



UNIVERSITÉ EVRY VAL D'ESSONNE

**Ecole Doctorale - Des Génomes Aux Organismes
(GAO)**



**Laboratoire : UNITE DE RECHERCHE EN GENOMIQUE VEGETALE
Equipe de recherche : Organisation et évolution des génomes
de plantes (OEGP)**

THÈSE

Présentée et soutenue publiquement le 16 juin 2014

Pour l'obtention du grade de

Docteur de l'Université d'Evry Val d'Essonne

Discipline ou Spécialité: Bioinformatique et Génomique

Par :

Harry Belcram

**Organisation, évolution et fonctionnement des
gènes majeurs de domestication (*Q/q*) chez
les blés polyploïdes**

COMPOSITION DU JURY

Malika AINOUCHE	Professeur, Université Rennes	Rapporteur
Marie-Angèle GRANDBASTIEN	Directeur de Recherche, INRA Versailles	Rapporteur
Sarah SAMADI	Professeur, Muséum National d'Histoire Naturelle	Examineur
Pierre ROUMET	Directeur de Recherche, INRA Montpellier	Examineur
Jérôme SALSE	Directeur de Recherche, INRA Clermont-Ferrand	Examineur
Boulos CHALHOUB	Directeur de Recherche, INRA Versailles	Directeur de thèse



« *Viva la re-evolution* »

Remerciements

Voilà ! La fin du début ou le début de la fin...est proche

Quand on arrive aux remerciements c'est que l'on a passé les principales turbulences dans la rédaction et que l'on devrait voir une courte période d'accalmie profiler à l'horizon. Mais la réalité vous rappelle que ce n'est pas fini et qu'il y a encore du pain sur la planche.

Alors je profite de cette accalmie pour remercier ceux qui m'ont soutenu, si ce n'est supporté...

Je commencerai par remercier les membres de mon jury de thèse (Marie-Angèle Grandbastien, Malika Ainouche, Sarah Samadi, Pierre Roumet et Jérôme Salse) qui ont accepté de juger mon travail.

Mon directeur de thèse, Boulos Chalhoub alias Boubou d'avoir accepté le dur labeur de me diriger dans ma thèse. Après 15 années de travail en commun, je suis content de voir que nos débats restent toujours aussi animés.

Mes remerciements vont également à mes amours, Katia, Ambre et Mayleen, qui commençaient à sérieusement perdre patience.

A mes amis, pour lesquels, Je commence à comprendre les «Tu fait une thèse ? C'est Super ! » Avec un sourire « qui en dit long». C'est ça les amis, ils ne disent rien sur l'instant, mais sont toujours présents, y compris aux bons «moments». Vivement les soirées entre poteaux avec le Jej, Agnés, Hélène, Soph et Sof, Pitchoune, Ben, MissTiti et tous les petits bouts.

A mes neveux et nièce, Jean-Yves, François-Joseph, José-Luc, Joé, et mes filleuls Adeline et Christophe qui ont supporté leur oncle comme des supporters d'une équipe de Basket. Ma mère qui ne pourra pas être présente, et le reste de ma famille qui est trop loin.

Je remercie les thésards et les PostDoc qui sont travaillés dans notre équipe, ou à l'URGV (Mat-Dieu (Mathieu Charles), JJ (Jérémie Just), Mimi (Imen ex Mestiri), Bella « Reine du logiciel End-Note » (Smahane Chalabi), Vivi (Vin Ha Din Thi), Dominique Arnaud, Houda Chelaïfa, Cléclé (Cléa Rouel), et monsieur Soleil (Cyril Zahabi).

Tout mes collègues et amis de l'URGV qui on su me soutenir dans les moments de solitudes à la paillasse ou devant l'écran. (Steph & Steph, Lulu, Alex, Sansan, Phiphi, Juju, Véro, Cécile, Isa, Aurélie, Jess, Christelle, Marion, Jean-Luc, Jean-Philippe, Laure, Seb, Marie-Laure, Arnaud Charpentier pour mes inscriptions à l'université.

Les personnes du CNS et du CNG qui me supportent depuis le début.

Une spéciale dédicace pour Audrey Didier du centre de ressources de Clermont-Ferrand qui à su trouver les lignées de blés manquantes pour mes analyses.

« Que la génomique soit avec vous ! »

Harry

Résumé

En 2006, après un siècle d'investigations, le gène majeur de domestication 5AQ, conférant de nombreux caractères comme la non-déhiscence et un battage facile des épillets des blés polyploïdes, a été identifié comme un facteur de transcription homologue au gène *Apetala2* d'*Arabidopsis*. Bien que ceci représente une avancée importante, le rôle des autres homéologues de ce gène, présents dans le blé tétraploïde (*Triticum turgidum*) et le blé hexaploïde (*T. aestivum*), reste à élucider. Dans ce contexte mon sujet de recherche porte sur l'appréciation de l'**organisation, l'évolution et le fonctionnement du gène majeur de domestication (Q) et de ses homéologues chez les blés polyploïdes**.

J'ai tout d'abord séquencé et analysé 11 régions génomiques (clones BAC) portant les copies du gène *Q/q* dans différents génomes de blés polyploïdes et diploïdes; constituant ainsi la plus grande analyse comparative réalisée aujourd'hui chez le blé. Les comparaisons entre les différents génomes et différents niveaux de ploïdie montrent que le gène *Q/q* est la seule séquence conservée, en commun dans les régions génomiques comparées, et que l'homéologue *5Bq* est pseudogénéisé dans les blés hexaploïdes. Les comparaisons montrent que le reste des séquences génomiques sont constituées d'environ 80% d'éléments transposables (TEs) qui sont entièrement différents quand on compare les génomes A, B, D et S entre eux. A l'inverse, les TEs sont relativement mieux conservés entre haplotypes du même génome et continuent leurs dynamiques d'insertions et de délétions différentielles, conduisant à 19 événements de rupture de synténie. Parmi ces événements, j'ai pu identifier le premier *Hélitron* actif du blé, inséré dans le pseudogène *5Bq* du cultivar Renan. La recherche de son origine par comparaison de séquences et l'étude de la variabilité haplotypique m'ont permis de confirmer l'insertion récente et l'origine commune de cet élément au blé sauvage *Aegilops ventricosa*. Cette espèce a été introgressée dans certaines variétés de blés hexaploïdes.

L'analyse fonctionnelle comparant les caractères de domestication ainsi que l'expression et les interactions entre les différents homéologues du gène *Q/q* dans le blé hexaploïde a été rendue possible par la caractérisation des « lignées de délétion », où une ou plusieurs copies homéologues ont été perdues ou substituées. J'ai pu ainsi caractériser l'hyper-fonctionnalisation de l'homéologue 5AQ, et la sous-fonctionnalisation des homéologues *5Dq* et plus étonnamment *5Bq*, pseudogénéisé ; les trois homéologues contribuent aux caractères de domestication et se régulent entre eux. Les comparaisons précises des séquences des allèles 5AQ et 5Aq pour plusieurs génotypes domestiqués et sauvages m'ont permis d'identifier une mutation SNP associée, dans le site d'adressage d'un micro RNA (miR172). L'utilisation d'une technique RACE-PCR semi-quantitative montre que la mutation dans l'allèle 5AQ conduit à moins d'ARNm clivés par les miR172 et donc à sa plus forte expression ; comparée à celle de l'allèle 5Aq. Ceci suggère un rôle des miR172 dans la régulation des différents homéologues du gène *Q/q*.

Mots-Clés : domestication, gène Q, blé, polyplœdie, élément transposable, *Hélitron*

Abstract

In 2006, and after a century of investigations, the major domestication gene in polyploid wheat (5AQ), involved in non free threshing and spike easy beating, among many other traits, has been identified as a homolog of *Apetala2* gene of *Arabidopsis*. While this represents an important breakthrough, nothing was yet known about the role of other homoeologs of the *Q/q* gene present in tetraploid (*Triticum turgidum*) and hexaploid (*T. aestivum*) wheat. In this context, my PhD thesis consists in characterizing **organization, evolution and function of the major wheat domestication (*Q/q*) gene and its homoeologs in polyploid wheat.**

I realized first comparative sequencing and analysis of 11 genomic regions (BAC clones) spanning the *Q/q* gene homolog's in different hexaploid, tetraploid and diploid wheat; constituting the most important comparative analysis done for this group of species. Comparisons show that only *Q/q* gene homologs are conserved in different genomes and across different ploidy levels and that the *5Bq* homoeolog is pseudogenized in hexaploid wheat. The remaining genomic sequences, constituted of ~80% of transposable elements (TEs) are completely different when comparing A, B, D and S genomes between each others. On the contrary, TEs are more conserved between different haplotypes of a same genome and continue their active insertion and deletion dynamic, leading to 19 identified synteny breaks. Among these, I identified the first active *Hélitron* in wheat inserted into the *5Bq* pseudogene of a hexaploid wheat cv. Renan. The *Hélitron* insertion was subsequently retraced as recently occurring whereas it could have been originated from the wild wheat *Aegilops ventricosa* which has been introgressed into hexaploid wheat.

Functional analysis comparing phenotype, domestication traits, expression and interaction between different *Q/q* homoeologs was rendered possible using series of “deletion lines”, where one or several homoeologs were deleted. This allows determining the hyper-functionalization of 5AQ and the subfunctionalization of *5Dq* and more interestingly the subfunctionalization of the pseudogene *5Bq*. All three homoeologs were shown to contribute to the domestication traits and regulate each others.

Precise sequence comparison of 5AQ and 5Aq alleles from different domesticated and wild genotypes allow identification of a SNP mutation, associated with domestication, in the target site of a micro RNA (miR172). Using an adapted semi-quantitative RACE-PCR, I showed that the mutation leads to less cleaved mRNA of the 5AQ gene by the miR172 and consequently its higher expression than the 5Aq allele. This also suggests a general role of miR172 in regulating the different homoeologs of the *Q/q* gene.

Keywords: *Q* gene, domestication, wheat, polyploidy, transposable element, *Hélitron*

Préambule

Il y a environ 19 000 ans, des tribus de nomades (chasseurs-cueilleurs) décident de s'installer autour de la mer de Galilée (site épipaléolithique de Ohalo II) dans le croissant fertile (le nord de l'Irak, le sud de la Turquie, le nord de la Syrie, le Liban, Israël et la Jordanie) pour se sédentariser. Cette histoire va se reproduire des dizaines de fois dans des sites différents au cours du temps. Vers 12 800 ans avant J.C., des Natoufiens, précurseurs de civilisations, vont se mettre à cultiver des plantes sauvages nécessaires à leurs alimentations, leurs soins et diverses activités. Un jour, à force d'observations, certains d'entre eux vont constater que dans une aire de culture, certaines plantes, comme le blé, ne sont pas comme les autres, elles ne sont pas déhiscentes (ne disséminent pas les grains à maturité) et ont des propriétés intéressantes, etc.... C'est le début de la sélection inconsciente et de la domestication de plantes dont les blés.

Bien plus tard, il y aura bientôt 100 ans, en 1917 un scientifique du nom de Nilsson-Ehle observe un champ de blé et se pose une simple question, qui aura de lourdes conséquences scientifiques pour la compréhension de la domestication du blé par la suite : Pourquoi dans ce champ de blés domestiqués, certaines plantes révertent vers un phénotype sauvage ? Pour répondre à cette question, il émet l'hypothèse qu'il y a probablement peu de gènes impliqués dans le passage, du phénotype domestiqué au phénotype sauvage et *vice versa*. C'est le début d'une recherche qui va progresser avec l'émergence de disciplines et de techniques scientifiques pour trouver des réponses (archéologie, cytologie, génétique, phylogénie...).

Il y aura de nombreux intervenants et de nombreuses hypothèses émises de la part de botanistes, évolutionnistes, archéologues, paléontologistes, cytologistes, généticiens, mathématiciens et informaticiens. Chacun d'entre eux placera des pièces du puzzle. Mon sujet de thèse « **Organisation, évolution et fonctionnement des gènes majeurs de domestication (Q/q) chez les blés polyploïdes** » est une des petites pièces de ce puzzle et bénéficie de plus d'un siècle de recherches antérieures.

Table des matières

INTRODUCTION	21
PREMIERE PARTIE : ETUDE BIBLIOGRAPHIQUE	27
CHAPITRE I : BLE ET ECONOMIE NATIONALE ET MONDIALE	29
I-1 IMPORTANCE ECONOMIQUE DU BLE EN FRANCE	31
I-1-2 La production nationale	31
I-1-3 Le secteur de la meunerie en France	31
I-2 Importance économique mondiale du blé.....	33
I-3 Conclusion.....	35
CHAPITRE II : ORIGINE DE LA DOMESTICATION DU BLE	37
II -1 Domestications et Civilisations.....	39
II-1-1 Définition	39
II-1-2 L'Homme et les plantes	39
II-2 Origine des plantes cultivées	41
II-2-1 Les théories d'Alphonse de Candolle sur l'origine des plantes cultivées	41
II-2-1-1 La méthode d'Alphonse de Candolle	41
II-2-1-2 Origine des blés	47
II-2-1-2-1 Dans la catégorie des blés facilement battables Vilmorin définit quatre groupes	47
II-2-1-2-2 Les blés non facilement battables de Henri Vilmorin (3 groupes)	49
II-2-1-3 Conclusion sur les travaux d'Alphonse de Candolle	49
II-2-2 Darwin et l'origine des plantes cultivées	49
II-2-2-1 Les conclusions de Darwin	51
II-2-3 Avancées depuis Alphonse de Candolle et Charles Darwin	53
II-2-4 Les connaissances actuelles sur l'origine des blés	53
II-3 Les changements apportés lors de la domestication	55
II-4 Les connaissances actuelles sur le processus de domestication du blé	55
II-4-1 Un processus rapide, localisé, comprenant de nombreuses espèces différentes	59
II-4-2 Un processus lent et graduel, avec des événements répétés dans des foyers multiples (géographiquement) pour des espèces différentes et de façon indépendante	63
II-4-3 Que peut-on conclure sur ces deux hypothèses ?	63
II-5 Conclusion des recherches sur la domestication	69
CHAPITRE III : GENOMIQUE DU BLE ET POLYPLOIDIE.....	71
III-1 Description et classification du blé	73
III-2 Cytologie du blé	75
III-3 La polyploïdie	75
III-3-1 Définitions de la polyploïdie	75
III-3-2 Voies de formation des polyploïdes	81
III-3-3 Prévalence et fréquence de la polyploïdie chez les eucaryotes	85
III-3-4 Changements génomiques chez les polyploïdes	87
III-3-4-1 Remaniements structuraux: dynamique et mécanismes	87
III-3-4-2 Les mécanismes responsables des réarrangements structuraux	91
III-3-4-3 Changements épigénétiques	93
III-3-4-4 Reprogrammation de l'expression des gènes	93
III-3-4-5 Devenir des gènes dupliqués dans le génome des plantes polyploïdes	95
III-4 L'aneuploïdie : effets sur la structure et l'expression des génomes	97
III-4-1 Travaux d'Ernest Sears (1910-1991)	99
III-4-2 Les lignées de délétions	101
CHAPITRE IV : IMPORTANCE DES ELEMENTS TRANSPOSABLES DANS LE GENOME DU BLE.....	105
IV-1 Description des éléments transposables	107
IV-1-2 Classification des éléments transposables	107
IV-1-3 Importance des TE dans les autres espèces	109
IV-1-4 Dynamique des éléments transposables	111
IV-1-4-1 Prolifération des TE	111

IV-1-4-2 Mécanismes d'insertion des TEs	113
IV-1-4-3 Mécanisme de transposition des rétrotransposons	113
IV-1-4-4 Mécanisme de transposition des rétrotransposons	115
IV-1-4-5 Mécanisme de transposition des transposons à ADN	115
IV-1-4-6 Mécanisme de transposition des <i>Hélitrons</i>	117
IV-1-4-7 Mécanismes d'élimination des TEs	119
IV-1-5 Rôle des Eléments Transposables	119
IV-1-6 Eléments Transposables et polyploidie	121
V-1 Pourquoi caractériser les gènes de domestication	125
V-2 Identification du gène <i>Q</i>	125
V-2-1 Les différents caractères sous l'influence du gène <i>Q</i>	127
V-2-2 Effet dose de la variation du nombre de copies <i>Q/q</i>	129
V-2-3 Comportement de <i>Q</i> chez les blés tétraploïdes	129
V-2-4 Comportement du gène <i>Q</i> dans les hexaploïdes	129
V-2-5 Identification et caractérisation moléculaire du gène <i>Q</i>	131
V-2-5-1 Carte génétique fine	131
V-2-5-2 "Chromosome walking" ou marche chromosomique sur le locus <i>Q</i>	133
V-2-5-3 Le clonage et la validation du gène <i>Q</i>	133
V-2-5-4 Dosage et effet pleiotropique du gène <i>Q</i>	133
Bilan de l'étude bibliographique	139
DEUXIEME PARTIE : ANALYSES ET RESULTATS	141
CHAPITRE VI: ROLE ET FONCTIONNEMENT DES DIFFERENTS HOMEOLOGUES DU GENE DE DOMESTICATION <i>Q/q</i> DANS LES BLES POLYPLOIDES	143
VI-1 Les questions posées	145
VI-2 Conclusions sur les travaux publiés dans l'article et orientations des recherches complémentaires	165
VI-2-1 Devenir et interactions entre les copies homéologues du gène <i>Q/q</i> dans le blé hexaploïde ..	165
VI-2-2 Dynamiques des éléments transposables	167
VI-2-3 Identification et caractérisation du premier <i>Hélitron</i> actif du blé	169
VI-3 Généralités sur les <i>Hélitrons</i>	171
VI-3-1 caractérisation de l' <i>Hélitron</i> actif du blé	173
VI-3-2 Identification des différentes catégories d' <i>Hélitrons</i> dans le blé hexaploïde Chinese Spring	177
VI-3-2-1 Les catégories d' <i>Hélitrons</i> identifiés	177
VI-3-2-2 Bilan des insertions par catégories	177
VI-3-2-3 Traçage des <i>Hélitrons</i> identifiés	181
VI-3-2-3-1 Matériel et méthodes	181
VI-3-2-3-2 Résultats sur la caractérisation des <i>Hélitrons</i> de CS et de Renan	181
VI-4 Recherche de l'origine de l'insertion de l' <i>Hélitron</i> actif de Renan	185
VI-4-1 Un peu d'histoire sur les collections de blés soviétiques	185
VI-4-2 Origine de la variété Mironovskaya 808	185
VI-5 Conclusion sur l'origine de l' <i>Hélitron</i> dans <i>5Bq</i>	189
CHAPITRE VII : REGULATION DU GENE MAJEUR DE LA DOMESTICATION	191
VII-1 La question qui demeure	193
VII-2 Les petits ARN	193
VII-2-1 Biogénèse, mécanisme et fonction des petits ARNs (siRNA et miRNA)	195
VII-2-1-1 Les siRNA	197
VII-2-1-2 Les miRNA	201
VII-3 Interactions et «Régulation-Croisée» des homéologues <i>5AQ/q</i> , <i>5Bq</i> et <i>5Dq</i> (ceRNA) via une population de mi172	205
VII-3 Régulation des transcrits des allèles <i>Q</i> et <i>q</i> du chromosome 5A du blé	207
VII-4 Matériel et méthodes	209
VII-4-1 Principe de la technique RACE PCR	209
VII-4-1-1 Dessins des amorces PCR	209
VII-4-1-2 Extraction des ARNs totaux	209

VII-4-1-3 Utilisation du kit GeneRacer TM RLM-RACE (RNA ligase-mediated Rapid Amplification of cDNA Ends) selon le protocole INVITROGEN [@]	209
VII-4-1-4 PCR semi-quantitatives	211
VII-4-1-4-1 Préparation des prémix PCR	211
VII-4-1-4-2 Conditions pour la PCR semi quantitative et le RACE PCR	213
VII-4-1-4-3 Observation des produits PCR	213
VII-5 Analyse des résultats et conclusions	213
DISCUSSION GENERALE ET PERSPECTIVES	219
ANNEXES.....	229
Annexe 1 : Classification des blés anciens.....	230
Annexe 2 : Alignements des séquences internes d' <i>Hélitrons</i> (CS et Renan) pour établir les groupes et vérifier la proximité de l' <i>Hélitron</i> de Renan.	231
Annexe 3 : Alignements des séquences internes et externe aux <i>Hélitrons</i> en 3'pour différencier les groupes établis et les insertions.....	232
Annexe 4 : position des amorces PCR pour RACE sur alignements de séquences <i>5AQ</i> , <i>5Bq</i> et <i>5Dq</i>	234
Annexe 5 : Tableau des amorces PCR et RACE-PCR pour l'étude de la régulation des gènes <i>5AQ/q</i> par le miR172.....	237
Annexe 6 : Liste des plantes et accèsions à l'origine de la variété Renan.....	238
Annexe 7 : Liste des plantes et accèsions Ukrainienne étudier dans la recherche de l'origine de l' <i>Hélitron</i> dans le gène <i>5Bq</i>	239
Annexe 8 : Photos des lignées Russes positives pour l'insertion de l' <i>Hélitron</i> dans le gène <i>5Bq</i> de cv. Renan.....	240
Annexe 9 : Article Charles, M. <i>et al.</i> (2008) Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. <i>Genetics</i> 180, 1071–86	241
Bibliographie.....	259
Ouvrages imprimés.....	259
Ouvrages électroniques.....	259
Sites web consultés	259
Articles de périodiques imprimés	260

Liste des figures :

<i>Fig. I-2 Production mondiale des céréales.....</i>	<i>32</i>
<i>Fig. I-3 Evolution de l'offre et de la demande de blé dans le monde sur la période 1990-2009.</i>	<i>34</i>
<i>Fig. II-1 Evolution de la diversité biologique.....</i>	<i>38</i>
<i>Fig. II-2 Alphonse de Candolle (1806-1893).....</i>	<i>40</i>
<i>Fig. II-3 Charles Darwin (1809-1882).</i>	<i>40</i>
<i>Fig. II-4 Nicolaï Vavilov (1887-1943).....</i>	<i>52</i>
<i>Fig. II-5 Les 12 principaux centres d'origines de la domestication des plantes dans le monde.</i>	<i>52</i>
<i>Fig. II-6 Le croissant fertile :.....</i>	<i>54</i>
<i>Fig. II-7 Les blés domestiqués et leurs liens avec les blés sauvages.</i>	<i>54</i>
<i>Fig. II-8 Localisations des sites de fouilles archéologiques dans le croissant fertile.</i>	<i>56</i>
<i>Fig. II-9 La domestication : Un processus rapide, localisé, comprenant de nombreuses espèces différentes :</i>	<i>58</i>
<i>Fig. II-10 La domestication : Un processus lent et graduel, avec des évènements répétés dans des foyers multiples (géographiquement) pour des espèces différentes de façon indépendante.</i>	<i>62</i>
<i>Fig. III-1 Description anatomique de l'épillet :</i>	<i>72</i>
<i>Fig. III-2 Description anatomique de la fleur du blé dans l'épillet.....</i>	<i>72</i>
<i>Fig. III-3 La ploïdie et la classification des blés :</i>	<i>74</i>
<i>Fig. III-4 Les évènements de polyploïdisation dans les différents règnes.</i>	<i>76</i>
<i>Fig. III-5 Les évènements de polyploïdisation pour quelques espèces.</i>	<i>76</i>
<i>Fig. III-6 Fondation d'une espèce polyploïde par hybridation de deux espèces différentes..</i>	<i>77</i>
<i>Fig. III-7 Schéma de formation des espèces polyploïdes suivant différentes voies :.....</i>	<i>80</i>
<i>Fig. III-8 Schéma de la méiose et de la mitose :</i>	<i>82</i>
<i>Fig. III-9 Modèle d'évolution de la structure des chromosomes chez les Poaceae.</i>	<i>90</i>
<i>Fig. III-10 Caryotype d'une cellule de blé hexaploïde euploïde (jeu complet).</i>	<i>96</i>
<i>Fig. III- 11 Ernest Sears (1910-1991).</i>	<i>98</i>
<i>Fig. III-12 Exemple schématique de lignées aneuploïde, de substitution disomiques et de délétion.....</i>	<i>98</i>
<i>Fig. IV-1 Les deux mécanismes d'insertion des éléments transposables à la base de leur classification :</i>	<i>106</i>
<i>Fig. IV-2 Classification des différents TEs sur 5 niveaux :</i>	<i>108</i>

Fig. IV-3 Divergence des espèces de la famille des Poaceae, évènements de polyploïdisation, taille des génomes et proportion des TEs :	110
Fig. IV-4 Mécanismes de transposition des principaux TEs de classe I :	112
Fig. IV-5 Mécanismes de transposition des principaux TEs de classe II :	114
Fig. IV-6 Mécanismes schématiques de délétion par recombinaison homologue inégale	118
Fig. V-1 Comparaison entre la forme « squarehead » (tétraédrique) et speltoïde de l'épi :	124
Fig. V-2 Croisement entre <i>T. aestivum</i> ssp. <i>aestivum</i> (5AQ) et <i>T. aestivum</i> ssp. <i>spelta</i> (5Aq) d'après Muramatsu (1962) :	126
Fig. V-3 Effet dose des gènes Q/q (épis) :	128
Fig. V-4 Effet dose du gene q (plantes) :	128
(Muramatsu, 1963)	128
Fig. V-5 Description des effets dose des gènes Q/q :	128
Fig. V-6 Expérience de l'effet dose du gène Q par transgénèse :	132
Fig. V-7 Corrélation entre niveau de transcription et phénotype obtenu :	132
Fig. V-8 Contrôle du nombre d'homéoallèles Q/q par génotypes :	134
Fig. V-9 Alignements des séquences des gènes Q et q de blés tétraploïdes et hexaploïdes. .	134
Fig. V-10 Différences phénotypiques entre engrain sauvage (<i>T. monococcum</i> ssp. <i>boeoticum</i>) et engrain domestiqué (<i>T. monococcum</i> ssp. <i>monococcum</i>) :	136
Fig. V-11 Les autres sélectionnés chez les blés :	136
Fig. V-12 Effet du gène c/C de compaction de l'épi :	136
Fig. V-13 Exemples de diversités de phénotypes observés chez les blés.	138
Fig. VI-1 Arbre réalisé à partir de 13 séquences 3' terminale (structure en épingle à cheveux) internes des différentes séquences d'Hélitron-like de Chinese Spring et l'Hélitron du 5ABq.	178
Fig. VI-2 Détection de différents Hélitrons du blé et de leurs insertions dans une collection de blés :	179
Fig. VI-3 Généalogie de la variété Renan :	183
Fig. VI-4 Etapes possibles du transfert de l'Hélitron de Renan depuis le blé Banatka Russe (1924) jusqu'à Renan (1989) :	188
Fig. VII-1 Les différentes voies d'action des petits ARN interférents :	198
Fig. VII-2 : Les zones d'hybridation préférentielles du miRNA sur le site d'adressage de l'ARNm pour les plantes et les animaux :	202
Fig. VII-3 : Interactions et « Régulation-Croisée » des homéologues 5AQ/q, 5Bq et 5Dq (ceRNA) via une population de mi172 suivant l'hypothèse de Salmena et al:	204

<i>Fig. VII-3 Analyses de l'hybridation du Tami172 sur différents facteurs de transcriptions :</i>	206
<i>Fig. VII-4 Technique RACE-PCR :</i>	208
<i>Fig. VII-5 PCR semi quantitative sur les ARNm clivé et non-clivé :</i>	212
<i>Fig. VII-6 Mutations identifiées entre les phénotypes sauvages /domestiqués (bleu) sur le site d'adressage du miR172 pour les gènes cly1 et Q :</i>	214
Liste des tableaux :	
<i>Tableau II-1 : Classification actuelle des principaux blés</i>	666
<i>Tableau VI-1 Amorces spécifiques de l'hélitron de Renan et des 3 différentes catégories d'hélitrons de Chinese Spring :</i>	176
<i>Tableau VI-2 Récapitulatif des 16 couples d'amorces sur les 24 génotypes de blés :</i>	180
<i>Tableau VI-3 Liste des blés d'origine ukrainienne sélectionnés pour le crible PCR avec les amorces Hélitron 5Bq spécifique :</i>	184

Abréviations principales (françaises et anglaises)

ADN : acide désoxy-ribonucléique

ADNc : ADN complémentaire

ADNr : ADN ribosomique

ARN : acide ribonucléique

ARNdb : ARN double-brin

ARNm : ARN messenger

ARNr : ARN ribosomique

ATP : adénosine tri-phosphate

BAC : bacterial artificial chromosome

B.P. : years before present

CDS : séquence codante

dNTP : désoxynucléotide tri-phosphate

EST : expressed sequence tag

ET : élément transposable

FDR : First Division Restitution (absence de première division de méiose)

kb : kilobase

LTR : long terminal repeat

Ma / Mya : million d'année / million years ago

Mb : mégabase

miRNA : micro ARN

ORF : open reading frame

pb : paire de bases

PCR : polymerase chain reaction

Ph1 : Pairing homeologous 1 (locus contrôlant l'appariement homéologue chez le blé)

PTGS : post-transcriptional gene silencing

Q ou q : squarehead - tétraédrique

QTL : Quantitative Trait Locus

RACE : rapid amplification of cDNA ends

RdDM : RNA-dependent DNA methylation

RFLP : Random Fragment Length Polymorphism (Marqueurs moléculaires basé sur le polymorphisme de sites de restrictions)

RNAi : ARN interférence

RT : reverse-transcription

SDR : Second Division Restitution (absence de seconde division de méiose)

siRNA : small interfering RNA (petits ARN interférents)

miRNA : micro ARN interférents

INTRODUCTION



Quand Hadès, souverain des morts, enleva Perséphone pour en faire son épouse, Déméter partit à sa recherche et négligea les récoltes de la Terre. En prenant la forme d'une vieille femme nommée «Doso», elle erra pendant neuf jours. Se rendant compte qu'une famine menaçait les mortels, Zeus se décida à envoyer Hermès au royaume d'Hadès pour lui demander de rendre Perséphone à sa mère.

Mais Perséphone avait mangé six pépins de la grenade offerte par Hadès, en guise de dernière ruse pour la garder avec lui, et la tradition voulait que quiconque mangeait dans le royaume des morts ne puisse le quitter. Zeus s'entendit pour que Perséphone passe l'hiver aux Enfers et le reste de l'année avec sa mère. C'est de ce mythe qu'est né le cycle des saisons dans la mythologie grecque.

Déméter enseigna aux humains le travail des semis et du labour. Durant son errance sous la forme de Doso, elle rencontra Céléos, roi d'Éleusis. Pour le remercier de son accueil, elle prit avec elle les fils du roi, Démophon et Triptolème, tenta de rendre le premier immortel et enseigna au second l'art de l'agriculture. Celui-ci devait en retour enseigner cet art au reste des humains. Certaines traditions mentionnent qu'elle lui aurait aussi donné des grains de blé afin qu'il les répandît sur la Terre.

Mais ceci n'est qu'une légende, à notre époque taper le mot « blé » ou « wheat » sur un navigateur internet permet de découvrir qu'il est profondément ancré dans notre société et notre culture. Le blé tient une place importante dans différents domaines : l'économie, la culture, les sciences et bien plus encore.

Dans la première partie du manuscrit pour situer mon sujet de recherche, il me paraissait important de présenter une synthèse des principales connaissances acquises jusqu'à nos jours qui auraient comme fil conducteur le « blé ». Cette première partie permet de diffuser deux informations importantes pour la compréhension de mon sujet de thèse. La première sur l'origine de la domestication du blé et du/des processus de domestication. En effet, si les blés sont les plus anciennes plantes cultivées,

le processus de domestication laisse apparaître un échelonnement des blés domestiqués liés à leurs polyploïdie. Les blés domestiqués sont chronologiquement diploïdes, allotétraploïdes puis allohexaploïdes. Le gène de domestication majeur 5AQ de mon sujet apparait dans des blés tetraploïdes.

Avant d'expliquer génétiquement son importance dans des blés domestiqués, J'aborde brièvement certaines particularités du génome du blé liées à sa taille. En effet, l'immensité du génome des blés hexaploïdes est principalement due à deux forces évolutives, une polyploïdie fréquente et récurrente et une prolifération répétée des éléments transposables. Dans deux chapitres qui suivent (« Génomique du blé et polyploïdie » (chapitre III) et « Importance des éléments transposables » (chapitre IV), je montre l'importance de la polyploïdie chez le blé et de la sélection possible de plantes ayant perdu ou gagné des chromosomes (lignées aneuploïde), ou perdu des segments génomiques (lignées de délétion). Ces plantes sont de puissants outils ayant servi à la compréhension des effets de la polyploïdie chez le blé. Pour mon sujet de thèse ce matériel a constitué un support important pour préciser l'implication des homéologues (5AQ/q, 5Bq et 5Dq) dans les caractères de domestication chez le blé polyploïde. L'autre chapitre bibliographique est consacré aux éléments transposables, pour lesquels je ne peux plus ignorer leur importance quantitative dans le génome du blé, mais également leur importance dans la diversification des espèces (liée à leurs dynamiques d'insertion/ délétion). Ce chapitre explique également les mécanismes de transposition des éléments transposables, leurs classifications et surtout les rôles possibles de l'*Helitron* actif, que j'ai découvert dans mes analyses, inséré dans l'homéologue 5Bq. Pour conclure cette partie bibliographique je réalise un bilan des recherches sur la génétique des caractères de domestication, le clonage, la caractérisation moléculaire des gènes *Q/q* comme étant des facteurs de transcription de la famille des AP2.

Dans la deuxième partie du manuscrit je présente mes hypothèses de travail et je décris mes résultats. Sachant que le gène majeur identifié 5AQ, est responsable de la mise en place des organes floraux et qu'il porte une mutation V329I associée aux caractères de domestication, je me suis interrogé sur le devenir et le rôle des autres homéologues présents dans les blés polyploïdes. Ceci m'a mené à m'interroger à trois niveaux.

1. Comment les gènes *Q* et *q* ont évolué par rapport aux orthologues et/ou homéologues dans les différents blés diploïdes et polyploïdes?
2. Quels sont les niveaux d'implications des différents homéologues dans les caractères de domestication des blés polyploïdes?
3. Comment les différents homéologues, interagissent-ils et se régulent entre eux?

Pour répondre à ces questions, j'ai utilisé différentes stratégies.

Pour la première question, j'ai réalisé l'analyse des séquences de 11 régions génomiques portant les différentes copies des gènes *Q/q* des génomes (A, B, D et S) et à différents niveaux de ploïdie des blés sauvages et domestiqués. Cette étude a été complétée par une analyse haplotypique par PCR d'une collection de 100 accèsions représentant aussi les différents génomes et les différents niveaux de ploïdie.

Pour la deuxième question, en collaboration avec l'équipe américaine de J. Faris (US Department of Agriculture, Northern Crop Science, Fargo), une collection de lignées de délétion et/ou de substitution pour les différents homéologues a été constituée. Cette collection nous a permis de caractériser l'implication et le rôle de chaque homéologue dans le phénotype de forme de l'épi de blé, mais également dans d'autres caractères liés au rendement, leurs implications au niveau fonctionnelles et leurs interactions au niveau d'expression. Les résultats et conclusions sur la première et deuxième question ont permis une publication présentée dans la partie analyses et résultats du manuscrit. En plus d'avoir pris en main entièrement la première partie, j'ai travaillé en interaction avec Zengcui Zhang afin d'étudier par PCR-Q l'expression et les interactions entre les différents homéologues du gène *Q/q*.

Pour répondre à la troisième partie et interpréter les interactions et la régulation croisée des différents homéologues, je me suis inspiré de l'hypothèse de Salmena et al, 2011, Les transcrits de tout types (codant, non codant, alternatifs, pseudogènes, etc...) ayant en commun une séquence site d'adressage spécifique pour un microARN interagissent et se régulent entre eux par l'intermédiaire de la voie des microARNs. Enfin, j'ai pu étudier le rôle régulateur des microARN (mi172) en montrant que par rapport à l'allèle sauvage 5Aq, une mutation dans la séquence du site d'adressage, également site de clivage par les miRNA dans l'allèle domestiqué 5AQ conduit à sa moindre reconnaissance et à son non clivage par les mi172. La mutation est donc responsable de la plus forte expression dans les blés domestiqués du facteur de transcription de la famille des AP2 permettant la mise en place des organes floraux.

A la fin de cette deuxième partie, analyses et résultats, je propose une discussion et les perspectives pour la continuité de mes travaux.

PREMIERE PARTIE : ETUDE BIBLIOGRAPHIQUE

CHAPITRE I :

BLE ET ECONOMIE

NATIONALE ET MONDIALE

La production céréalière (hors riz) en France
 9,3 Mha cultivés pour 63,7 Mt produits
 (Source: SSP, campagne 2011/2012)

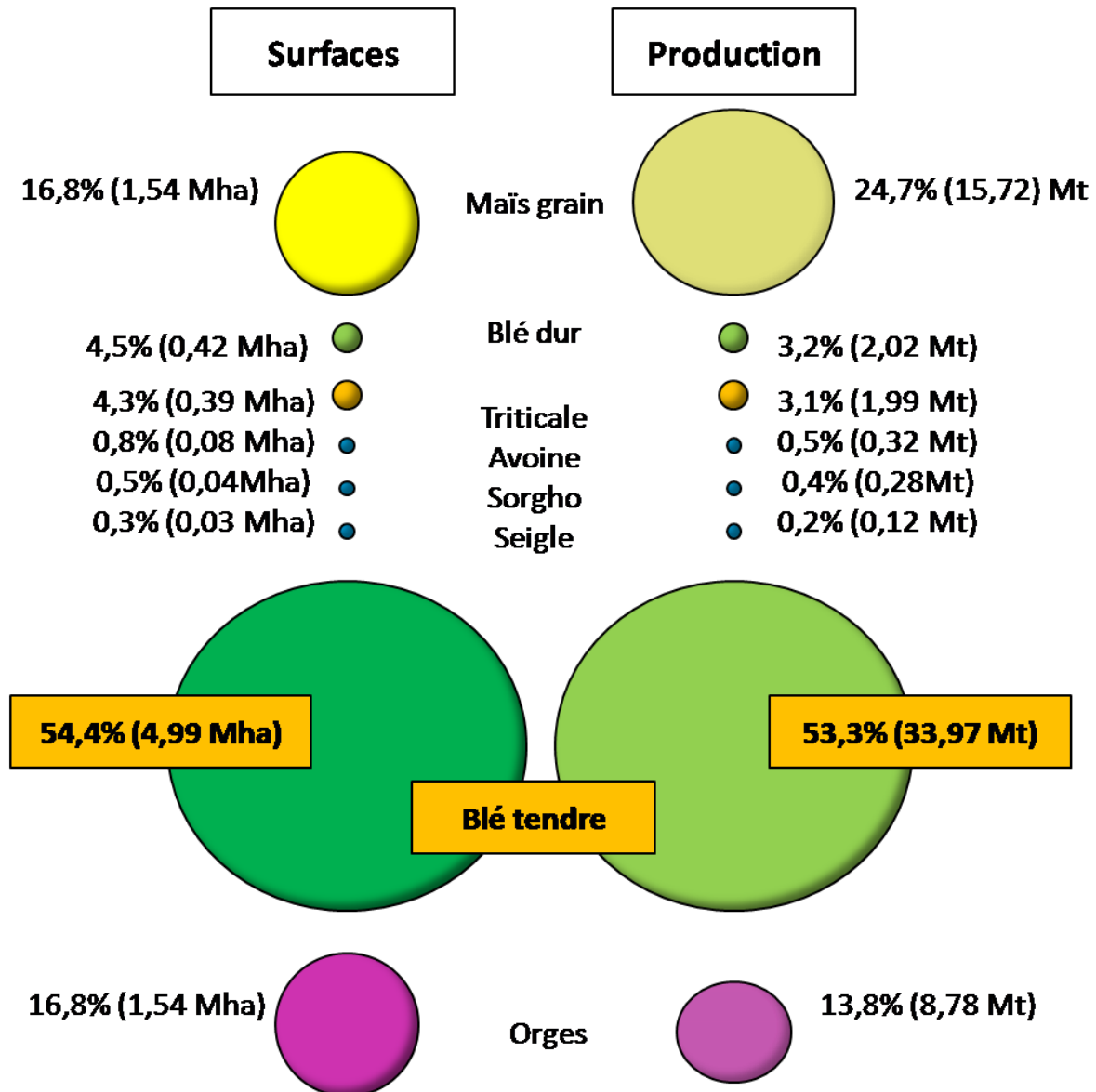


Fig. I-1 La production céréalière (hors riz) en France :
 (inspiré du site WWW.passioncereale.fr)

I-1 Importance économique du blé en France :

I-1-2 La production nationale :

C'est la céréale la plus importante en France (Fig. I-1), avec 35,9 millions de tonnes (Mt) produites en 2013 pour une surface cultivée de 5,3 Mha, soit 19,5% de la superficie agricole utilisée (FAO, <http://fao.org>). Ceci représente plus de 210 000 exploitations (soit 42% du total des exploitations françaises). L'exportation concerne 50% des récoltes nationales, les autres 50% sont destinée aux usages nationaux.

En France, les fabricants d'alimentation animale sont les premiers utilisateurs avec 34,5% de la production destinée aux usages nationaux. Pour l'alimentation humaine l'utilisation est de l'ordre de 32,7% et le pourcentage restant rentre dans les stocks. Or le stock comme le pratiquaient nos ancêtres, détermine le nombre de jours couverts en termes de consommation. Le blé tendre est utilisé pour la production de farines, produits de base dans l'industrie du pain, qui vient en deuxième position à égalité avec le secteur de l'amidonnerie-glutennerie (soit 19% des utilisations françaises). La biscuiterie (7%) arrive en troisième position, suivie par le secteur des agro carburants (6,7%).

Le blé dur trouve notamment son utilisation dans l'industrie des pâtes alimentaires et dans la production des semoules.

I-1-3 Le secteur de la meunerie en France :

L'industrie agroalimentaire (IAA) française compte environ 13 500 entreprises pour 415 000 salariés. La meunerie française regroupe 376 entreprises et plus de 6000 salariés. Les farines produites sont utilisées en boulangerie artisanale et industrielle. La boulangerie artisanale détient 60% de parts de marché du pain pour un chiffre d'affaire de 7 milliards d'euro. Elle est représentée par 34 000 boulangeries (130 000 personnes). La boulangerie industrielle détient les 40% restant avec 274 entreprises (35 000 salariés). Pour le secteur de la biscuiterie l'industrie représente un chiffre d'affaire de 400 millions d'euros et 11 300 emplois.

La production céréalière dans le monde
Pour la campagne 2011/2012
(Sources: USDA, campagne 2011/2012)

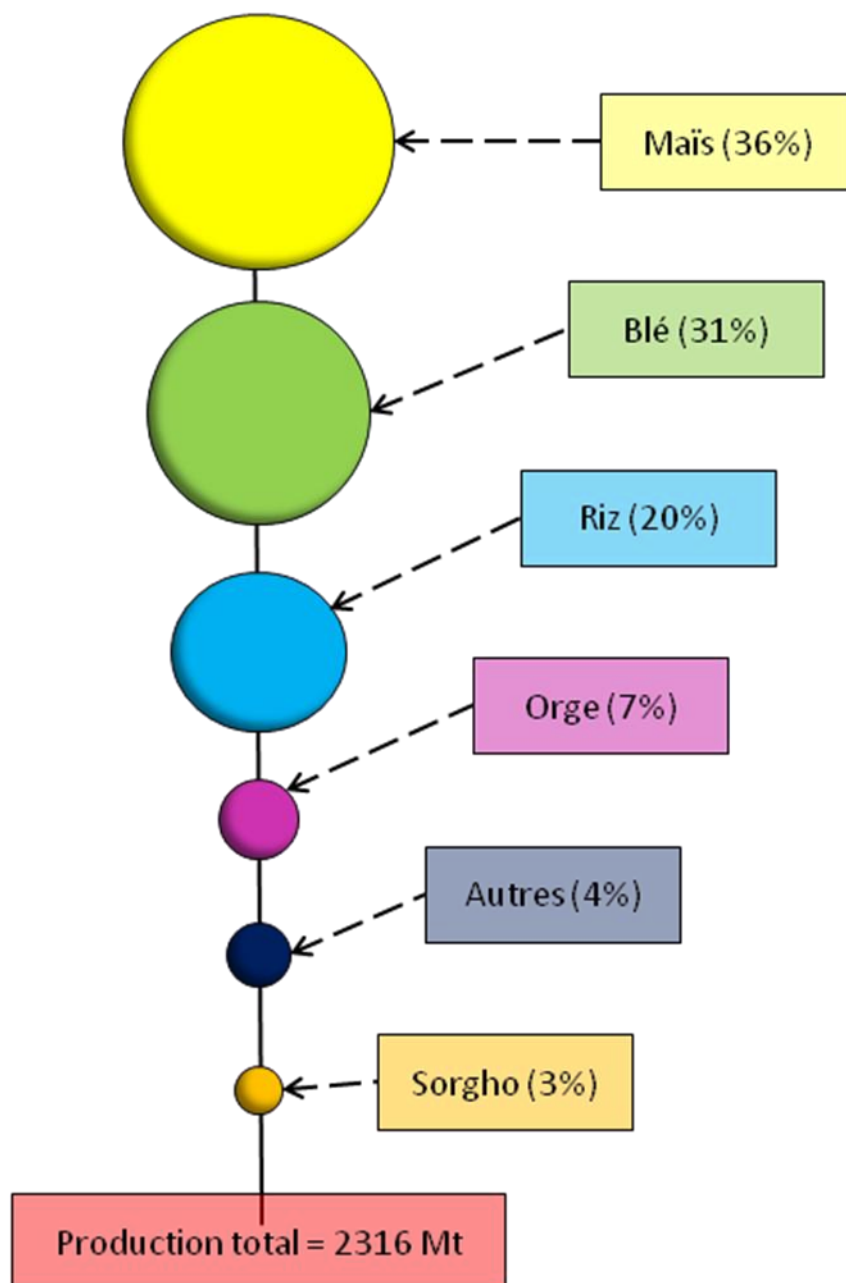


Fig. I-2 Production mondiale des céréales.

(source : WWW.FAO.org)

I-2 Importance économique mondiale du blé :

Avec le maïs et le riz, le blé est l'une des trois principales céréales mondiales. En 2013, sur 2,492 milliards de tonnes de céréales produites, le blé représente 28,2% de la production mondiale, le riz 20% (soit 704 et 500 millions de tonnes respectivement). Ces deux céréales constituent la base de la nutrition humaine. Le maïs représente 38,7% (soit 966 millions de tonnes), mais il est principalement utilisé pour l'alimentation animale (Fig. I-2).

Le blé occupe aujourd'hui, la deuxième place mondiale dans la production céréalière (selon les statistiques de la FAO), c'est également 2,8 fois plus que dans les années soixante. Cette hausse de la production est principalement due à une augmentation constante des rendements à l'hectare plutôt qu'à une augmentation des surfaces cultivées (239,2 millions d'hectares en 1981 et 216,8 millions d'hectares en 2010), mais elle demeure la plus grande superficie mondiale cultivée (15% de la surface agricole mondiale). Une grande part de la récolte mondiale de blé est produite par une dizaine de pays. En 1960, les cinq plus grands producteurs étaient l'URSS (25% de la production mondiale), l'Union européenne (16%), les Etats-Unis (16%), la Chine (9%) et le Canada (6%), représentant à eux seul 66% de la production. En 2011, dans le classement des cinq plus grands producteurs, l'Inde remplace le Canada et se place en troisième place. Les deux premiers étant maintenant l'union européenne (21%) et la Chine (18%).

Dans le rapport 'Evolution du marché mondial du blé au cours des cinquante dernières années' (Gavira et Burny, 2012), on trouve un certain nombre d'informations sur la situation économique du blé dans le monde. Notamment, la diminution des réserves mondiales de blé, surtout destiné à l'exportation. Quand on regarde le rapport stocks des principaux exportateurs sur l'utilisation totale, on passe de 21,7% en 2009/2010 à 14,6% en prévision pour 2013/2014. Or le prix du blé est indexé par l'offre et la demande, ce qui veut dire que les prix vont augmenter, car les exportateurs vont avoir de moins en moins de stocks disponibles

Exemple, en 2010 la Russie a souffert de sécheresse et d'incendies, le premier réflexe du gouvernement a été d'imposer un embargo à l'exportation de blé pour privilégier son marché interne. Elle est passée de 14% à 3%, par rapport au 20% mondiaux. L'impacte a été terrible pour les importateurs.

Actuellement l'UE est devenue le plus grand producteur, mais également le plus grand exportateur, le rapport exportation / production étant de 60% tandis que les Etats-Unis sont à 17%. Cela n'a pas toujours été ainsi, en 1960, l'EU importait 45% de sa consommation agricole, maintenant elle importe uniquement 4% de cultures autres que le blé (soit 4.7 millions de tonnes, de riz, d'oléagineux et d'aliments pour animaux). Malheureusement, la demande n'a fait que se déplacer. Actuellement, les régions les plus importatrices sont devenues, l'Afrique du Nord et le Moyen-Orient (32%), l'Asie de l'Est et du Sud-est (23%) et l'Afrique subsaharienne (11%).

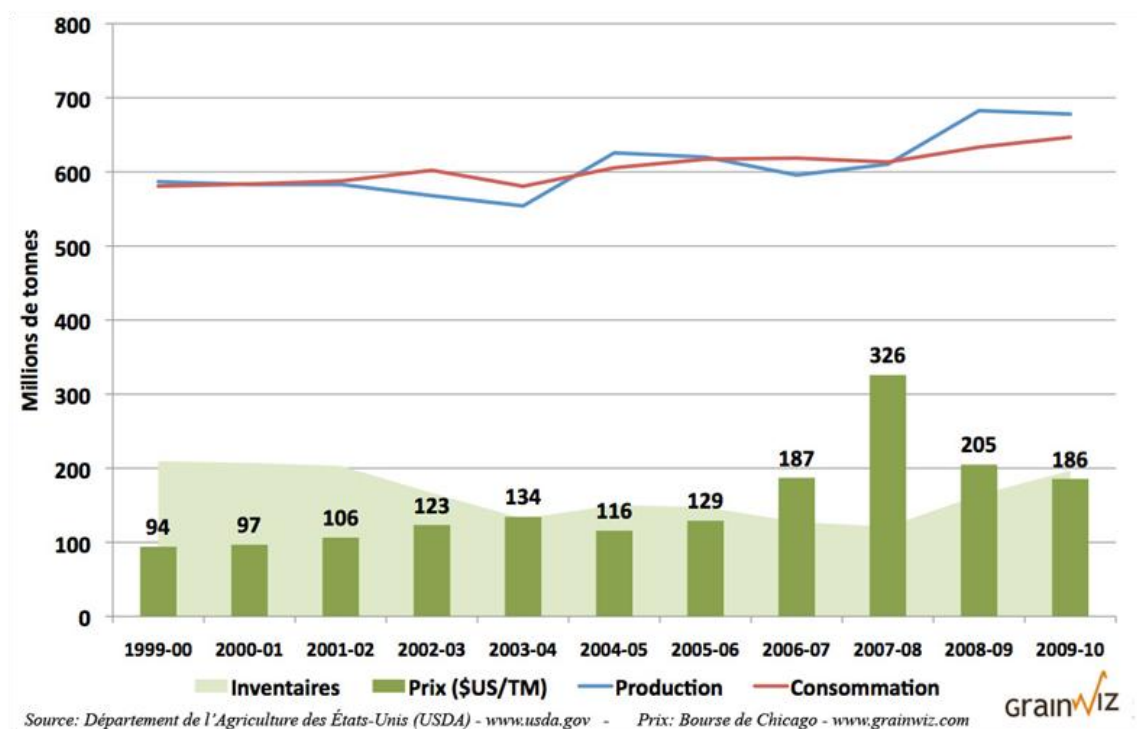


Fig. I-3 Evolution de l'offre et de la demande de blé dans le monde sur la période 1990-2009.

(Source : WWW.planetoscope.com)

I-3 Conclusion :

Pour conclure, en un demi-siècle la production mondiale de blé a triplé. Cela est principalement dû à la hausse des rendements. Les échanges internationaux ont su rester stables. Il faut chercher des solutions qui dépassent les limites physiques et biologiques pour maintenir cet équilibre (voir Fig. I-3).

Les niveaux de production vont de plus en plus avoir de difficultés à satisfaire la demande mondiale qui ne cesse de s'accroître avec l'augmentation de la population et les changements climatiques. Par conséquent, il est important d'améliorer le rendement et la qualité nutritionnelle du blé et cela dans une perspective d'agriculture durable. La compréhension de son génome et de son évolution est un des moyens pour y contribuer.

CHAPITRE II :

ORIGINE DE LA

DOMESTICATION DU BLE

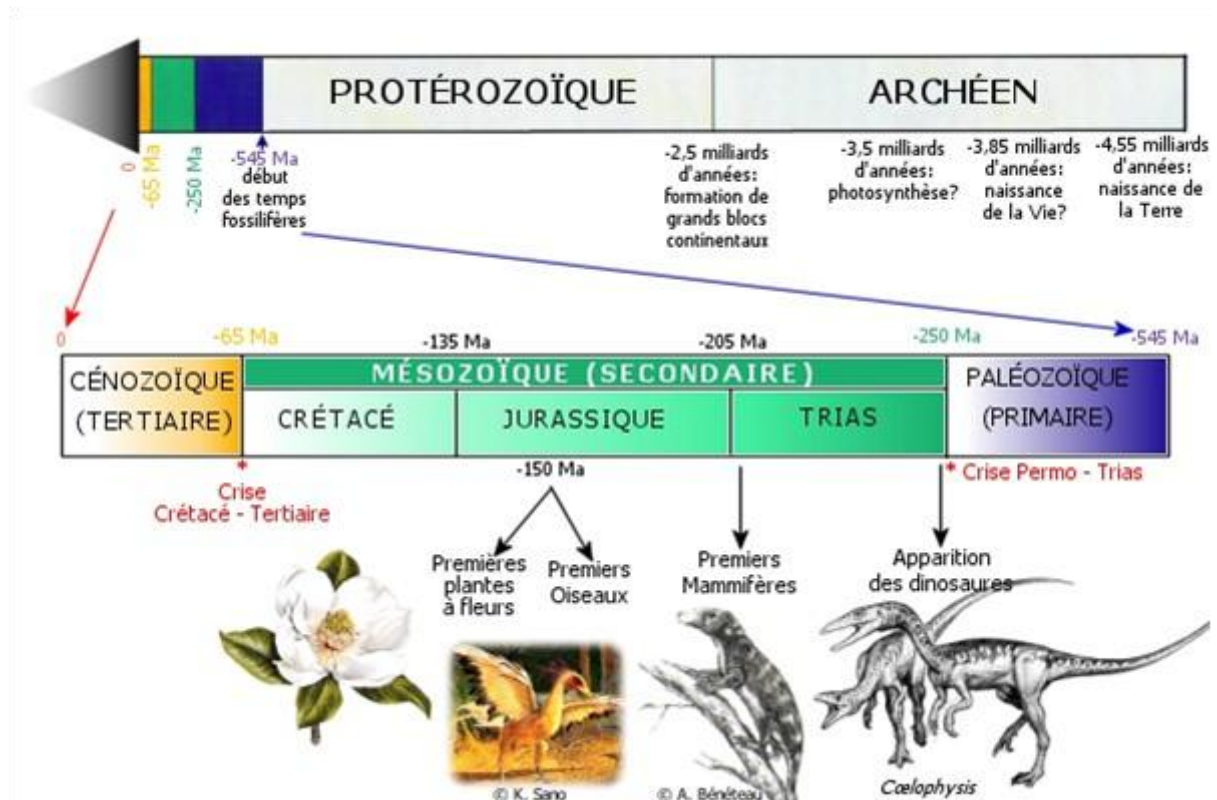


Fig. II-1 Evolution de la diversité biologique.

(Buffetaut, 2013).

II -1 Domestications et Civilisations :

II-1-1 Définition :

“La domestication d’une espèce, animale ou végétale, est l’acquisition, la perte ou le développement de caractères morphologiques, physiologiques ou comportementaux nouveaux et héréditaires, résultant d’une interaction prolongée, d’un contrôle voire d’une sélection délibérée de la part des communautés humaines”.

On comprend indéniablement, que l’on ne peut pas étudier l’évolution de la domestication sans s’intéresser à l’évolution de l’homme et de son passage du nomadisme au sédentarisme. Celui qui voudrait étudier une partie de notre histoire et de l’émergence de nos civilisations, devrait également intégrer l’histoire de la domestication des espèces.

Dans ce chapitre je vais brièvement résumer les travaux réalisés par quelques précurseurs, dont les méthodes et les résultats sont encore utilisés de nos jours.

II-1-2 L’Homme et les plantes :

Notre planète a environ 4.55 milliards d’années (Fig. II-1). La vie est apparue il y a 3,85 milliards d’années. Je ne citerai pas tous les événements qui se sont produits depuis, mais si on fait le parallèle entre l’évolution de l’homme et celui des plantes à fleurs (angiosperme), l’interaction majeure est la domestication des espèces par l’homme.

C’est le début d’un processus qui va permettre la fondation et l’essor des civilisations avec le début de l’agriculture et de l’élevage. En effet, l’homme qui a vécu dans le nomadisme depuis 2,5 millions d’années, opte pour un changement il y a 19 000 ans pour se sédentariser (site épipaléolithique d’Ohalo II au sud-ouest de la mer de Galilée - Israël).

Les archéologues définissent une civilisation suivant des critères d’organisation et des critères de réalisations matérielles. Les critères d’organisation sont par exemple, la sédentarisation (village et ville), une hiérarchie dans la population (classes), la création d’un état (organisation étatique), mais également le stockage des surplus de productions (céréales). Les critères de réalisations matérielles concernent, l’écriture, les monuments, le commerce, les connaissances scientifiques (arithmétique, géométrie, astronomie).

Les premières civilisations ont chacune apporté des espèces cultivées. Nous avons hérité de ces cultures et nous les perpétons. On remarquera qu’il a fallu adapter certaines espèces à d’autres environnements que ceux de leurs zones géographiques d’origine (par exemples le blé, le riz, le maïs, le pommier etc.).



Fig. II-2 Alphonse de Candolle (1806-1893).

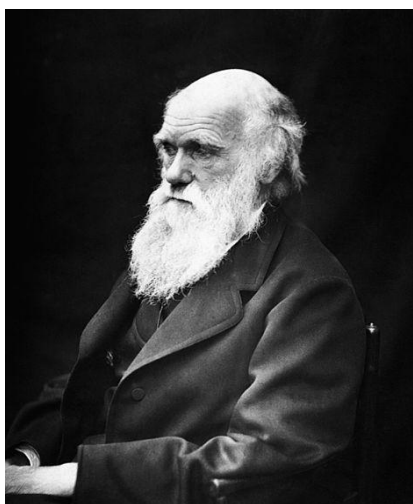


Fig. II-3 Charles Darwin (1809-1882).

La volonté de comprendre le(s) processus de domestication n'est pas simple. Il y a un grand nombre de scientifiques qui, aujourd'hui encore, cherchent à valider des théories élaborées il y a plus d'un siècle. Ces théories ont permis la naissance de différents courants de pensée et leurs défenseurs actuels utilisent tous les moyens possibles (technologiques y compris) pour les prouver. Commençons par le début.

II-2 Origine des plantes cultivées :

Quand on fait une étude bibliographique sur les origines des plantes cultivées, on ne peut pas ignorer les premières théories faites par Alphonse Pyrame de Candolle (1806-1893) (voir Fig. II-2) et Charles Robert Darwin (1809-1882) (Fig. II-3). Les deux personnages sont importants car ils sont précurseurs sur la découverte de l'origine des plantes cultivées.

II-2-1 Les théories d'Alphonse de Candolle sur l'origine des plantes cultivées :

II-2-1-1 La méthode d'Alphonse de Candolle :

Alphonse Pyrame de Candolle (1806-1893) est un botaniste genevois. En 1835, il est nommé professeur ordinaire à l'Académie de Genève, prenant la suite de son père, et va se consacrer à la botanique jusqu'à sa mort (Chauvet, 1998).

Quand on lit le livre d'Alphonse de Candolle intitulé « Origine des plantes cultivées, 1883 », il écrit dans son introduction: *“ Les traditions des anciens peuples, embellies par les poètes, ont attribué communément les premiers pas dans la voie de l'agriculture et l'introduction de plantes utiles à quelque divinité ou tout au moins à quelque grand empereur ou Inca. On trouve en réfléchissant que ce n'est guère probable, et l'observation des essais d'agriculture chez les sauvages de notre époque montre que les faits se passent tout autrement. En général, dans les progrès qui amènent la civilisation, les commencements sont faibles, obscurs et limités. Il y a des motifs pour que cela soit ainsi dans les débuts agricoles ou horticoles. Entre l'usage de récolter des fruits, des graines ou des racines dans la campagne et celui de cultiver régulièrement les végétaux qui donnent ces produits, il y a plusieurs degrés. ”*

Il propose également la raison pour laquelle les foyers de naissance de l'agriculture sont limités:

"Les causes variées qui favorisent ou contrarient les débuts de l'agriculture expliquent bien pourquoi certaines régions se trouvent, depuis des milliers d'années, peuplées de cultivateurs, tandis que d'autres sont habitées encore par des tribus errantes. Évidemment, le riz et plusieurs légumineuses dans l'Asie méridionale, l'orge et le blé en Mésopotamie et en Égypte, plusieurs Panicées (panicaceae), en Afrique, le maïs, la pomme de terre, et le manioc en Amérique ont été promptement et facilement cultivés, grâce à leurs qualités évidentes et à des circonstances favorables de climat. Il s'est formé

ainsi des centres d'où les espèces les plus utiles se sont répandues. Dans le nord de l'Asie, de l'Europe et de l'Amérique, la température est défavorable et les plantes indigènes sont peu productives; mais comme la chasse et la pêche y présentaient des ressources, l'agriculture a dû s'introduire tard, et l'on a pu se passer des bonnes espèces du midi sans souffrir beaucoup. Il en était autrement pour l'Australie, la Patagonie et même l'Afrique australe. Dans ces pays, des plantes des régions tempérées de notre hémisphère ne pouvaient pas arriver à cause de la distance, et celles de la zone intertropicale étaient exclues par la grande sécheresse ou par l'absence de températures élevées."

Dans cet ouvrage, le fait le plus marquant est la mise en place par Alphonse de Candolle d'un protocole d'investigations. Pour la première fois dans son domaine, il propose de coupler les méthodes de botaniques et les méthodes d'archéologies. Il considère un certain nombre de critères et de procédés pour chacune de ces méthodes, qu'il compare les unes aux autres. Au final, chaque méthode permet d'obtenir un résultat probable et c'est l'ensemble des voies pour un résultat identique qui valide ce résultat. Je commencerai par les différents critères qu'il décrit pour les méthodes botaniques. Ils sont au nombre de cinq, on trouve :

1. La géographie botanique : l'origine géographique d'une espèce cultivée peut-être établie par la recherche d'un état sauvage adapté à l'environnement. Mais il faut éviter un certain nombre de pièges tels que la dispersion de plantes cultivées qui deviennent des espèces adventives.
2. La date d'apparition d'une espèce indigène (la dissémination par les animaux, les hommes, etc.) pour cerner les espèces « naturalisées » récemment qui sont anciennes mais pas d'origine.
3. Les migrations : les espèces cultivées peuvent avoir changé de pays avant leur culture et/ou changé de formes.
4. La diversité pour une région, le nombre d'espèces contenues dans un genre permet de distinguer s'il y a eu transfert d'une espèce dans un autre lieu géographique.
5. L'observation du phénotype : il faut se débarrasser des nouvelles formes apportées par la sélection pour comparer des individus (taille des fruits, etc.). En effet une plante cultivée offre de la diversité principalement pour les parties pour lesquelles on la cultive.

Pour les méthodes faisant appel à l'archéologie, il écrit que la preuve la plus directe que l'on puisse imaginer de l'existence ancienne d'une espèce dans un pays est d'en voir des fragments reconnaissables dans de vieux édifices ou de vieux dépôts, d'une date plus ou moins certaine. La dernière méthode qu'il fait intervenir est la paléontologie. Pour cette méthode, il écrit que bien qu'antérieures à l'agriculture, l'utilisation des informations apportées par les différentes couches minérales permet de connaître la succession des formes végétales dans divers pays. Ces informations permettent de délimiter l'apparition de certaines espèces.

La découverte de documents historiques sert principalement à dater l'existence de cultures, découvrir leurs origines géographiques et les migrations qui ont eu lieu. Dans cet ouvrage, Alphonse de Candolle décrit méticuleusement la provenance des plantes cultivées en les classant suivant les parties d'intérêts agronomiques. On trouve :

- ✓ Les parties souterraines : racines (Radis, Navet), bulbes ou tubercules (Topinambour, Pomme de terre).
- ✓ Les tiges et les feuilles : légumes (Chou ordinaire, Epinard), fourrages (Trèfle, Luzerne...).
- ✓ Les fleurs ou les organes qui les enveloppent (Giroflier, Houblon ...).
- ✓ Les fruits (Pomme ; cannelle, Corossol...).
- ✓ Les graines (Cacaoyer, Litchi, Sarrasin, Blé ...).

Je ne décrirai pas tous les résultats obtenus par Alphonse de Candolle dans cet ouvrage, par contre lors de la lecture il ressort un certain nombre de conclusions extraordinaires par leurs précisions.

Sur 247 espèces cultivées étudiées (51 familles phanérogames), 199 provenaient du vieux continent, 45 d'Amérique et trois étaient non déterminées.

Bien qu'il existe des espèces cultivées avec une région d'origine unique, le plus souvent il existe plusieurs foyers tout comme pour les espèces non cultivées. Les espèces les plus anciennes étaient au nombre de 44 (il y a plus de 2000 ans) dont 22 annuelles (probablement les plus faciles et les plus rentables au début de l'agriculture). Quand on compare ces cultures par rapport aux plantes domestiquées depuis moins de 2000 ans avant JC (61 plantes à l'époque de l'auteur), aucune ne rivalisent en qualité nutritive et en rendement avec les plus anciennes (le blé, le riz, le maïs, la Pomme de terre, etc.).

L'auteur constate que si l'on divise les plantes cultivées en deux groupes, celles domestiquées il y a plus de 4000 ans et celles domestiquées il y a plus de 2000 ans, l'identification des parents à l'état sauvage aboutit à 63% de réussites pour les plus anciennes et 83% pour les plus récentes. Alphonse de Candolle est surpris par le fait que les espèces cultivées anciennes n'ont pas dévié d'avantage des espèces sauvages, ce qui aurait pu empêcher de découvrir les liens de parenté.

Il propose pour cela deux causes:

La période est trop courte pour observer une dérive génétique entre les deux espèces.

Une forte « naturalisation » (adaptation) des espèces cultivées qui se croisent avec les espèces sauvage.

Pour certaines espèces (le blé, le maïs, le tabac, les lentilles ...), il n'arrive pas à trouver le parent sauvage et constate qu'elles ne peuvent subsister sans la main de l'homme. De plus il n'y a pas d'adaptation au froid. Ce sont des décalages par rapport aux saisons dans chaque région qui ont permis les cultures vers le nord grâce à une adaptation à la photopériode.

Il dit que les botanistes ne devraient pas classer les variétés suivant des caractères agronomiques qui varient le plus (couleurs, formes, tailles, saveurs ...) mais les caractères fixes ou organes qui ne sont pas concernés par la sélection pour plus facilement identifier les espèces sauvages. Il conclut son œuvre en affirmant qu'il n'y a eu aucun échange de plantes cultivées entre l'ancien et le nouveau monde jusqu'à l'arrivée de Christophe Colomb.

II-2-1-2 Origine des blés :

Il me paraît important de présenter comment Alphonse de Candolle a retracé l'origine des blés, même si ceci a été mis à jour depuis. De Candolle utilise la classification des blés de Vilmorin (Essai d'un catalogue méthodique et synonymique des froments, Paris, 1850). Cette classification répartit les blés en deux groupes, les blés battables, possédant des grains facilement détachables de l'enveloppe et les non-battables demandant un effort mécanique conséquent pour obtenir le même résultat.

II-2-1-2-1 Dans la catégorie des blés facilement battables Vilmorin définit quatre groupes :

- ✓ Le froment ordinaire (*T. vulgare*, Villars - *T. hybernicum* et *T. aestivum*, Linné)
- ✓ Le gros blé, Pétanielle ou Poulard (*T. turgidum*, Linné)
- ✓ Le blé dur (*T. durum*, Desfontaines)
- ✓ Le blé de Pologne (*T. polonicum*, Linné)

Alphonse de Candolle pense que les quatre groupes n'en font qu'un pour une seule et même espèce.

Le froment ordinaire (*T. vulgare*, Villars - *T. hybernicum* et *T. aestivum*, Linné)

Ces recherches lui permettent de trouver deux sources d'indices capitales. Le premier vient d'un document (Syncelli) qui mentionne que « d'après le plus ancien de tous les historiens, Bérose, prêtre de Chaldée (III^e siècle av. J.C.), dont Hérodote a conservé des fragments, on voyait dans la Mésopotamie, entre le Tigre et l'Euphrate, le froment sauvage ».

Le deuxième document (Olivier, Voyages dans l'Empire Ottoman (1807), vol 3, p 460) mentionne les récits d'un explorateur, « Olivier étant sur la rive droite de l'Euphrate, au nord-ouest d'Anah, pays impropre à la culture, trouva dans une sorte de ravin le froment, l'Orge et l'Epeautre que nous avons déjà vus plusieurs fois en Mésopotamie ».

Alphonse de Candolle par sa méthode de recoupement, trouve deux informations de vingt-trois siècles d'écart aboutissant à une même région, il en vient à penser que la culture du froment date de la préhistoire car le nom se retrouve dans de nombreuses langues anciennes. Ce résultat permet de valider le protocole de recherche qu'il a mis en œuvre, c'est-à-dire l'addition et le recoupement d'informations issues de l'archéologie et de la botanique, il y a cent trente ans.

Le gros blé, Pétanielle ou Poulard (*T. turgidum* et *T. compositum*, Linné)

Ces recherches n'aboutissent pas car il ne trouve pas d'espèce sauvage. Il en conclut faussement que *T. turgidum* est un blé provenant de *T. aestivum*.

Le blé dur (*T. durum*, Desfontaines) et **le blé de Pologne** (*T. polonicum*, Linné)

L'incite à faire les mêmes conclusions que pour le gros blé par manque d'informations.

II-2-1-2-2 Les blés non facilement battables de Henri Vilmorin (3 groupes) :

- ✓ **L'Epeautre, Grande Epeautre, Dinkel** (*T. spelta*, Linné).

Il trouve quelques informations, mais ses recherches s'arrêtent en Europe orientale tempérée et dans une partie de l'Asie. Mais il conclut que l'épeautre serait une forme intermédiaire du froment issu d'une époque préhistorique pas très ancienne (car deux grains par épillet).

- ✓ **L'Amidonier, Emmer** (*T. turgidum* ssp. *dicoccum*, Schrank, *T. turgidum* ssp. *amyleum*, Seringe). Il conclut que c'est une ancienne race de l'épeautre (*T. aestivum* ssp. *spelta*) car deux grains par épillet, que l'on ne trouve qu'en Europe pour sa résistance au froid.

- ✓ **L'engrain, Petit Epeautre, Locular, Einkorn** (*T. monococcum* ssp. *monococcum*, Linné).

Il est considéré comme une espèce différente car il n'y a qu'un grain par épillet. De plus Henri de Vilmorin n'arrivait pas à le croiser avec d'autres épeautres.

Si on résume, Alphonse de Candolle pense que la séparation entre les froments à grains libres et les Epeautres serait antérieure aux données historiques et peut-être au commencement de toute agriculture. En conclusion, les froments proviendraient en premier d'Asie, tandis que les Epeautres proviendraient par la suite de l'Europe orientale et de l'Anatolie. L'Epeautre le plus ancien serait *T. monococcum*, les autres se seraient éloignés à la suite de plusieurs milliers d'années de culture et de sélection.

II-2-1-3 Conclusion sur les travaux d'Alphonse de Candolle :

Alphonse de Candolle ne possédait pas les moyens dont nous disposons actuellement, c'est la raison pour laquelle certaines de ces conclusions s'avèrent fausses en partie, même s'il essaie d'expliquer que toutes les espèces du blé sont liées. Mais l'impact de ses recherches et sa méthodologie restent plus que valables avec l'ajout d'informations issues d'autres disciplines comme la cytologie, la biochimie, la biologie moléculaire, etc...

II-2-2 Darwin et l'origine des plantes cultivées :

Maintenant, parlons des travaux de Charles Darwin (1809-1882), un autre grand évolutionniste de la même période qu'Alphonse de Candolle dans la recherche des origines des plantes cultivées. Les deux hommes sont contemporains et ont souvent été d'accord. Quand il y avait opposition d'idées entre les

deux hommes c'était sur des détails. Par exemple pour Alphonse de Candolle « le choix des espèces a plus d'importance que la sélection des variétés ». De plus il n'accepte pas l'idée qu'une espèce sauvage puisse être à l'origine d'une autre espèce sauvage.

II-2-2-1 Les conclusions de Darwin :

Dans les chapitres consacrés à la domestication des plantes cultivées dans son ouvrage de « De l'Origine des espèces » (1859), Charles Darwin dresse une longue liste de théories et leurs auteurs concernant l'évolution des êtres. Les premières études faites sur l'évolution des animaux domestiques et des plantes cultivées sont celles qui lui ont permis de développer sa théorie sur l'origine des espèces. Pour Charles Darwin, c'est « le meilleur champ de recherche pour expliquer cet obscur problème ». Il commence par rechercher les causes de la variabilité et conclut en écrivant que « Les conditions de la vie paraissent agir de deux façons distinctes : directement sur l'organisation entière ou sur certaines parties seulement, et indirectement en affectant le système reproducteur ».

L'action directe comporte deux facteurs « la nature de l'organisme et la nature des conditions ».

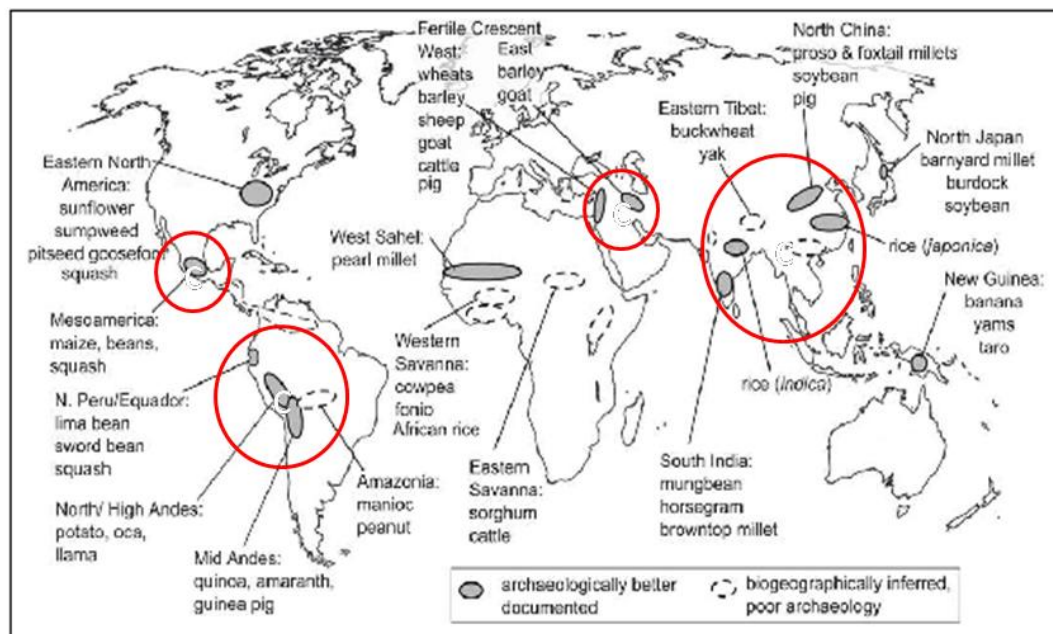
La nature de l'organisme est le plus important selon Darwin, car des variations presque semblables se produisent quelquefois dans des conditions différentes (cas de la domestication), mais il reconnaît qu'à l'inverse des variations différentes, se produisent dans des conditions uniformes. Les effets peuvent être définis (quand tous les descendants se modifient de la même manière) ou indéfinis (apparition d'innombrables particularités). Dans le cas de la domestication, les effets sont majoritairement indéfinis et ne proviennent pas des parents. Ces déviations ne concernent que quelques individus dans la population.

Pour l'action indirecte (modifications affectant le système reproducteur), selon Darwin le système reproducteur est sensible aux changements des conditions extérieures (les plantes sauvages cultivées ou les animaux en captivité sont moins féconds). Il s'interroge sur les mécanismes en jeux. Il dit que tout caractère, quel qu'il soit, se transmet par hérédité et que la non transmission est l'exception.

Il sous entend le rôle des gènes dominants et récessifs dans la transmission à la descendance. Darwin avait un sens de l'observation aiguisé.



Fig. II-4 Nicolai Vavilov (1887-1943).



2012 by the American Anthropological Association

Fig. II-5 Les 12 principaux centres d'origines de la domestication des plantes dans le monde.

(inspiré de (Fuller, 2011))

II-2-3 Avancées depuis Alphonse de Candolle et Charles Darwin :

Alphonse de Candolle et Charles Darwin ne tiennent pas compte des explications apportées par les lois de Mendel. En effet, les trois savants sont contemporains mais leurs théories (sur l'évolution pour Alphonse de Candolle et Charles Darwin, et sur la génétique pour Mendel) vont coexister durant 35 ans avant que d'autres scientifiques tels qu'Hugo de Vries, Carl Erich Correns et Erich Von Tschernak les englobent. Ce qu'il faut retenir, c'est que leurs idées sont devenues la base de nos recherches actuelles.

Avant de faire un bilan sur les connaissances actuelles, j'évoquerai rapidement les travaux d'un autre scientifique du siècle dernier, Nicolai Vavilov (1887-1943) (Fig. II-4), généticien et botaniste, il est à l'origine du concept qu'il y aurait sept « centres d'origines », sept berceaux de végétaux dans le monde pour lesquels il va constituer après prospection de vastes collections de plantes. Il s'est spécialisé dans l'identification des variétés cultivées sur des critères morphologiques, cytologiques, de résistance aux maladies et d'adaptation à l'environnement. Sa théorie était que le centre d'origine est la zone géographique ayant le plus de variétés d'une même espèce.

