

Durum wheat evolution : a genomic analysis

Ben-Abu Y, Tzfadia O., Maoz Y., Kachanovski D.E., Melamed-Bessudo C., Feldman M., Levy A.A.

in

Porceddu E. (ed.), Damania A.B. (ed.), Qualset C.O. (ed.).
Proceedings of the International Symposium on Genetics and breeding of durum wheat

Bari : CIHEAM

Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 110

2014

pages 31-45

Article available on line / Article disponible en ligne à l'adresse :

<http://om.ciheam.org/article.php?IDPDF=00007056>

To cite this article / Pour citer cet article

Ben-Abu Y, Tzfadia O., Maoz Y., Kachanovski D.E., Melamed-Bessudo C., Feldman M., Levy A.A.
Durum wheat evolution : a genomic analysis. In : Porceddu E. (ed.), Damania A.B. (ed.), Qualset C.O. (ed.). *Proceedings of the International Symposium on Genetics and breeding of durum wheat.* Bari : CIHEAM, 2014. p. 31-45 (Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 110)



<http://www.ciheam.org/>
<http://om.ciheam.org/>

Durum wheat evolution-- a genomic analysis

Yuval Ben-Abu^{1,2}, Oren Tzfadia¹, Yael Maoz¹, David E. Kachanovsky¹,
Cathy Melamed-Bessudo¹, Moshè Feldman¹, Avraham A. Levy^{1,2}

¹ Dept. of Plant Sciences, The Weizmann Institute of Science, Rehovot, Israel

² Dept. of Exact sciences and projects unit, Sapir Academic College, D. N. Hof Ashkelon, Israel

Abstract. Durum wheat appears in the archaeological record, very sporadically, ~7000 years before present, but becomes the dominant tetraploid wheat in the Levant and in the Mediterranean basin ~2500 years ago. Here, we discuss the archeological insights on durum wheat evolution and we focus on the analysis of the genomic changes that are correlated with the process of domestication and evolution of modern durum by comparing four genetic groups: wild emmer, domestic emmer, durum landraces and modern durum varieties. Changes in gene expression and copy number variation of genes and transposons were analyzed in the genetic groups. Genes were clustered based on their pattern of change during Durum evolution, e.g. gradual increase, or decrease, or increase at the onset of domestication and plateauing later on. There were not many genes that changed >2 fold in copy number. However, interestingly, the copy number of transposons increased with domestication, possibly reflecting the genomic plasticity that was required for adaptation under cultivation. Extensive changes in gene expression were seen in developing grains. For example, there was an enrichment for certain functions: genes involved in vesicle trafficking in the endosperm showed a gradual increase in expression during durum evolution and genes related to germination and germination inhibition increased in expression in the embryo, in the more recent stages of durum evolution. The approach described here enables better understanding of the genetic events that shaped modern wheat and identifies genes that can be used for crop improvement.

Keywords. Durum wheat – Evolution – Genomics domestication.

L'évolution du blé dur -- une analyse génomique

Résumé. Le blé dur a fait son apparition dans les archives archéologiques, très sporadiquement, il y a environ 7000 ans, mais il est devenu le blé tétraploïde prépondérant au Levant et dans le bassin méditerranéen il y a environ 2500 ans. Nous allons parcourir ici les connaissances archéologiques sur l'évolution du blé dur et focaliser l'attention sur l'analyse des changements génomiques liés au processus de domestication et d'évolution du blé dur moderne, en comparant quatre groupes génétiques : amidonnier sauvage, amidonnier domestique, variétés locales et variétés modernes de blé dur. Les changements de l'expression génique et la variation du nombre de copies de gènes et de transposons ont été analysés dans les groupes génétiques. Les gènes ont été regroupés en fonction de leur profil de changement au cours de l'évolution du blé dur, par exemple, augmentation ou bien réduction progressive, ou augmentation au début de la domestication et plafonnement successif. On n'a pas trouvé beaucoup de gènes montrant une variation du nombre de copies >2 fois. Cependant, il est intéressant de noter que le nombre de copies de transposons a augmenté au fur et à mesure de la domestication, ce qui indiquerait une certaine plasticité génomique nécessaire pour l'adaptation au système de culture. Des changements importants dans l'expression des gènes ont été observés au niveau des grains en développement. Par exemple, le renforcement de certaines fonctions : l'expression des gènes impliqués dans le transport vésiculaire de l'endosperme a augmenté progressivement au fil de l'évolution du blé dur et l'expression des gènes liés à la germination et à l'inhibition de la germination a augmenté au niveau de l'embryon dans les stades les plus récents de l'évolution. L'approche décrite permet de mieux appréhender les événements génétiques qui ont façonné le blé moderne et d'identifier les gènes utilisables en vue de l'amélioration de la culture.

Mots-clés. Blé dur – Évolution – Génomique de la domestication.

I – Introduction

Durum wheat, *Triticum turgidum* ssp. *durum*, is a tetraploid species whose genome is genetically very close to that of its progenitor, wild emmer wheat, *Triticum turgidum* ssp. *dicoccoides* ($2n=4X=28$, Genome BBAA) (Feldman 2001). The F1 hybrid between these two species is fully fertile and cytological analysis show that in most accessions of emmer wheat, chromosomes show full pairing with durum wheat. The first step in durum evolution was the domestication of wild emmer wheat through the loss of fragility of the spike, namely of its disarticulation into spikelets, the basic dispersal units (Feldman 2001; Salamini *et al.* 2002). This step was probably a gradual process as suggested from both genetic and the archeological evidence. Indeed, 2-3 major loci and several modifiers, with additive effects are involved in the control of fragility (Chen *et al.* 1998; Levy and Feldman 1989a; Millet *et al.* 2013; Nalam *et al.* 2006; Watanabe *et al.* 2002). It might thus have taken time for the appearance and fixation of the mutations and the full loss of fragility. In addition, the archeological record shows that in ancient sites where agriculture was practiced, a mixture of fragile and non-fragile types were found, and it took 3-4 thousand years until the non-fragile spikes became prominent in farming units in emmer wheat (Kislev 1984) and in einkorn wheat (Tanno and Willcox 2006). One interpretation of this observation is that the first mutants in the fragility locus were still partially fragile until a second and additional modifier mutations appeared, and/or that wild and domesticated spikes were grown in parallel. The loss of fragility gave rise to the first known domesticated wheat, *Triticum turgidum* ssp. *dicoccum*, or emmer wheat, which is grown to this day, albeit on a small scale (De Vita *et al.* 2006). Its spike is not fragile, however, like its wild progenitor, it has only 2 kernels per spikelet which are tightly wrapped in stiff glumes and it is not free threshing. How did durum wheat evolve from *dicoccum*? Did it get its naked kernels directly, through a mutation in the genes that control glume stiffness (the Q factor and the *Tenacious Glumes* (TG) locus) or, was there another intermediate step? The appearance of naked kernels was the second most important step in durum domestication after non-brittle spikes. The archeological record shows a very sporadic appearance of durum-like wheat ~ 7000 years BP (before the present) in the near east and is only ~ 2500 yrs ago that durum becomes a major crop in the Mediterranean basin (Feldman 2001). The question ‘why *durum* was not cultivated before the Hellenistic period?’ remains puzzling. Perhaps it was susceptible to some disease. On the other hand, *Triticum turgidum* ssp. *parvicoccum*, a tetraploid wheat “fossil” species was relatively abundant in the archeological record starting already 9000 yrs BP, but disappeared ~ 2000 yrs BP. It had a compact spike and was free-threshing, suggesting that it probably already contained the Q and *tg* mutations prior to durum (Feldman and Kislev 2007). This raises the possibility that *durum* received these mutations from *parvicoccum*, rather than evolving them independently from emmer wheat (Fig. 1). Durum may thus have derived from hybridization between *parvicoccum* and *dicoccum* receiving the free-threshing trait from *parvicoccum* and the large grains from *dicoccum*. The large grain of durum was probably preferred to the small grains of *parvicoccum* that lead to the prominence of durum as a tetraploid wheat and to the extinction of *parvicoccum*. While the origin of the free-threshing trait of durum, whether directly from *dicoccum* or via *parvicoccum*, remains uncertain, the molecular evidence suggests that there was a bottleneck in the formation of durum and it became isolated from its Near-Eastern emmer wheat center of origin (Oliveira *et al.* 2012; Ozkan *et al.* 2011).

