METHODS GUIDE

Methods for microbial single-genome sequencing

Accurately sequence known organisms for infectious disease and microbiology applications

Table of contents

01 Introduction

02 Method 1: Microbial whole-genome sequencing of cultured isolates

03 Method 2: Targeted amplicon sequencing

04 Method 3: Whole-genome amplicon sequencing

05 Support that never stops

06 Ordering information

07 References

01

Introduction

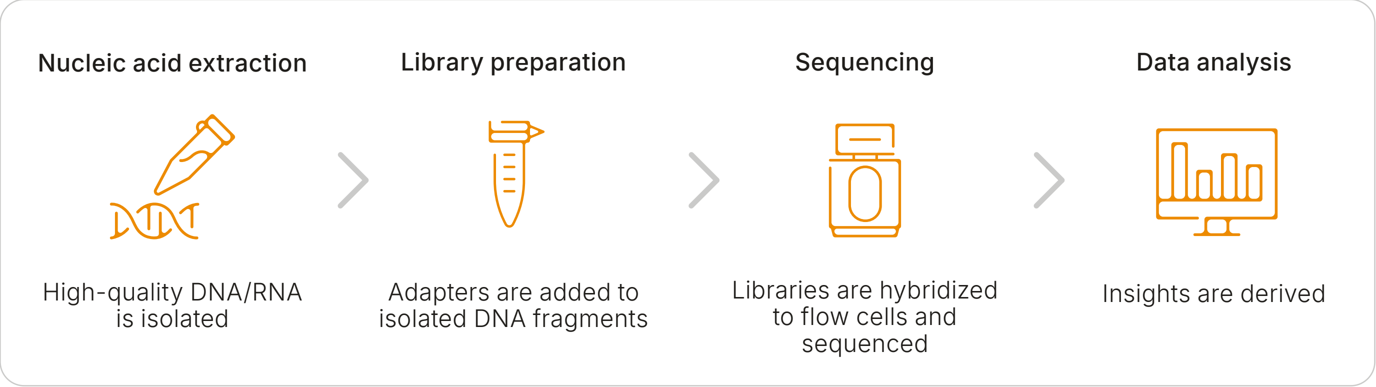
Next-generation sequencing (NGS) can rapidly sequence the genomes of microbes, allowing for a molecular-level understanding of their behavior, interactions with their hosts and the environment, evolution, and circulation within global populations. In contrast to methods like shotgun metagenomics sequencing, which enables comprehensive sequencing and identification of all microbes in a complex sample, microbial single-genome sequencing focuses on a single organism.

Whether the microbe is from an isolated, cultured bacterial colony, or a primary sample being surveilled for a determined virus of interest, microbial single-genome sequencing is best used to analyze a single, known microbe. With this approach, scientists can gain valuable insights into the microbe's evolution, virulence, and drug resistance. This information is vital for developing effective diagnostic tools, vaccines, and targeted therapies, and can aid in tracking and monitoring the spread of disease outbreaks. Use cases include tracking foodborne *E. coli*, confirming culture strains, identifying mutants in a colony, viral whole-genome sequencing (WGS) for strain typing, and more.

Overview of the microbial single-genome sequencing workflow

Depending on the microbe of interest, sample type, and data requirements, different NGS workflows and solutions can be used. Selection of the optimal workflow should be determined by the study goals. Single-genome, multipathogen, and discovery workflows can enable the full array of microbiology and infectious disease research applications. For each NGS workflow, Illumina offers powerful library preparation, sequencing systems, and corresponding data analysis apps that enable discoveries and provide insights to understand the microbe of interest better (Figure 1).

Figure 1: Microbial single-genome sequencing workflow.



STEP 1

Nucleic acid extraction

Many commercial DNA and RNA extraction and purification kits are available. Choose a kit or method that produces the highest quality nucleic acids possible for your specific targets and from your specific sample types. Illumina has tested a variety of commercially available kits, but other kits or methods work better for your targets and samples.

