



Institute of Genetics and Developmental Biology, CAS

Draft Genome of *Triticum urartu* and its Physical Mapping

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



Production of the Main Food Crops in the World and China

(Data from FAOSTAT 2011)



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.



- Bread wheat is an allopolyploid, containing A, B and D subgenomes, which derived from a common ancestor. They show a high similarity each other.
- Bread wheat has a huge genome size (~17 Gb). It is about 8 times larger than maize, 40 times than rice and 100 times than Arabidopsis thaliana.
- More than 85% of the genome sequence are repetitive DNA

These characters of wheat make genome study very difficulty. Although some progress has been made, sequencing of wheat genome is still a big challenge.

构



Institute of Genetics and Developmental Biology, CAS

Current Status of IWGSC





T. urartu Is the Progenitor of Wheat A Genome



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

2014-10-02



The A Genome is a Basic Genome of Wheat



The A genome is a basic genome of bread wheat and other polyploid wheats, plays a central role in wheat evolution, domestication and genetic improvement.



Sequencing and assembling of *T. urartu* genome will provide a diploid reference for analysis of polyploid wheat genomes

The genome sequence is also valuable for studying wheat evolution, domestication and even for genetic improvement of wheat







Sequencing Wheat A Genome



The whole genome of *T. urartu* accession G1812 were sequenced, collaborated with BGI-Shenzhen, using shotgun sequencing approach with Illumina's next-generation sequencing platforms.



Sequence Data

Library	Number 1	Number	· Average	Raw	Usable	Ffective	Physical	Physical
insert size	0f libraries	0f Janes	reads	data (Cb)	data (Cb)	depth*	data (Cb)	depth*
2001	15		111	212.00	164.00	22.20	201.00	(5.02
~200 bp) 15	22	111	213.96	164.88	33.38	321.20	65.03
~350 bp) 7	16	61	49.00	36.07	7.30	210.98	42.71
~500 bp) 10	19	82	121.49	91.87	18.60	576.65	116.73
~700 bp	9	13	79	95.43	64.98	13.15	573.49	116.09
2 kb	6	18	51	56.08	42.67	8.64	1653.53	334.72
5 kb	4	14	53	45.82	32.50	6.58	3046.77	616.76
10 kb	4	5	44	15.51	11.98	2.43	2723.43	551.30
20 kb	2	2	44	12.31	3.53	0.71	1603.93	324.68
Total	57	109	77	609.61	448.49	90.79	10710.05	2168.03

*Calculated with the estimated genome size of 4.94 Gb.

例

与な育

¥



Genome Assembly*

	Contig		Scaffo	ld
	Size (bp)	Number	Size (bp)	Number
N90	123	5,521,147	126	3,978,271
N80	127	4,715,825	232	410,092
N70	777	715,986	19,387	43,008
N60	1,903	400,253	42,913	27,515
N50	3,422	246,789	63,687	18,663
Longest	82,078		1,066,088	
Total size	3,922,395,337		4,660,785,691	
Total number (≥ 1 kb)		622,928		133,906
Total number (\geq 2 kb)		385,430		81,698
GC ratio (%)	45.388		45.388	

The average length of the contigs containing intact or partial genes was 9.91 kb.

*SOAPdenovo (v.1.05), (Li et al., 2010, Genome Res. 20: 265-272)





