

ゲノム解析を用いた
下水中溶血性細菌の分類に関する研究

池端建吾

概要

病原性細菌が有する溶血性毒素は、様々な疾患の増悪に関係している。例えば大腸菌の溶血毒 α ヘモリシン遺伝子(*hlyA*)の有無と、尿路感染症、結腸がん、潰瘍性大腸炎との関係が示唆されている。しかしながら、今後、溶血細菌と様々な疾患の関係を分子疫学的に解明するためには、溶血毒素遺伝子の有無だけではなく、溶血活性に関わる様々な遺伝マーカーを整備する必要がある。また、ヒトが曝露しうる溶血性細菌の網羅的リストの整備も必要である。都市下水中の細菌を調べることは、ヒト腸内細菌調査のサロゲートとして利用できるだけでなく、河川や湖沼などの水環境への排出源の調査としても重要である。本研究では、改良した高感度な血液寒天培地法を用いて、都市下水中の病原性細菌の検出と同定、さらに遺伝的特徴の解析を行った。

滋賀県の都市下水から血液寒天プレートによって明瞭な溶血斑を示す 77 株、不明瞭な溶血斑を示す 55 株、コロニー周辺が変色する 5 株を得た。これら全ての菌株について 16S rRNA 遺伝子解析を行い、一部の菌株については全ゲノムシーケンス解析を行った結果から、明瞭な溶血斑を示す株は病原性細菌としても知られる *Aeromonas* 属と大腸菌であることが分かった。不明瞭な溶血斑を示した株は、敗血症等の稀な感染症の報告のある *Streptococcus parasuis*、*Streptococcus lutetiensis*、*Streptococcus lactis*、*Leuconostoc lactis*、*Enterococcus casseliflavus* などが同定された。また、これまで溶血活性の報告がなかった *Ligilactobacillus salivarius*、*Lactococcus taiwanensis*、*Enterococcus lemanii* なども同定された。

Aeromonas 属については、大きな溶血斑を示す株と、小さな溶血斑を示す株の 2 種類に大別された。これらの違いは 16S rRNA 遺伝子アレルのパターンにより明確に区別することが可能であった。全ゲノムシーケンス解析の結果、溶血活性の大きい多型は *Aeromonas hydrophila* と同定され、溶血活性の小さな多型は *Aeromonas caviae* と同定された。*Aeromonas hydrophila* のもつ病原性遺伝子を解析したところ、*Aeromonas caviae* が保有しないエロリジン遺伝子 *aerA* や細胞外ヘモリシン遺伝子 *ahh1* を保有しており、これら遺伝子が強い溶血活性と関係していることが明らかとなった。

一方、*hlyA* 遺伝子を有する大腸菌についても、菌株により大きな溶血斑を示すものと小さな溶血斑を示すものが存在した。これら菌株の全ゲノムシーケンスに基づき、MLST 解析、FimH 分類、血清型分類、phyloptype 分類を行った。これら大腸菌株は全て、phyloptype B2 群に所属しており、ST95、FimH18、H7 血清型のグループ、ST73、H1 血清型のグループ、ST127、O6:H31 血清型のグループに分類され、溶血性大腸菌の遺伝的な特徴が明らかとなった。さらに溶血毒 α ヘモリシン産生に関わる遺伝子の SNP 解析をするため、SNP 解析ソフトウェア MMViewer (<https://github.com/KIkebata/mmviewer>) を開発した。MMViewer を用いた解析の結果、*hlyA* の SNP (c.841G>T : p.Gly281Cys) や *hlyCABD* オペロンの転写伸長因子 RfaH をコードする *rfaH* 遺伝子の SNP (c.50C>G : p.Ala17Gly) 等の SNPs が大腸菌の溶血活性に影響を及ぼしている可能性が示唆された。

目次

1. 序論	1
1.1. 都市下水から病原体を検出する意義.....	1
1.2. 細菌の検出手法.....	2
1.3. 溶血性細菌の病原性.....	3
1.4. 研究目的	5
1.5. 全ゲノムショットガンシーケンス法に基づく遺伝子保有の解釈.....	5
2. 方法	8
2.1. 実験で使用した培地.....	8
2.2. 下水に含まれる溶血性細菌の単離.....	8
2.3. 血液寒天プレート上における溶血活性の測定.....	8
2.4. 液体培地培養時の上清の溶血活性の測定、上清タンパク分析、大腸菌の <i>hlyA</i> mRNA の測定	10
2.5. 単離株の 16S rRNA 遺伝子 V3-V4 領域のシーケンシング	11
2.6. 16S rRNA 遺伝子 V3-4 領域の解析による単離細菌株の種の推定	12
2.7. 16S rRNA 遺伝子 V3-4 領域のアレル解析.....	12
2.8. 単離株の全ゲノムシーケンスとリードの前処理.....	12
2.9. 全ゲノムシーケンスを行った株の種の推定・同定.....	13
2.10. 大腸菌の分類.....	13
2.11. 大腸菌の各病原型に特徴的な病原遺伝子保有の確認.....	14
2.12. kSNP3.0 による系統解析	14
2.13. Missense Mutation Viewer (MMViewer) を用いた SNP 解析.....	15
2.14. 大腸菌以外の溶血性細菌の保有する溶血毒の予測.....	16
3. 結果	18
3.1. 血液寒天培地によって単離される溶血性細菌.....	18
3.1.1. 溶血性細菌の検出・単離.....	18
3.1.2. 溶血活性の測定.....	18
3.1.3. 液体培地を用いた既存の溶血活性測定手法との比較.....	19
3.1.4. 16S rRNA 遺伝子解析による単離株の種の推定	20

3.1.5.	16S rRNA 遺伝子 V3-4 領域のアレルと溶血活性.....	24
3.2.	溶血性大腸菌.....	31
3.2.1.	血液寒天培地で単離される大腸菌の遺伝的特徴.....	31
3.2.2.	溶血性大腸菌の溶血活性の決定因子.....	48
3.3.	大腸菌以外の溶血性細菌.....	60
3.3.1.	溶血斑の再確認.....	60
3.3.2.	種の同定.....	62
3.3.3.	保有する病原毒素遺伝子.....	64
4.	考察.....	65
4.1.	溶血性大腸菌の遺伝型と保有する病原性遺伝子.....	65
4.2.	下水中の溶血性細菌種と病原性毒素.....	68
4.3.	溶血性細菌の溶血活性決定因子.....	72
4.4.	血液寒天プレートの選択培地としての有用性と課題.....	74
5.	結論.....	76
	参考文献.....	77
	付録.....	92
1.	資料.....	92
1.1.	<i>Aeromonas</i> 属の 16S rRNA V3-V4 領域のアレル配列.....	92
1.2.	<i>Brenneria alni</i> 、 <i>Escherichia coli</i> 、 <i>Escherichia fergusonii</i> 、 <i>Shigella flexneri</i> 、 <i>Shigella sonnei</i> と推定された株の 16S rRNA V3-V4 領域のアレル配列.....	97
2.	図表.....	99
3.	Missense Mutation Viewer (MMViewer).....	101
	対象 CDS 領域の検索 : <code>get_target</code>	101
	アラインメント : <code>alignment</code>	102
	グラフの描画 : <code>gen_graph</code>	103
	ソースコード.....	105
	<code>__init__.py</code>	105
	<code>__main__.py</code>	105
	<code>__version__.py</code>	106
	<code>get_target.py</code>	106

alignment.py	110
gen_graph.py.....	111
utility.py	122

図表目次

図 1 (i) アラインメントに基づくアSEMBルと (ii) de novo アSEMBルに基づくアSEMBルと作成されるコンティグ.....	7
表 1 溶血斑を測定する際に使用した血液寒天プレート.....	9
図 2 溶血斑の撮影に使用した装置の概略.....	9
図 3 溶血斑幅の計算に使用した数値の計測項目.....	10
図 4. 16S rRNA 遺伝子 V3-4 領域のシーケンシングに使用したアンプリコン PCR プライマー.....	11
表 2 kmer ID を用いた解析で使用したリファレンス株.....	13
表 3 大腸菌の各病原型に特徴的な病原遺伝子のリファレンス配列.....	15
表 4 MMViwer を用いた SNP 解析で使用したリファレンス配列.....	17
図 5 血液寒天培地と下水流入水を混合して一晩培養したときの様子.....	18
図 6 単離した溶血性細菌の溶血斑の様子.....	19
図 7 (i) 液体培地で培養時の溶血活性を 450nm の波長の吸光度で測定して得られた溶血活性値 (percent hemolysis; A450) と血液寒天培地で測定した溶血活性 (hemolysis width) の関係、(ii) 液体培地で培養時の溶血活性を 595nm の波長の吸光度で測定して得られた溶血活性値 (percent hemolysis; A595) と血液寒天培地で測定した溶血活性 (hemolysis width) の関係.....	20
表 5 単離した溶血性細菌の 16S rRNA 遺伝子 V3-4 領域から推定した種とその数.....	23
図 8 単離細菌株の 16S rRNA 遺伝子 V3-4 領域の系統樹、推定種および明瞭な溶血斑を示す株の溶血活性の大きさ.....	24
表 6 それぞれの細菌種がゲノム上に保有する 16S rRNA 遺伝子の数.....	25
図 9 <i>Aeromonas</i> 属に分類された株の 16S rRNA 遺伝子アレルの構成割合 [%] とアレルの分子系統樹および単離株のアレル構成に基づくクラスタリング.....	27
表 7 <i>Aeromonas</i> 属細菌株の各アレルから推定された種.....	28
表 8 <i>Brenneria alni</i> 、 <i>Escherichia coli</i> 、 <i>Escherichia fergusonii</i> 、 <i>Shigella flexneri</i> 、 <i>Shigella sonnei</i> に分類された株の各アレルから推定された種.....	29
図 11 WWo108 の各血液寒天培地上における溶血斑の様子.....	33
表 9 全ゲノムシーケンスの対象株.....	34
表 10 単離株と大腸菌および赤痢菌の ANI 値.....	35
表 11 kmer ID で同定された単離株の種と MLST 解析の結果.....	36

図 12	kSNP3.0 を用いた系統解析の結果と、溶血活性の大きさ、 <i>hlyCABD</i> オペロンの保有、MLST 解析による分類、 <i>fimH</i> 分類、血清型 (serotype)、 <i>phylo</i> type の分類の結果	38
表 12	大腸菌の各病原型に特徴的な遺伝子 ⁸⁷ と UniProt データベース上で対応するタンパク質	43
図 13	大腸菌の各病原型に特徴的な遺伝子の塩基配列にアラインメントしたときの gene coverage	44
図 14	大腸菌のシーケンス解析の結果得られたコンティグと病原性遺伝子の塩基配列の一致率 [%] (下)、および病原性遺伝子の保有パターンのクラスター分析の結果 (上)	46
表 13	病原性遺伝子とその転写産物	47
図 15	液体培地で 2 時間培養した上清の溶血活性	49
図 16	上清の溶血活性の大きさと <i>hlyA</i> mRNA の相対発現量の関係	49
図 17	大腸菌培養液上清をポリアクリルアミド電気泳動 (SDS-PAGE) した結果	50
図 18	MMViewer の解析フローチャート	52
図 19	<i>hlyCABD</i> オペロン内のミスセンス変異とその上流 2000 bp の変異	55
図 20	<i>hlyE</i> とその上流 2000 bp の変異	55
図 21	<i>hlyF</i> とその上流 2000 bp の変異	56
図 22	α ヘモリシンの溶血活性に関わる遺伝子のミスセンス変異 (1/3)	57
図 23	α ヘモリシンの溶血活性に関わる遺伝子のミスセンス変異 (2/3)	58
図 24	α ヘモリシンの溶血活性に関わる遺伝子のミスセンス変異 (3/3)	59
図 25	血液寒天プレートタイプ III で確認した溶血斑の様子	61
表 14	16S rRNA V3-4 領域から推定した種と rMLST で推定した種	63
表 15	fastANI の結果	63
表 16	VFanalyzer ⁷¹ で予測された毒素遺伝子の保有と prokka ⁷² でヘモリシンとしてアノテーションされた遺伝子	64
表 S 1	VirulenceFinder2.0 で BLAST 検索を行いヒットした病原性遺伝子と、クエリー長/テンプレート長 (1/4)	99
表 S 2	VirulenceFinder2.0 で BLAST 検索を行いヒットした病原性遺伝子と、クエリー長/テンプレート長 (2/4)	99
表 S 3	VirulenceFinder2.0 で BLAST 検索を行いヒットした病原性遺伝子と、クエリー長/テンプレート長 (3/4)	99

表 S 4 VirulenceFinder2.0 で BLAST 検索を行いヒットした病原性遺伝子と、クエリー 長 / テンプレート長 (4/4)	100
図 S 1 MMViewer で出力されるグラフの概観.....	104

1. 序論

1.1. 都市下水から病原体を検出する意義

病原体を検出することは、公衆衛生にとって非常に重要であり、人々の健康を守る上で欠かせない。1876年、細菌学の父とも呼ばれるドイツの細菌学者、ロバート・コッホが、炭疽病が炭疽菌によって引き起こされることを発見し、細菌が病気の原因になることを明らかにした¹。その後、様々な病気が細菌によって引き起こされることが明らかとなってきた。感染症を引き起こす病原性細菌を検出することは多くの場面で重要で、例えば病院では、感染症を発症した患者から原因の病原性細菌を単離・同定することで適切な治療を行えるようになる。食中毒が発生した際には、食品中の食中毒細菌を検出することで食中毒の原因を明らかにすることができる。家畜や野生動物の糞便等からヒトに対する病原性細菌を検出することで、家畜や野生動物を中間宿主としてヒトに感染する感染症の感染源を特定することができる。

近年では、都市下水に含まれる病原体を検出することで、下水道に接続されているコミュニティの健康状態や市中感染の調査に役立てる試みが盛んに行われている。下水は、コミュニティを構成する多数の人々から排泄された糞便を含み、コミュニティに蔓延する病原体、環境曝露または毒性物質のレベルと量、使用されている様々な薬物の量、食事パターンなど様々な情報を提供する。下水から得られるデータは、下水道に接続されている人々の代表値である点、匿名性を持ってデータを収集できる点、低コストである点から、集団レベルの健康データを収集するのに適している²。このように下水をコミュニティレベルでの健康や病気のモニタリングに利用するというアイデアは古くから存在しており、現在では「下水疫学 (wastewater-based epidemiology)」という分野として確立している²。古くは1940年代、アメリカで小児麻痺の原因となるポリオウイルスを下水中から検出し、小児麻痺の症例報告と下水中ウイルスの検出頻度に弱い相関があることが報告された³。現在でも、排水監視によるポリオの疫学調査はポリオの撲滅に重要な手段として用いられている⁴。2019年に発生した新型コロナウイルス感染症 (COVID-19) の原因である SARS-CoV-2 ウイルスを下水から検出し、それを公衆衛生に役立てる試みは多くの国で行われ、下水疫学の分野を著しく成長させた²。さらに、下水道に含まれる病原体をモニタリングする疫学的アプローチは、生物兵器として利用される病原体の早期警告にも適している⁵。

都市下水に含まれる病原体は環境への排出源としても重要である。下水処理場で処理された水は最終的に塩素処理をはじめとする微生物の不活化プロセスを経て自然水に放流される。しかしながら、塩素処理された二次処理水からも *Bacteroides* や *Arcobacter*、*Clostridium* 等の腸管常在菌が検出されており⁶、病原性細菌が塩素耐性を獲得した場合これらの細菌が環境中に放出される可能性が考えられる。ヒトの体内、とくに腸内には様々な細菌が腸内細菌叢を形成しており、環境中に比べ多くの病原性細菌が存在する。実際、糞便によって汚染された水環境は、ほかの土壌環境などに比べて多くの病原性細菌が検出されることが報告されている⁷。

近年、地球温暖化とそれに伴う気候変動によって、これまでになかった病原体が出現する可能性が示唆されている。すでに現在、コレラ、マラリア、デング熱等のベクター媒介性の感染症を中心に、世界の小児疾患の34%、幼児死亡の36%は、環境要因の変化と関連していると推定されている⁸。また、地球温暖化に伴い何千年前から存在する氷の層が溶

解することで、凍った土壌に保存されている数千年前の細菌がよみがえり新たな感染症が引き起こされる可能性が危惧されている。実際、Bidle ら (2007) はアラスカで 3 万 2000 年前から凍結されていた細菌を蘇らせることに成功し⁹、Katayama ら (2007) はシベリアで 2 万 5000 年前から凍結されていた細菌を蘇らせることに成功しており、氷層から未知の病原性細菌が放出される可能性がある¹⁰。気候変動の影響で、すでにシベリアでは永久凍土に保存されていた人獣共通病原性微生物の炭疽菌が大気中に放出され、トナカイの大量死やヒトへの感染が報告されている^{11,12}。

時間の経過とともに新たに発生・増加する新興感染症や、未知の病原体や別の地域で流行していた病原体がこれまでに流行していなかった地域で新たに発生する再興感染症は、気候変動によって加速することが予想される。糞便汚染指標である大腸菌群数や大腸菌数を測定するだけでなく、環境中から直接これらの病原性細菌を検出し、新興感染症や再興感染症に対するリスク管理を行うことが望まれる。

1.2. 細菌の検出手法

既知の細菌を検出するのに PCR 法はよく用いられる手法である。PCR 法では、病原性細菌に固有の DNA 配列を増幅するプライマーを使用し、対象の病原性細菌の有無を高い感度と特異性をもって低コストで検出することができる。培養することが難しい株を検出できる上、培養にかかる時間も必要としないため、病院での検査だけでなく環境中の病原性細菌を検出する際にも広く用いられている¹³。しかしながら、PCR 法で病原性細菌を検出するにはいくつか欠点がある。最も大きな欠点は、未知の病原性細菌については検出することができない点である。PCR 法で病原性細菌を検出できるようにするために、病原性細菌の網羅的な解析と病原性に関わる分子生物学的マーカーの解明が望まれる。

未知の細菌を検出する手法として、現在はメタゲノム解析が広く用いられている。メタゲノム解析では、サンプル中の DNA を網羅的に解析することで、細菌集団を単離、培養することなく、培養の難しい細菌を含めた細菌集団全体から細菌を検出することができる。メタゲノム解析で明らかになった細菌種の 8 割は培養のできない株であるという報告もあり¹⁴、メタゲノム解析が非常に有用なツールであることは間違いない。しかしながら、メタゲノム解析においても欠点があり、例えばメタゲノム解析で用いるプライマーの感度が低いために優占種しか検出することができず、 10^5 CFU/g 未満の濃度で存在する細菌を検出できないことが報告されている¹⁵。この問題は「深度バイアス」として知られ、疫学的に重要な少数集団を見逃してしまう可能性がある。

近年では、メタゲノム解析では検出できない少数派の細菌を培養ベースの手法で網羅的に検出する「カルチュロミクス (culturomics)」と呼ばれる新たな細菌の検出手法が提案されている¹⁶。カルチュロミクスは、様々な培養条件で網羅的に細菌を培養することで細菌を増殖させて同定する手法で、メタゲノム解析では分からない細菌の少数集団 (10^5 CFU/g 未満) も検出することができる¹⁵。培養、単離することで単離細菌を用いた後段の実験に使用することができ、未知の細菌に対して薬剤耐性の有無の確認や全ゲノムシーケンスが行える。これらの利点から、培養ベースの手法は未知の細菌の少数集団を検出する手法として再び注目を集めている。

培養ベースの手法では、選択培地を使用することで目的の細菌のみを選択的に増殖させることもできる。選択培地は、基本培地に抗生物質や防腐剤、ナトリウム塩、化学物質、ファージ、染料などを追加して、目的の細菌集団以外の増殖を抑制、あるいは目的の細菌

集団を染色することで、目的の細菌を検出、単離することができる¹⁷。抗生物質を加えた培地は最も一般的に使用される選択培地で、近年注目の集まる薬剤耐性菌を検出するために用いられている。培地に防腐剤が使用されることは少ないが、結核菌 (*Mycobacterium tuberculosis*) の選択にクロルヘキシジンを加えた培地が用いられることや¹⁸、クロストリジウム・ディフィシル (*Clostridioides difficile*) などの芽胞形成細菌種の選択にエタノールが用いられることがある^{19,20}。好塩性細菌の選択培養には高濃度の塩化ナトリウムを加えたものが用いられる²¹。グラム陽性菌の増殖を抑制するため亜テルル酸カリや胆汁酸塩を加えたものもある¹⁷。また、ファージの持つ細胞溶解酵素を加えることで、喀痰サンプルから結核菌を選択的に検出する手法が提案されている²²。培地に色素を加えたものではXM-G培地等が知られる。糞便汚染指標として疫学的に重要な大腸菌群と大腸菌のコロニーがそれぞれ染色され、それぞれのコロニー数を同時に計数できることから、食品や環境中の大腸菌、大腸菌群の検出に広く用いられている

1.3. 溶血性細菌の病原性

溶血毒は、赤血球の細胞膜に孔をあける毒素で、様々な病原性細菌が保有している。溶血毒を産生する細菌の溶血活性は、血液寒天培地上での溶血の種類により、一般的に β 溶血（赤血球を完全に溶解）、 α 溶血（不完全な溶血）、 γ 溶血（溶血しない）に分けられる。赤血球を完全に溶解する β 溶血性細菌は一般的に病原体として考えられており、A群レンサ球菌 (*Streptococcus pyogenes*)、大腸菌 (*Escherichia coli*)、セレウス菌 (*Bacillus cereus*)、黄色ブドウ球菌 (*Staphylococcus aureus*)、ウェルシュ菌 (*Clostridium perfringens*)、リステリア菌 (*Listeria monocytogenes*)、コレラ菌 (*Vibrio cholerae*)、腸炎ビブリオ (*Vibrio parahaemolyticus*) の β 溶血を示すものは病原性との関連が示唆されている²³。

大腸菌の産生する α ヘモリシンは最もよく知られた溶血毒の一つで、血液寒天培地上で β 溶血を示し、大腸菌の病原性を増強させる。 α ヘモリシンを産生しない大腸菌に α ヘモリシン産生能を付加すると、ラットにおける大腸菌腹膜炎の致死率が高まる²⁴。また、マウスの腎臓感染モデルでは、 α ヘモリシンの働きにより腎障害が誘発され、腎臓に誘導されるマクロファージが増加する²⁵。女性の結腸腺がん患者糞便から単離した大腸菌には溶血活性を示す株割合が高いことも報告されている²⁶。また、動物を用いた実験では発がん性物質アゾキシメタン (AOM) で処理したメスのマウスに対して α ヘモリシン産生大腸菌を投与することで結腸腺がんの発症率を高め、長期にわたる実験では α ヘモリシン産生大腸菌の投与のみで結腸腺がんを発症することが報告されている²⁶。活動期の潰瘍性大腸炎患者においても、大腸粘膜における α ヘモリシン遺伝子 *hlyA* の量が健常者やクローン病患者と比べて約10倍高い濃度で存在する²⁷ことから、 α ヘモリシンとの関係性が示唆されている。

A群レンサ球菌 (*Streptococcus pyogenes*) は臨床的に重要な病原性細菌で、その初期のスクリーニングでは血液寒天培地上で溶血活性が調べられる。レンサ球菌は、グラム陽性、カタラーゼ陰性、コアグララーゼ陰性の球菌で、血液寒天培地上では種によって異なる溶血活性によって、 α 溶血性レンサ球菌、 β 溶血性レンサ球菌、 γ 溶血性レンサ球菌に分けられる。 β 溶血性レンサ球菌はさらにA群レンサ球菌 (*Streptococcus pyogenes*) とB群レンサ球菌 (*Streptococcus agalactiae*) に分類される。A群連鎖球菌は化膿レンサ球菌とも呼ばれ、溶菌菌感染症（急性咽頭炎）や、膿痂疹や蜂巣炎などの皮膚・軟部組織感染症を引き起こし²⁸、臨床的に重要な種である。A群レンサ球菌の主要な溶血毒はストレプトリ

シンOである²³。ストレプトリシンOは、マウスモデルを用いた実験から、壊死性筋膜炎、菌血症、皮膚・軟部組織感染における病原性に大きく寄与することが示唆されている²⁹。

セレウス菌 (*Bacillus cereus*) は、土壌、植物、食品などに存在する通性嫌気性の毒素産生グラム陽性菌で、吐き気、嘔吐、下痢等の症状を伴うセレウス菌食中毒と関連しており、敗血症や視力低下につながる眼内炎を引き起こす可能性もある³⁰。セレウス菌の主要な溶血毒はホスファチジルコリン特異的ホスホリパーゼCとスフィンゴミエリナーゼで、そのほかに Alveolysin、溶血性エンテロトキシン、エンテロトキシンBを保有する^{23,31}。黄色ブドウ球菌 (*Staphylococcus aureus*) は、健康な人口の約30%に存在する常在菌であると同時に病原性細菌でもある。菌血症、感染性心内膜炎、皮膚・軟部組織感染症（膿痂疹、毛囊炎、癬、カーバンクル、蜂巣炎、やけど皮膚症候群など）、骨髄炎、敗血症性関節炎、人工関節感染症、肺感染症（肺炎、肺水腫など）、胃腸炎、髄膜炎、毒素性ショック症候群、尿路感染症など、さまざまなヒトの感染症の原因菌となっている³²。溶血毒として、 α ヘモリシン、 γ ヘモリシン、 β ヘモリシン、 δ ヘモリシンを保有する³³。ウェルシュ菌 (*Clostridium perfringens*) は、食中毒、ガス壊疽、蜂巣炎、筋膜炎など様々な感染症の原因となる³⁴。ウェルシュ菌による敗血症は稀であるが、溶血毒 α 毒素が赤血球を損傷して溶血を起こすことが知られており、救急外来では急速に死に至ることが多い³⁵。リステリア菌 (*Listeria monocytogenes*) はリステリア症と呼ばれる食中毒の原因となることが知られ、髄膜炎や敗血症、流産等を引き起こす³⁶。コレラ菌 (*Vibrio cholerae*) は、重度の下痢症を引き起こすコレラの原因菌である³⁷。一部の株は溶血毒HlyAにより溶血活性を示すものがある³⁸。腸炎ビブリオ (*Vibrio parahaemolyticus*) は、河口域、海洋、沿岸環境に生息する好塩性のグラム陰性細菌で、十分に加熱されていない水産物を摂取したときに急性胃腸炎を引き起こすことが知られる。患者から単離されたほぼすべての腸炎ビブリオ (*Vibrio parahaemolyticus*) は熱安定性ダイレクトヘモリシン (TDH) やTDH関連ヘモリシン (TRH) に起因する β 溶血活性を有する³⁹。

多くの病原性細菌が溶血活性を持つことが知られるが、レンサ球菌や腸炎ビブリオのスクリーニング以外に血液寒天培地が選択培地として用いられることはあまり多くない。大きな理由としては、血液寒天プレートで溶血性細菌を培養するのに時間を必要とすることが挙げられる。食中毒などの感染症に対する治療を行う際には、できるだけ早く病原菌を同定する必要がある。培養法、特に血液寒天培地を用いて溶血活性を確認するには数日かかることがあり、現在ではPCR法等の培養によらない素早い識別法が確立しているため、培養ベースの同定法に置き換わって用いられている。また、血液寒天培地が検出する細菌種のスペクトルが広いことも、識別に用いることが難しい理由として考えられる。溶血毒は多くの細菌が保有する一般的な毒素であり、血液寒天培地を用いると溶血活性を示す多くの種が検出されると考えられる。また、大腸菌のように同じ種でも溶血活性を示さない株があり、血液寒天培地ですべての病原性細菌を検出することは難しく、感度の低さも理由として考えられる。

しかしながら、血液寒天培地は「溶血毒」という病原性因子を保有する病原性細菌を幅広く検出できることから、環境中から病原性細菌をスクリーニングするための選択培地として利用できると考えられる。環境中で血液寒天培地を適応した例としては、極地の細菌に対して血液寒天培地を用いて病原性細菌の単離を試みた例⁴⁰はあるが、環境中で血液寒天培地を選択培地として用い、病原性細菌を網羅的に単離した例はまだあまり多くない。

血液寒天培地は病原性細菌を単離するだけでなく、病原性因子である溶血毒の活性を直接測定できる。血液寒天培地はほかの選択培地とは異なり、溶血毒をコロニー周囲が透明になる溶血斑として示される。溶血斑の大きさは株によってことなり、溶血毒の種類や活性、分泌量によって変化すると考えられる。大腸菌の産生する α ヘモリシンをはじめとする多くの溶血毒は、溶血毒そのものが病気に関わる因子として働くことから、溶血斑の大きさは細菌の種類や毒性を示す指標として考えることができる。

1.4. 研究目的

本研究では、血液寒天培地を用いて都市下水から病原性細菌を網羅的に単離し、単離される細菌の遺伝的特徴を明らかにすることを目的とした。また、単離された溶血性細菌の溶血活性と遺伝的特徴の関係についても明らかにすることとした。

まず都市下水から病原性細菌を血液寒天プレートによって単離した。単離した溶血性細菌については、16S rRNA 遺伝子配列に基づき種を推定した。血液寒天プレート上の溶血斑の大きさから溶血活性を測定する手法の妥当性を、従来手法と比較して確認したのち、単離した溶血性細菌の溶血活性を血液寒天プレートで測定した。大腸菌と *Aeromonas* 属の細菌については、16S rRNA 遺伝子のアレル解析も行い、溶血活性と 16S rRNA 遺伝子配列の関係性を明らかにした。単離した一部の細菌については全ゲノムシーケンス解析を行った。得られたシーケンスデータをもとに、種の同定、保有する溶血毒遺伝子、病原遺伝子を明らかにした。大腸菌と同定された株については、河川から単離された溶血性大腸菌 3 株を加えて、MLST 解析に基づく分類、*fimH* 対立遺伝子に基づく分類、血清型分類、*phyloptype* に基づく分類、系統解析、保有する病原遺伝子に基づく病原型の推定、保有する病原遺伝子のパターン分類を行い、溶血性大腸菌の詳細な分類を行った。溶血性大腸菌の中で溶血活性が異なっていたことから、溶血毒 α ヘモリシンの mRNA 発現量およびタンパク分泌量、 α ヘモリシン産生に関わる遺伝子の変異について測定し、溶血活性の決定因子について調べた。大腸菌以外に同定された株については、種を同定し、保有する溶血毒遺伝子を明らかにした。

1.5. 全ゲノムショットガンシーケンス法に基づく遺伝子保有の解釈

本研究で使用した Illumina 社 (San Diego, CA) の MiSeq をはじめとする次世代シーケンサは、従来のキャピラリー電気泳動によるサンガー法と比べ、ハイスループットかつ低価格でのゲノム解析を可能とする。MiSeq を用いた解析を行った場合、最小 1×36 bp、最大 2×300 bp 長のリードを 12,000,000~50,000,000 リード得ることができる⁴¹。全ゲノムショットガンシーケンス法では、ランダムに断片化したリードを次世代シーケンサで分析することで、短いリードの配列データを大量に得ることができる。こうして得られた短いリードを組み合わせて元のゲノムとしての配列に戻す、すなわちアセンブルする方法には、アラインメント (あるいはマッピングと呼ばれる) と *de novo* アセンブリの 2 種類の方法がある⁴²。アラインメントは、既知のゲノム配列をリファレンスとして、次世代シーケンサで得られたショートリードをリファレンスゲノム配列上の相同性のある領域に貼り付けてアセンブルを行う (図 1 (i))。一方、*de novo* は、ラテン語で「最初から」や「改めて」という意味があり、*de novo* アセンブリではリファレンスを用いず、数百塩基の長さの配列からオーバーラップする領域を検索し、それらをつなぎ合わせることで 1 から元のゲノム配列をアセンブルする (図 1 (ii))。De novo アセンブリで作成された一続きの領域をコンティグとよび、一度のアセンブルで複数個のコンティグが作成される。

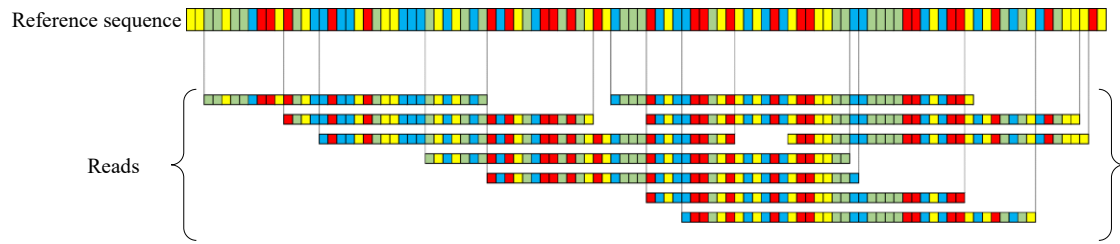
全ゲノムシーケンスの結果から遺伝子の保有を確認する場合、その解析方法と解釈はアセンブル手法によって異なる。アセンブルがアラインメントによって行われた場合、対象となる遺伝子領域にリードがアラインメントされるかどうかで判断することができる。この手法は、Center for Genomic Epidemiology (CGE) の提供する、全ゲノムシーケンスデータから病原性遺伝子の保有を調べるオンラインツール VirulenceFinder2.0

(<https://cge.cbs.dtu.dk/services/VirulenceFinder/>) で、リードの種類として「Raw Sequencing Reads (fastq)」を選択した際などで取られている⁴³。VirulenceFinder2.0 では、k-mer alignment (KMA) プログラム⁴⁴を内部で使用して、リファレンスとなる遺伝子配列にリードのアラインメントを行い遺伝子の保有を確認する。この手法によるメリットとしては、まず高速でできる点が挙げられる。一般的に de novo アセンブルを行うには膨大なメモリ (RAM) と時間を要するが、アラインメントに基づく方法ではこの操作を介さないため低スペックのコンピュータでも高速に解析を行える。また、de novo アセンブリによって、対象とする遺伝子領域でコンティグが作成されなかった場合においても、アラインメントに基づく手法では対象遺伝子を検出できる可能性がある。一方、デメリットとして相同性の低い遺伝子を検出することが難しい点が挙げられる。タンパク質をコードする遺伝子ではしばしばアミノ酸置換を伴わない変異 (silent mutation) が起こる。アミノ酸配列は一致しているが塩基配列が大きく異なるような場合、アラインメントに基づく遺伝子検出では偽陰性になる可能性が高いと考えられる。また、遺伝子周辺のゲノム構造がリファレンスと解析対象株のゲノムで異なる場合、検出した遺伝子のゲノム上での位置関係まではわからない。

一方、アセンブルが de novo アセンブリによって行われた場合は、リファレンスとなる遺伝子配列とコンティグ配列からよく似た配列を BLAST 等の相同検索プログラムを用いて検索することにより遺伝子保有を確認する。この手法は、VirulenceFinder2.0 で、リードの種類として「Assembled or Draft Genome/Contigs*」を選択した際に取りられており、VirulenceFinder2.0 では、BLAST 検索を内部で実行している⁴³。De novo アセンブリに基づく遺伝子検出によるメリットとしては、対象としたそれぞれの遺伝子のゲノム上での位置関係までわかる点が挙げられる。いくつかの遺伝子は同一のオペロンや、遺伝子アイランドと呼ばれる複数の遺伝子が特定の領域に集まった領域に存在することが知られ⁴⁵、これらの位置関係を見ることで、細菌の病原性推定などのヒントになりうる。また、リファレンス配列をアミノ酸配列に変換して検索することで、塩基配列の相同性は低いアミノ酸配列の相同性が高いホモログ遺伝子を検出できるようになる。一方、デメリットとしては前述のとおりコンティグ上に対象とする遺伝子領域が乗っていない場合は検出することができない点が挙げられる。また、コンティグの端に対象とする遺伝子領域が乗っている場合は、見かけ上高い相同性を示すことがあるが、部分的に一致しているだけの可能性がある。対象とする遺伝子の全長がコンティグ上にあることを確認したうえで遺伝子保有について議論する必要がある。設備上の問題としては、de novo アセンブリは計算量が膨大となるため、十分なメモリ量 (RAM) と CPU コアをもつコンピュータを使用することが望まれる。

アラインメントに基づくアセンブルでは対象とする遺伝子の相同性が低い場合検出できない可能性があり、de novo アセンブリに基づくアセンブルではコンティグ上に乗っていない場合に検出できない可能性があり、いずれの方法においても偽陰性になる可能性がある、それぞれ注意する必要がある。

(i) Mapping



(ii) De novo assembly

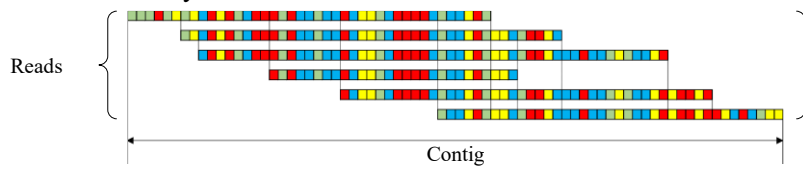


図1 (i) アラインメントに基づくアセンブルと (ii) de novo アセンブリに基づくアセンブルと作成されるコンティグ

2. 方法

2.1. 実験で使用した培地

LB 培地は、10g/L トリプトン、5g/L 酵母エキス、5g/L NaCl の Lennox LB 培地をオートクレーブして使用した。LB プレートは、LB 培地に寒天 1.5%を加えてオートクレーブし、直径 9 cm のディッシュに広げて固化させたものを使用した。軟寒天 LB 培地は、LB 培地に寒天 0.7%を加えてオートクレーブし、4°Cで保存した。使用直前に電子レンジで再融解させ、50°Cの湯浴で冷まして使用した。血液寒天培地は、LB 培地に寒天 1.5%を加えてオートクレーブし、4°Cで保存した。使用直前に再融解させ、50°Cの湯浴で冷ましたのち、使用直前に羊脱繊維血液 5%を加えて使用した。

2.2. 下水に含まれる溶血性細菌の単離

滋賀県大津市の下水処理場流入下水を日中採水した。流入下水は下水処理場からの返送汚泥を含まず、自動採水装置を用いて採水を行った。再融解させた軟寒天 LB 培地に下水を 10^{-3} ~ 10^{-4} %になるように希釈して加え、羊脱繊維血液 5%を加えてすぐに LB プレート上に 10mL 重層した。固化後、37°Cの好気条件下で 1 日培養した。培養後、溶血斑を示す、もしくは溶血斑のあることが疑われるコロニーを白金耳で触れて、別の LB プレートに画線塗抹しさらに一晚培養した。単一のコロニーからピックアップしたものを LB 培地に植種し、一晚振盪培養した。十分濁ったことを確認し、15%グリセロールを加え-80°Cで保存した。*Streptococcus* 属等の、一部の LB 培地で濁らなかった株は、Todd hewitt broth 培地 (Merck KGaA) に植種し、2 日間振盪培養した。十分濁ったことを確認し、15%グリセロールを加え-80°Cで保存した。

2.3. 血液寒天プレート上における溶血活性の測定

溶血斑の大きさに基づく溶血活性の測定は 3 種類のプレート (表 1) を使用した。血液寒天プレートタイプ I は、1 層目の LB プレートの上に細菌培養液もしくはプレート上のコロニーに触れた白金耳で画線塗抹し、その上に血液寒天培地を重層し、固化後さらに寒天 (1.5%) と塩化ナトリウム (1.0%) を融解させて 50°C程度冷ましたものを重層した。血液寒天プレートタイプ II は、約 50°Cの融解した血液寒天培地と細菌培養液の混合液を加えたものを LB プレート上に重層した。血液寒天プレートタイプ III は、1 層目の LB プレートの上に血液寒天培地を重層し、固化後さらに寒天 (1.5%) と塩化ナトリウム (1.0%) を溶解させて 50°C程度まで冷ましたものを重層した。固化したのち、細菌培養液、もしくはプレート培養で得られたコロニーに触れた白金耳で、プレートの上から穿刺した。いずれのプレートも 37°Cの好気条件下で培養し、24 時間あるいは 48 時間経過時の溶血斑の形成をそれぞれ目視で確認した。いずれのプレートも直径 90 mm のものを使用した。

溶血斑の形成が確認された株は、図 2 に示すような装置で溶血斑を撮影した。使用したカメラは、PowerShot G12 (Canon) で、いずれのサンプルも同一撮影条件下でプレートの透過光を撮影した。撮影した画像を画像解析ソフト ImageJ⁴⁶に取り込み、ディッシュの直径： d [pixel]、複数のコロニーサイズ： c [pixel]、およびその溶血斑のサイズ： h [pixel]を測定した (図 3)。コロニーサイズは、コロニーの形状が楕円形の場合はその最も長い長径をコロニーサイズとして測定した。培地表面に露出するまで成長したコロニー (図 3 の黄枠で囲んだようなコロニー) は、コロニーが横に大きく広がっていたため、培地表面に露出していないコロニーのみを対象としてコロニーサイズおよび溶血斑サイズを測定した。

ディッシュの直径は 90 mm で、以下の式で溶血斑幅：Hemolysis Width を計算した。各株の溶血斑幅を 5~7 個測定し、その平均値を溶血活性の値とした。

$$\{\text{Hemolysis Width}\} = \frac{\text{average}(h - c)}{2 \times d} \times 90 [\text{mm}]$$

表 1 溶血斑を測定する際に使用した血液寒天プレート

プレートタイプ		タイプ I	タイプ II	タイプ III
概観				
使用した培地	3層目	寒天 (1.5%) + NaCl (1.0%)	-	寒天 (1.5%) + NaCl (1.0%)
	2層目	血液寒天培地	血液寒天培地	血液寒天培地
	1層目	寒天 (1.5%) + LB培地	寒天 (1.5%) + LB培地	寒天 (1.5%) + LB培地
植菌方法		1層目の上に白金耳で画線塗抹	2層目に混和	白金耳で上から穿刺

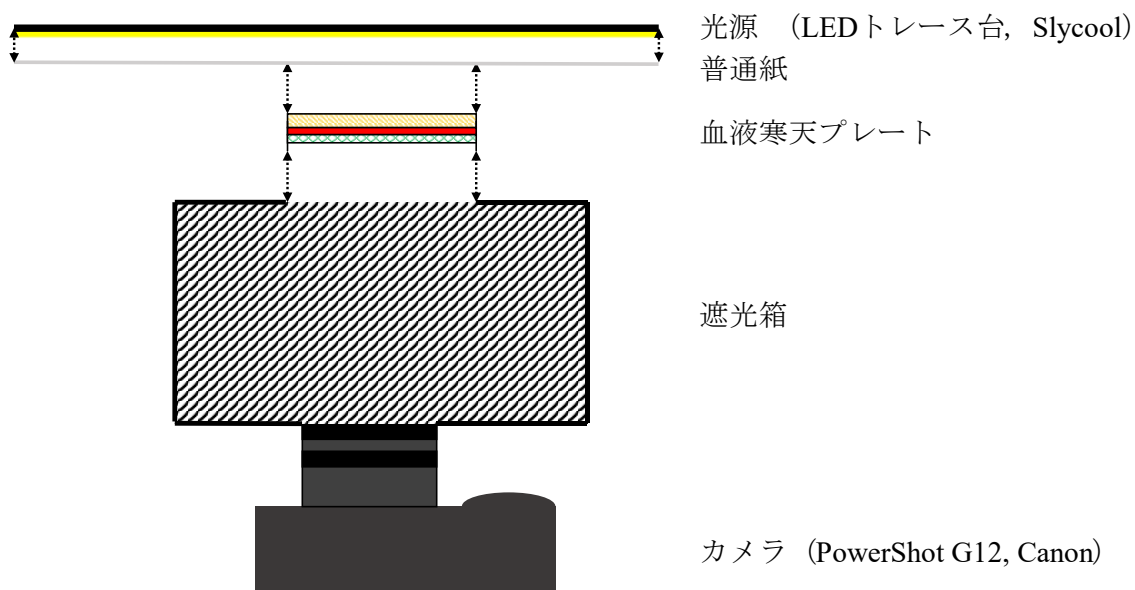


図 2 溶血斑の撮影に使用した装置の概略

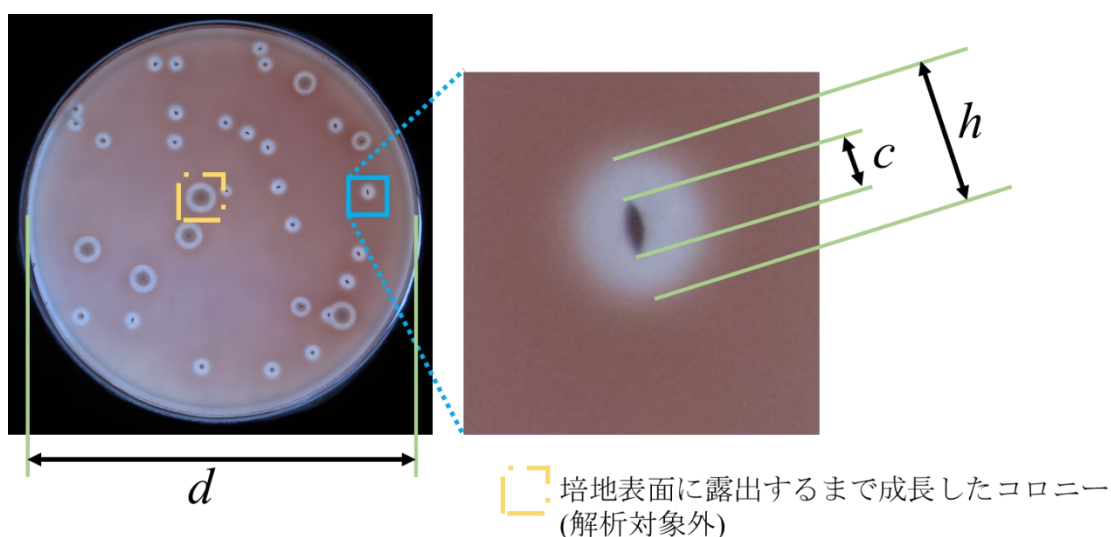


図3 溶血斑幅の計算に使用した数値の計測項目

2.4. 液体培地培養時の上清の溶血活性の測定、上清タンパク分析、大腸菌の *hlyA* mRNA の測定

大腸菌を LB 培地、37°C の好気条件下で一晩振盪培養した。LB 培地に 10 mM の塩化カルシウムを加えた培地に培養液を等量加え、2 時間培養した。OD595 を測定して大腸菌が増えていることを確認した。培養液の一部については、12,000×g で 10 分間遠心分離して、大腸菌培養液の上清を得た。

上清の溶血活性の測定は、Taneike ら (2002) の方法⁴⁷を参考に次のように行った。羊脱繊維血液に LB 培地を加え、軽く混合した後 800×g で 2 分間遠心分離し、上清を捨てる操作を数回繰り返して、得られた沈殿を洗浄済み赤血球として使用した。大腸菌培養液上清もしくはブランク (大腸菌を加えていない培地) に、洗浄済み赤血球 (5%)、塩化カルシウム (10 mM) をそれぞれ加え、軽く混和後 37°C で 1 時間静置した。その後 800×g で 2 分遠心分離し、上清の吸光度 (波長 450 nm・595 nm) を Microplate Reader Model 550 (Bio-Rad) で測定し、サンプルの吸光度 ($A_{450_{\text{sample}}}$ ・ $A_{595_{\text{sample}}}$) とブランクの吸光度 ($A_{450_{\text{blank}}}$ ・ $A_{595_{\text{blank}}}$) を得た。また、超純水に 5% 洗浄済み赤血球を加えて赤血球細胞膜を浸透圧の差で破壊したものを、800×g で遠心分離し、その上清と超純水の吸光度をそれぞれ測定し、100% 溶血時の吸光度 ($A_{450_{100\%}}$ ・ $A_{595_{100\%}}$) と水の吸光度 ($A_{450_{\text{water}}}$ ・ $A_{595_{\text{water}}}$) を得た。以下の式で溶血率 (Percent lysis) をそれぞれ測定した。

$$\text{Percent_lysis}_{450} = \frac{A_{450_{\text{sample}}} - A_{450_{\text{blank}}}}{A_{450_{100\%}} - A_{450_{\text{water}}}} \times 100[\%]$$

$$\text{Percent_lysis}_{595} = \frac{A_{595_{\text{sample}}} - A_{595_{\text{blank}}}}{A_{595_{100\%}} - A_{595_{\text{water}}}} \times 100[\%]$$

上清のタンパク分析では、まず上清タンパクの濃縮を行った。大腸菌培養液上清に氷冷したトリクロロ酢酸を 10% になるように加え、混和後氷上で 30 分静置したのち、12,000×g、4°C で遠心分離した。上清を取り除き、-30°C に冷やしたアセトンを少量加え、15,000×

g で 10 分遠心分離後、上清を取り除いた。残りを蒸発乾固させ、サンプルバッファーで溶解させた。サンプルバッファーには 1X NuPAGE® LDS Sample Buffer (Thermo Fisher Scientific)、100 mM ジチオスレイトール (DTT) を含む。濃縮したサンプルを 70°C で 10 分加熱して変性させた。変性後のサンプルを NuPAGE® Bis-Tris Gel (Thermo Fisher Scientific) にアプライした。XCell SureLock Mini-Cell Electrophoresis system (Thermo Fisher Scientific) を使用して、泳動槽を氷冷しながら 200V の一定電圧で 60 分間電気泳動した。泳動後のゲルは、銀染色 II キットワコー (富士フイルム和光純薬) を用いて染色した。

大腸菌の mRNA 測定では、まず大腸菌から RNA を抽出した。RNA の抽出は、RNAprotect® Bacteria Reagent (QIAGEN) 及び RNeasy Mini Kit (QIAGEN) を用いて行った。逆転写は PrimeScript™ II 1st strand cDNA Synthesis Kit (タカラバイオ) を用いて行った。定量 PCR では QuantiFast SYBR Green PCR Kit (QIAGEN) を用いて調整した反応液を、Thermal Cycler Dice® Real Time System II により 2 ステップ PCR (95°C 10 秒-60°C 30 秒、40 サイクル) で回した。反応後に融解曲線を描き、目的産物が増幅されていることを確認した。hlyA 遺伝子を対象として使用したプライマーは hlyA-F (5'-CGGCACAGCAGAGAACTC-3') と hlyA-R (5'-CACTGCCTGCCTTTCCTAAG-3') で、PCR 産物は 155 bp であった。内在コントロールとして gapA 遺伝子を用いた。プライマーは、Nhu ら (2019) ⁴⁸ の 4774-gapA-F (5'-CGTTAAAGGCGCTAACTTCG-3') と 4774-gapA-R (5'-ACGGTGGTCATCAGACCTTC-3') を使用し、PCR 産物は 119 bp であった。

