

厚生労働科学研究費補助金（食品の安全確保推進研究事業）
分担研究報告書

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

研究分担者 中村 佳司 九州大学・大学院医学研究院・講師

研究要旨

本研究班では、1) 反復配列多型解析法（MLVA法）運用の効率化・迅速化・精度向上およびMLVA型別データと発生動向を利用した腸管出血性大腸菌（EHEC）食中毒アラート検出システムの構築、2) EHECの全ゲノム配列（WGS）のデータベースの拡充化によるMLVA法との優劣性の解明およびWGS解析パイプラインの構築、3) 離島でのコホート研究によるEHEC感染症発生とその原因の関連性の評価、を行うことを目的としている。分担者は、研究代表者とともに2)の課題を担当するとともに、3)の課題で行われる動物分離菌株のWGSデータの収集を随時サポートし、WGSデータベースに組み込んで解析する役割を担っている。本年度は、2021年および2022年に国内で発生した集団食中毒事例（計39事例）に由来するEHEC菌株を対象に、WGS解析パイプラインによる解析結果を精査することで、本ツールの解析基準を詳細に検証すると共に、集団感染感知の基準となる指標を決定した。また、沖縄県衛生環境研究所で行われているゲノム解析の環境構築をサポートすると共に、九州地方の複数の衛生研究所から参加した研究員に対して、細菌のゲノム解析に関する情報提供を目的とした講演を行なった。さらに、ナノポアシーケンサのみを利用したEHEC菌株のゲノム構築に関する検討を行い、ナノポアシーケンサによる効率的な大腸菌の完全長ゲノム配列決定プロトコルを確立した。このプロトコルを用いて動物（牛）由来のEHECおよび腸管病原性大腸菌（EPEC）の完全長ゲノム配列を取得した。

A. 研究目的

腸管出血性大腸菌（EHEC）による食中毒は、その届出数や重症度と相まって日本の食中毒として最も警戒が必要な食中毒となっている。本感染症の事例調査のため分子型別法として、反復配列多型解析法（MLVA法）が迅速性等に優れていることから、国内では主にMLVA法を用いた解析が行われている。前身の研究班では、MLVA法によるEHEC菌株の全国的な分子疫学的解析に加えて、EHEC菌株の全ゲノム配列（WGS）のデータベース化を行った。その結果、ゲノム情報を利用してEHEC集団感染を早期探知するための基本的な情報基盤が構築でき、これを実際的なEHEC集団感染探知に利用できる段階に来ている。分担者は、本研究の柱の1つである、「EHECなどの検査のためのWGS解析法の確立に関する検討とEHEC株のWGSデータベースの拡充」を担当している。

本分担課題は、国際整合性の観点を考慮しつつ、研究代表者と連携して、EHEC食中毒探知の最前線である地方衛生研究所等で利用可能なゲノム解析パイプラインを構築することを目指している。このために、これまでに蓄積された国内分離株のゲノムデータに、EHEC感染症集団発生事例に由来する国内あるいは海外の菌株のデータを加えて、研究代表者のグループが開発したゲノム解析パイプラインによる解析を行い、このパイプラインの検証・解析結果を踏まえて、国際的に整合性のある同一クローンの判定基準等を確立することを目指している。また、これと並行して、新たに分離されたEHEC国内株のWGSデータを取得することで、前身の研究班から培われてきたEHECのWGS情報データベースをさらに拡充することも目的となる。これらに加え、本研究班では、MLVA及びWGS解析から得られたデータを利用して感染源や感染経路等に関する後方視的な疫学解析を行うことも目的としている。そのため、主感染源である食品からの分離菌株のWGSデータ収集を可能とする方策の検討と臨床分離株との比較、同様に動物由来株のデータ収集を可能とする方策の検討と事例調査への利用を試みることであり、本分担者は、動物（牛）由来株のWGSデータ収集を随時サポートし、必要に応じて、上記のWGS情報データベースに組み込んで解析する役割も担っている。

