Country capacity landscape assessment

Ruklanthi (Rukie) de Alwis
Deputy Director, Duke-NUS Center for Outbreak Preparedness
Asia Pathogen Genomics Initiative (Asia PGI)

WORKSHOP PARTNERS

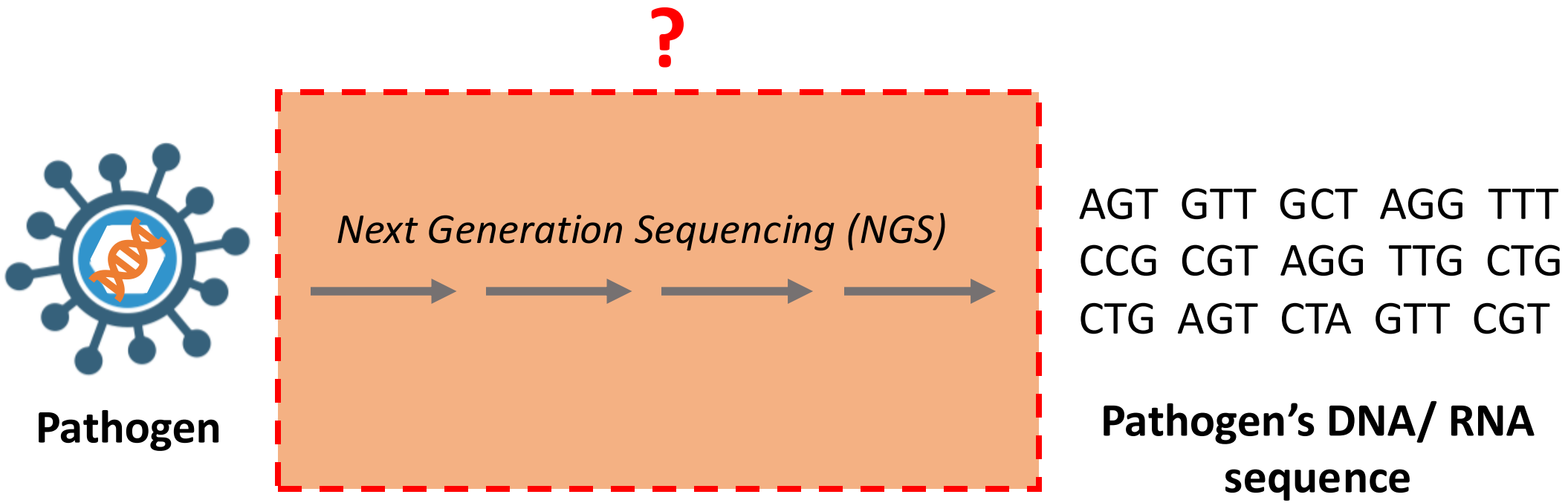
CDC U.S. CENTERS FOR DISEASE CONTROL AND PREVENTION

DukeNUS Medical School
Asia Pathogen Genomics Initiative

CENTRE FOR PATHOGEN GENOMICS

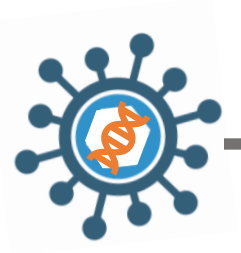
Sydney Infectious Diseases Institute
Centre for Infectious Diseases & Microbiology
WHO Southeast Asia Regional Office (SEARO)
WHO Western Pacific Regional Office (WPRO)
WHO International Pathogen Surveillance Network (IPSN)

What is Pathogen Genomics?



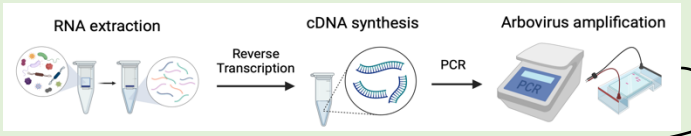
***NGS**: A modern and rapid method of DNA sequencing that can read a large volume of DNA sequences simultaneously*

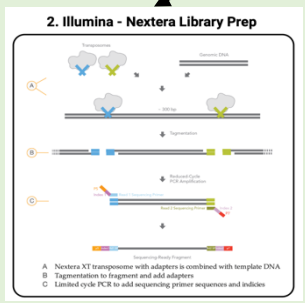
What is Pathogen Genomics?

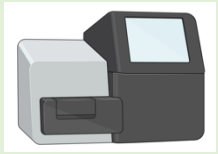


Pathogen

Laboratory

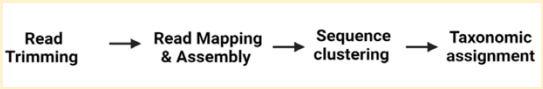


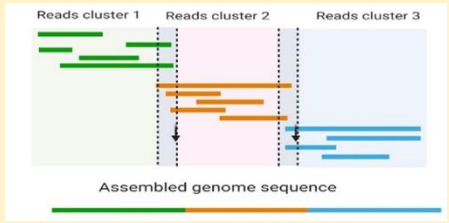




- Laboratory capacity/ infrastructure
- Equipment capacity
- Sequencing capacity
- Protocols
- Wet lab trained Human resources
- Accreditation and EQA

Bioinformatics





- Computing infrastructure/capacity
- Dry lab trained human resources
- cloud computing

AGT GTT GCT AGG
CCG CGT AGG TTG
CTG AGT CTA GTT

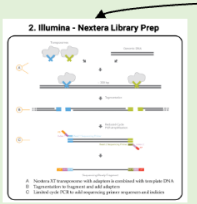
Pathogen's DNA/ RNA sequence

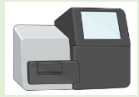
What is Pathogen Genomic Surveillance?

Sample

Pathogen

Laboratory

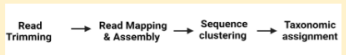




- Laboratory capacity/ infrastructure
- Equipment capacity
- Sequencing capacity
- Protocols
- Wet lab trained Human resources
- Accreditation and EQA

Supply Chain

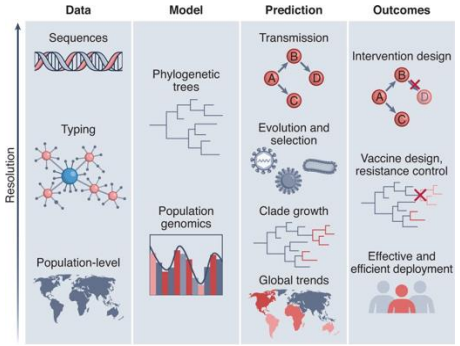
Bioinformatics



- Computing infrastructure/capacity
- Dry lab trained human resources
- cloud computing

