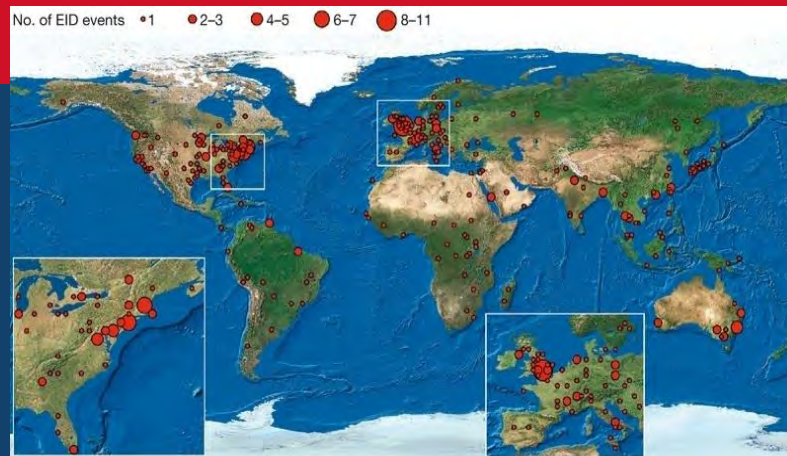


Next generation sequencing in diagnostic laboratories: opportunities and challenges

Vitali Sintchenko



Marie Bashir Institute
for
Emerging
Infectious Diseases
&
Biosecurity




- No conflict of interest to declare
-



THE UNIVERSITY OF
SYDNEY

We can sequence everything. Does it matter?

 Institute of Clinical Pathology and Medical Research <small>APA: Sydney West Area Health Service (SWAHS) 1053 Westmead Hospital, Hawkesbury Rd, Westmead 2145</small>		GENERAL ENQUIRIES: (02) 9845 6600		PATHOLOGY REQUEST FORM
		Result Enquiries: (02) 9845 8288		
Patient Details				
Medical Record No. 1130676	Location RMSH	Date of Birth 7/11/39	Sex M	
Last Name [REDACTED]		First Name [REDACTED]		
Street 36				
Clinical Information (Include drugs/therapy/travel that might affect results)				
<input type="checkbox"/> Fasting <input type="checkbox"/> Random <input type="checkbox"/> Pregnant				
[REDACTED]				
Swab Anal 13-Feb-17 NS1130676				
Client Details				
Name PALMS - MICROBIOLOGY				
PATHNET Code 0512		Phone [REDACTED]		Fax 94375746
Tests Requested / Specimen Type or Site				
<input type="checkbox"/> URGENT <input type="checkbox"/> Phone <input type="checkbox"/> Fax				
<i>whole genome sequencing.</i>				
SRN 20MAR 15:41				



Disruptive technology: NGS platforms

Platform	Output (Gb)	Read length (bp)	Reads (x10 ⁶)	Running time	Error rate
Illumina MiniSeq	0.6-7.5	2 x 150	25	4-24 h	0.1%
Illumina MiSeq	0.3-15	2 x 300	25	5-55 h	
Illumina NextSeq	20-120	2 x 150	130/400	12-32 h	
Illumina HiSeq	125-700	2 x 150	2500	1-6 days	
Ion PGM	0.03-2	200-400	0.4-5.5	2-7 h	1%
Ion S5/S5 XL	0.6-15	200-400	3-80	3-18 h	
Oxford Nanopore MinION	21-42	Up to 300,000	2.2-4.4	1min-48 h	4-12%
PacBio RSII	0.5-1	Up to 60,000	55,000	30min-4 h	?
PacBio Sequel	0.75-1.25	Up to 60,000	370,000	30min-6 h	



Relationship between quality and coverage depth

Coverage depth = Number of times a particular base pair is covered by sequenced reads

Reads

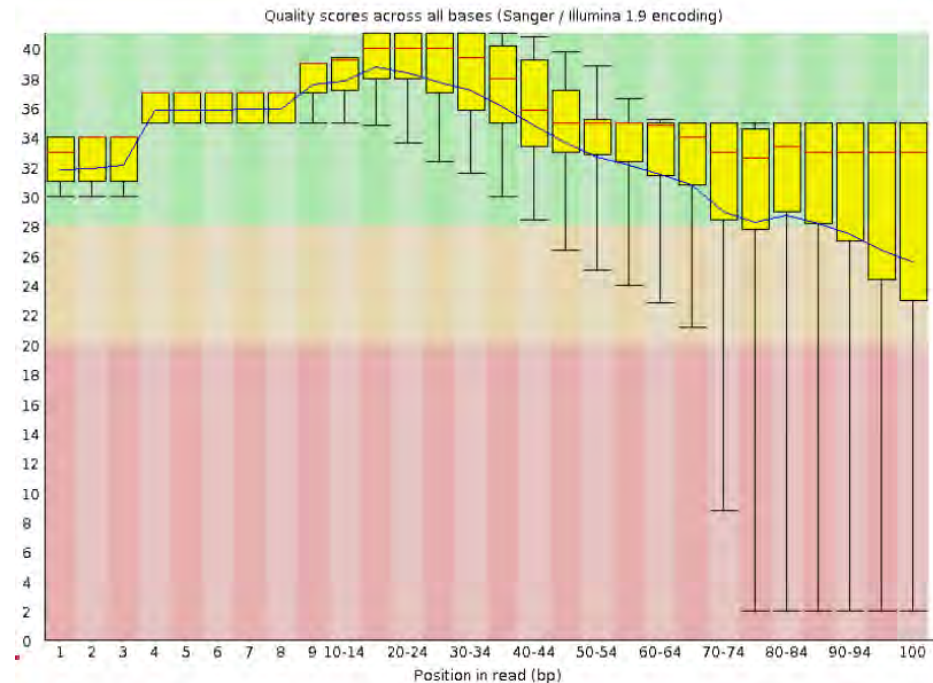
```

ATGGCATTGCAA
TGGCATTGCAATTTG
AGATGGTATTG
GATGGCATTGCAA
GCATTGCAATTTGAC
ATGGCATTGCAATTT
AGATGGTATTGCAATTTG
  