2.5. 単離株の 16S rRNA 遺伝子 V3-V4 領域のシーケンシング

LB 培地で一晚培養した単離細菌株を 95°C で 5 分間加熱して不活化した。不活性化した菌の 16S rRNA V3-V4 領域の DNA を KOD-Plus-Neo (東洋紡) と、Illumina 社の 16S Metagenomic Sequencing Library Preparation プロトコルで用いられるアンプリコン PCR プライマーを若干変更したプライマー (図 4) で増幅した。図 4 において、Illumina overhanging adaptor sequence は後段の操作で使用する Nextera XT index kit のプライマーとのオーバーハング領域を表し、16S rRNA targeting sequence はバクテリアと古細菌の 16S rRNA を最も広く認識するプライマーペアの配列 ⁴⁹ を表す。Index sequence は独自に設計した配列で、いくつかのサンプルで異なる Index sequence の組み合わせを使用した。シーケンシング後に index sequence の配列の組み合わせでサンプルを分類した。アンプリコン長は、16S rRNA のターゲット領域約 460bp に overhanging adaptor sequence と index sequence を加えて、約 540bp となった。Nextera XT Library Preparation Kit (Illumina) で調製したライブラリ MiSeq (Illumina) により、300 サイクルのペアドエンド反応でシーケンシングを行った。

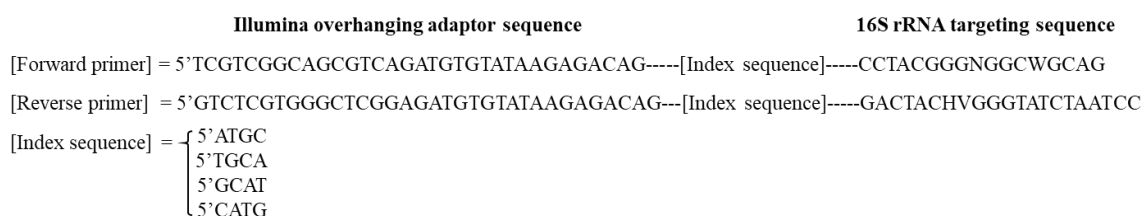


図 4. 16S rRNA 遺伝子 V3-4 領域のシーケンシングに使用したアンプリコン PCR プライマー

2.6. 16S rRNA 遺伝子 V3-4 領域の解析による単離細菌株の種の推定

16S rRNA 遺伝子のシーケンシングで得られたペアドエンドリードのマージを、PEAR-v0.9.6⁵⁰を用いて行った。マージされたリードの両端に含まれる Index sequence の組み合わせに従い、リードをサンプルごとに分類した。シーケンス時に読み取れなかった塩基を一つでも含むリード (Nを含むリード) や、シーケンス時のクオリティスコア (Q-Score) に5以下のものが含まれるリード、アンプリコン長と比べて明らかに短い400bp未満のリードについては、解析から除外した。サンプルごとにユニークな配列のリードを取り出し、それぞれのリード数をカウントした。最も多くカウントされたリードを、単離株の代表配列とした。また、配列比較のためのリファレンス配列をNCBIのBLASTデータベース (https://ftp.ncbi.nlm.nih.gov/blast/db/16S_ribosomal_RNA.tar.gz, 2021/6/21 アクセス) から得た。単離株の代表配列と、データベースに保存されている株の16S rRNA 遺伝子配列を、コマンドライン上からBLASTN (Nucleotide-Nucleotide BLAST 2.11.0+) ⁵¹で比較し、一致率 (pident) の値が最大となったときのリファレンス配列の種を、単離細菌株の推定種とした。

単離株の16S rRNA 遺伝子 V3-4 領域代表配列の系統解析では、MEGA X⁵²を用いてマルチプルアラインメントおよび系統樹の描画を行った。マルチプルアラインメントでは、ClustalW⁵³を使用し、系統解析では Neighbor-Joining 法⁵⁴で系統樹の描画、Kimura 2-parameter 法⁵⁵で進化距離の計算を行った。

2.7. 16S rRNA 遺伝子 V3-4 領域のアレル解析

16S rRNA 遺伝子 V3-V4 領域の種推定で近縁種に推定された株についてはアレル解析を行った。アレル解析では、CLC genomics workbenchを用いてトリム、マージしたリードの両端に含まれる Index sequence の組み合わせに従い、リードをサンプルごとに分類した。各株から検出されたユニークな配列 (アレル) をリストアップし、株ごとにアレルの内訳を割合として算出した。ユニークなアレルをMEGA X⁵²を用いてマルチプルアラインメントおよび系統樹の描画を行った。マルチプルアラインメントでは、ClustalW⁵³を使用し、系統解析では Neighbor-Joining 法⁵⁴で系統樹の描画、Kimura 2-parameter 法⁵⁵で進化距離の計算を行った。それぞれの株で検出されたユニークアレルの割合を ward 法⁵⁶によりパターン分析して、株をクラスタリングした。

2.8. 単離株の全ゲノムシーケンスとリードの前処理

グラム陰性菌のDNA抽出、精製には、DNeasy (QIAGEN)を使用した。ライブラリ調整はNextera XT Library Preparation Kit (Illumina)で行った。ライブラリ調整後、MiSeq (Illumina)により300サイクルのペアドエンド反応でシーケンシングを行った。

KKo007、KOr014、KKa004株のシーケンス結果はNCBIデータベース

(<https://www.ncbi.nlm.nih.gov/sra>) から得た。国際塩基配列データベース (the International Nucleotide Sequence Databases; INSD) アクセション番号はそれぞれKKo007

(DRR063477、DRR063383)、KOr014 (DRR063420、DRR063293)、KKa004

(DRR063293)であった。

Fastq形式で得られたリード配列を、fastQC⁵⁷と multiqc⁵⁸でクオリティチェックを行った。その後、アダプタートリミングとクオリティコントロールを、fastpプログラム⁵⁹で行った。fastQCとmultiqcで結果を確認したところ3'末端のクオリティが低くpolyGテール

を確認できたため、fastpでpolyGテールを取り除いた。アセンブラとしてSPAdes genome assembler (v3.15.2)⁶⁰を使用して、クオリティコントロール済みのリードをアセンブルしてコンティグを得た。

2.9. 全ゲノムシーケンスを行った株の種の推定・同定

PubMLST (<https://pubmlst.org/>)⁶¹の解析ツールの一つ、Ribosomal Multilocus Sequence Typing (rMLST)⁶²を用いて、種の推定を行った。PubMLST上のrMLSTウェブサービス上に、コンティグデータをアップロードして、種の推定を行った。

Average nucleotide index (ANI)の計算にはfastANIを使用した⁶³。各細菌種の基準株 (type strain) とのANI値が95%以上のものについて、対象株をtype strainと同一種であると決定した。大腸菌 (*Escherichia coli*) と赤痢菌 (*Shigella* 属) の識別には、kmer ID⁶⁴を使用して、表2に示すリファレンス株との類似度 (Similarity) を計算した。

表2 kmer IDを用いた解析で使用したリファレンス株

Genus	Species	Strain
<i>Shigella</i>	<i>S. boydii</i>	CDC 3083-94 Sb227
		<i>S. dysenteriae</i>
	<i>S. flexneri</i>	2457T 2002017
	<i>S. sonnei</i>	53G Ss046
		<i>Escherichia</i>
O7:K1 IAI39 (ExPEC)		
K-12 substr. W3110		
NA114 (UPEC)		
O111:H- 11128 (EHEC)		
O127:H6 E2348/69 (EPEC)		
O157:H7 TW14359 (EHEC)		
O26:H11 11368 (EHEC)		
SMS-3-5		
UM146 (AIEC)		
W		
	<i>E. fergusonii</i>	ATCC 35469

2.10. 大腸菌の分類

FimHの分類は、Fim Typer 1.0 (<https://cge.cbs.dtu.dk/services/FimTyper/>)⁶⁵にコンティグデータをアップロードして解析した。血清型の推定は、Serotype Finder 2.0

(<https://cge.cbs.dtu.dk/services/SerotypeFinder/>)⁶⁶にコンティグデータをアップロードして解析した。

Multi Locus Sequence Typing (MLST) の分類では、Center for Genomic Epidemiology のツールの Multi-Locus Sequence Typing (<https://bitbucket.org/genomicepidemiology/mlst.git>、2021/9/22 アクセス) をローカル環境にクローンして使用した。依存関係にあるツール KMA (<https://bitbucket.org/genomicepidemiology/kma.git>、2021/9/22 アクセス)⁴⁴およびデータベース (https://bitbucket.org/genomicepidemiology/mlst_db.git、2021/9/22 アクセス) も同様にローカル環境にクローンして使用した。解析では、Achtman スキーム⁶⁷、すなわち7種類のハウスキーピング遺伝子 (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, *recA*) の塩基配列の組み合わせに従い、シーケンスタイプ (ST) を決定した。

2.11. 大腸菌の各病原型に特徴的な病原遺伝子保有の確認

アダプタートリミングおよびクオリティコントロール済みのリードを、対象とする遺伝子とその前後 300 bp の塩基配列にアラインメントした。対象とした遺伝子のリファレンスとする塩基配列は、NCBI データベースから表 3 に示すゲノム配列をダウンロードして使用した。表では、対象とした遺伝子を保有する細菌株のシーケンスデータの INSD アクセス番号 (accession number)、それぞれの対象タンパク質をコードする位置、UniProt データベースから得られたアミノ酸配列と塩基配列の Full length pident を表記している。full length pident は、UniProt データベースから得られた遺伝子アミノ酸配列と NCBI データベースから得られたゲノム上の遺伝子配列の相同性を判断する指標として、次の式から算出した。なお、*pident*、*length*、*slen*、*send*、*sstart* の値は、BLASTX によって得られた値を使用した。

$$\{Full\ length\ pident\} = pident \times \frac{length}{slen + length - (send - sstart + 1)}$$

対象とする遺伝子領域および前後 300 bp の塩基配列に対し *bwa mem*⁶⁸ でアラインメントを行い、コーディング領域 (CDS) 内のアラインメントされた領域の割合 (gene coverage) を計算し、遺伝子の保有を確認した。

2.12. kSNP3.0 による系統解析

コンティグデータを用いて kSNP3.0 (v1.32)⁶⁹ で系統解析を行った。k-mer 値は kSNP3.0 に含まれる Kchooser 関数で最適値を計算して得られた「25」を使用した。バクテリアはウイルス等に比べ何千もの遺伝子を保有し、異なる遺伝子座の部位における変化の割合はほぼ同等とみなすことができ、このような場合においてはデフォルトの parsimony tree よりも maximum likelihood (ML) tree の方がより多くの SNP がノードにアラインメントされる⁷⁰ことから、系統樹の計算には ML tree を使用した。

表 3 大腸菌の各病原型に特徴的な病原遺伝子のリファレンス配列

Protein	Gene	UniProt ID	accession number	Nucleotide position		full length pident [%]
				start	end	
Pap fimbrial major pilin protein	<i>papA</i>	P04127	NZ_CP062228.1	4797030	4796467	81.58
S-fimbrial adhesin protein SfaS	<i>sfaS</i>	P13430	NC_022370.1	1114041	1114532	98.18
Type-1 fimbrial protein, C chain	<i>pilC</i>	P62605	NC_017631.1	1163034	1163576	98.90
Major structural subunit of bundle-forming pilus	<i>bfpA</i>	P58997	NC_010862.1	2646	3227	98.46
CFA/I fimbrial subunit B	<i>cfaB</i>	P0CK93	NC_017724.1	38571	39083	98.26
CS3 fimbrial subunit A	<i>N/A</i>	P15488	NC_014232.1	54992	55498	98.82
AAF/I fimbrial subunit	<i>aggA</i>	P46007	CP027392.1	72325	72828	68.21
Intimin	<i>eae</i>	O31000	NC_002695.2	4599387	4596583	88.90
E3 ubiquitin-protein ligase ipaH3	<i>ipaH3</i>	Q83RJ4	NC_004337.2	1423779	1422064	99.65
Probable E3 ubiquitin-protein ligase ipaH4.5	<i>ipaH4.5</i>	P18009	NC_002698.1	63876	65600	99.65
Probable E3 ubiquitin-protein ligase ipaH7.8	<i>ipaH7.8</i>	P18014	NC_002698.1	61751	63448	98.59
E3 ubiquitin-protein ligase ipaH9.8	<i>ipaH9.8</i>	Q8VSC3	NC_004851.1	181458	183095	99.63
Aerobactin synthase	<i>iucC</i>	Q47318	NC_007675.1	100818	102560	98.97
N-acylneuraminate cytidyltransferase	<i>neuA</i>	P13266	NC_010498.1	3304624	3303368	98.81
Polysialic acid biosynthesis protein P7	<i>neuC</i>	Q47400	NZ_WVVR01000003.1	341118	342293	99.49
KfiB protein	<i>KfiB</i>	Q6KCZ4	NC_017631.1	3416817	3415213	94.69
Heat-stable enterotoxin ST-IA/ST-P	<i>sta1</i>	P01559	NC_017722.1	57269	57051	97.30
Heat-stable enterotoxin II	<i>stiI</i>	P22542	NZ_UGEZ01000004.1	24248	24033	97.26
Shiga-like toxin 1 subunit A	<i>stxA</i>	P08026	NC_007606.1	1283865	1284812	99.05
Shiga-like toxin 1 subunit B	<i>stxB</i>	P69178	NC_007606.1	1284822	1285091	97.80
Shiga-like toxin 2 subunit A	<i>stxA2</i>	P09385	NC_000924.1	21462	22421	99.38
Shiga-like toxin 2 subunit B	<i>stxB2</i>	P09386	NC_000924.1	22433	22702	97.80
Heat-labile enterotoxin A chain	<i>eltA</i>	P06717	NC_017722.1	51370	50594	98.08
Heat-labile enterotoxin B chain	<i>eltB</i>	P32890	NC_017722.1	50597	50223	94.44
Heat-labile enterotoxin IIA, A chain	<i>N/A</i>	P13810	JQ031711.1	636	1415	99.23
Heat-labile enterotoxin IIA, B chain	<i>N/A</i>	P13812	NZ_VOCX01000183.1	510	139	98.40
Hemolysin, chromosomal	<i>hlyA</i>	P09983	NC_007946.1	4828096	4825022	97.27
Enterohemolysin	<i>EHEC-hlyA</i>	Q9LCS8	NC_007414.1	39822	42818	97.30
Cytotoxic necrotizing factor 1	<i>cnf1</i>	Q47106	NC_007946.1	4820427	4817383	99.61
EAST1 toxin	<i>EAST1</i>	Q53559	NC_007675.1	183520	184311	92.50

N/A: There is no gene named for correspond protein

2.13. Missense Mutation Viewer (MMViewer) を用いた SNP 解析

次世代シーケンサ解析で得られたショートリードからコーディング領域 (CDS) 内のミスセンス変異および CDS 外の変異を解析するツールである Missense Mutation Viewer (MMViewer) は筆者の開発したソフトで、GitHub (<https://github.com/KIkebata/mmviewer>) からダウンロードすることができる。

SNP 解析のリファレンスとして表 4 で示す株のゲノムを使用した。大腸菌 S65EC 株の配列は Nhu ら (2019) の研究⁴⁸ で得られたコンプリートゲノムを使用した。大腸菌 VREC0578 株はコンティグ ERS784353SCcontig000036 の配列を使用した。 α ヘモリシンの溶血活性に関わる遺伝子のアミノ酸配列データは UniProt データベース (<https://www.uniprot.org/>) から入手した。対象とした遺伝子のアミノ酸配列データについては表 4 の Product の列に示す。MMViewer を用いて、大腸菌のゲノム配列と対象とする遺伝子のアミノ酸配列を比較して、ゲノム上で対象とする遺伝子をコードする領域を探索し、これらの遺伝子を保有することを確認した。なお、アラインメントされた領域は Depth が 1 以上の領域とした。

2.14. 大腸菌以外の溶血性細菌の保有する溶血毒の予測

大腸菌以外の単離溶血性細菌株の溶血毒遺伝子を予測するため、VFAnalyzer2.0 (<http://www.mgc.ac.cn/VFs/>)⁷¹および prokka⁷²を使用した。VFAnalyzer2.0には、コンティグ配列をアップロードして保有する溶血毒遺伝子を確認した。prokkaを用いた解析では、コンティグデータに対してアノテーションを行い、「-lysin」、「RTX」を含む単語でアノテートされたCDSをリストアップした。

表 4 MMViwer を用いた SNP 解析で使用したリファレンス配列

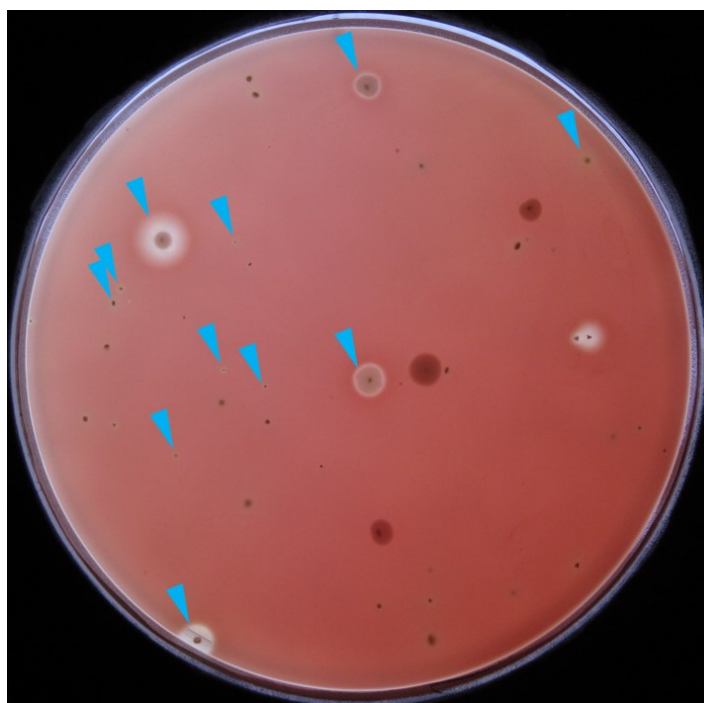
gene	Product		Strain	Accession	Target region		strand
	Protein	UniProt ID			start	end	
<i>hlyC</i>	Hemolysin-activating lysine-acyltransferase HlyC	P09985	<i>Escherichia coli</i> strain S65EC	CP036245.1	4813777	4814289	+
<i>hlyA</i>	Hemolysin	P09983	<i>Escherichia coli</i> strain S65EC	CP036245.1	4814301	4817375	+
<i>hlyB</i>	Alpha-hemolysin translocation ATP-binding protein HlyB	P10089	<i>Escherichia coli</i> strain S65EC	CP036245.1	4817446	4819569	+
<i>hlyD</i>	Hemolysin secretion protein D	P09986	<i>Escherichia coli</i> strain S65EC	CP036245.1	4819588	4821024	+
<i>hlyE</i>	Hemolysin E, chromosomal	P77335	<i>Escherichia coli</i> strain K-12 substr. MG1655	NC_000913.3	1229483	1230394	-
<i>hlyF</i>	HlyF	A0A0H2XJH4	<i>Escherichia coli</i> strain VREC0578	UDAB01000036.1	16394	17503	-
<i>acrR</i>	HTH-type transcriptional regulator AcrR	P0ACS9	<i>Escherichia coli</i> strain S65EC	CP036245.1	512415	513062	+
<i>dnaJ</i>	Chaperone protein DnaJ	P08622	<i>Escherichia coli</i> strain S65EC	CP036245.1	14049	15179	+
<i>dnaK</i>	Chaperone protein DnaK	P0A6Y8	<i>Escherichia coli</i> strain S65EC	CP036245.1	12044	13960	+
<i>hns</i>	DNA-binding protein H-NS	P0ACF8	<i>Escherichia coli</i> strain S65EC	CP036245.1	1446024	1446437	-
<i>rfaE</i>	Bifunctional protein HldE	P76658	<i>Escherichia coli</i> strain S65EC	CP036245.1	3492127	3493560	-
<i>rfaH</i>	Transcription antitermination protein RfaH	P0AFW0	<i>Escherichia coli</i> strain S65EC	CP036245.1	4332885	4333373	-
<i>rne</i>	Ribonuclease E	P21513	<i>Escherichia coli</i> strain S65EC	CP036245.1	1232002	1235187	-
<i>tolC</i>	Outer membrane protein TolC	P02930	<i>Escherichia coli</i> strain S65EC	CP036245.1	3472854	3474335	+
<i>waaC</i>	Lipopolysaccharide heptosyltransferase 1	P24173	<i>Escherichia coli</i> strain S65EC	CP036245.1	4079658	4080617	+
<i>waaF</i>	ADP-heptose--LPS heptosyltransferase 2	P37692	<i>Escherichia coli</i> strain S65EC	CP036245.1	4078608	4079654	+
<i>waaG</i>	Lipopolysaccharide core biosynthesis protein RfaG	P25740	<i>Escherichia coli</i> strain S65EC	CP036245.1	4089621	4090745	-

3. 結果

3.1. 血液寒天培地によって単離される溶血性細菌

3.1.1. 溶血性細菌の検出・単離

大津市の下水処理場流入水から溶血性細菌を単離した。血液寒天培地に下水処理場流入水を混ぜてプレート上で固化したものを、好気条件下 37°Cで一晩培養した。その結果、図5に示すようなプレートが得られた。溶血活性を示す株は周囲の赤血球を溶解（溶血）させ、図のような溶血斑を示した。溶血斑の大きさは一定ではなく、大小の差が見られた。コロニー間が十分に離れた溶血性細菌を単離した。ここでの単離は、できるだけ多くの溶血性細菌を単離するため、明瞭な溶血斑を示す株だけでなく、溶血斑とみられる透明部分が非常に小さいものも対象として単離を行った。一部の株は、後の溶血活性測定時に溶血斑を示さなかった。0.1 μLあたり約 10 株程度の溶血性細菌を単離することができた。



▼ 単離対象とした溶血性細菌

図5 血液寒天培地と下水流入水を混合して一晩培養したときの様子

3.1.2. 溶血活性の測定

単離した溶血性細菌は、細菌を LB プレートに画線塗抹した上から血液寒天培地を重層する血液寒天プレートタイプ II と、血液寒天プレートに白金耳で穿刺培養する血液寒天プレートタイプ III で様々な溶血斑を示した。本論文では、便宜上これらの溶血斑を「明瞭な溶血斑 (clear hemolysis)」、「不明瞭な溶血斑 (indistinct hemolysis)」、「変色した溶血斑 (discolored hemolysis)」、「溶血斑を示さない (no hemolysis)」の 4 つに分類した

(図 6)。「明瞭な溶血斑」は、血液寒天プレートタイプ II とタイプ III の両方で、コロニー周囲が透明で、溶血斑の境界は明瞭であった。「不明瞭な溶血斑」は、血液寒天プレートタイプ II ではコロニー周囲がわずかに透明になる溶血斑を示し、溶血斑の境界は不明瞭

であった。「不明瞭な溶血斑」を示した株を、血液寒天プレートに穿刺培養する血液寒天プレートタイプⅢで培養したところ、コロニー周囲の溶血斑が明瞭な境界を持っていることが確認できた。「変色した溶血斑」は、血液寒天プレートタイプⅡでコロニー周囲がわずかに黄色っぽく変色していることが確認できた。血液寒天プレートタイプⅢでは、黄色に変色した部分をより明瞭に確認することができた。これらのコロニーを、血液寒天プレートタイプⅢで48時間もしくは72時間培養すると変色した部分が透明になることも確認できた (data not shown)。一方、血液寒天培地を用いて単離した株の中には溶血斑を示さなかった株も含まれた。図6において、溶血斑を示さなかった株は血液寒天プレートタイプⅢで寒天中のコロニーが黒っぽくその周りが薄いグレーを示しているが、これはコロニーが培地表面まで増殖し、プレート表面でコロニーが薄く広がっていることによるもので、溶血斑は示していない。明瞭な溶血斑を示した77株、不明瞭な溶血斑を示した55株、変色した溶血斑を示した株5株、溶血斑を示さなかった30株を以降の解析で用いることとした。

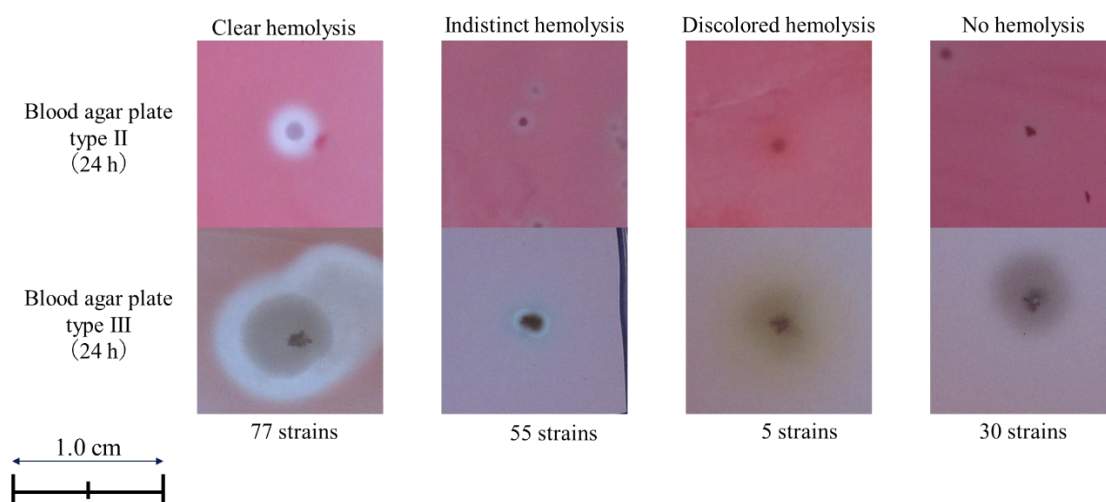


図6 単離した溶血性細菌の溶血斑の様子

3.1.3. 液体培地を用いた既存の溶血活性測定手法との比較

明瞭な溶血斑を示した株の溶血活性を測定した。溶血活性は、溶血斑幅 (hemolysis width) に基づき測定した。溶血斑幅は、コロニーの縁から溶血斑の縁までの大きさで定義した (図3)。

大腸菌をはじめとする細菌の溶血活性の一般的な測定法として、液体培地に血液と細菌もしくはその上清を加え、赤血球から放出されたヘモグロビン量で溶血活性を測定する方法がよく用いられる。Taneikeら (2002)⁴⁷やNhuら (2019)⁴⁸は、液体培地に塩化カルシウムと赤血球を加えて溶血性細菌を培養した。細菌の溶血作用により赤血球細胞膜が破壊され、ヘモグロビンが溶液中に放出される。彼らは、上清に含まれるヘモグロビン量を吸光度 (波長 540 nm) で測定することで溶血活性を測定した。溶血斑幅で溶血活性を測定する手法の妥当性を確かめるため、溶血斑幅に基づく溶血活性と、液体培地中で放出されるヘモグロビン量に基づく溶血活性に相関があることを確認することとした。なお当研究室で保有する装置の関係上、540 nmの波長の吸光度を測定することが難しく、450 nmと595

nm の波長の吸光度しか測定できなかったため、これらの2つの波長で測定することとした。

図7で、液体培地に塩化カルシウムと羊血液を加えて細菌を4時間培養し、遊離したヘモグロビン量を吸光度によって測定して溶血活性を測定する方法と、血液寒天プレート上で培養したときの溶血斑幅によって吸光度を測定する方法のそれぞれで得られた溶血活性値の関係を表す。吸光度の測定では、450 nm と 595 nm の2種類の波長で測定した。菌株は、溶血活性の異なる大腸菌 (*Escherichia coli*) 15 株と、*Aeromonas* 属の単離細菌株3株を用いた。培養血液寒天プレートでの培養は、血液寒天プレートタイプ II で24時間培養したときの溶血斑幅を測定した。

450 nm の波長と 595 nm の波長で吸光度を測定して得られた溶血活性の値と、溶血斑幅の測定で得られた溶血活性の値には高い正の相関が見られた (図7)。いずれの波長においても、大腸菌に比べ *Aeromonas* 属の方が、血液寒天プレートでの培養で溶血斑幅が大きく測定されていることが分かった。450 nm の波長で測定して得られた溶血活性と溶血斑幅との相関係数 R^2 値は、大腸菌では 0.8861、*Aeromonas* では 0.9998 と、595 nm の波長に比べそれぞれ 0.0761 ポイント、0.0865 ポイント高い値となり、波長 450 nm で特に溶血斑幅と高い相関を示した。液体培地で測定する溶血活性の測定方法と溶血斑幅の測定による溶血活性の測定方法には、同じ種であれば十分高い相関があることが示唆された。一方、*Aeromonas* は大腸菌に比べ、液体培地の測定による溶血活性に対して溶血斑幅が大きかった。種をまたぐ溶血斑幅の比較には注意する必要がある。

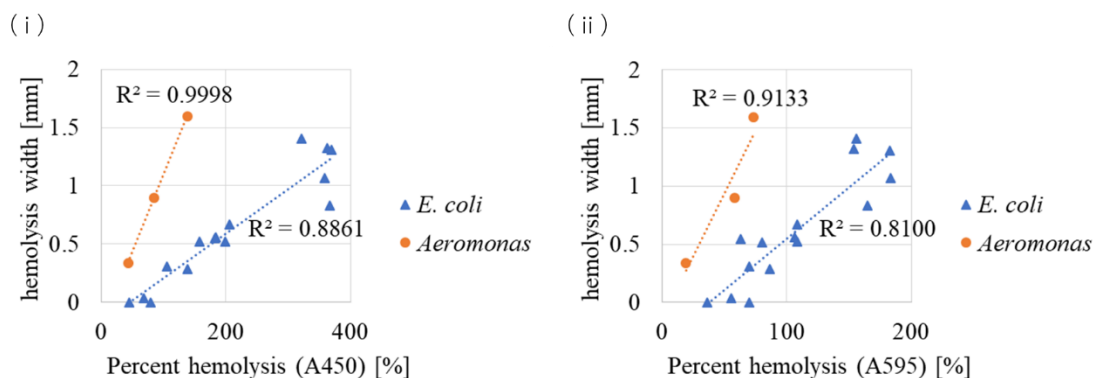


図7 (i) 液体培地で培養時の溶血活性を 450nm の波長の吸光度で測定して得られた溶血活性値 (percent hemolysis; A450) と血液寒天培地で測定した溶血活性 (hemolysis width) の関係、(ii) 液体培地で培養時の溶血活性を 595nm の波長の吸光度で測定して得られた溶血活性値 (percent hemolysis; A595) と血液寒天培地で測定した溶血活性 (hemolysis width) の関係

3.1.4. 16S rRNA 遺伝子解析による単離株の種の推定

病原性細菌を分類することは、古くからその病原性の評価や治療において非常に重要な操作であり、現在までに様々な細菌の分類手法が提案されてきた。分子生物学が発達する以前から、コロニー形状や色、運動性、基質の利用特性、グラム染色といった特徴によって細菌を分類することができ⁷³、それは現在においてもしばしば用いられる。しかし、この伝統的な手法では時間と手間がかかる点、環境条件の違いによる培養のばらつきがあい

まいな結果をもたらす可能性がある点、同定のために純粋培養が必要であり培養不可能な細菌等は同定できない点などから、分子生物学の発達した現在においては、表現型に基づくこれらの古典的な手法に代わる様々な手法が提案されている⁷⁴。現在細菌の識別に使用される手法には、①形態的な同定、②細胞壁の脂肪酸の構成に基づく同定、③16S rRNA 遺伝子の配列に基づく同定、④Matrix-Assisted Laser Desorption / Ionization Time-Of-Flight Mass Spectrometer (MALDI-TOF MS) によるタンパク質プロファイリングに基づく同定、⑤Biolog (Biolog 社) を利用した生化学的同定システムに基づく同定、⑥リボタイピングに基づく同定、⑦全ゲノムシーケンシングに基づく同定があげられる⁷⁴。

本研究では、単離した溶血性細菌について③の 16S rRNA 遺伝子配列に基づき種を推定した。16S rRNA は約 1500bp の遺伝子で、すべての細菌が保有する。種の中で 16S rRNA 遺伝子は高く保存されており、同一門であれば 80%以上、同一属であれば 95%以上一致、同一種であれば 97%以上一致していると一般的に考えられている⁷⁵。16S rRNA 遺伝子には、細菌間で配列が大きく異なる 9 つの可変領域が散在しており、それぞれ V1-V9 まで名前が付けられている。多くの研究では、V4 や V6 といった単一の可変領域や、V1-V3、V3-5、といった複数の可変領域を対象としてシーケンシングを行う。ただし、このような 16S rRNA 遺伝子の部分領域をターゲットとした場合、分類学上の十分な分解能は得られないことに留意する必要がある⁷⁶。

溶血斑形成の確認ができた 137 単離細菌株と、単離時には溶血活性を持つと判断したがその後溶血活性を確認できなかった 30 単離細菌株の種を対象とした。16S ribosomal RNA (16S rRNA) 遺伝子の V3-4 領域のシーケンシングを行い、16SrRNA 遺伝子データベースと BLAST 相同検索をして、対象株の種推定を行った。

BLAST の結果を表 5 に示す。推定種が複数あるものについては複数種を載せている。また、溶血斑の様子についても併せて記載しており、Clear (明瞭)、Indistinct (不明瞭)、Discolored (変色) のそれぞれの溶血斑を示した株および No hemolysis (単離後に溶血斑を再確認できなかった) 株の数を示している。

最も多く推定された種は大腸菌 (*Escherichia coli*) で、溶血性を示すものが 34 株推定された。ただし、同時に *Brenneria alni*、*Escherichia fergusonii*、*Shigella flexneri*、*Shigella sonnei* の 16S rRNA 遺伝子配列との相同性も同程度に高く、どの種であるかまでは絞り込むことはできなかった。複数種が推定された理由は、今回解析で使用した 16S rRNA の V3-4 領域が約 465 塩基と短い長さで構成される配列で、種間の配列の類似度がかなり高かったためであると考えられる。*Aeromonas* 属と推定された株は多く、73 株が推定された。特に *Aeromonas encheleia*、*Aeromonas hydrophila*、もしくは *Aeromonas molluscorum* と推定された溶血性細菌株が 31 株、*Aeromonas caviae*、*Aeromonas hydrophila*、もしくは *Aeromonas jandaei* と推定された溶血性細菌株が 23 株と多かった。その他に *Serratia* 属、*Citrobacter* 属、*Klebsiella* 属、*Leuconostoc* 属、*Enterobacter* 属、*Pseudomonas* 属、*Pseudoxanthomonas* 属、*Enterococcus* 属、*Streptococcus* 属、*Lactococcus* 属と推定された株が見られた。「明瞭な溶血斑」を示した株は、*Aeromonas* 属、もしくは、*Brenneria alni*、*Escherichia coli*、*Escherichia fergusonii*、*Shigella flexneri*、*Shigella sonnei* のいずれかと推定された。「変色した溶血斑」を示す株は、*Citrobacter freundii*、*Citrobacter braakii*、*Morganella morganii*、もしくは *Buttiauxella agrestis*、*Buttiauxella noackiae*、*Citrobacter braakii*、*Kluyvera intermedia* のいずれかと推定された。単離後に溶血斑を再確認できなかった株は、合計すると 30 株あり、最も多かったのは *Escherichia coli* 等に分類された株で 8 株

あった。この中には、単離時に誤って単離されてしまった溶血活性が全くない株や、わずかに溶血活性を持つが実験条件の違いが溶血活性に影響を与え、溶血斑を検出することが難しい株が含まれると考えられる。

さらに単離株の系統的な分類のため、溶血活性を示す単離細菌 136 株の 16S rRNA 遺伝子 V3-4 領域の系統解析を行った (図 8)。枝の横方向の長さは進化距離を表しており、サイトごとの塩基置換数を単位としている。各枝の信頼性を表す指標として、500 回の bootstrap 検定を行い得られた bootstrap 値をノードの横の数値として表す。BLAST 相同検索の結果推定された種名を末端ノードに記載する。複数の種の配列と高い相同性を示した株については、複数種を記載している。推定種が同一かつ系統樹上で隣の株はまとめて、括弧内の数値で末端ノードに含まれる株数を表す。明瞭な溶血活性を示した単離株については、血液寒天プレートタイプ I で 1 日培養したときの溶血斑幅の大きさを溶血活性の大きさとして、右上のストリッププロットで表す。明瞭な溶血斑を示した株が 8 株以上のグループについては箱ひげ図も重ねて表示する。図の右に、各グループ内で明瞭な溶血斑を示した株の数を表す。

系統解析の結果、単離した溶血性細菌株は *Aeromonas* 属、大腸菌 (もしくはその近縁株)、*Citrobacter* 属、*Pseudomonas* 属、*Pseudoxanthomonas* 属等を含むグループと、*Leuconostoc* 属、*Enterococcus* 属、*Lactococcus* 属、*Streptococcus* 属等を含むグループの 2 つに大別された。これは前者のグループがグラム陰性菌のグループで、後者のグループがグラム陽性菌のグループであった。明瞭な溶血斑を示した株は *Aeromonas* 属に分類される系統と、*Brenneria alni*、*Escherichia coli*、*Escherichia fergusonii*、*Shigella flexneri*、*Shigella sonnei* に分類された系統の 2 つに分類できた。変色した溶血斑を示した 4 株 (*Citrobacter freundii*、*Citrobacter braakii*、*Morganella morganii*、*Buttiauxella agrestis*、*Buttiauxella noackiae*、*Kluyvera intermedia* のいずれか) はいずれも、明瞭な溶血斑を示した *Brenneria alni*、*Escherichia coli*、*Escherichia fergusonii*、*Shigella flexneri*、*Shigella sonnei* のグループと系統的に近かった。

溶血活性に着目すると、*Aeromonas* 属の中でも *Aeromonas encheleia*、*Aeromonas hydrophila*、もしくは *Aeromonas molluscorum* と推定されたグループの株の溶血活性が特に大きかった。このグループに属する株は、最小で 0.39mm、最大で 3.77mm の溶血斑幅を示す株があり、グループ内でも溶血活性に差が見られた。また、前述したグループほどばらつきはないが、*Brenneria alni*、*Escherichia coli*、*Escherichia fergusonii*、*Shigella flexneri*、*Shigella sonnei* でも最小で 0.19mm、最大で 1.05mm の溶血斑幅を示し、ばらつきが見られた。

表 5 単離した溶血性細菌の 16S rRNA 遺伝子 V3-4 領域から推定した種とその数

Species	strain	Hemolysis			
		Clear	Indistinct	Discolored	No hemolysis
<i>Aeromonas encheleia</i> / <i>Aeromonas hydrophila</i> / <i>Aeromonas molluscorum</i>	n = 31	30	1	0	0
<i>Aeromonas popoffii</i>	n = 8	4	4	0	0
<i>Aeromonas sanarellii</i>	n = 1	1	0	0	0
<i>Aeromonas dhakensis</i> / <i>Aeromonas taiwanensis</i>	n = 6	4	2	0	0
<i>Aeromonas caviae</i> / <i>Aeromonas hydrophila</i> / <i>Aeromonas jandaei</i>	n = 27	8	15	0	4
<i>Serratia surfactantfaciens</i>	n = 1	0	1	0	0
<i>Citrobacter freundii</i>	n = 2	0	0	2	0
<i>Citrobacter braakii</i>	n = 1	0	0	1	0
<i>Buttiauxella agrestis</i> / <i>Buttiauxella noackiae</i> / <i>Citrobacter braakii</i> / <i>Kluyvera intermedia</i>	n = 1	0	0	1	0
<i>Enterobacter cloacae</i> / <i>Enterobacter kobei</i> / <i>Enterobacter ludwigii</i> / <i>Leclercia adecarboxylata</i> / <i>Pantoea agglomerans</i>	n = 1	0	1	0	0
<i>Klebsiella pneumoniae</i>	n = 1	0	1	0	0
<i>Enterobacter asburiae</i> / <i>Enterobacter cloacae</i> / <i>Enterobacter hormaechei</i>	n = 3	0	1	0	2
<i>Klebsiella pneumoniae</i> / <i>Klebsiella quasipneumoniae</i>	n = 4	0	1	0	3
<i>Proteus mirabilis</i>	n = 1	0	1	0	0
<i>Morganella morganii</i>	n = 1	0	0	1	0
<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>	n = 42	30	4	0	8
<i>Pseudomonas sesami</i>	n = 1	0	1	0	0
<i>Pseudoxanthomonas sacheonensis</i>	n = 1	0	1	0	0
<i>Leuconostoc lactis</i>	n = 1	0	1	0	0
<i>Enterococcus casseliflavus</i>	n = 2	0	2	0	0
<i>Enterococcus gallinarum</i>	n = 1	0	1	0	0
<i>Enterococcus lemanii</i>	n = 1	0	1	0	0
<i>Enterococcus durans</i> / <i>Enterococcus faecium</i> / <i>Enterococcus hirae</i> / <i>Enterococcus ratti</i> / <i>Enterococcus villorum</i>	n = 2	0	2	0	0
<i>Enterococcus faecalis</i>	n = 1	0	1	0	0
<i>Lactobacillus salivarius</i>	n = 1	0	1	0	0
<i>Streptococcus cristatus</i>	n = 1	0	1	0	0
<i>Lactococcus lactis</i>	n = 3	0	1	0	2
<i>Lactococcus taiwanensis</i>	n = 1	0	1	0	0
<i>Lactococcus formosensis</i> / <i>Lactococcus garvieae</i>	n = 1	0	1	0	0
<i>Streptococcus equinus</i> / <i>Streptococcus lutetiensis</i>	n = 1	0	1	0	0
<i>Streptococcus porcorum</i> / <i>Streptococcus sanguinis</i> / <i>Streptococcus sinensis</i>	n = 11	0	7	0	4
<i>Citrobacter cronae</i> / <i>Enterobacter cloacae</i> / <i>Enterobacter kobei</i> / <i>Enterobacter ludwigii</i> / <i>Leclercia adecarboxylata</i> / <i>Pantoea agglomerans</i>	n = 2	0	0	0	2
<i>Acinetobacter vivianii</i>	n = 1	0	0	0	1
<i>Raoultella ornithinolytica</i>	n = 2	0	0	0	2
<i>Phytobacter diazotrophicus</i> / <i>Pseudoescherichia vulneris</i> / <i>Yokenella regensburgei</i>	n = 1	0	0	0	1
<i>Kluyvera cryocrescens</i>	n = 1	0	0	0	1
Total	n = 167	77	55	5	30

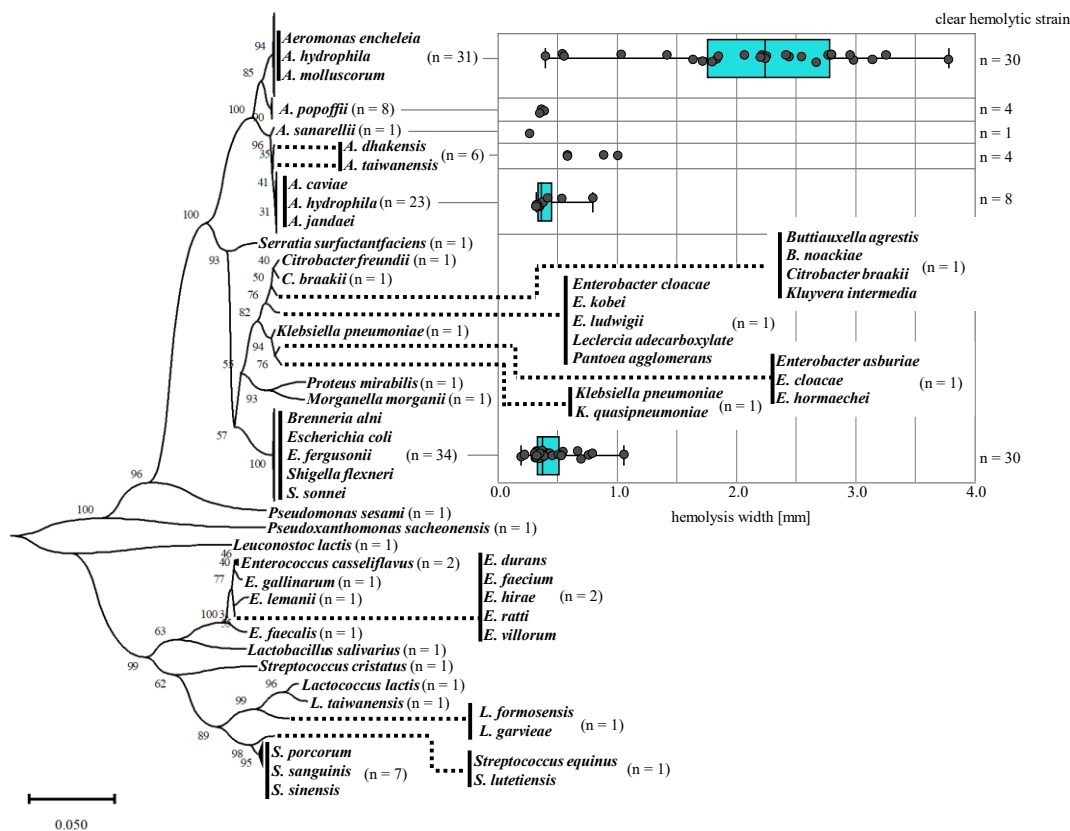


図8 単離細菌株の16S rRNA 遺伝子 V3-4 領域の系統樹、推定種および明瞭な溶血斑を示す株の溶血活性の大きさ

3.1.5. 16S rRNA 遺伝子 V3-4 領域のアレルと溶血活性

Aeromonas 属の細菌株グループや、大腸菌、*Brenneria alni*、*Escherichia fergusonii*、*Shigella flexneri*、*Shigella sonnei* と推定された細菌株グループ内で、16S rRNA 遺伝子配列はよく似ており、それぞれの株から最も多く検出された代表配列に基づく分類は、溶血活性の違いを説明できる十分な分解能を得ることができなかった。そこで、16S rRNA 遺伝子の対立遺伝子（アレル）に着目して、単離細菌株をさらに細かく分類することを試みた。大腸菌をはじめ多くの細菌は染色体上に同じ遺伝子を複数保有することがある。例えば、大腸菌は16S rRNA 遺伝子を染色体上に7個保有する⁷⁷。それぞれの遺伝子の塩基配列は必ずしも同一ではなく異なる配列を持つことがあり、このような遺伝子のバリエーションをアレルと呼ぶ。16S rRNA は細菌のタンパク質合成に関わる重要な遺伝子で、ほとんどの細菌は16S rRNA 遺伝子を複数保有する。16S rRNA 遺伝子配列は種で高く保存されることから、細菌の系統によって16S rRNA 遺伝子のアレルの構成割合も異なることが考えられる。16S rRNA 遺伝子アレルの構成割合に着目することで、単離細菌株をより細かく分類できると考えた。

16S rRNA 遺伝子の数は細菌種によって異なることから、全ゲノム配列が明らかとなっている *Escherichia coli* と *Aeromonas* 属の細菌株がゲノム上に保有する16S rRNA 遺伝子の数を調べた。次に、3.1.4 で *Brenneria alni*、*Escherichia coli*、*Escherichia fergusonii*、*Shigella flexneri*、*Shigella sonnei* のいずれかであると分類された菌株と、

Aeromonas に分類された菌株のそれぞれについて、16S rRNA 遺伝子のアレルの構成割合に基づきこれらの株を分類し、溶血活性との関係調べた。解析では、シーケンシングによって得られた 16S rRNA V3-V4 領域の塩基配列 (約 430 bp) からユニークな配列を取り出し、それぞれの株についてユニーク配列の構成割合を計算した。また、ユニーク配列間の類似性を見るため、ユニーク配列の系統解析を行った。さらに、ユニーク配列の構成割合を元に階層クラスタリングを行い、それぞれの株を分類した。

3.1.5.1. 大腸菌と *Aeromonas* 属の細菌株がゲノム上で保有する 16S rRNA 遺伝子の数

NCBI データベース上に登録されている、全ゲノム配列の明らかな大腸菌 2 株と *Aeromonas* 属 5 株の保有する 16S rRNA 遺伝子の保有数を調べたところ、表 6 のようになった。大腸菌は K-12 株と O157:H7 Sakai 株について調べた。いずれの株も 16S rRNA 遺伝子の保有数は 6 つであった。O157:H7 Sakai 株は 2 つのプラスミド pOSK1 と pO157 を保有することが知られるが、いずれのプラスミド上においても 16S rRNA 遺伝子は確認されなかった。*Aeromonas* 属は 5 つの種について調べた。*A. hydrophila*、*A. caviae*、*A. dhakensis*、*A. jandaei* は 9 つの 16S rRNA 遺伝子を保有することが確認された。*A. sanarellii* は 8 つの 16S rRNA 遺伝子を保有しており、他の 4 種より 1 つ少なかった。以上のことから、大腸菌は約 6 個、*Aeromonas* 属は約 8~9 個の 16S rRNA 遺伝子をゲノム上に保有することが示唆された。

表 6 それぞれの細菌種がゲノム上に保有する 16S rRNA 遺伝子の数

Species	Strain	Name	accession	Number of 16SrRNA
<i>Escherichia coli</i>	K-12 substr. MG1655	-	NC_000913.3	6
	O157:H7 str. Sakai	-	NC_002695.2	6
		pOSK1	NC_002127.1	0
		pO157	NC_002128.1	0
<i>Aeromonas hydrophila</i>	ASM1731021v1	-	NZ_CP050851.1	9
<i>Aeromonas caviae</i>	NCTC12244	-	NZ_LS483441.1	9
<i>Aeromonas sanarellii</i>	NS1	-	NZ_CP079751.1	8
<i>Aeromonas dhakensis</i>	1706-28330	-	NZ_CP054854.1	9
<i>Aeromonas jandaei</i>	FDAARGOS_986	-	NZ_CP066092.1	9

3.1.5.2. *Aeromonas* 属に分類された株

3.1.4 で *Aeromonas* 属に分類された株の、16S rRNA 遺伝子 V3-V4 領域のアレル構成割合を図 9 に示す。図のグラフは左から、単離株名、溶血活性、16S rRNA 遺伝子 V3-V4 領域のユニークアレルの構成割合、ユニークアレルの構成割合に基づく単離株のクラスタリングの結果を表す。溶血活性は、血液寒天プレートタイプ I で 24 時間培養したときの溶血斑幅の平均値 (溶血斑が不明瞭なものは Indistinct と記載) を表す。16S rRNA 遺伝子アレルは系統解析して並び替えた。系統解析の結果は図の上方に併せて記載する。ユニークアレルの構成割合に基づき、Ward 法で細菌株のクラスタリングを行い、細菌株の順を並び変えた。Ward 法によるクラスタリングの結果は図の右方に併せて記載する。ユニークアレルは、いずれかの株で構成割合が 5% 以上のものだけを表示する。それぞれのユニークアレ

