Prospects of WGS for Predicting Antimicrobial Resistance

Derrick Crook

Public Health England

University of Oxford

Oxford University Hospitals FT Trust

UNIVERSITY OF OXFORD

Public Health England

Conflicts of interest

- I receive an award from Jansen to support the Taiwanese mycobacterial laboratory contribute to the CRyPTIC consortium.

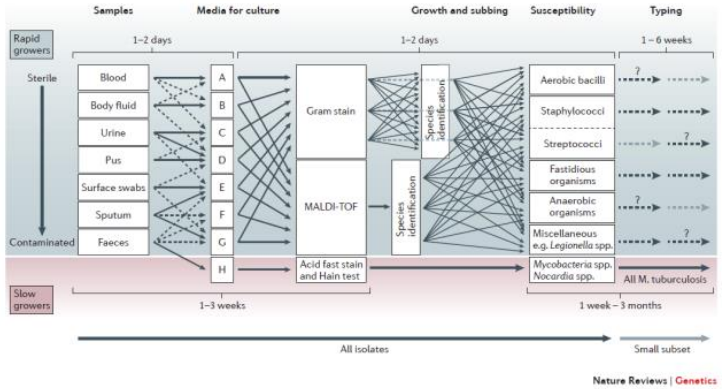
We can diagnose at two scales

- At the individual organism level
  - R vs S
    - Where it is difficult
    - Where it is straight-forward

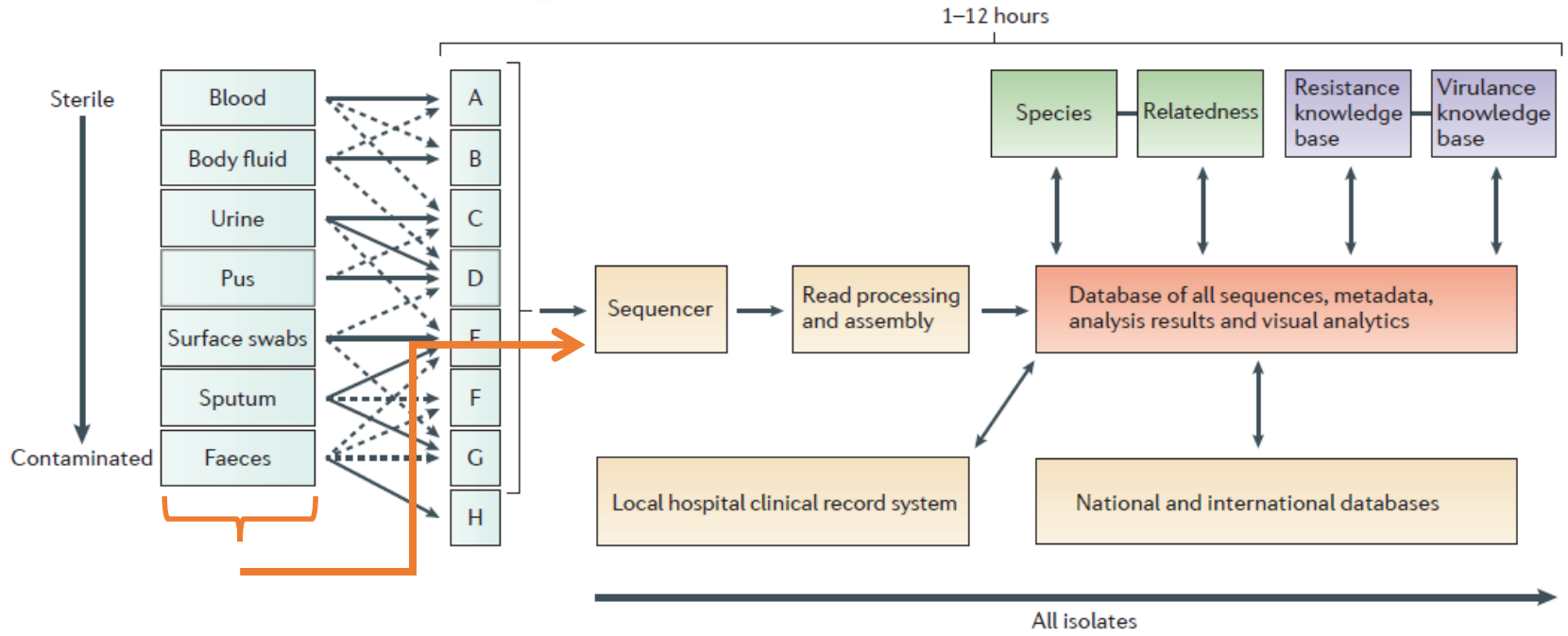
Exemplars to be presented

- Resistance prediction from difficult to more straight-forward
  - *Escherichia coli*
  - *Mycobacterium tuberculosis*
  - *Staphylococcus aureus*