```

Consensus Sequence **AGATGGCATTGCAATTTGAC**

x7

$$Q = -10 \log_{10} P$$



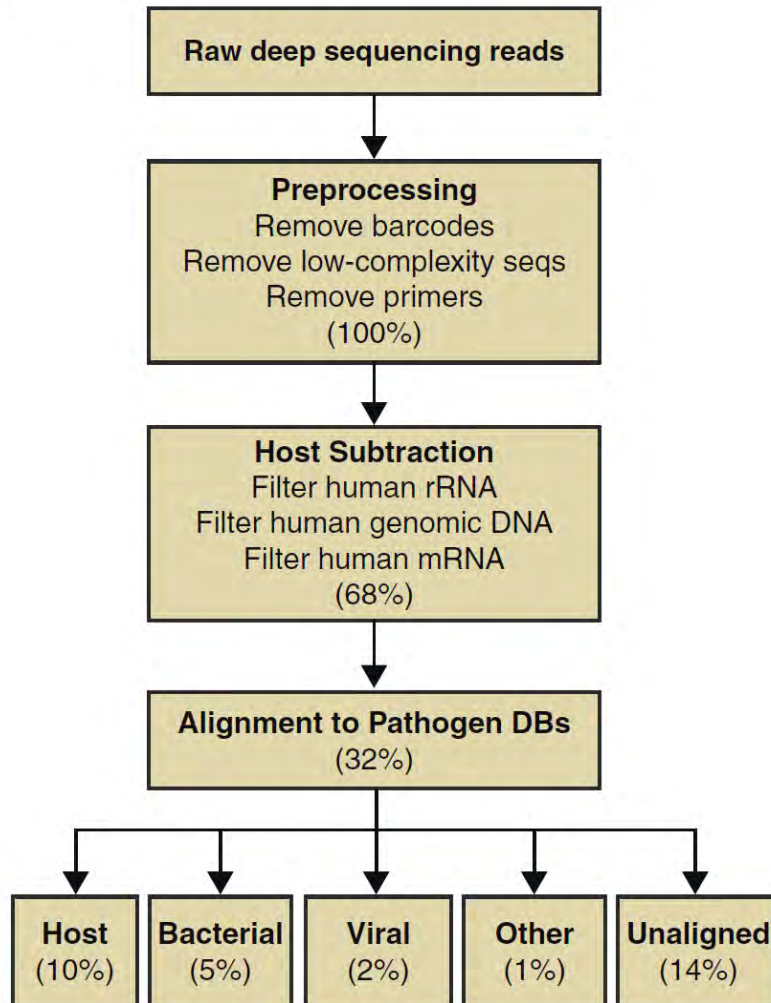
Phred score	Prob of incorrect base	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%
50	1 in 100000	99.999%

- Infectious disease genomics has been recognised as a subspecialty by NSW Health Pathology
- Applications of NGS
 - Pathogen discovery
 - Pathogen identification and characterization [e.g., detection of drug resistance associated substitutions (RAS)]
 - Public health surveillance
- Outdated databases and RAS catalogues can be dangerous for business





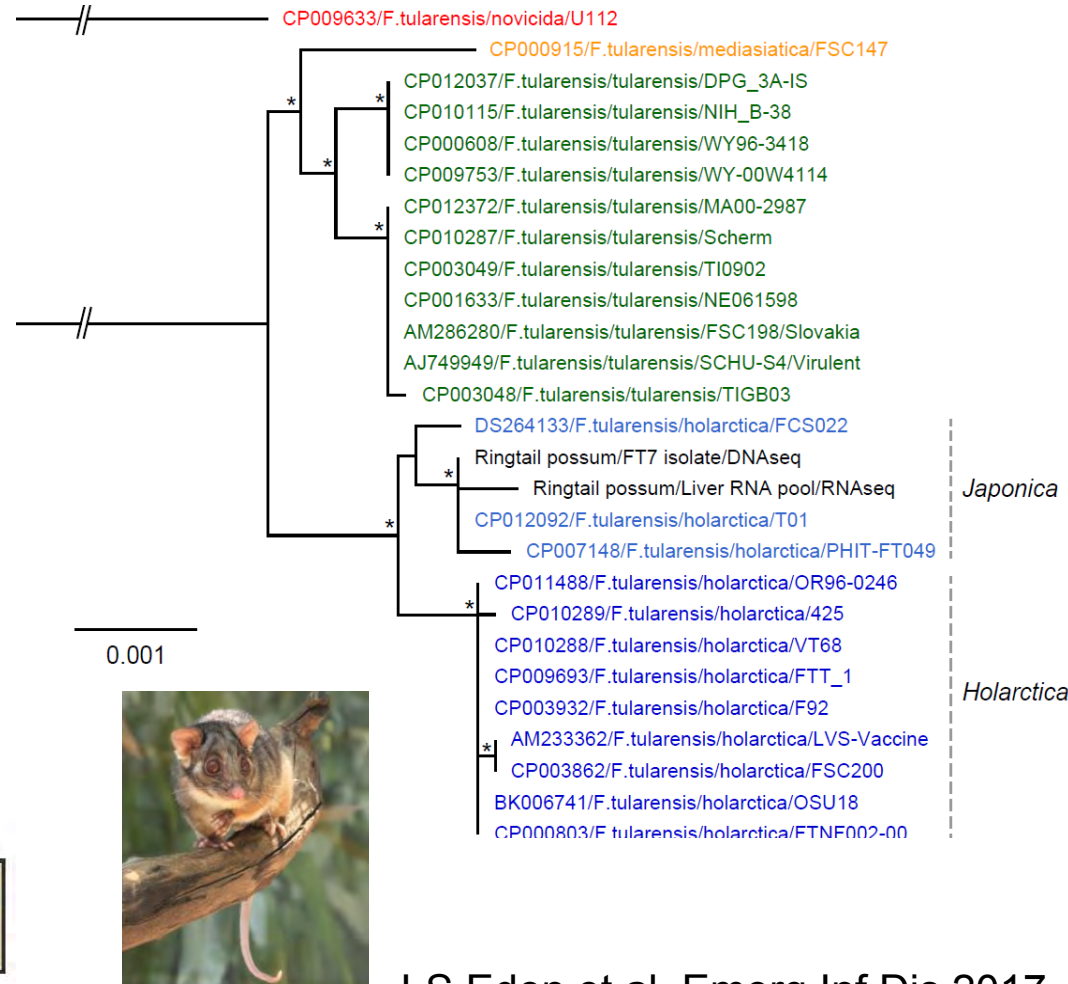
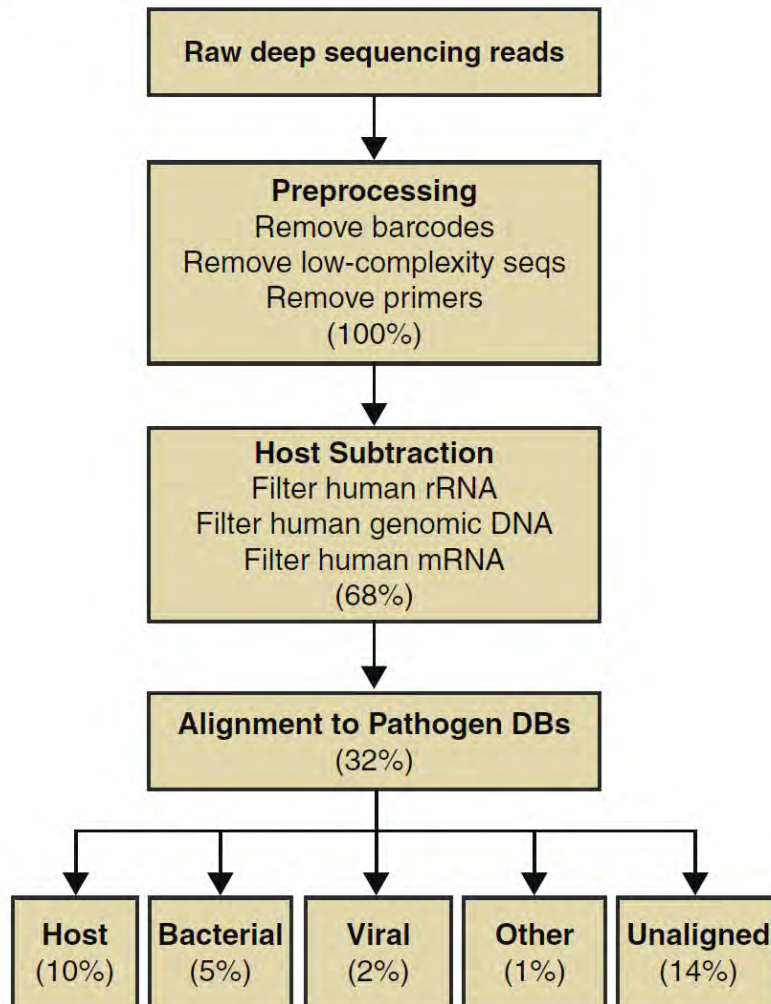
Power of unbiased pathogen discovery





