



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위논문

Long-read와 Short-read 시퀀싱을 통한
항생제 내성체 비교 분석

정지원

제주대학교 대학원

생명공학부 분자생명공학전공

2024년 2월



Long-read와 Short-read 시퀀싱을 통한

항생제 내성체 비교 분석

이 논문을 석사 학위논문으로 제출함

정 지원

제주대학교 대학원

생명공학부 분자생명공학전공

지도교수 이효연

정지원의 이학 석사 학위논문을 인준함

2023년 11월

심사위원장 운노 타쓰야 _____

위 원 이효연 _____

위 원 부경환 _____

목 차

목차	I-III
LIST OF FIGURES	IV-V
LIST OF TABLES	VI
국문 요약	VII-VIII
1. INTRODUCTION	1-5
2. MATERIALS & METHODS	6-13
2.1. Site description	6
2.1.1. Sample collection	6
2.1.2. DNA extraction	7
2.2. Sequencing	8-9
2.3. Data analysis	10-13
2.3.1. Trimming	10
2.3.2. Microbial community classification	10
2.3.3. ARG annotation	11-12
2.3.4. MGEs associate with ARGs detection	12-13

2.3.5. ARG-harboring plasmid annotation	13
2.4. Statistical analysis	13
3. RESULTS & DISCUSSION	14-47
3.1. Sequencing data summary	14-15
3.2. Microbial community analysis	16-24
3.2.1. Taxonomical classification	16-21
3.2.2. Alpha and beta diversity	22-24
3.3. Abundance of ARGs	25-33
3.3.1. Absolute abundance of ARGs	25-27
3.3.2. Relative abundance of ARGs	28-30
3.3.3. Overlapping ARGs	31-33
3.4. Mobile genetic elements (MGEs) associated with ARGs	34-43
3.5. ARG-harboring plasmids	44-47
3.5.1. Abundance of ARGs on plasmids	44-45
3.5.2. ARG-carrying reads and contigs in plasmids	46-47
4. CONCLUSION	48
ACKNOWLEDGMENT	49

REFERENCES 50-54

ABSTRACT 55-57

LIST OF FIGURES

Figure 1. Sampling sites on the Miho-river	6
Figure 2. Overview of experimental study.	9
Figure 3A. Heatmap analysis of microbial community (Phylum level)	18
Figure 3B. Heatmap analysis of microbial community (Class level)	18
Figure 3C. Heatmap analysis of microbial community (Order level)	19
Figure 3D. Heatmap analysis of microbial community (Family level)	20
Figure 3E. Heatmap analysis of microbial community (Genus level)	21
Figure 4A. Alpha diversity analysis (Chao1) for microbial community	23
Figure 4B. Alpha diversity analysis (Shannon) for microbial community	23
Figure 5. Beta diversity analysis (nMDS) for microbial community	24
Figure 6. Absolute abundance of ARGs	26
Figure 7A. Relative abundance of ARGs	29
Figure 7B. Principal component analysis (PCA) based on the relative abundance of ARGs	30

Figure 8A. Relative abundance of overlapping ARGs	32
Figure 8B. Principal component analysis (PCA) based on the relative abundance of overlapping ARGs	33
Figure 9. The length of ARG-carrying reads and contigs.....	34
Figure 10A. The number of ARG-MGE types across the sampling sites ..	36
Figure 10B. ARG-MGE types identified in both MinION and HiSeq sequencing	37
Figure 11A. Synteny analysis of ARG-MGE types at MH08.....	38
Figure 11B. Synteny analysis of ARG-MGE types at MH09	39
Figure 11C. Synteny analysis of ARG-MGE types at SY01	40-41
Figure 11D. Synteny analysis of ARG-MGE types at BR01	42
Figure 11E. Synteny analysis of ARG-MGE types at MH10.....	43
Figure 12. Relative abundance of ARGs on plasmids	45
Figure 13. Heatmap analysis of the ARG-carrying plasmids.....	47
Figure 14. Graphical abstract	57

LIST OF TABLES

Table 1. Advantages and disadvantages of MinION and HiSeq sequencing	5
Table 2. Sampling site information	7
Table 3. Summary of sequencing results	15
Table 4. Spearman's rho correlation coefficients for comparison between long- and short-read sequencing	27

Long-read와 Short-read 시퀀싱을 통한

항생제 내성체 비교 분석

정 지원

제주대학교 대학원 생명공학부 분자생명공학전공

요약

항생제 내성 유전자의 출현과 전파는 전 세계적인 공중 보건 문제이다. 환경에 존재하는 항생제 내성체에 대한 대부분의 연구에서 whole genome이 아닌 항생제 내성 유전자(antibiotic resistance gene, ARG)의 정량 분석을 통해 그 위험성을 밝혀왔으며, ARG의 이동과 전파에 대한 연구는 아직 충분히 이루어지지 않았다. 이에 본 연구에서는 기존에 수행되고 있던 short-read sequencing (HiSeq)과 비교적 최근 개발된 long-read sequencing (MinION)을 통해 환경에 존재하는 미생물 군집의 다양성을 조사하고, 항생제 내성체에 대한 비교 분석을 실시하였다. 미생물 군집을 조사한 결과, long-read sequencing과 short-read sequencing의

데이터로 각각 911, 841개의 박테리아 속(Genus)을 분류하였다. Long-read sequencing의 결과에서는 분류되지 않은(unclassified) 속이 평균 8.7 %를 차지하는 것에 비해 short-read sequencing의 결과에서는 63.3 %의 속이 unclassified 그룹에 속하는 것으로 나타났다. 두 플랫폼 간의 ARG abundance는 비슷한 경향을 보였으며(Spearman's rho = 0.655, p <0.05), multidrug resistance (MDR)를 가진 read가 가장 많은 것으로 나타났다. Read와 contig 내에서 5 kb 이내의 거리에 존재하는 ARG와 이동성 유전 인자(mobile genetic element, MGE) 쌍의 종류는 MinION과 HiSeq의 결과에서 각각 69, 25가지가 검출되었다. Plasmid로 예측되는 read와 contig가 가지고 있는 ARG는 MinION과 HiSeq의 결과에서 각각 183, 91개가 검출되었다. ARG를 가지고 있는 read와 contig 중에 sulfonamide/sulfone, aminoglycoside, tetracycline 계열에 내성을 가지고 있는 read와 contig가 대부분 plasmid인 것으로 나타났다. 따라서 본 연구에서는 미생물 군집의 다양성과 항생제 내성체 분석을 통해 long-read sequencing과 short-read sequencing의 차이점을 비교하고, 앞으로 ARG의 이동과 전파에 대한 연구의 방향성을 제시하고자 한다.

1. INTRODUCTION

항생제의 오남용으로 인한 항생제 내성 유전자(antibiotic resistance gene, ARG)의 출현과 환경으로의 전파는 전 세계적인 공중 보건 문제가 되었다 [1]. 지금까지 환경에 존재하는 항생제 내성체에 대한 수많은 연구가 이루어졌으며, 주로 real-time PCR (qPCR)과 next generation sequencing (NGS) 기술을 활용하여 환경에서 검출되는 ARG의 분포 양상을 조사하거나 ARG와 항생제 내성이 있는 숙주 박테리아와의 상관관계를 분석한 연구가 수행되었다 [2-5]. 대부분의 연구에서 whole genome이 아닌 ARG의 정량화를 통해 ARG의 위험성을 밝혀냈지만, 현재는 ARG의 이동과 전파에 중점을 뒀다 한다는 인식이 증가하고 있다 [6, 7].

박테리아 간 ARG의 이동은 주로 horizontal gene transfer (HGT)에 의해 일어나는데, 이는 ARG의 이동을 매개하는 효소와 단백질을 암호화한 mobile genetic elements (MGEs)와 밀접한 관련이 있다고 알려져 있다 [8, 9]. Integron, transposon, plasmid 등의 MGE는 HGT를 통한 ARG 전파에 중요한 역할을 하며, 유전자 내 ARG와의 거리에 따라 위험도가 달라진다 [6]. 특히 MGE 중에서도 plasmid는 여러 환경에서 ARG의 저장소 역할을 수행하며 ARG의 전파를

가속화하는 주요 원인으로 작용한다 [10-12].

이제까지 많은 연구자들은 whole genome sequencing (WGS)을 위해 Illumina사의 short-read sequencing (그 중에서도 주로 HiSeq)을 사용했다 [13-15]. HiSeq sequencing은 DNA를 증폭시킨 뒤 4개의 염기에 결합하는 형광물질의 색상에 따라 DNA 서열을 결정하는 sequencing이고, 최대 500 base pair의 short-read를 생성할 수 있다. Short-read sequencing을 통해 우리는 수백 개의 샘플을 동시에 sequencing 하여 대량의 데이터를 한꺼번에 얻을 수 있으며 [16], single-nucleotide substitutions (SNSs)와 small insertions or deletions (INDELs)을 확실하게 감지할 수 있어 sequence에 대한 정확도가 높다는 장점을 가지고 있다 [17]. 하지만 Illumina sequencing을 통해 생성된 short-read로는 박테리아의 반복서열까지 정확하게 알아내기 어렵다 [18]. DNA read의 길이가 짧을수록 반복서열의 위치와 유전자의 구조를 결정하기 어렵기 때문에 short-read를 contig로 만드는 assembly 과정에서 문제가 발생한다는 연구 결과가 있었다 [19]. *Kingsford, C.*를 비롯한 연구자들은 sequencing 기술이 빠른 속도로 발전하고 있지만, assembly 과정에 기술적으로 해결하기 어려운 근본적인 한계가 존재한다고 발표하였다. 또 다른 연구에서는 assembly 후에 만들어진 contig만으로는 whole genome을 완벽하게 파악하기 어렵고, contig 또한 DNA

조각에 불과하므로 중요한 유전자의 누락 등의 short-read sequencing이 가진 문제점을 발표하였다 [20]. Short-read sequencing만으로는 ARG가 chromosome과 plasmid 중 어느 곳에 존재하는지 구별하기 어렵다는 의견도 있었으며 [21, 22], 대부분의 short-read sequencing library 제작 과정에 포함된 polymerase chain reaction (PCR) 과정 때문에 GC contents에 의한 bias가 생길 수 있다는 점 또한 short-read sequencing이 가진 문제점 중 하나로 밝혀졌다 [23].

Pacific Biosciences (PacBio)는 2011년에 single molecule real-time (SMRT) 기술을 개발하였다. 이 기술은 PCR을 하지 않고 raw read의 sequence를 읽어 내기 때문에 유전자가 가진 GC contents 비율 차이에 의한 오류가 적고, long-read에 존재하는 유전자의 위치 정보와 후성유전학적 특성까지 분석할 수 있다는 장점을 가지고 있다 [24]. 2014년에 Oxford Nanopore Technologies (ONT)가 개발한 MinION은 nucleotide 염기가 nanopore로 이루어진 flow cell을 통과할 때 생기는 전류 신호의 차이를 읽어내는 sequencer이다. 최대 수백 kilobase의 long-read를 생산할 수 있으며, PacBio와 함께 third-generation sequencing으로 불린다. MinION sequencing은 전류 신호를 DNA 서열로 변환하는 과정(basecalling)에서 생기는 오류 때문에 short-read sequencing에

비해 sequence에 대한 정확도가 낮고, 생체 단백질로 만들어진 pore의 특성상 배송 및 보관 조건에 따라 flow cell 활성화에 영향을 받는다고 알려져 있다 [25]. Sequencing library를 제작할 때 1,000 ng의 DNA가 필요하지만 1회의 sequencing 당 얻을 수 있는 데이터의 양이 Illumina sequencing에 비해 상대적으로 적어 염기 당 비용이 많이 든다는 단점도 있다 [26]. 그럼에도 불구하고 MinION이 개발된 이후부터 최근까지 long-read sequencing으로 유전체 구조 분석을 통해 사람에게 치명적인 질병 관련 유전자를 찾아내는 등 long-read sequencing의 장점을 살린 연구가 활발히 진행 중이다 [27-29]. 최근 long-read와 short-read sequencing가 가진 각각의 단점을 보완하는 방법으로 hybrid sequencing 연구들이 이루어지고 있으며, 많은 연구자들은 이 방법으로 DNA 서열의 정확도를 높이고, 유전자의 위치 및 구조 분석 또한 가능하다고 발표했다 [30-32]. 그러나 short-read sequencing과 long-read sequencing은 개별적인 sequencing이며, MinION sequencing 기술이 정확도와 샘플 처리량을 높여 염기 당 비용을 줄이지 못한다면 short-read sequencing을 대체하기는 어려울 것으로 보인다 [33]. 따라서 본 연구에서는 다음과 같은 네가지 항목을 비교함으로써 short-read 기반의 HiSeq sequencing과 long-read 기반의 MinION sequencing의 차이점을 확인하고자 하였다: 1) 미생물 군집의 다양성, 2) ARG의

abundance, 3) 검출된 ARG와 근거리(5 kb 이내)에 존재하는 MGE의 종류, 4) ARG를 포함하고 있는(ARG-harboring) plasmid의 종류와 abundance. 네 가지 항목의 비교를 통해 앞으로 두 플랫폼의 데이터를 활용한 미생물 군집 분석과 항생제 내성체 분석 연구의 방향성을 제시하고자 한다.

Table 1. Advantages and disadvantages of MinION and HiSeq sequencing.

	MinION	HiSeq
Advantages	<ul style="list-style-type: none"> ● Using raw reads ● No limit to read length ● Fast library preparation ● Direct sequencing of RNA and detection of RNA modifications 	<ul style="list-style-type: none"> ● High accuracy, low error rates ● High throughput ● Relatively low cost per Gb ● Small amounts of material required for library preparation (~ 1 ng)
Disadvantages	<ul style="list-style-type: none"> ● Low accuracy, high error rates ● Large amount of material required for library preparation (1~2 µg) ● Frequent changes of software versions, flow cells, and kits 	<ul style="list-style-type: none"> ● No real long reads ● PCR amplification required for library preparation (bias from GC contents) ● No direct detection of epigenetic modifications

2. MATERIALS & METHODS

2.1. Site description

2.1.1. Sample collection

충청북도 청주시 미호강에 존재하는 미생물 군집과 항생제 내성체를 조사하기 위해 2023년 6월 중 3일 이내에 비가 오지 않은 날에 미호강 일대 5곳(MH08, MH09, SY01, BR01, MH10)에서 강물을 채수하였다(Fig. 1, Table 2). 각각의 site에서 1 L씩 3회에 걸쳐 총 3 L의 표면수를 멸균한 채수병에 담은 뒤, 아이스박스에 넣어 6시간 이내에 실험실로 운반하였다.

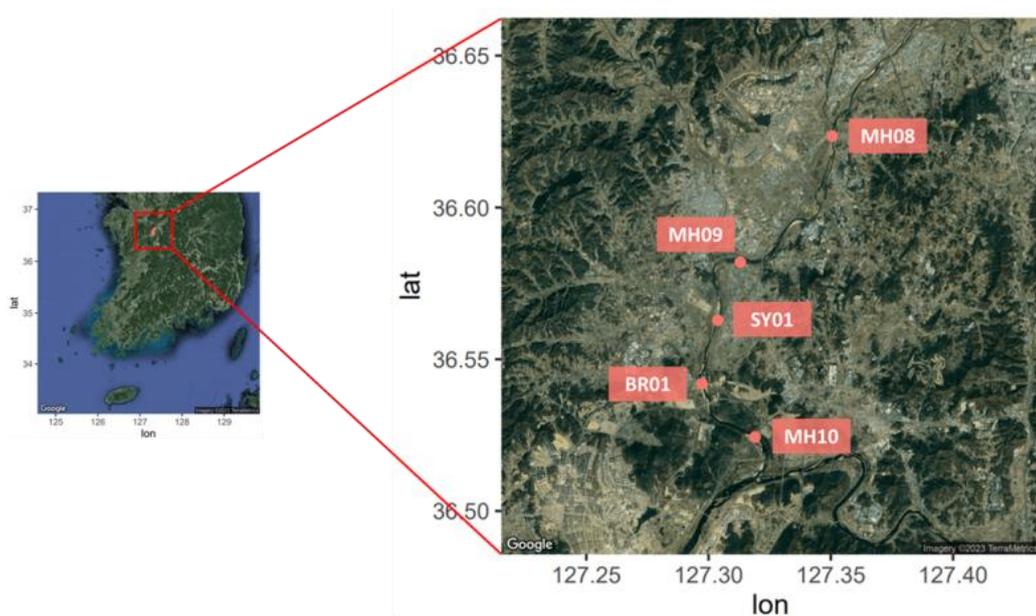


Figure 1. Sampling sites on the Miho-river. MH08: 8th of Miho-river site; MH09: 9th of Miho-river site; SY01: Songyong-ri; BR01: Borom-bridge; MH10: 10th of Miho-river site.

Table 2. Sampling site information.

Site	Address	Latitude	Longitude
MH08	청주시 흥덕구 강내면 탑연리 365	36.623532	127.350654
MH09	세종시 연동면 예양리 775-4	36.581986	127.312979
SY01	세종시 연동면 송용리 890-27	36.562871	127.303783
BR01	세종시 세종동 654-15 (보름교 위)	36.541954	127.297337
MH10	세종시 합강동 10-25	36.524442	127.319034

2.1.2. DNA extraction

Site 당 400 ml의 강물을 mixed cellulose ester (MCE) membrane filter (0.2 μm , Φ 47 mm) (ADVANTEC, Tokyo, Japan)로 거른 뒤, DNeasy PowerWater Kit (QIAGEN, Hilden, Germany)를 사용하여 DNA를 추출하였다. DNA의 농도는 dsDNA HS Assay Kit (ThermoFisher, MA, USA)를 사용하여 Qubit 2.0 Fluorometer (ThermoFisher, MA, USA)로 측정하였고, DNA의 순도(A260/280, A260/230)는 DS-11 FX+ spectrophotometer (DeNovix, DE, USA)로 측정하였다.

