Inferring the antibiogram from genomes

Derrick Crook

Public Health England

Oxford University Hospitals FT Trust

University of Oxford

UNIVERSITY OF OXFORD

Public Health England

Content

- Considerations for developing resistance prediction as a comprehensive replacement tool for DST

- Three examples:
  - *Staphylococcus aureus*
  - *Escherichia coli (Enterobacteriaceae)*
  - *Mycobacterium tuberculosis*

- Some initial pilot observations of using resistance prediction in a clinical setting

UNIVERSITY OF OXFORD

Public Health England

Basic components for a WGS resistance prediction solution

- A comprehensive knowledge base of variation conferring antibiotic resistance
  - Contain acquired genes encoding functions conferring resistance
  - Includes variation in genes through nucleotide substitution, insertion and deletion

- processing software

- Technical and clinical validation
  - Accreditation e.g. Very Major Errors and Major Errors of <1.5% and <3% /drug

S. aureus

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

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 phentypic 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

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

S. aureus – preliminary results WGS from blood cultures

	Phenotype S			Phenotype R			Phenotype I		
	WGS S	WGS R	WGS I	WGS S	WGS R	WGS I	WGS S	WGS R	WGS I
Ciprofloxacin (n=131)	126	0	1	1	3	0	0	0	0
Clindamycin (n=132)	105	3	0	1	22	0	1	0	0
Erythromycin (n=132)	104	4	0	2	22	0	0	0	0
Flucloxacillin (n=77)	75	2	0	0	0	0	0	0	0
Fusidic Acid (n=55)	49	1	0	0	5	0	0	0	0
Gentamicin (n=132)	116	13	0	2	1	0	0	0	0
Mupirocin (n=132)	129	3	0	0	0	0	0	0	0
Penicillin (n=55)	10	1	0	2	42	0	0	0	0
Rifampicin (n=132)	132	0	0	0	0	0	0	0	0
Tetracycline (n=132)	123	3	0	0	6	0	0	0	0
Vancomycin (n=132)	130	0	0	0	0	0	2	0	0

30/1242 (2%) major errors
(10% in gentamicin)

8/1242 (0.6%) very major errors

1204/1242 (97%) concordance

UNIVERSITY OF OXFORD

Public Health England

Escherichia 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.