STEP 2

Library preparation

Following nucleic acid extraction, scientists can prepare nucleic acid libraries—a collection of similarly sized fragments that have known oligonucleotide adapter sequences attached to the 5′ and 3′ ends of the strands that are loaded onto an instrument for sequencing.

STEP 3

Sequencing

After libraries are prepared, they are hybridized to flow cells that are loaded onto a system for sequencing. Illumina sequencing systems can help you find answers using simple push-button workflows. For microbial single-genome sequencing, researchers can use a benchtop sequencing system, such as the iSeq™ 100 System, the MiSeq™ i100 Series, or NextSeq™ 1000 and NextSeq 2000 Systems (Table 1).

Table 1: Select examples of sequencing systems[a]

System	iSeq 100 System	MiSeq i100 Series	NextSeq 1000 and NextSeq 2000 Systems
System overview	Illumina sequencing in an affordable, compact system	Simplified operations and fast, flexible sequencing	Expansive application breadth and proven performance
Output range	144 Mb–1.2 Gb	1.5–30 Gb	10–540 Gb
Reads per flow cell	4M	5–100M	100M–1.8B[b]
Maximum read length	2 × 150 bp	2 × 300 bp	2 × 300 bp

a. Review all systems at illumina.com/systems/sequencing-platforms.html.
b. Maximum specifications based on a P4 flow cell run; P4 flow cells are available for the NextSeq 2000 System only.

STEP 4

Data analysis

Illumina recommends DRAGEN™ secondary analysis for highly accurate and efficient data analysis. DRAGEN secondary analysis pipelines can be accessed in-cloud or on certain Illumina sequencing systems.

Microbial single-genome sequencing data can be easily and securely transferred, stored, and analyzed in BaseSpace™ Sequence Hub, an Illumina cloud-computing software. For users without bioinformatics expertise, BaseSpace Sequence Hub offers an intuitive interface, easy run setup and monitoring, and simplified push-button secondary analysis tools.

02

Method 1: Microbial whole-genome sequencing of cultured isolates

Microbial WGS is an important tool for mapping genomes of novel organisms, finishing genomes of known organisms, or comparing genomes across multiple samples for microbial identification with nucleotide-level resolution. The high accuracy of NGS can identify low-frequency variants and genome rearrangements that may be missed or are too expensive to identify using other methods.

Relevant applications

Foodborne pathogen surveillance

Foodborne illnesses are a significant threat to public health, with nearly 600 million individuals becoming ill and 420,000 deaths each year worldwide.[1] Public health surveillance and investigation is critical to determining the source of an outbreak, so that control measures can be implemented to prevent further spread and additional illnesses. Multiple public agencies, including Health and Human Services, the Centers for Disease Control and Prevention (CDC), the US Food and Drug Administration (FDA), and others, have adopted WGS for pathogen characterization and source attribution, given its higher resolution compared to traditional molecular methods.[1]

Healthcare-associated infections (HAI) monitoring

HAI can result in substantial patient morbidity and mortality and impose a significant financial burden on hospitals and healthcare systems. The impact of HAIs is difficult to estimate globally due to the lack of adequate data from many countries. In the US alone, the annual cost of treating HAI is estimated to be $28B–45B.[2] Timely detection and identification of HAI pathogens is critical for mitigating outbreaks. Real-time surveillance using WGS provides a cost-efficient and streamlined workflow for monitoring HAI outbreaks.[3]

Key benefits:

- **Comprehensive analysis:** Detect single nucleotide polymorphisms (SNPs), insertions and deletions (indels), structural variations (SVs), and other genomic variations
- **Deep insights:** Compare the sequence of a microbe of interest against a reference genome to determine species evolution, population characteristics, and more

Step-by-step overview

Microbial WGS has four basic steps: nucleic acid extraction, library preparation, sequencing, and data analysis (Figure 2).

