Biospectra-by-sequencing genetic analysis platform

Patrick Biggs[1], Rudiger Brauning[2], Mingshu Cao[2], Charles David[3], Marcus Davy[3], Helge Dzierzon[3], Rob Elshire[2], Ruy Jauregui[2], Aurelie Laugraud[2*], John McCallum[3], Alan McCulloch[2], Roger Moraga[2], Louis Ranjard[1], Roy Storey[3], Shane Sturrock[1]

[1] NZGL, New Zealand.

[2] AgResearch, New Zealand.

[3] Plant and Food, New Zealand.

*Aurelie.laugraud@agresearch.co.nz

All authors contributed equally to the work and are listed in alphabetical order.

Main Text of Abstract

The Biospectra-by-sequencing (BBS) project started in 2014. It is a collaborative multi-institutional effort to build a robust genetic analysis platform. It initially focused on open-source Genotyping-by-sequencing (GBS) technology. GBS reduces genome complexity by using restriction enzymes, making genotyping-by-sequencing cost effective. BBS goes beyond genotyping and extends the application of the reduced complexity sequencing to many other areas. Most importantly, our project aims to encourage a diverse range of end-users to adopt this technology.

Advances in next generation sequencing have consistently increased value for money thereby opening many new opportunities. However, the lack of appropriate analysis pipelines prevents scientists from taking full advantage of these. To realize this potential, we are developing a set of analytical platforms built around GBS. The first aim is to use proof-of-concept work to estimate their applicability across a wide range of biological systems. The second is to promote end-users to adopt these platforms through structured engagement.

The project brings together developers and users from different institutions in New Zealand. We work collaboratively using GitHub as our primary common platform. The core of the analysis engine uses the readily available TASSEL suite, developed at Cornell University. TASSEL takes sequence reads, maps them against a reference genome and then calls variants. We want to incorporate TASSEL in a more automated pipeline with emphasis on reproducible research. The frameworks we use ensure that the same analysis can be redone accurately and/or passed on, thereby facilitating better quality science. We also add pre- and post-processing components. The main focus of pre-processing is quality control.The post-processing part involves filtering after the variant calling. The multidisciplinary background of people in our group provides expertise in most BBS applications and ensures that our software can be used to study all organisms in a large range of experiments.