식품에서 분리된 A형 간염 바이러스 전장유전체 분석

Whole Genome Sequencing Analysis of Hepatitis A Virus from Food

여다슬, 최창순*

Daseul Yeo, Changsun Choi*

중앙대학교 식품영양학과

Department of Food and Nutrition, Chung-Ang University, 4726 Seodongdaero, Daeduck-myeon, Anseong-si, Gyeonggi-do 17546, Republic of Korea

서론

A형 간염 바이러스 (HAV)의 특징 및 감염 동향

HAV는 급성 간염을 일으키는 바이러스로 27 nm의 껍질이 없는 *Picornaviridae* 과의 Hepatovirus 속에 속하는 RNA 바이러스이다. 유전자는 7,500 bp 염기쌍으로 구성되고 크게 3가지(P1, P2, P3) 단백유전자로 구성된다. 7개의 유전형(Ⅰ[A, B], Ⅱ, Ⅲ[A, B], Ⅳ, Ⅴ, Ⅵ, Ⅶ)이 존재하며, 그 중 4개의 유전형(Ⅰ, Ⅱ, Ⅲ, Ⅶ)이 사람에게 감염을 일으킨다.

HAV는 주로 오염된 물이나 식품 섭취를 통해 경구 감염되며, 특히 완전히 조리되지 않은 조개류는 주된 원인 식품이다. 이외에도 HAV 감염자와 밀접한 접촉을 통해서도 감염 가능하다. 국내 HAV 감염 현황을 살펴보면 2015년부터 2020년 동안 HAV 감염자는 표 1과 같다.

표 1. 국내 HAV 감염 환자 수

발생 년도	2015	2016	2017	2018	2019	2020[1]
감염자 수	1,804	4,679	4,419	2,437	17,596	1,222

[1]2020년 5월 9일 기준

전 세계적으로 HAV 감염 현황을 살펴보면 그림 1과 같이 HAV I.A 형은 유럽과 아시아 등 전 세계 고루 분포하며, HAV I.B는 사하라 사막 이남의 아프리카에서 우세하고, 중남미는 HAV I.A, 남아시아는 HAV III.A, 북아프리카와 중동에서는 HAV I.A와 HAV I.B가 많이 분포하는 것으로 나타났다. 이러한 HAV 감염에 대한 분석은 주로 VP1/A2 region 혹은 VP3/VP4 region으로 유전자 variant가 높은 sequence를 주요 타겟으로 삼아 분석되고 있다.

전장 유전체 분석 (whole genome sequencing, WGS)의 발달로 바이러스의 더욱 정확한 유전자형 파악과 변이에 대해 연구되고 있다. 차세대 염기서열 분석(next generation sequencing, NGS)이 널리 활용됨에 따라 방대한 바이오 빅데이터가 쏟아져 나오며 이를 활용한

* Correspondence to: Changsun Choi, D.V.M., Ph.D.
Department of Food and Nutrition, School of Food Science and Technology College of Biotechnology and Natural Resources
Chung-Ang University, Anseong-Si, Gyeonggi-do 17546, Republic of Korea
Telephone: +82-31-670-4589 Fax: +82-31-676-8741 E-mail address: cchoi@cau.ac.kr