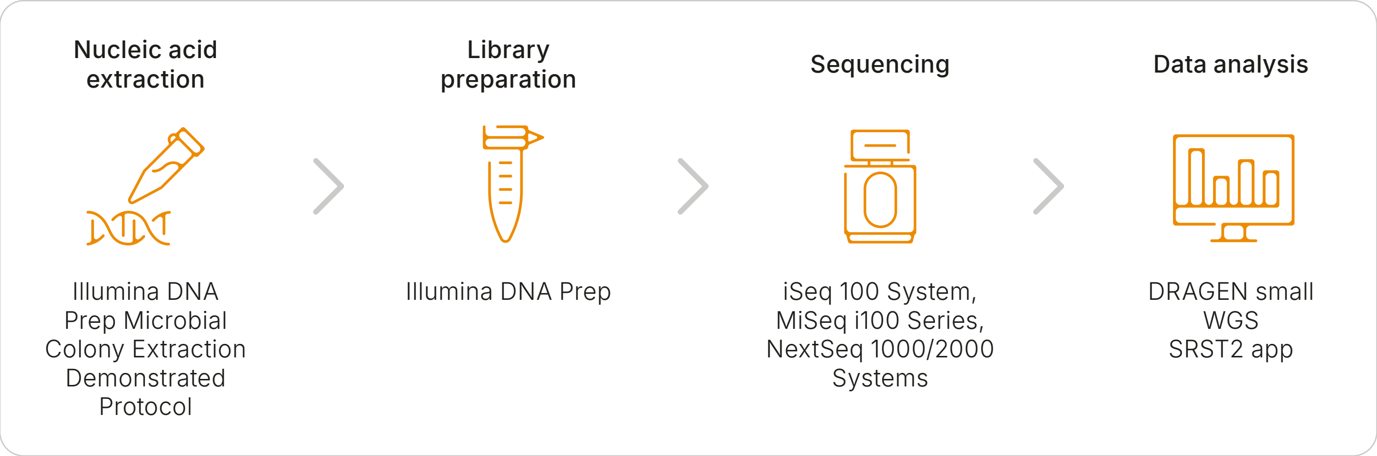


Figure 2: Microbial WGS workflow.

STEP 1

Nucleic acid extraction

For microbial WGS, it is crucial to use high-quality DNA as input. The Illumina DNA Prep Microbial Colony Extraction Demonstrated protocol is recommended.

STEP 2

Library preparation

Illumina DNA Prep delivers a fast, flexible workflow for preparing normalized, sequencing-ready libraries for microbial WGS (Table 2).

Table 2: Illumina DNA Prep specifications

Feature	Specification
DNA input amount	1–500 ng genomic DNA
Total assay time	~3–4 hr (from DNA extraction to normalized library)
Hands-on time	1–1.5 hr

STEP 3

Sequencing

Several sequencing systems can be used for microbial WGS of cultured isolates. The one chosen depends on several factors, including the study size, throughput requirements, and more (Table 3).

Table 3: Experimental parameters for performing microbial WGS on different sequencing systems[a]

System	Flow cell	Reads per flow cell	Maximum no. of samples per flow cell[b]	Recommended read length
iSeq 100 System	Reagent v2	4M	8	2 × 150 bp
MiSeq i100 Series	5M	5M	10	2 × 150 bp
	25M	25M	50	
	50M[d]	50M	100	
	100M[d]	100M	200	
NextSeq 1000 and NextSeq 2000 Systems	P1	100M	200	2 × 150 bp
	P2	400M	384[c]	

a. Experimental parameters provided in this table are a recommended starting point; optimal parameters should be determined empirically.
b. Based on 5 Mb genome size, 50×. coverage, and 2 × 150 bp read length.
c. Based on Illumina indexes, additional indexes can be added.
d. 50M and 100M flow cells are only available on the MiSeq i100 PlusSystem.

