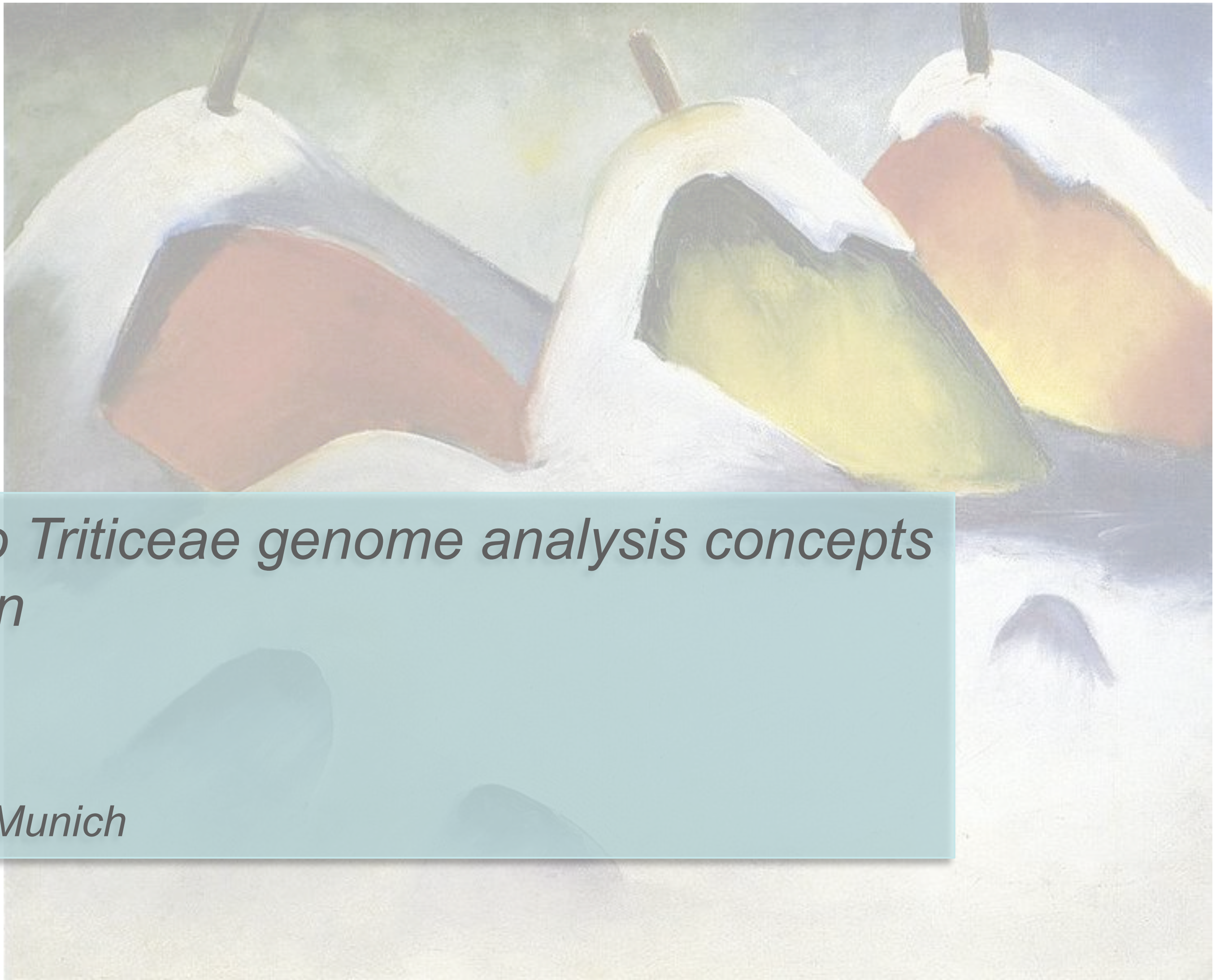
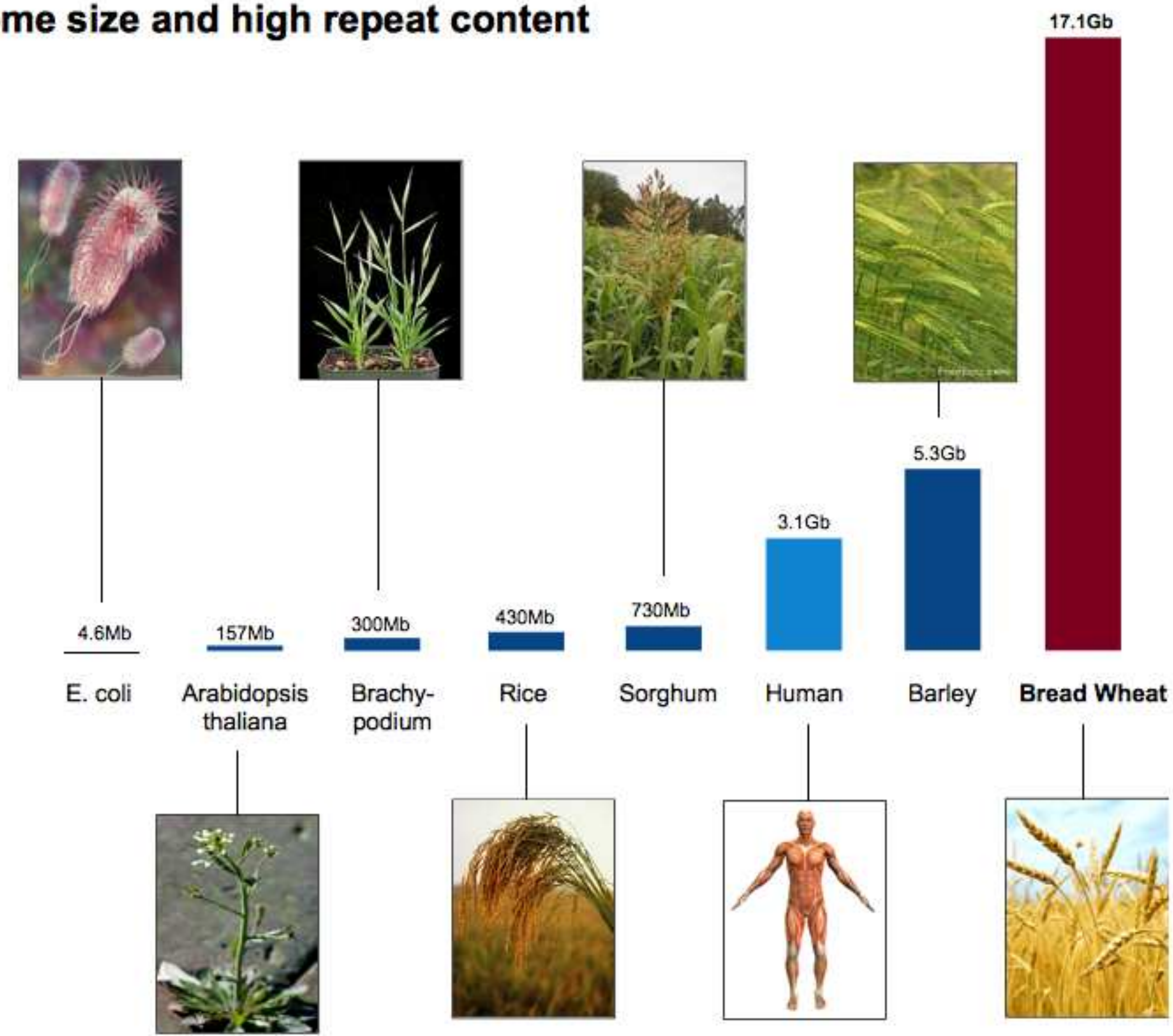


An Introduction into Triticeae genome analysis concepts and data generation

Klaus Mayer
MIPS, Helmholtz Center Munich

The Challenge

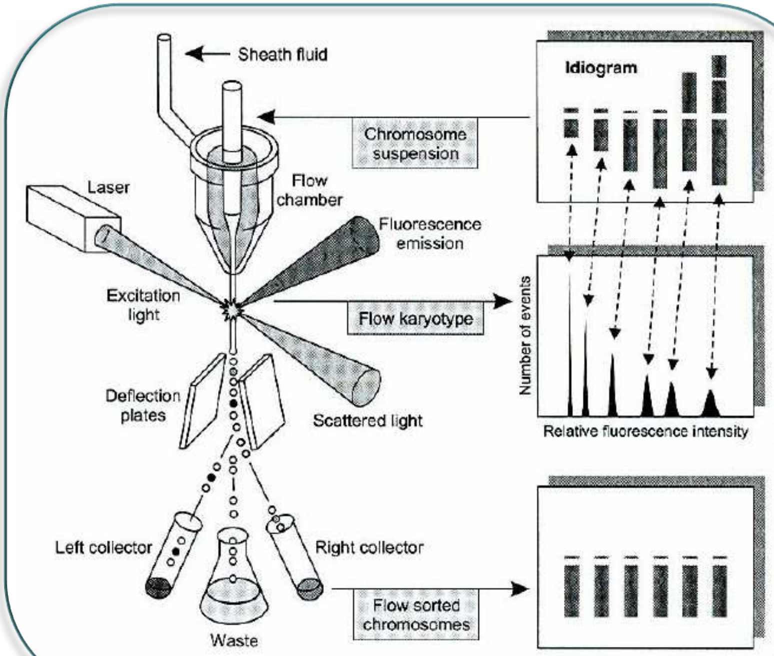
- ► **Big genome size and high repeat content**

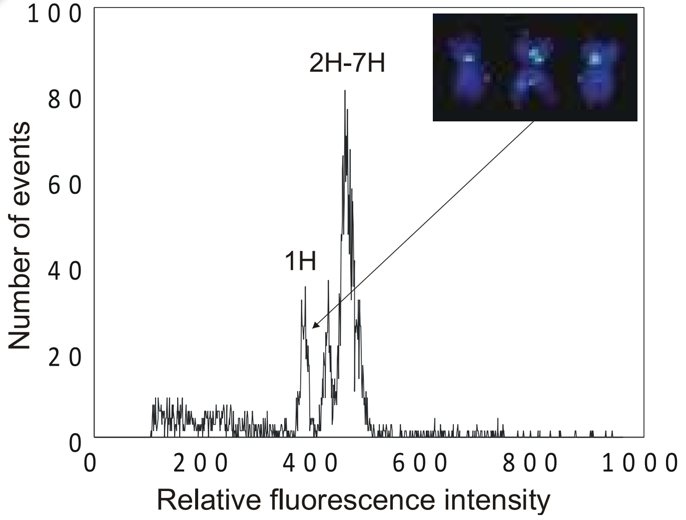


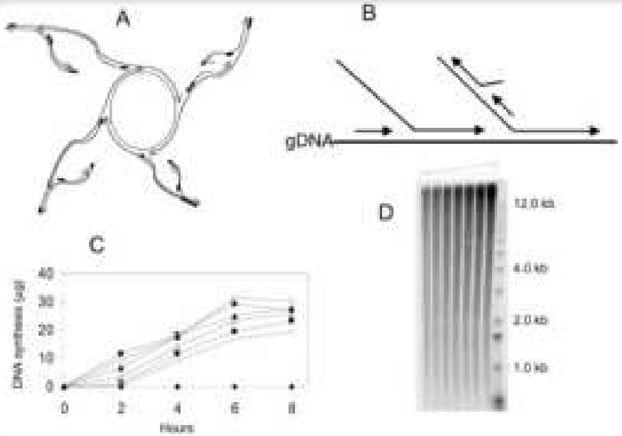
Outline

- Barley: genome zipping and genome stratification puts the pieces together
- Accessing wheat by WGS using „*in silico* exon capture“
- Rye, a highly rearranged *Triticeae* genome unzipped
- Rye B‘s: the path of the aliens

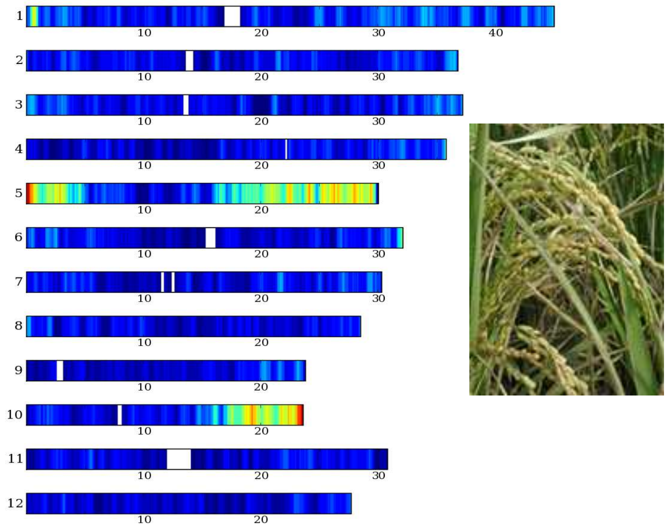
Reduction of complexity by chromosome sorting

