

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

「MLVA による EHEC 検出状況とゲノム解析パイプライン構築、評価」

研究分担者 明田幸宏（国立健康危機管理研究機構 国立感染症研究所細菌第一部）

研究協力者 泉谷秀昌（国立健康危機管理研究機構 国立感染症研究所細菌第一部）

李 謙一（国立健康危機管理研究機構 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 2023～2025 年分離株は 8,781 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。3 年間でデータを解析し型名を付与した株は約 4600 であった。また現在は広域食中毒の調査に全ゲノム配列解析の適応が進められており、これを用いるためには、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。試験的な配布を経て、2026 年 3 月には正式版（v1.0）を GitHub にてリリースした。SNP 解析機能のほかに、NCBI AMRFinderPlus による薬剤耐性遺伝子検出機能、およびカスタム GPT を活用した専用サポートチャットボット（SNPcaster-GPT）を実装した。さらに、主として地方衛生研究所職員を対象に利用者講習会を実施し、後日、講習内容の動画配信も行った。また、既存の SNP 解析プログラムとの比較検証を行い、本プログラムが高い精度を有することを明らかにした。

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。さらに、上記以外の O 群について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が、近年、高速シーケンサー（next-generation sequencing: 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）解析が取り入れられつつある。WGS を用いた実際の解析では、NGS よりデータを取得した後、生データのクオリティチェック（quality check: QC）を行い、さらに single nucleotide polymorphism（SNP）解析等を実施する必要がある。これらの解析に用いられるソフトウェアや解析サーバーは多数存在する一方で、使用方法やデータの評価基準等を体系的に記載したマニュアルは十分に整備されておらず、精度管理上の課題がある。さらに、Linux OS 上で動作するソフトウェアについては、利用する端末や環境によってインストールが困難となる場合がある。このため、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解析パイプラインの確立および共有化を進めた。

B. 研究方法

1. 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 2023～2025 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。

2. 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ョ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（図 1）。

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。

また、次の 2 種類の機能を追加した。1 つ目は、NCBI AMRFinderPlus による薬剤耐性遺伝子検索機能である。本機能では、主要な病原菌を対象として、薬剤耐性遺伝子の網羅的な検索を行うことができる。また、大腸菌など一部の菌種については、病原性遺伝子の検索も可能である。2 つ目は、SRA Toolkit を利用した、NCBI のショートリードデータベース (Sequence Read Archive: SRA) からのリードデータダウンロード機能である。本機能により、作成した SRA アクセッション番号リストに基づき、対象データを自動的にダウンロードすることが可能となった。これにより、データベース上に登録されている国内外のゲノムデータと、ローカルで保有するゲノムデータとの比較が可能となる。

3. WGS 解析パイプラインのマニュアル作製および試験的配布

解析パイプラインをインストールおよび操作するためのマニュアルを作

製した。これらのマニュアル、解析パイプラインおよびサンプルデータセットを、希望する自治体等に配布した。

さらに、マニュアルを補完する目的で、生成 AI を利用したチャットボット (SNPcaster-GPT) を作製した。これは ChatGPT のカスタム GPT 機能を利用したものであり、SNPcaster のマニュアルおよびプログラムを読み込ませることで、SNPcaster のインストール方法や使用方法に関する質問に対して、より正確な回答を提供できるようにした。

4. 講習会およびアンケートの実施

配布した解析パイプラインのインストールおよび操作方法の講習会を、2024 年 11 月 27 日および 2025 年 12 月 24 日に国立感染症研究所 戸山庁舎にておこなった。後日、説明動画を YouTube の限定公開機能を利用して配信した。

また解析パイプラインの配布時に同時に配布したアンケート結果の取りまとめを行った。アンケートは、解析に使用した情報解析端末のスペック (OS,CPU 等)、インストールや使用の可否等についての質問を設けた (図 1)。

5. 精度評価

SNPcaster の SNP 検出性能について、主要な SNP 解析ツールと比較することにより評価した。比較対象として、Snippy、Lyve-SET および Parsnp を用いた。まず、シミュレーションデータを用いて、各ツールの感度および陽性的中率 (positive predictive value: PPV) を算出し、比較した。対象菌種は、*Escherichia coli*、

Campylobacter jejuni、*Legionella pneumophila*、*Neisseria gonorrhoeae*、*Staphylococcus aureus* の 5 種とした。各菌種について、SimuG を用いて、参照配列との類似性が異なる 3 条件の配列データを作製し、入力データとした。具体的には、類似性 99.999% (菌株レベルの差異)、99.9% (菌株と菌種の間レベルの差異)、97% (菌種レベルの差異) の 3 条件とした。次に、実データを用いて SNP 数の比較を行った。本解析では、公開されている EHEC のベンチマークデータセット

(https://github.com/leech-rr/WGS_benchmark_dataset) を用いた。本データセットは、EHEC の集団感染事例 3 件 (O157 が 2 事例、O121 が 1 事例、各 5 株) から構成される。参照配列としては、解析対象菌株と系統的に遠い株として EHEC O26:H11 11368 株を用い、近縁な株として、O157 については Sakai 株、O121 については 51104 株の染色体配列を用いた。各条件で SNP 解析を実施し、各事例内における pairwise SNP 数を算出した。

