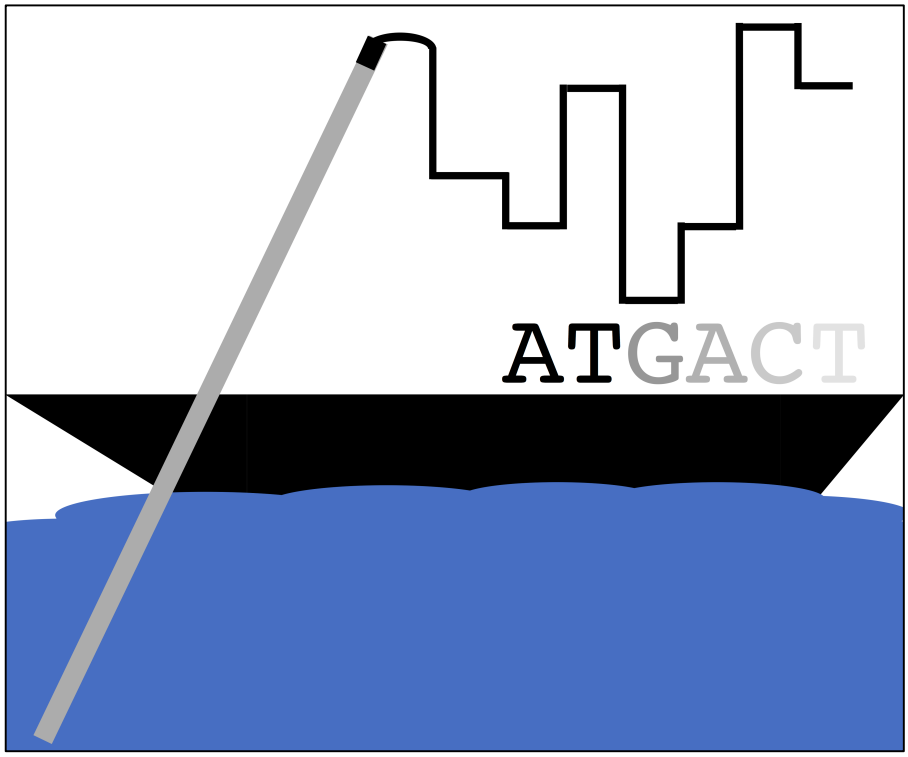
Application for an OpenPlant Grant 2017

PuntSeq

Chasing the invisible diversity of microbial life forms in freshwater with a portable DNA-sequencer



I. Team

Maximilian Stammnitz, mrs72@cam.ac.uk (main contact)
Cancer Genomics PhD candidate @ Transmissible Cancer Group, Department of Veterinary Medicine, Madingley Rd, University of Cambridge,
Experiences: DNA Sequencing, Data Mining, Fieldwork, Labwork

Philipp Braeuninger-Weimer, pab96@cam.ac.uk
Electrical Engineering Postdoc @ Centre of Advanced Photonics and Electronics, 9 JJ Thomson Ave, University of Cambridge,
Experiences: Microfluidics, Material Characterisation, Mechanical Design Engineering

Daniel Elías Martin-Herranz, dem44@cam.ac.uk
Bioinformatics PhD candidate @ Thornton Group, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, University of Cambridge,
Experiences: DNA and Methylome Sequencing, Ageing Research, Labwork

Lara Urban, lhu20@cam.ac.uk
Bioinformatics PhD candidate @ Stegle Group, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, University of Cambridge,
Experiences: Epigenomics, Cancer Research, Deep Learning

Surangi Perera, snp23@cam.ac.uk
Developmental Biology PhD candidate @ Baker Group, Department of Physiology, Development and Neuroscience, University of Cambridge,
Experiences: DNA and RNA work, Developmental Research, Neurobiology

Eirini Vamva, ev306@cam.ac.uk
Virology PhD candidate @ Lever Group, Department of Medicine, University of Cambridge,
Experiences: DNA and RNA work, Viral Genomics, HIV Research

Astrid Wendler, aw642@cam.ac.uk
Biology PhD candidate @ Tavaré Group, Cancer Research UK, Department of Clinical Neuroscience, University of Cambridge,
Experiences: DNA and RNA work, Neurobiology, Cancer Research

Daniel Kunz, djk44@cam.ac.uk
Mathematical Genomics PhD candidate @ Teichmann Group, Wellcome Trust Sanger Institute, Hinxton, University of Cambridge,
Experiences: Single Cell Sequencing, Physics, Math. Modelling, Computing

Christian Schwall, cs687@cam.ac.uk
Biochemistry and Plant Sciences PhD candidate @ Locke Group, Sainsbury Laboratory, Bateman St, University of Cambridge,
Experiences: Microbiology, Sampling and Measurement, Synthetic Biology

Meltem Gürel, mg773@cam.ac.uk
Computational Biology PhD candidate @ Tavaré Group, Cancer Research UK Cambridge Institute, Robinson Way, University of Cambridge,
Experiences: Nanopore DNA Sequencing, Data Mining, Math. Modelling

