Plan Overview

A Data Management Plan created using DMPonline

Title: PhD: Evolutionary models for describing regulatory diversity.

Creator: Peter Price

Principal Investigator: Alison Wright

Project Administrator: Alison Wright

Affiliation: The University of Sheffield

Funder: Natural Environment Research Council (NERC)

Template: NERC Template

Project abstract:

Adaptive evolution through changes in coding sequence is severely limited by functional constraints. As such, differential regulation of the genome is often central to the generation of novelty, complex body plans, and intraspecific diversity. We aim to explore the role of differential expression, differential splicing and gene regulatory network rewiring in the generation of sex differences across an avian lineage. As such we will use pre-existing and novel RNAseq datasets to answer questions regarding the evolutionary dynamics guiding regulatory diversity.

ID: 80600

Start date: 01-10-2020

End date: 01-10-2023

Last modified: 15-07-2021

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

PhD: Evolutionary models for describing regulatory diversity. - Outline DMP

Outline DMP

Project Title

Detecting signatures of selection in regulatory diversity.

Principal Investigator(s) / Grant Holder

Peter Price

Will the grant produce data?

- Yes

Yes. The project will generate Illumina paired RNAseq data and PacBio IsoSeq data for several avian species.

Nominated Data Centre(s)

- Environmental Information Data Centre (EIDC)

Briefly list the datasets that the project will produce. If the total is likely to be larger than 1TB please indicate.

Illumina Paired-end RNAseq data

Samples: five male and five female gonads across two developmental stages for five avian species

Species: Numida meleagris, Anas platyrhynchos, Meleagris gallopavo, Phasianus colchicus and *Taeniopygia guttata.*

Datatype: FASTQ

PacBio Isoseq data

Samples: pooled data from two male and two female gonads from two developmental stages from two species (one library per species)

Species: *Phasianus colchicus* and *Meleagris gallopavo*

Datatype: FASTA and BAM

PhD: Evolutionary models for describing regulatory diversity. - Full DMP

Project information

Project Name

Evolutionary models for describing regulatory diversity.

Project Number (NERC PIs only)

Question not answered.

Grant Reference

Question not answered.

Principal Investigator

Alison Wright

Organisation

Nominated Data Centre

- Environmental Information Data Centre (EIDC)

Data Centre Contact

Question not answered.

Project Data Contact

Question not answered.

Please specify any other team members with responsibility for data

Alison Wright

Roles and Responsibilities

Please state all roles and responsibilities throughout the project

Alison Wright was responsible for all embryonic disections.
Peter Price was responsible for RNA extraction from these samples, as well as processing for sequencing and management of data storage.

Data Generation Activities

What data will be created and how?

Illumina Paired-end RNAseq data
Samples: five male and five female gonads across two developmental stages for five avian species
Species: Numida meleagris, Anas platyrhynchos, Meleagris gallopavo, Phasianus colchicus and *Taeniopygia guttata.*
Datatype: FASTQ
PacBio Isoseq data
Samples: pooled data from two male and two female gonads from two developmental stages from two species (one library per species)
Species: *Phasianus colchicus* and *Meleagris gallopavo*
Datatype: FASTA and BAM

In-Project Data Management Approach

How will the data be managed?

Data has/will be downloaded instantly from the CGR servers and uploaded onto the SHARC computer servers and external hard drives.
Data storage will be managed to prevent accidental loss. Raw and processed sequencing data, as well as datasets and bioinformatics pipelines generated during the course of the work, will be securely backed up at multiple independent locations to avoid accidental loss through system failure at any single site. This includes internal servers at the University of Sheffield as well as external hard drives.

To ensure proper computational practices, I will also use Git (https://git-scm.com/) to version control my analytical pipelines. Laboratory work will be carefully recorded in notebooks.

Long-term data management will be achieved via online depositories, including NCBI Sequence Read Archive (SRA), Dryad and Github.

Metadata and Documentation

Outline plans for metadata, noting standards that will be used

All metadata will be stored with sequencing data including RNA extraction lab results.

Lab results include sample concentrations, purities, dates, and names of individuals whom collected the samples

To ensure proper computational practices, I will also use Git (https://git-scm.com/) to version control my analytical pipelines. Laboratory work will be carefully recorded in notebooks.

Data Quality

What procedures will be used to control data quality?

To ensure proper computational practices, I will also use Git (https://git-scm.com/) to version control my analytical pipelines. Laboratory work will be carefully recorded in notebooks.

I will follow standard protocols for the field to ensure data quality.

Exceptions or Additional Services

Clarify any support needed from data centres that exceeds the norm, and show this has been planned / costed.

Question not answered.

Data Management Plan Information

Author

Peter Price

Date

30th June 2021

Version Number

1

Approved by PI/PM

yes

Approved by (Data Centre)

Question not answered.

New Datasets

Digital Information

Dataset Description	Contact	Data Volume	Data Format	Issues	Delivery Date	Embargo Date	Reuse Scenario	Preservation Plan
PacBio Isoseq for Meleagris gallopavo and Phasianus colchicus	*Peter Price*	7.9GB	BAM and FASTA		*28th May 2021*			*Keep indefinitely*

Hardcopy Records

Question not answered.

Physical Collections & Samples

Question not answered.

Third Party/Existing Datasets

Third Party/Existing Datasets

Question not answered.