2.2. Sequencing

MinION sequencing을 위해 샘플 당 1000 ng의 DNA를 사용하였고, 6 kb의 길이로 fragmentation 하기 위해 100 μ l의 DNA 샘플을 g-TUBE (Covaris, MA, USA)에 넣고 30초 동안 11,000 rpm에서 총 2회 원심분리하였다. NEBNext FFPE RepairMix (NEB, Hitchin, UK)와 NEBNext Ultra II End Repair/dA-tailing Module (NEB, Hitchin, UK)를 사용하여 fragmented DNA의 양 끝을 보강하였다(end-prep). AMPure XP Beads (Beckman Coulter, CA, USA)를 사용하여 end-prepped DNA 이외의 불순물을 제거하였다. 제조사의 지시대로 Native Barcoding Kit 24 V14 (SQK-NBD114.24, Oxford Nanopore Technologies, UK), Blunt/TA Ligase Master Mix (NEB, Hitchin, UK)를 사용하여 native barcode를 end-prepped DNA에 ligation 하여 각각의 샘플을 구분하였다. NEBNext Quick Ligation Module (NEB, Hitchin, UK)를 사용하여 ONT에서 제공하는 Native adapter를 barcoded DNA에 ligation 하여 sequencing library를 제작하였다. 하나의 library에는 4개의 DNA 샘플이 pooling되었고, sequencing buffer, library beads와 함께 flow cell (R10.4.1, Oxford Nanopore Technologies, UK)에 loading한 뒤 72 시간동안 sequencing을 실시하였다.

Whole genome sequencing (WGS)를 위해 추출한 DNA를 Macrogen Inc.

(Seoul, Korea)에 의뢰하였고, TruSeq DNA Nano Library Prep Kit (Illumina Inc., CA, USA)으로 library를 제작한 뒤, HiSeqXten (Illumina Inc., CA, USA)으로 sequencing을 진행하였다.

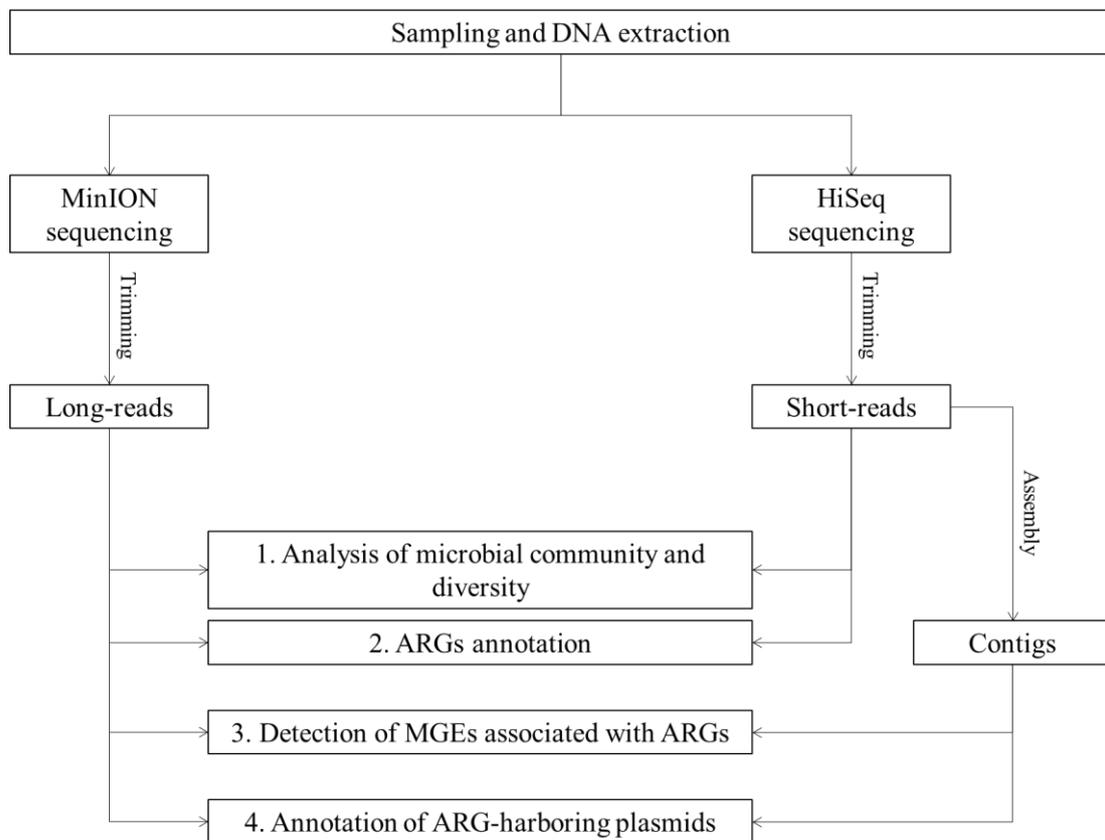


Figure 2. Overview of experimental study.

2.3. Data analysis

2.3.1. Trimming

MinION sequencing의 결과로 생성된 FAST5 파일은 ONT에서 제공하는 Dorado (v0.4.3) (<https://github.com/nanoporetech/dorado>)의 high accuracy mode로 basecalling하여 FASTQ 파일로 변환되었다. Basecalling의 과정에서 native barcode sequence는 제거되었고, Porechop (v0.2.4) (<https://github.com/rrwick/Porechop>)으로 native adapter sequence를 제거하였다. HiSeq sequencing의 결과로 생성된 FASTQ 파일은 FaQCs (v2.08) [34] 의 기본값으로 trimming되었다.

2.3.2. Microbial community classification

MinION sequencing 데이터는 trimming을 거친 뒤, Centrifuge [35] classifier를 기반으로 한 ARGpore2 [36] 를 사용하여 미생물 군집을 분석하였고, HiSeq sequencing 데이터는 trimming을 거친 뒤, Metaxa2 (v2.2.3) [37] 를 사용하여 미생물 군집을 분석하였다. 각각의 tool로 분류된 박테리아 read의 수는 16S rRNA 유전자가 포함되어 있는 read의 수로 normalization하였다.

2.3.3. ARG annotation

MinION sequencing 데이터에서 ARG를 가지고 있는 read를 선별하기 위해 Prodigal (v2.6.3) [38] 로 open reading frame (ORF)을 예측하여 amino acid sequence로 변환한 뒤, comprehensive antibiotic resistance database (CARD) [39] 와 DIAMOND [40] 를 이용하여 BLASTP [41] 를 실행하였다. 본 연구에서는 CARD의 최신 버전을 다운로드한 뒤, 유전자가 2개 이상의 항생제 계열에 내성을 가지고 있는 것을 모두 multidrug resistance (MDR)로 수정하였다. 75 %의 minimum identity (--id 75)로 선별한 ARG-carrying read의 abundance (gene copies/Gb)는 아래의 공식(Eq. (1))으로 계산하였다 [42]. Equation (1)에서 L은 length, 즉 sequence와 read의 길이를 의미하며, n과 m은 각각 total read와 ARG-like sequence의 개수를 의미한다.

$$ARG\ abundance = \frac{\sum_1^m \frac{L_{ARG-like\ sequence}}{L_{ARG\ reference\ sequence}}}{\frac{\sum_1^n L_{reads}}{10^9\ bp/Gb}} \quad (1)$$

HiSeq sequencing 데이터에서 ARG를 가지고 있는 read를 선별하기 위해 CARD [39] 와 DIAMOND [40] 를 이용하여 BLASTX [41] 를 실행하였다. 90 %의 minimum identity (--id 90)로 선별한 ARG-carrying read의 abundance는

아래의 공식 (Eq. (2))으로 계산하였다 [43]. Equation (2)에서 N은 sequence의 개수를 의미하며, L은 Equation (1)에서와 마찬가지로 sequence와 read의 길이를 의미한다. 마지막으로 n은 total ARG-like sequence의 개수를 의미한다.

$$ARG\ abundance = \sum_1^n \frac{\frac{N_{ARG-like\ sequence} \times L_{reads}}{L_{ARG\ reference\ sequence}}}{\frac{N_{16S\ sequence} \times L_{reads}}{L_{16S\ sequence}}} \quad (2)$$

2.3.4. MGEs associated with ARGs detection

먼저, HiSeq의 short-read를 MinION의 long-read와 비교하기 위해 MEGAHIT (v1.2.9) [44] 로 assembly하여 contig를 생성했다(--min-contig-len 300). ARG가 포함된 contig를 선별하기 위해 Prodigal (v2.6.3) [38] 로 ORF를 예측하여 amino acid sequence로 변환한 뒤, CARD [39] 와 DIAMOND [40] 를 이용하여 BLASTP [41] 를 실행하였다. MinION의 long-read와 마찬가지로 minimum identity를 75 %로 설정하였고, 선별된 ARG-carrying contig를 MGE detection과 plasmid annotation에 사용하였다.

ARG-carrying read와 contig에 포함된 MGE를 검출하기 위해 MobileGenetic-ElementDatabase [45] 를 다운로드 한 뒤, DIAMOND [40] 를 이용하여 identity 90%로 BLASTP [41] 를 수행하였으며, ARG와 MGE 사이의

거리가 5 kb 미만인 ARG-MGE 쌍을 골라 분석에 사용하였다.

2.3.5. ARG-harboring plasmid annotation

ARG-carrying read와 contig 중 plasmid의 형태를 가지고 있는 것은 Plasclass (<https://github.com/Shamir-Lab/PlasClass>)로 선별하였다. 그 중에서 plasmid라고 예측되는(value ≥ 0.5) read와 contig를 선별한 뒤, CARD [39] 와 DIAMOND [40] 를 이용하여 BLASTP [41] 를 실행하였다. Minimum identity를 75 %로 설정하였고, 선별된 read와 contig를 ARG-harboring plasmid annotation에 사용하였다.

2.4. Statistical analysis

미생물 군집의 알파다양성과 베타다양성은 Mothur (v.1.48.0) [46] 를 사용하여 계산하였다(Fig. 4, Fig. 5). ARG의 absolute abundance는 각 site 당 3개의 샘플에 대한 평균값과 표준편차를 Excel로 계산하여 Figure 6의 error bar로 표시하였다. Drug class에 대한 Spearman 상관계수는 R (v.4.2.3)을 사용하여 계산하였다 (Table 4). 또한, heatmap과 principal component analysis (PCA) (Fig. 7B), 그리고 synteny (Fig. 11)는 R (v.4.2.3)을 사용하여 시각화 하였다.

3. RESULTS & DISCUSSION

3.1. Sequencing data summary

본 연구에서는 미호강에서 채수한 15개 샘플의 DNA를 추출하여 MinION sequencing과 HiSeq sequencing을 실시하였다. 두 플랫폼의 결과로 얻은 read를 trimming하여 각각 14,360,979개의 read (41,412,971,565 bp)와 961,932,490개의 read (145,251,783,590 bp)를 분석에 사용하였다(Table 3).

Table 3. Summary of sequencing results.

SampleID	The No. of total reads		Total base pair (bp)		Total data (Gb)	
	MinION	HiSeq	MinION	HiSeq	MinION	HiSeq
MH08_1	1,378,903	64,683,517	2,477,781,213	9,767,209,515	5.2	24
MH08_2	1,215,505	62,843,851	3,540,207,889	9,489,420,046	7.1	22
MH08_3	1,361,967	50,792,965	2,756,530,215	7,669,736,540	5.7	17.8
MH09_1	838,413	70,469,921	2,456,961,539	10,640,956,381	4.9	26
MH09_2	907,278	56,332,600	2,808,461,926	8,506,221,321	5.6	18.8
MH09_3	839,318	59,991,696	2,533,372,299	9,058,744,699	5.1	22
SY01_1	1,213,953	53,548,841	2,668,562,104	8,085,873,731	5.5	18.8
SY01_2	1,057,868	64,397,078	2,588,116,704	9,723,957,280	5.3	24
SY01_3	744,751	61,193,029	2,573,574,711	9,240,145,937	5.1	22
BR01_1	562,618	65,308,491	2,219,292,900	9,861,580,570	4.4	24
BR01_2	635,499	66,808,443	2,143,291,573	10,088,073,299	4.3	24
BR01_3	542,105	73,770,559	1,956,272,532	11,139,352,773	3.9	26
MH10_1	1,078,302	75,381,927	3,945,544,189	11,382,669,201	7.8	28
MH10_2	893,709	77,103,111	3,086,785,789	11,642,568,011	6.2	28
MH10_3	1,090,790	59,306,461	3,658,215,982	8,955,274,286	7.4	22
Total	14,360,979	480,8961,932,490	41,412,971,565	145,251,783,590	83.5	348.4

3.2. Microbial community analysis

3.2.1. Taxonomical classification

MinION과 HiSeq sequencing의 결과에서 각각 32, 30개의 문(Phylum)이 확인되었고, 모든 site에서 Proteobacteria가 가장 많은 비율(각각 평균 53%, 41%)을 차지하는 것으로 나타났다(Fig. 3A). MinION의 결과에서는 Actinobacteria와 Bacteroidetes가 평균 11 %, 10.8 %로 그 뒤를 이었고, HiSeq의 결과에서는 Bacteroidetes와 Cyanobacteria가 평균 18.9 %, 17.6 %로 그 뒤를 이었다. 상위 6개의 문 중에 Proteobacteria, Bacteroidetes, Actinobacteria, Cyanobacteria는 주로 환경에 존재하는 박테리아 문으로 알려져 있으며, 본 연구의 결과는 이전에 발표된 연구들의 결과와 비슷하였다 [47-49]. MinION과 HiSeq의 결과에서 강(Class)은 각각 62, 56개, 목(Order)은 각각 156, 130개, 과(Family)는 각각 335, 276개, 속(Genus)은 각각 911, 841개가 검출되었다(Fig. 3). 속 수준에서 좀 더 자세히 살펴보자면, MinION의 결과에서는 Others 그룹이 61.09 %로 가장 많은 비율을 차지하였으며, *Limnohabitans*, *Pseudomonas*, *Burkholderia*가 각각 7.22 %, 2.77 %, 2.13 %로 그 뒤를 이었다. HiSeq의 결과에서는 Unclassified 그룹이 63.27 %로 가장 많은 비율을 차지하였으며, *Synechococcus*, *Limnohabitans*, *Flavobacterium*가 각각 8.71 %,

4.41 %, 2.78 %로 그 뒤를 이었다(Fig. 3E). 두 플랫폼에서 공통으로 검출된 *Limnohabitans* 속은 환경에 서식하는 박테리아 속으로 알려져 있고, 최근 발표된 논문에 따르면, *Limnohabitans* 속 박테리아가 인간에게 ARG를 전달할 가능성이 높다는 사실이 밝혀졌다 [50]. 2019년과 2021년에는 sequencing 플랫폼에 따라 분석 방법이 다양하기 때문에 미생물 군집 또한 다르게 나타난다는 연구 결과가 발표되었고 [18, 51], 이와 비슷하게 본 연구에서는 속 수준의 Unclassified 와 Others 그룹에서 sequencing 플랫폼의 차이가 뚜렷하게 나타났다. MinION의 결과에서는 Unclassified 그룹의 비율이 8.72 %에 불과한 반면, HiSeq의 결과에서는 Unclassified 그룹이 63.27 %를 차지하였다. 이는 read의 길이가 짧을수록 16S rRNA 유전자에 높은 identity로 align이 되지 않아 분류가 어려워진 것으로 해석될 수 있다 [52]. 뿐만 아니라, short-read의 경우 어떤 박테리아에서 유래했는지를 명확히 구분하기가 어려워, 많은 read가 Unclassified 그룹으로 분류되었을 것으로 추정된다 [53]. 이러한 차이는 MinION과 HiSeq의 데이터를 분석한 classifier 간의 차이와 read 길이에 따른 영향으로 인해 나타난 것으로 보인다. 본 연구의 결과를 통해 read의 길이는 미생물 군집 분석에서 고려해야 할 중요한 요소임을 시사한다.

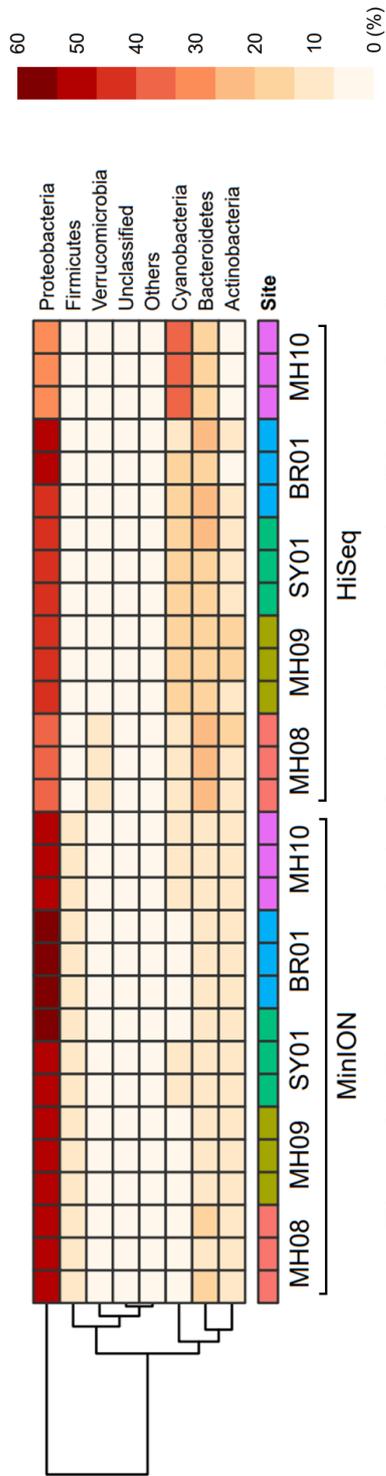


Figure 3A. Heatmap analysis of microbial community (Phylum level).

