高通量测序仪原理与实际操作

汇报：郝淋淋

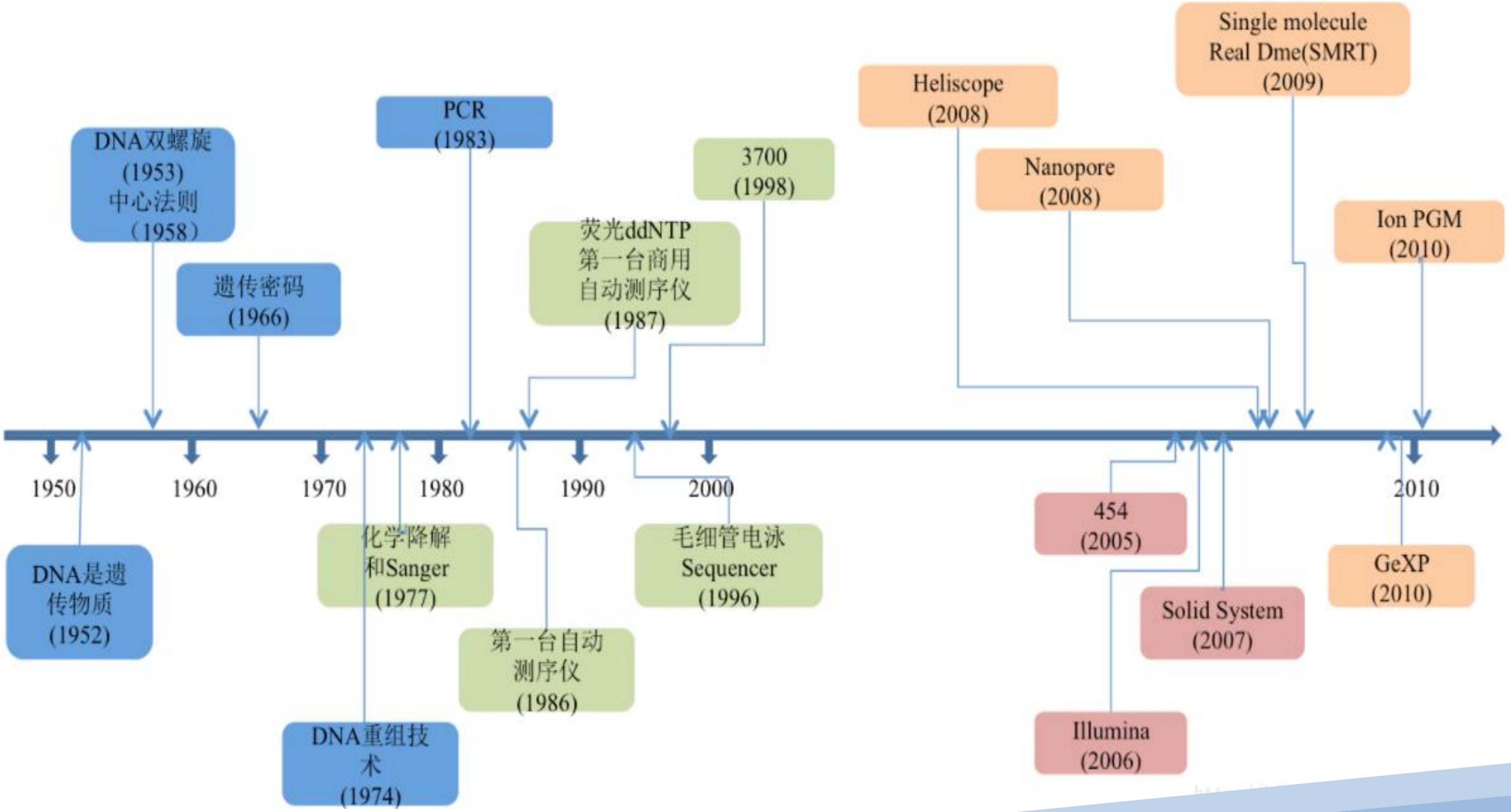
单位：中科院沈阳生态所

现代先进检测技术应用培训班

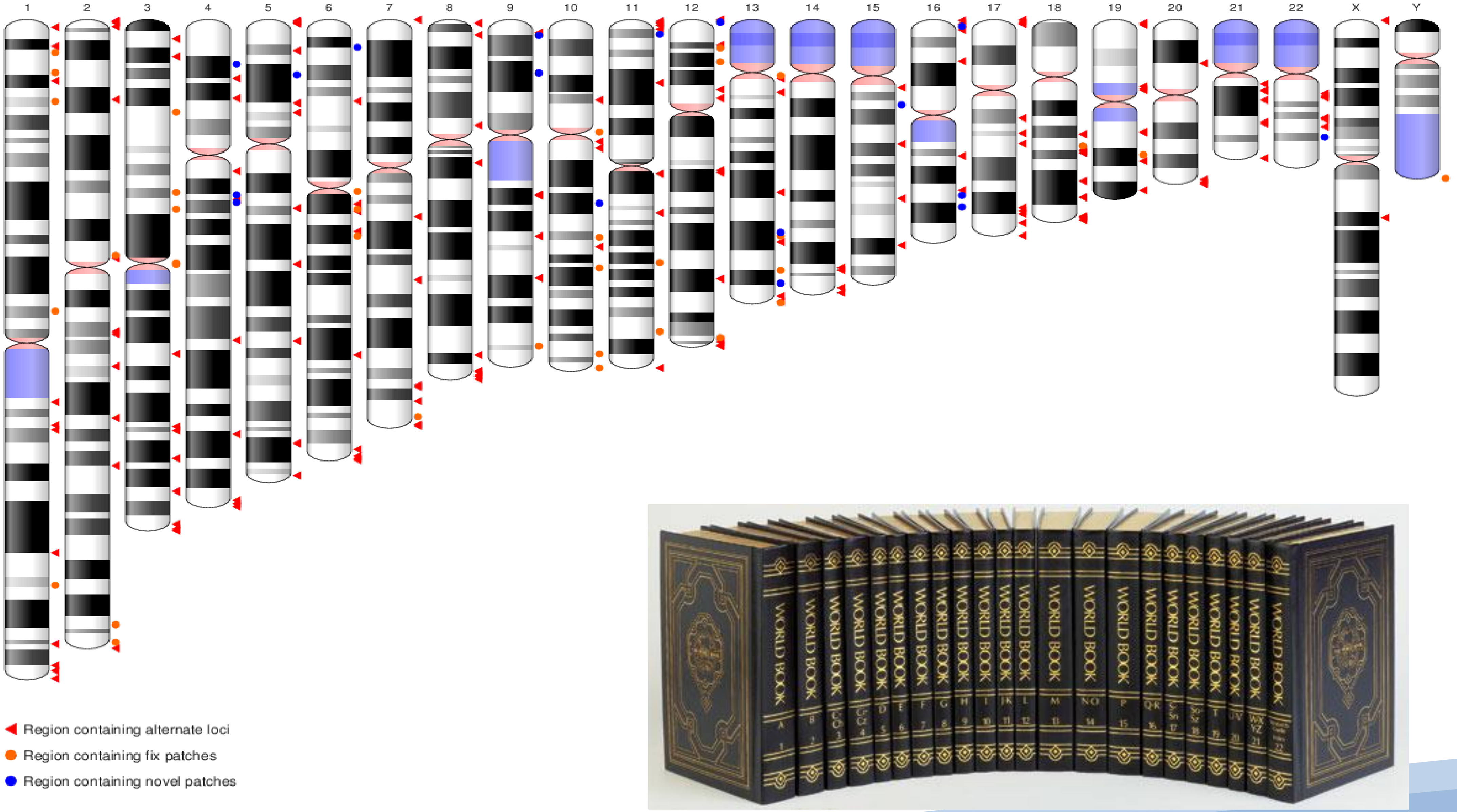
目录

CONTENTS

测序技术的发展历程


DNA双螺旋
(1953)
中心法则
（1958）
遗传密码
(1966)
PCR
(1983)
荧光ddNTP
第一台商用
自动测序仪
(1987)
3700
(1998)
Heliscope
(2008)
Nanopore
(2008)
Single molecule
Real Dme(SMRT)
(2009)
Ion PGM
(2010)
1950
1960
1970
1980
1990
2000
2010
DNA是遗
传物质
(1952)
化学降解
和Sanger
(1977)
DNA重组技
术
(1974)
第一台自动
测序仪
(1986)
毛细管电泳
Sequencer
(1996)
454
(2005)
Illumina
(2006)
Solid System
(2007)
GeXP
(2010)

人类基因组



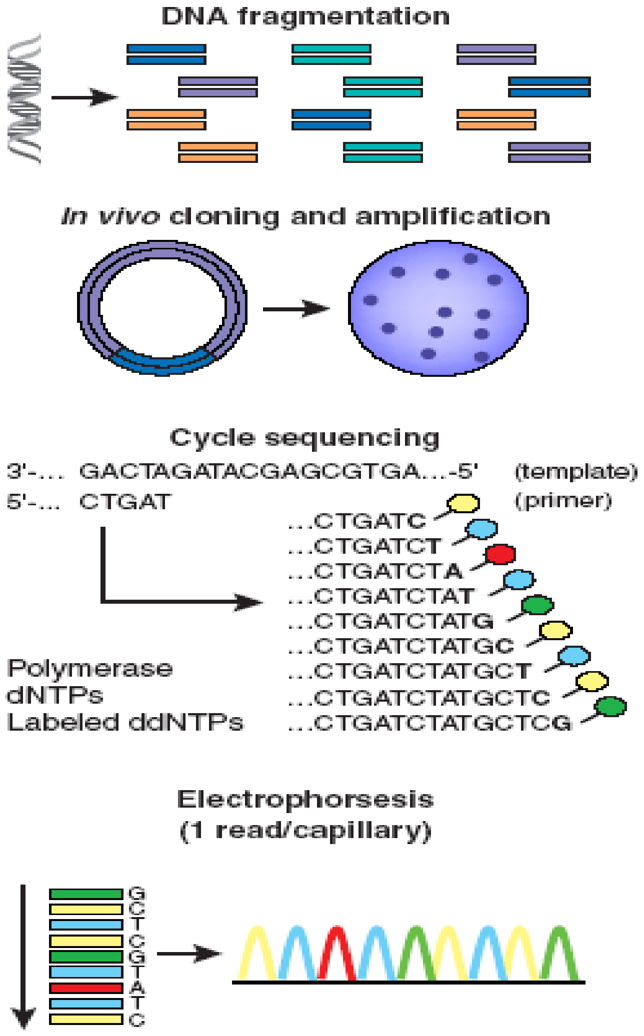
◀ Region containing alternate loci
● Region containing fix patches
● Region containing novel patches

举例：人类基因组测序

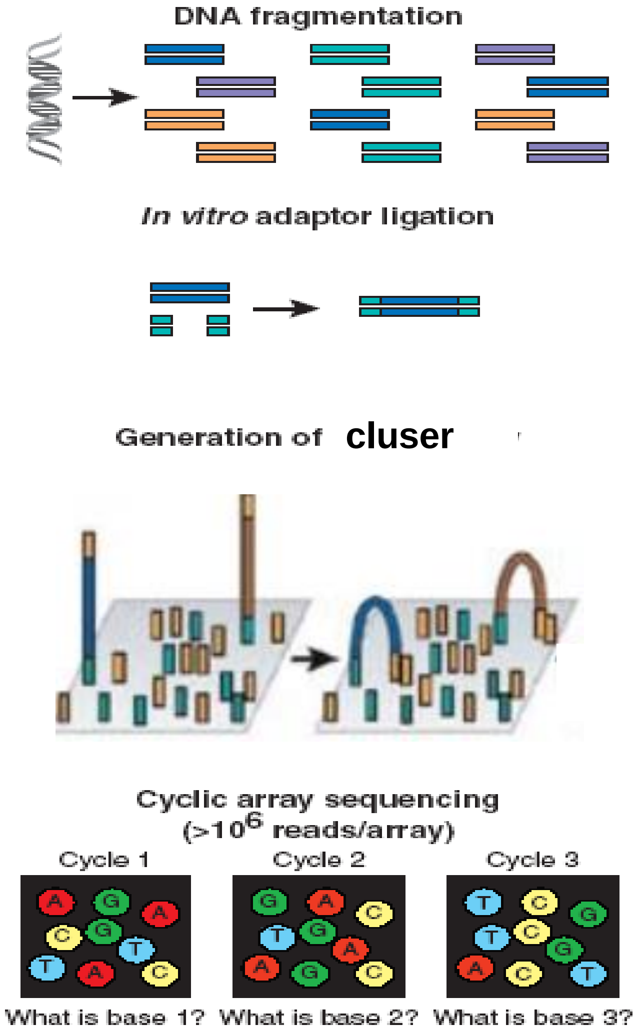
	1990 ~ 2003	2015
技术	Sanger	NGS
费用	50亿美金	几千美金
时间	13 年	3 天
人力	数百人	一人

二代测序-通量更大

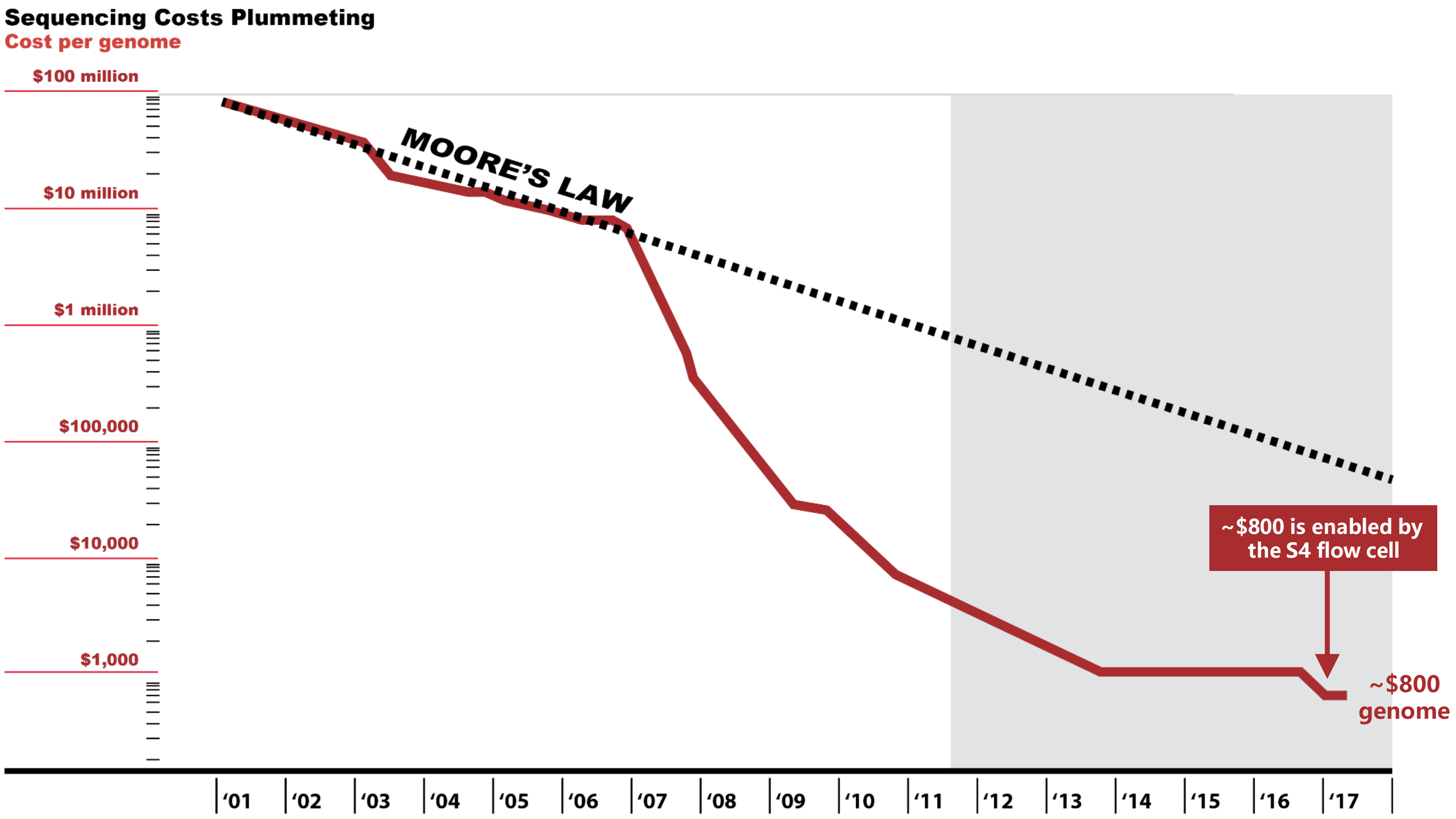
Sanger: 每条DNA序列单独测序



NGS: 多个DNA序列一起测序



二代测序-成本更低



Modified from: Hayden, E. C. (2014, March 19). Technology: The $1,000 genome. Retrieved June 16, 2017, from http://www.nature.com/news/technology-the-1-000-genome-1.14901

