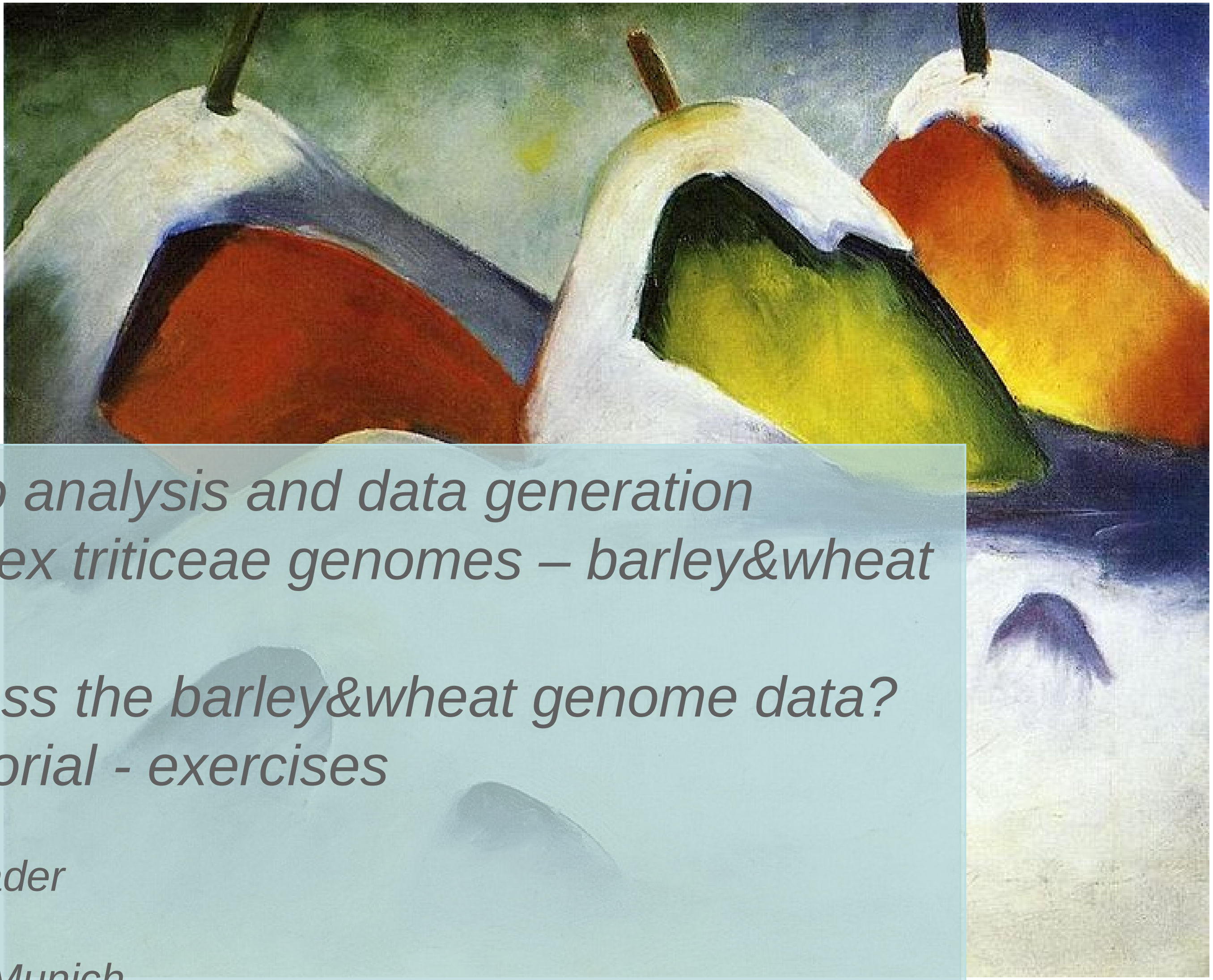
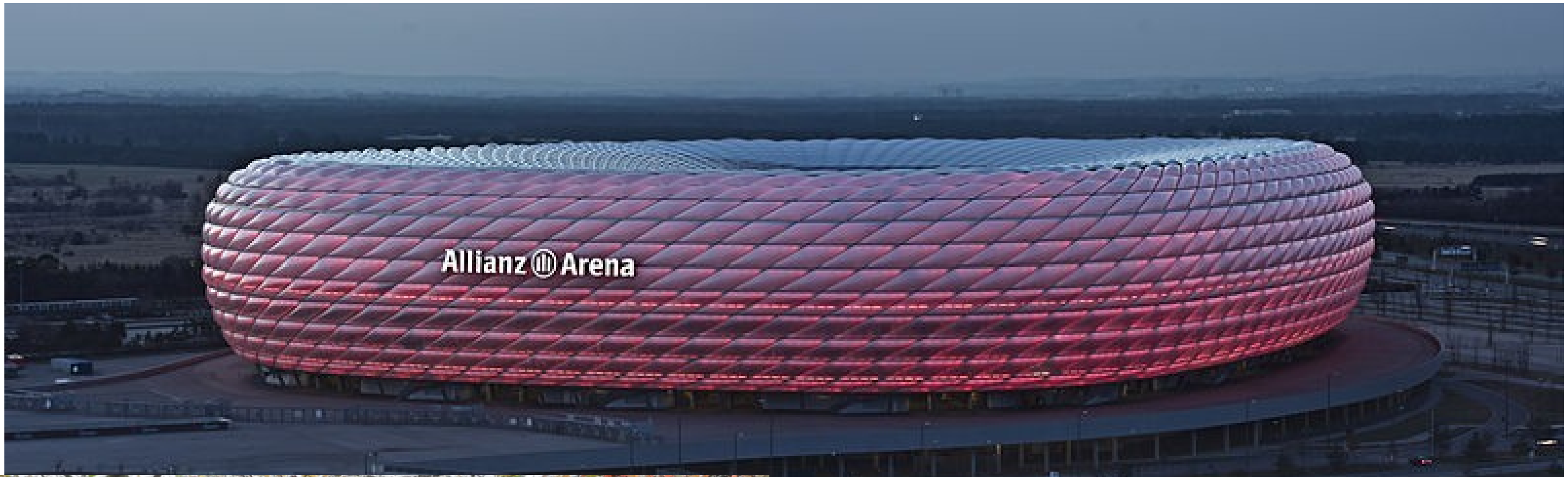


An Introduction into analysis and data generation concepts for complex triticeae genomes – barley&wheat

How/where to access the barley&wheat genome data? MIPS PlantsDB tutorial - exercises

Manuel Spannagl&Kai Bader
Klaus Mayer
MIPS, Helmholtz Center Munich

Who we are...



c/o Richard Bartz, wikimedia

Outline(1)

- Short introduction into data generation and analysis concepts for complex triticeae genomes: **the barley genome**:

  - Barley „genome zipper“
  - Barley genome sequencing, physical+genetic map integration
  - Gene prediction and annotation in barley
  - Comparative genomics in triticeae genomes

Outline(2)

- Analysing the 17 Gb **genome sequence of bread wheat** using NGS sequencing
- *Optional*: Access to the **Barley physical and genetic maps** – a tutorial intro (based on slides from Nils Stein, IPK)

Outline(3)

- **Barley&Wheat genome database resources**:

  - **MIPS PlantsDB tutorial**: how to access and analyse barley&wheat genome data within a comparative database framework – **interactive exercises**

  - **Homework exercises**: a „real-world“ use case accessing the barley genome databases – solutions provided

Bio*Green*formatics:
From Models to Crops, from Pets to Beasts

Arabidopsis thaliana ***(Nature, 2000)*** ***Medicago*** ***(Nature, 2011)*** ***Tomato*** ***(Nature, 2012)***

Sorghum ***(Nature, 2009)*** ***Brachypodium*** ***(Nature, 2010)*** ***Maize*** ***(Genome Research, 2006, Plant Phys. 2008, PNAS 2008)***

Barley ***(Plant Phys. 2009, Plant Cell 2011, Nature 2012)*** ***Arabidopsis lyrata*** ***(Nature Genetics 2011)***

Physcomitrella patens ***(Science 2008)*** ***Aegilops tauschii*** ***(Nature, under revision)***

Oryza ***(Genome Research 2001; in prep)*** ***Wheat*** ***(Plant Cell 2011;Plant Journal 2012,Nature 2012)***

Rye ***(PNAS 2012; in prep.)***

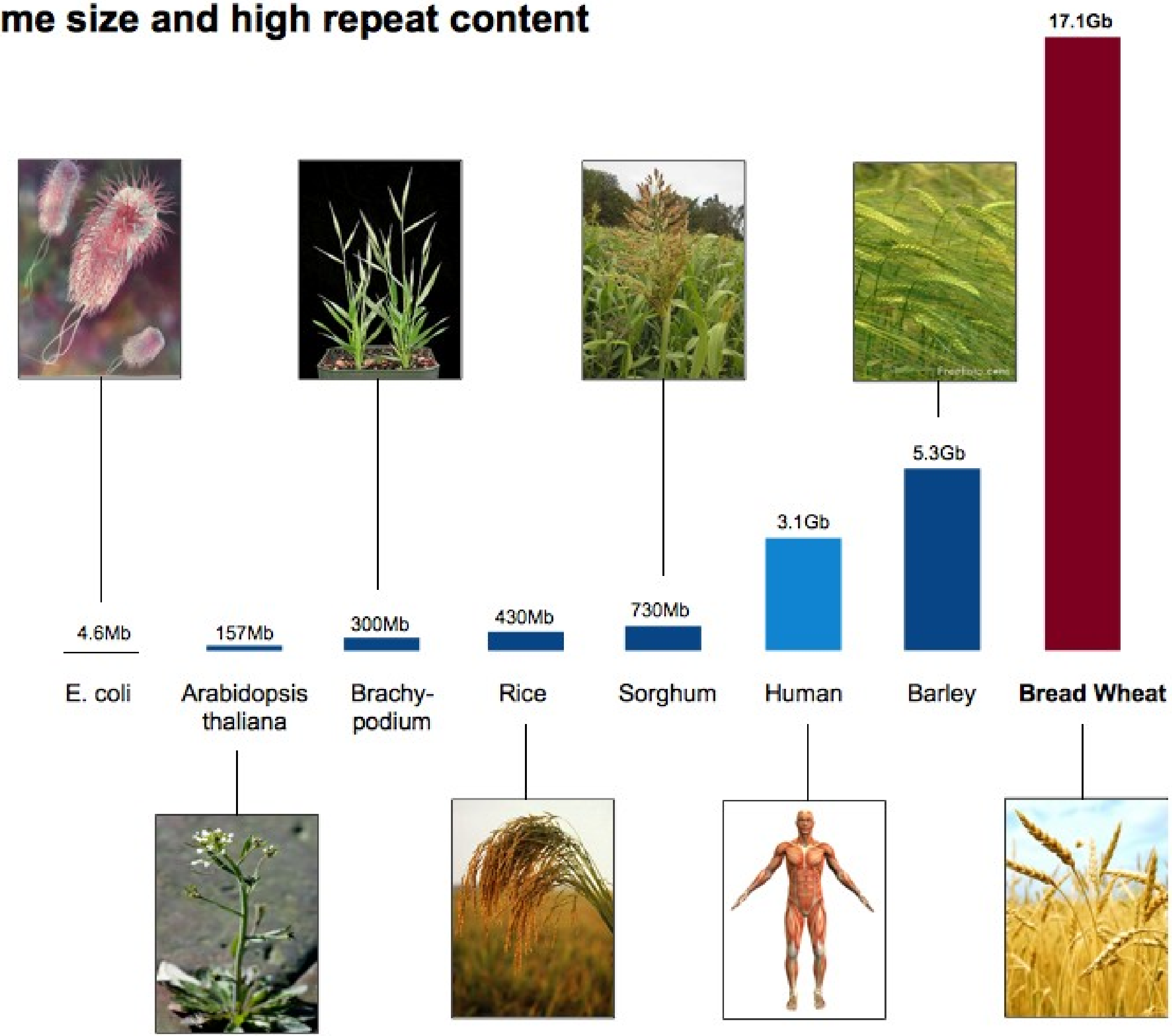
Lolium ***(submitted)*** ***Festuca*** ***(submitted)*** ***Spirodella*** ***(in prep.)*** ***Micromonas*** ***(Science 2009)***

Cardamine hirsuta

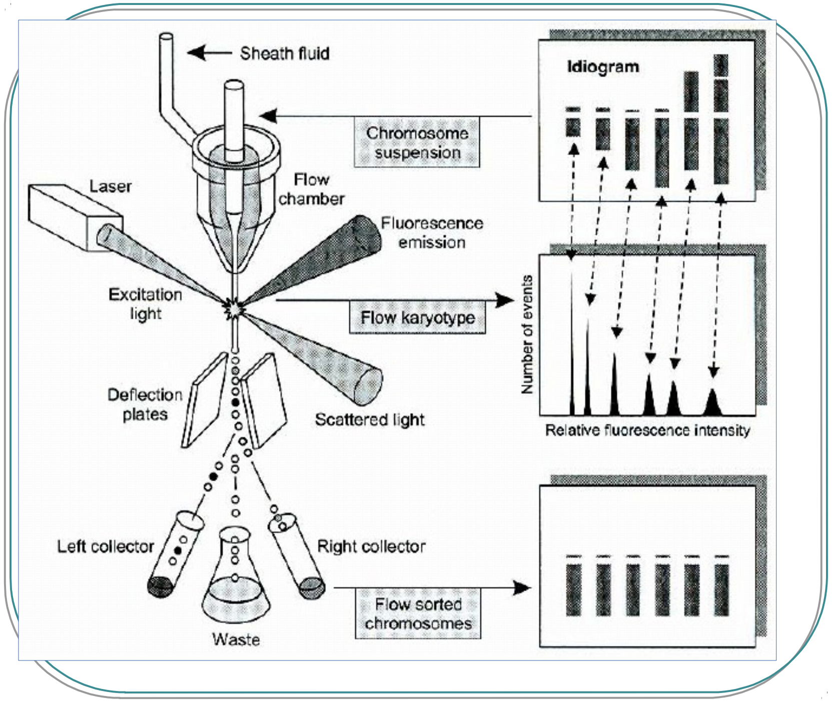
While NGS democratized sequencing the analytical bottleneck gets more pronounced.