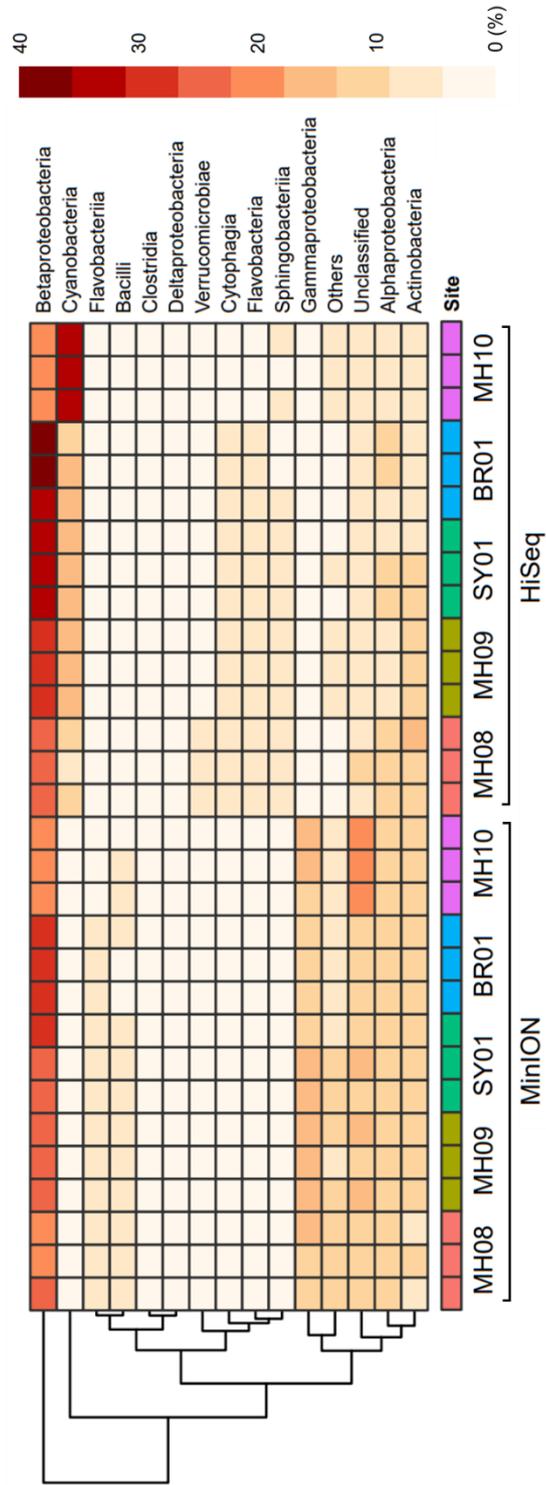


Figure 3B. Heatmap analysis of microbial community (Class level).

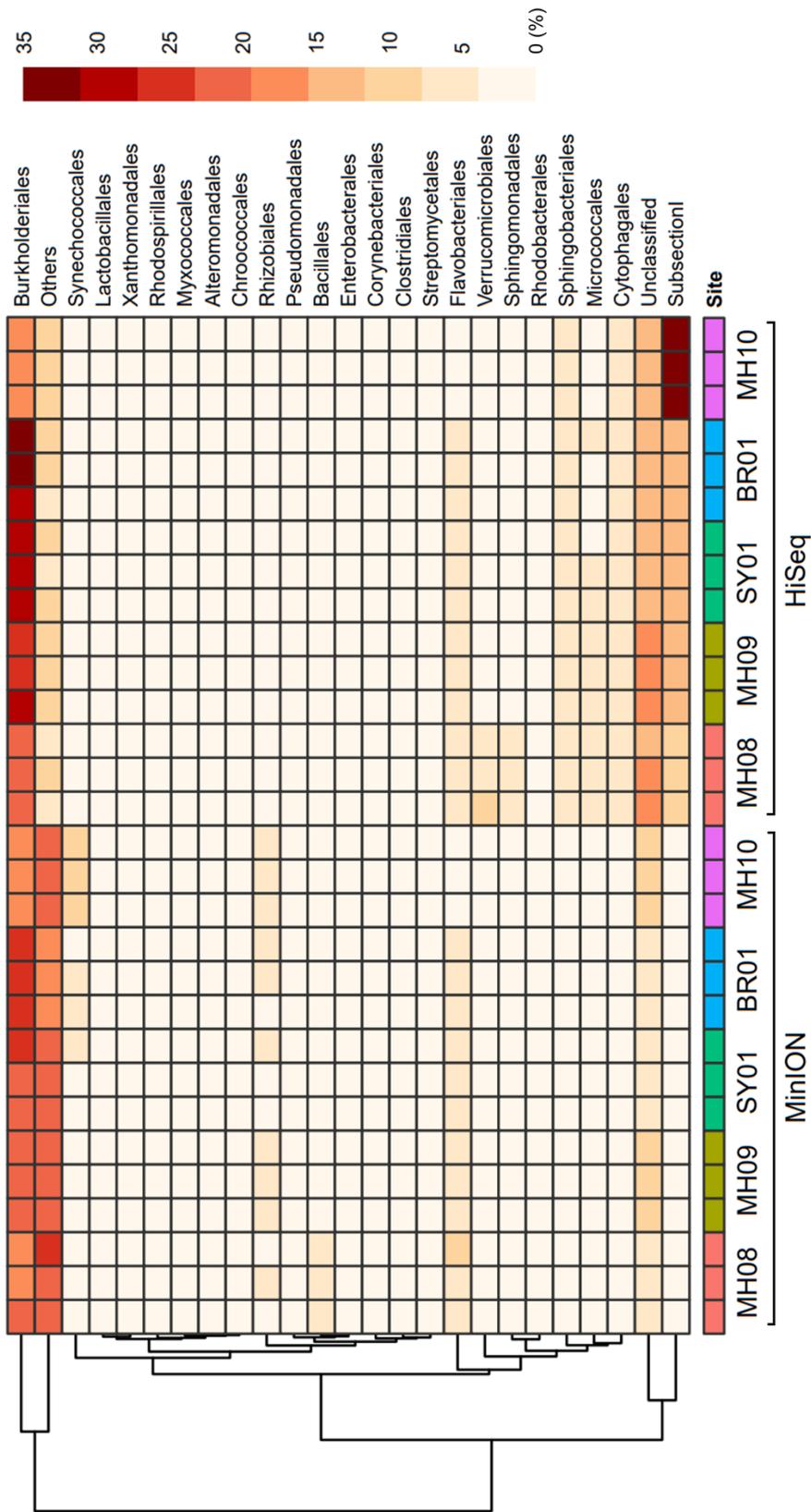


Figure 3C. Heatmap analysis of microbial community (Order level).

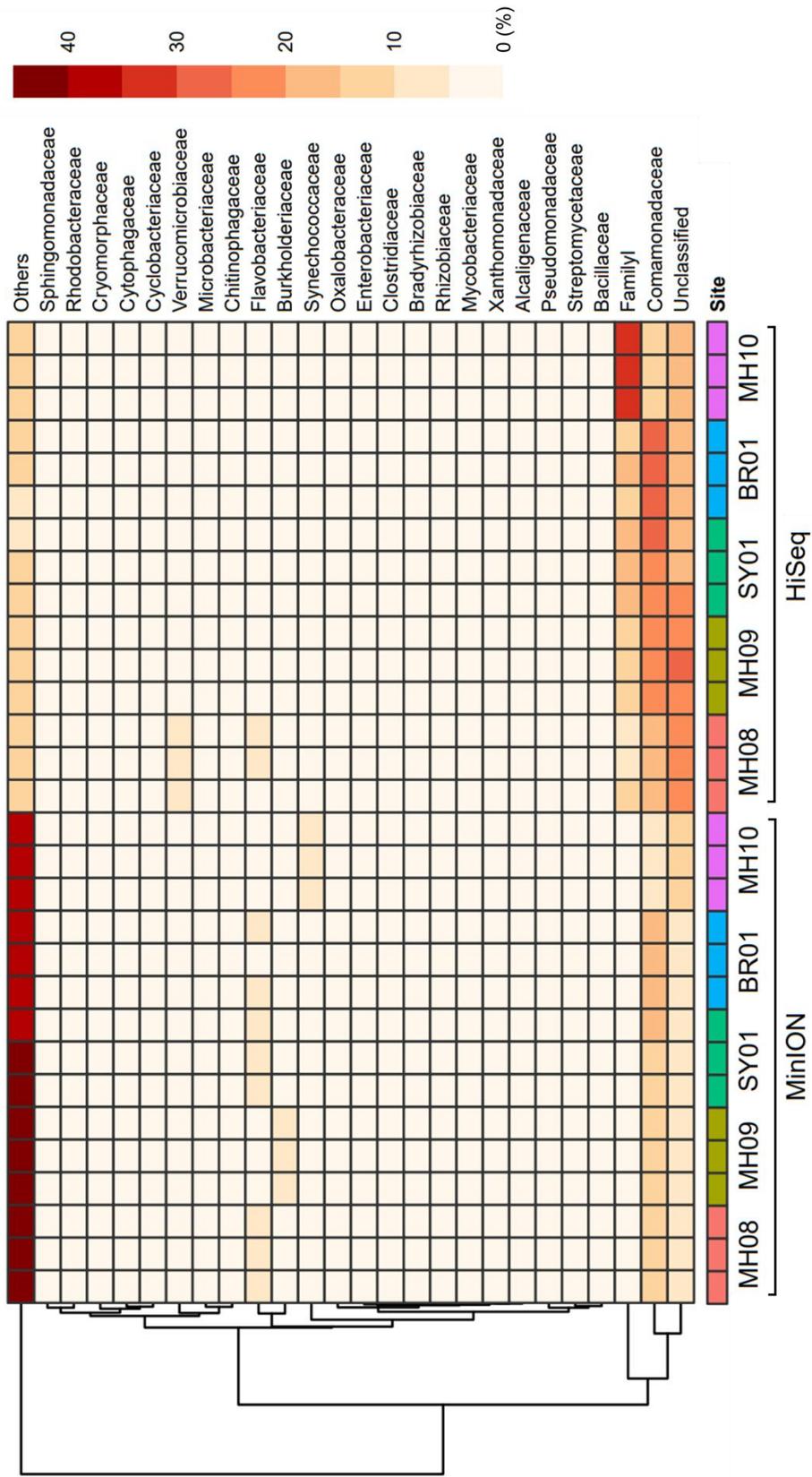


Figure 3D. Heatmap analysis of microbial community (Family level).

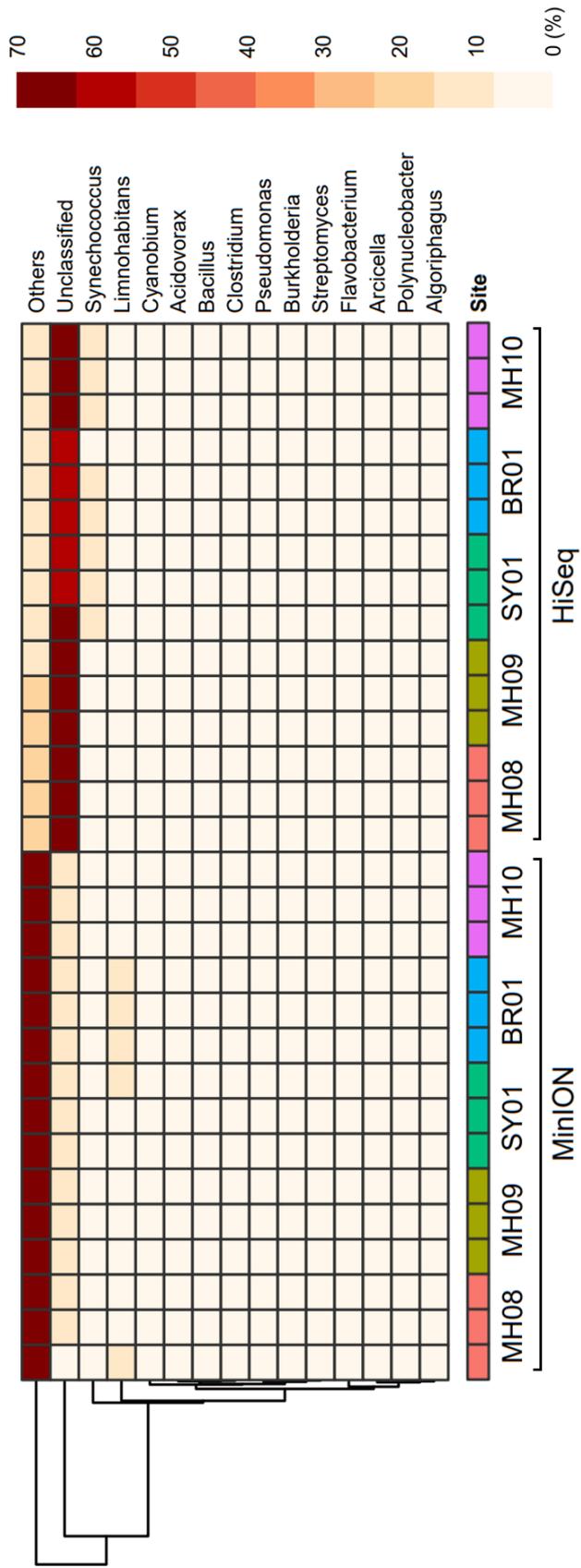


Figure 3E. Heatmap analysis of microbial community (Genus level).

3.2.2. Alpha and beta diversity

미생물 군집의 속 수준에서의 알파다양성은 Chao1과 Shannon 지수를 통해 확인하였고(Fig. 4), 베타다양성은 nMDS (Fig. 5) 분석을 통해 확인하였다. MinION과 HiSeq의 결과에서 Chao1 지수는 각각 평균 851.5과 318.6으로, Shannon 지수는 각각 평균 5.41과 2로 나타났다(Fig. 4). 이는 MinION sequencing이 보다 다양한 종류의 박테리아를 분류할 수 있음을 의미하며, MinION의 결과에서 문 수준에서 속 수준으로 갈수록 Others 그룹의 비율이 점차 증가한다는 본 연구의 결과(Fig. 3)를 뒷받침한다. 비배양성 박테리아가 많이 존재하는 환경 샘플의 경우, sequence 기반의 연구가 필수적이다 [54]. 이에, 앞으로 long-read sequencing을 활용한 미생물 분류는 환경 내에 존재하는 수많은 미생물 군집과 그들이 전파하는 항생제 내성체에 대한 연구에 기여할 것으로 기대된다.

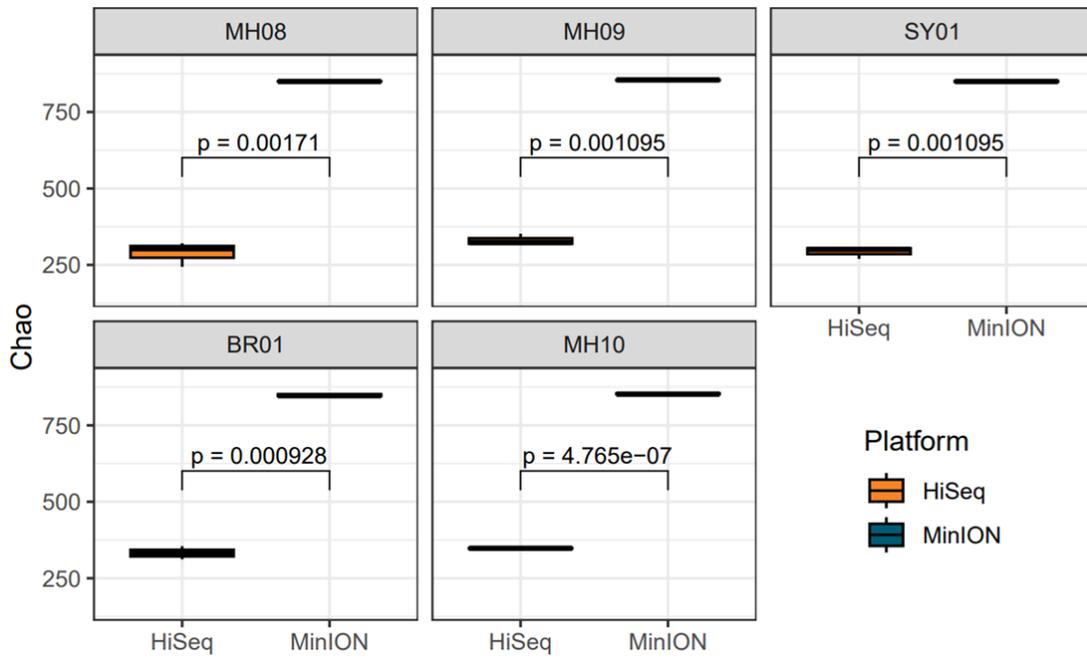


Figure 4A. Alpha diversity analysis (Chao1) for microbial community.

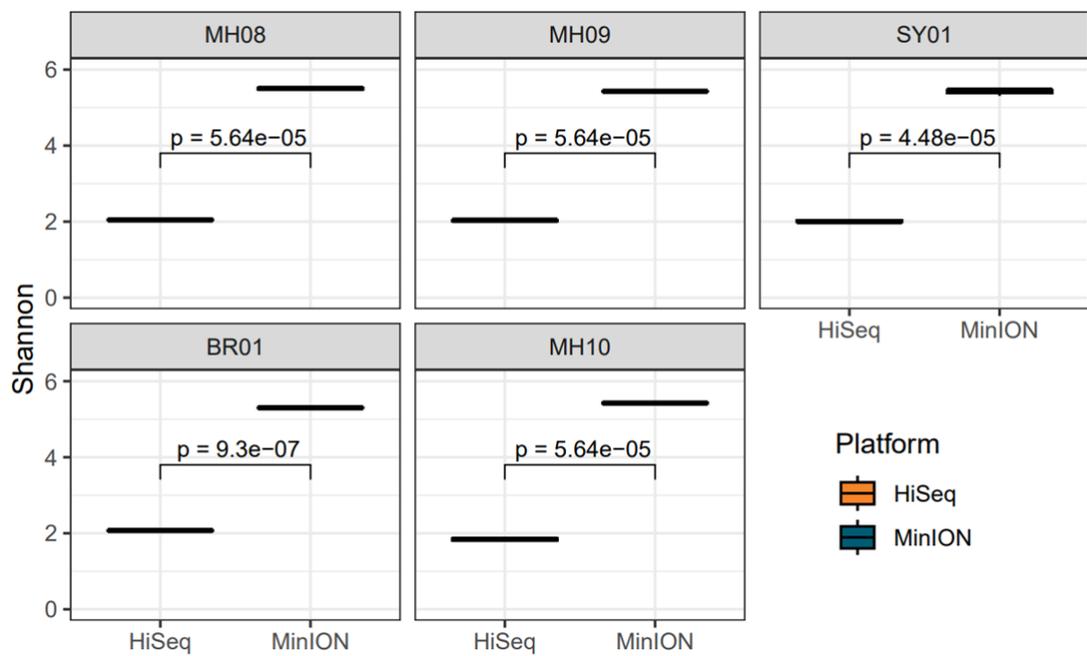


Figure 4B. Alpha diversity analysis (Shannon) for microbial community.