In addition to the above-mentioned classical domestication traits that were selected in the process of durum evolution, other domestication “syndrome” traits were selected that were advantageous to the farmer, such as plant erectness versus the wild grassy types, increased number of seeds per spikelet, and reduced seed dormancy (Feldman 2001). It is likely that many other traits were selected, including many QTLs, which are not easily visible to the eye, such as resistance to abiotic and biotic stresses, physiological parameters that contribute to yield (Peleg *et al.* 2009), as well as quality parameters (Levy and Feldman 1989c) and in recent decades, following the green revolution, adaptation to the new cultivation conditions including chemical fertilizers and mechanical harvest.

So far, with the notable exception of the Q locus, domestication genes identified in wheat have been characterized only through mapping. Two major genes that control spike fragility are *Brittle Rachis 2* and *Brittle Rachis 3* located on the short arms of chromosomes 3A and 3B, respectively (Nalam *et al.* 2006); in addition, another locus for spike brittleness was mapped to chromosome 2A (Peleg *et al.* 2011; Peng *et al.* 2003). The differences between studies mapping spike fragility suggest that there is diversity among wild accessions in the number and location of loci involved. Tenacious glumes, and Soft glumes are 2 independent loci that affect glume tenacity and spike threshability (Sood *et al.* 2009). Similarly, many domestication-related QTLs were mapped (Gegas *et al.* 2010; Peleg *et al.* 2009; Peng *et al.* 2003), but the underlying genes were not identified at the molecular level. The Q locus, located on chromosome 5A is the only one that was so far characterized at the molecular level. It is one of the most significant domestication loci as it controls spike compactness, glume tenacity and fragility. It encodes for the APETALA2-like transcription factor (Simons *et al.* 2006) and while the 5A homeoallele has the most significant contribution, other homeoalleles were also shown to be involved in the domestication traits (Zhang *et al.* 2011).

Among the traits that were affected by domestication are the storage proteins, in particular the high molecular weight (HMW) glutenins whose variability and amounts are higher in wild than in domesticated tetraploid wheat (Laido *et al.* 2013; Levy and Feldman 1988; 1989b). Recently, a *NAC* genes from emmer wheat that contributes to high protein percent, a trait that affects both the nutritive value and the processing of wheat and was lost during domestication, has been isolated (Uauy *et al.* 2006).

The identification of additional loci that control domestication-related traits will be facilitated by the new arsenal of genomic tools in wheat. Despite the complexity of the wheat genome, due to its polyploidy and to the large amount of transposons in its genome, there has been remarkable progress in the amount of datasets and tools for wheat genomics. To reduce complexity, several studies have chosen the strategy to sequence BAC libraries of single flow-sorted chromosomes. This led to a high-resolution map of chromosome 3B, the largest wheat chromosome (1Gb) (Paux *et al.* 2008) and more recently to a high density map of chromosome 1BL (Philippe *et al.* 2013) and of group 7 (Berkman *et al.* 2013). New mapping tools are available including a large number of SNPs spread across the genome that have been developed for sequence-based mapping for bread wheat (Saintenac *et al.* 2013) and more specifically for *durum* (van Poecke *et al.* 2013). In fact, SNP mapping in a broad collection of wheat landraces and modern varieties has indicated the genomic regions that underwent selection (selective sweep) during post-domestication wheat breeding (Cavanagh *et al.* 2013). Whole genome sequences are also available for the A (Ling *et al.* 2013) and D (Jia *et al.* 2013) genomes, however a good assembly of contigs is still missing. Recently, a major advance in *durum* transcriptome analysis was the development of tools for the discrimination of homeologues from the A and B genomes from expression sequence data such as RNA-Seq (Krasileva *et al.* 2013). Data sets from small RNAs are also becoming available (Kenan-Eichler *et al.* 2011; Yao and Sun 2012).

In order to identify the global genomic changes that occurred during wheat domestication, we performed a genomic analysis, using a microarray to measure gene copy number and expression patterns in ~ 40,000 genes of tetraploid wheat and ~ 400 transposable elements (TEs). The wheat lines represent a gradient of domestication including a collection of wild emmer wheat; of domesticated emmer wheat (*dicoccum*); of *durum* landraces; and of modern *durum* cultivars. Genes were sorted according to patterns of evolution, showing different modes of increase or decrease during *durum* wheat evolution.

