Basic of Next Generation Sequencing (NGS) and analysis

Yi-Mo Deng, PhD

WHO GISRS Influenza Bioinformatics workshop, Singapore, 2019

WHO Collaborating Centre for Reference and Research on Influenza
VIDRL

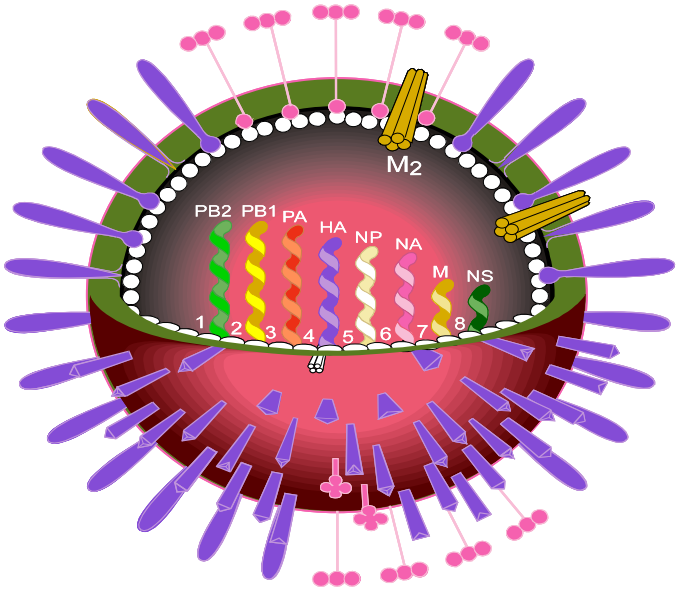
Doherty Institute
A joint venture between the University of Melbourne and Melbourne Health

Outline

- Influenza virus gene sequencing
- NGS technologies and platforms
- Influenza virus NGS analysis and pipelines
- Challenges
- WHO Guidance for NGS

Characteristics of influenza virus

- Influenza A and Influenza B
  - Segmented negative sense RNA genome
  - Highly variable surface glycoproteins
    - Haemagglutinin (HA)
    - Neuraminidase (NA)
  - Influenza A subclassified by HA and NA
    - 18 HA and 11 NA types
    - Many subtypes exist in avian, cross-species infection
  - Influenza B – two lineages
    - Yamagata lineage and Victoria lineage
  - High mutation rate - Antigenic drift
  - Potential for reassortment between influenza viruses - Antigenic shift, pandemic threat
    - H1N1pdm09



Information from influenza genome sequencing

- Diversity: Evolutionary trends
  - New clades/predominant clades
- Key Mutations: Antigenic sites, Receptor Binding sites
- Resistance markers for antivirals:
  - Neuraminidase inhibitors (Oseltamivir), M2 channel blocker (adamantanes), polymerase inhibitors (Baloxivir)
- Virus ID and Reassortment:
  - Novel subtypes H5Nx, H3N2v etc

Some of the key influenza gene substitutions

Segments	Example of significant amino acid substitution	Importance to virus characterisation
PB2 (polymerase basic protein 2)	E627K, D701N	Increased transmissibility to mammalian host/mammalian adaptation
PA (polymerase acidic protein)	I38T/M/F	Reduced susceptibility to Baloxavir
HA (hemagglutinin)	various antigenic sites 226	Antibody escape mutations Receptor specificity
NA (neuraminidase)	H275Y (N1)	Reduced susceptibility to oseltamivir
M1 &2 (matrix proteins 1 & 2)	L26F, V27A, A30V/T/S, **S31N,** G34E for M2	Reduced susceptibility to Admantene

Outline

- Influenza virus gene sequencing
- NGS technologies and platforms
- Influenza virus NGS analysis and pipelines
- Challenges
- WHO Guidance for NGS

NGS, revolution of sequencing technology

Massive parallel high-throughput sequencing, deep sequencing,

non-target specific sequencing technology

- Sequence by synthesis or direct single molecule sequencing

- Various platforms of NGS developed over the years

  -Illumina, Pacific Biosciences, Roche 454, Ion Torrent, MinION

Influenza viruses amplification for NGS

- Full genome amplification of influenza viruses Multi-RTPCR

JOURNAL OF VIROLOGY, Oct. 2009, p. 10309–10313
0022-538X/09/$08.00+0 doi:10.1128/JVI.01109-09

Vol. 83, No. 19

Single-Reaction Genomic Amplification Accelerates Sequencing and Vaccine Production for Classical and Swine Origin Human Influenza A Viruses▽