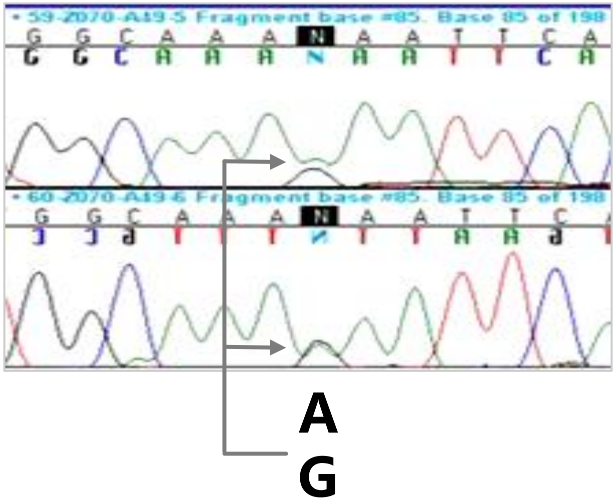
二代测序-灵敏度更高

二代测序能够检测稀有突变

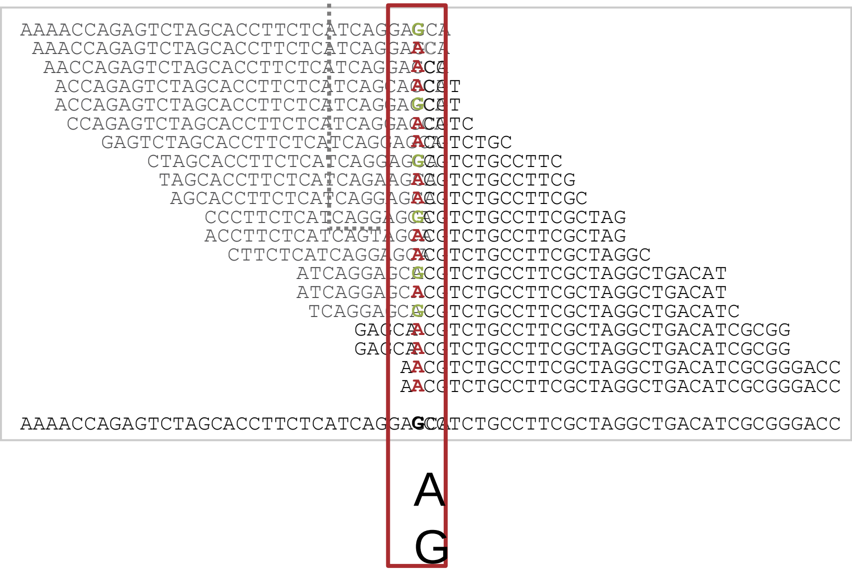
Sanger 测序	二代测序
最低检出限 ~20%	最低检出限 ~ 3-5%

Molecular Pathology Checklist

> “Detection of **20% variant allele frequency**, which is typically equivalent to a 40% proportion of mutation positive cells, is commonly measured as the **LOD for Sanger Sequencing**.”
>
> - College of American Pathologists (CAP)



覆盖度	30X	100X	1,000X	10,000X
A	28 – 29	95 – 97	950 – 970	9500 – 9700
G	1 – 2	3 – 5	30 – 50	300 – 500
最低检出限	~3 – 6%	3 – 5%	3 – 5%	3 – 5%



```
AAAACCAGAGTCTAGCACCTTCTCATCAGGAGCA
 AAACCAGAGTCTAGCACCTTCTCATCAGGAACA
  AACCAGAGTCTAGCACCTTCTCATCAGGAACG
   ACCAGAGTCTAGCACCTTCTCATCAGCAACGT
   ACCAGAGTCTAGCACCTTCTCATCAGGAGCGT
    CCAGAGTCTAGCACCTTCTCATCAGGAACGTC
       GAGTCTAGCACCTTCTCATCAGGAGACGTCTGC
           CTAGCACCTTCTCATCAGGAGGCGTCTGCCTTC
            TAGCACCTTCTCATCAGAAGAACGTCTGCCTTCG
             AGCACCTTCTCATCAGGAGAACGTCTGCCTTCGC
                CCCTTCTCATCAGGAGGCGTCTGCCTTCGCTAG
                ACCTTCTCATCAGTAGAACGTCTGCCTTCGCTAG
                  CTTCTCATCAGGAGAACGTCTGCCTTCGCTAGGC
                        ATCAGGAGCGCGTCTGCCTTCGCTAGGCTGACAT
                        ATCAGGAGCACGTCTGCCTTCGCTAGGCTGACAT
                         TCAGGAGCGCGTCTGCCTTCGCTAGGCTGACATC
                             GAGCAACGTCTGCCTTCGCTAGGCTGACATCGCGG
                             GAGCAACGTCTGCCTTCGCTAGGCTGACATCGCGG
                                 AACGTCTGCCTTCGCTAGGCTGACATCGCGGGACC
                                 AACGTCTGCCTTCGCTAGGCTGACATCGCGGGACC

AAAACCAGAGTCTAGCACCTTCTCATCAGGAGCATCTGCCTTCGCTAGGCTGACATCGCGGGACC
```

A
G

现有技术一览：竞争激烈、更新换代频繁

➢第一代（稳定需求）

- ✓ABI
  - ✓3130xL（停产）
  - ✓3730xL（停产）
  - ✓3500xL

➢第二代（国外竞争激烈、高速发展）

- ✓Roche
  - ✓Genome Sequencer FLX System（2015停产）
  - ✓GS Junior System（2015停产）
- ✓Illumina
  - ✓Genome Analyzer IIx（停产）
  - ✓MiSeq
  - ✓MiSeqDx（获FDA认证）
  - ✓HiSeq 1000/1500/2000（停产）
  - ✓HiSeq 2500
  - ✓HiSeq X Ten
  - ✓NestSeq 500
- ✓Thermo Fisher (Life Technologies) (ABi)
  - ✓5500 SOLiD™ System（停产）
  - ✓5500xL SOLiD™ System（停产）
  - ✓Ion Torrent PGM/Ion Proton/Ion Proton II
- ✓DanaherMotion
  - ✓Polonator G.007（停产）
- ✓MaxSeq（出售至Qiagen）
- ✓Complete Genomics（出售至华大基因）

➢第二代（国内迎头赶上）

- ✓无锡艾吉因生物信息技术有限公司
  - ✓AG-100
- ✓深圳华因康基因科技有限公司
  - ✓Pstar-II
- ✓中科院北京基因组所/半导体所/中科紫鑫
  - ✓BIGIS-1
  - ✓BIGIS-4

➢第三代（技术面临挑战）

- ✓Helicos Biosciences
  - ✓Helicos Genetic Analysis System（停产）
- ✓Pacific Biosciences
  - ✓RS System/RS II
- ✓Oxford Nanopore Technology
  - ✓GridION
  - ✓MinION

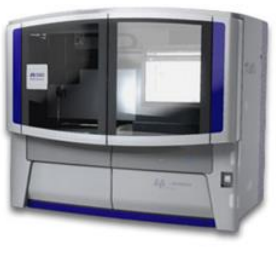
市场现存技术平台


3130xL


Genome Sequencer FLX System


HiSeq 2000/2500


NestSeq 500


5500 SOLiD


Personal Genome Machine

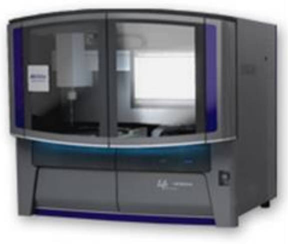

3730xL


MiSeq


MiSeqDX

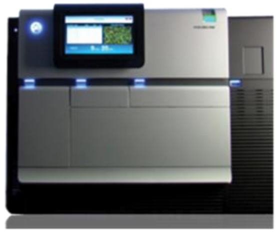

Ion Proton


Genome Sequencer Junior System


5500xL SOLiD


3500xL


HiSeq X Ten

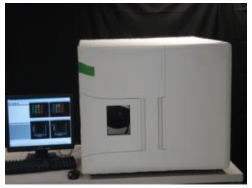

PacBio RS System


GridION & MinION


Polonator G.007


Complete Genomics/BGISEQ


Helicos Genetic Analysis System


MaxSeq

ARTICLE

Non-invasive prenatal measurement of the fetal genome

H. Christina Fan, Wei Gu, Jianbin Wang, Yair J. Blumenfeld, Yasser Y. El-Sayed & Stephen R. Quake

---

A High-Resolution View of Genome-Wide Transformation

Nicholas J. Croucher, Simon R. Harris, Lars Barquist, Julian Parkhill, Stephe

---

ARTICLE

Novel mutations target di subgroups of medulloblast

Transcriptome-wide Regulation of Pre-mRNA Splicing and mRNA Localization by Muscleblind Proteins

Eric T. Wang, Neal A.L. Cody, Sonali Jog, Michela Biancolella, Thomas T. Wang, Daniel J. Tre Gary P. Schroth, David E. Housman, Sita Reddy, Eric Lécuyer, and Christopher B. Burge

---

Human Molecular Genetics, 2012, Vol. 21, No. 11

Penetrance of biallelic *SMARCAL1* mutations is associated with environmental and genetic disturbances of gene expression

---

Immunology REVIEW ARTICLE

Rep-Seq: uncovering the immunological repertoire through next-generation sequencing

Jennifer Benichou, Rotem Ben-Hamo, Yoram Louzoun and Sol Efroni

---

Recurrent Hemizygous Deletions in Cancers May Optimize Proliferative Potential

---

OPEN ACCESS Freely available online PLoS PATHOGENS

Synergistic Parasite-Pathogen Interactions Mediated by Host Immunity Can Drive the Collapse of Honeybee Colonies

---

LETTER

Sequence analysis of mutations and translocations across breast cancer subtypes

---

frontiers in GENETICS TECHNOLOGY REPORT ARTICLE

Single-neuron RNA-Seq: technical feasibility and reproducibility

---

Cell

A Landscape of Driver Mutations in Melanoma

Sciencexpress Reports

Landscape of Somatic Retrotransposition in Human Cancers

Full-length mRNA-Seq from single-cel and individual circulating tumor cells

---

Infection, Genetics and Evolution

Isolation and molecular identification of Sunshine virus, a novel paramyxovirus found in Australian snakes

---

ORIGINAL ARTICLE

Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill

Olivia U Mason, Terry C Hazen, Sharon Borglin, Patrick SG Chain, Eric A Dubinsky, Julian L Fortney, James Han, Hoi-Ying N Holman, Jenni Hultman, Regina Lamendella, Rachel Mackelprang, Stephanie Malfatti, Lauren M Tom, Susannah G Tringe, Tanja Woyke, Jizhong Zhou, Edward M Rubin and Janet K Jansson

PART2

two

2

测序仪原理

高通量测序流程

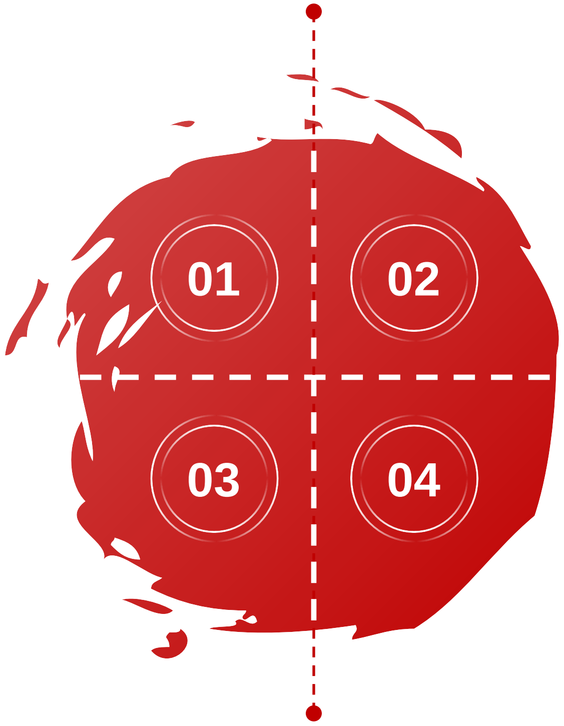
核酸提取

文库制备

测序

数据分析

样品前处理



样品采集

土壤、粪便或植物等样品参照采样标准，多点采样，混匀后装入无菌袋中，做好标记。

保存与运输

应立即置于干冰或者-80℃保存，保存及运输过程避免反复冻融。如需常温运输，宜将样品浸润与稳定剂中。

核酸提取

Qiagen，MP提取试剂盒提取样本基因组DNA。

核酸定量

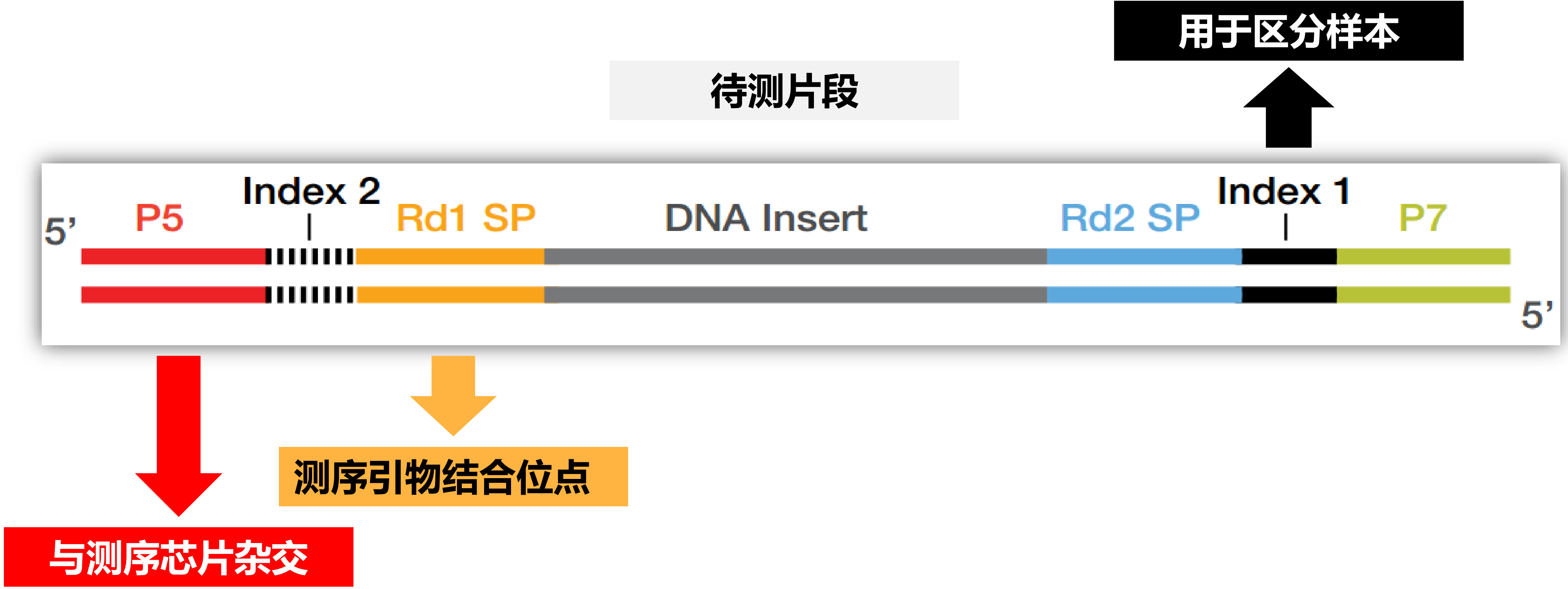
Quantify: Qubit, Nanodrop；

OD260/280: 1.8~2.0；

DNA质量对于二代测序至关重要！

文库制备

Illumina文库



用于区分样本

待测片段

测序引物结合位点

与测序芯片杂交

文库制备的目的：待测片段两侧连接测序接头

文库构建流程

核酸提取 (DNA 或 RNA)