II-2-4 Les connaissances actuelles sur l'origine des blés :

Actuellement, suite aux récentes découvertes, les conclusions sur l'origine géographique des plantes ou animaux domestiqués sont acceptées par toute la communauté scientifique. Les divergences portent sur le processus de domestication et la durée pendant laquelle l'homme est passé du statut de chasseur-cueilleur nomade, de sédentaire à celui de cultivateur. Pour comprendre les différents courants de pensées sur l'évolution de la domestication, je me concentrerai principalement sur les plantes. Pour les plantes domestiquées les plus anciennes, il y a quatre lieux géographiques qui ressortent (Brown et al., 2009) (Fig. II-5) :

- ✓ Le Proche-Orient (le croissant fertile) pour le blé, l'orge, les lentilles, les pois-chiches, etc.
- ✓ L'Extrême-Orient pour le riz, le millet, etc.
- ✓ L'Amérique centrale pour le maïs, la courge, le haricot, etc.
- ✓ L'Amérique du sud pour la pomme de terre, le quinoa etc.

Sur les 300 000 espèces de plantes estimées, 3000 espèces sont domestiquées dont 30 espèces représentent les principales cultures mondiales. Encore de nos jours beaucoup d'espèces sauvages sont encore utilisées à l'état sauvage comme médicaments ou nourritures.



Fig. II-6 Le croissant fertile :

Le croissant fertile couvre le Liban, la Syrie, le sud de la Turquie, l'Iran, l'Irak, la Jordanie et Israël.
(Source : Map courtesy of the university of Texas Libraries.)

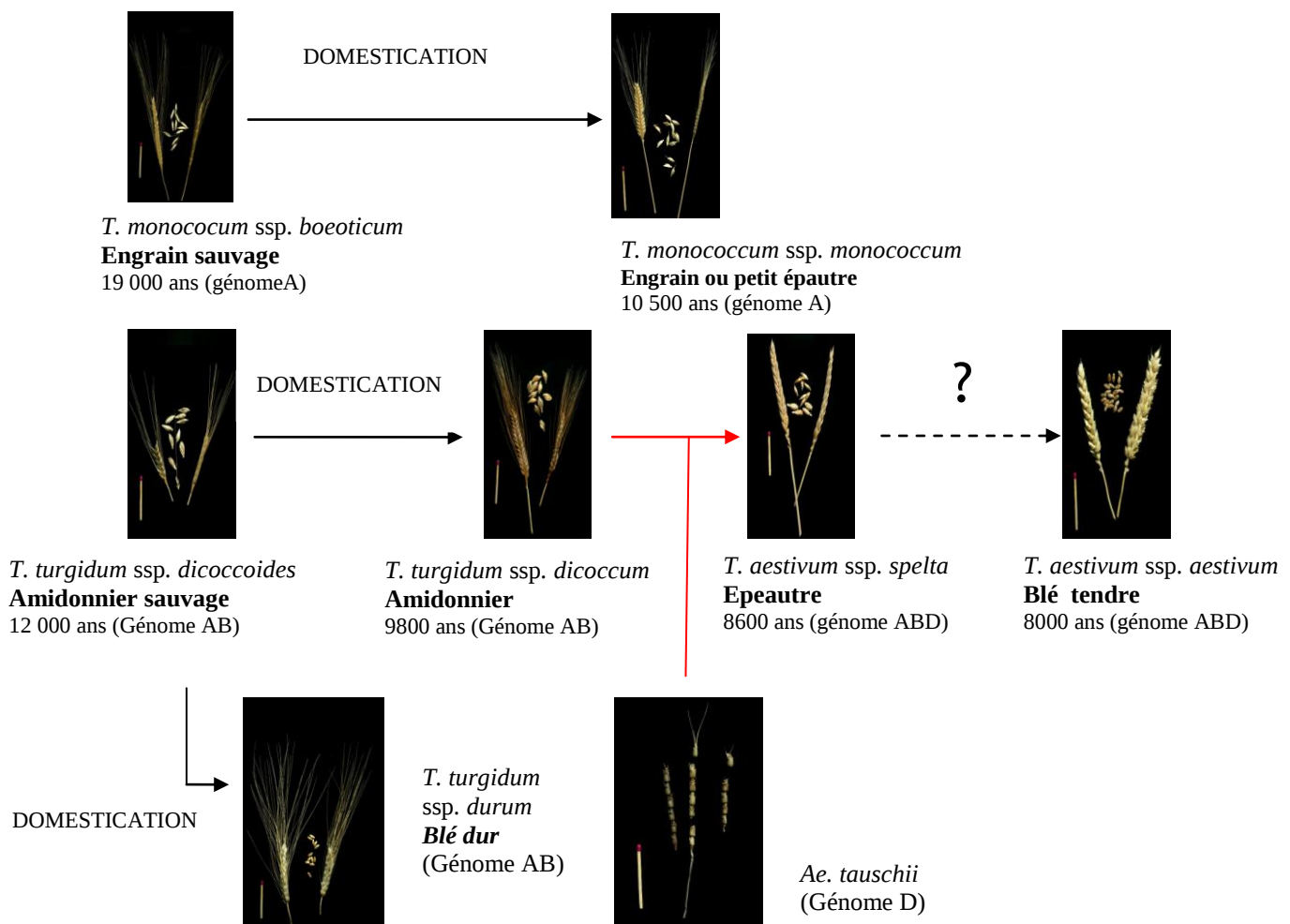


Fig. II-7 Les blés domestiqués et leurs liens avec les blés sauvages.

Les gènes majeurs de domestication (*Q/q*) chez les blés polyploïdes
Harry Belcram

II-3 Les changements apportés lors de la domestication :

Les différences entre les plantes sauvages et celles qui sont domestiquées sont qualitatives et quantitatives. Les plantes sauvages ont évolué en relation avec leur environnement (biotope), comme par exemple la nature du sol, le climat, la présence d'autres plantes et d'animaux. Chaque plante s'est adaptée à une niche écologique particulière de l'écosystème. On peut émettre l'idée que les plantes domestiques proviennent d'un ancêtre commun adapté à une niche écologique précise. Par conséquent, les plantes domestiquées et leurs ancêtres sauvages ont les mêmes difficultés face à la compétition dans une autre niche écologique, par exemple une forêt ou une prairie (Bunting, 1960 ; Harlan and De Wet, 1975). Cette théorie a été appliquée de nombreuses fois pour les plantes cultivées tel que le riz, le blé, les légumes, etc. Dans le cas de certaines plantes médicinales, on ne peut pas parler de domestication uniquement sur le fait qu'elles soient cultivées dans des jardins ou potagers. La domestication est avant tout la sélection de certains caractères provenant de mutations tels que la couleur, la taille, le goût, des fruits plus gros, plus sucrés ou encore la non dissémination des graines. On peut parler de la naissance, il y a plusieurs milliers d'années, du métier de sélectionneur inconscient (comme Darwin l'avait défini).

II-4 Les connaissances actuelles sur le processus de domestication du blé :

Actuellement, tout le monde s'accorde pour dire que la domestication du blé a pour origine le croissant fertile, c'est-à-dire une région qui comprend le Liban, la Syrie, l'Irak, l'Iran et le sud de la Turquie (Fig. II-6). La famille des *Poaceae* est apparue il y a environ 50-70 million d'années (Ma) (Kellogg, 2001) et la sous-famille des *Pooideae* (comprenant le blé, l'orge, le seigle ...) a divergé il y a environ 20 Ma (Inda et al., 2008). L'espèce du blé tétraploïde *T. Turgidum* ssp. *diccocoïdes* apparaît il y a environ 300 000-500 000 ans (Huang et al., 2002 ; Dvorak and Akhunov, 2005).

Les premiers blés cultivés étaient l'engrain sauvage (*T. monococcum* ssp. *beoticum*) il y a 19 000 ans avant J.C., l'engrain (domestique) (*T. monococcum* ssp. *monococcum*) il y a 10 500 ans, l'amidonnier sauvage (*T. turgidum* ssp. *dicoccoïdes*) il y a 12 000 ans (Feldman and Kislev, 2007), suivi de l'Amidonnier (domestique) (*T. turgidum* ssp. *dicoccum*) il y a 9 800 ans. L'amidonnier sauvage est également à l'origine du blé dur (*T. turgidum* ssp. *durum*) (Damania, 1998). Par la suite l'amidonnier domestiqué (*T. turgidum* ssp. *dicoccum*) s'est croisé avec *Ae. tauschii* pour donner l'épeautre (*T. aestivum* ssp. *spelta*) et le blé tendre (*T. aestivum* ssp. *aestivum*) (Dvorak et al., 1998 ; Matsuoka and Nasuda, 2004; Alo et al., 2011) (Fig. II-7).

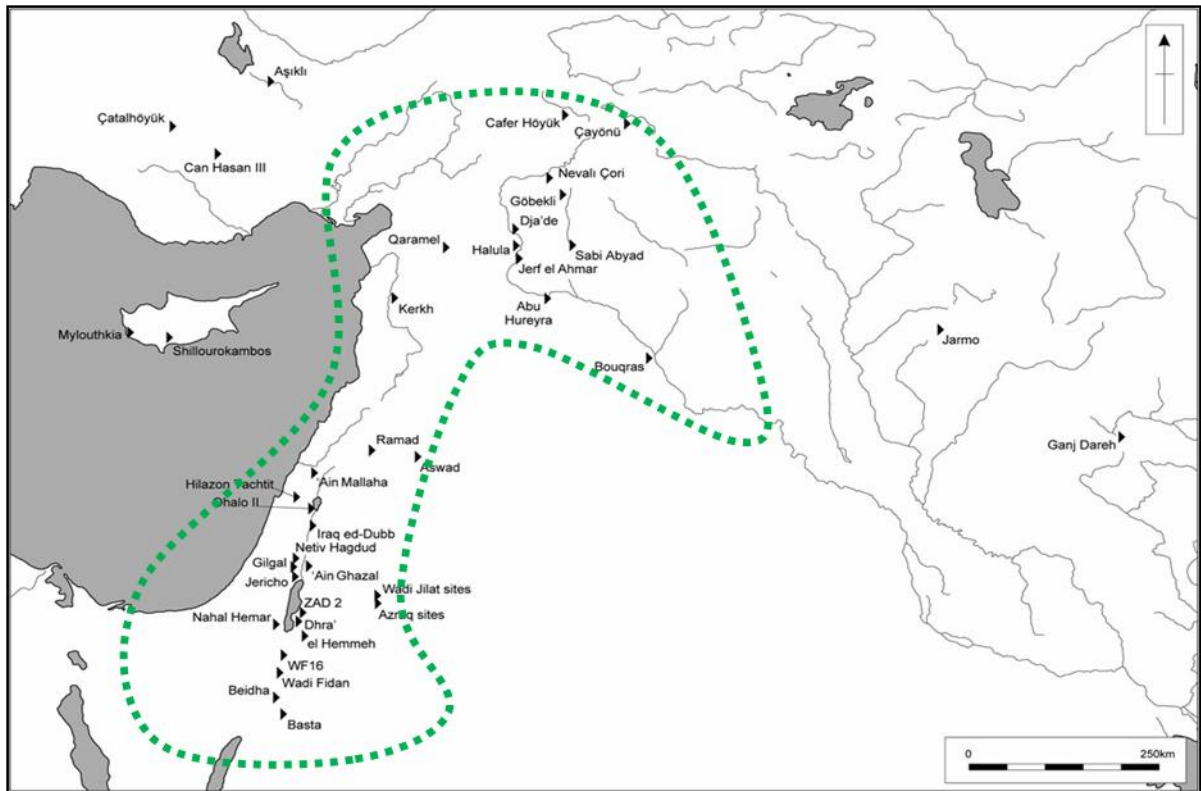


Fig. II-8 Localisations des sites de fouilles archéologiques dans le croissant fertile.
(Fuller, 2011)

Maintenant les idées divergent sur le processus de domestication lui-même. Il existe un certain nombre de paramètres pour lesquels il y a mésentente au sein de la communauté scientifique (Peleg et al., 2011).

Les paramètres répertoriés sur le processus de domestication permettent de faire une bonne synthèse des théories actuelles inspirées d'une méthode inventée par Claude Lévi-Strauss (Lévi-Strauss, 1962). La méthode consiste à créer des critères ou les réponses sont opposées (bidirectionnelles). Ces paramètres sont au nombre de cinq:

- ✓ La distribution (dans la région): Le processus de domestication provient d'un unique foyer ou de multiples foyers géographiques.
- ✓ la fréquence: Le processus de domestication s'est déroulé en une fois, ou c'est le résultat de plusieurs essais.
- ✓ la vitesse : Cela s'est produit rapidement ou c'est le résultat d'un processus lent et échelonné.
- ✓ l'échelle : C'est un processus qui englobe plusieurs espèces cultivées ou chaque espèce cultivée a subi une domestication indépendante.
- ✓ le catalyseur : le processus a été intentionnel (conscient) ou non intentionnel (inconscient).

En générale on observe deux grands courants de pensée où les critères peuvent être associés comme suit :

- Un processus rapide, localisé, comprenant de nombreuses espèces différentes (Lev-Yadun et al., 2000 ; Ozkan et al., 2002; Abbo et al., 2010b ; Abbo et al., 2010a).
- Un processus lent et graduel, avec des événements répétés dans des foyers multiples (géographiquement) pour des espèces différentes de façons indépendantes (Ladizinsky, 1998b ; Nesbitt, 2004 ; Tanno and Willcox, 2006 ; Weiss et al., 2006; Kilian et al., 2007 ; Morrell and Clegg, 2007 ; Allaby, 2010 ; Fuller et al., 2010).

Avant de détailler les principaux arguments des uns et des autres, il me paraît indispensable de faire un bilan des recherches archéologiques actuelles dans le «levant» (le nord de l'Irak, le sud de la Turquie, le nord de la Syrie, le Liban, Israël et la Jordanie). Il existe de nombreux sites de fouilles archéologiques dans certains pays (Fig. II-8) (Gopher et al., 2001; Fuller, 2011) tel que :

- ✓ Nemrik, Qermez Dere et M'lefaal dans Le nord de l'Irak
- ✓ Hallan Cemi Tepesi, Cayönü, Cafer Höyük, Nevali Cori, Göbekli Tepe, Dja'de et Karacadag dans le sud de la Turquie
- ✓ Jerf el Ahmar, Mureybet, Tell Abu Hureyera et Tell Aswad en Syrie
- ✓ Nemrik, M'lefaat et Qermez Dere dans le nord de L'Irak
- ✓ Ohalo II, Yiftahel et Mallaha en Israel
- ✓ Jericho, Netiv Hagdud, et Gilgal en Jordanie

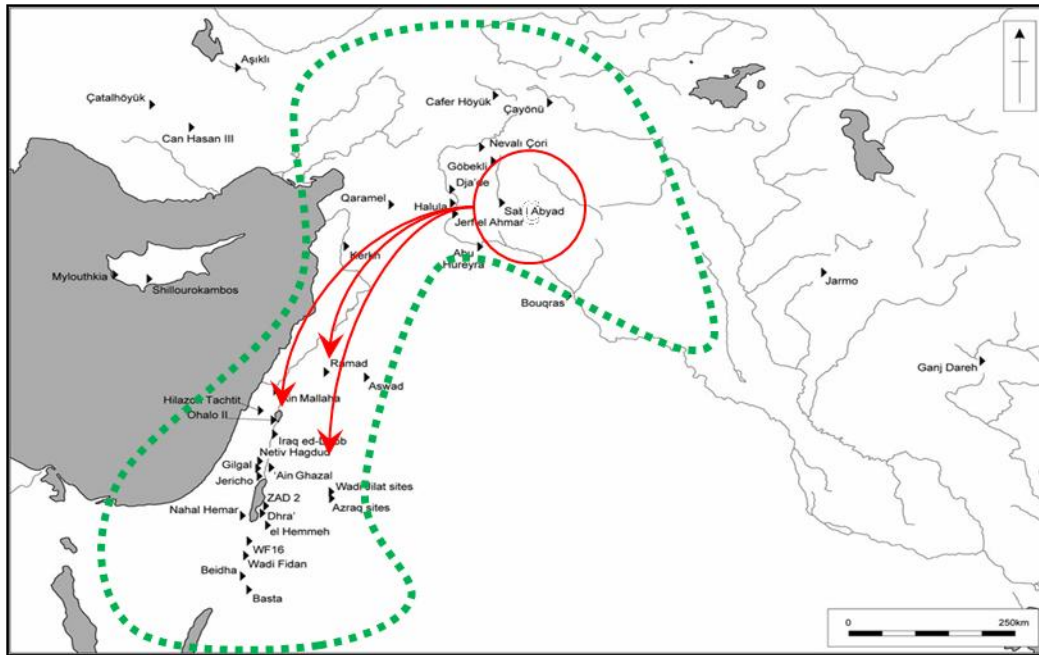


Fig. II-9 La domestication : Un processus rapide, localisé, comprenant de nombreuses espèces différentes :

Identification du “core area” décrit par (Zohary, 1996 ; Abbo et al., 2001 ; Gopher et al., 2001 ; Abbo et al., 2006 ; Abbo et al., 2010a).

Localisation des sites de fouilles archéologiques (Fuller, 2011)

A partir de l'ensemble des découvertes (outils, objets divers (meules, pilons, mortiers, faucilles, fondations d'habitats...)), l'identification d'une communauté « les Natoufiens » a pu être faite, ainsi qu'une chronologie de leurs pratiques sociales (chasse, cueillette, inhumation, etc.).

Entre la fin du paléolithique et le début du néolithique soit 10 000 ans av. J. C., des périodes intermédiaires ont été créées au fur et à mesure des découvertes archéologiques pour créer des points de repères temporelles. Le mot « natoufien » décrit également une période entre 12 800 et 10 000 ans av. J.-C. (épipaléolithique), caractérisée par les premières expériences de sédentarisation d'une communauté de nomades chasseurs-cueilleurs. Parmi les périodes créées, on trouve également le Néolithique précéramique A (PPNA ou pre-pottery neolithic A) entre 10 000 et 9000 ans av. J.C., avec les premières traces d'agriculture, puis le Néolithique précéramique B (PPNB) entre 9000 et 8000 ans av. J.-C., avec l'apparition de poteries, le développement de l'élevage et des habitations d'une architecture plus complexe (apparition d'un étage).

La communauté Natoufienne a souffert d'un changement climatique entre 10 800 et 9600 ans av. J.C., une baisse des températures de 7°C (mini glaciation). Cela a provoqué une diminution de la nourriture (chasse et cueillette) et un retour momentané au nomadisme. Après cette période, la communauté Natoufienne est retournée au sédentarisme et a continué son évolution par la création de sociétés de plus en plus complexes.

Maintenant, je décrirai les deux théories élaborées à partir des informations obtenues sur les différents sites.

II-4-1 Un processus rapide, localisé, comprenant de nombreuses espèces différentes :

Dans une série d'articles sur le processus de domestication des plantes dans le croissant fertile (Zohary, 1996 ; Abbo et al., 2001 ; Gopher et al., 2001 ; Abbo et al., 2006 ; Abbo et al., 2010a), les auteurs utilisent une méthode développée sur l'évolution de la relation homme-plante dans le temps, à l'aide de données issues de différentes disciplines (anthropologie, botanique, socio-économie, écologie, etc.). Les auteurs trouvent trois phases dans le processus de domestication (Harris and Hillman, 1989) :

- ✓ La cueillette
- ✓ La culture d'espèces sauvages : développement des techniques agricoles (« cultivation »)
- ✓ La culture d'espèces domestiquées : sélection des variétés (« domestication »)

Les trois phases se déroulent dans des régions différentes, mais la période de culture d'espèces sauvages se fait à partir d'une unique région (« core area ») entre deux fleuves (le Tigre et l'Euphrate) au sud de la Turquie qui serait à l'origine de la domestication « northern Levant » (Fig. II-9). Dans cette région il y aurait eu au départ la constitution d'une collection de plantes sauvages à l'origine d'espèces cultivées tel que l'Engrain, *Triticum monococcum* ssp. *monococcum* L. (Heun et al., 1997),

le pois, *Pisum sativum* (Zohary and Hopf, 2000), les lentilles, *Lens culinaris Medicus* (Ladizinsky et al., 1995) et le pois-chiche, *Cicer arietinum* (Ladizinsky et al., 1995). Puis par diffusion cette collection d'espèces sauvages s'est propagée vers d'autres régions plus au sud (« Southern levant »), et également les espèces domestiquées et sélectionnées par la suite. Un des arguments des auteurs serait que les espèces de cette collection étaient inséparables car ils présentaient l'avantage d'être complémentaire en acides aminés essentiels pour la nutrition. Les auteurs insistent sur le fait que c'est à partir de leur sédentarisation et le développement d'une culture d'échange que les natoufiens ont pu démarrer le processus de domestication des plantes sur 1500 ans. Ils disent qu'il y a autant d'évènements de domestication que de caractères recherchés chez une espèce, mais que la recherche étant intentionnelle (transfert d'idées, transfert de matériels, mariages entre communautés), la majorité des espèces ont été domestiquées sur une même période puis diffusées dans tout le croissant fertile.

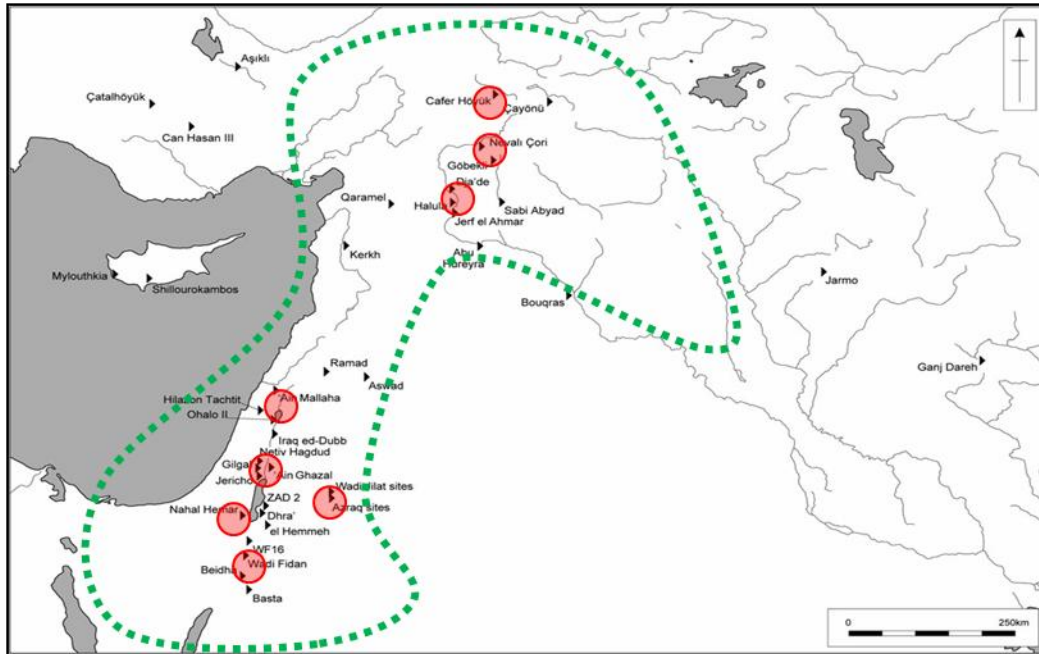


Fig. II-10 La domestication : Un processus lent et graduel, avec des évènements répétés dans des foyers multiples (géographiquement) pour des espèces différentes de façon indépendante.

(Harlan and Zohary, 1966 ; Ladizinsky, 1998a, 1998b ; Nesbitt, 2004 ; Tanno and Willcox, 2006 ; Morrell and Clegg, 2007 ; Allaby et al., 2008 ; Allaby, 2010 ; Fuller et al., 2010).

Localisation des sites de fouilles archéologiques (Fuller, 2011)

II-4-2 Un processus lent et graduel, avec des événements répétés dans des foyers multiples (géographiquement) pour des espèces différentes et de façon indépendante :

Cette théorie est en opposition avec la première citée, elle est menée par un ensemble de scientifiques (Harlan, 1973a ; Ladizinsky, 1998a, 1998b ; Nesbitt, 2004 ; Tanno and Willcox, 2006 ; Morrell and Clegg, 2007 ; Allaby et al., 2008 ; Allaby, 2010 ; Fuller et al., 2010). Par l'analyse des données pour chaque site de fouilles archéologiques, les auteurs arrivent à la conclusion qu'il y a de multiples centres d'origines du processus de domestication dans le nord et dans le sud du croissant fertile et qu'il y a eu différentes épisodes d'adaptation (Fig. II-10). Les différents sites donnent des résultats hétérogènes dans le processus de domestication lui-même, la cause principale serait les changements climatiques observés. Ces changements auraient amplifié la disparité des comportements (chasseurs-cueilleurs $\leftrightarrow$ sédentaires) au sein des communautés. Les auteurs pensent également que les collections d'espèces sauvages ont été constituées lors du regroupement d'individus sur des sites plus grands, ce qui a permis d'intensifier et de diversifier les cultures pour compenser l'augmentation de la population qui a pu être un élément moteur de la domestication.

II-4-3 Que peut-on conclure sur ces deux hypothèses ?

Les deux théories sont valables, le problème reste un manque d'informations pour confirmer l'une ou l'autre. Les deux principales sources de données pour l'argumentation sont la datation (des sites archéologiques et des éléments trouvés), la paléobotanique, les recherches faites en génétique (génomique comparée) entre les différentes espèces sauvages et les espèces domestiquées (après prospection) et la paléoclimatologie.

Pour la datation, en archéologie, hormis de rares documents, on utilise la datation par le carbone 14. Les âges sont exprimés en nombre d'années avec l'expression « calcul avant le présent », « calcul before present », ou la mention « cal BP ». L'inconvénient de la technique est que les résultats bruts des datations ne tiennent pas compte des fluctuations du taux de radiocarbone dans l'atmosphère au cours du temps. C'est la raison pour laquelle il faut corriger les résultats bruts. Il existe différents logiciels (ex : CalPal) pour cette correction qui n'est pas linéaire. Le résultat final est un intervalle avec une probabilité. Dans le cadre des recherches sur le processus de domestication des plantes dans le croissant fertile, la nature du sol empêche une datation précise et ne permet pas de conclure sur la chronologie des découvertes sur les sites de fouilles et l'évolution des occupants. Les balises deviennent l'apparition ou l'absence de certains objets (exemple la faucille, les céramiques, etc.) d'où la création de périodes artificielles.

Les données fournies par la génomique comparée entre les espèces sauvages et les espèces domestiquées, posent des problèmes qui ne permettent pas de conclure, car il existe une forte conservation de l'ensemble du génome entre les deux formes.

Si on fait référence aux avertissements faits par Alphonse Pyrame de Candolle l'origine géographique d'une espèce cultivée peut-être établie par la recherche d'un état sauvage adapté à l'environnement mais il faut éviter un certains nombres de pièges tels que leurs dispersions ou leurs disparitions dues aux phénomènes climatiques. Or dans les études climatologiques de la région, on observe une période de sécheresse, le "Younger Drays" entre 10 900 et 9600 ans av. J.C. (Alley et al., 1993 ; Sevringhaus et al., 1998), qui a modifié le comportement de la communauté Natoufienne sur certains sites, car des variétés ont pu disparaître ou ont du être abandonnées (Kislev, 1997).

De plus, il ne faut pas ignorer les migrations car les espèces cultivées peuvent avoir changé de régions avant leur culture et/ou changé de formes, aucun résultat n'a été validé. Il faut tenir compte des migrations, car celles-ci ont permis l'expansion des cultures dans d'autres régions.

Pour conclure sur les avertissements faits par Alphonse Pyrame de Candolle, une étude de la diversité pour une région, c'est-à-dire le nombre d'espèces contenues dans un genre permet de distinguer s'il y a eu transfert d'une espèce dans un autre lieu géographique, mais dans le cadre d'un changement climatique, ce qui est le cas entre le nord et le sud du croissant fertile, on obtient une sélection de certaines espèces ou de certaines variétés, ce qui fausserait les résultats de l'étude.

Ces exemples illustrent les difficultés que l'on rencontre dans l'étude du processus de domestication dans le « Levant », mais également dans d'autres régions du monde.

A la lecture d'articles ayant pour sujet « la domestication des plantes dans le croissant fertile » (Nesbitt, 2004 ; Abbo et al., 2006 ; Allaby et al., 2008 ; Abbo et al., 2010a ; Abbo et al., 2010b ; Aiello et al., 2010 ; Özkan et al., 2010 ; Fuller, 2011 ; Abbo et al., 2012 ; Asouti and Fuller, 2012 ; Willcox, 2012a) , on réalise également que la question fondamentale porte sur l'état de conscience des premières communautés. Ceux qui optent pour un processus conscient utilisent comme argument que l'homme en se sédentarisant a changé sa perception de la « nature ». Il passe d'une dépendance totale par rapport à son environnement pour survivre (nomadisme) à un état où il essaie de le contrôler (domestication) pour anticiper sa survie (Ladizinsky, 1987a, 1987b). Cet argument est utilisé pour expliquer que les Natoufiens entre 13 000 et 7000 calBP, se sont installés entre les fleuves « le Tigre » et « l'Euphrate » au sud de la Turquie. Ils auraient récolté, cultivé, établi une collection de plantes, puis ils auraient domestiqué des espèces sauvages et diffusé leur savoir dans tout le Proche-Orient ancien (Abbo et al., 2010b ; Abbo et al., 2010a).

Pour ceux qui, à l'opposé, pensent que l'homme n'était pas totalement conscient de ses actes, les arguments demeurent que par instinct de survie dans différentes régions (de conditions diverses), de petites communautés ont pratiqué la culture d'espèces sauvages. Dans chacune de ces régions, les espèces ont évolué sans que l'homme en aperçoive les variations phénotypiques et les moissonne en mélange (Purugganan and Fuller, 2009). Cette période porte également le nom de « pré-domestication », car inconsciente (Tanno and Willcox, 2006). Les auteurs démontrent également par

Tableau II-1 : Classification actuelle des principaux blés

Frequently used names in older literature	Current Names Based on van Slageren (Aegilops) or Waines & Barnhart (Triticum)	Genomes
A genome species		
Triticum monococcum	Triticum monococcum subsp. monococcum	Am
Triticum uartu	Triticum monococcum subsp. aegilopoides	Au
Triticum aegilopoides		
Triticum boeoticum		
Triticum durum	Triticum turgidum	BAu
Triticum dicoccoides		BAu
Triticum dicoccon		BAu
Triticum carthlicum		BAu
Triticum timopheevii	Triticum timopheevii	GAu
Triticum araraticum		GAu
Triticum zhukovskyi	Triticum zhukovskyi	GAmAu
Triticum aestivum	Triticum aestivum	BAuD
Triticum spelta		BAuD
B genome species		
Aegilops speltoides	Aegilops speltoides Tausch	B
	Aegilops speltoides Tausch var. speltoides	
	Aegilops speltoides var. ligustica (Savign.) Fiori	
Triticum durum	Triticum turgidum	BAu
Triticum dicoccoides		BAu
Triticum dicoccon		BAu
Triticum carthlicum		BAu
Triticum aestivum	Triticum aestivum	BAuD
Triticum compactum		BAuD
Triticum macha		BAuD
Triticum spelta		BAuD
Triticum sphaerococcum		BAuD
Triticum vavilovii		BAuD
Species with C genome		
Aegilops caudata	Aegilops markgrafii (Greuter) Hammer	C
Aegilops triuncialis	Aegilops triuncialis L.	
Aegilops triuncialis	Aegilops triuncialis L. var. triuncialis	UC
	Aegilops triuncialis var. persica	CU
Aegilops cylindrica	Aegilops cylindrica Host	DC

D genome species		
Aegilops tauschii	Aegilops tauschii Coss.	D
Aegilops cylindrica	Aegilops cylindrica	DC
Aegilops ventricosa	Aegilops ventricosa Tausch	DN
Aegilops crassa (4x)	Aegilops crassa Boiss .	DcM
Aegilops crassa (6x)		
Aegilops trivialis (6x)		DcDM, DMS
Aegilops juvenalis	Aegilops juvenalis (Thell.) Eig	DMU
Triticum syriacum	Aegilops vavilovii	DcSsX
Triticum aestivum		BAuD
M genome species		
Aegilops comosa	Aegilops comosa Sm.	M
	Aegilops comosa Sm. var. comosa	
	Aegilops comosa var. subventricosa Boiss.	
Aegilops biuncialis	Aegilops biuncialis Vis.	UM
Aegilops columnaris	Aegilops columnaris Zhuk.	UM
Aegilops crassa (4x)	Aegilops crassa Boiss .	DcM
Aegilops crassa (6x)		
Aegilops trivialis (6x)		DcDM, DMS
Aegilops juvenalis	Aegilops juvenalis (Thell.) Eig	DMU
Aegilops columnaris	Aegilops columnaris Zhuk.	UM
Aegilops recta	Aegilops neglecta	UMN
Triticum syriacum	Aegilops vavilovii	DcSsX
N genome species		
Aegilops uniaristata	Aegilops uniaristata	N
Aegilops ventricosa	Aegilops ventricosa	DN
Aegilops recta	Aegilops neglecta	UMN
S genome species		
Aegilops bicornis	Aegilops bicornis	Sb
Aegilops bicornis var. bicornis	Aegilops bicornis var. bicornis	
Aegilops bicornis var. anathera	Aegilops bicornis var. anathera	
Aegilops sharonensis	Aegilops sharonensis Eig	SI
Aegilops longissima	Aegilops longissima Schweinf. & Muschl.	SI
	Aegilops searsii Feldman & Kislev ex Hammer	Ss
Aegilops searsii	Aegilops kotschy Boiss.	SU
Aegilops kotschy	Aegilops peregrina Hack.) Mare & Weiller	SU; SIU
Aegilops variabilis	Aegilops peregrina Hack.) Mare & Weiller var. peregrina	
	Aegilops peregrina var. brachyathera (Boiss.) Maire & Weiller	

Aegilops crassa (4x)	Aegilops crassa Boiss .	DcM
Aegilops crassa (6x)		
Aegilops trivialis (6x)		
Triticum syriacum		DcSsX
Aegilops vavilovii		
T genome species		
Aegilops mutica	Amblyopyrum muticum	T
	Amblyopyrum muticum var. muticum	
	Amblyopyrum muticum var. loliaceum	
U genome species		
Aegilops umbellulata	Aegilops umbellulata	U
Aegilops biuncialis	Aegilops biuncialis Vis.	UM
Aegilops columnaris	Aegilops columnaris Zhuk.	UM
Aegilops ovata	Aegilops geniculata	MU
Aegilops kotschy	Aegilops kotschy Boiss.	SU
Aegilops recta	Aegilops neglecta	UMN
Aegilops variabilis	Aegilops peregrina Hack.) Mare & Weiller	US; USI
	Aegilops peregrina Hack.) Mare & Weiller var. peregrina	
	Aegilops peregrina var. brachyathera (Boiss.) Maire & Weiller	
	Aegilops triuncialis	
Aegilops triuncialis	Aegilops triuncialis L.	UC
	Aegilops triuncialis L. var. triuncialis	
	Aegilops triuncialis var. persica	CU
Triticum syriacum	Aegilops vavilovii	DcSsX
Aegilops mutica Boiss.; Triticum muticum(Boiss.) Hack.	Amblyopyrum muticum (Boiss.) Eig	T
	Amblyopyrum muticum (Boiss.) Eig var. muticum	
		Amblyopyrum muticum var. loliaceum

l'analyse des cicatrices à la base des épillets (d'orges et de blés) trouvés sur des sites archéologiques, une progression lente des proportions entre les céréales sauvages (dissémination naturelle), les céréales semi-sauvages (plus de dissémination naturelle) et les céréales domestiquées (battables) par rapport aux dates d'apparition des différents types.

II-5 Conclusion des recherches sur la domestication :

L'homme a cultivé ces différentes catégories de blés (sauvages et domestiques) en mélange sur une longue période (Tableau II-1). La période de pré-culture aurait duré plus de mille ans, et la période transitoire plus de deux mille ans. Par la suite, il y a eu une période de domestication, consciente cette fois, qui a permis la sélection des plantes ayant des caractères particuliers. C'est ainsi qu'au final, chaque région a fait évoluer différemment ces espèces cultivées. Cette théorie permet d'expliquer la variabilité des fonds génétiques.

CHAPITRE III :

GENOMIQUE DU BLE

ET POLYPLOIDIE

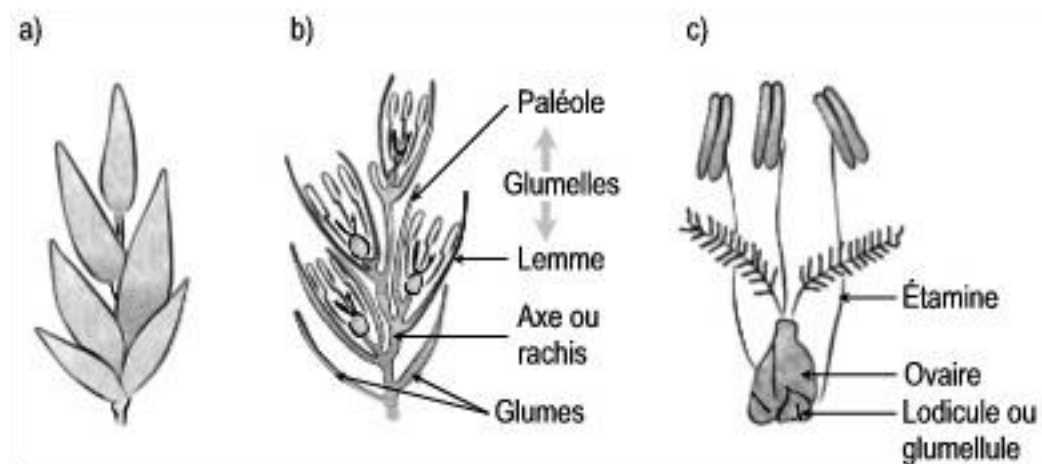


Fig. III-1 Description anatomique de l'épillet :

a - épi

b – coupe transversale de l'épi

c – description de la fleur

Source : floronet.pagesperso-orange.fr

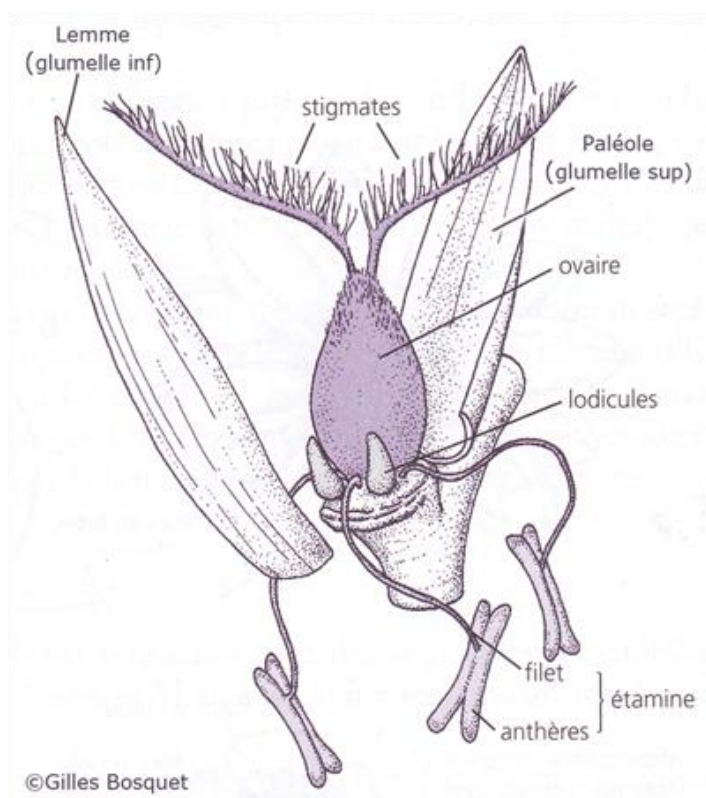


Fig. III-2 Description anatomique de la fleur du blé dans l'épillet.

Source : Gilles Bosquet

III-1 Description et classification du blé :

La description botanique et la classification du blé est la suivante (source: Plantae.ca) :

1. Règne: *Plantae*
2. Sous-règne : *Tracheobionta*
3. Division : *Spermatophyta*
4. Classe : *Angiospermae*
5. Sous-classe : *Monocotyledonae*
6. Super-ordre : *Commelinideae*
7. Ordre: *Poales*
8. Famille : *Poaceae*
9. Genres : il existe 668 genres (53 taxons), dont le genre *Triticum* et le genre *Aegilops*
10. Espèces : environ 10 000 espèces. Pour une distribution cosmopolite dans le monde.

La description botanique des *Poaceae* est faite suivant les formes des différentes parties de la plante (herbacées) (Fig. III-1) soit :

- ✓ Les feuilles : Alternes, sur deux rangs, limbe rubané. Parfois présence de ligule membraneuse ou poilue. Une gaine fendue formée par le bas de la feuille entourant la tige et insérée au niveau des nœuds.
- ✓ L'inflorescence : Epillet pouvant être pédicellé ou sessile.
- ✓ La fleur : Entourée de deux bractées appelées glumes. Vers l'intérieure, on trouve, une glumelle inférieure : *Lemma* (ou Lemme), une glumelle extérieure : *paléa* (ou paléole), 2 tépales (fusion de sépale et pétale), 3 (-6) étamines, 2 (-3) carpelles. Avec un ovaire supère portant deux stigmates. L'ensemble peut être réduit en une formule florale générale : $2T + 3-(3+3) A + 2(-3) G$. (Fig. III-2).
- ✓ Le fruit : Caryopse, correspondant au grain de blé.

Parfois traitée de rébarbative la classification des espèces (la taxonomie) est une discipline importante car elle apporte des informations qui peuvent être utilisées dans d'autres disciplines comme la systématique, l'étude de la diversité biologique et la mise en évidence des relations évolutives entre les divers organismes (la phylogénie). Depuis 1998, la classification de référence est celle réalisée par « Angiosperm Phylogeny Group » (APG). Elle intègre la classification de Cronquist (1981) et des données de biologie moléculaire. Cette classification phylogénétique (APG III) dépend principalement du passage de trois marqueurs moléculaires (deux chloroplastiques et un mitochondriale) (APG 1998; APG II, 2003 ; APG III, 2009 ; Chase et al., 2009b ; Chase et al., 2009a ; Haston et al., 2009).

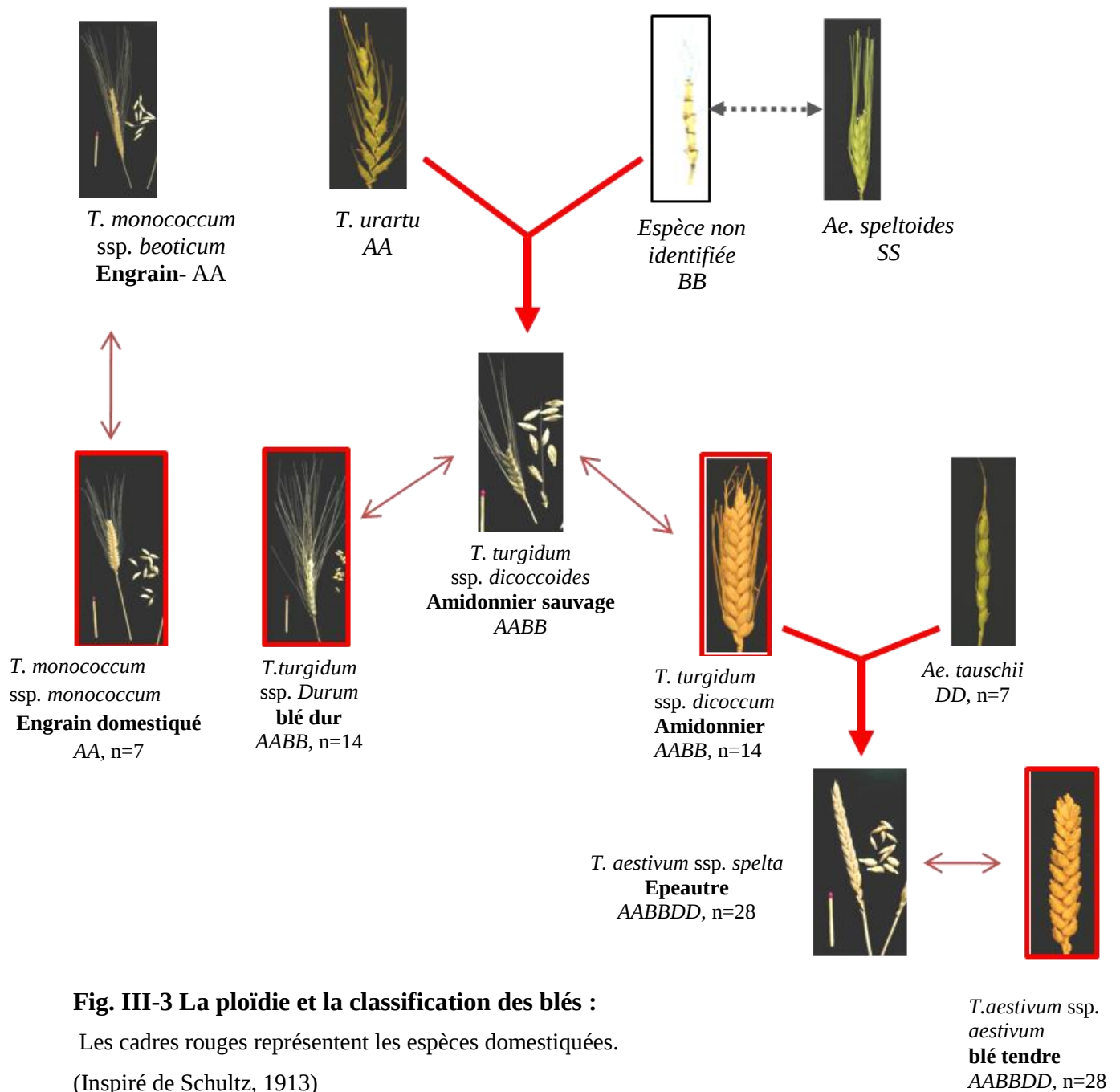


Fig. III-3 La ploïdie et la classification des blés :

Les cadres rouges représentent les espèces domestiquées.

(Inspiré de Schultz, 1913)

III-2 Cytologie du blé :

En 1913, Schultz, à partir d'observations phénotypiques divise les blés en trois groupes, le premier comprend l'engrain (car un grain), le deuxième l'amidonnier et le troisième l'épeautre (Fig. III-3). La cytologie permet de découvrir que dans chacun de ces groupes le nombre de chromosomes est respectivement $2n=14$, 28 et 42 (Sakamura, 1918). C'est la découverte de la polyploïdie chez le blé. Par la suite l'étude des hybrides entre différents groupes a permis de mettre en évidence l'allopolyploïde du blé (Kihara, 1919 ; Sax, 1922). L'engrain, baptisé *T. monococcum* (Morris and Sears, 1967), porte le génome A, l'amidonnier les génomes AB et l'épeautre les génomes ABD. Mais c'est *T. urartu* et non *T. monococcum* qui est reconnu pour être le parent donneur du génome A chez le tétraploïde (amidonnier) et l'hexaploïde (épeautre). Il a fallu attendre 1944, pour que le parent donneur du génome D (*T. tauschii*) soit identifié (Kihara, 1944 ; McFadden and Sears, 1944 ; 1946).

Personne n'a encore découvert l'espèce du blé diploïde, parent donneur du génome B. Beaucoup de théories ont été établies dont celle de l'extinction du parent (Zohary and Feldman, 1962 ; Kimberg and Saltee, 1973).

A partir de la mise en évidence de l'importance de la polyploïdie chez le blé, beaucoup de questions ont été posées.

Dans ma recherche du comportement des homéoallèles du gène majeur de domestication, je ne pouvais ignorer ces recherches ni les implications de la polyploïdie au sein du blé.

Dans la suite du chapitre, je décris la polyploïdie et son importance. Je parlerai également des travaux d'Ernest Sears (1910-1991) qui a su profiter de cette particularité du blé pour fabriquer des aneuploïdes (1954). Ces lignées sont de puissants outils encore largement utilisés aujourd'hui et sont à la base du travail que j'ai réalisé pour la validation des hypothèses que j'ai faites et des résultats que j'ai obtenus.

III-3 La polyploïdie :

III-3-1 Définitions de la polyploïdie :

La polyploïdie ou l'assortiment d'au moins deux jeux complets de chromosomes dans un même noyau, a joué un rôle essentiel dans l'évolution des eucaryotes (Fig. III-4), notamment chez les angiospermes qui ont tous connu au moins un événement de polyploïdisation au cours de leur évolution. Ainsi, les programmes de séquençage de génomes complets ont permis de mettre en évidence le fait que les espèces considérées comme diploïdes tel que le riz, le maïs, la vigne, *Arabidopsis* et même l'espèce basale des angiospermes *Amborella tricarpha*, comme démontré récemment, sont, en réalité, des anciens polyploïdes ou paléopolyploïdes (Fig. III-5).

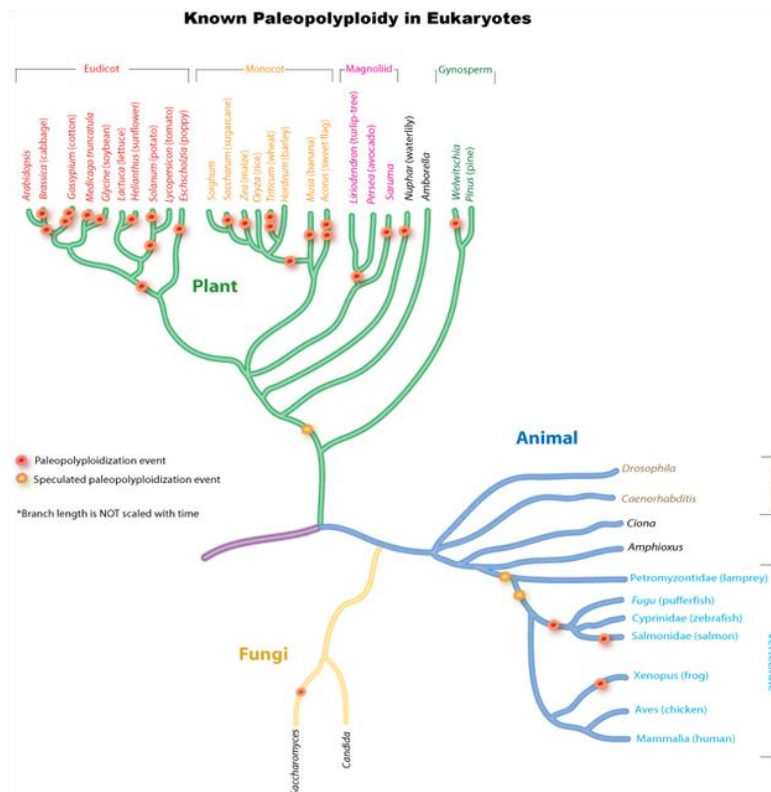


Fig. III-4 Les évènements de polyploïdisation dans les différents règnes.

(Wolfe, 2001 ; Adams and Wendel, 2005c ; Cui et al., 2006 ; Zhang et al., 2013)

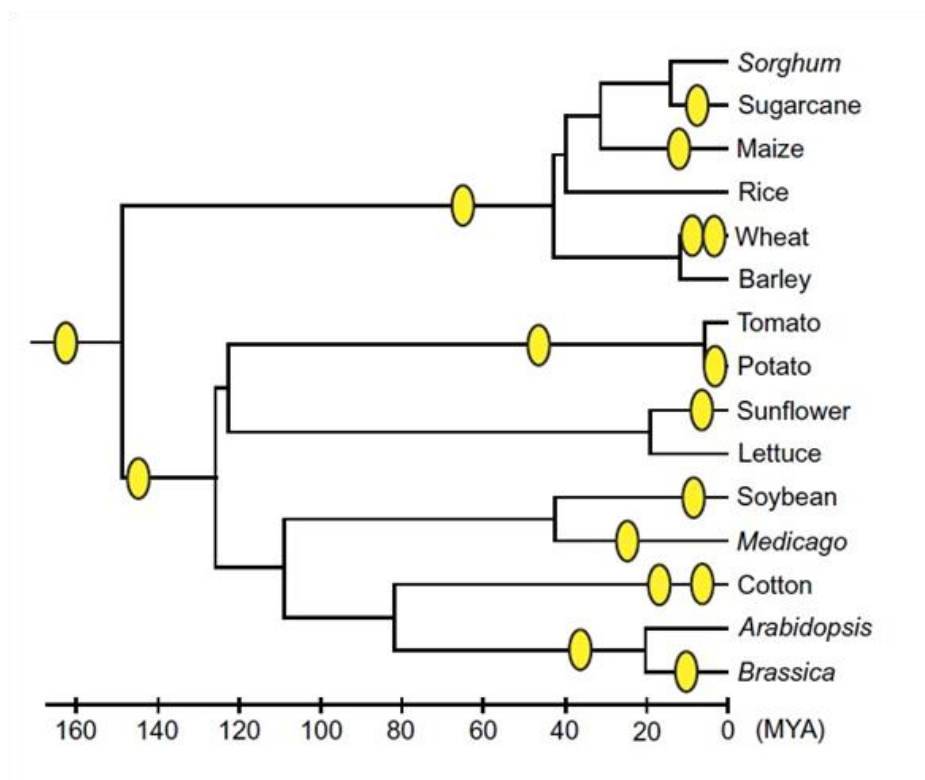


Fig. III-5 Les évènements de polyploïdisation pour quelques espèces.

(Zhang et al., 2013)

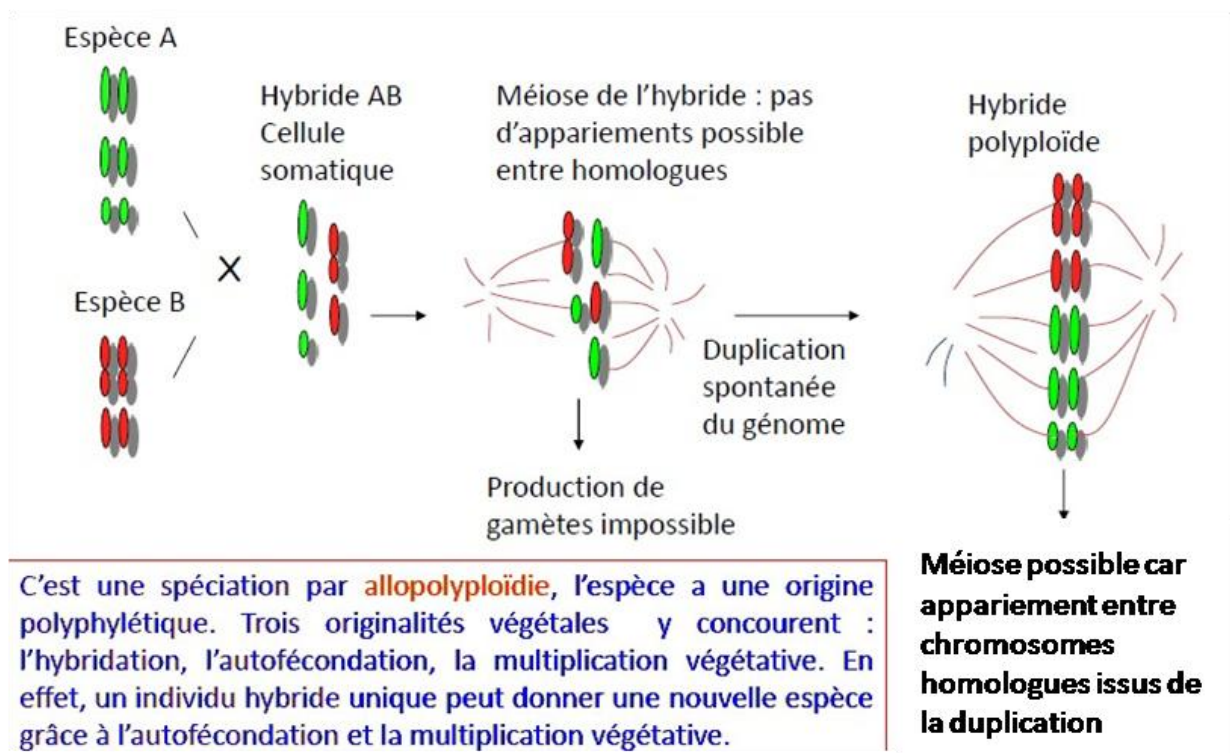


Fig. III-6 Fondation d'une espèce polyploïde par hybridation de deux espèces différentes.

(Iacouture, 2006)

Les raisons de ce succès évolutif ne sont pas encore bien élucidées et font l'objet de nombreuses recherches.

Les études menées à ce jour montrent que : (i) les espèces néo-polyploïdes doivent surmonter les biais d'appariement et de ségrégation chromosomique afin de s'établir comme nouvelles espèces ; (ii) le génome des espèces polyploïdes est très dynamique et subit, dès les premières générations de sa formation, puis au cours de l'évolution, une cascade de changements génétiques, fonctionnels et épigénétiques, générant de nouveaux phénotypes absents des espèces parentales. La nature, la fréquence et la durée de ces changements varient d'une espèce polyploïde à une autre.

La polyploïdie est importante dans l'évolution et la formation d'espèces du blé, qui constituent un modèle d'étude sur ses effets, mais également pour élucider les mécanismes impliqués dans la stabilisation des espèces polyploïdes. En effet, en plus de l'existence de différents polyploïdes à l'état naturel nous avons la possibilité de les néo-synthétiser en laboratoire (Fig. III-6). Chez le blé, ce système polyploïde est le seul pour lequel un locus à effet majeur empêchant l'appariement entre chromosomes homéologues lors de la méiose a pu être caractérisé. Ce locus, *Ph1* (*Pairing homoeologous 1*) présent seulement chez les blés tétraploïdes et hexaploïdes assure un appariement homologue strict, dit comportement de diploïde à la méiose, essentiel à la stabilité de ces espèces allopolyploïdes (Fig. III-6).

Les organismes diplophasiques (diploïdie dominante) ayant un mode de reproduction sexué passent, au cours de leur cycle de vie, par un état haploïde (mono- ou polyhaploïde) pendant lequel les cellules possèdent une seule copie du génome ($1n$). Ces cellules sont les gamètes. La fusion des gamètes mâles et femelles restaure le nombre de chromosomes somatiques ($2n$) en donnant un zygote possédant deux copies de chaque chromosome ou chromosomes homologues.

Ainsi, une espèce diploïde possède deux jeux de chromosomes ($2n=2x$ avec x : nombre de chromosomes dans un jeu) alors qu'une espèce polyploïde possède plus de deux jeux (triploïde : $2n=3x$, tétraploïde : $2n=4x$, hexaploïde : $2n=6x...$).

Il existe également un type particulier de polyploïdisation, l'endopolyploïdie qui touche certains tissus. Le processus de duplication de l'ADN pendant l'endopolyploïdie conduit à :

- la polyténie lorsque toutes les chromatides filles sont rattachées au même centromère sans modification du nombre de chromosomes.
- l'endomitose lorsque les chromosomes dupliqués sont indépendants et repartis dans l'ensemble du volume nucléaire.

Un organisme ou une cellule possédant un jeu complet de chromosomes sont dit euploïdes par opposition aux aneuploïdes qui présentent un jeu de chromosomes incomplet (perte d'un chromosome ou d'un bras de chromosome) ou surnuméraire (gain d'un chromosome ou d'un bras de chromosome).

- *T. monococcum*
(Engrain)
- *T. urartu*
- *Aegilops*

T. Turgidum
ssp. dicoccum
(Amidonnier)
ssp. dicoccoides
ssp. turgidum
ssp. carthlicum
ssp. durum
etc.

- *T. aestivum*
- *ssp. spelta* (Epeautre)
- *ssp. macha*
- *ssp. vavilovii*
- *ssp. eastivum*
- *ssp. compactum*
- etc.

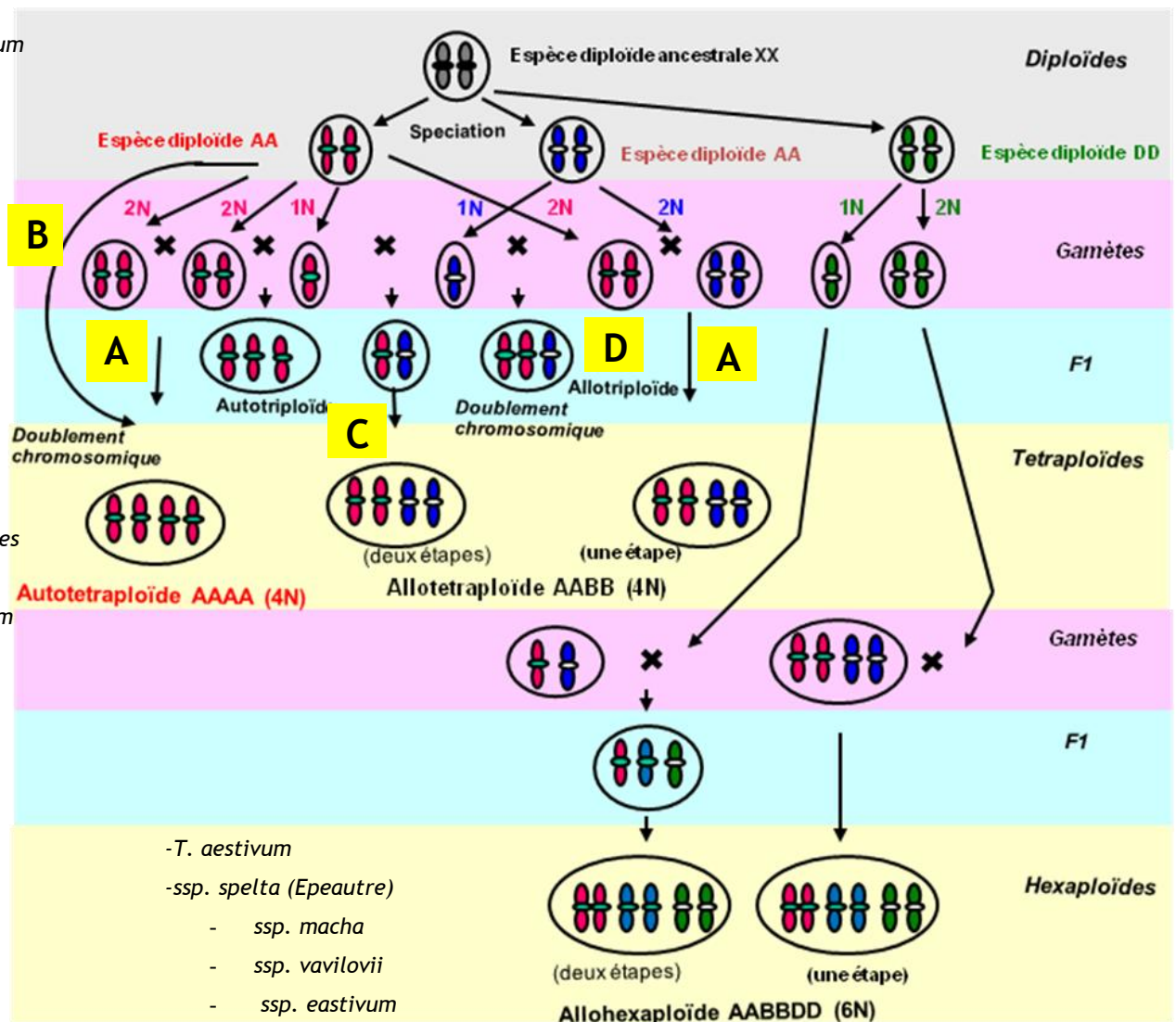


Fig. III-7 Schéma de formation des espèces polyploïdes suivant différentes voies :

un tétraploïde peut être formé :

A - de manière directe suite à la fusion de deux gamètes non-réduites produites par les deux espèces parentales diploïdes.

B - suite à un doublement somatique ayant lieu dans le zygote.

C - de manière indirecte après la fusion de deux gamètes non-réduites produites par l'hybride interspécifique.

D - suite à la formation d'un hybride triploïde qui, par autofécondation ou rétrocroisement avec l'une des espèces parentales, peut former un allotétraploïde.

(Raybould et al., 2008)

On distingue généralement les espèces autopolyploïdes et les espèces allopolyploïdes. Les autopolyploïdes possèdent des jeux de chromosomes provenant d'une même espèce alors que les espèces allopolyploïdes dérivent d'un ou de plusieurs événements d'hybridations interspécifiques, suivis ou précédés d'un doublement chromosomique (Stebbins, 1971).

Cette classification a souvent été sujette à controverse (revue dans (Ramsey and Schemske, 1998) ; (Ramsey and Schemske, 2002) puisque les autopolyploïdes peuvent être issus de croisements entre génotypes divergents de la même espèce. En d'autres termes, plus la divergence entre les espèces progénitrices est faible, plus un allopolyploïde tend vers un autopolyploïde.

Les espèces allopolyploïdes réunissent dans leurs noyaux des génomes ayant divergé de leurs ancêtres communs à la suite d'un événement de spéciation. Ces génomes sont qualifiés d'homéologues.

L'allopolyploïdie est un des mécanismes de spéciation le plus répandu chez les plantes, et conduit à la formation de nouvelles lignées-espèces, montrant des capacités adaptatives souvent supérieures à leurs parents diploïdes. Ces capacités, qui se traduisent dans la nature par une plus large expansion de ces espèces (dont certaines se montrent invasives), ont été largement exploitées par l'homme, qui a ainsi domestiqué un grand nombre d'espèces jouant aujourd'hui un rôle économique majeur comme le blé, le coton, ou le maïs (Ramsey and Schemske, 1998 ; Wendel and Doyle, 2005 ; Leitch and Leitch, 2008).

III-3-2 Voies de formation des polyploïdes :

Plusieurs modes de formation des polyploïdes ont été décrits chez les plantes (Ramsey and Schemske, 1998). En effet, les espèces polyploïdes (auto- ou allopolyploïdes) peuvent être formées de manières directes ou indirectes (formation d'hybrides) et suivant deux voies de doublement du stock chromosomique, la voie somatique et la voie méiotique (formation puis fusion de gamètes non réduits).

Le doublement somatique (Fig. III-7) se produit suite à une absence totale de division cellulaire lors de la mitose (alors que le jeu de chromosomes s'est répliqué) et peut aboutir à un organisme polyploïde s'il a lieu soit dans le zygote (œuf) ou le jeune embryon (post-zygotique), soit dans un tissu méristématique donnant naissance aux cellules germinales qui forment à leur tour les gamètes (pré-zygotique). La primevère polyploïde *Primula kewensis* dériverait ainsi d'un doublement somatique affectant une tige de l'hybride diploïde stérile *Primula floribunda* x *Primula verticellata* (Newton and Pellew, 1929). Bien que le doublement somatique soit couramment utilisé pour générer des polyploïdes synthétiques au laboratoire (doublement induit par un agent chimique tel que la colchicine ou le colcemide), il est admis que la plupart des polyploïdes ont été formés par la voie des gamètes non-réduits (Harlan and De Wet, 1975 ; Bretagnolle and Thompson, 1995 ; Ramsey and Schemske, 1998) (Fig. III-6 ; voies D et E).

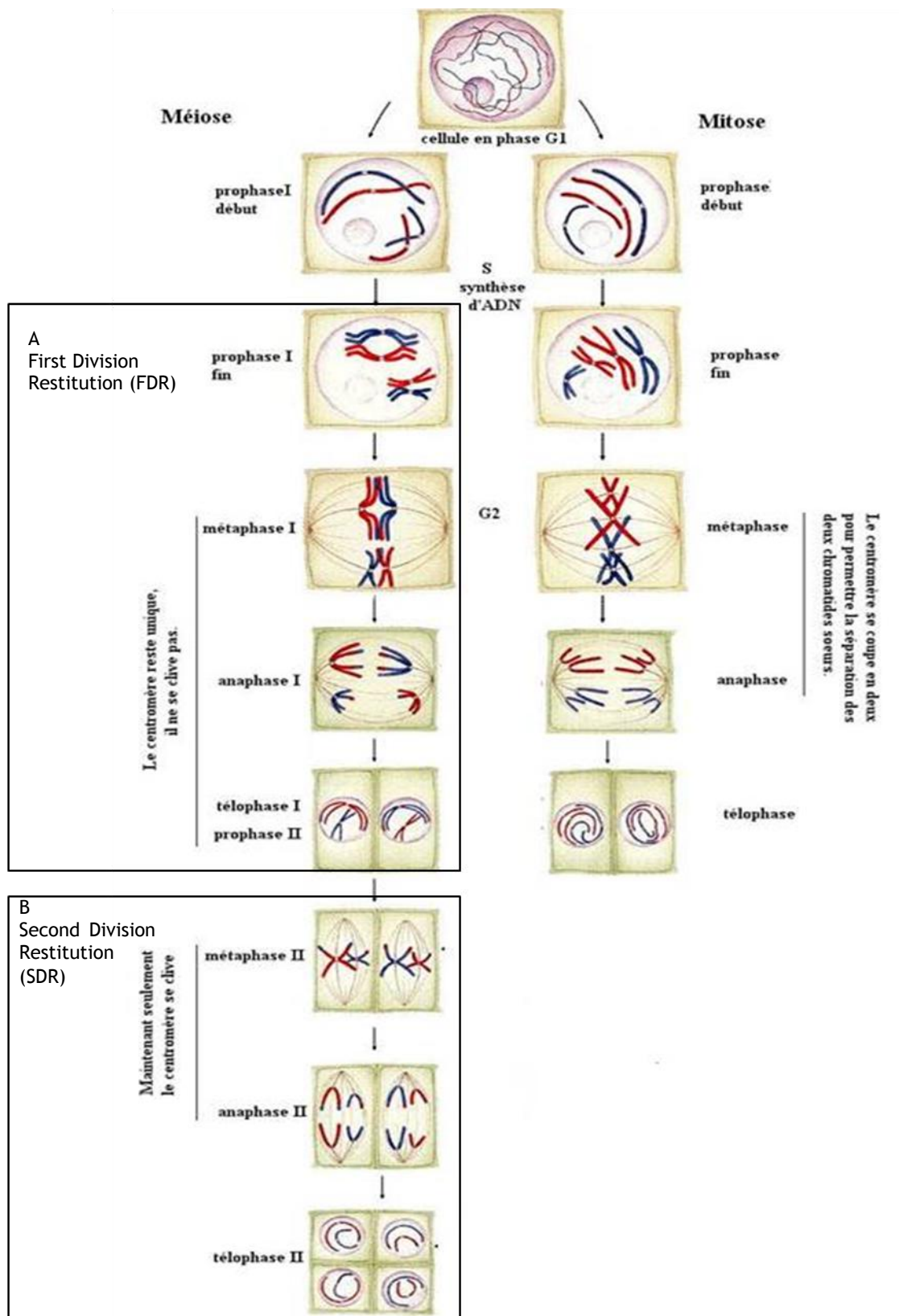


Fig. III-8 Schéma de la méiose et de la mitose :

A – First Division Restitution (FDR) en métaphase I de la méiose, normalement il y a séparation des chromosomes homologues. B – Second Division Restitution (SDR) en métaphase II, normalement il y a séparation des chromatides sœurs. (Source: edu.upmc.fr)

Les gamètes non réduits (gamètes $2n$ ou diplogamètes) contiennent le même nombre de chromosomes que la plante dont ils sont issus et trouvent leur origine dans les anomalies méiotiques, notamment l'absence de première ou deuxième division méiotique et l'organisation anormale des fuseaux (Fig. III-8). Plusieurs types de gamètes non-réduits (mécanismes cytologiques et conséquences génétiques différentes) ont été décrits dans la littérature (Bretagnolle and Thompson, 1995 ; Ramsey and Schemske, 1998) :

- ✓ Les gamètes issus de FDR (First Division Restitution), c'est-à-dire d'une absence de première division méiotique qui se traduit par la non-disjonction des chromosomes homologues. Les gamètes formés suite à une FDR sont diploïdes et hétérozygotes aux événements de recombinaison près (hétérozygotie au centromère et homozygotie au-delà du premier « crossing-over » (CO)). Ils contiennent une paire de chromatides homologues et donc l'intégralité de l'information génétique d'un ou des parents (aux remaniements près). C'est ce type de gamètes non-réduits qui a été observé chez les blés (Jauhar, 2007).
- ✓ Les gamètes issus de SDR (Second Division Restitution), l'absence de la seconde division méiotique aboutit à une non-disjonction des chromatides sœurs. Le processus aboutit à la formation de gamètes homozygotes sauf entre le premier (CO) et le télomère. La zone est hétérozygote et contient une paire de chromatides sœurs appartenant à l'un des jeux d'origine.
- ✓ Les gamètes issus d'IMR (Indeterminate Meiotic Restitution). Ce type de restitution méiotique implique la présence de chromosomes restés à l'état d'univalents (chromosomes non-appariés en métaphase I) dont la ségrégation se fait simultanément mais différemment de celle des chromosomes sous forme de bivalents (Lim et al., 2001).
- ✓ Les gamètes issus d'une absence totale de division (FDR suivie d'une SDR) et possédant donc les deux paires de chromosomes homologues à deux chromatides.

Les gamètes non réduits peuvent également être formés suite à une orientation anormale des fuseaux lors de la méiose. Ainsi, la formation de fuseaux parallèles lors de la métaphase II peut générer des gamètes non réduits de type FDR, comme il a été décrit pour les mutants *ps1* (*parallel spindle 1*) de pomme de terre ((Fukuda, 1927) cité dans (Andreuzza and Siddiqi, 2008)).

Très récemment, plusieurs gènes impliqués dans la formation des gamètes non réduits ont été isolés et caractérisés (d'Erfurth et al., 2009 ; Erilova et al., 2009 ; Brownfield and Kohler, 2010 ; d'Erfurth et al., 2010). Par exemple, une mutation dans le gène *AtPS1* (*Arabidopsis thaliana parallel spindle 1*) se traduit par la formation de fuseaux en orientation parallèle en métaphase II au lieu de la configuration tétraédrique normale aboutissant à la formation de dyades à la place de tétrades (d'Erfurth et al., 2009).

Ainsi, un allotétraploïde peut être formé (i) de manière directe suite à la fusion de deux gamètes non-réduites produites par les deux espèces parentales diploïdes ou suite à un doublement somatique ayant lieu dans le zygote (Fig. III-7 : A et B) ; (ii) de manière indirecte après la fusion de deux gamètes non-réduites produites par l'hybride interspécifique (Fig. III-7 : C) ou suite à la formation d'un hybride triploïde qui, par autofécondation ou rétrocroisement avec l'une des espèces parentales, peut former un allotétraploïde (Fig. III-7 : D).

Les travaux menés ces dernières années montrent que la plupart des espèces polyploïdes ont une origine multiple, l'événement de polyploïdisation s'étant produit de façon récurrente et indépendante au sein des populations (Soltis and Soltis, 2004). C'est notamment le cas des espèces polyploïdes du genre *Tragopogon* où, *T. mirus* et *T. miscellus* résulteraient respectivement de 9 et d'au moins 21 événements d'hybridation distincts (Soltis and Soltis, 1995). Le blé hexaploïde résulte d'un ou deux événements d'allopolypléidisation (Talbert et al., 1998).

III-3-3 Prévalence et fréquence de la polyploïdie chez les eucaryotes :

La polyploïdie a joué un rôle crucial dans l'évolution des eucaryotes, en particulier chez les angiospermes (Adams and Wendel, 2005a ; 2005b ; Chase et al., 2009b ; Van de Peer et al., 2009a ; Van de Peer et al., 2009b).

Tous les angiospermes ont eu un ou plusieurs cycles de polyploïdisation au cours de leur évolution (Muntzing, 1936 ; Darlington, 1937 ; Stebbins, 1950 ; Grant, 1981 ; Blanc and Wolfe, 2004 ; Cui et al., 2006 ; Jaillon et al., 2007 ; Soltis et al., 2009 ; Tang et al., 2010 ; Amborella Genome Project, 2013). Les gymnospermes ne montrent que 5% d'espèces polyploïdes alors que c'est le sous-embouchement le plus proche des angiospermes dans la classification des espèces. Le haut niveau de formation de polyploïdes naturels dans les angiospermes (1/100000 plantes) viendrait du taux élevé de gamètes non réduites produites (0,56%) dans ces espèces (Ramsey and Schemske, 1998).

Par ailleurs, Otto & Whitton (Otto and Whitton, 2000) ont estimé que la polyploïdie est impliquée dans 2 à 4% des événements de spéciation chez les angiospermes. De plus, les travaux de Soltis et al. (Soltis et al., 2009) suggèrent une forte corrélation entre la polyploïdie et le taux de diversification des espèces. Par exemple, l'événement de polyploïdisation chez les *Poaceae* coïncide avec l'origine et la divergence du «core des *Poaceae*», un clade contenant plus 10000 espèces actuellement.

La polyploïdie semble être moins répandue chez les animaux que chez les végétaux (Fig. III-4). En effet, uniquement 2 à 3 événements anciens ont eu lieu au cours des 500 millions d'années d'évolution des vertébrés et peu d'événement récents ont été documentés alors que plusieurs événements anciens ont eut lieu depuis la diversification des plantes à fleurs, il y a 170 millions d'années.

De plus, plusieurs évènements récents ont été trouvés chez les angiospermes. Il existe néanmoins des centaines d'espèces polyploïdes parmi les insectes et les vertébrés, notamment chez les reptiles, les amphibiens et les poissons (Fig. III-4). Des évènements de duplication du génome ont également été documentés chez des organismes unicellulaires. En effet, la première duplication ancestrale a été mise en évidence chez la levure *Saccharomyces cerevisiae* (Wolfe and Shields, 1997). Plus récemment, il a été montré que le protozoaire unicellulaire *Paramecium tetraurelia* a subi plusieurs évènements de polyploïdisation (Aury et al., 2006).

La polyploïdie a joué un rôle essentiel dans l'évolution et la domestication des plantes. Ce processus a un fort impact sur la dynamique de la biodiversité et le fonctionnement des écosystèmes car il est accompagné d'importants changements génétiques et épigénétiques qui génèrent de nouveaux phénotypes absents des espèces parentales (Adams and Wendel, 2005a, 2005b).