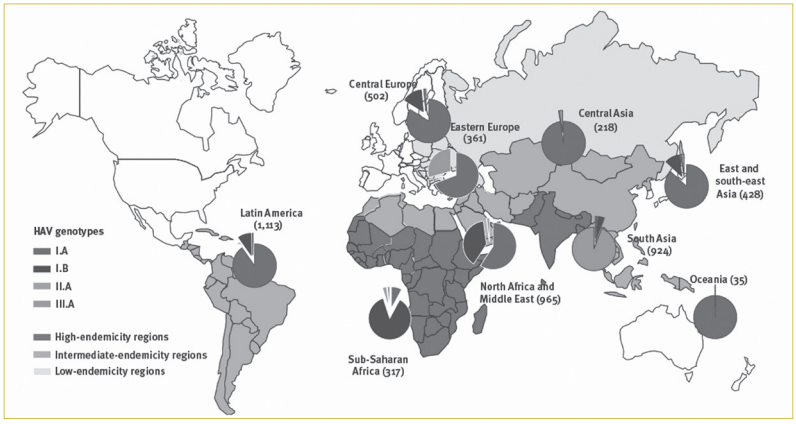

그림 1. 2010 – 2017년 국외 HAV genotype 별 감염 현황 (Annelies 외, 2018)

표 2. GenBank에 보고된 국가 및 기관별 HAV 전장유전체 분석 건수

국가	보고 건수 (%)
ATCC	3 (1.4 %)
중국	112 (51.6 %)
미국	22 (10.1 %)
일본	18 (8.3 %)
인도	10 (4.6 %)
남미	9 (4.1 %)
한국	7 (3.2 %)
몽골	6 (2.8 %)
유럽	6 (2.8 %)
인도네시아	6 (2.8 %)
아프리카	5 (2.3 %)
타이완	5 (2.3 %)
베트남	4 (1.8 %)
러시아	3 (1.4 %)
이란	1 (0.5 %)
합계	217 건

수많은 식중독 원인 병원체에 대한 연구들이 증가하는 추세이다. 바이러스 분야에서 WGS는 새로운 치료법 및 백신의 개발과 분자역학 (molecular epidemiology) 및 진화유전체학 (evolutionary genomics)에 있어 더욱 주목받는 연구 방법이다.

하지만, 표 2와 같이 국내에서 보고된 HAV의 전장 유전체는 7건으로 그 중 human HAV는 단 1건에 불과하여 reference gene으로 활용 가능한 HAV 유전자 database가 한정적인 현실이다. 특히, 국내에서 2019년 식품(조개젓)에 의한 HAV 감염자의 증가로 인하여 식품에서 분리된 HAV에 대해 더욱 빠르고 정확한 WGS이 필요하다.

본론

식품에서 바이러스 검출을 위한 전처리

식품에서 바이러스 검출은 크게 3단계로 "식품 탈리 · 농축 → 유전자 추출 → 유전자 검출" 과 같이 이뤄진다. 탈리 · 농축과 유전자 추출이 전처리 단계에 해당된다. 식품 시료에서 바이러스를 농축하는 방법으로는 guanidine isothiocyanate가 포함된 polyethylene glycol (PEG) 농축법, ultracentrifugation을 이용한 농축이 대표적으로 이용되는 방법이다. RNA extraction으로는 poly(dT) magnetic bead를 활용한 RNA extraction, ion chromatograph를 활용한 column법 및 phenol-chloroform을 이용한 방법들이 주로 이용된다. HAV는 RNA 바이러스로 검체로부터 RNA를 추출하여 유전자 검사에 사용된다.

1세대 유전자 분석 (First generation sequencing)

전처리 과정을 거쳐 추출된 RNA는 유전자 검출 과정을 통해 species를 추정 할 수 있으며 유전자 분석을 통해 genotype과 subgenotype을 판별한다. 유전자 정보 분석 (sequencing)을 위해 특정 바이러스 유전자의 region을 증폭한 후, 증폭 산물은 미생물 내에 클로닝 벡터 삽입을 통해서 *in vivo* 상태에서 복제 및 증폭하여 이들의 염기서열을 분석한다. 대표적인 1세대 유전자 분석 방법으로 sanger sequencing 가 있으며 높은 정확도와 상대적으로 긴 read length를 가지기 때문에 여전히 NGS와 함께 많이 사용되고 있다.

