Predicting origins and molecular mechanisms of antibiotic resistance: linking theory, experiment and clinical practice

A Data Management Plan created using DMPonline

Creator: Danna Gifford

Affiliation: University of Manchester

Funder: Wellcome Trust

Template: University of Manchester Generic Template

ORCID iD: 0000-0002-4990-8816

Project abstract:

Antimicrobial resistance evolution vastly outstrips antimicrobial discovery. There is therefore an urgent need for strategies that prevent resistance evolution. Combination antimicrobial therapy, using multiple drugs as a single treatment, is a strategy motivated by evolutionary biology. The rationale behind this approach is the assumption that pathogens must acquire multiple independent resistance mutations to resist the treatment. However, this assumption is violated by clinical infections, which often possesses standing genetic variation for resistance mutations, conjugative resistance plasmids and other mobile genetic elements. Further, the host microbiome may also provide a reservoir of resistance genes1. We have an insufficient understanding of how resistance evolution progresses in clinical pathogens during antibiotic treatment. This research aims to understand how standing genetic variation for resistance in clinical pathogens contributes to multi-drug resistance evolution during antibiotic treatment. To address this knowledge gap, I will use a combined approach of (1) experimental evolution of mixed populations of clinical pathogens (2) stochastic dynamic modelling of plasmid-mediated evolution, and (3) shotgun metagenomic sequencing of both infection and microbiome from patients given antibiotics.

Last modified: 22-01-2020

Predicting origins and molecular mechanisms of antibiotic resistance: linking theory, experiment and clinical practice

Manchester Data Management Outline

- None of the above
- No - only institution involved
- Acquire new data

Antibiotic resistance and bacterial growth curves, genomic data of bacteria

- University of Manchester Research Data Storage Service (Isilon)
- 1 - 8 TB
- No
- 5 - 10 years
- No sensitive or personal data

No confidential data will be collected

- Not applicable
- Not applicable
- Not applicable
- No

Danna Gifford

2019-07-29

Project details

This project will help to establish whether antibiotic combination therapy is a potential strategy for preventing the emergence of antibiotic resistance resistance in mixed populations of clinical pathogens. The potential benefits of this research will be two-fold: characterising whether combination therapy is effective, and providing a methodology for evaluating future antibiotic treatments. Further, for instances where combination treatment fails through resistance evolution, we will be able to characterise the genomic basis for resistance evolution, which can help identify new genetic targets against multi-drug resistance pathogens. Finally, our project also has a fundamental biology aspect, which will be to test hypotheses arising from from eco-evolutionary theory regarding the contribution of genetic diversity to evolution. This will help further our understanding of how populations of bacteria evolve in the real world.

The research is subject to the policies set out by the Wellcome Trust, currently published here:
https://wellcome.ac.uk/funding/guidance/data-software-materials-management-and-sharing-policy

Responsibilities and Resources

The research fellow, Dr Danna Gifford.

Backup storage provided by The University of Manchester Research IT (Isilon) and publicly accessible databases (European Nucleotide Archive).

Data Collection

Type of study

The studies encompass both laboratory experimental evolution with*Escherichia coli* bacteria, and clinical monitoring of the microbiome of lung infections and the gut. This will involve allowing bacterial populations to evolve and measuring associated changes in antibiotic resistance phenotype and genotype. Human microbiome data will involve metapopulation sequencing to detect resistance mutations and genes.

Types of data

a) Quantitative data on bacterial growth and population characteristics from laboratory experiments. This will include the frequencies of plasmids in bacteria within populations, bacterial phenotyping (e.g. growth rate produced by spectrophotometer and flurometer in the presence and absence of antibiotics).

b) Qualitative data on new resistance mutations arising during laboratory experiments. This will include genomic sequencing data produced by Illumina HiSeq.

c) Quantitative data on bacteria present in human patients. This will include metapopulation genomic sequencing data produced by Illumina HiSeq.

Format and scale of the data

Raw data will be stored in open formats (e.g. text-based CSV, R data objects, current Flow Cytometry Standard format (FCS3.1 or newer), FASTQ). Data initially output into proprietary formats will be immediately exported to open formats. Only open-source analysis tools will be used for downstream analysis of data to ensure reproducibility (e.g. R, breseq). New software generated will be stored in open-source repositories (e.g. GitHub). The use of open formats will facilitate data sharing and long-term data accessibility.

Methodologies for data collection / generation

Standards for data collection will be set at the beginning of the project, but will be continually reviewed to ensure that best practices are being followed. This will include e.g. how often data points are collected, the criteria for inclusion in the study, and how negative and positive controls will be included to detect potential mistakes in experimental work. A schema for associating laboratory notebooks with collected data will be made to ensure that the correct metadata is associated with raw data.

Data quality and standards

To ensure data quality, data will be collected by skilled researchers with the appropriate training to use relevant research equipment. The equipment used has checks to ensure data integrity at the point of collection. Data quality will further be maximised through the use of appropriate statistical experimental design to minimise the possibility of spurious results arising due to stochastic noise. At the point of collection, data will be collected by skilled researchers trained FASTQ and FCS3.1 format includes extensive metadata on the machine used for collecting data. Data checksums will be used to ensure that files copied from local RDM provisions to public repositories are done so faithfully.

Documentation and Metadata

Metadata standards and data documentation

Metadata includes documentation of methods and procedures used to conduct experiments and collect samples. This metadata will be stored. This will be stored alongside the databases mentioned in 3.1, which are flexible and allow free-form text documents to be stored alongside data formats e.g. CSV.

Ethics and Legal Compliance

Patient data that will be associated with the analysis for this research only concerns which treatment Arm the microbiome sample came from. Other patient data will be processed in accordance with standard policies governing according to the Medicines and Healthcare products Regulatory Agency ethics review.

Anonymised data will be released under Creative Commons Licence 3.0 (CC-BY). External users will be bound by this licence, which is designed to facilitate reuse without restrictions, as long as the original contributor is acknowledged.

Storage and backup

Data will be stored to meet the standards of GDPR. In the short and medium term (i.e. before publication), data will be stored using The University of Manchester's dedicated Research Data Storage (RDS) facility, which offers 8 TB of backed-up data free at the point of use to research groups. On publication, bacterial phenotyping data will be stored alongside publications in open access databases (e.g. Dryad or Mendeley Data), although there is no community agreed/formal data standard. Bacterial genomic data will be stored in the European Nucleotide Archive (ENA, http://ebi.ac.uk), which allows storage of project metadata. The ENA is one of the community agreed databases for genomic sequence data.

The research fellow (Dr Danna Gifford) on the project will make the decision to supply data. In principle, anonymised data will be freely accessible without a need for a formal request. Data will be stored in publicly accessible repositories and databases.

The main risk to confidentiality is through unauthorised access to raw data, which can occur if data is stored on a device accessible to the general public. This risk will be mitigated by encrypting the hard drives of laptop computers, preventing access to data without a username and password. Further, the use of VPN via Global Connect will be used to access data on RDM servers. Both of these procedures are part of The University of Manchester's standard IT policy.

Selection and Preservation

Upon publication, raw data will be made available in a public repository (e.g. Dryad or GitHub) as appropriate

Data will be maintained in an established repository

Data Sharing

Before publication, data will be made available upon request to the PI. Once published, data will be made available in a public repository with a doi made available in the publication. Data associated with patient microbiomes will be anonymised prior to sharing. Data sharing policies and potential risks will be described to participants as part of informed consent associated with participation.

There are no anticipated restrictions on sharing data generated.