

복잡한 샘플 내 미생물 군집의 연구를 위한 샷건 메타지노믹스 NGS 워크플로우

미생물 종의 동정에 필요한
정확성과 유연성을 제공하는
NextSeq™ 1000 및
NextSeq 2000 시스템과
600사이클 키트



복잡한 샘플의 균유전체(metagenome) 분류

샷건 메타지노믹스(Shotgun metagenomics)는 복잡한 샘플 내 미생물 다양성(diversity)을 연구할 때 16S 리보솜 RNA(ribosomal RNA, rRNA) 시퀀싱이나 internal transcribed spacer(ITS) rRNA 시퀀싱과 같은 앰플리콘(amplicon) 시퀀싱 방법에 대한 대안을 제시하는 시퀀싱 기법입니다. 차세대 시퀀싱(Next-generation sequencing, NGS) 기술을 활용하는 샷건 메타지노믹스는 앰플리콘 시퀀싱과는 달리 특정 샘플에 존재하는 모든 유기체에 대한 종합적인 유전체(genome) 정보를 포획합니다. 샷건 메타지노믹스는 이렇게 유전체 전체를 포획하는 능력을 바탕으로 앰플리콘 시퀀싱¹으로는 놓칠 수 있는 종(species)을 동정(identification)할 수 있으며, 앰플리콘 시퀀싱은 제공하지 않는 기능 정보도 데이터에 포함합니다.^{2,3}

이 Application Note는 대용량(high-throughput) 샷건 메타지노믹스 연구에 활용 가능한 NextSeq 1000, NextSeq 2000 및 MiSeq™ 시스템의 비등한 성능에 대해 기술합니다. 또 NextSeq 550 시스템으로 생성한 데이터를 사용했을 때 메타지노믹스 애플리케이션에서 일반적으로 사용되는 300사이클 키트보다 600사이클 키트가 더 큰 이점을 제공함을 설명합니다. 아울러 600사이클 키트의 뛰어난 속(genus) 및 종 동정 능력을 보여주는 인공 군집과 실제 샘플로부터 얻은 데이터도 함께 제시합니다.

방법

NextSeq 1000/2000 P1 Reagents(600 cycles) 및 NextSeq 1000/2000 P2 Reagents(600 cycles) 키트는 샷건 메타지노믹스 연구에 적합한 사양을 기반으로 NextSeq 1000 및 NextSeq 2000 시스템의 역량과 시퀀싱 아웃풋을 높여 줍니다. NextSeq 1000 및 NextSeq 2000 시스템은 기기에 내장된 유체(fluidics) 없이 바로 로딩만 하면 되는 시약을 사용하므로 워크플로우의 단계와 샘플 오염의 위험을 줄여 줍니다. 샷건 메타지노믹스 워크플로우는 라이브러리 준비 단계, 입증된 Illumina의 NGS 시스템을 통한 시퀀싱 단계 그리고 간편하게 버튼만 누르면 되는 2차 데이터 분석 단계로 구성된 마이크로바이옴(microbiome)의 발견을 돕는 완전한 솔루션을 제공합니다(그림 1).

라이브러리 준비

미생물 유전체 DNA(genomic DNA, gDNA) 샘플의 출처는 두 가지입니다. 첫 번째 샘플로는 상용 샘플인 American Type Culture Collection(ATCC)의 20 Strain Staggered Mix Genomic Material(카탈로그 번호: MSA-1003)이 사용되었습니다. 이 ATCC 샘플은 그람 염색(Gram stain), GC 함량, 포자 형성(sporulation)과 같은 속성을 기준으로 선택한 세균 균주에서 얻은 gDNA가 비균질하게 분포(staggered distribution)하는 모의 미생물 군집(mock microbial community)입니다. 두 번째 샘플은 앞서 언급한 실제 샘플⁴로, 분변 샘플을 준비했습니다.