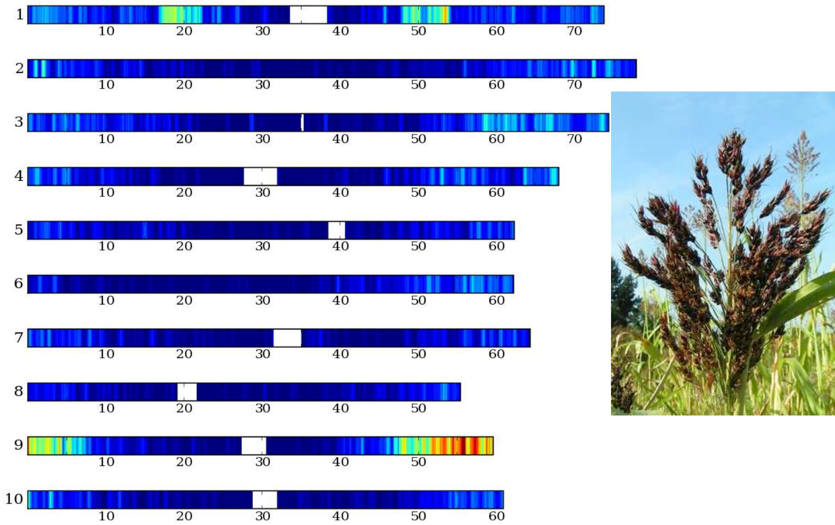


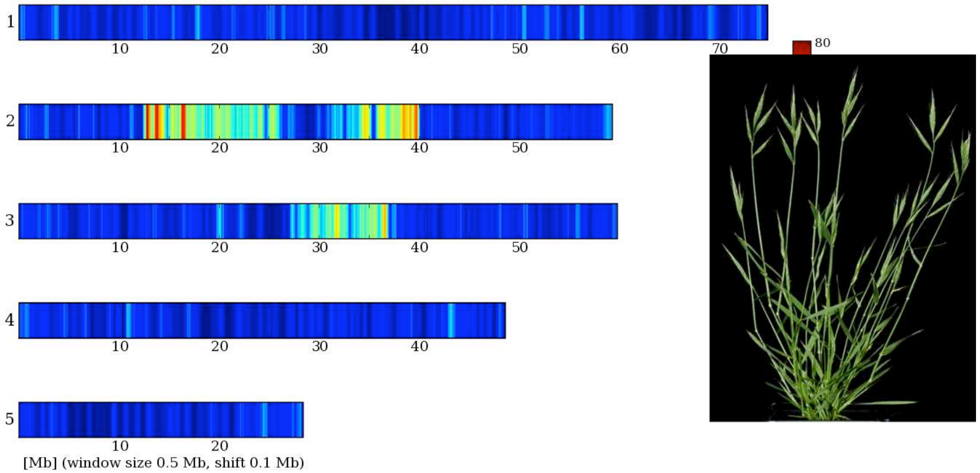




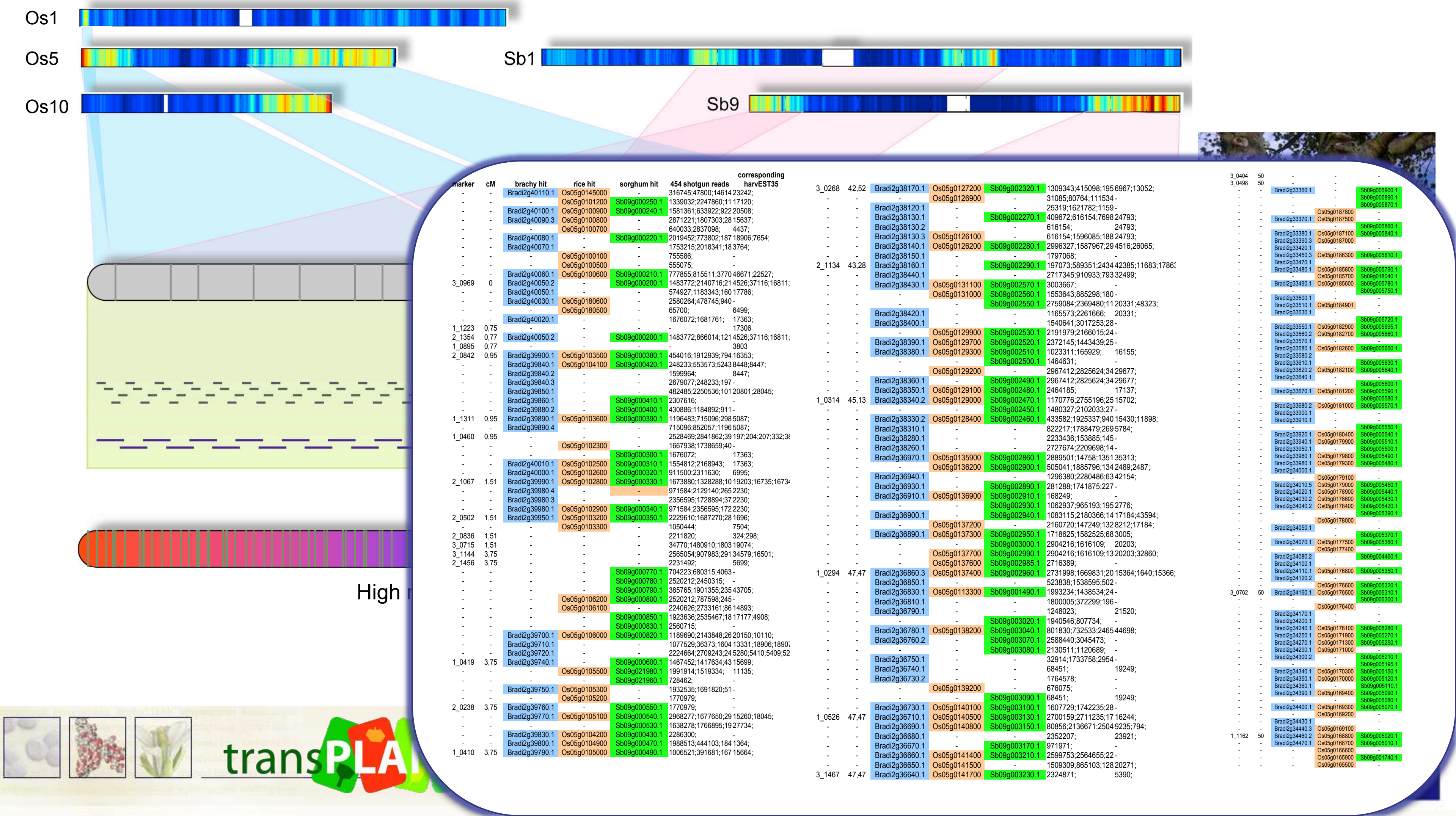
Barley vs *Brachypodium*, *Sorghum* and rice
-Synteny on a per gene resolution-





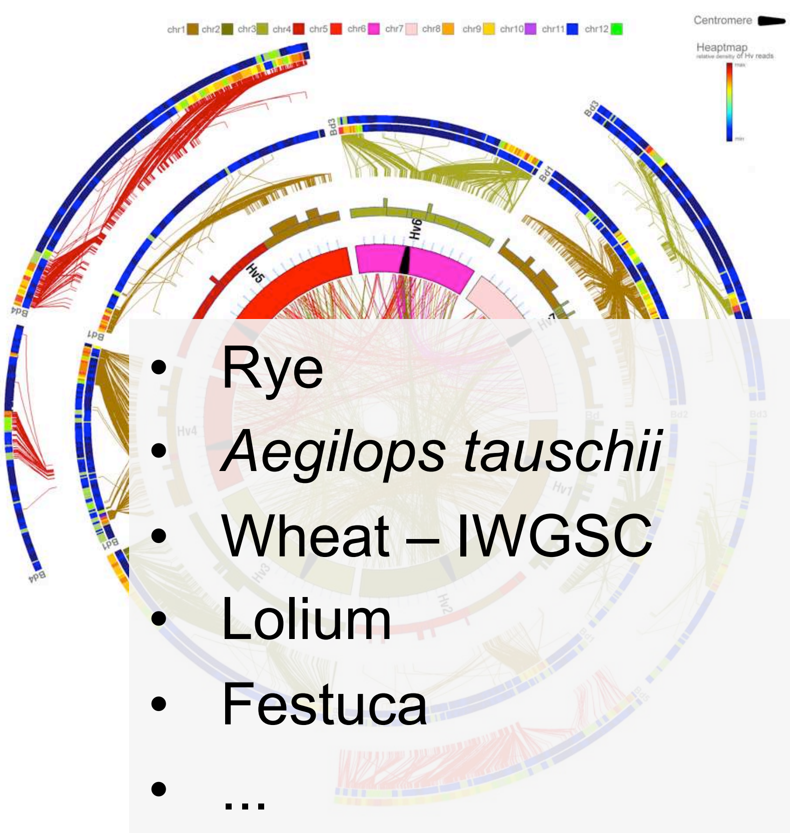


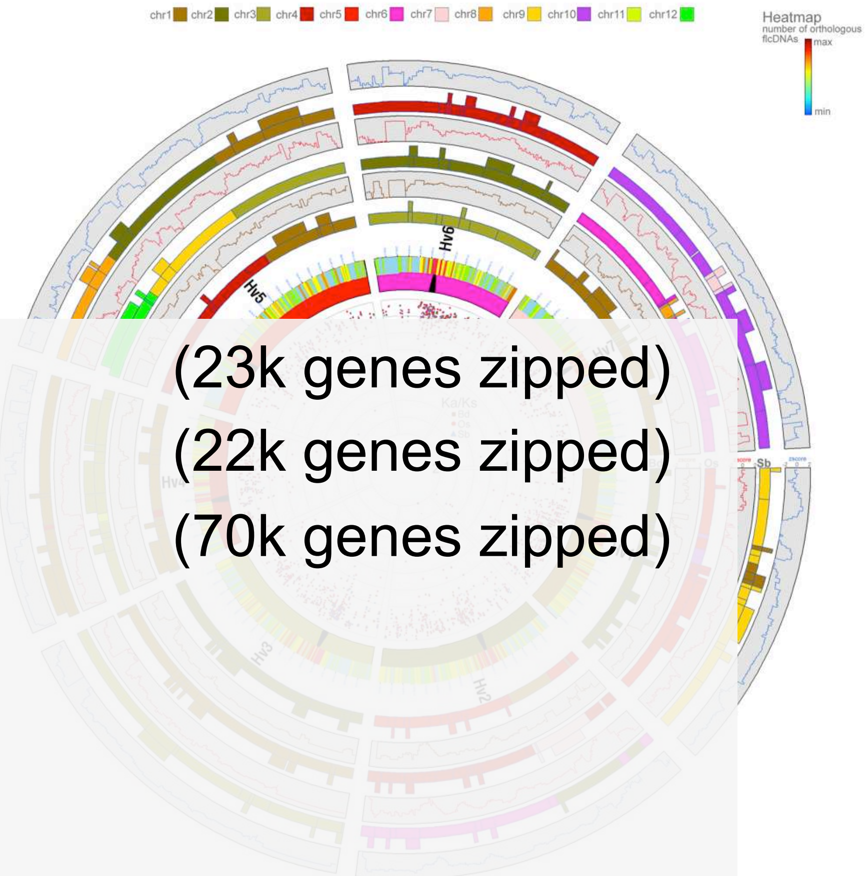
Syntenic Integration generates a „GenomeZipper“



