A short overview on the latest updates on cereal crop plant genome sequencing with an emphasis on cereal crops and their wild relatives

Zaeema Khan[1]   **Hikmet Budak[1*]**

[1] Biological Sciences and Bioengineering Program, Faculty of Engineering and Natural Sciences, Sabanci University, Orhanli 34956, Istanbul, Turkey

[*] Corresponding author e-mail: budak@sabanciuniv.edu

ABSTRACT

The advent of next generation sequencing has brought a revolution in the sequencing and availability of whole genome data for numerous plant species. However the genome sequencing of major staple food crops has been noticeably obscure and till relatively recently majorly unaccomplished. The obstacles for sequencing of genomes of the Poaceae grasses including sugarcane and the Triticeae wheat, barley and rye has been largely ascribed to the complex polyploid nature of their genomes, having undergone numerous evolutionary changes duplications and additions resulting in their huge modern genomes of today. Undertaking their sequencing has been a daunting task however due to the sequencing of wild grass relatives such as *Brachypodium* and *Aegilops* has been an encouraging step providing an essential framework and reference for deciphering the complex genomes particularly Triticum aestivum. This paper discusses the major challenges involved, the approaches taken and the up to date accomplished tasks for sequencing a few of the major large grass crop genomes.

Keywords: next generation sequencing, whole genome sequencing, plant genomes, grass crop genomes, *triticeae*, polyploid genomes

Introduction

Since the introduction of next generation sequencing large and complex plant genome projects have been undertaken and their complex genomes sequences deciphered. Of considerable importance in today's world with a population expected to be greater than 9 billion by 2050 (Foley et al. 2011), cereal plants have attained special attention in having their genomes sequenced. It all began after the model plant *Arabidopsis* was sequenced in 2000 (The Arabidopsis Genome Initiative 2000), followed by one of the three major cereal plants harbouring the smallest genome, rice (*Oryza sativa*) (Yu et al. 2002). Due to the small genome size of rice its entire genome sequence was unravelled by BAC to BAC sequencing. However the large genome size and high repeat content of the other grasses posed obstacles in their genome sequencing.

Only relatively recently due to the advancement and introduction of next generation sequencing has the rush towards crop genome sequencing been re-kindled. Several years after the genome sequencing of rice, with the improvement in technology other larger cereal grass species including sorghum (Paterson et al. 2009), *Brachypodium* (Vogel et al. 2010), barley (Mayer et al. 2012) and maize (Schnable et al. 2009) were sequenced and continues till today as a race for sequencing the larger genome grasses. This was the start of a difficult and laborious journey towards the initial steps towards genome sequencing of all the major cereal crops of the world notably the most important being bread wheat (*Triticum aestivum*) (Brenchley et al. 2012; Mayer et al. 2014). It is however interesting to note that the second cereal grass to be sequenced maize was incompletely sequenced in 2009 by BAC to BAC approach and not

whole genome shotgun sequencing, due to its high repeat content (Feuillet et al. 2011). This is similar to the problem posed by the bread wheat genome. This report sheds light on the very latest advancements of plant genome sequencing, its applications and milestones reached. For convenience only the latest research in cereal crop plants and the progress on the grass sugarcane will be discussed here.

Crop genome sequencing and the impact of NGS

The ever increasing human population coupled with fluctuation in the global climate all pose threats to global annual crop yield and demand (Foley et al. 2011). Conventional molecular breeding techniques for crop improvements would alone prove insufficient for meeting the ever increasing demands of the world population. The breakthrough in efforts for increasing plant yield came with next generation sequencing. The advent of NGS and its use for sequencing plant genomes revolutionized the approach towards crop genetics and genomics. With sequencing multiple reads in parallel NGS changed the face of functional genomics with its massive amount of output data in the form of sequence reads (Pareek et al. 2011). With considerable reduction in cost, and the large scale of this technology plant food species were sequenced by the dozen (Bolger et al. 2014). Combining NGS with precise phenotyping techniques result in rapid and powerful tools for genetic identification of agriculturally significant traits and the prediction of the breeding value of plant individuals in a population (Varshney et al. 2014).