B. 研究方法

(1) ゲノム解析パイプライン（SNPcaster）における一塩基多型（SNPs）の検出条件に関する検討

前年度に研究代表者と共にWGSデータを取得した、2021年および2022年に分離された集団食中毒由来のEHEC菌株のうち、疫学的関連性がある菌株を3株以上含む集団（合計39集団；合計402菌株）を解析の対象とした。各集団内の菌株間のSNPsの数を以下に示すSNPcasterの標準設定で算出

した：i) 参照配列として、EHEC O157:H7 Sakai株のリピート領域を除いたゲノム配列を指定 ii) 組み替え領域を検出し解析対象から除外するGubbinsプログラム (Croucher NJ, et al. *Nucleic Acid Res*, 2015) を使用。次に、i)の参照配列を大腸菌K-12 (MG1655株)に変更し、MG1655株のリピート配列を除去しない配列を参照配列として解析を行なった。さらに、ii) Gubbinsプログラムを使用しない条件で、Sakai株あるいはMG1655株を参照配列として解析を行った。以上の計4条件 (Sakai株参照配列・Gubbins有；MG1655株参照配列・Gubbins有；Sakai株参照配列・Gubbins無；MG1655株参照配列・Gubbins無) で解析したSNPcasterの結果を比較することで、EHEC菌株に対するSNPcasterの解析基準の検討を行った。

(2) ナノポアシーケンサを用いたEHEC菌株のゲノム配列の決定

公共データベースに完全長ゲノム配列が登録されているO157:H7 Sakai株 (Accession No. ASM886v2)を対象として、Sandersonら (Sanderson ND, et al. *Microb Genom*, 2024) の方法に従い、ナノポアシーケンサで得られたリード配列のみを用いて、Sakai株のゲノム配列を決定した。その後、MUMmer (Kurtz S, et al. *Genome Biol*, 2004) にパッケージされているdnadiffプログラムを用いて、決定した配列を公共データベース上のSakai株の配列と比較し、両配列間の差異を確認した。また、高精度なゲノム配列構築に必要なデータ量の検討のため、Rasusaプログラム (doi:10.5281/zenodo.3731394) を用いて、取得した塩基配列データをランダムに抽出した。抽出したデータを用いて、再度ゲノム配列を構築し、公共データベース上の配列との比較を行った。

EHEC菌株のゲノム配列決定法の効率化を目的として、Sandersonらの方法で用いられていたDNA抽出キット (Qiagen Genomic Tip 100/G; Genomic-tip)をPromega社のWizard® HMW DNA Extract ion Kit (HMW kit) に変更し、DNA精製を行った。DNA精製過程において、RNAをRNaseによって分解する必要があるが、HMW kitに同梱されているRNaseでは、RNAの分解が不十分であることが明らかとなったため、Invitrogen社のRNase Cocktail Enzyme Mixを代わりに使用した。また、ゲノム配列構築に用いるプログラムをHybracterパイプライン (Bouras G, et al. *Microb Genom*, 2024) に変更した。以上の改良プロトコルを用いて、Sakai株のDNAを抽出後、ナノポアシーケンシングを行い、得られたリード配列を用いてゲノム配列を構築した。

