

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

腸管出血性大腸菌（EHEC）感染症の事例調査のための分子型別法として、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、国内では本法が主に用いられてきた。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、これを実際のEHEC集団感染探知に利用できる段階に来ている。本研究班では、1) 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も担っている。本年度は、研究代表者と連携して、大腸菌の全ゲノム配列を用いた解析パイプラインを構築するとともに、集団感染感知の基準となる指標について検討した。また、臨床あるいは動物（牛）由来のEHEC菌株および健康牛から排菌される大腸菌の全ゲノム配列データ（ドラフトゲノム配列）の取得を行なった。

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、これを実際の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) 研究代表者が開発したゲノム解析パイプライン（SNPcaster）による一塩基多型（SNPs）の検出条件について、海外のEHEC感染症集団発生事例に由来する菌株のデータを用いて、i) 解析対象の配列に対する参照配列の設定、ii) 同一クローンの判定基準となるSNPs数の閾値を検証した。公共データベース（NCBIデータベース）からEHECの集団感染に関するデータセットのリストを抽出後、同一事例内に3株以上登録されているEHECのWGSデータをダウンロードし、合計5事例（イタリア1事例、フランス2事例、アメリカ2事例）のWGSデータを取得した。

(2) 研究代表者のグループが収集した、2021年・2022年に国内で発生したEHEC集団感染事例（5

検体以上の事例；分離源のヒトが全て無症状であった事例も含む）の分離株（計426株）のうち、226菌株のWGS情報を取得した。また、離島コホート研究の過程で収集された、2つの離島で飼育されている牛から分離された病原大腸菌（EHECおよび腸管病原性大腸菌（EPEC））および大腸菌群スクリーニング用の培地（XM-G培地）あるいは主要EHECの推定鑑別培地（CHROMagar STEC）から釣菌した細菌のWGS情報も取得した。WGS情報の取得には、イルミナ社のNovaSeqを利用し、ショートリード配列を取得した。離島コホート研究の分離菌については、shovillプログラムを用いて作成したドラフトゲノム配列の品質を、checkMプログラムを用いて評価した。その後、ECTyperプログラムを用いて菌種推定と大腸菌の血清型別を行うとともに、志賀毒素（*stx*）遺伝子およびエンチミン（*eae*）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）SNPcasterパイプラインの活用法についての検討：

i) 解析対象の配列に対する参照配列の設定：SNPcasterで行われるSNP解析には、BactSNPプログラムが採用されている。このプログラムでは、解析したい菌株のWGSをそれぞれ参照配列にアラインメントし、参照配列を含む全ての解析対象株に共通するゲノム領域（コアゲノム配列）に存在するSNPsを抽出する。EHEC血清型間にはゲノム多様性があることから、解析対象と遺伝的に近縁でない（例えば、異なる血清型の）菌株のゲノム配列を参照配列として設定すると、コアゲノム配列が縮小され、菌株間のSNPs数が過小評価される可能性がある。本研究で収集した海外5事例のEHEC血清型はO157（2事例）あるいはO26（3事例）であった。O157の菌株の解析にO26菌株のゲノムを、O26の菌株の解析にO157菌株のゲノムをそれぞれ参照配列とした場合、血清型が一致している場合と比較して、コアゲノム配列が0.3~0.4 Mb縮小され、抽出SNPsも最大で3 SNPs少なくなった。本研究では解析の解像度を最大にするため、解析対象の菌株の血清型と同一の血清型のゲノム（O157の場合はSakai株、O26の場合は11368株）を参照配列として、SNP解析を行なった。

ii) 同一クローンの判定基準となるSNPs数の閾値の検証：NCBIに登録されている菌株セットの管理ID（BioProject ID）のうち、イタリアおよびフランスの事例（計3事例、全てO26 EHEC）の管理ID内には、それぞれ単一事例に由来するEHEC菌株のWGSデータが登録されており、菌株間の最大SNPs数はいずれの事例においても5 SNPs以下であった。アメリカのO157による2事例のうち、1事例（BioProject ID: PRJNA517910）には、2018年の大規模集団事例時にウィスコンシン州で分離された菌株のWGSデータが登録されており、菌株同士の疫学的関連性が不明であったため、本解析から除外した。残りの1事例（BioProject ID: PRJNA215830）についても、複数の事例に由来する菌株のWGSデータが登録されていたが、菌株の疫学情報が論文報告されていた（Turabelidze G, et al. J Clin Microbiol, 2020）ため、菌株同士の疫学的関連性と遺伝的距離（菌株間のSNPs数）の対応を精査した。その結果、同一感染事例の原因と考えられる菌株同士の遺伝的距離は、1つの菌株を除いて5 SNPs以内であり、例外の1株とその他の菌株の遺伝的距離は11~12 SNPsであった。また、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が分かった。これらの結果は、海外のEHEC集団感染事例において、5 SNPsが同一クローン判定のひとつの基準となることを示唆している。