AGT GTT GCT AGG
CCG CGT AGG TTG
CTG AGT CTA GTT

Pathogen's DNA/ RNA sequence



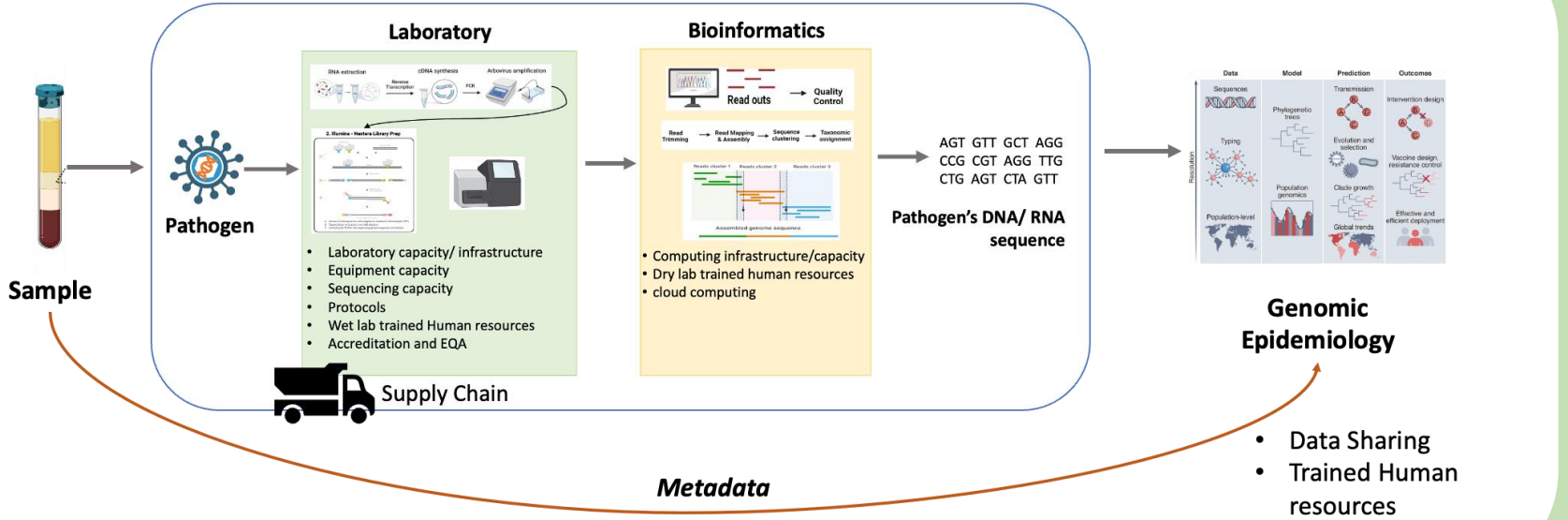
Genomic Epidemiology

Metadata

- Data Sharing
- Trained Human resources

What is Pathogen Genomic Surveillance?

Enabling Environment



- Country NGS partners
- Financial capacity
- Industry partners (technology platforms)
- Integrated surveillance plans

Types of sequencing approaches?

Table 1. Strengths and limitations of technologies and approaches commonly used for pathogen genomic sequencing

Pathogen genomic sequencing

Technology or approach	Description	Strengths	Limitations
Technology			
Second generation	massively parallel sequencing with read lengths of 75–300 bp; most commonly used platforms are from Illumina, Ion Torrent (Thermo Fisher), and MGI	higher per-read-sequencing accuracy, therefore better at identification of single-nucleotide variants (SNPs), small insertions, and deletions; short reads allow for greater read depth for a given GB of output; lower cost per GB at high throughput	repetitive/homologous regions and structural variants pose difficulty in sequence assembly; amplification bias
Third generation	sequencing of native DNA; capable of reaching read lengths up to and greater than 10 kb; long-read technologies are provided by Oxford Nanopore Technologies and PacBio	longer read lengths allow greater genome coverage; advantageous for *de novo* assembly and sequencing of novel pathogens better for sequencing repetitive regions and structural variants (such as large insertions, deletions, duplications, or translocations) advantageous for genomic resolution of plasmids (which often carry antibiotic-resistant genes)	lower per-read accuracy; more stringent requirements for input quality and quantity
Approach[a]			
Amplicon-based	a targeted sequencing approach involving PCR amplification of genes or genetic material from the pathogen of interest, followed by sequencing	usually the cheapest sequencing approach and often easiest to implement and integrate with existing laboratory processes; because of PCR amplification, low input sample material required and higher likelihood of obtaining sufficient depth; relatively straightforward sequence assembly and bioinformatic analysis; less data storage and processing infrastructure required	prior knowledge of infecting pathogen required; possible PCR bias; in situations where circulating strain differs in the primer-binding regions, this may lead to gaps in resulting genome
Probe-based	utilizes synthetic probes to capture genes or genomes of interested pathogens; captured genomic material is then sequenced	able to capture a range of pathogens need less prior knowledge of exact infecting pathogen; greater uniformity of coverage; relatively straightforward bioinformatic analysis	longer and more laborious workflows; can be more expensive than amplicon-based sequencing due to cost of probe sets
Metagenomics	a non-targeted approach that sequences all genetic material (i.e., all pathogens and host nucleic acid) in the sample; metagenomics workflows can involve some prior treatment steps to reduce host material	allows for discovery of highly divergent strains or novel and rare pathogens; relatively easy and less time-consuming laboratory workflow	complex and heavy bioinformatic analysis; high data storage and processing infrastructure needed; most costly sequencing approach per sample; privacy concerns around sequenced host human genomics data

[a]Strengths and limitations for amplicon-based, probe-based, and metagenomics approaches described here are for culture-free applications.

PCR vs tNGS

	qPCR	Targeted NGS
Benefits	• Familiar workflow • Accessible equipment available in most labs • Good screening test if you know what you are looking for • Lower cost	• Both identifies and sequences the target genomes • Higher discovery power - better for detecting mutations/variants • Good for **simultaneously identifying many targets** for **many pathogens** • Detects gene expression changes
Challenges	• Detects only known pathogen fragments • Does not sequence therefore limited discovery power • Limited throughput and mutation resolution • Usually test for single pathogens; multi-pathogen approaches are more complex and costly	• Costly if you are trying to detect a limited number of targets