차세대 유전자 분석(NGS)

1세대 유전자 분석과 NGS의 가장 큰 차이점은 상대적으로 많은 양의 데이터를 만든다는 점에 있다. 특히나 식품 안전 분야에서 식중독 원인 병원체 판별은 관련된 질병간의 판별, 원인 식품 판별 및 해당 식품 중재 전략 수립(intervention strategies)에 있어 중요하다. WGS를 이용한 바이러스 유전자 분석은 1세대 유전자 분석과 같이 단편적인 분석 기술로는 놓칠 수 있는 여러 질병간의 연관성을 분석 할 수 있다. 또한, 독성 유전자 (virulence gene) 검출 등 2차 분석으로도 활용 할 수 있다.

NGS에서 library 제작 단계는 3가지로 나눌 수 있다. 유전물질 단편화(fragmentation), 어댑터 부착(adaptor ligation) 그리고 library 정량 및 사이즈 확인 과정 (quality control, QC)으로 이뤄진다. 바이러스는 세균에 비해 유전체 사이즈가 작고 다양한 유전체형으로 존재하기 때문에 중합효소연쇄반응 (polymeras chain reaction, PCR) 산물이나, double stranded cDNA 형태로 합성한 후 library을 제작하는 경우가 많다.

• Metagenomic sequencing

Metagenomic sequencing은 분석 대상 시료에서 추출한 전체 RNA를 특정 유전자 증폭이 필요 없이 shotgun method 또는 long read RNA sequencing에 의해 library를 형성하여, *de novo* assembling를 통한 분석 방법이다. 이 방법은 비교적 간단하고 효율적인 시료 준비가 가능하며 새로운 바이러스의 발견 및 염기서열 분석이 가능하다. 필요한 PCR cycle수가 적어 증폭 도중 돌연변이가 거의 발생하지 않으며 새로운 바이러스 또는 염기서열 변이에 대한 primer 또는 probe 설계가 필요하지 않아, 새로운 바이러스 발견이나 분석에 용이하다. 하지만, 충분한 데이터를 얻기 위해 필요한 read의 수가 많아 sequencing에 드는 비용이 큰 편이다. 또한 무작위적인 library 형성 기반이기 때문에 host의 genomic DNA (gDNA)와 같은 비병원체 read의 비율이 높아 상대적으로 낮은 감도를 가진다.

• PCR amplification sequencing

PCR 증폭을 통한 sequencing은 알려진 nucleotide 서열에 상보적인 primer를 이용하여 바이러스 유전자에 대한 증폭을 통해 분석하는 방법으로 NGS 중 가장 일반적인 접근법이다. Sequencing 비용이 상대적으로 저렴하며 시료 내에 바이러스의 양이 적은 경우에도 적용 가능한 매우 높은 민감도를 가진다. 식품 시료와 같은 다량의 host genome을 보유한 시료에서는 PCR amplification sequencing이 가장 효과적인 바이러스 검출법으로 기대된다.

하지만, 다양한 종류의 바이러스 검출이 어려우며, 다양한 바이러스를 분석하고자 할 경우 PCR을 반복하기 위해 많은 양의 genome 샘플이 필요하다. 또한, 변이가 많은 바이러스나, 새로운 바이러스를 검출하는데 있어 PCR 반응은 primer 불일치에 의한 영향을 받을 수 있으며 PCR cycle 수가 많은 경우 증폭에 의한 돌연변이가 발생할 수 있다.

• Target enrichment sequencing

Target enrichment (full-down, capture, 또는 specific enrichment 법으로도 불린다)를 사용하여 사전 배양 또는 PCR 없이 시료에서 직접 전체 바이러스 genome을 sequencing 할 수 있다. 이 방법은 바이러스의 염기서열에 상보적인 small RNA 또는 DNA probe를 이용한다. Metagenomic sequencing보다 높은 특이성으로 sequencing에 소요되는 비용이 적으며, overlapping probe는 각각의 primer의 불일치를 줄일 수 있다. 또한, PCR amplification 방법에 비하여 적은 PCR cycle를 통한 증폭으로 유전자의 변이가 적어 바이러스 내 변이를 잘 반영 할 수 있다.

