The Genome of *Triticum urartu*, a Progenitor of Wheat A Genome

Hong-Qing Ling

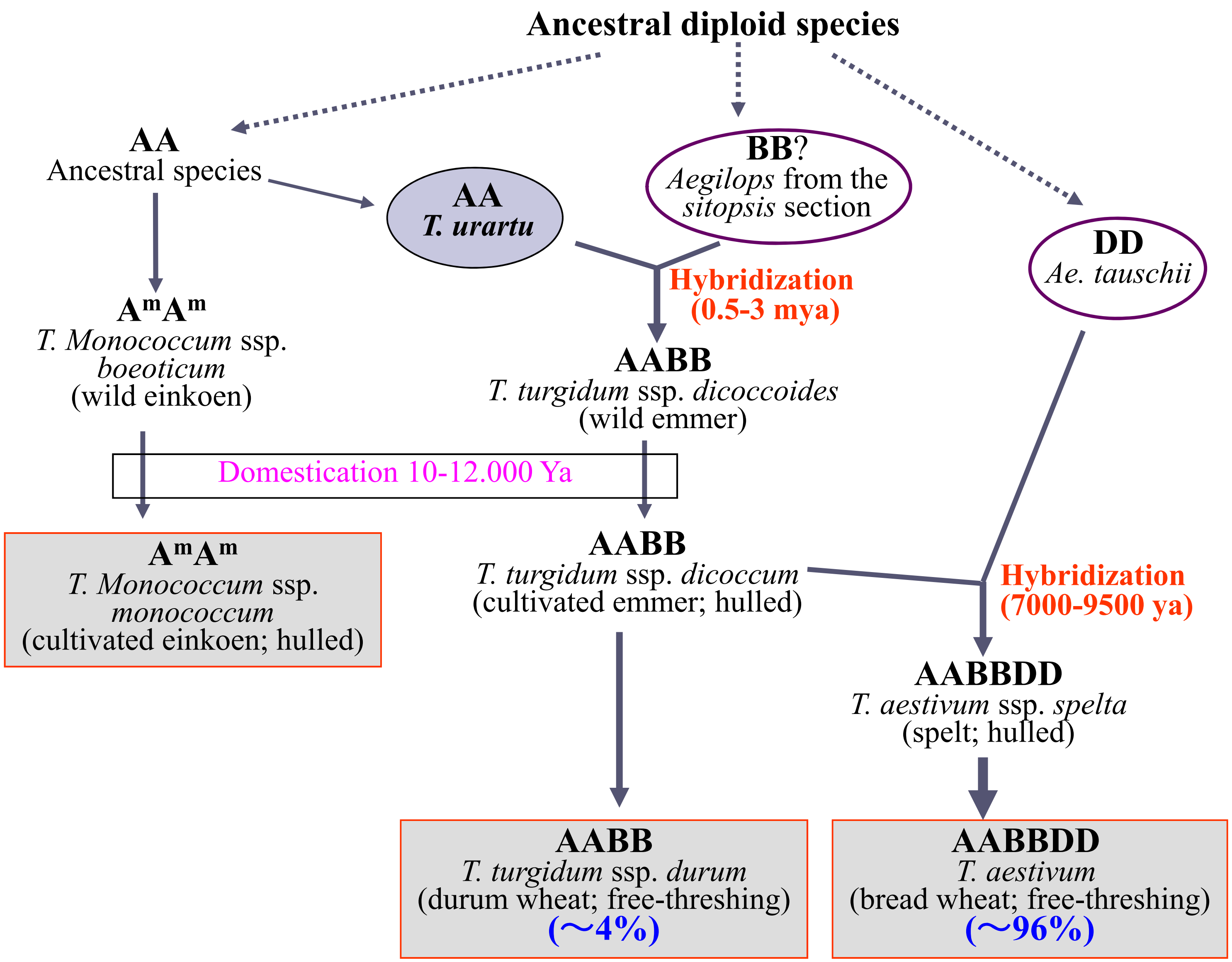
hqling@genetics.ac.cn

State Key Laboratory of Plant Cell and Chromosome Engineering
Institute of Genetics and Developmental Biology
Chinese Academy of Sciences

Bread Wheat

Wheat is one of the most important food crops in the world, feeding about 40% of the world population and providing 20% of total calories and protein in human nutrition.

Origin of Bread Wheat



Ancestral diploid species

AA
Ancestral species

AA
T. urartu

BB?
Aegilops from the *sitopsis* section

DD
Ae. tauschii

Hybridization (0.5-3 mya)

A^mA^m
T. Monococcum ssp. *boeoticum*
(wild einkoen)

AABB
T. turgidum ssp. *dicoccoides*
(wild emmer)

Domestication 10-12.000 Ya

A^mA^m
T. Monococcum ssp. *monococcum*
(cultivated einkoen; hulled)

AABB
T. turgidum ssp. *dicoccum*
(cultivated emmer; hulled)

Hybridization (7000-9500 ya)

AABBDD
T. aestivum ssp. *spelta*
(spelt; hulled)

AABB
T. turgidum ssp. *durum*
(durum wheat; free-threshing)
(∼4%)

AABBDD
T. aestivum
(bread wheat; free-threshing)
(∼96%)

Chantret et al. (2005), Plant Cell 17: 1033-1045

Bottleneck and Challenges in Wheat Research and Breeding

- Bread wheat is an allopolyploid, containing A, B and D subgenomes, which show a high similarity each other.
- The genome size of bread wheat is ~17 Gb. It is about 40 times of rice genome.
- More than 80% of wheat genome sequence are repetitive DNA.
- The genome sequencing of wheat is very difficult and is still a big challenge.
- Due to lacking the genome sequence, the basic research of wheat lags much behind rice and even other main crops, limiting its further improvement.