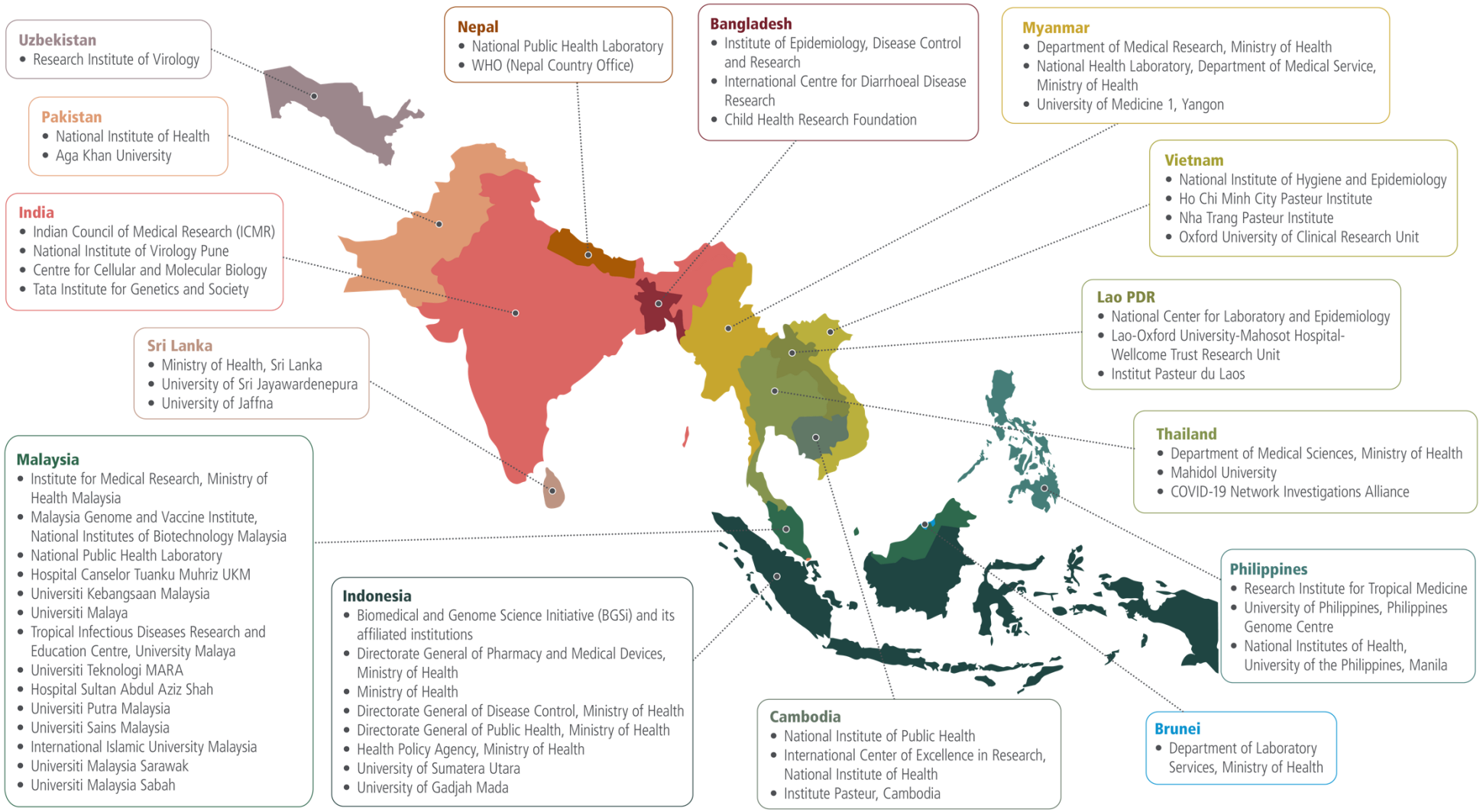
Pathogen size

More genetic material = more expensive to sequence

Species	*T2 phage*	*Escherichia coli*	*Drosophila melanogaster*	*Homo sapiens*	*Paris japonica*
Genome Size	170,000 bp	4.6 million bp	130 million bp	3.2 billion bp	150 billion bp
Common Name	 Virus	 Bacteria	 Fruit fly	 Human	 Canopy Plant

Asia PGI landscape assessment

ASIA PATHOGEN GENOMICS Initiative



End to end country capacity assessment

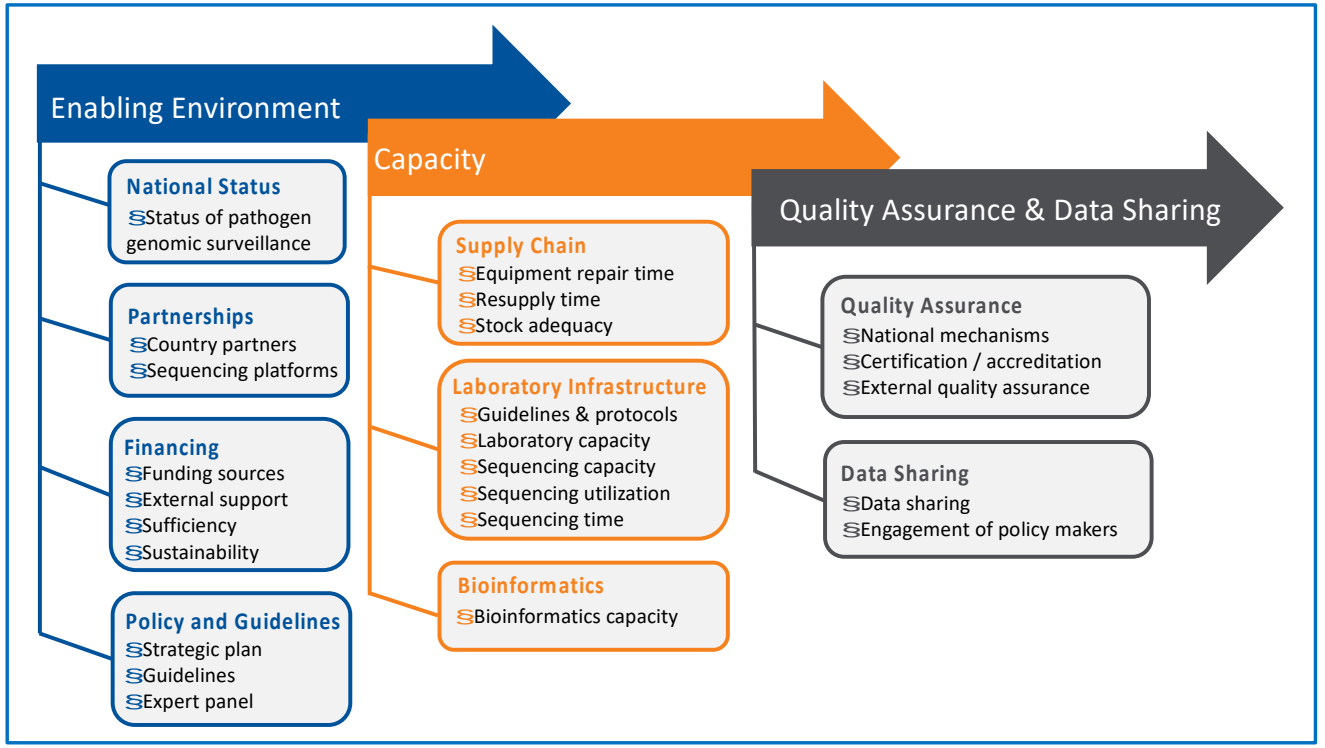
27 Source Documents

Synthesis of existing tools → *Extract relevant questions* → *Initial country consultations* → *Refine with partner feedback*

>100 Questions
(approx. 19 pages)

Source examples:

- UK New Variant Assessment Platform (NVAP) lab scoping survey
- FAO Laboratory Mapping Tool
- Regulatory System Profiling Instrument (RSPI), CoRE
- CDC Needs Assessment
- WHO: Genomic sequencing of SARS-CoV-2
- WHO: HSE GCR Laboratory Assessment Tool
- WHO: Global Influenza Surveillance and Response System
- WHO: GLASS Whole Genome Sequencing for surveillance of antimicrobial resistance
- WHO: Whole genomic sequencing for foodborne disease surveillance
- WHO: The use of NGS technologies for the detection of mutations associated with drug resistance in Mycobacterium tuberculosis complex: technical guide
- European Observatory on Health and Policies: Regulating the unknown, a guide to regulating genomics for health policy-makers (policy brief)
- FIND Next Generation Sequencing Global Capacity Mapping for SARS-CoV-2

