伝子の配列に対して完全に相補的な配列で、MM プローブは PM プローブの中央の塩基をその相補塩基に置換えた配列でバックグラウンド補正用に設計されたものである。本課題は、まずこのセット内のプローブシグナルから、遺伝子の発現量の差と塩基配列の差を別々に見積もる方法の確立を目的とする。また、アレイによって検出される塩基多型 (SFP: single feature polymorphism) は、ミュータントや QTL の原因遺伝子の単離、集団遺伝学的解析に利用可能なので、ゲノムハイブリダイゼーションによる SFP の検出方法の検討を行った。更にセット中のプローブの多くが SFP である遺伝子の発現量の推定方法を検討している。

[平成 20 年度の進捗]

遺伝子の発現量の差と塩基配列の差を別々に見積もる方法については、発現及び塩基多型の同時検出法 (SNEP: Simultaneous detection of nucleotide and expression polymorphisms) の開発を行った。SNEP はある遺伝子の塩基配列が品種間で同じで発現量が異なる時、PM プローブの対数シグナル値の品種間差がセット内で一定値であることに着目し、セット内で対数シグナル値

の品種間差がはずれ値となるプローブを探す事により遺伝子の塩基多型を検出し、対数シグナル値の品種間差が 0 より有意に異なる時に品種間で発現量の多型を検出する方法である。はずれ値によるマスク効果は、尤度比検定によるセット内の他のはずれ値そのものの検出力の低下と発現量多型の検出力の低下を招くが、統計的ロバスト法を利用し克服する検定手法を開発、統計言語 R によるソフトウェアを作製した。

今年度は、アレイ実験の特徴である繰り返し実験の回数が少ないが次元が多い検定に対し改良するため、ロバスト法にチューニングパラメーターを導入、更にアレイで 2 系統の発現遺伝子を比較し SNEP を行う時、両方の系統で塩基多型の検出が可能となる両側検定にも対応出来るよう改良した。本研究で開発した SNEP 法と尤度比検定による SFP 検出効率の比較を大麦の 5 種類の組織の mRNA のアレイのデータを用い、既存の方法との比較を受信者動作特性曲線 (ROC curve: receiver operating characteristic curve) の比較によって行った。図1に示した通りどの組織のデータにおいても SNEP 法による優位性が示された。また、イネやマウスの mRNA のアレイデータについてもロバスト法のチューニングパラメーターを導入する事により、SFP 検出の更なる効率化が図られた。

SNEP 法は mRNA のハイブリダイゼーションデータを用い SFP を検出する最良の方法で、ゲノムサイズに関わり無く適用できる優れた方法である。しかし、高発現遺伝子の SFP しか検出できない、遺伝的変異の多い (SFP がセット中に多数含まれる) 時は、SFP の検出力も発現量差の推定も厳しい欠点がある。今回のイネでは、対象とする高発現遺伝子は全遺伝子の 37% で、マウスでは 6.9% でしかない。

大麦などゲノムサイズの大きな種ではゲノムハイブリダイゼーションによる SFP を検出する事は出来ないが、アラビドプシス (~125 Mb) や出芽酵母 (~12 Mb) では、ゲノムハイブリダイゼーションで SFP を検出し、ミュータントや QTL の原因遺伝子の単離、集団遺伝学的解析に用いられている。イネ (~400 Mb) でゲノムハイブリダイゼーションによる SFP の検出を試みた。

イネのゲノムハイブリダイゼーションの実験条件の検討を行い最適な条件を求めた。この実験条件で塩基配列が既知のジャポニカとインディカの 2 品種を使って SFP 検出を、SNEP 法を含め既存の検出方法、バックグラウンド補正法、ノーマライゼーション法により検討した。その結果、バックグラウンド補正なしで、検出法としては SAM (significant analysis of microarray) による SFP の検出が最適であり、27,365 個の真の SFP を検出できた。ゲノムハイブリダイゼーションによる SFP の検出率は 38.6% と低い、解析対象とするプローブ数が多いため検出できた SFP の数は多い。同じ品種の組合せの young panicle の mRNA の SNEP 解析では、SFP 検

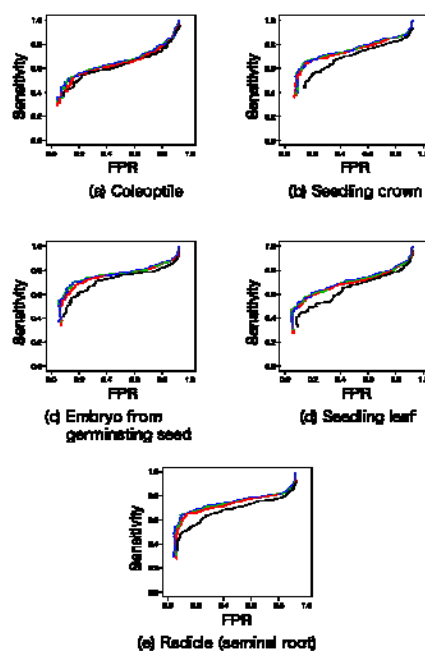


図1 ROC curve for SFP detection. The x - and y -axes are the FPR and sensitivity, respectively. The black, red, green and blue lines are based on LRT, SNEP ($\gamma = 0.2$), SNEP ($\gamma = 0.5$) and SNEP ($\gamma = 0.1$), respectively, using appropriate significance levels, which range from 10^{-20} to 1.

出率は 53.6%と高いが、解析対象となる高発現遺伝子の数が少ないため、真の SFP が 3,577 個しか検出できない。イネで SFP の検出に限ればゲノムハイブリダイゼーションの方が数多くの SFP の検出が可能である。

系統間の遺伝子発現量差を GeneChip にて推定する時、セット内の SFP の数が少ないときは SNEP によって正確に推定できるが、SPF の数が多いと推定する方法が無い。SFP プローブの塩基多型によるシグナル値の減少量が、どのような因子で説明できるか重回帰分析を進行中である。前もってゲノムハイブリダイゼーションの結果を利用して、SFP プローブの mRNA におけるシグナル値の減少量を推定する事ができれば、SPF の数が多いと判っているセットの遺伝子発現量を SFP シグナル減少量で補正することにより、本来の mRNA の発現量が推定可能となる。

サブテーマ 2 : 生物多様性データの統計的モデリングと解析システムの開発

課題 a. フーリエ記述子を用いた形態多様性データの統計的モデリング (田村, 江口, 城石, Mollah, 田中, 細谷)

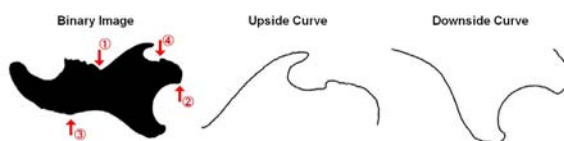
[研究目的]

本研究の目的は、生物形態の統計解析であるが、形を解析するためには、解析のためにふさわしい形の数値記述方法が必要である。形は、本質的に多次元の特質であり、それゆえ定量的評価が難しい。本研究において扱われる形とは、2次元空間における物体の内側と外側の境界線の形状であり、したがって、それ自身と交差しない一筆書きの閉曲線または開曲線の形状である。多くの場面で、一筆書きの曲線形状を記述するのに用いられてきた手法の一つにフーリエ記述子(FD)がある。この手法では、まず曲線上の各点に対し、その点における何らかの幾何学的情報が数値として対応づけられる。そして始点から終点に向けて並べられたその数値の列の離散フーリエ変換として図形は記述される。各点に対応づけられる幾何学的情報が異なるさまざまな FD が開発され、ある種の FD について、不規則形状に対する優れた特徴抽出能力をもつこと、不規則形状の統計解析において有用であること、などが示されてきた。特に閉曲線で表される図形の解析には、シンプルで、優れた特徴抽出能力を有する楕円フーリエ記述子が、最もよく用いられる。本研究では曲線形状の記述方法として FD に焦点をあてる。

2次元部分形状は開曲線で表されるが、形の統計解析において、これまで、部分形状の数値記述に用いられてきた開曲線記述子として、接線フーリエ記述子(TFD, P型フーリエ記述子)がある。TFDは、浮世絵の作者判別、ヒトの横顔輪郭の認識、ハナハスの花卉やイネの葉の形状の統計解析などに用いられ、それぞれの場面での有用性が示されてきたが、TFDは、曲線上の隣接点間の位置の差分系列のフーリエ変換であり、ノイズを増幅する傾向がある。

[平成20年度の進捗]

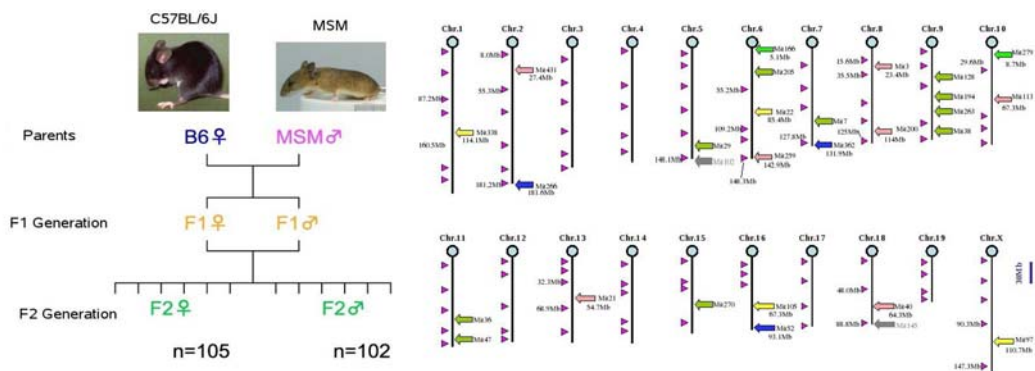
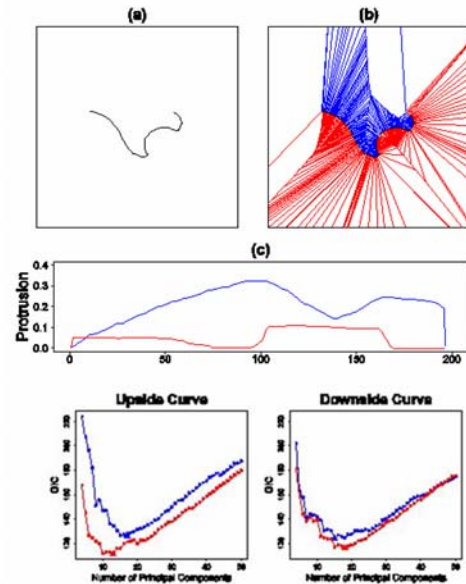
(1)まず、右図のごとく不規則で複雑な開曲線形状の微妙な特徴をとらえる能力について、



さらに優れた記述方法を求めて、protrusion フーリエ記述子(PFD)を考案した。PFD法では、

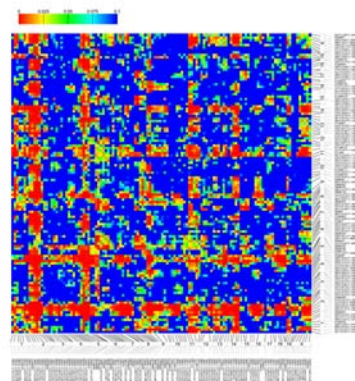
始点から終点への向きに対し左右からの protrusion (突出度合い) が、曲線上の各点において求められる左右からの protrusion をそれぞれ実数、虚数成分とする複素数が、曲線上の各点に対応付けられ、PFD は、始点から終点にむけて順序づけられたその複素数列のフーリエ変換として定義される。ちなみに曲線上の各点における protrusion は、曲線形状に対し一意に定まるスケルトン-垂線グラフと名付けられた場を用いて計算される。

(2)次に、不規則形状に対する特徴抽出能力に関し TFD と PFD を比較する実験を行った (右図)。用いたデータは、9つの系統 (合計 93 匹) のマウスの下顎骨から特定部位を表す 93 個の開曲線であり、まず、TFD、PFD それぞれを用いて曲線を数値表現した。系統の判別モデル (ロジスティック回帰モデル) を当てはめ、モデルの良さを一般化情報量基準 (GIC) を用いて情報論的に測った。もし用いた記述子が系統ごとの形状の差をより良くとらえるならば、データは系統ごとに明確に分離し、データに当てはめられた判別モデルの GIC 値はより低くなるはずである。結果として、TFD 使用時より PFD 使用



時の方が GIC 値は低かった。これは、PFD 使用時の方がより正確に系統を判別できること、つまり、PFD が TFD より不規則形状に対する特徴抽出能力において優れていることを示している。

(3)さらに、2つの系統を起源とする F2 世代 (合計 207 匹) のマウス (下図) を用い下顎骨の部分形状の遺伝解析を行った。本解析の目的は形態形成に対して影響力の大きい遺伝子の探索と、各遺伝子の影響の様子の視覚化による解釈である。全 207 個体について、ゲノム上の特定の 122 箇所において、分子マーカーを用いて遺伝型が調べられている。形は量的形質であり、一般に量的形質は、相互作用する複数の遺伝子の影響をうけている。最も単純な場合として 2つの遺伝子の相互作用 (エピスタシス) を考慮し、ゲノム上の位置 (の付近



にある遺伝子)の任意のペアについて、その下顎骨形状に対する影響力を測った。また、その影響力の測定方法として、本質的に多変量の特質である形を、多変量のまま解析するために、GIC値を検定統計量として用いる仮説検定を採用した。まず形のデータを、遺伝型の情報を用いずにランダムに分割し、それに判別モデルをあてはめ、モデルの良さをGIC値で測る、という計算を十分な回数くりかえし、GIC値のnull分布(形にその遺伝型が何の影響ももたない、という仮説の下でのnull分布)を用意しておいた。ゲノム上の任意の位置のペアについて、そのペアの遺伝型の情報を用いて形のデータを分割し、判別モデルを当てはめ、モデルの良さをGIC値で測る。そして、そのGIC値のnull分布に対する片側P値を求め、それをその遺伝子の形に対する影響力とみなした。本解析の結果、異なる部位に対し、形に対して大きな影響力をもつゲノム上の位置のペアが大きく異なること、さらに記述子からの曲線形状の視覚化により、各々の位置のペアの形に対する影響の仕方が、それぞれ異なることなどが分かった。

