

미생물법의학: 차세대염기서열분석 방법에 따른 MLVA 결과 비교 및 이를 활용한 DNA 감식

윤형석¹⁾ · 이승호¹⁾ · 임승현¹⁾ · 이대상¹⁾ · 구세훈¹⁾ · 김정은¹⁾ ·
정주환¹⁾ · 김성주¹⁾ · 허경행¹⁾ · 송동현^{*,1)}

¹⁾ 국방과학연구소 국방첨단과학기술연구원 Chem-Bio센터

Microbial Forensics: Comparison of MLVA Results According to NGS Methods, and Forensic DNA Analysis Using MLVA

Hyeongseok Yun¹⁾ · Seungho Lee¹⁾ · Seunghyun Lim¹⁾ · Daesang Lee¹⁾ · Sehun Gu¹⁾ ·
Jungeun Kim¹⁾ · Juhwan Jeong¹⁾ · Seongjoo Kim¹⁾ · Gyeunghaeng Hur¹⁾ · Donghyun Song^{*,1)}

¹⁾ Defense Advanced Science and Technology Research Institute, Agency for Defense Development, Korea

(Received 1 February 2024 / Revised 24 April 2024 / Accepted 8 May 2024)

Abstract

Microbial forensics is a scientific discipline for analyzing evidence related to biological crimes by identifying the origin of microorganisms. Multiple locus variable number tandem repeat analysis(MLVA) is one of the microbiological analysis methods used to specify subtypes within a species based on the number of tandem repeat in the genome, and advances in next generation sequencing(NGS) technology have enabled *in silico* analysis of full-length whole genome sequences. In this paper, we analyzed unknown samples provided by Robert Koch Institute(RKI) through The United Nations Secretary-General's Mechanism(UNSGM)'s external quality assessment exercise(EQAE) project, which we officially participated in 2023. We confirmed that the 3 unknown samples were *B. anthracis* through nucleic acid isolation and genetic sequence analysis studies. MLVA results on 32 loci of *B. anthracis* were analysed by using genome sequences obtained from NGS(NextSeq and MinION) and Sanger sequencing. The MLVA typing using short-reads based NGS platform(NextSeq) showed a high probability of causing assembly error when a size of the tandem repeats was grater than 200 bp, while long-reads based NGS platform(MinION) showed higher accuracy than NextSeq, although insertion and deletion was observed. We also showed hybrid assembly can correct most indel error caused by MinION. Based on the MLVA results, genetic identification was performed compared to the 2,975 published MLVA databases of *B. anthracis*, and MLVA results of 10 strains were identical with 3 unkonwn samples. As a result of whole genome alignment of the 10 strains and 3 unknown samples, all samples were identified as *B. anthracis* strain A4564 which is associated with injectional anthrax isolates in heroin users.

Key Words : Microbial Forensics(미생물법의학), *Bacillus Anthracis*(탄저균), MLVA(다중 유전자좌 가변수일렬반복 분석), NGS(차세대염기서열분석)

* Corresponding author, E-mail: swpia@add.re.kr

Copyright © The Korea Institute of Military Science and Technology

1. 서론

미생물법의학(Microbial forensics)은 미생물이나 그 독소가 무기로 사용되는 생물학적 범죄와 관련된 증거를 수집하고 분석하기 위해 발전된 분야이다^[1]. 생물 테러나 공격, 생물 범죄, 생물학적 작용제 및 독소의 인위적 조작과 부주의한 누출로부터 발생하는 증거를 확보하기 위한 과학적 접근 방법으로 미생물법의학이 사용되고 있다^[2]. 따라서 미생물법의학에서는 생물학적인 증거로부터 미생물의 출처나 생산자 또는 최초 발생지역을 특정하는데 있어 높은 수준의 정확도를 보이는 것을 목표로 한다^[1]. 따라서 이 분야는 미생물학을 포함하여 바이러스학, 생화학, 유전학, 생물정보학, 면역학, 분자생물학, 역학을 포함하는 다양한 전문분야에 대한 기초 지식을 필요로 한다.

미생물 유형화(Microbial typing) 기술은 미생물 군집의 역학 및 이와 관련된 진단학(Diagnostics), 유전체학(Genomics) 및 발병기전(Pathogenesis)을 밝히는 데 널리 사용되고 있을 뿐만 아니라 미생물법의학 분야에서 미생물의 출처를 규명하는 데도 사용될 수 있다^[3]. 이러한 목표를 위해 다중 유전자좌 가변수일렬반복 분석(Multiple Locus Variable number of tandem repeat Analysis, MLVA)등과 같이 분리된 균주를 식별할 수 있는 몇 가지 분석 방법이 널리 사용되어 왔다^[4].

