

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
「ロングリードシーケ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ケ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。2021 年および 2022 年に分離された EHEC のうち、5 株以上が分離された集団感染事例から 20 件、計 102 株の EHEC 解析株を決定した。本年度はそのうち約半数の 48 株について、ロングリードおよびショートリードによる WGS 解読を行った。DNA 抽出法およびライブラリー調整キットの条件検討を行った結果、地方自治体でも広く使われているカラム抽出キットおよび Rapid Barcoding Kit によって、十分なデータ量が取得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による影響で全国の地方自治体に次世代シーケンサー (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、疫学関連が既知の株について体系的な WGS 解読後、cgSNP 解析を行うことで、ロングリード型 NGS を用いた集団感染調査の判断基準の検討を行う。また、解析プログラムを構築し、地方自治体等への配布を目指す。今年度は、解析株の決定、約半数の WGS 解読、およびロングリード解析のための各種条件検討を行った。

B. 研究方法

2021 年および 2022 年に分離された EHEC について、疫学関連が既知の事例（以下、集団感染事例）のうち、5 株以上が分離された事例について、O 血清群ごとに株数を集計した。集計結果をもとに、計 102 株の WGS 解読株を決定した。決定した株について、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ケンシング (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 時間のシーケンス解析を行った。ロングリード解析では、上記の条件に加えて、一部の菌株で次の 2 条件での解析を行った：1) DNA 抽出を DNeasy Blood & Tissue Kit

(QIAGEN) に変更、2) ライブラリー調整を Native Barcoding Kit 24 V14 に変更。

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。

C. 研究結果

2021 年および 2022 年に分離された EHEC のうち、5 株以上が分離された集団感染事例は 39 件であり、O 血清群は 9 種に分かれた（表 1）。国内での EHEC 報告数の 95%以上は主要 8 血清群 (O157, O26, O111, O103, O121, O145, O91, O165) が占めることから、本研究では O157, O26, O111, O103 および O121 の計 20 事例、102 株を対象に解析を進めることにした。

本年度はこのうち、約半数の株 (48 株) についてショートリードおよびロングリードによる WGS 解読を行った。ロングリード解析では、基本条件 (MagMAX で DNA 抽出し、Rapid Barcoding Kit でライブラリー調整) では 1 枚のフローセルで 12 株を解析した場合、平均カバレッジは 179 となり、十分なデータ量が得られた（表 2）。ライブラリー調整キットに Native Barcoding Kit を用いた場合には、1 枚のフローセルで 15 株の解析を行った。その結果、データ量は基本条件よりも多くなった一方で、平均リード長はより短くなっていた。DNA 抽出に DNeasy キットを用いた場合にも、Native Barcoding Kit を用いた場合と同様に、基本条件と比べてより多くのデータ量、より短いリード長が認められた。

D. 考察

本研究で解析対象としたのは、2 年間に
おける国内で報告された主な EHEC 集団
感染事例の約半数であり、国内 EHEC を
網羅するものであると考えられる。これ
までの研究によって、系統や事例によっ
て認められる SNP の数が異なることが示
されており、多様な EHEC を解析するこ
とによって、より普遍的な知見が得られ
ると考えられる。

解析条件については、基本条件による
解析が最もリード長が長く、Native
Barcoding Kit によるライブラリー調整の
条件が最もデータ量が多かった。サンプ
ルあたりの費用はそれぞれ約 1500 円
(Rapid) および 3500 円 (Native) であっ
た。所要時間はそれぞれ約 1 時間 (Rapid)
および 3 時間 (Native) であった。DNA 抽
出法に関しては、広く用いられているス
ピンカラム型キットを用いてもロングリ
ードデータが得られることが示された。
地方自治体で ONT シークエンサーを用い
る際には、カラム抽出キットおよび Rapid
Barcoding Kit を用いることが多いと想定
されるが、その場合 1 日でシークエンス
開始が可能であり、20 株程度の大腸菌の
データが得られることが示された。

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を決定し、半数の
菌株のショートリードおよびロングリ
ード解読を行った。一般的な DNA 抽出お
よびライブラリー調整の条件で、約 20 株
の EHEC が一度に解析可能であることが明

らかとなった。

F. 健康危険情報

なし

G. 研究発表

1) 誌上発表

1. Y. Matsumoto, K. Lee, R. Akasaka,
H. Honjo, M. Koizumi, T. Sato, A.
Kubomura, N. Ishijima, Y. Akeda, M.
Ohnishi, S. Iyoda. 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. Appl Environ Microbiol. In press.

2) 学会発表等

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. 腸管出血性大腸菌(EHE
C)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.

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. 2022 および 2023 年に 5 株以上が分離された EHEC 集団感染事例件数および本研究における解析事例数

O 群	集団感染事例数	解析事例数
O157	20	10
O26	8	4
O111	3	3
O103	3	2
O121	1	1
O145	1	0
O172	1	0
O174	1	0
O5	1	0

表 2. DNA 抽出およびライブラリー調整キットの条件別にみたロングリード解析結果

DNA 抽出	ライブラリー調整	特徴	株数	平均	
				リード長 (kb)	Coverage (データ量)
MagMAX	Rapid	一般的な試薬	48	9.9	179
MagMAX	Native	手間・コストがかかる	15	3.8	313
DNeasy	Rapid	一般的な DNA 抽出	15	6.0	268