



HAL
open science

Wheat syntenome unveils new evidences of contrasted evolutionary plasticity between paleo- and neoduplicated subgenomes

Caroline C. Pont, Florent Murat, Sébastien S. Guizard, Raphaël-Gauthier R.-G. Flores, Séverine S. Foucrier, Yannick Y. Bidet, Umar Masood Quraishi Masood-Quraishi, Michael M. Alaux, Jaroslav J. Doležal, Tzion T. Fahima, et al.

► To cite this version:

Caroline C. Pont, Florent Murat, Sébastien S. Guizard, Raphaël-Gauthier R.-G. Flores, Séverine S. Foucrier, et al.. Wheat syntenome unveils new evidences of contrasted evolutionary plasticity between paleo- and neoduplicated subgenomes. *The Plant Journal*, 2013, pp.1-29. 10.1111/tpj.12366 . hal-00964421

HAL Id: hal-00964421

<https://hal.science/hal-00964421v1>

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Received Date : 27-Aug-2012
Revised Date : 01-Oct-2013
Accepted Date : 08-Oct-2013
Article type : Original Article

WHEAT SYNTENOME UNVEILS NEW EVIDENCES OF CONTRASTED EVOLUTIONARY PLASTICITY BETWEEN PALEO- AND NEODUPLICATED SUBGENOMES.

Caroline Pont^{1*}, Florent Murat^{1*}, Sébastien Guizard¹, Raphael Flores², Séverine Foucrier¹, Yannick Bidet³, Umar Masood Quraishi¹, Michael Alaux², Jaroslav Doležel⁴, Tzion Fahima⁵, Hikmet Budak⁶, Beat Keller⁷, Silvio Salvi⁸, Marco Maccaferri⁸, Delphine Steinbach², Catherine Feuillet¹, Hadi Quesneville², Jérôme Salse^{1**}.

¹**INRA/UBP UMR 1095, Centre de Clermont-Ferrand-Theix**, 5 Chemin de Beaulieu, 63100 Clermont Ferrand, France.

²**INRA/URGI, Centre de Versailles**, bâtiment 18, route de Saint Cyr Versailles, France.

³**Clermont Université/Plateforme GINA, Centre Jean Perrin**, 58 rue Montalembert, 63011 Clermont-Ferrand, France.

⁴**Centre of the Region Haná for Biotechnological and Agricultural research, Institute of Experimental Botany**, Olomouc, Czech Republic.

⁵**University of Haifa, Department of Evolutionary and Environmental Biology**, Faculty of Natural Sciences University of Haifa Mt. Carmel, Haifa 31905, Israel.

⁶**Sabancı University, Faculty of Engineering and Natural Sciences**, Orhanli, Tuzla-Istanbul. Turkey.

⁷**University of Zurich, Institute of Plant Biology**, Zollikerstrasse 107, 8008 Zurich, Switzerland.

⁸**University of Bologna, DiSTA - Agronomy**, Viale Fanin, 44, 40127 Bologna, Italy.

*These authors equally contributed to the work.

****Corresponding author:**

Jerome Salse

INRA-Université Blaise Pascal, UMR1095 Génétique, Diversité et Ecophysiologie des Céréales
234 Avenue du Brézet, 63100 Clermont-Ferrand, France.

Phone : (33) 4 73 62 4380.

Fax : (33) 4 73 62 4453.

Mail : jsalse@clermont.inra.fr

Running head: Wheat Genome Evolution.

Key words: Paleogenomics, Dominance, Partitioning, COS, SNP.

Manuscript type: Research article.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/tpj.12366

This article is protected by copyright. All rights reserved.

Web site : <http://urgi.versailles.inra.fr/synteny-wheat>.

SUMMARY

Bread wheat derives from a grass ancestor structured in 7 protochromosomes followed by a paleotetraploidization to reach a 12 chromosomes intermediate and a neohexaploidization (involving subgenomes A, B and D) event that finally shaped the 21 modern chromosomes. Insights into wheat syntenome in sequencing COS (Conserved Orthologous Set) genes unravelled differences in genomic structure (such as gene conservation and diversity) and genetical landscape (such as recombination pattern) between ancestral as well as recent duplicated blocks. Contrasted evolutionary plasticity is observed where the B subgenome appears more sensitive (*i.e.* plastic) in contrast to A as dominant (*i.e.* stable) in response to the neotetraploidization and D subgenome as supradominant (*i.e.* pivotal) in response to the neohexaploidization event. Finally, the wheat syntenome, delivered through a public web interface PlantSyntenyViewer at <http://urgi.versailles.inra.fr/synteny-wheat>, can be considered as a guide for accelerated dissection of major agronomical traits in wheat.

INTRODUCTION

Recent comparative genomics studies based on monocot genome sequences, including Panicoideae (sorghum, Paterson *et al.*, 2009; maize, Schnable *et al.*, 2009), Ehrhartoideae (rice, IRGSP 2005), and Pooideae (*Brachypodium*, IBI 2010), suggest that grasses derive from $n = 7$ (alternative scenario with $n=5$) to 12 ancestral karyotypes (named AGK for Ancestral Grass Karyotypes). Modern grass genomes were then shaped from this AGK through a shared whole genome duplication (WGD) followed by ancestral chromosome fusion (CF) events (for review Salse 2012). Polyploidization has been shown to be followed by genome-wide diploidization (also referenced as partitioning) through differential elimination of duplicated gene redundancy at the whole genome level (Wang *et al.*, 2005; Schnable *et al.*, 2012ab; Freeling *et al.*, 2012), leading to dominant (*i.e.* D, stable genomic compartment associated with low duplicated gene loss and reduced gene diversity) and sensitive (*i.e.* S, plastic genomic compartment associated with high duplicated gene loss and increased gene diversity) subgenomes in modern grass species (Woodhouse *et al.*, 2010; Schnable *et al.*, 2012ab; Abrouk *et al.*, 2012).

Bread wheat is a good plant model to study the impact of distinct rounds of WGD on the subgenome dominance phenomenon, as its genome comprises (i) seven ancestral paleoduplicated blocks corresponding to the shared paleotetraploidization event identified in all known cereal genomes and dating back to 65 million years ago (hereafter, mya), as well as (ii) two recent neopolyploidization events leading to *Triticum aestivum*, which originated from two hybridizations between *T. urartu* (A genome) and an *Aegilops. Speltoides* related species (B genome) 1.5 mya, forming *T. turgidum ssp. diccoides*; and between *T. turgidum ssp. durum* (genomes A-B) and *A. tauschii* (D genome) 10,000 years ago (Feldman *et al.*, 1995). Wheat diploidization (between A, B and D subgenomes) has been

partially investigated at the whole genome and gene levels where recent transcriptome analysis in hexaploid wheat suggested that 39 up to 46% of the wheat homoeologous genes may have either been lost or neo-/subfunctionalized within 1.5 my of evolution (Pont *et al.*, 2011). Such structural or functional diploidization brings unbalanced polyploidy gene systems into diploid-like mode of expression (Edger and Pires 2009).

The recent access to a large bread wheat genomic resources offers the opportunity to study in the same analysis not only the structural plasticity of paleoduplicated genes (during the last 50-70 million years of evolution) but also neoduplicated genes (during 0.1-1.5 million years of evolution) by comparing the conservation of A, B and D homoeologous gene copies, *i.e.* wheat homoeoalleles. Wheat genomics resources have been recently published with the release of the wheat genome shotgun sequences in hexaploid (Brenchley *et al.*, 2012) and diploids (D genome ancestor sequence in Jia *et al.*, 2013 and Luo *et al.*, 2013 as well as the A genome progenitor sequence in Ling *et al.*, 2013). However, these studies reported incomplete gene repertoire (*i.e.* whole-genome short reads assembled into gene models instead of whole chromosome pseudomolecules) either not ordered or partially ordered based on synteny relationships, mainly with *Brachypodium*. Independently, genome-wide diversity maps have been also made recently available in hexaploids (Allen *et al.*, 2011, 2013; Chao *et al.*, 2009; Winfield *et al.*, 2012; Lai *et al.*, 2012; Cavanagh *et al.*, 2013), tetraploids (Saintenac *et al.*, 2011; Trebbi *et al.*, 2011; Ren *et al.*, 2013) or diploid progenitors (Wang *et al.*, 2013; You *et al.*, 2011). However, most of these studies relied on transcriptome (*i.e.* exome) sequencing from different genotypes and then are dependent on both the tissues used and the expression of the three homoeologs on the considered tested tissues and developmental conditions, then potentially leading to possible bias in homoeoallele diversity estimation due to gene silencing in the tested biological conditions.

In contrast, the evolution of the homoeologous gene space in hexaploid wheat can be investigated through comparative genomics approaches trying to introduce no (or reduced) bias in gene loss and sequence polymorphism assessment (Bekaert *et al.*, 2011; Thomas *et al.*, 2006). Up to now very few genes are physically and genetically mapped in hexaploid bread wheat. The integration of both genomic and genetic resources, described in the previous section, offers the opportunity to provide the most accurate wheat syntenic (or also referenced as computed, Pont *et al.*, 2011) gene order (*i.e.* defined as syntenome) and test its accuracy as this will probably represent in a long term the wheat reference genome, until complete pseudomolecules will be publicly released for the 21 chromosomes. Such validated wheat syntenome will also offer the opportunity to perform a comprehensive analysis of the wheat gene space evolutionary plasticity during the last 100 million years and to investigate the impact of paleo- and neopolyploidization events on genome rearrangement and gene diversity. Despite insights into wheat subgenome evolution, the wheat syntenome can also be consider as an applied tool for the dissection of major traits in wheat.

RESULTS

WHEAT SYNTENOME DEFINES PALEO/NEODOMINANT AND SENSITIVE BLOCKS.

Grasses has been proposed to derive from a $n=7$ ancestor that has been duplicated to reach a $n=14$ intermediate followed by two chromosomal fusions to reach a $n=12$ ancestor (Figure 1A center of the circle and inner circle A1-A12) founder of all the modern grasses, Salse 2012. Comparing grass genome sequences, we identified a 17,317 Conserved Orthologous Set (*i.e.* COS, also referenced as protogenes) located on 12 syntenic blocks (*i.e.* Conserved Ancestral Regions, CARs) covering the rice, maize, sorghum, *Brachypodium* genomes (Figure 1A inner cereal circles and Table 1) and refining the previously reported conserved gene repertoire in grasses of 9,731 protogenes (Murat et al. 2010). The 17,317 COS has been used as a matrix to produce a wheat syntenome where on average 2,474 COS have been ordered on the seven wheat chromosome groups in respect to the position of their orthologous counterparts following the ordering priority of rice > *Brachypodium* > sorghum (*cf* wheat consensus circle on Figure 1A, Table 1). We made this wheat syntenome available through a public web interface named PlantSytenyViewer at <http://urgi.versailles.inra.fr/syteny-wheat> (illustrated in Figure 1B & with raw data available in Table S1). The wheat syntenome offered the opportunity to unravel the evolutionary fate of such COS in wheat in characterizing retained *vs.* lost duplicates in response to both ancestral shared (~65 mya) and recent specific (>1.5 mya) polyploidization events. To do this, we performed two complementary strategies, bypassing the requirement of a complete reference genome sequence, consisting in aligning the public wheat gene repertoire to the previous COS-based wheat syntenome (strategy #1 in Figure 1A referenced as ‘Wheat→COS’) as well as COS *de novo* sequencing in wheat (strategy #2 in Figure 1A referenced as ‘COS→Wheat’), Figure S1.