(3) EHECおよびEPEC菌株の完全長ゲノム配列決定

前年度および今年度の離島コホート研究によって分離されたEHECおよびEPEC (計43株) を解析対象とした。各菌株をLB培地で37°C、一晚培養後、Genomic Tip 20/G (Qiagen)を用いてDNAを精製した。Rapid Sequencing Kit V14 (Oxford Nanopore Technologies; ONT)を用いてDNAライブラリを調製後、MinION Flow Cell R10.4.1 (ONT)によるナノポアシーケンスを行なった。シーケンシングによって得られた波形データをDorado v.0.8.3 (ONT; Super accuracy basecalling modeおよびDNA model v.5.0.0を使用)を用いて塩基配列に変換 (ベースコール) した。生成された塩基配列データ (リード配列データ) から、Hybracterパイプラインを用いて、ゲノム配列を構築した。43株のうち2株については、NanoFilt v2.8.0 (De Coster W, et al. *Bioinformatics*, 2018)プログラムを用いて、Quality score (Q score)が8あるいは20を越えるリード配列を抽出し、ゲノム配列構築に利用した。構築したゲノム配列を用いて、ECTyper v.1.0.0 (Bessonov K, et al. *Microb Genom*, 2021)プログラムおよびmlst program v2.23.0 (<https://github.com/tseemann/mlst>; Achtmanのスキームを使用)によって各菌株の血清型およびSequence type (ST)を決定するとともに、BLASTNに基づいた方法 (Nakamura K, et al. *PLoS Pathog*, 2021)によって、志賀毒素 (*stx*)遺伝子およびインチミン (*ea*)遺伝子の保有の有無を解析した。

(倫理面への配慮)

