Public Health England

Whole genome sequencing for routine identification, drug resistance detection and epidemiology of *Salmonella*: A revolution in public health microbiology

Satheesh Nair
Salmonella Reference Service, GBRU,PHE
Colindale

April 2017

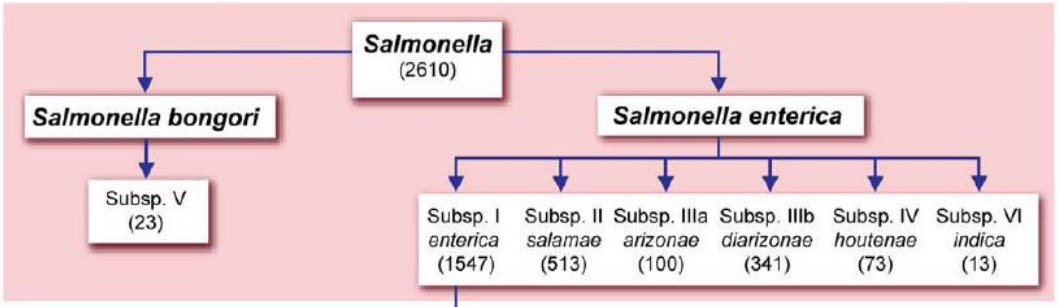
Public Health England

Salmonella classification is complicated

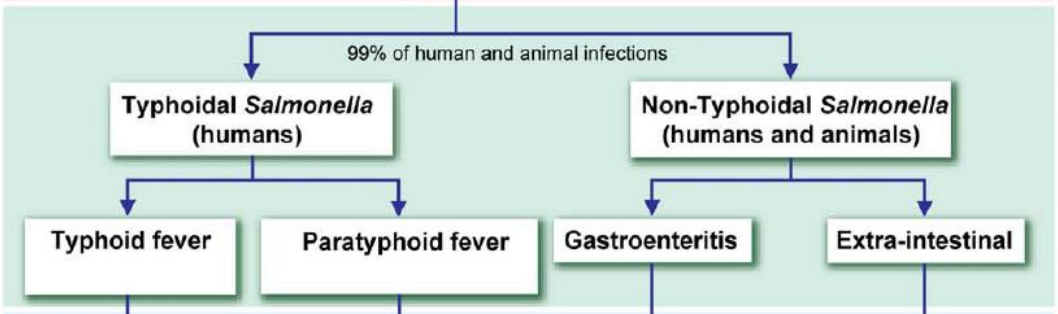
Genus

Species

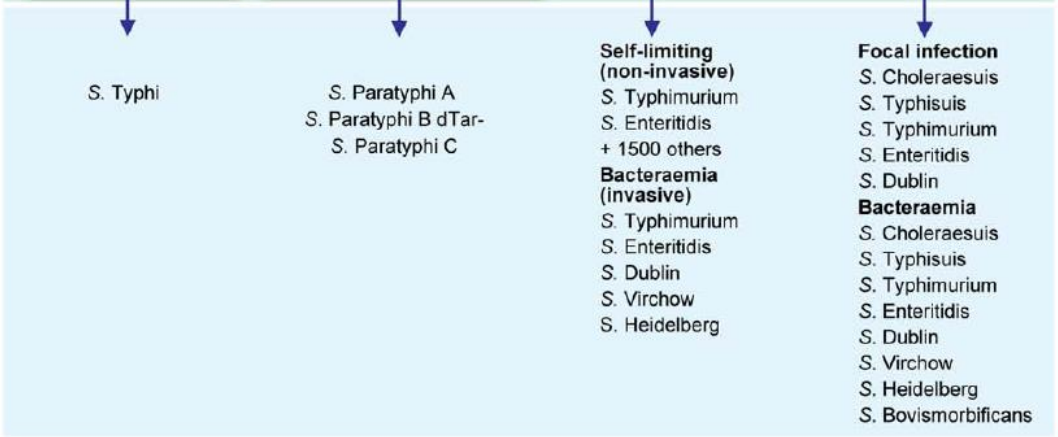
Subspecies



Species and subspecies were originally defined by DNA-DNA hybridisation, confirmed by MLEE and MLST and are currently differentiated by biochemistry and serology.



The split in typhoidal and non-typhoidal is based on the disease syndrome. Typhoid and paratyphoid fever is prolonged, whilst extra-intestinal infection is usually acute and metastatic. Gastroenteritis is characterised by diarrhoea.



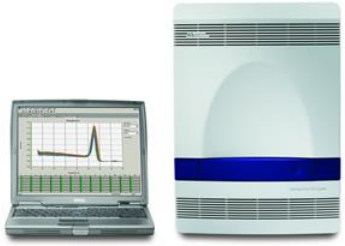
Differentiation of serovars is by agglutination with specific antisera against LPS (O), two phases of flagella (H1 and H2). There are 46 O, 85 H and 1 capsule (Vi) antigen which have been described in about 1,500 combinations within subspecies I.

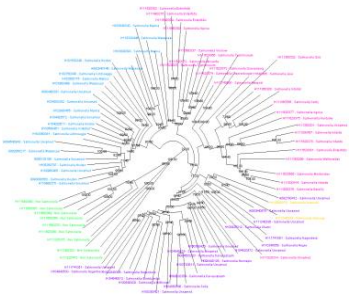
General overview of the current classification of *Salmonella* - Achtman, Nair *et al.,* 2012, PLoS Pathogens

Identification Methods for *Salmonella*

Public Health England

Subspeciation





Real time TaqMan® PCR assays
- target three different genes

Target			*Salmonella*
hil A	*tt* R	*lac* Z	
+	+	-	Subsp I
+	+	+	subsp III
-	+	-	*Salmonella* : sent to BIOLOG
-	-	-	non-*Salmonella*

OmniLog® ID System (Biolog)
- phenotypic microarray

Detection of Subspecies II and IV

Serotyping

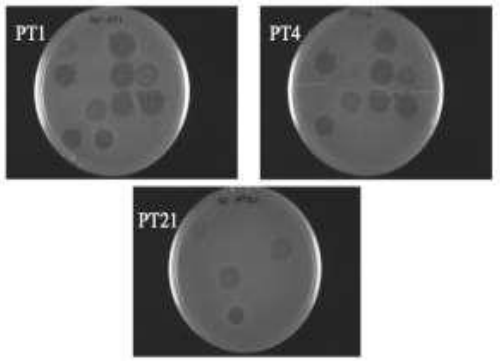
Agglutination with specific antisera against LPS & flagella (O & H antigens)

- **Slide agglutination**
- **Microtitre plates**
- **Dreyer's tubes**

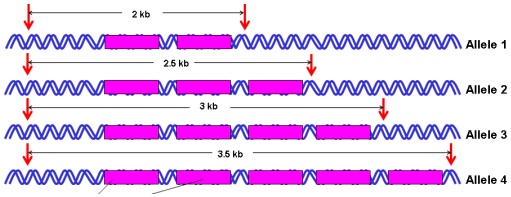
Serotype	O Antigen	H_1	H_2
S. Enteritidis	9,12	g,m	-
S. Typhimurium	1,4,[5],12	i	1,2
S. Newport	6,8	e,h	1,2
S. Kentucky	8,20	i	z6

Public Health England

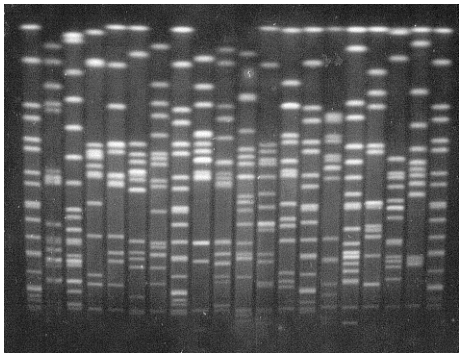
Sub-typing Methods for *Salmonella*



- **Phage typing**
  - **e.g. Typhimurium DT1, DT193**



- **Multi-locus Variable Number Tandem Repeat Analysis (MLVA)**
  - **e.g. 4-13-13-10-0211**



