International
Wheat Genome
Sequencing
Consortium

The IWGSC Bread Wheat Chromosome Survey Sequencing Initiative

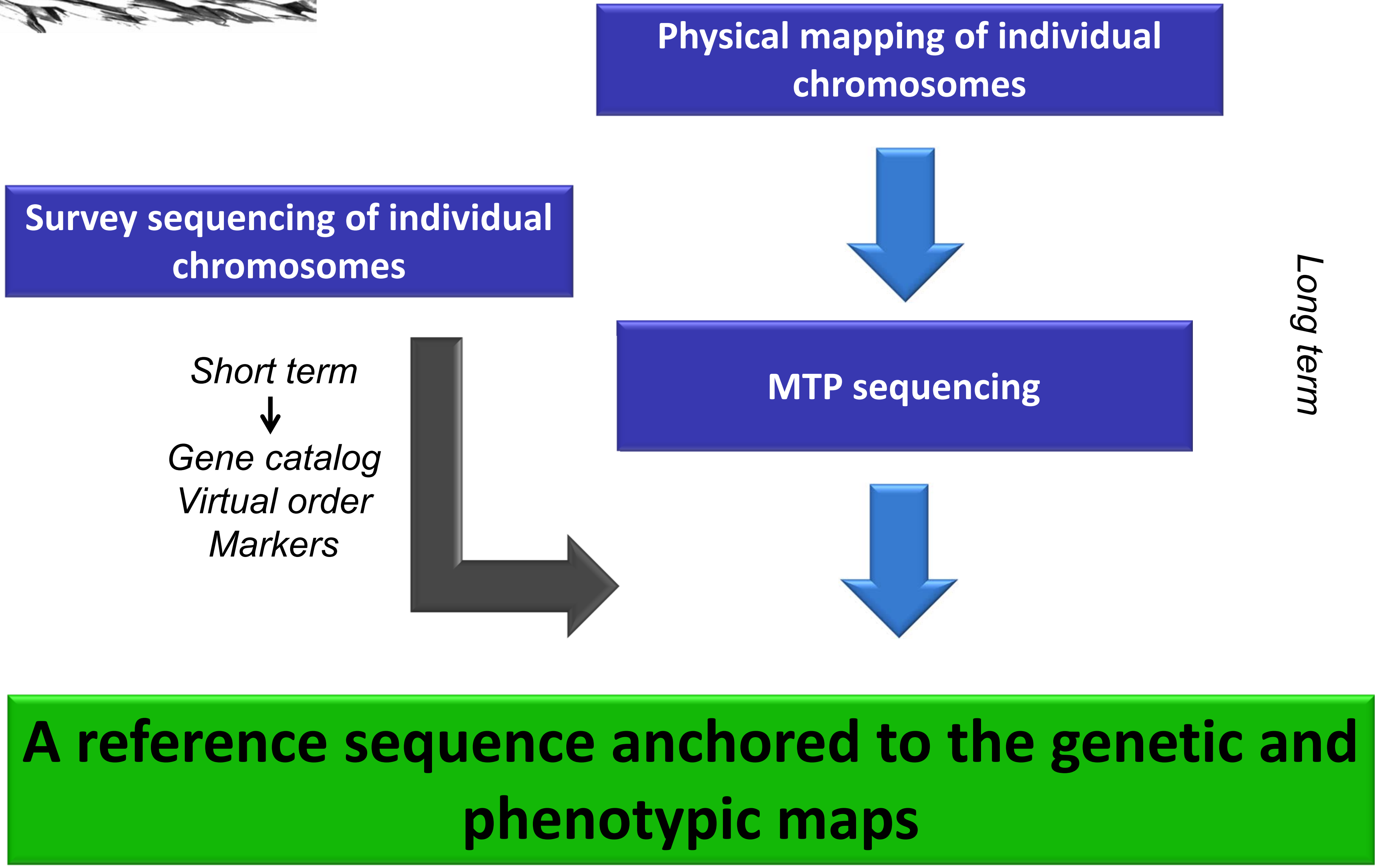
Jane Rogers

On behalf of the IWGSC

Plant and Animal Genomes
San Diego January 2014

Roadmap to the Wheat Genome Sequence

IWGSC



Physical mapping of individual chromosomes

Survey sequencing of individual chromosomes

Short term

↓

Gene catalog

Virtual order

Markers

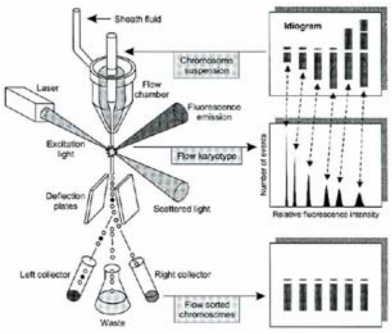
MTP sequencing

Long term

A reference sequence anchored to the genetic and phenotypic maps

Chromosome Survey Sequence Data

Institute of Experimental Botany of the AS CR, v. v. i.



Flow sorted chromosomes

Amplified DNA

Illumina Paired End Sequencing
30-200x

Sequence Assembly (ABySS)
1 per chromosome arm + 3B

TGAC The Genome Analysis Centre™

University of Saskatchewan

Kansas State University

The University of Queensland Australia

Genoscope

NIAS National Institute of Agrobiological Sciences

TGAC The Genome Analysis Centre™

Graminor

biogemma

INRA

BBSRC bioscience for the future

Punjab Agricultural University

ICARDA

Chromosome Survey Sequence Data

- Quality checked sequence reads (100-150bp)
- Sequence assemblies for 40 chromosome arms + chromosome 3B
  - Total length 10.2 Gb
  - 128Mb (1DS) – 639Mb (3B) assembled sequence per chromosome
  - contigs / chromosome arm range from 88k (6DS) – 508k (2DL)
  - N50 contig length = 2.4 kb (1.5kb - 4.3kb) after filtering contigs < 200bp
  - N50 contig length after repeat masking = 6.1kb (1.7kb-8.9kb)

Gene Calling

HelmholtzZentrum münchen
German Research Center for Environmental Health

Chromosome arm assemblies

- Chromosome arm sorted (excepting 3B): **41 sequence “bins”**
- 128Mb (1DS) – 639Mb (3B) assembled sequence per “bin”
- repeat masked
- Total sequence: **10.2Gb**

RNA-seq (E. Paux, INRA)