단점으로는 샘플 준비를 위해 비용이 많이 드는 편이며, 새로운 바이러스 염기서열 분석이 쉽지 않다. 분석하고자 하는 바이러스에 대한 probe 설계를 위해 잘 특성화 된 reference gene이 필요하고 이에 대한 비용과 시간이 신종 바이러스인 경우 더욱 많이 소요된다. 민감도는 PCR과 비

슷하지만 바이러스에 대한 coverage가 낮은 편에 속한다.

2세대 NGS

대규모의 병렬적인 분석을 기반으로 한 2세대 NGS는 2005년 Roche에 의해 454 GS20의 형태로 처음 소개되었고, 이어서 2007년에 Illumina의 genetic analyzer II가 등장하였다. 기종에 따라 약간의 차이는 있지만 기본적으로 DNA 서열에 대한 증폭을 하고 그 후 형광 표식 등을 카메라로 찍어 이미지 처리를 하는 과정을 거쳐 염기를 읽어낸다. PCR 증폭 방식에 따라서 emulsion PCR을 이용하는 Roche와 Life Technologies의 SOLiD system과 solid-phase amplification을 사용하는 Illumina의 genome analyser system으로 나눌 수 있다. 이러한 유전자 분석법은 다량의 병렬적 sequencing을 기반으로 염기 서열 합성(synthesis)에 의해 sequencing된다. 이는 주형 template이 증폭되면서 sequencing이 진행는 방식으로 인해 요구되는 starting material 양이 적어도 된다는 장점이 있어 식품에서 분리된 바이러스를 분석하기에도 적합하다. 또한, 1세대 분석과 같이 높은 정확도를 가진다. 하지만, 분석 되는 원리에 의해 상대적으로 적은 read length를 가지는 한계점이 있다.

• Roche platform

Roche paltform의 분석 원리는 "DNA 단편화(fragmentation), 증폭, pyrosequencing" 3단계로 나눌 수 있다. Nebulization을 이용하여 gDNA 단편을 만들고 sulphurylase와 luciferase가 포함된 enzyme bead에 하나의 DNA 사슬을 고정한 후, emulsion-based PCR로 gDNA를 증폭을 시킨다. 증폭 과정은 염기가 주형에 결합하는 sequencing-by-synthesis (SBS)에 기반을 두고 있다. 증폭이 끝나면 각 bead에는 수 백 만개로 복제된 동일한 DNA 서열이 덮이게 되고, 각 bead는 PicoTiterPlate라는 장치에서 하나의 구멍(well)에 하나씩 들어가서 pyrosequencing이 개시된다. 이는 한번의 running으로 상대적으로 긴 서열을 읽을 수 있다는 장점이 있지만 동일한 염기 서열이 길게 반복되는 homopolymer(동일단위체)가 생성 될 경우 정확한 서열을 알 수 없으며 indel error를 가진다.

가장 대표적인 454 시리즈는 reference 서열이 없이 read를 서로 조립을 통해 *de novo* sequencing에 유리하다. Read length는 400 bp에 이르며, 10시간의 run time으로 35 Mb의 SFF 형식의 데이터를 생산한다. 하지만, 한번의 run에서 생산되는 데이터양이 다른 기종에 비해서 상대적으로 적다.