- **Pulsed-field gel electrophoresis (PFGE) - e.g. SNWPXB.0010**

These 3 methods are absolute now at GBRU

Public Health England

Exploring the use of MLST for *Salmonella* Identification

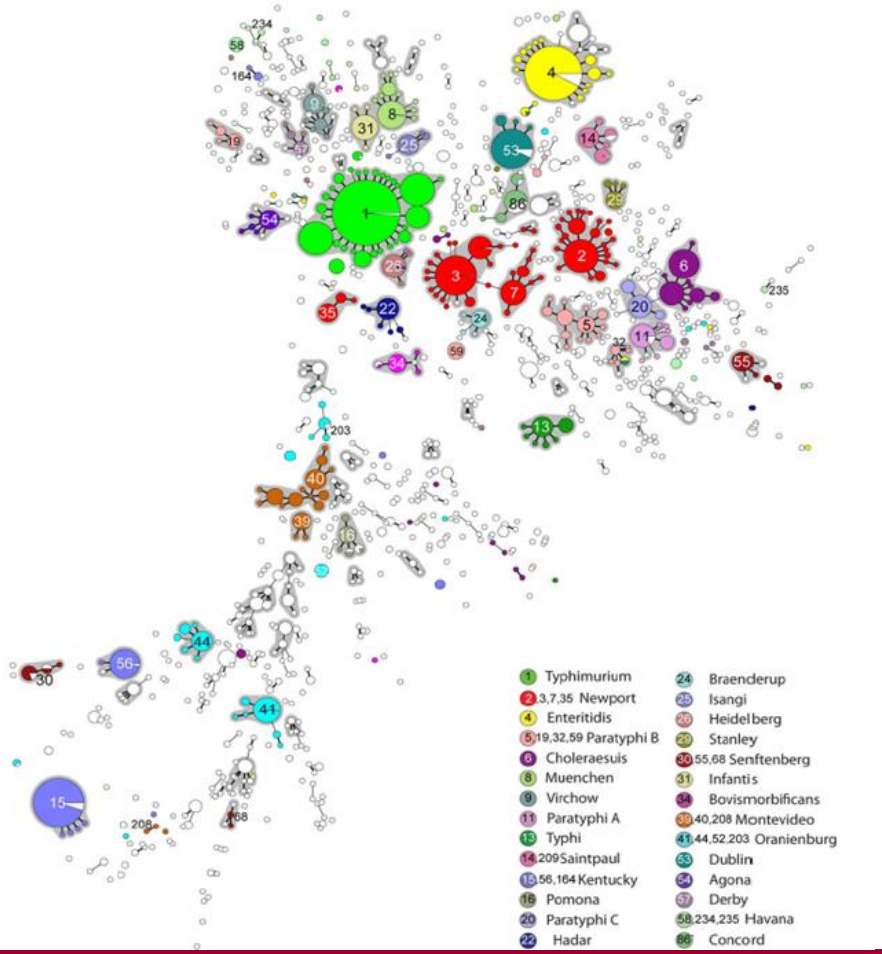
MLST based on sequences of 7 house keeping genes

MLST , effective in the identification of natural genetic clusters [**Sequence Types(ST)** and **e-Burst Groups(eBG)**]

In general : clusters defined by MLST correspond on a one to one basis with serovars
(Achtman, Nair et al: 2012)

e.g., ST19 – Typhimurium, ST 1 - Typhi

Minimal spanning tree of MLST data for *S. enterica*



Whole Genome Sequencing Era at PHE since 2012

Public Health England

PHE investment in WGS: financial, laboratory, bioinformatics, data handling, staff training