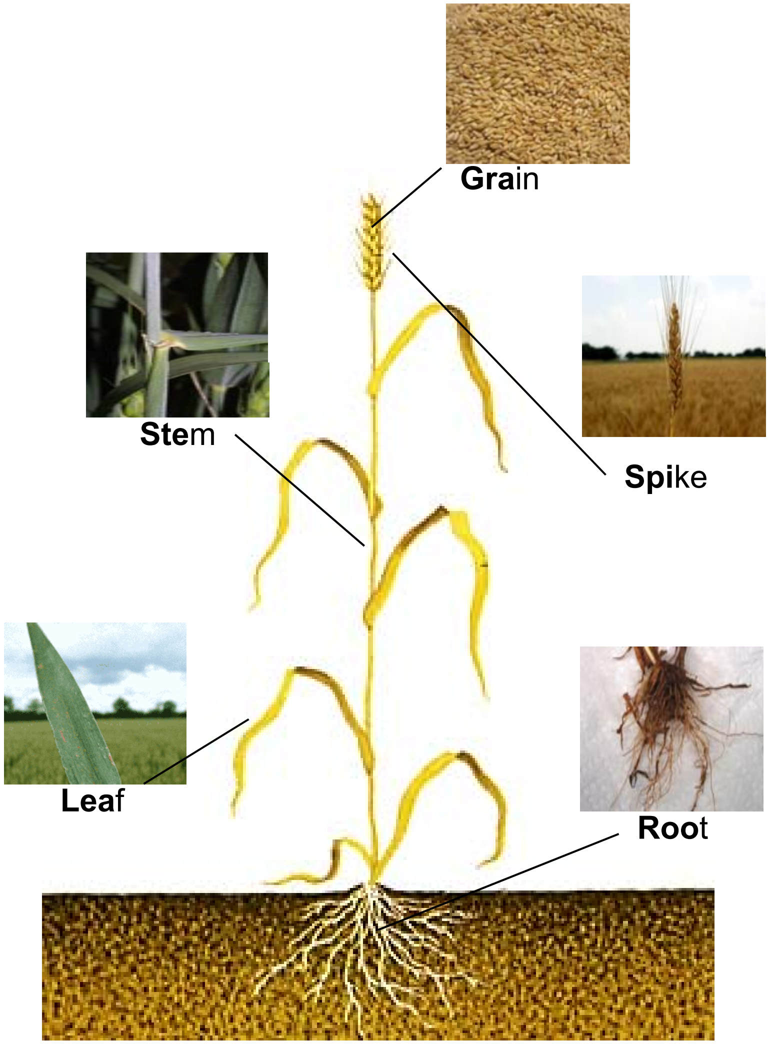
- 5 tissues
- 112 – 140mio reads per samples
- Single end, 101bp length
- 12- 14Gb sequence
- **∑615.3mio reads; 62Gb sequence**

Public wheat fl-cDNAs (TriFLDB)

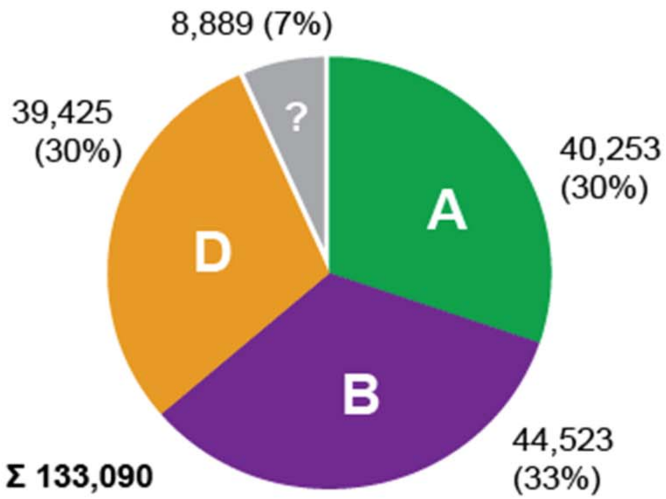
- 16,807 fl-cDNAs (16,607 with ORF)
- 15,245 fl-cDNAs - assigned to chromosome arms via *CARMA* (→1:1 relationship)