The first strategy (#1, black arrow Figure 1A) in investigating the retention of ancestral genes in bread wheat consisted in comparing the recently published wheat genome shotgun sequence (Brenchley *et al.*, 2012) to the previous wheat syntenome. When comparing the published wheat genome consisting in ~95K gene models with 18,483 of them assigned to the three subgenomes (15,996 on A, 15,244 on B and 16,880 on D) to the 17,317 ordered protogenes in wheat (mapped on chromosome groups 1 to 7), 6,423 (35%) ancestral grass genes appear retained in wheat (Table S1, Figure 1A). This reduced rate of gene conservation is consistent with the 17,093 genes in *A. tauschii* with 26.1% associated with orthologs in grasses (Luo *et al.*, 2013) as well as 42% (14,578 genes) of the 34,879 gene models from *T.urartu* reported as conserved with *Brachypodium* (Ling *et al.*, 2013). Such limited level of single gene orthologous relationships between the wheat gene repertoire and the ancestral retained grass genes in the other species, is largely due to wheat-specific gene content amplification though intra- and inter-chromosomal duplications (Figure S2). The distribution of the retained ancestral genes on the wheat subgenomes (4905, 158, 347, 652, 62, 187, 112 respectively located on A-B-D, A-B, B-D, A-D, B, D and A genomes specifically) showed an increased gene density at the subtelomeric regions compared to the centromeres (*cf* BIN-based gene density distribution in green on the Figure

1A). Homoeolog gene density appeared correlated to the recombination rate, as previously reported in the diploid *A. tauschii* (Jai *et al.*, 2013, Luo *et al.*, 2013) and *T. urartu* (Ling *et al.*, 2013) genome sequences. The density of non-collinear genes per bin was correlated with recombination rate, with non-collinear genes located in distal compartments of the chromosomes, then suggesting that non collinear genes may have survived in distal high recombination regions compared to proximal low-recombination regions.

The wheat syntenome illustrates that the seven triticeae chromosome groups originated from a $n=12$ ancestral species by five nested chromosome fusions (NCFs), involving chromosomes 1 (A5+A10), 2 (A4+A7), 4 (A3+A11), 5 (A12+A9+A3), 7 (A6+A8), Figure 1A. The previous chromosome-to-chromosome relationships, established at the BIN level between wheat and the sequenced grass genomes, allowed us to transfer the ancestral dominant (A1-2-3-4-9-11, associated with higher retention of ancestral genes) and sensitive (A5-6-7-8-10-12, associated with higher loss of ancestral genes) chromosomal blocks reported in grasses (Figure S3) into the 21 bread wheat chromosomes (Figure 1A with D and S blocks). Among the 157 BIN available on the 21 bread wheat chromosomes, 79 and 69 were respectively identified as dominant and sensitive according to the known nature of their orthologous counterparts in the sequenced grass species (Table S2). The Triticeae genomes, that derived from the $n=12$ ancestor (center on the circle on Figure 1A) through 5 ancestral chromosome fusions, can then be re-organized into paleo-D and paleo-S (*i.e.* paralogous Dominant (D) and Sensitive (S) blocks deriving from the ancestral WGD) chromosomal compartments where $w1(S)=A10(S)+A5(S)$, $w2(D)=A7(S)+A4(D)$, $w3(D)=A1(D)$, $w4(D)=A11(D)+A3(D)$, $w5(D)=A9(D)+A12(S)$, $w6(D)=A2(D)$, $w7(S)=A8(S)+A6(S)$.

Such high resolution BIN-based wheat syntenome offered the opportunity to investigate the retention of the ancestral genes in wheat following the paleotetraploidization event (by comparing D and S wheat blocks defined previously) as well as the neohexaploidization event (by comparing the A, B and D subgenomes), as illustrated in Figure 2A (top) with D chromosomes in blue and S chromosomes in red. Regarding the number of retained genes in D vs S blocks, we observed the expected higher retention ($p<5\%$) of protogenes in paleodominant blocks compared to paleosensitive counterparts (Figure 2B, left). This result confirm the existence of shared D and S blocks in wheat following the paleoteraploidization (~ 65 mya) as reported in rice, sorghum, *Brachypodium* and maize (Abrouk *et al.*, 2012). Surprisingly, when comparing the retention of protogenes in the A, B and D subgenomes, we also observed a bias retention ($p<5\%$, Figure 2B right) of ancestral genes, with the B subgenome showing the property of a so-called sensitive subgenome with higher loss of protogenes. This observation raised the hypothesis that the reported genome partitioning phenomenon following the ancestral shared paleotetraploidization event (bias retention of protogenes between D and S blocks confirmed in wheat as illustrated in Figure 2B left) may also acts between the A, B and D subgenomes deriving form the neohexaploidization, with A and D as dominant (*i.e.* stable) and B as sensitive (*i.e.* plastic), Figure 2B right. This confirms and largely complements recent evidences of differential gene

loss observed between the subgenomes of the wheat chromosome group 7 with increased gene numbers reported respectively on D>A>B (Berkman *et al.*, 2013).

However, the previous conclusions of wheat genome partitioning in response to both paleo- and neopolyploidization events were obtained in exploiting the public wheat genome shotgun sequence (Brenchley *et al.*, 2012) that consists only in a partial gene repertoire. We cannot exclude that missing (non-sequenced) genes/families/homoeologs in such resource may have impacted our conclusion regarding the observed structural plasticity of paleo- and neoduplicated compartments in modern hexaploid bread wheat. Consequently, in order to validate and complement the previous conclusions regarding the wheat subgenome dominance, we performed a second strategy (#2, black arrow Figure 1A, Figure S1) that consisted in *de novo* sequencing and mapping of COS genes in wheat deriving from the 17,317 protogenes but not included in the previous 6,423 wheat/grass ortholog repertoire (Figure S4). To do so, we applied a COS-finder tool (see method, Quraishi *et al.* 2009) to the 10,894 (17,317 total COS excluding the 6,423 COS identified from Brenchley *et al.*, 2012) ancestral grass genes not associated with a public wheat gene model sequence. We have then been able to develop a tremendous catalog of 6,033 primer pair set (primers pairs selection criteria in Method section) defining 5,234 COS relationships (Table S1). Such COS primers, selected on highly conserved exonic regions, offer all the guarantee to be useful in sequencing the wheat orthologs (and associated homoeologs), based on the observed transferability of such COS markers among monocot species (*i.e.* sorghum, maize, oat, rice, Triticale, wheat, *Brachypodium* and rye) at both the amplification (Figure S5A on agarose gels and Figure S5B on capillary sequencer) and sequence levels (Figure S5C with melting curve profiles supported by sequence-based haplotyping data). Overall, the characterized COS markers (17,317 COS with 5,234 selected COS for sequencing and 6,033 associated primer pairs) as well as the COS-finder software are made available at <http://urgi.versailles.inra.fr/synteny-wheat> (Figure 1B, Table S1).

In order to validate the orthologous, homoeologous and paralogous gene contents and retention in wheat paleohistory, raising the evolutionary model suggesting subgenome dominance between D>A>B, we performed the sequencing of the selected 5,234 COS (Method section and detailed strategy in Figure S6). COS markers have been sequenced in *Chinese Spring* using the 454 (Roche) technology, delivering 23,463 reference transcripts (RTs) covering 4,582 COS (with an average of 3 RTs per COS, putatively homoeoalleles, Figure S6). This result established that 88% of the COS (*i.e.* genes conserved in rice, *Brachypodium*, maize and sorghum) are conserved in wheat, in contrast to 35% of the 18,483 gene models (from Brenchley *et al.*, 2012) that have been associated with a COS gene in the previous section. This result reinforces the previous conclusion (Figure S2) that wheat has been enriched, in the course of evolution, by species and/or lineage specific genes. The capture (Agilent SureSelect Target Enrichment) of the 23,463 RTs in eight bread wheat genotypes (*Chinese Spring*, *Renan*, *Courtot*, *Alcedo*, *Brigadier*, *Genial*, *Hustler* and *Nicam*) and short read sequencing (Illumina HiSeq2000) delivered 9,969 high quality SNP covering 2,197 COS markers with a density

of 1.9 SNP / 500bp (Table S1, Figure S7). Such *de novo* COS sequences in wheat allowed us to confirm the bias retention of protogenes defining dominant as well as sensitive blocks for both paleo- and neopolyploidization events (Figure 2C). The number of sequenced COS in wheat observed in dominant and sensitive compartments confirmed our previous conclusions based on the public gene repertoire, where more retained ancestral genes (or orthologs) are observed in the dominant blocks (Figure 2C left). We also confirmed the bias retention of homoeologs between A, B and D subgenomes (Figure 2C right, Figure S8). Overall, we then proposed, based on a pure *in silico* experiment (alignment-based synteny analysis) complemented by a *de novo* sequencing strategy (COS characterization), that the ancient subgenome dominance process related to the shared paleotetraploidization in grasses is observed in wheat and has been eroded (p-values between $\sim 3.10^{e-2}$ to $\sim 5.10^{e-2}$) by the neohexaploidization event that may have led to a modern subgenome dominance where B is sensitive and A–D are dominant (p-values between $\sim 1.10^{e-3}$ to $\sim 8.10^{e-7}$).

GENOME PARTITIONING SHAPED WHEAT GENETIC LANDSCAPE.