4 MiSeq machines

+

2 HiSeq 2500 high-throughput machines

= Capacity ~ 3,000 genomes per week

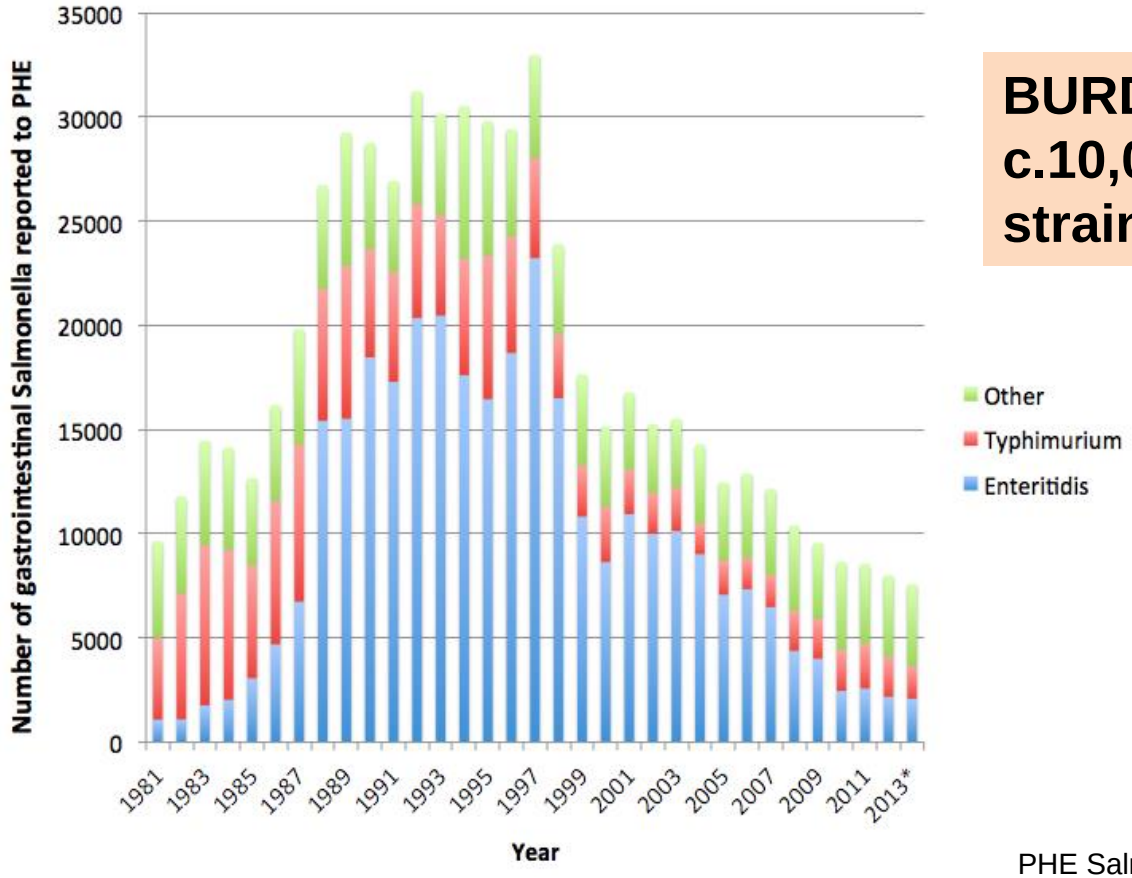
Infrastructure

Data storage warehouse
Generators & Coolers

Public Health England

Implementaion of WGS

- Priority organisms selected in 2013 – ***Salmonella phase 1 validation***



BURDEN :
c.10,000 *Salmonella* strains every year

PHE Salmonella surveillance data - http://bit.ly/1CdEOe0

Schematic representation of WGS methodology for DNA extraction in SRS

Samples arrive

Lysis & enzymatic steps

Run the QiaSymphony DNA extraction

Glomax - Promega

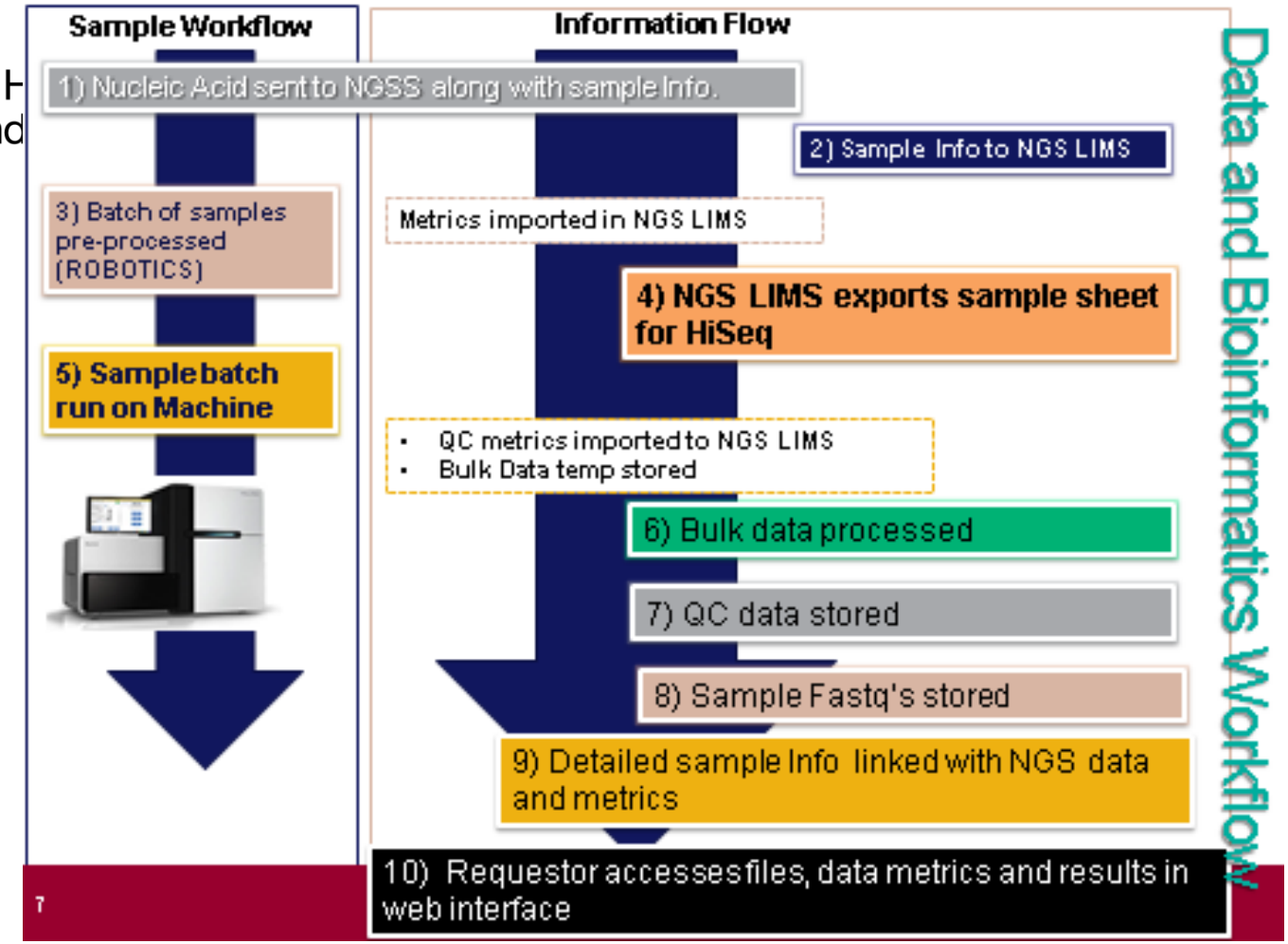
Quantify DNA extracts

LabChip – Perkin Elmer

Transfer DNA to Sequencing Service

GSU Bioinformatics

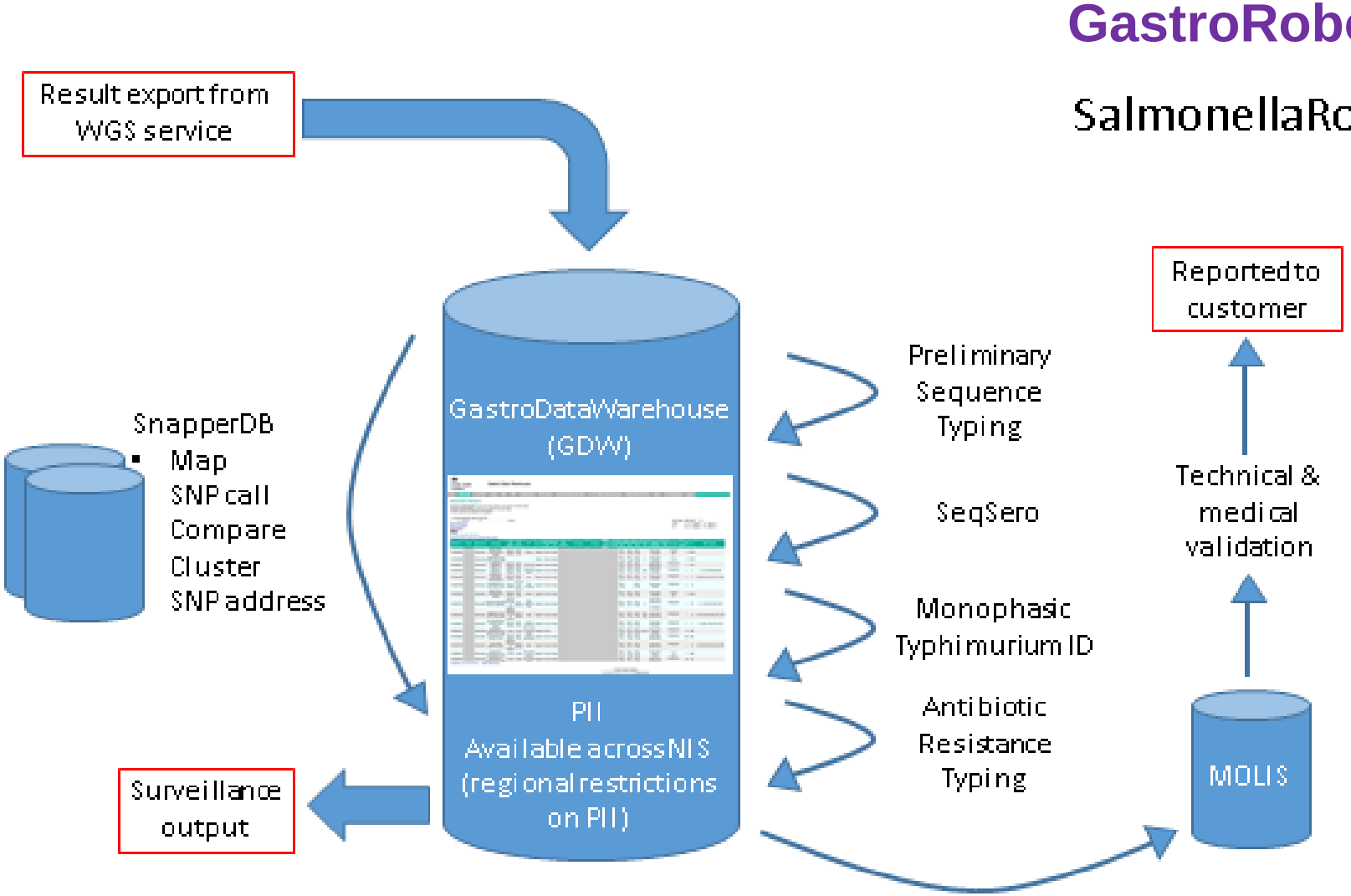
Public H
England



SRS Bioinformatician – GatroRobot (Salmonella Robot)

Jon Green

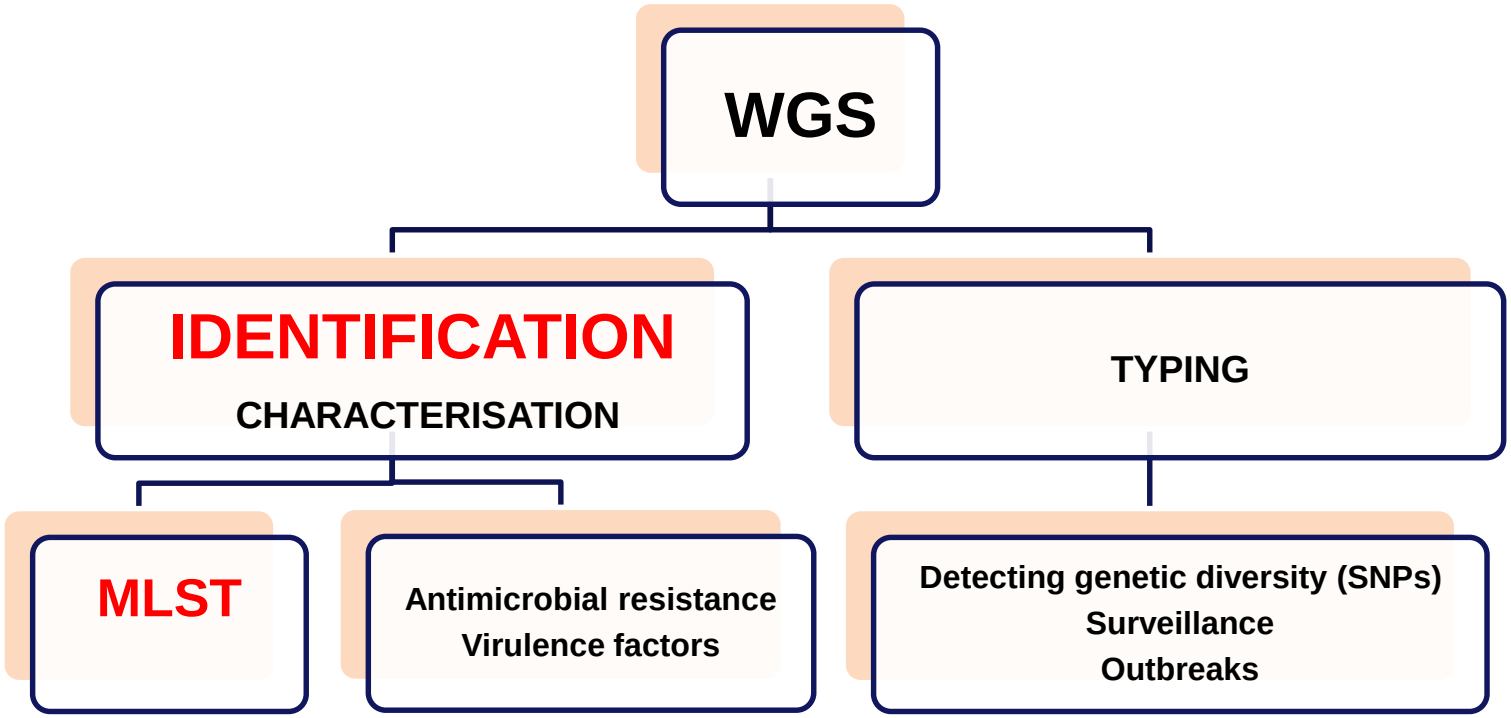
WGS has transformed routine microbiology