• Illumina platform

Illumina platform의 NGS 기법은 Roche platform와 같이 염기가 주형에 결합하는 SBS 증폭을 기본으로 한다. DNA를 단편으로 자른 뒤, 양쪽에 다른 종류의 adaptor를 붙여 만든 library를 이용한다. 제작된 library를 adaptor와 상보적인 oligo가 붙어있는 flow cell 슬라이드에 부착하여 bridge 증폭(bridge amplification)을 수행한다. Flow cell이 제조되는 과정에서, 표면은 통상적으로 "p5"와 "p7"으로 불리는 두 개의 올리고핵산염(oligonucleotide)이 코팅되어 있는데, 어댑터가 부착된 단일가닥 sequencing library가 flow cell에 로드(load)된다. 이후 library 내 각 분자는 올리고를 흘러가면서 상보적인 올리고에 결합하며 세워진다. 중합효소에 의해 올리고의 반대쪽(위쪽) 방향으로 중합이 일어나고, 변성 과정 후 원래의 sequencing library는 씻겨 진다. 남은 단일가닥은 bridge 증폭 과정을 통해 증폭되는데, 이 과정에서 단일가닥이 접혀지면서 어댑터 부위가 flow cell에 코팅된 다른 올리고핵산염과 결합한다. 중합효소에 의해 이 중가닥으로 중합된 bridge는 변성 후 두 개의 단일가닥으로 flow cell에 세워진다. 이러한 bridge 증폭 과정이 반복되어 수천만 개의 cluster가 생성된다. Bridge 증폭 후, 역방향 단일가닥은 분리되고 정방향 단일가닥만 남게 된다. Cluster amplification이 이뤄진 다음 SBS 과정을 통해 염기서열 분석이 이뤄지게 되는데, SBS technology는 4개의 labeling 된 nucleotide를 이용하여 flow cell

표면에 평행하게 세워져 있는 수 천 만개의 cluster를 sequencing 한다.

Illumina platform의 시리즈는 전반적으로 저렴한 running 가격으로 많은 데이터를 생산하는 특징이 있어 대용량의 sequencing을 대량으로 처리하는 연구에서 많이 사용된다. 또한 데이터 형식은 FASTQ로 분석 소프트웨어에서 대부분 치환가능하다. Illumina platform의 대표적인 기종인 HiSeq 2000는 2 X 100 bp read에서 8일까지의 run time으로 최대 540~600 Gb까지 데이터 생산이 가능하다. HiSeq의 기법 (chemistry)을 그대로 유지하면서 소형화에 성공한 MiSeq은 장비 크기와 가격을 줄이고 작업을 더 빠르고 간편하게 만들었다. MiSeq은 2 X 150 bp에서 최대 24시간의 run time으로 2.0~4.6 Gb의 데이터를 생산한다.

Illumina의 sequencing는 DNA 단편으로부터 진행되는 분석 방식으로 RNA 바이러스를 PCR amplicon을 기반으로 분석하기에 용이하다. 임상 시료에 미량 미량으로 존재하거나, 다량의 inhibitor를 가지는 식품 시료에서 분리되는 바이러스의 특성상 위와 같은 sequencing이 분석에 가장 적합할 것으로 보인다.

3세대 NGS

3세대 NGS는 단일 분자 시퀀서(single-molecule sequencer)로 single molecular에서 hairpin sequence를 만들고 이를 기반으로 consensus sequence를 만드는 기술이다. PacBio (Pacific Biosciences) 및 MinION (Oxford Nanopore)와 같은 단일 분자 시퀀서는 sequencing 가능한 read 길이가 매우 길어, 이론적으로는 단일 판독을 통해 바이러스 전체 유전자 분석이 가능하다. 단일 분자 시퀀서의 염기서열 분석은 기존 2세대 NGS의 경우 형광 dye의 감도를 이용한 것에 비해 전기적인 신호를 바로 감지하는 방식으로 sequencing 된다.

3세대 NGS를 이용하여 임상 시료 (혈청 등)로부터 사전 배양 후 에볼라 바이러스, 코로나19 바이러스, 헤르페스 바이러스 등을 분석한 연구 결과들이 발표되고 있다. HAV를 분석함에 있어 3세대 NGS platform과 같이 read length가 긴 platform을 이용할 경우 한번의 run으로 WGS를 할 수 있다는 장점이 있으나 식품 시료의 gDNA와 ribosomal RNA (rRNA) 및 fragmented DNA/RNA를 포함 할 경우 nanopore의 flow cell 막힘 현상을 유발하여 성공적인 분석이 이뤄지기 힘들다. 때문에 rRNA 제거 및 RNA purification과 같은 단계가 추가로 요구된다.