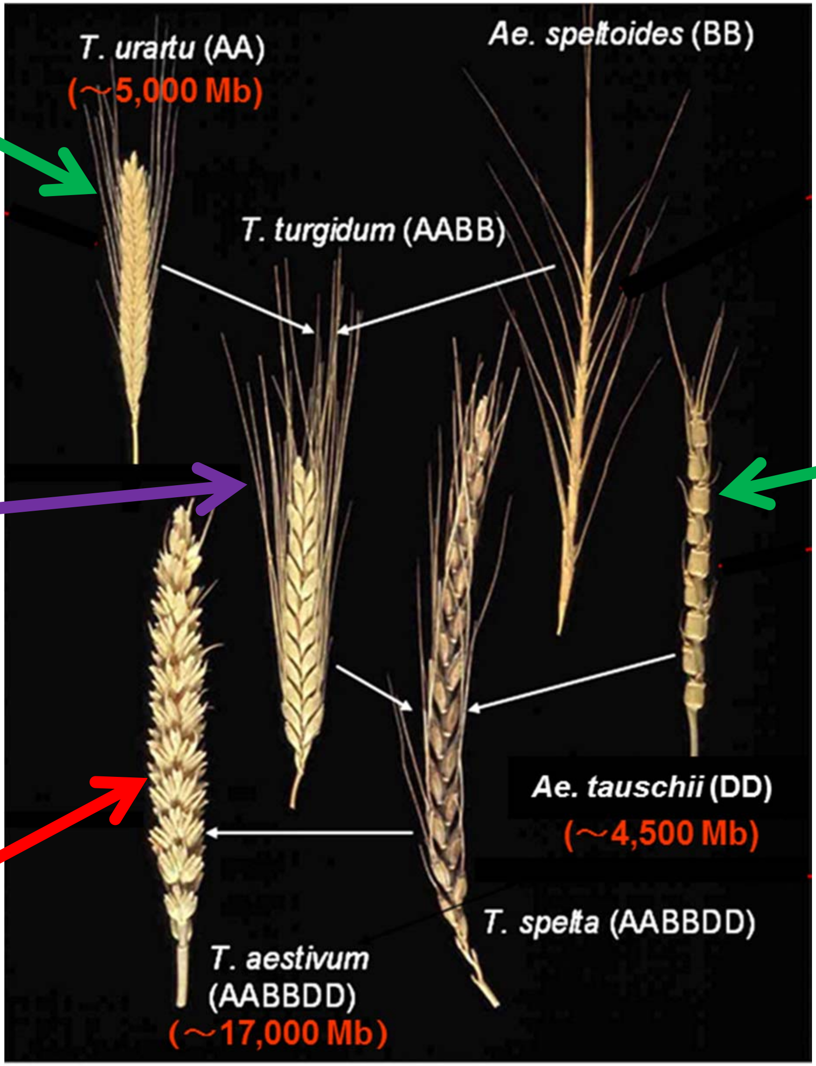
Wheat Genome Sequencing

IGDB, CAS: whole genome shotgun sequencing and BAC by BAC

WEWseq: *T. dicoccoides*, whole genome shotgun sequencing

UK: Chinese Spring, whole genome shotgun sequencing (WGS)

IWGSC: Chinese Spring, BAC by BAC, Chromosome-based WGS



ICS, CAAS: Whole genome shotgun sequencing

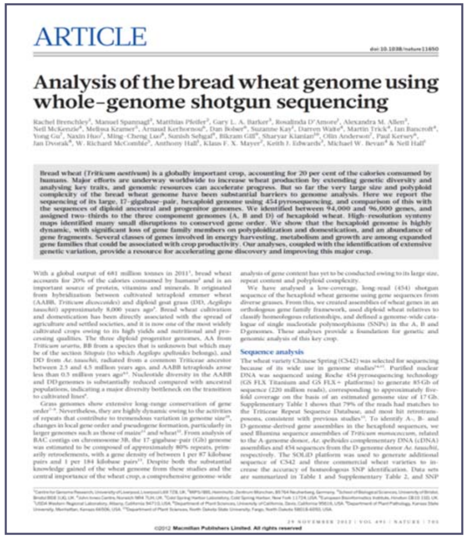
UC Davis: Physical map, BAC by BAC sequencing