라이브러리는 Illumina DNA Prep, (M) Tagmentation (24 Samples, IPB) 제품(Illumina, 카탈로그 번호: 20060060)과 IDT for Illumina DNA/RNA UD Indexes Set A, Tagmentation(96 Indexes, 96 Samples) 제품(Illumina, 카탈로그 번호: 20027213)을 사용해 준비했습니다. IDT for

라이브러리 준비



Illumina DNA Prep:
약 3.5시간 소요
Illumina DNA PCR-Free:
약 1.5시간 소요

라이브러리 시퀀싱



NextSeq 1000/2000 P1 Reagents
(600 cycles): 약 34시간 소요

데이터 분석



BaseSpace™ Sequence Hub
SPAdes Genome Assembler
Kraken2 Metagenomics 앱

그림 1: NextSeq 1000 및 NextSeq 2000 시스템을 사용한 샷건 메타지노믹스 NGS 워크플로우

Illumina DNA/RNA UD Indexes Set A, B, C 및 D 제품을 사용하면 384개의 샷건 메타지노믹스 라이브러리를 생성할 수 있습니다.

시퀀싱

준비한 라이브러리를 풀링(pooling)한 후 사전 충전된 NextSeq 1000/2000 P1 Reagents(600 cycles) 키트의 플로우 셀, MiSeq Reagent Kit v3(600 cycles)의 플로우 셀(Illumina, 카탈로그 번호: MS-102-3003) 및 NextSeq 500/550 High Output Kit v2.5(300 cycles)의 플로우 셀(Illumina, 카탈로그 번호: 20024908)에 로딩했습니다. 라이브러리는 각각 NextSeq 2000 시스템, MiSeq 시스템 및 NextSeq 550 시스템으로 시퀀싱했습니다. 모든 런(run)의 대표 시퀀싱 데이터는 [BaseSpace Sequence Hub 데모 데이터](#) 웹 페이지에서 확인 가능합니다.

분석

풀링된 라이브러리는 BaseSpace Sequence Hub 유전체학 클라우드 컴퓨팅 플랫폼에서 디멀티플렉싱(demultiplexing)했습니다. 이후 NextSeq 2000, MiSeq 및 NextSeq 550 시스템에서 생성한 데이터를 DRAGEN™ Metagenomics 파이프라인을 이용해 처리했습니다. 균유전체는 SPAdes Genome Assembler를 사용해 조립했습니다. 분류학적 분류(Taxonomic classification)에는 DRAGEN Metagenomics 파이프라인을 활용했습니다.

600사이클 키트로 생성한 NextSeq 2000 데이터와 300사이클 키트로 생성한 NextSeq 500 데이터를 비교하기 위해, BaseSpace Sequence Hub에서 제공되는 DRAGEN FASTQ Toolkit 앱을 사용해 NextSeq 2000 리드(read)를

트리밍(trimming)했습니다. 샘플 간 비교를 위해 DRAGEN FASTQ Toolkit 앱으로 각 샘플을 같은 수의 리드(각각 30M 개, 10M 개, 1M 개)로 다운샘플링(downsampling; 데이터가 많은 쪽을 적게 추출)했습니다. 다운샘플링은 한 애플리케이션을 통해 특정 샘플의 하위집합(subset)만 처리할 수 있는 경우(예: 메모리 제약이 있는 *de novo assembly*(드 노보 어셈블리)), 또는 한 샘플의 처리에 전체 데이터 세트가 필요하지 않은 경우(예: 다양한 유전체 커버리지 레벨에서 방법을 검증하려 할 때)에 필요합니다.

결과

향상된 주요 시퀀스 메트릭스

Q-Score(Quality score, 품질 점수)가 Q30 이상인 염기(base)의 비율은 NextSeq 2000 시스템에서 NextSeq 1000/2000 P1 Reagents(600 cycles)를 사용한 런이 MiSeq 시스템에서 MiSeq Reagent Kit v3(600 cycles)를 사용한 런보다 더 높은 것으로 측정되었습니다. 또한 NextSeq 1000/2000 P1 Reagents(600 cycles)는 최대 100M 개의 필터를 통과하는 싱글 엔드 리드(single-end read), 또는 최대 200M 개의 필터를 통과하는 페어드 엔드 리드(paired-end read)를 제공합니다. NextSeq 1000/2000 P1 Reagents(600 cycles)는 MiSeq Reagent Kit v3(600 cycles)보다 4배 많은 데이터를 생성(약 60 Gb 대 약 15 Gb)합니다. 또한 NextSeq 1000/2000 P1 Reagents(600 cycles)를 사용한 시퀀싱 런은 약 34시간이 소요되어, MiSeq Reagent Kit v3(600 cycles)를 사용한 시퀀싱 런에 비해 20시간가량 더 빨리 완료되었습니다(그림 2).

