Genome-wide association studies using high-throughput SNP analysis in tetraploid potato

Sanjeev Kumar Sharma[a], Micha Bayer[a], Karen McLean[a], Katrin McKenzie[b], Steve Daniels[c], Finlay Dale[a], Glenn Bryan[a]

[a]The James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK
[b]Biomathematics and Statistics Scotland, Invergowrie, Dundee DD2 5DA, UK
[c]Cygnet PB Ltd, Great Abington, Cambridge, CB21 6AS, UK

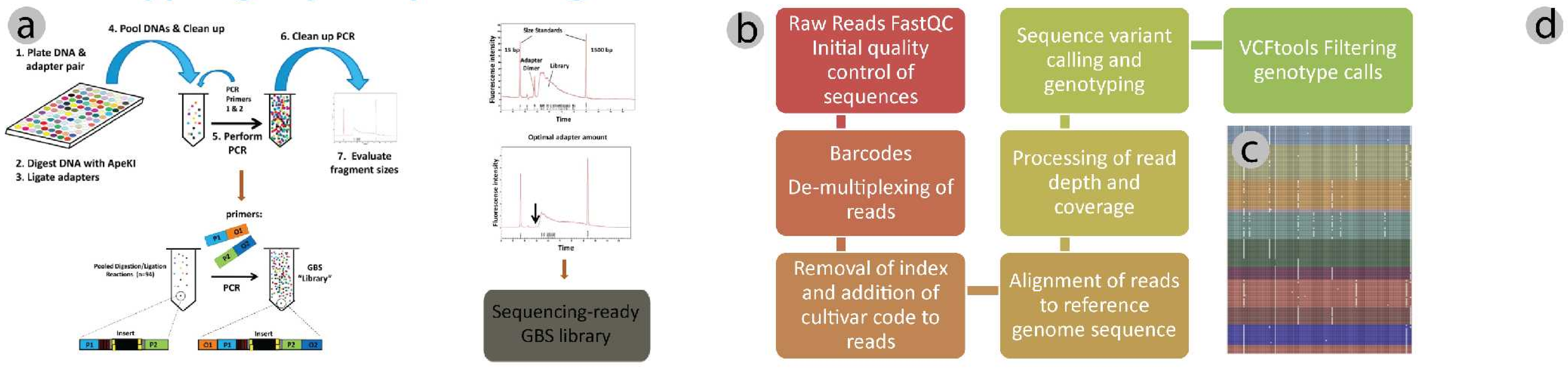
E-mail: sanjeev.sharma@hutton.ac.uk

The James Hutton Institute

Introduction

The cultivated potato is a highly heterozygous outbreeder, which exhibits tetrasomic inheritance. Due to the genetic intricacies posed by polyploidy potato genetics is often performed at the diploid level using bi-parental populations. Translating outcomes from diploid level studies to the commercially practised tetraploid level is not always straightforward, notably where allele dosage is important, such as the many economically important complex quality and physiological traits. The accurate dissection of traits at the tetraploid level across a wide range of germplasm is highly desirable. Genome-wide association studies (GWAS) hold great promise for studying the genetics of natural variation and agronomic traits. It also offers significant advantages, such as enhanced mapping resolution, increased numbers of segregating traits, and greater allelic diversity than traditional mapping using bi-parental crosses. However, a careful assessment of population structure in GWAS is required to avoid spurious associations arising from systematic variances in allele frequencies due to differences in sample ancestries. We have developed an association mapping panel of ~350 diverse autotetraploid potato cultivars and breeding lines which have been phenotyped in two environments over two growing seasons. The panel has been subjected to a Genotyping-by-Sequencing pipeline for simultaneous genotyping as well as marker discovery including a genome scan using the SolCAP 8k Infinium SNP array. These SNPs have been exploited to perform kinship and population stratification analyses, and to further perform an extensive GWAS study on a number of key agronomically and economically important traits. Insights from this investigation including an assessment of diversity, genomic complexity, population structure and a genome-wide survey of linkage disequilibrium are presented.

Genotyping-by-Sequencing



d

Clone	GBS-SNP1	GBS-SNP2	GBS-SNP3	GBS-SNP4	GBS-SNP5	GBS-SNP6	GBS-SNP7	GBS-SNP8	GBS-SNP9	GBS-SNP*n*
Adirondack	2	1	3	1	3	1	3	3	0	2
Alwara	2	1	2	1	2	2	1	1	1	4
Exquisa	3	2	4	0	4	0	2	1	1	4
Fianna	3	2	1	3	0	4	2	0	2	3
Fiftyfold	0	2	4	0	4	0	2	0	1	3