marker	cM	brachy hit	rice hit	sorghum hit	454 shotgun reads	corresponding harvEST35
-	-	Bradi2g40110.1	Os05g0145000	-	316745;47800;14614	23242;
-	-	-	Os05g0101200	Sb09g000250.1	1339032;2247860;11	17120;
-	-	Bradi2g40100.1	Os05g0100900	Sb09g000240.1	1581361;633922;922	20508;
-	-	Bradi2g40090.3	Os05g0100800	-	2871221;1807303;28	15637;
-	-	-	Os05g0100700	-	640033;2837098;	4437;
-	-	Bradi2g40080.1	-	Sb09g000220.1	2019452;773802;187	18906;7654;
-	-	Bradi2g40070.1	-	-	1753215;2018341;18	3764;
-	-	-	Os05g0100100	-	755586;	-
-	-	-	Os05g0100500	-	555075;	-
-	-	Bradi2g40060.1	Os05g0100600	Sb09g000210.1	777855;815511;3770	46671;22527;
3_0969	0	Bradi2g40050.2	-	Sb09g000200.1	1483772;2140716;21	4526;37116;16811;
-	-	Bradi2g40050.1	-	-	574927;1183343;160	17786;
-	-	Bradi2g40030.1	Os05g0180600	-	2580264;478745;940	-
-	-	-	Os05g0180500	-	65700;	6499;
-	-	Bradi2g40020.1	-	-	1676072;1681761;	17363;
1_1223	0,75	-	-	-	-	17306
2_1354	0,77	Bradi2g40050.2	-	Sb09g000200.1	1483772;866014;121	4526;37116;16811;
1_0895	0,77	-	-	-	-	3803
2_0842	0,95	Bradi2g39900.1	Os05g0103500	Sb09g000380.1	454016;1912939;794	16353;
-	-	Bradi2g39840.1	Os05g0104100	Sb09g000420.1	248233;553573;5243	8448;8447;
-	-	Bradi2g39840.2	-	-	1599964;	8447;
-	-	Bradi2g39840.3	-	-	2679077;248233;197	-
-	-	Bradi2g39850.1	-	-	482485;2250536;101	20801;28045;
-	-	Bradi2g39860.1	-	Sb09g000410.1	2307616;	-
-	-	Bradi2g39880.2	-	Sb09g000400.1	430886;1184892;911	-
1_1311	0,95	Bradi2g39890.1	Os05g0103600	Sb09g000390.1	1196483;715096;298	5087;
-	-	Bradi2g39890.4	-	-	715096;852057;1196	5087;
1_0460	0,95	-	-	-	2528469;2841862;39	197;204;207;332;3
-	-	-	Os05g0102300	-	1667938;1738659;40	-
-	-	-	-	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	1673880;1328288;10	19203;16735;1673
-	-	Bradi2g39980.4	-	-	971584;2129140;265	2230;
-	-	Bradi2g39980.3	-	-	2356595;1728894;37	2230;
-	-	Bradi2g39980.1	Os05g0102900	Sb09g000340.1	971584;2356595;172	2230;
2_0502	1,51	Bradi2g39950.1	Os05g0103200	Sb09g000350.1	2229610;1687270;28	1696;
-	-	-	Os05g0103300	-	1050444;	7504;
2_0836	1,51	-	-	-	2211820;	324;298;
3_0715	1,51	-	-	-	34770;1480910;1803	19074;
3_1144	3,75	-	-	-	2565054;907983;291	34579;16501;
2_1456	3,75	-	-	-	2231492;	5699;
-	-	-	-	Sb09g000770.1	704223;680315;4063	-
-	-	-	-	Sb09g000780.1	2520212;2450315;	-
-	-	-	-	Sb09g000790.1	385765;1901355;235	43705;
-	-	-	Os05g0106200	Sb09g000800.1	2520212;787598;245	-
-	-	-	Os05g0106100	-	2240626;2733161;86	14893;
-	-	-	-	Sb09g000850.1	1923636;2535467;18	17177;4908;
-	-	-	-	Sb09g000830.1	2560715;	-
-	-	Bradi2g39700.1	Os05g0106000	Sb09g000820.1	1189690;2143848;26	20150;10110;
-	-	Bradi2g39710.1	-	-	1077529;36373;1604	13331;18906;1890
-	-	Bradi2g39720.1	-	-	2224664;2709243;24	5280;5410;5409;52
1_0419	3,75	Bradi2g39740.1	-	Sb09g000600.1	1467452;1417634;43	15699;
-	-	-	Os05g0105500	Sb09g021980.1	1991914;1519334;	11135;
-	-	-	-	Sb09g021960.1	728462;	-
-	-	Bradi2g39750.1	Os05g0105300	-	1932535;1691820;51	-
-	-	-	Os05g0105200	-	1770979;	-
2_0238	3,75	Bradi2g39760.1	-	Sb09g000550.1	1770979;	-
-	-	Bradi2g39770.1	Os05g0105100	Sb09g000540.1	2968277;1677650;29	15260;18045;
-	-	-	-	Sb09g000530.1	1638278;1766895;19	27734;
-	-	Bradi2g39830.1	Os05g0104200	Sb09g000430.1	2286300;	-
-	-	Bradi2g39800.1	Os05g0104900	Sb09g000470.1	1988513;444103;184	1364;
1_0410	3,75	Bradi2g39790.1	Os05g0105000	Sb09g000490.1	1006521;391681;167	15664;
3_0268	42,52	Bradi2g38170.1	Os05g0127200	Sb09g002320.1	1309343;415098;195	6967;13052;
-	-	-	Os05g0126900	-	31085;80764;111534	-
-	-	Bradi2g38120.1	-	-	25319;1621782;1159	-
-	-	Bradi2g38130.1	-	Sb09g002270.1	409672;616154;7698	24793;
-	-	Bradi2g38130.2	-	-	616154;	24793;
-	-	Bradi2g38130.3	Os05g0126100	-	616154;1596085;188	24793;
-	-	Bradi2g38140.1	Os05g0126200	Sb09g002280.1	2996327;1587967;29	4516;26065;
-	-	Bradi2g38150.1	-	-	1797068;	-
2_1134	43,28	Bradi2g38160.1	-	Sb09g002290.1	197073;589351;2434	42385;11683;1786
-	-	Bradi2g38440.1	-	-	2717345;910933;793	32499;
-	-	Bradi2g38430.1	Os05g0131100	Sb09g002570.1	3003667;	-
-	-	-	Os05g0131000	Sb09g002560.1	1553643;885298;180	-
-	-	-	-	Sb09g002550.1	2759084;2369480;11	20331;48323;
-	-	Bradi2g38420.1	-	-	1165573;2261666;	20331;
-	-	Bradi2g38400.1	-	-	1540641;3017253;28	-
-	-	-	Os05g0129900	Sb09g002530.1	2191979;2166015;24	-
-	-	Bradi2g38390.1	Os05g0129700	Sb09g002520.1	2372145;1443439;25	-
-	-	Bradi2g38380.1	Os05g0129300	Sb09g002510.1	1023311;165929;	16155;
-	-	-	-	Sb09g002500.1	1464631;	-
-	-	-	Os05g0129200	-	2967412;2825624;34	29677;
-	-	Bradi2g38360.1	-	Sb09g002490.1	2967412;2825624;34	29677;
-	-	Bradi2g38350.1	Os05g0129100	Sb09g002480.1	2464185;	17137;
1_0314	45,13	Bradi2g38340.2	Os05g0129000	Sb09g002470.1	1170776;2755196;25	15702;
-	-	-	-	Sb09g002450.1	1480327;2102033;27	-
-	-	Bradi2g38330.2	Os05g0128400	Sb09g002460.1	433582;1925337;940	15430;11898;
-	-	Bradi2g38310.1	-	-	822217;1788479;269	5784;
-	-	Bradi2g38280.1	-	-	2233436;153885;145	-
-	-	Bradi2g38260.1	-	-	2727674;2209698;14	-
-	-	Bradi2g36970.1	Os05g0135900	Sb09g002860.1	2889501;14758;1351	35313;
-	-	-	Os05g0136200	Sb09g002900.1	505041;1885796;134	2489;2487;
-	-	Bradi2g36940.1	-	-	1296380;2280486;63	42154;
-	-	Bradi2g36930.1	-	Sb09g002890.1	281288;1741875;227	-
-	-	Bradi2g36910.1	Os05g0136900	Sb09g002910.1	168249;	-
-	-	-	-	Sb09g002930.1	1062937;965193;195	2776;
-	-	Bradi2g36900.1	-	Sb09g002940.1	1083115;2180366;14	17184;43594;
-	-	-	Os05g0137200	-	2160720;147249;132	8212;17184;
-	-	Bradi2g36890.1	Os05g0137300	Sb09g002950.1	1718625;1582525;68	3005;
-	-	-	-	Sb09g003000.1	2904216;1616109;	20203;
-	-	-	Os05g0137700	Sb09g002990.1	2904216;1616109;13	20203;32860;
-	-	-	Os05g0137600	Sb09g002985.1	2716389;	-
1_0294	47,47	Bradi2g36860.3	Os05g0137400	Sb09g002960.1	2731998;1669831;20	15364;1640;15366;
-	-	Bradi2g36850.1	-	-	523838;1538595;502	-
-	-	Bradi2g36830.1	Os05g0113300	Sb09g001490.1	1993234;1438534;24	-
-	-	Bradi2g36810.1	-	-	1800005;372299;196	-
-	-	Bradi2g36790.1	-	-	1248023;	21520;
-	-	-	-	Sb09g003020.1	1940546;807734;	-
-	-	Bradi2g36780.1	Os05g0138200	Sb09g003040.1	801830;732533;2465	44698;
-	-	Bradi2g36760.2	-	Sb09g003070.1	2588440;3045473;	-
-	-	-	-	Sb09g003080.1	2130511;1120689;	-
-	-	Bradi2g36750.1	-	-	32914;1733758;2954	-
-	-	Bradi2g36740.1	-	-	68451;	19249;
-	-	Bradi2g36730.2	-	-	1764578;	-
-	-	-	Os05g0139200	-	676075;	-
-	-	-	-	Sb09g003090.1	68451;	19249;
-	-	Bradi2g36730.1	Os05g0140100	Sb09g003100.1	1607729;1742235;28	-
1_0526	47,47	Bradi2g36710.1	Os05g0140500	Sb09g003130.1	2700159;2711235;17	16244;
-	-	Bradi2g36690.1	Os05g0140800	Sb09g003150.1	80856;2136671;2504	9235;794;
-	-	Bradi2g36680.1	-	-	2352207;	23921;
-	-	Bradi2g36670.1	-	Sb09g003170.1	971971;	-
-	-	Bradi2g36660.1	Os05g0141400	Sb09g003210.1	2599753;2564655;22	-
-	-	Bradi2g36650.1	Os05g0141500	-	1509309;865103;128	20271;
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	-		

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





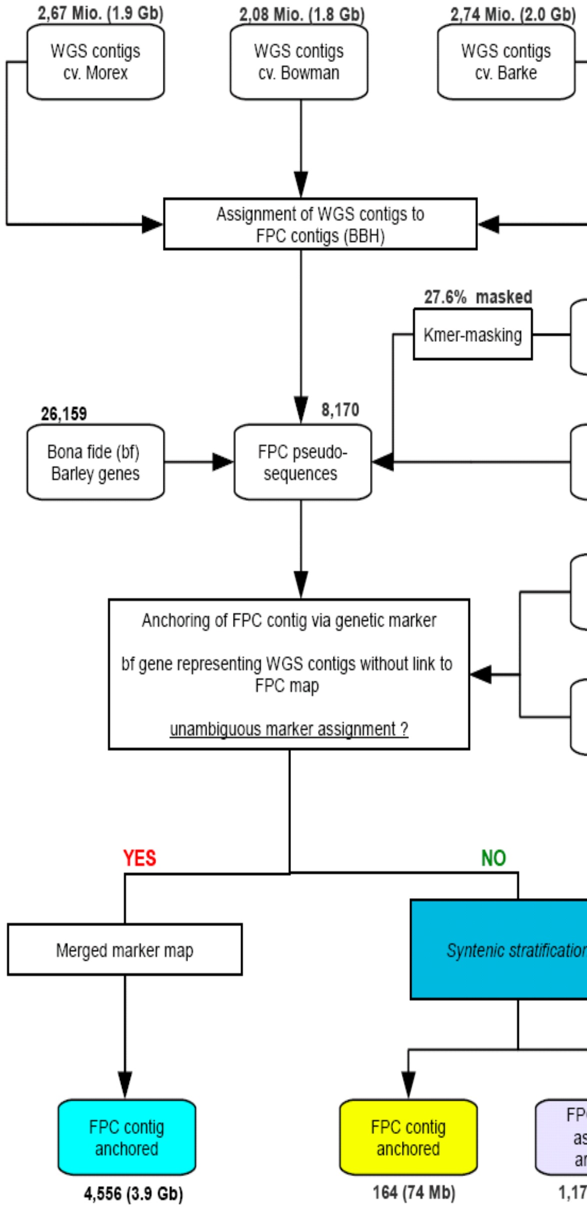
- Rye (23k genes zipped)
- *Aegilops tauschii* (22k genes zipped)
- Wheat – IWGSC (70k genes zipped)
- Lolium
- Festuca
- ...

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

26,159
Bona fide (bf) Barley genes

8,170
FPC pseudo-sequences

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

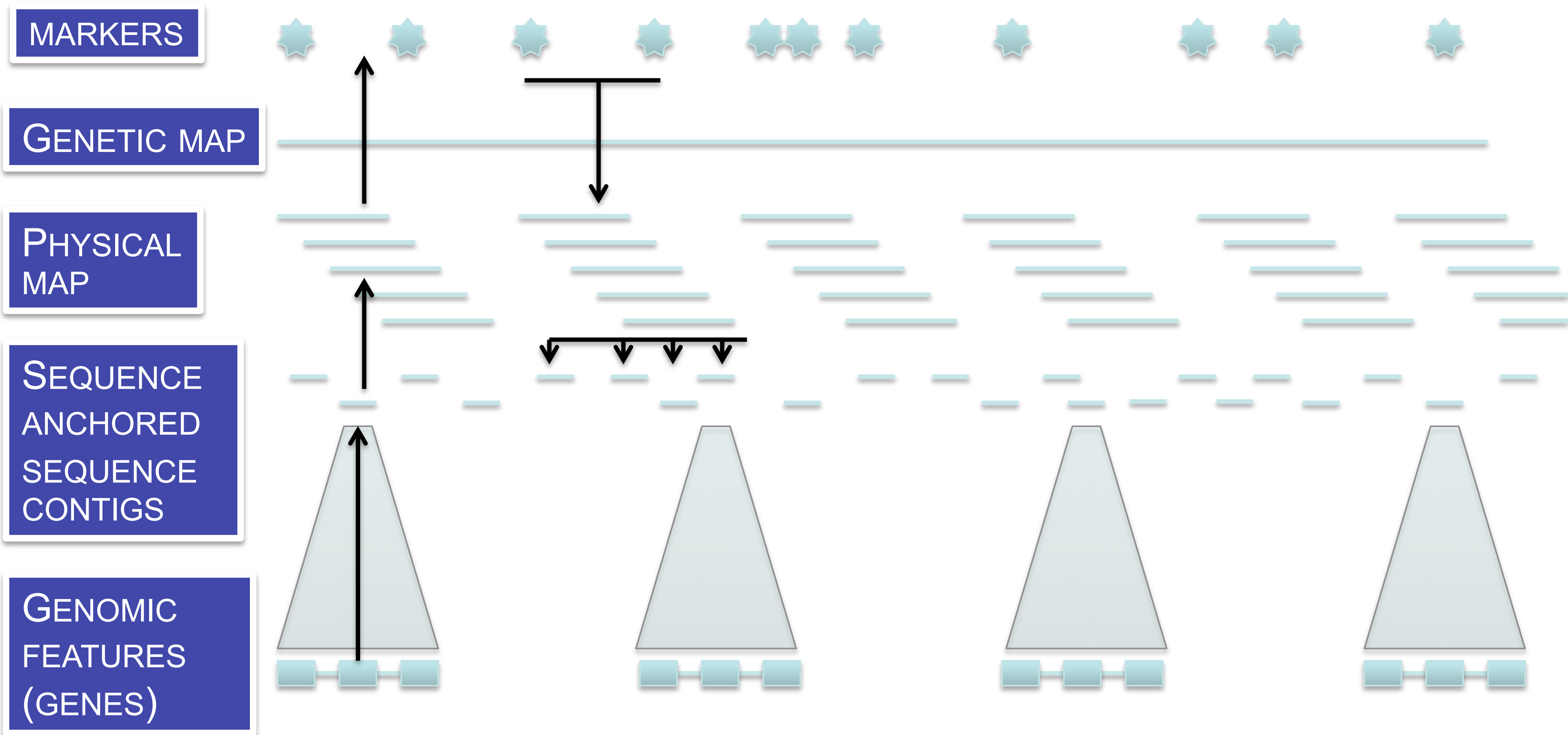
Syntenic stratificatio

FPC contig anchored
4,556 (3.9 Gb)

