

厚生労働省科学研究費補助金 食品の安全確保推進研究事業
「広域食中毒発生時の早期探知のための調査の迅速化及びゲノム解析技術を利用した
調査法の確立に資する研究」
(23KA0501)
研究分担報告書

分担研究課題

「2023 年度 EHEC 検出状況と EHEC の WGS 解析パイプライン構築」

研究分担者 明田幸宏 (国立感染症研究所 細菌第一部)

研究協力者 泉谷秀昌 (国立感染症研究所 細菌第一部)

李 謙一 (国立感染症研究所 細菌第一部)

柿田徹也 (沖縄県衛生環境研究所)

研究要旨

腸管出血性大腸菌 (*enterohemorrhagic Escherichia coli*: EHEC) のサーベイランスにおいて、現在、主に反復配列多型解析 (multi-locus variable-number tandem repeat analysis: MLVA) 法が使われている。MLVA データを基盤とするため継続的に全国の EHEC 分離株を解析した。MLVA 法により解析した 2023 年分離株は 2923 株であった。また、血清群 O157、O26、O111 については地方衛生研究所から直接 MLVA データが送付され MLVA 型の付与を実施した。今年度データを解析し型名を付与した株は約 1500 であった。

広域食中毒の調査に全ゲノム配列解析を用いるためには、標準的な解析法の整備が必要である。しかし、単一塩基多型 (SNP) 解析等の情報解析手法の共有化には、プログラムの依存関係等の問題があった。そこで、docker というプラットフォームを用いて、異なる OS 環境でも容易に使用するプログラム群をインストール可能な docker コンテナを構築した。同コンテナを用いることで、データのクオリティチェックから系統樹解析までを容易に実行可能なパイプラインを構築した。

A. 研究目的

腸管出血性大腸菌 (*enterohemorrhagic Escherichia coli*: EHEC) のサーベイランスではこれまでパルスフィールドゲル電気

泳動法 (PFGE) が主要な解析手法であったが、2018 年 6 月 29 日付の厚生労働省事務連絡「腸管出血性大腸菌による広域的な感染症・食中毒に関する調査について」

(2023年6月28日再周知)により、反復配列多型解析 (multi-locus variable-number tandem repeat analysis : MLVA) 法が血清群 O157、O26、O111 の統一手法として用いられている。本研究では全国の MLVA による解析結果の総括並びに、事務連絡に基づいて送付された地方衛生研究所からの MLVA データについて解析を行った。

腸管出血性大腸菌 (enterohemorrhagic *Escherichia coli*: EHEC) は広域食中毒の原因となるため、現在国内では分子疫学手法を用いたサーベイランスが行われている。分子疫学手法としては、8種のO群 (O157、O26、O111、O103、O121、O145、O165、O91) では multi locus variable tandem repeat analysis (MLVA) が、その他のO群ではパルスフィールドゲル電気泳動 (PFGE) 法が用いられている。近年、高速シーケンサー (NGS) の実用化により、集団感染等の調査に全ゲノム配列 (whole-genome sequence : WGS) を用いた解析が取り入れられつつある。実際の解析時には、NGS で WGS を解読した後に、生データのクオリティチェック (quality check, QC) を行い、その後 single nucleotide polymorphism (SNP) 解析等を行う必要がある。そのために用いるソフトウェアや解析サーバーは数多く存在するが、使用方法やデータの評価基準等が記された体系的なマニュアルは存在せず、精度管理上の課題がある。さらに、Linux OS で動作するソフトウェアの場合、端末によってインストールが困難な場合がある。このため本分担研究では、地方自治体等における集団感染調査の際に特に必要とされる、データの QC、アセンブルおよび SNP