대규모 병렬 염기서열분석법(Massive parallel sequencing)의 하나인 차세대염기서열분석법(Next Generation Sequencing, NGS)의 도입은 미생물의 전체 유전체(Whole genome) 분석 시간을 혁신적으로 감소 시킴으로써 미생물법의학 분야에서 상당한 발전을 이루는 주요한 기술이 되었다^[5,6]. 제한효소 기반의 DNA 지문분석(DNA fingerprinting)이나 DNA 증폭 기반의 MLVA, 염기서열분석 기반의 Multilocus Sequence Typing(MLST) 분석 기술이 *in-silico*로 분석이 가능해짐에 따라 더 높은 수준의 식별이 가능하게 되었다^[7].

유전체 상에서 반복되는 DNA 서열은 박테리아와 진핵생물을 포함한 대부분의 생명체에서 발견되며, 전체 유전체에서 여러 유전자좌에 발생하는 전이성 유전인자(Transposable element)에 의한 산재된 반복과 단일 유전자좌에서 발생하는 tandem repeat의 두 가지 범주로 그룹화할 수 있다^[8,9]. MLVA는 미생물 법의학에서 박테리아 균주의 아류형(Subtype)을 정하는 데 사용되는 분석 기술 중 하나로써 다중의 유전자좌에서 일렬반복의 개수를 측정하는 기술이다^[10]. NGS 플

랫폼들 가운데 가장 널리 사용되고 있는 Illumina사의 장비의 염기서열분석 오류율(base calling error)은 0.1 % 이하로 매우 낮은 편이나 리드(Reads)의 길이가 200 bp 이하로 비교적 짧은 short-reads를 생산한다^[11]. 유전체 서열에 short-reads의 길이보다 긴 반복되는 서열들이 존재할 경우, short-read를 조립(assembly)하여 만든 contigs 서열에서 조립 오류(assembly error)가 나타나는 근본적인 한계점을 가지고 있으며, 이로 인해 MLVA를 분석하는 데 있어 tandem repeat의 개수를 정확히 분석하기 어려울 수 있다^[12]. Oxford Nanopore Technologies사의 MinION 플랫폼의 경우 리드의 길이가 10 kb 이상으로 길어서 조립 오류를 해결할 수 있다는 장점이 있으나, 단일 통과 오류율(single-pass error rates)이 10-15 %이고 대부분이 뉴클레오티드의 삽입과 삭제로 이루어져 있어서 MLVA 결과 분석 시 반복되는 서열의 개수가 정확히 일치하지 않는 오류가 발생한 가능성이 있다^[13-16]. 이를 해결하기 위해 최근 연구에서는 조립 오류의 가능성이 낮지만 부정확한 long-read NGS 결과를 주형으로 하여 short-read NGS 결과로 연마(polishing)하는 하이브리드 조립(hybrid assembly) 방식의 접근법을 사용하여 정확도를 높일 수 있음이 확인된 바 있다^[21].

United Nations Secretary-General's Mechanism(UNSGM)은 생화학 테러의 위협을 감지하고 분석하기 위해 1987년 설립되었으며 참여국의 생물 분석 기준 실험실(Bio-analytical reference laboratories) 역량 강화를 위해 독일의 Robert Koch Institute(RKI) 연구소에서는 매년 외부 기관 검증 훈련(External Quality Assurance Exercise, EQAE)을 실시하고 있다. 본 논문에서는 UNSGM-EQAE 프로젝트의 활동으로 수분 보충 스프레이, 베이비파우더가 혼합된 환경 스왑(swab) 및 젤리 등 특정 표본에 미지의 핵산이 포함된 시료 3종을 대상으로 핵산 분리 및 염기서열 분석을 진행했으며 그 결과 *B. anthracis* 유전체로 특정했다. 확보한 서열을 대상으로 MLVA 결과를 분석하여 DNA 감식을 수행했으며, 그 과정에서 발견된 서열분석방법에 따른 tandem repeat의 조립 오류 및 단일 통과 오류율에 대한 분석, 그리고 하이브리드 조립 결과를 비교했다. NGS 플랫폼을 활용한 세가지 조립 결과 및 Sanger sequencing 서열 데이터, 그리고 전기영동결과를 조합하여 *B. anthracis*의 출처를 10개의 후보균으로 선별했으며, 전장유전체서열 분석을 통해 아류형을 A4564로 특정했다.

