

# 論文の内容の要旨

## 論文題目

### Bioinformatics for Understanding Animal Behavior (動物行動を理解するためのバイオインフォマティクス技術の開発)

氏名 福永 津嵩

## 1 General Introduction

動物が示す多様な行動は、古来より人々の広い関心を引いてきた。これらの動物行動を実験科学的に検証するためには、まず動物行動を定量化する事が必要になる。そのような行動の定量化方法として、特定の行動を行った回数を数える「目視カウントによる定量化」が頻繁に用いられてきた。しかしながらこの手法には、「行動に関する多くの情報が失われる」「観察者の主観が入り込む」といった問題点が存在している。

これらの問題点を解決する手法として、「動画解析による定量化」が近年注目を集めている。「動画解析による定量化」とは、ビデオカメラで撮影された動画を画像処理する事によって動物行動を定量化する手法であり、「多様な情報を同時に定量化出来る」「同一ソフトウェアを使う事で客観性が担保される」という長所を持つ。情報科学的手法を活用する事で動物行動の特徴解明を目指すこの新しい学問領域は“Computational Ethology”と呼ばれ、動物行動学にパラダイムシフトを引き起こす事が期待されている (図1)。

このように、動物行動を理解するためのバイオインフォマティクス研究が近年注目を集めている一方で、本分野には解決されていない重要な問題がいくつか存在している。第一の問題は、複数個体のトラッキングにおいて、個体同士が大きく接近した際に個体のidentityを正確に保持する事が出来ないという問題である。第二の問題は、教師なし学習に基づくトラッキングデータの解析手法の研究が発展していないという問題である。教師なし学習は人間の主観の影響を受けにくく未知の行動パターンを抽出するのに優れているが、得られた結果の解釈が難しいため未だ研究が限られている。第三の問題は、シーケンスデータなど他のオミックスデータ解析に基づく動物行動の分子基盤解明があまり進んでいないという問題である。

上記の背景に基づいて私は、1. 動物個体同士が重なり合うときでもトラッキング可能なシステムの開発、2. 教師なし学習に基づく解釈可能な行動パターン発見手法の開発、3. オミックスデータ解析による動物行動の分子基盤解明、という3つの研究課題について研究を行った。



図1 Computational Ethology における動物行動研究の概略図

## 2 GroupTracker: Video Tracking System for Multiple Animals under Severe Occlusion

まず、第一の課題である『動物個体同士が重なり合うときでもトラッキング可能なシステムの開発』について研究を行った。本研究では、動物個体を一つのガウス分布とみなし、複数の動物個体を混合ガウス分布とモデル化する事でトラッキングを行った。このようなトラッキングソフトウェアは、モデル生物を対象にいくつか存在している。しかしながら、混合ガウス分布の最尤推定法は本質的に不良設定問題であるため、個体同士が重なった時にはトラッキングする事が出来ないという問題点が存在していた。

本研究では、水中を三次元的に泳ぐために、個体間が大きく重なる事が多いメダカを対象として研究を行った。まず、複数匹のメダカを入れた水槽を上方からビデオカメラで撮影する事で動画データを得た (図2A)。次に、二値化及び統計的背景差分といった画像処理によって画像から動物個体のみを抽出した後 (図2B)、得られた画像を二次元混合ガウス分布としてモデル化しEMアルゴリズムによって各個体の位置を求めた。この時、一つ前のフレームの収束解をEMアルゴリズムの初期解として設定する事で、個体同士が離れている時は各個体のidentityは自然に保存される事となる。なお最初のフレームの個体の位置はK-means++アルゴリズムによって求めた。しかしながら、個体同士が重なった時は、一方のガウス分布の縮退が生じその尤度が無限大に発散してしまう事があり、適切な個体の位置を求める事が出来ないと言う問題が生じる。そこで本研究では、分散共分散行列の固有値 (動物の大きさを表す値) を固定し固有ベクトル (動物の向きを表す値) のみを更新する事で、ガウス分布の縮退を防ぐ手法を開発した。

