Plan Overview

A Data Management Plan created using DMPonline

Title: The genetics of bacterial evolution and biodiversity

Creator: Diarmaid Hughes

Principal Investigator: Diarmaid Hughes

Data Manager: Diarmaid Hughes

Project Administrator: Diarmaid Hughes

Affiliation: Uppsala University

Funder: Swedish Research Council

Template: Swedish Research Council Template

ORCID iD: 0000-0002-7456-9182

Project abstract:

The aim of the project is to significantly advance and deepen knowledge of the mechanisms driving evolution and biodiversity in the bacterial kingdom. Bacterial evolutionary change is traditionally explained by the combined effects of mutation and horizontal gene transfer events. However, analysis of the ongoing flood of bacterial genomic data has revealed at least two features of bacterial chromosomes that appear to be discordant with the standard model of genetic change. These are, the recent discoveries of bacterial strains with hybrid chromosomes, and the prevalence of chromosomal gene order rearrangements between bacterial species, neither of which is easily explained by the traditional model. By combining experimental and *in silico* approaches the project will explore hypotheses to explain the mechanisms driving these changes, their relevance in clinical and natural isolates, and the genetic and environmental features that influence the rates of change associated with these mechanisms. The outcomes will influence our understanding in several important areas, including evolution of virulence and antibiotic resistant pathogens, and the processes of lineage separation and speciation in bacteria. This project concept is highly original and could fundamentally expand our understanding of mechanisms and rates of bacterial evolution.01

ID: 90763

Start date: 01-01-2022

End date: 31-12-2025

Last modified: 20-12-2021

Grant number / URL: 2021-04814

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

The genetics of bacterial evolution and biodiversity

General Information

Project Title

The genetics of bacterial evolution and biodiversity

Project Leader

Diarmaid Hughes

Registration number/corresponding, date and version of the data management plan

2021-04814

Version

1

Date

2021-12-12

Description of data – reuse of existing data and/or production of new data

How will data be collected, created or reused?

The project will generate data from genetic and physiological measurements of bacterial conjugation rates and selection of mutants.

What types of data will be created and/or collected, in terms of data format and amount/volume of data?

Rates of conjugation will be measured and bacterial strains will be sequenced at the local gene level and the whole genome level. The data will be in the form of text data.

Documentation and data quality

How will the material be documented and described, with associated metadata relating to structure, standards and format for descriptions of the content, collection method, etc.?

All details of data acquisition and analysis will be provided in published papers and their accompanying supplemental materials. Bacterial genome sequence data will be uploaded to the NCBI Short Read Archive (SRA) in conjunction with publication.

How will data quality be safeguarded and documented (for example repeated measurements, validation of data input, etc.)?

The precision and accuracy of the data will be described in the published paper. Appropriate controls and number of replicate measurements will will be in accordance with the requirements for the specific methods used. Bacterial genome sequence data will be uploaded to the NCBI Short Read Archive (SRA) in conjunction with publication.

Storage and backup

How is storage and backup of data and metadata safeguarded during the research process?

Raw data, including whole genome sequence data, will be stored on local computers with backup on an external site at UU (argus.rudbeck.uu.se).

How is data security and controlled access to data safeguarded, in relation to the handling of sensitive data and personal data, for example?

This project will not generate any sensitive data.

Legal and ethical aspects

How is data handling according to legal requirements safeguarded, e.g. in terms of handling of personal data, confidentiality and intellectual property rights?

Intellectual property rights belong to the scientist according to Swedish university regulations.

How is correct data handling according to ethical aspects safeguarded?

This project does not require any ethical considerations.

Accessibility and long-term storage

How, when and where will research data or information about data (metadata) be made accessible? Are there any conditions, embargoes and limitations on the access to and reuse of data to be considered?

Raw data will be stored on local computers with a backup on argus.rudbeck.uu.se (an external storage site at Uppsala University). Processed and analysed data will be published, including as supplementary materials along with publications of papers.

In what way is long-term storage safeguarded, and by whom? How will the selection of data for long-term storage be made?

Processed data will be published and accessible in published papers and their supplementary materials. In addition, genome sequence data will be deposited at the NCBI SRA (Short read archive) in connection with publication.

Will specific systems, software, source code or other types of services be necessary in order to understand, partake of or use/analyse data in the long term?

No. All data data except genome sequencing data will be stored in human readable format in text files or excel sheets. Whole genome sequencing raw data will be stored as fast files that are readable using standard bioinformatic tools available for analysing such data.

How will the use of unique and persistent identifiers, such as a Digital Object Identifier (DOI), be safeguarded?

NCBI SRA provides and guards the unique identifiers for genome sequence data, DOI's for publications are maintained by the publishers.

Responsibility and resources

Who is responsible for data management and (possibly) supports the work with this while the research project is in progress? Who is responsible for data management, ongoing management and long-term storage after the research project has ended?

The PI (Diarmaid Hughes) is responsible for data management and ongoing management. According to Swedish law all data from a research project should be stored for 10 years. It will be the responsibility of the PI to maintain the data, and ensure that if necessary this responsibility will be passed to another researcher in the department working in a related area.

What resources (costs, labour input or other) will be required for data management (including storage, back-up, provision of access and processing for long-term storage)? What resources will be needed to ensure that data fulfil the FAIR principles?

The PI (Diarmaid Hughes) is paying to store all data generated at an external site (argus.rudbeck.uu.se) in addition to storage on local computers and external hard drives. This cost is small relative to the general costs of running an experimental group. The data is stored in a form that is open and accessible.