（2）国内分離株のWGSデータの収集：

計226の臨床由来EHEC菌株に関して、平均740 Mb（387~1,368 Mb）のショートリード配列を取得した。代表的なEHEC菌株であるO157 Sakai株ゲノムの全長（5.6 Mb）に対し、平均132xカバレッジ（69x~244x）のデータ量であったことから、WGS解析に十分なデータ量がいずれの菌株においても取得されていると考えられた。

離島コホート研究から分離された計155の牛由来の菌株に関して、平均806 Mb（560~1,225 Mb）のショートリード配列を取得した。ショートリード配列をアセンブルして得たドラフトゲノム配列のうち、8つについては複数の菌株のゲノム配列が含まれていると判断した。残りの147株のうち、XM-G培地で分離された菌株は122株含まれており、そのうち119株が大腸菌であることがWGSを用いた解析から推定された。CHROMagarで分離された9菌株は全て大腸菌であり、そのうち2株は*eae*遺伝子陽性であった。離島コホート研究を担当する研究者が事前に行なったPCR解析（*stx*, *eae*, およびEHEC主要O血清群）により、EHECあるいはEPECであると推定された16菌株のうち、15株は*stx*と*eae*遺伝子のどちらか、あるいは両方が陽性であった。残りの1株の血清型はO121:H11であり、O血清型はEHEC主要血清型のひとつであるO121:H19と共通するが、H血清型は異なっていた。

D. 考察

（1）SNPcasterパイプラインの活用法についての検討：

前身の研究班から既に報告されているとおり、WGS解析パイプラインは各国あるいは研究機関によって様々であり、集団事例に関連するクローンの判定基準についても未だに国際的なコンセンサスは得られていない。従って、国内で使用するべき解析パイプラインに関して、国際的に整合性のあるデータフォーマットや同一クローンの判定基準等を確立することは難しく、WGSの幅広い活用のために独自のパイプラインを構築する必要がある。今回、研究代表者のグループが開発したSNPcasterプログラム（試用版）では、参照配列をユーザー自身が指定することになっている

が、参照配列はSNPs数に直接影響することが今回の解析で明らかになったため、国内でのクローンの判定基準を統一するために、参照配列は指定しておくべきであると考えられる。国内で集団感染を引き起こすEHEC血清型は主に、O157:H7, O26:H11, O111:H8であることから、少なくともこれらの血清型については、代表菌株の参照配列を予め用意する必要がある。この3つの血清型以外の病原大腸菌に関しては、解析対象の菌株のうちのひとつのWGSを参照配列とすることによって、予め参照配列を指定したEHEC血清型と同様の判定を行うことができると考えられるが、同一クローン判定の基準を上記の3つの血清型と同じにして問題がないかどうかは、実際の事例の解析を通じて検討していく必要がある。

海外の集団発生事例に由来するEHEC菌株のWGS解析から、同一クローン判定の基準（5 SNPs）を明らかにした。この結果は、研究代表者が行なった国内のEHEC菌株の解析結果とほぼ一致しており、少なくともSNPcasterパイプラインを用いた場合にまず考慮すべき基準であると思われる。一方、アメリカの事例において、疫学的関連性がある菌株の間にも12 SNPsの遺伝的距離があり、疫学的関連性のない菌株同士であっても遺伝的距離が13 SNPsとなったことは、WGS解析の判定において、同一クローン判定の基準によっては無関係な事例の菌株同士を結びつける可能性があるという点で、注意すべき結果である。従って、今後国内株のWGS配列を用いて、同一のEHEC血清型であるが、疫学的関連性のない菌株同士の遺伝的距離の情報も蓄積していく必要があると考えられる。

(2) 国内分離株のWGSデータの収集：

計226の臨床由来EHEC菌株について、WGSデータを収集することができたため、今後、研究代表者と連携して、上述のSNPcasterパイプラインの検証にこのデータを活用するとともに、EHECデータベースのアップデートを行う。離島コホート研究によって牛から分離されたEHEC菌株、特にEHEC主要血清型ではないが、stx遺伝子陽性の菌株に関しては、参照配列あるいはlandmarkとなりうると考えられるため、完全長あるいはそれに近い高精度配列を準備するとともに、分担者が独自に開発した定量系を用いて、Stxの産生量を測定する予定である。牛から分離されたnon-EHEC菌株は、EHECデータベースの拡充には直接関連しないが、牛の腸管内でEHECと共存している可能性があるという点で重要であるため、典型的なEHEC菌株との系統関係を解析するとともに、離島コホート研究を分担する研究者とともに今後の研究方針を議論していく予定である。

E. 結論

国内独自のパイプラインであるSNPcasterについて、海外のEHEC集団感染事例のWGSデータを用いて、本パイプラインの使用に必要な参照配列の検討と集団感染事例に由来する菌株同士の遺伝的距離の決定を行い、SNPcasterにおける同一クローン判定基準を決定した。

WGS情報データベースの構築に関しては、2021年と2022年にEHEC集団感染事例の分離株と離島で飼育されている牛由来の分離株のWGSデータを収集した。さらに、EHECのリザーバーである牛の腸管内に存在する大腸菌のWGSデータも取得した。

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