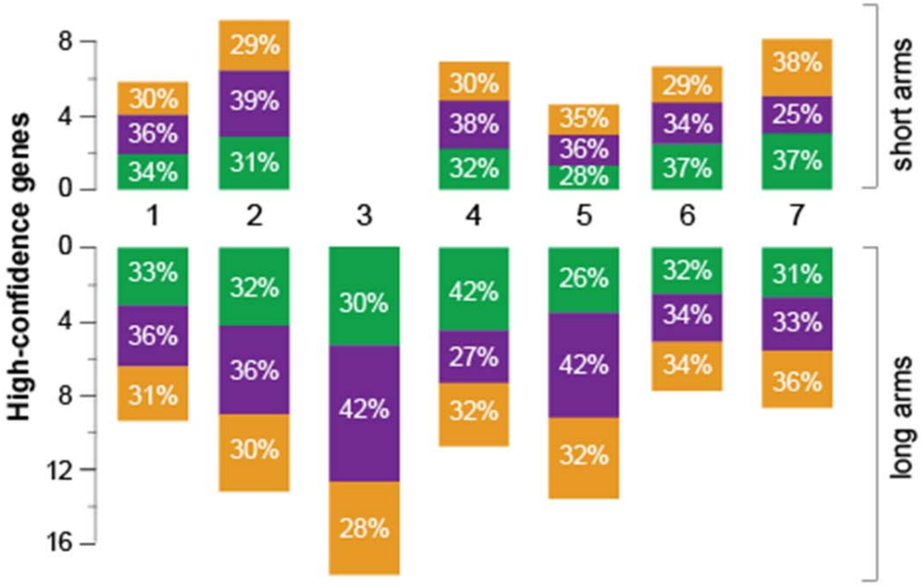
Reference annotations

- Barley (IBSC 2012), Brachypodium, Rice, Sorghum



Gene distribution across sub-genomes and chromosomes

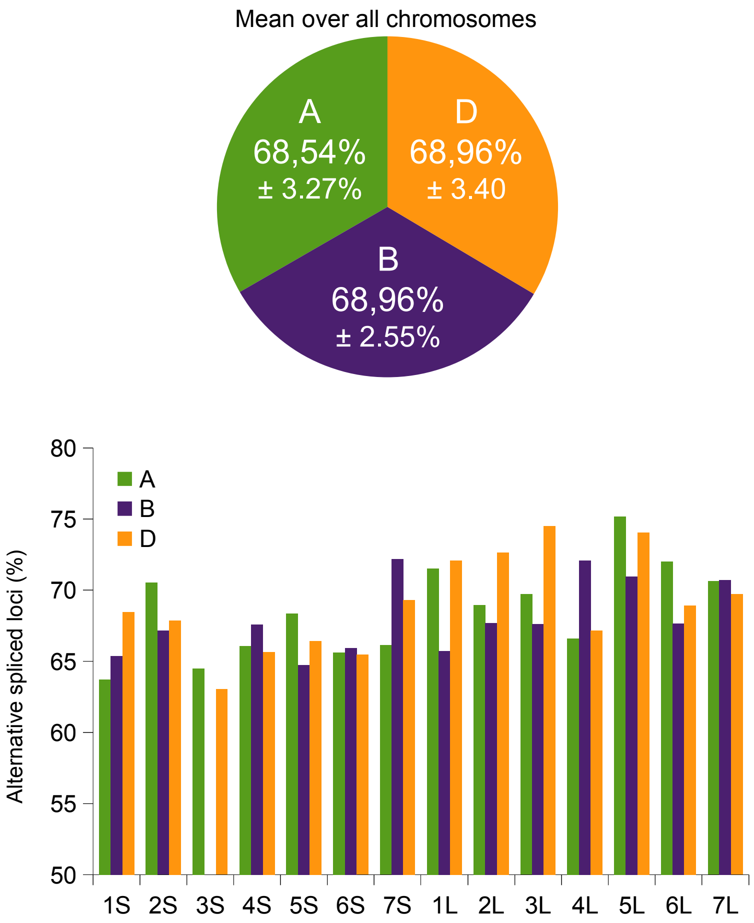




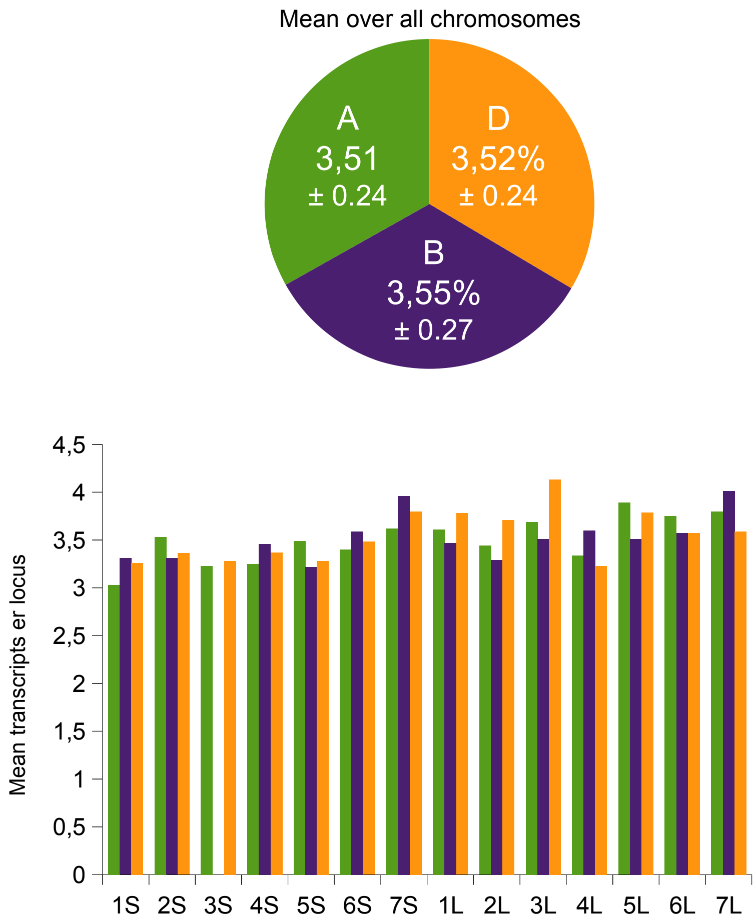
- 133,090 HC gene calls – 124,201 allocated to chromosomes
- Higher proportion of genes on the B genome
- Gene content but no order