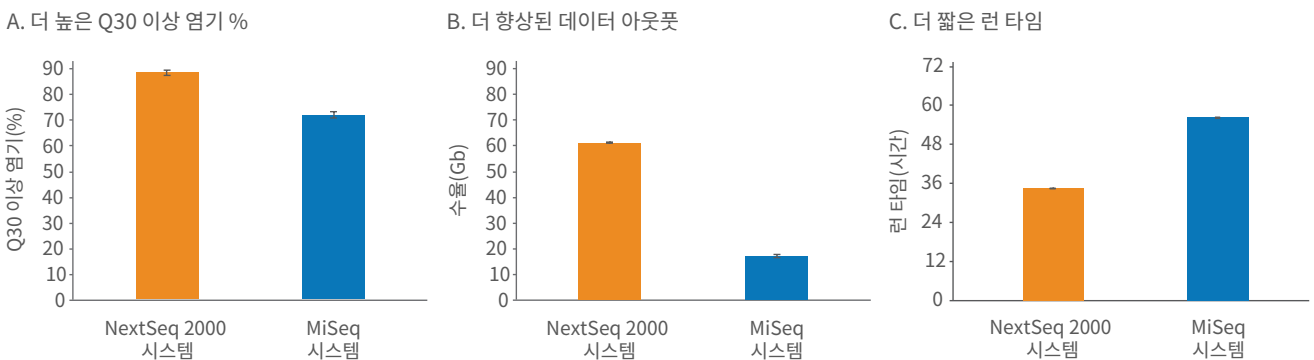


그림 2: NextSeq 2000 시스템과 MiSeq 시스템의 주요 성능 메트릭스 비교 — MiSeq 시스템에서 MiSeq Reagent Kit v3(600 cycles)를 사용한 런과 비교했을 때, NextSeq 2000 시스템에서 NextSeq 1000/2000 P1 Reagents(600 cycles)를 사용한 런은 (A) Q30 이상인 염기의 비율이 더 높고, (B) 4배 더 많은 데이터를 생성하며, (C) NextSeq P1 플로우 셀 사용 시 기기의 런 타임을 약 20시간 단축함.

NextSeq 500/550 High Output Kit v2.5(300 cycles) 제품(Illumina, 카탈로그 번호: 20024908)은 리드 길이(read length)가 2 × 150 bp일 때 약 29시간 안에 최대 120 Gb의 품질 높은 시퀀싱 데이터를 생성할 수 있습니다. 이 키트는 Q30 이상인 염기의 비율이 75%가 넘습니다(관련 데이터 포함하지 않음).

메타지노믹스 결과의 비교

NextSeq 2000, MiSeq 및 NextSeq 550 시스템의 성능을 비교하기 위해 각각의 시스템에서 20 Strain Staggered Mix Genomic Material을 시퀀싱했습니다. 분류학적 분류에 대한 상세한 정보를 얻기 위해 BaseSpace Sequence Hub에서 DRAGEN Metagenomics 앱을 사용해 후속 분석을 실시했습니다. 메타지노믹스 NGS를 통해 모의 미생물 군집의 모든 예상 구성원을 동정했고, 세 시스템 간 비슷한 결과가 관찰되었습니다(그림 3).

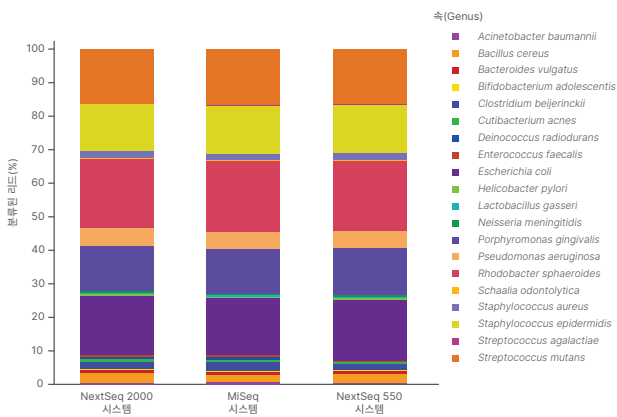


그림 3: NextSeq 2000, MiSeq 및 NextSeq 550 시스템으로 분석한 ATCC 20 Strain Staggered Mix Genomic Material의 미생물 조성 비교 — DRAGEN Metagenomics 앱의 미생물 조성 분석은 우수하고 재현 가능한 미생물 속의 정체(identity) 및 분포 데이터를 제공합니다.