C. 研究結果

1. 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、2023-2025 年分離株 10388 株について分子型別解析を実施した。このうち 8781 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 104 施設であった。各血清群における株数、同定された型数は以下のとおりであった。O157: 6190 株、2299 型; O26: 1120 株、451 型; O111: 464 株、223

型; O103: 631 株、197 型; O121: 90 株 50 型; O145: 112 株、46 型; O165: 14 株 14 型; O91: 160 株、115 型。得られたデータは毎年 5 月号の IASR の EHEC 特集号において公表された。

2023-2025 年分離株について、MLVA 型別を実施しデータを感染研に送付した地方衛生研究所は、延べ 55 施設であった。感染研において統一型名を付与した菌株数は 4592 株であった。このうち 2656 株については、菌株が感染研に送付された。感染研の結果と比較し精度確認が行われた。感染研の結果と一致したものは 89%、1 遺伝座違いが 10%を占めた。

2 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。

3. WGS 解析パイプラインのマニュアル作製および試験的配布

マニュアルでは、コマンド操作の基本、Docker Desktop のインストールを含め、情報解析初心者であってもインストールおよび使用可能なものとした(図 4)。また、Windows、Mac および Linux 別にソフトのインストール方法を記載し、使用する解析端末の OS を問わず利用可能なマニュアルとなった。

2024 年度に行ったアルファ版解析パイプラインは、16 自治体および 4 研究機関に配布した。これらの機関は四国・中国地方を除く全ての地方に位置していた。2025 年 4 月にはベータ版 (v0.9) を GitHub 上に公開し、2026 年 3 月には正式版 (v1.0) を公開した。

4. 講習会およびアンケートの実施

2024 年の講習会は対面および Zoom でのハイブリッド形式で行い、参加者は 14 の機関から、対面で 7 名、オンラインで 20 名以上であった。当日は解析パイプラインのインストール・使用方法に加えて、docker および conda の有償化条件や対処法についての情報提供を行った。また、追加機能の要望などの意見が出さ

れた。

アンケートは 11 機関から回答を得られた(図 5)。用いる情報解析端末については、多くの機関 (n=6) では用いる情報解析端末はプロキシ設定が不要であったが、一部機関で設定が必要であった。そのため、プロキシの設定についても詳細に記述する必要があると考えられた。インストールについては、多くの機関ではマニュアルのみでインストールが可能であった (n=7)。一方、自力でインストールできなかった機関では、docker のインストールが許可されない設定の解析端末を用いていたことが原因であった。同機関では、docker のインストールするために解析端末を購入した企業のサポートを受ける必要があった。また、メモリやプロキシの確認方法について調査する必要があった機関が存在した。プログラムについては、大部分の機関で全てのプログラムが動作しているとの回答が得られた (n=8)。1 機関では checkM が動作しない事例が報告された。調査の結果、Windows 上で docker を動作させる際のメモリの割り当てを設定する必要があることが判明した。機能については、全ての機関から必要な機能が揃っていたとの回答が得られた。マニュアルについては、7 機関から「分かりやすかった」との回答が得られた。追加機能については、BLAST 解析や、菌株のリスト化などのプログラムの要望が得られた。

2025 年の講習会は、対面および Zoom 配信によるハイブリッド形式で実施した。参加登録者は 75 機関から計 136 名であり、当日の Zoom における最大同

時参加者数は100名であった。講習では、プログラムの原理、インストール方法、使用方法および解析結果の解釈について説明した。また、講習会動画は、感染研YouTubeチャンネルにおいて限定公開した。

5. プログラム修正

講習会、アンケートおよび個別の問い合わせから、配布プログラムのバグが何点か同定された。そのため、バグ等の情報をGithub (https://github.com/leechrr/SNPcaster/blob/main/SNPcaster_bugs.md) 上で公開し、利用者との情報共有に努めた。

講習会およびアンケート等で得られた意見をもとに、プログラムの修正を行った。まず、上記のバグについての修正を行った。次に、Docker Desktopの有償化に伴い、無償で使えるソフトウェアとして Rancher Desktop および Docker Engine を標準のソフトウェアとし、マニュアルの改正を行った。また、ソフトウェアのバージョンによって解析結果が異なることから、全てのソフトウェアのバージョンを指定することで、使用端末を問わず同一の結果が得られるようにした。その他に、利用者から要望のあった機能の修正や追加を行った。追加機能として、NCBI AMRFinderPlus を用いた薬剤耐性遺伝子検索機能を導入し、主要病原菌における薬剤耐性遺伝子を自動的かつ網羅的に検索できるようにした。また、SRA Toolkit を利用したプログラムにより、SRA から FASTQ データを自動的にダウンロードできるようにした。

