

## DATA ACCESS AGREEMENT

This Data Access Agreement (hereinafter referred to as the “Agreement”) sets forth the terms governing access to and use of the raw sequencing data of samples in the Indonesian Microbiome Ecology and Evolution (IndoMEE) database (hereinafter referred to as the “Project Data”) by individuals (hereinafter referred to as the “User”) who are not part of the Data Producer (details of which are set out in Appendix 1).

This Agreement is made between the Principal Investigator and the Data Producer, represented by a member of the Data Access Committee. The User agrees as follows:

## DEFINITIONS

**Project Data:** Raw metagenomic sequencing data underlying the Indonesian Microbiome Ecology and Evolution (IndoMEE) database, the details of which are set out in Appendix 1.

**Data Producer:** A consortium of individuals listed in Appendix 1, responsible for the production, development, management, and oversight of the Project Data.

**Data Access Committee:** A group of individuals responsible for overseeing and governing access to and use of the Project Data on behalf of the Data Producer. A representative of the Data Access Committee acts on its behalf for the purpose of this Agreement.

**Study Participant:** The individual whose data contributes to the Project Data.

**User:** The individual requesting access to the Project Data, and who is not part of the Data Producer consortium.

**User’s Institution:** The organization, institution, or entity with which the User is formally affiliated and through which the User is requesting access to the Project Data.

**User’s Research Project:** An investigation aimed at discovering, interpreting, or revising facts, theories, or applications of existing knowledge, for which the User is requesting the access and use of the Project Data.

**Principal Investigator:** The lead researcher responsible for the overall design, conduct, and management of the User’s Research Project. The Principal Investigator can also be the User.

**Research Purposes:** Research activities conducted as part of the User’s Research Project.

**Publications:** Includes, without limitation, articles published in print and electronic journals, reviews, books, posters, and other written, verbal, and multimedia presentations of research.

## DATA ACCESS POLICY

The **User** and the **Principal Investigator** agree to be bound by the terms and conditions of access set out in this **Agreement**, as follows:

1. The **User** agrees to only use the **Project Data** for **Research Purposes** as outlined in Appendix 2.
2. The **User** and **User's Institution** will not make intellectual property claims on the **Project Data** and will not use intellectual property protection in ways that would prevent or block access to, or use of, any element of the **Project Data**, or conclusions drawn directly from the **Project Data**.
3. The **User** agrees to protect the confidentiality of **Study Participants** at all times. The **User** shall not attempt to re-identify participants or extract personal information from the **Project Data** through direct or indirect methods, including combining it with data from other studies or repositories where that may compromise confidentiality. Furthermore, the **User** agrees to take all reasonable measures to preserve the anonymity of the **Study Participants** in any data storage or publications that contain the **Project Data** or any data derived from it.
4. The **User** agrees not to share the **Project Data** with any individuals, whether within or outside the **User's Institution**, who are not listed in Appendix 2.
5. The **User** agrees to recognise and acknowledge the contribution of the **Data Producer** in all reports or publications resulting from the use of the **Project Data**.
6. The **User** has to notify the **Data Access Committee** as soon as it becomes aware of a breach of the terms or conditions of this Agreement.
7. The **User** acknowledges that the **Data Producer** is not legally responsible for the accuracy, completeness, or reliability of the **Project Data**. The **User** also agrees that the **Data Producer** is not liable for any damages or losses, direct or indirect, arising from the use of the **Project Data** and any interruptions in access to the **Project Data**.
8. The **User** confirms that the information provided in Appendix 2 is accurate and up-to-date. Should there be any changes to the name, affiliations, and research purposes of the **Users**, they agree to promptly notify the **Data Access Committee** and execute a new agreement reflecting these changes.
9. The **Data Producer** may terminate this agreement by written notice to the **User**. If this **Agreement** terminates for any reason, the **User** will be required to destroy any **Project Data** held, including copies and backup copies. This clause does not prevent the **User** from retaining these data for archival purposes in conformity with audit or legal requirements.
10. The **User** accepts that it may be necessary for the **Data Producers** to alter the terms of this **Agreement** from time to time. In the event that changes are required, the **Data Producers** or **Data Access Committee** will contact the **User** to inform it of the changes and the **User** may elect to accept the changes or terminate the **Agreement**.

