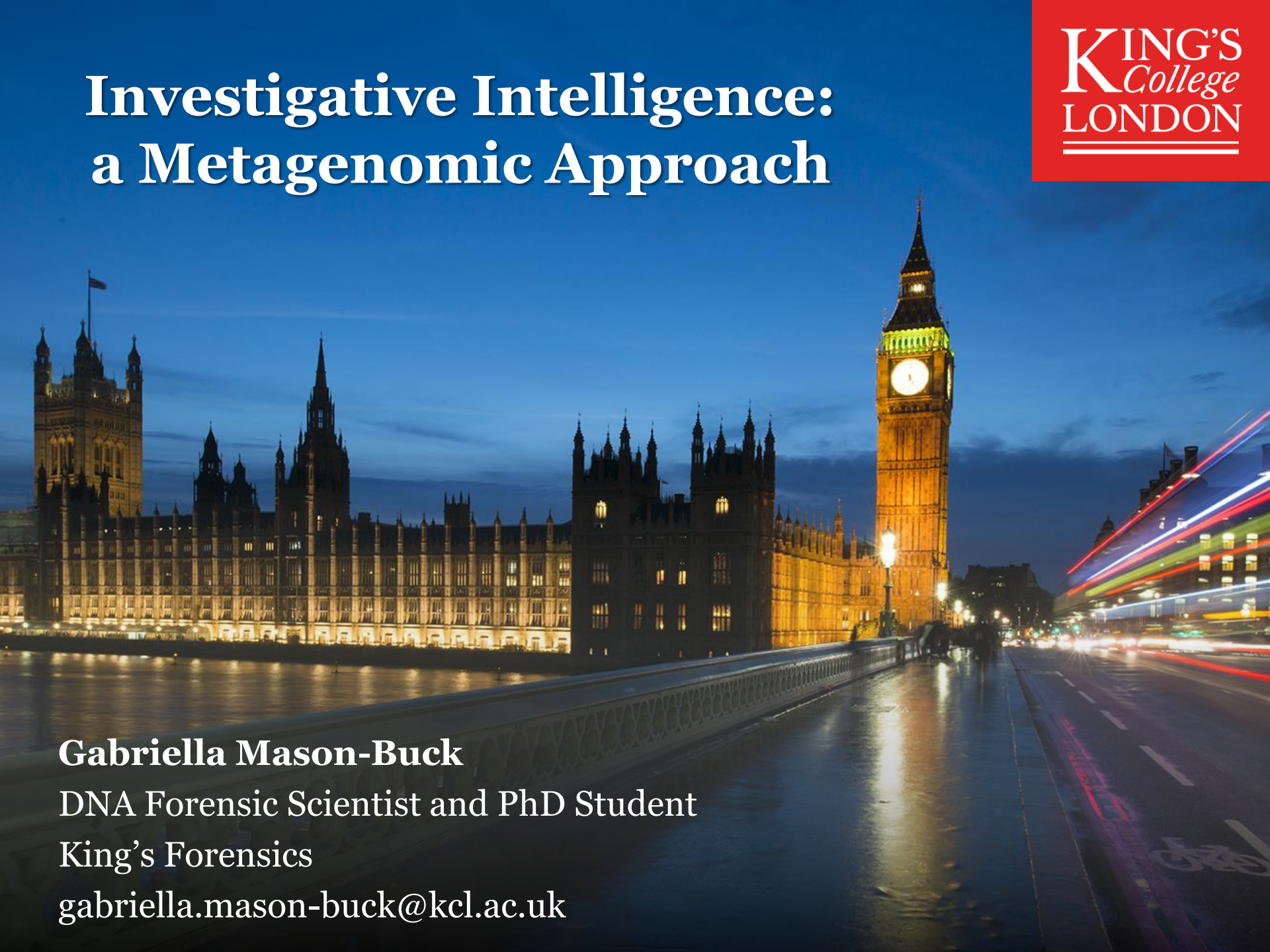
Investigative Intelligence: a Metagenomic Approach



Gabriella Mason-Buck
DNA Forensic Scientist and PhD Student
King's Forensics
gabriella.mason-buck@kcl.ac.uk

