Broadening the genetic basis of durum wheat

Jacques David, Muriel Tavaud, Pierre Roumet, Marie-Helene Muller, Sylvain Santoni, Gautier Sarah, Yan Holtz, Vincent Ranwez, Morgane Ardisson, Gerard Poux, et al.

Title : Broadening the genetic basis of durum wheat

Jacques David[1]**,** Muriel Tavaud, Pierre Roumet, Marie-Hélene Muller, Sylvain Santoni, Gautier Sarah, Yan Holtz, Vincent Ranwez, Morgane Ardisson, Gérard Poux, Constance Vagne

Abstract

A large reservoir of genetic diversity is available in the *Triticum turgidum* sub-species. This diversity can be used in breeding programs but the method by which it could be introgressed in the elite durum pool is challenging once the objective is to cover the whole genome. *T. turgidum* accessions are usually classified in sub-species or taxa while evidence are accumulating that a more complex genetic structure can be revealed using no *a priori* approaches. Wild and cultivated emmers are the most diverse compartment and some cultivated emmers are strongly differentiated from naked wheats while other emmers are closer to the current durum. Recent and successive bottlenecks (in the XXth century) explain a large part of the structuration of the modern durum pool and the loss of molecular diversity. The use of *T. turgidum* germplasm in classical genealogical breeding may lead to some disappointments if a quick return is researched while some advances in breeding elite lines can be achieved with more recombination and selection. Evolutionary Pre-Breeding appears as a valuable alternative. Building and managing composite cross populations for a long period yields in innovative genitors, with an enriched allelic diversity and reduced long range linkage disequilibrium.

Keywords

Triticum turgidum, durum, pre breeding, structuration, NGS, linkage disequilibrium

Introduction

Durum wheat belongs to the family of the AABB tetraploid wheats (*Triticum turgidum* spp.), where others cultivated forms such as emmer, polonicum and many others taxa co-exists (Bozzini, 1988), Nesbitt and Samuel 1996). A lot of work has been carried out to describe and explain this current diversity (Özkan et al, 2002, 2011, Kilian et al 2009, Luo et al. 2007, Thuillet et al. 2005, Haudry et al 2007). All these domestic forms derive from a common wild ancestor, the wild emmer, *Triticum dicoccoides* (Kilian et al, 2009). Among the cultivated forms, emmer, *T. dicoccum* appears as the current remnant form of the first domesticated taxa developed by the first farmers 10 000 years ago in the Middle East (Zaharieva et al, 2010). It presents hulled grains and has lost the seed dispersal habit of its wild ancestor. Grains have also increased in size and their shape have dramatically evolved from a triangular to a more round section with a reduction in the grain length (Zaharieva et al, 2010). Dicoccum spread from its center of domestication and spread westward and eastward giving rise to number of other different hulled taxa. A more recent transition (7000 years ago) has lead to the fixation in new taxa of an innovating trait, the free threshing of the kernels (Feldman and Kislev, 2007). The free threshing wheat group, sometimes improperly designed as naked wheats is large and morphologically diversified and the current durum wheat is its most representative and farmed member (Bozzini 1988, Kilian et al 2009).

Evolutionary factors such as spread and selection for the adaptation to new environments and varied farming practices, rise and fixation of new mutations, gene flows, seeds exchanges among and between farmers communities and lately modern breeding for adapting cultivars to the intensification and specialization of the crop production had a profound impact on cultivated plant growth, development and physiology (Alonso-Blanco et al, 2009) agronomic performance. Those evolutionary factors also deeply affect the level of genetic diversity between and among each group (Spilane and Gepts, 2001 for a review, Buckler and Thornsberry 2002, Luo et al., 2007, Thuillet et al, 2005, Haudry et al, 2007, Laidò et al, 2013 for *T. turgidum*).

The complex genetic landscape of old and modern *T. turgidum* wheats resulting of these 120 centuries of evolution has been partly elucidated. The major split appears between hulled wheats that constitutes diversified genetic pool and free threshing wheats, that tare comparatively very poor

1 UMR AGAP, Montpellier Supagro, 2 place Viala, F-34060 Montpellier, France

in allelic diversity (Thuillet et al, 2005, Haudry et al, 2007). Diversity is usually seen as an essential resource for breeding (Tanksley and Mc Couch, 1997, Acosta-Gallegos et al 2007, Cooper, Spillane and Hodgkin, 2001, Spilane and Gepts, 2001), especially for new and unpredictable environments. , Exploiting this "lost" or "neglected" diversity of hulled wheats for breeding elite cultivars is thus very attractive (FAO, 1996). This diversity is classically screened to seek for new resistance, find adaptation to harsh environments or to detect capacities to uptake more resources (water, light and nutrients) from the environment (Tanskley and McCouch, 1997).

But the exploitation of this diversity is very challenging since the gene pools spanning these 12000 years of evolution have been diverging for growth habits, adaptation to new cropping practices and to very different environments, from the harsh and competing condition in the wild to the highly controlled and fertile conditions of a modern field. In maize, it has been suggested that 4% of genes experienced a selective episode from the wild form to the crop (Wright et al 2005). In durum the transition from wild to domesticates also involved numerous QTLs (Peleg et al, 2011) and many physiological pathways have been differentially tuned (Papa, 2013). Mixing alleles selected in different conditions may therefore results in some physiological incompatibilities at the whole genome level, since many traits are constrained by contradictory trade-offs, e.g., the classical apparent and strong negative correlation between productivity and protein content (Bogard et al. 2010).

The knowledge of the genetic structure of diversity of the *T. turgidum* compartment taken as a whole, the creating of a pre-breeding germplasm gathering the diversity of the wild and the primary domesticated relatives, and the use of new technologies for its exploration and valorization is challenging. In this lecture note, we will first briefly expose recent work on the genetic structure and diversity of the *T. turgidum* gene pool without any prior on the different taxa. Then we will sum up some results obtained during a classical breeding program, lead in collaboration with French private companies, that included wild accessions, old landraces. Eventually, we will illustrate how the concept of evolutionary breeding (Suneson, 1956, 1969, Brown et al, 1990, Wolfe et al 2005, 2008) can be extended to the pre-breeding on durum wheat by creating and monitoring composite cross population to broaden the genetic basis of durum wheat. Finally we will detail how high-througput sequencing technologies can be used to detect the allelic diversity introgressed in such composite populations. More generally, we are convinced that current breakthroughs in massive DNA sequencing and in massive genotyping relying on thousands of single nucleotide polymorphisms (SNPs) (Kilian and Graner, 2012) are preparing an avenue for the use of this, so far neglected diversity in pre breeding activities (Hajjar and Hodgkin, 1997).