Power of unbiased pathogen discovery

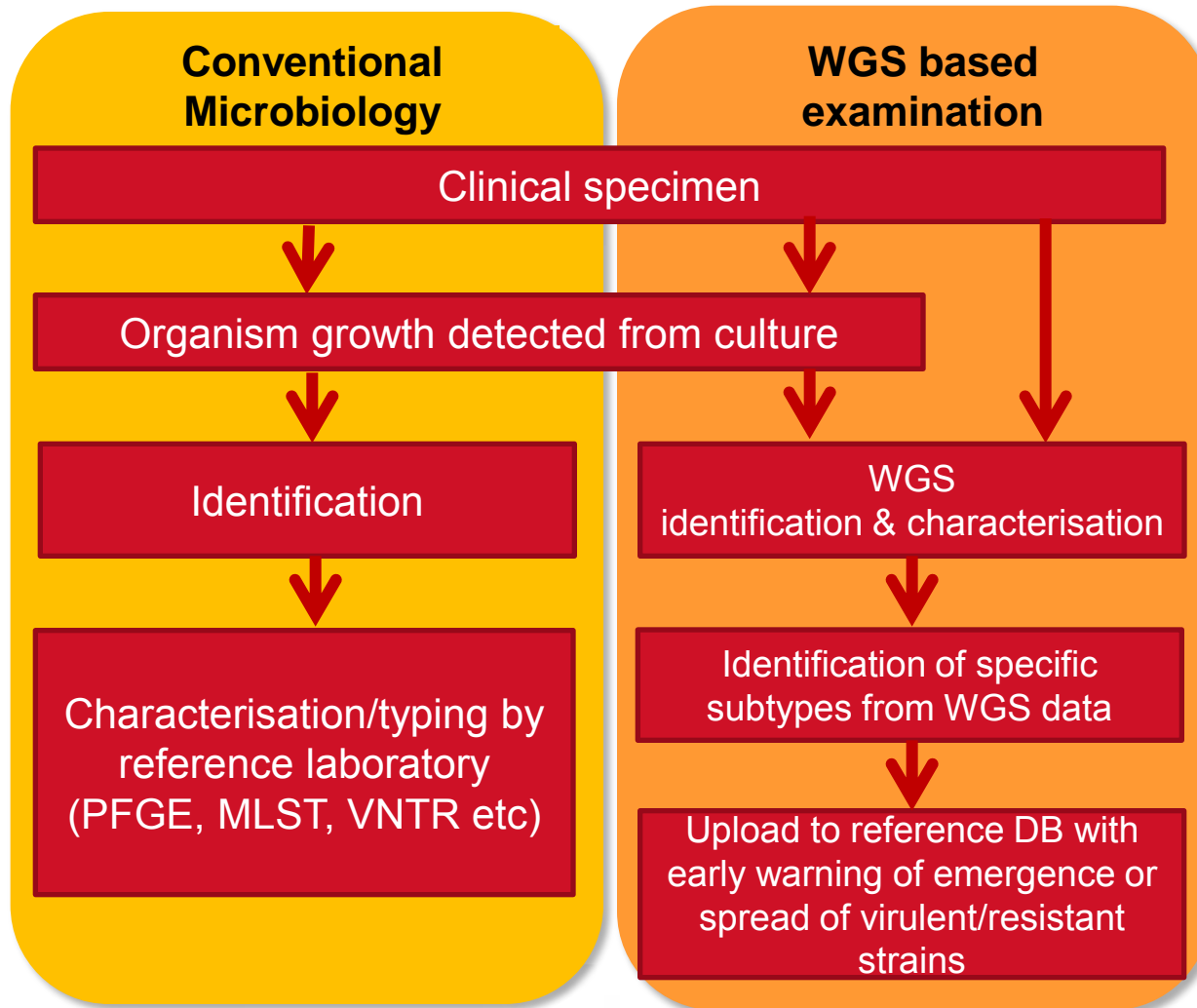
Dr Kerri Rose, NSW Wild Life Surveillance Program



J-S Eden et al. Emerg Inf Dis 2017



Pathogen Identification: From Pasteur to Watson





THE UNIVERSITY OF
SYDNEY

Genome sequencing as a lab automation

Modernising Medical Microbiology



UNIVERSITY OF
OXFORD



[Main](#) [News](#) [Patient and Public](#) [Projects](#) [People](#) [Papers](#) [Art](#) [Pathogens](#) [Media](#) [Funding](#) [Contact](#)

CRyPTIC Project

REHAB: The
Environmental Resistome

Infections in Oxfordshire
Research Database

CRyPTIC Project



BILL & MELINDA
GATES foundation

NSW Health Pathogen Genomics Partnership



Marie Bashir Institute for Emerging Infectious Diseases and Biosecurity (Tania Sorrell, Alicia Arnott)

Centre for Infectious Diseases and Microbiology-Public Health (Jon Iredell, Lyn Gilbert)

Pathology West-ICPMR (Dominic Dwyer, Sharon Chen, Peter Howard)

NSW Health (Vicky Sheppeard, Kirsty Hope, Chris Lowbridge)

