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
「ロングリードシークエンサーによる集団感染調査手法の確立」
（23KA3005）
研究分担報告書

分担研究課題「全ゲノム配列解読およびプログラム構築」
研究代表者　李　謙一（国立感染症研究所　細菌第一部）、

研究要旨

Oxford Nanopore 社のロングリードシークエンサーを用いた集団感染調査手法の確立のために、腸管出血性大腸菌（enterohemorrhagic *Escherichia coli* : EHEC）を材料とし、全ゲノム配列（whole-genome sequence, WGS）の解読を行った。既存の解析パイプラインを使用したところ、多数の偽陽性 SNP が認められた。そこで、新たな解析プログラムを構築したところ、高い精度で SNP が抽出可能であり、ショートリードデータと比較可能であることが明らかとなった。

A. 研究目的

病原体のサーベイランスにおける分子型別において、近年全ゲノム配列（whole-genome sequence：WGS）解析法が主流になりつつある。アメリカ合衆国 CDC においても腸管出血性大腸菌（enterohemorrhagic *Escherichia coli*: EHEC）等の食中毒菌のサーベイランスに用いる分子型別手法は、2019 年からパルスフィールドゲル電気泳動法から WGS 解析に移行している。日本においては、COVID-19 によるパンデミックの影響で全国の地方自治体に次世代シークエンサー（next generation sequencer: NGS）が配備されたものの、病原細菌における活用例は未だ少ない。その理由としては、WGS 解読後の情報解析が複雑で、難易度が高いことが挙げられる。

WGS の情報解析については、イルミナをはじめとするショートリード型 NGS に対応したプログラムは多数利用可能である。しかし、Oxford Nanopore 社のロングリード型 NGS から得られるデータを用いたプログラムは少ない。また、集団感染調査の際には、供試菌株が同一株であるかの判断を、コアゲノム単一塩基多型（core genome single nucleotide polymorphism, cgSNP）の数をもとに行うのが一般的である。ショートリード型データを用いた際の、同一由来と判断するための基準は多数の菌種で報告されているが、ロングリード型データを用いても同様の判断ができるか、体系的に検討した報告は少ない。

そこで本研究では、EHEC を材料に、ロングリードのみで SNP 解析を行うことができるプログラム構築を行う。本年度は、前年度までに開発したプログラムの改善および評価を行った。

B. 研究方法

1. 全ゲノム配列データの解読

前年度までに解析した株のうち、データが不足した株について、次の方法でデータを取得した MagMAX DNA Multi-Sample Ultra 2.0 Kit および KingFisher Duo Prime (ThermoFisher) を用いて DNA 抽出を行い、ショートリード型およびロングリード型 NGS を用いて WGS の解読を行った。

ショートリードについては、QIAseq FX DNA Library Kit（QIAGEN）を用いてライブラリー調製を行い、NovaSeq_X_Plus（illumina）によってペアエンドシークエンシング（150-mer×2）を行った。ロングリードについては、Rapid Barcoding Kit 24 V14、R10.4.1 フローセルおよび Mk1B を用いて 72 時間のシークエンス解析を行った。得られた配列データについて、porechop によるトリミング後、リード長や合計リード長等の統計量を、seqkit によって算出した。

2. 解析プログラムの構築

全ゲノム配列からの変異検出手法は、大きく分けて以下の 3 通りに分類される：ドラフトゲノムのアラインメント法、リードマッピング法、および k-mer 法（図 1）。本研究では、各手法の代表的なプログラム（アラインメント法：MUMmer、マッピング法：Clair3、k-mer 法：SKA2 および MINtyper）を用いて SNP 検出を行った。得られた結果を、ショートリードを用いて SNPcaster で抽出した SNP と比較し、各プログラムの精度を評価した。

上記いずれの方法でも良好な成績が得られなかったため、SNP 抽出プログラムを新規に開発した。基本アルゴリズムにはアラインメント法を採用し、BactSNP で用いられている MUMmer をもとに SNP 検出プログラムを構築した。この過程で、一部のメチル化モチーフにおいて偽陽性 SNP が認められたことから、dorado polish によるエラー修正を組み込んだ。得られた SNP 情報は、SNPcaster と同様に pseudogenome および vcf ファイルとして株別のフォルダに出力する仕様とし、これによりショートリードデータと併せて SNPcaster で解析することが可能となった（図 2）。

C. 研究結果

既存のプログラムを用いて SNP 検出を行った場合には、いずれのプログラムにおいても、ショートリードおよびロングリードデータ間で差異が認められた。そこで、新たなプログラムの構築を行うこととした。

新たなプログラムは、ロングリードデータを入力とし、自作シェルスクリプト(snpcaster-lp.sh)により以下のパイプラインで変異検出を行った。

まず、Flye(--nano-hq モード)で de novo アセンブリを行い、ドラフトゲノム配列を取得した。次に、Dorado aligner および Dorado polish を実行し、塩基精度を向上させた consensus 配列を得た。このエラー修正によって、メチル化が原因となる偽陽性 SNP を減らすことが可能であった。