Background

Metagenomics

- The study of all organisms present in a specific environment:
  - Influence
  - Combined function
- ‘Meta’ (μεΤα) – ‘transcendence’, the ability to go beyond ordinary limitations
- ‘Genomic’ total hereditary material within an organism
- *omics* refers to studying every aspect of a given area

The Microbiome

- Microbiome – human genome in its entirety and the microbial communities that colonise the body
- Walking petri dish, more microbe than human!
- 10 x more bacterial cells than human cells
  - More recent studies have reduced this estimate 3:1 or even the same number
- Human DNA contains viral parasites – 50% parasite DNA

Metagenomic Tools

Initially (specific approaches):

- 16S rRNA -Bacteria and Archaea
- 18S rRNA – Eukaryotes
- trnL intron – Plants
- ITS rDNA – Fungi

Now (shotgun approach):

- Bioinformatic analysis identifies sequences

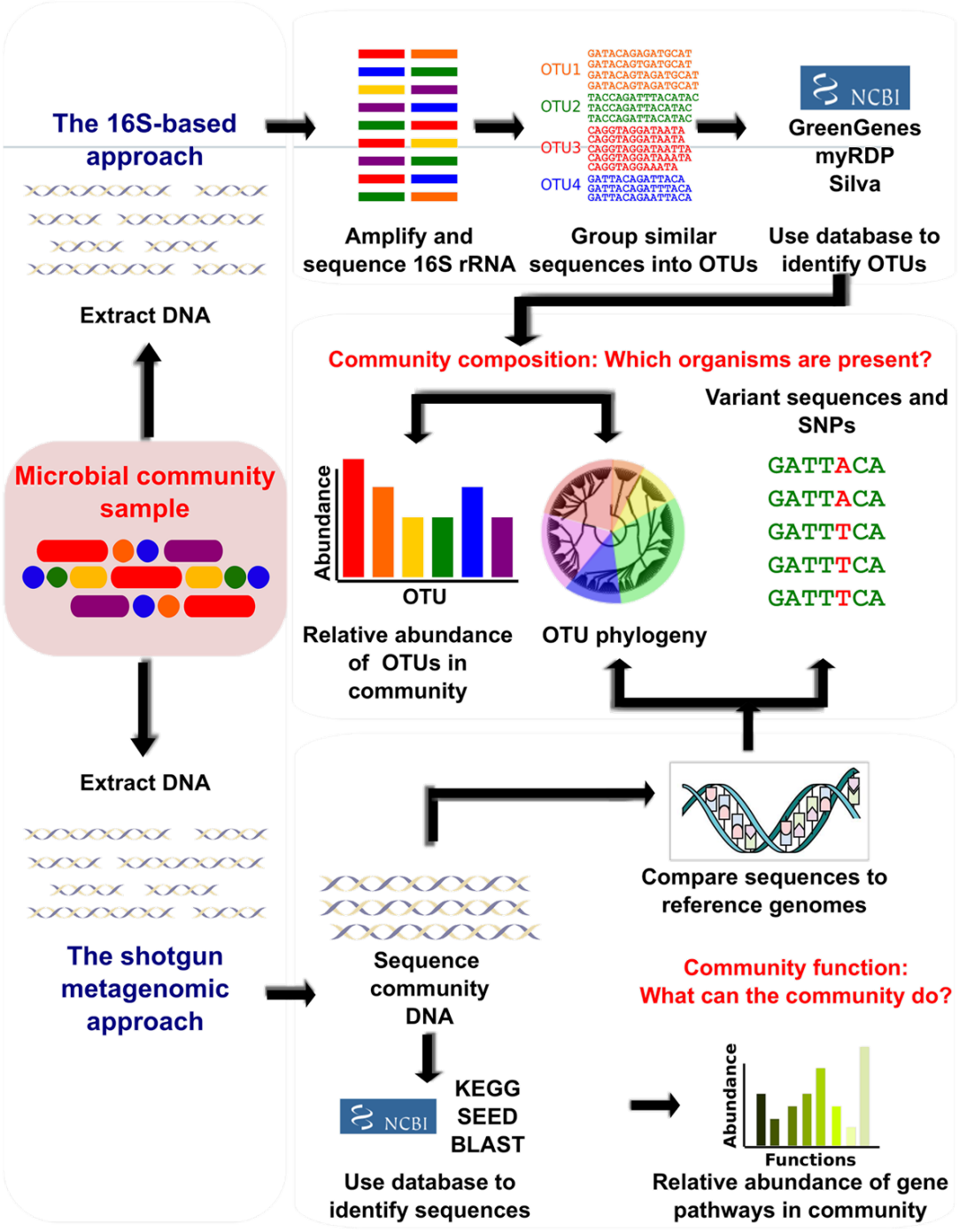


Image:
http://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1002808&type=printable

