

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(H<sup>2</sup>)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 型フリー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/qt1 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する。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 型フーリエ子の特徴量としてパタ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ている。

0 マッピングの再現性を確認するとともに相互作用検定量の再現性を確認し、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どうかを判定する。また、マーカー間隔の不均一性を確率的にモデル化した上で、平成 19 年度に導いたカイ 2 乗確率場の最大値分布と同様の解析式の導出を試みる。

課題 e. QTL 解析を SEM 解析する場合に問題なのは、従来の SEM 解析が連続値データだけを解析の対象しているのに対し、遺伝子型は離散値データであることである。本年度は、連続値と離散値が混在するデータにおける SEM 解析、グラフィカルモデルの、とくに推定のアルゴリズムを整理、開発する。また QTL 解析における影響度解析（各個体のデータが全体の解析結果にどのように影響するか）の方法論を開発する。

## 平成 21 年度

課題 a. 下顎骨の最終的の形状、形状の時間変化の違いが、どのような遺伝情報によっているかについて研究する。

課題 b. H20 年度に開発する隠れマルコフモデルを用い、目視観測による状態時系列データの存在しないマウス 2 個体の座標時系列データ(約 200 件)を自動的に処理し、遺伝的形質と社会的行動の関係を解析する。方向ベクトル解析においては、時間依存のパラメータ推定を本格的に進めると共に、ソフトウェアを公開するための準備を行う。

課題 c. 平成 20 年度までの議論を、複数のマウスのデータを同時にモデル化することができるよう、多変量ベイズ型動的線形モデルの枠組みを構築する。

課題 d. 分離頻度をゆがめるゲノム上の生殖的隔離障壁と細胞質との相互作用を調べる目的で、交雑親は同じで、雌雄を換え細胞質の異なる F<sub>2</sub> 集団にて連鎖地図を作製する。細胞質が異なる事によって生殖的隔離障壁がどの様になるか調べる。

課題 e. 前年度に引き続き。連続、離散変数が混在する SEM 解析、グラフィカルモデルの推測研究、ならびに QTL 解析の影響度解析の研究を行う。

## 平成 22 年度以降の展開

課題 a. 2 次元+表面テクスチャの数値化手法を発展させ、3 次元形態の数値化手法の研究を行うと共に、21 年度までの成果をソフトウェアとして公開する。このソフトウェアで、デジタルカメラ、2D(3D)スキャナ等で採取したデジタルデータを取り込めば QTL 解析まで、出来るようにする。

課題 b. H21 年度までに得られた成果に基づいて、マウス社会行動を総合的に解析するための統

合ソフトウェアを開発し、その公開を目指す。

課題c. 純粋系統に対して構築してきた解析システムを、混血系統に対しても適用し、意味ある結論が得られるようにシステムを改良する。判別問題としては、より難しくなると思われる。

課題d. 配偶体または接合体の遺伝子座と親の遺伝子座との相互作用によっても生殖的隔離障壁は起きる。正逆戻し交雑集団を用いて親の遺伝型との相互作用を検出する。

### 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年度見込： 24,560千円

### 4. 平成19年度の研究実施体制

[統計数理研究所] 江口真透 川崎能典 坂口隆之 種村正美 田村義保 土谷 隆  
藤澤洋徳 Nurul Haque Mollah

[国立遺伝学研究所] 梅森十三 小出 剛 岡 彩子 城石俊彦 田村 勝  
春嶋嘉章 高田豊行 前野哲輝 細谷正樹

[九州大学] 二宮嘉行

### 5. 平成19年度研究成果

#### (1) 成果物（知見・成果物・知的財産権等）

1. P型フリーエ記述子に関するソフトウェアをホームページに公開した。

#### (2) 成果発表等

##### <論文発表>

##### [学術論文]

1. Fujisawa, H., Isomura, H., Eguchi, S., Ushijima, M., Miyata, S., Miki, Y., Matsuura, M. Identifying haplotype block structure by using ancestor-derived model. *J. Human Genetics* 52 (9), 738-746, 2007.
2. Kanamori, T., Takenouchi, T., Eguchi, S. Murata, N. Robust loss functions for boosting. *Neural Computation* 19, 2183-2244, 2007.
3. Mollah, M. N. H., Minami, M. and Eguchi, S. Robust prewhitening for ICA by minimizing beta-divergence and its application to FastICA. *Neural Processing Letters* 25, 91-110, 2007.
4. Takenouchi, T., Ushijima, M. and Eguchi, S. GroupAdaBoost: accurate prediction and selection of important genes. *IPSI Transactions on Bioinformatics* 3, 1-8, 2007.
5. Ninomiya, S. and Fujisawa, H. A conservative test for multiple comparison based on highly correlated test statistics, *Biometrics*, 63, 1135-1142, 2007.
6. Takenouchi, T., Eguchi, S., Murata, N. and Kanamori, T. Robust boosting algorithm against

mislabeled in multi-class problems. To appear in Neural Computation, 2007.