Sampling diversity : links between genetic structure of T. turgidum and erosion of diversity

Many previous studies on the structure of T. turgidum genetic diversity relied on a priori assignation of the samples to the different taxa, based on discriminant morphological traits. This a priori classification was used to study the differentiation between taxa and to compare their levels of genetic diversity. But as the taxa discriminant traits may be based on very few major genes, they may not reflect the shared ancestry or the divergence within and between groups (i.e., common morphological traits may have arisen through different history), and the classification may conceal a very different genetic structure. Moreover, recent or ancient crosses may have altered the initial genetic structure, by introgressing new traits in the different taxa. We applied here a clustering method identifying groups of genetically related individuals without any a priori on the origin of those individuals (DAPC, discriminant analysis of principal components, Jombart, Devillard, & Balloux 2010). We then projected the individual taxon information on the groups obtained by the classification procedure.

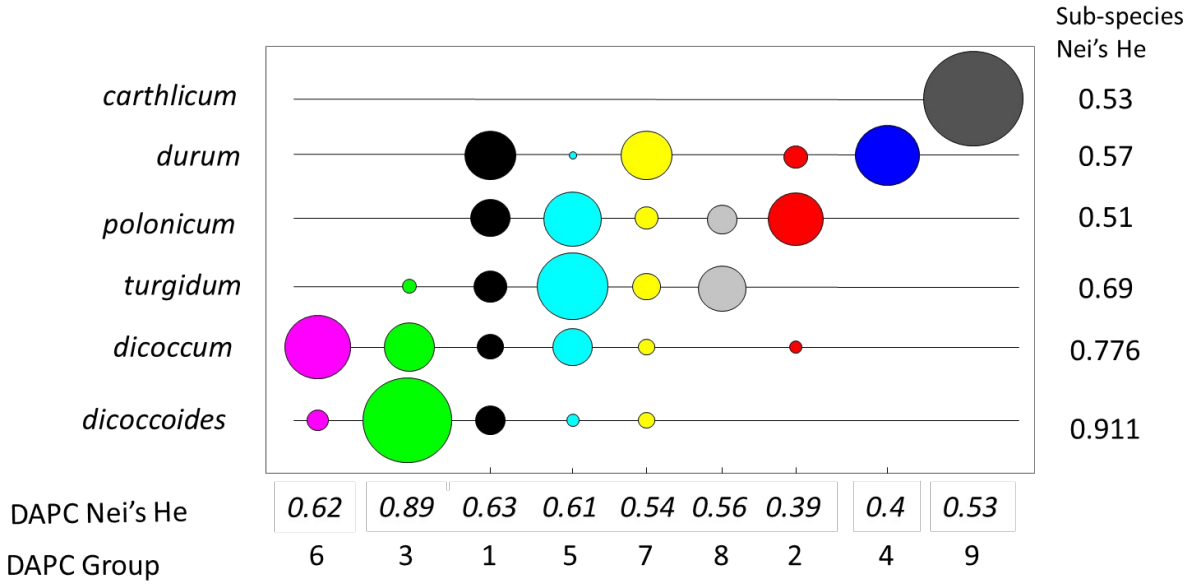
Our sample was made 492 individuals: 52 *T turgidum sp. dicoccoides (*DD), 52 *T. turgidum sp. dicoccum* (DC), 29 *T. turgidum sp. polonicum* (PO) , 33 *T. turgidum sp. turgidum* (TU) , 252 *T. turgidum sp. durum (*DR) covering traditional landraces and elite varieties mostly from the French catalog and 33 *T. turgidum sp. carthlicum* (CA) on which we firstly checked the ploidy level using flow cytometry. For these latter, we kept only 4X accessions since carthlicum accessions may count

(2n=4X=28 chromosomes, 4X) or (2n=6X=42 chromosomes, 6X) (see Thuillet et al. 2005 for details). Fourteen microsatellites locus (table 1) were used to genotype the whole sample on a capillary sequencer. The ADEGENET R package was used for the discriminant analysis of the groups (Jombart, 2008).

Locus	Chromosome location
Xgpw7577	1B
Xgwm312	2A
Xgwm257	2B
Xgwm374	2B
Xgwm413	2B
Xgwm2	3A
Xgwm285	3B
Xgwm601	4A
Xgwm495	4B
Xgwm234	5B
Xgwm193	6B
Xgpw2103	7A
Xgwm297	7B
Xgwm537	7B

Table : List and position of the 14 microsatellite locus used to genotype the 457 accessions.

Nine groups were detected using the procedures defined by Jombart et al. (2010), and the distribution of the different a priori taxa among groups is plotted figure 1. Taxa have been sorted according to their relative level of Nei's diversity . This suggests an historical interpretation.



Dicoccoides is mostly present in DAPC group 3, very few accessions being attributed to other groups for this sub-species. It may be seen as the basal group of the turgidum species with the highest level of diversity (He=0.89). A significant fraction of cultivated emmer accessions also belongs to this DAPC group 3, they could be considered as the closest cultivated emmer to the wild emmer and might be seen as the most ancient primitive source of domesticated wheats. The DAPC group 6 is mostly built on a portion of the sampled cultivated emmers and only a tiny portion of wild emmer accessions also belongs to this group. Note that some wild emmers in the DAPC group 6 could have been misclassified or be somewhat introgressed by domesticated emmers (Luo et al, 2007). No other sub-species contribute to this DAPC group 6 which appears then to be relatively disconnected of the rest of the cultivated turgidum sampled (graph not shown). This remarkable