2. 실험방법

미지시료는 0.22 µm 필터로 여과 멸균된 핵산이 0.85 % NaCl(S-1-2), 베이비파우더가 포함된 환경 스왑(S-2-1) 또는 젤리(S-3-1)에 미지의 농도로 접종된 형태로 RKI로부터 전달받았으며 Qiagen의 DNease Blood & Tissue Kits의 메뉴얼에 따라 핵산추출을 진행했다. 농도가 낮은 샘플의 경우, 염기서열을 효율적으로 확보하기 위해서 TruPrime Whole Genome Amplification Kit[®] (4basebio[™])을 이용하여 핵산 증폭을 수행하였다. Short-reads 기반 염기서열분석 플랫폼으로 Illumina의 NextSeq 500 장비를 이용하였고, long-reads 기반 염기서열분석 플랫폼으로 Oxford Nanopore의 MinION Mk1C 장비를 이용했다. short-reads 기반의 NextSeq 염기서열분석은 Illumina의 TruSeq Library Prep Kit를 이용하여 시료 전처리를 했으며, 샘플 당 각각의 dual index adaptor를 말단에 붙여 증폭시켰다. 각 단계는 Beckman Coulter의 AMPure Xp bead를 사용하여 cleanup을 진행했다. 생성된 라이브러리의 상태와 농도는 Agilent Technologies의 TapeStation 4200과 D1000 ScreenTape를 통해 1차 확인하였고, KAPA Biosystems의 KAPA Library Quantification Kit와 Thermo Fisher Scientific의 QuantStudio 6 Flex Real-time PCR system을 통해 최적의 라이브러리 농도를 확인했다. 최종적으로 4 pM의 라이브러리를 사용하여 Illumina의 10 % PhiX control v3와 함께 NextSeq 500 benchtop platform을 통해 염기서열분석을 진행했다. 염기서열분석을 통해 생성된 결과물은 adaptor trimming, FaQC를 이용한 low-quality reads 제거(Phred quality score < Q30), host reads 제거 단계를 거쳐서 SPAdes를 이용하여 *de novo* assembly를 수행했다.

Long-reads 기반의 MinION 염기서열분석은 초기 DNA 농도 100 ng/uL를 이용하였다. New England Biolabs의 NEBnext Ultra II End Repair/dA-tailing과 Oxford Nanopore Technologies의 Ligation Sequencing Kit(SQK-LSK109)를 이용하여 시료 전처리를 진행하였다. Ultra II End Repair/dA-tailing을 통하여 양말단을 blunt end로 복구 및 3'end에 A-tailing 후, 샘플의 양말단에 adaptor를 붙여서 라이브러리를 제작하였다. 각 단계의 clean-up은 Beckman Coulter의 AMPure Xp bead를 사용하였다. 생성된 라이브러리는 Qubit을 이용하여 농도를 확인하였다. 50-100 fmol의 농도 기준이 충족된 라이브러리는 Sequencing Buffer와 Loading Bead와 함께 mixture를 만들어서 flow cell(R9.4.1)에 넣고, MinION 장비인 Mk1C에 탑재하여

24시간 동안 염기서열분석을 수행했으며 생성된 FASTQ 파일은 adaptor trimming과 low-quality reads를 제거했다. 이후 long-read assembly 프로그램인 flye를 사용하여 *de novo* assembly를 수행하고 contig 서열을 산출했다.

Sanger sequencing을 위해 추출한 DNA는 Enzymomics 사의 nPfu-special 제품의 B buffer를 사용하여 메뉴얼에 따라 혼합물을 제작 및 PCR을 수행했다. PCR 조건은 95 °C 2분 반응 후 95 °C 15초, 56 °C 15초, 72 °C 2분 반응을 40회 수행한 후 72 °C 2분 반응 및 4 °C 로 유지했으며 각 유전자좌에 따른 PCR 프라이머는 Table 1을 참고했다. PCR 결과물은 agarose gel 전기영동 및 추출을 진행했으며 Table 1의 프라이머를 사용하여 각 산출물의 Sanger sequencing을 진행했다. Sanger sequencing은 Applied biosystems 사의 3500 genetic analyzer를 이용했으며 BigDye Terminator v3.1 cycle sequencing kit 및 BigDye XaTerminator purification kit을 사용해서 제조사 메뉴얼에 따라 염기서열분석을 수행했다.

Long-read와 short-read의 하이브리드 조립을 위해 CLC Genomics Workbench 24.0에서 MinION 염기서열 분석 결과로부터 *de novo* assembly long reads를 수행하여 contigs를 획득했다. 이후 map reads to contigs를 이용해 NextSeq 염기서열분석 결과로 polishing을 진행했다. MLVA 결과는 기존의 문헌의 분석 코드를 수정하여 확인했으며 CLC Genomics Workbench 24.0을 사용하여 교차검증했다^[2].