• Nanopore platform

Nanopore sequencer는 전자적으로 저항성이 있는 막에 직경 나노 (10억분의 1)미터의 극소의 구멍을 사이를 DNA 분자가 통과하면서 막에 흐르는 전류의 흐름을 변형시키는 원리를 기반으로 한다. 네 종류의 염기마다 서로 다른 세기의 전류가 흘러 각 염기에 대한 전류를 측정한다. 이러한 분석 방법은 fluorescence dye와 같은 광학적인 표지 혹은 분석 대상 바이러스의 증폭 없이 염기서열 분석이 가능한 방법으로 분석체를 직접적이고, 전자적인 방법으로 확인할 수 있다. 이러한 플랫폼의 유연성과 이식성으로 인해 분석 대상 바이러스 염기서열에 대한 실시간 검사에 적절하다.

MinION은 경량화 된 USB 3.0 인터페이스 장치를 기반의 Oxford Nanopore에서 판매하는 소형의 sequencing 장비로 RNA sequencing이 가능하다. MinION을 이용하여 RNA를 분석하는 경우 adapter attachment를 위해 primer annealing 시 poly A tail 부분에 primer가 부착하도록 디자인 되어 있기 때문에 mRNA를 사용하는 것이 권장된다. 분석 시료에 poly A tail이 포함되지 않는 경우에 따라서는 polymerase를 이용하여 poly A tailing을 진행한다.

RNA 분석에 있어서 쓰이는 sequencing 방법은 크게 3가지로 “Direct RNA sequencing”, “Direct cDNA sequencing”, “PCR-cDNA sequencing” 이 있다.

• Direct RNA sequencing

Reverse transcription (RT) 과정, PCR 과정 없이

input RNA 그대로 sequencing 하기 때문에 다른 방법에 비해 input mass가 많이 필요하다. Input되는 시료는 500 ng poly A RNA를 권장한다. 이론적인 read length는 RNA length와 같으며 read number는 1,000,000이다. Preparation time은 105분정도이다.

• Direct cDNA sequencing

RT 과정을 거친 후 double strand 양쪽 끝에 adapter를 붙여 sequencing한다. Input되는 시료는 100 ng poly A RNA를 권장한다. 이론적인 read length는 enriched 된 full cDNA length와 같으며 read number는 5,000,000 ~ 10,000,000 이다. Preparation time은 270 분정도이다.

• PCR cDNA sequencing

RT 과정을 거친 후 PCR 과정으로 strand를 증폭시킨 후 adapter를 붙인 뒤에 sequencing한다. 증폭이 가능하기 때문에 앞선 두 가지 방법과 달리 input mass가 적은 양으로 가능하고, 그렇기 때문에 total RNA를 이용하여 sequencing 할 수 있다 (total RNA 내 mRNA 1~2% 존재한다는 가정 하). 특히나, 식품 시료에 존재하는 inhibitor가 PCR을 통해 1차적으로 제거된 상태에서 sequencing이 진행되기 때문에 분석하고자 하는 시료의 origin에 따라 효율적으로 이용할 수 있다. Input되는 시료는 1 ng poly A RNA 또는 50 ng total RNA를 권장하며 이론적인 read length는 PCR에 의해 증폭된 DNA의 length와 같다. Read number는 7,000,000 ~ 12,000,000 이며 preparation time은 165 분정도이다.

• Nanopore RNA QC

올바른 분석을 위해 library 형성은 중요한 단계로 library 형성을 위해 적절한 RNA quality 가 확보되어야 한다. 화학적 불순물 (detergent, 변성제, chelating agents 혹은 고농도 염)은 효소 활성에 방해가 되어 library형성에 방해 될 수 있다. 그리고 DNA, 단백질 및 dye는 library 준비 단계의 효율을 감소시킬 수 있다.