Why do we need this tool?

Recommendation from the National Research Council:

There is a need to 'precisely identify and characterise microbes that have played a role in war, terrorism, and crime events, thus contributing to discovering the source of the microbes and the party responsible for their use'

Handelsman *et al* 2007

'The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet'

Microbial Forensics

"A scientific discipline dedicated to analyzing evidence from a bioterrorism act, biocrime, or inadvertent microorganism/toxin release for attribution purposes" - Budowle 2003

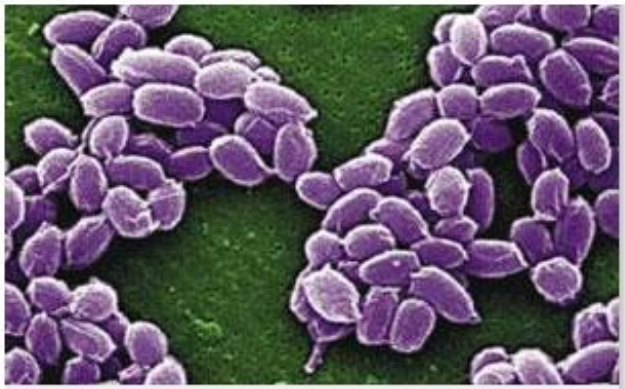
Published online 13 August 2008 | *Nature* **454**, 813 (2008) | doi:10.1038/454813a

News

Anthrax case ignites new forensics field

Biochemical method of tracking microbes hits the limelight.

