FUTURE PERSPECTIVES

Role of Bioinformatics in Infectious Diseases

Tolbert Sonda: PhD candidate

Kilimanjaro Clinical Research Institute
Academic Centre for Evidence Based Health Interventions
An Institution of the Good Samaritan Foundation

Outline

- Introductory bioinformatics
- Bioinformatics unit at KCMC/KCRI
- Bioinformatics as tool for resistance determination

Bioinformatics (1)
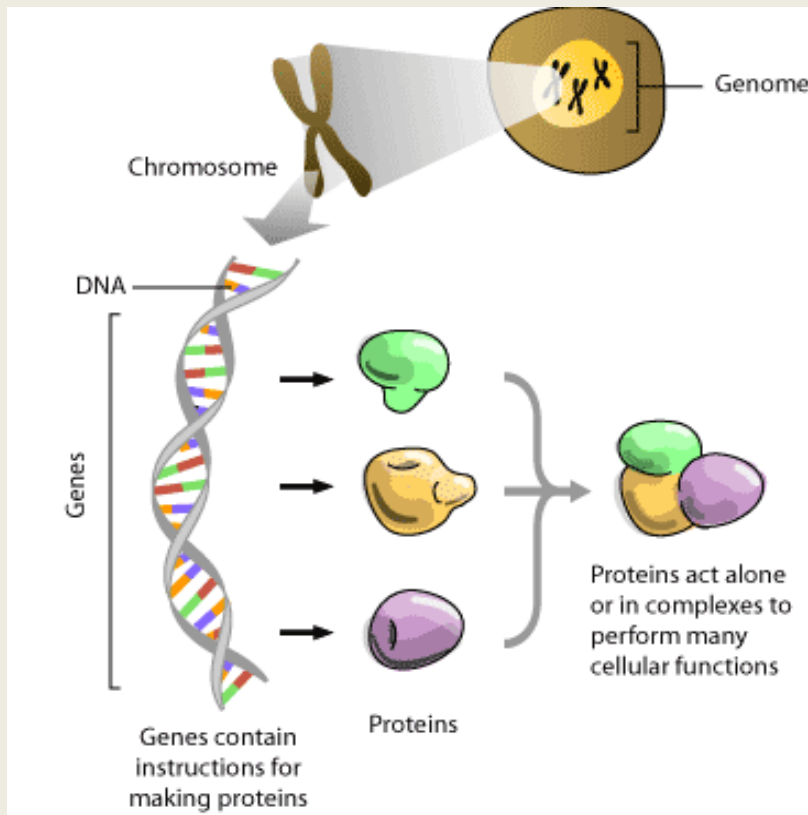
Genomics

A study of genes (DNA) and their functions

Bioinformatics

A computer analysis of biological data (DNA) i.e. decoding information stored in the genetic code
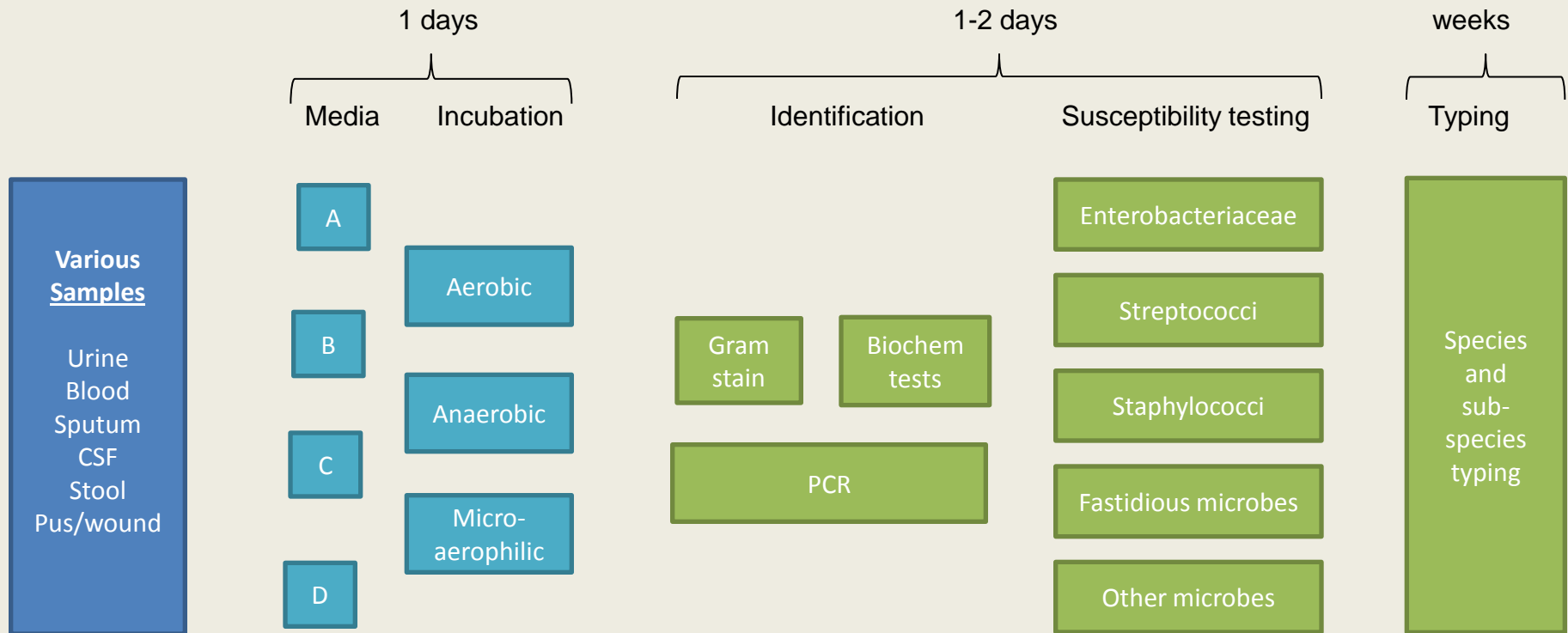
Double helix DNA