Whilst the genome sequencing of many non-cereal plant genomes underwent completion with the introduction of NGS, the main staple food crops remained hidden from mainstream sequencing efforts and initiatives. All three main food crops of the *Triticeae* namely, wheat and rye and until very recently barley, have not had their genomes readily sequenced and available for molecular breeding applications, contrary to the many non-plant species (Graph 1), (Bolger et al. 2014). As outstanding and popular as next generation sequencing has become in recent years undeniably due to its unique advantages and breakthrough technology, next generation sequencing platforms still have a long way to go before the final draft of the whole genome sequence of immensely essential staple crops such as bread wheat is completed. The second and third generation sequencers will have to undergo tremendous technological evolution similar to the way the cereal grasses underwent major evolutionary events to form into their giant present day genomes.

Barley (*Hordeum vulgare*) with a 5 GB genome was sequenced relatively recently in 2012 (Mayer et al. 2012). The diploidy of barley and three times smaller genome than *Triticum aestivum* are essentially contributing factors towards the availability of its genome sequence.

Rye (*Secale cereale*) a close relative of *Triticum aestivum* has an 8 Gb genome. Despite it also having a prominently vast genome, chromosome survey sequencing, high throughput transcript mapping alongwith exploiting the genome data of the sequenced grasses rice, sorghum and *Brachypodium,* resulted in a virtual linear gene order draft harbouring 31,008 rye genes. The application of sequenced grass genomes in syntenic analysis of huge plant genomes enables high-density genome wide comparative syntenic analysis. In rye this has enabled the identification of 17 conserved syntenic linkage blocks in both rye and barley and vivid dissimilarities in conserved syntenic gene content with an ancestral *Triticeae* genome (Martis et al. 2013).

Wheat with its gigantic allohexaploid genome consisting of 3 subgenomes A, B and D comprising a total of 17 Gb provides a huge obstacle in sequencing of its genome. Such a massive genome 5 times larger that of the human genome with an 80-90% repeat content (similar to rye and barley) is a daunting task. Still however efforts have been made to sequence the non-repetitive content of wheat to a 5X coverage. The sequence data of assembled Illumina reads of *Ae. tauschii* and *T. monococcum* were utilized for the gene assembly of the 5X coverage of *Triticum aestivum* cultivar Chinese Spring (Brenchley et al. 2012). Despite this, the sequencing and alignment of the uniform distribution of repetitive content in the wheat genome in long arrays and parallel copies is beyond the ability of next generation platforms and thus repetitive and intergenic remain un-assembled.

Accomplished projects of NGS

Despite the cumbersome genome of wheat and the shortcomings of current next generation sequencers in terms of sequencing large repetitive genomes, progress has been made in terms of reading the genomes and gene content of *Triticeae*. One important aspect here has been of the chromosome sorting with the isolation of individual purified chromosomes used in shotgun sequencing or in creating BAC libraries (Bolger et al. 2014). As aforementioned creating a reference genome sequence for *Tritcum aestivum* has been unrealised owing to the repetitive nature of its genome. With the availability of the EST and unigene and cDNA database for *Triticum aestivum* studies on microarray gene expression and targeted gene association have been

