Accelerating marker-assisted selection - from dirt to data

Dr. Jonathan Curry – Senior Scientist - Genomics
Introduction

• Who are LGC?

• Where do SNPs grow?

• How might they be harvested for breeding?

• You reap what you sow – applying technology
LGC Genomics is a part of the LGC Group

- LGC established in 1842 as a scientific testing laboratory for the UK government
  - Privatised in 1996
- Retains its role as UK National Measurement Institute
  - Standard bearer for chemical and bio-analytical measurement
- Global presence
  - 33 locations
  - 2,000+ staff
  - >3-fold growth since 2004

<table>
<thead>
<tr>
<th>LGC Health Sciences</th>
<th>LGC Genomics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leader in analytical testing services for food, drug, and consumer products</td>
<td>Genomic solutions for DNA/RNA extraction, genotyping, and sequencing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LGC Forensics</th>
<th>LGC Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>The world’s largest supplier of products and services for forensic analysis</td>
<td>Production and supply of chemical and biochemical reference materials</td>
</tr>
</tbody>
</table>

LGC Science & Technology
Home of the UK National Measurement Institute
LGC Genomics Labs - Hoddesdon
SNP genotyping project process – Kraken™ integration check points

Kragen - Correct project samples received/ assays match regions of interest / made correctly

Kragen – Correct samples to be used for project / sample template logged

Kragen – Correct Samples matched to assay. No machines can operate if no match

Kragen - Logs all results by plate and sample reference/barcode. All tracked to the unique project number

Project Manager / Project Design

Check

Samples Logged and Tracked

Check

Assay Run

Check

Results

Kraken™ integrated SNPlines checks at vital points of the process ensure the correct sample and project are always tied.
Distribution of Life in Kraken

- Plant: 63%
- Animal: 17%
- Fish: 7%
- Insect: 7%
- Bacteria / Yeast / Protozoan: 4%
- Human: 2%
Where do SNPs grow?
High-Throughput Marker Discovery
Discovery of valuable traits using Next-generation sequencing

- Using High-throughput sequencing to survey genomes in bulk
- Markers are used to identify differences between genomes.

- Use markers when cross breeding varieties.
- Once markers are selected they can be run over hundreds / thousands / millions of samples.
- Reduction of time to market – reduces cost / increases profitability
**KASP genotyping landscape**

- **SNP screening**
- **SNP QC**
- **SNP Validation**
- **GBS/GWAS: SNP ID**

- **Initial set of candidate SNPs identified**
- **High value targets screened against a much larger set of Samples – enabling MAS**
- **Initial candidates screened against a larger # of sample for greater statistical analysis. High value SNP markers validated.**

**Number of SNPs**
- 100
- 1,000
- 10,000
- 100,000

**Number of Samples**
- 100
- 1,000
- 10,000
- 100,000
Trait Identification

The reduction in cost of NextGen sequencing has meant a dramatic increase in possible loci to target.

Position of desired trait marker

Fine mapping with KASP genotyping

Design flexibility allows more markers to be targeted to saturate regions

By using more SNP markers (higher resolution mapping) you will find desired trait faster.

KASP assays allow higher resolution, flexible mapping of organisms.
High Throughput Marker Assisted Selection

Sequencing and informatics
Marker Identification

KASP

Development

Evaluate different generations in High - Throughput.
How might they be harvested for breeding?
- Assay mix - contains two allele specific primers and common reverse primer(s). This is often referred to as a KASP Assay.

- Universal Master mix - contains fluorophores, quenchers and Taq polymerase.
KASP – flexible for all platforms

KASP is very simple and scalable…

• Anyone who has a pipette, PCR machine and a way of reading FRET can run KASP!!

• System agnostic

• LGC Genomics service labs can run your project for you.

• Bi-allelic SNP data is general purpose marker currency – easy to exchange and work with.
Flexible design

• KASP is PCR
  – Can be designed as such
  – Can be optimised as such (Temp / Betaine / DMSO/ Mg^{2+})
  – 90% assay conversion
• Choose loci with a fair chance of designing primers.
• i.e. not:
  – Non-unique and highly spread throughout genome (although this can be solved)
  – Long homo-polymer repeats i.e. AAAAAAAA
  – VNTRs / micro-satellites – although Hybeacons can be used here.
  – Copy number
Solutions for homology and polyploidy

Want to save time and money on breeding / mapping projects with KASP?

• Answer?

  • BLAST, BLAST and BLAST some more!!!!!!!

• Very simple to do with available databases for many common crops.

• Turn off complexity filters and run Mega-BLAST then BLASTN.

• Even partially sequenced genomes can give some information about choice of markers to design.
Solutions for homology and polyploidy
Polyploid data – allele dosage

The central set of data points are from known heterozygotes.

These are wheat crosses with clear allele dosage effect seen centrally.

If using this kind of experiment then well normalised DNA is crucial.

Heterozygotes.
In a tetraploid organism these samples could have any of the following genotypes:

1:1:1:2
1:1:2:2
1:2:2:2
You reap what you sow – applying technology

Science
for a safer world
Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.)

Alexandra M. Allen¹,*, Gary L. A. Barker¹, Paul Wilkinson¹, Amanda Burridge¹, Mark Winfield¹, Jane Coghill¹, Cristobal Uauy², Simon Griffiths², Peter Jack³, Simon Berry⁴, Peter Werner⁵, James P. E. Melichar⁶, Jane McDougall⁷, Rhian Gwilliam⁷, Phil Robinson⁷ and Keith J. Edwards¹

¹School of Biological Sciences, University of Bristol, Bristol, UK  
²John Innes Centre, Norwich, UK  
³RAGT, Ickleton, Essex, UK  
⁴Limagrain, Woolpit, Suffolk, UK  
⁵KWS, Thriplow, Hertfordshire, UK  
⁶Syngenta Seeds Ltd, Whittlesford, Cambridge, UK  
⁷KBioscience Unit 7, Hertfordshire, UK
Wheat SNP validation – public projects

Bristol/JIC KASP assays

<table>
<thead>
<tr>
<th>Cross</th>
<th>Mapped Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avalon v Cadenza</td>
<td>3,028</td>
</tr>
<tr>
<td></td>
<td>1,122 A genome</td>
</tr>
<tr>
<td></td>
<td>1,520 B genome</td>
</tr>
<tr>
<td></td>
<td>386 D genome</td>
</tr>
<tr>
<td>Savannah x Rialto</td>
<td>1,543</td>
</tr>
<tr>
<td></td>
<td>790 A genome</td>
</tr>
<tr>
<td></td>
<td>594 B genome</td>
</tr>
<tr>
<td></td>
<td>159 D genome</td>
</tr>
<tr>
<td>Synthetic x Opata</td>
<td>201</td>
</tr>
<tr>
<td></td>
<td>70 A genome</td>
</tr>
<tr>
<td></td>
<td>73 B genome</td>
</tr>
<tr>
<td></td>
<td>49 D genome</td>
</tr>
</tbody>
</table>

- First report of a public linkage map for hexaploid wheat based on KASP to genotype wheat varieties and generate a linkage map.
- 67% polymorphic in varietal screen.
- 4% monomorphcic in hexaploid wheat, but polymorphic compared to diploid/tetraploid varieties.

GCP KASP assays

<table>
<thead>
<tr>
<th>GCP originator</th>
<th>Validated KASP Assays</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr Susanne Dreisigacker (CIMMYT)</td>
<td>1,864</td>
</tr>
</tbody>
</table>

- Validation on wheat cultivars originating from Australia, China, India and Mexico.
- Included high anther culture ability and disease resistant varieties.
Feed the world
Generation Challenge Program (GCP) panels

- **Select germplasm**
- **Genomic tools**
- **Identify and tag genes needed**
- **Select favourable alleles**
  - **IB Platform**

- **Maize RI**
- **Rice RI**
- **Sorghum RI**
- **Wheat RI**
- **Comparative genomics RI** (maize, rice, sorghum)
- **Legumes RI** (beans, chickpeas, cowpeas, groundnuts)
- **Cassava RI**

- **Capacity building**
- **Data management**

**GCP products**

- **Improved germplasm in farmers’ fields**

**Generation Challenge Programme**
Partnerships in modern crop breeding for food security
Aim is to reduce stunting by increasing the nutritional value and yields of 100 local African crops by training the best African breeders to use the best tools in the world.

