

Ⅱ 分担研究報告書

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

地方衛生研究所の連携による食品由来病原微生物の網羅的ゲノム解析を
基盤とする新たな食品の安全確保対策に関する研究

研究代表者

調 恒明（山口県環境保健センター）

研究分担者

猿木 信裕（群馬県衛生環境研究所）

研究協力者

井上 伸子、佐々木 佳子、丹羽 祥一、塚越 博之、塩野 雅孝、黒澤 肇
（群馬県衛生環境研究所）

関塚 剛史、黒田 誠（国立感染症研究所）

研究要旨

食品汚染による大規模アウトブレイクを早期に探知し、収束させることは「食の安全」に不可欠である。最近の遺伝子解析による病原体を特定する技術は急速に発展しており、特に、次世代シーケンサー（NGS）による塩基配列の網羅的解析は、パルスフィールドゲル電気泳動（PFGE）と比較して、大容量の情報を得ることが可能である。本研究では、NGS を活用し、広域的に活用できるデータベース構築のために、カンピロバクター属菌の網羅的遺伝子解析を行い、その有用性について評価を行った。一方、PFGE パターン解析による分類は、泳動像の識別が困難な場合や制限酵素による影響を受けることから、疫学解析に限界があると考えられた。今後は、様々な事例による株を加えることによって、広域的に活用できるデータベースを構築することが重要であると考えられた。また、NGS により得られたデータを集積することによって、質の高いデータベースの構築が可能であることが示唆された。

A. 研究目的

食中毒が発生した場合、発生原因に関する情報を迅速かつ正確に把握することが必要である。

これまで食品由来病原微生物の特定は、培養同定法及び生化学的性状によって行われてきた。また、食中毒菌の分子疫学解析には、パルスフィールドゲル電気泳動（PFGE）法が行われてきたが、病原体ゲノム解析を活用することにより、発生源の特定や被害拡大の抑制につながると期待されている。

最近の遺伝子解析により病原体を特定する

技術は急速に発展しており、特に、次世代シーケンサー（NGS）による塩基配列の網羅的解析は、PFGE に比較して大容量の情報を得ることが可能である。

本研究では、昨年度に引き続き食中毒由来病原微生物として、カンピロバクターを対象とした NGS 解析の有用性について、PFGE との比較・検討を行った。

B. 研究方法

対象は *Campylobacter jejuni* とし、菌株は当研究所において平成 20 年～27 年に発生した食中

毒事件において採取、保存されていた 16 株および平成 27 年に市販鶏肉から採取された 2 株を用いた。

PFGE は、菌株を制限酵素 *Sma* I、*Kpn* I で処理した後、遺伝子多型解析を実施した。

また、NGS では、菌体を増菌培養後、遠心により集菌し Gentra Puregene Yeast/Bact. Kit[®] (QIAGEN) によりゲノム DNA を抽出し、Qubit (Invitrogen) を用いて定量した。次に Nextera XT DNA sample prep kit[®] (Illumina) によりライブラリーを作成した。得られたライブラリーは、電気泳動し、目的とする DNA を切り出し、精製した後 DNA 量が確保できた 3 検体 (表 1) に対し、Miseq Reagent Kit v3[®] (600cycle) (Illumina) を使用して塩基配列の読み取りを行った。得られた塩基配列は、国立感染症研究所病原体ゲノム解析研究センターに解析を依頼した。

C. 研究結果

得られた遺伝子をアセンブリし、Clustal W2 を用いた NJ 法による系統樹解析を行った結果、3 株の食中毒由来 *Campylobacter jejuni* は異なるクラスターに属していた (図 1)。また、昨年度実施した牛由来株 *Campylobacter jejuni* (平成 25 年採取 19 株) との間での一致みられなかったが、平成 23 年に採取された食中毒由来株 1 株と牛由来株 *Campylobacter jejuni* 3 株との塩基配列が近似であった (図 1 緑色枠内)。これらは、PFGE による解析においても、類似した泳動像が得られていた (図 2)。

D. 考察

系統解析の結果 (図 1) から、全体的には菌株の由来ごとにクラスターを形成する傾向があった。一方、食中毒由来 1 株と牛由来 3 株に類似性がみられたが、この食中毒事件は鶏肉が原因であると報告され、牛肉・牛内臓等の喫食は、確認できなかったが、無関係な菌株においても類似性を有する例もあると思われた。

また、PFGE による解析においては、泳動像

の歪みや不明瞭な部分があり、一部の菌株においては、制限酵素 *Sma* I あるいは *Kpn* I 処理後にバンドが検出できない例もあり、解析が困難であった。NGS にて塩基配列が近似であった 4 株では、PFGE 泳動像においても、同分子量と推察されるバンドに「ずれ」が生じており、それが解析結果にも影響を与えていた (図 2)。これらのことから、PFGE による解析には限界があると思われた。

しかしながら NGS においても、今回の検査で増幅・精製後に塩基配列の読み取りが可能であった検体は、18 検体中 3 検体のみで、さらなる精製方法の技術向上が必要と思われた。

今回、NGS にて牛由来株と食中毒由来株を比較したが、さらに多様な株を加えることによって、広域的に活用できるデータベースを構築することが重要と考えられる。

E. 結論

本研究において、NGS を活用しカンピロバクターの詳細な解析を行うことにより、詳細なデータを得ることであった。これらのゲノムデータを蓄積し、その情報を共有することで、広域食中毒の早期探知のために役立てていく必要がある。

したがって、今後も多くの食中毒菌について、データベース構築のための解析を行っていく必要がある。

F. 参考文献

1. Gardy JL, Johnston JC, Ho Sui SJ, Cook VI, Shah L, *et al.* (2011) Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 364: 730–739.

G. 研究発表

なし。

H. 知的財産権の出願・登録状況

なし。

I. その他

謝辞

本研究の遂行にあたり、知識や技術的な御協

力を頂きました国立感染症研究所病原体ゲノム解析研究センターの皆様にご心より感謝申し上げます。

表 1. 使用菌株

No.	採取年月		由来
Cj-Gunma-Human-3	H21.6	食中毒	焼き鳥
Cj-Gunma-Human-4	H21.9	食中毒	牛生レバー、焼き鳥
Cj-Gunma-Human-16	H23.9	食中毒	鶏料理

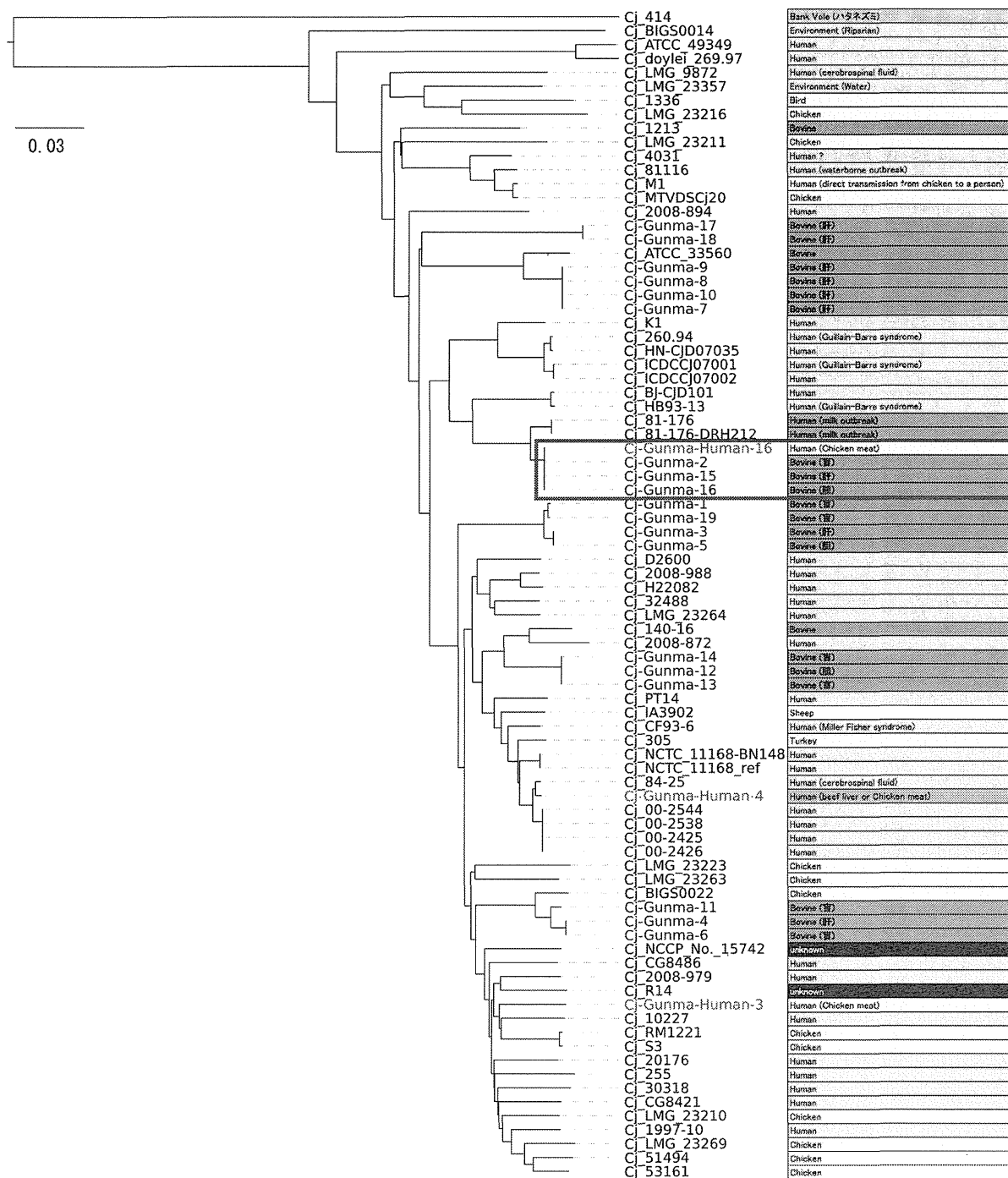
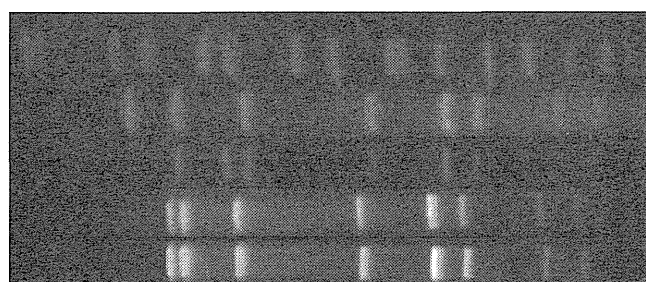


図1. NGS による解析結果



マーカー

Cj-Gunma-Human-16

Cj-Gunma-2

Cj-Gunma-15

Cj-Gunma-16

図2. 近似4株の PFGE 泳動像

地方衛生研究所の連携による食品由来病原微生物の網羅的ゲノム解析を基盤とする
新たな食品の安全確保対策に関する研究
－食品分離株及び臨床分離株のゲノム解析－

研究分担者 佐多徹太郎（富山県衛生研究所）
研究協力者 綿引 正則（富山県衛生研究所）
研究協力者 磯部 順子（富山県衛生研究所）
研究表力者 範本 志保（富山県衛生研究所）
研究協力者 木全 恵子（富山県衛生研究所）
研究協力者 三井 千恵子（富山県衛生研究所）
研究協力者 金谷 潤一（富山県衛生研究所）

研究要旨 大量の塩基配列データが得られる次世代シーケンサー（NGS）の公衆衛生分野、特に地方衛生研究所（地衛研）での活用方法について、検討してきた。最終年度は、食中毒原因菌として重要である、*Salmonella Infantis* 77 株のゲノム配列を取得し、比較ゲノム解析手法を用いて、遺伝的背景と保有遺伝子の状況の相関を調査した。しかし、遺伝的背景と保有遺伝子の関係は明らかにならなかった。また、食中毒事例で分離された *Campylobacter lari* のゲノム解析を実施し、病原性に関連する遺伝子の探索を行い、53 遺伝子が検出された。このように NGS の有用性は明らかになってきたが、得られた配列は、ドラフト配列であり、その活用法について注意が必要である。このドラフト配列を定量的に理解するために、既に完全配列が明らかになっている株を解析して、ドラフト配列の定量的な解析を試みた。その結果、コーディング領域の約 85% が一致した。以上の解析は、普段使い慣れた Windows 環境の利用を念頭に入れて実施した。このことは高い専門性が必要とされていたゲノム解析手法が、地衛研でも十分活用できるということを示しており、さらに公衆衛生学分野での活用に応用できると思われる。

A. 研究目的

近年の食品流通システムの変化により、先進国では、病原微生物に汚染された流通食品が原因となる大規模かつ広域化した集団食中毒の発生がしばしば報告されている。例えば腸管出血性大腸菌（EHEC）の食中毒の発生は重大な健康危機をもたらしている。その中で、地方衛生研究所（地衛研）は、法律にもとづき、集団発生時に食品あるいは臨床検体から病原体検索を行い、分離された病原体の解析から、原因究明に資する科学的根拠を提供する業務を行っている。しかし、原因となる病原体が検出できず、原因不明となる事例も少なくない。一方で、一度に大量の塩基配列データを簡単に得ることを可能となる次世代シーケンサー（NGS）が利用できるようになり、原因がこれまで不明であった事例でも原因究明のための検出技術として非常に有用であることが知られるようになってきた。

2011 年に NGS のデスクトップ型機器が発売され、また、この NGS で出力される膨大な塩基配列データを処理する Windows ベースのソフトウェアが充実し、NGS が利用しやすくなっていた。このような時期に、富山県を中心として、多数の重症

患者を含む腸管出血性大腸菌による集団食中毒事例が発生し、検査機能の充実を図るという目的で、地衛研として初めて富山県衛生研究所に NGS の一つであるイルミナ社の MiSeq が設置された。近年、感染症や食中毒もグローバル化し、従来の原因究明のための検査技術も多様化している。従って、地衛研の業務も世界的な視野に立って、遂行することが求められるようになり、新規な病原体や多様化する感染症、食中毒が発生したとき、迅速な原因究明に NGS の活用が期待される場所である。

本研究は、公衆衛生分野での感染症及び食中毒対応、すなわち、予防や蔓延防止等に資する MiSeq の活用法を検討することを目的としている。

研究班最終の今年度は、昨年度より引き続き課題である、サルモネラ属菌のひとつである *S. Infantis* について、複数株を対象に SNP 抽出を行い、その有用性を評価した。NGS より得られたこれらのゲノム配列は、ドラフト配列であるが、分子疫学マーカーとして SNP の有用性と分離菌の病原性に関する解析を試みたので報告する。これにより、複数株の同時ゲノム解析が可能であり、公衆衛生分野で活用できることが確認された。

さらに、食中毒事例で分離された *Campylobacter*

lari の起因菌としての特徴を、比較ゲノム生物学的な視点から考察したので報告する。最後に MiSeq で得られる「ドラフト配列」について定量的に考察するため、配列既知の細菌を解析して、このドラフトゲノム配列について、考察を加えた。今後、地衛研がどのように活用していくのか、使いなれた Windows PC 環境でどこまでできるのか、検討したので合わせて報告する。

