

平成 30 年 6 月 1 日現在

機関番号：82401

研究種目：挑戦的萌芽研究

研究期間：2016～2017

課題番号：16K12529

研究課題名(和文)多階層ゲノクスデータ流通におけるアジア拠点の構築

研究課題名(英文)Development of a genomics data sharing platform in Asia

研究代表者

川路 英哉(Kawaji, Hideya)

国立研究開発法人理化学研究所・情報基盤センター・開発ユニットリーダー

研究者番号：20525406

交付決定額(研究期間全体):(直接経費) 2,800,000円

研究成果の概要(和文):日本を含むアジア地域での大規模ゲノクス解析を支援するため、The UCSC Genome Browserデータベースの日本拠点を開発者とともに立ち上げた。立ち上げの後二年間で、月間2-30万回もアクセスされる程度の、日本において最も利用されているライフサイエンス・データベースの一つにまで成長した。

研究成果の概要(英文):We launched a mirror server of the UCSC Genome Browser Database in Japan, in a collaboration with its developer, to support a large-scale genome-wide analysis in Asia. With its stable service in two years, it grows as one of the largest life science databases in Japan, which have twenty to thirty thousands of accesses.

研究分野: バイオインフォマティクス

キーワード: ゲノクス バイオインフォマティクス データベース

1. 研究開始当初の背景

近年の分子生物学における技術革新、特に次世代シーケンサーの登場とその多様な応用技術開発によって、生物の分子の基盤を網羅的に調べるゲノミクス研究に目覚ましい前進があった。これらゲノミクス研究では、生命現象や原理に関する学術論文の他に、研究の過程で取得された網羅的データも科学的な財産として参照され、さらなる研究解析へと用いられている。

次世代シーケンサーより産出された生データは国際レポジトリ(SRA, ERA, DRA)に登録され共有されているが、利用には実験プロセスの適切な理解と大量の計算資源を要する前処理が必須である。したがって、前処理済のデータを共有することによって、より円滑なデータ再利用が実現できる。特に、リファレンス・ゲノムを軸として前処理が行われたデータの共有が広く行われており、その共有方式としてはカリフォルニア大学サンタクルス校の開発する **The UCSC Genome Browser** データベースを用いたものがデファクトスタンダードとなっている。

同データベースの内容を保持するサーバーは米国と欧州に存在するもののアジア地域には存在せず、日本を含むアジアの研究者が同方式で共有されたデータを利用する際に遅延が発生していた。少数のデータファイルを利用する程度の研究であれば大きな障害にはならないが、大規模ゲノミクス解析においては遅延によって大きな影響を受ける。一つの生命現象を読み解く上で、様々な側面のゲノミクスデータを積み重ねる研究が主流になりつつある現在、日本の研究コミュニティにおいてこの遅延の解消が一つの課題であった。

2. 研究の目的

The UCSC Genome Browser データベースの日本拠点を開発者と共に立ち上げる。これによって、利用者の視点に立った実際的かつ開かれたゲノミクスデータ流通を促すと共に、その利活用を推進する。

3. 研究の方法

同データベース拠点を構築する方針として、大きく (i) 公開されているソースコードを利用して独自に拠点を立ち上げる、(ii) 同データベースの開発者と共同して拠点を立ち上げる、というアプローチが考えられる。前者の方法論では、独自の運営方針を設定できるという利点はあるものの利用者への混乱を生みかねないこと、運用体制の確立と安定的な運用を実現する為のコストが大きくなること、データベースの更新に伴う遅れが

どうしても発生すること、等の不利益が考えられる。ユーザにとっても運営の観点からも後者の方法論が望ましいが、開発・運営者らと連携した国際的な共同研究の枠組みを立ち上げることが困難を解決することが必要である。この二者のうち、より困難な後者のアプローチによる立ち上げを試みた。

同データベースを開発・運用している研究者と独立に運用するのではなく、共同して運用する仕組みを模索する。これにより、運用にかかるコストを削減すると共に、サポートする機能の充実と安定的な運用を目指す。