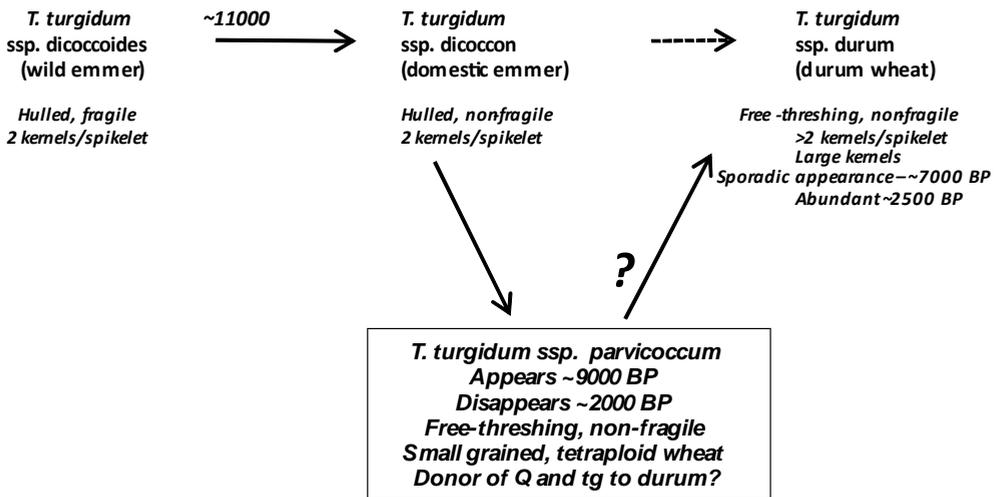


Figure 1. Evolution of durum wheat: The durum wheat wild progenitor, wild emmer wheat, *Triticum turgidum* ssp. *dicoccoides* was domesticated ~ 11,000 BP, giving rise to domestic emmer wheat, *Triticum turgidum* ssp. *dicoccon* through selection for non-brittle rachis. Naked kernels tetraploid wheat appears ~ 9000 BP in the near east. Its seeds and spike morphology are different from durum: the grains are small and it has relatively compact spike and short glumes. This sub-species, named *Triticum turgidum* ssp. *parvicoccum* disappears ~ 2000 BP. It might have a mutation in the Q factor and the *Tenacious Glume* genes, however, this cannot be confirmed as the species exists only in the archaeological record. The *parvicoccum* wheat might have contributed to the formation of durum wheat, providing it with the free-threshing trait.

Alternatively durum might have originated directly from emmer wheat through independent mutations in the Q and *Tg* genes.

II – Material and methods

1. Plant material

Thirty-six wheat lines that correspond to the various stages of domestication of the tetraploid level were grown and analyzed at the transcriptome and phenotypic levels. These wheat types consist of (a) the wild tetraploid varieties *T. turgidum* ssp. *dicoccoides* (11 lines), domesticated varieties including (b) primitive tetraploid wheat *T. turgidum* ssp. *dicoccon* (6 lines), (c) traditional tetraploid lines of *T. turgidum* ssp. *durum* (landraces) that were collected from traditional farmers, mostly from middle-eastern villages (7 lines), and (d) modern high yielding tetraploid macaroni wheat, *T. turgidum* ssp. *durum* varieties (5 lines). The plants were grown under the same conditions, in a net-house with 3 replicas per line, each replica being grown in a separate block. All plants were grown in 3-liter pots during the winter. The lines analyzed were of a broad range of eco-geographical origins (Table 1) in order to cover as much as possible of the variation typical of the subgroup analyzed.

Table 1. Tetraploid wheat lines used in this study (lab number and origin).

Wild emmer <i>Triticum turgidum</i> ssp. <i>dicoccoides</i>	Domestic emmer <i>Triticum turgidum</i> ssp. <i>dicoccum</i>	Durum Landraces <i>Triticum turgidum</i> ssp. <i>durum</i>	Durum modern <i>Triticum turgidum</i> ssp. <i>durum</i>
TTD12 (Ammiad.) Galilee, Israel	TTC1 (cv. Farrum, Italy)	TTR25 (cv. Ma'ari, Beit Sira Judean Mt., Israel)	TTR2 (cv. Hazera 163 Nursit, Israel)
TTD24, (Bet-Meir Judean Mts. Israel)	TTC4 (cv. Nigro-ajar)	TTR6 (cv. Beladi, Hevron Israel)	TTR19 (cv. Cappelli, Italy)
TTD28 (Northern Samaria, Israel)	TTC2 (cv. Khapli, India)	TTR265 (Mehola, Jordan Valley, Israel)	TTR1 (cv. camara, Portugal)
TTD31 (Lebanon)	TTC8 (cv. Submajus, India)	TTR333 (Turkey)	TTR298 (cv. Westbred 881, USA)
TTD32 (Turkey)	TTC6 (origin unknown)	TTR86 (Israel)	TTR16 (cv. Langdon, North Dakota, USA)
TTD35 (Turkey)	TTC7 (cv. macro antherium)	TTR60 (Israel)	
TTD 37 (Iran)		TTR42 (Israel)	
TTD48 (Shahabad-Illam Iran)			
TTD49 (Rosh Pinna-Zefat Rd. Easten Galilee, Israel)			
TTD64 (Diyarbakir, Turkey)			
TTD150 (Northern Iraq)			

2. RNA extraction and quality control

RNA was extracted from each replica from the developing seed (embryo or endosperm) (14 days after anthesis). Embryos were manually dissected from ~6 developing seed in each spike and the endosperm “milky” liquid was collected separately for each seed.

All tissue samples were immediately frozen in liquid nitrogen, and total RNA was extracted from 1.0 g of each pool tissue type using the Trizol® Plus RNA Purification Kit (Invitrogen, Carlsbad, CA) with an on-column DNase treatment.

Total RNA integrity was assessed using RNA 6000 Nano Lab Chip on the 2100 Bioanalyzer (Agilent, Palo Alto, CA) following the manufacturer’s protocol. Total RNA purity was assessed by the NanoDrop® ND-1000 UV-Vis Spectrophotometer (Nanodrop technologies, Rockland, USA). We considered RNA to be of good quality based on the 260/280 values (Nanodrop), rRNA 28S/18S ratios and RNA integrity number (RIN) (Bioanalyzer).

3. Labeling and microarray hybridization

For each block, an equal amount of RNA was pooled for each genetic group, namely we had 4 RNA samples, one for each of the Wild, Primitive (*dicoccum*), Land race durum, and modern durum lines. Microarray experiments were performed with 2 biological replicas (one series of 4 samples for each block).

The samples were labeled using Agilent Quick Amp Kit (Part number: 5190-0442). 500ng of total RNA was reverse transcribed using oligo-dT primer tagged to T7 promoter sequence. cDNA thus obtained was converted to double stranded cDNA in the same reaction. Further the cDNA was converted to cRNA in the in-vitro transcription step using T7 RNA polymerase enzyme and Cy3 dye was added into the reaction mix. During cRNA synthesis Cy3 dye was incorporated into the newly synthesized strands. cRNA obtained was cleaned up using Qiagen RNeasy columns (Qiagen, Cat No: 74106). Concentration and amount of dye incorporated were determined using Nanodrop. Samples that pass the QC for specific activity were taken for hybridization. 600 ng of labeled cRNA were hybridized on the custom Microarray Wheat 8x60K designed by Genotypic Technology Private Limited (AMADID: 037650) using the Gene Expression Hybridization kit (Part Number 5190-0404; Agilent) in Sure hybridization Chambers (Agilent) at 65° C for 16 hours. Hybridized slides were washed using Agilent Gene Expression wash buffers (Part No: 5188-5327). The hybridized, washed microarray slides were then scanned on a G2505C scanner (Agilent Technologies).

