

**The Institute of Environmental Medicine at the Helmholtz Center Munich, in collaboration with the Biomedical Statistics and Data Science group at the Computational Health Center, Helmholtz Center Munich, is offering a B.Sc./M.Sc. thesis project on the topic:**

### **Cross-contamination and chimera formation in microbiome data**

We are studying the interaction of humans and their environment in the context of health and disease, with a strong focus on the human microbiome and its bioinformatic and statistical analysis.

**Project background:** Changes in the human microbiome have been associated with a range of disorders, such as obesity, autism and atopic dermatitis. However, using microbiome data in a clinical context remains problematic due to various errors and biases present in microbiome sequencing data. Some of these errors are particularly under-investigated, and solutions to specifically address them are lacking. Two of such under-addressed problems are chimeras (incorrectly joined sequences) and cross-contamination (mixed up sequences between samples). Both problems lead to the distortion of true microbiome sample compositions and blur real biological signals.

**Objective:** We would like to explore two independent datasets where we found a phenomenon of peculiarly joined sequences, and to investigate the cause and potential applications of this phenomenon.

The M.Sc. thesis should cover the following aspects:

- Analysis of the phenomenon and whether it is related to cross-contamination or chimera formation
- Further development of the idea into an approach to measure and correct for the specific bias it is related to
- Additional aspects could be: review and testing of existing approaches to identify and remove cross-contaminating and/or chimeric sequences, validation of the phenomenon in additional mock or real microbiome samples

A desirable outcome would be an explanation of the phenomenon. In the best case, the thesis proposes an experimental setup to validate and actively induce the phenomenon in mock or real microbiome data.

**We are looking for** a motivated student with a strong interest technical and bioinformatic microbiome sequencing workflows, who is keen on critically and creatively exploring a given challenge. Experience with (or strong interest in learning) a programming language is required. We are providing an open and supportive working atmosphere with scientists of diverse backgrounds.

**If you are interested**, please contact [luise.rauer@tum.de](mailto:luise.rauer@tum.de).