得られた polish 済みアセンブリと参照ゲノムとの間で、BactSNP と同様の whole-genome alignment ベースの変異検出を実施した。具体的には、MUMmer の nucmer で両ゲノムをアラインメントし、delta-filter のマッピング(-r -q、identity ≥ 90 %、length ≥ 100 bp)で品質の低い領域を除外した。フィルタ後の delta ファイルから SNP・indel を抽出するとともに、indel 周辺 5 bp 以内に位置する SNP(偽陽性となりやすい)をマスクして除外した。最終出力として、(i) リファレンス座標に整列した pseudogenome 配列(snps.aligned.fa)、(ii) 変異リスト(snps.vcf)、(iii) アラインメント統計(alignment_stats.txt)を得た。

本パイプラインは、ショートリード由来アセンブリを対象として開発された BactSNP の手法をロングリードアセンブリに適用したものであり、参照ゲノムにマッピングできない領域は pseudogenome 上で N(欠損)として扱われるため、core 領域に基づく株間比較が可能である。

本パイプラインでショートリードおよびロングリードによる SNP 解析結果を比較したところ、93 株中 82 株(88.2%)で両者の差異がゼロとなった（図 3)。残りの 9 株についても、ごく少数(1-3 か所)の SNP のみが認められた。

D. 考察

本研究では、前年度までに得られた知見をもとにプログラムの改良を行い、実用レベルの解析プログラムを構築した。

既存のプログラムには、いずれも偽陽性 SNP が存在していた。その原因として、マッピングおよび k-mer 法によるものでは、クオリティの低いマッピング（ソフトクリップ）が要因となっていることが明らかとなった。ソフトクリップとは、シークエンスリードを参照配列にマッピングする際に、リードの末端部分が参照配列と一致しない場合に、当該領域をアラインメントから除外しつつリード自体は保持する処理である。ソフトクリップは、組換え領域や菌株間で**差異**の大きい領域に高頻度で発生する。そのため、参照配列が遠縁の株である場合には、このような低品質マッピングに起因する偽陽性 SNP が頻発することとなる。

本プログラムでは、このソフトクリップを抑制する目的でアラインメント法を採用した。しかし、アラインメント**法**でもアラインメントの端で偽陽性 SNP が発生することが明らかとなった。そのため、BactSNP と同様に、アラインメントの端から一定領域（デフォルトで 5 bp、変更可能）の SNP を除外するアルゴリズムを採用した。

また、一部の株ではメチル化部位における偽陽性 SNP が認められた。これはナノポアシークエンサーに見られるベースコールのミスに起因するものであり、polishing ツールである dorado polish を適用することで修正可能であった。

このようにして開発したプログラムは、従来の SNPcaster と同様の形式のファイルを出力する。これにより、ショートリードとロングリードの両方を一つのプログラムで解析することが可能となった。今後は、ショートリードおよびロングリードで認められたごく少数の SNP の検

証を行うとともに、プログラムのさらなるアップデートを経て SNPcaster プログラムへの統合を進めることで、ロングリード解析の社会実装につながると考えられる。

E. 結論

国内で分離された EHEC 集団感染事例を網羅する菌株セット計 102 株菌株のショートリードおよびロングリード解読を行い、ロングリードを用いてショートリードと同精度の SNP 解析を行えるパイプラインを構築した。

F. 健康危険情報

なし

G. 研究発表

1) 誌上発表

1. 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.
2. 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.
3. 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.
4. 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. **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.
2. **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.
3. **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.
4. 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.
5. 小髙優人, 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 W

orking 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.

6. 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.

7. 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.

8. 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.

9. 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.

10. 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(SNP caster)の開発およびと共有化方法の構築. 第98回日本細菌学会総会. 石川, 2025.