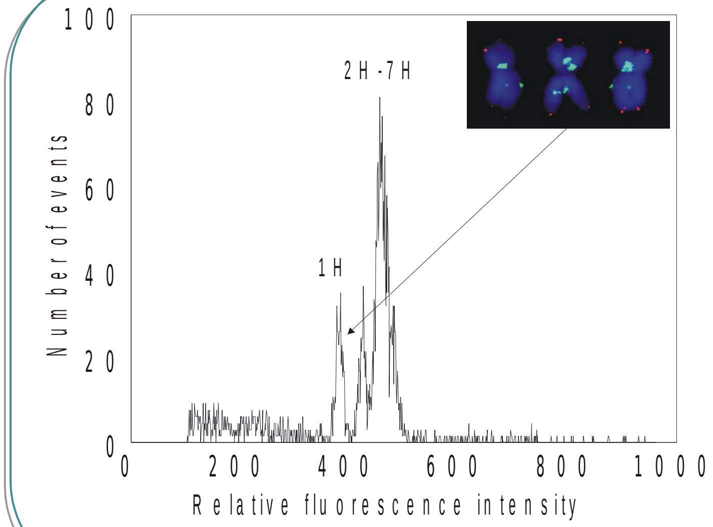
The Challenge

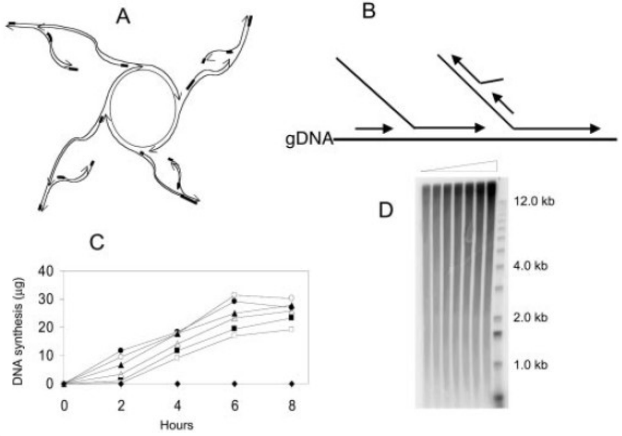
- Big genome size and high repeat content



Reduction of complexity by chromosome sorting







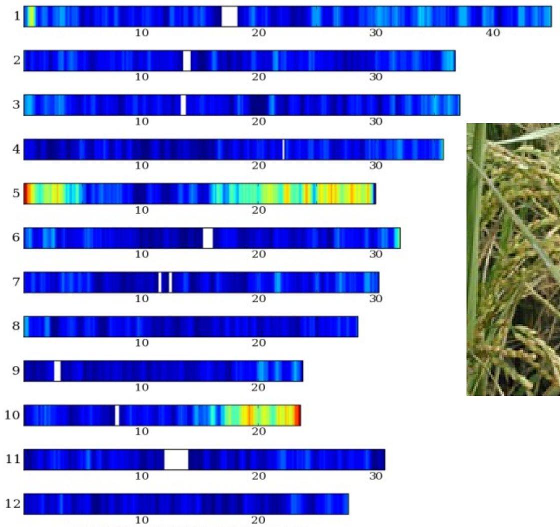
Barley WGS sequencing

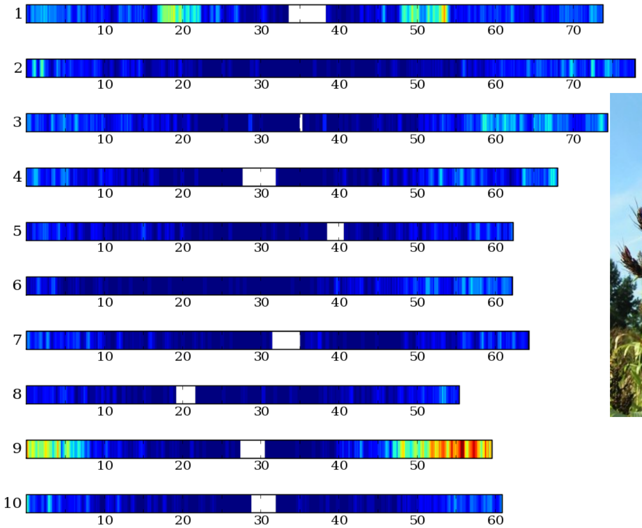
Barley reference sequence - Illumina

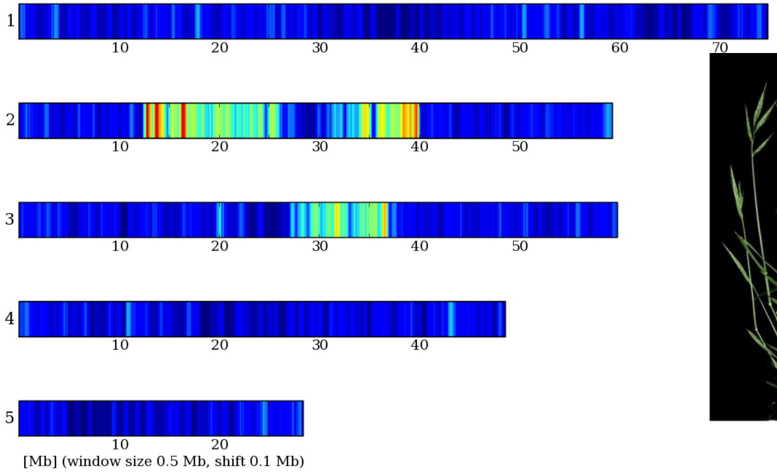
Morex 50x WGS assembly 3, repeat masked

- 2,670,738 contigs
- 1,868,648,155 bp sequence (min 200bp, max 36 kbp, mean 700bp, N50 1,425 bp)
- 936,664,164 bp (50.13%) masked sequence
- chromosome arm sorting available (CarmA)
- add. varieties available: Barke, Bowman, …
- SNP/SNVs called and available

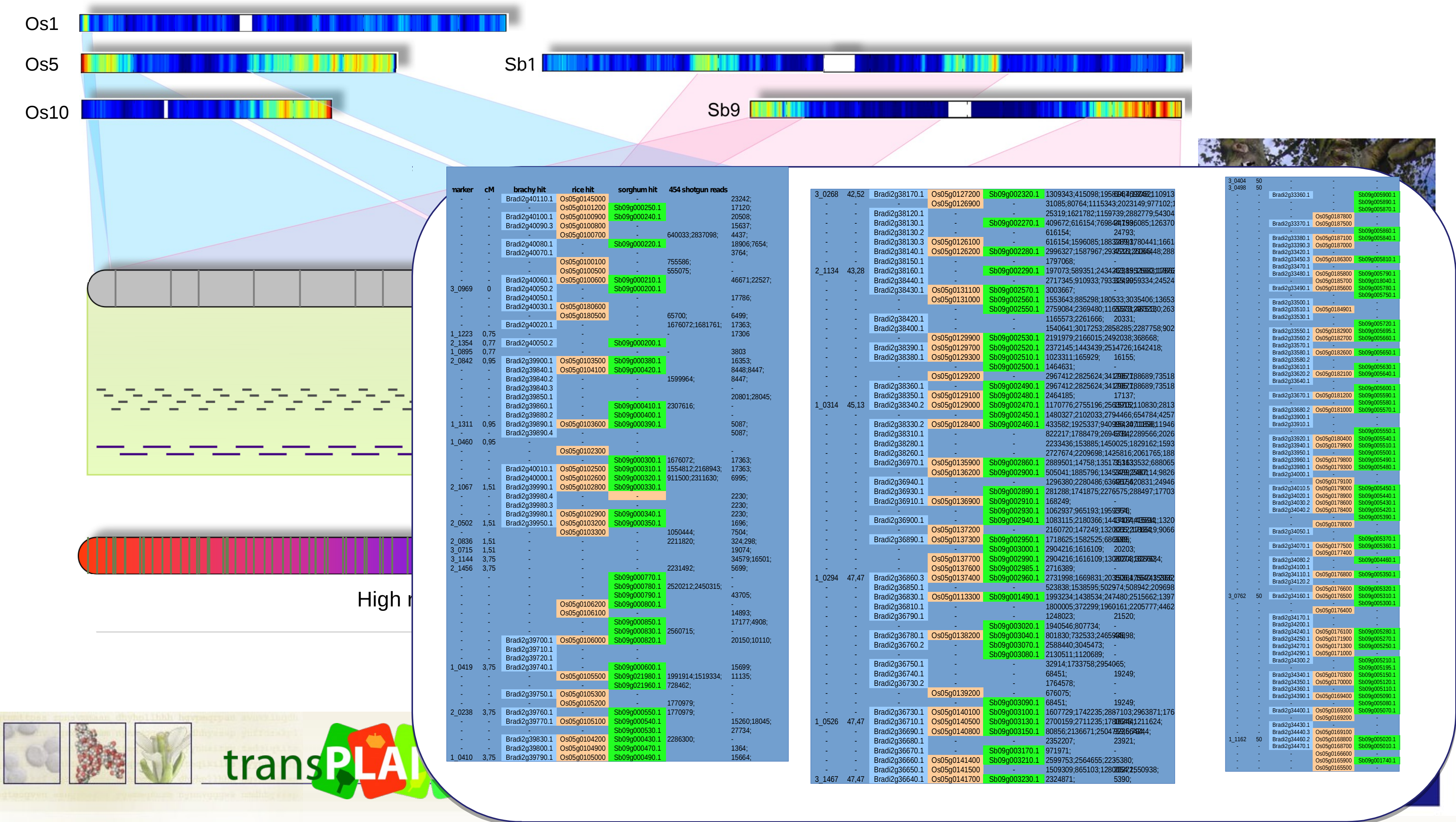
Barley vs *Brachypodium*, *Sorghum* and rice
-Synteny on a per gene resolution-







Syntenic Integration generates a „GenomeZipper“



marker	cM	brachy hit	rice hit	sorghum hit	454 shotgun reads	
-	-	Bradi2g40110.1	Os05g0145000	-		23242;
-	-	-	Os05g0101200	Sb09g000250.1		17120;
-	-	Bradi2g40100.1	Os05g0100900	Sb09g000240.1		20508;
-	-	Bradi2g40090.3	Os05g0100800	-		15637;
-	-	-	Os05g0100700	-	640033;2837098;	4437;
-	-	Bradi2g40080.1	-	Sb09g000220.1		18906;7654;
-	-	Bradi2g40070.1	-	-		3764;
-	-	-	Os05g0100100	-	755586;	-
-	-	-	Os05g0100500	-	555075;	-
-	-	Bradi2g40060.1	Os05g0100600	Sb09g000210.1		46671;22527;
3_0969	0	Bradi2g40050.2	-	Sb09g000200.1		
-	-	Bradi2g40050.1	-	-		17786;
-	-	Bradi2g40030.1	Os05g0180600	-		-
-	-	-	Os05g0180500	-	65700;	6499;
-	-	Bradi2g40020.1	-	-	1676072;1681761;	17363;
1_1223	0,75	-	-	-	-	17306
2_1354	0,77	Bradi2g40050.2	-	Sb09g000200.1		
1_0895	0,77	-	-	-	-	3803
2_0842	0,95	Bradi2g39900.1	Os05g0103500	Sb09g000380.1		16353;
-	-	Bradi2g39840.1	Os05g0104100	Sb09g000420.1		8448;8447;
-	-	Bradi2g39840.2	-	-	1599964;	8447;
-	-	Bradi2g39840.3	-	-		-
-	-	Bradi2g39850.1	-	-		20801;28045;
-	-	Bradi2g39860.1	-	Sb09g000410.1	2307616;	-
-	-	Bradi2g39880.2	-	Sb09g000400.1		-
1_1311	0,95	Bradi2g39890.1	Os05g0103600	Sb09g000390.1		5087;
-	-	Bradi2g39890.4	-	-		5087;
1_0460	0,95	-	-	-		
-	-	-	Os05g0102300	-		-
-	-	-	-	Sb09g000300.1	1676072;	17363;
-	-	Bradi2g40010.1	Os05g0102500	Sb09g000310.1	1554812;2168943;	17363;
-	-	Bradi2g40000.1	Os05g0102600	Sb09g000320.1	911500;2311630;	6995;
2_1067	1,51	Bradi2g39990.1	Os05g0102800	Sb09g000330.1		
-	-	Bradi2g39980.4	-	-		2230;
-	-	Bradi2g39980.3	-	-		2230;
-	-	Bradi2g39980.1	Os05g0102900	Sb09g000340.1		2230;
2_0502	1,51	Bradi2g39950.1	Os05g0103200	Sb09g000350.1		1696;
-	-	-	Os05g0103300	-	1050444;	7504;
2_0836	1,51	-	-	-	2211820;	324;298;
3_0715	1,51	-	-	-		19074;
3_1144	3,75	-	-	-		34579;16501;
2_1456	3,75	-	-	-	2231492;	5699;
-	-	-	-	Sb09g000770.1		-
-	-	-	-	Sb09g000780.1	2520212;2450315;	-
-	-	-	-	Sb09g000790.1		43705;
-	-	-	Os05g0106200	Sb09g000800.1		-
-	-	-	Os05g0106100	-		14893;
-	-	-	-	Sb09g000850.1		17177;4908;
-	-	-	-	Sb09g000830.1	2560715;	-
-	-	Bradi2g39700.1	Os05g0106000	Sb09g000820.1		20150;10110;
-	-	Bradi2g39710.1	-	-		
-	-	Bradi2g39720.1	-	-		
1_0419	3,75	Bradi2g39740.1	-	Sb09g000600.1		15699;
-	-	-	Os05g0105500	Sb09g021980.1	1991914;1519334;	11135;
-	-	-	-	Sb09g021960.1	728462;	-
-	-	Bradi2g39750.1	Os05g0105300	-		-
-	-	-	Os05g0105200	-	1770979;	-
2_0238	3,75	Bradi2g39760.1	-	Sb09g000550.1	1770979;	-
-	-	Bradi2g39770.1	Os05g0105100	Sb09g000540.1		15260;18045;
-	-	-	-	Sb09g000530.1		27734;
-	-	Bradi2g39830.1	Os05g0104200	Sb09g000430.1	2286300;	-
-	-	Bradi2g39800.1	Os05g0104900	Sb09g000470.1		1364;
1_0410	3,75	Bradi2g39790.1	Os05g0105000	Sb09g000490.1		15664;

3_0268	42,52	Bradi2g38170.1	Os05g0127200	Sb09g002320.1	1309343;415098;195[illegible]110913	
-	-	-	Os05g0126900	-	31085;80764;1115343;2023149;977102;1	
-	-	Bradi2g38120.1	-	-	25319;1621782;1159-39;2882779;54304	
-	-	Bradi2g38130.1	-	Sb09g002270.1	409672;616154;7698[illegible]596085;126370	
-	-	Bradi2g38130.2	-	-	616154;	24793;
-	-	Bradi2g38130.3	Os05g0126100	-	616154;1596085;188[illegible]780441;1661	
-	-	Bradi2g38140.1	Os05g0126200	Sb09g002280.1	2996327;1587967;29[illegible]448;288	
-	-	Bradi2g38150.1	-	-	1797068;	-
2_1134	43,28	Bradi2g38160.1	-	Sb09g002290.1	197073;589351;2434[illegible]	
-	-	Bradi2g38440.1	-	-	2717345;910933;793[illegible]59334;24524	
-	-	Bradi2g38430.1	Os05g0131100	Sb09g002570.1	3003667;	-
-	-	-	Os05g0131000	Sb09g002560.1	1553643;885298;180533;3035406;13653	
-	-	-	-	Sb09g002550.1	2759084;2369480;116[illegible]0;263	
-	-	Bradi2g38420.1	-	-	1165573;2261666;	20331;
-	-	Bradi2g38400.1	-	-	1540641;3017253;2858285;2287758;902	
-	-	-	Os05g0129900	Sb09g002530.1	2191979;2166015;249-2038;368668;	
-	-	Bradi2g38390.1	Os05g0129700	Sb09g002520.1	2372145;1443439;251-4726;1642418;	
-	-	Bradi2g38380.1	Os05g0129300	Sb09g002510.1	1023311;165929;	16155;
-	-	-	-	Sb09g002500.1	1464631;	-
-	-	-	Os05g0129200	-	2967412;2825624;34[illegible]88689;73518	
-	-	Bradi2g38360.1	-	Sb09g002490.1	2967412;2825624;34[illegible]88689;73518	
-	-	Bradi2g38350.1	Os05g0129100	Sb09g002480.1	2464185;	17137;
1_0314	45,13	Bradi2g38340.2	Os05g0129000	Sb09g002470.1	1170776;2755196;256[illegible]110830;2813	
-	-	-	-	Sb09g002450.1	1480327;2102033;2794466;654784;4257	
-	-	Bradi2g38330.2	Os05g0128400	Sb09g002460.1	433582;1925337;940[illegible]11946	
-	-	Bradi2g38310.1	-	-	822217;1788479;269[illegible]289566;2026	
-	-	Bradi2g38280.1	-	-	2233436;153885;1450-025;1829162;1593	
-	-	Bradi2g38260.1	-	-	2727674;2209698;1425816;2061765;188	
-	-	Bradi2g36970.1	Os05g0135900	Sb09g002860.1	2889501;14758;1351[illegible]3532;688065	
-	-	-	Os05g0136200	Sb09g002900.1	505041;1885796;134[illegible]14;9826	
-	-	Bradi2g36940.1	-	-	1296380;2280486;63[illegible]20831;24946	
-	-	Bradi2g36930.1	-	Sb09g002890.1	281288;1741875;2276575;288497;17703	
-	-	Bradi2g36910.1	Os05g0136900	Sb09g002910.1	168249;	-
-	-	-	-	Sb09g002930.1	1062937;965193;195[illegible];	
-	-	Bradi2g36900.1	-	Sb09g002940.1	1083115;2180366;144[illegible];1320	
-	-	-	Os05g0137200	-	2160720;147249;1320[illegible]9;9066	
-	-	Bradi2g36890.1	Os05g0137300	Sb09g002950.1	1718625;1582525;68[illegible];	
-	-	-	-	Sb09g003000.1	2904216;1616109;	20203;
-	-	-	Os05g0137700	Sb09g002990.1	2904216;1616109;130[illegible]4;	
-	-	-	Os05g0137600	Sb09g002985.1	2716389;	-
1_0294	47,47	Bradi2g36860.3	Os05g0137400	Sb09g002960.1	2731998;1669831;203[illegible]	
-	-	Bradi2g36850.1	-	-	523838;1538595;502974;508942;209698	
-	-	Bradi2g36830.1	Os05g0113300	Sb09g001490.1	1993234;1438534;247-480;2515662;1397	
-	-	Bradi2g36810.1	-	-	1800005;372299;1960-161;2205777;4462	
-	-	Bradi2g36790.1	-	-	1248023;	21520;
-	-	-	-	Sb09g003020.1	1940546;807734;	-
-	-	Bradi2g36780.1	Os05g0138200	Sb09g003040.1	801830;732533;24659[illegible]98;	
-	-	Bradi2g36760.2	-	Sb09g003070.1	2588440;3045473;	-
-	-	-	-	Sb09g003080.1	2130511;1120689;	-
-	-	Bradi2g36750.1	-	-	32914;1733758;2954-065;	
-	-	Bradi2g36740.1	-	-	68451;	19249;
-	-	Bradi2g36730.2	-	-	1764578;	-
-	-	-	Os05g0139200	-	676075;	-
-	-	-	-	Sb09g003090.1	68451;	19249;
-	-	Bradi2g36730.1	Os05g0140100	Sb09g003100.1	1607729;1742235;2887103;2963871;176	
1_0526	47,47	Bradi2g36710.1	Os05g0140500	Sb09g003130.1	2700159;2711235;178[illegible]211624;	
-	-	Bradi2g36690.1	Os05g0140800	Sb09g003150.1	80856;2136671;2504[illegible]4;	
-	-	Bradi2g36680.1	-	-	2352207;	23921;
-	-	Bradi2g36670.1	-	Sb09g003170.1	971971;	-
-	-	Bradi2g36660.1	Os05g0141400	Sb09g003210.1	2599753;2564655;2235380;	
-	-	Bradi2g36650.1	Os05g0141500	-	1509309;865103;128[illegible]50938;	
3_1467	47,47	Bradi2g36640.1	Os05g0141700	Sb09g003230.1	2324871;	5390;