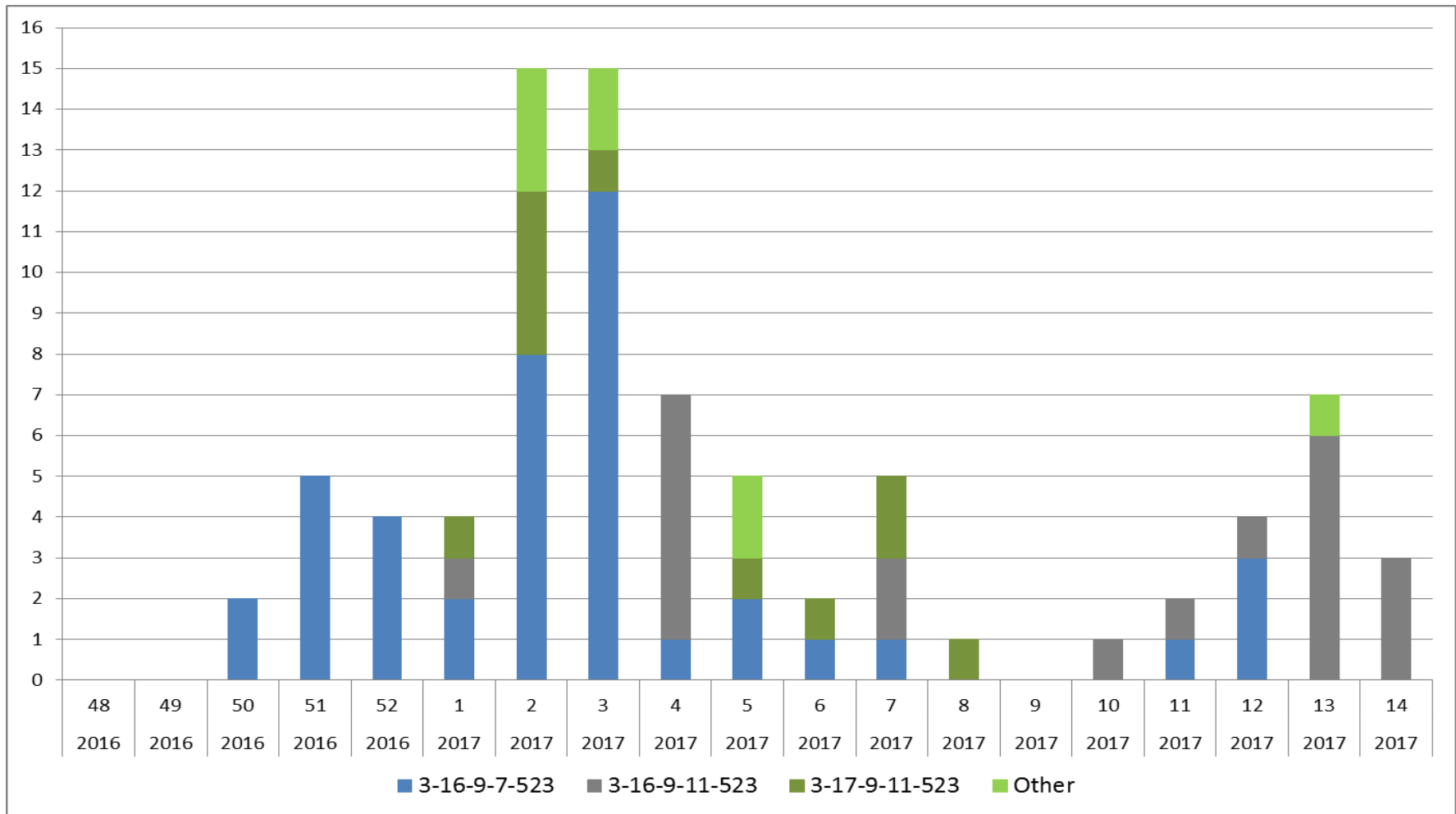
NSW Health TRG

Translating Pathogen Genomics into Improved Public Health Outcomes: Prospective evaluation of the effectiveness of genome sequencing-guided investigation of outbreaks





Genomics-guided laboratory surveillance



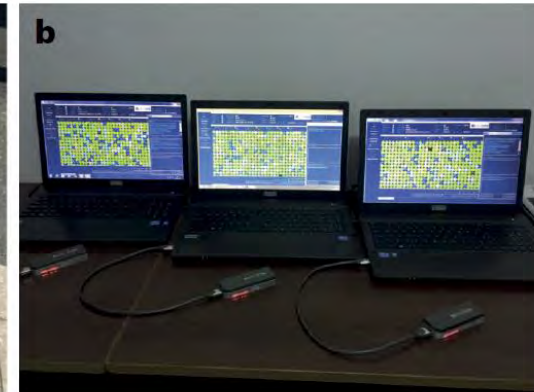


LETTER

doi:10.1038/nature16996

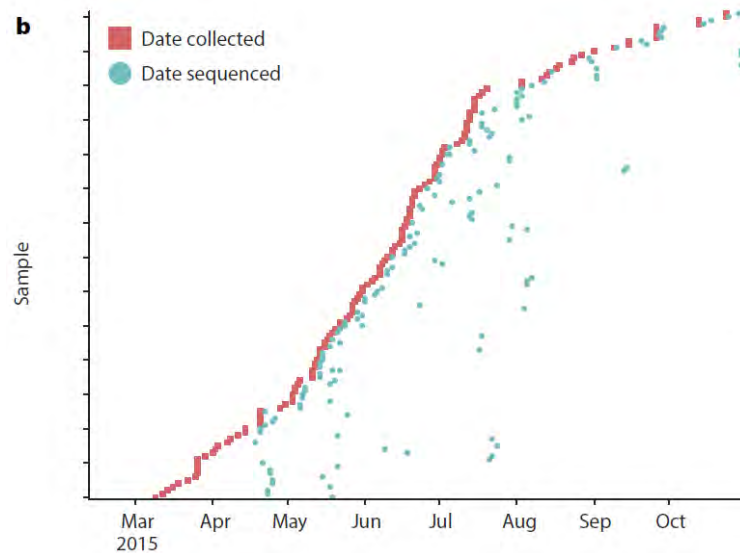
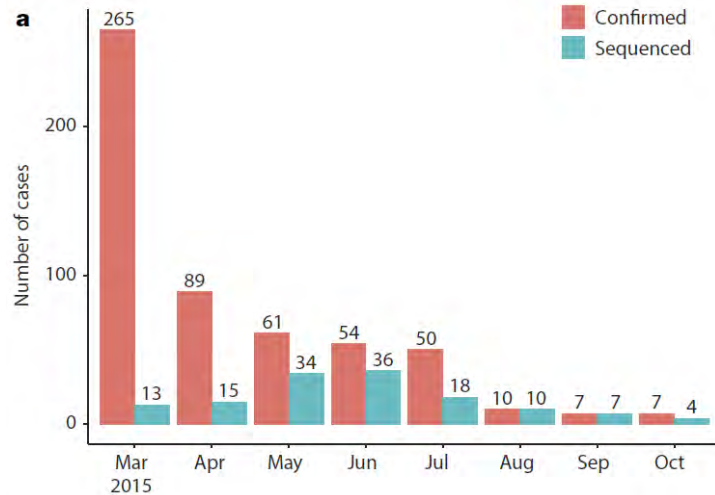
Real-time, portable genome sequencing for Ebola surveillance

