高解像度メタバーコーディング手法で描く 琵琶湖・淀川水系における病原細菌群の全容

京都大学大学院 工学研究科　中西智宏

１. はじめに

近年の温暖化や気象の極端化を鑑みれば、水道水源における微生物リスク管理の重要性が増している。飲料水の微生物リスクを評価するためには、定量的微生物リスク評価(Quantitative Microbial Risk Assessment; QMRA)手法が用いられる。QMRA ではまず対象とする病原体を決定し、水源等での濃度を測定/推定したあとに感染確率の推定を行う。しかし、水源には多種多様な病原体が存在しており、始めに対象微生物を絞ることで本当に重要な病原体を見逃してしまう可能性がある。

一方、これまで琵琶湖・淀川水系においては *Campylobacter* [1] や *Legionella* [2]といった細菌類、薬剤耐性菌 [3]、ウイルス [4,5]、原虫類 [6]が測定され、個々の病原体の分布調査や QMRA が行われてきた。しかし、「琵琶湖・淀川水系には潜在的にどれだけの種類の病原微生物が存在するのか？」ということは未だ明らかになっておらず、重要な病原体についての情報も乏しい。合理的な微生物リスク管理のためには、制御すべき病原微生物の多様性を正しく捉え、流域での分布特性を把握しておくことが望ましい。

上記を踏まえ、本研究では琵琶湖・淀川水系において病原細菌を網羅的に検出し、病原細菌の多様性や分布特性を把握することを目的とした。ここで、微生物の一斉検出には次世代シーケンサーを用いた網羅的な遺伝子解析が有用だが、細菌の病原性の有無を判断するには種レベルまでの高い分類解像度を備えた分析技術が必要である。そこで、本研究では 16S rRNA 遺伝子(16S rDNA)全長を解読対象としたメタバーコーディング手法に着目している。

2021 年度にはロングリード型次世代シーケンサー(NGS)を用いて分析したものの、環境水から検出される病原種の一部には偽陽性が疑われ、信頼性の向上が課題であった。その原因として、ロングリード型 NGS による塩基配列の解読エラー（読み間違い）が考えられた。そこで、2022 年度は解読エラーを補正する手法 [7]を用いることとし、方法を改善した上で改めて桂川・淀川水系に適用することとした。

２. 研究方法

本研究で用いた解読エラー補正手法 [7]の概要を図 1 に示す。解析対象となる DNA 分子ごとに異なるランダム塩基（Unique Molecular Identifier; UMI）を付加した後に PCR 増幅を行い、ロングリード型 NGS で配列データを取得した後、UMI 配列によって取得リードをまとめ、コンセンサス配列を得ることで解読エラーを補正するものである。以降、このエラー補正によって得られた配列を単にコンセンサス配列と表記する。

2.1 標準試料を用いたエラー補正の精度チェック

上記の手法によるエラー補正の精度を確認するために、15 種の既知の細菌種が既知濃度で混合された標準試料を用いた。具体的には、独立行政法人 製品評価技術基盤機構 (NBRC）から標準 DNA（DNA-Mock-001）を購入した。その細菌種の組成と混合割合を表 1 に示す。これを鋳型として 16 塩基の縮重塩基を含む UMI 配列が付加されたプライマー対(27F/1492R)