3_0404	50	-	-	-
3_0498	50	-	-	-
-	-	Bradi2g33360.1	-	Sb09g005900.1
-	-	-	-	Sb09g005890.1
-	-	-	-	Sb09g005870.1
-	-	-	Os05g0187800	-
-	-	Bradi2g33370.1	Os05g0187500	-
-	-	-	-	Sb09g005860.1
-	-	Bradi2g33380.1	Os05g0187100	Sb09g005840.1
-	-	Bradi2g33390.3	Os05g0187000	-
-	-	Bradi2g33420.1	-	-
-	-	Bradi2g33450.3	Os05g0186300	Sb09g005810.1
-	-	Bradi2g33470.1	-	-
-	-	Bradi2g33480.1	Os05g0185800	Sb09g005790.1
-	-	-	Os05g0185700	Sb09g018040.1
-	-	Bradi2g33490.1	Os05g0185600	Sb09g005780.1
-	-	-	-	Sb09g005750.1
-	-	Bradi2g33500.1	-	-
-	-	Bradi2g33510.1	Os05g0184901	-
-	-	Bradi2g33530.1	-	-
-	-	-	-	Sb09g005720.1
-	-	Bradi2g33550.1	Os05g0182900	Sb09g005695.1
-	-	Bradi2g33560.2	Os05g0182700	Sb09g005660.1
-	-	Bradi2g33570.1	-	-
-	-	Bradi2g33580.1	Os05g0182600	Sb09g005650.1
-	-	Bradi2g33580.2	-	-
-	-	Bradi2g33610.1	-	Sb09g005630.1
-	-	Bradi2g33620.2	Os05g0182100	Sb09g005640.1
-	-	Bradi2g33640.1	-	-
-	-	-	-	Sb09g005600.1
-	-	Bradi2g33670.1	Os05g0181200	Sb09g005590.1
-	-	-	-	Sb09g005580.1
-	-	Bradi2g33680.2	Os05g0181000	Sb09g005570.1
-	-	Bradi2g33900.1	-	-
-	-	Bradi2g33910.1	-	-
-	-	-	-	Sb09g005550.1
-	-	Bradi2g33920.1	Os05g0180400	Sb09g005540.1
-	-	Bradi2g33940.1	Os05g0179900	Sb09g005510.1
-	-	Bradi2g33950.1	-	Sb09g005500.1
-	-	Bradi2g33960.1	Os05g0179800	Sb09g005490.1
-	-	Bradi2g33980.1	Os05g0179300	Sb09g005480.1
-	-	Bradi2g34000.1	-	-
-	-	-	Os05g0179100	-
-	-	Bradi2g34010.5	Os05g0179000	Sb09g005450.1
-	-	Bradi2g34020.1	Os05g0178900	Sb09g005440.1
-	-	Bradi2g34030.2	Os05g0178600	Sb09g005430.1
-	-	Bradi2g34040.2	Os05g0178400	Sb09g005420.1
-	-	-	-	Sb09g005390.1
-	-	-	Os05g0178000	-
-	-	Bradi2g34050.1	-	-
-	-	-	-	Sb09g005370.1
-	-	Bradi2g34070.1	Os05g0177500	Sb09g005360.1
-	-	-	Os05g0177400	-
-	-	Bradi2g34080.2	-	Sb09g004460.1
-	-	Bradi2g34100.1	-	-
-	-	Bradi2g34110.1	Os05g0176800	Sb09g005350.1
-	-	Bradi2g34120.2	-	-
-	-	-	Os05g0176600	Sb09g005320.1
3_0762	50	Bradi2g34160.1	Os05g0176500	Sb09g005310.1
-	-	-	-	Sb09g005300.1
-	-	-	Os05g0176400	-
-	-	Bradi2g34170.1	-	-
-	-	Bradi2g34200.1	-	-
-	-	Bradi2g34240.1	Os05g0176100	Sb09g005280.1
-	-	Bradi2g34250.1	Os05g0171900	Sb09g005270.1
-	-	Bradi2g34270.1	Os05g0171300	Sb09g005250.1
-	-	Bradi2g34290.1	Os05g0171000	-
-	-	Bradi2g34300.2	-	Sb09g005210.1
-	-	-	-	Sb09g005195.1
-	-	Bradi2g34340.1	Os05g0170300	Sb09g005150.1
-	-	Bradi2g34350.1	Os05g0170000	Sb09g005120.1
-	-	Bradi2g34360.1	-	Sb09g005110.1
-	-	Bradi2g34390.1	Os05g0169400	Sb09g005090.1
-	-	-	-	Sb09g005080.1
-	-	Bradi2g34400.1	Os05g0169300	Sb09g005070.1
-	-	-	Os05g0169200	-
-	-	Bradi2g34430.1	-	-
-	-	Bradi2g34440.3	Os05g0169100	-
1_1162	50	Bradi2g34460.2	Os05g0168800	Sb09g005020.1
-	-	Bradi2g34470.1	Os05g0168700	Sb09g005010.1
-	-	-	Os05g0166600	-
-	-	-	Os05g0165900	Sb09g001740.1
-	-	-	Os05g0165500	-

GenomeZipper…what is it? For what?

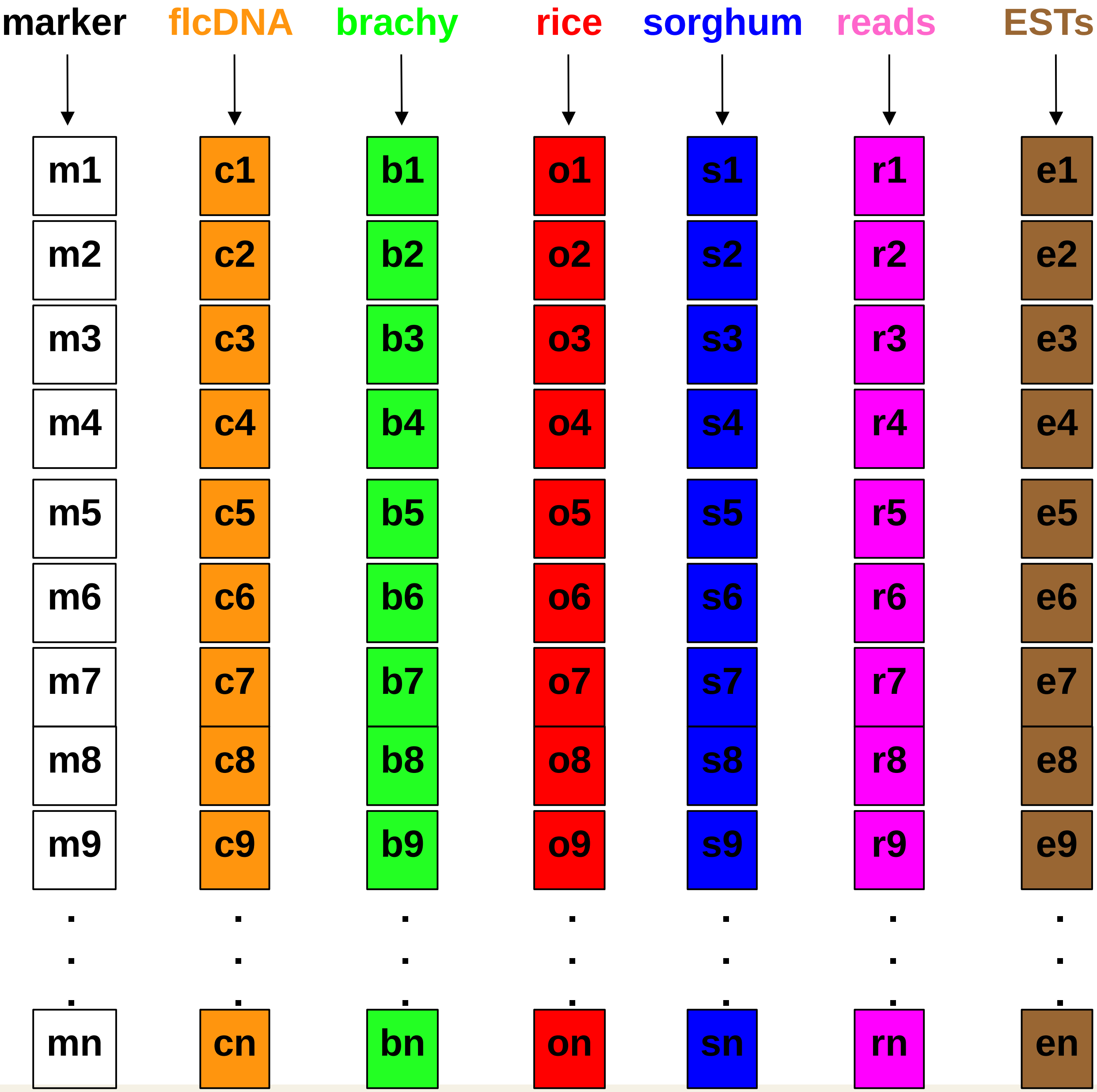
- Is an approach developed to create an ordered virtual gene map for a chromosome

- It smartly combines chromosome sorting, next generation sequencing, genetic maps, flcDNAs and systematic exploitation of conserved synteny with model grasses

- It provides a valuable surrogate for the gene space of the analyzed chromosome/genome

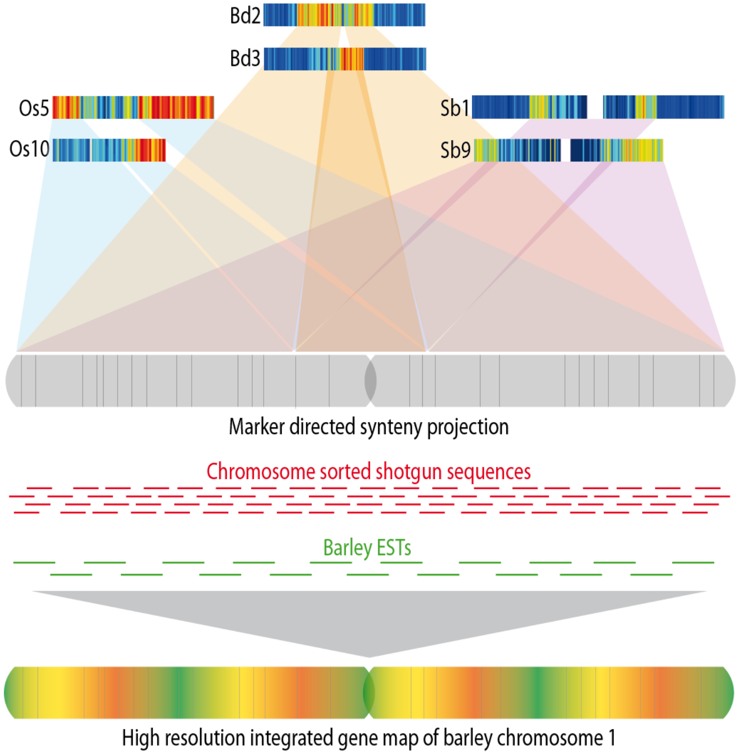
- Requirements:
    - Masked 454 reads/contigs
    - Orthologs from syntenic regions