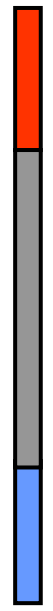
核酸片段化

添加测序接头

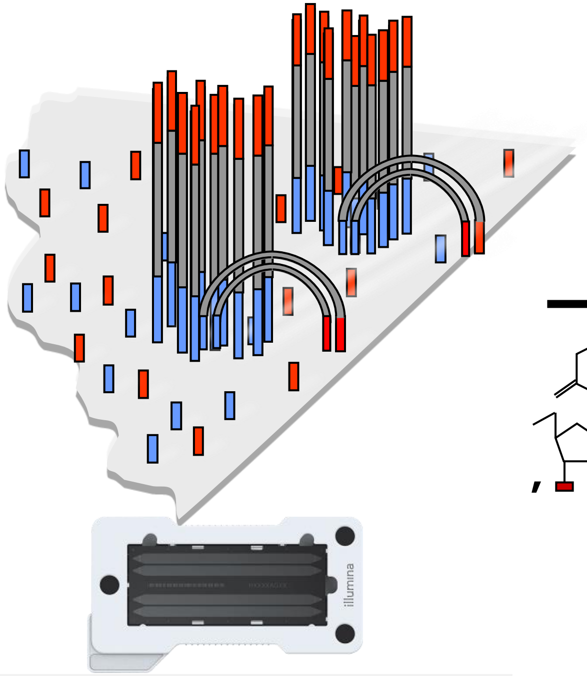
Illumina 测序原理

边合成边测序（SBS）

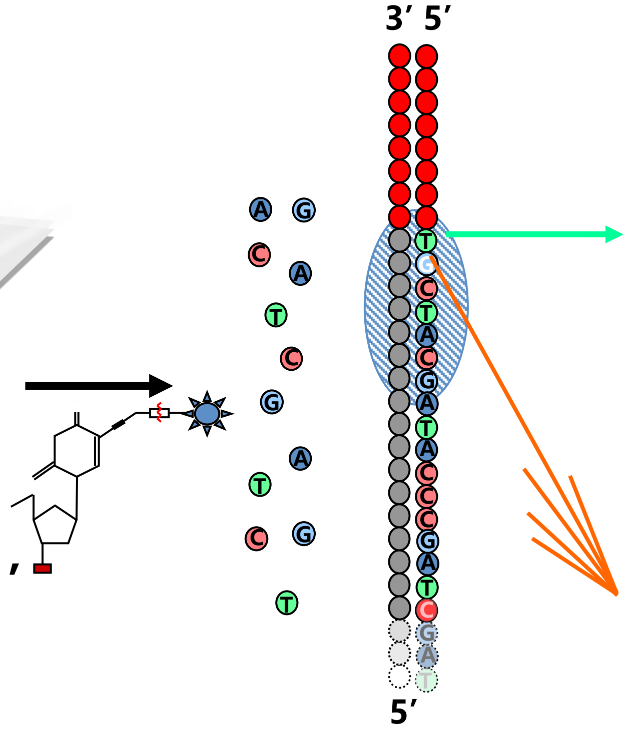
DNA
(0.1-1.0 ug)



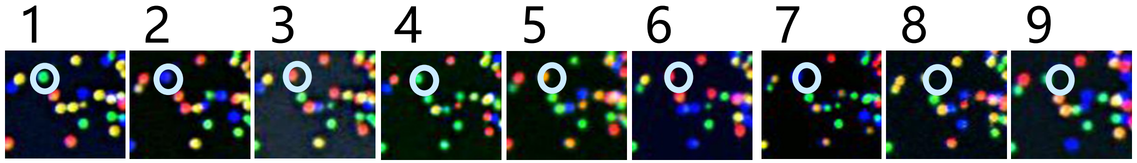
样本制备



簇生成



边合成边测序



→ TGCTACGAT…

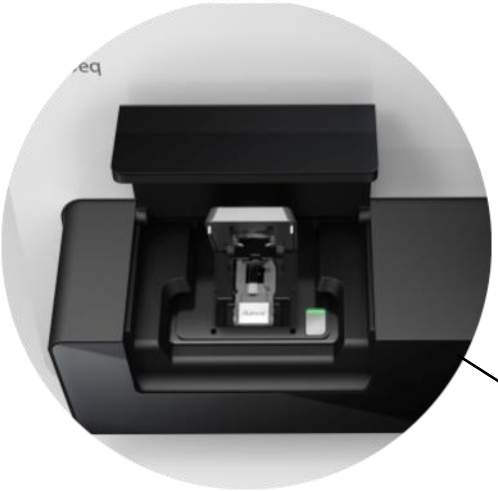
碱基读取

成像及数据处理

MiSeq® 测序仪

PART3

MiSeq 简单的操作



- 预装试剂
- 自动进样
- 整合了簇生成和测序工作

MiSeq 提供灵活的测序读长

25
Million Reads

15
Million Reads

4
Million Reads

1
Million Reads

MiSeq Core Consumables Version 3

- 600 cycles
- 150 cycles

MiSeq Core Consumables Version 2

- 500 cycles
- 300 cycles
- 50 cycles
- 36 cycles

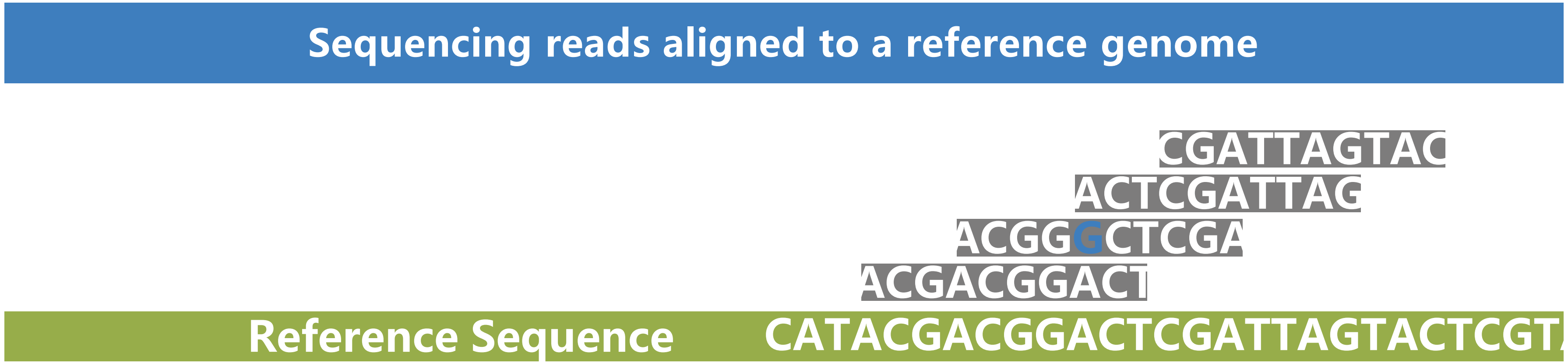
MiSeq Core Consumables Version 2 Micro

- 300 cycles (Micro)

MiSeq Core Consumables Version 2 Nano

- 500 cycles (Nano)
- 300 cycles (Nano)

数据分析



Sequencing reads aligned to a reference genome

CGATTAGTAC
ACTCGATTAG
ACGGGCTCGA
ACGACGGACT

Reference Sequence CATACGACGGACTCGATTAGTACTCGT

PART4

4

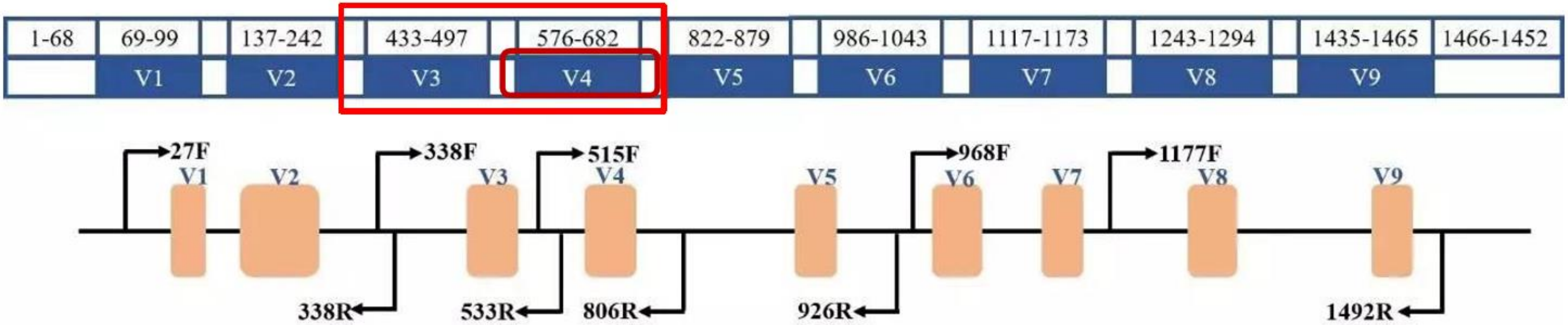
Miseq测序仪的应用

Miseq测序仪的应用

MiSeq应用

16S Metagenomics Sequencing Workflow

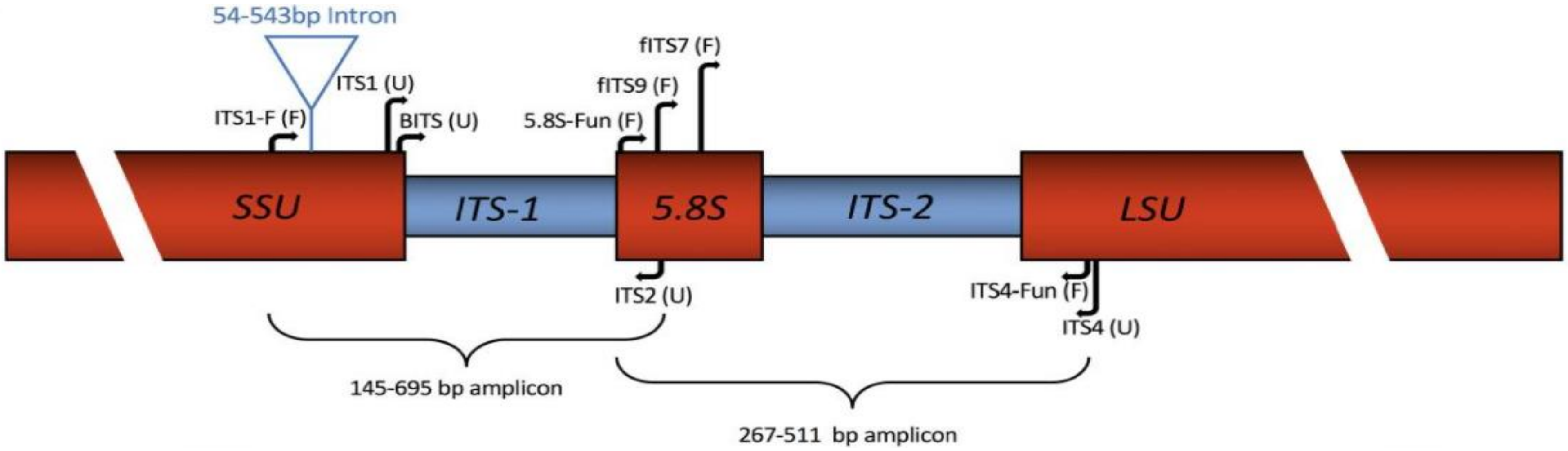
- 16S Metagenomics
- – Study of genetic material recovered directly from environmental samples
- – Uses conserved region of 16S Ribosomal RNA gene to PCR amplify DNA samples



MiSeq应用

Fungal sequencing and classification with the ITS Metagenomics

- ITS Metagenomics
- – fungal microbiome, is an important component in the environment and in human health.
- – Uses conserved region of ITSI and ITS2 to PCR amplify DNA samples .



Step 1: Order Amplicon Primers

Primers target 16S rRNA gene V3-V4 region

- A single amplicon of approximately 460bp is created
- Primers include overhang adaptor sequences
- Illumina recommends using standard desalting purification when ordering oligo primer sets

Forward Primer:

5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG

Overhang Adapter Sequence: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG

Locus-Specific Sequence *16S V3-V4*: CCTACGGGNGGCWGCAG

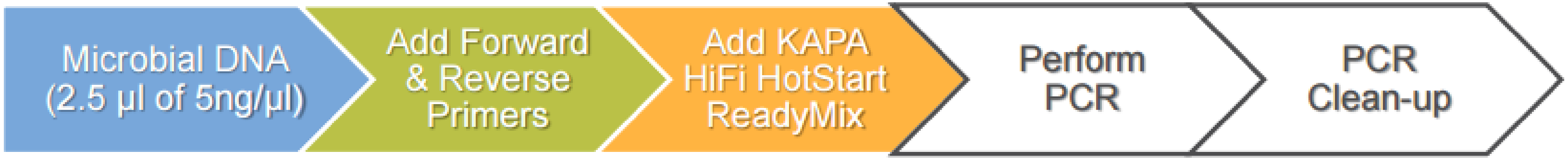
Reverse Primer:

5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC

Overhang Adapter Sequence: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

Locus-Specific Sequence *16S V3-V4*: GACTACHVGGGTATCTAATCC

Step2: PCR Amplify V3 and V4 regions of 16S rRNA gene



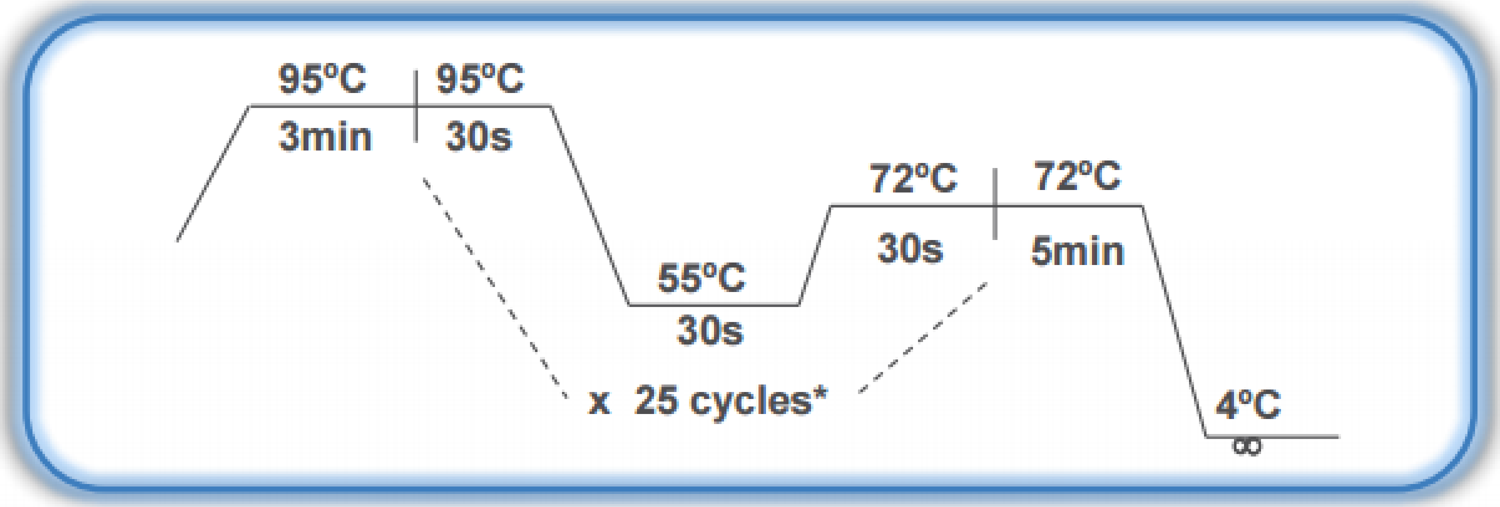
Microbial DNA (2.5 µl of 5ng/µl)