Genome Annotation and Comparative Analysis



Analysis of Repetitive Elements

	Percentage of genome (%)			Length (bp)		
	Bd	Sb	Os	Zm	Tu	Tu
Class I: Retrotransposon	21.58	50.77	21.00	76.35	49.07	1,765,277,214
LTR-Retrotransposon	18.38	49.70	19.85	75.52	46.66	1,678,595,438
LTR/Gypsy	13.77	42.85	16.39	48.43	36.57	1,315,436,369
LTR/Copia	4.46	6.81	3.08	26.55	9.89	355,762,130
Other	0.15	0.04	0.38	0.54	0.21	7,396,939
Non-LTR Retrotransposon	3.20	1.07	1.16	0.84	2.41	86,681,776
SINE	0.26	0.08	0.05	0.03	0.07	2,566,147
LINE	2.94	0.98	1.11	0.80	2.34	84,115,629
Class II: DNA Transposon	5.33	7.17	5.82	5.39	9.77	351,279,176
DNA Transposon Superfamily	3.32	4.73	2.75	3.37	7.33	263,725,407
DNA-CACTA	1.44	3.67	2.38	2.06	5.44	195,685,200
hAT	0.43	0.26	0.27	0.75	0.37	13,201,783
Harbinger	0.26	0.20	0.08	0.22	0.20	7,238,463
Tc1/Mariner	1.19	0.61	0.03	0.07	0.56	19,993,205
MITE	1.95	2.31	3.07	0.77	1.88	67,769,240
Tourist	0.28	1.47	1.11	0.12	0.32	11,503,170
Stowaway	0.14	0.09	0.60	0.00	0.05	1,839,920
Other	1.53	0.74	1.37	0.66	1.51	54,426,150
Helitron	0.06	0.13	0.00	0.54	0.01	270,967
Tandem repeat	1.89	2.49	2.90	0.86	1.21	43,630,554
Low complexity	0.27	0.19	0.82	0.12	0.10	3,489,764
Unclassified	8.41	5.21	0.23	0.74	8.04	289,349,611
Total content	37.48	65.83	30.78	82.48	66.88	2,405,906,001

Bd, B. distachyon; Sb, S. bicolor; Os, O. sativa; Zm, Z. mays; Tu, T. urartu.



Ŀ

RNA-Seq and Transcriptome Assembly

Organ	Usable data (Gb)	Transcripts	Average length (bp)	Maximum length (bp)	Total size of transcripts (bp)
2mR	9.38	54,601	1,096	11,937	59,830,667
YS	18.98	81,950	1,422	15,381	116,559,256
2mL	13.57	55,750	1,311	15,365	73,098,093
5dS	8.58	47,397	1,275	1,437	60,429,264
10dR	4.44	43,350	1,101	15,213	47,732,145
10dL	3.47	36,133	1,045	12,185	37,770,880
7wL	4.02	42,950	1,205	15,205	51,757,961
CL	4.70	50,838	1,208	23,741	61,427,836
Integration	67.14	92,868	1,256	15,378	116,650,180
454-est	0.34	49,935	406	6,088	20,290,712

녉

な育

物

Ž



Institute of Genetics and Developmental Biology,

CAS

いもなう

生物資料

Flowchart of Gene Prediction



In total, 34,879 protein-coding gene models have been predicted

2014-10-02



Comparison of Gene Features



Institute of Genetics and Developmental Biology, CAS 学院送供与发育生物学研 家所



Institute of Genetics and Developmental Biology,

CAS

カ

育

教育

Gene Family Comparison



68464 genes were clustered together in 9836 gene families, which were shared by all five monocots, with 1103 gene families (3425 genes) being specific to the *T. urartu* genome.

2014-10-02



Pfam Domain Comparison



2,067 Pfam domains were shared among the five sequenced monocot species. Of them, 14 Pfam domains had differences in member numbers in *T. urartu* compared to the other four grasses.

2014-10-02



R Proteins in *T. urartu* and Other Grass Genomes

Predicted Protein Domain	T. urartu	Brachypodium	Rice	Maize	Sorghum
CC-NBS-LRR	174	121	246	44	96
NBS-LRR	247	40	134	27	92
CC-NBS	69	24	40	19	13
NBS	103	12	40	16	20
TIR-NBS-LRR	0	0	0	0	0
Total	593	197	460	106	211

2014-10-02

与な言

ł

粉香料

Identification of miRNAs



412 conserved and 24 new miRNAs distributed into 116 families were identified. Comparison to the miRNAs of five monocots and five dicots showed that 73 miRNA families were monocot specific, of which 23 were uniquely present in *T. urartu*.





Ŀ

例

与な育

ł

粉管研

爽

Predicted Target Genes of New miRNAs

miRNA name	Target gene number	InterPro	Description of target gene
Sr126	15	IPR000863	Sulfotransferase domain
		IPR003690	Mitochodrial transcription termination factor-related
		IPR002866	Maturase, MatK
Sr197	9	IPR002885	Pentatricopeptide repeat
Sr2072	14	IPR000767	Disease resistance protein
		IPR002182	NB-ARC
		IPR012871	Protein of unknown function DUF1677
		IPR000058	Zinc finger, AN1-type
Sr2145	1	IPR000432	DNA mismatch repair protein MutS, C-terminal domain
Sr26942	4	IPR009057	Homeodomain-like
		IPR003100	Argonaute/Dicer protein, PAZ
Sr4181	1	IPR007749	Protein of unknown function DUF677





Scaffold Anchoring, Synteny Analysis



Assigning Scaffolds on Chromosome Bins

The scaffolds and gene models of *T. urartu* were assigned to chromosome bins using genetically mapped bread wheat ESTs20 as queries to search for homologous sequences in the *T. urartu* assembly. A total of 8,715 scaffolds, harboring 14,578 genes (41.8% of the total predicted genes) were mapped to 45 chromosomal bins of the wheat A genome.