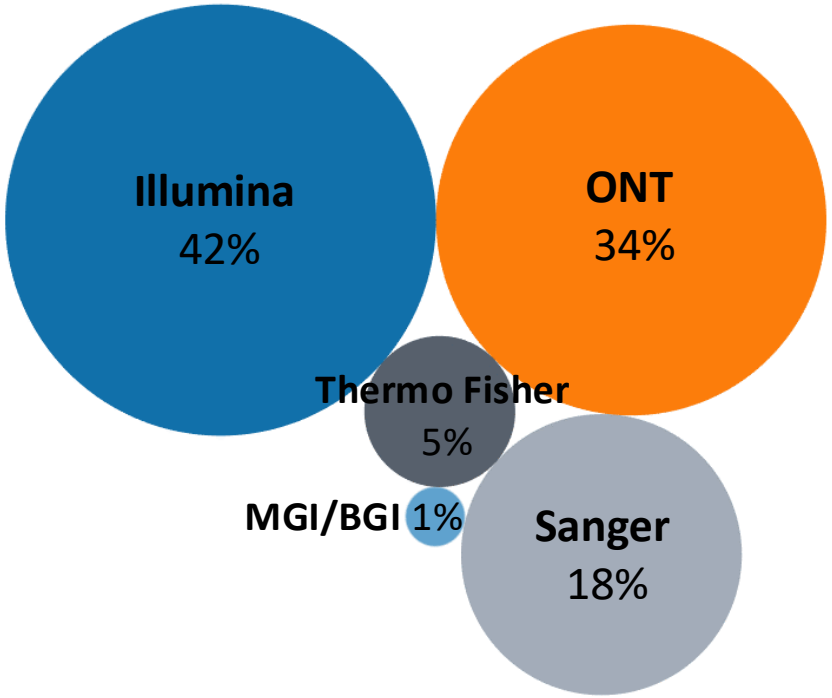


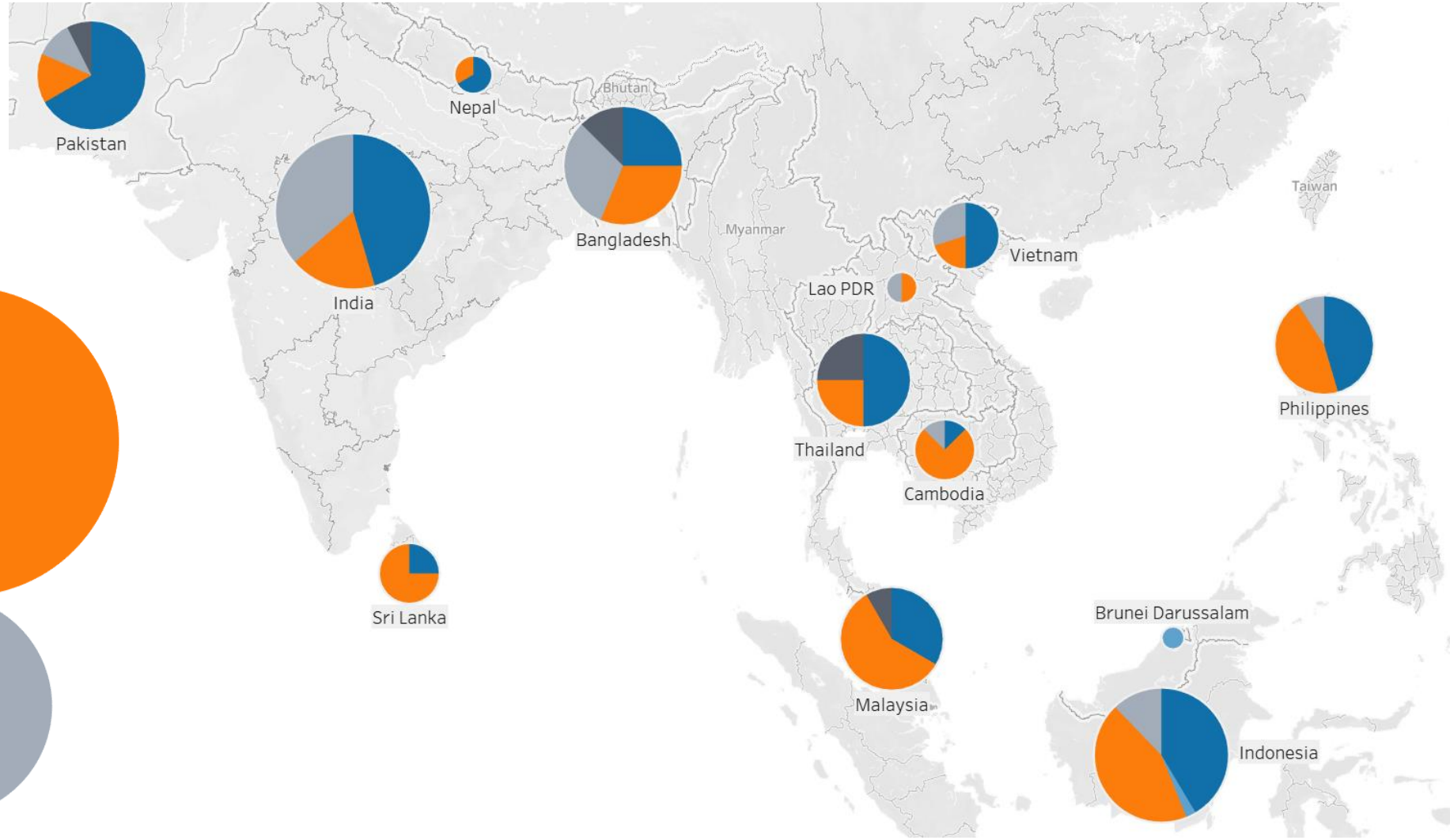
Enabling Environment: Sequencing exist across Asia