Add Forward & Reverse Primers

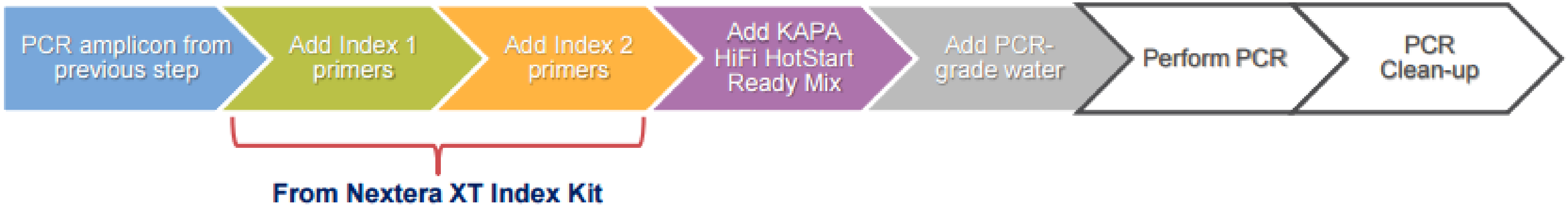
Add KAPA HiFi HotStart ReadyMix

Perform PCR

PCR Clean-up



Step3: 2nd PCR to Add Indices and Adaptors



PCR amplicon from previous step

Add Index 1 primers

Add Index 2 primers

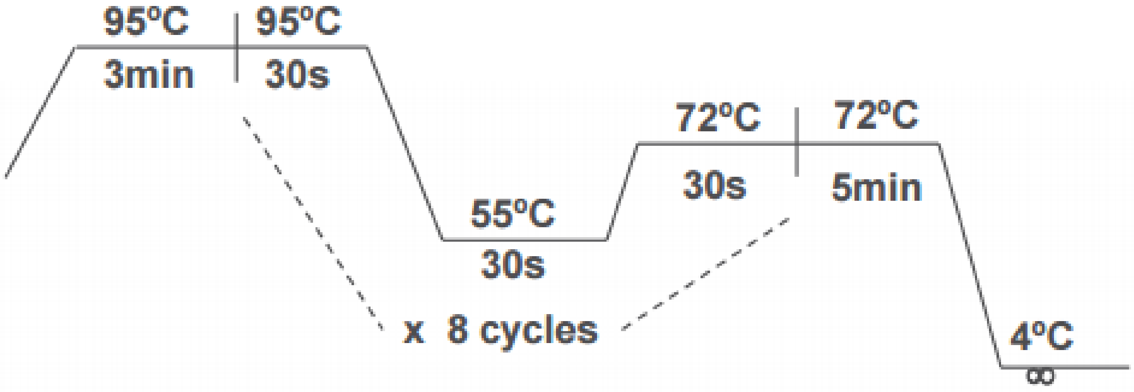
Add KAPA HiFi HotStart Ready Mix

Add PCR-grade water

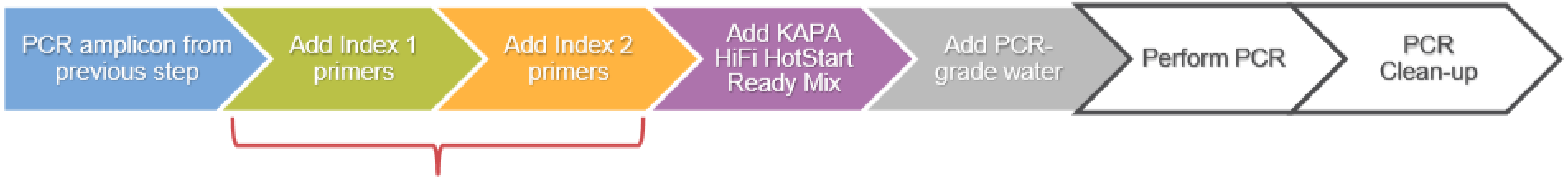
Perform PCR

PCR Clean-up

From Nextera XT Index Kit



Step3: 2nd PCR to Add Indices and Adaptors



PCR amplicon from previous step

Add Index 1 primers

Add Index 2 primers

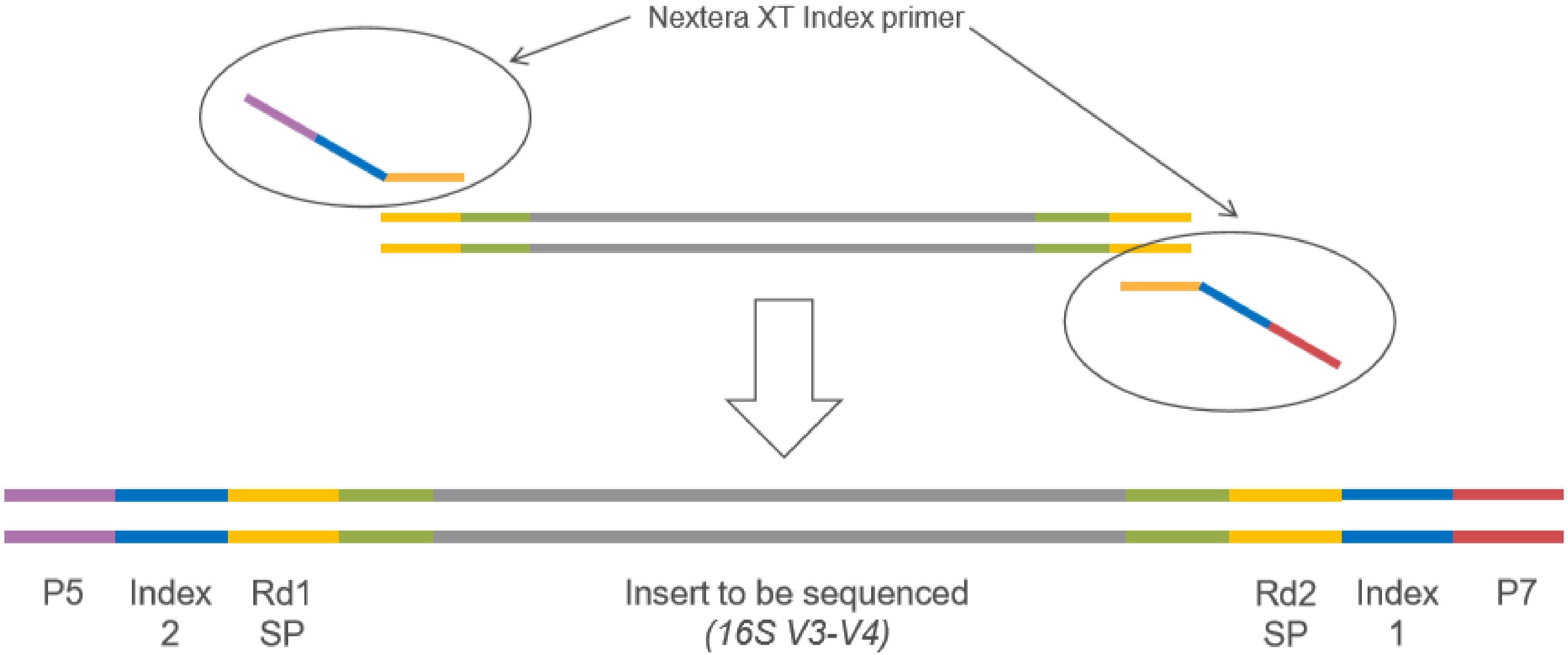
Add KAPA HiFi HotStart Ready Mix

Add PCR-grade water

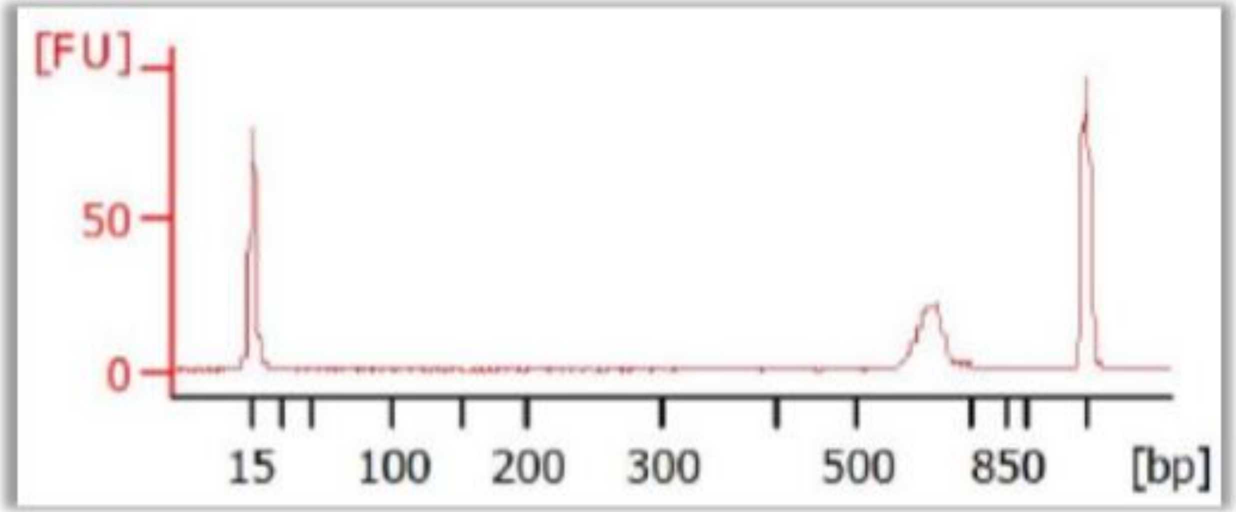
Perform PCR

PCR Clean-up

From Nextera XT Index Kit



Step4:Library QC



Example Bioanalyzer Trace of Final Library

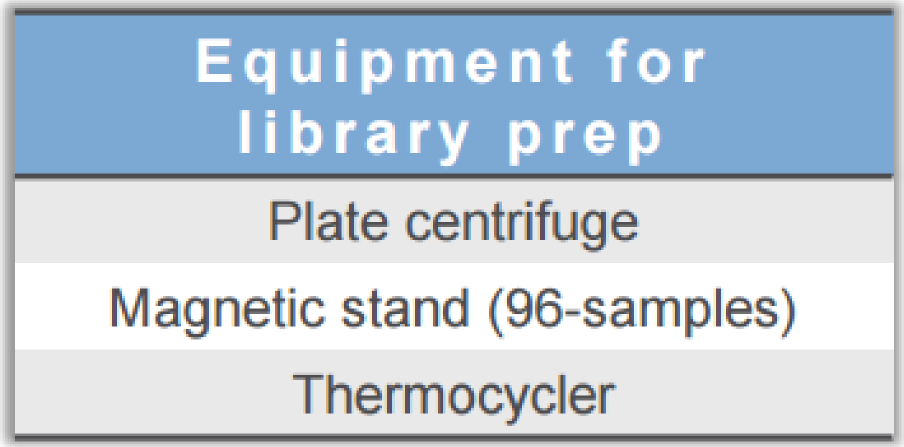
Quantity

- Use dsDNA-specific fluorescent dye (Qubit or PicoGreen)
- Optimal cluster density: 800-1000k/mm^2

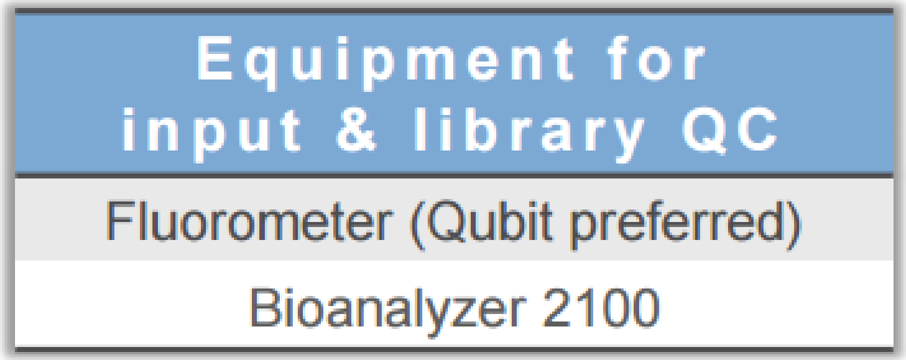
Quality

- Run on Agilent Bioanalyzer DNA 1000 chip
- Expected library size ≈ 630bp

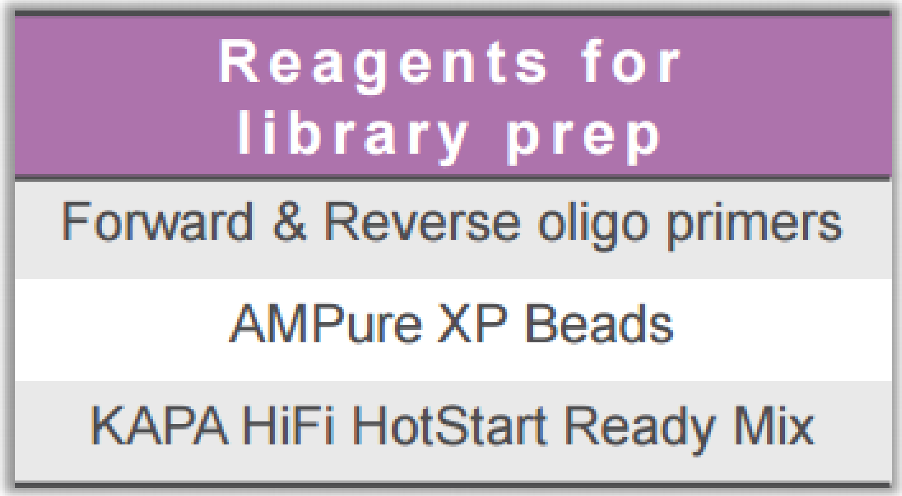
16S Metagenomics需准备的仪器与耗材



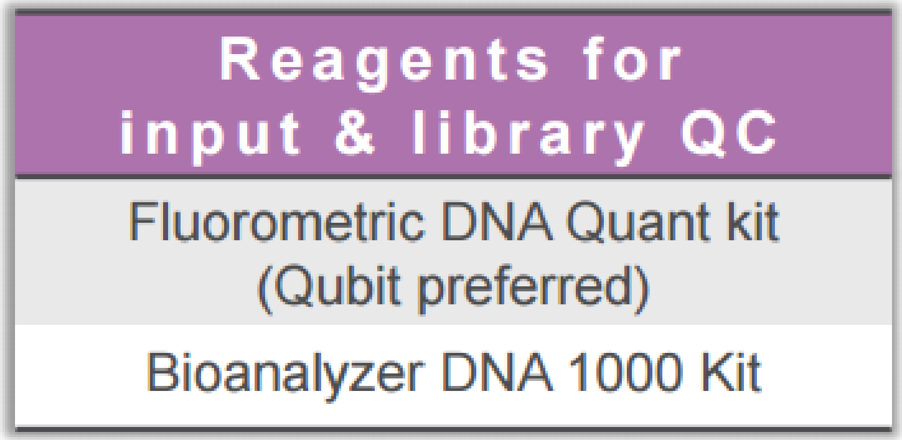
Equipment for library prep
Plate centrifuge
Magnetic stand (96-samples)
Thermocycler



Equipment for input & library QC
Fluorometer (Qubit preferred)
Bioanalyzer 2100



Reagents for library prep
Forward & Reverse oligo primers
AMPure XP Beads
KAPA HiFi HotStart Ready Mix



Reagents for input & library QC
Fluorometric DNA Quant kit (Qubit preferred)
Bioanalyzer DNA 1000 Kit