Concept for ideal whole genome sequencing solution



In one step generate the complete diagnostic, typing and surveillance information



Nature Reviews Genetics 13, 601-612 (September 2012)

UNIVERSITY OF OXFORD

Public Health England

Resistance prediction from WGS

Iterative method of development

- A derivation set: compare genotypic prediction vs a gold-standard phenotypic susceptibility test
- Refine the catalogue and software
- A replication set: re-evaluate resistance prediction vs phenotype recording **very major** and **major errors**
- Analyse discrepant and improve the software, knowledge base and (if necessary) phenotypic methodology
- Test the revised algorithm with a fresh set of samples

E. coli

Sensitivity and specificity of genotypic resistance predictions versus gold standard "reference" phenotype results for 74 *Escherichia coli* bloodstream isolates

Table 3. Sensitivity and specificity of genotypic resistance predictions versus comparison with standard phenotype results for 74 *E. coli* bloodstream isolates.

	Susceptible by comparison standard phenotype		Resistant by comparison standard phenotype			
Antibiotic	susceptible by genotype (row %)	resistant by genotype (row %; major error)	susceptible by genotype (row %; very major error)	resistant by genotype (row %)	Sensitivity (95% CI)	Specificity (95% CI)
Amoxicillin	23 (31)	1 (1)	0 (0)	50 (68)	1.00 (0.91–1.00)	0.96 (0.77–1.00)
Co-amoxiclav	46 (62)	0 (0)	0 (0)	28 (38)	1.00 (0.85–1.00)	1.00 (0.90–1.00)
Gentamicin	60 (81)	0 (0)	0 (0)	14 (19)	1.00 (0.73–1.00)	1.00 (0.93–1.00)
Ciprofloxacin	48 (65)	0 (0)	0 (0)	26 (35)	1.00 (0.84–1.00)	1.00 (0.91–1.00)
Ceftriaxone	43 (58)	1 (1)	1 (1)	29 (39)	0.97 (0.81–1.00)	0.98 (0.87–1.00)
Ceftazidime	43 (58)	11 (15)	1 (1)	19 (26)	0.95 (0.73–1.00)	0.80 (0.66–0.89)
Meropenem	74 (100)	0 (0)	0 (0)	0 (0)	—	1.00 (0.94–1.00)
Total	337 (65)	13 (3)	2 (0.3)	166 (32)	0.99 (0.95–1.00)	0.96 (0.94–0.98)

J. Antimicrob. Chemother. (2013)

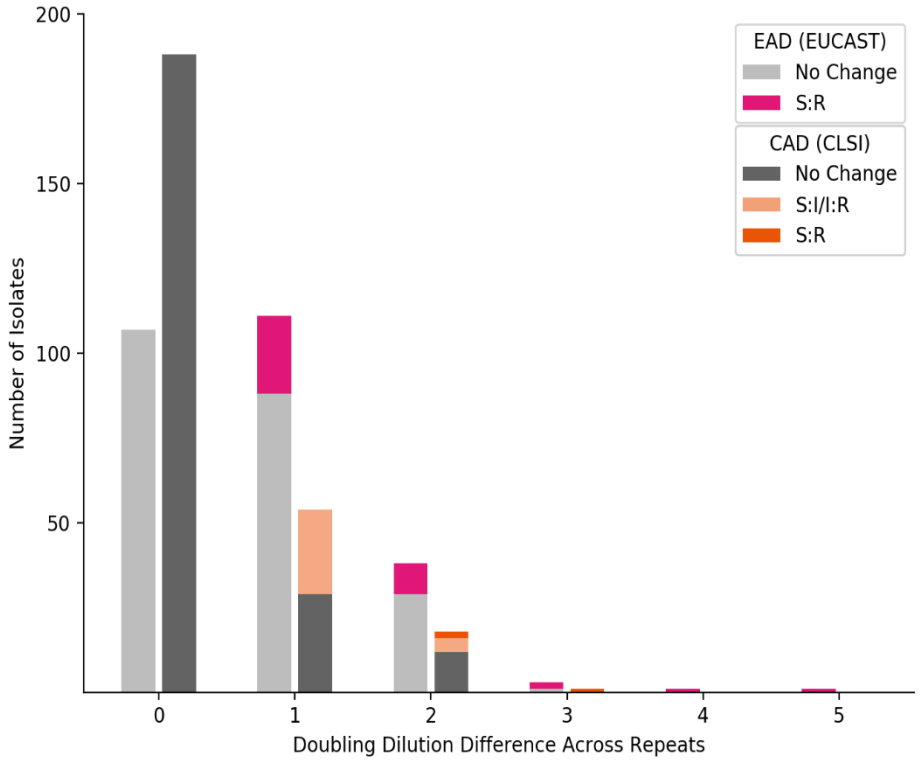
UNIVERSITY OF OXFORD

Public Health England

Co-amoxiclav reproducibility 261 isolates UKAS vs CLSI

- Significant within sample variation, worse using EUCAST guidelines

- Potential call changes
  - Worst Case Scenario
    - 76 EUCAST
    - 48 CLSI S:NS (I or R)
    - 6 CLSI S:R

