

# 生理・代謝機能ポテンシャル評価システム Genomaple-ES4 の運用

課題責任者

古市 幹人 海洋研究開発機構 付加価値情報創生部門 数理科学・先端技術研究開発センター

著者

山岸 保子\*<sup>1</sup>, 古市 幹人\*<sup>1</sup>, 石渡 千恵子\*<sup>1</sup>, 高見 英人\*<sup>1,2</sup>

\*<sup>1</sup> 海洋研究開発機構 付加価値情報創生部門 数理科学・先端技術研究開発センター, \*<sup>2</sup> 株式会社 C02 資源化研究所

キーワード：微生物, メタゲノム, 生理代謝機能, 機能ポテンシャル解析, web application

## 1. 序論

海洋研究開発機構が開発した Genomaple は、ゲノムデータベースを検索することで、ゲノム・メタゲノム配列から様々な環境に生息する微生物(群)が有する生理・代謝機能ポテンシャルを評価する解析システムである。機構では、本システムを所内大型計算機 DA システムに搭載し、webUI を提供することで外部に公開していた。Web から容易に解析が可能であったため、46 カ国、約 400 人を超える研究者によりゲノム・メタゲノム解析に活用されていたが、機構で発生したセキュリティインシデントの影響により 2021 年からその公開は中断を余儀なくされた。一方で、メタゲノム解析を主要な解析手法とした論文は年々増加している。そして Genomaple システムの開発において発表された解析原理に関する論文[1]、並びに解析システムの自動化や解析例に関する論文[2]、さらにシステムの高速化に関する論文[3]は、現在までに数多くの研究論文で引用されており、その注目度が高いことがわかる。よって Genomaple の再公開は、バイオサイエンスにおいて極めて重要と考えられる。そのため本課題では、再公開を目指して、Genomaple を地球シミュレータ (ES4) に移植して、Web 関連システムを含めて様々な改修を行った。その結果、2023 年 7 月に Genomaple-ES4 として再び公開することが出来た。Genomaple-ES4 がどのようにゲノム・メタゲノム解析に使用されるのか、その役割の概念を図 1 に示す。現在は運用を行いながら本システムを全世界のバイオサイエンティストに提供している。

## 2. Genomaple-ES4 の性能

本課題では ES4 へ移植を行う上で以下の点に留意した。一つ目は、マルチアーキテクチャシステムである ES4 の特性を活かすこと。二つ目は外部公開を前提としているためセキュリティを強化することである。一つ目の ES4 の特性を活かすために GPU を活用するための改修を行なった。その結果、データ処理を高速化し同時処理データ数も増加することができた。これは DA システムと比較して約 2 倍の高速化であり、また処理が可能なデータ数の限界値も 300 万配列から 1000 万配列と大幅に上昇した。二つ目のセキュリティの強化であるが、まずユーザーのログインに対して二要素認証システムを導入することで不正ログイン対策を行なった。また内部のシステム構成として、Web システムを搭載したフロントサーバと ES4 の間に中継サーバを設置することで、外部と ES4 のネットワーク領域を完全に遮断した。また、フロントおよび中継サーバの管理体制を厳しくするなど、様々な対策を行なった。GPU の利用やセキュリティ対策以外にも、ユーザーの利便性を高めるための様々な改修を行なった。一例を挙げると、システムに投入された各クエリ配列がどのような代謝・生物機能に関与するのかわかりやすく整理するとともに、ユーザーがそれらの関連データファイルをダウンロード出来るようにした。計算結果の UI 上の表示の改良も行なった。そして、既知の不具合についても修正を行なった。

## 3. 利用者数の推移

様々な改修を経て 2023 年 7 月に ES4 に搭載した Genomaple システムを外部に一般公開した。DA システム上で稼働していた 2 年半の運用では月平均 200 件の利用実績があり、前述の通り登録ユーザー数は 400 人以上であった。一方、Genomaple-ES4 における 2023 年 7 月から 12 月までの新規登録ユーザーは約 30 人、総ユーザー数は約 300 人弱であり、ES4 の定期メンテナンスなどが無かった月の利用実績は約 120~140 件であった。DA システム搭載時と比較すると、サービス再開時に非アクティブアカウントの整理を行った結果総ユーザー数が 100 名ほど減少していることを踏まえると、利用頻度は中断前とほぼ遜色がなく、サービスに対する需要が依然として高いことが伺える。