For copy number variation (Comparative genome hybridization - CGH) and gene (as well as transposons) expression profiling, we used a custom designed Agilent microarray chips of ~160,00 probes for the CGH (four for each EST, from which we choose the best probe in terms of quality for further analysis), and 60,000 probes for the gene expression analyses. The transposon fraction was assembled using data from the TREP database, which contains a collection of repetitive DNA sequences from different *Triticeae* species. The 10th version of this database, which was used here, contains a list of 477 sequences composed of DNA transposons, retrotransposons and other, non-classified repetitive sequences (<http://wheat.pw.usda.gov/ITMI/Repeats/>). Four Oligos were selected for each TE type from conserved regions that are representative of the TE family.

4. Microarray Feature Extraction and Data Analysis

Data extraction from Images was done using Feature Extraction software of Agilent V-10.7.3.1. Feature extracted data was analyzed using GeneSpring GX Version 11 software from Agilent. Normalization of the data was done in GeneSpring GX using the 75th percentile shift. Percentile shift normalization is a global normalization, where the locations of all the spot intensities in an array are adjusted. This normalization takes each column in an experiment independently, and computes the nth percentile of the expression values for this array, across all spots (where n has a range from 0-100 and n=75 is the median). Fold change expression values in test samples were obtained with respect to the specific control samples. Significant genes up and down regulated within the group of samples were identified. Statistical t-test was calculated based on volcano plot. For differential expression and clustering we used the EXPANDER and Cluster Identification via Connectivity Kernels (CLICK) algorithms (Sharan *et al.* 2003).

5. Calculation of copy number variation (CNV)

We first calculate the median signal (gMedianSignal) and background (gBGMedian) signal for each array from the raw data files along with probe names. We then averaged the background and subtracted gMedianSignal by gBGMedianSignal and then convert to log base 2 for all arrays. In each sample the log transformed intensity values for each probe is subtracted by the calculated 75th percentile value of the respective array and expression values are obtained like so:

$$R = (P/100) * (N+1)$$

Where: R = Rank; P = Percentile; N = Number of Entities (Rows).

III – Results and Discussion

1. Changes in gene expression of the developing kernel during durum evolution

To examine the genes that show the most significant change in gene expression in embryo or endosperm of developing kernels, during durum wheat evolution, we divided all the probes on the chip by their expression patterns in the different genetic groups (Wild, Primitive, Landrace and Modern) using the CLICK clustering solution (Sharan *et al.* 2003) and we selected genes which showed > 3 fold change. We chose to analyze significant clusters that have > 150 probes and used a stringent averaged homogeneity Pearson correlation coefficient r-value > 0.7. In Figure 2 we present such gene clusters for embryonic tissues dissected from two-week old seedlings.

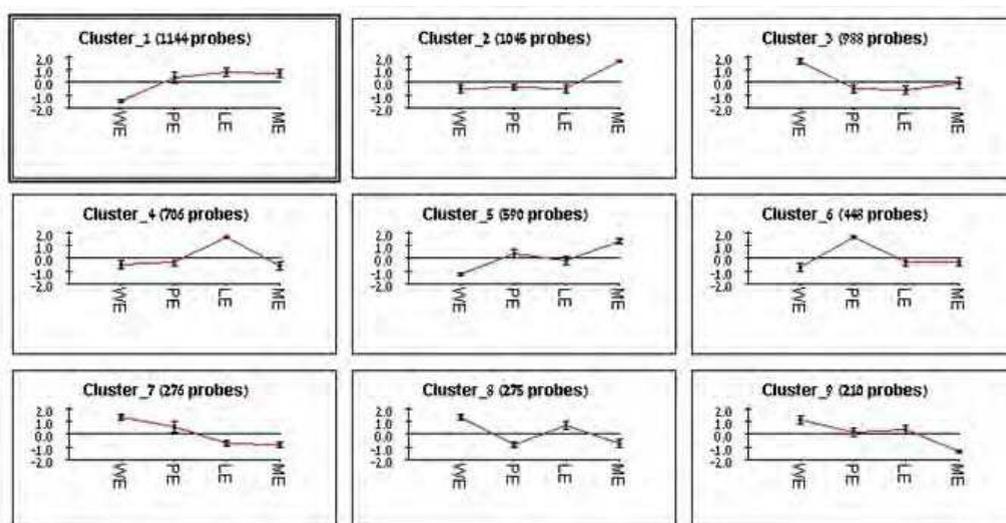


Figure 2. Clustering of probes corresponding to patterns of altered gene expression in the embryo during durum domestication. The X axis corresponds to the different stages of domestication, starting from the wild progenitor (WE= *T. turgidum* ssp. *dicoccoides*), the primitive non-fragile, non-free threshing tetraploid wheat (PE= *T. turgidum* ssp. *dicoccum* the *T. turgidum* ssp. *durum* landraces (LE) and modern durum cultivars (ME). The Y axis corresponds to the log value of fold change.

Within the embryonic tissue dataset, clusters 1, 2, 3, 5, 7 and 9 showed expression profiles in consistent with durum domestication and breeding. Cluster1 showed a gradual increase in expression during evolution with most changes occurring between the wild and primitive stage, namely at the onset of domestication. Cluster 2 showed most changes that occurred during recent amelioration (between landraces and modern varieties). These clusters were examined using BLASTx, Blast2GO (Conesa and Gotz 2008) and Ontologizer (Bauer *et al.* 2008) suites for GO enrichment. In this way, the biological processes could be examined on a larger scale for each cluster. Interestingly, we found that for gene expression in embryonic tissue, Cluster 2 has

a significant number of probes (p-value 0.001) associated with alpha-amylase inhibitor (Figure 2 and Table 2). This is interesting in light of its role in the regulation of seed dormancy, a trait that was counter-selected in domesticated wheat (Feldman 2001).

Alpha-amylase is an enzyme which aids in the breakdown of starch into maltose by hydrolyzing bonds between glucose molecules (Tanaka and Akazawa 1970). Regulation of alpha-amylase would allow for control over the availability of sugars in the embryo needed for germination (Garcia-Maya *et al.* 1990). As uniform germination would be a trait that would have been selected for during domestication, this becomes highly relevant. Other functions that were enriched in Cluster 2 are annotated as 'extracellular regions'. In this case the role for selection for enhanced expression in such genes is less clear but the analysis provides leads on potentially interesting genes.

Table 2. Biological processes, corresponding to Cluster2 pattern of evolution that were significantly enriched within embryonic tissues.

