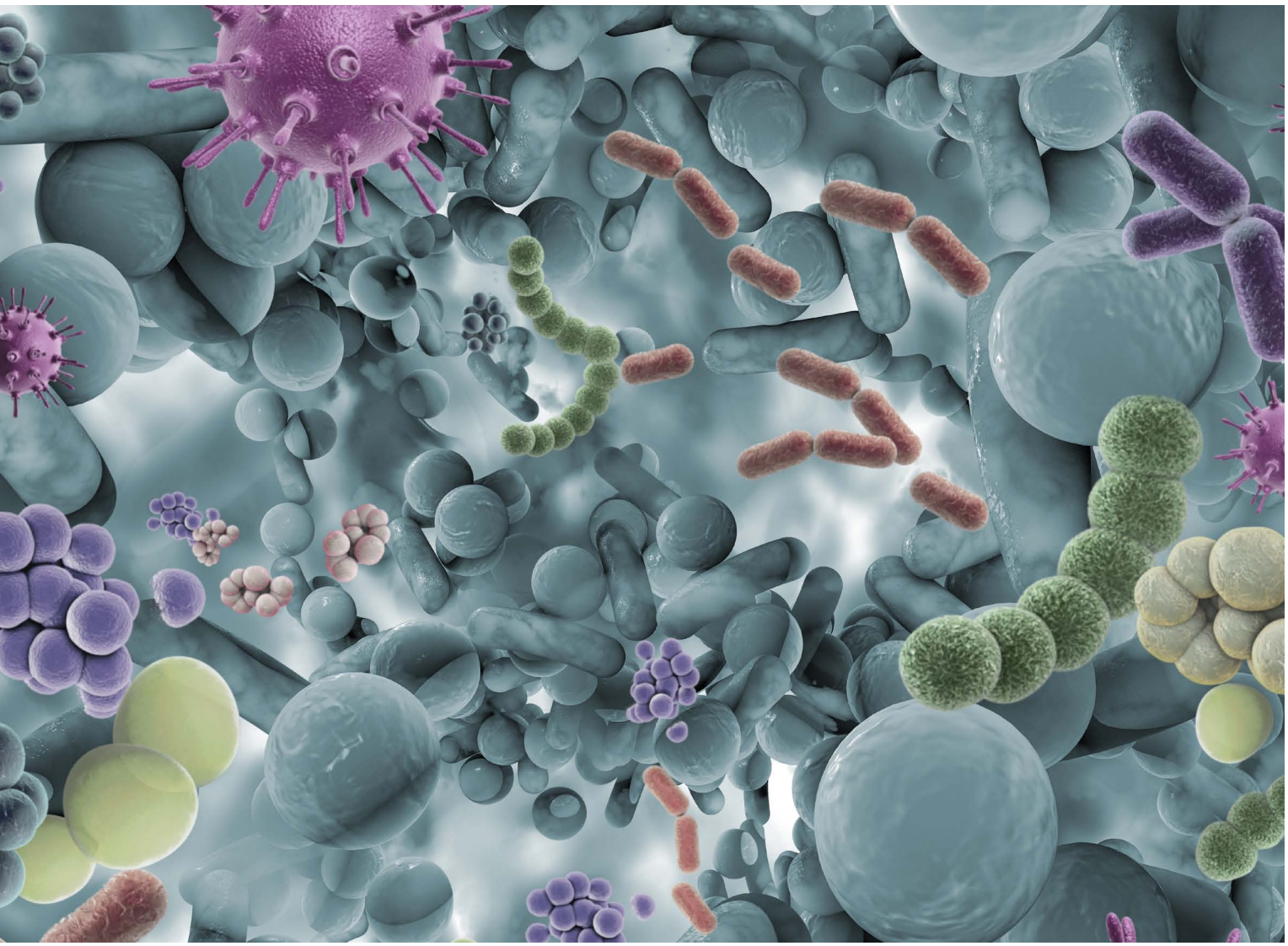
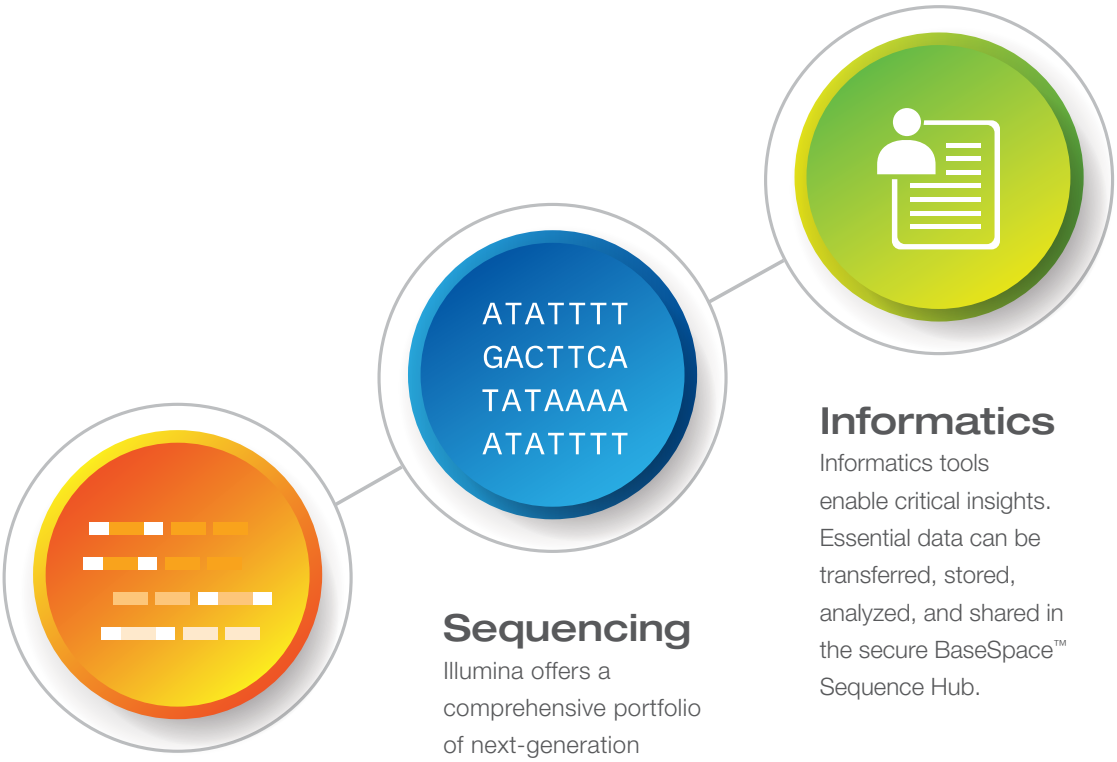
illumina®

Methods Guide for Microbiology





Informatics

Informatics tools enable critical insights. Essential data can be transferred, stored, analyzed, and shared in the secure BaseSpace™ Sequence Hub.

Sequencing

Illumina offers a comprehensive portfolio of next-generation sequencing (NGS) instruments that are accessible and scalable for every lab.

Library prep

Illumina library preparation kits enable a broad range of microbiology-specific methods and applications.

For Research Use Only. Not for use in diagnostic procedures.

INTRODUCTION

Next-generation sequencing for microbiologists

Overview of microbiology workflows

METHODS

Whole-genome sequencing

Targeted resequencing

AmpliSeq™ for Illumina Panels

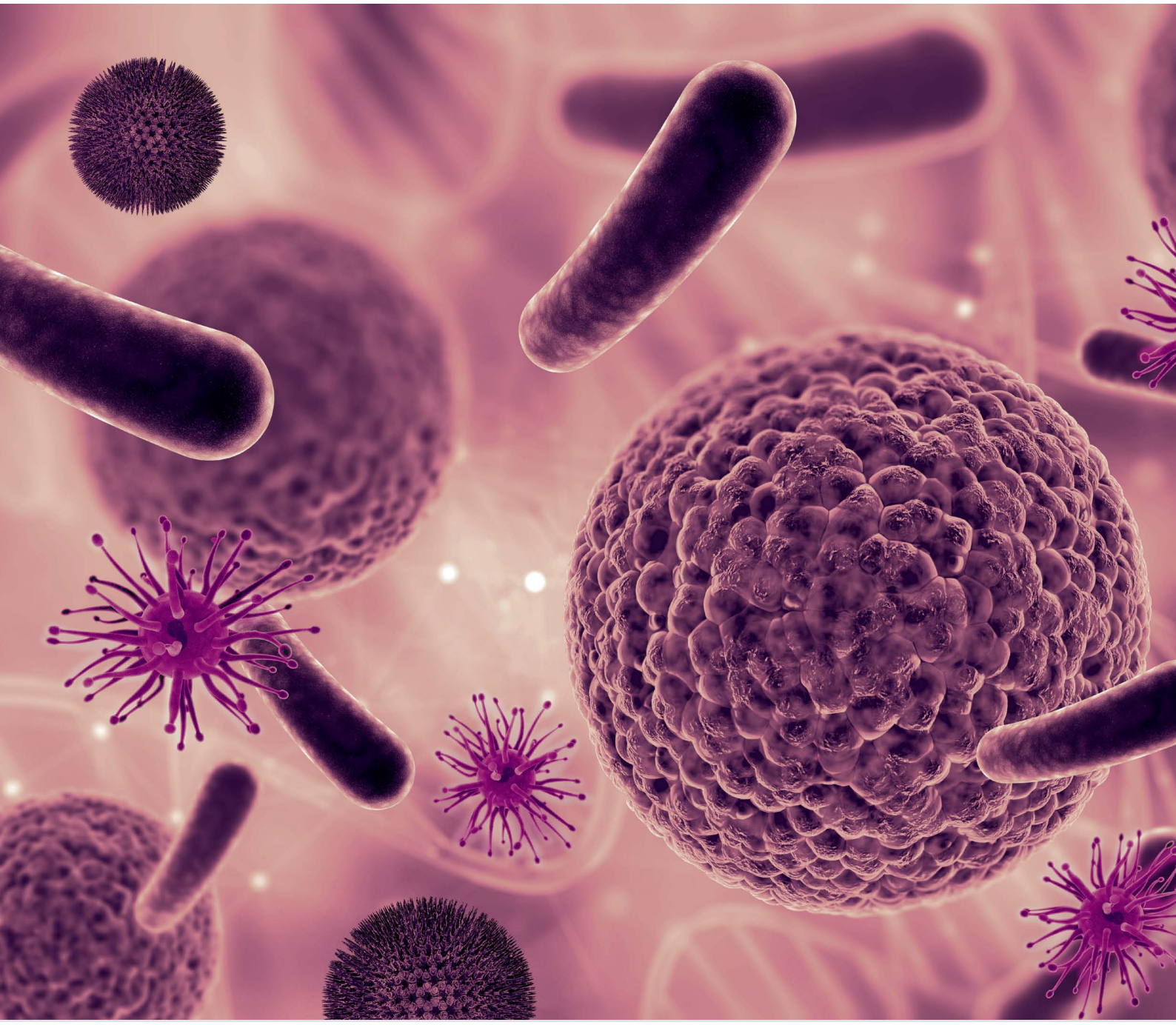
DATA ANALYSIS

Research informatics

SERVICES

Instrument services, training, and consulting

Introduction



Illumina next-generation sequencing for microbiologists

Take your research to the next level

The introduction of next-generation sequencing (NGS) has transformed the way scientists study biological and microbiological systems. With clear benefits such as reduced time and cost compared to legacy technologies, as well as the capacity to scale from small studies to population-level throughput, NGS opens the door to a broad range of research capabilities.