Joshua Quick^{1*}, Nicholas J. Loman^{1*}, Sophie Duraffour^{2,3*}, Jared T. Simpson^{4,5*}, Ettore Joseph Akoi Bore², Raymond Koundouno², Gytis Dudas⁸, Amy Mikhail⁷, Nobila Ouédraogo^{2,11}, Amadou Bah^{2,11}, Jonathan H. J. Baum^{2,3}, Beate Becker-Ziaja^{2,3}, Jan Peter Boettcher^{2,12}, M. Alvaro Camino-Sánchez², Lisa L. Carter^{2,13}, Juliane Doerrbecker^{2,3}, Theresa Enkirch^{2,14}, Nicole Hetzelt^{2,12}, Julia Hinzmann^{2,12}, Tobias Holm^{2,3}, Liana Eleni Kafetzopoulou^{2,16}, Mikaela Eeva Kuisma^{2,10}, Christopher H. Logue^{2,10}, Antonio Mazzarelli^{2,19}, Sarah Meisel^{2,3}, Marc Didier Ngabo^{2,10}, Katja Nitzsche^{2,3}, Elisa Pallasch^{2,3}, Livia Victoria Patrono^{2,3}, Jasmine P. Natasha Y. Rickett^{2,15,23}, Andreas Sachse^{2,12}, Katrin Singethan^{2,24}, Inês Vitoriano^{2,10}, Rafaela Elsa G. Zekeng^{2,15,23}, Trina Racine²⁵, Alexander Bello²⁵, Amadou Alpha Sall²⁶, Ousmane N'Faly Magassouba²⁷, Cecelia V. Williams^{28,29}, Victoria Amburgey^{28,29}, Linda Winona^{28,29}, Frank Washington^{29,30}, Vanessa Monteil³¹, Marine Jourdain³¹, Marion Bererd³¹, Alimou Ca Abdoulaye Camara³¹, Marianne Gerard³¹, Guillaume Bado³¹, Bernard Baillet³¹, Deborah De Abdoulaye Diarra³⁴, Yacouba Savane³⁴, Raymond Bernard Pallawo³⁴, Giovanna Jaramillo Isabelle Roger³⁴, Christopher J. Williams^{6,37}, Facinet Yattara¹⁷, Kuiama Lewandowski¹⁰, Ja Daniel J. Turner³⁹, Georgios Pollakis^{15,23}, Julian A. Hiscox^{15,23}, David A. Matthews⁴⁰, Matt Andrew McD. Johnston⁴¹, Duncan Wilson⁴¹, Emma Hutley⁴², Erasmus Smir⁴³, Antonino I. Kilian Stoecker^{2,44}, Erna Fleischmann^{2,44}, Martin Gabriel^{2,3}, Simon A. Weller³⁸, Lamine K. Sakoba Keita¹⁷, Andrew Rambaut^{8,46,47}, Pierre Formenty³⁴, Stephan Günther^{2,3} & Miles W.