Phil Ashton/Tim Dallman

Public Health England

As of 1st April 2015 WGS has been adopted for routine use in SRS



- WGS to replace lengthy laboratory methods (serology,PFGE) and improve safety, quality
- WGS can provide identification and typing in a single method

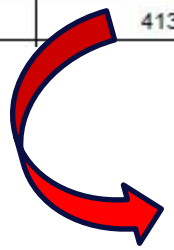
Public Health England

Automated report for *Salmonella* ID

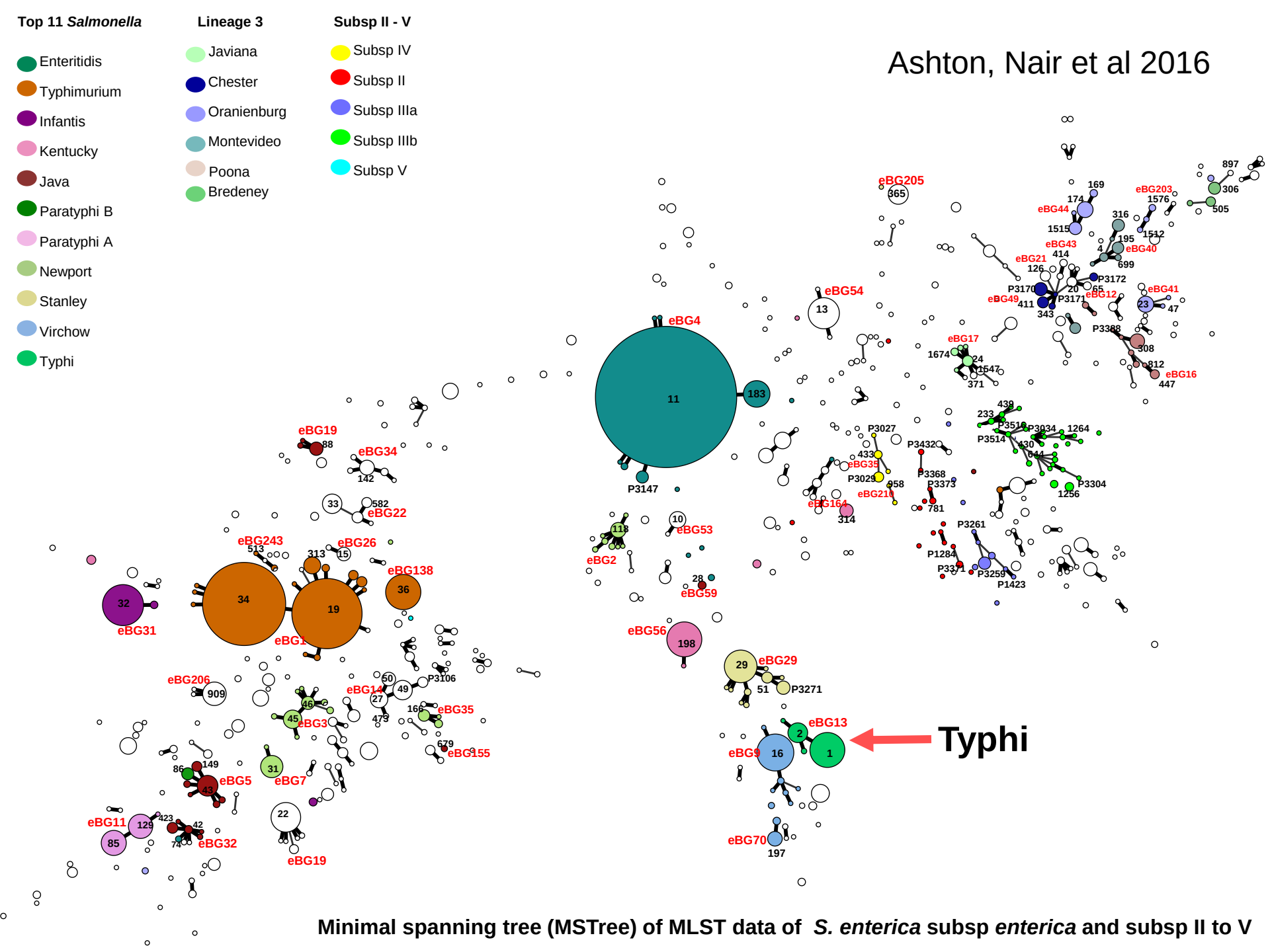
Results for SRS : salmonella-typing

Submitter: SRS
Workflow: salmonella-typing
Date-Time of report: 13/09/2014-22:19.49

Sample	MLST ST	MLST profile	QC mean cons depth	QC max % non cons base	QC % coverage	QC min cons depth	predicted serotype
H14354083501-1	NOVEL_allele	191,22,*27,22,18,85,169	28.65	7.1	100	12	no ST-serotype
H14354083601-1	1541	197,187,10,234,8,65,22	37.86	5.3	100	15	Corvallis 1
H14354083701-1	592	189,70,68,132,175,9,172	26.79	9.4	100	12	Worthington 4
H14354083801-1	413	15,70,93,78,113,6,68	31.94	6.3	100	15	Mbandaka 15