Alternative splicing per chromosome and genome

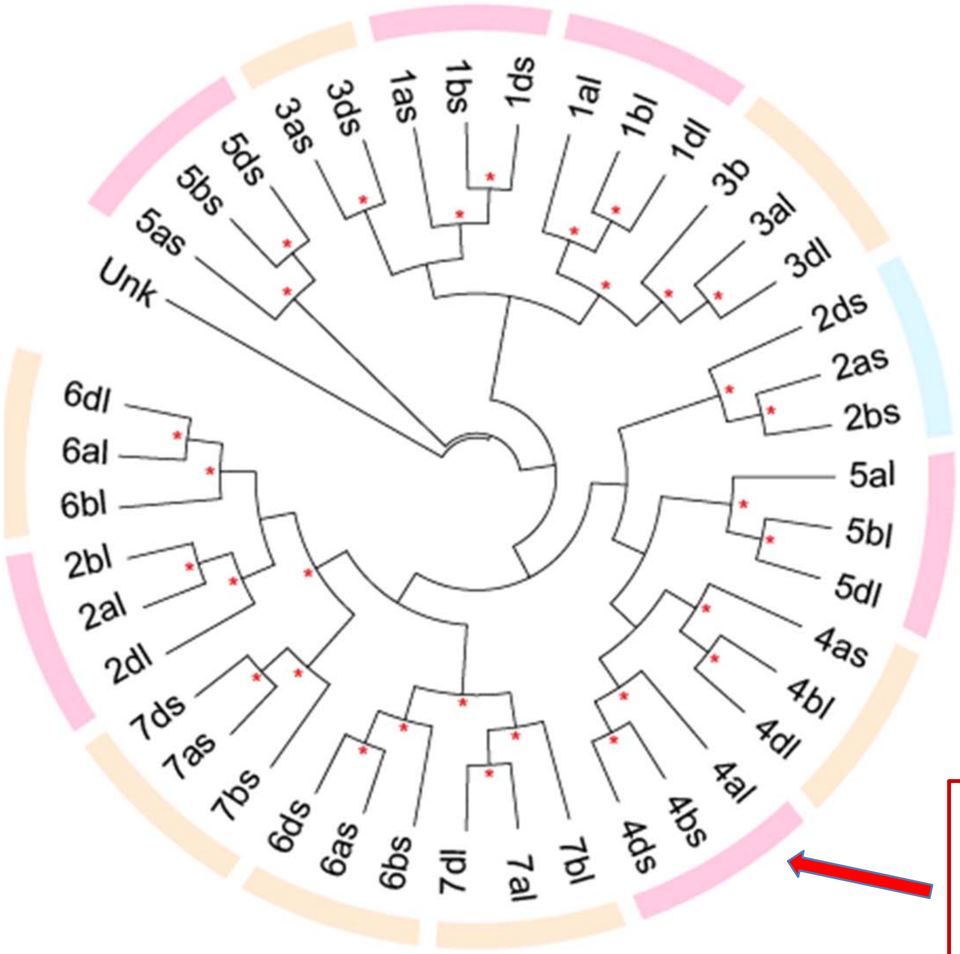
Alternative spliced loci



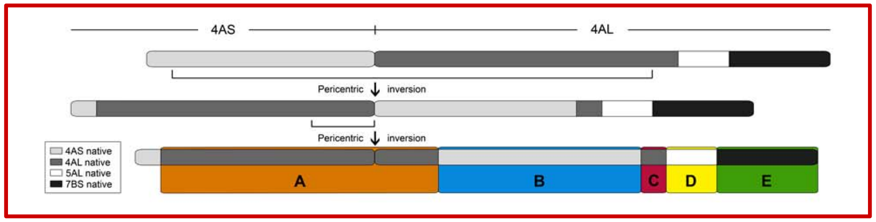
Predicted transcripts per locus



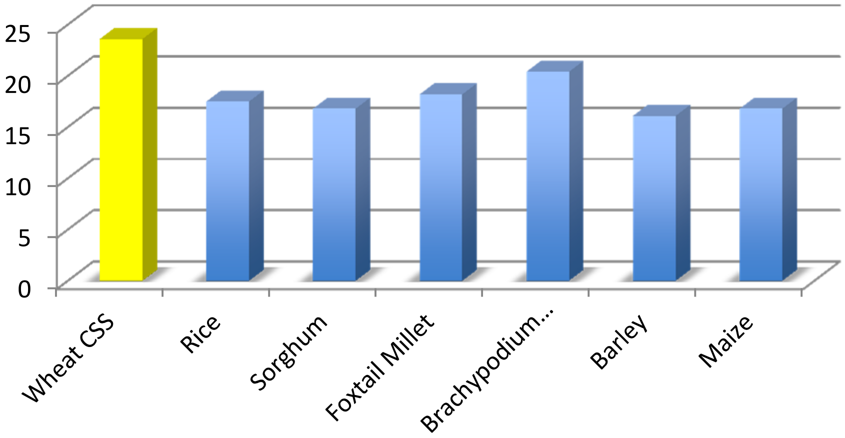
Hierarchical clustering of genes among homoeologous chromosomes



Genes represented in each of the sub-genomes cluster with genes on homoeologous chromosomes:
every topology is found