中国科学院遗传与发育生物学研究所

Institute of Genetics and Developmental Biology, CAS

Wheat Genome Sequencing

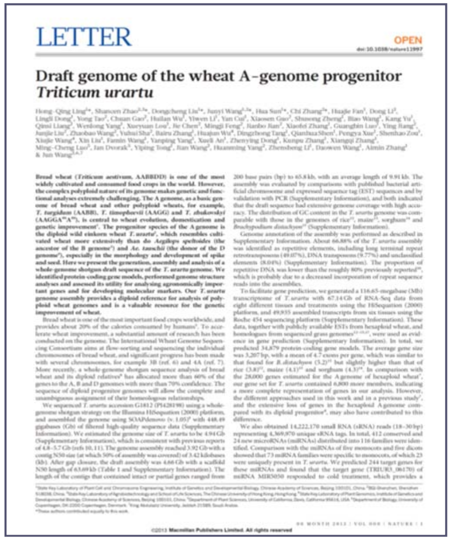
Brenchley et al., Nature, 2012



ARTICLE

Analysis of the bread wheat genome using whole-genome shotgun sequencing

Ling, et al., Nature, 2013



LETTER OPEN

Draft genome of the wheat A-genome progenitor *Triticum urartu*

Jia et al., Nature, 2013

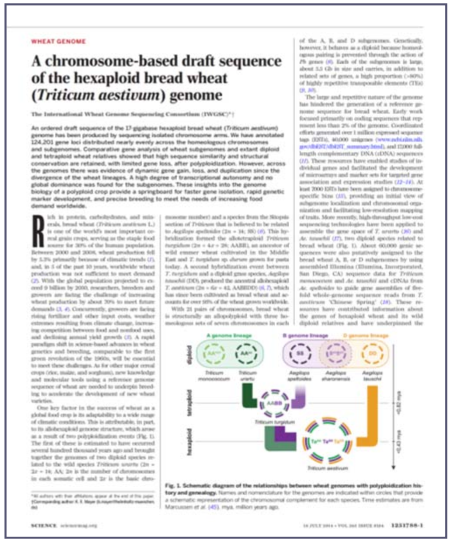
LETTER OPEN

***Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation**

Luo et al., PNAS, 2013

A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor

IWGSC, Science, 2014

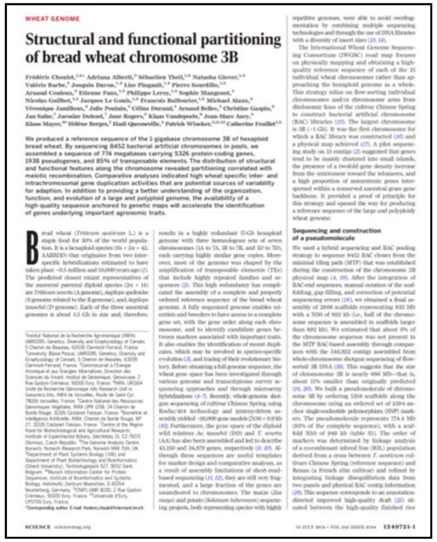


WHEAT GENOME

A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome

The International Wheat Genome Sequencing Consortium (IWGSC)*

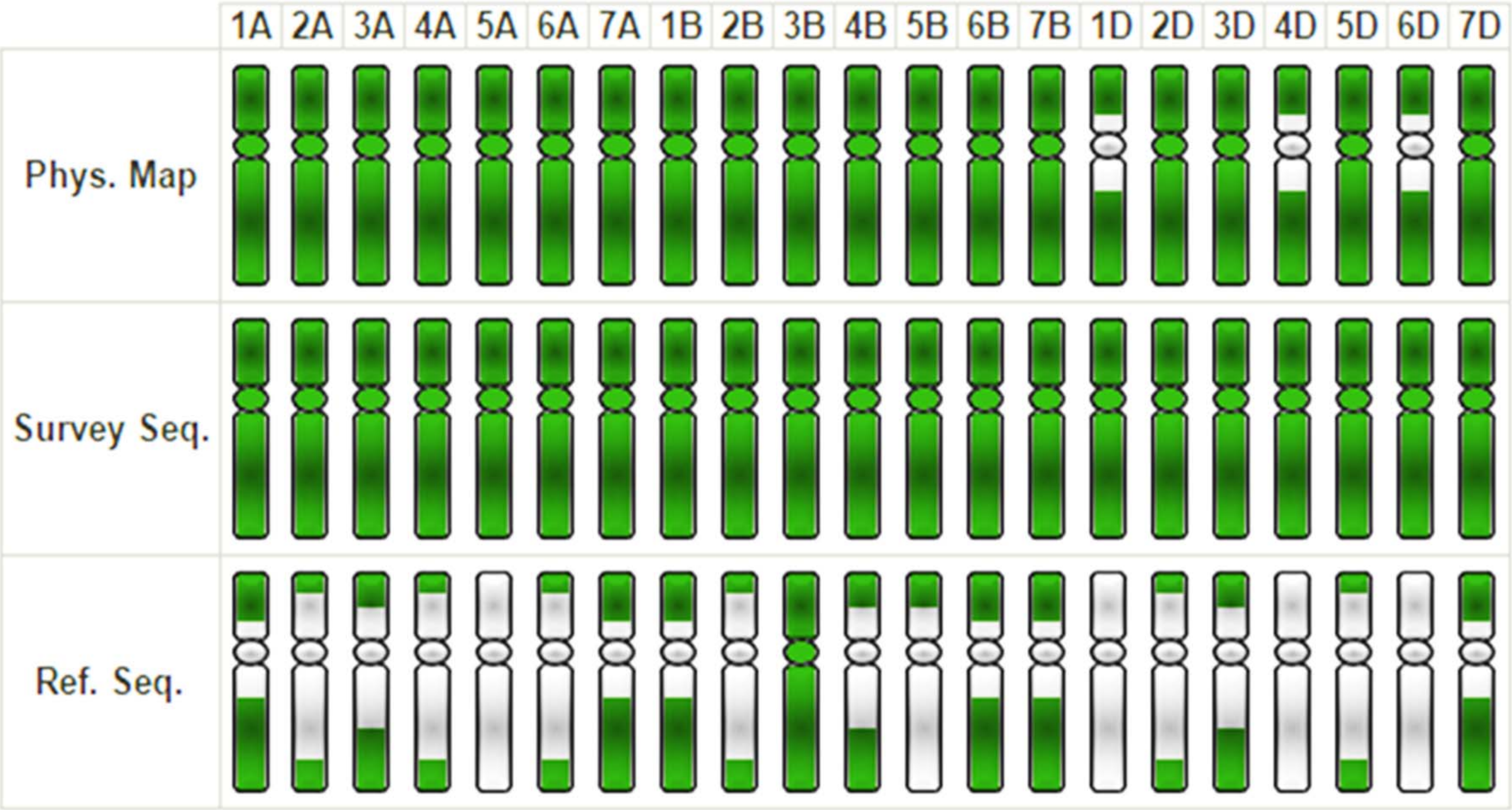
Choulet et al., Science, 2014



WHEAT GENOME

Structural and functional partitioning of bread wheat chromosome 3B

Status of IWGSC

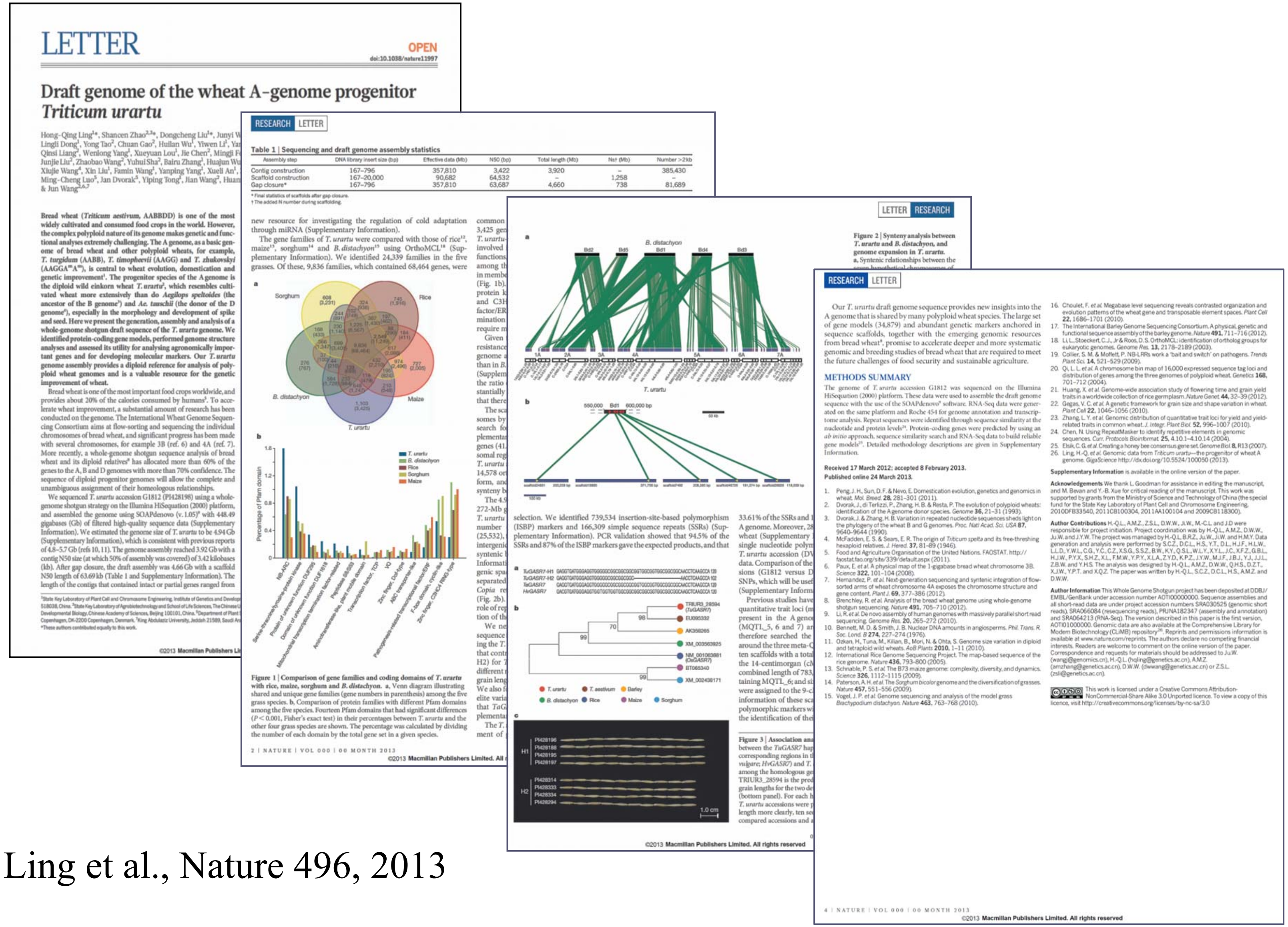


IWGSC: International wheat genome sequencing consortium

T. urartu Is a Good Plant for Genome Sequencing in Wheats

- ***Triticum urartu*** **is a diploid and the progenitor of wheat A genome.**
- **Its genome size (5,000 Mb) is about one third of bread wheat.**
- **Sequencing and assembling of *T. urartu* genome will provide a diploid reference for analysis of polyploid wheat genomes, simplify the sequencing problems.**
- **The genome sequence of *T. urartu* is also valuable for studying wheat evolution, domestication and even for genetic improvement of wheat**

Draft Genome of *Triticum urartu*