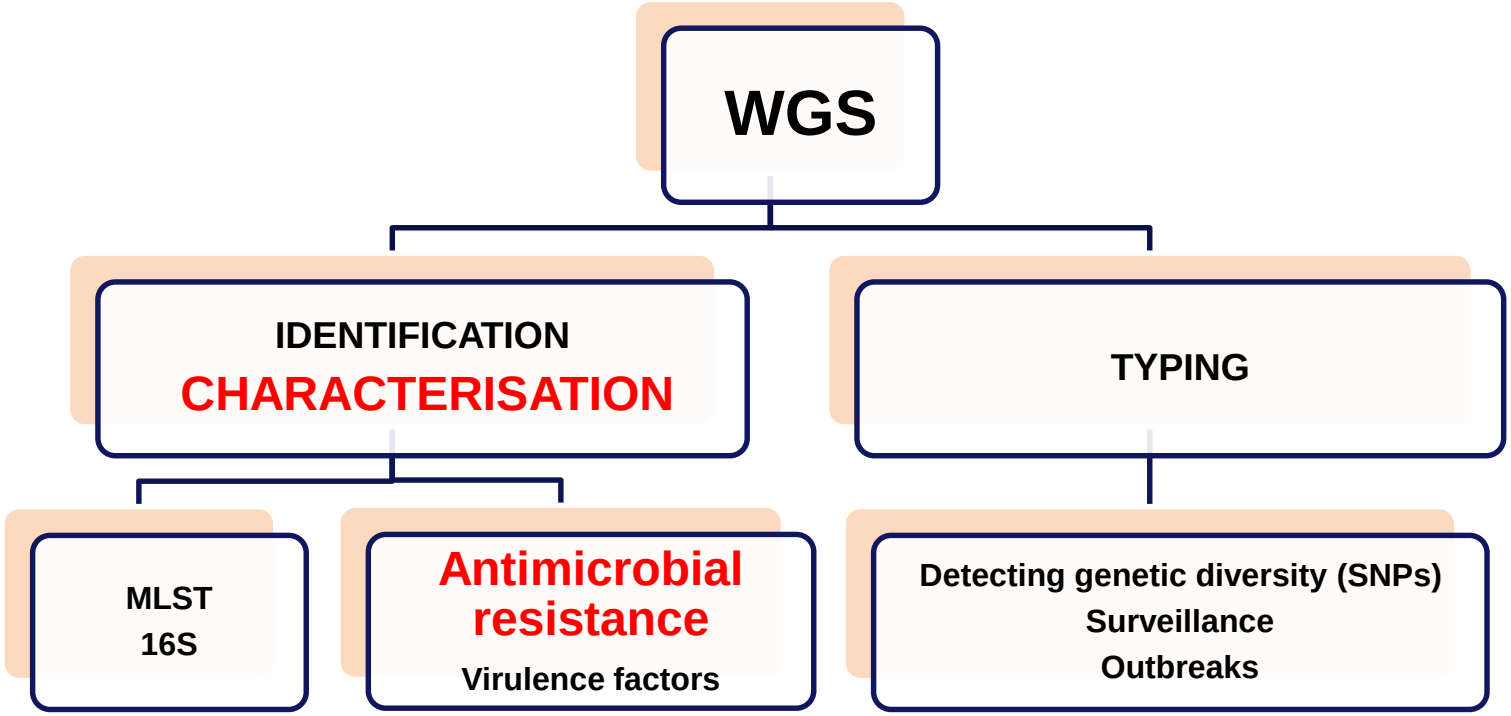


- ST 413 = *Salmonella* Mbandaka


Top 11 Salmonella
Enteritidis
Typhimurium
Infantis
Kentucky
Java
Paratyphi B
Paratyphi A
Newport
Stanley
Virchow
Typhi
Lineage 3
Javiana
Chester
Oranienburg
Montevideo
Poona
Bredeney
Subsp II - V
Subsp IV
Subsp II
Subsp IIIa
Subsp IIIb
Subsp V
Ashton, Nair et al 2016
Typhi
Minimal spanning tree (MSTree) of MLST data of S. enterica subsp enterica and subsp II to V

Public Health England

Salmonella WGS Project



GeneFinder (antimicrobial resistance gene finder)

Developed by Michel Doumith

Database

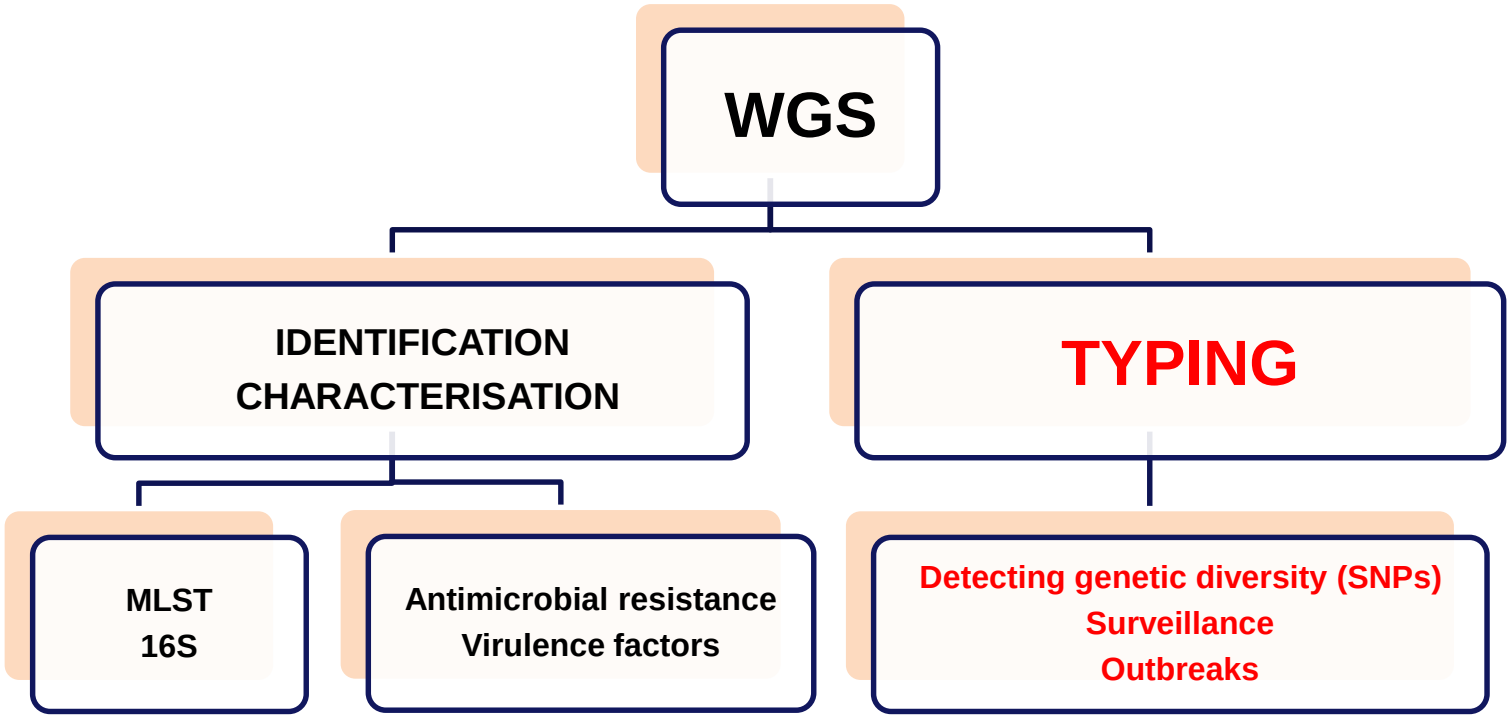
- Acquired : ~ 1600 resistance genes
- Chromosomal : *gyrA*, *parC* and *rpoB*

molis_id	mlst_st	predicted_serotype	rtype_str	rep_t	carb	b_lac	amn_c	flu_c	gly_c	mls_c	trm	fos	tet	sul	chl	rif	mup	otr	fus
H1439406	P3075	no ST-serotype	"SUL256,TET8,TMP2,NAL16,CIP0.064"	P	-	-	aac(6')-Iaa[v]	gyrA_SET[87:D-Y];parC_SET[57:T-S]	-	mph-(A)	dfrA-5	-	tet(A)-1[v]	sul-1[v]	-	-	-	-	-
H1418203	52	('Blockley', 4)	"CHL8,CHL16,STR16,TET8,NAL16,CIP0.064"	-	-	-	aph(6)-Id,strB;aac(6')-Iy[v]	gyrA_SET[87:D-G];parC_SET[57:T-S]	-	mph-(A)	-	-	tet(A)-1[v]	-	catA-2[v]	-	-	-	-
H1438405	516	(Give, 11), (, 1)	"AMP8,TET8,NAL16,CIP0.064,CIP0.5"	-	-	TEM-215	aac(6')-Iy[v]	gyrA_SET[83:S-Y];parC_SET[57:T-S];qnrS-1[v]	-	mph-(A)	-	-	tet(A)-1[v]	-	-	-	-	-	-
H1439207	29	(Stanley, 24), (Sarajane, 1)	"AMP8,CHL16,SUL256,STR16,TET8,TMP2,CIP0.064,CAZ1,CAZ2,CTX0.5,CTX1,FOX8"	HI2[v];N[v]	-	CMY-2;TEM-1	aadA-1b[v];aac(6')-Iy[v];aph(6)-Id,strB	parC_SET[57:T-S];qnrS-1	-	mph-(A);	dfrA-12	-	tet(M)[v];tet(A)-1[v]	sul-3;sul-1[v]	floR[v]	-	-	terF;terE;terD[v];terC;terB	-