III-3-4 Changements génomiques chez les polyploïdes :

Les études menées à ce jour montrent que le génome des espèces polyploïdes est très dynamique et subit, dès sa formation et au cours de l'évolution, une cascade de changements génétiques, épigénétiques et fonctionnels. La nature, le timing et la fréquence de ces changements varient d'une espèce polyploïde à une autre. Des descriptions exhaustives de ces changements et des mécanismes s'y rapportant ont fait l'objet de revues récentes (Feldman and Levy, 2005 ; Chen and Ni, 2006 ; Chen, 2007 ; Doyle et al., 2008 ; Hegarty and Hiscock, 2008 ; Leitch and Leitch, 2008 ; Soltis and Soltis, 2009).

III-3-4-1 Remaniements structuraux: dynamique et mécanismes :

Les données moléculaires suggèrent que les génomes polyploïdes subissent des changements dynamiques et fréquents de l'ADN en réponse au « choc génomique » (McClintock, 1984) que constituent l'hybridation et la duplication du génome. Toutefois, des résultats contrastés ont été obtenus quant à la fréquence et à la durée des remaniements structuraux induits par la polyploïdie (et l'hybridation interspécifique). Song et al. (Song et al., 1995) ont été les premiers à mettre en évidence des changements majeurs dans l'organisation des génomes des allotétraploïdes synthétiques de Brassica et ont détecté (en utilisant des RFLP) des profils non additifs des fragments parentaux chez ces synthétiques. De même, d'importants remaniements structuraux ont été observés chez les allopolyploïdes synthétiques (et naturels) de blé (Feldman et al., 1997 ; Liu et al., 1998 ; Ozkan et al., 2001), alors que des études faites au laboratoire sur les blés hexaploïdes synthétiques ont conclu au faible taux de remaniement (Mestiri et al., 2010).

Différents types de remaniements structuraux se produisent et s'accumulent au fil du temps entraînant ainsi une évolution structurale des génomes polyploïdes. Ainsi, des translocations inter-génomiques sont souvent mises en évidence chez les allopolyploïdes naturels de tabac (Leitch and Bennett, 1997), de colza (Udall et al., 2005) et de blé (Badaeva et al., 2007).

Outre ces remaniements chromosomiques majeurs, des réarrangements affectant des loci particuliers ont été documentés chez les allopolyploïdes naturels de blé, par exemple au niveau du locus Ph1 (Griffiths et al., 2006 ; Al-Kaff et al., 2008) et du locus Ha (Chantret et al., 2005 ; Li et al., 2008 ; Ragupathy and Cloutier, 2008).

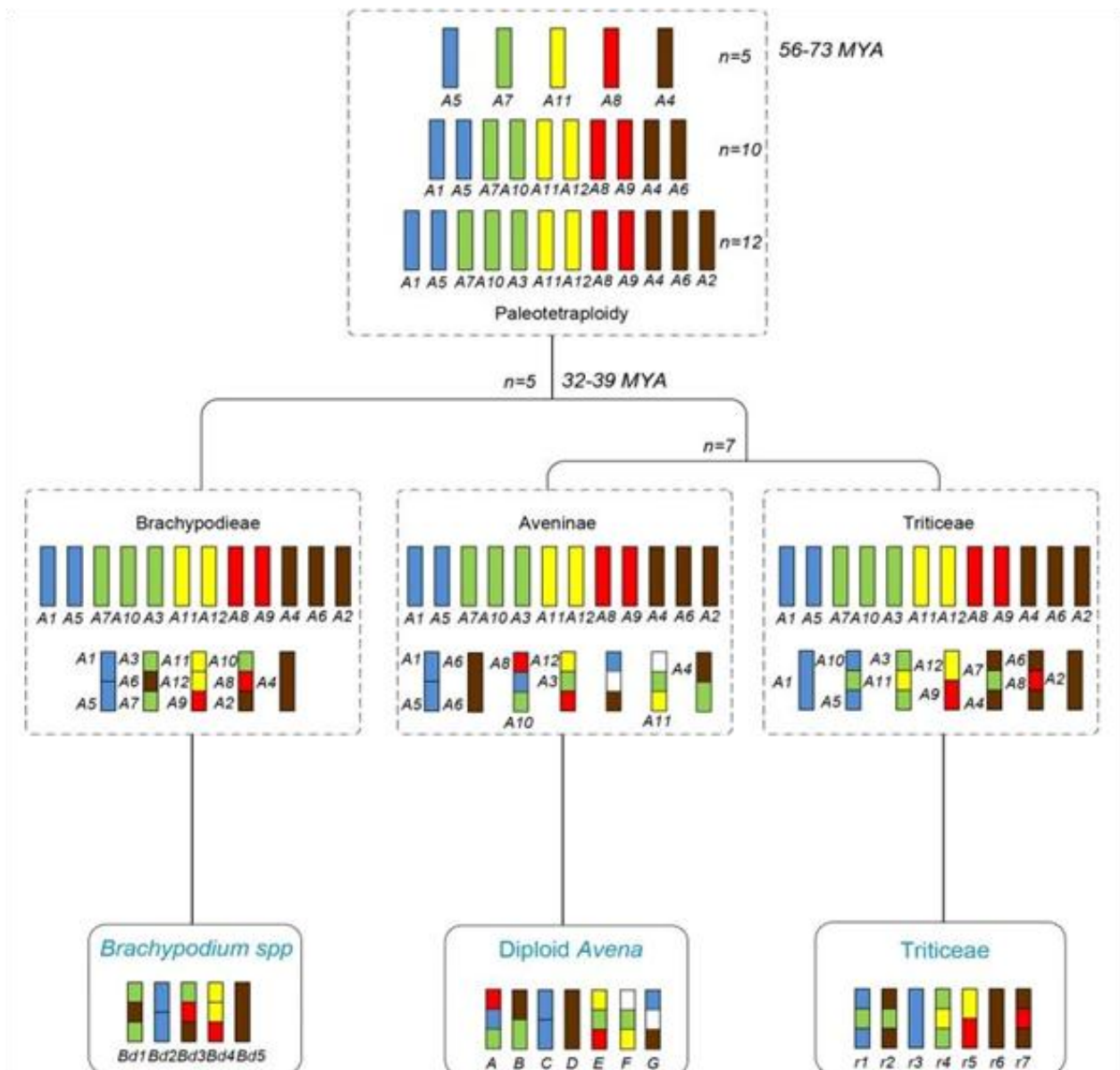


Fig. III-9 Modèle d'évolution de la structure des chromosomes chez les *Poaceae*.
(Salse et al., 2009)

Par ailleurs, l'accumulation de ces remaniements chromosomiques peut entraîner une restructuration importante du caryotype. Ainsi, afin de reconstituer l'évolution structurale d'*Arabidopsis thaliana* à partir du génome ancêtre, Lysak et al. (Lysak et al., 2006) ont pu montrer que la réduction du nombre de chromosomes de 8 à 5 depuis la divergence avec *Arabidopsis lyrata*, il y a environ 5 millions d'années (Ma), résultait de la succession d'évènements de type inversion, translocation et fusion.

Ces modifications structurales, dont la fréquence et la nature est très variable selon les taxons, pourraient contribuer aux changements de l'expression des gènes homéologues observés chez les polyploïdes.

III-3-4-2 Les mécanismes responsables des réarrangements structuraux :

Plusieurs hypothèses quant aux mécanismes moléculaires susceptibles d'être à l'origine des remaniements structuraux observés ont été émises (Fig. III-9). Ces remaniements semblent être liés à l'occurrence de cassures doubles brins pouvant être dues à des évènements de recombinaisons et à des anomalies méiotiques mais aussi à des évènements de transpositions. Ainsi, les échanges intergénomiques (translocations réciproques ou non réciproques entre chromosomes homéologues) au cours de la méiose sont probablement à l'origine de la plupart des pertes de séquences observées chez les *Brassica* synthétiques (Osborn et al., 2003a ; Udall et al., 2005 ; Nicolas et al., 2007 ; Gaeta and Chris Pires, 2010).

Les conversions géniques ou les crossing-overs inégaux sont les mécanismes les plus souvent invoqués pour expliquer l'évolution, parfois concertée (Wendel et al., 1995 ; Belyayev et al., 2000 ; Wendel, 2000) des séquences d'ADN répétées en tandem (comme les ADN ribosomaux).

Des anomalies méiotiques ont été détectées chez les allotétraploïdes d'*Arabidopsis* mais leur lien avec la recombinaison homéologue n'a pas été démontré (Comai et al., 2003; Madlung et al., 2005).

Les réarrangements peuvent également se produire par le biais des éléments transposables. Ces éléments répétés, de part leur transposition et leur dispersion dans le génome, peuvent servir de matrice pour des recombinaisons ectopiques (Chantret et al., 2005) générant ainsi des réarrangements chromosomiques.

Les importants remaniements structuraux observés chez le blé peuvent être dus à des évènements de transpositions. Toutefois, aucune explosion de ces évènements de transposition après la polyploïdisation n'a été mise en évidence chez ces allopolyploïdes. En revanche, les évènements de recombinaisons illégitimes sont à l'origine des remaniements observés chez les blés polyploïdes naturels comme pour le locus *Ha* (Chantret et al., 2005 ; Charles et al., 2009).

III-3-4-3 Changements épigénétiques :

Conjointement ou indépendamment des remaniements génétiques, des changements épigénétiques, notamment la méthylation de l'ADN ont été mis en évidence dans les différents modèles polyploïdes. La méthylation de l'ADN joue un rôle important dans la stabilisation du génome en (i) régulant l'expression des gènes, (ii) ciblant les séquences répétées et notamment les éléments transposables (ET), dont elle réprime l'activité, (iii) coordonnant et stabilisant (grâce à la méthylation des CG) la mémoire de l'information épigénétique afin que celle-ci soit transmise fidèlement au cours des générations (Mathieu et al., 2007).

III-3-4-4 Reprogrammation de l'expression des gènes :

La caractérisation de l'expression des gènes dans différents modèles allopolyploïdes indique que plusieurs gènes se comportent selon un modèle additif, c'est-à-dire que l'expression des gènes chez les polyploïdes correspond à une combinaison des profils d'expression parentaux. D'autre part, une proportion de gènes, variable selon les polyploïdes, montre une déviation par rapport à ce modèle et de nouveaux profils d'expression (non additifs) sont alors observés. Ceci concerne environ 5% des gènes chez les allotétraploïdes synthétiques d'*Arabidopsis* (Wang et al., 2006), 1 à 6% chez le coton (Adams et al., 2004 ; Rapp et al., 2009) et environ 5% chez *Senecio* (Hegarty et al., 2005).

La reprogrammation de l'expression des gènes chez les blés allohexaploïdes a été bien étudiée. Chez les allohexaploïdes synthétiques de blé, l'utilisation de la technique cDNA-AFLP montre que 7.7% des transcrits ont un profil d'expression différent de celui de leurs géniteurs (He et al., 2003).

En utilisant une puce Affymetrix GeneChip® blé (55,049 transcrits), Chagué et al. (2010) ont observés que l'expression de plus de 93% des transcrits est additive. L'expression de plus de 2000 gènes (environ 7%) se ferait de manière non additive chez ces allohexaploïdes synthétiques. Cette étude effectuée dans notre laboratoire suggère que la reprogrammation de l'expression des gènes est rapidement établie lors de l'allopolyploïdie et hautement conservée à travers les générations, comme le montre la comparaison faite entre les allohexaploïdes synthétiques et naturels de blé (Chelaifa et al., ; Chague et al., 2010).

Par ailleurs, tandis que chez les allopolyploïdes d'*Arabidopsis* une répression concernant les gènes surexprimés dans l'un des géniteurs par rapport à l'autre a été observée (Wang et al., 2006), chez le coton, il a été suggéré que l'un des génomes parentaux contribue plus à l'expression dans l'allopolyploïde que l'autre génome (Rapp et al., 2009). En revanche, aucun phénomène de «dépendance» d'expression envers un des génomes parentaux ou de dominance d'un des génomes parentaux n'a été observée chez le blé (Chague et al., 2010).

III-3-4-5 Devenir des gènes dupliqués dans le génome des plantes polyploïdes :

Les gènes dupliqués suite à un événement de polyploïdie peuvent avoir plusieurs « destins » (Lynch and Force, 2000 ; Adams et al., 2004 ; Soltis and Soltis, 2004 ; Comai, 2005 ; Adams and Wendel, 2005b ; Chen and Ni, 2006) :

- ✓ La non fonctionnalisation, l'inactivation par l'élimination ou l'altération de la séquence (pseudogénéisation) d'une copie aboutissant à une perte de fonction. Une des explications serait que la redondance apportée par la polyploïdie entraîne le relâchement de la pression de sélection sur une des deux copies, favorisant ainsi l'accumulation de mutations, puis son élimination.
- ✓ La néo-fonctionnalisation, constitue l'acquisition d'une nouvelle fonction.
- ✓ La sous-fonctionnalisation, une spécialisation et/ou une expression différente selon les tissus ou le stade de développement.
- ✓ La « redondance fonctionnelle » ou le maintien de la fonction des gènes dupliqués.

Quelque soit l'option il y a la mise en place d'un réseau de régulation et d'interaction plus complexe pour un effet dosage.

Bien que certains cas de néo-fonctionnalisation aient récemment été identifiés chez les cotons synthétiques (Chaudhary et al., 2009), il est plus vraisemblable que l'évolution de la fonction des gènes, et des réseaux de gènes participant à une même voie métabolique, soit un processus lent car elle nécessite la mutation et la co-évolution de nombreux gènes (Adams and Wendel, 2005a).

Des travaux réalisés sur *Arabidopsis* suggèrent que la polyploidie module et diversifie l'expression des gènes en réponse à des chocs environnementaux (Ha et al., 2009a). Les auteurs ont montré que la proportion des gènes dupliqués ayant une expression non additive chez les polyploïdes (naturel et synthétiques) est significativement plus élevée que celle des gènes présents en copie unique.

Les travaux réalisés sur trois gènes homéologues à MADS Box de la classe E chez le blé hexaploïde, illustrent bien certains de ces mécanismes (Shitsukawa et al., 2007). En effet, alors que l'homéologue sur le génome A subit une altération structurale de sa séquence (insertion d'une large séquence au niveau du domaine K) aboutissant à une pseudogénéisation, la structure des deux autres homéologues (B et D) reste intacte. Toutefois, l'homéologue B est préférentiellement inactivé suite à la méthylation des cytosines (sous fonctionnalisation). Ainsi, les auteurs ont montré que les trois homéologues sont régulés différemment par des mécanismes génétiques et épigénétiques.

En plus de ces changements génétiques et épigénétiques, d'autres mécanismes peuvent causer des variations de l'expression des homéologues; par exemple des «interactions régulatrices» (Comai et al., 2000 ; Birchler et al., 2003 ; Riddle and Birchler, 2003 ; Osborn et al., 2003b ; Adams and Wendel, 2005c ; Hovav et al., 2008 ; Chaudhary et al., 2009 ; Pang et al., 2009 ; Ha et al., 2009a).

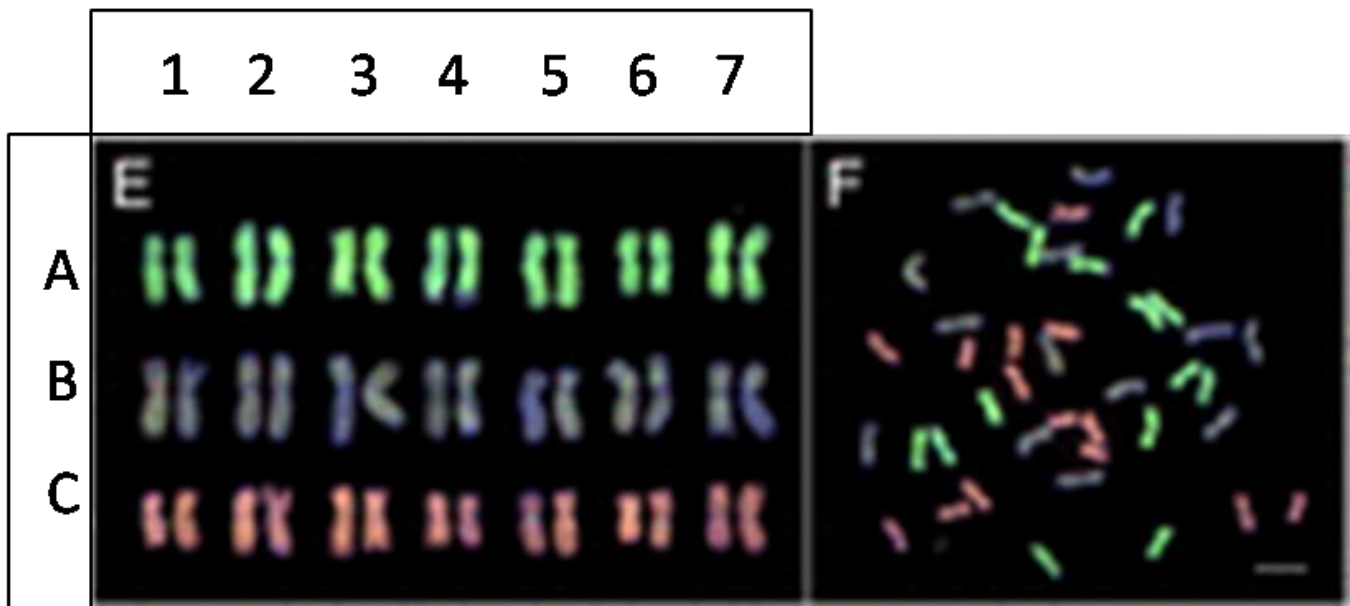


Fig. III-10 Caryotype d'une cellule de blé hexaploïde euploïde (jeu complet).

On retrouve les sept paires de chromosomes de chacun des génomes A, B et D.

(Zhang et al., 2013)

Il peut y avoir des mécanismes indirects *via* l'activation transcriptionnelle des rétroéléments et leur action sur des gènes adjacents. C'est le cas du rétrotransposon à LTR Wis2-1A inséré dans l'UTR 3' du gène similaire aux puroindolines, le résultat est son inactivation par une transcription antisens. A l'inverse, l'insertion d'un élément de la même famille dans le bon sens de transcription dans l'UTR 5' du gène aboutit à une sur activation de ce gène (Kashkush et al., 2003). Une autre équipe, Madlung et al. (2005) ont mis en évidence une activation transcriptionnelle de plusieurs éléments transposables suite à leur déméthylation mais leur transposition semble limitée.

Dans les mécanismes d'actions indirectes on trouve également les ARN interférents qui jouent un rôle important dans la régulation de l'expression des gènes dupliqués, notamment dans l'inactivation réciproque des gènes homéologues ou dans la sous-fonctionnalisation des copies homéologues (Chen and Ni, 2006).

Il a récemment été suggéré par Ha et al. (2009b) que les siRNA associés aux éléments répétés jouent un rôle dans le maintien d'une stabilité génomique et chromatinienne alors que les miRNA assurent la diversité de l'expression des gènes.

III-4 L'aneuploïdie : effets sur la structure et l'expression des génomes :

Nous avons parlé de l'impact de la polyploidie sur le génome, la majorité des exemples concerne des organismes ou les cellules sont euploïdes (Fig. III-10). Mais on obtient parfois des cas de figures où les cellules sont aneuploïdes, il y a eu la perte (hypoaneuploïde) ou le gain (hyperaneuploïde) de plusieurs chromosomes ou de fragments de chromosomes. Plusieurs anomalies méiotiques peuvent être à l'origine de l'aneuploïdie, la non-disjonction des chromosomes et/ou des chromatides pendant la méiose en étant la principale source.

L'aneuploïdie est peu tolérée chez les animaux. Chez l'homme, elle est souvent associée à des retards développementaux tels que la trisomie 21 et à certains cancers (Matzke et al., 2003). L'aneuploïdie n'est cependant pas toujours délétère et peut être dans certains cas persistante. L'exemple le plus connu concerne les chromosomes B. Ces chromosomes surnuméraires ont été observés chez diverses espèces végétales et animales et semblent avoir leur propre voies évolutives (Camacho et al., 2000). Il existe d'autres cas d'aneuploïdie non délétère comme par exemple l'occurrence de tissus aneuploïdes dans le cerveau qui semblent jouer un rôle fonctionnel (Kingsbury et al., 2005).

Les plantes semblent être plus tolérantes à l'aneuploïdie. Des individus aneuploïdes peuvent apparaître spontanément au sein des populations diploïdes (Khush, 1973). Les 12 lignées trisomiques possibles de *Datura stramonium* à l'origine de diverses variations phénotypiques constituent un excellent exemple de tolérance et de variabilité phénotypique induite par l'aneuploidie. Ainsi, pour chacun des chromosomes trisomiques un phenotype différent a été observé (Blakeslee, 1922).



Fig. III- 11 Ernest Sears (1910-1991).

Cytogénéticien, il a travaillé dans deux grands domaines, la systématique et l'étude de la structure cytogénétique (formation des polyploïdes).

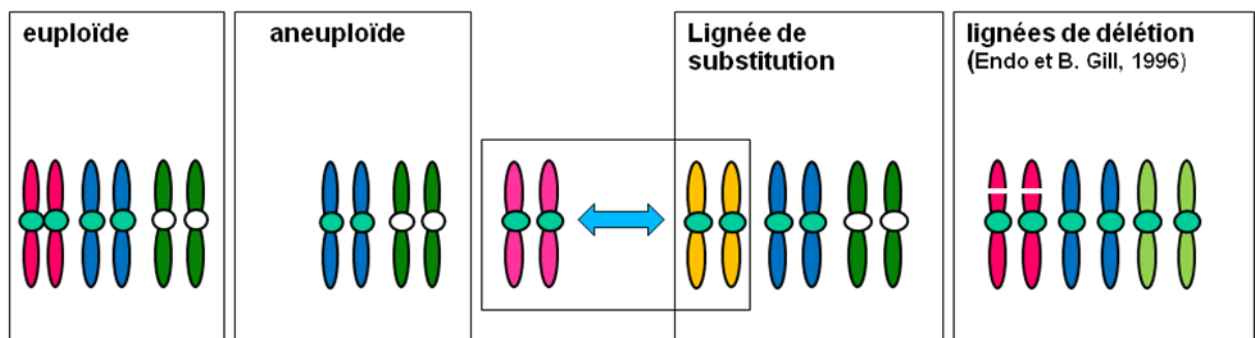


Fig. III-12 Exemple schématique de lignées aneuploïde, de substitution disomiques et de délétion.

- Plante euploïde (jeu complet)
- Plante aneuploïde (perte ou addition de chromosomes)
- Lignée de substitution (substitution d'une paire de chromosomes)
- Lignée de délétion (perte d'un segment chromosomique)

Chez les espèces polyploïdes plus ou moins récentes, la fréquence des individus aneuploïdes est nettement plus importante et varie d'une espèce à l'autre (Henry et al., 2006). Ainsi, le pourcentage d'aneuploïdes dans les populations tétraploïdes atteint 50% pour le maïs alors qu'il est < 6% chez la luzerne. Par ailleurs, Henry et al., (2006) ont montré que 25% de la descendance des tétraploïdes d'*A. thaliana* est aneuploïde (la majorité présentant un chromosome surnuméraire). Les aneuploïdes sont couramment rencontrés dans les populations naturelles de blés hexaploïdes (Riley and Kimber, 1961) et causent des problèmes d'homogénéité variétale retardant ou empêchant le processus d'inscription des variétés commerciales. De plus, la fréquence d'aneuploïdie est variable au sein des lignées et des variétés hexaploïdes (Storlie and Talbert, 1993 ; Worland and Law, 1985).

Cette bonne aptitude à tolérer l'aneuploïdie a été exploitée par Sears (1954) pour développer les lignées monosomiques, nulli-tétrasoniques, ditélosomiques et double ditélosomiques issues du cultivar Chinese Spring et qui ont été depuis largement utilisées dans les études génétiques (Law et al., 1975). Plus récemment, Endo & Gill (1996) ont développé une série de lignées présentant des délétions de longueur variable pour chacun des 21 chromosomes.

Il a été suggéré que l'aneuploïdie a un impact plus important sur le phénotype que la polyploïdie (Birchler and Veitia, 2007). En effet, les processus cellulaires sont sensibles aux changements de la « balance de gènes » : le bon fonctionnement d'un complexe protéique, par exemple, nécessite une relation stœchiométrique entre les produits des différents gènes impliqués et l'altération de la « dose » de chacun ou d'un de ces gènes perturbe cet équilibre (Birchler and Veitia, 2007). L'aneuploïdie constitue une altération de cette « balance de gènes » à l'échelle chromosomique et a de ce fait un effet majeur sur la structure et l'expression des génomes. C'est probablement le changement de la balance des gènes qui serait à l'origine des modifications phénotypiques observées.

III-4-1 Travaux d'Ernest Sears (1910-1991) :

Ernest Sears (Fig. III-11) a travaillé dans deux grands domaines, la systématique et l'étude de la structure cytogénétique du blé (formation des polyploïdes) (Ryley, 1995). C'est en étudiant la structure cytogénétique des hexaploïdes, en 1939, dans la variété Chinese Spring (*T. aestivum* alias *T. vulgare*), qu'il commence à collectionner les lignées monosomiques, nullisomiques, trisomiques et tétrasomiques issues de plantes haploïdes (croisement blé-seigle). C'est à partir de la lignée Nullisomique 3B qu'il a pu obtenir les autres, car son absence augmente le taux de mésappariement. Il a été jusqu'à l'obtention de lignées de substitutions inter-variétale (Fig. III-12). Ceci a permis la localisation de certains gènes, et de connaître leurs récessivités ou leurs dominances. C'est dans son laboratoire, à partir de 1278 plantes issues de la fécondation de plantes émasculées et monosomiques pour le chromosome 5B avec des grains de pollens euploïdes irradiés au rayon X, que le gène *Ph1* fut identifié chez un mutant *ph1* déficient pour le gène.

Un grand nombre de chercheurs sont passés dans son laboratoire et sont maintenant reconnus tel que K. Tsunewaki, Gordon Kimber, Bikram Gill, M. Muramatsu et tant d'autres. Ils ont su développer de nouveaux outils notamment des lignées de délétions qui offrent une localisation plus précise sur les chromosomes.

III-4-2 Les lignées de délétions :

Les lignées de délétions (Endo and Gill, 1996) ont été obtenues en plusieurs étapes :

1. La réalisation de croisements entre *T. aestivum* ssp. *aestivum* var Chinese spring ($2n=42$, AABBDD) avec *Ae. cylindrica* ($2n=28$, CCDD), *Ae. triuncialis* ($2n=28$, UUCC) ou *Ae. speltoides* ($2n=14$, SS). Cette étape permet d'obtenir des plantes monosomiques pour un chromosome homéologue du groupe 2 d'*Ae. cylindrica* ou *Ae. triuncialis* (on parle de la présence d'un chromosome « alien »). Pour les plantes obtenues avec *Ae. speltoides* on obtient des plantes avec une translocation d'un segment du bras court du chromosome homéologue du groupe 2 d'*Ae. speltoides* sur le chromosome 2B de Chinese Spring (la plante est donc T2BS.2BL-2SL). Le chromosome ou le segment « alien » serait porteur d'un gène important provoquant les cassures double-brins au niveau des chromosomes du blé (gametocidal gene).
2. Par rétrocroisement de ces plantes avec le cv. Chinese spring du blé hexaploïde, on obtient soit des plantes avec des délétions de segments d'ADN ('bin' de délétion ou « regions delimited by neighboring deletion breakpoint ») dans les chromosomes appartenant à Chinese Spring avec en plus le chromosome « alien » ou des plantes uniquement avec des délétions dans les chromosomes de Chinese Spring. Dans ces deux cas les plantes sont hétérozygotes pour les délétions. Ces plantes dans les cas extrêmes présentent des translocations de chromosomes ou sont aneuploïdes.
3. La dernière étape est l'autofécondation des plantes hétérozygotes (pour les délétions) afin d'obtenir des plantes homozygotes (collection de lignées de délétion). Ou le croisement des plantes hétérozygotes pour la délétion avec des lignées ditélosomiques (Sears and MS., 1978) ou nullisomiques-tétrahélosomiques (Sears, 1966) pour obtenir des lignées de délétion hémizygotes.

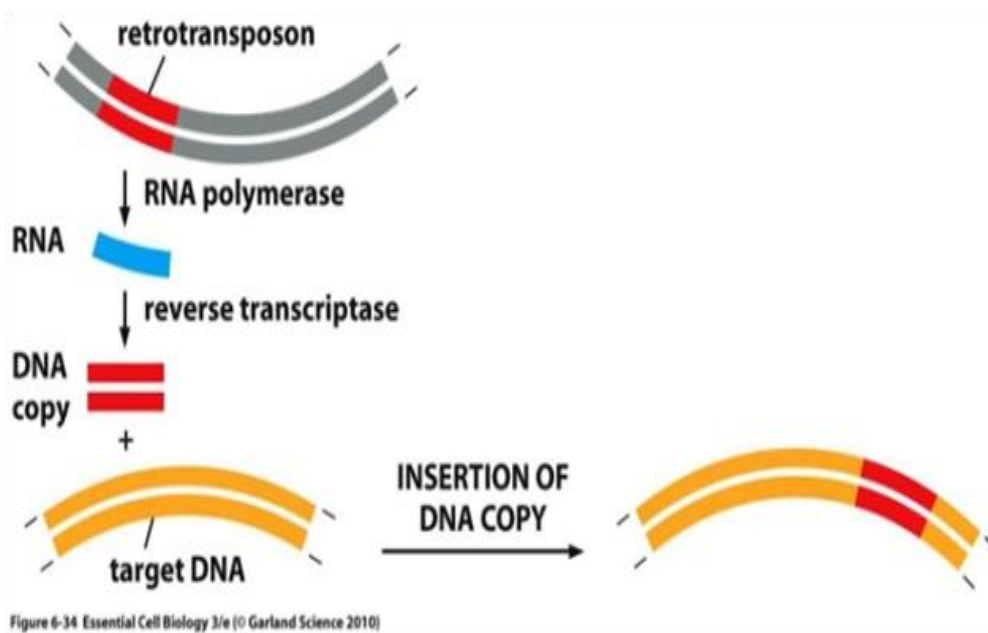
Par ces méthodes, 436 lignées de délétions ont été obtenues avec des délétions simples, doubles, triples ou quadruples dans une lignée (Endo and Gill, 1996). Les segments d'ADN chromosomiques délétés ont été identifiés par la technique de coloration Giemsa (Nilsson, 1973; Gill and Kimber, 1974b). Ces lignées ont été régulièrement utilisées pour la cartographie, pour l'hybridation *in situ*, des marqueurs biochimiques (Gill et al., 1991 ; Mukai et al., 1991 ; Yamamori et al., 1994 ; Delaney et al., 1995a ; 1995b).

Pour étudier la fonction des homéologues du gène de domestication (Q/q) dans un contexte de polyploidie, il était plaisant d'avoir à ma disposition des lignées de délétions ayant un ou plusieurs gènes Q/q homéologues absents ou substitués. Ces lignées de délétions (Endo et Gill, 1996) ont permis de décomposer l'implication fonctionnelle de chacune des copies.

CHAPITRE IV :

IMPORTANCE DES ELEMENTS TRANSPOSABLES DANS LE GENOME DU BLE

TE Classe I, Rétroelement, AND -> ARN -> ADN



TE Classe II, Transposon à ADN, ADN -> ADN

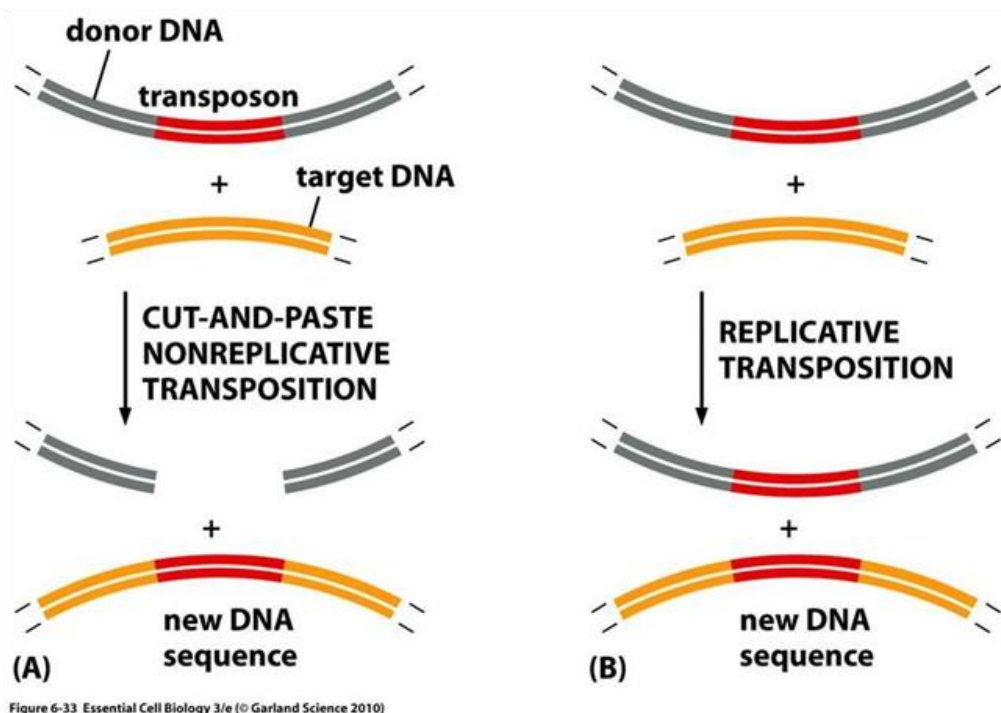


Fig. IV-1 Les deux mécanismes d'insertion des éléments transposables à la base de leur classification :

- en haut le mécanisme du copier-coller des TE de classes I
- en bas le mécanisme du couper-coller ou de transposition au moment de la réplication de l'ADN des TE de classe II

(Kwan, 2009)

IV-1 Description des éléments transposables :

Au début des années 80, les éléments transposables (TEs pour « Transposable Elements ») étaient considérés comme des parasites du génome, de l'ADN poubelle égoïste (Orgel and Crick, 1980 ; Doolittle and Sapienza, 1980.) ne servant en rien la cellule ou l'organisme et n'ayant que peu d'effets notables au niveau phénotypique. Pourtant le couple d'éléments mobiles (Activator /Dissociator) avaient été désignés comme 'éléments régulateurs', à leur découverte dans le maïs par Barbara McClintock en 1950. Mais ce n'est qu'à la fin des années 90, après la découverte de leur quasi-ubiquité et l'analyse de leurs effets sur les génomes grâce à la disponibilité croissante de leur séquence, que leur statut a peu à peu changé, il est passé de 'parasites' à 'partenaire symbiotique'.

Les TEs sont des fragments d'ADN génomique, qui ont la capacité de s'insérer, se déplacer et de modifier le nombre de leurs copies au sein de leur génome hôte. Depuis les années 80, de nombreux éléments très différents ont été découverts dans les génomes, rendant indispensable leur classification.

IV-1-2 Classification des éléments transposables :

On trouve des TEs dans tous les organismes vivants, mais leur importance relative dans le génome et leur diversité peuvent grandement varier selon l'espèce considérée. La première classification proposée (Finnegan, 1990 ; Capy et al., 1998) repose sur les différents modes et mécanismes de répllication observés (transposition). On distingue ainsi les TEs de classe I qui utilisent un intermédiaire ARN pour transposer et les TEs de classe II qui utilisent un intermédiaire ADN.

Ces deux mécanismes correspondent aussi à deux modes différents de propagation (Fig. IV-1). Les éléments de classe I génèrent une nouvelle copie de l'élément à partir de la copie originale qui va s'insérer à une autre place dans le génome (mécanisme de 'copier-coller'). Pour les éléments de classe II, la copie originale est excisée du génome puis réintégrée à un endroit différent (mécanisme de 'couper-coller').

Les TEs de classe I sont appelés rétroéléments en rapport avec la transcriptase reverse dont ils se servent pour copier l'ARN en ADN. On distingue deux catégories de rétroéléments: les rétrotransposons et les rétrovirus. Les rétrotransposons sont reconnaissables par leurs LTRs (Long terminal Repeat) qui correspondent à des longues séquences répétées au début et à la fin de l'élément. Les éléments de classe II sont appelés transposons à ADN.

L'arrivée massive de séquences depuis une dizaine d'année, provenant de génomes différents, a permis l'identification de nombreux nouveaux TEs. Un nouveau modèle de classification plus organisé est devenu nécessaire, notamment avec la découverte dans le maïs d'éléments (*Hélitrons*) utilisant un nouveau mécanisme de transposition (dit par cercle roulant ou 'rolling-circle') (Kapitonov et Jurka 2001).

ORDRE	SUPER FAMILLE	STRUCTURE	TSD	CODE	ESPECE
Classe I (rétroéléments : rétrotransposons et rétroposons)					
LTR	Copia	→ GAG AP INT RT RH →	4-6	RLC	P, M, F, O
	Gypsy	→ GAG AP RT RH INT →	4-6	RLG	P, M, F, O
	Bel-Pao	→ GAG AP RT RH INT →	4-6	RLB	M
	Retrovirus	→ GAG AP RT RH INT ENV →	4-6	RLR	M
	ERV	→ GAG AP RT RH INT ENV →	4-6	RLE	M
DIRS	DIRS	→ GAG AP RT RH YR ←	0	RYD	P, M, F, O
	Ngaro	→ GAG AP RT RH YR →	0	RYN	M, F
	VIPER	→ GAG AP RT RH YR →	0	RYV	O
PLE	Penelope	← RT EN →	Variable	RPP	P, M, F, O
LINE	R2	RT EN	Variable	RIR	M
	RTE	APE RT	Variable	RIT	M
	Jockey	ORF1 APE RT	Variable	RIJ	M
	L1	ORF1 APE RT	Variable	RIL	P, M, F, O
	I	ORF1 APE RT RH	Variable	RII	P, M, F
SINE	tRNA		Variable	RST	P, M, F
	7SL		Variable	RSL	P, M, F
	5S		Variable	RSS	M, O
Classe II (transposons à ADN) - Sous-classe 1					
TIR	Tc1-Mariner	→ Tase* ←	TA	DTT	P, M, F, O
	hAT	→ Tase* ←	8	DTA	P, M, F, O
	Mutator	→ Tase* ←	9-11	DTM	P, M, F, O
	Merlin	→ Tase* ←	8-9	DTE	M, O
	Transib	→ Tase* ←	5	DTR	M, F
	P	→ Tase ←	8	DTP	P, M
	PiggyBac	→ Tase ←	TTAA	DTB	M, O
	PIF- Harbinger	→ Tase* ORF2 ←	3	DTH	P, M, F, O
	CACTA	→ Tase ORF2 ←	2-3	DTC	P, M, F
Crypton	Crypton	→ YR ←	0	DYC	F
Classe II (transposons à ADN) - Sous-classe 2					
Helitron	Helitron	→ RPA // Y2 HEL ←	0	DHH	P, M, F
Maverick	Maverick	→ C-INT ATP // CYP POL B ←	6	DMM	M, F, O

Fig. IV-2 Classification des différents TEs sur 5 niveaux :

Classe, Sous-classe, Ordre, Super-famille, Famille d'après Wicker et al. (2007a). Le dernier niveau, sous-famille, n'est pas illustré sur cette figure. Les éléments caractéristiques de chaque catégorie sont mentionnés, ainsi que les espèces dans lesquelles ils ont été trouvés. La colonne TSD indique la taille de ceux-ci pour les éléments de cette famille. Le code indique le préfixe à faire figurer devant l'élément pour sa nomenclature. ORF1 et ORF2 représentent des protéines dont la fonction est inconnue.

Un modèle hiérarchique (Fig. IV-2) a été récemment proposé (Wicker et al., 2007a) pour classer les TEs d'eucaryotes en fonction de leur mode de transposition (au niveau mécanistique et enzymatique). Il prend en compte d'autres modèles (Jurka et al., 2005), mais aussi des règles pratiques venant des retours d'experts en annotation de TE. On peut citer la règle dite des '80-80-80' basée sur des similarités de séquences: deux éléments appartiennent à la même famille s'ils ont plus de 80% d'identité sur 80% de leur séquence sur un minimum de 80 pb. Le système binaire (classe I / classe II) a été conservé et enrichi de 5 niveaux : sous-classe, ordre, superfamille, famille et sous-famille.

- ✓ Les classes restent définies par la présence ou non d'un intermédiaire ARN lors de la transposition (transposons à ARN ou transposons à ADN).
- ✓ Les sous-classes séparent les transpositions par 'copier-coller' des transpositions par 'couper-coller'. Tous les éléments de classe I appartiennent donc à la même sous-classe ('copier-coller'), alors que les éléments de classe II peuvent appartenir à l'une ou à l'autre.
- ✓ Les ordres séparent les transpositions ayant des caractéristiques enzymatiques et organisationnelles différentes.
- ✓ Les superfamilles différencient des éléments ayant une même stratégie de réplication, mais une conservation au niveau protéique très limitée (exemple superfamille des *Copia* et des *Gypsy*).
- ✓ Les familles, où les éléments ont une forte conservation au niveau protéique (>80%).

Certaines familles ont des membres se regroupant en sous-familles sur des critères de similarité (conservation nucléique ou regroupement dans des arbres phylogénétiques). Exemple : Wis et Angela sont deux sous-familles de BARE-1.

Cette proposition de classification, même si elle n'est pas exempte de défauts, permet d'organiser simplement les TEs et aide à annoter les nouveaux éléments (décrits pour la première fois).

IV-1-3 Importance des TE dans les autres espèces :

La proportion des TEs est aussi très variable chez les plantes dicotylédones comme, le coton (Hawkins et al. 2006), ou par rapport à Arabidopsis (Arabidopsis and Initiative, 2000)(Fig. IV-3). Cette différence des proportions existe également dans le règne animal, pour l'homme (Lander et al., 2001) ils occupent 45% de la séquence du génome tandis que pour la drosophile l'occupation n'est que de 15% (Kaminker et al., 2002).

La proportion des TEs de classes I et II est également très variable selon les espèces (Pritham, 2009). Les éléments de classe I représentent plus de 60% du génome du blé, mais seulement 10% du génome de la drosophile. Ils sont globalement très abondants dans les plantes (Fig. IV-4). Les éléments de classe II sont près de 10 fois plus importants dans le génome du blé (20%) que celui du maïs (3%).

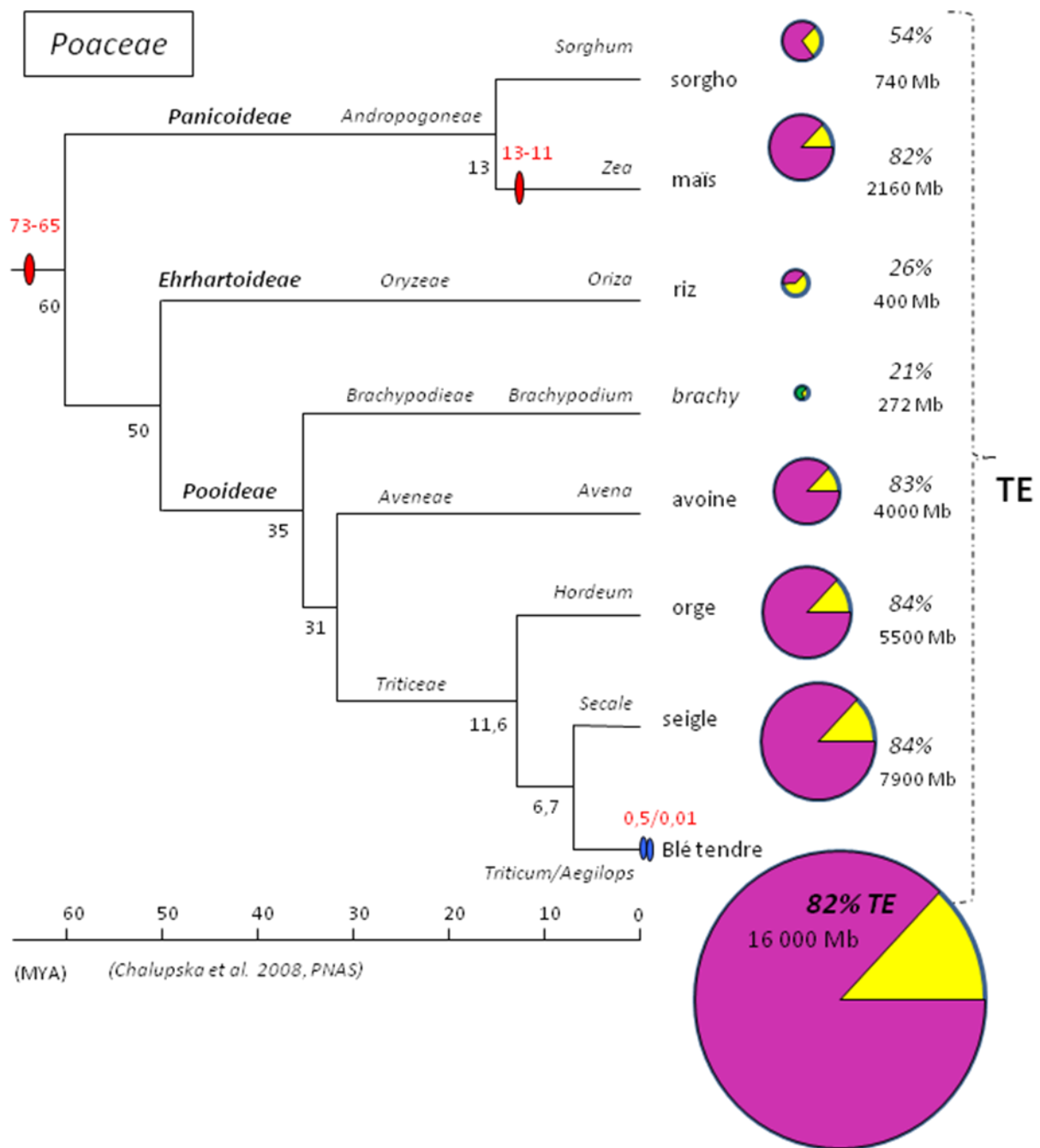


Fig. IV-3 Divergence des espèces de la famille des Poaceae, évènements de polyploïdisation, taille des génomes et proportion des TEs :

Divergences indiquées (Chalupska et al., 2008), la taille des génomes (Lander et al., 2001 ; Kaminker et al., 2002 ; International Rice Genome Sequencing Project, 2005 ; Paterson and al., 2009), la proportion des TEs (Hawkins et al., 2006 ; Paux et al., 2006 ; Piegu et al., 2006 ; Charles et al., 2008 ; Pri-
tham, 2009).

La taille des TEs peut également varier de quelques dizaines de paires de bases (pb) à quelques dizaines de milliers de paires de bases (kb), selon la famille des éléments. Certaines espèces contiennent plusieurs centaines de familles d'éléments alors qu'une seule famille d'éléments peut représenter la grande majorité des TEs présents dans un génome (exemple de l'élément Alu chez l'homme ou BARE chez l'orge).

IV-1-4 Dynamique des éléments transposables :

IV-1-4-1 Prolifération des TE :

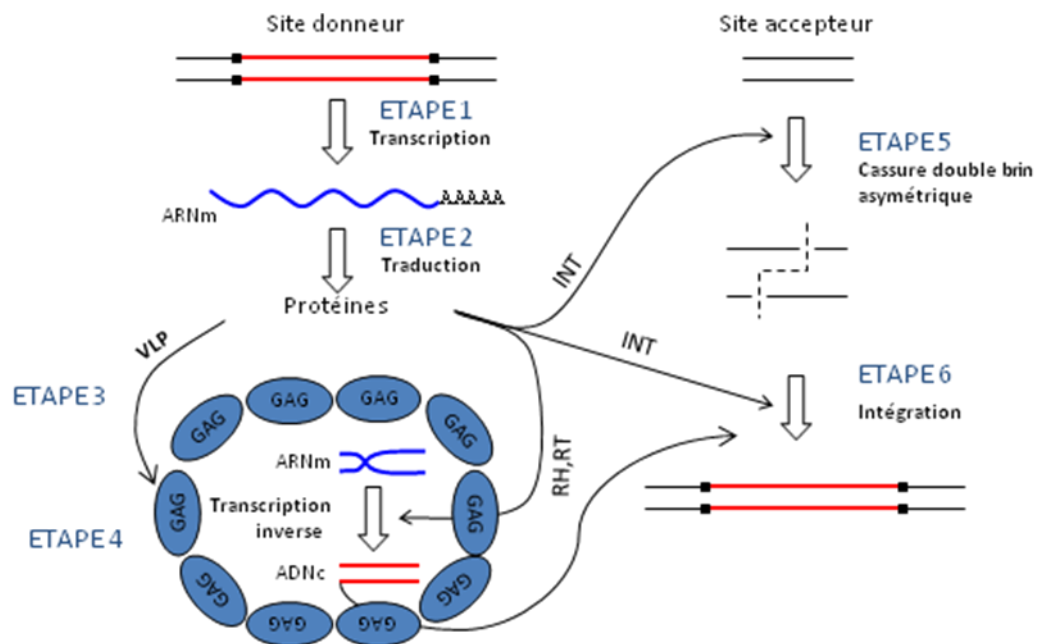
Les variations de la composition en TEs, trouvées dans les espèces, illustrent des dynamiques de prolifération très différentes. Cette prolifération est la résultante de deux forces d'évolution antagonistes : leur activité insertionnelle d'une part et leur élimination d'autre part (SanMiguel et al., 1996 ; Bennetzen and Kellogg, 1997 ; Petrov et al., 2000 ; Kidwell, 2002 ; Wendel et al., 2002 ; Bennetzen, 2002a ; Petrov, 2002a ; Bennetzen, 2002b ; Bennetzen et al., 2005 ; Hawkins et al., 2006 ; Piegu et al., 2006 ; Zuccolo et al., 2007). La disponibilité et l'abondance des séquences de TEs à l'échelle génomique ont permis une caractérisation de ces deux forces (SanMiguel et al., 1998 ; SanMiguel et al., 2002 ; Wicker et al., 2003b ; Gao et al., 2004 ; Ma and Bennetzen, 2004a ; Wicker et al., 2005 ; Du et al., 2006 ; Piegu et al., 2006 ; Wicker et al., 2007a ; Wicker and Keller, 2007b). Il est en effet devenu possible de déterminer :

- ✓ la proportion des TEs et des différentes familles dans le génome
- ✓ la proportion des copies complètes, ou tronquées
- ✓ la distribution des TEs le long des chromosomes
- ✓ l'estimation des dates d'insertion des rétrotransposons ayant leur deux LTR

La prolifération des TE n'est pas constante au cours de l'évolution des espèces, ni homogène le long des chromosomes. On distingue des périodes d'activité très fortes et des périodes de plus faible intensité. De plus, elle semble différente d'une espèce à une autre et d'une famille de TE à une autre. Chez les angiospermes, les estimations des dates d'insertions des rétrotransposons n'excèdent pas 3 Ma alors que chez les gymnospermes, la plupart des estimations sont supérieures à 35 Ma (Nobuta et al., 2008). De plus, un grand nombre de copies apparaissent tronquées suggérant des mécanismes d'élimination (délétion) des TE.

Il est difficile de séparer ces deux forces lorsque l'on étudie la dynamique globale des TEs dans un génome. Par exemple, des pics apparents d'insertions de TEs ('bursts') dans un génome peuvent résulter d'une forte activité insertionnelle et/ou d'une faible vitesse d'élimination. J'ai contribué à l'étude de ces deux forces d'évolution des TEs dans le blé au cours de mon activité de recherche en pré-thèse au laboratoire ((Charles et al., 2008), article en annexe).

A – Transposition du rétrotransposons (Classe I) :



B – Transposition du rétroposon (Classe I) :

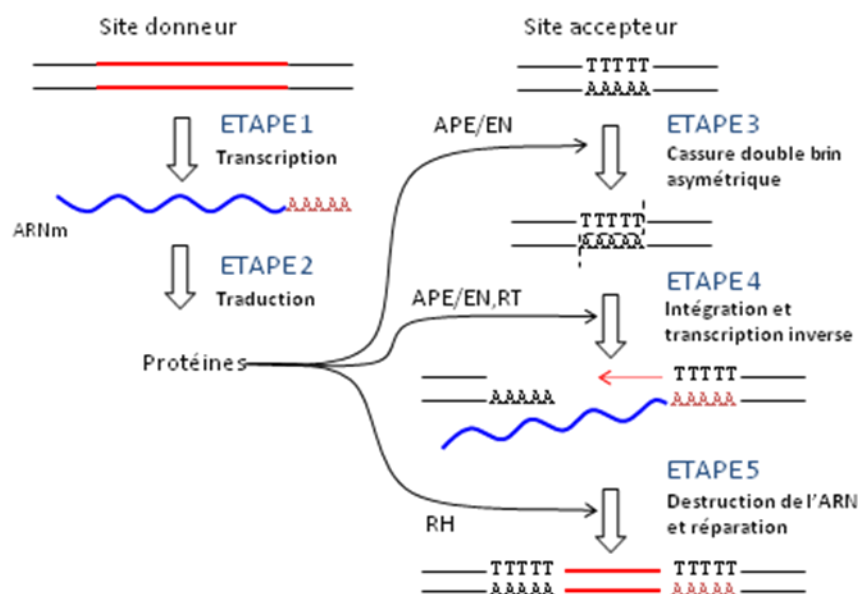


Fig. IV-4 Mécanismes de transposition des principaux TEs de classe I :

- Transposition des rétrotransposons. Les carrés noirs indiquent les « target site of duplication » ou TSD.

IV-1-4-2 Mécanismes d'insertion des TEs :

Les TEs qui peuvent coder les protéines essentielles pour leur transposition sont dits autonomes. Ils ont néanmoins besoin de la machinerie de la cellule hôte pour traduire ces protéines. Lorsqu'un élément autonome est actif, il peut transposer mais aussi induire la transposition d'éléments de la même famille dits non-autonomes qui n'encodent pas toutes les protéines essentielles pour leur transposition. Un élément non-autonome peut avoir conservé les sites de reconnaissance (peu spécifiques) de la machinerie de transposition et être mobilisé en 'trans'.

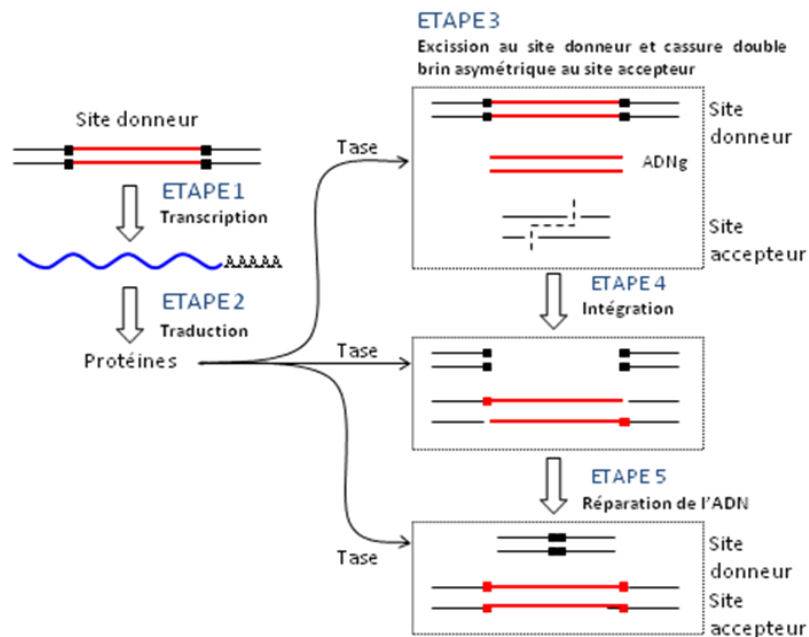
Une région codante parfaitement conservée ne garantit pas l'activation et la transposition de l'élément et inversement une région codante dégradée ne va pas forcément empêcher la transposition. Un faible nombre de copies autonomes actives dans un génome peut permettre la transposition d'un grand nombre de copies non-autonomes ('trans' activation). Les mécanismes de réplication des TEs de classes I et II sont très différents. Les éléments de classe I passent par un intermédiaire ARN. Le mode de transposition par 'copier-coller' des éléments de classe I les rend potentiellement très invasifs pour un génome. Il n'est donc pas surprenant de les voir représenter parfois une grande partie d'un génome. C'est particulièrement le cas des rétrotransposons, plus fréquents chez les plantes que chez les animaux, occupant jusqu'à 60% des grands génomes de céréales (blé, orge ou maïs). Il est intéressant de constater que quelques familles de TEs peuvent représenter une grande partie de la multitude de TEs trouvés dans un génome.

IV-1-4-3 Mécanisme de transposition des rétrotransposons :

Les rétrotransposons sont les éléments les plus répandus dans les génomes du blé (près de 60% de la séquence). Les sous-familles Wis et Angela de la famille *BARE-1* (Copia) sont les plus fréquentes et représentent à elles seules de 10 à 20% du génome du blé (Charles et al., 2008). Les familles Sabrina (*Athila*) et Fatima (*Gypsy*) occupent chacune près de 7% du génome.

Les rétrotransposons autonomes encodent une polyprotéine comprenant deux domaines : GAG et Pol. Les deux domaines sont transcrits en une fois (Fig. IV-4A, étape 1). La traduction du messager donne la protéine GAG, protéine capsidique, et le complexe de protéines Pol, divisé en 4 protéines distinctes au niveau post-traductionnel, RT : transcriptase inverse, INT : intégrase, RH : RnaseH, AP : protéase aspartique (Fig. VI-4A, étape 2). Les protéines GAG vont se polymériser dans le cytoplasme pour former des VLP, Virus Like Particules, (Fig. IV-4A, étape 3). Une partie de l'ARNm va entrer dans ces VLP, se dimériser et, sous l'action de RH et RT, être retro-transcrit en ADN double brin (Fig. IV-4A, étape 4). L'intégrase, INT, va former une cassure double brin de l'ADN génomique et y intégrer l'ADN du rétrotransposon (Fig. IV-4A, étape 5 et 6). La réparation de cette cassure asymétrique va former les TSDs, Target Site Duplication, signatures caractéristiques laissées par l'insertion de la plupart des TEs.

A – Transposition du Transposon à ADN (Classe II) :



B – Transposition de l'Hélitron (Classe II) :

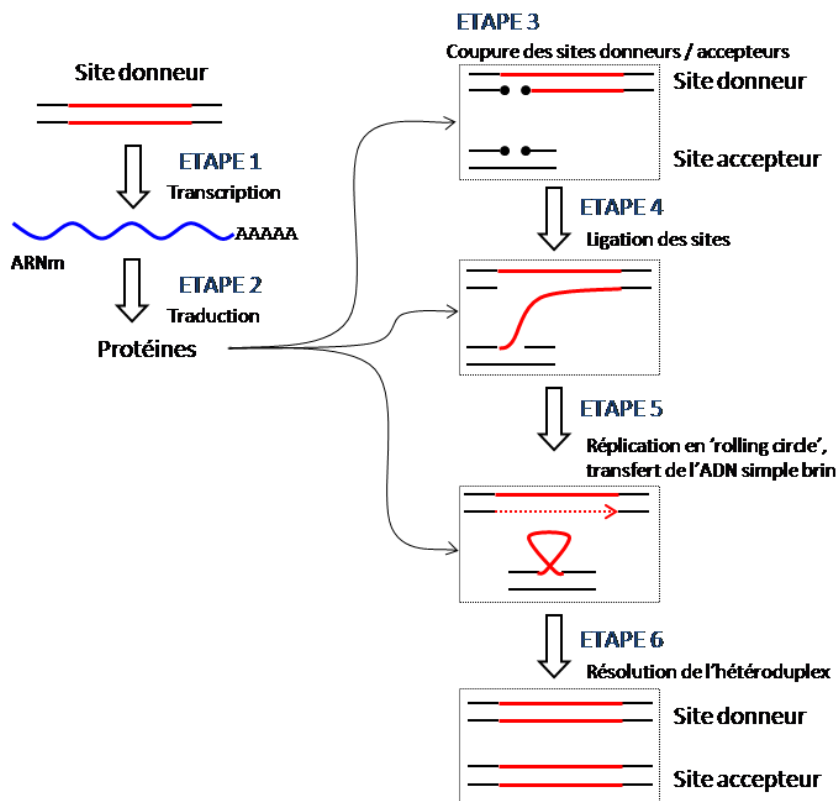


Fig. IV-5 Mécanismes de transposition des principaux TEs de classe II :

(A) Transposition des transposons à ADN de type *CTA*. Les carrés noirs et rouges indiquent les « tandem inverted repeat » ou TIR. (B) Transposition des hélitrons. Les points noirs indiquent le site de clivage entre A et T. Les différentes étapes des transpositions sont détaillées dans le texte correspondant. (inspiré des schémas des articles (Sabot et al., 2004) et (Kapitonov and Jurka, 2006)).

Le cycle de transposition supposé des rétrotransposons, est calqué sur celui des rétrovirus. Ils ont, en effet, une origine commune et gardent de nombreux points communs, même si l'absence de protéine d'enveloppe chez les rétrotransposons reste une différence notable, les confinant dans la cellule et empêchant leur transfert horizontal.

IV-1-4-4 Mécanisme de transposition des rétroposons :

Les rétroposons sont peu fréquents dans les génomes du blé (autour de 2%) et sont essentiellement de la famille des *LINES*.

Les rétroposons autonomes (comme les *LINES*) ont un mécanisme de transposition assez différent de celui des rétrotransposons. Pourtant, le principe général de la transposition reste le même : transcription puis traduction des parties codantes de l'élément par la machinerie de l'hôte (Fig. IV-4B, étape 1 et 2), transcription reverse de l'ARN en ADN, cassure asymétrique double brin de l'ADN au site cible suivi par l'intégration de l'élément. Une des différences principales entre les deux mécanismes vient de la transcription inverse qui se fait sur le lieu d'insertion pour les rétroposons, et pas dans une VLP. Un rétroposon complet a généralement deux cadres ouverts de lecture (ORFs pour Open Reading Frames) : ORF1 et ORF2. Le premier code pour une protéine pouvant se lier à l'ADN et le deuxième pour une endonucléase (APE ou EN), une transcriptase inverse (RT) et parfois une RnaseH (RH).

L'endonucléase va permettre de lier un ARNm de l'élément avec la partie 3' d'un fragment d'ADN libre après une cassure double brin asymétrique, provoquée ou non par l'endonucléase (Fig. IV-4B, étapes 3 et 4). La transcription inverse (action de RT) commence alors sur place, dans le sens 3' vers 5' (Fig. IV-4B, étape 4). Elle se déroule cependant rarement en entier (10% des cas), produisant donc de nombreuses copies partielles de l'élément d'origine. Une fois la transcription terminée, l'ARNm est digéré (action de RH) et la cassure double brin se répare, dupliquant ainsi le rétroposon sur le deuxième brin d'ADN (Fig. IV-4B, étape 5).

IV-1-4-5 Mécanisme de transposition des transposons à ADN :

Les transposons à ADN sont généralement moins invasifs que les rétrotransposons dans les génomes de plantes, car ils transposent par un système de 'couper-coller'. Cependant, ils constituent dans le blé jusqu'à 16% du génome dont la majorité appartient à la superfamille des *CACTA*. Pour expliquer cette amplification importante, il a été suggéré que leur transposition à des moments précis du cycle cellulaire permettrait d'augmenter leur nombre (décrit ci-après).

Les transposons à ADN complet de type *CACTA* ont deux ORFs, le premier encodant pour une transcriptase (Tase) et le second pour une protéine pouvant se lier à l'ADN mais dont le rôle reste à préciser (Wicker et al. 2003a), les deux protéines sont transcrites et traduites par la machinerie de l'hôte (Fig. IV-5, étape 1 et 2). La transcriptase reconnaît spécifiquement les TIRs (Tandem Inverted Repeat)

présents aux extrémités de ces TEs et catalyse toutes les étapes de la transposition, de l'excision à l'intégration (Fig. IV-5A, étape 3, 4 et 5). La transposition se fait donc par un mécanisme de 'couper-coller', conservatif au niveau du nombre de copies. Mais si la transposition se produit en phase S du cycle cellulaire, en aval de la fourche de réplication, la cassure double brin provoquée par le transposon va se réparer en utilisant la chromatide sœur comme modèle possédant encore l'élément. L'élément est donc copié à un autre endroit du génome tout en gardant une copie à sa position d'origine.

IV-1-4-6 Mécanisme de transposition des *Hélitrons* :

Les *Hélitrons*, découverts récemment (Kapitonov and Jurka, 2001), utilisent un mécanisme singulier, dit de 'rolling-circle' en référence à un mécanisme similaire observé chez les gémivirus pour leur transposition. Ces éléments, transposant sans intermédiaire ARN, par un mécanisme original, et par 'copier-coller', étaient difficiles à décrire selon les critères de la première classification. Ils ont ainsi contribué à la formation d'une nouvelle classification, présentée précédemment. Contrairement à la plupart des autres éléments de classe II, ils n'ont pas de TIRs, motifs reconnus spécifiquement par les transposases. Ils commencent par TC, finissent par CTRR (R étant A ou G) et ont aussi une séquence palindromique, de 16-20 pb situé une dizaine de pb en amont du CTRR. Ils s'insèrent entre les bases A et T d'un site accepteur. Selon les espèces, la composition des *Hélitrons* autonomes peut être assez variable, mais ils ont tous en commun un ORF codant pour la protéine 'RepHel' (Kapitonov and Jurka, 2006), constituée des domaines Rep (Réplicase) et Hel (Helicase).

Pour la transposition d'un *Hélitron*, suite à la transcription puis traduction des parties codantes de l'élément par la machinerie de l'hôte (Fig. IV-5B, étape 1 et 2), le domaine Rep se lie aux sites donneur et accepteur, coupe l'ADN (simple brin) entre les bases A et T des deux sites (Fig. IV-5B, étape 3) et fait une ligation entre la partie 3' du site donneur et la partie 5' du site accepteur (Fig. IV-5B, étape 4). Le domaine Hel catalyse ensuite la synthèse de l'*Hélitron* par l'ADN polymérase de l'hôte au niveau du site donneur (Fig. IV-5B, étape 5). En fin de synthèse, l'ADN simple brin correspondant à l'*Hélitron* original est transféré au site accepteur formant ainsi un hétéroduplex (Fig. IV-5B, étape 5). Lors d'une prochaine réplication de l'ADN, cet hétéroduplex va être résolu par la réplication de l'*Hélitron* sur le deuxième brin du site donneur (Fig. IV-5B, étape 6). Le palindrome à la fin de l'*Hélitron* joue le rôle de terminateur de la réplication par 'rolling-circle'. S'il n'est pas bien reconnu, la synthèse continue sur le site donneur, si bien que le brin d'ADN transféré au site accepteur peut transporter non seulement l'*Hélitron*, mais aussi des séquences en 3' du site donneur. Ce phénomène a été notamment observé dans le maïs (Morgante et al., 2005) où des fragments de gènes voire des gènes entiers sont ainsi transportés à d'autres endroits du génome.

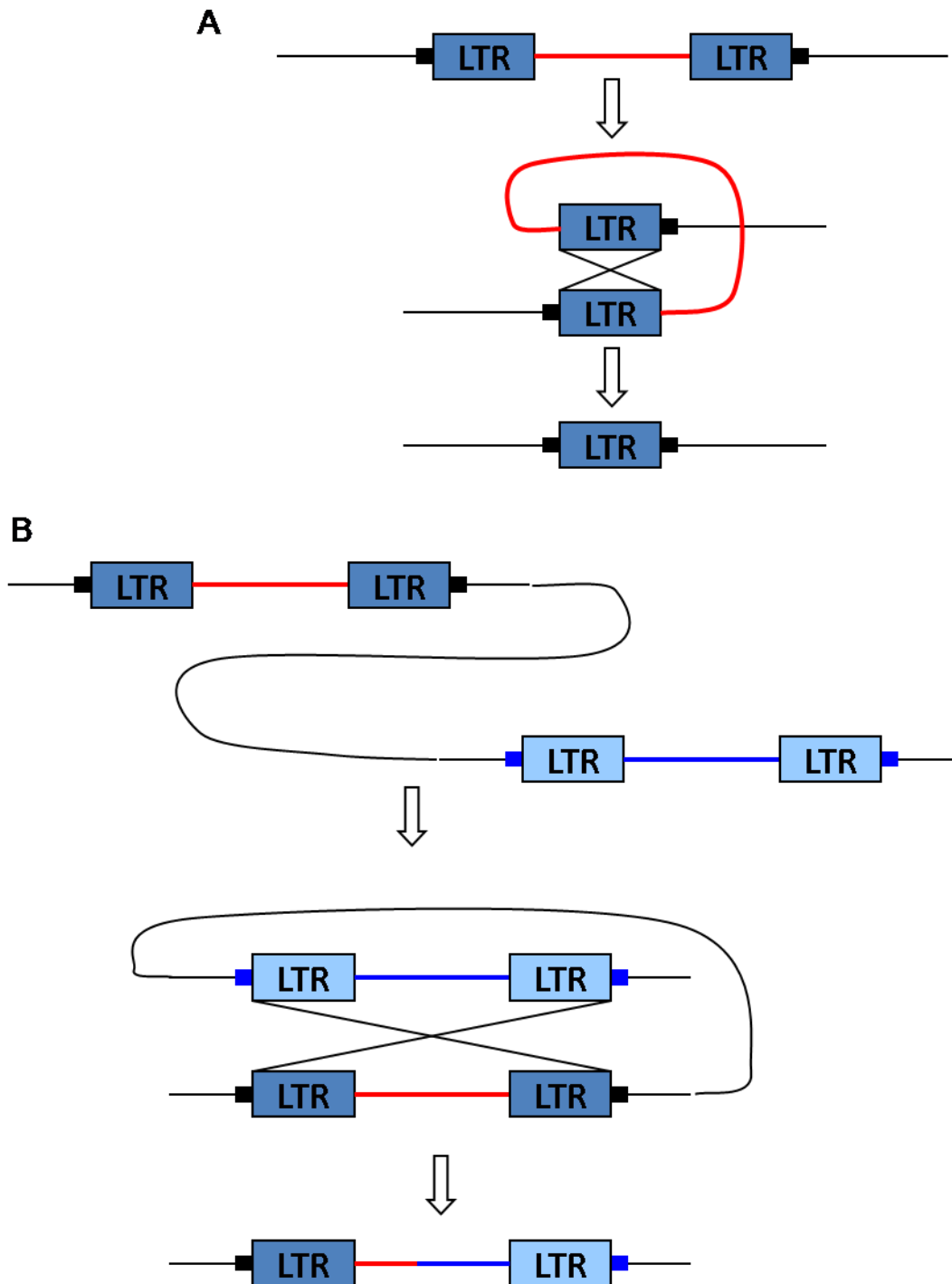


Fig. IV-6 Mécanismes schématiques de délétion par recombinaison homologue inégale

- A. Formation d'un Solo-LTR avec TSD par recombinaison homologue inégale entre les deux LTRs d'un rétrotransposon sur le même brin d'ADN.
- B. Formation d'un rétrotransposon complet mais chimérique, sans TSD, par recombinaison entre deux éléments suffisamment similaires, sur le même brin d'ADN. Toute la séquence d'ADN entre les deux éléments est éliminée.

IV-1-4-7 Mécanismes d'élimination des TEs :

Les TEs ne restent pas intacts dans le génome hôte après leurs insertions, comme le montre le nombre important de copies tronquées et dégénérées trouvées dans les séquences génomiques disponibles. Les deux principaux mécanismes responsables de cette élimination des TEs sont les recombinaisons homologues inégales et les recombinaisons illégitimes (Devos et al., 2002 ; Ma and Bennetzen, 2004a).

Les recombinaisons homologues inégales font intervenir deux séquences suffisamment longues (généralement plusieurs centaines de pb) et similaires (>85-100% d'identité, selon la taille). Les rétrotransposons sont donc particulièrement concernés par ce type de recombinaison puisque leurs LTRs remplissent très bien ces critères. La recombinaison peut ainsi avoir lieu entre les deux LTRs d'un même élément (formation d'un Solo-LTR) (Fig. IV-6A) ou des LTRs d'éléments de la même famille (Fig. IV-6B), aboutissant à la délétion plus ou moins importante de toute la séquence entre les deux éléments. Le taux de recombinaisons homologues, révélées par la présence de Solo-LTR et d'éléments complets sans TSD, n'est pas très important dans le blé (taux de 1/50, (Charles et al., 2008)).

Les recombinaisons qui ne sont pas homologues sont dites 'illégitimes'. Le mécanisme des ces recombinaisons illégitimes est encore mal caractérisé. Elles impliquent souvent des motifs de quelques pb, conservés dans des orientations variables (directs, complémentaires, anti-sens). On observe ainsi des délétions allant de quelques pb à plusieurs dizaines de kb (Chantret et al., 2005). Ces recombinaisons sont très fréquemment associées avec les TEs (intra-éléments ou inter-éléments).

IV-1-5 Rôle des Eléments Transposables :

Une des conséquences de leur activité est la génération de diversité au sein d'une population et d'une espèce par une action directe ou indirecte. L'insertion d'un TE dans la séquence codante d'un gène va le plus souvent l'inactiver. Son insertion dans la zone promotrice du gène ou la présence d'un LTR, contenant des régions promotrices peut aussi altérer l'expression du gène (Kashkush et al., 2003). La méthylation par l'organisme d'un TE peut également s'étendre jusqu'à un gène proche et conduire à l'inactivation transcriptionnelle ('silencing') de celui-ci. Les TEs peuvent aussi participer à la création de nouveaux exons ou même de nouveaux gènes. En effet, des études ont montré que certains TEs (*Hélitrons*, *PACK-Mule*, *CACTA* éléments de classe II) peuvent emporter des exons (ou même la totalité des gènes adjacents) lorsqu'ils transposent (Le et al., 2000 ; Jiang et al., 2004 ; Lai et al., 2005). Ce phénomène 'd'exon-shuffling', aboutit le plus souvent à des pseudogènes. Parfois, ces fragments sont capturés par un gène existant (néo-fonctionnalisation) ou peuvent même former, dans de rares cas, un gène complètement nouveau (Jiang, 2004 ; Morgante et al., 2005).

La première ‘domestication’ (recrutement) d’un TE par l’organisme a été identifiée chez *Drosophila* (Levis et al., 1993). Les TEs de type LINE (HET-A et TART) sont laissés libres de s’insérer dans les télomères remplaçant ainsi l’activité d’une télomérase. Depuis, de nombreux autres cas ont été détectés chez les animaux comme chez les végétaux où l’organisme utilise les TE d’une façon qui lui est bénéfique (Agrawal et al., 1998 ; Lynch and Tristem, 2003 ; Kapitonov and Jurka, 2004 ; Gao and Voytas, 2005 ; Kapitonov and Jurka, 2006). Chez l’homme, ils ont contribué à la formation de notre système immunitaire (Agrawal et al., 1998).

Plus les TEs sont étudiés, plus on découvre le rôle important qu’ils ont joué ou jouent dans les organismes. Les études récentes ont montré leur implication au niveau de la forme et de la fonction des chromosomes par leur insertion dans l’hétérochromatine autour des centromères et des télomères mais aussi au niveau de la régulation de l’expression et les modifications de la chromatine par leurs liens avec les ARN interférents, RNAi (revue dans (Slotkin and Martienssen, 2007)).

IV-1-6 Eléments Transposables et polyploïdie :

Cette activation des TEs peut avoir un impact majeur sur l’organisation et le fonctionnement des génomes polyploïdes. En effet, il est aujourd’hui admis que leur transposition s’accompagne de divers remaniements structuraux (cassures chromosomiques, délétions...) et/ou d’un changement de l’expression des gènes (Slotkin and Martienssen, 2007). L’insertion d’un TE dans une séquence codante peut générer un allèle nul. Par ailleurs, l’activation transcriptionnelle d’un élément, inséré à proximité d’un gène, peut entraîner l’activation ou l’inactivation de ce gène adjacent. De plus, les TEs sont régulés par des mécanismes épigénétiques, notamment la méthylation de l’ADN qui peut être altérée chez les espèces polyploïdes.

Il n’est pas encore prouvé que la polyploïdisation entraîne une activation d’insertion des éléments transposables ou le « burst » d’éléments transposables. Mais des changements de l’état de méthylation des séquences de TEs ont été constatés suite à la polyploïdisation. Mon travail de pré-thèse montre que les « burst » ont eu lieu (~3Ma) bien avant la polyploïdisation chez le blé tétraploïde (Charles et al, 2008).

CHAPITRE V :

LES GENES DE DOMESTICATION



Fig. V-1 Comparaison entre la forme « squarehead » (tétraédrique) et speltoïde de l'épi :

La photo montre (A1) la forme « squarehead » (tétraédrique) et (A2) le battage facile de l'épi du blé domestiqué, (B1) la forme speltoïde et (B2) la déhiscence des épillets de l'épi du blé non domestiqué. (Zhang et al., 2011)

V-1 Pourquoi caractériser les gènes de domestication :

Avant de développer les résultats que j'ai obtenus durant ma thèse, il me paraît important de citer quelques travaux précurseurs qui ont été réalisés sur le sujet et qui sont complémentaires aux résultats obtenus sur l'origine des blés cultivés. En effet, découvrir les mécanismes ou les gènes différenciant les espèces sauvages des espèces domestiquées ont toujours suscité beaucoup d'engouements.

« Il était une fois, un homme qui s'est mis à cultiver le blé sauvage, puis le blé sauvage et domestiqué puis uniquement le blé domestiqué »

Cette phrase fait allusion aux travaux menés sur l'apparition des blés domestiqués à partir des blés sauvages que j'ai développés dans le chapitre consacré à l'origine des blés cultivés et le processus de domestication (Tanno and Willcox, 2006).

En effet, la transition entre les blés cultivés sauvages, l'engrain sauvage (*T. monococcum* ssp. *beotium*), l'amidonnier sauvage (*T. turgidum* ssp. *dicoccoides*), aux blés domestiqués cultivés, l'engrain (*T. monococcum* ssp. *monococcum*), l'amidonnier (*T. turgidum* ssp. *dicoccum*), le blé tendre (*T. aestivum* ssp. *aestivum*), serait liée à une lente sélection de la perte de la faculté de dissémination (déhiscence) parmi d'autres caractères tels que le rendement.

