

ゲノム解析初心者のための支援アプリケーションの提案

Proposal of support application for genome analysis beginners

佐々木 郁人†
Ayato Sasaki大沢 勇統†
Yuto Ohzawa高橋 篤‡
Atsushi Takahashi大星 直樹†
Naoki Ohboshi

1. 序論

近年、ゲノムシーケンシング技術の発展により、ゲノム解析では数十億もの大規模な塩基配列情報を扱う必要性が高まってきており、これらの情報を効率的に扱うには情報処理技術が必要不可欠である。しかし、そういったデータを扱うことができる研究者は少なく、人材を教育する環境や体制も十分に整っていない。

人材不足を解決する試みとして、2014年9月、バイオサイエンスデータベースセンター (NBDC) によるバイオインフォマティクス人材育成カリキュラム¹⁾が実施された。しかし、セミナーは不定期開催であるため、定常的にゲノム解析について学習する機会が提供されておらず、より解析について学習するためには個人で環境を用意する必要がある。

現状、ゲノム解析における学習手段として、入門書が挙げられるが、計算機についての操作の基礎から学習する必要があり、環境構築が複雑といった問題がある。また、ゲノム解析において使用されるツールのほとんどがコマンドラインによる入力によって解析を行うものである。そのため、生物学を専門とする研究者などコンピュータの操作に慣れていないユーザに対してハードルの高さを感じさせてしまい新たにゲノム解析について学習する際の足枷となっている。

本研究では、コンピュータの操作に不慣れな研究者においても、簡単な操作で視覚的にゲノム解析が学習可能なアプリケーションを提案する。

本稿では、開発アプリケーションにおける「解析ツール選択画面」を用いて学生10名に対して評価実験を行った結果、得られた知見についても報告する。

2. 開発アプリケーション

本アプリケーションは、ゲノム解析をより簡単な操作で視覚的に理解することを目的としたGUIアプリケーションである。ゲノム解析にて使用するツールの選択から実行までをマウス操作のみで行う。ツール選択の際に処理の内容をアニメーションで表示することにより、データに対して施す処理の内容をユーザが理解できる機能を実装した。

現在、データ解析環境としてOS Xが広く活用されているため²⁾、Macでの使用を想定して開発した。また、開発環境にはPythonおよびそのGUIツールキットであるwxPythonを採用した。

2.1 アプリケーション構成

本アプリケーションは「解析ツール選択画面」、「ワークフロー作成画面」、「ワークフロー実行画面」の3つから構成されており、タブ操作にて各画面へ切り替える。以下にそれぞれの画面について説明する。本アプリケーションにおける操作の流れを図1に示す。

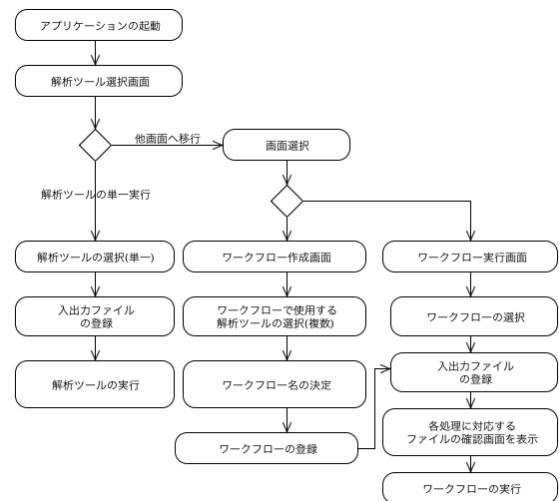


図1. アプリケーションのフロー図

・解析ツール選択画面

解析ツール選択画面(図2)では、リストボックスよりゲノム解析で使用するツールとその処理の組み合わせ(16種類)を選択し、実行できる。

ゲノム解析で、より精度の高い解析を行うためには、どのような処理が施されているのかを理解することは重要である。そのため、ツール選択の際にその処理の内容を説明文およびアニメーションとしてユーザに与える機能を用意した。

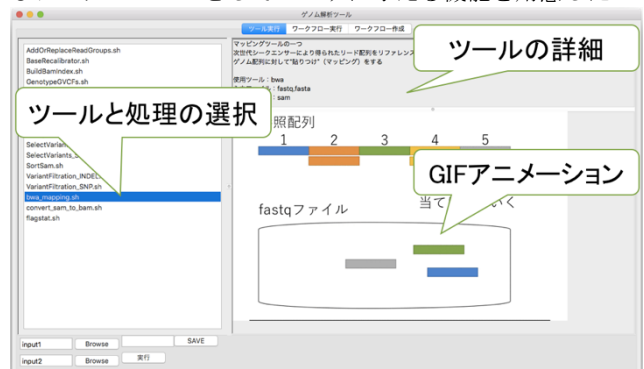


図2. 解析ツール選択画面

・ワークフロー作成画面, ワークフロー実行画面

ゲノム解析において、解析全体の流れを理解することも精度の高い解析のために重要である。そのため、処理の順序をワークフローとして保存するワークフロー作成画面、それを実行する機能をワークフロー実行画面として用意した。保存の際にフローの詳細をコメントとして保持することが可能となっている。