課題 b. マウス社会行動の統計的モデリング (種村, 土谷, 小出, 杉本)

[研究目的]

集団中で連続的な多様性を示す高次機能の理解は、体質や性格などの個人差をもたらす機構の解明という点からも現在注目されている研究分野の一つである。その高次機能の中で、行動表現型は形として残らないため、その計測記録システムおよび解析には困難がつきまとう。また、行動は時間軸に沿った発現をするため、その定量的な評価は難しい。本研究では、マウス行動を定量的に計測し、得られたデータを数理的な手法を用いて解析して、将来の研究において発展性のある行動成分を抽出し、その形質の背後にある特性をモデル化により明らかにすることを目的とする。

[平成20年度の進捗]

(i) 複数の動物個体が示す社会行動は、その一般的な定量化が難しく、解析が著しく困難である。本研究では、2個体のマウスを同時にオープンフィールドに置いた際の10分間の社会行動の画像データから、自動トラッキングにより、2個体の時系列位置座標データを抽出し、それらのデータに基づいて2個体の行動を統計数理モデルで特徴づける解析を進めた。20年度は50ペアについて、画像の目視観察法により行動項目分析時系列データを取得した。また、マウス24系統、461ペアのオープンフィールド社会行動について、動画画像データとその時系列移動座標データを抽出し、基礎解析を行った。

(ii) マルコフ係数によるマウスの社会性の定量化についてより詳細に検討を進めた。これは、2個体のマウスの画像時系列を5つの状態に分解してその状態時系列に基づくマルコフ行列を構成し、人間による目視観察による各画像の社会性の得点(8人の観察者の得点の平均;より詳しくはB6とMSMの画像を見たのちに、コンソミックマウスによる画像がB6に近いMSMに近いかを得点化したもの(B6とMSMをその行動パターンより社会性が低いタイプと高いタイプの象徴例と見なしている))から社会性を定量化しようというものである。クロ

スバリデーションによって推定の安定性を確認した。また、AIC により、要因としてとりあげるマルコフ係数の個数を減らすことも試みた。目視による社会性得点を横軸、マルコフ行列による社会性推定得点を縦軸として描いたものが図 1 である。また、図 2 に各画像データについて目視得点の箱ひげ図とマルコフ行列による得点推定値の関係を示す。これらの結果より概ねマルコフ行列による推定値が一定の妥当性を有すると考えている。特に、図 1 から、概ね推定値の方が 60

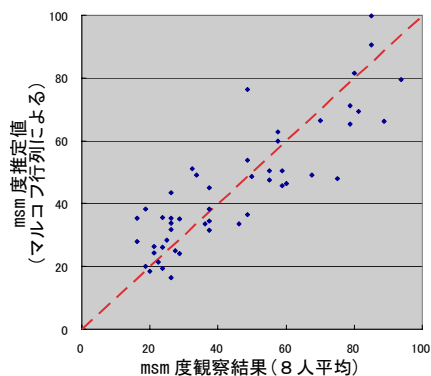


図 1

点以上のものであれば、社会性が高いと判定できるのではないかと考えられる。MSM 度が高いデータについて、目視観察による結果よりも推定値が低くなる傾向があるが、この点は今後モデルを改良することで改善したい。来年度は各状態の自動推定を隠れマルコフモデルにより行い、411 ペアの画像データを自動解析する。

8 人評価点の箱ひげ図とマルコフ行列による推定値

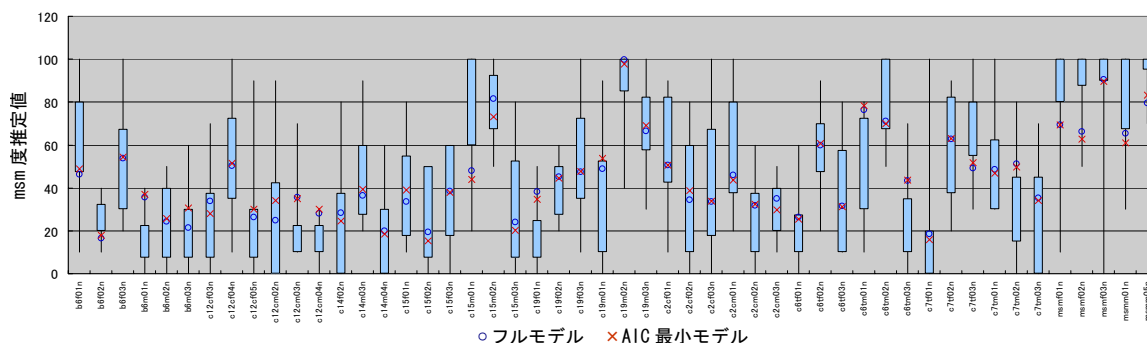


図 2

(iii) 2次元 unit vector chain による解析 (平成 18 年度研究成果報告書 p.38 参照) では、マウス 2 個体の行動の時系列に対応して 2 個体間の角度を割り当てるアルゴリズムを一層改良した。図 4(a), (b) は改良したアルゴリズムによって割り当てられた角度の時系列の例である。図 4(a) および (b) はそれぞれ B6 雌および MSM 雄に対応する。得られた角度時系列から B6 および MSM の雄・雌ごとに方向相互作用パラメータの推定値を求めることによって、B6 と

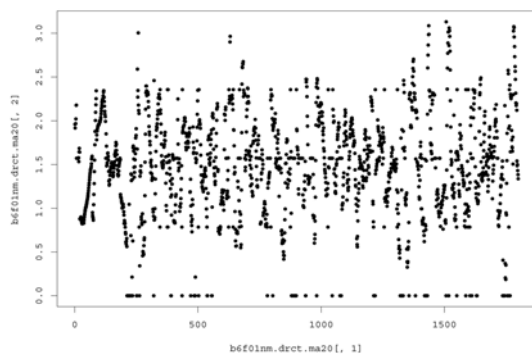


図 4(a)

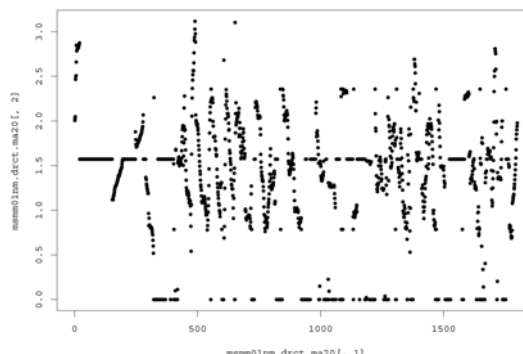


図 4(b)

MSM の間の社会行動に明確な違いが現れること、また、性別による行動の違いも定量化できることが分かった。今年度に新しく整理されたマウス 24 系統 461 ペアに対する社会行動トラッキングデータについても予備的解析を行った。

課題 c. 多変量時系列モデルによるマウス自発行動の特徴抽出 (川崎, 小出, 杉本)

【研究目的】

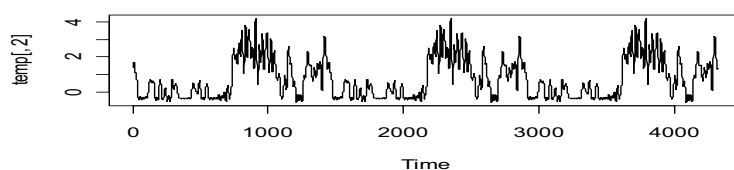
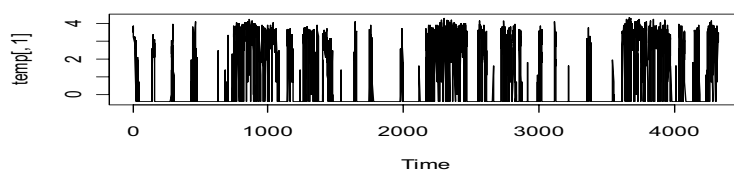
遺伝研において開発維持している多様な野生由来マウス系統は、高次機能の多様性を解析するための有用なリソースとなる。その高次機能の中で、行動表現型は形としては残らないため、その計測記録システムおよび解析には困難が伴う。また、行動は時間軸に沿った発現をするため、その定量的な評価は難しい。本研究では、マウス行動を定量的に計測し、得られたデータを数理的な手法を用いて解析して、将来の研究において発展性のある行動成分を抽出し、その形質の背後にある特性をモデル化により明らかにすることを目的とする。そのため、統計数理研究所と共同で、マウスの自発活動性や社会性などの行動を時系列に沿って効果的に解析するシステムの確立を進める。

【平成 20 年度の進捗】

自発活動性は、動物にとってテリトリーの確保や餌の探索といった生存に深く関わる重要な行動形質の一つである。その活動には周期性が見られ、概日周期についてはすでによく知られている。しかし、動物は活動期においても一様に活動しているわけではなく短い時間で活動と休息を繰り返しているように見える。このような超日周期については、まだほとんど研究が進んでいない。その原因は、超日周期を解析するための有効な解析手法がこれまで確立されていなかったためである。そこで、遺伝的に異なったマウス系統である MSM と C57BL/6 (B6)を用いて、その活動性について解析系の確立を目指した。

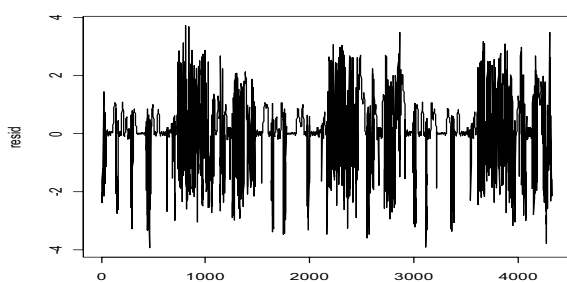
解析に用いたマウスは、照明を 8:00 から 20:00 までを明期、20:00 から 8:00 までを暗期としてコントロールし、室温 $22 \pm 2^{\circ}\text{C}$ 、湿度 $50 \pm 10\%$ に保った飼育室内で維持した。解析には MSM と C57BL/6 の生後 8~15 週齢の雌雄を解析に用いた。各マウスはテストの前に個別に飼育ケージに移し 1 日間飼育した。その後、ACTIVITY SENSOR (O'hara, Co., Ltd) の測定ケージにそれぞれ移し、4 日間連続でその活動量を自動計測した。最初の 1 日間は馴化期間として解析からは除外した。各個体の自発活動量は 1 分毎に表す時系列データとして記録した。

H20 年度の進捗は以下の通り。マウスの自発行動を記録したデータは非負整数値による計数



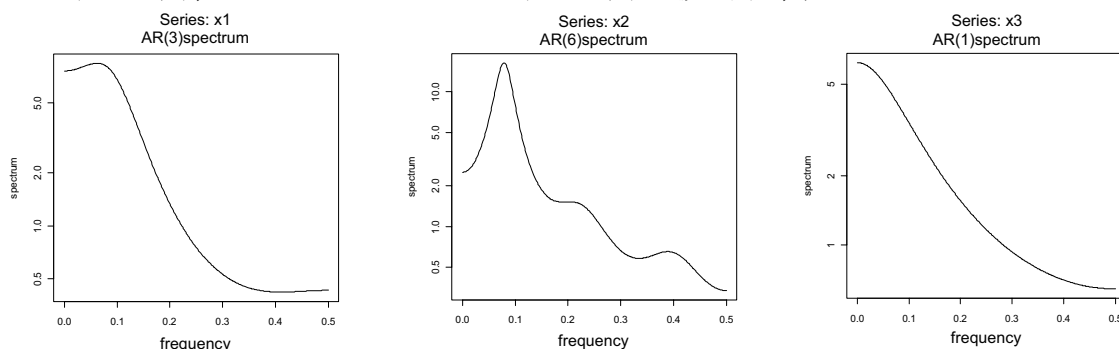
時系列で極めてゼロが多く、上下の非対称性が強い。興味があるのは活動期の短期的変動の特徴であるので、1分値で計数0は欠測として非ゼロのデータを対数・巾関数等で線形化し、欠測値を許す線形ガウス型カルマンフィルタで、日周期まわりの自己回帰モデルをあてはめることにより、特徴的な活動周波数を推定する方法を研究した。

右上の図は、B6 オス一頭から得られた活動データ（上側パネル）と、そのポアソン強度のうち周期関数（sin, cos）の組み合わせで説明できる部分のプロット（下側パネル）である。周期関数で説明しきれない「残差」の部分から活動の特徴量を抽出できるかどうかが本題である。



その残差のプロットが左図である。縦軸は強度ではなく対数強度になっていることに注意する。一見して、暗期（活動期）と明期（休息期）では対数強度の分散は異なり、休息期であっても時折活動の増減が見られる。ここではとりあえず、消灯直後の20時以降でなるべく活動が

継続している期間を取り出して、その期間にARモデルをあてはめて周波数特性を探る。データは1分値、3日間で4320時点あるが、730時点から1000時点（1日目夜）、2170時点から2440時点（2日目夜）、3610時点から3880時点（3日目夜）を抜き出し、個別にモデルをあてはめた。



モデルの推定値から導かれるスペクトル密度が上の図である。左から1日目、2日目、3日目であり、それぞれAR次数は3次、6次、1次である。一例だけから一般化はできないが、にもかかわらずこの結果は興味深い。まず、2日目は鋭いピークを周波数0.1付近で記録しており、10分弱に1回程度の変動が活動性を支配していることが示されている。1日目と3日目の結果は、10分より長い周期が存在している可能性を示しているが、これは局所的なトレンドに引っ張られている可能性もある。更にもう一段のトレンド除去を行えば、2日目のようなスペクトル特性が得られるかもしれない。まずは一頭に関して3日間である程度共通のスペクトル構造が取り出せるような枠組みを考え、然る後固体間での比較、系統間での比較に持ち込むのが今後の目標である。H21年度は、当座は局所定常ARモデルを上で言う残差系列にあてはめ、データドリブンで決めた時の区間区分とそのもとでのスペクトル特性を観察する。