LETTER OPEN

doi:10.1038/nature11997

Draft genome of the wheat A-genome progenitor *Triticum urartu*

Ling et al., Nature 496, 2013

Wheat A Genome Sequencing

Goal:

Completing the genome sequencing of *T. urartu* using new sequencing technologies, and providing a diploid reference genome for analysis of polyploid wheat genomes, and for studying wheat evolution and even for genetic improvement of wheat.

Strategy:

- Constructing physical map, sequencing BAC clones, and completing the genome sequencing.
- Constructing a high density SNP map, and anchoring the genes and scaffolds on chromosomes.

Physical Mapping

◆ BAC Library Construction

Three genomic BAC libraries of *T. urartu* G1812 were constructed using *Hind*III, *Eco*RI and *Mbo*I. They contain 470,000 BAC clones with an average insert size of 120 kb.

◆ Physical Mapping

In collaboration with Keygene, 451,584 BAC clones were analyzed using the whole genome profiling approach. Of them, 345,233 BAC clones were available and used for constructing BAC contigs. The total size of the BACs was 8x of *T. urartu* genome.

BAC Contigs of *T. urartu*

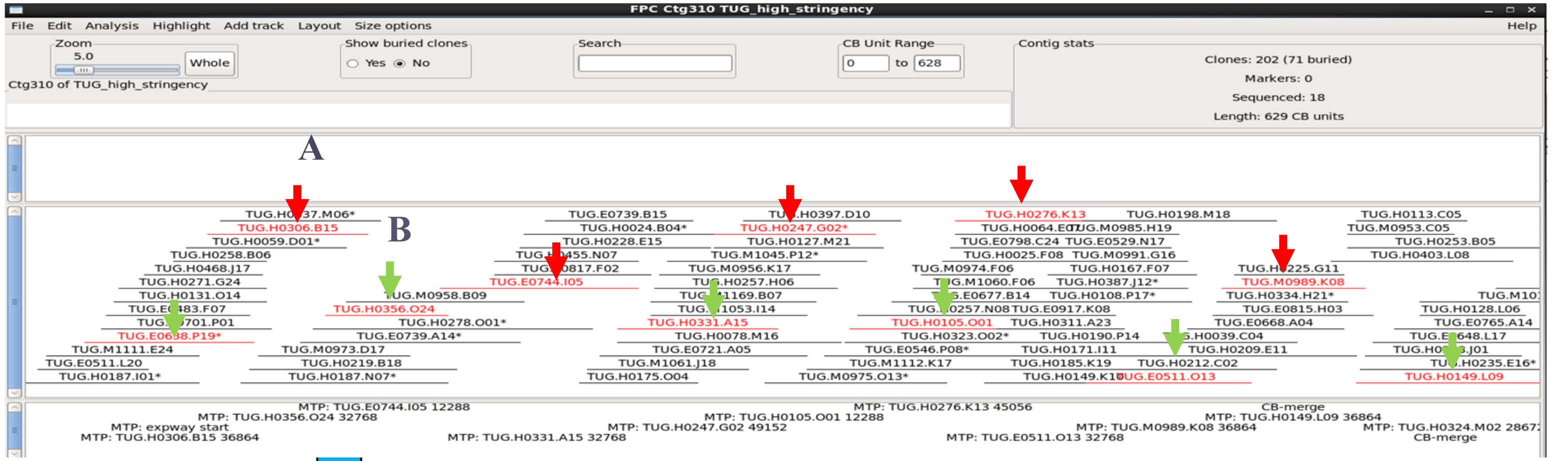
	HS-WGP	**LS-WGP**
Total No. of BACs in FPC	345,233	345,233
Contig No.	**20,702**	**12,137**
No. and % BACs in contigs	274,728 (80%)	323,058 (94%)
No. and % Singleton BACs	70,505 (20%)	22,175 (6%)
Coverage (Mbp)	5522	4688
Average contig size (BACs)	13.3	26.6
N50 contig size (BACs)	24	55
Average contig size (kb)	**267**	**386**
N50 contig size (kb)	**340**	**656**

HS-WGP: High stringency WGP map assembly
LS-WGP: Reduced stringency WGP map assembly