4. 研究成果

(1) 立ち上げ

同データベースを開発・運用している研究者と議論し、正式な共同研究の枠組みを立ち上げた。その後運用方針の概要を定め、サーバーの導入とシステム環境の構築を行った上で、同データベースに関するアジア・ミラーの立ち上げを行った。本立ち上げに関しては、本サイトにおいても言及され広く周知が図られた(図1)

Jun. 15, 2016 New supported Asian mirror site released

We are pleased to announce the introduction of a new supported Genome Browser mirror to serve our users in Asia. The genome-asia server (<http://genome-asia.ucsc.edu>), which is administered by UCSC, is physically located at the **RIKEN Yokohama Campus** in Japan.

Genome-asia is intended as an alternate, faster access point for those Browser users who are geographically closer to Asia than to the western United States. The mirror is functionally identical to the U.S.-based server, with one exception. Custom tracks residing on the U.S. server are not automatically accessible from the genome-asia mirror; therefore, the use of these custom tracks and saved sessions containing them will require some manual intervention.

Genome Browser users located in Asia will be automatically redirected to genome-asia when they click the "Genomes" menu item on the Genome Browser home page, but will have the option to remain on the U.S.-based server. In addition to the genome-asia mirror, we continue to support a Europe-based Genome Browser mirror (<http://genome-euro.ucsc.edu>) in Bielefeld, Germany. A third-party mirror in Aarhus, Denmark, is available as an emergency backup in the rare event that our three supported sites are offline.

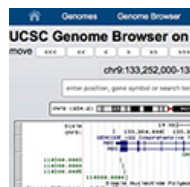
We wish to express our special thanks to Dr. Hideya Kawaji at the **RIKEN Advanced Center for Computing and Communication** for making this possible. We would also like to acknowledge the following members of the UCSC Genome Browser team for their efforts: Galt Barber, Robert Kuhn, Jorge Garcia, and Luvina Guruvadoo.

図1. 本サイトにおける周知

そのほか、本立ち上げに関しては理化学研究所のホームページでも紹介された(図2)

2016年6月17日

[ゲノムブラウザ・データベースの公式ミラーを開設](#)



理研は6月16日、カリフォルニア大学サンタクルス校(UCSC)と共同し、広く用いられているUCSCゲノムブラウザ・データベースの日本ミラーを開設しました。[続きを見る...](#)

図2. 理研ホームページにおける紹介

(2) 運用

2016年の立ち上げ以来およそ二年間の間、安定的な運用を実現できた。この間、次のよ

うな機能更新を実施した。

- ソフトウェア・アップデート (およそ 30 回)
- データ追加や機能拡張 (およそ 50 回)

いずれの更新においてもメイン・データベースと同時に機能更新が実施され、ユーザが気付かない程度に円滑な移行であった。その他、インターネット・プロトコルの中でも当初は HTTP のみのサポートであったが、HTTPS を用いた安全な通信をサポートする機能拡張については、日本側でのネットワーク管理と整合性を図りつつ実現した。

(3) 活用への呼びかけ

本データベース活用を呼びかけるため、ライフサイエンスの研究者が多く集まる分子生物学会、ライフサイエンスにおけるデータベース利活用をテーマにしているトーゴの日シンポジウム、等を中心に 5 件の学会発表を実施した。

(4) 利用例の実証

利用例の実証として、研究代表者らが別個に運用・開発を行っているデータベース (FANTOM5 web resource) のデータに関する共有を、本データベースを用いて実施した。これは 3,000 以上ものゲノクスデータファイルを扱うものであり、その利用が極めて円滑に行えることを示した (図 3、発表論文)。

(5) アクセス数

現在では月間 2~30 万回もアクセスされており、立ち上げ 2 年目にして日本のライフサイエンス・データベースの中でも最も利用されているものの一つにまで成長した。

5. 主な発表論文等

[雑誌論文] (計 1 件)