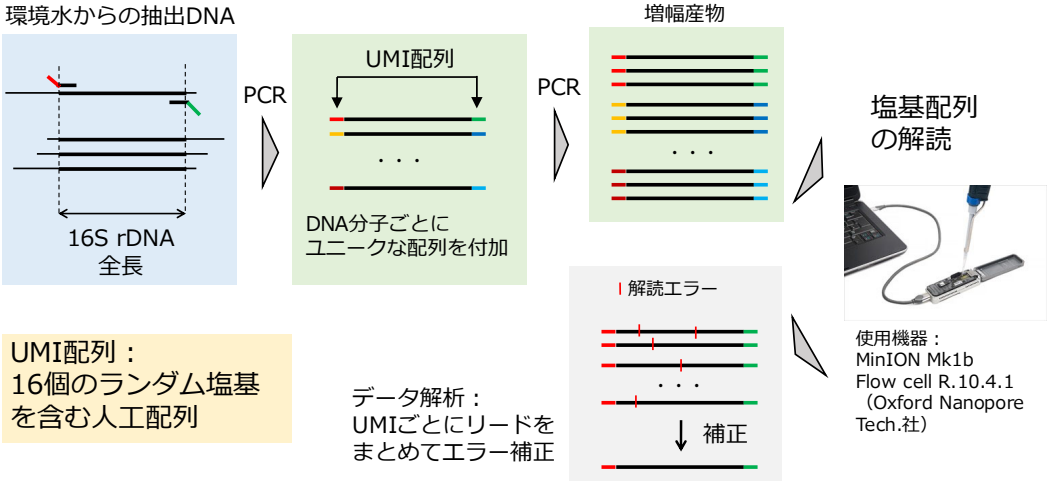

図１ 解読エラーの補正方法の概要（既往研究[7]による方法を採用）

表１ 標準試料に含まれる細菌種の学名と16S rDNAコピー数の存在割合

学名	NBRC 番号	16S rDNAの存在割合（理論値） (%)
Bacillus subtilis subsp. subtilis	13719T	10.5
Bifidobacterium pseudocatenulatum	113353	5.3
Clostridium butyricum	13949T	11.6
Corynebacterium striatum	15291T	4.2
Cutibacterium acnes subsp. acnes	107605T	3.2
Lactobacillus delbrueckii subsp. delbrueckii	3202T	8.4
Staphylococcus epidermidis	100911T	6.3
Streptococcus mutans	13955T	5.3
Acinetobacter radioresistens	102413T	6.3
Bacteroides uniformis	113350	4.2
Enterocloster clostridioformis	113352	5.3
Comamonas terrigena	13299T	7.4
Escherichia coli (K-12株)	3301	7.4
Parabacteroides distasonis	113806	7.4
Pseudomonas putida	14164T	7.4

で 2 サイクルの PCR 反応を行い、16S rDNA 分子の両末端に UMI 配列を付加させた。精製後、さらに PCR 増幅とライブラリ調製を行い、MinION Mk1B と Flow Cell R10.4.1(共に Oxford Nanopore Technologies 社)を用いて配列データを取得した。PCR 反応液の組成や反応条件、ライブラリ調製は既報[7, 8]に従った。得られたデータを Guppy ver. 6.3.4 でベースコールした後、Karst ら[7]の解析パイプラインを用いて UMI 配列ごとにリードをまとめ、コンセンサス配列を生成した。標準試料に含まれる細菌種の 16S rDNA 参照配列を公共データベース（NCBI Refseq）から取得し、コンセンサス配列を比較することで解読エラーの補正の度合いを確認した。

2.2 桂川・淀川水系の水試料に対する手法の適用

採水地点を図 2 に示す。桂川～淀川水系に沿って、保津峡付近の渓流水、保津川、宮前橋、枚方大橋地点（三川合流後）での河川水を採取した。また、保津川～宮前橋の間に位置する下水処理場放流水も採取した。孔径 0.22 μm のろ紙で検水をろ過濃縮後、PowerSoil kit (Qiagen