B. 研究方法

B-1. 供試菌株

S. Infantis 株は、富山県で 2000 年～2013 年にかけて分離された株で、うち 37 株は臨床分離株であり、40 株は、市販されていた鶏肉から分離された株である。

Campylobacter lari (*C. lari*) は、平成 20 年に発生した集団食中毒事例で、共通食材としての鯨肉と複数の患者便から分離された、いずれも PFGE パターンは一致していた。ここでは、鯨肉から分離された *C. lari* 分離株をゲノム解析対象株 (Toyama07) とした。

また、腸管出血性大腸菌 O157:H7 株は、既に完全なゲノム配列が報告されていた堺株 (O157 堺株、大阪大学微生物研究所から入手) を用いた。

B-2. 供試菌のゲノム DNA の抽出と解析

供試菌は、*S. Infantis* 株と O157 堺株については、トリプチケースソイ寒天培地 (TSA)、*C. lari* 株については、羊血液寒天培地上で一晩培養した新鮮コロニーから Genra Puregene Yeast/Bact. Kit (キアゲン) を用いて、製品添付プロトコールに従って、ゲノム DNA を調製した。調製した DNA は、Qubit® 2.0 Fluorometer (ライフテクノロジー社) を用いて、DNA 濃度を測定した。

B-3. DNA ライブラリーの調製と MiSeq による解析

MiSeq で解析する DNA ライブラリーは、Nextera XT DNA サンプル調製キット (イルミナ) を用いて、調製した (これを「DNA ライブラリー」と称する)。DNA 分画は、1% アガロース電気泳動を行い、泳動後、臭化エチジウムで染色し、LED ランプ照射下、アガロースを切り出し、Montage DNA Gel Extraction Kit (ミリポア) で、溶出後、溶出液を濃縮した。KAPA Library Quantification Kits を用いて定量後、MiSeq Reagent Kit, V3 600PE を用いて MiSeq から塩基配列リードを得た。

B-4. MiSeq データの情報処理

MiSeq からの出力データは、fastq 形式とし、CLC Genomics Workbench (CLC) を用いて、アセンブルを行なった。

アセンブルしたデータを RAST (<http://rast.nmpdr.org/>) を使用して、自動注釈 (アノテーション) を行い、遺伝子の機能単位である、蛋白質コーディング領域 (CDS) 等の情報を得た。また、比較ゲノム学手法である Blast atlas 解析は、Gview Server (<https://server.gview.ca/>) を利用した。病原性因子の探索は、Virulence Searcher (http://www.hpa-bioinfotools.org.uk/pese/virfactfind_small.html) を利用した。

blast 検索は、公的データベースを利用する場合は、DDBJ の Web サービスを、In-house データベースを利用する場合は、NCBI blast+ を Windows PC にインストールして実施した。

S. Infantis 77 株のゲノム配列から得られた CDS の比較には、RAST により得た CDS 情報をデータベースソフトである、File Maker Pro (ファイルメーカー) を用いて、すべての CDS をデータベース化し、ソフト上で比較した。

(倫理面への配慮)

本研究は、臨床情報のない分離菌株を用いるものであり、解析対象にはヒトゲノム配列情報等は含まれていないため、倫理上の問題は発生しない。

C. 研究結果

C-1. *Salmonella Infantis* のゲノム解析

C-1-1. *S. Infantis* 77 株の MiSeq による解析データの品質評価

S. Infantis 77 株を MiSeq で解析した。取得した配列は、CLC でアセンブルを実施した。得られた塩基配列の品質を評価するためにその概要を表 1 に示した。今回 MiSeq から得られた配列は、カバー率 (全リード塩基数/コンティグ長の塩基数)、最大 169、最小 33、平均 100 であった。アセンブル後のコンティグ数は、最大 567、最小 64、平均 138 となり、得られた情報としては、若干、バラツキがあるものの、これ以降の解析に耐えうる値であると判断した。

C-1-2. *S. Infantis* 77 株の SNP データによる系統解析

S. Infantis 77 株の MiSeq による解析リードを、完全ゲノム配列が明らかになっていた *S. Infantis* CM001274 のゲノム配列にマッピングし、SNP を

抽出し、77 株の SNP 情報を統合して、最終的に、タンパク質をコードする領域のみとして、104,586 塩基の SNP を得た。これらの配列から系統樹を作成した(図 1、2)。この結果、臨床分離株 17 株からなる大きなクラスター (C5) が検出された。また、臨床分離株 13 株と食品分離株 36 株の混在した大きなクラスターを形成した(IC-1)。

C-1-3. 比較ゲノム学的手法による MiSeq データの利用の試み—クラスター特異的遺伝子の探索

図 2 の系統樹で検出された、クラスター C5 は、食品分離株の混在がなく特徴的なグループであった。そこで今回は、複数株のゲノム解析の試みとして、臨床分離株に特異的な遺伝子(CDS)が存在するかどうかを検討した。

方法としては、解析するクラスター内の各 *S. Infantis* ゲノム配列について、アセンブル後、RAST により自動アノテーションを行った。そこで、得られたすべての Genbank ファイル (拡張子 =gbk) を用いて、クラスター毎に GViewServer の Pangenome analysis を実施し、そのクラスター全体で検出された CDS を統合した Pangenome 配列を得た。今回の解析では、クラスター C5 (17 株)、クラスター IC-1 (49 株)、及び IC-1 に含まれる食品由来株 (36 株) の Pangenome 配列を得た。これら 3 つの Pangenome 配列の gbk ファイルを作成し、GViewServer の Signature Analysis を選択し、クラスター C5 と IC-1 にともに存在し、IC-1 の食品由来株に存在しない、臨床分離株に特異的な CDS の検出を試みた。その結果、56 の候補遺伝子を得た。この 56 遺伝子の内訳は、機能が付与された 22 遺伝子と機能未知遺伝子(hypothetical protein)、34 遺伝子であった(表 2)。

この 56 遺伝子が、臨床分離株特異的な遺伝子かどうかを検証した。方法は、今回、配列解析した *S. Infantis* 77 株のアセンブル後の配列を、RAST で自動アノテーションを行い、得られた CDS の塩基配列の fasta ファイルのデータをすべて統合した、*S. Infantis* 77 株の CDS データベースを作成した。次に、この CDS データベースに対して、先の 56 遺伝子をクエリー配列として、blast 検索を実施した。その結果を図 3 に示した。この図は、系統樹に示された株について、56 遺伝子の有無を示したマトリックスとして示した。このマトリックスの結果からは、臨床分離株特異的な遺伝子の特定には至らなかった。

C-2. *Campylobacter lari* 分離株のゲノム解析

C-2-1. *C. lari* 分離株の MiSeq データの取得

本食中毒事例は、平成 20 年に富山県で発生した

もので、2 つの飲食店で提供された料理に含まれていた鯨肉が共通食であり、患者数 61 名を数えた事例であり、鯨肉残品と患者便から分離された *C. lari* は、PFGE で一致した。そこで、今回、鯨肉から分離された *C. lari* 株 (Toyama 07) のゲノム解析を実施した。

MiSeq からの解析リードのアセンブルした後の統計数字は、表 3 に示した。今回の解析では、コンティグ数 38、カバー率約 400、ゲノムサイズ 1,617,445 塩基のドラフト配列を得た。

C-2-2. *C. lari* Toyama07 の他のカンピロバクター属とのゲノム配列の比較

カンピロバクター属のゲノム解析は既に複数の株で実施されており、入手可能であった。*C. lari* については、RM2100 株、LMG11760 株の配列が既に報告されている。また、*C. jejuni*、*C. coli* 及び *C. fetus* の配列が報告されており、各 2 株ずつ選び、今回、解析した *C. lari* Toyama07 株の配列と比較した。

ゲノム配列の比較は、*C. lari* RM2100 株の配列を参照配列として、GViewServer の Blast atlas(図 4) 及び Pangenome Analysis (Pangenomic blast atlas) を行った(図 5)。

Blast atlas の結果から、*C. lari* Toyama07 株は、報告されている *C. lari* のゲノム配列と類似していた。また、参照株と異なるゲノム配列上の GAP が検出されたが、この GAP の遺伝子は、細胞表面の LPS の多型及びファージ遺伝子の有無に関係していた。また、Pangenome Analysis の結果から、*C. jejuni* と *C. coli* は、ゲノム配列の 2/3 はほぼ共通なゲノム配列であるが、*C. lari* と *C. fetus* のゲノム配列は互いに相同性が低いことがわかる。また、*C. lari* Toyama07 株に特異的な配列があることが判明した。

C-2-3. 病原性遺伝子の検索

病原性遺伝子の検索は、Virulence Searcher を用いた。Web から、得られた *C. lari* Toyama07 のドラフト配列をアップロードし、解析してところ、53 遺伝子(図 7) が病原性遺伝子として検出された。*Campylobacter* 属菌でもっとも解析が進んでいるのは *C. jejuni* である。主要な病原因子である定着因子、CadF, JipA, MOMP, PEBA, 毒素 CdtABC、侵入因子 CiaB について、今回の *C. lari* Toyama07 株では、CadF, MOMP, CdtABC および CiaB がゲノムデータから検出された(表 4)。

C-3. 既知ゲノム解析株の MiSeq から得られたゲノム配列と報告配列との比較

O157 堺株のゲノム配列を MiSeq で解析した。

MiSeq の 2 回のランにより得られた解析リードを個別にアッセンブルしたコンティグ SK01, SK02, 及びこの 2 つのランデータを統合したコンティグ SK12 を得た。これらのアッセンブル後の統計データを表 5 に示した。この結果、SK01, SK02, および SK12 のシーケンスのカバー率は、それぞれ 43.5、50.9、及び 94.1 であり、MiSeq より得られた解析リードの品質を評価することが出来るこれらの統計データから、ドラフト配列の定量的な解析に耐えうる品質であると判断した。そこで、これらのコンティグを用いて、蛋白コーディング領域 (CDS) を抽出し、既知配列と比較した。

O157 堺株については、既にその完全なゲノム配列が明らかにされており、詳細な解析結果が報告されている (Hayashi, T., et al. *DNA Res.*, 8:11, 2001)。報告では、O157 堺株は、2 つのプラスミドを保有しているため、今回解析した配列中には、プラスミド由来の配列が当然含まれることになる。そこで、今回決定した SK01, SK02, および SK12 は、それぞれコンティグ数は、203, 181, および 182 であり (表 6)、そのうち、スキヤフォールド領域は (ペアエンドで解析していることから、配列が不明のまま、長さのみ情報としたリードが存在しているコンティグのこと)、それぞれ 17, 19, および 20 か所検出された (データ未提示)。先の論文では、大きさ 150 塩基対以上の CDS の数は、5,447 個として報告されているが、今回、MiSeq で得られた解析リードのアノテーションと比較できるように、O157 堺株の既知である完全ゲノム配列を改めて、RAST によるアノテーション情報を得たところ、プラスミドの CDS と合わせると、総計で、5,310 個の CDS が検出された (これを以後、「Sakai」と呼ぶ)。この値は、Hayashi らの報告数より若干少ない。この差は、RAST の CDS の抽出法の問題もあるが、アノテーションについては、最終的には、ひとつひとつの CDS を詳細に検討することが必要となる。しかし、今回、ほぼ、同じ数字であること、完全な配列は得ていないことから、この後は、RAST で得られた 5,310 の CDS をそのまま用いることとした。そして、MiSeq により得られた O157 堺株の配列である SK01, SK02 および SK12 も同じく RAST を用いて、アノテーションを行い、比較した。

得られた CDS 情報は、File Maker 上で比較した。実際には、Sakai と SK01, SK02 および SK12 それぞれ比較し、共通に存在する CDS、あるいはどちらか一方にのみ存在する CDS を数え、その結果を図 7 に示した。

解析リードのカバー率が約 50 と 100 で比較したところ、コンティグ数、Web サービスによる自動アノテーションの結果から、カバレッジ 50 程度でも充分、ドラフト配列として、活用できることが

示された。また、報告された配列と MiSeq から得られた配列から、CDS を比較したところ、83%~87%であった。

D. 考察

NGS から出力される解析配列は、大量であり、従って、ゲノム解析は、情報処理を適切に行うことが要求される。この大量のゲノム情報の解析技術のほとんどが、我々が日常的に利用している WindowsPC 環境ではないので、高い専門性が要求されることになり、多くは情報処理の専門家が担当していることが多い。これは地衛研で、NGS 技術を検査や感染症、食中毒の原因究明のひとつの方法として活用する場合、大きな壁となる。しかし、最近、Windows 環境も新しい OS の出現や、これまで Unix, Linux や MacOS の利用を前提としていたアプリケーションに、Windows 環境でも利用できるソフトが少しずつ増えているのも事実である。従って、最終年である今年度は、情報処理の専門家でない、研究者が Windows 環境でのゲノム解析がどこまで出来るか検討することも目的とした。

S.*Infantis* 77 株の解析については、臨床分離株と食品分離株を含めて、SNP を抽出し、SNP による系統関係を明らかにした。その結果、臨床分離株 17 株の特徴的なクラスターが検出された。この原因としては、我々が選択した食品由来株は、市販の鶏肉がほとんどであり、分離元としては偏っており、今回、検出された臨床分離株由来のクラスターの感染源は、鶏肉ではない別な感染源の可能性が高いと思われた。

さらに、S.*Infantis* の人に対する病原性に関する情報を今回の 77 株を用いて、抽出できるかどうかについて検討した。臨床分離株と食品分離株のゲノム配列上の遺伝子を比較 (GViewSever) し、臨床株に特異的な候補遺伝子として、56 遺伝子を選択した (表 2)。今回、この 56 遺伝子の評価については、S.*Infantis* 77 株から得られた全 CDS をデータベース化し、その保有状況を調べることで実施した。もし、臨床分離株特異的な遺伝子であれば、その保有分布については、臨床分離株にのみ検出されるはずである。しかし、今回、そのような結果にはならなかった (図 3)。今後、改めて、抽出方法等を検討し、評価することが必要である。

食中毒事例で分離された *C.lari* のゲノム解析に関しては、病原遺伝子を検出する Web サービスを利用することが出来たので、ある程度の品質の配列リードが得られれば、簡単に病原遺伝子のスクリーニングは可能となる。そして、その細菌の病原性遺伝子が検出されれば、起因菌判断が難しいときに、起因菌とするかどうかの判断に役に立つ

情報になると思われる。今回の結果では、*C.lari* のこれまでの報告と矛盾しない結果であった。しかし、一部で *C.jejuni* で保有している因子が *C.lari* で検出できなかったことから、このことが、所謂「ドラフト配列」を用いた結果、見逃しているのか、あるいは、*C.lari* による病態が比較的 *C.jejuni* よりも軽症傾向にあることと関連があるのか、興味もたれる。

今年度の *S.Infantis*77 株のゲノム解析と食中毒事例で分離された *C.lari* のゲノム解析は、MiSeq で取得したリードのみで実施したもので、完全な配列でない