With NGS, deep sequencing provides the sensitivity needed to detect low abunance members of complex populations. NGS can rapidly advance microbial identification studies, including those based on complex microbial communities, such as human microbiome or environmental surveillance studies. Furthermore, studies performed with NGS do not rely on *a priori* knowledge of the genomes in the system. High-resolution NGS data (eg, nucleotide level sequence data) can be instrumental for examining pathogenesis, horizontal gene transfer, pangenomes, and the co-evolution of hosts and symbionts/parasites. NGS delivers a wealth of information for mutational studies of all kinds including directed evolution strategies, phylogenetic analysis, or the study of temporal and spatial dynamics of disease transmission.

This accumulation of sequence information has greatly expanded our appreciation of the dynamic nature of microbial populations and their impact on human, animal, and environmental health. With the advantages of speed, sensitivity, and scale, NGS can take your microbiology and metagenomics research to the next level.

Overview of microbiology workflows

Simple, comprehensive workflows for a broad range of microbiology applications

Library prep and accessories

Illumina library preparation kits provide all the reagents and enzymes needed to convert samples of interest into ready-to-sequence libraries. Choose from application specific kits including small whole-genome sequencing, targeted, or custom panel sequencing.

Sequencing

Choose a sequencer to match your needs from the low-throughput iSeq™ 100 System to the production-scale NovaSeq™ 6000 System.

Data analysis

Manage, analyze, and share data in the cloud securely with BaseSpace Sequence Hub. Data is automatically converted to standard file formats (FASTQ) for analysis using a broad suite of microbiology applications (Apps).

Step 1. Library preparation

Illumina offers a comprehensive NGS sequencing workflow, from library preparation to final data analysis. Library preparation kits are available for a range of NGS methods including small whole-genome sequencing (eg, bacteria, archaea, viruses, plasmids) targeted panel sequencing, 16S and ITS rRNA metagenomics, and shotgun metagenomics.

Whole-genome sequencing (from isolates)	Shotgun metagenomic sequencing
Nextera™ DNA Flex Library Prep Kit	Nextera DNA Flex Library Prep Kit
Nextera XT DNA Library Prep Kit	TruSeq™ DNA PCR-Free Library Prep Kit
Targeted metagenomic sequencing	**Targeted sequencing panels**
16S rRNA Library Prep Protocol (Demonstrated Protocol)	AmpliSeq for Illumina Antimicrobial Resistance Research Panel
ITS rRNA Library Prep Protocol (Demonstrated Protocol)	AmpliSeq for Illumina TB Research Panel

Step 2. Sequencing

Illumina offers a full portfolio of sequencing platforms, from the benchtop iSeq 100 System to the production-scale NovaSeq 6000 System, that deliver the right level of speed, capacity, and cost for various laboratory or sequencing center requirements. Illumina has pioneered major advances in sequencing simplicity, flexibility, and platform performance. Experiments that once required complex workflows now use simple push-button workflows.

iSeq 100 System | MiniSeq System | MiSeq System | NextSeq 550 System | NovaSeq 6000 System

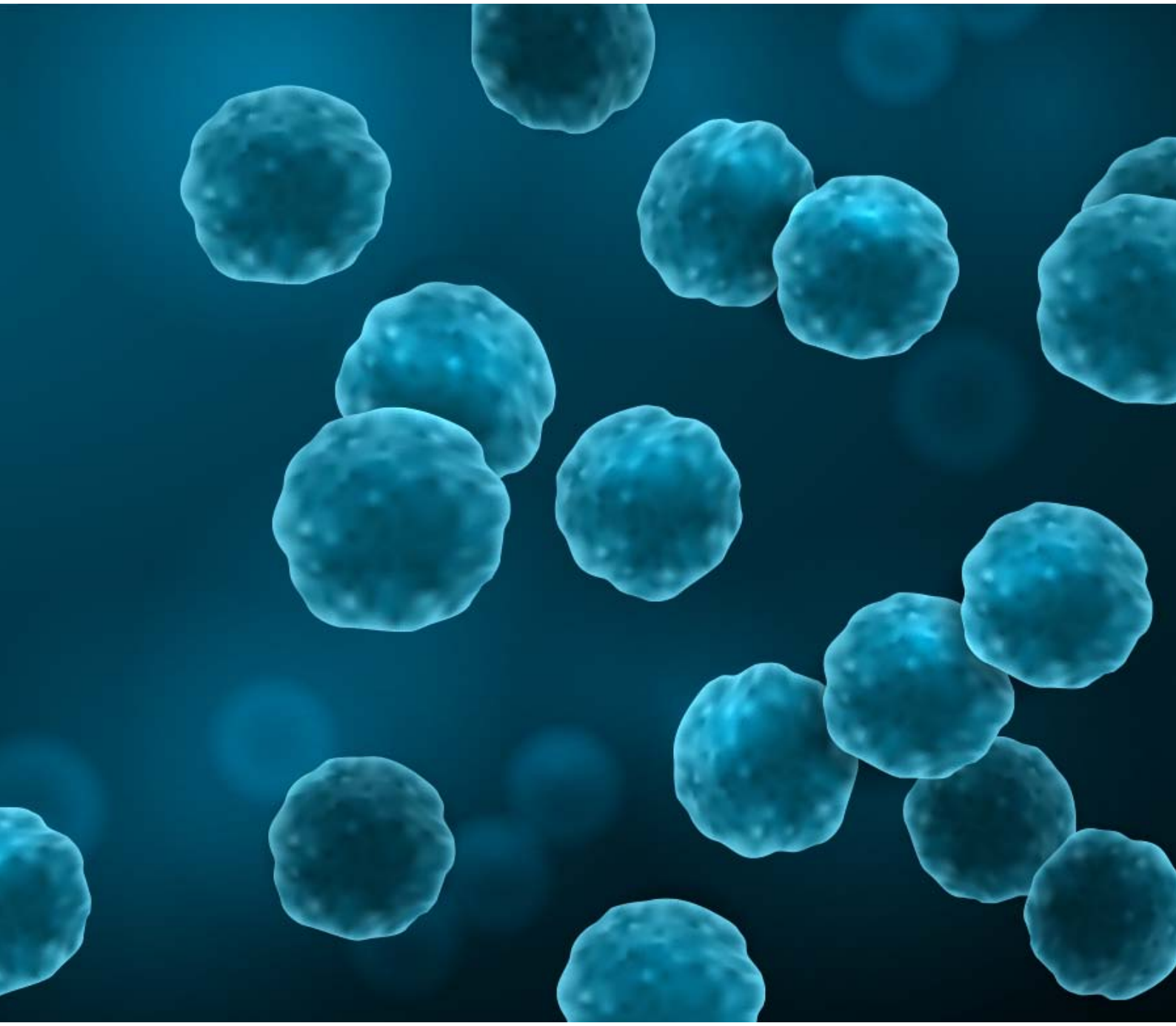
Step 3. Data Analysis

Illumina microbiology workflows provide user-friendly data analysis tools that are easily accessible through the web with BaseSpace Sequence Hub, on instrument, or through onsite servers. BaseSpace Sequence Hub, the Illumina online genomics computing and data storage platform, is integrated directly with Illumina sequencing systems. On-instrument access to BaseSpace enables integration of many workflow steps including library prep planning and sample management, run set-up and chemistry validation, and real-time data monitoring and data transfer to computing and analysis modules. Data analysis proceeds seemlessly with BaseSpace Apps, which offer a wide variety of analysis pipelines from Illumina as well as 3rd party developers.

www.illumina.com/systems/sequencing-platforms.html
www.illumina.com/basespace

For Research Use Only. Not for use in diagnostic procedures.

Methods



Sequencing and assembly of genomes from isolates

Sequence virtually any small genome and make new discoveries with scalable throughput

Key benefits

- Allows comprehensive analysis of microbial genomes to better understand the variations both within and between species (eg, food safety and healthcare associated infections)
- Differentiates between organisms with nucleotide-level precision
- Provides comprehensive analysis of 20 Mb or smaller genomes
- Generate accurate reference genomes

Prepare library	Sequence	Analyze and Report
		
Nextera DNA Flex Library Prep Kit Nextera XT DNA Library Prep Kit	**iSeq™ 100 System** **MiniSeq™ System** **MiSeq™ System** **NextSeq™ 550 System**	**BaseSpace Sequence Hub** BWA Aligner, Pisces Variant Caller, SRST2, SPAdes, Bacterial Analysis Pipeline, Velvet *de novo* Assembly, and Prokka Genome Annotation
	Illumina Informatics Solutions Local Run Manager, BaseSpace Sequence Hub Run management | Analysis | Reporting	

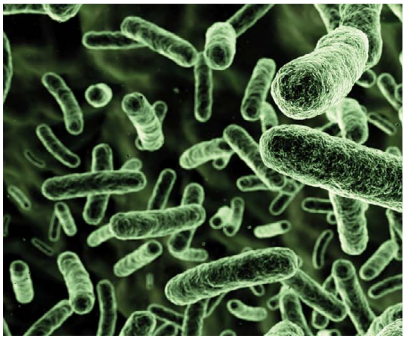
Training	Service contracts	Professional services
Customer site training Nextera DNA Flex Library Prep Kit Nextera XT DNA Library Prep Kit	**Tiered service plans** iSeq 100 System MiniSeq System MiSeq System NextSeq 550 System	**Proof-of-Concept Service** Standard application functional testing with your samples

www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods/microbial-whole-genome-sequencing.html

For Research Use Only. Not for use in diagnostic procedures.