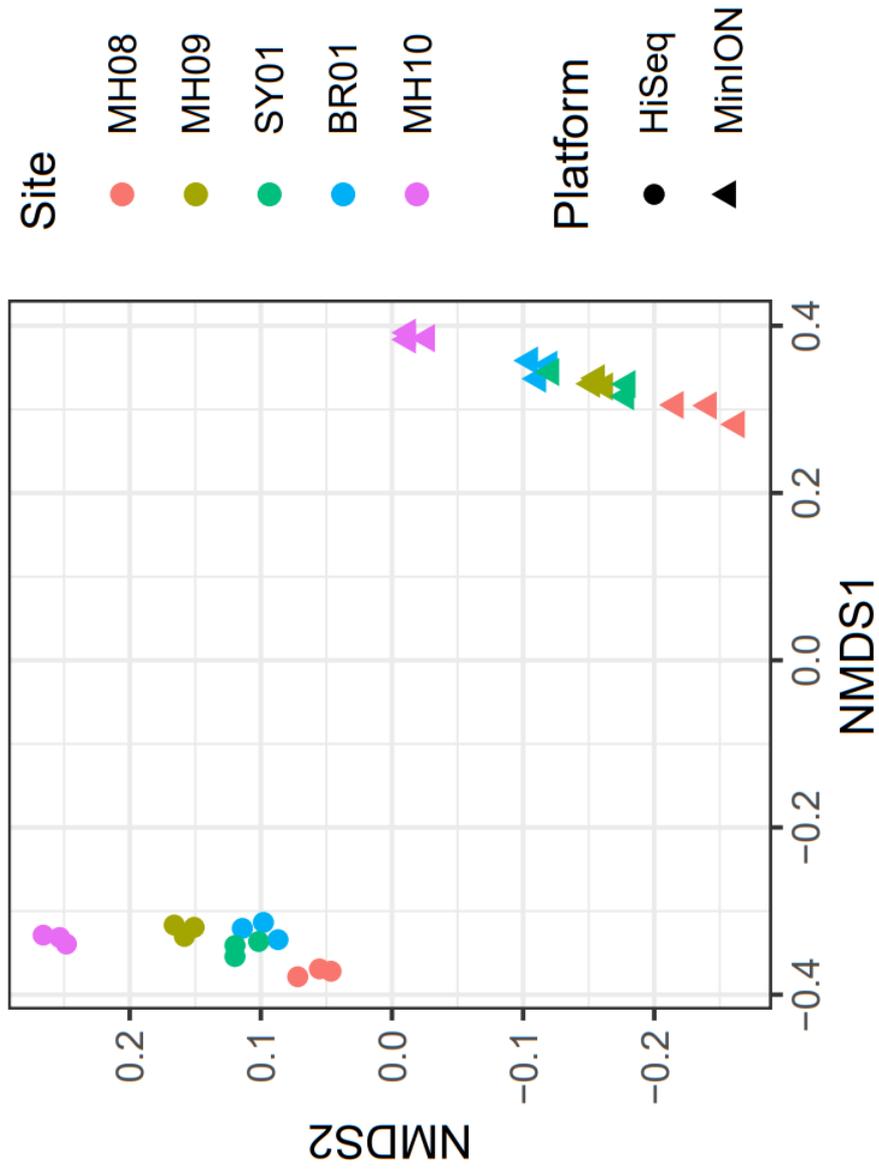


Figure 5. Beta diversity analysis (nMDS) for microbial community.

3.3. Abundance of ARGs

3.3.1. Absolute abundance of ARGs

미호강에 존재하는 ARG를 검출하기 위해 sequencing 데이터를 modified CARD에 align하였다. Illumina sequencing은 Nanopore sequencing에 비해 ARG를 높은 정확도로 검출할 수 있다는 연구 결과가 있었으며 [32], 각 플랫폼의 결과를 CARD에 align하는 과정에서 minimum identity를 조절하여 ARG의 가지수를 비슷하게 검출할 수 있도록 하였다. 그 결과, MinION의 데이터에서는 18개의 drug class에 속한 300가지의 ARG가 검출되었으며, HiSeq의 데이터에서는 16개의 drug class에 속한 395가지의 ARG가 검출되었다. 각 플랫폼에 적합한 공식(Eq. (1), Eq. (2))을 사용하여 abundance를 계산한 결과, long-read에서 검출된 ARG의 abundance가 short-read에서 검출된 ARG의 abundance보다 큰 것으로 나타났다(Fig. 6). 또한, MH09_1 그룹을 제외한 모든 그룹에서 Spearman 상관계수가 0.5 이상으로 나타났으며, 이는 두 플랫폼 간의 drug class 별 abundance가 일반적인 양의 상관관계를 갖고 있음을 의미한다 (p-value <0.05) (Table. 4). Long-read와 short-read에서 검출된 ARG의 절대량은 차이가 있었으나, ARG의 종류와 분포 양상은 두 플랫폼의 결과가 비슷한 것을 확인하였다.

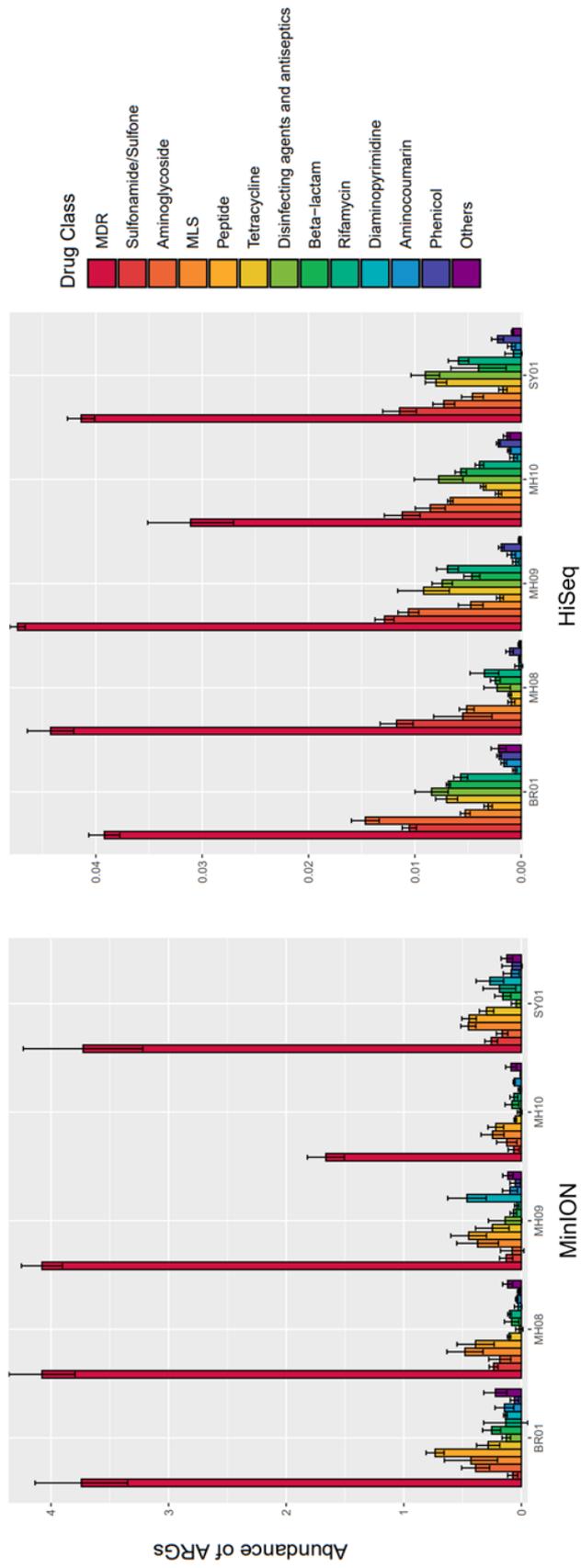


Fig. 6. Absolute abundance of ARGs. The error bars represent the standard deviation.

Table 4. Spearman rho correlation coefficients for comparison between long- and short-read sequencing.

Group	Spearman (rho)	p-value
MH08_1	0.7204752	0.0007444
MH08_2	0.7109574	0.000941
MH08_3	0.7533673	0.0003066
MH09_1	0.472367	0.04777
MH09_2	0.6155754	0.006536
MH09_3	0.6384758	0.004348
SY01_1	0.5618217	0.01525
SY01_2	0.5696493	0.01359
SY01_3	0.7666848	0.0002059
BR01_1	0.6244835	0.005599
BR01_2	0.681916	0.001826
BR01_3	0.7046051	0.001095
MH10_1	0.7311288	0.0005662
MH10_2	0.6844008	0.00173
MH10_3	0.5875648	0.01034

3.3.2. Relative abundance of ARGs

MinION과 HiSeq의 결과에서 각각 18개와 16개의 drug class가 확인되었고, 평균 1% 미만으로 검출된 것은 Others 그룹으로 분류되었다. MinION과 HiSeq의 결과에서 가장 많이 검출된 drug class는 multidrug resistance (MDR)로, 각각 평균 62.13 %, 43 %를 차지하였다(Fig. 7A). MinION 결과에서는 peptide 및 macrolides-lincosamides-streptogramins (MLS) 계열에 내성을 가진 ARG-carrying read가 각각 7.96 %, 7.44 %로 MDR 다음으로 많이 나타났다. HiSeq의 결과에서는 sulfonamide/sulfone, aminoglycoside 계열에 내성을 가진 ARG-carrying read가 각각 12.23 %, 9.53 %로 MDR 다음으로 많이 나타났다. Long-read는 short-read에 비해 다양한 class의 ARG를 갖게 될 가능성이 높아 대부분 MDR로 분류된 것으로 보인다 [6]. 본 연구 결과를 통해 sequencing data를 활용하여 ARG를 검출하는 과정에서 read의 길이에 따라 다양한 결과가 나타나는 것을 확인하였으며, 이는 환경 내에 존재하는 ARG의 abundance를 정확하게 이해하고 해석하기 위해 read의 길이를 고려하는 것이 필수적임을 시사한다.

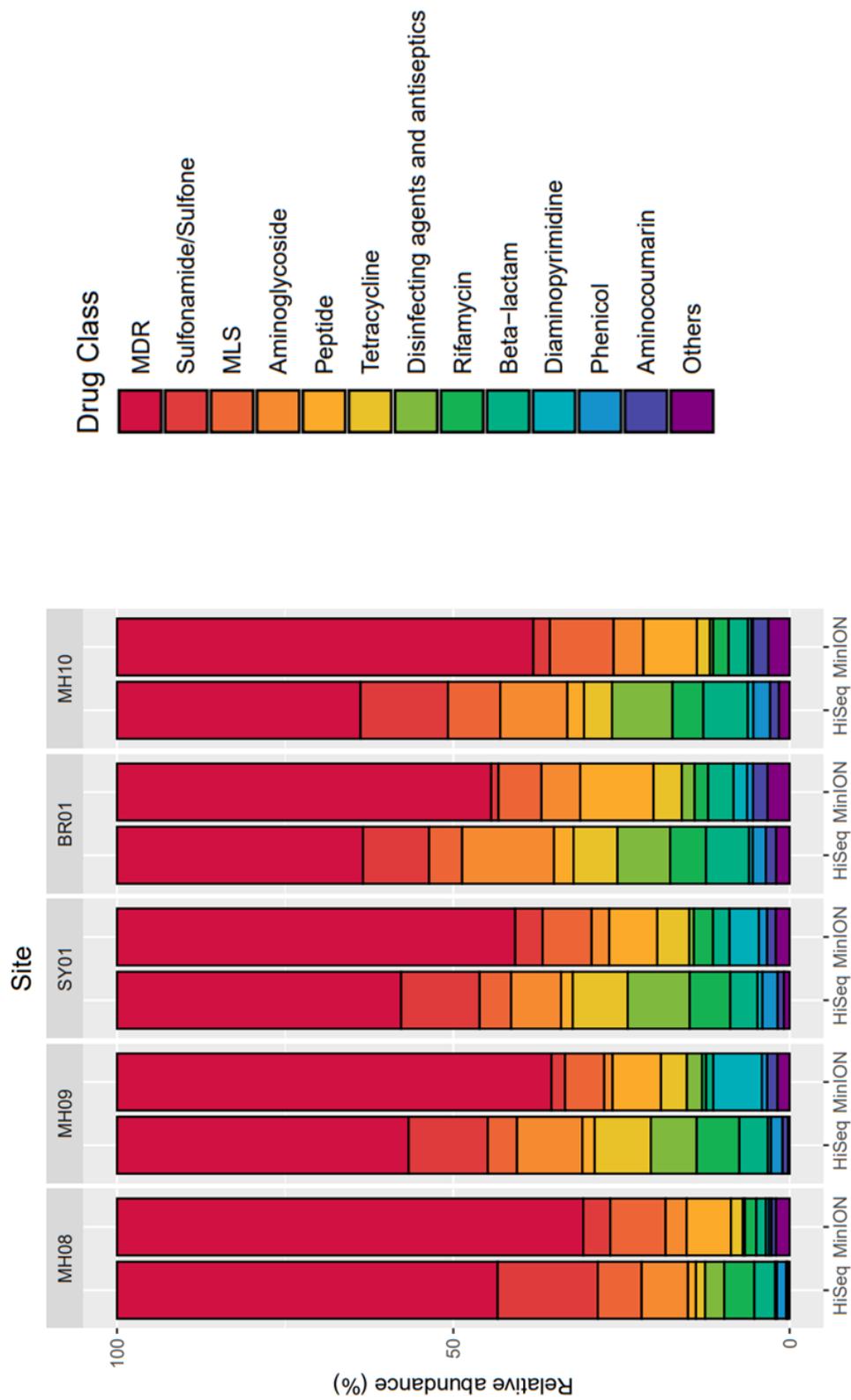


Figure 7A. Relative abundance of ARGs.

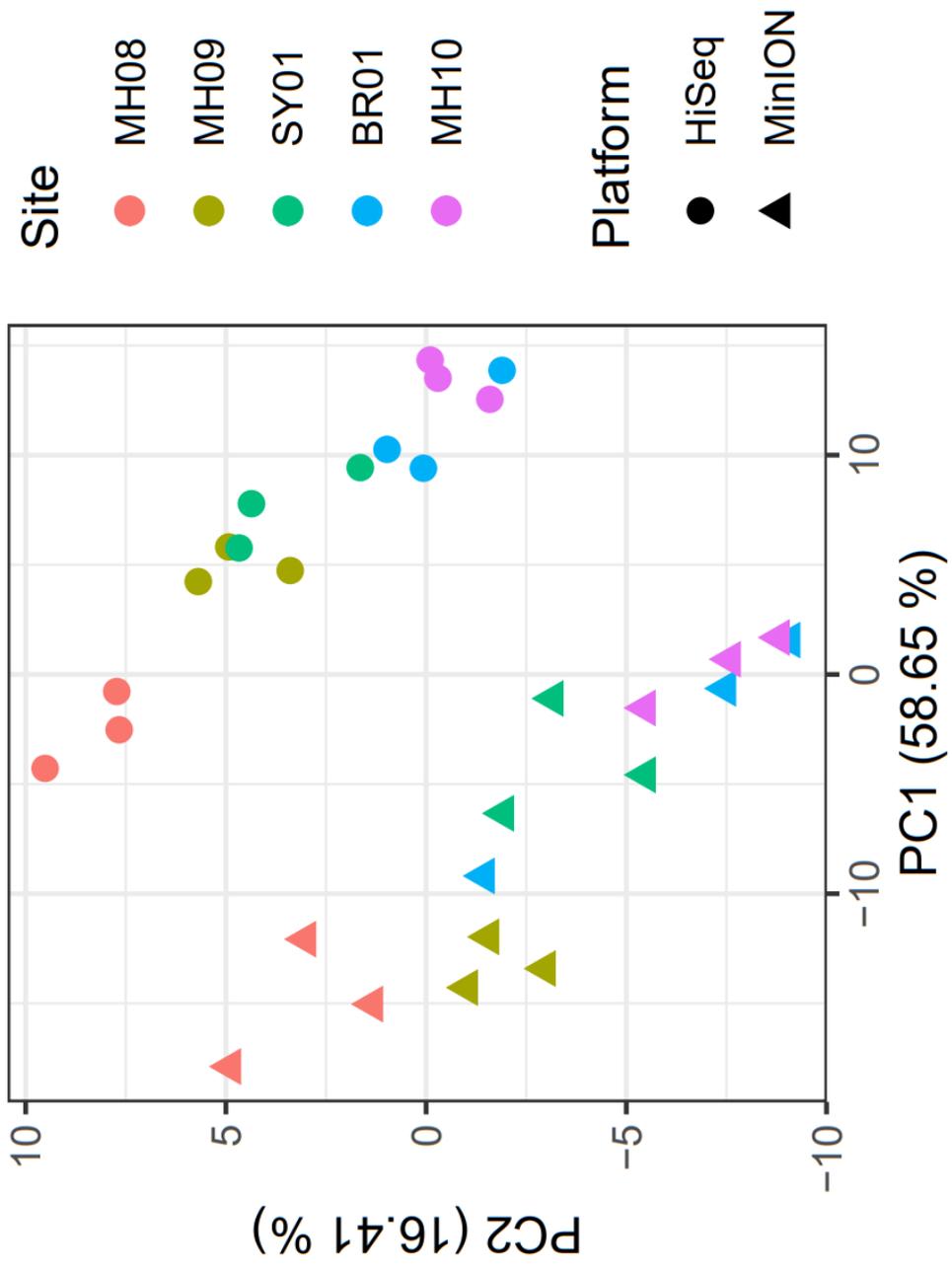


Figure 7B. Principal component analysis (PCA) based on the relative abundance of ARGs.