ルの BLAST 相同検索によって推定される種については表 7 に記載する。また、アレルのそれぞれの配列は付録資料 1.1 に記載する。

解析の結果、*Aeromonas* 属に分類された株では、いずれかの株で構成割合が 5%以上のアレルは 26 種類得られた。ほとんどの株でアレル 00 もしくはアレル 01 を保有していた。細菌株のクラスタリングの結果、アレル 01 を保有する株とアレル 01 を保有しない株で大きく 2 つに分類された。アレル 01 を保有しない株のほとんどはアレル 00 を保有していた。興味深いことに、アレル 01 を保有する株は溶血活性の大きい株が多かったのに対して、アレル 01 を保有しない株は溶血活性の小さい株が多かった。表 7 より、アレル 01 を保有する株は *Aeromonas encheleia*、*Aeromonas hydrophila*、もしくは *Aeromonas molluscorum* と推定された。アレル 00 を多く保有する株は *Aeromonas caviae*、*Aeromonas hydrophila*、もしくは *Aeromonas jandaei* と推定された。また、WWo048 をはじめ複数の株でアレル 05 が検出された。アレル 05、アレル 22、アレル 45 は *Aeromonas popoffii* と推定されるアレルで、これらの構成割合の高い WWo048 や WWo065、WWo161 等は 3.1.4 で *Aeromonas popoffii* と推定された。また、16S rRNA 系統解析の結果、アレル 05 に近いアレル 22 を保有する WWo007 も *Aeromonas popoffii* に推定されていた。アレル 02 の割合が高い WWo011、WWo012、WWo039、WWo062 はいずれも *Aeromonas dhakensis* もしくは *Aeromonas taiwanensis* と推定された株であった。これらの株の多くはアレル 00 も保有しており、アレル 00 を保有する株と系統的に近いことが示唆される。また、WWo096、WWo044、WWo038、WWo041 等のいくつかの株では、アレル 02、アレル 00 に加えてアレル 05 を保有していた。アレル 02 は *Aeromonas dhakensis* もしくは *Aeromonas taiwanensis* と推定されるアレルで、アレル 00 は *Aeromonas caviae*、*Aeromonas hydrophila*、もしくは *Aeromonas jandaei* と推定されるアレルで、アレル 05 は *Aeromonas popoffii* と推定されるアレルで、すべてのアレルで推定される種が異なっていた。各アレルの検出割合は同程度の値であることがあり、例えば WWo040 の場合、アレル 02、00、05 はそれぞれ 36%、30%、32% で検出されいずれも近い値を示した。これらのアレルを複数種類保有する株は 16S rRNA 遺伝子 V3-V4 領域配列から種を推定した場合誤って推定される可能性が示唆される。

アレル 08 は、他のアレルと異なり、*Escherichia coli*、*Escherichia fergusonii*、*Pseudoescherichia vulneris*、*Shigella flexneri*、もしくは *Shigella sonnei* と推定されるアレルで、*Aeromonas* 属以外の細菌種に推定されるアレルであった。これを保有する株は WWo066 のみであった。異なる種に推定される 16S rRNA 遺伝子が同一株から検出された例はこれまでにないため、WWo066 についてはほかの菌株が混入し、誤って検出された可能性が高い。

3.1.5.1 で、*Aeromonas* 属の保有する 16S rRNA 遺伝子の個数が 8-9 個であることが示唆された。*Aeromonas* 属が保有する 16S rRNA 遺伝子の個数が 8-9 個であると仮定した場合、16S rRNA 遺伝子 1 個あたりの 16S rRNA の構成割合は 11.1-12.5% になる。このことから、たとえば WWo094 はアレル 01 と 42 の構成割合がそれぞれ 85% と 11% であるが、アレル 01 を 7、もしくは 8 個、アレル 42 を 1 個保有することが示唆される。

アレルの系統解析の過程で行ったマルチプルアラインメントの結果を見ると 100 番目から 119 番目の塩基で変異のバリエーションが多く見られた (付録資料 1.1)。アレル 00 とアレル 01 は分子系統樹の中で離れた位置に存在するが、100 番目から 119 番目の塩基変異のバリエーションによって、分けられたと考えられる。

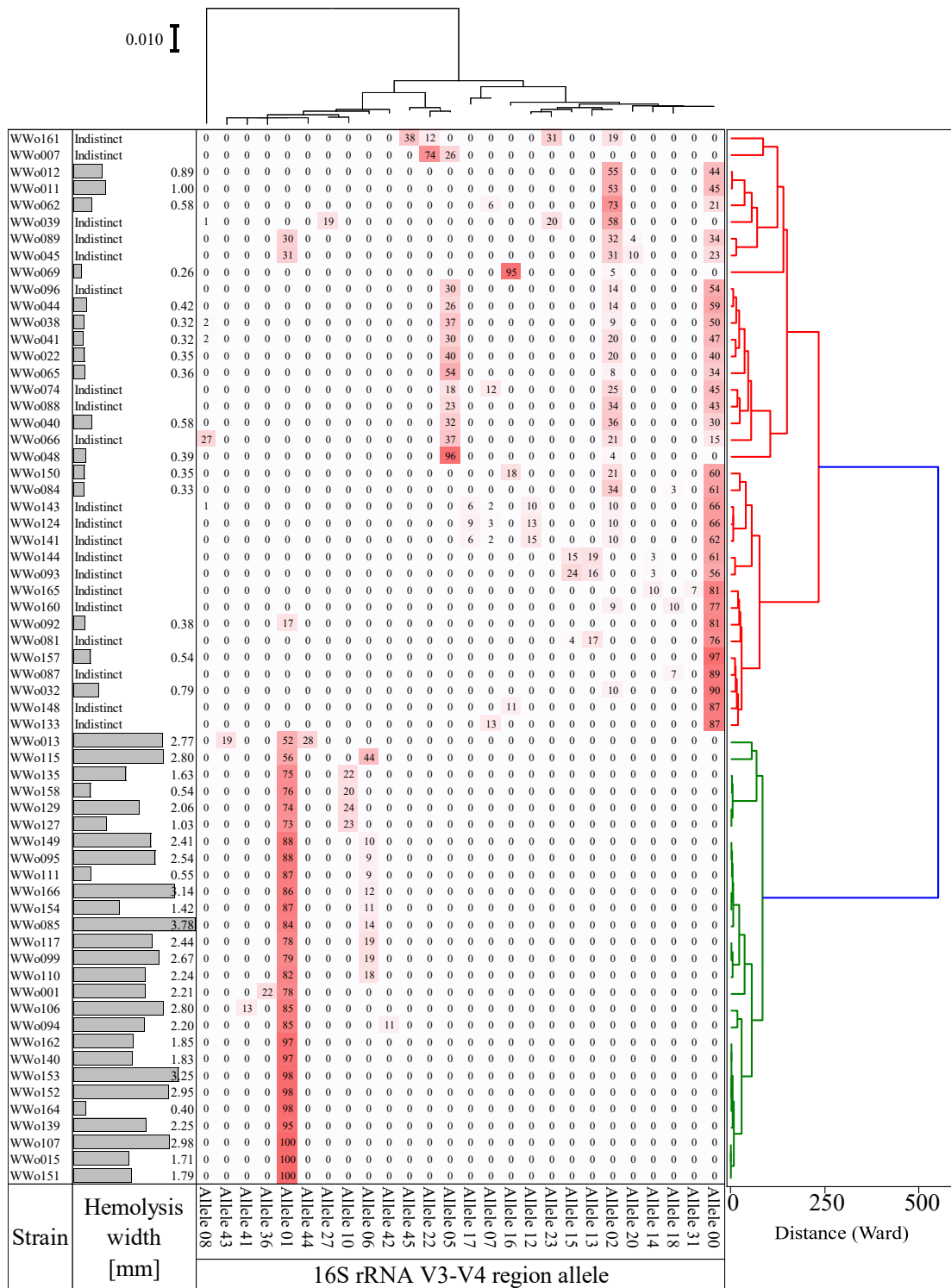


図9 *Aeromonas* 属に分類された株の 16S rRNA 遺伝子アレルの構成割合 [%] とアレルの分子系統樹および単離株のアレル構成に基づくクラスタリング

表 7 *Aeromonas* 属細菌株の各アレルから推定された種

Allele	Predicted species
Allele 00	<i>Aeromonas caviae</i> / <i>Aeromonas jandaei</i> / <i>Aeromonas hydrophila</i>
Allele 31	<i>Aeromonas caviae</i> / <i>Aeromonas jandaei</i> / <i>Aeromonas hydrophila</i>
Allele 18	<i>Aeromonas caviae</i> / <i>Aeromonas jandaei</i> / <i>Aeromonas hydrophila</i>
Allele 14	<i>Aeromonas enteropelogenes</i>
Allele 20	<i>Aeromonas dhakensis</i> / <i>Aeromonas taiwanensis</i>
Allele 02	<i>Aeromonas dhakensis</i> / <i>Aeromonas taiwanensis</i>
Allele 13	<i>Aeromonas dhakensis</i> / <i>Aeromonas sanarellii</i> / <i>Aeromonas taiwanensis</i>
Allele 15	<i>Aeromonas dhakensis</i> / <i>Aeromonas enteropelogenes</i> / <i>Aeromonas sanarellii</i> / <i>Aeromonas taiwanensis</i>
Allele 23	<i>Aeromonas dhakensis</i> / <i>Aeromonas enteropelogenes</i> / <i>Aeromonas taiwanensis</i>
Allele 12	<i>Aeromonas caviae</i> / <i>Aeromonas jandaei</i> / <i>Aeromonas hydrophila</i>
Allele 16	<i>Aeromonas sanarellii</i>
Allele 07	<i>Aeromonas veronii</i>
Allele 17	<i>Aeromonas veronii</i>
Allele 05	<i>Aeromonas popoffii</i>
Allele 22	<i>Aeromonas popoffii</i>
Allele 45	<i>Aeromonas popoffii</i>
Allele 42	<i>Aeromonas hydrophila</i>
Allele 06	<i>Aeromonas encheleia</i> / <i>Aeromonas hydrophila</i> / <i>Aeromonas molluscorum</i>
Allele 10	<i>Aeromonas aquatica</i> / <i>Aeromonas media</i> / <i>Aeromonas tecta</i>
Allele 27	<i>Aeromonas aquatica</i> / <i>Aeromonas encheleia</i> / <i>Aeromonas media</i> / <i>Aeromonas molluscorum</i> / <i>Aeromonas tecta</i>
Allele 44	<i>Aeromonas encheleia</i> / <i>Aeromonas hydrophila</i> / <i>Aeromonas molluscorum</i>
Allele 01	<i>Aeromonas encheleia</i> / <i>Aeromonas hydrophila</i> / <i>Aeromonas molluscorum</i>
Allele 36	<i>Aeromonas encheleia</i> / <i>Aeromonas hydrophila</i> / <i>Aeromonas molluscorum</i>
Allele 41	<i>Aeromonas encheleia</i> / <i>Aeromonas hydrophila</i> / <i>Aeromonas molluscorum</i>
Allele 43	<i>Aeromonas encheleia</i> / <i>Aeromonas hydrophila</i> / <i>Aeromonas molluscorum</i>
Allele 08	<i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Pseudoescherichia vulneris</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>

3.1.5.3. *Brenneria alni*, *Escherichia coli*, *Escherichia fergusonii*, *Shigella flexneri*, *Shigella sonnei* と分類された株

3.1.4 で *Brenneria alni*, *Escherichia coli*, *Escherichia fergusonii*, *Shigella flexneri*, *Shigella sonnei* と分類された株の、16S rRNA 遺伝子 V3-V4 領域のアレル構成割合を図 10 に示す。*Aeromonas* 属のアレル構成の図と同じように、図のグラフは左から、単離株名、溶血活性、16S rRNA 遺伝子 V3-V4 領域のユニークアレルの構成割合、ユニークアレルの構成割合に基づく単離株のクラスタリングの結果を表す。溶血活性は、血液寒天プレートタイプ I で 24 時間培養したときの溶血斑幅の平均値（溶血斑が不明瞭なものは Indistinct と記載）を表す。16S rRNA 遺伝子 V3-V4 領域のアレルは系統解析して並び替えて表示してある。系統解析の結果は図の上方に併せて記載する。ユニークアレルの構成割合に基づく単離株のクラスタリングでは、Ward 法で細菌株のクラスタリングを行い、行を並び変えて表示してある。Ward 法によるクラスタリングの結果は図の右方に併せて記載する。それぞれのユニークアレルから推定される種は図 10 に記載する。また、アレルのそれぞれの配列は付録資料 1.2 に記載する。

解析の結果、いずれかの株で構成割合が 5%以上のユニークなアレルが 7 種類得られた。単離株はいずれもアレル 100 を保有することが分かった。溶血活性に着目すると、アレルの構成割合でクラスタリングした結果と、溶血活性の大小に関する関係性を見つけることはできなかった。

3.1.5.1 の結果から、これらの株が大腸菌で染色体上に 16S rRNA 遺伝子を 7 個保有すると仮定すると、1 個あたりの 16S rRNA の構成割合は 14.3%になる。このとき、WWo167、WWo024、WWo147、WWo075 は、アレル 100 を 6 個、アレル 100 以外のアレルを 1 個保

有していることが示唆される。また、WWo112とWWo003はアレル100を5個、アレル100以外のアレルを2種類、それぞれ1個ずつ保有していることが示唆される。2個あたりの16S rRNAの構成割合は28.6%となることから、WWo025はアレル104を2個保有していることが示唆される。3個あたりの16S rRNAの構成割合は42.9%となることから、WWo050、WWo083はアレル104を3個保有していることが示唆される。4個あたりの16S rRNAの構成割合は57.1%となることから、WWo023はアレル104を4個保有していることが示唆される。

アレルの系統解析の結果、アレル100とアレル104、106、107、108、114、109のそれぞれの間の進化距離は同じであった。系統解析の過程で行うマルチプルアラインメントの結果を見ると、アレル100に対する塩基置換は1塩基にとどまっていた（付録資料1.2）。アレル100以外のアレルは、アレル100に対する塩基置換で考えた時に最も置換数が小さくなることから、いずれのアレルもアレル100に由来して変異の入ったアレルであると考えられる。いずれのアレルからも推定される種は同じであった（表8）。アレルの構成割合と溶血活性の関係については、今回の解析からは見つからなかった。

表8 *Brenneria alni*、*Escherichia coli*、*Escherichia fergusonii*、*Shigella flexneri*、*Shigella sonnei*に分類された株の各アレルから推定された種

Allele	Predicted species
Allele 109	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Pseudescherichia vulneris</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>
Allele 114	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Pseudescherichia vulneris</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>
Allele 108	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Pseudescherichia vulneris</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>
Allele 107	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Pseudescherichia vulneris</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>
Allele 106	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Pseudescherichia vulneris</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>
Allele 100	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Pseudescherichia vulneris</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>
Allele 104	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Pseudescherichia vulneris</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>

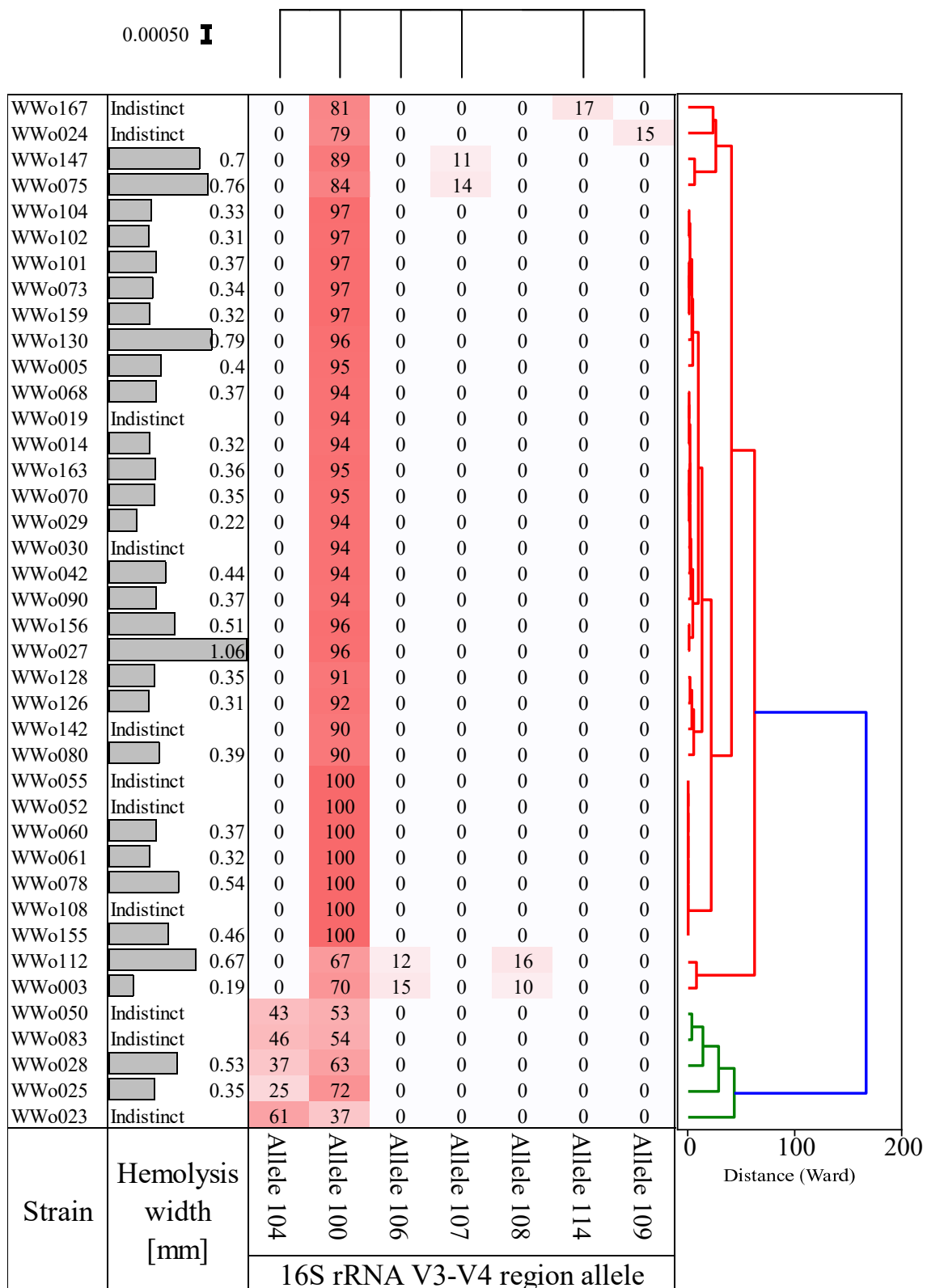


図 10 *Brenneria alni*、*Escherichia coli*、*Escherichia fergusonii*、*Shigella flexneri*、*Shigella sonnei* と分類された溶血性細菌株の 16S rRNA 遺伝子アレルの構成割合 [%] とアレルの分子系統樹および単離株のアレル構成に基づくクラスタリング

3.2. 溶血性大腸菌

3.1.4で *Brenneria alni*、*Escherichia coli*、*Escherichia fergusonii*、*Shigella flexneri*、*Shigella sonnei* に分類された溶血性細菌は 34 株あった。大腸菌 (*Escherichia coli*) は、糞便汚染指標としても広く使われ、下水から最も多く単離される細菌の一つで、単離したこれらの株の多くは大腸菌である可能性が高い。溶血性大腸菌は、前述のとおり大腸がんや尿路感染症など様々な病気に関わっていることが知られている。また、大腸菌は生物学におけるモデル生物として、遺伝子解析をはじめ様々な研究が行われており、保有する病原遺伝子の違いにより細かく分類することができる。ここでは、一部の大腸菌と推定された単離株を対象として大腸菌の分類、種の同定、保有する病原性遺伝子について解析した結果について報告する。

3.2.1. 血液寒天培地で単離される大腸菌の遺伝的特徴

3.2.1.1. 全ゲノムシーケンスに基づく種の同定

3.1.4で述べたように、16S rRNA 遺伝子の V3-V4 領域の解析では分類学上の十分な解像度を得ることはできなかった。そこで、全ゲノムシーケンスを行った株については前述の方法とは異なる手法で種を同定することとした。

細菌の種同定のゴールドスタンダードは、全ゲノム DNA-DNA ハイブリダイゼーションである⁷⁸。DNA-DNA ハイブリダイゼーションの結果 70%以上のものが同一種と判断される。しかし、DNA-DNA ハイブリダイゼーションは実験の手間がかかり、技術的にも困難であるため、この手法は現代ではあまり用いられることはない。そこで、16S rRNA 遺伝子配列の類似性が 97%を下回る株が、DNA-DNA ハイブリダイゼーションで 70%を超える類似性を示す例がないことから、同一種とは DNA-DNA 結合率が 70%より大きい、もしくは 16S リボソーム RNA (rRNA) 遺伝子配列の同一性が 97%以上であること⁷⁹と考えられるようになり、16S rRNA 遺伝子配列が調べられるようになった。近年においては、Average Nucleotide Identity (ANI) が DNA-DNA ハイブリダイゼーションの結果と強い相関関係があり、ANI 値が 95%以上であれば従来の DNA-DNA ハイブリダイゼーションの 70%閾値に相当する⁸⁰ことから、ANI が種同定のスタンダードとして用いられることが多い。

一部の株について全ゲノムシーケンスを行い、正確な種の同定を行った。全ゲノムシーケンスを行った株を表 9 に示す。これらの株はまず、Ribosomal Mutilocus Sequence Typing (rMLST)⁶² を用いて大まかな種の推定を行い、推定された種の基準株との ANI を算出した。本研究では ANI の計算には fastANI⁶³ を使用した。BLAST 等のアラインメントベースの手法が用いた ANI の計算は、計算に時間がかかることが問題とされていたが、fastANI では 80-100%の値の範囲において、従来の ANI 計算手法と同等の値を算出しながら計算速度が大幅に改善されている⁶³。さらに、ANI では区別のできない *Escherichia coli* と *Shigella* 属については、Kmer ID を用いた解析⁶⁴ と MLST 解析を行い、それぞれの種の識別を行った。

全ゲノムシーケンスを行った株を表 9 に示す。3.1.4で *Brenneria alni*、*Escherichia coli*、*Escherichia fergusonii*、*Shigella flexneri*、*Shigella sonnei* に分類された溶血性細菌では WWo003、WWo005、WWo014、WWo027、WWo055、WWo070、WWo073、WWo083、WWo102、WWo108、WWo147、WWo156 の 12 株を対象とした。WWo055 と WWo108 を除き明瞭な溶血斑を確認することができた。WWo055 は単離後に溶血斑を確認することができなかった。WWo102 は、血液寒天プレートタイプ I とタイプ II で溶血斑を確認するこ

とができなかったが、血液寒天培地タイプ III でわずかに溶血斑を確認することができたので、不明瞭な溶血斑を持つと判断した (図 11)。KKo007、KOr014、KKa004 株は、五味良太先生 (京都大学) から譲渡して頂いた、河川から単離された大腸菌株でいずれも溶血活性を持つ。溶血活性は、血液寒天プレートタイプ II で 24 時間培養したときの溶血斑幅を表している。

rMLST による種の推定

まず、Ribosomal Multilocus Sequence Typing (rMLST) ⁶² を用いて、種を推定した。rMLST では、53 個の *rps* 遺伝子のバリエーションから種を推定する。今回対象とした 15 株いずれの株のシーケンスデータにおいても、53 個すべての *rps* 遺伝子の保有が確認された。rMLST で推定された種は、WWo003, WWo005, WWo014, WWo027, WWo055, WWo070, WWo073, WWo083, WWo102, WWo108, WWo147, WWo156, KKo007, KOr014, KKa004 のいずれの株も *Escherichia coli* であった。

ANI の計算に基づく種の同定

次に、average nucleotide index (ANI) の計算に基づいて種の同定を行った。単離株のゲノム配列と、*Escherichia coli* の基準株 (type strain) 2 株およびその他の *Escherichia* 属 5 種、*Shigella* 属 4 種の基準株の配列を比較して、それぞれの ANI を計算した。ANI は、種の境界を評価する堅実な手段として近年頻繁に用いられており、同一種に属する生物は通常 95-96% 以上の ANI を示すことが知られている ⁸⁰⁻⁸²。ただし、*Shigella* 属と *Escherichia coli* は ANI によって区別することができないため、後述の kmer ID によって識別した。

ANI の計算の結果、いずれの株も *Escherichia coli*、もしくは *Shigella* 属と推定された。*Escherichia coli* と *Shigella* 属の type strain との ANI 値はいずれも 96% 以上であった (表 10)。一方、*Escherichia coli* 以外の *Escherichia* 属の基準株との ANI 値は、*Escherichia alba* ではいずれも約 80%、そのほかの *Escherichia albertii*、*Escherichia fergusonii*、*Escherichia marmotae*、*Escherichia russiae* では 90-93% で、95% を下回る値であった。このことから、単離株はいずれも *Escherichia coli*、もしくは *Shigella* 属と推定された。

Kmer ID と MLST 解析による *Escherichia coli* と *Shigella* 属の識別

Escherichia coli と *Shigella* 属の識別のため、kmer ID ⁶⁴ を用いた解析と MLST 解析を行った。*Shigella* 属は、生化学的特徴や臨床的関連性に基づき大腸菌から分類されている。系統的には大腸菌であり、分子生物学的には同一種としてみることもできるため、その識別が難しい。全ゲノムシーケンスを用いた *Escherichia coli* と *Shigella* 属の識別方法では、k-mers や SNPs の解析に基づく方法が、高い識別能力を持つことが指摘されている ⁸³。特に k-mers を用いた識別では、k-mers を株間の進化的距離として算出でき、種のレベルで分類することができる。kmer ID を用いた方法 ⁶⁴ では、98% 以上の精度で *Escherichia coli* と *Shigella* 属の識別が可能であるが、一部の株で誤認される。kmer ID で誤認される株は、MLST 解析で ST270、ST1751、ST1767 に分類される。ST270 は、腸管侵襲性大腸菌 (EHEC) であり、kmer ID により *Shigella flexneri* と誤認される。ST1751 と ST1767 は *Shigella boydii* で、kmer ID により *Shigella dysenteriae* と誤認される可能性がある。これは、生化学的同定法及び血清型による同定法において用いられる *S. boydii* 血清型 9 (ST1751) と *S. boydii* 血清型 12 (ST1767) が、系統的に *S. dysenteriae* と *S. boydii* の中間に位置するためであると考えられている ⁶⁴。本研究では、kmer ID による分類を行い、MLST 解析で上記の ST 番号に属していないことを確認した。

表 11 で、kmer ID を用いた解析結果と MLST 解析の結果を表す。Strain の列は単離株名、Reference strain with highest similarity の列は *Escherichia coli* 及び *Shigella* 属の基準株との類似度 (similarity) が最も高かった基準株の株名、Species predicted by kmer ID の列は基準株の種名、Similarity の列は類似度の値、ST の列は MLST 解析で明らかとなった ST 番号を表す。kmer ID を用いた解析で最も類似度の高い基準株の種を同定種とした。

kmer ID を用いた解析の結果、下水由来単離細菌 12 株および河川由来溶血性大腸菌 3 株はいずれも *Escherichia coli* と同定された (表 11)。kmer ID で最も高い類似度を示した基準株は、WWo055 と WWo102、WWo156 を除き、UM146 株であった。WWo102 と WWo156 は大腸菌 NA114 と最も類似度が高かった。単離後に溶血活性を持つことが確認できなかった WWo055 は大腸菌 W 株と最も類似度が高かった。UM146 株は、adherent invasive *Escherichia coli* (AIEC) に分類される大腸菌で、クローン病患者の回腸から単離された株である。ゲノム配列の解析から、ヒトの尿路感染性大腸菌 (UPEC) である CFT073 株と最も近縁であることが報告されている⁸⁴。NA114 株は前立腺炎患者の尿から単離された UPEC である⁸⁵。今回の解析では、基準株としては UPEC の他に、EAEC の O44:H18 042 株、腸管出血性大腸菌 (EHEC) の O111:H- 11128 株、O157:H7 TW14359 株、O26:H11 11368 株等を基準株として類似度を計算した (表 2)。下水から血液寒天培地を用いて単離された溶血性大腸菌の多くは、類似度の最も高い基準株が AIEC、もしくは UPEC であることから、溶血性大腸菌株の多くは AIEC や UPEC に近い仲間であることが示唆された。

MLST 解析の結果を表 11 に示す。WWo005、WWo014、WWo070、WWo073、WWo108、KOr014 はいずれも ST95 に分類された。WWo027、WWo156、KKo007 は ST127 に分類された。WWo003 と WWo147 は ST73 に分類された。WWo083 は ST117、KKa004 は ST961 にそれぞれ分類された。溶血活性を示さなかった WWo055 は ST5073 に分類された。WWo102 は、ST83 に最も近かったが、*fumC* 遺伝子配列が既知の配列とは異なっていた。

いずれの株も kmer ID での解析で誤認される可能性のある、ST270、ST1751、ST1767 には分類されなかった。以上の結果から、今回単離した WWo003、WWo005、WWo014、WWo027、WWo055、WWo070、WWo073、WWo083、WWo102、WWo108、WWo147、WWo156 の 12 株はいずれも大腸菌 (*Escherichia coli*) と同定された。

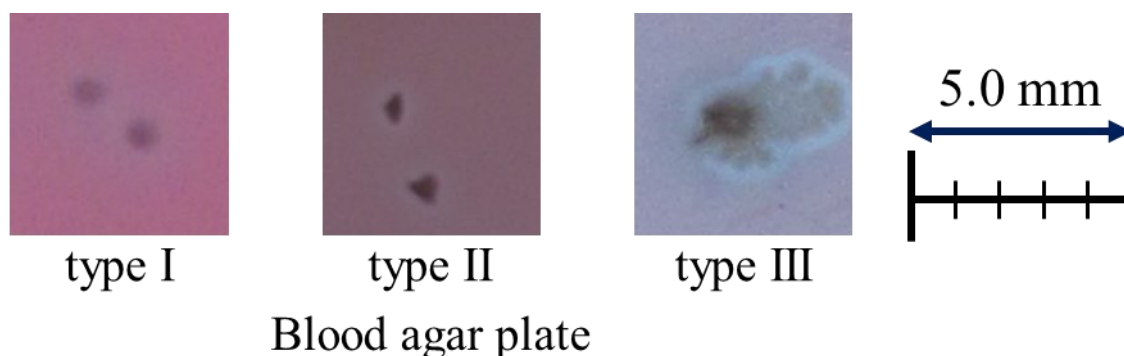


図 11 WWo108 の各血液寒天培地上における溶血斑の様子

表9 全ゲノムシーケンスの対象株

Strain	Species predicted by 16S rRNA genes	Hemolysis type	Hemolysis* ¹
WWo003	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>	Clear hemolysis	0.3
WWo005	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>	Clear hemolysis	0.5
WWo014	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>	Clear hemolysis	0.55
WWo027	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>	Clear hemolysis	1.4
WWo055	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>	No hemolysis	0
WWo070	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>	Clear hemolysis	0.65
WWo073	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>	Clear hemolysis	0.5
WWo083	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>	Clear hemolysis	0.05
WWo102	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>	Clear hemolysis	0.55
WWo108	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>	Indistinct hemolysis	0
WWo147	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>	Clear hemolysis	1.3
WWo156	<i>Brenneria alni</i> / <i>Escherichia coli</i> / <i>Escherichia fergusonii</i> / <i>Shigella flexneri</i> / <i>Shigella sonnei</i>	Clear hemolysis	0.8
KKo007	(<i>Escherichia coli</i>)	Clear hemolysis	1.1
KOr014	(<i>Escherichia coli</i>)	Clear hemolysis	0.3
KKa004	(<i>Escherichia coli</i>)	Clear hemolysis	1.35
WWo011	<i>Aeromonas dhakensis</i> / <i>Aeromonas taiwanensis</i>	Clear hemolysis	0.9
WWo148	<i>Aeromonas caviae</i> / <i>Aeromonas hydrophila</i> / <i>Aeromonas jandaei</i>	Clear hemolysis	0.35
WWo140	<i>Aeromonas encheleia</i> / <i>Aeromonas hydrophila</i> / <i>Aeromonas molluscorum</i>	Clear hemolysis	1.6
WWo190	<i>Streptococcus porcorum</i> / <i>Streptococcus sanguinis</i> / <i>Streptococcus sinensis</i>	Indistinct hemolysis	0
WWo197	<i>Streptococcus equinus</i> / <i>Streptococcus lutetiensis</i>	Indistinct hemolysis	0
WWo204	<i>Leuconostoc lactis</i>	Indistinct hemolysis	0
WWo207	<i>Lactococcus taiwanensis</i>	Indistinct hemolysis	0
WWo210	<i>Lactobacillus salivarius</i>	Indistinct hemolysis	0
WWo220	<i>Enterococcus casseliflavus</i>	Indistinct hemolysis	0
WWo221	<i>Enterococcus lemanii</i>	Indistinct hemolysis	0

*1; Error bars indicate standard error

表 10 単離株と大腸菌および赤痢菌の ANI 値

Genus	<i>Escherichia</i>							<i>Shigella</i>				
	<i>E. coli</i>	<i>E. coli</i>	<i>E. alba</i>	<i>E. albertii</i>	<i>E. fergusonii</i>	<i>E. marmotae</i>	<i>E. ruyisiae</i>	<i>S. boydii</i>	<i>S. dysenteriae</i>	<i>S. flexneri</i>	<i>S. sonnei</i>	
Species	K-12											
Strain	Sakai substr.	substr.										
	RIMD 0509952	MG1655	B35	EC06-170	RHB19-C05	NCTC11133	OPT1704	DMB SH136	HNCMB 20080	301	ECH+12	
strain	WWo003	96.5	96.8	80.8	90.2	91.0	91.2	92.7	96.6	96.1	96.3	96.8
	WWo005	96.5	96.7	80.6	90.2	91.1	91.1	92.7	96.6	96.0	96.2	96.7
	WWo014	96.6	96.7	80.7	90.1	91.0	91.1	92.7	96.6	96.0	96.2	96.6
	WWo027	96.6	96.7	80.8	90.0	91.2	91.1	92.7	96.6	96.2	96.3	96.8
	WWo055	97.7	98.4	80.9	90.1	91.2	91.2	92.9	98.3	97.1	97.8	99.0
	WWo070	96.5	96.7	80.7	90.2	91.0	91.1	92.7	96.7	96.0	96.2	96.7
	WWo073	96.6	96.7	80.6	90.2	91.0	91.1	92.7	96.6	96.0	96.1	96.7
	WWo083	97.0	97.1	80.9	90.2	91.2	91.3	92.8	96.9	96.3	96.6	97.1
	WWo102	96.6	96.8	80.8	90.3	91.0	91.2	92.7	96.7	96.1	96.4	96.7
	WWo108	96.5	96.7	80.6	90.2	91.1	91.1	92.7	96.6	96.0	96.2	96.6
	WWo147	96.5	96.7	80.8	90.1	91.0	91.2	92.8	96.6	95.9	96.2	96.7
	WWo156	96.5	96.7	80.8	90.3	90.9	91.1	92.7	96.6	96.0	96.2	96.8
	KKo007	96.5	96.7	80.7	90.1	91.0	91.1	92.7	96.6	96.0	96.3	96.7
	KOr014	96.4	96.8	80.7	90.2	91.0	91.1	92.7	96.6	95.9	96.2	96.6
	KKa004	96.5	96.8	80.7	90.2	90.9	91.3	92.7	96.6	96.0	96.3	96.7

表 11 kmer ID で同定された単離株の種と MLST 解析の結果

Strain	Reference strain with highest similarity	Species predicted by kmer ID	Similarity	ST
WWo003	UM146	<i>Escherichia coli</i>	81.22%	73
WWo005	UM146	<i>Escherichia coli</i>	96.59%	95
WWo014	UM146	<i>Escherichia coli</i>	99.11%	95
WWo027	UM146	<i>Escherichia coli</i>	83.36%	127
WWo055	W	<i>Escherichia coli</i>	82.18%	5073
WWo070	UM146	<i>Escherichia coli</i>	97.09%	95
WWo073	UM146	<i>Escherichia coli</i>	96.96%	95
WWo083	UM146	<i>Escherichia coli</i>	62.38%	117
WWo102	NA114	<i>Escherichia coli</i>	79.17%	83*
WWo108	UM146	<i>Escherichia coli</i>	97.98%	95
WWo147	UM146	<i>Escherichia coli</i>	81.77%	73
WWo156	NA114	<i>Escherichia coli</i>	81.51%	127
KKo007	UM146	<i>Escherichia coli</i>	83.96%	127
KOr014	UM146	<i>Escherichia coli</i>	99.23%	95
KKa004	UM146	<i>Escherichia coli</i>	81.01%	961

* alleles with less than 100% identity found

* *fumC* : Novel allele, ST may indicate nearest ST.

3.2.1.2. 溶血性大腸菌の系統と溶血毒 α ヘモリシンの保有

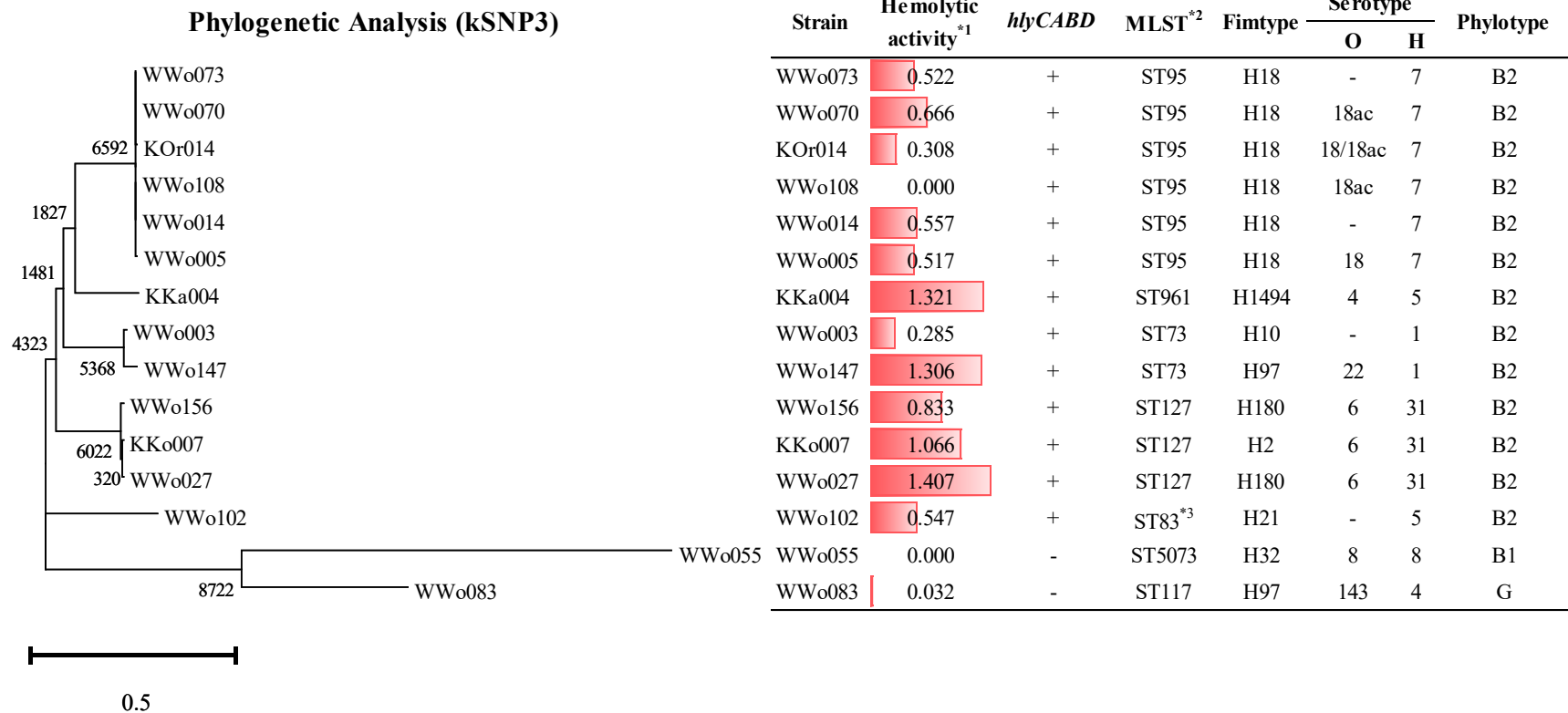
全ゲノムシーケンスを行った大腸菌株の kSNP3.0 を用いた系統解析⁶⁹、MLST 解析、phylogroup、fimH 分類、血清型 (serotype) および溶血毒 α ヘモリシンの保有について調べた。病原性を予測するため、大腸菌には様々な分類方法が存在する。kSNP3.0 を用いた系統解析や MLST 解析は、大腸菌のゲノム全体、もしくは特定の遺伝子上の SNPs から進化上の距離を算出することで、進化論的観点から大腸菌を分類することができる。kSNP3.0 は全ゲノム中の SNPs を検出して系統解析を行う。MLST 解析は複数の遺伝子上の SNPs に基づきシーケンスタイプに分類する。Phylogroup 分類は、特定の遺伝子の保有の有無に従い大腸菌を分類する。fimH 分類は、*fimH* 遺伝子の型に従い分類する。血清型 (serotype) 分類では、大腸菌の保有する抗原によって分類する。血清型分類では 186 種類の O-抗原と 53 種類の H-抗原の組み合わせで分類される。特定の溶血性大腸菌の血清型は志賀毒素産生大腸菌 (STEC) の O157:H7 や O103:H21 などの特定の病原型と関連していることも多く、病原性を推定するうえで重要なヒントになりうる。溶血毒 α ヘモリシンは大腸菌の保有する溶血毒で、 α ヘモリシンとその分泌等に関わる遺伝子は *hlyCABD* オペロンにコードされる。ここでは *hlyCABD* オペロンの保有について、大腸菌株のシーケンス解析で得られたコンテイングと *hlyC*、*hlyA*、*hlyB*、*hlyD* 遺伝子配列のそれぞれを BLAST 相同検索で保有を確かめた。

kSNP3.0 を用いた系統解析の結果を図 12 の左に表す。また、*hlyCABD* オペロンの保有、シーケンスタイプ (ST)、FimH 分類、血清型 (serotype)、phylogroup を溶血活性と併せて図 12 の右に示す。kSNP3.0 を用いた系統解析の結果は MLST や fimtype、serotype、phylogroup の分類や *hlyCABD* オペロンの保有との関係が見られた。WWo055 と WWo083 を除くすべての株が *hlyCABD* オペロンを保有しており、いずれも B2 phylogroup に分類された。ST95 に分類された WWo073、WWo070、KOr014、WWo108、WWo014、WWo005 は、kSNP3.0 の系統樹で近縁であった。ST95 の株はいずれも fimtype が H18、血清型の H

抗原は H7、O 抗原が分かったものは O18 もしくは O18ac であった。ST95 の株は系統解析の結果から近縁株であることが示されたが、溶血活性はいくつかの株で大きく異なっていた。いずれも *hlyCABD* オペロンを保有するが、溶血活性は WWo073、WWo070、WWo014、WWo005 が 0.5~0.7 mm であったのに対し、KOr014 株は約 0.3 mm とやや小さく、WWo108 は血液寒天プレートタイプ II では溶血活性を示さず、血液寒天プレートタイプ III でわずかに溶血活性を示した。KKa004 は kSNP3.0 の系統樹の中で ST95 と近いが、fimtype が H1494、血清型の H 抗原は H18、O 抗原が O4 であり、ST95 とは遺伝的に異なっていることが示唆される。また、ST95 の株に比べて溶血活性が約 2 倍以上大きいことも特徴的であった。

ST73 に分類された WWo003 と WWo147 は、系統解析の結果からも近縁種であったが、溶血活性は約 4.6 倍異なっていた。血清型の O 抗原型や fimtype も異なっていた。ST127 に分類された WWo156、KKo007、WWo027 は溶血活性が 0.8~1.4 mm と比較的溶血活性の大きい株が分類された。Fimtype は KKo007 株だけ H2 で残りの 2 株は H180 であった。血清型はいずれも O6:H31 であった。系統解析の結果からこれらの株は近縁であることが示唆される。WWo102 は、該当する MLST が存在せず、最も近いのが ST83 であった。H 抗原の血清型は H5 で KKa004 と同じであったが。系統解析の結果ではこの 2 株は遺伝的にはやや離れていることが分かる。

hlyCABD オペロンを保有する大腸菌はいずれも B2 phylotype に分類され、kSNP3.0 を用いた系統解析でも近いクラスターに分類された一方、*hlyCABD* オペロンを保有しない WWo055 と WWo083 は全ゲノム系統樹では異なるクラスターに分類され、phylotype もそれぞれ B1、G で、ほかの B2 phylotype の株とは異なっていた。WWo055 は溶血活性を示さなかった株で、ST5073、H32 fimtype、血清型 H8:O8 を示した。WWo083 は溶血活性を示した株であったが *hlyCABD* オペロンの保有は確認できなかった。このことは、WWo083 が α ヘモリシン以外の溶血毒を保有することを示唆する。WWo083 は、ST117、H97 fimtype、血清型 O147:H4 を示した。



Nodes are labeled with the number of SNPs that are present in all of the descendants of that node and nowhere else

*1: hemolytic activity is measured by hemolytic width on blood agar plate type II (24 hr)

*2: Alleles with less than 100 % identity found. fumC: Novel allele, ST may indicate nearest ST.

