Collectively Improving Robustness and Reproducibility in Genotyping-by-Sequencing (GBS) Experiments

Marcus Davy[1], Cecilia H. Deng[2], Helge Dzierzon[3], Robert J Elshire[4], Stephanie Galla[5], Elena Hilario[2], Robyn Johnston[4], Tammy Steeves[5], Roy Storey[1], and Susan Thomson[6]

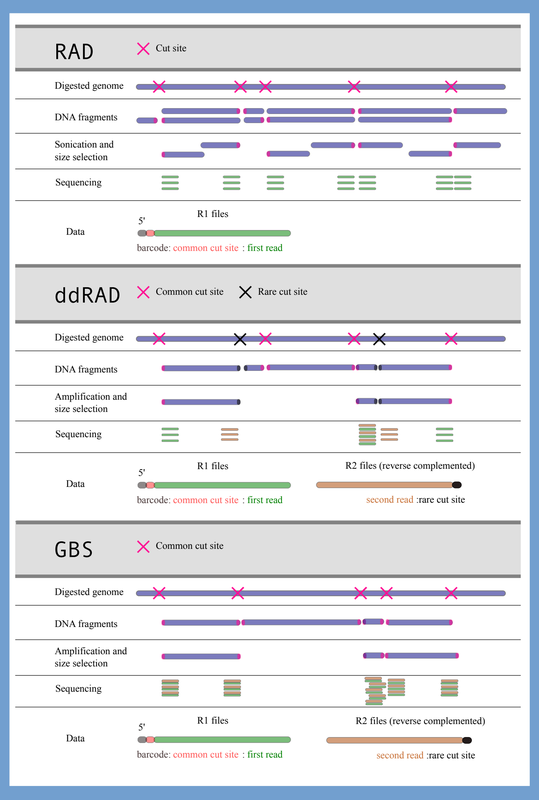
[1]The New Zealand Institute for Plant & Food Research Limited, Te Puke, New Zealand, [2]The New Zealand Institute for Plant & Food Research Limited, Auckland, New Zealand, [3]The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand, [4]The Elshire Group Limited, Palmerston North, New Zealand, [5]University of Canterbury, Christchurch, New Zealand, [6]The New Zealand Institute for Plant & Food Research Limited, Lincoln, New Zealand

Overview

Many researchers have used genotyping-by-sequencing to generate marker data since the method was published in PLoS ONE in 2011. Others have made modifications to the method resulting in different, but related, types of GBS data generated. Additional analysis pipelines have been developed, many of which are licensed under Free / Libre and Open Source licenses -- allowing them to be inspected, tested and improved upon. Taken together these modifications and analysis pipelines demonstrate the power of the scientific method when combined with modern genomics laboratory techniques, open access publishing and open source software to rapidly and democratically advance our ability to conduct research. It has also left researchers wanting to use the technology with several questions which need to be answered before they adopt the approach. First, what laboratory method should one use to make GBS libraries? Second, what sequencing platform should be used and how? Third, what pipeline is best suited for the data generated and the genetics of the system being studied?

A group of like-minded scientists has come together to build software tools and an information repository to help others answer these questions for their own experiments in a project called Biospectra-by-Sequencing. The aim of the project is developing an informational Wiki for geneticists and genomicists as well as a containerized software test rig to evaluate and run many GBS pipelines.

Wet Lab

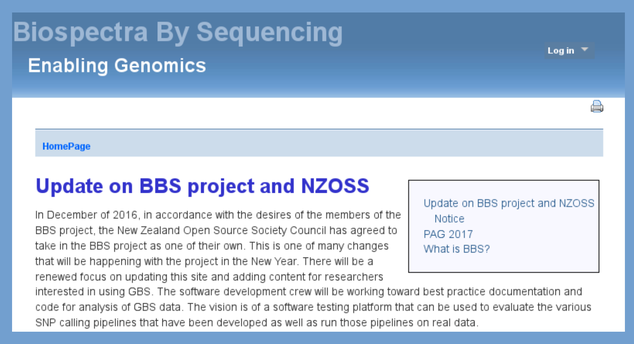


Data Types image courtesy of Dean Eaton (CC-BY)

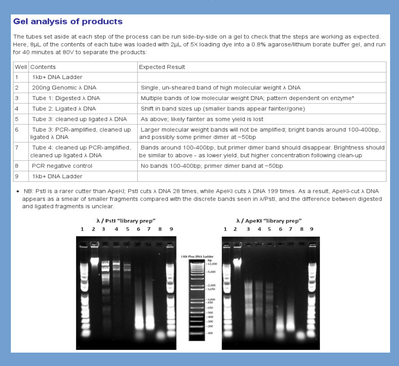
Representative Methods

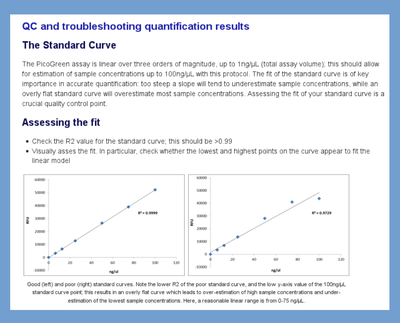
Author	DOI	Year	Enzyme Number	Size Selection	Selective Bases	Enzyme Type	DNA source
Elshire et. al.	10.1371/journal.pone.0019379	2011	1	no	no	II	native
Poland et. al.	10.1371/journal.pone.0032253	2012	2	no	no	II	native
Sonah et. al.	10.1371/journal.pone.0054603	2013	1	no	yes	II	native
Peterson et. al	10.3390/d6040665	2014	2	yes	no	II	native
Pan et. al.	10.1111/1755-0998.12342	2015	4	yes	no	II	native
Hilario et al.	10.1371/journal.pone.0143193	2015	1	no	no	II	amplified