Innovation. Discovery. Application.

Bacterial discovery and evolution (public health and epidemiology)
Bacterial species change over time through various mechanisms, such as horizontal gene transfer and coevolution with other organisms. High-resolution gene sequencing can inform a wide range of studies, including mutagenesis, directed evolution, the spatial and temporal dynamics of epidemics, and mechanisms of disease transmission.



Pathogen surveillance (public health and epidemiology)
Microbiologists often need to screen human, animal, environmental, and agricultural samples to identify and confirm the presence of specific microbes. Researchers use WGS to identify the bacterial genomes of pathogens, increasing the accuracy of microbial testing. WGS can be used as a valuable tool in investigating outbreaks and tracing the cause back to the source.

Risk monitoring (biopharmaceutical development)
Risk monitoring is a key aspect of biopharmaceutical research and product development. WGS allows researchers to closely track cell line stability and conduct biosafety monitoring.

STEP 1
Recommended library prep methods


Product	Nextera DNA Flex Library Prep Kit	Nextera XT DNA Library Prep Kit
Most important to me	Broad DNA input range Fewest number of workflow steps High uniformity of coverage	Low input for small genomes and large insert size
Input quantity	1–500 ng DNA	1 ng DNA
Multiplexing	384 samples	384 samples
Assay time[a]	3–4 hrs	5 hrs
Target insert size	~350 bp	300 bp–1.5 kb
Fragmentation method	Enzymatic	Enzymatic

a. Denotes total workflow time, including DNA extraction, library preparation, and library normalization/pooling steps

STEP 2
Recommended sequencers




Product	iSeq 100 System	MiniSeq System	MiSeq System	NextSeq 550 System
Most important to me	Affordability and efficiency	Simplicity and instrument affordability	Longest read lengths in Illumina portfolio, well-suited for microbiology applications	Capacity for whole genome metagenomics and production power for small genome sequencing applications
Bacterial genome samples processed/flow cell [a,b]	1–4	1–18	1–24	1–384
Recommended read length for WGS (isolates)	2 × 150 bp	2 × 150 bp	2 × 250 bp	2 × 150 bp
Run mode/kit type	Standard	Mid-output/ High-output	Research mode and Dx	Mid-output/ High-output
Max total output per flow cell	1.2 Gb	2.4 Gb (Mid-output) 7.5 Gb (High-output)	0.5 Gb (Nano v2) 1.2 Gb (Micro v2) 7.5 Gb (Standard v2) 15 Gb (Standard v3)	39 Gb (Mid-output v2.5) 120 Gb (High-output v2.5)
Flow cells processed/run	1	1	1	1

a. Sample numbers may depend on genome size, read depth, and specific flow cell output.
b. Calculations based on 5 Mb genome and 1 million reads.

STEP 3
Data analysis

Analyze whole-genome data with Local Run Manager Software or BaseSpace Sequence Hub Apps

Pathogen detection

BWA* Aligner B and Pisces Variant Caller B
- These 2 apps perform read alignment and mutation detection for any genome with a known reference sequence. If none is available, the pipeline can still be used to analyze bacteria, fungi, or viruses after *de novo* assembly

*BWA: Burrows-Wheeler Aligner

SRST2 B
- Reports the presence of STs (sequence types) from a MLST database and/or reference genes from a database of sequences for virulence genes, resistance genes, and plasmid replicons

Bacterial Analysis Pipeline B
- Identifies species, multilocus sequence type, plasmids, virulence - and antimicrobial resistance genes in bacteria

Isolate sequencing and assembly

Assembly workflow L
- Assembles small genomes (< 20 Mb) and best suits the assembly of bacterial genomes, such as *E. coli*
- Uses the EMBL Velvet algorithm and writes assembly results in FASTA format

SPAdes Genome Assembler B
- Assembles genomes from standard bacterial isolates and single-cell multiple displacement amplification (MDA) preparations

Gene annotation

Prokka Genome Annotation B
- Rapidly annotates genes and identifies coding sequences in prokaryotic genomes
- Does not annotate eukaryotic genomes

Commercial 3rd party apps

bioMérieux EPISEQ
- Comprehensive, integrated applications for microbiological research and routine applications. Bionumerics combines information from various genomic and phenotypic sources into one global database and conducts combined analyses
www.applied-maths.com/applications

ChunLab TrueBac ID
- A cloud based system that incorporates the “gold standard” of bacterial identification by using whole genome sequences. TrueBac ID can identify >12,000 species for microbial applications in research, applied and industrial settings
www.truebacid.com

L Local Run Manager Software B BaseSpace Sequence Hub

www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods/microbial-whole-genome-sequencing.html

Library prep ordering information

Nextera DNA Flex Library Prep Kits

Save time and reduce hands-on touch points with On-Bead Tagmentation, producing sequencing-ready libraries in less than three hours. Simplify daily operations with a kit that supports a broad DNA input range (1–500 ng), multiple DNA input types, and flexible for use with small (bacteria, archaea, viruses, plasmids and other microbial isolates) to large genomes (human, plant, mouse). Nextera DNA Flex Library Prep Kits deliver consistent insert sizes, uniform coverage, and optimized performance, regardless of DNA input amount or genome size. On-bead tagmentation minimizes bias and opportunities for error, resulting in highly robust data without pre-library prep DNA quantification steps, or post-library prep quantification steps such as library normalization and pooling. Nextera DNA Flex libraries are compatible with all Illumina sequencers.

Ordering information

Product	No. of indexes	No. of samples	Catalog no.
Nextera DNA Flex Library Prep Kit	-	24	20018704
Nextera DNA Flex Library Prep Kit	-	96	20018705
Flex Lysis Reagent Kit	-	-	20018706
Nextera DNA CD Indexes	24	24	20018707
Nextera DNA CD Indexes	96	96	20018708
IDT for Illumina Nextera DNA UD Indexes Set A	96	96	20027213
IDT for Illumina Nextera DNA UD Indexes Set B	96	96	20027214
IDT for Illumina Nextera DNA UD Indexes Set C	96	96	20027215
IDT for Illumina Nextera DNA UD Indexes Set D	96	96	20027216
Training			
Nextera DNA Flex Library Prep Kit training at customer site	-	-	20022900

Nextera XT DNA Library Prep Kits

Generate sequencing-ready libraries from small genomes (bacteria, archaea, viruses), amplicons, and plasmids. Nextera XT DNA Library Prep Kits simultaneously fragment input DNA and tag the fragments with sequencing adapters in a single-tube enzymatic reaction. Nextera XT DNA Library Prep Kits require as little as 1 ng of input, supporting a wide array of sample types. Bead-based normalization eliminates the need for library quantification before pooling and sequencing. Nextera XT libraries are compatible with all Illumina sequencing systems.

Ordering information

Product	No. of indexes	No. of samples	Catalog no.
Nextera XT DNA Library Prep Kit	-	24	FC-131-1024
Nextera XT DNA Library Prep Kit	-	96	FC-131-1096
Nextera XT Index Kit	24	96	FC-131-1001
TruSeq Dual Index Sequencing Primer Kit, single read	Single-use kit	-	FC-121-1003
TruSeq Dual Index Sequencing Primer Kit, paired-end read	Single-use kit	-	PE-121-1003
Nextera XT Index Kit v2, set A	96	384	FC-131-2001
Nextera XT Index Kit v2, set B	96	384	FC-131-2002
Nextera XT Index Kit v2, set C	96	384	FC-131-2003
Nextera XT Index Kit v2, set D	96	384	FC-131-2004
IDT for Illumina Nextera DNA UD Indexes Set A	96	96	20027213
IDT for Illumina Nextera DNA UD Indexes Set B	96	96	20027214
IDT for Illumina Nextera DNA UD Indexes Set C	96	96	20027215
IDT for Illumina Nextera DNA UD Indexes Set D	96	96	20027216
Training			
Nextera XT DNA Library Prep Kit training at customer site	-	-	TR-204-0009

Proof-of-Concept Service

The POC Service will run a subset of your samples with your desired system and application in a real-world demonstration of how Illumina NGS can empower your laboratory. Receive in-person consultation pre-POC to establish POC needs and goals, POC sequencing services of your samples at the Illumina Service Lab, and in-person consultation post-POC to review sequencing run QC, data, and reports based on your samples. Contact your local Illumina representative for more information.

Ordering information

Product	Catalog no.
iSeq 100 POC Service	20023613
MiniSeq POC Service	20003924
MiSeq POC Service	SP-801-1002
NextSeq POC Service	SP-801-1003
NovaSeq POC Service	20016091

Shotgun metagenomics

Culture-free sequencing of complex microbial samples to identify emerging pathogens, classify microbial species, and analyze functional characteristics of microbial communities

Key benefits

- Enables culture-free identification and classification of complex microbial communities
- Accelerates discovery of new markers and functional genes that translate to virulence or antibiotic resistance
- Promotes accurate, species-level evaluation of microbial biodiversity and abundance
- *De novo* discovery and characterization of novel species and assembly of new genomes
- Sequence multiple genomes in parallel

Prepare library	Sequence	Analyze and Report
		
Nextera DNA Flex Library Prep Kit **TruSeq DNA PCR-Free Library Prep Kit**	**NextSeq 550 System** **NovaSeq 6000 System**	**BaseSpace Sequence Hub** Kraken2 Metagenomics CosmosID Metagenomics SPAdes Genome Assembler MetaPhlan
	Illumina Informatics Solutions BaseSpace Sequence Hub Run management | Analysis | Reporting	

Training	Service contracts	Professional services
Customer site training Nextera DNA Flex Library Prep Kit TruSeq DNA PCR-Free Library Prep Kit	**Tiered service Plans** NextSeq 550 System NovaSeq 6000 System	**Proof of Concept Service** Standard application functional testing with your samples

www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods/shotgun-metagenomic-sequencing.html

For Research Use Only. Not for use in diagnostic procedures.

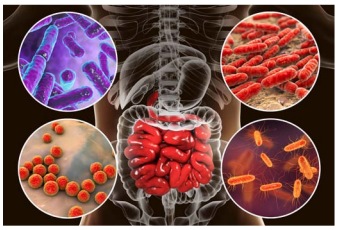
Innovation. Discovery. Application.

Microbiome (agrigenomics)

Bacteria, viruses, and fungi all affect agriculturally important plant and animal species. Metagenomics allows researchers and farmers to manage the health of their herds and plant crops to maximize food production and quality.