facilitated. The availability of the genome sequence of *Triticum urartu, Ae. tauschii* the progenitors of A and D wheat genomes, through high throughput sequencing also proved to be a hallmark in the progress on unravelling bread wheat genome. Through the relentless efforts of the International Wheat Genome Sequencing Consortium of which we are a small part a draft sequence of prepared of *T. aestivum* has been prepared approximately more than 95% of the genes of Chinese Spring cultivar of bread wheat. However an indepth detailed sequence of only one chromosome 3B is available. This draft sequence of wheat was prepared through sequencing of the individual flow-sorted chromosome arms. 124,201 gene loci have been annotated throughout the homeologous subgenomes. For survey sequencing each chromosome arm of the genome was sequenced with Illumina platform to a depth between 30X and 241X. These sequence assemblies cover roughly 61% of the genome in the form of survey sequences. The repetitive DNA comprised of 24 to 26% of the sequence reads and contained high copy number repeats. From the raw reads 81% and from the assembled sequences 76.6% contained repeats. Notably genome A contained more retroelements (Class I elements) and a pronounced abundance of LTR retrotransposons in comparison to genome B or D. From the protein coding genes a total of 44%, (55,249) were termed as high confidence from those assigned to the chromosome. (Mayer et al. 2014).

One of the recently accomplished resequencing of genomes has been of sorghum. Although initially sequenced in 2009, lately a high coverage resequencing of genomes of 44 lines of sorghum from diverse geographical origins has been presented, depicting the primary gene pool. The genome of *S. propinquum* was resequenced for the first time and 8M high quality SNPs were identified along with 1.9M indels indicating distinctive events of gene loss and gain. From the representation of the largest high-quality indel and SNP data for sorghum intricate domestication events were observed along with a large pool of diversity (Mace et al. 2013).

Similar to the resequencing of diverse racial accessions of sorghum, deep sequencing of 6 divergent lines of *Brachypodium distachyon* was undergone to analyse polymorphisms and gene expression. mRNA-Seq was performed under normal conditions and drought stress through which 300 genotype dependent genes were identified. A de novo transcriptome assembly was created with the most divergent line with the mRNA-Seq dataset. This remarkably resulted in more than 2400 previously unannotated transcripts along with hundreds of newly discovered gene absent in the reference genome (Gordon et al. 2014).

Though not a cereal, but a major food, grass crop and a relative of sorghum nevertheless sugarcane is also a complex genome crop whose genome is too complex for the whole genome shotgun approach. Sugarcane also harbours a largely repetitive and complex genome with a monoploid genome size of 930Mb. Interspecific crosses generating hybrid cultivars of sugarcane having complex polyploidy and aneuploidy produce genomes with great variation in their repetitive content and regions. Therefore gene enrichment using methyl filtration in order to enrich euchromatic regions was used for genome sequencing and assembly preparation. The availability of the sorghum genome sequence has facilitated the sequencing of sugarcane genome with conserved sequences having greater than 85% similarity between orthologs and the methyl filtered assembly obtained covered 98.4% of the sorghum coding sequences. This highly novel sequencing approach opens doors for sequencing of complex genomes with hypomethylated gene regions (Grativol et al. 2014). A complete list of the major food grasses with the approaches and accomplished milestones in genome sequencing is listed in Table 1.

Advantages and applications of plant genome sequencing

As mentioned earlier the whole genome sequence of crop species largely made possible due to NGS provide a not only a reference genome for unsequenced and/or large and complex grass genomes but also are reservoirs of genomic information to be manipulated for plant breeding strategies (Kurtoglu et al. 2014). The easy availability of relatively small noncomplex plant genome sequences by the progression in next generation sequencing has catapulted crop domestication studies, particularly in understanding the phenotype-genotype interaction. Genetic mapping of desirable traits has been facilitated by genotyoing by NGS through genome-wide SNP analysis. This has implications in GWAS studies, biparental crosses and intercrosses between parental lines of diverse origin. Genome resequencing can also identify genomic regions with low nucleotide diversity and linkage disequilibrium as genomic regions selected during domestication (Olsen and Wendel 2013).

Despite the limitations of next generation sequencing in sequencing the cumbersome wheat genome, the previous few years have witnessed a substantial increase in the amount of wheat genomic sequence data available publicly. Integrating whole genome sequencing and physical mapping will lead to a huge reservoir of wheat sequence data upon which a reliable reference genome sequence can be

Table 1. Chronological Order of Important Cereal Crop genomes sequenced (including the grass sugarcane).