5. 精度評価

シミュレーションデータを用いた精度評価の結果、SNPcaster は高い SNP 検出精度を示した。陽性的中率で評価した場合、いずれのデータセットにおいても、SNPcaster、Snippy および Lyve-SET は高い値を示した (図1)。特に、SNPcaster および Snippy では偽陽性 SNP は認められなかった。一方、Parsnp は他のプログラムと比較して低い値を示した。感度で評価した場合、類似性 99.999% および 99.9% のデータでは、いずれのプログラムも高い値を示した。一方、類似性 97% のデータでは、Snippy および Lyve-SET の値は約 0.75 まで低下し、Parsnp では大きく低下した (図2)。

実データを用いた検証では、SNPcaster は参照配列にかかわらず安定した値を示した。参照配列として解析対象データと同一血清型の株を用いた場合、Snippy 以外のプログラムでは pairwise SNP 数は10か所以内に収まった (図3)。一方、参照配列として遺伝的に遠縁な株を用いた場合、SNPcaster および Parsnp では pairwise SNP 数に大きな変化は認められなかったものの、Lyve-SET では10か所を超える pairwise SNP を有するペアが認められた。Snippy では、平均して500か所以上の pairwise SNP が認められた。

また、クラスタード SNP、リピート領域およびファージ領域に存在する SNP を除外した場合、いずれのプログラムでも類似した結果が得られた。近縁株を参照配列とした場合には、いずれのプログラムにおいても pairwise SNP 数は10

か所以内であった。参照配列が遺伝的に遠縁な株であった場合には、Snippy 以外のプログラムでは pairwise SNP 数は 5 か所以内であった。Snippy では、検出される SNP 数は大幅に減少したものの、最大で 30 か所の pairwise SNP が検出された。

D. 考察

MLVA 法により解析した菌株数は 3 年間で 10000 株を超え、前期と比べて 4 割増であった。解析結果は定期的に厚生労働省 NESFD に MLVA リストとして掲載された。地方衛生研究所から送付された MLVA データは 3 年間で約 4600 株で前期のほぼ倍となった。このうち 6 割の株が後日感染研に送付され、感染研で実施した MLVA データと比較された。結果としては一致もしくは 1 遺伝子座違いのものが 99%を占めた。これらの地衛研については技術的な問題は概ねないと考えられた。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。一方で、マニュアルだけではインストールできなかった事例や、得られた結果の細部が異なっている事例も認められた。いずれの事例についても、原因を特定し、プログラムおよびマニュアルの修正を行うことができた。GitHub 上で一般公開後は、より広範な利用を促進したため、新たなバグ等が確認され、フィードバックをもとに修正を行うことで、よりインストールしやすいプログラムへと改良することができた。講習会には 100 名以上が参加し、本プログラムに対する関心の高さがうかがえた。また、学会や研究会等においても本プログラムの使用例が散見されており、地方衛生研究所における標準的な解析ツールになりつつあると考えられる。シミュレーションデータを用いた検証では、本プログラムが解析対象株の遺伝的多様性にかかわらず、高い精度を示すことが明らかとなった。偽陽性の少なさを示す陽性的中率は、いずれのデータにおいても 100%であった。正確性が求められる行政判断において、SNP 検出精度の高さは重要な要素である。また、感度についても他のプログラムより高い値を示したことから、データの種類にかかわらず、より多くの有用な情報を取得することが示唆された。実データを用いた検証では、参照配列にかかわらず安定し

た結果が得られることが示された。実データにおいて、検出された SNP が偽陽性であると断定することは困難である。しかし、参照配列が遺伝的に遠い場合に、近い場合よりも多数の SNP が認められることは、偽陽性 SNP の存在を示唆している。本研究では、Snippy および Lyve-SET において、遺伝的に遠い株を参照配列とした際に、多数の偽陽性と考えられる SNP が出現した。過去の研究においても、これらのプログラムでは参照配列が遺伝的に遠い場合に偽陽性 SNP が検出されることが示されており、本研究の結果と一致している。偽陽性 SNP の主な原因は、ファージ領域などの組換え領域であると考えられる。本研究においても、組換えの結果と考えられるクラスタード SNP、ならびに組換えが起こりやすいリピート領域およびファージ領域に存在する SNP を除外した結果、検出される SNP 数が減少し、プログラム間の差異も小さくなった。組換え領域の存在は、リードマッピングの際にソフトクリップと呼ばれる不完全なアラインメントを生じさせ、その結果、偽陽性 SNP が検出されることが考えられている。SNPcaster において SNP 検出を担っている BactSNP は、アラインメントおよびリードマッピングの両方に基づく変異検出を用いることで、ソフトクリップ等に起因する偽陽性 SNP 検出を大幅に低減していると考えられる。

E. 結論

MLVA の導入によって、迅速な菌株情報の集約が実現可能となってきた。食中毒事例等への対応には、MLVA を実施する

地衛研での精度維持向上、並びに地衛研—感染研—厚労省—自治体間の連携が不可欠である。

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。プログラムの一般公開を行ったことで地方衛生研究所等の公的な機関において利用が広がっている。精度についても高精度であることが確認された。今後、地方衛生研究所等において SNP 解析の標準的なツールとなることが期待される。

F. 健康危険情報

なし

G. 研究発表

1) 誌上発表

1. 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
2. 泉谷秀昌、李謙一、伊豫田淳、明田幸宏：2023 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 45 巻、80-82、2024 年 5 月
3. 泉谷秀昌、李謙一、伊豫田淳、明田幸宏：2024 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 46 巻、103-104、2025 年 5 月
4. 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.