11. 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

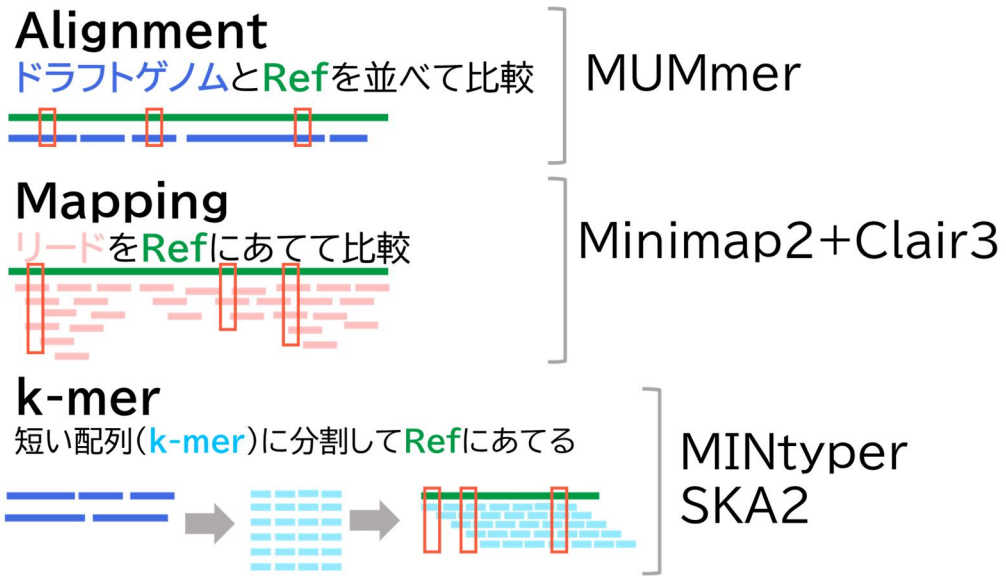


図 1. 変異検出アルゴリズムおよび使用したプログラム

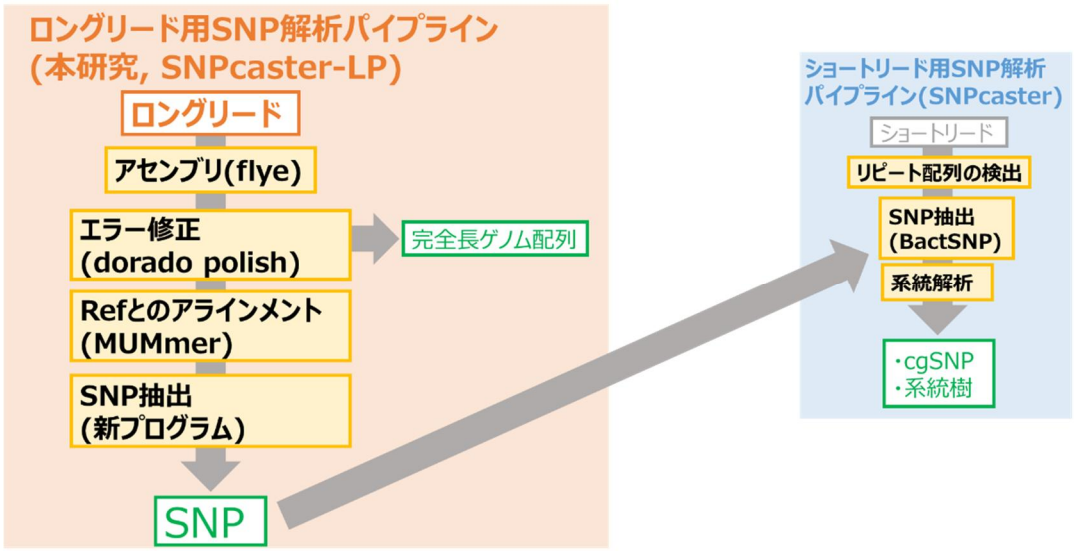


図 2. 解析パイプライン概要

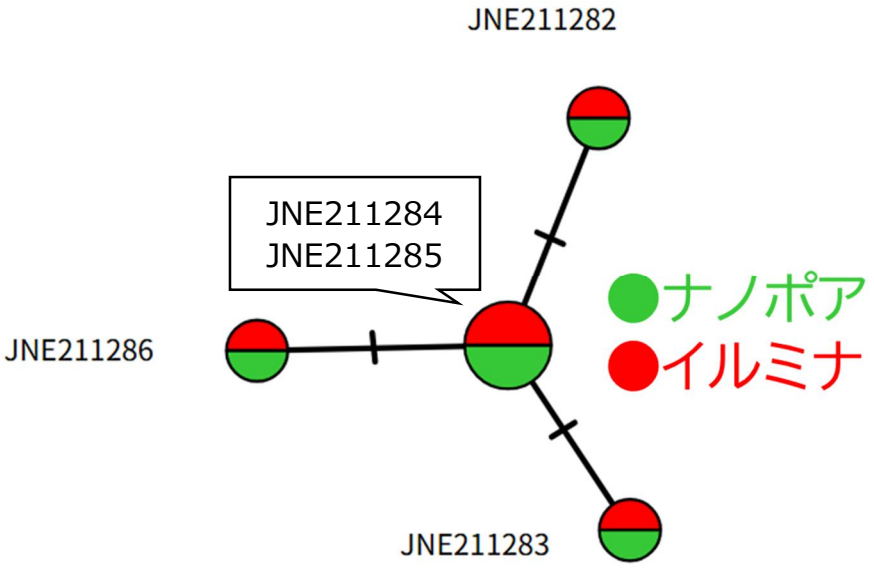


図 3. ナノポアおよびイルミナシークエンサーを用いて抽出した SNP によるハプロタイプネットワーク図例
丸印は各菌株を示す。緑はナノポアシークエンサーによるデータ、赤はイルミナシークエンサーによるデータを示す。