3. 결과

Short-reads 기반의 플랫폼(NextSeq)과 long-reads 기반의 플랫폼(MinION)을 이용해서 미지시료 3종으로부터 분리된 핵산에 대한 염기서열분석을 수행했다. 획득한 서열의 BLAST(Basic Local Alignment Search Tool)를 수행한 결과 미지시료에 포함된 핵산은 모두 *B. anthracis*으로 확인되었으며 DNA 감식을 수행하기 위해 미지시료 S-1-2에 대하여 32개의 유전자좌에 대한 MLVA 결과 분석을 *in silico*로 수행했다(Table 1). 이때 참조서열 *B. anthracis* strain Ames Ancestor에 대하여 200 bp 이하의 가변수열반복(Variable Number Tandem Repeat, VNTR)이 포함된 15개의 MLVA 유전자좌를 short VNTR(shVNTR)로, 나머지를 long VNTR(loVNTR)로 그룹을 나누었다.

Table 1. List of *B. anthracis* VNTR markers^[18]

Locus	Foward primers	Reverse primers
vrrA	CACAAC TACCACCGATGGCACA	GCGCGTTTCGTTTGATT CATA C
vrrB1	ATAGGTGGTTTTCCGCAAGTTATTC	GATGAGTTTGATAAAGAATAGCCTGTG
vrrB2	CACAGGCTATTCTTTATCAAAC TCATC	CCCAAGGTGAAGATTGTTGT TGA
vrrC1	GAAGCAAGAAAGTGATGTAGTGGAC	CATTCCTCAAGTGCTACAGG TTC
vrrC2	CCAGAAGAAGTGGAACCTGTAGCAC	GTCTTTCCATTAATCGCGCTCTATC
CG3	TGTCGTTTTACTTCTCTCTCCAATAC	AGTCATTGTTCTGTATAAAGGGCAT
pXO1-aat	CAATTTATTAACGATCAGATTAAGTTCA	TCTAGAATTAGTTGCTTCATAATGGC
pXO2-at	TCATCCTCTTTTAAGTCTTGGGT	GTGTGATGAACTCCGACGACA
bams01	GTTGAGCATGAGAGGTACCTTGTCTTTTT	AGTTCAAGCGCCAGAAGGTTATGAGTTATC
bams03	GCAGCAACAGAAAAC TTTCTCTCCAATAACA	TCCTCCCTGAGAACTGCTATCACCTTTAAC
bams05	GCAGGAAGAACAAAAGAAACTAGAAAGACA	ATTATTAGCAGGGCCTCTCTGCATTACC
bams07	GAATATTCGTGCCACCTAACAAAACAGAAA	TGTCAGATCTAGTTGGCCCTACTTTTCCTC
bams13	AATTGAGAAATTGCTGTACCAAAC T	CTAGTGCATTTGACCCTAATCTTGT
bams15	GTATTTCCCCAGATACAGTAATCC	GTGTACATGTTGATTCATGCTGTTT
bams21	TGTAGTGCCAGATTTGCTTCTGTGA	CAAATTTTGAGATGGGAGTTTACT
bams22	ATCAAAAATCTTGGCAGACTGA	ACCGTTAATTCACGTTTAGCAGA
bams23	CGGTCTGTCTCTATTATTCAGTGGT	CCTGTTGCTCCTAGTGATTCTTAC
bams24	CTTCTACTTCCGTACTTGAAATTGG	CGTCACGTACCATTTAATGTTGTTA
bams25	CCGAATACGTAAGAAATAAATCCAC	TGAAAGATCTTGAAAAACAAGCATT
bams28	CTCTGTTGTAACAAAATTTCCGTCT	TATTAACCAGGCGTTACTTACAGC
bams30	GCATAATCACCTACAACCTGGTA	CAGAAAATATTGGACCTACCTCC
bams31	GCTGTATTTATCGAGCTTCAAAATCT	GGAGTACTGTTTGTGTAATGTTGTTT
bams34	CAGCAAAATCAATCGAATCAAA	TGTGCTAAATCATCTTGCTTGG
bams44	GCGAATTAATTGCTCCTCAAAT	GCACTTGAATATTGGCGGTAT
bams51	ATTCCTGAAGCAGGTTGTGTT	TGCATCTAACAATGCAGAACAA
bams53	GAGGTGTGTTAGGTGGGCTTAC	CATATTTTCACCTTAATTTTGAAG
vntr12	CGTACGAAGTAGAAGTCATTAA	GCATATAATTGCACCTCATCTAG
vntr16	CTCTTGAAAATATAAACGCA	GAATAATAAGGGTTCTCATGGTAT
vntr17	TAGGTAAACAAATTTTCGTAATC	GATCGTACAACAGCAATTATCAT
vntr19	GTGATGAAATCGGACAAGTTAGGAG	GAAATATTTTATTAACATGCTTTCCATCC
vntr23	TTTAGAAACGTTATCACGCTTA	GTAATACGTATGGTTCATTCCC
vntr35	AAATAATATGTTCCTTTTGCTG	GTCCTGAAATAAATGCTGAAT