GO ID	Ontology Description	Total genes in GO category and annotated genes= 20772)	Genes in GO category (% total in Cluster2 and % total genes in cluster=539	Fold enrichment (p value)
GO:0019012	virion	30 (0.14%)	8 (1.4%)	10 (0.005)
GO:0005576	extracellular region	640 (3.08%)	32 (5.93%)	1.92 (0.002)
GO:0045735	nutrient reservoir activity	243 (1.17%)	75 (13.91%)	11.89 (0)
GO:0030234	enzyme regulator activity	398 (1.19%)	32 (5.94%)	4.99 (0.00004)
GO:0015066	alpha-amylase inhibitor activity	39 (0.19%)	11 (2.04%)	10.74 (0.001)
GO:0009405	pathogenesis	34 (0.16%)	6 (1.11%)	6.94 (0.018)

2. Changes in endosperm tissue gene expression during durum evolution

GO enrichment was also performed on clusters of genes consistent with patterns of changes relevant to durum evolution for endosperm tissue (Figure 3). Enrichment for cytoplasmic vesicle in cluster 5, (Table 3), a group of genes that showed gradual increase during durum evolution, is of interest since it might point to selection made for better starch and protein highways in the endosperm during domestication. Plant seeds accumulate starch in starch granules, providing sugars to the germinating embryo, and storage proteins which are a source of amino acids for use during germination are deposited into protein bodies (Takahashi *et al.* 2005). Both starch granules and protein bodies require vesicles for their packaging. In addition unique precursor-accumulating vesicles are known to mediate a transport pathway for insoluble aggregates of storage proteins directly to protein storage vacuoles (Hara-Nishimura *et al.* 1998). Better starch and protein trafficking in the developing seed could increase the ability to act as an efficient sink, which might be beneficial for both quality and yield.

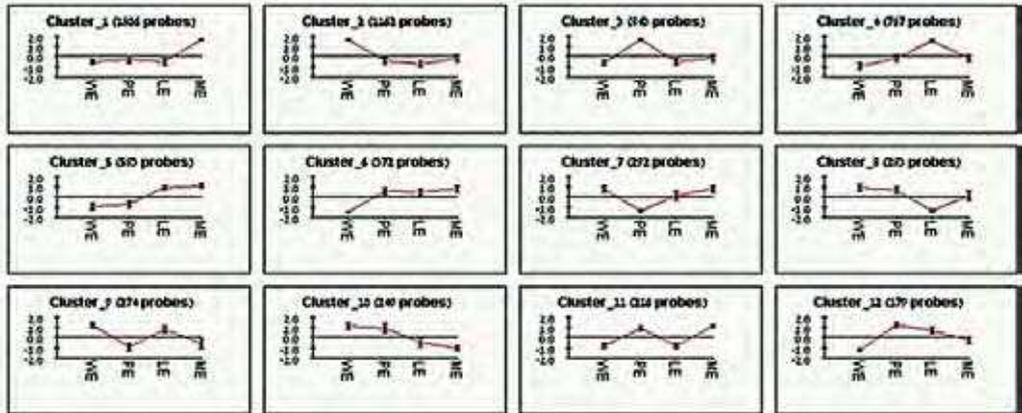


Figure 3. Clustering of probes in the endosperm, corresponding to patterns of altered gene expression during durum domestication. The X axis corresponds to the different stages of domestication, starting from the wild progenitor (WE= *T. turgidum* ssp. *dicoccoides*), the primitive non-fragile, non-free threshing tetraploid wheat (PE= *T. turgidum* ssp. *dicoccum* the *T. turgidum* ssp. *durum* landraces (LE) and modern durum cultivars (ME). The Y axis corresponds to the log value of fold change.

Within the embryonic tissue dataset, clusters 1, 2, 3, 5, 7 and 9 showed expression profiles in consistent with durum domestication and breeding. Cluster1 showed a gradual increase in expression during evolution with most changes occurring between the wild and primitive stage, namely at the onset of domestication. Cluster 2 showed most changes that occurred during recent amelioration (between landraces and modern varieties). These clusters were examined using BLASTx, Blast2GO (Conesa and Gotz 2008) and Ontologizer (Bauer *et al.*, 2008) suites for GO enrichment. In this way, the biological processes could be examined on a larger scale for each cluster. Interestingly, we found that for gene expression in embryonic tissue, Cluster 2 has a significant number of probes (p-value 0.001) associated with alpha-amylase inhibitor (Figure 2 and Table 2). This is interesting in light of its role in the regulation of seed dormancy, a trait that was counter-selected in domesticated wheat (Feldman, 2001).

Alpha-amylase is an enzyme which aids in the breakdown of starch into maltose by hydrolyzing bonds between glucose molecules (Tanaka and Akazawa 1970). Regulation of alpha-amylase would allow for control over the availability of sugars in the embryo needed for germination (Garcia-Maya *et al.* 1990). As uniform germination would be a trait that would have been selected for during domestication, this becomes highly relevant. Other functions that were enriched in Cluster 2 are annotated as ‘extracellular regions’. In this case the role for selection for enhanced expression in such genes is less clear but the analysis provides leads on potentially interesting genes.

3. Changes in copy number variation during durum evolution

The microarray was hybridized with genomic DNA to determine changes in copy number of genes or of TEs. With a few exceptions listed in the pie chart (Figure 5), there were no major changes in gene copy number during the various stages of durum evolution. Nevertheless, a cluster of 396 probes that exhibited > 2 fold change increase in copy number was detected between the wild emmer wheat and the modern durum cultivars. The pattern of copy number accumulation for these probes suggests a gradual accumulation of gene copies over the course of domestication (Figure 4).

Table 3. Biological processes, corresponding to Cluster5 pattern of evolution that were significantly enriched within endosperm tissues.

GO ID	Ontology Description	Total genes in GO category and (%total annotated genes =20772)	Genes in GO category in Cluster5 and (% total genes in cluster=240)	Fold enrichment and (p-value)
GO:0016023	cytoplasmic membrane bounded vesicle	4012 (19.3%)	99 (41.2%)	2.13 (0.0000)
GO:0065009	regulation of molecular function	555 (2.67%)	18 (7.5%)	2.81 (0.002)
GO:0050790	regulation of catalytic activity	549 (2.64%)	18 (7.5%)	2.84 (0.002)
GO:0044092	negative regulation of molecular function	338 (1.62%)	17 (7.08%)	4.37 (0.0005)
GO:0043086	negative regulation of catalytic activity	337 (1.62%)	17 (7.08%)	4.37 (0.0005)
GO:0030234	enzyme regulator activity	398 (1.92%)	18 (7.5%)	3.91 (0.0005)
GO:0004857	enzyme inhibitor activity	270 (1.3%)	17 (7.08%)	5.45 (0.0002)
GO:0016787	hydrolase activity	3517 (16.91%)	64 (26.67%)	1.58 (0.0003)
GO:0052689	carboxylic ester hydrolase activity	197 (0.95%)	14 (5.83%)	6.14 (0.0006)
GO:0030599	pectinesterase activity	71 (0.34%)	13 (5.42%)	15.94 (0.0003)

These genes cover all levels of plant cell functions and maintenance. There was no obvious enrichment for any particular function in the cluster of genes whose pattern of CNV is as shown in Figure 5, namely, a gradual increase. It seems that if there was selection for copy number increase of specific characteristics it was done through the modulation of broad cellular mechanisms complexes.