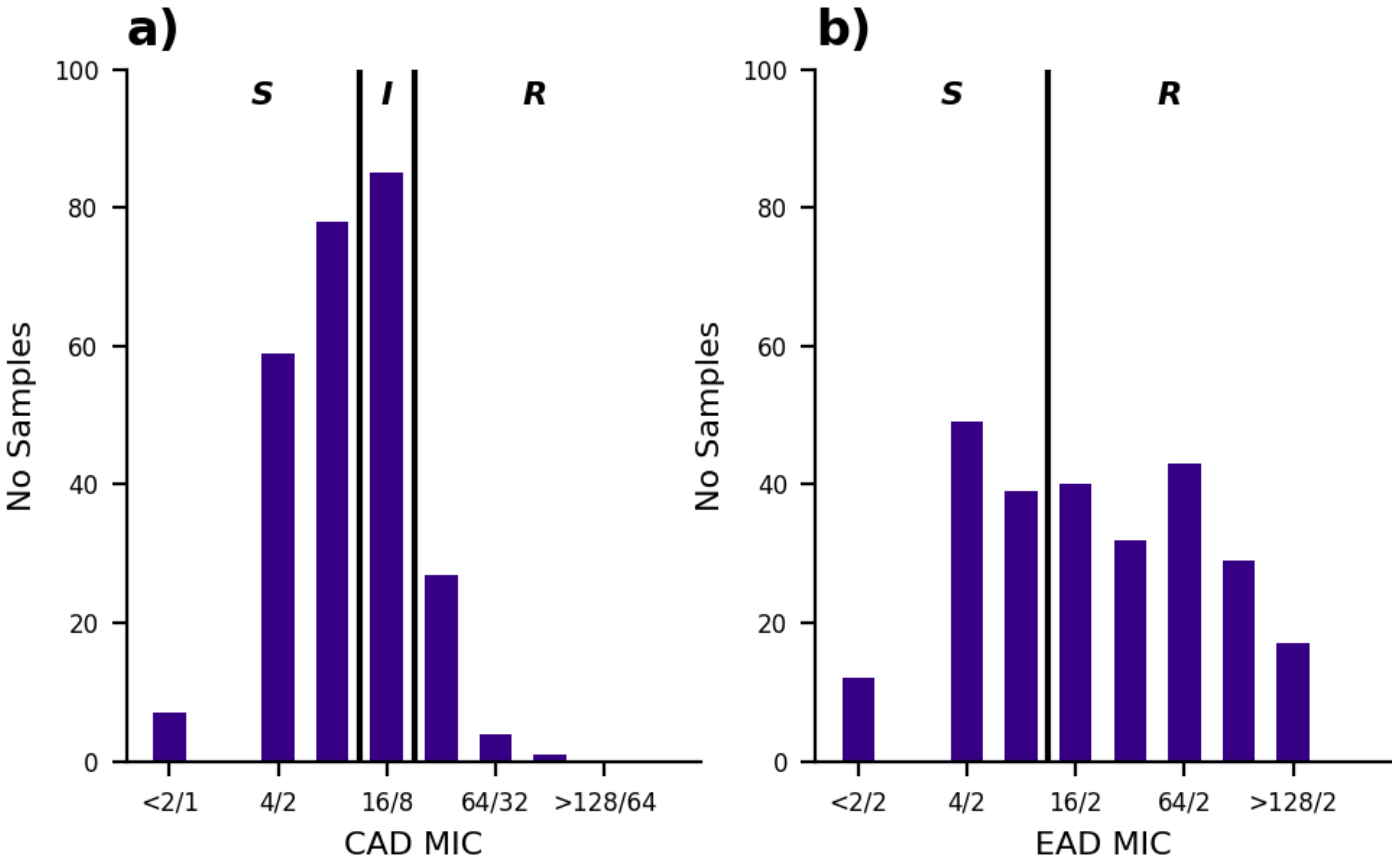


UNIVERSITY OF OXFORD

Public Health England

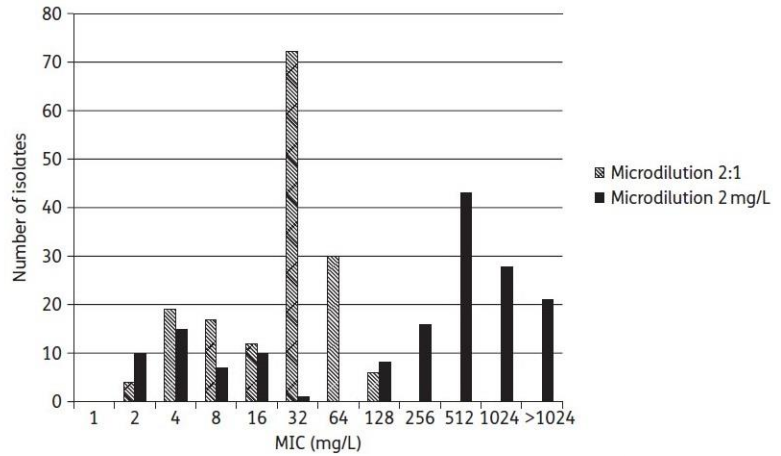
Disparity in Coamoxiclav phenotype UKAS vs CLSI

261 isolates by agar incorporation MIC in triplicate



This fails categorical agreement

Diez-Aguilar et al. JAC (2015)