다만 ATCC 샘플이 높은 재현성을 보인다 해도, 모의 샘플로 얻은 결과는 실제 샘플을 사용했을 때의 성과 차이가 있을 수 있습니다. 따라서 Illumina는 실제 분변 샘플을 이용해 유기체 검출 성능을 확인했습니다. 분변 샘플에서 검출된 20가지 대표적인 속을 가지고 NextSeq 2000, NextSeq 550 및 MiSeq 시스템 간 성능을 비교했습니다(그림 4). 이러한 복잡한 샘플은 검출 결과가 그다지 균일하지 않지만, 모든 실제 분변 샘플의 메타지노믹스 군집 프로파일은 NextSeq 2000, MiSeq 및 NextSeq 550 시스템 간 일치성이 매우 높았습니다. 이는 연구자가 세 가지 시스템에서 다양한 종류의 샘플을 사용해 신뢰할 수 있는 메타지노믹스 연구를 수행할 수 있음을 의미합니다.

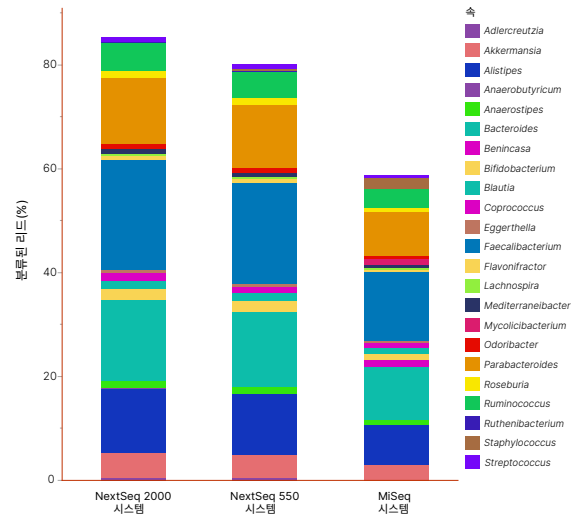


그림 4: 메타지노믹스 연구에서 세 가지 시스템 간 확인된 대표적인 실제 분변 샘플의 유사한 군집 조성 — DRAGEN Metagenomics 앱을 통해 NextSeq 2000, NextSeq 550 및 MiSeq 시스템으로 시퀀싱한 실제 분변 샘플에서 검출된 대표적인 20가지 미생물 속의 조성을 분석한 데이터로, 복잡한 샘플을 사용했음에도 세 가지 플랫폼 간 비슷한 속 커버리지가 관찰됨.

더 긴 리드로 향상되는 샘플 특성 확인 능력

300사이클 플로우 셀도 유의미한 메타지노믹스 데이터를 제공하지만, 복잡한 샘플에서 개별 종을 동정할 때는 600사이클 키트가 더 유용합니다. 더 긴 리드가 균유전체 조립에 미치는 영향을 확인하기 위해 DRAGEN FASTQ Toolkit 앱을 사용해 NextSeq 2000 시스템으로 생성한 리드를 600사이클에서 300사이클로 트리밍했습니다. 600사이클 또는 300사이클일 때 샘플별 분류된 리드의 비율은 k-mer 기반의 분류학적 분류기(taxonomic classifier)인 Kraken2*를 사용해 계산했습니다(그림 5). 해당 데이터에 따르면, 다양한 환경 샘플 분석 시 더 긴 리드를 선택하면 k-mer 기반의 분류학적 분류 성능이 약간 향상되는 것이 확인되었습니다.

* Kraken2 Metagenomics 분류학적 분류는 Illumina의 DRAGEN Metagenomics BaseSpace 앱을 통해 실행 가능.