13 countries

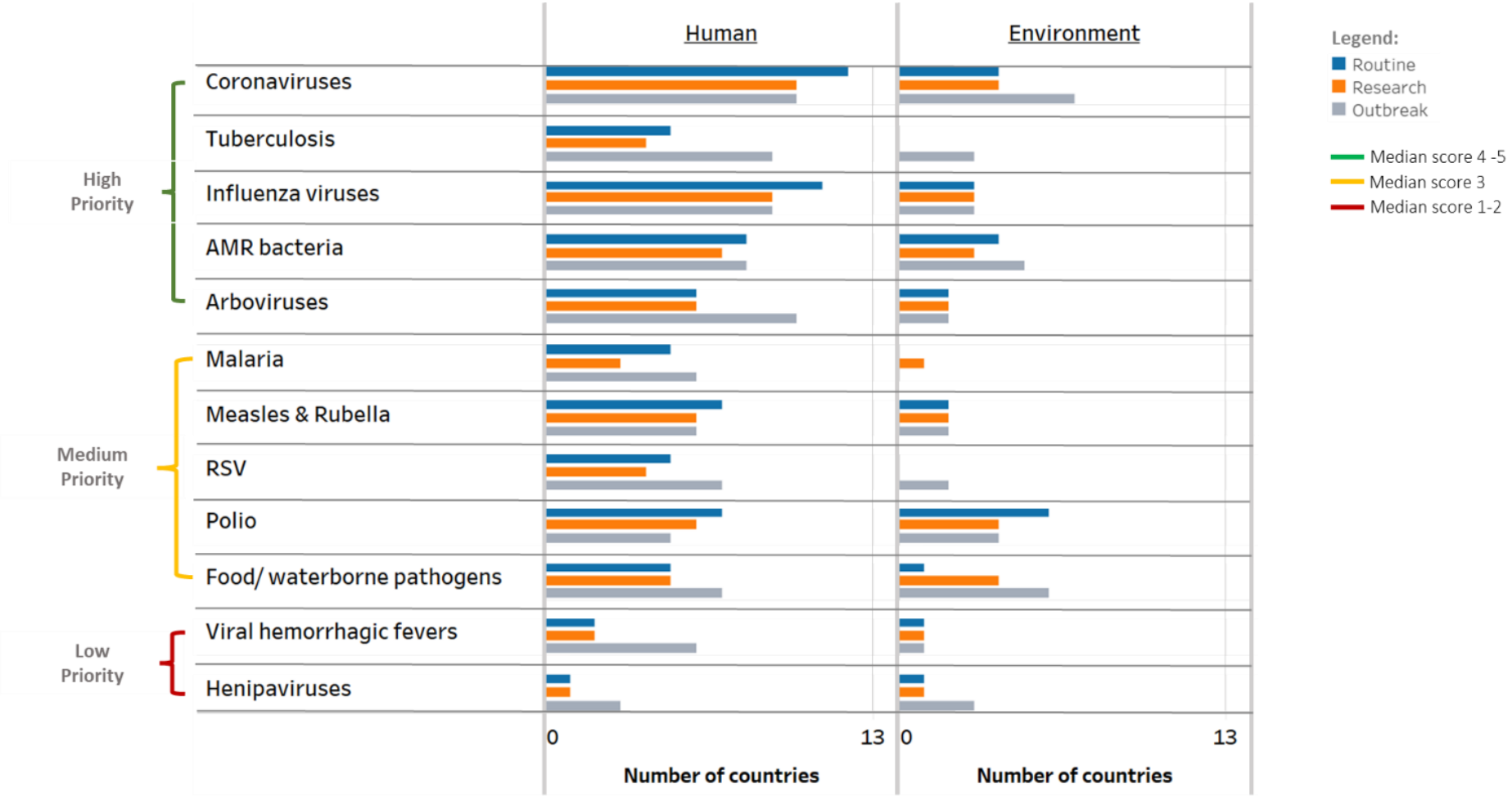
253 sequencers

156 laboratories





Enabling Environment: National status of Pathogens sequenced



Capacity Lab: Wide range in sequencing efficiency and turnaround time

46%

Efficiency: Sequencing output vs total capacity

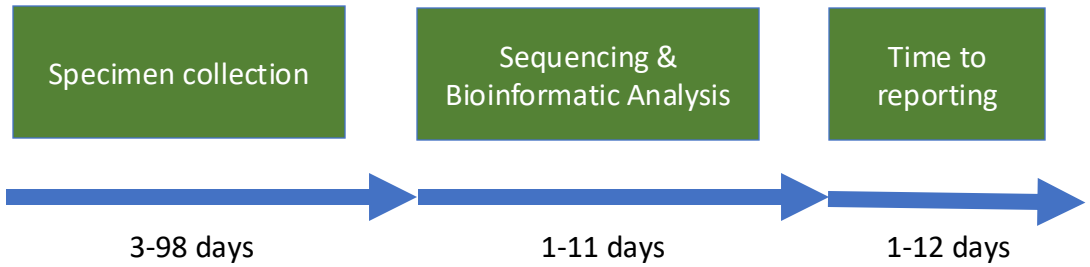
Cross country variation in reported sequencing efficiency