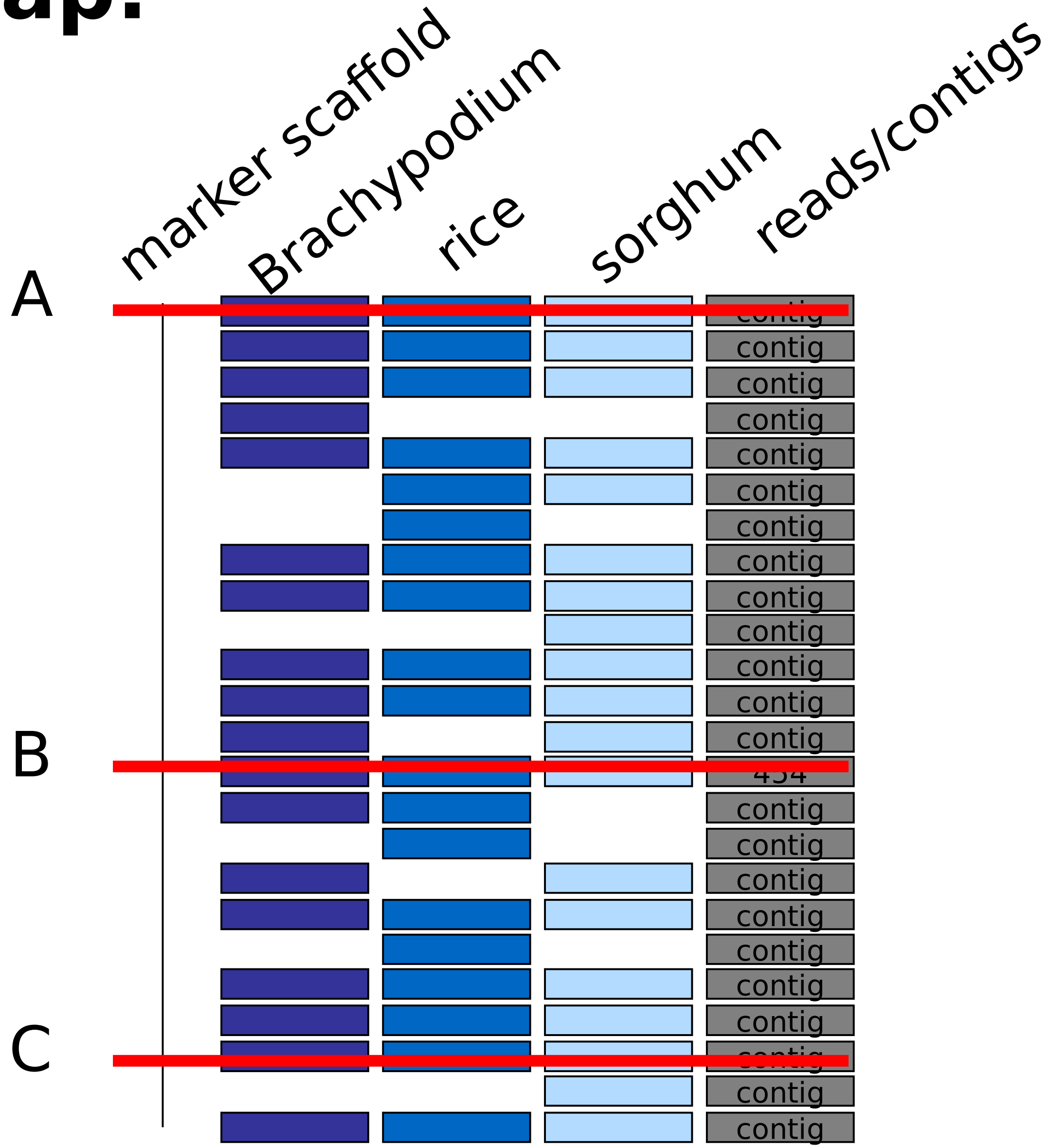
GZipper Input



marker	flcDNA	brachy	rice	sorghum	reads	ESTs
m1	c1	b1	o1	s1	r1	e1
m2	c2	b2	o2	s2	r2	e2
m3	c3	b3	o3	s3	r3	e3
m4	c4	b4	o4	s4	r4	e4
m5	c5	b5	o5	s5	r5	e5
m6	c6	b6	o6	s6	r6	e6
m7	c7	b7	o7	s7	r7	e7
m8	c8	b8	o8	s8	r8	e8
m9	c9	b9	o9	s9	r9	e9
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
mn	cn	bn	on	sn	rn	en

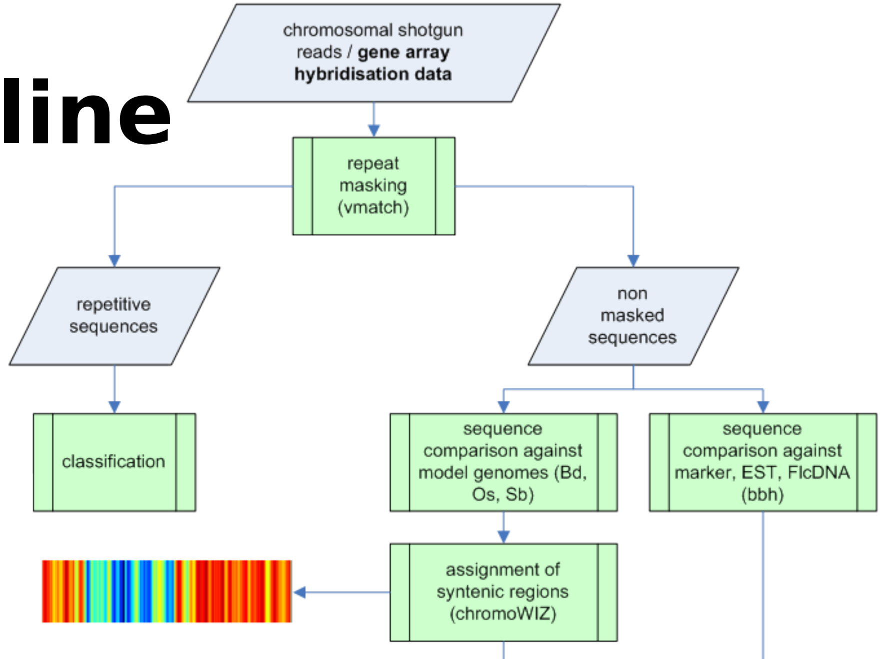
Virtual Gene Map: Syntenic Integration

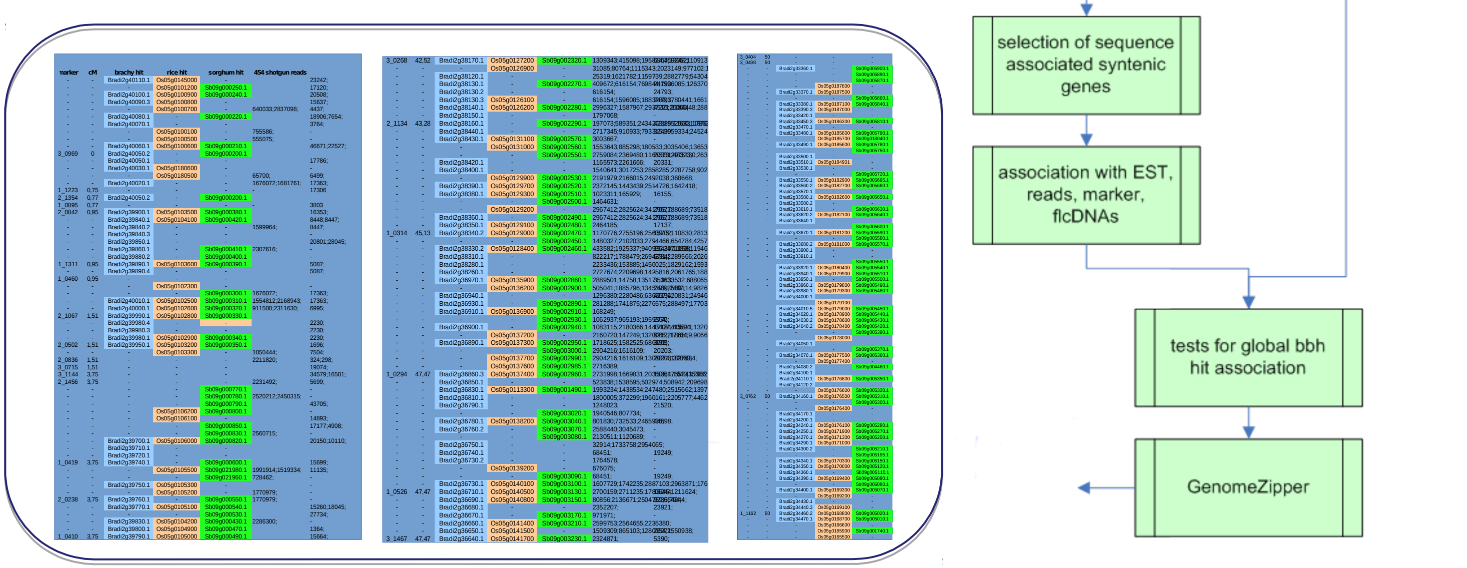

Bd2
Bd3
Os5
Os10
Sb1
Sb9
Marker directed synteny projection
Chromosome sorted shotgun sequences
Barley ESTs
High resolution integrated gene map of barley chromosome 1



	marker scaffold	Brachypodium	rice	sorghum	reads/contigs
A		■	■	■	contig
		■	■	■	contig
		■	■	■	contig
		■			contig
		■	■	■	contig
			■	■	contig
			■		contig
		■	■	■	contig
		■	■	■	contig
				■	contig
		■	■	■	contig
		■	■	■	contig
		■		■	contig
B		■	■	■	454
		■	■		contig
			■		contig
		■		■	contig
		■	■	■	contig
			■		contig
		■	■	■	contig
		■	■	■	contig
C		■	■	■	contig
				■	contig
		■	■	■	contig

GenomeZipper Pipeline


chromosomal shotgun reads / gene array hybridisation data
repeat masking (vmatch)
repetitive sequences
non masked sequences
classification
sequence comparison against model genomes (Bd, Os, Sb)
sequence comparison against marker, EST, FlcDNA (bbh)
assignment of syntenic regions (chromoWIZ)

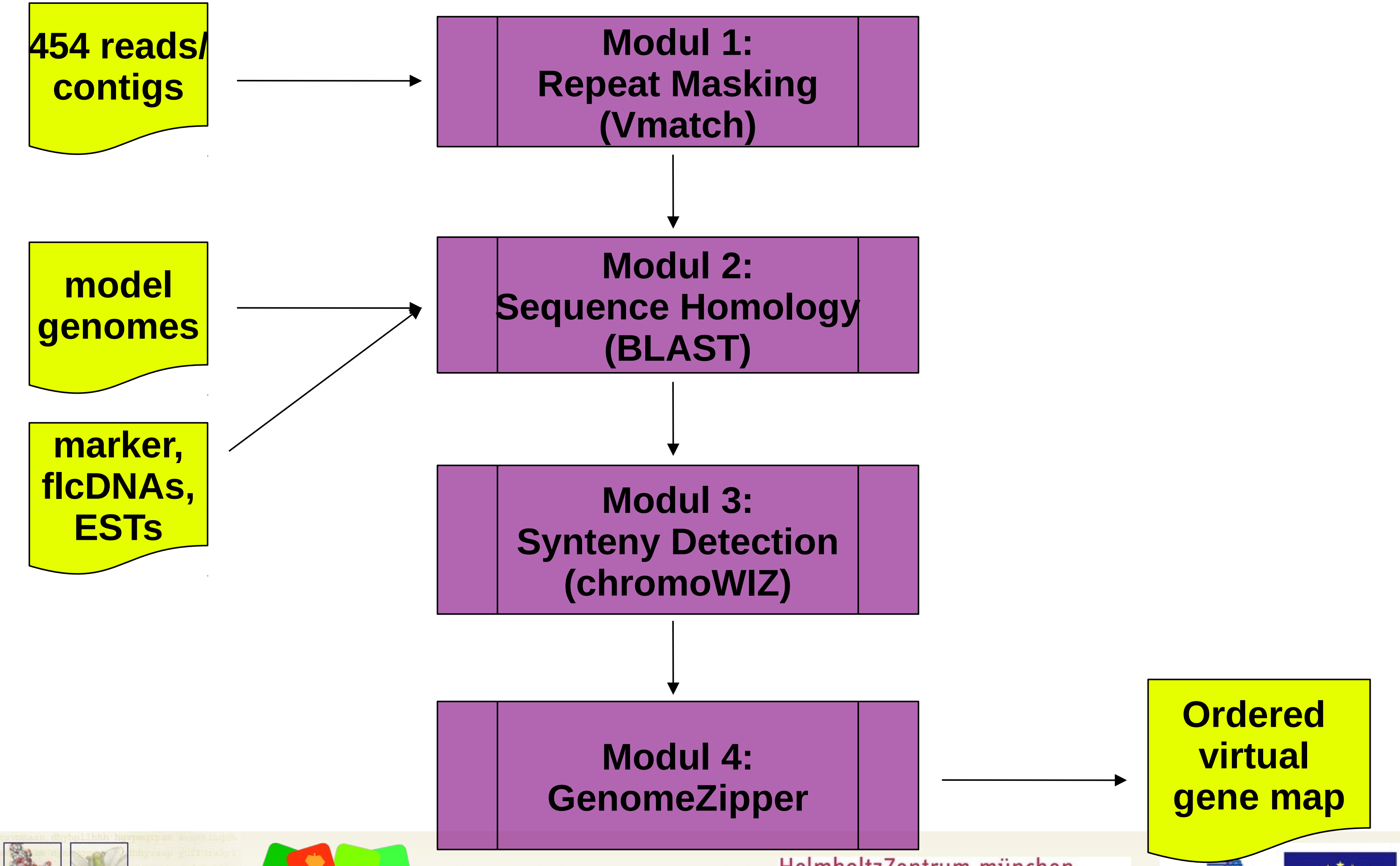

selection of sequence associated syntenic genes
association with EST, reads, marker, flcDNAs
tests for global bbh hit association
GenomeZipper

marker	cM	brachy hit	rice hit	sorghum hit	454 shotgun reads	
-	-	Bradi2g40110.1	Os05g0145000	-		23242;
-	-	-	Os05g0101200	Sb09g000250.1		17120;
-	-	Bradi2g40100.1	Os05g0100900	Sb09g000240.1		20508;
-	-	Bradi2g40090.3	Os05g0100800	-		15637;
-	-	-	Os05g0100700	-	640033;2837098;	4437;
-	-	Bradi2g40080.1	-	Sb09g000220.1		18906;7654;
-	-	Bradi2g40070.1	-	-		3764;
-	-	-	Os05g0100100	-	755586;	-
-	-	-	Os05g0100500	-	555075;	-
-	-	Bradi2g40060.1	Os05g0100600	Sb09g000210.1		46671;22527;
3_0969	0	Bradi2g40050.2	-	Sb09g000200.1		
-	-	Bradi2g40050.1	-	-		17786;
-	-	Bradi2g40030.1	Os05g0180600	-		-
-	-	-	Os05g0180500	-	65700;	6499;
-	-	Bradi2g40020.1	-	-	1676072;1681761;	17363;
1_1223	0,75	-	-	-	-	17306
2_1354	0,77	Bradi2g40050.2	-	Sb09g000200.1		
1_0895	0,77	-	-	-	-	3803
2_0842	0,95	Bradi2g39900.1	Os05g0103500	Sb09g000380.1		16353;
-	-	Bradi2g39840.1	Os05g0104100	Sb09g000420.1		8448;8447;
-	-	Bradi2g39840.2	-	-	1599964;	8447;
-	-	Bradi2g39840.3	-	-		-
-	-	Bradi2g39850.1	-	-		20801;28045;
-	-	Bradi2g39860.1	-	Sb09g000410.1	2307616;	-
-	-	Bradi2g39880.2	-	Sb09g000400.1		-
1_1311	0,95	Bradi2g39890.1	Os05g0103600	Sb09g000390.1		5087;
-	-	Bradi2g39890.4	-	-		5087;
1_0460	0,95	-	-	-		
-	-	-	Os05g0102300	-		-
-	-	-	-	Sb09g000300.1	1676072;	17363;
-	-	Bradi2g40010.1	Os05g0102500	Sb09g000310.1	1554812;2168943;	17363;
-	-	Bradi2g40000.1	Os05g0102600	Sb09g000320.1	911500;2311630;	6995;
2_1067	1,51	Bradi2g39990.1	Os05g0102800	Sb09g000330.1		
-	-	Bradi2g39980.4	-	-		2230;
-	-	Bradi2g39980.3	-	-		2230;
-	-	Bradi2g39980.1	Os05g0102900	Sb09g000340.1		2230;
2_0502	1,51	Bradi2g39950.1	Os05g0103200	Sb09g000350.1		1696;
-	-	-	Os05g0103300	-	1050444;	7504;
2_0836	1,51	-	-	-	2211820;	324;298;
3_0715	1,51	-	-	-		19074;
3_1144	3,75	-	-	-		34579;16501;
2_1456	3,75	-	-	-	2231492;	5699;
-	-	-	-	Sb09g000770.1		-
-	-	-	-	Sb09g000780.1	2520212;2450315;	-
-	-	-	-	Sb09g000790.1		43705;
-	-	-	Os05g0106200	Sb09g000800.1		-
-	-	-	Os05g0106100	-		14893;
-	-	-	-	Sb09g000850.1		17177;4908;
-	-	-	-	Sb09g000830.1	2560715;	-
-	-	Bradi2g39700.1	Os05g0106000	Sb09g000820.1		20150;10110;
-	-	Bradi2g39710.1	-	-		
-	-	Bradi2g39720.1	-	-		
1_0419	3,75	Bradi2g39740.1	-	Sb09g000600.1		15699;
-	-	-	Os05g0105500	Sb09g021980.1	1991914;1519334;	11135;
-	-	-	-	Sb09g021960.1	728462;	-
-	-	Bradi2g39750.1	Os05g0105300	-		-
-	-	-	Os05g0105200	-	1770979;	-
2_0238	3,75	Bradi2g39760.1	-	Sb09g000550.1	1770979;	-
-	-	Bradi2g39770.1	Os05g0105100	Sb09g000540.1		15260;18045;
-	-	-	-	Sb09g000530.1		27734;
-	-	Bradi2g39830.1	Os05g0104200	Sb09g000430.1	2286300;	-
-	-	Bradi2g39800.1	Os05g0104900	Sb09g000470.1		1364;
1_0410	3,75	Bradi2g39790.1	Os05g0105000	Sb09g000490.1		15664;