On sait également que le tétraploïde domestiqué (amidonnier) est à la base des blés hexaploïdes qui se sont formés au cours de la domestication du blé (Dvorak et al., 2011). Depuis près d'un siècle les scientifiques essayent d'expliquer les bases génétiques différenciant les blés sauvages des blés domestiqués. Les premiers objectifs étaient principalement de caractériser le(s) gène(s) responsable(s) du passage du phénotype sauvage, des épis speltoïdes (épis allongés, grains carénés par une glume solide) au phénotype « domestiqué » ou « squarehead » (épi tétraédrique, condensé ou compact, glumes facilement détachables) indispensables, respectivement, pour la non-déhiscence des grains et pour un battage plus facile du blé (Fig. V-1).

Ces résultats représentent un travail de longue haleine mené par des scientifiques qui a débuté au début du siècle dernier que je résumerai par la suite, car ils ont été importants dans ma réflexion sur mon sujet de thèse.

V-2 Identification du gène Q :

En 1917, Nilsson-Ehle observe qu'il apparaît au sein d'une culture de blé des «variants» qui présentent un phénotype speltoïde (sauvage), probablement révertants (Nilsson-Ehle, 1917). En 1929, Winge établit que le phénotype speltoïde (Fig. V-1) observé est dû à la présence d'un facteur sur une paire de chromosomes qu'il baptisera 'C' (Winge, 1929). En 1958, Sears les identifie comme les chromosomes 5A.



Fig. V-2 Croisement entre *T. aestivum* ssp. *aestivum* (5AQ) et *T. aestivum* ssp. *spelta* (5Aq) d'après Muramatsu (1962) :

- (a) et (b) : A(QQ) B(qq) D(qq) → *T. aestivum* ssp. *aestivum* var. Chinese Spring (CS)
- (e) et (f) : A(qq) B(qq) D(qq) → *T. aestivum* ssp. *spelta*
- (c) et (d) : A(QQ) B(qq) D(qq) → hybride F1 (CS x *T. aestivum* ssp. *spelta*)

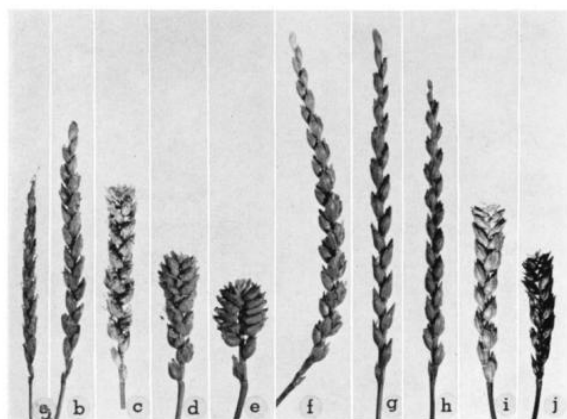
(Muramatsu, 1963)

Watkins suggère la coexistence de deux gènes qu'il nomme 'k' et 'q' ; 'k' serait responsable du grain non-caréné et de la suppression du phénotype speltoïde, tandis que 'q' contrôlerait la forme compacte de l'épi (Watkins, 1927). Pendant longtemps on a pensé que les deux gènes étaient liés sur le même chromosome, jusqu'à ce que MacKey démontre qu'il s'agit d'un seul et même gène qu'il baptisera le gène *Q* du mot anglais "Squarehead" qui signifie 'tétraédrique', la forme compacte de l'épi pour les blé domestiqués (MacKey, 1954).

V-2-1 Les différents caractères sous l'influence du gène *Q* :

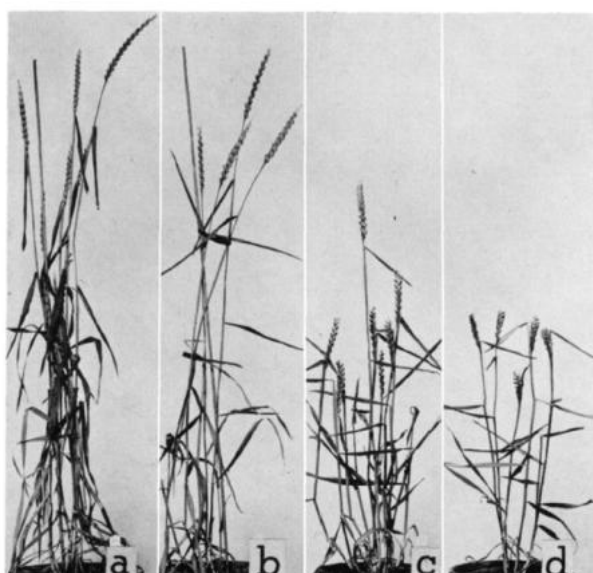
En plus du contrôle des caractères cités plus haut, le gène *Q* influence d'autres caractères contrôlés par d'autres gènes qui varient selon le fond génétique. Pour valider cette théorie, Tsunewaki et Jenkins utilisent des lignées monosomiques et démontrent que ces gènes se trouvent également sur le chromosome 5A. Ils baptiseront la région génomique trouvée le locus *Q* (Tsunewaki and Jenkins, 1961).

- En 1963, Muramatsu confirme son existence.
- En 1969, en irradiant des blés domestiqués, Singh découvre tout un éventail de variations dans d'autres caractères (taille de l'épi, taille de l'épillet, etc.) chez des mutants au phénotype speltoïde.
- En 1963, Muramatsu montre l'influence du fond génétique qui affecte plus ou moins l'expression des gènes influencés par le gène *Q*. Mais il démontre que quelque soit le fond génétique, si l'effet dose du gène *Q* ou *q* est suffisamment forte, tout les caractères contrôlés par le gène *Q/q* donnent un phénotype domestiqué tandis qu'en dessous d'une certaine dose le phénotype est sauvage (Fig. V-2). Il en arrive à la conclusion que pour le cultivar Chinese Spring la forme domestiqué "squarehead" est complètement récessive tandis que la ténacité de la glume et la fragilité du rachis sont majoritairement dominants.
- En 1971, Leighty and Boshnakian montrent que la ténacité de la glume et la fragilité du rachis sont sous l'influence du locus *Q*, mais c'est en 1974 que Kerber et Roland montrent que le gène *Tg* (*Tough Glume*) sur le chromosome 2D d'*Aegilops tauschii* est épistatique au gène *Q*. En croisant un tétraploïde (*QQ*) et le diploïde *Ae. tauschii* (*Tg*), ils obtiennent des blés hexaploïdes synthétiques avec un phénotype 'spelta' et non battable. Leur conclusion est qu'un génotype *QQtgtg* est nécessaire pour qu'un blé soit battable. En général, le gène *Q* est considéré partiellement dominant par rapport à *q*, les plantes *Qq* ont une morphologie intermédiaire entre la forme sauvage *qq*, speltoïde, et la forme domestiquée *QQ*, squarehead (Fig. V-2).

**Fig. V-3 Effet dose des gènes *Q/q* (épis) :**

- a. nullisomic 5A [A(--) B(qq) D(qq)]
- b. monosomic 5A [A(-Q) B(qq) D(qq)]
- c. disomic 5A (CS) [A(QQ) B(qq) D(qq)]
- d. trisomic 5A [A(QQQ) B(qq) D(qq)]
- e. tetrasomic 5A [A(QQQQ) B(qq) D(qq)]
- f. *T. spelta* [A(qq) B(qq) D(qq)]
- g. trisomic spelta 5A [A(qqq) B(qq) D(qq)]
- h. tetrasomic spelta 5A [A(qqqq) B(qq) D(qq)]
- i. trisomic spelta 5A + iso 5A [A(qqqqq) B(qq) D(qq)]
- j. tetrasomic spelta 5A + iso 5A [A(qqqqq) B(qq) D(qq)]

(Muramatsu, 1963)

Fig. V-4 Effet dose du gène *q* (plantes) :

- a. Disomic 5A [A(qq) B(qq) D(qq)]
- b. Tetrasomic 5A [A(qqqq) B(qq) D(qq)]
- c. Trisomic 5A + iso 5A [A(qqqq) B(qq) D(qq)]
- d. Tetrasomic 5A + iso 5A [A(qqqqq) B(qq) D(qq)]

(Muramatsu, 1963)

Fig. V-5 Description des effets dose des gènes *Q/q* :

Tableau recapitulatif des lignées obtenues par Muramatsu et des phénotypes observés.

(Muramatsu, 1963)

TABLE 3

Dosage effect of the vulgare gene *Q* and the spelta gene *q* in Chinese Spring background

Status of chromosome 5A	2n	Genotype	Dosage of <i>q</i> (<i>Q</i> = 2.5 <i>q</i>)	Phenotype
Nullisomic	40	—* —	0	speltoid
Monosomic	41	<i>Q</i> —	2.5	speltoid
Disomic	42	<i>Q q</i>	3.5	speltoid
Disomic	42	<i>Q Q</i>	5	squarehead
Trisomic	43	<i>Q Q Q</i>	7.5	subcompactoid
Tetrasomic	44	<i>Q Q Q Q</i>	10	compactoid
Mono- <i>spelta</i> -5A	41	<i>q</i> —	1	speltoid, (SEARS unpublished)
<i>spelta</i> -5A	42	<i>q q</i>	2	speltoid
Tri- <i>spelta</i> -5A	43	<i>q q q</i>	3	speltoid
Tetra- <i>spelta</i> -5A	44	<i>q q q q</i>	4	speltoid
Tri-5A(<i>spelta</i>) plus iso-5A(<i>spelta</i>)	44	<i>q q q q q</i>	5	squarehead
Tetra-5A(<i>spelta</i>) plus iso-5A(<i>spelta</i>)	45	<i>q q q q q q</i>	6	transitional type, squarehead-subcompactoid

* — indicates absence of one whole chromosome 5A.

V-2-2 Effet dose de la variation du nombre de copies *Q/q* :

D'après des études menées par Huskins, Sears et Muramatsu sur des lignées nullisomiques, monosomiques, disomiques, trisomiques, et tétrasomiques pour le chromosome 5A, conduisant respectivement à 0, 1, 2, 3 et 4 copies de l'allèle domestiqué *Q* dans le blé hexaploïde, on obtient des phénotypes différents (Huskins, 1946 ; Sears, 1952 ; Muramatsu, 1963). En effet, les formes des épis sont respectivement, speltoïdes, « squarehead », sub-compactes et hyper-compactes (Fig. V-3a, b, c, d et e). Muramatsu (1963) montre que si on augmente le nombre de chromosome 5A portant l'allèle sauvage *q* jusqu'à quatre copies (tétrasomique), on observe que l'épi conserve une forme sauvage. Néanmoins, quand on intègre un cinquième chromosome portant l'allèle sauvage *q* (pentasomique), on observe une bascule vers un phénotype domestiqué (Fig. V-4 et V-5). Cette découverte indiquerait que *q* n'est pas un gène déficient, mais un allèle avec les mêmes effets que *Q* à un degré moindre, leur conclusion est que *q* est hypomorphique à *Q*.

Il est à signaler que suite à l'identification et au clonage du gène *Q*, 40 ans après, la caractérisation et les comparaisons des différents allèles dans certains blés hexaploïdes ont montré des blés de type spelta qui portent l'allèle *Q* ou *q* selon qu'ils soient de type européens ou iraniens (Luo et al., 2000).

V-2-3 Comportement de *Q* chez les les blés tétraploïdes :

De 1928 à 1940, Watkins annonce qu'à l'exception de *T. turgidum* ssp. *carthlicum* (*QQ*), tout les blés tétraploïdes sont carénés et possèdent le gène sauvage *K*, qui plus tard sera renommé *q* par MacKey en 1954. Face à cette contradiction avec les tétraploïdes domestiqués qui sont connus pour être facilement battables. De 1978 à 1986, Muramatsu conduit une expérience avec plusieurs tétraploïdes pour déterminer s'ils portent le gène *Q* ou *q*. Dans la pratique, il a observé le comportement du chromosome 5A de ces tétraploïdes dans le fond génétique de Chinese Spring, via le biais de lignées de substitution. Cela lui a permis de démontrer que *T. turgidum* ssp. *carthlicum* possède *Q*, mais également *T. turgidum* ssp. *polonicum*, ssp. *durum* et ssp. *dicoccum* cv. *liguliforme*. Cette dernière ayant un phénotype compact avec des grains carénés, un rachis fragile. En plus, les expériences confirment que *T. turgidum* ssp. *dicoccum* cv. *farrum* (ammidonier) possède *q*. Il conclut en 1986, qu'il existe des variations du phénotype sauvage dans des lignées possédant *Q*, et que ces variations sont plus visibles en l'absence du gène *Q*.

V-2-4 Comportement du gène *Q* dans les hexaploïdes :

Due à leur récente origine, l'étude des blés hexaploïdes est plus simple par rapport à celle des tétraploïdes. Tout les hexaploïdes qui ont *Q*, n'ont pas forcément des glumes non carénées. Tout comme pour les tétraploïdes on ne peut conclure sur la présence du gène *Q* ou *q* uniquement par le phénotype

observé, car la modification du fond génétique joue également un grand rôle, tout comme le phénotype ‘ ‘ squarehead ’ ’ de l’épi ne concerne pas uniquement les lignées possédant *Q*.

Pour conclure, de 1963 à 1986, Muramatsu explique que la nature pleiotropique des gènes *Q/q* et les différents fonds génétiques sont à l’origine de l’éventail des différents phénotypes observés chez le blé, c’est-à-dire entre le blé *T. aestivum* ssp. *aestivum* (facilement battable 5AQ) et les blés *T. aestivum* ssp. *spelta*., ssp. *macha* et ssp. *vavilovii* (difficilement battable 5Aq).

Muramatsu émet l’hypothèse qu’il doit exister un génotype possédant l’allèle *Q* avec un phénotype speltoïde due au fond génétique. C’est le blé *T.aestivum* ssp. *spelta* européen (Luo et Dvorak en 2000), mais c’est un croisement entre *T. turgidum* ssp. *dicoccum* et *T. aestivum* ssp. *compactum*.

V-2-5 Identification et caractérisation moléculaire du gène *Q* :

Avant d’arriver au clonage positionnel et à la caractérisation moléculaire du gène *Q*, il a fallu un certain nombre d’étapes telles que l’établissement de cartes génétiques denses et d’une carte physique autour du gène sur le chromosome 5A. L’obtention des cartes génétiques faisait appel à deux stratégies, la première était l’utilisation de lignées de délétion, avec la perte de fragments du chromosome 5A autour du gène *Q* (Miller and Reader, 1982 ; Endo and Gill, 1988 ; Tsujimoto and Noda, 1989 ; Ogihara et al., 1994 ; Endo and Gill, 1996), la deuxième stratégie était l’utilisation de lignées de recombinaison (Kojima and Ogihara, 1998 ; Kato et al., 1999).

Une fois un consensus de ces cartes établis, il demeurait un manque de précisions pour le locus *Q*, dû à une densité trop faible de marqueurs et des populations génétiques ségréguantes de tailles insuffisantes.

V-2-5-1 Carte génétique fine :

Les résultats insuffisants des premières cartes consensus obligeaient l’élaboration d’une carte génétique fine de la région comprenant le locus *Q* en combinant d’autres approches moléculaires (RFLP et AFLP).

L’utilisation des techniques RFLP (restriction fragment length polymorphism) et AFLP (Amplification fragment length polymorphism) permettaient d’obtenir 18 marqueurs liés au gène *Q* à partir de deux populations (Fans et al., 2003). La première population était issue du croisement entre Chinese Spring et Chinese Spring-substitué appelé (CS-DIC 5A), la substitution est le remplacement du chromosome 5A (5AQ) portant l’allèle *Q* par le chromosome 5A (5Aq) portant l’allèle sauvage *q* de *T. turgidum* ssp. *Dicoccoides*. La deuxième population était issue du croisement entre Chinese Spring-Cheyenne chromosome 5A disomic substitution (CS-CNN 5A) et CS-DIC 5A (Fans et al., 2003).

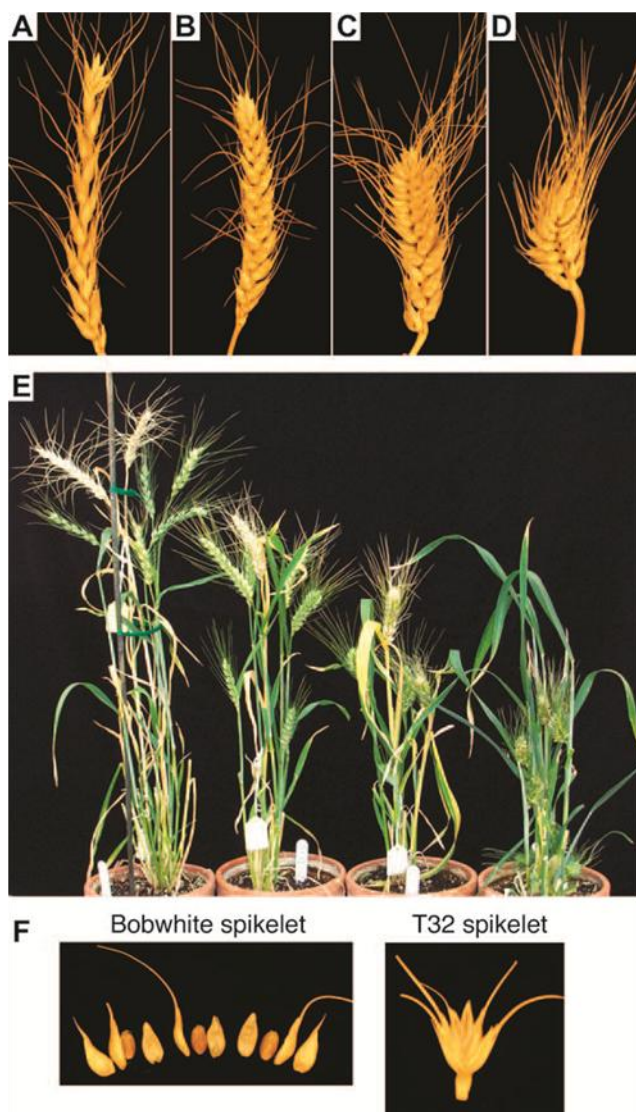


Fig. V-6 Expérience de l'effet dose du gène Q par transgénèse :

Analyse des plantes transgéniques T1.

- A. T1 transgénique a l'épi speltoïde (T32, Q mis sous silence « silenced »).
- B. Bobwhite non-transformé épi « squarehead ».
- C. T1 transgénique épi subcompacté (T30).
- D. T1 transgénique épi compacté (T39).
- E. Vue d'ensemble des plantes transformées. De la gauche vers la droite: T32, Bobwhite, T30, and T39. (F) Différents profils de battage entre un épillet de Bobwhite non-transformé (QQ) (à gauche) et un épillet de plante transgénique T1 au profil speltoïde (T32, Q silenced) (à droite). Bobwhite (Q) possède un rachis solide, une cassure mécanique à la base de l'épillet provoque la libération des grains. Quand l'allèle Q est mis sous silence et dans les géotypes *qq* l'abscission se produit entre le rachis et rachilla libérant l'épillet contenant des grains carénés par un tégument (non-battable). L'analyse par southern sur la descendance T1 montre une à plusieurs copies insérées (Simons et al., 2006).

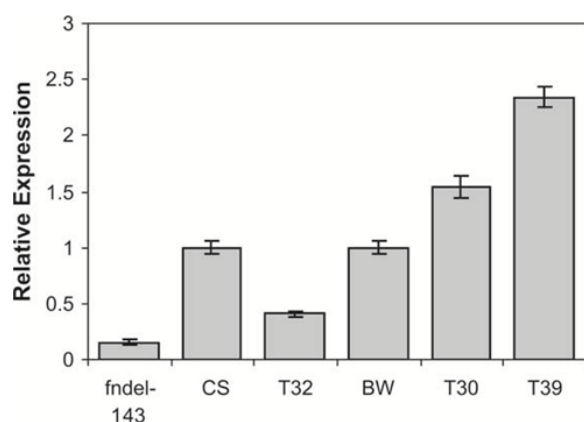


Fig. V-7 Corrélation entre niveau de transcription et phénotype obtenu :

Les différents niveaux d'expression du gène Q mesurés dans les plantes. De gauche à droite : le mutant *fndel-143* (fast-neutron induced deletion of Q), CS (QQ), T32 (plante transgénique mis sous silence), la variété Bobwhite, T30 (plante transgénique, épi subcompacté), and T39 (plante transgénique, épi compacté). Les niveaux d'expression dans CS (Q) et le mutant *fndel-143* (Q absent) sont mis comme témoins. Les barres représentent l'erreur standard (Simons et al., 2006).

V-2-5-2 “Chromosome walking” ou marche chromosomique sur le locus Q :

La marche chromosomique sur le génome de blé était réalisable car il s’agissait d’une région riche en gènes. Le matériel de départ était la banque BAC de *T. monococcum* (Lijavetsky et al., 1999) pour identifier quatre clones BAC pour un “contig” d’une longueur de 300kb portant le gène (Fans et al., 2003). Les marqueurs plus précis obtenus ont été utilisés sur une population de mutants au phénotype speltoïdes obtenue par bombardement aux neutrons, « fast-neutron induced » (Fans et al., 2003).

V-2-5-3 Le clonage et la validation du gène Q :

L’annotation a permis d’identifier un gène de 3,5kb similaire au gène *ids* (*indeterminant spikelet gene*) du maïs (Chuck et al., 1998), mais également au gène *AP2* (*APETALA2*) d’*Arabidopsis* (Jofuku et al., 1994). Ce sont des facteurs de transcriptions agissant au niveau de l’homéostasie florale et possèdent des motifs DNA-binding propre aux plantes appelés domaines AP2. Ces gènes ont un large spectre d’action car ils sont impliqués dans le développement du méristème floral (Irish and Sussex, 1990; Bowman et al., 1993), dans la mise en place des organes floraux (Komaki et al., 1988 ; Bowman et al., 1989 ; Kunst et al., 1989 ; Jofuku et al., 1994) et dans la régulation spatiale et temporelle des gènes impliqués dans l’homeostasie florale (Drews et al., 1991).

A partir des séquences du gène *5Aq* de *T. monococcum*, des PCR de longues tailles et des RT-PCR ont permis de récupérer l’allèle *Q* (Simons et al., 2006). Le gène a été validé par complémentation et par le crible du phénotype speltoïde sur deux populations mutantes (traitement à l’EMS). Puis par l’annotation des séquences après RT-PCR pour trois mutants positifs au crible, confirmant le rôle du gène pour les phénotypes observés.

V-2-5-4 Dosage et effet pleiotropique du gène Q :

Quarante deux années après l’utilisation de lignées de substitutions (chromosomes entiers) pour décrire les effets pleiotropiques du gène *Q* (Muramatsu, 1963) (Fig. V-3,V-4 et V-5). Une expérience similaire mais pas identique est réalisée en transformant un blé domestique *T. aestivum* ssp. *aestivum* var Bobwhite avec l’allèle *Q* de *T. turgidum* ssp. *durum* var Langdon (LDN) (Fig. V-6 et V-7) (Simons et al., 2006). Les plantes transgéniques obtenues (variant en nombre de copies du gène insérées) présentent des différentes morphologies similaires à celles obtenues en variant le nombre de copies *5AQ* (Huskins, 1946; Sears, 1952; Muramatsu, 1963) (Fig. V-6A, 6B, 6C et 6D). Les épis sont de plus en plus compacts suivant l’augmentation du nombre de copies insérées et de l’augmentation du niveau d’expression. L’efficacité des transformations leurs permettent d’aller jusqu’à la mise sous silence du gène qui aboutit à une forme speltoïde. Les plantes transgéniques qui subissent une mise sous silence du gène *Q* présentent des glumes très coriaces, des rachis plus fragiles (Fig. V-6F).

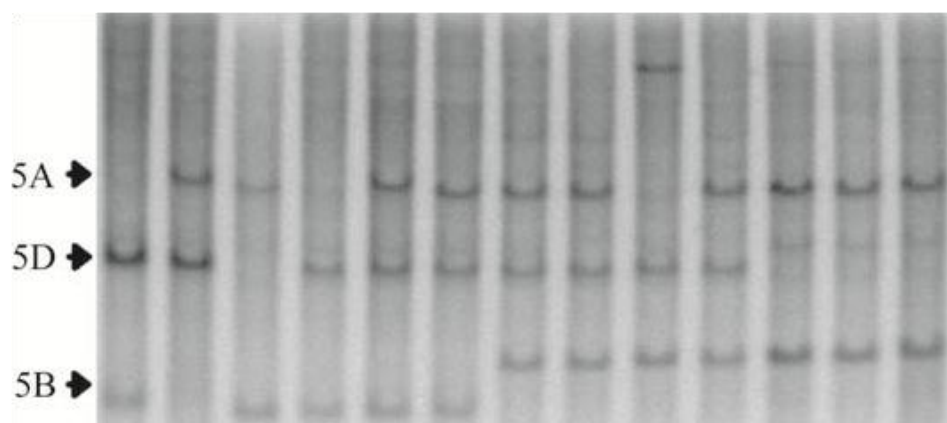


Fig. V-8 Contrôle du nombre d'homéoallèles Q/q par génotypes :

Analyse par Southern de l'hybridation d'un fragment de WAP2 (Q) sur l'ADNg de divers blés. De gauche à droite: N5AT5D, N5BT5D, N5DT5B, fndel-143 (Q-null), *T. aestivum* spp. *aestivum* cv. CS (Q), CS-DIC 5A (q), *T. aestivum* spp. *macha* (q), European *T. aestivum* spp. *spelta* (q; TA2603), Iranian *T. aestivum* spp. *spelta* (Q), European *T. aestivum* spp. *spelta* (q; DS 5A Europe), *T. turgidum* spp. *carthlicum* (Q), *T. turgidum* spp. *durum* cv. LDN (Q), and *T. turgidum* spp. *polonicum* (Q).

(Simons et al., 2006)

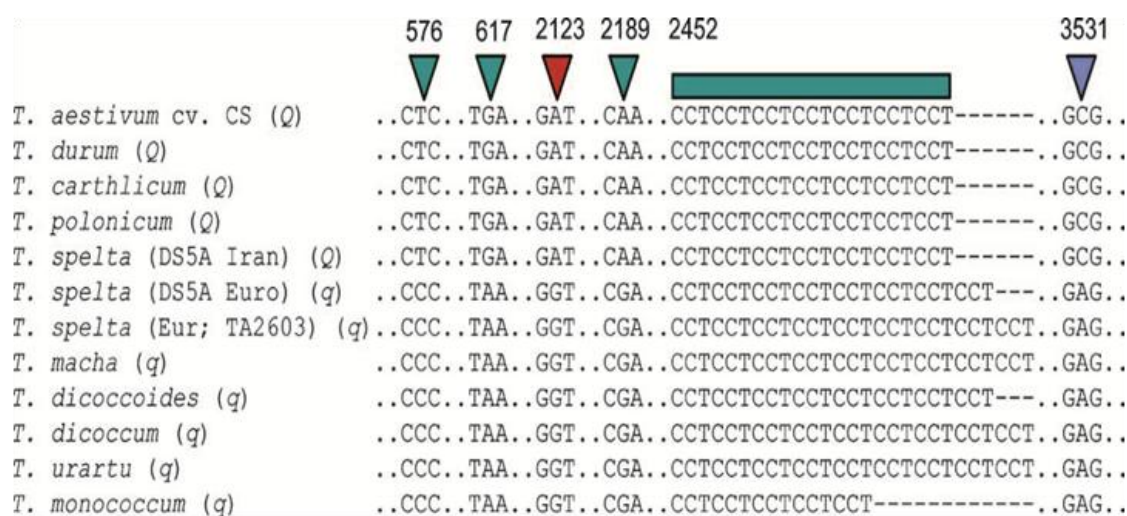


Fig. V-9 Alignements des séquences des gènes Q et q de blés tétraploïdes et hexaploïdes.

Localisation des six mutations nucléotidiques différenciant l'allèle Q de l'allèle q, dans les séquences génomiques de différents génotypes. Les flèches indiquent les SNPs (Single Nucleotide Polymorphisms) entre Q et q. (la couleur verte pour les introns; le rouge pour les exons; le bleu pour le 3'UTR). Le SNP en rouge indique également la différence d'acide aminé à la position V329I entre les protéines prédites pour l'allèle Q et l'allèle q. La barre verte représente la variation d'une séquence microsatellite dans l'intron 9.

(Simons et al., 2006)

V-2-5-5 Analyses structurale et fonctionnelle du gène Q :

Comparaison des allèles Q et q :

Pour départager les deux théories sur la différence entre Q et q, la première étant la duplication de q (Kuckuck, 1959; Swaminathan, 1963) et la deuxième une mutation de q en Q (Maramatsu 1963). Un fragment de q (*T. monococcum*) est hybridé sur le génome de blés hexaploïdes et tétraploïdes. Les intensités d'hybridation confirment qu'il n'y a qu'une copie par chromosome 5A, ce résultat oriente la recherche vers la caractérisation de mutations, le fait qu'uniquement les blés domestiques possèdent l'allèle Q et que les blés sauvages possèdent l'allèle q, donne raison à la théorie de Muramatsu, selon laquelle le gène Q est la forme mutée du gène q. (Fig. V-8) (Simons et al., 2006).

La comparaison de 12 séquences, cinq de l'allèle Q et sept de l'allèle q de blés domestiqués et de blés sauvages a mis en évidence six SNPs (Single nucleotid Polymorphism) dont une est associée au phénotype « squarehead » provoquant un changement de codon, d'acide aminé (Fig. V-9).

Les auteurs sont directement partis sur des expériences doubles hybrides pour valider que le changement d'acide aminé (isoleucine en valine) en position 329, était la raison pour laquelle le gène Q permet d'obtenir un phénotype domestiqué. En effet ce changement d'acide aminé provoque la dimérisation des protéines (homodimères) dans leurs expériences. Les auteurs émettent également deux théories, la première serait que l'homodimère agirait directement sur le promoteur du gène Q, provoquant une sur-expression, la deuxième qu'un miRNA pourrait jouer un rôle dans la régulation de la traduction via une boîte AASSGF, observé pour le gène AP2 chez *Arabidopsis* (Chen, 2004). Ils rejettent cette dernière hypothèse car la boîte AASSGF n'est pas modifiée entre l'allèle q et l'allèle Q.

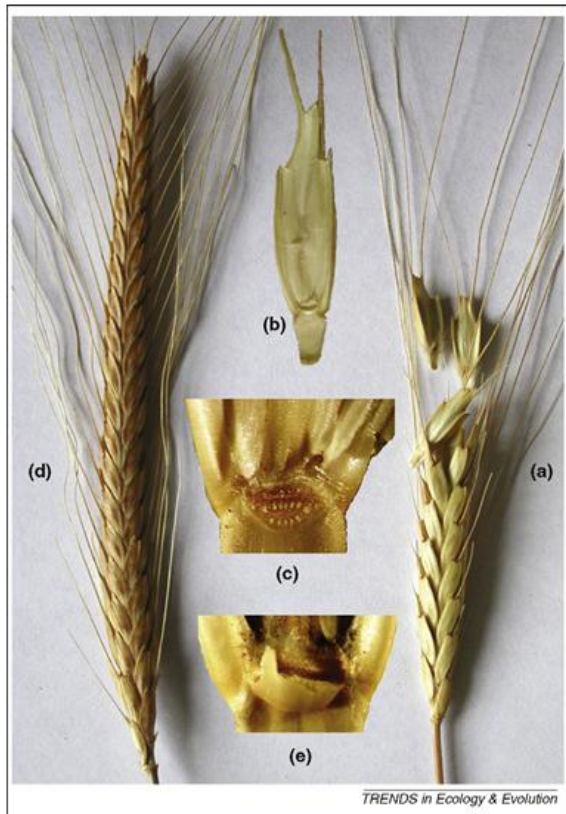


Fig. V-10 Différences phénotypiques entre engrain sauvage (*T. monococcum* ssp. *boeoticum*) et engrain domestiqué (*T. monococcum* ssp. *monococcum*) :

Différences morphologiques entre un épi de blé domestiqué (*T. monococcum* ssp. *monococcum*) et un épi de blé sauvage (*T. monococcum* ssp. *boeoticum*). (a) La déhiscence des épillets (b) d'un épi d'engrain sauvage à maturité, chaque épillet se détache en laissant une cicatrice (c) à la base du rachis. (d) Par contraste les épillets de l'engrain domestiqué ne sont pas déhiscent et restent sur l'épi à maturité. Une action mécanique est nécessaire, ce qui provoque une cassure (e) à la base de chaque épillet.

(Brown et al., 2009)

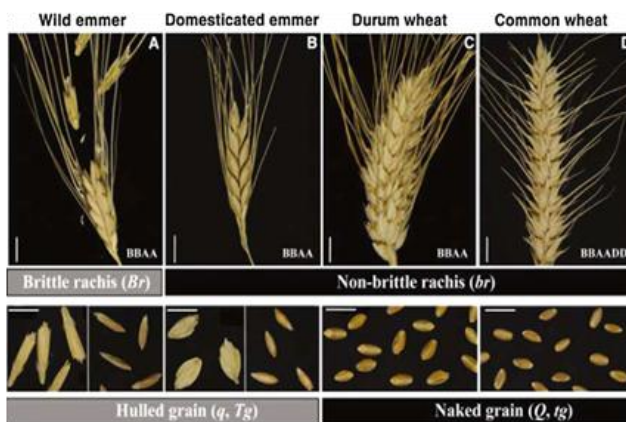


Fig. V-11 Les autres sélectionnés chez les blés :

Rôle des gènes de fragilité du rachis (*Br/br*) et du carénage du grain (*Tg/tg*) dans les blés domestiqués.

(Peng et al., 2011)

Fig. V-12 Effet du gène *c/C* de compaction de l'épi :

La différence de compaction des épis entre *T. aestivum* ssp. *aestivum* (ABD) et *T. aestivum* ssp. *compactum* (ABD) est due au gène *c/C*.

(Filatenko and Kurkiev, 1975).



Actuellement, on sait que les caractères de domestication (non dissémination des grains et un battage facile, etc...) impliquent d'autres gènes (Fig. V-10) :

- ✓ Le gène *Br* (*brittle rachis*) pour la fragilité du rachis, il permet la désarticulation ou abscission à la base de l'épillet des blés lorsque l'épi est mature (déhiscence). Chez les blés domestiqués le gène est muté (*br*), c'est la raison pour laquelle les épillets restent sur l'épi (Zohary and Hopf, 2000 ; Peng et al., 2003 ; Konishi and al., 2006), le battage consiste à casser mécaniquement le rachis (Fig. V-11). Le gène a été localisé sur le bras court du chromosome 3A de l'engrain, *T. monococcum* (Sharma and Waynes, 1980), puis les homéoallèles sur les chromosomes 3AS (*Br2*) et 3BS (*Br3*) de l'amidonnier, *T. turgidum* ssp. *dicoccoïdes* (Cao et al., 1997 ; Watanabe, 2005 ; Nalam et al., 2006 ; Kurarparthy et al., 2007).
- ✓ Le gène *Tg* (*Tough glume*) pour le carénage de la graine (tégument). Il a été identifié sur le bras court du chromosome 2D (2DS) sur une population issue du croisement entre *T. aestivum* ssp. *spelta* (épeautre) et *T. turgidum* ssp. *tibetanum*, blé semi-sauvage tibétain (Kerber and Rowland, 1974). Le blé semi-sauvage tibétain diffère des blés communs pour la ténacité de la glume (Cao et al., 1997 ; Kurarparthy et al., 2007). Pour les blés domestiqués, le gène est muté (*tg*), ce qui permet un détachement facile des glumes. Depuis, les homéologues ont été également identifiés sur les chromosomes 2A et 2B. Il est à noter que le gène *sog* (*soft glume*) de *T. monococcum* (engrain) n'est pas localisé au même endroit sur le chromosome 2AS, ce qui mène à penser qu'il y a deux processus différents en terme évolutif.
- ✓ Le gène *C* (*club*), pour un épi compact (Fig. V-12), a été identifié sur le bras long du chromosome 2D (2DL) lors d'une première étude sur des blés hexaploïdes notamment *T. aestivum* ssp. *compactum*, puis sur *T. monococum* ssp. *sinskajae* (Nilsson-Ehle, 1917 ; Swaminathan, 1966 ; Rao, 1973 ; Mitrofanova, 1997). Chez la plus part des blés tendres étudiés, le gène *C* est muté et porte le nom de gène *c*. Le blé, *T. monococum* ssp. *sinskajae* est un mutant naturel de *T. monoccocum* ssp. *monococcum* (Filatenko and Kurkiev, 1975).

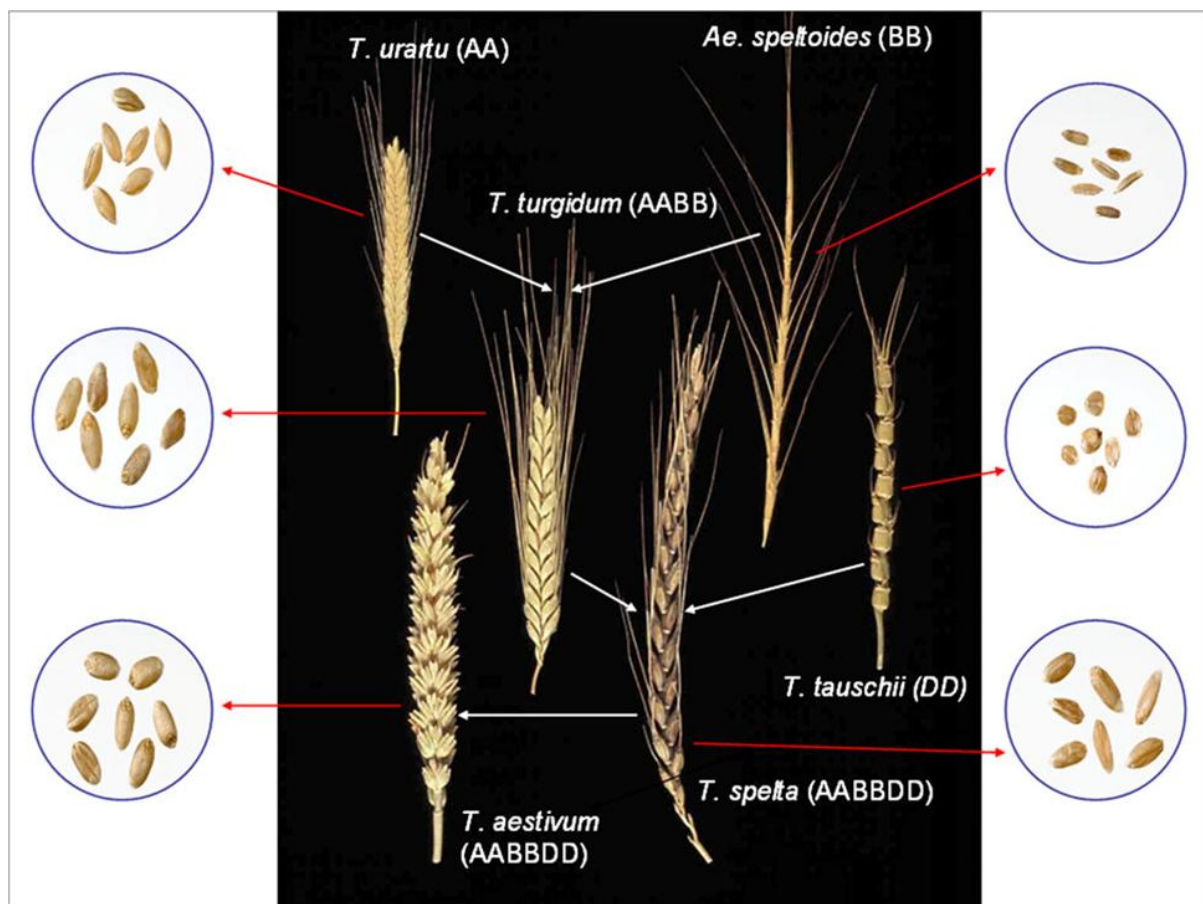


Fig. V-13 Exemples de diversités de phénotypes observés chez les blés.
(Shewry, 2009)

Bilan de l'étude bibliographique :

Le résultat de ces recherches bibliographiques par rapport à l'objectif de ma thèse m'ont permis de comprendre que le blé, un complexe d'espèces (Fig. V-13) ayants différents niveaux de ploïdies (di, tetra, hexa), a toujours intéressé un grand nombre de scientifiques. Parmi les diverses approches possibles, j'ai choisi d'orienter et d'approfondir mes connaissances sur les gènes de domestication.

Maintenant après diverses études nous savons qu'il y a un gène majoritairement responsable de ce passage. Une simple mutation en position 329 de la séquence en acides aminés conduit la plante d'un phénotype non-domestiqué à un phénotype domestiqué suivant le fond génétique.

Nous savons que ce gène est un facteur de transcription, qu'il a un effet pléiotrope, que son action est corrélée à son niveau d'expression (expériences de transgénèse). Or dans les études récentes, on trouve des gènes cibles qui correspondent à un caractère (la fragilité du rachis, la solidité du tégument, la compaction de l'épi) et qui ont été identifiés dans des génotypes sauvages. Sachant que les premières études ont été faites sur des blés domestiqués et que les nouveaux gènes ont été identifiés dans des génotypes sauvages, il est nécessaire de faire un lien entre ces différentes informations.

Malheureusement, la plus part de ces gènes n'ont pas été caractérisés au niveau moléculaire, ou sont en cours d'identification. Dans un avenir proche, je pense que l'on pourra faire ce travail et mieux comprendre les différentes combinaisons et les interactions entre ces différents gènes.

DEUXIEME PARTIE :

ANALYSES ET RESULTATS

CHAPITRE VI:

ROLE ET FONCTIONNEMENT DES DIFFERENTS HOMEOLOGUES DU GENE DE DOMESTICATION Q/q DANS LES BLES POLYPLOIDES

VI-1 Les questions posées :

Le clonage du gène majeur 5AQ (Simons et al., 2006) représente une avancée importante. De nombreuses questions subsistent sur ses effets au niveau du phénotype, sa régulation et ses interactions avec les autres copies homéologues dans les autres sous-génomes des blés polyploïdes:

1. Comment les copies homéologues ont évolué et notamment par rapport aux orthologues et/ou homéologues dans les différents blés diploïdes et polyploïdes?
2. Quels sont les niveaux d'implications des différents homéologues dans les caractères de domestication des blés polyploïdes?
3. Est ce que les différentes copies homéologues interagissent entre elles?

La caractérisation des différents orthologues et homéologues du gène de domestication aux niveaux structural, phénotypique et expressionnelle a été investie en collaboration avec l'équipe de Justin Faris (USDA). Cette caractérisation représente un immense travail de génomique comparée, structurale et fonctionnelle. Les résultats de cette caractérisation ont fait l'objet d'une publication dans le journal *Proceedings of National Academy of Science (PNAS)* (**Titre : Duplication and partitioning in evolution and function of homoeologous Q loci governing domestication characters in polyploids wheat**) que je présente ici avec des résultats complémentaires (**Supporting information**) décrivant le matériel et méthodes, l'évolution de l'espace des TEs, et la caractérisation du premier *Hélitron* actif identifié dans le blé.

Je présente également des résultats et analyses complémentaires obtenus depuis cette publication au niveau de la caractérisation du premier *Hélitron* actif du blé, la régulation de l'expression par le micro ARN 172 (miR172) et les différences d'expression observées entre les deux formes alléliques 5AQ et 5Aq dans le blé hexaploïde.

Duplication and partitioning in evolution and function of homoeologous *Q* loci governing domestication characters in polyploid wheat

Zengcui Zhang^{a,1}, Harry Belcram^{b,1}, Piotr Gornicki^c, Mathieu Charles^b, Jérémy Just^b, Cécile Huneau^b, Ghislaine Magdelenat^d, Arnaud Couloux^d, Sylvie Samain^d, Bikram S. Gill^e, Jack B. Rasmussen^a, Valérie Barbe^d, Justin D. Faris^{f,2}, and Boulos Chalhou^{b,2}

^aDepartment of Plant Pathology, North Dakota State University, Fargo, ND 58102; ^bOrganization and Evolution of Plant Genomes, Unité de Recherche en Génétique Végétale, Unité Mixte de Recherche, Institut National de la Recherche Agronomique 1165, Centre National de la Recherche Scientifique 8114, Université d'Evry Val d'Essonne, 91057 Evry Cedex, France; ^cDepartment of Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL 60637; ^dCommissariat à l'Energie Atomique, Institut de Génétique GENOSCOPE, 91057 Evry Cedex, France; ^eWheat Genetic and Genomic Resources Center, Department of Plant Pathology, Kansas State University, Manhattan, KS 66506; and ^fUS Department of Agriculture, Agricultural Research Service, Cereal Crops Research Unit, Northern Crop Science Laboratory, Fargo, ND 58102

Edited* by Jeffrey L. Bennetzen, University of Georgia, Athens, Georgia, and approved October 5, 2011 (received for review July 7, 2011)

The *Q* gene encodes an AP2-like transcription factor that played an important role in domestication of polyploid wheat. The chromosome 5A *Q* alleles (5AQ and 5Aq) have been well studied, but much less is known about the *q* alleles on wheat homoeologous chromosomes 5B (5Bq) and 5D (5Dq). We investigated the organization, evolution, and function of the *Q/q* homoeoalleles in hexaploid wheat (*Triticum aestivum* L.). *Q/q* gene sequences are highly conserved within and among the A, B, and D genomes of hexaploid wheat, the A and B genomes of tetraploid wheat, and the A, S, and D genomes of the diploid progenitors, but the intergenic regions of the *Q/q* locus are highly divergent among homoeologous genomes. Duplication of the *q* gene 5.8 Mya was likely followed by selective loss of one of the copies from the A genome progenitor and the other copy from the B, D, and S genomes. A recent *V*₃₂₉-to-I mutation in the A lineage is correlated with the *Q* phenotype. The 5Bq homoeoalleles became a pseudogene after allotetraploidization. Expression analysis indicated that the homoeoalleles are coregulated in a complex manner. Combined phenotypic and expression analysis indicated that, whereas 5AQ plays a major role in conferring domestication-related traits, 5Dq contributes directly and 5Bq indirectly to suppression of the speltoid phenotype. The evolution of the *Q/q* loci in polyploid wheat resulted in the hyperfunctionalization of 5AQ, pseudogenization of 5Bq, and subfunctionalization of 5Dq, all contributing to the domestication traits.

fate of duplicated genes | hyperfunctionalization | subfunctionalization

Polyploidy is an important evolutionary feature widespread in the plant kingdom leading to whole-genome duplication (1–3), after which individual genes follow one of many possible evolutionary fates, including nonfunctionalization (deletion or pseudogenization), neofunctionalization, or subfunctionalization (4, 5), often modified by genetic and epigenetic interactions between homoeoalleles (1, 6). Various mechanisms affect regulation of gene expression and the fate of homoeologous genes, including altered or incompatible regulatory interactions, epigenetic modifications, gene dosage changes, partitioning, and compensation (6–11).

Wheat was among the first plant species to be domesticated and was instrumental in spawning the agricultural revolution and the establishment of human civilization. The widespread cultivation of wheat resulted from founder evolutionary events including the acquisition of domestication traits such as a nonbrittle rachis to prevent spikelet shattering and soft glumes and hull-less seed for ease of threshing, as well as allopolyploidization events (12) that resulted in species with better agronomic performance and wide adaptability.

The widely cultivated allohexaploid *Triticum aestivum* L. ($2n = 6x = 42$, AABBDD), known as common or bread wheat, originated as the result of two separate amphiploidization events (Fig. S1) (13–16). The tetraploid wheat *Triticum turgidum* L., ($2n = 4x = 28$, AABB), or pasta wheat, arose <0.5 million years ago as a result of hybridization between *Triticum urartu* Tumanian ex Gandylian ($2n = 2x = 14$, AA) and an unidentified diploid *Aegilops* species of the section *Sitopsis*, thought to be *Aegilops speltoides* Tausch ($2n = 2x = 14$, SS) or a close relative thereof, as the donor of the B genome (12, 13, 17–22). A spontaneous hybridization that occurred between a primitive domesticated tetraploid *T. turgidum* subspecies and the diploid goatgrass *Aegilops tauschii* Coss ($2n = 2x = 14$, DD) about 10,000 years ago gave rise to *T. turgidum* (14, 15).

The wheat *Q* gene, which arose through a spontaneous mutation in allopolyploid wheat, is arguably the most important domestication gene in cultivated wheat because it confers the free-threshing character (Fig. 1) and influences a number of other domestication-related traits such as rachis fragility, glume toughness, spike architecture, flowering time, plant height, and others (23, 24). The *q* allele in wild wheats confers a speltoid spike phenotype characterized by a lax head with elongated rachis and non-free-threshing seed (Fig. 1). The mutation that gave rise to *Q* had a profound effect on agriculture because it allowed early farmers to efficiently harvest their grain on a grand scale, one of the features that led to rapid spread of polyploid wheat cultivation around the world.

The *Q* gene, encoding a member of the AP2 family of transcription factors (23), resides on the long arm of wheat chromosome 5A (5AQ) (24, 25). Homoeologous *q* loci reside on chromosome arms 5BL (5Bq) and 5DL (5Dq). The *Q* allele mutation resulted in a single-amino-acid substitution changing properties of the transcription factor (23), probably affecting expression of multiple genes, thus explaining its pleiotropic nature. The free-threshing phenotype, however, is only expressed in polyploid wheat (26), indicating that genome composition, cross-talk be-

Author contributions: J.D.F. and B.C. designed research; Z.Z., H.B., P.G., M.C., C.H., G.M., A.C., S.S., J.B.R., V.B., J.D.F., and B.C. performed research; Z.Z., H.B., P.G., M.C., J.J., G.M., A.C., S.S., V.B., J.D.F., and B.C. analyzed data; and Z.Z., H.B., P.G., M.C., J.J., B.S.G., J.B.R., J.D.F., and B.C. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. JF01612–JF01622.)

¹Z.Z. and H.B. contributed equally to this work.

²To whom correspondence may be addressed. E-mail: chalhou@evry.inra.fr or justin.faris@ars.usda.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1110552108/-DCSupplemental.

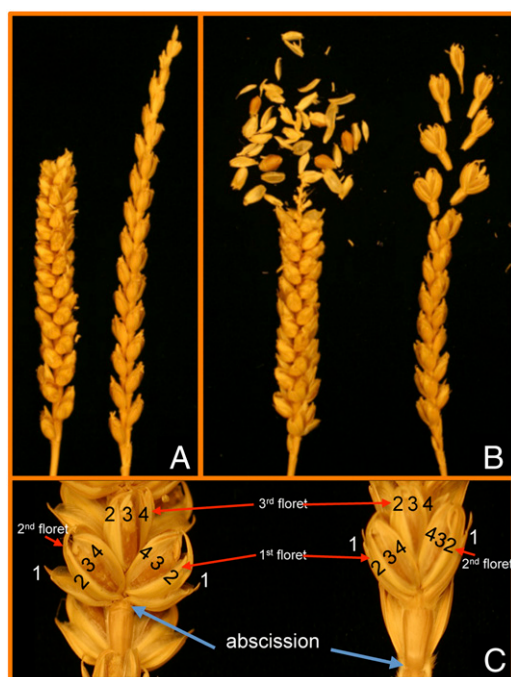


Fig. 1. Comparison of effects of 5AQ on morphological and domestication-related traits of hexaploid wheat cv. CS spikes. (A) Presence of 5AQ confers a square-shaped spike (Left), and the absence of 5AQ (in the deletion mutant CS del143) (Right) results in a speltoid spike characterized by a much longer and spear-shaped spike compared with the spike of CS primarily due to the elongation of rachis segments. (B) Seed of CS with 5AQ (Left) is free threshing because the seed is loosely held and abscission occurs at the base of the glumes, leaving the rachis intact. CS lacking the 5AQ gene (Right) is not free threshing because the glumes hold the seed more tightly and abscission occurs within the rachis, causing disarticulation of the entire spikelet with a portion of the rachis left attached to the base of the spikelet. (C) Detailed illustration of a single spikelet of CS when 5AQ is present (Left) and absent (Right). Both spikelets have three fertile florets (first, second, and third) and each spikelet has a pair of glumes (1). Each floret has a lemma (2), seed (3), and palea (4). Note the looseness of the seeds due to the difference in the position of the glumes in 5AQ compared with the mutant lacking 5AQ. The abscission zones are indicated by the blue arrows (the lower spikelet in the 5AQ mutant was removed so that the abscission zone could be observed).

tween homoeoalleles, or other changes caused by polyploidization, are essential as well.

Here, we evaluate the structure, evolution, and function of the homoeologous *Q/q* loci to establish a molecular basis for their contribution to the domestication phenotype in the genetic background of diploid and allopolyploid wheat.

Results

Organization of the *Q/q* Gene Locus in Different Wheat Genomes and Species. The sequences of 11 homologous and homoeologous genomic regions spanning the *Q/q* loci from five wheat species (A, B, S, and D genomes and three ploidy levels) were compared: the A-genome locus from the diploid *T. urartu* (URA) accession TMU138, the tetraploid *T. turgidum* cv. Langdon (LDN) and the hexaploid *T. turgidum* cvs. Renan (RE) and Chinese Spring (CS); the B-genome locus from LDN, RE, and CS; the S-genome locus from the diploid *Ae. speltooides* (SPE) accession SPELT1; and the D-genome locus from the diploid *Ae. tauschii* (TAU) accession AL8/78, as well as from RE and CS (Fig. 2 and Fig. S2).

Q/q was the only gene identified in the sequenced fragments of the B- and S-genome loci. A 23S ribosomal protein-like gene was identified 40 kb downstream of *5Dq* in the D genomes and 100 kb

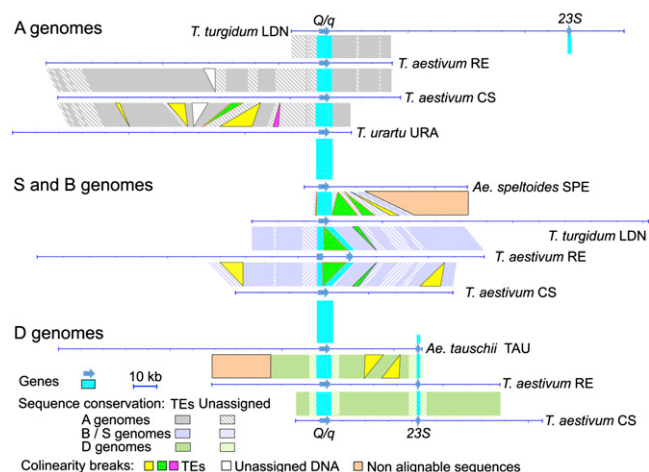


Fig. 2. Schematic presentation of the 11 sequenced homologous and homoeologous genomic regions, spanning the *Q/q* loci from five wheat species. Line diagrams illustrate sequenced DNA segments. Color-coded block diagrams illustrate results of pairwise sequence comparisons. Additional information is provided in [Fig. S2](#).

downstream of *5Aq* in the A genome of LDN. This gene is located 55 kb from the *5Aq* gene in *Triticum monococcum* and 9 kb from the *q* ortholog in rice chromosome 3 (27).

Eighty percent of the sequenced genomic regions consist of nested transposable elements (TEs) interspersed with stretches of unassigned sequences (Fig. 2 and Fig. S2). TEs are different when comparing homoeologous loci, i.e., between A, B/S, and D loci, highlighting their dynamic nature. Most TEs are conserved among A, B, and D homologous loci where 19 large genome-specific indels and breaks in genome colinearity due to recent events such as transposon insertions were revealed (Fig. 2 and Fig. S2). TE annotation and insertion dates estimated for 52 retrotransposons with complete long-terminal repeats (LTRs) are presented in Fig. S2 and Table S1.

All *Q/q* homoeoalleles Are Transcriptionally Active, but *5Bq* Is a Pseudogene That Does Not Encode a Full-Length *q* Protein. The in-

tron/exon structure of the *5AQ* gene (Fig. 3) was previously established (23). Sequences of the *5AQ* coding regions of LDN and CS were identical and differed from the RE sequence by only two substitutions in introns. The coding sequence of URA *5aq* differed from *5AQ* coding sequence of CS, RE, and LDN by five substitutions including two nonsynonymous substitutions, one being the V329I mutation (numbering as in *5AQ*) proposed to be responsible for the phenotypic differences between *5AQ* and *5aq* alleles (23).

For *5Bq*, two transcript variants, *5Bq.1* and *5Bq.2*, were identified for CS (Fig. 3). *5Bq.1* contains only six exons, corresponding to exons 1 and 5–10 of *5AQ*. The first exon ends 45 bp upstream and then continues from 79 bp upstream of the corresponding 3'- and 5'-splice sites of exons 1 and 5 of *5AQ*, respectively, due to a 531-bp deletion including sequences corresponding to exons 2–4 as well as parts of exon 1 and intron 4 of *5AQ*. *5Bq.2* has the same structure as *5Bq.1*, except that the sequence corresponding to intron 7 of *5AQ* is not spliced out. *5Bq.1* and *5Bq.2* encode truncated 354-bp ORFs with a stop codon located at a position equivalent to nucleotides 63–61 upstream of the corresponding 5'-splice site of *5AQ* exon 5.

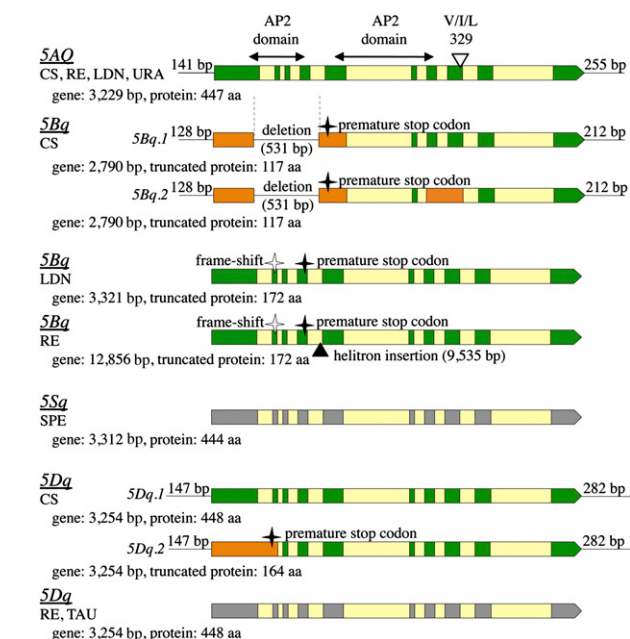


Fig. 3. An overview of the intron/exon structure of the *Q/q* homoeoalleles determined by RT-PCR cloning and sequencing. Exons conforming to the splicing pattern found for the 5AQ gene are shown in green (deduced from cDNA sequences) and gray (predicted), alternative exons are shown in orange, and introns are shown in yellow. 5'- and 3'-UTR lengths (L) by RACE are shown for CS mRNAs. Gene sizes shown are from the translation start to the stop codon. Exons encoding the AP2 domains and the position of the codon for amino acid residue 329 in 5AQ (I), and the corresponding codons in 5AQ (V) and 5Dq (L) are also shown.

letion is present in hexaploid wheats and in most domesticated tetraploids, but it is much less prevalent among wild tetraploids (*T. turgidum* ssp. *dicoccoides*), suggesting that it occurred after tetraploidization and was perhaps selected upon domestication. The 531-bp deletion was only found in some hexaploid wheat lines, suggesting that it occurred after allohexaploidization.

For LDN, one *5Bq* transcript type with the 5AQ splicing pattern was confirmed (Fig. 3). For RE, five different transcript types were detected by RT-PCR cloning and sequencing. These transcripts include two to four additional exons spliced out of a 9,535-bp helitron inserted two bp upstream of the 5'-splice site of the predicted exon 5 of the *5Bq* gene (Fig. 3). The entire helitron sequence was presumably transcribed. Some *5Bq* transcripts from RE lacked the canonical exon 5 or used an alternative 3' splicing site for intron 4. One transcript lacked the canonical exon 9. The RE sequence differed from CS by one synonymous substitution and they both differed from LDN by two more synonymous and one nonsynonymous substitutions, if the 5AQ splicing pattern was assumed and the frameshift-causing deletions were ignored.

5Bq of LDN, CS, and RE are transcriptionally active, but do not encode full-length functional *q* proteins. Therefore, *5Bq* is a pseudogene in all three species.

The 9,536-bp helitron transposon in RE is the first complete helitron identified in wheat (more information is provided in Fig. S3). The helitron insertion was traced in the RE pedigree to Mironovskaya 808, a *T. turgidum* winter wheat cultivar of Russian origin. This insertion has not been found in any other hexaploid nor in any wild or domesticated tetraploid wheat line (Table S2), suggesting that it occurred only very recently.

Two transcript variants were identified for the CS *5Dq* gene. *5Dq.1* results from splicing of 10 exons matching the splicing pattern of 5AQ, but for transcript *5Dq.2*, retention of the first intron leads to a premature stop codon and a 495-bp truncated

ORF (Fig. 3). The coding parts of *5Dq* of CS and REN are identical and differ from the TAU sequence by only one synonymous substitution and three intron substitutions. *5Sg* in SPE encodes a full-length *q* protein, on the basis of the splicing pattern of 5AQ (Fig. 3).

A *q* Gene Duplication in the Wheat Ancestral Genome Was Probably Followed by Differential Loss of One Copy. Full-length *Triticeae Q/q* coding sequences were used in the phylogenetic analysis (Fig. 4). For the *5Bq* pseudogenes, the frameshifts caused by the deletions were adjusted to generate coding sequences corresponding to full-length ORFs found in other *Q/q* genes, even though such mRNAs and corresponding proteins are not made. The null hypothesis of equal evolutionary rate (third codon positions) throughout the tree was not rejected under any of several models tested. The molecular clock was calibrated using 11.6 million years ago (Mya) for the divergence time between barley and wheat (12).

It has been shown previously that 5AQ and 5Aq alleles and corresponding phenotypes are correlated with a V329I mutation (23). This mutation occurred very recently and is present in the A genome of only some wheats (Fig. 4). V is the ancestral state at residue 329 and it mutated to L in the B/D/S *q* lineage.

The 2.6 ± 0.8 Mya divergence time between the D and B/S genomes (S genome appears to be more closely related to B than to D genome) (Fig. 4) is consistent with our previous estimates (12, 13, 28). The divergence time of *5Aq* of *T. monococcum* from 5AQ/*q* of *T. urartu* and polyploid wheats was estimated at 2.6 ± 0.7 Mya, an earlier date than similar estimates for the *Acc-1* gene (13). In the absence of any evidence of an accelerated substitution rate, an ancestral *q* gene duplication some 5.8 Mya (Fig. 4) followed by a selective loss of one of the copies in the A-genome progenitor and the other one in the progenitor of the B, D, and S genomes is the

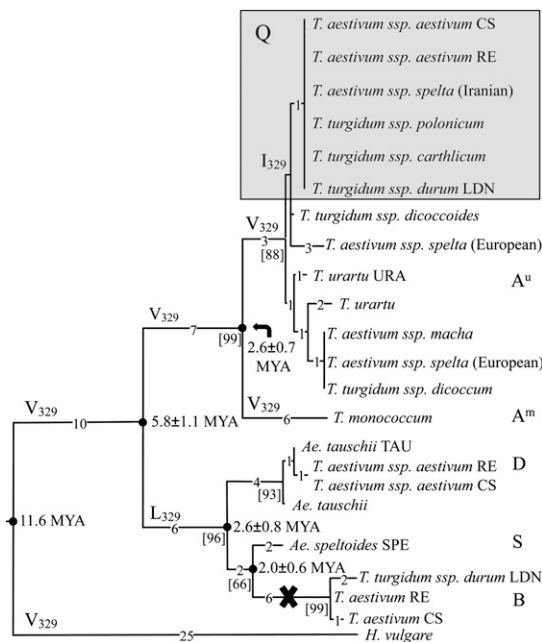


Fig. 4. Phylogenetic relationships and divergence times of the *Q/q* genes. The phylogenetic tree for synonymous substitutions was constructed by the neighbor-joining method. Branch lengths are shown by the number of substitutions. Bootstrap values (%) are shown in square brackets. The molecular clock was calibrated using 11.6 Mya for the time of divergence between barley and wheat (12). Species with the *Q* allele carrying I₃₂₉ are shown on a gray background. Genomes are indicated by letters. The cross indicates that *5Bq* is a pseudogene. Numbering of amino acid residues is as in 5AQ.

Role of *Q/q* in Controlling Morphology and Domestication Traits of Hexaploid Wheat. Correlation between *Q/q* loci and various traits revealed pleiotropic effects of *5AQ* influencing glume toughness, threshability, rachis fragility, spike length and shape, spike emergence time, and plant height (Fig. 6 and Table S3). Deletion of *5AQ* (CS-del143) or replacement with *5Aq* (CS-DIC 5A) had similar effects on all traits compared with CS. CS has soft, papery glumes and free-threshing seed, whereas CS-DIC 5A and CS-del143 have tougher, thicker, and more heavily keeled glumes tightly enveloping the seed that was not free threshing. CS-DIC 5A and CS-del143 plants are taller and have longer, lax, and tapering (speltoid) spikes compared with CS (Fig. 6 and Table S3). Spike emergence is 7 d earlier in CS-del143 and 12 d earlier in CS-DIC 5A than CS, which is consistent with the previously identified role of *5AQ* in delaying heading time (23). CS-DIC 5A has more spikelets and seeds per spike than CS-del143 and CS, but grain yield per plant is similar for all three due to more tillers in CS and CS-del143.

The spikes of CS 5BL-14 (*5Bq* deletion) and CS 5DL-5 (*5Dq* deletion) showed slightly more speltoid phenotypes compared with CS, primarily due to reduced spikelet density in the lower portions of the spikes (Fig. 6). The shape of the upper portion of the spikes of CS 5BL-14 and CS 5DL-5 was intermediate between the speltoid spike shape of CS-del143 and the square-headed spike shape of CS. Glumes of CS 5BL-14 and CS 5DL-5 were tougher and more keeled compared with those of CS, but they were not as tough as the CS-del143 glumes. Disarticulation in CS 5BL-14 and CS 5DL-5 occurred at the base of the glumes, as it does in CS. Compared with CS, the number of spikelets per spike and seed weight were reduced in CS 5BL-14 and CS 5DL-5 (Fig. 6 and Table S3). The number of seeds per spike was reduced by 67–75% due to low fertility. All these changes led to substantially reduced grain yield. These observations suggest that *5Dq* and *5Bq* contribute to threshability and the suppression of the speltoid characters, but to a lesser degree than *5AQ*.

Compared with the parental lines CS-del143 and CS 5DL-5, the spikes of the double deletion line CS-5A/5Ddd, which produces no functional *Q* or *q* protein, were much more speltoid in appearance (Fig. 6) with extremely tough glumes suggesting that the genetic effects of *5AQ* and *5Dq* are additive. The double deletion line had fewer spikelets per spike and lower fertility than its parental lines (Table S3), which resulted in the lowest grain yield among all of the stocks tested. Spikes of CS 5DL-5 and CS-5A/5Ddd emerged more than 30 d later than spikes of CS, most likely caused by the absence of gene(s) other than *q* in the terminal deletion of chromosome 5D.

Discussion

It has long been understood that the *Q* allele on wheat chromosome 5A has had a profound influence on wheat domestication. Here, we characterize the structure, evolution, and function of the *Q/q* loci on chromosomes 5A, 5B (and 5S), and 5D, to understand their contribution to domestication traits.

First, the intergenic sequences flanking the *Q/q* loci, composed primarily of TEs, have profoundly diverged by nested insertion and deletion among the A, B, D, and S genomes. This is a common feature of the intergenic regions of the wheat genome (12, 28–31). This important comparative study showed thus no putative regulating conserved sequence, except the ~2 and ~1 kb at 5'- and 3' ends of the *Q/q* gene, respectively (Fig. 2).

Second, *5AQ/q* is a cryptoparalog of the *q* gene on B, S, and D genomes, a result of gene duplication in the progenitor of the diploid wheats followed by selective loss of one copy of the gene. Functional interactions of paralogs may be substantially different from interactions of less divergent homoeologs.

Third, *5AQ/q*, *5Dq*, and *5Sq* are expected to produce functional *Q/q* proteins, but *5Bq* is a pseudogene encoding truncated proteins without the AP2 DNA-binding domains. It is unknown

whether these truncated proteins serve a function. Both *5Bq* and *5Dq* produce multiple transcript variants. One of the *5Dq* transcripts encodes a full-length *q* protein, whereas the other transcript encodes a truncated protein without the AP2 domains. About 20% of plant genes produce multiple transcripts, but their functions often remain unknown (32).

Fourth, the *5Bq* pseudogene remains transcriptionally active. Plant pseudogenes derived from recently duplicated genes often remain transcriptionally active, presumably due to insufficient time for promoter degradation (33). We showed however that the *5Bq* pseudogene participates in regulation of expression of the *5AQ* and *5Dq* genes at the transcript level, suggesting an active interaction between the homoeoalleles (see below).

Fifth, the *5Bq* pseudogenization occurred after allotetraploidization, possibly upon selection imposed by domestication. This suggests that *5Bq* plays a role in regulating expression of the domestication traits in tetraploid and hexaploid wheat.

Given the current findings, we propose that expression of the free-threshing phenotype and other domestication traits associated with the *Q* allele requires not only the V329I amino acid substitution in the *Q* protein, but also a combination of the divergent *Q/q* loci, with features we have identified, brought together in polyploid wheat. This also provides an explanation for the specificity of the free-threshing character to polyploid wheat and its absence among the diploids (26).

To investigate expression regulation and interactions between different *Q/q* homoeoalleles and their roles in domestication of hexaploid wheat, we analyzed different wheat lines altered for the *Q/q* homoeoalleles. Whereas the deletion of *5AQ* includes a small genomic segment of ~2 Mb (34), deletions of *5Bq* and *5Dq* encompass genomic segments representing 25% and 24% of chromosomes 5B and 5D, respectively (35), suggesting that deletion of other genes could also influence *Q/q* expression or the morphological traits that we evaluated. However, loci affecting these traits have not been reported on the deleted segments of chromosomes 5B or 5D. Good evidence of the appropriateness of the genetic material to study interactions between the different *Q/q* homoeoalleles was provided by the observation that the transcript levels and phenotypes for lines lacking *5Bq* and *5Dq*, which both down-regulate expression of *5AQ*, were nearly the same.

Our transcriptional expression analysis indicated complex interactions among the *Q/q* homoeoalleles. Interpretation was complicated by the fact that *5Bq* is a transcriptionally active pseudogene (no *q* protein is made) and *5AQ/q* is a cryptoparalog of *5Bq* and *5Dq*. Clearly, deletion of *5Dq* or *5Bq* leading to reduced expression of *5AQ* suggests that they indirectly up-regulate expression of *5AQ*, whereas the latter plays a repressor role for both *5Bq* and *5Dq* expression. Our results, consistent with previous research (23), demonstrate that *5AQ* is transcribed at higher levels than *5Aq*. The V329I substitution was shown to be responsible for enhanced homodimer formation in *5AQ*, possibly leading to self up-regulation of *5AQ* transcription (23). Our results show that the expression of *5Bq* and *5Dq* increases when *5AQ* is deleted or replaced by *5Aq*, suggesting that the V329I mutation in *5Aq/Q* is also responsible for the down-regulation of *5Bq* and *5Dq* (comparison of expression in lines CS, CS-del143, and CS-DIC 5A). This result could be explained by dosage balance or complementary effects in regulation among the *Q/q* homoeoalleles (36), but more work is needed to decipher the mechanism of this transcriptional regulation.

It has been recently proposed that interactions between mRNAs, transcribed pseudogenes, and long noncoding RNAs form a large-scale regulatory network across the transcriptome using microRNA response elements (11). A putative miRNA172 binding site in exon 10 of the *Q/q* gene (23) points to a possible role of miRNA in regulation of *Q/q* gene expression. The miRNA172 binding site is also present in the *5Bq* pseudogene transcripts.

Interesting correlations were observed between expression variation of *Q/q* homoeoalleles and variations of domestication and agronomic traits. Our study confirms previous findings that the *5AQ/q* genes influence numerous domestication traits (23, 25, 34, 37, 38), and it shows that along with the *5AQ/q* alleles, the *5Dq* and *5Bq* alleles contribute to speltoid suppression, but their function depends on the allelic state of *5AQ/q*. The contribution of *5Bq* to these traits is indirect because it is a pseudogene, but we demonstrated that its deletion significantly reduced transcription of *5Dq*. This apparent coregulation of the *q* genes largely explains the similar phenotype observed when *5Dq* (CS 5DL-5) or *5Bq* (CS 5BL-14) are deleted. The phenotypic differences correlated with expression variations observed between CS-del143 and CS-5A/5Ddd suggest that *5Dq* plays a more important role in controlling spike shape, spikelet number, and glume toughness when *5AQ* is absent.

Subfunctionalization (evolution of partitioned ancestral functions among alleles or homoeoalleles) and neofunctionalization (evolution of novel functions among alleles or homoeoalleles) were suggested as mechanisms leading to the preservation of the duplicated genes after polyploidization (5). Here, we provide a rare example of the fate of a set of homoeologous genes, associated with plant morphology and domestication in hexaploid

wheat. A recent mutation in *5Aq* led to “hyperfunctionalization” of *5AQ*; *5Bq* underwent pseudogenization, but remains transcriptionally active and influences regulation of expression of the homoeologs; and *5Dq* has been subfunctionalized because it contributes to many of the domestication traits.

Materials and Methods

A full description of the materials and methods used in this work is provided in *SI Materials and Methods*. This information includes BAC library screening, sequencing, and annotation; cDNA amplification and sequencing; transcript and expression analysis; phylogenetic analysis; phenotypic analysis; sequences of primer (Table S4); and descriptions of the wheat genotypes used in this research.

ACKNOWLEDGMENTS. We thank Joseph Jahier (Institut National de la Recherche Agronomique) and Moshe Feldman (Weizmann Institute of Science) for valuable discussions and for providing some of the wheat samples. This project was supported by US Department of Agriculture (USDA), Agricultural Research Service, Current Research Information Systems Project 5442-22000-033-00D and the National Center for Sequencing (Centre National de Séquençage-Génoscope)/APCNS2003 Project: *Triticum* Species Comparative Genome Sequencing in Wheat. PCR-based tracing of retrotransposon insertions was funded by the Agence Nationale pour la Recherche (ANR) Biodiversité Project (ANR-05-BDIV-015) and the ANR-05-Blanc Project impact of transposable elements on gene regulation.

- Adams KL, Wendel JF (2005) Novel patterns of gene expression in polyploid plants. *Trends Genet* 21:539–543.
- Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. *Annu Rev Plant Biol* 60:561–588.
- Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K (2009) The flowering world: A tale of duplications. *Trends Plant Sci* 14:680–688.
- Prince VE, Pickett FB (2002) Splitting pairs: The diverging fates of duplicated genes. *Nat Rev Genet* 3:827–837.
- Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154:459–473.
- Comai L (2005) The advantages and disadvantages of being polyploid. *Nat Rev Genet* 6:836–846.
- Chen ZJ, Ni Z (2006) Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *Bioessays* 28:240–252.
- Birchler JA, Auger DL, Riddle NC (2003) In search of the molecular basis of heterosis. *Plant Cell* 15:2236–2239.
- Chaudhary B, et al. (2009) Reciprocal silencing, transcriptional bias and functional divergence of homoeologs in polyploid cotton (gossypium). *Genetics* 182:503–517.
- Hovav R, et al. (2008) Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proc Natl Acad Sci USA* 105: 6191–6195.
- Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP (2011) A ceRNA hypothesis: The Rosetta Stone of a hidden RNA language? *Cell* 146:353–358.
- Chalupska D, et al. (2008) Acc homoeoloci and the evolution of wheat genomes. *Proc Natl Acad Sci USA* 105:9691–9696.
- Huang S, et al. (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc Natl Acad Sci USA* 99:8133–8138.
- Kihara H (1944) Discovery of the DD-analyser, one of the ancestors of vulgare wheats. *Ag. Hort. (Tokyo)* 19:889–890.
- McFadden ES, Sears ER (1946) The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *J Hered* 37:81–89, 107.
- Nesbitt M, Samuel D (1996) From the staple crop to extinction? The archaeology and history of hulled wheats. *Hulled Wheats. Proceedings of the First International Workshop on Hulled Wheats*, eds Padulosi S, Hammer K, Heller J (International Plant Genetic Resources Institute, Rome), pp 41–100.
- Blake NK, Lehfelddt BR, Lavin M, Talbert LE (1999) Phylogenetic reconstruction based on low copy DNA sequence data in an allopolyploid: The B genome of wheat. *Genome* 42:351–360.
- Dvorák J, Terlizzi P, Zhang H-B, Resta P (1993) The evolution of polyploid wheats: Identification of the A genome donor species. *Genome* 36:21–31.
- Dvorák J, Zhang HB (1990) Variation in repeated nucleotide sequences sheds light on the phylogeny of the wheat B and G genomes. *Proc Natl Acad Sci USA* 87:9640–9644.
- Riley R, Unrau J, Chapman V (1958) Evidence on the origin of the B genome of wheat. *J Hered* 49:91–98.
- Takumi S, Nasuda S, Liu YG, Tsunewaki K (1993) Wheat phylogeny determined by RFLP analysis of nuclear DNA. I. Einkorn wheat. *Jpn J Genet* 68:73–79.
- Talbert LE, Blake NK, Storlie EW, Lavin M (1995) Variability in wheat based on low-copy DNA sequence comparisons. *Genome* 38:951–957.
- Simons KJ, et al. (2006) Molecular characterization of the major wheat domestication gene *Q*. *Genetics* 172:547–555.
- Faris JD, Simons KJ, Zhang Z, Gill BS (2005) The wheat super domestication gene *Q*. *Frontiers of wheat Bioscience. Wheat Information Service* 100(Memorial Issue): 129–148.
- Faris JD, Gill BS (2002) Genomic targeting and high-resolution mapping of the domestication gene *Q* in wheat. *Genome* 45:706–718.
- Sharma HC, Waines JG (1982) Attempted gene transfer from tetraploids to diploids in *Triticum*. *Can J Genet Cytol* 23:639–645.
- Faris JD, Zhang Z, Fellers JP, Gill BS (2008) Micro-colinearity between rice, Brachypodium, and *Triticum monococcum* at the wheat domestication locus *Q*. *Funct Integr Genomics* 8:149–164.
- Salse J, 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* 9:555.
- Chantret N, et al. (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *Plant Cell* 17:1033–1045.
- Dvorak J, Akhunov ED, Deal KR, Luo MC (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* 23:1386–1396.
- Isidore E, Scherrer B, Chalhoub B, Feuillet C, Keller B (2005) Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels. *Genome Res* 15:526–536.
- Barbazuk WB, Fu Y, McGinnis KM (2008) Genome-wide analyses of alternative splicing in plants: opportunities and challenges. *Genome Res* 18:1381–1392.
- Zou C, et al. (2009) Evolutionary and expression signatures of pseudogenes in *Arabidopsis* and rice. *Plant Physiol* 151:3–15.
- Faris JD, Fellers JP, Brooks SA, Gill BS (2003) A bacterial artificial chromosome contig spanning the major domestication locus *Q* in wheat and identification of a candidate gene. *Genetics* 164:311–321.
- Endo TR, Gill BS (1996) The deletion stocks of common wheat. *J Hered* 87:295–307.
- Birchler JA, Riddle NC, Auger DL, Veitia RA (2005) Dosage balance in gene regulation: Biological implications. *Trends Genet* 21:219–226.
- Kato K, Miura H, Sawada S (1999) QTL mapping of genes controlling ear emergence time and plant height on chromosome 5A of wheat. *Theor Appl Genet* 98:472–477.
- Muramatsu M (1963) Dosage effect of the spelta gene *Q* of hexaploid wheat. *Genetics* 48:469–482.