Microbiome (environmental studies)

Complex environmental samples often include organisms that cannot be cultured in the laboratory. With shotgun metagenomics sequencing, researchers evaluate bacterial, fungal, and viral diversity and detect the abundance of microbes in various environments, including unculturable microorganisms. In addition to gene identification, shotgun metagenomics provides gene function information which is key to understanding complex microbial communities.



Human microbiome research

The human microbiome varies widely among individuals and populations, and can affect health status. Leveraging shotgun metagenomics, scientists can identify the impact of human health or disease based on variations and biodiversity of the microbiome. Shotgun metagenomics also enables detection of anti-microbial resistance genes and virulence genes that impact human health.

STEP 1
Recommended library prep methods


Product	Nextera DNA Flex Library Prep Kit	TruSeq DNA PCR-Free Library Prep Kit
Most important to me	Broad DNA input range Fewest number of workflow steps High uniformity of coverage	Lowest bias to achieve highest coverage[a]
Input quantity	1–500 ng DNA	1 µg DNA
Assay time	3–4 hrs	5 hrs
Diversity	≥ 1 billion unique fragments	≥ 1 billion unique fragments
Genome coverage	Complete	Complete
Fragmentation process	Enzymatic	Mechanical

a. Comparisons are within the Illumina targeted sequencing portfolio

STEP 2
Recommended sequencers


Product	NextSeq 550 System	NovaSeq 6000 System
Most important to me	Desktop system with mid-level throughput and instrument affordability	Scalable throughput, longer read lenghts, and discovery power [a]
Run mode/kit type	Mid-output/ High-output	SP-S4
Shotgun metagenomic samples per run[b]	10 samples (Mid-output) 16 samples (High-output)	32 samples (SP flow cell)
Recommended read length for shotgun metagenomics	2 × 150 bp	2 × 250 bp
Max output per flow cell	39 Gb (Mid-output v2.5) 120 Gb (High-output v2.5)	400 Gb (SP flow cell) 500 Gb (S1 flow cell)
Flow cells processed/run	1	1 or 2

a. Comparison among Illumina sequencing portfolio

b. Calculations based on soil metagenomic samples which contains high species diversity. A minimum 25 million reads per sample is recommended for sensitive species detection and functional identification. Other sample types and metagenomics analysis may vary.

www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods/shotgun-metagenomic-sequencing.html

STEP 3
Data analysis

Analyze shotgun metagenomics data with Local Run Manager Software or BaseSpace Sequence Hub Apps

Taxonomic classification/ Community profiling

Kraken2 Metagenomics B
- Assigns taxonomic labels to short DNA sequences with high sensitivity and speed using exact k-mer alignments and a novel classification algorithm
- Utilizes RefSeq reference genomes as the database for classification

CosmosID Metagenomics B
- Easy to use third party application that performs microbial identification at the species, subspecies, or strain level
- Features a private, curated database www.cosmosid.com

MetaPhlAn B
- Profiles the composition of microbial communities from metagenomic shotgun sequencing data
- Metagenomic microbial community profiling relies on unique clade-specific marker genes identified from reference genomes

Sequencing and *de novo* assembly

SPAdes Genome Assembler B
- Assembles genomes from standard bacterial isolates and single-cell multiple displacement amplification (MDA) preparations

L Local Run Manager Software B BaseSpace Sequence Hub

www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods/shotgun-metagenomic-sequencing.html

Library prep ordering information

Nextera DNA Flex Library Prep Kits

Save time and reduce hands-on touch points with On-Bead Tagmentation, producing sequencing-ready libraries in less than three hours. Simplify daily operations with a kit that supports a broad DNA input range (1–500 ng), multiple DNA input types, and flexible for use with small (bacteria, archaea, viruses, plasmids and other microbial isolates) to large genomes (human, plant, mouse). Nextera DNA Flex Library Prep Kits deliver consistent insert sizes, uniform coverage, and optimized performance, regardless of DNA input amount or genome size. On-bead tagmentation minimizes bias and opportunities for error, resulting in highly robust data without pre-library prep DNA quantification steps, or post-library prep quantification steps such as library normalization and pooling. Nextera DNA Flex libraries are compatible with all Illumina sequencers.

Ordering information

Product	No. of indexes	No. of samples	Catalog no.
Nextera DNA Flex Library Prep Kit	-	24	20018704
Nextera DNA Flex Library Prep Kit	-	96	20018705
Flex Lysis Reagent Kit	-	-	20018706
Nextera DNA CD Indexes	24	24	20018707
Nextera DNA CD Indexes	96	96	20018708
IDT for Illumina Nextera DNA UD Indexes Set A	96	96	20027213
IDT for Illumina Nextera DNA UD Indexes Set B	96	96	20027214
IDT for Illumina Nextera DNA UD Indexes Set C	96	96	20027215
IDT for Illumina Nextera DNA UD Indexes Set D	96	96	20027216
Training			
Nextera DNA Flex Library Prep Kit training at customer site	-	-	20022900

TruSeq DNA PCR-Free Library Prep Modular Kits

TruSeq DNA PCR-Free Library Prep Kits provide uniform coverage for whole-genome library prep for organisms ranging from bacteria to human. The kits offer shortened gel-free workflows, the ability to sequence the most challenging regions, and the power to identify a large number of variants. Libraries prepared with TruSeq DNA PCR-Free Library Prep Kits are compatible with all Illumina sequencing systems. TruSeq DNA PCR-Free Library Prep Kits are available as single components, allowing users to take advantage of proven Illumina chemistry and use indexes from other sources, such as Integrated DNA Technologies (IDT). This also offers users the opportunity to increase assay plexity and enables accurate assignment of reads and more efficient use of flow cells. The library preparation protocol requires one kit from each component, depending on sample requirements.

Ordering information

Product	No. of indexes	No. of samples	Catalog no.
Library Prep Component			
TruSeq DNA PCR-Free Low-Throughput Library Prep Kit	-	24	20015962
TruSeq DNA PCR-Free High-Throughput Library Prep Kit	-	96	20015963
Index Adapter Component			
TruSeq DNA Single Indexes Set A	12	24	20015960
TruSeq DNA Single Indexes Set B	12	24	20015961
TruSeq DNA CD Indexes	96	96	20015949
IDT for Illumina - TruSeq DNA UD Indexes	24	96	20020590
IDT for Illumina - TruSeq DNA UD Indexes	96	96	20022370
Training			
TruSeq DNA PCR-Free Library Prep Kit training at customer site	-	-	TR-204-0011

www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods/shotgun-metagenomic-sequencing.html

16S and ITS rRNA metagenomics

Culture-free sequencing of bacterial or fungal communities to identify and classify microbial populations

Key benefits

- Enables culture-free bacterial and fungal identification and comparison within mixed population samples
- Promotes accurate evaluation of microbial biodiversity and relative abundance measurement

Prepare library	Sequence	Analyze and Report
		
16S rRNA Metagenomic Sequencing Protocol **ITS rRNA Metagenomic Sequencing Protocol**	**iSeq 100 System** **MiniSeq System** **MiSeq System**	**Local Run Manager Software** 16S Metagenomics (iSeq, MiSeq) **BaseSpace Sequence Hub** 16S Metagenomics ITS Metagenomics

Illumina Informatics Solutions
BaseSpace Sequence Hub
Run management | Analysis | Reporting

Service contracts	Professional services
Tiered service plans iSeq 100 System MiniSeq System MiSeq System	**Proof of Concept Service** Standard application functional testing with your samples

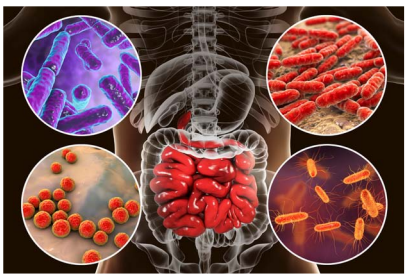
www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods.html

For Research Use Only. Not for use in diagnostic procedures.

Innovation. Discovery. Application.

Microbiome

16S and ITS sequencing are established methods for achieving genus level surveillance of bacterial and fungal populations, respectively, in human, animal, and environmental microbiome studies.

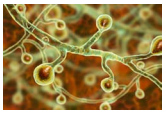


Human microbiome research

The human gut microbiome varies widely among individuals and populations, and can affect health status. A culture-free approach to evaluating human microbiome samples allows researchers to study the presence of bacterial and/or fungal communities and its association to human health and disease.

www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods.html

STEP 1
Recommended library prep methods



	16S Metagenomic Sequencing Protocol[a]	ITS Metagenomic Sequencing Protocol[a]
Input quantity	10-15 ng DNA	10-15 ng DNA
Assay time	~6 hrs	~6 hrs
Read length	2 × 250	2 × 150
Multiplexing capacity	384[b]	384[b]
Genome coverage	V3 and V4 regions of the 16S rRNA gene	ITS1 region of rRNA cistron
Sequencing method	Amplicon sequencing	Amplicon sequencing
View protocol	16S protocol	ITS protocol

a. Demonstrated protocol

b. Multiplexing with Nextera XT v2 indexes kits A-D

STEP 2
Recommended sequencers

Product	iSeq 100 System	MiniSeq System	MiSeq System
Most important to me	Affordability and efficiency	Simplicity and instrument affordability	Longest read lengths in Illumina portfolio, well-suited for microbiology applications
16S metagenomic samples per flow cell at 100K clusters per sample[a]	40	250	250
ITS metagenomic samples per flow cell at 100K clusters per sample[a]	40	250	250
Recommended read length for targeted metagenomics	2 × 150 bp	2 × 150 bp	2 × 250 bp
Run mode/kit type	Standard	Mid-output/ High-output	Research mode and Dx
Max output per flow cell	1.2 Gb	2.4 Gb (Mid-output) 7.5 Gb (High-output)	0.5 Gb (Nano v2) 1.2 Gb (Micro v2) 7.5 Gb (Standard v2) 15 Gb (Standard v3)
Flow cells processed/run	1	1	1

a. Refer to demonstrated protocol for range of clusters per sample

www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods.html

STEP 3

Data analysis

Analyze 16S and ITS rRNA data for bacterial and fungal community profiling with Local Run Manager or BaseSpace Sequence Hub Apps

Apps for 16S and ITS metagenomics

16S Metagenomics L B

- The 16S Metagenomics app performs taxonomic classification of 16S rRNA targeted amplicon reads using an Illumina-curated version of the GreenGenes taxonomic database
- The app provides interactive visualizations and raw classification output for per-sample and aggregate analyses

ITS Metagenomics B

- The ITS Metagenomics analysis app performs taxonomic classification of Fungal rRNA targeted amplicon reads using the UNITE taxonomic database
- The app provides interactive visualizations and raw classification output for per-sample and aggregate analyses

QIIME 2 B

- An open-source bioinformatics pipeline for performing microbiome analysis from raw DNA sequencing data

Commercial 3rd party apps

ChunLab EzBioCloud Microbiome B

Taxonomic Profiling

- The comprehensive EzBioCloud 16S database contains >2300 species represented by accurate, full-length 16S rRNA sequences allowing species-level profiling using closed-reference OTU picking strategy
  www.ezbiocloud.net

L Local Run Manager Software B BaseSpace Sequence Hub

www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods.html

Library prep information

16S Metagenomic Sequencing Protocol (Demonstrated Protocol)

Variable regions of 16S rRNA are frequently used in phylogenetic classifications such as genus or species in diverse microbial populations. This protocol describes a method for preparing samples for sequencing the variable V3 and V4 regions of the 16S rRNA gene. When combined with a benchtop sequencing system, on-board primary analysis, and secondary analysis using the 16S Metagenomics App available on iSeq 100 and MiSeq or analyzing data with BaseSpace Sequence Hub, this protocol provides a comprehensive workflow for 16S rRNA amplicon sequencing.