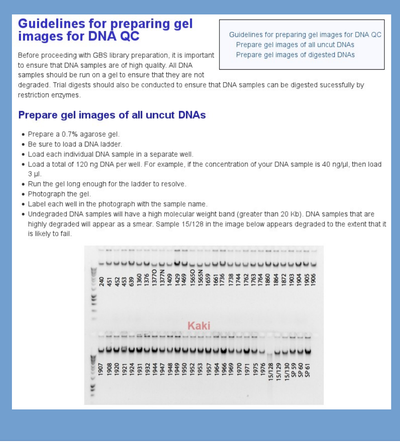
BBS Wiki – http://biospectrabysequencing.org

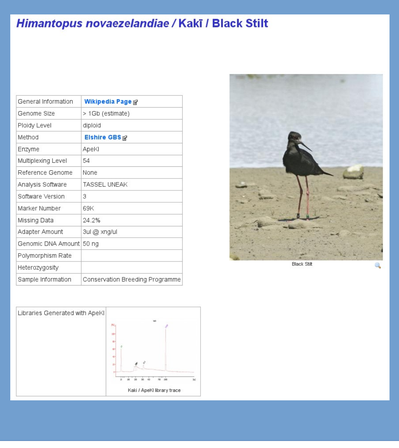


Wiki Resources



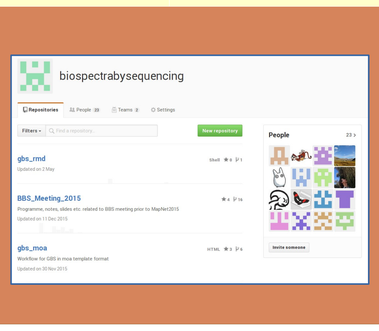






Dry Lab

Pipeline Name`	DOI	Year Published	Denovo / Reference	Demultiplexer	Trimmer / Read QC Filter	Aligner	SNP Caller	Hardcoded enzyme	Output file types
TASSEL UNEAK	10.1371/journal.pgen.1003215	2013	Denovo	Built In	None	NA	Built In	Yes	hmp
STACKS	10.1111/mec.12354	2013	Denovo / RefereZnce	Built In	Built In	BWA / Bowtie2 / GSnap	Built In	Yes	Database with export
IGST-GBS	10.1371/journal.pone.0054603	2013	Reference	FastX toolkit	FastX toolkit	BWA	sam tools	Unknown	vcf
TASSLE 3	10.1371/journal.pone.0090346	2014	Reference	Built In	None	BWA / Bowtie 2	Built In	Yes	hmp
pyRAD	10.1093/bioinformatics/btu121	2014	Denovo	Built In	Built in	Muscle	Built In	No	Text (.loci, .phy, .nex, .snps, .vcf and others),
AftrRad	10.1111/1755-0998.12378	2015	Denovo	Built In	Built in	Mafft	Built In	No	Text (translation scripts)
GBS-SNP-CROP	10.1186/s12859-016-0879-y	2016	Denovo / Reference	Built In	Trimmomatic	BWA	sam tools	No	SNP matrix, TASSEL hmp, PLINK tped
GibPSs	10.1111/1755-0998.12510	2016	Denovo	Built In	Built in	Matching Built in	Built In	No	Database with export.
NGSEP	10.1186/s12864-016-2827-7	2016	Reference	Built In	Built in	Bowtie2	Built In	Unknown	vcf
FastGBS	10.1186/s12859-016-1431-9	2017	Reference	Sabre	Cutadapt	BWA	Platypus	Not used	vcf
ipyRAD	NA	NA	Denovo, Reference, Denovo+Reference	Built In	Built in	BWA / SMALT	sam tools / bed tools	No	Vcf + many others
TASSEL 5	NA	NA	Reference	Built In	Built In	BWA / Bowtie2	Built In	Yes	Database with export



Collaboration

The Biospectra-by-Sequencing project welcomes participation by like minded scientist from around the world regardless of skill level. One way to participate is to use the resources we are developing and provide feedback so they can be improved. If you have conducted GBS experiments you can add information about those experiments to the wiki (http://biospectrabysequencing.org). By doing that, you will be helping create a resource for other scientists and yourself. For those with coding skills, we encourage you to assist with the pipeline testing rig. There is plenty to do there with the twelve pipelines in the table above and more likely to come along. Find us on github (https://github.com/biospectrabysequencing). For more information or general questions about BBS, email Rob at Rob@ElshireGroup.co.nz.

NZOSS

Plant & Food RESEARCH RANGAHAU AHUMARA KAI

The Elshire Group Ltd.

UC UNIVERSITY OF CANTERBURY Te Whare Wānanga o Waitaha CHRISTCHURCH NEW ZEALAND