解析を自動で行える解析パイプラインの構築と、同パイプラインの docker コンテナ化を目的とした。

B. 研究方法

MLVA 法の有効性の検証・精度管理手法の確立

感染研に送付された腸管出血性大腸菌 2023 年分離株に対して MLVA 法により解析した。方法は Izumiya ら (2008、2020) の方法に従って実施した。血清群 O157、O26、O111 については 17 か所、O103、O121、O145、O165、O91 については 43 か所の遺伝子座を用いた。地方衛生研究所から MLVA 型付与のために送付された MLVA データ (血清群 O157、O26、O111) も併せて解析を行った。

EHEC ゲノム解析パイプライン (データ QC、アセンブル、SNP 解析パイプラインの構築)

これまでに感染研・細菌第一部で使用実績のあるプログラムを中心に、解析パイプラインの構築を行った。アセンブリに関しては、Guide to bacterial genome assembly

(<https://github.com/rrwick/Tricycler/wiki/Guide-to-bacterial-genome-assembly>) を参考に、広く用いられている SPAdes と SKESA の両方を使用可能とした。QC に関しては、CDC PulseNet が公表している SOP

(https://www.aphl.org/programs/food_safety/Pages/PulseNet-International-SOPs.aspx) を参考に、カバレッジやコンタミネーシ

オン等をチェックする機能を加えた。SNP 解析に関しては、高精度な解析プログラムである BactSNP と、迅速な再解析が可能な Snippy を組み合わせた新たなパイプライン (SNPcaster) を構築した。加えて、遺伝子組換えについて検出できる Gubbins も使用できるようにした。

上記の解析パイプラインについて、docker コンテナ化を行った。Docker は、アプリケーションを開発、配布、実行するためのプラットフォームである。コンテナと呼ばれる仮想化技術を利用して、アプリケーションとその依存関係をパッケージ (コンテナ) 化することで、異なる OS 環境でも同じプログラムを実行できるようになる。上記解析パイプラインについて、docker コンテナ化を行うと共に、インストールおよび操作マニュアルの作製を行った。

地衛研におけるゲノム解析

沖縄県において 2006 年から 2024 年までに収集された腸管出血性大腸菌株のうち、NESID 情報が取得可能な菌株を対象とした。ゲノム DNA は DNeasy Blood & Tissue kit (Qiagen) にて抽出し、Qiaseq FX DNA library kit (Qiagen) にてライブラリを作成した。iSeq 及び Novaseq (外部委託) にてペアエンドシーケンシングを実施した。得られたリードについて、shovill にてアセンブルを実施した。

C. 研究結果

MLVA 法の有効性の検証・精度管理手法の確立

感染研細菌第一部において、3308 株について分子型別解析を実施した。このう

ち 2923 株について MLVA 法による解析を実施した。解析依頼施設数は延べ 94 施設であった。各血清群において同定された型数は、O157 が 862、O26 が 148、O111 が 72、O103 が 62、O121 が 16、O145 が 12、O165 が 3、O91 が 38 であった。得られたデータは 2024 年 5 月号の IASR の EHEC 特集号において公表される。

MLVA 型別を実施しデータを送付した地方自治体は、48 施設であった。MLVA データを送付し、感染研において統一型名を付与した菌株数は約 1500 株であった。

2023 年度は 55 施設から約 1500 株分のデータについて照会があり、感染研で精度確認を行った (上記型名付与を行った株のデータを含む)。全遺伝子座における正解率 (平均値) は約 98% であった。株毎の正解率 (平均値) は約 89% であった。

EHEC 菌株解析パイプライン (データ QC、アセンブル、SNP 解析パイプラインの構築)

解析パイプラインについては、感染研・細菌第一部内および九州大学 中村佳司博士の協力のもと、機能の評価や改善を行った。その結果、事前に設定した値 (カバレッジやコンタミネーションの割合等) に対して自動的に QC を行い結果を表示する機能を追加した (表 1)。また、SNP 解析に関しては、結果の評価に必要なコアゲノムサイズ等を自動的に計算してレポートを出力する機能を追加した (図 2)。