Supporting Information

Zhang et al. 10.1073/pnas.1110552108

Supporting Information Corrected December 29, 2011

SI Materials and Methods

BAC Library Screening and Sequencing of the *Q/q* Locus. *Q/q*-containing BACs from the A and B genome of *Triticum turgidum* cv. LDN (1), and from *Aegilops tauschii* accession AL8/78 (2) were isolated by Southern hybridization using probe Tmap2 as previously described (3). One *T. turgidum* BAC of each type was sequenced by Myriad Genetics, and one *Ae. tauschii* BAC was sequenced at the Washington University Genome Sequencing Center. All other BACs were isolated using the PCR method with *Q/q*-specific primers (Table S4). *Q/q*-containing BACs from A, B, and D genomes of *Triticum aestivum* cvs. RE and CS were isolated from libraries described previously (4, 5). *Triticum urartu* (URA) and *Aegilops speltoides* (SPE) BACs were isolated from pooled libraries as described previously (6, 7). Shotgun sequencing of RE, CS, URA, and SPE BACs was performed at the Centre National de Sequençage as previously described (8).

Identification of Transposable Elements (TEs). TE annotation and prediction were essentially done as previously described (9). TEs were identified by BLASTn searches against two databases of repetitive elements: TREP (<http://wheat.pw.usda.gov/ITMI/Repeats/index.shtml>) (10) and Repbase (http://www.girinst.org/Repbase_Update.html) (11). Core domains (nucleic acid coordinates of known elements) were identified through BLASTn alignments against TREPn. Long-terminal repeats (LTRs) and limits were identified using BLASTn and CENSOR (12) alignments against Repbase and TREP databases. Putative polyproteins were identified by BLASTx alignments against TREPprot. No a priori cutoff was imposed for BLASTx and BLASTn. A complete reconstruction of nested insertions of TEs and split elements was conducted. TE prediction and classification followed the 80-80-80 rule recommended by the unified classification system for eukaryotic TEs: a query element belongs to the same family as a subject element referenced in Repbase or TREP if it is longer than 80 bp, and its sequence is at least 80% identical over at least 80% of their coding or internal domain, or in the terminal repeat regions, or both (13).

Novel TE families include elements that did not fit the 80-80-80 rule (13) showing weak or no similarity with the referenced TE families. Novel TEs were confirmed and analyzed by LTR_STRUC software (14), and/or BLASTx similarity searches against the NCBI nr database (<http://www.ncbi.nlm.nih.gov/>), as well as using the DOTTER program (15). When possible (i.e., for complete TEs), target-site duplications were identified to confirm an insertion by transposition. Each BAC sequence was also analyzed using the DOTTER program to identify or confirm direct repeats, LTRs, local duplications, and deletion events as well as MITEs. TEs were named and classified according to the previously suggested nomenclature (i.e., element name, BAC name, discovery appearance rank) and designated as complete, truncated, or degenerated as previously suggested (13) with two exceptions: the Sukkula large retrotransposon derivatives (LARDs) were considered to be *Gypsy*-like because their LTRs showed sequence BLASTn similarity to Erika (*Gypsy*-like) TEs. We maintained the *Athila* designation for those TEs that were initially referenced in TREP (<http://wheat.pw.usda.gov/ITMI/Repeats/>) to distinguish them from other *Gypsy*-like retrotransposons. They were designated Ga (*Gypsy* retrotransposons, which were initially designated as *Athila*) in the nomenclature and annotation files.

Annotation of Other Repetitive DNA. Short repeats were identified either as inverted repeats (by using EINVERTED with default parameters; <http://bioweb.pasteur.fr/seqanal/interfaces/einverted.html>) or tandem repeats (Tandem Repeat Finder, with default parameters; <http://tandem.bu.edu/trf/trf.advanced.submit.html>). Repeats longer than 100 bp were included in the annotation files.

Parameters; <http://tandem.bu.edu/trf/trf.advanced.submit.html>. Repeats longer than 100 bp were included in the annotation files.

Gene Structure Analysis. Gene prediction analysis was conducted for the 18.5% non-TE and nonrepeated DNA, using the FGENESH gene prediction program (with the Monocot matrix) as well as BLASTn, BLASTx, and tBLASTx alignments against dbEST (<http://www.ncbi.nlm.nih.gov/>), SwissProt (<http://expasy.org/sprot/>), and the rice genome databases.

Estimation of Retrotransposon Insertion Dates. Retrotransposons with both 5' and 3' LTRs and target-site duplications (TSDs) were considered to correspond to original insertions. The number of transitions and transversions in pairs of LTRs was calculated using MEGA3 software (16). Retrotransposon insertion dates were calculated using the substitution rate of 13×10^{-8} substitutions/site/year (17). SEs were estimated using the formula $T = 1/4 K2P/2r$ (18). All retrotransposon insertion date estimates are presented in Table S1.

Primer Design and PCR-Based Analysis of Helitron-Insertion Haplotype Diversity. The Primer3 program (19) was used to design PCR primers (Table S4) on the basis of the sequence of gene-TE or TE-unassigned DNA junctions as previously described (20). Several pairs of primers were often used. PCR primers targeting internal TE sites were used in control amplifications. PCR amplification was carried out in a final volume of 10 μ L with 200 nM of each dNTP, 500 nM each of forward and reverse primers, and 0.2 units of Taq polymerase (Perkin-Elmer) using the following "touchdown" procedure: 14 cycles of 30 s at 95 $^{\circ}$ C, 30 s at 72 $^{\circ}$ C minus 1 $^{\circ}$ C for each cycle, 30 s at 72 $^{\circ}$ C followed by 30 cycles of 30 s at 95 $^{\circ}$ C, 30 s at 55 $^{\circ}$ C, 30 s at 72 $^{\circ}$ C, and 1 cycle of 10 min at 72 $^{\circ}$ C.

cDNA Amplification and Sequencing. Total RNA was isolated from immature spikes of CS, LDN, and RE (with one-tenth of the length of the full-grown spike) using on-column digestion (RNAeasy Plant mini kit, Qiagen) following the manufacturer's instructions. cDNA was prepared using TaqMan reverse transcription reagents with Oligo d(T)16 primer and MultiScribe Reverse Transcriptase (Applied Biosystems) as recommended by the manufacturer. *5Bq* and *5Dq* cDNAs were PCR amplified using primers based on the predicted LDN *5Bq* and *Ae. tauschii* *5Dq* cDNA sequences (Table S4). Gene-specific primers were designed from *5Bq* and *5Dq* cDNA sequences for rapid amplification of cDNA ends (RACE) using the BD SMART RACE cDNA amplification kit (BD Biosciences). The 5' RACE and 3' RACE ready cDNAs were prepared according to the manufacturer's instructions and PCR amplified using primers shown in Table S4. Sequence alignments and amino acid sequence prediction were performed using CLUSTALW and SIXFRAME provided by San Diego Supercomputer Center (SDSC) biology workbench (<http://workbench.sdsc.edu/>).

Transcript Analysis. Steady-state mRNA level in immature spikes collected from CS, CS del143 (*5AQ* deleted), CS 5BL-14 (*5Bq* deleted), CS 5DL-5 (*5Dq* deleted), and the *5AQ/5Dq* double deletion line (CS-5A/5Ddd; both *5AQ* and *5Dq* deleted) was measured by relative quantitative (RQ)-PCR with primers for *5AQ*, *5Bq* (*5Bq.1* and *5Bq.2*), and *5Dq* (*5Dq.1* and *5Dq.2*) using a 7500 Real-Time PCR system (Applied Biosystems). CS-del143 was generated by fast-neutron mutagenesis and contains an interstitial

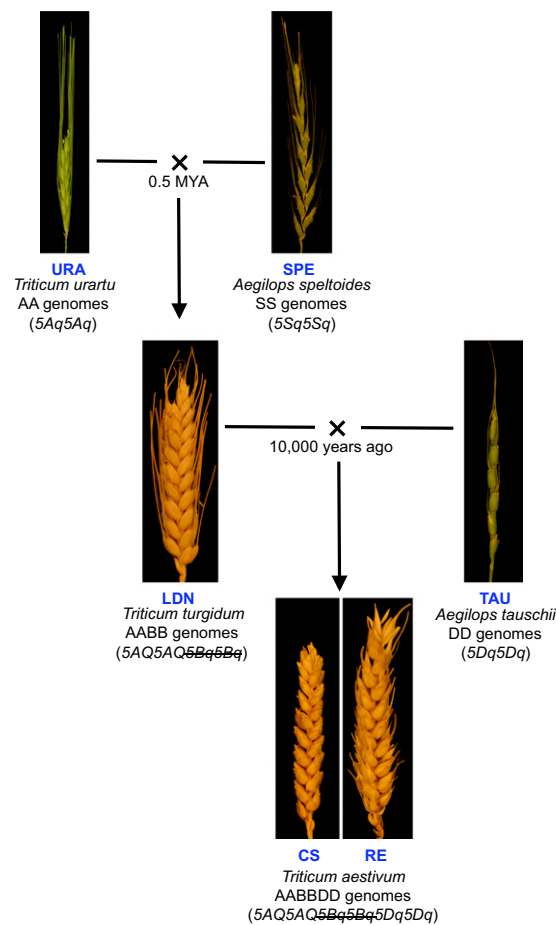


Fig. S1. The amphiploidization events involved in the evolution of common bread wheat illustrated with photographs of spikes of species/cultivars used for BAC sequence analysis of the *Q/q* loci. Designation of accessions used in this research is indicated in blue. Taxonomical names, genome constitutions, and *Q* loci genotypes are shown in parentheses with strikethrough indicating a pseudogene.

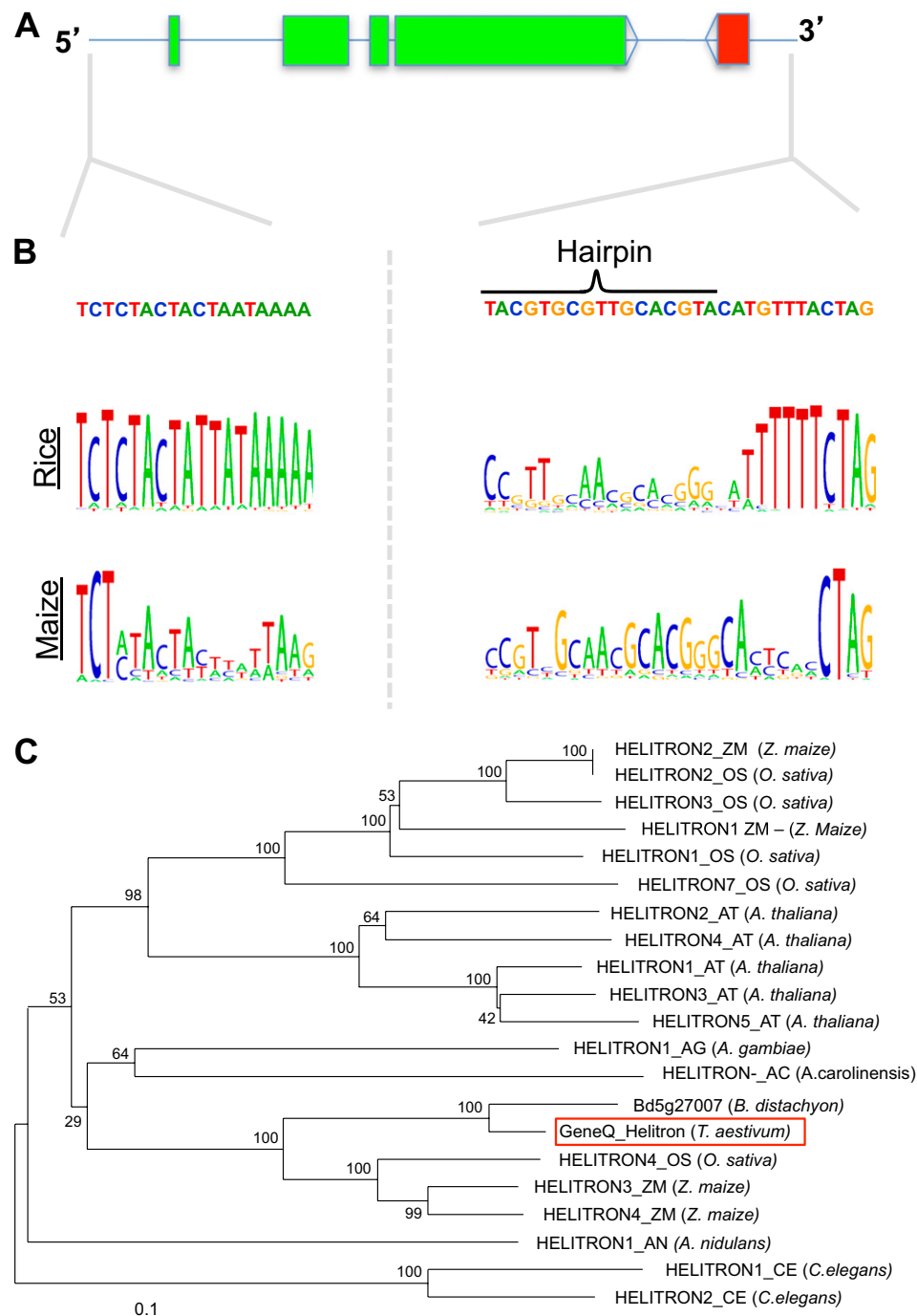


Fig. S3. Analysis of the 9,535-bp helitron inserted in opposite orientation in the *5Bq* gene of *T. aestivum* cv. RE. (A) Intron/exon structure of the helitron, predicting a putative gene of four exons (shown in green), where the putative RepHel protein is of 1,405 amino acids. A gene encoding a 128-amino-acid fragment of a glyoxalate reductase-like protein interrupted by a frameshift is also present in the helitron (shown in red) inserted in the same orientation as *q*. (B) Comparison of the 5'- and 3'-end sequences of the identified helitron shows a typical AT insertion site, the 5' TC, the 3' CTAG, and the sequences leading to the hairpin structure formation at 3' ends, show high conservation with rice and maize helitron consensus sequences. (C) Bootstrapped neighbor-joining tree of predicted RepHel proteins encoded by helitrons from various species [found in rebase (<http://www.girinst.org/rebase/update/index.html>)]. Reference helitron sequences and species are indicated. The protein encoded by the *T. aestivum* cv. RE *5Bq* helitron is shown in the red box. It is 76% similar to RepHel encoded by Helitron4 Os of rice (*AC105746.1*) and 87% similar to a putative *Brachypodium distachyon* helicase (*Bradi5g20077*). The insertion site, presence of the RepHel-like gene, and fragments of other genes are characteristic features of helitrons (1, 2).

1. Du C, Fefelova N, Caronna J, He L, Dooner HK (2009) The polychromatic Helitron landscape of the maize genome. *Proc Natl Acad Sci USA* 106:19916–19921.
2. Yang L, Bennetzen JL (2009) Distribution, diversity, evolution, and survival of Helitrons in the maize genome. *Proc Natl Acad Sci USA* 106:19922–19927.

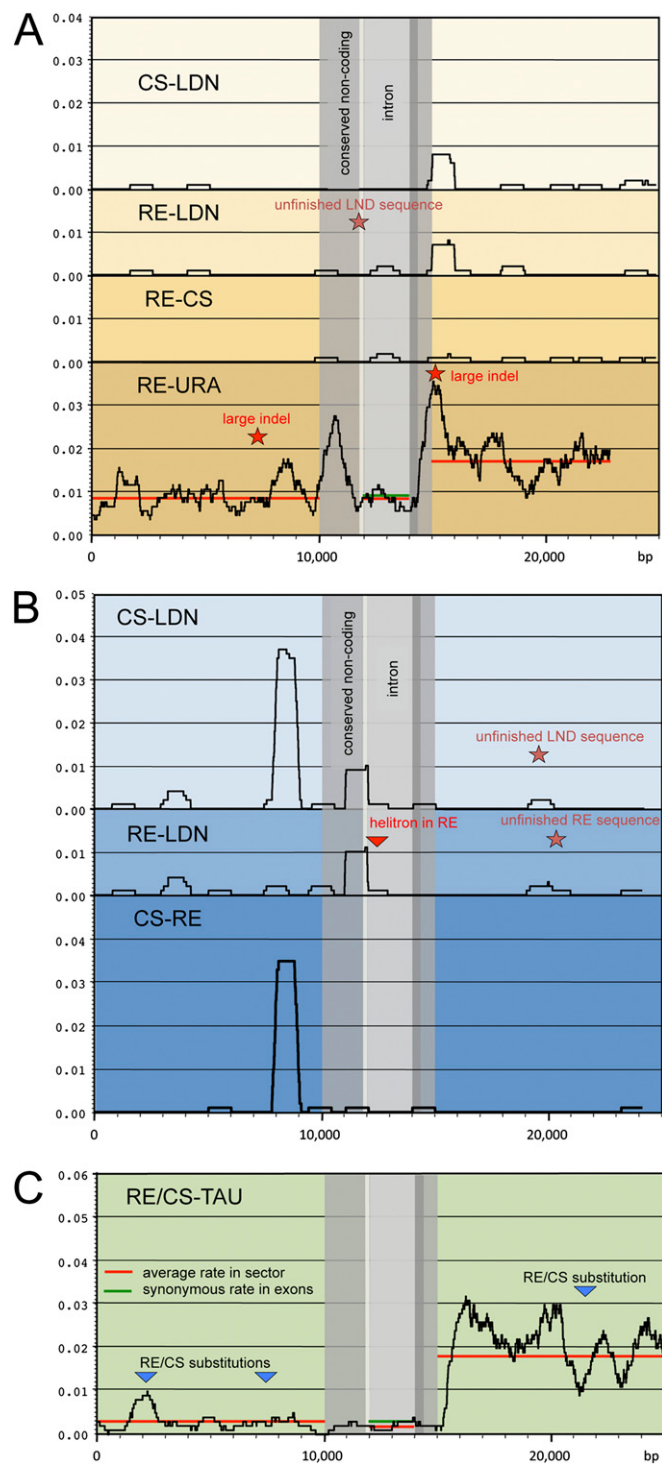


Fig. S4. Pairwise comparisons of the noncoding parts of the *Q/q* genes and the neighboring intergenic regions. Nucleotide substitution rates in a 1-kb window sliding by 1 bp in *5AQ/q*, *5Bq*, and *5Dq* for different pairs of wheats are shown in *A*, *B*, and *C*, respectively. Introns, light gray; UTRs, dark gray; and conserved sequences at the 5' and 3' ends of the *Q/q* gene, medium gray. Red lines indicate average substitution rates for the marked region (intergenic or intron) and green lines indicate synonymous nucleotide substitution rates for the *Q/q* coding sequence.

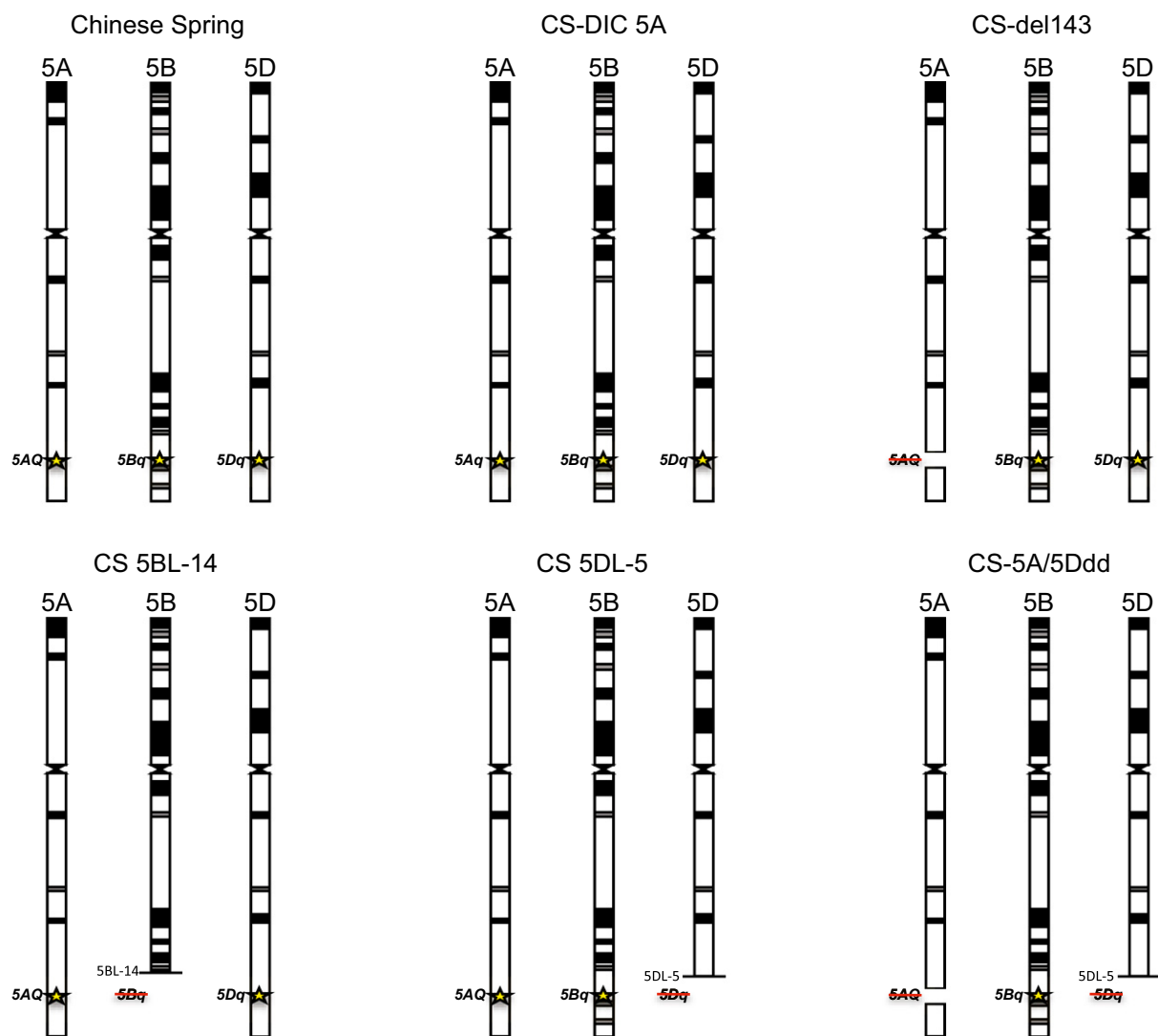


Fig. S5. Homoeologous group 5 chromosome constitutions for the genetic stocks used for *Q/q* expression and phenotypic analysis. Locations of homoeologous *Q/q* loci on the long arms of group 5 chromosomes are indicated by yellow stars. Absence of a *Q/q* locus is indicated by a red strike through. Black and gray regions along the chromosomes indicate dark and light cytological C bands, respectively.

Table S1. Divergence and insertion dates estimated for LTR retrotransposons with both left and right LTRs, and target site duplications identified

TE name	Insertion date (Mya)	Divergence	Std error
A_ContigQRG1_Ale1.fasta	0.66	0.0172	0.0084
A_ContigQRG1_Fatima1.fasta	1.25	0.0326	0.0083
A_ContigQRG1_Melina1.fasta	1.07	0.0279	0.0048
A_ContigQRG1_Wis1.fasta	1.06	0.0276	0.004
A_ContigQURA_Ale1.fasta	0.66	0.0172	0.0083
A_ContigQURA_Claudia1.fasta	0.28	0.0073	0.0022
A_CS12224M17Q_Ale1.fasta	0.66	0.0172	0.0085
A_CS12224M17Q_Fatima1.fasta	1.25	0.0326	0.0083
A_CS12224M17Q_Melina1.fasta	1.01	0.0263	0.0047
A_CS12224M17Q_Sabrina2.fasta	1.12	0.029	0.0046
A_CS12224M17Q_Wis1.fasta	1.06	0.0276	0.0039
A_LDN376H15_Angela1.fasta	0.36	0.0093	0.0025
A_LDN376H15_Angela2.fasta	1.95	0.0507	0.0055
A_LDN376H15_Fatima1.fasta	1.06	0.0275	0.0072
A_LDN376H15_Melina1.fasta	1.23	0.032	0.0053
A_LDN376H15_Sabrina1.fasta	1.49	0.0387	0.005
A_LDN376H15_Sabrina2.fasta	1.30	0.0338	0.005
A_LDN376H15_Sabrina3.fasta	0.93	0.0241	0.0039
A_LDN376H15_Wis1.fasta	1.02	0.0264	0.0039
B_1255P23QRG4_Angela1.fasta	1.11	0.0288	0.0046
B_1255P23QRG4_Derami1.fasta	2.42	0.0629	0.0067
B_1255P23QRG4_Jeli1.fasta	1.20	0.0312	0.0075
B_1255P23QRG4_Wis1.fasta	0.15	0.004	0.0016
B_1255P23QRG4_Wis2.fasta	0.33	0.0086	0.0022
B_CS102N4QF_Angela1.fasta	1.14	0.0296	0.0045
B_LDN1004P5_Angela1.fasta	1.11	0.0289	0.0043
B_LDN1004P5_Barbara1.fasta	1.28	0.0334	0.0043
B_LDN1004P5_Derami1.fasta	2.52	0.0654	0.0073
B_LDN1004P5_Fatima3.fasta	1.31	0.034	0.0079
B_LDN1004P5_Laura1.fasta	0.36	0.0094	0.0015
B_LDN1004P5_Wis1.fasta	0.33	0.0086	0.0022
D_AeT20P19_Angela2.fasta	1.08	0.0282	0.0043
D_AeT20P19_Angela3.fasta	1.02	0.0265	0.0037
D_AeT20P19_Barbara1.fasta	1.02	0.0266	0.0042
D_AeT20P19_Derami1.fasta	1.09	0.0283	0.0042
D_AeT20P19_Maximus1.fasta	1.01	0.0262	0.0041
D_AeT20P19_NewQ04.fasta	0.67	0.0174	0.0068
D_AeT20P19_Valerie1.fasta	0.88	0.023	0.0034
D_AeT20P19_Valerie2.fasta	0.03	0.0009	0.0006
D_ContigQRG5_Ifis1.fasta	1.17	0.0305	0.0077
D_ContigQRG5_Ifis2.fasta	0.55	0.0142	0.005
D_ContigQRG5_Maximus1.fasta	0.87	0.0225	0.0037
D_ContigQRG5_Sabrina1.fasta	0.53	0.0139	0.0031
D_ContigQRG5_Wis1.fasta	0.34	0.0088	0.0022
D_CS201F2QGD_Angela3.fasta	1.04	0.0271	0.0043
D_CS201F2QGD_Ifis1.fasta	0.52	0.0136	0.005
D_CS201F2QGD_Maximus1.fasta	0.84	0.0218	0.0038
D_CS201F2QGD_NewQ04.fasta	0.78	0.0204	0.0076
D_CS201F2QGD_Wis1.fasta	0.32	0.0082	0.0022
S_SH120_23G16_Quinta2.fasta	0.16	0.0041	0.0021
S_SH120_23G16_Quinta3.fasta	0.88	0.0228	0.0047
S_SH120_23G16_Quinta4.fasta	0.83	0.0217	0.0046

Insertion dates were calculated using the substitution rate of 13×10^{-9} substitutions/site/year. SEs were estimated using the formula $T \pm 1/4 K2P/2r$ (1).

1. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120.

Table S2. Helitron is found only in hexaploid wheat cvs Renan and Mironovskaya 808

	Helitron	+	-	+	-
Haplotype	Primers	21F / 21R	22F / 22R	32F / 32R	32F / 32bR
44 - AABBD-D. aestivum cv. Arminda			1		1
45 - AABBD-D. aestivum cv. Chinese Spring			1		1
46 - AABBD-D. aestivum ssp. macha 1793MA86112			1		1
47 - AABBD-D. aestivum ssp. compactum rufulum 71VREB68368			1		1
48 - AABBD-D. aestivum ssp. compactum crebicum 72VREB68369			1		1
49 - AABBD-D. aestivum ssp. compactum erinaceum 75VREB68372			1		1
50 - AABBD-D. aestivum ssp. spaerococcum tumidum perciv globosum REB77050			1		1
51 - AABBD-D. aestivum cv. Recital			1		1
52 - AABBD-D. aestivum cv. Renan		1		1	
52 - AABBD-D. aestivum cv. Mironovskaya 808		1		1	
53 - AABBD-D. aestivum cv. Soisson			1		1
54 - AABBD-D. aestivum ssp. spelta SP81300			1		1
55 - AABBD-D. aestivum ssp. spelta SP86054			1		1
56 - AABBD-D. aestivum cv. Vilmorin			1		1
13 - AABBD-D. aestivum cv. Chopin			1		1
5 - AAB-B. T. turgidum ssp. dicoccoides DIC 118			1		1
6 - AAB-B. T. turgidum ssp. dicoccoides DIC 119			1		1
7 - AAB-B. T. turgidum ssp. dicoccoides DIC 120			1		1
8 - AAB-B. T. turgidum ssp. dicoccoides DIC 132			1		1
9 - AAB-B. T. turgidum ssp. dicoccoides DIC 196			1		1
10 - AAB-B. T. turgidum ssp. dicoccoides DIC 44			1		
11 - AAB-B. T. turgidum ssp. dicoccoides DIC 46			1		
12 - AAB-B. T. turgidum ssp. dicoccoides DIC 54			1		
13 - AAB-B. T. turgidum ssp. dicoccoides DIC 55			1		1
14 - AAB-B. T. turgidum ssp. dicoccoides DIC 87			1		
15 - AAB-B. T. turgidum ssp. dicocum DIC 298			1		1
16 - AAB-B. T. turgidum ssp. dicocum DIM-05			1		1
17 - AAB-B. T. turgidum ssp. dicocum DIM-15			1		1
18 - AAB-B. T. turgidum ssp. dicocum DIM-78			1		1
19 - AAB-B. T. turgidum ssp. durum HTTH01			1		1
20 - AAB-B. T. turgidum ssp. dicocum TTC01			1		1
21 - AAB-B. T. turgidum ssp. dicocum TTC02			1		1
22 - AAB-B. T. turgidum ssp. dicoccoides TTD01			1		1
23 - AAB-B. T. turgidum ssp. dicoccoides TTD03			1		1
24 - AAB-B. T. turgidum ssp. dicoccoides TTD06			1		1
25 - AAB-B. T. turgidum ssp. dicoccoides TTD09			1		1
26 - AAB-B. T. turgidum ssp. dicoccoides TTD121			1		1
27 - AAB-B. T. turgidum ssp. dicoccoides TTD125			1		1
28 - AAB-B. T. turgidum ssp. dicoccoides TTD15			1		1
29 - AAB-B. T. turgidum ssp. dicoccoides TTD20			1		1
30 - AAB-B. T. turgidum ssp. dicoccoides TTD23			1		1
32 - AAB-B. T. turgidum ssp. dicoccoides TTD48			1		1
33 - AAB-B. T. turgidum ssp. dicoccoides TTD58			1		
34 - AAB-B. T. turgidum ssp. dicoccoides TTD72			1		
35 - AAB-B. T. turgidum ssp. dicoccoides TTD86			1		
36 - AAB-B. T. turgidum ssp. dicoccoides TTD88			1		
37 - AAB-B. T. turgidum ssp. dicoccoides TTD95			1		
38 - AAB-B. T. turgidum ssp. polinicum TTP03			1		
39 - AAB-B. T. turgidum ssp. durum TTR04			1		
40 - AAB-B. T. turgidum ssp. durum TTR05			1		
41 - AAB-B. T. turgidum ssp. durum TTR16			1		1
42 - AAB-B. T. turgidum ssp. durum TTR19			1		1
43 - AAB-B. T. turgidum ssp. durum TTT02			1		1
11 - AAB-B. T. turgidum ssp. durum cv. Langdon			1		1
12 - AAB-B. T. turgidum ssp. durum cv. Capelli			1		1
1 - AA-T. monococcum TMB02			1		
2 - AA-T. urartu TMU38			1		
1 - AA - T. beoticum-no_ref			1		
2 - AA- T. beoticum-no_ref			1		
3 - AA- T. beoticum-no_ref			1		

Table S2 Cont.

Haplotype	Helitron	+		–	
	Primers	21F / 21R	22F / 22R	32F / 32R	32F / 32bR
6 - AA- <i>T. urartu</i> -no_ref			1		
7 - AA- <i>T. beoticum</i> -no_ref			1		
8 - AA- <i>T. beoticum</i> -no_ref			1		
9 - AA- <i>T. urartu</i> -no_ref			1		
70 - AAGG- <i>T. timopheevii</i> ssp. <i>armeniicum</i> TIA02			1		
71 - AAGG- <i>T. timopheevii</i> ssp. <i>armeniicum</i> TIA06			1		
72 - AAGG- <i>T. timopheevii</i> ssp. <i>armeniicum</i> TIA28			1		
73 - AAGG- <i>T. timopheevii</i> ssp. <i>armeniicum</i> TIA50			1		
74 - AAGG- <i>T. timopheevii</i> ssp. <i>timopheevii</i> TIM01			1		
75 - AAGG- <i>T. timopheevii</i> ssp. <i>timopheevii</i> <i>viticulosum</i> TIM05			1		
77 - AAGG- <i>T. timopheevii</i> ssp. <i>timopheevii</i> zhuk (k38555) TIM07			1		
78 - AAGG- <i>T. timopheevii</i> TIM09			1		
79 - AAGG- <i>T. timopheevii</i> ssp. <i>typicum</i> TR103			1		
80 - AAGG- <i>T. araraticum</i> TR94			1		
81 - AAGG- <i>T. araraticum</i> TR95			1		
82 - AAGG- <i>T. araraticum</i> TR96			1		
83 - AAGG- <i>T. araraticum</i> TR99			1		
14 - AAGG- <i>T. timopheevii</i> ssp. <i>timopheevii</i> var <i>typica</i> -no_ref			1		
15 - AAGG- <i>T. timopheevii</i> ssp. <i>araticum</i> var <i>Schulze-motel</i> -no_ref			1		
16 - AAGG- <i>T. timopheevii</i> ssp. <i>militinae</i> -no_ref			1		
18 - AAGG- <i>T. timopheevii</i> ssp. <i>timopheevii</i> -no_ref			1		
84 - DD-Ae. <i>kotschy</i> var. <i>palaestina</i> AEG 506_1			1		

Empty cells, no PCR amplification; cells showing number 1, PCR amplification confirming helitron insertion.

Table S3. Traits of genetic stocks with various combinations of *Q/q* homoeoalleles and chromosomal deletions including the *Q/q* locus

Wheat line <i>Q/q</i> genotype (AABBDD)	Spike length (mm)	Spike emergence (days)	Plant height (cm)	Yield components					Grain yield (g/plant)
				Tillers/ plant	Spikelets/ spike	Seeds/ spike	Seed weight (mg)		
CS 5DL-5 (<i>QQ qq</i> –)	67 (63–72)	114 (112–117)	102 (96–108)	11.4 (9.2–13.6)	20 (18.4–21.2)	12 (6–19)	20 (18–23)	3.0 [3.0] (1.2–4.7)	
CS (<i>QQ qq qq</i>)	72 (69–75)	80 (78–82)	118 (116–120)	10.3 (9.3–11.2)	25 (24.6–25.5)	36 (30–42)	31 (28–33)	11 [11] (9.6–13.0)	
CS 5BL-14 (<i>QQ – qq</i>)	62 (57–67)	76 (72–79)	98 (88–107)	8.6 (7.1–10.2)	20 (18.4–21.8)	9 (4–14)	18 (16–19)	1.6 [1.4] (0.5–2.7)	
CS-5A/5Ddd (– <i>qq</i> –)	81 (78–83)	113 (110–115)	110 (105–115)	12.1 (9.7–14.6)	17 (16.1–17.9)	6 (5–7)	18 (17–20)	1.4 [1.3] (1.0–1.7)	
CS-del143 (– <i>qq qq</i>)	101 (98–105)	73 (71–75)	123 (119–127)	10.8 (9.7–11.8)	25 (24.6–26.1)	37 (32–42)	29 (28–30)	12 [12] (9.8–13.3)	
CS-DIC 5A (<i>qq qq qq</i>)	110 (107–114)	68 (66–69)	129 (122–134)	9.9 (8.2–11.5)	28 (27.2–28.3)	47 (43–50)	29 (27–31)	13 [13] (11.6–14.7)	

~~q~~, pseudogene; –, chromosomal deletion including the *Q/q* locus; 95% confidence intervals are shown in parentheses; grain yield calculated on the basis of the yield components is shown in brackets.

VI-2 Conclusions sur les travaux publiés dans l'article et orientations des recherches complémentaires :

VI-2-1 Devenir et interactions entre les copies homéologues du gène *Q/q* dans le blé hexaploïde :

Dans cette publication, une attention particulière a été portée sur la contribution, les interactions et la partition des copies homéologues du gène *Q/q* dans le blé hexaploïdes et leur évolution par rapport aux espèces parentales diploïdes et tétraploïdes. L'utilisation des lignées de délétion, où une ou plusieurs copies homéologues sont délétées a été déterminante pour apprécier le devenir de ces copies dupliquées. Ceci m'a permis de conclure que les trois homéologues se régulent entre eux dans le blé hexaploïde et contribuent (à différents degrés) aux caractères de domestication: Une hyperfonctionnalisation de l'homologue 5AQ qui a l'effet majeur sur les caractères de domestication; une

sous-fonctionnalisation de l'homéologue *5Dq* et une pseudogénéisation de l'allèle *5Bq* qui est aussi sous-fonctionnalisé car il continue à contribuer au phénotype et à réguler les autres homéologues.

Le fait que l'homéologue *5Bq* ne code plus pour une protéine et continue à réguler l'expression des autres copies homéologues suggère une interaction à un autre niveau que celui des protéines. En effet les résultats d'interactions entre les différents homéologues (pseudogène compris) des allèles *Q/q* pourrait correspondre à un mécanisme de régulation suggéré entre les gènes codants, les gènes à ARN (non codants), et les transcrits de pseudogènes (Salmena et al., 2011). A partir d'études réalisées sur des cellules cancéreuses, les auteurs émettent l'hypothèse qu'il existe une régulation des différentes voies métaboliques (jusqu'à la mise en place de tissus ou d'organes) par des interactions entre les transcrits de gènes codant, de pseudogènes et de gènes à ARN (précurseurs des miRNA). Il a été établi que les miRNA régulent les quantités d'ARNm transcrites ou bloquent la traduction. Entre les ARNm, il existerait une compétition endogène pour la fixation de l'ensemble des miARNs présent dans la cellule, que les auteurs nomment « ceRNAs » (competitive endogenous RNAs). Les miRNA seraient accrochés dans la région 3'UTR sur des sites que les auteurs nomment « microRNA Response Element » (MREs). Il existerait autant de MREs que de miRNA différents dans l'organisme, chaque ARNm possède différents MREs le long de sa séquence en 3'UTR. L'hypothèse est que les ARNm d'un gène peuvent modifier la quantité et la traduction d'un ou d'autres ARNm issu d'autres gènes par l'intermédiaire de micro ARNs via des MREs communs. Cette hypothèse amène à trois constats. Le premier constat est le rôle des transcrits de pseudogènes dans le processus de régulation d'autres ARNm en fixant une quantité de miARNs. Le deuxième constat est de supposer l'existence d'un vaste réseau de régulation entre transcrits de différentes voies métaboliques qui interagiraient par l'intermédiaire de miARNs et de MREs communs. Le troisième constat est que la régulation de la concentration en miARN fait parti du réseau de régulation car les transcrits (pré-miRNA) précurseurs de miARNs peuvent également porter des sites MREs. Cette théorie a orienté mes recherches vers le rôle des miARNs et leurs actions avec les transcrits des gènes *Q/q*.

VI-2-2 Dynamiques des éléments transposables :

Le séquençage des différents clones BAC couvrant les copies orthologues et homéologues du gène *Q/q* m'a permis de réaliser une analyse précise et comparative d'un total de 11 régions différentes des espèces de blé diploïdes, tetraploïdes et hexaploïdes, ce qui représente, à ma connaissance, l'analyse comparée la plus importante réalisée sur le blé.

Je présente ici les principales conclusions sur les résultats d'annotation et de dynamique des éléments transposables (TEs) en comparant les différents génomes A, B/S et D dans différents niveaux de ploïdie (cf Supporting Information, Fig. S2, S3 et Table S1).

(i) - Les comparaisons montrent que l'espace des TEs, représentant environ 80% du génome, est complètement différente entre les génomes A, B/S et D (Fig. S2). Les TEs ont donc été complètement remplacés depuis la divergence de ces quatre génomes par une dynamique d'insertion et de délétion, en une période relativement courte, soit 2,5 à 4 millions d'années séparant ces génomes du blé (Chalupska et al., 2008 ; Salse et al., 2008a ; Salse et al., 2008b). Ce résultat corrobore avec des résultats antérieurs obtenus au laboratoire, auxquels j'ai contribué en étudiant d'autres régions du blé (Chantret et al., 2005 ; Gu et al., 2006 ; Chalupska et al., 2008 ; Chantret et al., 2008 ; Salse et al., 2008a).

(ii) - Ces analyses permettent également la comparaison des différents sous-génomes A, B/S ou D entre eux ; provenant d'espèces diploïdes ou polyploïdes, ayant divergé il y a moins longtemps : 20 000-300 000 ans pour les génomes A; 10 000-100 000 ans pour les D et 2,5 Ma pour les B/S).

a- L'espace des TEs se trouve plus conservée entre les différents génomes A comparés entre eux. Les comparaisons m'ont permis d'identifier sept événements de rupture de synténie, correspondant tous à des événements d'insertions/délétion de TEs, qui ont eu lieu depuis la divergence des génomes A (Fig. S2, Table S1)

b- L'espace des TEs est également majoritairement conservée entre les différents génomes D comparés, avec trois événements majeurs de dynamique d'insertion/délétion différentielles de TEs.

c- La conservation de l'espace des TEs n'est pas le cas pour les génomes B des blés hexaploïde et tetraploïdes comparés au génome S d'*Ae. Speltoides*. En effet, l'espace TEs est complètement différente et reflète la divergence plus importante entre les génomes S et B (Chalupska et al., 2008; Salse et al., 2008a ; Salse et al., 2008b). Le génome S n'est donc pas le progéniteur direct du génome B bien qu'il soit le génome diploïde le plus proche. Parmi les 9 événements de rupture de synténie observés entre ces génomes (Fig.S2), j'ai pu dater des insertions de retrotransposons récentes (0,1-0,2 Ma) (Fig.S2, Table S1) et des insertions différentielles de retrotransposons beaucoup plus anciennes (2,02-2,28 Ma). Ceci démontre que la divergence entre les deux génomes est plus ancienne.

VI-2-3 Identification et caractérisation du premier *Hélitron* actif du blé :

Parmi les événements de rupture de synténie observés en comparant le génome B entre plusieurs cultivars du blé, j'ai pu identifier et caractériser que l'événement N°13 correspond à un *Hélitron* inséré dans le gène *5Bq* de *T.aestivum* ssp. *aestivum* cv. Renan (Fig. S2, S.3). Son insertion serait récente car, en plus d'être complet, il est absent dans la plupart des autres génotypes de blés analysés (Table S2). C'est le premier *Hélitron* actif découvert chez le blé.

Ceci amène un certain nombre de questions sur son origine et sa transposition dans la variété Renan.

Parmi la diversité des TEs existants, les *Hélitrons* ont une incidence particulière. C'est la raison pour laquelle je me suis focalisé d'avantage sur sa caractérisation.

Je présente tout d'abord quelques généralités bibliographiques afin de situer le contexte et ma démarche.

VI-3 Généralités sur les *Hélitrons* :

Les *Hélitrons*, découverts relativement récemment dans le maïs (*Zea mays*) (Kapitonov and Jurka, 2001), puis dans les séquences d'autres génomes complets (*Arabidopsis thaliana*, *Oriza sativa* et *Caenorhabditis elegans*) (Kapitonov and Jurka, 2001), utilisent pour leur transposition, un mécanisme singulier, dit de 'rolling-circle' en référence à un mécanisme similaire observé chez les geminivirus. Ils ont probablement une origine virale, car il y a une forte similitude avec des geminivirus (Mendiola et al., 1994; Murad et al., 2004), mais ces virus n'ont pas encore été trouvés chez les procaryotes et dans le règne animal.

A partir des séquences disponibles, la construction de trois séquences consensus a permis d'identifier 2 à 23 familles différentes sur la base des séquences en acides aminés de la protéine RPA (ssDNA-binding replication protein A), un des deux domaines protéiques identifiés dans (Kapitonov and Jurka, 2001). Les *Hélitrons* représentent en moyenne 2% du génome (*Arabidopsis thaliana*, 1.30% ; *Oriza sativa*, 2% ; *Sorgum bicolor*, 3%), mais il est possible qu'un processus d'élimination biaise les résultats (Yang and Bennetzen, 2009).

Le palindrome ou « hairpin » à la fin de la séquence de l'*Hélitron* (Fig. S3-B) joue le rôle de terminateur de la réplication par 'rolling-circle'. S'il n'est pas bien reconnu, la synthèse continue sur le site donneur, si bien que le brin d'ADN transféré au site accepteur peut transporter non seulement l'*Hélitron*, mais aussi des séquences adjacentes en 3' du site donneur. Ce phénomène a été notamment observé dans le maïs (Morgante et al., 2005) où des fragments de gènes, voire des gènes entiers sont ainsi transportés à d'autres endroits du génome.

Suite à cette découverte, les recherches se sont d'avantages orientées sur le rôle des *Hélitrons* dans le remodelage du génome et de leurs facultés à créer de nouveaux gènes. Comparativement à MULE, un autre transposon à ADN dans le génome du riz qui agirait comme vecteur dans des transferts horizontaux, les *Hélitrons* seraient spécialisés dans les transferts intra-espèce et ne provoqueraient pas la destruction de l'allèle sauvage (Jiang et al., 2004). Dans l'espèce *C.elegans* un lien a été établi entre les *Hélitrons* et les séquences microsatellites adjacentes dans le même ordre et la même direction que les *Hélitrons* insérés.

La première insertion caractérisée dans un gène est celui d'un *Hélitron* de 12kb dans l'intron 11 (sur 16 exons) du gène *Sh2* du maïs (Lal et al., 2003). Tout comme pour celui trouvé dans le gène *5Bq* de

Renan, les auteurs démontrent l'existence de transcrits avec la totalité de l'*Hélitron* (2 types de clones isolés, avec ou sans l'intron 9).

Il est admis que les *Hélitrons* sont responsables de la capture de gène. Si ce mécanisme n'apporte rien il y a pseudogénéisation du gène. Sinon il est conservé car bénéfique pour le transposon et au moins non nuisible pour l'hôte. Ces intégrations sont à l'origine d'une variabilité de 40 à 50% en séquence entre les différentes familles d'*Hélitrons* connues. Deux et six ans après, les mêmes auteurs présentent dans d'autres études (Lal and Hannah, 2005 ; Lal et al., 2009), un exemple synthétique, concret du transport et du remodelage de certains gènes par les *Hélitrons*.

Dans le maïs la fonction de nombreux gènes est influencée par la présence d'un *Hélitron* (exemple du locus *bz* (bronze)). Des études sur des lignées recombinantes montrent que 20% des 20 656 gènes étudiés ont été remodelés par les *Hélitrons* (Feschotte and Pritham, 2009). Suite à tous ces résultats une partie de la communauté scientifique s'est offusquée, car toutes ces études demeuraient *in silico*, basées sur l'étude et les comparaisons des séquences. Pour pallier à cela, une expérience d'excision *in vitro* d'un *Hélitron* dans le maïs a été réalisée pour mimer une activité. Puis à partir de marqueurs PCR flanquant des *Hélitrons* insérés connus, l'auteur a fait un crible PCR, le séquençage et l'analyse des séquences de lignées de descendants. Il observe plusieurs cas de Fig.s, des lignées où il n'y a pas d'insertion et des lignées où il y a insertion d'un *Hélitron*. Une preuve supplémentaire de leur activité (Li et al., 2009).

Jusqu'à maintenant il n'y avait pas de preuve directe de leur activité dans le blé, mais le fait de trouver cette insertion uniquement dans quelques lignées de blé (Table S2) conforte cette idée.

VI-3-1 caractérisation de l'*Hélitron* actif du blé :

Dans le but de caractériser l'*Hélitron* (9535pb) inséré dans le gène *5Bq* de cv. Renan quatre approches ont été réalisées.

(i) La première analyse permet d'identifier la protéine RepHel (Fig. S3 A) propre au *Hélitron* (Kapitonov and Jurka, 2001). Sa structure est faite de quatre exons pour une protéine prédite de 1405 acides aminés. On remarque comme cela a déjà été décrit, un mécanisme de capture de gène ou d'exons chez les *Hélitrons* par la présence d'un fragment de pseudogène correspondant à une glyoxylate reductase-like inséré dans le même sens que le gène *q* (Fig. S3 A).

(ii) Pour la deuxième analyse (Fig. S3. B), j'ai comparé les extrémités 5' et 3' de l'*Hélitron* avec les extrémités des *Hélitrons* d'autres espèces (riz et maïs). On retrouve le site préférentiel d'insertion AT ainsi que les signatures TC en 5' et CTAG en 3' de l'élément avec un haut niveau de conservation avec les *Hélitrons* du maïs et du riz. De plus on retrouve la structure en épingle à cheveux (29pb) dans la région 3' de l'*Hélitron* qui joue un rôle dans l'arrêt de la réplication de l'élément.

(iii) La troisième étude est la réalisation d'un arbre (Fig. S3. C) suivant la méthode du « neighbor-joining » avec les séquences de la protéine RepHel d'*Hélitrons* de différentes espèces. La somme des tirages (Bootstrap) nous montre que la protéine RepHel de l'*Hélitron* inséré dans le gène *5Bq* de Renan a 76% de similarité avec la protéine RepHel de l'*Hélitron4_OS* du riz (AC107746.1) et 87% de similarité avec une Helicase de *B. distachyon* (Bradi5g20077). L'ensemble de ces données permettent de valider que l'élément identifié correspond à un *Hélitron* des *Poaceae*.

(iv) Dans la quatrième approche, j'ai retracé l'insertion de cet élément dans le *5Bq* en analysant une collection de blés diploïdes, tétraploïdes et hexaploïdes (Table S2). J'ai montré que cette insertion est trouvée uniquement dans Renan et dans un autre cultivar hexaploïde (Moronoskaya 808) d'origine ukrainienne.

Cette insertion serait donc très récente, car seulement présente dans deux cultivars du blé hexaploïde.

A partir des informations apportées sur l'*Hélitron* du gène *5Bq*, j'ai voulu caractériser d'avantage son origine dans le blé *T. aestivum* cv. Renan. Une opportunité s'est présentée avec les données issues du séquençage du génome entier de 5X en 454 de la variété Chinese Spring du blé hexaploïde (Brenchley et al., 2012). Je me suis servi de ces séquences dans le but de distinguer les catégories d'*Hélitrons* qui pouvaient être présentes dans le génome du blé hexaploïde et analyser leur proximité avec l'*Hélitron* actif de Renan. Dans une seconde étape, j'ai retracé par PCR la présence/absence des catégories trouvées dans Chinese Spring ainsi que l'*Hélitron* actif de Renan dans d'autres espèces et génotypes du blé.

Groupe	Numéro	Identifiant 454	Taille (pb)	Couples Amorces PCR internes hélitron	Produit PCR (pb)	Couples Amorces PCR interne hélitron /externe	Produit PCR (pb)
Renan	[0]	Hélitron dans 5Bq		A75	266	A21	501
	[10]	GG1VXPK01B4N5R	501				
2	[2]	GIB3ZOT04JGZ0J	527	A79	83	A86	182
	[6]	GG1VXPK04C9ACN	501				
	[7]	F7IO22A0214JP8	488				
3	[4]	GKVVV9KP08IV6ON	667	A83	94	A82	247
	[9]	GIB3ZOT04H35YQ	433				
	[8]	GJZO2FP01C141V	397				
1	[1]	F2NKJR10215F83	470	A74	139	A73	219
	[5]	GGSKSHR01CKBUM	572				
	[12]	GCITVBT01BPGBP	540				
	[3]	GIIP3QJH01BYXJE	499				
	[13]	GJ20ROV01AI9QU	497				
	[11]	GGWFHOC02FTJO2	511				

Tableau VI-1 Amorces spécifiques de l'hélitron de Renan et des 3 différentes catégories d'hélitrons de Chinese Spring :

Il y a deux couples différents (interne-interne et interne-externe) dessiné par catégorie d'Hélitrons identifiée par comparaison de séquences. En gris les séquences non-utilisable pour dessiner les amorces.

VI-3-2 Identification des différentes catégories d'*Hélitrons* dans le blé hexaploïde Chinese Spring :

Dans ma démarche, j'ai évité une recherche de similarité à partir des domaines protéiques réplicase-hélicase (RepHel) pour ne pas confondre, dans ma recherche, les vrais gènes *hélicases* du blé et la protéine RepHel des *hélitrons*, les séquences de Chinese Spring n'étant pas assemblé en contigs assez longs. J'ai donc utilisé la séquence couvrant la structure en épingle à cheveux ('hairpin', signature des *Hélitrons*) à la fin de l'élément complet de Renan, soit 29pb en 3', vue sa conservation avec des *hélitrons* du maïs et du riz. A partir d'une stringence de BLASTN relativement souple (E-value=10), j'ai récupéré 13 séquences de Chinese Spring (Tableau VI-1), ayant une forte similarité avec la séquence 'Hairpin' de l'*Hélitron* de Renan. Les 13 séquences sont hétérogènes en longueur (Tableau VI-1) Pour chacune d'entre elles j'ai réalisé deux jeux de séquences:

Le premier jeu de séquences représente uniquement les séquences des *hélitrons* couvrant la structure hairpin jusqu'à la fin exacte de l'*Hélitron*. Ce jeu m'a permis de définir les différentes catégories d'*Hélitron* trouvées dans le cultivar Chinese Spring.

Le deuxième jeu de séquences représente pour chaque la partie 3' de l'*Helitron* et la séquence génomique adjacente. Chaque séquence plus courte dans l'élément, débute de l'hairpin et continue en 3' après la fin de l'*Hélitron* pour une longueur de 170pb. Ce jeu doit permettre d'identifier les insertions différentes et/ou la redondance de séquence.

VI-3-2-1 Les catégories d'*Hélitrons* identifiés :

La comparaison des 13 séquences des *Hélitron* de Chinese Spring (CS) entre elles montrent qu'elles se regroupent en trois catégories, j'ai utilisé la séquence d'un *Hélitron* du maïs (*Z. mays*) en « out group » pour construire un arbre (Fig. VI-1) à partir de l'alignement des séquences (Annexe 3). Les groupes 1, 2 et 3 contiennent respectivement 6, 3 et 4 séquences.

Quand on ajoute la séquence de l'*Hélitron* de Renan dans l'alignement et que l'on analyse l'arbre, on constate que l'élément inséré dans Renan est plus proche des *Hélitrons* de Chinese Spring (CS) que de ceux du Maïs. Néanmoins, il ne rentre dans aucuns des trois groupes définis pour CS. Ceci suggère fortement que l'*Hélitron* actif de Renan n'est pas présent dans Chinese Spring. A t-il une autre origine que le blé hexaploïde?

VII-3-2-2 Bilan des insertions par catégories :

A partir de l'arbre réalisé sur les séquences internes spécifique des l'*Hélitrons* et face à l'hétérogénéité des tailles de séquences, j'ai éliminé trois séquences du groupe 1 et une séquence du groupe 3. Ceci m'a permis de recommencer l'alignement des différentes séquences 56pb en 5' avant la fin de

l'élément, et 235pb flanquant l'*Hélitron* en 3'. La conclusion est qu'il n'y aurait que trois séquences différentes (une pour chacun des 3 groupes identifiés) flanquant l'*Hélitron* en 3' (Annexe 3).

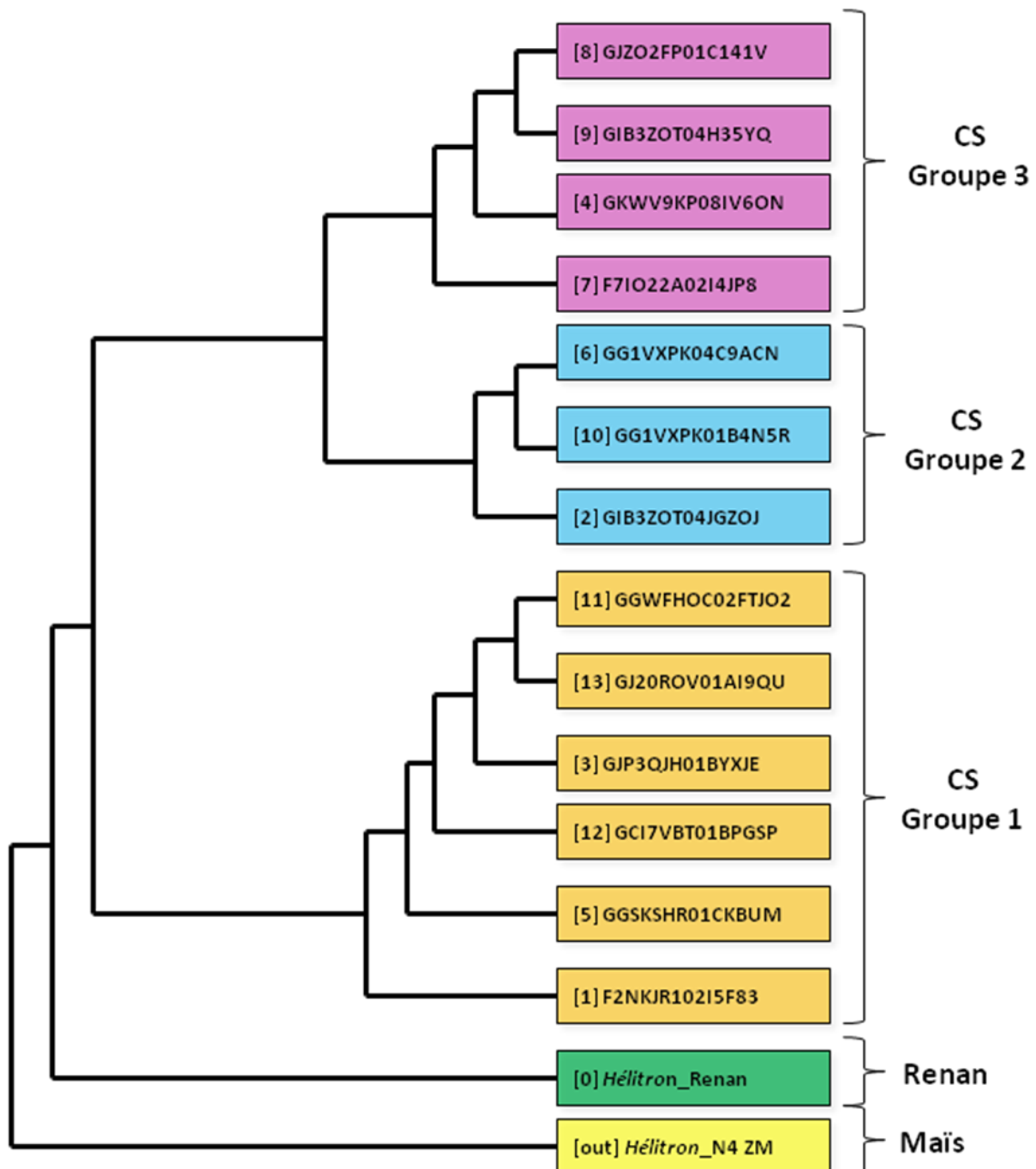


Fig. VI-1 Arbre réalisé à partir de 13 séquences 3' terminale (structure en épingle à cheveux) internes des différentes séquences d'*Hélitron*-like de Chinese Spring et l'*Hélitron* du 5ABq.

Pour valider que chaque hélitron et que chaque insertion sont bien différentes dans le génome, des marqueurs ont été dessinés, suivant la taille des séquences disponibles.

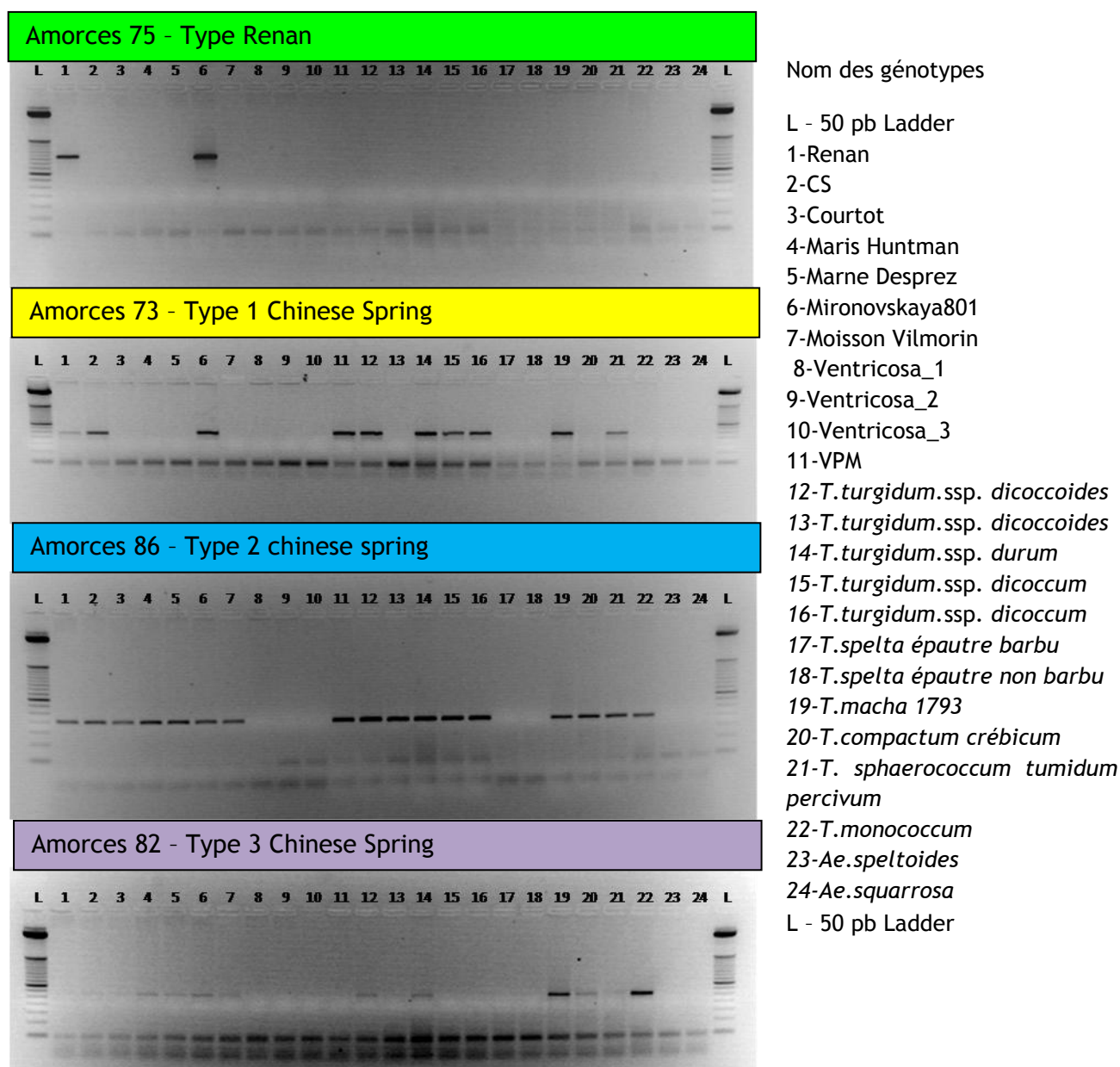


Fig. VI-2 Détection de différents Hélitrons du blé et de leurs insertions dans une collection de blés :

Sur 24 génotypes de blés, on remarque la spécificité des insertions suivant les différentes espèces. Les Hélitrons trouvés dans Chinese Spring n'existent pas dans les blés hexaploïdes anciens tel que les *Triticum aestivum ssp. spelta*. Les Hélitrons sont également absents d'*Ae. speltoides* et d'*Ae. tauschii*.

Nom	Accession	Genomes	Type 1 A	Type 1 B	Type 2 A	Type 2 B	Type 3 A	Type 3 B	Type Renan A	Type Renan B
Couple d'amorce			A74	A73	A79	A86	A83	A82	A75	A21
1-Renan	6086	ABD								
2-CS		ABD								
3-Courtot	2358	ABD								
4-Maris Huntman	4594	ABD								
5-Marne Desprez	4625	ABD								
6-Mironovskaya801	4856	ABD								
7-Moisson Vilmorin	4907	ABD								
8-Ventricosa_1	26258	DN								
9-Ventricosa_2	26259	DN								
10-Ventricosa_3	26260	DN								
11-VPM	26839									
12-T.turgidum.ssp dicoccoides	TTD09	AB								
13-T.turgidum.ssp dicoccoides	TTD06	AB								
14-T.turgidum.ssp durum	TTR16	AB								
15-T.turgidum.ssp dicoccum	TTT02	AB								
16-T.turgidum.ssp dicoccum	TTC02	AB								
17-T.spelta épautre barbu	SP81300	ABD								
18-T.spelta épautre non barbu	SP84011	ABD								
19-T.macha 1793	MA86112	ABD								
20-T.compactum crébicum	REB68369	ABD								
21-Sphaerococcum tumidum percivum	REB77	ABD								
22-T.monococcum	TMB02	A								
23-Ae.speltoides	SPE479-SS	S								
24-Ae.squarrosa	SQUA12	D								

Tableau VI-2 Récapitulatif des 16 couples d'amorces sur les 24 génotypes de blés :

(A) Amorces internes spécifiques de l'*Hélitron* (colonne A),

(B) Amorces spécifiques de l'insertion de l'*Hélitron* (colonne B).

On remarque la spécificité des insertions suivant les différents génotypes. Les *Hélitrons* trouvés n'existent pas dans les blés hexaploïdes anciens tel que les *T. aestivum* ssp. *spelta*, ni dans les espèces d'*Aegilops* diploïdes. Par contre on remarque que l'*Hélitron* de Renan est présent dans les *Ae. ventricosa*, mais il est absent des ammidoniers sauvages et domestiqués ainsi que l'épautre.

VI-3-2-3 Traçage des *Hélitrons* identifiés :

VI-3-2-3-1 Matériel et méthodes :

Choix des amorces PCR :

Des couples d'amorces spécifiques ont été dessinés sur les séquences, du gène *5Bq* contenant l'*Hélitron* de *T. aestivum* ssp. *aestivum* cv. Renan et sur les différentes séquences d'*Hélitron* récupérées dans la base de séquences 'shotgun' de *T. aestivum* ssp. *aestivum* var Chinese Spring.

A l'aide du logiciel OLIGO7[®] conçue pour dessiner des amorces, les paramètres retenus sont une température d'hybridation des amorces comprise entre 55-60°C et une taille des produits PCR comprise entre 100 et 300pb. A partir de l'alignement (Annexe 3) réalisé entre les séquences (internes et externe à l'élément) de Chinese Spring, une seule séquence est retenue pour chaque groupe d'*Hélitron*. A cela j'ajoute des amorces spécifiques de l'*Hélitron* inséré dans le gène *5Bq* de Renan pour la région équivalente. Pour chacune de ces séquences, j'ai dessiné deux couples d'amorces PCR spécifiques, un couple dans la séquence de l'*Hélitron* (interne à l'élément) pour détecter sa présence dans d'autres blés et un couple composé d'une amorce dans la séquence de l'*Hélitron* et l'autre dans la séquence 3' adjacente, pour détecter la même région d'insertion dans la collection des génotypes de blé (Tableau VI-1 et Annexe 4).

Extraction d'ADN génomique et réalisation des PCR :

L'extraction de l'ADN génomique des génotypes supplémentaires et les PCRs incluses dans cette analyse complémentaire ont été réalisés selon les mêmes protocoles décrit dans le papier précédemment (Zhang et al., 2011).

VI-3-2-3-2 Résultats sur la caractérisation des *Hélitrons* de CS et de Renan :

J'ai utilisé ces marqueurs sur 24 génotypes représentant différents niveaux de ploïdies, soit di-, tétra- et hexaploïdes (Fig. VI-2 et Tableau VI-2). Les résultats montrent une spécificité des insertions localisées suivant les différents génotypes. Les trois catégories d'*Hélitrons* identifiés dans CS ne sont pas présentes dans le blé *T. aestivum* ssp. *spelta*.

On trouve les trois catégories d'*Hélitrons* de Chinese Spring dans les génomes A et B, mais tous les *Hélitrons* n'ont pas été identifiés dans le génome S d'*Ae. speltoïdes* et ni dans le génome D d'*Ae. tauschii*.

Comme observé dans la première étude (Table S2) l'*Hélitron* de Renan est absent dans les blés tétraploïdes (*T. turgidum* ssp. *dicoccum*, ssp. *dicoccoides*) et dans certains blés hexaploïdes (*T. aestivum* ssp. *spelta*, ssp. *sphaerococcum*). On remarque cependant que l'*Hélitron* de Renan est le seul présent dans les *Ae. ventricosa*.

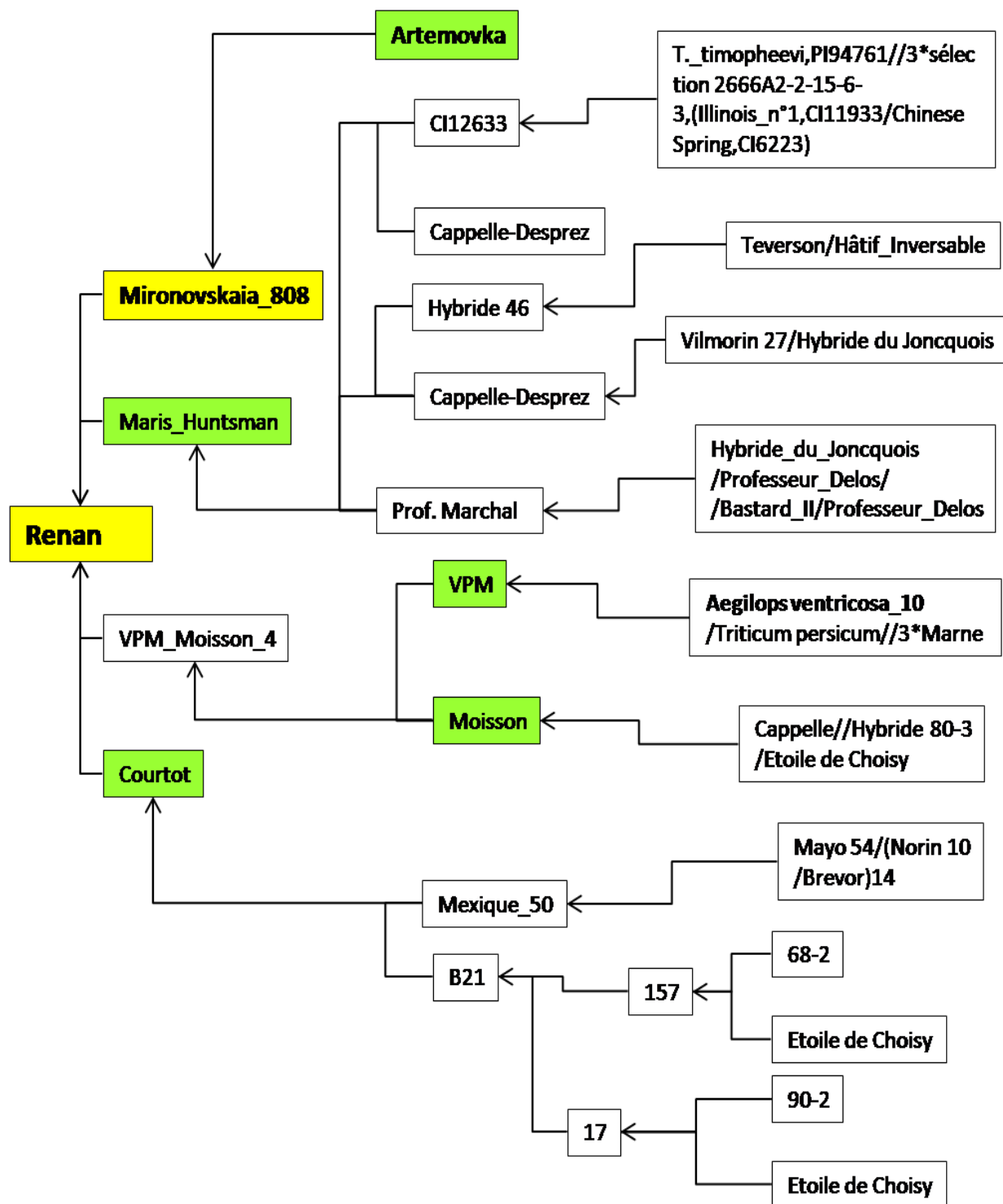


Fig. VI-3 Généalogie de la variété Renan :

Les rectangles verts et jaunes représentent les génotypes testés avec le marqueur Hélitron spécifique du gène 5Bq de Renan. Les rectangles jaunes sont les génotypes positifs pour l'insertion de l'Hélitron dans le gène 5Bq.

(Joseph Jahier – INRA de rennes)

ACCESSION	NOM	PAYS	INSCRIPTION	GENOME
29940	CRIMEAN	UKR	1873	ABD
24029	KRYMKA	UKR	1875	ABD
4102	KHARKOF	UKR	1900	ABD
24026	BANATKA POZDNYAYA	BLR	1908	ABD
24028	BANATKA	RUS	1911	ABD
1198	BANATKA RAWSKA A	POL	1920	ABD
24039	BANATKA	RUS	1924	ABD
24043	BANATKA	UKR	1924	ABD
24047	CRIMEAN	HUN	1924	ABD
24043	BANATKA	UKR	1924	ABD
7099	UKRAINKA	UKR	1929	ABD
8056	NOVOKRYMKA 102	UKR	1929	ABD
13379	UKRAMCA	UKR	1929	ABD
24038	UKRAINKA	UKR	1929	ABD
24044	DYURABL	UKR	1929	ABD
24046	ZEMKA	UKR	1929	ABD
24048	KOOPERATORKA	UKR	1929	ABD
1196	BANATKA	HUN	1930	ABD
24071	BANATKA	POL	1930	ABD
24085	BANATKA	YUG	1930	ABD
24074	FERRUGINEUM 1239	UKR	1933	ABD
24078	SANDOMIRKA	UKR	1933	ABD
4301	LESOSTEPKA	UKR	1938	ABD
7344	VESELOPODOLYANSKAIA	UKR	1938	ABD
7345	VESELOPODOLYANSKAIA	UKR	1938	ABD
29964	ARTEMOVKA	UKR	1945	ABD
5437	ODESSKAYA 12	UKR	1947	ABD
24200	BANATKA	MDA	1951	ABD
5438	ODESSKAYA 16	UKR	1953	ABD
36085	VESELOPODOLIANSKAJA 499	UKR	1954	ABD
4864	MIRONOVSKAIA 264	UKR	1960	ABD
33970	BELOTSERKOVSKAJA-23	UKR	1962	ABD
4866	MIRONOVSKAYA 808	UKR	1963	ABD
4870	MIRONOVSKAIA 808 MUT.96-3-1	UKR	1963	ABD
4861	MIRONOVSKAIA 25	UKR	1980	ABD
13673	MIRONOVSKAIA 61	UKR	1989	ABD
6086	RENAN	FRA	1989	ABD
20074	MIRLEBEN	UKR	1993	ABD
	NARODNAYA			AB

Tableau VI-3 Liste des blés d'origine ukrainienne sélectionnés pour le crible PCR avec les amorces *Hélitron* 5Bq spécifique :

Les lignes jaunes et vertes sont les lignées positives pour l'insertion de l'*Hélitron* dans le pseudogène 5Bq. La ligne bleue montre la lignée Artemovka qui serait à l'origine de la variété Mironovskaya 808, mais qui ne possède pas l'insertion.

VI-4 Recherche de l'origine de l'insertion de l'*Hélitron* actif de Renan :

L'insertion de l'*Hélitron* actif du cultivar Renan dans le 5Bq est également trouvée dans le cultivar Mironovskaya 808, d'origine ukrainienne (Table S1 et Table S2). J'ai essayé de pousser l'investigation pour identifier les lignées ayant en commun une insertion identique au cultivar Renan. J'ai pu remonter les différents croisements qui ont servi à la création de la variété Renan (Fig. VI-2) et constater que Mironovskaya 808 fait parti des lignées à l'origine de la variété Renan. Pour retracer l'insertion dans d'autres blés je me suis intéressé à l'histoire du cultivar hexaploïde Mironovskaya 808 (Tableau VI-3 et annexe 6).