中国科学院遗传与发育生物学研究所

Institute of Genetics and Developmental Biology, CAS

BAC Selection and Sequencing



FPC Ctg310 TUG_high_stringency

Zoom 5.0

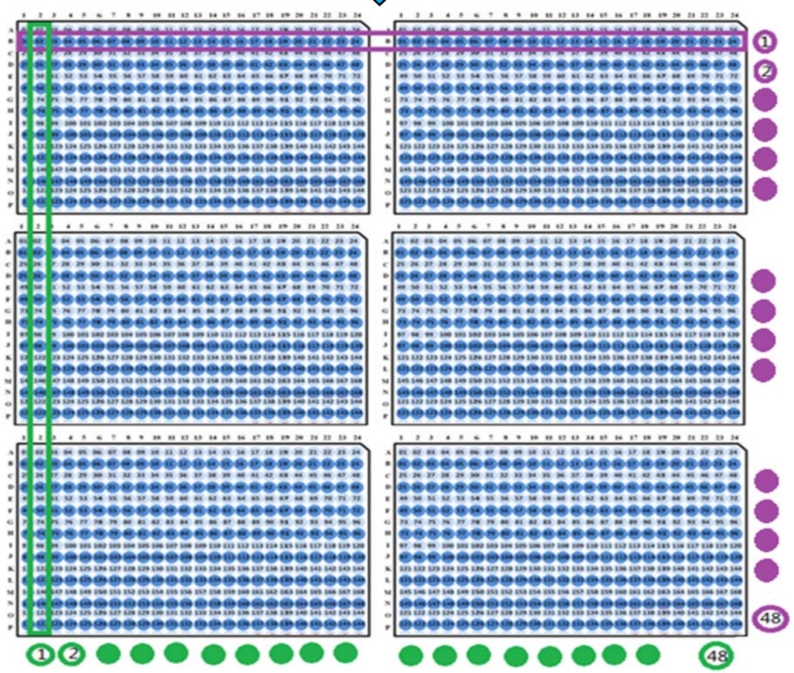
Show buried clones: ○ Yes ● No

CB Unit Range: 0 to 628

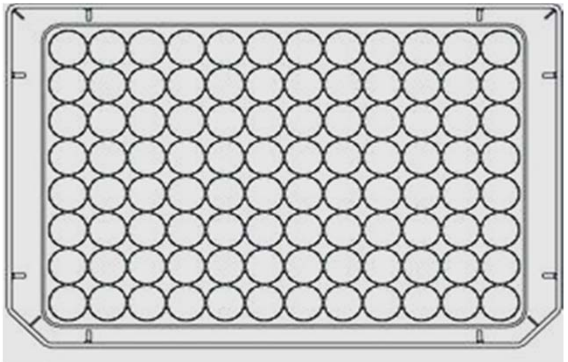
Ctg310 of TUG_high_stringency

Contig stats

Clones: 202 (71 buried)
Markers: 0
Sequenced: 18
Length: 629 CB units



48X48 Pooling

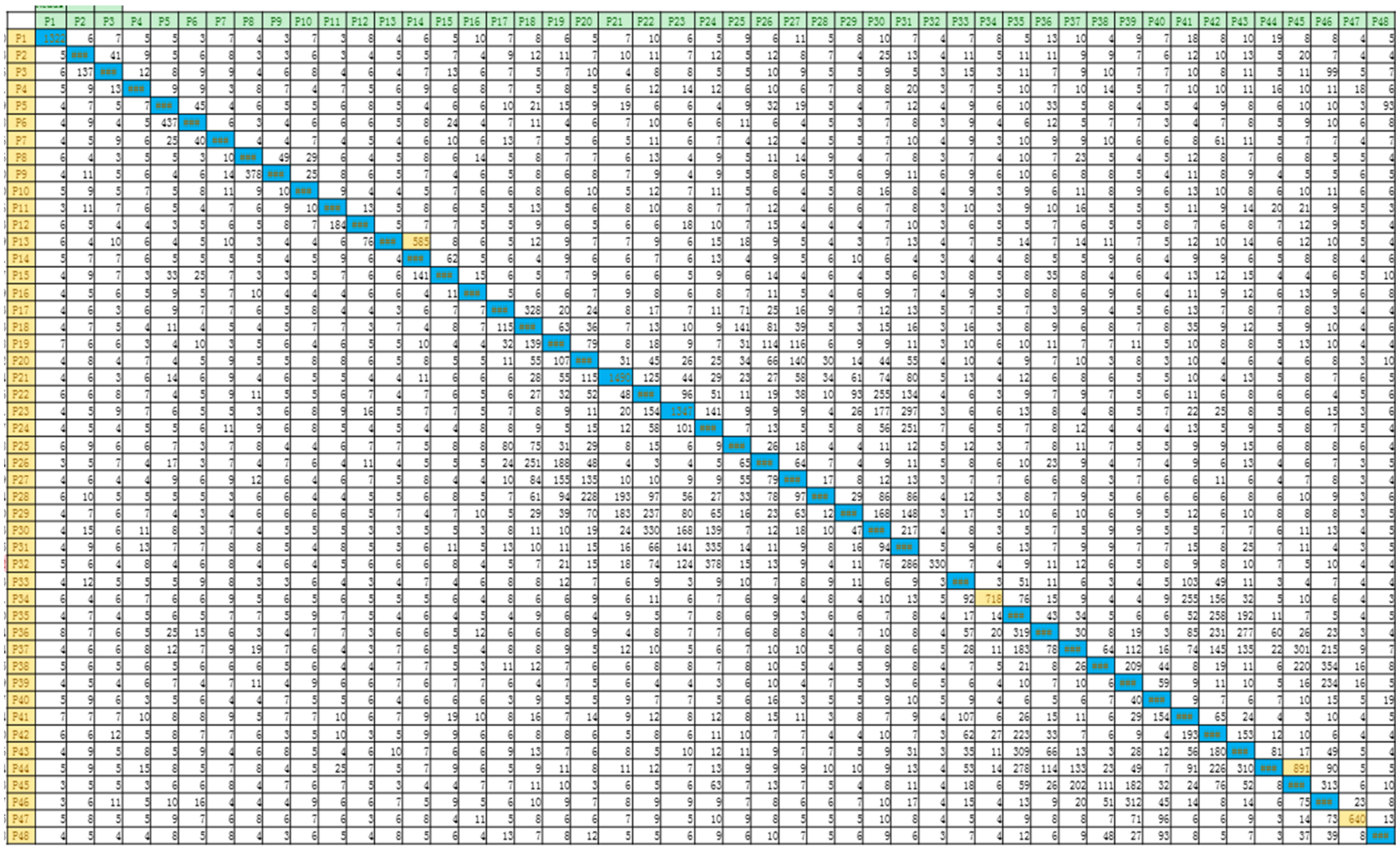


- **Selected BACs: 47223** (26400/A, 20823/B)
- **BAC pools: 2023**
- **Libraries: 23** (96 pools/lib)
- **5 runs/Hiseq2500** (2x150 bp)

