

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

「2024 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 2025 年分離株は 3,168 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 1700 であった。

また全ゲノム配列解析を用いて広域食中毒の調査に活用可能な解析パイプライン（SNPcaster）を構築した。前年度までに開発した試験版を改良し、ベータ版として GitHub 上で公開するとともに、試験運用を開始した。その後、得られたフィードバックをもとに改良を重ね、正式版（v1.0）をリリースした。追加機能として、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。

また近年、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。実際の解析では、NGS により WGS 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、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解析パイプラインの確立および共有化を目的とした。本年度は、前年度までに作製した解析パイプラインのアップデートおよび精度検証を行った。

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 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. プログラムアップデートおよび機能追加

前年度までに得られたフィードバックをもとに、各種バグ修正等を行い、解析パイプラインをベータ版（v0.9）として GitHub 上に公開した。主に地方衛生研究所職員を対象に周知を行い、試験的な運用を開始した。

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

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。

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。

3. 講習会の実施

配布した解析パイプラインのインストール方法および操作方法に関する講習会を、2025 年 12 月 24 日に国立感染症研究所戸山庁舎にて開催した (対面および Zoom 配信)。また、後日、説明動画を YouTube の限定公開機能を利用して配信した。

4. 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、3,793 株について分子型別解析を実施した。このうち 3,168 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 861、O26 が 164、O111 が 69、O103 が 97、O121 が 21、O145 が 21、O165 が 3、O91 が 40 であった。得られたデータは 2026 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、延べ 46 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 1700 株であった。

2025 年は 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 99%であった。株毎の正解率（平均値）は約 88%であった。

2. プログラムアップデートおよび機能追加

本年度は、GitHub 上で v0.9 を公開した後、地方衛生研究所向けに積極的な周知を行い、得られたフィードバックをもとに修正および機能追加を実施した。

追加機能として、NCBIAMRFinderPlus 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。

3. 講習会の実施

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。

4. 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昨年より約 25%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。この中には食中毒事例関連株、有症苦情関連株も含まれ、厚労省及び各自治体における対応に活用された。

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 1700 株に上った。感染研で行った精度確認では菌株ベースで 88%、遺伝子座ベースで 99% 一致しており、昨年度とほぼ同レベルで推移した。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。

ゲノム解析パイプライン構築について、本年度は、SNPcaster を GitHub 上で一般公開し、より広範な利用を促進した。その結果、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でき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、迅速な菌株情報集約の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。

また SNPcaster 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：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
2. 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.

3. 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.

4. 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.

5. 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：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。
2. 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
3. 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。
4. 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.
5. 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.
 6. 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.
 7. 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.
 8. 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.
 9. 李 謙一. ゲノム解析が切り拓く腸管出血性大腸菌の新たなサーベイランスと高病原性株の特定. 微生物ウィーク2025. 東京, 2025.
 10. 岡田由美子, 西田智子, 山本詩織, 李 謙一, 泉谷秀昌, 西野由香里, 井田美樹, 下島優香子, and 都丸亜希子. 国内で分離された患者及び食品由来リステリア・モノサイトゲネスの分子疫学的解析. 第98回日本細菌学会総会. 石川, 2025.
 11. 石嶋 希, 岩瀬忠行, 児玉年央, 李 謙一, 大岡唯祐, 後藤義幸, 八尋錦之助, 明田幸宏, and 伊豫田 淳. HUS患者から分離されたstx2f遺伝子保有E. albertiiの病原性解析. 第98回日本細菌学会総会. 石川, 2025.
 12. 李 謙一, 森田昌知, 伊豫田 淳, 泉谷秀昌, 中村佳司, 山岸敏明, and 明田幸宏. 細菌ゲノム解析パイプライン(SNPcaster)の開発および共有化方法の構築. 第98回日本細菌学会総会. 石川, 2025.
 13. 大岡唯祐, 玉村雪乃, 桃木杏奈, 李 謙一, 平井晋一郎, 伊豫田 淳, 楠本正博, and E. albertii Working Group. 国内で分離されたEscherichia albertiiの進化系統およびゲノム特性. 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