J'ai découvert que c'est une variété d'origine ukrainienne datant de l'ancien régime soviétique. J'ai décidé de collectionner certaines lignées ukrainiennes trouvées dans la lecture de documents mentionnant l'utilisation ou la création de la variété Mironovskaya 808. J'ai complété cette liste avec d'autres lignées inscrites, uniquement par rapport à leurs origines ukrainiennes communes pour retracer l'insertion de l'*Hélitron* dans un spectre plus large (voir Tableau VI-3 et Annexe 7).

VI-4-1 Un peu d'histoire sur les collections de blés soviétiques :

La recherche d'informations concernant les blés russes est compliquée. Entre 1898 et 1912, le gouvernement russe a lancé un programme de sélection agricole (Morgunov, 1992). La première étape de sélection de blé s'est arrêtée en 1930, les variétés sélectionnées ont été cultivées jusqu'en 1950. Mais entre-temps le pays a souffert de la première guerre mondiale, de la famine et de guerre civile. En 1917, le gouvernement soviétique prend le pouvoir, jusqu'en 1930 un scientifique, du nom de Nicolai Ivanovich Vavilov, enrichi une collection monumentale au sein de l'institut de recherche des plantes de Léninegrad en prospectant dans le monde entier. Mais en 1930 sous Staline, un autre scientifique du nom de Trofim Lysenko avec l'appui du gouvernement lance une campagne de rééducation du comportement des plantes. En effet il ne croit pas en la génétique et à la notion d'hérédité. Jusqu'en 1953 (à la mort de Staline), le seul processus de sélection consistait à mettre des plantes en culture au champ et d'attendre leurs adaptations à de nouvelles conditions. Suite à cette mauvaise expérience, les russes ont décidé de revenir à des méthodes scientifiques. Ils ont lancé de vastes programmes de croisement et de sélection avec les blés collectionnés par N. I. Vavilov. Le retour au travail de sélection génétique s'est déroulé sans lui, car il est décédé suite à son arrestation dans la prison de Saratov en 1940.

VI-4-2 Origine de la variété Mironovskaya 808 :

De tout les génotypes ayant permis la création de Renan, un seul possède l'insertion de l'*Hélitron*. Ce cultivar Mironovskaya 808 a un historique particulier car il valide la théorie de Lysenko, en effet la légende veut que Mironovskaya 808 est issu de « l'éducation » de la variété de printemps Artemovka.

En 1950, un sélectionneur, qui donnera son nom à un institut, Remeslo aurait semé à l'automne un champ de blé de printemps du nom d'Artemovka, après le passage de l'hiver toute les cultures auraient péri sauf une plante qui deviendra Mironovskaya 808, la variété de blé d'hiver la plus cultivée en Russie soit 7 millions ha de 1963 à 1968 (Dorofeev et al., 1987).

Est-ce que l'insertion de l'*Héliron* a eu lieu dans cette plante ou est ce que cette plante provenait d'un grain ayant contaminé le lot de semis de la variété Artemovka ? La majorité de la communauté scientifique pensent que c'est le résultat d'une fécondation croisée avec une variété de blé d'hiver.

Entre 1930 et 1980, il existe plus de deux cents publications, aux résultats contradictoires sur le renouvellement de cette expérience (Rajki, 1985). Certains scientifiques essaient encore aujourd'hui de valider le processus. Parmi les expériences, une en faveur d'une adaptation possible du blé de printemps en blé d'hiver affirme qu'il existe trois gènes importants pour le processus de vernalisation, VRN1 pour les blés de printemps, VRN2 de la famille des Zinc fingers CCT pour les blés d'hiver et VRN3 qui contrôlerait l'expression des deux autres. Le passage du blé d'hiver au blé de printemps et *vice versa* serait lié à l'état de méthylation de l'ADN, autrement la vernalisation serait sous contrôle de phénomènes épigénétiques, ce qui expliquerait le passage de la variété Artemovka à la variété Mironovskaya 808 (Li and Liu, 2010 ; Kippes, 2014).

Les résultats que j'ai obtenus avec le marqueur spécifique de l'insertion de l'*Hélitron* de Renan montrent que les lignées qui possèdent l'insertion (Tableau IIX-4) sont :

- ✓ Banatka-RUS (1924)
- ✓ Ukrainka (1929), Kooperatorka (1929)
- ✓ Lesostepka (1938)
- ✓ Mironovskaya 808 (1963)
- ✓ Mironovskaya 25 (1980)
- ✓ Mironovskaya 61 (1989)
- ✓ Renan (1989)

Artemovka mis en catalogue en 1945, qui est censé être à l'origine de Mironovskaya 808 (1963) ne possède pas l'insertion de l'*Hélitron*, mais on trouve d'autres variétés plus anciennes qui la possède. Les recherches que j'ai faites sur l'origine de ces différentes variétés (Tableau VI-3 et Fig. VI-4) montrent que le cultivar Mironovskaya 808 aurait hérité de séquences provenant soit de la variété Ukrainka, Kooperatorka ou Lesostepka inscrites à la même époque (1929). On retrouve l'insertion également dans une variété plus ancienne du nom de Banatka russe qui peut avoir servi de nombreuses fois pour la création de variétés (Ukrainka, Kooperatorka, Lesostepka, Mironovskaya 808, Mironovskaya 25, Mironovskaya 61 et Renan) à des époques différentes, ou que chacune de ces variétés a été utilisée pour la création de la suivante jusqu'à l'aboutissement de la variété Renan si on suit la chronologie des inscriptions en catalogue.

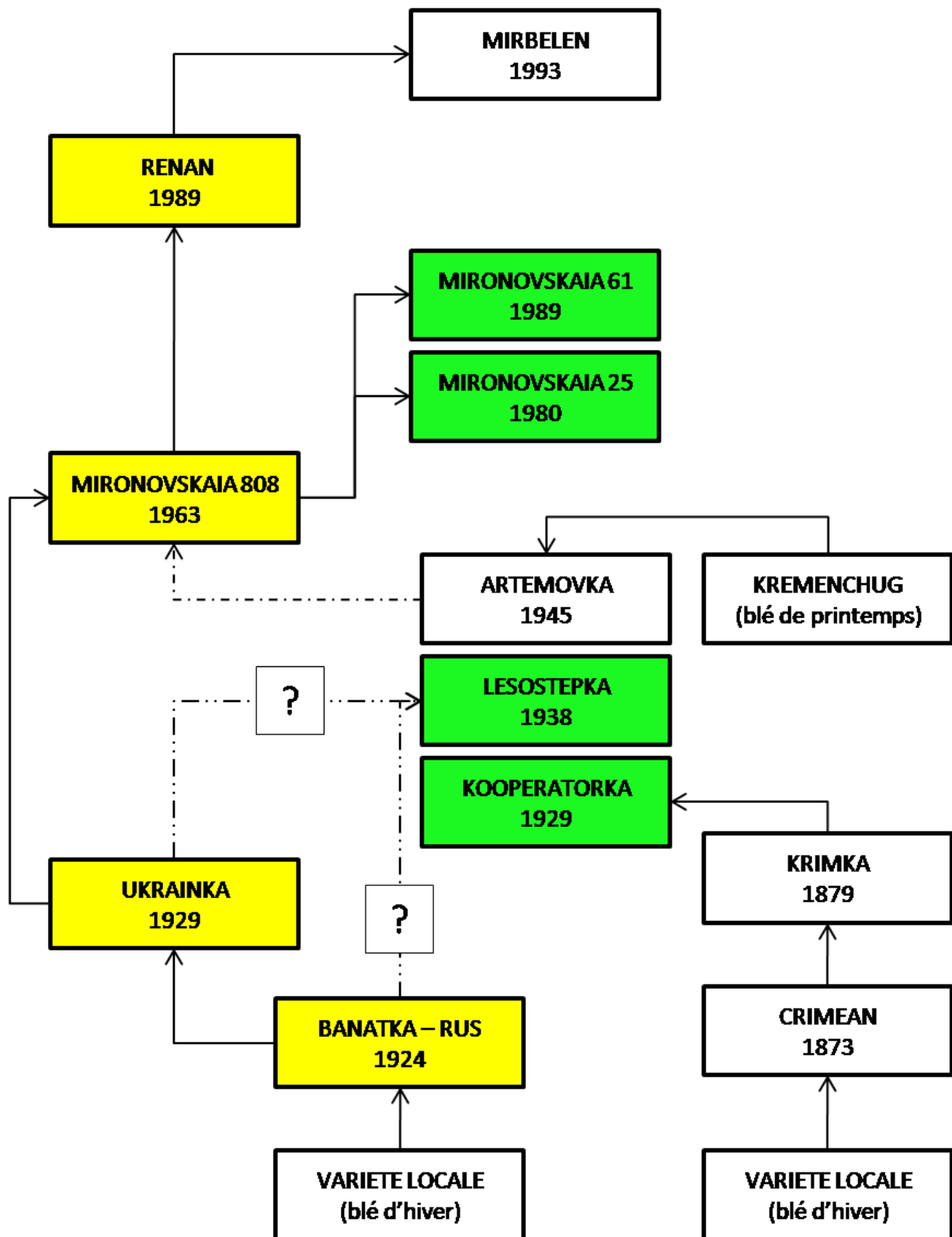


Fig. VI-4 Etapes possibles du transfert de l'*Hélitron* de Renan depuis le blé Banatka Russe (1924) jusqu'à Renan (1989) :

- en jaune les génotypes qui portent l'insertion de l'*Hélitron* et qui sont à l'origine de Renan
- en vert les génotypes qui portent l'insertion à l'occasion d'autres croisements intermédiaires.

VI-5 Conclusion sur l'origine de l'*Hélitron* dans *5Bq* :

Les marqueurs issus de séquences d'*Hélitrons* sont des outils suffisamment robustes pour suivre le transfert de la région d'une variété à une autre lors de croisement. Mes travaux montrent qu'il est fort possible que la variété Mironovskaya 808 ait profité d'un transfert de séquence issu de la variété Banatka Russe directement ou indirectement par l'intermédiaire d'autres variétés issues elles-mêmes de la variété Banatka Russe (Fig. VI-4 et photo en annexe 8). Ceci signifie que contrairement aux archives laissées, la rééducation de la variété Artemovka (blé de printemps) n'est pas le mécanisme de sélection pour l'obtention de la variété Mironovskaya 808 (blé d'hivers). Il est plus que probable, comme l'ont émis certains scientifiques, que la « rééducation » est bénéficiée d'un transfert par hybridation de la variété Banatka Russe ou d'une autre variété issue d'un croisement avec elle (Ukrainka, Kooperatorka, Lesostepka).

Ces travaux retracent l'insertion de l'*Hélitron* dans le *5Bq* à moins de 90 ans, ce qui est relativement récent. Les variétés russes ainsi qu'*Ae. Ventricosa* ont servi dans la généalogie de Renan ; il est possible que son origine soit dans les espèces apparentées (types *Ae. Ventricosa*) et que l'insertion dans le *5Bq* a eu lieu lors des croisements interspécifiques à l'origine de la variété Banatka.

Cet *Hélitron* représente le seul élément complet et actif identifié dans le blé. Les conditions de son activation restent encore à déterminer.

CHAPITRE VII : REGULATION DU GENE MAJEUR DE LA DOMESTICATION

VII-1 La question qui demeure :

Plusieurs questions demeurent sur la régulation des gènes *Q/q*. En effet, si la mutation V329I est responsable de l'hyperfonction du gène 5AQ, pourquoi la quantité de ses transcrits est beaucoup plus élevée que celle du gène 5Aq et des autres homéologues (5Bq et 5Dq). Aussi comment le pseudogène 5Bq continue à contribuer aux caractères de domestication alors qu'il ne code plus pour une protéine et comment les trois homéologues se régulent entre eux dans le blé hexaploïde ?

Il est à signaler que Muramatsu (1962), et sans avoir fait des caractérisations moléculaires, émet la conclusion d'un effet dose indépendant de la mutation : « il faut l'expression de deux gènes et demi *q* (sauvage) pour obtenir les effets d'un gène *Q* (domestiqué) et obtenir un épi «squarehead» domestiqué ». Comment la mutation d'un simple acide aminé peu démultiplier l'activité d'une protéine ? Existe-t-il d'autres facteurs ou d'autres mécanismes impliqués dans l'expression du gène *Q* ?

Simons et al. (2006) ont identifié une mutation dans la séquence du 3'UTR, mais ils ne l'ont pas identifiée comme le site d'accrochage du miRNA172 (position 3531 sur la séquence nucléotidique). La mutation est présente dans tous les allèles 5AQ et absente des allèles 5Aq comparés. Les auteurs ont omis que cette mutation dans le site d'adressage du miRNA172 peut jouer un rôle en perturbant la régulation de l'expression du gène par le miRNA172 et ce sans avoir besoin de changement d'acide aminé. Ce type de perturbation de la régulation a été observé pour le gène *cly1* impliqué dans le caractère de cléistogamie chez l'orge. Le gène appartient également à la famille des AP2 (Nair et al. 2010).

Pour valider cette hypothèse il m'a fallu étudier les mécanismes de régulation des ARNm par les miRNA. Avant de développer la méthode utilisée et les résultats obtenus, je propose un bilan des connaissances actuelles sur les petits ARN qui m'ont permis de tester cette hypothèse.

VII-2 Les petits ARN :

Depuis la découverte de l'ARN par Friedrich Miesher en 1869, nous progressons à grand pas dans la compréhension de l'organisation du vivant. Si je résume les découvertes faites, il y a plus d'un siècle à nos jours, je dirais qu'il y a quatre grandes périodes parfois chevauchantes. La première est la caractérisation biochimique des constituants et de la structure de l'ADN et de l'ARN (1869-1965), la deuxième la découverte des relations ADN, ARN et protéine et les mécanismes enzymatiques (1950-1965), la troisième la découverte des ARNs non codant (le premier étant l'ARN de transfert (tRNA) (1954-1965), et la quatrième l'étude des mécanismes de régulation de l'expression (1965 à nos jours). Actuellement il existe un grand nombre de publications traitant des petits ARNs, il est parfois assez compliqué pour un non averti de tout synthétiser. J'ai répertorié cinq processus où ils interviennent :

- ✓ Dans l'épissage des pre-ARNm (snoRNA du spliceosome)
- ✓ Dans la traduction des ARNm (ARNr du ribosome et ARN de transfert (ARNt))
- ✓ Dans la régulation des ARNm (siRNA et miRNA)
- ✓ Dans un mécanisme de défense contre des ARN exogènes (Virus)

Je ne parlerai pas de l'épissage et de la traduction des ARNm, car ils n'interviennent pas dans mon hypothèse sur la régulation des gènes *Q/q*.

VII-2-1 Biogénèse, mécanisme et fonction des petits ARNs (siRNA et miRNA) :

Les premiers petits ARNs identifiés interviennent dans la régulation des ARNm, par un phénomène d'ARN interférence (RNAi). Historiquement le phénomène d'ARN interférence a été découvert par Rochard Jorgensen dans les années 1990 lors de ses recherches sur les mécanismes de coloration de Pétunia. Dans le but de renforcer la coloration des fleurs, les auteurs ont transformé la plante avec une copie d'un gène intervenant sur la pigmentation. Mais ils ont obtenus des fleurs blanches (Napoli et al., 1990). Cette expérience de co-suppression a permis de mettre en évidence un mécanisme très répandu chez les plantes le PTGS (Post Transcriptional Gene Silencing).

Ce mécanisme intervient dans la protection du génome contre les virus et les éléments mobiles à ARN (Fagard et al., 2000). Par la suite, il a été découvert que l'introduction d'un ARN double-brin (ARNdb) dans les cellules de *C. elegans* pouvait réduire spécifiquement la traduction des protéines en se liant à leurs ARNm, jusqu'à plusieurs générations successives. Cela a permis de comprendre la fonction de nombreuses protéines (Fire et al., 1998). Chez les animaux, ce phénomène est appelé ARN interférence ou siRNA (short interference RNA). Le phénomène siRNA est très ressemblant au phénomène PTGS observé chez les plantes.

L'ARN interférence (RNAi) est un terme qui se réfère à une mise sous silence, « silencing », génétique post-transcriptionnel induit soit par l'arrêt de la traduction, soit par la dégradation de l'ARNm cible (Werk and al., 2005).

Chez les plantes, l'introduction d'un ARNdb induit la méthylation des séquences génomiques homologues, qui a pour conséquence l'inactivation de la transcription. Chez les ciliés il existe une voie qui va jusqu'à l'élimination de l'ADN. La fonction naturelle de l'interférence chez les plantes est la protection du génome contre l'invasion par des éléments génétiques mobiles comme les transposons et les virus qui produisent de l'ARNdb quand ils deviennent actifs (Malinsky et al., 2000). Chez les mammifères, il a fallu attendre un peu pour transposer le modèle, car l'injection d'un long ARNdb conduit à la production d'interférons, ceux-ci bloquent toutes les synthèses protéiques. Mais ils ont réussi à identifier le même système chez les animaux, le processus est relayé par des petits fragments de 21-22 nucléotides qui correspondent aux produits de clivage de grands ARNdb (Elbashir et al., 2001), appelés siRNA.

Les différentes actions impliquant les petits ARNs :

Depuis ces premières découvertes, un grand nombre de processus impliquant les petits ARN ont été découverts. Ils sont actifs pour :

- ✓ La répression de la traduction
- ✓ Le clivage des ARNm et la répression de la traduction
- ✓ L'inhibition de la transcription via la méthylation de l'ADN
- ✓ L'élimination de l'ADN

En reprenant une des classifications proposées, on distingue deux catégories de petits ARNs, ceux qui ont pour précurseur des ARNs simple brin formant des structures tiges-boucles (Hairpin RNA, hpRNA), c'est la voie des miRNAs (Axtell, 2013). La deuxième catégorie concerne les siRNA, les petits ARNs ayant pour précurseur des ARNdb (double-strand RNA, dsRNA) d'origines diverses que je détaillerai par la suite.

VII-2-1-1 Les siRNA :

Les siRNA (short interfering RNA) représentent des ARNdb de 21-22 nucléotides générés à partir de long double-brin d'ARN (dsRNA) provenant du génome ou d'ADN exogène (virus, bactérie, transformation). Ce sont ces origines diverses qui sont la cause d'une nomenclature complexe, on trouve :

- ✓ Les tasiRNA (trans-actin short-interfering RNA) de 21-22 nt, dont le mode d'action se finalise par le clivage des ARNm. Le précurseur est un gène transcrit sur les deux brins pour générer un ARNdb complémentaire (cis) ou des gènes de régions éloignées ayant une partie commune et complémentaire. Cette voie permet également la synthèse du brin complémentaire d'un ARNm cible par le complexe tasiRNA et RdRP (RNA dependent RNA polymerase) pour amplifier la réponse de dégradation.
- ✓ Les rasiRNA (repeat-associated short-interfering RNA) de 24-27 nt, jouant un rôle dans la régulation de la transcription des gènes, via la méthylation des Histones et de l'ADN. Ce processus peut-être à l'origine de l'inactivation des TEs.
- ✓ Les scnRNA (small scan RNA) de 28 nt, leurs existences n'ont été prouvées que chez les ciliates, ils participent au processus d'élimination de l'ADN par la méthylation des histones dans le nouveau macronoyau formé.
- ✓ Les tncRNA (tiny non-coding RNA) représentent une nouvelle classe découverte chez *C. elegans*, mais leurs rôles restent inconnus.
- ✓ Les smRNA (small regulatory RNA) découverts en 2004 chez la souris, ils seraient impliqués dans la régulation de l'expression de gènes spécifiques dans les neurones (Novina and Sharp, 2004).

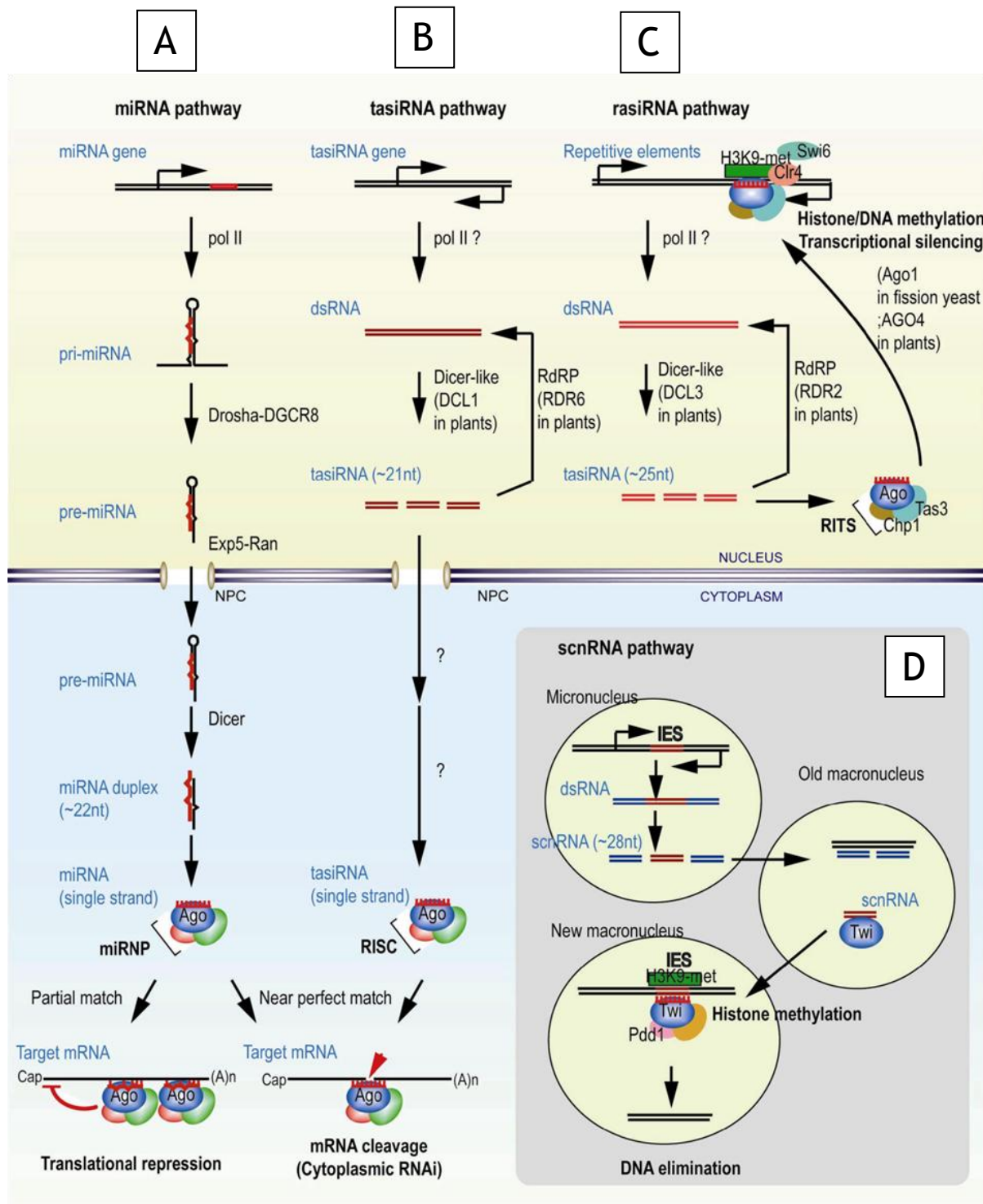


Fig. VII-1 Les différentes voies d'action des petits ARN interférents :

A - Voie des miRNA → blocage de la traduction ou clivage des ARNm

B - Voie des rasiRNA → clivage des ARNm

C - Voie des rasiRNA → méthylation de l'ADN

D - Voie des scnRNA → élimination de l'ADN

(Kim, 2005)

Mode d'action des siRNA :

Je ne parlerai que du mécanisme des tasiRNA et des rasiRNA qui a été validé à multiples reprises chez de nombreuses espèces, et du mécanisme des scnRNA chez les ciliés.

Voie des tasiRNA (voir Fig. VII-1B): Un long brin d'ARN double-brin (ARNdb) provenant d'un gène codant, d'un virus, d'un transposon ou d'un transgène est clivé par une enzyme appelée Dicer (RNAses de type III) en tasiRNA. Le brin sens du tasiRNA est dégradé tandis que le brin antisens est utilisé pour cibler l'ARNm pour la mise sous silence.

On distingue deux voies, la première celle de la drosophile et des mammifères où les brins antisens sont directement intégrés dans un complexe RISC («RNA induced silencing complex») pour cibler la séquence complémentaire de l'ARNm et pour la détruire, la deuxième celle de *C. elegans* et des plantes, une partie des brins antisens suit la première voie, une autre partie va se lier à une enzyme appelée RdRP (RNA dependent RNA polymerase). Le complexe s'apparie à un ARNm qui lui sert de matrice pour la synthèse d'ARN complémentaire. L'ARNdb nouvellement synthétisé entre dans le processus ARN interférence (RNAi) et permet une réponse amplifiée. Le résultat est pour les deux voies une destruction de la cible (Novina and Sharp, 2004) et un blocage de la traduction.

De plus on distingue deux classes de siRNA, les symétriques et les asymétriques. Les siRNA symétriques ont les deux parties terminales stables et identiques tandis que les asymétriques ont une partie terminale moins stable que l'autre. Cette différence entre les deux classes de siRNA, joue un rôle dans l'efficacité de formation du complexe RISC, donc de la régulation.

Voie des rasiRNA (voir Fig. VII-1C) :

Un brin long d'ARN double-brin (ARNdb) provenant par exemple d'un élément transposable est clivé par une enzyme appelée Dicer (RNAses de type III) en rasiRNA. Le brin sens du rasiRNA est dégradé tandis que le brin antisens est directement intégré dans un complexe RITS (« RNA-induced initiation of transcriptional silencing complex ») pour méthyliser l'ADN et les histones cibles, et ainsi bloquer la transcription.

Voie des scnRNA (voir Fig. VII-1D) :

Chez les ciliés, le génome du micronoyau est transcrit dans les deux sens, ce qui génère des ARNs double-brin (ARNdb). Ils sont clivés par une enzyme appelée Dicer-like (RNAses de type III) en scnRNA. Les scnRNA vont dans l'ancien macronoyau trier et reconnaître les séquences à éliminer, puis ils sont dirigés dans le nouveau macronoyau, ils incorporent un complexe RITS pour repérer les séquences d'ADN et les histones à méthyliser. Les histones méthylées permettent le recrutement des protéines nécessaires à l'élimination de l'ADN.

VII-2-1-2 Les miRNA :

Rôle des miRNA :

Les miRNA sont une classe d'ARN simple brin de 19-25 nucléotides, ils ont été mis en évidence dans le règne animal et végétal. Leur nombre est très variable dans les cellules, quelques centaines à des dizaines de milliers (40 000) (Novina and Sharp, 2004). Ils sont codés par le génome de la plupart des organismes multicellulaires mais ne produisent pas de protéines, ils font donc partie des gènes à ARN au même titre que les ARNr et ARNt. Ils sont régulés pendant le développement et interviennent dans la régulation de la synthèse protéique. Leur action va du blocage de la traduction au clivage de l'ARNm. Les miRNA représentent entre 0.5 et 1% du génome humain.

Les miRNA sont formés à partir d'un ARN en épingle à cheveux, ce qui veut dire qu'ils ont dans leur biogénèse une étape ARNdb après le clivage par l'enzyme Dicer.

Voie des miRNA (voir Fig. VII-1A) :

Au départ il y a la transcription d'un gène MIR (gène à ARN) issu d'une région intergénique (IGR) par l'ARN polymérase II, le transcrit est un pri-miRNA (primary micro RNA en épingle à cheveux).

Ce pri-miRNA est clivé par une enzyme nucléaire appelé Drosha en pre-miRNA (ARNdb de 70 nucléotides). Chaque pre-miRNA forme avec un Dicer et des cofacteurs (DGCR8/Pasha) un complexe appelé «microprocessor complex» de 500-650 KDa (Schwarz et al., 2003). Ce complexe est exporté du noyau vers le cytoplasme par l'exportine 5 (Exp5) ou un Dicer et une hélicase vont cliver les pre-miRNA et séparer les brins pour donner des miRNA matures simple-brin de 21-22 nucléotides. Le miRNA forme un complexe ribonucléoprotéique avec AGO (miRNP). Le complexe ainsi formé peut à son tour fixer l'ARNm pour bloquer la traduction ou le cliver en 3', c'est-à-dire l'inactiver (Bernstein et al., 2003).

Il existe deux RNase de type III essentielles dans ce système : Dicer et Drosha. Ce sont des protéines de haut poids moléculaire avec des domaines catalytiques en tandem et un domaine de liaison à l'ARN dans la partie C terminale. Dicer est une protéine très conservée, on trouve des homologues chez la levure (Dcr), chez l'homme (Dicer), chez la Drosophile (DCR1 et DCR2) et quatre Dicer-like chez Arabidopsis (DCL1, DCL2, DCL3 et DCL4) (Kim, 2005) qui interviennent dans des voies différentes.

	miRNA/target pairing pattern	Prevalence	Mechanisms of target repression
A	Canonical plant pairing	Common in plants; very rare in animals	Slicing-dependent (and perhaps slicing-independent) reduction in mRNA accumulation, and/or translational repression
B	Seed pairing (canonical animal pairing)	No evidence in plants to date; common in animals	Translational repression followed by slicing-independent reduction in mRNA accumulation

Fig. VII-2 : Les zones d'hybridation préférentielles du miRNA sur le site d'adressage de l'ARNm pour les plantes et les animaux :

A – les plantes (12 nucléotides)

B – les animaux (6 nucléotides)

(Axtell, 2013)

Particularité des miRNA :

Il y a une particularité chez les miRNA, quand le brin antisens se lie à sa cible, la complémentarité de séquence peut varier. Dans le cas d'un appariement parfait, l'ARNm cible est clivé entre les nucléotides 10 et 11 de la partie 5' terminale du miRNA. Dans le cas d'une complémentarité très basse, la réponse est une répression de la transcription. Pour l'inhibition de la traduction, il peut y avoir plusieurs sites reconnus sur l'ARNm. De plus on note une différence entre les plantes et les animaux pour la région d'appariement du miRNA sur l'ARNm ciblé. En effet pour les animaux la région d'hybridation est comprise entre les nucléotides 1 et 8 (Fig. VII-2A), tandis que pour les plantes la région d'hybridation est comprise entre 1 et 14 nucléotides.

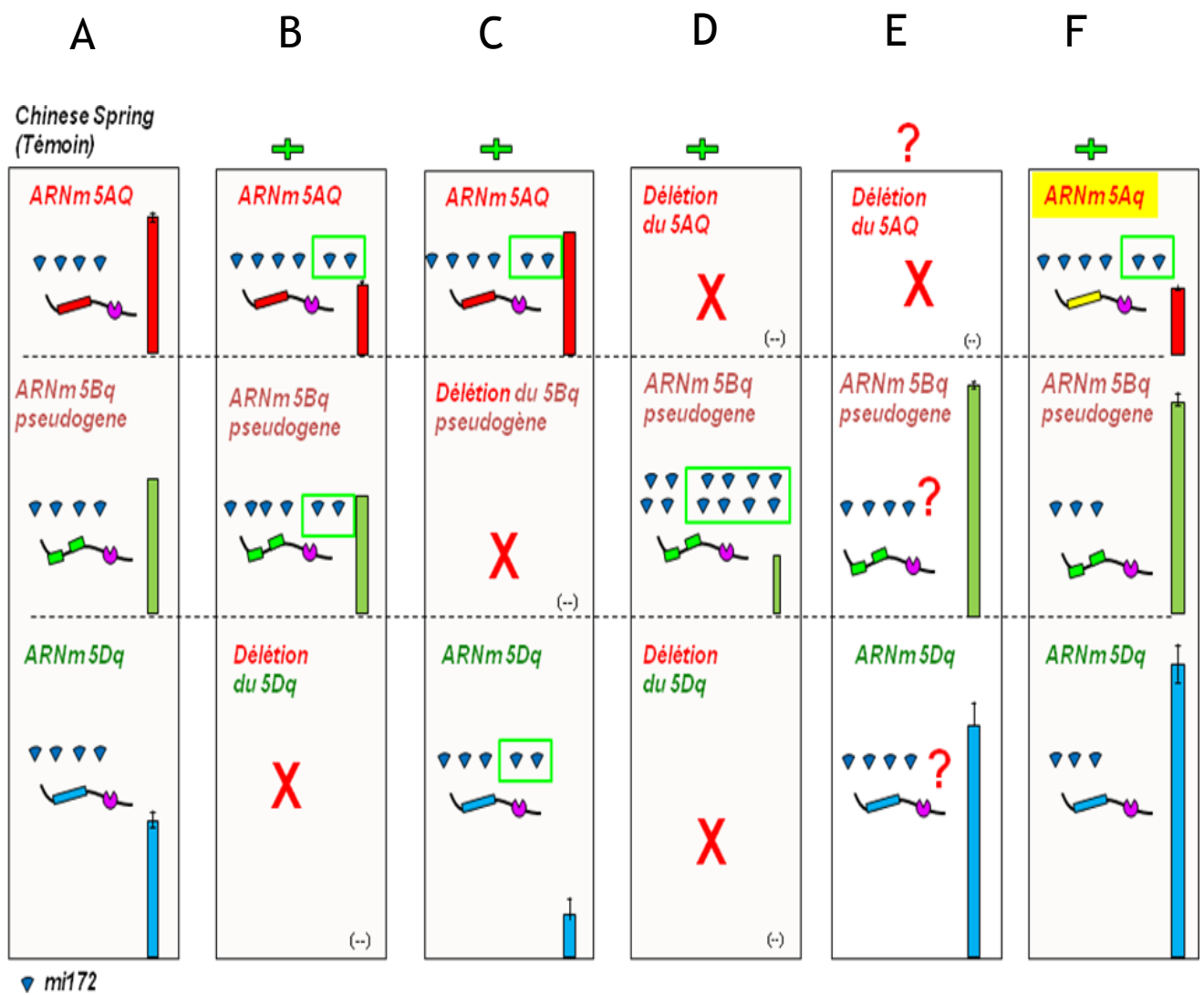


Fig. VII-3 : Interactions et «Régulation-Croisée» des homéologues 5AQ/q, 5Bq et 5Dq (ceRNA) via une population de mi172 suivant l'hypothèse de Salmena et al:

Cette interprétation permet de superposer l'hypothèse d'un réseau de régulation et d'une compétition entre différents transcrits possédant un site d'adressage ou MRE (microRNA Response Element) pour fixer une population de microARN et les résultats obtenus lors de l'étude de l'expression des différents homéologues dans les lignées de délétion ou de substitution.

A est le témoin Chinese Spring, les trois homéologues s'expriment y compris le pseudogène 5Bq.

B, C et D montrent que l'absence d'un homéologue (ou deux) conduit à une diminution des homéologues restant, probablement due à une plus grande disponibilité de mi172.

E – montre que la substitution de l'allèle 5AQ domestiqué par l'allèle 5Aq sauvage non muté dans le site d'adressage modifie l'équilibre d'utilisation de la population de mi172, par rapport au témoin Chinese Spring.

F – montre que l'absence de l'homéologue 5AQ muté dans le site d'adressage, n'est pas ou peu concerné par l'hypothèse de compétition entre homéologues. On parle sur un modèle classique où uniquement la délétion d'un des homéologues restant (5Bq ou 5Dq) peut influencer l'expression de l'autre.

VII-3 Interactions et «Régulation-Croisée» des homéologues 5AQ/q, 5Bq et 5Dq (ceRNA) via une population de mi172 :

Pour interpréter les résultats obtenus lors de l'étude de l'expression des différents homéologues dans les lignées de délétion ou de substitution, je me suis inspiré d'un modèle décrivant un réseau d'interaction et de régulation de l'expression des gènes par l'intermédiaire des transcrits entre-eux ayant en commun un site d'adressage pour un microARN. Les transcrits agissent en compétition pour fixer une population de microARN.

Si on élargi le modèle de Salmena au différents homéologues 5AQ (domestiqué), 5Aq (sauvage), le pseudogène 5Bq et l'homéologue 5Dq ayant en commun le site d'adressage dans la région 3'UTR pour le mi172.

On peut interpréter les résultats montrant une diminution des autres homéologues en l'absence d'un homéologue ainsi, pour une population de mi172, l'absence d'un homéologue dans le blé hexaploïde rend disponible une plus grande quantité de mi172. Cette quantité supplémentaire va augmenter l'efficacité de fixation des mi172 dans les sites d'adressage (ou MRE, pour microRNA Response Element) et donc augmenter le nombre de transcrits clivé (diminution de l'expression) des homéologues restant. Cette interprétation est valable quand les homéologues 5Bq ou 5Dq sont délétés et pour la double délétion 5AQ / 5Dq (Fig. VII-3, B, C et D).

Mais les résultats en l'absence de l'homéologue 5AQ (domestiqué) ou sa substitution par l'allèle 5Aq (sauvage) ne suivent pas la théorie de Salmena sur une compétition endogène pour les mi172.

Je me suis intéressé à ce résultat, et une des interprétations possible est une différence de régulation entre les deux homéologues peut-être lié à une autre mutation.

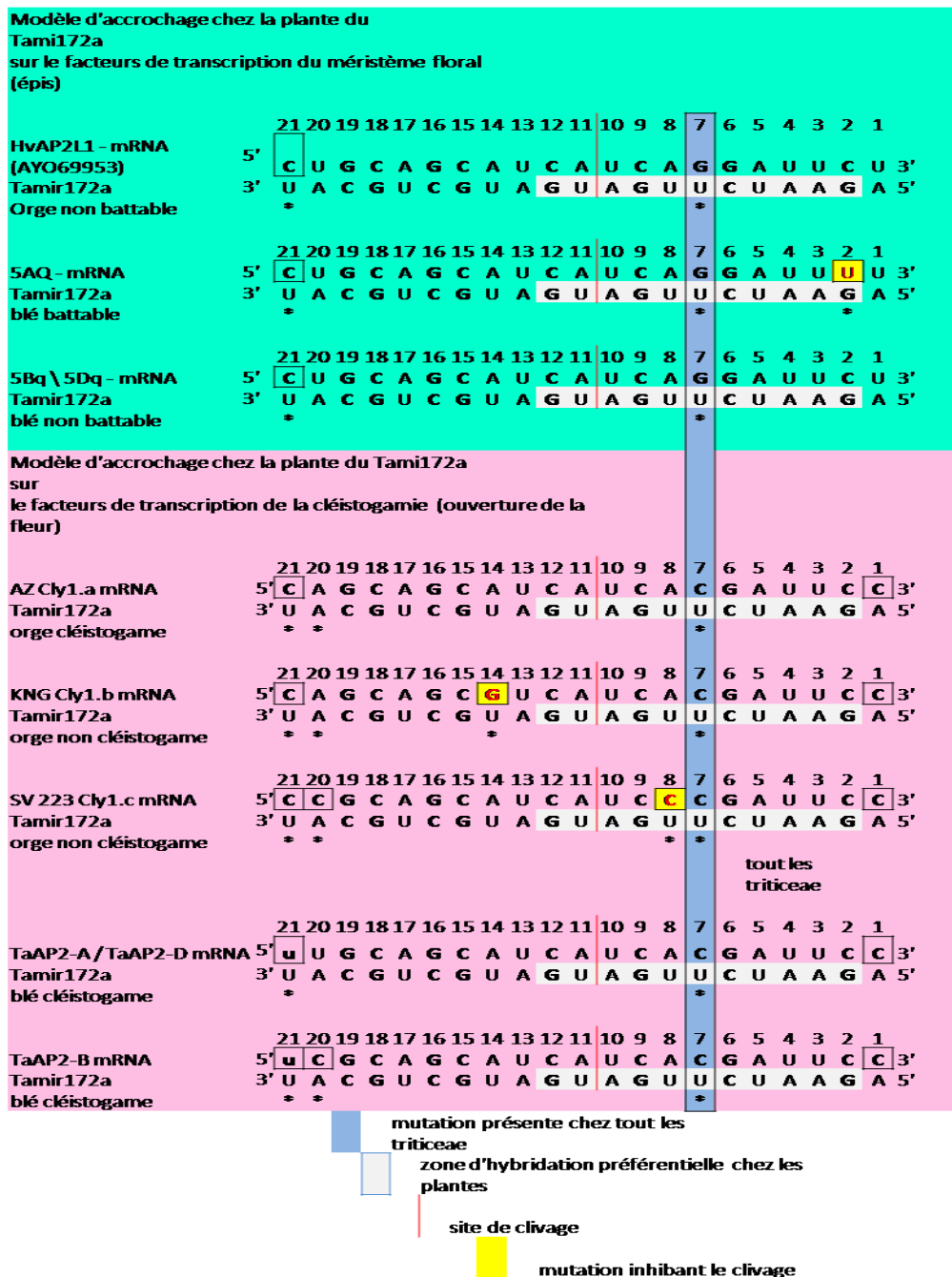


Fig. VII-3 Analyses de l'hybridation du Tami172 sur différents facteurs de transcriptions :

Dans la zone verte les facteurs de transcriptions ayant une action sur le méristème floral (formation de l'épi), le gène *Q/q* du blé et le gène *HvAP2-like* de l'orge. Dans la zone rose les facteurs de transcriptions ayant une action sur la cléistogamie (ouverture de la fleur), les gènes *Cly1/cly1* de l'orge et les gènes *TaAP2* du blé. Il existe deux mutations différentes inhibant le clivage des ARNm. Les mésappariements observés sont pour l'un hors du site d'accrochage préférentiel à la 14ème position (G-U) en 5' dans le site d'adressage et pour l'autre à la 8ème position (C-U). Il existe d'autres mésappariements entre le Tami172a et les différents sites d'adressage en première et en dernière position, il est probable que leurs présences soient tolérées (Nair et al., 2010).

VII-3 Régulation des transcrits des allèles *Q* et *q* du chromosome 5A du blé :

Entre les séquences des allèles 5AQ et les séquences des allèles 5Aq, il y a 6 SNPs identifiés. Les SNPs dans l'intron 1 (position 576) et dans l'intron 9 (2452) du gène, ne modifient pas la protéine traduite. Deux SNPs, correspondent à des substitutions synonymes, dans les exons 2 (617), 8 (2189) et sont sans conséquences sur la protéine traduite. Dans l'exon 9 (2123), il y a une substitution non-synonyme qui correspond à la mutation V329I dans la protéine associée au phénotype domestiqué (Simons et al., 2006). Le dernier SNP se situe dans le 3'UTR du transcrit (position 3531), l'UTR n'étant pas traduit, il n'y a aucune incidence sur la protéine, mais il est dans le site d'adressage du miR172.

En reprenant les alignements des séquences entre plusieurs allèles *Q* et *q* sur le chromosome 5A, les SNPs en position 2123 (V329I) et 3531 sont toujours associées. Sans écarter l'implication de la mutation 2123 (V329I) (Simons et al., 2006) dans la protéine traduite de l'allèle *Q*, on peut émettre l'hypothèse que le SNP à la position 3531 du gène joue un rôle dans la régulation des transcrits. Car cette mutation dans le site d'adressage du miR172 pourrait modifier le clivage des ARNm de l'allèle 5AQ.

De plus l'alignement des séquences disponibles des ARNm, produits des allèles *Q* et *q*, y compris les homéologues *q* des chromosomes 5B et 5D, montre que la mutation n'est présente que dans l'allèle 5AQ (Fig. VII-3). Pour vérifier si la mutation dans le site d'adressage du miR172 joue un rôle dans la régulation de l'allèle 5AQ et conduit au moindre clivage du transcrit ; j'ai entrepris de comparer la dégradation des transcrits des gènes 5AQ (CS) et 5Aq (CS-DIC5A) par la technique RACE (Rapid Amplification of cDNA Ends) suivi d'une PCR semi-quantitative.

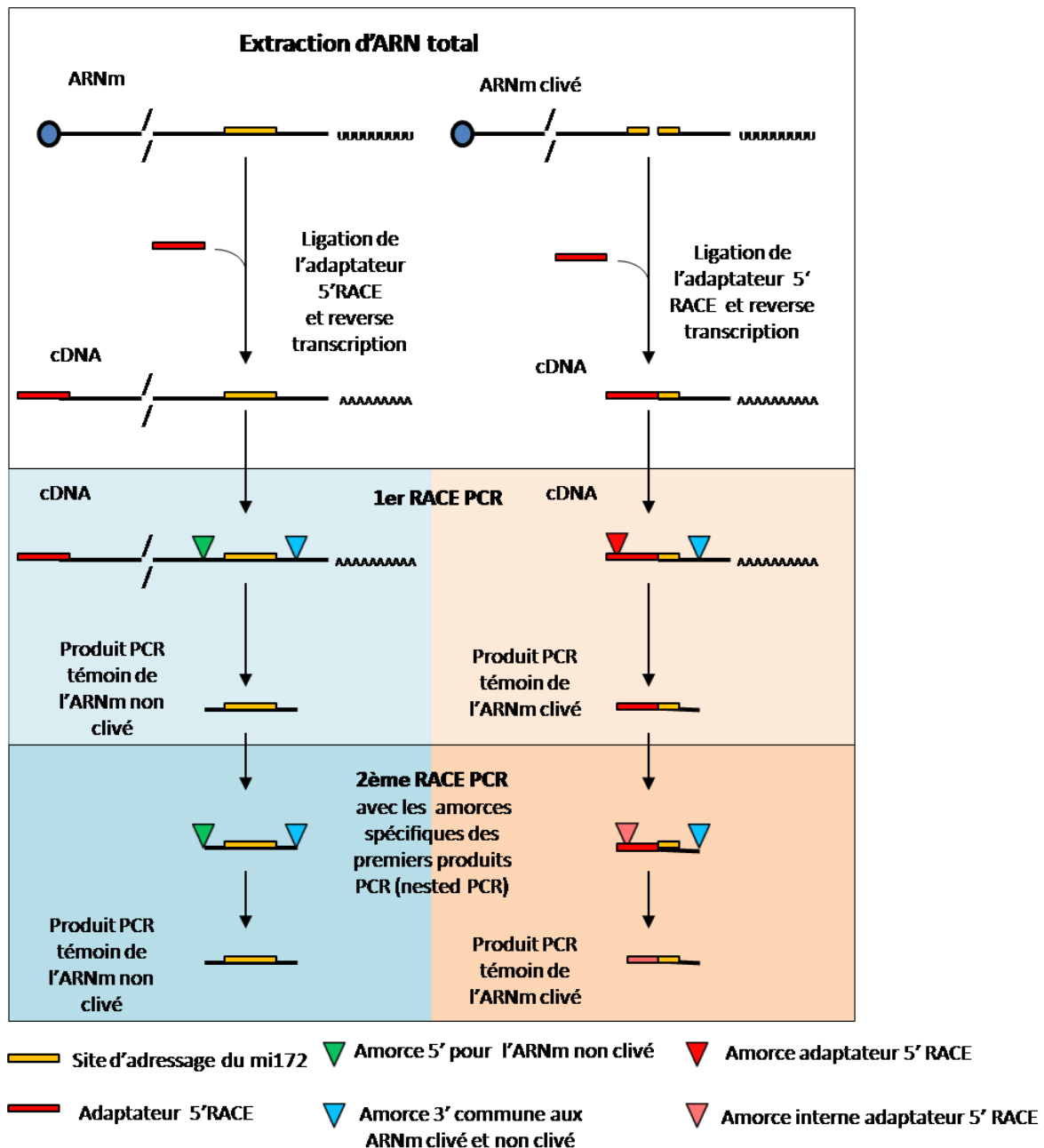


Fig. VII-4 Technique RACE-PCR :

1. Extraction de l'ARN total et réverse transcription en ADNc
2. Ligation de l'adaptateur pour la PCR
3. PCR avec les amorces spécifiques de l'adaptateur et du gène *Q/q*
4. PCR de confirmation avec des amorces internes aux premiers produits PCR

VII-4 Matériel et méthodes :

VII-4-1 Principe de la technique RACE PCR :

La technique RACE (Rapid Amplification of cDNA Ends) PCR (Fig. VII-4) consiste à liguer des adaptateurs (5'RACE) en 5' des ARNm, puis de synthétiser des cDNA par une transcription inverse. Pour analyser les ARNm clivés à un site donné (exemple le site d'adressage du miR172) par rapport aux ARNm non clivés, on peut comparer les produits PCR obtenus par exemple avec : (i) un couple d'amorces spécifiques de l'ARNm non clivé (encadrant le site de clivage) et (ii) un couple d'amorces constitué d'une amorce spécifique de l'adaptateur RACE ligué en 5' et d'une autre amorce en 3' du site de clivage étudié. Ce dernier couple d'amorce devrait en principe amplifier les différents ARNm du gène. On reconnaîtra la forme clivée en fonction de la taille attendue. Aussi une deuxième PCR-interne aux premiers produits PCR ou « nested » pourrait confirmer la spécificité.

VII-4-1-1 Dessins des amorces PCR :

Des couples d'amorces spécifiques encadrant le site d'adressage du miR172 ont été dessinés sur la séquence du gène 5AQ de *T. aestivum* ssp. *aestivum* cv. Chinese Spring à l'aide du logiciel OLIGO7[®] (OLIGO Primer Analysis Software Version 7). Les paramètres retenus pour le dessin de chaque amorce est une température d'hybridation comprise entre 55-60°C, pour un produit PCR compris entre 200 et 400pb. La spécificité des amorces a été obtenue en concaténant les séquences des différents homéoallèles et en ciblant une fenêtre autour du site d'adressage de l'homéologue 5AQ (Annexe 4 et 5).

VII-4-1-2 Extraction des ARNs totaux :

L'extraction des ARNs totaux des génotypes supplémentaires inclus dans cette analyse complémentaire a été réalisée selon les protocoles décrit dans le papier Zhang et al. (2011).

VII-4-1-3 Utilisation du kit GeneRacer[™] RLM-RACE (RNA ligase-mediated Rapid Amplification of cDNA Ends) selon le protocole INVITROGEN[®] :

A- Décoiffage des ARNm :

Mélanger 7µl d'ARN extrait, 1µl de 10X TAP (Tobacco Acid Pyrophosphatase) buffer, 1µl de RNase Out (40u/µl) et 1µl de TAP (0,5 u/µl), incuber 1heure à 37°C, mettre sur glace, ajouter 90µl d'eau RNase free (SIGMA) et 100µl de phénol/chloroforme, agiter 30s et centrifuger à 10 000g/5min à température ambiante. Récupérer 100µl, ajouter 2µl de glycogène Mussel (10mg/ml), 10µl d'acétate de sodium 3M à pH 5,2 (SIGMA) et 220 µl d'éthanol 96°. Laisser 10min sur glace, puis centrifuger à

10 000g / 20min / 8°C. Laver à l'éthanol 75°, sécher et reprendre le culot dans 8µl d'eau avec du DEPC (diethyl pyrocarbamate) fourni avec le kit.

B- Ligation de l'adaptateur 5'RNA Oligo gene RACER :

Mettre les 8µl d'ARN décoiffés dans un tube contenant l'adaptateur 5'Gene RACER Oligo RNA, incuber à 65°C / 5min, mettre sur glace 2min et ajouter 1µl de ligase buffer 10X, 1µl de mix ATP (10mM), 1µl d'RNase Out (40u/µl), 1µl de T4 RNA ligase (5u/µl). Incuber à 37°C/1 heure, centrifuger à 10 000g /5min, ajouter 90µl d'eau RNase free + 100µl de phénol/chloroforme/isoamyl alcohol (25:24:1), agiter 30s et centrifuger à 10 000g/5min à température ambiante. Récupérer 100µl, ajouter 10µl d'acétate de sodium 3M à pH 5,2 (SIGMA) et 220 µl d'éthanol 96° (CARLO ERBA). Centrifuger à 10 000g/5min à température ambiante et reprendre le culot dans 10µl d'eau + DEPC.

C- Transcription inverse des ARNm clivés/non-clivés :

Dans les 10µl d'ARN ligués ajouter 1µl de d'amorces Gene RACE Oligo dt primer, 1µl de dNTP mix, 1µl d'eau RNase free, incuber à 65°C/5min, ajouter 4µl de First strand buffer 5X, 1µl de DTT (0,1 M), 1µl de RNase Out (40u/µl), 1 µl de Taq Super Scrip III RT sur glace. Incuber à 50°C/1 heure et inactiver la Taq Super Scrip III RT à 70°C/15min. Éliminer l'ARN avec 1µl de RNaseH (2u)/37°C/20min. Conserver les échantillons à -20°C.

VII-4-1-4 PCR semi-quantitatives :

Le couple d'amorce PCR spécifique (A46F/44R) du transcrit non clivé pour le gène 5AQ (CS) ou 5Aq (CS-DIC5A) doit amplifier un produit de 327pb. (Fig. VII-5B). Les amorces étant choisies en fonction des différences de séquences entre les homéologues 5AQ/q, 5Bq et 5Dq.

Le couple d'amorce PCR spécifique (Amorce 5'RACE/A44R) des transcrits clivés au site miR172 pour le gène 5AQ (CS) ou 5Aq (CS-DIC5A), doit amplifier un produit de 335pb (Fig. VII-5A).

VII-4-1-4-1 Préparation des prémix PCR :

A- Le premix pour les ARNm clivés (colonne A de la Fig. IIX-7 A) :

Oligo Forward de l'adaptateur « gene RACER 5' »	3µl
Oligo Reverse A44 (en 3' du site d'adressage du miR172)	1,5µl
Échantillon cDNA-RACE	1µl
10X buffer	5µl
Solution de dNTPs	2µl
MgSO ₄	1µl
Taq platinum pfx	0,5µl
H2O bidistillée	35µl

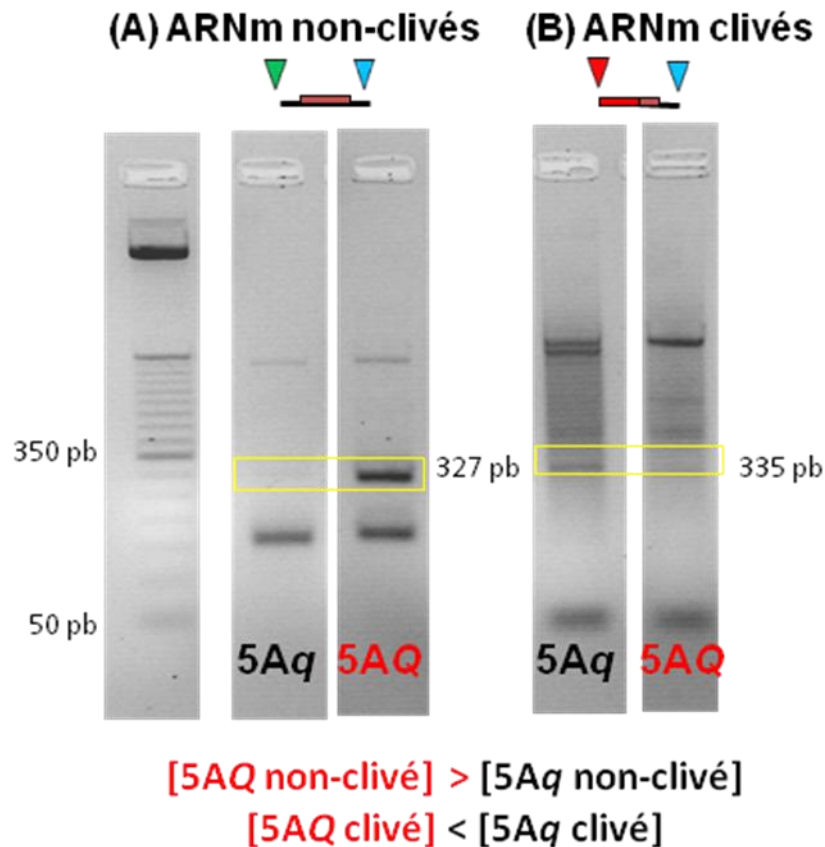


Fig. VII-5 PCR semi quantitative sur les ARNm clivé et non-clivé :

A - Vérification des niveaux d'expressions des gènes 5AQ et 5Aq par l'intermédiaire d'une PCR couvrant le site d'adressage. Le produit PCR des ADNc non clivés avec une taille attendu de 327 pb montre une plus expression plus élevée de l'allèle 5AQ par rapport à l'allèle 5Aq.

B - Produit PCR des fragments clivés au site d'adressage du miR172 avec une taille prévue de 335pb.

On retrouve le résultat attendu, c'est-à-dire un niveau d'expression plus élevé pour le gène 5AQ avec un faible clivage des ARNm. Alors que le gène 5Aq présente un niveau d'expression en dessous de celui du 5AQ, mais un clivage des ARNm bien plus important.

B- Le premix pour les ARNm non-clivés (colonne B de la Fig. IIX-7 B) :

Oligo Forward A46 (en 5' du site d'adressage du miR172)	1,5µl
Oligo Reverse A44 (en 3' du site d'adressage du miR172)	1,5µl
Échantillon cDNA-RACE	1µl
10X buffer	5µl
Solution de dNTPs	2µl
MgSO ₄	1µl
Taq platinum pfx	0,5µl
H ₂ O bidistillée	36,5µl

VII-4-1-4-2 Conditions pour la PCR semi quantitative et le RACE PCR :

Les conditions PCR sont les suivantes: 94°C 2 min, 5 cycles (94°C 30s, 72°C 1min), 5 cycles (94°C 30s, 70°C 1min), 15 cycles (94°C 30s, 60°C 30s, 72°C 1min), 72°C 10min.

VII-4-1-4-3 Observation des produits PCR :

Les produits PCR sont observés dans un gel d'agarose standard 1% (SIGMA) avec pour tampon du TAE 1X (Accugene), plus BET (Bromure d'éthydiu). Le volume de charge par échantillon était de 20µl plus 4µl de bleu de charge (6X).

Les conditions de migration sur gel étaient à 300mA/45min/14°C. Les prises d'images sur table UV ont été réalisées avec l'appareil GEL DOC 2000 (BIORAD). Le marqueur de taille est du 50pb Ladder (INVITROGEN).

VII-5 Analyse des résultats et conclusions :

L'observation et les comparaisons des produits PCR à la taille attendue, 327pb pour les ARNm non clivés (Fig. VII-5B) et 335pb pour les ARNm clivés (Fig. VII-5A) montrent :

1. Entre les ARNm non clivés des allèles 5AQ et 5Aq, qu'il y a plus d'ARNm non clivés de l'allèle 5AQ que de l'allèle 5Aq (Fig. VII-5B) confirmant donc l'expression plus forte de 5AQ par rapport à 5Aq.
2. Entre les ARNm clivés des allèles 5AQ et 5Aq, qu'il y a plus d'ARNm clivés de l'allèle 5Aq par rapport à l'allèle 5AQ, environ un facteur 2 (Fig. VII-5A).
3. Entre les ARNm clivés et les ARNm non clivés de l'allèle 5AQ, les résultats montrent que la quantité d'ARNm clivés de l'allèle 5AQ est inférieure en intensité à la quantité d'ARNm non clivés (Fig. VII-5 CS(A et B)).
4. Entre les ARNm clivés et les ARNm non clivés de l'allèle 5Aq, les résultats montrent que la quantité d'ARNm clivés de l'allèle 5Aq est supérieure en intensité à la quantité d'ARNm non clivés (Fig. VII-5 CS-DIC5A (A et B)).

HvAP2	Cly1. A	... CAGCAGCAUCAUCCAGGAUUC C...	<i>H. vulgare</i>
	cly1. B	... CAGCAGC G UCAUCCAGGAUUC C ...	<i>H. vulgare</i>
	cly1. C	... C CGCAGCAUCAUCC C GGAUUC C ...	<i>H. vulgare</i>
TaAP2	TaAP2.A	...CUGCAGCAUCAUCCAGGAUUC C ...	<i>T. aestivum</i> ssp. <i>aestivum</i>
	TaAP2.B	...C C GCAGCAUCAUCCAGGAUUC C ...	<i>T. aestivum</i> ssp. <i>aestivum</i>
	TaAP2.C	...CUGCAGCAUCAUCCAGGAUUC C ...	<i>T. aestivum</i> ssp. <i>aestivum</i>
HvAP2Like	AY069953	...CUGCAGCAUCAUCCAGGAUUCU...	<i>H. vulgare</i>
5AQ	AY702959	... CUGCAGCAUCAUCCAGGAUUUU ...	<i>T. turgidum</i> ssp. <i>carthlicum</i>
	AY702955	... CUGCAGCAUCAUCCAGGAUUUU ...	<i>T. turgidum</i> ssp. <i>durum</i>
	JF701620	... CUGCAGCAUCAUCCAGGAUUUU ...	<i>T. turgidum</i> ssp. <i>durum</i>
	AY714339	... CUGCAGCAUCAUCCAGGAUUUU ...	<i>T. turgidum</i> ssp. <i>polonicum</i>
	JF701614	... CUGCAGCAUCAUCCAGGAUUUU ...	<i>T. aestivum</i> ssp. <i>aestivum</i>
	AY702956	... CUGCAGCAUCAUCCAGGAUUUU ...	<i>T. aestivum</i> ssp. <i>aestivum</i>
	JF701619	... CUGCAGCAUCAUCCAGGAUUUU ...	<i>T. aestivum</i> ssp. <i>aestivum</i>
5Aq	AY170867	...CUGCAGCAUCAUCCAGGAUUCU...	<i>T. monococcum</i>
	JF701616	...CUGCAGCAUCAUCCAGGAUUCU...	<i>T. urartu</i>
	AY702958	...CUGCAGCAUCAUCCAGGAUUCU...	<i>T. urartu</i>
	AY702957	...CUGCAGCAUCAUCCAGGAUUCU...	<i>T. turgidum</i> ssp. <i>dicoccoides</i>
	AY714343	...CUGCAGCAUCAUCCAGGAUUCU...	<i>T. turgidum</i> ssp. <i>dicoccum</i>
	AY702960	...CUGCAGCAUCAUCCAGGAUUCU...	<i>T. aestivum</i> ssp. <i>spelta</i>
	AY714342	...CUGCAGCAUCAUCCAGGAUUCU...	<i>T. aestivum</i> ssp. <i>macha</i>
5Bq	JF701622	...CUGCAGCAUCAUCCAGGAUUCU...	<i>Ae. speltoides</i>
	JF701621	...CUGCAGCAUCAUCCAGGAUUCU...	<i>T. turgidum</i> ssp. <i>durum</i>
	DQ123819	...CUGCAGCAUCAUCCAGGAUUCU...	<i>T. turgidum</i> ssp. <i>durum</i>
	JF701612	...CUGCAGCAUCAUCCAGGAUUCU...	<i>T. aestivum</i> ssp. <i>aestivum</i>
	JF701617	...CUGCAGCAUCAUCCAGGAUUCU...	<i>T. aestivum</i> ssp. <i>aestivum</i>
5Dq	JF701613	...CUGCAGCAUCAUCCAGGAUUCU...	<i>Ae. tauschii</i>
	EU350482	...CUGCAGCAUCAUCCAGGAUUCU...	<i>Ae. tauschii</i>
	JF701615	...CUGCAGCAUCAUCCAGGAUUCU...	<i>T. aestivum</i> ssp. <i>aestivum</i>
	JF701618	...CUGCAGCAUCAUCCAGGAUUCU...	<i>T. aestivum</i> ssp. <i>aestivum</i>
miRNA			
	TamiR172a	U ACGUCGUAGUAGU U CUAAGA	
	TamiR172b	U ACGUCGUAGUAGU U CUAAG	
	TamiR172c	U ACGUCGUAGUAGU U CUAAG U	
		3' 5'	

Fig. VII-6 Mutations identifiées entre les phénotypes sauvages /domestiqués (bleu) sur le site d'adressage du miR172 pour les gènes *cly1* et *Q* :

En jaune, les SNPs identifiés entre les différentes séquences des sites d'adressage du miR172 des gènes AP2.

En gris, les séquences du site d'adressage du miR172 des gènes AP2 de l'orge (chromosome 2).

En bleu, les séquences du site d'adressage du miR172 des gènes AP2 de plantes domestiquées.

Les baret orange symbolisent l'hybridation des TamiR172 sur les sites d'adressages.

L'expérience ayant été répétée sur 6 plantes différentes de chacun de et CS-DIC5A.

Pour éliminer toute ambiguïté sur les résultats obtenus pour les ARNm clivés, j'ai réalisé une deuxième PCR, avec un couple d'amorces internes (« nested » PCR) aux produits des premières PCR semi-quantitatives sur les ARNm clivés (Amorce geneRACER_5nested/ A46R) pour un produit de 258pb.

Malgré la présence d'un bruit de fond sur gel, probablement dû à la ligation de l'adaptateur RACE sur des fragments d'ADNc de différentes tailles, on constate un clivage bien plus important des ARNm de l'allèle 5AQ par rapport aux ARNm de l'allèle 5Aq. Le résultat confirme donc que la mutation dans le site d'adressage du miR172 de l'allèle Q conduit à moins de clivage et par conséquent à une plus forte transcription observée de l'allèle Q par rapport à l'allèle q.

Pour interpréter les résultats de cette étude j'ai réalisé l'alignement des différentes séquences d'ARNm de la famille des AP2 avec les miR172 (du blé) identifiés (Yao, 2009). Il existe 3 miR172 identifiés chez le blé et baptisés Tami172 (a,b, et c). Des trois, Tami172a présente la séquence la plus proche de la séquence du site d'adressage du miR172 des gènes Q et q avec deux mésappariements probablement tolérés et toujours présent chez les blés (Nair et al., 2010). Pour les deux autres, Tami172c possède une séquence avec 1 SNP de plus de différences soit trois au final tandis que Tami172b possède un SNP commun et un nucléotide de moins. On observe des mésappariements pour toutes les séquences Q ou q, homéologue q compris. En privilégiant le modèle d'hybridation préférentiel chez les végétaux (12 nucléotides) du TamiR172a sur les sites d'adressage du miR172 des différentes séquences de gènes de la famille des AP2 (Fig. VII-6), on peut émettre l'hypothèse que le mésappariement issue de la mutation 3531 du gène Q est dans la zone d'accrochage préférentielle observée chez les plantes. Ce mésappariement (U à la place de C) dans l'ARNm à la 20^{ème} position en 5' dans le site d'adressage, pourrait être à l'origine de la diminution du clivage des ARNm provenant du gène 5AQ, laissant apparaître une surexpression (Fig. VII-5 B).

Il est possible que la mutation provoquant le changement d'acide aminé en position 329 de la protéine joue un rôle dans le phénotype domestiqué observé (Simons et al. 2006). Mes expériences montrent que l'expression plus forte de l'allèle 5AQ corréle avec un non clivage des ARNm 5AQ par le miR172 suite au mésappariement dû à la mutation dans le site d'adressage. L'idéal serait de pouvoir dissocier les deux mutations et observer les phénotypes correspondant. Peut-être qu'il existe un génotype où il n'y a qu'une des deux mutations mais je ne l'ai pas trouvé.

Ces expériences de validation par PCR semi-quantitative que j'ai réalisées, ne représentent qu'une étape préliminaire. Elles devraient être validées et complétées par d'autres approches, telles que des « northern », des PCRQ avec des génotypes présentant ou pas un phénotype domestiqué (Fig. VII-6). Dans ce cadre nous devrions observer une nette différence des produits de clivage entre les gènes Q/q des blés domestiqués et des blés aux phénotypes speltoïdes. Il est aussi important d'étudier et de

quantifier les produits de clivage des différents ARNm des homéologues *Q/q* des trois génomes A, B et D dans CS, dans les différentes lignées de délétion CS, et rechercher d'autres éventuelles séquences de miR172 pour comprendre les différents profils d'expression obtenus. Car il est possible que des mutations dans les gènes homéologues MIR172 non codants jouent un rôle dans la régulation des gènes homéologues codants.

Mais ce résultat permet d'interpréter les raisons pour lesquelles l'absence de l'homéologue 5AQ ou sa substitution par l'homéologue 5Aq provoque une augmentation de l'expression des deux autres homéologues 5Bq et 5Dq, écartant l'hypothèse d'une interaction de régulation par l'intermédiaire d'une compétition endogène pour une population de mi172 de Salmena.

Pour la substitution de l'homéologue 5AQ par l'homéologue 5Aq, le site de clivage de l'homéologue 5Aq est normale, il fixe donc correctement les mi172. Se qui se traduit par une diminution des mi172 disponibles pour les deux autres homéologues par rapport au témoin Chinese Spring qui possède l'homéologue 5AQ.

L'explication pour l'augmentation de l'expression des deux autres homéologues (5Bq et 5Dq) en l'absence de l'homéologue 5AQ, est la suivante, l'absence de transcrit ayant une mutation sur le site d'adressage conduit à une régulation plus classique, telque déjà décrite. Dans ce cas l'absence d'un des autres homéologues conduit à l'augmentation de l'autre pour compenser.

DISCUSSION GENERALE ET PERSPECTIVES

Le blé, originaire du croissant fertile (Israël, le Liban, la Syrie, le sud de la Turquie, le nord de l'Irak et le nord de l'Iran), est la première céréale domestiquée par l'homme, permettant entre autre l'émergence des premières civilisations dans la région.

Il y a 19 000 ans avant J.C., des chasseurs-cueilleurs, anciens nomades, se sédentarisent et cultivent l'en grain sauvage (*T. monococcum* ssp. *beoticum*). Environ 8500 ans plus tard, ils sélectionnent l'en grain (*T. monococcum* ssp. *monococcum*) ou les épillets ne sont plus déhiscents, sur deux sites différents. Il serait intéressant d'étudier les orthologues des gènes *Q/q* et des gènes *Br/br* dans les génotypes domestiqués et les parents sauvages des engrains (*T. monococcum*) afin de déterminer si les gènes impliqués et les mécanismes sont similaires.

Le processus de domestication se renouvelle plus tard avec les blés polyploïdes. L'amidonnier sauvage tétraploïde (*T. turgidum* ssp. *dicoccoïdes*) a été cultivé il y a 15 000-10 000 ans avant J.-C. (Feldman and Kislev, 2007) puis l'amidonnier domestiqué non déhiscent (*T. turgidum* ssp. *dicoccum*), il y a environ 12 000-9 000 ans. La culture des espèces tétraploïdes de blés et leurs disséminations a permis la formation de blés hexaploïdes (*T. aestivum*), par croisement entre l'amidonnier domestiqué (*T. turgidum* ssp. *dicoccum*) avec une espèce diploïde de blé sauvage *Ae. tauschii* (Dvorak et al., 1998 ; Matsuoka and Nasuda, 2004). Les premiers blés tétraploïdes et hexaploïdes (épautre) domestiqués n'étaient pas facilement battables. Il a fallu attendre l'arrivée des blés tétraploïdes à grains nus, (*T. turgidum* ssp. *durum*) (Damania, 1998) et des blés hexaploïdes à grains nus, (*T. aestivum* ssp. *aestivum*) (Dvorak et al., 2011) qui sont facilement battables. Ces blés ont été suggérés comme issus, pour les premiers, de l'amidonnier sauvage (*T. turgidum* ssp. *dicoccoïdes*) et pour les deuxièmes d'un autre croisement entre un allotétraploïde intermédiaire *T. turgidum* ssp. *dicoccon* et *Ae. tauschii* (Dvorak et al., 2011). Pour chacune de ces formes de blé domestiqué, la sélection d'autres caractères de domestication, comme l'augmentation des rendements, la panification, la résistance aux maladies, s'est poursuivie. L'étude que j'ai menée autour des gènes de domestication montre des relations et des échanges complexes de gènes (y compris les gènes *Q/q*) entre ces différentes espèces et formes de blés.

Ce n'est pas l'unique gène de domestication du blé mais la ou les mutation(s) ayant conduit au gène 5AQ des blés domestiqués sont les principales responsables du caractère de non-déhiscence et d'un battage facile des épillets. Ce gène est un facteur de transcription. Il intervient dans le développement du méristème floral, il possède deux motifs AP2 (motifs DNA-binding propre aux plantes). En tant que facteur de transcription son spectre d'action serait large puisqu'il gouverne la mise en place des organes floraux par l'intermédiaire d'une régulation spatiale et temporelle des gènes impliqués dans l'homéostasie florale (Drews et al., 1991). Il exerce une action pleiotropique sur différents gènes notamment, *Tg/tg*, *Tough Glumes* (Kerber and Rowland, 1974 ; Cao et al., 1997 ; Gill et al., 2006), *Br/br*,

Brittle Rachis (Sharma and Waynes, 1980 ; Zohary and Hopf, 2000 ; Peng et al., 2003 ; Konishi and al., 2006) et *C/c, Club* (Nilsson-Ehle, 1911 ; Swaminathan, 1966 ; Rao, 1973 ; Mitrofanova, 1997), ces différents gènes sont responsables du carénage de la graine par le tégument (*Tg*), la fragilité du rachis pour la déhiscence des épillets (*Br*) et la compaction de l'épi (*C*). Ces gènes ont du être sélectionnés pour que les caractères majeurs de domestication soient visibles.

Pour étudier le devenir des homéologues du gène *5AQ* dans le blé hexaploïde, j'ai caractérisé leurs organisations et évolution ainsi que leurs implications dans le phénotype domestiqué.

L'étude structurale des séquences génomiques, comparant 11 régions de cinq espèces différentes représentant les trois niveaux de ploïdie, montre une dynamique importante d'insertion et d'élimination des TEs dans le génome du blé. Cette dynamique a conduit à un espace TEs complètement différent entre les principaux génomes A, B, D et S; et à des ruptures de syntenie importantes entre haplotypes d'un même génome. Cette étude m'a permis d'identifier le premier *Hélitron* actif du blé qui s'est inséré dans l'homéologue *5Bq* du cv de Renan. En recherchant son origine, j'ai découvert que la séquence de cet *Hélitron* est absente de l'amidonnier (*T. turgidum* ssp. *dicoccoïdes*), de l'amidonnier domestique (*T. turgidum* ssp. *dicoccum*), d'*Ae. tauschii* et encore plus surprenant de la majorité des génotypes du blé hexaploïde. Le traçage de cet *Hélitron* actif et de son insertion dans le *5Bq* de Renan m'a permis également de « remonter le temps » à travers l'histoire des différents croisements de blés hexaploïdes. Cette recherche m'a permis de le tracer l'insertion dans des variétés du blé d'origine ukrainienne et de trouver des éléments similaires dans le blé sauvage *Ae. ventricosa*. Ces blés ayant servis dans la sélection du cv Renan suggèrent que l'*Hélitron* actif de Renan ne proviendrait probablement pas des espèces progénitrices du blé hexaploïde mais des espèces et génotypes qui ont été introgressés dans le blé hexaploïde lors de sélections variétales. Avoir un *Hélitron* actif du blé n'est pas sans conséquences quand on regarde l'impact des *Hélitron* dans l'évolution et la génération rapide de diversité haplotypique dans le maïs ou d'autres espèces ((Morgante et al., 2005; Lal et al., 2009). Il serait important d'approfondir et de confirmer les origines de l'*Hélitron* actif ainsi que les conditions de son activation par la génération et l'analyse des séquences haut-débits de ces différentes espèces de blé et sur des descendants pour repérer des éventuelles transpositions (Wang and Dooner, 2006b).

L'analyse de l'implication des homéoallèles *5Bq* et *5Dq* dans les caractères de domestication m'a été rendue possible par la caractérisation phénotypique et fonctionnel du blé hexaploïde et surtout des lignées de délétions, ayant perdu un ou plusieurs homéologue(s) du gène *Q/q*. Les résultats montrent que les trois homéologues jouent un rôle dans le phénotype domestiqué observé et régulent l'expression des uns et des autres. Ce résultat est surprenant quand on sait que l'homéologue *5Bq* est pseudogénéisé dans le blé hexaploïde à cause d'un décalage de la fenêtre de lecture.

L'interaction entre homéologues avec des niveaux d'implications différents dans le phénotype domestiqué, m'a amené à la conclusion de sous-fonctionnalisation des trois homéologues; avec une hyperfonctionnalisation du gène *5AQ* et une hypo-fonctionnalisation des homéologues *5Dq* et plus étonnamment du *5Bq* (spseudogène). Cette interaction entre les trois homéologues se ferait donc au niveau transcriptionnel.

Bien qu'aucune expérience sur l'expression n'ait été faite à l'époque, mes travaux sur le rôle et les interactions entre différents homéologues rappellent des constatations anciennes très pertinentes, montrant qu'il faut une quantité de copies au moins deux fois et demi supérieures de l'allèle *5Aq* (Mura-mastu, 1962) pour égaler la copie *5AQ*. Il serait fort intéressant de reprendre ce matériel et de caractériser l'expression des homéologues *Q/q* et leur régulations

Les différents homéologues du gène *Q/q*, sont homologues à *AP2* dont l'expression se régule par les miRNAs, et possèdent tous le site d'adressage miR172. Leurs interactions peut s'interpréter par un modèle de régulation récemment suggéré (Salmena et al., 2011) selon lequel les séquences d'ARNm codants, d'ARN longs non-codants, et les transcrits de pseudogènes se régulent entre eux au niveau transcriptionnel par l'intermédiaire de miRNA. Salmena et al., (2011) proposent en effet qu'entre les ARNs ayant les même sites d'adressage de miRNA, il existerait une compétition endogène pour la fixation de l'ensemble des miARNs, qu'ils nomment « ceRNAs » (competitive endogenous RNAs). Les miRNA seraient accrochés dans la région 3'UTR sur des sites que les auteurs nomment « microRNA Response Element » (MREs). Il existerait autant de MREs que de miRNA différents dans l'organisme. La perturbation engendré par la capture d'un pool de miRNA dans la cellule par l'ARNm d'un gène peut modifier la quantité d'un ou d'autres ARNm issus d'autres gènes ayant des MREs communs (Salmena et al., 2011). Ce model d'interaction pourrait être une réponse pour identifier un vaste réseau de régulation via le transcriptome et la mise en place des tissus ou organes.

J'ai testé ce modèle, en partie, en caractérisant les allèles *5AQ* et *5Aq*. Ces deux allèles diffèrent par six SNP, deux présents dans des introns (1 et 9), deux sont des substitutions synonymes dans des exons (2 et 8), un dans l'exon 9 correspond à une substitution non-synonyme. C'est la mutation V329I (Isoleucine à la place de la Valine en position 329) dans la protéine qui a été identifiée comme responsable du caractère de domestication. L'étude des interactions protéine-protéine a montré qu'elle engendre la formation de homodimères et que ceux-ci amplifient leurs propres transcriptions, fut la première conclusion (Simons et al., 2006). En reprenant les alignements des séquences entre les allèles *5AQ* et les allèles *5Aq* de différentes espèces et génotypes du blé, j'ai constaté qu'une mutation dans le site d'adressage du miR172 (dans la séquence 3' UTR) était également toujours associée aux phénotypes domestiqués/sauvages observés. En étudiant les produits de clivage entre l'allèle *5AQ* et l'allèle *5Aq*, j'ai validé que la mutation sur le site d'adressage du miR172 est corrélée à un faible taux de clivage des transcrits de l'allèle *5AQ* par le complexe miR172.