Evaluation of BAC Sequencing Reads

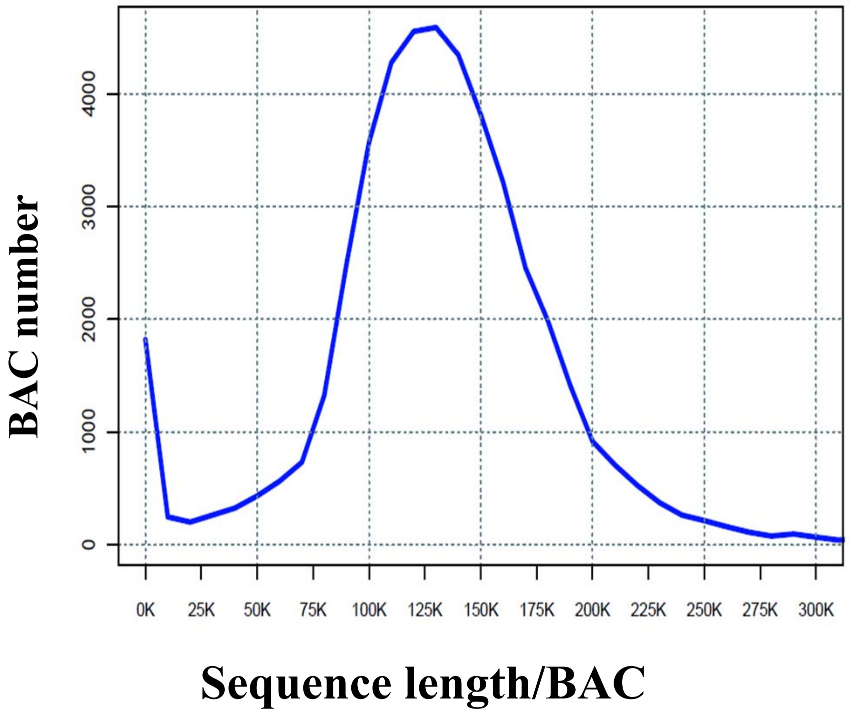
Sequenced reads

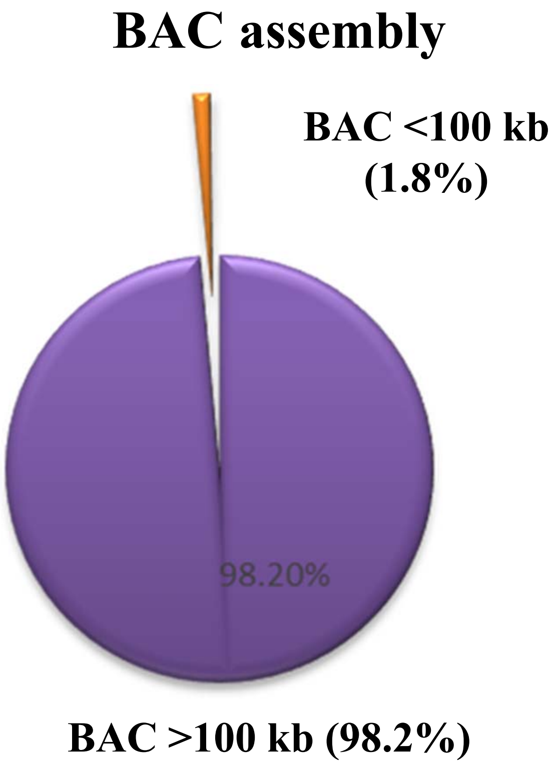
Physical map tags



	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24	P25	P26	P27	P28	P29	P30	P31	P32	P33	P34	P35	P36	P37	P38	P39	P40	P41	P42	P43	P44	P45	P46	P47	P48
P1	1322	6	7	5	5	3	7	4	3	7	3	8	4	6	5	10	7	8	6	5	7	10	6	5	9	6	11	5	8	10	7	4	7	8	5	13	10	4	9	7	18	8	10	19	8	8	4	5
P2	5	###	41	9	5	6	8	3	3	6	3	4	5	5	7	4	9	12	11	7	10	11	7	12	5	12	8	7	4	25	13	4	11	5	11	11	9	9	7	6	12	10	13	5	20	7	4	4
P3	6	137	###	12	8	9	9	4	6	8	4	6	4	7	13	6	7	5	7	10	4	8	8	5	5	10	9	5	5	9	5	3	15	3	11	7	9	10	7	7	10	8	11	5	11	99	5	7
P4	5	9	13	###	9	9	3	8	7	4	7	5	6	9	6	8	7	5	8	5	6	12	14	12	6	10	6	7	8	8	20	3	7	5	10	7	10	14	5	7	10	10	11	16	10	11	18	6
P5	4	7	5	7	###	45	4	6	5	5	6	8	5	4	6	6	10	21	15	9	19	6	6	4	9	32	19	5	4	7	12	4	9	6	10	33	5	8	4	5	4	9	8	6	10	10	3	93
P6	4	9	4	5	437	###	6	3	4	6	6	6	5	8	24	4	7	11	4	6	7	10	6	6	11	6	4	5	3	7	8	3	9	4	6	12	5	7	7	3	4	7	8	5	9	10	6	8
P7	4	5	9	6	25	40	###	4	4	7	4	5	4	6	10	6	13	7	5	6	5	11	6	7	4	12	4	5	5	7	10	4	9	3	10	9	9	10	6	6	8	61	11	5	7	7	4	7
P8	6	4	3	5	5	3	10	###	49	29	6	4	5	8	6	14	5	8	7	7	6	13	4	9	5	11	14	9	4	7	8	3	7	4	10	7	23	5	4	5	12	8	7	6	8	5	5	4
P9	4	11	5	6	4	6	14	378	###	25	8	6	5	7	4	6	5	8	6	8	7	9	4	9	5	8	6	5	6	9	11	6	9	6	10	6	8	8	5	4	11	8	9	4	5	5	6	5
P10	5	9	5	7	5	8	11	9	10	###	9	4	4	5	7	6	6	8	6	10	5	12	7	11	5	6	4	5	8	16	8	4	9	3	9	6	11	8	9	6	13	10	8	6	10	11	6	6
P11	3	11	7	6	5	4	7	6	9	10	###	13	5	8	6	5	5	13	5	6	8	10	8	7	7	12	4	6	6	7	8	3	10	3	9	10	16	5	5	5	11	9	14	20	21	9	5	3
P12	6	5	4	4	3	5	6	5	8	7	184	###	5	7	7	5	5	9	6	5	6	6	18	10	7	15	4	4	4	7	10	3	6	5	5	7	6	5	5	8	7	6	8	7	12	9	5	4
P13	6	4	10	6	4	5	10	3	4	4	6	76	###	585	8	6	5	12	9	7	7	9	6	15	18	9	5	4	3	7	13	4	7	5	14	7	14	11	7	5	12	10	14	6	12	10	5	4
P14	5	7	7	6	5	5	5	5	4	5	9	6	4	###	62	5	6	4	9	6	6	7	6	13	4	9	5	6	10	6	4	3	4	4	8	5	5	9	6	4	9	9	6	5	8	8	4	6
P15	4	9	7	3	33	25	7	3	3	5	7	6	6	141	###	15	6	5	7	9	6	6	5	9	6	14	4	6	4	5	6	3	8	5	8	35	8	4	6	4	13	12	15	4	4	6	5	10
P16	4	5	6	5	9	5	7	10	4	4	4	6	6	4	11	###	5	6	6	7	9	8	6	8	7	11	5	4	6	9	7	4	9	3	8	8	6	9	6	4	11	9	12	6	13	9	6	8
P17	4	6	3	6	9	7	7	6	5	8	4	4	3	6	7	7	###	328	20	24	8	17	7	11	71	25	16	9	7	12	13	5	7	5	7	3	9	4	5	6	13	7	8	7	8	3	4	4
P18	4	7	5	4	11	4	5	4	5	7	7	3	7	4	8	7	115	###	63	36	7	13	10	9	141	81	39	5	3	15	16	3	16	3	8	9	6	8	7	8	35	9	12	5	9	10	4	6
P19	7	6	6	3	4	10	3	5	6	4	6	5	5	10	4	4	32	139	###	79	8	18	9	7	31	114	116	6	9	9	11	3	10	6	10	11	7	7	11	5	10	8	8	5	13	10	4	4
P20	4	8	4	7	4	5	9	5	5	8	8	6	5	8	5	5	11	55	107	###	31	45	26	25	34	66	140	30	14	44	55	4	10	4	9	7	10	3	8	3	10	4	6	4	6	8	3	10
P21	4	6	3	6	14	6	9	4	6	5	5	4	4	11	6	6	6	28	55	115	1490	125	44	29	23	27	58	34	61	74	80	5	13	4	12	6	8	6	5	5	10	4	13	5	8	7	6	5
P22	6	6	8	7	4	5	9	11	5	5	6	7	4	7	6	5	6	27	32	52	48	###	96	51	11	19	38	10	93	255	134	4	6	3	9	7	9	7	5	6	11	6	8	6	6	4	5	6
P23	4	5	9	7	6	5	5	3	6	8	9	16	5	7	7	5	7	8	9	11	20	154	1347	141	9	9	9	4	26	177	297	3	6	6	13	8	4	8	5	7	22	25	8	5	6	15	3	6