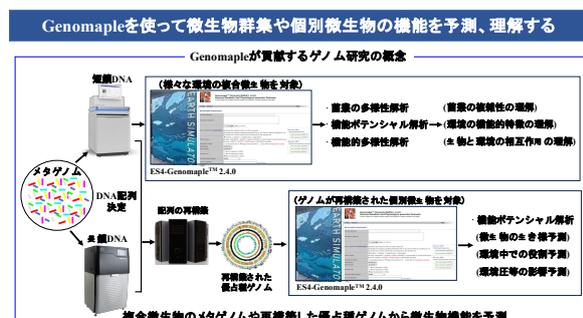


図 1. Genomaple-ES4 のゲノム・メタゲノム解析における役割

#### 4. ES4 の活用事例としての Genomable-ES4

Web システムを備えた Genomable-ES4 は、ES4 を Web アプリケーションのバックエンドサーバとして利用する初のシステムである。よって、ES4 の新たな利用手段を提示したと考えられ、ES4 の活用促進の一助になると思われる。一方で、全体のシステム構成は複雑化し複数のネットワーク領域を横断している。そのため、移植改修から運用に移行した Genomable-ES4 システムであるが、挙動が不安定になる場合がある。よって現在は、不安定挙動の原因の追及と、システム間の通信を減らすための改修を行なっている。不安定挙動の原因が明らかになれば、ユーザの利便性は高まり利用実績もさらに向上すると考えられる。また、今後同様に ES4 を Web アプリケーションのバックエンドとして利用したいシステムがあれば、その開発にも貢献が可能である。

#### 5. Genomable-ES4 が創生する学術成果

DA システム上での稼働時を含めて Genomable システムを利用して数多くの研究成果が発表されており、今年度も複数の論文が発表された[4-5]。近年爆発的に増加する自然環境や腸内環境などのメタゲノム配列とメタゲノム配列から再構築された個別生物種のゲノム配列解析や比較解析には、莫大な計算資源が必要となっている。この状況下において本課題で公開・運用する Genomable-ES4 は、自然科学のあらゆる分野で用いられるゲノム、メタゲノム解析、特に機能的ゲノム、メタゲノム解析に必要な計算資源と計算システムを提供し、バイオサイエンスの発展に大きく寄与すると考えられる。

#### 謝辞

Genomable システムの ES4 への移植並びに Web 関連システムの再構築および外部公開は、当機構の付加価値情報創生部門・地球情報科学技術センターの中川剛史氏と情報セキュリティ・システム部情報セキュリティ統括課の古川向直氏並びに豊村鉄男氏の多大な協力を得て実施した。ここに記し深く感謝する。

#### 文献

- [1] H. Takami et al., “Evaluation method for potential functionome harbored in genome and metagenome”, *BMC Genomics* 13, 699 (2012)
- [2] H. Takami et al., “An automated system for evaluation of the potential functionome: MAPLE version 2.1.0”, *DNA Res.* 23, 467-475 (2016).
- [3] W. Arai et al., “MAPLE 2.3.0: An improved system for evaluating the functionomes of genomes and metagenomes”, *Biosci. Biotech. Biochem.* 82(9), 1515-1517 (2018).
- [4] T. Shiozaki et al., “Distribution and survival strategies of endemic and cosmopolitan diazotrophs in the Arctic Ocean”, *ISME J* 17, 1340-1350 (2023).

- [5] S. Jiang et al., “Variations in Physiology and Genomic Function of *Prochlorococcus* Across the Eastern Indian Ocean”, *J. Geophys. Res. Oceans* 128(10), 2169-9275 (2023).