Le constat est une plus forte expression de l'allèle 5AQ, comparé à celle de l'allèle 5Aq, ayant ses transcrits plus clivés. Cette mutation induit une plus forte expression du 5AQ, la conséquence est une répercussion sur l'expression des autres homéologues 5Dq et 5Bq. Afin d'établir les différentes interactions, il serait intéressant d'étudier par diverses approches, les populations d'ARNm clivés ou non, pour les différents homéologues des gènes *Q/q*, ainsi que les populations miR172 et la régulation du gène MIR172 du blé. Il est à noter que l'on trouve trois Tamir172 différents dans le blé, il est probable qu'ils représentent chacun un des trois gènes homéologues. Or leurs séquences montrent qu'un seul pourrait s'apparier efficacement sur le site d'adressage du miR172 des transcrits, cela pourrait signifier que la sélection du blé concernerait également les gènes longs à ARN non codants qui joueraient un rôle dans le blé pour obtenir le phénotype de domestication.

L'action du gène *Q*, me rappelle dans un contexte autre que celui de la domestication, la théorie de Goldschmidt (1878-1958), qui affirmait qu'il est possible qu'une simple mutation qui touche un gène régulateur d'autres gènes, qu'il donne « monstre prometteur » soit capable de provoquer un bon évolutif des individus qui le possèdent.

Pour conclure, mes travaux de thèse m'ont permis d'apprécier l'importance de la polyploïdie et de la dynamique des TEs dans l'organisation, l'évolution et le fonctionnement des génomes du blé. J'ai pu également, apprécier, en analysant les homéologues des gènes *Q/q*, le devenir des gènes dupliqués au niveau structural, fonctionnel et de la régulation. Mon travail ne représente qu'une petite pièce du puzzle ayant démarré il y a un siècle et qui vise à comprendre la polyploïdie et la domestication du blé. J'espère en poursuivant ce travail pouvoir rajouter d'autres pièces manquantes.

ANNEXES

Annexe 1 : Classification des blés anciens.

grains	Genome	Nom d'espèce		nom allemand	Nom français	Nom anglais
diploid (2n=14)	A	<i>Triticum monococcum</i> ssp. <i>beotium</i>	ssp. <i>Aegilopoides</i> Schiem.	wildes Einkorn	engrain sauvage	wild einkorn (1 grained)
dur à battre	A	<i>Triticum monococcum</i> ssp. <i>beotium</i>	ssp. <i>Thaoudar</i> Schiem.	wildes Einkorn	engrain sauvage	wild einkorn (2 grained)
grains carénés	A	<i>Triticum monococcum</i> ssp. <i>monococcum</i>	L.	Einkorn	engrain	einkorn
tetraploid (2n=28)	AB	<i>Triticum turgidum</i> ssp. <i>dicoccoides</i>	Körn	wilder Emmer	amidonnier sauvage	wild emmer wheat
dur à battre	AB	<i>Triticum turgidum</i> ssp. <i>araraticum</i>	Jakubz	wilder Emmer	amidonnier sauvage	wild emmer wheat
grains carénés	AB	<i>Triticum turgidum</i> ssp. <i>dicoccum</i> var <i>farum</i>		Emmer	amidonnier	emmer wheat
	AB	<i>Triticum turgidum</i> ssp. <i>dicoccum</i> var <i>lugiliform</i>	Schübl.	Emmer	amidonnier	emmer wheat
		<i>Triticum turgidum</i> ssp. <i>palaeocolchicum</i>	Men.	*	*	Kolchic emmer wheat
	AB	<i>Triticum turgidum</i> ssp. <i>tibetanum</i>		*	*	
	AG	<i>Triticum turgidum</i> ssp. <i>timophevii</i>	Zhuk.	*	*	Timophevii wheat
tetraploid (2n=28)	AB	<i>Triticum turgidum</i> ssp. <i>durum</i>	Desf.	Hartweizen	blé dur	hard wheat
facilement battable	AB	<i>Triticum turgidum</i> ssp. <i>turgidum</i>	L.	Rauhweizen	blé poulard	rivet wheat
grains nus	AB	<i>Triticum turgidum</i> ssp. <i>persicum</i> / <i>carthanicum</i>	Vav. (= <i>T. carthanicum</i> Nevski)	Persischer Weizen	*	Persian wheat
	AB	<i>Triticum turgidum</i> ssp. <i>polonicum</i>	L.	Polnisher Weizen	*	Polish wheat
	AB	<i>Triticum turgidum</i> ssp. <i>turanicum</i>	Jakubz. (= <i>T. orientale</i> Perc.)	Khorassan Weizen	*	Khorassan wheat
	AB	<i>Triticum turgidum</i> ssp. <i>parvicoccum</i>	Kisley	*	*	*
hexaploid (2n=42)	ABD	<i>Triticum aestivum</i> ssp. <i>spelta</i>	L. (Iranian)	Dinkel (Spelz)	épeautre	spelt wheat
dur à battre	ABD	<i>Triticum aestivum</i> ssp. <i>spelta</i>	(European) <-- <i>T. turgidum</i> ssp. <i>dicoccum</i> (q) x <i>T. aestivum</i> ssp. <i>compactum</i> (Q)	Dinkel (Spelz)	épeautre	spelt wheat
grains carénés	ABD	<i>Triticum aestivum</i> ssp. <i>macha</i>	Derk. et Men.	*	*	Makha wheat
	ABD	<i>Triticum aestivum</i> ssp. <i>vavilovii</i>	(Tum.) Jakubz	*	*	Vavilov's wheat
hexaploid (2n=42)	ABD	<i>Triticum aestivum</i> ssp. <i>aestivum</i>	L. (= <i>T. aestivum</i> ssp. <i>vulgare</i> vill.)	Saatweizen	blé tendre	bread wheat
facilement battable	ABD	<i>Triticum aestivum</i> ssp. <i>compactum</i>	Host (= <i>T. aestivum</i> grex <i>aestivo-compactum</i> Schiem.)	Zwergweizen	blé compact	club wheat
grains nus	ABD	<i>Triticum aestivum</i> ssp. <i>sphaerococcum</i>	Blé nain indien	Indischer Kugelweizen		*

Les gènes majeurs de domestication (Q/q) chez les blés polyploïdes
Harry Belcram

Annexe 2 : Alignements des séquences internes d'*Hélitrons* (CS et Renan) pour établir les groupes et vérifier la proximité de l'*Hélitron* de Renan.

(L'arbre correspondant est la Fig. VII-2)

CLUSTAL W (1.83) multiple sequence alignment

```

[8] lcl | GJZO2FP01C141V      ----AATCTTGAAACAAATAGTATTAGTCGAGACGTGCGTTGCACGTACA
[9] lcl | GIB3ZOT04H35YQ      ----AATCTTGAAACAAATAGTATTAGTCGAGACGTGCGTTGCACGTACA
[4] lcl | GKWV9KP08IV6ON      ----AATCTTGAAACAAATAGTATTAGTCGAGACGTGCGTTGCACGTACA
[7] lcl | F7IO22A02I4JP8      ----AATCTTGAAACAAATAGTATTAGTCGAGACGTGCGTTGCACGTACA
[6] lcl | GG1VXPK04C9ACN      AACGAATTTT---ATGTGT-GTATTGGTGGAGACGTGTGTTGCACGTGCA
[10] lcl | GG1VXPK01B4N5R     AACGAATTTT---ATGTGT-GTATTGGTGGAGACGTGTGTTGCACGTGCA
[2] lcl | GIB3ZOT04JGZOJ      AACGAATTTT---ATGTGT-GTATTGGTGGAGACGTGTGTTGCACGTGCA
[13] lcl | GJ20ROV01AI9QU     ----TATTAACAATTATTTGATGTCGATTGAGACGTGCGTTGCACGTACA
[11] lcl | GGWFHOC02FTJO2     ----TATTAACAATTATTTGATGTCGATTGAGACGTGCGTTGCACGTACA
[3] lcl | GJP3QJH01BYXJE      ----TATTAACAATTATTTGATGTCGATTGAGACGTGCGTTGCACGTACA
[12] lcl | GCI7VBT01BPGSP     ----TATTAACAATTATTTGATGTCGATTGAGACGTGCGTTGCACGTACA
[5] lcl | GGSKSHR01CKBUM      ----TATTAACAATTATTTGATGTCGATTGAGACGTGCGTTGCACGTACA
[1] lcl | F2NKKJR102I5F83     ----TATTAACAATTATTTGATGTCGATTGAGACGTGCGTTGCACGTACA
[0] Hélitron-Renan           ---GTATGACCATA-ATTTAATTTTATAATATACGTGCGTTGCACGTACA
[out] Hélitron-N4_ZM         ----TATCGTAGTATAC-TGGTATTATATGTTCCCGTTGCAACGCACGGG
                               **          *  *  *          *      *  *  *

[8] lcl | GJZO2FP01C141V      TGTTTAC-TAG
[9] lcl | GIB3ZOT04H35YQ      TGTTTAC-TAG
[4] lcl | GKWV9KP08IV6ON      TGTTTAC-TAG
[7] lcl | F7IO22A02I4JP8      TGTTTAC-TAG
[6] lcl | GG1VXPK04C9ACN      CGCTTAC-TAG
[10] lcl | GG1VXPK01B4N5R     CGCTTAC-TAG
[2] lcl | GIB3ZOT04JGZOJ      CGCTTAC-TAG
[13] lcl | GJ20ROV01AI9QU     TGTTTAC-TAG
[11] lcl | GGWFHOC02FTJO2     TGTTTAC-TAG
[3] lcl | GJP3QJH01BYXJE      TGTTTAC-TAG
[12] lcl | GCI7VBT01BPGSP     TGTTTAC-TAG
[5] lcl | GGSKSHR01CKBUM      TGTTTAC-TAG
[1] lcl | F2NKKJR102I5F83     TGTTTAC-TAG
[0] Hélitron-Renan           TGTTTAC-TAG
[out] Hélitron-N4_ZM         CACTCACCTAG
                               *  **  ***

```

Annexe 3 : Alignements des séquences internes et externe aux *Hélitrons* en 3' pour différencier les groupes établis et les insertions.

CLUSTAL W (1.83) multiple sequence alignment

```
[8] lcl|GJZO2FP01C141V      --ACGTGCGTTGCACGTAC-ATGTTTACTAGTAGTATTAGATGCGTGTTT
[9] lcl|GIB3ZOT04H35YQ      --ACGTGCGTTGCACGTAC-ATGTTTACTAGTAGTATTAGATGCGTGTTT
[4] lcl|GKWW9KP08IV6ON      --ACGTGCGTTGCACGTAC-ATGTTTACTAGTAGTATTAGATGCGTGTTT
[1] lcl|F2NJKR102I5F83      --ACGTGCGTTGCACGTAC-ATGTTTACTAGTTGGATAAAAAAGGAGAGAA
[5] lcl|GGSKSHR01CKBUM      --ACGTGCGTTGCACGTAC-ATGTTTACTAGTTGGATAAAAA-GGAGAGAA
[12] lcl|GCI7VBT01BPGSP     --ACGTGCGTTGCACGTAC-ATGTTTACTAGTTGGATAAAAA-GGAGAGAA
[6] lcl|GG1VXPK04C9ACN      --ACGTGTGTTGCACGTGC-ACGCTTACTAGTTAATCTAAATGTTTCATG
[10] lcl|GG1VXPK01B4N5R     --ACGTGTGTTGCACGTGC-ACGCTTACTAGTTAATCTAAATGTTTCATG
[2] lcl|GIB3ZOT04JGZOJ      --ACGTGTGTTGCACGTGC-ACGCTTACTAGTTAATCTAAATGTTTCATG
[out] Hélitron-N4_ZM        TATCGTAGTATACTGGTATTATATGTTCCCGTTGCAACGCACGGGCACTC
[0] Hélitron-Renan          --TTTTATAATATACGTAC-ATGTTTACTAGTCTGGCAAAAAGCACCACC
                        *      *      **      *      *      **      *      *
```

```
[8] lcl|GJZO2FP01C141V      GCTCTCCGTGGTATCCGAGTTAATTTGTCTGTTACCGGCTGGATGAACATA
[9] lcl|GIB3ZOT04H35YQ      GCTCTCCGTGGTATCCGAGTTAATTTGTCTGTTACCGGCTGGATGAACATA
[4] lcl|GKWW9KP08IV6ON      GCTCTCCGTGGTATCCGAGTTAATTTGTCTGTTACCGGCTGGATGAACATA
[1] lcl|F2NJKR102I5F83      GCCCCAC----TGCCCCACTGGGCCTCTCGCTCAGCTGGCCACTAAACTG
[5] lcl|GGSKSHR01CKBUM      GCCCCAC----TGCCCCACTGGGCCTCTCGCTCAGCTGGCCACTAAACTG
[12] lcl|GCI7VBT01BPGSP     GCCCCAC----TGCCCCACTGGGCCTCTCGCTCAGCTGGCCACTAAACTG
[6] lcl|GG1VXPK04C9ACN      GTTTAGA----ATTCATGCAAGGTATGTTTAGTACTCCTTCCGTTAACCT
[10] lcl|GG1VXPK01B4N5R     GTTTAGA----ATTCATGCAAGGTATGTTTAGTACTCCTTCCGTTAACCT
[2] lcl|GIB3ZOT04JGZOJ      GTTTAGA----ATTCATGCAAGGTATGTTTAGTACTCCTTCCGTTAACCT
[out] Hélitron-N4_ZM        ACCTAG-----
[0] Hélitron-Renan          ACCACCCACAGGATATCAACCATCGGACAAACCACCCA-TATATTATCCG
```

```
[8] lcl|GJZO2FP01C141V      GTGT--CTATTTACTGTTTTTGTGTATGTGTTAACAAAACAAGCACTGCC
[9] lcl|GIB3ZOT04H35YQ      GTGT--CTATTTACTGTTTTTGTGTATGTGTTAACAAAACAAGCACTGCC
[4] lcl|GKWW9KP08IV6ON      GTGT--CTATTTACTGTTTTTGTGTATGTGTTAACAAA-CAAGCACTGCC
[1] lcl|F2NJKR102I5F83      GCCCAGCCGGGCAACCCACTAGCCCAGCCCCC-TCCTCCCTTAACCCC
[5] lcl|GGSKSHR01CKBUM      GCCCAGCCGGGCAACCCACTAGCCCAGCCCCCCTCACTCCCTTAACCCC
[12] lcl|GCI7VBT01BPGSP     GCCCAGCCGGGCAACCCACTAGCCCAGCCCCC-CCTCCCTACTACCCCT
[6] lcl|GG1VXPK04C9ACN      AGGT-GAGGGTTAACCTACGACTCAGGGATTGGGAGAGCCTCAGATCCGG
[10] lcl|GG1VXPK01B4N5R     AGGT-GAGGGTTAACCTACGACTCAGGGATTGGGAGAGCCTCAGATCCGG
[2] lcl|GIB3ZOT04JGZOJ      AGGT-GAGGGTTAACCTACGACTCAGGGATTGGGAGAGCCTCAGATCCGG
[out] Hélitron-N4_ZM        -----
[0] Hélitron-Renan          CGCG-CTGACTACCTGGTAGTACAATCTTCTACCCAGATTATAGCTTAT
```

```
[8] lcl|GJZO2FP01C141V      A-CATGGACGATATTTTC-ACGCACGAGTTCGACAGTTAATAAACCCAGA-
[9] lcl|GIB3ZOT04H35YQ      A-CATG-ACGATATTTTC-ACGCACGAGTTCGACAGTTAATAAACCCAGA-
[4] lcl|GKWW9KP08IV6ON      A-CATGGACGATATTTTCTACGCACGAGTTCGACAGTTAATAAACCCAGA-
[1] lcl|F2NJKR102I5F83      T-CACCGGGGGAACCTAGCC-ACCAACTCCCC-ACTCCC--ACGTCA-
[5] lcl|GGSKSHR01CKBUM      T-CACCGGGGGAATCCTAGCCCACCAACTCCCCCACTCCCCGACGTCA-
[12] lcl|GCI7VBT01BPGSP     T-AACCCCTTA--CCCGACCGGGGAACCGGGTACTACCGTACCCGA-
[6] lcl|GG1VXPK04C9ACN      C-TATTTCCGGCGA-CCAGAGCCACAGCACAGATCACAGAAAGATGGGAA
[10] lcl|GG1VXPK01B4N5R     C-TATTTCCGGCGA-CCAGAGCCACAGCACAGATCACAGAAAGATGGGAA
[2] lcl|GIB3ZOT04JGZOJ      C-TATTTCCGGCGA-CCAGAGCCACAGCACAGATCACAGAA-GATGGGAA
[out] Hélitron-N4_ZM        -----
[0] Hélitron-Renan          AGTAGCTTAGATGATTACCTGCTTCAAATCCTCCTCATAGTCGCTCAGAT
```

```
[8] lcl|GJZO2FP01C141V      --TCCGGAGACT--TCCC-ATCTTTCTGTGATCTGTGCTGTGGCTCTGG
[9] lcl|GIB3ZOT04H35YQ      --TCCGGAGACT--TCCCCATCTTTCTGTGATCTGTGCTGTGGCTCTGG
[4] lcl|GKWW9KP08IV6ON      --TCCGGAGACT--TCCC-ATCTTTCTGTGATCTGTGCTGTGGCTCTGG
[1] lcl|F2NJKR102I5F83      --CTC--ACTCTCCCTCCC-ATCGATCCANATCGGGG-CTCGACCCCGCA
[5] lcl|GGSKSHR01CKBUM      --CTC--ACTCTCCCTCCCCATCGATCCATATCGGGGGCTCGACCCCGCA
[12] lcl|GCI7VBT01BPGSP     --CCCCAACCCCTAACCCCTACC--CCCTACCCCTACCCCGTACCGTA
[6] lcl|GG1VXPK04C9ACN      GTCTCCGGATCTGGTTTATTAAGTGTC-GAACTCGTGCGTGAAATATCG
[10] lcl|GG1VXPK01B4N5R     GTCTCCGGATCTGGTTTATTAAGTGTC-GAACTCGTGCGTGAAATATCG
```

Les gènes majeurs de domestication (*Q/q*) chez les blés polyploïdes
Harry Belcram

[2] lcl GIB3ZOT04JGZOJ	GTCTCCGGATCTGGTTTATTAACTGTC-GAACTCGTGCGTGAAAAATATCG
[out] <i>Hélitron</i> -N4_ZM	-----
[0] <i>Hélitron</i> -Renan	TGAAGTTGATGTCGGCCTCCAGCCCCGGAAGTTGATCGCCGCGCGATCG
[8] lcl GJZO2FP01C141V	TCGCCGGAAATAGCCGGATCTGAGG-----
[9] lcl GIB3ZOT04H35YQ	TCGCCGGAAATAGCCGGATCTGAGG-----
[4] lcl GKWV9KP08IV6ON	TCGCCGGAAATAGCCGGATCTGAGG-----
[1] lcl F2NKJR102I5F83	GCTCCTCGCCCGAACGCCCGCCCGCCTCCTA
[5] lcl GGSKSHR01CKBUM	GCTCCTCGCCCGAACGCC-GCCCGCC-----
[12] lcl GCI7VBT01BPGSP	CCTACCTACCTACCTTCCCCCTCCCCCCTA-
[6] lcl GG1VXPK04C9ACN	TCCATGTGGCAGTGCTTGTTTT-----
[10] lcl GG1VXPK01B4N5R	TCCATGTGGCAGTGCTTGTTTT-----
[2] lcl GIB3ZOT04JGZOJ	TCCATGTGGCAGTGCTTGTTTTG-----
[out] <i>Hélitron</i> -N4_ZM	-----
[0] <i>Hélitron</i> -Renan	TAGGCCCTGCGCCACG-----

Annexe 4 : position des amorces PCR pour RACE sur alignements de séquences 5Aq, 5Bq et 5Dq.

CS102N4Q est la séquence du BAC portant l'homéologue 5Bq

CS201F2Q est la séquence du BAC portant l'homéologue 5Dq

CS12224M17Q est la séquence du BAC portant le gène 5Aq

CLUSTAL 2.0.12 multiple sequence alignment (60 nucleotides par lignes)

```

5Bq_CS102N4QEXONS      ATGCCATAGACGCGACCCCACTCATCGGTCCAGGTCG----CTCCGAGCCGAGCGGCGGC
5Dq_CS201F2QEXONS      ATGCCATAGACGCGACCCCACTCATCGGTCCAGGTCGGTTCGCTCCGAGGCGAGCGGCGGC
5Aq_CS12224M17Q        ATGCCATAGACGCGACCCCACTCATCGGTCCAGGTCGCTCGGAGCCGAGCGGCGGC
*****
5Bq_CS102N4QEXONS      GGGCGGGGC---GGGTCAAGCGAGTTTCGGAGGCGCCCGGGGCCCCGGATCGGAGATG
5Dq_CS201F2QEXONS      GGGAGAGGCGGTGGGTCAAGCGAGTTTCGGAGGCGCCCGGGGCGCTCGGATCGGAGATG
5Aq_CS12224M17Q        GGGCGGGC-----GAGGAGTGCCTTTATTTCGTCCCGGCGGGCTCGGATCGGAGATG
*** * *                *   * * * * * * * * * * * * * * * * * * * * *
5Bq_CS102N4QEXONS      GTGCTGGATCTCAATGTCGAGTCGCCGCGGACTCCGGCACGTCCAGCTCGTCCGTGCTC
5Dq_CS201F2QEXONS      GTGCTGGATCTCAATGTCGAGTCGCCGCGGACTCCGGCACGTCCAGCTCGTCCGTGCTC
5Aq_CS12224M17Q        GTGCTGGATCTCAATGTCGAGTCGCCGCGGACTCCGGCACGTCCAGCTCCTCCGTGCTC
*****
5Bq_CS102N4QEXONS      AACTCCGCGGACGCCGGCG-----GCTTCCGGTTCGGCTGCTCGGGAGCCCCGAT
5Dq_CS201F2QEXONS      AACTCCGCGGACGCCGGCGCGCGCGCGGCTTCAGGTTCCGGCTGCTCGGGAGCCCCGAT
5Aq_CS12224M17Q        AACTCCGCGGACGCCGGTGGCGGCG---GCTTCCGGTTCGGCTGCTCGGGAGCCCTGAT
***** *                * * * * * * * * * * * * * * * * *
5Bq_CS102N4QEXONS      GACGACGACTGGTCCGCGAGCCGCGCGCCGCGCTCCGGGTTTCGTGACGAGGCAGCTC
5Dq_CS201F2QEXONS      GATGACGACTGCTCCGCGAGCCGCGCGCGGTCGGGTTCCGGGTTTCGTGACGAGGCAGCTC
5Aq_CS12224M17Q        GATGACGACTGCTCCGCGAGCCGCGCGCGGTCGGGCCCGGGTTTCGTACGAGGCAGCTC
** ***** * * * * * * * * * * * * * * * * * * * * *
5Bq_CS102N4QEXONS      TTCCCCGCGTCGCCGCCCG-----CGCCCGGGATGATGATGGGGCAGCAGGCC
5Dq_CS201F2QEXONS      TTCCCCGCGTCGCCGCCCGGTACGCGGGCGCGCCCGGGATGATGACGGGCAGCTGGCA
5Aq_CS12224M17Q        TTCCCCGCGTCGCCGCCCGGGACGCGGGCGCGCCCGGGGTGACGATGGGGCAGCAGGCC
*****
5Bq_CS102N4QEXONS      CCGGCGCC---GCCGATGGCCCCCGTGTGGCAGCCGCGCCGCGCCGAGGAGCTGGTCGCG
5Dq_CS201F2QEXONS      CCGGCGCC---GCCGATGGCCCCCGTGTGGCAGCCGCGGCGCGCCGAGGAGCTCGTGATG
5Aq_CS12224M17Q        CCGGCGCCTGCGCCGATGGCGCCCCGTGTGGCAGCCGCGGCGCGCCGAGGAGCTCCTCGTG
***** * * * * * * * * * * * * * * * * * * * * *
5Bq_CS102N4QEXONS      GCGCAGCGGGTGGCGCCCGCGAAGAAGACGCGGCGCGGGCCGAGGTCGCGGAGCTCGCAG
5Dq_CS201F2QEXONS      GCGCAGCGGGTGGCGCCCGCGAAGAAGACGCGGCGGGGGCCGAGGTCGCGGAGCTCGCAG
5Aq_CS12224M17Q        GCGCAGCGGATGGCGCCCGCGAAGAAGACGCGGCGGGGCCGAGGTCGCGCAGCTCGCAG
*****
5Bq_CS102N4QEXONS      TACAG-----
5Dq_CS201F2QEXONS      TACAGGGGCGTCACCTTCTACCGCAGGACCGGCCGGTGGGAGTCGCACATCTGGGGGACT
5Aq_CS12224M17Q        TACAGGGGCGTCACCTTCTACCGCAGGACCGGCCGGTGGGAGTCGCACATCTGGGGGATT
*****
5Bq_CS102N4QEXONS      -----
5Dq_CS201F2QEXONS      GCGGAAGCAGGTCTACTTAGGTGAGGTGGTTTCGACACTGCGCACGCGGCCGCAAGGAG
5Aq_CS12224M17Q        GCGGAAGCAGGTCTACTTGGGTGAGGTGGTTTCGACACTGCGCACGCGGCCGCAAGGAG
-----
5Bq_CS102N4QEXONS      -----
5Dq_CS201F2QEXONS      GGCCTACGATCGCGCGCGATCAAGTTCGGGGGCTGGAGGCCGACATCAACTTCAATCT
5Aq_CS12224M17Q        GGCCTACGATCGCGCGCGATCAAGTTCGGGGGCTGGAGGCCGACATCAACTTCAATCT

```

Les gènes majeurs de domestication (Q/q) chez les blés polyploïdes
Harry Belcram

5Bq_CS102N4QEXONS -----CAGATGAGGAACTGGACCAAGGAGGAGTTCGT
5Dq_CS201F2QEXONS GAGCGATTATGAGGAGGATTTGAAGCAGCAGATGAGGAACTGGACCAAGGAGGAGTTCGT
5AQ_CS12224M17Q GAGCGACTACGAGGAGGATTTGAAGCAGCAGATGAGGAACTGGACCAAGGAGGAGTTCGT

5Bq_CS102N4QEXONS GCACATCCTCCGCCGCCAGAGCACGGGGTTCGCCAGGGGAGCTCCAAGTACCGCGGCGT
5Dq_CS201F2QEXONS GCACATCCTCCGCCGCCAGAGCACGGGGTTCGCCAGGGGAGCTCCAAGTACCGCGGCGT
5AQ_CS12224M17Q GCACATCCTCCGCCGCCAGAGCACGGGGTTCGCCAGGGGAGCTCCAAGTACCGCGGCGT

5Bq_CS102N4QEXONS CACGCTCCACAAGTGC GGCCGCTGGGAGGCAAGGATGGGCCAGCTGCTCGGCAAGAAGTA
5Dq_CS201F2QEXONS CACGCTCCACAAGTGC GGCCGCTGGGAGGCAAGGATGGGCCAGCTGCTCGGCAAGAAGTA
5AQ_CS12224M17Q CACGCTCCACAAGTGC GGCCGCTGGGAGGCAAGGATGGGCCAGCTGCTCGGCAAGAAGTA

5Bq_CS102N4QEXONS GTACATATATCTTGGCCTCTTTGACAGCGAAGTTGAAGCTGCAAGGAGGGCGTACGACAG
5Dq_CS201F2QEXONS GTACATATATCTTGGCCTCTTTGACAGCGAAGTTGAAGCTGCAAGGAGGGCGTACGACAG
5AQ_CS12224M17Q GTACATATATCTTGGCCTCTTTGACAGCGAAGTTGAAGCTGCAAGGAGGGCGTACGACAG

5Bq_CS102N4QEXONS GGCGGCGATTTCGCTTCAATGGGAGGGAAGCTGTGACTAACTTTGAGAGCAGCTCCTACAA
5Dq_CS201F2QEXONS GGCGGCGATTTCGCTTCAATGGGAGGGAAGCTGTGACTAACTTTGAGAGCAGCTCCTACAA
5AQ_CS12224M17Q GGCGGCGATTTCGCTTCAATGGGAGGGAAGCTGTGACTAACTTTGAGAGCAGCTCCTACAA

5Bq_CS102N4QEXONS TGGGGATGCTCCACCCGACGCCGAAAATGAGGGCAATTGTTGATGCTGATGCTCTTGACT
5Dq_CS201F2QEXONS TGGGGATGCTCCACCCGACGCCGAAAATGAGGGCAATTGTTGATGCTGATGCTCTTGACT
5AQ_CS12224M17Q TGGGGATGCTCCACCCGACGCCGAAAATGAGGGCAATTGTTGATGCTGATGCTCTTGACT

5Bq_CS102N4QEXONS TGGATCTTCGGATGTCGCAACCTACCGCGCACGATCCCAAGCGGGACAACATCGTCGCCG
5Dq_CS201F2QEXONS TGGATCTACGGATGTCGCAACCTACCGCGCACGATCCCAAGCGGGACAACATCATCGCCG
5AQ_CS12224M17Q TGGATCTGCGGATGTCGCAACCCACCGCGCACGATCCCAAGAGGGACAACATCATCGCCG

Amorce 5AQ/q-A43F

5Bq_CS102N4QEXONS GCCTTCAGTTAACTTTTGA**TTCCCTGAATCGTCAACCAC**AATGCTCTCTTCTCAGAGCC
5Dq_CS201F2QEXONS GCCTTCAGTTAACTTTTGAATTCCTTGAATCGTCAACCACAATGCTCTCTTCTCAGAGCC
5AQ_CS12224M17Q GCCTTCAGTTAACTTTTGAATTCCTTGAATCGTCAACCACAATGATCTCTTCTCAGAGCC

5Bq_CS102N4QEXONS AATGAGCTCATCTTCTTCCAGTGGCCTGTGCATCAACATGGCAGGCGAGTACCACCTCA
5Dq_CS201F2QEXONS AATGAGCTCATCTTCTTCCAGTGGCCTGTGCATCAACATGGCAGGCGAGTAGCACCTCA
5AQ_CS12224M17Q AATGAGCTCATCTTCGTCCAGTGGCCTGTGCATCAACATGGCAGGCGAGTAGCACCTCA

Amorce 5AQ/q-A45F

5Bq_CS102N4QEXONS GCAGCACCAGCGTTTGTACCCATCTG**CTGTGTCATGGCTTCTACCCGAA**CGTACAGGTACAG
5Dq_CS201F2QEXONS GCAGCACCAGCGTTTGTACCCATCTGCTTGTGTCATGGCTTCTACCCGAACGTACAGGTACAG
5AQ_CS12224M17Q GCAGCACCAGCGTTTGTACCCATCTGCTTGTGTCATGGCTTCTACCCGAACGTACAGGTACAG

5Bq_CS102N4QEXONS GTGCAGGTGCAGGAGAGGCCGATGGAGGCAAGGCCCCCTGAGCCGCCGTCGTGCTTCCCC
5Dq_CS201F2QEXONS GTGCAGGTGCAGGAGAGGCCCATGGAGGCAAGGCCCCCTGAGCCGCCGTCGTGCTTCCCC
5AQ_CS12224M17Q GTGCAGGTGCAGGAGAGGCCCATGGAGGCAAGGCCCCCTGAGCAGCCGTCGTCTTCCCC

Amorce 5AQ/q-A46F-A47F

5Bq_CS102N4QEXONS GGCTGGGGGTGGCAAGCGCAAGCCACGCCGCCGGCTCCTCCACTCGCCGTTGCTTTAC
5Dq_CS201F2QEXONS AGCTGGGGGTGGCAAGCGCAAGCCATGCCGCCGGCTCCTCCACTCGCCGTTGCTTTAC
5AQ_CS12224M17Q GGCTGGGGGTGGCAAGCGCAAGCCATGCCGCCGG**CTCCTCCCACTCGCCGTTGCTTTAC**

miR172 site d'adressage du MIR172

5Bq_CS102N4QEXONS **GCTGCAGCATCATCAGGATTCT**TACCGCCGCCG---GCGCGACCCCGCCCGCCGCCG
5Dq_CS201F2QEXONS **GCTGCAGCATCATCAGGATTCT**TACCGCCGCCG---GCGCGACCCCGCCCGCCGCCG
5AQ_CS12224M17Q **GCTGCAGCATCATCAGGATT**TACCGCCGCCGCCGCCGCGCAACCTCGCCCGCCGCCG

Annexe 5 : Tableau des amorces PCR et RACE-PCR pour l'étude de la régulation des gènes 5AQ/q par le miR172.

En bleu pour le génotypage des *Hélitrons*, en jaune pour le RACE-PCR

Hélitron/Gene	Nom des amorces	Sequence 5'→3'
Q-PCR		
5AQ/q Renan interne	Hel-Renan_A21F	CTGATGATGATGCGTGATT
5AQ/q Renan interne	Hel-Renan_A21R	TACTGTATATGGCATCTGATTAGAG
A73 groupe 1 externe	Hel_CS-G1_A73F	GGTGTCCGTAAC TAGCAACAA
A73 groupe 1 externe	Hel_CS-G1_A73R	GGCTTCTCTCCTTTTATCCAAC
A74 groupe 1 interne	Hel_CS-G1_A74F	GTGTCCGTAAC TAGCAACAA
A74 groupe 1 interne	Hel_CS-G1_A74R	ATAGTTTTTCATCCGTACATACC
A75 Renan externe	Hel_CS-G1_A75F	CTCCATCGCCGTAATCGAC
A75 Renan externe	Hel_CS-G1_A75R	AACATGAACCAAATGACGGAA
A79 groupe 2 interne	Hel_CS-G1_A79F	CGATTTAATCTTTGTCATGCCTT
A79 groupe 2 interne	Hel_CS-G1_A79R	ACACGTCTCCACCAATACACA
A82 groupe 3 externe	Hel_CS-G1_A82F	TTTCTTTCAGGCTTCAATCTGG
A82 groupe 3 externe	Hel_CS-G1_A82R	AACTGTCGAAC TCGTGCCTA
A83 groupe 3 interne	Hel_CS-G1_A83F	GAGACTAATATTATCAGGCACTC
A83 groupe 3 interne	Hel_CS-G1_A83R	TGGCATGAACCAA ACTTAGACA
A86 groupe 2 externe	Hel_CS-G1_A86F	GCATATTGTTGTGCATCG
A86 groupe 2 externe	Hel_CS-G1_A86R	ACGGAAGGAGTACTAAACA
5AQ/q	5AQ_HB43F	TTCCCCTGAATCGTCAACCAC
5AQ/q	5AQ_HB43R	AGACCAGCCAGTAGTGTCACC
5AQ/q	5AQ_HB44F	TCCCCTGAATCGTCAACCAC
5AQ/q	5AQ_HB44R	CTAATGAACAGCCATGATCCG
5AQ/q	5AQ_HB45F	CTTGTCATGGCTTCTACCCGAAC
5AQ/q	5AQ_HB45R	AAGTAGAACC GGTGGTGTC
5AQ/q	5AQ_HB46F	CTCCTCCC ACTCGCCGTTG
5AQ/q	5AQ_HB46R	AGACCAGCCAGTAGTGTCACC
5AQ/q	5AQ_HB47F	TCCTCCC ACTCGCCGTTGCTT
5AQ/q	5AQ_HB47R	GCACCCACCGCCGTCTGGTCA
RACE OligodT	GCTGTCAACGATACGCTACGTAACGGCATGACAGTGTTTTTTTTTTTTTTTTTTT	
RACE RNA-Oligo	CGACUGGAGCACGAGGACACUGACAUGGACUGAAGGAGUAGAAA	
RACE GeneRacer_5		CGACTGGAGCACGAGGACACTGA
RACE GeneRacer_5nested		GGACACTGACATGGACTGAAGGAGTA
RACE GeneRacer_3		GCTGTCAACGATACGCTACGTAACG
RACE GeneRacer_3nested		CGCTACGTAACGGCATGACAGTG

Annexe 6 : Liste des plantes et accèsions à l'origine de la variété Renan.

Grains fournis par le Centre de Ressources Génétiques (CRG) INRA des Céréale à Paille de Clermont-Ferrand.

Code ERGE	Nom de la variété
2358	COURTOT
4594	MARIS HUNTSMAN
4625	MARNE DESPREZ
4856	MIRONOVSKAIA
4907	MOISSON VILMORIN
26258	Ventricosa - 1
26259	Ventricosa - 2
26260	Ventricosa - 3
26261	Ventricosa - 4
26262	Ventricosa - 5
26263	Ventricosa - 6
26264	Ventricosa - 7
26265	Ventricosa - 8
26266	Ventricosa - 9
26267	Ventricosa - 10
26839	CA.25V (T. carthlicum)

Annexe 7 : Liste des plantes et accessions Ukrainienne étudié dans la recherche de l'origine de l'*Hélitron* dans le gène *5Bq*

Grains fournis par le Centre de Ressources Génétiques (CRG) INRA des Céréale à Paille de Clermont-Ferrand.

ERGE	Nom	Pays	Obtenteur	Inscription
29940	CRIMEAN	UKR		1873
24029	KRYMKA	UKR		1875
4102	KHARKOF	UKR		1900
24043	BANATKA	UKR		1924
7099	UKRAINKA	UKR		1929
8056	NOVOKRYMKA 102	UKR		1929
13379	UKRAMCA	UKR		1929
24038	UKRAINKA	UKR	INST. MIRONOVSKYI	1929
24044	DYURABL	UKR		1929
24046	ZEMKA	UKR		1929
24048	KOOPERATORKA	UKR		1929
24074	FERRUGINEUM 1239	UKR		1933
24078	SANDOMIRKA	UKR		1933
4301	LESOSTEPKA	UKR		1938
7344	VESELOPODOLYANSKAIA	UKR	INST. POLTAVA	1938
7345	VESELOPODOLYANSKAIA	UKR		1938
29964	ARTEMOVKA	UKR		1945
5437	ODESSKAYA 12	UKR		1947
5438	ODESSKAYA 16	UKR	INST. ODESSA	1953
36085	VESELOPODOLYANSKAJA 499	UKR	INST. POLTAVA	1954
4864	MIRONOVSKAIA 264	UKR	INST. MIRONOVSKYI	1960
33970	BELOTSEKOVSKAJA-23	UKR		1962
4866	MIRONOVSKAYA 808	UKR	INST. MIRONOVSKYI	1963
4870	MIRONOVSKAIA 808 MUT.96-3-1	UKR		1963
4861	MIRONOVSKAIA 25	UKR		1980
13673	MIRONOVSKAIA 61	UKR		1989
20074	MIRLEBEN NARODNAYA	UKR	INST. MIRONOVSKYI	1993

ERGE	Nom	Statut	Pays	Type	inscription
1196	BANATKA	Variété ancien	HUN	Hiver	1930
1197	BANATKA KRASSAWA	Variété modern	YUG	Hiver	*
1198	BANATKA RAWSKA A	Variété ancien	POL	Hiver	1920
5295	NOVA BANATKA	Variété modern	YUG	Hiver	1973
6708	STARA BANATKA	Variété ancien	YUG	Hiver	*
24026	BANATKA POZDNYAYA	Variété ancien	BLR	Hiver	1908
24028	BANATKA	Variété ancien	RUS	Hiver	1911
24039	BANATKA	Variété ancien	RUS	Hiver	1924
24043	BANATKA	Variété ancien	UKR	Hiver	1924
24047	BANATKA	Variété ancien	HUN	Hiver	1924
24071	BANATKA	Variété ancien	POL	Hiver	1930
24085	BANATKA	Variété ancien	YUG	Hiver	1930
24200	BANATKA	Variété ancien	MDA	Hiver	1951

Annexe 8 : Photos des lignées Russes positives pour l'insertion de l'*Hélitron* dans le gène *5Bq* de cv. Renan



T. aestivum cv. Renan
Accession 6086 - inscrit en 1986



T. aestivum cv. Mironovskaya 808
Accession 4866 - inscrit en 1963



T. aestivum cv. Lesostepka
Accession 4301 – inscrit en 1938
Accession 24039 – inscrit en 1924



T. aestivum cv. Ukrainka
Accession 7099 – inscrit en 1929



T. aestivum cv. Kooperatorka
Accession 24048 – inscrit en 1929



T. aestivum cv. banatka

Annexe 9 : Article **Charles, M. *et al.*** (2008) Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. *Genetics* **180**, 1071–86

Dynamics and Differential Proliferation of Transposable Elements During the Evolution of the B and A Genomes of Wheat

Mathieu Charles,* Harry Belcram,* Jérémy Just,* Cécile Huneau,* Agnès Viollet,[†]
Arnaud Couloux,[†] Béatrice Segurens,[†] Meredith Carter,[‡] Virginie Huteau,[§]
Olivier Coriton,[§] Rudi Appels,[‡] Sylvie Samain[†] and Boulos Chalhoub^{*,1}

*Organization and Evolution of Plant Genomes, Unité de Recherche en Génomique Végétale, UMR: INRA-1165, CNR-S8114, 91057 Evry Cedex, France, [†]CEA: Institut de Génomique GENOSCOPE, 91057 Evry Cedex, France, [‡]State Agricultural Biotechnology Centre and Centre for Comparative Genomics, Murdoch University, Perth, Western Australia 6150, Australia and [§]Unité Mixte de Recherches INRA, Agrocampus Rennes Amélioration des Plantes et Biotechnologies Végétales, 35653 Le Rheu, France

Manuscript received June 6, 2008

Accepted for publication August 7, 2008

ABSTRACT

Transposable elements (TEs) constitute >80% of the wheat genome but their dynamics and contribution to size variation and evolution of wheat genomes (*Triticum* and *Aegilops* species) remain unexplored. In this study, 10 genomic regions have been sequenced from wheat chromosome 3B and used to constitute, along with all publicly available genomic sequences of wheat, 1.98 Mb of sequence (from 13 BAC clones) of the wheat B genome and 3.63 Mb of sequence (from 19 BAC clones) of the wheat A genome. Analysis of TE sequence proportions (as percentages), ratios of complete to truncated copies, and estimation of insertion dates of class I retrotransposons showed that specific types of TEs have undergone waves of differential proliferation in the B and A genomes of wheat. While both genomes show similar rates and relatively ancient proliferation periods for the *Athila* retrotransposons, the *Copia* retrotransposons proliferated more recently in the A genome whereas *Gypsy* retrotransposon proliferation is more recent in the B genome. It was possible to estimate for the first time the proliferation periods of the abundant *CACTA* class II DNA transposons, relative to that of the three main retrotransposon superfamilies. Proliferation of these TEs started prior to and overlapped with that of the *Athila* retrotransposons in both genomes. However, they also proliferated during the same periods as *Gypsy* and *Copia* retrotransposons in the A genome, but not in the B genome. As estimated from their insertion dates and confirmed by PCR-based tracing analysis, the majority of differential proliferation of TEs in B and A genomes of wheat (87 and 83%, respectively), leading to rapid sequence divergence, occurred prior to the allotetraploidization event that brought them together in *Triticum turgidum* and *Triticum aestivum*, <0.5 million years ago. More importantly, the allotetraploidization event appears to have neither enhanced nor repressed retrotranspositions. We discuss the apparent proliferation of TEs as resulting from their insertion, removal, and/or combinations of both evolutionary forces.

GENOMES of higher eukaryotes, and particularly those of plants, vary extensively in size (BENNETT and SMITH 1976, 1991; BENNETT and LEITCH 1997, 2005). This is observed not only among distantly related organisms, but also between species belonging to the same family or genus (CHOOI 1971; JONES and BROWN 1976). More than 90% of genes are conserved in sequenced plant genomes (BENNETZEN 2000a; SASAKI *et al.* 2005; JAILLON *et al.* 2007) and thus differences in gene content explain only a small

fraction of the genome size variation. It is widely accepted that whole-genome duplication by polyploidization (BLANC *et al.* 2000; PATERSON *et al.* 2004; ADAMS and WENDEL 2005) and differential proliferation of transposable elements (TEs) are the main driving forces of genome size variation. The differential proliferation of TEs results from their transposition (SANMIGUEL *et al.* 1996; BENNETZEN 2000b, 2002a,b; KIDWELL 2002; BENNETZEN *et al.* 2005; HAWKINS *et al.* 2006; PIEGU *et al.* 2006; ZUCCOLO *et al.* 2007) as well as the differential efficiency of their removal (PETROV *et al.* 2000; PETROV 2002a,b; WENDEL *et al.* 2002).

Polyploidization and differential proliferation of TEs are particularly obvious in the case of wheat species belonging to the closely related *Triticum* and *Aegilops* genera. Rice (*Oryza sativa*), *Brachypodium*, and diploid *Triticum* or *Aegilops* species underwent the same whole-genome duplications (ADAMS and WENDEL 2005; SALSE

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AM932680, AM932681, AM932682, AM932683, AM932684, AM932685, AM932686, AM932687, AM932688, AM932689.

¹Corresponding author: Organization and Evolution of Plant Genomes, Unité de Recherche en Génomique Végétale, UMR: INRA-1165, CNR-S8114, 2 rue Gaston Crémieux, 91057 Evry Cedex, France.
E-mail: chalhoub@evry.inra.fr

et al. 2008), but *Triticum* or *Aegilops* genomes are >10 times larger (BENNETT and SMITH 1991), mainly due to proliferation of repetitive DNA, which represents >80% of the genome size (SMITH and FLAVELL 1975; VEDEL and DELSENY 1987). Diploid wheat species can differ in their genome sizes by hundreds or even thousands of megabases (BENNETT and SMITH 1976, 1991; <http://data.kew.org/cvalues/homepage.html>). For example, the genome size of *Triticum monococcum* (6.23 pg) is 1.3 pg greater than that of *Triticum urartu* (4.93 pg) (BENNETT and SMITH 1976, 1991), although these species diverged <1.5 million years ago (MYA) (DVORAK *et al.* 1993; HUANG *et al.* 2002; WICKER *et al.* 2003b). Similarly, the calculated size of the B genome of polyploid wheat species (7 pg) is higher than that of any diploid wheat species (<http://data.kew.org/cvalues/homepage.html>).

The genome size variation within wheat is also accentuated by frequent allopolyploidization events, among which two successive events have led to the formation of the allohexaploid bread wheat *Triticum aestivum* ($2n = 6x = 42$, AABBDD). The first event led to the formation of the allotetraploid *Triticum turgidum* ($2n = 4x = 28$, AABB) and occurred <0.5–0.6 MYA between the diploid species *T. urartu* ($2n = 2x = 14$, AA), donor of the A genome, and an unidentified diploid species of the Sitopsis section, donor of the B genome (FELDMAN *et al.* 1995; BLAKE *et al.* 1999; HUANG *et al.* 2002; DVORAK *et al.* 2006). The second allopolyploidization event occurred 7000–12,000 years ago, between the early domesticated tetraploid *T. turgidum* ssp. *dicoccum* and the diploid species *Aegilops tauschii* ($2n = 14$), donor of the D genome, resulting in hexaploid wheat (FELDMAN *et al.* 1995).

The amount of available wheat genomic sequences is very limited, compared to other organisms (reviewed by SABOT *et al.* 2005; STEIN 2007; <http://genome.jouy.inra.fr/triannot/index.php> and <http://www.ncbi.nlm.nih.gov/>). Individual bacterial artificial chromosome (BAC) clones, selected primarily because they contained genes of agronomic interest, have been sequenced. Analyses of randomly chosen BAC clones from wheat have been also performed (DEVOS *et al.* 2005), and 2.9 Mb of sequences from a whole-genome shotgun library of *Ae. tauschii* were analyzed by LI *et al.* (2004). More recently, a detailed analysis of 19,400 BAC-end sequences of chromosome 3B, representing a cumulative sequence length of nearly 11 Mb (1.1% of the estimated chromosome length) was reported (PAUX *et al.* 2006). Altogether, these sequencing efforts have confirmed previous estimates of the amount of repetitive DNA in the wheat genome (~80%) (SMITH and FLAVELL 1975; VEDEL and DELSENY 1987) and have identified the major types of TEs (WICKER *et al.* 2002; SABOT *et al.* 2005).

Because of the limited genomic sequence information, the extent to which various TEs contribute to the

wheat genome and affect its size variation, or how they are distributed among different genomes, remains unexplored. Little is known about the dynamics of TEs, their proliferation processes, and whether they proliferated gradually or in waves of sudden bursts of insertions. In this study, 10 genomic regions from wheat chromosome 3B were sequenced and used to constitute, along with three other genomic sequences, 1.98 Mb of sequence from the wheat B genome. Transposable element dynamics and proliferation in these B-genome sequences were analyzed and compared to those in 3.63 Mb of sequence from 19 genomic regions of the wheat A genome. Our study provides novel insights into the dynamics and differential proliferation of TEs as well as their important role in the evolution and divergence of the wheat B and A genomes.

MATERIALS AND METHODS

Plant material and genomic DNA isolation: Hexaploid wheat deletion lines used to map the 10 BAC clones on different deletion bins of chromosome 3B (see RESULTS) were originally described by Qi *et al.* (2003) and kindly provided by Catherine Feuillet (INRA, Clermont-Ferrand, France). Hexaploid wheat genotypes were kindly provided by Joseph Jahier (INRA, Rennes, France). Tetraploid wheat genotypes were kindly provided by Moshe Feldman (Weizmann Institute). Genomic DNA was extracted from leaves as described by GRANER *et al.* (1990).

Primer design and PCR-based tracing of retrotransposon insertions: The program Primer3 (ROZEN and SKELETSKY 2000) was used to design oligonucleotide primers on the basis of TE–TE or TE-unassigned DNA junctions. We often designed and used several couples (including nested) of PCR primers. Internal controls (PCR primers designed within the TE) were also used. Primer sequences are given in supplemental Table 1. PCR reactions were carried out in a final volume of 10 μ l with 200 μ M of each dNTP, 500 nM each of forward and reverse primers, 0.2 units Taq polymerase (Perkin Elmer). PCR amplification was conducted using the following “touchdown” procedure: 14 cycles (30 sec 95°, 30 sec 72° minus 1° for each cycle, 30 sec 72°), 30 cycles (30 sec 95°, 30 sec 55°, 30 sec 72°), and one additional cycle of 10 min 72°. Amplification products were visualized using standard 2% agarose gels.

BAC sequencing, sequence assembly, and annotation: BAC shotgun sequencing was performed at the Centre National de Sequencage (Evry, France) essentially as described by CHANTRET *et al.* (2005). Genes, TEs, and other repeats were identified by computing and integrating results on the basis of BLAST algorithms (ALTSCHUL *et al.* 1990, 1997), predictor programs, and different software and procedures, detailed below. Cross-analysis of the information obtained for genes and TEs as well as for repeats and unassigned DNA was integrated into ARTEMIS (RUTHERFORD *et al.* 2000). Sequence annotation and analysis were performed as described in supplemental Method 1. The 10 BAC clone sequences were submitted to EMBL and under the following accession nos.: TA3B54F7, AM932680; TA3B63B13, AM932681; TA3B63B7, AM932682; TA3B81B7, AM932683; TA3B95C9, AM932684; TA3B95F5, AM932685; TA3B95G2, AM932686; TA3B63C11, AM932687; TA3B63E4, AM932688; TA3B63N2, AM932689. Accession numbers for the three publicly available genomic sequences

from the wheat B genome (SABOT *et al.* 2005; GU *et al.* 2006; DVORAK *et al.* 2006) are CT009588, AY368673, DQ267103.

Publicly available genomic sequences from the wheat A genome: The retained publicly available A-genome sequences consist of 19 sequenced and well annotated BAC clones or contigs (SANMIGUEL *et al.* 2002; YAN *et al.* 2002, 2003; WICKER *et al.* 2003b; CHANTRET *et al.* 2005; ISIDORE *et al.* 2005; DVORAK *et al.* 2006; GU *et al.* 2006; MILLER *et al.* 2006), representing >3.5 Mb. Accession numbers for the analyzed BAC sequences are the following: diploid A genome—AF326781, AF488415, AY146588, AY188331, AY188332, AY188333, AY491681, AY951944, AY951945, DQ267106, AF459639; tetraploid A genome—AY146587, AY485644, AY663391, CT009587, DQ267105; hexaploid A genome—AY663392, CT009586, DQ537335.

Chromosome 3B BAC clones and fluorescent *in situ* hybridization: The 10 BAC clones and/or their subclones were originally mapped by fluorescence *in situ* hybridization (FISH) on flow-sorted 3B chromosomes using the Cot-1 fraction as blocking DNA to suppress hybridization of repeated sequences (DOLEZEL *et al.* 2004; SAFAR *et al.* 2004; M. KUBALAKOVA and J. DOLEZEL, personal communication). Further FISH hybridization experiments were conducted, without Cot-1 DNA, on mitotic metaphase chromosomes of hexaploid wheat (*T. aestivum*) cv. Chinese Spring. The FISH hybridization protocol is presented in supplemental Method 2.

Estimation of Long Terminal Repeat-retrotransposon insertion dates: For all genomic sequences of the B and A genomes of wheat, retrotransposon copies with both 5' and 3' long terminal repeats (LTRs), and target-site duplications (TSD) were considered as corresponding to original insertions and analyzed by comparing their 5' and 3' LTR sequences. The two LTRs were aligned and the number of transition and transversion mutations was calculated using MEGA3 software (KUMAR *et al.* 2004). A mutation rate of 1.3×10^{-8} substitutions/site/year (SANMIGUEL *et al.* 1998; MA *et al.* 2004; MA and BENNETZEN 2004; WICKER *et al.* 2005; GU *et al.* 2006) was used. The insertion dates and their standard errors (SE) were estimated using the formula $T = K2P/2r$ (KIMURA 1980).

Statistical analysis: All statistical analyses and the different tests (Kolmogorov–Smirnov, Bootstrap, and probability density functions) were done with the R-package (<http://www.r-project.org>). Kolmogorov–Smirnov tests (FÉRIGNAC 1962) were applied to check whether the distribution of insertion dates of retrotransposons deviates from uniformity, and whether they are different when comparing different TE families or superfamilies within and between the B and A genomes. Probability density of TE insertion dates was estimated using Gaussian kernel density estimation (SILVERMAN 1986), taking into account measured standard deviation for each individual insertion date (KIMURA 1980).

RESULTS

Constitution of a genomic sequence data set representative of the wheat B genome—analysis of 10 BAC sequences from the wheat chromosome 3B: Only three large well-annotated genomic sequences (BAC clones), representing 0.55 Mb of sequence, were available for the wheat B genome (SABOT *et al.* 2005; DVORAK *et al.* 2006; GU *et al.* 2006). To obtain more representative genomic sequences, we sequenced and annotated 10 BAC clones of wheat chromosome 3B, representing 0.15% of the chromosome length (1.43 Mb) (Figure 1). Detailed

annotation files are deposited at EMBL/GenBank Data Libraries.

These sequenced genomic regions show a high proportion of TEs, which represent 79.1% of the cumulative sequence length (Figure 1, supplemental Table 2). Other repeated DNA sequences represent 2.4% and unassigned DNA sequences account for 17.5% of the cumulative sequence length.

We conducted gene prediction analysis for the remaining 18.5% non-TEs and nonrepeated DNA, using different search programs (see supplemental Method 1 and supplemental Text 1 for detailed description). Genes of known and unknown functions or putative genes were defined on the basis of predictions and the existence of rice or other Triticeae homologs. Hypothetical genes were identified on the basis of prediction programs only. Pseudogenes were not well predicted and frameshifts need to be introduced within the coding sequences (CDS) structure to better fit a putative function on the basis of BLASTX (mainly with rice). Truncated pseudogenes (genes disrupted by large insertion or deletion) and highly degenerated CDS sequences were considered as *gene-relics*. Combined together, all these types of gene sequence information (GSI) account for only 1.0% of the sequence and are present in seven BAC clones (one or two genes per clone) while the remaining three BAC clones (TA3B95C9, TA3B95G2, TA3B63N2) contain no genes (indicated in Figure 1A and detailed in supplemental Text 1, supplemental Table 3, and supplemental Table 4).

Six genes (of known or unknown function) and two putative genes were identified using the FGENESH prediction software (<http://www.softberry.com>) and by identification of homologs in rice (Figure 1A, supplemental Table 3). Six additional “gene-relics” or “pseudogenes” were also identified on the basis of colinearity with rice (Figure 1A, supplemental Table 3). Finally, 10 CDS, designated as “hypothetical genes,” were identified according to the FGENESH prediction program only (Figure 1A, supplemental Table 4).

TE prediction, annotation, classification, and nomenclature were performed essentially as suggested by the unified classification system for eukaryotic TEs (WICKER *et al.* 2007) with two modifications. The *Athila* retrotransposons were analyzed separately from the other *Gypsy* retrotransposons (see also supplemental Methods 1). The *Sukkula* retrotransposons were considered as belonging to the *Gypsy* superfamily because of similarities with the *Erika* (*Gypsy*) elements. The 79.1% of TEs were shown to be composed of a wide variety of TEs, distributed as follows: 61.9% class I (171 TEs from 48 families), 16.2% class II (113 TEs from 28 families), and 1.0% unclassified TEs (18 TEs from 9 families) (Figure 1). The *CACTA* TEs represent the majority (96%) of class II TEs. More details about the TE composition in the 10 different BAC clones of wheat chromosome 3B are provided in supplemental Text 2.

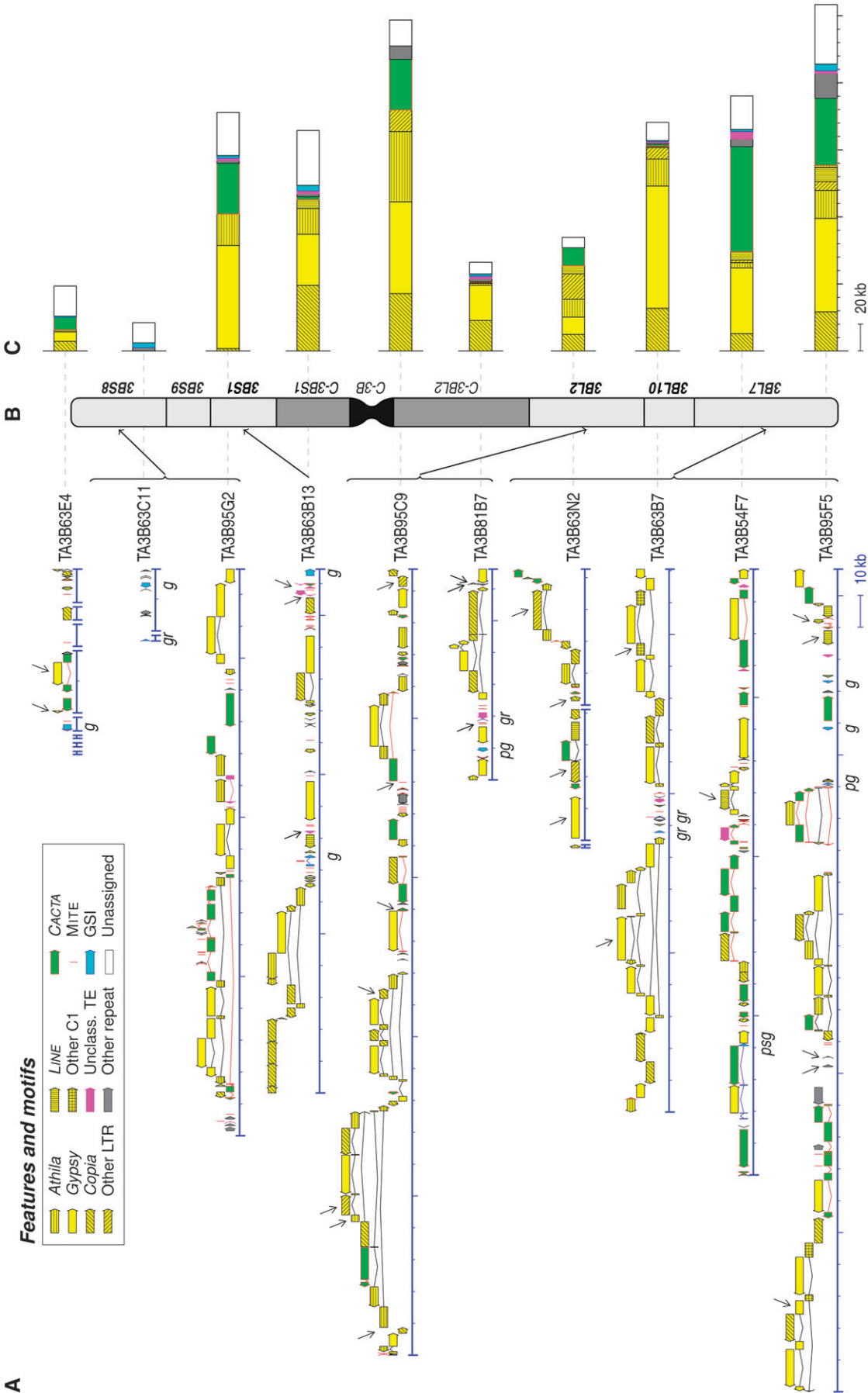


FIGURE 1.—Detailed annotation, BIN map positions, and sequence composition of 10 sequenced BAC clones of wheat chromosome 3B. (A) Detailed annotations of the 10 sequenced BAC clones. Main TEs, other repeats, and gene sequence information (GSI) are represented with distinct features and motifs (detailed in the “features and motifs” key). *g*, genes; *pg*, putative genes; *gr*, gene relics; and *pg*, pseudogenes. For nested insertions of TEs, the newly inserted TE is presented above the split one. Complete reconstruction of split TEs was done and the different parts are linked with a line to visualize the entire element. Some BAC clones are represented by several unordered contigs (TA3B63E4, TA3B63C11, TA3B63N2). EMBL BAC clone references and annotation files are given in MATERIALS AND METHODS. Detailed coding sequence and TE descriptions are supplied in supplemental Text 1 and supplemental Table 2. Arrows indicate novel TEs identified in this study and described in supplemental Text 2 and supplemental Table 5. (B) BIN map position of nine of the BAC clones. The wheat chromosome 3B bins are according to Qi *et al.* (2003). Details of the genotyping results are given in supplemental Table 6. (C) Proportions of the main sequence classes and types. See “features and motifs” in A for an explanation of colors. Details are given in supplemental Table 2.

TABLE 1

Details of TEs from the four most represented superfamilies in 13 genomic regions of the wheat B genome, compared to publicly available sequences from 19 genomic regions of the wheat A genome

	13 genomic regions of the wheat B genome (1.98 Mb) ^a				19 publicly available genomic regions of the wheat A genome (3.63 Mb) ^b			
	<i>Athila</i>	<i>Copia</i>	<i>Gypsy</i>	<i>CACTA</i>	<i>Athila</i>	<i>Copia</i>	<i>Gypsy</i>	<i>CACTA</i>
Observed number of TEs	54	57	79	70	72	149	123	53
Sequence proportion (means $\pm$ SE) % ^c	10.8 $\pm$ 1.6	14.2 $\pm$ 2.5	28.1 $\pm$ 3.8	13.4 $\pm$ 3.3	10.4 $\pm$ 1.8	21.8 $\pm$ 1.8	19.7 $\pm$ 2.9	9.4 $\pm$ 1.9
Bootstrap means deviation ^d	-0.07	+0.02	+0.02	-0.05	+0.01	-0.02	-0.03	-0.09
Complete TEs with TSD (%)	13	18	39	19	19	60	38	32
Incomplete (truncated) TEs	41	39	40	51	53	89	85	21
LTR-mediated homologous recombination								
Entire TE without TSD	3	7	0	—	0	4	0	—
Solo LTR	4	2	2	—	5	15	9	—
Illegitimate recombination	34	30	38	51	48	70	76	21
Complete TEs/incomplete (truncated) TEs	0.32	0.46	0.98	0.37	0.36	0.67	0.45	1.52

^aThis corresponds to 1.43 Mb from the 10 genomic regions sequenced in this study and 0.55 Mb from three other publicly available genomic regions from SABOT *et al.* (2005), GU *et al.* (2006), and DVORAK *et al.* (2006). See MATERIALS AND METHODS for BAC clone sequence references.

^bNineteen genomic regions available for the A genome (SANMIGUEL *et al.* 2002; YAN *et al.* 2002, 2003; WICKER *et al.* 2003b; CHANTRET *et al.* 2005; ISIDORE *et al.* 2005; DVORAK *et al.* 2006; GU *et al.* 2006; MILLER *et al.* 2006). See MATERIALS AND METHODS for BAC clone sequence references.

^cRelative to cumulative sequence length. SE, standard errors for estimated means.

^dDifferences between arithmetic means (line above) and bootstrap analysis (EFRON 1979) with 10,000 resamplings.

Twenty-one transposable element families, some of which are present in several copies, were identified for the first time in this study (Figure 1A, indicated by arrows). They account for 9.8% by number and 7.9% by length of the overall sequences. Class I retrotransposons are the category for which we found the majority of novel TE families (17). Description of these novel TEs, their features, and the suggested nomenclature are presented in supplemental Text 2 and supplemental Table 5.

The 10 sequenced BAC clones or their subclones were originally mapped by FISH on flow-sorted 3B chromosomes, using the *C₀t* - 1 fraction as blocking DNA to suppress hybridization of repeated sequences (DOLEZEL *et al.* 2004; SAFAR *et al.* 2004; M. KUBALAKOVA and J. DOLEZEL, personal communication). As described by DEVOS *et al.* (2005) and PAUX *et al.* (2006), specific PCR markers, based on TE-TE or TE-unassigned DNA junctions, were used to confirm the different BAC clone map positions on the deletion bins (Q1 *et al.* 2003) of chromosome 3B (except TA3B63E4) (Figure 1B). Details of PCR markers and genotyping results are given in supplemental Table 6.

Representation of transposable elements and the wheat B genome: Five BAC clone sequences were publicly available from the B genome of wheat (SABOT *et al.* 2005; DVORAK *et al.* 2006; GU *et al.* 2006). Four of these were sequenced for two orthologous regions in tetraploid and hexaploid wheat species (one BAC clone

per region and per species) (SABOT *et al.* 2005; GU *et al.* 2006). As they share nearly identical sequences (99%) with common TE insertions, they were considered as redundant in our study and only the longest BAC clone sequences (three in total) were counted in calculation and appreciation of TE proliferation. These, added to the above-described 10 genomic region sequences of wheat chromosome 3B, constitute 1.98 Mb of sequence from the wheat B genome. Four main TE superfamilies occupy 66.5% of the analyzed B-genome loci: the *Athila* superfamily (54 elements), the *Copia* superfamily (57 elements), the *Gypsy* superfamily (79 elements), and the *CACTA* superfamily (70 elements) (Table 1). Interestingly, proportions of the *Athila*, *Copia*, and *Gypsy* retrotransposons (respectively, 10.8, 14.2, and 28.1%) (Table 1) are very similar to estimates based on 11 Mb of the chromosome 3B sequence BAC end (PAUX *et al.* 2006). The major deviation concerns the proportion of *CACTA* class II TEs, which is higher in the 13 genomic regions (13.4%) than in the overall BAC-end sequences (4.9%), probably due to their clustering in some BAC clones that we have sequenced, such as TA3B54F7 (40.5% of *CACTA* TEs) (Figure 1).

The 13 sequences represent only ~0.03% of the B genome. However, statistical tests, using SE as well as a bootstrap analysis with 10,000 resamplings, confirm the robustness of estimations of sequence proportions of the *Gypsy*, *Copia*, *Athila*, and *CACTA* TE superfamilies (Table 1). We also evaluated the variation of mean

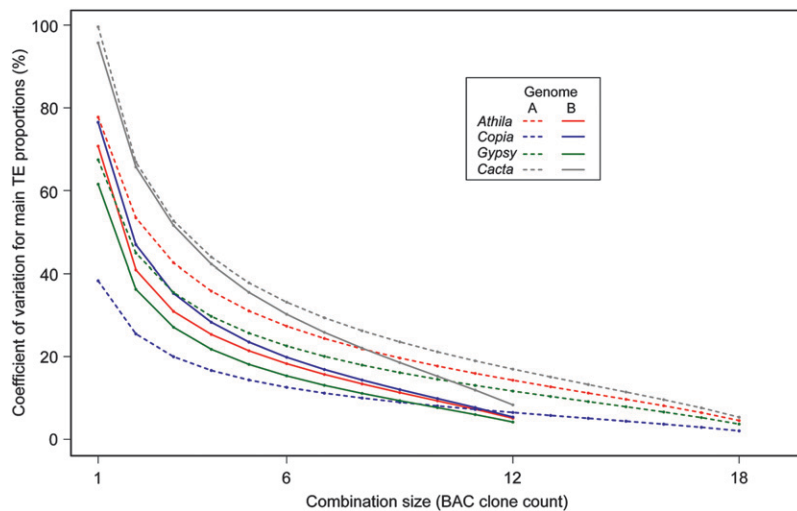


FIGURE 2.—Changes of the coefficient of variation of proportions (in percentages) of the main transposable element superfamilies calculated over all possible BAC clone combinations and simulated over a size varying from 1 to 12 BAC clones for the wheat B genome and 1 to 18 for the wheat A genome (combination size). For each number of considered BAC clones (x -axis), sequence proportions (in percentages) were calculated for all possible BAC clone combinations, and the coefficient of variation between these proportions was calculated (y -axis).

sequence proportions estimated for the four TE superfamilies by comparing all possible clone number representations and combinations (from 1 to 12 BAC clones) (Figure 2). Results show that representing the wheat B genome with a low number of BAC clones results in very variable proportions of the TE sequences (Figure 2). These variations decrease significantly by increasing the number of considered BAC clones (Figure 2). This confirms the usefulness of our effort in sequencing more BAC clones for better representation of the wheat B genome.

It is also interesting to note that direct FISH hybridization, using the whole BAC clone as a probe, resulted in dispersed and mostly homogenous signals across all wheat chromosomes for 8 of all 10 BAC clones of wheat chromosome 3B (except TA3B63C11 and TA3B54F7) (SAFAR *et al.* 2004 and supplemental Figure 1), thus confirming sequencing results that show high TE composition.

Constitution of a genomic sequence data set representative of the wheat A genome: The publicly available A-genome sequences that we were able to use are more abundant and consist of 20 sequenced and well-annotated BAC clones or contigs. Ten of these were comparatively sequenced for five orthologous regions of the wheat A genome at the diploid, tetraploid, and/or hexaploid levels and were partially overlapping (WICKER *et al.* 2003b; CHANTRET *et al.* 2005; ISIDORE *et al.* 2005; DVORAK *et al.* 2006; GU *et al.* 2006), while others were determined at only one ploidy level (mostly diploid) (SANMIGUEL *et al.* 2002; YAN *et al.* 2002, 2003; MILLER *et al.* 2006). Comparisons show that no shared TE insertions were observed between orthologous regions (from two ploidy levels), except in the region of the high-molecular-weight (HMW) glutenin gene, the sequences of which were nearly identical at the tetraploid and hexaploid levels (GU *et al.* 2006). Thus, we used only the sequence from hexaploid wheat to

represent the HMW glutenin gene region and considered all the other different orthologous regions (from different ploidy levels) separately. This led to 19 BAC clones, representing 3.63 Mb of sequence, that were analyzed for the wheat A genome.

The *Gypsy* TEs were found to occupy 19.7%, the *Athila* TEs 10.4%, the *Copia* TEs 21.8%, and the *CACTA* TEs 9.4% of the cumulative sequence length (Table 1). Similarly, for the B-genome sequences, we also analyzed and validated the robustness of the estimation of sequence proportions of the main TE superfamilies and their representation of the A genome (Figure 2). Similar proportions of the *Gypsy*, *Copia*, *Athila*, and *CACTA* TEs were found whether the 11 genomic sequences from the diploid A genome or those determined from A genomes of tetraploid (six regions) and hexaploid (three regions) wheat species were considered separately or combined (data not shown).

Comparison of TE sequence proportions and ratios of complete to truncated copies: Our analysis showed a significantly higher number of *Gypsy* retrotransposons in the wheat B-genome sequences than in the A genome (Table 1). Conversely, a higher proportion of *Copia* retrotransposons is observed in genomic sequences of the wheat A genome than in the B genome (Table 1). Proportions of the *Athila* and *CACTA* TEs were not statistically different between the two genomes (Table 1).

Major differences were found between the three main retrotransposon superfamilies in the ratio of complete (intact) copies, defined as having both LTRs and target-site TSD, as compared to degenerated and truncated copies that resulted from LTR-mediated unequal homologous recombinations or illegitimate DNA recombination (DEVOS *et al.* 2002; MA *et al.* 2004; MA and BENNETZEN 2004; VITTE and BENNETZEN 2006) (Table 1). In the B-genome sequences, the *Athila* and *Copia* retrotransposons show low ratios of complete to incomplete retrotransposons (respectively, 0.32 and 0.46),

whereas the *Gypsy* retrotransposons show the highest ratio (0.98) (Table 1). In comparison, the 3.63 Mb of genomic sequence of the wheat A genome shows a lower ratio (0.45) of complete to incomplete *Gypsy* retrotransposons whereas proportions of intact *Copia* retrotransposons are relatively higher than those observed in the B genome (0.67) (Table 1). The *Athila* retrotransposon ratio in the A genome is comparable to the ratio in the B genome (0.36 and 0.32, respectively).

CACTA TE original insertions are characterized by the “CACTA” sequence and 3-bp TSD sequence motifs surrounding terminal inverted repeats (TIR) at both ends. We used these signatures to define complete *CACTA* copies, where the “CACTA,” TIR, and TSD sequence motifs are observed at both ends, and truncated copies, where the “CACTA” and TSD motifs are absent from one or both ends. The ratio of complete to incomplete copies of the *CACTA* class II TEs was about five times lower in the wheat B genome (ratio of 0.37) than in the A genome (ratio of 1.52) (Table 1).

Insertion dates and proliferation of LTR retrotransposons: To understand differences in sequence proportions and the ratios of complete to truncated copies between retrotransposon superfamilies, as well as between the B and A genomes, we compared TE proliferation periods and rates.

The two LTRs are identical at the time of retrotransposon insertion and their sequence divergence reflects time lapsed since the insertion (SANMIGUEL *et al.* 1998). Several studies have shown that LTRs evolve at approximately twice the rate of genes and UTR regions, and we used a rate of 1.3×10^{-8} substitutions/site/year (MA *et al.* 2004; MA and BENNETZEN 2004; WICKER *et al.* 2005; GU *et al.* 2006).

We calculated the LTR divergence and dates of insertion of the *Athila*, *Copia*, and *Gypsy* retrotransposon (complete copies with both LTRs and TSD) found in the wheat B and A genomes (Figure 3). Such TE insertion dates offer a very important insight into the relative timing of various events, regardless of the approaches used to estimate nucleotide substitution rates or the molecular clock calibration points used in these calculations.

The vast majority of complete retrotransposons in the B and A genomes of wheat (86 and 92%, respectively) were estimated to be <3 million years old (Figure 2) in agreement with several previous studies of grasses and other plants species (SANMIGUEL *et al.* 1998, 2002; WICKER *et al.* 2003b, 2005; GAO *et al.* 2004; MA *et al.* 2004; DU *et al.* 2006; PIEGU *et al.* 2006; WICKER and KELLER 2007). This is explained by the fact that LTR retrotransposons are continuously removed by unequal homologous recombination and illegitimate DNA recombination as new ones are inserted (VICIENT *et al.* 1999; DEVOS *et al.* 2002; MA *et al.* 2004; PEREIRA 2004). Insertion of the *Egug* element (RLGa_ *Egug*_TA3B95C9-1 ~5 MYA; divergence of 0.131) is the oldest such event

found in our study and the most recent one is the *Sukkula* insertion (RLG_ *Sukkula*_TA3B63B7-2) for which only a 1-base indel differentiates the two LTRs of 4192/4193 bp.

Comparison of LTR divergence dates revealed that different LTR–retrotransposon superfamilies and families proliferated at different periods and rates during evolution of the wheat B and A genomes (Figure 3). We applied Kolmogorov–Smirnov tests to check whether within the last 3 million years (0.078 divergence) the distribution of insertion dates of retrotransposons deviates from uniformity (thus confirming a burst of higher proliferation), and whether these dates are different when comparing different retrotransposon families or superfamilies within and between the wheat B and A genomes (thus illustrating differential proliferation). This was done for all complete copies of the three main retrotransposon superfamilies as well as for the most abundant retrotransposon families (nine) that have five or more complete copies in the B and/or A genomes (Figure 3).

Superfamily level comparison: The combination of all complete retrotransposon copies at the superfamily level (Figure 3A) indicated that the distribution of the *Gypsy* retrotransposon insertion dates in both B and A genomes and that of *Copia* retrotransposons in the A genome were significantly different from uniform (P -value <0.01) because of their higher proliferation during the last 2 million years (Figure 3A). Proliferation of the *Copia* retrotransposons in the B genome was uniform and low all across the 3-million-year period, whereas proliferation of the *Athila* retrotransposons was different from a uniform distribution in both genomes at P -value <0.1.