P24	4	5	4	5	5	6	11	9	6	8	5	4	5	4	4	8	8	9	5	15	12	58	101	###	7	13	5	5	8	56	251	7	6	5	7	8	12	4	4	4	13	5	9	5	8	7	5	4
P25	6	9	6	6	7	3	7	8	4	4	6	7	7	5	8	8	80	75	31	29	8	15	6	9	###	26	18	4	4	11	12	5	12	3	7	8	11	7	5	5	9	9	15	6	8	8	6	4
P26	3	5	7	4	17	3	7	4	7	6	4	11	4	5	5	5	24	251	188	48	4	3	4	5	65	###	64	7	4	9	11	5	8	6	10	23	9	4	7	4	9	6	13	4	6	7	3	5
P27	4	6	4	4	9	6	9	12	6	4	6	7	5	8	4	4	10	84	155	135	10	10	9	9	55	79	###	17	8	12	13	3	7	7	6	6	8	3	7	6	6	11	6	4	7	8	3	4
P28	6	10	5	5	5	5	3	6	6	4	4	5	5	6	8	5	7	61	94	228	193	97	56	27	33	78	97	###	29	86	86	4	12	3	8	7	9	5	6	6	6	6	8	6	10	9	3	6
P29	4	7	6	7	4	3	4	6	6	6	6	5	7	4	7	10	5	29	39	70	183	237	80	65	16	23	63	12	###	168	148	3	17	5	10	6	10	6	9	5	12	6	10	8	8	8	3	5
P30	4	15	6	11	8	3	7	4	5	5	5	3	3	5	5	3	8	11	10	19	24	330	168	139	7	12	18	10	47	###	217	4	8	3	5	7	5	9	9	5	5	7	7	6	11	13	4	5
P31	4	9	6	13	7	7	8	8	5	4	8	5	5	6	11	5	13	10	11	15	16	66	141	335	14	11	9	8	16	94	###	5	9	6	13	7	9	9	7	7	15	8	25	7	11	4	3	5
P32	5	6	4	8	4	6	8	4	6	4	5	6	6	6	8	4	5	7	21	15	18	74	124	378	15	13	9	4	11	76	286	330	7	4	9	11	12	6	5	8	9	8	10	7	5	10	4	4
P33	4	12	5	5	5	9	8	3	3	6	4	3	4	7	4	6	8	8	12	7	6	9	3	9	10	7	8	9	11	6	9	3	###	3	51	11	6	3	4	5	103	49	11	3	4	7	4	4
P34	6	4	6	7	6	6	9	5	6	5	6	5	5	5	6	4	8	6	6	9	6	11	6	7	6	9	4	8	4	10	13	5	92	718	76	15	9	4	4	9	255	156	32	5	10	6	4	3
P35	4	7	4	5	6	5	7	7	5	9	7	5	4	6	4	5	4	9	6	4	9	5	7	8	6	9	3	6	6	7	8	4	17	14	###	43	34	5	6	6	52	258	192	11	7	5	4	5
P36	8	7	6	5	25	15	6	3	4	7	7	3	6	6	5	12	6	6	8	9	4	8	7	7	6	8	8	4	7	10	8	4	57	20	319	###	30	8	19	3	85	231	277	60	26	23	3	4
P37	4	6	6	8	12	7	9	19	7	6	4	3	7	6	5	4	8	8	9	5	12	10	5	6	7	10	10	5	6	8	6	5	28	11	183	78	###	64	112	16	74	145	135	22	301	215	9	7
P38	5	6	5	6	5	6	6	6	5	6	4	4	7	7	5	3	11	12	7	6	6	8	8	7	8	10	5	4	5	9	8	4	7	5	21	8	26	###	209	44	8	19	11	6	220	354	16	8
P39	4	5	4	6	7	4	7	11	4	9	6	6	7	6	7	7	6	4	7	5	6	4	4	3	6	10	4	7	5	3	6	5	6	4	10	7	10	6	###	59	9	11	10	5	16	234	16	5
P40	5	9	6	3	5	6	4	4	7	5	5	6	4	6	6	6	3	9	5	5	9	7	7	5	6	16	3	5	5	9	10	5	9	4	6	5	6	7	40	###	9	7	6	7	10	15	5	19
P41	7	7	7	10	8	8	9	5	7	7	10	6	7	9	19	10	8	16	7	14	9	12	8	12	8	15	11	3	8	7	9	4	107	6	26	15	11	6	29	154	###	65	24	4	3	10	4	5
P42	6	6	12	5	8	7	7	6	3	5	10	3	5	9	9	6	6	8	8	6	5	8	6	11	10	7	7	4	4	10	7	3	62	27	223	33	7	6	9	4	193	###	153	12	10	6	4	4
P43	4	9	5	8	5	9	4	6	8	5	4	6	10	7	6	6	9	13	7	6	8	5	10	12	11	9	7	7	5	9	31	3	35	11	309	66	13	3	28	12	56	180	###	81	17	49	5	4
P44	5	9	5	15	8	5	7	8	4	5	25	7	5	7	9	6	5	9	11	8	11	12	7	13	9	9	9	10	10	9	13	4	53	14	278	114	133	23	49	7	91	226	310	###	891	90	5	5
P45	3	5	5	3	6	6	8	4	7	6	7	5	3	4	4	7	7	11	10	3	6	5	6	63	7	13	7	5	4	8	11	4	18	6	59	26	202	111	182	32	24	76	52	8	###	313	6	10
P46	3	6	11	5	10	16	4	4	4	9	6	6	7	5	9	5	6	10	9	7	8	9	9	9	7	8	6	6	7	10	17	4	15	4	13	9	20	51	312	45	14	8	14	6	75	###	23	8
P47	5	8	5	5	9	7	6	8	6	7	6	3	6	8	4	11	5	8	6	6	7	9	5	10	9	8	5	5	5	10	8	4	5	4	9	8	8	7	71	96	6	6	9	3	14	73	640	13
P48	4	5	4	4	8	5	8	4	3	6	5	4	8	5	6	4	13	7	8	12	5	5	6	9	6	10	7	5	6	9	6	3	7	4	12	6	9	48	27	93	8	5	7	3	37	39	8	###