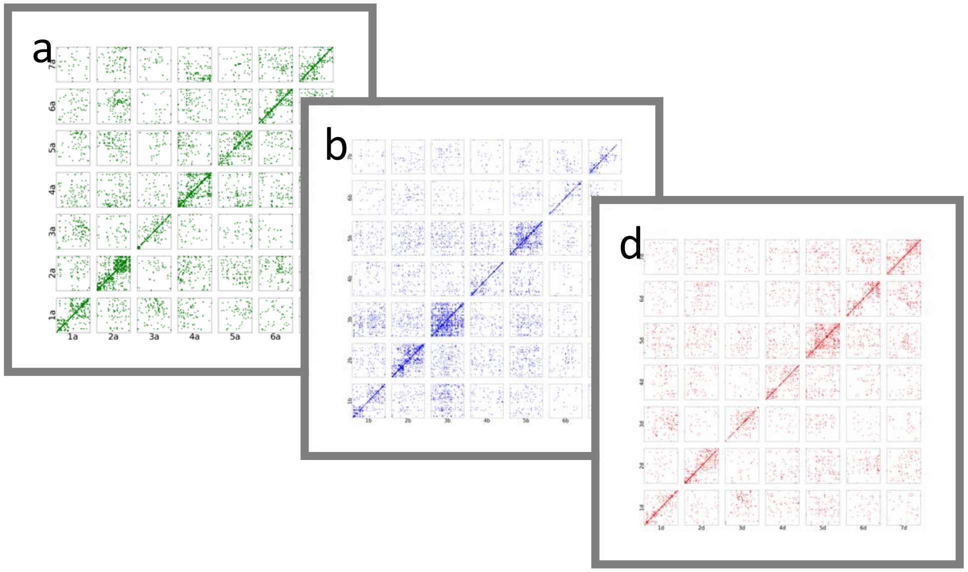


Gene duplications are abundant in wheat



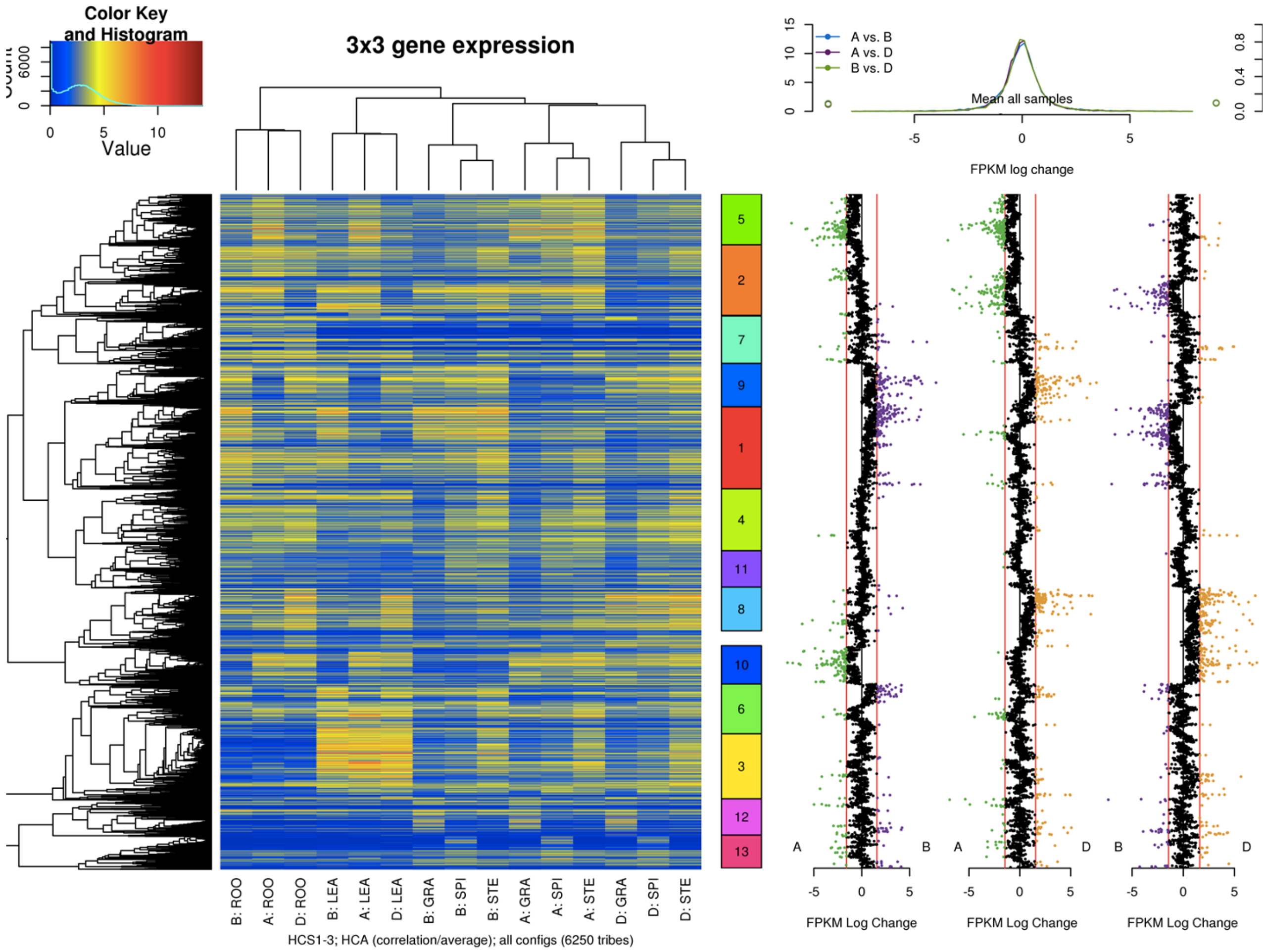
Gene duplication in grass genomes

Duplications in wheat underestimated due to sequence fragmentation



73% of duplicates are located on the same chromosome arm

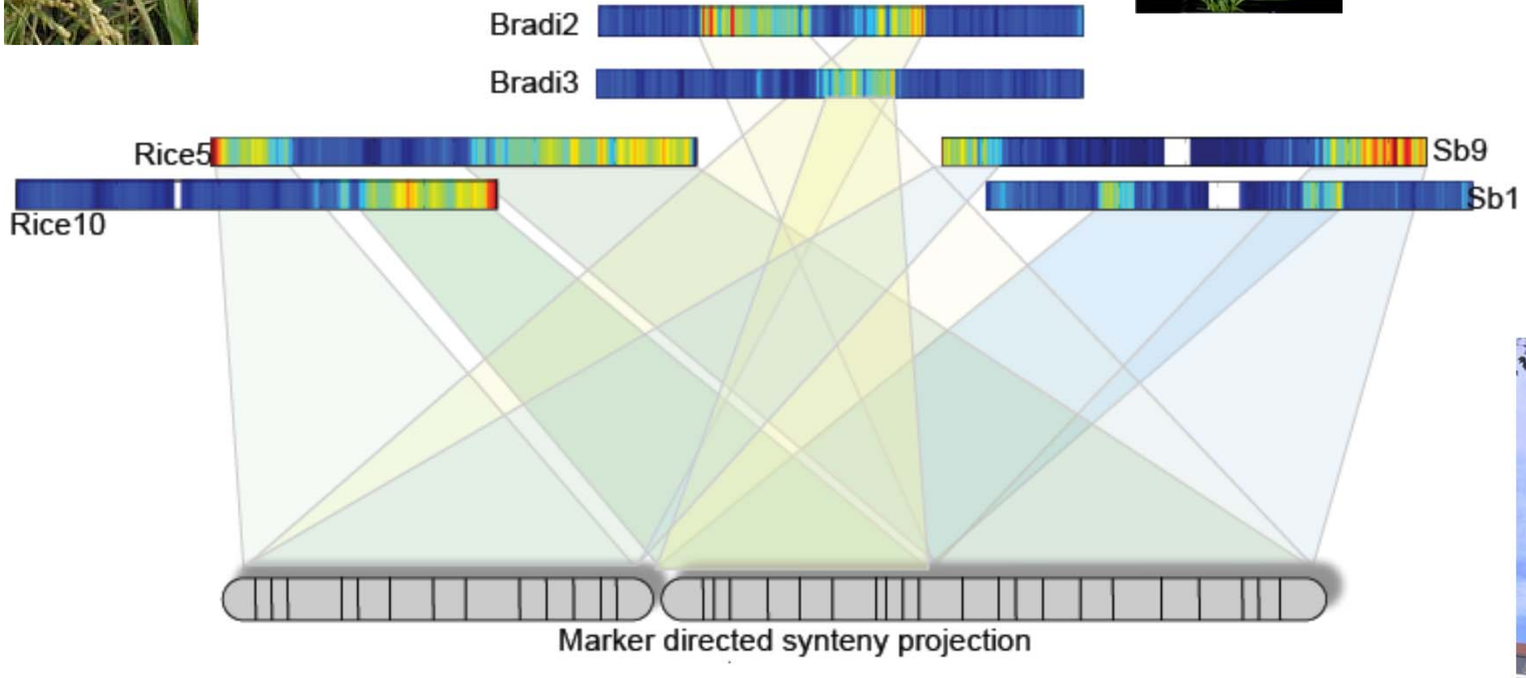
Differential gene expression in individual selected tissues



See Matthias Pfeiffer – Sunday workshop

Syntenic alignment and integration generates a ‘GenomeZipper’