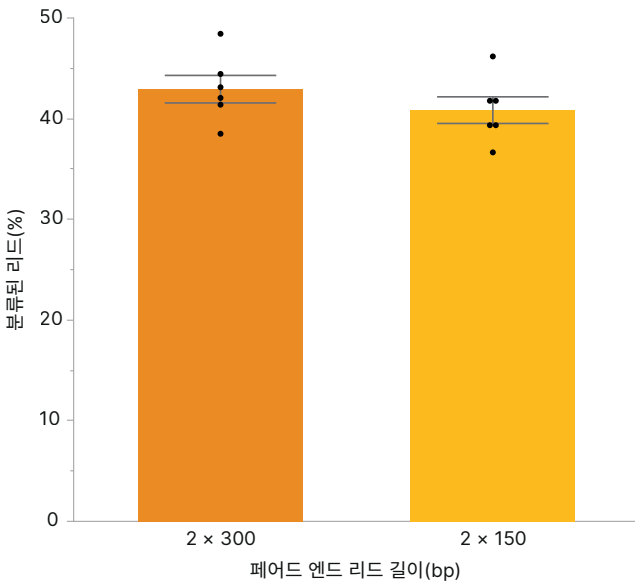


그림 5: 더 긴 리드 적용 시 향상되는 NextSeq 2000으로 시퀀싱한 실제 분변 샘플의 리드 분류 성능 — DRAGEN FASTQ Toolkit 앱을 사용해 NextSeq 2000 시스템으로 생성한 리드를 600사이클(2 x 300 bp)에서 300사이클(2 x 150 bp)로 트리밍한 후, Kraken2를 사용해 샘플별로 분류된 리드의 비율을 확인함. 이 분석의 리드 뎁스(read depth)는 30M 개의 리드로 설정함. 오차 막대는 평균으로부터 한 표준 오차(one standard error from the mean)를 나타냄.

또한 동일한 트리밍된 데이터를 사용해 리드 길이가 샘플 풍부도 (richness; 샘플에서 검출되는 종의 수) 및 샤논 지수(Shannon index; 샘플에서 검출되는 종을 비례적으로 표현)에 미치는 영향도 측정했습니다. 이러한 매트릭스를 살펴보면, 리드가 길수록 분변 샘플의 미생물 다양성은 증가하는 것으로 관찰된 반면 샤논 지수로 정량화한 검출된 종의 비율은 예상한 바와 같이 상대적으로 변화가 없었습니다(그림 6).

더 높은 시퀀싱 뎁스로 향상되는 샘플 특성 확인 능력

이어서 실제 분변 샘플의 군집 다양성 측정 시 리드 뎁스의 중요성을 살펴보았습니다. NextSeq 2000 시스템으로 생성한 리드는 DRAGEN FASTQ Toolkit 앱을 사용해 30M 개, 10M 개 및 1M 개의 리드로 다운샘플링했고, 풍부도와 샤논 지수는 DRAGEN Metagenomics 앱으로 계산했습니다. 다양성 매트릭스를 보면, 시퀀싱 뎁스가 높을수록 분변 샘플의 미생물 다양성은 증가한 반면 샤논 지수로 표현한 종의 비율은 상대적으로 변화가 없었습니다(그림 7).

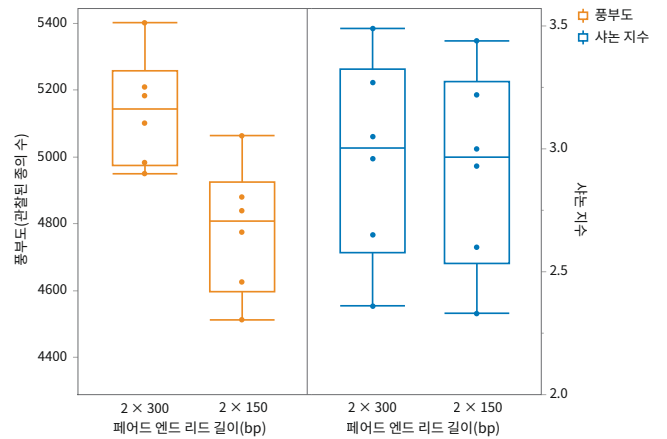


그림 6: 리드가 길수록 증가하는 분변 샘플의 미생물 풍부도 — DRAGEN FASTQ Toolkit 앱을 사용해 NextSeq 2000 시스템으로 생성한 리드를 600사이클(2 x 300 bp)에서 300사이클(2 x 150 bp)로 트리밍한 후, DRAGEN Metagenomics 앱을 사용해 풍부도와 샤논 지수를 계산함으로써 검출된 종의 수와 비례적인 다양성을 각각 정량화함. 리드가 길수록 풍부도는 증가했지만 샤논 지수는 상대적으로 변화 없이 유지됨. 이 분석의 리드 뎁스는 30M 개의 리드로 설정함.