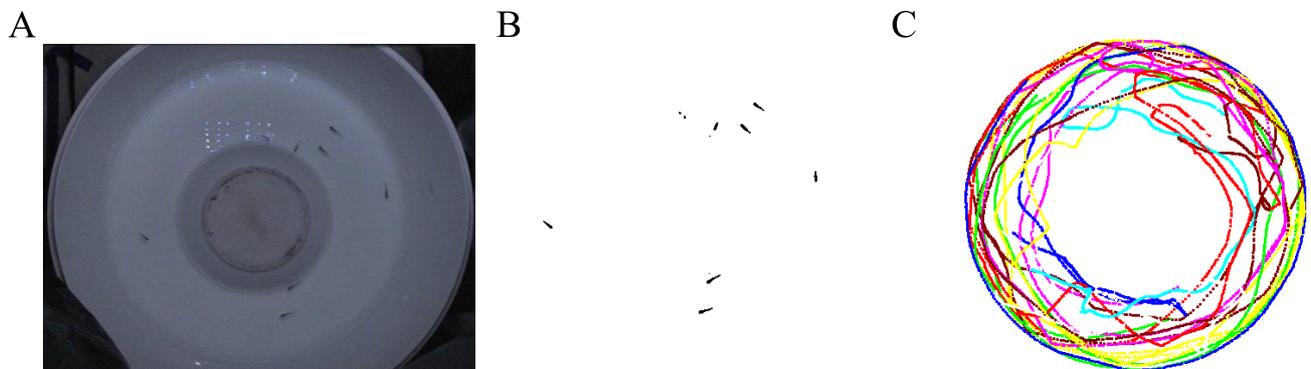


図2(A) 解析に用いた画像の例 (B) 画像処理によって動物個体のみを抽出した結果 (C) GroupTrackerによる複数の個体のトラッキング結果

撮影した動画データに対してGroupTrackerを適用した所、複数匹のメダカが同時に撮影されている条件下でも個々のメダカを追尾し、高い精度でメダカの位置と角度を取得出来る事を確認した (図2C)。精度評価の結果、一つの個体の行動軌跡が別の個体の行動軌跡に合流してしまうloss of identity errorは一回も発生していなかった。一方、交差後に2つの個体の行動軌跡が入れ替わってしまうswapping of identity errorについては92%と高い精度で交差後のidentityを保持している事がわかった。

### 3 Postural motif detection algorithm for revealing relationships among genes and behavior

次に、第二の課題である『教師なし学習に基づく解釈可能な行動モチーフ発見手法の開発』について研究を行った。本研究では、遺伝学や神経科学のモデル生物として幅広く研究が行われている線虫*C. elegans*を対象に研究を行った。Brownらは2013年に、300種を超える*C. elegans*の変異株と野生株の行動データを教師なし学習により解析する事で、特徴的な姿勢変化のパターンであるbehavioral motifを抽出し、これらのmotifに基づいて遺伝子機能ごとに変異株をクラスタリング出来る事を示した。この研究は、Computational Ethologyにおいて教師なし学習を利用した先駆的な研究である。一方でこの研究では、(1) 時系列を伴わない姿勢情報のみでの変異株のクラスタリング性能を評価していない (2) motifの有無ではなくmotifからの距離に基づいてクラスタリングを行っている、という問題点が存在している。以上の背景を踏まえて本研究では、先行研究において解析不十分であった点を再解析した上で、より適切なmotif抽出アルゴリズムを開発する事、またそのアルゴリズムを用いて行動と遺伝子パターンの関係性を直接結びつけて解釈する事を目標とした。

まず、一フレームのmotifでの変異株のクラスタリング性能について評価を行った。クラスタリング性能の評価基準として、(1) 変異株内の個体のクラスタリング性能 (2) 同一遺伝子機能を持つ変異株のクラスタリング性能の二点を用いて評価した所、一フレームのmotifは40-800フレームのmotifを利用した先行研究と同等のクラスタリング性能である事がわかった。次に、ランダムに生成した姿勢をmotifとみなしてクラスタリングの性能を評価した所、同じく先行研究と同等のクラスタリング性能を示した。これは、先行研究が motif の有無ではなくmotif からの距離によってクラスタリングをしているためであると考えられ、また同時に、先行研究で抽出されたmotif を生物学的に解釈する事が困難である事を示している (図3A, B)。そこで本研究では新しくmotifの有無に基づいてクラスタリングをする手法を考案した。具体的には、姿勢データの集合を混合ガウス分布でモデル化し、推定された各分布の平均値をmotifと定義した。また、各個体における分布の混合比を各motifの出現確率とみなした。精度確認を行った所、提案する手法は先行研究と同等の高いクラスタリング性能を持つ事を確認した (図3C)。

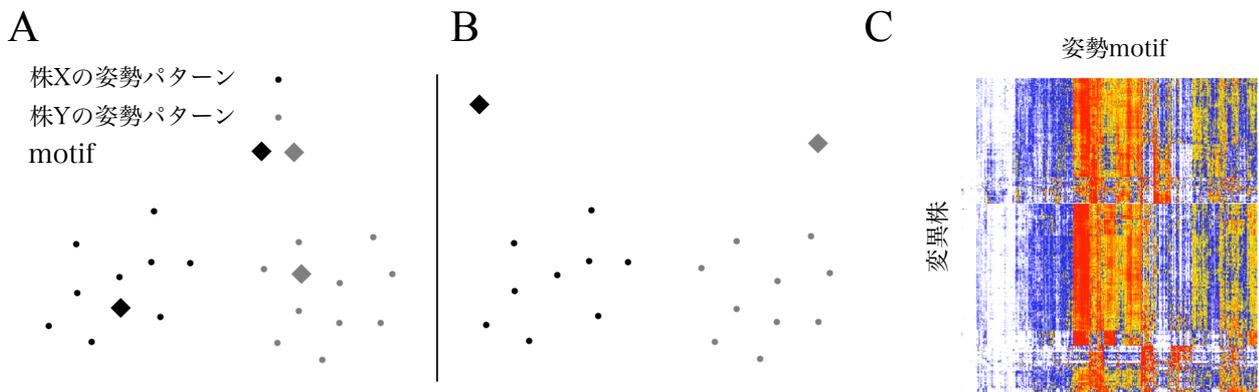


図3 (A)(B) 黒いmotifから近く灰色motifから遠い個体が株 X、その逆が株 Y と分類される場合の模式図 (C) 変異株と姿勢 motif の対応関係を示すヒートマップ