shVNTR을 포함하는 MLVA 결과를 세부적으로 확인했을 때 NextSeq와 Sanger sequencing의 결과는 모두 일치함을 확인했다(Table 2). MinION을 이용한 염기서열분석 결과 pXO2-at, bams25, bams53 그리고 vntr12에 대한 MLVA 결과가 부정확했는데, indel에 의한 무작위적인 오류인 것으로 확인되었다. 네 개의 MLVA 유전자좌 모두 1 bp 결손이 일어났으나 pXO2-at와 vntr12의 tandem repeat 서열 길이가 2 bp로 짧아서 MLVA 결과값이 0.5 감소하는 큰 차이로 나타났다. 이러한 차이에도 MLVA 결과는 정수로 표현되기 때문에 실제 분석에 사용한 결과는 괄호에 반올림하여 표현했다. MinION 결과의 indel 오류를 바로잡기 위해 NextSeq 염기서열분석 결과로 연마를 수행했을 때 indel에 의한 무작위적인 오류가 모두 수정되었음을 확인했다(Table 2, Hybrid).

NextSeq을 이용한 염기서열분석은 0.1 % 이하의 매우 낮은 오류율을 보이지만 리드의 길이가 250 bp 이내로 짧기 때문에 200 bp 이상의 tandem repeat에서 조

립 오류가 발생 할 확률이 높아진다. 미지시료 S-1-2에 대하여 loVNTR을 포함하는 17개의 MLVA 유전자좌에 대한 NextSeq 결과를 분석했을 때 2개의 유전자좌에서 Sanger sequencing과 상이한 결과를 관찰했으며 이는 조립 오류에 의한 차이임을 확인했다(Table 3). 반면 MinION의 리드 길이는 10-40 kbp로 200 bp 이상의 tandem repeat이 포함된 MLVA 유전자좌의 염기서열을 비교적 성공적으로 분석했으나, MinION의 경우 11-15 %의 단일 통과 오류율을 보이고 오류의 대부분이 뉴클레오티드의 무작위적인 삽입과 결손으로 이루어져 있어서 tandem repeat 길이가 부정확한 경우를 관찰했다.

하이브리드 조립 결과 MinION에서 부정확했던 vrrC1, vrrC2, bams07 그리고 bams28의 MLVA 결과의 오류가 수정되었음을 확인했다(Table 3). 하지만 연마과정에서 bams03, bams15 그리고 bams31의 MLVA 결과값에 변

Table 2. Analysis results of MLVA including shVNTR for *B. anthracis* isolated from the DNA of the S-1-2 sample

Locus	NextSeq	MinION	Hybrid	Sanger
vrrA	4	4	4	4
vrrB1	20	20	20	20
vrrB2	14	14	14	14
CG3	2	2	2	2
pXO1-aat	10	10	10	10
pXO2-at	7	6.5(7)	7	7
bams05	5	5	5	5
bams25	12	11.9(12)	12	12
bams53	8	7.9(8)	8	8
vntr12	6	5.5(6)	6	6
vntr16	20	20	20	20
vntr17	4	4	4	4
vntr19	5	5	5	5
vntr23	4	4	4	4
vntr35	4	4	4	4

Table 3. Analysis results of MLVA including loVNTR for *B. anthracis* isolated from the DNA of the S-1-2 sample

Locus	NextSeq	MinION	Hybrid	Sanger
vrrC1	57	56	57	57
vrrC2	21	22.4(22)	21	21
bams01	13	13	13	13
bams03	27	30	29.4(29)	30
bams07	99	98.9(99)	99	99
bams13	30	30	30	30
bams15	45	45	45.3(45)	45
bams21	10	10	10	10
bams22	16	16	16	16
bams23	11	11	11	11
bams24	11	11	11	11
bams28	14	13.8(14)	14	14
bams30	69	74.9(75)	74.9(75)	74.9(75)
bams31	64	64	63	64.1(64)
bams34	9	9	9	9
bams44	8	8	8	8
bams51	9	9	9	9

Table 4. MLVA analysis of *B. anthracis* isolated from the DNA of the 3 unknown sample. ^aMLVA results using gel electrophoresis