FPC contig anchored
164 (74 Mb)

FP
as
ar
1,17

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



MARKERS

GENETIC MAP

PHYSICAL MAP

SEQUENCE ANCHORED SEQUENCE CONTIGS

GENOMIC FEATURES (GENES)

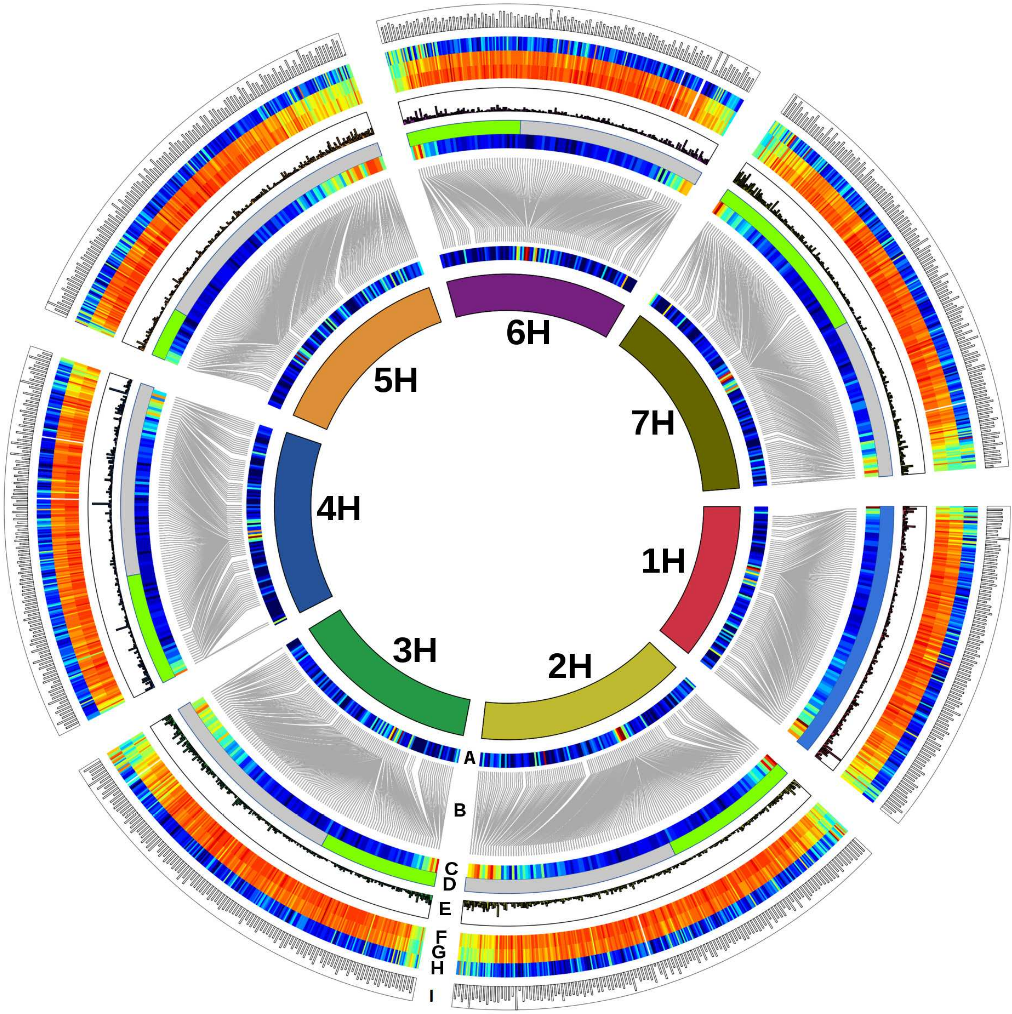
Genome stratification in barley cont.

The Barley **Gene-ome**:

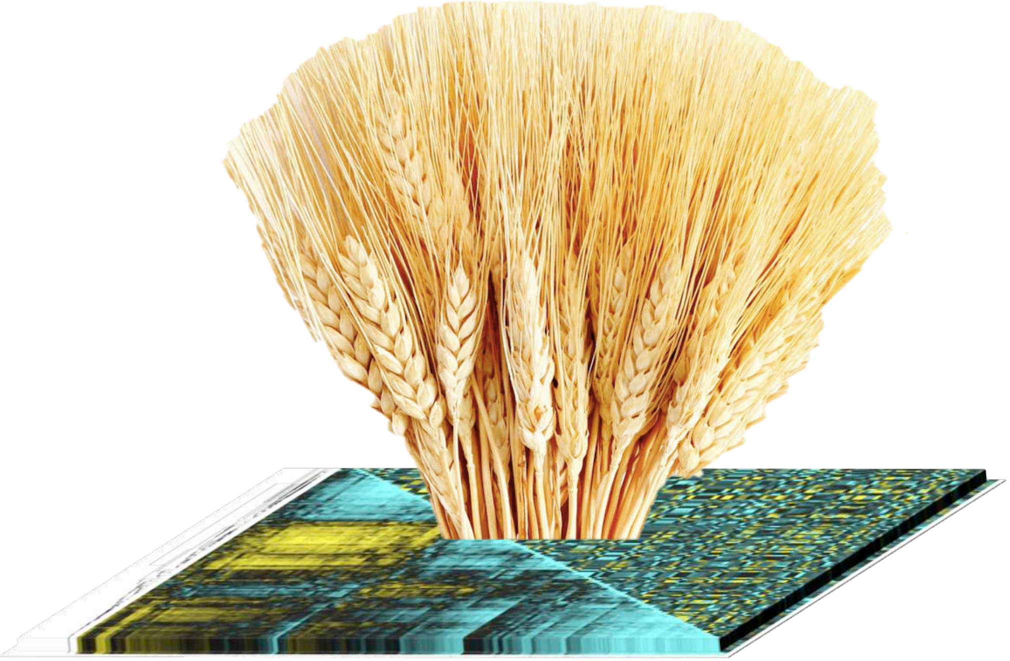
A **physical**, **genetic** and

functional sequence

assembly

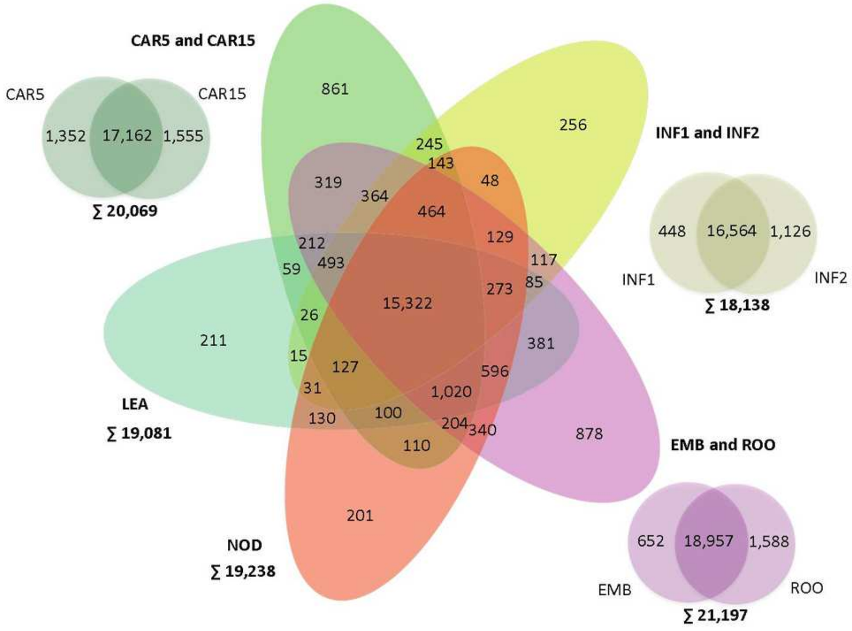


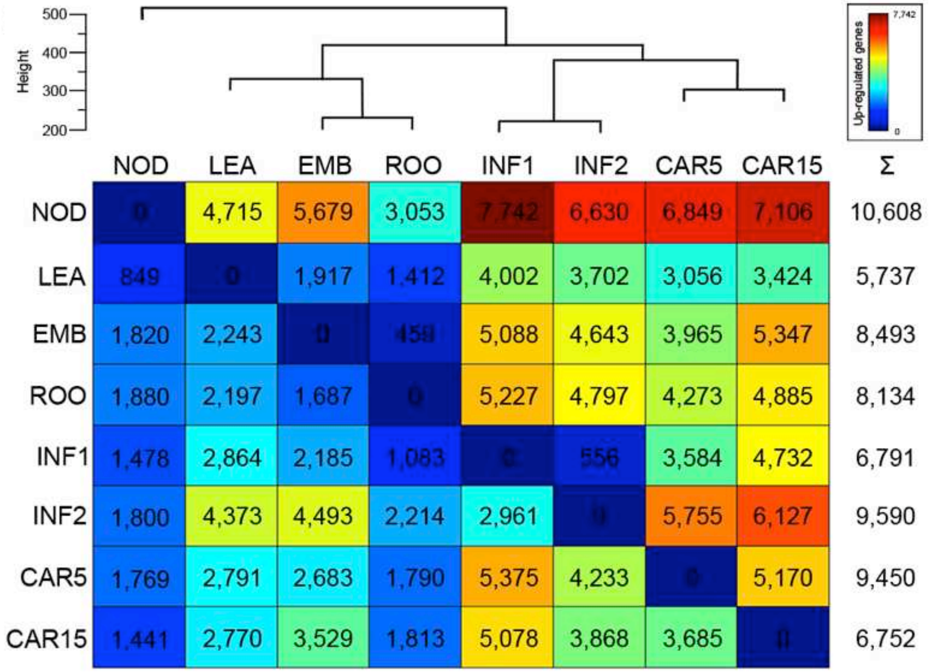
The Transcriptome of a Cereal



The Transcriptome of a Cereal

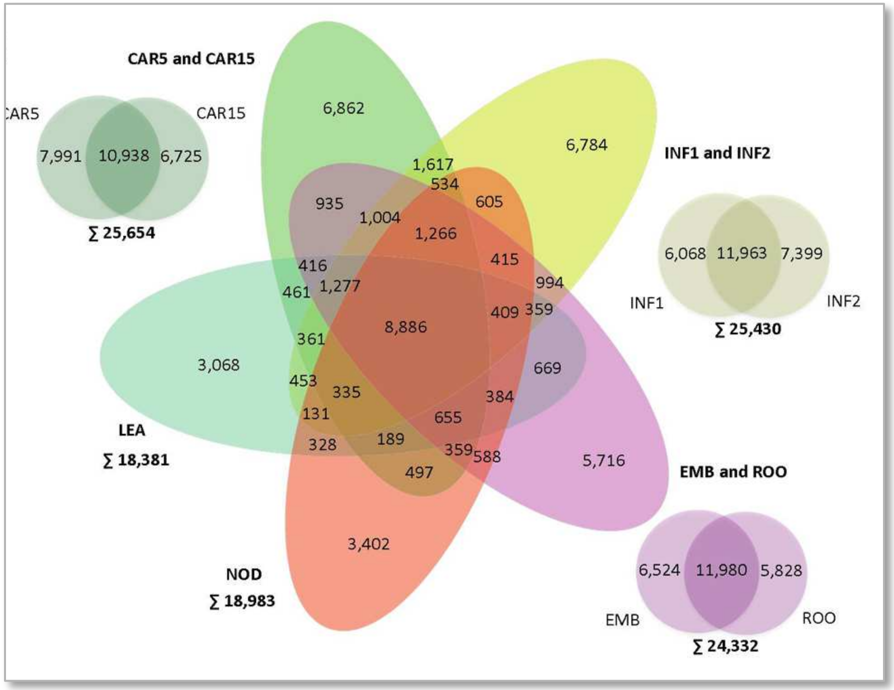
- Overlap and Dynamics in Expression -



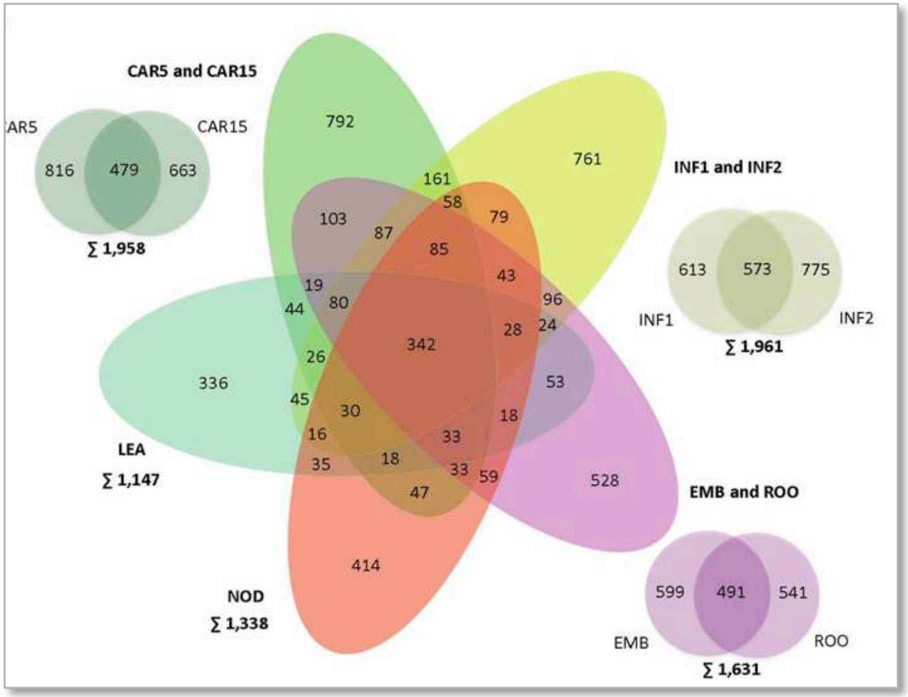


	NOD	LEA	EMB	ROO	INF1	INF2	CAR5	CAR15	Σ
NOD	0	4,715	5,679	3,053	7,742	6,630	6,849	7,106	10,608
LEA	849	0	1,917	1,412	4,002	3,702	3,056	3,424	5,737
EMB	1,820	2,243	0	458	5,088	4,643	3,965	5,347	8,493
ROO	1,880	2,197	1,687	0	5,227	4,797	4,273	4,885	8,134
INF1	1,478	2,864	2,185	1,083	0	556	3,584	4,732	6,791
INF2	1,800	4,373	4,493	2,214	2,961	0	5,755	6,127	9,590
CAR5	1,769	2,791	2,683	1,790	5,375	4,233	0	5,170	9,450
CAR15	1,441	2,770	3,529	1,813	5,078	3,868	3,685	0	6,752

The Transcriptome of a Cereal
- A pronounced RNA based Regulatory Layer (?)-



Pronounced Differential Alternative Splicing



Nonsense Mediated Decay (NMD) and Premature Termination Codons (PTC) are subject to translational and transcriptional regulation

Wheat – a WGS approach
Aims and Goals

Challenges

- hexaploid structure AABBDD
- 7 chromosomes per genome (42 chrom.)
- 17.1Gb
- ~80% repeat content

Aims

- ▶ Overview of the wheat gene content
- ▶ Gene retention rate
- ▶ Contracted / expanded gene families
- ▶ Pseudogenization
- ▶ Relationship of homeologous genes

WGS assembly using „*in silico* exon capture“

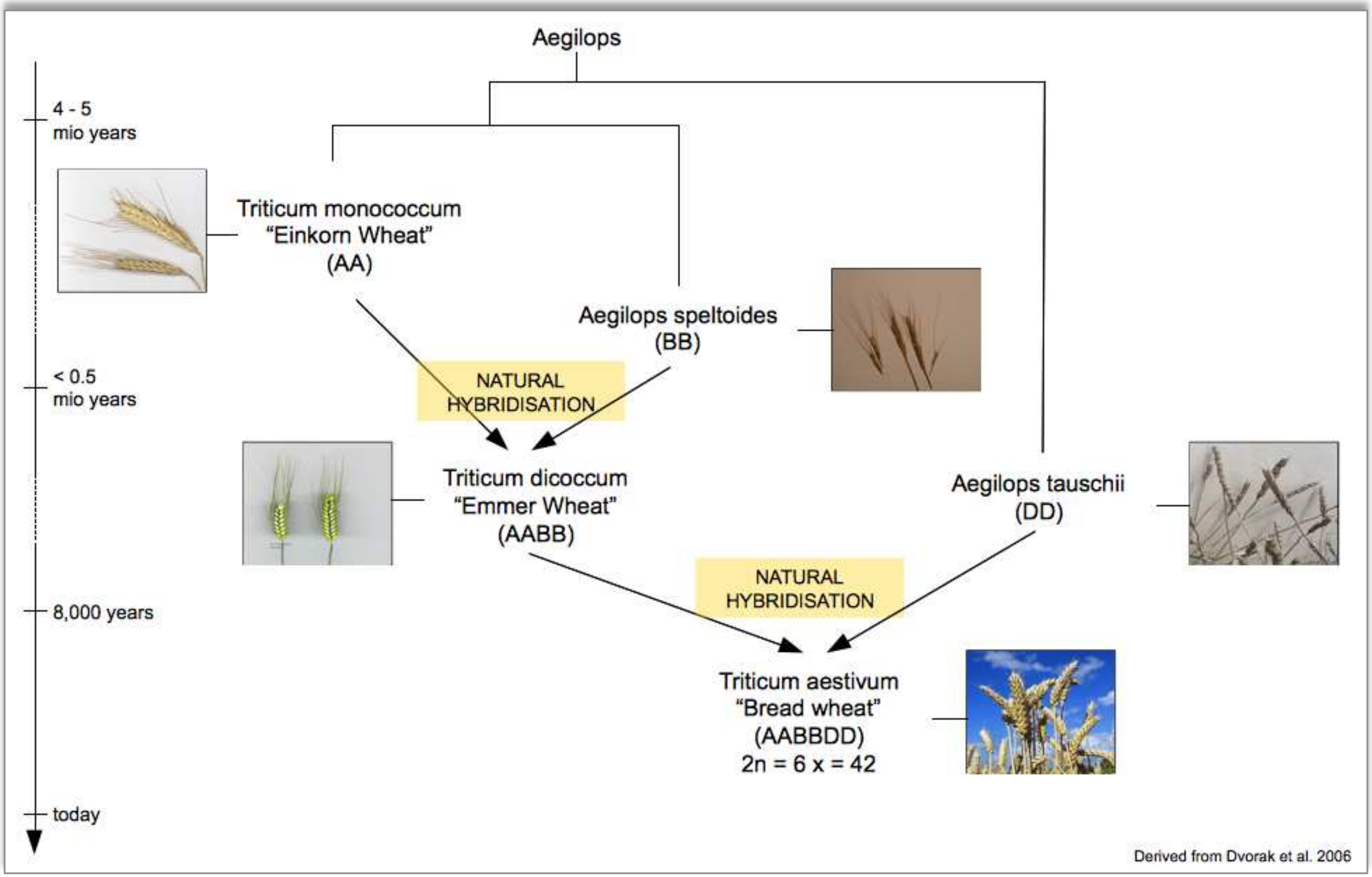
① Use fully sequences and analysed reference genomes (rice, *Brachypodium*, sorghum)

② Group genes into families (Orthologous Groups)

③ Use the orthologous group representatives as sequence baits to capture corresponding sequence reads.

④ Do sub-assembly for each „orthologous bin“ seperatly

Bread Wheat Genaology



Aegilops

4 - 5
mio years

< 0.5
mio years

8,000 years

today

Triticum monococcum
"Einkorn Wheat"
(AA)

Aegilops speltoides
(BB)

NATURAL
HYBRIDISATION

Triticum dicoccum
"Emmer Wheat"
(AABB)

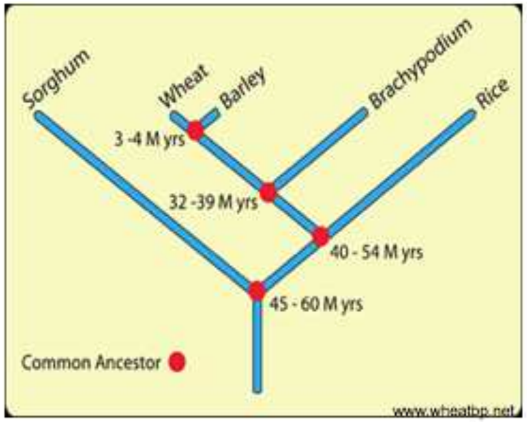
Aegilops tauschii
(DD)