7. Ushijima, M., Miyata, S., Eguchi, S., Kawakita, M., Yoshimoto, M., Iwase, T., Akiyama, F., Sakamoto, G., Nagasaki, K., Miki, Y., Noda, T., Hoshikawa Y. and Matsuura, M. Common peak approach using mass spectrometry data sets for predicting the effects of anticancer drugs on breast cancer. *Cancer Informatics* 3, 285-293, 2007.
8. Zheng, Z., Iwata, H., Hirata, H. and Tamura, Y. Quantitative evaluation of the degree of sprout leaf bending of rice cultivars using P-type Fourier descriptors and principal component analysis, *Euphytica*, 2008, DOI 10.1007/s10681-007-9642-9 (to appear)(online 版  
<http://www.springerlink.com/content/513u046251521246/>)

#### [解説・総説]

1. 江口真透, 統計的パターン認識: 線型判別からアダブーストまで。日本化学会情報化学部会誌 25, 68, 2007。
2. 江口真透, リスク科学のための人工知能的方法と統計的思考。人工知能学会誌 22 巻 5 号特集「定量的リスク科学」656-662, 2007 年 9 月。

#### <会議発表等>

##### [招待講演]

1. Eguchi, S. Boosting learning approach to association studies in bioinformatics, First joint meeting between Institute of Statistical Science, November 29, 2007, Academia Sinica, Taiwan and the Institute of Statistical Mathematics.
2. Eguchi, S. Boosting methods for association studies in bioinformatics, International Conference on Multiple Decision Theory, Statistical Inference and Applications, December 28, 2007, Taipei, Republic of China.
3. 江口真透. 表現形の相関研究: パタン認識の先に何が見えるか。春季統計学会, 2008 年 3 月 1 日。
4. Kuriki, S. Multiplicity Adjustments in Detecting Reproductive Barriers caused by Loci Interactions, First joint meeting between Institute of Statistical Science, Academia Sinica, Taiwan and the Institute of Statistical Mathematics, November 29, 2007, Tokyo.

##### [一般講演]

1. 江口真透, 竹之内高志, マルチクラスの判別問題におけるミスラベルのモデリング, 統計関連学会, 2007 年 9 月 7 日
2. 江口真透, 観察データの推測の限界—揺らぎモデルアプローチ—広島大学理学部, 科研費研究集会「統計的データ解析手法の評価と開発」2008 年 1 月 16 日, 広島。
3. 川崎能典, 多変量計数時系列データのモデリング, 統計数理セミナー, 2007 年 4 月 18 日, 統

計数理研究所, 東京。

4. 栗木 哲, 春島嘉章, 藤澤洋徳, 倉田のり, 遺伝子座間の相互作用による生殖的隔離障壁の検出と多重性調整, 統計関連学会連合大会, 神戸大学, 2007年9月6日, 神戸。
5. 栗木 哲, 格子点上カイ2乗確率場の最大値分布の近似とその連鎖解析への応用, 科研費研究集会「統計的モデリングの方法と理論」, 一橋大学, 2007年11月27日, 東京。
6. 栗木 哲, 生殖的隔離に関わるエピスタシス(相互作用)検出と多重性調整, 融合研究シンポジウム「地球と生命の新パラダイム創造への挑戦」ハイライト講演, 2007年11月29日, 東京。
7. 春島嘉章, 栗木哲, 水多陽子, 藤澤洋徳, 倉田のり「イネF2集団における生殖的隔離障壁の相互作用の検出」第79回日本遺伝学会, 岡山大学, 2007年9月21日, 岡山。
8. 春島嘉章, 倉田のり「栽培イネの第3染色体の雄性配偶体型生殖的隔離障壁と相互作用する第6染色体の雌性親遺伝子のポジショナルクローニング」日本育種学会112回講演会, 山形大学, 2007年9月22日, 鶴岡。
9. 堀内陽子, 藤澤洋徳, 川喜田雅則, 望月孝子, 春島嘉章, 坂口隆之, 江口真透, 倉田のり「Rice Genome Arrayを用いたSFP検出手法の開発」日本育種学会112回講演会, 山形大学, 2007年9月23日, 鶴岡。
10. 堀内陽子, 藤澤洋徳, 川喜田雅則, 望月孝子, 春島嘉章, 坂口隆之, 江口真透, 倉田のり「Affymetrix Rice Genome Arrayを用いたSFP検出手法の開発」第30回日本分子生物学会年会・第80回日本生化学会大会 合同大会, パシフィコ横浜, 2007年12月13日, 横浜。
11. 水多陽子, 春島嘉章, 倉田のり「イネ亜種間交雑における生殖的隔離障壁遺伝子のポジショナルクローニング」2007年イネ分子遺伝学ワークショップ, 名古屋大学, 2007年7月3日, 名古屋。
12. 水多陽子, 春島嘉章, 倉田のり「イネ雑種花粉で相互作用する2遺伝子座に起因する生殖的隔離」第79回日本遺伝学会, 岡山大学, 2007年9月20日, 岡山。
13. 水多陽子, 春島嘉章, 倉田のり「イネ雑種花粉で作用する生殖的隔離障壁遺伝子のポジショナルクローニング」日本育種学会112回講演会, 山形大学, 2007年9月23日, 鶴岡。
14. 水多陽子, 春島嘉章, 倉田のり「イネ雑種花粉で作用する生殖的隔離障壁の単離・機能解析」特定領域研究「植物ゲノム障壁」第1回若手の会, 茂庭荘, 2007年11月5日, 仙台。

### (3) その他の成果発表

#### 国際会議

1. Umemori J., Nishi A., Takahashi A., Kawasaki Y., Lionikas A., Blizard D., Koide, T. QTL analysis of differences of activities in home-cage and open-field between KJR and B6 mouse strains. 6<sup>th</sup> Annual Meeting of the Complex Trait Consortium, May 26-29, 2007, Braunschweig, Germany.