In order to address not only the bias in gene retention (detailed in the previous section) but also in gene diversity between modern and ancient dominant vs. sensitive blocks, we used the overall depth of read coverage across COS as well as the depth of read coverage at variable sites to detect not only sequence variants (*i.e.* 9,969 SNPs, Figure S7) but also putative structural variants such as PAVs (*i.e.* Presence/Absence Variations), CNVs (*i.e.* Copy Number Variations) in contrast to NSV (*i.e.* Non Structural Variation, where all genotypes are covered for the considered COS). Figure S9 illustrates three distinct COS markers associated with (i) sequences for the eight sequenced genotypes [COS-5368, observed in 80% (4,286 COS) of the cases], (ii) missing sequences [COS-542, PAV observed in 6% (304 COS) of the cases] or (iii) extra copy sequences [COS-3070, CNV observed in 14% (644 COS) of the cases]. We then provide a complete view of the gene diversity distribution in wheat based on the precise characterisation of both non-structural (SNPs) as well as structural variants (InDels, CNVs and PAVs). The wheat COS sequences offered the opportunity to test the impact of the reported biased chromosomal plasticity between dominant and sensitive blocks on gene diversity through differential SNP/InDel contents observed in such wheat genomic compartments. Figure S10 illustrates the observed distribution of InDel and SNP polymorphisms in wheat where the more frequently observed (61% of the cases) sequence variations of SNPs correspond to transitions (A/G or T/C) compared to transversions (A/T, A/C, T/G or G/C). The most frequently observed (52% of the cases) sequence variations of InDels correspond to a single nucleotide. The distribution of SNPs on D and S blocks as well as A, B and D subgenomes shows that the gene diversity increases in the sensitive blocks (either paleo-S or B genome, Figure 2D, plain box-plots). Wheat chromosome groups derived five fusions of the 12 ancestral chromosomes then juxtaposing high-gene-density terminal paleoregions near low-gene-density paleocentromeric regions in the post-fusion neo-chromosome (Figure 1A dotted connecting black lines Murat *et al.*, 2010). Consequently, despite bias distribution of sequence variants between dominant and sensitive subgenomes, we observed a higher gene (COS)

diversity (*i.e.* SNP/bp density) in telomeric regions and synteny breakpoints (SBP) regarding the wheat chromosome groups that derived from the centromeric nested insertion of ancestral chromosomes (Figure 1A black stars and Figure S11).

Until the wheat genome is entirely sequenced and available (in delivering complete pseudomolecules) the discovery and application of genome polymorphism for wheat crop improvement and evolutionary investigation remains a real challenge, where the provided wheat syntenome constitutes a tremendous matrix for these purposes. Based on the previous *in silico* syntenome and SNP-based diversity map, we addressed differences in gene content and plasticity between dominant and sensitive fragments for both paleo- and neopolyploidization events. As the previous conclusions derived from pure *in silico* syntenome and COS-SNPs characterization, we validated the delivered computed wheat gene order (17,317 ordered COS) as well as associated sequence variants (9,969 SNPs) through COS-SNP mapping (Figure S4). As a validation procedure, we have randomly selected a subset of 1,135 (22%) among the 5,234 COS markers to test the accuracy of the characterized *in silico* COS-SNPs as well as computed gene order of such COS. We have identified 807 *in silico* SNPs from 986 (out of 1135, 87%) COS markers that have been successfully sequenced in genotypes for which mapping populations are available (see experimental procedure section). 540 high quality scoring (Quality score A to D) and 267 low quality scoring (Quality score classes E and F) *in silico* SNPs (based on SNP calling criterion described in Figure S7) have been tested. 357 (66%) and 63 (24%) respectively have been validated as polymorphic between bread wheat cultivars *Chinese Spring*, *Courtot*, *Arche*, *Recital* and *Renan*. 375 (89%) out of the 420 (357 high quality + 63 low quality scores) validated SNP have been successfully mapped (Illumina approach) on the three available mapping populations (*Courtot x Chinese Spring*, *Arche x Recital*, *Renan x Recital*, see experimental procedure section), cf Figure 1A (red bars on the wheat chromosome circles) and Table S1. Such observed validation rate ($420/807=52\%$) of *in silico* SNPs is consistent with the one (60%) reported by Allen *et al.*, 2011 as well as in Trick *et al.*, 2012 (56 to 58% of validation accuracy). Overall, based on the previous genotyping data, we provided in the current analysis 9,969 high confidence SNPs (deriving from 5,234 sequenced COS markers) for which 4,613 ($9,969 \times 0.52 \times 0.89$) are expected to be converted efficiently in any functional assays. The remaining non-polymorphic *in silico* SNPs may be related to (i) heterogeneity observed in using different genotyping approach (for example five SNPs failed to be mapped using Illumina technology but recovered with KASpar) (ii) *in silico* intervarietal high quality SNPs that happened to be homoeoSNPs (iii) SNPs associated with the same quality score that can either be validated or not through Illumina genotyping approach (Figure S12).

COS-SNPs have been mapped and tested on Illumina, KASpar, SSCP, HRM, LNA detection (see experimental procedure section) to test the accuracy of the provided computed wheat gene order based on the observed COS genetic position, Figure 2B. Overall, taking into account the mapped COS-SNP information over the 21 bread wheat chromosomes, 89% (335) of mapped COS were retained at the conserved (*i.e.* expected orthologous position) loci whereas 40 are non-syntenic (Table

1, figure S13). Whereas 147 and 184 COS-SNPs have been mapped on the A and B genomes, only 44 have been positioned on the D genome (Table 1 and illustrated as ‘mapped COS rate’ on Figure 2D, dashed box-plots) supporting the higher plasticity of the B (sensitive) compared to the A (dominant) related to long term paleo-tetraploidization evolutionary event; and between A/B and D subgenomes due to the low level of genetic diversity in the D genome associated with recent origin of hexaploid wheat. Comparison of the *in silico* COS-SNP density (Figure 2D, plain box-plots representing the observed ‘SNP/500COS ratio’ from 9,969 *in silico* COS-SNPs) and mapped COS-SNP rate (observed from the 375 mapped COS-SNPs) it appeared that the D genome diversity may have been overestimated *in silico* through the possible mis-association of SNPs either on the A or D genomes, especially for short sequence reads not harbouring clear and multiple homoeoSNPs (Quality score classes E and F, cf Figure S7). Interestingly the 11% (40 genes) of COS-SNP markers mapped at non-orthologous positions (that can be considered as transposed genes) are mostly located (60%) on the B genome, reinforcing the structural plasticity nature of such genomic compartment (Table 1). As expected, due to the considered time frame (~65 mya), bias in gene diversity (expressed as SNPs/bp) cannot be observed for the ancestral paleotetraploidization event (Figure 2D left). The same observation is made when investigating the impact of paleo-dominance and paleo-sensitivity on the recombination pattern where no difference can be observed between paleo-D vs. paleo-S blocks ($p>5\%$) but a higher cM/marker ratio was observed for the B compared to the A ($p=2,4.10^{-2}$) and the D ($p=1.10^{-3}$) subgenomes (Figure 2E). The observed reduced cM/marker ratio of D subgenome compared with the A and B counterparts has been attributed to the loss of polymorphism during the genetic bottleneck that accompanied the development of modern elite cultivars. Overall, taking into account the provided COS-SNP markers associated with reference public wheat genetic maps, we then produced the most complete composite bread wheat genetic map (consisting in 7,520 molecular markers covering 4,318.03 cM with a marker density of one marker every 0.78 cM) of the 21 chromosomes and including mapped RFLP (1,687), AFLP (712), STS (262), SSR (2,315), DaRTs (1,246) and COS-SNP (375) gene markers, named Wheat Composite Genetic Map ‘WCGM2013’ (available as Table S3 and detailed provided in Figure S14).

Genome partitioning in delivering D and S fragments (shown as associated to bias in gene content, diversity and recombination in the current study) following polyploidy has been suggested to be epigenetically driven (Doyle *et al.*, 2008). To test this hypothesis, we investigated C-band (from Hutchinson *et al.*, 1982) differences between paleo- and neodominant and sensitive blocks (Figure 2F). Where differences are observed between ancient D and S compartments $p=2,7.10^{-2}$, Figure 2F left), the difference is more visible between the neodominant (A-D subgenomes) and neosensitive (B subgenome, $p=1,210^{-2}$ and 2.10^{-3} respectively) compartments, Figure 2F right). The transition nature (A/G or T/C) of the reported 9,969 SNPs may also indicate the historic methylation profile of each subgenomes as the transition abundance bias is commonly observed for SNPs reflecting the high

Accepted Article

frequency of C to T mutation following methylation (Coulondre *et al.*, 1978). Such a bias observed in polyploid wheat is greater than that observed in its diploid Triticea relative barley (Duran *et al.*, 2009) and may reflect a higher level of methylation in polyploid wheat genome as part of the diploidization process leading to D (low methylation) and S (high methylation) compartments. Both observations (C-band differences as SNP transition enrichments) lead to the opening question of epigenetic driving the observed paleo- and neodominance in wheat as well as in other grasses. Overall, the clear characterisation of ancient and recent D-S compartments, in relation respectively to the paleotetraploidization and the neohexaploidization events, suggest that they may have shaped the modern bread wheat genomic and genetic landscapes with bias in gene number (*i.e.* higher gene retention in dominant blocks), gene diversity (*i.e.* higher variant per genes in sensitive blocks), recombination (*i.e.* lower cM/Marker in sensitive blocks) and heterochromatin structure (*i.e.* higher C-band and possibly methylation in sensitive blocks).

BREAD WHEAT EVOLUTIONARY MODEL.

Based on the previous observations, we propose an evolutionary scenario that has shaped the modern bread wheat genome during the last 65 my of evolution in response to three polyploidization events (Figure 3A). The $n=7$ AGK has been paleotraploidized leading to seven pairs of dominant (red) and sensitive (bleu) chromosomes. Five ancestral chromosome fusions took place to build the seven Triticeae chromosomes and wheat ancestors (*i.e.* *T. urartu*, *A. speltoides* and *A. tauschii*), where T3-T4 are entirely sensitive, T1-T7 are entirely dominant and T2-T5-T6 are structured as a mosaic of D and S blocks. The first neotetraploidization event (1.5 mya) led to a subgenome dominance where the A subgenome became dominant and the B subgenome sensitive. The second neohexaploidization event (0.1 mya) led to a supra-dominance where the tetraploid became sensitive (subgenomes A and B) and the D subgenome dominant. Thus, the modern bread wheat genome can be divided into five supra-dominant chromosomes (also referenced as pivotal) deriving from surimposed dominances (*i.e.* chromosomes 3D, 2D-S/L, 4D, 5D-L, 6D-S/L) in contrast to five supra-sensitive ones (*i.e.* chromosomes 1B, 2B-C, 5B-S, 6B-C, 7B), and the remaining regions are associated with opposite D and S following the successive duplication events (*i.e.* with S for short arm, L for long arm and C for centromeric region), see Figure 3A (with the Stability/Plasticity scale at the right).

In order to investigate precisely supra-dominant and supra-sensitive chromosomal architecture, we focussed on rice chromosomes 1 (dominant) and 5 (sensitive) associated with differences in gene content, orthologous/ancestral gene retention, chromosome length as well as transposable element (mainly class II) repertoire (Figure 3B left). The orthologous counterparts in wheat (Group 1 as sensitive and Group 3 as dominant) display the same bias in number of orthologs (*i.e.* COS number higher on the dominant group 1) as well as gene diversity (*i.e.* SNP/COS density higher on the sensitive group 3). In our scenario, among the six investigated chromosomes, 3D is supra-D in

contrast to 1B as supra-S. Consequently the chromosome 1B displays the lowest gene retention, the highest gene diversity, lowest cM/marker ratio, the highest number of transposed genes and highest number of C-band.