† 近畿大学大学院総合理工学研究科

Graduate School of Science and Engineering, Kindai University

‡ 国立循環器病研究センター

National Cerebral and Cardiovascular Center

3. 評価実験

3.1 実験概要

開発アプリケーションの「操作面」および「学習面」について評価実験を行った。実験の手順は以下の通りである。

1. ゲノム解析についての簡単な説明
2. アプリケーションの操作方法を説明
3. ゲノム解析演習
4. 各処理の内容を答える確認問題の解答
5. 評価シートにてアプリケーションを評価

ゲノム解析についての簡単な説明

- ・ゲノムとはなにか
- ・ゲノム解析ではなにを解析するのか
- ・解析の結果がどのように貢献されるのか
- ・実際にゲノム解析を行う場合の流れ
- ・実験で行う演習問題の説明

以上の5項目についてのプレゼンテーションを行った。

アプリケーションの操作方法を説明

ゲノム解析の開始作業であるマッピングを練習問題として与え被験者とともにアプリケーションを操作しながら解答してもらった。

ゲノム解析演習

ゲノム解析操作を問題形式で被験者に問題シートとして与え、取り組んでもらった。被験者の負荷を考慮し、実験にて行う処理を、実装した16種類の内5種類に限定した。

各処理の内容を答える確認問題の解答

被験者がツールの動きを正しく理解できているのかを確認するため、確認問題に解答してもらった。

評価シートにてアプリケーションを評価

評価シートでは学習面での支援効果を評価するため、

- ・練習問題は難しかったか
- ・確認問題は難しかったか

を3段階で

- ・処理の内容を理解できたか（理解度）
- ・アニメーションはわかりやすかったか

を5段階で、また、操作面での評価のため

- ・アプリケーションの操作性

を5段階で評価してもらった。

最後に本アプリケーションについて主観的な意見を自由記述にて回答してもらった。

3.2 実験結果

評価実験の結果について以下にまとめる。

- ・操作性に関する回答

操作面に対する評価を表1に示す。表1から評価結果の平均は4.1と高い数値を得た。また、自由記述において「使いやすかった」との回答を得ることができた。

表1. 操作面に関する評価の集計

	平均	標準偏差	中央値
操作面に対する評価	4.1	1.3	4.5

- ・学習面に関する評価

確認問題の正答率を表2にまとめる。表2の結果から処理1,3に関しては高い正答率を得ることができたが、処理2,4,5に関する正答率が低く、処理4,5においては理解度及びアニメーション評価のどちらにおいても他に比べ標準偏差が大きくなっている。

表2. 各処理におけるアンケート結果の集計

	処理1	処理2	処理3	処理4	処理5
正答率	70%	40%	100%	30%	40%
理解度(平均)	4.3	4.2	4.3	3.4	3.6
理解度(標準偏差)	0.78	0.75	0.64	1.01	1.02
アニメ評価(平均)	4.6	4.5	4.4	3.7	3.6
アニメ評価(標準偏差)	0.49	0.67	0.91	1.10	1.23

4. 考察

操作面に対する評価および自由記述の結果から、開発アプリケーションは「ゲノム解析を簡単な操作で実行する」という操作面に対する支援は十分にできているといえる。しかし、確認問題の正答率にばらつきがあることから「視覚的に処理の流れを理解」という学習面に対する支援が十分ではないと考えられる。特に処理2（インデックス作成）においては理解度及びアニメーション評価が高いにもかかわらず、正答率が低くなっている。また、処理4,5（ローカルリアライメント）においては理解度及びアニメーション評価の標準偏差が大きいため改善が必要と考えられる。以下にそれぞれのツールにおける正答率低下の原因をまとめる。

処理4,5では他のツールに比べ理解度及びアニメーション評価が低く、標準偏差も大きくなっている。これは、自由記述にて、「アニメーションに専門用語が多くわかりにくい」、「アニメーションが早すぎる」という意見が見受けられたことから、アニメーションによる理解が困難であったことが原因として考えられる。その結果、他に比べアニメーション評価および理解度が低下したと考えられる。また、処理2の解析データに対するインデックスファイル作成では、理解度およびアニメーション評価は高いが、正答率が低い結果となっている。原因として、アニメーションでデータそのものにインデックスを付加するような表現を用いてしまった結果、ユーザに本来の動作とは異なった理解を促したと考えられる。

以上の結果から、開発アプリケーションは現段階では学習面に対する支援という点において改善が必要であると結論づけた。

5. まとめ

本研究では、計算機の扱いが不得意な研究者においても簡単な操作でゲノム解析が実行でき、視覚的に処理の内容を理解できることを目指したアプリケーションを提案した。その結果、提案アプリケーションは操作面に対する支援はできているが、学習面に対する支援が十分でないことが確認できた。今後の課題として、アニメーションの改善や提供する処理の動作情報をアニメーションだけでなく、他の方法を用いて情報提供することが挙げられる。

参考文献

- 1) National Bioscience Database Center, 人材育成講習会 バイオインフォマティクス人材育成カリキュラム, <https://events.biosciencedbc.jp/training/ajacs47>(参照 2017.07.24)
- 2) 清水厚志, 坊農秀雅:次世代シークエンサーDRY 解析教本, pp 17 (2015).