Amber Dance

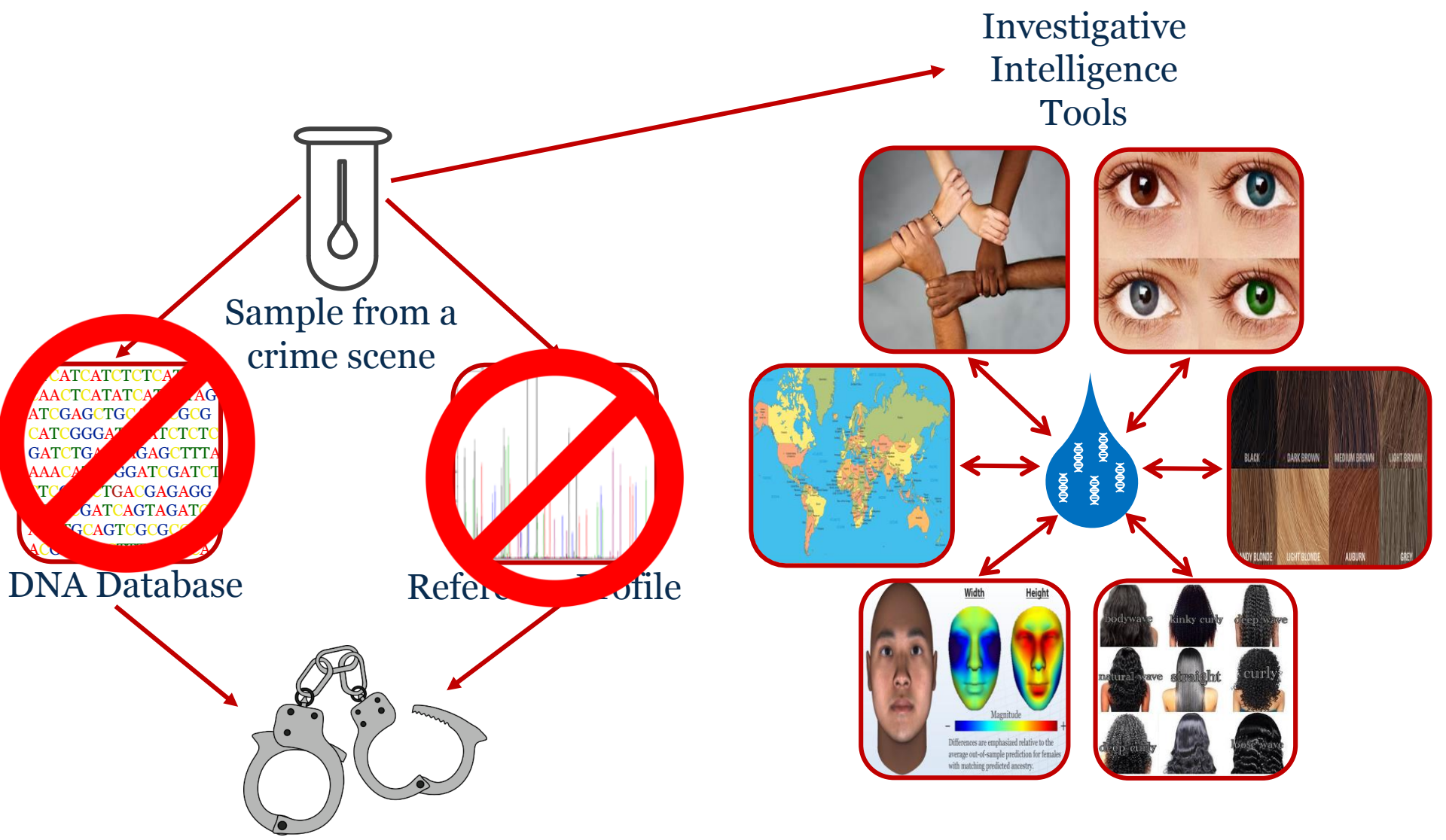


Anthrax spores were forensically analysed.

J. HANEY CARR/CDC

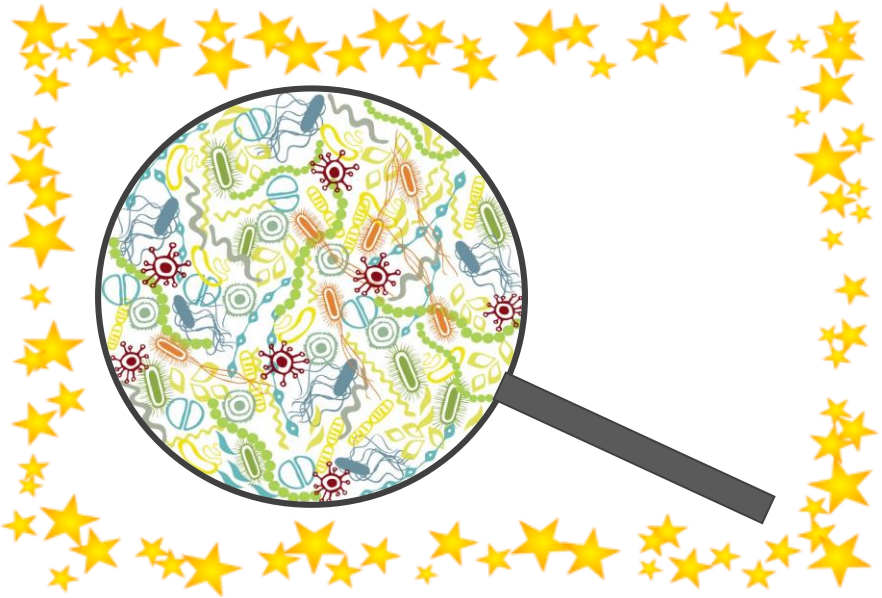
The messy tragedy surrounding the 2001 US anthrax attacks and the suicide of chief suspect Bruce Ivins has thrown the emergent field of microbial forensics into the spotlight. The forensic techniques proved vital in allowing the Federal Bureau of Investigation (FBI) to make its case that the anthrax used in the attacks came from a particular sample in Ivins's lab.

