Overcoming Data Science Hurdles in Complex Analyses of Symbiotic Reef-Building Corals

<u>Background:</u> Climate change is rapidly altering the structure and function of important ecosystems and outpacing the natural adaptive capacity of organisms required to survive. In an age where collecting large -omic data sets is more accessible than ever, data collection has outpaced the development of analytical techniques to analyze this data and understand how organisms respond to stress. An excellent case for addressing these challenges is coral reefs, which are built by coral animals that depend on multi-partner symbiotic relationships to build the ecosystem foundation. The field of coral biology utilizes multiple -omic approaches, but sound data science techniques to conduct multi-omic analyses that consider complex data structures are severely lacking. This proposal facilitates a new collaboration between UW eScience Data Science postdoctoral fellows Huffmyer (marine molecular biologist) and Kessler (data scientist and statistician), to address this challenge.

<u>Objectives</u>: This proposal seeks funding to support the cost of an interdisciplinary collaboration between Huffmyer and Kessler, two UW Data Science postdoctoral fellows.
- Aim (1): Conduct analysis of existing highly dimensional, multi-omic and time series coral reef biology datasets
- Aim (2): Produce a perspectives publication piece to showcase the utilization of advanced statistical techniques for biological interpretation in a non-model system

The proposed expenses in this application will support the cost of a collaborative visit from Dr. Peter W. MacDonald from McGill University to accomplish analytical goals and expand the knowledge of the postdoctoral fellows (Aim 1), and the cost of a publication in which Huffmyer and Kessler will discuss data science approaches for complex biological systems (Aim 2).

<u>Approach</u>: We will develop novel data science approaches to investigate how corals respond to environmental variation using two existing datasets collected by Huffmyer. The first data set is a physiological time series of three coral species at multiple reef environments on the island of Moorea, French Polynesia, collected in 2020. In this project, physiological metrics and genetic identities of hosts and symbionts were measured in individual coral colonies across one year, generating a data set comprising large matrices that are inherently nested within genetics, time, and location. Together, Kessler & Huffmyer will develop a matrix-variate mixed effects framework in order to exploit rich structure and properly account for dependence among measurements.The second data set is a developmental time series led by Huffmyer in a reef-building coral in Hawaii, USA. This project includes measurements of metabolic rates, physiological characteristics, host gene expression, host metabolomic response, and symbiont and bacterial community composition during multiple early life stages. Our goal is to conduct an integrative analysis that identifies strongly correlated responses across biological levels to provide a robust understanding of developmental changes in coral function. This project involves matrix-variate data that is expected to evolve smoothly over time. Kessler & Huffmyer, along with collaborator MacDonald, will develop methods for functional matrix-variate data adapted from past work by MacDonald that addressed a mathematically-related problem for functional network data (arXiv:2210.07491).

How does the proposed expense enhance your postdoctoral research and/or career development? The proposed expenses support an interdisciplinary collaboration between Kessler, an expert in statistical analyses of complex data, and Huffmyer, an expert in ecophysiology, to conduct data science research together. Huffmyer will gain knowledge and practical experience applying complex statistical models to the field of coral biology. Kessler will gain experience translating analytical knowledge by applying data science approaches to a complex biological system. For both fellows, this collaborative relationship will form a foundation for future research and ongoing data science knowledge exchange, both in our present and future work. The proposed expenses will provide opportunities for both fellows to learn from Dr. MacDonald from McGill University in approaches that are relevant to our respective fields. Further, publication funding strengthens our publication records and demonstrates advancements in our knowledge and data science.

Explain how/why this expense will provide something that you would not otherwise be able to attend/purchase/do. Huffmyer and Kessler do not have access to funds to support this collaborative visit and publication of productions of this collaboration from existing funding sources. This funding will allow us to expand our network and add valuable expert perspectives to this work by facilitating a collaborative visit from Dr. Peter W. MacDonald from McGill University to consult with us and guide us in our data analysis approach. Furthermore, access to funds to support the cost of publication ($2625) provides a pathway to formalize our analyses and generate a product that will be of valuable use to the broader scientific community, especially those conducting similar biological and data science research.

Discuss how this work is related to eScience areas of interest and/or our overarching mission to advance data-intensive discovery across all fields? This work supports an interdisciplinary collaboration between Huffmyer and Kessler, two eScience Data Science Postdoctoral Fellows, which directly addresses a core goal of the postdoctoral program to facilitate collaboration between fellows. We will formalize the connections made through this program through a collaboration to advance implementation of data science principles in a biological field of study. This collaboration will advance both our biological knowledge and analytical capacity to use multi-omic and highly dimensional data sets to inform biological interpretations. This work directly complements the mission of the eScience Institute by using large, complex, and interdisciplinary datasets to understand foundational questions about an important ecosystem. In addition, we plan to arrange a seminar for the eScience audience to be given by collaborator MacDonald during his visit to Seattle, which will provide the broader eScience postdoctoral with an opportunity to learn from a data science expert.

Item	Justification	Projected Cost
Round trip travel	Airfare & ground transit for collaborator (Montreal <--> Seattle)	$600
Visit to UW by collaborator	7 nights lodging in Seattle for collaborator, including per diems	$1745.50
Publication	Estimated fee for open access publication in Functional Ecology	$2625 (APC)
	Total Requested Funding	**$4970.50**