

The New Zealand Institute for Plant & Food Research Limited



Cecilia Deng, David Chagné, Elena Hilario, Chris Kirk

Overview



- » Pipfruit in New Zealand
- » What is GBS?
- » Pipfruit GBS workflow @PFR
- » How we manage the workflow & best practice tips
- » Conclusions



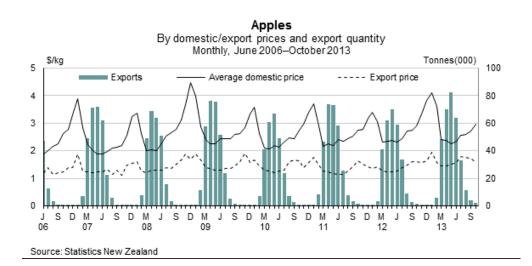
Pipfruit in New Zealand



» Apple is the 2nd largest fresh fruit export from NZ

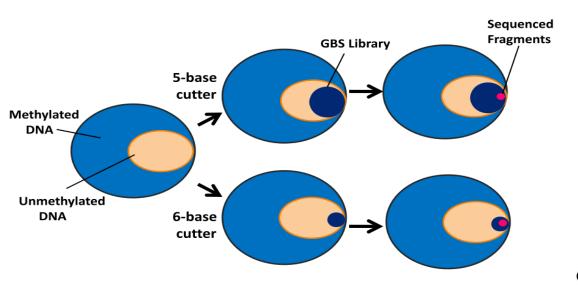






What is GBS?







Spruce





Maize Sorghum Rice Barley Switchgrass Bracypodium **Pearl Millet** Teosinte Andropogon Fonio Finger Millet

Reed Canary Grass Tripsacum Sugar cane Rye grass Iris Banana

Oil Palm



Strawberry Silene Sunflower Safflower Soybean Goldenberry Jatropha Willow Turnip rape Canola Plum Monkey Flower Onion

Peanut Cowpea Squash Tomato Clover Corn Salad Almond



Grape Cassava Cacao Watermelon Apple Hop **Eucalyptus** Cashew Auracaria **Guinea Yam** Horseradish Cotton