該当しない

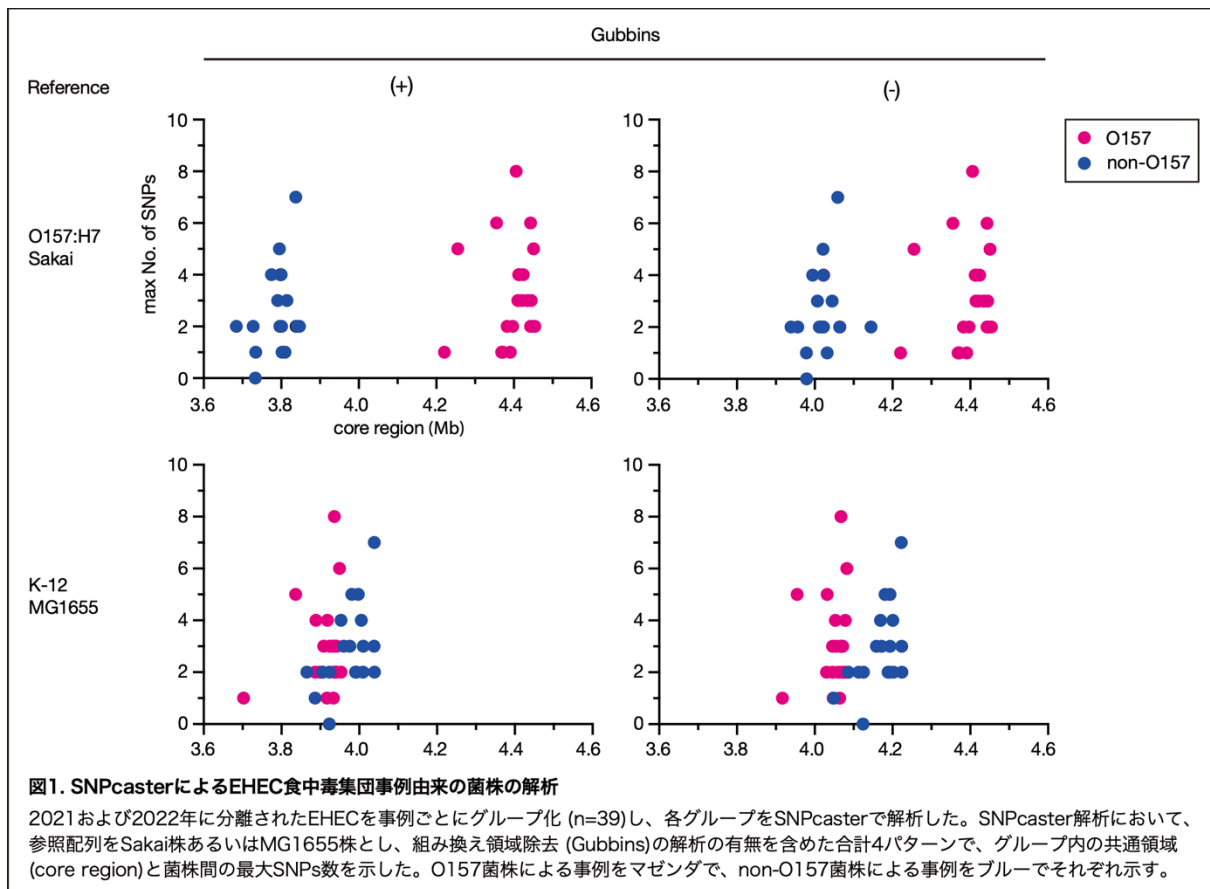
C. 研究結果

(1) 国内の食中毒事例に由来する菌株を用いたSNPcasterパイプラインの解析条件の検討

2021年および2022年に発生したEHEC集団食中毒事例から分離された菌株を対象として、SNPcaster解析を行った (図1)。解析セットには、EHEC主要血清型のO157:H7による事例 (n=21)のほか、non-O157として、O26:H11 (n=8), O111:H- (n=3), O103:H2 (n=3)などによる事例が含まれていた。SNPcasterの解析では、参照配列 (Reference)を含む全ての解析対象株に共通するゲノム領域 (core region) に存在するSNPsを抽出する。SNPcasterの標準設定 (図1左上)において、O157による事例を解析した場合には、事例内の菌株間のcore regionが4.2~4.4 Mbであったのに対し、non-O157による事例では、core regionが3.6~3.8 Mbとなったことから、O157とnon-O157ではcore regionに明らかな差が見られることがわかった。この一方で、参照配列としてK-12 MG1655株を用いた場合、O157およびnon-O157事例におけるcore regionの間に著しい差は見られなかった。core regionに含まれる組み替え領域を推定し、これを除去するGubbinsプログラムがcore regionに与える影響を調べた。参照株にO157 Sakai株を用いた場合、血清型が同一のO157による事例のcore regionは1事例を除いて全く変化しなかったが、non-O157による事例では、core regionに0.2 Mb程度の違いが見られた。参照配列としてK-12 MG1655株を用いた場合は、事例にかかわらず、Gubbinsを使用することでcore regionが0.2 Mb程度低下した。このことから、SNPcasterで参照株として標準設定されて

いる菌株と血清型のレベルで同一ではない菌株による事例では、参照株と解析株との間に組み替え領域と判定される領域 (0.2 Mb程度)がcore regionから除去されていることが明らかとなった。

各事例における菌株間の最大SNPs数を解析したところ、4つの解析パターンのうち、いずれの場合においても、最大SNPs数が10を越える事例は存在しなかった。解析菌株は疫学的関連性に基づいて事例ごとに分類されていることから、国内で発生している食中毒事例において疫学的関連性のある菌株をSNPcasterで解析する際の遺伝的な違いは、10 SNPs以下となることが示唆された。



(2) 地方衛生研究所のゲノム解析環境の構築に関する支援活動

WGS解析パイプラインが配布される地方衛生研究所におけるゲノム解析環境の構築のサポートを目的として、分担研究者 (大西)の所属する沖縄県衛生環境研究所を訪問した。本所では解析PCや使用する主要な解析プログラムの設定は終了していたので、所属研究員にゲノム解析における注意点、具体的には煩雑になる菌株のデータ管理の効率化についてアイデアを提案するとともに、個々の解析プログラムの使用方法や結果の解釈などを、研究員の疑問に答える形で説明した。また、九州地方の衛生研究所の研究員を対象として開催された、ゲノム配列解析の研修会に参加し、「病原細菌の全ゲノム解析とゲノム研究の概要」というタイトルで講演を行なった。研修会に参加したのはゲノム解析の設備が整っていない研究所に所属する、ゲノム解析の経験のほとんどない研究員であったため、講演では病原細菌のゲノム解析に必要な設備や、将来的なSNPcasterの配布を想定したゲノム解析に関する基礎知識の解説に焦点を当てた。講演終了後にそれぞれの研究員と個別に相談する機会を得ることができたため、ゲノム解析環境構築前の地方衛生研究所の状況に関する情報を収集した。

