Plan Overview

A Data Management Plan created using DMPonline

Title: Suppressing mutation-mediated resistance through antibiotic combination treatment

Creator: Danna Gifford

Principal Investigator: Danna Gifford

Data Manager: Danna Gifford

Project Administrator: Danna Gifford

Affiliation: University of Manchester

Template: University of Manchester Generic Template

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.

ID: 77997

Start date: 01-03-2022

End date: 29-02-2024

Last modified: 11-06-2021

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Suppressing mutation-mediated resistance through antibiotic combination treatment

Manchester Data Management Outline

1. Will this project be reviewed by any of the following bodies (please select all that apply)?

- None of the above

2. Is The University of Manchester collaborating with other institutions on this project?

- No - only institution involved

3. What data will you use in this project (please select all that apply)?

- Acquire new data

Antibiotic resistance phenotypes (CSV) and bacterial growth curves (CSV), genomic data of bacteria (FASTQ and derivative analysis files), flow cytometry data (FCS3)

4. Where will the data be stored and backed-up during the project lifetime?

- University of Manchester Research Data Storage Service (Isilon)

5. If you will be using Research Data Storage, how much storage will you require?

- 1 - 8 TB

Genomic data and flow cytometry data are large files, justifying the need for more than 1 TB of space, despite the short duration of the project.

6. Are you going to be working with a 3rd party data provider?

- Yes

MicrobesNG (genomics company) stores the data for at least 1 year after acquisition.
Genomic DNA will also be uploaded to the European Nucleotide Archive, as is standard practice in the research field.

7. How long do you intend to keep your data for after the end of your project (in years)?

- 5 - 10 years

Questions about personal information

Personal information, also known as personal data, relates to identifiable living individuals. Special category personal data is more sensitive information such as medical records, ethnic background, religious beliefs, political opinions, sexual orientation and criminal convictions or offences information. If you are not using personal data then you can skip the rest of this section.

Please note that in line with data protection law (the General Data Protection Regulation and Data Protection Act 2018), personal information should only be stored in an identifiable form for as long as is necessary for the project; it should be pseudonymised (partially de-identified) and/or anonymised (completely de—identified) as soon as practically possible. You must obtain the appropriate ethical approval in order to use identifiable personal data.

8. What type of personal information will you be processing (please select all that apply)?

- No sensitive or personal data

9. Please briefly outline how you plan to store, protect and ensure confidentiality of the participants' information.

No participant data, including any confidential data, will be collected

10. If you are storing personal information (including contact details) will you need to keep it beyond the end of the project?

- Not applicable

11. Will the participants’ information (personal and/or sensitive) be shared with or accessed by anyone outside of the University of Manchester?

- Not applicable

12. If you will be sharing personal information outside of the University of Manchester will the individual or organisation you are sharing with be outside the EEA?

- Not applicable

13. Are you planning to use the personal information for future purposes such as research?

- No

14. Who will act as the data custodian for this study, and so be responsible for the information involved?

Danna Gifford

15. Please provide the date on which this plan was last reviewed (dd/mm/yyyy).

2021-05-21

Project details

What is the purpose of your research project?

This project will help to establish whether antibiotic combination therapy, involving antibiotics where resistance is achieved by mutation, is a potential strategy for preventing the emergence of antibiotic resistance resistance in 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 (efflux pumps, most likely), which can help identify new genetic targets against multi-drug resistance pathogens.

What policies and guidelines on data management, data sharing, and data security are relevant to your research project?

The research is subject to the policies set out by the Academy of Medical Sciences, which requires reporting on:
1. What data outputs will your research generate and what data will have value to other researchers?
2. When will you share the data?

3. Where will you make the data available?
4. How will other researchers be able to access the data?
5. Are any limits to data sharing required - for example, either to safeguard research participants or to gain appropriate intellectual property protection?
6. How will you ensure that key datasets are preserved to ensure their long-term value?
7. What resources will you require to deliver your plan?
8. Does your Institution have a data repository that is available to you?

Responsibilities and Resources

Who will be responsible for data management?

The PI, Dr Danna Gifford.

What resources will you require to deliver your plan?

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

Data Collection

What data will you collect or create?

Type of study

The studies encompass both laboratory experimental evolution with *Escherichia coli* bacteria. This will involve allowing bacterial populations to evolve and measuring associated changes in antibiotic resistance phenotype and genotype.

Types of data

a) Quantitative data on bacterial growth and population characteristics from laboratory experiments. This will include the frequencies of mutations 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.

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.

How will the data be collected or created?

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

What documentation and metadata will accompany the data?

Metadata standards and data documentation

Metadata includes documentation of methods and procedures used to conduct experiments and collect samples. This metadata will be stored with the data, and also available in all resulting publications. 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

How will you manage any ethical issues?

We do not anticipate any ethical issues arising from the data. Any ethical issues will be managed through referral to departmental or institutional ethics committees.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

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

How will the data be stored and backed up?

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.

How will you manage access and security?

The research fellow (Dr Danna Gifford) on the project will make the decision to supply data. In principle 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

Which data should be retained, shared, and/or preserved?

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

What is the long-term preservation plan for the dataset?

Data will be maintained in an established repository (European Nucleotide Archive for genomic data, Dryad for other types of data, GitHub for software pipelines).

Data Sharing

How will you share the data?

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.

Are any restrictions on data sharing required?

There are no anticipated restrictions on sharing data generated.