

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  
「食中毒調査の迅速化・高度化及び広域食中毒発生時の早期探知等に資する研究」  
(20KA0501)  
総合研究報告書

研究代表者

明田幸宏 (国立感染症研究所)

研究分担者

林哲也	(九州大学大学院医学研究院細菌学分野 教授)
砂川富正	(国立感染症研究所実地疫学研究センター センター長)
工藤由起子	(国立医薬品食品衛生研究所衛生微生物部 部長)
寺嶋 淳	(岩手大学農学部共同獣医学科 教授)
平井晋一郎	(国立感染症研究所感染症危機管理研究センター主任研究官)

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 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。2020-2022 年の間に約 7700 株について MLVA による型別を実施し、厚生労働省 NESFD において MLVA リストの基盤データを提供した。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 型の付与が行われている。2020-2022 年の間に約 2200 株について解析し型名を付与した。このうち約 6 割の株について感染研でも MLVA 法による解析を行い、データの精度確認を行った。

国内 EHEC のサーベイランスを全ゲノム配列解析で効率的に行うために、core genome single nucleotide polymorphism (cgSNP) および core genome multilocus sequence typing (cgMLST) パイプラインを構築するとともに、国内株ゲノムデータのデータベース化を行った。同パイプラインを用いることで、新たに EHEC 分離株ゲノムが得られた際に、国内外の大腸菌ゲノムとの比較解析を迅速に行える体制が確立した。加えて、同環境を地方衛生研究所にも試験的に導入した。また、Oxford Nanopore 社 (ONT) シークエンサーによるデータ解析を地方衛生研究所と協同して行った。この結果、全ゲノム配列解析をサーベイランス等で用いる際の課題として、データおよび解析手法の共有、ONT データの解析法の確立が必要であることが明らかとなった。

## 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」により、反復配列多型解析 (multi locus variable 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。

近年、EHEC 菌株の解析にゲノム配列 (whole-genome sequence : WGS) が広く用いられており、Enterobase (<https://enterobase.warwick.ac.uk/species/index/ecoli>) などの公共データベースには25万件以上の大腸菌ゲノム情報が登録されている。また、感染研・細菌第一部ではこれまでに9,000株以上のEHECゲノムを解読しているが、それらゲノム既読株と新規ゲノムとの網羅的な比較には、数日単位での時間がかかるなどの課題があった。そこで、新規ゲノムを入手した際に国内外の大腸菌ゲノムと迅速な比較解析を可能にする手法の確立を行った。さらに、最適な単一塩基多型 (single nucleotide polymorphism : SNP) 解析条件を明らかにする目的で、参照配列と解析するデータセットといった条件が型別能に与える影響について、EHEC O76:H7 および O111:H8 の WGS データをもとに検討し

た。

また近年、多くの地方衛生研究所において次世代シーケンサーが導入されたが、Oxford Nanopore 社 (ONT) シーケンサーについては、SNP 等を用いた菌株間の関連性解析の知見が少ない。そこで、沖縄県衛生環境研究所 (沖縄衛研) と協同して、ONT データを用いた病原性大腸菌事例の解析を行った。さらに、イルミナシーケンサーによるデータも含めた情報解析環境の構築を、沖縄衛研にて試験的に行った。

## B. 研究方法

### 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2020年~2022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。

### 腸管出血性大腸菌等の検査法(全ゲノム解析)の開発

Core genome (cg)SNP は、解析時に供試する全株に共通する領域 (コアゲノム) を抽出する必要があるため、計算機への負荷が高く、時間がかかるため、snippy (<https://github.com/tseemann/snippy>) および BactSNP (<http://platanus.bio.titech.ac.jp/bactsnp>) を

組み合わせた解析パイプラインを構築した。事前の検討では、snippy は再解析が容易であるものの一部のデータで多数の SNP を誤って検出することが示されている。一方、BactSNP は精度の高い SNP 抽出が可能であるものの、SNP のデータベース化や数百株以上の解析が困難であった。そこで、BactSNP で得られた SNP 情報を snippy で利用できる形に変換するプログラムを作製することで、高精度な SNP のデータベース化と迅速な再解析が可能となるパイプラインを構築した。一方、cgMLST は大腸菌の大部分の株が保有する 2,513 遺伝子を対象とした MLST であり、Enterobase 上で手法が公開されている。そこで、各遺伝子のアリアル情報をダウンロードし、BLAST 解析をベースにした自製 cgMLST 解析パイプラインの構築を行った。以上のパイプラインを利用し、これまでに細菌第一部で WGS を解読した EHEC 計 2,248 株の cgSNP および cgMLST データベース化を行った。さらに、新たに EHEC のゲノム情報が得られた際に、迅速に国内外の大腸菌ゲノムとの比較を容易にするパイプライン構築を行った。