図 12 kSNP3.0 を用いた系統解析の結果と、溶血活性の大きさ、*hlyCABD* オペロンの保有、MLST 解析による分類、fimH 分類、血清型 (serotype)、phylotype の分類の結果

3.2.1.3. 大腸菌の各病原型に特徴的な病原性遺伝子の保有

病原性大腸菌は様々な病気を引き起こすが、引き起こす病気の種類によってそれぞれの病原性大腸菌は病原型で分類される。ヒトに感染する病原性大腸菌はその感染する臓器によって、腸管内病原性大腸菌 (intestinal pathogenic *Escherichia coli*; InPEC) と腸管外病原性大腸菌 (extraintestinal pathogenic *Escherichia coli*; ExPEC) の二つに大別される。InPECはさらに、志賀毒素産生大腸菌 (Shiga toxin-producing *Escherichia coli*; STEC)、腸管病原性大腸菌 (enteropathogenic *Escherichia coli*; EPEC)、腸管侵入性大腸菌 (enteroinvasive *Escherichia coli*; EIEC)、腸管凝集性大腸菌 (enteroaggregative *Escherichia coli*; EAEC)、腸管毒素原性大腸菌 (ETEC)、分散接着性大腸菌 (diffusely adherent *Escherichia coli*; DAEC) 等に分類される。ExPECは、尿路感染性大腸菌 (uropathogenic *Escherichia coli*; UPEC)、新生児髄膜炎起因大腸菌 (neonatal meningitis *Escherichia coli*; NMEC)、敗血症関連大腸菌 (sepsis-associated *Escherichia coli*; SEPEC) 等に分類される⁸⁶⁻⁸⁸。

それぞれの病原型には、それぞれを特徴づけるマーカーとなる病原性因子があり、いくつかの病原型は1つ以上の病原性マーカーを持つことで定義されている⁸⁸。STEC (EHEC)、ETEC、EIEC、EPECはそれぞれの病原型を決定する分子マーカーが証明されている。EAECやDAEC、AIEC、ExPECについては病原性におけるマーカーの役割が証明されていない⁸⁸ため、病原性因子による病原型の明確な定義はまだされていないが、いくつかの病原性因子が各病原型に特徴的なマーカーとして挙げられる。

STECはその名の通り志賀毒素を産生し、病原性の決定因子として志賀毒素1と志賀毒素2をそれぞれコードする *stx1*、*stx2* 遺伝子を保有する。STECの中で病原性の高いものはEHECと呼ばれる。

ETECは、1つ以上の腸内毒素を産生する細菌の能力によって定義される⁸⁸。腸内毒素には熱感受性腸管毒素 (heat-labile enterotoxin; LT) と熱安定性腸管毒素 (heat-stable enterotoxin; ST) に分類され、それぞれのサブタイプが存在する⁸⁹。ETECで最も主要な熱安定性毒素はSTa (もしくはSTIとして知られる) とSTb (もしくはSTIIとして知られる) で、ヒトに対しては特にSTaが重要である^{89,90}。

EIECは *Shigella* に近い仲間、いずれも pINV と呼ばれるプラスミドを持つ。pINVには、タイプIII分泌システムと複数のエフェクタータンパクがコードされる⁹¹。ipaH 遺伝子はEIECもしくは *Shigella* であることを判断する際にしばしばターゲットとされる遺伝子で、pINV上及び染色体上に複数存在する⁸⁸。保存領域と可変領域で構成されており、制限酵素 HindIII で処理した際の断片サイズによって、ipaH1.4、ipaH2.5、ipaH4.5、ipaH7.8、ipaH9.8のように命名されている^{92,93}。

EPECは、腸球排出遺伝子座 (LEE) と呼ばれる病原性遺伝子アイランドを持つことが特徴である。この40kbpのアイランドには、*eae* 遺伝子にコードされるインチミンと呼ばれる外膜接着タンパク、タイプIII分泌システム、インチミンの転移受容体である Tir タンパクを含む複数のタイプIII分泌エフェクタータンパクがコードされる⁸⁸。

EAECには、その病原性を決定する因子は明らかにはなっていない。EAECの表現型の特徴として凝集性があげられるが、この特徴はEAEC病原性プラスミド (pAA) や類似のプラスミドにコードされるAAF/I、AAF/II、AAF/III、AAF/IV と呼ばれる疎水性凝集性フィブリナのいずれかによって発現していると考えられている^{88,94}。また、pAAにコードされた細胞毒素 (Pet)、pAAにコードされた熱安定性腸内毒素である enteroaggregative

stable toxin (EAST-1) (ただし EAEC 以外にも保有することがある)、推定腸内毒素 ShET1 (ただし *Shigella flexneri* も保有する) 等が推定病原因子としてあげられるが、病原性関連決定因子としては明らかとなっていない^{88,94,95}。

ExPEC も、その病原性を決定する因子は明らかになっていない。Johnson ら⁹⁶は、大腸菌株に 5 つの病原性マーカー (*papA/papC*, *sfa/foc*, *afa/dra*, *iutA*, *kpsMTII*) を PCR で調べ、これらのうち 2 つが存在する場合を ExPEC とみなすことで、ExPEC を判断する分子学的アプローチを提案している。しばしばこの定義に基づき ExPEC の判別が行われているが、ExPEC としての病原性を決定する因子についてはまだ明らかではない。ExPEC には、ヒトに感染する UPEC、NMEC、SEPEC 以外に、鳥類に感染する鶏病原性大腸菌 (Avian pathogenic *Escherichia coli*; APEC) がある。これらの病原性大腸菌の保有する病原性因子は共通するものが多数存在する^{97,98}ため、ExPEC の中で病原性を区別することは難しい。

すべての病原性を決定するゴールドスタンダードは存在しないが、前述のとおり各病原型のマーカー遺伝子がいくつか存在する。Kuhnert ら (2000)⁸⁷は、ETEC、EPEC、STEC/EHEC、EIEC、EAEC、UPEC、NMEC に特徴的な遺伝子を一つに取りまとめた。本研究では、Kuhnert らがとりまとめた各病原型に特徴的な遺伝子の保有を確認した。

表 12 に、Kuhnert らが取りまとめた各病原型に特徴的な各遺伝子と、各遺伝子に対応する UniProt ID を示す。本研究では、これらの遺伝子の保有について調べた。*ipaH* は種類の異なる *ipaH3*, *ipaH4.5*, *ipaH7.8*, *ipaH9.8* の複数について調べた。*stx1*, *stx2*, *eltI*, *eltIIIa* については転写産物がサブユニットによって構成されることからサブユニットを構成する複数の遺伝子をターゲットとした。いくつかの遺伝子については原論文に記載されている名前と UniProt データベース上に記載されている名前が異なっていることがあり、それぞれの遺伝子名についても併せて載せている。Kuhnert らは UPEC と NMEC のそれぞれに特徴的な遺伝子を記述したが、UPEC に特徴的な *papA*, *pilC*, *hlyA*, *cnfI* 遺伝子は NMEC も保有することがあり⁹⁹、そのほか複数の病原性因子も共有していることが多い⁹⁸。そのため本研究では、UPEC や NMEC をまとめて ExPEC として、各病原型に特徴的な病原性遺伝子の保有について調べた。遺伝子保有は、各遺伝子に対応する塩基配列にショットガンシーケンスで得られたリードをアラインメントすることで確かめた。

アラインメントした結果を図 13 に示す。図では、大腸菌の全ゲノムシーケンスによって得られたデータをもとに、各病原型に特徴的な遺伝子配列に対してアラインメントしたときの *gene coverage* を示しており、1 に近いほどアラインメントされている領域の割合が高い。各株は図 12 の kSNP3.0 を用いた系統解析で得られたクラスターの順に並べて表示している。

今回調べた大腸菌は、InPEC に特徴的な病原性因子をほとんど保有しなかった。ETEC の決定因子である、*eltB*, *stx1*, *stx2*, *HL-IIa*, *HL-IIAb*, EPEC の決定因子である *eae*, STEC/EHEC の決定因子である *stxA*, *stxB*, *stxA2*, *stxB2*, EIEC の決定因子である *ipaH*, EAEC の推定病原因子である *aggA* の *gene coverage* はいずれの株も 0 に近い値を示し、これらの遺伝子を保有していないことが示唆された。InPEC の推定病原因子として挙げられる、*EAST1*, *hlyA*, *iucC* の *gene coverage* はいくつかの株で高い値を示した。溶血活性を持たなかった WWo055 と不明瞭な溶血斑を示した WWo083 以外の株で、STEC/EHEC に特徴的な *hlyA* 遺伝子の *gene coverage* は 1 に近い値を示したが、*hlyA* は ExPEC にも特徴的な遺伝子でもあった。ETEC や EPEC、STEC/EHEC に特徴的な *EAST1* 遺伝子について、WWo003, WWo055, WWo083, WWo102, WWo147 の *gene coverage* は 0.15-0.63 と他の株

と比べてやや高い値となった。*EASTI* 遺伝子には部分的にアラインメントされていたため、リファレンスで使用した *EASTI* 遺伝子配列と部分的に類似している領域が存在していた可能性が考えられる。EIEC に特徴的な遺伝子の一つである *iucC* は ExPEC にも特徴的な遺伝子である。

一方、今回調べた溶血性大腸菌の多くは ExPEC に特徴的な病原性因子を複数保有していた。*papA*、*sfaS*、*pilC*、*neuA*、*neuC*、*hlyA*、*cnfI* は複数の溶血性大腸菌株で高い gene coverage を示し、*iucC* と *KfiB* は WWo003 でのみ高い gene coverage を示した。

hlyA 遺伝子と *cnfI* 遺伝子は、溶血活性を持たなかった WWo055 と溶血活性の小さな WWo083 を除くすべての株で gene coverage が高く、この遺伝子を保有していると考えられる。また、これらの株は *papA* の gene coverage も高い値を示した。*papA* は P フィンブリアの構成タンパク質の一つをコードする遺伝子である。P フィンブリアは UPEC の病態に関与しており、*in vivo* で尿路上皮細胞への付着に重要な働きをする¹⁰⁰。また、尿路敗血症患者から単離された UPEC においては高い割合（約 77%）で P フィンブリアを保有する¹⁰¹。

sfaS 遺伝子と *pilC* 遺伝子（別名 *foc*）はいずれも、WWo073、WWo070、KOr014、WWo014、WWo005、KKa004、WWo003、WWo147、KKo007、WWo102 で gene coverage が高かった。*sfaS* 遺伝子と *pilC* 遺伝子は、それぞれ S フィンブリアと 1 型フィンブリアの構成タンパク質の一つをコードする遺伝子である。S フィンブリアは、UPEC の 30-60% が保有している¹⁰²⁻¹⁰⁴。1 型フィンブリアは、*fim* オペロンにコードされており、UPEC のほとんどの株が保有する。尿路上皮細胞に定着した際に発現し、UPEC の感染力を高める¹⁰⁵⁻¹⁰⁷。なお、これらのフィンブリアを保有しなくても尿路上皮細胞へ浸潤することも報告されている¹⁰²。

neuA と *neuC* は WWo073、WWo070、KOr014、WWo108、WWo014、WWo005 において、gene coverage が高かった。これらの株は MLST 解析の結果 ST95、ST961、もしくは ST83 に近いと分類された株で、いずれも系統的に近いグループであった。*neuA* と *neuC* は K1 カプセルタンパク質の合成に関わる。大腸菌のカプセルは 80 以上が確認されており、宿主の免疫細胞による貪食や血清に含まれる補体から菌体を保護する^{108,109}。UPEC や NMEC では K1 や K5 カプセルを保有することが多い⁸⁷。K1 カプセルについては、血清耐性に関わっている可能性が示唆されているが、株によって耐性への寄与は異なっていることや¹¹⁰、血清耐性以外に大腸菌に感染するある種のフェージに対する防御としての機能を持つ等の報告もあり¹¹¹、血清耐性に対する役割についてはまだ明確にはなっていない。

iucC と *KfiB* は WWo003 のみにおいて gene coverage が高かった。*iucC* は、シデロフォアであるエアロバクチンの生合成に関わる遺伝子である。エアロバクチンは、鉄の少ない環境下において産生され、鉄のキレート剤として作用することで、大腸菌の鉄獲得に寄与する¹¹²。病原性細菌はコロニー形成や感染時の増殖に鉄を必要とするため、宿主によって管理されている鉄を獲得しなければならない。しかしながら、細菌が利用できる第二鉄 (Fe^{3+}) は宿主であるヒトの体内における pH では不溶性である上、遊離した鉄のほとんどが宿主のタンパク質と結合して厳重に管理されており、宿主体内で容易に鉄を獲得することは難しい^{113,114}。そこで、細菌はエアロバクチンをはじめとするシデロフォアと呼ばれる鉄のキレート剤を産生することで鉄を封じ込め、鉄を取り込む¹¹⁴⁻¹¹⁶。シデロフォアによる鉄の獲得によって、感染時の増殖および生存を促進する^{117,118}。エアロバクチンは *iucC* 遺伝子を含む *iucABCD* 遺伝子と *iutA* 遺伝子で構成されるオペロンによって産生される¹¹⁹。女性の尿路感染症患者に感染した大腸菌でエアロバクチンインポーター遺伝子

(*iutA*) の転写 RNA が検出された報告がある¹²⁰。また、尿路感染症を再発させる UPEC 株では、エアロバクチンの保有率が高いことも報告されている^{121,122}。*KfiB* は、K5 カプセルタンパクの合成に関わる。K5 カプセルは、宿主の免疫細胞である多形核白血球や単球による貪食作用を抑制する働きがある¹²³。

表 12 大腸菌の各病原型に特徴的な遺伝子⁸⁷と UniProt データベース上で対応するタンパク質

Type	Gene (Kuhnert <i>et al.</i> 2000)	Gene (UniProt)	UniProt ID	Protein	InPEC					ExPEC			
					ETEC	EPEC	STEC/EHEC	EIEC	EAEC	UPEC	NMEC		
Adhesins	<i>papA</i>	<i>papA</i>	P04127	Pap fimbrial major pilin protein							++		
	<i>sfaS</i>	<i>sfaS</i>	P13430	S-fimbrial adhesin protein SfaS							++	+	
	<i>foc</i>	<i>pilC</i>	P62605	Type-1 fimbrial protein, C chain							++		
	<i>bfpA</i>	<i>bfpA</i>	P58997	Major structural subunit of bundle-forming pilus								++	
	<i>cfa/I</i>	<i>cfaB</i>	P0CK93	CFA/I fimbrial subunit B	+								
	<i>cfa/III</i>	N/A	P15488	CS3 fimbrial subunit A	++								
	<i>aaf/I</i>	<i>aggA</i>	P46007	AAF/I fimbrial subunit								++	
	<i>eae</i>	<i>eae</i>	O31000	Intimin									
invasins	<i>ipaH</i>	<i>ipaH3</i>	Q83RJ4	E3 ubiquitin-protein ligase ipaH3									
		<i>ipaH4.5</i>	P18009	Probable E3 ubiquitin-protein ligase ipaH4.5									
		<i>ipaH7.8</i>	P18014	Probable E3 ubiquitin-protein ligase ipaH7.8									
		<i>ipaH9.8</i>	Q8VSC3	E3 ubiquitin-protein ligase ipaH9.8									
iron acquisition	<i>iucC</i>	<i>iucC</i>	Q47318	Aerobactin synthase				+			++	+	
Capsule	<i>neuA+neuC</i>	<i>neuA</i>	P13266	N-acylneuraminate cytidyltransferase							+	++	
		<i>neuC</i>	Q47400	Polysialic acid biosynthesis protein P7								+	
toxin	<i>KfiB</i>	<i>KfiB</i>	Q6K CZ4	KfiB protein								+	
	<i>stla/stIb</i>	<i>stI</i>	P01559	Heat-stable enterotoxin ST-IA/ST-P	++								
	<i>stI</i>	<i>stI</i>	P22542	Heat-stable enterotoxin II	++*a								
	<i>stx1</i>	<i>stxA</i>	P08026	Shiga-like toxin 1 subunit A									++
		<i>stxB</i>	P69178	Shiga-like toxin 1 subunit B									
	<i>stx2</i>	<i>stxA2</i>	P09385	Shiga-like toxin 2 subunit A									++
		<i>stxB2</i>	P09386	Shiga-like toxin 2 subunit B									
	<i>eltI</i>	<i>eltA</i>	P06717	Heat-labile enterotoxin A chain	++								
		<i>eltB</i>	P32890	Heat-labile enterotoxin B chain									
	<i>eltIIa</i>	N/A	P13810	Heat-labile enterotoxin IIA, A chain	++*a								
		N/A	P13812	Heat-labile enterotoxin IIA, B chain									
	<i>hlyA</i>	<i>hlyA</i>	P09983	Hemolysin, chromosomal									++*b
<i>Ehx</i>	<i>EHEC-hlyA</i>	Q9LC58	Enterohemolysin									++*c	
<i>cnfI</i>	<i>cnfI</i>	Q47106	Cytotoxic necrotizing factor 1									+	
<i>EASTI</i>	<i>EASTI</i>	Q53559	EAST1 toxin	+	+							++	

+ occasionally detected, ++ normally detected

N/A: There is no gene named for correspond protein

a: Primarily in animal isolates

b: In porcine isolates

c: Predominantly strains of the EHEC sub-pathotype

strain	<i>eltA</i>	<i>eltB</i>	<i>stx1</i>	<i>stx2</i>	<i>HL-IIAa</i>	<i>HL-IIAb</i>	<i>cfaB</i>	<i>C3Stx</i>	<i>EAST1</i>	<i>eae</i>	<i>bfpA</i>	<i>EAST1</i>	<i>stxA</i>	<i>stxB</i>	<i>stxA2</i>	<i>stxB2</i>	<i>eae</i>	<i>hlyA</i>	<i>EHEC-hlyA</i>	<i>EAST1</i>	<i>ipaH3</i>	<i>ipaH4.5</i>	<i>ipaH7.8</i>	<i>ipaH9.8</i>	<i>iucC</i>	<i>aggA</i>	<i>papA</i>	<i>sfaS</i>	<i>pilC</i>	<i>iucC</i>	<i>neuA</i>	<i>neuC</i>	<i>KfiB</i>	<i>hlyA</i>	<i>cnf1</i>
WWo073	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	1.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.98	1.00	0.33	0.00	0.55	0.83	0.00	1.00	0.83
WWo070	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	1.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	1.00	1.00	0.33	0.00	0.80	0.90	0.00	1.00	0.94
KOr014	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	1.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	1.00	1.00	0.33	0.00	0.97	1.00	0.00	1.00	1.00
WWo108	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.99	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	1.00	0.00	0.00	0.00	0.94	0.93	0.00	0.99	0.99
WWo014	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	1.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	1.00	1.00	0.37	0.00	1.00	0.99	0.00	1.00	0.99
WWo005	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	1.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	1.00	1.00	0.31	0.00	0.89	0.80	0.00	1.00	0.98
KKa004	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.21	1.00	0.00	0.00	0.00	0.00	1.00	0.98
WWo003	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.34	0.01	0.00	0.34	0.00	0.00	0.00	0.00	0.01	1.00	0.00	0.34	0.00	0.00	0.00	0.01	1.00	0.00	0.96	0.26	1.00	1.00	0.00	0.00	0.74	1.00	0.85
WWo147	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.37	0.01	0.00	0.37	0.00	0.00	0.00	0.00	0.01	0.99	0.00	0.37	0.00	0.00	0.00	0.01	0.00	0.00	0.22	0.46	0.36	0.00	0.00	0.00	0.00	0.99	0.91
WWo156	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.01	0.00	0.15	0.00	0.00	0.00	0.00	0.01	0.99	0.00	0.15	0.00	0.00	0.00	0.01	0.00	0.00	0.64	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.93
KKo007	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	1.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.72	1.00	0.35	0.00	0.00	0.00	0.00	1.00	0.99
WWo027	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	1.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.64	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.84
WWo102	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.34	0.01	0.00	0.34	0.00	0.00	0.00	0.00	0.01	1.00	0.00	0.34	0.00	0.00	0.00	0.01	0.00	0.00	0.20	0.23	1.00	0.00	0.00	0.00	0.00	1.00	0.95
WWo055	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.63	0.01	0.00	0.63	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.63	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
WWo083	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.19	0.01	0.00	0.19	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01

*1. Gene coverage are calculated from the formula; {Mapped region [bp]} / {Gene length [bp]}

図 13 大腸菌の各病原型に特徴的な遺伝子の塩基配列にアラインメントしたときの gene coverage

3.2.1.4. その他の病原性遺伝子の保有

溶血性大腸菌の保有する病原性遺伝子を VirulenceFinder2.0

(<https://cge.cbs.dtu.dk/services/VirulenceFinder/>)^{43,66}を用いて調べた。結果を図 14 に示す。また、各病原性遺伝子の転写産物を表 13 に示す。VirulenceFinder2.0 では、コンティグ配列をクエリー、病原性遺伝子配列をテンプレートとして、それぞれの配列を BLAST で比較する。図 14 の下側のヒートマップで、VirulenceFinder2.0 を用いた解析の中で BLAST プログラムを用いて得られた一致率 (Identity) を表す。大腸菌株は、3.2.1.2 の kSNP3.0 を用いた系統解析でクラスタリングされた順に並べている。また、大腸菌株間で保有パターンがよく似た病原性遺伝子を明らかにするため、病原性遺伝子が大腸菌の保有パターンでクラスタリングを行い、病原性遺伝子を並び替えた。クラスタリングの結果は図の上方に併せて示している。

コンティグに対して BLAST 検索を行っている点に注意する必要がある。つまり、コンティグ上に乗らなかった遺伝子は検出できていない。そのため、ヒットしていない場合でも実際には遺伝子を保有する可能性がある。また、コンティグの端に遺伝子が存在する場合は、部分的な配列比較となってしまうため実際には部分的でしか配列が一致していない場合でも一致率は高い値を示してしまう。テンプレート長と、それにヒットしたクエリー長については表 S 1、表 S 2、表 S 3、表 S 4 に示す。

溶血活性に関わると考えられる病原性遺伝子では、*hlyF* と *hlyE* がヒットした。 α ヘモリシンをコードする *hlyA* 遺伝子は VirulenceFinder2.0 のデータベース上にはなかったため、ここではヒットしていない。*hlyF* 遺伝子は WWo055 が保有していた。 α ヘモリシン産生に関わる *hlyA* 遺伝子を保有する株 (WWo055、WWo083 以外の株) は、*papC*、*kpsE*、*yfcV*、*clbB*、*cnf1*、*stiA*、*iss*、*terC*、*hra*、*fyuA*、*chuA*、*irp2*、*ompT* でヒットした。また、*papC* や *vat* も *hlyA* を保有する株のほとんどが保有していた。ST95 に分類された WWo073、WWo070、KOr014、WWo108、WWo014、WWo005 は、病原性遺伝子で共通するものが多く存在した。*neuC* は ST95 の株でのみヒットしたがほかの株でこれらの遺伝子にヒットした株はなかった。この結果は、3.2.1.3 のアラインメントに基づく遺伝子保有を確認した結果とも一致する。

kpsM 遺伝子は VirulenceFinder2.0 で使用するデータベースの中でいくつかの種類が記載されている。*kpsM* 遺伝子は 大腸菌カプセルタンパクをコードしており、血清型分類における K 抗原に対応している。ST95 に分類された WWo073、WWo070、KOr014、WWo108、WWo014、WWo005 は、*kpsMIII_K1* がヒットした。このことからこれらの株は K1 表現型であると推定される。一方、ST961 の KKa004、ST73 の WWo003、WWo147 は *KpsMIII_K5* でヒットした。このことから、これら 3 株は K5 表現型であると推定される。ST127 の WWo156、KKo007、WWo027 と ST83 に近い WWo102 は *kpsMIII* にヒットしたが遺伝型の推定まではできなかった。

KKa004 株は kSNP3.0 を用いた系統解析で ST95 の株と近縁であったが、ST95 の株 (WWo073、WWo070、KOr014、WWo108、WWo014、WWo005) と病原性遺伝子のヒットしたパターンが異なっていた。KKa004 株は *mcmA*、*mchF*、*mchC*、*cea*、*mchB* でヒットしたが、ST95 の株はこれらの遺伝子でヒットしなかった。一方、ST73 の WWo003 と ST83 に最も近い WWo102 はこれらの遺伝子にヒットした。

表 13 病原性遺伝子とその転写産物

Gene	Product
<i>mcmA</i>	Microcin M part of colicin H
<i>mchF</i>	ABC transporter protein MchF
<i>mchC</i>	MchC protein
<i>cea</i>	Colicin E1
<i>mchB</i>	Microcin H47 part of colicin H
<i>sat</i>	Secreted autotransporter toxin
<i>iutA</i>	Ferric aerobactin receptor
<i>iha</i>	Adherence protein
<i>iucC</i>	Aerobactin synthetase
<i>papA_F12</i>	Major pilin subunit F12
<i>pic</i>	serine protease autotransporters of Enterobacteriaceae (SPATE)
<i>focG</i>	F1C adhesion
<i>kpsMIII_K5</i>	Polysialic acid transport protein; Group 2 capsule
<i>astA</i>	EAST-1 heat-stable toxin
<i>etsC</i>	Putative type I secretion outer membrane protein
<i>hlyF</i>	Hemolysin F
<i>hlyE</i>	Avian E.coli haemolysin
<i>ireA</i>	Siderophore receptor
<i>lpfA</i>	Long polar fimbriae
<i>mcbA</i>	Bacteriocin microcin B17
<i>sfaE</i>	S fimbrial/F1C minor subunit
<i>tcpC</i>	Tir domain-containing protein
<i>kpsMII</i>	Polysialic acid transport protein; Group 2 capsule
<i>papA_F48</i>	Major pilin subunit F48
<i>gad</i>	Glutamate decarboxylase
<i>senB</i>	Plasmid-encoded enterotoxin
<i>traT</i>	Outer membrane protein complement resistance
<i>papC</i>	Outer membrane usher P fimbriae
<i>usp</i>	Uropathogenic specific protein
<i>kpsE</i>	Capsule polysaccharide export inner-membrane protein
<i>yfcV</i>	Fimbrial protein
<i>clbB</i>	Hybrid non-ribosomal peptide / polyketide megasynthase
<i>cnf1</i>	Cytotoxic necrotizing factor
<i>sitA</i>	Iron transport protein
<i>iss</i>	Increased serum survival
<i>terC</i>	Tellurium ion resistance protein
<i>vat</i>	Vacuolating autotransporter toxin
<i>hra</i>	Heat-resistant agglutinin
<i>fyuA</i>	Siderophore receptor
<i>chuA</i>	Outer membrane hemin receptor
<i>irp2</i>	High molecular weight protein 2 non-ribosomal peptide synthetase
<i>ompT</i>	Outer membrane protease (protein protease 7)
<i>focCsfaE</i>	S fimbrial/F1C minor subunit
<i>iroN</i>	Enterobactin siderophore receptor protein
<i>sfaD</i>	S fimbrial/F1C minor subunit
<i>sfaS</i>	S-fimbriae minor subunit
<i>papA_F43</i>	Major pilin subunit F43
<i>ibeA</i>	Invasin of brain endothelial cells
<i>kpsMIII_K1</i>	Polysialic acid transport protein; Group 2 capsule
<i>neuC</i>	Polysialic acid capsule biosynthesis protein

3.2.2. 溶血性大腸菌の溶血活性の決定因子

単離した溶血性大腸菌の多くは、*hlyCABD* オペロンを保有していた。*hlyCABD* オペロンを保有する株の中でも溶血活性が異なり、ST95 のように同じ ST 番号であるのにもかかわらず、溶血活性が大きく異なっているものも見つかった。*hlyCABD* オペロン内の *hlyA* 遺伝子は、溶血毒として知られる α ヘモリシンをコードする。溶血活性が異なる理由として、 α ヘモリシンをコードする *hlyA* 遺伝子から転写される mRNA 量の違いや、分泌される α ヘモリシタンパク量の違い、 α ヘモリシンの遺伝子配列の違い、そのほかの溶血毒との相互作用の有無等が考えられる。そこで、*hlyA* の mRNA 発現量とタンパク分泌量の測定、 α ヘモリシンの産生に関わる遺伝子の保有とその一塩基多型 (SNPs)、そのほかの溶血毒 (α ヘモリシン E、 α ヘモリシン F) 遺伝子の保有とその SNPs をそれぞれ確認し、溶血活性との関係について調べた。

3.2.2.1. 溶血毒の発現量と溶血活性

単離大腸菌を LB 培地で 2 時間培養して、上清の溶血活性、*hlyA* mRNA 発現量、上清に含まれる α ヘモリシタンパク量を測定した。上清の溶血活性は、上清に血液を加え、遊離したヘモグロビン量を吸光光度計で測定して求めた。*hlyA* mRNA 量は RT-qPCR によって測定した。 α ヘモリシタンパク分泌量は、濃縮した培養液上清をポリアクリルアミド電気泳動 (SDS-PAGE) によって分離し、 α ヘモリシタンパク量を目視で確認したが、該当するバンドを確認することはできなかった。

大腸菌を 2 時間培養した上清の溶血活性の大きさ (波長 450 nm で測定) を図 15 に示す。縦軸に溶血活性値として percent hemolysis を表す。液体培地で 2 時間培養したときの溶血活性は血液寒天プレートで 24 時間培養したときとは異なり、WWo027、WWo147、WWo156、KKa004、KKo007 の溶血活性が大きく、そのほかの株の溶血活性は小さかった。

RT-qPCR で測定した結果得られた *hlyA* mRNA 量と溶血活性の大きさの関係を図 16 に示す。溶血活性の大きかった株の内、WWo027、KKo007 の *hlyA* mRNA 発現量は、上清の溶血活性の小さかった WWo003 と比べて約 5 倍大きく、ほかの株と比べても発現量が大きかった。一方、同じように溶血活性の大きかった KKa004、WWo147、WWo156 の相対発現量は 1.8~2.5 で、溶血活性の小さかった WWo070、WWo073、WWo014 の相対発現量 1.8~2.9 と大きな差は見られなかった。WWo055、WWo083 は *hlyA* mRNA の発現量が 0 に近く、*hlyCABD* オペロンを保有しないという 3.2.1.2 の結果と一致した。

上清をポリアクリルアミド電気泳動した結果を図 17 に示す。操作上の問題により、KKo007、KOr014、KKa004 株については測定することができなかった。 α ヘモリシタンパクの分子量は 109.8kDa で、該当する分子量に対応するバンドを探したが、確認することはできなかった。

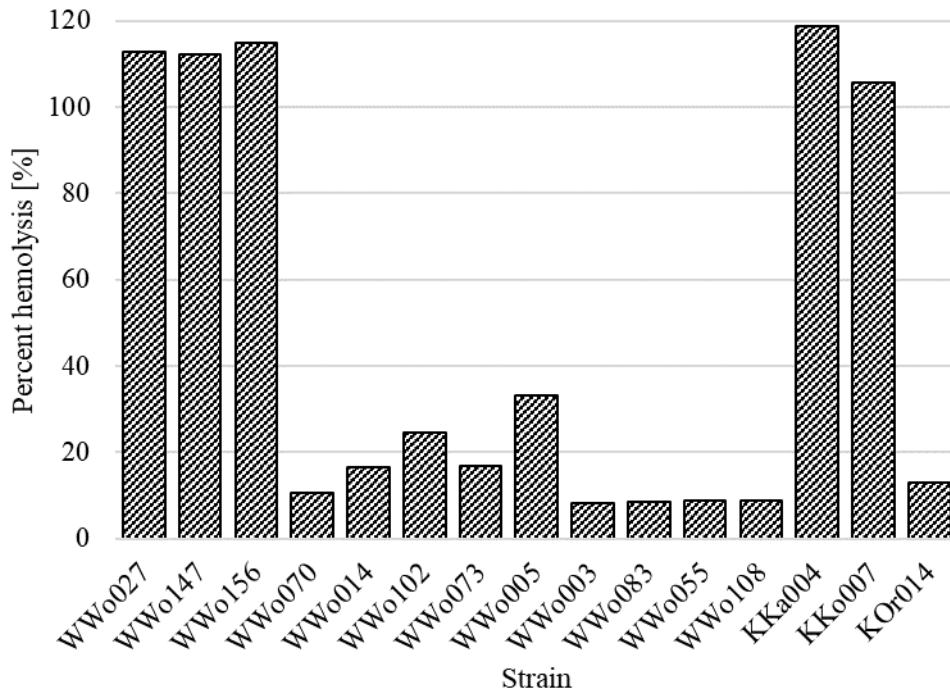


図 15 液体培地で 2 時間培養した上清の溶血活性

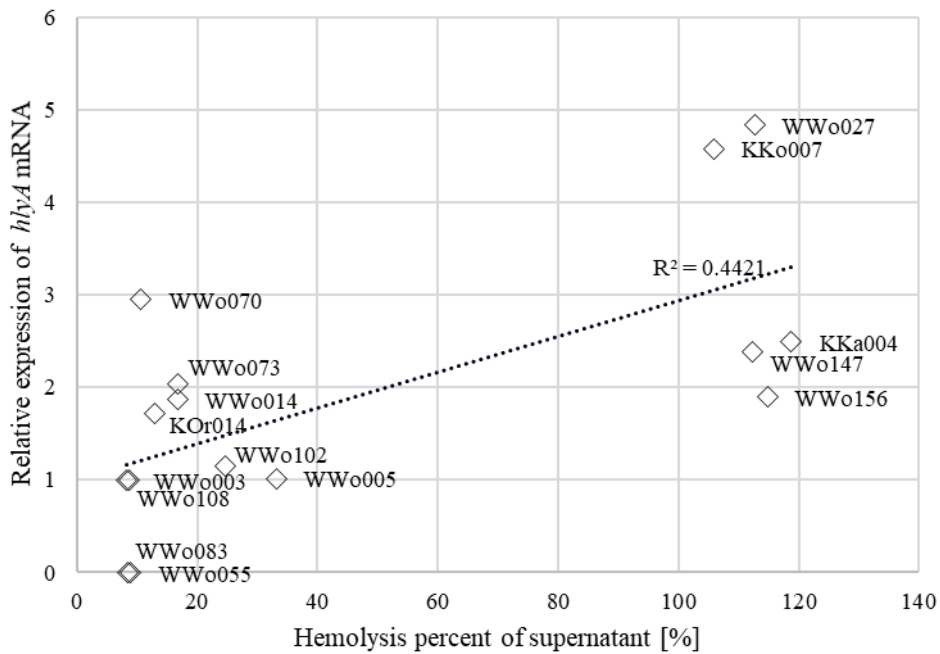


図 16 上清の溶血活性の大きさと *hlyA* mRNA の相対発現量の関係

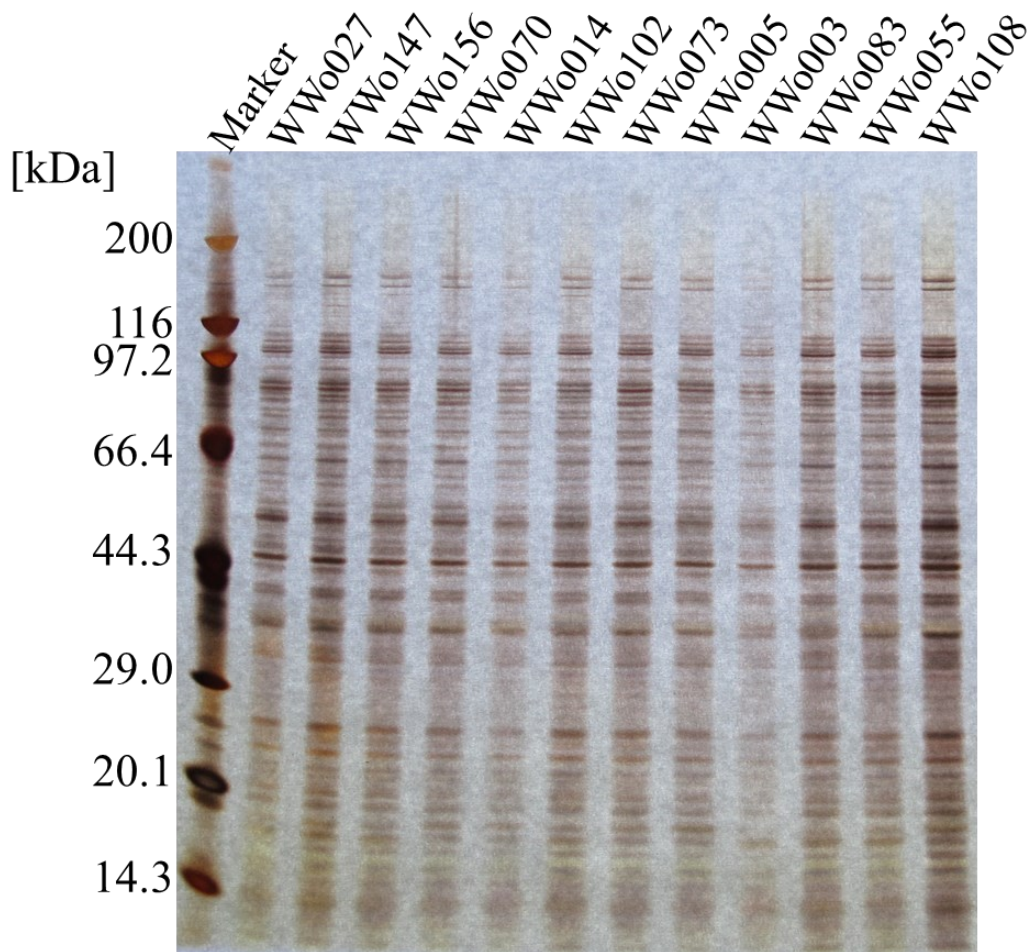


図 17 大腸菌培養液上清をポリアクリルアミド電気泳動 (SDS-PAGE) した結果

3.2.2.2. MMViewer を用いた溶血毒遺伝子保有の確認と変異解析

MMViewer の開発

溶血活性に関わる遺伝子の変異と溶血活性の関係を明らかにするため、溶血活性に関わる遺伝子の SNP 解析を行った。アラインメントに基づき変異を検出するプログラムとしていくつかのパイプラインが存在するが、多くのプログラムは十分な **depth** が確保できない領域については SNP を見落とす可能性があり、サンプル間の比較ができないことがある。また、変異データをグラフィカルなデータとして確認するツールの多くは、アミノ酸置換を伴わないナンセンス変異とアミノ酸置換を伴うミスセンス変異を一目で区別できない。そこで、十分な **depth** が確保できていない領域の変異も確認でき、サンプル間の変異データの比較が行え、ミスセンス変異を検出することができ、シーケンスで得られたデータに補正をかけることなく変異データをグラフィカルに確認できる変異解析ソフト Missense Mutation Viewer (MMViewer) を開発した。

MMViewer は、対象とする CDS 領域内のミスセンス変異や、その上流、下流の変異をグラフで表示するソフトウェアで、linux OS 上で動作する。MMViewer の解析に必要なデー

タは、「クオリティコントロール済みのショートリードデータ (fastq フォーマット)」、「リファレンスとして使用するコンプリートゲノムもしくはコンティグ (fasta フォーマット)」、「変異の有無を確認したい遺伝子の塩基配列もしくはアミノ酸配列 (fasta フォーマット)」の3つである。MMViewerは3つの機能「対象 CDS 領域の探索」、「アラインメント」、「グラフの描画」から構成される。

MMViewerは、図18のフローチャートに従い変異解析を行う。対象のCDS領域の探索では、prodigal¹²⁴を使用してリファレンスとして使用するコンプリートゲノムもしくはコンティグ上のCDS領域を予測する。得られたCDSの配列と、変異を確認したい遺伝子の塩基配列もしくはアミノ酸配列に対して、BLASTを用いた相同性検索を行い、リファレンス配列上の解析対象となるCDSを記録する。続いて、bwa mem⁶⁸を用いて、リファレンス配列に次世代シーケンサの解析で得られたクオリティコントロール済みのショートリードをアラインメントする。アラインメントされたデータは、fixmate情報の追加、ソート、PCRによる重複リードのタグ付けプロセスを経て、bam形式のファイルとして出力される。bamファイルはindexが追加される。出力されたbamファイルを元にbcftools mpileup¹²⁵を用いて変異データがリストアップされる。リストアップされた変異データはsnpEff¹²⁶を用いてアノテーションされたのち、一塩基置換 (SNP: snp) (例: A => T)、複数塩基置換 (MNP: mnp) (例: GC => AT)、挿入 (Insertion: ins) (例: ATT => AGTT)、欠失 (Deletion: del) (例: ACGG => ACG)、SNPやMNPの組み合わせ (complex) (例: ATTC => GTTA)の5種類に分類される。また、bamファイルからsamtools depth¹²⁵を用いてdepthを計算する。最後に、以上の解析から得られたデータをもとにグラフを作成し、svg形式で出力する。なお、MMViewerの使用例やソースコード等については付録3に記載する。MMViewerは<https://github.com/KIkebata/mmviewer>からダウンロードして使用することができる。

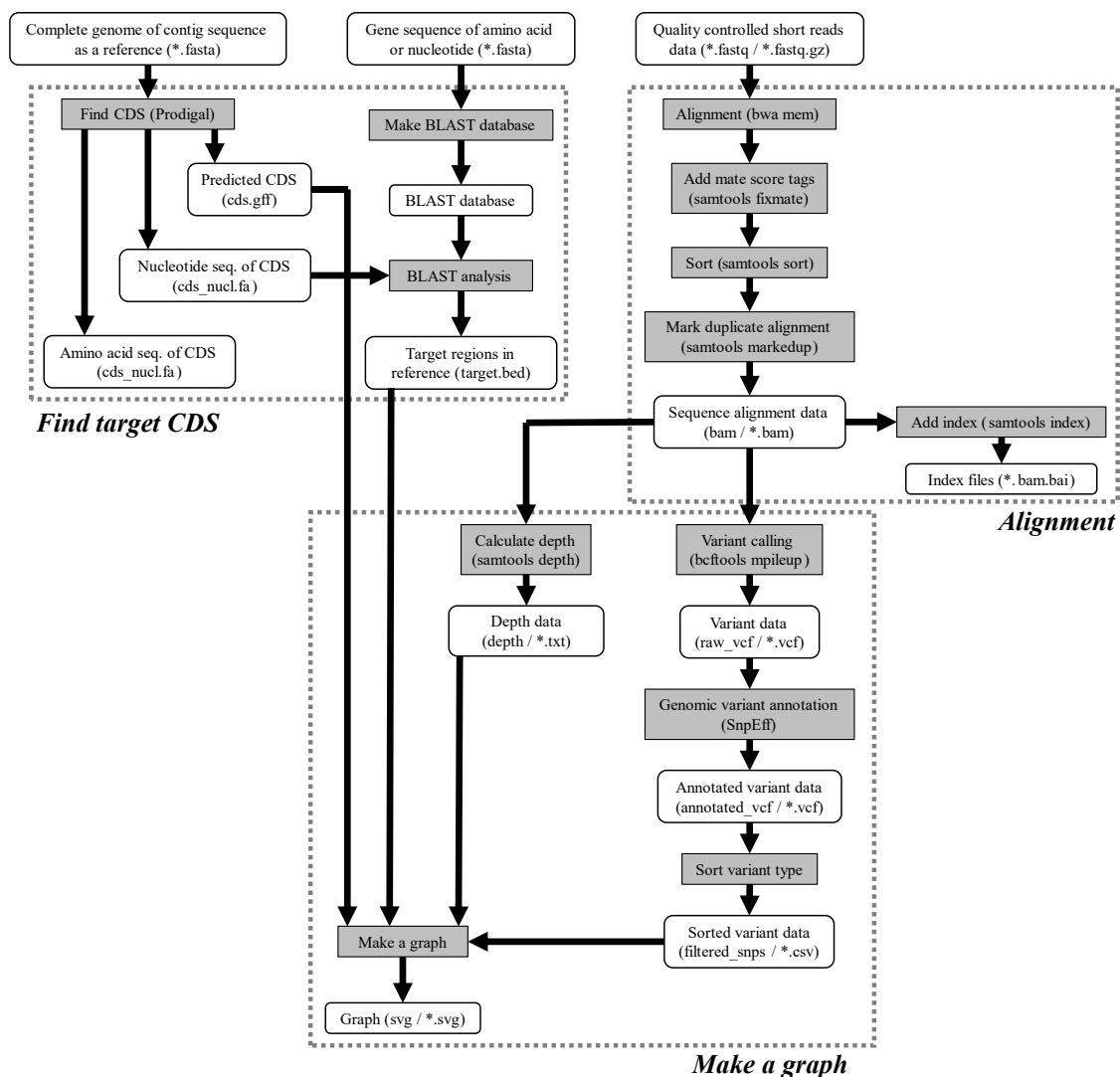


図 18 MMViewer の解析フローチャート

MMViewer を用いた SNP 解析

遺伝子のミスセンス変異は、産生されるタンパクのアミノ酸配列が変わる変異を指し、産生されるタンパクの活性が変わる可能性があることから重要な変異と考えられる。 α ヘモリシンやその産生に関わるタンパク質をコードする *hlyCABD*、ヘモリシン E をコードする *hlyE*、ヘモリシン F をコードする *hlyF* の各遺伝子のミスセンス変異と、それぞれの遺伝子の上流 2,000bp の変異について、MMViewer を用いて解析した。 α ヘモリシンの溶血活性に関わる遺伝子としては、 α ヘモリシン産生に関わる遺伝子 (*tolC*、*rfaH*、*hns*)、LPS 内核の生合成に関与する遺伝子 (*waaC*、*waaF*、*waaG*、*rfaE*)、細胞質シャペロン活性に関与する遺伝子 (*dnaK*、*dnaJ*)、そのほかの遺伝子 (*acrR*、*rne*) が知られる⁴⁸。これらの遺伝子のミスセンス変異についても、MMViewer を用いて解析した。

hlyCABD、*hlyE*、*hlyF* の SNP 解析の結果はそれぞれ図 19、図 20、図 21 のようになった。図中の矢印は対象とした CDS 領域とその方向を示している。マーカーはリファレンス

配列に対する変異を表している。CDS 領域内はミスセンス変異のみをマーカーで示しており、ナンセンス変異（アミノ酸変異を伴わない変異）は表示していない。CDS 領域外についてはすべての変異をマーカーで示している。変異の種類は一塩基置換（SNP: snp）

（例：A => T）、複数塩基置換（MNP: mnp）（例：GC => AT）、挿入（Insertion: ins）

（例：ATT => AGTT）、欠失（Deletion: del）（例：ACGG => ACG）、SNP や MNP の組み合わせ（complex）（例：ATTC => GTTA）に分けてマーカーの形状で分類した。また、アラインメントされなかった領域はグレーで塗りつぶされている。さらに、変異の検出頻度をマーカーの大きさで表している。

hlyCABD オペロンは、上流から *hlyC*、*hlyA*、*hlyB*、*hlyD* の順で遺伝子が並んでいる。WWo055 と WWo083 を除く株で *hlyCABD* にアラインメントされた。このことは 3.2.1.2 で確認した *hlyCABD* オペロンの保有の結果と一致した。*hlyCABD* 遺伝子をコードする領域の変異のパターンは [WWo073、WWo070、KOr014、WWo108、WWo014、WWo005、KKa004、WWo156、KKo007、WWo027] のグループ 1、[WWo003、WWo147] のグループ 2、[WWo102] のグループ 3、の 3 グループに分けることができる。グループ 1 は、ST95 (WWo073、WWo070、KOr014、WWo108、WWo014、WWo005)、ST961 (KKa004)、ST127 (WWo156、KKo007、WWo027) が属する 10 株が分類された。グループ 2 は ST73 (WWo003、WWo147) に属する 2 株が分類された。グループ 3 の WWo102 は ST83 に近い株で、他の株のシーケンスタイプとは異なっていた。今回 *hlyCABD* オペロンのリファレンスとして使用した S65EC 株は ST131 に分類され⁴⁸、いずれの株でもリファレンスに対する変異が見つかった。グループ 1 の中で KOr014 のみにおいて *hlyA* で SNP (c.841G>T : p.Gly281Cys) が確認された。KOr014 は同じグループ 1 に属するほかの株（ただし WWo108 を除く）に比べ、溶血活性がやや小さかった（図 12）。KOr014 の溶血活性は *hlyA* の SNP (c.841G>T : p.Gly281Cys) によって小さくなっている可能性が示唆される。

リファレンスとして使用した S65EC 株の *hlyCABD* の転写開始点は、*hlyC* 遺伝子の 1,616 bp 上流、ゲノム上の 1,812,160 の位置に存在する。*hlyCABD* の転写は、*hlyC* 遺伝子の上流に位置する JUMP Start 配列や転写制御に関わる -10 領域や -35 領域によって制御されると考えられている⁴⁸。リファレンス配列として用いた S65EC 株のゲノム上で、JUMP Start 配列は *hlyC* 遺伝子の 634 bp 上流の 4,813,142 の位置から 4,813,103 の間に位置する。また転写開始点は *hlyC* の 1,616 bp 上流の 4,812,160 の位置で、-10 領域は 1,812,150 付近、-35 領域は 1,812,125 付近に位置する。これらの領域での変異はいずれの株においても見つからなかった。

α ヘモリシンの溶血活性に関わる *tolC*、*rfaH*、*hns*、*waaC*、*waaF*、*waaG*、*rfaE*、*dnaK*、*dnaJ*、*acrR*、*rne* 遺伝子はすべての株で保有することが確認できた（図 22、図 23、図 24）。WWo003 と WWo147 はいずれも ST73 に分類されたが、溶血活性は大きく異なっていた。この 2 株間で異なる変異パターンは、*tolC* 遺伝子と *rfaE* 遺伝子で見られた。*tolC* 遺伝子については、WWo147 に 5 つの SNPs (c.52T>A : p.Ser18Thr、c.663A>T : p.Glu221Asp、c.697A>G : p.Thr233Ala、c.1421T>C : p.Val474Ala、c.1432A>G : p.Thr478Ala) が確認された一方、WWo003 ではこれらの SNPs は確認されなかった（図 22）。*rfaE* 遺伝子については、WWo147 で 2 つの SNPs (c.734C>A : p.Ala245Glu、c.1418T>A : p.Leu473Gln) が確認されたが、WWo003 でこれらの SNPs は確認されなかった。ST73 に分類される大腸菌では、*tolC* の SNPs (c.52T>A : p.Ser18Thr、c.663A>T : p.Glu221Asp、c.697A>G : p.Thr233Ala、c.1421T>C : p.Val474Ala、c.1432A>G :

p.Thr478Ala) あるいは *rfaE* の SNPs (c.734C>A : p.Ala245Glu, c.1418T>A : p.Leu473Gln) が溶血活性に影響を与える可能性が示唆される。

WWo108 は ST95 に分類された株の中で、他の ST95 の大腸菌株よりも溶血活性が小さな株であった。WWo108 に特徴的な変異として、*rfaH* での SNP (c.50C>G : p.Ala17Gly) が確認された。*rfaH* の変異が溶血活性に関わる可能性がある。KKa004 も ST95 に分類された株で、他の ST95 の株よりも溶血活性が大きな株であった。KKa004 に特徴的な変異は、*rfaH*、*rne*、*waaG* で確認された。*rfaH* で特徴的な変異は SNP (c.95T>C : p.Met32Thr) であった。*rne* では、他の ST95 に分類された株で SNP (c.705A>C : p.Lys235Asn) が確認されたが、KKa004 ではこの SNP は確認されなかった。*waaG* は、他の ST95 に分類された株で、Complex (c.252T>C; c.248G>A : p.Val84Val; p.Arg83Lys) が確認されたが、KKa005 では確認されなかった。ST95 に分類される大腸菌では、*rfaH* の SNP (c.50C>G : p.Ala17Gly) あるいは *rne* の SNP (c.705A>C : p.Lys235Asn) が溶血活性に影響を与えている可能性が示唆される。

ST127 に分類された株は、いずれも溶血活性の大きな株で、*hlyCABD* の配列は ST95 とほぼ同じであった。ST95 と異なる点として、*tolC* の SNP (c.805A>G : p.Thr269Ala) や *rne* の SNPs、*waaG* の SNP (ST95 でのみ c.248G>A : p.Arg83Lys, c.252T>C : p.Val84Val) が確認できた。ST95 に比べ ST127 に分類される大腸菌の溶血活性が大きい理由として、これらの SNPs が関わっている可能性が示唆される。

WWo102 は、*hlyCABD* オペロン、*waaC*、*rfaE*、*rne* でほかのシーケンスタイプの大腸菌とは異なる変異パターンを示した。*waaC* では、他の *hlyCABD* オペロン保有株が 2 つの SNPs (c.235G>A : p.Glu85Lys, c.301A>G : p.Thr101Ala) を確認できたが、WWo102 では確認されなかった。*rfaE* では、1 つの SNP (c.1418T>A : p.Leu473Glu) が確認され、WWo147 で確認された SNPs の 1 つと同様の変異であった。*rne* では多くの変異が他のシーケンスタイプとは異なっていた。

hlyE 遺伝子の保有と変異は図 20 のようになった。*hlyE* 遺伝子は、WWo055 を除くすべての株で一部分のみがアラインメントされた。WWo055 は c.494delA : p.Tyr165fs のフレームシフト変異を引き起こしており、*hlyE* は不活化されていると考えられる。3.2.1.4 で *hlyE* は WWo083 でのみヒットしており、この解析では *hlyE* に対してはアラインメントされなかった。本解析ではリファレンスとして K-12 substr. MG1655 株の *hlyE* 遺伝子配列を使用した。K-12 substr. MG1655 株と WWo083 の *hlyE* 遺伝子の塩基配列の差が大きくアラインメントがうまくできなかった可能性がある。

hlyF 遺伝子の保有と変異は図 21 のようになった。*hlyF* には WWo055 のみがアラインメントされ、この結果は 3.2.1.4 の結果と一致した。リファレンス配列に対するミスセンス変異は 4 つの SNPs (c.162A>T : p.Lys54Asn, c.162A>T : p.Lys54Asn, c.694A>G, p.Asn232Asp, c.1015C>T : p.Pro339Ser) が確認された。*hlyF* 遺伝子の上流約 400 bp はアラインメントされなかった。アラインメントされなかったことは、その領域の塩基配列が大きく異なることや、その領域が存在しない可能性が考えられる。遺伝子上流は転写制御等に関わる可能性があり、リファレンスとして用いた大腸菌 VREC0578 株とは異なる転写制御によって発現が制御されている可能性が示唆される。