(3) ナノポアリードのみで構築したEHEC菌株の配列精度

SNPcasterにおける参照株として用いられているSakai株のゲノムは、2001年に完全長配列が決定され、2018年に配列修正が行われていることから、非常に高精度な配列データとなっている。Sakai株のゲノム配列をナノポアリードのみで構築し、その配列を登録済みの配列と比較することで、ナノポアリードのみで構築するゲノム配列の精度を検証した。ナノポアリードのみを用いて構築されたゲノム配列において、Sakai株の染色体と2つのプラスミド (pO157およびpOSAK1)の全てが環状化された。また、それぞれの

表1. ナノポアリードのみで構築したSakai株のゲノム配列のサイズと参照配列との違い

	染色体		pO157		pOSAK1	
	サイズ (bp)	SNV/InDel*数	サイズ (bp)	SNV/InDel数	サイズ (bp)	SNV/InDel数
参照配列†	5,498,578		92,721		3,306	
ナノポア	5,498,579	2/7	92,722	8/1	3,306	0/0

* SNV, Single Nucleotide Variation; InDel, Insertion and Deletion

† PacBioリード (ロングリード)とイルミナリードを用いて構築した配列 (Accession No. ASM886v2)

ゲノムサイズは2018年に登録された配列とほぼ一致しており、両者の配列の違いについても、染色体でわずか2塩基ミスマッチと7塩基の欠失・挿入、pO157で8塩基ミスマッチと1塩基の欠失・挿入であり、pOSAK1に関しては完全に一致していた（表1）。以上の結果は、ナノポアリードのみでSakai株のゲノムを高精度に構築できることを示唆している。次に、Sakai株ゲノム配列の全長の160倍（160xカバレッジ）に相当する全ナノポアリードからランダムにリードを抽出することで、ゲノム配列構築に使用するデータ量を調節し、高精度なゲノム構築のために必要なデータ量を検討した（表2）。その結果、50xカバレッジよりもデータ量が多い場合には、十分な精度のゲノム配列が構築でき、50xカバレッジを下回ると、染色体の環状化に影響が見られることが明らかになった。また、pOSAK1に関しては同一配列の重複に起因するゲノムサイズの拡大が確認されたが、これは使用するデータ量の調節のために行ったナノポアリード配列のランダムサンプリングが原因であると考えられた。

表2. ナノポアリードのデータ量とSakai株のゲノムサイズ・ゲノム環状化との関係

データ量*	染色体		pO157		pOSAK1	
	サイズ (bp)	環状化	サイズ (bp)	環状化	サイズ (bp)	環状化
160x†	5,498,578	○	92,720	○	3,306	○
100x	5,498,578	○	92,720	○	9,219	○
50x‡	5,498,586	○	92,721	○	9,682	○
40x	2 frags§	×	92,721	○	16,429	○
30x	5,498,599	○	92,721	○	16,497	○
20x	5 frags	×	92,721	○	3,581	○
10x	26 frags	×	92,725	○	9,069	○

* Sakai株のゲノムサイズ (5.6 Mb) を基準として算出したナノポアリードのデータ量
† ナノポアリードの全データ量 (結果の比較のため、表1で行った配列修正処理をしていない)
‡ 参照配列との違いは、染色体で2-SNVs/10-InDels, pO157で8-SNVs/1-InDelであった。
§ 染色体に該当するゲノム配列がN個に断片化 (n frags)