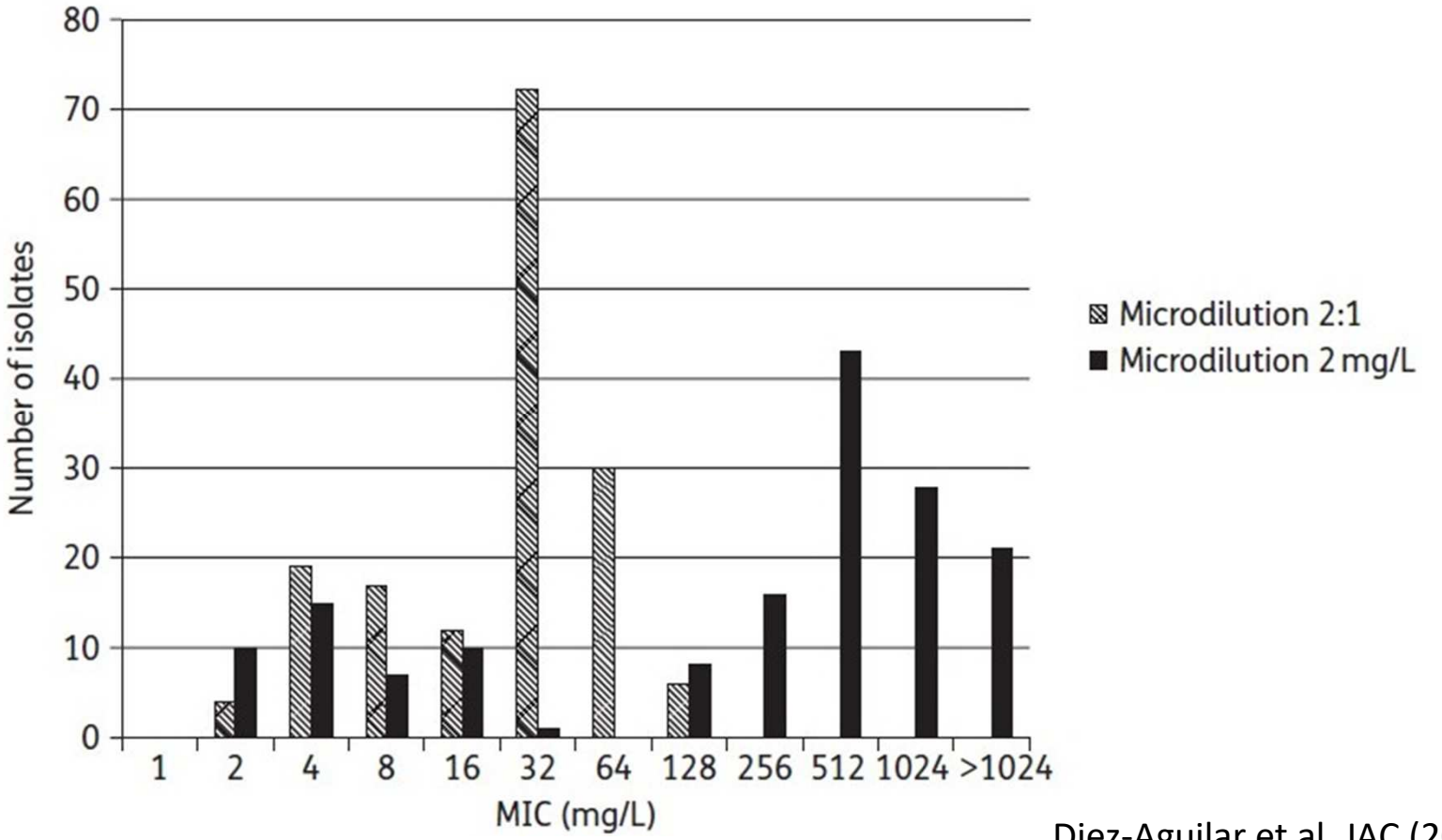
	Susceptible by comparison standard phenotype		Resistant by comparison standard phenotype			
Antibiotic	susceptible by genotype (row %)	resistant by genotype (row %; major error)	susceptible by genotype (row %; very major error)	resistant by genotype (row %)	Sensitivity (95% CI)	Specificity (95% CI)
Amoxicillin	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

Disparity in Coamoxiclav phenotype



Diez-Aguilar et al. JAC (2015)

UNIVERSITY OF OXFORD

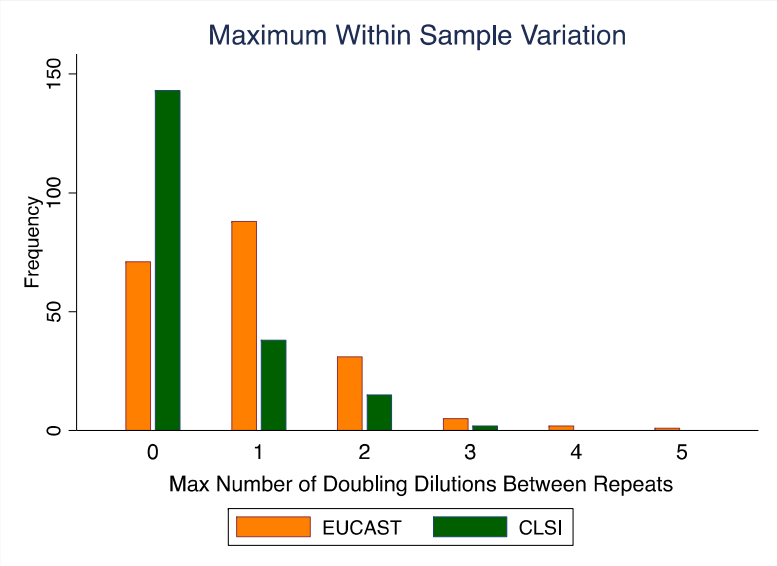
Public Health England

Samples

- Stratified random subsample 290 – Currently, 230 have complete tests.
- Agar Dilution (AD)
  - Triplicate
    - Amoxicillin
    - CLSI co-amoxiclav
    - EUCAST co-amoxiclav
- Excluded if unable to be found/mixed on culture/assembly
- Co-amoxiclav phenotype obtained for 198 isolates