STEP 4

Data analysis

For analysis of microbial WGS data, Illumina recommends using the SRST2 or DRAGEN small whole-genome apps, available on BaseSpace Sequence Hub (Table 4).

Table 4: Illumina microbial WGS analysis software

Pipeline	Application	Input	Access point
SRST2	Reports the presence of sequence types (STs) from a MLST database and/or reference genes from a sequence database of virulence genes, resistance genes, and plasmid replicons	FASTQ	• BaseSpace Sequence Hub
DRAGEN small WGS	Maps microbial whole genomes to a reference; some genomes are preloaded onto the app, custom genomes can be added	FASTQ	• BaseSpace Sequence Hub • Onboard MiSeq i100 Series

03 Method 2: Targeted amplicon sequencing

Amplicon sequencing is a highly targeted approach that enables analysis of genetic variation in specific genomic regions. The deep sequencing of PCR products (amplicons) enables efficient variant identification and characterization. Amplicon sequencing is ideal for targeting specific regions of large genome microbial species (ie, bacteria, fungi, parasites, viruses).

Relevant applications

Tuberculosis (TB) detection and surveillance

TB remains a significant threat to public health, with almost half a million people affected with drug-resistant TB (DR-TB) worldwide.[4] Surveillance and identification of specific DR-TB strains using conventional methods are challenging, as they have limited accuracy, poor reproducibility, and lengthy turnaround times. Targeted NGS panels offer a rapid workflow for highly accurate detection and identification of DR-TB.[4]

Antimicrobial resistance (AMR) characterization

AMR is a major threat to public health, with nearly 700,000 deaths globally each year.[5] With wastewater treatment plants being known hotspots of AMR, there is a significant need for improved AMR surveillance at these sites to increase understanding of how new antimicrobial resistance genes develop and spread into indigenous microbial communities. Given its low cost and production of an easily managed data set, targeted NGS with a multiplexed amplicon approach can be a powerful tool for surveying and tracking AMR at wastewater treatment sites.[5]

Key benefits:

- **Culture-free method**— Access a streamlined method that prepares sequencing libraries directly from primary samples without labor-intensive cloning steps, saving time and simplifying the workflow
- **Species-level identification**— Sequence targeted regions of pathogen species with larger genomes (eg, *Mycobacterium tuberculosis*) for accurate identification
- **Drug resistance profiling**—Stay ahead of emerging resistance markers through resistome monitoring; gain insights into AMR development to prevent further AMR resistance

Step-by-step overview

Targeted amplicon sequencing has four basic steps: nucleic acid extraction, library preparation, sequencing, and data analysis (Figure 3).