#### EHEC 調査における最適な SNP 解析法の検討

EHEC O76:H7 および O111:H8 の菌株を対象に、より高精度な SNP 解析法の検討を行った。SNP 解析時には、参照配列および解析を行うデータセットが重要と考えられたため、それぞれに複数の条件を設けて解析を行った。O76:H7 では、参照配列として、近縁 (O76:H7 JNE132847

株の完全長ゲノム配列) または遠縁 (O157:H7 Sakai 株) の 2 株を用いた。データセットとしては、多様な大腸菌を含む 34 株 (O76:H7 8 株を含む) または O76:H7 8 株のみでの解析、の 2 条件で行った。O111:H8 の解析では、計 878 株分のゲノムデータを用いた。この中には 79 事例のクラスター (疫学関連のある事例) が含まれており、このうち 10 株以上のクラスター 15 事例を対象とした。参照配列としては、O111:H8 の代表株 (11128 株)、クラスター内の株、または O157:H7 Sakai 株の 3 種を用いた。データセットとしては、O111:H8 878 株をまとめて解析、またはクラスター内の株のみで解析、の 2 条件を設けた。SNP 解析は、1 年度目で確立した BactSNP および snippy などを用いた解析パイプラインを用いて行った。

#### 地方衛生研究所と協同した WGS 解析

沖縄県で 2022 年に発生した腸管病原性大腸菌 (EPEC) O153 感染事例について、協同して WGS 解析を行った。本事例では、家族内感染 2 株および散発例 1 株が近い時期に分離されていた。そこで、沖縄衛研にて 3 株のゲノム DNA を抽出し、Rapid Barcoding Kit (Oxford Nanopore) でライブラリー抽出した後に、MinIONにて R9.4 のフローセル (Oxford Nanopore) を用いてシーケンス解析を行った。得られた Fast5 ファイルを感染研に送付後、Guppy basecallerにて、3 種のモデル (Rapid, High accuracy, Super accuracy) でベースコールを行った。それぞれの条件で得られた FastQ ファイルを用いて、Tricyclerにてアセンブリを行った。アセンブリによ

って得られたコンプリートゲノムから ART を用いて疑似ショートリードを作製し、Lee ら (2021. Emerg Infect Dis 27:1509-1512) の方法を用いて cgSNP を抽出した。また、解析を行った 3 株については、MiSeq (Illumina) でも全ゲノム配列の解読を行った。上記と同様の方法で cgSNP を抽出し、ONT データによる結果との比較を行った。

### 地方衛生研究所における WGS 解析環境の構築

沖縄県衛生環境研究所における WGS 解析環境を構築するために、同所のワークステーションへの Linux 仮想環境の構築および感染研・細菌第一部使用プログラムのインストールおよび動作確認を行った。

### C. 研究結果

#### 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、2020-2022 年分離株 7665 株について分子型別解析を実施した。このうち 6695 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 102 施設であった。各血清群における株数、同定された型数は以下のとおりであった。O157: 4373 株、1644 型; O26: 1266 株、501 型; O111: 363 株、176 型; O103: 384 株、121 型; O121: 120 株 59 型; O145: 70 株、37 型; O165: 17 株 15 型; O91: 102 株、83 型。得られたデータは毎年 5 月号の IASR の EHEC 特集号において公表された。

2020-2022 年分離株について、MLVA 型

別を実施しデータを感染研に送付した地方自治体は、36 施設であった。感染研において統一型名を付与した菌株数は 2209 株であった。このうち 1323 株については、菌株が感染研に送付された。感染研の結果と比較し精度確認が行われた。感染研の結果と一致したものは 93%であり、それ以外の株もほとんどすべてが 1 遺伝子座違いであった。