3.3.3. Overlapping ARGs

미호강 상류 지점인 MH08에서부터 하류 지점인 MH10까지 모든 site에서 한 번 이상 검출된 ARG 중 두 플랫폼에서 공통으로 나타난 ARG는 총 29개였다(Fig. 8A). 이 중에서 MDR로 분류된 ARG가 17개로 가장 많이 검출되었으며, 그 다음으로는 Peptide 계열 ARG가 많이 검출되었다. 이 29개의 overlapping ARG의 relative abundance는 두 플랫폼에서 비슷한 비율로 검출되었지만, *rdpA*, *sul1*, *sul2*, *tet(C)* 유전자에서 약간의 차이가 나타났다(Fig. 8A). Sulfonamide/sulfone 및 Tetracycline 계열 ARG는 특히 integron과 transposon과 함께 빈번하게 이동한다는 연구 결과가 있었으며 [55, 56], long-read sequencing의 결과에서는 두 계열의 항생제 내성 유전자가 종종 나타나지 않을 수 있다고 보고되었다 [32]. 본 연구에서도 이전의 연구들과 비슷한 결과가 확인되었으며, 이는 ARG를 검출하는 데에 있어서 read의 길이가 영향을 미칠 수 있다는 점을 시사한다.

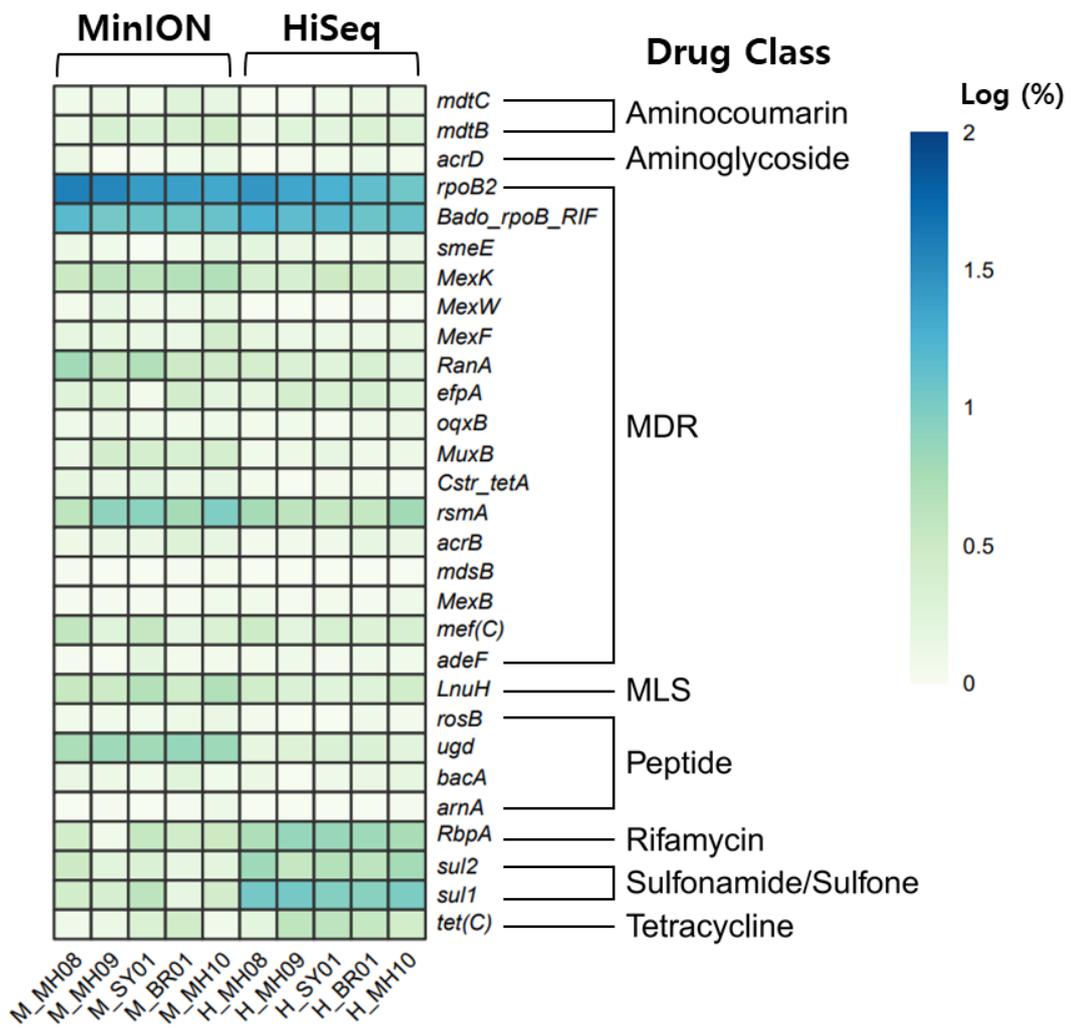


Fig. 8A. Relative abundance of overlapping ARGs.

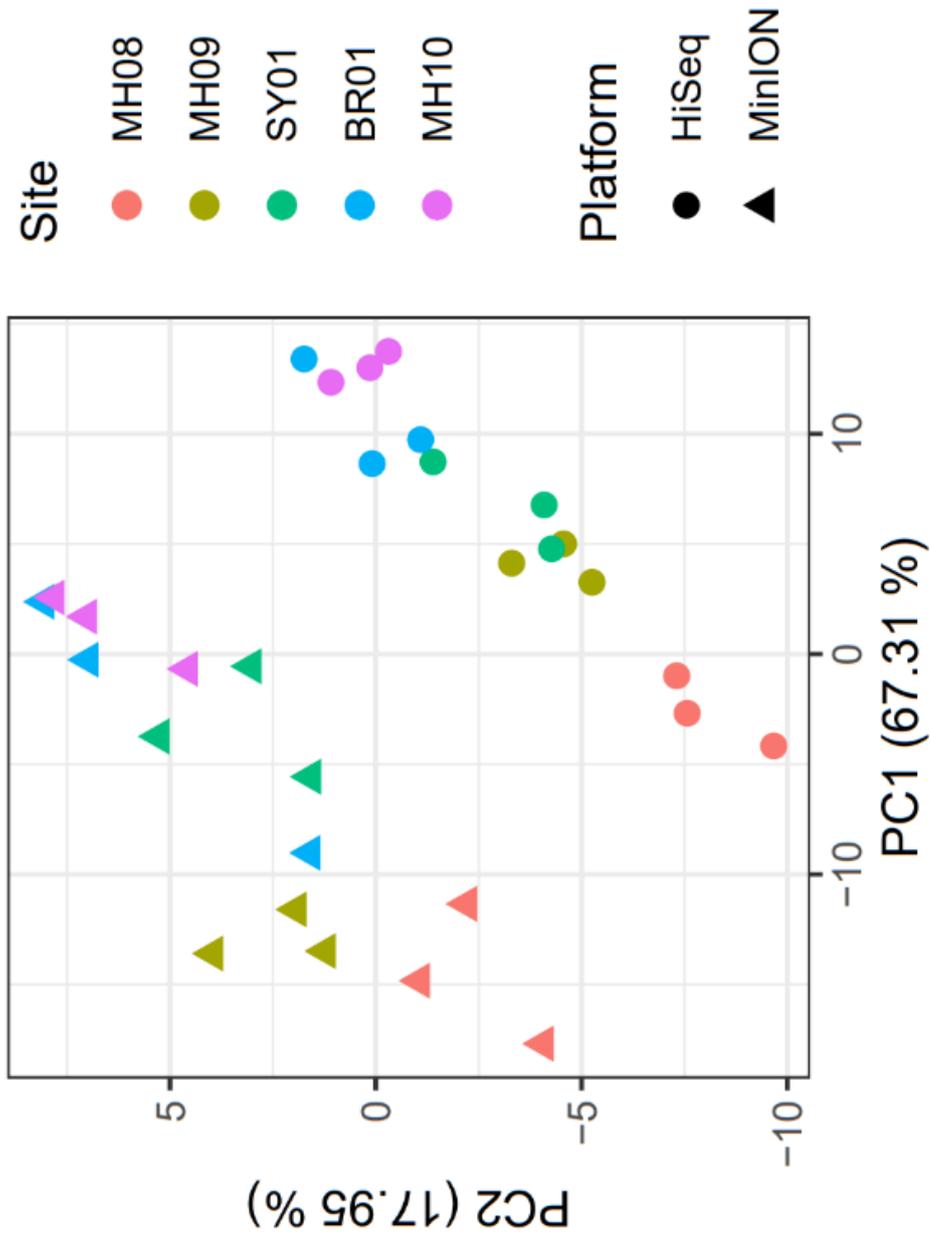


Fig. 8B. Principal component analysis (PCA) based on the relative abundance of overlapping ARGs.

3.4. Mobile genetic elements (MGEs) associated with ARGs

본 연구에서는 ARG-MGE 쌍을 검출할 때 assembly 과정의 유무에 따른 차이를 비교하기 위해 HiSeq의 short-read를 이용하여 contig를 생성한 뒤, 이를 MinION의 long-read와 비교하였다. MinION의 long-read는 대부분 contig보다 길다는 결과를 확인할 수 있었으며(Fig. 9), 이는 contig를 조립하는 과정에서 유전자 내 반복서열로 인해 short-read의 정확한 위치를 결정하기가 어렵기 때문인 것으로 추정된다.

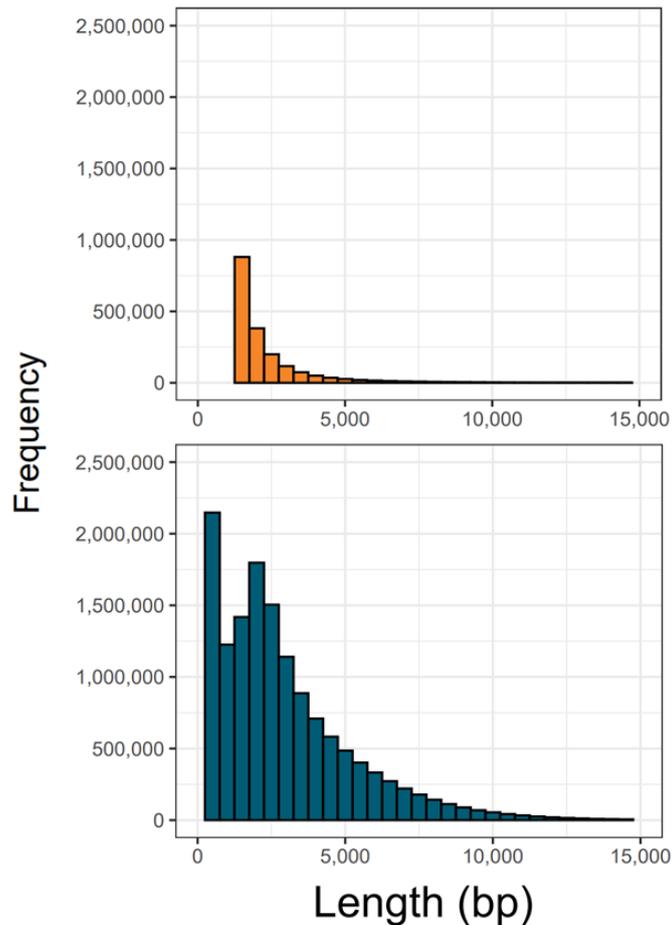


Figure 9. The length of ARG-carrying reads and contigs.

ARG를 가지고 있는 read와 contig를 선별하여 MGEs database에 align한 뒤, ARG와 MGE 사이의 거리가 5 kb 미만인 ARG-MGE 쌍에 대해 조사하였다. MinION의 결과에서는 54개의 read에서 69가지 종류의 ARG-MGE 쌍이 발견되었으며, HiSeq의 결과에서는 54개의 contig에서 25가지 종류의 ARG-MGE 쌍이 발견되었다. 두 플랫폼의 데이터에서 동시에 나타난 ARG-MGE 쌍은 총 13개이며, 이 중 8개 (*qacEdelta1*-qacEdelta, *sul1*-int11, *sul1*-qacEdelta, *sul1*-tnpA, *sul2*-IS91, *sul2*-tnpA, *tet(A)*-tnpA, *tet(G)*-qacEdelta)가 각 site마다 동시에 확인되었다(Fig. 10). Long-read와 contig 차이는 assembly 과정에서 contig가 MinION의 long-read와 다르게 조립되면서 발생한다 [57]. 이러한 이유로, long-read가 가진 ARG-MGE 쌍이 contig가 가진 ARG-MGE 쌍보다 더 다양하다는 결과가 나타난 것으로 예상된다. 또한, long-read를 활용한 분석에서는 raw read를 직접 사용했기 때문에 ARG와 MGE의 유전자 내 위치 정보가 보존되어 있어 여러 종류의 ARG-MGE 쌍이 검출된 것으로 보인다. 두 플랫폼에서 모두 검출된 8개의 ARG-MGE 쌍을 분석한 결과를 synteny로 시각화 하였다(Fig. 11). *Int11* 유전자는 유전자 이동에 관여하는 MGE이며 [58], 과거부터 박테리아 간 유전자 이동에서 중요한 역할을 하는 것으로 알려져 있다 [56, 59]. 본 연구 결과에 따르면 *int11* 유전자는 모든 site에서 ARG와 가까운 위치에 존재하는 것으로

확인되었다(Fig. 11). 특히, *sull* 유전자와 항상 같은 read 또는 같은 contig에 존재하는 것으로 나타났으며, 이는 이전의 다른 연구 결과와 비슷하다 [56, 60]. 또한, MGE 중 transposon의 한 종류인 *tnpA* 유전자는 MH10을 제외한 모든 site에서 검출되었으며, 주로 *sull* 유전자(Fig. 11A, 11C, 11D) 또는 *tet(A)* 유전자(Fig. 11B)와 같은 read 또는 contig에 존재하는 것으로 나타났다.

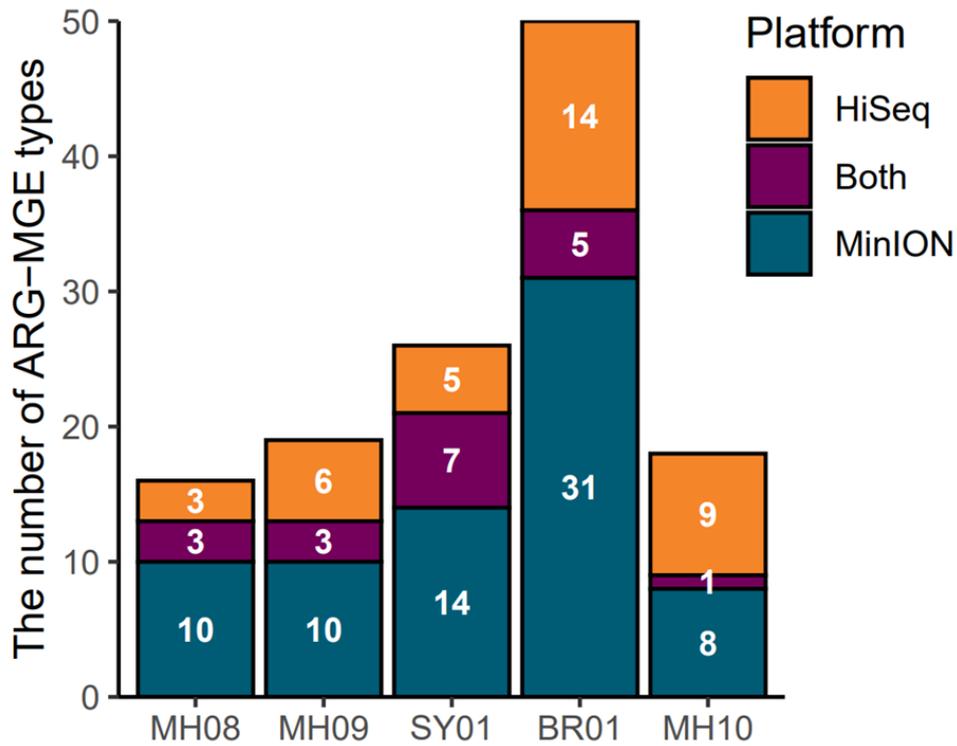


Figure 10A. The number of ARG-MGE types across the sampling sites.

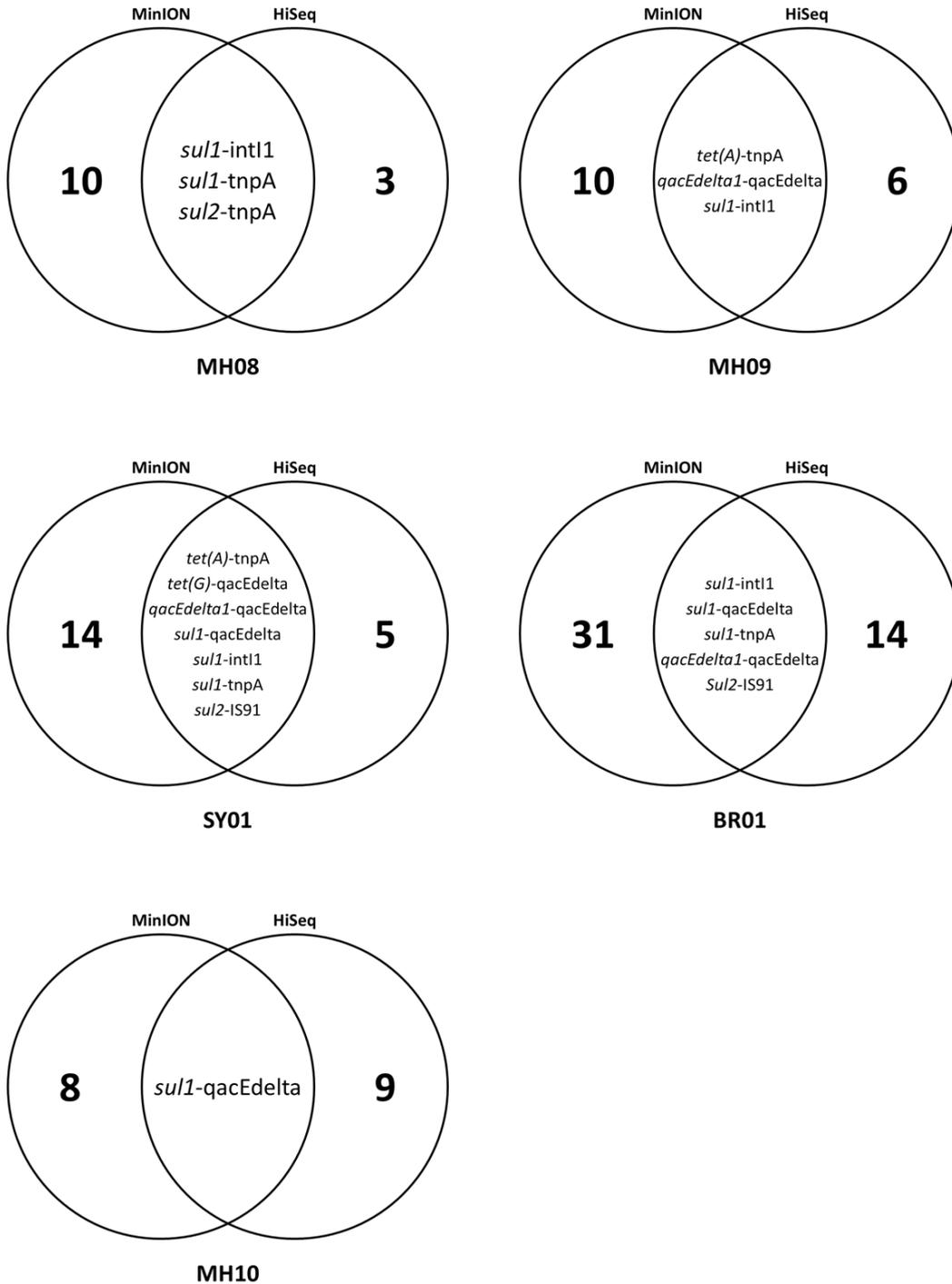


Figure 10B. ARG-MGE types identified in both MinION and HiSeq sequencing.