構築したパイプラインについて、docker コンテナ化を行った。Windows および Linux 環境下でインストールおよび

動作確認を行った。通常インストールにはエラー対応を含め数時間を要するが、**docker** を利用した場合には一時間程度でインストール可能であった。プログラムの実行には、**JupyterLab** を利用した。**JupyterLab** は、ウェブブラウザ上で動作するため、OS に関わらず利用可能である。また、ノートブックと呼ばれるファイル内ではコードの編集、実行および結果の表示が可能である。**JupyterLab** のノートブック上でマニュアルと解析コードを一体化したファイルを作製することで、情報解析初心者でも容易に実行可能な環境が構築された (図 3)。

地衛研における EHEC 菌株のゲノム解析

R5 年度において沖縄県で分離され、NESID 情報が付加されている腸管出血性大腸菌株 322 株についてゲノム配列を取得した。アセンブル方法を検討するため、**shovill** にて 22 株のアセンブルを実施した。completeness 99.93 以上、contamination 0.43%以下の良好な結果を得た。これらの 22 株のドラフトゲノム配列は DDBJ に登録した。

D. 考察

MLVA 法により解析した菌株数は昨年より約 28%と大幅に増加した。解析結果は定期的あるいは不定期に厚生労働省 NESFD に MLVA リストとして掲載された。地方衛生研究所から型名付与のために送付された MLVA データは約 1500 株に上った。感染研で行った精度確認では菌株ベースで 89%、遺伝子座ベースで 98%一致しており、多くの地衛研において

技術的な問題が解消されつつあると考えられた。MLVA データ送付にあたっては、地衛研におけるデータの信頼性が重要であり、今後も引き続きモニタリングしていく必要がある。

食中毒事例等の集団感染等の調査で WGS 解析を用いるの情報解析手法は、正確かつ他機関でも利用・検証可能であることが望ましい。SNP 解析に用いるプログラムは数多く利用可能であるが、標準的な解析手法は現在までに確立されていない。その理由としては、SNP 解析には、参照配列におけるリピート配列の検出、コアゲノムサイズの同定や、組換え領域の検出・除去等の複数の解析過程が存在することが挙げられる。これらについて、一度に行えるプログラムは利用可能でなかった。また、プログラム間の依存関係のために、インストールが困難な場合も存在した。本研究では、生データ (WGS リードデータ) の QC から SNP 解析、系統解析までを行えるプログラムを、単一の **docker** コンテナに含めることで、インストール時の課題を解決することができた。今後、地方衛生研究所等に試験的に配布し、評価を行うことで、プログラムおよびマニュアルの改善を行う予定である。

一方で、感染研細菌第一部の協力のもと、地方衛生研究所の協力研究者が自力で DDBJ まで登録できたことは大きな成果であり、この取り組みを更に拡大する試みを進める。

E. 結論

主に SNP 解析を目的とした、WGS 解析プログラムを確立し、OS に依らずイン

ストール可能な docker コンテナ化を行った。

F. 健康危険情報

なし

G. 研究発表

1) 誌上発表

1. Lee K., Iguchi A., Terano C., Hataya H., Isobe J., Seto K., Ishijima N., Akeda Y., Ohnishi M., and Iyoda S. Combined usage of serodiagnosis and O antigen typing to isolate Shiga toxin-producing *Escherichia coli* O76:H7 from a hemolytic uremic syndrome case and genomic insights from the isolate. 2024. Microbiology spectrum 12:e0235523.
2. 泉谷秀昌、李謙一、伊豫田淳、明田幸宏：2022年に分離された腸管出血性大腸菌のMLVA法による解析。IASR、第44巻、72-73、2023年5月
3. Kakita T, Lee K, Morita M, Okuno M, Kyan H, Okano S, Maeshiro N, Ishizu M, Kudeken T, Taira H, Teruya M, Ogura Y, Akeda Y, Ohnishi M, Isolation and whole-genome sequencing analysis of *Escherichia fergusonii* harboring a heat-labile enterotoxin gene from retail chicken meat in Okinawa, Japan. Microbiol Immunol 2024, 68 115-121. Epub 2024 Jan 20.

