

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）
総括研究報告書

全ゲノム情報を用いた腸管出血性大腸菌サーベイランス実用化に関する研究

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

A．研究目的

腸管出血性大腸菌（enterohemorrhagic *Escherichia coli*: EHEC）は人に下痢などを主徴とする感染症を起こす食中毒細菌である。同菌感染症の重症例では血便や溶血性尿毒症症候群を経て死に至らしめ、国内では年間3,000名以上の感染症患者が報告されることから公衆衛生上の脅威となっている。同菌は菌体の抗原性によって多数のO血清群に分けられ、重症患者から分離される株の9割以上がO157、O26、O111、O103、O121、O145、O165（以下、メジャー血清群）に属する。一方で、O69、O5、O76、O177といった血清群でも、感染者の重症化が一定数存在することが我々のこれまでの研究から明らかになっている。現在当研究所では、地方衛生研究所等から送付されたEHECについて、multiple-locus variable number tandem repeat analysis (MLVA)法やパルスフィールドゲル電気泳動 (PFGE)法といった分子型別手法を用いたサーベイランスによって、集団感染の検出や伝播経路の解明を試みている。

EHEC感染症の多くは、汚染食品が原因であると考えられるが、現在ほとんどの事例において原因食品は特定されていない。

近年実用化されつつある全ゲノム配列（whole genome sequence: WGS）をサーベイランスに用いれば、より高精度の型別および系統情報から原因食品および伝播経路の推測を行うことが可能となる。そこで本研究では、WGSを用いたEHECサーベイランスの実用化を目指し、次の3点の研究を行った：1) 菌株及び汚染食品保存中に出現する高率変異株（*mutS*欠失株）の頻度および欠失機構の解明、2) 非メジャー血清群（O69など）のEHECの完全長ゲノム配列の決定およびSNP解析手法の確立、3) WGSの自動解析パイプラインの構築。これらの研究から、効率的・高精度なEHECサーベイランスおよび地方衛生研究所等とのWGS情報共有化の仕組みを構築し、食中毒事例における原因究明および食品の安全確保に資することを目的とした。

B．研究方法

1. *mutS*欠失機構の解明

EHEC O157、O111およびO113の各血清群の1株について、単一コロニーを37°Cで一晩培養後、室温、-4°Cおよび-80°Cで保存した。これらの培地10 µlを1か月後および6か月後に採取し、段

階希釈後、LB寒天平板に塗抹した。出現コロニーを100株釣菌し、*mutS*の全長および*stx*遺伝子を対象としたPCR(表1)によって、両遺伝子の欠失を確認した。

2. 非メジャー血清群の解析

国内において重症例が報告され、EHECの病原性に重要なIII型分泌装置を有するEHEC O69 NIID150949株の完全長ゲノム配列の決定を試みた。まず、M9最小培地で37 °C、16時間培養後にQIAGEN Genomic-tip 100/G を用いてDNAを抽出した。抽出DNAをPacBio RSIIシーケンサーを用いてロングリードシーケンス解析を行った。得られたロングリード配列による *de novo* アセンブリは、Hierarchical Genome Assembly Process 3 (HGAP3)およびUnicycler v.0.4.4 (Wick et al., PLoS Comput Biol, 2017.)を用いて行った。また、国内で分離されたEHEC O69、計6株の系統解析を行うためにMiSeq (Illumina)による全ゲノムのペアエンド解析(300×2bp)を行った。得られたショートリードを用いてコアゲノムSNP解析を行った。その際、参照配列として、EHEC O157 Sakaiの完全長ゲノム配列または本研究で完全長ゲノム配列を決定したNIID150949を用いて、両者による解析結果の比較を行った。

3. WGSを用いたEHECサーベイランス法の確立

ドラフトゲノムからコアゲノムMLST (cgMLST)を簡便・迅速に行う事ができる解析パイプラインをlocal BLASTおよびPerlを使用して構築した。また、*in silico*でのribosomal MLSTおよびcgMLST解析の結果は、*de novo* アセンブリーソフトウェアによって大きく異なる。そこで、cgMLST解析に適したアセンブリーの検討を行った。EHEC O26およびO121の10

株(表2)について3種(SPAdes v.3.11.1、A5-miseq v.20160825 および Platanus v.1.2.4)のアセンブラーでコンティグを作製し、コンティグ長などの統計値を集計すると共にcgMLSTを行った。この結果を EnteroBase (<https://enterobase.warwick.ac.uk/>) によるcgMLST結果と比較した。