5. Matsumoto Y., Lee K., Akasaka R., Honjo H., Koizumi M., Sato T., Kubomura A., Ishijima N., Akeda Y., Ohnishi M., and Iyoda S. Increased resistance against tellurite is conferred by a mutation in the promoter region of uncommon tellurite resistance gene *tehB* in the *ter*-negative Shiga toxin-producing *Escherichia coli* O157:H7. 2024. Appl Environ Microbiol 90:e0228323.

6. Kubomura A, Lee K, Iyoda S, Akeda Y. 2025. Complete genome sequence of three atypical diarrheagenic *Escherichia coli* O166:H15 strains from foodborne outbreaks. Microbiol Resour Announc 14:e0074625.

7. Ooka T, Arai S, Lee K, Gotoh Y, Kubomura A, Imuta N, Hara-Kudo Y, Iyoda S, Hayashi T, Nishi J. 2025. Prevalence, sequence diversity, and amplification of an IS-associated enterotoxin gene, *astA*, in *Escherichia coli*. Front Microbiol 16:1635769.

8. Imai Y, Okuno M, Hoshiko Y, Kaneko H, Sato T, Noguchi A, Lee K, Yamamoto T, Nakayama-Imaohji H, Kuwahara T, Iyoda S, Ogura Y. 2025. Genomics of Shiga toxin-producing *Escherichia coli* from asymptomatic carriers in Japan highlights risk-adaptive control. Commun Biol 9.

9. Morita M, Lee K, Sugiyama A, Kojima N, Kawai Y, Izumiya H, Akeda Y. 2026. A case of laboratory-acquired *Salmonella* Typhi infection due to phage typing in Japan: whole-genome sequencing confirms the source of infection. Access Microbiol 8.

2) 学会発表

1. 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。

2. 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。

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（2023 年）。第 45 回日本食品微生物学会学術総会、2024 年 9 月、青森県青森市。

4. 泉谷秀昌：細菌性食中毒の概要、発生動向、集団事例、広域事例等について。第 64 回東北ブロック 食品衛生・環境衛生監視員研修会、2024 年 9 月、岩手県盛岡市。

5. 泉谷秀昌：MLVA をはじめとした分子疫学解析と食中毒。令和 6 年度岐阜県食品衛生監視員等研修会、2024 年 11 月、岐阜県岐阜市。

6. 泉谷秀昌：外部精度管理事業（EHEC）。令和 5 年度希少感染症診断技術研修会、2024 年 12 月オンライン。

7. 泉谷秀昌：EHEC O157、O26、O111 の MLVA。令和 6 年度腸管出血の MLVA 研修会、2025 年 1 月、東京都

8. 泉谷秀昌：3 類感染症。令和 7 年度希少感染症・重点感染症診断技術研修会、2026 年 2 月オンライン。
9. 泉谷秀昌：EHEC O157、O26、O111 の MLVA。令和 7 年度腸管出血の MLVA 研修会、2026 年 2 月、東京都
10. 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.
11. 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、2024.
12. 李 謙一, 伊豫田 淳, 泉谷秀昌, 小泉充正, 仙田 隆一, 菅原 庸, 菅井基行, 明田幸宏. 輸入馬刺しによる食中毒事例に関連した腸管出血性大腸菌 O26:H11 の全ゲノム配列解析. 第 26 回腸管出血性大腸菌感染症研究会. 茨城, 2024.
13. 鈴木麻友, 瀬戸順次, 的場洋平, 池田辰也, 水田克巳, 大貫典子, 李 謙一, 泉谷秀昌. 2023 年の馬刺しによる腸管出血性大腸菌食中毒の原因追究. 令和 6 年度東北地区獣医師大会・令和 6 年度獣医学術東北地区学会. 2024.
14. 李 謙一. 腸管出血性大腸菌における全ゲノム配列解析のサーベイランス等への応用. 第 167 回日本獣医学会学術集会. 北海道, 2024.
15. 李 謙一, 伊豫田 淳, 泉谷秀昌, 小泉充正, 仙田隆一, 明田幸宏. 輸入馬刺しによる食中毒事例株を含む腸管出血性大腸菌 O26:H11 流行株の全ゲノム配列解析. 第 167 回日本獣医学会学術集会. 北海道, 2024.
16. 鈴木麻友, 瀬戸順次, 的場洋平, 池田辰也, 水田克巳, 大貫典子, 李 謙一, 泉谷秀昌. 2023 年の馬刺しによる腸管出血性大腸菌食中毒の原因追究. 獣医学術学会年次大会. 2024.
17. 李 謙一, 伊豫田 淳, 泉谷秀昌, 明田幸宏. 腸管出血性大腸菌の 大規模集団事例解析にもとづく SNP 解析条件の最適化. 第 99 回日本細菌学会総会. 広島, 2026.
18. 李 謙一, 森田昌知, 伊豫田 淳, 泉谷秀昌, 中村佳司, 山岸敏明, 明田幸宏. SNPcaster: 細菌ゲノム解析パイプライン. 第 46 回日本食品微生物学会学術総会. 川崎、神奈川, 2025.
19. 李 謙一, 森田昌知, 伊豫田 淳, 泉谷秀昌, 中村佳司, 山岸敏明, 明田幸宏. SNPcaster: 細菌ゲノムアセンブリ・系統解析パイプライン. 第 168 回日本獣医学会学術集会. 宮崎, 2025.
20. 原田哲也, 若林友騎, 李 謙一, 河合高生, 伊豫田 淳. 腸管出血性大腸菌 O157 clade 分類のためのリアルタイム PCR 法の確立. 第 27 回腸管出血性大腸菌感染症研究会. 2025.
21. 小高優人, 上村直輝, 岩瀬忠行, 李 謙一, 石嶋 希, 中根大介, 菅井基行, 菅原庸, 明田幸宏, 伊豫田 淳, EHEC Working Group. 腸管出血性大腸菌における多細胞鎖形成機構の分子遺伝学的解析. 第 27 回腸管出血性大腸菌感染症研究会. 2025.

