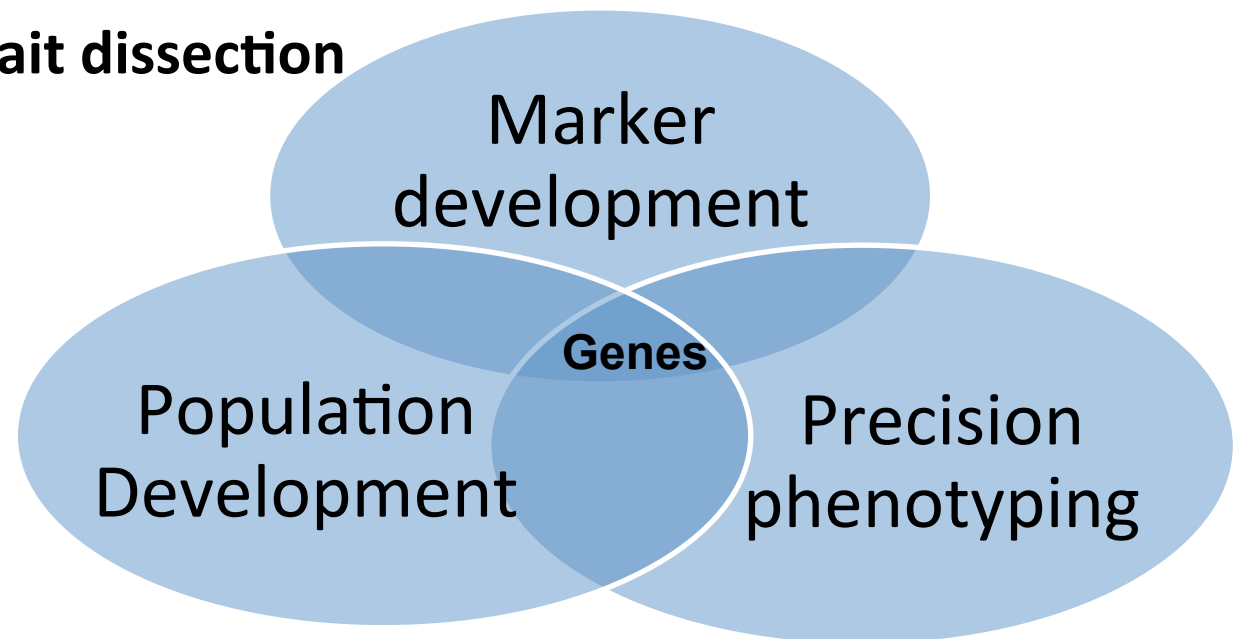


What can we learn from diploid oats? The IBERS Diploid Avena Genomics Resource (DAGR) and its application to hexaploid oat breeding

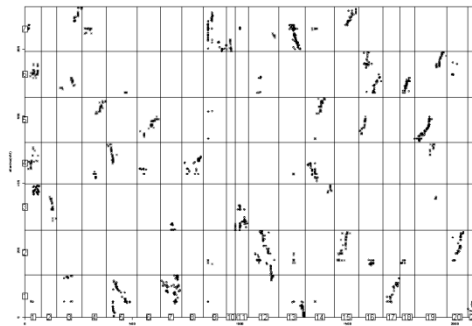
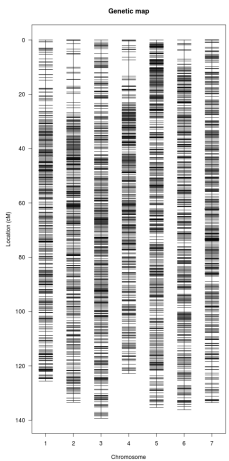
**Catherine Howarth, Rob Vickerstaff, Tim Langdon, Matt
Hegarty, Irene Griffiths, Sandy Cowan, Athole Marshall
IBERS, Aberystwyth University, UK**

Developing tools and resources for oat breeding

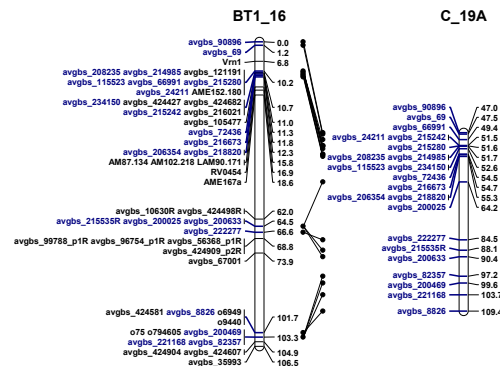
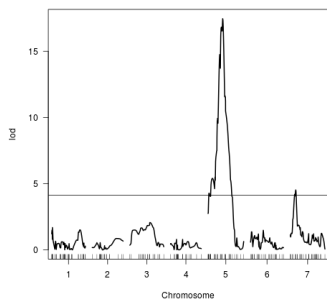
- **Marker, map and genome resource development**
- **Population, germplasm and comparative genetic resource development**
- **Phenotyping and trait dissection**



Working across scales- from field to genome

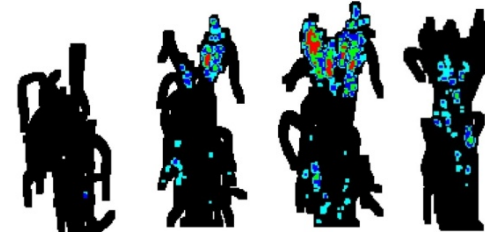


Population
development



Diploid sequencing
mapping
QTL analysis

Hexaploid mapping
QTL analysis
Comparative genomics



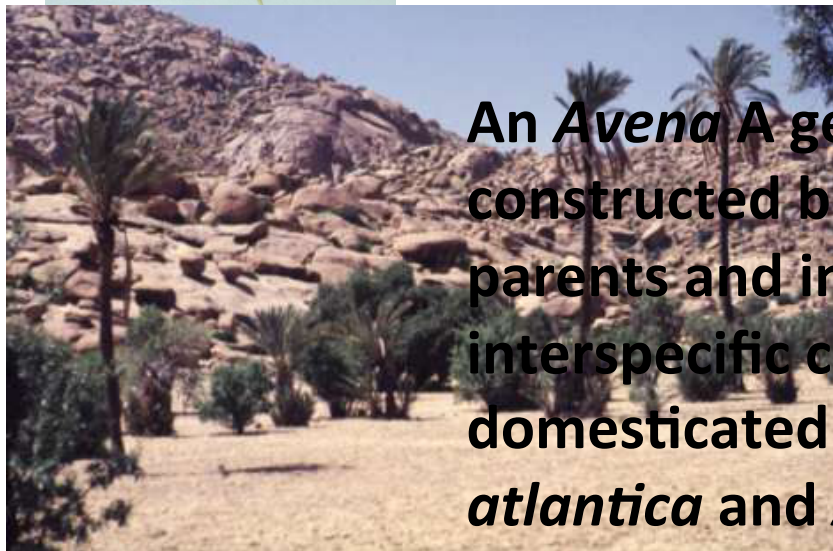
Trait dissection
Association
with markers
MAS

Avena comprises a polyploid series of species

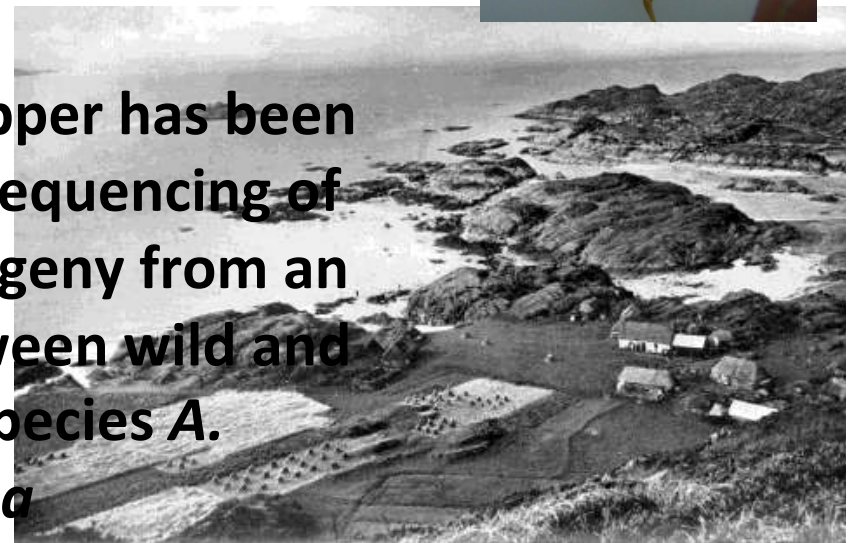


- Diploids (14 chromosomes)
- Tetraploids (28)
- Hexaploids (42) ~13000Mb

The cultivated oat is hexaploid, containing AA, CC and DD genomes



An *Avena* A genome zipper has been constructed based on sequencing of parents and inbred progeny from an interspecific cross between wild and domesticated diploid species *A. atlantica* and *A. strigosa*



A. atlantica genome assembly

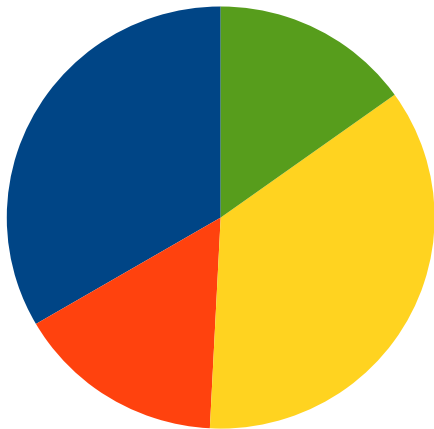


Illumina HiSeq
 2000



```
[rov@bert avena]$ ls atl* -lh
-r--r----- 1 rov quats 2.1G Aug 22 2013 atlan01-R1.fq.gz
-r--r----- 1 rov quats 2.2G Aug 22 2013 atlan01-R2.fq.gz
-r--r----- 1 rov quats 3.0G Aug 22 2013 atlan02-R1.fq.gz
-r--r----- 1 rov quats 3.1G Aug 22 2013 atlan02-R2.fq.gz
-r--r----- 1 rov quats 8.7G Aug 22 2013 atlan03-R1.fq.gz
-r--r----- 1 rov quats 8.7G Aug 22 2013 atlan03-R2.fq.gz
-r--r----- 1 rov quats 7.9G Aug 22 2013 atlan04-R1.fq.gz
-r--r----- 1 rov quats 7.9G Aug 22 2013 atlan04-R2.fq.gz
-r--r----- 1 rov quats 7.3G Aug 22 2013 atlan05-R1.fq.gz
-r--r----- 1 rov quats 7.2G Aug 22 2013 atlan05-R2.fq.gz
-r--r----- 1 rov quats 4.6G Aug 22 2013 atlan06-R1.fq.gz
-r--r----- 1 rov quats 4.5G Aug 22 2013 atlan06-R2.fq.gz
-r--r----- 1 rov quats 4.3G Aug 22 2013 atlan07-R1.fq.gz
-r--r----- 1 rov quats 4.3G Aug 22 2013 atlan07-R2.fq.gz
-r--r----- 1 rov quats 259M Aug 22 2013 atlan08-R1.fq.gz
-r--r----- 1 rov quats 279M Aug 22 2013 atlan08-R2.fq.gz
-r--r----- 1 rov quats 7.5G Aug 22 2013 atlan09-R1.fq.gz
-r--r----- 1 rov quats 7.7G Aug 22 2013 atlan09-R2.fq.gz
-r--r----- 1 rov quats 7.4G Aug 22 2013 atlan10-R1.fq.gz
-r--r----- 1 rov quats 7.7G Aug 22 2013 atlan10-R2.fq.gz
-r--r----- 1 rov quats 14G Oct 2 2013 atlan11-R1.fq.gz
-r--r----- 1 rov quats 14G Oct 2 2013 atlan11-R2.fq.gz
```

4.1 gigabase genome

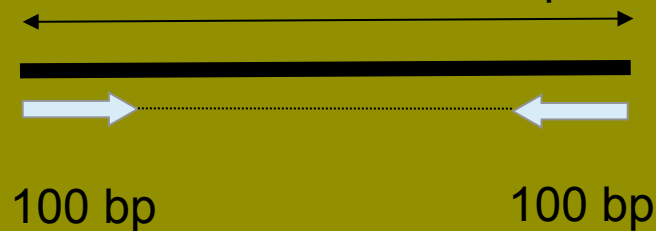


- unassembled (33%)
- scaffolding (16%)
- repetitive sequence (36%)
- unique sequence (15%)



157 gigabase
 data
 ~40x coverage

200/350/700/2000 bp





Genetic Mapping

10000 GBS markers, 92 RILs



genetic map

A. strigosa



x

A. atlantica



F1

.

.

selfing

.

.

.

186 F6 individuals (RILs)

Genome Assembly

Illumina reads
40x coverage PE/MP
6x coverage PE
2x coverage PE of 15 RILs

atlantica genome ssembly
strigosa SNP catalog

Genome Annotation

- Compare sequence to known grass genes
 - Brachypodium, rice, sorghum, barley
 - Uniprot Poaceae proteiins
- gmap - align oat transcripts
 - *A. barbata*
 - *A. sativa*
- ab initio gene prediction
 - Augustus
- CEGMA (conserved eukaryotic) genes
 - 84% complete
 - 96% at least partial
- identify repeats
 - repeatmodeler, repeat masker

Contig browser (Gbrowse)

gene models

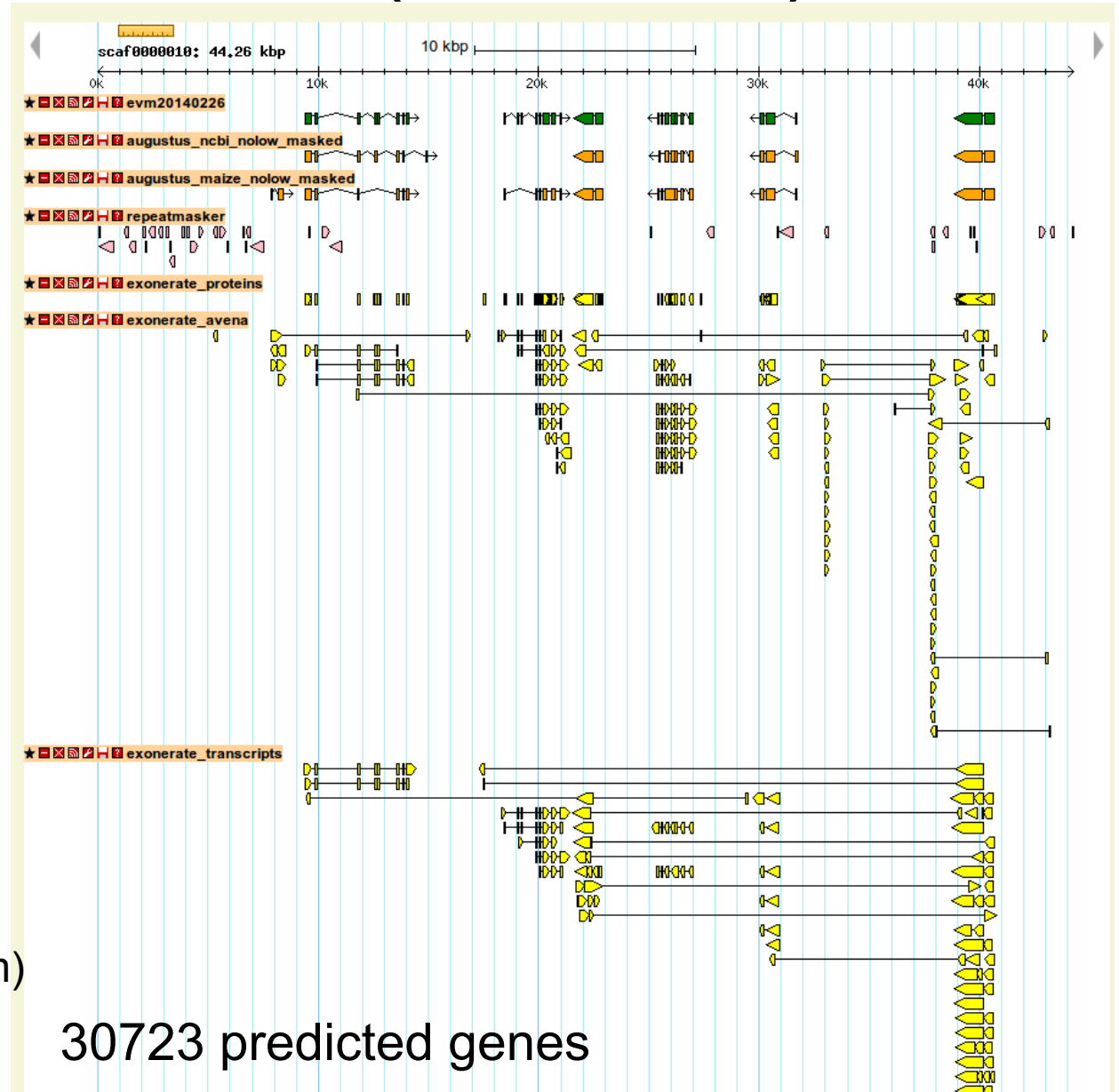
ab initio predictions

repetitive sequences

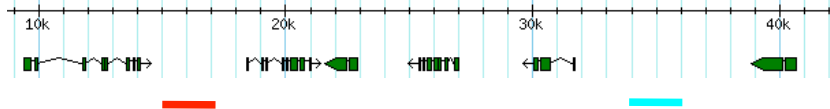
aligned grass proteins

aligned oat
transcripts

aligned grass transcripts
(*Brachypodium*, rice, sorghum)

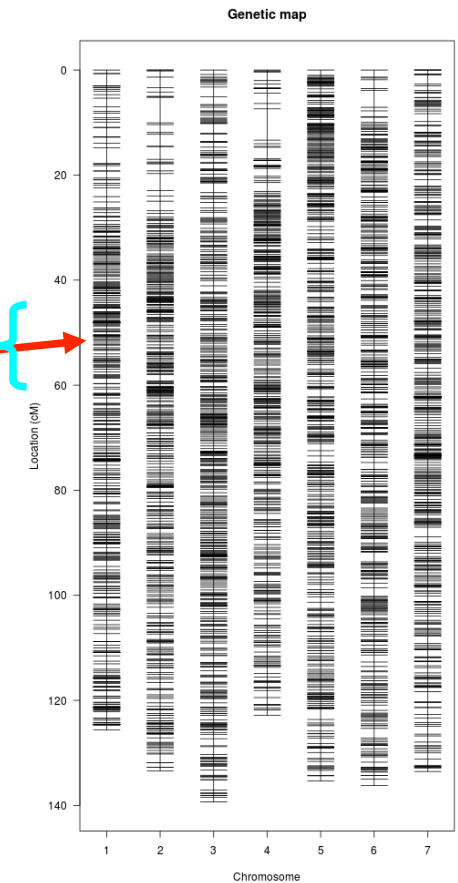


Placing Genes on the Map

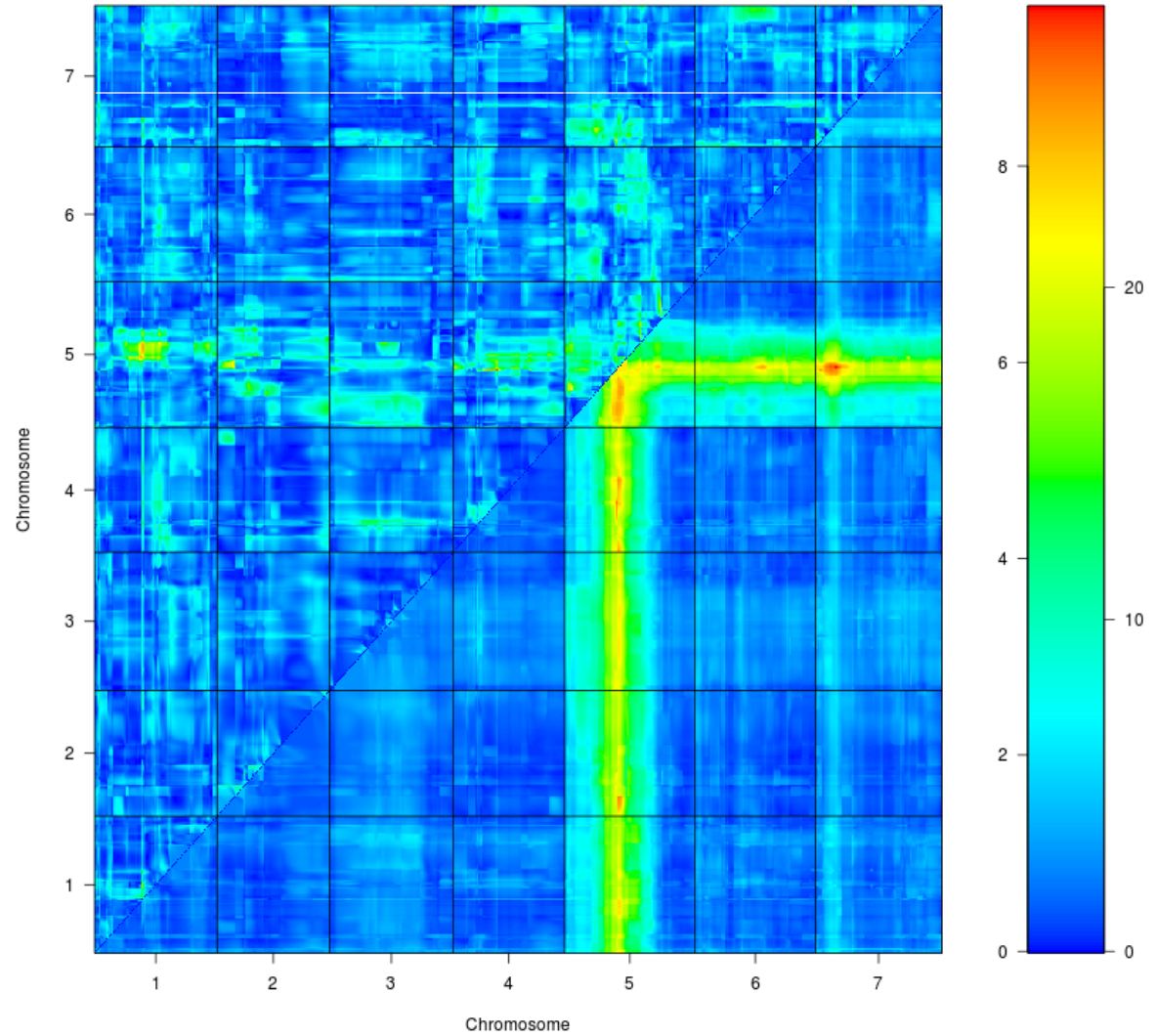
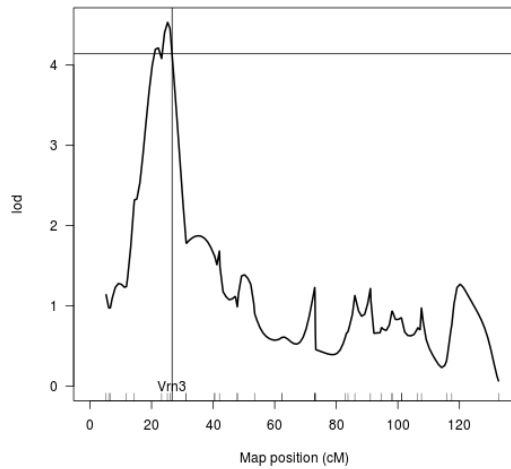
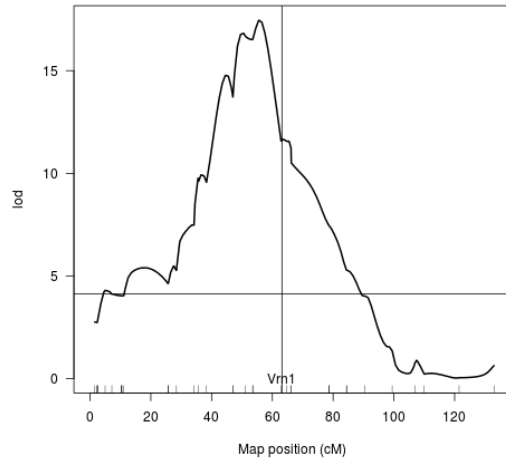


10000 GBS
markers
tested on parents
and 92 RILs

4.2 million WGS
markers tested
on parents and
15 selected RILs

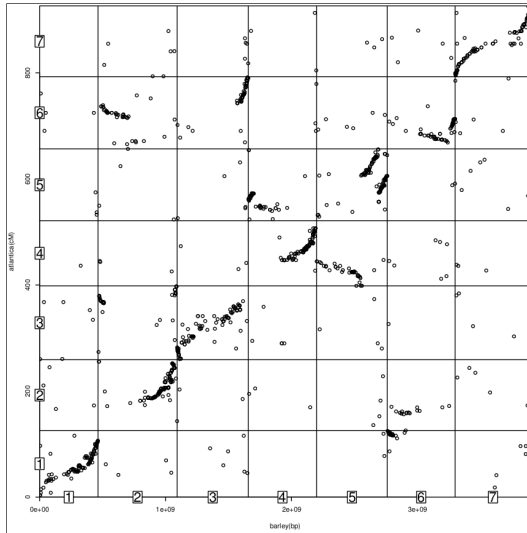


Flowering Time QTL

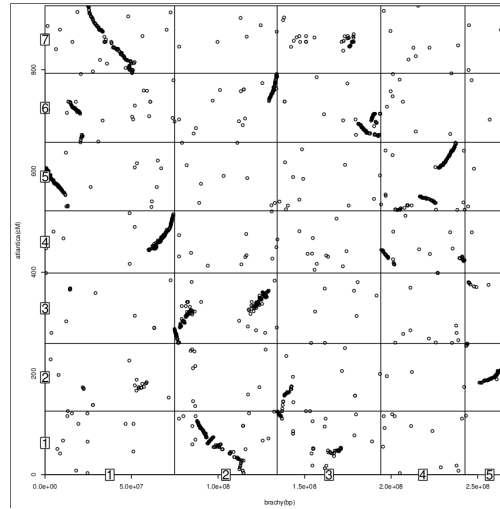


Comparison to Other Genomes

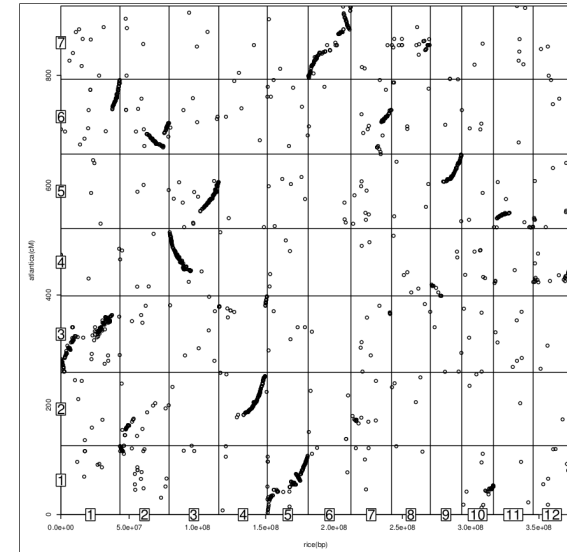
Avena atlantica



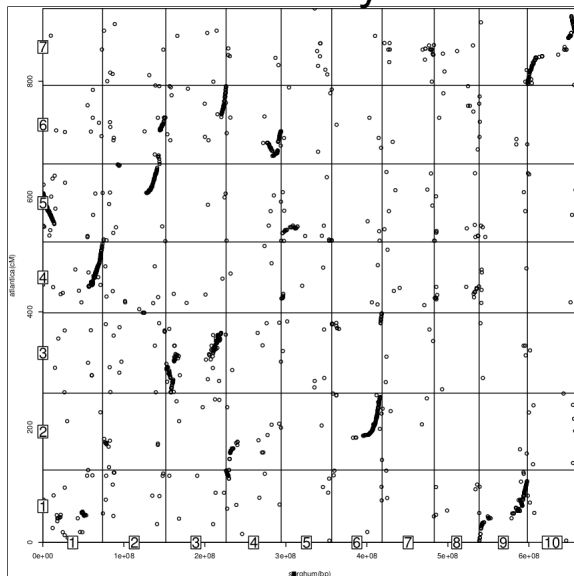
barley



brachypodium



rice

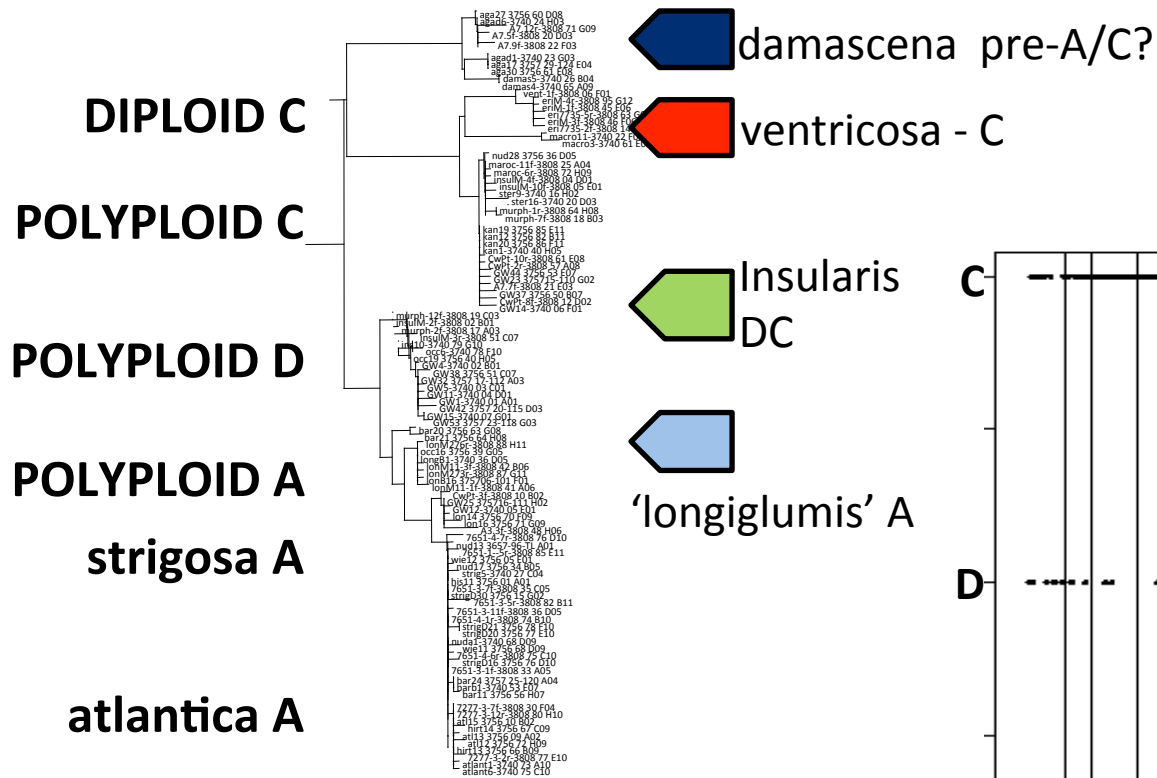


sorghum

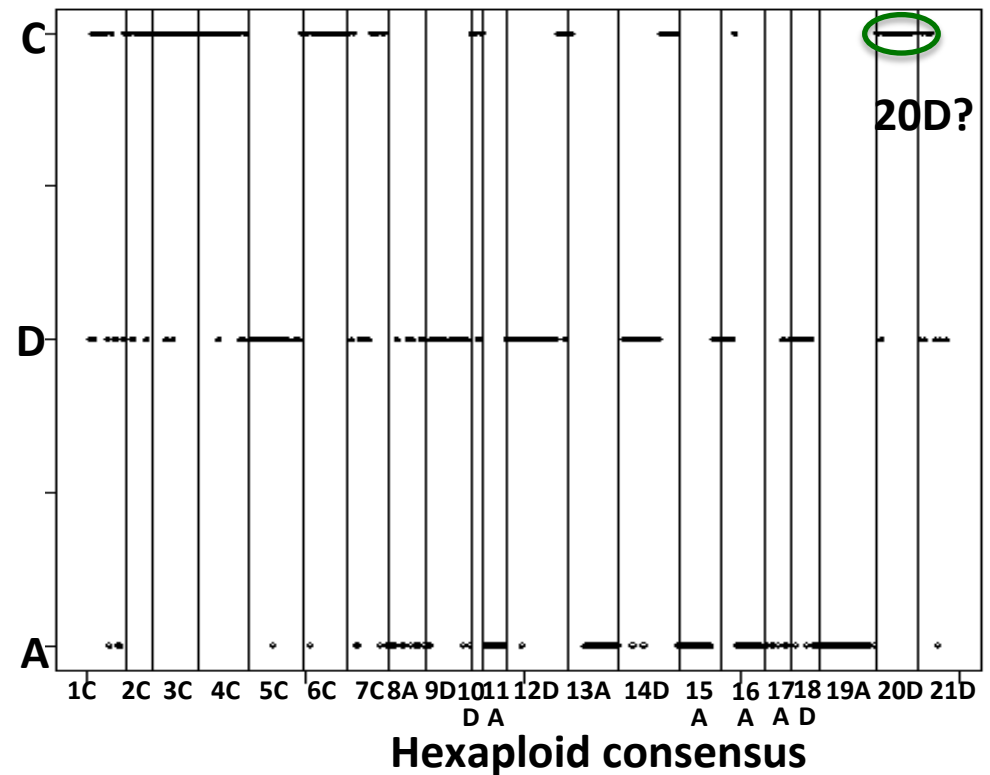
**putative orthologous genes
identified using stringent two-way
BLAST homology search**



Other 'reference' genomes?



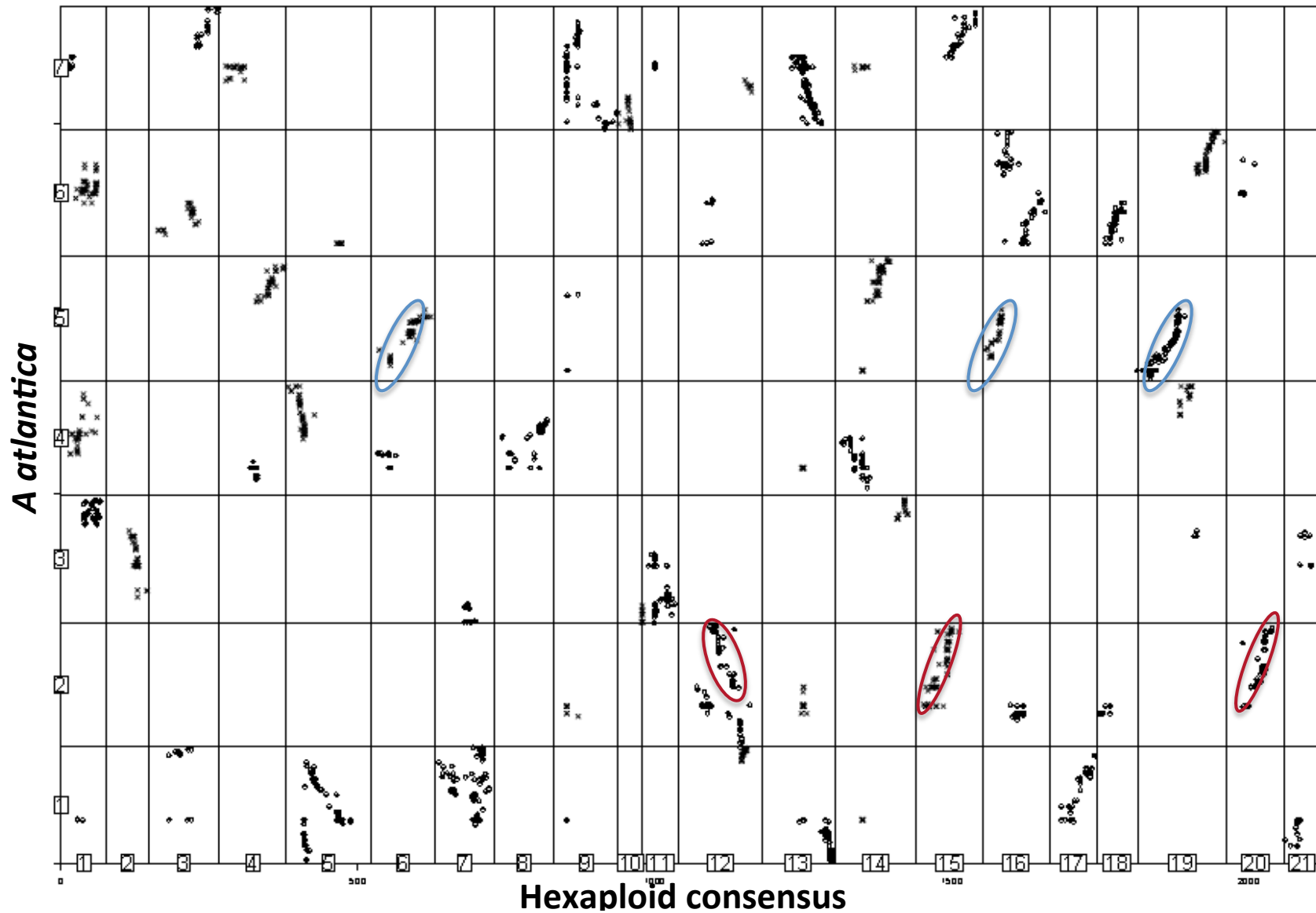
Use for "eGISH"





IBERS

Athrofa y Gwyddorau Biologol, Amgylcheddol a Gwledig



Beyond diploids- uses of the genomics reference information in the oat crop

Integration with hexaploid maps will allow rapid identification of candidate genes and development of functional markers linked to phenotype

Need to validate candidates, dissect traits and prioritise breeding targets?

Population development

- 
- **Bi-parental**
 - **Association mapping (including spring oat population and European landrace collections)**
 - **Wild relatives (diploid, tetraploid and hexaploid)**
 - **MAGIC population**
 - **Nested Association Mapping (NAM)**
 - **TILLING**
 - **QTL-NILs**
 - **Breeding programme crosses for testing/ validation of MAS and genomic selection**

Winter oat mapping family

Tardis

- conventional height
- good kernel content
- large grains
- high resistance to mildew from PC54.

Buffalo

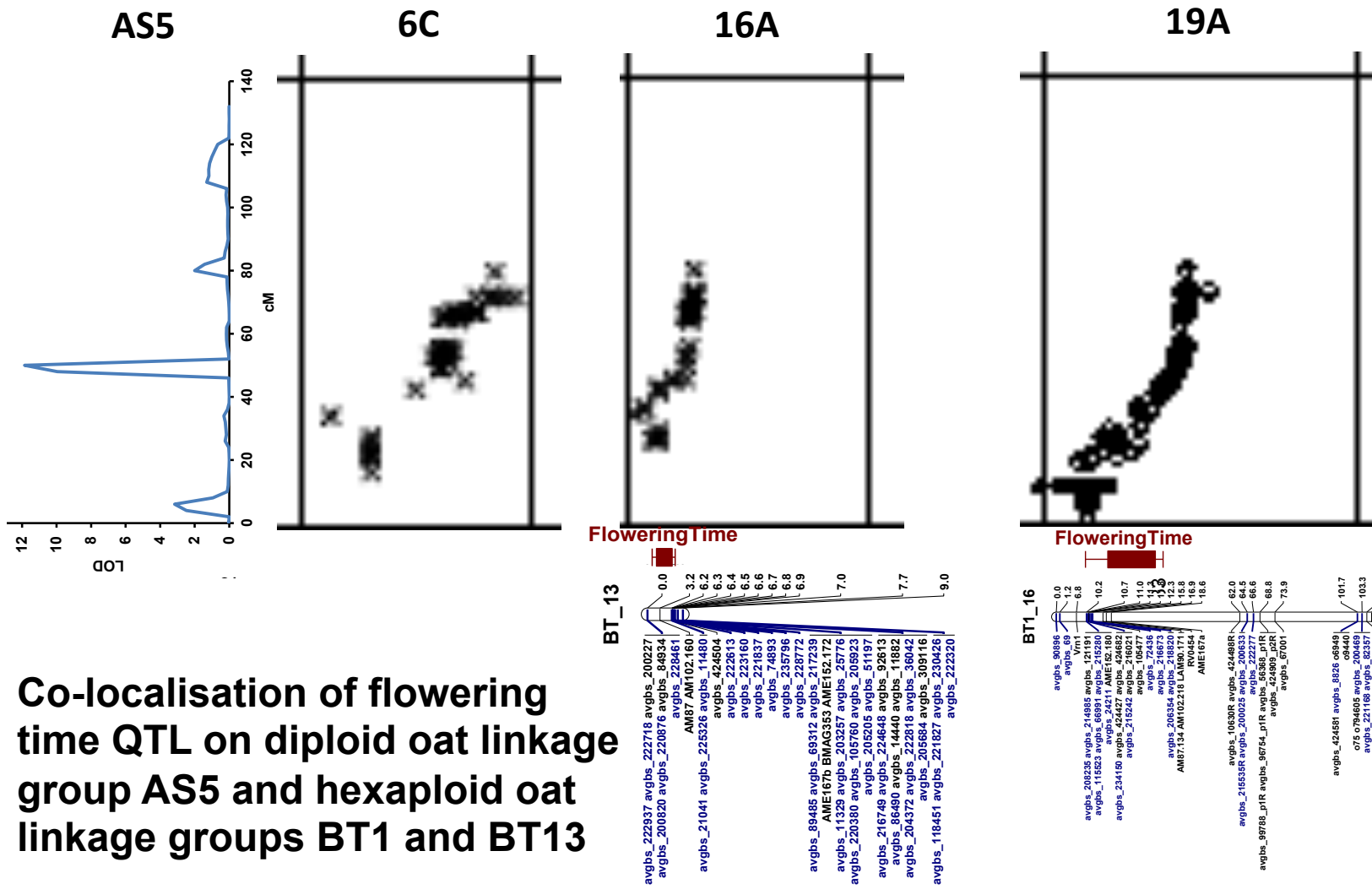
- dwarf variety
- low kernel content
- small grains
- poor resistance to mildew
- High NUpE

239 RILs genotyped with SSRs, DArT and GBS



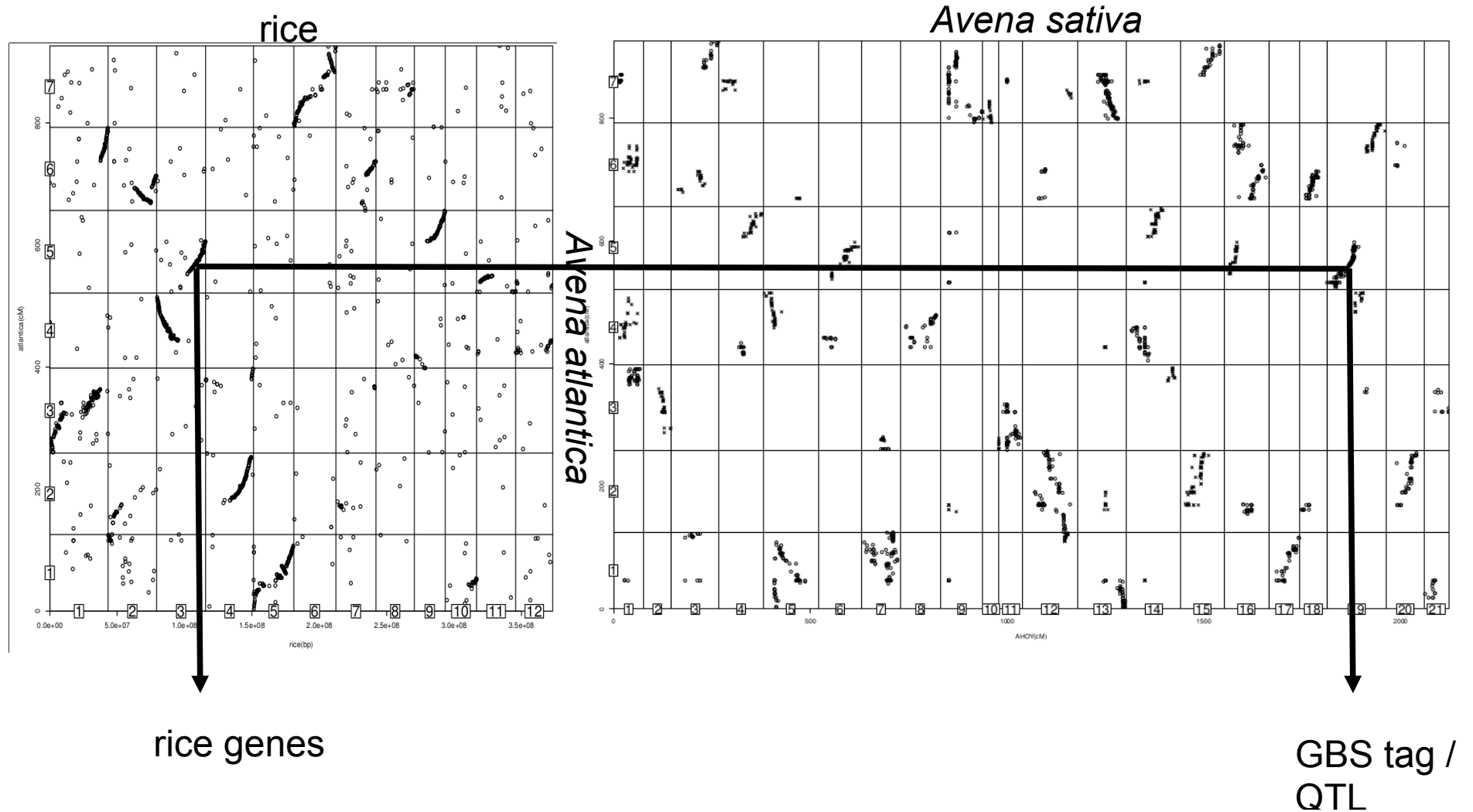
IBERS

Athrofa y Gwyddorau Biologol, Amgylcheddol a Gwledig
Institute of Biological, Environmental and Rural Sciences

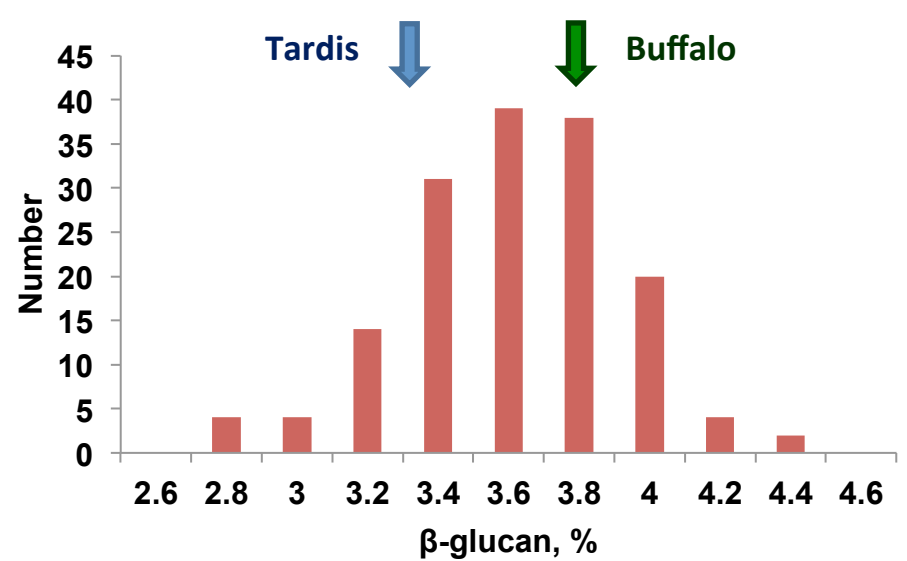
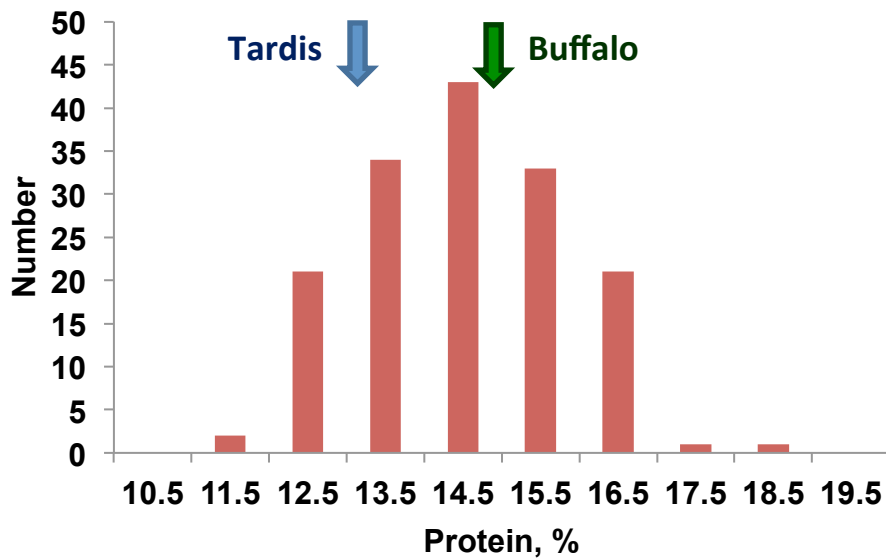
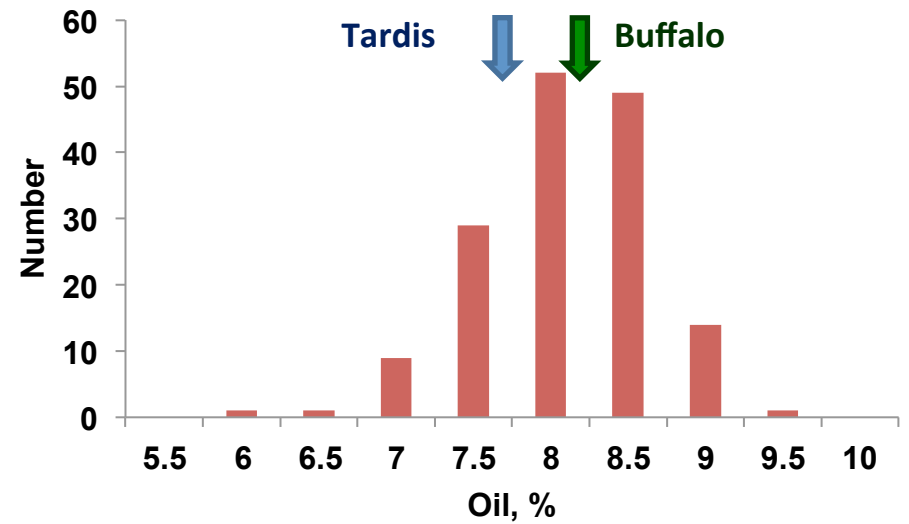
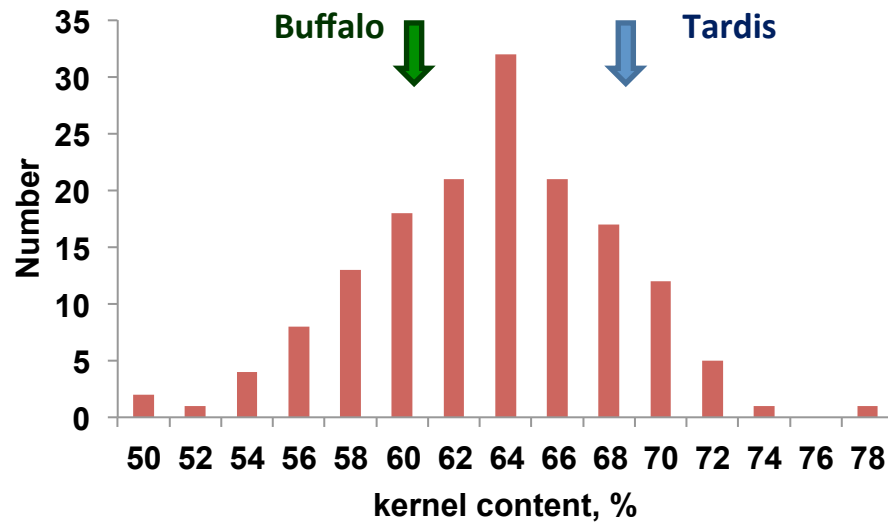


Co-localisation of flowering time QTL on diploid oat linkage group AS5 and hexaploid oat linkage groups BT1 and BT13

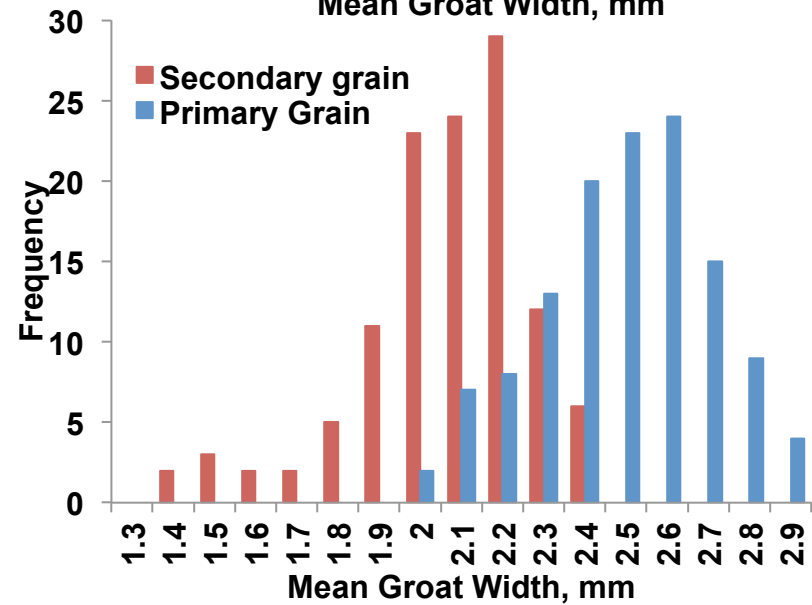
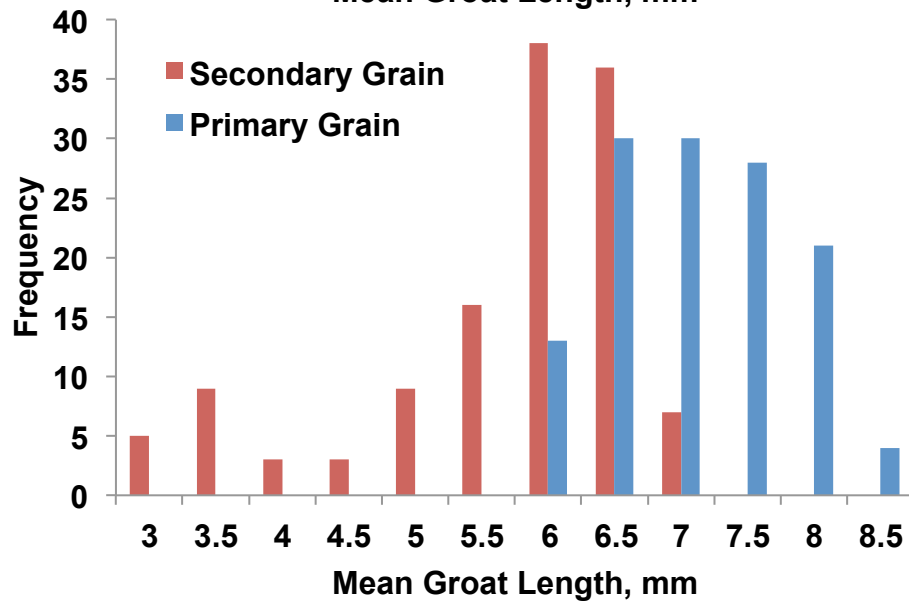
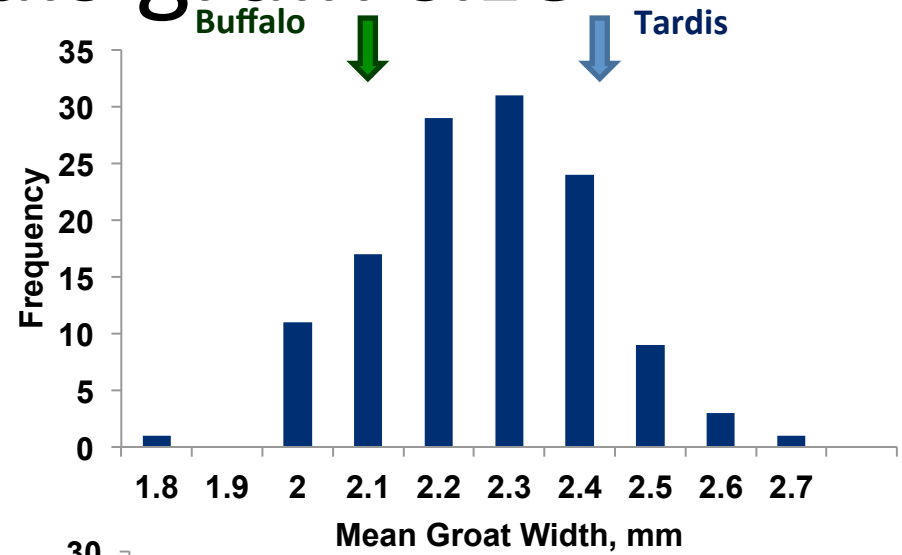
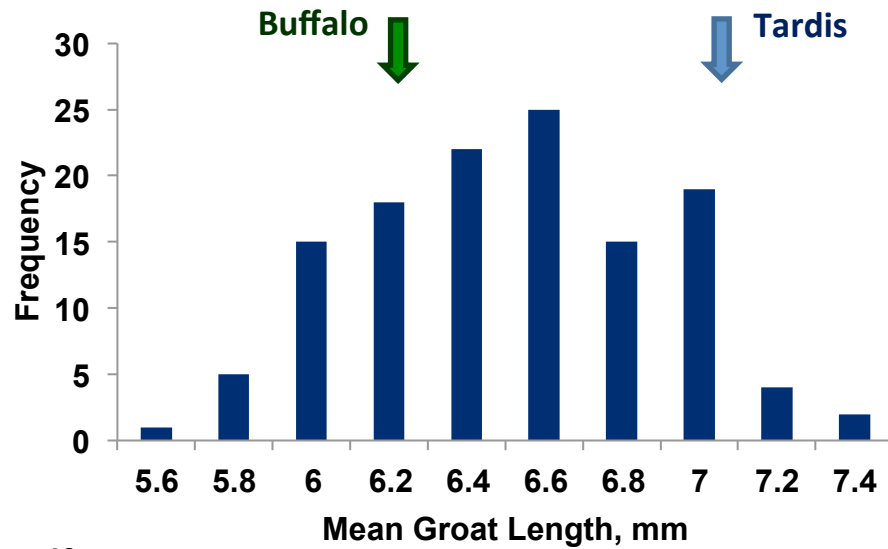
Bridging to Other Genomes



Buffalo x Tardis grain quality

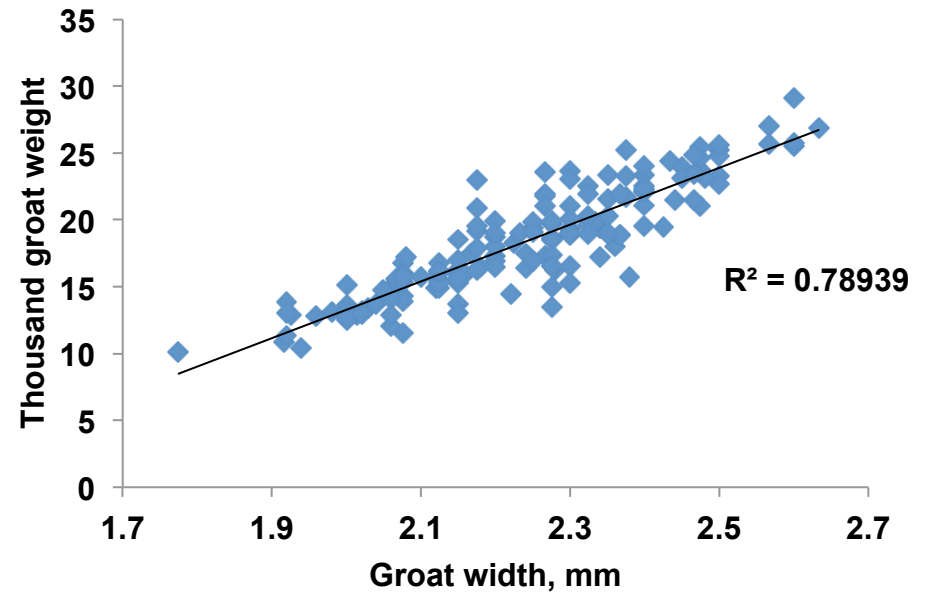
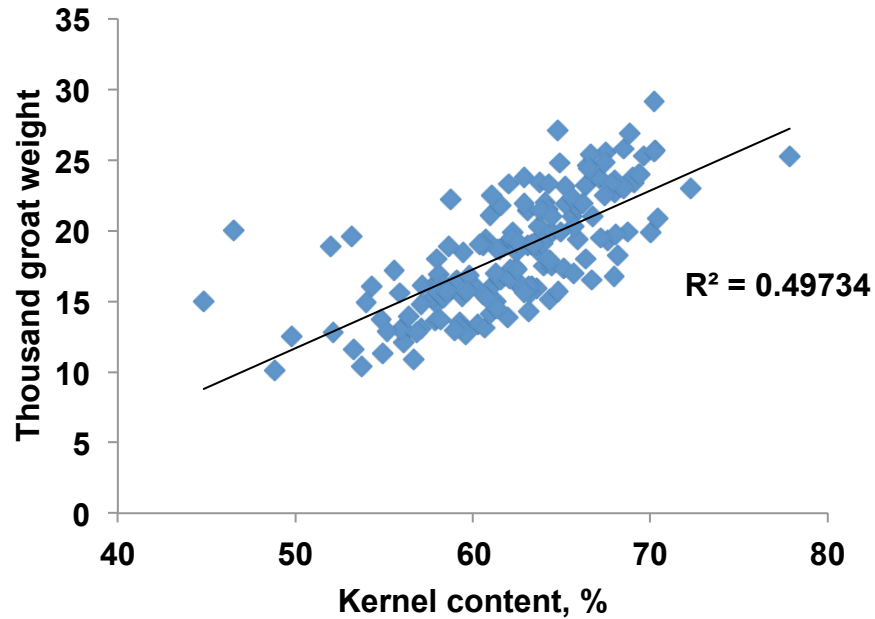


Buffalo x Tardis grain size

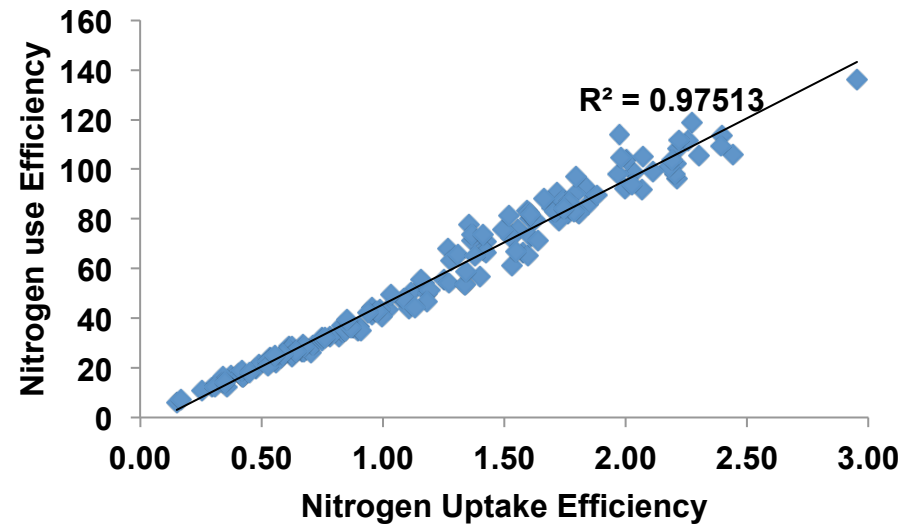
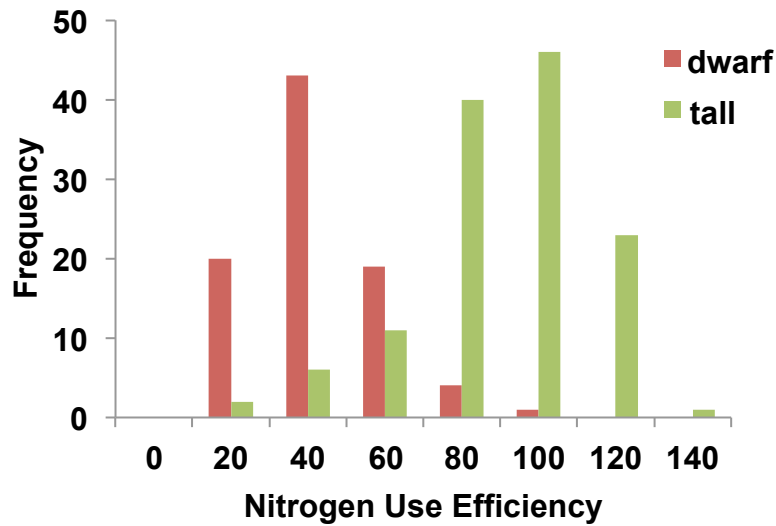
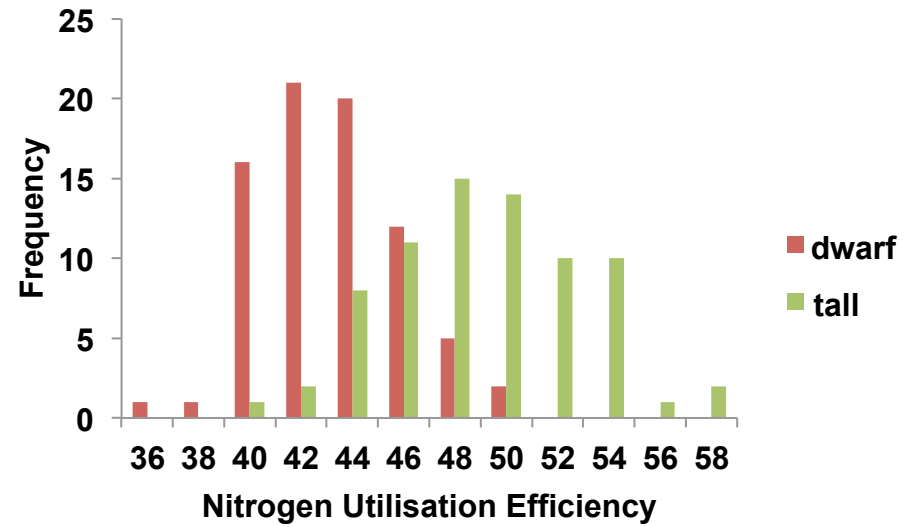
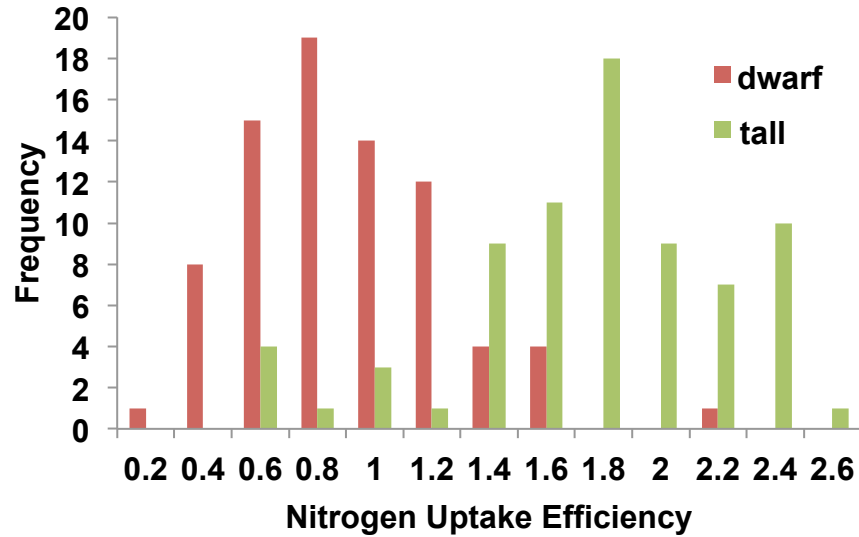




Relationship between TGW and kernel content



Buffalo x Tardis NUE



Buffalo x Tardis RIL population and MAGIC parents in phenomics centre



NPPC Capabilities:

A platform for non-destructive dynamic imaging of plant growth & development



- Conveyor based system
- c900 radio-tagged carriages
- Automated delivery to imaging stations



- Climate controlled glasshouses
- State-of-the art imaging stations
- High performance computational facilities to allow storage and retrieval of datasets
- Bio-informatics/ontology framework
- Flexible layout:- randomisation in time and place



IBERS

Athrofa y Gwyddorau Biolegol, Amgylcheddol a Gwledig
Institute of Biological, Environmental and Rural Sciences



1: Trial plots & Field Phenotyping



2: SmartHouse



3: Nutrient Flow GH



4: Photo-Physiology
a. PAM fluorimeters
b. Chlorophyll Imager
c. Gas analyser



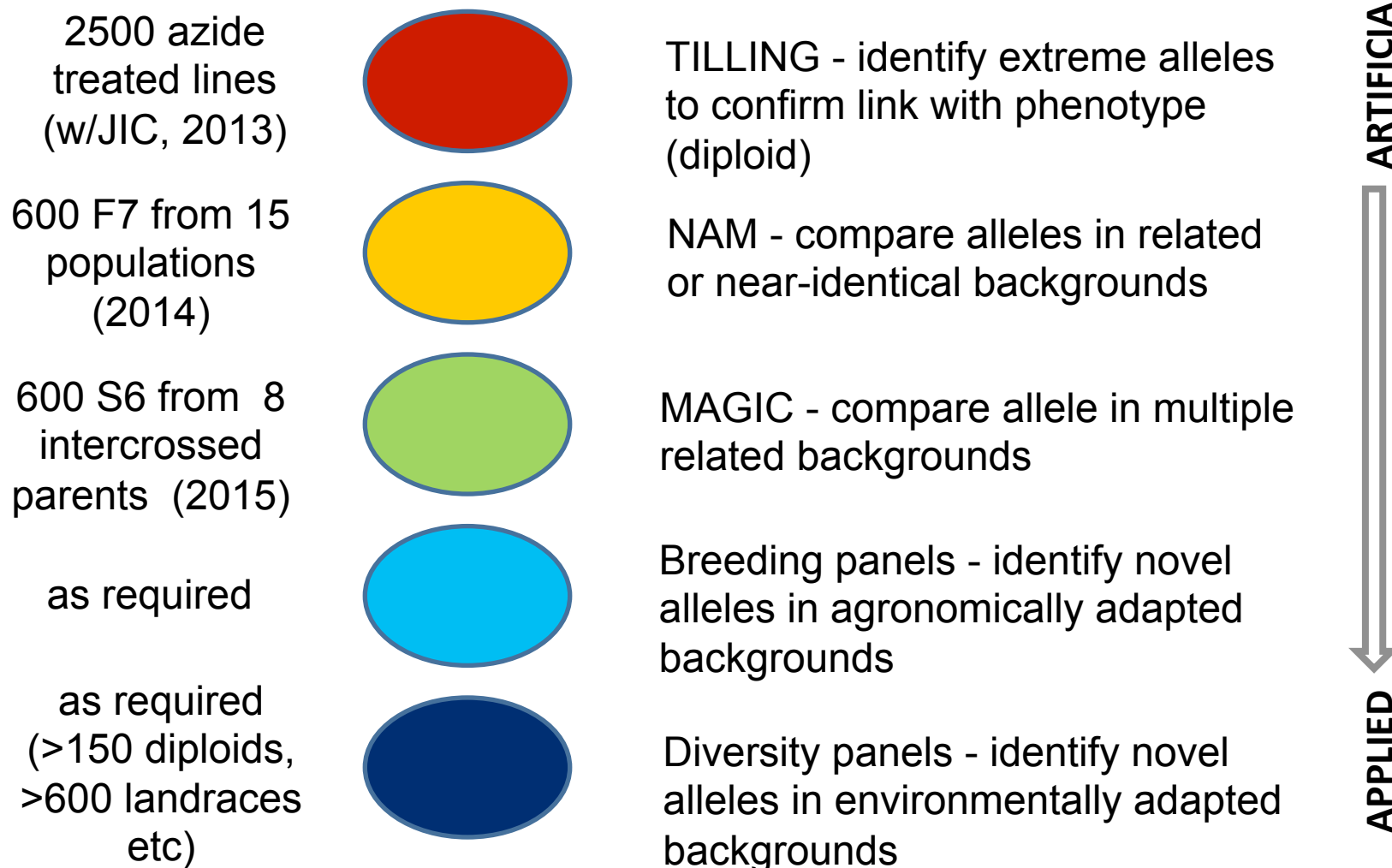
5: Organ, Cell & Molecular Analysis
a. HiSeq
b. Mass spec
c. FTIR/NIR, Micro-Raman
d. Molecular Devices
Microscopy
e. Marvin grain analyser
f. CT scanning



Integrating High Content Plant Phenotyping

**6: Bio-informatics
Data Integration and Analysis**

Genetic resources for functional genomics

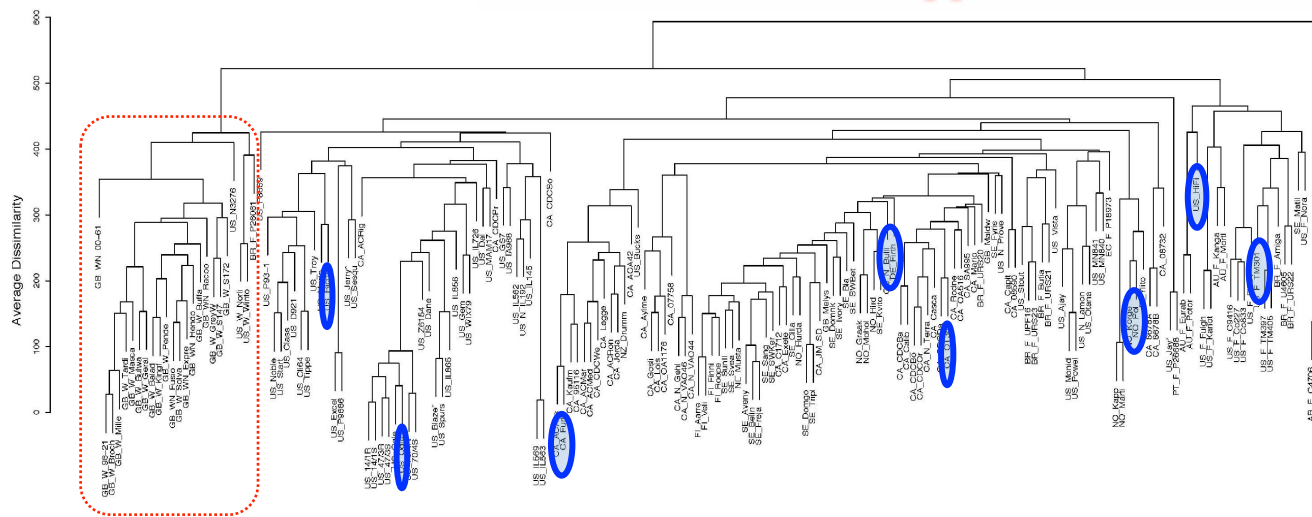
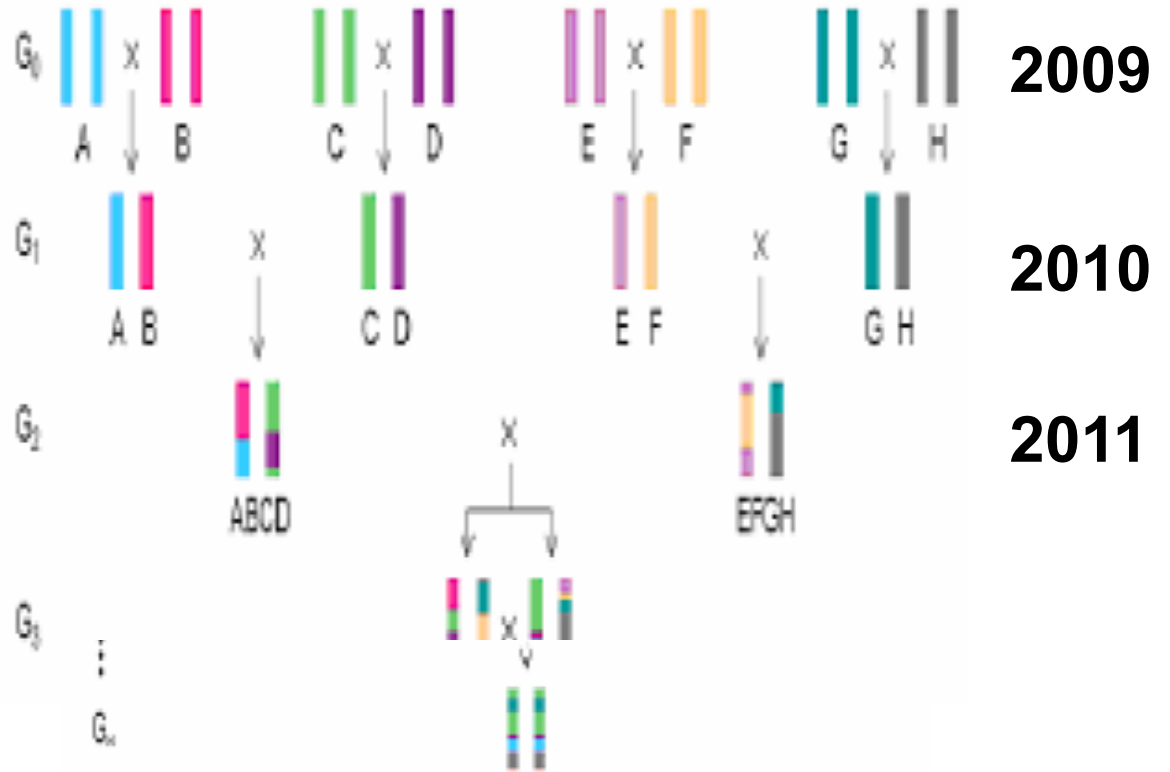




Visible phenotypes in the *A. strigosa* TILLING population (UK RevGen collaboration)

Spring oat MAGIC population development

8 spring oats chosen to sample world-wide genetic diversity (highlighted in blue in dendrogram from results from DArT analysis)



Stages in MAGIC population development

2009: 1st generation of crosses successfully completed (28 x 2 way crosses)

	1	2	3	4	5	6	7	8
	Ogle	TAM O-301	Ac Assiniboia	HiFi	CDC Dancer	Firth	Pol	CDC SolFi
1 Ogle								
2 TAM O-301	12							
3 Ac Assiniboia	13	23						
4 HiFi	14	24	34					
5 CDC Dancer	15	25	35	45				
6 Firth	16	26	36	46	56			
7 Pol	17	27	37	47	57	67		
8 CDC SolFi	18	28	38	48	58	68	78	



Female panicle selected and emasculated



Set-up cross



Male pollen donor

Progress in MAGIC population development

2009: 1st generation of crosses successfully completed (28 x 2 way crosses)

2010: 2nd generation of crossing successfully completed (28 crosses combining 4 genotypes)

2011: 3rd generation of crossing successfully completed (42 crosses combining 8 genotypes). 1st generation of single seed descent (SSD) harvested October 2011

2012: 2nd generation of SSD harvested June ; 3rd generation of SSD harvested Dec 2012

2013: 4th generation of SSD sown February 2013 and population size increased
5th generation of SSD harvested 2014

2014: seed multiplication in progress along with preliminary phenotyping and DNA extraction, ready for sowing in field 2015



9 QTL-NIL populations developed using MAS targeting introgression of key QTL into either a Buffalo or Tardis background

- Dwarf
- Flowering time/
vernalisation/ photoperiod
- Crown rust resistance
- Mildew resistance
- Grain size
- TGW



Buffalo Tardis Tardis Buffalo
 + Buffalo + Buffalo
 QTL on LG13 QTL on LG13

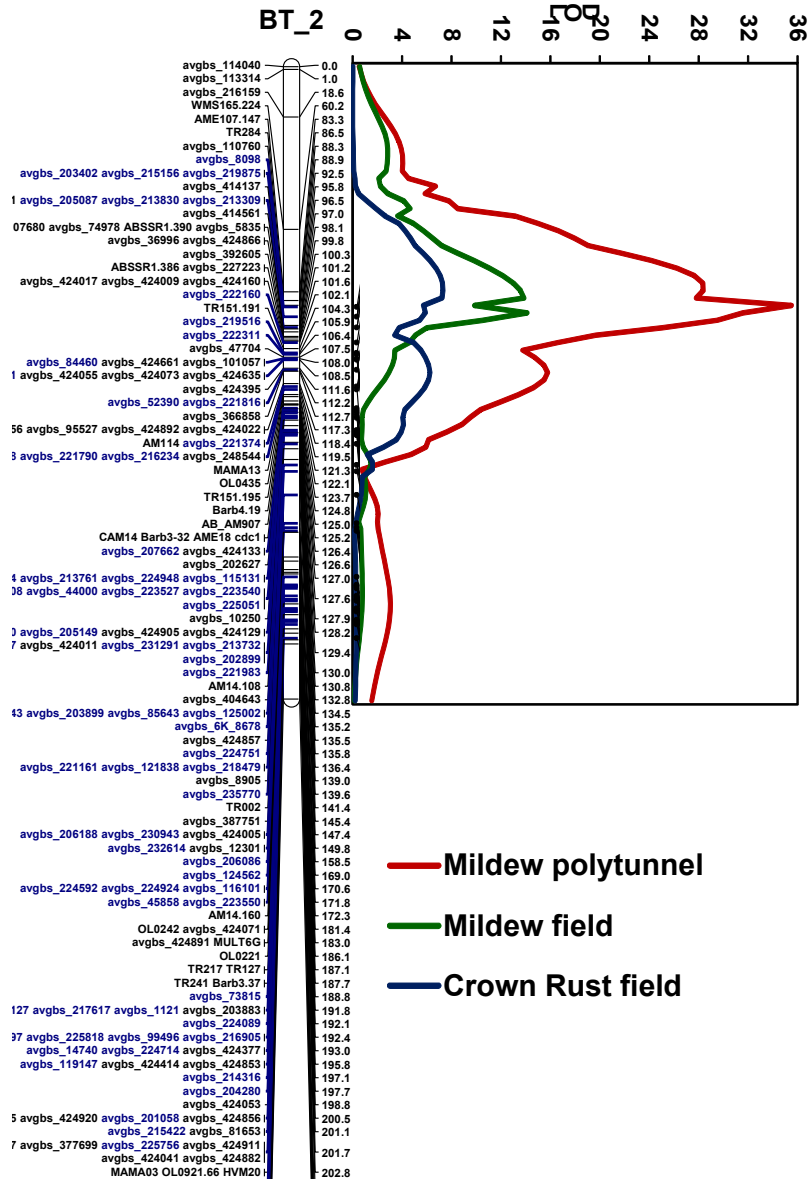
Using both foreground and background selection

How markers are being used:

- Assessment of genetic diversity of breeding programme
- Selection of appropriate parents to use in the crossing programme
- Characterising novel sources of traits
- Identifying/ verifying marker-trait associations
- Back-cross introgression of exotic alleles into a UK adapted background
- Fixing multiple disease resistance alleles at an early stage in the breeding programme
- Identification of individuals at an early stage in breeding programme containing desired allele combinations in conjunction with agronomic assessments.
- Checking uniformity of advanced breeding lines

Three general strategies for molecular breeding

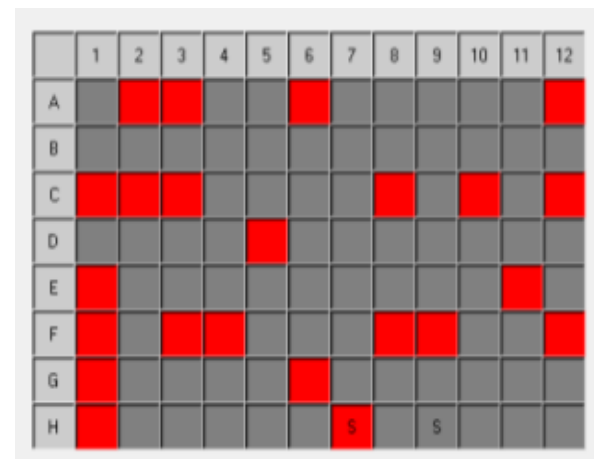
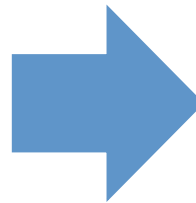
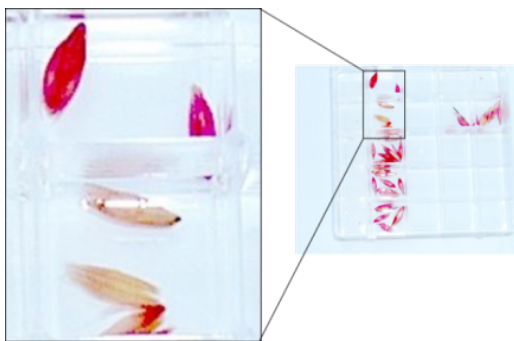
- Fix large-effect QTL from specific sources identified through bi-parental mapping
- Target medium-effect QTL with common alleles found through genome-wide association (GWAS)
- Enrich small-effect QTL through genomic selection (GS)



Example: Identification of markers linked to QTL associated with disease resistance

Diagnostic markers now routinely used in breeding programme to develop novel disease resistant varieties.

Low lignin husk is being incorporated into new varieties to improve digestibility and enhance feed value



Low lignin can be detected by colorimetric methods
-time consuming and conducted post harvest

SNP assay
Rapid, accurate and uses leaf tissue

Thanks to Pete Eckstein et al. CDC, Saskatoon

The future

- Current molecular markers used in breeding programme are in the main linked to traits conditioned by a small number of genes or for one component of a multi-gene trait
- Decreased costs of sequencing mean that whole genome analysis now a cost-effective reality
- Extend into multi-gene traits such as yield and to combine multiple traits

Acknowledgements

QUOATS (DEFRA sustainable arable LINK)

CORE

Nick Tinker, Jesse Poland

Web site <http://www.QUOATS.org>