C. 研究結果

1. *mutS*欠失機序の解明

1か月後および6か月後の保存菌株からは、いずれの株においても*mutS*および*stx*の欠失は認められなかった。

2. 非メジャー血清群の解析

ロングリードシーケンサーから得られたリードをアセンブルした結果、完全長ゲノム配列として5.6 Mbの染色体および3種のプラスミド(それぞれ96.6, 8.3, 6.7 kb)を得ることができた(表3および図1)。国内分離株のコアゲノム SNP系統解析を行った結果、NIID150397 および NIID150949株は参照株に関わらず2か所のSNPのみ存在し、関連性の高い株であることが示唆された(表4)。一方で、他の株間は66か所以上のSNPが存在し、直接的な関連性は低いことが示唆された。また、参照株として同一血清型のNIID150949を用いた際には、O157 Sakaiを用いた際より多くのSNPが検出可能であった。

3. WGSを用いたEHECサーベイランス法の確立

3種のアセンブルによって得られたコンティグ(ドラフトゲノム)を比較した結果、コンティグ数及び最大コンティグ長はA5-miqおよびSPAdesで同等であり、Platanusでは他の2種のアセンブラーに劣っていた(表2)。総塩基長、GC割合およ

びN50の値はいずれのアsemblerも同等であった。cgMLSTの結果をアsembler別に比較したところ、SPAdesでは9株でEnterobaseによる結果と一致したのに対して、A5-miseqおよびPlatanusではそれぞれ4および0株のみの一致であった(表5)。相同配列が存在しなかったアレルの数は、SPAdesが株によって0-5であったのに対して、他のアsemblerではより多くのアレルがドラフトゲノム中に存在しなかった。

D. 考察

1. *mutS*欠失機序の解明

現時点での研究期間(6か月間)では使用した3菌株で、*mutS*遺伝子の欠失が認められなかった。今後、培養開始から1年および1年半の時点でも同様の試験を行い、同遺伝子の欠失状況を確認する予定である。一方、他の集団感染事例由来株の中には1年程度の保存で*mutS*欠失が起こることが示唆される株があり、今後そのような株でも同内容の試験を予定している。

2. 非メジャー血清群の解析

EHEC O69の1株について完全長ゲノム配列を決定した結果、コアゲノムおよび96.6kbのプラスミドはEHEC O26:H11と相同性が高く、他のプラスミドはその他のEHECのプラスミドと相同性が高いことが明らかとなった。今後、ドラフトゲノムでは全長の解析が困難なStxフェージやlocus of enterocyte and effacement領域について、他のEHECと比較解析を行うことで同血清型の病原性を明らかになることが期待される。コアゲノム SNPを用いた系統解析からは、国内分離株から遺伝的に非常に近縁な株(NIID150949およびNIID153097)が見つかった。両者のSNPの差異(2か所)は、先行研究から、同一の感染源も示唆される。両株は同一の自治

体で分離日も3日しか離れていないことから、同一の感染源が疑われる。その他の株間ではSNPの差異は大きく、直接的な関連性は低いと考えられた。また、SNP解析の際に同一血清型の株を参照株として用いることによって、より多くのSNPが検出され解析の精度を高めることができた。

3. WGSを用いたEHECサーベイランス法の確立

cgMLSTを含む、自動解析パイプライン確立のために、まずより信頼性の高いアsemblerの検討を行った。この結果、SPAdesを用いることによって、より信頼性の高いドラフトゲノムを得ることが確認できた。今後は同アsemblerを用いたドラフトゲノムを使用し、*in silico*での病原遺伝子検出やMLSTおよびrMLSTを行える解析パイプラインの確立を目指す。

E. 結論

今までに報告されていなかったEHEC O69の完全長ゲノム配列を決定したことは、WGSを用いたサーベイランスの高精度化とともに同血清型の病原性解明につながると思われる。本年度確立したcgMLST解析パイプラインは、実用的なサーベイランスツールとなることを目指し、今後病原遺伝子検出などの機能を搭載することを計画している。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

1. Ogura Y, Gotoh Y, Itoh T, Sato M, Seto K, Yoshino S, Isobe J, Etoh Y, Kurogi M, Kimata K, Maeda E, Pierard D, Kusumoto M, Akiba M, Tominaga K, Kirino Y, Kato

Y, Shirahige K, Ooka T, Ishijima N, Lee K, Iyoda S, Mainil JG, Hayashi T. 2017. Population structure of *Escherichia coli* O26:H11 with recent and repeated *stx2* acquisition in multiple lineages. Microb Genom 3.

2. Ishijima N, Lee K, Kuwahara T, Nakayama-Imahiji H, Yoneda S, Iguchi A, Ogura Y, Hayashi T, Ohnishi M, Iyoda S. 2017. Identification of a new virulent clade in enterohemorrhagic *Escherichia coli* O26:H11/H- sequence type 29. Sci Rep 7:43136.