Ordering information

Product	No. of indexes	No. of samples	Catalog no.
Nextera DNA CD Indexes	24	24	20018707
Nextera DNA CD Indexes	96	96	20018708
Nextera XT Index Kit v2, set A	96	384	FC-131-2001
Nextera XT Index Kit v2, set B	96	384	FC-131-2002
Nextera XT Index Kit v2, set C	96	384	FC-131-2003
Nextera XT Index Kit v2, set D	96	384	FC-131-2004
IDT for Illumina Nextera DNA UD Indexes Set A	96	96	20027213
IDT for Illumina Nextera DNA UD Indexes Set B	96	96	20027214
IDT for Illumina Nextera DNA UD Indexes Set C	96	96	20027215
IDT for Illumina Nextera DNA UD Indexes Set D	96	96	20027216
PhiX Control Kit v3	-	-	FC-110-3001

ITS Metagenomic Sequencing Protocol (Demonstrated Protocol)

The ITS Metagenomic Sequencing Protocol describes a method for preparing libraries that allows researchers to sequence the Internal Transcribed Spacer Region 1 (ITS) of the fungal ribosomal RNA cistron. The ITS region of the rRNA cistron is a useful DNA barcode for characterization of fungi in complex sample types such as stool, sputum, skin, soil, and water. This protocol details a workflow for the amplification and analysis of the ITS1 region to provide accurate classification resolution for a broad range of taxa. Primary and secondary analysis using ITS Metagenomics App on BaseSpace Sequence Hub provides a comprehensive worklfow for ITS rRNA amplicon sequencing.

Ordering information

Product	No. of indexes	No. of samples	Catalog no.
Nextera DNA CD Indexes	24	24	20018707
Nextera DNA CD Indexes	96	96	20018708
Nextera XT Index Kit v2, set A	96	384	FC-131-2001
Nextera XT Index Kit v2, set B	96	384	FC-131-2002
Nextera XT Index Kit v2, set C	96	384	FC-131-2003
Nextera XT Index Kit v2, set D	96	384	FC-131-2004
IDT for Illumina Nextera DNA UD Indexes Set A	96	96	20027213
IDT for Illumina Nextera DNA UD Indexes Set B	96	96	20027214
IDT for Illumina Nextera DNA UD Indexes Set C	96	96	20027215
IDT for Illumina Nextera DNA UD Indexes Set D	96	96	20027216
PhiX Control Kit v3	-	-	FC-110-3001

www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods.html

AmpliSeq for Illumina Panels

Prepare high-quality libraries for microbiology applications quickly and simply using a multiplexed PCR-based workflow.

AmpliSeq for Illumina panels for microbiology

AmpliSeq chemistry allows researchers to analyze hundreds of genes simultaneously with ultra-high multiplexed PCR with 12 to more than 24,000 amplicons in a single panel. AmpliSeq for Illumina works with DNA and RNA samples and requires as little as 1 ng of input. AmpliSeq panels can accommodate high-quality samples such as blood, cell culture, or fresh frozen tissue and also challenging samples such as formalin-fixed paraffin-embedded (FFPE) tissue. Select from Community Panels or Custom Panels for a range of microbiology research applications.

Panel types

- **Community Panels**—AmpliSeq for Illumina Community Panels are made-to-order. The designs are provided for your convenience; however, they do not have associated performance metrics. AmpliSeq for Illumina Community Panels contain content selected and designed with input from leading disease researchers. You can order Community Panels as is or customize the panel to meet your needs, using DesignStudio™ Software to add or remove amplicons.

- **Custom Panels**—If your genes of interest are not available as ready-to-use or community panels, consider designing a custom AmpliSeq for Illumina panel. Include just the genes and targets you're interested in. For faster design, you can easily modify genes or targets from AmpliSeq Community Panels or AmpliSeq Ready-to-Use Panels using DesignStudio Software.

www.illumina.com/products/by-brand/ampliseq.html

Library prep ordering information

Ordering information

Product	Product description	Panel type	Catalog no.
AmpliSeq for Illumina Antimicrobial Resistance (AMR) Research Panel	Contains 2 pools with a total of 815 amplicons to assess the presence of 478 antimicrobial resistance genes. Developed in collaboration with a team from Lawrence Livermore National Laboratory, University of Chicago, Argonne National Laboratory, Los Alamos National Laboratory and Naval Research Laboratory.	Community Panel	20020495
AmpliSeq for Illumina Ebola Research Panel	Contains 150 amplicons in 2 pools targeting the Ebola genome. The reference genome came from the UCSC Ebola virus Sierra Leone 2014 genome portal.	Community Panel	20020496
AmpliSeq for Illumina TB Reseach Panel	Includes 109 amplicons in 2 pools from 8 genes related to antimicrobial resistance (*embB*, *eis*, *gyrA, inhA, katG, pncA, rpoB, and rpsL*). Enables researchers to identify gene variants correlating with antimicrobial resistance in *Mycobacterium tuberculosis*.	Community Panel	20020495
AmpliSeq Custom DNA Panel for Illumina	Flexible and tunable designs from 12 to over 12,000 amplicons, and multiple amplicon size options. Design custom content for specific genes or targets of interest with DesignStudio Software.	Custom Panel	20020495
AmpliSeq for Illumina Custom RNA Panel	Measures gene expression in 12 to 1200 gene targets in a single assay. Design custom targeted panels from a menu of over 20,000 human RefSeq genes using DesignStudio Software.	Custom Panel	20020496
AmpliSeq UD Indexes for Illumina (24 Indexes, 24 Samples)	Kit contains 24 unique dual indexes in a plate format. Sufficient for 24 samples.	-	20019104
AmpliSeq CD Indexes for Illumina Set A - D (384 Indexes, 384 Samples)	Kit includes 4 sets of AmpliSeq CD indexes for Illumina: Set A, Set B, Set C, and Set D. Each set contains 96 dual indexes in a plate, sufficient for 96 samples. Sets can also be purchased individually.	-	20031676

Data analysis

Analyze AmpliSeq for Illumina data with 3rd party tools.

Analysis of AmpliSeq for Illumina AMR Research Panel

One Codex

- Analyzes NGS data generated from AmpliSeq for Illumina Antimicrobial Resistance Research Panel
- Analyzes multiple samples and delivers a PDF report that aggregates data into gene-level presence/absence calls for each AMR gene.
  docs.onecodex.com/docs/intro-getting-started-ampliseq-amr

Analysis of AmpliSeq for Illumina TB Research Panel

genTB

- A genomics tool for analyzing and predicting drug resistances to tuberculosis (TB)
  gentb.hms.harvard.edu/

bioMérieux EPISEQ

- Comprehensive, integrated applications for microbiological research and routine applications. Bionumerics combines information from various genomic and phenotypic sources into one global database and conducts combined analyses.
  www.applied-maths.com/applications

Data Analysis

Data analysis

Local Run Manager

Local Run Manager is an easy-to-use data management and analysis software for Illumina desktop sequencers. It provides a range of data analysis tools and generates comprehensive PDF reports for multiple applications, such as amplicon sequencing, targeted resequencing, small RNA sequencing, and more. With Local Run Manager, both expert and novice users can harness the power of next-generation sequencing right from their own lab. After a sequencing run is complete, Local Run Manager automatically launches data analysis using one of the application-specific analysis modules.

Local Run Manager features:

- Integrates with the instrument control software when run on the instrument computer
- Records samples and input files, such as manifests and reference genomes.
- Provides a run setup interface specific to the selected analysis module.
- Performs data analysis on completed sequencing runs.
- Performs a series of analysis steps specific to the selected analysis module.
- Shows analysis metrics upon completion of analysis.

Local Run Manager is included on the iSeq 100, MiniSeq, MiSeq, and NextSeq 550 Systems.

BaseSpace Sequence Hub

As the cost of sequencing decreases, the volume of NGS-generated data increases, creating new challenges such as secure data storage and management, complex data analysis, and sharing results with collaborators. BaseSpace Sequence Hub is a genomics cloud computing platform designed to bring simplified data management and analytical sequencing tools directly to investigators.

BaseSpace Sequence Hub feature:

- Real-time data upload and run monitoring
- One-click analysis with over 80 bioinformatics tools
- Configure analysis workflows to remove touchpoints and streamline data processing
- Easy global collaboration and data sharing

www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps.html

BaseSpace Sequence Hub Apps for microbiology: Powerful yet simple

Analysis of complex sequencing data sets is a challenge at any scale. BaseSpace Sequence Hub provides a continuously growing list of powerful apps, allowing researchers to set up and perform complex data analyses. A simple interface links data sets directly to bioinformatics pipelines based on open source and commercial tools. In addition to Illumina developed apps, BaseSpace Sequence Hub hosts third-party premium apps developed by commercial partners.