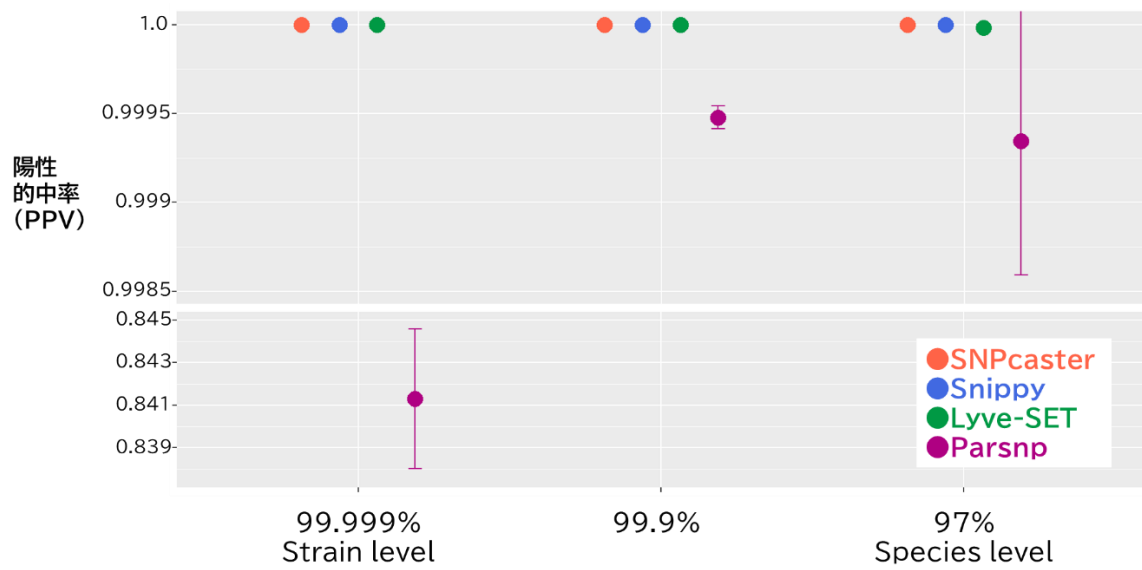


図 1. SNP 検出プログラム別にみたシミュレーションデータにおける陽性的中率の比較

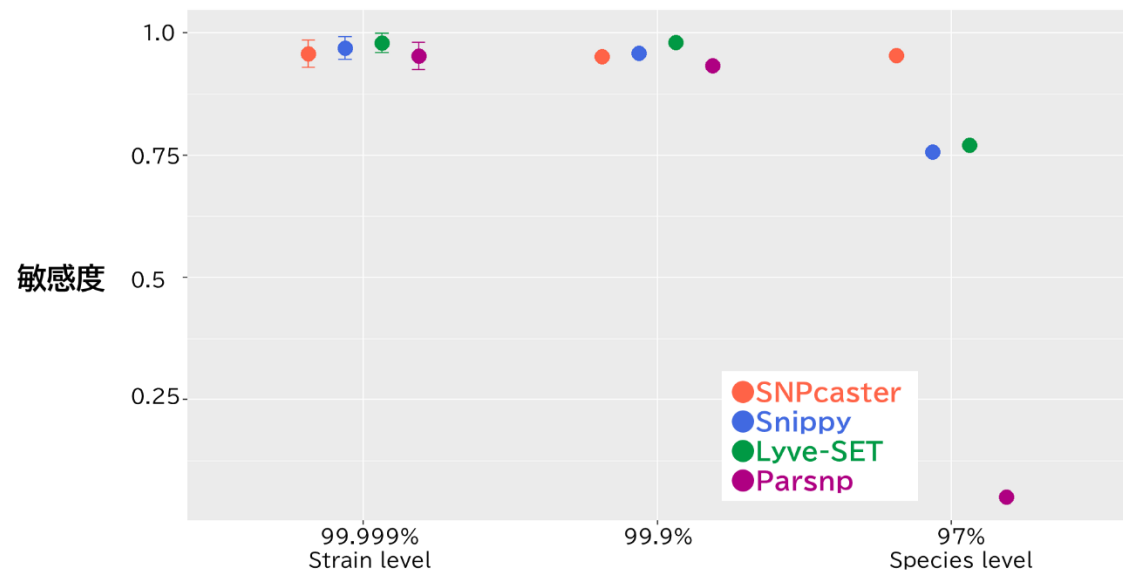


図 2. SNP 検出プログラム別にみたシミュレーションデータにおける
敏感度の比較

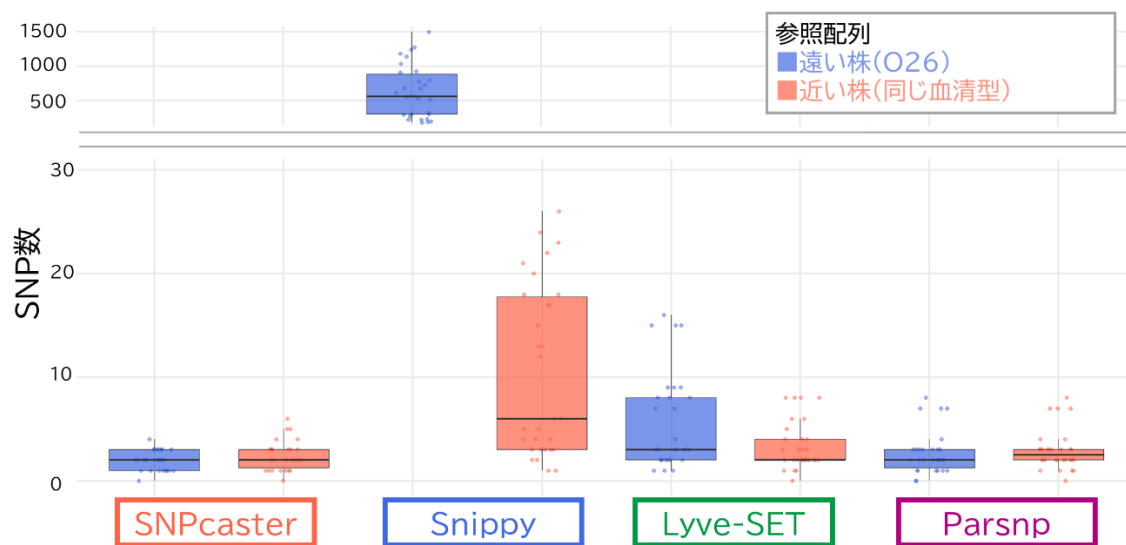