Step5: Library Pooling

Library Pooling Considerations

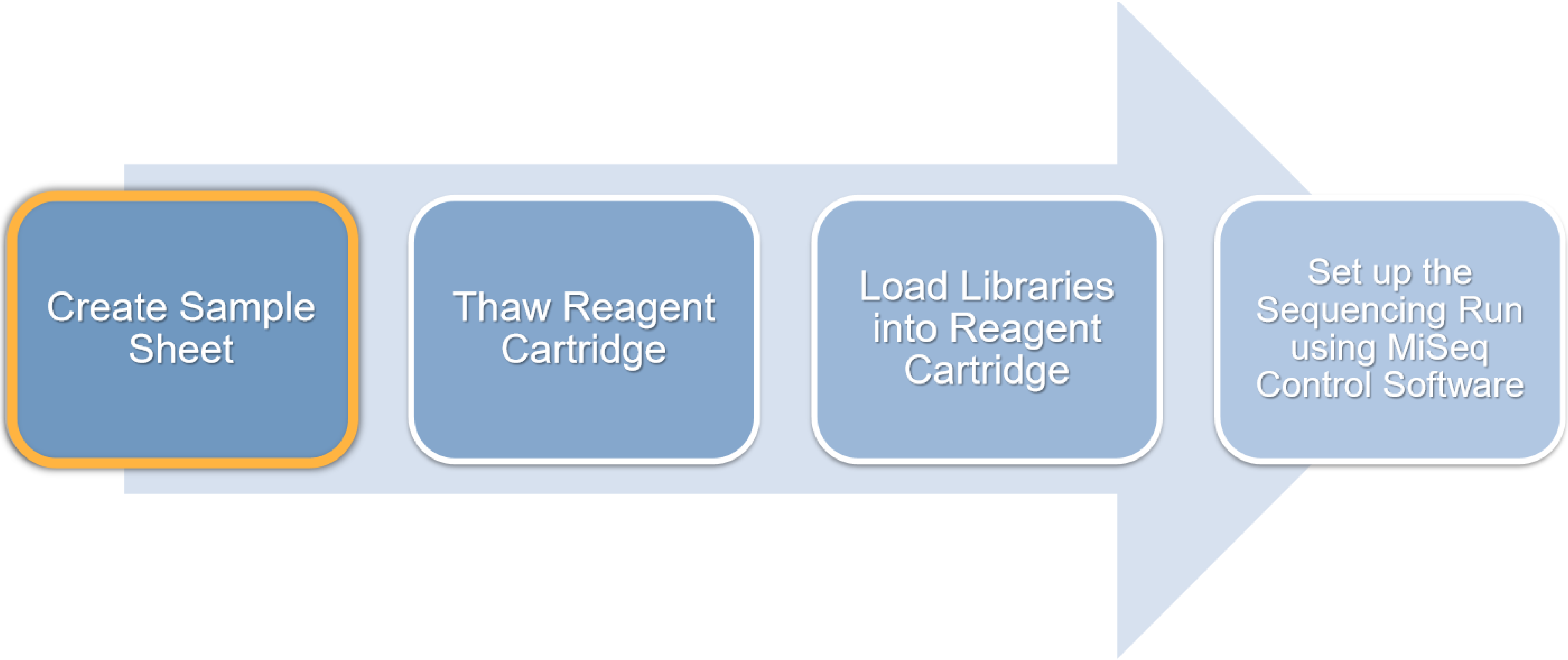
- Pool a maximum of 384 sample libraries per MiSeq v3 run
- Total MiSeq v3 run output is > 20 million reads
- This will generate > 30,000 reads per sample which is commonly recognized to be sufficient for metagenomic surveys

Sequencing Considerations

- Optimal raw cluster density is 800-1000k/mm^2
- Cluster densities below 500k/mm^2 and above 1000k/mm^2 is not recommended
- Spike-in minimum 5% PhiX library to serve as an internal control for low-diversity library

Sequence on MiSeq

How to Start a Run Workflow



1. Create Sample Sheet
2. Thaw Reagent Cartridge
3. Load Libraries into Reagent Cartridge
4. Set up the Sequencing Run using MiSeq Control Software

Create Sample Sheet

Thaw Reagent Cartridge

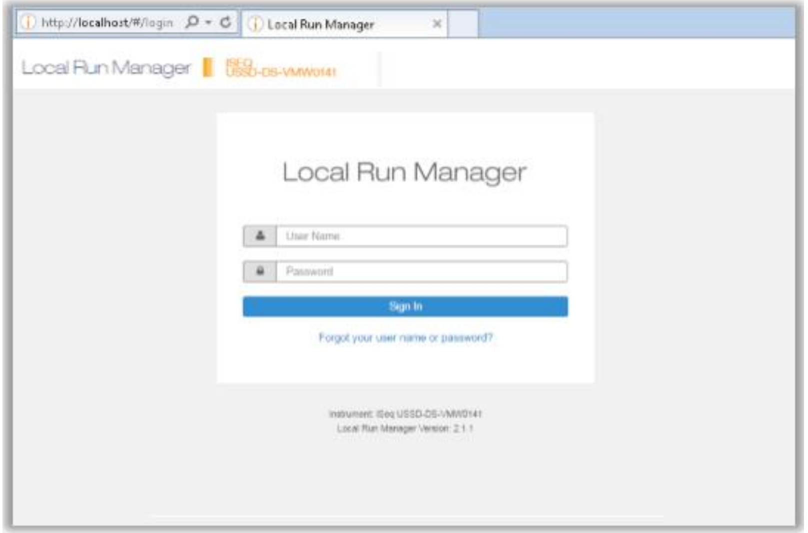
Load Libraries into Reagent Cartridge

Set up the Sequencing Run using MiSeq Control Software

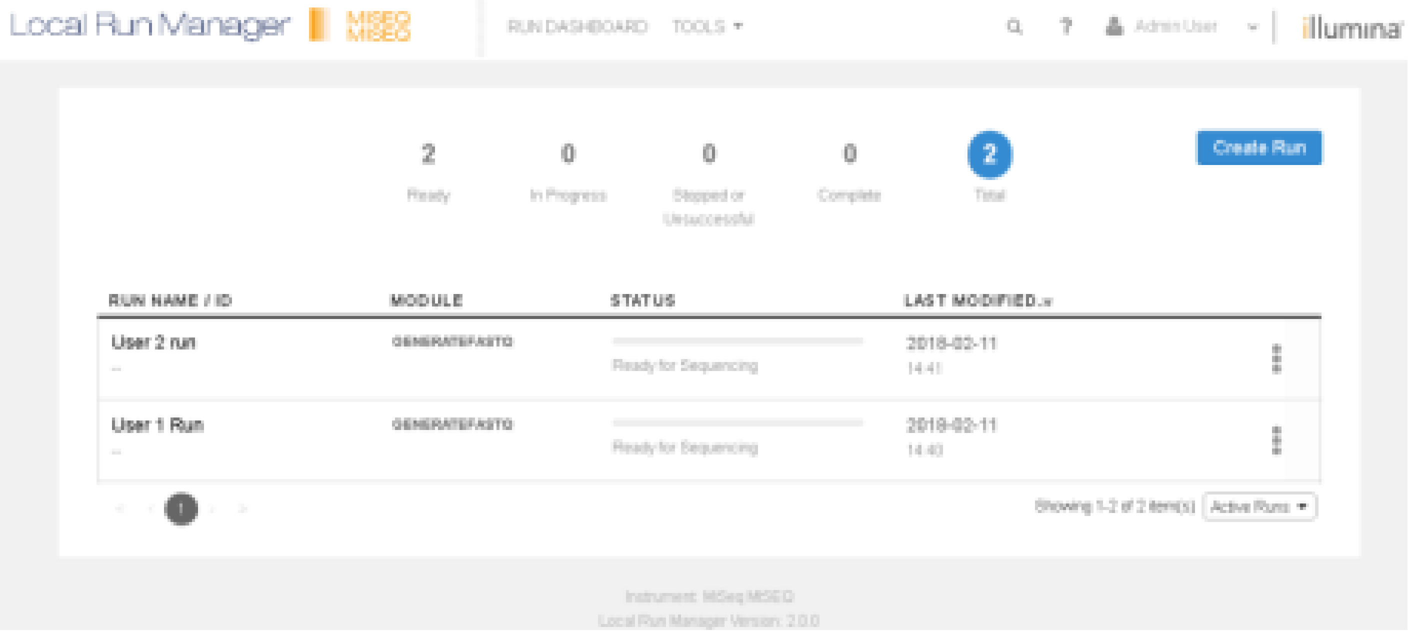
Create sample sheet

- 创建run,我们会用到Local Run Manager 软件

❓
确定要做多少个样本
哪些文库混合在一起
读长及数据量是多少



Instrument: iSeq USSD-DS-VMW0141
Local Run Manager Version: 2.1.1



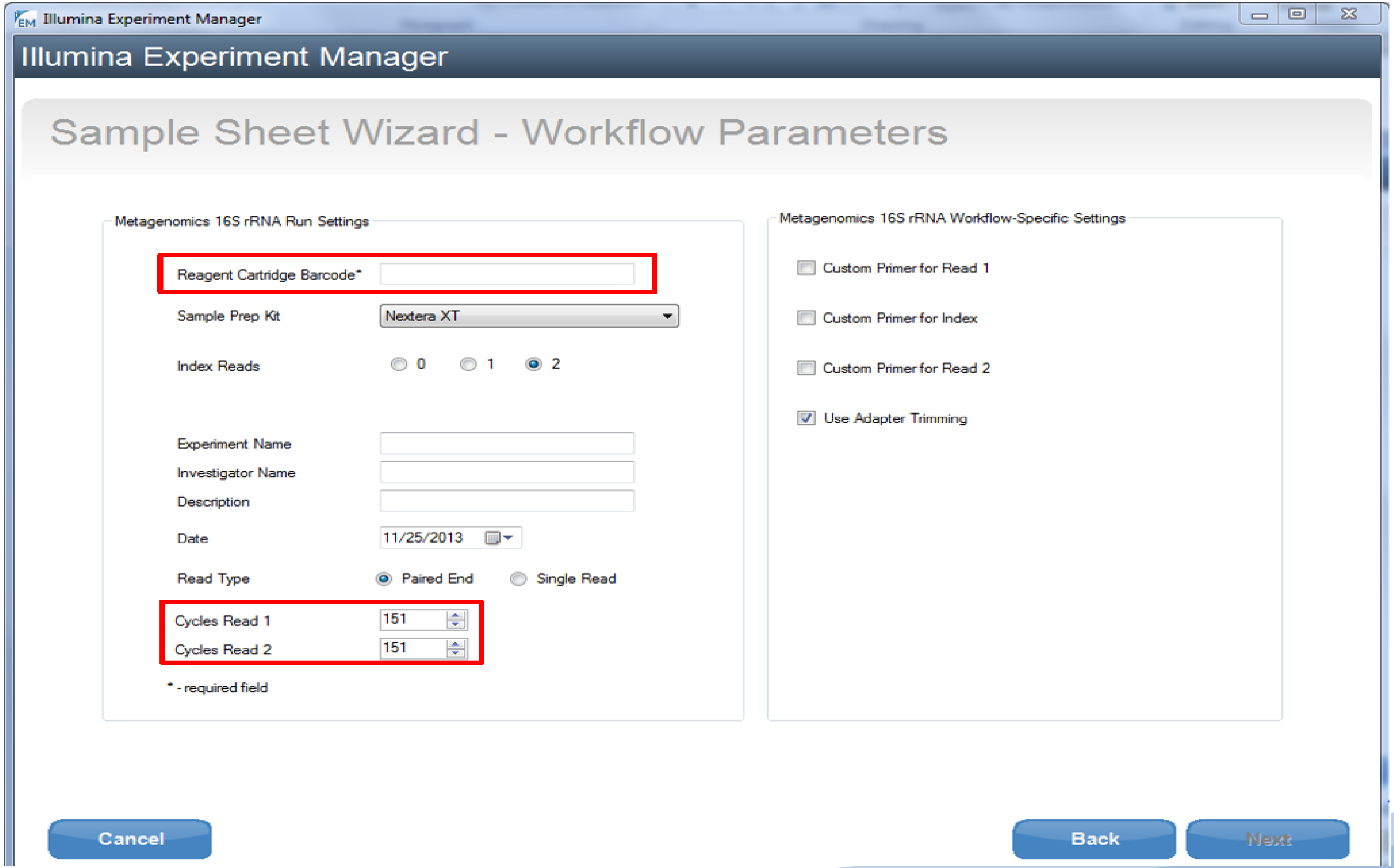
2	0	0	0	2
Ready	In Progress	Stopped or Unsuccessful	Complete	Total

RUN NAME / ID	MODULE	STATUS	LAST MODIFIED
User 2 run --	GENERATEFASTQ	Ready for Sequencing	2018-02-11 14:41
User 1 Run --	GENERATEFASTQ	Ready for Sequencing	2018-02-11 14:40