Year	Grass	Significance	Sequencing Platform	Seq-Approach	Genome Size	Protein Coding Genes	Total Coverage
2002	Oryza sativassp japonica (Goff et al., 2002)	Long grain rice	MegaBACE capillary DNA sequencers	(random fragment) whole-genome shotgun sequencing	420 Mbp	14,345 high evidence	93%
2002	Oryzasativa ssp indica (Yu et al., 2002)	paternal cultivar of super-hybrid rice, Liang-You-Pei-Jiu (LYP9),short grain rice most widely cultivated in China	High throughput capillary machine MegaBACE 1000	Whole-genome Shotgun sequencing	466 Mb	53,398 prediction	92% functional coverage
2009	Sorghum bicolor (Paterson et al., 2009)	African grass related to maize and sugarcane	ABI 3730, Mega, Sanger	Whole genome sequencing, Sanger	700-772 Mb	27,640 bona fide protein-coding genes	13.5X clone, 8.50x sequence, 11x BAC library coverage
2009	Zea mays ssp mays (Schnable et al., 2009)	Maize genome B73	-----N/A----	BAC-by-BAC shotgun sequencing	2,300Mbp 2.3Gb	109,563 annotated loci	~38%
2010	Brachypodium distachyon (Vogel et al., 2010)	Small Foot, model organism for monocots, wild relative of wheat	Illumina GAIIx	Whole genome Shotgun, Deep sequencing	272.1 Mb	25,532 loci	9.43 X
2012	Hordeum vulgare (cv.) Morex (Klaus F X Mayer et al., 2012)	Barely, food crop	Illumina GAIIx	Whole genome Shotgun, RNA-seq	5.1 Gb	14,481 low-confidence genes	55.4 fold haploid genome coverage
2012	Triticum aestivum (Brenchley et al., 2012)	Bread wheat, main staple food of the world, hexaploid, landrace Chinese Spring	Roche 454 pyrosequencing/ Illumina	Whole genome sequencing, sanger	17 Gb	54,368 (~56%)	between 23X and 83X of non-repetitive region
2013	Aegilops tauschii (Jia et al., 2013)	Goat Grass Wild relative of T. aestivum, D genome progenitor, Accession AL8/78	Roche 454 (long reads) Illumina	Whole genome Shotgun RNA-seq Sanger sequencing	4.36 Gb	34,498	76X
2013	Oryza brachyantha	Wild species of Oryza genus	Illumina GA II platform	Whole genome Shotgun	297 Mb	32038	104 fold
2013	Triticum urartu	Wild wheat relative donor of genome	Illumina HiSequation (2000) platform	Whole genome Shotgun sequencing	4.94Gb	34,879 protein-coding gene models	94%
2013	Sorghum bicolor; Sorghum propinquum (Mace et al., 2013)	44 lines of African Sorghum allopatric Asian species	HiSeq 2000 Illumina platform	Whole genome resequencing	700-772 Mbp	19348	16–45
2014	Saccharum. spontaneum and S. officinarum (Grativol et al., 2014)	Wild sugarcane species	Illumina GAII machine, HiSeq2000 machin	genome sequencing by methylation filtration	930 Mb (one monoploid genome)	98.4% of sorghum protein sequences	134X
2014	Brachypodium distachyon (Gordon et al., 2014)	6 divergent lines	Illumina sequencing	Deep sequencing	272 Mb	33,626	92.6–96.8% of the reference genome
2014	Triticum aestivum (K. F. X. Mayer et al., 2014)	Chromosome based draft sequence of Chinese spring cultivar	Illumina sequencing	Individual chromosome arms	17Gb	124,201 gene loci	61% of genome sequence

drafted. Through the availability of wheat genomic data RNA-seq and exome capture have facilitated SNP identification and thus genome specific markers which can facilitate precise mapping of grain iron and zinc traits by marker assisted selection. This can result in availing all the genomic data resources in order to biofortify crops such as wheat with zinc and iron (Borrill et al. 2014).

sequences of wheat provide a clear picture of DNA markers and genes in the vicinity of these markers and thus creating more precision for molecular breeding. Nevertheless the complete high quality genome sequence is essential for pin pointing the precise gene loci of a trait. This would facilitate in creating considerably tolerant and superior crop varieties (Pennisi 2014).