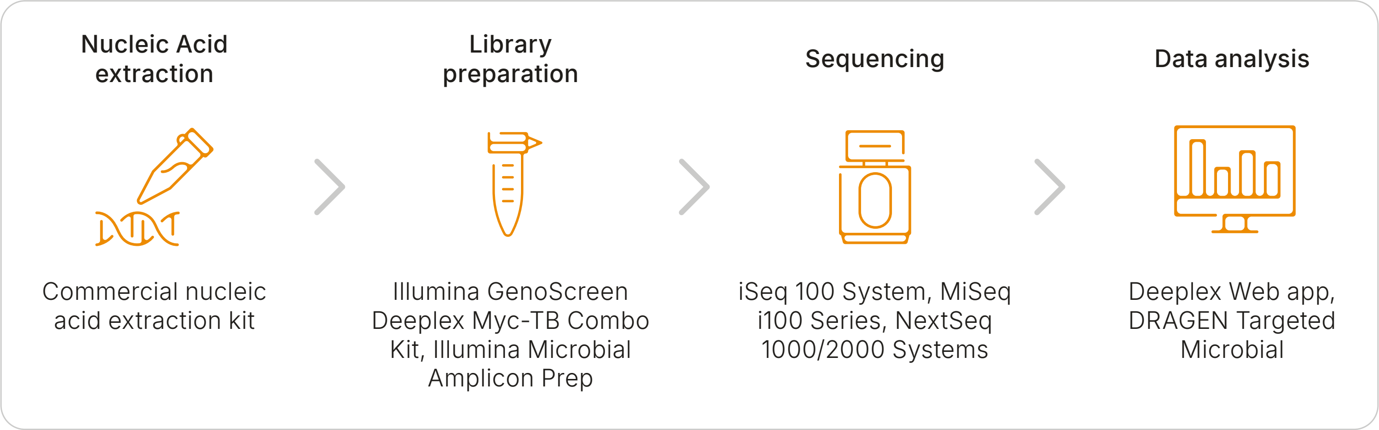


Figure 3: Targeted amplicon sequencing workflow.

STEP 1

Nucleic acid extraction

For targeted amplicon sequencing, RNA or DNA isolated from various sample types can be used as input, depending on the library preparation kit. Customers can choose from several commercially available options for nucleic acid isolation.

STEP 2

Library preparation

Illumina offers several library preparation kits for targeted amplicon sequencing (Table 5). Illumina Microbial Amplicon Prep can be used with custom, published, or commercially available primer sets (primer oligos are not included) and is compatible with RNA or DNA extracted from various sources. It can be used for various infectious disease, public health surveillance, and microbial research applications, including AMR marker analysis, bacterial and fungal identification. The Illumina and GenoScreen Deeplex® Myc-TB Combo Kit identifies *Mycobacterium tuberculosis* complex (MTBC) strains, mycobacterial species identification, genotyping, and prediction of drug resistance mutations of MTBC strains.

Table 5: Selected library prep kit specifications

Parameter	Illumina Microbial Amplicon Prep	Illumina and Genoscreen Deeplex Myc-TB Combo Kit
DNA input amount	Varies depending on sample source	9 pg DNA
Total assay time	< 9 hr	~15 hr from sample to prepared library
Hands-on time	~3 hr for 48 samples	< 4 hr

STEP 3

Sequencing

The sequencing system used depends on several factors, such as study size, throughput requirements, and more (Table 6).

Table 6: Experimental parameters for Illumina Microbial Amplicon Prep on different platforms[a]

System	Flow cell	Reads per flow cell	Maximum no. of samples per flow cell[b]	Recommended read length
iSeq 100 System	Reagent v2	4M	4	2 × 150 bp
MiSeq i100 Series	5M	5M	5	2 × 150 bp
	25M	25M	25	
	50M[c]	50M	50	
	100M[c]	100M	100	
NextSeq 1000 and NextSeq 2000 Systems	P1	100M	100	2 × 150 bp
	P2	400M	384[d]	

[a] Experimental parameters provided in this table are a recommended starting point, and optimal parameters should be determined empirically.

[b] Based on 1M reads per sample.

[c] 50M and 100M flow cells are only available on the MiSeq i100 Plus System.

[d] Based on Illumina indexes. Additional indexes can be added.

STEP 4

Data analysis

Illumina recommends using the Deeplex Web App and DRAGEN Microbial Amplicon pipeline for analysis of sequencing data generated from Illumina and Genoscreen Deeplex Myc-TB Combo Kit and Illumina Microbial Amplicon Prep libraries, respectively (Table 7). The Deeplex Web App is available on the cloud-based Illumina Connected Analytics. DRAGEN Microbial Amplicon is available on the cloud-based BaseSpace Sequence Hub.

Table 7: Illumina targeted amplicon sequencing analysis software

Pipeline	Application	Input	Access point
Deeplex Web App	Performs mycobacterial species identification and drug resistance prediction with interactive results visualization and reporting	FASTQ	• Illumina Connected Analytics
DRAGEN Microbial Amplicon pipeline	Generates consensus sequences from tiled amplicon panels and provides a best estimate of the population of targeted sequences present	FASTQ	• BaseSpace Sequence Hub

04 Method 3: Whole-genome amplicon sequencing

WGS via an amplicon approach is ideal for known viruses with small genomes. This method involves analyzing genomic regions of interest with ultradeep sequencing of PCR amplicons. It enables researchers to detect and fully characterize a known virus.