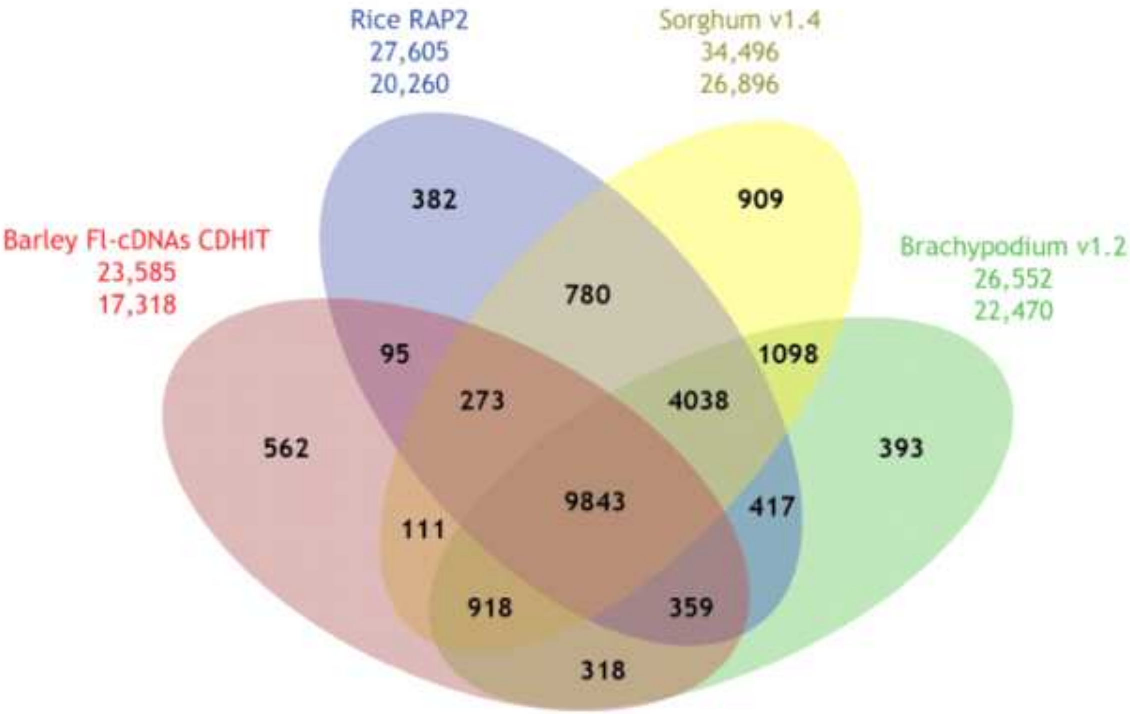
NATURAL
HYBRIDISATION

Triticum aestivum
"Bread wheat"
(AABBDD)
2n = 6 x = 42

Derived from Dvorak et al. 2006

Ortholome directed assembly circumvents limitations faced by WGS assembly



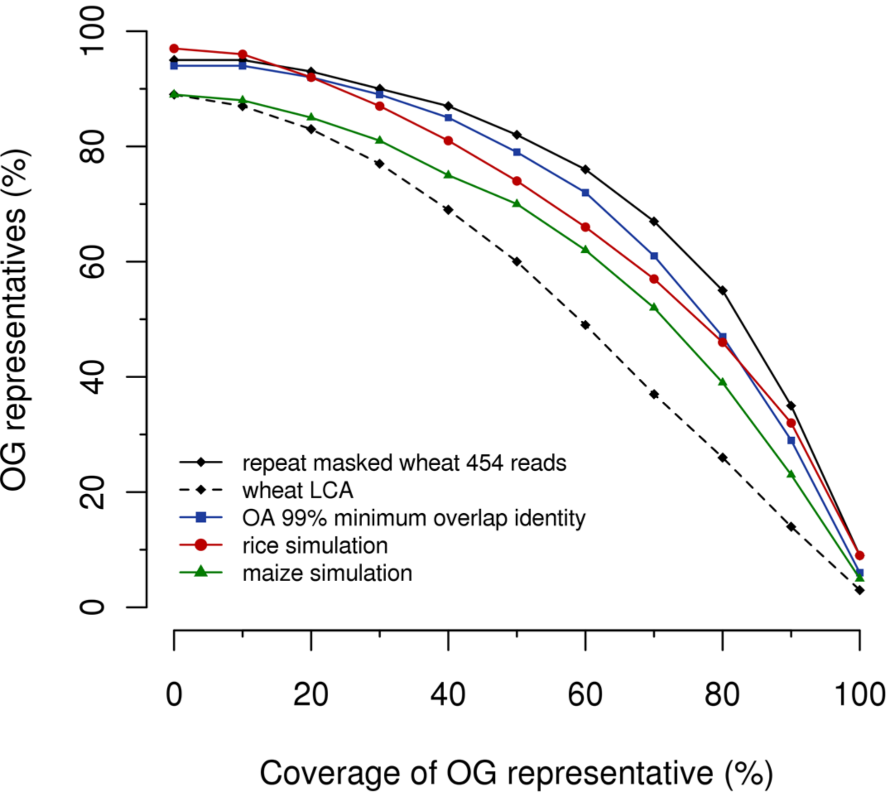


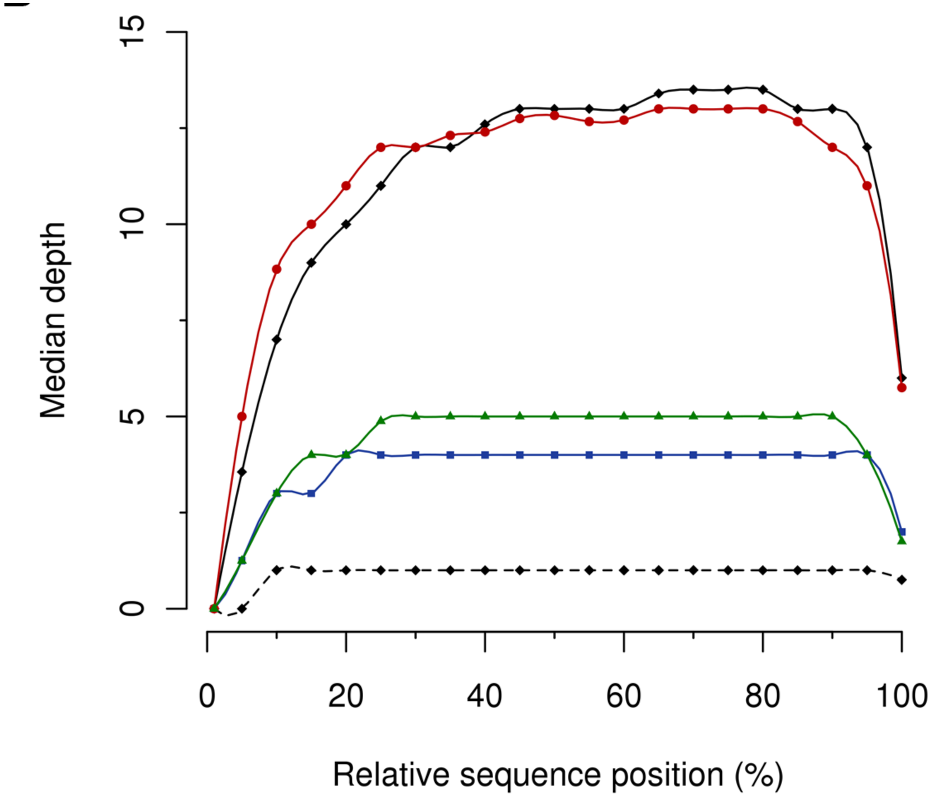
7,996 Brachypodium genes
3,582 Sorghum genes
5,337 Rice genes
5,337 Barley FL-cDNAs

20,051 orthologous gene (Og) representatives

used as templates to guide stringent assemblies

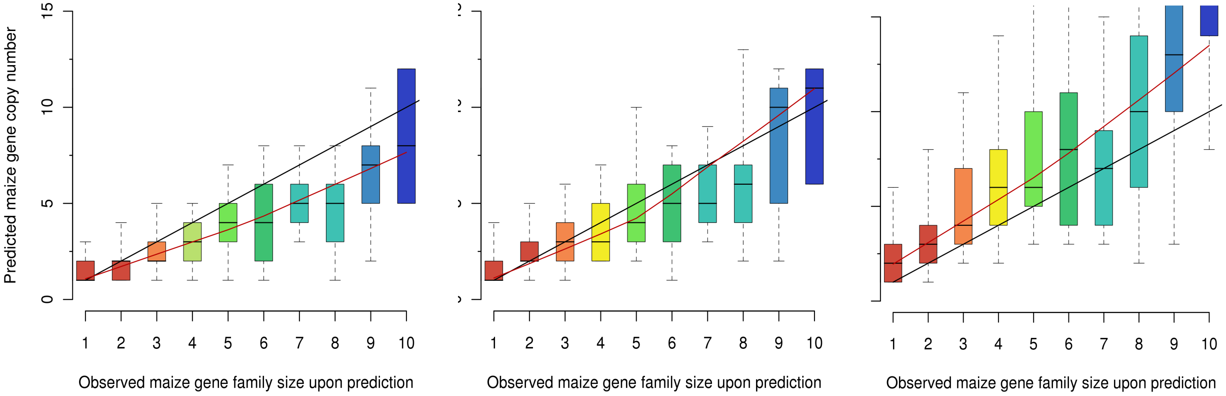
Coverage of Orthologous Group





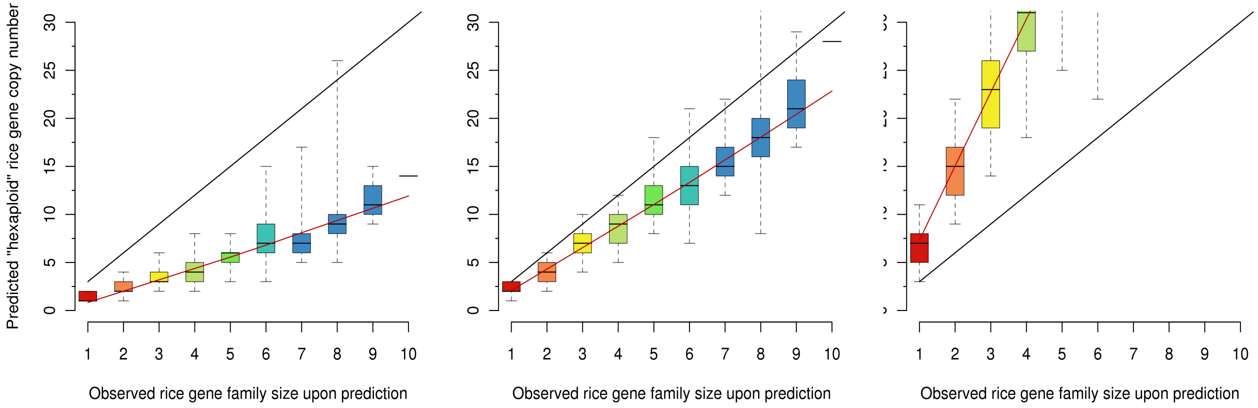
Gene Copy Retention after Polyploidization
- Calibration of the method-

Maize

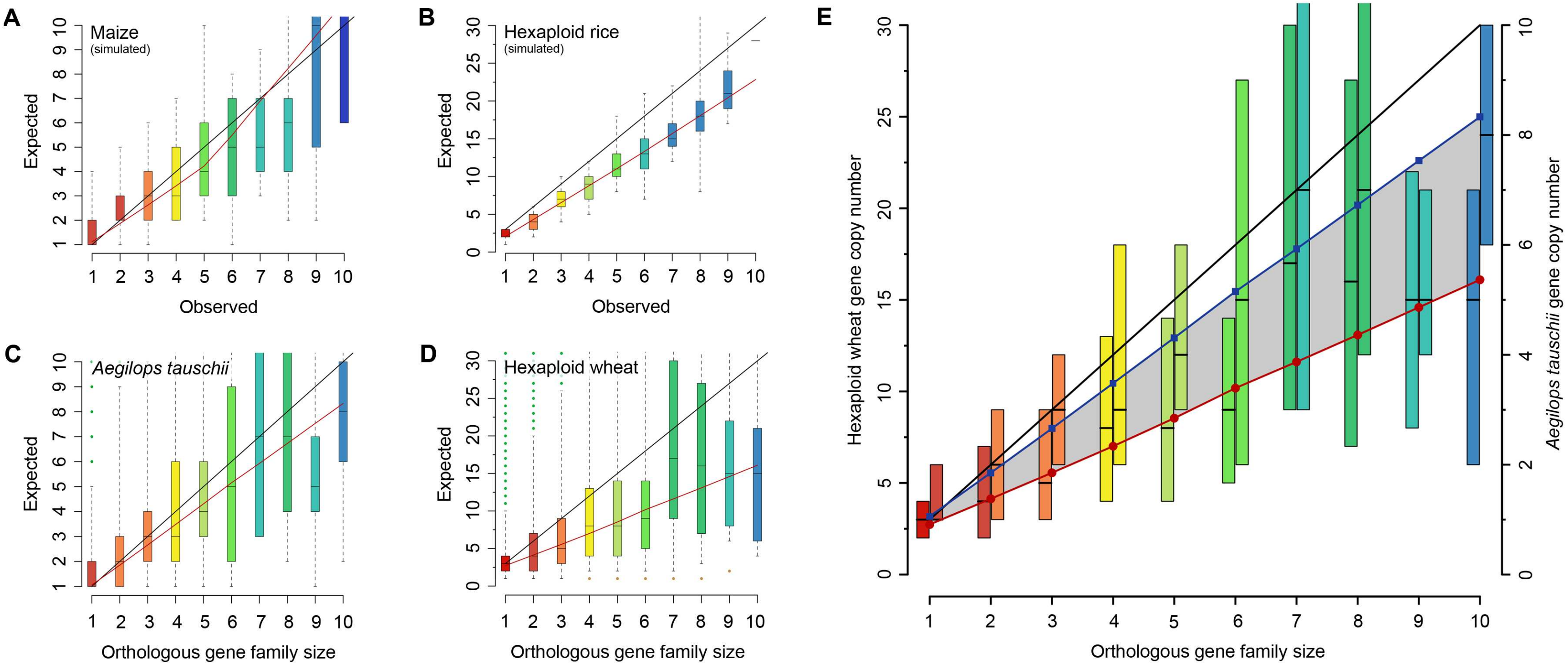


97% **99%** **100%**

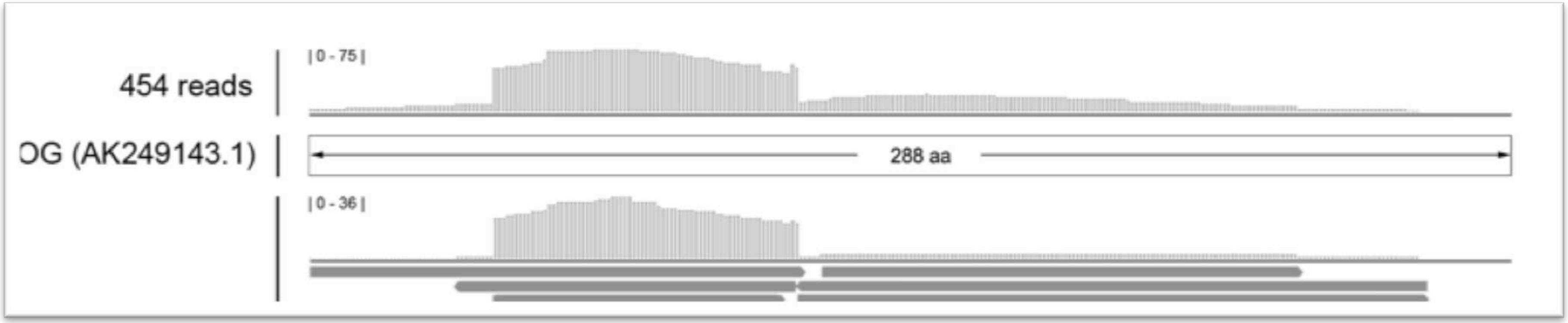
Hexaploid Rice
„TRice“



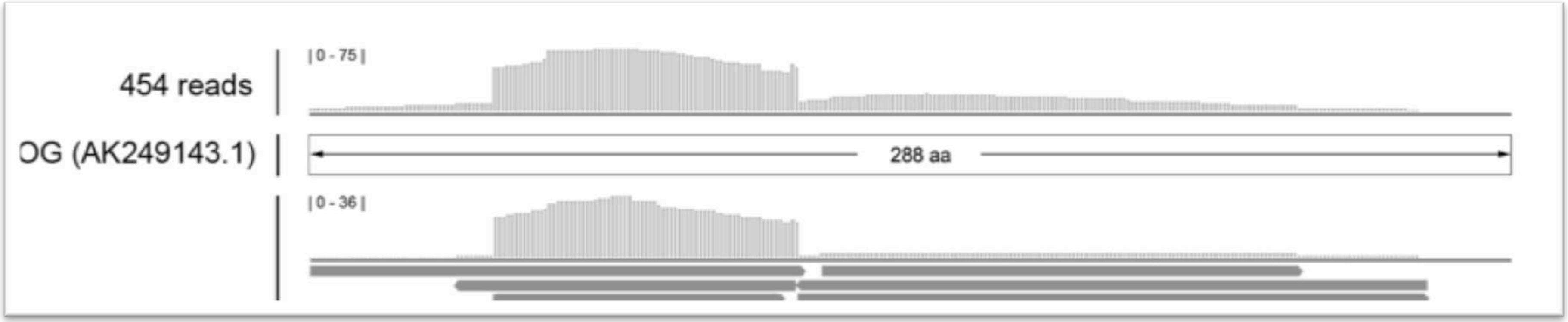
Gene Copy Retention after Polyploidization