더 긴 리드로 향상되는 분류학적 동정 능력

연구자들이 다양한 환경 미생물 군집을 프로파일링할 때 직면하는 어려움 중 하나는 배양이 불가능한 여러 희귀 종에 대한 완전한 참조 유전체(reference genome)의 부재입니다. 일반적으로 매우 다양한 미생물 군집의 경우 리드가 길수록 조립된 콘티그(contig)의 수가 더 많은데, 이는 조립된 전체 길이가 더 긴 것을 보면 알 수 있습니다(그림 8). Illumina의 2 x 301 bp 시퀀싱 리드 길이로 샷건 메타지노믹스 수행 시 일반적으로 환경 샘플에서 얻은 균유전체의 *de novo* assembly 성능이 향상되며, 이는 각각의 조립된 균유전체의 전반적인 완전성(completeness)에도 크게 기여합니다.

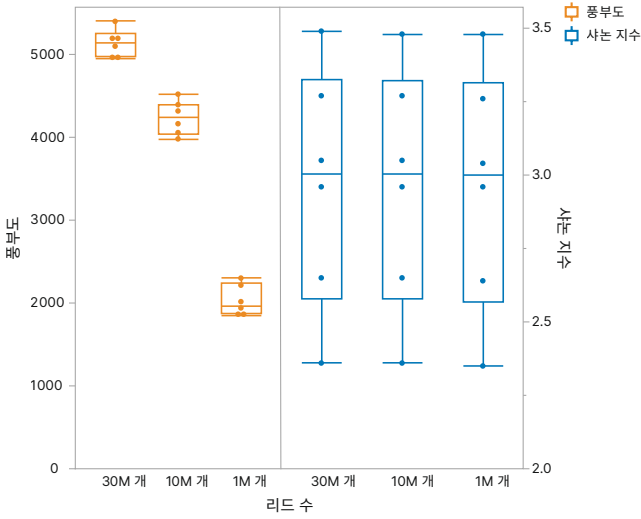


그림 7: 시퀀싱 데프스가 높을수록 증가하는 분변 샘플의 미생물 풍부도 — DRAGEN Metagenomics 앱으로 풍부도와 사슴 지수를 계산해 30M 개, 10M 개 및 1M 개의 리드 시퀀싱을 통해 검출된 종의 수와 비례적인 다양성을 측정 한 결과, 예상대로 다운샘플링한 데이터에서 풍부도는 감소한 반면 사슴 지수는 거의 동일하게 유지됨.

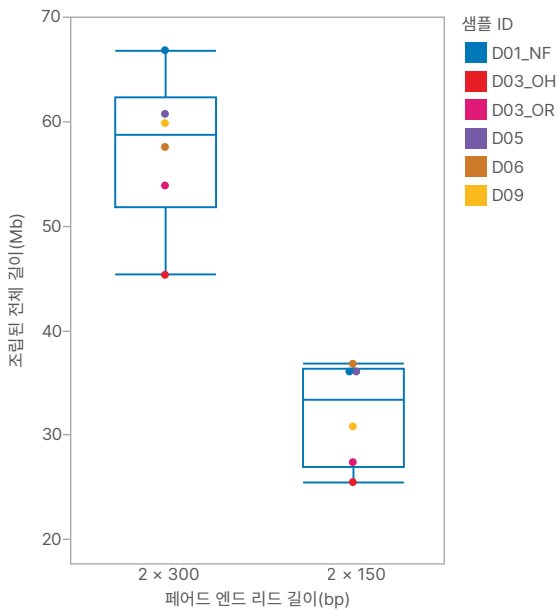


그림 8: 리드가 길수록 전반적인 증가가 관찰된 NextSeq 2000 시스템으로 시퀀싱한 실제 분변 샘플의 조립된 염기 수 — DRAGEN FASTQ Toolkit 앱을 사용해 NextSeq 2000 시스템으로 생성한 메타지노믹스 데이터를 1M 개의 리드로 그리고 2 x 300 bp에서 2 x 150 bp의 리드 길이로 트리밍함. 비교를 위해 SPAdes Genome Assembler를 사용해 2 x 300 bp 및 2 x 150 bp로 설정된 시퀀싱 리드로부터 전체 콘티그 길이를 얻음.