Locus	S-1-2	S-2-1	S-3-2	Ref.
vrrA	4	4	4	4
vrrB1	20	20	20	20
vrrB2	14	14	14	14
vrrC1	57	57a	57	57
vrrC2	21	21	21	21
CG3	2	2	2	2
pXO1-aat	10	10	10	10
pXO2-at	7	7	7	7
bams01	13	13	13	13
bams03	30	30a	30	30
bams05	5	5a	5	5
bams07	99	99a	99	99
bams13	30	30a	30	30
bams15	45	45a	45	45
bams21	10	10	10	10
bams22	16	16	16	16
bams23	11	11a	11	11
bams24	11	11a	11	11
bams25	12	12	12	12
bams28	14	14	14	14
bams30	75	75a	75	75
bams31	64	64a	64	64
bams34	9	9	9	9
bams44	8	8	8	8
bams51	9	9a	9	9
bams53	8	8	8	8
vntr12	6	6	6	6
vntr16	20	20	20	20
vntr17	4	4	4	4
vntr19	5	5	5	5
vntr23	4	4	4	4
vntr35	4	4	4	4

등이 있었으며, 이 중 bams03 과 bams31의 경우 정수로 치환했을 때 결과값이 Sanger sequencing 결과와 일치하지 않는 것을 확인했다. 이러한 결과는 하이브리드 조립이 *in silico* 기반의 MLVA에 있어서 현재까지 최고의 정확도를 보이지만 여전히 실제 MLVA 결과와 정확하게 일치하지 않는다는 최근의 논문과 일치하는 결과였다^[21]. 또한 본 결과를 바탕으로 연마과정을 반복하여 수행했을 때 bams03의 MLVA 결과가 short-read MLVA 수치인 27에 가까워지는 것을 확인함으로써 여전히 반복서열에 대한 부정확성이 존재함을 확인했다.

미지시료 S-2-1과 S-3-1 또한 NGS(NextSeq, MinION 그리고 Hybrid)와 Sanger sequencing을 이용하여 MLVA 분석을 수행했다(Table 4). S-2-1의 경우 시료의 농도가 낮아 전체 유전체 증폭(whole genome amplification)을 수행했음에도 NextSeq 결과 참조서열 *B. anthracis* strain Ames Ancestor에 대한 coverage가 93.23 %로, 99.99 %인 다른 시료들에 비해 유전체서열 획득에 어려움이 있었으며 MinION 및 Sanger sequencing 결과 또한 MLVA를 분석하기 어려운 수준이었다. 차선책으로 vrrC1을 포함한 11개의 MLVA 유전자좌는 전체 유전체 증폭(whole genome amplification)을 수행한 후 전통적인 방법으로 PCR 및 전기영동을 통해 결과를 확

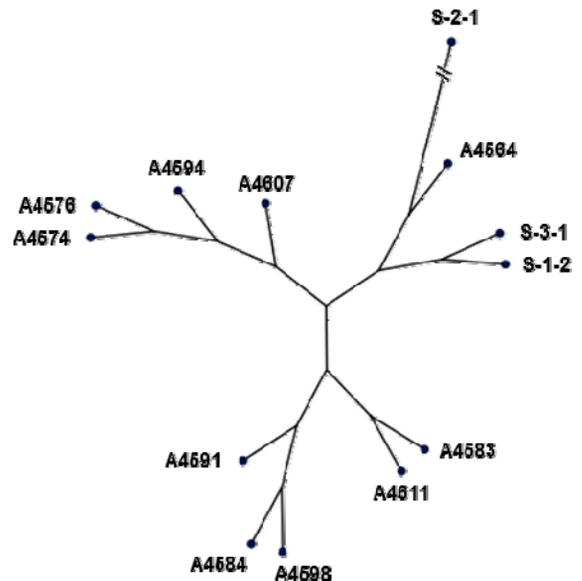


Fig. 1. Average nucleotide identity based phylogenetic tree of 10 *B. Anthracis* and 3 unknown samples

인했으며 대조군으로 S-1-2와 S-3-2를 사용했다.

미지시료의 출처를 특정하기 위해 MLVA 데이터베이스에 존재하는 2,975개의 공개된 *B. anthracis* MLVA 결과와 미지시료 3종에 포함된 핵산으로부터 분석한 염기서열의 MLVA 결과를 비교했을 때 미지시료에 포함된 핵산은 헤로인 투여 후 감염 집단인 injectional anthrax에서 분리된 *B. anthracis* 중 10개의 strain과 같은 계열임을 확인했다(Table 4, Ref.)^[19-20]. NCBI(National Center for Biotechnology Information) 데이터베이스에서 10개의 *B. anthracis* 후보의 NGS 결과를 확보한 후 Ames Ancestor를 참조서열로 하여 전장유전체분석 결과 미지시료 3종은 A4564임을 특정할 수 있었다(Fig. 1).