Nobel Prize winners, 1962

Wilkins Steinbeck Kendrew Perutz Crick Watson

Bioinformatics (2): example data file

```
>NODE_1_length_4413_cov_26.455925
TTATCGGGGCGAAGTGCCTCCGGTAAAAATACCAGAACAAGAGCCATTACAGGCATTGAC
GGTGATATCCGAATCAACAAAGCATTATGGATGATTGCTGAACAGTTCAGGGAGTGGAAG
TCATGATAATATAAGTAAAGCCCCGAAAAATTTTTCGGGGCTTTACTTATATTATCATCC
TGCGGCATTGCTGTGTCGTCTGATATGCCTCAATTTCATCAACTGGCTTTTCGGTAGTCT
GTAAATATTGTGCCCGCGAAAAATGGTCTGGCAGGAGGCTGCGCAGTTACCGGATATATA
TAAGATGGGGTCCCGCTTTGCGGGGCTTCGGCCAGTCAGCATTATCACTCATATGATATT
TTTGTGTGTGGCCTTCCATGCCGCTGTCGCGGCATAACCGGTATTTGACAATGTATACAA
TATTAATTGTTCTGCACACTGTTAATTTGGGAAAGTGATTTCTTCGTGGTCGGAACCCGG
ATGCCTTTACTGCTCTTTATCCAGGTATAAATCCGGTAAACCTCTGATGCGTAGGCAAGA
CATCTCTGGTTCTGGCGTTCGGTCTTCCTGAATCCGGCATTGTATGCACCAACGGCCTCC
CAGGAGACGCCCCATTTTTTAAAGGCTATTGCCAGATAATAAGCACCGGTATAAATGTTC
ATGCAGGGATCTGTTGTCAGATGTTCCGGCTTAATTCCATAGCGGGCCAGTTCGTTAAAA
TGCTGGGAATCTACCTGCATCAGTCCGCTGCCATATCCCGTTACCGGATTAATACCGATG
GCATTAACCCGGTAACGGGATTCTTTCCATGATATCGCTCTCAGTAAATCCGGGTCTATT
TTGTAATCCCGACCTGCAAGATCAAAGCAATCGGCGGCTTCGCAGATCCCATTTATAAAC