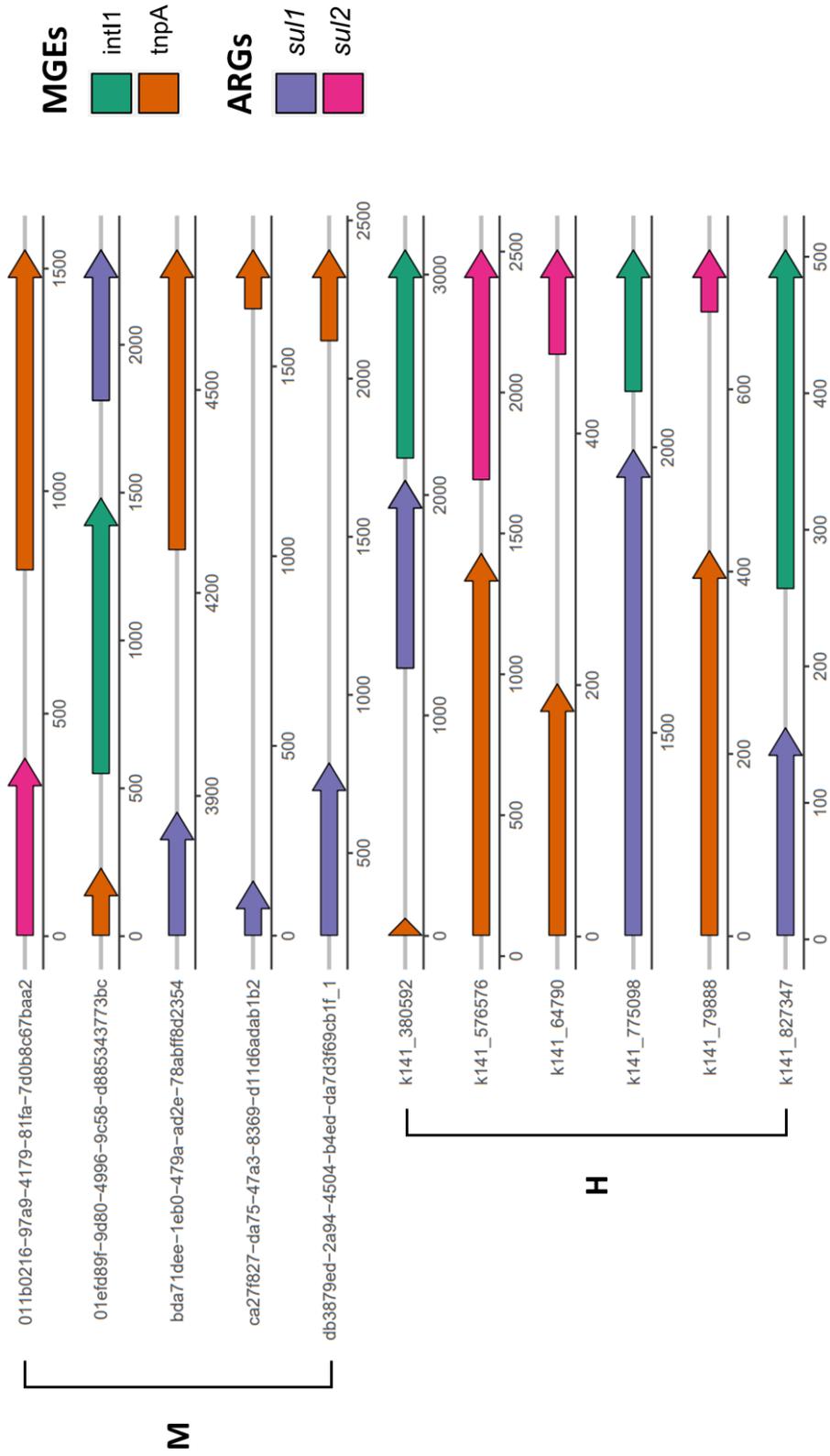


Figure 11A. Synteny analysis of ARG-MGE types at MH08.

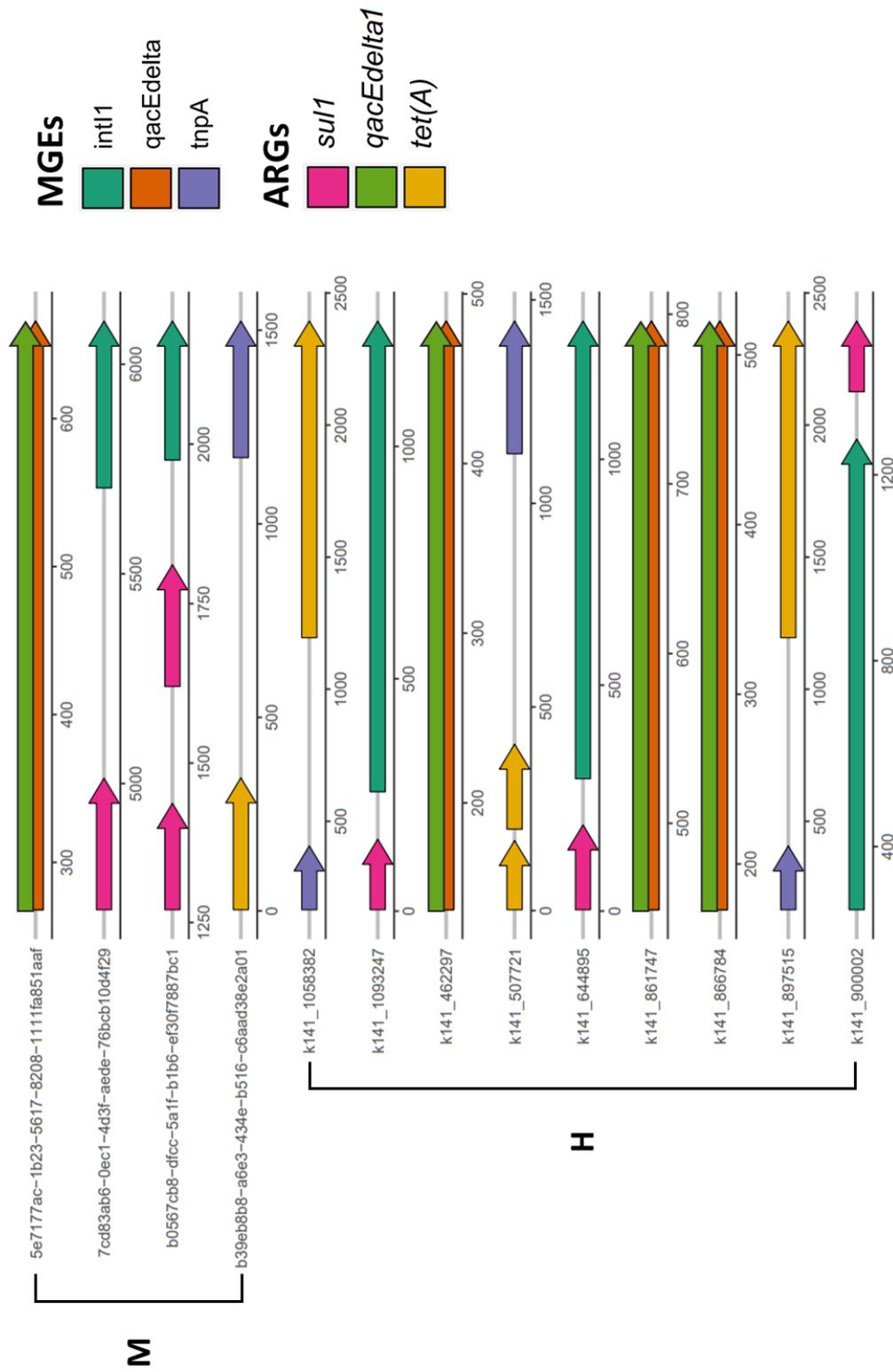


Figure 11B. Synteny analysis of ARG-MGE types at MH09.

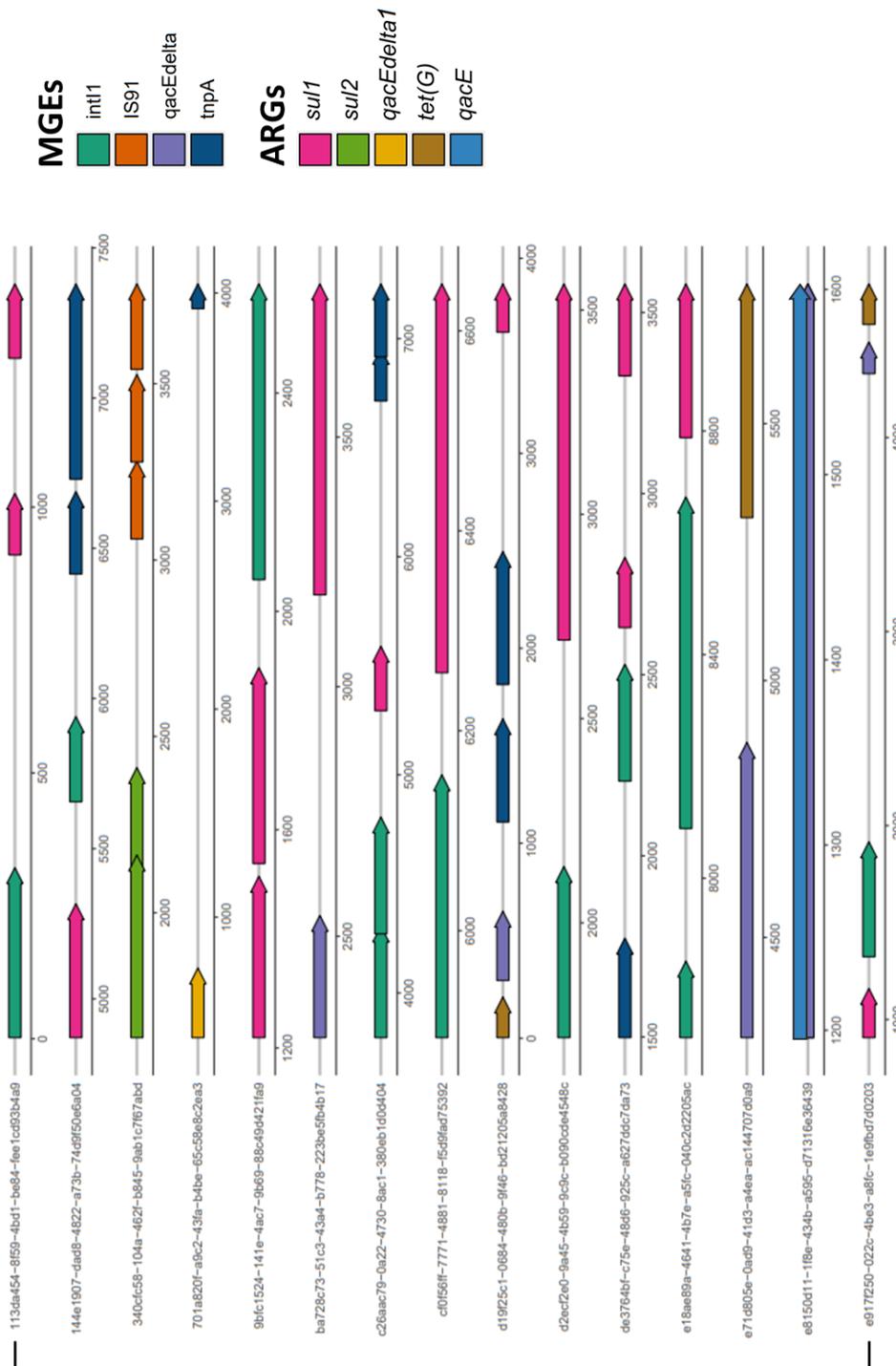


Figure 11C. Synteny analysis of ARG-MGE types at SY01.

M

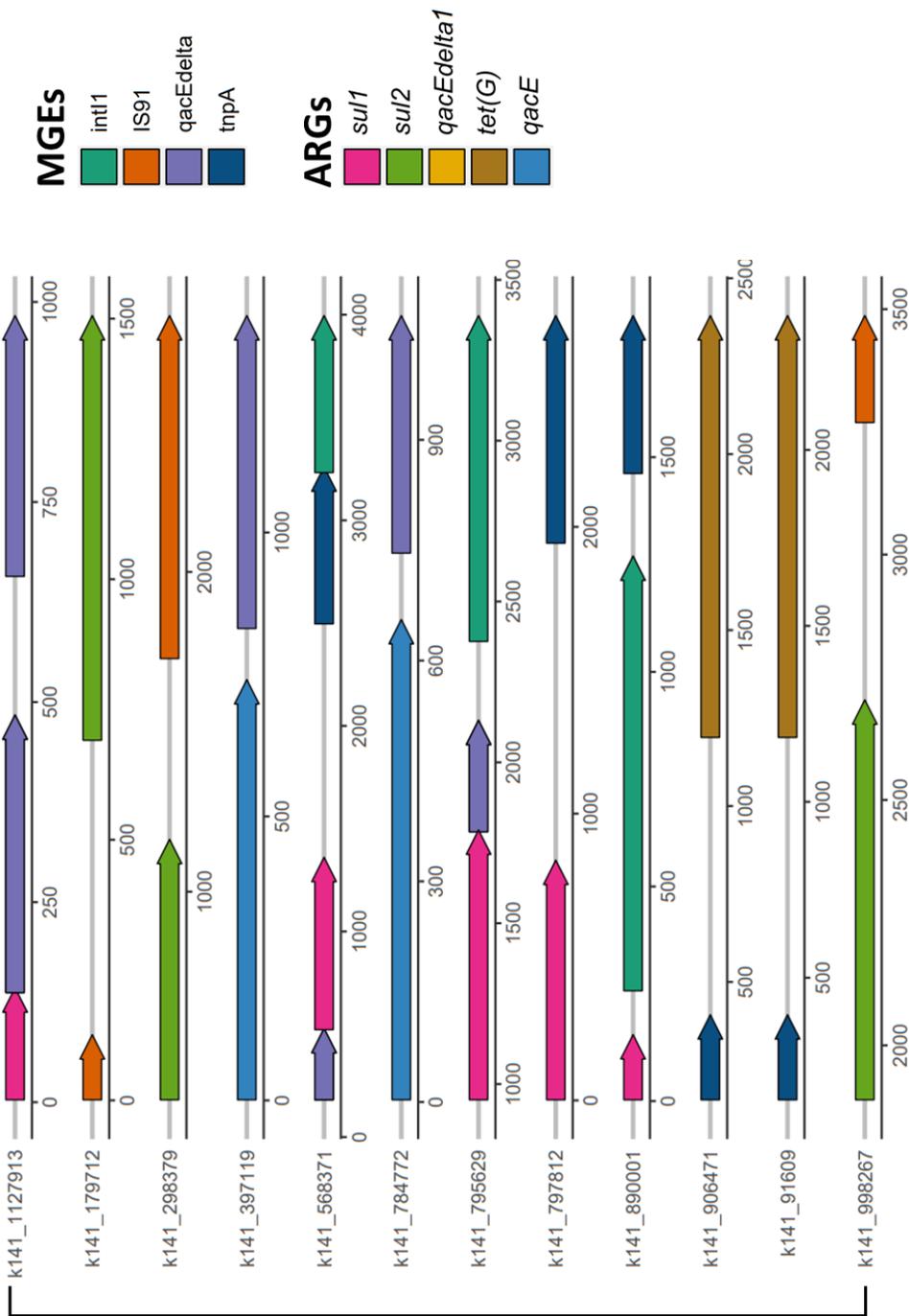


Figure 11C. Synteny analysis of ARG-MGE types at SY01.