Current DNA identification tools



Forensic Metagenomics

Project Aim: to simultaneously amplify and sequence trace taxonomic material from bacteria, fungi, viruses, archaea and protozoa along with human and animal DNA for forensically relevant purposes



Can we obtain information at the activity level, such as:

- Where has someone been?
- What have they touched leading up to, during and after an event?

Taxonomic Classification

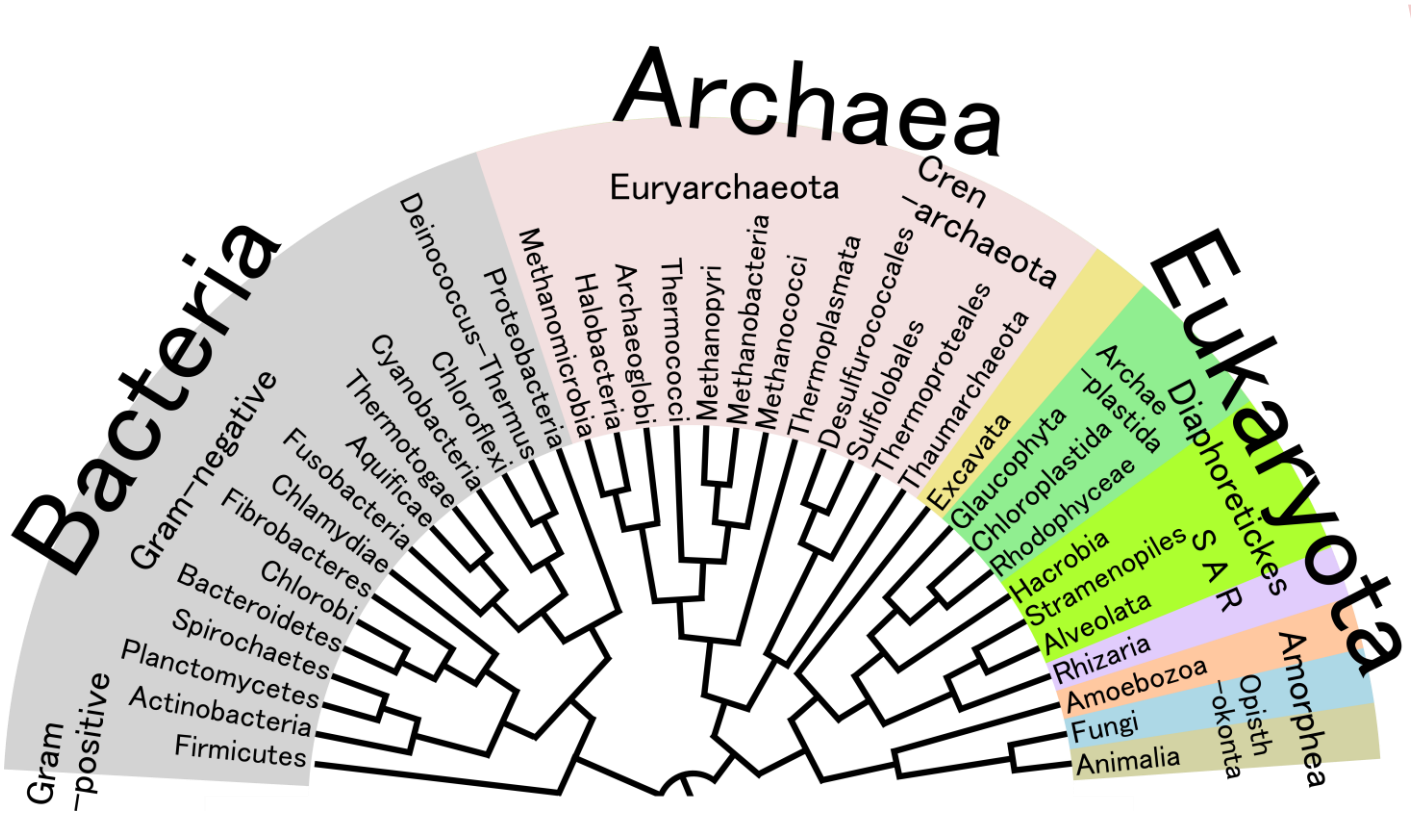


Figure: Adl, Sina M.; Simpson, Alastair G. B.; et al. (2006). "Toward Automatic Reconstruction of a Highly Resolved Tree of Life". *Science* **311** (5765): 1283–1287. DOI:10.1126/science.1123061. PMID 16513982