ATCAGGCAGATTGCTAACATCCATTTTTTTCATTTTTCCACCTCTGGTGACTTTATCCGTA
AATAATTTAACCCACTCCACAAAAAAGGCTCAACAGGTTGGTGGTTCTCACCACCAAAAG
CACCACACCCCACGCAAAAACAAGTTTTTGCTGATTTGCTATTTGAATCATTAACTTATA
TTTTAAATAATGTATTTTAATTTATTTTACATTATAAAAAGGATTCATTGGTGAATCGCA
TATGATTCACCAATGAATTTAACGCAACTCATAGATGAATCAGTAATAATAAAACAACGA
GTTAAAGCGAATTGACTCTAGATTCAATTCGTGATGTGATTTGCTGCATGAATCGTAATT
TTCGTAACTTGCGTCACGATTCGATTCATCGATGAATTGTCATTTTTTATATGTGAAATA
```
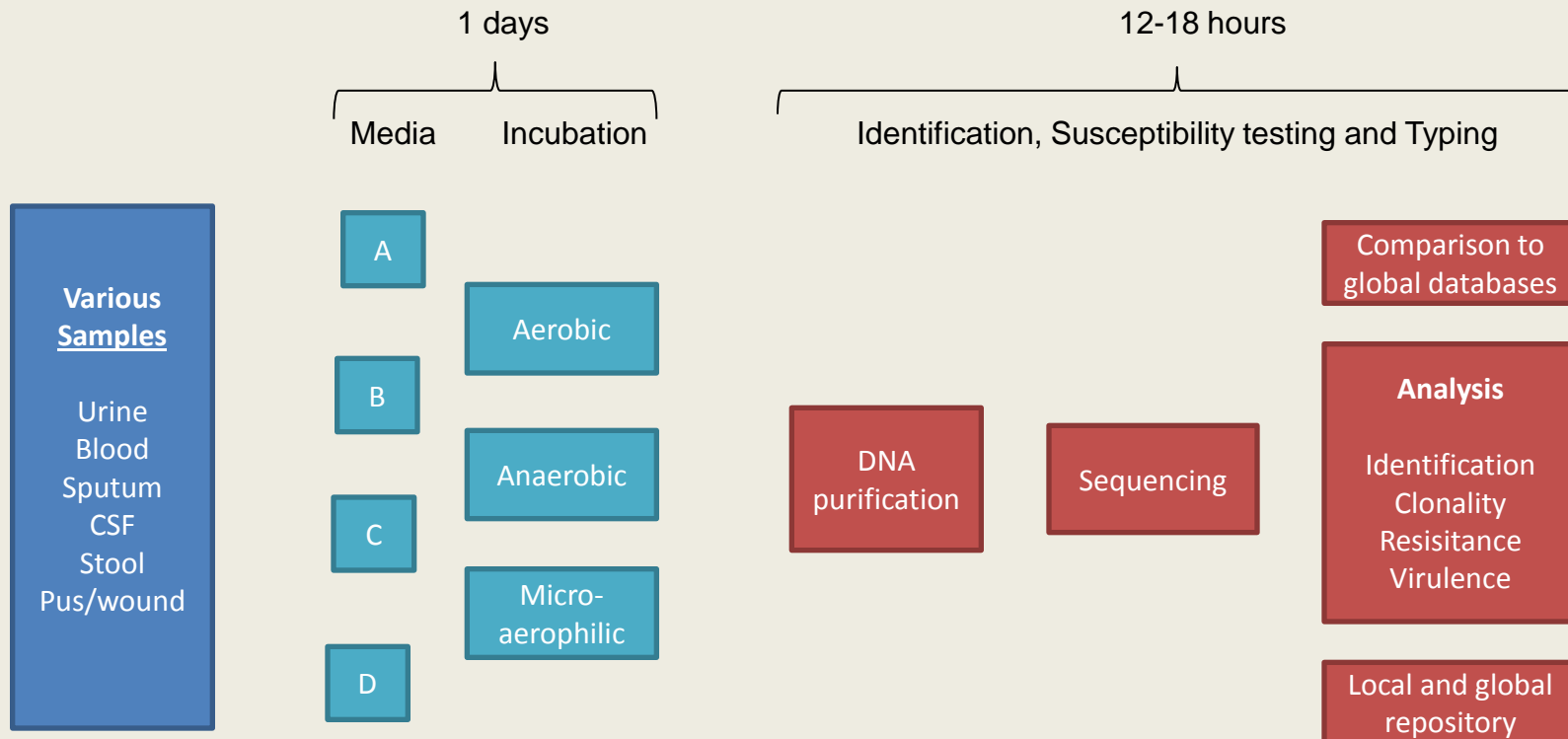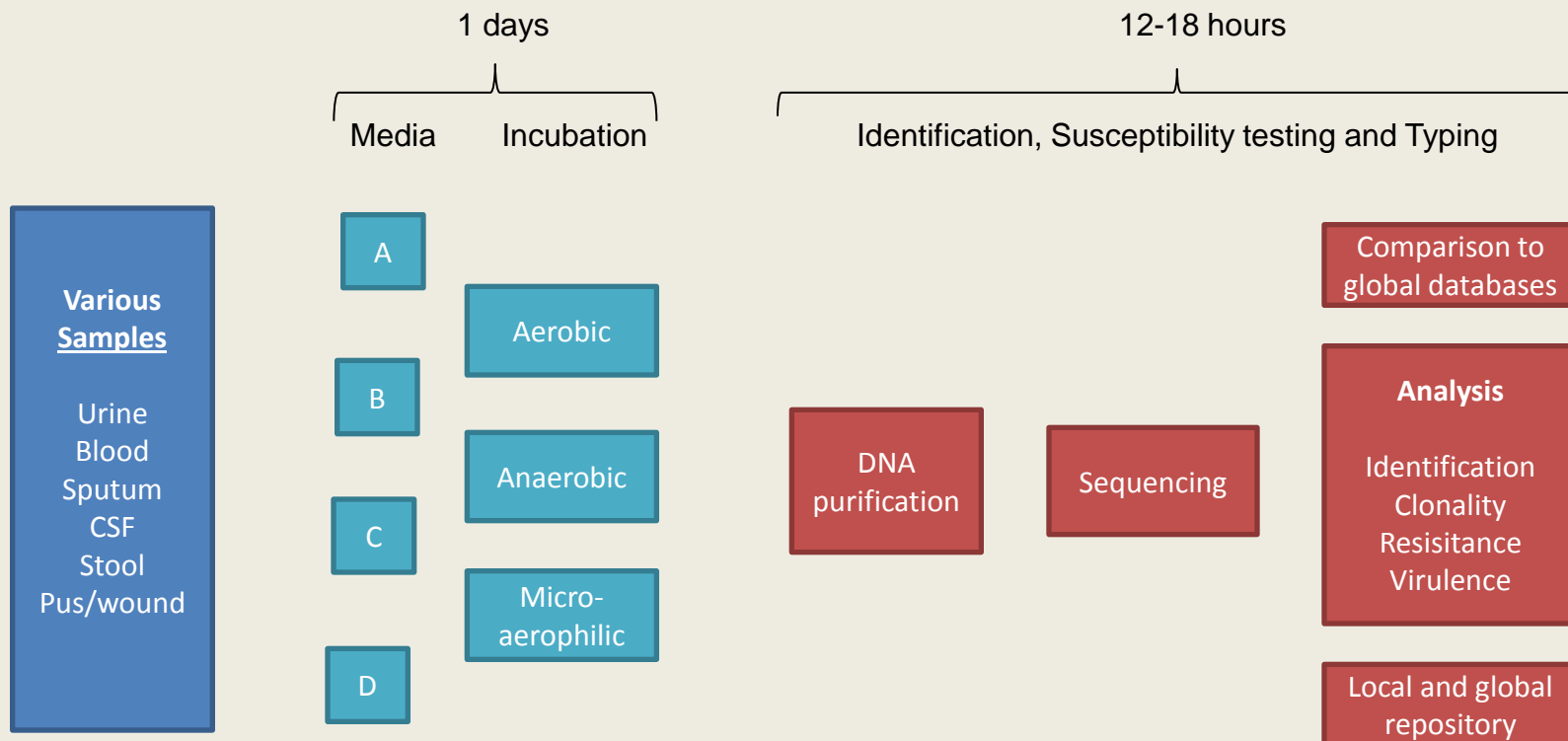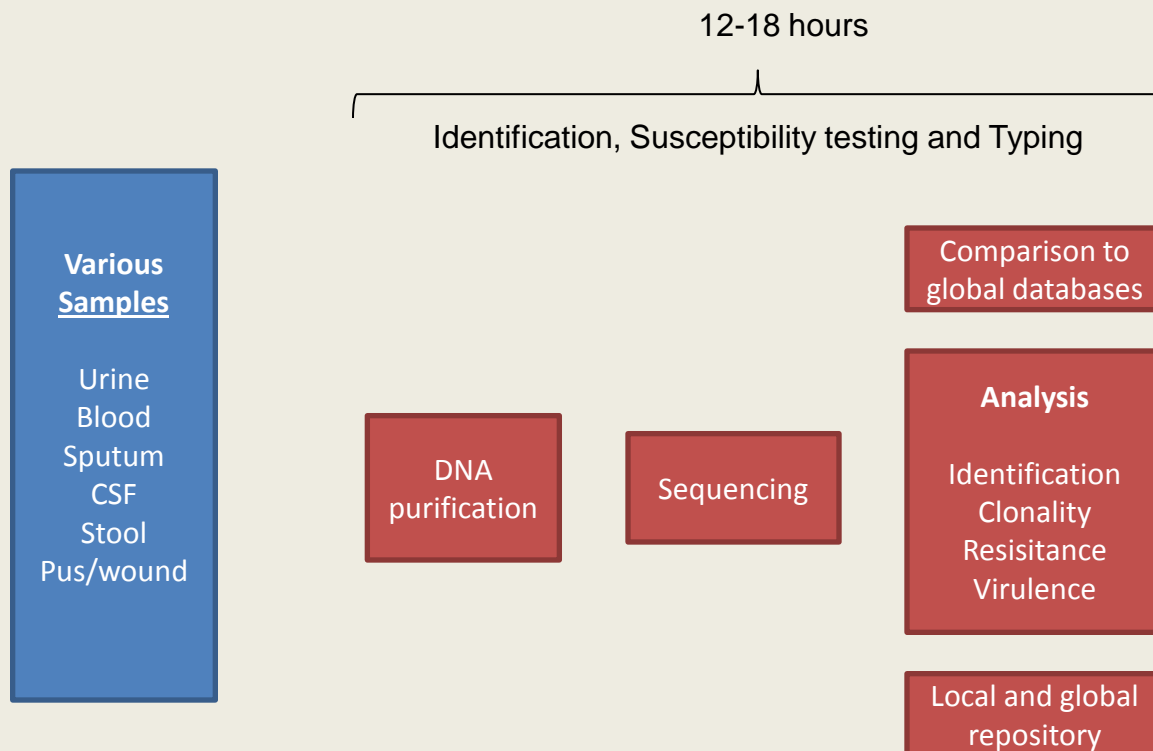
Bioinformatics (3)