WGS to predict Genotype resistance *vs* Phenotype resistance – 1% error

Public Health England

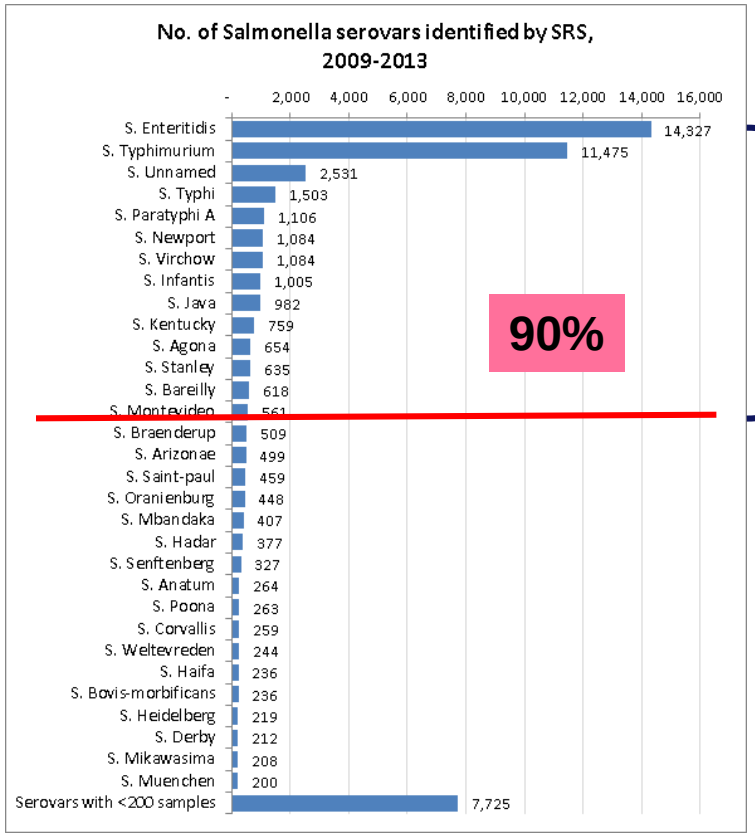
Salmonella WGS Project



Higher resolution SNP typing for surveillance

Public Health England

SnapperDB(SNPdb) of top 14 serovars seen in the UK

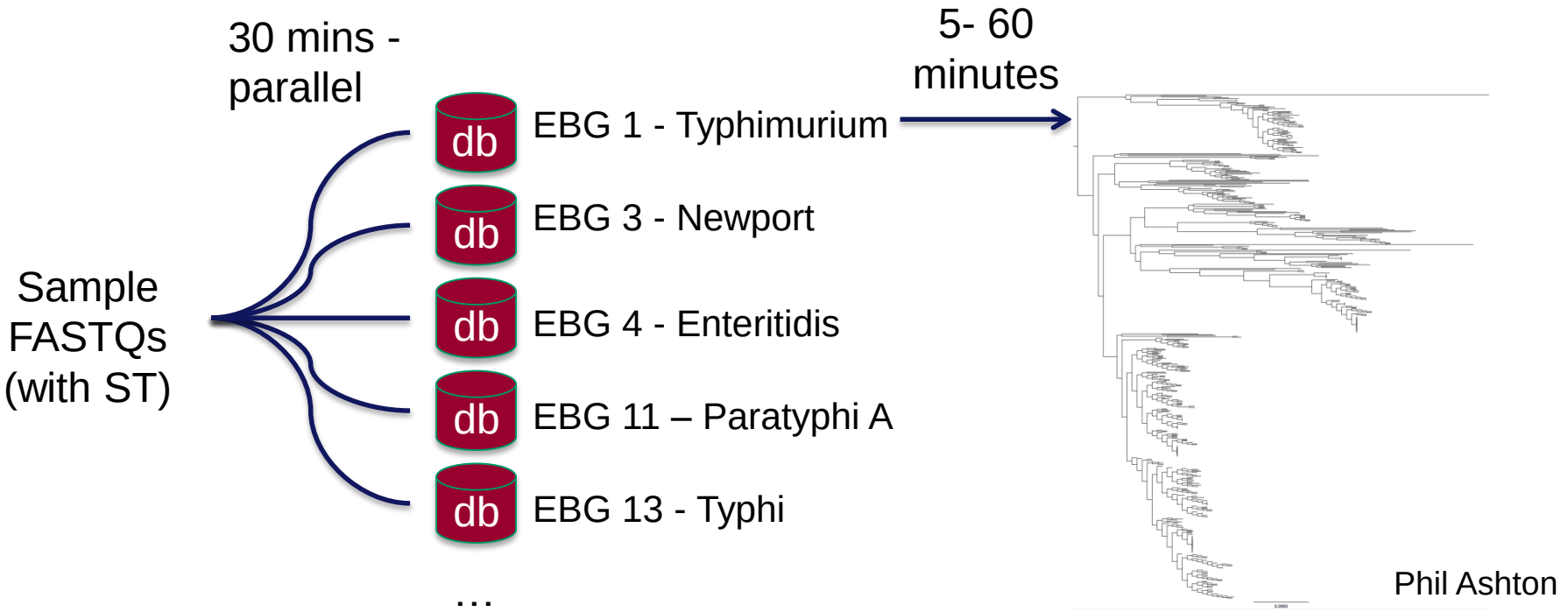


Enteritidis
Typhimurium
Unnamed
Typhi
ParatyphiA
Newport
Virchow
Infantis
Java
Kentucky
Agona
Stanley
Bareilly
Montevideo

David Powell

Public Health England

Rapid hands-off analysis of hundreds of strains a week - *generate phylogenetic trees*



Public Health England

Salmonella Enteritidis 14B

National / International outbreak in 2014

Orange – Austria
Blue - France
Red - Germany
Black - U.K

Tim Dallman

WGS : *Salmonella*

1. For Identification/classification

2. Antimicrobial resistance detection

3. Typing for surveillance – detection of outbreaks (international/national – e.g PT14b), local outbreak (Gold coast)

WGS : ***Salmonella***

Detection of novel regions (e.g Azithromycin drug island in Blockley)

Acquired resistance to macrolides

Several different mechanisms involved :

1. Target site modification – *erm* genes
2. **Modifying enzymes – *ere*A, B and *mph*A,B,D**
3. Efflux pumps – *mef*A and *msr*A (mainly in gram positive)
4. Mutations in the *rrl* and *rpl* genes (gram positive)

In Enterobacteriaceae - presence of ***mphA*** is sufficient to cause high resistance to azithromycin (MIC ≥ 16ug/ml)

Azithromycin is being used now for Enteric fever and invasive NTS in Asia

Macrolide resistance in *S*.Blockley

During the GeneFinder validation :

In 19 sequenced *S*. Blockley between 2012 – 2015

- **9 *mph*A positive and conferred high AZT resistance**
- **10 *mph*A negative sensitive**

Chromosomally or plasmid mediated ?

Discovery of Drug resistant island in *Blockley*

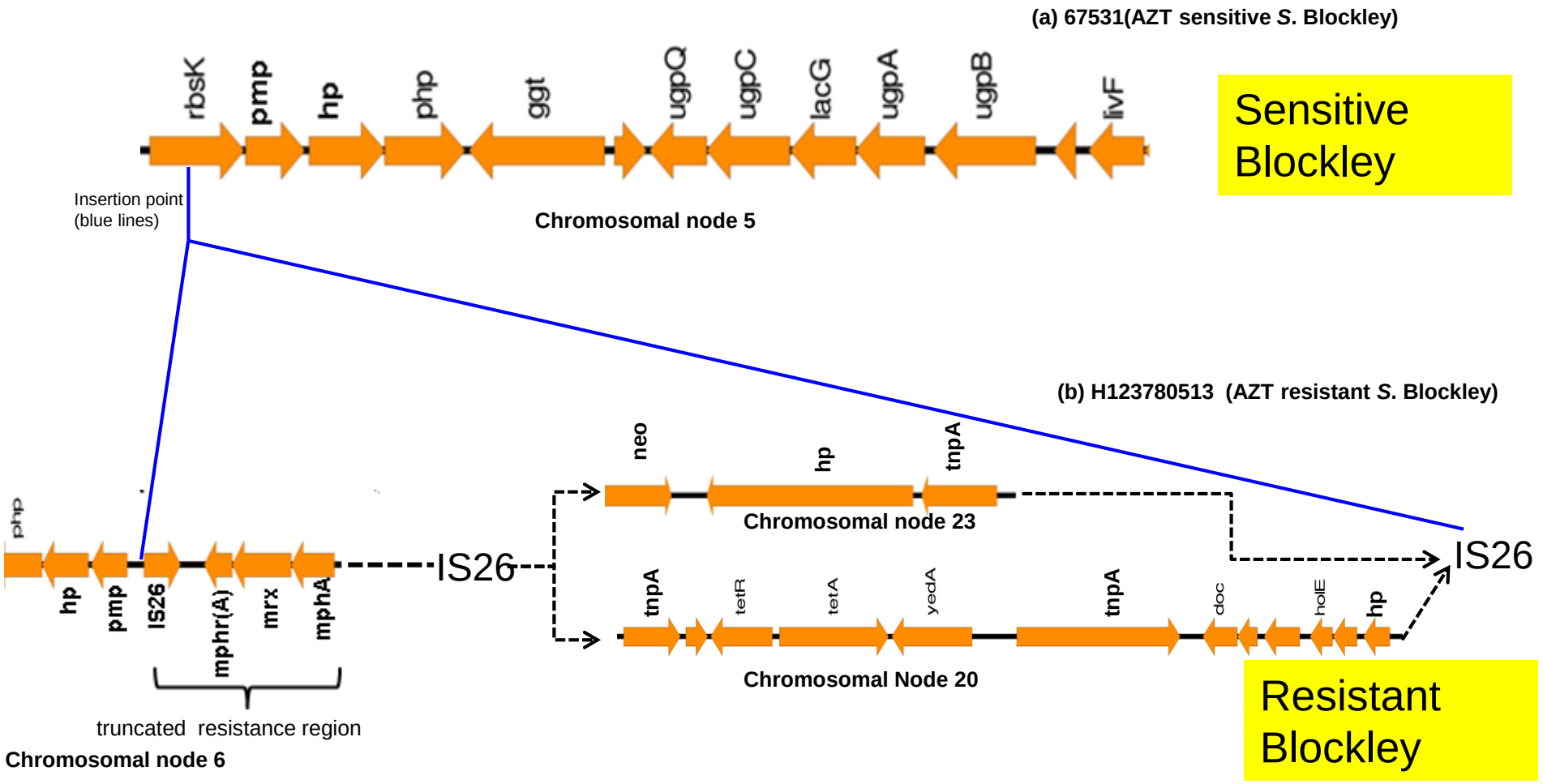
Illumina HiSeq

MinION

Illumina HiSeq

- High throughput
- Cost : £45-55
- Average size footprint
- Read lengths : c.100bp
- Accuracy : c.90%