3_0268	42,52	Bradi2g38170.1	Os05g0127200	Sb09g002320.1	1309343;415098;195[illegible]110913	
-	-	-	Os05g0126900	-	31085;80764;1115343;2023149;977102;1	
-	-	Bradi2g38120.1	-	-	25319;1621782;1159739;2882779;54304	
-	-	Bradi2g38130.1	-	Sb09g002270.1	409672;616154;7698[illegible]6085;126370	
-	-	Bradi2g38130.2	-	-	616154;	24793;
-	-	Bradi2g38130.3	Os05g0126100	-	616154;1596085;188[illegible]780441;1661	
-	-	Bradi2g38140.1	Os05g0126200	Sb09g002280.1	2996327;1587967;29[illegible]448;288	
-	-	Bradi2g38150.1	-	-	1797068;	-
2_1134	43,28	Bradi2g38160.1	-	Sb09g002290.1	197073;589351;2434[illegible]	
-	-	Bradi2g38440.1	-	-	2717345;910933;793[illegible]9334;24524	
-	-	Bradi2g38430.1	Os05g0131100	Sb09g002570.1	3003667;	-
-	-	-	Os05g0131000	Sb09g002560.1	1553643;885298;180533;3035406;13653	
-	-	-	-	Sb09g002550.1	2759084;2369480;11[illegible]0;263	
-	-	Bradi2g38420.1	-	-	1165573;2261666;	20331;
-	-	Bradi2g38400.1	-	-	1540641;3017253;2858285;2287758;902	
-	-	-	Os05g0129900	Sb09g002530.1	2191979;2166015;2492038;368668;	
-	-	Bradi2g38390.1	Os05g0129700	Sb09g002520.1	2372145;1443439;2514726;1642418;	
-	-	Bradi2g38380.1	Os05g0129300	Sb09g002510.1	1023311;165929;	16155;
-	-	-	-	Sb09g002500.1	1464631;	-
-	-	-	Os05g0129200	-	2967412;2825624;34[illegible]88689;73518	
-	-	Bradi2g38360.1	-	Sb09g002490.1	2967412;2825624;34[illegible]88689;73518	
-	-	Bradi2g38350.1	Os05g0129100	Sb09g002480.1	2464185;	17137;
1_0314	45,13	Bradi2g38340.2	Os05g0129000	Sb09g002470.1	1170776;2755196;256[illegible]110830;2813	
-	-	-	-	Sb09g002450.1	1480327;2102033;2794466;654784;4257	
-	-	Bradi2g38330.2	Os05g0128400	Sb09g002460.1	433582;1925337;940[illegible]11946	
-	-	Bradi2g38310.1	-	-	822217;1788479;269[illegible]2289566;2026	
-	-	Bradi2g38280.1	-	-	2233436;153885;1450025;1829162;1593	
-	-	Bradi2g38260.1	-	-	2727674;2209698;1425816;2061765;188	
-	-	Bradi2g36970.1	Os05g0135900	Sb09g002860.1	2889501;14758;1351[illegible]3532;688065	
-	-	-	Os05g0136200	Sb09g002900.1	505041;1885796;134[illegible]14;9826	
-	-	Bradi2g36940.1	-	-	1296380;2280486;63[illegible]20831;24946	
-	-	Bradi2g36930.1	-	Sb09g002890.1	281288;1741875;2276575;288497;17703	
-	-	Bradi2g36910.1	Os05g0136900	Sb09g002910.1	168249;	-
-	-	-	-	Sb09g002930.1	1062937;965193;195[illegible]	
-	-	Bradi2g36900.1	-	Sb09g002940.1	1083115;2180366;14[illegible];1320	
-	-	-	Os05g0137200	-	2160720;147249;1320[illegible]9;9066	
-	-	Bradi2g36890.1	Os05g0137300	Sb09g002950.1	1718625;1582525;68[illegible]	
-	-	-	-	Sb09g003000.1	2904216;1616109;	20203;
-	-	-	Os05g0137700	Sb09g002990.1	2904216;1616109;130[illegible]4;	
-	-	-	Os05g0137600	Sb09g002985.1	2716389;	-
1_0294	47,47	Bradi2g36860.3	Os05g0137400	Sb09g002960.1	2731998;1669831;203[illegible]	
-	-	Bradi2g36850.1	-	-	523838;1538595;502974;508942;209698	
-	-	Bradi2g36830.1	Os05g0113300	Sb09g001490.1	1993234;1438534;247480;2515662;1397	
-	-	Bradi2g36810.1	-	-	1800005;372299;1960161;2205777;4462	
-	-	Bradi2g36790.1	-	-	1248023;	21520;
-	-	-	-	Sb09g003020.1	1940546;807734;	-
-	-	Bradi2g36780.1	Os05g0138200	Sb09g003040.1	801830;732533;2465[illegible]98;	
-	-	Bradi2g36760.2	-	Sb09g003070.1	2588440;3045473;	-
-	-	-	-	Sb09g003080.1	2130511;1120689;	-
-	-	Bradi2g36750.1	-	-	32914;1733758;2954065;	
-	-	Bradi2g36740.1	-	-	68451;	19249;
-	-	Bradi2g36730.2	-	-	1764578;	-
-	-	-	Os05g0139200	-	676075;	-
-	-	-	-	Sb09g003090.1	68451;	19249;
-	-	Bradi2g36730.1	Os05g0140100	Sb09g003100.1	1607729;1742235;2887103;2963871;176	
1_0526	47,47	Bradi2g36710.1	Os05g0140500	Sb09g003130.1	2700159;2711235;17[illegible]1211624;	
-	-	Bradi2g36690.1	Os05g0140800	Sb09g003150.1	80856;2136671;2504[illegible]44;	
-	-	Bradi2g36680.1	-	-	2352207;	23921;
-	-	Bradi2g36670.1	-	Sb09g003170.1	971971;	-
-	-	Bradi2g36660.1	Os05g0141400	Sb09g003210.1	2599753;2564655;2235380;	
-	-	Bradi2g36650.1	Os05g0141500	-	1509309;865103;128[illegible]550938;	
3_1467	47,47	Bradi2g36640.1	Os05g0141700	Sb09g003230.1	2324871;	5390;

3_0404	50	-	-	-
3_0498	50	-	-	-
-	-	Bradi2g33360.1	-	Sb09g005900.1
-	-	-	-	Sb09g005890.1
-	-	-	-	Sb09g005870.1
-	-	-	Os05g0187800	-
-	-	Bradi2g33370.1	Os05g0187500	-
-	-	-	-	Sb09g005860.1
-	-	Bradi2g33380.1	Os05g0187100	Sb09g005840.1
-	-	Bradi2g33390.3	Os05g0187000	-
-	-	Bradi2g33420.1	-	-
-	-	Bradi2g33450.3	Os05g0186300	Sb09g005810.1
-	-	Bradi2g33470.1	-	-
-	-	Bradi2g33480.1	Os05g0185800	Sb09g005790.1
-	-	-	Os05g0185700	Sb09g018040.1
-	-	Bradi2g33490.1	Os05g0185600	Sb09g005780.1
-	-	-	-	Sb09g005750.1
-	-	Bradi2g33500.1	-	-
-	-	Bradi2g33510.1	Os05g0184901	-
-	-	Bradi2g33530.1	-	-
-	-	-	-	Sb09g005720.1
-	-	Bradi2g33550.1	Os05g0182900	Sb09g005695.1
-	-	Bradi2g33560.2	Os05g0182700	Sb09g005660.1
-	-	Bradi2g33570.1	-	-
-	-	Bradi2g33580.1	Os05g0182600	Sb09g005650.1
-	-	Bradi2g33580.2	-	-
-	-	Bradi2g33610.1	-	Sb09g005630.1
-	-	Bradi2g33620.2	Os05g0182100	Sb09g005640.1
-	-	Bradi2g33640.1	-	-
-	-	-	-	Sb09g005600.1
-	-	Bradi2g33670.1	Os05g0181200	Sb09g005590.1
-	-	-	-	Sb09g005580.1
-	-	Bradi2g33680.2	Os05g0181000	Sb09g005570.1
-	-	Bradi2g33900.1	-	-
-	-	Bradi2g33910.1	-	-
-	-	-	-	Sb09g005550.1
-	-	Bradi2g33920.1	Os05g0180400	Sb09g005540.1
-	-	Bradi2g33940.1	Os05g0179900	Sb09g005510.1
-	-	Bradi2g33950.1	-	Sb09g005500.1
-	-	Bradi2g33960.1	Os05g0179800	Sb09g005490.1
-	-	Bradi2g33980.1	Os05g0179300	Sb09g005480.1
-	-	Bradi2g34000.1	-	-
-	-	-	Os05g0179100	-
-	-	Bradi2g34010.5	Os05g0179000	Sb09g005450.1
-	-	Bradi2g34020.1	Os05g0178900	Sb09g005440.1
-	-	Bradi2g34030.2	Os05g0178600	Sb09g005430.1
-	-	Bradi2g34040.2	Os05g0178400	Sb09g005420.1
-	-	-	-	Sb09g005390.1
-	-	-	Os05g0178000	-
-	-	Bradi2g34050.1	-	-
-	-	-	-	Sb09g005370.1
-	-	Bradi2g34070.1	Os05g0177500	Sb09g005360.1
-	-	-	Os05g0177400	-
-	-	Bradi2g34080.2	-	Sb09g004460.1
-	-	Bradi2g34100.1	-	-
-	-	Bradi2g34110.1	Os05g0176800	Sb09g005350.1
-	-	Bradi2g34120.2	-	-
-	-	-	Os05g0176600	Sb09g005320.1
3_0762	50	Bradi2g34160.1	Os05g0176500	Sb09g005310.1
-	-	-	-	Sb09g005300.1
-	-	-	Os05g0176400	-
-	-	Bradi2g34170.1	-	-
-	-	Bradi2g34200.1	-	-
-	-	Bradi2g34240.1	Os05g0176100	Sb09g005280.1
-	-	Bradi2g34250.1	Os05g0171900	Sb09g005270.1
-	-	Bradi2g34270.1	Os05g0171300	Sb09g005250.1
-	-	Bradi2g34290.1	Os05g0171000	-
-	-	Bradi2g34300.2	-	Sb09g005210.1
-	-	-	-	Sb09g005195.1
-	-	Bradi2g34340.1	Os05g0170300	Sb09g005150.1
-	-	Bradi2g34350.1	Os05g0170000	Sb09g005120.1
-	-	Bradi2g34360.1	-	Sb09g005110.1
-	-	Bradi2g34390.1	Os05g0169400	Sb09g005090.1
-	-	-	-	Sb09g005080.1
-	-	Bradi2g34400.1	Os05g0169300	Sb09g005070.1
-	-	-	Os05g0169200	-
-	-	Bradi2g34430.1	-	-
-	-	Bradi2g34440.3	Os05g0169100	-
1_1162	50	Bradi2g34460.2	Os05g0168800	Sb09g005020.1
-	-	Bradi2g34470.1	Os05g0168700	Sb09g005010.1
-	-	-	Os05g0166600	-
-	-	-	Os05g0165900	Sb09g001740.1
-	-	-	Os05g0165500	-

GenomeZipper pipeline



454 reads/ contigs

model genomes

marker, flcDNAs, ESTs

Modul 1: Repeat Masking (Vmatch)

Modul 2: Sequence Homology (BLAST)

Modul 3: Synteny Detection (chromoWIZ)