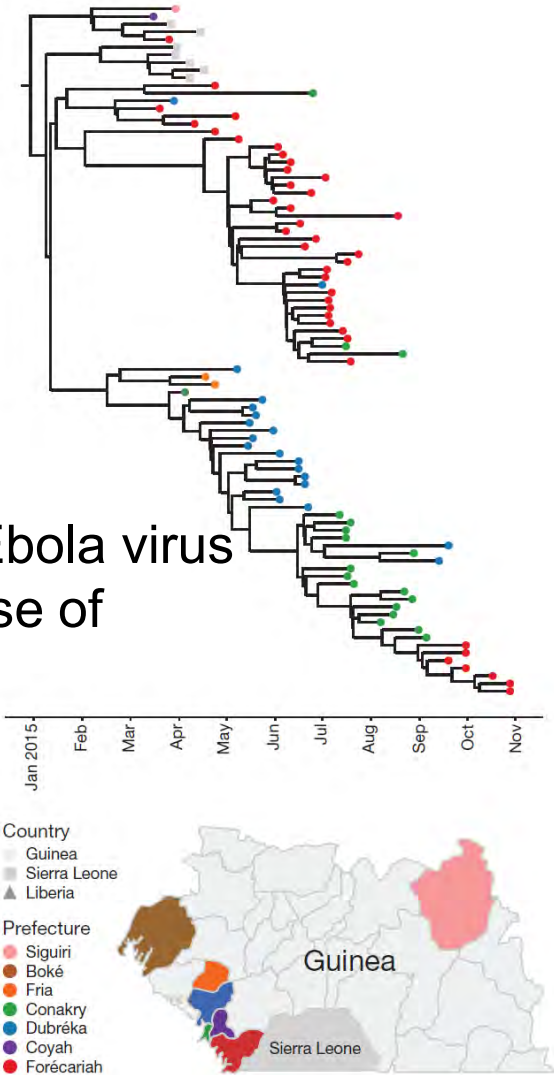




Monitoring Ebola outbreaks

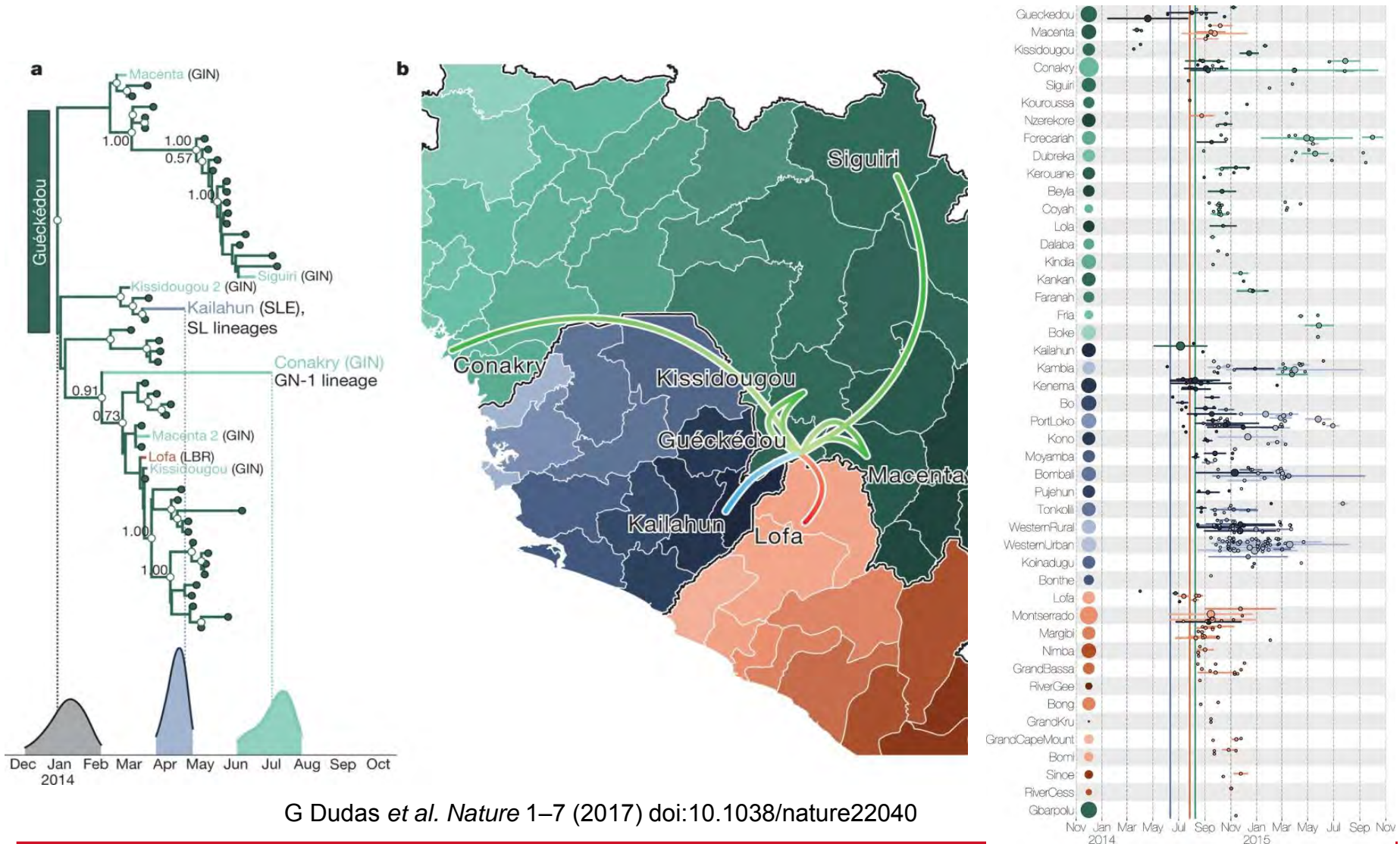


Evolution of Ebola virus
over the course of
epidemic



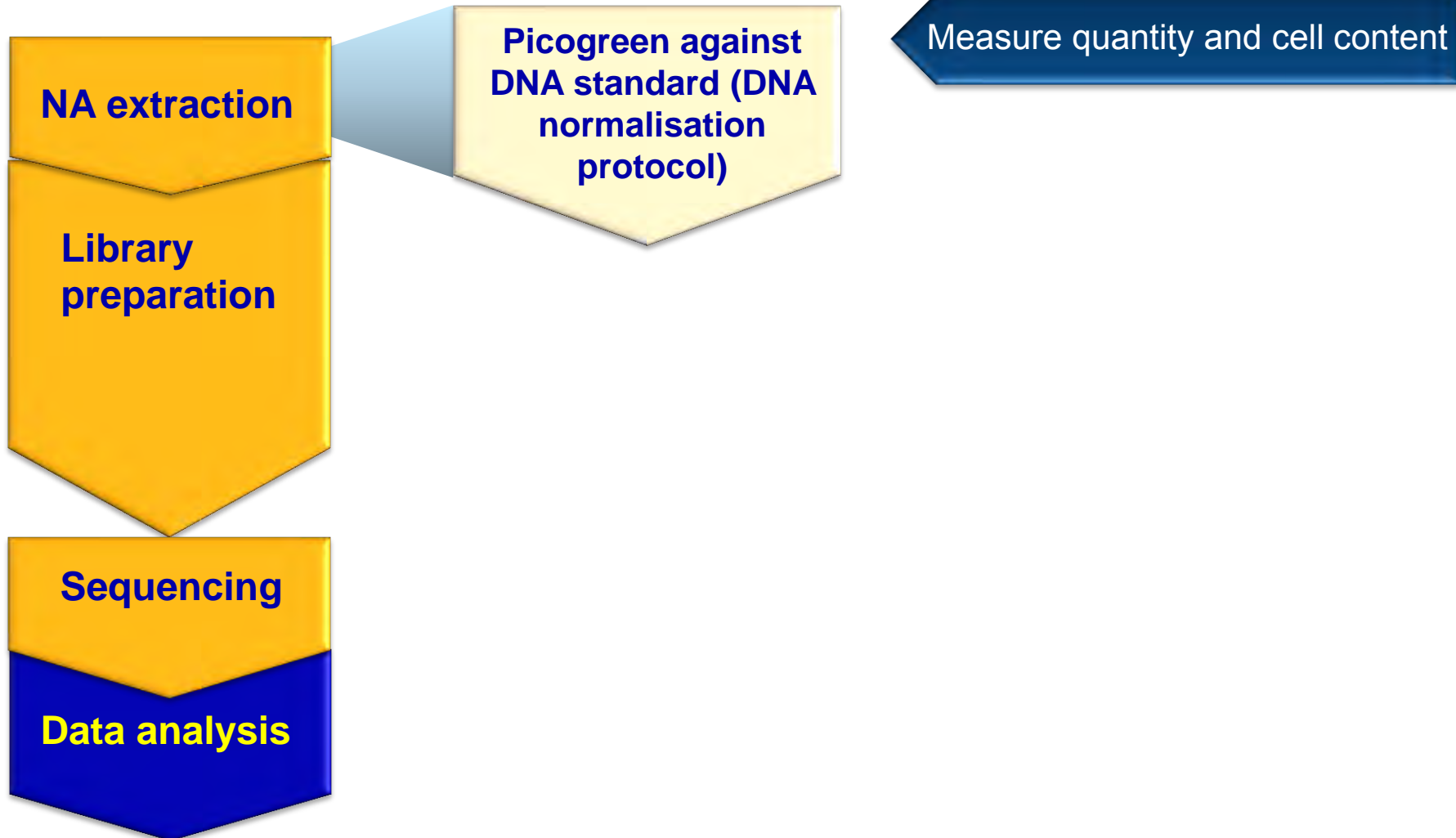


Temporal phylogeny and EBOV dispersal events



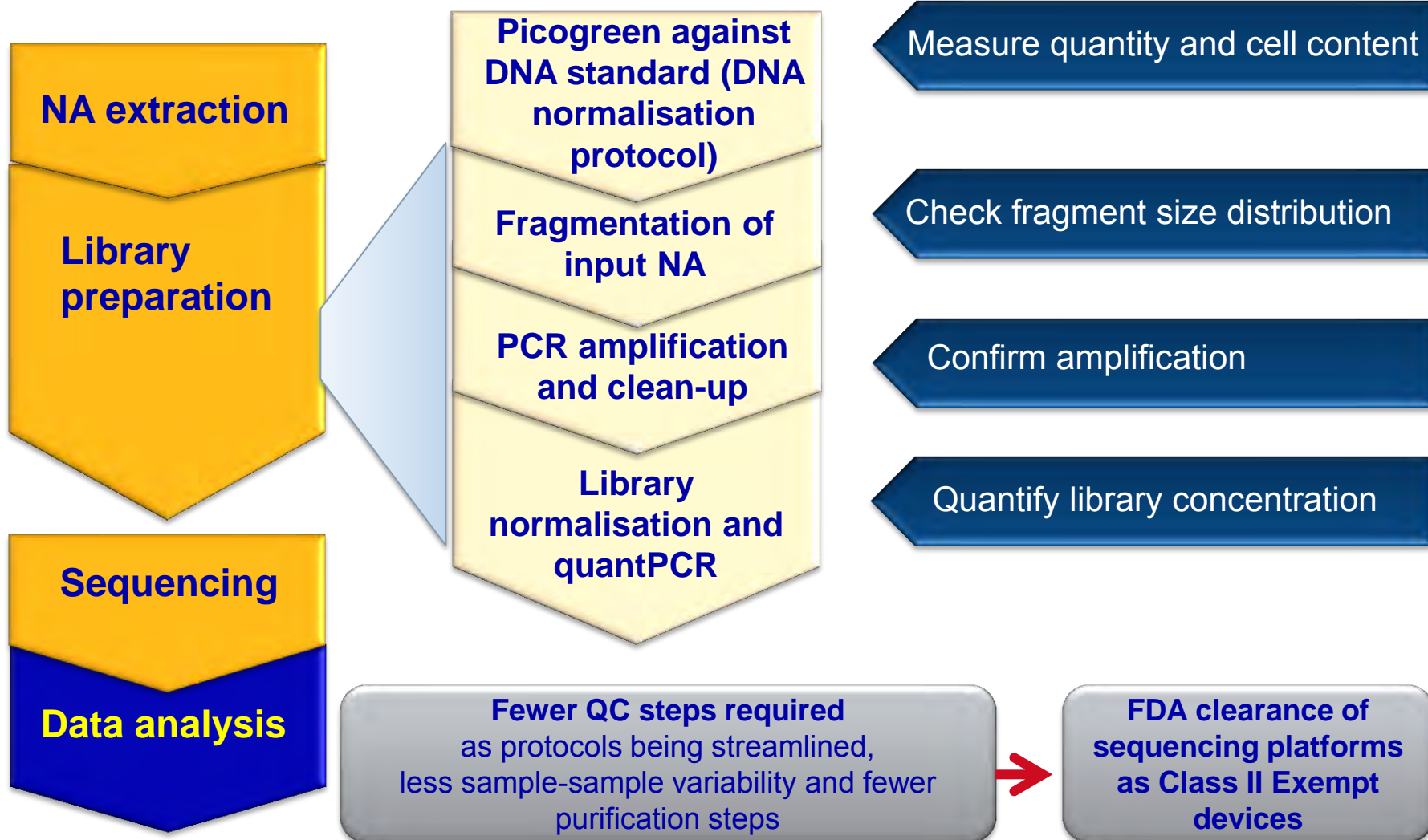


Good Laboratory Practice: 'Wet lab' NGS



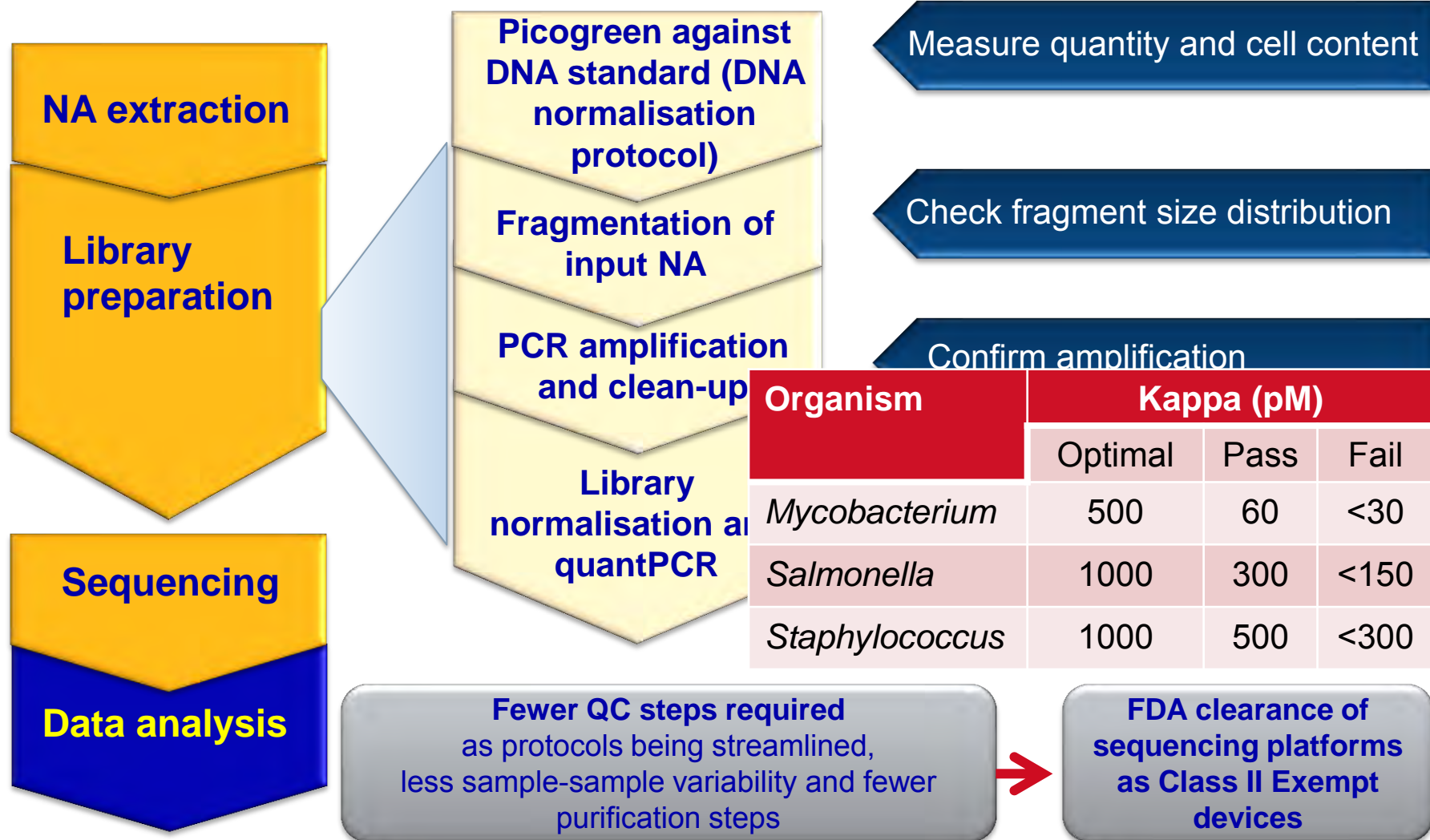


Good Laboratory Practice: 'Wet lab' NGS





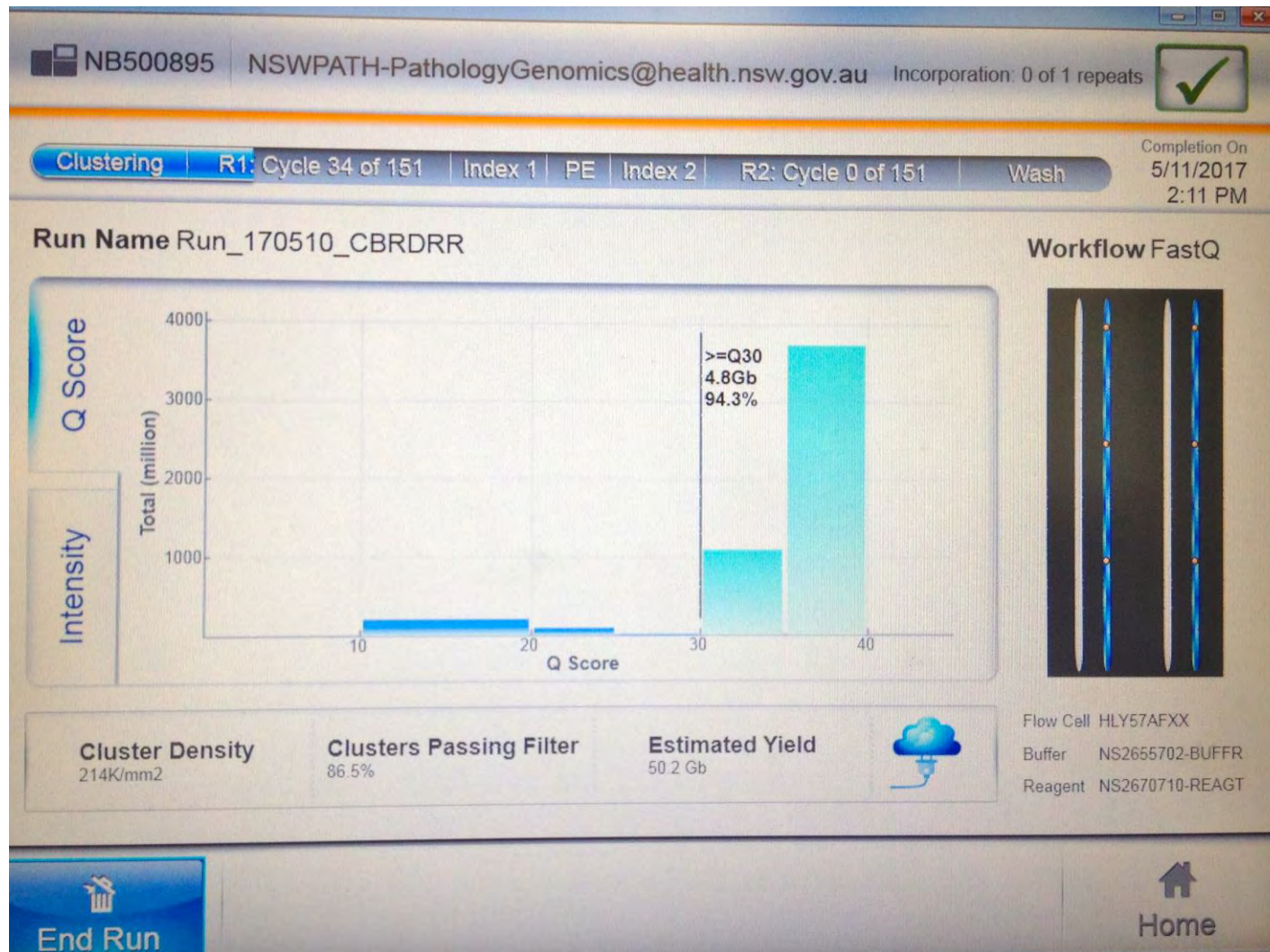
Good Laboratory Practice: 'Wet lab' NGS





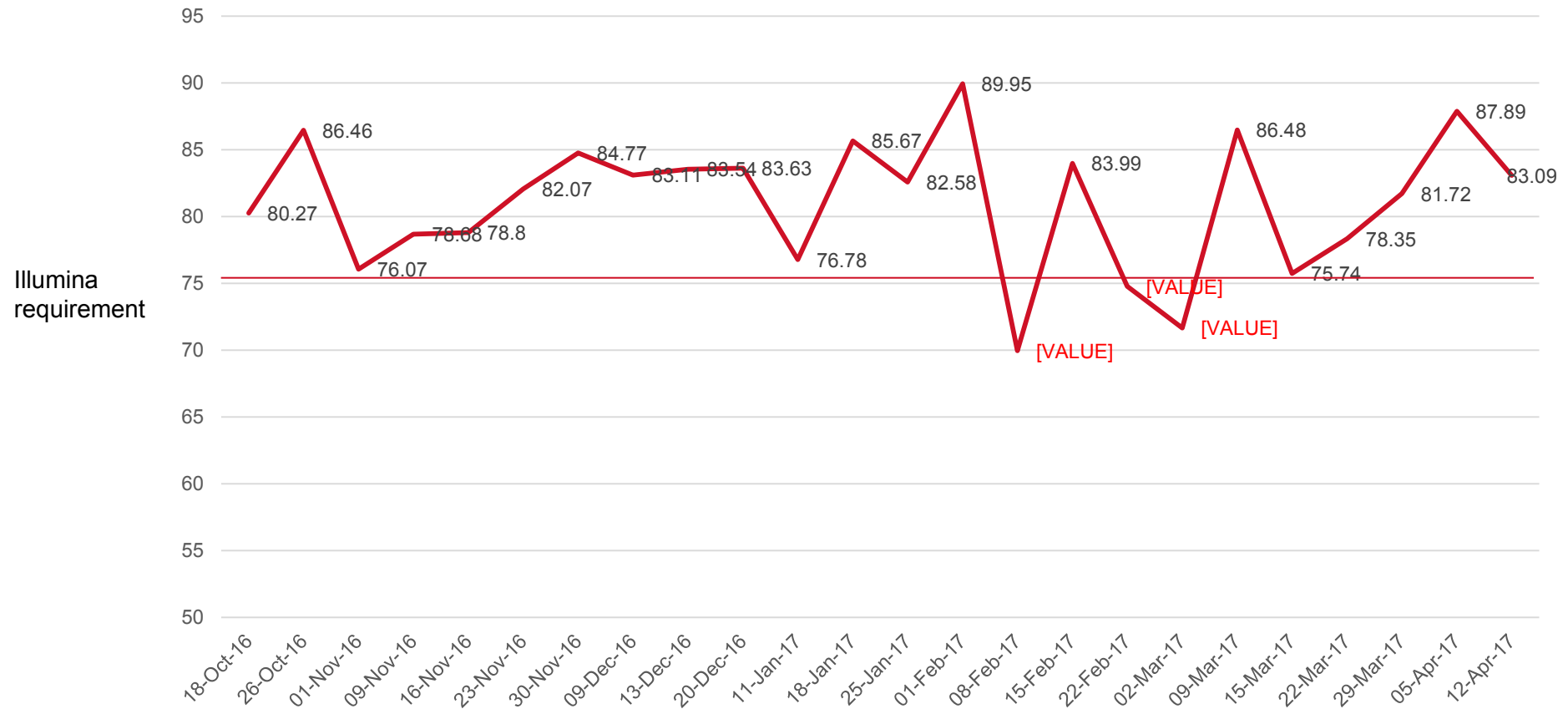
THE UNIVERSITY OF
SYDNEY

Assess sequencing run metrics