2014-2021 年に発生した集団事例のうち、菌株数 10 以上の事例 106 件、約 2600 株について、各事例内で各遺伝子座の標準偏差を算出した。その分布から、事例内におけるリピート数の変動は EHC-6、O157-37 を除きほとんどなく、95 パーセントイル値で 0.4 以下であった。EHC-6 は 5.9、O157-37 は 2.2 であった。

#### 腸管出血性大腸菌等の検査法(全ゲノム解析)の開発

BactSNP および snippy を利用した解析パイプラインの構築により、SNP のデータベース化が可能となった。SNP の抽出には 1 株あたり 20 分程度を要する。既に SNP 情報が存在する株を対象とした再解析では、100 株の解析が 19 秒、500 株の解析が 75 秒で可能であった。また、cgMLST 解析では、98%がデータベース上に同一型が存在しない新規 ST となった。cgSNP および cgMLST データベース化を行った 2,248 株の血清型は、表 X1 のような割合であった。

#### EHEC 調査における最適な SNP 解析法の検討

O76:H7 を用いた解析における条

件ごとの最尤法系統樹を図 X1 に示す。最も型別能が高い（系統樹の枝の長さが長い、株間の SNP 数が多い）条件は、参照配列として O76:H7 を参照配列として用い、同血清型のみで解析を行った場合であった（図 X1D）。しかし、O76:H7 を参照配列として用いた場合でも、遠縁な株を含めて解析した場合には最も型別能が低くなった（図 X1C）。一方で、参照配列が別血清型（O157:H7 Sakai）であっても、O76:H7 のみで解析を行った際には、2 番目に高い型別能となった（図 X1B）。

O111:H8 の解析では、同血清型代表株を参照配列として全株を解析した条件を基準に、他の条件との比較を行った。その結果、クラスター内の株のみで解析した場合には参照配列が代表株の場合でも、クラスター内の場合でも SNP 数の増加分に大きな違いは見られなかった（図 X2）。一方で、参照配列を別血清型（O157:H7 Sakai）とした場合には、参照配列が O111:H8 代表株である場合と比べて型別能が劣る例も認められた。

#### 地方衛生研究所と協同した WGS 解析

Oxford Nanopore シークエンサーによるデータを解析した結果、ベースコールモデルによって結果が大きく異なっていた（図 X3）。すなわち、精度が最も低い Rapid によるベースコールでは、最も多数の SNP が認められた。精度が中間の High accuracy および精度が最高の Super accuracy によるデータでは、菌株 22-13 および 22-18 間の SNP の数は同一であり、ゲノム中の位置も概ね同一であった。High accuracy では 22-17 および 22-18 間

の SNP は検出されなかったが、Super accuracy では両株間に 2 か所の SNP が認められた。イルミナシークエンサーによる cgSNP 解析では、22-13 および 22-17/22-18 間で 2 か所の SNP が認められた。22-17 および 22-18 間での差異は認められなかった。これらの SNP の位置は、ONT における High accuracy データと概ね一致していた。各条件で得られたコンプリート配列を比較したところ、Super accuracy およびイルミナ（ONT データで得られた配列をイルミナデータで修正した配列）によって得られた配列間では 47 か所、High accuracy およびイルミナ間では 241 か所の違いが認められた。これらの大部分はホモポリマー（同一塩基が連続する配列）上に存在していた。

#### 地方衛生研究所における WGS 解析環境の構築

沖縄衛研の解析ワークステーションに、基本的な WGS 解析（アセンブリおよび cgSNP 解析）を行うことが可能な環境を構築した。同環境は、現在沖縄衛研においてセレウス菌等の食中毒調査に活用されている。

#### D. 考察

MLVA 法により解析した菌株数は 3 年間で 7000 株を超えた。解析結果は定期的に厚生労働省 NESFD に MLVA リストとして掲載された。地方衛生研究所から送付された MLVA データは 3 年間で 2200 株余りに上った。このうち 6 割の株が後日感染研に送付され、感染研で実施した MLVA データと比較された。結果として

は 93%が一致し、一致しなかった株についても 1 若しくは 2 遺伝子座の違いのみであった。これらの地衛研については技術的な問題は概ねないと考えられた。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。