29 days

Timeliness: Average time between sample collection and reporting

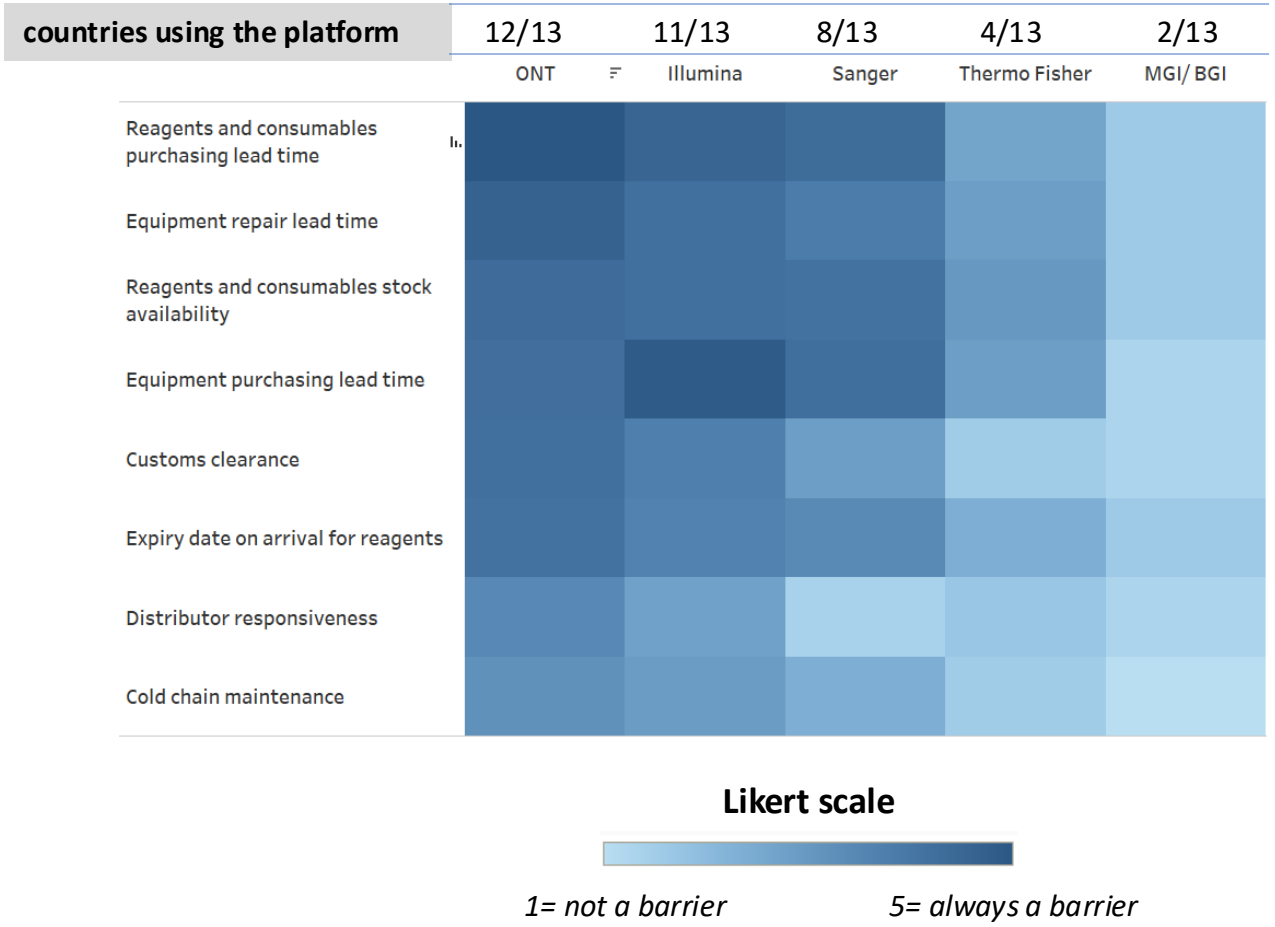
Range - 8-113 days

Delays between sample collection and sequencing

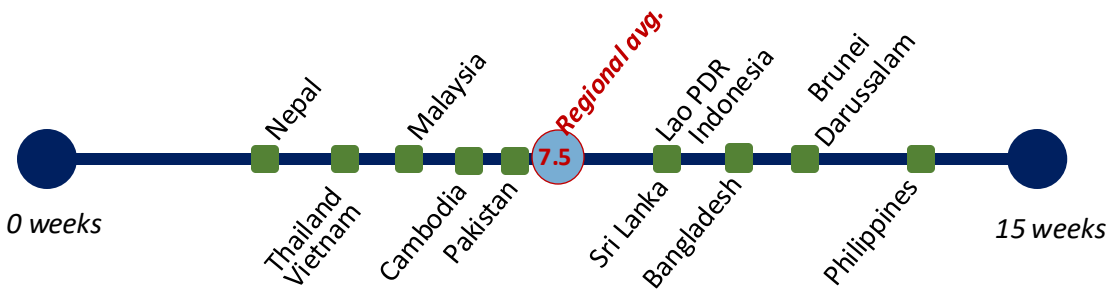


Capacity Lab: Supply chain capacity of NGS reagents

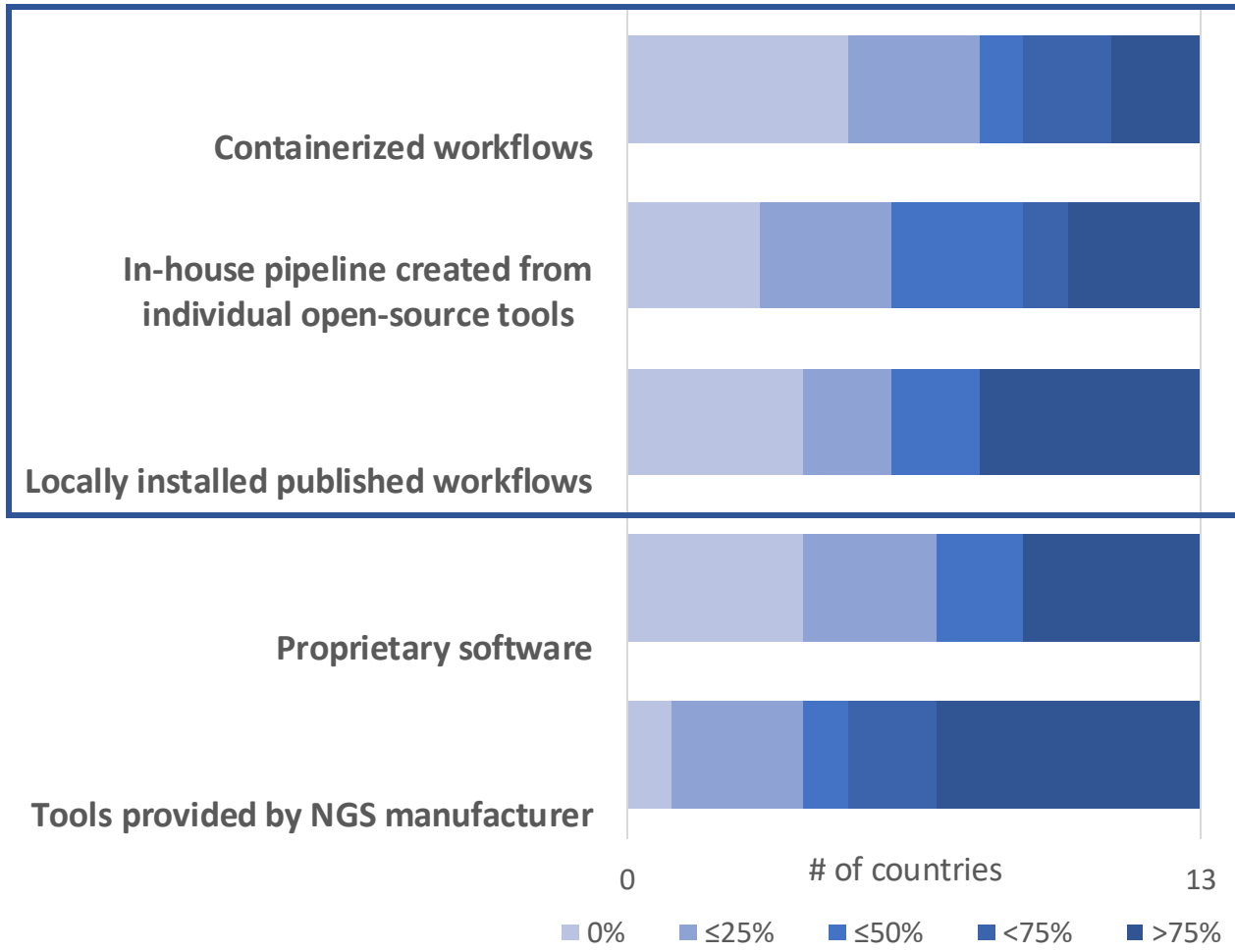
Supply chain barriers



Average re-supply time = 7.5 weeks



Capacity Bioinformatics: High dependence on proprietary software



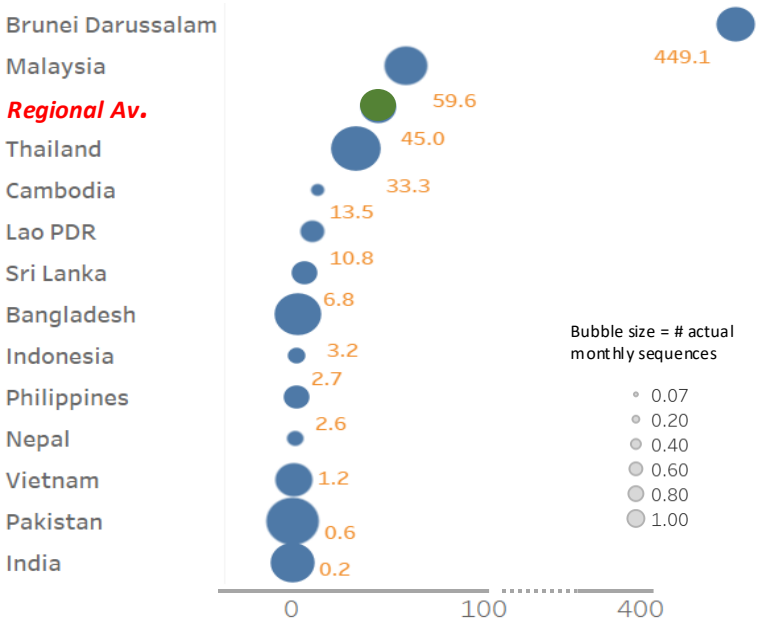
76% countries perform bioinformatics using either containerized workflows, in-house pipeline or locally installed public workflows, however only **46% uses them to perform >75% of the work**

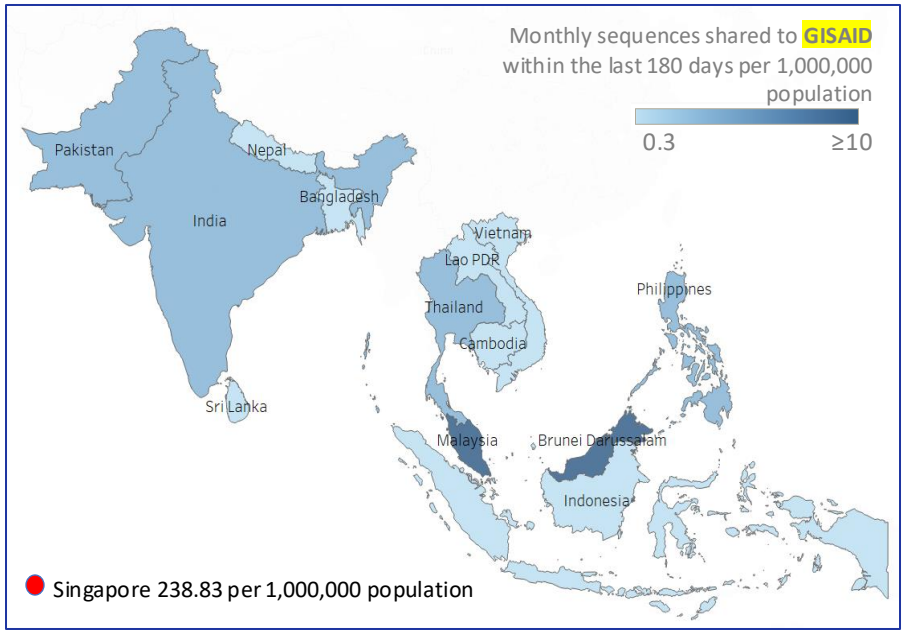
92% (12/13) countries use tools provided by NGS manufacturer for bioinformatic analysis

Data sharing: Sequencing capacity and data sharing varies across the region

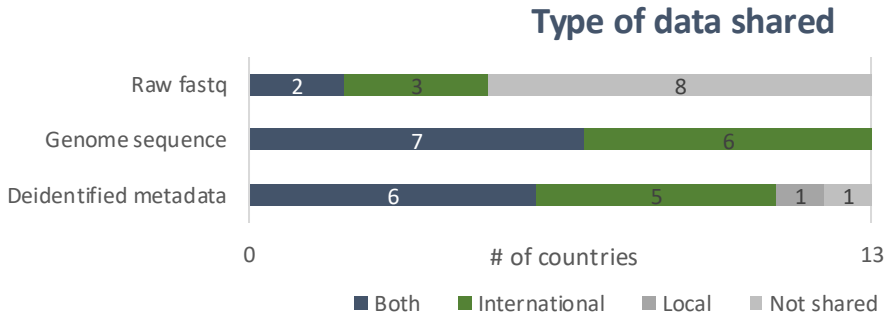
Number of Sequences/ 1 million population

- **45 sequences/ 1,000,000 population** is the average of monthly pathogen sequences generated in past year in the region.