課題 d. 生殖隔離障壁に関わるエピスタシスの統計学的検出と多重性調整（栗木、藤澤、倉田、春島、岡、城石、二宮）

[研究目的]

本研究課題では雑種の自殖集団 (F₂) におけるマーカー遺伝子型の分離データを用い、異なる遺伝子座が生殖的隔離障壁として作用し特定の遺伝子型組合せを持つ配偶体又は接合体が選択を受けたか否かの検出することを目的とする。

栽培イネ (*Oryza sativa*) には、ジャポニカ型とインディカ型があり両者間で生殖的隔離が観測出来る。ジャポニカ型日本晴とインディカ型 Kasalath との雑種 (F₁) の自殖後代 (F₂) 186 個体の集団を用い、イネ高密度連鎖地図が作成されており、そのマーカーの遺伝子型分離頻度から、異なる遺伝子座間で遺伝子型分離の独立性検定を行うことにより、生殖的隔離障壁間の相互作用の検出を目指す。

配偶体及び接合体内の生殖的隔離障壁間の相互作用検出のため、マーカー分離の独立検定を行ったところ、連鎖しているマーカー間だけではなく多くの相互作用を示すと考えられるピークが認められた。相互作用検出のため独立性検定を全てのマーカー間で行うことは、組合せ数が膨大になるため重大な多重性を引き起こす。多重性調整を行い調整 p 値によると、全ての相互作用のピークは有意水準 0.05 以下の非有意な結果となった。しかし実験的に再現性の有るピークも存在することから、真の生殖的隔離障壁間の相互作用を検出する別のアプローチが必要となった。

自由度 4 の独立性カイ 2 乗検定統計量は、自由度 1 の 4 個の独立なコンポーネントへ分解が可能である。そこで生殖的隔離障壁間の相互作用について単純な致死遺伝的モデルをたて、どのような相互作用カイ 2 乗統計量及び各コンポーネント量を与えるを考察した。分離ゆがみのゆがみ方やコンポーネントへの分配のされ方に適合するモデルがあるかを詳細に調べることにより、相互作用による独立性検定量のピークとランダムに生じたピークとの判別することが可能と推定された。

[平成 20 年度の進捗]

前年度までは、相互作用による完全致死かつホモ型同士の遺伝型組合せもヘテロ型を含む遺伝型組合せも致死率は同じとする単純なモデルを想定していた。今年度は浸透率 (penetrance) を考慮し、更にヘテロ型を含む遺伝型組合せがホモ型同士の組合せの半分の致死率となる場合も含め、より一般的な場合に遺伝的相互作用致死モデルを拡張した。配偶体では遺伝子型が 2 通りなので、2 遺伝子座の遺伝型で致死となる組合せの場合の数は 5 通り考えられる。また接合体では遺伝子型が 3 通り 2 遺伝子座の遺伝型で致死となる組合せの場合の数は単純な致死モデルでは 35 通りだが、ヘテロ型を含む遺伝型組合せがホモ型同士の組合せの半分の致死率となる場合を考慮すると、致死となる組合せの場合の数は 161 通りとなる。F₂ 集団の独立性検定にて観測されたカイ 2 乗値のピークが配偶体及び接合体の合計 166 個のいずれかの場合に、浸透率を考慮すれば当てはまるものがあるかどうかを検定する手法の開発は次年度行う予定である。

また、上述の自殖後代 (F₂) 186 個体によるデータ解析 (自由度 4 の独立性検定) の結果を踏まえ、次のような追加実験を行った。F₂ と同じ組合せ親の F₁ を雌性親に用いた戻し交雑集団 (BCF) と F₁ を雄性親に用いた戻し交雑集団 (BCM) をそれぞれ 236 個体を用い 160 個のマ-

カーで連鎖地図を作成し、マーカー遺伝子型分離比の分離の自由度 1 の独立性検定を行った。検定統計量の等高線を図 1 に示す。なお 160 個のマーカーは、F₂ 186 個体の解析で用いたマーカーのサブセットである。最初の実験で F₂ 集団で配偶体で相互作用があるモデルと一致した第 1 染色体と第 1 染色体間のピークのカイ 2 乗値は 39.01 であり、非線形再生理論を用いた多重性調整 p 値は 2.88 × 10⁻⁶ であった。またこのマーカーの組み合わせは一方の集団の BCM でのみ観測され、雄性配偶体での相互作用による生殖隔離と結論づけることができた。

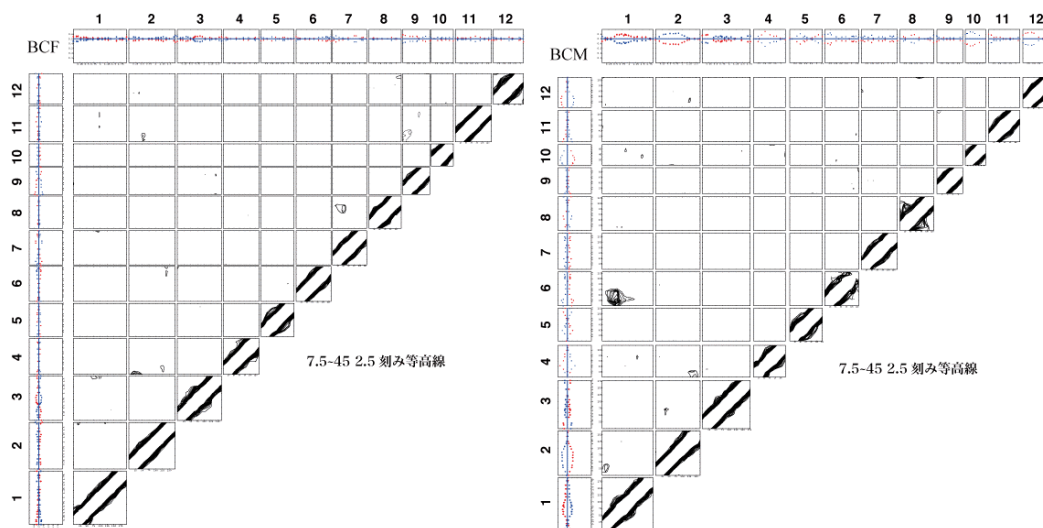


図 1 BCF 集団における独立性。カイ 2 乗統計量

課題 e. QTL 解析法の最適化 (栗木, 江口, 藤澤, 坂口, Mollah, Dou, 城石, 前野, 高田, 岡, 小出, 杉本, 原, 二宮)

e1. 課題「ロバスト QTL」

ロバスト QTL についても特に標準的に広く使われている複合区間マッピング (CIM) の尤度に基づく解析法のロバスト化について研究した。提案する方法はベータべきダイバージェンスに基づく方法で、従来の尤度関数に代わるベータべき尤度関数を以下のように定義し、

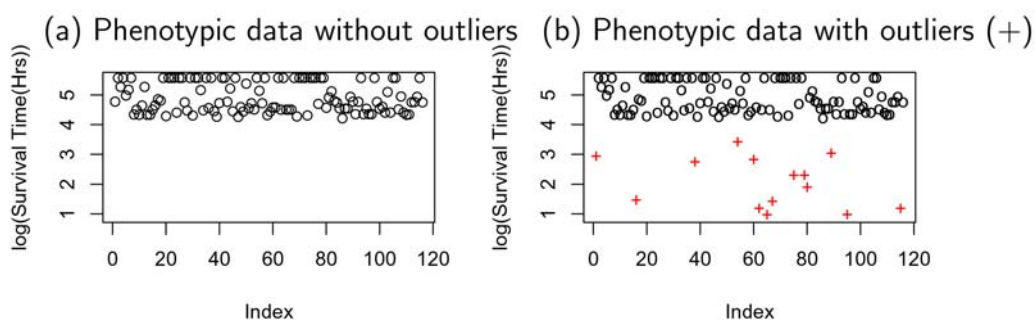
$$L_{\beta}(\theta|Y, X) = \frac{1}{\beta} \left[\frac{1}{nl_{\beta}(\theta, X)} \sum_{j=1}^n \{f(y_j|\theta, X_j)\}^{\beta} - 1 \right]$$

検定統計量を次のように定義するものである。

$$LOD_{\beta} = 2n \left\{ \sup_{\theta} L_{\beta}(\theta|Y, X) - \sup_{\theta_0} L_{\beta}(\theta|Y, X) \right\}$$

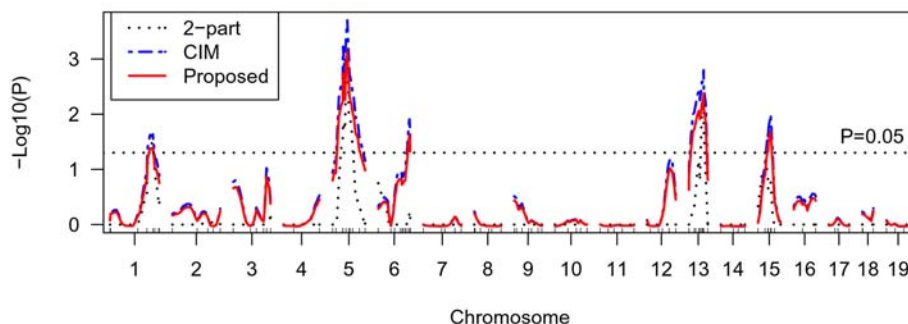
このべき指数 β が 0 になると従来法の対数尤度法に帰着されることに注意する。

ここでは試験的に公開されたデータについて従来法との比較を行った。量的形質はリステリア・モノサイトゲネス病原微生物に感染した F₂ マウスの 116 個体の対数生存時間である [図 (a) 参照]。この対数生存時間にたいして人為的に図 (b) 赤でプロットされるように 14 個体に対して外れ値を混入した。

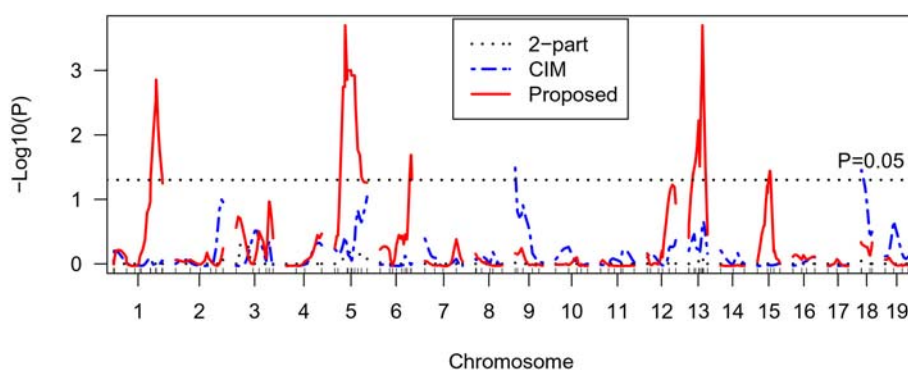


この2つのデータセットに対して共通の遺伝子形のデータに基づいて従来法と提案法の解析結果を比較した。外れ値がない公開データで解析した結果は図(e)となり、人工的に外れ値を混入した解析結果は図(f)となった。ここで従来法のロッド曲線を青線で表し、提案法のロッド曲線を赤線で表した。

(e) QTL mapping in absence of outliers.



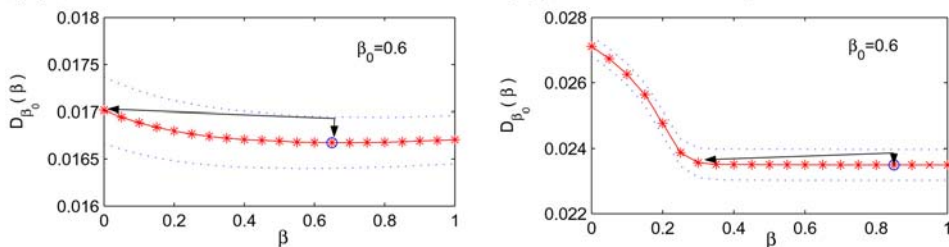
(f) QTL mapping in presence of outliers.



このように外れ値が無いときは従来の CIM 法と提案された CIM 法のロッド曲線はほとんど重なっている。しかし、外れ値が混入されたときは従来の CIM 法は大きく変化して異なる染色体にピークを示唆している。一方で提案された CIM 法は外れ値の影響をほとんど受けてない。このように提案手法がうまくいった理由はべき指数 β を K 重クロス・バリデーション法を援用してデータに適応的に次のように選択したことによる。

実際、データ (a) に対してはべき指数 β が 0 と選択されることから、定義から従来法に等価

(c) β -selection in absence of outliers (d) β -selection in presence of outliers



となる。一方で、データ (b) に対しては 0.3 と選択されているので従来法とかけ離れた解析法を行っていることになる。

結論としてこのデータ解析においては従来の方法は外れ値の影響を大きく受けることがあるが提案された手法は非常に安定した性能を示すことが示された。今後、城石研究グループで得られた QTL データについても提案手法の解析を加えて外れ値の探索について検討することが課題である。また提案された方法から影響解析についても新たな方法が検討する課題が残っている。

e2. 課題「QTL 解析の影響分析」

[研究目的]

個体の形質を規定する遺伝子 QTL を探索するための QTL 解析では、ロッドスコア (尤度関数) の極大点を探索することによって QTL の位置を推測する。しかし、同じ染色体に 2 つ以上のロッドスコアの極大値が存在するときに、その 2 つの極大値が 2 つの QTL がもたらすものか、あるいは単なる統計誤差によるものかどうかは遺伝学的に重要であり、慎重に吟味する必要がある。そのためには、ロッドスコアの形状に影響を与える個体を検出し、その個体のデータが信頼できるものか、何らかの誤りをふくむものかどうかを再検討する必要がある。この目的のため従来は、ロッドスコアの極大点付近で遺伝子型が変化するような個体の洗い出しを行っていた。しかしながら、この方法は非効率であるばかりでなく、表現型データを参照していないという意味で不十分な方法である。本研究ではロッドスコアの形状に影響する個体を系統的に検出する方法の開発を目的としている。