Scaffolds on Chromosome 1

	145	1-0. 47-0. 86	(76)	
caffold97192	scaffold59285	scaffold363	scaffold21259	scaffold61734
affold96181	scaffold58908	scaffold35667	scaffold18922	scaffold61699
ffold87008	scaffold55238	scaffold35040	scaffold17867	scaffold61049
Told86748	scaffold53717	scaffold34518	scaffold16529	scaffold40713
fold83905	scaffold53572	scaffold32522	scaffold16436	scaffold40664
fold8254	scaffold52747	scaffold32244	scaffold150163	scaffold3991
affold7375	scaffold5232	scaffold32226	scaffold14708	scaffold36822
affold71963	scaffold52070	scaffold3196	scaffold143370	scaffold26853
affold71900	scaffold50529	scaffold20026	scaffold13089	scaffold25981
affold7153	scaffold4655	scaffold2847	scaffold116121	scaffold2306
affold7102	scaffold465	scaffold27522	scaffold11234	scaffold104387
ffold67281	scaffold41501	scaffold273645	scaffold110143	scaffold100036
affold65635	scaffold41064	scaffold27017	scaffold1096	C167360911
167032996				
	17 scaffo]	d41939	scaffold32	677
		100050		
CIA	scaffol	d72350 scaf	Fo1d29044	
	C-	1AL1-0.17 (13)	
caffold88814	scaffold32593	scaffold41587	scaffold60344	scaffold27674
caffold71072	scaffold31684	scaffold13968	scaffold50701	scaffold18186
scaffold6597	scarrold27996	scaffold131558		

1		1AL1	0. 17-0. 61	(150)	
	scaffold957	scaffold56021	scaffold38156	scaffold27780	scaffold13794
	scaffold9488	scaffold55302	scaffold37814	scaffold27552	scaffold137564
	scaffold93020	scaffold54710	scaffold36267	scaffold26931	scaffold13694
	scaffold9148	scaffold5392	scaffold36257	scaffold26834	scaffold13314
	scaffold90402	scaffold53670	scaffold36029	scaffold26741	scaffold131040
	scaffold90037	scaffold53193	scaffold3582	scaffold24836	scaffold13091
	scaffold88034	scaffold52676	scaffold35728	scaffold24531	scaffold129878
	scaffold87089	scaffold52266	scaffold35068	scaffold24208	scaffold12985
	scaffold8499	scaffold51676	scaffold3501	scaffold23885	scaffold12879
	scaffold82914	scaffold50412	scaffold34578	scaffold228825	scaffold128060
	scaffold82091	scaffold4964	scaffold34444	scaffold22680	scaffold12605
	scaffold8113	scaffold48740	scaffold33414	scaffold224796	scaffold124497
	scaffold8087	scaffold48264	scaffold33328	scaffold22410	scaffold1235
	scaffold77342	scaffold4786	scaffold33308	scaffold217098	scaffold122782
	scaffold74731	scaffold47696	scaffold33094	scaffold20878	scaffold12182
	scaffold72938	scaffold46429	scaffold327706	scaffold20712	scaffold11517
	scaffold68800	scaffold46360	scaffold326200	scaffold204864	scaffold111776
	scaffold66553	scaffold46267	scaffold32233	scaffold20400	scaffold1101
	scaffold66347	scaffold46230	scaffold32008	scaffold2033	scaffold11001
	scaffold65499	scaffold46110	scaffold30961	scaffold20149	scaffold109834
	scaffold65237	scaffold45467	scaffold3025	scaffold198392	scaffold10903
	scaffold64458	scaffold45028	scaffold30149	scaffold19710	scaffold108158
	scaffold62587	scaffold4185	scaffold293308	scaffold18218	scaffold10782
	scaffold62211	scaffold4114	scaffold29077	scaffold175984	scaffold107064
	scaffold61640	scaffold42258	scaffold2887	scaffold1627	scaffold104427
	scaffold60896	scaffold41830	scaffold28720	scaffold162524	scaffold102555
	scaffold5984	scaffold39916	scaffold286438	scaffold161630	scaffold101933
	scaffold5907	scaffold39251	scaffold28333	scaffold161176	scaffold101037
	scaffold59009	scaffold39209	scaffold27902	scaffold160604	C166821380
	scaffold56368	scaffold39027	scaffold27891	scaffold1450	C155818264