図 3. SNP 検出プログラム別にみた実データにおける pairwise SNP 数の比較(クラスタード SNP、リピートおよびファージ領域の除去なし)

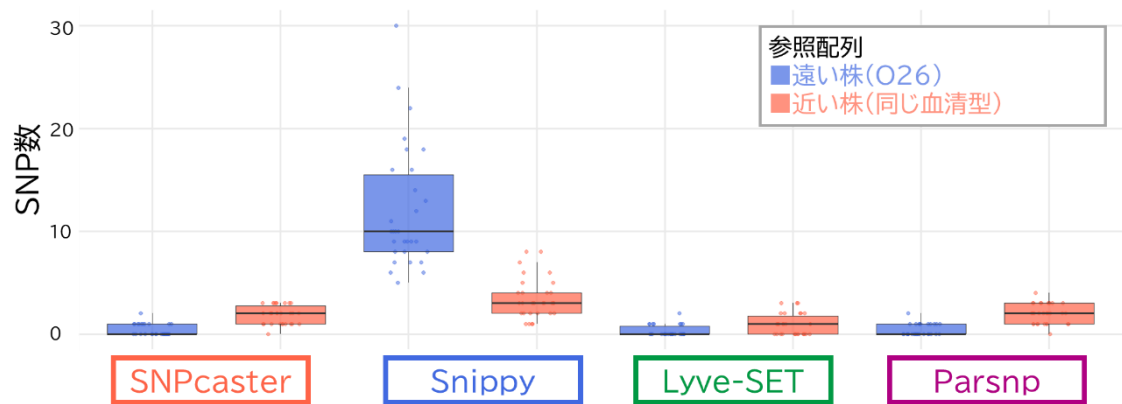


図 4. SNP 検出プログラム別にみた実データにおける pairwise SNP 数の比較(クラスタード SNP、リピートおよびファージ領域の除去あり)