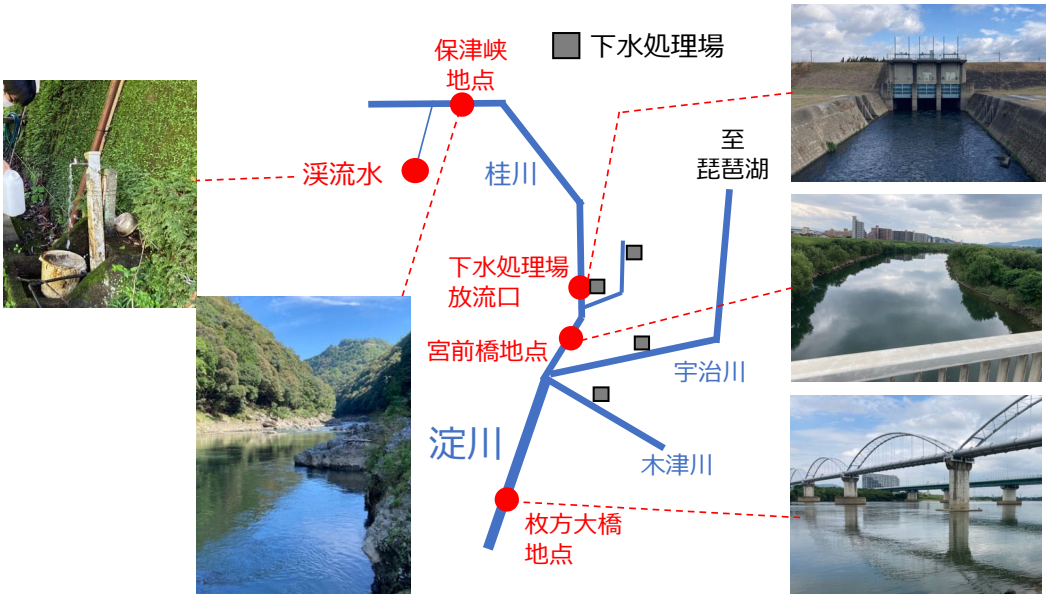

図 2 採水地点

社)で核酸抽出した。得られた核酸抽出液に対して 2.1 と同様の操作でライブラリ調製を行い、シーケンス解析、コンセンサス配列の生成を行った。

環境水から得られたコンセンサス配列を *rrn* オペロン配列のデータベース MIrROR[9] に対して BLAST 検索した。各問い合わせ配列について、上位の複数ヒットを用いて最小共通祖先を推定した。属レベルまで分類群が推定できたコンセンサス配列のうち、病原属（ヒトに対して病原性を示す種が 1 つでも含まれる細菌属）として分類されたものを抽出した。

次に、病原属と推定されたコンセンサス配列の種レベルまでの分類群を高い信頼度で推定するために、さらなる BLAST 検索を行った。16S rDNA 配列データベース SILVA version138[10] において基準株(Type strain)あるいは培養株(Cultured strain)として登録されているエントリーから、配列の信頼性などの観点から品質の良いものだけを抽出し、新たなデータベースを作成した。これを照合先として、病原属と推定されたコンセンサス配列を問い合わせ配列として BLAST 検索を行った。各問い合わせ配列について、配列類似性 99%以上の複数ヒットで最小共通祖先を推定し、種レベルまでの分類群の推定を行った。最後に、1409 種からなる病原種リストと照合し、病原種を決定した。

また、大腸菌および嫌気性芽胞菌（原虫汚染指標）による病原細菌の指標性を確認するため、各採水試料についてこれらを測定し、上記のシーケンス解析結果と比較した。

３. 結果と考察

3.1 標準試料を用いた解読エラー補正のチェック

標準試料に対して 16S rDNA 全長を対象とした DNA メタバーコーディング法を適用し、シーケンシングエラーを補正することで計 34678 本のコンセンサス配列が得られた。これと標準試料に含まれる細菌種 15 種の参照配列との類似性を集計したものを図 3 に示す。参照配列はいわば「正解」とも言える真の配列であり、これと高い類似性を示すほど、正しく塩基配列を解読できていることを示す。コンセンサス配列の 95%以上が参照配列と 99.8%以上の配列類似性を示した。ロングリード型 NGS である MinION の生リードのエラー率は 1～数%と言われているが、これに比べて大幅に解読エラーが低減できていることが確認された。