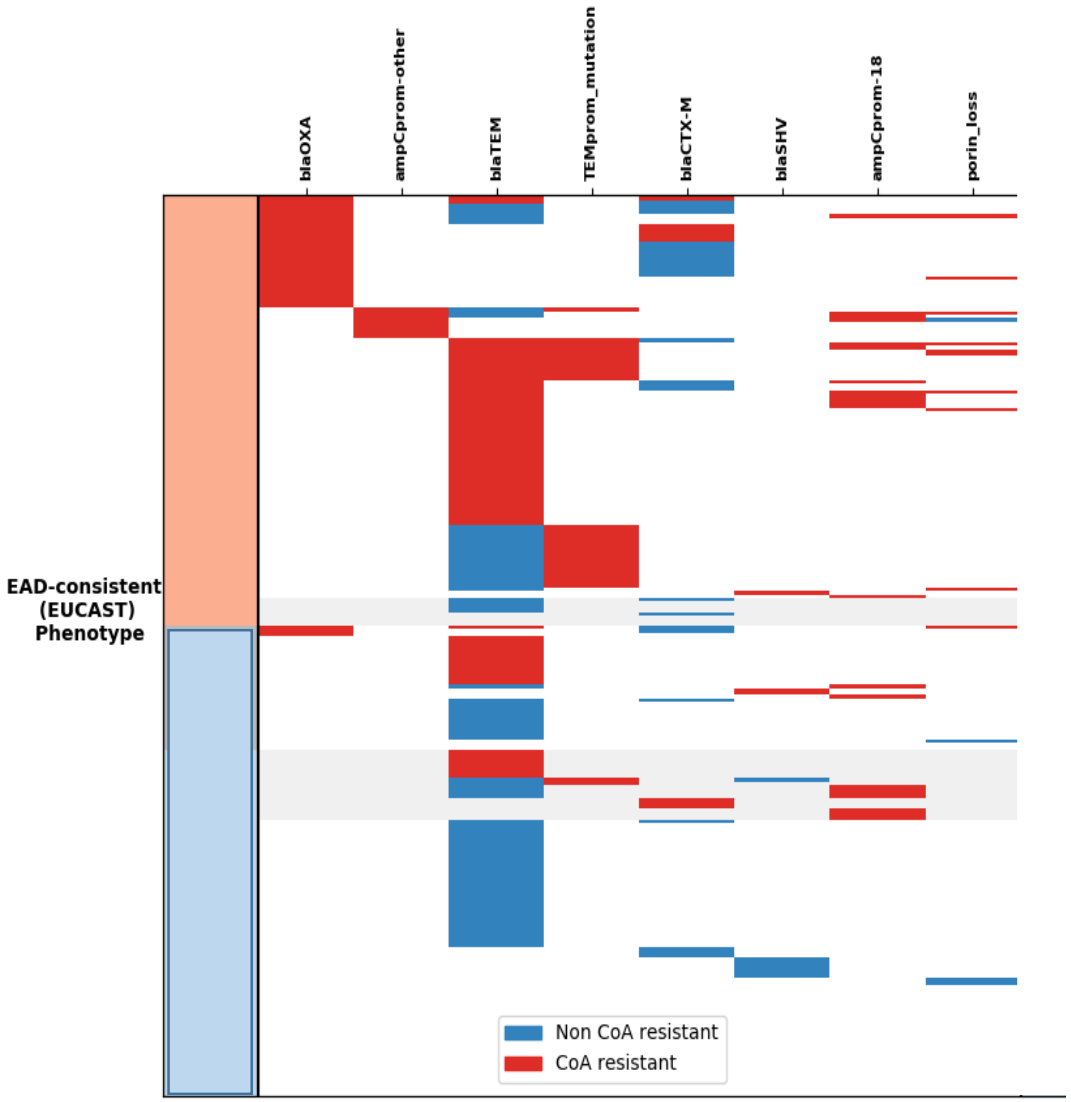
UNIVERSITY OF OXFORD

Public Health England

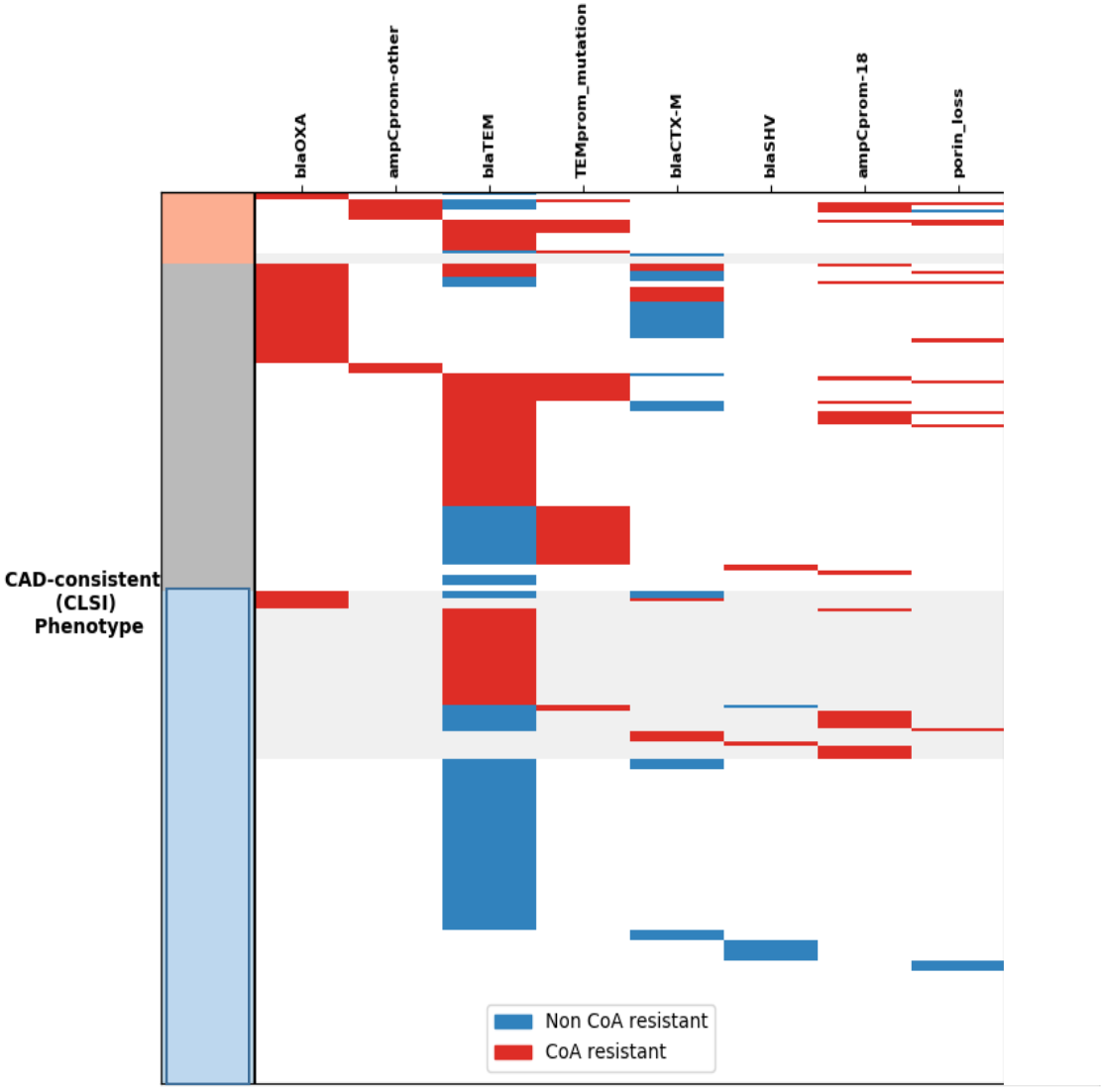
The catalogue (knowledge base) of variation

Slide removed

Depiction of categorical results vs geno-prediction



Very major error > 10 % Major error >10%



Very major error > 3% Major error >10%

OXFORD

Public Health England

Multivariate model investigating independent effects of each mechanism

Slide removed

Conclusion

- We don’t have a reference standard for genotypic prediction
- There is large uncertainty about the truth

Mycobacterium tuberculosis

Anti-tuberculosis drug resistance prediction

- Arguably 15 drugs are available for treating TB with more new drugs in development
- Is genomic variation which confers resistance limited to somewhere between 20 to 30 genes?
- Current knowledge indicates molecular prediction of INH, rifampicin resistant or pan-susceptible isolates is ~ 95% accurate
- The knowledge base of variation conferring resistance to ‘all drugs’ is incomplete

Can we discover explanatory variation in TB?


Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study

Timothy M Walker*, Thomas A Kohl*, Shaheed V Omar*, Jessica Hedge*, Carlos Del Ojo Elias, Phelim Bradley, Zamin Iqbal, Silke Feuerriegel, Katherine E Niehaus, Daniel J Wilson, David A Clifton, Georgia Kapatai, Camilla L C Ip, Rory Bowden, Francis A Drobniewski, Caroline Allix-Béguec, Cyril Gaudin, Julian Parkhill, Roland Diel, Philip Supply, Derrick W Crook, E Grace Smith, A Sarah Walker, Nazir Ismail†, Stefan Niemann†, Tim E A Peto†, and the Modernizing Medical Microbiology (MMM) Informatics Group

- Investigation of 3651 isolates :
  - Using a heuristic method of predicting resistance
- divided into
  - a 2099 derivation set
  - a 1552 validation set
- Resistance is conferred by genomic variation:
  - Non-synonymous mutations , deletions and insertions in relevant genes – 23 genes
  - Arises mostly de-novo in a non-recombining genome leading to homoplasy

Lancet Infect Dis 2015; 15: 1193–1202

TB drug resistance prediction in a validation set

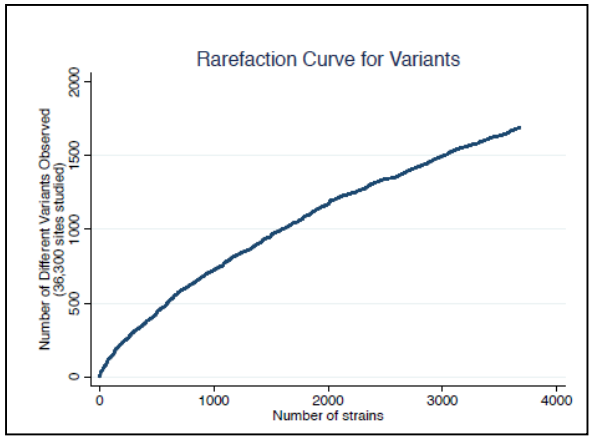
	Phenotypically Resistant Genotype					Phenotypically Sensitive Genotype					All		Excluding Unclassfied		
	R	S_0	S_s	U	Total	R	S_0	S_s	U	Total	Sensitivity	Specificity	Sensitivity	Specificity	% Unclassifed
Isoniazid	310	18	1	35	364	19	1,065	52	52	1188	85.2	98.4	94.2	98.3	5.6
Rifampicin	275	8	1	16	300	10	1,200	4	38	1252	91.7	99.2	96.8	99.2	3.5
Ethambutol	158	7	1	26	192	67	1003	79	210	1359	82.3	95.1	95.2	94.2	15.2
Pyrazinamide	43	27	5	104	179	2	1,218	67	83	1370	24.0	99.9	57.3	99.8	12.1
Streptomycin	284	6	9	49	348	11	970	34	189	1204	81.6	99.1	95.0	98.9	15.3
Ofloxacin	5	4	2	0	11	0	489	134	38	661	45.5	100.0	45.5	100.0	5.7
Amikacin	52	5	0	2	59	3	427	38	140	608	88.1	99.5	91.2	99.4	21.3
Total	1127	75	19	232	1453	112	6372	408	750	7642	77.6	98.5	92.3	98.4	10.8

Table 1: Genotypic predictions in the validation-set based on: R (resistance-determinant); S0 (zero non-synonymous variants/SNPs present); Ss (only sensitive variants present); U (unclassified variants present). Weighted mean sensitivity and specificity given for all phenotypes, and with the 10.8% of phenotypes associated with previously unclassified variation (U) excluded.