3. Lee, K., M. Kusumoto, T. Iwata, S. Iyoda, and M. Akiba, Nationwide investigation of Shiga toxin-producing *Escherichia coli* among cattle in Japan revealed the risk factors and potentially virulent subgroups, Epidemiol. Infect., 145, 1557-1566, 2017.

2. 学会発表

(発表誌名巻号・頁・発行年等も記入)

1. 李 謙一, 伊豫田 淳, 小椋 義俊, 林 哲也, 大西 真, EHEC Working Group. 腸管出血性大腸菌O115における国内分離株の系統解析および病原性評価. 第91回日本細菌学会総会. 2018年3月. 福岡.

2. 石嶋 希, 李 謙一, 勢戸 和子, 大西 真, 伊豫田 淳. 混合感染が確認された HUS 症例2例から分離された腸管出血性大腸菌の性状解析. 第91回日本細菌学会総会. 2018年3月.福岡.

3. 李 謙一, 伊豫田 淳., 泉谷秀昌, 大西 真, EHEC working group,. 腸管出血性大腸菌O157サーベイランスにおける反復配列多型解析と全ゲノム配列解析の分子型別能の比較. 第21回腸管出血性大腸菌感染症研究会. 2017年11月. 鹿児島.

4. 木全恵子, 磯部順子, 綿引正則, 勢戸和子, 尾畑浩魅, 小西典子, 李 謙一, 伊豫田 淳, 大西 真. LEEを保有しない腸管出血性大腸菌のゲノム解析. 第21回腸管出血性大腸菌感染症研究会. 2017年11月. 鹿児島.

5. 伊豫田 淳, 李 謙一, 石嶋 希, 勢戸和子, 齊藤剛仁, 大西 真. HUS発症例における血清診断とEHECの分離同定について. 第21回腸管出血性大腸菌感染症研究会. 2017年11月. 鹿児島.

6. 小原敦美, 松本一俊, 近藤ひとみ, 原田誠也, 大迫英夫, 李 謙一, 伊豫田 淳, 大西 真. HUS患者から分離された EHEC O76:H7(*stx2* 陽性)について. 第21回腸管出血性大腸菌感染症研究会. 2017年11月. 鹿児島.

7. 李 謙一, 伊豫田 淳, 泉谷秀昌, 大西 真, EHEC Working Group. 腸管出血性大腸菌O157における反復配列多型解析と全ゲノム配列解析の比較. 第160回日本獣医学会学術集会. 2017年9月. 鹿児島.

H . 知的財産権の出願・登録状況

なし

表1. *mutS* 欠失頻度測定に用いたプライマー

Primer	Target gene	Sequence (5' → 3')	Amplicon size (bp)	Reference
mutS-F	<i>mutS</i>	ATGAGTGCAATAGAAAATTT	951 ^a	this study
mutS-R1		ATCGCGCACTGGCATATGCA		this study
mutS-F1		ACCCGCGTGTGCTTGAGCG	1611	this study
mutS-R		TTACACCAGACTCTTCAAGCGATAAA		this study
LP30	<i>stx1</i>	CAGTTAATGTGGTGGCGAAGG	348	Cebula et al., J. Clin. Microbiol., 1995, p248-250
LP31		CACCAGACAATGTAACCGCTG		
LP43	<i>stx2</i>	ATCCTATTCCCGGGAGTTTACG	584	1995, p248-250
LP44		GCGTCATCGTATACACAGGAGC		

^aExpected size in EHEC O157 Sakai

表2. EHEC O69 NIID150949株 コンプリートゲノム配列の概要

Statistics	Chromosome	Plasmid 1	Plasmid 2	Plasmid 3
Total Length (bp)	5,587,898	96,568	8,284	6,673
GC Content (%)	50.7%	48.0%	44.0%	50.2%
N50	5,587,898	96,568	8,284	6,673
No. of CDSs	5,404	99	8	8
No. of rRNA	22	0	0	0
No. of tRNA	98	0	0	0
No. of CRISPRS	2	0	0	0
Coding Ratio (%)	86.70%	78.70%	46.10%	61.80%

表3. 参照配列にEHEC O157 SakaiおよびO69 NIID150949を用いた際のcore genome中のpairwise SNP数

参照株	菌株名	No. of Pairwise SNP				
		NIID120784	NIID132128	NIID153097	NIID160444	NIID110900
O157 Sakai						
	NIID120784					
	NIID132128	211				
	NIID153097	168	239			
	NIID160444	65	226	181		
	NIID110900	66	235	198	95	
	NIID150949	166	237	2	179	196
NIID150949						
	NIID120784					
	NIID132128	1437				
	NIID153097	1394	753			
	NIID160444	1028	501	454		
	NIID110900	1052	537	504	138	
	NIID150949	1392	751	2	452	502