H

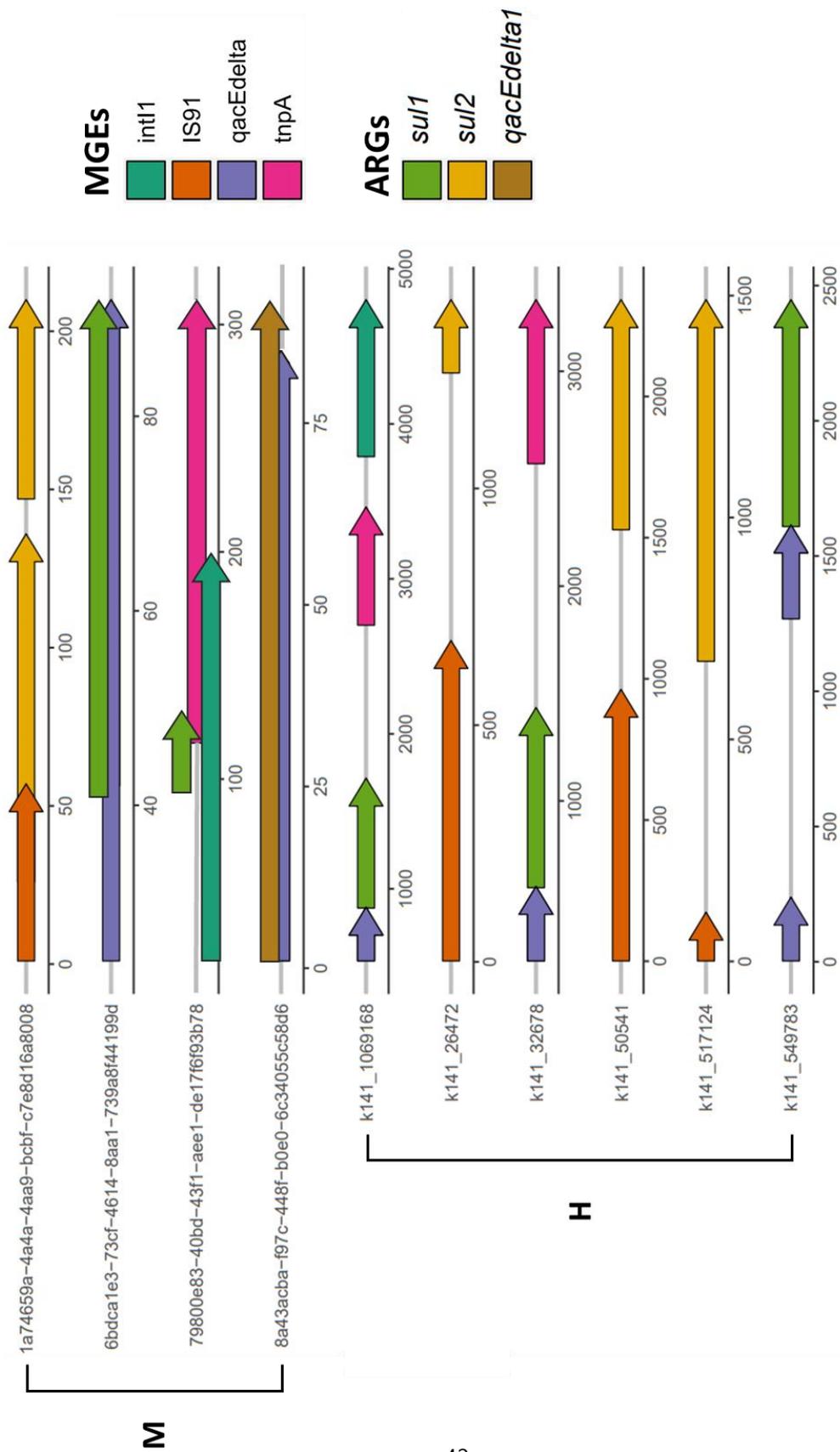


Figure 11D. Synteny analysis of ARG-MGE types at BR01.

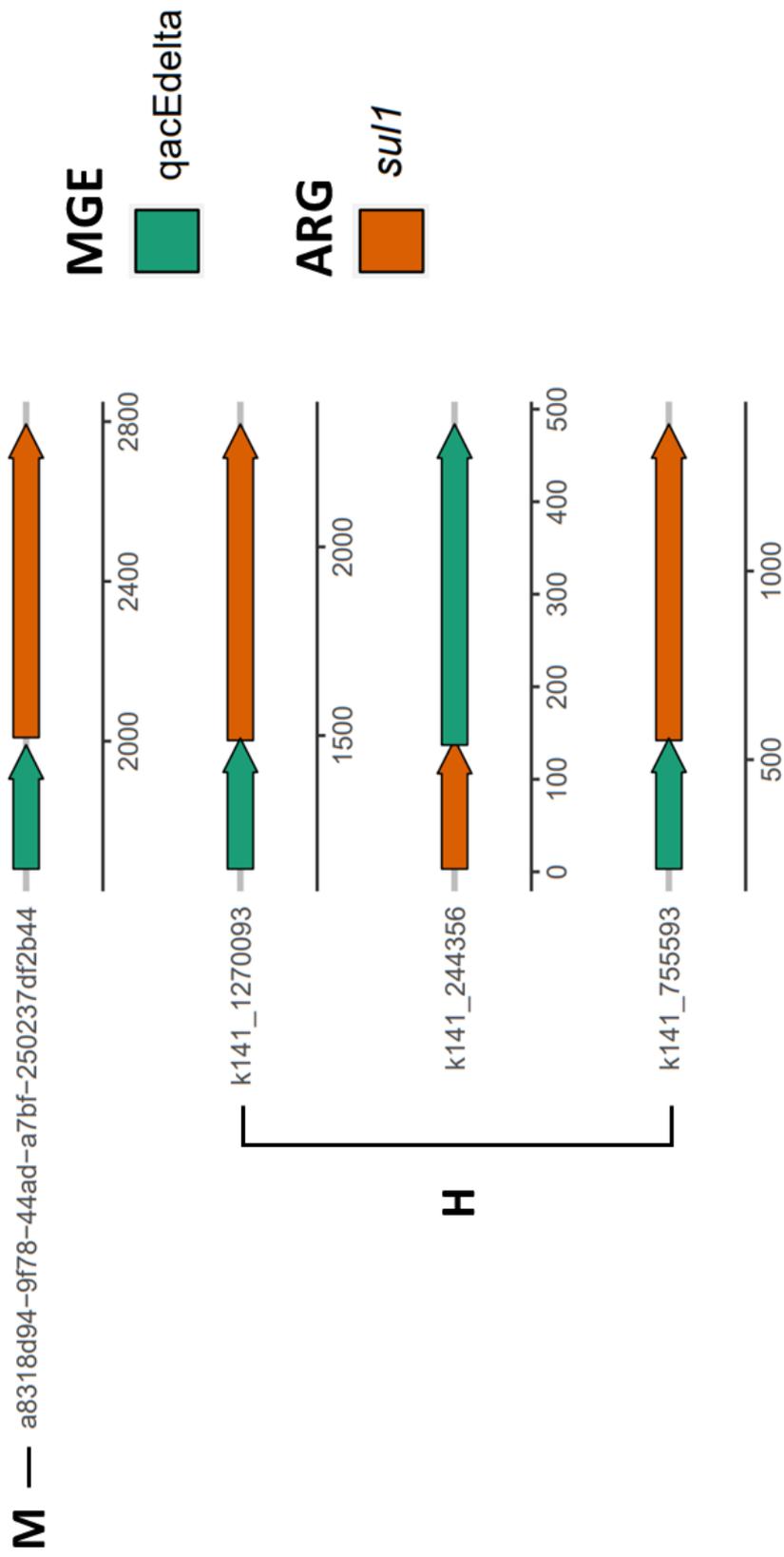


Figure 11E. Synteny analysis of ARG-MGE types at MH10.

3.5. ARG-harboring plasmids.

3.5.1. Abundance of ARGs on plasmids

본 연구에서는 plasmid에 존재하는 ARG의 종류를 확인하기 위해 PlasClass를 사용하여 ARG를 가지고 있는 read와 contig 중 plasmid의 형태를 띄는 것을 선별하였다. MinION의 결과에서는 183개의 ARG를 포함하는 1718개의 read가 plasmid인 것으로 나타났으며, HiSeq의 결과에서는 91개의 ARG를 포함하는 749개의 contig가 plasmid인 것으로 확인되었다. 19개의 ARG가 평균 1 % 이상으로 검출되었으며, 이 중 8개의 ARG의 class가 MDR인 것으로 확인되었다(Fig. 12). 특히 *rpoB2* 유전자는 MinION과 HiSeq의 데이터에서 각각 평균 33.52 %, 45.03 %를 차지하였다. *rpoB2* 유전자는 주로 담수에서 검출되며, 수많은 박테리아와 상관관계를 가지는 것으로 알려져 있다 [61]. Plasmid는 박테리아 사이에 직접적인 이동이 어려운 transposon이나 integron에 비해 ARG를 효과적으로 전달할 수 있는 MGE로 알려져 있다 [9]. 이러한 결과를 종합적으로 고려한다면, *rpoB2* 유전자는 plasmid를 통해 빠르게 환경으로 전파될 것으로 예상된다.

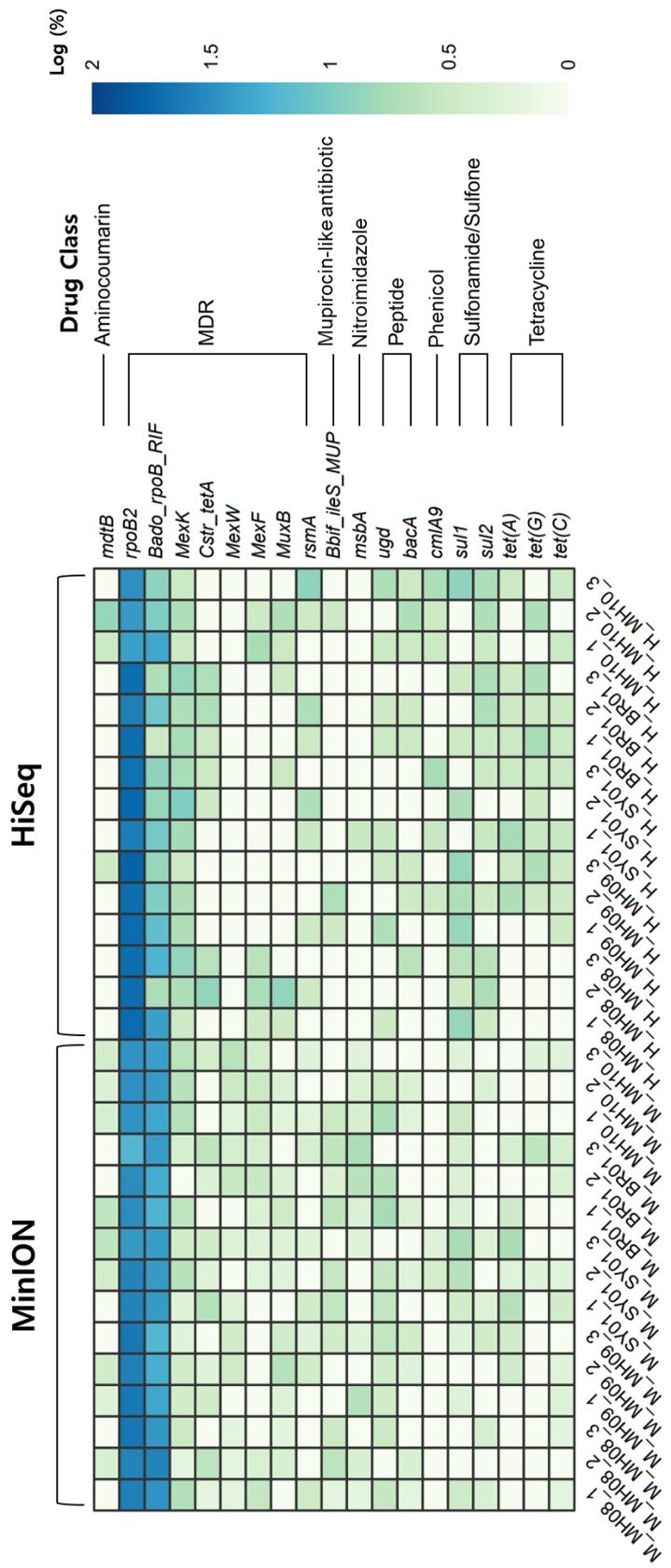


Figure 12. Relative abundance of ARGs on plasmids.

3.5.2. ARG-carrying reads and contigs in plasmids.

ARG-carrying read와 contig 중에 plasmid로 판별된 read와 contig를 drug class 별로 확인하였다(Fig. 13). MinION의 데이터에서는 ARG-carrying read 중에 sulfonamide/sulfone, tetracycline, aminoglycoside 계열의 순서대로 각각 평균 81.3 %, 75.4 %, 74.1 %의 read가 plasmid인 것으로 확인되었으며, HiSeq의 데이터에서는 전체 ARG-carrying contig 중에 aminoglycoside, tetracycline, sulfonamide/sulfone 계열의 순서대로 각각 평균 93.2 %, 84.3 %, 77.8 %의 contig가 plasmid인 것으로 확인되었다(Fig. 13). Aminocoumarin, mupirocin-like antibiotic, nitroimidazole 계열에서 MinION과 HiSeq 결과 간에 차이가 나타났다. MinION의 long-read는 가공되지 않은 read이기 때문에 polishing 과정을 여러 번 거쳐야 하는 short-read에 비해 상대적으로 plasmid 서열을 정확하게 파악할 수 있다 [62]. 이러한 이유로 인해, MinION의 결과를 활용한다면, ARG가 포함된 plasmid를 더 효과적으로 선별할 수 있을 것으로 예상된다. 따라서 long-read sequencing은 plasmid를 기반으로 하는 ARG의 전파를 모니터링하는 데에 유용한 도구로 활용될 수 있을 것으로 보인다.

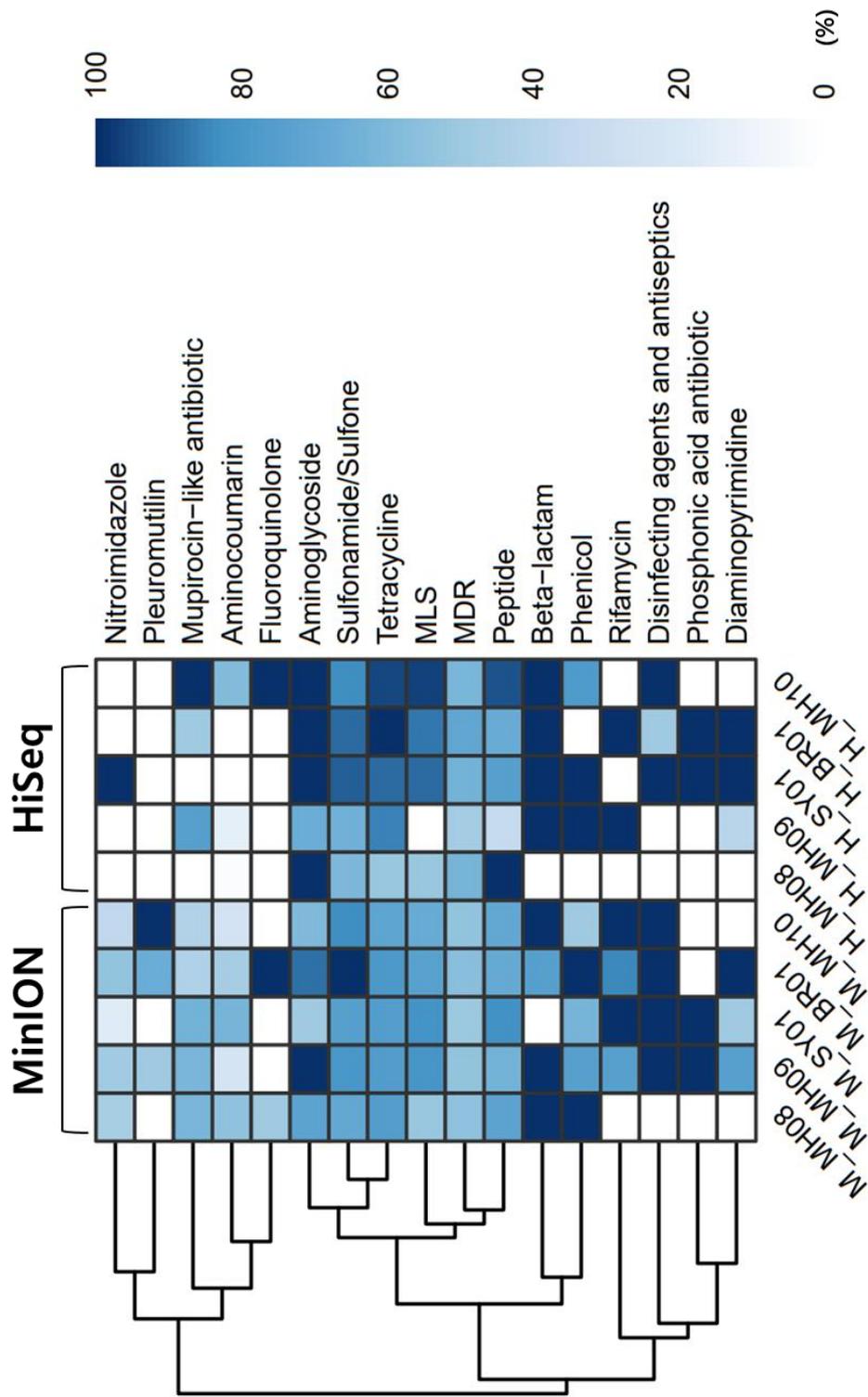


Figure 13. Heatmap analysis of the ARG-carrying plasmids.

4. CONCLUSION

본 연구에서는 충청북도 청주시에 있는 미호강 일대 5곳에서 채수한 강물의 DNA로 long-read sequencing (MinION)과 short-read sequencing (HiSeq)을 실시하였다. 미생물 군집의 경우, 문 수준에서는 두 플랫폼의 결과가 비슷하였으나, 속 수준으로 갈수록 HiSeq의 데이터에서 Unclassified 그룹의 비율이 점차 증가하였다. ARG의 absolute abundance를 분석한 결과, 두 플랫폼 간의 차이가 있었으나, drug class 별 경향성은 비슷하였다. Relative abundance 분석 결과, MinION과 HiSeq의 데이터에서 MDR을 가진 read와 contig가 각각 62.13 %, 43 %의 비율을 차지하는 것으로 나타났다. 모든 site에서 검출된 overlapping ARG의 relative abundance는 두 플랫폼의 결과에서 비슷하게 나타났다. ARG-MGE 쌍을 확인해 본 결과, *sul1-int11*와 *tet(A)-tnpA*가 대부분의 site에서 모두 검출되었다. Sulfonamide/sulfone, aminoglycoside, tetracycline 계열에 내성을 가진 read와 contig가 대부분 plasmid로 확인되었다. 이러한 연구 결과를 통해 앞으로의 ARG의 이동과 전파에 대한 연구에서 sequencing read의 길이를 고려한 미생물 군집 및 항생제 내성체 분석에 대해 정확한 이해와 깊은 고찰의 필요성이 강조된다.

ACKNOWLEDGEMENT

이 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아
수행된 연구임(No. 2021R1A2C1094426).