Modul 4: GenomeZipper

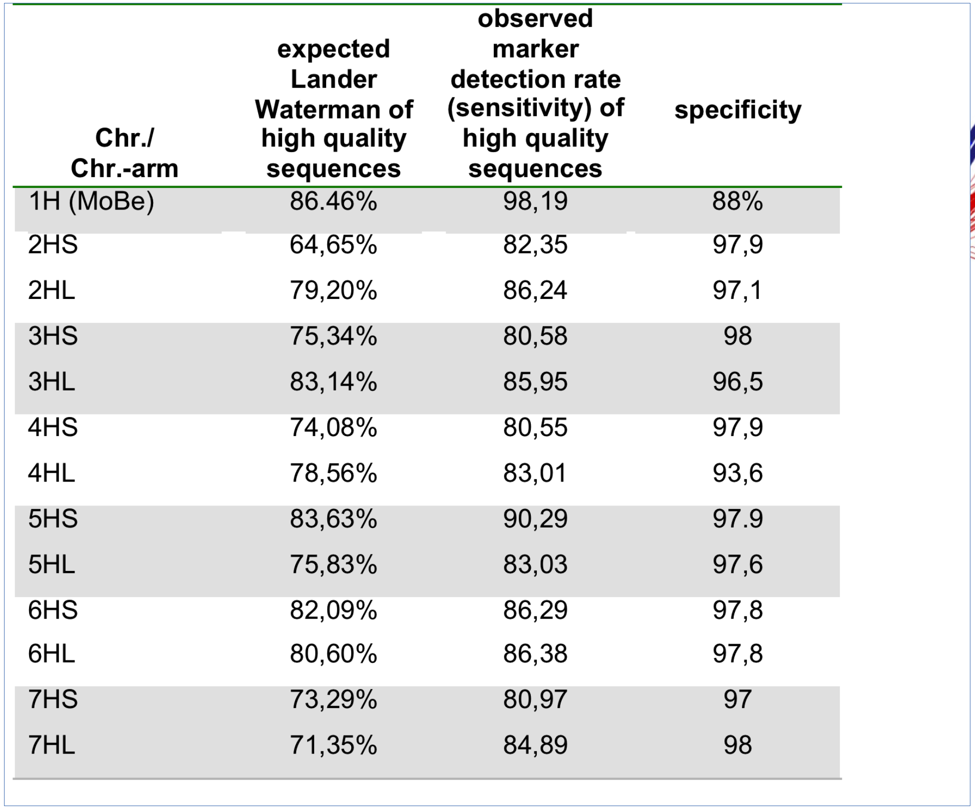
Ordered virtual gene map

GenomeZipper: Barley Chromosomes

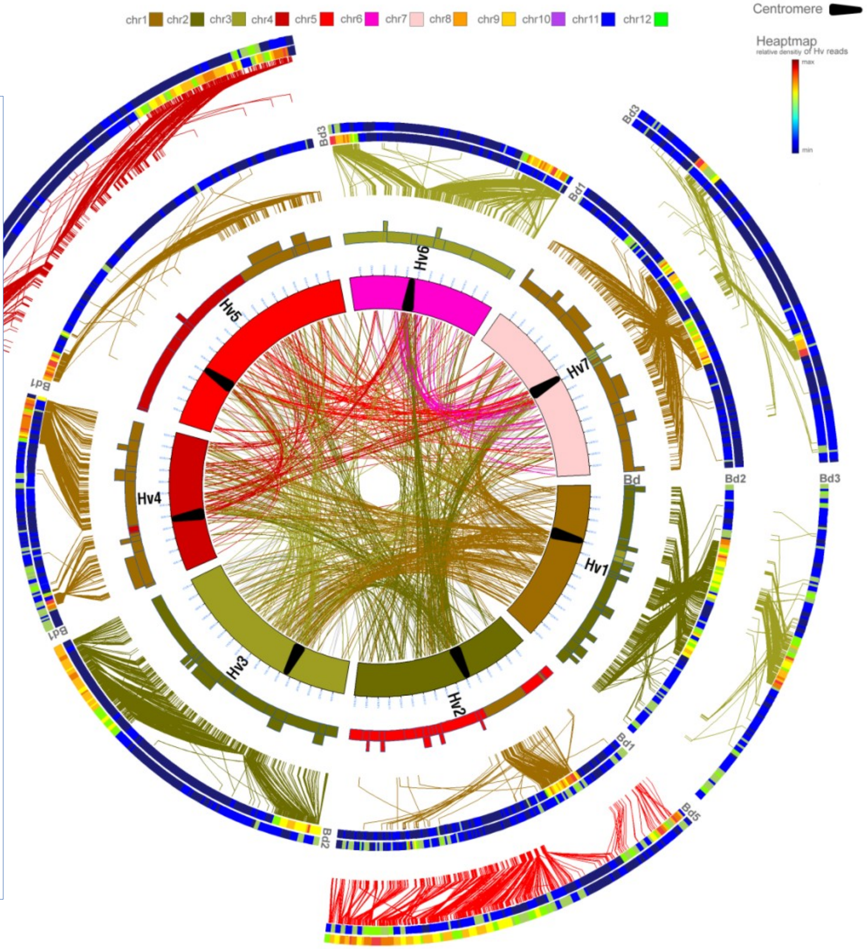
Data Sets	1H mobe	2H	3H	4H	5H	6H	7H	All
# nonredundant anchored gene loci	3,331	3,616	3,394	2,709	3,208	2,304	3,204	21,766
% markers with associated gene from ref. genome(s)	63.25	61.1	66.29	69.1	60.77	58.75	53.9	61.72
# matched barley fl-cDNAs	1,676	1,619	1,628	1,255	1,474	1,058	1,395	10,105
# nonredundant sequence reads & array hybridization probes	52,704	31,294	32,078	22,644	27,197	20,943	24,423	211,283
# nonredundant ESTs	3,543	3,678	3,392	2,605	3,354	2,387	3,120	22,079
# Brachypodium genes	2,141	2,379	2,363	1,876	2,159	1,588	1,915	14,421
# rice genes	1,845	2,073	2,016	1,614	1,576	1,348	1,621	12,093
# sorghum genes	1,833	1,946	2,039	1,284	1,695	1,369	1,721	11,887

Mayer et al. (2011): Unlocking the Barley Genome by Chromosomal and Comparative Genomics; The Plant Cell; DOI: 10.1105/tpc.110.082537

Barley – a high resolution genome scaffold



Chr./ Chr.-arm	expected Lander Waterman of high quality sequences	observed marker detection rate (sensitivity) of high quality sequences	specificity
1H (MoBe)	86.46%	98,19	88%
2HS	64,65%	82,35	97,9
2HL	79,20%	86,24	97,1
3HS	75,34%	80,58	98
3HL	83,14%	85,95	96,5
4HS	74,08%	80,55	97,9
4HL	78,56%	83,01	93,6
5HS	83,63%	90,29	97.9
5HL	75,83%	83,03	97,6
6HS	82,09%	86,29	97,8
6HL	80,60%	86,38	97,8
7HS	73,29%	80,97	97
7HL	71,35%	84,89	98



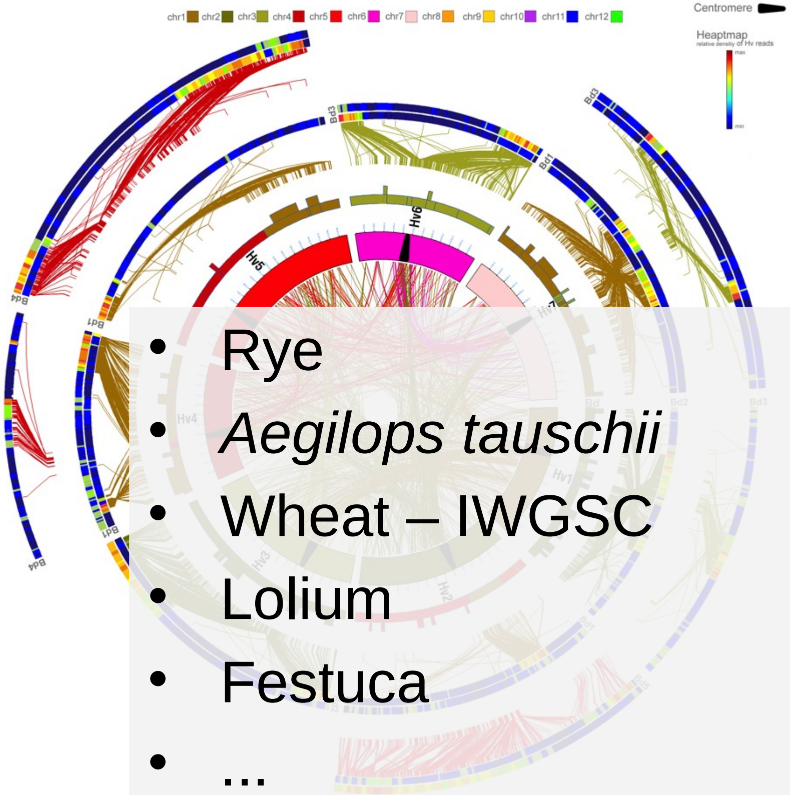
Barley Genome Zipper summary

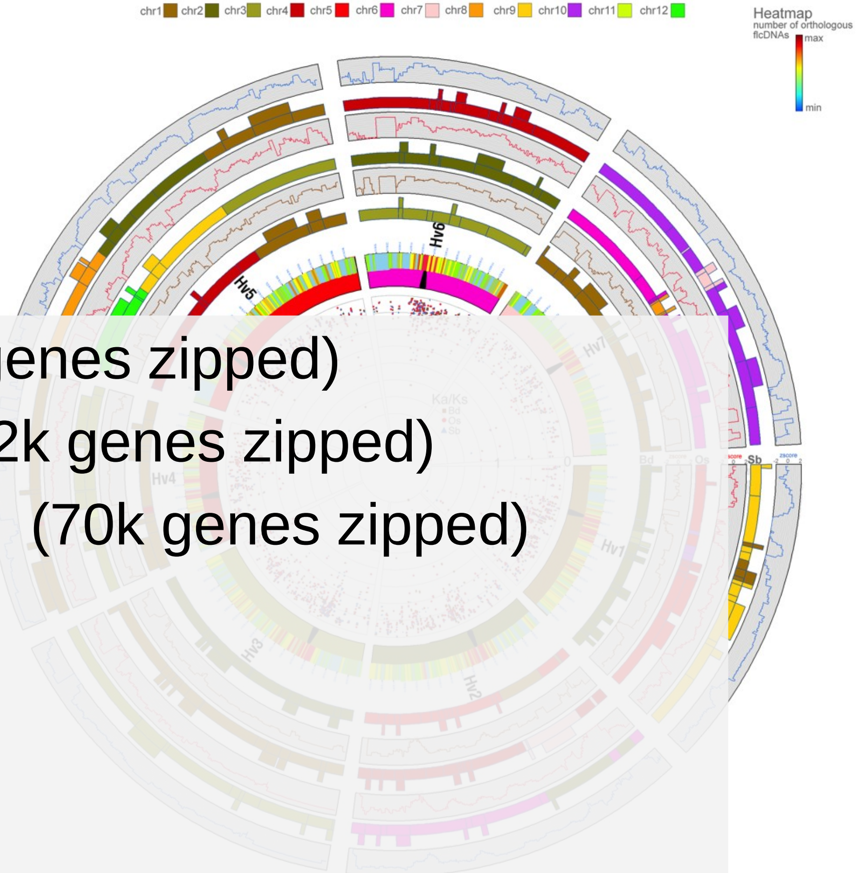
- **22k barley genes** sequence tagged **positionally ordered and** in part associated with flcDNA & EST
- **Additional 6k genes with chromosome arm assignment**
- Resolution of appr. 0,05 (0,1) cM; 20 loci (9,3 fl-cDNAs) per cM
- >3000 (14%) genes are located in low/non-recombining regions
- All but 9 ordered and assigned to short and long arm respectively

Shortcomings:

- Can‘t resolve small local rearrangements
- Can‘t position genes that are out of syntenic context
- Pseudogenes, tandem duplicates, ...

A powerful shortcut towards an ordered gene map of the huge *Triticeae* genomes





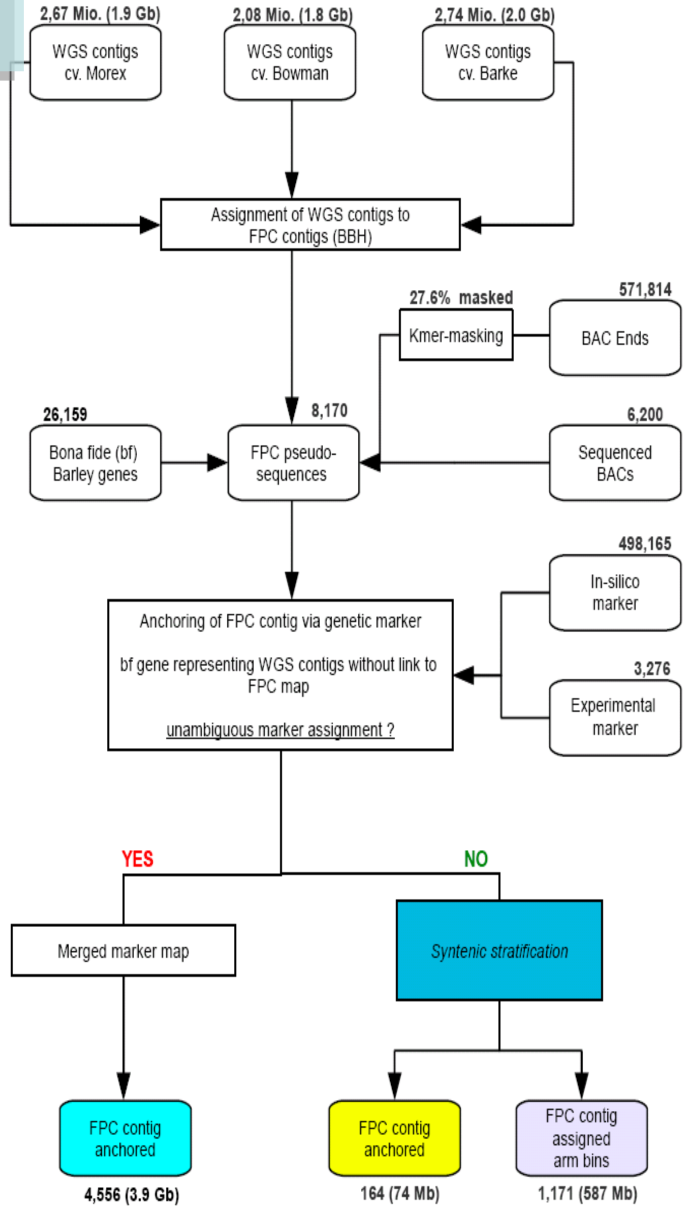
- Rye (23k genes zipped)
- *Aegilops tauschii* (22k genes zipped)
- Wheat – IWGSC (70k genes zipped)
- Lolium
- Festuca
- ...

Genome stratification in barley

Combined genetic and physical map
build scaffold

+ Sequence enriched via 80x
WholeGenomeShotgun
6,200 BACs,
570,000 BacEndSeqs
500,000 genetic markers
250 Gb RNA Seq

=>3,9Gb (76%) anchored
+ 650 Mb (13%) chr. arm associated



2,67 Mio. (1.9 Gb)
WGS contigs cv. Morex

2,08 Mio. (1.8 Gb)
WGS contigs cv. Bowman

2,74 Mio. (2.0 Gb)
WGS contigs cv. Barke

Assignment of WGS contigs to FPC contigs (BBH)

27.6% masked
Kmer-masking

571,814
BAC Ends

26,159
Bona fide (bf) Barley genes

8,170
FPC pseudo-sequences

6,200
Sequenced BACs

498,165
In-silico marker

3,276
Experimental marker

Anchoring of FPC contig via genetic marker

bf gene representing WGS contigs without link to FPC map

unambiguous marker assignment ?

YES

NO

Merged marker map

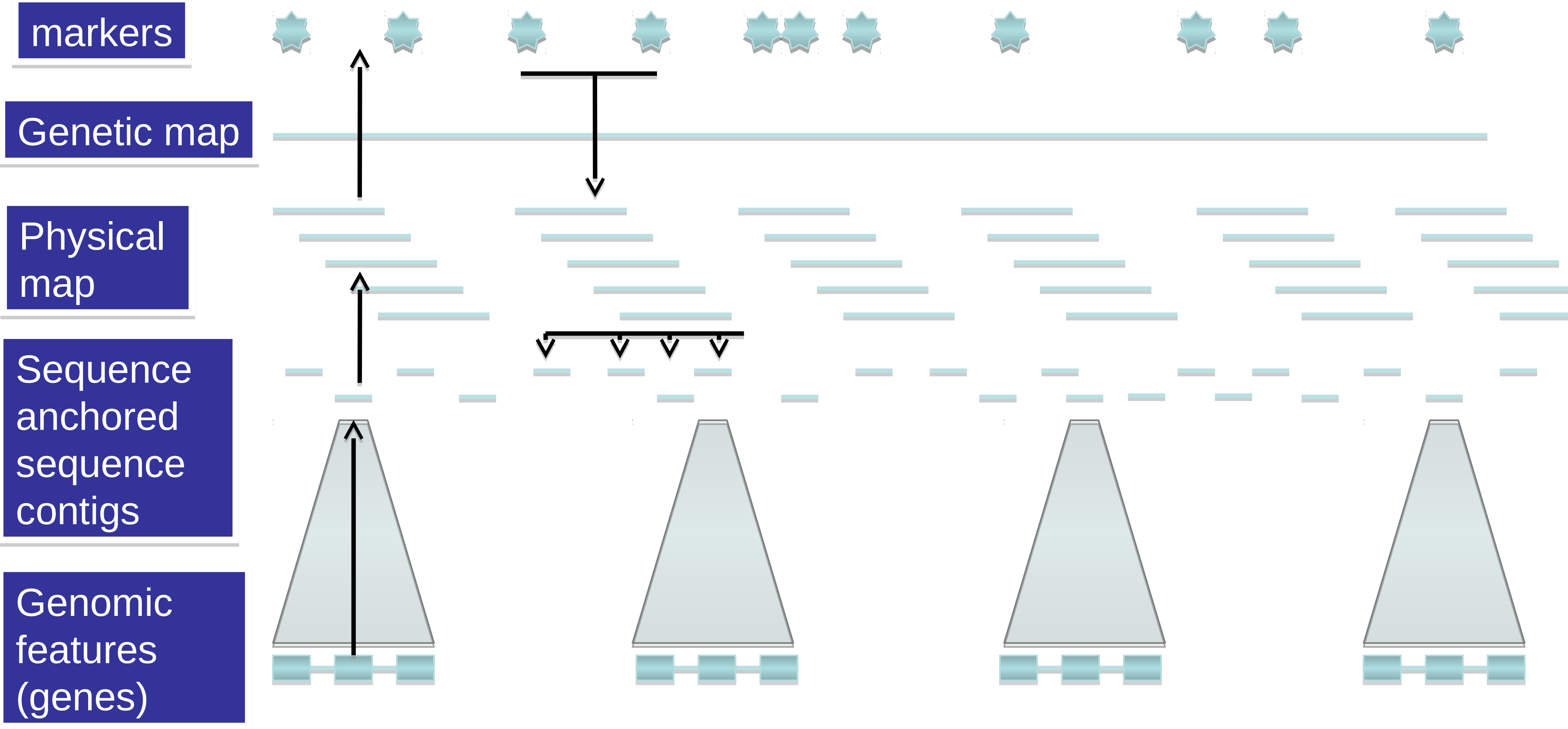
Syntenic stratification

FPC contig anchored
4,556 (3.9 Gb)

