

## Plan Overview

---

*A Data Management Plan created using DMP Tool*

**Title:** A digital learning hub dedicated to learning molecular evolution, drawing insights from Ziheng Yang's seminal books Computational Molecular Evolution and Molecular Evolution: A Statistical Approach.

**Creator:** Sishuo Wang - **ORCID:** [0000-0002-7220-7305](https://orcid.org/0000-0002-7220-7305)

**Affiliation:** The Chinese University of Hong Kong

**Principal Investigator:** Sishuo Wang, Jianhao Lv

**Data Manager:** Sishuo Wang

**Project Administrator:** Sishuo Wang

**Template:** CUHK Data Management Plan Template

### Project abstract:

This project aims to provide details solutions to the around 100 exercises from the two prominent books (bibles in the field molecular evolution), Computational Molecular Evolution and Molecular Evolution: A Statistical Approach, authored by Professor Ziheng Yang, FRS.

The solutions will be made available online for researchers at no cost in the form of an online learning platform (<https://github.com/evolbeginner/Solutions-manual-for-CME2006-and-MESA2014>), allowing readers to comment and contribute their problem-solving approaches. Key objectives of the project include providing detailed answers from multiple perspectives for each exercise and establishing a global networked learning platform for scholars to study and provide feedback. The current study seeks to construct an **online educational platform** for biologists by offering detailed solutions to exercises from the field of molecular evolution.

**Start date:** 01-01-2024

**End date:** 12-31-2024

**Last modified:** 07-08-2024

**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 customize 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

---

**A digital learning hub dedicated to learning molecular evolution, drawing insights from Ziheng Yang's seminal books Computational Molecular Evolution and Molecular Evolution: A Statistical Approach.**

## **Data Collection**

---

**Will you create or collect data in your study?**

- Yes

**If yes, by what means will you create or collect the data?**

- Data repository

**What type of data will you create or collect?**

- Code

**What are the data formats?**

- Image - .jpg
- Image - .png
- Tabular - .xlsx
- Text - .pdf
- Text - .docx

**How will the data be organized?**

The data will be organized by content (substitution models, tree reconstruction, advanced phylogenetics) and by the types of the exercises (analytical derivation, coding, software operation, etc.).

## **Storage and Backup**

---

**How will the data be backed up?**

- On desktop / laptop
- Cloud storage
- On external harddisk

**How will the data be recovered in the event of an incident?**

- By secondary backup

## **Selection and Preservation**

---

**How will you decide what data to be kept or destroyed?**

- By value of data

### **How long will the data be retained and preserved?**

- Permanently

### **Where will the data be preserved?**

- CUHK Research Data Repository
- Others

github, figshare, and other online data repositories. DOI will be assigned so that everyone can access the any versions of the data at any time.

### **Will the data repository charge for depositing data?**

- No

## **Data Sharing**

---

### **Will you share the data created or collected in the study?**

- Yes

### **When will you make the data available for sharing?**

- Upon project completion

The data and results will be available once they are generated.

### **With whom will you share the data?**

- Public

The aim of the project is to inspire a community-effort to help people learn molecular evolution so anyone can easily and freely access the data.

### **How will potential users find out about your data?**

- Via data repository
- Via search engine

The data will be available by github, figshare. A pdf-formatted will also be posted to preprint server such as arxiv or hal such that a doi will be generated and the results can be permanently maintained online.

### **Will data sharing be restricted?**

- No

## **Documentation and Metadata**

---

**What documentation and metadata will be provided to help others discover and understand the data?**

- readme.txt

**What metadata standard will be used?**

- Dublin Core

## **Ethics and Legal Compliance**

---

**Will human participants be involved in your study?**

- No

**What are the risks to data security?**

- Others

Not applicable.

**How will confidential or sensitive data be handled to ensure it is stored and transferred securely?**

- Others

Not applicable.

**Who own(s) the data generated in your study?**

- Data creator

**How will the data be licensed for reuse?**

- CC BY-NC

**Are there any restrictions on the reuse of secondary data that were created by others?**

- Yes

**If yes, what are the restrictions?**

They need to follow CC BY-NC 4.0.

## **Responsibilities and Resources**

---

**Who will be responsible for the data management activity?**

- Principal investigator

- Data manager

**Will additional specialist expertise (or training for existing staff) be required to deliver your data management plan?**

- No

**Do you require hardware or software which is additional or exceptional to existing institutional provision?**

- Yes

**If yes, what hardware or software will you need?**

We'd like to buy math software like matlab and mathtype. We would also like to buy 2 hard drives to deposit the data for large-scale bioinformatics and statistical computation.

---

Planned Research Outputs

Data paper - "A Solutions Manual for The Exercises of Ziheng Yang's Computational Molecular Evolution, and Molecular Evolution: A Statistical Approach"

This project provides a solutions manual for all excercises in the Ziheng Yang's two seminal books on statistical molecular evolution: Computational Molecular Evolution (CME2006) [Yang 2006] and Molecular Evolution: A Statistical Approach (MESA2014) [Yang 2014]. We hope the solutions manual plays a tiny role in helping better understand Ziheng's books and appreciate molecular phylogenetics from a statistics perspective.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
A Solutions Manual for The Exercises of Ziheng Yan ...	Data paper	2024-11-30	Open	None specified	100 MB	Creative Commons Attribution Non Commercial 4.0 International	None specified	No	No