요약

이 Application Note는 600사이클 키트가 NextSeq 2000 시스템과 MiSeq 시스템에서 비슷한 성능을 보임을 설명합니다. NextSeq 550 시스템 또한 정확한 균유전체 분석이 가능하지만 600사이클 옵션이 제공되지 않습니다. 세 가지 시스템 모두 사용한 기기에 관계없이 정확한 *de novo* assembly 성능을 보였습니다. 일관성 있는 균유전체 프로필, 그중에서도 특히 관찰된 속의 비례적인 다양성이 세 가지 시스템 모두에서 확인되었습니다.

NextSeq 2000 시스템은 추가적인 리드 길이 및 리드 뎀스를 지원하므로 다양한 미배양(culture-free) 샘플로 풍부도 세부 정보를 분석할 때 MiSeq 시스템과 NextSeq 550 시스템보다 더 큰 이점을 제공합니다.[†] 이러한 장점은 분변 샘플이나 환경 샘플과 같이 복잡한 샘플을 사용하는 메타지노믹스 연구에 특히 유용합니다.

요약하면, NextSeq 1000 및 NextSeq 2000 시스템에서 NextSeq 1000/2000 P1 Reagents(600 cycles) 및 NextSeq 1000/2000 P2 Reagents(600 cycles) 키트를 사용하면 MiSeq 시스템에서 MiSeq Reagent Kit v3(600 cycles)를 사용할 때보다 더 짧은 시간 안에 높은 품질의 시퀀싱을 수행할 수 있습니다. 앞서 확인된 바와 같이, NextSeq 1000 및 NextSeq 2000 시스템에서 600사이클 키트 사용 시 높은 Q30 점수와 낮은 오류율을 유지할 수 있습니다. 또한 NextSeq 1000/2000 P1 Reagents(600 cycles)와 NextSeq 1000/2000 P2 Reagents(600 cycles)는 중간 규모나 작은 규모의 유전체 시퀀싱 실험에 적합한 유연한 데이터 아웃풋을 제공합니다. NextSeq 1000 및 NextSeq 2000 시스템에서 600사이클 키트를 사용하면 이미 성능이 입증된 MiSeq 시스템이 제공하는 데이터 품질은 그대로 유지하면서 애플리케이션을 확장하고 운영의 간편성을 제고할 수 있습니다.

[†] NextSeq 1000 시스템은 기능 면에서 NextSeq 2000 시스템과 유사하며 동일한 600사이클 키트 사용 시 비슷한 결과를 기대할 수 있음.

상세 정보

Illumina의 시퀀싱 플랫폼

NextSeq 1000 및 NextSeq 2000 시스템

NextSeq 1000/2000 시약

Illumina DNA Prep

Illumina DNA Prep crude lysate protocol for NGS

참고 문헌

1. Peterson D, Bonham KS, Rowland S, Pattanayak CW; RESONANCE Consortium, Klepac-Ceraj V. [Comparative Analysis of 16S rRNA Gene and Metagenome Sequencing in Pediatric Gut Microbiomes](#). *Front Microbiol*. 2021;12:670336. Published 2021 Jul 15. doi:10.3389/fmicb.2021.670336
2. Durazzi F, Sala C, Castellani G, Manfreda G, Remondini D, De Cesare A. [Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota](#). *Sci Rep*. 2021;11(1):3030. Published 2021 Feb 4. doi:10.1038/s41598-021-82726-y
3. Stothart MR, McLoughlin PD, Poissant J. [Shallow shotgun sequencing of the microbiome recapitulates 16S amplicon results and provides functional insights](#). *Mol Ecol Resour*. 2023;23(3):549-564. doi:10.1111/1755-0998.13713
4. Illumina. 16S rRNA sequencing on NextSeq 1000 and NextSeq 2000 Systems. [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/nextseq-600c-16s-rrna-application-note-m-gl-01146/nextseq-600c-16s-rrna-application-note-m-gl-01146.pdf](https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/nextseq-600c-16s-rrna-application-note-m-gl-01146/nextseq-600c-16s-rrna-application-note-m-gl-01146.pdf). Published 2023. Accessed February 6, 2024.



무료 전화(한국) 080-234-5300
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. All rights reserved.
모든 상표는 Illumina, Inc. 또는 각 소유주의 자산입니다.
특정 상표 정보는 www.illumina.com/company/legal.html을 참조하십시오.
M-GL-01147 v1.0 KOR