影響分析の方法としては Cook (1996) の局所影響関数の方法がよく知られている。これは、モデルの適合度に影響を与えるパラメータの摂動方向を探索するものであるが、QTL 解析のようにモデルの当てはめによらず統計的検定を解析の目的とする場合はその考え方を直接適用できない。ここでは、複数の遺伝子座におけるロッドスコアの値 (ベクトル値) に対する標本影響関数行列 (Li, et al. 1997; Tanaka, 1999) を求め、主成分分析と類似の考え方によってロッドスコアに影響を与える個体のグループを検出する方法を考える。

[平成 20 年度の進捗]

マウスの肥満の原因となる遺伝子を探索するため、肥満と関連性が深いとされる血中アディポネクチンに着目し、標準的マウス近交系である B6 と、日本産亜種由来の MSM 系統の F2 雑種

4170 個体に対して解析を行う。各マウスの 20 対の染色体における 119 箇所の遺伝子型が三つの値 (A, H, B) として観測されている。ここでは、血中アディポネクチンの対数を目的変数とし、遺伝子型の加法効果、優性効果ならび性別を変数とする線形モデルを仮定し、QTL 解析 (単一マーカー分析) を行う。得られたロッドスコアを図 1 に示す。

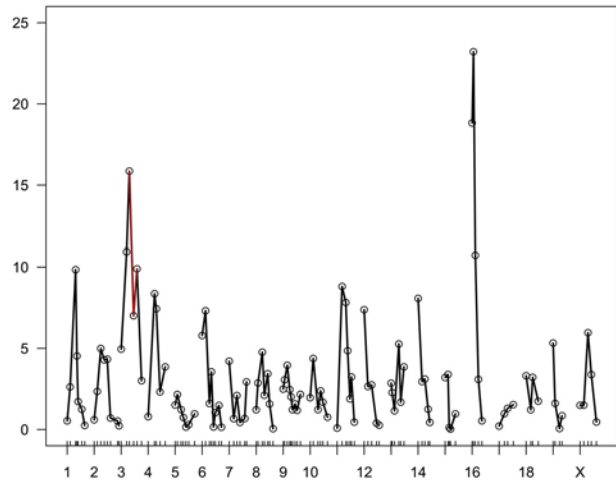


図 1 ロッドスコア

図 1 において、第 3 染色体上にロッドスコアの極大値が二つ存在することが見て取れる (双峰)。このような曲線の形が、少数

の外れ値によるものかどうかを調べたい。そのために、複数の遺伝子座におけるロッドスコアの標本影響関数を求め、ロッドスコア曲線の形に影響を与える個体のグループを検出する。

第 i 座のロッドスコアの $\frac{1}{n}$ 倍 (これを T とおく) は経験分布関数の汎関数であるので、その標本影響関数が定義できる。

$$\text{SIF}(F_n, x_s) = \frac{T \left((1 - \varepsilon) F_n + \varepsilon \delta_{x_s} \right) - T(F_n)}{\varepsilon} \Bigg|_{\varepsilon = -\frac{1}{n-1}}$$

$$= -(n-1)(T(F_n^{(s)}) - T(F_n))$$

ここで、 $F_n = \frac{1}{n} \sum_{t=1}^n \delta_{x_t}$ は表現型と座 i の遺伝子型データ $x_t = (y^{(t)}, z_i^{(t)})$, ($t = 1, \dots, n$) から計算される経験分布関数、 $F_n^{(s)} = \frac{1}{n-1} \sum_{t \neq s} \delta_{x_t}$ は s 番目の個体を除いた経験分布関数である。 T_i の標本影響関数を SIF_i とし、興味のある遺伝子座 $I = i_1, \dots, i_k$ に対してロッドスコア曲線の変動に影響する個体を検出する。ロッドスコアの、QTL がどこにも存在しないという帰無仮説の下の漸近共分散行列を Σ とおく (詳細略)。

ロッドスコアの形状の変動を考えるためには、 $k \times k$ 行列 $J = I_k - \frac{1}{1' \Sigma^{-1} 1} \mathbf{1} \mathbf{1}' \Sigma^{-1}$ を用いて平均の変動を除いて影響する個体を見つけ出す。ここでは、

$$(\text{SIF}_{i_1}, \dots, \text{SIF}_{i_k}) J' \{ J \Sigma J' \}^{-1} J (\text{SIF}_{i_1}, \dots, \text{SIF}_{i_k})' \quad (1)$$

に対して固有値解析を行う。このとき、行列 (1) の固有ベクトルはロッドスコア曲線の何らかの形状的な特性を意味するものと考えられ、固有ベクトルに大きな値をもつ個体はその特性に対する影響集合と判断できる。行列 (1) について固有値解析を行った結果を図 2 に示す。

図 2 における点線は $\text{SIF}_{(i_1)}, \dots, \text{SIF}_{(i_k)}$ を平均 0 の nk 次元正規分布に従うと考え、分散共分散行列の推定値を用いて生成された正規乱数に基づく 99% 信頼区間で

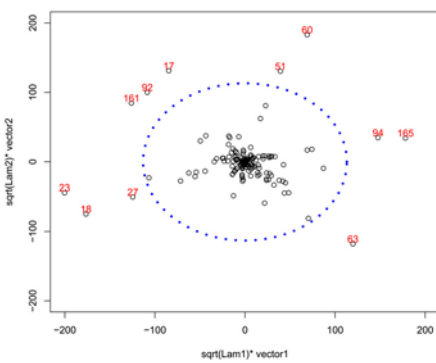


図 2 固有ベクトルの要素のペアのプロット

ある。図2においては、円の外にある番号のマウスが第3染色体のロッドスコア曲線の形に影響する個体として判断される。

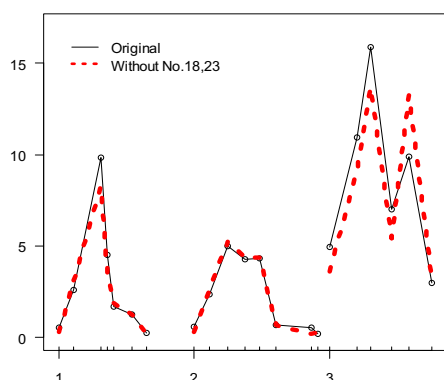


図3 個体No. 18, 23を除いて得られたロッドスコア (第1~3染色体)

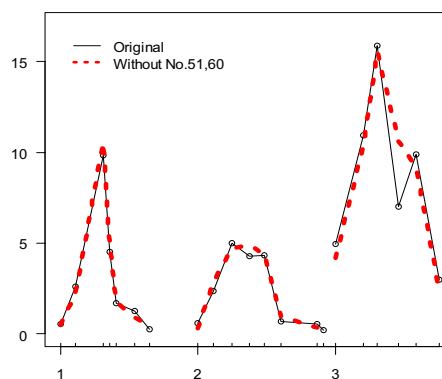


図4 個体No. 51, 60を除いて得られたロッドスコア (第1~3染色体)

図2において第1軸方向に大きな絶対値をとる個体18と23及び第2軸方向に大きな値をとる個体51と60を除いて得られたロッドスコア曲線はそれぞれ図3と図4のように示される。これらの図から図2の横軸(第1固有ベクトル)は全体の傾きを表し、縦軸(第2固有ベクトル)は凸性または谷の有無に対応することがわかる。

(まとめ) ロッドスコア曲線の形状に影響を与える個体の集合を特定するために、複数の遺伝子座での影響関数を用いる方法を提案した。この方法によってロッドスコア曲線の傾き、凸性及び平均移動などの特性に影響するそれぞれの個体集合を検出できた。

7. 平成20年度の研究成果

(1) 成果物(知見・成果物・知的財産権等)

サブテーマの報告書に記載した。

(2) 成果発表等

<論文発表>

[学術論文]

1. Ito, N. and Kurata, N. Disruption of KNOX gene suppression in leaf by introducing its cDNA in rice. *Plant Science*. 174: 279-289. 2008.
2. Thirumurugan, T., Ito, Y., Kubo, T., Serizawa, A. and Kurata, N. Identification, characterization and interaction of HAP family genes in rice. *Mol Gen Genomics*. 279: 279-289. 2008.
3. Suzuki, T., Eiguchi, M., Kumamaru T., Satoh, H., Matsusaka, H., Moriguchi, K. and Kurata, N. MNU-induced mutant pools and high performance TILLING enable finding of any gene mutation in rice. *Mol Gen Genomics*. 279:213-223. 2008.
4. Suwabe, K. Suzuki, G. Takahashi, H. Shiono, K. Endo, M. Yano, K. Fujita, M. Masuko, H. Saito, H.

- Fujioka, T. Kaneko, F. Kazama, T. Mizuta, Y. Separated transcriptomes of male gametophyte and tapetum in rice: validity of a laser microdissection (LM) microarray. *Plant Cell Physiol.*49:1407-1416. 2008.
5. Hobo, T. Suwabe, K. Aya, K. Suzuki, G. Yano, K. Ishimizu, T. Fujita, M. Kikuchi, S. Hamada, K. Miyano, M. Fujioka, T. Kaneko, F. Kazama, T. Mizuta, Y. Takahashi, H. Shiono, K. Nakazono, M. Tsutsumi, N. Nagamura, Y. Kurata, N. Watanabe, M. Matsuoka, M. Various spatiotemporal expression profiles of anther-expressed genes in rice. *Plant Cell Physiol.*49:1417-1428. 2008.
 7. Araki K, Takeda N, Yoshiki A, Obata Y, Nakagata N, Shiroishi T, Moriwaki K, Yamamura K. Establishment of germline-competent embryonic stem cell lines from the MSM/Ms strain. *Mamm Genome.* 2009 20(1): 14-20.
 8. Sakuraba Y, Kimura T, Masuya H, Noguchi H, Sezutsu H, Takahasi KR, Toyoda A, Fukumura R, Murata T, Sakaki Y, Yamamura M, Wakana S, Noda T, Shiroishi T, Gondo Y. Identification and characterization of new long conserved noncoding sequences in vertebrates. *Mamm Genome.* 2008 19(10-12): 703-712.
 9. Kaminuma E, Masuya H, Miura I, Motegi H, Takahasi KR, Nakazawa M, Matsui M, Gondo Y, Noda T, Shiroishi T, Wakana S, Toyoda T. Objective evaluation measures of genetic marker selection in large-scale SNP genotyping. *J Bioinform Comput Biol.* 2008 6(5): 905-17.
 10. Fujii T, Tamura M, Tanaka S, Kato Y, Yamamoto H, Mizushina Y, Shiroishi T. Gasdermin D (Gsdmd) is dispensable for mouse intestinal epithelium development. *Genesis.* 2008 46(8): 418-23.
 11. Liu YH, Takahashi A, Kitano T, Koide T, Shiroishi T, Moriwaki K, Saitou N. Mosaic genealogy of the *Mus musculus* genome revealed by 21 nuclear genes from its three subspecies. *Genes Genet Syst.* 2008 83(1): 77-88.
 12. Tanaka, E., Tamura, Y., Hosoya, M. and Shiroishi, T., "Protrusion Fourier Descriptor: Skeleton-based Representation of Open Curves", *Forma*, Vol. 23, 9-18, 2008.
 13. Tanaka, E. and Tamura, Y., "Skeleton-based Fourier Descriptor of Open Curves", *Pacific Science Review*, Vol. 10, No. 10, 194-198, 2008.
 14. Takahashi A., Shiroishi T., Koide T.: Multigenic factors associated with a hydrocephalus-like phenotype found in inter-subspecific consomic mouse strains. *Mammalian Genome*, 29:333-338, 2008.
 15. Takahashi A., Nishi A., Ishii A., Shiroishi T., Koide T.: Systematic analysis of emotionality in consomic mouse strains established from C57BL/6J and wild-derived MSM/Ms. *Genes, Brain and Behavior.* 7:849-858, 2008.
 16. Kawakita, M., Eguchi, S., Boosting method for local learning in statistical pattern recognition. *Neural Computation* 20, 2792-2838 (Nov., 2008)
 17. Fujisawa, H., Eguchi, S., Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis* 99, 2053-2081 (Oct., 2008)

18. Takenouchi, T., Eguchi, S., Murata, N., Kanamori, T., Robust Boosting Algorithm Against Mislabeling in Multiclass Problems. *Neural Computation* 20, 1596-1630 (June, 2008)

〔会議録〕

1. 渡辺正夫, 倉田のり 国立遺伝学研究所研究会「高等植物の生殖過程を制御する因子の多様性と生殖隔離」, 国立遺伝学研究所, 2008年11月5日～6日, 三島.
2. 長戸康郎, 倉田のり 国立遺伝学研究所研究会「イネ分子遺伝学の展開」, 国立遺伝学研究所, 2008年12月19～20日, 三島
3. Mollah, Md. Nurul Haque, Eguchi, Shinto Robust Composite Interval Mapping for QTL Analysis by Minimum β -Divergence Method. *IEEE International Conference on Bioinformatics and Biomedicine 2008*, 115-120, 978-0-7695-3452-7, 2008.11

〔解説・総説〕

1. Koide T, Catanesi CI, Nishi A, Shiroishi T, Kasai S, Ikeda K, Takahashi A. Systematic mapping of pain-related QTL using consomic mouse strains: Advantage of using wild-derived strains. *Brain Res J.* in press.
2. 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, II, 小西貞則, 国友直人(編), 東京大学出版会, 東京, 315-356, 2008.7

<会議発表等>

〔招待講演〕

1. 城石俊彦: マウス亜種間コンソミック系統によるゲノム機能解析. 第44回高血圧関連疾患モデル学会学術総会, 2008年11月22日, 出雲.
2. 城石俊彦: 糖尿病研究における forward genetics アプローチ. 第1回疾患モデルシンポジウム, 2008年12月3日, 東京.
3. 江口真透. タンパク質構造と進化と情報幾何. 数理研短期共同研究集会「離散力学系の分子細胞生物学への応用数理」. 京都, 日本. 2009.1.8
4. Eguchi, Shinto. Information divergence geometry and its application to machine learning. The 1st MSJ-SI, Probabilistic Approach to Geometry, Kyoto, Japan. 2008.8.4

〔一般講演〕

1. Horiuchi Y., Harushima Y., Mochizuki T., Fujisawa H., Eguchi., Kawakita M., Kurata N. Detection of nucleotide and expression polymorphisms between rice strains using affymetrix rice genome array. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
2. Harushima Y., Kuriki S., Mizuta Y., Kurata N. Detection of pairs of interactive reproductive barriers in rice. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
3. Mizuta Y., Harushima Y., Kurata N. Positional cloning of a pair of interactive genes causing reproductive barrier in the hybrid pollen of rice. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.