4. 결론

본 연구에서는 서열분석방법에 따른 tandem repeat의 조립 오류 및 indel에 의한 오류 가능성을 확인하고 MLVA 결과를 바탕으로 한 유전자 감식을 수행했다. NextSeq을 이용한 short-reads 기반의 전장유전체서열 분석은 비용적인 효율성과 높은 정확도로 인해 현재는 가장 널리 사용되고 있다. MinION을 활용한 long-reads 기반의 전장유전체서열분석은 short-reads 기반 전장유전체서열분석 대비 높은 오류율을 보이기는 하지만 리드의 길이가 길어서 tandem repeat과 같이 조립과정에서 오류가 발생할 수 있는 경우에는 이를 활용하여 정확도를 높일 수 있다. 또한 최근의 연구결과에 따르면 long-read로부터 획득한 서열을 기반으로 하여 short-read 서열로 오류를 바로잡는 하이브리드 조립의 정확도가 가장 높음이 보고된 바 있다.

본 연구 결과 NGS(NextSeq, MinION 그리고 Hybrid)와 Sanger sequencing으로부터 얻은 shVNTR의 MLVA 결과 MinION에서 indel에 의한 부정확성이 관찰되었으나 반올림하여 정수로 치환했을 때 분석 방법에 따른 정확도는 100 %였다(Fig. 2). loVNTR에 대하여 NextSeq, MinION의 정확도는 앞선 이유와 마찬가지로의 이유로 88.2 %로 동일했다. 하이브리드 조립 시 MinION에서 나타나는 indel 오류가 대부분 수정되었으나 bams03과 bams31에서 부정확한 결과를 나타내어 정확도는 88.2 %였다. 이는 법의학적 측면에서 *in silico* 분석만으로 MLVA 결과를 근거로 한 미생물 출처 식별에는 한계가 있음을 의미하며 Sanger sequencing을 포함한 전통적인 방법이 병행되어야 할 것으로 보

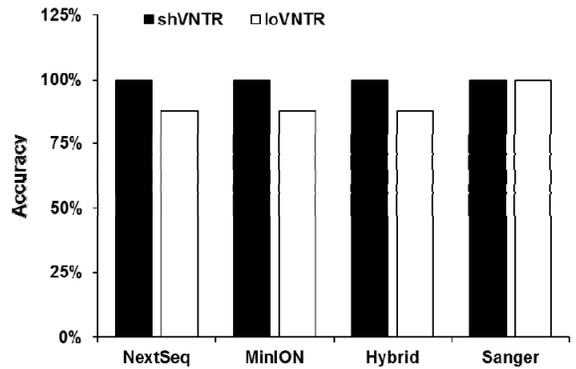


Fig. 2. Accuracy of MLVA results depending on sequencing platform

인다. 또한 미지시료 S-2-1과 같이 환경 스왑시료의 경우 핵산의 농도 및 순도가 매우 제한적이기 때문에 NGS를 활용한 전장유전체서열 분석 방법으로는 식별의 한계가 명확했으며 MLVA를 활용한 1차 선별이 효과적으로 사용되었다.

MLVA를 활용한 유전자 감식은 전장유전체서열의 1 % 이하의 정보만으로 유전자 감식을 수행하므로 전장유전체서열기반 분석 대비 낮은 해상도를 보일 수 있다. 그럼에도 불구하고 전장유전체서열이 알려지지 않았거나, 제한된 시료의 양이나 농도 문제로 NGS 품질이 제한적인 경우, 과거의 MLVA 결과만 존재할 때 등 법의학적 증거를 바탕으로 미생물 감식을 수행할 시 본 연구를 바탕으로 한 MLVA 결과 분석이 적용 가능할 것으로 판단되며, 본 연구결과로부터 획득한 기술을 활용하여 생물테러 등 군사적으로 사용된 미지시료의 출처 특정에 활용이 가능할 것으로 기대한다.

후 기

본 논문은 2024년 정부의 재원으로 국방과학연구소에서 수행한 연구결과임(912A01201).

References

[1] Schmedes S, Budowle B. Microbial Forensics. Encyclopedia of Microbiology. 2019:134-45. doi: 10.1016/B978-0-12-801238-3.02483-1.