Lancet Infect Dis 2015; 15: 1193–1202

Filling the resistance gap

Comprehensive Resistance Prediction for Tuberculosis: an International Consortium (CRyPTIC)



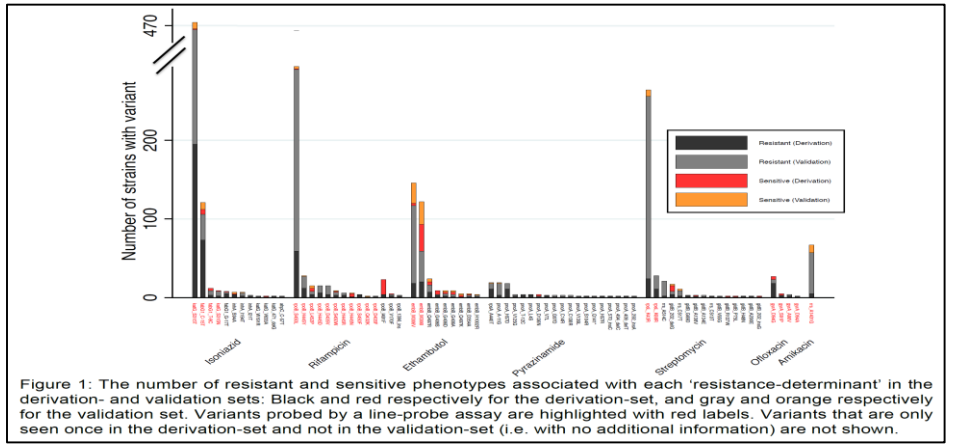
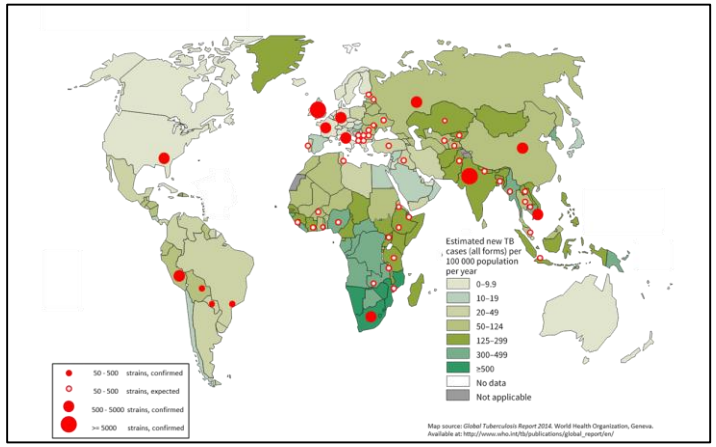


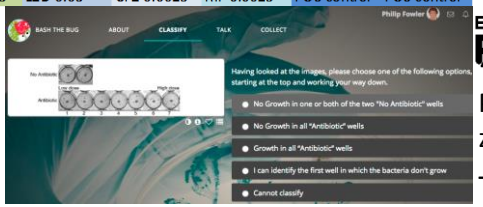
Figure 1: The number of resistant and sensitive phenotypes associated with each 'resistance-determinant' in the derivation- and validation sets: Black and red respectively for the derivation-set, and gray and orange respectively for the validation set. Variants probed by a line-probe assay are highlighted with red labels. Variants that are only seen once in the derivation-set and not in the validation-set (i.e. with no additional information) are not shown.



Phenotyping

BDQ 2	KAN 16	KAN 8	KAN 4	KAN 2	KAN 1	ETH 8	ETH 4	ETH 2	ETH 1	ETH 0.5	ETH 0.25
BDQ 1	AMI 8	EMB 8	INH 1.6	LEV 8	MXF 4	DLM 1	LZD 2	CFZ 4	RIF 4	RFB 2	PAS 4
BDQ 0.5	AMI 4	EMB 4	INH 0.8	LEV 4	MXF 2	DLM 0.5	LZD 1	CFZ 2	RIF 2	RFB 1	PAS 2
BDQ 0.25	AMI 2	EMB 2	INH 0.4	LEV 2	MXF 1	DLM 0.25	LZD 0.5	CFZ 1	RIF 1	RFB 0.5	PAS 1
BDQ 0.125	AMI 1	EMB 1	INH 0.2	LEV 1	MXF 0.5	DLM 0.125	LZD 0.25	CFZ 0.5	RIF 0.5	RFB 0.25	PAS 0.5
BDQ 0.06	AMI 0.5	EMB 0.50	INH 0.1	LEV 0.5	MXF 0.25	DLM 0.06	LZD 0.125	CFZ 0.25	RIF 0.25	RFB 0.125	PAS 0.25
BDQ 0.03	AMI 0.25	EMB 0.25	INH 0.05	LEV 0.25	MXF 0.125	DLM 0.03	LZD 0.06	CFZ 0.125	RIF 0.125	RFB 0.0625	PAS 0.125
BDQ 0.015	EMB 0.0625	EMB 0.125	INH 0.025	LEV 0.125	MXF 0.0625	DLM 0.015	LZD 0.03	CFZ 0.0625	RIF 0.0625	POS control	POS control

Pyrazinamide will be done by MGIT liquid culture



BASH THE BUG

People powered research
zooniverse.org

Twitter: @bashthebug

Genotypic characterisation

- 100,000 WGS TB pledged
- ~ 40,000 with extensive DST
- Analysis:
  - Heuristic approach
  - GWAS
  - Machine Learning
  - Thermodynamic modelling of proteins
  - Molecular genetic characterisation