## 4 CapR: Revealing Structural Specificities of RNA-binding Protein Target Recognition using CLIP-seq Data

最後に、第三の課題である『オミックスデータ解析による動物行動の分子基盤解明』について研究を行った。神経疾患の多くは行動に異常をきたす事が知られているが、その分子生物学的メカニズムは未だ多くの謎に包まれている。特に近年、RNA結合タンパク質(RBP)やそのターゲットの変異が、脆弱X症候群を始めとする神経疾患に関わっている事が明らかとなってきた。しかしながら、RBPが標的とする遺伝子転写産物や結合領域の特徴を把握することは、これまでの分子生物学的手法では困難であり、網羅的にデータを解析する情報科学的手法の開発が望まれていた。

RBPはモチーフ配列のみならず結合するRNAのループ構造も認識して結合する(図4A)。そのため、RBPによるRNA構造認識メカニズムを理解するためには、ターゲット領域が形成するループ構造を明らかにする必要がある。しかしながら、RNA二次構造に関するこれまでの情報科学的研究では、ループ構造に関する研究は行われてこなかった。その理由として、各ループ構造を取る確率(構造プロファイル)は統計力学的に定式化できるが、RNAの取りうる構造は長さに応じて指数関数的に増大するため、網羅的に構造を数え上げる事が不可能だったことが挙げられる。以上の背景を踏まえて本研究では、RNAの構造プロファイルを計算するソフトウェアCapRを開発し(図4B)、その手法を用いてRBP結合領域のRNAループ構造を明らかにする事を本研究の目的とした。

CLIP-seqデータセットから得られたRBP結合部位にCapRを適用する事で、RBP結合領域の構造プロファイルを解析したところ、脆弱X症候群の原因遺伝子であるFMRPが、Internal/Bulge に特異的に結合する事を初めて明らかにし(図4C)、FMRPのInternal/Bulgeとの相互作用の阻害が神経疾患の原因となっている可能性を示唆した。また、傍腫瘍性神経症候群に関与する遺伝子であるNovaでは、モチーフ領域の5'がExterior loopまたはMultibranch loopを形成する必要がある事を示した。NovaとRNAの共結晶構造を解析した所、モチーフ領域5'端はNovaのC末端アミノ酸残基と結合している事がわかった。このC末端アミノ酸残基を欠失させるとNovaがRNAに結合しなくなる事が実験的に明らかになっている事から、Novaが正しい結合サイトを認識するために結合サイトの5'端の二次構造はExterior loopまたはMultibranch loopを形成しているのではないかと考えられた。これらの成果により今後、RBPの変異と神経疾患機序の関係性の理解、そして行動の分子生物学的メカニズムの理解が進展すると期待される。

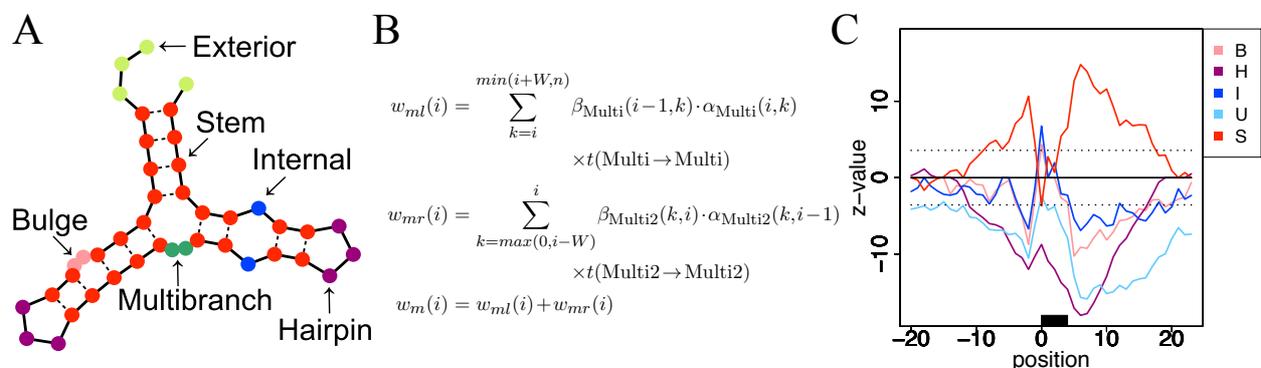


図4 (A) RNAの取る六種のループ構造 (B) 構造プロファイルを求めるアルゴリズム (C) FMRPはInternal/Bulge loop特異的に結合する