集団事例内の MLVA のリピート数のばらつきはほとんどなく、95 パーセントイル値で 0.4 以下であった。95 パーセントイル値が高かった EHC-6 と O157-37 はプラスミドにあるとされる遺伝子座で、当該座位の欠失などの影響から変動が大きかったことが推察された。これ以外の遺伝子座については集団事例内でほとんど変化しないことが示唆され、今後のデータ解析の指標になることが期待される。

cgSNP および cgMLST 解析手法の確立を行い、国内 EHEC 計 2,248 株の WGS データのデータベース化を行った。本手法によって、新たに集団感染株等が得られた際に迅速に近縁株を抽出することが可能となった。特に、O157 以外の血清型では MLVA が利用可能でない O 群や、利用可能な O 群であっても十分な型別能を有さない事例が存在する。そのような事例では、WGS 解析が特に有効であると考えられる。

サーベイランスに用いる SNP 解析法の検討では、近縁な株を参照配列とし、近縁な株のみで解析することでより型別能の高い解析が行えることが示された。O111:H8 の解析からは、参照配列として用いるのが同一血清型であれば、クラスター内の株を参照配列として用いた場

合と比べて同等の型別能が認められることが明らかとなった。完全長ゲノム配列を参照配列に用いることの利点は、組換え領域等をエラーが起りやすい領域をより正確に除けることであるが、本解析では型別能に大きな影響は与えなかった。また、用いるデータセットの影響は、参照配列の影響よりも大きいことが O76:H7 の解析から明らかとなった。これは、遠縁な株を解析に含めることでコアゲノムサイズが小さくなる結果であると考えられる。このため、cgSNP のデータベース化では参照株を EHEC O157 Sakai 株に統一しているが、解析菌株を限定することで cgSNP の減少を最小限にとどめることができることが示された。集団感染等の調査の際には、まず cgSNP データベースを用いた全体的な系統解析を行い、次に近縁株を参照配列として解析をすれば、効率的で精度の高い解析が行えることが明らかとなった。

地方衛生研究所と協同した WGS 解析の結果、ONT シークエンサーによって得られたデータは、High accuracy または Super accuracy のベースコールを行うことで、イルミナデータと同等の cgSNP 解析が行える可能性が示唆された。ONT によるデータは精度が低いとされる。実際にホモポリマー部分ではイルミナデータとの齟齬が認められたが、このような部分は insertion/deletion (indel) とされ cgSNP 解析には含まれない。このため、cgSNP 解析結果には影響がなかったと考えられる。今後、より多数の事例の解析によって、ONT データ解析の妥当性を評価する必要がある。一方、ONT データは容量が非常

に大きい（数 G byte 以上）ため、送付時にデータが破損する例が見られた。このため、地方衛研である程度の解析を行った後に情報を共有するのが効率的と考えられた。

データ解析環境の構築では、感染研・細菌第一部で使用中のプログラムのインストールを行ったが、一部専門知識が必要であり、情報解析に慣れない担当者が行うのは困難であることが判明した。今後、より簡易に解析環境を構築可能な体制を整える必要がある。

## E. 結論

MLVA 法は迅速性、精微性に優れた分子型別手法であり、MLVA リストを共有することで広域食中毒等の早期探知に有効である。地衛研から送付される MLVA データは迅速性の向上に重要であり、今後も当該データの信頼性確保のため、モニタリング等の方策が必要である。

本研究では、国内 EHEC ゲノムのデータベース化と国内外データベース株との効率的な比較解析パイプラインの構築を行った。今後、地方衛生研究所とも協同して、データ、解析手法の共有化や同手法の検証を行っていく必要がある。

## F. 健康危険情報

なし

## G. 研究発表

### 1) 誌上発表

1. Kimata K, Lee K, Watahiki M, Isobe J, Ohnishi M, Iyoda S. Global distribution of epidemic-related Shiga toxin 2 encoding

phages among enteroaggregative *Escherichia coli*. *Sci Rep*. 2020 Jul 16;10(1):11738. doi: 10.1038/s41598-020-68462-9.

2. Kato H, Yahata Y, Hori Y, Fujita K, Ooura N, Kido T, Yoshimoto K, Matsui T, Izumiya H, Ohnishi M, Oishi K. A shigellosis outbreak associated with a sports festival at a kindergarten in Kitakyusyu City, Japan. *J Infect Chemother*. 2020 Jul 2:S1341-321X(20)30191-4. doi: 10.1016/j.jiac.2020.06.005.