本手法によって標準試料に含まれている 15 種全てが検出されており、感度は良好であっ

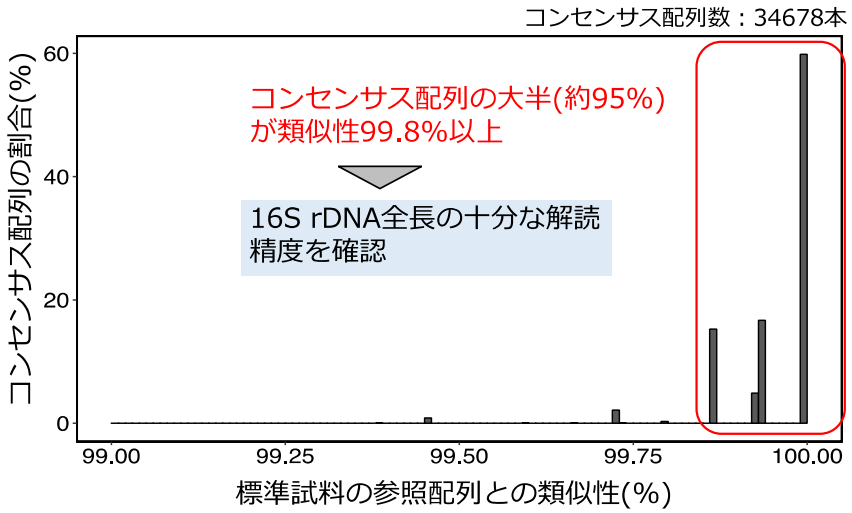

図 3 コンセンサス配列と真の配列との類似性

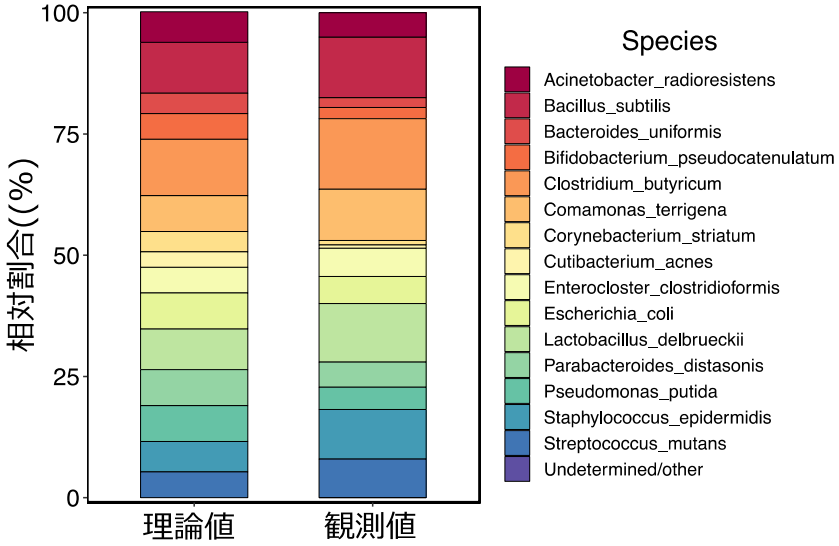

図 4 本手法による標準試料中の細菌種の組成の推定結果

た。また、異なる種として誤検出、あるいは種レベルの分類群の推定ができなかったコンセンサス配列は 34678 本中 5 本のみであり、検出精度も極めて高かった。続いて、図 4 に各細菌種の相対存在割合を示す。図中左側の「理論値」は標準試料のメーカー公表値であり、右側の「観測値」の積み上げグラフがこれに近いほど真の群集構造を正しく推定できていることを意味する。これより、特定の細菌種だけが極端に高い割合を示すような傾向も見られなかった。すなわち、ライブラリ調製時における PCR によってバイアスがかかることもなく、ある程度正しく細菌組成（全細菌に占める相対割合）を推定できることが確認できた。

3.2 桂川・淀川水系における病原細菌種の一斉検出

図 5 に各採水試料における病原種の検出状況を示す。ヒートマップ中の色の濃さは各試料から得られたコンセンサス配列の総数に占める割合を示す。また、分類群を属レベルでまとめて表示している。全 15 検体から計 37 属 81 種が病原種として推定された。下水処理場放流水とその放流地点より下流（桂川宮前橋地点、淀川枚方大橋地点）の河川水では、*Streptococcus mutans*、*Klebsiella pneumoniae*、*Bacteroides uniformis* 等の動物の腸管等に寄生する病原細菌が