Bin Zhou,[1,2] Matthew E. Donnelly,[1] Derek T. Scholes,[1] Kirsten St. George,[1] Masato Hatta,[3] Yoshihiro Kawaoka,[3,4] and David E. Wentworth[1,2]*

Wadsworth Center, New York State Department of Health,[1] and Department of Biomedical Sciences, School of Public Health, State University of New York,[2] Albany, New York 12201; Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin—Madison, Madison, Wisconsin 53706[3]; and Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan[4]

Received 31 May 2009/Accepted 8 July 2009

Pandemic influenza A viruses that emerge from animal reservoirs are inevitable. Therefore, rapid genomic analysis and creation of vaccines are vital. We developed a multisegment reverse transcription-PCR (M-RTPCR) approach that simultaneously amplifies eight genomic RNA segments, irrespective of virus subtype. M-RTPCR amplicons can be used for high-throughput sequencing and/or cloned into modified reverse-genetics plasmids via regions of sequence identity. We used these procedures to rescue a contemporary H3N2 virus and a swine origin H1N1 virus directly from human swab specimens. Together, M-RTPCR and the modified reverse-genetics plasmids that we designed streamline the creation of vaccine seed stocks (9 to 12 days).

JCM
Journals.ASM.org

Universal Influenza B Virus Genomic Amplification Facilitates Sequencing, Diagnostics, and Reverse Genetics

Bin Zhou,[a] Xudong Lin,[a] Wei Wang,[a] Rebecca A. Halpin,[a] Jayati Bera,[a] Timothy B. Stockwell,[a] Ian G. Barr,[b] David E. Wentworth[a]
Virology, J. Craig Venter Institute, Rockville, Maryland, USA[a]; WHO Collaborating Centre for Reference and Research on Influenza, Melbourne, Australia[b]

Although human influenza B virus (IBV) is a significant human pathogen, its great genetic diversity has limited our ability to universally amplify the entire genome for subsequent sequencing or vaccine production. The generation of sequence data via next-generation approaches and the rapid cloning of viral genes are critical for basic research, diagnostics, antiviral drugs, and vaccines to combat IBV. To overcome the difficulty of amplifying the diverse and ever-changing IBV genome, we developed and optimized techniques that amplify the complete segmented negative-sense RNA genome from any IBV strain in a single tube/well (IBV genomic amplification [IBV-GA]). Amplicons for >1,000 diverse IBV genomes from different sample types (e.g., clinical specimens) were generated and sequenced using this robust technology. These approaches are sensitive, robust, and sequence independent (i.e., universally amplify past, present, and future IBVs), which facilitates next-generation sequencing and advanced genomic diagnostics. Importantly, special terminal sequences engineered into the optimized IBV-GA2 products also enable ligation-free cloning to rapidly generate reverse-genetics plasmids, which can be used for the rescue of recombinant viruses and/or the creation of vaccine seed stock.

- Partial genome amplification of influenza A and B viruses multi-RTPCR

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

VIROLOGY

Multiplex Reverse Transcription-PCR for Simultaneous Surveillance of Influenza A and B Viruses

Bin Zhou,[a,b] Yi-Mo Deng,[c] John R. Barnes,[d] October M. Sessions,[e] Tsui-Wen Chou,[a] Malania Wilson,[d] Thomas J. Stark,[d] Michelle Volk,[a] Natalie Spirason,[c] Rebecca A. Halpin,[b] Uma Sangumathi Kamaraj,[e] Tao Ding,[a] Timothy B. Stockwell,[b] Mirella Salvatore,[f] Elodie Ghedin,[a,g] Ian G. Barr,[c] David E. Wentworth[b,d]