II. Summary

In this proposal we present *PuntSeq*, a toolbox and workflow to facilitate real-time monitoring of bacterial and viral diversity in aquatic field work situations. Our team combines knowledge and know-how from very different study backgrounds with the curiosity to learn about and share experiences on cutting-edge portable DNA sequencing technology. We will design and openly distribute a template for production of a suitable mini-lab for the Oxford nanopore MinION sequencing device and its complementary equipment, in addition to writing and adapting software for processing large volumes of DNA sequencing data. Combined, this will serve in the study and categorisation of aquatic microbial life communities at different sites of our city's river, the Cam. The core of this project is to make the complex field of DNA sequencing accessible to non-life scientists, by enabling a simple hands-on experience. We will hold regular *PuntSeq* workshops for interested individuals and organisations to teach them about state of the art DNA sequencing, its vast applications and benefits in a broader picture.

III. Introduction

Nearly any accessible spot on earth is colonised with microbial life. From the antarctic ice shield to deep sea hot springs, from Mount Everest to the human gut and to the root tips of soy beans: a plethora of invisible, highly adaptive bacteria persist and evolve in almost any described ecological niche (Torsvik et al., 2002). Single-cell prokaryotes often co-locate in complex symbiotic arrangements of increased taxonomical diversity. Moreover, bacteria and archaea share comparatively much smaller genome sizes than most higher multicellular organims, thereby rendering themselves as cost-effective to

sequence and compare. By studying the origins of bacterial DNA in field samples, we can therefore use 'metagenomics' tools to aim at reconstructing substrates' (i) microbial diversity, (ii) localised proportions of taxa, and (iii) functional niche and environmental condition (Fierer and Jackson, 2006; Juul et al., 2015; Venter et al., 2004).

Over the past two decades, DNA sequencing costs have been dropping at a record rate, from the first 3,000,000,000 $ human genome draft to a <1000 $ high-quality sequence in 2017. Technological milestones in next-generation DNA sequencing have also preceeded substantial improvements in organisational logistics, processing speed and automated analysis (http://www.nature.com/news/technology-the-1-000-genome-1.14901).

Projects such as the EMBL Tara Oceans expeditions (https://www.embl.de/tara-oceans/) underline the scale of opportunity that has meanwhile arisen for discoveries of aquatic bacterial forms of life on earth. Over the past two years, sequencing has become even more comfortable and affordable for daily genomics: smaller than hand-sized, portable devices now enable sampling and real-time ultra long read sequencing in the same location outside of a big laboratory (Jain et al., 2016). First remarkable applicational successes of the market's pioneer instrument, Oxford Nanopore's MinION, involved the live-tracking of viral evolution during the recent West-African EBOLA outbreak (Quick et al., 2016), *in-situ* identification experiments of microbial pathogens in human urine and blood (Greninger et al., 2015; Schmidt et al., 2017), field-work profiling of metagenomics from an arctic glacier and the bacterial decomposition of a two-weeks standing glass of raw cow milk (Edwards et al., 2016; Juul et al., 2015). We wish to bring this technology to the broader community.

IV. Aims and Methods

Our interdisciplinary team of Cambridge PhD students is composed of several passionate outdoor activists and sportsmen. For the past three years, we have been regularly swimming, rowing, and punting together on our river Cam. During this time, we have witnessed several river-related illnesses of

our fellow students, primarily due to bacterial infections assigned to water contact of open wounds or leaking drinking bottles. In a few cases known to us, these diseases have been life threatening: septic infections that were only addressable by several courses of antibiotics, local surgery and weeks-lasting stationary treatment. Hence, our simple question: 'what are these (terrifying) bugs?' Repeated enquiries with Public Health England and the Environment Agency have yielded no conclusive picture regarding the abundance and diversity of microbial pathogens in the Cam water and, to our best knowledge, no public monitoring scheme of such biological entities in Cambridge's local freshwater sources has been put in place (Philipp Braeuninger-Weimer, personal communication).

Here, we aim at exploring the microbial search space of our river Cam through a cheap and transparent metagenomics approach. By designing a punt-suited mini lab (P. Braeuninger-Weimer, D. Kunz, C. Schwall) which at its heart contains a portable MinION DNA sequencer wired to a small laptop or tablet, developing and integrating standalone bioinformatic analysis protocols for DNA basecalling, alignment, and phylogenetic analysis, (L. Urban, D. E. Martin-Herranz, D. Kunz, M. Gürel), taking water samples from a number of sites of the river, and isolating heavy-weight bacterial DNA with different methods (S. Perera, A. Wendler, E. Vamva, L. Urban, M. Stammnitz), we count on a broad range of skills and strengths that our members contribute from several different fields of research. Although some members of our team have already been involved with some aspects of DNA sequencing studies, none of us can claim a broader knowledge of all aspects: (i) aquatic fieldwork sampling, (ii) DNA-extraction, (iii) MinION usage, (iv) MinION raw data processing, (v) MinION data interpretation. Together, we intend to perform *PuntSeq*, a real world metagenomics pipeline for field work in the context of freshwater analysis.