DISCUSSION

WHEAT GENOME ARCHITECTURE HAS BEEN SHAPED BY SUBGENOME PARTITIONING FOLLOWING PALEO- AND NEOPOLYPLOIDIZATION.

Common anchors (COS genes) to compare genomes and reconstruct conserved ancestral regions (CARs) are necessary to unravel plant genome paleohistory (Salse 2012). It has been shown that CARs duplication in the grass paleohistory has been followed by a structural partitioning in defining post-duplication dominant regions (defined as structurally stable with higher retention of protogenes) in contrast to sensitive paralogous counterparts (defined as structurally plastic with higher loss of protogenes), Schnable *et al.*, 2012ab; Abrouk *et al.*, 2012. Based on the chromosome-to-chromosome synteny relationships established between the seven bread wheat chromosome groups and the rice, sorghum, *Brachypodium* and maize genomes, it was then possible to produce a partial wheat gene-based physical map (*i.e.* syntenome) including 17,317 COS. However, in the absence of large scale gene mapping data in wheat, how relevant such computed gene order is robust, remains an open question and is of importance if one wants to investigate the evolution of wheat gene content in comparison of the sequenced grasses genomes. Our current study aiming at sequencing and mapping ancestral COS genes in bread wheat established that the synteny-based computed gene order delivered is currently correct for up to 89% of the simulated gene order and that lineage specific rearrangement in wheat may account for less than 11% of the ordered genes *in silico*.

Such wheat syntenome, that deliver *in silico* synteny-based ordered genes conserved in close relatives, is a perfect matrix to characterized genes that have been conserved or lost in bread wheat evolution in response to paleotetraploidization (~65 mya) and neohexaploidization (>1.5 mya) events. Using complementary strategies consisting in aligning the public wheat genome repertoire (Brenchley *et al.*, 2012) as well by *de novo* COS gene sequencing in wheat, we identified respectively 35% of wheat genes associated with an ortholog in the other grasses and 88% of COS successfully sequenced in wheat, suggesting that lineage-specific gene amplification and shuffling mechanisms have shaped the wheat genome in its recent evolution. More interestingly, more conserved genes are observed in dominant wheat regions in contrast to sensitive ones arising from the ancestral shared paleotetraploidization event as reported in the other grass species (Abrouk *et al.*, 2012). Moreover, such bias in gene conservation is also observed between the three wheat subgenomes deriving from the neohexaploidization, with more orthologs detected in D>A>B. This subgenome dominance phenomenon may then explain the observed differences in genetic and physical size reported between subgenomes, the B (plastic) as the largest and the D (stable) the smallest (Furuta *et al.*, 1986). Bias retention of ancestral genes has been observed between paleoduplicated (plasticity of S>D) as well

between neoduplicated (plasticity of subgenomes B>A>D) genes, suggesting that ancient as well as recent polyploidization events are followed by a diploidization mechanism consisting in the structural partitioning of the paralogous fragments.

Despite the characterization of conserved genes in the wheat subgenomes leading to the identification of paleo- as neodominant and sensitive chromosomal regions, we investigated non-syntenic genes in wheat corresponding (i) at the genome level to lineage specific intra- (18%) and interchromosomal (82%) duplications, and (ii) at the gene level to CNVs (14%), PAVs (6%) as well as gene transposition (20%). Since the COS markers have been selected as single copy conserved genes, it is unlikely that all cases of non-syntenic genes were due to erroneous mapping of duplicates or homologs. That raises the possibility of a small scale (*i.e.* few genes) transposition/movement mechanism despite large scale ones (such as translocations and inversions) involving DNA segments carrying several genes, operating in the decay of synteny between grass genomes and then driving non-conserved genes in wheat (*i.e.* ~50% initially reported in wheat from Pont *et al.*, 2011; and ~60% in the current analysis). More transposed genes have been observed on the B subgenome (sensitive according to the current analysis), that is consistent with the conclusion raised on chromosome group 1 (Wicker *et al.*, 2011), based on partial sequence investigations. Gene movement may appear as a particularly active phenomenon in polyploidy wheat and may act then preferentially on the sensitive compartments.

REVISITING THE B GENOME PROGENITOR ENIGMA BASED ON CONTRASTED PLASTICITY FOLLOWING POLYPLOIDIZATION.

Several wheat phylogeny studies have tried to identify the progenitor of the B genome of polyploid wheat based on cytology (Zohary and Feldman 1962), nuclear and mitochondrial DNA sequences (Dvorak *et al.*, 1989; Dvorak and Zhang 1990; Terachi *et al.*, 1990) as well as chromosome rearrangement studies such as common translocation events (Feldman 1966ab; Hutchinson *et al.*, 1982; Gill and Chen 1987; Naranjo *et al.*, 1987, 1990; Jiang and Gill 1994; Devos *et al.*, 1995; Maestra and Naranjo 1999). More recent and representative molecular comparisons, at the whole genome level using germplasm collections, have shown that the B genome could be related to several *A. speltoides* lines but not to other species of the *Sitopsis* section (Kilian *et al.*, 2007; Salina *et al.*, 2006). At the *SPA* locus level (Salse *et al.*, 2008), close relationships between the *A. speltoides* and the hexaploid B subgenome has been reported based on both coding and non-coding sequence comparisons, but with a lower conservation compared to the A subgenome and its *T. urartu* progenitor at the PSR920 region (Dvorak and Akhunov 2005, Dvorak *et al.*, 2006). The greater genetic diversity of the B genome compared to the A genome of hexaploid wheat was also reported based on SNPs (previous references), SSRs (Roder *et al.*, 1998ab), RFLPs (Liu *et al.*, 1991), SNPs (Akhunov *et al.*, 2010), as well as in tetraploid (Thuillet *et al.*, 2005, Ren *et al.*, 2013). Such contrasted conservation between B subgenome and *A. speltoides* compared to A genome and its *T. urartu* progenitor is classically explained with two hypothesis where (i) the progenitor of the B

genome is a unique *Aegilops* species that remains unknown (*i.e.* monophyletic origin and ancestor closely related to *A. speltoides* from the *Sitopsis* section) or (*ii*) this genome resulted from an introgression of several parental *Aegilops* species (*i.e.* polyphyletic origin) that need to be identified from the *Sitopsis* section. Overall the B genome is therefore with greater divergence compared to the A and D subgenomes relative to their putative diploid progenitors, with the D subgenome as the most stable in the course of evolution. The contrasted genomic (gene loss and diversity) and genetic (recombination and Linkage Disequilibrium) plasticity reported for the B>A>D (from the most plastic to more stable subgenomes) have been tentatively explained so far through hypotheses relying on diploid progenitors differences. However, these hypotheses rely on the assumption of a constant and similar evolutionary rate between subgenomes.

In contrast, we propose an alternative scenario where such particular conservation of the B subgenome with *A. speltoides* at the sequence level is the consequence of a differential evolutionary plasticity of the B subgenome compared to the other A and D subgenomes in response to polyploidization events. We then propose an evolutionary scenario where the modern bread wheat genome has been shaped through a first neotetraploidization event (1.5 mya) leading to a subgenome dominance where the A subgenome was dominant and the B subgenome sensitive. The second neohexaploidization event (0.1 mya) led to a supra-dominance where the tetraploid became sensitive (subgenomes A and B) and the D subgenome dominant (*i.e.* pivotal). Following our scenario, wheat subgenome architecture has then been polyploidy-driven (*i.e.* pure post-polyploidization mechanism), where genome doubling is followed by genome portioning into dominant (stable) and sensitive (plastic) compartments, instead of entirely explained by pre-existing differences between founder (A, B and D) progenitors (*i.e.* pure pre-polyploidization mechanism). We propose in the current study that subgenome dominance is not only active in paleopolyploids but also in modern neopolyploids such as bread wheat. In our bread wheat evolutionary model, we propose that the differences in genomic structure and genetical landscape between subgenomes resulted directly from differential modes of evolution between dominant and sensitive blocks, where B subgenome acts as a sensitive in contrast to A as dominant in response to the neotetraploidization event 2.5 mya and D as dominant and A-D subgenomes as sensitive in regard to the neohexaploidization event 10,000 years ago.

This scenario is in agreement with the differential plasticity (at the gene content and diversity levels) reported in the current analysis in regards to gene loss and SNP/gene, higher in the B subgenome compared to the two others. Moreover, our wheat subgenome dominance model provides highlight into earlier studies of genome rearrangement (Zhang *et al.*, 2013) or gene loss (Ozkan *et al.*, 2001ab) in newly synthesized polyploids as well as the reported biased erosion of genetic diversity during domestication (Cavanagh *et al.*, 2013). The B genome, sensitive subgenome in regard to the paleotetraploidization, is associated to a pronounced genome plasticity related to the number of

paralogous loci (reported higher than in the A and D genomes in Akhunov *et al.*, 2003), translocation (Kota and Dvorak 1988), and other large-scale structural changes (Zhang *et al.*, 2013). Zhang *et al.*, (2013) using a set of 16 independently synthesized allohexaploid wheat lines produced >1000 individual plants at different selfing generations (S1>S20) after allohexaploidization that have been karyotyped. Of the three consistent subgenomes, B (sensitive) showed the highest frequency of chromosome rearrangements, followed by A (dominant); and the D (supra-dominant, referenced also as ‘pivotal’ in Zhang *et al.*, 2013) subgenome was largely observed with the higher stability. Overall, these data suggest that DNA rearrangements at the chromosome as well as gene levels, leading to the reported diploidization-driven subgenome dominance, occurred immediately or within a few generations following polyploidization.

Both scenarios aiming at explaining the structural asymmetry characterized between the wheat subgenomes need to be now considered, *i.e.* either a pre-polyploidization mechanism where the evolutionary history of the cross-pollinator *A. speltoides* progenitor (B donor) had a more diverse genome than self-pollinators *T. urartu/A. tauschii* (A/D donors) or a post-polyploidization mechanism with an accelerated plasticity of the B subgenome in contrast to the homoeologous counterparts.

SYNTENOME OF COMPLEX POLYPLOID SPECIES CAN BE USED AS A GUIDE FOR TRAIT DISSECTION.