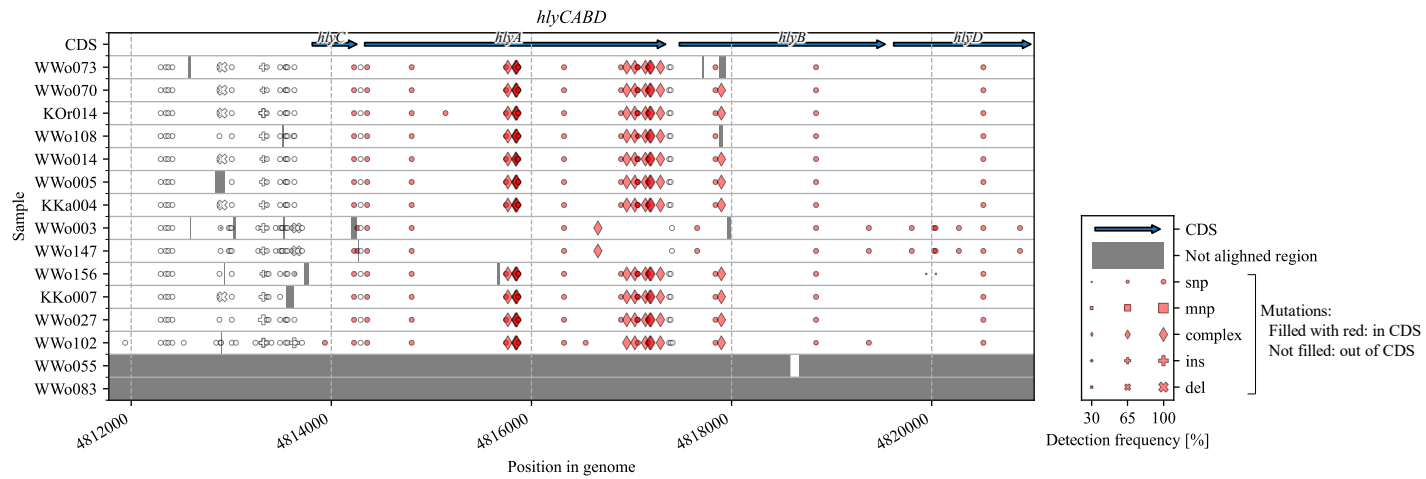


図 19 *hlyCABD* オペロン内のミスセンス変異とその上流 2000 bp の変異

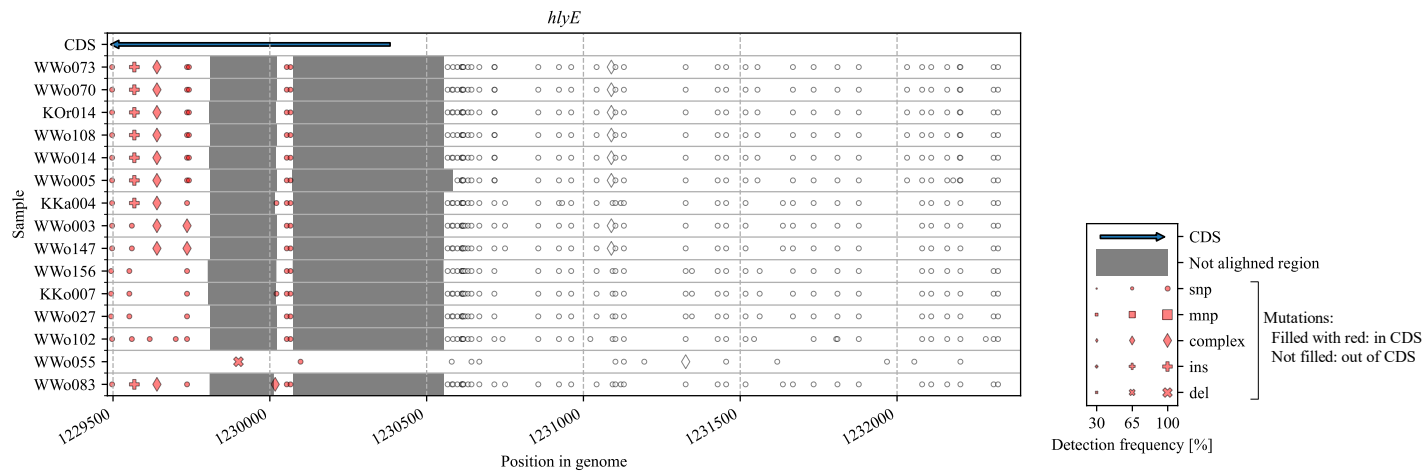


図 20 *hlyE* とその上流 2000 bp の変異

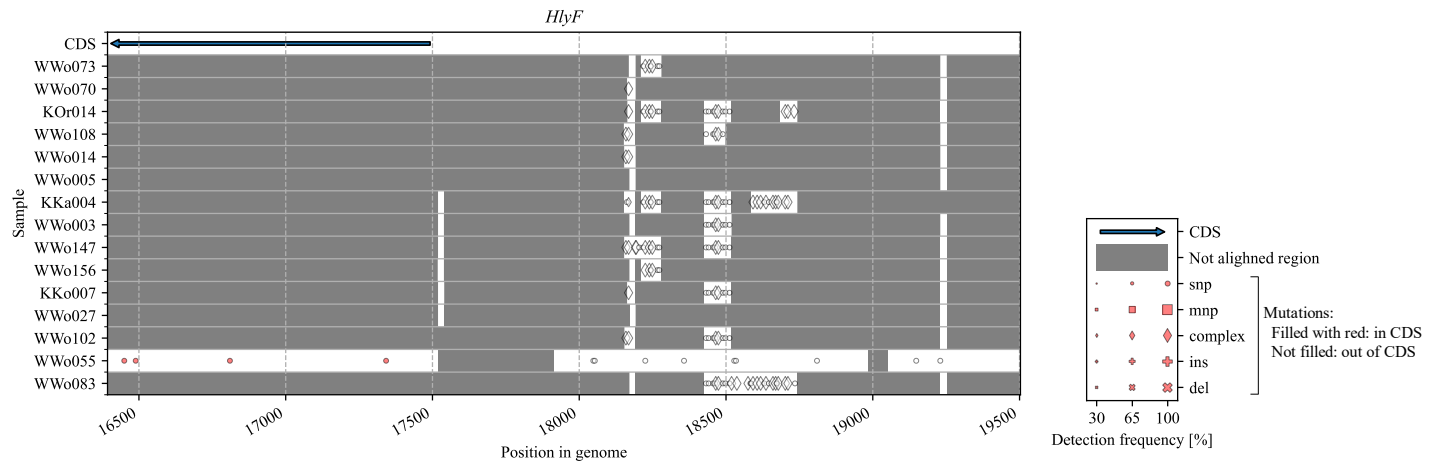


図 21 *hlyF* とその上流 2000 bp の変異

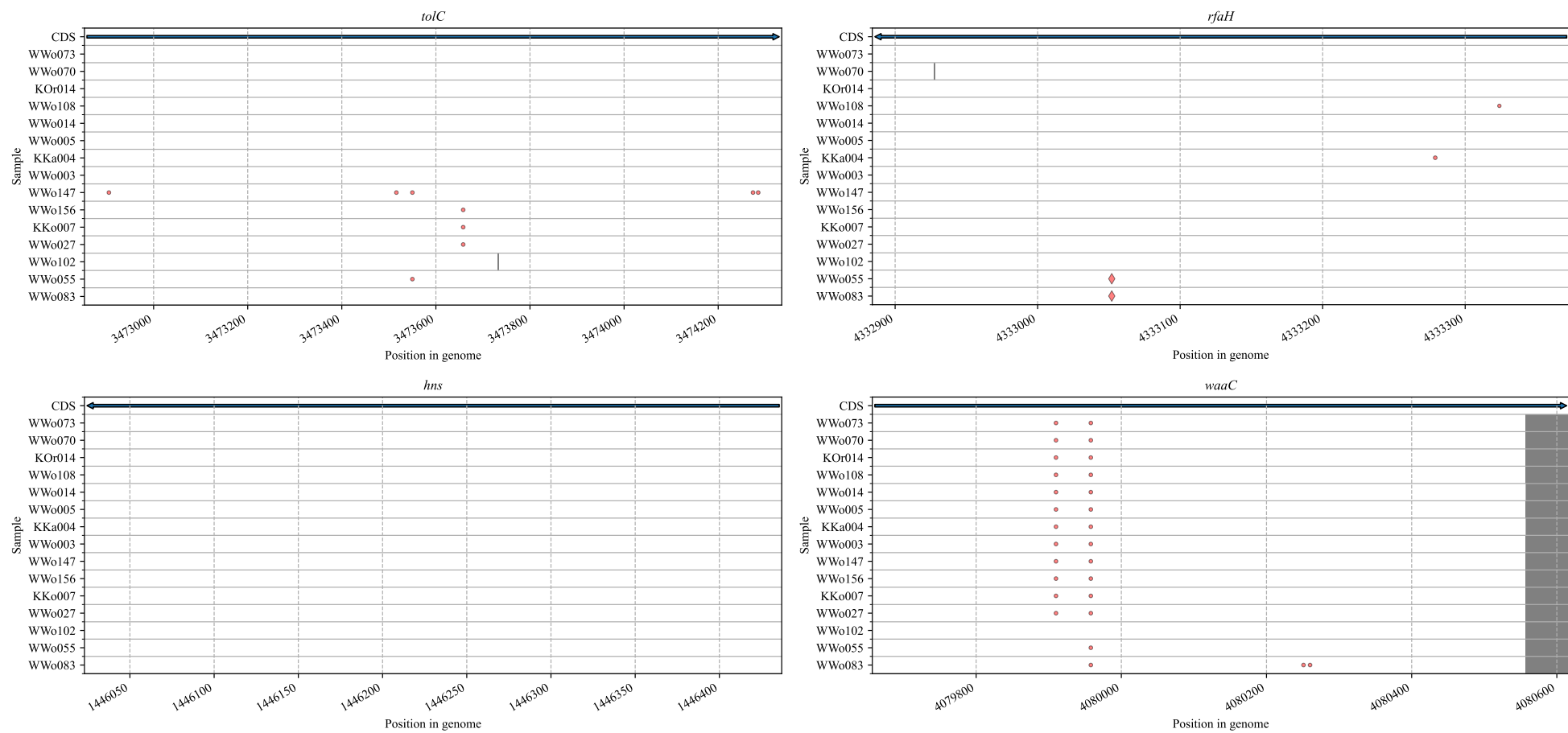


図 22 α ヘモリシンの溶血活性に関わる遺伝子のミスセンス変異 (1/3)

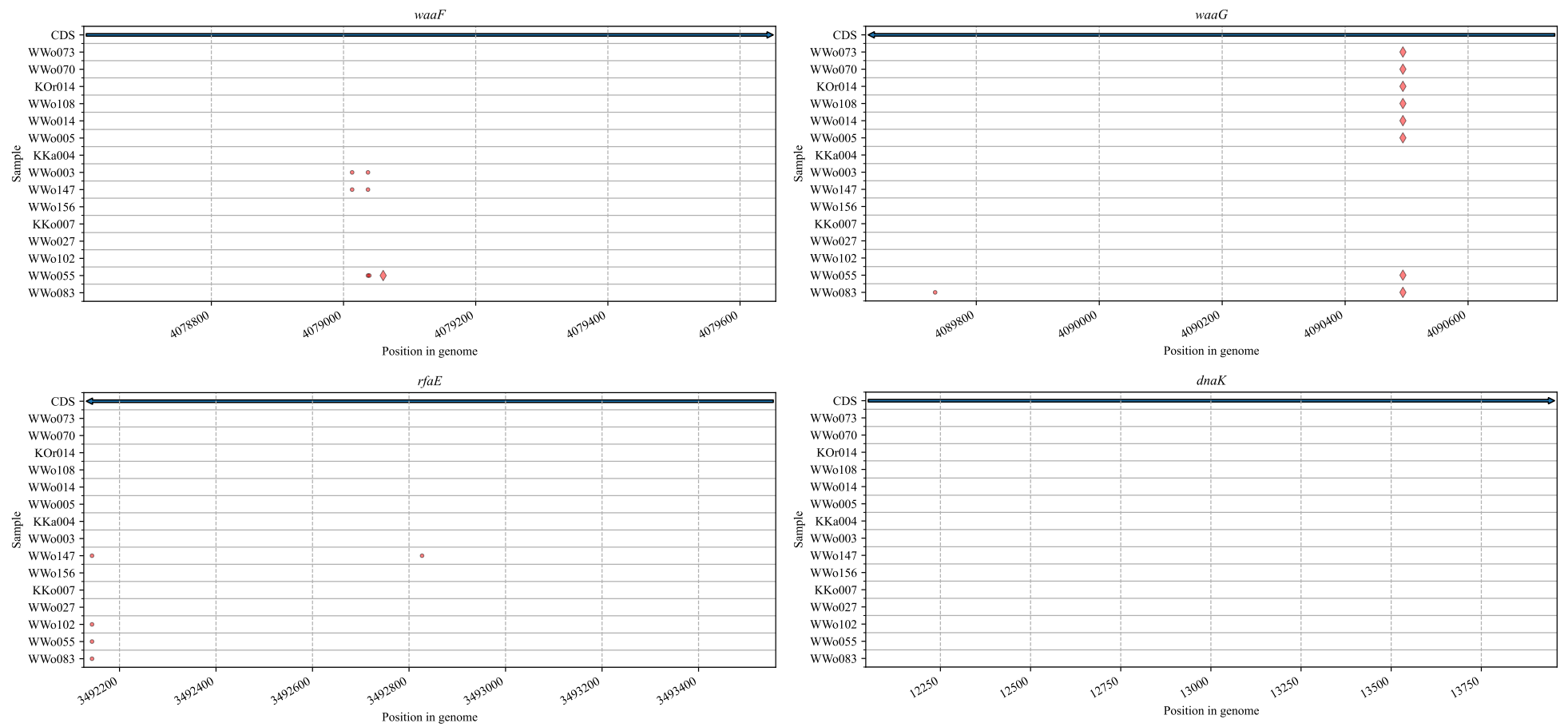


図 23 α ヘモリシンの溶血活性に関わる遺伝子のミスセンス変異 (2/3)

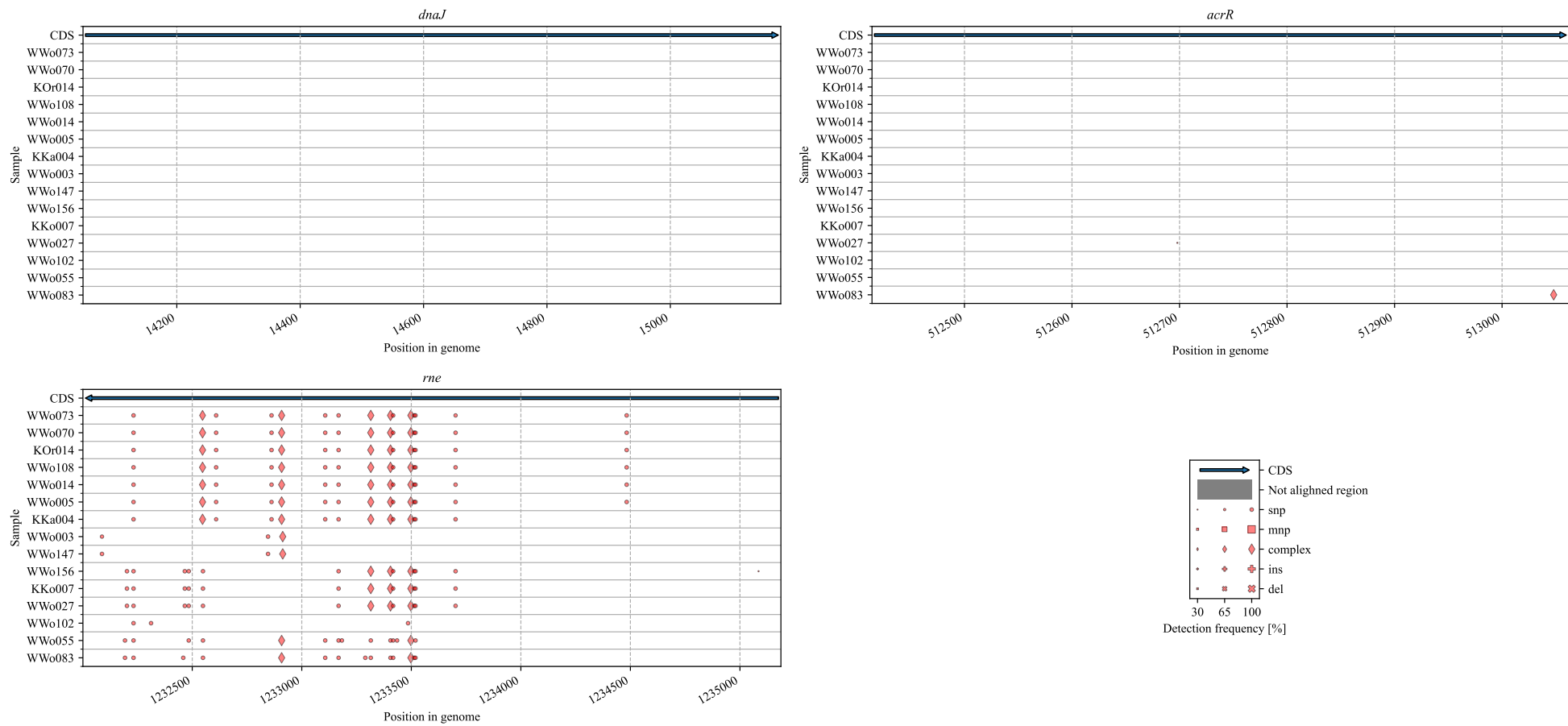


図 24 α ヘモリシンの溶血活性に関わる遺伝子のミスセンス変異 (3/3)

3.3. 大腸菌以外の溶血性細菌

大腸菌以外に、複数の溶血性細菌を単離することができた。本研究では、表 9 に示す *Aeromonas* 属を含む 10 株の全ゲノムシーケンスを行った。ここでは、大腸菌以外の溶血性細菌の種、保有する溶血毒遺伝子、溶血毒以外に保有する病原性遺伝子について述べる。

3.3.1. 溶血斑の再確認

Aeromonas 属以外の 7 株は不明瞭な溶血斑を示していたため、溶血斑を血液寒天プレートタイプ III で再確認した。培養後 24 時間、48 時間、72 時間の溶血斑の様子を図 25 に示す。いずれの株も溶血斑を示すことが確認された。

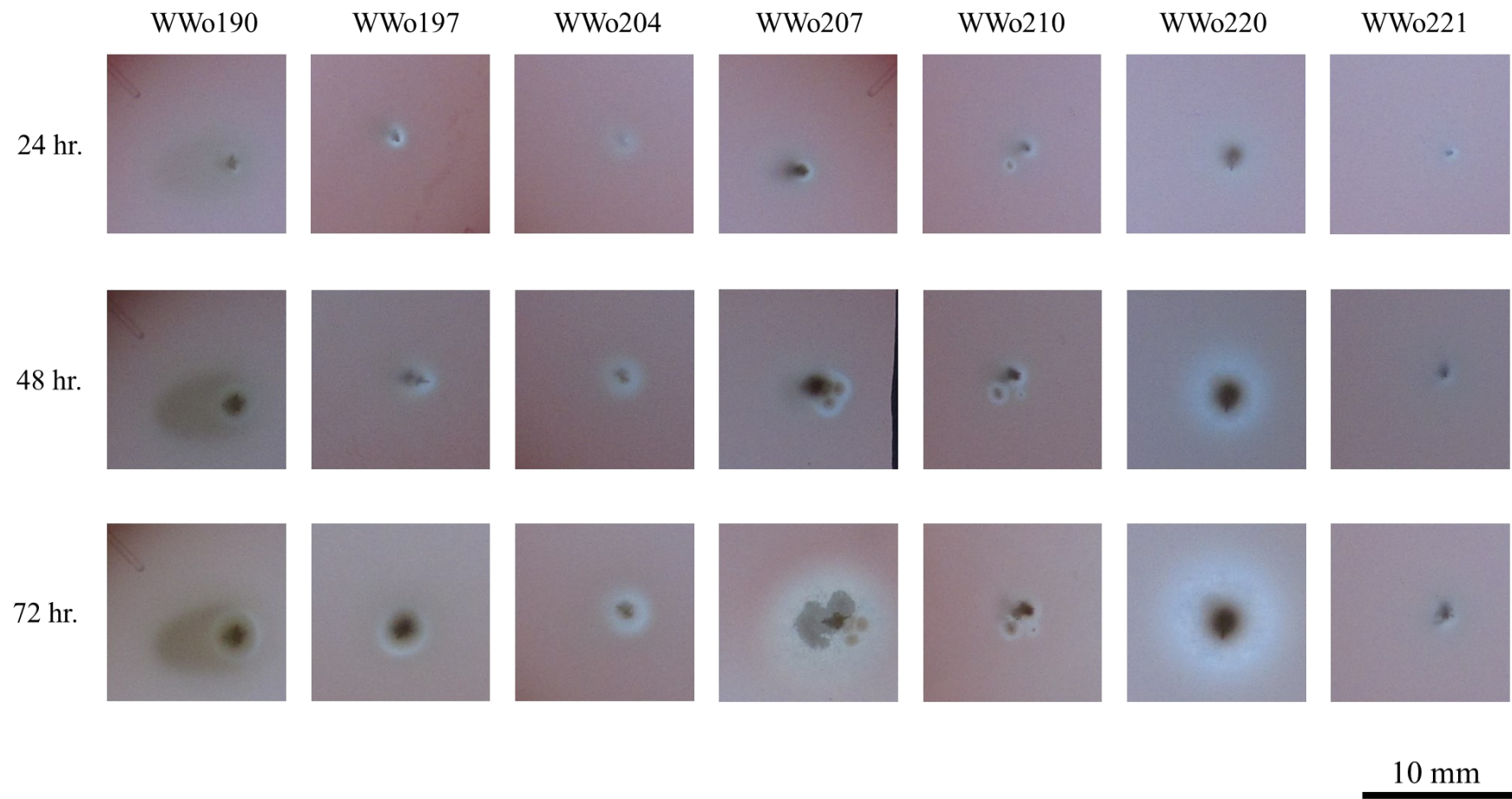


図 25 血液寒天プレートタイプ III で確認した溶血斑の様子

3.3.2. 種の同定

3.1.4 で、都市下水から単離した溶血性細菌の 16S rRNA 遺伝子を解析することにより単離株の種推定を行ったが、V3-4 領域だけでは種の識別としては十分な長さではない。そこで、全ゲノムシーケンスを行った株のシーケンスデータをもとに種の同定を行った。

まず、Ribosomal Multilocus Sequence Typing (rMLST) を用いて、種の推定を行った⁶²。表 14 は全ゲノムシーケンシングを行った単離溶血性細菌 10 株の、2.6 で 16S rRNA 遺伝子 V3-4 領域解析から推定された種と、rMLST で推定された種を表示している。rMLST ではリボソームタンパクサブユニットをコードする 53 個の *rps* 遺伝子のバリエーションから細菌種を推定する。rMLST の Support は、53 個の *rps* 遺伝子のうち単離株のシーケンスデータから見つかった *rps* 遺伝子の割合を表す。

16S rRNA 遺伝子 V3-4 領域の解析で推定された種と rMLST で推定された種の多くは異なっていた (表 14)。WWo011 と WWo140、WWo148 はいずれも *Aeromonas* 属である点は一致していたが、16S rRNA 遺伝子解析による推定では複数の種が推定されており、rMLST 解析の結果と一致していなかった。WWo190 は、rMLST の support 値が 18% と低く、rMLST では正確な推定を行うことができずに 2 つの種が推定された。WWo197 は 16S rRNA V3-V4 領域の解析の結果 2 種が推定されたが、rMLST 解析の結果、そのうちのひとつである *Streptococcus lutetiensis* が推定された。*Leuconostoc lactis* (WWo204)、*Lactococcus taiwanensis* (WWo207)、*Ligilactobacillus salivarius* (WWo210)、*Enterococcus casseliflavus* (WWo220) は、16S rRNA V3-4 領域の解析から推定された種と、RMLST で推定された種は同じであった。WWo221 は、support 値が 2% と一つの *rps* 遺伝子しか見つからず、rMLST では正確な推定を行うことはできず、属レベルでしか推定することができなかった。

続いて、rMLST で推定された種の type strain の配列と単離株のゲノム配列の average nucleotide index (ANI) を計算し、種の同定を行った。ANI の計算結果を表 15 に示す。rMLST で推定された種を species の列、その時の ANI の値を ANI value の列に表している。また、ANI が 95% 以上で type strain と同一と判断できる種について、species の列と ANI value の列のそれぞれの値を下線付きの太字で表している。Assembly accession の列は、ANI のリファレンスとして使用した基準株 (type strain) の INSD アクセッション番号を表す。

WWo011、WWo140、WWo148、WWo197、WWo204、WWo207、WWo210、WWo220 は、rMLST で推定された種と同一の種として同定された。WWo190 は、rMLST 解析の結果 *Streptococcus suis* と *Streptococcus parasuis* の両方に推定されたが、それぞれの ANI 値は 86.5% と 96.9% であり、*Streptococcus parasuis* と同定できた。WWo221 は、rMLST 解析の結果 *Streptococcus* 属までしか推定できなかったため、*Streptococcus* 属に含まれる 62 種のすべてについて、それぞれの type strain との ANI 値を計算した。その結果、*Enterococcus lemanii* が最も高い ANI 値を示し、98.3% であった。ほかの種では *Enterococcus aquimarinus* と *Enterococcus alcedinis* がともに 78.6% を示し、それ以外の種は 70% 未満であったため、WWo221 は *Enterococcus lemanii* と同定された。

表 14 16S rRNA V3-4 領域から推定した種と rMLST で推定した種

Strain ID	16S rRNA V3-V4 region analysis		rMLST	
	Genus	Species	Taxon	Support
WWo011	<i>Aeromonas</i>	<i>Aeromonas dhakensis</i> <i>Aeromonas taiwanensis</i>	<i>Aeromonas caviae</i>	100%
WWo140	<i>Aeromonas</i>	<i>Aeromonas encheleia</i> <i>Aeromonas hydrophila</i> <i>Aeromonas molluscorum</i>	<i>Aeromonas hydrophila</i>	100%
WWo148	<i>Aeromonas</i>	<i>Aeromonas caviae</i> <i>Aeromonas hydrophila</i> <i>Aeromonas jandaei</i>	<i>Aeromonas caviae</i>	100%
WWo190	<i>Streptococcus</i>	<i>Streptococcus porcorum</i> <i>Streptococcus sanguinis</i> <i>Streptococcus sinensis</i>	<i>Streptococcus suis</i> <i>Streptococcus parasuis</i>	18% 18%
WWo197	<i>Streptococcus</i>	<i>Streptococcus equinus</i> <i>Streptococcus lutetiensis</i>	<i>Streptococcus lutetiensis</i>	100%
WWo204	<i>Leuconostoc</i>	<i>Leuconostoc lactis</i>	<i>Leuconostoc lactis</i>	95%
WWo207	<i>Lactococcus</i>	<i>Lactococcus taiwanensis</i>	<i>Lactococcus taiwanensis</i>	91%
WWo210	<i>Lactobacillus</i>	<i>Lactobacillus salivarius</i>	<i>Ligilactobacillus salivarius</i>	100%
WWo220	<i>Enterococcus</i>	<i>Enterococcus casseliflavus</i>	<i>Enterococcus casseliflavus</i>	100%
WWo221	<i>Enterococcus</i>	<i>Enterococcus lemanii</i>	<i>Enterococcus</i>	2%

表 15. fastANI の結果

strain ID	Species	Support	Assembly accession	ANI value
WWo011	<u><i>Aeromonas caviae</i></u>	100%	GCF_000783715.2	98.2
WWo140	<u><i>Aeromonas hydrophila</i></u>	100%	GCF_017310215.1	96.9
WWo148	<u><i>Aeromonas caviae</i></u>	100%	GCF_000783715.2	97.9
WWo190	<i>Streptococcus suis</i>	18%	GCF_000026725.1	86.5
	<u><i>Streptococcus parasuis</i></u>	18%	GCF_004283785.1	96.9
WWo197	<u><i>Streptococcus lutetiensis</i></u>	100%	GCF_900475675.1	99.3
WWo204	<u><i>Leuconostoc lactis</i></u>	95%	GCF_002287365.1	97.3
WWo207	<u><i>Lactococcus taiwanensis</i></u>	91%	GCF_017068355.1	98.1
WWo210	<u><i>Ligilactobacillus salivarius</i></u>	100%	GCF_001011095.1	98.3
WWo220	<u><i>Enterococcus casseliflavus</i></u>	100%	GCF_000157355.2	98.2
WWo221	<i>Enterococcus aquimarinus</i>	2%	GCF_001885765.1	78.6
	<u><i>Enterococcus lemanii</i></u>	2%	GCF_016909125.1	98.3
	<i>Enterococcus alcedinis</i>	2%	GCF_014635985.1	78.6
	Other <i>Enterococcus</i> sp. (total 59)	2%		<70.0

3.3.3. 保有する病原毒素遺伝子

溶血性を示した細菌の保有する溶血毒遺伝子について調べた。溶血性細菌の保有する病原性遺伝子を VFAnalyzer (<http://www.mgc.ac.cn/VFs/>)⁷¹を用いて探索した。また、溶血性細菌のコンティグに対して prokka⁷²を用いてアノテーションを行った。VirulenceFinder2.0でヒットした溶血毒と、prokkaによってヘモリシンとしてアノテーションされた遺伝子を表 16 にリストアップした。

VFAnalyzer を用いた探索により、*Aeromonas* の WWo011、WWo148、WWo140 はいずれもヘモリシン HlyA、ヘモリシン III、熱安定性ヘモリシン (TH) を保有することが分かった。また、*Aeromonas hydrophila* の WWo140 はこれらの溶血毒遺伝子に加え、エロリジン (AerA)、細胞外ヘモリシン AHH1、リピートイントキシン (RTX) を保有していた。RTX は *rtxACHBDE* の 6 遺伝子から構成され、そのうちの 5 遺伝子 (*rtxAHBDE*) で保有が確認された。prokka⁷²によるアノテーションでは、これらの遺伝子のなかで WWo140 が保有するエロリジン (AerA)、ヘモリシン HlyA の 2 つがアノテートされた。*Streptococcus* の WWo190 と WWo197、*Leuconostoc* の WWo204、*Ligilactobacillus* の WWo220、*Enterococcus* の WWo221 は、prokka によるアノテーションによって、ヘモリシン A をコードする *tlyA* 遺伝子がアノテートされた。加えて、WWo197 は VFAnalyzer の探索により Cytolysin を保有していた。*Enterococcus casseliflavus* の WWo221 は溶血毒遺伝子を確認することはできなかった。

表 16 VFAnalyzer⁷¹ で予測された毒素遺伝子の保有と prokka⁷² でヘモリシンとしてアノテーションされた遺伝子

Strain	Toxins predicted by VFAnalyzer (gene)	Hemolysins annotated by prokka (gene)
BC011 (<i>Aeromonas caviae</i>)	Hemolysin HlyA (<i>hlyA</i>) Hemolysin III (Undetermined) Thermostable hemolysin; TH (Undetermined)	
BC148 (<i>Aeromonas caviae</i>)	Hemolysin HlyA (<i>hlyA</i>) Hemolysin III (Undetermined) Thermostable hemolysin; TH (Undetermined)	
BC140 (<i>Aeromonas hydrophila</i>)	Aerolysin AerA/Cytotoxic enterotoxin (<i>aerA/act</i>) Extracellular hemolysin AHH1 (<i>ahh1</i>) Hemolysin HlyA (<i>hlyA</i>) Hemolysin III (Undetermined) The repeat in toxin; RTX (<i>rtxAHBDE</i>) Thermostable hemolysin; TH (Undetermined)	Aerolysin AerA/Cytotoxic enterotoxin (<i>aerA/act</i>) Hemolysin HlyA (<i>hlyA</i>)
BC190 (<i>Streptococcus parasuis</i>)		Hemolysin A (<i>tlyA</i>)
BC197 (<i>Streptococcus lutetiensis</i>)	Cytolysin	Hemolysin A (<i>tlyA</i>)
BC204 (<i>Leuconostoc lactis</i>)		Hemolysin A (<i>tlyA</i>)
BC207 (<i>Lactococcus taiwanensis</i>)		Hemolysin A (<i>tlyA</i>)
BC210 (<i>Ligilactobacillus salivarius</i>)	Hemolysin (<i>Clostridium</i>)	Hemolysin A (<i>tlyA</i>)
BC220 (<i>Enterococcus casseliflavus</i>)		
BC221 (<i>Enterococcus lemanii</i>)		Hemolysin A (<i>tlyA</i>)

4. 考察

4.1. 溶血性大腸菌の遺伝型と保有する病原性遺伝子

VirulenceFinder2.0 を用いた遺伝子保有の確認、病原型推定で用いた遺伝子保有の確認、遺伝子変異解析で得られた結果の間でいくつか齟齬が見られた。これは 1.5 で述べたリードのアセンブル方式の違い等によって生じたものであると考えられる。まず EAST-1 熱耐性毒素は、病原型解析の結果、WWo003、WWo147、WWo156、WWo102、WWo055、WWo083 で *EAST1* 遺伝子の一部の領域にアラインメントされた。一方、VirulenceFinder2.0 を用いた解析では WWo055 のみが EAST-1 熱耐性毒素をコードする *estA* 遺伝子の保有が確認された。WWo055 は病原型解析の際も *EAST-1* 遺伝子領域の **gene coverage** は 0.63 で最も高い値を示していた。アラインメントでは、対象遺伝子周辺の DNA 配列が異なる場合、遺伝子の末端領域に対してリードがアラインメントされず、**gene coverage** の値が 100% よりも小さい値として検出されることがある。また、アミノ酸配列が同じ場合でも塩基配列が異なる場合があり、このような場合には **gene coverage** の値は小さい値になると考えられる。一方、VirulenceFinder2.0 を用いた解析では、コンティグ上の遺伝子を探索する。WWo003、WWo147、WWo156、WWo102、WWo083 は、*estA* と部分的に同じ塩基配列をもつ別の遺伝子を保有していた可能性と、コンティグに乗らなかったために VirulenceFinder2.0 で検出できなかった可能性がある。

papA は、病原型推定で WWo073、WWo070、KOr014、WWo108、WWo014、WWo005、KKa004、WWo003 の **gene coverage** は 1 に近い値であったが、WWo147、WWo156、KKo007、WWo027、WWo102 ではやや小さい値となった。ここで 1 に近い値を示した株は VirulenceFinder2.0 を用いた解析ではすべて *papA_F43* で高い一致率を示していた。一方、それ以外の株は *papA_F12* や *papA_F48* と高い一致率を示した。これは、*papA* 遺伝子に複数のアレルが存在してその相違性によって *papA_F43* 以外の株は病原型推定時に **gene coverage** の値が小さくなったものと考えられる。

hlyE の保有を調べると、VirulenceFinder2.0 を用いた解析では WWo083 だけが保有していたが、変異解析では WWo055 のみがアラインメントされた。WWo055 は途中でフレームシフト変異が入っていたことから、WWo055 で *hlyE* は不活化していると考えられる。WWo083 でアラインメントされなかった理由としては、WWo083 の保有する *hlyE* 遺伝子の塩基配列と、リファレンスとして使用した大腸菌 K-12 株 substr. MG1655 の *hlyE* 遺伝子の塩基配列の間の違いが大きく、アラインメントが適切に行われなかった可能性がある。*hlyE* 遺伝子の塩基配列のパターンを複数用意し、改めてアラインメントを行い *hlyE* 遺伝子の保有を確認する必要がある。

hlyCABD オペロンを保有する大腸菌株はいずれも B2 phylotype を示した。MLST 解析の結果、これらの株は ST95、ST73、ST127、ST961、ST117 と、ST83 に近縁のシーケンスタイプに分類された。ST95、ST73、ST127、ST117 は主要な ExPEC 株として報告されている¹²⁷。一方、ST83、ST5073 は ExPEC としての報告は多くない。ExPEC としての報告が最も多く薬剤耐性遺伝子を保有することから臨床的に重要とされる ST131 は¹²⁷、今回全ゲノムシーケンスの対象とした 15 株の大腸菌に含まれなかった。

単離された溶血性大腸菌の中では ST95 に分類されたものが最も多く、6 株が ST95 に分類された。いずれの株も *FimH* の分類が H18 で血清型の H 抗原が 7、O 抗原の分かったものでは 18ac もしくは 18 で、kSNP3 系統樹でも近縁の株であった (図 12)。ST95 は、尿路

感染症や新生児髄膜炎、菌血症との関係が報告されている^{99,128-130}。ST95やST73は抗生物質耐性を持たない尿路感染性大腸菌に多く見られるシーケンスタイプ(ST)として知られる。これらのシーケンスタイプの株はプラスミドを保有しづらいつと考えられており、薬剤耐性遺伝子の組み込みも少ないと考えられている¹³¹。

ST95には、ヒトを主な宿主とするExPECとAPECが含まれる¹²⁸。ST95内で、APECとヒトのExPECを区別することは難しいが、5つの病原性遺伝子(*ibeA*、*sfaS*、*sfa*、*cnf1*、*papG-III*)はヒトのExPECで独占的に見られる遺伝子で、これらの遺伝子を保有する株の*fimH*タイプはH18であることが報告されている¹²⁸。ST95に分類された6株が*cnf1*、*ibeA*を保有しており、*sfaS*もWWo108を除く5株で確認された(図14)。また*fimH*タイプはH18であったことから、これらの株はヒトに優先的に感染するExPECであることが示唆される。

KKa004株はST961に分類された。ST961は、臨床的に病変のある豚の腸管外臓器や、臨床的に病変のあるイヌから単離された報告があり、保有する病原毒素遺伝子から人獣共通病原体の可能性が示唆されている^{132,133}。ST95の株とkSNP3.0を用いた系統解析において近い位置にあり、ST95の株と共通する病原性遺伝子を数多く保有していた。加えて*mcmA*、*mchC*、*mcmF*、*mchB*、*cea*、*papA_F12*、*tcpC*を保有していた。*mcmA*、*mchC*、*mcmF*、*mchB*はWWo003やWWo102も保有していた。*mcmA*、*mchF*、*mchC*、*mchB*はいずれもシデロフォア-マイクロシン遺伝子アイランドに存在しており、*mcmA*はシデロフォア-マイクロシンの前駆体、*mcfF*はシデロフォア-マイクロシンの分泌に関わるABCトランスポーターを構成するタンパク質、*mchC*はシデロフォア-マイクロシンの翻訳後修飾に関わるタンパク質、*mchB*はシデロフォアマイクロシンMccH47の前駆体をそれぞれコードしている¹³⁴。遺伝子アイランドは移動可能な遺伝的要素として知られ、ダイレクトリピート構造と呼ばれる繰り返し配列によって挟まれている¹³⁵。移動可能な遺伝的要素であることから、ST番号に関わらずこれらの遺伝子が移動することで同じSTでもこれらの遺伝子の保有、非保有の差が生まれた可能性もある。また、繰り返し配列が存在するとショートリードからコンティグを作成することは困難になる¹³⁶。実際、KKo007株は*mcmA*のみがヒットしていたが、ヒットしたコンティグを調べてみると長さ2336bpのコンティグ上の1540番目の塩基から1818番目でヒットしていた(data not shown)。WWo003とWWo102で得られたコンティグ上で、*mchF*、*mchC*、*mchB*は、*mcmA*からそれぞれ254bp、4248bp、6070bp離れており(data not shown)、それぞれの遺伝子が近い位置に存在することが確認された。KKo007株では*mcmA*遺伝子周辺の配列が十分にアセンブルされておらず、シデロフォア-マイクロシン遺伝子アイランド周辺のコンティグ作成がうまくいっていない可能性が示唆された。

ST73は、WWo003とWWo147が分類された。ST73は、ブラジルの尿路感染症患者から単離された大腸菌の中では最も主要なシーケンスタイプであり¹³⁷、尿路感染症の重要な病原性細菌と考えられる。ヒトと家畜の両方から大腸菌を検出した研究において、ST73はヒトからのみ単離されたとの報告がある¹³⁸。WWo003の特徴としてはエアロバクチンの合成に関わる*iucC*遺伝子およびエアロバクチン受容体をコードする遺伝子*iutA*を保有する点が挙げられる。エアロバクチンシステムは、*iucABCDiutA*の5つのオープンリーディングフレームからなる遺伝子群によって構成されている¹³⁹。実際、WWo003で*iutA*と*iucC*は同じコンティグ上で1356bpだけ離れた位置に存在していた(data not shown)。エアロバクチンはシデロフォアの一つで、大腸菌の鉄獲得に関与する¹⁴⁰。

ST127は、尿路感染症、菌血症、新生児壊死性腸炎に関与している^{129,141,142}。イングランド北西部における調査によれば、尿路感染症患者から単離された尿路完成性大腸菌の中で6番目に多いシーケンスタイプであった¹⁴¹。抗生物質に対し広い感受性を示すが、ほかのシーケンスタイプに比べ多くの病原性遺伝子を保有して高い病原性を有することが報告されている¹⁴¹。ST127については、血清型 O6 が菌血症患者や尿路感染症患者から最も多く単離される血清型で¹²⁹、下水から単離された ST127 の3株が菌血症や尿路感染症を引き起こす可能性が示唆される。

hlyCABD オペロンを保有していた ST95、ST73、ST127、ST961 の株が共通して保有する病原性遺伝子に、*usp*、*clbB*、*cnfI* 等があり、*hlyCABD* オペロンを保有していない WWo055 と WWo083 ではこれらの遺伝子の保有は確認されなかった。*cnfI* のコードされている位置は、*hlyCABD* オペロンの近傍に位置していることが確認され (data not shown)、*hlyCABD* オペロンを含む病原性遺伝子アイランド内に存在することが示唆された。*cnfI* 遺伝子は *hlyA* 遺伝子とともに、尿路感染症における病態に寄与する¹⁴³。今回単離した大腸菌のコンティグ上において *cnfI* 遺伝子と *hlyCABD* オペロンが近い位置にあることから、これらの両遺伝子が協同的に働く可能性が示唆される。*clbB* 遺伝子は、*pks* アイランド内に存在する遺伝子で、遺伝毒性物質として知られるコリバクチン産生に関わる¹⁴⁴。*usp* 遺伝子はコリシン様 *usp* をコードする遺伝子で、この物質も遺伝毒性物質であることが知られる¹⁴⁵。Jinら (2016)²⁶によれば、女性の結腸がん患者の糞便から検出される *hlyA* 遺伝子を持つ大腸菌の割合が健常者よりも多いことが報告されている。また、同論文で、メスのマウスに対しては α ヘモリシンを産生する大腸菌を経口投与した場合、結腸がんが引き起こされることが報告されている。遺伝毒性物質は発がんに寄与すると考えられており、 α ヘモリシンを産生するこれらの大腸菌はヒトの腸内でコリバクチンやコリシン様 *usp* といった遺伝毒性物質とともに結腸がんの発症や進行に寄与している可能性がある。

既存のシーケンスタイプには分類されず、ST83 に最も近かった WWo102 は、kSNP3 系統解析においてもほかの溶血性細菌とは遺伝的に離れていることが示された (図 12)。ほかの ExPEC 同様、ExPEC に特徴的な病原性遺伝子を保有しているほか、KKa004 株や WWo003 が保有するシデロフォア-ミクロシン産生に関わる遺伝子 *mcmA*、*mchF*、*mchC*、*mchB* を保有していた。ST83 は、尿路感染症に感染したと疑われる猫の尿から単離された報告がある^{146,147}。Ksiezarekら (2021) の単離した大腸菌 Ec_151217 株の血清型は O83:H5 で¹⁴⁶、WWo102 と H 抗原が同じであった。ST83 O83:H5 の大腸菌はヒトからも単離されているが¹⁴⁶、ヒトに対する病原性は不明である。興味深いことに、*hlyCABD* オペロンを2つ保有することは稀であるが、Ec_151217 株では珍しく2つ保有する¹⁴⁶。WWo102 で2つのオペロンを保有しているかどうかは不明であるため、更なる解析が望まれる。

WWo055 は、B1 phylotype に分類され、シーケンスタイプは ST5073、血清型は H8:O8 であった。毒素遺伝子として、EAST-1 耐熱性毒素をコードする *astA* 遺伝子及び溶血毒 *hlyF* をコードする *hlyF* 遺伝子を保有していた。大腸菌 H8:O8 による感染症の例としては、2014年4月と2016年9月に、大分県の2つの寮で発生したアウトブレイクが挙げられる。両アウトブレイクでは、ほとんどの患者が下痢と腹痛を訴え、発熱、嘔吐、頭痛などの症状が見られた。保健所の実施した微生物検査では、下痢症等を引き起こす既知の病原性細菌は検出されず、代わりに病原性大腸菌としてはあまり知られていない大腸菌 O8 が患者の便検体のほとんどから分離された¹⁴⁸。H8:O8 の中には下痢症を引き起こす大腸菌が含まれることが示唆され、WWo055 も病原性をもつ可能性が示唆される。また、保有する病原性遺伝子には、*iutA*、*hlyF*、*iss*、*iroN* があり、これらの遺伝子は鶏のコリバシラス症を発症す

る APEC に特徴的な遺伝子である¹⁴⁹。特に *hlyF* 遺伝子は、ColV プラスミドにコードされる遺伝子の一つで、APEC や NMEC の疫学マーカーであることが示されている¹⁵⁰⁻¹⁵²。*hlyF* 遺伝子を持つ大腸菌は、羊血液寒天培地上で弱い溶血活性を示す¹⁵³。ニワトリ胚の感染モデルで *hlyF* 遺伝子を保有する大腸菌は、LB 培地やヒトの尿中に比べて *hlyF* の転写が強く亢進されており（LB 培地で培養したときの約 29 倍）、*hlyF* が鳥類の腸管外感染の成立に関与している可能性が示唆されている。病原型の分類においては ExPEC に分類されなかったが、WWo055 は APEC に分類される可能性が示唆される。

WWo083 は、ほかの溶血性細菌とは異なり、G phylotype に分類された。G phylotype は ST117 のために 2018 年に新たに追加された phylotype である¹⁵⁴。G phylotype は B2 phylotype と F phylotype の間に等距離に位置している。G phylotype の大腸菌が報告された例として、中国の青海チベット高原に生息する野生の Marmot *Marmota himalayana* 116 頭の腸管内容物から単離された大腸菌が挙げられる¹⁵⁵。また、血清型 O143:H4 は牛の鼻腔から単離される大腸菌で最も多い血清型であるとの報告がある¹⁵⁶。ST117 は尿路感染症患者から単離される大腸菌では稀であるが、食肉製品から単離される大腸菌ではこれが最も優勢を占める¹⁵⁷。WWo083 は、各大腸菌病原型に特徴的な病原性遺伝子のほとんどを保有しないが、病原性遺伝子として *hlyE*、*ireA*、*lpfA*、*gad*、*traT*、*iss*、*terC*、*vat*、*hra*、*fyuA*、*chuA*、*irp2*、*ompT* を保有していた（図 14）。ごくまれに ST117 大腸菌による尿路感染症が報告されている¹⁵⁷ ことから、WWo083 は潜在的に尿路感染症を引き起こす可能性が示唆されるが、大腸菌 ST117 についての知見が不足しているため更なる調査が求められる。

4.2. 下水中の溶血性細菌種と病原性毒素

16S rRNA 遺伝子 V3-V4 領域の解析の結果、溶血活性と細菌種に関係性があることが示唆された。特に 24 時間の培養で明瞭な溶血斑を示した株は *Aeromonas* 属の細菌種と、*Escherichia coli* もしくは 16S rRNA 遺伝子 V3-V4 領域の配列が *Escherichia coli* に近い種であることが分かった。これらの細菌が多く単離された理由として、今回単離には 24 時間培養を行い、そこで溶血活性を示した細菌株を単離したため、増殖が速く 24 時間で明瞭な溶血斑を示す株が選択的に単離されたためであると考えられる。

16S rRNA 遺伝子解析では、*Escherichia coli* と *Escherichia fergusonii*、*Shigella flexneri*、*Shigella sonnei*、*Brenneria alni* を区別することができなかった。このうち *Brenneria alni* はグラム陰性の通性嫌気性細菌で、植物のハンノキの樹皮病を引き起こす¹⁵⁸⁻¹⁶⁰。ハンノキから単離されることがほとんどであり、羊の赤血球に対して反応性を示すことや、ヒトの糞便を主に含む下水から単離されたことから、血液寒天培地を用いて単離した株は *Brenneria alni* である可能性は低いと考えられる。全ゲノムシーケンスを行った株はいずれも *Escherichia coli* に同定されたため、そのほかの株の大部分も *Escherichia coli* であると推定される。しかし、*Escherichia coli*、*Escherichia fergusonii*、*Shigella flexneri*、*Shigella sonnei* はいずれも近縁の株であるため、今回の実験結果から区別することは難しい。

Aeromonas 属は、グラム陰性桿菌（0.3-1.0 × 1.0-3.5 μm）、オキシダーゼおよびカタラーゼ陽性で、硝酸塩を亜硝酸塩に分解する能力を持ち、ブドウ糖を発酵し、少数の例外を除いて O/129（2,4-Diamino-6,7-di-iso-propylpteridine phosphate）に耐性を示す¹⁶¹。*Aeromonas* 属には 36 種が報告されており、そのうち少なくとも 19 種がヒトに対する病原体と考えられている¹⁶¹。ヒトからの単離が多く報告されている種として *Aeromonas caviae*、*Aeromonas dhakensis*、*Aeromonas veronii*、*Aeromonas hydrophila* が挙げられる¹⁶¹。

Aeromonas 属はもともと免疫力の低下したヒトの日和見病原体と考えられていたが、世界各地で腸管および腸管外の疾患との関係が報告されるケースが増えており、宿主の免疫状態に関わらない新たなヒトの病原体として注目されている。*Aeromonas* 属が引き起こす感染症としては、胃腸炎、菌血症、敗血症、創傷感染症が挙げられる¹⁶¹。2004年にタイ南部で発生した津波では、外傷を負って皮膚・軟部組織感染症に罹患した患者305名から*Aeromonas* 属が最も多く単離された¹⁶²。幼児の溶血性尿毒症症候群患者からの単離や¹⁶³、骨折をした小児の発症した重度の壊死性筋膜炎患者からの単離も報告されている¹⁶⁴。

血液寒天プレートで単離された*Aeromonas* 属の16S rRNA 遺伝子解析をした結果、溶血活性と検出されたアレルパターンに関係性が見られた(図9)。*Aeromonas* 属は16S rRNA 遺伝子配列が非常に類似していることが指摘されている。種間類似度は96.7~100%で、16S rRNA 遺伝子に基づく種同定は適さないとされており¹⁶¹、代わりに average nucleotide identity (ANI) や *in silico* DNA-DNA ハイブリダイゼーションの使用が推奨されている¹⁶⁵。しかしアレルの構成に着目すると、16S rRNA 遺伝子のアレルには多くのバリエーションがあり、その保有パターンが複数存在した。アレル01を保有する株は溶血活性の大きい株が多かった。WWo140株はアレル01の検出割合が97%で、全ゲノムシーケンスの結果*Aeromonas hydrophila*であった。アレル01の検出割合の高いほかの株も*Aeromonas hydrophila*である可能性が示唆される。アレル01を保有する株の内いくつかの株はアレル06やアレル10を検出していた。これらのアレルは、アレル01とは系統的に近く、アレル01に由来する可能性が示唆される。WWo011はアレル00とアレル02を保有しており、それぞれの構成割合は45%と53%であった。このような株を従来の解析手法で種推定した場合、アレル02のみを用いて種推定を行うことになり、アレル02のみを保有する株と区別されることはなかった。このことは、16S rRNA 遺伝子のアレル構成割合に基づく解析が、*Aeromonas* 属に対して有用であることを示唆する。*Aeromonas* 属の種分類に16S rRNA 遺伝子アレルの検出割合を用いることができる可能性があり、今後の調査が望まれる。アレル00を保有するそのほかの株は溶血活性が小さいが、WWo011とWWo011のアレル構成と似ているWWo012は、溶血斑幅がそれぞれ1.0 mmと0.89 mmとやや大きい値を示したことは興味深い。全ゲノムシーケンスに基づく種同定の結果、WWo011とWWo148は*Aeromonas caviae*に同定されたことから、アレル00を保有するそのほかの株も*Aeromonas caviae*である可能性が示唆されるが、WWo011とWWo012は遺伝的背景がやや異なるかもしれない。

全ゲノムシーケンスを行った*Aeromonas* 属3株は、いずれも複数の溶血毒遺伝子を保有していた。ヘモリシンHlyAをコードする*hlyA* 遺伝子は、全ゲノムシーケンスを行った*Aeromonas* 属3株すべてが保有する溶血毒遺伝子であった。また、*Aeromonas hydrophila*のWWo140は、溶血毒エロリジン産生に関わる*aerA* 遺伝子を保有していた。*Aeromonas* 属では、*hlyA* 遺伝子もしくは*aerA* 遺伝子を持っている株は18時間の培養でβ溶血を示すことが報告されている¹⁶⁶。ヘモリシンHlyAは*Aeromonas hydrophila*等によって産生される溶血毒で、*Vibrio cholerae*の産生するHlyAに類似した溶血毒である¹⁶⁷。エロリジンは463アミノ酸残基からなり、分子量は53.8 kDaである¹⁶⁸。エロリジンの構造はよく知られており、7量体で膜貫通型のβバレルを形成し、これが細胞膜に挿入することで細胞膜に穴をあける¹⁶⁹。マウスを用いたエロモナスの経口投与モデルでは、*aerA*と*hlyA*の両方の不活化により毒性が低下する¹⁶⁸ことから、*aerA*と*hlyA*はエロモナスの病原性に大きく関わっていると考えられる。*Aeromonas hydrophila*の強毒株は*hlyA*と*aerA*の両方を保有しており¹⁶⁶、WWo140も強毒性株である可能性が示唆される。