Q30 (i.e. 99.9% base call accuracy) values for TRG sequencing





Good Laboratory Practice: 'Dry lab' NGS

Automation

Primary analysis

Base calling
Data trimming

Read length, % reads with ~Q30
or above (e.g. FastQC)
*(performance specifications
established by manufacturer)*

FASTQ

Secondary analysis

De-multiplexing
Alignment [to a reference]
Assembly
Id of sequence variants

N50
No of reads mapped to reference
Coverage
% of SNPs in clusters
Multiplicity and parsimony

BAM and
VCF files

Tertiary analysis

Annotation
Filtering and classification
Reporting and retention of
lab results

% of intact MLST genes
Phylogenetic tree shapes
Concordance among prediction
programs
Credentialing of pipelines
Concordance among labs in PT

'Clinical-
grade'
VCF



	[Standard] Genomics	Pathogen discovery (RNASeq)
Pre-processing	Genomic DNA extraction & library prep	Additional host DNA depletion & microbial DNA enrichment
NGS output and TAT	1-2 GB 26-30 h	40 GB 14 days
Pathogen/host reads ratio	Low	High
Taxonomic accuracy	Species and subtypes	Species
Differentiation of virulent/MDR strains	+	+/-
Detection of point-source outbreaks, estimation of transmission pathways	+	?

Thank You!

Vitali.Sintchenko@sydney.edu.au