High resolution and large-scale comparative genomics studies offer a tremendous set of gene-based markers that can be used directly as founder resource for genome mapping (physical or genetic) and ultimately trait dissection. In the present work, we deliver a wheat syntenome consisting in 17,317 ordered COS genes in wheat and the associated sequence variations (SNP and Indels). Overall, the sequence capture of the COS markers associated with NGS methods appear to be a powerful technique for the large-scale discovery of gene-associated SNPs in any plant of interest. Moreover, such sequencing efforts within a large scale intra-specific background will provide a complete set of putative causal SNPs, PAVs, CNVs, as functional diagnostic markers in the near future. Associated with public RFLP, AFLP, STS, SSR, DaRTs markers, COS-SNP characterization and mapping effort in wheat allowed us to provide here the most complete genetic map consisting in 7,520 molecular markers. Such resolution including synteny-based COS markers allows to refine metaQTL intervals and immediately benefits from the COS markers to access robust links with sequence genomes and precise orthologous candidate genes as we precisely illustrated in Quraishi *et al.*, (2011ab) and Dibari *et al.*, (2012). Moreover the wheat syntenome has been also used as a matrix for physical map construction (Lucas *et al.*, 2013) prior QTL dissection.

The proposed impact of polyploidization-based subgenome partitioning on contrasted gene content and diversity in dominant and sensitive blocks may need to reconsider in agronomic traits dissection. First of all, the sensitive chromosomal compartments that appeared most plastic (in term of polymorphism as well as recombination pattern) may be more accessible to QTL cloning. It would be then interesting to investigate the differential efficiency in trait or gene cloning observed in both D and S compartments. Moreover, the difference in subgenome stability (with higher plasticity observed

for B>A>D) may also lead to the hypothesis that homoeologous groups may support different type of agronomic traits. Feldman *et al.*, (2012ab) reported that genome A was found to control morphological traits while genome B in allotetraploid to control reactions to biotic and abiotic factors. This difference in trait natures is also consistent with the subgenome dominance hypothesis where the sensitive compartment (B subgenome in wheat as defined in the current analysis) has been suggested to control adaptative traits in contrast to more stable processes driven by the dominant counterpart (A and D in hexaploid wheat), Abrouk *et al.*, 2012. Moreover, QTL partitioning following polyploidy has been also suggested recently with 21% homoeologous fiber quality QTLs in cotton (Rong *et al.*, 2007) and 23% homoeologous fruit quality QTLs in strawberry (Lerceteau-Köhler *et al.*, 2012) characterized, then suggesting that the majority of QTL are no longer maintained on the duplicated blocks as putatively a direct consequence of the diploidization mechanism. To what extent the hexaploid wheat adaptation (in particular regarding adaptation in responses to biotic and abiotic stresses) is possibly partitioned in the genome, between the currently defined dominant and sensitive chromosomal compartments, remains an open question that still needs to be addressed in the future.

EXPERIMENTAL PROCEDURE

The experimental procedures may be found in the online version of this article in detailing (i) *Wheat syntenome protocol* (COS identification; CAR identification; Computational gene order in wheat; D-S blocks identification), (ii) *Wheat COS sequencing protocol* (COS-primer design; COS-SNP sequencing and clustering); (iii) *SNP genotyping protocol* (SSCP, Illumina, Size polymorphism, Sequencing, Kaspar, LNA® Dual-Labeled Fluorogenic Probes), (vi) *Wheat comprehensive consensus genetic map protocol* (Selection of public reference genetic maps; Construction of WCGM2013).

FUNDING AND ACKNOWLEDGEMENT

This work has been supported by grants from INRA ('Génétique et Amélioration des Plantes, *ref*: 'Appel d'Offre AIP Bioressources'), the Auvergne region ('Pôle de compétitivité : Céréales Vallée' *ref*: program 'Semences de demain'), the Agence Nationale de la Recherche (Programs ANRjc-PaleoCereal, *ref*: ANR-09-JCJC-0058-01 and program ANR Blanc-PAGE, *ref*: ANR-2011-BSV6-00801), and the European program 7th Framework (program 'TRITICEAE GENOME', *ref*: FP7-212019). The author would like to thank Emmanuelle Legendijk, Nils Stein, Laura Rossini for the coordination of the COS sequencing and Grain Fiber Content QTL characterization initiatives as well as Sébastien Faure for providing the wheat lines in the frame of the TRITICEAE GENOME project.

SUPPORTING INFORMATION

Tables S1-3 as well as Figures S1-13 may be found in the online version of this article.

Table S1: Wheat syntenome raw data.

Table S2: List of dominant (D) and sensitive (S) chromosomal blocks in wheat.

Table S3: Wheat Composite Genetic Map (WCGM 2013).

Figure S1: Wheat syntenome characterization flow chart.

Figure S2: Wheat duplicated genes (homoeologs & paralogs) characterization.

Figure S3: Grass dominant and sensitive compartments following ancestral WGD.

Figure S4: Strategy for COS-SNP marker development and exploitation.

Figure S5: COS-SNP transferability in Grasses.

Figure S6: *De novo* COS sequencing in wheat.

Figure S7: HomoeoSNP vs. SNP calling through sequence coverage criterion.

Figure S8: Characterization of wheat homoeologs (A, B and D hamoeoalleles).

Figure S9: Identification of PAVs and CNVs.

Figure S10: COS-SNP and InDels typology in wheat.

Figure S11: Genome-wide distribution of the COS-SNP diversity in wheat.

Figure S12: False COS-SNP origins.

Figure S13: Validation of COS computed gene order.

Figure S14: Detailed characteristics of the WCGM 2013.

Figure S15: Wheat Synteny Viewer characterization flow chart.

REFERENCES

Abrouk, M., Zhang R., Murat, F., Li A., Pont, C., Mao, L. and Salse, J. (2012) Grass microRNA Gene Paleohistory Unveils New Insights into Gene Dosage Balance in Subgenome Partitioning after Whole Genome Duplication. *Plant Cell* **24**(5), 1776-92.

Akhunov, E.D., Akhunova, A.R., Linkiewicz, A.M. *et al.* (2003) Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates. *Proc Natl Acad Sci U S A.* **100**(19), 10836-41.

Akhunov, E.D., Akhunova, A.R., Anderson, O.D. *et al.* (2010) Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics* **11**, 702.

Allen, A.M., Barker, G.L., Berry, S.T. *et al.* (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). *Plant Biotechnol J.* **9**(9), 1086-99.

Allen, A.M., Barker, G.L., Wilkinson, P. *et al.* (2013) Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Plant Biotechnol J.* **11**(3), 279-95.

- Bekaert, M., Edger, P.P., Pires, J.C. and Conant, G.C.** (2011) Two-phase resolution of polyploidy in the Arabidopsis metabolic network gives rise to relative and absolute dosage constraints. *Plant Cell* **23**(5), 1719-28.
- Berkman, P.J., Lai, K., Lorenc, M.T. and Edwards, D.** (2012) Next-generation sequencing applications for wheat crop improvement. *Am J Bot.* **99**(2), 365-71.
- Berkman, P.J., Visendi, P., Lee, H.C. et al.** (2013) Dispersion and domestication shaped the genome of bread wheat. *Plant Biotechnol J.* **11**(5), 564-71
- Brenchley, R., Spannagl, M., Pfeifer, M. et al.** (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**(7426), 705-10.
- Cavanagh, C.R., Chao, S., Wang, S. et al.** (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc Natl Acad Sci U S A.* **110**(20), 8057-62.
- Chao, S., Zhang, W., Akhunov, E., Sherman, J., Ma, Y., Luo, M. and Dubcovsky, J.** (2009) Analysis of gene-derived SNP marker polymorphism in wheat (*Triticum aestivum* L.). *Molecular Breeding* **23**, 23-33.
- Coulondre, C., Miller, J.H., Farabaugh, P.J. and Gilbert, W.** (1978) Molecular basis of base substitution hotspots in Escherichia coli. *Nature* **274**(5673), 775-80.
- Devos, K.M., Dubcovsky, J., Dvorák, J., Chinoy, C.N. and Gale, M.D.** (1995) Structural evolution of wheat chromosomes 4A, 5A and 7B and its impact on recombination. *Theor Appl Genet* **91**, 282–288.
- Dibari, B., Murat, F., Chosson, A. et al.** (2012) Deciphering the genomic structure, function and evolution of carotenogenesis related phytoene synthases in grasses. *BMC Genomics* **13**, 221.
- Doyle, J.J., Flagel, L.E., Paterson, A.H., Rapp, R.A., Soltis, D.E., Soltis, P.S. and Wendel, J.F.** (2008) Evolutionary genetics of genome merger and doubling in plants. *Annu Rev Genet.* **42**, 443-61.
- Duran, C., Edwards, D. and Batley, J.** (2009) Genetic maps and the use of synteny. *Methods Mol Biol.* **513**, 41-55.
- Dvorak, J., Zhang, H.B., Kota, R.S. and Lassner, M.** (1989) Organization and evolution of the 5S ribosomal RNA gene family in wheat and related species. *Genome* **32**, 1003–1016.
- Dvorak, J. and Zhang, H.B.** (1990) Variation in repeated nucleotide sequences sheds light on the phylogeny of the wheat B and G genomes. *Proc Natl Acad Sci U S A.* **87**(24), 9640-4.
- Dvorak, J. and Akhunov, E.D.** (2005) Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploid evolution in the *Aegilops-Triticum* alliance. *Genetics* **171**(1), 323-32.
- Dvorak, J., Akhunov, E.D., Akhunov, A.R., Deal, K.R. and Luo, M.C.** (2006) Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. *Mol Biol Evol.* **23**(7), 1386-96.