Clone	GBS-SNP1	GBS-SNP2	GBS-SNP3	GBS-SNP4	GBS-SNP5	GBS-SNP6	GBS-SNP7	GBS-SNP8	GBS-SNP9	GBS-SNP*n*
Adirondack	AABB	ABBB	AAAB	ABBB	AAAB	ABBB	AAAB	AAAB	BBBB	AABB
Alwara	AABB	ABBB	AABB	ABBB	AABB	AABB	ABBB	ABBB	ABBB	AAAA
Exquisa	AAAB	AABB	AAAA	BBBB	AAAA	BBBB	AABB	ABBB	ABBB	AAAA
Fianna	AAAB	AABB	ABBB	AAAB	BBBB	AAAA	AABB	BBBB	AABB	AAAB
Fiftyfold	BBBB	AABB	AAAA	BBBB	AAAA	BBBB	AABB	BBBB	ABBB	AAAB

Figure 1: (a) GBS procedure adapted from Elshire *et al.* (2011) and Poland *et al.* (2012), PLoS-One, (b) Bioinformatics pipeline for processing GBS data and variant calling, (c) GBS *PstI* tag displaying varying levels of heterozygosity in tetraploid potato. Each colour panel represents a different clone and each white rectangle indicates alternate base (SNP) in comparison to the reference DM potato genome (PGSC, 2011 Nature), and (d) GBS SNP genotype calls represented in numerical and genotypic format following a tetraploid allelic dosage model.

Characterization of Association Panel

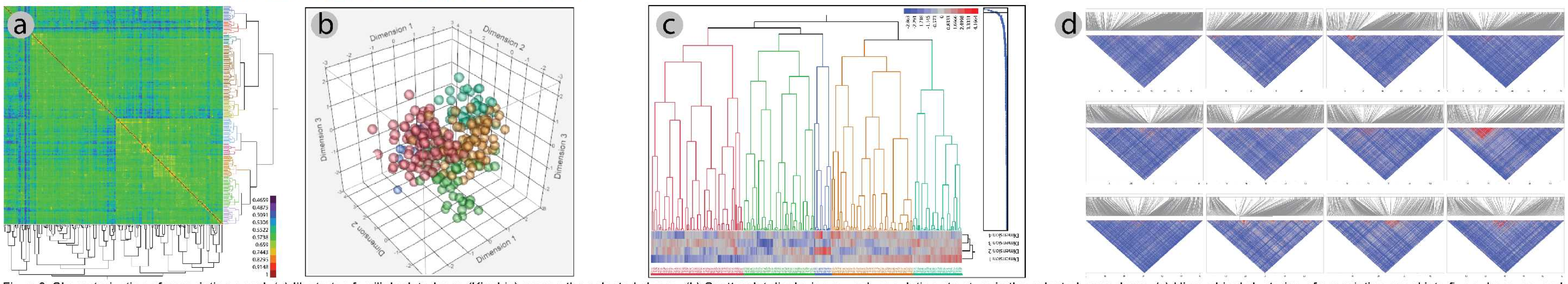


Figure 2: Characterization of association panel: (a) Illustrates familial relatedness (Kinship) among the selected clones, (b) Scatterplot displaying a weak population structure in the selected germplasm, (c) Hierarchical clustering of association panel into five subgroups, and (d) genome-wide linkage disequilibrium scan.

Phenotyping

Figure 3: Field trials for phenotypic evaluation of targeted traits at two different locations over two consecutive years.