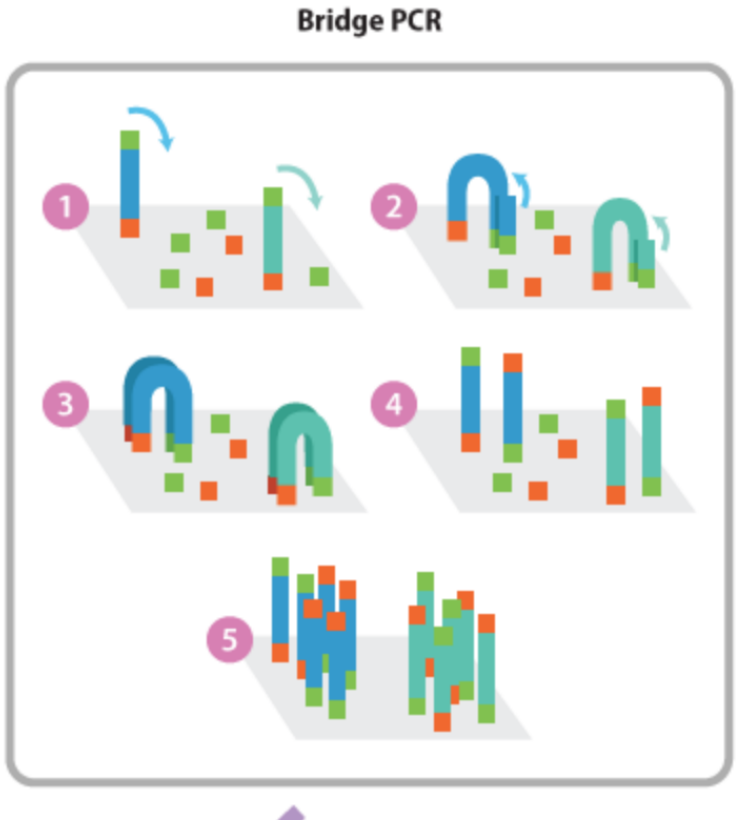
Center for Genomics and Systems Biology, Department of Biology, New York University, New York, New York, USA[a]; J. Craig Venter Institute, Rockville, Maryland, USA[b]; World Health Organization Collaborating Centre for Reference and Research on Influenza, The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia[c]; Influenza Division, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA[d]; Program in Emerging Infectious Diseases, Duke-NUS Graduate Medical School Singapore, Singapore[e]; Department of Medicine and Department of Healthcare Policy and Research, Weill Cornell Medical College of Cornell University, New York, New York, USA[f]; College of Global Public Health, New York University, New York, New York, USA[g]

NGS workflow

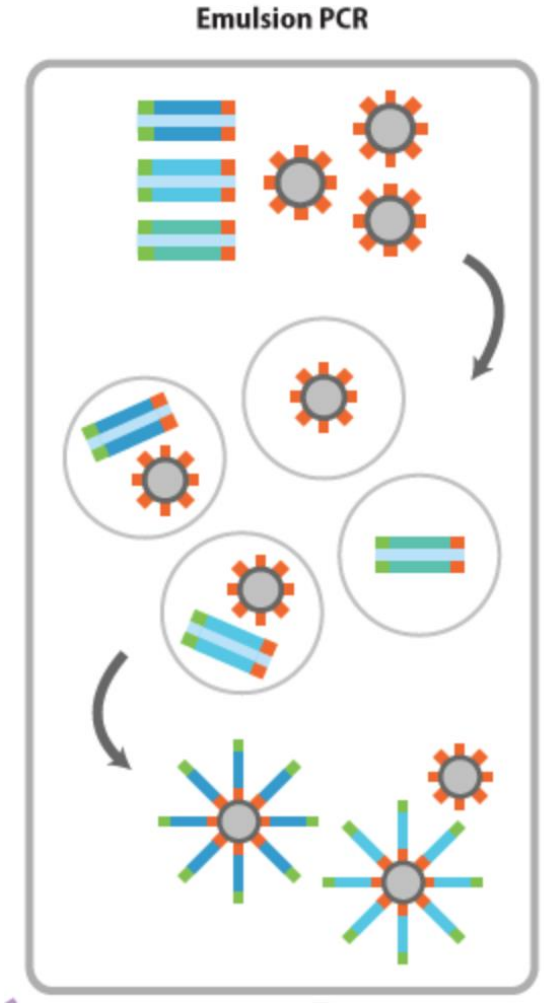
1. Library preparation
   - amplicon based: generate multi segments PCR products
   - Fragmentation and adaptor ligation (bar-coded)

NGS workflow

2. Clonal amplification



Illumina



Ion Torrent

NGS workflow

3. Sequencing millions of the same short reads within each clone

- Different NGS platforms using different sequencing methods
- Each has its own advantages and limitations