通常、完全なゲノム配列を得るためには、ドラフト配列を元にして、「finishing」という工程が必要であり、通常は、膨大なコストと時間が必要となる。その意味で、地衛研での finishing 作業は、事実上、不可能であり、ドラフト配列を利用した活用方法が主流となると思われる。その意味で、SNP による系統解析は十分利用価値が高いことは、既にこれまでの検討から明らかになっている。しかし、病原性に関する検討には、完全ゲノム配列が理想ではあるが、「ドラフト配列」でもある程度は可能であることは、先の *C.lari* のゲノム解析でも示唆されている。従って、この「ドラフト配列」がどの程度のものであるか、定量的に理解することを試みた。その結果は、図 7 に示した。既に完全なゲノム配列が報告されている腸管出血性大腸菌 O157 堺株を MiSeq で解析し、報告配列、MiSeq で取得した配列を比較することで「ドラフト配列」の評価を試みた。その結果、この数字をどう考えるかについては、Hayashi らの論文の報告と較べてみると、O157 堺株の全 CDS は 5,447 であり、そのうち、大腸菌 K-12 のコアゲノムとしては、3,729 個であり、全体の 70% を占めており、残りは、株特異的な遺伝子となり、病原性に関与する遺伝子が含まれると推定される。このようなことから、83%~87% という数字が示す、一致度は、MiSeq の解析結果としては順当なところであろうと推定された。

E. 結論

NGS の一つである MiSeq を用いた解析は、細菌のドラフトゲノムデータであっても、高精度分子疫学的マーカーとしての SNP の利用価値が高いこと、さらに完全な配列が明らかなゲノム株を reference 株として全コード領域を対象とした Blast 解析を実施すること (Blast Atlas 解析) により、病原性や株の特徴などゲノムワイドな視点から知ることが可能であり、公衆衛生分野で利用価値は高いと思われた。

また、Windows PC 環境においても、インターネ

ット環境の整備と、NGS 解析が出来るソフトウェアの導入、Windows 環境で動作するソフトウェアの導入、Web サービスの利用で、専門性の高い Linux 等を用いなくとも、充分解析が可能であることが示された。

さらに、公的データベースで実施している NGS データの解析支援プログラムを利用することにより、予算的に厳しくても、インターネット環境があれば、Windows 環境でも安価に NGS 解析が出来る可能性がある。このような状況から、地衛研において、MiSeq 等の NGS を所持に関係なく、膨大な NGS データを利用できる環境が整いつつあるなかで、NGS の活用について、検討することは、それほど高いハードルではないと考えられる。

F. 健康危機情報

なし

G. 研究発表

1. 論文発表

なし

2. 学会発表

- ① 鯨肉の刺身を原因とする食中毒事例から分離されたカンピロバクター・ラリのゲノム解析と病原因子の検出
磯部順子、金谷潤一、木全恵子、三井千恵子、範本志保、佐多徹太郎、綿引正則
第 49 回腸炎ビブリオシンポジウム (東京都渋谷区、平成 27 年 10 月 16 日)

H. 知的財産権の出願・登録状況

なし

I. 参考資料

以下、巻末に参考資料として添付した。

1. Windows PC 環境における有用な NGS データ解析用ソフト、アプリケーションのリスト

ここで示したソフト等は、一例を示しただけであり、他にも多くのソフトがある。予算や解析目的に沿ったソフトを導入し、評価しながら使うことが重要である。

2. Windows PC で活用することが有用と思われるソフトの導入例 (チュートリアル)

参考まで、ゲノム解析のために使用頻度が高い、あるいは、汎用性の高いソフトを導入のためのチュートリアルを参考資料として紹介する。ソフトの正式な使い方は、それぞれ入手して確認していただきたいが、ここでは、Windows PC

環境の導入のために必要なポイントをまとめ、失敗しないで導入するための参考としていただきたい。

- GView Server の使い方
- NCBI Blast+の Windows 環境での利用について

3. MiSeq を使用する場合のコスト (試算結果例)

MiSeq を使って、ゲノム解析する場合、試薬が高いので、失敗をしないようにすることが重要である。我々は、複数の細菌のドラフト配列を得ているが、最適なクラスターを形成するために、DNA

ライブラリーの定量が非常に重要なステップとなる。複数の株を同時に解析するとどうしても、カバー率のばらつきが出るので、たくさんの株を解析する場合には、同じ DNA ライブラリーを 2 回ランすることとしている。1 回目のランで、クラスター形成のばらつきを判定し、2 回目のランでは、そのばらつきを解消するように DNA ライブラリーのインプット量を調整する。この方法は失敗が少なくなるので、推奨したい。ここではこの方法を用いたときの消耗品、試薬のコスト計算をしたので、紹介する。

表 1. *S. Infantis* 77 株のゲノム配列データのアセンブル後の概要

Salmonella Infantis 77 株

Miseq Reagent Kit v3 (600 cycle)

		平均値	最大値	最小値
No. of reads		2,047,210	3,835,990	603,348
Total base		490,074,777	825,754,204	158,906,599
Coverage		100	169	33
Contig	Total base	4,891,271	5,158,601	4,588,724
	Count	138	567	64
	Maximum	1,051,615	1,241,747	480,288
	N50	290,813	521,185	159,289

表 2. 比較ゲノム学的手法を用いて、臨床分離 *S. Infantis* 株に特異的に存在することが期待された 56 の候補遺伝子

No.	Seq_ID	Hit Product
1	pj039-01	hypothetical protein
2	pj039-02	hypothetical protein
3	pj039-03	hypothetical protein
4	pj039-04	hypothetical protein
5	pj039-05	hypothetical protein
6	pj039-06	HTH-type transcriptional regulator hdfR
7	pj039-07	SanA protein
8	pj039-08	Cytidine deaminase
9	pj039-09	hypothetical protein
10	pj039-10	hypothetical protein
11	pj039-11	hypothetical protein
12	pj039-12	Ferredoxin
13	pj039-13	Mobile element protein
14	pj039-14	hypothetical protein
15	pj039-15	hypothetical protein
16	pj039-16	hypothetical protein
17	pj039-17	ClpB protein
18	pj039-18	hypothetical protein
19	pj039-19	hypothetical protein
20	pj039-20	Glutamate racemase
21	pj039-21	Glutamate racemase
22	pj039-22	Vitamin B12 receptor BtuB
23	pj039-23	hypothetical protein
24	pj039-24	hypothetical protein
25	pj039-25	5-deoxy-glucuronate isomerase
26	pj039-26	Methylmalonate-semialdehyde dehydrogenase
27	pj039-27	Cellulose synthase catalytic subunit
28	pj039-28	Cellulose synthase, putative
29	pj039-29	Putative cytoplasmic protein
30	pj039-30	hypothetical protein
31	pj039-31	hypothetical protein
32	pj039-32	Preprotein translocase subunit SecE
33	pj039-33	Transcription antitermination protein NusG
34	pj039-34	Cytochrome c heme lyase subunit CcmL / CcmH
35	pj039-35	hypothetical protein
36	pj039-36	hypothetical protein
37	pj039-37	hypothetical protein
38	pj039-38	hypothetical protein
39	pj039-39	Colicin V production protein
40	pj039-40	hypothetical protein
41	pj039-41	hypothetical protein
42	pj039-42	hypothetical protein
43	pj039-43	hypothetical protein
44	pj039-44	hypothetical protein
45	pj039-45	hypothetical protein
46	pj039-46	hypothetical protein
47	pj039-47	Polymyxin resistance protein AmC
48	pj039-48	hypothetical protein
49	pj039-49	hypothetical protein
50	pj039-50	hypothetical protein
51	pj039-51	putative cytoplasmic protein
52	pj039-52	Type III secretion protein SsaB
53	pj039-53	hypothetical protein
54	pj039-54	hypothetical protein
55	pj039-55	hypothetical protein
56	pj039-56	Mobile element protein

表3. 食中毒事例で分離された *C.lari* 株のゲノム配列データのアセンブル後の概要

	カウント数	平均長(bp)	全塩基数(bp)
リード数	2,311,897	277.25	641,005,075
Contigs数	38	42,564	1,617,445 (396 coverages)

表4. 食中毒事例で分離された *C.lari* 株のゲノム配列から推定した主要な病原因子の保有状況

	<i>C.lari</i> Toyama 07	<i>C.lari</i> RM2100	<i>C. jejuni</i> All strains
定着因子	CadF	+	+
	JipA	-	+
	MOMP	+	+
	PEBA	-	+
毒素	CdtABC	+	+
	CiaB	+	+

表5. 腸管出血性大腸菌 O157 堺株のゲノム解析結果の概要

SK01		Summary statistics	Count	average length	Total bases
Contig measurements (including scaffolded regions)		Reads	802,464	296	237,427,329
N75	52,302	Matched	788,226	296	233,224,139
N50	124,190	Not Matched	14,238	295	4,203,190
N25	235,262	Contigs	203	26,429	5,365,229
Minimum	500	Reads in pairs	660,940	487	
Maximum	314,054	broken paired reads	127,286	272	
average	26,430				Coverage=43.5
Count	203				
SK02		Summary statistics	Count	average length	Total bases
Contig measurements (including scaffolded regions)		Reads	941,258	296	278,710,760
N75	84,323	Matched	927,039	296	274,484,347
N50	146,270	Not Matched	14,219	297	4,226,413
N25	224,569	Contigs	183	29,450	5,389,464
Minimum	512	Reads in pairs	786,502	490	
Maximum	374,972	broken paired reads	140,537	271	
average	29,451				Coverage=50.9
Count	183				
SK12 (=SK01+SK02)		Summary statistics	Count	average length	Total bases
Contig measurements (including scaffolded regions)		Reads	1,743,722	296	516,138,089
N75	73,375	Matched	1,716,252	296	507,995,197
N50	148,380	Not Matched	27,470	296	8,142,892
N25	327,284	Contigs	182	29,670	5,399,993
Minimum	537	Reads in pairs	1,442,975	488	
Maximum	374,974	broken paired reads	273,278	272	
average	29,670				Coverage=94.1
Count	182				

表6 O157 堺株の配列の評価

	EHEC O157 Sakai from Public site				EHEC O157 Sakai by MiSeq		
	Chromosome	pO157	pOSAK1	Total	SK01	SK02	SK12
Length of sequence (bp)	5,498,450	92,721	3,306	5,594,477	5,365,229	5,389,464	5,399,993
Contigs	1	1	1	3	203	183	182
ORF (=CDS, >150)	5,361	83	3	5,447			
RNA	110	0	0	110			
<small>Hayashi, T., et al. <i>DNA Res.</i>, 8:11(2001)</small>							
after Re-annotation by RAST				after annotations by RAST			
ORF (=CDS, >150)	5,208	99	3	5,310	5,091	5,190	5,138
RNA	124	0	0	124	107	137	107

de novo assembly

リードをレファレンス配列(アノテーション情報あり)に
マッピング

(ない場合は RAST で取得)