RNA의 품질은 nanodrop을 이용하여 검사가 가능하며 분석에 권장되는 RNA quality는 260/280 값 2.0 미만이며, 260/230 2.0~2.2 이다. 260/280 값이 2.0 미만인 경우 시료에 DNA가 오염된 가능성이 크며, 260/280 값이 1.8 이하는 경우 단백질이나 페놀이 오염된 가능성이 크다. 또한, 260/230 값이 2.0~2.2이하인 경우 RNA의 순도가 떨어져 올바른 분석이 이뤄지기 어려우며, 이는 별도의 purification이 필요하다.

HAV WGS 분석

NGS 플랫폼은 RNA/DNA preparation, sequencing, imaging 및 분석에 대해 다른 접근 방식을 가진다 (표 3). 각 분석 platform의 차이에 따라 식품 시료에서 HAV를 분석하기 위한 효과적인 전략 수립이 필요하다.

식품 시료에서 분리된 HAV는 다량의 sequenciung inhibitor와 wild type의 사전 배양법 부재, 낮은 RNA purity 등 RNA sequencing에 어려움이 있다. 이러한 한계점으로 인해 식품에서 분리된 HAV를 분석할 경우 Illumina와 같은 DNA 증폭을 기반과 낮은 error rate (0.58 %)을 가지는 sequencing이 더욱 유리할 것으로 보인다. HAV 분석을 위해 target이 되는 HAV RNA를 long PCR을 통해 증폭하고, 증폭된 PCR 산물을 이용하여 분석하는 workflow를 가질 것으로 예상된다 (그림 2).

3세대 NGS 장비를 이용하면 HAV 유전자 length를 충분히 cover 할 수 있는 장점이 있지만, sequencing 오류를 수정하고 variant를 정확히 분석하기 위해서 매우 높은 coverage가 요구된다.

결론

HAV로 인한 식중독은 증상이 나타나기까지 2~6 주가 소요되어 식중독의 원인이 되는 식품을 파악하기에 어려움이 있다. NGS를 이용한 WGS으로부터 생성된 데이터

표 3. NGS platform에 따른 RNA 바이러스 분석 시 장 · 단점

NGS platform		장점	단점
1세대 유전자분석		높은 정확도	Input 시료량이 많음
			다량의 데이터 확보 어려움
NGS			
	Metagenomic	시료 준비 용이	낮은 감도
		낮은 돌연변이 발생률	
		primer/probe 설계 불필요	
		새로운 바이러스 발견 용이	
	PCR amplification	높은 민감도	Multiplexing 적용 어려움
		적은 input 시료량	새로운 바이러스 발견 어려움
	Target enrichment	적은 PCR cycle	Workflow가 복잡함
		적은 input 시료량	새로운 바이러스 발견 어려움
		높은 특이성	낮은 바이러스 coverage
2세대 NGS			
	454 GS	적은 input 시료량	낮은 throughput
			Indel error
	Illumina Miseq	적은 input 시료량	짧은 read length
		낮은 error rate	
		짧은 running time	
		Multiplexing 적용 용이 및 높은 정확도	
3세대 NGS			
	MinION Nanopore	긴 read length	낮은 정확도
		높은 유연성 및 이식성 분석 대상 바이러스 증폭 불필요	높은 시료 purity