Taxonomic classification

16S Metagenomics L B

- Performs taxonomic classification of 16S rRNA targeted amplicon reads using an Illumina-curated version of the GreenGenes taxonomic database
- The app provides interactive visualizations and raw classification output for per-sample and aggregate analyses

ITS Metagenomics B

- The ITS Metagenomics analysis app performs taxonomic classification of Fungal rRNA targeted amplicon reads using the UNITE taxonomic database
- The app provides interactive visualizations and raw classification output for per-sample and aggregate analyses

QIIME 2 B

- An open-source bioinformatics pipeline for performing microbiome analysis from raw DNA sequencing data

ChunLab EzBioCloud Microbiome B Taxonomic Profiling

- The comprehensive EzBioCloud 16S database contains >2300 species represented by accurate, full-length 16S rRNA sequences allowing species-level profiling using closed-reference OTU picking strategy

www.ezbiocloud.net

Kraken2 Metagenomics B

- Assigns taxonomic labels to short DNA sequences with high sensitivity and speed using exact alignments of k-mers and a novel classification algorithm

Pathogen detection (isolates)

BWA* Aligner B and Pisces Variant Caller B

- These 2 apps perform read alignment and mutation detection for any genome with a known reference sequence. If none is available, the pipeline can still be used to analyze bacteria, fungi, or viruses after *de novo* assembly

*BWA: Burrows-Wheeler Aligner

SRST2 B

- Reports the presence of STs (sequence types) from a MLST database and/or reference genes from a database of sequences for virulence genes, resistance genes, and plasmid replicons

Bacterial Analysis Pipeline B

- Identifies species, multilocus sequence type, plasmids, virulence - and antimicrobial resistance genes in bacteria

Virology

DeepChek (HIV-1, HCV, HBV) B

- Performs deep sequencing analyses and reports on subtyping, genotyping, and inferred levels of resistance.

Kraken2 Metagenomics B

- Assigns taxonomic labels to short DNA sequences with high sensitivity and speed using exact alignments of k-mers and a novel classification algorithm

One Codex

- Provides a fast and easy-to-use analysis platform for viral metagenomics from the curated One Codex genome database

www.onecodex.com/

L Local Run Manager Software B BaseSpace Sequence Hub

For Research Use Only. Not for use in diagnostic procedures.

De novo assembly

Assembly workflow L

- Assembles small genomes (< 20 Mb) and best suits the assembly of bacterial genomes, such as *E. coli*
- Uses the EMBL Velvet algorithm and writes assembly results in FASTA format

SPAdes Genome Assembler B

- Assembles genomes from standard bacterial isolates and single-cell multiple displacement amplification (MDA) preparations

Shotgun metagenomics

CosmosID Metagenomics B

- Provides rapid, accurate bacterial identification at the species, subspecies, and/or strain level from the curated CosmosID genome database

Kraken2 Metagenomics B

- Assigns taxonomic labels to short DNA sequences with high sensitivity and speed using exact alignments of k-mers and a novel classification algorithm

MetaPhlan B

- A computational tool for profiling the composition of microbial communities from metagenomic shotgun sequencing data

Gene annotation

Prokka Genome Annotation B

- Rapidly annotates genes and identifies coding sequences in prokaryotic genomes
- Does not annotate eukaryotic genome

Commercial 3rd party apps

bioMérieux EPISEQ

- Comprehensive, integrated applications for microbiological research and routine applications. Bionumerics combines information from various genomic and phenotypic sources into one global database and conducts combined analyses. www.applied-maths.com/applications

ChunLab TrueBac ID

- A cloud based system that incorporates the "gold standard" of bacterial identification by using whole genome sequences. TrueBac ID can identify >12,000 species for microbial applications in research, applied and industrial settings www.truebacid.com

ChunLab EzBioCloud Microbiome B
Taxonomic Profiling

- The comprehensive EzBioCloud 16S database contains >2300 species represented by accurate, full-length 16S rRNA sequences allowing species-level profiling using closed-reference OTU picking strategy www.ezbiocloud.net

CosmosID Metagenomics

- Provides rapid, accurate bacterial identification at the species, subspecies, and/or strain level from the curated CosmosID genome databas www.cosmosid.com/

One Codex

- Provides a fast and easy-to-use analysis platform for viral metagenomics from the curated One Codex genome database www.onecodex.com/

Services

Instrument services, training, and consulting

We provide solutions so you can focus on the big discoveries

Illumina offers an integrated service solution that combines product care with opportunities for training and consulting support to help boost your lab efficiency and productivity, and optimize your workflow. With our flexible offering, you can select the services you need, so you can focus on what matters in your lab.

Illumina Product Support Services

In addition to the one-year basic service warranty included with your new purchase, Illumina offers instrument maintenance, repair, and qualification solutions. From basic to fully dedicated onsite options, we strive to anticipate your needs and exceed your expectations, including reagent replacement if one of our instruments fails during its service plan coverage.

Illumina University

Get high-quality results on Illumina technology even faster. Whether you want to maximize the effectiveness of your Illumina system, train new employees, or learn the latest techniques and best practices, we have a huge assortment of instructor-led, hands-on courses and web-based options for you at Illumina University.

Illumina Consulting

Illumina Consulting offers a suite of workflow, operation, and bioinformatics consulting services that expedite lab startup and maximize NGS and microarray workflow efficiencies so you can find the answers you seek. Harness our global network of expertise in genomics applications, IT/networking, data management, high-throughput operations, and bioinformatics.

www.illumina.com/ProductServices

For your Illumina next-generation solutions from library prep, to sequencing and data analysis, we are here to provide you with the resources you need to accelerate your microbiology research.

Industry-leading solutions.

A community of support.

From library prep, arrays, and sequencing to informatics, Illumina genomic solutions empower researchers and clinical researchers across the globe to find the answers they seek.

When you join the Illumina community, you become part of a dynamic scientific movement that includes thousands of researchers and industry thought leaders. Throughout the year, we host user group meetings, symposia, consortia, online forums, and other initiatives—all designed to bring the best minds together to share ideas and advance science.

In addition to on-site training, ongoing support, and phone consults, we offer webinars and courses at various Illumina locations. We're here with all the resources you need to accelerate your research.

A global genomics leader, Illumina delivers next-generation sequencing workflow solutions to the basic and translational research communities. More than 90% of the world's sequencing data is generated using Illumina sequencing by synthesis technology.* Through collaborative innovation, Illumina is fueling groundbreaking advances in oncology, reproductive health, genetic disease, agriculture, and microbiology.

* Data calculations on file. Illumina, Inc., 2018.

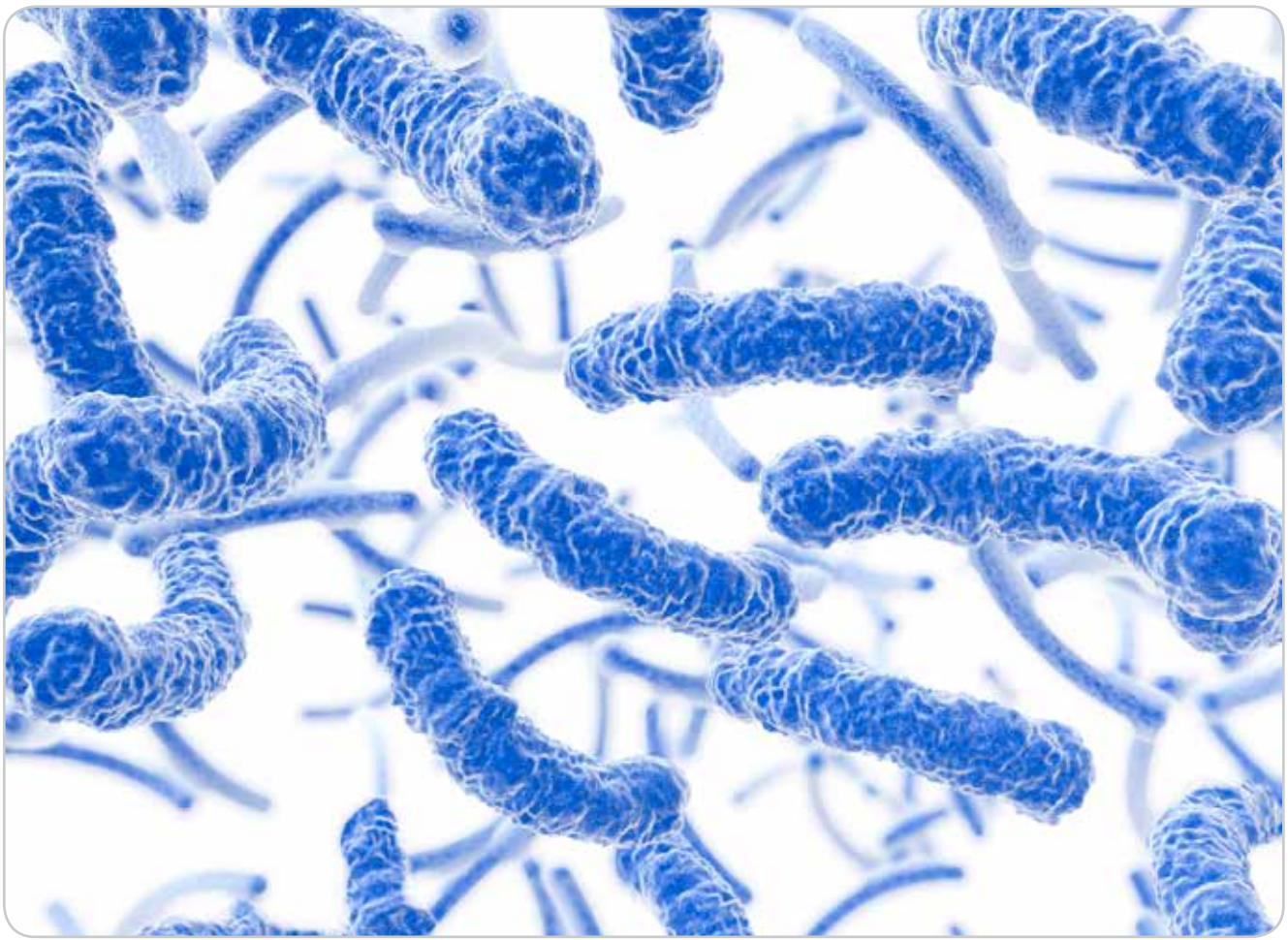
Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

For Research Use Only. Not for use in diagnostic procedures.

illumina®

An Introduction to Illumina Next-Generation Sequencing Technology for Microbiologists

Deciphering DNA sequences is essential for virtually all branches of biological research. Capillary electrophoresis (CE)-based sequencing has enabled scientists to elucidate genetic information from almost any organism or biological system. Although this technology has become widely adopted, inherent limitations in throughput, scalability, cost, speed, and resolution can hinder scientists from obtaining essential genomic information. To overcome these barriers, an entirely new technology was developed—next-generation sequencing (NGS), a fundamentally different approach to sequencing that has triggered numerous ground-breaking discoveries. The years since the introduction of NGS have seen a major transformation in the way scientists extract genetic information from biological systems, revealing insight about the genome, transcriptome, and epigenome. This introduction will highlight the benefits of using NGS for microbiology research.