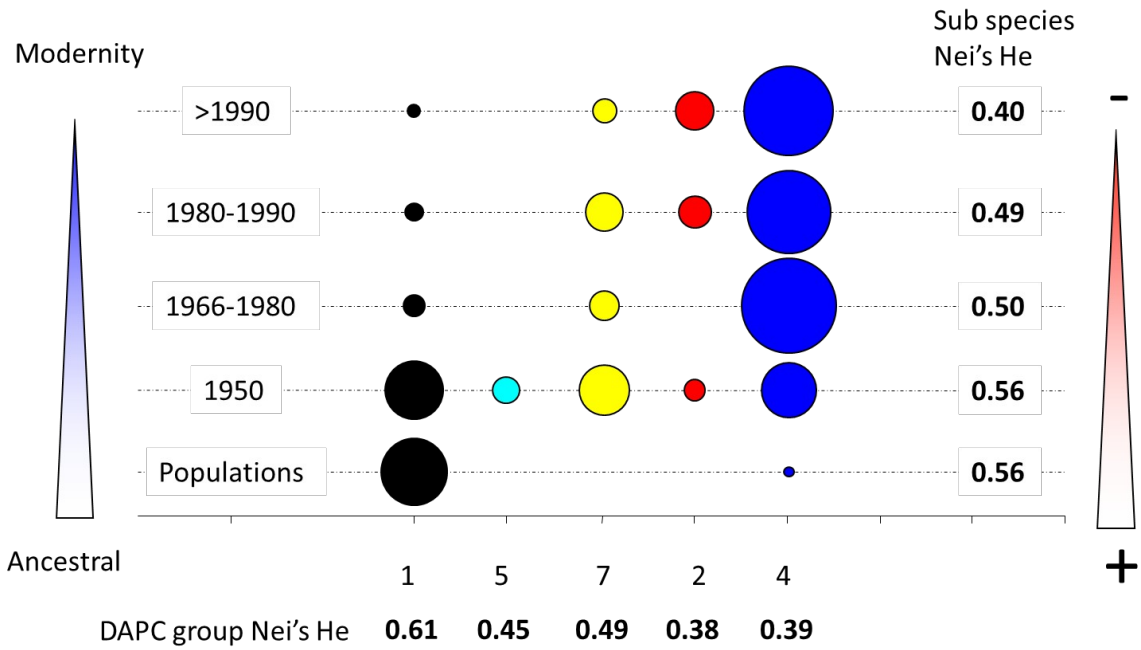
result suggests that this group did not participate to the emergence of the free-threshing wheat and remained isolated from the other cultivated sub-species. It has a relatively high level of diversity (He= 0.62). No obvious geographic localization could explain this structuration among cultivated emmers. Recent work support a polyphyletic origin of domesticated emmer from different sources of wild emmer (Civáň et al, 2013) and our results are somewhat congruent with this assumption. The groups 1, 5, 7 and 8 have a complex composition which underlines the difficulty to resolve the taxonomy of t. turgidum in terms of history, using molecular markers. Namely, ssp. polonicum, turgidum and durum do not correspond to clear and distinguishable genetic entities.). The group 1 appears relatively polymorphic (He=0.63) and spans all the sub species with the exception of carthlicum. This group may descent from the original domesticated genepool from which evolved the free threshing forms. Its complex structure is closed to that of the group 7 except that this latter has a reduced level of diversity. The group 5 collects accessions from dicoccum, and the largest part of the polonicum and turgidum accessions.

The DAPC groups 2 and 4 are made almost exclusively with durum accessions and show a strong reduction of diversity (He=0.4). All *carthlicum* are grouped in a specific goup, DAPC group 9, with a medium level of diversity.). These results are in strong agreement with the recent work of Laidò et al ; (2013). Our data does not permit to elucidate the origin of this group. *Triticum carthlicum* spikes resemble those of Triticum aestivum L. rather than those of free-threshing tetraploid wheats (Haque et al, 2011). The existence of 6X accessions in co existence with 4X accessions suggests that this sub-species has had a specific evolutionary pattern and it may result from recurrent intercrossing between 4X and 6X specific gene pools in Georgia, Armenia, Azerbaijan, northern Iraq, and Iran where it is still cultivated (Metakovski et al, 1989)

In brief, like other authors in recent works (Civáň et al. 2013, Laidò et al. 2013) we found that *T. turgidum spp* diversity should not be based only on keys determining their sub species status. Like in molecular phylogeny, morphological resemblance or difference may be or may be not linked to a common or divergent evolutionary history. More works should be dedicated to a fine analysis of the origin of the different emmers, their potential and respective implication in the formation of the naked wheats. The understanding of origin and evolution of carthlicum also deserves deeper and appropriate sampling. Indeed these wheats are a very valuable source of traits for durum breeding. They have resistance to drought, frost, and resistance to ergot infection.

Impact of modern breeding on Durum wheat structure and genetic diversity

Focusing on durum wheat, a more precise and recent pattern appears. We split the sample before and after 1950. After 1950, varieties were distributed by decades according to their registration to the French catalog. Their distribution between the different DAPC groups and their relative Nei's heterozygosity are plotted on figure 2. Landraces clearly belong to the DAPC group 1, the somewhat undifferentiated group described before. In 1950, two main groups 7 and 4 appear and a minor group (group 2) as well. These groups clearly experienced a strong reduction of diversity, the group 4 being the less diverse. A temporal evolution is also observed from 1950 to the post 1990's varieties. If the Nei heterozygosity was around 0.56 in landraces and in the 1950's varieties, it regularly decreased and is now as low as 0.4, less than half of that found in dicoccoides (group 3). Modern breeding for the transition to short stature but probably also more recent effort for developing varieties with high quality standards (e.g., selection on the gliadin profile) led to a strong and continuous reduction in genetic diversity. Selection in interaction with genetic drift, probably at the whole genome level (selective sweep like in bread wheat (Cavanagh et al., 2013)) is likely responsible of this dramatic reduction of genetic diversity in modern cultivars. This confirms previous results (Thuillet et al, 2005) and more recent work on durum (Laidò et al, 2013). This continuing erosion of genetic diversity is alarming.



M

Use in pedigree breeding

Diversity per se may not be an interest even the pledge for new alleles is attracting. Introgressing dicoccoides alleles in elite germplasm is usually realized after an identification of promising parents, a cross and successive backcrosses to eliminate undesirable chromosome fragments from the donor. This method of backcross has demonstrated a real efficiency to transfer monogenic traits, mostly resistance. In a larger view, broadening the genetic basis of a crop necessitates another, less targeted approach. Observing that the loss of diversity in modern durum wheat is really strong and assuming that wild and exotic germplasm can carry a number of valuable alleles for many traits, not easy to evaluate, or even not easy to identify at their sub species level, methods of non targeted introgressions, guided by the idea of a broad and non targeted introgression of new diversity have been proposed as a new pre-breeding challenge. With the help of a guilde of French durum wheat breeders (GIE Blé dur) we investigated the interests of the use of such germplasm.