Short sequence reads – cannot identify repetitive insertion sequences that flank horizontal acquired genes (antibiotic genes and bacterial virulence genes)



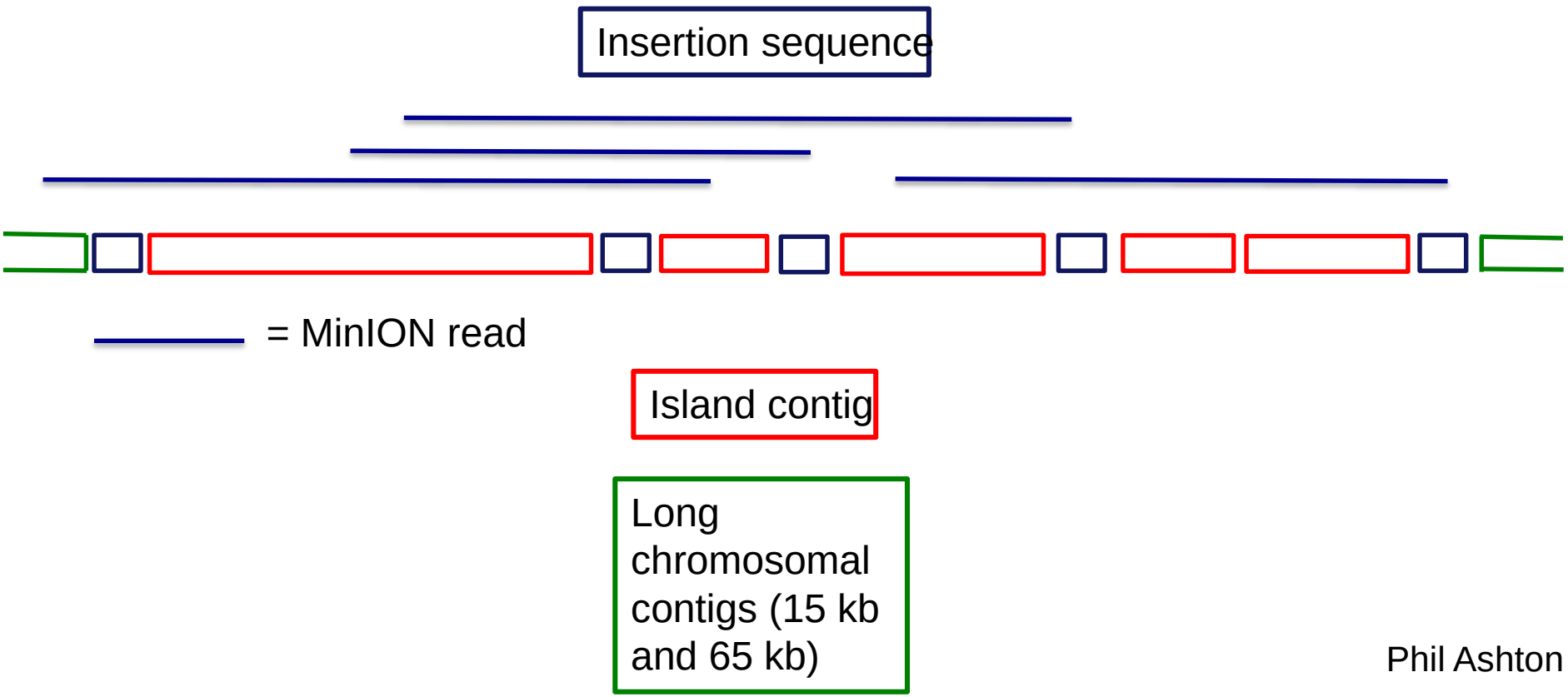
Predicted chromosomal drug island based on Ilumina sequencing

MinION sequencing

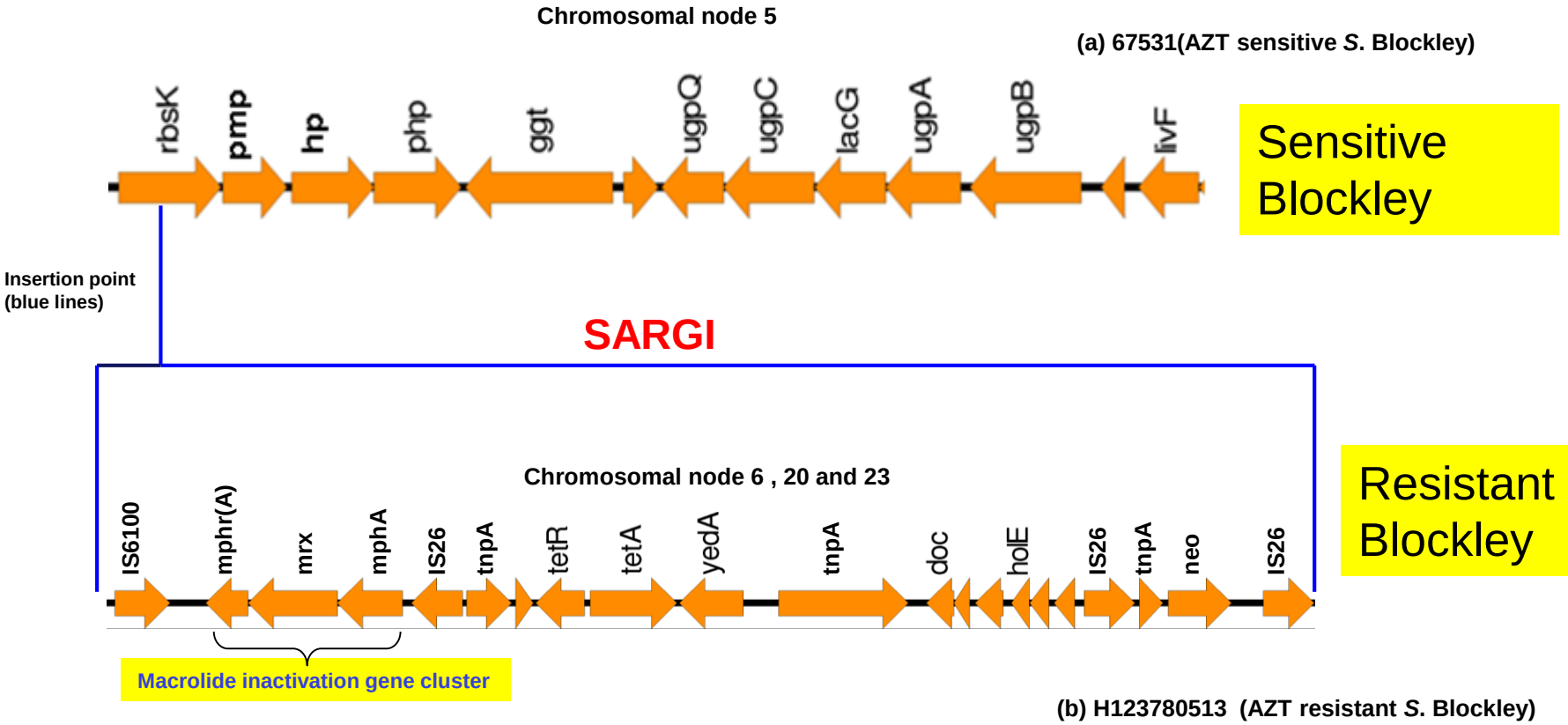
- Cost : Company has not set a price yet
- Small footprint
- Longer reads c.65kb
- Accuracy c.70%

Process for hybrid genome assembly

- To resolve the sequence gaps in Illumina sequencing
- Longer MinION sequence reads as a scaffold for the shorter Illumina sequence reads



Schematic of how the MinION reads allowed the scaffolding of the Illumina contigs.