表4. アセンブラー別のドラフトゲノム統計量の比較

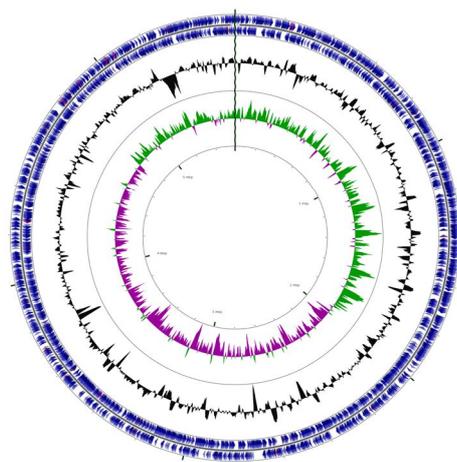
Strain	Statistics	a5-miseq	Platanus	SPAdes
NIID080884	No. of contigs	240	949	298
	Largest contig	233,668	74,228	230,709
	Total length	5,774,041	5,426,394	5,640,933
	GC (%)	50.5	50.7	50.4
	N50	85,581	14,860	96,948
NIID112183	No. of contigs	266	520	291
	Largest contig	239,463	151,002	230,459
	Total length	5,788,789	5,587,039	5,697,461
	GC (%)	50.4	50.3	50.4
	N50	63,142	38,866	104,236
NIID131037	No. of contigs	285	918	279
	Largest contig	224,285	70,162	274,329
	Total length	5,546,492	5,153,567	5,427,786
	GC (%)	50.6	50.8	50.5
	N50	59,185	12,636	91,514
NIID141341	No. of contigs	170	463	266
	Largest contig	425,834	160,850	426,152
	Total length	5,390,950	5,158,973	5,281,975
	GC (%)	50.5	50.4	50.5
	N50	138,303	54,581	134,860
NIID141423	No. of contigs	220	1,138	244
	Largest contig	266,864	48,488	224,910
	Total length	5,536,382	5,202,592	5,424,078
	GC (%)	50.5	51.0	50.5
	N50	102,360	9,603	105,136
NIID141424	No. of contigs	205	810	245
	Largest contig	278,076	65,675	282,947
	Total length	5,620,003	5,285,318	5,492,102
	GC (%)	50.5	50.6	50.4
	N50	95,769	18,116	112,155
NIID150174	No. of contigs	161	422	218
	Largest contig	322,740	129,437	319,751
	Total length	5,371,528	5,166,136	5,265,524
	GC (%)	50.4	50.4	50.4
	N50	127,612	44,439	134,860
NIID150373	No. of contigs	144	355	223
	Largest contig	436,921	229,130	435,274
	Total length	5,365,043	5,292,171	5,260,459
	GC (%)	50.4	50.4	50.4
	N50	161,233	84,618	134,860
NIID150400	No. of contigs	148	357	217
	Largest contig	442,834	257,870	434,980
	Total length	5,324,964	5,255,106	5,220,650
	GC (%)	50.5	50.5	50.5
	N50	148,043	97,486	134,860
NIID150616	No. of contigs	151	447	226
	Largest contig	435,449	152,402	435,193
	Total length	5,356,998	5,192,610	5,259,933
	GC (%)	50.4	50.4	50.4
	N50	161,602	35,235	134,860

表5. アセンブラー別のcgMLST結果の比較

Strain	ST in EnteroBase	a5-miseq		Platanus		SPAdes	
		ST	No. of null ^a	ST	No. of null	ST	No. of null
NIID080884	44924	44924	0	- ^b	23	44924	0
NIID112183	44920	-	164	-	18	44920	1
NIID131037	44911	-	1	-	32	-	3
NIID141341	46459	-	21	-	9	46459	1
NIID141423	44928	44928	5	-	58	44928	5
NIID141424	44914	44914	0	-	15	44914	0
NIID150174	46487	-	19	-	6	46487	1
NIID150373	46499	46499	1	-	17	46499	1
NIID150400	46502	-	2	-	9	46502	1
NIID150616	46492	-	8	-	14	46492	1

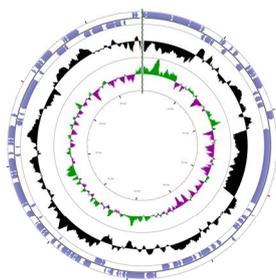
^aNumber of null alleles, ドラフトゲノム中で相同な配列が存在しなかったアレルの数