<http://rast.nmpdr.org/>

SNP 検出

77 株の SNP 情報を統合

CDS の SNP のみ抽出

16S rRNA、prophage、pathogenicity island
にある SNP は除去

系統樹作成

CLC genomic workbench

エクセル

MEGA

図 1. *S. Infantis* 77 株からの Single Nucleotide Polymorphism (SNP) の抽出と系統樹作成工程

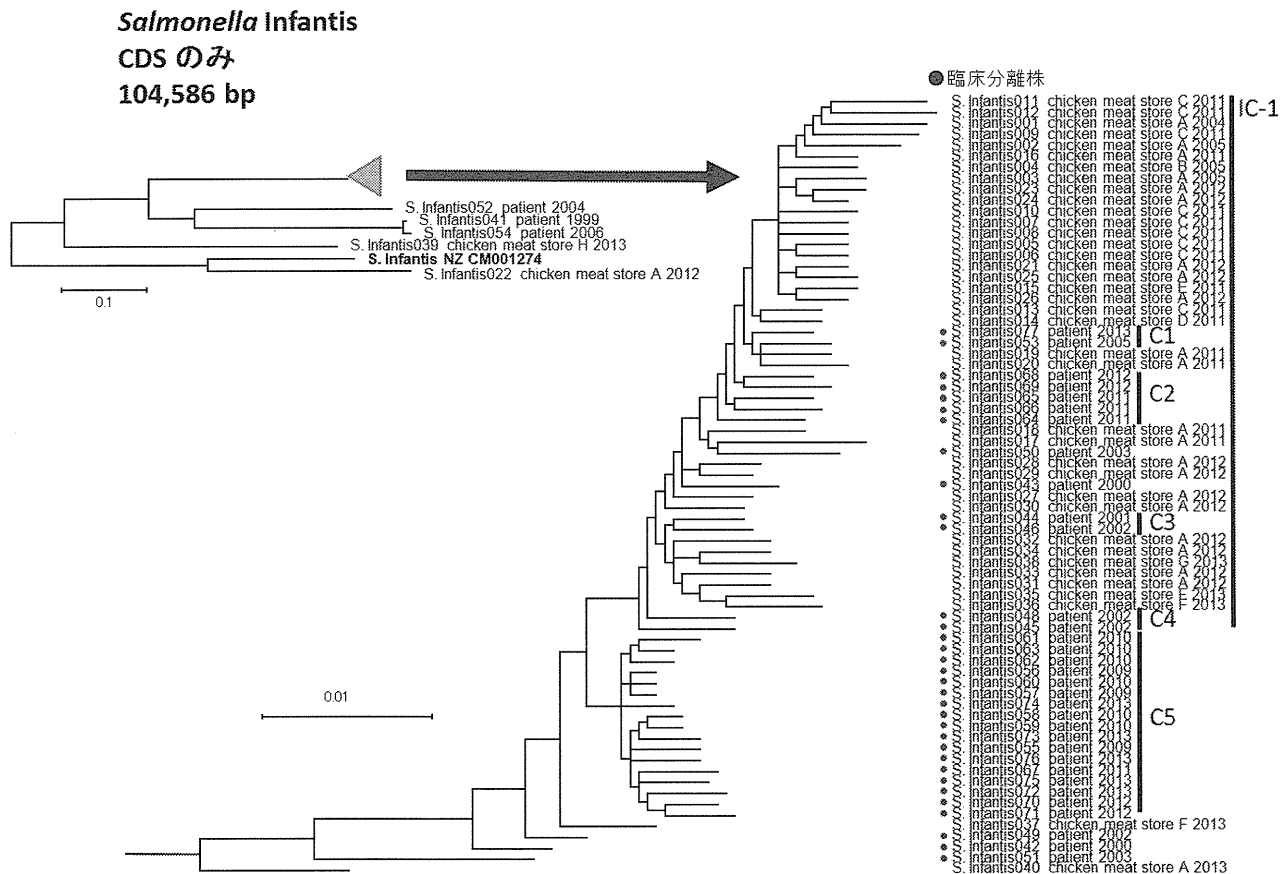


図 2. *S. Infantis* 77 株からの Single Nucleotide Polymorphism (SNP) による系統樹

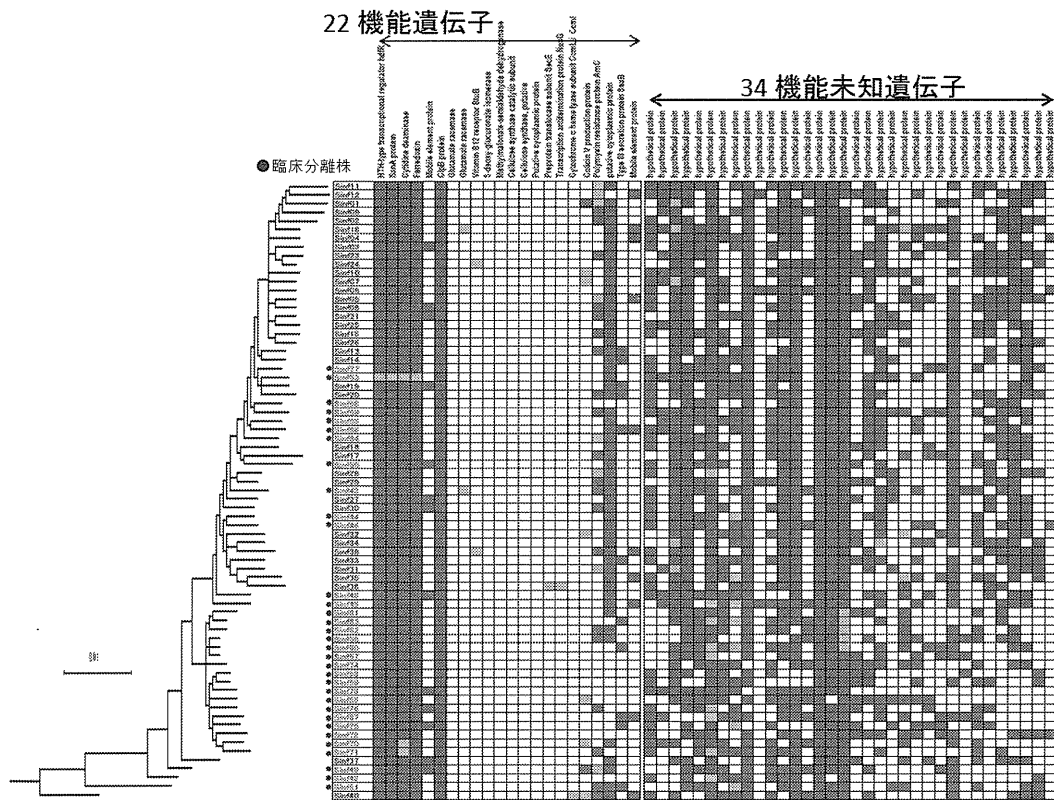


図 3. 臨床分離株特異的と推定された 56 遺伝子の確認

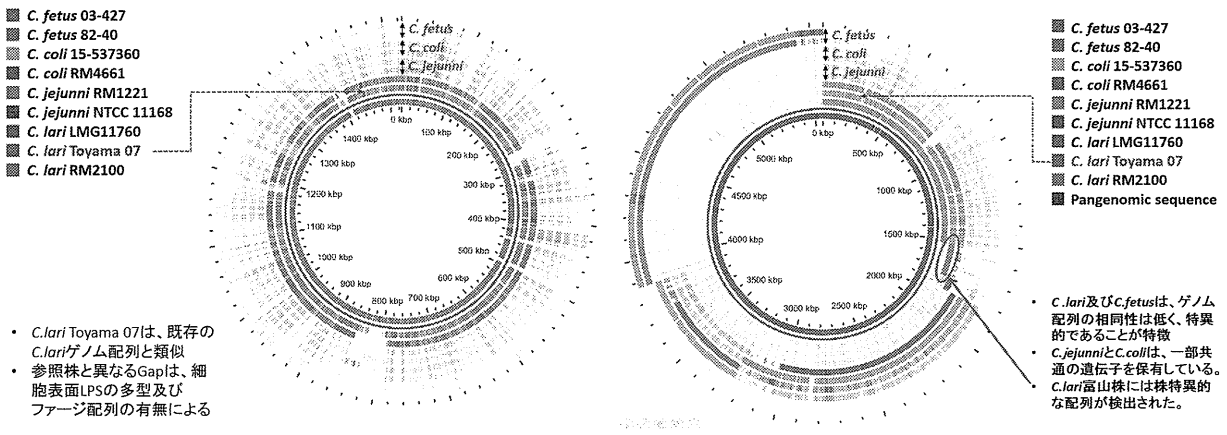


図 4. ゲノム解析した *C.lari* 分離株の Blast atlas 結果

図 5. ゲノム解析した *C.lari* 分離株の Pangenome blast atlas 結果

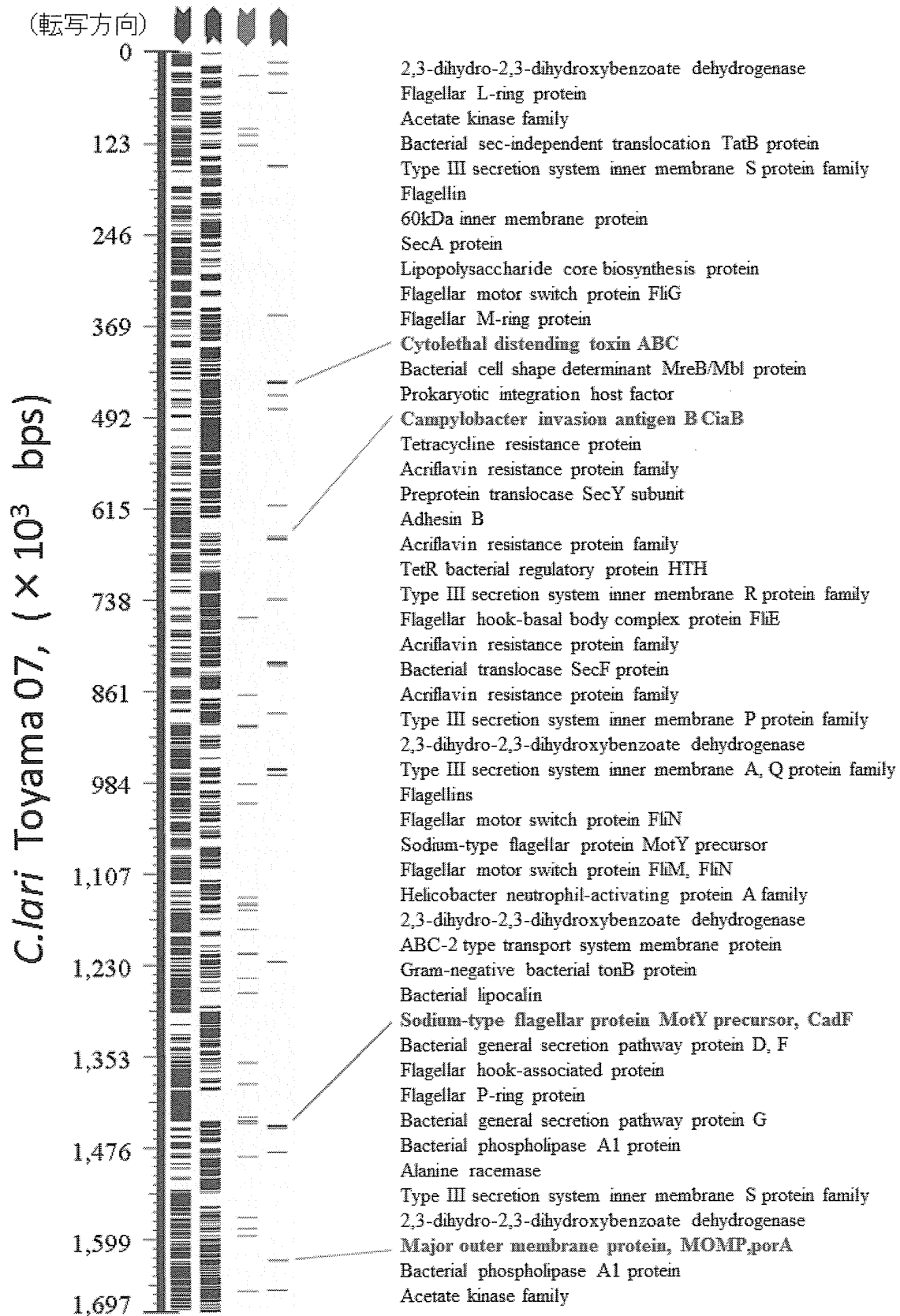
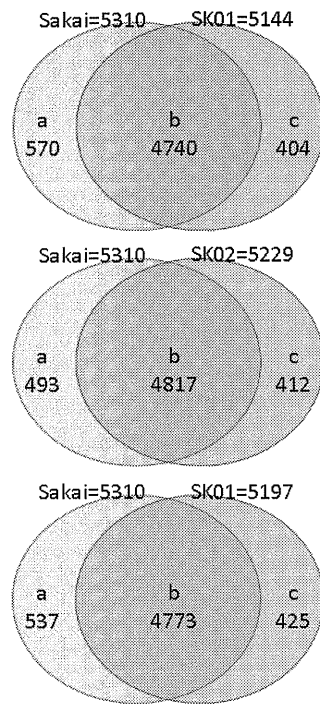


図6. *C.lari* Toyama07 の病原因子の検索結果

左図は、*C.lari* Toyama07 のゲノム上の遺伝子のマッピングしたもので、Virulence Searcher (Web サービス)を利用して検出された53遺伝子のゲノム上の位置

CDS, >150bp



a, Sakaiのみ検出
b, SakaiとSKに共通
c, SKのみ検出

Comparison of RAST Annotation Data between Public Sequence and MiSeq Draft Data

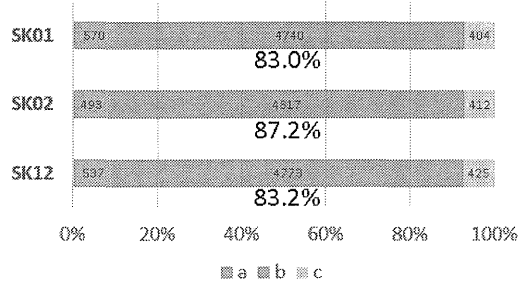


図7. 腸管出血性大腸菌 O157 堺株を用いた MiSeq データのドラフト配列の定量的な理解
ゲノム配列既知の O157 堺株を MiSeq でゲノム配列を得て、報告されている配列と比較して、
CDS の一致度を評価した。

I. 参考資料 1. WindowsPC環境における有用なNGS解析のためのソフト、アプリケーションのリスト

用途	ソフト名	説明	PC	利用形態	必要条件等	コスト	URL	コメント
NGS用総合解析	CLC genomic workbench	NGSデータを様々な解析が可能。De novo assemble、各種マッピング等可能。また、ゲノム解析にしばしば利用されるLinux系のソフトで利用できるファイル形式を読み込んだり、書き出しなど可能である。このソフトがあれば、ほぼ、なんとかなる。但し、高額である。	Windows	独立		有料	フィルジェン株式会社 http://www.filgen.jp/	公的データベースや、論文等でも引用されている。
NGS	DDBJ Read Annotation Pipeline	Web環境が必要だが、de novo assemble等が可能。	Windows	ウェブ		無料	https://p.ddbj.nig.ac.jp/pipeline/Login.do	登録必要/登録しているが、使用経験なし。
自動アノテーション	RAST	アップロードされたゲノム配列について、遺伝子データベースに対して、blast解析を実施し、蛋白質をコードする領域等を推定し、注釈するサービス。解析結果は、様々な様式で入手可能で、細菌のゲノム解析では有用な解析サービスである。	Windows	ウェブ	firefox	無料	http://rast.nmpdr.org/	登録必要/スタートゴドンの選択が時々、異なっており、注意が必要であるが、様々な様式のファイルがダウンロードできるので、非常に有用なサービス。
自動アノテーション	MiGAP	DDBJ japanが提供している細菌ゲノム配列の注釈サービス。RASTと同じような目的で利用されるWebサービスである。	Windows	ウェブ		無料	http://www.migap.org/	登録必要/使用頻度は高くありません。
比較ゲノム解析結果の比較	GviewServer	細菌ゲノムの比較解析を実施してくれるサービス。複数のゲノムを環状、線状形式で、ゲノム上のすべての遺伝子を対象として、視覚化できる。	Windows	ウェブ	firefox	無料	https://server.gview.ca/	登録必要/使用頻度は高い。
ゲノムの可視化	Gview	genebank形式のファイルを見やすい図として表示できる。環状、線状の図として表示し、CDSの色分け等、カスタマイズできる。また、Genbank形式やgffファイルを読み込ませ、より複雑な図として、カスタマイズできる。GVewServerで得られたblast atlasの結果を表示する場合も使われるソフトである。	Windows	独立	Java	無料	https://www.gview.ca/wiki/GView/Web_Home	使用頻度は高い。
手動アノテーション	Artemis	塩基配列の特徴を見やすく表示してくれる。また、遺伝子のアノテーションを行うことができる。EMBL及びGENBANK形式の配列データを読み込むことができる。	Windows	独立	Java	無料	http://www.sanger.ac.uk/science/tools/artemis	使用頻度は高い。
比較ゲノム解析	ACT	Artemisをベースにゲノムワイドな配列を比較し、その結果を表示する。その結果から、大きな領域での逆位、欠失、挿入などを視覚化してくれる。	Windows	独立	Java	無料	http://www.sanger.ac.uk/science/tools/artemis-comparison-tool-act	大きなGAPや逆位等が視覚化できる。
ABIシーケンサーデータのAssemble	Sequencher	ABIシーケンサーで解析した配列の確認、Assembleソフト	Windows	独立		有料	タカラバイオ	MiSeqデータではなく、ABIシーケンサーからのデータの解析のため、使用頻度は高い。

ABIシーケンサーデータのAssemble	GeneStudio	ABIシーケンサーで解析した配列の確認、Assembleソフト。波形を確認して編集可能。系統樹解析も可能。	Windows	独立		無料	http://genestudio.com/download	複数のPCで使用できるので利用価値はあるが、それほど使用頻度は高くない。
相同性解析	NCBI blast+	ペアワイズな塩基配列の局所的なアライメント／相同性検索(ホモロジー検索)を高速に行うプログラム	Windows	独立		無料	http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download	Windows10で使うと非常に使いやすい。使用頻度高い。2つのゲノムを比較するとき、ACTを使用するときに必要なcomparison fileを作成するblasstnを実施できる。 ゲノム解析した配列のアノテーション作業には必須の環境を提供する。
DNA配列のアライメント及び遺伝子の系統解析	MEGA	DNA配列あるいは蛋白のアミノ酸配列の並列化ソフト。配列の編集及び配列の系統分類。	Windows	独立		無料	http://megasoftware.net/	遺伝子の並列化、系統解析等、使用頻度は高い。
細菌ゲノムから病原性解析	The virulence factor database (VFDB)	病原細菌の病原性因子のデータベースが提供されている。	Windows	ウェブ		ウェブ	http://www.hpa-bioinfotools.org.uk/pise/	各種病原性細菌の病原性遺伝子が確認できる
細菌ゲノムから病原性解析	Virulence Searcher	NGSから得られたde novo assemblyデータから、病原性遺伝子の検索	Windows	ウェブ		ウェブ	http://www.hpa-bioinfotools.org.uk/pise/virfactfind_small.html	病原性遺伝子の検索