Among these genes we found 13 genes related to ubiquitin and E3 ligase members, which are part of the autophagy mechanism. Particularly interesting is the *Opaque-2* transcription factor which appeared in both the copy number variation data set and the differential gene expression data set. In Maize, the *Opaque-2* has been shown to be involved in the regulation of expression of major storage proteins and other important genes involved in seed development. It is a major regulator in the balancing of starch and protein in maize seeds (Zhang *et al.* 2012). As it is regulated in a phosphorylation/dephosphorylation manner it is likely to be closely involved with kinases and phosphatases (Guo *et al.* 2012) which also appear in our results.

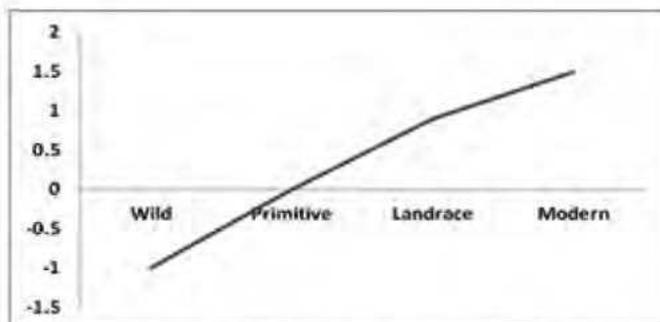


Figure 4. Cluster of 396 probes that increase > 2 fold in copy number during Durum wheat domestication. Samples of wild emmer wheat (Wild); domesticated emmer wheat (Primitive); of durum landraces (Landrace); and of modern durum cultivars (Modern) were tested using CGH on a custom designed microarray. Probes were clustered according to patterns of copy number variation (calculated as follows: the background was averaged and subtracted $gMedianSignal$ by $gMedianSignal - gBGMedianSignal$ and then converted to log base 2) This cluster shows a gradual increase in gene copy number over the course of tetraploid wheat domestication.

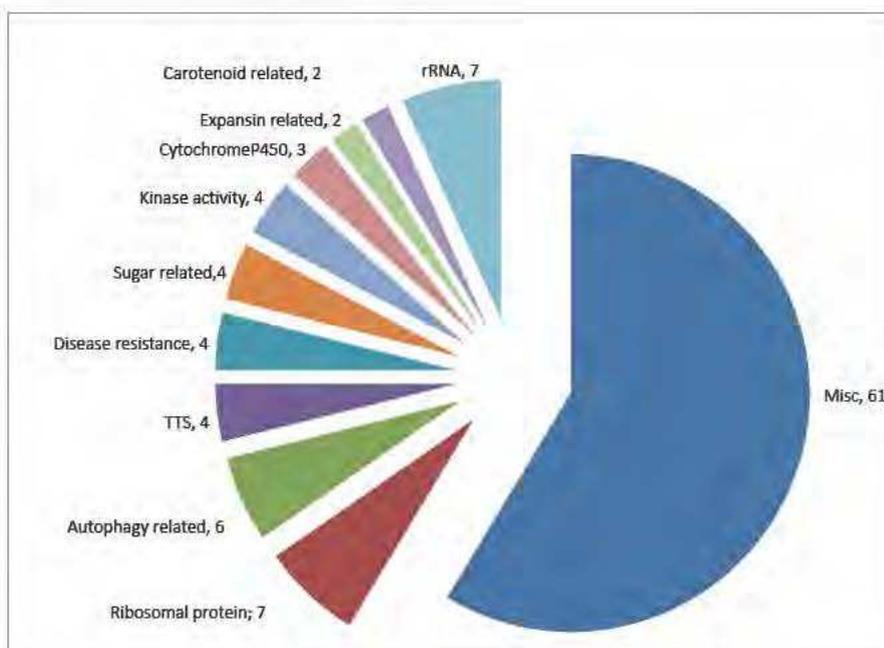


Figure 5. Representation of the functional annotation of genes found in the copy number module. The relative percent is given for each category.

When examining copy number variation on its own, we observed an over-representation of carotenoid biosynthesis related genes (carotenoid cleavage dioxygenase 1; CCD1, Chlorophyll synthase, viviparous-14; *vp14* and a light-harvesting complex I protein; *Lhca1*) (p-value 0.020). These five genes showed an increase in copy number between wild and modern tetraploid

wheat lines. Previous work has shown a link between domestication of maize and wheat and the accumulation of high levels of β -carotene in various tissues normally devoid of carotenoids (Rodriguez-Concepcion and Stange 2013). There are several reasons why the carotenoid pathway may have been selected for during domestication. Carotenoids are known to contribute to the stress response via ROS quenching and as a precursor to the plant hormone abscisic acid (Bradbury *et al.* 2012). They are also a supporting mechanism for chlorophyll biosynthesis and the photosynthetic pathway in general.

As with any broad data set, there are several classes/families of genes which one would expect to be present on the bases of statistics alone. Many classes of genes participate in diverse functional mechanisms through their many members/constituents. Therefore it was not surprising to see that within our dataset we identified 16 ribosomal genes, six Cytochrom P450 genes and several protein kinases with diverse functions (data not shown).

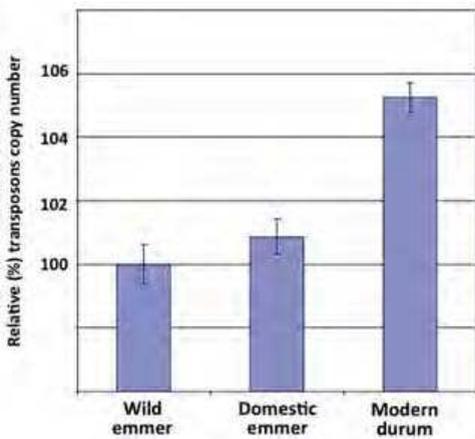


Figure 6. Relative change in transposable elements copy number during durum evolution. The copy number is expressed relatively to the wild emmer wheat. The results are for 477 sequences including DNA transposons and retrotransposons.

IV – Conclusions

While earlier studies on wheat domestication have mostly mapped the typical domestication syndrome genes (e.g. fragility and free-threshing traits), we have studied the genomic analysis of durum wheat evolution. We studied genome-wide changes in copy number and gene expression during the first stage of domestication, namely the transition from wild to domesticated emmer wheat, then the stages from *dicoccum* to *durum*, free-threshing landraces and modern varieties. For this purpose we have used a broad collection of lines (Table 1) from varied eco-geographical origins.

We have analyzed gene expression in developing kernels, in embryonic and endosperm tissues and we have classified genes according to different patterns of evolution. This analysis sheds light on genes whose expression was up regulated or down regulated at the various stages of evolution. For example, we could discover non-obvious targets of evolution, such as an enrichment for genes whose expression is related to vesicles and trafficking in the endosperm. It is tempting to speculate that the up-regulation we have observed was the result of human selection for types better adapted to agriculture conditions. There are many such examples, each of which requiring a deeper analysis to understand the functional significance of the observed changes.