4. Harushima Y., Yano M., Kurata N. Identification of areproductive barrier working in the process of pollen competition in rice. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
5. Tsuda K., Ito Y., Miyao A., Hirochika H., Kurata N. Identification and analysis of rice mutants misexpressing NKOX genes in leaves. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
6. Takada, T., Ebata, T., Shin-I, T., Narita, T., Abe, K., Sakaki, Y., Toyoda, A., Obata, Y., Moriwaki, K., Kohara, Y., Shiroishi, T. Two wave intersubspecific introgression built up genome framework of the classical laboratory mouse strains. 22th International Mammalian Genome Conference, 2008. 11.2-5, Prague, Czech Republic.
7. Takada, T., Mita, A., Maeno, A., Moriwaki, K., Yonekawa, H., Shiroishi, T. Mouse inter-subspecific consomic strains uncovers additive and non-additive genetic effects on complex traits. 22th International Mammalian Genome Conference, 2008. 11.2-5, Prague, Czech Republic.
8. Takada, T., Mita, A., Moriwaki, K., Yonekawa, H., Shiroishi, T. Mouse inter-subspecific consomic strains for the study of energy metabolism-related traits. CBI Annual Meeting 2008 International Symposium, 2008. 10.22-24, Tokyo, Japan.
9. Oka, A., Takagi, N., Moriwaki, K., and Shiroishi, T. Genetic Study of the Reproductive Isolation between Two Mouse Subspecies, *Mus musculus domesticus* and *M. m. molossinus* XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
10. Tanaka, E. and Tamura, Y. "Skeleton-based Fourier Descriptor of Open Curves" APCOM2008, Kokushikan University, September 3, 2008
11. Takahashi, A., Sugimoto, H., Kimura, S., Tomihara, K., Tsuchiya, T., Kakihara, S., Tanemura, M., Shiroishi, T., Koide, T.: Complex genetic architecture of social interaction and aggressive behavior clarified using consomic strains derived from MSM and C57BL/6. 22nd International Mammalian Genome Conference, Prague, November 2-5, 2008.
12. Takahashi, A., Tomihara, K., Shiroishi, T., Koide, T.: Genetic analysis of inter-male aggression using consomic mouse strains established from C57BL/6J and MSM. 10th Annual Meeting of the International Behavioural and Neural Genetics Society, Portland, May 5-9, 2008.
13. Kuriki, S., Harushima, Y., Fujisawa, H., and Kurata, N., Multiplicity Adjustments in Detecting Reproductive Barriers caused by Loci Interactions, BIRS Workshop 09w5040 Random Fields and Stochastic Geometry, 2009.2.26, Banff, Canada
14. Md. Nurul H.M., Eguch, S., Robust QTL Analysis by the Minimum β -Divergence Method. International Association for Statistical Computing 2008, Yokohama, 2008.12.6
15. Pritchard, M., Eguchi, S. Finding Optimal Gene Set for Classification from Multiple Predictive Gene Sets. International Association for Statistical Computing, Yokohama, 2008.12.6
16. 米田典央, 倉田のり, 野々村賢一「減数分裂第一分裂前期に特異的な染色体挙動の観察」イ

- ネ遺伝学・分子生物学ワークショップ 2008, 九州大学, 2008年7月5日, 福岡.
17. 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障壁となる重複遺伝子の解析」日本育種学会 114 回講演会, 滋賀県立大学, 2008年10月11日, 彦根.
 18. 津田勝利, 伊藤幸博, 宮尾安藝雄, 廣近洋彦, 倉田のり「イネのシュート形成における極長鎖脂肪酸 (LCFA) の機能解析」日本育種学会 114 回講演会, 滋賀県立大学, 2008年10月11日, 彦根.
 19. 久保貴彦, 水多陽子, 新濱充, 春島嘉章, 倉田のり「イネ生殖隔離機構の解析から見えてくるもの」国立遺伝学研究所研究会「高等植物の生殖過程を制御する因子の多様性と生殖隔離」, 国立遺伝学研究所, 2008年11月5日, 三島.
 20. 望月孝子, 菊池俊介, 濱田和輝, 加藤大貴, 大木伸彦, 藤田雅丈, 堀内陽子, 倉田のり, 矢野健太郎「OryzeExpress: イネのゲノム・アノテーションとオミックス統合データベース」第31回日本分子生物学会年会, 第81回日本生化学会大会 合同大会, 神戸ポートアイランド, 2008年12月10日, 神戸.
 21. 津田勝利, 伊藤幸博, 倉田のり「イネにおける KNOX 遺伝子を介した SAM の維持および葉の分化の研究」国立遺伝学研究所研究会「イネ分子遺伝学の新展開」, 国立遺伝学研究所, 2008年12月20日, 三島.
 22. 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を用いた肥満関連表現型の遺伝解析. 日本遺伝学会, 2008, 9.3-5, 名古屋.
 23. 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を用いたエネルギー代謝関連形質の遺伝解析. 第55回日本実験動物学会総会, 2008, 5.15-17, 仙台.
 24. 岡彩子, 高田幸, 古関明彦, 森脇和郎, 城石俊彦: マウス亜種間における生殖隔離と減数分裂期のチェックポイント機構との関連. 日本遺伝学会, 2008, 9.3-5, 名古屋.
 25. 田中英希, 田村義保「新しい開曲線記述子を用いた2次元部分形状の定量的評価」日本計算機統計学会 第22回大会, 秋田文化会館, 平成20年5月23日
 26. 江口真透. バイオインフォマティクスにおける統計的課題について. 科研費研究集会「高次元データの統計解析」博多. 2008.11.21
小森理, 江口真透. 1クラスラベルに注目したブースティング. 統計関連学会連合大会, 横浜. 2008.9.8
 27. プリチャード真理, 江口真透. マイクロアレイにおける遺伝子選択と判別能力の関係. 統計関連学会連合大会, 横浜. 2008.9.9

<著書等>

1. Koide T, Catanesi CI, Nishi A, Shiroishi T, Kasai S, Ikeda K, Takahashi A. Advantage of using wild-derived mouse strains for a variety of pain-related studies: Genetic diversity and new genetic tools. In: Acute Pain (Columbus F, ed), Nova Science Publishers, New York, in press.

2. Eguchi, S. Information Divergence Geometry and the Application to Statistical Machine Learning. Eds. Frank Emmert-Streib and Matthias Dehmer 'Information Theory and Statistical Learning', Springer, New York, 10.1007/978-0-387-84816-7_13

プロジェクト名： 生物多様性解析

サブテーマ名： 「表現型数値計測システムの開発と実験データの収集」

研究代表者： 城石俊彦 [国立遺伝学研究所]

1. 研究目標

生命システムの理解には、生物多様性を客観的に記載する数値計測技術と得られた表現型としての計測数値をゲノム多型に関連づけるための統計解析手法の融合が必要である。そこで、国立遺伝学研究所が保有する遺伝的多様性に富んだマウスやイネ等の多数のモデル生物系統の多様性を客観的に評価するために、表現形質の数値計測システムの開発を行う。さらに計測された数値データを統計解析に適用するため、生物系統や遺伝的交配個体からの表現型データ収集においても拡充する。イネ・マウスの形態多様性、脂肪分布などの生体内部の形質についての多様性データは特に重視する。また、マイクロアレイデータからの SNP 推定プログラムや推定された SNP に配慮した遺伝子発現解析のための統計解析システムについてもその完成を目指す。

この研究目標の達成のために、本年度は次の4つの課題を設定した。

課題 a. 3D 画像による体脂肪計測法の開発とデータマイニング

課題 b. マウスエネルギー代謝関連表現型計測法の開発と実験データの収集

課題 c. イネストレス耐性・穂形質の系統データ収集と遺伝子発現データとの相関解析法の開発

課題 d. マイクロアレイによるイネ・マウス遺伝子発現データの収集と解析

2. 年度研究計画

平成16年度（予備研究）

統数研と遺伝研の二研究所間で融合研究の実施内容について数回の打ち合わせを行い協議した。マウス・イネの表現型、遺伝子間相互作用の数理解析法について具体的な研究方針の検討を行った。イネのマイクロアレイによる遺伝子発現解析と生殖的隔離障壁の研究について協議した。マウス形態多様性の数理解析を目的として、国立遺伝学研究所においてマウス下顎骨の骨格標本を野生マウス由来系統から作製した。また、これらの標本を使って統計数理研究所が開発した P 型フーリエ記述子を用いた形態の数理的計測を行った。さらに主成分分析によって、形態多様性に係わる成分解析を行った。

平成17年度

形態多様性解析として、マウス下顎骨の画像データについて P 型フーリエ記述子を用いた形態数値化とそれを用いた主成分分析が系統間の多様性解析に有効であることを示した。X 線 CT 値による分析において、マウス内蔵脂肪と皮下脂肪を自動的に判別して各脂肪量を定量化するためのソフトウェアの開発に着手した。マウス行動パターンの内、社会行動と自発活動の日周期変動について、客観的な数値計測化と統計モデルによるシミュレーションを行った。遺伝的距離の大

きな生物系統において、一方の系統のゲノム DNA のプローブセットを用いたマイクロアレイの統計解析についての検討を行い、SNP 由来の見せかけのシグナル強度を判定するための方法論を検討した。

平成18年度

課題 a. マウス皮下・内臓脂肪を自動測定し数値化するシステムを構築するため、(B6 x JF1)F2 世代の複数個体の CT 画像データの取得と、それらを用いた各組織の CT 値の画像毎のばらつき抽出と、皮下・内臓脂肪を自動測定するため腹筋線の特徴点選定を行った。

課題 b. マウス肥満に関連する表現型に関する責任遺伝子座を検出するため、(MSM x B6)F2 交配世代を用いた QTL 解析を行った。また、新規に (JF1 x B6) F2 交配世代個体について、エネルギー代謝関連表現型を中心としたパラメータの計測を行った。

課題 c. イネについて、野生イネ・コアコレクション 46 系統で穂形質（穂型、穎花長、葯長、粒重、1 穂穎花数、稔性）を解析した。各形質の相関関係の検定を行い、今後のアソシエーション解析に適した形質と系統の検討を行った。

課題 d. イネとマウスの二つの生物種において、SNP に配慮したマイクロアレイによる遺伝子発現の統計解析手法開発のための実験データを生産した。また、アフィメトリクス社製 GeneChip を用いて、シグナル強度の比較から SNP を検出するために開発したプログラムを検証するため、このプログラムにより検出された SNP と、塩基配列比較から検出された SNP（すでに DB 化した）の比較検討を行った。

平成19年度

課題 a. マウス F2(JF1/B6)個体の BMI により分類した多様な CT 画像データを脂肪量自動抽出に必要な腹筋線検出のための各種モデルに当てはめ、最適となるモデルの選定を行う。マウスの全身 CT 画像から腹筋線および 2 種類の脂肪組織（内臓脂肪と皮下脂肪）の自動抽出を行う画像処理プログラム (ActScan) を作成した。

課題 b. JF1-B6 系統間のエネルギー代謝関連表現型差異を規定する遺伝的要因を探索するため、両系統より作出した F2 世代の当該表現型を中心としたパラメータの計測を充実させる。さらに全ゲノムスキャンによる各種表現型を規定する責任領域の抽出を行う。

課題 c. 野生イネ系統の集団を用いて穂形質特性についての集団内系統別形質データを収集し、さらに系統別の遺伝子発現をマイクロアレイデータとして取得した。特に種内で形質が大きく 2 つのグループに分かれた *O. rufipogon* に target を絞って、形質データを収集した。

課題 d. (i) イネゲノム SNP とハイブリシグナル強度差との相関解析の結果をもとに、SNP 推奨プログラムの最適化を行い、論文を作成した。

(ii) 日本産野生由来近交系統のゲノム情報から SNP により影響を受けているプローブの抽出をさらに推し進める。B6 系統、日本産野生由来近交系統の発現データを使用し、プローブのシグナル値から SNP effect を検出するアルゴリズムによる遺伝子発現差異

の検証を行う。マウス SNP 推定アルゴリズムにより、B6-MSM 系統間の SNP 情報とシグナル強度の関連を検証し、論文を作成した。

平成20年度

課題 a. マウス脂肪領域判定について、画像のピクセル間に定義したネットワーク形状の解析に基づく ActCuts 法を開発した。ActScan 法と比較して皮下脂肪と内臓脂肪の判定に関して精度向上を確認した。

課題 b. JF1 系統と B6 系統の F2 世代個体を用いたエネルギー代謝関連形質の QTL 解析を行い、全ゲノムの 30 か所以上に高い LOD 値 (2.5 以上) を示す QTL を検出した。

課題 c. 野生イネ 46 系統で穂形質情報を収集した。野生イネ O. rufipogon 由来系統の形質情報調査・整備した。また、遺伝子多型、発現多型解析用サンプルを収集し、マイクロアレイデータを一部測定した。

課題 d. 前年度に開発した発現及び塩基多型の同時検出法 (SNEP: Simultaneous detection of nucleotide and expression polymorphisms) を改良した結果、検出率や擬陽性などのパフォーマンスが改善され、platform strain でない場合にも対応可能となった。イネの genome DNA のアレイへのハイブリを用いた塩基多型検出頻度の最適化を行った。mRNA のハイブリで SNP を含むプローブについて、SNP によるシグナル減少量の推定を様々な観測量の重回帰分析量から推定することを試みた。

平成21年度

課題 a. 腹筋線自動抽出アルゴリズムを実装したソフトウェアの完成と公開を行う。完成版ソフトウェアを使用した各種脂肪組織の測定とデータマイニングを行い、その有効性を検討する。