遺伝子解析ユーティリティ	Sequence Assistant	非常に軽い遺伝子解析ソフト。Genbankから、ダウンロードしたファイルから、塩基配列だけ抽出する、相補鎖変換等も可能。	Windows	独立		有料	http://www2s.biglobe.ne.jp/~haruta/ http://www.vector.co.jp/soft/ から入手可能	使用頻度は高い。塩基配列の文字列の抽出、結合、相補鎖変換等が簡単に行える。細菌ゲノムサイズの取扱いも可能。Artemisによる手動アノテーションと一緒に使うと便利。
遺伝子解析ユーティリティ	Format Assistant	遺伝子解析ユーティリティ。他の文字が含まれる文字列から塩基、アミノ酸を表す文字のみを抽出することが出来る。	Windows	独立		無料	http://www2s.biglobe.ne.jp/~haruta/ http://www.vector.co.jp/soft/ から入手可能	塩基配列、アミノ酸配列文字列のの編集に利用。あまり長鎖のDNA配列は利用できないようである(60,00塩基程度が限界?)
遺伝子解析ユーティリティ	秀丸	テキストエディタ	Windows	独立		有料	http://hide.maruo.co.jp/software/hidemaru.html	ゲノム解析に用いられるファイル形式は、基本的にテキストベースであり、テキストを直接、加工できるテキストエディタは便利である。特に本ソフトは、1億行まで扱えるため、ある程度の巨大ファイルの編集が
遺伝子解析ユーティリティ	7-zip	解凍ソフト	Windows	独立		無料	https://sevenzip.osdn.jp/ (日本語サイト)	巨大な圧縮ファイルの解凍が出来る。
遺伝子解析ユーティリティ	GENOMICS	ゲノム解析法の紹介	Windows	独立		無料	http://molbiol-tools.ca/Genomics.htm	ゲノム解析手法のリストが紹介されている。

GView Server の使用方法と注意事項

用途：微生物ゲノムを比較解析し、可視化する。環状形式と線状形式、遺伝子の転写の方向、等、カスタマイズ可能である。既に複数の完全ゲノム配列が報告されている細菌種が利用できる状況であり、NGS で得られた細菌の *de novo assemble* した *contigs* を、完全ゲノム配列を参照配列 (Ref. seq) として、遺伝子単位で比較 (*blastn* or *blastp*) する。

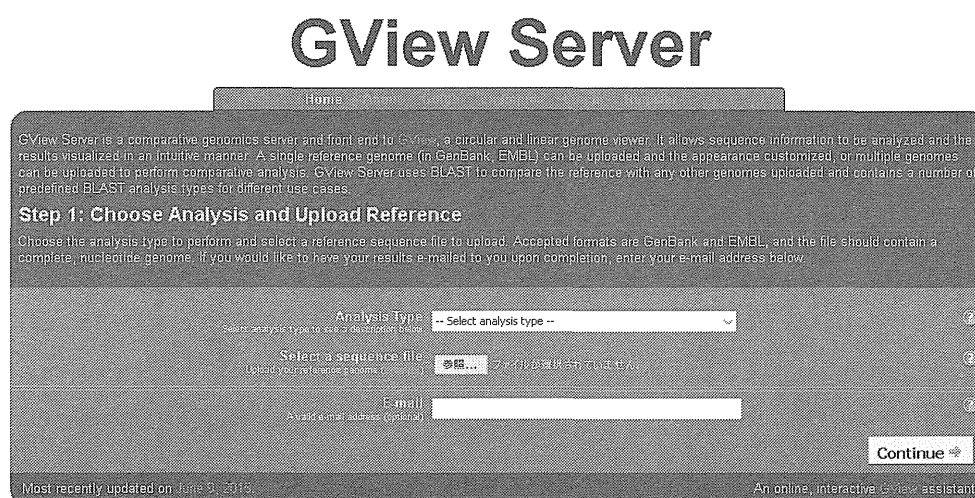
以下の解析が可能である。

1. BLAST Atlas
2. Pangenome Analysis
3. Core Analysis
4. Accessory Analysis
5. Unique Analysis
6. Signature Analysis
7. Reciprocal Analysis
8. Pangenome Phage Analysis

準備：以下の環境を確認

1. OS: Windows7 以上、Windows10 を推奨。インターネット環境必須。
2. ブラウザ：FireFox
3. <https://server.gview.ca/> (以下、トップページ)

Register タブから、氏名とメールアドレスを入力し、登録する。



使用方法：

1. Ref seq は、完全配列が報告されている株を用いる。

2. Refseq は、注釈情報が付いた gbk ファイルを用いると後で便利。付いていない場合には、RAST 等を利用して注釈付きの gbk ファイルを作成するとよい。
3. Query は、gbk でも fasta 形式でもよい。
4. gbk file は、RAST で autoannotation を行い、merged タイプを使うと、artemis で読み込んでくれるので、後で遺伝子構成を見たり、遺伝子やアミノ酸配列を抽出するのが楽となる。

GviewServer への file の upload は、以下の図中に示す、4 ステップからなり、条件を入力していく。以下は、Pangenomic Analysis したときの例である。Final Step の入力が終わったら、念のため、Job ID はコピーしておき、log out する。解析が終わると登録したメールアドレスに、連絡がある。そのときの解析の込み具合にもよるが、翌日の連絡が多い。

右図は、記録のため、FileMaker を使い、解析条件等を記録する画面である。Web 上で、ファイルや条件を入力する順番となっている。この例は、Pangenome analysis を選択したときの例示である。その他の解析をする場合には、この図では、解析タイプを選択できるようになっている。

New	Date	2015/11/03
Step 1	Analysis Type	Pangenome Analysis
	Reference genome	Sinf27-merged.gbk
Step 2	Upload Query Files	
	Blast Program & File features	Sinf27 - Sinf_Meat1_18core-pj033.gbk CDS-CDS
Step 3	Blast Parameters	
	Filter Low complexity sequences?	<input type="checkbox"/> on <input checked="" type="checkbox"/> off
	Expect value cutoff	1e-10
	Genetic code	Bacterial and Plant Plastid
	Alignment length cutoff	100
	Percent identity cutoff	80
	Final Step Map Title	SInfantis2015 Pj035
	Description	Meat1 core seeking Pangenome 80%
	Display reference feature	<input checked="" type="checkbox"/> CDS <input type="checkbox"/> -10_signal <input type="checkbox"/> misc_RNA <input type="checkbox"/> protein_bind <input type="checkbox"/> rRNA <input type="checkbox"/> -35_signal <input type="checkbox"/> gene <input type="checkbox"/> repeat_region <input type="checkbox"/> RBS <input type="checkbox"/> tRNA
	Show GC content	<input type="checkbox"/> on <input checked="" type="checkbox"/> off
	Show GC skew	<input type="checkbox"/> on <input checked="" type="checkbox"/> off
	Show COG categories	<input type="checkbox"/> on <input checked="" type="checkbox"/> off
	Show legends	<input type="checkbox"/> on <input checked="" type="checkbox"/> off
	Select a font	Arial
	Options	one strand line
	Backbone	large
	JobID	D96519393986949C7F47FEC56385728

解析が終わったという連絡が来たら、Web からログインして、以下のような画面まで行く。

結果の viewing は、2 通りの方法がある。

WebStart と結果ファイルを download して、中にあるバッチファイルから、表示する方法である。いずれも JAVA が予めインストールされていることが必要。

GView Server

GView Results for Job D965193939986949C7F47FEC56385728

View



Launch GView Webstart
Circular



Launch GView Webstart
Linear

The above images are previews of the GView Server analysis. Clicking on the image will display a full size version.

Clicking **Launch GView Webstart** will open the results in GView, an interactive circular and linear genome viewer, where you can pan, zoom, and interact with the results. GView requires Java 1.5 or greater and at least 512MB of memory.

The **Pangenome** analysis type displays a pangenome BLAST Atlas of a set of genomes. The reference genome consists of a pangenome constructed by concatenating all the unique regions among each of the uploaded query genomes. A BLAST atlas is then constructed for each query genome. The features in these slots show regions where there is a BLAST hit between the query genome and the constructed pangenome. An example of how to interpret a Pangenome BLAST Atlas can be found on the examples page at [Pangenome of the E. coli](#).

Download

The below table lists all downloadable files containing the resulting data of this analysis. There are two types of data that can be downloaded: the GView zip package, and the BLAST results table. The zip package contains all the necessary information to launch GView offline. The BLAST Results table contains the results of the BLAST analysis that were performed on the uploaded data in a tabular format. To launch GView offline, please execute the appropriate QuickLaunch program, either `QuickLaunch.bat` or `QuickLaunch.sh`.

File	Type	Description
gviewall-D965193939986949C7F47F47FEC56385728.zip	GView Data + Executable	Contains the resulting data plus the GView executable.
gview-D965193939986949C7F47F47FEC56385728.zip	GView Data	Contains only the resulting data from the job, not the gview executable.
blastresults-D965193939986949C7F47F47FEC56385728.csv	BLAST Results Table	A table of the resulting BLAST hits, in csv format. The results may take a while to download for large data sets.
blastresults-D965193939986949C7F47F47FEC56385728.xls	BLAST Results Table	A table of the resulting BLAST hits, in excel format. The results may take a while to download for large data sets.

Logged in as: [M4W4ts](#) Most recently updated on: June 3, 2019. An online, interactive GView assistant.

- ①JAVAstartのスイッチがリンクされている。ここをクリックすると、うまくするとJAVAstartが開始され、結果のファイルが読み込まれて、結果が図で表示される。しかし、PCにより、エラーになる場合が多く、使いにくい。うまく、表示されても、色を変更したりしたい場合には、結果のファイルを一度、PCにdownload (DL)して、Viewing softであるGviewで表示した方が使いやすいので、ほとんどは、②で行っている。
- ②のJobIDをダブルクリックして、DownlordからGview Data+ Excutable typeのfile (ZIP)をPCにDLする。DLしたファイルは圧縮されているので、解凍する。

解凍すると解凍ホルダ内には以下のようなファイル構成となっている。

JAVA上で動く、gviewのソフトが入っているのので、インストールする必要はない。勿論、インストールしても利用できる。

名前	更新日時	種類	サイズ
task_1	2015/11/03 18:17	ファイル フォルダ	
gview.jar	2014/01/03 13:09	Executable Jar File	6,746 KB
QuickLaunch.bat	2015/11/03 3:16	Windows バッチ ファ...	1 KB
QuickLaunch.sh	2015/11/03 3:16	SH ファイル	1 KB
results.gff	2015/11/03 0:50	GFF ファイル	1,598 KB
style.gss	2015/11/03 0:50	GSS ファイル	4 KB

Task_1ホルダの中には、gbkファイルが入っている。

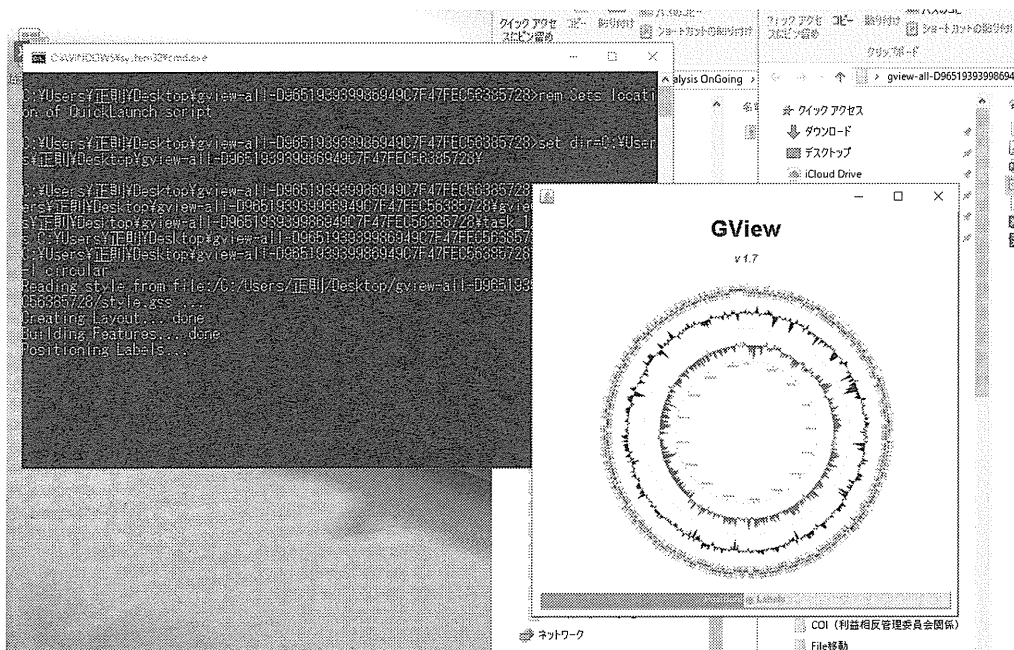
クイックランチャーのバッチファイルが装備されており、まずは、試しにダブルクリックしてみる。運が良ければ、プログラムが稼働し、blast atlasの画面が表示される。

名前	更新日時	種類	サイズ
panggenome.gbk	2015/11/03 0:47	GBK ファイル	8,947 KB

クイックランチャーのバッチファイルをダブルクリックすると、一瞬、何かウインドウが現れるが、直ぐに消える現象が現れる。これが現れるとうまく、表示してくれない。

これは、インストールされている JAVA アプリの状態に依存するものと考えられ、Windows 環境でこれを設定することは、素人には難しい。これは、Linux でもそうであるが、作業ホルダのディレクトリを理解する必要があり、Windows で慣れている人間にはなかなか理解しにくいところ。おそらく、Gview Server から DL したクイックランチャーのバッチファイルには、相対パスではなく、絶対パスで記述されており、これが WindowsPC 上で合致しないとプログラムが動作しないと考えられた。そこで、動作するディレクトリーを探したところ、デスクトップでうまく行くことが分かった。