3. Izumiya H, Lee K, Ishijima N, Iyoda S, Ohnishi M. Multiple-locus variable-number tandem repeat analysis for non-O157 Shiga toxin-producing *Escherichia coli*: focus on serogroups O103, O121, O145, O165, and O91. *Jpn J Infect Dis*. 2020 Nov 24;73(6):481-490. doi: 10.7883/yoken.JJID.2020.095.

4. Nakamura K., Seto K., Lee K., Ooka T., Gotoh Y., Taniguchi I., Ogura Y., Mainil J. G., Piérard D., Harada T., Etoh Y., Ueda S., Hamasaki M., Isobe J., Kimata K., Narimatsu H., Yatsuyanagi J., Ohnishi M., Iyoda S., and Hayashi T. Global population structure, genomic diversity and carbohydrate fermentation characteristics of clonal complex 119 (CC119), an understudied Shiga toxin-producing *E. coli* (STEC) lineage including O165:H25 and O172:H25. 2023. *Microbial genomics* 9.

5. 泉谷秀昌、李謙一、伊豫田淳、大西真：2020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42巻、96-97、2021年5月

6. 李謙一、伊豫田淳、泉谷秀昌、大西真：国内で分離された腸管出血性大腸

菌の全ゲノム配列データベース化と食中毒事例の解析. IASR、第42巻、97-98、2021年5月

7. 泉谷秀昌, 李 謙一, 伊豫田 淳, 大西 真. 2021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43巻、108-109、2022年5月

## 2) 学会発表

1. 李 謙一, 綿引正則, 磯部順子, 大西 真, 伊豫田 淳, 大規模集団感染由来O104:H4と同一のStx2aファージを有する志賀毒素産生性腸管凝集性大腸菌 (Stx-EAEC) O86の解析, 第163回日本獣医学会学術集会、2020年9月14-30日、Web開催.

2. 李 謙一、井口 純、宇田和宏、松村 壮史、宮入烈、石倉健司、大西 真、伊豫田淳、EHEC Working Group in Japan. 小児重症例から分離された腸管出血性大腸菌新規血清群OX18および関連株のゲノム解析 第94回日本細菌学会総会、2021年3月23-25日、岡山（オンライン）.

3. 泉谷秀昌、李謙一、伊豫田淳、大西真、明田幸宏：腸管出血性大腸菌のMLVAによる分子疫学解析（2018-2020年）。第42回日本食品微生物学会学術総会、2021年9月、オンライン。

4. 泉谷秀昌, 李 謙一, 伊豫田 淳, 大西 真, 明田幸宏. 腸管出血性大腸菌のMLVAによる分子疫学解析 (2018-2020年) . 第42回日本食品微生物学会学術総会. 2021年9月、オンライン.

5. 泉谷秀昌, 李 謙一, 伊豫田 淳, 明田幸宏. 腸管出血性大腸菌のMLVAによる分子疫学解析. 第43回日本食品微生物学会学術総会. 2021年9月、東京.

物学会学術総会. 2021年9月、東京.

6. 泉谷秀昌、李謙一、伊豫田淳、大西真、明田幸宏：腸管出血性大腸菌のMLVAによる分子疫学解析（2018-2020年）。第42回日本食品微生物学会学術総会、2021年9月、オンライン。

7. 泉谷秀昌：分子疫学解析の現状と問題点、課題など（総括）。令和3年度希少感染症診断技術研修会、2022年2月オンライン。

8. 泉谷秀昌、李謙一、伊豫田淳、大西真、明田幸宏：腸管出血性大腸菌のMLVAによる分子疫学解析（2021年）。第43回日本食品微生物学会学術総会、2022年9月、東京都。