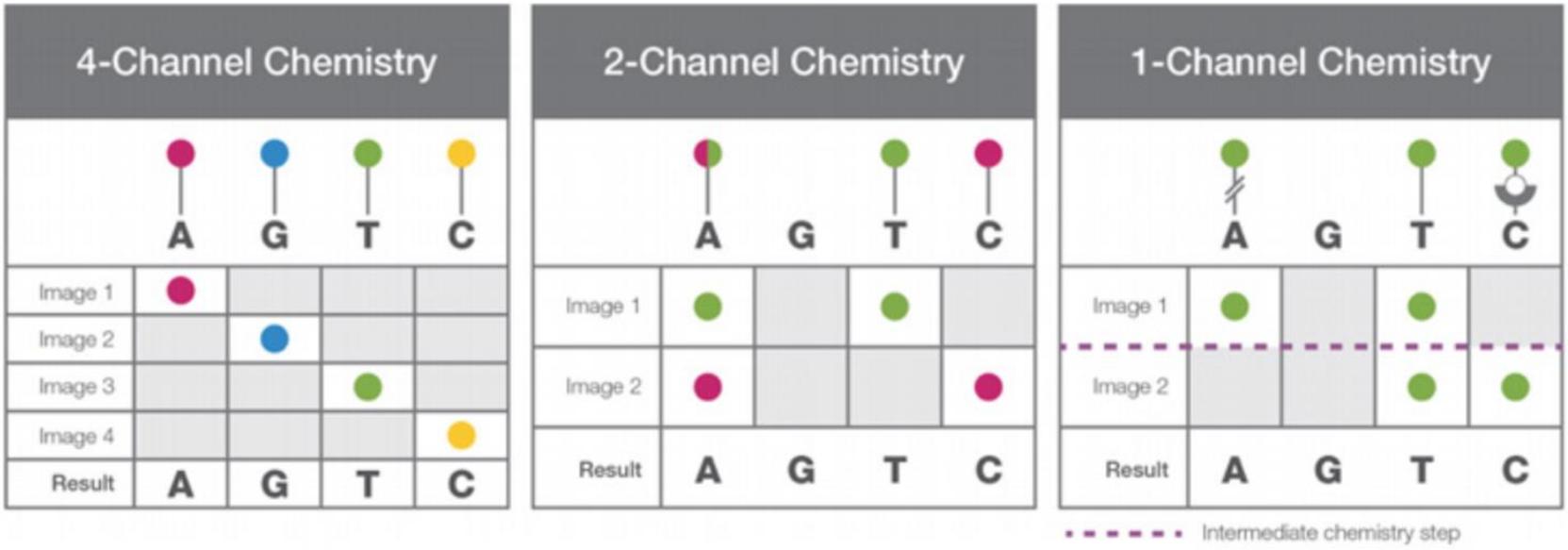
Illumina Platform

Evolution of Illumina Sequencing Chemistry

HiSeq/NextSeq/MiSeq

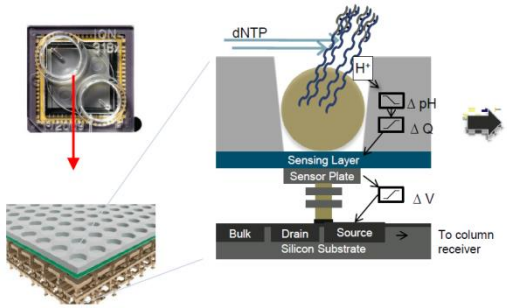
MiniSeq

iSeq

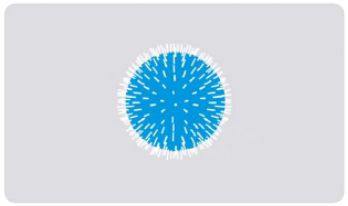


Ion Torrent platform

- Semi-conductor technology
- pH detector



Copy DNA

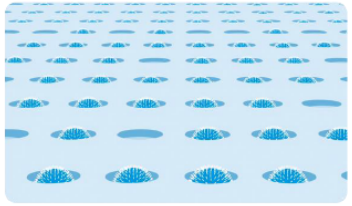


A sample of DNA is cut into millions of fragments, and each fragment is attached to its own bead

The fragment is copied until it covers the bead

This automated process produces millions of beads covered with millions of different fragments

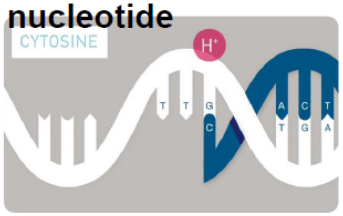
Load chip



The beads are then flowed across the chip, each being deposited into a well

Then the chip is flooded with one of the four nucleotides

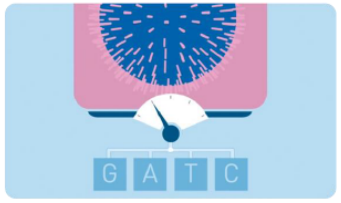
Incorporate nucleotide



If the next base on the DNA strand is complementary to this nucleotide, a nucleotide will be incorporated and a hydrogen ion will be released

