

Web ベース統計解析システムの細胞遺伝学への応用  
A web-based statistical analysis system for Cytogenetics.

宮本 句子\*, 宿久 洋\*, 末永勝征\*\*

\*鹿児島大学理学部, \*\*鹿児島純心女子短期大学

Junko MIYAMOTO, Hiroshi YADOHISA, Katsuyuki SUENAGA

Faculty of Science, Kagoshima University, Korimoto, Kagoshima, JAPAN

Kagoshima Immaculate Heart College, Toso, Kagoshima, JAPAN

## 1. はじめに

生物の細胞内にある染色体(クロモソーム)の形をS-PLUSを用いて統計解析するWebベースシステムについて報告する。

真核生物の細胞の核内には,生命の設計図である遺伝子を含むDNAが蛋白質と結合したクロマチンという状態で存在する。細胞分裂が起こる時,クロマチンは何重にも巻き上がってコンパクトな桿状または円柱形にまとまる。これを染色体(クロモソーム)といい,顕微鏡下で可視的に観察し画像として記録することができる。染色体の本数や大きさや側面から見た形は生き物の種類によって一定である。例えばヒトの体細胞には大小22対の常染色体と1対の性染色体の合計46本の染色体が含まれるのに対して,ゴリラは23対の常染色体と1対の性染色体の合計48本の染色体を持つ。どのような長さのどのような形の染色体を何本持っているかという情報は核型(カリオタイプ)と呼ばれ,近縁な生き物では,基本の数や形に類似点が見られるため,生物の進化の研究に利用されてきた。しかし,いろいろな生物間のカリオタイプを比較するには,経験を積んだ研究者の職人技や直勘に頼る面が大きかった(図1)。

近年,顕微鏡写真撮影機材のデジタル化や画像の自動計測プログラムおよび描画ソフトウェアの汎用化によって,カリオタイプの模式図(イデオグラム)を描くためにコンピュータを利用する研究者や技術者が増えた。特に遺伝病などを調べる目的でのヒトの染色体自動解析のみを指向したコンピュータシステムの開発の試みは世界的にも少なくない。しかし,あらゆる動物や植物でのイデオグラムの比較には,あいかわらずトランプゲームの「神経衰弱」のような絵合わせや,手作業による個々の変量の計測や比較が行われてきた。

私達は,様々な生物の染色体イデオグラムの統計的解析を簡単に行うことができるWebベースのコンピュータシステムがあれば,より客観的に,共通の手順によって比較をすることができるのではないかと考えた。

## 2. システム概要

システムは、Web によるデータのアップロードとダウンロードを行う部分、個々の画像データから解析対象を認識し計測する部分、計測値を統計解析する部分からなっている。システムをサーバ上におき、ユーザは手元の端末から Web を介して利用する。ユーザ側から見ると、手元の PC に比較したい複数のイデオグラムをデジタル画像として取り込んでおき、まずそれらをアップロードする段階、画像から1本1本の染色体の模式図を構成する四角形を認識する段階、計測対象を補正する段階、長さを自動計測する段階、計測した数値を統計解析する段階、計算結果を図表として出力する段階がある。このうち統計解析システムとして S-PLUS を用い、その部分の機能としては、クラスター分析、ヒストグラム作成、散布図作成、要約統計量表作成を含んでいる。