## 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

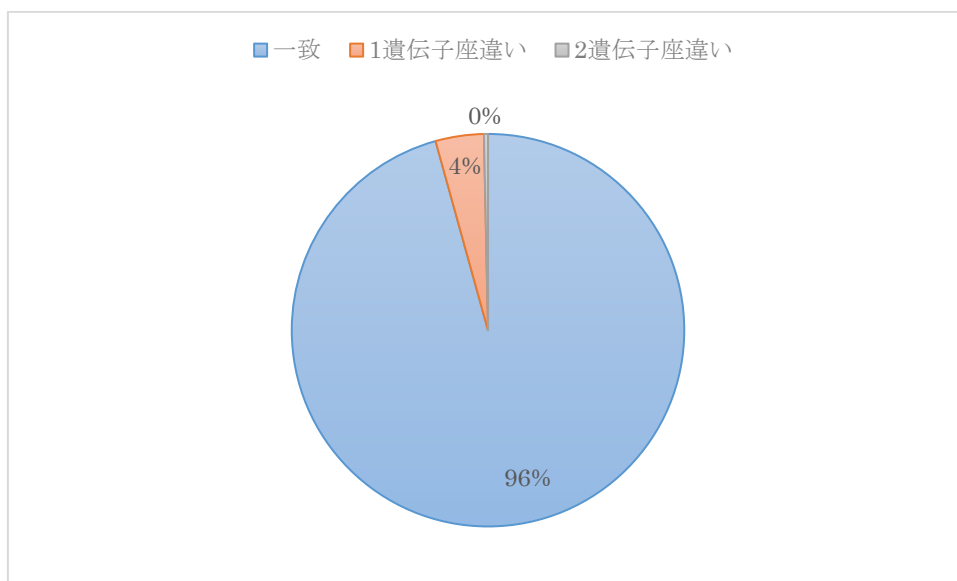


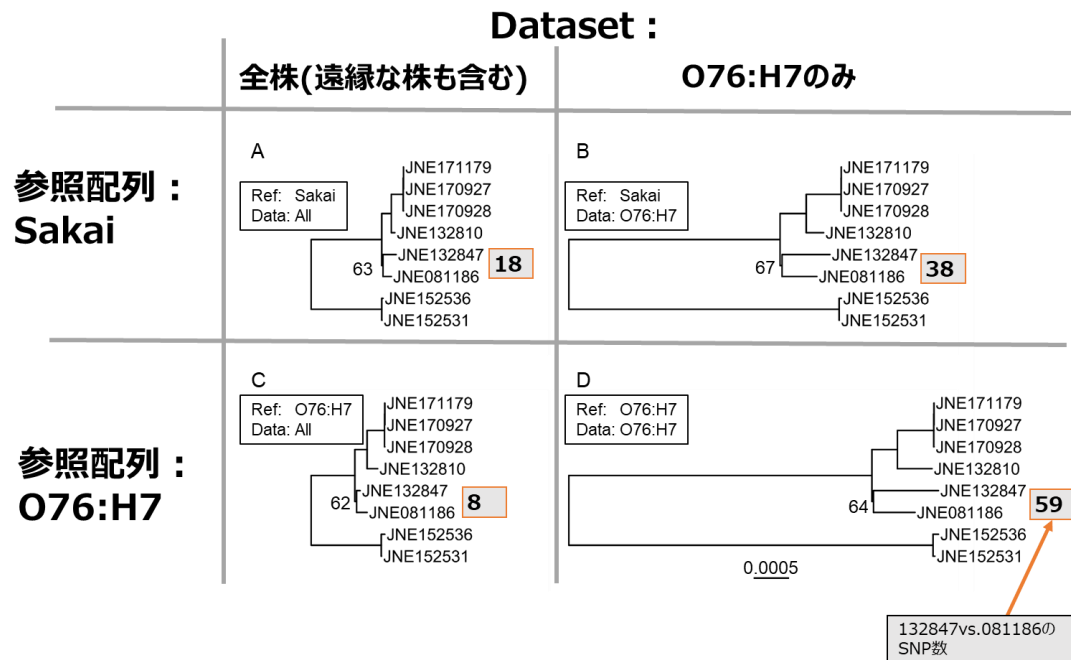
**表 X1. データベース化を行った菌株の血清型別菌株数**

血清型	菌株数
O111:H8	1,020
O26:H11	274
O103:H2	176
O121:H19	153
O146:H21	62
O103:H25	41
O145:H28	41
O91:H14	39
O165:H25	37
O5:H9	36
O115:H10	29
O111:HUT	27
O103:H11	23
OX18:H19	20
O123:H2	16
O177:H25	14
OUT:H8	13
O69:H11	11
O76:H19	10
その他*	206
計	2,248