- Edger, P.P. and Pires, J.C.** (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res.* **17**(5), 699-717.
- Feldman M** (1966a) Identification of unpaired chromosomes in F1 hybrids involving *Triticum aestivum* and *T. timopheevii*. *Can J Genet Cytol.* **8**, 144–151.
- Feldman M** (1966b) The mechanism regulating pairing in *Triticum timopheevii*. *Wheat Inf Serv.* **21**, 1–2.
- Feldman, M., Lupton, F.G.H. and Miller, T.E.** (1995) Wheats. In *Evolution of crop plants*. 2nd edition. J. Smartt and N. W. Simmonds, eds. Pp. 184–192. Harlow: Longman Scientific & Technical.
- Feldman, M., Levy, A.A., Fahima, T. and Korol, A.** (2012a) Genomic asymmetry in allopolyploid plants: wheat as a model. *J Exp Bot.* **63**(14), 5045-59.
- Feldman, M. and Levy, A.A.** (2012b) Genome evolution due to allopolyploidization in wheat. *Genetics.* **192**(3), 763-74
- Freeling, M., Woodhouse, M.R., Subramaniam, S., Turco, G., Lisch, D. and Schnable, J.C.** (2012) Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Curr Opin Plant Biol.* **15**(2), 131-9.
- Furuta, Y., Nishikawa, K. and Yamaguchi, S.** (1986) Nuclear DNA content in diploid wheat and its relatives in relation to the phylogeny of tetraploid wheat. *Japanese Journal of Genetics* **61**, 97-105.
- Gill, B.S. and Chen, P.D.** (1987) Role of cytoplasm specific introgression in the evolution of the polyploid wheats. *Proc Natl Acad Sci U S A.* **84**, 6800–6804.
- Hutchinson, J., Miller, T.E., Jahier, J. and Shepherd, K.W.** (1982) Comparison of the chromosomes of *Triticum timopheevii* with related wheats using the techniques of C-banding and in situ hybridization. *Theor Appl Genet* **64**, 31–40.
- International Rice Genome Sequencing Project** (2005) The map-based sequence of the rice genome. *Nature* **436**, 793-800.
- International Brachypodium Initiative** (2010) Genome Sequencing and Analysis of the Model grass *Brachypodium distachyon*. *Nature* **463**(7282), 763-768.
- Jia, J., Zhao, S., Kong, X. et al.** (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**(7443), 91-5.
- Jiang, J. and Gill, B.S.** (1994) Different species-specific chromosome translocations in *Triticum timopheevii* and *T. turgidum* support the diphyletic origin of polyploid wheats. *Chromosome Res.* **2**, 59-64.
- Kilian, B, Ozkan, H, Deusch, O, Effgen, S, Brandolini, A, Kohl, J, Martin, W, Salamini, F** (2007) Independent wheat B and G genome origins in outcrossing *Aegilops* progenitor haplotypes. *Mol Biol Evol.* **24**, 217-227.
- Kota, R.S. and Dvorak, J.** (1988) Genomic Instability in Wheat Induced by Chromosome 6b(s) of *Triticum Speltoides*. *Genetics* **120**(4), 1085-94.

- Lai, K., Duran, C., Berkman, P.J. et al.** (2012) Single nucleotide polymorphism discovery from wheat next-generation sequence data. *Plant Biotechnol J.* **10**(6), 743-749.
- Lerceteau-Köhler, E., Moing, A., Guérin, G., Renaud, C., Petit, A., Rothan, C. and Denoyes, B.** (2012) Genetic dissection of fruit quality traits in the octoploid cultivated strawberry highlights the role of homoeo-QTL in their control. *Theor Appl Genet.* **124**(6), 1059–1077.
- Ling, H.Q., Zhao, S., Liu, D. et al.** (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* **496**(7443), 87-90.
- Liu, Y.G. and Tsunewaki, K.** (1991) Restriction fragment length polymorphism (RFLP) analysis in wheat. II. Linkage maps of the RFLP sites in common wheat. *Jpn J Genet.* **66**(5), 617-33.
- Lucas, S.J., Akpınar, B.A., Kantar, M. et al.** (2013) Physical Mapping Integrated with Syntenic Analysis to Characterize the Gene Space of the Long Arm of Wheat Chromosome 1A. *PLoS One* **8**(4), e59542.
- Luo, M.C., Gu, Y.Q., You, F.M. et al.** (2013) A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. *Proc Natl Acad Sci U S A.* **110**(19), 7940-5.
- Maestra, B. and Naranjo, T.** (1999) Structural chromosome differentiation between *Triticum timopheevii* and *T. turgidum* and *T. aestivum*. *Theor Appl Genet.* **98**, 744-750.
- Murat, F., Xu, J.H., Tannier, E., Abrouk, M., Guilhot, N., Pont, C., Messing, J. and Salse, J.** (2010) Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. *Genome Res.* **20**(11), 1545-57.
- Naranjo, T., Roca, A., Goicoechea, P.G. and Giráldez, R.** (1987) Arm homoeology of wheat and rye chromosomes. *Genome* **29**, 873–882.
- Naranjo T** (1990) Chromosome structure of durum wheat. *Theor Appl Genet.* **79**, 397–400.
- Ozkan, H., Levy, A.A. and Feldman, M.** (2001a) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* **13**(8), 1735-47.
- Ozkan, H. and Feldman, M.** (2001b) Genotypic variation in tetraploid wheat affecting homoeologous pairing in hybrids with *Aegilops peregrina*. *Genome* **44**(6), 1000-6.
- Paterson, A.H., Bowers, J.E., Bruggmann, R. et al.** (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**(7229), 551-6.
- Pont, C., Murat, F., Confolent, C., Balzergue, S. and Salse, J.** (2011) RNA-seq in grain unveils fate of neo- and paleopolyploidization events in bread wheat (*Triticum aestivum* L.). *Genome Biol.* **12**(12), R119.
- Quraishi Masood, M., Abrouk, M. and Bolot, S. et al.** (2009) Genomics in cereals: From genome-wide conserved orthologous set (COS) sequences to candidate genes for trait dissection. *Functional & Integrative Genomics* **9**(4), 473-84.

- Quraishi, U.M., Murat, F., Abrouk, M. et al.** (2011a) Combined meta-genomics analyses unravel candidate genes for the grain dietary fiber content in bread wheat (*Triticum aestivum* L.). *Functional & Integrative Genomics* **11**(1), 71-83.
- Quraishi, U.M., Abrouk, M., Murat, F. et al.** (2011b) Cross-genome map based dissection of a nitrogen use efficiency ortho-metaQTL in bread wheat unravels concerted cereal genome evolution. *Plant J.* **65**(5), 745-56.
- Ren, J., Sun, D., Chen, L. et al.** (2013) Genetic diversity revealed by single nucleotide polymorphism markers in a worldwide germplasm collection of durum wheat. *Int J Mol Sci.* **14**(4), 7061-88.
- Roder, M.S., Korzun, V., Gill, B.S. and Ganal, M.W.** (1998a) The physical mapping of microsatellite markers in wheat. *Genome*, **41**, 278-283
- Roder, M.S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M.H., Leroy, P. and Ganal, M.W.** (1998b) A microsatellite map of wheat. *Genetics* **149**, 2007-2023.
- Rong, J., Feltus, F.A., Waghmare, V.N. et al.** (2007) Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. *Genetics.* **176**(4): 2577-88.
- Saintenac, C., Jiang, D. and Akhunov, E.D.** (2011) Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biol.*, **12**(9), R88.
- Salina, E.A., Lim, K.Y., Badaeva, E.D., Shcherban, A.B., Adonina, I.G., Amosova, A.V., Samatadze, T.E., Vatolina, T.Y., Zoshchuk, S.A. and Leitch, A.R.** (2006) Phylogenetic reconstruction of Aegilops section Sitopsis and the evolution of tandem repeats in the diploids and derived wheat polyploids. *Genome* **49**(8):1023-35.
- Salse, J., Chagué, V., Bolot, S. et al.** (2008) New insights into the origin of the B genome of hexaploid wheat: evolutionary relationships at the SPA genomic region with the S genome of the diploid relative Aegilops speltoides. *BMC Genomics* **25**, 9:555.
- Salse, J.** (2012) *In silico* archeogenomics unveils modern plant genome organization, regulation and evolution. *Curr Opin Plant Biol.* **15**(2), 122-30.
- Schnable, P.S., Ware, D., Fulton, R.S. et al.** (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**(5956), 1112-5.
- Schnable, J.C., Freeling, M. and Lyons, E.** (2012a) Genome-wide analysis of syntenic gene deletion in the grasses. *Genome Biol Evol.* **4**(3), 265-77.
- Schnable, J.C., Wang, X., Pires, J.C. and Freeling, M.** (2012b) Escape from preferential retention following repeated whole genome duplications in plants. *Front Plant Sci.* **3**, 94.
- Terachi, T., Ogiwara, Y. and Tsunewaki, K.** (1990) The molecular basis of genetic diversity among cytoplasm of Triticum and Aegilops. 7. Restriction endonuclease analysis of mitochondrial DNA from polyploid wheats and their ancestral species. *Theor Appl Genet.* **80**, 366–373.

- Thomas, B.C., Pedersen, B. and Freeling, M.** (2006) Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res.* 16(7), 934-46.
- Thuillet, A.C., Bataillon, T., Poirier, S., Santoni, S. and David, J.L.** (2005) Estimation of long-term effective population sizes through the history of durum wheat using microsatellite data. *Genetics* 169(3), 1589-99.
- Trebbi, D., Maccaferri, M., de Heer, P., Sørensen, A., Giuliani, S., Salvi, S., Sanguineti, M.C., Massi, A., van der Vossen, E.A. and Tuberosa, R.** (2011) High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.). *Theor Appl Genet.* 123(4), 555-69.
- Trick, M., Adamski, N.M., Mugford, S.G., Jiang, C.C., Febrer, M. and Uauy, C.** (2012) Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol.* 26, 12:14.
- Wang, X., Shi, X., Hao, B., Ge, S. and Luo, J.** (2005) Duplication and DNA segmental loss in the rice genome: implications for diploidization. *New Phytol.* 165(3), 937-46.
- Wang, J., Luo, M.C., Chen, Z., You, F.M., Wei, Y., Zheng, Y. and Dvorak, J.** (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 Phytol.* 198(3), 925-37.
- Wicker T, Mayer KF, Gundlach H et al.** (2011) Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. *Plant Cell* 23(5), 1706-18.
- Winfield, M.O., Wilkinson, P.A., Allen, A.M. et al.** (2012) Targeted re-sequencing of the allohexaploid wheat exome. *Plant Biotechnol J.* 10(6), 733-42.
- Woodhouse, M.R., Schnable, J.C., Pedersen, B.S., Lyons, E., Lisch, D., Subramaniam, S. and Freeling, M.** (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. *PLoS Biol.* 8(6), e1000409.
- You, F.M., Huo, N., Deal, K.R., Gu, Y.Q., Luo, M.C., McGuire, P.E., Dvorak, J. and Anderson, O.D.** (2011) Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence. *BMC Genomics* 12, 59.
- Zhang, H., Bian, Y., Gou, X., Zhu, B., Xu, C., Qi, B., Li, N., Rustgi, S., Zhou, H., Han, F., Jiang, J., von Wettstein, D. and Liu, B.** (2013) Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. *Proc Natl Acad Sci U S A.* 110(9), 3447-52.
- Zohary, D. and Feldman, M.** (1962) Hybridization between amphidiploids and the evolution of polyploids in the wheat (*Aegilops-Triticum*) group. *Evolution* 16, 44-61.

FIGURE LEGENDS

Figure 1: Wheat syntenome circles.