REFERENCES

1. O'Neill, J., *Tackling drug-resistant infections globally: final report and recommendations*. 2016.
2. Keenum, I., et al., *A framework for standardized qPCR-targets and protocols for quantifying antibiotic resistance in surface water, recycled water and wastewater*. *Critical Reviews in Environmental Science and Technology*, 2022. **52**(24): p. 4395-4419.
3. Waseem, H., et al., *Contributions and challenges of high throughput qPCR for determining antimicrobial resistance in the environment: a critical review*. *Molecules*, 2019. **24**(1): p. 163.
4. Han, Z., et al., *Antibiotic resistomes in drinking water sources across a large geographical scale: multiple drivers and co-occurrence with opportunistic bacterial pathogens*. *Water Research*, 2020. **183**: p. 116088.
5. An, X.-L., et al., *Tracking antibiotic resistome during wastewater treatment using high throughput quantitative PCR*. *Environment international*, 2018. **117**: p. 146-153.
6. Qian, X., et al., *Long-read sequencing revealed cooccurrence, host range, and potential mobility of antibiotic resistome in cow feces*. *Proceedings of the National Academy of Sciences*, 2021. **118**(25): p. e2024464118.
7. Martínez, J.L., T.M. Coque, and F. Baquero, *What is a resistance gene? Ranking risk in resistomes*. *Nature Reviews Microbiology*, 2015. **13**(2): p. 116-123.
8. Forster, S.C., et al., *Strain-level characterization of broad host range mobile genetic elements transferring antibiotic resistance from the human microbiome*. *Nature Communications*, 2022. **13**(1): p. 1445.
9. Partridge, S.R., et al., *Mobile genetic elements associated with antimicrobial resistance*. *Clinical microbiology reviews*, 2018. **31**(4): p. 10.1128/cmr.00088-17.
10. Peng, K., et al., *Long-read metagenomic sequencing reveals that high-copy small plasmids shape the highly prevalent antibiotic resistance genes in animal fecal microbiome*. *Science of The Total Environment*, 2023: p. 164585.
11. Li, A.-D., L.-G. Li, and T. Zhang, *Exploring antibiotic resistance genes and metal resistance genes in plasmid metagenomes from wastewater treatment plants*. *Frontiers in microbiology*, 2015. **6**: p. 1025.

12. Meng, M., Y. Li, and H. Yao, *Plasmid-mediated transfer of antibiotic resistance genes in soil*. *Antibiotics*, 2022. **11**(4): p. 525.
13. Metcalf, B., et al., *Short-read whole genome sequencing for determination of antimicrobial resistance mechanisms and capsular serotypes of current invasive *Streptococcus agalactiae* recovered in the USA*. *Clinical Microbiology and Infection*, 2017. **23**(8): p. 574. e7-574. e14.
14. Huptas, C., S. Scherer, and M. Wenning, *Optimized Illumina PCR-free library preparation for bacterial whole genome sequencing and analysis of factors influencing de novo assembly*. *BMC research notes*, 2016. **9**(1): p. 1-14.
15. Quainoo, S., et al., *Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis*. *Clinical microbiology reviews*, 2017. **30**(4): p. 1015-1063.
16. Hu, T., et al., *Next-generation sequencing technologies: An overview*. *Human Immunology*, 2021. **82**(11): p. 801-811.
17. Marin, M., et al., *Benchmarking the empirical accuracy of short-read sequencing across the *M. tuberculosis* genome*. *Bioinformatics*, 2022. **38**(7): p. 1781-1787.
18. Goldstein, S., et al., *Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing*. *BMC genomics*, 2019. **20**(1): p. 1-17.
19. Kingsford, C., M.C. Schatz, and M. Pop, *Assembly complexity of prokaryotic genomes using short reads*. *BMC bioinformatics*, 2010. **11**: p. 1-11.
20. Lee, H., et al., *Third-generation sequencing and the future of genomics*. *BioRxiv*, 2016: p. 048603.
21. van Almsick, V., et al., *The use of long-read sequencing technologies in infection control: Horizontal transfer of a blaCTX-M-27 containing IncFII plasmid in a patient screening sample*. *Microorganisms*, 2022. **10**(3): p. 491.
22. Partridge, S.R., et al., *Gene cassettes and cassette arrays in mobile resistance integrons*. *FEMS microbiology reviews*, 2009. **33**(4): p. 757-784.
23. Bohlin, J., et al., *The nucleotide composition of microbial genomes indicates differential patterns of selection on core and accessory genomes*. *BMC genomics*, 2017. **18**(1): p. 1-11.
24. Nakano, K., et al., *Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area*. *Human Cell*, 2017. **30**(3): p. 149-161.
25. Maestri, S., et al., *A rapid and accurate MinION-based workflow for tracking species biodiversity in the field*. *Genes*, 2019. **10**(6): p. 468.
26. Wang, Y., et al., *Nanopore sequencing technology, bioinformatics and applications*. *Nature biotechnology*, 2021. **39**(11): p. 1348-1365.

27. Leggett, R.M., et al., *Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens*. *Nature Microbiology*, 2020. **5**(3): p. 430-442.
28. Midha, M.K., M. Wu, and K.-P. Chiu, *Long-read sequencing in deciphering human genetics to a greater depth*. *Human genetics*, 2019. **138**(11-12): p. 1201-1215.
29. Matsuo, Y., et al., *Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution*. *BMC microbiology*, 2021. **21**: p. 1-13.
30. De Maio, N., et al., *Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes*. *Microbial genomics*, 2019. **5**(9): p. e000294.
31. Zhang, H., C. Jain, and S. Aluru, *A comprehensive evaluation of long read error correction methods*. *BMC genomics*, 2020. **21**: p. 1-15.
32. Maboni, G., et al., *Three Distinct Annotation Platforms Differ in Detection of Antimicrobial Resistance Genes in Long-Read, Short-Read, and Hybrid Sequences Derived from Total Genomic DNA or from Purified Plasmid DNA*. *Antibiotics*, 2022. **11**(10): p. 1400.
33. Adewale, B.A., *Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years?* *African journal of laboratory medicine*, 2020. **9**(1): p. 1-5.
34. Lo, C.-C. and P.S. Chain, *Rapid evaluation and quality control of next generation sequencing data with FaQCs*. *BMC bioinformatics*, 2014. **15**(1): p. 1-8.
35. Kim, D., et al., *Centrifuge: rapid and sensitive classification of metagenomic sequences*. *Genome research*, 2016. **26**(12): p. 1721-1729.
36. Wu, Z., et al., *Nanopore-based long-read metagenomics uncover the resistome intrusion by antibiotic resistant bacteria from treated wastewater in receiving water body*. *Water Research*, 2022. **226**: p. 119282.
37. Bengtsson-Palme, J., et al., *METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data*. *Molecular ecology resources*, 2015. **15**(6): p. 1403-1414.
38. Hyatt, D., et al., *Prodigal: prokaryotic gene recognition and translation initiation site identification*. *BMC bioinformatics*, 2010. **11**: p. 1-11.
39. Alcock, B.P., et al., *CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database*. *Nucleic acids research*, 2023. **51**(D1): p. D690-D699.
40. Buchfink, B., C. Xie, and D.H. Huson, *Fast and sensitive protein alignment using DIAMOND*. *Nature methods*, 2015. **12**(1): p. 59-60.
41. Johnson, M., et al., *NCBI BLAST: a better web interface*. *Nucleic acids research*, 2008. **36**(suppl_2): p. W5-W9.

42. Dai, D., et al., *Long-read metagenomic sequencing reveals shifts in associations of antibiotic resistance genes with mobile genetic elements from sewage to activated sludge*. *Microbiome*, 2022. **10**(1): p. 20.
43. Yang, Y., et al., *ARGs-OAP: online analysis pipeline for antibiotic resistance genes detection from metagenomic data using an integrated structured ARG-database*. *Bioinformatics*, 2016. **32**(15): p. 2346-2351.
44. Li, D., et al., *MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph*. *Bioinformatics*, 2015. **31**(10): p. 1674-1676.
45. Pärnänen, K., et al., *Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements*. *Nature communications*, 2018. **9**(1): p. 3891.
46. Schloss, P.D., et al., *Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities*. *Applied and environmental microbiology*, 2009. **75**(23): p. 7537-7541.
47. Niestępski, S., et al., *Environmental fate of Bacteroidetes, with particular emphasis on Bacteroides fragilis group bacteria and their specific antibiotic resistance genes, in activated sludge wastewater treatment plants*. *Journal of hazardous materials*, 2020. **394**: p. 122544.
48. Anandan, R., D. Dharumadurai, and G.P. Manogaran, *An introduction to actinobacteria*, in *Actinobacteria-basics and biotechnological applications*. 2016, IntechOpen.
49. Wang, Z., et al., *Critical roles of cyanobacteria as reservoir and source for antibiotic resistance genes*. *Environment International*, 2020. **144**: p. 106034.
50. Wang, J., et al., *Supercarriers of antibiotic resistome in a world's large river*. *Microbiome*, 2022. **10**(1): p. 1-19.
51. Maggiori, C., et al., *MinION sequencing from sea ice cryoconites leads to de novo genome reconstruction from metagenomes*. *Scientific reports*, 2021. **11**(1): p. 21041.
52. Wommack, K.E., J. Bhavsar, and J. Ravel, *Metagenomics: read length matters*. *Applied and environmental microbiology*, 2008. **74**(5): p. 1453-1463.
53. Ashton, P.M., et al., *MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island*. *Nature biotechnology*, 2015. **33**(3): p. 296-300.
54. Vaz-Moreira, I., et al., *Culture-dependent and culture-independent diversity surveys target different bacteria: a case study in a freshwater sample*. *Antonie Van Leeuwenhoek*, 2011. **100**: p. 245-257.

55. Leclercq, S.O., et al., *Diversity of the tetracycline mobilome within a Chinese pig manure sample*. Applied and Environmental Microbiology, 2016. **82**(21): p. 6454-6462.
56. Chaturvedi, P., et al., *Occurrence of emerging sulfonamide resistance (sul1 and sul2) associated with mobile integrons-integrase (int1 and int2) in riverine systems*. Science of The Total Environment, 2021. **751**: p. 142217.
57. Frank, J.A., et al., *Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data*. Scientific reports, 2016. **6**(1): p. 25373.
58. Stokes, H.t. and R.M. Hall, *A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons*. Molecular microbiology, 1989. **3**(12): p. 1669-1683.
59. Chen, B., et al., *The role of class I integrons in the dissemination of sulfonamide resistance genes in the Pearl River and Pearl River Estuary, South China*. Journal of Hazardous materials, 2015. **282**: p. 61-67.
60. Makowska, N., R. Koczura, and J. Mokracka, *Class 1 integrase, sulfonamide and tetracycline resistance genes in wastewater treatment plant and surface water*. Chemosphere, 2016. **144**: p. 1665-1673.
61. Zhang, T., et al., *Metagenomic insights into the antibiotic resistome in freshwater and seawater from an Antarctic ice-free area*. Environmental Pollution, 2022. **309**: p. 119738.
62. Zhao, W., et al., *Oxford nanopore long-read sequencing enables the generation of complete bacterial and plasmid genomes without short-read sequencing*. Frontiers in Microbiology, 2023. **14**: p. 1179966.

Comparative Analysis of Antibiotic Resistome between Long- and Short-read Sequencing

Jiwon Jeong

Department of Biotechnology

The Graduate School

Jeju National University

Abstract

The emergence and dissemination of antibiotic resistance genes (ARGs) have become a matter of public health in the world. Most studies on antibiotic resistance in the environment have revealed its risk through quantifying antibiotic resistance genes (ARGs) rather than whole genomes, suggesting the need for research on the mobility and transfer of ARGs. In this study, we investigated the diversity of microbial communities present in the environment using both short-read sequencing (HiSeq) and the recently developed long-read sequencing (MinION). We conducted a comparative

analysis on the antibiotic resistome. As a result of investigating microbial communities, 911 and 841 genera were classified based on long-read sequencing and short-read sequencing data, respectively. In the long-read sequencing results, unclassified genera accounted for an average of 8.7%, while the short-read sequencing results showed that 63.3% of genera belonged to the unclassified group. The abundance of antibiotic resistance genes (ARGs) between the two platforms shows a similar trend (Spearman's $\rho = 0.655$, $p < 0.05$), and multidrug resistance (MDR) was found to be the most common type (62.13%, 43%). In MinION and HiSeq results, 69 and 25 types of ARG-mobile genetic element (MGE) pairs present within 5 kb in a read or a contig were detected, respectively, and 183 and 91 ARGs were detected in reads or contigs predicted to be plasmids. Therefore, in this study, the differences between long- and short-read sequencing were compared through microbial community diversity and antibiotic resistance analysis, and suggestions for the future direction of research on antibiotic resistance analysis were proposed.

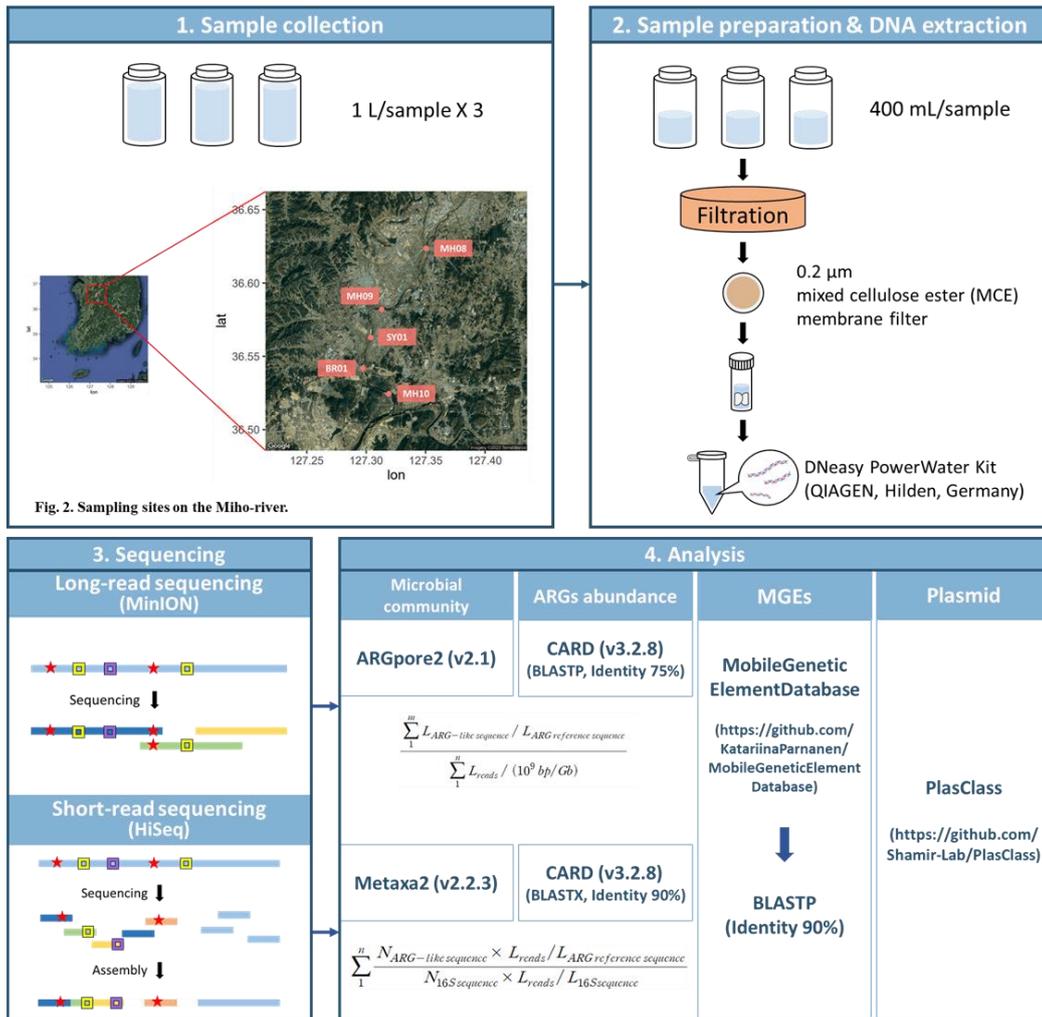


Figure 14. Graphical abstract.

감사의 글

2년 6개월이라는 시간이 정말 빠르게 지나갔습니다. 먼저, 석사과정 동안 부족한 저에게 기초부터 세심하게 가르쳐 주신 운노 타쓰야 교수님께 감사드립니다. 또한 실험실 이사라는 큰 이슈에도 걱정하지 말라며 흔쾌히 지도 교수님이 되어 주신 이효연 교수님께 감사드립니다. 마지막으로 바쁘신 와중에도 시간을 내어 저의 졸업 논문을 심사해주신 부경환 교수님께도 감사드립니다.

늦은 나이에 공부를 시작한 딸에게 한결 같은 응원과 전폭적인 지지를 보내주신 부모님께 감사드리고, 무심하지만 항상 걱정해준 동생 정규현과 언제나 그리운 강아지 강산이... 고맙습니다.

MBL의 한 사람으로 지내는 동안 많은 일들이 있었지만, 실험실 식구들이 있어 참 즐거웠습니다. 저를 포함한 MBL 학생들의 정신적 지주였던 김정만 박사님, R이 재밌다고 느끼게 해준 송호경 박사님, Dr. Farooq Adeel께 감사드립니다. 6시 이후 대학원생의 삶을 함께 해준 전다빈, 이선우, 정유진 선배님과 짧은 기간이었지만 많은 얘기를 들려준 조혜준, 고광표 선배님에게도 감사의 말씀을 전하고 싶습니다. 든든한 후배로서 많은 실험을 도와준 Aprajita와 현수진에게도 감사드립니다. 학부 연구생으로 실험과 샘플링을 도와준 김민우, 천영진, 최 건 학생에게도 감사드립니다. 실험실이 낯설었던 저에게 MinION sequencing을 가르쳐준 소윤정 선생님께도 감사드립니다.

학부생 시절부터 항상 응원해준 제주도 COFFEE 99.9 순금 언니와 민석 오빠, 언제나 멋지다며 응원해준 오랜 친구들 김상우, 김영룡, 김진희, 박재현, 양소정, 정다올, 하은진 언니, 한혜주, 한재현에게 감사드립니다.

석사과정 막바지에 만나 언제나 믿고 의지했던, 굴보다는 굴나무가 되라고 말해주신 이종서에게도 감사드립니다.

저는 한번 더 용기를 얻어 도전하려고 합니다. 그 끝이 어디인지는 모르겠으나, 언젠가 <그것이 알고 싶다>라는 프로그램에 출연하고 싶다는 그 마음을 잊지 않겠습니다.

2024년 1월, 청주의 한복판에서

정 지 원