具体的な解析手順は次の通りである。

- 1) 染色体は途中に括れがある円柱形をしている。染色体の括れの両側にある短い方の円柱を短腕、長い方の円柱を長腕という。側面からは2本の角の丸い細長い四角形が括れを挟んで結合しているように見える。1細胞中の染色体を長さの順に並べ、それらの側面図を、四角形等を組み合わせた模式的に描画した図をイデオグラムという。腕に構造的に異なる節がある時は、縞模様等を腕の中に描くこともある。これをデジタル画像として、ユーザ側の端末 PC に取り込む。
- 2) 端末 PC からイデオグラムのデジタル画像をシステムへアップロードする。
- 3) システムが自動認識した四角形をカラー化して表示する。このとき、縞模様があれば最少単位の四角形として認識し、腕ごとに集計して中程度の四角形として認識し、短腕と長腕を合わせた全体を1本の染色体として認識する(図2)。
- 4) 万一、システムが認識し表示した四角形が不適切であった場合は、ユーザ側 PC から、削除、修正、追加する。
- 5) ユーザ側 PC からの指示により、縞模様または腕の長さを自動計測してシステム内に記憶する。
- 6) 計測対象のイデオグラムについて、上記の1)～5)の操作を繰り返す。
- 7) ユーザ側 PC からの指示により、計測したイデオグラム全部のデータに基づいて、基礎統計量の算出とクラスタリングを行う(図3)。
- 8) ユーザ側 PC からの指示により、計測した数値情報を表示させる。
- 9) ユーザ側 PC からの指示により、算出結果のグラフ表示を行う(図4)、さらに指示を行い、結果をユーザ側 PC へダウンロードする。

### 3. まとめ

以上の機能によって、今まで1種類1種類のイデオグラムを描き並べ、人の目で

染色体の形を判断し、どの特徴を比較に用いるかを主観的に決め、染色体の1細胞中の本数、腕の長さ、腕同士の比率などの様々な特徴量をひとつひとつ入力して比較していたが、人手を介さずに自動的に高精度に抽出し、統計解析までの作業を Web を介して実施することができるようになった。

学術的な特徴としては、生命情報学(バイオインフォマティクス)分野では分子や遺伝子の文字情報の解析利用が氾濫している中で、画像から数値情報を引き出して統計解析し、結果を画像で提供するという流れを提示し、より視覚的な面を強調したこと、細胞分類学という極めて古典的分野とコンピュータサイエンスという現在もっとも進展速度が著しい分野の融合的課題を具体的なシステムとして実現したことである。

技術面でも、学術面でも、生命科学系と情報科学系の共同作業、あるいは企業と大学の共同作業があっただけで実現したシステムでもある。高度な統計解析機能やグラフ出力がS-PLUSの利用で容易に実装できたことで、今後の発展的改良も機能面、生産性の両面で可能性がある。DHTML技術によりユーザがWebページ上の画像を動的に操作できる対話的な環境を実現できたので、ユーザは遠隔地から手元のデータを簡単にサーバにアップロード、ダウンロードして解析でき、自分のPCには解析システムを置く必要がない。また、四角形の自動認識とその計測や統計処理と連動した部分については、染色体模式図だけでなく、植物の葉や茎の四角形の細胞や、さらには航空写真におけるビルディングの配置や、その他の画像内の様々な箱状の物体の平面的配置の統計解析にも応用可能と考えられる。

最後に、当該システムの開発に当たっては、株式会社数理システムS-PLUSグループ、および南日本マイクロコンピュータ株式会社に技術的協力をいただいた。ここに記して謝す。本稿において紹介したシステムは、平成12年度～14年度文部科学省科学技術振興調整費総合研究において発表者等によるデータベースシステム開発の際に着想を得、平成15年度鹿児島大学理学部奨学寄付金により構築したシステムの機能に基づく。

図1. 植物染色体顕微鏡画像と従来のハンドライトイデオグラム.

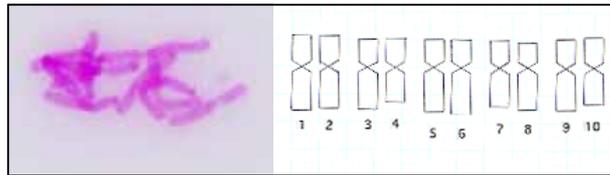


図2. 自動認識した四角形の表示例.

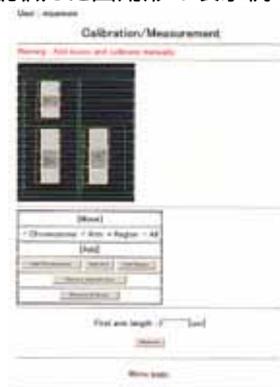


図3. 統計解析画面の例.



図4. クラスタリング結果のグラフ表示の例.