Conventional microbiology

Various Samples

Urine
Blood
Sputum
CSF
Stool
Pus/wound

1 days

Media | Incubation

A

B

C

D

Aerobic

Anaerobic

Micro-aerophilic

1-2 days

Identification | Susceptibility testing

Gram stain | Biochem tests

PCR

Enterobacteriaceae

Streptococci

Staphylococci

Fastidious microbes

Other microbes

weeks

Typing

Species and sub-species typing

Credits to Henrik Hasman et al.

Bioinformatics (4)

Whole genome sequencing

1 days

12-18 hours

Media | Incubation

Identification, Susceptibility testing and Typing

Various Samples

Urine
Blood
Sputum
CSF
Stool
Pus/wound

A

B

C

D

Aerobic

Anaerobic

Micro-aerophilic

DNA purification

Sequencing

Comparison to global databases

Analysis

Identification
Clonality
Resisitance
Virulence

Local and global repository

Credits to Henrik Hasman et al.

Whole community sequencing

KCRI-DTU Genomics project

Objectives

1. To set up genomics/bioinfo units at KCMC/KCRI
2. To use the genomics/bioinfo tools for diagnosis
3. To share the genomics/bioinfo tools with others

Workflow infrastructure

Sequencing

Storage server
Local repository

Analysis server
(Genie)

Global repositories

Web server

Sample preparation

External user

KCRI user
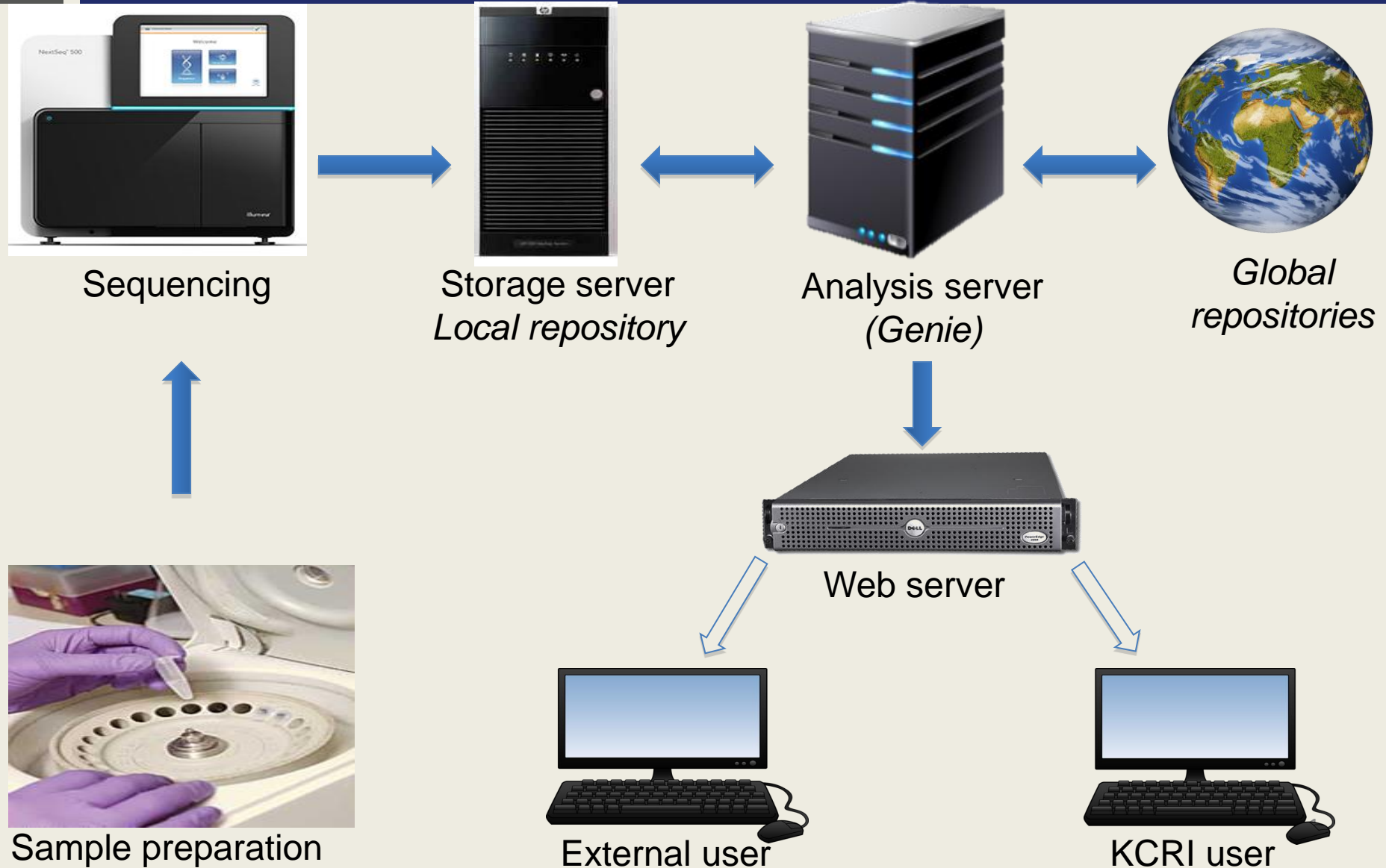
The portal

kcri applications portal

home   info sharing   data management   **bioinformatics**   contact

Quick Links

OpenClinica

HighRif 2
OEBA
WGS
TanZamBo
Muleba
Haydom
HR2 Test
WGS Test
Mul/TZB Test

Genie

Genie Homepage
MLST
PlasmidFinder
KmerFinder
pMLST
SpeciesFinder
ResFinder
NDtree
snpTree
Assembler

Alfresco

Bioinformatics Tools

The KCRI Bioinformatics Unit operates *"Genie"*, an application server offering a range of applications for genomic analysis. The Genie server was originally configured by researchers at the Center for Biological Sequence Analysis of DTU Copenhagen. It will be further developed jointly by KCRI and DTU, in the context of their collaboration programme focused on microbial genomics.