2) 学会発表

1. 李謙一. 病原細菌における全ゲ

ノム配列解析の基礎. 衛生微生物技術協議会第43回研究会. 岐阜, 2023.

2. 李謙一. EHECにおけるWGS解析. 令和5年度 北海道・東北・新潟ブロック腸管出血性大腸菌検査担当者研修会. 岩手県、2024.
3. 泉谷秀昌、李謙一、伊豫田淳、明田幸宏：腸管出血性大腸菌の MLVA による分子疫学解析（2022年）。第44回日本食品微生物学会学術総会、2023年9月、大阪府堺市。
4. 泉谷秀昌：2023年 EHEC の状況まとめ。令和5年度北海道東北新潟ブロック腸管出血性大腸菌検査担当者研修会、2024年1月、岩手県盛岡市。

H. 知的財産権の出願・登録状況

1. 特許取得

なし

2. 実用新案登録

なし

表 1 自動クオリティチェック (QC) 評価の結果例

strain	no.of_reads	total_read_length	coverage	#contigs	Largest_contig	Total_length	Marker_lineage	Completeness	Contamination	qc_results	qc_coverage	qc_contigs	qc_Total_length	qc_Completeness	qc_Contamination
DRR147067	1,154,412	335,803,419	61.1	181	434,806	5,120,788	f_Enterobacteriaceae(UJD5162)	99.93	0.31	PASS	pass	pass	pass	pass	pass
DRR147069	1,250,569	364,082,817	66.2	204	434,985	5,186,658	f_Enterobacteriaceae(UJD5162)	99.93	0.36	PASS	pass	pass	pass	pass	pass
DRR147075	1,032,502	299,954,394	54.5	189	435,066	5,185,030	f_Enterobacteriaceae(UJD5162)	99.93	0.36	PASS	pass	pass	pass	pass	pass
DRR147079	1,172,467	341,279,434	62.1	227	435,066	5,361,884	f_Enterobacteriaceae(UJD5162)	99.93	0.51	PASS	pass	pass	pass	pass	pass