22. 李 謙一, 森田昌知, 伊豫田 淳, 泉谷
秀昌, 中村佳司, 山岸敏明, 久保田寛顕,
木全恵子, 若林友騎, 野本竜平, 柿田徹也,
and 明田幸宏. 細菌ゲノム解析パイプ
ライン(SNPcaster)の開発. 第 27 回腸管出血
性大腸菌感染症研究会. 2025.

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 (UID5162)	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 (UID5162)	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 (UID5162)	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 (UID5162)	99.93	0.51	PASS	pass	pass	pass	pass	pass

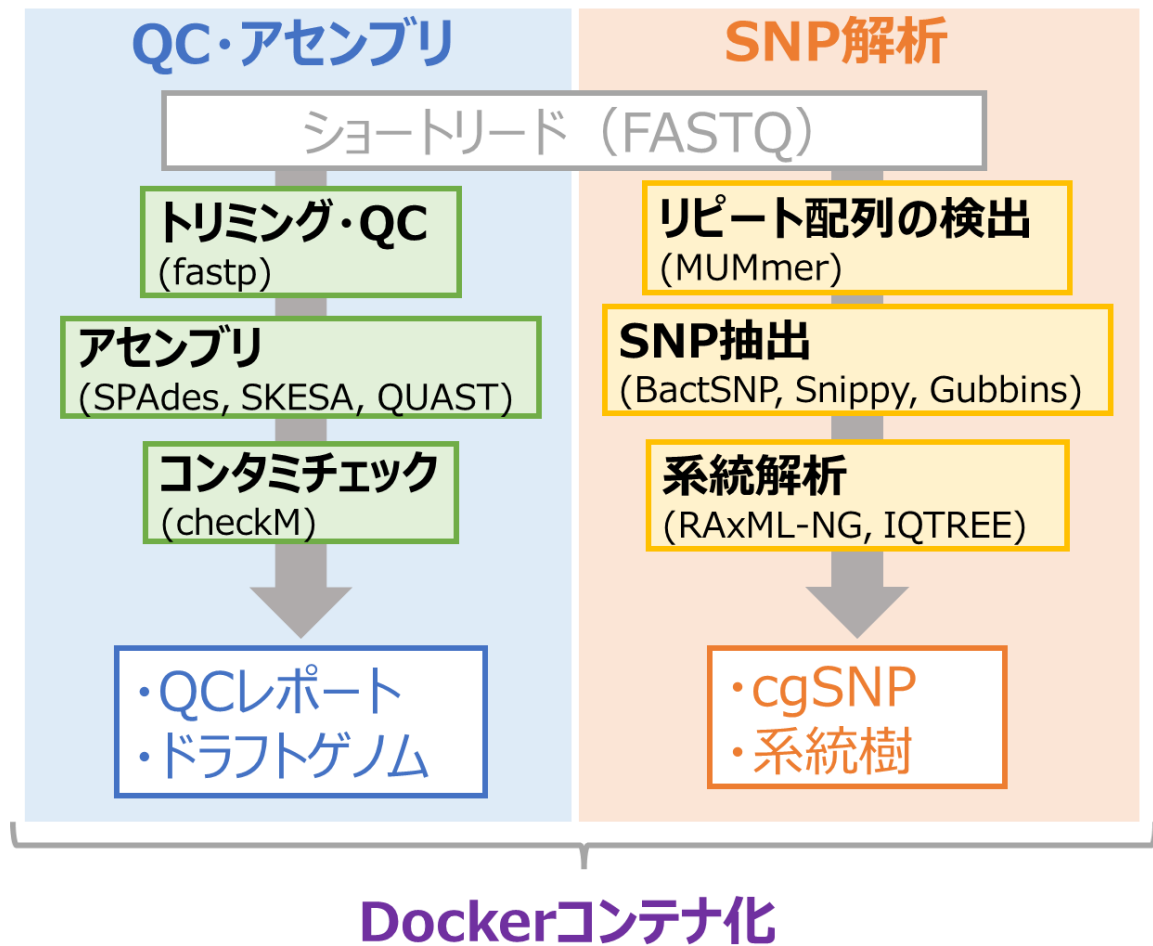


図 1. 解析パイプラインの概要

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 recombino-genic 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