Genie's homepage is here. It offers the following services:

Typing and Identification

- Multi-locus sequence typing (MLST)
- Plasmid Finder
- K-mer Finder
- Plasmid MLST (pMLST)
- Species Finder

Resistance determination

- Resistance Finder (ResFinder)

Phylogeny analysis

- Nucleotide Difference (NDtree)
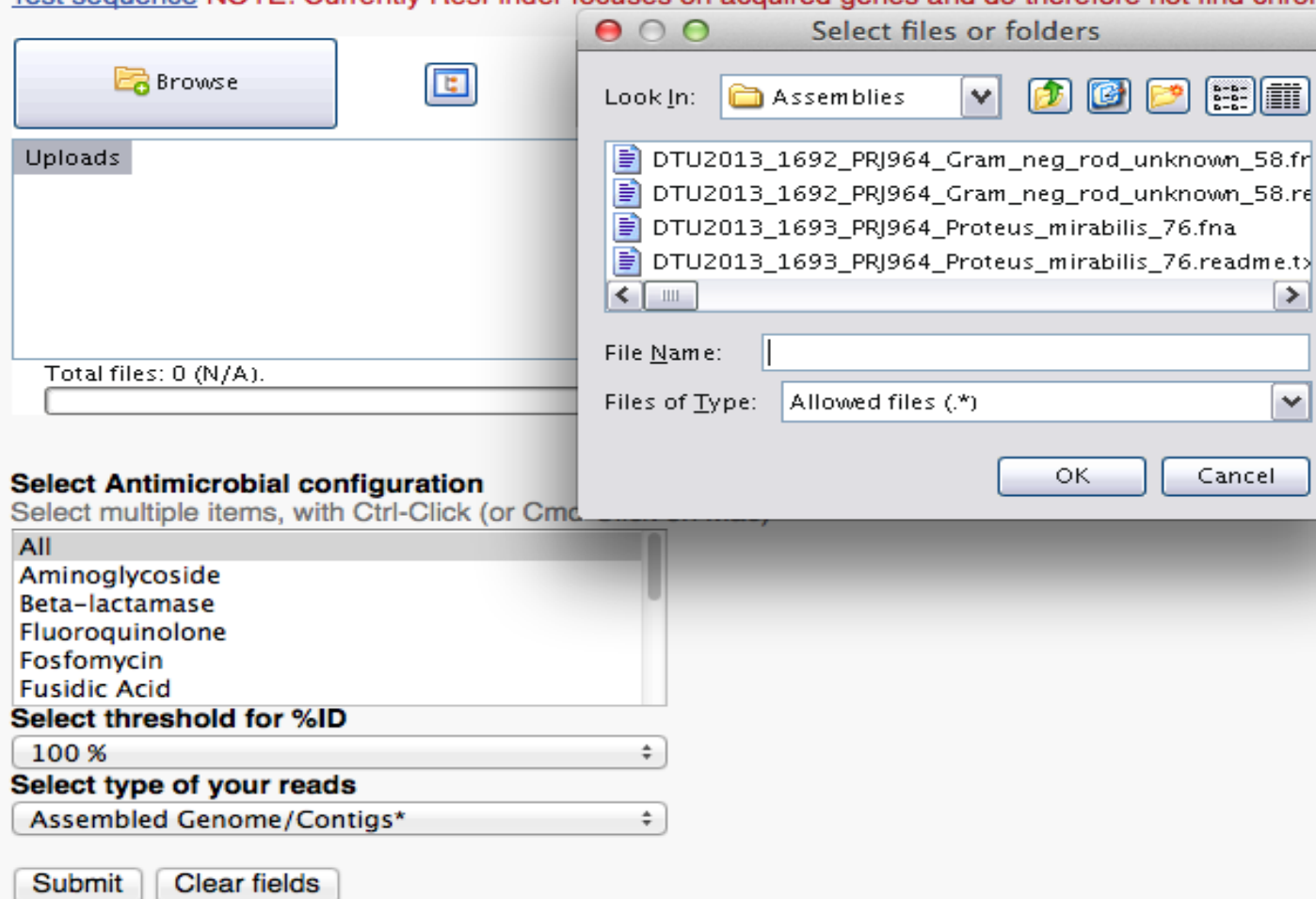- Single Nucleotide Polymorphisms (snpTree)

Determining resistance

Resistance results

ResFinder-1.3 Server - Results

| Aminoglycoside						
Resistance gene	%Identity	HSP/Query length	Contig	Position in contig	Predicted phenotype	Accession number
aac(6')-Iaa	100.00	438 / 438	NODE_46_length_597357_cov_13.450850	389085..389522	Aminoglycoside resistance	NC_003197
aadA1	100.00	792 / 792	NODE_136_length_15228_cov_8.325453	6404..7195	Aminoglycoside resistance	JX185132
strA	100.00	804 / 804	NODE_136_length_15228_cov_8.325453	2952..3755	Aminoglycoside resistance Alternate name; aph(3")-Ib	AF321551
strB	100.00	837 / 837	NODE_136_length_15228_cov_8.325453	3755..4591	Aminoglycoside resistance Alternate name; aph(6)-Id	M96392

Beta-lactam						
Resistance gene	%Identity	HSP/Query length	Contig	Position in contig	Predicted phenotype	Accession number
blaTEM-1	100.00	861 / 861	NODE_51_length_1988_cov_8.507042	607..1467	Beta-lactam resistance Alternate name; RblaTEM-1	JF910132

Fluoroquinolone
No resistance genes found.