During several years, more than 200 crosses have been realized between a core collection of tetraploid accessions and a set of elite genitors provided by the private partners. The core collection has been build from a 600 accessions sample by maximizing allelic richness on a set of 30 microsatellite locus used in diversity study (Thuillet et al. , 2005). F2 seeds were distributed in a multi-site network of public (INRA) and French private partners (DESPREZ, SERASEM-RAGT, EURODUR-LIMAGRAIN, SYNGENTA, BENOIST, GAE) , and a classical pedigree breeding method has been carried out to start the fixation of valuable lines that were generally used as genitors for backcrossing on durum elite lines. Several hundreds of thousands of individual F2 plants were evaluated by the partners. High throughput phenotyping for quality traits (protein content, yellow colour, semolina yield) were applied in the F4 selected families. Multi-site evaluation for frost and rust resistance were carried out on F5 families. The positive qualities of this material lied mostly in disease resistance (leaf rust, head blight, Mosaic virus (Wheat Spindle Streak Mosaic Virus, Soilborne Cereal Mosaic Virus) and a large morphological diversity. After the removal of the unfavorable undomesticated or primitive traits such as brittle rachis, hulled kernels and tall tillers, the main caveats of this material were defaults in the kernel size and color in crosses involving *T. dicoccoides* in their genealogy, a lack of productivity and possibly a unefficient remobilization of nitrogen from leaves to the kernel during the senescence period. *T. polonicum* appeared as a very good source of kernel colour and some accessions had good roots implantation. These primary and empirical observations justifies new studies about the impact of domestication

and recent breeding on the *durum* plant physiology, its N economy during its whole lifespan from uptake to remobilization. Domesticated and elite favorable alleles at some key locus, yet to be identified, may be important to explain a good end-use quality and productivity. Due to this lack of productivity, most of private breeders finally stopped this program since the agronomical level of the advanced bred lines was not sufficient to register elite varieties for the current fertilization and treatment practices. To keep up recombination and pre-breeding effort, INRA and Agri-Obtention went forward for more years with a policy of recombining several promising lines together as a priority instead of backcrossing recurrently on elite durum. The resulting most promising material increased in productivity, kept a good level of protein content and show a high level of leaf rust resistance in untreated, and in treated as well, experiments (fig 3). Two lines are currently (in 2013) following the French registering procedure and we hope they will finally be registered in 2014-2015. Their advantage seems to root in their good level of leaf rust resistance even under treated conditions, that may itself come from the introgression of new major genes of resistance to this disease. In this case, this advantage might only last some years until new virulent strains of leaf rust become adapted to these new sources if the new lines finally succeed in being cultivated on a a large area. This demonstrates that genetic advance for yield and quality relies on complex interactions between sanitary aspects, potential productivity in varying environments. These results clearly indicate that pursuing efforts in long term recombination and selection could permit a valorization of exotic germplasm for durum wheat breeding. Productivity and quality traits can be improved, either via better exploitation of the resources and adaptation to harsh environment but also by using new patterns of genetic resistance to main disease. In the present case, investigations are necessary to identify the genetic basis of this enhanced leaf rust resistance in order to assess their sustainability. If finally registered, these lines will constitute the demonstration that the erosion of genetic diversity in the modern elite pool of durum wheat can be stopped and that new diversity will be available for all breeders. Whole genome investigation will rapidly permit to estimate in which chromosomic regions new alleles are brought by these new lines.

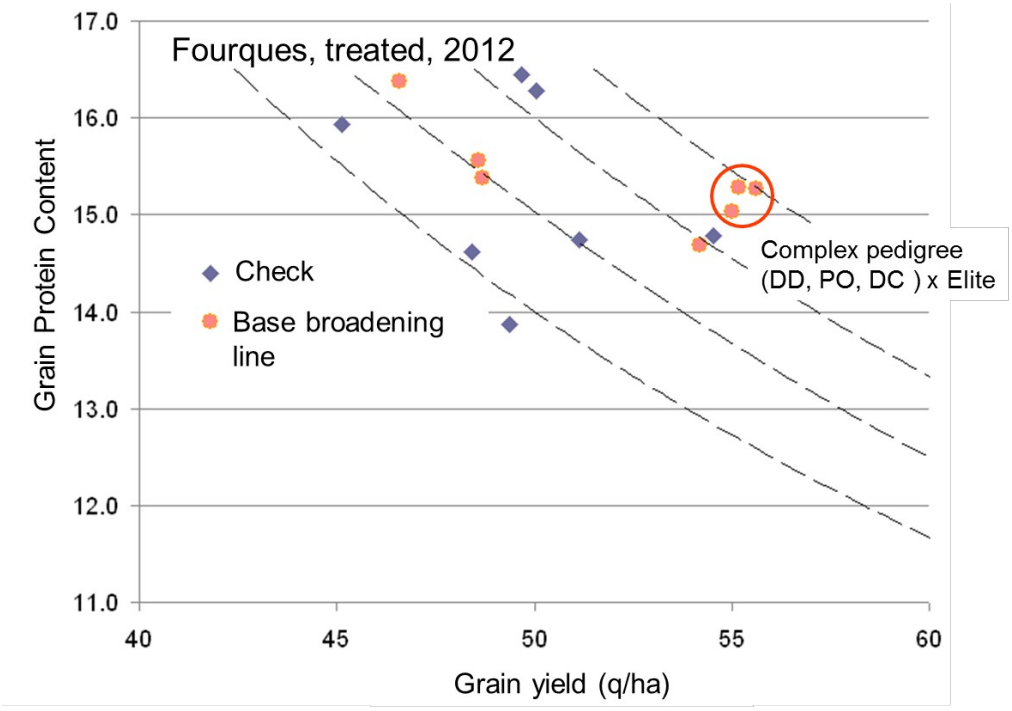


Figure 3. Comparison of agronomic performances of check lines (recent elite French durum cultivars) in blue versus lines derived from the pedigree base broadening program led by INRA Montpellier. Data are from Fourques in 2012 in a treated experiment. Dashed lines are the gradient of the grain yield x protein content product, i.e., the yield in protein/ha. The circled lines apply currently for a registration to the French durum wheat catalog.

Evolutionary Pre-breeding : presentation of the pre-breeding population of durum wheat

The term "pre-breeding" refers to the transfer of genes from related wild ancestors or from ancient varieties to breeding material (FAO, 1996). Pre-breeding activities spans a very large set of methods, from interspecific crosses followed by recurrent back crosses to the management on the long term of composite cross populations. In this latter case, recombination and soft selection are used to introgress exotic material in a elite gene pool. Barley composite cross, started by Suneson (1956, 1969) and whose evolution was described by Allard and many others brought information about the very dynamic evolutionary processes at work in such long term monitored composite cross. More recent work on bread wheat (Goldringer) confirmed that heterogeneous gene pools can adapt rapidly to different situations including climate gradients (Leboulch', David) , pathogen pressures (PAillard,) maintaining their genetic diversity (Enjalbert). The lessons drawn from such experiments are that natural selection leads to positive evolutions : adaptation to local condition (climate, disease)but competition between different architectural traits, such as plant height may drive the population on the fixation of bad alleles for productivity, e.g., in all populations semi-dwarf alleles disappeared and the harvest index evolved negatively, which is negatively correlated with the farmer's interest, except if he is interested by very long straw (leboucl'h) .