Method Trial

Experimental

Sampling

Extraction

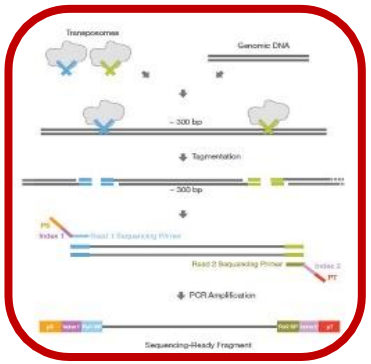
Heptamer of random nucleotides

5'- G T T C T A C A C G T C A C T G C A G N N N N N N N T G G C - 3'

Adaptor sequence

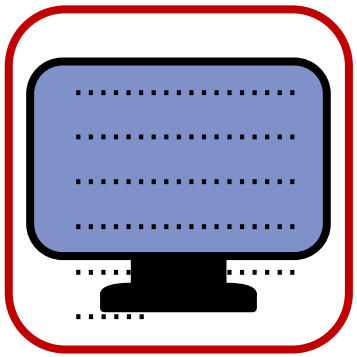
Clamp

Amplification

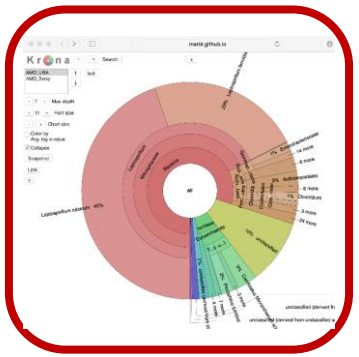


Library preparation

Massively Parallel Sequencing



Bioinformatic analysis



Data visualisation

Results

General findings

- Reads aligning to the human genome are most abundant
- Propionibacterium acnes (*P.acnes*) is the most abundant species on touched areas – different strains observed
- Sample content can infer sampling location. Eucalyptus found from sample taken from a bathroom – soap/cleaning product?

Issues

- Variability in the number of taxa identified across 3 different bioinformatic pipelines
- Taxa observed within negative controls and across all samples, background contamination – 'The Kitome'

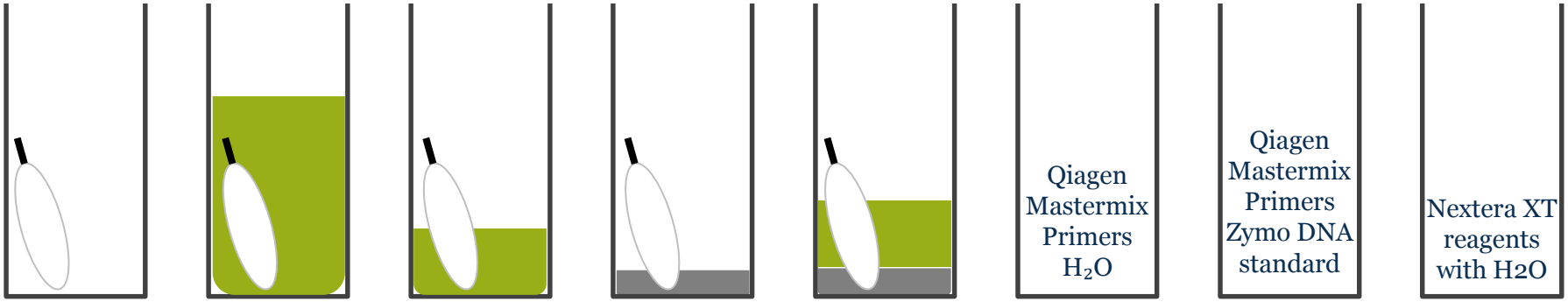
Exploring the issues

Experimental

(Sample number = 8 in triplicate = 24)

= Zymo Microbial Community

= Biological sample (saliva)



A – Extraction negative control
B – Extraction positive control
C – Extraction positive reduced volume
D – Sample only
E – Internal positive control
F – Amplification negative control
G – Amplification positive control
H – Library preparation control