(4) ナノポアリードを用いたEHEC菌株のゲノム配列決定方法の検討

ナノポアシーケンスには長鎖DNAが必要であり、Qiagen社のGenomic-tipはDNA精製のゴールドスタンダードとして利用されている。しかしながら、Genomic-tipでは、陰イオン交換樹脂が充填されたカラムに菌体破砕液を添加することによりDNAを分離する、一般的なDNA精製キットとは異なる手法を採用しているため、多検体を処理することが困難であることに加え、菌体によっては破砕液がカラムに詰まることでDNA分離に手間がかかり、想定よりも長時間の作業を要することがある。そこで、Genomic-tipに代わるDNA精製キットとして、試薬添加と遠心分離のみで作業が完結する、Promega社のHMW kitでSakai株のDNAを精製し、これを用いて取得したナノポアリードの品質をGenomic-tipを用いて得たナノポアリードの品質と比較した。Genomic-tipによって精製したDNAに由来するナノポアリードの平均長は5,625-bpであり、シーケンスリードの長さの分布を示すN50 (値が大きいほど長いリード配列が得られていることを示す指標)は14,299-bpであった。一方、HMW kitによって精製したDNAに由来するナノポアリードでは、平均長が3,696-bp、N50が8,308-bpであり、いずれの値もGenomic-tipを下回ったことから、HMW kitで精製されたDNAはGenomic-tipによるDNAよりも剪断されており、その結果、ナノポアリード長が短くなったことが示唆された。この一方で、HMW kitによって得られたナノポアリードを用いて、Genomic-tipによるものと遜色ないSakai株のゲノム配列を構築することができた。このことは、HMW kitを用いて精製されたDNAであっても十分な精度のゲノム配列が構築できることを示している。

Sandersonらの方法では、ナノポアシーケンスを用いたゲノム配列の構築にはFlyeプログラム (K olmogorov M, et al. Nat Biotech, 2019) が使用されているが、本プログラムはプラスミドの配列構築が不得手であることが指摘されており (Johnson J, et al. Microb Genom, 2023)、ユーザー自身が構築されたプラスミド配列の正確性を検証する必要がある。また、Flyeプログラム後に、Medakaプログラム (ONT) を用いたゲノム配列の修正をユーザー自身が操作して実行し、修正前後の配列を比較して再度プログラムを使用するかを判断することとなる。以上のことから、既報の方法ではナノポアリードの取得後、配列構築のために複数の操作と解析に関する知識が必要となる。そこで、本ステップの簡略化およびプラスミド構築の高度化を目的として、Hybracterパイプラインの使用を検討した。本パイプラインでは、一度の操作によって、Flyeプログラムに続いてMedakaプログラムが実行され、配列修正の必要性も自動で判断される。また、プラスミド配列構築についてはPlassemblerプログラム (Bouras G, et al. Bioinformatics, 2023) が実行され、プラスミドの配列精度を高めている。Hybracterパイプラインを用いて、Sakai株のゲノム配列とプラスミドがナノポアリードから問題なく構築できることを確認した。また、後述するEHECおよびEPEC菌株においても、染色体およびプラスミドが問題なく構築されていることを確認した。このことから、ナノポアリードを用いたゲノム配列構築においてHybracterパイプラインが利用可能であると考えられた。

以上の検討結果を踏まえ、図2に示すゲノム配列プロトコルを構築した。

1. Genomic-tipあるいはHMW kitを用いたDNA精製
2. ナノポアシーケンス用ライブラリ調製
3. MinIONシーケンス
4. Hybracterパイプラインを用いたゲノム配列の構築

図2. EHECゲノム配列構築手順の概要

(5) ウシ由来EHECおよびEPEC菌株のゲノム配列決定

(4) で構築したプロトコルを用いて、離島コホート研究によって分離された計43のEHECおよびEPEC菌株の完全長ゲノム配列を決定した。DNA精製に関しては、HMW kitの検討中であったため、Genomic-tipを使用した。この菌株セットには、25種類の血清型あるいは24種のSTの大腸菌が含まれていた。染色体のゲノムサイズは4.8–5.8 Mbであった。プラスミドを保有していない菌株は存在しておらず、1つの菌株あたりの保有プラスミド数は1-11であった。以上のことから、今回