Relevant applications

Variant identification

Due to its rapid spread and potential for debilitating epidemics, chikungunya virus (CHIKV) is a serious public health concern. Surveillance of CHIKV and its evolutionary trajectory using WGS in Argentina found significant patterns of virus dissemination, highlighting key areas of transmission and potential factors contributing to its spread.[6]

Wastewater surveillance for variant tracking

Detecting viral RNA in wastewater offers a way to track infections at a community scale. Wastewater data can also predict surges of infections up to two weeks before clinical cases present in the healthcare system.[1] Early warning from wastewater data can prompt local action to prevent the spread of disease and prepare hospitals for imminent surges.[7] A study in Utah used the Illumina COVIDSeq™ Assay to monitor SARS-CoV-2 lineages across various wastewater collection sites. Importantly, sequencing results identified two Omicron lineages among the different variants detected.[8]

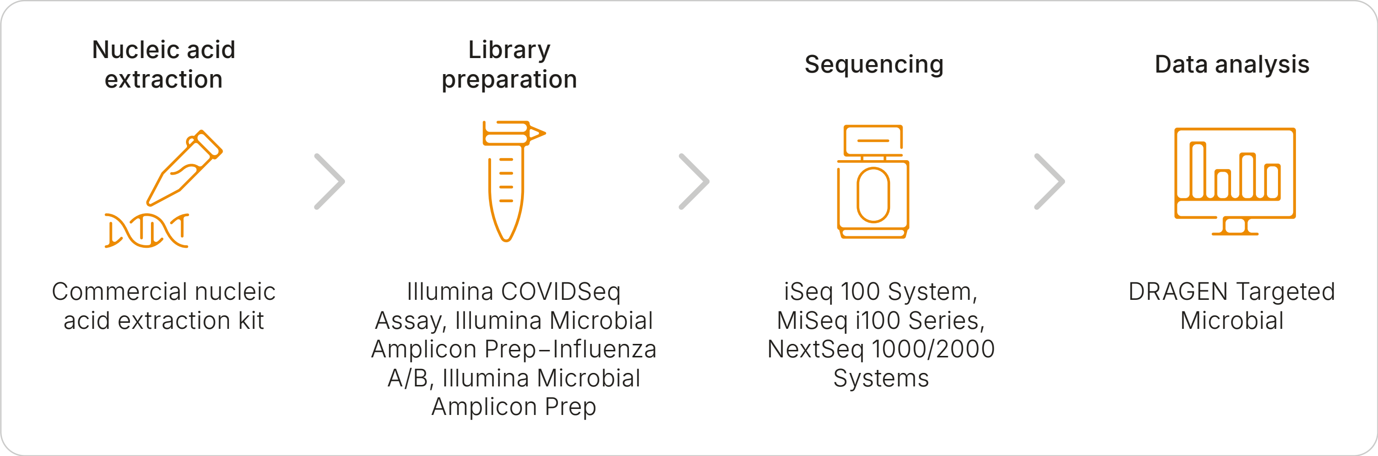
Key benefits:

- **Direct from primary samples**—Access a streamlined, culture-free method without labor-intensive cloning steps, saving time and simplifying the workflow
- **Whole-genome characterization**—Generate rapid and accurate genome sequence information that is critical for detecting low-frequency mutations, finding key deletions and insertions, and discovering other genetic changes

Step-by-step overview

Whole-genome amplicon sequencing has four basic steps: nucleic acid extraction, library preparation, sequencing, and data analysis (Figure 4).

Figure 4: Whole-genome amplicon sequencing workflow.