Pepper

Cucumber

Snap pea

Gourd

Arabidopsis

Willow

Potato

Raspberry

Peach

Knautia

Tomatillo

Corn Salad

Alder

Apricot

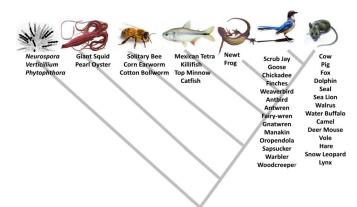
Bean

Pigeon Pea

Chickpea Lentil



Flowering Plants



GBS Method Overview, Sharon E. Mitchell **Cornell University**



Pipfruit GBS Workflow @PFR





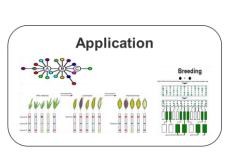
Plant tracking & phenotyping



Sample collection







gDNA extraction

GBS lib preparation •



Sequencing

Key file building

Data analysing





Assess & Evolve

Result Sharing



Sequencing Data: /input



1304KHS-0092/CKVAopDA4_2.fastq.gz

racking of Metabata 1506KHS-0054_44_283_C738AACXX_Lane4_[1

D1B7JACXX_SD1_29_3_GCCAAT_L002

1312AHS-0044/H9_[1/2].fastg_c

3_1.fastq.gz

4_2.fastq.gz

Effective x_L004_R1.fastq.gz









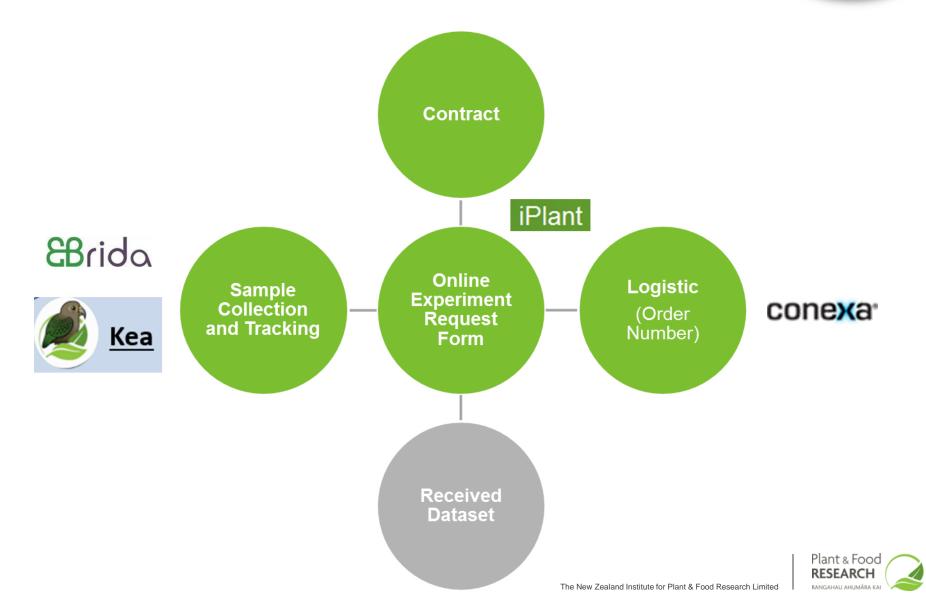






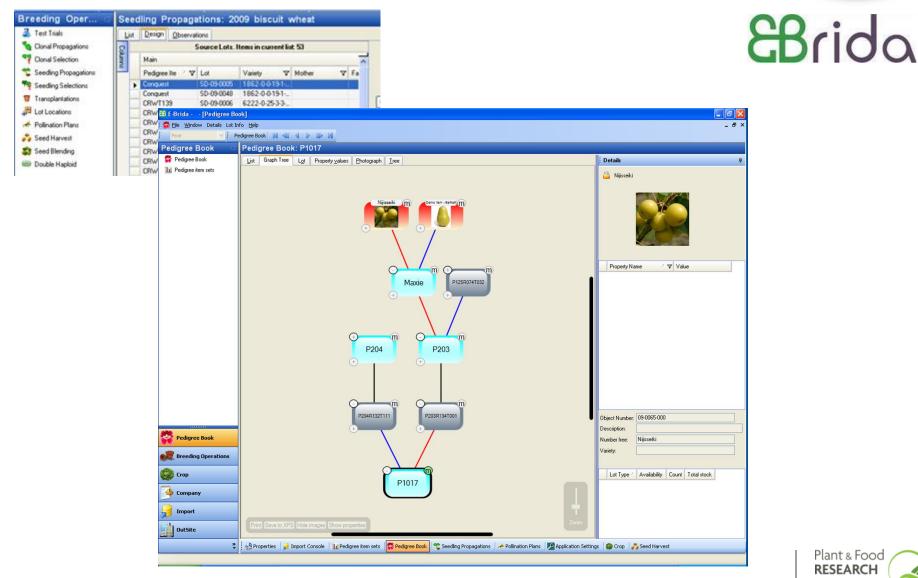
Best Practice: Link Science, Management and Data