Length Distribution of Assembled BACs





Sequenced Data

Data Type	Library Insert Size	Library	Lane	Cells	Average Read Length (bp)	Raw Data（Gb）	Usable Data（Gb）	Effective Depth
BAC	～300 bp	34	33		150	1,630	1,150	230
Pacbio				109	8,000	97	97	19
WGS	～500 bp	1	2		250	130	107	21
Total		35	35	109		1,857	1,354	271

Calculated with estimated genome size of 5Gb

Assembly Results

	Contig			Scaffold	
	Size (bp)	Number		Size (bp)	Number
N95	69,250	21,282		172,395	5,563
N90	99,508	18,422		262,382	4,436
N80	135,055	14,272		480,285	3,097
N70	167,418	11,044		710,158	2,271
N60	202,973	8,400		929,688	1,677
N50	241,179	6,201		1,169,259	1,213
Longest	1,570,203			6,267,833	
Total Size	4,863,000,000			4,883,161,420	
Total number（≥1Kb）		28,535			8,023
GC ratio（%）	45.931			45.930	

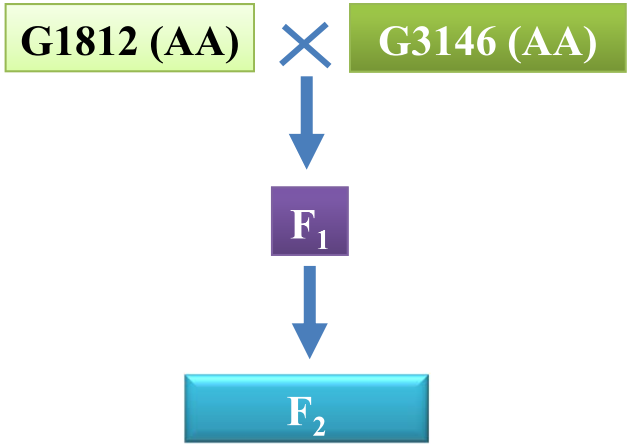
Characters of the Predicted Genes

Gene number	**42,488**
Gene length (Avg.)	**3,379**
Transcript length (Avg.)	**1,497**
Max CDS length	**15,036**
CDS length (Avg.)	**998**
Protein length (Avg.)	**331**
Exon length (Avg.)	**274**
Intron length (Avg.)	**524**
Exon number(Avg.)	**5.5**
Intron number(Avg.)	**4.5**

Constructing a high density SNP map and anchoring the Sequences on chromosomes.

Construction of High Density SNP Map

Mapping population and RAD-sequencing



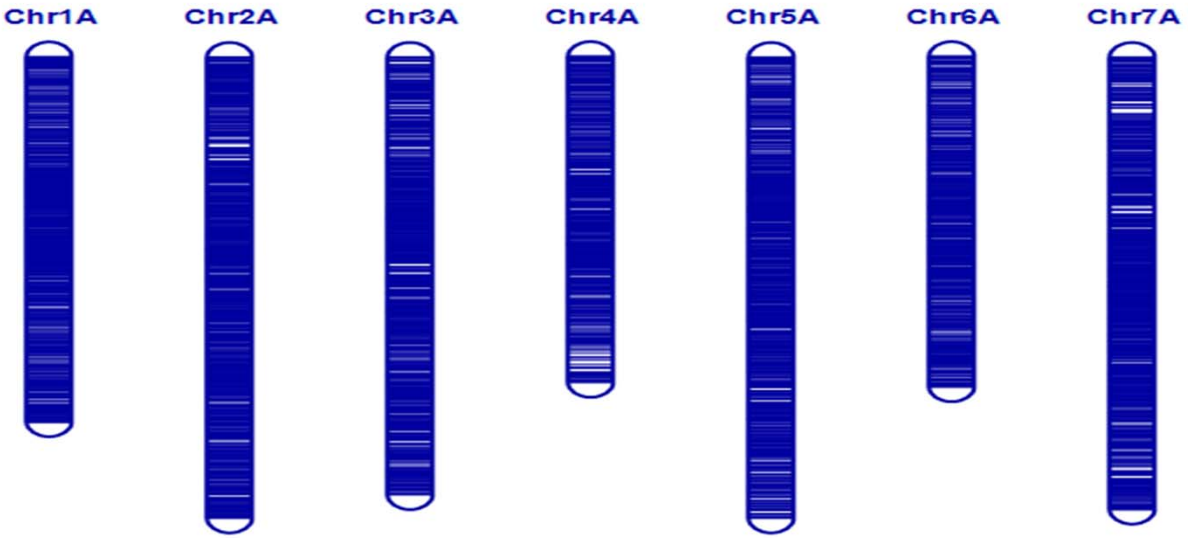
G1812 (AA) × G3146 (AA)

F_1

F_2

- RAD-Sequencing of 471 F_2 lines.
- Sequencing depth: 8x coverage per line.
- 200,293 Indels and 579,126 SNPs were identified.

High density SNP map



Assigning sequences on chromosomes

Chromosome	Sequence Length	Gene Number
Chr1	546,352,135	4,639
Chr2	743,055,976	6,348
Chr3	745,435,511	5,979
Chr4	608,737,882	4,033
Chr5	645,961,361	5,735
Chr6	566,958,958	4,473
Chr7	715,110,063	6,364
Total	**4,571,611,886**	**37,571**

Summary

- The A genome is a basic genome of bread wheat and other polyploid wheats, played a central role in wheat evolution, domestication and genetic improvement.
- *T. urartu* is the progenitor of the wheat A genome. We have generated its high quality genome sequence with 4.86 Gb by sequencing BACs. **The contig N50 reaches to 241kb, and scaffold N50 is 1.17 Mb**.
- We predicted 42,488 protein-coding gene models with assembled sequence.
- A high density SNP genetic map has been constructed. With it, more than 93% of the sequences have been anchored onto its corresponding chromosome.

Acknowledgments:

Dr. Chengzhi Liang

Dr. Hui Liu

Dr. Hua Sun

Mr. Bin Ma

Other colleagues worked on this project

Collaboration:

BGI-Shenzhen

Keygene N.V. (Dr. Michiel van Eijk)

Funding: **Chinese Ministry of Science and Technology**

中国科学院遗传与发育生物学研究所

Institute of Genetics and Developmental Biology, CAS

Thanks !