(A) The center of the circle illustrates the wheat chromosomes origin from a grass ancestor structured in 7 protochromosomes followed by a paleotetraploidization to reach a 12 chromosomes of known dominance (D) and sensitivity (S) blocks. Integration of the seven bread wheat chromosome groups (wheat consensus circle, ‘wheat cons’) with the rice (12 chromosomes, ‘R’ circle), sorghum (10 chromosomes, ‘S’ circle) and *Brachypodium* (5 chromosomes, ‘B’ circle) chromosomes is illustrated as inner circles with 17,317 COS (black connecting lines between inner circles), 5,234 COS markers selected for sequencing (grey connecting lines between inner circles) and 9,969 SNPs (shown as orange distribution bars on the ‘wheat cons’ circle) according to their orthologous positions on rice, sorghum and *Brachypodium* chromosomes. The three external concentric circles (referenced ‘A’, ‘B’ and ‘D’ circles) illustrate the 21 bread wheat chromosomes based on the BIN locations (‘BIN’ circle), the genetic map (‘cM’ circle where 375 mapped COS-SNPs are illustrated as red bars) and 6,423 homoeologous genes (green curves). The recombination density heatmap is illustrated as red (for BIN where physical size is higher than cM size) or blue (for BIN where physical size is lower than cM size) BIN intervals. The characterized dominance and sensitivity of the modern wheat genomic compartments are mentioned on the external circle. (B) Screen capture of the PlantSytenyViewer web tool [<http://urgi.versailles.inra.fr/syteny-wheat>] visualizing the syteny between wheat, *Brachypodium*, rice, maize, sorghum and providing the COS marker information (ID, primer pairs, SNP in wheat).

Figure 2: Wheat paleo- and neosubgenomes.

(A) Illustration of the dominant (blue) and sensitive ancestral chromosome derived from a $n=7$ AGK that has been duplicated (WGD) to reach a $n=12$ intermediate. The Triticeae ancestor (T1-T7) derived from five NCF that took place between the 12 ancestral chromosomes (left). The modern hexaploid bread wheat derived from the polyploidization between 3 diploid progenitors (A, B and D sub genomes) structured in seven protochromosomes (right). Distribution and associated statistical significance are illustrated as box-plots (considering individually the 7 wheat chromosomal groups) for the number of orthologs (B), number of COS (C), SNP (aSNP/500COS ratio from 9,969 *in silico* COS-SNPs and mapped COS rate from 375 mapped COS-SNPs) data (D), cM/marker ratio (E), number of C-band (F) between ancestral D and S blocks (left) as well as modern A, B and D subgenomes (right) in wheat. Pair T-test p values are referenced between paleo-D and paleo-S blocks (at the left, with A11-A12 pair excluded in parenthesis highlighted with *), and between A-B and B-D subgenomes (at the right, with binomial ($n=2$) p-value highlighted with **).

Figure 3: Wheat evolutionary model.

(A) Evolutionary model of the modern wheat genome from a $n=7$ (AGK), 12 (duplicated AGK), 7 (Triticeae) ancestors illustrated as dominant (blue bars) and sensitive (red bars) blocks. The subgenome dominance took place in wheat paleohistory when the diploid progenitors hybridized.

Genome A (blue outlines) was dominant and B (red outlines) sensitive after the first polyploidization event 1.5 mya. The tetraploid (A-B genome) was sensitive and D dominant after the second polyploidization event 0.1 mya. The modern 21 bread wheat chromosomes are then illustrated as dominant, sensitive, supra-dominant and supra-sensitive fragments according to the colored scale at the right defining distinct degree of genomic stability and plasticity. **(B)** Distribution of gene number, ortholog number, chromosome length, transposable element observed between paleo-S and paleo-D fragments in rice (left). Distribution of COS number, SNP/COS ratio, cM/marker ratio, number of transposed genes, C-band number observed between chromosomes 1A (sensitive), 1B (supra-sensitive), 1D (sensitive), 3A (dominant), 3B (dominant), 3D (supra-dominant), right.

TABLE

Table 1: Wheat syntenome and derived COS-SNP makers.

| Ch r. | Gene repertoire | | | | COS genes | | | | COS-SNP markers | | |
|-------|-------------------------------|------------------------------|------------------------------|------------------------------|-------------|-------------|------------------------|----------------------|-------------------------|---------------|-----------------------------|
| | Genes /group | gA | gB | gD | Proto genes | COS primers | COS-seq ¹ | COS-SNP ² | COS Mapped ³ | COS conserved | COS transposed ³ |
| 1 | 1945 | 655 | 600 | 690 | 2001 | 698 | 634 (2792/42/64) | 1492 (299) | 76 (36/28/12) | 72 (94%) | 4 (0/4/0) |
| 2 | 2964 | 882 | 942 | 1028 | 2963 | 1020 | 882 (4139/41/125) | 2077 (386) | 64 (27/33/4) | 57 (89%) | 7 (6/1/0) |
| 3 | 2830 | 755 | 880 | 993 | 2733 | 924 | 755 (3298/40/105) | 1507 (323) | 82 (25/47/10) | 77 (93%) | 5 (1/3/1) |
| 4 | 2553 | 770 | 809 | 889 | 3019 | 776 | 685 (3035/39/79) | 861 (243) | 18 (7/11/0) | 12 (66%) | 6 (4/2/0) |
| 5 | 2519 | 788 | 804 | 878 | 2023 | 941 | 873 (4012/59/103) | 1339 (352) | 28 (5/22/1) | 15 (53%) | 13 (1/12/0) |
| 6 | 2156 | 625 | 681 | 763 | 2097 | 717 | 625 (2769/42/69) | 1147 (264) | 51 (22/22/7) | 49 (96%) | 2 (0/1/1) |
| 7 | 2423 | 780 | 756 | 850 | 2481 | 957 | 780 (3418/41/99) | 1546 (330) | 56 (25/21/10) | 53 (94%) | 3 (2/1/0) |
| Tot. | 17390 (48120) ⁴ | 5827 (15996) ⁴ | 5472 (15244) ⁴ | 6091 (16880) ⁴ | 17317 | 6033 | 5234 (4286/304/644) | 9969 (2197) | 375 (147/184/44) | 335 (89%) | 40 (14/24/2) |

¹: in parenthesis: number of NSV, PAV and CNV.

²: in parenthesis: number of COS corresponding to the number of SNP.

³: in parenthesis: number of mapped and transposed COS respectively in A, B and D subgenomes.

⁴: in parenthesis: total number of wheat homoeologous genes from Brenchley *et al.*, 2012.

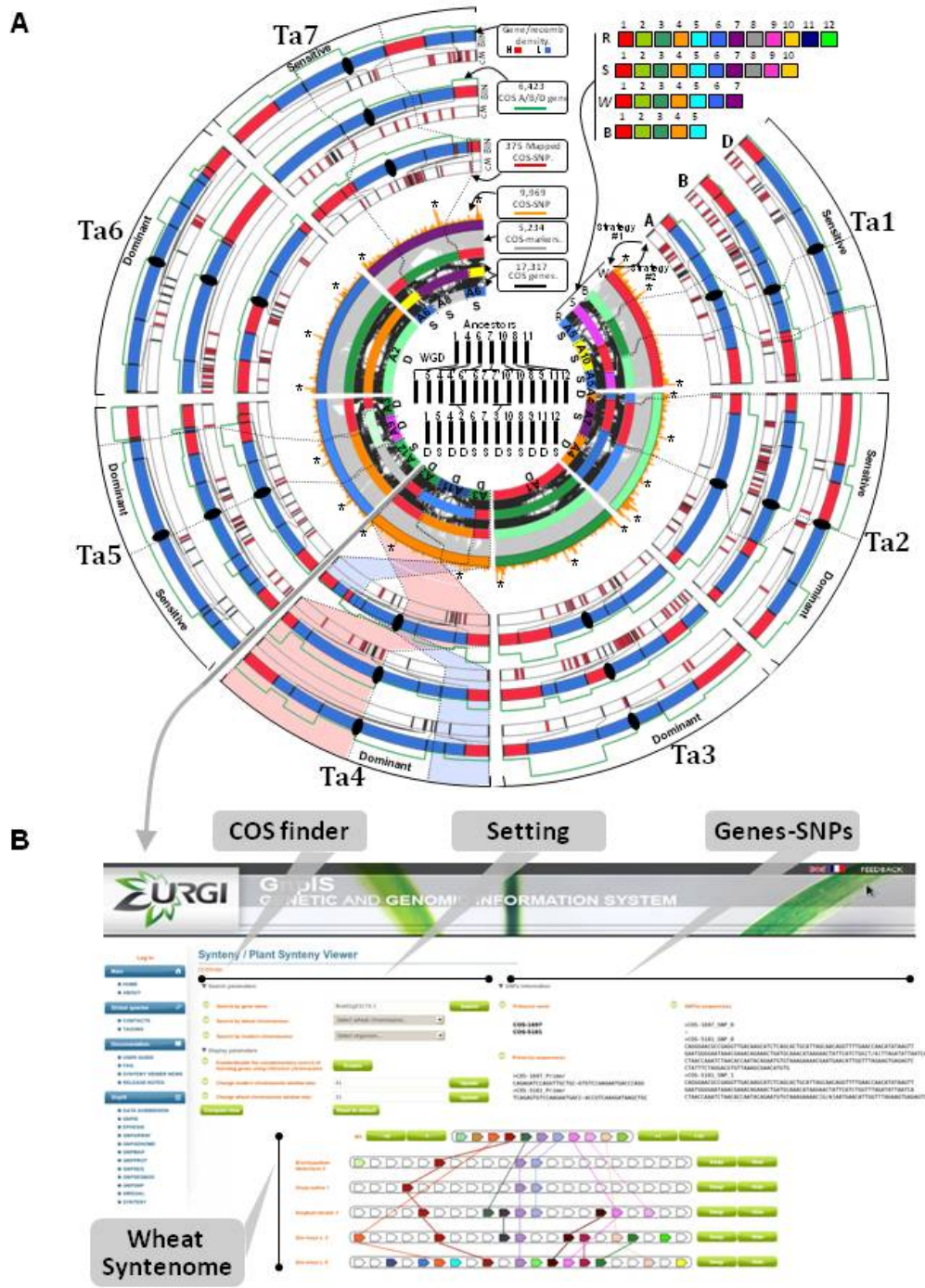


Figure 1: Wheat syntenome circles.