Creating and managing composite cross on the long range could be a really interesting pre-breeding method but methodological work should be devoted to understand the interplay between recombination, natural and human selection and genetic drift in order to set empirical and efficient rules for managing and improve such composites on the long range. Their role is to introgress massively interesting diversity at a whole genomic level. This needs that recombination and selection are finely tuned and like the CC cross of barley they have to be maintained for an indefinite number of generations and constitute an evolving reservoir of diversity (Henry et al. 1991). On durum wheat, our laboratory launched a pilot experiment ; we used a population of durum wheat donated by a former French INRA scientist, François Kaan, in which a nuclear male sterility gene segregates. The male fertile allele Ms is dominant on the male sterile allele. Plants can

be either hermaphrodites (Ms/Ms or ms/ms) or male sterile (ms/ms). A collection of flowering dicoccoides, dicoccum, polonium accessions were crossed in 1997 on male sterile plants of this population. The resulting seeds were used to found a pre-breeding composite cross, the INRA Pre breeding durum wheat population (hereafter named IPBDWP). Our aim was to combine recombination by promoting outcrossing and rapid fixation of favourable combination by permitting selfing. The population is thus monitored under a mixed mating system thanks to the male sterility gene. This population is being reproduced as follows : every year, once the flowering starts and until harvest, the tallest tillers are eliminated to avoid a detrimental evolution of the IPBDWP due to competition of tall plants on short plants (Leboulc'h et al,) , male sterile spikes are identified by their wide glume opening at the blooming stage and marked by a red twist. These marked spiked are harvested and threshed in bulk separately from the selected fertile spikes. Hermaphrodites spikes are chosen visually at harvest for their shape, vigor and health status and then threshed in bulk. The new generation is composed then by 20% of seeds coming from the marked male ms/ms sterile spikes (outcrossing portion), 70% of the selected hermaphrodite spikes Ms/ms and Ms/Ms (selfing portion) and 10% coming from the best lines selected in the pedigree selection scheme presented above to bring new diversity and agronomic performance. The population has this introgressed a new diversity and is experiencing recombination at each generation, fixation of new combinations under the combined effect of anthropic selection for a return to agronomical conditions, natural selection for adaptation to the environment and of course random genetic drift. The restricted amount of outcrossing (20%) reduced also the selective pressure to adapt to allogamy which can be the major evolutionary force in such population of usually selfing crops (David et al. 1993). The project is now to verify the interest of such resources for breeding, either as a source of new alleles or gene combination or as a tool for deciphering the genetic basis of traits.

Recently interest in genome wide association studies (GWAS) pointed out the worth of diversified panels to accurately detect chromosomal segments carrying valuable alleles for interesting agronomical traits (Ref durum also) . As most of these panels are assembled from large and diverse collections, genetic structuration among accessions may lead to a high level of false positive associations and even if several methods to take into account the level of structuration have been proposed, coping with genetic structuration remains a challenge (macafferi) . The interest of evolving composite populations in the GWA approach is that the population can be seen as a reproductive unit and after several generations of partial outcrossing and effective recombination, a reduction of the genetic structuration and a consequent reduction of the statistical linkage between locus, especially those that are not closely physically linked is expected. Consequently the False Discovery Rate (FDR*), i.e.,* the ratio with spurious association between a polymorphism and a variation of a trait should decrease substantially in a composite cross compared to a panel made of lines from different geographical areas, from different periods or different breeding programs. The other interesting aspect of GWA in a diversified panels compared to biparental segregating population comes from a more robust estimation of allelic effects.

In the following, we investigated for the first time the genetic diversity content of a composite cross of durum IPBDWP and estimated the extent of linkage disequilibrium along chromosomal segments to determine whether such populations might be a good support for GWAS studies.

First genomic investigations in the IPBDWP

Using New Generation Sequencing, information on thousands of candidate genes and candidate regions can be harnessed for thousands of individuals to sample genetic diversity within and between germplasm pools, to map Quantitative Trait Loci (QTLs), to identify individual genes and to determine their functional diversity (Kilian and Graner, 2012). Here we applied for the first time in durum wheat such an approach on our population.

Data production and SNP detection : In 2009, 500 spikes were randomly harvested in IPBDWP and entered a 2 year fixation process. Hundred and six (106) of these lines were used to investigate the level of genetic diversity available in this composite. Seeds were germinated in growth chamber in standardized conditions and young coleoptiles were used to extract RNAs. CDNAs libraries were

produced and tagged for each of these 106 genotypes. These 106 libraries were pooled, either by 24 samples or by 48 samples and sequenced on a HiSeq 2000 to produce 70 pb read pairs. Finally 813,110,268 cleaned reads were used to produce a *de novo* assembly using a bio-informatic pipeline (publication in prep). To separate homeologs between their A and B copies, we used an algorithm based on unbalanced expression ratio between the two copies implemented in the Homeosplitter software (Ranwez et al, 2013). The good split of copies were verified when possible by mapping reads sequences on *T. urartu and Ae. speltoides* transcriptomes produced and assembled by the same protocols. Finally, only good quality SNPs with no heterozygous excess were used in this preliminary study to evaluate the level of diversity and the extent of the linkage disequilibrium in IPBDWP.

Nucleotide diversity (π) was estimated as proposed by Tajima (Tajima 1989). To estimate the decay of linkage disequilibrium, it was first necessary to obtain the position of the SNP on a reference genetic map. In this preliminary work, no segregation data were available for these SNPs in durum wheat and we used external and public data from bread wheat. Contigs containing SNPs were blasted against the sequence of the 9K SNP array defined and mapped on bread wheat polymorphism using several segregating populations (Cavanagh et al , 2013). To eliminate possible errors and to ensure appropriate genome localization, we kept only SNPs for which the genome localization was identified identically by mapping on bread wheat (Cavanagh et al, 2013) and properly assigned to a donor diploid species in our data (*T. urartu* for the A ge,ome , *Ae. speltoides* for the S genome). Pairwise linkage disequilibrium was then computed and plotted against genetical distance between locus, distance estimated from the bread wheat data (Cavanagh et al, 2013)

Diversity and linkage

Finally 13 911 SNPs on 5980 locus fulfilled the conditions to be kept for the study, *i.e*., no excess of heterozygous individuals. The nucleotide diversity π was computed for these 13, 911 SNP and the obtained values vary between $\pi \sim 0.5$ to $0.9\ 10^{-3}$ per base pair. From previous evaluation on 21 genes (Haudry et al, 2007) , estimations for the wild dicoccoides is $\pi \sim 2.5\ 10^{-3}$, $\pi \sim 1.3\ 10^{-3}$ for dicoccum and $\pi \sim 0.4\ 10^{-3}$ for all durum. It seems thus that if the population has effectively a good level of diversity compared to the whole durum sub-species it is still far from what it could have been if a large part of diversity from wild and cultivated emmer had been successfully introgressed in IPBDWP.