scaffold7153 scaffold4655 scaf C-1AS1-0.47 scaffold7102 scaffold465 scaf scaffold67281 scaffold41501 scaf scaffold65635 scaffold41064 scaf C167032886 C-1ASI-0.47 scaffold41 C-C1A C1A scaffold72 C-1AL scaffold88814 scaffold32593 scaf C-1AL1-0.17 scaffold71072 scaffold31684 scaf scaffold6597 scaffold27996 scaf 1AL1-0.17-0.61 1AL3-0.61-1.00 (109) scaffold96731 scaffold72316 scaffold49307 scaffold32882 scaffold15669 scaffold928 scaffold70570 scaffold48896 scaffold30877 scaffold156523 scaffold91997 scaffold69548 scaffold4832 scaffold275364 scaffold145620 scaffold88153 scaffold68270 scaffold46549 scaffold27294 scaffold137180 scaffold88145 scaffold68139 scaffold4193 scaffold27238 scaffold135382 scaffold87963 scaffold66584 scaffold41796 scaffold27061 scaffold134565 scaffold87106 scaffold64481 scaffold4145 scaffold26730 scaffold132870 scaffold8606 scaffold64141 scaffold41042 scaffold2625 scaffold132286 scaffold85770 scaffold63206 scaffold40605 scaffold25963 scaffold12774 scaffold844 scaffold61195 scaffold4009 scaffold25037 scaffold12325 scaffold83946 scaffold5892 scaffold3956 scaffold2321 scaffold123200 scaffold83806 scaffold5825 scaffold39550 scaffold23003 scaffold1211 scaffold83193 scaffold57164 scaffold39484 scaffold2292 scaffold11537 scaffold82996 scaffold5631 scaffold38686 scaffold21722 scaffold11427 scaffold80366 scaffold55701 scaffold38363 scaffold21050 scaffold111377 1AL3-0.61-1.00 scaffold799 scaffold55533 scaffold3835 scaffold194443 scaffold110533 scaffold79752 scaffold55329 scaffold38278 scaffold17797 scaffold107418 scaffold79275 scaffold53563 scaffold3690 scaffold17737 scaffold10713 scaffold7692 scaffold51037 scaffold36102 scaffold17640 scaffold10112 scaffold7570 scaffold50579 scaffold34149 scaffold17467 scaffold100235 scaffold74829 scaffold50050 scaffold34042 scaffold173 C167338323

scaffold72501 scaffold49854 scaffold33754 scaffold15858



1AS3-0.86-1.00

1AS1-0.47-0.86

1A



Phylogenetic Relationship of *T. urartu* and Four Sequenced Grasses





Institute of Genetics and Developmental Biology, CAS

与发育生物学研

Synteny Analysis between *B. distachyon* and *T. urartu*



Triticum urartu

Wheat A genome

Genome Expansion

Species	Genome size	Fold
T. urartu	5,000 Mb	
B. distachyon	272 Mb	18.4
O. sativa	466 Mb	10.7
S. bicolor	730 Mb	6.8
Z. mays	2,300 Mb	2.2



Analysis of Synteny Blocks in Gene-rich Regions between *T. urartu* and *B. distachyon*

Region	T. urartu (bp)	B. distachyon (bp)	Ratio
Intergenic	24,128,306	10,827,354	2.23
Gene	20,118,863	18,821,875	1.07
CDS	7,848,156	8,380,692	0.94
Intron	12,270,707	10,441,183	1.18
Total	108,613,201	78,120,333	1.39

The 2,498 synteny blocks contained 7,559 genes, indicating that 21% of *T. urartu* genes had similarly sized intergenic spaces to those of in *B. distachyon*.



