Contribution ID: **7**   Type: **Talk**

Best of both worlds: unlocking the potential of data science for biological research

Thursday 4 April 2024 14:35 (15 minutes)

As the life sciences are becoming increasingly reliant on big data and an ever-growing number of computational strategies, leveraging a data science approach can substantially enhance the efficacy and robustness of biological data analyses. Yet, for these data manipulation and processing methods to reach their full potential for biological applications, it is critical for biologists and computer scientists to work together and combine biological knowledge of the research system with technical expertise to craft appropriate solutions. I will present two stories from the ʿomics field that demonstrate the power of data science when applied with biological insights to otherwise difficult problems. First, I will describe a simple technical heuristic based upon our knowledge of the extent of diversity in protein families to reduce the number of false negatives in homologous gene annotation, which in some cases enables the annotation of up to 16% more functions in a given microbial genome. Second, I will describe how adding a novel normalization step could explain the drivers of the loss of microbial diversity in individuals with inflammatory bowel disease (IBD), revealing a potential ecological explanation for something that puzzled microbial ecologists for almost two decades. In both cases, finding these solutions required a combination of technical and biological expertise, highlighting that the path forward for enhancing the scope and utility of computational work within the biological sciences will depend upon effective communication between scientists in these two fields.

Primary author: VESELI, Iva (Alfred Wegner Institute)

Presenter: VESELI, Iva (Alfred Wegner Institute)

Session Classification: Session 1: Capacity Building and AI

Track Classification: Capacity buidling and AI