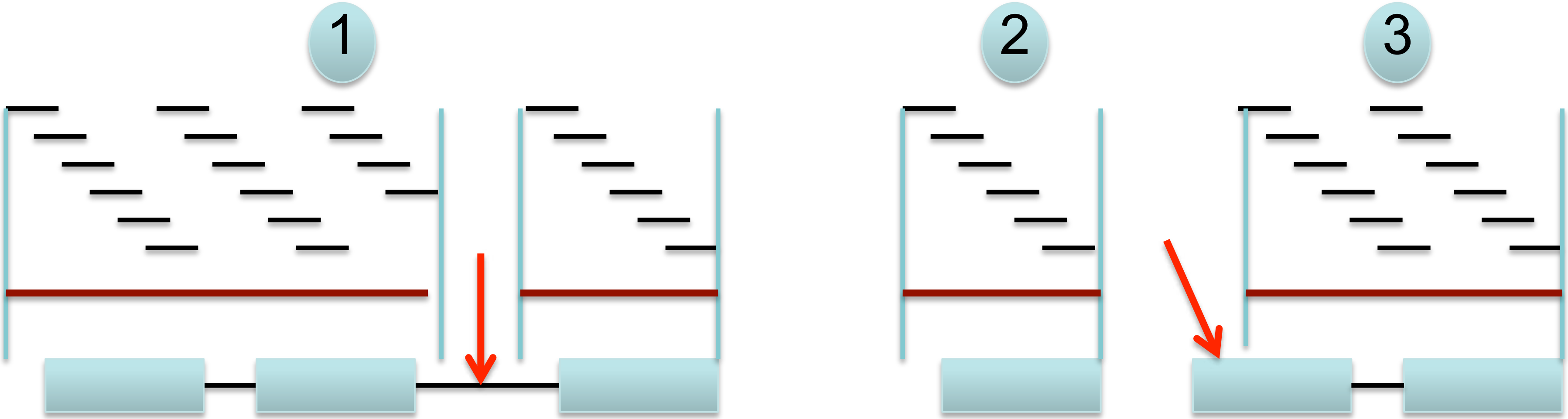


The ortholome direct assembly delivers ordered segments

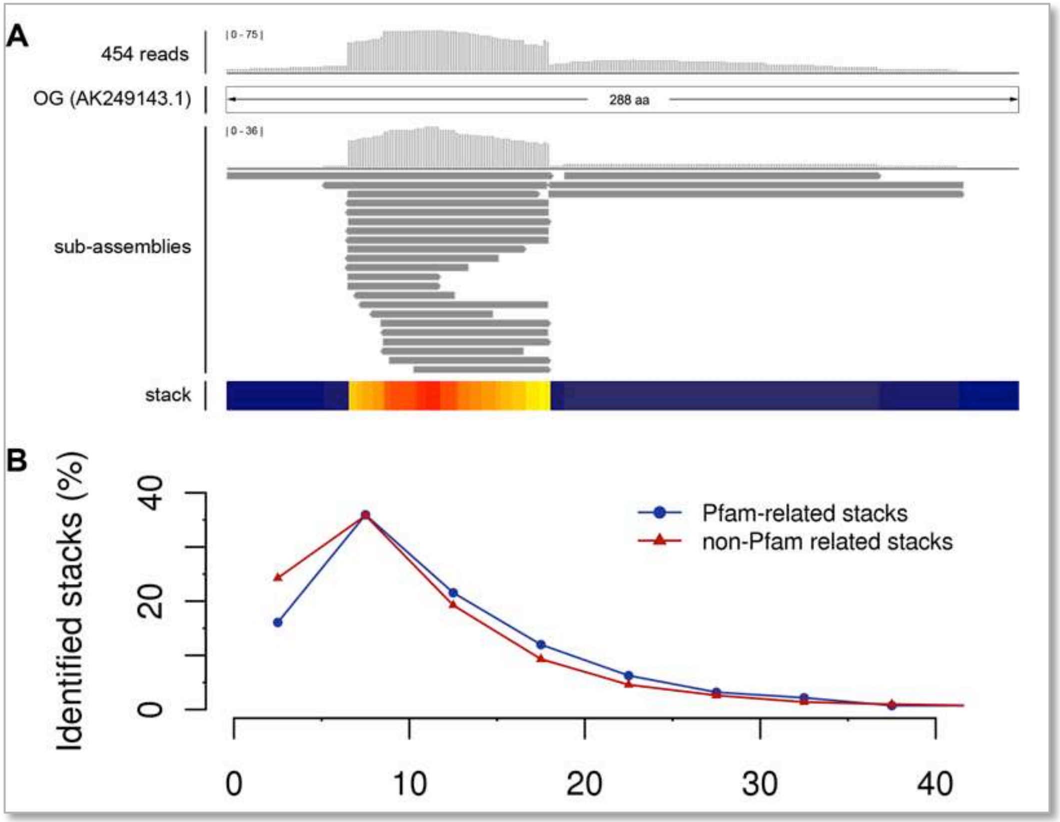


The ortholome direct assembly delivers ordered segments II

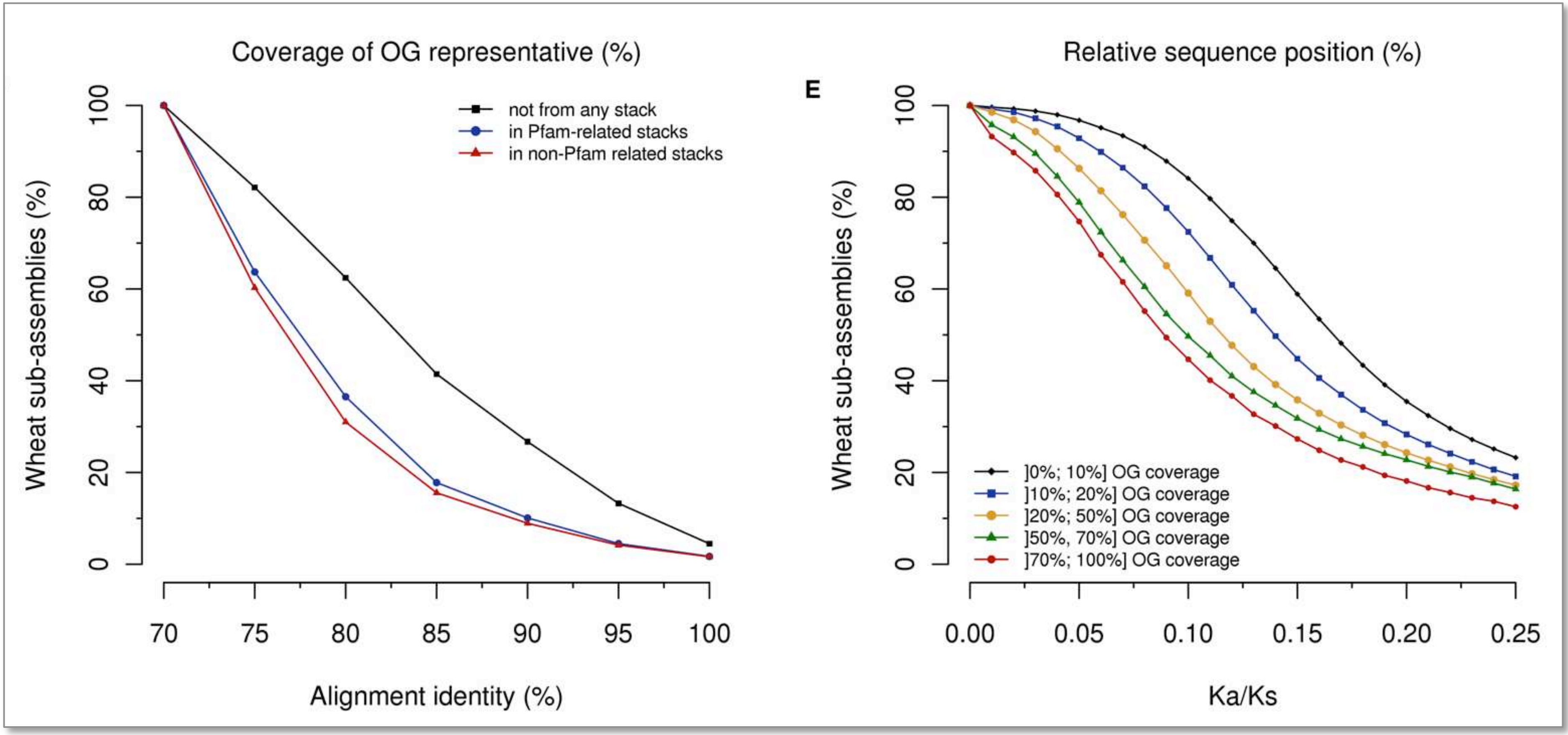




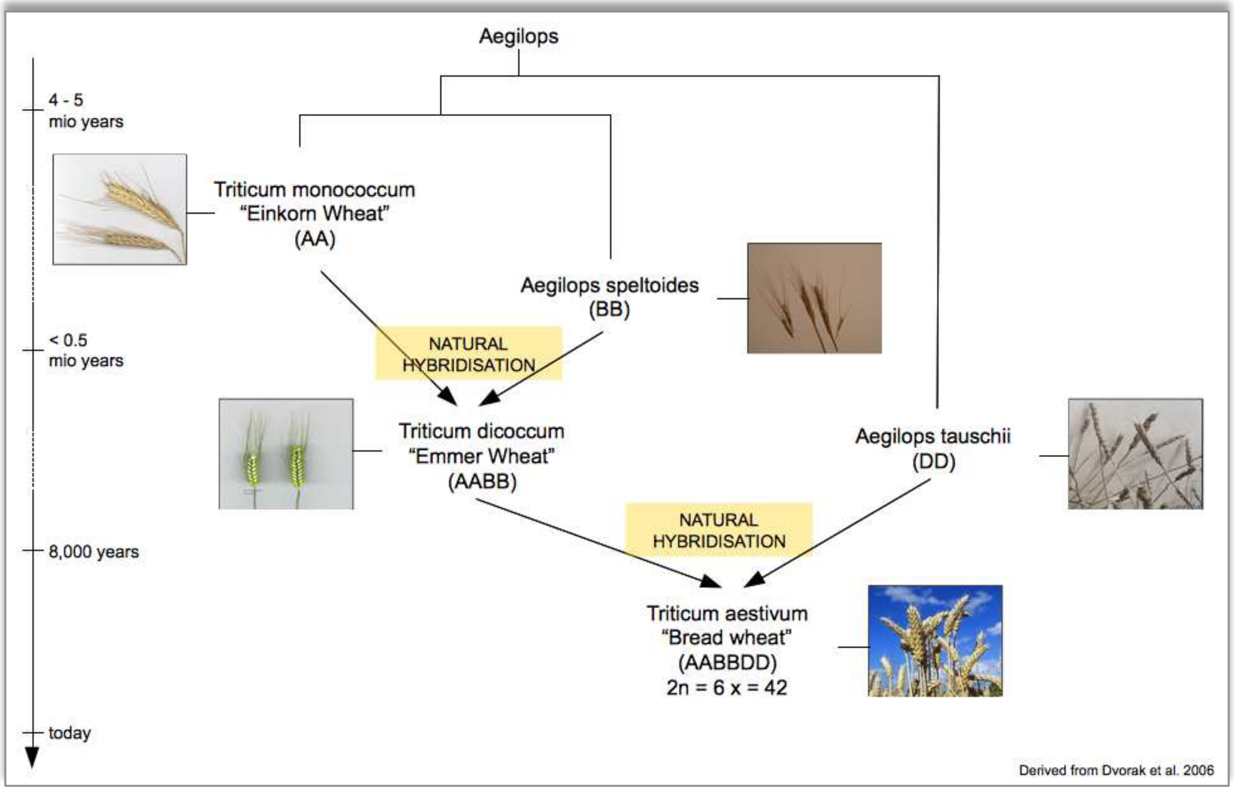
Gene fragments are abundant in wheat



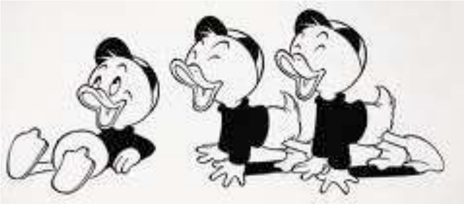
Gene fragments are abundant in the wheat genome