① Lizio M, Harshbarger J, Abugessaisa I, Noguchi S, Kondo A, Severin J, Mungall C, Arenillas D, Mathelier A, Medvedeva YA, Lennartsson A, Drabløs F, Ramilowski JA, Rackham O, Gough J, Andersson R, Sandelin A, Ienasescu H, Ono H, Bono H, Hayashizaki Y, Carninci P, Forrest AR, Kasukawa T, Kawaji H. Update of the FANTOM web resource: high resolution transcriptome of diverse cell types in mammals. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D737-D743, 査読有

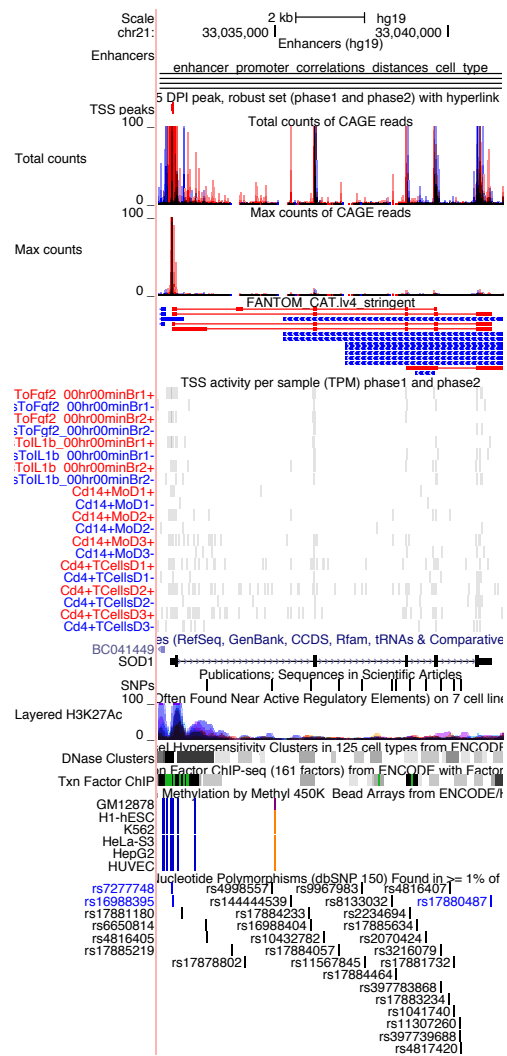


図 3. データベース利用例。横軸にゲノム(染色体)の塩基対が並んでおり、それらに対応するデータが縦に描画されている。上側は FANTOM5 web resource の提供するデータ、下側は The UCSC Genome Browser Database に保存されているデータであり、これらが有機的に統合表示されていることが示されている。

[学会発表] (計 5 件)

- ① 川路英哉、“国際的ゲノクスデータ統合のアジア拠点: The Asian Mirror of the UCSC Genome Browser Database”, トーゴの日シンポジウム 2017(2017)
- ② 川路英哉、“国際的ゲノクスデータ統合のアジア拠点: The Asian Mirror of the UCSC Genome Browser Database”, 2017 年度生命科学系学会合同年次大会 (2017)
- ③ 川路英哉、“国際的ゲノクスデータ統

合のアジア拠点: The Asian Mirror of the UCSC Genome Browser Database”, トーゴの日シンポジウム 2016 (2016)

④ 川路英哉、 “国際的ゲノムクスデータ統合のアジア拠点: The Asian Mirror of the UCSC Genome Browser Database”, NGS 現場の会第五回研究会 (2016)

⑤ 川路英哉、 “国際的ゲノムクスデータ統合のアジア拠点: The Asian Mirror of the UCSC Genome Browser Database”, 2016 年度生命科学系学会合同年次大会 (2016)

[図書] (計 0 件)

[産業財産権] (計 0 件)

[その他]

① 本研究で立ち上げたミラーサイト
<http://genome-asia.ucsc.edu/>

② 本データベースのメインサイト
<http://genome.ucsc.edu/>

6. 研究組織

(1) 研究代表者

川路 英哉 (KAWAJI, Hideya)

国立研究開発法人理化学研究所・
情報基盤センター・開発ユニットリーダー
研究者番号: 20525406