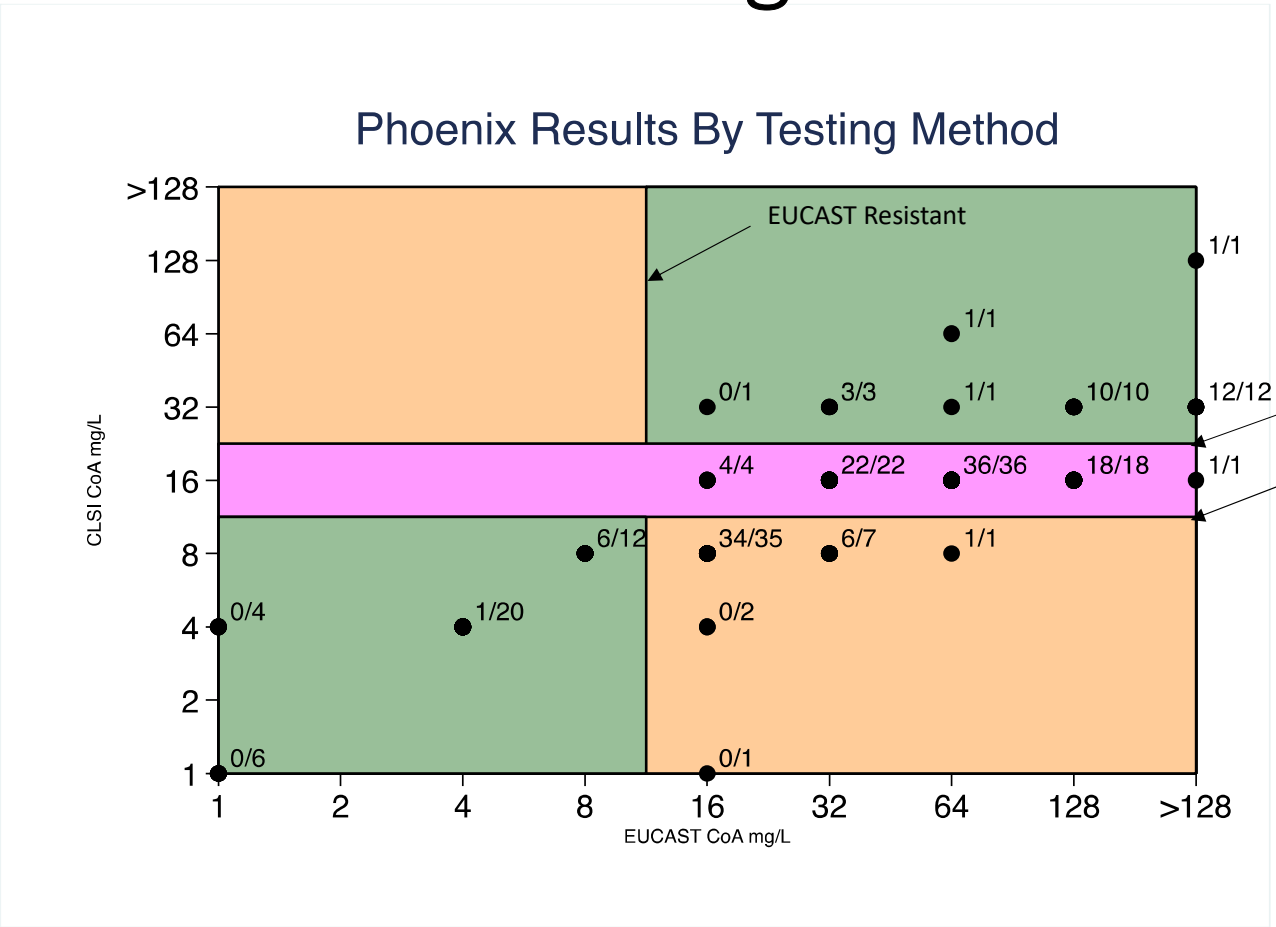
Co-amoxiclav repeatability

- 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

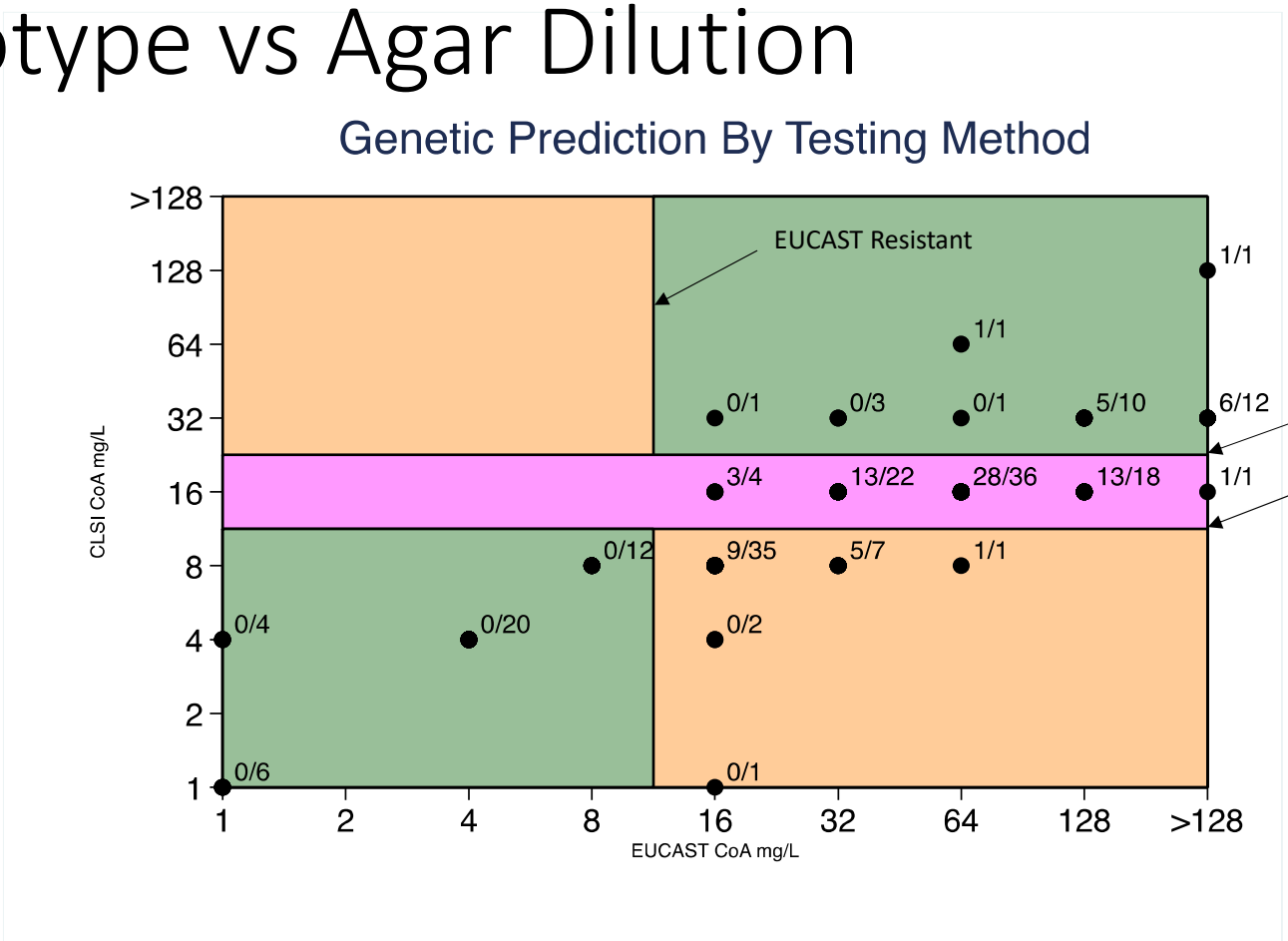
Phoenix vs Agar Dilution



UNIVERSITY OF OXFORD

Public Health England

Genotype vs Agar Dilution



UNIVERSITY OF OXFORD

Public Health England

Accuracy of WGS predictions

- Resistance

Method	No of Resistant Samples	Resistance Correctly Predicted (%) (95% CI)
BD Phoenix (Whole Cohort)	275	45.8% (39.8 – 51.9)
BD Phoenix (Sample)	157	54.8% (46.6 – 62.7)
CLSI Non-Susceptible	110	64.5% (54.9 – 73.4)
CLSI Resistant	29	44.8% (26.4 – 64.3)
EUCAST Resistant	156	55.1% (47.0 – 63.1)