www.illumina.com/microbiology

illumina®

Welcome to Next-Generation Sequencing

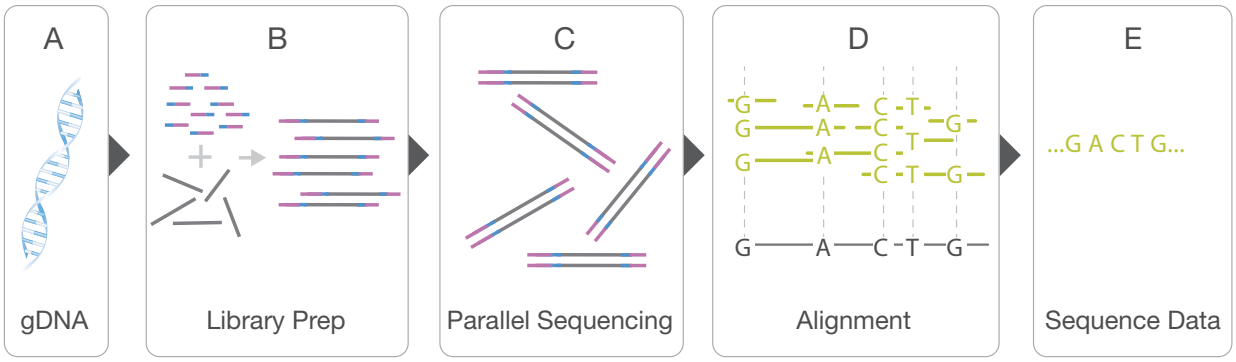
Advances in next-generation sequencing have been instrumental in advancing scientific fields from human disease research to environmental and evolutionary science. NGS lends itself particularly well to the microbial laboratory, where the genomes are small and the data analysis is relatively simple. One major advantage of using NGS over other genome interrogation methods is the ability to measure changes anywhere in the genome without prior knowledge, such as for unculturable organisms. Single-base resolution allows tracking of microbial adaptation over short periods of time, both in the laboratory and in the environment.

As evidence of the uptake of this technology, NGS data output has increased at a phenomenal rate, more than doubling each year since it was introduced. In 2007, a single sequencing run could produce about one gigabase (Gb) of data. By 2012, that rate had jumped 1000×, to one terabase (Tb) of data in a single sequencing run. With this tremendous increase in output has come a 10^5-fold decrease in the cost of determining the genome of a microorganism. In 1995, sequencing the 1.8 megabase (Mb) genome of *Haemophilus influenzae* with CE technology cost ~1 million US dollars, taking over one year. Today, sequencing the 5 Mb genome of *Escherichia coli* with Illumina NGS can be done in one day at a fraction of the cost.

Basic Concepts of NGS

In principle, NGS is similar to Sanger-based, or CE sequencing. The bases of a small fragment of DNA are sequentially identified from signals emitted as each fragment is re-synthesized from a DNA template strand. NGS extends this process across millions of reactions in a massively parallel fashion, rather than being limited to a single or a few DNA fragments. This advance enables rapid sequencing of large stretches of DNA, with the latest instruments capable of producing hundreds of gigabases of data in a single sequencing run. To illustrate how this process works, consider a single genomic DNA (gDNA) sample. The gDNA is first fragmented into a library of small segments and sequenced. The newly identified strings of bases, called reads, are then reassembled using a known reference genome as a scaffold (resequencing), or assembled together using advanced computational techniques if no reference genome is available (de novo sequencing). The full set of aligned reads reveals the entire genomic sequence of the sample (Figure 1). Once the sample library is prepared, all of the sequencing steps through data analysis can be performed on a single instrument, facilitating rapid turnaround with minimal hands-on time.

Figure 1: Concepts of Next-Generation Sequencing



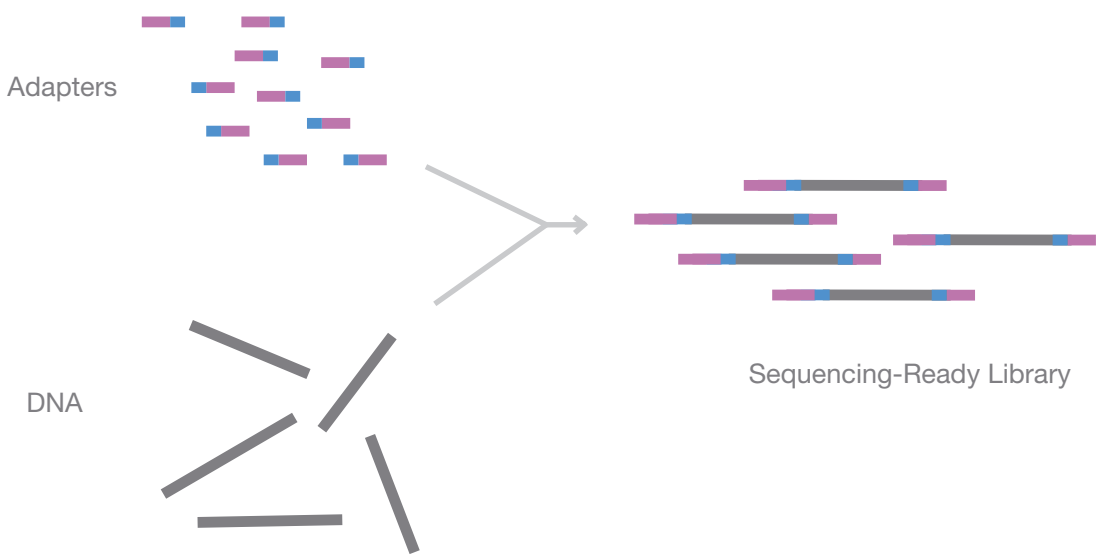
A. Extracted gDNA.
B. Sample preparation fragments genomic DNA and adds adapters to generate a library.
C. DNA fragments within the library are each sequenced in parallel.
D. Individual sequence reads are reassembled by aligning to a reference genome.
E. The whole genome sequence is derived from the consensus of aligned reads.

Sample Preparation

How NGS is used experimentally is largely dictated by the way sequencing libraries are prepared and the way the data is analyzed, with the actual sequencing steps remaining fundamentally unchanged. A growing number of library preparation kits provide complete reagent sets and protocols for sequencing whole genomes, small genomes, mRNA, targeted regions such as whole exomes, custom-selected regions, protein-binding regions, and more. To address specific research objectives, researchers have developed many novel protocols to isolate specific regions of the genome associated with a given biological function.

Sample preparation protocols for NGS are generally more rapid and straightforward than those for CE-based Sanger sequencing. With NGS, researchers can start directly from a gDNA or cDNA library. The DNA fragments are then ligated to specific oligonucleotide adapters needed to perform the sequencing biochemistry, requiring as little as 90 minutes with Illumina's Nextera® technology (Figure 2). In contrast, CE-based Sanger sequencing requires genomic DNA to be fragmented first and cloned into either bacterial artificial chromosomes (BACs) or yeast artificial chromosomes (YACs). Then, each BAC/YAC must be further subcloned into a sequencing vector and transformed into the appropriate microbial host. Template DNA is then purified from individual colonies or plaques prior to sequencing. This process can take days or even weeks to complete.

Figure 2: NGS Sample Preparation Overview



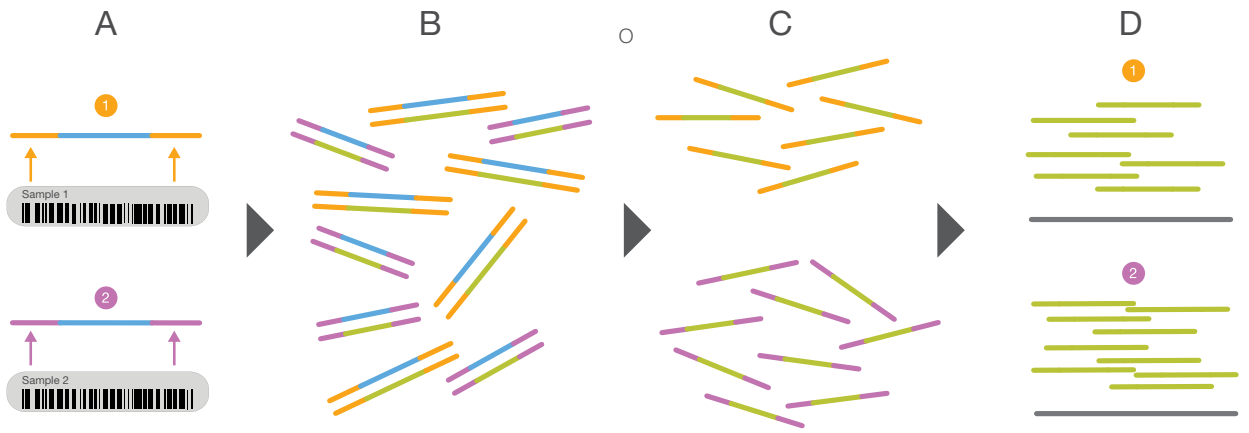
NGS library preparation starts directly from fragmented genomic DNA. Platform-specific oligonucleotide adapters, needed to perform the sequencing biochemistry and index samples, are ligated to each end of the fragments to yield the sequencing-ready library.

Scalable Studies Enabled by Multiplexing

For sequencing small bacterial/viral genomes, a researcher can choose to use a lower output instrument and process a smaller number of samples per run, or can opt to process a large number of samples. Multiplexing enables large numbers of samples to be simultaneously sequenced during a single experiment (Figure 3). To accomplish this, individual "barcode" sequences are added to each sample so they can be differentiated during the data analysis.

With multiplexing, NGS dramatically reduces the time to data for large numbers of samples. Processing hundreds of amplicons using CE technology generally requires several weeks or months. The same number of samples can now be sequenced in a matter of hours and fully analyzed within two days using NGS. With highly automated, easy-to-use protocols, researchers can go from experiment to data to publication faster and easier than ever before.

Figure 3: Sample Multiplexing Overview



A. Two representative DNA fragments from two unique samples, each attached to a specific barcode sequence that identifies the sample from which it originated.
B. Libraries for each sample are pooled and sequenced in parallel. Each new read contains both the fragment sequence and its sample-identifying barcode.
C. Barcode sequences are used to de-multiplex, or differentiate reads from each sample.
D. Each set of reads is aligned to the reference sequence.