課題 b. MSM-B6 系統ならびに JF1-B6 系統間のエネルギー代謝関連表現型を規定する原因遺伝子、エピスタシスに関与する遺伝子群の探索を行う。有力な候補遺伝子についてはノックアウトマウスやトランスジェニックマウス作製などの逆遺伝学的実験手法による検証を行う。

課題 c. 野生イネ複数系統の発現データをマイクロアレイにより取得、前年度までに課題 d にて開発したシステムにて解析可能かどうか検討する。さらに、これらの結果に基づいて構造化されていない自然集団における e-QTL 及び形質データの QTL 解析の具体的な問題点の洗い出しを行う。

課題 d. SNP が多く含まれる遺伝子群の発現量の推定法を開発し、他の解析方法 (RT-PCR 法など) により検証する。野生イネの一部系統については、発現遺伝子およびゲノム塩基配列について、次世代シーケンサーによる読み取りと解析を行い、より簡便かつ確実な統計的方法論の導入について考察する。

平成22年度以降の展開

課題 a. 開発した体脂肪測定法の検証と実測定への応用を行うため、多数のマウス系統を用いた

脂肪蓄積量の測定を実施し、解剖により計測した実測値と比較する。また、本課題により開発したソフトウェアによる C57BL/6J やコンソミック系統など特定系統の脂肪蓄積の時系列データを収集する。また、脂肪蓄積測定については、外部研究機関とデータを共有して解析するための共同研究ネットワークを構築する。

課題 b. MSM/JF1-B6 系統間のエネルギー代謝関連表現型を規定する原因遺伝子群、さらにはエピスタシスに關与する遺伝因子群を同定する。本課題により明らかにした関連エネルギー代謝表現型について広くマウス系統を調査し、当該遺伝子の多型と表現型の関連について解析する。対応するヒト相同遺伝子が肥満や脂肪蓄積の責任遺伝子となっているかどうかを外部の医学系研究機関との共同研究によって明らかにする。

課題 c. 複数の系統間の SNP、遺伝子発現量、形質多型について、相互の関連をアソシエーション解析によってどの程度の信頼性で特定できるかは、それぞれのデータの信頼度と系統の選択の仕方によってかなり異なると思われる。よって、この研究で得られた結果を基に、さらにどのような遺伝的および統計的改良が可能であるかを、次のステップでいくつかの例として示す必要がある。

課題 d. 我々の開発した手法により、SNP を考慮した遺伝子発現解析をマイクロアレイにより行うことが可能になりつつある。今後は、次世代シーケンサーの利用も含め、同種の異なる系統間の遺伝子発現多様性と形質多様性の相関解析についてゲノムシャッフル系統群等を利用して一層推進する。さらに、一連の解析から得られる豊富な e-QTL 情報を利用することにより、エピスタシスや遺伝子発現ネットワークの解析への新たな展開を図る。

3. 研究経費の推移

平成 17 年度実績 (サブテーマ 1 と 2 の総計) : 157,820 千円

平成 18 年度実績 : 115,370 千円

平成 19 年度実績 : 94,210 千円

平成 20 年度実績 : 111,750 千円

平成 21 年度見込 : 107,180 千円

4. 平成 20 年度の研究実施体制

[国立遺伝学研究所] 倉田のり 春島嘉章 田村 勝 高田豊行 前野哲輝 岡(木曾)彩子
久保貴彦 永口 貢 堀内陽子 小出 剛

[国立情報学研究所] 北本朝展 藤山秋佐夫 佐藤真一

[統計数理研究所] 江口真透 池田思朗 藤澤洋徳 川崎能典 坂口隆之

Nurul Haque Mollah

[東京大学 大学院新領域創成科学研究科 情報生命科学専攻] 中谷明弘 廣田晃士

[長浜バイオ大学 バイオサイエンス学科 生命情報科学コース] 阿部貴志

5. 平成20年度研究成果

(1) 成果物 (知見・成果物・知的財産権等)

- 1) 皮下脂肪・内蔵脂肪を分離する新規アルゴリズム ActCuts。
- 2) イネ穂形成・多年生形質の系統別データ
- 3) 遺伝子発現・SNP 同時検出用のプログラム SNEP の改良版

(2) 成果発表等

<論文発表>

[学術論文]

1. Ito, N. and Kurata, N. Disruption of KNOX gene suppression in leaf by introducing its cDNA in rice. *Plant Science*. 174: 279-289. 2008.
2. Thirumurugan, T., Ito, Y., Kubo, T., Serizawa, A. and Kurata, N. Identification, characterization and interaction of HAP family genes in rice. *Mol Gen Genomics*. 279: 279-289. 2008.
3. Suzuki, T., Eiguchi, M., Kumamaru T., Satoh, H., Matsusaka, H., Moriguchi, K. and Kurata, N. MNU-induced mutant pools and high performance TILLING enable finding of any gene mutation in rice. *Mol Gen Genomics*. 279:213-223. 2008.
4. Suwabe, K. Suzuki, G. Takahashi, H. Shiono, K. Endo, M. Yano, K. Fujita, M. Masuko, H. Saito, H. Fujioka, T. Kaneko, F. Kazama, T. Mizuta, Y. Separated transcriptomes of male gametophyte and tapetum in rice: validity of a laser microdissection (LM) microarray. *Plant Cell Physiol*. 49:1407-1416. 2008.
5. Hobo, T. Suwabe, K. Aya, K. Suzuki, G. Yano, K. Ishimizu, T. Fujita, M. Kikuchi, S. Hamada, K. Miyano, M. Fujioka, T. Kaneko, F. Kazama, T. Mizuta, Y. Takahashi, H. Shiono, K. Nakazono, M. Tsutsumi, N. Nagamura, Y. Kurata, N. Watanabe, M. Matsuoka, M. Various spatiotemporal expression profiles of anther-expressed genes in rice. *Plant Cell Physiol*. 49:1417-1428. 2008.
7. Araki K, Takeda N, Yoshiki A, Obata Y, Nakagata N, Shiroishi T, Moriwaki K, Yamamura K. Establishment of germline-competent embryonic stem cell lines from the MSM/Ms strain. *Mamm Genome*. 2009 20(1): 14-20.
8. Sakuraba Y, Kimura T, Masuya H, Noguchi H, Sezutsu H, Takahasi KR, Toyoda A, Fukumura R, Murata T, Sakaki Y, Yamamura M, Wakana S, Noda T, Shiroishi T, Gondo Y. Identification and characterization of new long conserved noncoding sequences in vertebrates. *Mamm Genome*. 2008 19(10-12): 703-712.
9. Kaminuma E, Masuya H, Miura I, Motegi H, Takahasi KR, Nakazawa M, Matsui M, Gondo Y, Noda T, Shiroishi T, Wakana S, Toyoda T. Objective evaluation measures of genetic marker selection in large-scale SNP genotyping. *J Bioinform Comput Biol*. 2008 6(5): 905-17.
10. Fujii T, Tamura M, Tanaka S, Kato Y, Yamamoto H, Mizushina Y, Shiroishi T. Gasdermin D (Gsdmd) is dispensable for mouse intestinal epithelium development. *Genesis*. 2008 46(8): 418-23.

11. Liu YH, Takahashi A, Kitano T, Koide T, Shiroishi T, Moriwaki K, Saitou N. Mosaic genealogy of the *Mus musculus* genome revealed by 21 nuclear genes from its three subspecies. *Genes Genet Syst.* 2008 83(1): 77-88.
12. Tanaka, E., Tamura, Y., Hosoya, M. and Shiroishi, T., "Protrusion Fourier Descriptor: Skeleton-based Representation of Open Curves", *Forma*, Vol. 23, 9-18, 2008.
13. Tanaka, E. and Tamura, Y., "Skeleton-based Fourier Descriptor of Open Curves", *Pacific Science Review*, Vol. 10, No. 10, 194-198, 2008.
14. Takahashi A., Shiroishi T., Koide T.: Multigenic factors associated with a hydrocephalus-like phenotype found in inter-subspecific consomic mouse strains. *Mammalian Genome*, 29:333-338, 2008.
15. Takahashi A., Nishi A., Ishii A., Shiroishi T., Koide T.: Systematic analysis of emotionality in consomic mouse strains established from C57BL/6J and wild-derived MSM/Ms. *Genes, Brain and Behavior*. 7:849-858, 2008.
16. Kawakita, M., Eguchi, S., Boosting method for local learning in statistical pattern recognition. *Neural Computation* 20, 2792-2838 (Nov., 2008)
17. Fujisawa, H., Eguchi, S., Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis* 99, 2053-2081 (Oct., 2008)
18. Takenouchi, T., Eguchi, S., Murata, N., Kanamori, T., Robust Boosting Algorithm Against Mislabeling in Multiclass Problems. *Neural Computation* 20, 1596-1630 (June, 2008)

[会議録]

1. 渡辺正夫, 倉田のり 国立遺伝学研究所研究会「高等植物の生殖過程を制御する因子の多様性と生殖隔離」, 国立遺伝学研究所, 2008年11月5日～6日, 三島.
2. 長戸康郎, 倉田のり 国立遺伝学研究所研究会「イネ分子遺伝学の新展開」, 国立遺伝学研究所, 2008年12月19～20日, 三島
3. Mollah, Md. Nurul Haque, Eguchi, Shinto Robust Composite Interval Mapping for QTL Analysis by Minimum β -Divergence Method. *IEEE International Conference on Bioinformatics and Biomedicine 2008*, 115-120, 978-0-7695-3452-7, 2008.11

[解説・総説]

1. Koide T, Catanesi CI, Nishi A, Shiroishi T, Kasai S, Ikeda K, Takahashi A. Systematic mapping of pain-related QTL using consomic mouse strains: Advantage of using wild-derived strains. *Brain Res J.* in press.
2. 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, II, 小西貞則, 国友直人(編), 東京大学出版会, 東京, 315-356, 2008.7

<会議発表等>

[招待講演]

1. 城石俊彦：マウス垂種間コンソミック系統によるゲノム機能解析. 第 44 回高血圧関連疾患モデル学会学術総会, 2008 年 11 月 22 日, 出雲.
2. 城石俊彦:糖尿病研究における forward genetics アプローチ. 第 1 回疾患モデルシンポジウム, 2008 年 12 月 3 日, 東京.
3. 江口真透. タンパク質構造と進化と情報幾何. 数理研短期共同研究集会「離散力学系の分子細胞生物学への応用数理」. 京都, 日本. 2009.1.8
4. Eguchi, Shinto. Information divergence geometry and its application to machine learning. The 1st MSJ-SI, Probabilistic Approach to Geometry, Kyoto, Japan.2008.8.4

〔一般講演〕

1. Horiuchi Y., Harushima Y., Mochizuki T., Fujisawa H., Eguchi., Kawakita M., Kurata N. Detection of nucleotide and expression polymorphisms between rice strains using affymetrix rice genome array. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
2. Harushima Y., Kuriki S., Mizuta Y., Kurata N. Detection of pairs of interactive reproductive barriers in rice. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
3. Mizuta Y., Harushima Y., Kurata N. Positional cloning of a pair of interactive genes causing reproductive barrier in the hybrid pollen of rice. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
4. Harushima Y., Yano M., Kurata N. Identification of a reproductive barrier working in the process of pollen competition in rice. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
5. Tsuda K., Ito Y., Miyao A., Hirochika H., Kurata N. Identification and analysis of rice mutants misexpressing NKOX genes in leaves. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
6. Takada, T., Ebata, T., Shin-I, T., Narita, T., Abe, K., Sakaki, Y., Toyoda, A., Obata, Y., Moriwaki, K., Kohara, Y., Shiroishi, T. Two wave intersubspecific introgression built up genome framework of the classical laboratory mouse strains. 22th International Mammalian Genome Conference, 2008. 11.2-5, Prague, Czech Republic.
7. Takada, T., Mita, A., Maeno, A., Moriwaki, K., Yonekawa, H., Shiroishi, T. Mouse inter-subspecific consomic strains uncovers additive and non-additive genetic effects on complex traits. 22th International Mammalian Genome Conference, 2008. 11.2-5, Prague, Czech Republic.
8. Takada, T., Mita, A., Moriwaki, K., Yonekawa, H., Shiroishi, T. Mouse inter-subspecific consomic strains for the study of energy metabolism-related traits. CBI Annual Meeting 2008 International Symposium, 2008. 10.22-24, Tokyo, Japan.
9. Oka, A., Takagi, N., Moriwaki, K., and Shiroishi, T. Genetic Study of the Reproductive Isolation between Two Mouse Subspecies, *Mus musculus domesticus* and *M. m. molossinus* XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.