の菌株セットは多様な病原大腸菌を含んでいることが明らかとなり、少なくとも病原大腸菌においては、今回のプロトコルでほぼ確実に完全長ゲノム配列の構築が可能であることが示唆された。

D. 考察

細菌の進化の過程では、ファージなどの可動性遺伝因子の獲得や、染色体内において互いに類似した配列間で相同組み替えが起きることがある。これらに起因するSNPsは突然変異に由来するSNPsとは本質的に異なることから、系統解析において細菌を区別する情報として利用されない。SNPcasterで行われるSNP解析においても、標準設定では、解析対象株に共通するゲノム領域（コアゲノム配列）から、可動性遺伝因子領域と、Gubbinsプログラムを用いて推定された組み替え領域を除去している。しかしながら、様々な年代に、様々な由来から分離された細菌を解析セットに含めることが多い系統解析とは異なり、集団食中毒事例に由来する菌株間の異同を判定する場合は、原因（食品など）から分離された菌株と患者から分離された菌株の間は遺伝的に著しく近縁であるという仮定に基づいて解析するため、これらの分離菌株間で、可動性遺伝因子の獲得や相同組み替え発生による違いが発生する可能性は低い。実際に、病院内の同一患者から継続的に分離された菌株のセットを用いた解析により、可動性遺伝因子領域や組み替え領域は菌株間の異同の判定に影響をほとんど及ぼさないことが報告されている（Gorrie CL, et al. Lancet Microbe, 2021）。国内のEHEC集団感染事例の菌株を用いた今回の解析においても、菌株間の最大SNPs数はGubbinsプログラム使用の影響を受けなかった。Gubbinsプログラムでは複雑な計算が行われるため、解析に使用するPCに与える負担が他のプログラムと比較して大きく、そのため結果の出力に要する時間に著しく影響することがある。地方衛生研究所での用いられるPCの性能によっては、Gubbinsプログラムが解析の負担となる可能性もあるため、SNPcaster配布時にGubbinsプログラムに関する正確な情報提供が必要であると考えられた。今回、可動性遺伝因子の影響は検討しなかったが、SNPcasterではSakai株以外の参照配列を使用する場合は、ユーザー自身が除去すべき領域として可動性遺伝因子を指定することになるため、こちらについても可動性遺伝因子領域指定の必要性に関しての状況説明が必要であると思われる。

SNPcasterでは、解析したい菌株のWGSをアラインメントするための参照配列が必要であり、標準設定ではO157:H7 Sakai株が指定されている。今回の解析により、Sakai株を標準株に選んだ場合に、O157とnon-O157の集団事例の間で、SNPcaster解析の出力結果のひとつであるcore regionのサイズに大きな違いがあることが明らかとなった。これは、Sakai株がO157 EHECであることから、Sakai株と同一血清型のO157菌株を解析する場合には、Sakai株と解析株の間で共有される領域が広がるため、結果的にcore regionのサイズが大きくなることに起因している。一方、MG1655を用いた場合には、O157とnon-O157のcore regionサイズは大差がなかった。MG1655は大腸菌の一般的な実験用の株で、非病原性であり、病原性大腸菌とは遺伝的に異なる。このため、今回の解析対象株のいずれとも、同程度のcore region（おそらく大腸菌株をSNPcasterで解析した場合の最低限の領域）を確保したと考えられる。プログラムのオプション設定を除けば、参照配列株はユーザーが唯一指定しなければいけないパラメータであり、実際に、九州の衛生研究所職員を対象とした研修会において、参照株には何を選択するべきかという質問を受けた。Sakai株を設定した場合、解析対象がO157かそれ以外かで、少なくともcore regionのサイズの標準値が異なる。これに対し、MG1655株では、ほとんどの解析対象で類似したcore regionとなると予想され、その標準値から大きく異なる値が出力された場合は、何らかのエラーを推測できる。また、菌株同士の異同の判定基準となる最大SNPsについて、今回の解析セットでは、Sakai株とMG1655株の間に明確な差は確認できなかった。従って、大腸菌に関してはMG1655株を参照株に設定することをオプションの一つとして提案できると考えられた。