Fosfomycin
No resistance genes found.

Molecular relatedness
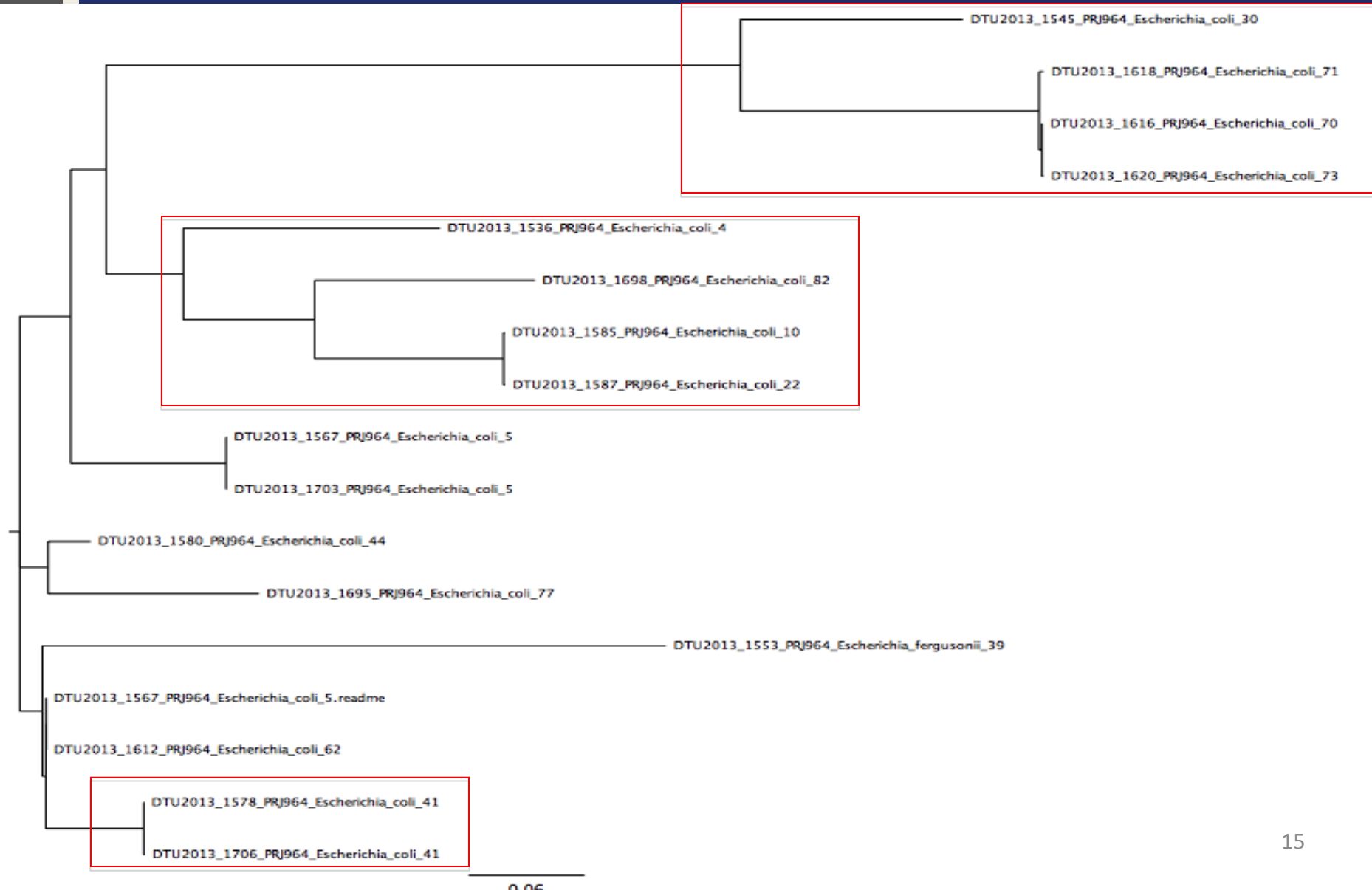
Determination of resistance due to ESBL-producing bacteria at KCMC