ヘモリシン III と熱安定性毒素もエロモナス株 3 株すべてで保有が確認された。ヘモリシン III は、*Aeromonas hydrophila* ML09-119 株のゲノム上 4207405-4208040 に位置する 636 bp からなる ORF をリファレンスとする配列で、211 アミノ酸からなるタンパク質が翻訳されると考えられる。レビューされているタンパク質の中では *Escherichia coli* の保有する内膜タンパク YpfA や *Bacillus cereus* の保有するヘモリシン III との相同性が最も高く、BLAST による一致率 (pident) はそれぞれ 67% と 49% であった (data not shown)。熱安定性ヘモリシン (TH) は、*Aeromonas hydrophila* ML09-119 株のゲノム上 3924899-3925543 の 645bp からなるタンパク質が翻訳されると考えられる。*Aeromonas* におけるヘモリシン III と熱安定性ヘモリシン (TH) の働きはまだよく知られておらず、今後の調査が望まれる。

WWo140 で確認された RTX は 6 つの遺伝子 (*rtxA*CHBDE) からなるオペロンにコードされており、これは *Vibrio* 属の RTX によく似た構造である¹⁷⁰。*Vibrio anguillarum* において、*rtxA* は外毒素 RtxA、*rtxC* は RtxA の活性化酵素、*rtxH* は保存された推定タンパク、*rtxBDE* は ABC トランスポーターをそれぞれコードする¹⁷¹。*Aeromonas* では、RtxA は宿主細胞と接触した際に発現量が増加し、RtxACD は宿主細胞に対してアクチン細胞骨格の破壊とアポトーシスを誘導する¹⁷⁰。*Vibrio vulnificus* のマウス感染モデルにおいて、RTX を欠損させた細菌株は宿主のマクロファージにより容易に排除されるようになり病原性が低下する¹⁷²。これらのことから、RTX は宿主の免疫細胞による貪食から回避するのに役立つ可能性が示唆されている。

VFAnalyzer による病原性毒素の予測では、*Streptococcus lutetiensis* に同定された WWo197 で *Enterococcus* の持つ細胞溶解毒をコードする *cylR2* 遺伝子と高い相同性を示す遺伝子を保有していた。*cylR2* は、*Enterococcus faecalis* の染色体上の病原性アイランドに存在するサイトリシンオペロンにコードされる遺伝子の一つである¹⁷³。サイトリシンは *cylR1* と *cylR2* によって発現が制御される。*cylR2* は毒素そのものではなく、発現を制御する因子であるため、必ずしもこれがあることによって病原性を発揮するとは限らない。

WWo190、WWo197、WWo204、WWo207、WWo210、WWo221 はいずれも不明瞭な溶血斑を示した株で、prokka を用いた解析ではヘモリシン A をコードする *thyA* 遺伝子を保有していた。*thyA* 遺伝子は、様々な病原性細菌で発現しており、コロニー形成、宿主の免疫応答への影響、溶血、抗生物質耐性等さまざまな役割を担っている¹⁷⁴。以前はサイトトキシシン/ヘモリシンファミリーとのアミノ酸配列の類似性からヘモリシンとして分類されていたが、多くの研究から、ヘモリシン A には RNA メチルトランスフェラーゼの指標となるモチーフが含まれ、16S rRNA ヌクレオチド C1409 及び 23S rRNA ヌクレオチド C1929 をメチル化することが分かっている¹⁷⁵。一方で、*thyA* にコードされるタンパクは溶血酵素としても機能することが示唆されている。*Helicobacter pylori* が保有する *thyA* 遺伝子に変異を加えることで血液寒天プレート上での溶血活性が小さくなることが報告されている¹⁷⁶。また、結核菌 (*Mycobacterium tuberculosis*) の保有する *thyA* 遺伝子を非溶血性大腸菌に組み込むことで大腸菌が溶血活性を持つようになる¹⁷⁷。*thyA* 遺伝子をコードする WWo190、WWo197、WWo204、WWo207、WWo210、WWo221 の溶血活性は、*thyA* 遺伝子による可能性が示唆される。

Streptococcus parasuis は、*Streptococcus suis* の血清型 20、22、26 が再分類された種である^{178,179}。血液寒天プレートを用いて単離された溶血性細菌の中では WWo190 がこれに同定された。*Streptococcus parasuis* はヒトや豚に炎症性疾患や侵襲性疾患を引き起こす人獣共通感染症の病原体として知られ、何らかの疾患をもつ豚や牛から単離される¹⁷⁹。健康な

豚の唾液中にも含まれるため、菌自体の病原性は低いと考えられている¹⁷⁸。しかしながら稀に菌血症患者からも単離されることがあり、免疫力の低下したヒトの健康を脅かす可能性がある¹⁸⁰。

Streptococcus lutetiensis は、以前は *Streptococcus infantarius* に分類されていた種で¹⁸¹、*Streptococcus bovis* / *Streptococcus equinus* complex (SBSEC) と呼ばれるグループに属する。血液寒天プレートを用いて単離された溶血性細菌の中で WWo197 がこれに分類された。SBSEC は、動物やヒトの消化管に生息する複数の種で構成されており、乳製品由来の発酵食品から頻りに分離されるものや、人獣共通感染症の原因になるものが含まれる^{182,183}。*Streptococcus lutetiensis* は菌血症患者の血液から単離されたとの報告や¹⁸⁴、乳房炎にかかった乳牛から単離された報告がある¹⁸⁵。*Streptococcus lutetiensis* は推定溶血毒 *cylZ* 遺伝子を保有することがある¹⁸⁶。また、SBSEC に属する *Streptococcus gallolyticus* や *Streptococcus pasteurianus*、*Streptococcus macedonicus*、*Streptococcus infantarius* はヘモリシン TLY III やヘモリシン A ファミリータンパクを保有することがある¹⁸³。これらの遺伝子が溶血活性にどのような影響を与えるかはまだ明らかではない。WWo197 は、VFAnalyzer の解析の結果 *cylZ* 遺伝子の保有は確認されなかった。ヘモリシン TLY III 及びヘモリシン A ファミリータンパクについては VFAnalyzer で使用したデータベースに含まれていなかったため、本研究では保有を確認していない。

WWo204 の同定された *Leuconostoc lactis* は、カタラーゼおよびオキシダーゼ陰性のグラム陽性の通性嫌気性球菌で、野菜、豆類、果物、肉等に含まれ、食品業界では乳製品、ワイン、糖類の製造に使用されている¹⁸⁷。かつてはヒトに対する病原性はないと考えられていたが、近年 *Leuconostoc lactis* が原因と思われる脳室炎¹⁸⁸、敗血症^{187,189} が稀に報告されている。溶血性を調べた研究ではキムチから単離された *Leuconostoc lactis* の溶血活性を調べたものがあるが、溶血活性を持つ株は認められていない¹⁹⁰。一方、敗血症患者から単離された *Leuconostoc lactis* は α 溶血を示すことがあり¹⁹¹、溶血活性と病原性に何らかの関係があるかもしれない。

WWo207 の同定された *Lactococcus taiwanensis* は、2013 年に新種として提案された種で、当初は *Lactococcus lactis* に分類されていた¹⁹²。*Lactococcus taiwanensis* は台湾の伝統的な発酵料理 pobuzihi (発酵した *cummingcordia*) から単離された株が新種として提案された。現在までで、*Lactococcus taiwanensis* の病原性や溶血活性に関する報告はない。*Lactococcus taiwanensis* と同定された株が溶血活性を持つことは初めて明らかになった。

WWo210 の同定された *Ligilactobacillus salivarius* は、ラマのミルクから単離されたプロバイオティクス効果のある乳酸菌がこの種であることが報告されている¹⁹³。以前の報告では *ligilactobacillus salivarius* は溶血活性を示しておらず¹⁹⁴、今回単離された株はプロバイオティクス効果のある以前のものとは異なる表現型を示す可能性が示唆される。*Ligilactobacillus salivarius* と同定された株が溶血活性を持つことは初めて明らかになった。

WWo220 は、唯一溶血毒に関わる遺伝子が検出されなかった。同定種の *Enterococcus casseliflavus* は、免疫不全患者や慢性疾患患者に重篤な感染症を引き起こす可能性がある^{195,196}。*Enterococcus* 属の保有する溶血毒としては *Enterococcus faecalis* の保有するサイトリシン等が挙げられる¹⁹⁷。*Enterococcus casseliflavus* の保有する病原性遺伝子や溶血毒についてはあまり多くの報告はない。WWo220 は、血液寒天プレートタイプ III で培養したときに溶血活性が確認できた。WWo220 の保有する溶血毒については、今後の解析が望まれる。

WWo221 の同定された *Enterococcus lemanii* は、グラム陽性球菌の腸球菌 (*Enterococci*) で、以前の報告では溶血活性を示さない¹⁹⁸。今回、WWo221 は血液寒天プレートタイプ III で培養したときにわずかに溶血活性を持つことが確認された。*Enterococcus lemanii* が保有する溶血毒はまだ知られていないが、*prokka* によって保有が確認されたヘモリシン A によるものかもしれない。*Enterococcus lemanii* が溶血活性を示すことは初めて明らかになった。

4.3. 溶血性細菌の溶血活性決定因子

16S rRNA 遺伝子 V3-V4 領域の解析の結果、溶血活性と細菌種に関係性があることが示唆された。特に 24 時間の培養で明瞭な溶血斑を示した株は *Aeromonas* 属の細菌種と、*Escherichia coli* もしくは 16S rRNA 遺伝子 V3-V4 領域の配列が *Escherichia coli* に近い種であることが分かった。これらの細菌が多く単離された理由として、今回単離には 24 時間培養を行い、そこで溶血活性を示した細菌株を単離したため、増殖が早く 24 時間で明瞭な溶血斑を示す株が選択的に単離されたためであると考えられる。コロニー周辺が変色したのは *Citrobacter* 属で、残りの種に推定された株はすべて不明瞭な溶血斑を示した。

Aeromonas 属の溶血活性は、16S rRNA V3-V4 領域のアレルの構成で説明され、アレル 01 を保有する株は溶血活性が大きい傾向にあり、それらに比べてアレル 00、アレル 17、アレル 02 のいずれかを保有するエロモナスは溶血活性が小さい傾向にあった。16S rRNA 遺伝子は種で高く保存されることを考慮すると、*Aeromonas* 属の溶血活性は種によって異なることが示唆される。全ゲノムシーケンスに基づき種を同定した結果、アレル 01 を保有して溶血活性の比較的大きい WWo140 は *Aeromonas hydrophila* と同定され、エロリジン、ヘモリシン HlyA、細胞外ヘモリシン AHH1、熱安定性ヘモリシン、ヘモリシン III を保有していた。一方、*Aeromonas caviae* と同定された WWo011 と WWo148 はエロリジン、細胞外ヘモリシン AHH1 を保有していなかった。*Aeromonas hydrophila* と *Aeromonas caviae* で、*hlyA* と *aerA* の両方を保有する割合はそれぞれ 75.4%と 29.4%と *Aeromonas hydrophila* の方が高いことが報告されている¹⁶⁶。種による溶血活性の差は、保有する溶血毒の種類によって説明される可能性があり、更なる調査が望まれる。

Escherichia coli の溶血活性は、16S rRNA V3-V4 領域のアレルの構成では説明されなかった。16S rRNA 遺伝子 V3-V4 領域の解析で *Escherichia coli* もしくは 16S rRNA 遺伝子 V3-V4 領域の配列が *Escherichia coli* に近い種と推定された株のすべてがアレル 100 を保有しており、アレル 100 を保有する株の中で溶血活性の大きい株と小さい株が存在した。全ゲノムシーケンスを行った結果、溶血毒 α ヘモリシン産生に関わる *hlyCABD* オペロンをしない株で溶血活性が小さいことが分かった。一方で、*hlyCABD* オペロンを保有する大腸菌では株間の溶血活性にばらつきが見られた。kSNP3.0 を用いた系統樹、MLST、FimH、血清型、Phylotype でそれぞれ分類したところ、シーケンスタイプが ST127 の株で溶血活性が大きい傾向があったが、ST95 の株をはじめほかのシーケンスタイプでは溶血活性が異なっており、溶血活性はシーケンスタイプだけで決まるわけではなかった。 α ヘモリシン発現量によって溶血活性が異なると考え、*hlyA* mRNA 量と溶血活性の関係を調べたところ、KKo007 株と WWo027 は *hlyA* mRNA 量発現量が高く溶血活性も大きかった (図 16)。しかし、KKa004、WWo147、WWo156 の溶血活性は大きかったが *hlyA* mRNA 発現量は、溶血活性の小さかった WWo070 や WWo073、WWo014、KOr014 と差がほとんど見られず、*hlyA* の発現量だけで溶血活性が説明できないことが示唆された。 α ヘモリシンタンパク分

泌量を SDS-PAGE で確認しようとしたが、分泌量が少なく確認することはできなかった。 α ヘモリシタンパクの測定には、より感度の高い手法を選択する必要がある。

α ヘモリシン産生に関わる遺伝子の SNP 解析を行ったところ、溶血活性に関わる遺伝子にいくつかミスセンス変異が見つかった。 α ヘモリシンは様々な遺伝子が関わって溶血活性を示す。まず、 α ヘモリシタンパク前駆体 HlyA は *hlyCABD* オペロンの *hlyA* 遺伝子にコードされている。*hlyCABD* オペロンには *hlyC*、*hlyA*、*hlyB*、*hlyD* の 4 遺伝子がこの順でコードされ、SNP 解析のリファレンスとして用いた S65EC 株の場合、*hlyC* 遺伝子の 1,616 bp 上流から転写が開始される⁴⁸。RNA ポリメラーゼ複合体は、転写開始点の約 15 bp 上流の -15 領域と、約 35 bp 上流の -35 領域を認識して転写を開始する。転写開始点と *hlyCABD* オペロンの間には、operon polarity suppressor (*ops*) エlement と呼ばれる共通配列 (5'-GGCGGTAG-3') を含む、JUMP Start 配列が存在する⁴⁸。RNA ポリメラーゼは JUMP Start 配列が作る RNA ヘアピン構造により *ops* 要素で転写伸長を一時停止するが、*ops* 要素は *rfaH* 遺伝子にコードされる RfaH タンパクをリクルートする。RfaH タンパクは *ops* 配列上で RNA ポリメラーゼ伸長複合体と結合し、RNA ポリメラーゼのバックトラックの抑制と RNA ヘアピン構造安定化を抑制することで、RNA ポリメラーゼの転写伸長を促進する¹⁹⁹⁻²⁰¹。

hlyCABD オペロンから転写、翻訳された α ヘモリシン前駆体 HlyA は、*hlyC* 遺伝子にコードされる HlyC タンパクにより翻訳後修飾を受け、2 か所のリジン (Lys-563、Lys-689) がアシル化される。このアシル化は、HlyA が宿主の細胞膜内でオリゴマー化して溶血活性を示すのに必要である^{202,203}。アシル化された HlyA は、*hlyB* にコードされる ABC トランスポーター HlyB、*hlyD* にコードされる膜融合/チャネルタンパク HlyD、*tolC* にコードされる外膜タンパク TolC の 3 種類のタンパクによって構成されるタイプ I 分泌システムによって細胞外に放出される²⁰⁴。

分泌された α ヘモリシタンパクはカルシウムイオン存在下で赤血球と結合するようになる²⁰⁵。 α ヘモリシンはリピートイントキシン (RTX) とよばれる毒素ファミリーの一種で、X-(L/I/F)-X-G-G-X-G-(N/D)-D (X は任意の残基) というコンセンサス配列を含む RTX モチーフの繰り返し構造 (RTX リピート) を持つ²⁰⁴。RTX リピートはパラレル β ロールと呼ばれる構造を形成し、カルシウムイオンと結合する²⁰⁶。 α ヘモリシンは RTX リピートを 11 個含み、カルシウムイオンと結合することで溶血活性を示す²⁰⁷。さらに、分泌された α ヘモリシタンパクは大腸菌の産生するリポ多糖 (LPS) と複合体を形成することで、毒素の安定性を高め、また α ヘモリシタンパクの自己凝集能を抑制する^{208,209}。LPS は負電荷を帯びているためカルシウムを貯蔵し、溶血活性に必要なカルシウムの貯蔵庫としても働いていると考えられている²⁰⁸。

ST95 に分類された WWo073、WWo070、KOr014、WWo108、WWo014、WWo005、KKa004 の内、KOr014 はやや溶血活性が小さく、*hlyA* 遺伝子に c.841G>T (p.Gly281Cys) の SNP が見つかった。*hlyA* 遺伝子は α ヘモリシン前駆体 HlyA をコードする遺伝子であることから、KOr014 は c.841G>T (p.Gly281Cys) の SNP が α ヘモリシタンパクの溶血活性に直接影響を及ぼした可能性が示唆される。ST95 に分類された株の内、WWo108 はさらに溶血活性が小さかった。WWo108 は、*rfaH* 遺伝子で c.50C>G (p.Ala17Gly) の SNP が確認された。転写伸長に関わる RfaH タンパクをコードする *rfaH* 遺伝子の中で、*ops* 要素を認識する Lys-10、Arg-16、His-20、Thr-72、Arg-73 は RfaH の活性に重要なアミノ酸で、これらのアミノ酸が置換されると RfaH タンパクの *ops* 要素へのリクルートは阻害される²¹⁰。

c.50C>G (p.Ala17Gly) の SNP は Arg-16 の隣に位置し、この位置のアミノ酸置換により RfaH タンパクの構造変化が起きる可能性がある。これにより RfaH の ops 要素認識活性が低下し、転写伸長作用に影響を及ぼしているかもしれない。逆に、ST95 の中で KKa004 はほかの株に比べて溶血活性が大きかった。KKa004 は、*rfaH* で c.95T>C (p.Met32Thr) の SNP と *rne* で c.705A>C (p.Lys235Asn) の SNP が確認された。KKa004 の溶血活性が大きいことから、WWo108 とは逆に、c.95T>C (p.Met32Thr) の SNP によって RfaH の活性が大きくなり、*hlyCABD* オペロンの転写が活性化している可能性がある。*rne* と α ヘモリシンの関係については、*rne* 遺伝子にトランスポゾン挿入させたところ溶血活性が小さくなることが報告されているが⁴⁸、*rne* 遺伝子が溶血活性に影響を与える作用機序についてはまだ明らかではない。*rfaH* の SNPs は KKa004 と WWo108 で確認されたが、液体培地で2時間培養したときの *hlyA* mRNA 量は他の ST95 と大きな差は見られなかった。今回見つかった *rfaH* の SNPs が転写にどのような影響を及ぼすか、今後さらなる解析が望まれる。

ST127 に分類された株はいずれも溶血活性が大きく、ST95 とは *rne*、*waaG*、*tolC* で SNPs のパターンが異なっていた。*waaG* は LPS の合成に関わるタンパクをコードする遺伝子で、*waaG* を欠損すると溶血活性が低下する⁴⁸。 α ヘモリシンは LPS 存在下で安定化し、自己凝集能が抑制され、カルシウムイオンの貯蔵庫としても働く^{208,209}。また、*tolC* のコードする TolC タンパクは α ヘモリシンの分泌を行うタイプ I 分泌システムを構成する²⁰⁴。SNPs によって生じる LPS 合成活性の差、あるいは TolC により構成されるタイプ I 分泌システムの差により、ST127 は ST95 と比べて溶血活性が高い可能性が示唆される。

4.4. 血液寒天プレートを選択培地としての有用性と課題

血液寒天プレートを用いて下水サンプルを培養したところ、明瞭な溶血斑を示す株、不明瞭な溶血斑を示す株、変色した溶血斑を示す株、溶血斑を示さない株が存在した。できるだけ多くの種類の溶血性細菌を単離するため、目視で溶血活性の有無を判断するのが難しい不明瞭な溶血斑を示すコロニーからも単離を行った。そのため、単離後に溶血活性を再測定すると一部の株は溶血活性を示さなかった。これは、単離時に溶血活性を示したが条件が変わり溶血活性を示さなくなった株と、初めから溶血活性を示さなかった株の両方が含まれると考えられる。

溶血活性の小さな株に対しては、使用する培地を工夫することで溶血活性の検出感度が上がった。本研究では3種類の血液寒天プレートを用いた(表1)。溶血活性の大きい株についてはいずれの血液寒天プレートでも溶血斑を確認することができたが、WWo108のように溶血活性の小さい株の場合、血液寒天プレートタイプ I、II では溶血斑を確認することは難しく、血液寒天プレートタイプ III で、培養時間を長くすることで溶血斑を確認することができた。感度の高い血液寒天プレート III で溶血活性を測定することでこれまで溶血活性を持つことが知られていない種でも溶血斑を確認することができた。WWo207、WWo210、WWo221 がそれぞれ同定された、*Lactococcus taiwanensis*、*Ligilactobacillus salivarius*、*Enterococcus lemanii* はこれまでに溶血活性を持つことが報告されてこなかった種であったが、今回血液寒天プレートタイプ III を用いて溶血斑を確認することで溶血活性を持つことが明らかとなった。また、*Streptococcus parasuis* や *Leuconostoc lactis*、*Enterococcus casseliflavus* は稀に菌血症等の感染症を引き起こすことが知られている。血液寒天プレートタイプ III は、わずかな溶血活性を示すような弱毒性の溶血性細菌の溶血活性を確認できることから、免疫不全患者などに対する潜在的な病原性細菌を検出できる可能性が示唆された。

血液寒天培地を選択培地として用いるのにいくつか課題も残る。一つは培養条件によって溶血活性が異なる点である。本研究ではLB培地に寒天と羊血液を加えて培地を作成した。しかし、培養に用いられる培地は多種多様かつ血液にもいくつか種類がある。Robertsonら(2006)²¹¹は*Bacteroides fragilis*の産生する溶血毒の溶血活性について調べたが、溶血活性が培地成分によって異なることや、ヒツジ、ヒト、ウマといった赤血球の由来によって溶血活性を示す場合と示さない場合があること、そして好気性では溶血活性を持たないが嫌気性では溶血活性を持つ場合があることを報告している。条件によって検出される細菌の種類や数が変わると考えられることから、培養条件と検出される病原性細菌の関係について明らかにする必要がある。もう一つは感度の低さである。今回、血液寒天培地に下水を混和して固めたものを用いて培養した。この際、溶血活性を持たないものと持つものを区別することが難しく、再度溶血活性を確認した際に全く溶血活性を示さない株が多く存在した。培養時間を長くすることで溶血斑は大きくなるが、株によって成長速度に差があることや長時間の培養で培地中の血液が黒くなり溶血斑を確認しづらくなるといった問題が生じる。血液寒天プレートタイプIIIは、感度は高いが単離時に使用することは難しいため、感度のよい単離法について改めて検討する必要がある。

今回使用した血液寒天プレートで、明瞭な溶血斑を示すコロニーから腸管外病原性大腸菌と病原性の*Aeromonas*属の細菌を単離でき、不明瞭な溶血斑を示すコロニーからは稀に病原性を発揮する病原性細菌をそれぞれ単離できた。今回初めて明らかになった溶血性細菌種にはこれまで病原性細菌として考えられてこなかった種が含まれ、病原性等に関して更なる調査が望まれる。既知の病原性細菌を単離できるだけでなく、稀な病原性細菌やこれまで病原性細菌として考えられてこなかったが溶血活性を示す細菌を単離できることから、血液寒天プレートは未知の病原性細菌の単離法として期待される。

5. 結論

血液寒天培地を用いて都市下水から網羅的に病原性細菌を単離することで、主に大腸菌と *Aeromonas* 属が検出された。これらの株の多くは 24 時間の培養で明瞭な溶血斑を示した。溶血性大腸菌の多くは溶血毒 α ヘモリシン産生に関わる *hlyCABD* 遺伝子を保有しており、いずれも B2 phylotype に属する腸管外病原性大腸菌 (ExPEC) に分類され、尿路感染症や敗血症、結腸がんとの関わりが示唆された。これらの株は ST95、FimH18、H7 血清型のグループ、ST73、H1 血清型のグループ、ST127、O6:H31 血清型のグループに大別された。SNP 解析ソフト MMViewer を開発し SNP 解析を行ったところ、*hlyA* 遺伝子の SNP : c.841G>T (p.Gly281Cys) や、*rfaH* 遺伝子の SNP : c50C>G (p.Alal7Gly) 等の SNPs が溶血活性に影響を与えている可能性が示唆された。*hlyCABD* 遺伝子を保有しない溶血性大腸菌は、溶血毒 HlyF をコードする *hlyF* 遺伝子を保有しており、鶏病原性大腸菌である可能性が示唆された。*Aeromonas* 属は大きな溶血斑を示す株と小さな溶血斑を示す株があり、この違いは 16S rRNA 遺伝子アレルのパターンにより、明確に区別することができた。溶血活性の大きな株は *Aeromonas hydrophila* であり、強毒性であることが示唆された。不明瞭な溶血斑を示した細菌は、敗血症等の稀な感染症の報告のある *Streptococcus parasuis*、*Streptococcus lutetiensis*、*Leuconostoc lactis*、*Enterococcus casseliflavus* などが同定された。また、これまで溶血活性の報告がなかった *Ligilactobacillus salivarius*、*Lactococcus taiwanensis*、*Enterococcus lemanii* なども同定された。

参考文献

- 1 Kaufmann SHE. Robert Koch, the Nobel Prize, and the Ongoing Threat of Tuberculosis. *N Engl J Med* 2005; **353**: 2423–2426.
- 2 Larsen DA, Green H, Collins MB, Kmush BL. Wastewater monitoring, surveillance and epidemiology: a review of terminology for a common understanding. *FEMS Microbes* 2021; **2**: xtab011.
- 3 Melnick JL. Poliomyelitis virus in urban sewage in epidemic and in nonepidemic times. *American Journal of Epidemiology* 1947; **45**: 240–253.
- 4 Brouwer AF, Eisenberg JNS, Pomeroy CD, Shulman LM, Hindiyeh M, Manor Y *et al*. Epidemiology of the silent polio outbreak in Rahat, Israel, based on modeling of environmental surveillance data. *Proc Natl Acad Sci USA* 2018; **115**: E10625–E10633.
- 5 Sinclair RG, Choi CY, Riley MR, Gerba CP. Pathogen Surveillance Through Monitoring of Sewer Systems. *Adv Appl Microbiol* 2008; **65**: 249–269.
- 6 Zerva I, Remmas N, Kagalou I, Melidis P, Ariantsi M, Sylaios G *et al*. Effect of Chlorination on Microbiological Quality of Effluent of a Full-Scale Wastewater Treatment Plant. *Life* 2021; **11**: 68.
- 7 Li X. Metagenomic screening of microbiomes identifies pathogen-enriched environments. *Environ Sci Eur* 2019; **31**: 37.
- 8 Thompson AA, Matamale L, Kharidza SD. Impact of Climate Change on Children’s Health in Limpopo Province, South Africa. *Int J Environ Res Public Health* 2012; **9**: 831–854.
- 9 Bidle KD, Lee S, Marchant DR, Falkowski PG. Fossil genes and microbes in the oldest ice on Earth. *Proc Natl Acad Sci U S A* 2007; **104**: 13455–13460.
- 10 Katayama T, Tanaka M, Moriizumi J, Nakamura T, Brouchkov A, Douglas TA *et al*. Phylogenetic Analysis of Bacteria Preserved in a Permafrost Ice Wedge for 25,000 Years. *Appl Environ Microbiol* 2007; **73**: 2360–2363.
- 11 Timofeev V, Bahtejeva I, Mironova R, Titareva G, Lev I, Christiany D *et al*. Insights from *Bacillus anthracis* strains isolated from permafrost in the tundra zone of Russia. *PLoS One* 2019; **14**: e0209140.
- 12 Liskova EA, Egorova IY, Selyaninov YO, Razheva IV, Gladkova NA, Toropova NN *et al*. Reindeer Anthrax in the Russian Arctic, 2016: Climatic Determinants of the Outbreak and Vaccination Effectiveness. *Frontiers in Veterinary Science* 2021; **8**: 486.
- 13 Toze S. PCR and the detection of microbial pathogens in water and wastewater. *Water Research* 1999; **33**: 3545–3556.
- 14 Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M *et al*. Diversity of the Human Intestinal Microbial Flora. *Science* 2005; **308**: 1635–1638.
- 15 Lagier J-C, Armougom F, Million M, Hugon P, Pagnier I, Robert C *et al*. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clinical Microbiology and Infection* 2012; **18**: 1185–1193.

- 16 Lagier J-C, Dubourg G, Million M, Cadoret F, Bilen M, Fenollar F *et al.* Culturing the human microbiota and culturomics. *Nat Rev Microbiol* 2018; **16**: 540–550.
- 17 Bonnet M, Lagier JC, Raoult D, Khelaifia S. Bacterial culture through selective and non-selective conditions: the evolution of culture media in clinical microbiology. *New Microbes and New Infections* 2020; **34**: 100622.
- 18 Asmar S, Drancourt M. Chlorhexidine decontamination of sputum for culturing *Mycobacterium tuberculosis*. *BMC Microbiol* 2015; **15**: 155.
- 19 Riley TV, Brazier JS, Hassan H, Williams K, Phillips KD. Comparison of alcohol shock enrichment and selective enrichment for the isolation of *Clostridium difficile*. *Epidemiol Infect* 1987; **99**: 355–359.
- 20 Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD *et al.* Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. *Nature* 2016; **533**: 543–546.
- 21 Adamiak J, Otlewska A, Gutarowska B, Pietrzak A. Halophilic microorganisms in deteriorated historic buildings: insights into their characteristics. *Acta Biochim Pol* 2016; **63**: 335–341.
- 22 Subramanyam B, Sivaramakrishnan GN, Dusthacker A, Nagamiah S, Kumar V. Phage lysin as a substitute for antibiotics to detect *Mycobacterium tuberculosis* from sputum samples with the BACTEC MGIT 960 system. *Clinical Microbiology and Infection* 2012; **18**: 497–501.
- 23 Sum R, Swaminathan M, Rastogi SK, Piloto O, Cheong I. Beta-Hemolytic Bacteria Selectively Trigger Liposome Lysis, Enabling Rapid and Accurate Pathogen Detection. *ACS Sens* 2017; **2**: 1441–1451.
- 24 May AK, Gleason TG, Sawyer RG, Pruett TL. Contribution of *Escherichia coli* Alpha-Hemolysin to Bacterial Virulence and to Intraperitoneal Alterations in Peritonitis. *Infect Immun* 2000; **68**: 176–183.
- 25 Wang C, Li Q, Lv J, Sun X, Cao Y, Yu K *et al.* Alpha-hemolysin of uropathogenic *Escherichia coli* induces GM-CSF-mediated acute kidney injury. *Mucosal Immunol* 2020; **13**: 22–33.
- 26 Jin Y, Tang S, Li W, Ng SC, Chan MWY, Sung JJY *et al.* Hemolytic *E. coli* Promotes Colonic Tumorigenesis in Females. *Cancer Research* 2016; **76**: 2891–2900.
- 27 Bückner R, Schulz E, Günzel D, Bojarski C, Lee I-FM, John LJ *et al.* α -Haemolysin of *Escherichia coli* in IBD: a potentiator of inflammatory activity in the colon. *Gut* 2014; **63**: 1893–1901.
- 28 Kanwal S, Vaitla P. *Streptococcus Pyogenes*. In: *StatPearls*. StatPearls Publishing: Treasure Island (FL), 2021 <http://www.ncbi.nlm.nih.gov/books/NBK554528/> (accessed 3 Dec2021).
- 29 Zhu L, Olsen RJ, Lee JD, Porter AR, DeLeo FR, Musser JM. Contribution of Secreted NADase and Streptolysin O to the Pathogenesis of Epidemic Serotype M1 *Streptococcus pyogenes* Infections. *The American Journal of Pathology* 2017; **187**: 605–613.
- 30 McDowell RH, Sands EM, Friedman H. *Bacillus Cereus*. StatPearls Publishing, 2021 <https://www.ncbi.nlm.nih.gov/books/NBK459121/> (accessed 4 Dec2021).

- 31 Pomerantsev AP, Kalnin KV, Osorio M, Leppla SH. Phosphatidylcholine-Specific Phospholipase C and Sphingomyelinase Activities in Bacteria of the *Bacillus cereus* Group. *Infect Immun* 2003; **71**: 6591–6606.
- 32 Taylor TA, Unakal CG. *Staphylococcus Aureus*. StatPearls Publishing, 2021 <https://www.ncbi.nlm.nih.gov/books/NBK441868/> (accessed 4 Dec2021).
- 33 Vandenesch F, Lina G, Henry T. Staphylococcus aureus Hemolysins, bi-component Leukocidins, and Cytolytic Peptides: A Redundant Arsenal of Membrane-Damaging Virulence Factors? *Front Cell Inf Microbio* 2012; **2**: 12.
- 34 Sakaue M, Ota K, Nakamura E, Nitta M, Oka M, Oishi Y *et al*. Type A fulminant Clostridium perfringens sepsis indicated RBC/Hb discrepancy; a case report. *BMC Infect Dis* 2019; **19**: 719.
- 35 Kurasawa M, Nishikido T, Koike J, Tominaga S, Tamemoto H. Gas-forming liver abscess associated with rapid hemolysis in a diabetic patient. *WJD* 2014; **5**: 224–229.
- 36 Allerberger F. *Listeria* : growth, phenotypic differentiation and molecular microbiology. *FEMS Immunology & Medical Microbiology* 2003; **35**: 183–189.
- 37 Hsiao A, Zhu J. Pathogenicity and virulence regulation of *Vibrio cholerae* at the interface of host-gut microbiome interactions. *Virulence* 2020; **11**: 1582–1599.
- 38 Fan Y, Li Z, Li Z, Li X, Sun H, Li J *et al*. Nonhemolysis of epidemic El Tor biotype strains of *Vibrio cholerae* is related to multiple functional deficiencies of hemolysin A. *Gut Pathog* 2019; **11**: 38.
- 39 Broberg CA, Calder TJ, Orth K. *Vibrio parahaemolyticus* cell biology and pathogenicity determinants. *Microbes and Infection* 2011; **13**: 992–1001.
- 40 Mogrovejo DC, Perini L, Gostinčar C, Sepčić K, Turk M, Ambrožič-Avguštin J *et al*. Prevalence of Antimicrobial Resistance and Hemolytic Phenotypes in Culturable Arctic Bacteria. *Front Microbiol* 2020; **11**: 570.
- 41 Ravi RK, Walton K, Khosroheidari M. MiSeq: A Next Generation Sequencing Platform for Genomic Analysis. In: DiStefano JK (ed). *Disease Gene Identification*. Springer New York: New York, NY, 2018, pp 223–232.
- 42 Ng PC, Kirkness EF. Whole Genome Sequencing. In: Barnes MR, Breen G (eds). *Genetic Variation*. Humana Press: Totowa, NJ, 2010, pp 215–226.
- 43 Malberg Tetzschner AM, Johnson JR, Johnston BD, Lund O, Scheutz F. *In Silico* Genotyping of *Escherichia coli* Isolates for Extraintestinal Virulence Genes by Use of Whole-Genome Sequencing Data. *J Clin Microbiol* 2020; **58**: e01269-20.
- 44 Clausen PTL, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics* 2018; **19**: 307.
- 45 Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* 2009; **33**: 376–393.

- 46 Rasband WS. ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA. <https://imagej.nih.gov/ij/> (accessed 15 Dec2021).
- 47 Taneike I, Zhang H-M, Wakisaka-Saito N, Yamamoto T. Enterohemolysin operon of Shiga toxin-producing *Escherichia coli*: a virulence function of inflammatory cytokine production from human monocytes. *FEBS Lett* 2002; **524**: 219–224.
- 48 Nhu NTK, Phan M-D, Forde BM, Murthy AMV, Peters KM, Day CJ *et al.* Complex Multilevel Control of Hemolysin Production by Uropathogenic *Escherichia coli*. *mBio* 2019; **10**: e02248-19.
- 49 Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 2013; **41**: e1.
- 50 Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 2014; **30**: 614–620.
- 51 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* 2009; **10**: 421.
- 52 Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution* 2018; **35**: 1547–1549.
- 53 Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**: 4673–4680.
- 54 Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; **4**: 406–425.
- 55 Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980; **16**: 111–120.
- 56 Ward JH. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 1963; **58**: 236–244.
- 57 Andrews S. FastQC: A Quality Control tool for High Throughput Sequence Data. 2010. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed 16 Jul2021).
- 58 Ewels P, Magnusson M, Lundin S, Källér M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016; **32**: 3047–3048.
- 59 Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018; **34**: i884–i890.
- 60 Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics* 2020; **70**: e102.
- 61 Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 2018; **3**: 124.

- 62 Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM *et al.* Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* 2012; **158**: 1005–1015.
- 63 Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018; **9**: 5114.
- 64 Chattaway MA, Schaefer U, Tewolde R, Dallman TJ, Jenkins C. Identification of *Escherichia coli* and *Shigella* Species from Whole-Genome Sequences. *J Clin Microbiol* 2017; **55**: 616–623.
- 65 Roer L, Tchesnokova V, Allesøe R, Muradova M, Chattopadhyay S, Ahrenfeldt J *et al.* Development of a Web Tool for *Escherichia coli* Subtyping Based on *fimH* Alleles. *J Clin Microbiol* 2017; **55**: 2538–2543.
- 66 Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and Easy *In Silico* Serotyping of *Escherichia coli* Isolates by Use of Whole-Genome Sequencing Data. *J Clin Microbiol* 2015; **53**: 2410–2426.
- 67 Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH *et al.* Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* 2006; **60**: 1136–1151.
- 68 Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013; **arXiv**: 1303.3997v2 [q-bio.GN].
- 69 Gardner SN, Slezak T, Hall BG. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 2015; **31**: 2877–2878.
- 70 Gardner SN, Hall BG. When Whole-Genome Alignments Just Won't Work: kSNP v2 Software for Alignment-Free SNP Discovery and Phylogenetics of Hundreds of Microbial Genomes. *PLoS ONE* 2013; **8**: e81760.
- 71 Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Research* 2019; **47**: D687–D692.
- 72 Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**: 2068–2069.
- 73 Christopher K, Bruno E. Identification of Bacterial Species. In: O'Donnell MA (ed). *Tested Studies for Laboratory Teaching; Proceedings of the 24th Workshop/Conference of the Association for Biology Laboratory Education (ABLE)*. The Association, 2003, pp 103–130.
- 74 Phumudzo T, Ronald N, Khayaletu N, Fhatuwani M. Bacterial species identification getting easier. *Afr J Biotechnol* 2013; **12**: 5975–5982.
- 75 Schloss PD, Handelsman J. Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. *Appl Environ Microbiol* 2005; **71**: 1501–1506.
- 76 Johnson JS, Spakowicz DJ, Hong B-Y, Petersen LM, Demkowicz P, Chen L *et al.* Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun* 2019; **10**: 5029.

- 77 Asai T, Zaporozhets D, Squires C, Squires CL. An *Escherichia coli* strain with all chromosomal rRNA operons inactivated: Complete exchange of rRNA genes between bacteria. *Proceedings of the National Academy of Sciences* 1999; **96**: 1971–1976.
- 78 Moore WEC, Stackebrandt E, Kandler O, Colwell RR, Krichevsky MI, Truper HG *et al.* Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *International Journal of Systematic and Evolutionary Microbiology* 1987; **37**: 463–464.
- 79 Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ *et al.* Re-evaluating prokaryotic species. *Nat Rev Microbiol* 2005; **3**: 733–739.
- 80 Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* 2007; **57**: 81–91.
- 81 Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *PNAS* 2009; **106**: 19126–19131.
- 82 Kim M, Oh H-S, Park S-C, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology* 2014; **64**: 346–351.
- 83 Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Inbanathan FY, Veeraraghavan B. Accurate differentiation of *Escherichia coli* and *Shigella* serogroups: challenges and strategies. *New Microbes and New Infections* 2018; **21**: 58–62.
- 84 Krause DO, Little AC, Dowd SE, Bernstein CN. Complete Genome Sequence of Adherent Invasive *Escherichia coli* UM146 Isolated from Ileal Crohn’s Disease Biopsy Tissue. *J Bacteriol* 2011; **193**: 583–583.
- 85 Avasthi TS, Kumar N, Baddam R, Hussain A, Nandanwar N, Jadhav S *et al.* Genome of Multidrug-Resistant Uropathogenic *Escherichia coli* Strain NA114 from India. *J Bacteriol* 2011; **193**: 4272–4273.
- 86 Bekal S, Brousseau R, Masson L, Prefontaine G, Fairbrother J, Harel J. Rapid Identification of *Escherichia coli* Pathotypes by Virulence Gene Detection with DNA Microarrays. *J Clin Microbiol* 2003; **41**: 2113–2125.
- 87 Kuhnert P, Boerlin P, Frey J. Target genes for virulence assessment of *Escherichia coli* isolates from water, food and the environment. *FEMS Microbiol Rev* 2000; **24**: 107–117.
- 88 Robins-Browne RM, Holt KE, Ingle DJ, Hocking DM, Yang J, Tauschek M. Are *Escherichia coli* Pathotypes Still Relevant in the Era of Whole-Genome Sequencing? *Front Cell Infect Microbiol* 2016; **6**. doi:10.3389/fcimb.2016.00141.
- 89 Qadri F, Svennerholm A-M, Faruque ASG, Sack RB. Enterotoxigenic *Escherichia coli* in Developing Countries: Epidemiology, Microbiology, Clinical Features, Treatment, and Prevention. *Clin Microbiol Rev* 2005; **18**: 465–483.
- 90 Taxt A, Aasland R, Sommerfelt H, Nataro J, Puntervoll P. Heat-Stable Enterotoxin of Enterotoxigenic *Escherichia coli* as a Vaccine Target. *Infect Immun* 2010; **78**: 1824–1831.

- 91 Marteyn B, Gazi A, Sansonetti P. Shigella: A model of virulence regulation in vivo. *Gut Microbes* 2012; **3**: 104–120.
- 92 Venkatesan MM, Goldberg MB, Rose DJ, Grotbeck EJ, Burland V, Blattner FR. Complete DNA Sequence and Analysis of the Large Virulence Plasmid of *Shigella flexneri*. *Infect Immun* 2001; **69**: 3271–3285.
- 93 Buysse JM, Hartman TAB, Strockbine N, Venkatesan M. Genetic polymorphism of the ipaH multicopy antigen gene in *Shigella* spp. and enteroinvasive *Escherichia coli*. *Microb Pathog* 1995; **19**: 335–349.
- 94 Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB. Recent Advances in Understanding Enteric Pathogenic *Escherichia coli*. *Clin Microbiol Rev* 2013; **26**: 822–880.
- 95 Okeke IN, Nataro JP. Enteroaggregative *Escherichia coli*. *Lancet Infect Dis* 2001; **1**: 304–313.
- 96 Johnson JR, Murray AC, Gajewski A, Sullivan M, Snippes P, Kuskowski MA *et al*. Isolation and Molecular Characterization of Nalidixic Acid-Resistant Extraintestinal Pathogenic *Escherichia coli* from Retail Chicken Products. *Antimicrob Agents Chemother* 2003; **47**: 2161–2168.
- 97 Sarowska J, Futoma-Koloch B, Jama-Kmiecik A, Frej-Madrzak M, Ksiazczyk M, Bugla-Ploskowska G *et al*. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: recent reports. *Gut Pathog* 2019; **11**: 10.
- 98 Ewers C, Li G, Wilking H, Kiebling S, Alt K, Antao E *et al*. Avian pathogenic, uropathogenic, and newborn meningitis-causing *Escherichia coli*: How closely related are they? *International Journal of Medical Microbiology* 2007; **297**: 163–176.
- 99 Wijetunge DSS, Gongati S, DebRoy C, Kim KS, Couraud PO, Romero IA *et al*. Characterizing the pathotype of neonatal meningitis causing *Escherichia coli* (NMEC). *BMC Microbiol* 2015; **15**: 211.
- 100 Lane MC, Mobley HLT. Role of P-fimbrial-mediated adherence in pyelonephritis and persistence of uropathogenic *Escherichia coli* (UPEC) in the mammalian kidney. *Kidney International* 2007; **72**: 19–25.
- 101 Johnson JR, Stell AL. Extended Virulence Genotypes of *Escherichia coli* Strains from Patients with Urosepsis in Relation to Phylogeny and Host Compromise. *J INFECT DIS* 2000; **181**: 261–272.
- 102 Miyazaki J, Ba-Thein W, Kumao T, Obata Yasuoka M, Akaza H, Hayashi H. Type 1, P and S fimbriae, and afimbrial adhesin I are not essential for uropathogenic *Escherichia coli* to adhere to and invade bladder epithelial cells. *FEMS Immunology & Medical Microbiology* 2002; **33**: 23–26.
- 103 Blanco M, Blanco JE, Alonso MP, Mora A, Balsalobre C, Munoz F *et al*. Detection of pap, sfa and afa adhesin-encoding operons in uropathogenic *Escherichia coli* strains: Relationship with expression of adhesins and production of toxins. *Research in Microbiology* 1997; **148**: 745–755.

- 104 Yamamoto S, Tsukamoto T, Terai A, Kurazono H, Takeda Y, Yoshida O. Distribution of Virulence Factors in *Escherichia coli* Isolated from Urine of Cystitis Patients. *Microbiology and Immunology* 1995; **39**: 401–404.
- 105 Stærk K, Khandige S, Kolmos HJ, Møller-Jensen J, Andersen TE. Uropathogenic *Escherichia coli* Express Type 1 Fimbriae Only in Surface Adherent Populations Under Physiological Growth Conditions. *J Infect Dis* 2016; **213**: 386–394.
- 106 Connell I, Agace W, Klemm P, Schembri M, Marild S, Svanborg C. Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. *Proceedings of the National Academy of Sciences* 1996; **93**: 9827–9832.
- 107 Wright KJ, Seed PC, Hultgren SJ. Development of intracellular bacterial communities of uropathogenic *Escherichia coli* depends on type 1 pili. *Cellular Microbiology* 2007; **9**: 2230–2241.
- 108 Weiss J, Victor M, Cross AS, Elsbach P. Sensitivity of K1-Encapsulated *Escherichia coli* to Killing by the Bactericidal/Permeability-Increasing Protein of Rabbit and Human Neutrophils. *Infect Immun* 1982; **38**: 5.
- 109 Howard CJ, Glynn AA. The Virulence for mice of Strains of *Escherichia coli* related to the Effects of K Antigens on their Resistance to Phagocytosis and Killing by Complement. *Immunology* 1971; **20**: 767–777.
- 110 Sarkar S, Ulett GC, Totsika M, Phan M-D, Schembri MA. Role of Capsule and O Antigen in the Virulence of Uropathogenic *Escherichia coli*. *PLoS ONE* 2014; **9**: e94786.
- 111 Scholl D, Adhya S, Merrill C. *Escherichia coli* K1's Capsule Is a Barrier to Bacteriophage T7. *Appl Environ Microbiol* 2005; **71**: 4872–4874.
- 112 Robinson AE, Heffernan JR, Henderson JP. The iron hand of uropathogenic *Escherichia coli* : the role of transition metal control in virulence. *Future Microbiology* 2018; **13**: 745–756.
- 113 Pantopoulos K, Porwal SK, Tartakoff A, Devireddy L. Mechanisms of Mammalian Iron Homeostasis. *Biochemistry* 2012; **51**: 5705–5724.
- 114 Correnti C, Strong RK. Mammalian Siderophores, Siderophore-binding Lipocalins, and the Labile Iron Pool. *Journal of Biological Chemistry* 2012; **287**: 13524–13531.
- 115 Sandy M, Butler A. Microbial Iron Acquisition: Marine and Terrestrial Siderophores. *Chem Rev* 2009; **109**: 4580–4595.
- 116 Fischbach MA, Lin H, Liu DR, Walsh CT. How pathogenic bacteria evade mammalian sabotage in the battle for iron. *Nat Chem Biol* 2006; **2**: 132–138.
- 117 Bachman MA, Oyler JE, Burns SH, Caza M, Lépine F, Dozois CM *et al.* *Klebsiella pneumoniae* Yersiniabactin Promotes Respiratory Tract Infection through Evasion of Lipocalin 2. *Infect Immun* 2011; **79**: 3309–3316.
- 118 Russo TA, Shon AS, Beanan JM, Olson R, MacDonald U, Pomakov AO *et al.* Hypervirulent *K. pneumoniae* Secretes More and More Active Iron-Acquisition Molecules than “Classical” *K. pneumoniae* Thereby Enhancing its Virulence. *PLoS ONE* 2011; **6**: e26734.

- 119 Lorenzo VD, Bindereif A, Paw BH, Neilands JB. Aerobactin Biosynthesis and Transport Genes of Plasmid CoIV-K30 in *Escherichia coli* K-12. *Journal of Bacteriology* 1986; **165**: 9.
- 120 Hagan EC, Lloyd AL, Rasko DA, Faerber GJ, Mobley HLT. *Escherichia coli* Global Gene Expression in Urine from Women with Urinary Tract Infection. *PLoS Pathog* 2010; **6**: e1001187.
- 121 Johnson JR, O'Bryan TT, Delavari P, Kuskowski M, Stapleton A, Carlino U *et al.* Clonal Relationships and Extended Virulence Genotypes among *Escherichia coli* Isolates from Women with a First or Recurrent Episode of Cystitis. *J INFECT DIS* 2001; **183**: 1508–1517.
- 122 Soto SM, Smithson A, Horcajada JP, Martinez JA, Mensa JP, Vila J. Implication of biofilm formation in the persistence of urinary tract infection caused by uropathogenic *Escherichia coli*. *Clinical Microbiology and Infection* 2006; **12**: 1034–1036.
- 123 Burns SM, Hull SI. Loss of Resistance to Ingestion and Phagocytic Killing by O⁻ and K⁻ Mutants of a Uropathogenic *Escherichia coli* O75:K5 Strain. *Infect Immun* 1999; **67**: 6.
- 124 Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**: 119.
- 125 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* 2021; **10**: giab008.
- 126 Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* 2012; **6**: 80–92.
- 127 Manges AR, Geum HM, Guo A, Edens TJ, Fibke CD, Pitout JDD. Global Extraintestinal Pathogenic *Escherichia coli* (ExPEC) Lineages. *Clin Microbiol Rev* 2019; **32**: e00135-18.
- 128 Jørgensen SL, Stegger M, Kudirkiene E, Lilje B, Poulsen LL, Ronco T *et al.* Diversity and Population Overlap between Avian and Human *Escherichia coli* Belonging to Sequence Type 95. *mSphere* 2019; **4**: e00333-18.
- 129 Ciesielczuk H, Jenkins C, Chattaway M, Doumith M, Hope R, Woodford N *et al.* Trends in ExPEC serogroups in the UK and their significance. *Eur J Clin Microbiol Infect Dis* 2016; **35**: 1661–1666.
- 130 Adams-Sapper S, Diep BA, Perdreau-Remington F, Riley LW. Clonal Composition and Community Clustering of Drug-Susceptible and -Resistant *Escherichia coli* Isolates from Bloodstream Infections. *Antimicrob Agents Chemother* 2013; **57**: 490–497.
- 131 Bengtsson S, Naseer U, Sundsfjord A, Kahlmeter G, Sundqvist M. Sequence types and plasmid carriage of uropathogenic *Escherichia coli* devoid of phenotypically detectable resistance. *Journal of Antimicrobial Chemotherapy* 2012; **67**: 69–73.
- 132 Zhu Y, Dong W, Ma J, Yuan L, Hejair HMA, Pan Z *et al.* Characterization and virulence clustering analysis of extraintestinal pathogenic *Escherichia coli* isolated from swine in China. *BMC Veterinary Research* 2017; **13**: 94.