Out of the 5980 contigs , 553 blasted on the sequences of 9k bread wheat SNP arrays, giving a total of 1858 SNPs for studying the linkage disequilibrium decay. Discarding ambiguous SNPs, attributed to different genomes by the bread wheat mapping and by the proximity to one of the diploid ancestor, 1577 SNPs could be eventually used to evaluate the decay of the linkage disequilibrium in IPBDWP. Figure 4 illustrates this decay on the chromosome 1A, the other chromosomes showing very similar patterns. As expected, the disequilibrium values between pairs of SNPs located in the same contig have the highest value, but their average value is far from the maximum value of 1, which would have mean complete linkage within genes and a low haplotype complexity. In this presentation paper, this apparent lack of linkage has not been fully investigated but it could mean that introgression of wild and exotic accessions has effectively enriched the haplotype diversity at very short genetic distance. Nevertheless, if some *de novo* assembled contigs are still chimeric between the A and B genome, low spurious r^2 values between some pairs of homeologous SNPs could decrease artificially the within contig linkage estimation. Between different contigs the decay of the linkage idsequilibrium is decreasing very fast and is lower than the value found by Maccaferi et al 2005 using microsatellites. A threshold value for r^2 around 0.1 is found after 70 cM very close to the value (dashed line) found for SNPs located on different chromosomes. Naturally, deeper investigations are needed to ascertain these linkage disequilibrium patterns but this preliminary data suggest that evolving composite cross such as IPBWDWP could have very good and interesting properties for detecting markers closely linked to causal polymorphim. They could constitute then very good alternative to association panels.

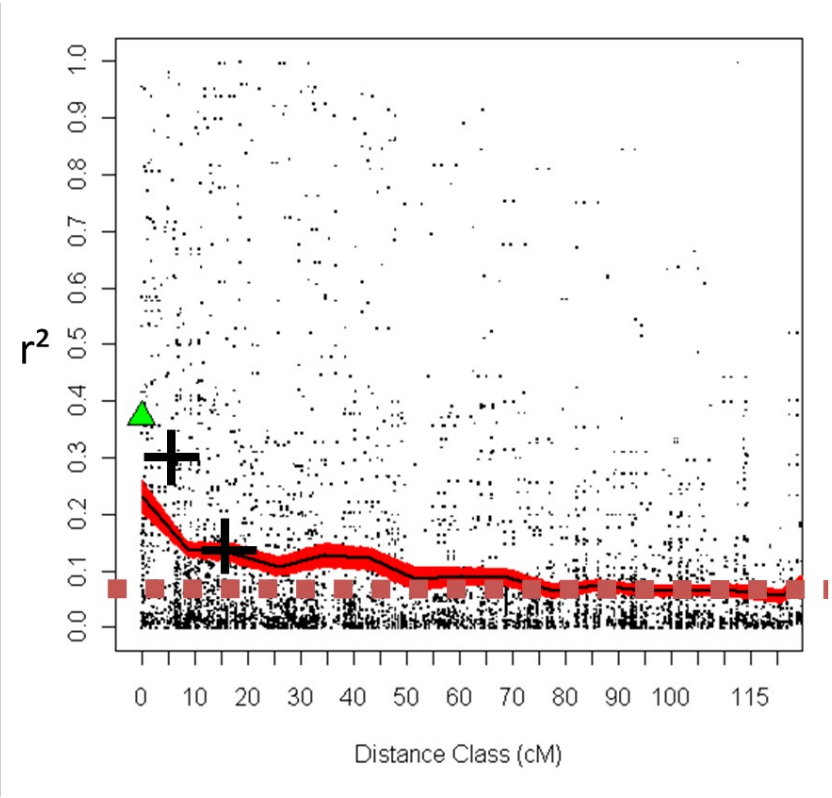


Figure 3. Evolution of linkage disequilibrium (r^2) between pair of SNPs located at different distances (cM) on the chromosome 1A in the INRA pre breeding Durum wheat population (IPBDWP). Mapping positions of the SNPs on the chromosome 1A were predicted by blasting contig sequence containing the SNPs on the sequence of the mapped markers of the 9K bread wheat micro array (Cavanagh et al. 2013). The green triangle is the average value of r^2 when the two SNP of a pair are located in the same contig. Black crosses are the values of r^2 estimated by Maccaferi et al, 2005 at similar distance classes in a durum wheat panel. In red, the within segmental average value of r^2 for 20 subsequent windows of equal genetic distance spanning the whole chromosome. The dotted brown line is the average value of pairwise r^2 when the two SNPs were assigned to different chromosomes.

Conclusions & Perspectives

The modern elite pool of durum wheat has experienced several severe reduction of genetic diversity, and there is evidence that this genetic erosion is still continuing. The recent cultivars share a lot of common alleles and their deviation from the historical genetic background of the species seems accelerating at least until the end of the XXth century and for the French elite catalog. The same trend has been observed in the sample of durum recently investigated by Laido et al. (2013). If a lack of diversity in the elite pool is susceptible to impair future advances in the development of a sustainable durum wheat production, for which disease resistances, effective nutrient uptake capacities, growth and flowering in harsher environmental conditions will be needed, the use of the genetic diversity of the whole *T. turgidum* species may be a key element of a germplasm development and integration strategy. But this will be a real challenge. If many studies demonstrated the worth of the genetic diversity hold in genebank in the whole *Turgidum subsp*, especially in the wild and cultivated emmers, the use of this valuable diversity is not easy and may not be really successful if one excepts the use of valuable alleles at some major genes, such as disease resistance. In a collobarative program between INRA and GIE Blé dur, classical breeding led to some results by using intensive back crossing after the initial cross but the selected lines were not sufficiently productive in the first place to be registered as elite varieties in the French catalog by the Private Breeders. Nevertheless, some success was obtained by persevering in recurrently crossing advanced lines with introgressed backgrounds. Productivity eventually increased and some

lines might become registered varieties in a close future, probably thanks to their good level of resistance to brown rust. This tolerance to rust probably provided a yield advantage to these advanced lines in an experiment where rust attack was important. This success should be confirmed on the long range since a quick overcome of the allelic of resistance is likely in the case of their commercial development

As an alternative to this quick use of valuable germplasm, long term evolutionary pre-breeding programs may be of a great interest for creating new germplasm, integrating new alleles, promoting recombination and soft selection in populations of reasonable population effective size. In this paper, we reported the very first results on a composite cross population of a durum wheat population with a broaden genetic basis monitored for 12 years under a 20% outcrossing mating system. This current IPBDW population appears as an interesting resource for GWAS because of its reasonable level of genetic diversity, reduced long distance linkage disequilibrium and large phenotypic variation (data not shown). We are currently accumulating phenotypic data on a large number of traits (morphology, phenology, N status of leaves and grains) to verify if the sequencing effort yielded sufficient data to detect associations. RNAseq data obtained here will be directly be used as a genotyping method (Genotyping by sequencing, GBS) but a number of missing data arose since gene expression may greatly vary among individuals. The coverage of RNA seq for each individual, the standardization of the growing conditions before RNA extraction and the development of adapted bio informatics pipelines are key elements for the success of a RNA seq GBS approach. The detected SNPs here can also contribute to the assembly of a specific durum polymorphism database that can be used to develop a micro array chip within a durum wheat consortium.