Showing 1-2 of 2 item(s) Active Runs

Instrument: MiSeq MISEQ
Local Run Manager Version: 2.0.0

Workflow Parameters



Sample Sheet Wizard - Workflow Parameters

Metagenomics 16S rRNA Run Settings

Reagent Cartridge Barcode*

Sample Prep Kit: Nextera XT

Index Reads ○ 0 ○ 1 ● 2

Experiment Name

Investigator Name

Description

Date: 11/25/2013

Read Type ● Paired End ○ Single Read

Cycles Read 1: 151

Cycles Read 2: 151

* - required field

Metagenomics 16S rRNA Workflow-Specific Settings

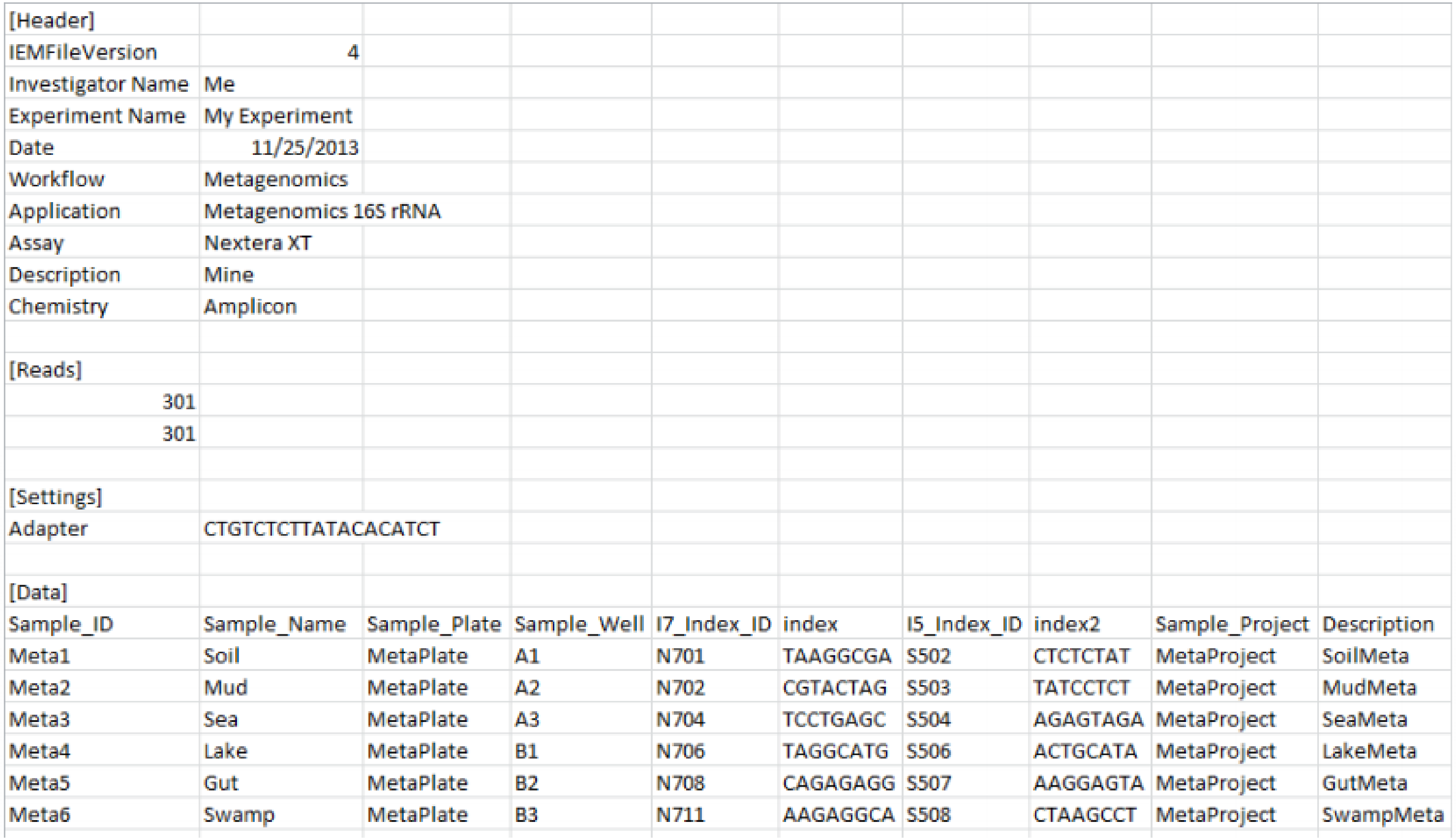
☐ Custom Primer for Read 1

☐ Custom Primer for Index

☐ Custom Primer for Read 2

☑ Use Adapter Trimming

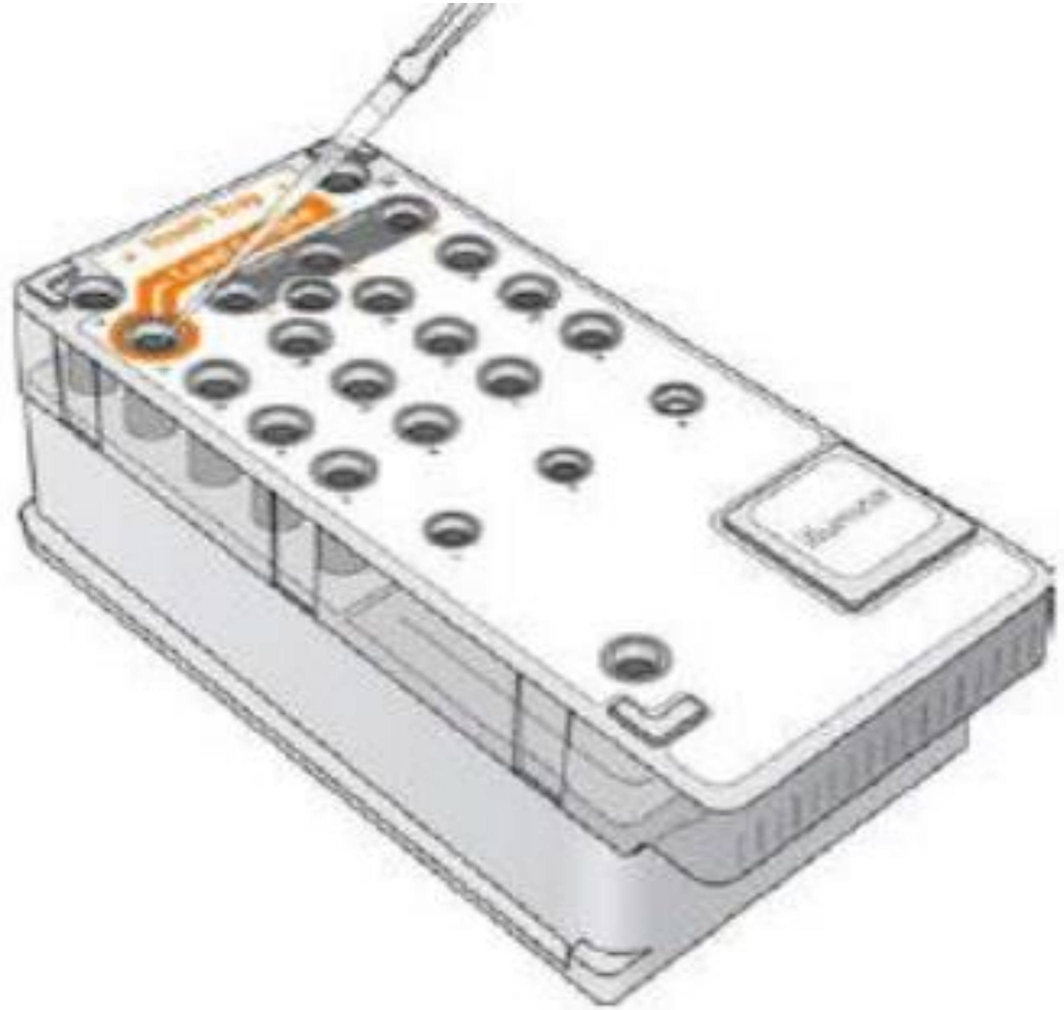
Sample sheet



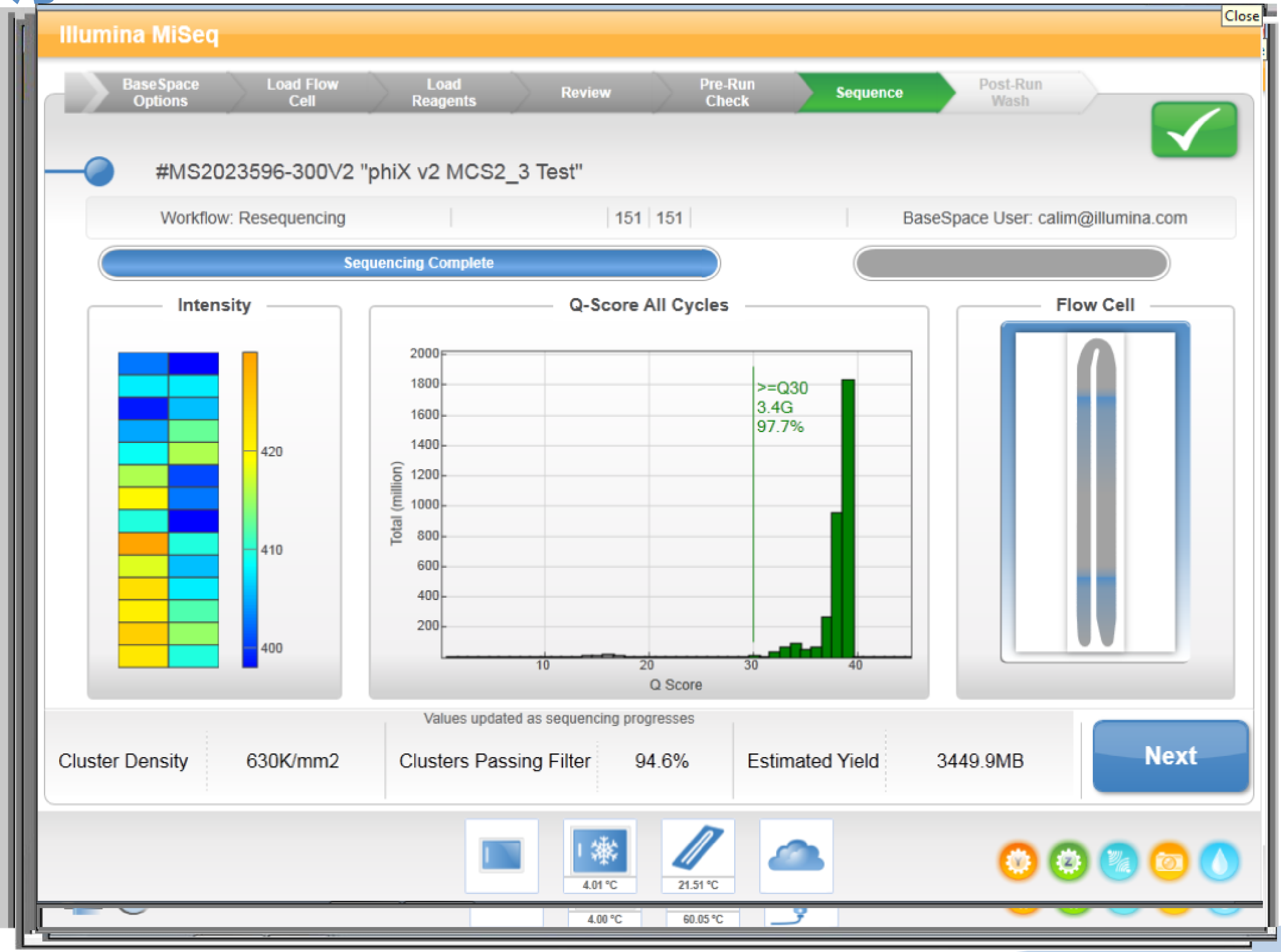
[Header]									
IEMFileVersion	4								
Investigator Name	Me								
Experiment Name	My Experiment								
Date	11/25/2013								
Workflow	Metagenomics								
Application	Metagenomics 16S rRNA								
Assay	Nextera XT								
Description	Mine								
Chemistry	Amplicon								
[Reads]									
301									
301									
[Settings]									
Adapter	CTGTCTCTTATACACATCT								
[Data]									
Sample_ID	Sample_Name	Sample_Plate	Sample_Well	I7_Index_ID	index	I5_Index_ID	index2	Sample_Project	Description
Meta1	Soil	MetaPlate	A1	N701	TAAGGCGA	S502	CTCTCTAT	MetaProject	SoilMeta
Meta2	Mud	MetaPlate	A2	N702	CGTACTAG	S503	TATCCTCT	MetaProject	MudMeta
Meta3	Sea	MetaPlate	A3	N704	TCCTGAGC	S504	AGAGTAGA	MetaProject	SeaMeta
Meta4	Lake	MetaPlate	B1	N706	TAGGCATG	S506	ACTGCATA	MetaProject	LakeMeta
Meta5	Gut	MetaPlate	B2	N708	CAGAGAGG	S507	AAGGAGTA	MetaProject	GutMeta
Meta6	Swamp	MetaPlate	B3	N711	AAGAGGCA	S508	CTAAGCCT	MetaProject	SwampMeta

Miseq试剂盒准备

- 放到室温的水槽里融化60-90min，确保所有试剂都融化，注意不要超过最低水位线。
- 手动翻转试剂盒10次。
- 拍试剂槽移除气泡
- 融化后，在上机前一直放在冰上保存。



上机测序



#MS2023596-300V2 "phiX v2 MCS2_3 Test"

Workflow: Resequencing | 151 | 151 | BaseSpace User: calim@illumina.com

Sequencing Complete

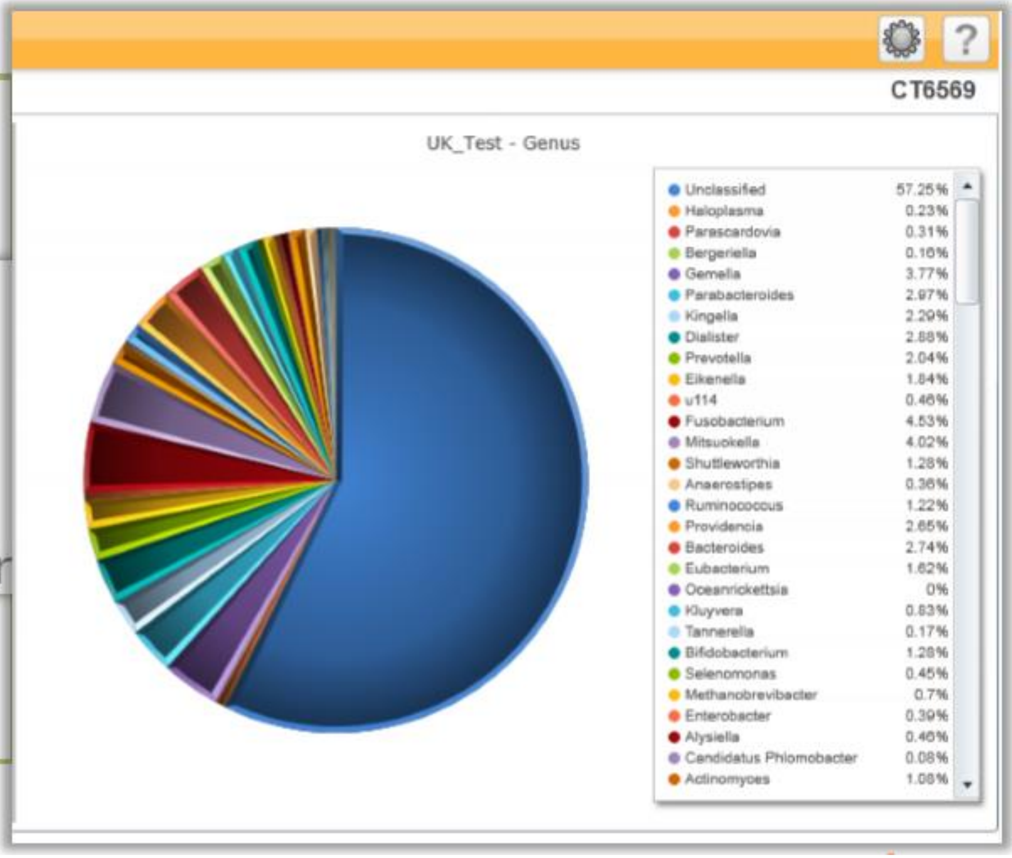
Values updated as sequencing progresses

Cluster Density	630K/mm2	Clusters Passing Filter	94.6%	Estimated Yield	3449.9MB

16S Metagenomics Data Analysis

MiSeq Reporter

- On-instrument software
- Demultiplex indices/samples
- Aligns reads and classifies organisms



Unclassified	57.25%
Haloplasma	0.23%
Parascardovia	0.31%
Bergeriella	0.16%
Gemella	3.77%
Parabacteroides	2.97%
Kingella	2.29%
Dialister	2.88%
Prevotella	2.04%
Eikenella	1.84%
u114	0.46%
Fusobacterium	4.53%
Mitsuokella	4.02%
Shuttleworthia	1.28%
Anaerostipes	0.36%
Ruminococcus	1.22%
Providencia	2.65%
Bacteroides	2.74%
Eubacterium	1.62%
Oceanrickettsia	0%
Kluyvera	0.83%
Tannerella	0.17%
Bifidobacterium	1.28%
Selenomonas	0.45%
Methanobrevibacter	0.7%
Enterobacter	0.39%
Alysiella	0.46%
Candidatus Phlomobacter	0.08%
Actinomyces	1.08%