Future prospects

This massive and continually increasing reservoir of plant genome sequence data is a huge step forwards in terms of speed and technology for plant breeders. They have become reliant on DNA marker assessment in seedlings for rapid elucidation of desired traits, rather than laboriously wasting time for a plant to mature. Although the progress is considerable in non-cereal plants and some cereals like rice, maize, *Brachypodium* and sorghum but still even the survey

Abbreviations

BAC	Bacterial Artificial Chromosome
EST	Expressed Sequence Tag
Gb	Giga basepair
GWAS	Genome Wide Association Studies
LTR	Long Terminal Repeat
MB	Mega basepair
mRNA-Seq	mRNA Sequencing
NGS	Next Generation Sequencing
SNP	Single Nucleotide Polymorphism

Graph 1.This graph depicts the progress over recent years with most of the progress in crop genome sequencing skewed between 2009-2014 (correlating with the progress in next generation sequencing). Note the size of wheat genome as compared to all the rest of the grasses. Identical grasses are depicted in the same colour.
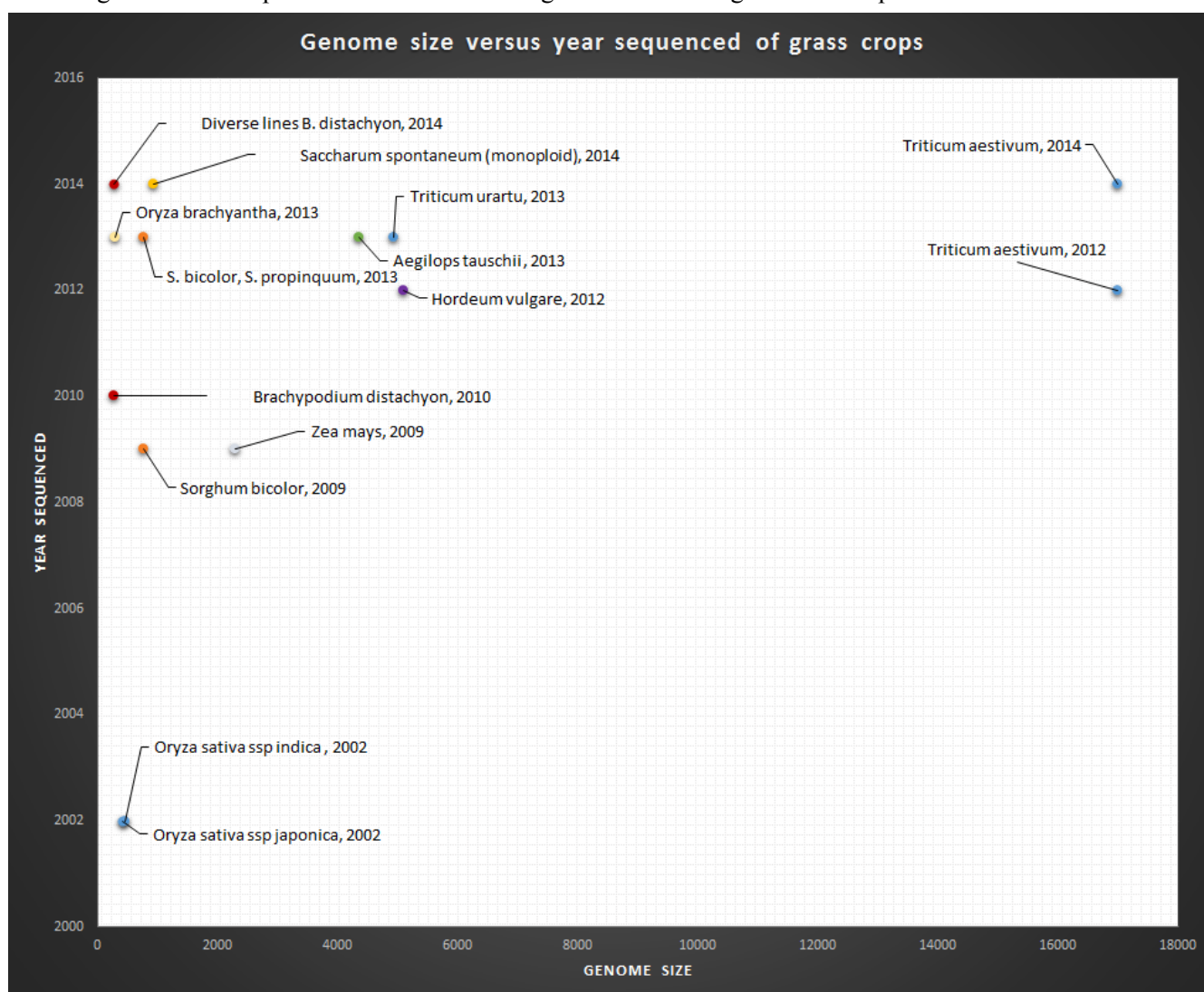
References