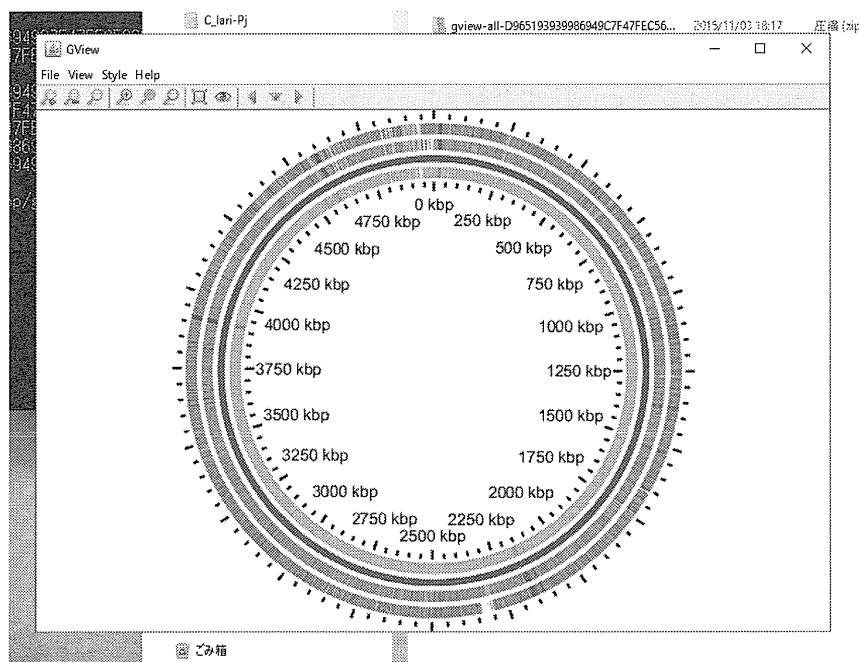
解凍したホルダをデスクトップ画面にドラッグして、クイックランチャーバッチファイルをダブルクリックすると、以下のような画面が現れて、動作が始まる。もし、ここでうまく表示されない場合には、C:ドライブの直下等に解凍したフォルダをコピーして、試してみると開く場合がある。このことは、使用する PC 環境、固有の条件となっている可能性があるため、GViewServer が使用できるディレクトリーを決定しておくことが重要である。



以下のように、うまく表示されるはずですが。

ここまで来れば、後は比較的簡単に、カスタマイズできるようになる。詳細は GView の help を参照のこと。

gbk ファイルを使っているので、カーソルを重ねると遺伝子名が表示される、拡大が出来る、blast 検索のカットオフ値が、ファイルの upload 時に変更できるなど、便利にゲノム全体を比較することが出来る。



(Ver1.1)

Windows PC を用いたスタンドアロン BLAST のセットアップと Blast 検索

はじめに

NCBI では、BLAST 検索サービスを Web 上のサービスとして提供しているだけでなく、Windows PC を用いて、Web を介さない BLAST 検索ができる環境も提供している。この検索は、DOS コマンドウインドウから、コマンドラインで操作するもので、グラフィカルインターフェイスではない。

これに使用するプログラムはblast+と呼ばれ、NCBI C++ based BLAST programs をWindows PCにインストールして使用する。これまでのWindows バージョンでは、Windows PCのDOSコマンドの入力には、コピー&ペーストは利用できなかったため、作業領域として、pathを理解する必要があることと正確にコマンドスクリプトをタイプする必要があるため、非常に使いにくかったが、最新のバージョンであるWindows10から、コピー&ペーストが出来るようになり、非常に簡単にコマンドが使えるようになった。これまで、blast検索は、blastnやblastpなどを中心として、DDBJやNCBIのWeb service を用いた利用がほとんどであったが、Windows PC単独で、NCBI blast+ packageが使えるようになれば、blast+で提供されているすべてのプログラムが利用できるようになり、その解析の幅が大きく広がることが期待される。また、NGSから得られたリードやContigsを利用した検索に利用できるため、Windows PCを用いたNGSデータの解析の幅が広がることを期待できる。

1. ソフトの入手と準備

Blast Packages:

http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download

BLAST+ executables の Installers and source code are available from

<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/>.

から、最新版の

ncbi-blast-2.2.30+-win64.exe (64-bit PC 用)をダウンロードする。

(バージョンアップされていれば、そのバージョンとして明記されたファイル名となっています)

今回インストールした PC 環境

Windows 10 Home, Intel Core i3-4160 CPU 3.60 GHz

16.0 GB RAM, 64-bit operating system

インストール

ncbi-blast-2.2.30+-win64.exe をダブルクリックすると自動的にインストールが始まり、終了する。これでWindows 環境下、提供されている blast プログラムがすべて利用できるようになる。

作業フォルダの設定

C:\blast\db を作成する。今後、この db フォルダを使って、blast 検索を実施する。

2. プログラム一覧 (一部の説明文は、ウィキペディアから抜粋)

1	blastdbcheck	BLAST データベースの完全性を確認する。
2	blastdbcmd	BLAST データベースから塩基配列やその他の情報を検索する。
3	blastdb_aliastool	
4	blastn	塩基配列 (ヌクレオチドのシーケンス) をクエリシーケンスとして、ユーザが指定した塩基配列データベースから類似するシーケンスを検索する。
5	blastp	アミノ酸配列 (タンパク質のシーケンス) をクエリとして、ユーザが指定したアミノ酸配列データベースから類似するシーケンスを検索する。
6	blastx	塩基配列のクエリシーケンスを、6 フレームで翻訳して、アミノ酸配列データベースから類似するシーケンスを検索する。
7	blast_formatter	
8	convert2blastmask	
9	deltablast	
10	dustmasker	
11	legacy_blast.pl	
12	makeblastdb	blast 検索をするために、FASTA ファイルを BLAST データベース (db) に変換する。
13	makembindex	
14	makeprofiledb	
15	psiblast	プロファイルを利用したホモロジー検索を行う。 blastp よりも遠縁種の検索が得意。アミノ酸配列をクエリとして、ユーザが指定したアミノ酸配列データベースから類似するシーケンスを検索する。最近開発された BLAST プログラムの一つ。このプログラムはタンパク質の遠い系統的関連性を発見するために使われる。まず、クエリのアミノ酸配列と近い関連性をもつタンパク質の一覧を作成する。そしてこの一群のタンパク質を統合して「プロファイル」(平均化したシーケンスのようなもの) を作る。このプロファイルを使ってアミノ酸配列データベースに対する検索を行い、プロファイルのもととなったアミノ酸配列より多い件数のアミノ酸配列を見つける。見つかったアミノ酸配列でさらにプロファイルを作成する。この一連の処理を繰り返し行う。検索を行う際に関連するタンパク質を考慮するため、遠い系統的関連性を発見する場合に、 PSI-BLAST は blastp よりもはるかに正確な処理を行う。
16	rpsblast	blast のアルゴリズムを使ったプロファイル検索プログラム。
17	rpsblastn	
18	segmasker	
19	tblastn	アミノ酸配列をクエリとして、塩基配列データベースを 6 フレームで翻訳しながら、類似するシーケンスを検索する。
20	tblastx	塩基配列のクエリシーケンスを、6 フレームで翻訳して、アミノ酸

		配列に翻訳された塩基配列データベースから類似するシーケンスを検索する。このプログラムは BLAST のプログラムの中で最も処理速度が遅い。このプログラムの使用目的は、塩基配列のシーケンス間の非常に遠い関連性を発見することである。
21	update_blastdb.pl	
22	windowmasker	

空欄は、ほとんど使うことはないが、今後、利用法について研究したい。また NCBI の BLAST データベースは日々、アップデートされているらしい。

Databases

BLAST databases are updated daily and may be downloaded via FTP from <ftp://ftp.ncbi.nlm.nih.gov/blast/db/>. Database sets may be retrieved automatically with `update_blastdb.pl`, which is part of the BLAST+ suite. Please refer to the [BLAST database documentation](#) for more details.

3. 公的データベースのダウンロード(公的データベースの利用が前提の場合)

1. 次のリンクをクリックして NCBI の ftp サーバーにアクセスする。
<ftp://ftp.ncbi.nih.gov/blast/db/>
2. ダウンロードしたいデータベースファイル (たとえば 16SMicrobial.tar.gz) をダブルクリックする。
3. ファイルを C:\blast\db に保存する。
4. ダウンロードしたファイルを展開する。 解凍ソフトとして 7-zip が使いやすい (無料)。

解凍したファイルは、

```
16SMicrobial.nsq  
16SMicrobial.nsi  
16SMicrobial.nsd  
16SMicrobial.nog  
16SMicrobial.nni  
16SMicrobial.nnd  
16SMicrobial.nin
```

16SMicrobial.nhr の 8 つファイルが存在する。blastn を実施するとき、この 8 つの一つのセットとして、そのまま使用可能である。

4. In-house データベースの作成と blast サーチ

準備

1. blast 検索に使用したいデータベースは、FASTA ファイル形式として作成する。このままでは、blast のデータベースとしては使用できないので、データベースとして使用できるように様式を変換する必要がある。
2. コマンドプロンプトを開く。(「管理者用のコマンドプロンプト」も存在するが、これは使用しない)
3. [cd 半角スペース]入力し、C:\blast ウィンドウ内のフォルダ db をコマンドプロンプト・ウィンドウにマウスを使って、ドラッグする。Enter
4. C:\blast\db>
表示になる。
5. 点滅カーソル状態で、[dir] Enter
6. XXXXX.fas が表示されていることを確認する。
7. この後、query 配列とデータベースは全て、このディレクトリを使用する。また、blast の結果は、全てこのディレクトリに作成される。

データベースの作成

1. DNA 配列かアミノ酸配列の FASTA ファイルとして作成する。ファイル名は、DNA の場合には、XXXXXnuc.fas、蛋白質の場合には、XXXXXaa.fas などと区別するように命名するとよい。ここで、作成した FASTA ファイル、例えば、XXXXXnuc.fas は、C:\blast\db フォルダにコピーする。

次にこの“XXXXXnuc.fas”という名前の fasta ファイルをデータベースとして利用する様式に変換する。

2. “makeblastdb -in XXXXXnuc.fas -out XXXXXnucDB -dbtype nucl”と入力し、**Enter**を押す。
3. 「XXXXXnucDB」は、データベース名である。「-dbtype nucl」は、入力必須項目である。因みに蛋白質の場合は、「-dbtype prot」と入力する。
4. “db” ディレクトリに以下の3つのファイルが作成されたことを確認する。

```
XXXXXnucDB.nhr  
XXXXXnucDB.nin  
XXXXXnucDB.nsq
```

5. blastn

query に” dna-query.fas” ファイルを使用する場合、

“C:\blast\db>”の後に、以下のように入力する。

```
blastn -query dna-query.fas -db XXXXXDB -out dna-query.out
```

と入力して、**Enter**を押すところであるが、この入力に、コピー&ペーストを利用する。

C:\blast\db>blastn -query まで入力して、db フォルダ（のアクティブウインドウ）から、「dna-query.fas」をドラッグすると、以下ようになる。

```
C:\blast\db>blastn -query C:\blast\db\dna-query.fas 、そこで続けて
```

```
C:\blast\db>blastn -query C:\blast\db\dna-query.fas -db と入力し、db  
フォルダから、データベースとして、作成した上記3つの「XXXXXDB」のうち、  
いずれか一つをドラッグする（今回は、「XXXXXnucDB.nhr」を用いる）と、
```

```
C:\blast\db>blastn -query C:\blast\db\dna-query.fas -db
```

```
C:\blast\db\XXXXXDB.nhr
```

となる。そこで、データベースファイルの3つに共通な部分である

「XXXXXuncDB」だけを残すため、「.nhr」の文字を `BackSpaceKey` で削除する。

そして、その後に結果の出力ファイル名として、「-out dna-query.out」と入力して、`Enter`を押す。

“db”内に、瞬時に”dna-query.out”が作成される。

このコマンドの意味は、クエリー (-query) に「dna-query.fas」という fasta 形式のファイルにある配列 (単独でもマルチでも可) を用いて、データベース (-db) として、[XXXXXDB] (上記の3つのファイルを指定するために、最初の共通の文字列のみを入力する) を使用して、その結果を (-out) [dna-query.out] というファイル名として、書き出す、ことである。

Comments & Options

① option switch を指示しなければ、デフォルトでは、500 行と 250 アラインメントの表示となる (従って、長大なファイルとなる)。これを変更する場合には、以下の通り、表示 option スイッチと表示数を加える

```
>blastn -query [dna-query.fas] -db [XXXXXDB] -out [dna-query.out] (-  
num_descriptions <500> -num_alignments <250>)
```

②表形式結果表示 ⇒ エクセルでの表形式で結果を示すことができるので、利用価値は高い。

```
>blastn -query [dna-query.fas] -db [XXXXXDB] -out [dna-query.out] -  
outfmt [String]
```

*String=6: tabular (ACT で利用する Comparison file の作製に必要)

*String=7: tabular with comment lines

*String=10: Comma-separated values

その他、多数の formatting option が利用可能である。この説明は、[-help] で確認することができる。

(Version 1.2 2016.02.29)

NGS解析のコスト(計算例)

24細菌分離株、カバー率、>100とし、均一で高品質のデータを得るために、MiSeqRunを2回実施を前提として、1検体(=1細菌株)当たりのコストを試算した。

試薬キット名	規格	納入価	1検体(株)当り	24検体当り
Qiagen PureGene Yeast/Bact Kit B	200mL	¥35,100	¥176 (1mL)	¥4,224
Qubit dsDNA HS Assay Kit	500回	¥30,000	¥60	¥3,200 (2回分+α)
Qubit Assay Tube	500回	¥5,900	¥12	¥624
Nextera XT DNA Library Kit	24回分	¥103,680	¥4,320	¥51,840 (1/2 scale)
Wizard SV Gel and PCR Clean-up System	250回分	¥33,372	¥134	¥3,204
KAPA Library Quantification Kit	500回分	¥135,000	¥270	¥13,500 (50回分)
MiSeq reagent Kit v3 600PE		¥270,000		¥540,000 (2 runs)
			合計 1検体当り	¥616,592 ¥25,691

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

地方衛生研究所の連携による食品由来病原微生物の網羅的ゲノム解析を基盤とする

新たな食品の安全確保対策に関する研究

研究代表者 調 恒明 山口県環境保健センター・所長

愛媛県内で、患者、食材、家畜から分離されたサルモネラ株の
次世代シーケンサーによるゲノム解析 (3)

研究分担者	四宮 博人	愛媛県立衛生環境研究所
研究協力者	仙波 敬子	愛媛県立衛生環境研究所
	園部 祥代	愛媛県立衛生環境研究所
	大塚 有加	愛媛県立衛生環境研究所
	菅 美樹	愛媛県立衛生環境研究所
	山下 まゆみ	愛媛県立衛生環境研究所
	横山 栄二	千葉県衛生研究所
	黒田 誠	国立感染症研究所
	関塚 剛史	国立感染症研究所
	山下 明史	国立感染症研究所
	加藤 健吾	国立感染症研究所

研究要旨

サルモネラ属菌による食中毒・感染性胃腸炎の起因菌として、*Salmonella Infantis* は上位を占める確かな感染制御が望まれる。愛媛県内で分離された、患者由来株 31 株、食材（鶏肉、豚肉）由来株 29 株、家畜（豚）由来株 10 株の計 70 株の *S. Infantis* の全ゲノム配列を次世代シーケンサー(NGS)で決定し、詳細な系統解析を実施した。その結果、3つの主要なクラスター、クラスター1（患者株のみ）、クラスター3（患者株と動物由来株）、クラスター5（患者株と食材由来株）、が得られた。これらのクラスター内で、患者由来株と食材・動物由来株からなる近縁のペアが認められ、感染経路を示唆する。なかでも、同一のSNV(single nucleotide variation)を示した鶏肉由来と患者由来の2株は、同一クローンの可能性が高く、原因食材である可能性を強く示唆する。また、クラスター1は、愛媛株では患者株のみであったが、公開されている *S. Infantis* のゲノム情報と比較すると、他の都道府県で分離された鶏卵由来株と同一のクラスターを形成した。さらに、*S. Infantis* 株のうち薬剤耐性を示す株（患者由来株、食材由来株）は全てクラスター5に属していた。これらの薬剤耐性株はメガプラスミドの遺伝子マーカーである *irp2* 遺伝子を保有していたことから、本プラスミド上に薬剤耐性遺伝子が存在することが示唆された。以上の結果は、NGSによる迅速ゲノム解析が病原体の高精度同定や分子疫学に基づく感染症対策に極めて有用であることを示している。