\* O103:H8 等 82 種の血清型

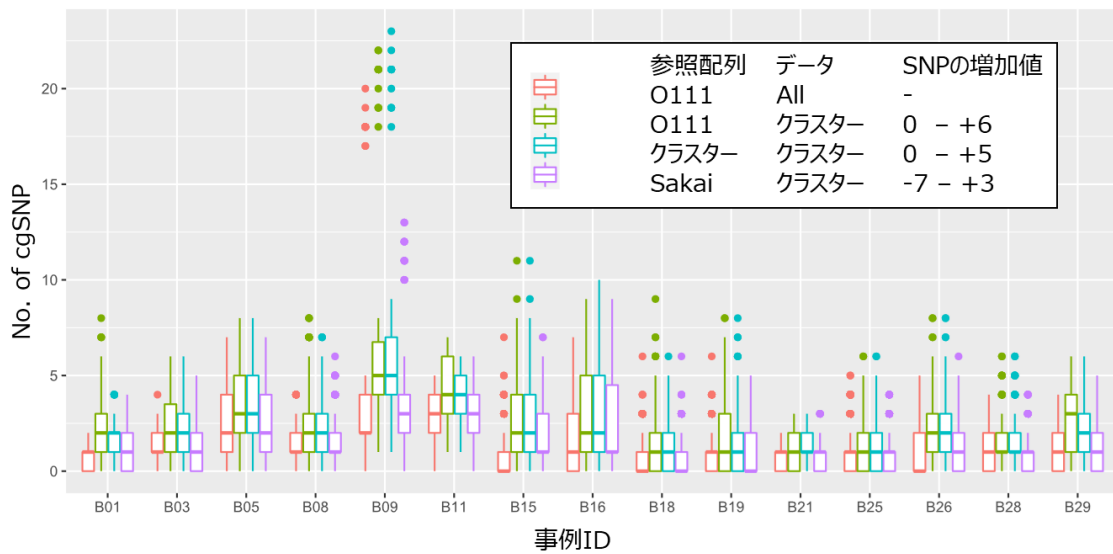
図 1. 送付 MLVA データと菌株データの結果の比較





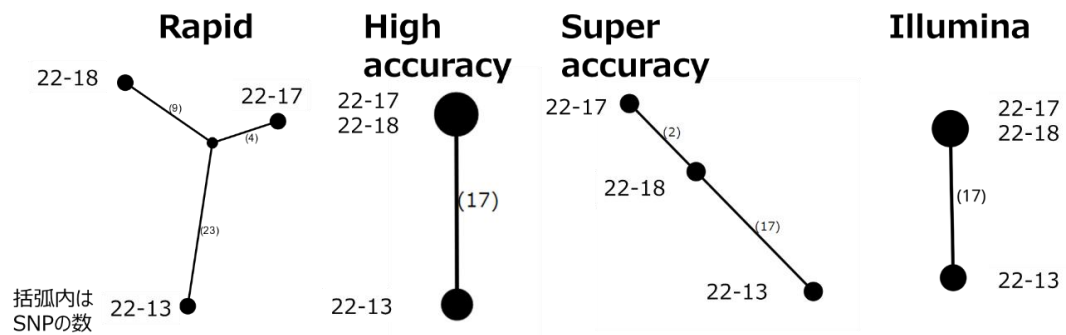
**図 X1. SNP 抽出条件ごとの O76:H7 の最尤法系統樹**

参照配列としては、O76:H7 代表株 (JNE132847 株) および O157:H7 Sakai 株の 2 種を用いた。解析データは、多様な大腸菌を含むデータセットおよび O76:H7 のみの 2 条件を用いた。四角内の数字は、型別能の例として JNE132847 および JNE081186 株間の SNP 数を示した。



**図 X2. O111:H8 における集団感染クラスター別の SNP 数を示した箱ひげ図**

参照配列としては、O111:H8 代表株（11128 株）、クラスター内の株、O157:H7 Sakai 株の 3 種を用いた。解析データは、O111:H8 全株（878 株）およびクラスター内の株のみの 2 条件を用いた。注釈中の SNP の増加値とは、参照配列として O111:H8 代表株を用いて全株で解析を行った際のデータと比べた時の、SNP 数の差異を示したものである。



**図 X3. ベースコールモデルおよびシーケンサー別にみた EPEC O153 における SNP による median joining tree**

各ノード（丸印）は菌株を示す。カッコ内は SNP の数を示す。Rapid、High accuracy、および Super accuracy は Oxford Nanopore シーケンサーデータのベースコールモデルを示す。Illumina は、Illumina HiSeqX によって得られたデータを示す。