STEP 1

Nucleic acid extraction

Nucleic acid isolated from various sample types can be used as input, depending on the library preparation kit. Customers can choose from several commercially available options for nucleic acid isolation.

STEP 2

Library preparation

Illumina offers several library preparation kits for whole-genome amplicon sequencing (Table 8). The Illumina COVIDSeq Assay (96 samples) provides uniform viral genome coverage for identification and characterization of novel SARS-CoV-2 variants. Illumina Microbial Amplicon Prep–Influenza A/B enables WGS of influenza A/B viruses for public health surveillance and influenza virus research. As stated previously, Illumina Microbial Amplicon Prep offers customizable library prep for various microbial research applications. Illumina has tested protocols for Chikungunya, Dengue, Mpox, RSV, and Zika viruses, and customers have demonstrated compatibility with many other viruses.

Table 8: Selected library prep kit specifications

Library prep kit	Illumina COVIDSeq Assay (96 samples)	Illumina Microbial Amplicon Prep–Influenza A/B	Illumina Microbial Amplicon Prep
Input type	RNA	RNA	DNA, RNA
Total assay time	< 9 hr	< 9 hr	< 9 hr
Hands-on time	< 1 hr with automation	~ 3 hr for 48 samples	~ 3 hr for 48 samples

STEP 3

Sequencing

The sequencing system used depends on several factors, including study size, throughput requirements, and more (Table 9).

Table 9: Experimental parameters for performing whole-genome amplicon sequencing on different platforms[a]

System	Flow cell	Reads per flow cell	Sample no. per flow cell[a]	Recommended read length
iSeq 100 System	Reagent v2	4M	8	2 × 150 bp
MiSeq i100 Series	5M	5M	10	2 × 150 bp
	25M	25M	50	
	50M[c]	50M	100	
	100M[c]	100M	200	
NextSeq 1000 and NextSeq 2000 Systems	P1	100M	200	2 × 150 bp
	P2	400M	384[d]	

a. Experimental parameters shown are for sequencing libraries prepared with the Illumina COVIDSeq Assay (96 samples) with a maximum of 384 indexes. Parameters may vary for other libraries.
b. Based on 500K reads per sample.
c. The 50M and 100M flow cells are only available on the MiSeq i100 Plus System.
d. Based on Illumina indexes. Additional indexes can be added.

STEP 4

Data analysis

Illumina recommends using the DRAGEN Microbial Amplicon pipeline for analysis of sequencing data (Table 10). DRAGEN Microbial Amplicon is available on the cloud-based BaseSpace Sequence Hub.

Table 10: Illumina targeted amplicon sequencing analysis software

Pipeline	Application	Input	Access point
DRAGEN Microbial Amplicon pipeline	Generates consensus sequences from tiled amplicon panels and provides a best estimate of the population of targeted sequences present	FASTQ	• BaseSpace Sequence Hub

05 Support that never stops

Illumina strives to be the best partner possible. With a global presence, you can rely on our support to facilitate your success. Technical support is available via phone five days a week or access online support 24/7, worldwide and in multiple languages, with rapid response time near most major metropolitan areas. Illumina provides excellent product consistency, supply, and quality enabled by a mature global manufacturing infrastructure.

Trusted technology, trusted partner

As a preferred NGS platform provider, Illumina has shipped over 20,000 sequencing systems globally. Illumina NGS technology is cited in over 421,000 peer-reviewed publications—5× more than all other NGS technologies combined.[9] Building on decades of expertise, Illumina has a relentless commitment to innovation and building future NGS capabilities and applications.