Best Practice: Track Plants & Phenotypes

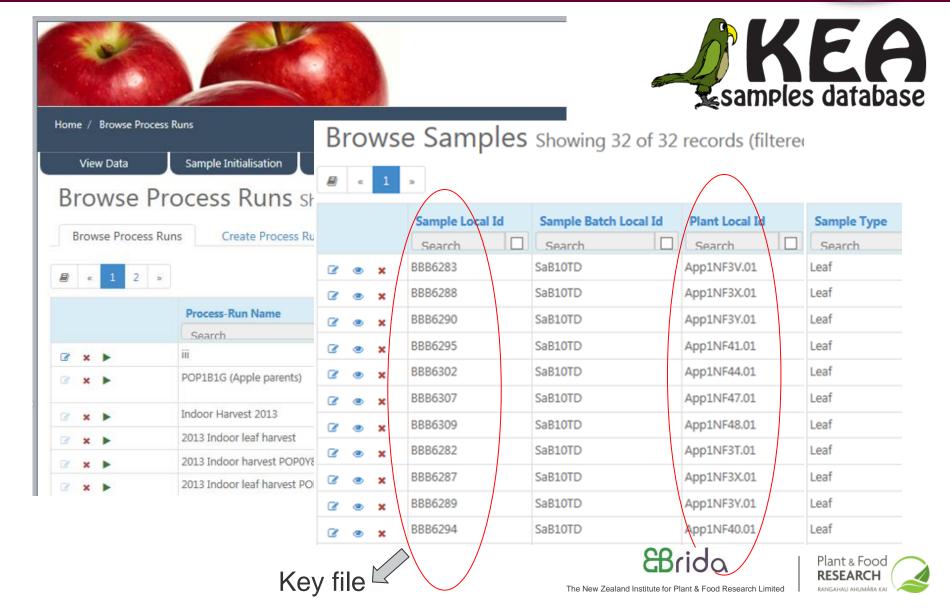






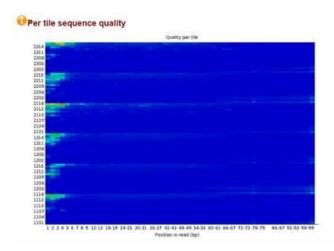
Best Practice: Track Sample Collection and Processing

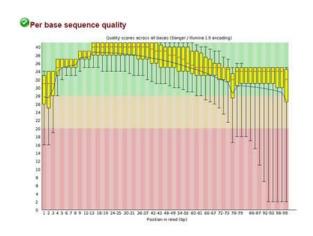


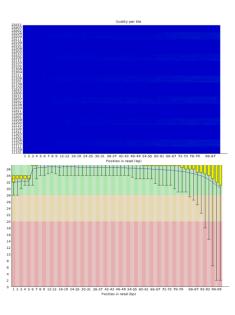


Best Practice: Data QC









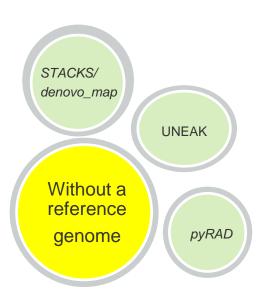
Cutoff score for barcode bases	Q10	Q20	Q30
Malus2015 A P10	42.15%	34.14%	25.43%
Malus2014 A L2	79.22%	58.61%	33.93%
Malus2014 B L1	90.96%	86.35%	78.65%
A2015 B L3*	91.46%	77.73%	70.91%



Best Practice: Choose a Variant Calling Tool









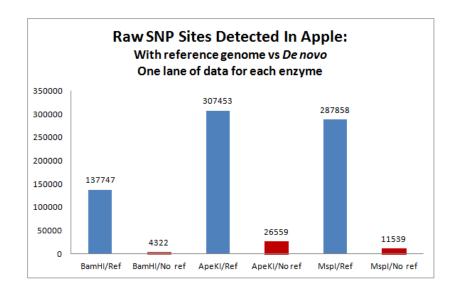
Best Practice: Get a Reference Genome





Enzyme	BamHI
Samples	13 pear accessions
	P. communis assembly v1,
Ref genome	142,083 scaffolds
Raw SNPs detected	
(using ref genome)	205,730
De novo variants	
detection	1176
%	0.57%







Best Practice: Track Data Analysis Workflow





PlantandFoodResearch / analysis-workflows PRIVATE

Branch: master ▼

analysis-workflows / Malus / GBS / Malus_GBS_pilot /

Analysis Malus GBS with TASSEL

- 06_bowtie.sh
- 07_samConverter.sh
- 08_FqToTBT.sh
- 09_mergeTagByTaxa.sh
- 10_tagToSnp.sh
- 11_mergeSNPs.sh
- 12_filter.sh
- 13_check_missing_data_by_sample.sh
- 14_compare_SNP_sites.sql
- 15_compress_raw_SNPs.sh