- Clavulanate alone – MIC on all samples 16-32mg/L except one at 64mg/L

- Above calls include mechanisms under dispute
  - TEM “copy number” , TEM presence + omp truncation (32)
  - If inexact matching (+12)

UNIVERSITY OF OXFORD

Public Health England

E. coli/Klebsiella – preliminary results

	Phenotype S		Phenotype R		Phenotype I	
	WGS S	WGS R	WGS S	WGS R	WGS S	WGS R
Amikacin	47	2	0	1	0	0
Amoxicillin/ampicillin	16	0	2	32	0	0
Aztreonam	43	1	0	6	0	0
Ceftazidime	42	1	0	6	1	0
Ceftriaxone	43	1	0	6	0	0
Ciprofloxacin	40	0	0	9	1	0
Co-amoxiclav	30	0	14	6	0	0
Co-trimoxazole	34	1	1	13	1	0
Carbapenems	50	0	0	0	0	0
Gentamicin	43	0	1	5	1	0
Pip-taz	44	1	4	0	1	0
Trimethoprim	33	0	2	14	1	0

7/600 (1%) major errors

24/600 (4%) very major errors (58% in co-amoxiclav)

563/600 (94%) concordance

UNIVERSITY OF OXFORD

Public Health England

TB Resistance prediction

UNIVERSITY OF OXFORD

Public Health England

What is the molecular basis?

- A nucleotide substitution:
  - Best recovered using mapping
    - In a coding sequence
    - In an intergenic region (upstream of a coding sequence)
- An insertion
- A deletion
  - Best identified using *de novo* methods e.g. cortex
- (Genes may be ‘inaccessible’ using short read sequencing)

Can all/most genomic variation conferring anti-tuberculosis drug resistance be discovered?

- Many resistance determinants have been discovered, are there many more?
- How best to discover more?
- Can we dispense with routine phenotyping
  - Need a complete knowledge base
  - Requires a software to process sequences

UNIVERSITY OF OXFORD

Public Health England

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‡*

Lancet Infect Dis 2015;
15: 1193–1202

Published Online
June 24, 2015
http://dx.doi.org/10.1016/
S1473-3099(15)00062-6

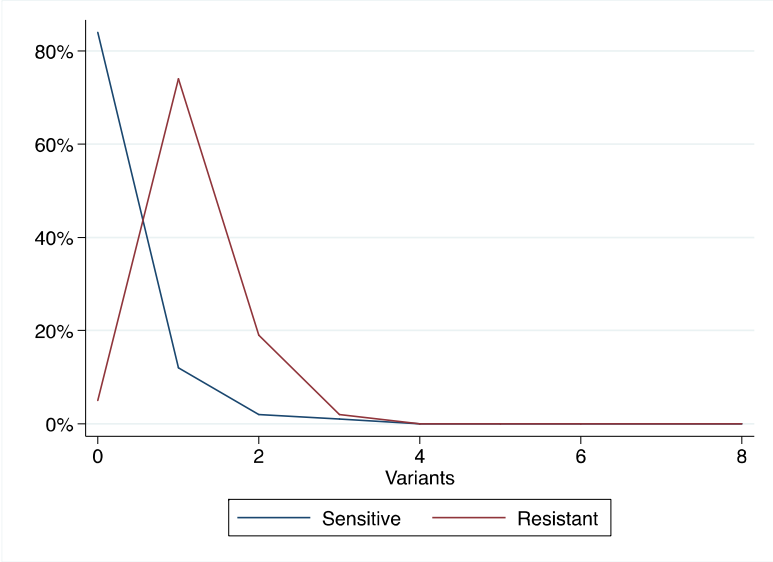
UNIVERSITY OF OXFORD

Public Health England

Can we discover explanatory variation?