Intergenic Space Expansion in *T. urartu* Genome



A synteny block containing 5 genes in a 50-kb region on chromosome 1 of *B. distachyon*. Their orthologous genes of *T. urartu* were separated in five scaffolds with a total length of >1091 kb. The intergenic spaces were expanded >21 fold in *T. urartu* by repetitive DNA (*Gypsy, Copia*), compared to *B. distachyon*.

Ž



Institute of Genetics and Developmental Biology, CAS

もなう育

生物資研

薮

Assessing the Utility of the Genome Sequence

TuGASR7, a Gene Controlling Seed Length







Institute of Genetics and Developmental Biology, CAS 例 与な言 生物堂研 薮

Homologous gene of rice GW2





- 36 homologs of *Pm3b* in *T. urartu* genome have been identified. Fifteen of them revealed a high sequence identity (Evalue = 0).
- 21 of the 36 *Pm3b*-homologs have been assigned to chromosome. Of them, 9 homologs were assigned to the chromosome 1AS of wheat.





国际

Molecular Marker Development

Marker type	Number	A-specific
SSR	166,309	33.6%
ISBP	739,534	10.2%
SNP (G1812/DV2138)	3,422,189	-







窈









Institute of Genetics and Developmental Biology, CAS

科学院透传与发育生物学研究

LETTER

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

Hong-Qing Ling¹*, Shancen Zhao^{2,3}*, Dongcheng Liu¹*, Junyi Wang^{1,2}*, Hua Sun¹*, Chi Zhang²*, Huajie Fan¹, Dong Li², Lingli Dong¹, Yong Tao², Chuan Gao², Huilan Wu¹, Yiwen Li¹, Yan Cui¹, Xiaosen Guo², Shusong Zheng¹, Biao Wang¹, Kang Yu¹, Qinsi Liang², Wenlong Yang¹, Xueyuan Lou¹, Jie Chen², Mingji Feng², Jianbo Jian², Xiaofei Zhang¹, Guangbin Luo¹, Ying Jiang², Junjie Liu², Zhaobao Wang², Yuhui Sha², Bairu Zhang¹, Huajun Wu⁴, Dingzhong Tang¹, Qianhua Shen¹, Pengya Xue¹, Shenhao Zou¹, Xiujie Wang⁴, Xin Liu¹, Famin Wang¹, Yanping Yang¹, Xueli An¹, Zhenying Dong¹, Kunpu Zhang¹, Xiangqi Zhang¹, Ming-Cheng Luo⁵, Jan Dvorak⁵, Yiping Tong¹, Jian Wang², Huanming Yang², Zhensheng Li¹, Daowen Wang¹, Aimin Zhang¹ & Jun Wang^{2,6,7}

Bread wheat (Triticum aestivum, AABBDD) is one of the most widely cultivated and consumed food crops in the world. However, the complex polyploid nature of its genome makes genetic and functional analyses extremely challenging. The A genome, as a basic genome of bread wheat and other polyploid wheats, for example, T. turgidum (AABB), T. timopheevii (AAGG) and T. zhukovskyi (AAGGA^mA^m), is central to wheat evolution, domestication and genetic improvement¹. The progenitor species of the Agenome is the diploid wild einkorn wheat T. urartu², which resembles cultivated wheat more extensively than do Aegilops speltoides (the ancestor of the B genome3) and Ae. tauschii (the donor of the D genome4), especially in the morphology and development of spike and seed. Here we present the generation, assembly and analysis of a whole-genome shotgun draft sequence of the T. urartu genome. We identified protein-coding gene models, performed genome structure analyses and assessed its utility for analysing agronomically important genes and for developing molecular markers. Our T. urartu genome assembly provides a diploid reference for analysis of polyploid wheat genomes and is a valuable resource for the genetic 200 base pairs (bp) to 65.8 kb, with an average length of 9.91 kb. The assembly was evaluated by comparisons with published bacterial artificial chromosome and expressed sequence tag (EST) sequences and by validation with PCR (Supplementary Information), and both indicated that the draft sequence had extensive genome coverage with high accuracy. The distribution of GC content in the *T. urartu* genome was comparable with those in the genomes of rice¹², maize¹³, sorghum¹⁴ and *Brachypodium distachyon*¹⁵ (Supplementary Information).