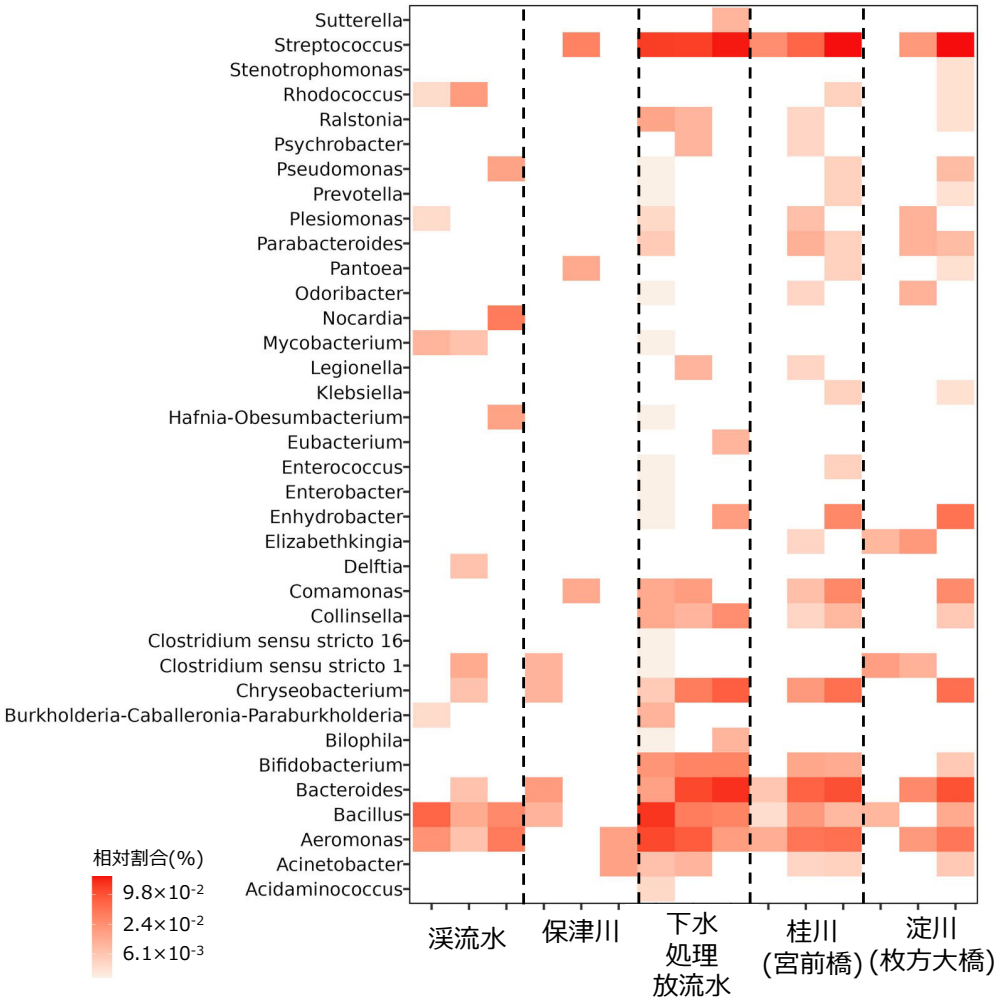

図 5 桂川水系における病原細菌の検出結果（同属の病原種をまとめ、属名で記載）

比較的検出されやすい傾向にあった。一方、渓流水や保津川（下水処理水の放流による影響が少ないと思われる地点）ではこれらの糞便由来の病原種の検出率が低く、*Pseudomonas putida*、*Chryseobacterium indologenes*、*Pseudomonas fluorescens*、*Aeromonas hydrophila*、*Comamonas aquatica* といった土壌、水環境中の常在菌として知られる細菌が中心であった。なお、これらは日和見感染菌（高齢者などの免疫の低い人々に対して感染性を示す病原菌）であり、それほど病原性は強くないものと考えられる。次に、図 6 に各地点で検出された病原種の数を示す。病原種の数は下水処理水やその放流後の河川水で多くなる傾向が見られた。以上より、下水処理水は病原細菌、特に糞便由来の病原細菌の主要なソースであり、河川水の微生物リスクの増大に寄与していることが明らかとなった。