- 133 Valat C, Drapeau A, Beurlet S, Bachy V, Boulouis H-J, Pin R *et al.* Pathogenic *Escherichia coli* in Dogs Reveals the Predominance of ST372 and the Human-Associated ST73 Extra-Intestinal Lineages. *Front Microbiol* 2020; **11**: 580.
- 134 Massip C, Oswald E. Siderophore-Microcins in *Escherichia coli*: Determinants of Digestive Colonization, the First Step Toward Virulence. *Front Cell Infect Microbiol* 2020; **10**: 381.
- 135 Azpiroz MF, Bascuas T, Laviña M. Microcin H47 System: An *Escherichia coli* Small Genomic Island with Novel Features. *PLoS ONE* 2011; **6**: e26179.
- 136 Paszkiewicz K, Studholme DJ. De novo assembly of short sequence reads. *Briefings in Bioinformatics* 2010; **11**: 457–472.
- 137 de Souza da-Silva AP, de Sousa VS, Martins N, da Silva Dias RC, Bonelli RR, Riley LW *et al.* *Escherichia coli* sequence type 73 as a cause of community acquired urinary tract infection in men and women in Rio de Janeiro, Brazil. *Diagn Microbiol Infect Dis* 2017; **88**: 69–74.
- 138 Manges AR, Harel J, Masson L, Edens TJ, Portt A, Reid-Smith RJ *et al.* Multilocus Sequence Typing and Virulence Gene Profiles Associated with *Escherichia coli* from Human and Animal Sources. *Foodborne Pathogens and Disease* 2015; **12**: 302–310.
- 139 Carbonetti NH, Williams PH. A cluster of five genes specifying the aerobactin iron uptake system of plasmid ColV-K30. *Infect Immun* 1984; **46**: 7–12.
- 140 Di Lorenzo M, Stork M. Plasmid-Encoded Iron Uptake Systems. *Microbiology Spectrum* 2014; **2**: 10.1128/microbiolspec.PLAS-0030–2014.
- 141 Gibreel TM, Dodgson AR, Cheesbrough J, Fox AJ, Bolton FJ, Upton M. Population structure, virulence potential and antibiotic susceptibility of uropathogenic *Escherichia coli* from Northwest England. *J Antimicrob Chemother* 2012; **67**: 346–356.
- 142 Ward DV, Scholz M, Zolfo M, Taft DH, Schibler KR, Tett A *et al.* Metagenomic Sequencing with Strain-Level Resolution Implicates Uropathogenic *E. coli* in Necrotizing Enterocolitis and Mortality in Preterm Infants. *Cell Rep* 2016; **14**: 2912–2924.
- 143 Garcia TA, Ventura CL, Smith MA, Merrell DS, O'Brien AD. Cytotoxic Necrotizing Factor 1 and Hemolysin from Uropathogenic *Escherichia coli* Elicit Different Host Responses in the Murine Bladder. *Infect Immun* 2013; **81**: 99–109.
- 144 Iftekhhar A, Berger H, Bouznad N, Heuberger J, Boccellato F, Dobrindt U *et al.* Genomic aberrations after short-term exposure to colibactin-producing *E. coli* transform primary colon epithelial cells. *Nat Commun* 2021; **12**: 1003.
- 145 Nipič D, Podlesek Z, Budič M, črnigoj M, Žgur-Bertok D. *Escherichia coli* Uropathogenic-Specific Protein, Usp, Is a Bacteriocin-Like Genotoxin. *The Journal of Infectious Diseases* 2013; **208**: 1545–1552.
- 146 Ksiezarek M, Novais Â, Felga H, Mendes F, Escobar M, Peixe L. Phylogenomic analysis of a highly virulent *Escherichia coli* ST83 lineage with potential animal-human transmission. *Microbial Pathogenesis* 2021; **155**: 104920.
- 147 Liu X, Thungrat K, Boothe DM. Multilocus Sequence Typing and Virulence Profiles in Uropathogenic *Escherichia coli* Isolated from Cats in the United States. *PLoS ONE* 2015; **10**: e0143335.

- 148 Ishimaru K, Sasaki M, Narimatsu H, Arimizu Y, Gotoh Y, Nakamura K *et al.* Escherichia coli O8:H8 Carrying a Novel Variant of the Heat-Labile Enterotoxin LT2 Gene Caused Outbreaks of Diarrhea. *Open Forum Infect Dis* 2020; **7**: ofaa021.
- 149 Johnson TJ, Wannemuehler Y, Doetkott C, Johnson SJ, Rosenberger SC, Nolan LK. Identification of Minimal Predictors of Avian Pathogenic Escherichia coli Virulence for Use as a Rapid Diagnostic Tool. *J Clin Microbiol* 2008; **46**: 3987–3996.
- 150 Dissanayake DRA, Octavia S, Lan R. Population structure and virulence content of avian pathogenic Escherichia coli isolated from outbreaks in Sri Lanka. *Veterinary Microbiology* 2014; **168**: 403–412.
- 151 Ahmed AM, Shimamoto T, Shimamoto T. Molecular characterization of multidrug-resistant avian pathogenic Escherichia coli isolated from septicemic broilers. *International Journal of Medical Microbiology* 2013; **303**: 475–483.
- 152 Kaczmarek A, Budzyńska A, Gospodarek E. Prevalence of genes encoding virulence factors among Escherichia coli with K1 antigen and non-K1 E. coli strains. *Journal of Medical Microbiology* 2012; **61**: 1360–1365.
- 153 Morales C, Lee MD, Hofacre C, Maurer JJ. Detection of a novel virulence gene and a Salmonella virulence homologue among Escherichia coli isolated from broiler chickens. *Foodborne Pathog Dis* 2004; **1**: 160–165.
- 154 Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O. ClermonTyping: an easy-to-use and accurate in silico method for Escherichia genus strain phylotyping. *Microbial Genomics* 2018; **4**. doi:10.1099/mgen.0.000192.
- 155 Lu S, Jin D, Wu S, Yang J, Lan R, Bai X *et al.* Insights into the evolution of pathogenicity of Escherichia coli from genomic analysis of intestinal E. coli of Marmota himalayana in Qinghai–Tibet plateau of China. *Emerg Microbes Infect* 2016; **5**: e122.
- 156 El-Hamed HAA, Ibrahim GA. Molecular, bacteriological and clinical pathological studies on pneumonic calves with special reference to antibiotic resistance genes. *Assiut Veterinary Medical Journal* 2017; **63**: 144–160.
- 157 Yamaji R, Friedman CR, Rubin J, Suh J, Thys E, McDermott P *et al.* A Population-Based Surveillance Study of Shared Genotypes of Escherichia coli Isolates from Retail Meat and Suspected Cases of Urinary Tract Infections. *mSphere* 2018; **3**: e00179-18.
- 158 Magistro G, Mugnai L, Pastorelli R, Giovannetti L, Stead DE. Erwinia alni, a New Species Causing Bark Cankers of Alder (Alnus Miller) Species. *International Journal of Systematic Bacteriology* 1996; **46**: 720–726.
- 159 Hauben L, Moore ERB, Vauterin L, Steenackers M, Mergaert J, Verdonck L *et al.* Phylogenetic Position of Phytopathogens within the Enterobacteriaceae. *Systematic and Applied Microbiology* 1998; **21**: 384–397.
- 160 Moradi-Amirabad Y, Khodakaramian G. *Brenneria alni*, causal agent bark canker of *Alnus subcordata*. *J Phytopathol* 2020; **168**: 516–523.
- 161 Fernández-Bravo A, Figueras MJ. An Update on the Genus Aeromonas: Taxonomy, Epidemiology, and Pathogenicity. *Microorganisms* 2020; **8**: 129.

- 162 Hiransuthikul N, Tantisiriwat W, Lertutsahakul K, Vibhagool A, Boonma P. Skin and Soft-Tissue Infections among Tsunami Survivors in Southern Thailand. *Clinical Infectious Diseases* 2005; **41**: e93–e96.
- 163 Bogdanović R, Sarjanovifil Č, Marković M, Nikolić V, Ognjanović M, Sarjanović L *et al.* Haemolytic-uraemic syndrome associated with *Aeromonas hydrophila* enterocolitis. *Pediatr Nephrol* 1991; **5**: 293–295.
- 164 Abuhammour W, Hasan RA, Rogers D. Necrotizing Fasciitis Caused by *Aeromonas hydrophila* in an Immunocompetent Child: *Pediatric Emergency Care* 2006; **22**: 48–51.
- 165 Colston SM, Fullmer MS, Beka L, Lamy B, Gogarten JP, Graf J. Bioinformatic Genome Comparisons for Taxonomic and Phylogenetic Assignments Using *Aeromonas* as a Test Case. *mBio* 2014; **5**: e02136-14.
- 166 Heuzenroeder MW, Wong CYF, Flower RLP. Distribution of two hemolytic toxin genes in clinical and environmental isolates of *Aeromonas* spp.: correlation with virulence in a suckling mouse model. *FEMS Microbiol Lett* 1999; **174**: 131–136.
- 167 Erova TE, Sha J, Horneman AJ, Borchardt MA, Khajanchi BK, Fadl AA *et al.* Identification of a new hemolysin from diarrheal isolate SSU of *Aeromonas hydrophila*. *FEMS Microbiology Letters* 2007; **275**: 301–311.
- 168 Wong CYF, Heuzenroeder MW, Flower RLP. Inactivation of two haemolytic toxin genes in *Aeromonas hydrophila* attenuates virulence in a suckling mouse model. *Microbiology* 1998; **144**: 291–298.
- 169 Peraro MD, van der Goot FG. Pore-forming toxins: ancient, but never really out of fashion. *Nat Rev Microbiol* 2016; **14**: 77–92.
- 170 Suarez G, Khajanchi BK, Sierra JC, Erova TE, Sha J, Chopra AK. Actin cross-linking domain of *Aeromonas hydrophila* repeat in toxin A (RtxA) induces host cell rounding and apoptosis. *Gene* 2012; **506**: 369–376.
- 171 Li L, Rock JL, Nelson DR. Identification and Characterization of a Repeat-in-Toxin Gene Cluster in *Vibrio anguillarum*. *Infect Immun* 2008; **76**: 2620–2632.
- 172 Lo H-R, Lin J-H, Chen Y-H, Chen C-L, Shao C-P, Lai Y-C *et al.* RTX Toxin Enhances the Survival of *Vibrio vulnificus* During Infection by Protecting the Organism From Phagocytosis. *The Journal of Infectious Diseases* 2011; **203**: 1866–1874.
- 173 Coburn PS, Gilmore MS. The *Enterococcus faecalis* cytolysin: a novel toxin active against eukaryotic and prokaryotic cells. *Cellular Microbiology* 2003; **5**: 661–669.
- 174 Sałamaszyńska-Guz A, Serafińska I, Baçal P, Douthwaite S. Virulence properties of *Campylobacter jejuni* are enhanced by displaying a mycobacterial TlyA methylation pattern in its rRNA. *Cell Microbiol* 2020; **22**: e13199.
- 175 Monshupanee T. Increased Bacterial Hemolytic Activity is Conferred by Expression of TlyA Methyltransferase but not by its 2'-O-methylation of the Ribosome. *Curr Microbiol* 2013; **67**: 61–68.

- 176 Martino MC, Stabler RA, Zhang ZW, Farthing MJG, Wren BW, Dorrell N. Helicobacter pylori Pore-Forming Cytolysin Orthologue TlyA Possesses In Vitro Hemolytic Activity and Has a Role in Colonization of the Gastric Mucosa. *Infect Immun* 2001; **69**: 1697–1703.
- 177 Rahman A, Srivastava SS, Sneha A, Ahmed N, Krishnasastri MV. Molecular characterization of tlyA gene product, Rv1694 of Mycobacterium tuberculosis: A non-conventional hemolysin and a ribosomal RNA methyl transferase. *BMC Biochem* 2010; **11**: 35.
- 178 Yamada R, Tien LHT, Arai S, Tohya M, Ishida-Kuroki K, Nomoto R *et al.* Development of PCR for identifying *Streptococcus parasuis*, a close relative of *Streptococcus suis*. *The Journal of Veterinary Medical Science* 2018; **80**: 1101–1107.
- 179 Tien LHT, Nishibori T, Nishitani Y, Nomoto R, Osawa R. Reappraisal of the taxonomy of *Streptococcus suis* serotypes 20, 22, 26, and 33 based on DNA–DNA homology and sodA and recN phylogenies. *Veterinary Microbiology* 2013; **162**: 842–849.
- 180 Wang J, Yi X, Liang P, Tao Y, Wang Y, Jin D *et al.* Investigation of the Genomic and Pathogenic Features of the Potentially Zoonotic *Streptococcus parasuis*. *Pathogens* 2021; **10**: 834.
- 181 Poyart C, Quesne G, Trieu-Cuot P. Taxonomic dissection of the *Streptococcus bovis* group by analysis of manganese-dependent superoxide dismutase gene (sodA) sequences: reclassification of ‘*Streptococcus infantarius* subsp. coli’ as *Streptococcus lutetiensis* sp. nov. and of *Streptococcus bovis* biotype II.2 as *Streptococcus pasteurianus* sp. nov. *Int J Syst Evol Microbiol* 2002; **52**: 1247–1255.
- 182 Jans C, Boleij A. The Road to Infection: Host-Microbe Interactions Defining the Pathogenicity of *Streptococcus bovis*/*Streptococcus equinus* Complex Members. *Frontiers in Microbiology* 2018; **9**: 603.
- 183 Papadimitriou K, Anastasiou R, Mavrogonatou E, Blom J, Papandreou NC, Hamodrakas SJ *et al.* Comparative genomics of the dairy isolate *Streptococcus macedonicus* ACA-DC 198 against related members of the *Streptococcus bovis*/*Streptococcus equinus* complex. *BMC Genomics* 2014; **15**: 272.
- 184 Almuzara M, Bonofiglio L, Cittadini R, Vera Ocampo C, Montilla A, del Castillo M *et al.* First Case of *Streptococcus lutetiensis* Bacteremia Involving a Clindamycin-Resistant Isolate Carrying the InuB Gene. *J Clin Microbiol* 2013; **51**: 4259–4261.
- 185 Chen P, Qiu Y, Liu G, Li X, Cheng J, Liu K *et al.* Characterization of *Streptococcus lutetiensis* isolated from clinical mastitis of dairy cows. *Journal of Dairy Science* 2021; **104**: 702–714.
- 186 Jin D, Chen C, Li L, Lu S, Li Z, Zhou Z *et al.* Dynamics of fecal microbial communities in children with diarrhea of unknown etiology and genomic analysis of associated *Streptococcus lutetiensis*. *BMC Microbiol* 2013; **13**: 141.
- 187 Yang C, Wang D, Zhou Q, Xu J. Bacteremia Due to Vancomycin-Resistant *Leuconostoc lactis* in a Patient With Pneumonia and Abdominal Infection. *Am J Med Sci* 2015; **349**: 282–283.
- 188 Deye G, Lewis J, Patterson J, Jorgensen J. A Case of *Leuconostoc* Ventriculitis with Resistance to Carbapenem Antibiotics. *Clinical Infectious Diseases* 2003; **37**: 869–870.

- 189 Deng Y, Zhang Z, Xie Y, Xiao Y, Kang M, Fan H. A mixed infection of *Leuconostoc lactis* and vancomycin-resistant *Enterococcus* in a liver transplant recipient. *Journal of Medical Microbiology* 2012; **61**: 1621–1624.
- 190 Jeong D-W, Lee J-H. Antibiotic resistance, hemolysis and biogenic amine production assessments of *Leuconostoc* and *Weissella* isolates for kimchi starter development. *LWT - Food Science and Technology* 2015; **64**: 1078–1084.
- 191 Kikuchi K, Totsuka K, Shimizu K, Yoshida K, Kobayashi M, Tomonaga O *et al.* Microbiological and Clinical Studies of Vancomycin Resistant *Leuconostoc* spp. and *Pediococcus* spp. Isolated from Septicemia Patients. *kansenshogakuzasshi* 1994; **68**: 1084–1092.
- 192 Chen Y, Chang C, Pan S, Wang L, Chang Y, Wu H *et al.* *Lactococcus taiwanensis* sp. nov., a lactic acid bacterium isolated from fresh cummingcordia. *Int J Syst Evol Microbiol* 2013; **63**: 2405–2409.
- 193 Quilodrán-Vega S, Albarracín L, Mansilla F, Arce L, Zhou B, Islam MA *et al.* Functional and Genomic Characterization of *Ligilactobacillus salivarius* TUCO-L2 Isolated From Lama glama Milk: A Promising Immunobiotic Strain to Combat Infections. *Frontiers in Microbiology* 2020; **11**: 3123.
- 194 Balasubramanian B, Soundharajan I, Al-Dhabi NA, Vijayaraghavan P, Balasubramanian K, Valan Arasu M *et al.* Probiotic Characteristics of *Ligilactobacillus salivarius* AS22 Isolated from Sheep Dung and Its Application in Corn-Fox Tail Millet Silage. *Applied Sciences* 2021; **11**: 9447.
- 195 Vasilakopoulou A, Vourli S, Siafakas N, Kavatha D, Tziolos N, Pournaras S. *Enterococcus casseliflavus* Bacteraemia in a Patient with Chronic Renal Disease. *Infect Dis Rep* 2020; **12**: 70–73.
- 196 Reid KC, Cockerill FR, Patel R. Clinical and Epidemiological Features of *Enterococcus casseliflavus/flavescens* and *Enterococcus gallinarum* Bacteremia: A Report of 20 Cases. *Clinical Infectious Diseases* 2001; **32**: 1540–1546.
- 197 Mundy LM, Sahm DF, Gilmore M. Relationships between Enterococcal Virulence and Antimicrobial Resistance. *Clin Microbiol Rev* 2000; **13**: 513–522.
- 198 Cotta MA, Whitehead TR, Falsen E, Moore E, Lawson PA. Two novel species *Enterococcus lemanii* sp. nov. and *Enterococcus eurekensis* sp. nov., isolated from a swine-manure storage pit. *Antonie van Leeuwenhoek* 2013; **103**: 89–98.
- 199 Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH *et al.* Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biology* 2010; **11**: R31.
- 200 Artsimovitch I, Landick R. The Transcriptional Regulator RfaH Stimulates RNA Chain Synthesis after Recruitment to Elongation Complexes by the Exposed Nontemplate DNA Strand. *Cell* 2002; **109**: 193–203.
- 201 Kang JY, Mooney RA, Nedialkov Y, Saba J, Mishanina TV, Artsimovitch I *et al.* Structural basis for transcript elongation control by NusG family universal regulators. *Cell* 2018; **173**: 1650-1662.e14.

- 202 Soloaga A, Ostolaza H, Goñi FM, De La Cruz F. Purification of Escherichia coli Pro-Haemolysin, and a Comparison with the Properties of Mature α -haemolysin. *European Journal of Biochemistry* 1996; **238**: 418–422.
- 203 Herlax V, Maté S, Rimoldi O, Bakás L. Relevance of Fatty Acid Covalently Bound to Escherichia coli α -Hemolysin and Membrane Microdomains in the Oligomerization Process. *J Biol Chem* 2009; **284**: 25199–25210.
- 204 Linhartová I, Bumba L, Mašín J, Basler M, Osička R, Kamanová J *et al.* RTX proteins: a highly diverse family secreted by a common mechanism. *FEMS Microbiol Rev* 2010; **34**: 1076–1112.
- 205 Boehm DF, Welch RA, Snyder IS. Calcium Is Required for Binding of Escherichia coli Hemolysin (HlyA) to Erythrocyte Membranes. *Infect Immun* 1990; **58**: 1951–1958.
- 206 Baumann U, Wu S, Flaherty KM, McKay DB. Three-dimensional structure of the alkaline protease of Pseudomonas aeruginosa: a two-domain protein with a calcium binding parallel beta roll motif. *EMBO J* 1993; **12**: 3357–3364.
- 207 Ludwig A, Jarchau T, Benz R, Goebel W. The repeat domain of Escherichia coli haemolysin (HlyA) is responsible for its Ca²⁺-dependent binding to erythrocytes. *Mol Gen Genet* 1988; **214**: 553–561.
- 208 Herlax V, de Alaniz MJT, Bakás L. Role of lipopolysaccharide on the structure and function of α -hemolysin from Escherichia coli. *Chemistry and Physics of Lipids* 2005; **135**: 107–115.
- 209 Bauer ME, Welch RA. Pleiotropic effects of a mutation in rfaC on Escherichia coli hemolysin. *Infect Immun* 1997; **65**: 2218–2224.
- 210 Belogurov GA, Sevostyanova A, Svetlov V, Artsimovitch I. Functional regions of the N-terminal domain of the antiterminator RfaH. *Mol Microbiol* 2010; **76**: 286–301.
- 211 Robertson KP, Smith CJ, Gough AM, Rocha ER. Characterization of Bacteroides fragilis Hemolysins and Regulation and Synergistic Interactions of HlyA and HlyB. *Infect Immun* 2006; **74**: 2304–2316.

付録

1. 資料

1.1. *Aeromonas* 属の 16S rRNA V3-V4 領域のアレル配列

単離した *Aeromonas* 属細菌株の 16S rRNA 遺伝子 V3-V4 領域のアレル配列は以下の通りであった。いずれかの株で検出割合が 5%を超えたアレルを記載する。いずれかのアレルで変異の見つかった塩基を網掛け表示している。なお、一行あたり 50 塩基で表記する。

>Allele00

TGGGGAATATTGCACAATGGGGGGAAACCCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTC
AGTAGCTAATACTCTGCTGGCTGTGTACGTTACTCGCAGAAGAAGCACCCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAATTGCATTTAAAACTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele31

TGGGGAATATTGCACAATGGGGGGAAACCCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTC
AGTAGCTAATACTCTGCTGGCTGTGTACGTTACTCGCAGAAGAAGCACCCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAATTGCATTTAAAACTGTCCAGCTAGAGT
CTTGTAGAGGGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele18

TGGGGAATATTGCACAATGGGGGGAAACCCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGCC
AGTAGCTAATACTCTGCTGGCTGTGTACGTTACTCGCAGAAGAAGCACCCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAATTGCATTTAAAACTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele14

TGGGGAATATTGCACAATGGGGGGAAACCCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTC
AGTAGCTAATACTCTGCTGGCTGTGTACGTTACTCGCAGAAGAAGCACCCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAATTGCATTTAAAACTGTCCGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele20

TGGGGAATATTGCACAATGGGGGGA^AAA^CCCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTC
AGTAGCTAATACTCTGCTGACTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCAITTA^AAA^AACTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGTCCCCTGGACA^AAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA-

>Allele02

TGGGGAATATTGCACAATGGGGGGA^AAA^CCCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTC
AGTAGCTAATACTCTGCTGACTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCAITTA^AAA^AACTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACA^AAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA-

>Allele13

TGGGGAATATTGCACAATGGGGGGA^AAA^CCCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTC
AGTAGCTAATACTCTGCTGACTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCAITTA^AAA^AACTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACA^AAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA-

>Allele15

TGGGGAATATTGCACAATGGGGGGA^AAA^CCCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTC
AGTAGCTAATACTCTGCTGACTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCAITTA^AAA^AACTGTCCGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACA^AAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA-

>Allele23

TGGGGAATATTGCACAATGGGGGGA^AAA^CCCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTC
AGTAGCTAATACTCTGCTGACTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCAITTA^AAA^AACTGTCCGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACA^AAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA-

>Allele12

TGGGGAATATTGCACAATGGGGGGA^AAA^CCCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTC
AGTAGCTAATACTACTGGCTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCATTTAAA^AACTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACA^AAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA-

>Allele16

TGGGGAATATTGCACAATGGGGGGA^AAA^CCCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTC
AGTAGCTAATACTGCTGGCTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCATTTAAA^AACTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACA^AAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA-

>Allele07

TGGGGAATATTGCACAATGGGGGGA^AAA^CCCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTT
GGTAGCTAATAACTGCCAGCTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCATTTAAA^AACTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACA^AAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA-

>Allele17

TGGGGAATATTGCACAATGGGGGGA^AAA^CCCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTT
GGTAGCTAATAACTGCCAGCTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCATTTAAA^AACTGTCCAGCTAGAGT
TTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACA^AAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA-

>Allele05

TGGGGAATATTGCACAATGGGGGGA^AAA^CCCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTT
GTTGGCTAATA^CCCAACAGCTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCATTTAAA^AACTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACA^AAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA-

>Allele22

TGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTT
GTTGGCTAATATCCAACAGCTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCATTTAAAAGTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA-

>Allele45

TGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTT
GTTGGCTAATATCCAACAGCTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCATTTAAAAGTGTCCGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA-

>Allele42

TGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTT
GATACCTAATACCGTATCAACTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCATTTAAAAGTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele06

TGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTT
GGTGCCTAATACCGTATCAACTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCATTTAAAAGTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA-

>Allele10

TGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTT
GATGCCTAATACCGCATCAGCTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAITGCATTTAAAAGTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA-

>Allele27

TGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTT
GATGCCTAATACGTATCAGCTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAATTGCATTTAAAAGTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele44

TGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTT
GATGCCTAATACGTATCAACTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AATCCCCGGGCTCAACCTGGGAAATTGCATTTAAAAGTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele01

TGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTT
GATGCCTAATACGTATCAACTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAATTGCATTTAAAAGTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele36

TGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAAGTT
GATGCCTAATACGTATCAACTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAATTGCATTTAAAAGTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele41

TGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAAGGTT
GATGCCTAATACGTATCAACTGTGACGTTACTCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGATAAGTTAGATGTGA
AAGCCCCGGGCTCAACCTGGGAAATTGCATTTAAAAGTGTCCAGCTAGAGT
CTTGTAGAGGGGGG-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele43

TGGGGAATATTGCACAATGGGGGGA~~A~~AA~~C~~CCTGATGCAGCCATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAG~~C~~ACTTTCAGCG~~A~~GGAGGAA~~A~~GG~~T~~
GATGCCTAATAC~~C~~G~~T~~ATCA~~A~~CT~~T~~GTGACGTTAC~~T~~CGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGGA~~A~~TAAGT~~I~~AG~~G~~TGTGA
AA~~G~~CCCCGGGCTCAACCTGGGAA~~T~~TGCAT~~T~~T~~A~~AA~~A~~ACTG~~T~~CC~~A~~GCT~~A~~AGAGT
CT~~T~~GTAGAGGGGGG~~G~~-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGG~~C~~CCCCCTGGAC~~A~~AAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA~~A~~

>Allele08

TGGGGAATATTGCACAATGGGG~~C~~G~~C~~AA~~G~~CCTGATGCAGCCATGCCGCGTGT
A~~T~~GAGAAGGCCTTCGGGTTGTAAAG~~T~~ACTTTCAGCG~~G~~GGAGGAA~~A~~GG~~G~~AG~~G~~
TAAAG~~T~~AATA~~C~~CT~~T~~TGCT~~C~~AT~~T~~GACGTTAC~~C~~CGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTT~~T~~G~~T~~TAAGT~~C~~AG~~A~~TGTGA
AA~~T~~CCCCGGGCTCAACCTGGGAA~~C~~TGCAT~~C~~T~~G~~A~~T~~ACTG~~G~~CA~~A~~GCT~~T~~GAGT
CT~~C~~GTAGAGGGGGG~~G~~-TAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGA
TCTGGAGGAATACCGGTGGCGAAGGCGG~~C~~CCCCCTGGAC~~G~~AAGACTGACGC
TCAGGTGCGAAAGCGTGGGGAGCAAACA~~A~~

1.2. *Brenneria alni*, *Escherichia coli*, *Escherichia fergusonii*, *Shigella flexneri*, *Shigella sonnei* と推定された株の 16S rRNA V3-V4 領域のアレル配列

単離大腸菌の 16S rRNA 遺伝子 V3-V4 領域のアレル配列は以下の通りであった。いずれかの株で検出割合が 5%を超えたアレルを記載する。いずれかのアレルで変異の見つかった塩基を網掛け表示している。なお、一行あたり 50 塩基で表記する。

>Allele109

TGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGT
ATGAAGAAGGCCTT~~C~~G~~A~~GTTGTAAAGTACTTTCAGCGGGGAGGAAGGGAG
TAAAGTTAATACCTTT~~C~~G~~T~~CATTGACGTTACCCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGA
AATCCCCGGGCTCAACCTGGGAA~~A~~CTGCATCTGATACTGGCAAGCTTGAAGT
CTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGAT
CTGGAGGAATACCGGTGGCGAAGGCGGCCCCCT~~G~~GACGAAGACTGACGCT
CAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele114

TGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGT
ATGAAGAAGGCCTT~~A~~G~~C~~GTTGTAAAGTACTTTCAGCGGGGAGGAAGGGAG
TAAAGTTAATACCTTT~~C~~G~~T~~CATTGACGTTACCCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGA
AATCCCCGGGCTCAACCTGGGAA~~A~~CTGCATCTGATACTGGCAAGCTTGAAGT
CTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGAT
CTGGAGGAATACCGGTGGCGAAGGCGGCCCCCT~~G~~GACGAAGACTGACGCT
CAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele108

TGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGT
ATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCAGCGGGGAGGAAGGGAG
TAAAGTTAATACCTTTGCTCATTGACGTTACCCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTAAAGTCAGATGTGA
AATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGT
CTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGAT
CTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTTGACGAAGACTGACGCT
CAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele107

TGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGT
ATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCAGCGGGGAGGAAGGGAG
TAAAGTTAATACCTTTGCTCATTGACGTTACCCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTAAAGTCAGATGTGA
AATCCCCGGGCTCAACCTGGGAGCTGCATCTGATACTGGCAAGCTTGAGT
CTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGAT
CTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCT
CAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele106

TGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGT
ATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCAGCGGGGAGGAAGGGAG
TAAAGTTAATACCTTTGCACATTGACGTTACCCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTAAAGTCAGATGTGA
AATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGT
CTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGAT
CTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCT
CAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele100

TGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGT
ATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCAGCGGGGAGGAAGGGAG
TAAAGTTAATACCTTTGCTCATTGACGTTACCCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTAAAGTCAGATGTGA
AATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGT
CTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGAT
CTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCT
CAGGTGCGAAAGCGTGGGGAGCAAACA

>Allele104

TGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGT
ATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCAGCGGGGAGGAAGGGAG
TAAAGTTAATACCTTTACTCATTGACGTTACCCGCAGAAGAAGCACCGGC
TAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCG
GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTAAAGTCAGATGTGA
AATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGT
CTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGAT
CTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCT
CAGGTGCGAAAGCGTGGGGAGCAAACA

2. 図表

表 S 1 VirulenceFinder2.0 で BLAST 検索を行いヒットした病原性遺伝子と、クエリー長 / テンプレート長 (1/4)

Strain	<i>mcmA</i>	<i>mchF</i>	<i>mchC</i>	<i>cea</i>	<i>mchB</i>	<i>sat</i>	<i>iutA</i>	<i>iha</i>	<i>iucC</i>	<i>papA F12</i>	<i>pic</i>	<i>focG</i>	<i>kpsMII K5</i>	<i>astA</i>
WWo073														
WWo070														
KOr014														
WWo108														
WWo014														
WWo005														
KKa004	279/279	2115/2115	1551/1551	240/240	294/294					504/504		504/504	777/777	
WWo003	279/279	2115/2115	1551/1551	240/240	294/294	3888/3888	2202/2202	2091/2091	1743/1743		4116/4116	504/504	777/777	
WWo147										504/504		4116/4116	504/504	777/777
WWo156											3510/4119			
KKo007	279/279													
WWo027														
WWo102	279/279	2115/2115	1551/1551	240/240	294/294					504/504				
WWo055														117/117
WWo083														

表 S 2 VirulenceFinder2.0 で BLAST 検索を行いヒットした病原性遺伝子と、クエリー長 / テンプレート長 (2/4)

Strain	<i>etsC</i>	<i>hlyF</i>	<i>hlyE</i>	<i>ireA</i>	<i>lpfA</i>	<i>mcbA</i>	<i>sfaE</i>	<i>tcpC</i>	<i>kpsMII</i>	<i>papA F48</i>	<i>gad</i>	<i>senB</i>	<i>traT</i>
WWo073													
WWo070													
KOr014						210/210					1110/1401	1176/1176	118/177
WWo108												1176/1176	735/735
WWo014												1176/1176	735/735
WWo005													
KKa004								924/924					
WWo003											1401/1401		
WWo147											1401/1401	1176/1176	735/735
WWo156								574/924	777/777	552/552	1401/1401		
KKo007						210/210	696/696	556/924	777/777	552/552	845/1401	1176/1176	118/177
WWo027								924/924	777/777	552/552	1401/1401	1176/1176	735/735
WWo102									777/777		1401/1401		
WWo055	1371/1371	1110/1110		1535/2049	573/573						1401/1401		732/732
WWo083			903/903	2049/2049	573/573						1401/1401		735/735

表 S 3 VirulenceFinder2.0 で BLAST 検索を行いヒットした病原性遺伝子と、クエリー長 / テンプレート長 (3/4)

Strain	<i>papC</i>	<i>usp</i>	<i>kpsE</i>	<i>yfeV</i>	<i>clbB</i>	<i>cnf1</i>	<i>sitA</i>	<i>iss</i>	<i>terC</i>	<i>vat</i>	<i>hra</i>	<i>fyuA</i>	<i>chuA</i>
WWo073	2511/2511	1041/1041	1149/1149	750/750	9621/9621	3045/3045	915/915	342/342	714/714	4132/4131	741/741	2022/2022	1983/1983
WWo070	2511/2511	1041/1041	1149/1149	750/750	9621/9621	3045/3045	915/915	342/342	714/714	4132/4131	741/741	2022/2022	1983/1983
KOr014	2511/2511	1041/1041	1149/1149	750/750	9621/9621	3045/3045	915/915	342/342	714/714	4132/4131	741/741	2022/2022	1983/1983
WWo108	2511/2511	1041/1041	1149/1149	750/750	9621/9621	3045/3045	915/915	342/342	714/714	4132/4131	741/741	2022/2022	1983/1983
WWo014	2511/2511	1041/1041	1149/1149	750/750	9621/9621	3045/3045	915/915	342/342	714/714	4133/4131	741/741	2022/2022	1983/1983
WWo005	2511/2511	1041/1041	1149/1149	750/750	9621/9621	3045/3045	915/915	342/342	714/714	4132/4131	741/741	2022/2022	1983/1983
KKa004	2511/2511	1041/1041	1149/1149	567/567	9621/9621	3045/3045	915/915	294/294	714/714	4132/4131	741/741	2022/2022	1983/1983
WWo003		1782/1782	1149/1149	567/567	9621/9621	2524/3045	915/915	342/342	714/714	4132/4131	741/741	2022/2022	1983/1983
WWo147	2520/2520	1782/1782	1149/1149	567/567	9621/9621	3045/3045	915/915	342/342	714/714		697/741	2022/2022	1983/1983
WWo156	2511/2511	1041/1041	1149/1149	567/567	9621/9621	3045/3045	915/915	342/342	714/714	4131/4131	741/741	2022/2022	1983/1983
KKo007	2091/2511	1041/1041	1149/1149	567/567	9621/9621	2550/3045	915/915	342/342	714/714	4098/4131	482/741	2022/2022	1983/1983
WWo027	2511/2511	1041/1041	1149/1149	567/567	9621/9621	3045/3045	915/915	342/342	714/714	4131/4131	482/741	2022/2022	1983/1983
WWo102	2520/2520	1043/1041	1149/1149	567/567	9621/9621	3045/3045	915/915	342/342	714/714	4131/4131	741/741	2022/2022	1983/1983
WWo055							915/915	309/309	959/966				
WWo083								294/294	714/714	2729/4131	747/747	2022/2022	1983/1983

表 S 4 VirulenceFinder2.0 で BLAST 検索を行いヒットした病原性遺伝子と、クエリー長 / テンプレート長 (4/4)

Strain	<i>irp2</i>	<i>ompT</i>	<i>focC</i> / <i>sfaE</i>	<i>iroN</i>	<i>sfaD</i>	<i>sfaS</i>	<i>papA</i> <i>F43</i>	<i>ibeA</i>	<i>kpsMII</i> <i>K1</i>	<i>neuC</i>
WWo073	6108/6108	954/954	696/696	2178/2178	540/540	492/492	565/564	1371/1371	777/777	741/1176
WWo070	6108/6108	954/954	696/696	2178/2178	540/540	492/492	565/564	1371/1371	777/777	1176/1176
KOr014	6108/6108	954/954	696/696	2178/2178	540/540	492/492	564/564	1371/1371	777/777	1176/1176
WWo108	6108/6108	954/954					564/564	1371/1371	777/777	1176/1176
WWo014	6108/6108	954/954	696/696	2178/2178	540/540	492/492	564/564	1371/1371	777/777	1073/1176
WWo005	6108/6108	954/954	696/696	2178/2178	540/540	492/492	564/564	1371/1371	777/777	1176/1176
KKa004	6108/6108	954/954	696/696	2178/2178	447/447		564/564			
WWo003	6108/6108	954/954	696/696	2178/2178	363/447		543/564			
WWo147	6108/6108	954/954	696/696	2178/2178	447/447					
WWo156	6108/6108	954/954								
KKo007	6108/6108	954/954		2178/2178	540/540	492/492				
WWo027	6108/6108	954/954								
WWo102	6108/6108	954/954	696/696	2178/2178	447/447			1371/1371		
WWo055				2178/2178						
WWo083	6108/6108	954/954								

3. Missense Mutation Viewer (MMViewer)

本研究のため開発した変異解析ソフト Missense Mutation Viewer (MMViewer) v0.0.9 の
使用法及びソースコードをここに記載する。ソースコードやインストール方法について
は、GitHub (<https://github.com/KIkebata/mmviewer>) にアップロードしている。

MMViewer では、リファレンスとするコンプリートゲノムもしくはコンティグに対して
ショートリードをアラインメントして、対象とする CDS 内のミスセンス変異と、CDS の
上流、下流の変異をグラフで表示するソフトウェアである。MMViewer は linux OS 上で動
作する。MMViewer の解析に必要なデータは、「クオリティコントロール済みのショート
リードデータ (fastq フォーマット)」、「リファレンスとして使用するコンプリートゲノ
ムもしくはコンティグ (fasta フォーマット)」、「変異の有無を確認したい遺伝子の塩基
配列もしくはアミノ酸配列 (fasta フォーマット)」の3つである。MMViewer は3つの機
能「対象 CDS 領域の探索」、「アラインメント」、「グラフの描画」から構成される。

対象 CDS 領域の検索 : get_target

「対象 CDS 領域の探索」では、「リファレンスとして使用するコンプリートゲノムもし
くはコンティグ (fasta フォーマット)」、「変異の有無を確認したい遺伝子の塩基配列も
しくはアミノ酸配列 (fasta フォーマット)」を元に、リファレンスのコンプリートゲノム
上もしくはコンティグ上から解析対象となる CDS 領域を、BLAST プログラムを用いて検
索する。実行コマンドは次の通りである。

```
$ mmviewer get_target -c reference.fasta -g genes.fasta  
-o output/directory -t prot
```

-c リファレンスとして使用するコンプリートゲノムもしくはコンティグ (fasta フォー
マット) のファイルパスを指定する (required)

-g 変異の有無を確認したい遺伝子の塩基配列もしくはアミノ酸配列 (fasta フォーマッ
ト) のファイルパスを指定する (required)

-o 出力先のディレクトリのパスを指定する (required)

-t -g で指定した fasta フォーマットのファイルがアミノ酸配列を記載しているものであ
る場合は prot、塩基配列を記載しているものである場合は nucl を指定する (required)

-u 対象とする遺伝子の上流の変異を確認したい場合はここに数値を入力する。

例 : -u 2000 (対象遺伝子の上流 2000 bp の変異を確認したい場合)

(optional, default 0)

-l 対象とする遺伝子の下流の変異を確認したい場合はここに数値を入力する。

例 : -l 2000 (対象遺伝子の下流 2000 bp の変異を確認したい場合)

(optional, default 0)

上記のコマンドを実行すると、-o で指定したディレクトリ内に target ディレクトリが作
成される。すでに target ディレクトリが存在する場合は target_1 が作成され、target_1 が存

在する場合は `target_2` が新たに作成される。`target` ディレクトリ内には、`blastdb` ディレクトリ、`blastx_out.tsv` もしくは `blastn_out.tsv` ファイル、`cds.gff` ファイル、`cds_nucl.fa` ファイル、`cds_prot.fa` ファイル、`target.bed` ファイルが作成される。`cds_prot.fa`、`cds_nucl.fa`、`cds.gff` ファイルはいずれも `prodigal` によって予測された CDS が記録されている。`blastdb` は、`-g` で指定した遺伝子配列を元に作成した BLAST データベースファイルが格納されている。BLAST データベースに対して、`cds_nucl.fa` ファイルの BLAST 相同性検索を行い、リファレンスゲノム上の対象遺伝子を探索する。これにより見つかった対象の遺伝子領域の位置を `target.bed` ファイルとして記録される。`target.bed` ファイルは、「`chrom`」、「`chromStart`」、「`chromEnd`」、「`name`」、「`score`」、「`strand`」、「`CDS_Start`」、「`CDS_End`」の 8 列からなるファイルである。`gen_graph` 実行時には、`target.bed` ファイル 1 行あたり 1 つのグラフが出力される。`chrom` の列でリファレンスゲノムの名前を指し、`chromStart` と `chromEnd` はグラフ作成時に対象とする領域、`name` は対象領域の名前を表す。`score` は `get_target` 解析における BLAST 検索で得られた `full length pident` の値に 1000 をかけて得られた整数値で、値が 1000 に近いほど `-g` で指定した遺伝子配列に近い配列であることを意味する。`strand` は翻訳の向きを表す。`CDS_Start` と `CDS_End` は CDS の領域を表す。`target.bed` ファイルは、グラフ作成時に使用するファイルで、解析を行いたい遺伝子領域を `score` の値に応じたフィルタリングや、`name` を分かりやすい名前に変更する、といった編集を必要に応じて行う。なお、`name` が同一の場合、`gen_graph` では同一ファイル上に出力され、グラフ領域はその `name` の行の `chromStart` と `chromEnd` の最小値と最大値に挟まれた領域を表示する。

アラインメント : alignment

「アラインメント」では、「クオリティコントロール済みのショートリードデータ (`fastq` フォーマット)」、「リファレンスとして使用するコンプライトゲノムもしくはコンティグ (`fasta` フォーマット)」を元に、アラインメントを実施する。実行前に、2 列もしくは 3 列の `csv` ファイルであるアラインメントコンフィグファイルを作成する必要がある。ペアドエンドリードの場合、アラインメントコンフィグファイルは、1 行目に (`sample_name,fq1,fq2`) 等のヘッダー、2 行目以降では 1 列目にはそれぞれのサンプル名、2 列目と 3 列目に「クオリティコントロール済みのショートリードデータ (`fastq` フォーマット)」の `forward` と `reverse` のそれぞれのファイルパスを入力して保存しておく。シングルエンドリードの場合、アラインメントコンフィグファイルは 1 行目に (`sample_name,fa`) 等のヘッダー、2 行目以降では 1 列目にはそれぞれのサンプル名、2 列目に「クオリティコントロール済みのショートリードデータ (`fastq` フォーマット)」のファイルパスを入力して保存しておく。フォーマットは `csv` ファイル (カンマ区切り) としておく。実行コマンドは次のとおりである。

```
$ mmviewer alignment -a alignment_config.config -c reference.fasta  
-o output/directory
```

- a アラインメントコンフィグファイルのファイルパスを指定する (required)
- c リファレンスとして使用するコンプライトゲノムもしくはコンティグ (`fasta` フォーマット) のファイルパスを指定する (required)
- o 出力先のディレクトリのパスを指定する (required)

上記のコマンドを実行すると、`-o`で指定したディレクトリ内に `alignment` ディレクトリが作成される。すでに `target` ディレクトリが存在する場合は `alignment_1` が作成され、`alignment_1` が存在する場合は `alignment_2` が新たに作成される。`alignment` ディレクトリ内には、`reference` ディレクトリ、`bam` ディレクトリ、`graph_config.csv` ファイルが生成される。`reference` ディレクトリには、`-c`で指定した `fasta` フォーマットのファイルのコピーと `bwa index` プログラムで作成されたインデックスファイルが格納される。`bam` ディレクトリには、`bwa mem` プログラムによって `-c`で指定されたりファレンスゲノムにアラインメントされた `bam` フォーマットのファイルが格納される。`graph_config.csv` ファイルは、「sample」、「bam」の2列で構成される `csv` ファイルである。「sample」の列にはサンプル名、`bam`の列には作成された `bam` ファイルのパスが入力される。`gen_graph` では、`graph_config.csv` ファイルの行の順でグラフが作成されるため、作成されるグラフのサンプルを並び替えて表示したい場合は、この `graph_config.csv` ファイルの行を並び替えるとよい。また、「sample」名を分かりやすい名前に変更することも可能で、必要に応じて編集するとよい。

グラフの描画 : `gen_graph`

「グラフの描画」では、「リファレンスとして使用するコンプライートゲノムもしくはコンティグ (`fasta` フォーマット)」、「`alignment` で作成した `graph_config.csv` ファイル」、「`get_target` で作成した `target.bed` ファイル」、「`get_target` で作成した `cds.gff` ファイル」を元に、グラフの描画を行う。実行コマンドは次のとおりである。

```
$ mmviewer gen_graph -c reference.fasta -o output/directory  
-a path/to/graph_config.csv -b path/to/target.bed -d path/to/cds.gff -p 5
```

`-c` リファレンスとして使用するコンプライートゲノムもしくはコンティグ (`fasta` フォーマット) のファイルパスを指定する (required)

`-o` 出力先のディレクトリのパスを指定する (required)

`-a` `mmviewer alignment` で作成した `graph_config.csv` ファイルのパスを指定する (required)

`-b` `mmviewer get_target` で作成した `target.bed` ファイルのパスを指定する

`-p` 変異を検出する最小 `depth` を指定する

上記のコマンドを実行すると、`-o`で指定したディレクトリ内に `missense_mut_graph` ディレクトリが生成される。すでに `missense_mut_graph` ディレクトリが存在する場合は `missense_mut_graph_1` が生成され、`missense_mut_graph_1` が存在する場合は `missense_mut_graph_2` が生成される。`missense_mut_graph` ディレクトリ内には、`annotated_vcf` ディレクトリ、`depth` ディレクトリ、`filtered_snps` ディレクトリ、`raw_snps` ディレクトリ、`raw_vcf` ディレクトリ、`ref` ディレクトリ、`snpEff.config` ファイル、`legends.svg` ファイルが生成される。`snpEff.config` ファイルは、SNPs のアノテーションに使用した `snpEff` ソフトウェアの解析に必要としたファイルである。`depth` ディレクトリには、アラインメントの結果得られた `bam` ファイルから、リファレンスゲノム上にアラインメントされたリードの `depth` を計算して出力された結果が記録されている。`raw_vcf` ディレクトリには、`bam` ファイルから変異をリストアップして `vcf` ファイルにまとめた結果が記録されて

いる。refディレクトリに含まれるファイルと snpEff.config ファイルは、snpEffを用いて変異のアノテーションを行う際に必要としたファイルである。annotated_vcfディレクトリに含まれているファイルにはアノテーション済みの変異データが記録されている。生成されたグラフはsvgディレクトリにsvg形式で保存される。グラフの外観は図S1のようになる。マーカーの形状で変異の種類、矢印で対象のCDS領域およびその翻訳方向、グレーの網掛けでアラインメントされなかった領域を表す。

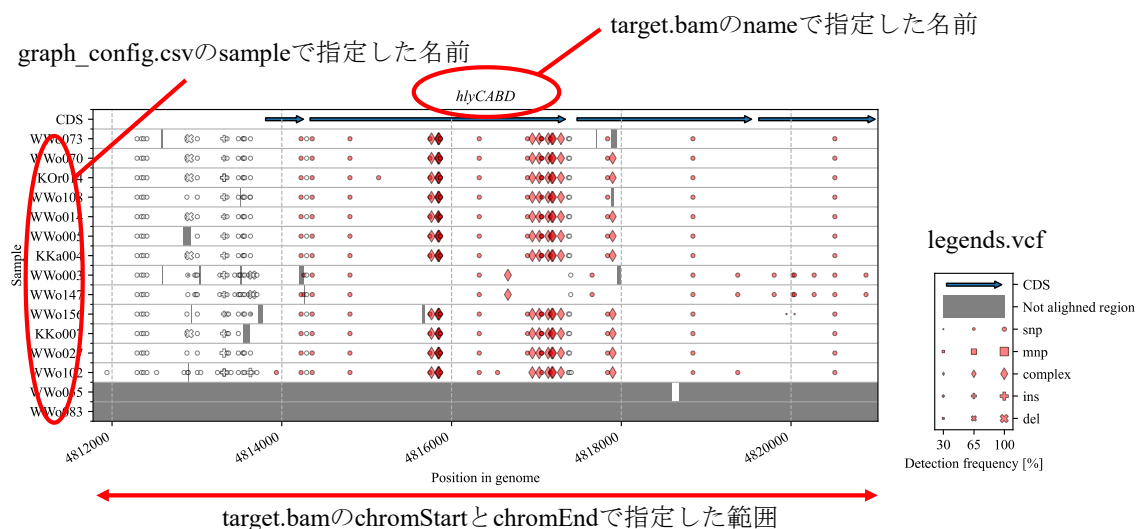


図 S 1 MMViewer で出力されるグラフの概観

ソースコード

MMViewerは7つのソースコードファイル `__init__.py`、`__main__.py`、`__version__.py`、`alignment.py`、`gen_graph.py`、`get_target.py`、`utility.py` からなる。それぞれのソースコードは次のとおりである。

`__init__.py`

```
#!/usr/bin/env python3
from __version__ import __version__

__copyright__ = 'Copyright (C) 2022 Kengo Ikebata'
__license__ = 'GPLv3'
__author__ = 'Kengo Ikebata'
__author_email__ = 'kikebata623@gmail.com'
__url__ = 'https://github.com/KIkebata/mmviewer'
```