HelmholtzZentrum münchen
German Research Center for Environmental Health



Genome zipper generates a virtual gene order for the wheat genome

Data Sets	1AS	1AL	2AS	2AL	3AS	3AL	4AS	4AL	5AS	5AL	6AS	6AL	7AS	7AL	∑
No. of markers	1123	947	1780	1075	786	1033	544	2669	2205	1079	1257	804	1534	1373	**18209**

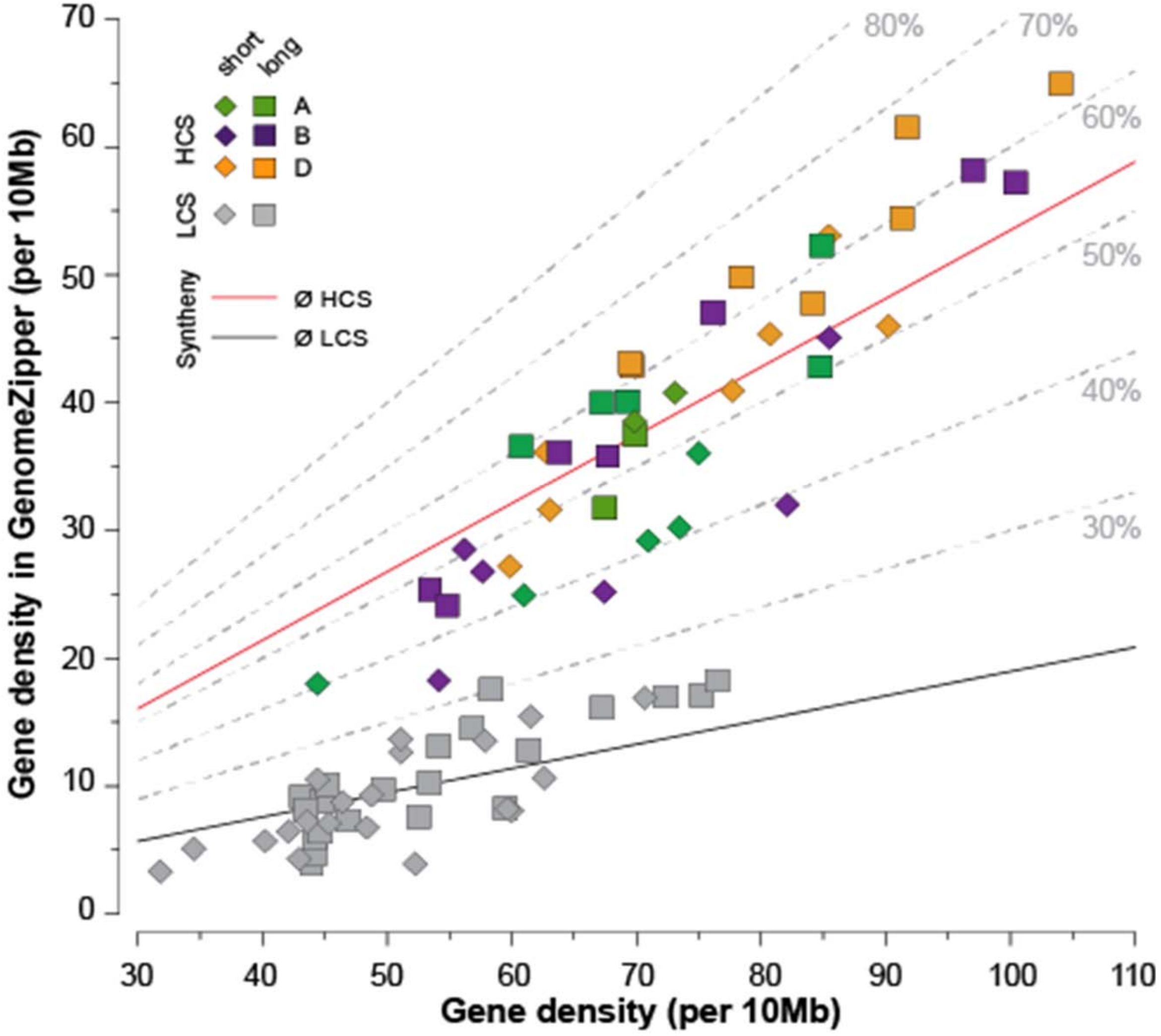
Data Sets	1BS	1BL	2BS	2BL	3B	4BS	4BL	5BS	5BL	6BS	6BL	7BS	7BL	∑
No. of markers	1668	1136	1459	2606	5064	1321	673	813	2858	1355	1030	869	2323	23175
No. of Bd genes	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Data Sets	1DS	1DL	2DS	2DL	3DS	3DL	4DS	4DL	5DS	5DL	6DS	6DL	7DS	7DL	∑
No. of markers	914	669	806	1516	515	1186	194	450	638	1146	1040	630	1492	955	12151
No. of Bd genes	423	1147	895	1567	634	1246	571	1123	392	1724	462	949	855	907	12895
No. of Os genes	295	1000	808	1451	514	1201	475	1031	307	1256	354	878	836	747	11153
No. of Sb genes	329	1083	774	1318	489	1164	483	1102	280	1467	380	916	856	815	11456
No. of non-redundant anchored gene loci in genomeZipper	611	1762	1380	2385	880	1845	869	1583	561	2544	676	1349	1435	1356	**19236**

64,000 markers used (GBS; J. Poland and Eduard Akhunov)

19,000 -20,000 genes per sub-genome zipped

Synteny-based chromosomal ordering of genes



On average 53% of HC genes are virtually ordered

Genome zipper available at MIPS and at URGI

Bread Wheat vs. di-and tetraploid ancestors

Triticum monococcum

- Illumina
- Assembly (MIPSv3):

Triticum urartu

- Illumina
- Assembly (TGACv1)

Aegilop. speltoides

- Illumina
- Assembly (TGACv1)

Aegilops sharonensis

- Illumina
- Assembly (B. Steuernagel)

Triticum durum (CN)
Triticum durum (IT)

- Illumina
- Assemblies (TGAC v1)

Aegilops tauschii

- Assembly (BGI): 4.1Gb (N50: 50kb)