Learn more

Illumina microbial single-genome sequencing

06 Ordering information

Library preparation

Product	Catalog no.
Illumina DNA Prep, (M) Tagmentation (24 samples, IPB)	20060060
Illumina DNA Prep, (M) Tagmentation (96 samples, IPB)	20060059
Illumina and Genoscreen Deeplex Myc-TB Combo IDP Kit (48 samples)	20099512
Illumina and Genoscreen Deeplex Myc-TB Combo XT Kit (48 samples)	20090205
Illumina Microbial Amplicon Prep (48 samples)	20097857
Illumina Microbial Amplicon Prep–Influenza A/B (48 samples)	20106305
Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 Indexes, 96 Samples)[a]	20091654
Illumina DNA/RNA UD Indexes Set B, Tagmentation (96 Indexes, 96 Samples)[a]	20091656
Illumina DNA/RNA UD Indexes Set C, Tagmentation (96 Indexes, 96 Samples)[a]	20091658
Illumina DNA/RNA UD Indexes Set D, Tagmentation (96 Indexes, 96 Samples)[a]	20091660

a. 384 indexes total available when combining sets A, B, C, and D.

Sequencing systems

System	Catalog no.
iSeq 100 System	20021532
MiSeq i100 System	Coming soon
MiSeq i100 Plus System	20115695
NextSeq 1000 System	20038898
NextSeq 2000 System	20038897

Data analysis

Product	Catalog no.
BaseSpace Sequence Hub Professional Annual Subscription[a]	20042109
BaseSpace Sequence Hub Enterprise Annual Subscription[a]	15066411
Illumina Connected Analytics Professional Annual Subscription	20044876
Illumina Connected Analytics Enterprise Annual Subscription	20038994

a. BaseSpace Sequence Hub subscriptions include complimentary iCredits for running analysis apps and data storage. Additional iCredits are available for purchase.

07 Appendix

References

1. World Health Organization. Estimating the burden of foodborne diseases. who.int/activities/estimating-the-burden-of-foodborne-diseases. Accessed August 11, 2024.
2. Gidey K, Gidey MT, Hailu BY, Gebreamlak ZB, Niriayo YL. Clinical and economic burden of healthcare-associated infections: A prospective cohort study. *PLoS One*. 2023;18(2):e0282141. Published 2023 Feb 23. doi:10.1371/journal.pone.0282141
3. Waggle KD, Griffith MP, Rokes AB, et al. Methods for cost-efficient, whole genome sequencing surveillance for enhanced detection of outbreaks in a hospital setting. *medRxiv*. 2024;02.16.24302955. doi: https://doi.org/10.1101/2024.02.16.24302955
4. Mansoor H, Hirani N, Chavan V, et al. Clinical utility of target-based next-generation sequencing for drug-resistant TB. *Int J Tuberc Lung Dis*. 2023;27(1):41-48. doi:10.5588/ijtld.22.0138
5. Gibson C, Kraemer SA, Klimova N, Guo B, Frigon D. Antibiotic resistance gene sequencing is necessary to reveal the complex dynamics of immigration from sewers to activated sludge. *Front Microbiol*. 2023;14:1155956. Published 2023 Apr 26. doi:10.3389/fmicb.2023.1155956
6. Fabbri C, Giovanetti M, Luppo V, et al. Tracing the evolution of the chikungunya virus in Argentina, 2016-2023: independent introductions and prominence of Latin American lineages. *Emerg Microbes Infect*. 2024;13(1):2362941. doi:10.1080/22221751.2024.2362941
7. Iwamoto R, Yamaguchi K, Katayama K, et al. Identification of SARS-CoV-2 variants in wastewater using targeted amplicon sequencing during a low COVID-19 prevalence period in Japan. *Sci Total Environ*. 2023;887:163706. doi:10.1016/j.scitotenv.2023.163706
8. Gupta P, Liao S, Ezekiel M, et al. Wastewater Genomic Surveillance Captures Early Detection of Omicron in Utah. *Microbiol Spectr*. 2023;11(3):e0039123. doi:10.1128/spectrum.00391-23
9. Data calculations on file. Illumina, Inc. 2022.

illumina

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-02702 v1.0