`__main__.py`

```
#!/usr/bin/env python3
from .get_target import get_target_position
from .alignment import make_alignment
from .utility import get_option
from .gen_graph import gen_graph
from __version__ import __version__

def main():
    arg_option = get_option()
    if arg_option.program_name=='get_target':
        get_target_position(
            ref_fa=arg_option.complete_seq,
            out_dir=arg_option.output,
            gene_seq_type=arg_option.gene_seq_type,
            gene_seq=arg_option.gene_sequence,
            f_interval=arg_option.upper_interval,
            r_interval=arg_option.lower_interval
        )
    elif arg_option.program_name=='alignment':
        make_alignment(
            al_config_csv=arg_option.alignment_config_file,
            ref_fa=arg_option.complete_seq,
            out_dir=arg_option.output
        )
    elif arg_option.program_name=='gen_graph':
        gen_graph(
            out_dir=arg_option.output,
            graph_config_csv=arg_option.graph_config,
            target_bed=arg_option.target_bed,
            ref_fa=arg_option.complete_seq,
            cds_gff=arg_option.cds_gff,
            min_depth=arg_option.min_depth
        )

if __name__ == '__main__':
    main()
```

```

__version__.py
#!/usr/bin/env python3

__version_info__ = (0, 0, 9)
__version__ = '.'.join(map(str, __version_info__))

get_target.py
#!/usr/bin/env python3
import subprocess, os, shutil, glob, re
import pandas as pd
from .utility import make_directory

def prodigal_py(in_ref_fa, out_dir):
    """Predict CDS of fasta file

    :param in_ref_fa: A fasta file as input
    :type in_ref_fa: str
    :param out_dir: Path to output directory
    :type out_dir: str
    :return: output gff file path, predicted cds (nucl.) fa file path, predicted cds (prot.) fa
    file path
    :rtype: (str, str, str)
    """
    gff_file = os.path.join(out_dir, 'cds.gff')
    nucl_fa_file = os.path.join(out_dir, 'cds_nucl.fa')
    prot_fa_file = os.path.join(out_dir, 'cds_prot.fa')
    subprocess.run(' '.join([
        'prodigal', '-i', in_ref_fa,
        '-o', gff_file,
        '-f', 'gff',
        '-d', nucl_fa_file,
        '-a', prot_fa_file
    ]), shell=True)
    return gff_file, nucl_fa_file, prot_fa_file

def makeblastdb_py(seq_type, ref_fa, out_dir):
    """Run makeblastdb

    :param seq_type: ['prot' | 'nucl']
    :type seq_type: str
    :param ref_fa: Path to reference.fasta
    :type ref_fa: str
    :param out_dir: Path to output directory
    :type out_dir: str
    :return: File path to blast database
    :rtype: str
    """
    blastdb_dir = os.path.join(out_dir, 'blastdb')
    os.mkdir(blastdb_dir)
    shutil.copy2(src=ref_fa, dst=blastdb_dir)
    blastdb_path = glob.glob(os.path.join(blastdb_dir, '*'))[0]
    subprocess.run(' '.join([
        'makeblastdb', '-in', blastdb_path, '-dbtype', seq_type
    ]), shell=True)
    return blastdb_path

def blast_py(query_fa, seq_type, blastdb_path, out_dir, anno_fmt):
    """Run blast
    If seq_type is 'nucl', run BLASTN. If seq_type is 'prot', run BLASTX.

    :param query_fa: Path to cds.fasta

```

```

:type query_fa: str
:param seq_type: ['nucl' | 'prot']
:type seq_type: str
:param blastdb_path: File path to blast database
:type blastdb_path: str
:param anno_fmt: result format
:type anno_fmt: list
:param out_dir: Path to output directory
:type out_dir: str
:return: Path to blast output
:rtype: str
"""
if seq_type == 'prot':
    blast_out = os.path.join(out_dir, 'blastx_out.tsv')
elif seq_type == 'nucl':
    blast_out = os.path.join(out_dir, 'blastn_out.tsv')
else:
    return None
out_fmt = "%6 '+' '.join(anno_fmt) + '"
if seq_type == 'prot':
    subprocess.run(' '.join([
        'blastx', '-query', query_fa,
        '-db', blastdb_path,
        '-out', blast_out,
        '-outfmt', out_fmt,
        '-evalue', '1e-5',
        '-num_threads', str(os.cpu_count())
    ]), shell=True)
elif seq_type == 'nucl':
    subprocess.run(' '.join([
        'blastn', '-query', query_fa,
        '-db', blastdb_path,
        '-out', blast_out,
        '-outfmt', out_fmt,
        '-evalue', '1e-5',
        '-num_threads', str(os.cpu_count())
    ]), shell=True)
return blast_out

def get_bed_from_blast(gff_file, nucl_fa_file, blast_out, anno_fmt, out_dir, f_interval=0,
r_interval=0):
    """Make bed file from gff file from prodigal, cds_nucl.fa from prodigal, and blast result

    :param gff_file: gff file path made by prodigal
    :type gff_file: str
    :param nucl_fa_file: cds_nucl.fa file path made by prodigal
    :type nucl_fa_file: str
    :param blast_out: blast output file path
    :type blast_out: str
    :param anno_fmt: blast annotation list
    :type anno_fmt: list
    :param out_dir: Path to output directory
    :type out_dir: str
    :param f_interval: upper region you want to analyze than objective cds , defaults to 0
    :type f_interval: int
    :param r_interval: lower region you want to analyze than objective cds, defaults to 0
    :type r_interval: int
    """
    out_bed = os.path.join(out_dir, 'target.bed')
    gff_fmt = ['seqname', 'source', 'feature', 'start', 'end', 'score', 'strand', 'frame',
'attribute']
    gff_df = pd.read_table(gff_file, names=gff_fmt, comment='#')

```

```

tmp_list = [re.sub(r'\;$', '', re.search(r'ID\=.+?\;?', ll).group()) for ll in
gff_df['attribute']]
gff_df2 = pd.concat([gff_df, pd.DataFrame({'ID':tmp_list})], axis=1)
with open(nucl_fa_file, 'r') as f:
    lines = f.read().splitlines()
tmp_list = [re.sub(r'^\>', '', ll) for ll in lines if ll.startswith('>')]
seqid_list = [ll.split(' # ')[0] for ll in tmp_list]
id_list = [re.sub(r'\;$', '', re.search(r'ID\=.+?\;?', ll).group()) for ll in tmp_list]
nucl_head_df = pd.DataFrame({'qseqid':seqid_list, 'ID':id_list})
gff_df3 = pd.merge(left=gff_df2, right=nucl_head_df, on='ID', how='left')
blast_out_df = pd.read_table(blast_out, names=anno_fmt)
blast_out_df2 = pd.merge(left=blast_out_df, right=gff_df3, on='qseqid', how='left')
bed_score = (10 * blast_out_df2['pident'] * blast_out_df2['length'] / \
(blast_out_df2['slen'] + blast_out_df2['length'] - (blast_out_df2['send'] -
blast_out_df2['sstart'] + 1))).astype(int)
c_start, c_end = [], []
for j in range(len(blast_out_df2)):
    if blast_out_df2['strand'][j] == '+':
        c_start += [blast_out_df2['start'][j] - 1 - f_interval]
        c_end += [blast_out_df2['end'][j] + r_interval]
    elif blast_out_df2['strand'][j] == '-':
        c_start += [blast_out_df2['start'][j] - 1 - r_interval]
        c_end += [blast_out_df2['end'][j] + f_interval]
    if c_start[j] < 0:
        c_start[j] = 0
    # consider end if possible
bed_df = pd.DataFrame({
    'chrom':blast_out_df2['seqname'],
    'chromStart':c_start,
    'chromEnd':c_end,
    'name':[re.sub(r'[\|/|:|?|"|<|>|\||\s|\|=]', '_', ll) for ll in
blast_out_df2['stitle']],
    'score':bed_score,
    'strand':blast_out_df2['strand'],
    'CDS_Start':blast_out_df2['start'] - 1,
    'CDS_End':blast_out_df2['end']
})
bed_df.to_csv(out_bed, sep='\t', index=False, header=False)
with open(out_bed) as f:
    s = f.read()
s = '# ' + '\t'.join(bed_df.columns.values.tolist()) + '\n' + s
with open(out_bed, 'w') as f:
    f.write(s)
return out_bed

def add_fasta_entry(ref_fa, gff_file):
    """add fasta entry at the last of gff_file

    :param ref_fa: path to reference.fasta
    :type ref_fa: str
    :param gff_file: gff file path made by prodigal
    :type gff_file: str
    """
    with open(gff_file, 'a') as f:
        with open(ref_fa) as ref_f:
            f.write('###\n##FASTA\n')
            f.write(ref_f.read())

def get_target_position(ref_fa, out_dir, gene_seq_type, gene_seq, f_interval=0, r_interval=0):
    """Generate target.bed file to indicate target region in genome

    :param ref_fa: Path to a file of complete genome or contig in fasta format as a reference

```

```

:type ref_fa: str
:param out_dir: Path to output directory
:type out_dir: str
:param gene_seq_type: ['nucl' | 'prot'] : a type of gene_seq file (nucleotide or animo acid)
:type gene_seq_type: str
:param gene_seq: Path to a file of target gene in fasta format
:type gene_seq: str
:param f_interval: upper region you want to analyze from cds (corresponding to gene_seq),
defaults to 0
:type f_interval: int
:param r_interval: lower region you want to analyze from cds (corresponding to gene_seq),
defaults to 0
:type r_interval: int
"""
print_sep = '#' * 40
out_dir = make_directory(dir_name='target', out_dir=out_dir)
anno_fmt=['qseqid', 'sseqid', 'stitle', 'evaluate', 'slen', 'sstart', 'send', 'qlen',
'qstart', 'qend', 'pident', 'nident', 'mismatch', 'length', 'qseq']
print(print_sep)
print('Running prodigal.....')
gff_file, ncds_file, pcds_file = prodigal_py(
    in_ref_fa=ref_fa,
    out_dir=out_dir
)
print(print_sep)
print('Running makeblastdb.....')
blastdb_path = makeblastdb_py(
    seq_type=gene_seq_type,
    ref_fa=gene_seq,
    out_dir=out_dir
)
print(print_sep)
print('Running BLAST program.....')
blast_out = blast_py(
    query_fa=ncds_file,
    seq_type=gene_seq_type,
    blastdb_path=blastdb_path,
    out_dir=out_dir,
    anno_fmt=anno_fmt
)
print(print_sep)
print('Converting files into bed format.....')
out_bed = get_bed_from_blast(
    gff_file=gff_file,
    nucl_fa_file=ncds_file,
    blast_out=blast_out,
    anno_fmt=anno_fmt,
    out_dir=out_dir,
    f_interval=f_interval,
    r_interval=r_interval
)
add_fasta_entry(
    ref_fa=ref_fa,
    gff_file=gff_file
)
print(print_sep)
print('Finish.')

```

alignment.py

```
#!/usr/bin/env python3
import subprocess, os, shutil, glob
import pandas as pd
import sys
from .utility import make_directory

def bwa_mem_py(ref_fa, sample_names, out_dir, in_fq1_list, in_fq2_list=None):
    """Run bwa mem to generate sorted bam.

    :param ref_fa: Path to input reference file in fasta format.
    :type ref_fa: str
    :param sample_names: Sample name list (same length with in_fq1_list).
    :type sample_names: list
    :param out_dir: Path to output directory.
    :type out_dir: str
    :param in_fq1_list: List of trimmed reads file path (fq format or fq.gz format)
    :type in_fq1_list: list
    :param in_fq2_list: List of trimmed reads file path for paired end reads, defaults to None
    (optional)
    :type in_fq2_list: list, None
    :return: Output sorted bam file list.
    :rtype: list
    """
    if in_fq2_list is None:
        fq_list = in_fq1_list
    else:
        fq_list = [fq1 + ' ' + fq2 for fq1, fq2 in zip(in_fq1_list, in_fq2_list)]
    output_bam_dir = os.path.join(out_dir, 'bam')
    os.mkdir(output_bam_dir)
    bam_files = [os.path.join(output_bam_dir, ll + '.bam') for ll in sample_names]
    subprocess.run(' '.join([
        'bwa', 'index', ref_fa
    ]), shell=True)
    for i in range(len(sample_names)):
        subprocess.run(' '.join([
            'bwa', 'mem', '-t', str(os.cpu_count()), ref_fa, fq_list[i], '|',
            'samtools', 'view', '-@', str(os.cpu_count()), '-bS', '|',
            'samtools', 'fixmate', '-m', '-@', str(os.cpu_count()), '-', '-', '|',
            'samtools', 'sort', '-@', str(os.cpu_count()), '|',
            'samtools', 'markdup', '-@', str(os.cpu_count()), '-r', '-', bam_files[i]
        ]), shell=True)
        subprocess.run(' '.join([
            'samtools', 'index', bam_files[i]
        ]), shell=True)
    return bam_files

def make_alignment(al_config_csv, ref_fa, out_dir):
    """Make alignment by using bwa mem
    Before running this function, prepare alignment_config.csv file.
    Alignment_config.csv file consists of 2 or 3 columns with a header in first row.
    1st column is sample name.
    2nd column is path to trimmed read in fastq format or its compressed format (.gz)
    3rd column is path to trimmed read in fastq format or its compressed format (.gz) of reverse
    reads if paired end read (option)

    :param al_config_csv: path to alignment_config.csv
    :type al_config_csv: str
    :param ref_fa: path to a reference sequence in fasta format
    :type ref_fa: str
    :param out_dir: path to output directory
```

```

:type out_dir: str
"""
print_sep = '#' * 40
out_dir = make_directory(dir_name='alignment', out_dir=out_dir)
ref_fa_dir = os.path.join(out_dir, 'reference')
os.mkdir(ref_fa_dir)
shutil.copy2(src=ref_fa, dst=ref_fa_dir)
ref_fa = glob.glob(os.path.join(ref_fa_dir, '*'))[0]
al_config_df = pd.read_csv(al_config_csv)
if len(al_config_df.columns) == 2:
    in_fq2_list = None
elif len(al_config_df.columns) == 3:
    in_fq2_list = al_config_df.iloc[:, 2].values.tolist()
else:
    print()
    sys.exit('Error! Invalid alignment config file.')
sample_names = al_config_df.iloc[:, 0].values.tolist()
in_fq1_list = al_config_df.iloc[:, 1].values.tolist()
print(print_sep)
print('Running bwa mem.....')
bam_files = bwa_mem_py(
    ref_fa=ref_fa,
    sample_names=sample_names,
    out_dir=out_dir,
    in_fq1_list=in_fq1_list,
    in_fq2_list=in_fq2_list
)
pd.DataFrame({'sample': sample_names, 'bam':bam_files}).to_csv(os.path.join(out_dir,
'graph_config.csv', ), index=False)
print(print_sep)
print('Finish.')

```

gen_graph.py

```

#!/usr/bin/env python3
import re, subprocess, os, shutil
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.patches as patches
from matplotlib.font_manager import FontProperties
from joblib import Parallel, delayed
from .utility import make_directory, make_dir_files
import vcf
from Bio import SeqIO

def mpileup(bam_files, sample_list, ref_fa, out_dir, bq_threshold=0):
    """Run mpileup to generate vcf files by using bcftools

    :param bam_files: list of bam files which are fixmate, sort, and markdup processed.
    :type bam_files: list
    :param sample_list: sample name list (same length with bam_files).
    :type sample_list: list
    :param ref_fa: path to input reference file in fasta format.
    :type ref_fa: str
    :param out_dir: path to output directory
    :type out_dir: str
    :return: output raw_vcf files list.
    :rtype: list
    """
    raw_vcf_out_dir = os.path.join(out_dir, 'raw_vcf')
    os.mkdir(raw_vcf_out_dir)
    raw_vcf_files = [os.path.join(raw_vcf_out_dir, ll + '.vcf') for ll in sample_list]

```

```

def process(bam_file, vcf_file):
    subprocess.run(' '.join([
        'bcftools', 'mpileup', '--min-BQ', str(bq_threshold), '-Ou', '-a',
        'FORMAT/AD,INFO/AD','-f', ref_fa, bam_file, '|',
        'bcftools', 'call', '-mv', '-Ov', '-o', vcf_file
    ]), shell=True)
    Parallel(n_jobs=-1, verbose=3)([delayed(process)(bam_file, vcf_file) for bam_file, vcf_file
in zip(bam_files, raw_vcf_files)])
    return raw_vcf_files

def snpEff_py(out_dir, cds_gff, raw_vcf_files, sample_list):
    """annotate mutation using snpEff

    :param out_dir: path to output directory
    :type out_dir: str
    :param cds_gff: path to gff file to indicate translated region of reference strain (e.g.
    calculated by prodigal) and must include fasta sequence.
    :type cds_gff: str
    :param raw_vcf_files: list of path to raw_vcf_files calculated by mpileup()
    :type raw_vcf_files: list
    :param sample_list: sample name list
    :type sample_list: list
    :return: list of vcf files
    :rtype: list
    """
    snps_ref_dir = os.path.join(out_dir, 'ref')
    os.mkdir(snps_ref_dir)
    shutil.copy2(
        src=cds_gff,
        dst=os.path.join(snps_ref_dir, 'genes.gff')
    )
    conf_file = os.path.join(out_dir, 'snpEff.config')
    with open(conf_file, 'w') as f:
        f.write('ref.genome : ref')
    subprocess.run(' '.join([
        'snpEff', 'build', '-gff3', '-c', conf_file,
        '-dataDir', out_dir, 'ref'
    ]), shell=True)
    vcf_dir, vcf_files = make_dir_files(dir_name='annotated_vcf', out_dir=out_dir,
    prefix_list=sample_list, ext='.vcf')
    def process(raw_vcf_file, vcf_file):
        subprocess.run(' '.join([
            'snpEff', 'ann', '-c', conf_file, '-dataDir', out_dir,
            '-no-downstream', '-no-upstream', '-no-utr',
            '-i', 'vcf', 'ref', raw_vcf_file, '>', vcf_file
        ]), shell=True)
    Parallel(n_jobs=-1, verbose=3)([delayed(process)(raw_vcf_file, vcf_file) for raw_vcf_file,
    vcf_file in zip(raw_vcf_files, vcf_files)])
    return vcf_files

def judge_mut_type(ref, alt):
    """type mutation

    Type   Name   Example;
    snp    Single Nucleotide Polymorphism   A => T
    mnp    Multiple Nuclotide Polymorphism   GC => AT
    ins    Insertion   ATT => AGTT
    del    Deletion   ACGG => ACG
    complex    Combination of snp/mnp   ATTC => GTTA

    :param ref: reference base
    :type ref: str

```



```

:param alt: alternated base
:type alt: str
:return: snp: mutation type
:rtype: str
"""
if len(ref) == len(alt):
    if len(ref) == 1:
        ans = 'snp'
    else:
        diff_all = sum([aa != bb for aa, bb in zip(ref, alt)])
        if diff_all:
            ans = 'complex'
        else:
            ans = 'mnp'
else:
    if len(ref) - len(alt) > 0:
        ans = 'del'
    else:
        ans = 'ins'
return ans

def combine_snps(vcf_df, ref_fa):
    """joint snps to mnp if they are in next amino acid in the same protein, and classify
    mutation type

    :param vcf_df: dataframe for vcf file (for csv file)
    :type vcf_df: pd.core.frame.DataFrame
    :param ref_fa: path to a file of complete genome or contig in fasta format as a reference
    :type ref_fa: str
    :return: newly annotated dataframe for vcf file (for csv file)
    :rtype: pd.core.frame.DataFrame
    """
    seq_list = [record.seq for record in SeqIO.parse(ref_fa, 'fasta')]
    prot_pos_list = [[int(jj) for jj in re.findall(r'\d+', ll)] for ll in vcf_df.PROT_EFFECT]
    n_chrom_list, n_pos_list, n_ftype_list, n_ref_list, n_alt_list = [vcf_df.CHROM[0]],
    [vcf_df.POS[0]], [[vcf_df.FTYPE[0]], [str(vcf_df.REF[0])], [str(vcf_df.ALT[0])]]
    n_nucl_effect_list, n_prot_effect_list, n_depth_list, n_frequency_list =
    [str(vcf_df.NUCL_EFFECT[0])], [str(vcf_df.PROT_EFFECT[0])], [vcf_df.DEPTH[0]],
    [vcf_df.FREQUENCY[0]]
    n_type_list = [mut_type_select(REF=n_ref_list[0], ALT=n_alt_list[0])]
    for i in range(1, len(vcf_df)):
        if (vcf_df.CHROM[i - 1] == vcf_df.CHROM[i]) & (vcf_df.POS[i] - vcf_df.POS[i - 1] < 30) & \
            (len(prot_pos_list[i - 1]) > 0) & (len(prot_pos_list[i]) > 0):
            n_type = mut_type_select(REF=str(vcf_df.REF[i]), ALT=str(vcf_df.ALT[i]))
            if (difference_pos(prot_pos_list[i - 1], prot_pos_list[i]) == 1) & \
                (n_type_list[len(n_type_list) - 1] not in ['ins', 'del']) & \
                (n_type not in ['ins', 'del']):
                gap_seq = str(seq_list[0][(vcf_df.POS[i] - 1):(vcf_df.POS[i] - 1 +
                len(n_ref_list[len(n_ref_list) - 1])]))
                n_ref_list[len(n_ref_list) - 1] += gap_seq + str(vcf_df.REF[i])
                n_alt_list[len(n_alt_list) - 1] += gap_seq + str(vcf_df.ALT[i])
                n_nucl_effect_list[len(n_nucl_effect_list) - 1] += ';' +
                str(vcf_df.NUCL_EFFECT[i])
                n_prot_effect_list[len(n_prot_effect_list) - 1] += ';' +
                str(vcf_df.PROT_EFFECT[i])
                n_depth_list[len(n_depth_list) - 1] = min(n_depth_list[len(n_depth_list) - 1],
                vcf_df.DEPTH[i])
                n_frequency_list[len(n_frequency_list) - 1] =
                min(n_frequency_list[len(n_frequency_list) - 1], vcf_df.FREQUENCY[i])
                n_ftype_list[len(n_ftype_list) - 1] = n_ftype_list[len(n_ftype_list) - 1] +
                [vcf_df.FTYPE[i]]

```

```

        n_type_list[len(n_type_list) - 1] =
mut_type_select(REF=n_ref_list[len(n_ref_list) - 1], ALT=n_alt_list[len(n_alt_list) - 1])
    else:
        n_chrom_list += [vcf_df.CHROM[i]]
        n_pos_list += [vcf_df.POS[i]]
        n_fctype_list += [[vcf_df.FTYPE[i]]]
        n_ref_list += [str(vcf_df.REF[i])]
        n_alt_list += [str(vcf_df.ALT[i])]
        n_nucl_effect_list += [vcf_df.NUCL_EFFECT[i]]
        n_prot_effect_list += [vcf_df.PROT_EFFECT[i]]
        n_depth_list += [vcf_df.DEPTH[i]]
        n_frequency_list += [vcf_df.FREQUENCY[i]]
        n_type_list += [mut_type_select(REF=str(vcf_df.REF[i]), ALT=str(vcf_df.ALT[i]))]
    else:
        n_chrom_list += [vcf_df.CHROM[i]]
        n_pos_list += [vcf_df.POS[i]]
        n_fctype_list += [[vcf_df.FTYPE[i]]]
        n_ref_list += [str(vcf_df.REF[i])]
        n_alt_list += [str(vcf_df.ALT[i])]
        n_nucl_effect_list += [vcf_df.NUCL_EFFECT[i]]
        n_prot_effect_list += [vcf_df.PROT_EFFECT[i]]
        n_depth_list += [vcf_df.DEPTH[i]]
        n_frequency_list += [vcf_df.FREQUENCY[i]]
        n_type_list += [mut_type_select(REF=str(vcf_df.REF[i]), ALT=str(vcf_df.ALT[i]))]
for i in range(len(n_fctype_list)):
    if n_prot_effect_list[i] == '':
        n_fctype_list[i] = 'intergenic_region'
    elif not sum([ll != 'synonymous_variant' for ll in n_fctype_list[i]]):
        n_fctype_list[i] = 'synonymous_variant'
    else:
        n_fctype_list[i] = 'missense_variant'
new_vcf_df = pd.DataFrame({
    'CHROM':n_chrom_list,
    'POS':n_pos_list,
    'TYPE':n_type_list,
    'FTYPE':n_fctype_list,
    'REF':n_ref_list,
    'ALT':n_alt_list,
    'NUCL_EFFECT':n_nucl_effect_list,
    'PROT_EFFECT':n_prot_effect_list,
    'DEPTH':n_depth_list,
    'FREQUENCY':n_frequency_list
})
return new_vcf_df

def mut_type_select(REF, ALT):
    """Classify mutation type

    snp: one base variant. (e.g. C -> G)
    mnp: multi bases variant. (e.g. CA -> GT)
    complex: combination of snps and mnp with a short gap. (e.g. ATGC -> CAGC)
    ins: insertion. (e.g. CG -> CAG)
    del: deletion. (e.g. AT -> A)

    :param REF: reference sequence
    :type REF: str
    :param ALT: alternated sequece
    :type ALT: str
    :return: ['snp', 'mnp', 'complex', 'ins', 'del']
    :rtype: [type]
    """
    if (len(REF) == 1) & (len(ALT) == 1) & (REF != ALT):

```

```

        mut_type = 'snp'
    elif len(REF) > len(ALT):
        mut_type = 'del'
    elif len(REF) < len(ALT):
        mut_type = 'ins'
    else:
        if sum([aa == bb for aa, bb in zip(REF, ALT)]) == 0:
            mut_type = 'mnp'
        else:
            mut_type = 'complex'
    return mut_type

def difference_pos(a_list, b_list):
    return min(abs(max(a_list) - min(b_list)), abs(max(b_list) - min(a_list)))

def gen_snps_csv(vcf_files, sample_list, out_dir, min_depth, min_freq, ref_fa):
    """Convert vcf files into csv files and filter them.

    :param vcf_files: list of vcf files which are annotated by snpEff
    :type vcf_files: list
    :param sample_list: sample name list (same length with vcf_files)
    :type sample_list: list
    :param out_dir: path to output directory
    :type out_dir: str
    :param min_depth: SNPs filtering parameter: minimum depth
    :type min_depth: int
    :param min_freq: SNPs filtering parameter: minimum frequency of the SNPs (0.0 - 1.0)
    :type min_freq: float
    :return: raw SNPs.csv file list, filtered SNPs.csv file list
    :rtype: (list, list)
    """
    raw_snps_csv_dir, raw_snps_csv_files = \
        make_dir_files(dir_name='raw_snps', out_dir=out_dir, prefix_list=sample_list,
ext='.csv')
    filt_snps_csv_dir, filt_snps_csv_files = \
        make_dir_files(dir_name='filtered_snps', out_dir=out_dir, prefix_list=sample_list,
ext='.csv')
    def process(vcf_file, raw_snps_csv_file, filt_snps_csv_file):
        vcf_gen = vcf.Reader(filename=vcf_file)
        vcf_list = list(vcf_gen)
        chrom_list = [ll.CHROM for ll in vcf_list]
        pos_list = [ll.POS for ll in vcf_list]
        dp_list = [ll.INFO['DP'] for ll in vcf_list]
        ad_list = [ll.INFO['AD'] for ll in vcf_list]
        ref_list = [ll.REF for ll in vcf_list]
        alt_list = [ll.ALT for ll in vcf_list]
        ann_list = [ll.INFO['ANN'] if 'ANN' in ll.INFO.keys() else '' for ll in vcf_list]
        alt_list2 = [alt[ad[1:].index(max(ad[1:]))] for ad, alt in zip(ad_list, alt_list)]
        freq_list = [max(ad[1:]) / dp for ad, dp in zip(ad_list, dp_list)]
        type_list = [judge_mut_type(ref=ref, alt=alt) for ref, alt in zip(ref_list, alt_list)]
        ftype_list = [ann[ad[1:].index(max(ad[1:]))].split('|')[1] if ann != '' else '' for ann,
ad in zip(ann_list, ad_list)]
        effect_nt_list = [ann[ad[1:].index(max(ad[1:]))].split('|')[9] if ann != '' else '' for
ann, ad in zip(ann_list, ad_list)]
        effect_pr_list = [ann[ad[1:].index(max(ad[1:]))].split('|')[10] if ann != '' else '' for
ann, ad in zip(ann_list, ad_list)]
        vcf_df = pd.DataFrame({
            'CHROM':chrom_list,
            'POS':pos_list,
            'TYPE':type_list,
            'FTYPE':ftype_list,
            'REF':ref_list,

```

```

        'ALT':alt_list2,
        'NUCL_EFFECT':effect_nt_list,
        'PROT_EFFECT':effect_pr_list,
        'DEPTH':dp_list,
        'FREQUENCY':freq_list
    })
    vcf_df.to_csv(raw_snps_csv_file, index=False)
    filt_vcf_df = vcf_df[(vcf_df['DEPTH'] >= min_depth) & (vcf_df['FREQUENCY'] >=
min_freq)].reset_index(drop=True)
    filt_vcf_df = combine_snps(vcf_df=filt_vcf_df, ref_fa=ref_fa)
    filt_vcf_df.to_csv(filt_snps_csv_file, index=False)
    Parallel(n_jobs=-1, verbose=3)([delayed(process)(vcf_file, raw_snps_csv_file,
filt_snps_csv_file)\
    for vcf_file, raw_snps_csv_file, filt_snps_csv_file\
    in zip(vcf_files, raw_snps_csv_files, filt_snps_csv_files)])
    return raw_snps_csv_files, filt_snps_csv_files

def calc_depth(out_dir, sample_list, bam_files):
    """calculate mapping depth from bam files

    :param out_dir: path to output directory
    :type out_dir: str
    :param sample_list: sample name list (same length with bam_files)
    :type sample_list: list
    :param bam_files: list of bam files path
    :type bam_files: list
    :return: list of depth files (tab-separated)
    :rtype: list
    """
    depth_dir, depth_files = make_dir_files(dir_name='depth', out_dir=out_dir,
prefix_list=sample_list, ext='.txt')
    def process(bam_file, depth_file):
        subprocess.run(' '.join([
            'samtools', 'depth', bam_file, '>', depth_file
        ]), shell=True)
    Parallel(n_jobs=-1, verbose=3)([delayed(process)(bam_file, depth_file)\
    for bam_file, depth_file in zip(bam_files, depth_files)])
    return depth_files

def get_not_aligned_area(depth_files, left_pos, right_pos, min_depth=1):
    """get list of not aligned region

    :param depth_files: list of depth files calculated by calc_depth
    :type depth_files: list
    :param left_pos: [description]
    :type left_pos: [type]
    :param right_pos: [description]
    :type right_pos: [type]
    :param min_depth: [description], defaults to 1
    :type min_depth: int, optional
    """
    def process(depth_file):
        depth_df = pd.read_table(depth_file, header=None, names=['CHROM', 'POS', 'DEPTH'])
        depth_df[(depth_df['POS'] >= left_pos) & depth_df['POS'] <=
right_pos].reset_index(drop=True)
        no_depth_pos_list = []
        tmp_left, tmp_right = 0, 0
        for j in range(left_pos, right_pos+1):
            if (len(depth_df[depth_df['POS'] == j]) == 0)\
            | len(depth_df[(depth_df['POS'] == j) & (depth_df['DEPTH'] < min_depth)]) == 1:
                if j - tmp_right > 1:
                    tmp_left = j

```

```

        tmp_right = j
        no_depth_pos_list += [[tmp_left, tmp_right]]
    else:
        tmp_right = j
        no_depth_pos_list[len(no_depth_pos_list) - 1][1] = tmp_right
    no_depth_pos_list2 = []
    for j in range(len(no_depth_pos_list)):
        no_depth_pos_list2 += [tuple(no_depth_pos_list[j])]
    return no_depth_pos_list2
print('getting not aligned region.....')
no_depth_list = Parallel(n_jobs=-1, verbose=3)([delayed(process)(depth_file) for depth_file
in depth_files])
return no_depth_list

def get_target_snps(filt_snps_csv_file, cds_regions, start_pos, end_pos):
    """filter snps

    In CDS region; rows with 'synonymous_variant' in 'FTYPE' column are removed
    Out of CDS region; values in 'FTYPE' column converted to 'intergenic_region'

    :param filt_snps_csv_file: path to filtered_snps file generated from gen_snps_csv()
    :type filt_snps_csv_file: str
    :param cds_regions: cds region [(start - 1, end), (start - 1, end), .....]
    :type cds_regions: list
    :param start_pos: start - 1 (target region)
    :type start_pos: int
    :param end_pos: end (target region)
    :type end_pos: int
    :return: processed filtered_snps data frame
    :rtype: pandas.core.frame.DataFrame
    """
    filt_vcf_df = pd.read_csv(filt_snps_csv_file)
    filt_vcf_df = filt_vcf_df[(filt_vcf_df['POS'] > start_pos) & (filt_vcf_df['POS'] <=
end_pos)].reset_index(drop=True)
    in_cds = [False] * len(filt_vcf_df)
    pos_list = filt_vcf_df['POS'].values.tolist()
    for i in range(len(in_cds)):
        for j in range(len(cds_regions)):
            if (pos_list[i] > cds_regions[j][0]) & (pos_list[i] <= cds_regions[j][1]):
                in_cds[i] = True
            else:
                pass
    # convert ftype to 'nan' out of cds
    filt_vcf_df['FTYPE'] = [ftype if in_cds_v == True else 'nan' for ftype, in_cds_v in
zip(filt_vcf_df['FTYPE'].values.tolist(), in_cds)]
    # remove synonymous_variant in cds
    filt_vcf_df = filt_vcf_df[filt_vcf_df['FTYPE'] !=
'synonymous_variant'].reset_index(drop=True)
    return filt_vcf_df

def ref_dict(input_list, input_dict):
    """input_list を input_dict に対応した値に変換して出力

    :param input_list: 変換したい値の入ったリスト
    :type input_list: list
    :param input_dict: 変換用辞書。{'org':[], 'new':[]}で記述する。
    :type input_dict: dict
    :return: 変換後のリスト
    :rtype: list
    """
    value_list = []
    for i in range(len(input_list)):

```

```

        value_list += [input_dict[input_dict['org']==input_list[i]]['new'].values.tolist()[0]]
    return value_list

def gen_legends(out_dir):
    """generate legends

    :param out_dir: path to output directory
    :type out_dir: str
    :return: output file path
    :rtype: str
    """
    out_file = os.path.join(out_dir, 'legends.svg')
    fig = plt.figure(figsize=(1, 2))
    ax = fig.add_subplot(111)

    # set xlim and ylim
    ax.set_xlim(left=20, right=110)
    ax.set_ylim(bottom=0, top=14)

    # set font properties
    font = FontProperties(family='serif')
    font.set_name('Times New Roman')

    # add ticks
    ax.yaxis.tick_right()
    ax.set_yticks(np.linspace(1, 13, 7))
    ax.set_yticklabels(['del', 'ins', 'complex', 'mnp', 'snp', 'Not aligned region', 'CDS'])
    ax.set_xticks([30, 65, 100])
    ax.set_xticklabels(['30', '65', '100'])
    for x_tick_labels in ax.get_xticklabels():
        x_tick_labels.set_fontproperties(font)
    for y_tick_labels in ax.get_yticklabels():
        y_tick_labels.set_fontproperties(font)

    # show labels
    ax.set_xlabel('Detection frequency [%]', fontproperties=font)

    # show cds arrow
    arrow = patches.FancyArrowPatch((30, 13), (100, 13), mutation_scale=10)
    ax.add_patch(arrow)

    # mark mutations
    shape_list = ['X', 'P', 'd', 's', '.']
    yvalue_list = [1, 3, 5, 7, 9]
    xvalue_list = [30, 65, 100]
    m_size_list = [(6.0 * 0.3) ** 2, (6.0 * 0.65) ** 2, (6.0 * 1.0) ** 2]
    m_col = 'red'
    for i in range(len(yvalue_list)):
        ax.scatter(
            x=xvalue_list,
            y=[yvalue_list[i]] * 3,
            s=m_size_list,
            marker=shape_list[i],
            c=m_col,
            edgecolor = 'k',
            linewidth=0.5,
            alpha=0.5
        )

    # show not aligned region
    rect = patches.Rectangle(xy=(30, 10), width=70, height=2, fc='black', alpha=0.5,
        linewidth=0.)

```

```

ax.add_patch(rect)

fig.savefig(out_file, bbox_inches='tight')
return out_file

def make_graph(sample_list, target_bed, filt_snps_csv_files, depth_files, out_dir, min_depth=5):
    """generate graph

    target_bed
        If the values in name column are same, target regions are combined.

    :param sample_list: sample name list
    :type sample_list: list
    :param target_bed: path to target.bed file
    :type target_bed: str
    :param filt_snps_csv_files: list of path to filtered_snps files generated from
gen_snps_csv()
    :type filt_snps_csv_files: list
    :param depth_files: list of depth files calculated by calc_depth()
    :type depth_files: list
    :param out_dir: output directory
    :type out_dir: str
    :return: [description]
    :rtype: [type]
    """
    y_value_dict = pd.DataFrame({'org':sample_list, 'new':[2 * (len(sample_list) - 1) - 1 for
ll in range(len(sample_list))])})
    m_shape_dict = pd.DataFrame({'org':['snp', 'mnp', 'complex', 'del', 'ins'], 'new':['.', 's',
'd', 'X', 'P']})
    m_col_dict = pd.DataFrame({'org':['nan', 'missense_variant'], 'new':['white', 'red']})
    target_df = pd.read_table(
        target_bed, comment='#',
        names=['chrom', 'chromStart', 'chromEnd', 'name', 'score', 'strand', 'CDS_Start',
'CDS_End']
    )
    unique_name = list(set(target_df['name'].values.tolist()))
    # グラフへの描画
    strain_count = len(sample_list)
    svg_dir, svg_files = make_dir_files(dir_name='svg', out_dir=out_dir,
prefix_list=unique_name, ext='.svg')
    for i in range(len(unique_name)):
        filt_target_df = target_df[target_df['name'] == unique_name[i]].reset_index(drop=True)
        filt_target_df = filt_target_df.sort_values(by='CDS_Start').reset_index(drop=True)
        cds_regions = [(cds_start, cds_end) for cds_start, cds_end in
zip(filt_target_df['CDS_Start'].values.tolist(), filt_target_df['CDS_End'].values.tolist())]
        start_pos = min(filt_target_df['chromStart'].values.tolist())
        end_pos = max(filt_target_df['chromEnd'].values.tolist())
        fig = plt.figure(figsize=(10, 4))
        ax = fig.add_subplot(111)

        # グラフ領域の指定
        ax.set_xlim(
            left=min(filt_target_df['chromStart'].values.tolist()),
            right=max(filt_target_df['chromEnd'].values.tolist())
        )
        ax.set_ylim(bottom=0, top=strain_count*2+2)

        # font 設定
        font = FontProperties(family='serif')
        font.set_name('Times New Roman')
        title_font = FontProperties(family='serif')
        title_font.set_name('Times New Roman')

```

```

title_font.set_style('italic')
title_font.set_size(11)

# タイトルの追加
ax.set_title(unique_name[i], fontproperties=title_font)

# 目盛りの設定
ax.set_yticks(np.linspace(1, strain_count * 2 + 1, int(strain_count + 1)))
ax.set_yticklabels(sample_list[::-1] + ['CDS'])
ax.set_yticks(np.linspace(0, strain_count * 2 + 2, int(strain_count * 2 + 3)),
minor=True)
ax.get_xaxis().get_major_formatter().set_scientific(False)
ax.get_xaxis().get_major_formatter().set_useOffset(False)
for x_tick_labels in ax.get_xticklabels():
    x_tick_labels.set_rotation(30)
    x_tick_labels.set_fontproperties(font)
    x_tick_labels.set_horizontalalignment('right')
for y_tick_labels in ax.get_yticklabels():
    y_tick_labels.set_fontproperties(font)

# グリッドの表示
ax.grid(b=True, which='major', axis='x', linestyle='--')
ax.grid(b=True, which='minor', axis='y')

# ラベルの表示
ax.set_xlabel('Position in genome', fontproperties=font)
ax.set_ylabel('Sample', fontproperties=font)

# cds 領域の描画
arrow_y = strain_count * 2 + 1
for j in range(len(filt_target_df)):
    if filt_target_df['strand'][j] == '+':
        arrow = patches.FancyArrowPatch(
            (filt_target_df['CDS_Start'][j], arrow_y),
            (filt_target_df['CDS_End'][j], arrow_y),
            mutation_scale=10
        )
    elif filt_target_df['strand'][j] == '-':
        arrow = patches.FancyArrowPatch(
            (filt_target_df['CDS_End'][j], arrow_y),
            (filt_target_df['CDS_Start'][j], arrow_y),
            mutation_scale=10
        )
    ax.add_patch(arrow)
# mutation の描画
for j in range(strain_count):
    filt_vcf_df = get_target_snps(
        filt_snps_csv_file=filt_snps_csv_files[j],
        cds_regions=cds_regions,
        start_pos=start_pos,
        end_pos=end_pos
    )
    for k in range(len(filt_vcf_df)):
        y_value = ref_dict([sample_list[j]], y_value_dict)
        x_value = [filt_vcf_df['POS'][k]]
        m_size = [(6.0 * filt_vcf_df['FREQUENCY'][k]) ** 2]
        m_shape = ref_dict([filt_vcf_df['TYPE'][k]], m_shape_dict)[0]
        m_col = ref_dict([filt_vcf_df['FTYPE'][k]], m_col_dict)[0]
        ax.scatter(x=x_value, y=y_value, s=m_size, marker=m_shape, c=m_col, edgecolor =
'k', linewidth=0.5, alpha=0.5)
# アライメントされなかった部分の描画
print('Processing not aligned area...')

```



```

no_depth_list = get_not_aligned_area(
    depth_files=depth_files,
    left_pos=start_pos,
    right_pos=end_pos,
    min_depth=min_depth
)
rect_height = 2
for j in range(len(no_depth_list)):
    for k in range(len(no_depth_list[j])):
        x_value = no_depth_list[j][k][0]
        y_value = ref_dict([sample_list[j]], y_value_dict)[0] - 1
        rect_width = no_depth_list[j][k][1] - x_value
        rect = patches.Rectangle(xy=(x_value, y_value), width=rect_width,
height=rect_height, fc='black', ec='black', alpha=0.5, linewidth=0.)
        ax.add_patch(rect)
    # fig.show()
    fig.savefig(svg_files[i], bbox_inches='tight')
    print('Completed saving graph to'+ svg_files[i])
return svg_files

def gen_graph(out_dir, graph_config_csv, target_bed, ref_fa, cds_gff, min_depth=5):
    """gen_graph main function.

    graph_config.csv file is generated from alignment.make_alignment.
    target.bed is generated from get_target.get_target_position.
    If the values in name column are same, target regions are combined.
    cds.gff file is generated from get_target.prodigal_py.

    :param out_dir: output directory
    :type out_dir: str
    :param graph_config_csv: path to graph_config.csv
    :type graph_config_csv: str
    :param target_bed: path to target.bed
    :type target_bed: str
    :param ref_fa: path to a file of complete genome or contig in fasta format as a reference
    :type ref_fa: str
    :param cds_gff: path to cds.gff
    :type cds_gff: str
    :param min_depth: minimum depth of sequence
    :type min_depth: int
    """
    out_dir = make_directory(dir_name='missense_mut_graph', out_dir=out_dir)
    graph_conf_df = pd.read_csv(graph_config_csv)
    sample_list = graph_conf_df.iloc[:, 0].values.tolist()
    bam_files = graph_conf_df.iloc[:, 1].values.tolist()
    print('generating vcf files.....')
    row_vcf_files = mpileup(
        bam_files=bam_files,
        sample_list=sample_list,
        ref_fa=ref_fa,
        out_dir=out_dir,
        bq_threshold=0
    )
    print('annotating vcf files.....')
    vcf_files = snpEff_py(
        out_dir=out_dir,
        cds_gff=cds_gff,
        raw_vcf_files=row_vcf_files,
        sample_list=sample_list
    )
    print('converting vcf files into csv files.....')
    raw_snps_csv_files, filt_snps_csv_files = gen_snps_csv(

```

```

    vcf_files=vcf_files,
    sample_list=sample_list,
    out_dir=out_dir,
    min_depth=min_depth,
    min_freq=0.3,
    ref_fa=ref_fa
)
print('calculating depth.....')
depth_files = calc_depth(
    out_dir=out_dir,
    sample_list=sample_list,
    bam_files=bam_files
)
print('making graph.....')
svg_files = make_graph(
    sample_list=sample_list,
    target_bed=target_bed,
    filt_snps_csv_files=filt_snps_csv_files,
    depth_files=depth_files,
    out_dir=out_dir,
    min_depth=min_depth
)
legend_file = gen_legends(
    out_dir=out_dir
)
print('Finish.')
return None

```

utility.py

```

#!/usr/bin/env python3
import os
from argparse import ArgumentParser
from __version__ import __version__

def get_option():
    super_argparser = ArgumentParser(
        prog='mmviewer',
        description='Missense Mutation Viewer (MMViewer) v' + str(__version__) + ' (K. Ikebata,
2021)'
    )
    super_argparser.add_argument(
        '-v', '--version', action='version', version='%(prog)s v' + str(__version__)
    )
    sub_argparser = super_argparser.add_subparsers(
        dest='program_name',
        help="""First, run get_target program to get target.bed file and cds.gff file. \
Second, run alignment program to get graph_config.csv file. \
Third, run gen_graph program to generate graph."""
    )
    # definition of sub arg
    gt_argparser = sub_argparser.add_parser(
        'get_target', help='Run get_target program to get target.bed describing target region
and cds.gff file.'
    )
    al_argparser = sub_argparser.add_parser(
        'alignment', help='Run alignment program to get bam files.'
    )
    gg_argparser = sub_argparser.add_parser(
        'gen_graph', help='Run gen_graph program to generate mutation graph.'
    )
    # Set gt_argparser
    rq_gt_group = gt_argparser.add_argument_group('required arguments')

```

```

op_gt_group = gt_argparser.add_argument_group('graph region (optional)')
rq_gt_group.add_argument(
    '-c', '--complete_seq', type=str, required=True,
    help='Path to a file of complete genome or contig in fasta format as a reference.
(required)')
)
rq_gt_group.add_argument(
    '-g', '--gene_sequence', type=str, required=True,
    help='Path to a file of target gene in fasta format. (required)')
)
rq_gt_group.add_argument(
    '-o', '--output', type=str, required=True,
    help='Output directory. (required)')
)
rq_gt_group.add_argument(
    '-t', '--gene_seq_type', type=str, default='prot', choices=['nucl', 'prot'],
required=True,
    help='A type of gene_sequence file (nucleotide or animo acid), default = prot.
(optional)')
)
op_gt_group.add_argument(
    '-u', '--upper_interval', type=int, required=False, default=0, metavar='0',
    help='The bp number of additional upper region from cds, defaults to 0. (optional)')
)
op_gt_group.add_argument(
    '-l', '--lower_interval', type=int, required=False, default=0, metavar='0',
    help='The bp number of additional lower region from cds, you want to analyze, defaults
to 0. (optional)')
)
# Set al_argparser
rq_al_group = al_argparser.add_argument_group('required arguments')
rq_al_group.add_argument(
    '-a', '--alignment_config_file', type=str, required=True,
    help='''Alignment_config.csv file consists of 2 or 3 columns with a header in first row.
1st column: sample name.
2nd column: path to trimmed read file in fastq format or its compressed format
(.gz).
3rd column (option): path to trimmed read file in fastq format or its compressed
format (.gz) of reverse reads if paired end read.'''
)
rq_al_group.add_argument(
    '-c', '--complete_seq', type=str, required=True,
    help='Path to a file of complete genome or contig in fasta format as a reference.
(required)')
)
rq_al_group.add_argument(
    '-o', '--output', type=str, required=True,
    help='Output directory. (required)')
)
# Set gg_argparser
rq_gg_group = gg_argparser.add_argument_group('required arguments')
op_gg_group = gg_argparser.add_argument_group('optional arguments')
rq_gg_group.add_argument(
    '-c', '--complete_seq', type=str, required=True,
    help='Path to a file of complete genome or contig in fasta format as a reference.
(required)')
)
rq_gg_group.add_argument(
    '-o', '--output', type=str, required=True,
    help='Output directory. (required)')
)
rq_gg_group.add_argument(

```

```

        '-a', '--graph_config', type=str, required=True,
        help="""Path to graph_config.csv file generated by \'mmviewer alignment\'.
        It consists of sample_name column and bam file path column with header in first row.
        Sample order of output graph follows the order in this file."""
    )
    rq_gg_group.add_argument(
        '-b', '--target_bed', type=str, required=True,
        help="""Path to target.bed file generated by \'mmviewer get_target\'.
        It consists of 8 columns (chrom, chromStart, chromEnd, name, score, strand,
        CDS_Start, CDS_End).
        Users can edit this file.
        If the values in name column are same, regions in these rows will be shown in the
        same graph sheet.
        """)
    )
    rq_gg_group.add_argument(
        '-d', '--cds_gff', type=str, required=True,
        help='Path to cds.gff file generated by \'mmviewer get_target\'.')
    )
    op_gg_group.add_argument(
        '-p', '--min_depth', type=int, required=False, default=5, metavar='5',
        help='Minimum depth of reads as a mapped region. (default=5)')
    )
    return super_argparser.parse_args()

def make_directory(dir_name, out_dir):
    """Make new directory and confirm that there's no same directy name.
    If new file_name already exist, add '_1' at the last.

    :param dir_name: new directory name
    :type file_name: str
    :param out_dir: path to directory where you want to put the new directory
    :type out_dir: str
    :return: path to new directory
    :rtype: str
    """
    i=0
    tmp_dir=os.path.join(out_dir, dir_name)
    while True:
        if not os.path.isdir(tmp_dir):
            break
        else:
            i += 1
            tmp_dir = os.path.join(out_dir, dir_name + '_' + str(i))
    if os.path.join(out_dir, dir_name) != tmp_dir:
        print('warning! ' + out_dir + ' already exist!')
        print(tmp_dir + ' was generated instead.')
    os.mkdir(tmp_dir)
    return tmp_dir

def make_dir_files(dir_name, out_dir, prefix_list, ext):
    """Make new directory and generate files path list

    :param dir_name: new directory name
    :type file_name: str
    :param out_dir: path to directory where you want to put the new directory
    :type out_dir: str
    :param prefix_list: prefix list of files
    :type prefix_list: list
    :param ext: file extension (e.g. '.txt')
    :type ext: str
    :return: directory name, file list

```

```

:rtype: (str, list)
"""
i=0
tmp_dir=os.path.join(out_dir, dir_name)
while True:
    if not os.path.isdir(tmp_dir):
        break
    else:
        i += 1
        tmp_dir = os.path.join(out_dir, dir_name + '_' + str(i))
if os.path.join(out_dir, dir_name) != tmp_dir:
    print('warning! ' + out_dir + ' already exist!')
    print(tmp_dir + ' was generated instead.')
os.mkdir(tmp_dir)
files = [os.path.join(tmp_dir, ll + ext) for ll in prefix_list]
return tmp_dir, files

```