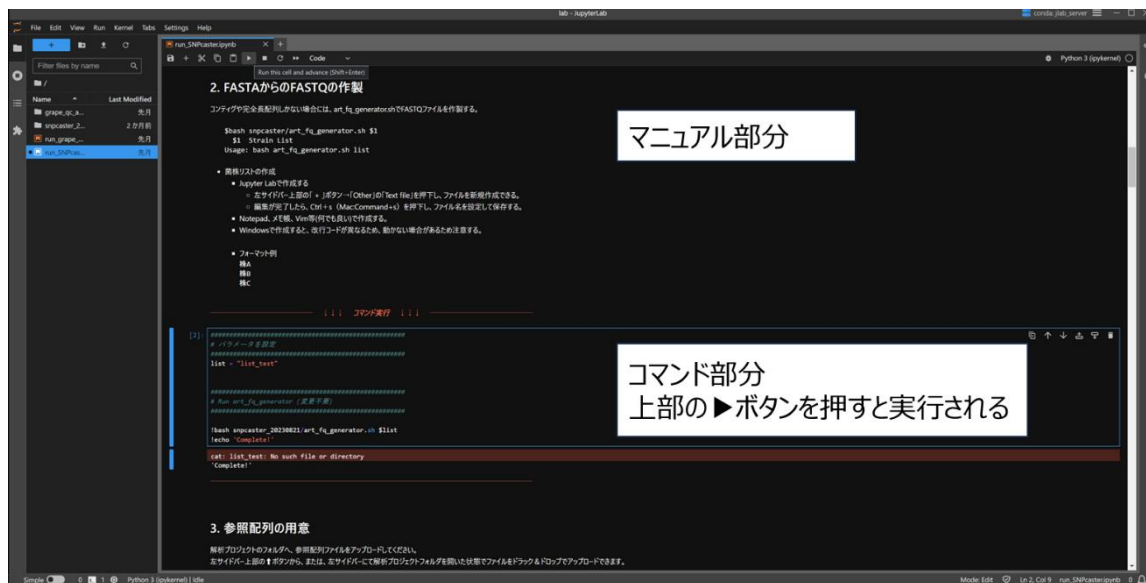


図 3. JupyterLab ノートブック例

A

目次

1. 基本的な使い方	5
1.1. コマンド入力アプリ	6
1.2. コマンドライン操作	9
1.3. ディレクトリ説明、移動	9
2. インストール	15
2.1. Dockerのインストール	15
2.2. プロキシサーバの設定	15
2.3. SNPCaster・grape_qc_assembly Notebook のインストール	32
2.4. 解析端末への SNPCaster インストール	35
3. 起動・停止方法	38
4. 解析実行手順	40
4.1. フォルダ構成	40
4.2. Jupyter Lab の使い方を学ぶ	41
4.3. プロジェクト作成	44
4.4. 解析の実行	47
5. アップデート(非推奨)	51
5.1. Jupyter Notebook の修正	51
5.2. 解析スクリプトの修正	52
5.3. Dockefile の修正	53
6. コマンドライン操作	54
7. トラブルシューティング、Q&A	56
今後の検討課題	58

B

1.1.1.3. Linux

上述の Windows と同様に、フォルダ上のコンテキストメニューから「Open in Terminal」で、表示されているフォルダをカレントディレクトリとしたターミナルが開きます。