国内の合計39の集団事例を用いた解析の結果、疫学的関連性のある菌株間のSNPs数は8以内に収まることが明らかとなった。前年度行った海外事例（計5事例）の菌株を用いた解析では、5 SNPsが同一クローン判定のひとつの基準となることが示唆されていた。また、前年度の解析により、互いに疫学的関連性のない菌株同士の遺伝的距離は13 SNPsよりも大きくなることも明らかとなっている。これらのことから、国際的に整合性のある同一クローンの判定基準の候補として、10 SNPsはある程度妥当な数値であると考えられた。

ナノポアシーケンサによって得られる配列はロングリード配列と言われ、イルミナシーケンサによるショートリード配列と比較して、リードあたりの情報量が飛躍的に増加するため、細菌の染色体およびプラスミドのゲノム配列を完全に構築する点において、非常に有用となる。この一方で、ナノポアリードの塩基配列の精度が課題となり、ナノポアリードを使用してゲノム配列を決定する場合にはイルミナリードと併用する必要があった。しかしながら、近年の技術革新により、ナノポアリードのみで細菌の高精度な完全長ゲノム配列を構築可能であることが報告されている（Sanderson ND, et al. Microb Genom, 2024）。ナノポアシーケンサのための設備に係る費用は、地方衛生研究所に配備されているイルミナシーケンサ(iSeq 100)よりも圧倒的に安価である

（現時点の価格で約150万円程度）。また、iSeq 100の販売は2025年9月末で終了し、システムサポートは2029年末で終了することが告知されている。このことから、将来的なイルミナからナノポアへの切り替えの可能性を想定し、ナノポアシーケンサを用いたゲノム配列決定プロトコルを構築しておくことは重要であると考えられる。本研究により、大腸菌に関しては、ナノポアシーケ

ンスのみで精度の高いゲノム配列を安定して構築できることが明らかとなった。DNA精製において特に多検体を処理する必要がある場合は、HMW kitを利用することで作業を効率的に進めることが可能となるが、Genomic-tip使用時と同様のゲノム配列を常に構築できるかどうかは、Sakai株以外の様々なEHEC菌株を用いて検証する必要がある。また、先述の通り、ナノポアシーケンスには長鎖DNAが必要であり、長鎖DNAをいかに調製するかがシーケンスの成否に大きく影響する。長鎖DNAの調製法は菌種によって異なることが予想されるため、大腸菌以外の食中毒起因菌に関してはWGSデータベースの構築に加えて、ナノポアシーケンスを利用する場合はDNA精製方法を最適化することが将来的な課題のひとつであると考えられる。

E. 結論

SNPcasterパイプラインの標準設定および解析基準について、以下の条件を決定し代表者に提案した：

- ・ 参照配列として、MG1655株 (NCBI accession No. NC_000913.3)の染色体のすべての領域を使用する。
 - ・ Gubbinsプログラムは、解析環境に応じて使用しない設定でも大きな問題は発生しない。
 - ・ 解析後に得られるcore regionは、Gubbinsプログラム有では3.7-4.0 Mb、Gubbinsプログラム無では3.9-4.2 Mbとなることが予想される。
 - ・ 菌株間のSNPsが10以下の場合、同一クローンに由来する可能性が非常に高いと判定できる。
- また、ナノポアシーケンサのみを用いたゲノム配列決定法を確立し、その有用性を確認した。これを用いて、離島コホート研究で得られた病原大腸菌のゲノム配列を決定した。ゲノム配列決定に関する成果については、来年度論文報告する予定である。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

なし

2. 学会発表

なし

H. 知的財産権の出願・登録状況（予定を含む。）

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし