Oat Genomics at IBERS
Minor crop – major genome

- 17Gb, 2n=6x
- 13Gb, 2n=6x
- 5.1Gb, 2n=2x
- 3.3Gb
- 0.43Gb
Major genome – minor budget...

> $100M

< $1M
Why genomics?

Aka Why does this take so long and cost so much?

A. Easy things

B. Avena
“Non-genic” – too variable for a reference?

Comparison of maize ‘bronze’ region
Random samples of polymorphisms will often target non-genic areas (repeat derived?)

- Need a reference map of genes to align to marker maps
Synteny – conserved gene content
Synteny – conserved gene content*

*may contain artefacts
Genome Analysis

Gene Content and Virtual Gene Order of Barley Chromosome 1H}{1C}[W][OA]

Klaus F.X. Mayer, Stefan Taudien, Mihaela Martis, Hana Šimková, Pavla Suchánková, Heidrun Gundlach, Thomas Wicker, Andreas Petzold, Marius Felder, Burkhard Steuemagel, Uwe Scholz, Andreas Graner, Matthias Platzer, Jaroslav Doležel, and Nils Stein*

Munich Information Center for Protein Sequences/Institute for Bioinformatics and Systems Biology, Helmholtz Zentrum Munich, German Research Center for Environmental Health, 85764 Neuherberg, Germany (K.F.X.M., M.M., H.G.); Leibniz Institute for Age Research, Fritz Lipmann Institute, 07745 Jena, Germany (S.T., A.P., M.F., M.P.); Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, 77200 Olomouc, Czech Republic (H.S., P.S., J.D.); Institute of Plant Biology, University of Zurich, CH-8008 Zurich, Switzerland (T.W.); and Leibniz Institute of Plant Genetics and Crop Plant Research, 06466 Gatersleben, Germany (B.S., U.S., A.G., N.S.)
Genome Analysis

Gene Content and Virtual Gene Order of Barley Chromosome 1H[1C][W][0A]

Klaus F.X. Mayer, Stefan Taudien, Mihaela Martis, Hana Šimková, Pavla Suchánková, Heidrun Gundlach, Thomas Wicker, Andreas Petzold, Marius Felder, Burkhard Steuemagel, Uwe Scholz, Andreas Graner, Matthias Platzer, Jaroslav Doležel, and Nils Stein*

Munich Information Center for Protein Sequences/Institute for Bioinformatics and Systems Biology, Helmholtz Zentrum Munich, German Research Center for Environmental Health, 85764 Neuherberg, Germany (K.F.X.M., M.M., H.G.); Leibniz Institute for Age Research, Fritz Lipmann Institute, 07745 Jena, Germany (S.T., A.P., M.F., M.P.); Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, 77200 Olomouc, Czech Republic (H.S., P.S., J.D.); Institute of Plant Biology, University of Zurich, CH-8008 Zurich, Switzerland (T.W.); and Leibniz Institute of Plant Genetics and Crop Plant Research, 06466 Gatersleben, Germany (B.S., U.S., A.G., N.S.)
**Avena zipper - Genetic Mapping**

**Diploid species**

A. strigosa  
A. atlantica

\[ \text{A. strigosa} \times \text{A. atlantica} \]

\[ \text{F}_1 \]

selfing

\[ \text{186 F}_6 \]

individuals (RILs)

\[ \text{SSRs and 3000 'DArT-seq' codominant markers, 92 RILs} \]

\[ \text{dense genetic map} \]

2772 markers
Avena zipper - NGS based mapping anchor contigs based on parent-of-origin in RILs

Using barcodes for specific recombinant inbred lines (RILs) will allow chromosomal 'bin' assignments of mate-pair assemblies. The parental origin of specific regions of a single chromosome is indicated above for nine RIL progeny (most recombination events occur in early generations, leading to relatively large contiguous segments with the same parental origin). 3kb mate-pair templates, barcoded to identify the RIL of origin, may be used for further assembly of initial parental HiSeq contigs. Sufficient coverage of each RIL for >1 read per assembled contig will also yield addresses which will be rare or unique depending on the number and choice of RILs.
Figure A. Methods used to place sequences on the genetic map.

(A) Placement of scaffolds. Out of a total of 849314 scaffolds 137187 (16.2%) are placed by at least one method.