SARGI –Salmonella Azithromycin Resistance Genomic Island

1st chromosomally mediated macrolide drug island in *Salmonella*

Nair et al 2016. JAC

Other findings from this current study :

- **Azithromycin found in multiple *Salmonella* serovars in the UK since 2015**
- **Multiple drug regions involved – both plasmid and chromosomal mediated**
- **Care in using Azithromycin for treatment**
- **Manuscript in process**

WGS : *Salmonella*

1. For Identification/classification

2. Antimicrobial resistance detection

3. Typing for surveillance – detection of outbreaks (international/national – e.g PT14b), local outbreak (Gold coast)

WGS : ***Salmonella***

Detection of novel regions (e.g Azithromycin drug island in Blockley)

WGS : *Salmonella*

Surveillance of emerging virulent pathogens - Typhimurium ST313

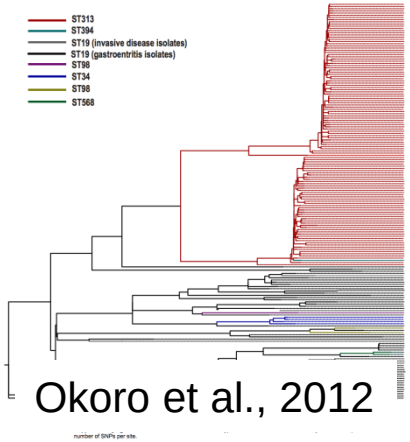
Public Health England

Salmonella Typhimurium ST313

Do we see ST313 in the UK?

Public Health England

Salmonella Typhimurium ST313



Okoro et al., 2012

Lineage 2 – three SRS isolates
Ghana – Blood
Nigera – Blood
Unknown – Blood/ bronchial

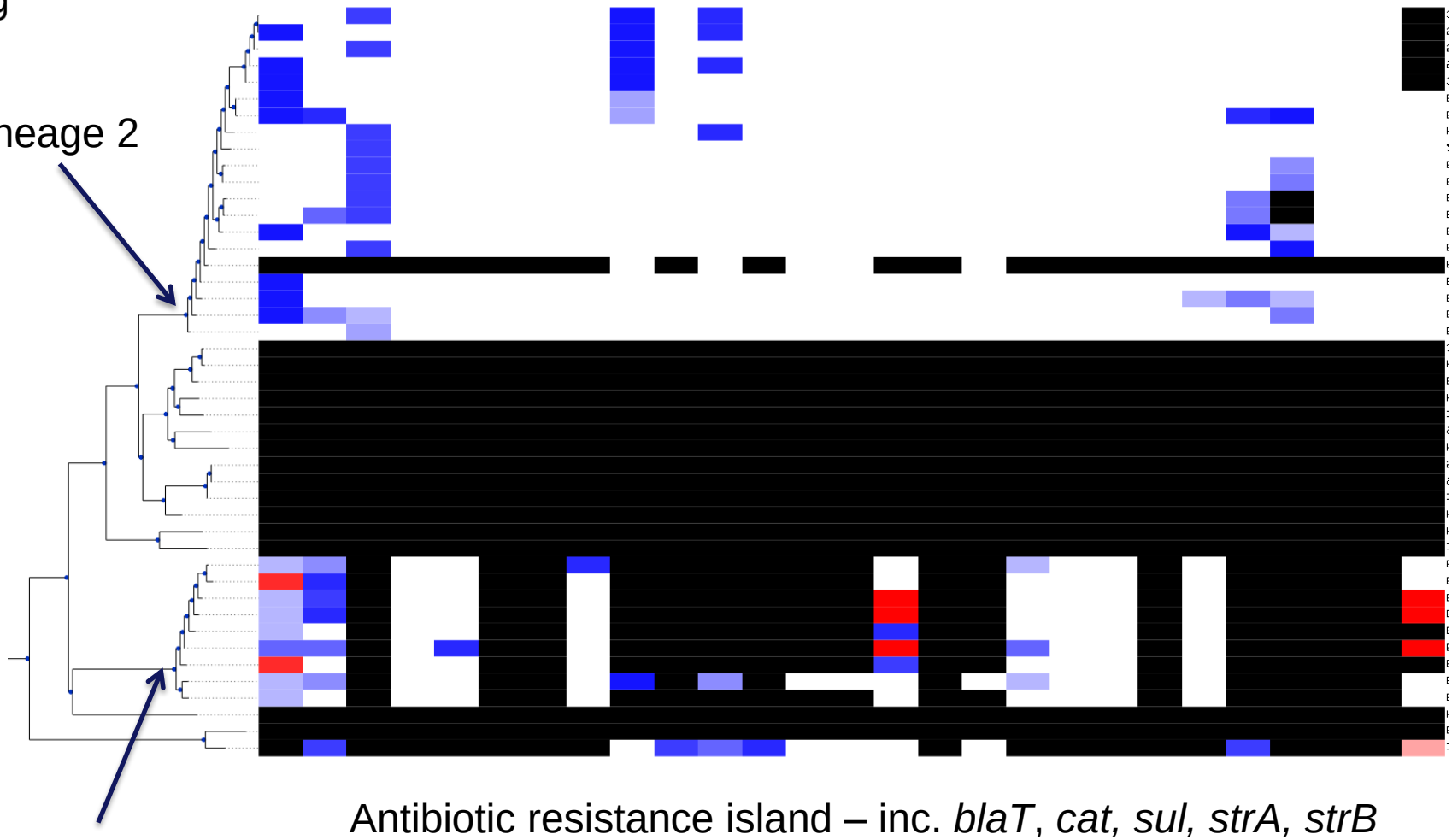
Lineage UK? – 12 SRS isolates (gastroenteritis)
No travel

Lineage 1 – no SRS isolates

Phil Ashton

- **Phylogenetic analysis shows diversity of UK ST313 gastroenteritis isolates from invasive isolates**

Public Health England



Antibiotic resistance island – inc. *blaT*, *cat, sul, strA, strB*
Black = absence
White = presence

Phil Ashton

Public Health England

General conclusions : *Salmonella* WGS

- **Rapid evolution in sequencing technologies and lower prices allows us to :**

  1) Carry out high throughput identification of bacteria

  2) Type bacteria – higher resolution SNP based typing for surveillance

  3) Identify virulence and drug resistant regions that were previously challenging to characterise

Conclusions (2):

WGS allows the development of :

- PCR based assays to differentiate Enterics from NTS
- Real-time PCR assays to detect drug resistance (e.g. *mphA*)
- Routine and diagnostics tools to take out to 1st line labs/field

Public Health England

Sequence data – Public Domain

- Public deposition of data

Uploading data into short read archive
(SHARING DATA)

NCBI BioProject accession: PRJNA248064

Public Health England Pathogen Sequencing

Whole genome sequencing data from Public Health England.

Project Type: Umbrella project

Relevance: Medical

SRA Data Details

Parameter	Value
Data volume, Gbases	1
Data volume, Mbytes	248

This project encompasses the following sub-project:

Project Type	Number of Projects
Genome sequencing *Highest level of assembly :* SRA or Trace	1

BioProject accession	Assembly level	Name	Title
PRJNA248792	SRA or Trace	Public Health England - Gastrointestinal Bacteria Reference Unit pathogens Genome sequencing	Public Health England - Gastrointestinal Bacteria Reference Unit pathogens Genome sequencing (Public Health England)

Submission:

Registration date: 19-May-2014

Public Health England

Global tracking of important foodborne pathogens

Acknowledgements

Public Health England

Salmonella Reference Service

Every member of this lab (past and present)

GSU WGS Group

Cath Arnold and team

Bioinformatics

Tim Dallman

Phil Ashton

Anthony Underwood

Rediat Tewolde

Jonathon Green

GBRU

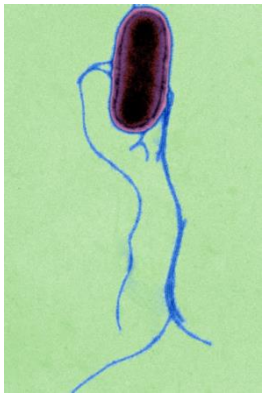
Kathie Grant

Claire Jenkins

WGS Core Group

WGS Implementation Group

David Powell – Data management



Salmonella by Dave Goulding