16S Metagenomics Data Analysis

- **16S分析的主流软件**

- QIIME2: https://qiime2.org

- 将多个程序整合到Python平台
- 现有程序的包装
- 可扩展性
- 命令行模式，工作流程更灵活

- Mothur: https://www.mothur.org

- 单个程序
- 易于安装设置
- 不太可扩展性
- 工作流程适合内部工具

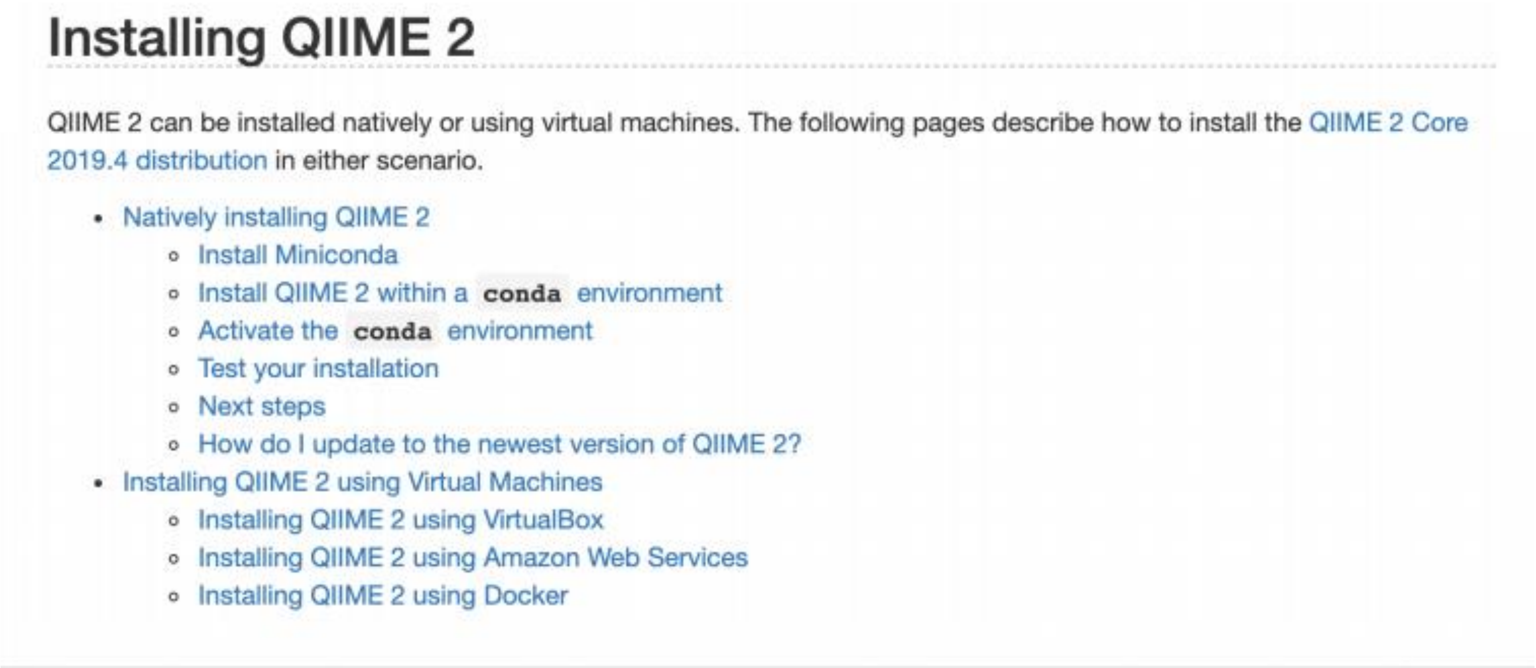
Qiime2特点

1. 容易安装，提供多种安装方法

□ Miniconda软件包管理器（Lincx或Mac平台，官方推荐）

□ Dorker方式安装（Linux需要管理员权限）

□ Virtualbox虚拟机安装（windows 7及以上）



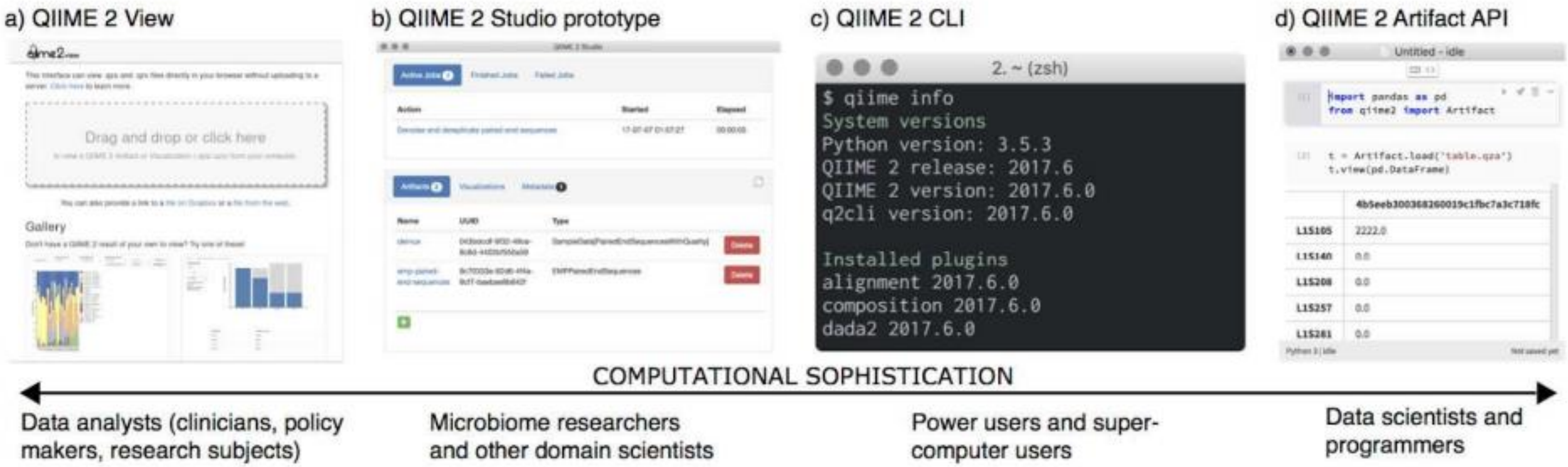
Installing QIIME 2

QIIME 2 can be installed natively or using virtual machines. The following pages describe how to install the QIIME 2 Core 2019.4 distribution in either scenario.

- Natively installing QIIME 2
  - Install Miniconda
  - Install QIIME 2 within a `conda` environment
  - Activate the `conda` environment
  - Test your installation
  - Next steps
  - How do I update to the newest version of QIIME 2?
- Installing QIIME 2 using Virtual Machines
  - Installing QIIME 2 using VirtualBox
  - Installing QIIME 2 using Amazon Web Services
  - Installing QIIME 2 using Docker

Qiime2特点

2. 多种使用界面



```
$ qiime info
System versions
Python version: 3.5.3
QIIME 2 release: 2017.6
QIIME 2 version: 2017.6.0
q2cli version: 2017.6.0

Installed plugins
alignment 2017.6.0
composition 2017.6.0
dada2 2017.6.0
```

```python
import pandas as pd
from qiime2 import Artifact
```

```python
t = Artifact.load('table.qza')
t.view(pd.DataFrame)
```

```
        4b5eeb300368260019c1fbc7a3c718fc
L1S105  2222.0
L1S140  0.0
L1S208  0.0
L1S257  0.0
L1S281  0.0
```

a. QIIME2 View：查看数据或结果的网页工具，无需安装软件，交互式可视化结果；

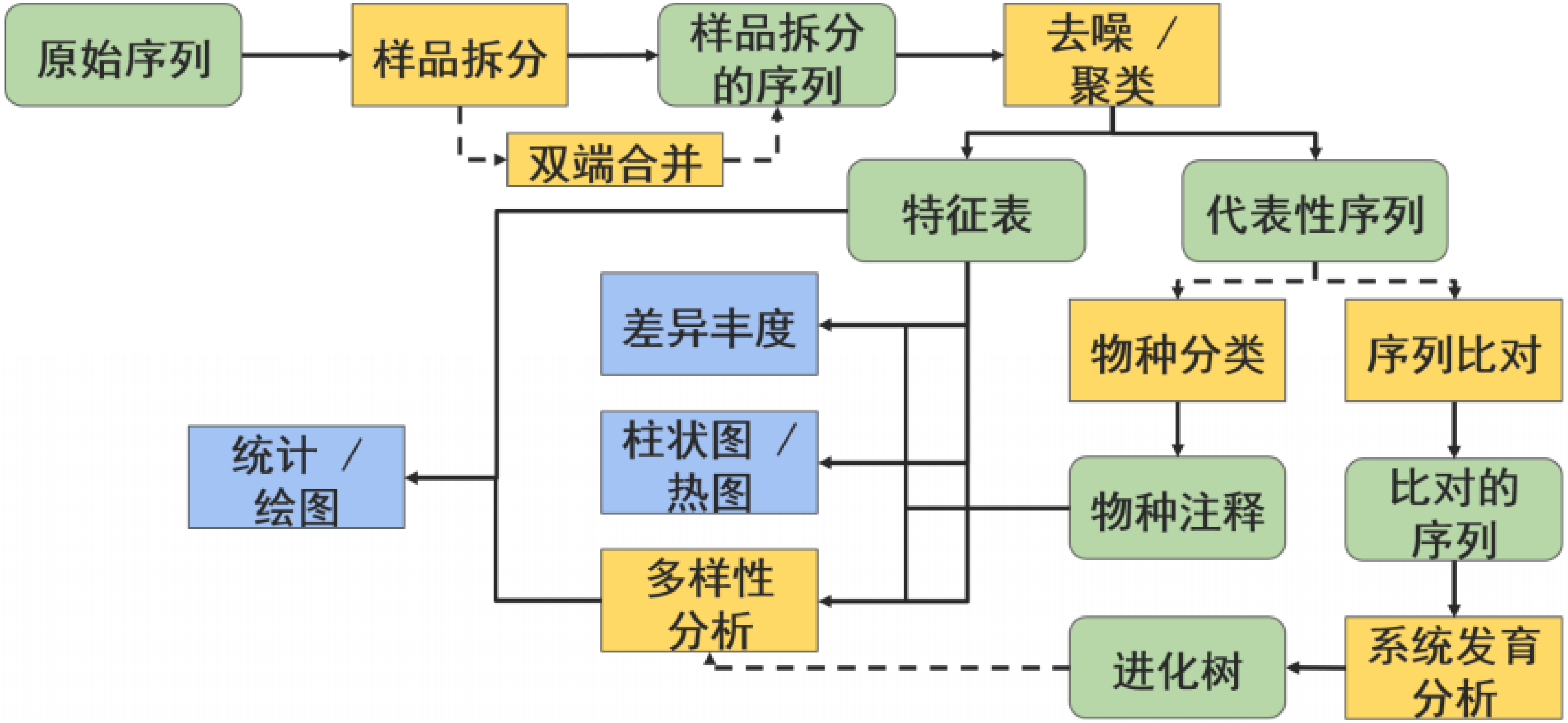
b. 图型界面QIIME 2 Studio：无需命令行或编程技巧；

c. 命令行界面 q2cli；熟悉Linux命令行，计算集群使用的用户

d. Python 3接口的artifact API：数据科学家、程序员；

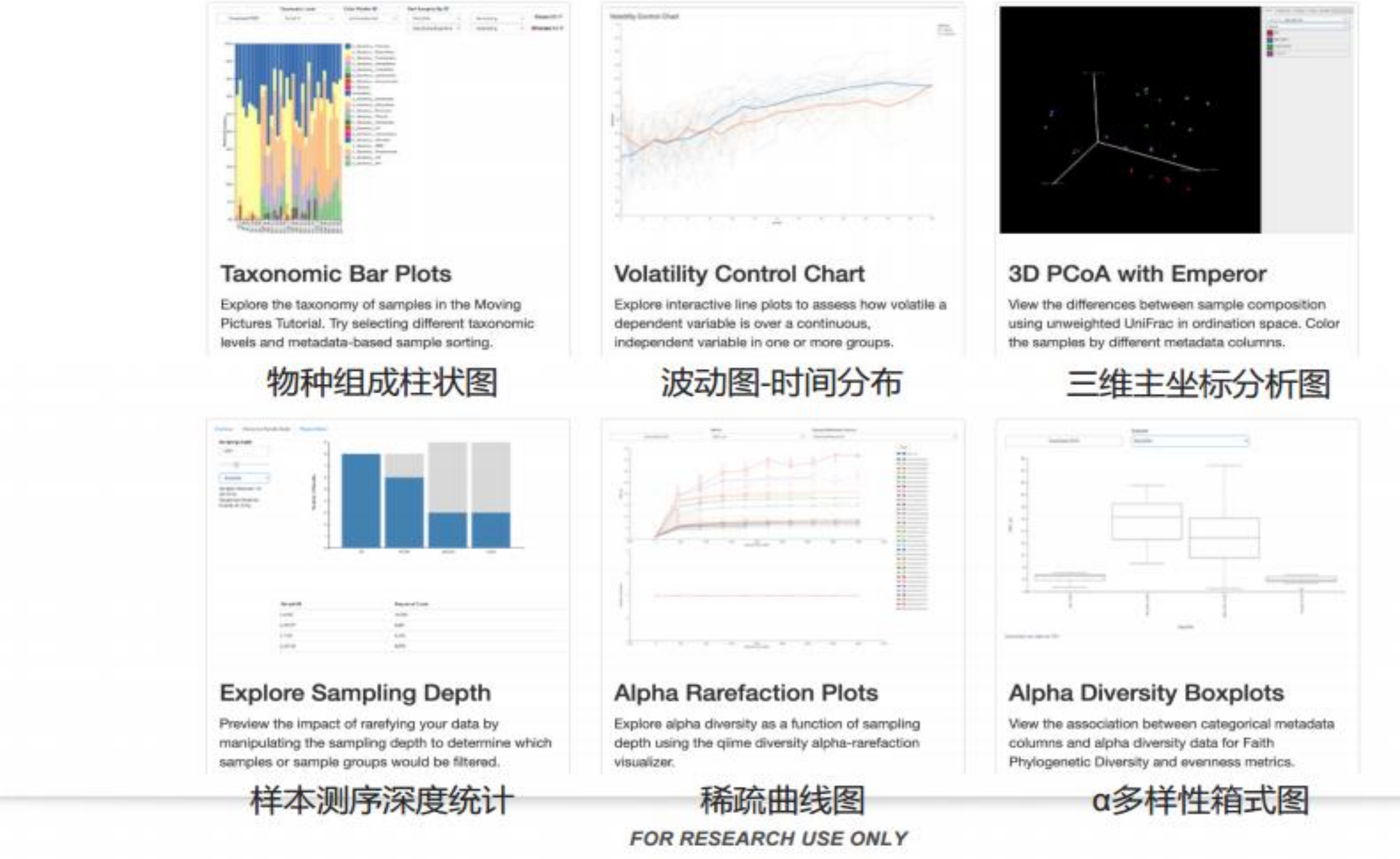
Qiime2特点

3. 分析流程化



Qiime2特点

4. 结果可视化（https://view.qiime2.org/）



Taxonomic Bar Plots

Explore the taxonomy of samples in the Moving Pictures Tutorial. Try selecting different taxonomic levels and metadata-based sample sorting.

物种组成柱状图

Volatility Control Chart

Explore interactive line plots to assess how volatile a dependent variable is over a continuous, independent variable in one or more groups.

波动图-时间分布

3D PCoA with Emperor

View the differences between sample composition using unweighted UniFrac in ordination space. Color the samples by different metadata columns.

三维主坐标分析图

Explore Sampling Depth

Preview the impact of rarefying your data by manipulating the sampling depth to determine which samples or sample groups would be filtered.

样本测序深度统计

Alpha Rarefaction Plots

Explore alpha diversity as a function of sampling depth using the qiime diversity alpha-rarefaction visualizer.

稀疏曲线图

Alpha Diversity Boxplots

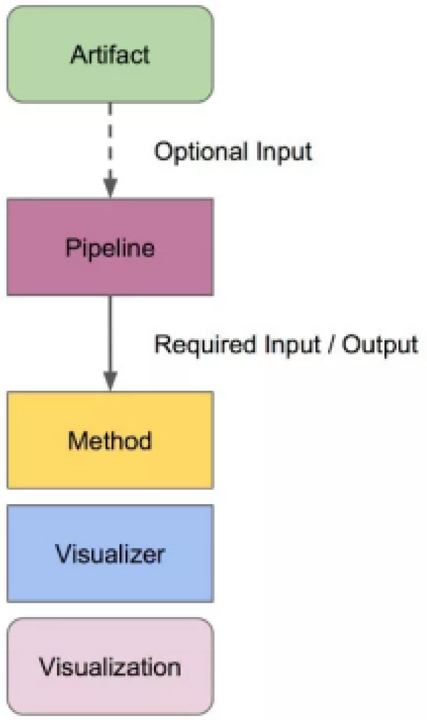
View the association between categorical metadata columns and alpha diversity data for Faith Phylogenetic Diversity and evenness metrics.