^bNo known ST was assigned

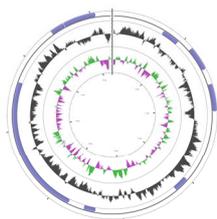


Chromosome: 5.6 Mbp

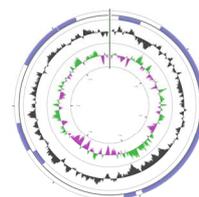
- CDS
- tRNA
- rRNA
- Other
- GC content
- GC skew+
- GC skew-



Plasmid1: 96.6kbp



Plasmid2: 8.3kbp



Plasmid2: 6.7kbp

図1. EHEC O69 NIID150949株の完全長ゲノム配列

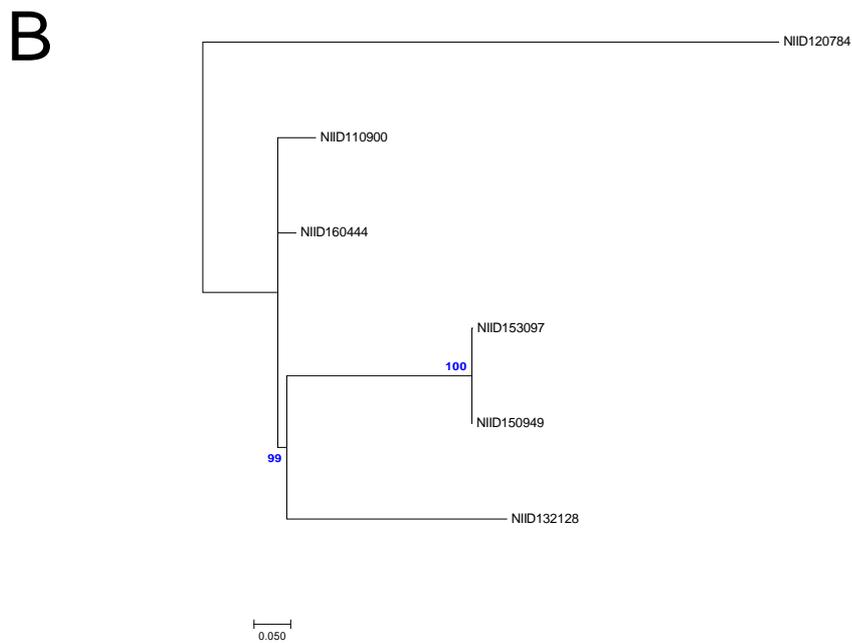
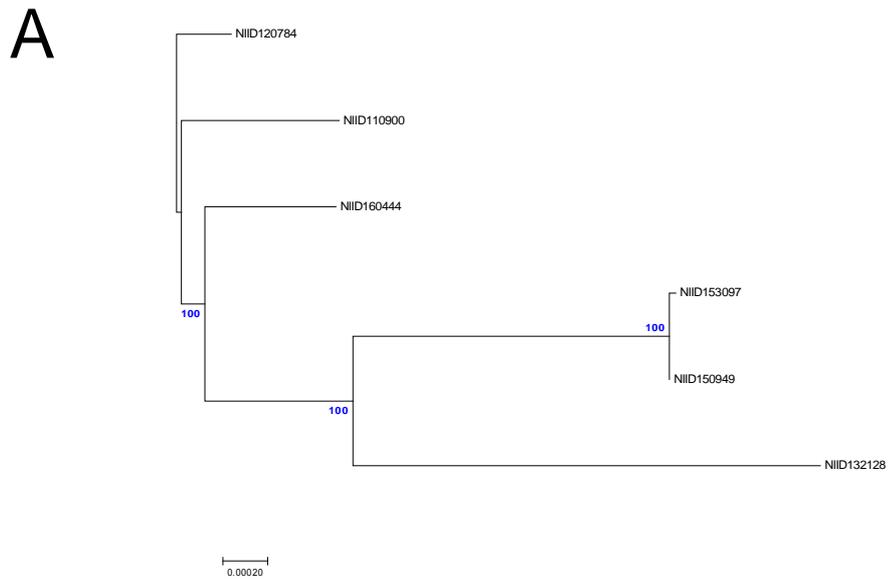


図2. EHEC O69 国内分離株のcore genome SNPによる最尤法系統樹. (A)参照配列にEHEC O157 Sakaiを用いた場合および(B)参照配列にEHEC O69 NIID150949を用いた場合. 図中の数字は1000回試行時のブートストラップ値を示す.