FPC contig anchored
164 (74 Mb)

FPC contig assigned arm bins
1,171 (587 Mb)

Different hierarchies of feature connection and different layers to start the navigation



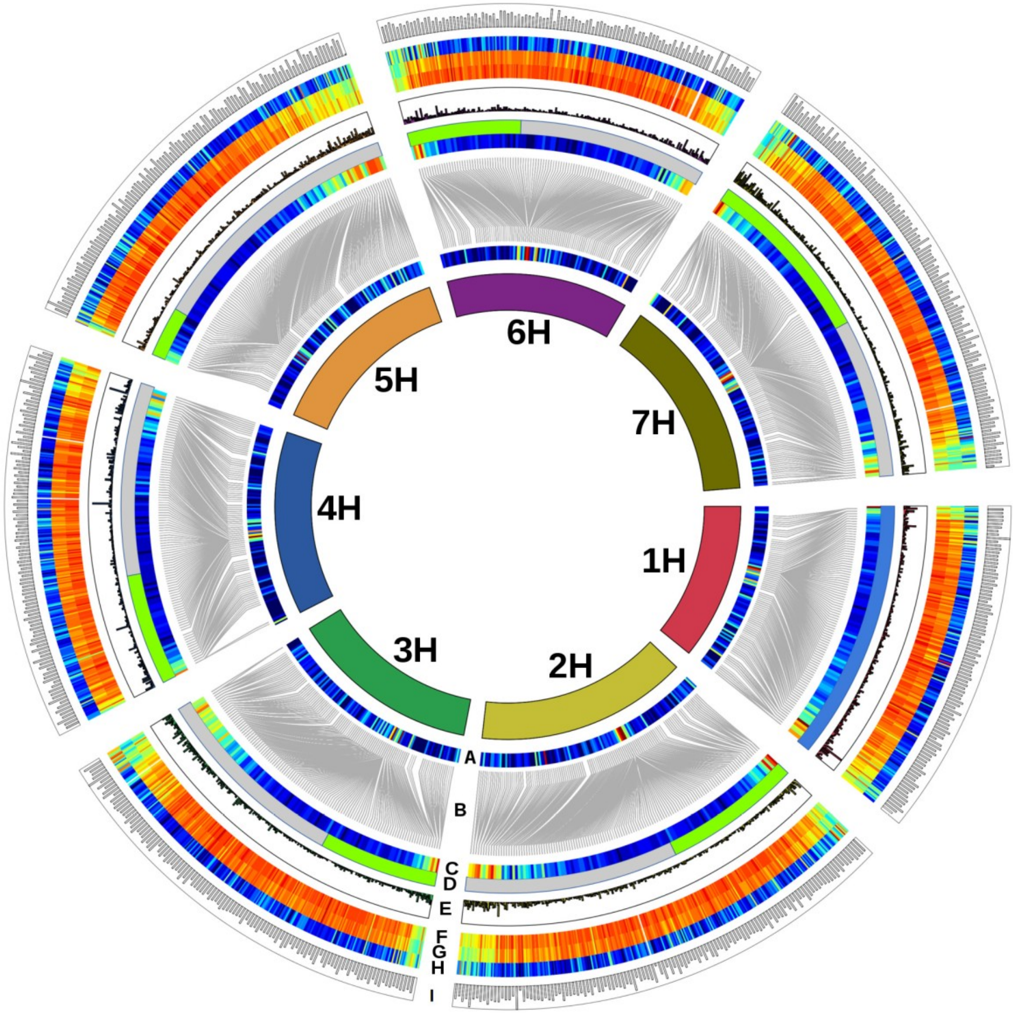
Genome stratification in barley cont.

The Barley **Gene-ome**:

A **physical**, **genetic** and

functional sequence

assembly



More on barley physical and genetic map (integration) in seperate presentation!

Gene prediction in barley

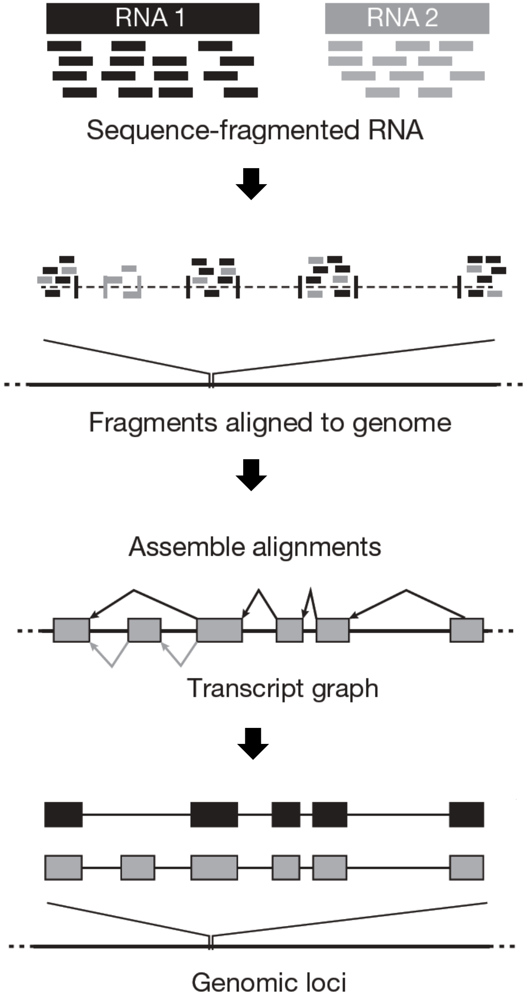
Barley RNA-seq data from SCRI

Platform	Read Length [bp]	Paired End	Material	Genotype	Reads [#]	Sequence [bp]
Illumina GA2 SE	76	no	4 days germination embryo	Morex	23,250,889	1,767,067,564
Illumina GA2 SE	76	no	4 days germination embryo	Quench	26,946,706	2,047,949,656
Illumina GA2 SE	76	no	4 days germination embryo	Optic	23,252,182	1,767,165,832
Illumina GA2 SE	76	no	4 days germination embryo	Barke	25,663,186	1,950,402,136
Illumina GA2 SE	76	no	4 days germination embryo	Tocada	23,868,881	1,814,034,956
Illumina GA2 SE	76	no	4 days germination embryo	Betzes	22,204,022	1,687,505,672
Illumina GA2 SE	76	no	4 days germination embryo	Sergeant	24,480,462	1,860,515,112
					Σ 169,666,328	**Σ 12,894,640,928**

=> Barley reference assembly sequence from IPK

Morex 50x WGS assembly 3, repeat masked

Gene prediction in barley



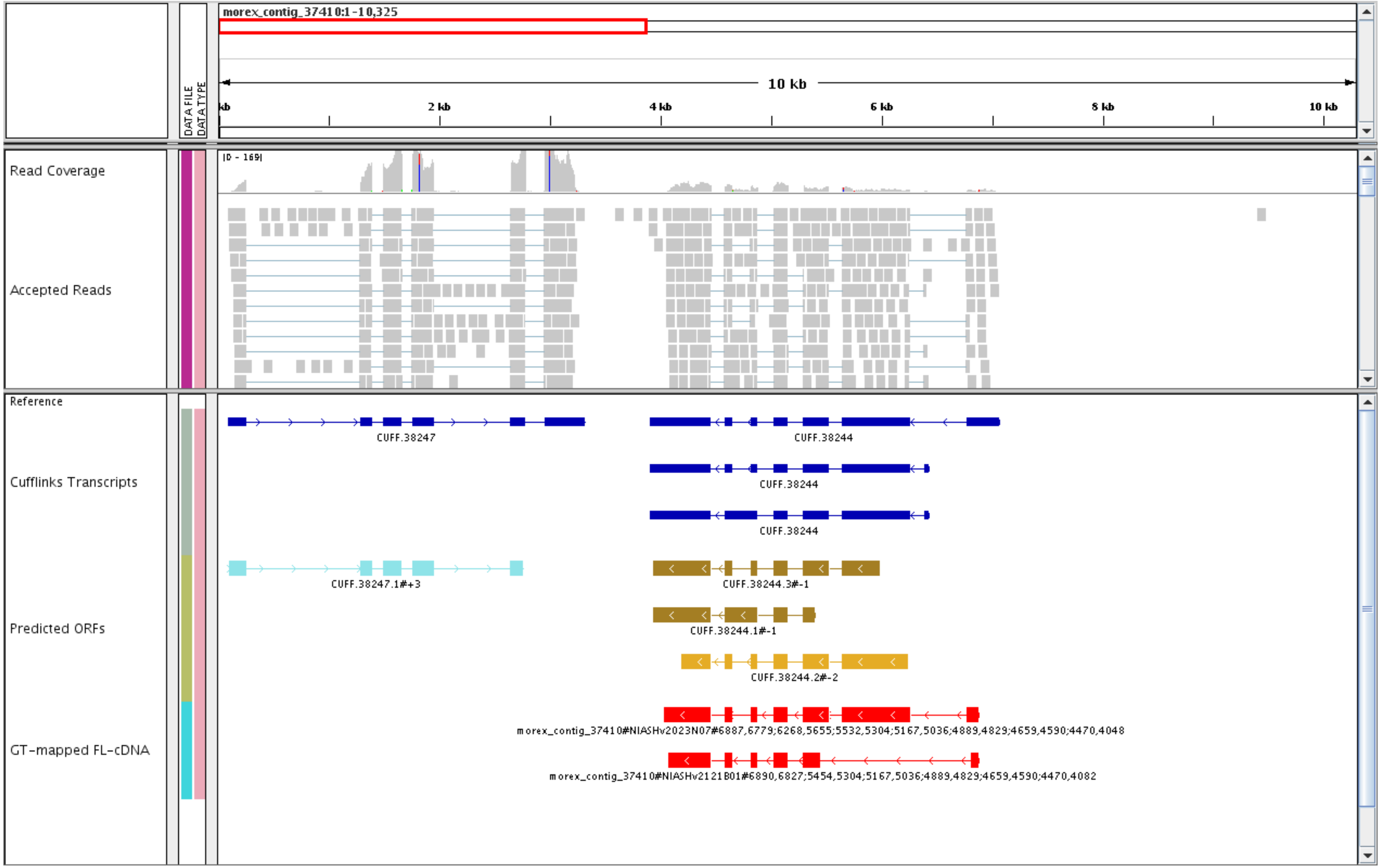
adapted from Garber et al 2011

Bowtie/Tophat

Alignment of reads to reference genome and identification of splice junctions

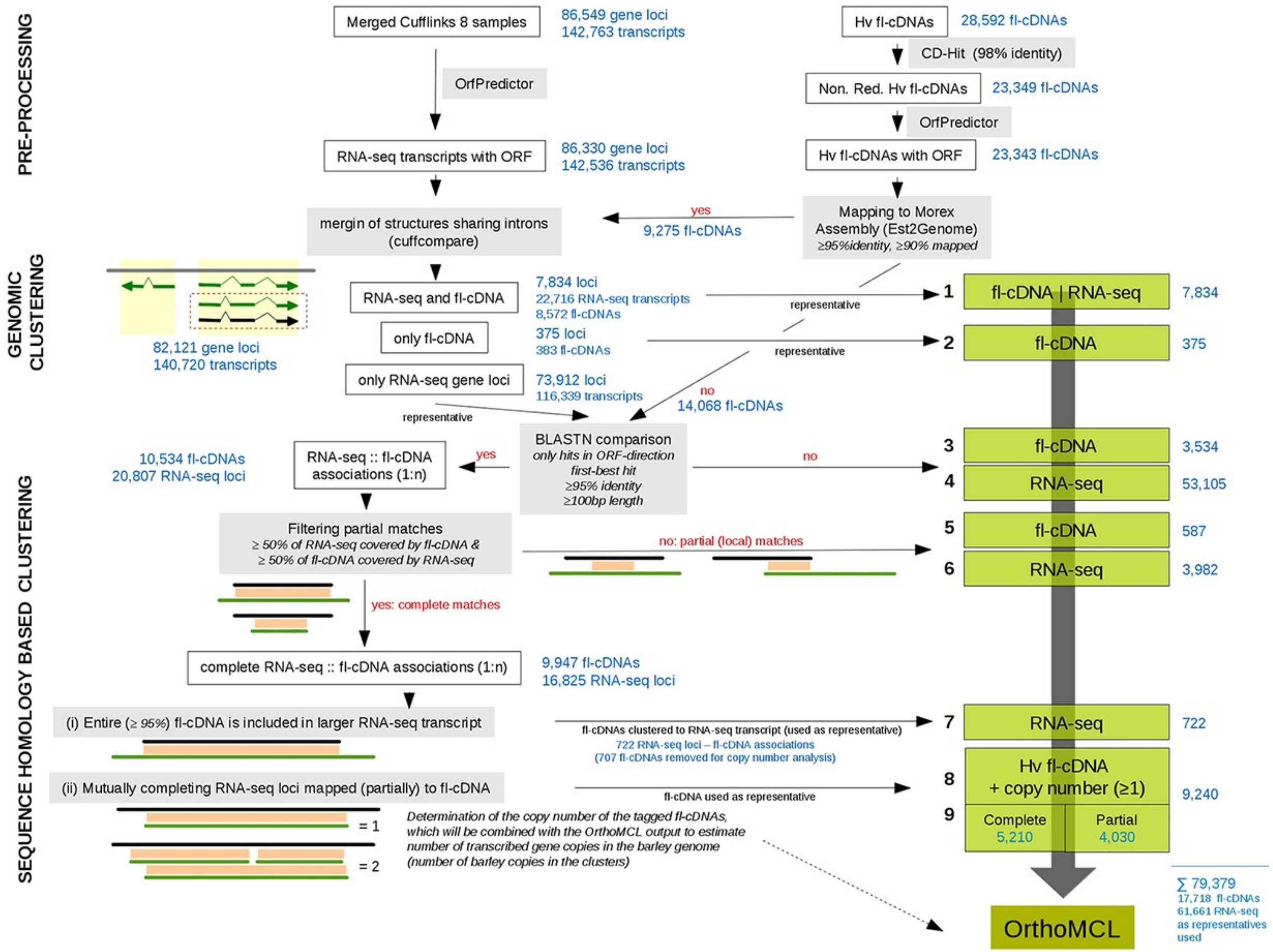
Cufflinks

Identification of genes and transcripts based on the location of the alignments of spliced reads