A 研究目的

サルモネラ属菌はヒトの急性胃腸炎・食中毒の原因菌であり、数多くの血清型が存在する。非チフス性のサルモネラ属菌による感染性胃腸炎は、世界的には年間 1 億人近い患者が発生している。日本においても、食中毒事例は、毎年全国で多数発生し、発生件数、患者数ともに細菌性食中毒の原因菌の上位を占めている²⁾。

一方、食中毒として扱われない散発性サルモネラ感染症については、「感染症の予防及び感染症の患者に対する医療に関する法律」における全数把握対象疾患とされていないため、分離菌の性状、薬剤耐性、感染経路等に不明な点が多く、公衆衛生上の問題となっている。

サルモネラ分離株の分子疫学的解析は感染を制御するうえで重要である。我々は、患者、食材、家畜由来株の血清型分類、薬剤感受性試験、パルスフィールドゲル電気泳動(PFGE)解析を実施してきた³⁾。分離株の中で、*Salmonella* *Infantis*(*S. Infantis*)は、患者のみならず食材や家畜からも高率に分離されている。また、*S. Infantis* は食中毒・感染性胃腸炎の原因菌として、*S. Enteritidis* に次ぐ血清型である(2012年第2位、2013年第3位、2014年第2位、2015年第4位)⁴⁾。

近年、次世代シーケンサー(NGS)の開発により、短時間の全ゲノム解析が可能になり、より詳細な分子疫学解析ができるようになった。今回、我々は、患者、食材(鶏肉、豚肉)、家畜(豚)から分離された *S. Infantis* の全ゲノム解析を基に分子系統解析を実施し、分離株間の分子疫学的な関連性を検討したので報告する。

B 材料と方法

1 サルモネラ菌株

愛媛県内で2008年～2010年に、患者、食材、家畜等から収集した菌株を用いた。患者由来株は、協力医療機関3機関、検査センター2機関、保健所及び当所から分離された。食材由来株は、当所

及び保健所で実施した愛媛県食中毒菌汚染実態調査及び委託検査から得られた。家畜由来株として、と畜場に搬入された豚(盲腸便、皮剥前腎・腹部)からの分離株を収集した。分離株はサルモネラと同定された後、O抗原、H抗原の組み合わせによって、Kauffmann-Whiteの様式により血清型を分類し、*S. Infantis*を同定した³⁾。

2 薬剤感受性試験

薬剤感受性試験は、CLSIの抗菌薬ディスク感受性試験実施基準に基づき感受性ディスク(BD)を用いて実施した。供試薬剤はアンピシリン(ABPC)、ストレプトマイシン(SM)、カナマイシン(KM)、ゲンタマイシン(GM)、テトラサイクリン(TC)、クロラムフェニコール(CP)、ST合剤(ST)、サルファ剤(SUL)、セフトキシム(CTX)、シプロフロキサシン(CPFX)、ナリジクス酸(NA)、ホスホマイシン(FOM)、セフトジジム(CAZ)、イミペネム(IPM)を用いた。

3 ゲノム DNA の抽出

S. Infantis のゲノム DNA は、菌体から粗抽出した DNA 分画から DNeasy 96 Blood & Tissue Kit 用いて抽出した。

4 次世代シーケンサーによる DNA シーケンス

ライブラリは Nextera XT DNA Sample Prep Kit (illumina) を用いて作成し、1% アガロースゲル電気泳動で得られた 750bp～1000bp のバンドを切り出し、Wizard SV Gel and PCR Clean-Up System で抽出した。ライブラリのシーケンスは、MiSeq あるいは NextSeq 500 (illumina) を用いて行った⁵⁾。得られたリードの *de novo* アセンブリは国立感染症研究所病原体ゲノム解析研究センターで実施した⁶⁾。得られた *S. Infantis* のゲノム配列から SNV(single nucleotide variation)を抽出し、RAxML (最尤

法)、bootstrap 1000 回で系統樹を作製し、クラスター解析を実施した。

5 *S. Infantis* 株の *irp2* 遺伝子の検出

S. Infantis 株のメガプラスミド (megaplasmid) 保有を検出するため、本プラスミドの遺伝子マーカーである *irp2* 遺伝子の特異プライマー (Forward primer, 5'-AAGGATTCGCTGTTACCGGAC-3'; Reverse primer, 5'-TCGTCGGGCCAGCGTTTCTTCT-3') を用いて PCR 法にて検出した^{7, 8, 9)}。

(倫理面への配慮)

本研究では、特定の研究対象者は存在せず、倫理面への配慮は不要である。

C 研究結果

1 *S. Infantis* の NGS ゲノム解析

患者、食材 (鶏肉、豚肉)、家畜 (豚) から分離された *S. Infantis* 70 株のゲノム DNA を抽出し NGS 解析を実施した。NGS 解析によって見出された SNV に基づき、系統解析を実施し (図 1)、大きく 5 つのクラスターに分かれた。クラスター 1 は全て患者由来株 9 株であった。クラスター 2 は患者株 1 株であった。クラスター 3 は屠畜場で豚直腸便から採取された豚由来株 10 株と患者由来株 3 株を含んでいた。クラスター 4 は患者株 3 株を含んでいた。クラスター 5 は鶏肉由来株 28 株、豚肉由来株 1 株、患者由来株 16 株を含んでいた。クラスター 1~5 に属さない株が 1 株 (E208H22) 存在した。興味深いことに、何らかの薬剤耐性を示した菌株は、全てクラスター 5 に含まれ、クラスター 1~4 には薬剤耐性菌は存在しなかった。

NGS 解析による SNV に基づく系統解析から、70 株の *Infantis* 株中に同一あるいは極めて近縁の菌株が見出された。クラスター 1 の E48H23, 22-179, E207H22, クラスター 3 の H152H20~E150H20, E151H21~E145H21, ク

ラスター 5 の 21-267 と 22-286, E28H22~22-358, 20-132 と 20-157 等である。それぞれの分離株の背景を考慮して解析することにより、菌株間の関係を推定することが可能であり、その結果は感染経路の解明等に科学的根拠を提供すると考えられる。

今回得られた *S. Infantis* 70 株の NGS 解析結果を既に公表されている *S. Infantis* のゲノム配列情報と比較した (図 2)。クラスター 1 は、愛媛株では患者由来株のみのクラスターであったが、他都道府県から報告されていた鶏卵由来株と同じクラスターを形成することが示された。また、クラスター 5 は愛媛株では患者由来株と食材由来株が認められたが、他都道府県から報告されていた食材 (鶏肉) 由来株及びヒト (患者、無症候保菌者) 由来株も同じクラスターを形成し、このクラスターに属する株は薬剤耐性を示した。

最近、*S. Infantis* が保有する約 300 kb のメガプラスミド (megaplasmid) が同定され、本プラスミドが薬剤耐性にも関与することが明らかにされつつある^{8, 10)}。今回の解析で、薬剤耐性を示す株がクラスター 6 にのみ認められ、明瞭な分布を示したことから (図 2)、今回用いた *S. Infantis* 70 株のメガプラスミド保有を調べた。このメガプラスミドの遺伝子マーカーである *irp2* (図 3) に特異的なプライマーを用いて PCR 法で検出したところ、薬剤耐性を示した *S. Infantis* 株は全て *irp2* 陽性で、メガプラスミドを保有することが示唆された (表 1)。

D 考察

愛媛県において感染性胃腸炎患者、食材、家畜から分離されたサルモネラ株の血清型分類と薬剤感受性試験を実施した。これらの分離株のうち、*S. Infantis* については、患者、食材 (鶏肉、豚肉)、家畜 (豚) から分離された 70 株を NGS ゲノム解析することにより、食材・家畜由来株と患者由来株との関係を詳細に解析した (図 1)。PFGE 解析のパターンが同一でも、NGS 解析では相違

が認められ、より高分解能の解析が可能となった。NGS 解析において、同一あるいは非常に近縁の組み合わせが7対以上見出され、菌株間の関係や感染源の由来が見えてくる可能性がある。特にクラスター5の同一SNPを示した鶏肉由来20-132株と患者由来20-157株の2株は、同一クローンの可能性が高く、直接の感染源であることが強く示唆される。また、鶏肉由来22-286株と患者由来21-267株が近縁であり、22-286株が存在していた同じ鶏舎で飼育されたものから感染した可能性がある。さらに、鶏肉由来のE28H22株と21-235株および患者由来22-358株が近い系統で、E28H22株と21-235株が存在していた鶏と同じ鶏舎由来のものから感染した可能性がある。

NGS 解析のクラスター1に属する患者由来株9株を他の都道府県で分離されたものと比較すると、鶏卵由来株と近縁であることが判明した(図2)。一般に、PFGEのような電気泳動に基づくバンドパターンは、異なった施設間での比較が難しい場合があるが、NGS 解析によるシーケンスに基づく比較はそのような曖昧さを排除できるため、全国的な比較が可能である。この関係が示されたことは、クラスター1の患者由来株9株について鶏卵の摂食との関連を調べる根拠となる。また、クラスター3は豚由来株が多く、畜舎でまとまっている傾向が認められた。このクラスターに属する患者由来株も存在することから、豚由来株にヒトの感染が関連する経路も示唆される。

NGS ゲノム解析は、菌株間の近縁性を測るだけでなく、病原性因子や薬剤耐性因子等の感染や流行につながる菌株の特徴についての情報を遺伝子解析によって提供できる。この点もPFGEにはない長所である。

今回得られた結果、および関連する最近の研究は¹¹⁾、NGSによる迅速ゲノム解析が、サルモネラ等病原体の高精度同定や分子疫学に基づく感染症対策において極めて有用であることを示している。

E 結 論

愛媛県で分離された患者、食材(鶏肉、豚肉)、家畜(豚)から分離された*S. Infantis* 70株のNGS ゲノム解析することにより、食材・家畜由来株と患者由来株との関係をより詳細に解析した。特に、同一SNVを示した鶏肉由来と患者由来の2株は、同一クローンの可能性が高く、原因食材である可能性を強く示唆する。これらの結果は、NGSによる迅速ゲノム解析が、病原体の同定や分子疫学に基づく感染症対策に極めて有用であることを示している。

参考文献

- 1) Majowicz SE et al., The global burden of nontyphoidal *Salmonella* gastroenteritis. *Clin Infect Dis.* (2010)50: 882-9.
- 2) 厚生労働省 食中毒統計資料 <http://www.mhlw.go.jp/topics/syokuchu/04.html#4-2>
- 3) 松本純子ほか、愛媛県内におけるサルモネラ感染症の発生動向調査と分子疫学的解析、愛媛県立衛生環境研究所年報 (2013) 14:1-9.
- 4) 病原体検出情報 サルモネラ血清型割合、2012 ~ 2016 年 <http://www0.nih.go.jp/niid/idsc/iasr/Byogentai/Pdf/data48j.pdf>
- 5) Kuroda M et al., Characterization of quaspecies of pandemic 2009 influenza A virus (A/H1N1/2009) by *de novo* sequencing using a next-generation DNA sequencer. *PLoS One.* (2010) 5(4):e10256.
- 6) Sekizuka T et al., Complete sequencing of the *bla*(NDM-1)-positive IncA/C plasmid from *Escherichia coli* ST38 isolate suggests a possible origin from plant pathogens. *PLoS One.* (2011) 6(9):e25334.
- 7) Karch H et al., A genomic island, termed

high-pathogenicity island, is present in certain non-O157 Shiga toxin-producing *Escherichia coli* clonal lineages. *Infect Immun.* (1999) 67(11):5994-6001.

- 8) Aviv G et al., A unique megaplasmid contributes to stress tolerance and pathogenicity of an emergent *Salmonella enterica* serovar Infantis strain. *Environ Microbiol.* (2014) 16(4):977-994.
- 9) Yokoyama E et al., Phylogenetic and population genetic analysis of *Salmonella enterica* subsp. *enterica* serovar Infantis strains isolated in Japan using whole genome sequence data. *Infect Genet Evol.* (2014) 27:62-68.
- 10) Franco A et al., Emergence of a clonal lineage of multidrug-resistant ESBL-producing *Salmonella* Infantis transmitted from broilers and broiler meat to humans in Italy between 2011 and 2014. *PLoS One.* (2015)10(12):e0144802
- 11) Octavia S et al., Delineating community outbreaks of *Salmonella enterica* serovar Typhimurium using whole genome sequencing: insight into genomic variability within an outbreak. *J Clin Microbiol.* (2015) doi:10.1128/JCM.03235-14

G 研究発表

1. 論文発表

- 1) 四宮博人, 勢戸和子, 川瀬 遵, 有川健太郎, 船渡川圭次, 鈴木匡弘, 久保田寛顕, 調 恒

明: 地方衛生研究所における細菌学的検査・研究の最新事情. *日本細菌学雑誌* 70(2):309-318, 2015.