Our 20% outcrossing mating regimes clearly reduced the long distance linkage disequilibrium in the population and also probably also reduced greatly the within population genetic structure that usually creates spurious association in GWA studies using panels assembled from different genetic sources. More methodological work is needed to set the most efficient value of the outcrossing rate in order to promote effective reduction of haplotype length, reduction of kinship structuration but also to promote a rapid fixation of valuable homozygous individuals in the population If IPBWP appears as genetically diverse compared to a durum wheat panel, its nucleotide diversity is still much lower than the potential diversity available in the exotic parents of the composite. Selection for plant height, removal of plants showing genetic incompatibility and other unidentified selective pressures for adaptation to climate and local pathogens may explain a strong loss of diversity by linkage drag and selective sweeps around the domesticated alleles at locus determining minimum agronomical values. If the decay of linkage disequilibrium is rather steep in the population, low levels of linkage are still present at 50 – 70 cM. This suggests that effective recombination, led by the 20% outcrossing level, was not sufficient to break rapidly and efficiently mix the elite haplotypes with the introgressed ones. Furthermore, if many genes, and not only some major genes major responsible for dramatic and apparent changes in morphology and shape, (e.g., brittle rachis), have been involved in domestication and further improvement of durum quality and agronomic performance, it is likely at the whole genome level that valuable alleles in the exotic germplasm have good chance to be regularly associated with unfavourable alleles. In this case, a more appropriate method to enrich the allelic diversity of such pre-breeding populations would have been first to promote 100% outcrossing and recombination during the first generations and to start conscious massal selection for a return to a “durum” like morphology compatible with modern agronomical practices only after high recombination generations. Such a strategy should reduce the number and strength of the selective sweeps.

In conclusion we claim that new composite populations should be created by controlled crosses of male sterile plants with wild and cultivated emmers, traditional durum, polonicum and turgidum landraces and carthlicum as well sampled to cover the whole diversity of the genetic groups described in this paper (figure 1) or discovered elsewhere. From our first experience in IPBDW, outcrossing rate should be increased to promote effective and rapid recombination to avoid strong selective sweeps. The interplay between outcrossing and selection practices should be theoretically

investigated. Our selection practices in IPBDW were probably too strong to eliminate wild traits such as dispersal or asynchronous growth habits, tallness and hulled kernels. Accepting that these traits co-exist for longer period along with the domesticated phenotypes could be a key for a good introgression of larger levels of diversity in valuable pre-breeding composites. This claims for theoretical approaches to deliver methodological recipes to create, monitor and use of evolutionary pre-breeding populations.
If our population evolved in only one environment (Montpellier; Southern France), such prebreeding composites can be used to create a network of connected populations evolving in contrasted environments. Diversifying selection on a similar genetic background may help to detect chromosomal regions involved in dfferent adaptations patterns (Goldringer et Bataillon, Enjalbert) and are very well adapted to an international collaboration. IPBDW is available for distribution, lines and associated molecular data will soon be released.

References

Acosta-Gallegos, J. A., Kelly, J. D., & Gepts, P. (2007). Pre-breeding and genetic diversity in common bean (Phaseolus vulgaris). In *Proceedings of international plant breeding symposium, Ciudad de México. Crop Sci* (Vol. 47, No. S3, pp. S44-S59).

Alonso-Blanco, C., Aarts, M. G., Bentsink, L., Keurentjes, J. J., Reymond, M., Vreugdenhil, D., & Koornneef, M. (2009). What has natural variation taught us about plant development, physiology, and adaptation?. *The Plant Cell Online*, *21*(7), 1877-1896.

Bhullar, N. K., Street, K., Mackay, M., Yahiaoui, N., & Keller, B. (2009). Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the Pm3 resistance locus. *Proceedings of the National Academy of Sciences*, *106*(23), 9519-9524.

Bogard, M., Allard, V., Brancourt-Hulmel, M., Heumez, E., Machet, J. M., Jeuffroy, M. H., ... & Le Gouis, J. (2010). Deviation from the grain protein concentration–grain yield negative relationship is highly correlated to post-anthesis N uptake in winter wheat. *Journal of experimental botany*, *61*(15), 4303-4312.

Bozzini, A. 1988. Origin, distribution, and production of durum wheat in the world. Pages 1-16 in: Durum Wheat: Chemistry &Technology. G. Fabriani and C. Lintas, eds. Am. Assoc. Cereal Chem.: St. Paul, MN.

Brown, A. H. D., Clegg, M. T., Kahler, A. L., & Weir, B. S. (1990). *Plant population genetics, breeding, and genetic resources*. Sinauer Associates Inc..

Buckler IV, E. S., & Thornsberry, J. M. (2002). Plant molecular diversity and applications to genomics. *Current opinion in plant biology*, *5*(2), 107-111.

Caldwell K.S., Dvorak J., Lagudah S.J., Akhunov E, Luo M.C., Wolters P. and W. Powell; 2004. Sequence Polymorphism in Polyploid Wheat and Their D-Genome Diploid Ancestor. Genetics June 1, 2004 vol. 167 no. 2 941-947.

Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., Kiani, S., ... & Akhunov, E. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences*, *110*(20), 8057-8062.

Civáň P, Ivaničová Z, Brown TA (2013) Reticulated Origin of Domesticated Emmer Wheat Supports a Dynamic Model for the Emergence of Agriculture in the Fertile Crescent. PLoS ONE 8(11): e81955. doi:10.1371/journal.pone.0081955

Cooper, H. D., Spillane, C., & Hodgkin, T. (Eds.). (2001). *Broadening the genetic base of crop production*. CABI.

Diamond J. 2000. De l'inégalité parmi les sociétés - Essai sur l'homme et l'environnement dans l'histoire, Gallimard

FAO (1996). Report of the State of the World's Plant Genetic Resources for Food and Agriculture (Rome).

Feldman, M., and Kislev, M. E. 2007. Domestication of emmer wheat and evolution of free-threshing tetraploid wheat Special Issue: A Century of Wheat Research-From Wild Emmer Discovery to Genome Analysis. Israel Journal of Plant Sciences. 55:207-221.