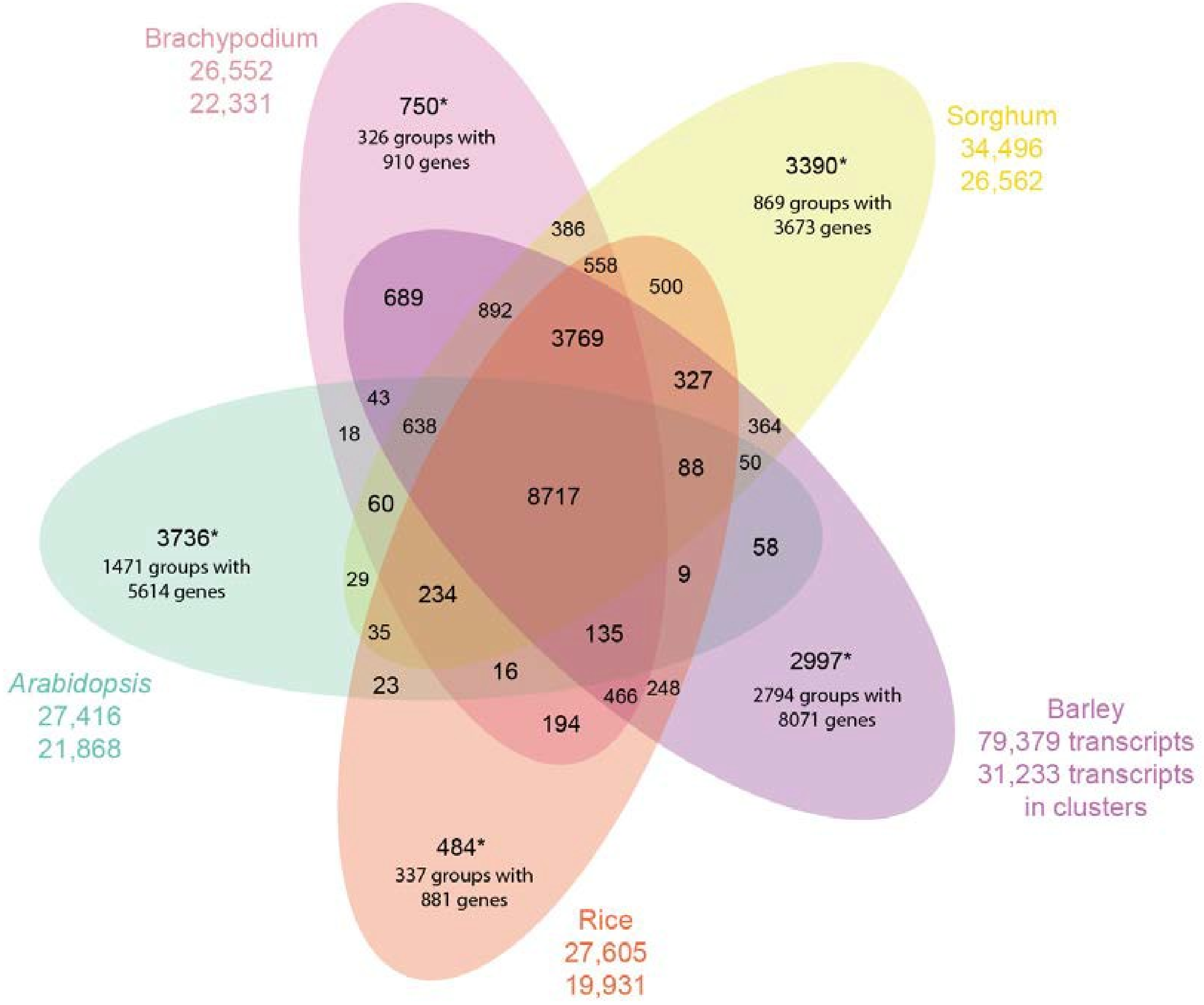


=> 86,330 barley CuffLink loci

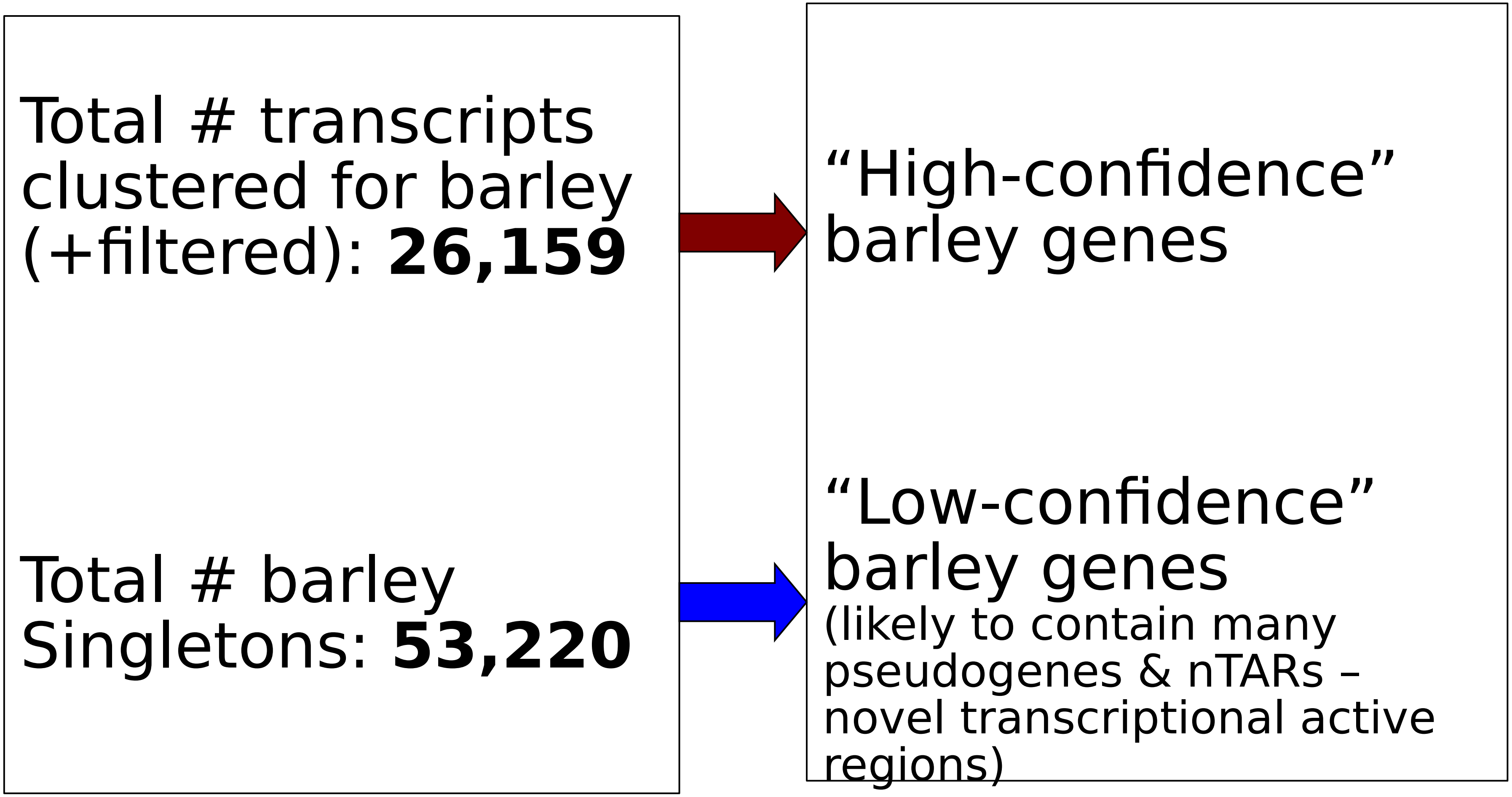
Gene prediction in barley - pipeline



PRE-PROCESSING
GENOMIC CLUSTERING
SEQUENCE HOMOLOGY BASED CLUSTERING
Merged Cufflinks 8 samples
86,549 gene loci
142,763 transcripts
OrfPredictor
RNA-seq transcripts with ORF
86,330 gene loci
142,536 transcripts
Hv fl-cDNAs
28,592 fl-cDNAs
CD-Hit (98% identity)
Non. Red. Hv fl-cDNAs
23,349 fl-cDNAs
OrfPredictor
Hv fl-cDNAs with ORF
23,343 fl-cDNAs
Mapping to Morex Assembly (Est2Genome)
≥95%identity, ≥90% mapped
yes
9,275 fl-cDNAs
mergin of structures sharing introns (cuffcompare)
82,121 gene loci
140,720 transcripts
RNA-seq and fl-cDNA
7,834 loci
22,716 RNA-seq transcripts
8,572 fl-cDNAs
representative
only fl-cDNA
375 loci
383 fl-cDNAs
representative
only RNA-seq gene loci
73,912 loci
116,339 transcripts
no
14,068 fl-cDNAs
representative
BLASTN comparison
only hits in ORF-direction
first-best hit
≥95% identity
≥100bp length
yes
no
10,534 fl-cDNAs
20,807 RNA-seq loci
RNA-seq :: fl-cDNA associations (1:n)
Filtering partial matches
≥ 50% of RNA-seq covered by fl-cDNA &
≥ 50% of fl-cDNA covered by RNA-seq
no: partial (local) matches
yes: complete matches
complete RNA-seq :: fl-cDNA associations (1:n)
9,947 fl-cDNAs
16,825 RNA-seq loci
(i) Entire (≥ 95%) fl-cDNA is included in larger RNA-seq transcript
fl-cDNAs clustered to RNA-seq transcript (used as representative)
722 RNA-seq loci – fl-cDNA associations
(707 fl-cDNAs removed for copy number analysis)
(ii) Mutually completing RNA-seq loci mapped (partially) to fl-cDNA
fl-cDNA used as representative
= 1
= 2
Determination of the copy number of the tagged fl-cDNAs, which will be combined with the OrthoMCL output to estimate number of transcribed gene copies in the barley genome (number of barley copies in the clusters)
1 fl-cDNA | RNA-seq 7,834
2 fl-cDNA 375
3 fl-cDNA 3,534
4 RNA-seq 53,105
5 fl-cDNA 587
6 RNA-seq 3,982
7 RNA-seq 722
8 Hv fl-cDNA + copy number (≥1) 9,240
9 Complete 5,210
Partial 4,030
∑ 79,379
17,718 fl-cDNAs
61,661 RNA-seq
as representatives used
OrthoMCL


Brachypodium
26,552
22,331
750*
326 groups with
910 genes
Sorghum
34,496
26,562
3390*
869 groups with
3673 genes
386
558
500
689
892
3769
327
43
18
638
364
50
88
8717
60
58
3736*
1471 groups with
5614 genes
29
234
9
35
135
23
16
2997*
2794 groups with
8071 genes
466
248
194
Arabidopsis
27,416
21,868
Barley
79,379 transcripts
31,233 transcripts
in clusters
484*
337 groups with
881 genes
Rice
27,605
19,931

Gene prediction in barley - results



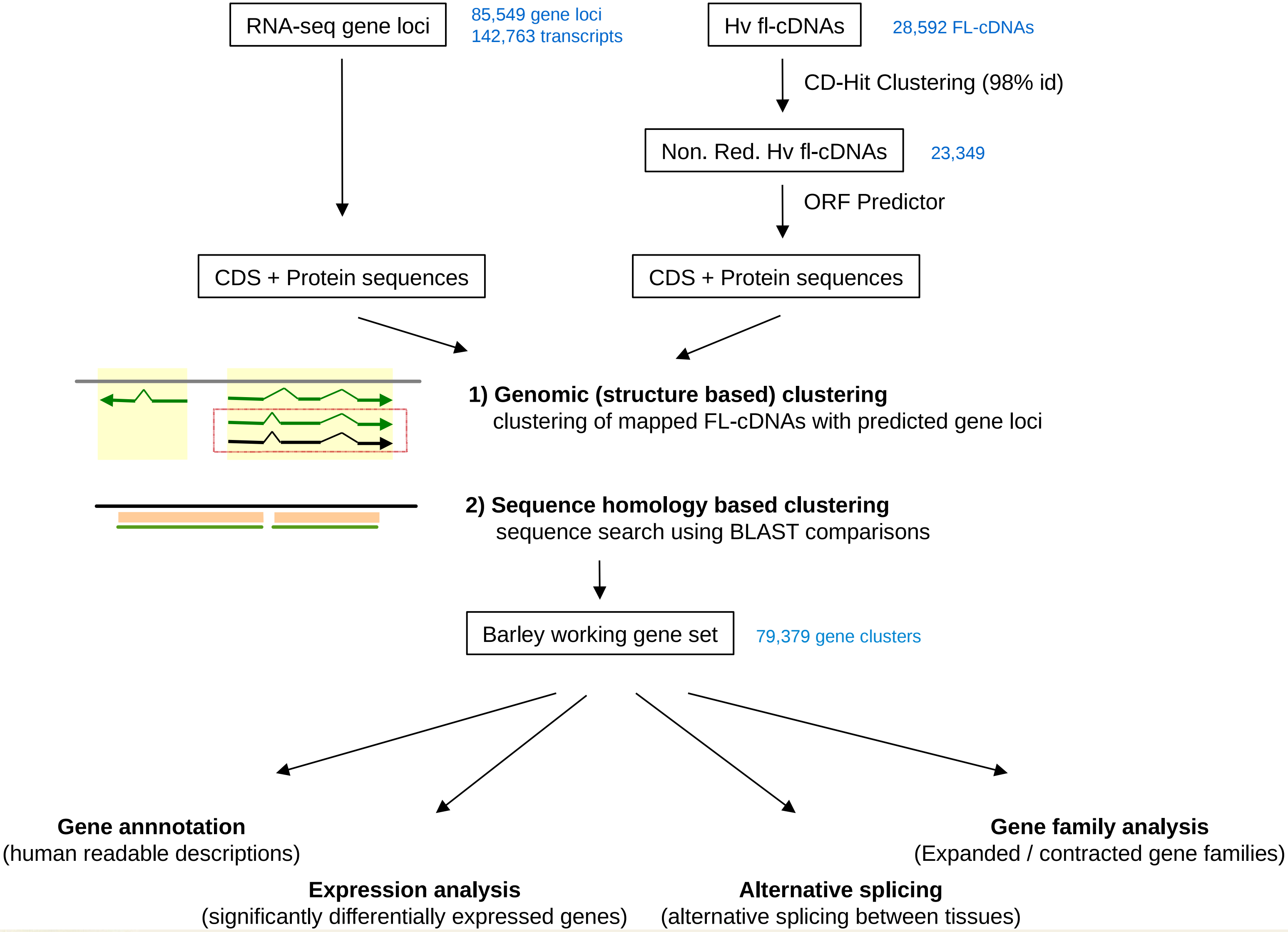
Total # transcripts clustered for barley (+filtered): **26,159**

Total # barley Singletons: **53,220**

“High-confidence” barley genes

“Low-confidence” barley genes
(likely to contain many pseudogenes & nTARs – novel transcriptional active regions)

Gene prediction in barley - summary



RNA-seq gene loci

85,549 gene loci
142,763 transcripts

Hv fl-cDNAs

28,592 FL-cDNAs

CD-Hit Clustering (98% id)

Non. Red. Hv fl-cDNAs

23,349

ORF Predictor

CDS + Protein sequences

CDS + Protein sequences

1) Genomic (structure based) clustering
clustering of mapped FL-cDNAs with predicted gene loci

2) Sequence homology based clustering
sequence search using BLAST comparisons

Barley working gene set

79,379 gene clusters

Gene annnotation
(human readable descriptions)

Expression analysis
(significantly differentially expressed genes)

Alternative splicing
(alternative splicing between tissues)

Gene family analysis
(Expanded / contracted gene families)

acknowledgements

MIPS
Thomas Nussbaumer
Matthias Pfeifer
Klaus Mayer
Mihaela Martis
Sapna Sharma
Tsuyoshi Tanaka
Heidrun Gundlach
Stephan Roessner
Michael Seidel
Georg Haberer

IPK
Nils Stein
Andreas Graner
B Steuernagel
Uwe Scholz

FLI Jena
Stefan Taudien
Marius Felder
Andreas Petzold
Matthias Platzer

IEB Olomouc
Hana Simkova
Pavla Suchankova
Jaroslav Dolezel

WZW/ Technical University Munich
Eva Bauer
Grit Haseneyer
Chris Schön

The IWGSC

SCRI
Pete Hedley
Hui Liu
Robbie Waugh

NIAS/Okayama
Takashi Matsumoto
Tsuyoshi Tanaka
Takeshi Itoh
Kaz Sato et al.

The UK Wheat Consortium
Mike Bevan
Neil Hall
Anthony Hall
Keith Edwards
Rachel Brenchley

The IBSC

CSHL
Dick McCombie

UC Davis & USDA Albany
Jan Dvorak
Mincheng Luo
Olin Anderson

Kansas State University
Bikram Gill
Sunish Segal

EBI
Paul Kersey
Dan Bolser

Bundesministerium für Bildung und Forschung
gabi
Deutsche Forschungsgemeinschaft DFG
SEVENTH FRAMEWORK PROGRAMME
Triticeae Genome