Only a few genes showed some trends in copy number variation, but unlike for expression, these were few and overall rarely changed beyond 2 fold. Transposons seem to have increased in copy number. The increase was only of ~5% but considering that these correspond to a large fraction of the genome, this may have slightly affected genome size. More importantly, it is possible that the selection pressure of the new habitat of agriculture has served as a stress that activated copy number, or conversely, that only lines where transposons were active could provide the new mutations controlling the traits needed for evolution. Combining mapping data that should be soon available from whole genome sequences, together with a genomic analysis should point to new targets for further breeding of durum wheat.

References

- Bauer S., Grossmann S., Vingron M., Robinson P.N., 2008.** Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics*, 24, pp. 1650-1651.
- Berkman P.J., Visendi P., Lee H.C., Stiller J., Manoli S., Lrenc M.T., Lai K., Batley J., Fleury D., Simkova H., Kubalaková M., Weining S., Dolezel J., Edwards D., 2013.** Dispersion and domestication shaped the genome of bread wheat. *Plant Biotechnol J.*, 11, pp. 564-571.
- Bradbury L.M., Shumskaya M., Tzfadia O., Wu S.B., Kennelly E.J., Wurtzel E.T., 2012.** Lycopene cyclase paralog CruP protects against reactive oxygen species in oxygenic photosynthetic organisms. *Proc. Natl. Acad. Sci. USA*, 109: E1888-1897.
- Cavanagh C.R., Chao S., Wang S., Huang B.E., Stephen S., Kiani S., Forrest K., Saintenac C., Brown-Guedira G.L., Akhunova A., See D., Bai G., Pumphrey M., Tomar L., Wong D., Kong S., Reynolds M., da Silva M.L., Bockelman H., Talbert L., Anderson J.A., Dreisigacker S., Baenziger S., Carter A., Korzun V., Morrell P.L., Dubcovsky J., Morell M.K., Sorrells M.E., Hayden M.J., Akhunov E., 2013.** Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl. Acad. Sci. USA*, 110, pp. 8057-8062.
- Charles M., Belcram H., Just J., Huneau C., Viollet A., Couloux A., Segurens B., Carter M., Huteau V., Coriton O., Appels R., Samain S., Chalhou B., 2008.** Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. *Genetics*, 180, pp. 1071-1086.
- Chen Q.F., Yen C., Yang J.L., 1998.** Chromosome location of the gene for brittle rachis in the T betan weedrace of common wheat. *Genet. Resour. Crop Evol.*, 45, pp. 407-410.
- Conesa A., Gotz S., 2008.** Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics*, pp. 619-832.
- De Vita P., Riefolo C., Codianni P., Cattivelli L., Fares C., 2006.** Agronomic and qualitative traits of *T-turgidum* ssp. *dicoccum* genotypes cultivated in Italy. *Euphytica*, 150, pp. 195-205.
- Feldman M., 2001.** The origin of cultivated wheat. In: *The World Wheat Book*. Bonjean A., Angus W. (eds). Lavoisier Tech. & Doc., Paris, pp. 3-56.
- Feldman M., Kisslev M.E., 2007.** Domestication of emmer wheat and evolution of free-threshing tetraploid wheat. *Israel J. Plant Sci.*, 55, pp. 207-221.
- Garcia-Maya M., Chapman J.M., Black M., 1990.** Regulation of alpha-amylase formation and gene expression in the developing wheat embryo : Role of abscisic acid, the osmotic environment and gibberellin. *Planta*, 181, pp. 296-303.
- Gegas V.C., Nazari A., Griffiths S., Simmonds J., Fish L., Orford S., Sayers L., Doonan J.H., Snape J.W., 2010.** A genetic framework for grain size and shape variation in wheat. *Plant Cell*, 22, pp. 1046-1056.
- Guo X., Ronhovde K., Yuan L., Yao B., Soundararajan M.P., Elthon T., Zhang C., Holding D.R., 2012.** Pyrophosphate-dependent fructose-6-phosphate 1-phosphotransferase induction and attenuation of Hsp gene expression during endosperm modification in quality protein maize. *Plant Physiol.*, 158, pp. 917-929.
- Hara-Nishimura I., Shimada T., Hatano K., Takeuchi Y., Nishimura M., 1998.** Transport of storage proteins to protein storage vacuoles is mediated by large precursor-accumulating vesicles. *Plant Cell*, 10, pp. 825-836
- Jia J., Zhao S., Kong X., Li Y., Zhao G., He W., Appels R., Pfeifer M., Tao Y., Zhang X., Jing R., Zhang C., Ma Y., Gao L., Gao C., Spannagl M., Mayer K.F., Li D., Pan S., Zheng F., Hu Q., Xia X., Li J., Liang Q., Chen J., Wicker T., Gou C., Kuang H., He G., Luo Y., Keller B., Xia Q., Lu P., Wang J., Zou H., Zhang R., Xu J., Gao J., Middleton C., Quan Z., Liu G., Wang J., Yang H., Liu X., He Z., Mao L., Wang J., 2013.** *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*, 496, pp. 91-95.