- 2) 菅 美樹, 四宮博人, 北尾孝司: 市販鶏レバーおよび臨床材料から分離した基質特異性拡張型 β -ラクタマーゼ産生 *Escherichia coli* および *Klebsiella pneumoniae* が保有する *bla*_{CTX-M} 型別に関する検討. *感染症学雑誌* 印刷中

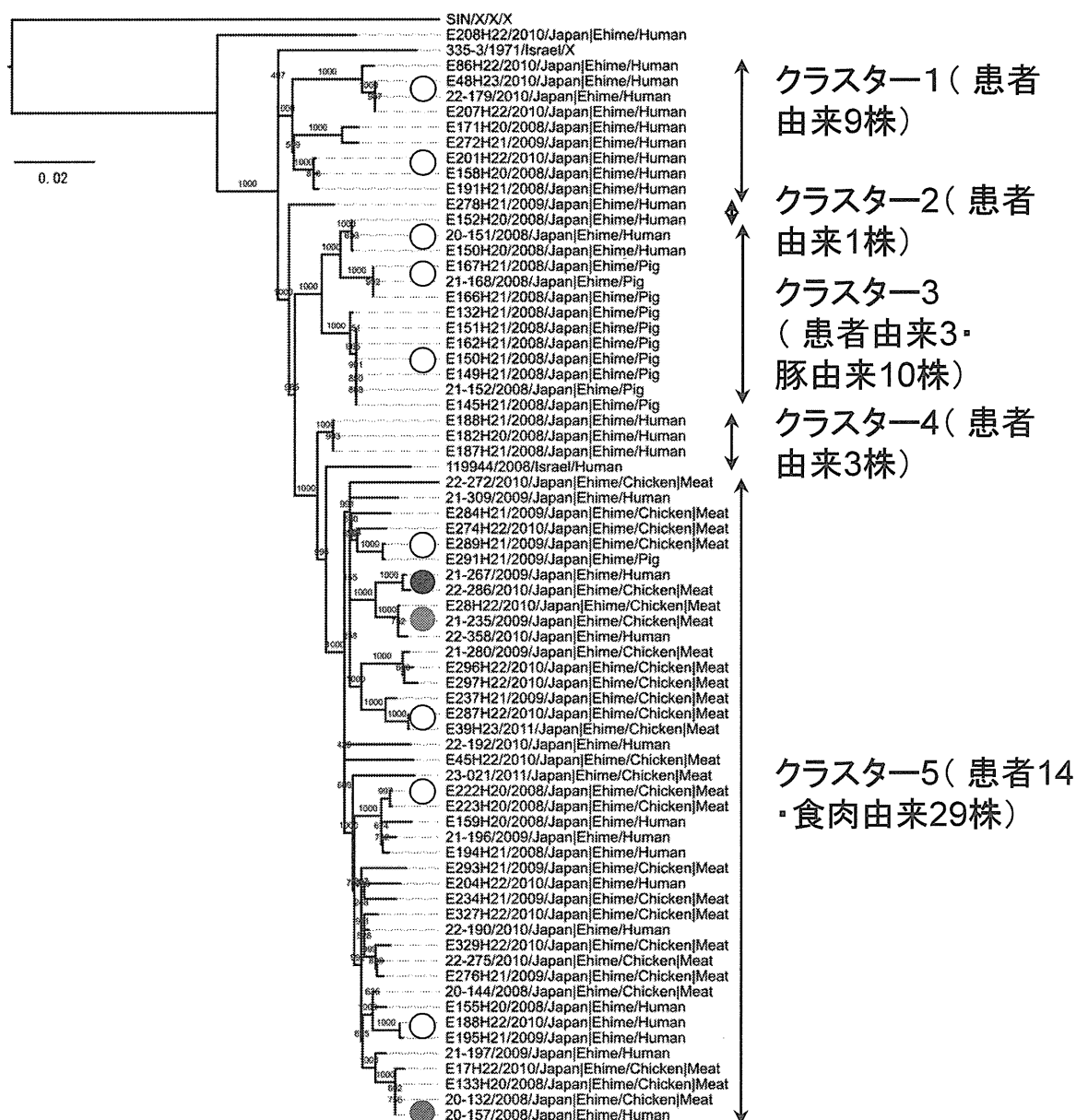
2. 学会発表

- 1) 仙波敬子, 園部祥代, 木村俊也, 大倉敏裕, 鳥谷竜哉, 四宮博人: 地研における薬剤耐性菌解析の取り組み, 衛生微生物技術協議会第 36 回研究会, 2015. 7. 23-24, 仙台
- 2) 木村千鶴子, 仙波敬子, 園部祥代, 木村俊也, 四宮博人: 小児感染性胃腸炎患者から分離された腸管凝集付着性大腸菌の性状について, 第 68 回日本細菌学会中国・四国支部総会, 2105.10.3-4, 岡山
- 3) Keiko Semba, Mayumi Yamashita, Sachiyo Sonobe, Eiji Yokoyama, Tsuyoshi Sekizuka, Komei Shirabe, Makoto Kuroda, and Hiroto Shinomiya: Whole genome analysis of *Salmonella* isolates from foods and patients reveals their detailed relationships. シンポジウム 7 「ゲノム解析手法の最前線」、第 89 回日本細菌学会総会, 2016.3.23-25, 大阪 (予定)

H 知的財産権の出願・登録状況

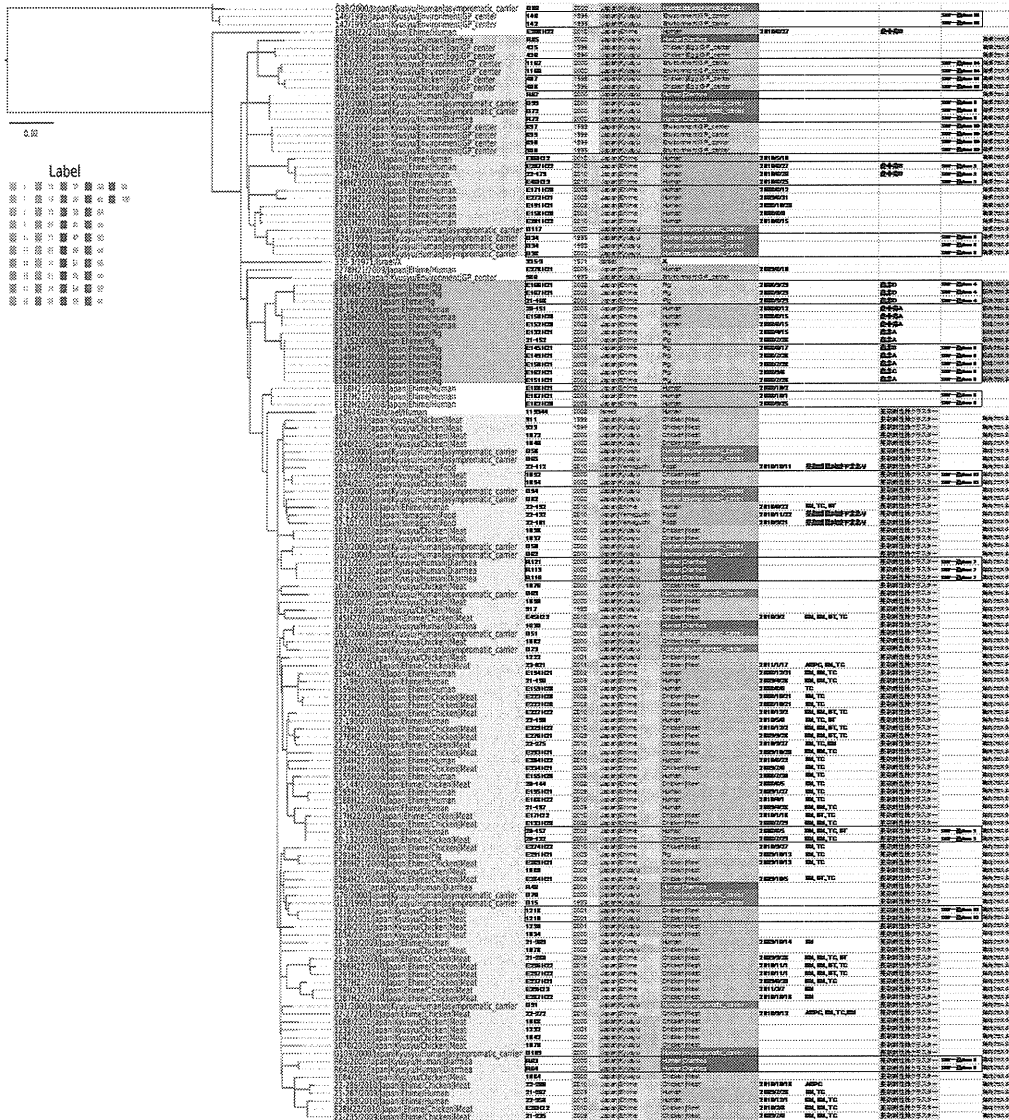
なし

図1. S. Infantis愛媛70株の次世代シーケンサーを用いたゲノム解析に基づく系統解析



NGS解析において、同一あるいは非常に近縁の組み合わせが10対以上見出された(○)。特にクラスタ-5の同一SNPを示した鶏肉由来20-132株と患者由来20-157株の2株は(●)、同一クローンの可能性が高く、直接の感染源であることが強く示唆される。また、鶏肉由来22-286株と患者由来21-267株が近縁であり(●)、22-286株が存在していた同じ鶏舎で飼育されたものから感染した可能性がある。さらに、鶏肉由来のE28H22株と21-235株および(●)患者由来22-358株が近い系統で、E28H22株と21-235株が存在していた鶏と同じ鶏舎由来のものから感染した可能性がある。

図2. 愛媛株及び他都道府県株を含む141株の解析



鶏卵クラスター

豚クラスター

鶏肉クラスター

NGS解析のクラスター1に属する愛媛の9株を他の都道府県で分離されたものと比較すると、鶏卵由来株と近縁であることが判明した(上図)。

この関係が示されたことは、クラスター1の患者由来9株について鶏卵の摂食との関連を調べる根拠となる。

また、他の都道府県のヒト由来株、鶏肉由来株も愛媛株のクラスター5と同じクラスターを形成し、薬剤耐性菌は全てこのクラスターに属していた。

図3

Environ Microbiol 16: 977-994, 2013

Aviv G et al., A unique megaplasmid contributes to stress tolerance and pathogenicity of an emergent *Salmonella enterica* serovar Infantis strain.

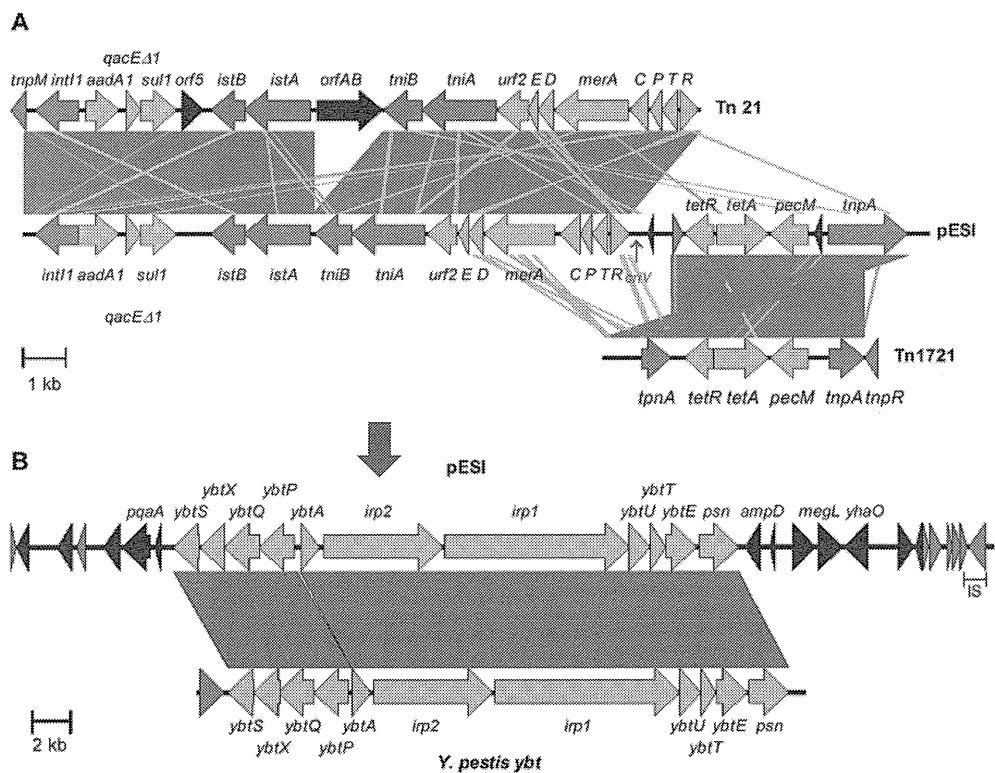


表1. *S. Infantis* 株における *irp2* 遺伝子保有と薬剤耐性との相関

NGS解析に基づくクラスター	菌株の由来	薬剤耐性	<i>irp2</i> 保有*
クラスター1	患者9	—	no
クラスター2	患者1	—	no
クラスター3	患者3, 動物10	—	no
クラスター4	患者3	—	no
クラスター5	患者14, 食材29	ABPC, KM, SM, TC, ST	yes

*PCRによる検出

Ⅲ 研究成果の刊行に関する一覧表

研究成果の刊行に関する一覧表

著者名	論文タイトル名	発表雑誌	巻号ページ	出版年
Yoshikawa T, Shimojima M, Fukushi S, Tani H, Fukuma A, Taniguchi S, Singh H, Suda Y, Shirabe K, Toda S, Shimazu Y, Nomachi T, Gokuden M, Morimitsu T, Ando K, Yoshikawa A, Kan M, Uramoto M, Osako H, Kida K, Takimoto H, Kitamoto H, Terasoma F, Honda A, Maeda K, Takahashi T, Yamagishi T, Oishi K, Morikawa S, Saijo M.	Phylogenetic and Geographic Relationships of Severe Fever With Thrombocytopenia Syndrome Virus in China, South Korea, and Japan.	J Infect Dis.	15;212(6):88 9-98	2015
Shinomiya H, Seto K, Kawase J, Arikawa K, Funatogawa K, Suzuki M, Kubota H, Shirabe K.	Current status of bacteriological studies at prefectural and municipal public health institutes in Japan	Nihon Saikingaku Zasshi.	70(2):309-18.	2015
Kimura H, Saitoh M, Kobayashi M, Ishii H, Saraya T, Kurai D, Tsukagoshi H, Shirabe K, Nishina A, Kozawa K, Kuroda M, Takeuchi F, Sekizuka T, Minakami H, Ryo A, Takeda M.	Molecular evolution of haemagglutinin (H) gene in measles virus.	Sci Rep.	5: 11648..	2015
Kawase J, Etoh Y, Ikeda T, Yamaguchi K, Watahiki M, Shima T, Kameyama M, Horikawa K, Fukushima H, Goto R, Shirabe K.	Improved multiplex real-time SYBR Green PCR assay for analysis of 24 target genes from 16 bacterial species in fecal DNA samples from patients with foodborne illnesses.	Jpn J Infect Dis.	[Epub ahead of print]	2015

<p>Kobayashi M, Yoshizumi S, Kogawa S, Takahashi T, Ueki Y, Shinohara M, Mizukoshi F, Tsukagoshi H, Sasaki Y, Suzuki R, Shimizu H, Iwakiri A, Okabe N, Shirabe K, Shinomiya H, Kozawa K, Kusunoki H, Ryo A, Kuroda M, Katayama K, Kimura H.</p>	<p>Molecular Evolution of the Capsid Gene in Norovirus Genogroup I.</p>	<p>Sci Rep..</p>	<p>5:13806</p>	<p>2015</p>
<p>Obuchi M, Yagi S, Oguri A, Takizawa T, Kimura H, Sata T.</p>	<p>Outbreak of human bocavirus 1 infection in young children in Toyama,</p>	<p>Jpn J Infect Dis.</p>	<p>68(3):259-61.</p>	<p>2015</p>
<p>Yamasaki E, Watahiki M, Isobe J, Sata T, Nair GB, Kurazono H.</p>	<p>Quantitative Detection of Shiga Toxins Directly from Stool Specimens of Patients Associated with an Outbreak of Enterohemorrhagic Escherichia coli in Japan—Quantitative Shiga toxin detection from stool during EHEC outbreak.</p>	<p>Toxins (Basel).</p>	<p>27;7(10):438 1-9.</p>	<p>2015</p>