Source: sequence shared data from https://gisaid.org/submission-tracker-global/ and 2021 population data from the World Development Indicators

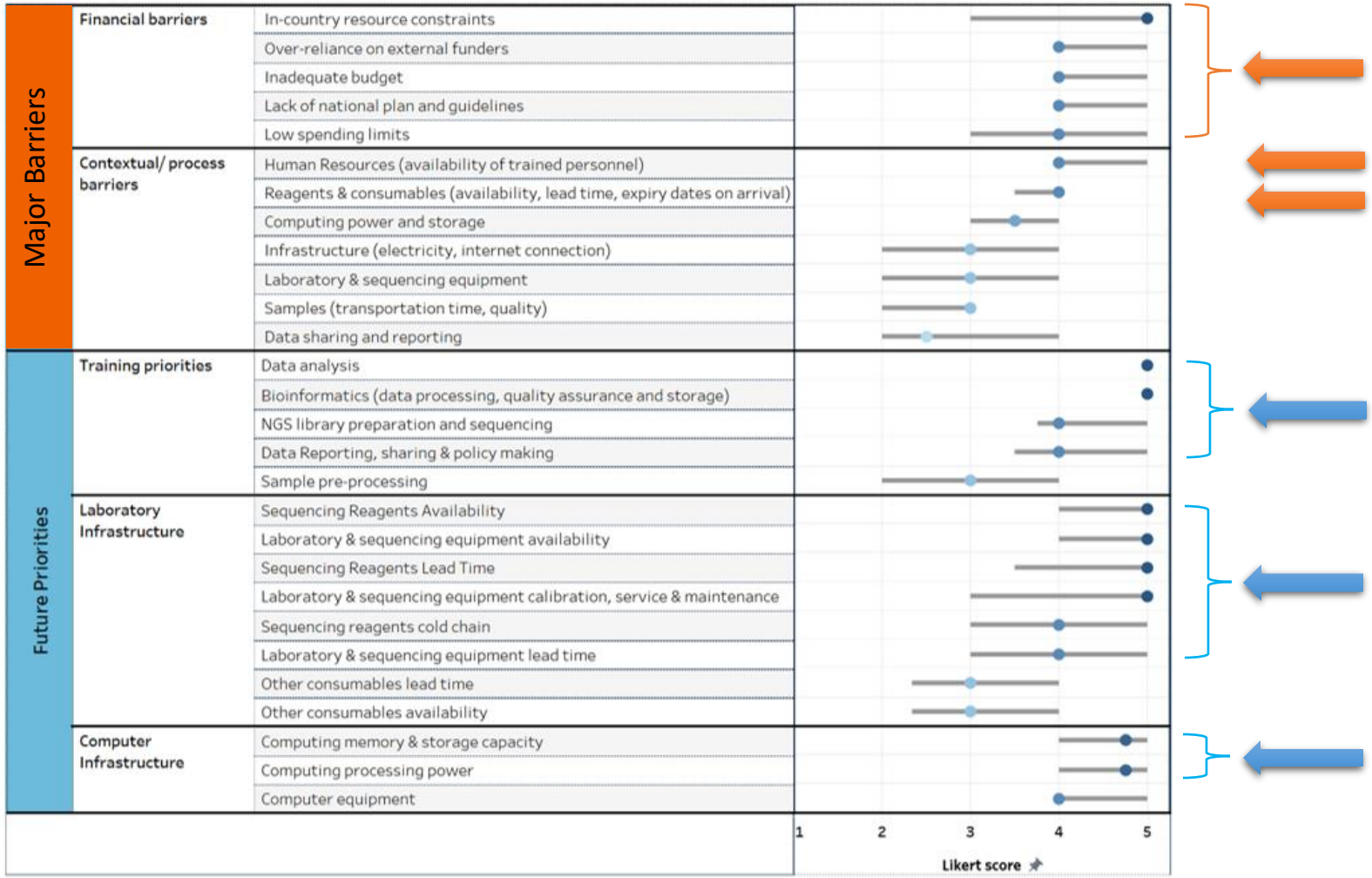


Majority of countries shared **deidentified metadata (92%)** and **genome sequence data (100%). 39% shared Raw fastq**.

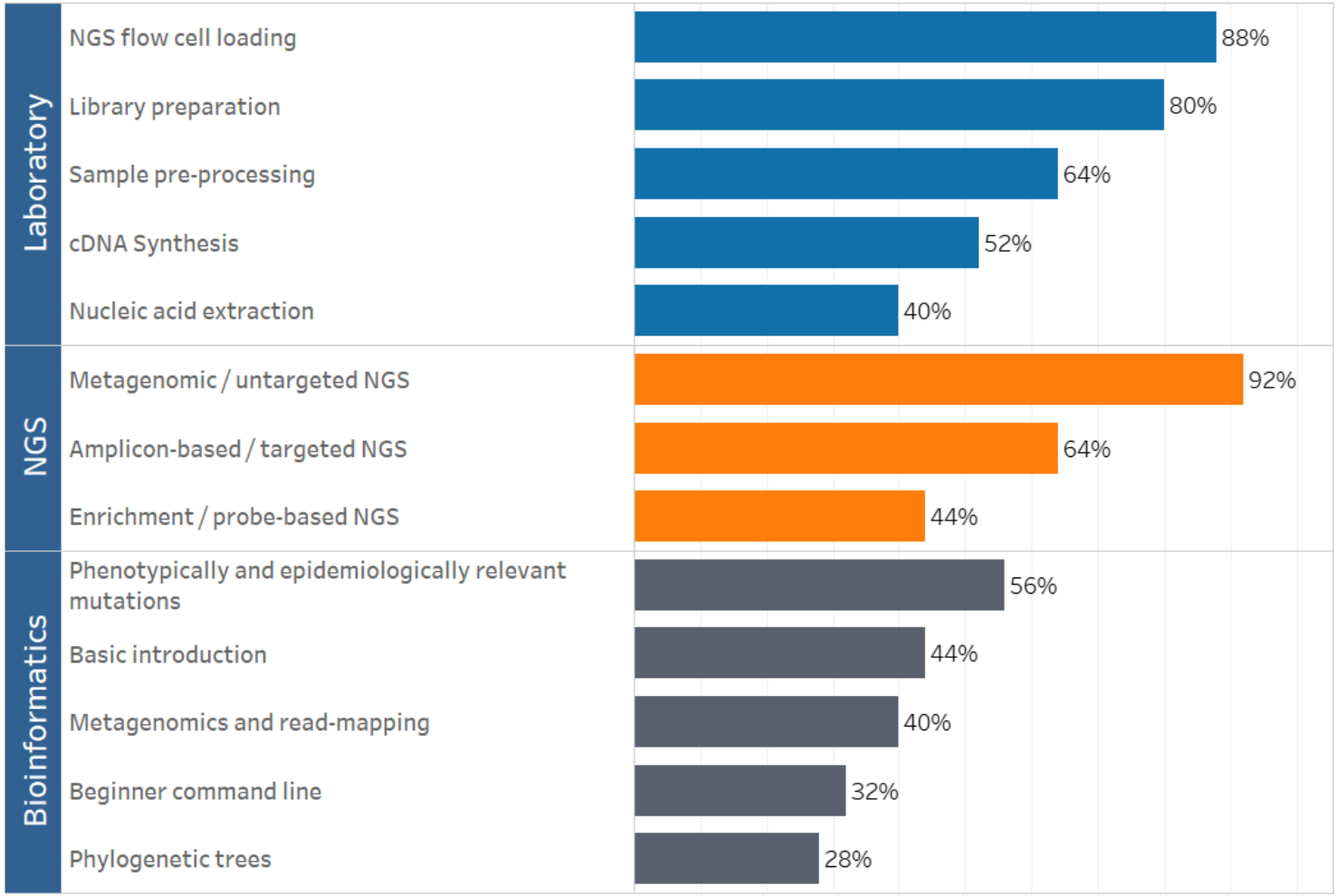
Major Barriers and Future Priorities for pathogen genomic surveillance