10. Tanaka, E. and Tamura, Y. "Skeleton-based Fourier Descriptor of Open Curves" APCOM2008, Kokushikan University, September 3, 2008
11. Takahashi, A., Sugimoto, H., Kimura, S., Tomihara, K., Tsuchiya, T., Kakihara, S., Tanemura, M., Shiroishi, T., Koide, T.: Complex genetic architecture of social interaction and aggressive behavior clarified using consomic strains derived from MSM and C57BL/6. 22nd International Mammalian Genome Conference, Prague, November 2-5, 2008.
12. Takahashi, A., Tomihara, K., Shiroishi, T., Koide, T.: Genetic analysis of inter-male aggression using consomic mouse strains established from C57BL/6J and MSM. 10th Annual Meeting of the International Behavioural and Neural Genetics Society, Portland, May 5-9, 2008.
13. Kuriki, S., Harushima, Y., Fujisawa, H., and Kurata, N., Multiplicity Adjustments in Detecting Reproductive Barriers caused by Loci Interactions, BIRS Workshop 09w5040 Random Fields and Stochastic Geometry, 2009.2.26, Banff, Canada
14. Md. Nurul H.M., Eguchi, S., Robust QTL Analysis by the Minimum β -Divergence Method. International Association for Statistical Computing 2008, Yokohama, 2008.12.6
15. Pritchard, M., Eguchi, S. Finding Optimal Gene Set for Classification from Multiple Predictive Gene Sets. International Association for Statistical Computing, Yokohama, 2008.12.6
16. 米田典央, 倉田のり, 野々村賢一「減数分裂第一分裂前期に特異的な染色体挙動の観察」イネ遺伝学・分子生物学ワークショップ 2008, 九州大学, 2008年7月5日, 福岡.
17. 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障壁となる重複遺伝子の解析」日本育種学会 114 回講演会, 滋賀県立大学, 2008年10月11日, 彦根.
18. 津田勝利, 伊藤幸博, 宮尾安藝雄, 廣近洋彦, 倉田のり「イネのシュート形成における極長鎖脂肪酸(LCFA)の機能解析」日本育種学会 114 回講演会, 滋賀県立大学, 2008年10月11日, 彦根.
19. 久保貴彦, 水多陽子, 新濱充, 春島嘉章, 倉田のり「イネ生殖隔離機構の解析から見えてくるもの」国立遺伝学研究所研究会「高等植物の生殖過程を制御する因子の多様性と生殖隔離」, 国立遺伝学研究所, 2008年11月5日, 三島.
20. 望月孝子, 菊池俊介, 濱田和輝, 加藤大貴, 大木伸彦, 藤田雅丈, 堀内陽子, 倉田のり, 矢野健太郎「OryzeExpress: イネのゲノム・アノテーションとオミックス統合データベース」第 31 回日本分子生物学会年会, 第 81 回日本生化学会大会 合同大会, 神戸ポートアイランド, 2008年12月10日, 神戸.
21. 津田勝利, 伊藤幸博, 倉田のり「イネにおける KNOX 遺伝子を介した SAM の維持および葉の分化の研究」国立遺伝学研究所研究会「イネ分子遺伝学の新展開」, 国立遺伝学研究所, 2008年12月20日, 三島.
22. 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を用いた肥満関連表現型の遺伝解析. 日本遺伝学会, 2008, 9.3-5, 名古屋.
23. 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を

- 用いたエネルギー代謝関連形質の遺伝解析. 第55回日本実験動物学会総会, 2008, 5.15-17, 仙台.
24. 岡彩子, 高田幸, 古関明彦, 森脇和郎, 城石俊彦: マウス亜種間における生殖隔離と減数分裂期のチェックポイント機構との関連. 日本遺伝学会, 2008, 9.3-5, 名古屋.
25. 田中英希, 田村義保「新しい開曲線記述子を用いた2次元部分形状の定量的評価」日本計算機統計学会 第22回大会, 秋田文化会館, 平成20年5月23日
26. 江口真透. バイオインフォマティクスにおける統計的課題について. 科研費研究集会「高次元データの統計解析」博多. 2008.11.21
小森理, 江口真透. 1クラスラベルに注目したブースティング. 統計関連学会連合大会, 横浜. 2008.9.8
27. プリチャード真理, 江口真透. マイクロアレイにおける遺伝子選択と判別能力の関係. 統計関連学会連合大会, 横浜. 2008.9.9

<著書等>

1. Koide T, Catanesi CI, Nishi A, Shiroishi T, Kasai S, Ikeda K, Takahashi A. Advantage of using wild-derived mouse strains for a variety of pain-related studies: Genetic diversity and new genetic tools. In: Acute Pain (Columbus F, ed), Nova Science Publishers, New York, in press.
2. Eguchi, S. Information Divergence Geometry and the Application to Statistical Machine Learning. Eds. Frank Emmert-Streib and Matthias Dehmer 'Information Theory and Statistical Learning', Springer, New York, 10.1007/978-0-387-84816-7_13

プロジェクト名： 生物多様性解析

サブテーマ名： 「生物多様性データの統計的モデリングと解析システムの開発」

研究代表者： 栗木 哲 [統計数理研究所]

1. 研究目標

現代遺伝学研究の目的の一つは、生物個体の多様性をゲノム多型に結びつけて理解することである。その目的のためには、生物の多様性とゲノム多型の双方を計量化し、それらから必要な情報を抽出し、連関を探ることが必要となる。ところで遺伝現象はそれ自体確率的な事象であり、そのためそのモデル化において統計モデルを用いることはきわめて有効である。このことは遺伝学と統計学はそれらの初期において未分化であったことから伺い知ることができる。そのような背景を踏まえて、本サブテーマでは遺伝学研究所で得られたデータを解析するための統計的手法の研究を行う。データ解析の結果として遺伝学の分野で重要な知見を得るとともに、開発した統計解析の方法論をより広い分野へ適用するための普遍化を行うことが本サブテーマの研究目標である。

この研究目標の達成のために、本年度は次の5つの課題を設定した。

課題 a. フーリエ記述子を用いた形態多様性データの統計的モデリング

課題 b. マウス社会行動の統計的モデリング

課題 c. 多変量時系列モデルによるマウス自発行動の特徴抽出

課題 d. 生殖隔離障壁に関わるエピスタシスの統計学的検出と多重性調整

課題 e. QTL 解析法の最適化

2. 年度研究計画

平成16年度（予備研究）

課題 b. マウス2系統（C57BL/6 と MSM）を用いて活動性と社会行動解析の基礎となる行動データを得た。更に、これらのデータから数理・統計的手法を用いた解析を行うための、データ処理等の準備を進めた。

課題 e. マウスの2系統（C57BL/6 と KJR）を用いてホームケージ内活動性の行動データを得た。この2系統間では顕著な活動性の差が存在することから、この2系統を交配して得られたF2集団の活動性データをもとにQTL解析の準備を進めた。また、新規な場面であるオープンフィールドにおける活動性についても同様にデータの準備を進めた。

平成17年度

課題 a. P型フーリエ記述子を用いてマウス下顎骨の形を数値化し、得られた数値データの多変量解析を実施した。

課題 b. マウス2系統の活動性データから日内の短時間での変化の周期性を見つけ出すため、点

過程モデルにより解析を進めた。また、社会行動については動きの方向ベクトルに関する尤度推定を行うことで社会性の数値化を進めた。また、2 個体間の距離の分布に基づく社会行動の定量的評価法の開発に向けて、ヒストグラムによる予備的解析を行った。

課題 c. 個々のマウス活動性データに点過程モデルを適用し、得られた強度関数の主成分分析から、系統特有の日周期パターンを探索した。

課題 d. 配偶体または接合体内の異なる遺伝子座の生殖的隔離障壁間の相互作用検出のため、マーカー分離の独立検定を行った。独立性カイ 2 乗検定統計量を格子点上の自由度 4 の確率場と捉え、マーカーの連鎖による相関構造を確定し、調整 p 値の計算をシミュレーションにより行った。

課題 e. ケージ内での活動性とオープンフィールド活動性 (OFA) に関わる行動要素をより明らかにするために、F2 集団のケージ内活動性について、従来の総活動性 (THA) に加え、時間的要素である活動時間 (AT) と量的要素である平均活動量 (AA) に分離した測度を用いた解析を行った。それぞれの形質について、広義の遺伝率 (H2) を調べた結果、何れの形質も 0.5 以上の高い値を示したことから、これらの形質は高い遺伝性を示すことがわかった。

平成 18 年度

課題 a. P 型フーリエ記述子を用いてマウス下顎骨の形を数値化し、得られた数値データの多変量解析を実施するためのソフトウェアを整備した。

課題 b. マウスの自発活動性や社会性などの行動を時系列に沿って効果的に解析するシステムの確立を進めた。活動性に関しては、点過程で活動をモデル化し、活動頻度を強度関数により推定する手法を用いて各系統に共通の日周期を抽出することを目指した。社会行動については、2 次元 unit vector chain を用いた解析と 2 個体間の距離情報を用いた解析を進めることで、社会行動の評価を目指した。

課題 c. 30 分ごとの活動に集約した上で、系統ごとの日周期プラス個別変動に活動量を分解し、日周期の系統差・短期的変動の系統特性を定量的に分析した。

課題 d. 独立性カイ 2 乗検定統計量の自由度 1 の 4 個の独立なコンポーネントへ分解と、配偶体または接合体内の異なる遺伝子座の生殖的隔離障壁間の相互作用の遺伝的モデルとの対応関係を明らかにした。また、カイ 2 乗確率場の最大値分布の逐次解析理論による解析式を導出し、実用的な多重性調整 p 値の計算法を与えた。

課題 e. これらの F2 個体群の測度を用いた composite interval mapping を行った結果、THA, AT, AA のそれぞれについて関与する QTL 遺伝子座 (Hylaq1, Hylaq2, Hylaq3) を同定した。更に、これら遺伝子座の AT および AA に対する独立の効果の詳細に調べるため、R/qrtl による共変量を用いた QTL interval mapping を行い、それぞれの行動要素間の関連を明らかにした。

平成19年度

課題 a. H18 年度に作成したデータ，開発してソフトウェアを用いて，下顎形態の数値化を行った。形状の数値化のための新たな方法を研究した。また，マウスの形態の解析のプロジェクトに参加した。B6 マウスと MSM マウスの下顎形態の特徴を決定している因子を探索した。田村たちが行った P 型フーリエ記述子を特徴量としてパターン認識をアダプテストで行うとシャープな特徴量が発見された。この特徴量を量的形質とする QTL 解析を行った結果 13 番染色体に QTL 遺伝子形が示唆された。

課題 b. 複数の動物個体が示す社会行動は，生存において重要な意味を持つが，その定量化が難しく解析を困難にしている。ここでは，2 個体のマウスを同時にオープンフィールドにおいた際の 10 分間の社会行動を時系列のトラッキングデータに基づき，統計モデルで特徴づける解析を進めた。社会的行動が異なる B6 および MSM の 2 近交系統の特徴を，移動行動のベクトル間の相互作用をもとに尤度推定値を算出することで数値化した。特に H19 年度は移動行動からベクトルを算出する方法に改良を加えて相互作用のパラメータを尤度推定することで，B6 と MSM の 2 近交系統の社会行動の違いを明確に捉えることを可能にした。また 2 個体のマウスのビデオ画像をトラッキングデータと同期させて表示させ，観測者がマウスの状態を目視により効率的に記録することができるソフトウェアを開発した。このソフトウェアを用いてマウスの行動の時系列パターンにマルコフモデルを当てはめて予備的解析を行った。

課題 c. マウス活動量データ（赤外センサー交差回数時系列）に，三角関数列を説明変数とするポアソン回帰モデルを適用し，24 時間周期から 2 分周期まで全ての可能な周波数の有意性を検出する方法を提示した。

課題 d. 相互作用の真偽判定の妥当性を吟味するため，以前構築したイネの高密度連鎖地図と同じ組合せ親による別集団で連鎖地図作製を行った。

課題 e. C57BL/6 系統と KJR 系統は顕著に異なった活動性を示す。この活動量の違いに関わる遺伝的基盤を解明する目的で，F2 集団を用いて時間的要因とスピードの要因を表現型とする QTL 解析を行い複数の遺伝子座を得た。その結果に基づき，Structural Equation Model (SEM) 解析を行い，各遺伝子座と各活動量指標との関係を構成するシステムを明らかにした。

平成20年度

課題 a. 部分形状の統計解析で部分形状の数値記述に用いられてきた開曲線記述子である接線フーリエ記述子 (TFD) は，曲線上の隣接点間の位置の差分系列のフーリエ変換でありノイズを増幅する傾向がある。そこで新たな記述子として protrusion フーリエ記述子 (PFD) を提案した。次に，2 系統を起源とする F2 世代（計 207 匹）のマウスを用い下顎骨の部分形状の遺伝解析を行った。異なる部位に対し形に影響する遺伝子が大きく異なること，さらに記述子からの曲線形状の視覚化により，各々の遺伝子の形に対する影響の仕方がそれぞれ異なることなどが分かった。

課題 b. (1) 基礎的データ解析: 2 個体のマウスを同時にオープンフィールドに置いた際の 10 分間

の社会行動時系列のトラッキングデータを抽出し、それらのデータに基づいて2個体の行動を統計数理モデルで特徴づける解析を進めた。(2) 方向データ解析：マウス2個体の社会行動トラッキングデータから、2個体同士の社会行動時系列を移動のベクトル系列に変換し、方向相互作用のパラメータを尤度推定することで移動行動の特徴を数値化している。今年度は、移動行動からベクトルを算出するアルゴリズムに改良を加え、新しく得られたマウス24系統461ペアに対する予備的解析を行った。(3) マルコフモデルによる解析：マウス2個体の画像データから目視によりマウスの状態を5つに分類した状態時系列を解析し、この時系列にマルコフ行列をあてはめて得られる遷移行列の各要素の重みつき和によって、適切な行動の社会性の尺度が構築できることを確認した。

課題c. マウスの自発行動を記録したデータは非負整数値による計数時系列で極めてゼロが多く、上下の非対称性が強い。興味があるのは活動期の短期的変動の特徴であるので、1分値で計数0は欠測として非ゼロのデータを対数・巾関数等で線形化し、欠測値を許す線形ガウス型カルマンフィルタで、日周期まわりの自己回帰モデルをあてはめることにより、特徴的な活動周波数を推定する方法を研究した。

課題d. 致死遺伝子相互作用の遺伝的モデルについて浸透度を考慮した一般化、正逆戻し交雑集団を使った真の相互作用の検証を行った。また正逆戻し交雑集団から観測された多数の検定統計量データについて、過去に開発された多重性調整を適用した。

課題e. ロバストQTLについても幾つかの検討を加えた。従来の方法は外れ値の影響を大きく受けることがあるが提案された手法は非常に安定した性能を示すことが示された。また、QTL解析における影響診断の方法を検討した。単にロッドスコアのピークの大きさだけでなく、ロッドスコアが形状(単峰であるか二峰であるかなど)に影響を与える個体を、グラフィカルに図示する方法の開発を行った。

平成21年度

最終年度であるため、昨年度の課題a～eを継続して行い、成果発表を行う。

課題a. 形状の数値化のための新たなフーリエ記述子について研究する。前年度までに開発したprotrusionフーリエ記述子(PFD)と新たな記述子を用いて、遺伝子の交互作用についての研究を進める。