Introduction

- ESBL
  - Enzymes that mediate resistance to extended-spectrum (third generation) cephalosporins (e.g., ceftazidime, cefotaxime, and ceftriaxone) by hydrolyzing β-lactam ring http://www.cdc.gov/hai/settings/lab/lab_esbl.html
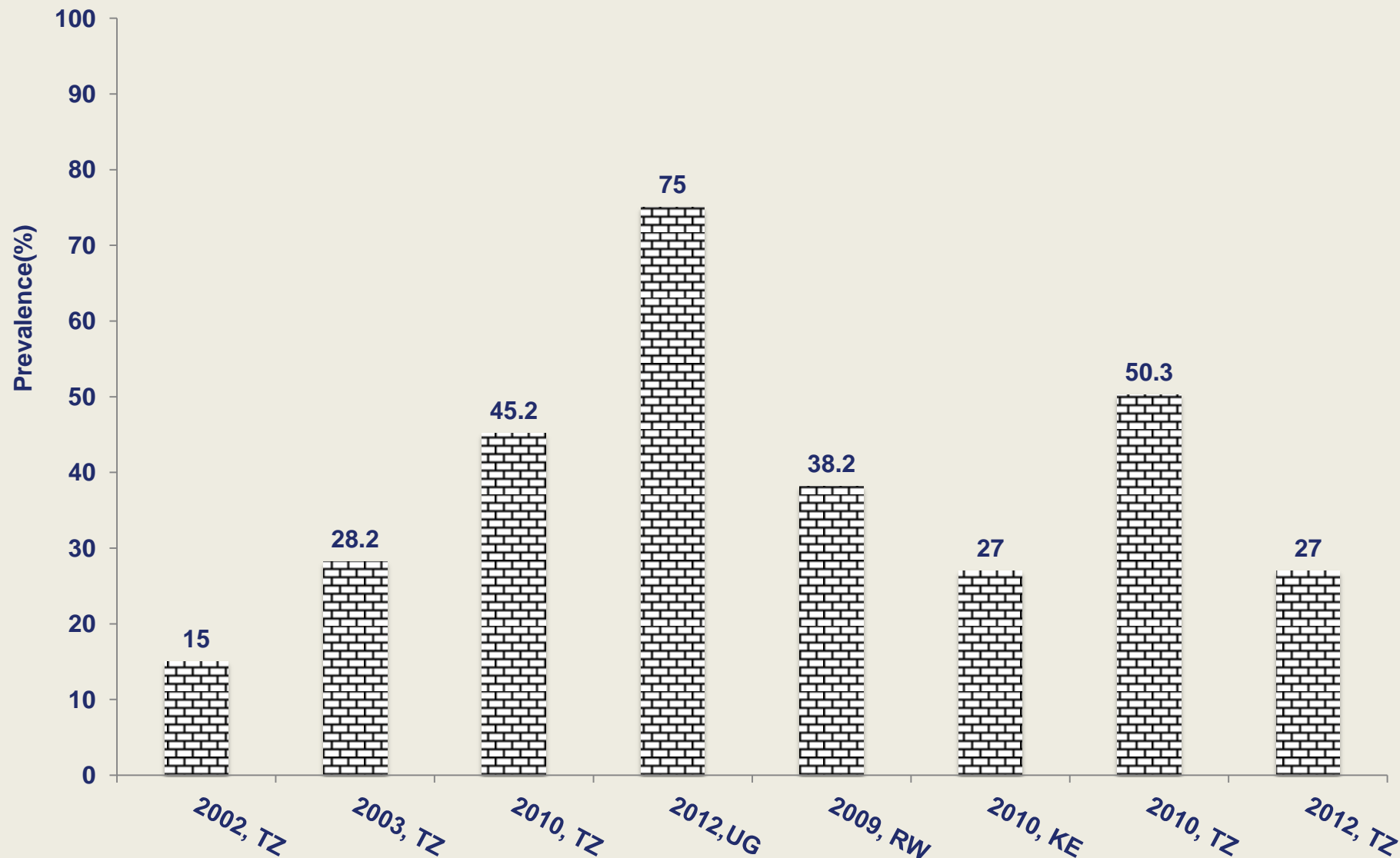- Limited choice of drugs for treatment
- More health-care resources due to
  - Long hospitalisation time
  - Drug costs: 15 times more expensive
- Increased risk of clinical outcomes and death

Lit. on ESBL in east African hospitals

Methods

- Whole genome sequencing
  - MiSeq system (Illumina®)
- Bioinfo analysis
  - *ResFinder*
    - Threshold set: 100% ID
    - Minimum length of nucleotides: 60%

Results

- Overall ESBL burden: 54.2%
- Highest prevalent gene: blaOXA (30.1%)
- Lowest prevalent gene: blaSHV 6.0%

Conclusion

- This highlights the need for
  - ESBL screening in our hospital
  - Robust(one-in-all) diagnostics that can timely and accurately identify infectious disease agents

Genomics/bioinfo publications

Rapid Whole-Genome Sequencing for Detection and Characterization of Microorganisms Directly from Clinical S...

Rapid Whole-Genome Sequencing for Investigation of a Suspected Tuberculosis Outbreak

Whole-genome sequences of Chlamydia trachomatis from clinical samples without culture

...Harris,[1] Rachel J. Skilton,[2] Frans M. Radebe,[3] ...[6] Paul Scott,[1] Lesley T. Cutcliffe,[2] ...Catherine A. Ison,[8]

Whole-genome sequencing to establish relapse or re-infection with Mycobacterium tuberculosis: a retrospective observational study

Josephine M Bryant, Simon R Harris, Julian Parkhill, Rodney Dawson, Andreas H Diacon, Paul van Helden, Alex Pym, Aziah A Mahayiddin, Charoen Chuchottaworn, Ian M Sanne, Cheryl Louw, Martin J Boeree, Michael Hoelscher, Timothy D McHugh, Anna L C Bateson, Robert D Hunt, Solomon Mwaigwisya, Laura Wright, Stephen H Gillespie, Stephen D Bentley

Acknowledgment

Collaborators