Positive Control - ZymoBIOMICS™

Species	Gram stain	Theoretical composition (%) Genomic DNA	Genome size (Mb)
Pseudomonas aeruginosa	-	12.0	6.77
Eschericia coli	-	12.0	5.47
Salmonella enterica	-	12.0	4.83
Lactobacillus fermentum	+	12.0	2.08
Enterococcus faecalis	+	12.0	3.01
Staphylococcus aureus	+	12.0	2.93
Listeria monocytogenes	+	12.0	2.95
Bacillus subtilis	+	12.0	3.98
Saccharomyces cerevisiae	Yeast	2.0	13.3
Cryptococcus neoformans	Yeast	2.0	18.9

- Mock microbial community
- Quality control
- Well defined composition
- Available as a DNA standard

Source: ZymoBIOMICS™ Microbial Community Standards, zymoresearch.com

Results


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A1
A2
A3
B1
B2
B3
C1
C2
C3
D1
D2
D3
E1
E2
E3
F1
F2
F3
G1
G2
G3
H1
H2
H3
Acinetobacter
Alicycliphilus
Alkalilimnicola
Anabaena
Atopobium
Bacillus
Bifidobacterium
Burkholderia
Campylobacter
Citrobacter
Corynebacterium
Cronobacter
Dickeya
Enterobacter
Enterococcus
Erwinia
Escherichia
Eubacterium
Haemophilus
Klebsiella
Kribbella
Lactobacillus
Listeria
Micrococcus
Microvirus
Mycoplasma
Nakamurella
Neisseria
Olsenella
Pectobacterium
Porphyromonas
Prevotella
Propionibacterium
Pseudomonas
Raoultella
Rhodomicrobium
Rothia
Saccharomonospora
Salmonella
Saprospira
Selenomonas
Serratia
Shigella
Staphylococcus
Streptococcus
Tannerella
Thermoanaerobacter
Thermus
Veillonella
Xylanimonas

Conclusion – Bioinformatic Pipelines

EDGE

- Open-access (free!)
- Quick analysis time
- Quick to learn
- No need for ‘supercomputer’

- Bacterial and viral genomes
- 6- tool ‘compendium approach’
- Strain level analysis

[dstl]

- Tailor-made and adjustable
- Slower analysis time
- Coding knowledge required
- Specific computer specification

- All taxonomy
- MEGAN 6 visual output
- Species level analysis

Conclusion – Background Contamination

- Most contaminating taxa introduced at the extraction stage
- Contaminating taxa found in high abundance in low-level samples and found in lower abundance in DNA-rich samples
- *Corynebacter* found in high level samples but originates only from sample H (library preparation)
- Contaminating species found within the Zymo positive control (*Shigella flexneri*, various phage, *etc.*)

MetaSUB

MetaSUB

Metagenomics and Metadesign of the Subways and Urban Biomes

Aims:

- To create geospatial metagenomic and forensic genetic maps
- To identify and track antimicrobial resistance markers (AMRs) in the urban built environment
- To identify novel biosynthetic gene clusters (BCGs) for drug discovery

Weill Cornell Medicine

Date: 21st June 2018

Number of stations: 272

Number of samplers: 35

Sampling areas:

- bench on the platform
- handrail or elevator button
- ticket machine

Bakerloo
Central
Circle
District
District open weekends, public holidays and some Olympia events
Hammersmith & City
Jubilee
Metropolitan
Northern
Piccadilly
Victoria
Waterloo & City
DLR
London Overground
Emirates Air Line
Interchange stations
Step-free access from street to train
Step-free access from street to platform
National Rail
Riverboat services
Tramlink
Airport
Emirates Air Line

Acknowledgements

DNA analysis at King's:

Prof. Denise Syndercombe Court

Dr. David Ballard

Rebecca Smith

Ye Jee Roh

Laurence Devesse

Dstl:

Dr. Carl Mayers

Kathy Gammon

Dr. James Taylor

Contact details

Gabriella Mason-Buck
+44 (0)20 7848 4161
gabriella.mason-buck@kcl.ac.uk
www.kcl.ac.uk/http://www.kcl.ac.uk/lsm/research/divisions/aes/research/forensic/index.aspx