The hydrogen ion changes the pH of the solution in the well

Detect and call



An ion-sensitive layer beneath the well measures that pH change and converts it to voltage

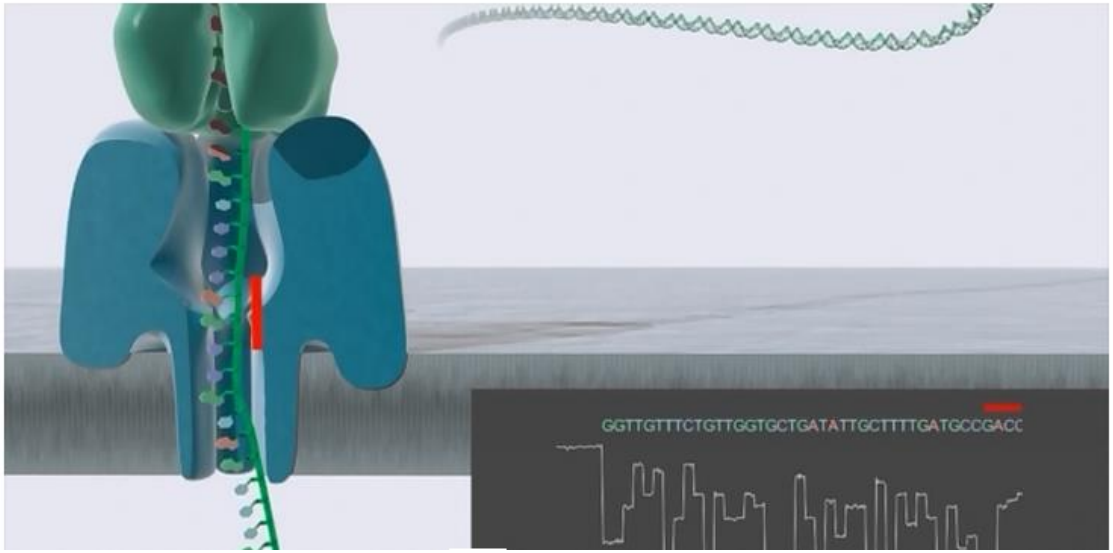
This voltage change is recorded, indicating the nucleotide has been incorporated and the base is called

This process happens simultaneously in millions of wells

life technologies

MinION platform

- Nanopore-based electronic systems for analysis of single molecules
- Capable of sequencing RNA directly
- Portable sequencer



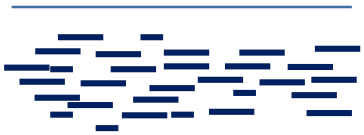
4. Data analysis

Reconstruct the genome from the fragmented reads using various of programs

- PGM Suite
- IGV
- CLC Genomics
- Geneious
- Custom made Pipeline: IRMA, FluLINE…