- 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
- Investigation of 3651 isolates :
  - Using a heuristic method of predicting resistance
- divided into
  - a 2099 derivation set
  - a 1552 validation set

Distribution of number of variants across candidate-genes and their promoter-regions in phenotypically susceptible and resistant isolates



Number of variants	Sensitive phenotype	%	Resistant phenotype	%
0	7,566	84	33	5
1	1,111	12	518	74
2	211	2	130	19
3	61	1	16	2
4	21	0	3	0
5	6	0	0	0
6	1	0	1	0
8	1	0	0	0

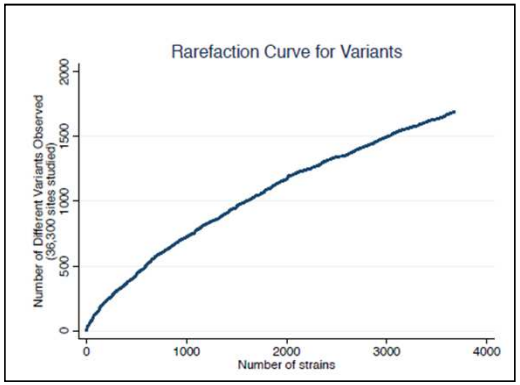
Resistance prediction in a validation set

	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.

Filling the resistance gap

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



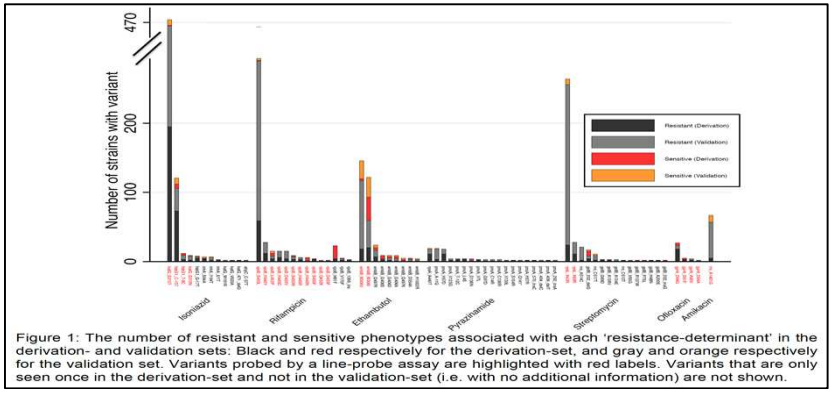
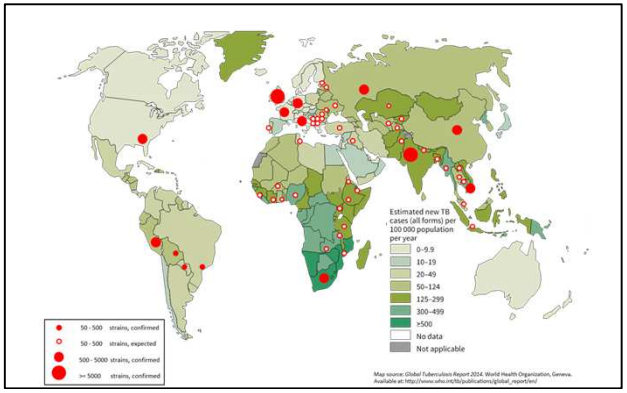


Figure 1: The number of resistant and sensitive phenotypes associated with each 'resistance-determinant' in the derivation- and validation sets: Black and red respectively for the derivation-set, and gray and orange respectively for the validation set. Variants probed by a line-probe assay are highlighted 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.



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

- 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

Public Health England

Head-to-head evaluation of WGS for routine implementation

- 2171 samples processed:
  - 778 TB, 31 failed 747 available for analysis
  - 1323 NTMs
- Species identification WGS vs Hain
  - Concordant with 97% (96%-98%)
- Resistance prediction WGS vs Hain
  - % Fails: 10.4% (any antibiotic)
  - Sensitivity: 98.8% (93-100)
  - Specificity: 99.8% (100-100)
  - Total: 99.8% (99-100)
- MIRU VNTR vs WGS
  - Analysis being completed – MIRU - high level of falsely predict transmission in ~ 50% of cases

UNIVERSITY OF OXFORD

Public Health England

Acknowledgements

- Sarah Walker
- Rosalind Harding
- Tim Peto
- Mark Wilcox – Leeds
- Grace Smith – Birmingham
- Philip Monk - Leicester
- John Paul – Brighton
- Martin Llewellyn – Brighton
- Research Fellows (6)
- Julian Parkhill
- 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
- Adam Geiss
- 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