Besides the conceptualisation of and tinkering on a reproducible workflow that involves a state-of-the-art high tech instrument, we intend to draw an accurate picture of the microbial life communities in a river setting. Our proof-of-concept experiment involves a comparative phylogenetic analysis of the fluvial

surface water microbiome by repeated sampling in multiple spots along the Cam's course.

With this project, our team wishes to bring the ease and usage of DNA sequencing closer to our Cambridge community of scientists and naturalists. Despite tremendous improvements in quality, nanopore sequencing devices are yet in their beta stage. Thus, they hold great educational and developmental potential in the right hands of technology enthusiasts, synthetic biology practicioners, curious students and teachers alike.

V. Benefits and outcomes

DNA sequencing and modification are controversially discussed in our society, however few people have been personally involved or exposed to such experimental research settings without year-long training. Yet, research in the fields of Genomics and Synthetic Biology is progressing fast.

Until recently, DNA sequencing was only possible with expensive equipment, operated by specialists. The commercialisation of the Oxford nanopore MinION is about to change this paradigm: sequencing will become available to non-specialists without a dedicated laboratory facility, for example enabling doctors to do rapid diagnostics on their patients. We think that the public, particularly in low-resource settings, should therefore gain access to learning opportunities for understanding what DNA sequencing is, how cheap, portable devices can be used at home, and what the beneficial applications and implications of its rapid advancements will be in the near future.

Besides our ambition of determining the extent and variety of pathogenic prokaryotes in our river Cam and learning about the technical challenges with a MinION-instrument, the main outcome of this project will be an open source outreach program that provides hands-on experience in DNA sequencing and analysis for non-scientists. The open source nature of our *PuntSeq* workflow, along with inexpensive components which are further expected to decrease in

price and accessibility over time, could even be used in schools to educate young students about cutting edge technology transitions and to also teach ourselves about the invisible biodiversity within and around us. Last, our team is very excited and passionate about *PuntSeq* and its potential impact on biological water monitoring schemes in a wider geographical context, we are curious to pioneer ahead, learn about and from a potentially revolutionary technology; all of us are very much looking forward to this project!

VI. Project support and cost-code centre

Dr. Elizabeth Murchison
Principal Investigator,
Transmissible Cancer Group,
Department of Veterinary Medicine,
University of Cambridge,
epm27@cam.ac.uk

I confirm that I have the full support of the sponsor listed above and that they can be added to the OpenPlant Fund mailing list to receive project updates (to which they can unsubscribe at any time).

VII. Budget

We budget our project as follows:

- 800£ Oxford Nanopore MinION Starter Kit
- 1200£ Additional Nanopore Instrument flow cells and chemicals
- 1000£ DNA Isolation Kits
- 500£ Material Costs

Additional support for the project will be provided as follows:

- In case of a successful application, Maximilian Stammnitz would apply for additional funding from the Gates Cambridge Trust to attend a 1-week nanopore instrument handling workshop

- Dr. Elizabeth Murchison agrees to provide a laboratory at the Department of Veterinary Medicine available for DNA-isolation steps and preliminary technical tests of the nanopore sequencer

VIII. References

Edwards, A., Debbonaire, A.R., Sattler, B., Mur, L.A.J., and Hodson, A.J. (2016). Extreme metagenomics using nanopore DNA sequencing: a field report from Svalbard, 78 °N. bioRxiv.

Fierer, N., and Jackson, R.B. (2006). The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A *103*, 626-631.

Greninger, A.L., Naccache, S.N., Federman, S., Yu, G., Mbala, P., Bres, V., Stryke, D., Bouquet, J., Somasekar, S., Linnen, J.M.*, et al.* (2015). Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med *7*, 99.

Jain, M., Olsen, H.E., Paten, B., and Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol *17*, 239.

Juul, S., Izquierdo, F., Hurst, A., Dai, X., Wright, A., Kulesha, E., Pettett, R., and Turner, D.J. (2015). What's in my pot? Real-time species identification on the MinION. bioRxiv.

Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., Bore, J.A., Koundouno, R., Dudas, G., Mikhail, A.*, et al.* (2016). Real-time, portable genome sequencing for Ebola surveillance. Nature *530*, 228-232.

Schmidt, K., Mwaigwisya, S., Crossman, L.C., Doumith, M., Munroe, D., Pires, C., Khan, A.M., Woodford, N., Saunders, N.J., Wain, J.*, et al.* (2017). Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. J Antimicrob Chemother *72*, 104-114.

Torsvik, V., Ovreas, L., and Thingstad, T.F. (2002). Prokaryotic Diversity - Magnitude, Dynamics, and Controlling Factors. Science *296*, 1064-1066.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W.*, et al.* (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. Science *304*, 66-74.