α多样性箱式图

FOR RESEARCH USE ONLY

Qiime2特点

5. 分析结果可重现



对象/数据，该软件定义的文件格式（.qza）；存储数据和分析结果；压缩文件

流程(pipelines)，即一条命令运行多个单独的命令。为了简洁，在一些流程图中，流程被标记为箱体 —封装多个内部运行的功能。

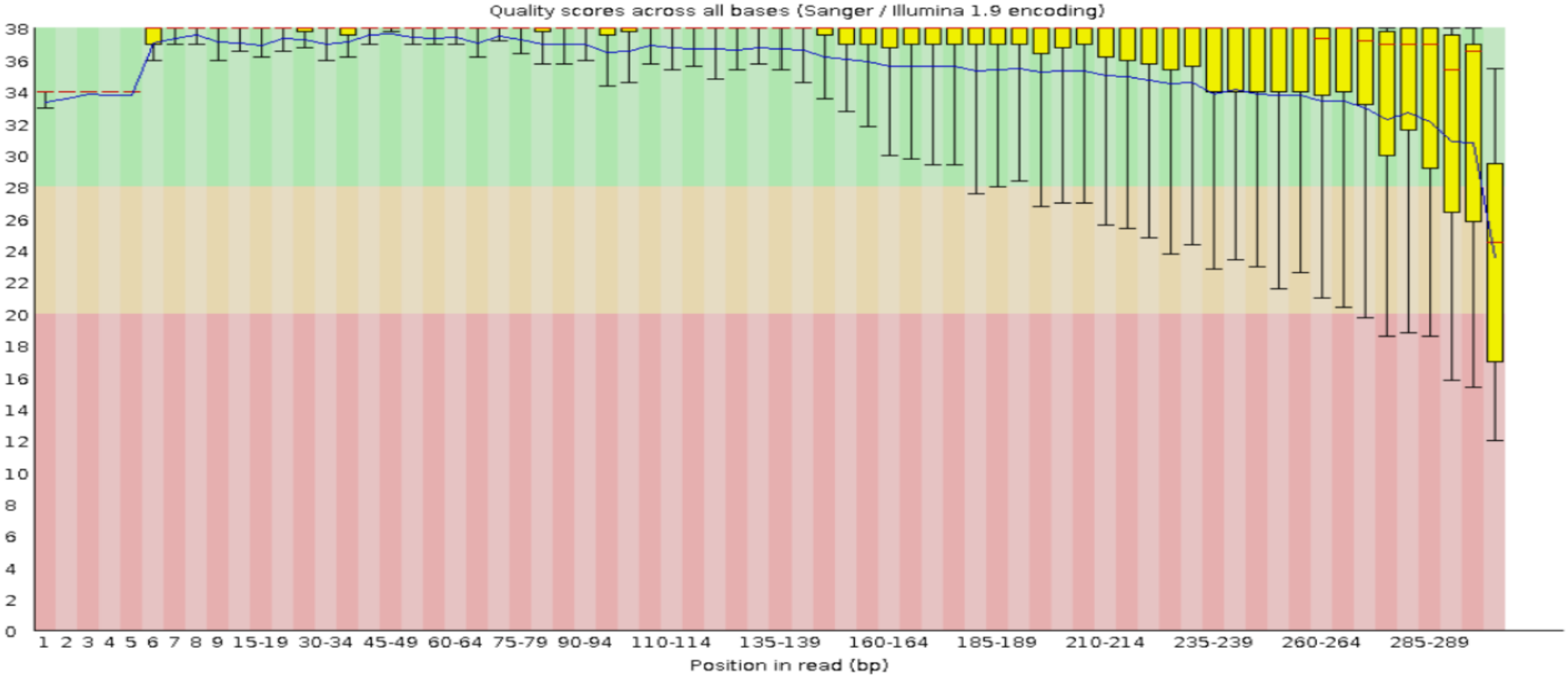
对Artifact 分析的方法，可拓展。

可视化工具，将分析结果可视化的软件。

可视化（.qzv）： 把数据绘制成突变方便查看和分析规律。

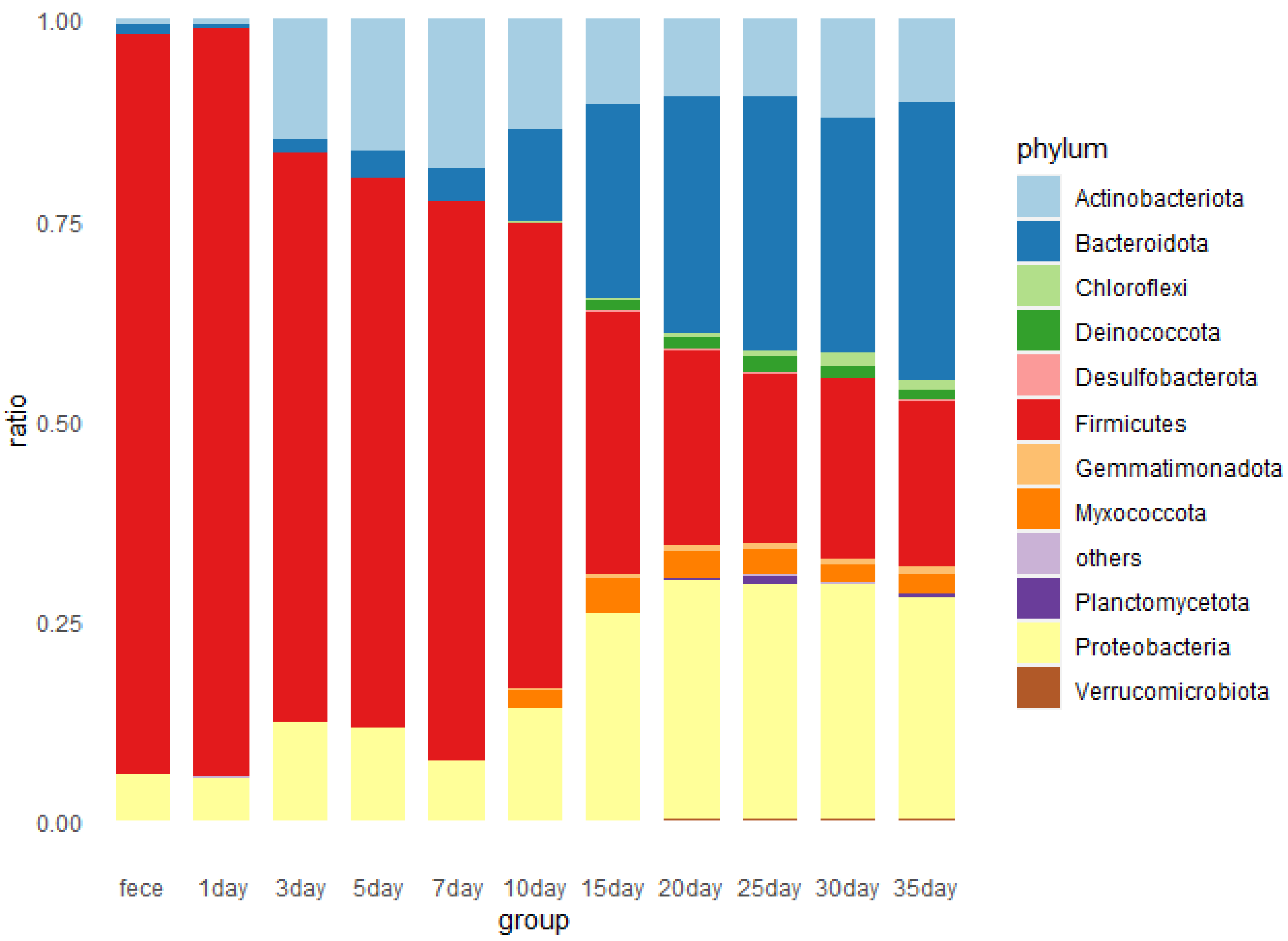
分析测试中心Miseq实例

数据质量 Q30>80%



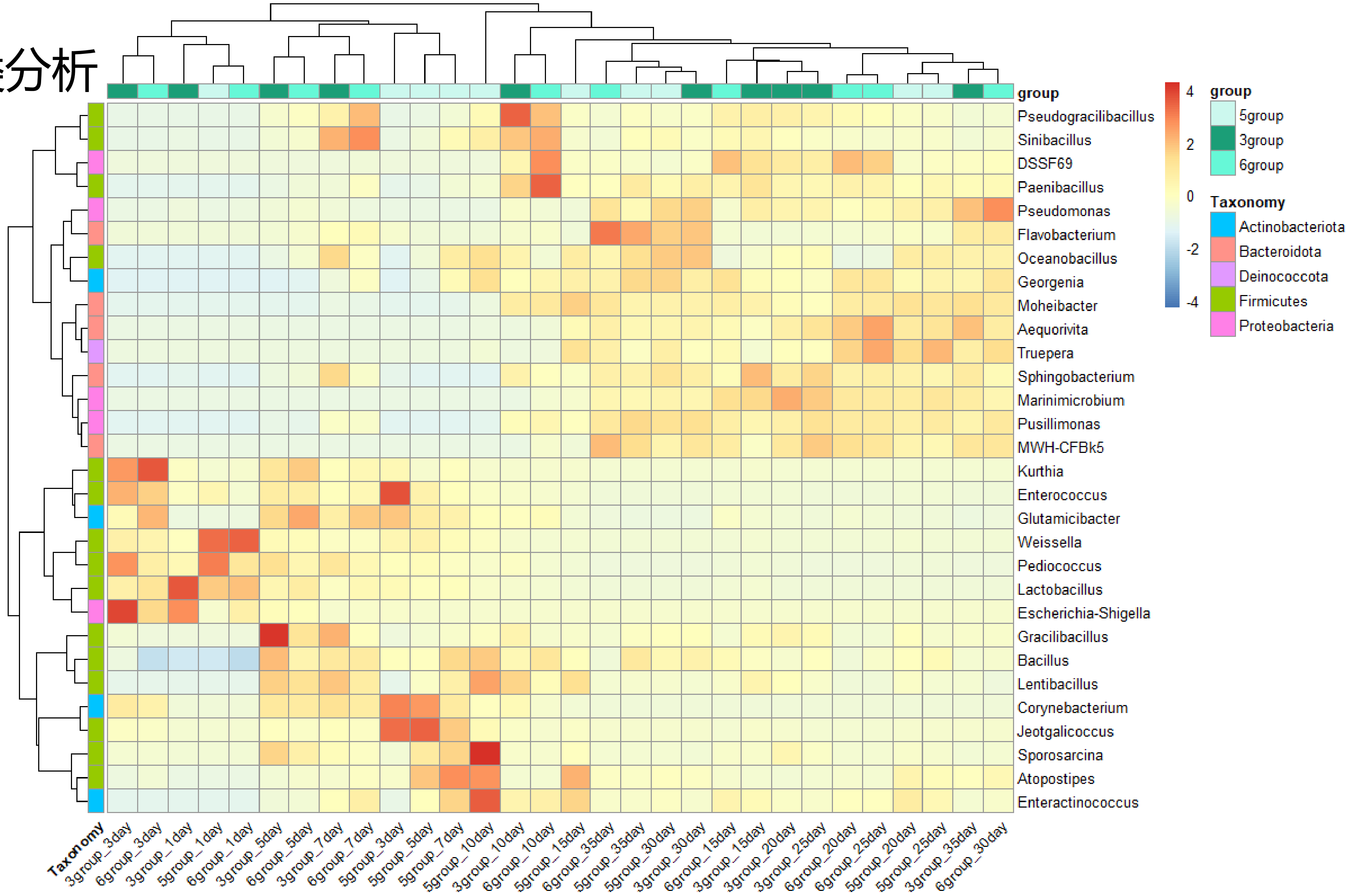
分析测试中心Miseq实例

物种分类分析

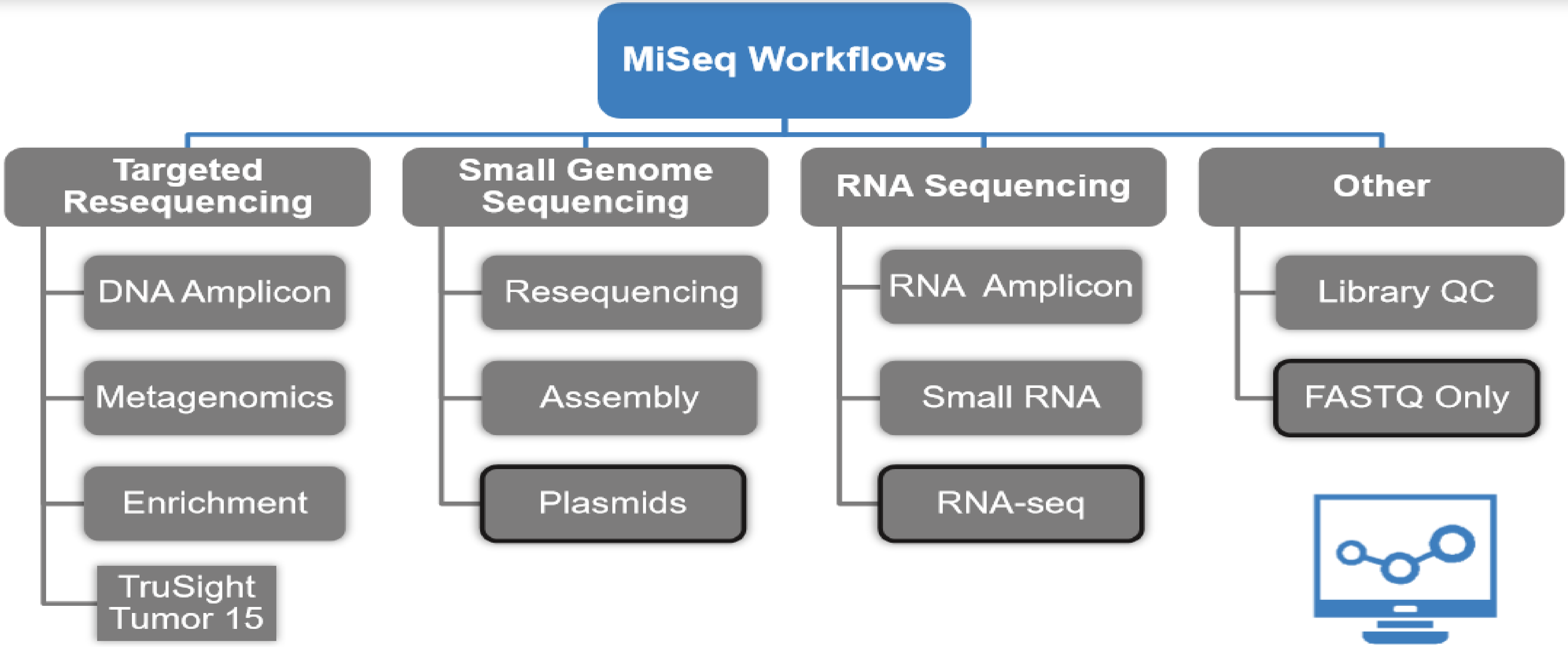


分析测试中心Miseq实例

物种分类分析



MiSeq在其他方面的应用



MiSeq Workflows

- **Targeted Resequencing**
  - DNA Amplicon
  - Metagenomics
  - Enrichment
  - TruSight Tumor 15
- **Small Genome Sequencing**
  - Resequencing
  - Assembly
  - Plasmids
- **RNA Sequencing**
  - RNA Amplicon
  - Small RNA
  - RNA-seq
- **Other**
  - Library QC
  - FASTQ Only

Note: Refer to the Local Run Manager support page to verify that your workflow is supported

Highlighted applications produce FASTQ files only

课题服务

- **所内服务**

1. 袁海生老师课题组，样本：植物根系（800多），服务项目：细菌v4区和真菌ITS1/ITS2区。
2. 孙浩老师，180个核酸DNA，测细菌V3V4区。
3. 王超老师组，300多个土壤样本，测细菌v4和真菌ITS1区。
4. 稳定同位素组，不到100个土壤样品，测细菌V3V4区。
5. 土壤与环境科学组，300多个肥料样本，测细菌v4和真菌ITS1区。
6. 检测中心项目组，土壤样品几百个，测细菌v4和真菌ITS1/ITS2区。

- **所外服务**

1. 安徽农业大学，300多个土壤样品，测细菌v4和真菌ITS1区。
2. 土地资源微生物多样性考察项目等。

2021

谢谢观看

Thank You For Watching