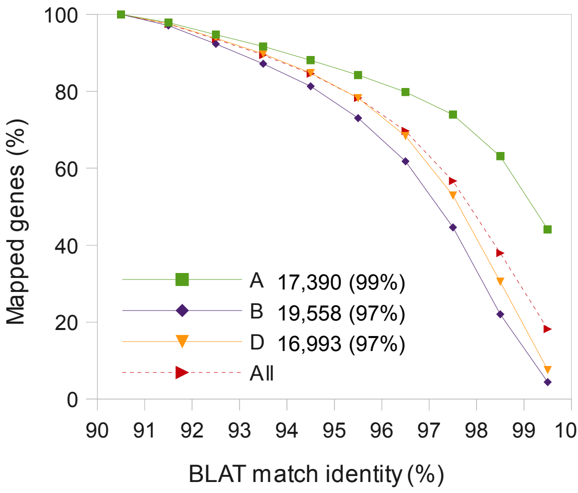
Tm	Triticum monococcum
Tu	Trititcum uruatu
Td	Triticum durum
Asp	Aegilops speltoides
Ash	Aegilops sharonensis
At	Aegilops tauschii
Ta	Triticum aestivum

Bread Wheat vs. Diploid ancestors

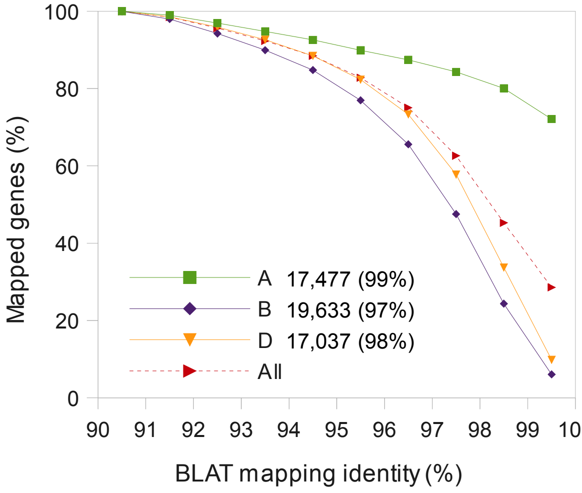
-alignment identity of wheat genes vs. genome assemblies of A,B and D donors-

AA

vs. *Triticum monococcum*

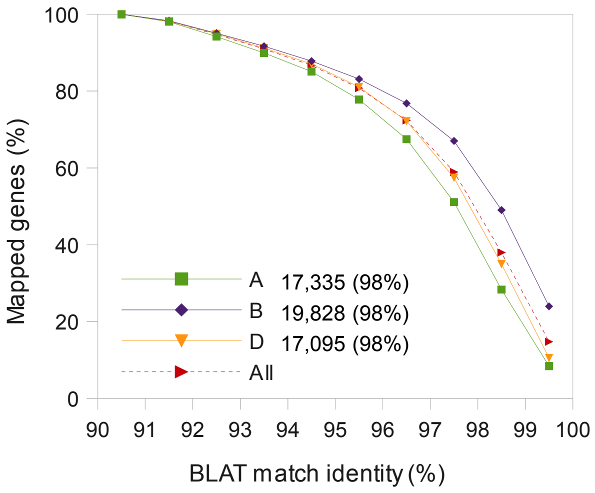


vs. *Triticum urartu*

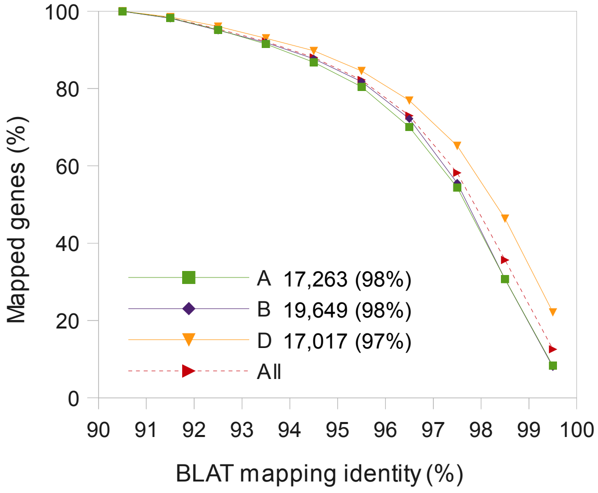


BB

vs. *Aegilops speltoides*

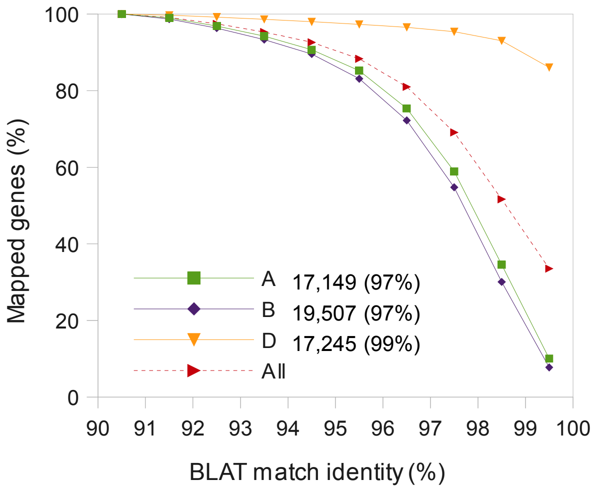


vs. *Aegilops sharonensis*



DD

vs. *Aegilops tauschii*



- BLAT: HCS 1 CDS vs. Assemblies
- Best match
  - i. max coverage
  - ii. max match identity

17,635 HCS1 A
20,144 HCS1 B
17,470 HCS1 D

55,249 HCS1

Summary

- Almost full wheat gene complement identified and allocated to chromosome arms
- On average, 53% of genes virtually ordered along chromosomes
- High level of inter- and intrachromosomal duplication

- Over 3.5 M markers mapped to contigs (1.3M wheat markers + 2.3M SNPs) - SSR, EST, DArT, SNP (90k) markers...