Kilian, B., & Graner, A. (2012). NGS technologies for analyzing germplasm diversity in genebanks. *Briefings in functional genomics*, *11*(1), 38-50.

Hajjar R, Hodgkin T. 2007. The use of wild relatives in crop improvement: a survey of developments over the last 20 years. Euphytica, 156: 1-13.

Haque, M. A., Takayama, A., Watanabe, N., & Kuboyama, T. (2011). Cytological and genetic mapping of the gene for four-awned phenotype in Triticum carthlicum Nevski. *Genetic Resources and Crop Evolution*, *58*(7), 1087-1093.

Haudry A, A. Cenci, C. Ravel, T. Bataillon, D. Brunel, C. Poncet, I. Hochu, S. Poirier, S. Santoni, S. Glémin and J. David 2007. Grinding up Wheat: A Massive Loss of Nucleotide Diversity since Domestication Molecular Biology and Evolution 2007 24(7):1506-1517.

Horvath A., Didier A., Koenig J. and F. Exbrayat 2009. Analysis of diversity and linkage disequilibrium along chromosome 3B of bread wheat (Triticum aestivum L.). Theor Appl Genet (2009) 119:1523–1537

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070-3071.

Jombart T: adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 2008, 24:1403-1405.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, *11*(1), 94.

Kilian, B., Özkan, H., Pozzi, C., & Salamini, F. (2009). Domestication of the Triticeae in the Fertile Crescent. In *Genetics and Genomics of the Triticeae* (pp. 81-119). Springer US.

Laidò G, Mangini G, Taranto F, Gadaleta A, Blanco A, et al. (2013) Genetic Diversity and Population Structure of Tetraploid Wheats (***Triticum turgidum*** L.) Estimated by SSR, DArT and Pedigree Data. PLoS ONE 8(6): e67280. doi:10.1371/journal.pone.0067280

Luo, M. C., Yang, Z. L., You, F. M., Kawahara, T., Waines, J. G., and Dvorak, J. 2007. The structure of wild and domesticated emmer wheat populations gene flow between them, and the site of emmer domestication. Theor. Appl. Genet. 114:947–959.

Nesbitt, M., and Samuel, D. 1996. From stable crop to extinction? The archaeology and history of the hulled wheats. Pages 41–100 In: Hulled Wheats. Proceedings of the first international workshop on hulled wheats, Castelvecchio, Pascoli, Tuscany. S. Padulosi, K. Hammer, J. Heller (Eds.). International Plant Genetic Resources Institute, Rome.

Özkan H., Willcox G., Graner A., Salamini F. and Kilian B. 2011. Geographic distribution and domestication of wild emmer wheat (*Triticum dicoccoides*) Genet Resour Crop Evol 58:11–53.

Özkan H.• Willcox G.• Graner A., Salamini F., and B. Kilian 2011. Geographic distribution and domestication of wild emmer wheat (Triticum dicoccoides). Genet Resour Crop Evol (2011) 58:11–53.

Özkan, H., Brandolini, A., Schaefer-Pregl, R., and Salamini, F. 2002. AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. Mol. Biol. Evol. 19:1797–1801.

Papa, R. (2013, January). Evolutionary Metabolomics of Durum Wheat Domestication. In *Plant and Animal Genome XXI Conference*. Plant and Animal Genome.

Peleg, Z., Fahima, T., Korol, A. B., Abbo, S., & Saranga, Y. (2011). Genetic analysis of wheat domestication and evolution under domestication. *Journal of experimental botany*, *62*(14), 5051-5061.

Phillips, S. L., & Wolfe, M. S. (2005). Evolutionary plant breeding for low input systems. *The Journal of Agricultural Science*, *143*(04), 245-254.

Roussel V., Leisova L., Exbrayat F., Stehno Z. and F. Balfourier 2005. SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. Theor Appl Genet (2005) 111: 162–170

Salamini F, Özkan H, Brandolini A, Schäfer-Pregl R, MartinW (2002) Genetics and geography of wild cereal domestication in the Near East. Nat Rev Genet 3:429–441

Spillane, C., & Gepts, P. (2001). Evolutionary and genetic perspectives on the dynamics of crop genepools. *Broadening the genetic base of crop production*, 25-70.

Suneson, C. A. (1956). An evolutionary plant breeding method. *Agronomy Journal*, *48*(4), 188-191.

Suneson, C. A. (1969). Evolutionary plant breeding. *Crop Science*, *9*(2), 119-121.

Tanksley SD, McCouch SR, 1997.*Seed banks and molecular maps: unlocking genetic potential from the wild. Science 1997;277:1063-6.*

Tanno K, Willcox G (2006a) How fast was wild wheat domesticated? Science 311:1886

Thuillet AC, Bataillon T, Poirier S, Santoni S., David JL 2005 Estimation of Long-Term Effective Population Sizes Through the History of Durum Wheat Using Microsatellite DataGenetics 169: 1589-1599. .

Valkoun, J. (2001). Wheat pre-breeding using wild progenitors. In *Wheat in a global environment* (pp. 699-707). Springer Netherlands.

Wang J. Luo MC, Chen Z, You F.M., Weil y., Zheng Y. and J. Dvorak 2013. Aegilops tauschii single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat, New Phytologist (2013) 198: 925–937.

Wolfe, M. S., Baresel, J. P., Desclaux, D., Goldringer, I., Hoad, S., Kovacs, G., & Van Bueren, E. L. (2008). Developments in breeding cereals for organic agriculture. *Euphytica*, *163*(3), 323-346.

Wright, S. I., Bi, I. V., Schroeder, S. G., Yamasaki, M., Doebley, J. F., McMullen, M. D., & Gaut, B. S. (2005). The effects of artificial selection on the maize genome. *Science*, *308*(5726), 1310-1314.

Yahiaoui N. Brunnert S., Keller B 2006. Rapid generation of new powdery mildew resistance genes after wheat domestication. The Plant Journal, Volume 47, Issue 1, pages 85–98.

Zaharieva M, Geleta Ayana N., Al Hakimi A., Misra S.C. and P. Monneveux, 2010 Cultivated emmer wheat (*Triticum dicoccon* Schrank), an old crop with promising future: a review. Genet Resour Crop Evol (2010) 57:937–962

Metakovsky, E. V., Kudryavtsev, A. M., Iakobashvili, Z. A., & Novoselskaya, A. Y. (1989). Analysis of phylogenetic relations of durum, carthlicum and common wheats by means of comparison of alleles of gliadin-coding loci. *Theoretical and applied genetics*, *77*(6), 881-887.