(A) The center of the circle illustrates the wheat chromosomes origin from a grass ancestor structured in 7 protochromosomes followed by a paleotetraploidization to reach a 12 chromosomes of known dominance (D) and sensitivity (S) blocks. Integration of the seven bread wheat chromosome groups (wheat consensus circle, 'wheat cons') with the rice (12 chromosomes, 'R' circle), sorghum (10 chromosomes, 'S' circle) and *Brachypodium* (5 chromosomes, 'B' circle) chromosomes is illustrated as inner circles with 17,317 COS (black connecting lines between inner circles), 5,234 COS markers selected for sequencing (grey connecting lines between inner circles) and 9,969 SNPs (shown as orange distribution bars on the 'wheat cons' circle) according to their orthologous positions on rice, sorghum and *Brachypodium* chromosomes. The three external concentric circles (referenced 'A', 'B' and 'D' circles) illustrate the 21 bread wheat chromosomes based on the BIN locations ('BIN' circle), the genetic map ('cM' circle where 375 mapped COS-SNPs are illustrated as red bars) and 6,423 homoeologous genes (green curves). The recombination density heatmap is illustrated as red (for BIN where physical size is higher than cM size) or blue (for BIN where physical size is lower than cM size) BIN intervals. The characterized dominance and sensitivity of the modern wheat genomic compartments are mentioned on the external circle. **(B)** Screen capture of the PlantSyntenyViewer web tool [<http://urgi.versailles.inra.fr/synteny-wheat>] visualizing the synteny between wheat, *Brachypodium*, rice, maize, sorghum and providing the COS marker information (ID, primer pairs, SNP in wheat).

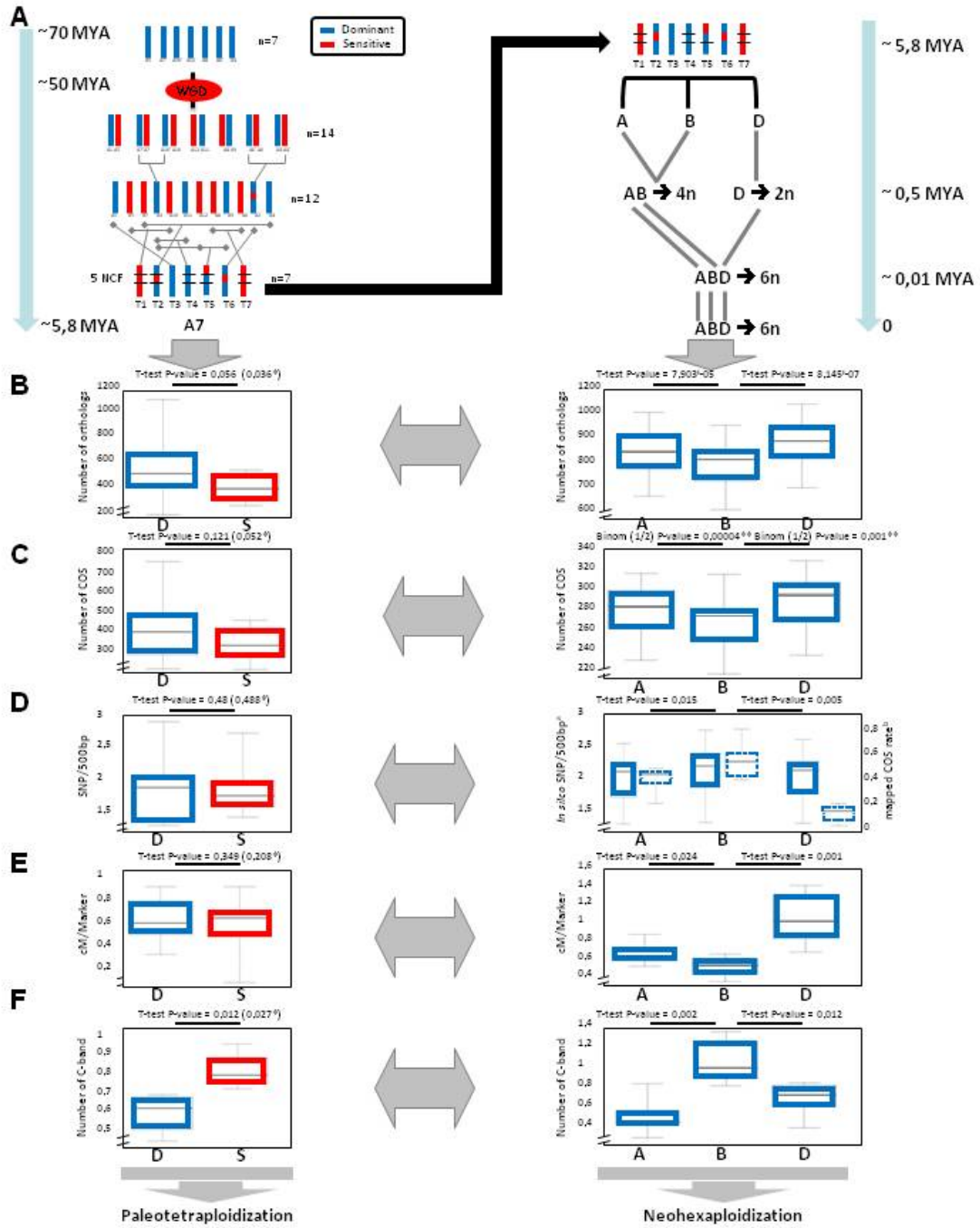


Figure 2

Figure 2: Wheat paleo- and neosubgenomes.

(A) Illustration of the dominant (blue) and sensitive ancestral chromosome derived from a $n=7$ AGK that has been duplicated (WGD) to reach a $n=12$ intermediate. The Triticeae ancestor (T1-T7) derived from five NCF that took place between the 12 ancestral chromosomes (left). The modern hexaploid bread wheat derived from the polyploidization between 3 diploid progenitors (A, B and D sub genomes) structured in seven protochromosomes (right). Distribution and associated statistical significance are illustrated as box-plots (considering individually the 7 wheat chromosomal groups) for the number of orthologs (B), number of COS (C), SNP (#SNP/500COS ratio from 9,969 *in silico* COS-SNPs and mapped COS rate from 375 mapped COS-SNPs) data (D), cM/marker ratio (E), number of C-band (F) between ancestral D and S blocks (left) as well as modern A, B and D subgenomes (right) in wheat. Pair T-test p values are referenced between paleo-D and paleo-S blocks (at the left, with A11-A12 pair excluded in parenthesis highlighted with *), and between A-B and B-D subgenomes (at the right, with binomial ($n=2$) p-value highlighted with **).

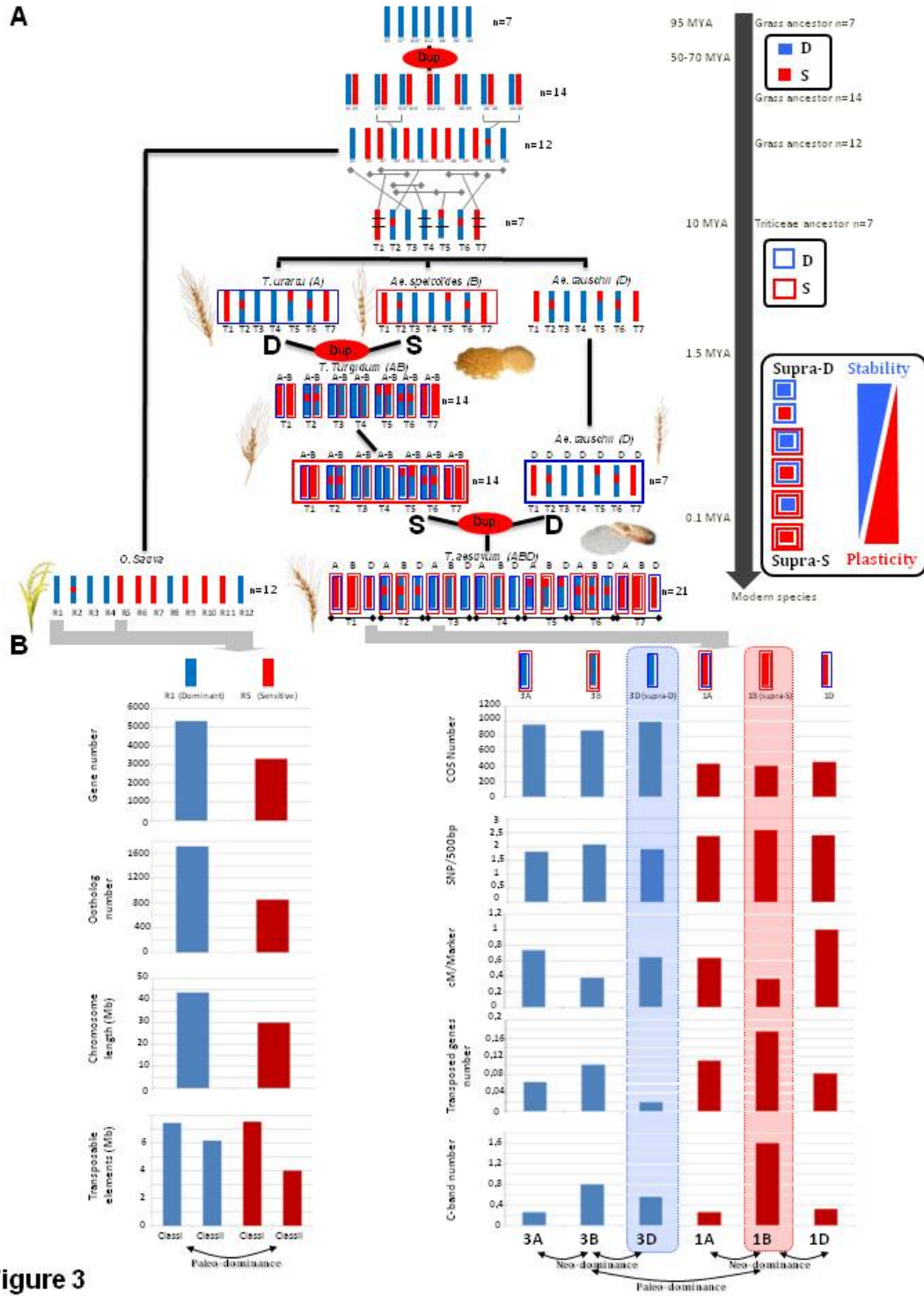


Figure 3: Wheat evolutionary model.

(A) Evolutionary model of the modern wheat genome from a $n=7$ (AGK), 12 (duplicated AGK), 7 (Triticeae) ancestors illustrated as dominant (blue bars) and sensitive (red bars) blocks. The subgenome dominance took place in wheat paleohistory when the diploid progenitors hybridized. Genome A (blue outlines) was dominant and B (red outlines) sensitive after the first polyploidization event 1.5 mya. The tetraploid (A-B genome) was sensitive and D dominant after the second polyploidization event 0.1 mya. The modern 21 bread wheat chromosomes are then illustrated as dominant, sensitive, supra-dominant and supra-sensitive fragments according to the colored scale at the right defining distinct degree of genomic stability and plasticity. **(B)** Distribution of gene number, ortholog number, chromosome length, transposable element observed between paleo-S and paleo-D fragments in rice (left). Distribution of COS number, SNP/COS ratio, cM/marker ratio, number of transposed genes, C-band number observed between chromosomes 1A (sensitive), 1B (supra-sensitive), 1D (sensitive), 3A (dominant), 3B (dominant), 3D (supra-dominant), right.