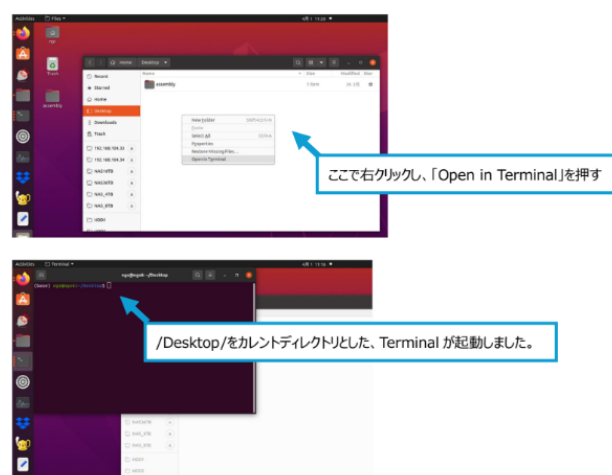


図 4. 解析パイプラインマニュアルの目次（A）および内容の一部（B）

1. 選択形式の設問

設問	回答数				
docker pullによるsnpcaster イメージのダウンロードをどこで行 いましたか?	職場の解析端末 (プロキシ設定必要)	職場の解析端末 (プロキシ設定必要なし)	職場の別の端末 (プロキシ設定必要)	職場の別の端末 (プロキシ設定必要なし)	自宅のパソコンで
	2	6	0	1	2
インストール	自力ではインストールでき なかった	マニュアルに書いていないことを調 べる必要があった	マニュアルだけを読んでイン ストールできた		
	1	3	7		
プログラムが動いたか	全てのプログラムが動作し なかった	一部のプログラムが動作せず、解 決しなかった	一部のプログラムが動作せ ず、調べる必要があった	全てのプログラムが問題なく 動いた	
	0	1	2	8	
機能	業務で必要だが、含まれ ていない機能がかった	必要な機能は揃っていた			
	0	11			
マニュアル	自力ではインストールでき なかった	マニュアルに書いていないことを調 べる必要があった	分かりやすかった		
	0	4	7		

2. 自由記載の設問

設問	回答
追加して欲しい機能	・菌株リストの作製 ・アノテーション ・アセンブリ結果を1つのフォルダにまとめて欲しい ・MLST ・血清型別 ・病原因子の検索 ・トリミングファイルを残して欲しい
改善して欲しい部分	・系統解析結果をまとめて欲しい ・実行中ログに関する注意事項の追加 ・repeatsファイルが移動してしまう ・データ解釈について学べる機会が必要