The **User** and **Principal Investigator** hereby acknowledge that they have read and understood the content and policies of this **Agreement**. By signing this document, the **Principal Investigator** accepts responsibility for ensuring the **User's** compliance with these policies. This **Agreement** is considered executed upon the signature of the **Principal Investigator**, with the approval of the **Data Access Committee** representative, who also signs this document.

**Signed by:**

**Principal Investigator**

Signature:

---

Name :

Date :

**Approved by:**

**Data Access Committee Representative**

Signature:

---

Name :

Date :

Institution :

Institutional email :

## APPENDIX 1 – DATASET DETAILS

This dataset comprises 124 human faecal metagenomic samples collected across Indonesia between 2015 and 2021, together with 6 controls. It was collected to study patterns of human microbiome variation across Indonesia's diverse geography and lifestyles, and to generate the Indonesian Microbiome Ecology and Evolution (IndoMEE) metagenome-assembled genome (MAG) reference database. The dataset consists of 250 bp paired-end metagenomic reads from 124 faecal samples (including eight technical replicates), acquired from 116 individuals, at ca. 10.6 Gb / sample, together with 2 positive controls and 4 negative controls. Library preparation was carried out using Illumina's TruSeq DNA Nano DNA kit (Illumina, Catalogue No. 20015965), and sequencing was performed on the Illumina NovaSeq 6000 platform.

Samples were voluntarily donated by Study Participants from various communities across Indonesia. Collection was performed cross-sectionally at different time points, from the following sites: Denpasar in 2015, Respen Tubu and Long Sule in 2018, Teluk Sumbang, Punan Batu, and Asmat in 2019, Rian Tubu in 2020, and Pedawa in 2021. Only adults (18 years and older) were included in this study.

The project received ethical approval from several institutions. The Eijkman Institute for Molecular Biology Research Ethics Commission approved the project under numbers 80/2014, 119/2019, and 126/2019. Approval was also granted by the Udayana University Faculty of Medicine and Sanglah Hospital Ethics Commission under numbers 1286/UN.14.2/Litbang/2014 and 1742/UN14.2.2.VII.14/LT/2021. Additionally, the Mochtar Riady Institute for Nanotechnology Ethics Committee approved the study under number 006/MRIN-EC/ECL/III/2022, and the University of Cambridge Human Biology Ethics Committee approved it under HBREC.2021.01.

The Data Producer is a consortium consisting of Safarina G. Malik, Pradiptadjati Kusuma, Herawati Sudoyo, Clarissa A. Febinia, and Lidwina Priliani from the Mochtar Riady Institute for Nanotechnology (MRIN), Tangerang, Indonesia; along with Guy S. Jacobs and Hirzi Luqman from the Department of Archaeology, University of Cambridge, United Kingdom.

The management and stewardship of this dataset are overseen by the Data Access Committee, which includes Safarina G. Malik and Pradiptadjati Kusuma from MRIN, and Guy S. Jacobs from the University of Cambridge. Any one of these individuals may act as a representative of the Data Access Committee for the purposes of this Agreement.

## APPENDIX 2 – USER’S RESEARCH PROJECT DETAILS

### 1. Details of dataset requested i.e., EGA Study and Dataset Accession Number

--

### 2. Brief abstract of the Project in which the Data will be used (500 words max)

--

### 3. List of Individuals in the User Institution to be named as registered users

Name of Registered User	Email	Job Title	Supervisor

### 4. List of Individuals that should have an account created at the EGA

Name of Registered User	Email	Job Title