Plan Overview

A Data Management Plan created using DMPonline

Title: Microbial Mutation: Mechanisms, Measurements, Models

Creator: Danna Gifford

Principal Investigator: Chris Knight

Data Manager: Danna Gifford, Chris Knight

Project Administrator: Chris Knight

Affiliation: University of Manchester

Funder: Biotechnology and Biological Sciences Research Council (BBSRC)

Template: University of Manchester Generic Template

ORCID iD: 0000-0001-9815-4267

Project abstract:

Understanding and Predicting Mutation Availability in Evolution

Mutations are the raw material of evolution. In many cases, a single genetic change can determine whether an organism survives or perishes—whether a bacterium resists an antibiotic, a cancer cell evades treatment, or a species adapts to environmental change. However, mutations are rare, and only a small fraction are beneficial. Evolution, therefore, does not always act on the "best" possible mutation, but rather on the mutations that happen to arise.

Despite its importance, we do not fully understand what determines when and where mutations occur. Traditional models have treated mutation rates as fixed, but recent advances in genomics and single-cell studies reveal a far more dynamic picture. Mutation rates can vary between individuals in a population, across different environments, and even due to interactions between species. This project aims to develop a predictive framework for mutation availability by integrating theory and cutting-edge experimental approaches.

The Challenge

Current evolutionary models lack the ability to predict mutation availability with precision. While we can measure mutation rates, we do not yet have a comprehensive understanding of the factors that shape them. Cellular repair processes, environmental conditions, and even interactions between microbes can all influence mutation rates, but how these factors come together to shape evolution remains unclear. Without this knowledge, we struggle to anticipate how pathogens evolve resistance, how species adapt to environmental change, or how to control evolution for beneficial outcomes in synthetic biology.

Project Aims and Approach

Our research will establish fundamental rules for mutation availability by combining multiple experimental and theoretical approaches:

- Single-cell microfluidics (Platform A) will allow us to directly observe mutation-generating

processes in live bacterial cells, using fluorescence microscopy and mathematical modeling to understand how mutation rates vary in space and time.
- Large-scale microbial cultures (Platform B) will capture mutation dynamics in different growth conditions, providing a population-level perspective on mutation availability.
- Experimental evolution and sequencing (Platform C) will connect observed mutation patterns to their long-term evolutionary consequences, showing how mutations emerge and persist in adapting populations.

By integrating findings across these platforms, we will bridge gaps between molecular mechanisms, individual cell behavior, and large-scale evolutionary outcomes.

ID: 169818

Start date: 01-09-2025

End date: 30-09-2030

Last modified: 03-02-2025

Copyright information:

Microbial Mutation: Mechanisms, Measurements, Models

Manchester Data Management Outline

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

- Funder

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

- Yes - Part of a collaboration and owning or handling data

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

- Re-use existing data (please list below)
- Acquire new data

Single-cell microscopy images, single-protein tracking 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?

- > 8 TB

Microscopy data, genomic data and flow cytometry data are large files, justifying the need for more than 8 TB of space.

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).

2025-02-03

Project details

What is the purpose of your research project?

The project aims to determine the dynamics of multidimensional selection in bacteria, which will provide insight into evolutionary processes with implications for combating antibiotic resistance and advancing understanding in evolutionary biology.

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

From: https://www.ukri.org/wp-content/uploads/2021/07/data-sharing-policy-v1.22.pdf BBRSC Data management plan

All applications seeking research grant funding from BBSRC must submit a data management plan. This should include concise plans for data management and sharing as part of the research grant proposal, or provide reasons why data sharing is not possible or appropriate.

The plan will be included in applications as a separate mandatory attachment.

The page limit for the plan is maximum one side of A4.

You must use this document to cover the plans for data management and sharing. Use of this space allocation for any other purpose will result in withdrawal of the application. BBSRC reserves the right to withdraw proposals that do not adhere to these guidelines.

What to include

BBSRC recognises that plans for sharing data will vary according to the type of data collected. Data sharing should be driven by scientific benefit and should also be cost effective. Data should be shared using established standards and existing resources where this is possible.

You may wish to include details of:

- data areas and data types – the volume, type and content of data that will be generated, for example experimental measurements, models, records and images
- standards and metadata – the standards and methodologies that will be adopted for data collection and management and why these have been selected
- relationship to other data available in public repositories
- secondary use – further intended and/or foreseeable research uses for the completed datasets
- methods for data sharing – planned mechanisms for making these data available, for example through deposition in existing public databases or on request, including access mechanisms
- proprietary data – any restrictions on data sharing due to the need to protect proprietary or patentable data
- timeframes – timescales for public release of data
- format of the final dataset.

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 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 short read sequencing.

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 PI (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. Useful unpublished data will be made available in a public repository on conclusion of the project.

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.