図 5. 解析パイプラインアンケート結果

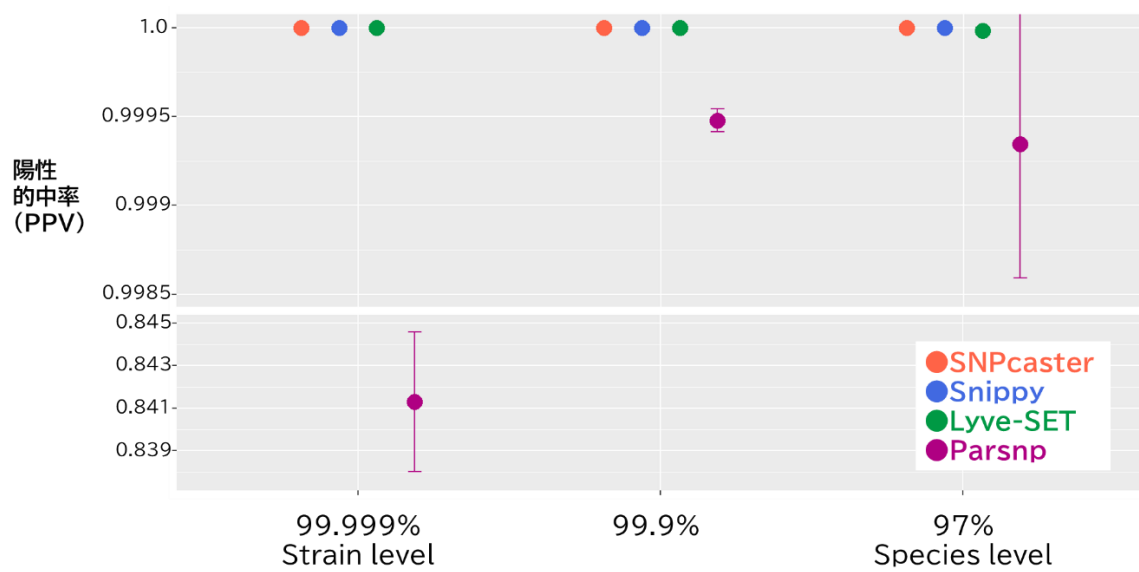


図 6. SNP 検出プログラム別にみたシミュレーションデータにおける陽性的中率の比較

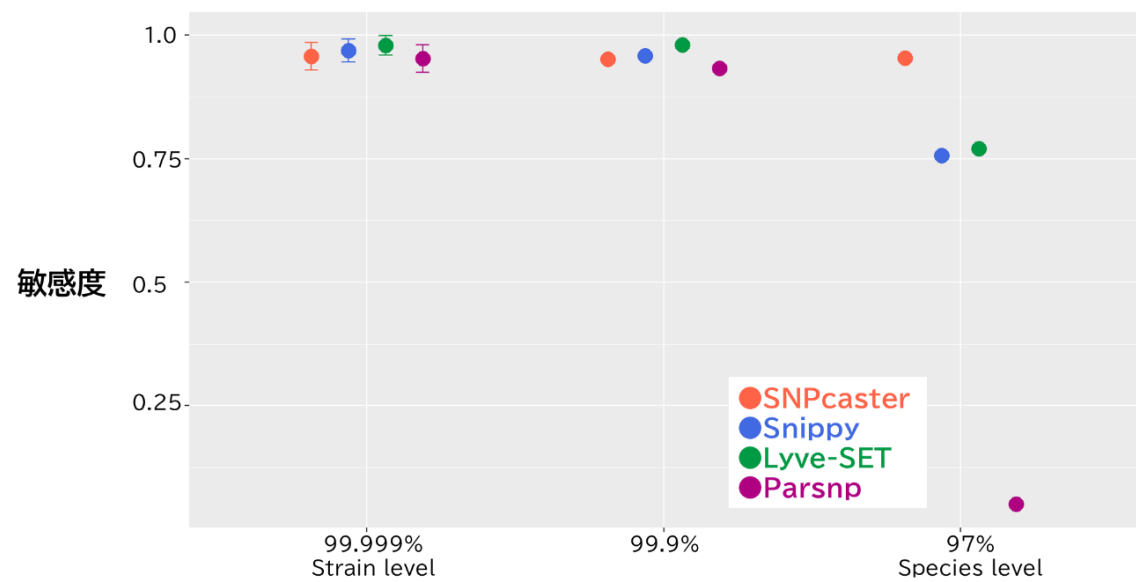


図 7. SNP 検出プログラム別にみたシミュレーションデータにおける敏感度の比較

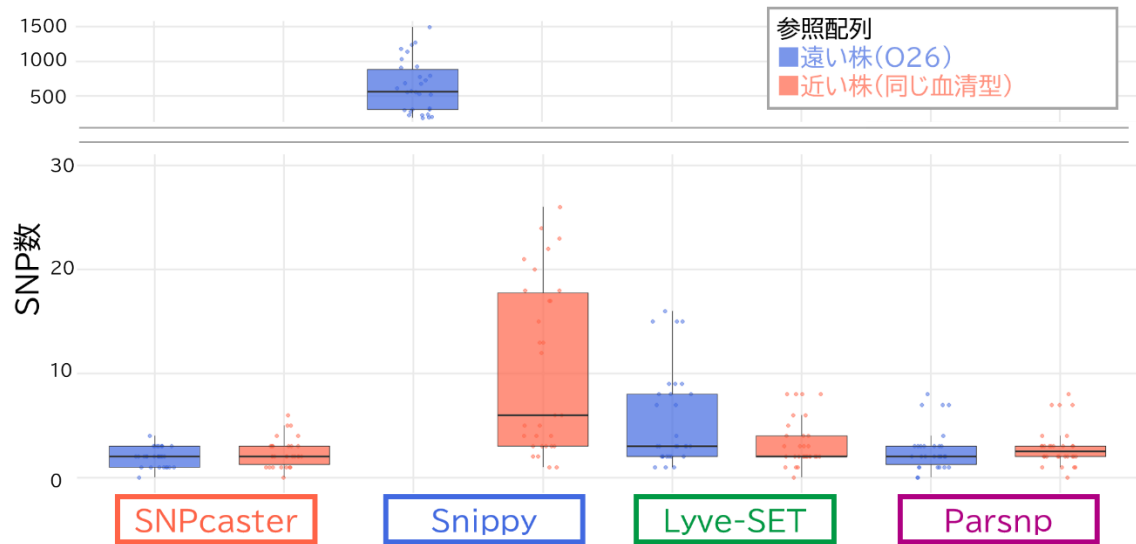


図 8. SNP 検出プログラム別にみた実データにおける

pairwise SNP 数の比較(クラスタード SNP、リピートおよびファージ領域の除去なし)

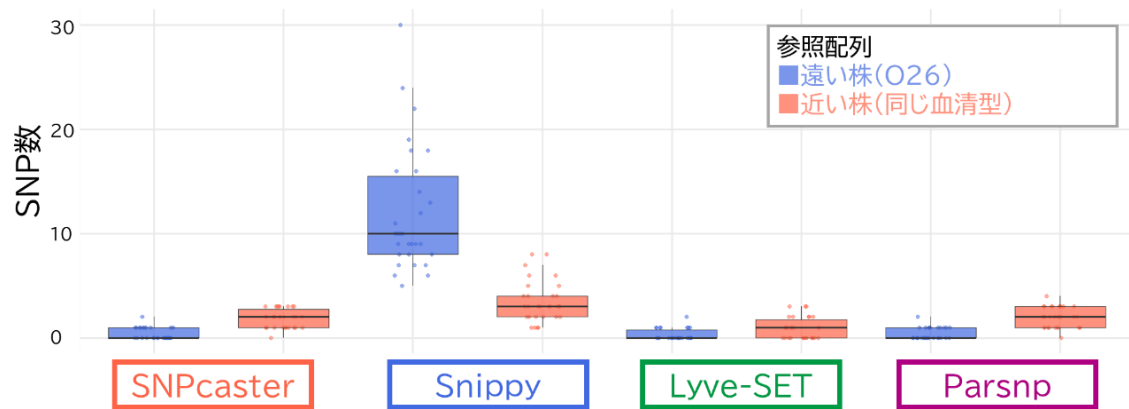


図 9. SNP 検出プログラム別にみた実データにおける

pairwise SNP 数の比較(クラスタード SNP、リピートおよびファージ領域の除去あり)