最後に、図 7 に大腸菌、嫌気性芽胞菌の測定結果を示す。大腸菌濃度は各地点で大きく変動し、渓流水のみ他地点より低濃度となる様子が確認された。大腸菌は土壌由来のものも存在し、それらも含めて測定されるために下水処理水の放流の影響が見られなかったものと思われる。一方、嫌気性芽胞菌は下水処理水の放流によって急激に濃度が上昇した。これより、本調査の範囲では、大腸菌よりも嫌気性芽胞菌のほうが図 6 のような病原種の検出状況をうまく表現できているといえ、ある程度の指標性が期待できるものと考えられた。

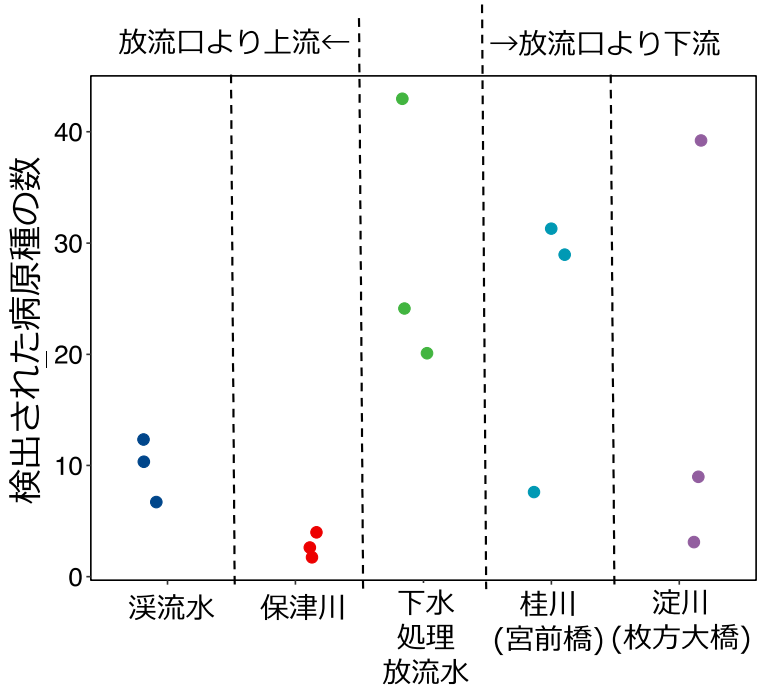

図 6 桂川水系の流下過程における病原細菌種の検出数の変化

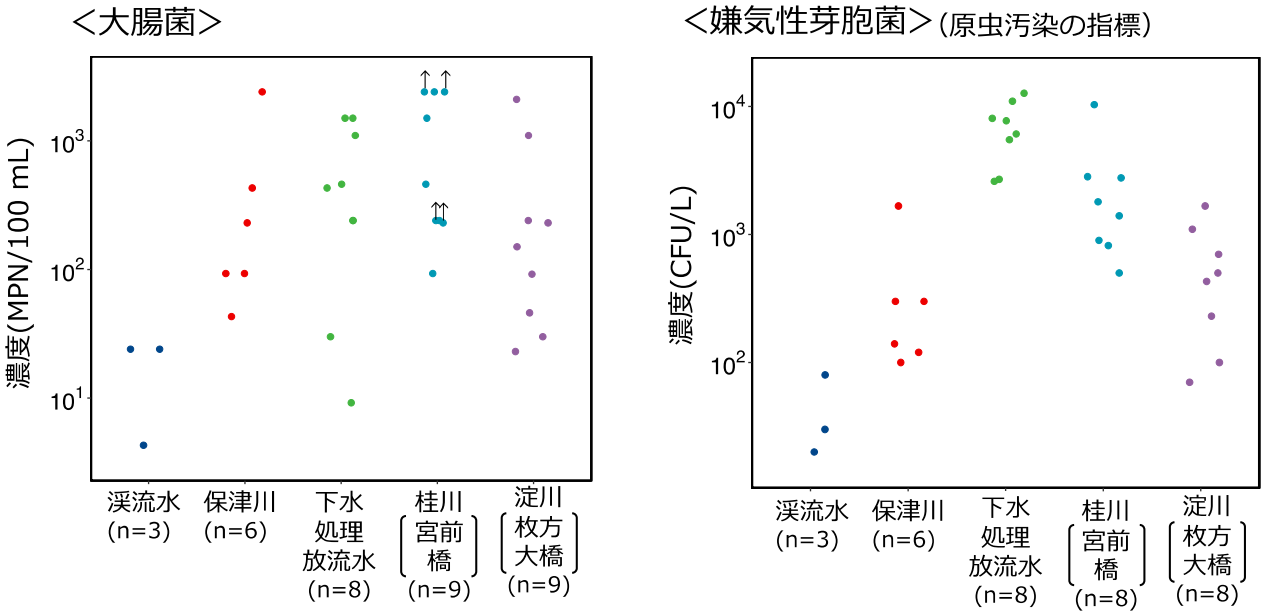

図 7 桂川水系の流下過程における指標微生物濃度の変化（左：大腸菌、右：嫌気性芽胞菌）
（図中の↑は濃度が検出上限を超過したことを意味し、検出上限値でプロットしている）

４．おわりに

本研究では、水道原水中の病原細菌のスクリーニングのために完全長 16S rDNA を対象とした DNA メタバーコーディング手法を確立した。さらにこれを桂川・淀川水系に適用した結果、病原種として全 15 検体から計 37 属 81 種が抽出でき、下水処理水の放流によって病原細菌の多様性が増し河川水の微生物リスクの増大に寄与していることを示しえた。

本手法によって病原細菌の大まかなスクリーニング（存在する可能性のある病原細菌をな

るべく多く同時検出すること）は可能となったものの、病原遺伝子を直接検出しているわけではないため、偽陽性の可能性は依然として残ることには注意が必要である。特に、病原性大腸菌に代表される一部の細菌は 16S rDNA 配列のみでは病原性を全く判定できないことが知られている。今後、本手法が適用できる病原細菌の種類を整理することが望ましい。また、冒頭で示したリスク評価（QMRA）にこの手法を利用するためには、検出される病原細菌の定量値や生死状態の把握も必要となる。これらは今後の課題である。