LGC are primary partner of the program and took part in training the 1st 25 breeders in Nairobi 2014.
Plant sampling essentials
AOCC Nairobi 2014
Outline protocol

LGC leaf sampling kit

Collecting leaf samples
Repeat this process for each plant you want to sample, using a new tube each time.

Cut leaf discs
- Place the leaf to be sampled on the cutting mat.
- Uncap the cutting tool.
- Cut a disc out of the leaf by pushing the cutting tool into the leaf; twist the tool as you push to make the tool pick the disc up.

Fill tubes with samples
- Insert the end of the tool into one of the racked tubes, and depress the plunger to dispense the disc.
- The tubes are in strips of 12, and can be removed from the rack if needed at this step.
- Add 1-12 discs per tube, depending on the project (see ‘How many disks are required?’)

Wash cutting tool
- Place the end of the cutting tool in some clean water and depress the plunger a few times.
- Flick the tool until completely dry.
- Please ensure the rack is labelled appropriately.

Preparing samples for transport

Seal tubes
- Place the perforated Strip caps on top of the tubes.
- Press firmly to ensure caps are secured.
- Remove the desiccant from its small sealed bag.

Add desiccant
- Place the desiccant sachet on top of the rack of tubes and replace lid.
- Place completed rack inside the larger (labelled) bag.
- Force the air out of the bag and then seal it.

Prepare for shipping
- Place the sealed bag in a suitable container.
- Our address: LGC, Extractions department, Units 1 & 2 Trident Industrial Estate, Pindar Road, Hoddesdon, Hertfordshire, UK, EN11 0WZ.
- Provide a description of the contents for customs.

How many disks are required?

![Graph showing how many disks are required for different numbers of samples.](image)

Kit contents:
- 1 x 96-well tube storage rack with lid
- 12 x perforated strip caps
- 1 x 50g desiccant sachet (in a bag)
- 1 x larger (labelled) sealable bag
- 1 x cutting tool & 1 x cutting mat

If multiple kits have been requested, only 1 x cutting tool and 1 x cutting mat will be sent to you.

We may be able to handle projects outside the specifications given in this document; if this is the case please contact us directly to discuss your requirements:

+44 (0)1902 470757  info.uk@lgcgenomics.com

© LGC Limited, 2014. All rights reserved.
Agricultural crop improvement case study: The story of the Pigeon pea revolution

- KASP Genotyping assay run by LGC genotyping service lab
- Marker-assisted breeding for improved crops

Pigeon pea circa. 1992
- Slow growing
- Low yield
- No commercial value
- Neglected by crop breeders
- Un-developed potential

Pigeon pea circa. 2013
- Breeding program sponsored
- Accelerated MAS using SNPs
- SNPs analysed with KASP
- Drought & disease resistant
- Excellent food supply
- Better market value
- Established life changing market value chain for growers in Tanzania
Colorado State University
Wheat Breeding and Genetics Program

- Scott Healy’s group use KASP to identify and monitor traits for Colorado’s harsh climate

- They also use KASP to find traits for improving human health such as starch quality and antioxidants.
Development of gene specific KASP markers: a toolbox for marker-assisted selection in wheat.

Gina Brown-Guedira 1, Neelam Kumari 2, Susan Dreisigacker 3, Peter Sharp 4, Catherine Ravel 5, and Cristobal Uauy 6.

1 USDA–ARS Plant Science Research, North Carolina State University, Raleigh, NC, USA; 2 Department of Crop Science, North Carolina State University, Raleigh, NC, USA; 3 International Maize and Wheat Improvement Center (CIMMYT), Mexico; 4 Plant Breeding Institute, University of Sydney, Narellan NSW, Australia; 5 INRA, Université Blaise Pascal, Genetics, Diversity and Ecophysiology of Cereals, Clermont-Ferrand, France; and 6 Department of Crop Genetics, John Innes Centre, Colney, Norwich, United Kingdom.