Bolger ME, Weisshaar B, Scholz U, Stein N, Usadel B, Mayer KFX (2014) Plant genome sequencing - applications for crop improvement. Curr Opin Biotechnol 26:31–7. doi: 10.1016/j.copbio.2013.08.019

Borrill P, Connorton JM, Balk J, Miller AJ, Sanders D, Uauy C (2014) Biofortification of wheat grain with iron and zinc: integrating novel genomic resources and knowledge from model crops. Front Plant Sci 5:53. doi: 10.3389/fpls.2014.00053

Brenchley R, Spannagl M, Pfeifer M, Barker GLA, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo M-C, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KFX, Edwards KJ, Bevan MW, Hall N (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705–10. doi: 10.1038/nature11650

Feuillet C, Leach JE, Rogers J, Schnable PS, Eversole K (2011) Crop genome sequencing: lessons and rationales. Trends Plant Sci 16:77–88. doi: 10.1016/j.tplants.2010.10.005

Foley JA, Ramankutty N, Bennett EM, Brauman KA, Carpenter SR, Cassidy E, Gerber J, Hill J, Johnston M, Monfreda C, Mueller ND, O'Connell C, Polasky S, Ray DK, Rockström J, Sheehan J, Siebert S, Tilman D, West PC, Zaks DPM (2011) Solutions for a cultivated planet: Addressing our global food production and environmental sustainability challenges. Nature 478:337–342.

Goff S a, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun W, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296:92–100. doi: 10.1126/science.1068275

Gordon SP, Priest H, Des Marais DL, Schackwitz W, Figueroa M, Martin J, Bragg JN, Tyler L, Lee C-R, Bryant D, Wang W, Messing J, Manzaneda AJ, Barry K, Garvin DF, Budak H, Tuna M, Mitchell-Olds T, Pfender WF, Juenger TE, Mockler TC, Vogel JP (2014) Genome diversity in Brachypodium distachyon: deep sequencing of highly diverse inbred lines. Plant J 79:361–74. doi: 10.1111/tpj.12569

Grativol C, Regulski M, Bertalan M, McCombie WR, Da Silva FR, Zerlotini Neto A, Vicentini R, Farinelli L, Hemerly AS, Martienssen RA, Ferreira PCG (2014) Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus Saccharum. Plant J 79:162–172. doi: 10.1111/tpj.12539

Jia J, Zhao S, Kong X, Li Y, Zhao G, He W (2013) Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation.