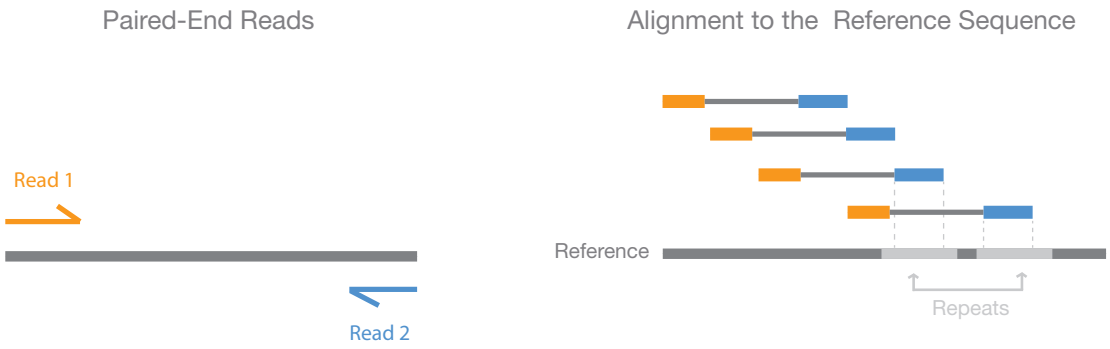
Paired-End Sequencing

Paired-end (PE) sequencing, where both ends of a DNA fragment are sequenced (Figure 4) allows long range positioning of the DNA fragment. Because the distance between each paired read is known, alignment algorithms can use this information to precisely map the reads, resulting in superior alignment across difficult-to-sequence or repetitive genome regions. Illumina NGS offers the flexibility of variable insert sizes and read lengths (35–300 bp), allowing high resolution characterization of any genome.

Analyze, Store, and Share in Illumina's BaseSpace® Cloud

Data analysis is an important factor to consider for sequencing applications. One of the biggest challenges with NGS systems has been the requirement for a high-performance computing infrastructure, enterprise-level storage, and highly skilled bioinformatics and IT staff. While complex, primary data processing including alignment and variant calling happens seamlessly and behind the scenes as the sequencing run progresses. Depending on the application, most subsequent analyses can be run directly on optimized software installed on the sequencer's internal computer, or in BaseSpace, Illumina's unique cloud computing environment. Essentially, push-button informatics solutions simplify the analysis, allowing researchers to focus on the biology. By storing and analyzing data in the cloud, BaseSpace users can instantly share data with collaborators across the hallway or across the globe. The BaseSpace Application Store will provide seamless access to a wide variety of commercial software tools, as well a collection of well-known and open-source algorithms from academic institutions. These tools will provide biological interpretation and insights to further your research.

Figure 4: Paired-End Sequencing and Alignment



Paired-end sequencing enables both ends of the DNA fragment to be sequenced. Because the distance between each paired read is known, alignment algorithms can use this information to map the reads over repetitive regions more precisely. This results in much better alignment of the reads, especially across difficult-to-sequence, repetitive regions of the genome.

End-to-End Solution

Only Illumina NGS provides a fully supported solution from DNA to results, with specialized sample prep choices for the application you are working on, to robust and proven sequencing reagents, and a wide range of simple data analysis tools (Figure 5).

Figure 5: Illumina's End-to-End NGS Workflow on the MiSeq Benchtop Sequencer

NGS workflow on the MiSeq system includes sample preparation, massively parallel sequencing, automated data analysis, and cloud-enabled data analysis, storage, and sharing.

Microbiology Applications

Genomic sequencing can further any microbial identification study, including those based on known organisms, and those with incomplete or no information, as in microbiome studies or environmental surveillance. High-resolution genome data can be instrumental for examining pathogenesis, horizontal gene transfer, pangenomes, and co-evolution of hosts and symbionts/parasites. The wealth of information enabled by NGS is beneficial for mutational studies of all kinds, including directed evolution strategies, lab adaptation analyses, mutagenesis screens, or studying the temporal and spatial dynamics of epidemics and transmission.

Another important advantage of NGS is the abundance of sequence information. Deep sequencing makes it possible to detect very low abundant members of complex populations. As a result, the ability to detect low abundance populations can profoundly impact the interpretation of microbiological changes. Sequencing of microbial genomes has become routine and individual cells have been sequenced. This accumulation of sequence information has greatly expanded our appreciation of the dynamic nature of microbial populations and their impact on the environment and human health.

Whole-Genome Sequencing of Small Genomes

Until recently, sequencing an entire genome was a major endeavor. While NGS is commonly associated with sequencing large genomes, the scalability of the technology makes it just as useful for small viral or bacterial genomes. By contrast, whole-genome sequencing using CE-based Sanger technology requires significant time and resources, even for compact genomes. The ability of NGS platforms to produce a large volume of data in a short period of time makes it a powerful tool for whole-genome sequencing. The power and speed of NGS was demonstrated during the 2011 enteroaggregative *E. coli* outbreak in Europe, which prompted a rapid scientific response. Using NGS data, researchers were able to quickly generate a high-quality, whole-genome sequence of the bacterial strain, enabling them to better understand the genetic mutations conferring the increased virulence.

Table 1: Illumina NGS and CE-Based Sanger Sequencing for Small Genomes

Parameter	MiSeq System	Sanger Sequencing
Strains to be sequenced	4*	4
Genome size	3 Mb	3 Mb
Time for sample prep	1.5 hours	weeks
Sequencing time	2 days	243 days**
Price per genome†	$249 USD	$84,000 USD
Project price†	< $995 USD	> $336,000 USD
Coverage depth per genome	> 583 x	7 x
On instrument data analysis?	Yes	No

* Can sequence up to 75 strains per run at 30x coverage
**Assumes single 3730xl instrument running 24 hr/day
† Excluding sample prep

Example: *De novo* Sequencing

One challenge associated with sequencing small genomes is the lack of reference genomes available for most species. This means that whole-genome sequencing must often be done *de novo*, where the reads are assembled without aligning to a reference sequence. Paired-end reads and increasing reads lengths up to 300 bp result in good alignment across regions containing repetitive sequences and produce longer contigs for *de novo* sequencing by filling gaps in the consensus sequence, resulting in more complete coverage. Compared to CE sequencing, NGS enables researchers to simultaneously analyze many strains in one experiment, at significant time and cost savings (Table 1).

Targeted Sequencing

With targeted sequencing, only a subset of genes or defined regions in a genome are sequenced, allowing researchers to focus time, expenses, and data storage resources on the regions of the genome in which they are most interested. Amplicon sequencing refers to sequencing selected regions of the genome spanning hundreds of base pairs. The latest NGS amplicon library preparation kits allow researchers to perform rapid in-solution amplification of custom-targeted regions from genomic DNA. Using this approach, thousands of amplicons spanning multiple samples can be simultaneously prepared and indexed in a matter of hours. With the ability to process numerous amplicons and samples on a single run, NGS enables researchers to simultaneously analyze all genomic content of interest in one experiment, at fraction of the time and cost of conventional CE sequencing.

Table 2. Illumina NGS and CE-Based Sanger Sequencing for Targeted Applications

Parameter	MiSeq System	Sanger Sequencing
Samples in project	96	96
Number of amplicons	12	12
Target panel size	~5 kb	~5 kb
Time for sample prep	< 3 hours	< 3 hours
Sequencing time	1 day	6 days
Price per amplicon*	$1 USD	$4 USD
Project price*	< $2000 USD	> $4500 USD
Coverage depth per amplicon	> 13,000 x	2 x**
On instrument data analysis?	Yes	No

* Excluding PCR amplification.
** Including bidirectional sequencing.

Example: 16S Metagenomic Sequencing

A common amplicon sequencing application is comparing the bacterial 16S rRNA gene, a widely used method for studying phylogeny and taxonomy. This method has been used to evaluate bacterial diversity in many environments, allowing researchers to characterize microbiomes from samples that are otherwise difficult or impossible to study. NGS, with its ability to sequence thousands of organisms in parallel, is uniquely suited to this application. The ability to pool samples and obtain high sequence coverage during a single run allows NGS to identify rarer variants that are missed, or too expensive to identify, using CE-based sequencing approaches (Table 2). It is worth mentioning that the example in Table 2 is based on only 96 samples for the purpose of comparison. True metagenome sequencing studies, comprising hundreds or thousands of possible genomes, are cost- and labor-prohibitive with CE sequencing, and only possible with high-throughput NGS systems such as Illumina's HiSeq and NextSeq™ systems.

Take Your Research to the Next Level

The advent of NGS has enabled researchers to study biological systems at a level never before possible. With clear benefits over Sanger-based CE sequencing, next-generation sequencing can transform your microbiology research, opening new avenues to explore. To identify the sequencing platform that is optimal for your research needs, visit www.illumina.com.

From Innovation to Publication

As NGS technology continues to evolve, researchers are making fascinating discoveries in a number of biological fields, unlocking answers never before possible in all fields of research. As a result, there has been an explosion in the number of peer-reviewed scientific publications, including over 4,500 featuring Illumina sequencing technology. Selected recent examples relevant to microbiology are listed below.

Whole-Genome Sequencing

1. Toprak E, Veres A, Michel JB, Chait R, Hartl DL, et al. (2011) Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. Nat Genet 44: 101–105.
2. Chua, KYL, Seemann T, Harrison PF, Monagle S, Korman TM, et al. (2011) The dominant Australian community-acquired methicillin-resistant *Staphylococcus aureus* clone ST93-IV [2B] is highly virulent and genetically distinct. PLoS ONE 6:
3. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, et al. (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30(5): 434–9.

De novo Sequencing

4. Chitsaz H, Yee-Greenbaum JL, Tesler G, Lombardo MJ, Dupont CL, et al. (2011) Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. Nat Biotechnol 29: 915–921.
5. Rodrigue S, Malmstrom R, Berlin A, Birren B, Henn M, et al. (2009) Whole genome amplification and de novo assembly of single bacterial cells PLoS One (4)9 e6864.

Metagenomics

6. Caporaso JG, Lauber CL, Walkers, WA, Berg-Lyons D, Lozupone CA et al. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA 108:4516–22.
7. Mackelprang, R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, et al. (2011) Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. Nature 480: 368–371.

FOR RESEARCH USE ONLY

Pub No. 770-2012-023 Current as of 20 December 2013

illumina®