課題b. これまでに確立した二つの社会行動評価法を用いて、実際の24系統461ペアの時系列移動行動データから実際のマウス社会性の評価を行う。こうして数理的解析により得られた社会性評価の結果と、実際のビデオ観察法による行動項目の定量結果とを比較し、これらの解析手法の総合的な検討を行う。

課題c. H20年度に開発したプロトタイプをB6とMSMの全データに適用し、系統共通の基調的活動周期が存在するかどうかを調べる。一本の時不変係数自己回帰モデルではあてはまりが不十分である可能性もあるので、個体の異質性をうまく捉えるモデリング方法について研究する。個体差をうまくコントロールするモデルを発見し、系統ごとの特徴抽出を行うことが目標

である。

課題 d. 生殖隔離障壁を引き起こす遺伝子座相互作用の検索において、真の相互作用を再現性実験や別集団による確認なしに検出する手法を開発する。またマーカー間隔が均一でないことをモデルに取り込んだ大偏差型の近似公式を検討する。

課題 e. 昨年度、開発されたロバスト QTL 法について更に発展させる。特に外れ値の影響にロバストでありながらデータの情報を効率良く取り込むために、データに応じてロバスト QTL 法を選択する方法を考える。これらの成果を持って、このプロジェクトで得られたデータに実際に適用する。また、影響診断については、単一マーカー分析だけでなく、区間マッピング法によって得られたロッドスコアについて、その大きさや形状に影響を与える個体を特定する方法を開発する。また近年高次元の表現型データに関する QTL 解析の需要が高まっている。eQTL 解析は、その最たるものである。多変量の連続変量である表現型と、多変量の離散変量である遺伝子型を関連づける QTL 解析のためのグラフィカルモデルの推測、モデル比較の方法を検討する。

平成 22 年度以降の展開

課題 a. (画像データの統計モデリングによるゲノム機能解析) 部分形状 (開曲線) の記述には接線ベクトル系列のフーリエ変換があるが、両端点間と急カーブでの系列の不連続のため情報縮約に劣る。改善策として、急カーブ近傍では速度を落とすような速度ベクトル系列のフーリエ変換 (速度フーリエ記述子) を考案し、これを用いた形態の遺伝解析を行う。さらに 3 次元形状への拡張を図る。

課題 b. 隠れマルコフ行列を用いた 2 個体画像の状態時系列解析による社会性尺度構築法を一般化し、従来定量化が困難であったような (ヒトを含む) 動物行動解析や大量データからの自動異常行動抽出に利用する。また、得られた知見を、行動と遺伝的形質を結び付ける研究に活用していく。

課題 c. 遺伝的に純粋系統の個体でも行動の個体差は大きく、解析結果はばらつくので、個体固有の変動を表現する要因を組み込んだ階層型統計モデルの適用可能性を研究する。

課題 d. 非線形再生理論 (逐次解析) を用いた QTL 解析・生殖隔離機構検出の p 値多重性調整法は、マーカーがほぼ等間隔に配置されているという前提が必要であった。マーカー間隔もランダムな確率変数であるというモデルの下で、同様の多重性調整が可能でないかどうかを検討する。

課題 e. ロバスト QTL は個体の尤度への寄与に応じて、重み付けロッドスコアを構成する。この重み関数を利用してロバストな影響診断の方法を検討する。従来の影響診断法の弱点である複数の個体の影響を探索できる方法を開発したい。eQTL については表現形が高次元になるため遺伝形との関連を直接に視覚化する方法を階層クラスタリングを参考にして開発したい。

3. 研究経費の推移

平成17年度実績： 157,820千円（プロジェクト全体予算）
平成18年度実績： 18,800千円
平成19年度実績： 24,560千円
平成20年度実績： 25,000千円
平成21年度見込： 20,820千円

4. 平成20年度の研究実施体制

[統計数理研究所] 江口真透 奥田将己 川崎能典 坂口隆之 田中英希 種村正美
田村義保 土谷 隆 藤澤洋徳 Dou Xiaoling Haque Mollah
[国立遺伝学研究所] 小出 剛 岡 彩子 城石俊彦 杉本大樹 高田豊行 田村 勝
春島嘉章 前野哲輝 細谷正樹
[東京大学] 原 尚幸
[九州大学] 二宮嘉行

5. 平成20年度研究成果

(1) 成果物（知見・成果物・知的財産権等）

(2) 成果発表等

<論文発表>

[学術論文]

1. Tanaka, E., Tamura, Y., Hosoya, M. and Shiroishi, T., "Protrusion Fourier Descriptor: Skeleton-based Representation of Open Curves", *Forma*, Vol. 23, 9-18, 2008.
2. Tanaka, E. and Tamura, Y., "Skeleton-based Fourier Descriptor of Open Curves", *Pacific Science Review*, Vol. 10, No. 10, 194-198, 2008.
3. Takahashi A., Shiroishi T., Koide T.: Multigenic factors associated with a hydrocephalus-like phenotype found in inter-subspecific consomic mouse strains. *Mammalian Genome*, 29:333-338, 2008.
4. Takahashi A., Nishi A., Ishii A., Shiroishi T., Koide T.: Systematic analysis of emotionality in consomic mouse strains established from C57BL/6J and wild-derived MSM/Ms. *Genes, Brain and Behavior*. 7:849-858, 2008.
5. Ito, N. and Kurata, N. Disruption of KNOX gene suppression in leaf by introducing its cDNA in rice. *Plant Science*. 174: 279-289. 2008.
6. Thirumurugan, T., Ito, Y., Kubo, T., Serizawa, A. and Kurata, N. Identification, characterization and interaction of HAP family genes in rice. *Mol Gen Genomics*. 279: 279-289. 2008.
7. Suzuki, T., Eiguchi, M., Kumamaru T., Satoh, H., Matsusaka, H., Moriguchi, K. and Kurata, N.

- MNU-induced mutant pools and high performance TILLING enable finding of any gene mutation in rice. *Mol Gen Genomics*.279:213-223. 2008.
8. Suwabe, K. Suzuki, G. Takahashi, H. Shiono, K. Endo, M. Yano, K. Fujita, M. Masuko, H. Saito, H. Fujioka, T. Kaneko, F. Kazama, T. Mizuta, Y. Separated transcriptomes of male gametophyte and tapetum in rice: validity of a laser microdissection (LM) microarray. *Plant Cell Physiol*.49:1407-1416. 2008.
 9. Hobo, T. Suwabe, K. Aya, K. Suzuki, G. Yano, K. Ishimizu, T. Fujita, M. Kikuchi, S. Hamada, K. Miyano, M. Fujioka, T. Kaneko, F. Kazama, T. Mizuta, Y. Takahashi, H. Shiono, K. Nakazono, M. Tsutsumi, N. Nagamura, Y. Kurata, N. Watanabe, M. Matsuoka, M. Various spatiotemporal expression profiles of anther-expressed genes in rice. *Plant Cell Physiol*.49:1417-1428. 2008.
 10. Kawakita, Masanori., Eguchi, Shinto., Boosting method for local learning in statistical pattern recognition. *Neural Computation* 20, 2792-2838 (Nov., 2008)
 11. Fujisawa, Hironori., Eguchi, Shinto., Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis* 99, 2053-2081 (Oct., 2008)
 12. Takenouchi, Takashi., Eguchi, Shinto., Murata, Noboru., Kanamori, Takafumi., Robust Boosting Algorithm Against Mislabeling in Multiclass Problems. *Neural Computation* 20, 1596-1630 (June, 2008)

[会議録]

1. 渡辺正夫, 倉田のり 国立遺伝学研究所研究会「高等植物の生殖過程を制御する因子の多様性と生殖隔離」, 国立遺伝学研究所, 2008年11月5日～6日, 三島.
2. 長戸康郎, 倉田のり 国立遺伝学研究所研究会「イネ分子遺伝学の新展開」, 国立遺伝学研究所, 2008年12月19～20日, 三島
3. Mollah, Md. Nurul Haque, Eguchi, Shinto Robust Composite Interval Mapping for QTL Analysis by Minimum β -Divergence Method. *IEEE International Conference on Bioinformatics and Biomedicine 2008*, 115-120, 978-0-7695-3452-7, 2008.11

[解説・総説]

1. Koide T, Catanesi CI, Nishi A, Shiroishi T, Kasai S, Ikeda K, Takahashi A. Systematic mapping of pain-related QTL using consomic mouse strains: Advantage of using wild-derived strains. *Brain Res J*. in press.
2. 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, II, 小西貞則, 国友直人(編), 東京大学出版会, 東京, 315-356, 2008.7

<会議発表等>

[招待講演]

1. 江口真透. タンパク質構造と進化と情報幾何. 数理研短期共同研究集会「離散力学系の分子細胞生物学への応用数理」. 京都, 日本. 2009.1.8

2. Eguchi, Shinto. Information divergence geometry and its application to machine learning. The 1st MSJ-SI, Probabilistic Approach to Geometry, Kyoto, Japan.2008.8.4

〔一般講演〕

1. 田中英希, 田村義保「新しい開曲線記述子を用いた2次元部分形状の定量的評価」日本計算機統計学会 第22回大会, 秋田文化会館, 平成20年5月23日
2. Tanaka, E. and Tamura, Y.“Skeleton-based Fourier Descriptor of Open Curves” APCOM2008, Kokushikan University, September 3, 2008
3. Takahashi, A., Sugimoto, H., Kimura, S., Tomihara, K., Tsuchiya, T., Kakihara, S., Tanemura, M., Shiroishi, T., Koide, T.: Complex genetic architecture of social interaction and aggressive behavior clarified using consomic strains derived from MSM and C57BL/6. 22nd International Mammalian Genome Conference, Prague, November 2-5, 2008. (ポスター)
4. Takahashi, A., Tomihara, K., Shiroishi, T., Koide, T.: Genetic analysis of inter-male aggression using consomic mouse strains established from C57BL/6J and MSM. 10th Annual Meeting of the International Behavioural and Neural Genetics Society, Portland, May 5-9, 2008.
5. Horiuchi Y., Harushima Y., Mochizuki T., Fujisawa H., Eguchi., Kawakita M., Kurata N. Detection of nucleotide and expression polymorphisms between rice strains using affymetrix rice genome array. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
6. Harushima Y., Kuriki S., Mizuta Y., Kurata N. Detection of pairs of interactive reproductive barriers in rice. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
7. Mizuta Y., Harushima Y., Kurata N. Positional cloning of a pair of interactive genes causing reproductive barrier in the hybrid pollen of rice. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
8. Harushima Y., Yano M., Kurata N. Identification of a reproductive barrier working in the process of pollen competition in rice. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
9. Tsuda K., Ito Y., Miyao A., Hirochika H., Kurata N. Identification and analysis of rice mutants misexpressing NKOX genes in leaves. XX International congress of Genetics, Berlin, Germany, July 12-17, 2008.
10. 米田典央, 倉田のり, 野々村賢一「減数分裂第一分裂前期に特異的な染色体挙動の観察」イネ遺伝学・分子生物学ワークショップ2008, 九州大学, 2008年7月5日, 福岡.
11. 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障壁となる重複遺伝子の解析」日本育種学会114回講演会, 滋賀県立大学, 2008年10月11日, 彦根.
12. 津田勝利, 伊藤幸博, 宮尾安藝雄, 廣近洋彦, 倉田のり「イネのシュート形成における極長鎖脂肪酸(LCFA)の機能解析」日本育種学会114回講演会, 滋賀県立大学, 2008年10月11日, 彦根.
13. 久保貴彦, 水多陽子, 新濱充, 春島嘉章, 倉田のり「イネ生殖隔離機構の解析から見えてく

- るもの」 国立遺伝学研究所研究会「高等植物の生殖過程を制御する因子の多様性と生殖隔離」, 国立遺伝学研究所, 2008年11月5日, 三島.
14. 望月孝子, 菊池俊介, 濱田和輝, 加藤大貴, 大木伸彦, 藤田雅丈, 堀内陽子, 倉田のり, 矢野健太郎「OryzeExpress: イネのゲノム・アノテーションとオミックス統合データベース」第31回日本分子生物学会年会, 第81回日本生化学会大会 合同大会, 神戸ポートアイランド, 2008年12月10日, 神戸.
 15. 津田勝利, 伊藤幸博, 倉田のり「イネにおける KNOX 遺伝子を介した SAM の維持および葉の分化の研究」 国立遺伝学研究所研究会「イネ分子遺伝学の新展開」, 国立遺伝学研究所, 2008年12月20日, 三島.
 16. Kuriki, S., Harushima, Y., Fujisawa, H., and Kurata, N., Multiplicity Adjustments in Detecting Reproductive Barriers caused by Loci Interactions, BIRS Workshop 09w5040 Random Fields and Stochastic Geometry, 2009.2.26, Banff, Canada
 17. 江口真透. バイオインフォマティクスにおける統計的課題について. 科研費研究集会「高次元データの統計解析」博多. 2008.11.21
 18. Md. Nurul Haque Mollah., Eguchi, Shinto., Robust QTL Analysis by the Minimum β -Divergence Method. International Association for Statistical Computing 2008, Yokohama, 2008.12.6
 19. Pritchard, Mari., Eguchi, Shinto. Finding Optimal Gene Set for Classification from Multiple Predictive Gene Sets. International Association for Statistical Computing, Yokohama, 2008.12.6
 20. 小森理, 江口真透. 1クラスラベルに注目したブースティング. 統計関連学会連合大会, 横浜. 2008.9.8
 21. プリチャード真理, 江口真透. マイクロアレイにおける遺伝子選択と判別能力の関係. 統計関連学会連合大会, 横浜. 2008.9.9

<著書等>

1. Koide T, Catanesi CI, Nishi A, Shiroishi T, Kasai S, Ikeda K, Takahashi A. Advantage of using wild-derived mouse strains for a variety of pain-related studies: Genetic diversity and new genetic tools. In: Acute Pain (Columbus F, ed), Nova Science Publishers, New York, in press.
2. Eguchi, Shinto. Information Divergence Geometry and the Application to Statistical Machine Learning. Eds. Frank Emmert-Streib and Matthias Dehmer 'Information Theory and Statistical Learning', Springer, New York, 10.1007/978-0-387-84816-7_13