Kurtoglu KY, Kantar M, Budak H (2014) New wheat microRNA using whole-genome sequence. Funct Integr Genomics. doi: 10.1007/s10142-013-0357-9

Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, Campbell BC, Hu W, Innes DJ, Han X, Cruickshank A, Dai C, Frère C, Zhang H, Hunt CH, Wang X, Shatte T, Wang M, Su Z, Li J, Lin X, Godwin ID, Jordan DR, Wang J (2013) Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nat Commun 4:2320. doi: 10.1038/ncomms3320

Martis MM, Zhou R, Haseneyer G, Schmutzer T, Vrána J, Kubaláková M, König S, Kugler KG, Scholz U, Hackauf B, Korzun V, Schön C-C, Dolezel J, Bauer E, Mayer KFX, Stein N (2013) Reticulate evolution of the rye genome. Plant Cell 25:3685–98. doi: 10.1105/tpc.113.114553

Mayer KFX, Rogers J, Dole el J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski a. J, Sourdille P, Endo TR, Kubalakova M, Ihalikova J, Dubska Z, Vrana J, Perkova R, Imkova H, Febrer M, Clissold L, McLay K, Singh K, Chhuneja P, Singh NK, Khurana J, Akhunov E, Choulet F, Alberti A, Barbe V, Wincker P, Kanamori H, Kobayashi F, Itoh T, Matsumoto T, Sakai H, Tanaka T, Wu J, Ogihara Y, Handa H, Maclachlan PR, Sharpe A, Klassen D, Edwards D, Batley J, Olsen O -a., Sandve SR, Lien S, Steuernagel B, Wulff B, Caccamo M, Ayling S, Ramirez-Gonzalez RH, Clavijo BJ, Wright J, Pfeifer M, Spannagl M, Martis MM, Mascher M, Chapman J, Poland J a., Scholz U, Barry K, Waugh R, Rokhsar DS, Muehlbauer GJ, Stein N, Gundlach H, Zytnicki M, Jamilloux V, Quesneville H, Wicker T, Faccioli P, Colaiacovo M, Stanca a. M, Budak H, Cattivelli L, Glover N, Pingault L, Paux E, Sharma S, Appels R, Bellgard M, Chapman B,

bisab
bitki islahçıları alt birliği
www.bisab.org.tr

Nussbaumer T, Bader KC, Rimbert H, Wang S, Knox R, Kilian A, Alaux M, Alfama F, Couderc L, Guilhot N, Viseux C, Loaec M, Keller B, Praud S (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science (80-) 345:1251788–1251788. doi: 10.1126/science.1251788

Mayer KFX, Waugh R, Brown JWS, Schulman A, Langridge P, Platzer M, Fincher GB, Muehlbauer GJ, Sato K, Close TJ, Wise RP, Stein N (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–6. doi: 10.1038/nature11543

Olsen KM, Wendel JF (2013) Crop plants as models for understanding plant adaptation and diversification. Front Plant Sci 4:290. doi: 10.3389/fpls.2013.00290

Pareek C, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. J Appl Genet 52:413–35. doi: 10.1007/s13353-011-0057-x

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev I V, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556. doi: 10.1038/nature07723

Pennisi El (2014) Agriculture. Harvest of genome data for wheat growers. Science 345:251. doi: 10.1126/science.345.6194.251

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves T a, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh C-T, Emrich SJ, Jia Y, Kalyanaraman A, Hsia A-P, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia J-M, Deragon J-M, Estill JC, Fu Y, Jeddeloh J a, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen R a, Clifton SW, McCombie WR, Wing R a, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–5. doi: 10.1126/science.1178534

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796–815. doi: 10.1038/35048692

Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS Biol 12:e1001883. doi: 10.1371/journal.pbio.1001883

Vogel J, Garvin D, Mockler T (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463:763–8. doi: 10.1038/nature08747

Yu J, Hu S, Wang J, Wong GK-S, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296:79–92. doi: 10.1126/science.1068037