These resources are facilitating:

- New phylogenetic analyses of wheat genome evolution
- Definition of core and pan gene sets for *Triticeae* by comparing di-tetra- and hexaploid genomes
- Homoeologue-specific gene expression analyses
- Wheat haplotype map – E. Akhunov
- TILLING projects in *T. durum* and bread wheat – UC Davis, JIC,RRes
- Gene identification and cloning

Chromosome Survey Sequence Access

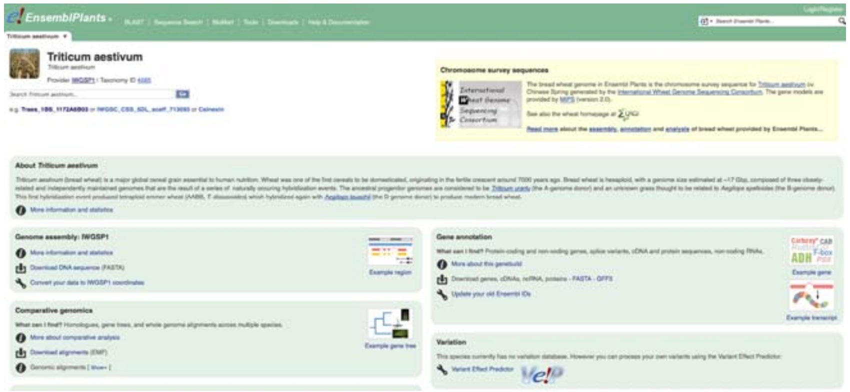
- Resources accessible at URGI, Versailles
  http://wheat-urgi.versailles.inra.fr/Seq-Repository/
- Raw sequence reads in the SRA
- Assemblies available at EBI since September

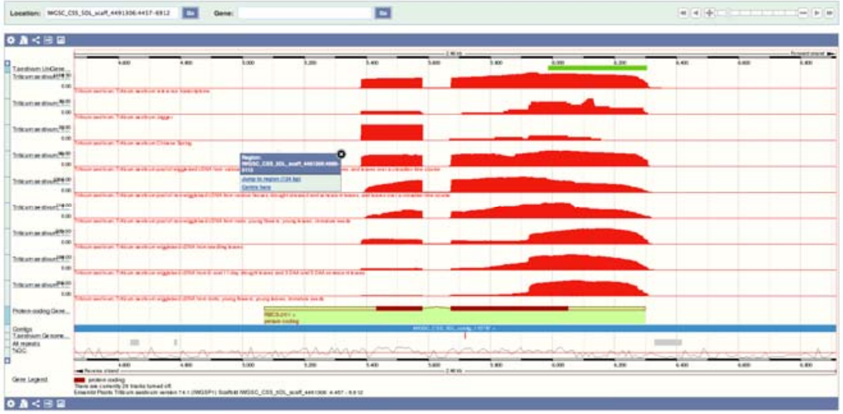
- ☐ **272 institutions**
- ☐ **40 countries**

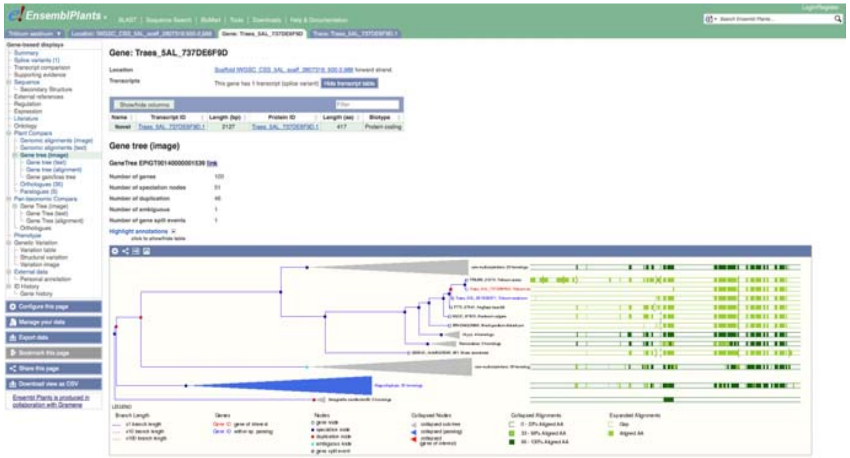
- ☐ **52,000 BLAST searches**
- ☐ **2.500 downloads**

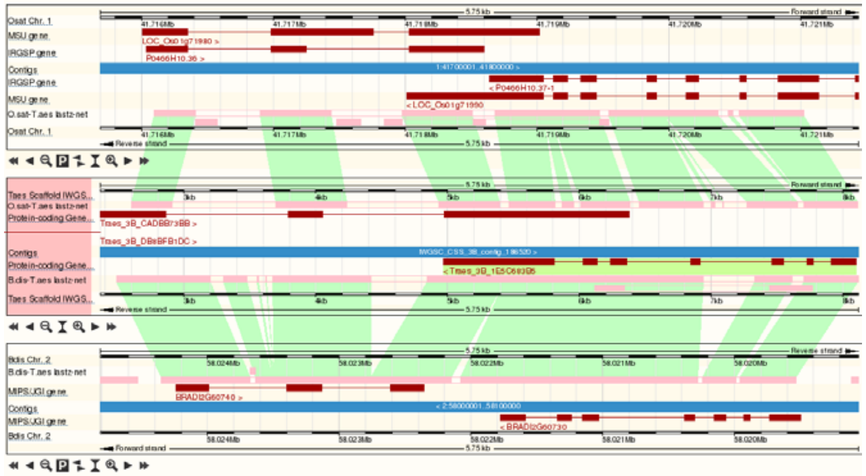
EnsemblPlants

http://plants.ensembl.org/Triticum_aestivum









Acknowledgements

IWGSC

Jaroslav Dolezel
& the IEB team

IWGSC

Kellye Eversole

TGAC
The Genome Analysis Centre™

Mario Caccamo
Sarah Ayling
Jon Wright
Ricardo Ramirez-Gonzalez

Hirokazu Handa
David Edwards
Burkhard Steuernagel
Jesse Poland
Eduard Akhunov
Luigi Cattivelli
Simen Sandve
Odd-Arne Olsen
Andrew Sharpe
& Many More....

HelmholtzZentrum münchen
German Research Center for Environmental Health

Klaus Mayer & team
Matthias Pfeiffer
Manuel Spannagl
Mihaela Martis

University of Saskatchewan

Curtis Pozniak

UMR 1095 GDEC INRA

Catherine Feuillet
& the 3BSeq Group

URGI

Michael Alaux

IWGSC Sponsors

Arcadia Biosciences, ACPFG, biogemma, Bayer CropScience, CSIRO, CIMMYT, Céréales Vallée, Dow, Dow AgroSciences, DuPont The miracles of science, evogene, Florimond Desprez Créations Variétales, GRDC Grains Research & Development Corporation, Graminor, Heartland Plant Innovations, INRA, Lochow, KWS, Together, we are Kansas Wheat, Limagrain from earth to life, Monsanto, RAGT, syngenta

IWGSC