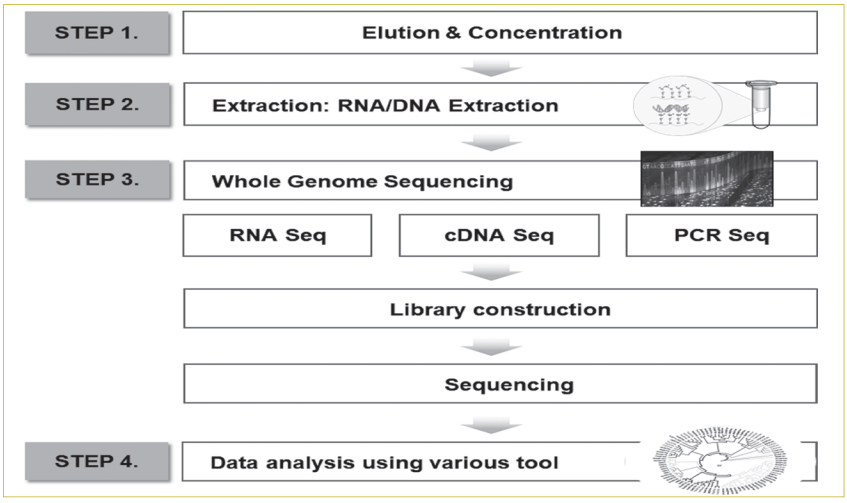

그림 2. 식품에서 분리된 바이러스 NGS 분석 workflow

는 계통 발생적 관계를 밝히는 데 사용될 수 있으며 질병을 일으키는 특정 strain에 대한 추적 및 모니터링이 가능하다.

NGS를 이용한 분석 시 RNA 바이러스의 경우 DNA 바이러스나 균에 비하여 1~3 bases deletion이 잦은 편이기 때문에, sequencing error를 줄이기 위해 높은 coverage가 필요하다. 2세대 NGS는 read length가 짧다는 한계가 있으나 높은 정확도를 가

지고, 3세대 NGS는 read length는 길지만 낮은 정확도를 가진다. 이러한 특징으로 RNA 바이러스 분석에는 단일 장비를 이용하기보단, 2개 이상의 장비를 이용한 hybrid sequencing으로 한계점을 보완하는 것이 적절하다.

위와 같은 NGS를 이용한 HAV 전장 유전체 분석은 국내에서 발생하는 HAV 감염 사고에 대한 방대한 분석과 reference gene 확립으로 HAV에 의한 식중독 사고 예방을 기대한다.

감사의 글

본 연구는 2020년도 식품의약품안전처의 연구개발비(20162MFDS033)로 수행되었으며 이에 감사드립니다.

참고 문헌

1. 박근준. NGS (Next Generation Sequencer) 기법 소개. 생화학분자학회 2012, 32(2), 22-30.

2. 양효진, 최현정, 채희열, 강연호. 병원체 염기서열 생산을 위한 라이브러리 제작 기법 소개. 질병관리본부., 주간 건강과 질병 2017, 10(44), 1178-1182.

3. 정은경. 2019년도 A형 간염 유행에 따른 심층역학조사 결과 발표. 2019. https://www.cdc.go.kr/board.es?mid=a20501000000&bid=0015&act=view&list_no=364837&tag=&nPage=1

4. Annelies K, Rita S, Linda V, Marion K, Harry V, on behalf of the HAVNet network. Usability of the international HAVNet hepatitis A virus database for geographical annotation, backtracing and outbreak detection. Eurosurveillance 2018, 23, 1700802.

5. Charlotte H, Mathew B, Judith, B. Clinical and biological insights from viral genome sequencing. Nature review Microbiology 2017, 15, 183-192.

6. Jianping X. Next-generation Sequencing (current technologies and applications). Caister Academic Press. 2014.

7. Michael L, Metzker. Sequencing technologies-the next generation. Nature Reviews Genetics 2010, 11, 31-46.

8. Omana N, Guoliang X, Gilberto V, Harold M. Diagnosis of Hepatitis A Virus Infection: a Molecular Approach. Clinical Microbiology 2006, 19(1), 63-79.

9. Ronholm J, Nasheri N, Petronella N, Pagotto F. Navigating Microbiological Food Safety in the Era of Whole-Genome Sequencing. Clinical Microbiology Reviews 2016, 29, 837-857.

10. Steven H, Phillip O, Daniel S. Next-generation Sequencing Method and Protocols. Humana Press. 2018.