The Three Nephews: the A, B and D‘s of wheat



Derived from Dvorak et al. 2006

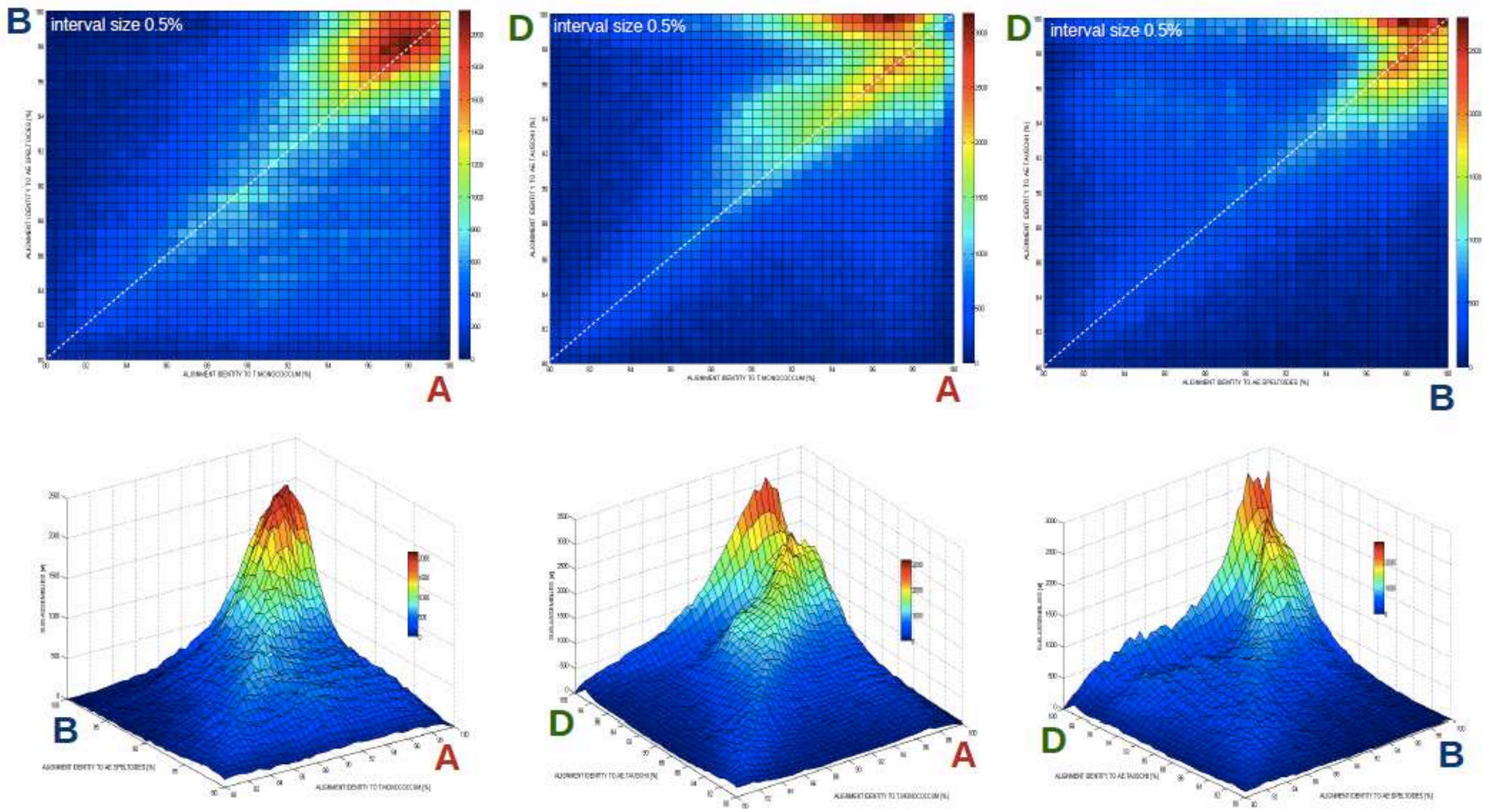


Shotguns (Illumina 80x (*T.monococcum*)) and 454 (3x (*Ae.tauschii*))

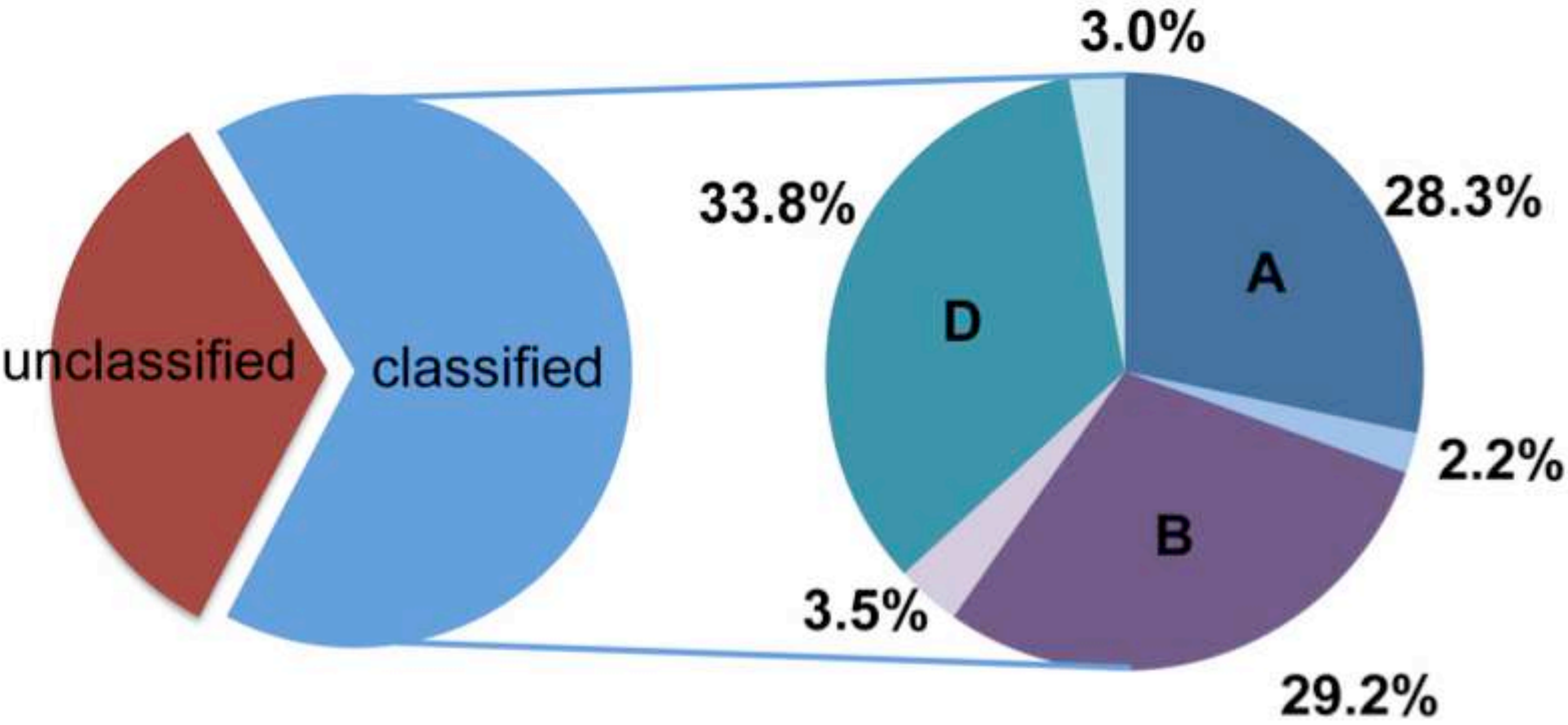
cDNA seq‘s from the *Ae. speltoides* group (B)

Can A and D genome shotgun data be used to dissect the ABD of wheat?

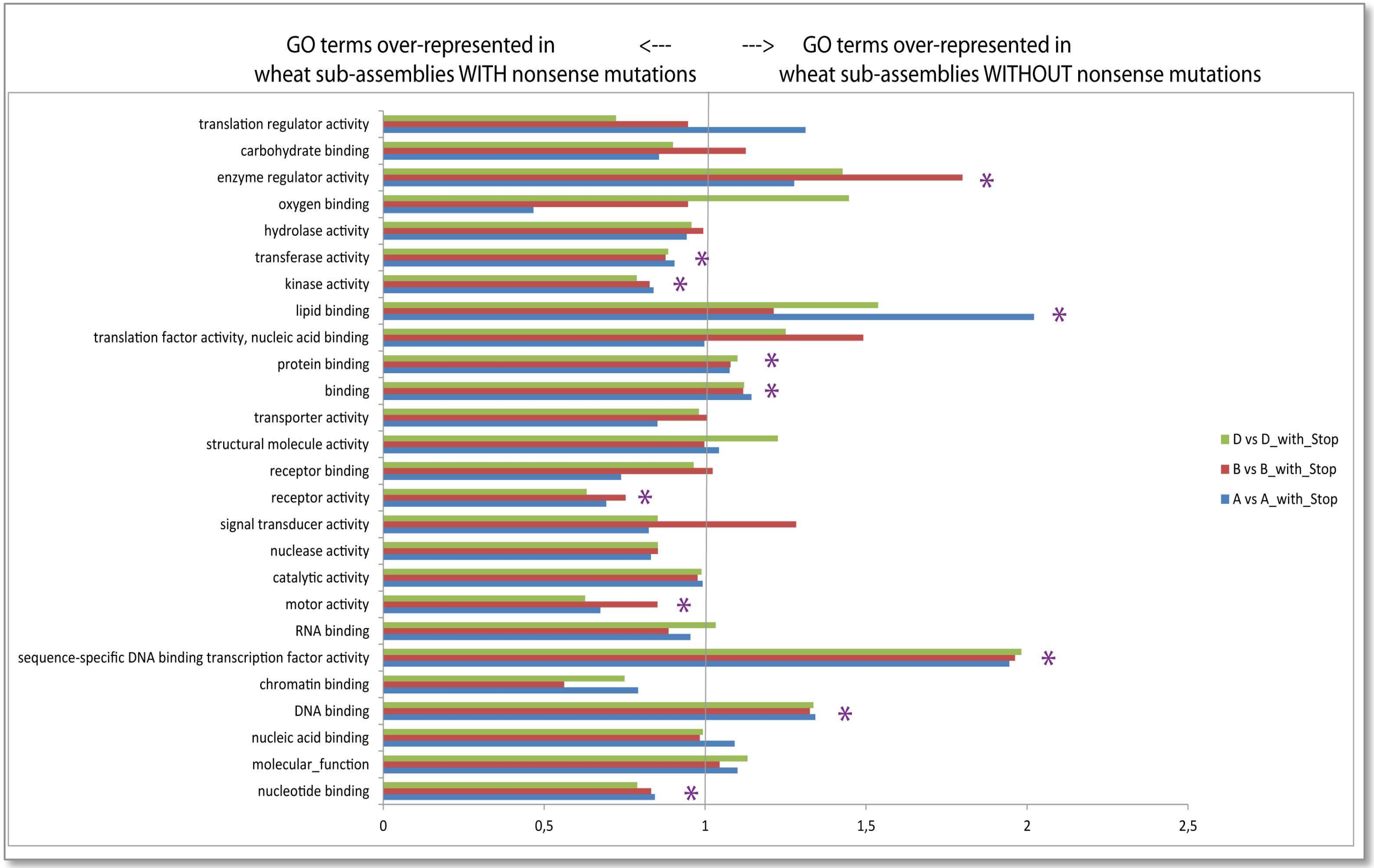
The Three Nephews: Similarity on a Sequence Basis



Wheat A, B and D Assignment using Machine Learning (SVM)



Particular Gene Categories are preferentially retained



trans PLANT

HelmholtzZentrum münchen
German Research Center for Environmental Health

Haystack -omics

Almost full gene complement detected and structured

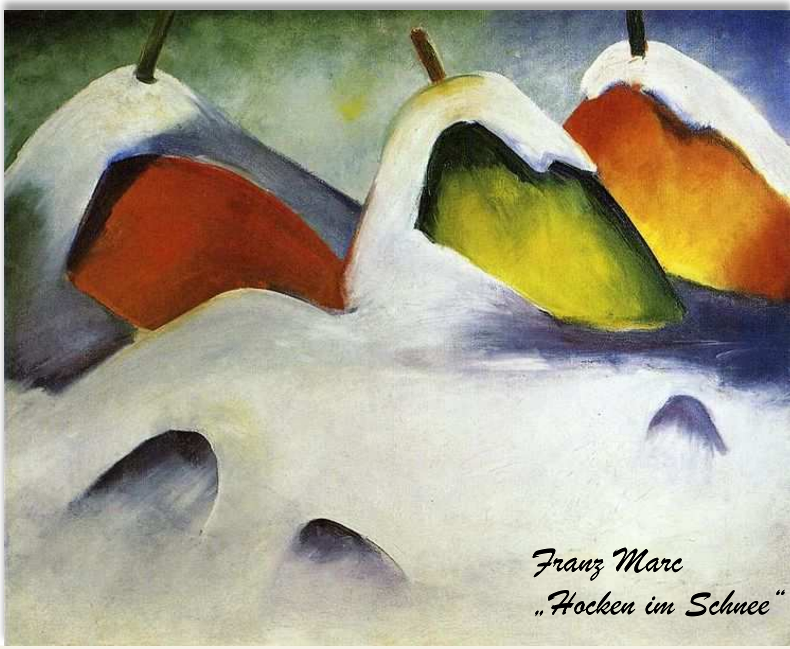
10000s of pseudogenes detected

Separation of A, B and D using machine learning with > 75% accuracy

Complementary to chromosome sorting approaches

Applicable to polyploids in general to get genome overview

Rapid and economic approach to pragmatically cope with limitations in sequence technology



Franz Marc
„Hocken im Schnee“

Take home messages

- A shortcut to access average sized plant genomes can be accessed using WGS, WCS and Genome Zipping
- Join of heterogeneous but complementary datasets delivers *Triticeae* „gene-omes“
- Provide jumboards for gene isolation, meta-analysis, genome wide functional analysis, comparative analysis, the $x10^3$ barley‘s/wheat‘s/rye‘s, ...
- B‘s are likely products of WGD/introgression events

acknowledgements

MIPS
Thomas Nussbaumer
Matthias Pfeifer
Manuel Spannagl
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

The IBSC

CSHL
Dick McCombie

UC Davis & USDA Albany
Jan Dvorak
Mincheng Luo
Olin Anderson

Kansas State University
Bikram Gill
Sunish Segal

Bundesministerium
für Bildung
und Forschung

gabi

Deutsche
Forschungsgemeinschaft
DFG

SEVENTH FRAMEWORK
PROGRAMME

Triticeae Genome