UNIVERSITY OF OXFORD

BILL & MELINDA GATES foundation

wellcome trust

Public Health England

How good can we get when we analyse >10,000 isolates to the 4 first line drugs

Predicting susceptibility to four 1st line drugs

Slide removed

Sequence based prediction for four first line drugs is highly specific

- Error rates (i.e. falsely predicting susceptibility) are very low <2% with very tight confidence intervals for the 4 first line drugs
- Now need to establish the status of the other drugs

S. aureus

S. aureus: Resistance prediction algorithm

- Derivation set of 501 samples
- Algorithm was refined after the derivation set.
- Many of the discrepant results were found to be phenotypic errors in the routine laboratory.
- Other discrepants were resolved by improvements in the bio-informatics software
- The improved algorithm was tested against a further 487 isolates (the ‘validation’ set).

Gordon et al J Clin Microbiol. 2014 Feb 5

Blinded validation study of resistance prediction from WGS *Staphylococcus aureus* (478)

	Phenotype: resistant		Phenotype: susceptible		Error Rates	
	Genotype		Genotype		ME	VME
Antimicrobial	Susceptible	Resistant	Susceptible	Resistant	(%)	(%)
Penicillin	2	398	84	3	3.4	0.5
Methicillin	0	55	432	0	0.0	0.0
Ciprofloxacin	2	64	421	0	0.0	3.0
Erythromycin	1	80	404	2	0.5	1.2
Clindamycin	1	76	2	0	0.0	1.3
Tetracycline	0	18	467	2	0.4	0.0
Vancomycin	0	0	491	0	0.0	n/a
Fusidic acid	1	39	445	0	0.0	2.6
Trimethoprim	0	2	200	1	0.5	0.0
Gentamicin	1	2	484	0	0.0	33.3
Mupirocin	0	2	485	0	0.0	0.0
Rifampicin	0	5	482	0	0.0	0.0
Total	8	741	4397	8	0.2	1.1

Gordon et al J Clin Microbiol. 2014 Feb 5

UNIVERSITY OF OXFORD

Public Health England

Previous phenotyping studies

Study	Comparison	no of isolates	Categorical agreement (%)	ME rate (%)	VME rate (%)
Ligozzi 2002	Vitek 2 vs agar dilution	100	94-100	0	0
Fahr 2003	BD Phoenix vs broth dilution plus mecA PCR	116	97.6	1.2	1.7
Nonhoff 2005	Vitek 2 vs agar dilution	273	-	1.5	0.7
Carroll 2006	BD Phoenix vs agar dilution	232	98.2	0.3	0.4
Giani 2012	BD Phoenix vs broth dilution	95	98	1.3	2.1
Bobenchik 2014	Vitek 2 vs broth dilution	134	98.9	0.1	1.4
This study	WGS vs combined disc diffusion / BD Phoenix	491	98.8	0.2	1.1

Resistance prediction is looking very promising

- Combination of β-lactam and β-lactamase inhibitor is a major problem for phenotyping and resistance prediction
- For TB, further extensive work on discovering all the variation conferring resistance needs to be done
- For *S. aureus* further validation is needed, but results appear very good

Acknowledgements

- Sarah Walker
- Rosalind Harding
- Tim Peto
- Neil Woodford
- Mark Wilcox – Leeds
- Grace Smith – Birmingham
- Philip Monk - Leicester
- John Paul – Brighton
- Martin Llewellyn – Brighton
- Research Fellows (6)
- Tim Davies
- Amy Mathers - Uva, USA

Microbiology, DNA preparation

- Dai Griffiths
- Kate Dingle
- Nicole Stoesser
- Alison Vaughan
- Bernadette Young
- Claire Gordon

International

- Stefan Niemann Tom Rogers
- Nazir Ismail Philip Supply
- Jennifer Gardy

Oxford High Throughput Sequencing Hub team

People participating in the studies

Informatics

- David Wyllie
- John Finney
- Milind Achyria
- Laura Madrid
- Infections in Oxfordshire Research Database Team

Bioinformatics and Population Biology

- Danny Wilson
- Carlos del Ojo Elias
- Saheer Gharbia
- Tanya Golubchik
- Anna Sheppard
- Dilrini de Silva
- Xavier Didelot
- Jess Hedge
- Vasiliki Kostiou
- Tonya Votintseva
- Luke Anson
- Teresa Street

Oxford Biomedical Research Centre | MRC Medical Research Council | Newton Fund | Bill & Melinda Gates foundation | NERC Science of the Environment | Public Health England | wellcome | Department of Health | Animal & Plant Health Agency | NHS National Institute for Health Research