Whole-genome, SNP detection technologies now available in construction of linkage maps, genome-wide association mapping these technologies are not suited for marker-assisted breeding the same polymorphism on tens of thousands of plants in theative markers amenable to high-throughput genotyping are needed and in combination with other marker-assisted breeding approach of an international collaboration to develop and make publically wheat. Stream-lined homogeneous assays were developed using assay design, making it well-suited for use in polyploid wheat methods. Reported sequence variation (SNPs and indels) were the reduced height, vernalization, and photoperiod-response for cloned disease resistance and end-use quality genes. In some the causal gene sequence. Also, a number of genes in wheat assays for these, associated SSR and STS markers were evaluaWheat Infinium Assay. SNP in linkage disequilibrium (r²>0.9) approach, markers were developed that are highly predictive the Sr36 and Sbm1 resistance genes, and alleles at the Glu-D1 association mapping experiments, coupled with SNP genotyping and current sequencing projects in wheat, will result in identification of numerous sequence targets for development of new predictive, homogeneous assays for important genes. We encourage researchers to contribute sequences for development of additional publicly available KASPassays for use in wheat improvement programs world-wide.

“KASP provides flexibility in assay design, making it well-suited for use in polyploid wheat, and is low cost”

“We encourage researchers to contribute sequences for development of additional publicly available KASP assays for use in wheat improvement programs world-wide”
<table>
<thead>
<tr>
<th>Marker</th>
<th>Trait</th>
<th>Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-hydroxy-3-methylglutaryl-coenzyme A reductase</td>
<td>Ggc grey ground colour</td>
<td>1</td>
</tr>
<tr>
<td>40S ribosomal protein S31, chloroplastic</td>
<td>Yc cotyledon colour</td>
<td>1</td>
</tr>
<tr>
<td>40S ribosomal protein S15-4</td>
<td>Ggc grey ground colour</td>
<td>2</td>
</tr>
<tr>
<td>40S ribosomal protein S15a-1</td>
<td>Tgc tan ground colour</td>
<td>2</td>
</tr>
<tr>
<td>40S ribosomal protein S18</td>
<td>Tgc tan ground colour</td>
<td>2</td>
</tr>
<tr>
<td>40S ribosomal protein S18</td>
<td>Tgc tan ground colour</td>
<td>3</td>
</tr>
<tr>
<td>60S ribosomal protein L31, chloroplastic</td>
<td>Tgc tan ground colour</td>
<td>3</td>
</tr>
<tr>
<td>60S acidic ribosomal protein P1-2</td>
<td>Seed diameter, plumpness</td>
<td>3</td>
</tr>
<tr>
<td>60S ribosomal protein L27a-2</td>
<td>Seed diameter, plumpness</td>
<td>3</td>
</tr>
<tr>
<td>ABSCISIC ACID-INSENSITIVE 5-like protein 5</td>
<td>Seed thickness</td>
<td>4</td>
</tr>
<tr>
<td>ABSCISIC ACID-INSENSITIVE 5-like protein 5</td>
<td>Seed thickness</td>
<td>4</td>
</tr>
</tbody>
</table>

**Genetic**

- **Chromosome:**

**Physical**

- **Mapping Min:**
- **Mapping Max:**
- **Distance/Mapping Interval:**

**Plant Assay Cart**

Your cart is empty.

[Add to cart]
Web assay search tool

- A web based tool for locating functionally validated SNPs for crops.
- Started out for human validated SNP (more than 100 k!)
- We thought that it might be useful for breeding communities.
- Not there to compete but compliment projects
- We aim for it to:
  - Deposit markers for everyone to use – Can be quickly (minutes) with new markers for sharing.
  - Links to original material
  - Gives basic information about the marker (i.e. sequence or map position (cM))
  - An easy way to order assays
Assay Search Tool – Submitting

- You can add information linking to reference
- Point me towards where the information is and I’ll do the rest.
  - OR
- Provide either map units (cM – reference map) or nucleotide positions (provide the reference genome name).
- Provide marker reference name / sequence
- I’ll do the rest
What will be available?

- Know Pulse – Lentil – core 154 initial assays but more to come

- Sol Cap - All Tomato validated 384 assays and 7k *in silico* designed

- Wheat – cerealsDB assays – around 9000 validated

- More to come!!
Any Questions?