(B) Placement of high confidence gene models. Of 30723 genes, 28916 (94.1%) are placed by at least one method.
Transcriptome database, 11 tissues assembled and annotated
Synteny and lineage-specific evolution of key genes

Synteny plot of reciprocal best blastp hits, showing base pair position in (A) brachypodium, (B) rice, (C) sorghum and (D) barley versus genetic map position in A. atlantica with positions of candidate CsLF genes and their likely ortholog(s) shown with black circles. Orthology was based on gene phylogeny.
Additional reference genomes (>20x cover)

Lagurus outgroup

Provides full Avena gene set, comparative resource for functional analysis, proxies for hexaploid sub-genomes
Hexaploid sequences can be seen to be derived from divergent sub-genomes – software cannot always resolve into correct haplotypes

- Diploid progenitors help software
- Ideally physical sorting of templates?
Illumina HiSeq reads mapped to FISHIS R3 UPL contig

RNAseq (Firth developing grain)

FISHIS (Firth)

R8 – C-genome
R9 – A-genome
R3 – D-genome
Histogram of fluorescence intensity (flow karyotype) obtained after flow-cytometric analysis of 4’,6-diamino-phenylindole stained chromosomes of Morex.

FISHIS - Fluorescent *In Situ* Hybridisation in Suspension

(A) (ACT)5-fitc FISH on metaphase spread

(B) Unlabelled chromosome suspension

(C) GBS tag content of FISHIS fraction R4

---

Debora Giorgi and Sergio Lucretti, ENEA, Rome
Expected synteny of R8, R9 and R3 fractions
Illumina HiSeq reads mapped to FISHIS R3 UPL contig

RNAseq (Firth developing grain)

FISHIS (Firth)

R8 – C-genome
R9 – A-genome
R3 – D-genome
Assemblies helped significantly by using diploid sequence as guide
Novel population development; NAM (Nested Association, common single parent)

Common parent, Firth, European spring

F7s, F8s of 600 lines grown in field
GbS mapping underway
POPSEQ?

Ethiopian tetraploid
GENE CONTENT & EXPRESSION

ATLANTICA ZIPPER

FIRTH FISHIS

NAM Grain & meristem

GENOME ASSEMBLIES & CONTIG MAPPING

ATL

ATLANTICA ZIPPER

FIRTH FISHIS

FIRTH

X

X

strig

DIVERSE

COMPARATIVE RESOURCES

damascena canariensis longiglumis

VENTRICOSA CONTIGS

AVENEAE SPP

GBS dBs
ATLANTICA ZIPPER
VENTRICOSA CONTIGS
damascena
canariensis
longiglumis
FIRTH FISHIS
FIRTH FISHIS
FIRTH FISHIS
FIRTH FISHIS
GRAIN
ATLANTICA
11 tissues
ATLANTICA
11 tissues
ATLANTICA
11 tissues
ATLANTICA
11 tissues
ATLANTICA
11 tissues
ATLANTICA
11 tissues
ATLANTICA
11 tissues
ATLANTICA
11 tissues
ATLANTICA
11 tissues
ATLANTICA
11 tissues
ATLANTICA
11 tissues
ATLANTICA
11 tissues
NAM
Grain &
meristem
ATLANTICA ZIPPER
 ATL x STRIG
FIRTH FISHIS
FIRTH FISHIS
FIRTH FISHIS
FIRTH FISHIS
FIRTH X DIVERSE
GENE CONTENT & EXPRESSION
GENOME ASSEMBLIES & CONTIG MAPPING
COMPARATIVE RESOURCES
AVENEAE SPP
GBS dBs
Phenotype
Genetic resources for functional genomics

2500 azide treated lines (w/JIC, 2013)

TILLING - identify extreme alleles to confirm link with phenotype (diploid)

600 F7 from 15 populations (2014)

NAM - compare alleles in related or near-identical backgrounds

600 S5 from 8 intercrossed parents (2015)

MAGIC - compare allele in multiple related backgrounds

as required

Breeding panels - identify novel alleles in agronomically adapted backgrounds

as required (>150 diploids, >1500 landraces etc)

Diversity panels - identify novel alleles in environmentally adapted backgrounds

Genetic resources for functional genomics

ARTIFICIAL

APPLIED
DNA stocks from viable and dead germplasm

IBERS ‘landrace’ panels
Identify likely progenitor populations
Identify ‘adaptations’ missing from ‘progenitors’
Environmental adaptation or population structure?
IBERS resources funded by public/private collaboration, linked with other public funded researchers and projects - Aim to integrate with T3, metabolite dBs, et al
IBERS

Rob Vickerstaff - Bioinformatics
Catherine Howarth - Mapping, model populations, trait dissection
Maciej Bisaga – RNAseq, current LINK project
Sandra Prusaka, Chris Creavey – transcriptome database
Adriana Ravagnani – Pathway annotation
Matt Hegarty - Genomics facility
Athole Marshall - Head of Public Good Plant Breeding

ENEA

Debora Giorgi, Anna Farina, Sergio Lucretti – FISHIS

CRA

Primetta Faccioli – miRNA analysis

AAFC Ottawa

Nick Tinker – GbS

Laval

Brian Boyle – GbS

UNCC

Schlueter lab – hexaploid references