New bioinformatic and statistical approaches to longitudinal metagenomics

Primary supervisor: Josie Bryant (WSI)
Secondary supervisor: John Lees (EBI)

PROJECT DESCRIPTION

Microbiomes play a crucial role in human health and are important mediators of disease. Metagenomics provides an unparalleled way of sampling the full complement of genetic diversity found in these microbial communities through capture of both intra and interspecies genomic diversity. However, the majority of microbiome studies rely on data drawn from single time points, so are unable to capture temporal changes in microbial communities. There is increasing evidence that the temporal fluxes in microbial species and their component subclones are important for disease progression in both the gut and the respiratory tract. In addition, a lack of longitudinal datasets has been cited as a key factor in our current inability to define causal relationships between microbiomes and health[1].

The respiratory microbiome is thought to be highly important for disease progression in cystic fibrosis, where bacterial densities are high and community dysbiosis is associated with advanced disease[2]. Despite intensive investigation however, a mechanistic understanding of this system is lacking and the bacterial drivers of significant clinical events such as acute pulmonary exacerbations are still unknown[3]. Previous studies which have been impeded by sparse longitudinal sampling, lack the temporal resolution to untangle the dynamics of the relevant bacterial communities. In an effort to address this unmet need, the Bryant lab has established large sample collections (n=20,000 daily metagenomic samples) with dense longitudinal sampling from 200 people with cystic fibrosis. By combining this data with extensive clinical measurements and daily physiological monitoring, we aim to uncover microbial triggers of observed clinical deterioration events.

However, longitudinal metagenomic analysis is a new area, and so presents unique statistical and bioinformatic challenges. Previously in the Bryant lab, we have developed methods which allow the inference of subclonal population structure from longitudinal deep sequencing data[4] of respiratory bacteria. But we don't yet have methods to jointly infer both intra-species and interspecies diversity. Furthermore, longitudinal datasets often suffer from missing data-points, non-independence of data, and sample-to-sample noise due to the imperfect nature of clinical sampling. The Lees lab has extensive experience in processing, modelling, and drawing biological insights from large and complex bacterial datasets. They have specialised in inference of population structure from genomic data and the development of new bioinformatic tools[5,6].

This ESPOD project provides an exciting opportunity for a postdoctoral fellow to develop new and creative approaches to existing datasets to uncover bacterial signals of clinical deterioration in chronic lung disease. This project would also provide an opportunity to cement a new collaboration between the Bryant and Lees labs.

Our hypothesis is that development of new statistical and bioinformatic approaches tailored to longitudinal data are required to uncover novel associations between the respiratory microbiome and the progression of chronic lung disease. **The aims of the ESPOD project are**:

(1) Leverage information gain across longitudinal samples to develop bioinformatic methods which jointly estimate both intra-species (subclonal) and inter-species diversity in patients over time.

(2) Using this data, create hierarchical Bayesian models to account for sample bias and variability to infer true biological temporal fluctuations.

(3) Integrate inferred fluctuations with clinical metadata to robustly identify statistical associations between microbial diversity and changes in clinical phenotype.

The project will primarily focus on data generated for cystic fibrosis patients, but it is anticipated that the methods will be widely applicable to other longitudinal metagenomic studies of human disease. High-coverage, unbiased metagenomic data is already available for ~300 samples and a further 5,000 will be in place by the start of the project.

OUTPUTS

This is a unique dataset with unprecedented sampling density, yielding significant opportunities for novel method development, and generating important insights into disease progression. We therefore anticipate that this project will lead to the identification of new biological insights and high impact publications.

Any significant or novel microbial signals identified during the project would need to be further characterized and validated through experimentation. We anticipate this to be beyond the scope of the project and could provide further avenues of research for the candidate. In addition, the methods developed in this project would be directly applicable to longitudinal studies of gut metagenomes so there are many ways that the work generated during this fellowship could be developed further by the applicant.

TRAINING AND MENTORSHIP

This project provides an ideal opportunity to develop interdisciplinary skills. The Bryant lab has extensive wet-lab and bioinformatics expertise in the respiratory microbiome with currently three postdocs working on distinct but complementary projects in this area. The Lees lab provides unique experience in bioinformatic analysis of complex microbial datasets, mathematical modelling of genome data, method and tool development. The ESPOD fellow would benefit from the expertise through integration into both labs and frequent meetings with both supervisors.

We envisage that the proposed ESPOD fellow would be primarily based in the Bryant lab at the WSI but will attend all of the Lees lab meetings and will be paired with existing members of the Lees lab to facilitate the transfer of skills. Monthly meetings will be held between the two supervisors and the fellow to discuss the project progress as a group.

The co-supervisors will endeavour to provide the necessary training and mentorship to develop the candidate to ensure they are successful in the next stages in their career. They will be encouraged to apply for the postdoctoral Lab Leadership and postdoctoral Accelerator Award schemes at the WSI, and the Transition to Independence grants at EMBL.

1. Schmidt, T. S. B., Raes, J. & Bork, P. The Human Gut Microbiome: From Association to Modulation. *Cell* **172**, 1198–1215 (2018).
2. Cuthbertson, L. *et al.* Lung function and microbiota diversity in cystic fibrosis. *Microbiome* **8**, 45 (2020).
3. Caverly, L. J. & LiPuma, J. J. Cystic fibrosis respiratory microbiota: unraveling complexity to inform clinical practice. *Expert Rev Respir Med* **12**, 857–865 (2018).
4. Bryant, J. M. *et al.* Stepwise pathogenic evolution of Mycobacterium abscessus. *Science* **372**, eabb8699 (2021).
5. Lees, J. A., Harris, S. R., Tonkin-Hill, G., Gladstone, R. A., Lo, S. W., Weiser, J. N., Corander, J., Bentley, S. D., & Croucher, N. J. (2019). Fast and flexible bacterial genomic epidemiology with PopPUNK. Genome Research, 29(2), 304–316.
6. Horsfield, S. T., Tonkin-Hill, G., Croucher, N. J., & Lees, J. A. (2023). Accurate and fast graph-based pangenome annotation and clustering with ggCaller. Genome Research, 33(9), 1622–1637.