参考文献

1) 浅田安廣：淀川流域における感染性病原微生物の実態把握と遊泳等による感染リスク推定モデルの構築, 平成 26 年度水質保全研究助成 成果報告書.
2) 枝川 亜希子：淀川水系に生息するレジオネラ属菌と宿主となる自由生活性アメーバに関する研究, 平成 30 年度水質保全研究助成 成果報告書.
3) 東 剛志：臨床領域で対策が求められている薬剤耐性菌の河川環境中における動態, 平成 31 年度水質保全研究助成 成果報告書.
4) 浅田安廣：淀川流域における病原ウイルスの存在把握ならびに健康影響評価, 平成 28 年度水質保全研究助成 成果報告書.
5) 浅田安廣：淀川流域におけるアデノウイルス血清型の網羅的探索に基づいた水系感染による健康影響の推定, 平成 29 年度水質保全研究助成 成果報告書.
6) 大西義博・大瀧博文：大容量の淀川水系河川水からの病原性原虫類の検出, 平成 28 年度水質保全研究助成 成果報告書.
7) Karst, Søren M., et al. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nature methods* 18.2 (2021): 165-169.
8) Oxford Nanopore Technologies 社, プロトコル CPU_910 7_v109_revD_09Oct2020, 2023 年 1 月閲覧
9) Seol, Donghyeok, et al. Microbial identification using rRNA operon region: database and tool for metataxonomics with long-read sequence. *Microbiology Spectrum* 10.2 (2022): e02017-21.
10) Quast, Christian, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* 41.D1 (2012): D590-D596.