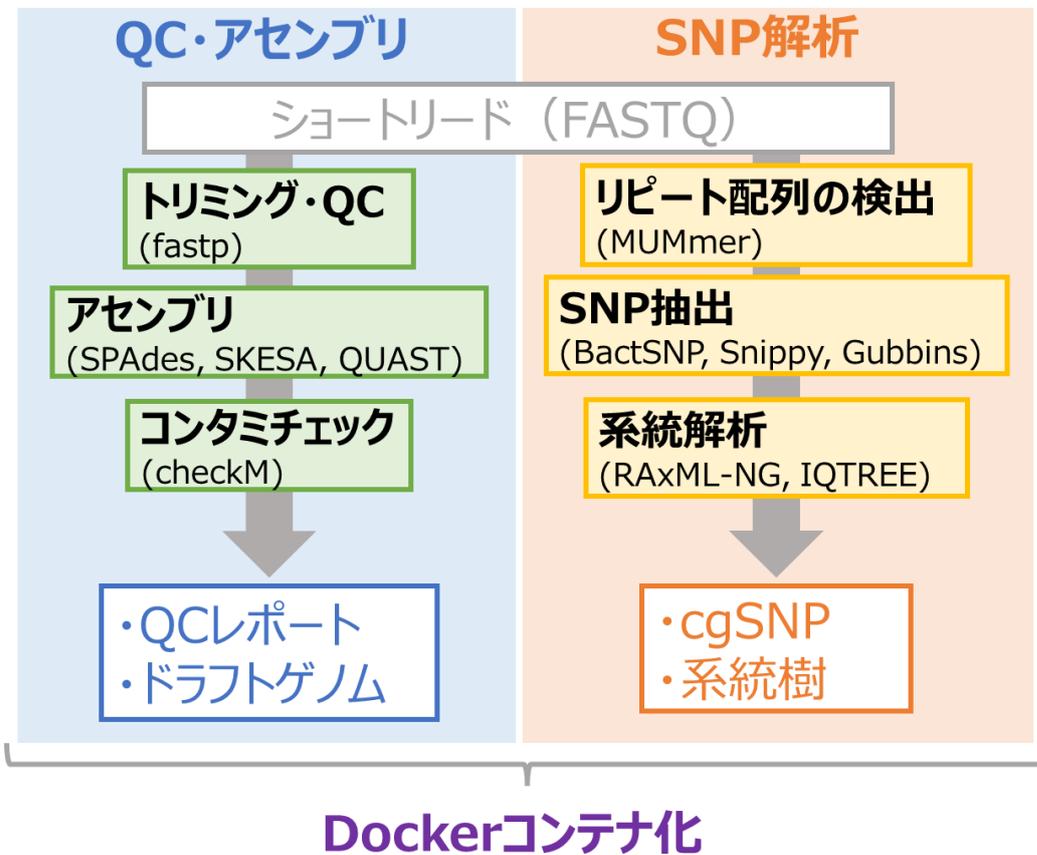


図 1. 解析パイプラインの概要
括弧内は主要なプログラム名を示す。

This is SNPcaster
version :2023-06-05
list_name : list_test
repeats_file_name : repeats.tsv
Reference file name : Sakai_BA000007.fas
Clustered SNP removal (bp) : 10
Recombinogenic region detection (Gubbins) : ON

Removed SNPs in repeat region : 2928
Removed SNPs in recombinogenic region : 3136
Number of informative site before Gubbins : 35781
Number of informative site after Gubbins : 32645
Core_genome size before Gubbins (bp) : 4012769
Core_genome size after Gubbins (bp) : 3803669

図 2. SNPcaster プログラムでの結果レポートの例

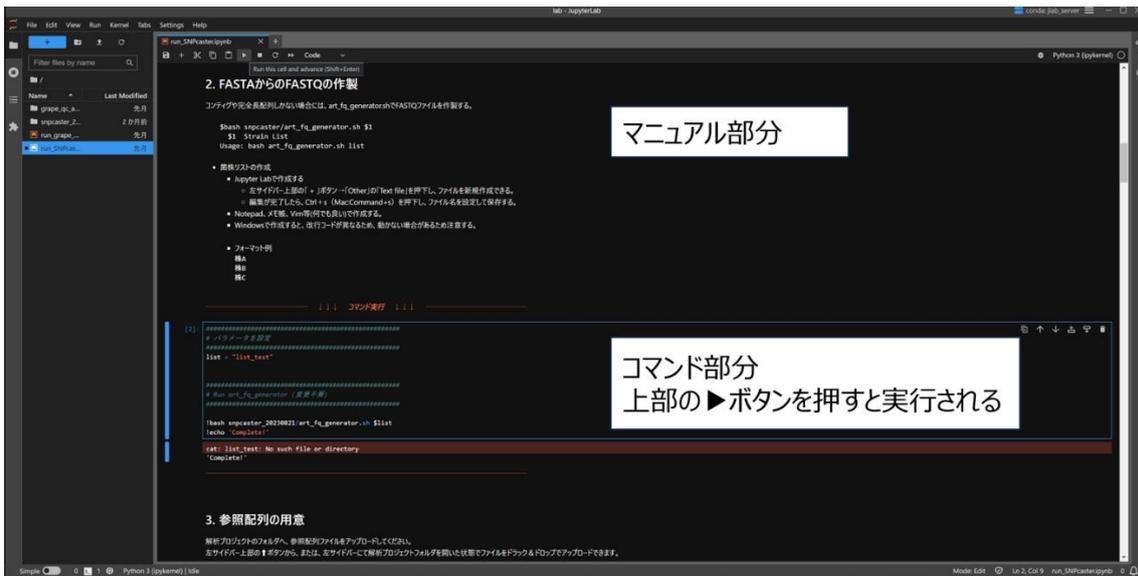


図 3. JupyterLab ノートブック例