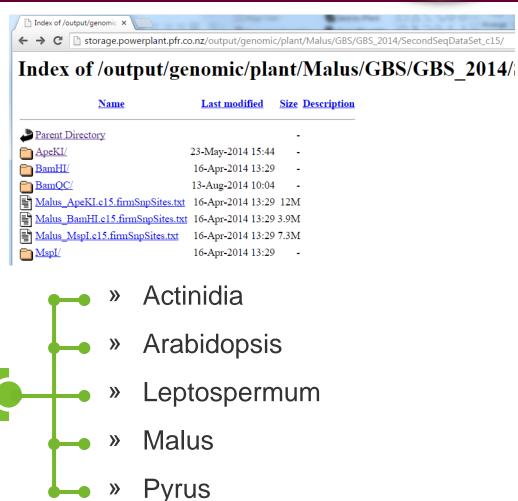




Best Practice: Share Result at /output



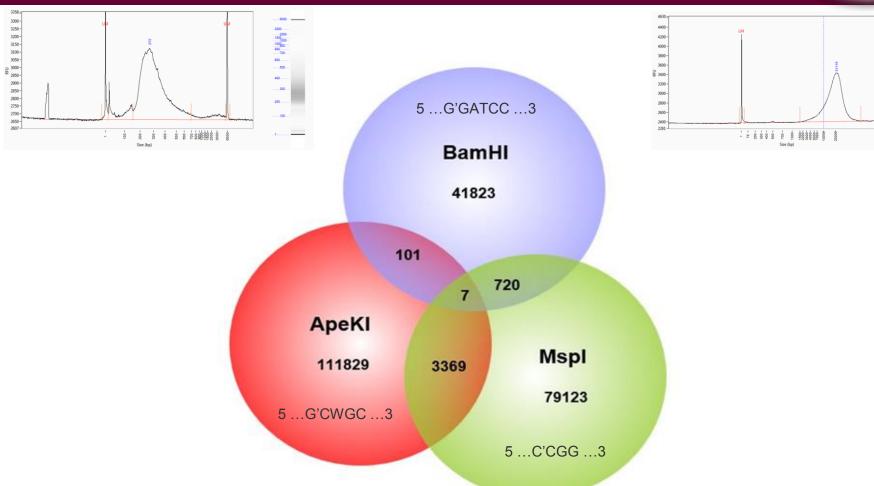
- » Bacterial
- » Fish
- » Fungal
- » Homo
- » Insect
- » Metagenomics
- » NCBI
- » Plant
- » TrainingData
- » Viral





Best Practice: Test Enzymes for New Crop





SNP Sites Detected

(Tags supported by at least 15 reads)



Conclusions





Track
as much
metadata as
you can in a
searchable
way



Acknowledgements



The Bioinformatics team



Scientists

- David Chagné
- Elena Hilario
- Chris Kirk
- Sue Gardiner
- Satish Kumar
- John McCallum

IKS

- Eric Burgueno
- Nicolas Babot
- Zane Gilmore
- Matthew Lawrenson



Questions and Discussions



Thank You!









The New Zealand Institute for Plant & Food Research Limited



Cecilia.Deng@plantandfood.co.nz

Best Practice: Choose a Variant Calling Tool



» TASSEL/GBS

[hrachd	rachd@genome7 03_HapMapFilteredSNPs]\$ head tmp.txt							
rs#	alleles chrom	pos strand	assembly#	center protLSI	D assayL	SID panelL	SID QCcode	GgDNA 100
A_107	GgDNA_108	GgDNA_109	GgDNA_112	GgDNA_113	GgDNA_114	GgDNA_115	GgDNA_116	GgDNA 118
A_123	GgDNA 124	GgDNA 126	GgDNA 127	GgDNA 128	GgDNA 129	GgDNA 12	GgDNA 130	GgDNA 131

» STACKS

ld	SNP	Consensus	Matches	Ratio
> 9 annotate	1	GATCCGTTTCATTGCTTCCTCTGACAACTGCTTTCACTGTATGTA	48	aa: 17 (35.4%) ab: 9 (18.8%) abc: 18 (37.5%) abcd: 3 (6.2%) bc: 1 (2.1%)

» FreeBayes or Samtools

```
##fileformat=VCFv4.1
##fileDate=20150908
##source=freeBayes v0.9.21
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT BAR9135
208 BAR9210 BAR9216 BAR9228 BAR9230 BAR9234 BAR9238 BAR9248 BAR9254 BAR9256
```