- Kenan-Eichler M., Leshkowitz D., Tal L., Noor E., Melamed-Bessudo C., Feldman M., Levy A.A., 2011.** Wheat hybridization and polyploidization results in deregulation of small RNAs. *Genetics*, 188, pp. 263-272.
- Kislev M.E., 1984.** Emergence of wheat agriculture. *Paleorient*, 10, pp. 61-70.
- Krasileva K.V., Buffalo V., Bailey P., Pearce S., Ayling S., Tabbita F., Soria M., Wang S., Consortium I., Akhunov E., Uauy C., Dubcovsky J., 2013.** Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biol.*, 14, R66.
- Laido G., Mangini G., Taranto F., Gadaleta A., Blanco A., Cattivelli L., Marone D., Mastrangelo A.M., Papa R., De Vita P., 2013.** Genetic Diversity and Population Structure of Tetraploid Wheats (*Triticum turgidum* L.) Estimated by SSR, DaRT and Pedigree Data. *Plos One*, 8.
- Levy A.A., Feldman M., 1988.** Ecogeographical distribution of HMW glutenin alleles in populations of the wild tetraploid wheat *Triticum turgidum* var. *dicoccoides*. *Theor. Appl. Genet.*, 75, pp. 651-658.
- Levy A.A., Feldman M., 1989a.** Genetics of morphological traits in wild wheat, *Triticum turgidum* var. *dicoccoides*. *Euphytica*, 40, pp. 275-281.
- Levy A.A., Feldman M., 1989b.** Location of genes for high grain protein percentage and other quantitative traits in wild wheat *Triticum turgidum* var. *dicoccoides*. *Euphytica*, 41, pp. 113-122.
- Levy A.A., Feldman M., 1989c.** Location of Genes for High Grain Protein Percentage and Other Quantitative Traits in Wild Wheat *Triticum turgidum* var. *dicoccoides*. *Euphytica*, 41, pp. 113-122.
- Ling H.Q., Zhao S., Liu D., Wang J., Sun H., Zhang C., Fan H., Li D., Dong L., Tao Y., Gao C., Wu H., Li Y., Cui Y., Guo X., Zheng S., Wang B., Yu K., Liang Q., Yang W., Lou X., Chen J., Feng M., Jian J., Zhang X., Luo G., Jiang Y., Liu J., Wang Z., Sha Y., Zhang B., Wu H., Tang D., Shen Q., Xue P., Zou S., Wang X., Liu X., Wang F., Yang Y., An X., Dong Z., Zhang K., Zhang X., Luo M.C., Dvorak J., Tong Y., Wang J., Yang H., Li Z., Wang D., Zhang A., Wang J.M., 2013.** Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature*, 496, pp. 87-90.
- Millet E., Rong J.K., Qualset C.O., McGuire P.E., Bernard M., Sourdille P., Feldman M., 2013.** Production of chromosome-arm substitution lines of wild emmer in common wheat. *Euphytica*, 190, pp. 1-17.
- Nalam V.J., Vales M.I., Watson C.J., Kianian S.F., Riera-Lizarazu O., 2006.** Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (*Triticum turgidum* L.). *Theor. Appl. Genet.*, 112, pp. 373-381.
- Oliveira H.R., Campana M.G., Jones H., Hunt H.V., Leigh F., Redhouse D.I., Lister D.L., Jones M.K., 2012.** Tetraploid wheat landraces in the Mediterranean basin: taxonomy, evolution and genetic diversity. *Plos One*, 7: e37063.
- Ozkan H., Willcox G., Graner A., Salamini F., Kilian B., 2011.** Geographic distribution and domestication of wild emmer wheat (*Triticum dicoccoides*). *Genet. Resour. Crop Evol.*, 58, pp. 11-53.
- Paux E., Sourdille P., Salse J., Sainetnac C., Choulet F., Leroy P., Korol A., Michalak M., Kianian S., Spielmeier W., Lagudah E., Somers D., Kilian A., Alaux M., Vautrin S., Berges H., Eversole K., Appels R., Safar J., Simkova H., Dolezel J., Bernard M., Feuillet C., 2008.** A physical map of the 1-gigabase bread wheat chromosome 3B. *Science*, 322, pp. 101-104.
- Peleg Z., Fahima T., Korol A.B., Abbo S., Saranga Y., 2011.** Genetic analysis of wheat domestication and evolution under domestication. *J. Exp. Bot.*, 62, pp. 5051-5061.
- Peleg Z., Fahima T., Krugman T., Abbo S., Yakir D., Korol A.B., Saranga Y., 2009.** Genomic dissection of drought resistance in durum wheat x wild emmer wheat recombinant inbred line population. *Plant. Cell Environ.*, 32, pp. 758-779.
- Peng J., Ronin Y., Fahima T., Roder M.S., Li Y., Nevo E., Korol A., 2003.** Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat. *Proc. Natl. Acad. Sci. USA*, 100, pp. 2489-2494.
- Philippe R., Paux E., Bertin I., Sourdille P., Choulet F., Laugier C., Simkova H., Safar J., Bellec A., Vautrin S., Frenkel Z., Cattonaro F., Magni F., Scalabrin S., Martis M.M., Mayer K.F., Korol A., Berges H., Dolezel J., Feuillet C., 2013.** A high density physical map of chromosome 1BL supports evolutionary studies, map-based cloning and sequencing in wheat. *Genome Biol.*, 14, pp. R64.
- Rodriguez-Concepcion M., Stange C., 2013.** Biosynthesis of carotenoids in carrot: An underground story comes to light. *Archives of biochemistry and biophysics*, 539, pp. 110-116.
- Sainetnac C., Jiang D., Wang S., Akhunov E., 2013.** Sequence-based mapping of the polyploid wheat genome. *G3 (Bethesda)*, 3, pp. 1105-1114.
- Salamini F., Ozkan H., Brandolini A., Schafer-Pregl R., Martin W., 2002.** Genetics and geography of wild cereal domestication in the near east. *Nat. Rev. Genet.*, 3, pp. 429-441.
- Sharan R., Maron-Katz A., Shamir R., 2003.** CLICK and EXPANDER: a system for clustering and visualizing gene expression data. *Bioinformatics*, 19, pp. 1787-1799.
- Simons K.J., Fellers J.P., Trick H.N., Zhang Z., Tai Y.S., Gill B.S., Faris J.D., 2006.** Molecular characterization of the major wheat domestication gene Q. *Genetics*, 172, pp. 547-555.

- Sood S., Kuraparthi V., Bai G., Gill B.S., 2009.** The major threshability genes soft glume (sog) and tenacious glume (Tg) of diploid and polyploid wheat, trace their origin to independent mutations at non-orthologous loci. *Theor. Appl. Genet.*, 119, pp. 341-351.
- Takahashi H., Saito Y., Kitagawa T., Morita S., Masumura T., Tanaka K., 2005.** A novel vesicle derived directly from endoplasmic reticulum is involved in the transport of vacuolar storage proteins in rice endosperm. *Plant Cell Physiol.*, 46, pp. 245-249.
- Tanaka Y., Akazawa T., 1970.** Alpha-Amylase Isozymes in Gibberellic Acid-treated Barley Half-seeds. *Plant Physiol.*, 46, pp. 586-591.
- Tanno K., Willcox G., 2006.** How fast was wild wheat domesticated? *Science*, 311, pp. 1886.
- Uauy C., Distelfeld A., Fahima T., Blechl A., Dubcovsky J., 2006.** A NAC Gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science*, 314, pp. 1298-1301.
- van Poecke R.M., Maccaferri M., Tang J., Truong H.T., Janssen A., van Orsouw N.J., Salvi S., Sanguineti M.C., Tuberosa R., van der Vossen E.A., 2013.** Sequence-based SNP genotyping in durum wheat. *Plant Biotechnol. J.*, 11, pp. 809-817.
- Watanabe N., Sugiyama K., Yamagishi Y., Sakata Y., 2002.** Comparative telosomic mapping of homoeologous genes for brittle rachis in tetraploid and hexaploid wheats. *Hereditas*, 137, pp. 180-185.
- Yao Y., Sun Q., 2012.** Exploration of small non coding RNAs in wheat (*Triticum aestivum* L.). *Plant Mol. Biol.*, 80, pp. 67-73.
- Zhang N., Qiao Z., Liang Z., Mei B., Xu Z., Song R., 2012.** *Zea mays* Taxilin protein negatively regulates opaque-2 transcriptional activity by causing a change in its sub-cellular distribution. *Plos One*, 7: e43822.
- Zhang Z., Belcram H., Gornicki P., Charles M., Just J., Huneau C., Magdelenat G., Couloux A., Samain S., Gill B.S., Rasmussen J.B., Barbe V., Faris J.D., Chalhou B., 2011.** Duplication and partitioning in evolution and function of homoeologous Q loci governing domestication characters in polyploid wheat. *Proc. Natl. Acad. Sci. USA*, 108, pp. 18737-18742.