- [2] Microbial Forensics: A Scientific Assessment: This report is based on a colloquium sponsored by the American Academy of Microbiology held June 7-9, 2002, in Burlington, Vermont. Washington (DC): American Society for Microbiology; 2003.
- [3] Boers SA, van der Reijden WA, Jansen R. High-throughput multilocus sequence typing: bringing molecular typing to the next level. *PLoS One*. 2012;7(7):e39630.
- [4] Foley SL, Lynne AM, Nayak R. Molecular typing methodologies for microbial source tracking and epidemiological investigations of Gram-negative bacterial foodborne pathogens. *Infect Genet Evol*. 2009 Jul;9(4):430-40.
- [5] Kuiper I. Microbial forensics: next-generation sequencing as catalyst: The use of new sequencing technologies to analyze whole microbial communities could become a powerful tool for forensic and criminal investigations. *EMBO Rep*. 2016 Aug;17(8):1085-7.
- [6] Schmedes SE, Sajantila A, Budowle B. Expansion of Microbial Forensics. *J Clin Microbiol*. 2016 Aug;54(8):1964-74.
- [7] Perez-Losada M, Arenas M, Castro-Nallar E. Microbial sequence typing in the genomic era. *Infect Genet Evol*. 2018 Sep;63:346-359.
- [8] Paco A, Freitas R, Vieira-da-Silva A. Conversion of DNA Sequences: From a Transposable Element to a Tandem Repeat or to a Gene. *Genes(Basel)*. 2019 Dec 5;10(12):1014.
- [9] Richard G.F., Kerrest A., Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev*. 2008;72:686-72
- [10] Lindstedt BA. Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis*. 2005 Jun;26(13):2567-82.
- [11] Glenn TC. Field guide to next-generation DNA sequencers. *Mol Ecol Resour*. 2011 Sep;11(5):759-69.
- [12] Bentley DR, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008 Nov 6;456(7218):53-9.
- [13] Olasagasti F, Lieberman KR, Benner S, Cherf GM, Dahl JM, Deamer DW, Akeson M. Replication of individual DNA molecules under electronic control using a protein nanopore. *Nat Nanotechnol*. 2010 Nov;5(11):798-806.
- [14] Eid J, et al. Real-time DNA sequencing from single polymerase molecules. *Science*. 2009 Jan 2;323(5910):133-8.
- [15] Weirather JL, de Cesare M, Wang Y, Piazza P, Sebastiano V, Wang XJ, Buck D, Au KF. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res*. 2017 Feb 3;6:100.
- [16] Rhoads A, Au KF. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*. 2015 Oct;13(5):278-89.
- [17] Appelt S, Rohleder AM, Invernizzi C, Mikulak R, Brinkmann A, Nitsche A, Kruger M, Dorner MB, Dorner BG, Scholz HC, Grunow R. Strengthening the United Nations Secretary-General's Mechanism to an alleged use of bioweapons through a quality-assured laboratory response. *Nat Commun*. 2021 May 25;12(1):3078.
- [18] Thierry S, Tourterel C, Le Fleche P, Derzelle S, Dekhil N, Mendy C, Colaneri C, Vergnaud G, Madani N. Genotyping of French Bacillus anthracis strains based on 31-loci multi locus VNTR analysis: epidemiology, marker evaluation, and update of the internet genotype database. *PLoS One*. 2014 Jun 5;9(6):e95131.
- [19] Grunow R, Verbeek L, Jacob D, Holzmann T, Birkenfeld G, Wiens D, von Eichel-Streiber L, Grass G, Reischl U. Injection anthrax--a new outbreak in heroin users. *Dtsch Arztebl Int*. 2012 Dec;109(49):843-8.
- [20] Keim P, Grunow R, Vipond R, Grass G, Hoffmaster A, Birdsell DN, Klee SR, Pullan S, Antwerpen M, Bayer BN, Latham J, Wiggins K, Hepp C, Pearson T, Brooks T, Sahl J, Wagner DM. Whole Genome Analysis of Injectional Anthrax Identifies Two Disease Clusters Spanning More Than 13 Years. *EBioMedicine*. 2015 Oct 6;2(11):1613-8.

- [21] Linde J, Brangsch H, H?lzer M, Thomas C, Elschner MC, Melzer F, Tomaso H. Comparison of Illumina and Oxford Nanopore Technology for genome analysis of *Francisella tularensis*, *Bacillus anthracis*, and *Brucella suis*. *BMC Genomics*. 2023 May 12;24(1):258.
- [22] Shevtsov V, Kairzhanova A, Shevtsov A, Shustov A, Kalendar R, Abdrakhmanov S, Lukhnova L, Izbanova U, Ramankulov Y, Vergnaud G. Genetic diversity of *Francisella tularensis* subsp. *holarctica* in Kazakhstan. *PLoS Negl Trop Dis*. 2021 May 17;15(5):e0009419.