GWAS Models

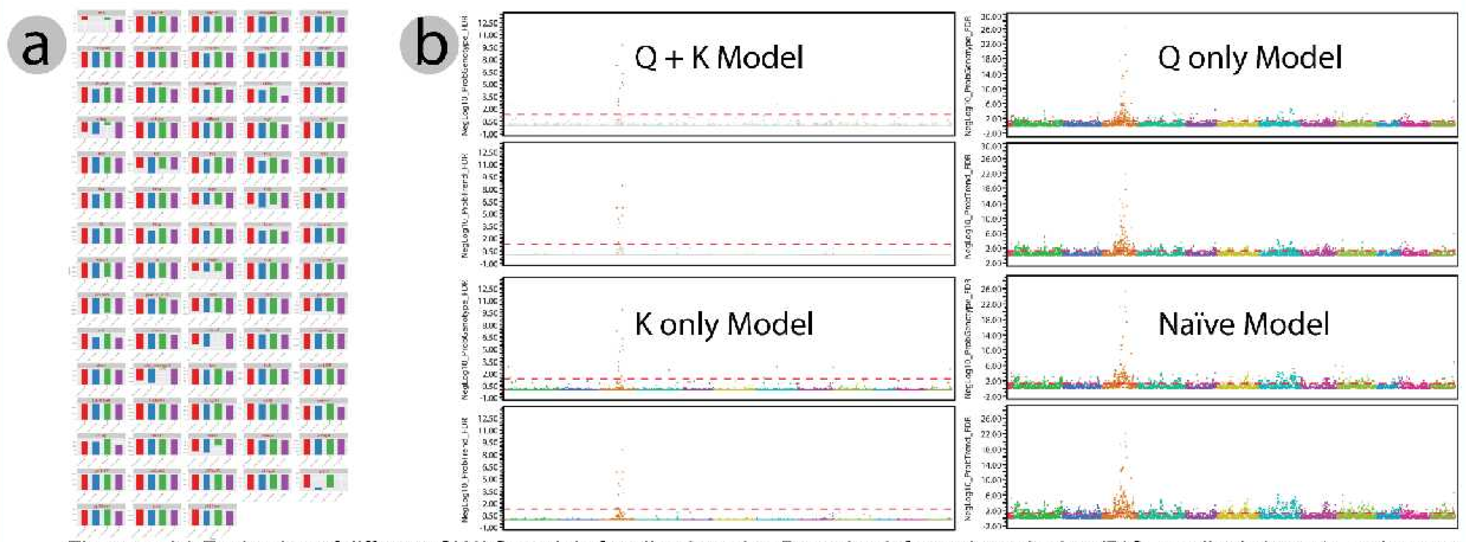


Figure 4: (a) Evaluation of different GWAS models for all traits using Bayesian information criterion (BIC, smaller is better) - red: naïve model, blue: kinship (k) only, green: structure (Q) only, purple: kinship and structure both included in the mixed model, (b) representation of these models using tuber flesh colour GWAS. Top and bottom panels in each model are from genotype and trend (additive coding) association tests, respectively.

Conclusions

- GBS validated in tetraploid material
- Availability of a set of ~45k robust SNPs and a well characterized association panel
- Weak population structure in tetraploid material but strong enough to carefully address in GWAS studies
- Good associations with simple and more complex traits
- Potential for developing diagnostic molecular markers for keys marker-trait associations

GWAS and Development of Diagnostic Markers

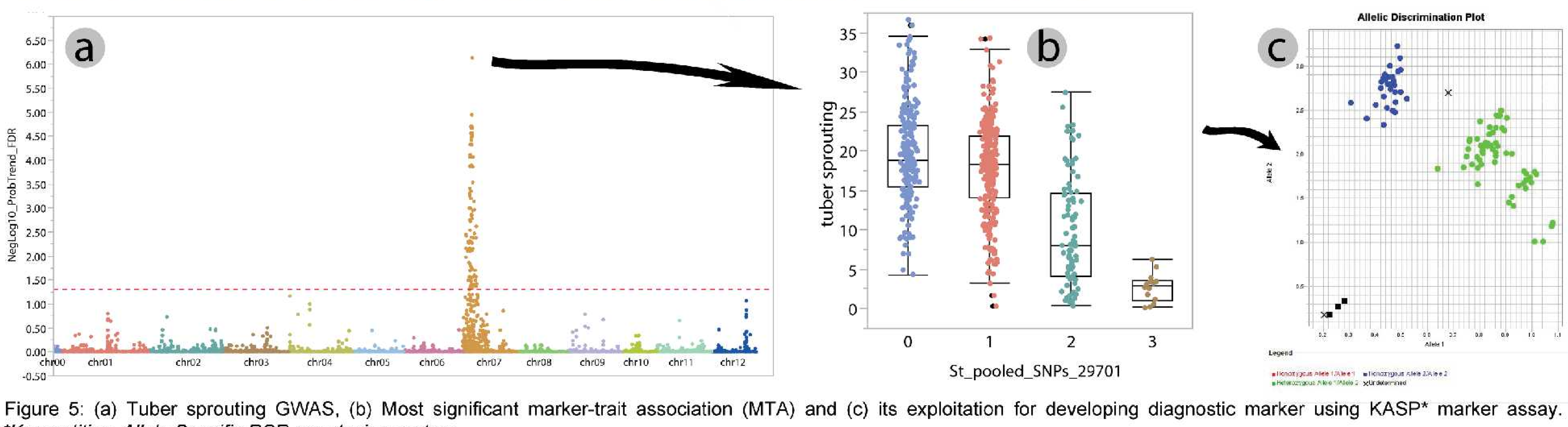


Figure 5: (a) Tuber sprouting GWAS, (b) Most significant marker-trait association (MTA) and (c) its exploitation for developing diagnostic marker using KASP* marker assay.
*Kompetitive Allele Specific PCR genotyping system

Acknowledgements

This project was jointly funded by the UK Technology Strategy Board and Industrial partners, namely Cygnet Potato Breeders Ltd., McCain Foods (GB) Ltd. and James Hutton Ltd.

References

Elshire *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS One 6: e19379

Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475: 189-195

Poland *et al.* (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 15:e3225