Genome annotation of the assembly was performed as described in Supplementary Information. About 66.88% of the *T. urartu* assembly was identified as repetitive elements, including long terminal repeat retrotransposons (49.07%), DNA transposons (9.77%) and unclassified elements (8.04%) (Supplementary Information). The proportion of repetitive DNA was lower than the roughly 80% previously reported¹⁶, which is probably due to a decreased incorporation of repeat sequence reads into the assemblies.

To facilitate gene prediction, we generated a 116.65-megabase (Mb) transcriptome of *T. urartu* with 67.14 Gb of RNA-Seq data from eight different tissues and treatments using the HiSequation (2000)

Nature 469: 87-90 (2013)



Physical Mapping and Superscaffold Construction



BAC Libraries and Physical Mapping

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 the physical map. The total size was 8x of *T*. *urartu* genome.



į.

きょぼ

1

与发育

家好

Physical Mapping Results

	LS-WGP
Total No. of BACs in FPC	345,233
Contig No.	12,137
No. and % BACs in contigs	323,058 (94%)
No. and % Singleton BACs	22,175 (6%)
Coverage (Mbp)	4,688
Average contig size (BACs)	26.6
Average contig size (kb)	386
N50 contig size (BACs)	55
N50 contig size (kb)	656



Evaluation of BAC contigs

(10 largest BAC contigs)

	Minimum tiling path	Sequencing & assembling	%
BAC Nos	154	152	99%
Total length of BACs (TLBS)	26,357,760	20,505,754	78%
Average length of BACs	171,154	134,906	79%
Nos of overlapped BACs	144	115	80%
Total length of overlapped sequence (TLOBS)	4,419,584	3,163,170	72%
Average length of overlapped sequences	30,692	27,506	90%
TLOBS/TLBS	17%	15%	92%

ね

育

羧

Ž



Institute of Genetics and Developmental Biology,

CAS

ž

Whole Genome Profiling (Keygene N.V.)

Sequence-based physical mapping BAC clones using Illumina Genome Analyzer II / HiSeq 2000





Integration of WGP and WGS Data







Institute of Genetics and Developmental Biology, CAS

7

物

資料



臣

例

ų,

Superscaffold Building Results

	Scaffolds	Superscaffolds
Number	133,687	13,899
Total Bases with Ns	3,597,407,679	4,975,825,774
Total Bases without Ns	2,864,495,294	2,531,764,367
Largest (nt)	1,066,088	19,663,984
Smallest (nt)	1,001	8,416
Average (nt)	26,909	357,999
N50 size (nt)	89.810	509,168
N50 index	11,635	2,850
N80 size (nt)	39.151	252,480
N80 index	29,490	6,994
N95 size (nt)	8,856	126,240
N95 index	54,258	11,036



- 1) Constructing a high resolution genetic map with SNPs
- Generated 500 F₂ plants by crossing of *T. urartu* G1812 with G3146
- Sequencing the 500 F2 plants to identify SNPs
- Constructing a high resolution genetic map
- 2) Anchoring the BAC contigs/scaffolds on the genetic map to finish the physical mapping of wheat A genome.
- 3) Sequencing BACs (70,000) to complete the wheat A genome.



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.
- We sequenced, assembled and annotated the wheat A genome using a whole-genome shotgun sequencing strategy. 34,789 protein-coding gene models has been predicted.
- We analyzed the genome structure, and assessed its utility for analyzing agronomically important genes and for developing molecular markers.
 - BAC contigs have been constructed using WGP technology. We integrated the WGS and WGP data and constructed superscaffolds. The average size of supperscaffolds was 358 kb, and its N50 size reached to 509 kb.
 - **T. urartu** genome assembly provides a diploid reference for analysis of polyploid wheat genome and is a valuable resource for the genetic improvement of wheat.



Acknowledgments:

Academician Zhensheng Li

Dr. Daowen Wang

Dr. Aimin Zhang

All peoples worked on this project

Collaboration: BGI-Shenzhen

UC Davis (Prof. Jan Dvorak, Dr. Mingcheng Luo)

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

Sequencing Center of IGDB (Dr. Chengzhi Liang)

Funding: Chinese Ministry of Science and Technology



P

学院送传与发育生物学研究所

Institute of Genetics and Developmental Biology, CAS