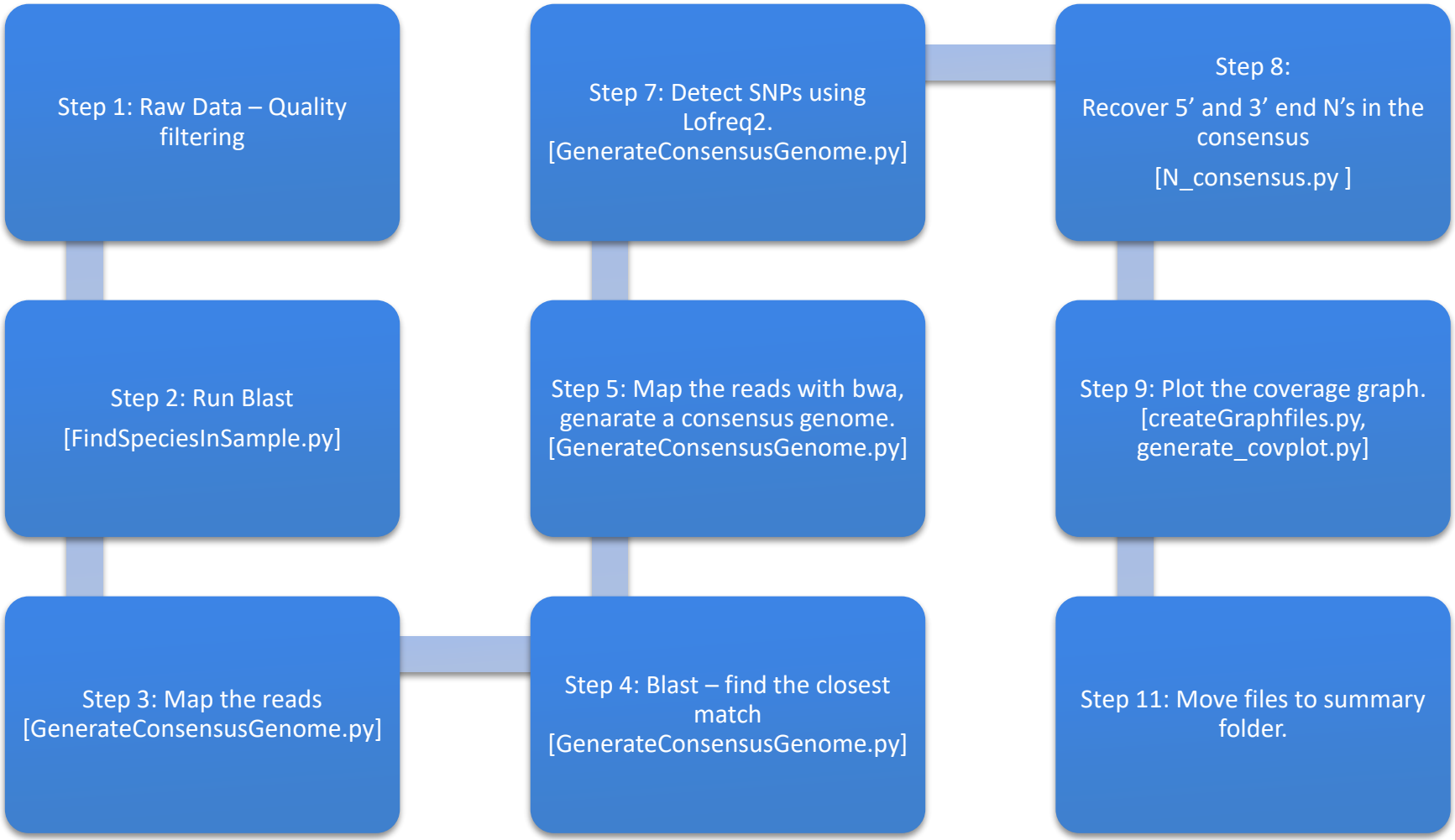
Barcode 001

Barcode 002



Barcode 003

FluLINE – a pipeline for Influenza NGS data analysis



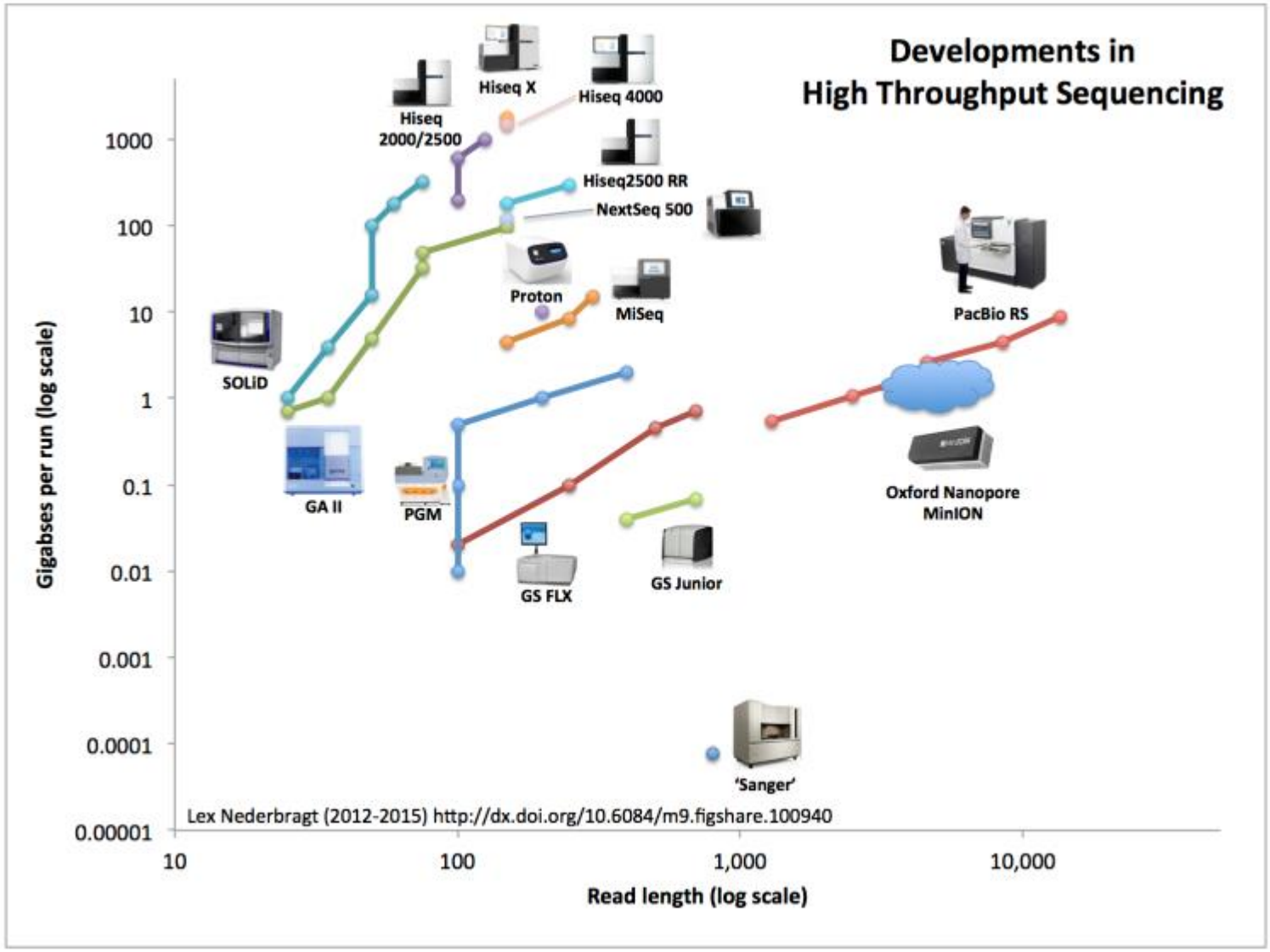
Outline

- Influenza virus gene sequencing
- NGS technologies and platforms
- Influenza virus NGS analysis and pipelines
- Challenges
- WHO Guidance for NGS

Challenges

- Which chemistry/platform to choose
- Original specimens
  - Host gene/other pathogens background
  - Enrichment of specific target
- Defective interfering (DI) virus particles
  - Partial sequences for certain genes
- Bottleneck: data analysis (time-consuming, often requires bioinformatics expertise)
- Common problems for data analysis:
  - gaps, uneven coverage, deletion/insertion detection, reference selection
  - De novo analysis for unknown influenza
- A pipeline is useful, but is not the answer for everything

Read length vs Data output



flxlexblog.wordpress.com

	 iSeq 100*	 MiniSeq	 MiSeq*†	 NextSeq*†
Output Range	1.2 Gb	1.8–7.5 Gb	0.3–15 Gb	20–120 Gb
Run Time	9–17.5 hr	4–24 hr	5–55 hr	11–29 hr
Reads per Run	4 million	8–25 million	1–25 million	130–400 million
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp
Samples per Run§	1–48	1–96	1–96	12–36
Relative Price per Sample§	Higher Cost	Mid Cost	Mid Cost	Lower Cost
Relative Instrument Price§	Lower Cost	Mid Cost	Mid Cost	Higher Cost
Downloads	Spec Sheet	Spec Sheet	Spec Sheet	Spec Sheet
System Overview	iSeq Overview ›	MiniSeq Overview ›	MiSeq Overview ›	NextSeq Overview ›

Outline

- Influenza virus gene sequencing
- NGS technologies and platforms
- Influenza virus NGS analysis and pipelines
- Challenges
- WHO Guidance for NGS

WHO Guidance on NGS

- Objective:

  To provide practical information and guidance to

  - Evaluate and choose sequencing technologies – Sanger/NGS

- Target Audience:

  NICs who are considering to implement NGS technology

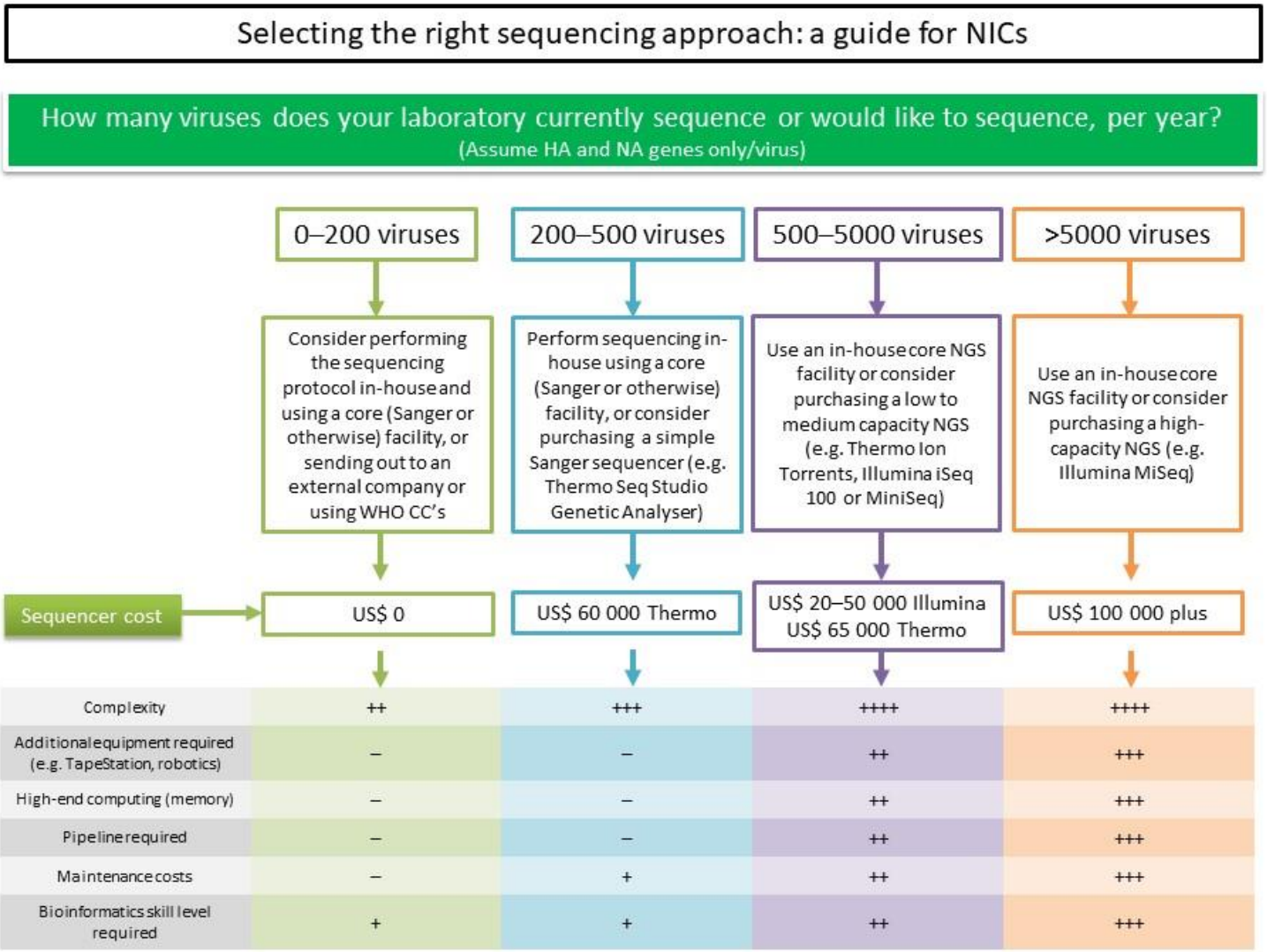
- Goal:

  For genetic characterisation and surveillance of influenza viruses

Main considerations

- Sample size and collection pattern
- Budget
  - Initial instrument cost
  - Ongoing annual maintenance cost
  - Reagent/kit cost
  - Other equipment cost (robotic, extra data storage capacity)
- Local technical support, stable reagent supply
- Bioinformatics expertise/support
- Reagent storage
- Resources (computational, stable power supply)

Sample size vs sequencing platform



Summary

- Powerful tool for new influenza virus discovery and ongoing surveillance in a timely manner
- High throughput for routine sequencing, cheaper than Sanger sequencing if a large amount of viruses sequenced simultaneously
- Bioinformatics analysis is challenging, even using a pipeline
- Evaluate and plan well if consider to get an NGS platform