Likert scores (1-5) and interquartile ranges are displayed above, with scores for:

- Major Barriers ranging from 1 (not a barrier) to 5 (always a barrier)
- Future priorities from 1 (not a priority) to 5 (essential).



Human Resource: Capacity development priorities



Recommendations for accelerating pathogen genomics

FINANCING

- National investment cases for pathogen genomic surveillance
- Global Financing - prioritise early detection

POLICY & GUIDELINES

- National surveillance plans that include pathogen genomics
- Multi-partner national coordination mechanisms – Expert technical panels

SUPPLY CHAIN

- Pooled procurement mechanisms for genomic surveillance commodities
- Regional supply chain solutions – manufacturing, warehousing, distribution, customs

LABORATORY CAPACITY

- Coordinated laboratory training hubs for novel and endemic pathogens
- Joint training for human-animal laboratories

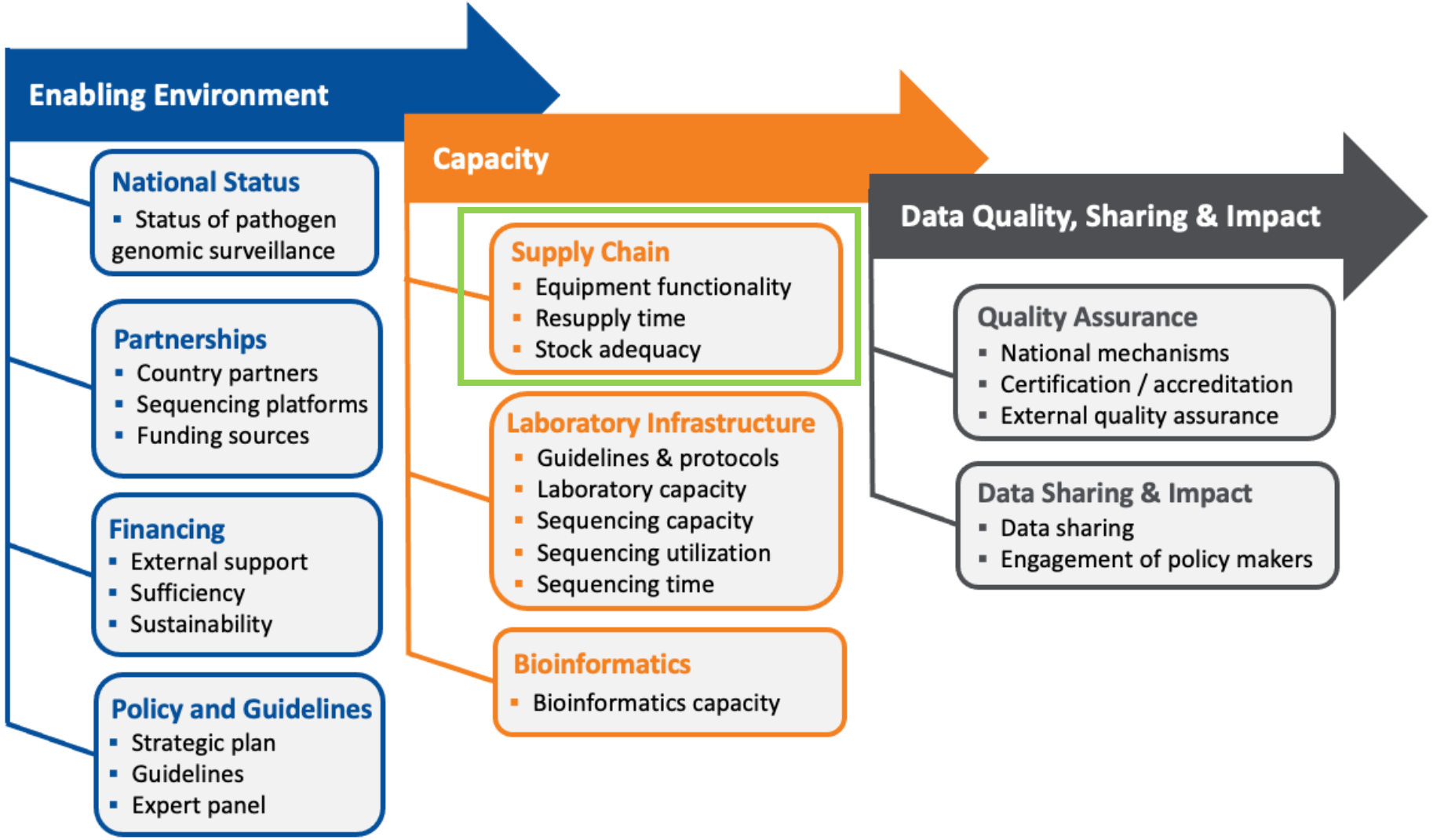
QUALITY ASSURANCE

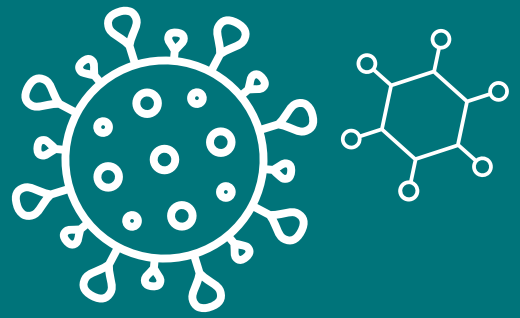
- Laboratory accreditation standards for pathogen genomics
- Establish low-cost regional external quality assurance (EQA) mechanisms

BIOINFORMATICS & DATA SHARING

- Regional bioinformatics capacity for in-house pipeline development
- Advance meta-data standards for pathogen genomics; align with global best-practice

NEXT: Pathogen Genomic Surveillance Capacity Discussion





THANK YOU!

WORKSHOP PARTNERS

CDC
U.S. CENTERS FOR DISEASE CONTROL AND PREVENTION

DukeNUS Medical School
ASIA PATHOGEN GENOMICS Initiative

CENTRE FOR PATHOGEN GENOMICS

THE UNIVERSITY OF SYDNEY
Sydney Infectious Diseases Institute

CIDM
centre for infectious diseases & microbiology
Public Health