One possible reason for the non-uniform distributions of retrotransposon insertion dates within the 3-million-year period is because older insertions are more likely to be removed (completely or partially) from the genome (see above). Therefore, we checked whether distributions of insertions are significantly different from a uniform distribution for the most recent period of evolution during which the impact of DNA removal should be lower. To carry out this analysis, we divided the LTR–retrotransposon insertions according to the median (of their distribution) that varies depending on the retrotransposon superfamily and family (Figure 3A, gray circle). Kolmogorov–Smirnov (FÉRIGNAC 1962) tests were then conducted on half of the complete copies, which show the most recent insertion dates. Distribution of insertion dates of the *Gypsy* retrotransposons in the wheat B genome and that of the *Copia* retrotransposons in the B and A genomes can be considered as uniform (P -value >0.05, Figure 3A), indicating that they have constantly proliferated during this most recent period. In contrast, the distribution of *Athila* retrotransposons in the wheat B and A genomes and that of *Gypsy* retrotransposons in the A genome

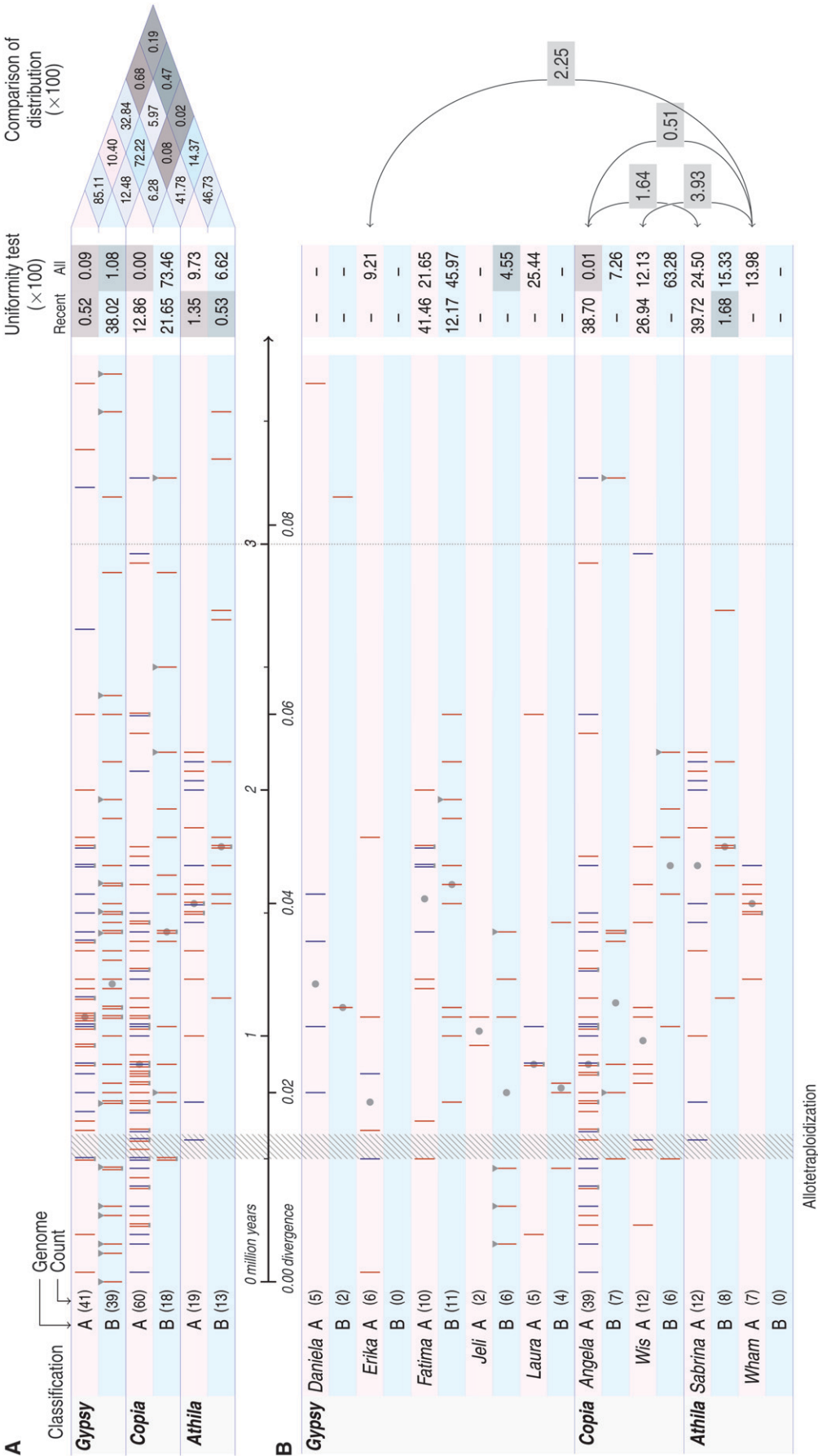


FIGURE 3.—Distribution of insertion dates estimated for LTR retrotransposons in the B and A-genome sequences of wheat (divergence and MYA). (A) All dated LTR retrotransposons combined at the three main superfamily levels (*Athila*, *Gypsy*, and *Copia*). (B) The most abundant retrotransposon families, showing five or more dated copies in at least one of the A or B wheat genomes. Mean insertion dates calculated for retrotransposons are represented by vertical bars. For the A-genome sequences, blue indicates retrotransposons detected from the diploid and red from the polyploid genomic sequences. The genomic sequences of the B genome (red) were obtained from the polyploid wheat. Copies of a given retrotransposon superfamily or family showing identical mean insertion dates are presented by adjacent vertical bars that are joined with a lower horizontal gray bar. The number within parentheses corresponds to the total number of considered retrotransposon copies. Gray triangles indicate retrotransposon insertions that have been traced using PCR in a collection of genotypes of *T. aestivum* and *T. turgidum*. The interval period of the allotetraploidization event (0.5–0.6 MYA, divergence 0.013–0.016) is highlighted in gray. “Uniformity test” refers to Kolmogorov–Smirnov (FERIGNAC 1962) tests determining probabilities (*P*value) that the distribution of insertion dates of retrotransposons deviates from uniformity (thus confirming a burst of higher proliferation): “All” refers to the last 3 million years (0.078 divergence); “Recent” refers to the most recent periods, estimated when dividing the LTR–retrotransposon insertions by the median (indicated by gray circles). Tests were done on families that show five copies or more. “Comparison of distribution” indicates the same Kolmogorov–Smirnov tests determining probabilities that distributions of insertion dates for the last 3 million years (0.078 divergence) are different in the retrotransposon superfamilies and families as well as the in B and A genomes of wheat.

are not uniform (P -value <0.05 , Figure 3A), consistent with a decreasing proliferation during the most recent period.

Comparison of the proliferation of the three retrotransposon superfamilies shows that distribution of the *Athila* retrotransposons is statistically different from that of the *Gypsy* retrotransposons (Figure 3A, P -value <0.05) in the B genome. The *Athila* distribution is significantly different from that of the *Gypsy* and *Copia* retrotransposons (Figure 3A, P -value <0.05) in the A genome.

Comparison of the distributions of the three retrotransposon superfamilies between the B and A genomes shows that *Copia* distributions are significantly different (Figure 3A, P -value = 0.628) due to their higher proliferation and more recent insertions in the A genome. Both genomes show similar old distribution of the *Athila* retrotransposons (Figure 3A). Distributions of the *Gypsy* retrotransposons were not statistically different between the two genomes for the entire 3-million-year period (Figure 3A, P -value >0.05). However, separate Kolmogorov–Smirnov tests for the most recent period show that these have proliferated less in the wheat A genome (P -value = 0.052, Figure 3A), unlike in the wheat B genome (P -value = 0.38, Figure 3A).

Distribution of the most abundant retrotransposon families: Some specific retrotransposon families were abundant in the B and/or A genomes. This is the case of the *Angela* and *Wis* families, together representing 72 and 85% of the *Copia* superfamily in the B and A genomes, respectively (Figure 3B). This is also the case of the *Sabrina* family representing 62 and 63% of the *Athila* superfamily in the B and A genomes, respectively (Figure 3B). There are more families that compose the *Gypsy* retrotransposon superfamily, the most abundant being *Fatima*, representing 25% in both genomes (Figure 3B).

Kolmogorov–Smirnov tests show nonsignificant deviations (P -value >0.05) from uniform distributions for all nine retrotransposon families (with five or more observed complete copies in at least one genome), with the exception of the *Jeli* (*Gypsy*) elements in the B genome and the *Angela* (*Copia*) elements in the A genome, which have more recently proliferated (Figure 3B). Separate analysis for the most recent period, corresponding to half of the complete copies, shows that, as expected from the superfamily-level analysis, the *Wham* family in the A genome and the *Sabrina* family in the B genome have not recently proliferated (P -value <0.05 , Figure 3B).

Distribution of insertion dates of the *Wham* and *Sabrina* families is different from almost all the other seven families within and between the B and A genomes (P -value <0.05). Distribution of insertion dates of the *Angela* family in the wheat A genome is statistically different (P -value <0.05) from that of the *Fatima* family in both genomes. Distributions of insertion dates of the

remaining families do not show statistical differences (P -value >0.05) within and between the wheat B and A genomes (Figure 3B).

Moreover, some retrotransposon families were abundant and present in several complete copies in only one genome (*Romani*, *Daniela*, *Erika*, and *Wham* for the A genome; *Egug* and *Jeli* for the B genome) but absent or presenting few copies in the other (Figure 3B). It is likely that this corresponds to differential proliferation of the considered retrotransposons, as different copies were detected in different genomic regions of wheat B or A genomes.

LTR-retrotransposon proliferation was neither enhanced nor repressed by the allotetraploidization event: The allotetraploidization event that brought the B and A genomes of wheat together in one nucleus was estimated to occur no more than 0.5–0.06 MYA (HUANG *et al.* 2002; DVORAK *et al.* 2006; CHALUPSKA *et al.* 2008). This corresponds to a divergence interval of 0.013–0.016, using the corrected rate of 1.3×10^{-8} substitutions/site/year for more rapid divergence of LTRs (MA *et al.* 2004; MA and BENNETZEN 2004; DVORAK *et al.* 2006).

Comparisons show that retrotransposon insertions continued in wheat B and A genomes during the last 0.5–0.6 million years, apparently without being enhanced nor repressed by the allotetraploidization event (Figure 3). For example, analysis of genomic sequences available from the three ploidy levels of the A genome does not show differences in proliferation periods and rates of retrotransposons (Figure 3).

To check the accuracy of these observations and to calibrate the divergence rate used for coding sequences, on one hand, and that used for LTRs of retrotransposons, on the other hand, we traced several retrotransposons for their insertion prior or posterior to the allopolyploidization event. A PCR-based tracing strategy, derived from the retrotransposon-based insertion polymorphism method (FLAVELL *et al.* 1998; DEVOS *et al.* 2005; PAUX *et al.* 2006), was developed for 21 retrotransposon insertions from the B genome, sampled as having different estimated insertion dates (Figure 3, indicated by gray triangles). It simply relies on primers designed in both the retrotransposon and its flanking sequences (either unassigned DNA or an older preinserted TE sequence) so that PCR amplification will be specific to the retrotransposon insertion. As the diploid wheat species donor of the B genome is unknown (FELDMAN *et al.* 1995; BLAKE *et al.* 1999; HUANG *et al.* 2002), we analyzed the occurrence (*i.e.*, presence or absence) of the 21 retrotransposon insertions in hexaploid (*T. aestivum*) and tetraploid (*T. turgidum*) wheat genotypes, which carry the wheat B genome. Examples of PCR-based tracing of the 21 original retrotransposon insertions in the wheat genotypes compared with their estimated insertion dates ($\pm$ SE) are presented in Figure 4. Full tracing results are supplied in supplemental Table 7 and sequences of

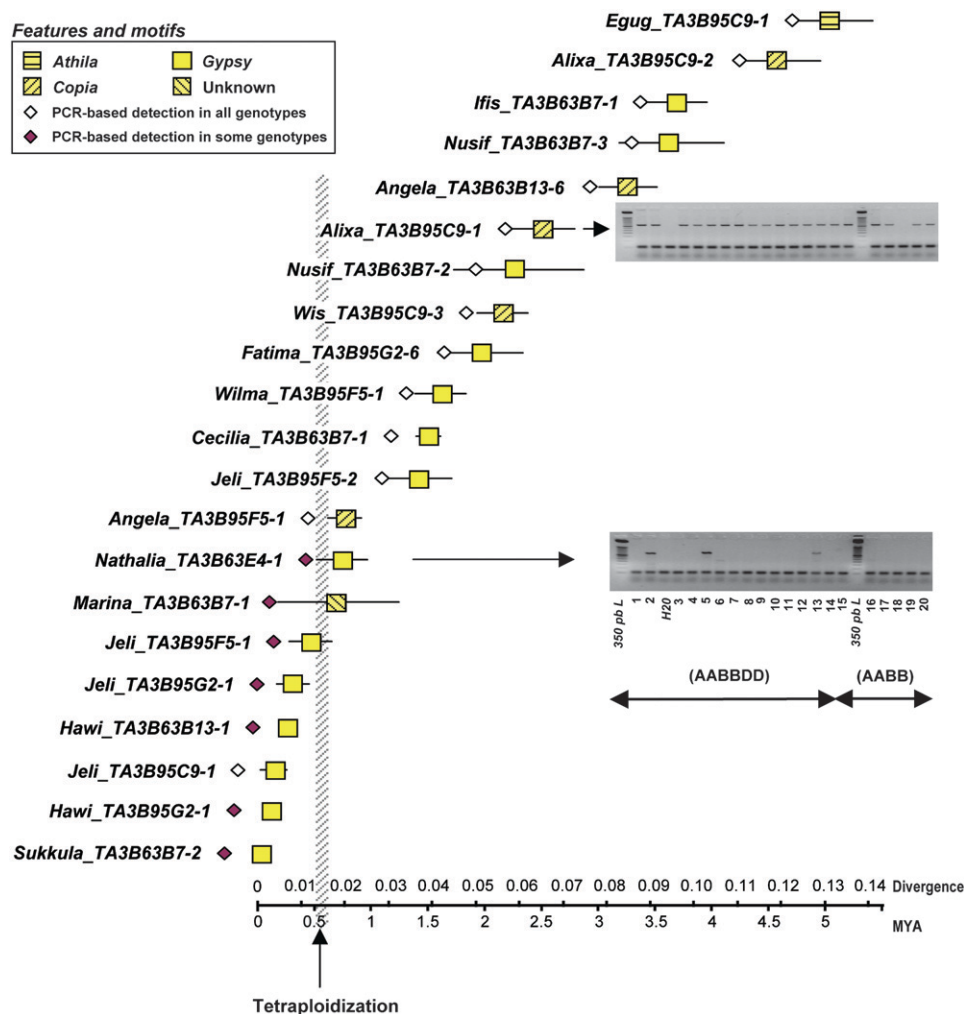


FIGURE 4.—PCR-based tracing of series of retrotransposons, inserted at different dates in wheat chromosome 3B across a collection of wheat tetraploid (*T. turgidum*) and hexaploid (*T. aestivum*) genotypes. Averages and intervals (means $\pm$ SE) of retrotransposon insertion dates are presented. The interval of the allotetraploidization event (0.5–0.6 MYA), calculated according to gene sequence divergence (HUANG *et al.* 2002; DVORAK *et al.* 2006; CHALUPSKA *et al.* 2008), is in gray. Retrotransposons for which insertion dates were estimated on the basis of divergence of their LTR prior to the tetraploidization event were generally detected in almost all genotypes, whereas those posterior to the tetraploidization event were detected in only some genotypes. Gels show PCR-based detection of insertion of RLC_*Alixa_TA3B95C9-1* (into DTC_*Caspar_TA3B95C9-1*) in all tested genotypes, except one and insertion of RLG_*Nathalia_TA3B63E4-1* (into DTC_*Vincent_TA3B63E4-2*) in some genotypes. Primer sequences, details of insertion dates (averages and intervals), and PCR-based detection in different wheat genotypes are given in supplemental Table 1 and supplemental Table 7. AABBDD (hexaploid wheat accessions): -1—*T. aestivum* cv. Renan;

-2—*T. aestivum* cv. Chinese Spring; -3—*T. aestivum* spelta, Erge 27216; -4—*T. aestivum* spelta, Erge 2776; -5—*T. aestivum* spelta, Erge 2771; -6—*T. aestivum* spelta Rouquin, Erge 6329; -7—*T. aestivum* macha 1793, Erge 27240; -8—*T. aestivum* compactum rufulum 71V, Erge 26786; -9—*T. aestivum* compactum crebicum 72V, Erge 26787; -10—*T. aestivum* compactum clavatum 73V, Erge 26788; -11—*T. aestivum* compactum icterinum 74V, Erge 26789; -12—*T. aestivum* compactum erinaceum 75V, Erge 26790; -13—*T. aestivum* sphaerococcum tumidum perciv globosum, Erge 27016; -14—*T. aestivum* cv. Soisson. AABB (*T. turgidum*, tetraploid wheat accessions): -15—*T. turgidum* durum cv. Langdon; -16—*T. turgidum* durum; -17—*T. turgidum* dicoccum; -18—*T. turgidum* dicoccoides; -19—*T. turgidum* polinicum; -20—*T. turgidum* turgidum.

the PCR primers in supplemental Table 1. With the exception of *Jeli_TA3B95C9-1*, all the other 7 most recently inserted retrotransposons, which have calculated insertion date intervals (means $\pm$ SE) equal to or less than the 0.5–0.6 MYA interval (divergence 0.013–0.016), were detected in some but not all genotypes carrying the B genome, suggesting their occurrence after the tetraploidization event (Figure 4 and supplemental Table 7). In contrast, all 13 retrotransposon insertions, which have calculated insertion intervals (means $\pm$ SE) >0.7 MYA, were detected in all tested genotypes carrying the B genome, suggesting their occurrence prior to the allotetraploidization event (Figure 4 and supplemental Table 7). Given the uncertainty in calculating intervals of insertion dates, the PCR-based tracing method confirms the calibration of LTR divergence on that of gene di-

vergence. More importantly, it also confirms that retrotranspositions (insertions) were not enhanced or repressed by the allotetraploidization event.

Relative proliferation periods of the CACTA class II transposable elements: The CACTA class II DNA TEs represent an important proportion of the B- and A-genome sequences (13.4 and 9.4%, respectively). As for the main LTR-retrotransposon superfamilies, ratios of complete to truncated copies are very different for B (0.37) and A (1.52) genomes (Table 1). In contrast to LTR retrotransposons, the CACTA TEs do not have long repeats or other features, which would allow determination of their insertion dates on the basis of sequence divergence. Therefore, their proliferation periods and rates were evaluated indirectly, relative to their level of insertions into or by other CACTA TEs and, more

TABLE 2

Associations of *CACTA* transposable elements with the four most represented TE superfamilies and other DNA sequence classes in 13 genomic regions of the wheat B genome and 19 publicly available genomic sequences of the wheat A genome

DNA sequence classes	13 genomic regions of the wheat B genome (1.98 Mb) ^a		19 publicly available genomic regions of the wheat A genome (3.63 Mb) ^b	
	<i>CACTA</i> TEs inserted into other DNA sequences ^c	Other DNA sequences inserted into <i>CACTA</i> TEs ^c	<i>CACTA</i> TEs inserted into other DNA sequences	Other DNA sequences inserted into <i>CACTA</i> TEs
<i>Athila</i> TEs	12: 4/8	6: 0/6	7: 7/0	0: 0/0
<i>Copia</i> TEs	1: 0/1	5: 0/5	10: 7/3	8: 6/2
<i>Gypsy</i> TEs	1: 1/0	6: 0/6	5: 4/1	6: 5/1
<i>CACTA</i> TEs	7: 4/3	6: 1/5	4: 4/0	3: 3/0
Other TEs	3: 3/0	0: 0/0	2: 1/1	0: 0/0
Unclear TE associations ^d	2: 0/2	2: 0/2	4: 0/4	4: 0/0
Unassigned DNA	44: 7/37	—	21: 9/12	—
Total	70: 19/51	25: 1/24	53: 32/21	21: 14/7

^a This corresponds to 1.43 Mb from the 10 genomic regions sequenced in this study and 0.55 Mb from three other publicly available genomic regions from SABOT *et al.* (2005), GU *et al.* (2006), and DVORAK *et al.* (2006). See MATERIALS AND METHODS for BAC clone sequence references.

^b Nineteen genomic regions available for the A genome (SANMIGUEL *et al.* 2002; YAN *et al.* 2002, 2003; WICKER *et al.* 2003b; CHANTRET *et al.* 2005; ISIDORE *et al.* 2005; DVORAK *et al.* 2006; GU *et al.* 2006; MILLER *et al.* 2006). See MATERIALS AND METHODS for BAC clone sequence references.

^c Results are as follows: total *CACTA* TE copies: complete *CACTA* TE copies/truncated *CACTA* TE copies.

^d From cases where we cannot be certain that a *CACTA* TE is inserted into or by another TE element.

importantly, by elements of the three main LTR-retrotransposon superfamilies for which proliferation periods and rates were evaluated on the basis of the dates of insertions (described above). This was calculated for all *CACTA* TE copies as well as for complete and truncated copies separately (Table 2).

In the wheat B genome, the majority of *CACTA* TE insertions (mainly those detected as truncated copies) occurred in DNA annotated as unassigned (Table 2). For the rest, significantly higher insertions of *CACTA* TEs into *Athila* and other *CACTA* TEs than into *Copia* and *Gypsy* retrotransposons were observed. The two latter retrotransposon superfamilies were significantly more inserted into, rather than by, *CACTA* TEs (Table 2). These observations indicate that proliferation of the *CACTA* TEs in the B genome of wheat started before, and continued during and after *Athila* retrotransposon proliferation, whereas very few insertions occurred during the last waves of high proliferation of *Copia* and *Gypsy*.

Similarly, a high level of insertions into unassigned DNA was observed for the *CACTA* TEs in the A genome. However, for the remaining insertions, no clear period of proliferation could be determined as these show similar levels of insertions into or by all other TE superfamilies (Table 2). These observations, combined with the observed higher level of complete copies (Table 1), suggest that the *CACTA* TE proliferation continued in the wheat A genome during the last waves of proliferation of *Copia* and *Gypsy*, unlike those in the B genome.

DISCUSSION

To constitute representative genomic sequences of the wheat B genome, in this study we have sequenced 10 BAC clones of the chromosome 3B, representing the most important number of genomic regions sequenced for a single wheat chromosome and a cumulative sequence length of 1.429 Mb (0.15% of the chromosome length). As expected, TE proliferation was pronounced (representing 79.1%). Five of these were revealed as gene-containing BAC clones at a density of one or two genes per clone; two other BAC clones contain gene relics or pseudogenes, whereas the three remaining BAC clones were missing genes. This confirms the previous conclusion about the more random distribution of genes on the wheat genome (Devos *et al.* 2005). Interestingly and in comparison with rice, a high level of “truncated genes” was revealed [six gene relic or pseudogenes, several of which because of TE insertions (three confirmed cases)]. If the confirmed gene number (excluding hypothetical genes) identified in the 1.43-Mb sequences (eight) is extrapolated to the whole wheat chromosome 3B of 1 Gb estimated size, then 5594 genes might be present. A slightly higher number (6000) was calculated from BAC-end sequence analysis (PAUX *et al.* 2006).

Representation of transposable elements: In this study, TE dynamics, proliferation, and evolutionary pathways were analyzed and compared in 1.98 Mb of sequence from 13 BAC clones of the wheat B genome and 3.63 Mb of sequence from 19 BAC clones of the

wheat A genome. These genomic sequences represent very small fractions (<0.03%) of their respective genomes. Nevertheless, it has been argued that, for studying abundant repeats, sequencing and annotation of a small proportion of the genome can be representative (BRENNER *et al.* 1993; VITTE and BENNETZEN 2006; LIU *et al.* 2007). We have been able to confirm the adequate representation where less variation in the proportion of the main TE superfamilies was observed when analyzing a large number of BAC clones (Figure 2). Interestingly, TE proportions observed in the 13 genomic regions of the B genome of wheat are similar to those obtained from 11 Mb of BAC-end sequences of wheat chromosome 3B (PAUX *et al.* 2006). Similarly, TE proportions were not significantly different for the wheat A genome when they were compared with the different ploidy levels (see RESULTS).

Although they are representative of abundant wheat TEs available in the TREP database (WICKER *et al.* 2002; <http://wheat.pw.usda.gov/ITMI/Repeats>), the class I and class II TEs observed in the genomic sequences of the wheat B and A genomes may not cover all wheat TEs. It is expected that more wheat TEs will be identified, as more wheat genomic sequences will become available. This is particularly supported by the identification in this study of >21 different novel TE families, most of which (17) are retrotransposons. We also believe that low-copy TEs and those that tend to “compartmentalize” in specific regions, such as pericentromeric heterochromatin regions (which is not the case in our regions), would be missed, over-, or underrepresented in this study (MA and BENNETZEN 2006; LIU *et al.* 2007). This could be the case for the *CACTA* TEs, which show the highest variation in sequence proportion between regions because they tend to be clustered in the Triticeae genomes (our unpublished results and WICKER *et al.* 2003a, 2005).

Transposable elements proliferated differentially in the B and A genomes of wheat: Abundance of TEs varies widely across different organisms. Human (*Homo sapiens*) DNA is composed of 45% (LANDER *et al.* 2001) repetitive sequences, *Drosophila melanogaster* of 3.9% (KAMINKER *et al.* 2002), and maize of 67% (HABERER *et al.* 2005; LIU *et al.* 2007) whereas TE content in the wheat genomic sequences analyzed in this study or in other studies (LI *et al.* 2004; GU *et al.* 2006; PAUX *et al.* 2006) is ~80%. Proportions of different classes of TEs also vary among organisms. Class II TEs are almost >10 times less abundant than class I TEs and constitute a small fraction (<2%) of the human, rice (PIEGU *et al.* 2006), maize (KRONMILLER and WISE 2008), *Arabidopsis*, and cotton (HAWKINS *et al.* 2006) genomes. In comparison, class II TE abundance is relatively high in the wheat B and A genomes (14.1 and 9.9%, respectively), the majority of which (95%) are *CACTA* TEs, which are particularly abundant in the Triticeae genomes (WICKER *et al.* 2003a, 2005). Class I retrotrans-

poson abundance is relatively high in several plant genomes, 58.7 and 56.6% estimated in this study for the wheat B and A genomes, respectively; 40–50% in cotton species (HAWKINS *et al.* 2006); 35–60% in rice species (PIEGU *et al.* 2006); and 64% in maize (LIU *et al.* 2007; KRONMILLER and WISE 2008).

In this study, combination of TE sequence analysis and classification, comparison of proportions of complete to incomplete copies, TE insertion date estimations, and PCR-based tracing of insertions allow us to compare TE proliferation periods and rates in the wheat B and A genomes (Figure 5). It is evident that TEs appear to proliferate differentially in waves of high activity followed by periods of low activity (Figure 5). Both genomes show similar rates and relatively old proliferation periods for the *Athila* retrotransposons (Figure 5). However, the *Copia* retrotransposons have proliferated relatively more recently in the A genome whereas a more recent *Gypsy* proliferation is observed in the B genome. Due to their biology and replication mechanism, it was not possible to directly estimate the *CACTA* class II TE insertion dates. We have estimated their proliferation periods and rates relative to that of the three main LTR retrotransposon superfamilies. In the wheat B genome, the *CACTA* TE high proliferation period started before and overlaps with that of the *Athila* retrotransposons. In the wheat A genome, in addition to the relatively old proliferation similar to that in the B genome, *CACTA* TEs continued to proliferate during the same period as *Gypsy* and *Copia* retrotransposons. Determining the ancient proliferation periods of *CACTA* TEs partially explains why *CACTA* TEs often tend to be clustered together (see RESULTS and WICKER *et al.* 2003a, 2005), although they were detected in almost all analyzed BAC clones. Differential proliferation of TEs provides a valid explanation for the size variation of closely related wheat genomes (BENNETT and SMITH 1976, 1991; <http://data.kew.org/cvalues/homepage.html>).

Four families (*Angela*, *Wis*, *Sabrina*, *Fatima*) were abundant, representing the majority of LTR retrotransposons in the B and A genomes of wheat, some of which proliferated differentially (see RESULTS). Proliferation of specific types of TEs in specific genomes (or species), leading to rapid genome size variation and sequence divergence, has also been observed in other plant species. Analysis of maize (*Zea mays*) genomic sequences suggests that the high percentage of LTR retrotransposons is due to proliferation of only a few families of TEs (MEYERS *et al.* 2001; LIU *et al.* 2007; KRONMILLER and WISE 2008). Similarly, comparison of TE proportions between various cotton species (*Gossypium* species) revealed differential lineage-specific expansion of various LTR–retrotransposon superfamilies and families, leading to threefold genome size differences (HAWKINS *et al.* 2006). Species-specific differential retrotransposon expansions are also the

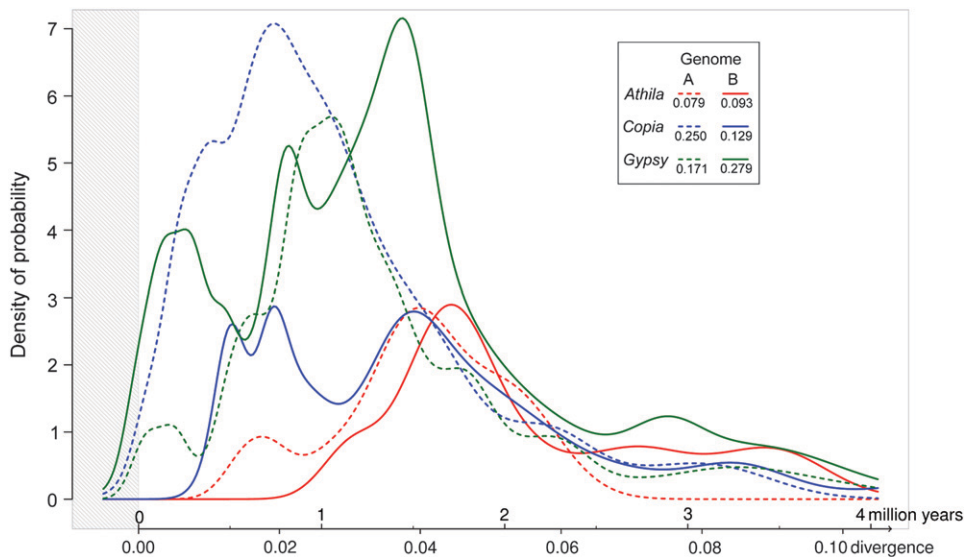


FIGURE 5.—Proliferation periods and rates of the main retrotransposon superfamilies in the wheat B and A genomes. Expressed as probability density functions, where the area under each curve was calculated on the basis of the estimated insertion dates of retrotransposons (in Figure 3) and their corresponding standard errors, using Gaussian kernel density estimation (SILVERMAN 1986). The curves have been scaled with respect to the number of observations, so that the sum of their areas (given for each retrotransposon superfamily in the key) equals the probability of 1 and comparisons between genomes and retrotransposon superfamilies can be performed. When calculated standard errors were very low, a minimum value of 80,000 years (corresponding to 0.002 divergence) was used. The shaded field is due to uncertainty in very recent insertion date estimations.

cause of the size doubling of the *Oryza australiensis* genome as compared to cultivated rice (*O. sativa*) (PIEGU *et al.* 2006).

This is the first time that dynamics as well as proliferation periods and rates of TEs have been compared between two closely related wheat genomes. This was possible only because in this study we sequenced 10 different genomic regions that constituted a genomic sequence data set representative of the wheat B genome. For the wheat A genome, more representative genomic sequence data were rendered publicly available. There have been initial attempts to evaluate TE proliferation in the wheat genomes. LI *et al.* (2004) analyzed the D genome of the diploid *Ae. tauschii* and showed that the copy number of most TEs have increased gradually following polyploidization. However, they used dot blots, which are not very accurate. SABOT *et al.* (2005) have updated TE annotation in wheat genomic sequences and reported their composition and distribution in relation to genes. They suggested that *Copia* TEs have been most active in the wheat A, B, and D genomes, combined together (SABOT *et al.* 2005). Accurate comparison of dynamics as well as proliferation periods and rates between individual genomes of wheat could not be conducted in the study of SABOT *et al.* (2005) as, in the genomic sequences available at that time, the A genome was overrepresented whereas the B genome was underrepresented. By using more representative genomic sequences in this study, we showed the more recent activation of the *Copia* and *CACTA* TEs in the wheat A genome but not in the B genome in which a more recent *Gypsy* proliferation is observed. Overrepresentation of the A-

genome sequences in the study of SABOT *et al.* (2005) may explain the reason why they found that *Copia* TEs have been most active in the wheat A, B, and D genomes combined together. Thus our analysis, using representative sequence data sets, for the first time shows differential proliferation of TEs between the wheat A and B genomes and illustrates the inadequacy of combining sequence data sets from different genomes as was previously done.

Neither enhancement nor repression of transposable element proliferation following allotetraploidization: As estimated from their insertion dates and confirmed by PCR-based tracing analysis, the majority of the differential proliferation of TEs in B and A genomes of wheat (87% and 83, respectively) occurred prior to the allotetraploidization event that brought them together in *T. turgidum* and *T. aestivum* <0.5 MYA (HUANG *et al.* 2002; DVORAK *et al.* 2006; CHALUPSKA *et al.* 2008). More importantly, the allotetraploidization event appears to have neither enhanced nor repressed retrotranspositions. We suggest that, in addition to the *Ph1* gene preventing homeologous recombination (GRIFFITHS *et al.* 2006), differential proliferation of TEs has also contributed to the rapid divergence of the B and A genomes of the wheat diploid progenitors and the relative stability of the natural wheat allopolyploids that occurs thereafter.

Different levels of stability, estimated as elimination of DNA sequences, were observed in newly synthesized wheat allopolyploids, depending on wheat genome combinations (FELDMAN and LEVY 2005 and our unpublished results). The natural wheat allopolyploids combining the B and A genomes are relatively stable

and cannot be exactly resynthesized because the diploid progenitor of the B genome is unidentified (FELDMAN *et al.* 1995; BLAKE *et al.* 1999; HUANG *et al.* 2002; DVORAK *et al.* 2006). Nevertheless, by studying a synthetic wheat allotetraploid combining the A and S genomes (the closest identified diploid relatives to the progenitors of the A and of the B genomes of natural wheat polyploids), KASHKUSH *et al.* (2003) reported on transcriptional activation of the *Wis* LTR retrotransposon but not its transposition following allotetraploidization. This is in agreement with the lack of enhancement of transpositions observed in this study in wheat natural allopolyploids combining the A and B genomes. Comparatively, less TE proliferation, estimated as the increased rate of deletions and the decreased rate of insertions, was recently observed in the cotton polyploid species *Gossypium hirsutum* as compared to its diploid progenitors *Gossypium arboreum* and *Gossypium raimondii* (GROVER *et al.* 2008).

Apparent transposable element proliferation as a balance between two evolutionary forces: TEs “transposition” and also their removal: As in this study, the vast majority of complete retrotransposons studied so far were also estimated to be <3 million years old (SANMIGUEL *et al.* 1998, 2002; WICKER *et al.* 2003b, 2005; GAO *et al.* 2004; MA *et al.* 2004; DU *et al.* 2006; PIEGU *et al.* 2006; WICKER and KELLER 2007). These findings imply that there are mechanisms of active deletion of LTR retrotransposons from the genome, such as unequal homologous recombination and illegitimate recombination (VICIENT *et al.* 1999; DEVOS *et al.* 2002; MA *et al.* 2004; PEREIRA 2004). Proliferation periods and rates estimated for TEs at a given evolutionary period are the result of both antagonist evolutionary forces: TE insertion activity (transpositions) (BENNETZEN and KELLOGG 1997) and the removal of TEs (PETROV *et al.* 2000; PETROV 2002a). Thus, it is not clear whether the insertions and/or truncation (removal) rates of TEs are constant or vary during genome evolution. The “burst of insertions” described for TEs could correspond to periods of (i) high insertion activity, (ii) low rates of TE removal, and/or (iii) combinations of both evolutionary forces.

The fact that *Copia* retrotransposons have been active until recently in the *Arabidopsis thaliana* genome allowed PEREIRA (2004) to calculate the rate of their elimination (or half-life) as 472,000 years, outside of centromeric regions. Using this method and assuming that repetitive sequences are removed from the genome at a constant rate, a higher half-life (79,000 years) was calculated for *Copia* removal in rice (WICKER and KELLER 2007). As the insertion-date distribution of *Copia* retrotransposons in Triticeae (wheat and barley) is not exponential, WICKER and KELLER (2007) suggested that their half-life is much longer than in rice, thus representing a major difference between small and large genomes of plants. Similar distributions are observed in our study for all three retrotransposon superfamilies in both B and A

genomes of wheat. Our analysis suggests that lower proliferation of the LTR retrotransposons during the most recent period could account for these apparent nonexponential distributions of insertion dates (including *Copia* retrotransposons) (Figure 5).

Our study clearly shows that, during their evolution, specific types of TEs have undergone differential proliferation in specific wheat genomes (or species) but not in others, leading to rapid sequence divergence. Little is known about the mechanistic causes that lead to differential proliferation of a single or related group of TEs across the genome of a specific species. These rapid TE expansions could correspond to periods of relaxed selection pressure such as genome duplication, interspecific hybridizations (although this was not revealed in our study), or stress conditions. It is also possible that TE proliferation could be caused by advantageous mutations in the TE sequence. A third alternative is differential deregulation of epigenetic silencing that allows specific TE families to proliferate in specific genomes.

We sincerely thank J. Dolezel and M. Kubalaková (Institute of Experimental Botany, Olomouc, Czech Republic) for providing FISH mapping information for BAC clones B95G2, B95C9, B63B7, and B54F7; Joseph Jahier [Institut National de la Recherche Agronomique (INRA), Rennes, France] and Moshe Feldman (Weizmann Institute of Science) for valuable discussions and for providing wheat genotypes; Catherine Feuillet (INRA, Clermont-Ferrand, France) for providing the wheat deletion lines; Thomas Wicker (Zurich University) for valuable advice on novel transposable element classifications and *CACTA* TE evolution; Piotr Gornicki (University of Chicago) and anonymous reviewers for valuable discussion and constructive criticisms; and Heather McKhann (Centre National de Génotypage, Etude du Polymorphisme Génomique Végétal-INRA, Evry, France) for valuable discussion and revision of the manuscript. This project was supported by the National Center for Sequencing (Centre National de Séquençage-Génoscope)/APCNS2003-Project: Triticum species comparative genome sequencing in wheat (<http://www.genoscope.cns.fr/externe/English/>). PCR-based tracing of retrotransposons insertions was funded by the Agence Nationale pour la Recherche Biodiversité Project (ANR-05-BDIV-015) and the ANR-05-Blanc project-ITEGE.

LITERATURE CITED

- ADAMS, K. L., and J. F. WENDEL, 2005 Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* **8**: 135–141.
- ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFER, J. ZHANG, Z. ZHANG *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- BENNETT, M. D., and I. J. LEITCH, 1997 Nuclear DNA amounts in angiosperms: 583 new estimates. *Ann. Bot.* **80**: 169–196.
- BENNETT, M. D., and I. J. LEITCH, 2005 Plant genome size research: a field in focus. *Ann. Bot.* **95**: 1–6.
- BENNETT, M. D., and J. B. SMITH, 1976 Nuclear DNA amounts in angiosperms. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **274**: 227–274.
- BENNETT, M. D., and J. B. SMITH, 1991 Nuclear DNA amounts in angiosperms. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **334**: 309–345.
- BENNETZEN, J. L., 2000a Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. *Plant Cell* **12**: 1021–1029.
- BENNETZEN, J. L., 2000b Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* **42**: 251–269.

- BENNETZEN, J. L., 2002a Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* **115**: 29–36.
- BENNETZEN, J. L., 2002b The rice genome: opening the door to comparative plant biology. *Science* **296**: 60–63.
- BENNETZEN, J. L., and E. A. KELLOGG, 1997 Do plants have a one-way ticket to genomic obesity? *Plant Cell* **9**: 1509–1514.
- BENNETZEN, J. L., J. MA and K. M. DEVOS, 2005 Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* **95**: 127–132.
- BLAKE, N. K., B. R. LEHFELDT, M. LAVIN and L. E. TALBERT, 1999 Phylogenetic reconstruction based on low copy DNA sequence data in an allopolyploid: the B genome of wheat. *Genome* **42**: 351–360.
- BLANC, G., A. BARAKAT, R. GUYOT, R. COOKE and M. DELSENY, 2000 Extensive duplication and reshuffling in the Arabidopsis genome. *Plant Cell* **12**: 1093–1101.
- BRENNER, S., G. ELGAR, R. SANDFORD, A. MACRAE, B. VENKATESH *et al.*, 1993 Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature* **366**: 265–268.
- CHALUPSKA, D., H. Y. LEE, J. D. FARIS, A. EVRARD, B. CHALHOUB *et al.*, 2008 Acc homoeoloci and the evolution of wheat genomes. *Proc. Natl. Acad. Sci. USA* **105**: 9691–9696.
- CHANTRET, N., J. SALSE, F. SABOT, S. RAHMAN, A. BELLEC *et al.*, 2005 Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *Plant Cell* **17**: 1033–1045.
- CHOOI, W. Y., 1971 Variation in nuclear DNA content in the genus *Vicia*. *Genetics* **68**: 195–211.
- DEVOS, K. M., J. K. BROWN and J. L. BENNETZEN, 2002 Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. *Genome Res.* **12**: 1075–1079.
- DEVOS, K. M., J. MA, A. C. PONTAROLI, L. H. PRATT and J. L. BENNETZEN, 2005 Analysis and mapping of randomly chosen bacterial artificial chromosome clones from hexaploid bread wheat. *Proc. Natl. Acad. Sci. USA* **102**: 19243–19248.
- DOLEZEL, J., M. KUBALAKOVA, J. BARTOS and J. MACAS, 2004 Flow cytogenetics and plant genome mapping. *Chromosome Res.* **12**: 77–91.
- DU, C., Z. SWIGONOVA and J. MESSING, 2006 Retrotranspositions in orthologous regions of closely related grass species. *BMC Evol. Biol.* **6**: 62.
- DVORAK, J., P. DITERLIZZI, H.-B. ZHANG and P. RESTA, 1993 The evolution of polyploid wheats: identification of the A genome donor species. *Genome* **36**: 21–31.
- DVORAK, J., E. D. AKHUNOV, A. R. AKHUNOV, K. R. DEAL and M. C. LUO, 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**: 1386–1396.
- EFRON, B., 1979 Bootstrap methods: another look at the jackknife. *Ann. Stat.* **7**: 1–26.
- FELDMAN, M., and A. A. LEVY, 2005 Allopolyploidy: a shaping force in the evolution of wheat genomes. *Cytogenet. Genome Res.* **109**: 250–258.
- FELDMAN, M., F. G. H. LUPTON and T. E. MILLER, 1995 Wheats, pp.184–192 in *Evolution of Crops*, Ed. 2, edited by J. SMARTT and N. W. SIMMONDS. Longman Scientific, London.
- FÉRIGNAC, P., 1962 Test de Kolmogorov-Smirnov sur la validité d'une fonction de distribution. *Rev. Stat. Appl.* **10**: 13–32.
- FLAVELL, A. J., M. R. KNOX, S. R. PEARCE and T. H. ELLIS, 1998 Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *Plant J.* **16**: 643–650.
- GAO, L., E. M. MCCARTHY, E. W. GANKO and J. F. McDONALD, 2004 Evolutionary history of *Oryza sativa* LTR retrotransposons: a preliminary survey of the rice genome sequences. *BMC Genomics* **5**: 18.
- GRANER, A., H. SIEDLER, A. JAHOOOR, R. G. HERRMAN and G. WENZAL, 1990 Assessment of the degree and the type of restriction fragment length polymorphism in barley (*Hordeum vulgare*). *Theor. Appl. Genet.* **80**: 826–832.
- GRIFFITHS, S., R. SHARP, T. N. FOOTE, I. BERTIN, M. WANOUS *et al.*, 2006 Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. *Nature* **439**: 749–752.
- GROVER, C. E., Y. YU, R. A. WING, A. H. PATERSON and J. F. WENDEL, 2008 A phylogenetic analysis of indel dynamics in the cotton genus. *Mol. Biol. Evol.* **25**: 1415–1428.
- GU, Y. Q., J. SALSE, D. COLEMAN-DERR, A. DUPIN, C. CROSSMAN *et al.*, 2006 Types and rates of sequence evolution at the high-molecular-weight glutenin locus in hexaploid wheat and its ancestral genomes. *Genetics* **174**: 1493–1504.
- HABERER, G., S. YOUNG, A. K. BHARTI, H. GUNDLACH, C. RAYMOND *et al.*, 2005 Structure and architecture of the maize genome. *Plant Physiol.* **139**: 1612–1624.
- HAWKINS, J. S., H. KIM, J. D. NASON, R. A. WING and J. F. WENDEL, 2006 Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res.* **16**: 1252–1261.
- HUANG, S., A. SIRIKHACHORNKIT, X. SU, J. FARIS, B. GILL *et al.*, 2002 Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc. Natl. Acad. Sci. USA* **99**: 8133–8138.
- ISIDORE, E., B. SCHERRER, B. CHALHOUB, C. FEUILLET and B. KELLER, 2005 Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels. *Genome Res.* **15**: 526–536.
- JAILLON, O., J. M. AURY, B. NOEL, A. POLICRITI, C. CLEPET *et al.*, 2007 The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**: 463–467.
- JONES, R. N., and L. M. BROWN, 1976 Chromosome evolution and DNA variation in *Crepis*. *Heredity* **36**: 91–104.
- KAMINKER, J. S., C. M. BERGMAN, B. KRONMILLER, J. CARLSON, R. SVIRSKAS *et al.*, 2002 The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol.* **3**: RESEARCH0084.
- KASHKUSH, K., M. FELDMAN and A. A. LEVY, 2003 Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat. Genet.* **33**: 102–106.
- KIDWELL, M. G., 2002 Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**: 49–63.
- KIMURA, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**: 111–120.
- KRONMILLER, B. A., and R. P. WISE, 2008 TE nest: automated chronological annotation and visualization of nested plant transposable elements. *Plant Physiol.* **146**: 45–59.
- KUMAR, S., K. TAMURA and M. NEI, 2004 MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5**: 150–163.
- LANDER, E. S., L. M. LINTON, B. BIRREN, C. NUSBAUM, M. C. ZODY *et al.*, 2001 Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- LI, W., P. ZHANG, J. P. FELLERS, B. FRIEBE and B. S. GILL, 2004 Sequence composition, organization, and evolution of the core *Triticeae* genome. *Plant J.* **40**: 500–511.
- LIU, R., C. VITTE, J. MA, A. A. MAHAMA, T. DHLIWAYO *et al.*, 2007 A GeneTrek analysis of the maize genome. *Proc. Natl. Acad. Sci. USA* **104**: 11844–11849.
- MA, J., and J. L. BENNETZEN, 2004 Rapid recent growth and divergence of rice nuclear genomes. *Proc. Natl. Acad. Sci. USA* **101**: 12404–12410.
- MA, J., and J. L. BENNETZEN, 2006 Recombination, rearrangement, reshuffling, and divergence in a centromeric region of rice. *Proc. Natl. Acad. Sci. USA* **103**: 383–388.
- MA, J., K. M. DEVOS and J. L. BENNETZEN, 2004 Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res.* **14**: 860–869.
- MEYERS, B. C., S. V. TINGEY and M. MORGANTE, 2001 Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res.* **11**: 1660–1676.
- MILLER, A. K., G. GALIBA and J. DUBCOVSKY, 2006 A cluster of 11 CBF transcription factors is located at the frost tolerance locus *Fr-Am2* in *Triticum monococcum*. *Mol. Genet. Genomics* **275**: 193–203.
- PATERSON, A. H., J. E. BOWERS and B. A. CHAPMAN, 2004 Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci. USA* **101**: 9903–9908.
- PAUX, E., D. ROGER, E. BADAIEVA, G. GAY, M. BERNARD *et al.*, 2006 Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. *Plant J.* **48**: 463–474.

Bibliographie

Ouvrages imprimés

1. Prat D., Raynald-Roques A., Roguenant A. *Peut-on classer le vivant (Linné et la systématique aujourd'hui)*. Paris : BELIN, février 2007, 437p.
2. De Candolle A. *L'origine des plantes cultivées*. Paris : Diderot Multimedia. (Coll. Latitudes, 18). 488p.
3. Morgunov A.I. and Associate Scientist. *Wheat and Wheat Breeding in the Former USSR. CIMMYT Wheat Program Wheat Special report No 13*, (1992). 35p

Ouvrages électroniques

1. Darwin C. *De l'origine des espèces*. Schleicher Frères, 1896 traduit Ed.Barbier [référence Avril 2004], 708p. <http://www.ebooksgratuits.com>

Sites web consultés

The Emergence of Civilisation: From Hunting and Gathering to Agriculture ... - Charles Keith Maisels
- Google Livres

<http://books.google.fr/>

Wikipédia

<https://fr.wikipedia.org/>

Livre:Alphonse de Candolle - Origine des plantes cultivées, 1883

<http://fr.wikisource.org/>

Plantae, Botanique systématique des plantes du monde.

<http://www.plantae.ca/>

ClustalW2<MultipleSequenceAlignment<EMBL-EBI

http://www.ebi.ac.uk/Tools/services/web_clustalw2/toolform.ebi

BLAST: Basic Local Alignment Search Tool

<http://www.ncbi.nlm.nih.gov/blast/>

Statistiques mondiales écologiques

<http://www.planetoscope.com/cartes-statistiques.php>

Poacées - Graminées - Epillet

<http://botarela.fr/Poaceae/Description-detail/Epillet.html>

Ernest Robert Sears, October 15, 1910—February 15, 1991 | By Ralph Riley | Biographical Memoirs

<http://www.nap.edu/readingroom.php-book=biomems&page=esears.html>

Articles de périodiques imprimés

- Abbo, S., Lev-Yadun, S., and Ladizinsky, G. (2001). Tracing the wild genetic stocks of crop plants. *Genome* **44**, 309-310.
- Abbo, S., Lev-Yadun, S., and Gopher, A. (2010a). Agricultural Origins: Centers and Noncenters; A Near Eastern Reappraisal. *Critical Reviews in Plant Sciences* **29**, 317-328.
- Abbo, S., Lev-Yadun, S., and Gopher, A. (2010b). Origin of Near Eastern plant domestication: homage to Claude Levi-Strauss and "La Pensée Sauvage". *Genetic Resources and Crop Evolution* **58**, 175-179.
- Abbo, S., Lev-Yadun, S., and Gopher, A. (2012). Plant Domestication and Crop Evolution in the Near East: On Events and Processes. *Critical Reviews in Plant Sciences* **31**, 241-257.
- Abbo, S., Gopher, A., Peleg, Z., Saranga, Y., Fahima, T., Salamini, F., and Lev-yadun, S. (2006). The ripples of "The Big (agricultural) Bang": the spread of early wheat cultivation **863**, 861-863.
- Adams, K., and Wendel, J. (2005a). Allele-specific, bi-directional silencing of an alcohol dehydrogenase gene in different organs of interspecific diploid cotton hybrids. *Genetics* **171**, 2139-2142.
- Adams, K., and Wendel, J. (2005b). Novel patterns of gene expression in polyploid plants. . *Trends Genet* **21**, 539-543.
- Adams, K.L., and Wendel, J.F. (2005c). Polyploidy and genome evolution in plants. *Current opinion in plant biology* **8**, 135-141.
- Adams, K.L., Percifield, R., and Wendel, J.F. (2004). Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics* **168**, 2217-2226.
- Agrawal, A., Eastman, Q., and Schatz, D. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394**, 744-751.
- Aiello, B.L.C., Peggy, B., Sanday, R., Fuller, B.D.Q., and Whitman, W. (2010). Going Home : An the Origins of Agriculture **17**.
- Al-Kaff, N., Knight, E., Bertin, I., Foote, T., Hart, N., Griffiths, S., and Moore, G. (2008). Detailed Dissection of the Chromosomal Region Containing the Ph1 Locus in Wheat *Triticum aestivum*: With Deletion Mutants and Expression Profiling. *Annals of Botany* **101**, 863-872.
- Allaby, R. (2010). Integrating the processes in the evolutionary system of domestication. *J Exp Bot* **61**, 935-944.
- Allaby, R.G., Fuller, D.Q., and Brown, T.a. (2008). The genetic expectations of a protracted model for the origins of domesticated crops. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 13982-13986.
- Alley, R., Meese, D., Shuman, C., and Gow, A. (1993). Abrupt increase in snow accumulation at the end of the Younger Dryas event. *Nature* **362**, 527-529.
- Alo, F., Furman, B.J., Akhunov, E., Dvorak, J., and Gepts, P. (2011). Leveraging genomic resources of model species for the assessment of diversity and phylogeny in wild and domesticated lentil. *J Hered* **102**, 315-329.
- Amborella Genome Project. (2013). The Amborella genome and the evolution of flowering plants. *Science* **342**, 1241089.

- Andreuzza, S., and Siddiqi, I.** (2008). Spindle positioning, meiotic nonreduction, and polyploidy in plants. *PLoS Genet* **4**, e1000272.
- APG , A.P.G.** (1998). An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* **85**, 531-553.
- APG II, A.P.G.** (2003). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. *Botanical Journal of the Linnean Society* **141**, 399-436.
- APG III.** (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161**, 105-121.
- Arabidopsis, T., and Initiative, G.** (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815.
- Asouti, E., and Fuller, D.Q.** (2012). From foraging to farming in the southern Levant: the development of Epipalaeolithic and Pre-pottery Neolithic plant management strategies. *Vegetation History and Archaeobotany* **21**, 149-162.
- Aury, J.M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Segurens, B., Daubin, V., Anthouard, V., Aiach, N., Arnaiz, O., Billaut, A., Beisson, J., Blanc, I., Bouhouche, K., Camara, F., Duharcourt, S., Guigo, R., Gogendeau, D., Katinka, M., Keller, A.M., Kissmehl, R., Klotz, C., Koll, F., Le Mouel, A., Lepere, G., Malinsky, S., Nowacki, M., Nowak, J.K., Plattner, H., Poulain, J., Ruiz, F., Serrano, V., Zagulski, M., Dessen, P., Betermier, M., Weissenbach, J., Scarpelli, C., Schachter, V., Sperling, L., Meyer, E., Cohen, J., and Wincker, P.** (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**, 171-178.
- Axtell, M.J.** (2013). Classification and comparison of small RNAs from plants. *Annual review of plant biology* **64**, 137-159.
- Badaeva, E.D., Dedkova, O.S., Gay, G., Pukhalskyi, V.A., Zelenin, A.V., Bernard, S., and Bernard, M.** (2007). Chromosomal rearrangements in wheat: their types and distribution. *Genome* **50**, 907-926.
- Belyayev, A., Raskina, O., Korol, A., and Nevo, E.** (2000). Coevolution of A and B genomes in allotetraploid *Triticum dicoccoides*. *Genome* **43**, 1021-1026.
- Bennetzen, J.L.** (2002a). Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* **115**, 29-36.
- Bennetzen, J.L.** (2002b). The rice genome: opening the door to comparative plant biology. *Science* **296**, 60-63.
- Bennetzen, J.L., and Kellogg, E.A.** (1997). Do plants have a one-way ticket to genomic obesity? *Plant Cell* **9**, 1509-1514.
- Bennetzen, J.L., Ma, J., and Devos, K.M.** (2005). Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* **95**, 127-132.
- Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., and Hannon, G.J.** (2003). Dicer is essential for mouse development. *Nat. Genet.* **35**, 215-217.
- Birchler, J.A., and Veitia, R.A.** (2007). The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell* **19**, 395-402.
- Birchler, J.A., Auger, D.L., and Riddle, N.C.** (2003). In search of the molecular basis of heterosis. *Plant Cell* **15**, 2236-2239.
- Blakeslee, A.** (1922). Variation in *Datura* due to changes in chromosome number. *Am Nat.* **56**, 16-31.

- Blanc, G., and Wolfe, K.H.** (2004). Widespread Paleopolyploidy in Model Plant Species Inferred from Age Distributions of Duplicate Genes **16**, 1667-1678.
- Bowman, J., Smyth, D., and Meyerowitz, E.** (1989). Genes directing flower development in Arabidopsis. *Plant cell* **1**, 37-52.
- Bowman, J., Alvarez, J., Weigel, D., Meyerowitz, E., and Smyth, D.** (1993). Control of flower development in Arabidopsis thaliana by APETALA1 and interacting genes. *Development* **3**, 721-743.
- Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G.L.a., D'Amore, R., Allen, A.M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M.-C., Sehgal, S., Gill, B., Kianian, S., Anderson, O., Kersey, P., Dvorak, J., McCombie, W.R., Hall, A., Mayer, K.F.X., Edwards, K.J., Bevan, M.W., and Hall, N.** (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**, 705-710.
- Bretagnolle, F., and Thompson, J.D.** (1995). Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopoloid plants. *New Phytol* **129**, 1-22.
- Brown, T.a., Jones, M.K., Powell, W., and Allaby, R.G.** (2009). The complex origins of domesticated crops in the Fertile Crescent. *Trends in ecology & evolution* **24**, 103-109.
- Brownfield, L., and Kohler, C.** (2010). Unreduced gamete formation in plants: mechanisms and prospects. *J Exp Bot* **62**, 1659-1668.
- Buffetaut, E.** (2013). Enquête sur la disparition des dinosaures.
- Bunting, G.S.** (1960). A revision of Spathiphyllum(Araceae). *Mem. New York bot. Gard.* **10**, 1-54.
- Camacho, J.P.M., Sharbel, T.F., and Beukeboom, L.W.** (2000). B-chromosome evolution. . *Phil. Trans. R. Soc. Lond* **355**, 163-178.
- Cao, W., Scoles, G.J., and Hucl, P.** (1997). The genetics of rachis fragility and glume tenacity in semi-wild wheat. *Crop Science*, 119-124.
- Capy, P., Bazin, C., Higuët, D., and Langin, T.** (1998). Dynamics and evolution of transposable elements. (Austin: Library of Congress).
- Chague, V., Just, J., Mestiri, I., Balzergue, S., Tanguy, A.M., Huneau, C., Huteau, V., Belcram, H., Coriton, O., Jahier, J., and Chalhoub, B.** (2010). Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. *New Phytol* **187**, 1181-1194.
- Chalupska, D., Lee, H.Y., Faris, J.D., Evrard, a., Chalhoub, B., Haselkorn, R., and Gornicki, P.** (2008). Acc homoeoloci and the evolution of wheat genomes. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 9691-9696.
- Chantret, N., Salse, J.r.m., Sabot, F.o., Bellec, A., Laubin, B., Dubois, I., Dossat, C., Sourdille, P., Joudrier, P., Gautier, M.-F.o., Cattolico, L., Beckert, M., Aubourg, S.b., Weissenbach, J., Caboche, M., Leroy, P., Bernard, M., and Chalhoub, B.** (2008). Contrasted microcolinearity and gene evolution within a homoeologous region of wheat and barley species. *Journal of molecular evolution* **66**, 138-150.
- Chantret, N., Salse, J., Sabot, F., Rahman, S., Bellec, A., Laubin, B., Dubois, I., Dossat, C., Sourdille, P., Joudrier, P., Gautier, M.F., Cattolico, L., Beckert, M., Aubourg, S., Weissenbach, J., Caboche, M., Bernard, M., Leroy, P., and Chalhoub, B.** (2005). Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (Triticum and Aegilops). *Plant Cell* **17**, 1033-1045.
- Charles, M., Tang, H., Belcram, H., Paterson, A., Gornicki, P., and Chalhoub, B.** (2009). Sixty million years in evolution of soft grain trait in grasses: emergence of the softness locus in the common ancestor

of Pooideae and Ehrhartoideae, after their divergence from Panicoideae. *Molecular biology and evolution* **26**, 1651-1661.

- Charles, M., Belcram, H., Just, J.r.m., Huneau, C.c., Viollet, A.s., Couloux, A., Segurens, B.a., Carter, M., Huteau, V., Coriton, O., Appels, R., Samain, S., and Chalhou, B.** (2008). Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. *Genetics* **180**, 1071-1086.
- Chase, M.W., Reveal, J.L., Hortorium, L.H.B., Biology, P., and Building, M.** (2009a). A phylogenetic classification of the land plants to accompany APG III, 122-127.
- Chase, M.W., Fay, M.F., Reveal, J.L., Soltis, D.E., Soltis, P.S., Peter, F., Anderberg, A.A., Moore, M.J., Olmstead, R.G., Rudall, P.J., and Kenneth, J.** (2009b). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants : APG III, 105-121.
- Chaudhary, B., Flagel, L., Stupar, R.M., Udall, J.A., Verma, N., Springer, N.M., and Wendel, J.F.** (2009). Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (gossypium). *Genetics* **182**, 503-517.
- Chauvet, M.** (1998). L'origine des plantes cultivées d'Alphonse de Candolle. Comment lire ce classique plus d'un siècle après sa parution. In http://www.tela-botanica.org/page:article_opcac.
- Chelaifa, H., Chague, V., Chalabi, S., Mestiri, I., Arnaud, D., Deffains, D., Lu, Y., Belcram, H., Huteau, V., Chiquet, J., Coriton, O., Just, J., Jahier, J., and Chalhou, B.** Prevalence of gene expression additivity in genetically stable wheat allohexaploids. *New Phytol* **197**, 730-736.
- Chen, X.** (2004). AmicroRNAas a translational repressor of APETALA2 in Arabidopsis. *Science* **303**, 2022–2025.
- Chen, Z.J.** (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol* **58**, 377-406.
- Chen, Z.J., and Ni, Z.** (2006). Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *Bioessays* **28**, 240-252.
- Chuck, G., Meeley, R., and Hake, S.** (1998). The control of maize spikelet meristem fate by the APETALA2 -- like gene *Genes indeterminate spikelet1*. *Gene. Devel* **12**, 1145-1154.
- Comai, L.** (2005). The advantages and disadvantages of being polyploid. *Nature reviews. Genetics* **6**, 836-846.
- Comai, L., Tyagi, A.P., and Lysak, M.A.** (2003). FISH analysis of meiosis in Arabidopsis allopolyploids. *Chromosome Res* **11**, 217-226.
- Comai, L., Tyagi, A.P., Winter, K., Holmes-Davis, R., Reynolds, S.H., Stevens, Y., and Byers, B.** (2000). Phenotypic instability and rapid gene silencing in newly formed arabidopsis allotetraploids. *Plant Cell* **12**, 1551-1568.
- Cui, L., Wall, P.K., Leebens-Mack, J.H., Lindsay, B.G., Soltis, D.E., Doyle, J.J., Soltis, P.S., Carlson, J.E., Arumuganathan, K., Barakat, A., Albert, V.A., Ma, H., and dePamphilis, C.W.** (2006). Wide-spread genome duplications throughout the history of flowering plants. *Genome Res* **16**, 738-749.
- d'Erfurth, I., Jolivet, S., Froger, N., Catrice, O., Novatchkova, M., and Mercier, R.** (2009). Turning Meiosis into Mitosis. *PLoS Biol.* **7**, e1000124.
- d'Erfurth, I., Cromer, L., Jolivet, S., Girard, C., Horlow, C., Sun, Y., To, J.P., Berchowitz, L.E., Copenhaver, G.P., and Mercier, R.** (2010). The cyclin-A CYCA1;2/TAM is required for the meiosis I to meiosis II transition and cooperates with OSD1 for the prophase to first meiotic division transition. *PLoS Genet* **6**, e1000989.

- Damania, A.** (1998). Diversity of major cultivated plants domesticated in the Near East. . In *The origins of agriculture and crop domestication.*, A. Damania, J. Valkoun, G. Willcox, and C. Qualset, eds (ICARDA, Aleppo, Syria.: Proceedings of the Harlan Symposium.), pp. 51-64.
- Darlington, C.** (1937). Recent Advances in Cytology.
- Delaney, D.E., Nasuda, S., Endo, T.R., Gill, B.S., and Hulbert, S.H.** (1995a). Cytogenetically based physical maps of the group 2 chromosomes of wheat. . *Theor. Appl. Genet.* **91**, 568–573.
- Delaney, D.E., Nasuda, S., Endo, T.R., Gill, B.S., and Hulbert, S.H.** (1995b). Cytogenetically based physical maps of the group 3 chromosomes of wheat. . *Theor. Appl. Genet.* **91**, 780–782.
- Devos, K.M., Brown, J.K., and Bennetzen, J.L.** (2002). Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res* **12**, 1075-1079.
- Doolittle, W.F., and Sapienza, C.** (1980.). Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603.
- Dorofeev, V.F., Udachin, R.A., Semenova, L.V., and al., e.** (1987). *Wheats of the World: Leningrad.* Kolos, 560 p.
- 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**, 443-461.
- Drews, G.N., Bowman, J.L., and Meyerowitz, E.M.** (1991). Negative regulation of the *Arabidopsis* homeotic gene *AGAMOUS* by the *APETALA2* product. *Cell* **65**, 991–1002.
- Du, C., Swigonov, Z., and Messing, J.** (2006). Retrotranspositions in orthologous regions of closely related grass species. *BMC evolutionary biology* **6**, 62-62.
- 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**, 323-332.
- Dvorak, J., Luo, M., and Akhunov, E.** (2011). N.I. Vavilov's theory of centers of diversity in light of current understanding of wheat diversity, domestication and evolution. *Czech. J Genet Plant Breed.*
- Dvorak, J., M.-C., L., Z.-L., Y., and H.-B., Z.** (1998). The structure of the *Aegilops tauschii* gene pool and the evolution of hexaploid wheat. *Theoretical and Applied Genetics* **97**, 657-670.
- Elbashir, S., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T.** (2001). Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster*. *J Exp Bot* **20**, 6877–6888.
- Endo, T.R., and Gill, B.S.** (1988). The Deletion Stocks of Common Wheat.
- Endo, T.R., and Gill, B.S.** (1996). The deletion stocks of common wheat. *J of Heredity* **87**, 295-307.
- Erilova, A., Brownfield, L., Exner, V., Rosa, M., Twell, D., Mittelsten Scheid, O., Hennig, L., and Köhler, C.** (2009). Imprinting of the polycomb group gene *MEDEA* serves as a ploidy sensor in *Arabidopsis*. *PLoS Genetics* **5**, e1000663.
- Fagard, M., Boutet, S., Morel, J.B., Bellini, C., and Vaucheret, H.** (2000). AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proc. Natl. Acad. Sci. USA* **97**, 11650–11654.
- Fans, J.D., Simons, K.J., Zhang, Z., and Gill, B.S.** (2003). The wheat super domestication gene. *Frontiers of Wheat Bioscience* 129-148.

- Feldman, M., and Levy, A.A.** (2005). Allopolyploidy--a shaping force in the evolution of wheat genomes. *Cytogenet Genome Res* **109**, 250-258.
- Feldman, M., and Kislev, M.** (2007). Domestication of emmer wheat and evolution of free-threshing tetraploid wheat. *Israel J. Plant Sci.* **55**, 207-221.
- Feldman, M., Liu, B., Segal, G., Abbo, S., Levy, A.A., and Vega, J.M.** (1997). Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. *Genetics* **147**, 1381-1387.
- Feschotte, C.d., and Pritham, E.J.** (2009). A cornucopia of Helitrons shapes the maize genome. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 19747-19748.
- Filatenko, A., and Kurkiev, U.** (1975). Sinskaya wheat.(in Russian). *Trudy po prikl. botan., genet. i selektsii* **54**, 239-241.
- Finnegan, D.** (1990). Transposable elements and DNA transposition in eukaryotes. *Curr Opin Cell Biol* **2**, 471-477.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and al., e.** (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806-811.
- Fukuda, Y.** (1927). Cytological studies on the development of pollen grains in different races of solanum. *Bot Mag Tokyo* **41**, 459-474.
- Fuller, D.** (2011). The Archaeobotanist, <http://archaeobotanist.blogspot.fr/2011/12/de-centering-fertile-crescent.html>, ed.
- Fuller, D.Q., Asouti, E., and Purugganan, M.D.** (2010). Cultivation as slow evolutionary entanglement: comparative data on rate and sequence of domestication. *Vegetation History and Archaeobotany* **21**, 131-145.
- Gaeta, R.T., and Chris Pires, J.** (2010). Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytol* **186**, 18-28.
- Gao, L., McCarthy, E.M., Ganko, E.W., and McDonald, J.F.** (2004). Evolutionary history of *Oryza sativa* LTR retrotransposons: a preliminary survey of the rice genome sequences. *BMC Genomics* **5**, 18.
- Gao, X., and Voytas, D.** (2005). A eukaryotic gene family related to retroelement integrases. *Trends Genet* **21**, 133-137.
- Gill, B., Friebe, B., and Endo, T.** (1991). Standard karyotype and nomenclature system for description of chromosome bands and structural aberrations. *Genome* **34**, 830-839.
- Gill, B., Friebe, B., Raupp, W., Wilson, D., Cox, T., Sears, R., Brown-Guedira, G., and Fritz, A.** (2006). Wheat genetics resource center: the first 25 years. *Adv Agron* **89**, 73-136.
- Gill, B.S., and Kimber, G.** (1974b). Giemsa C-banding and the evolution of wheat. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 4086-4090.
- Gopher, A., Abbo, S., and Lev-Yadun, S.** (2001). The "when", the "where" and the "why" of the Neolithic revolution in the Levant. *Doc Praehist.*
- Grant, V.** (1981). Plant speciation.
- Griffiths, S., Sharp, R., Foote, T.N., Bertin, I., Wanous, M., Reader, S., Colas, I., and Moore, G.** (2006). Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. *Nature* **439**, 749-752.

- Gu, Y.Q., Salse, J.r.m., Coleman-Derr, D., Dupin, A., Crossman, C., Lazo, G.R., Huo, N., Belcram, H., Ravel, C., Charmet, G., Charles, M., Anderson, O.D., and Chalhou, B.** (2006). Types and rates of sequence evolution at the high-molecular-weight glutenin locus in hexaploid wheat and its ancestral genomes. *Genetics* **174**, 1493-1504.
- Ha, M., Kim, E.D., and Chen, Z.J.** (2009a). Duplicate genes increase expression diversity in closely related species and allopolyploids. . *Proc Natl Acad Sci U S A* **106**, 2295-2300.
- Harlan, J.** (1973a). *Crops and man*. (Madison: American Society of Agronomy).
- Harlan, J., and De Wet, J.** (1975). On Ö.Winge and a prayer: the origins of polyploidy. *Bot. Rev.* **41**, 361-390.
- Harlan, J.R., and Zohary, D.** (1966). Distribution of wild wheats and barley. *Science* **153**, 1074-1080.
- Harris, D., and Hillman, G.** (1989). an evelutionary continuum of people-plant interaction. In *Foraging and farming: the evolution of plant exploitation* (london: unwin hyman), pp. 11-26.
- Haston, E., Richardson, J.E., Stevens, P.F., Chase, M.W., Harris, D.J., Garden, M.B., Box, P.O., and Louis, S.** (2009). The Linear Angiosperm Phylogeny Group (LAPG) III : a linear sequence of the families in APG III, 128-131.
- Hawkins, J.S., Kim, H., Nason, J.D., Wing, R.a., and Wendel, J.F.** (2006). Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome research* **16**, 1252-1261.
- He, P., Friebe, B.R., Gill, B.S., and Zhou, J.M.** (2003). Allopolyploidy alters gene expression in the highly stable hexaploid wheat. *Plant Mol Biol* **52**, 401-414.
- Hegarty, M.J., and Hiscock, S.J.** (2008). Genomic clues to the evolutionary success of polyploid plants. *Curr Biol* **18**, R435-444.
- Hegarty, M.J., Jones, J.M., Wilson, I.D., Barker, G.L., Coghill, J.A., Sanchez-Baracaldo, P., Liu, G., Buggs, R.J., Abbott, R.J., Edwards, K.J., and Hiscock, S.J.** (2005). Development of anonymous cDNA microarrays to study changes to the *Senecio* floral transcriptome using hybrid speciation. . *Mol Ecol* **14**, 2493-2510.
- Henry, I.M., Dilkes, B.P., and Comai, L.** (2006). Molecular karyotyping and aneuploidy detection in *Arabidopsis thaliana* using quantitative fluorescent polymerase chain reaction. *Plant J* **48**, 307-319.
- Heun, M., Scha, R., Klawan, D., Castagna, R., Accerbi, M., Borghi, B., and Salamini, F.** (1997). Site of Einkorn Wheat Domestication Identified by DNA Fingerprinting. *Science* **278**, 1312-1314.
- Hovav, R., Udall, J.a., Chaudhary, B., Rapp, R., Flagel, L., and Wendel, J.F.** (2008). Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 6191-6195.
- Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R., and Gornicki, P.** (2002). Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 8133-8138.
- Huskins, C.** (1946). Fatuoid, speltoid and related mutations of oats and wheat. *Botan Rev* **12**, 457- 514.
- Inda, L.A., Segarra-Moragues, J.G., Muller, J., Peterson, P.M., and Catalan, P.** (2008). Dated historical biogeography of the temperate Loliinae (Poaceae, Pooideae) grasses in the northern and southern hemispheres. *Mol Phylogenet Evol* **46**, 932-957.
- International Rice Genome Sequencing Project.** (2005). The map- based sequence of the rice genome. *Nature* **436**, 793–800.

- Irish, V., and Sussex, I.** (1990). Function of the APETALA-1 gene during Arabidopsis floral development. *Plant cell* **8**, 8.
- Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S.b., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Huguency, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyat, C.m., Billault, A., Segurens, B.a., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V.r., Vico, V., Del Fabbro, C., Alaux, M.l., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A.l., Le Clainche, I., Malacrida, G., Durand, E.o., Pesole, G., Laucou, V.r., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F.o., Pää, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.-F.o., Weissenbach, J., Quatier, F., and Wincker, P.** (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463-467.
- Jauhar, P.** (2007). Meiotic restitution in wheat polyploid (amphihaploids): a potent evolutionary force. *J Hered* **98**, 188-193.
- Jiang, N., Bao, Z., Zhang, X., Eddy, S., and Wessler, S.** (2004). Pack-MULE transposable elements mediate gene evolution in plants. *Nature* **431**, 569-573.
- Jofuku, K., den Boer, B., Van Montagu, M., and Okamoto, J.** (1994). Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *Plant. Cell* **6**, 1211-1225.
- Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J.** (2005). Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462-467.
- Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., and al., e.** (2002). The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol.* **3**.
- Kapitonov, V., and Jurka, J.** (2004). Harbinger transposons and an ancient HARBI1 gene derived from a transposase. *DNA Cell Biol* **23**, 311-324.
- Kapitonov, V., and Jurka, J.** (2006). Self-synthesizing DNA transposons in eukaryotes. *Proc Natl. Acad. Sci.* **103**, 4540-4545.
- Kapitonov, V.V., and Jurka, J.** (2001). Rolling-circle transposons in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 8714-8719.
- Kashkush, K., Feldman, M., and Levy, A.A.** (2003). Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics* **33**, 102-106.
- Kato, K., Miura, H., and Sawada, S.** (1999). QTL mapping of genes controlling ear emergence time and plant height on chromosome 5A of wheat. *Theor Appl Genet* **98**, 472-476.
- Kellogg, E.** (2001). Evolutionary history of the grasses. *Plant Physiol.* **125**, 1198-1205.
- Kerber, R., and Rowland, G.** (1974). Origin of the threshing character in hexaploid wheat. *Can J Genet Cytol* **16**, 145-154.
- Khush, G.S.** (1973). *Cytogenetics of aneuploids* (New York.).
- Kidwell, M.G.** (2002). Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**, 49-63.
- Kihara, H.** (1919). Ueber cytologische Studien bei einigen Getreidearten. 1. Spezies-Bastarde des Weizens und Weizenroggen-Bastard. *Bot. Mag.* **32**, 17-38.
- Kihara, H.** (1944). Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare* (Japanese). *Agric. Hort.* **19**, 13-14.

- Kilian, B., Ozkan, H., Walthe, A., Kohl, J., Dagan, T., Salamini, F., and Martin, W.** (2007). Molecular diversity at 18 loci in 321 wild and 92 domesticated lines reveal no reduction of nucleotide diversity during *Triticum monococcum* (einkorn) domestication: Implications for the origin of agriculture. *Mol Biol Evol* **24**, 2657-2668.
- Kim, V.N.** (2005). Minireview Molecules and Small RNAs : Classification , Biogenesis , and Function **19**, 1-15.
- Kimberg, G., and Sallee, P.** (1973). The regulation of chromosome pairing in hybrids of wheat. *Cereal Research Communication* **3**, 5-12.
- Kingsbury, M., Friedman, B., McConnell, M., Rehen, S., and Yang, A.e.a.** (2005). Aneuploid neurons are functionally active and integrated into brain circuitry. *Proc Natl Acad Sci U S A* **102**, 6143-6147.
- Kippes, N. et al.** (2014). Fine mapping and epistatic interactions of the vernalization gene VRN-D4 in hexaploid wheat. *Molecular genetics and genomics : MGG* **289**, 47–62.
- Kislev, M.** (1997). Early agriculture and palaeoecology of Netiv Hagdud. In *An early Neolithic village in the Jordan valley. Part I: the archaeology of Netiv Hagdud*, O. Bar-Yosef and A. Gopher, eds (Cambridge: Peabody Museum of Archaeology and Ethnology), pp. 203-230.
- Kojima, T., and Ogihara, Y.** (1998). High-resolution RFLP map of the long arm of chromosome 5A in wheats and its synteny among cereals. *Genes Genet Syst* **73**, 51- 58.
- Komaki, M., Okada, K., Nishino, E., and Shimura, Y.** (1988). Isolation and characterization of novel mutants of *Arabidopsis thaliana* defective in flower development. *Development* **104**, 195.
- Konishi, S., and al., e.** (2006). An SNP caused loss of seed shattering during rice domestication. *Science* **312**, 1392–1396.
- Kuckuck, H.** (1959). Neuere Arbeiten zur Entstehung der hexaploiden Kulturweizen. *Z. Pflanzenzucht* **41**, 205-226.
- Kunst, L., Klenz, J.E., Martinez-Zapater, J., and Haughn, G.W.** (1989). AP2 gene determines the identity of perianth organs in flowers of *Arabidopsis thaliana*. *Plant Cell* **1**, 1195–1208.
- Kurarporthy, V., Sood, S., Dhaliwal, H., Chhuneja, P., and Gill, B.S.** (2007). Identification and mapping of a tiller inhibition gene (*tin3*) in wheat. *Theor Appl Genet* **114**, 285-294.
- Ladizinsky, G.** (1987a). pulse domestication before cultivation. *Econ. Bot.* **41**, 60-65.
- Ladizinsky, G.** (1987b). Collection of wild cereals in the upper Jordan valley. . *Econ. Bot.* **29**, 264-267.
- Ladizinsky, G.** (1998a). Plant evolution under domestication.
- Ladizinsky, G.** (1998b). How many tough-rachis mutants gave rise to domesticated barley. *genetic Resources and Crop Evolution* **45**, 411-414.
- Ladizinsky, G., Smartt, J., and Simmonds, N.** (1995). *Evolution of crop plants.* (Harlow: Longman).
- Lai, J., Li, Y., Messing, J., and Dooner, H.K.** (2005). Gene movement by Helitron transposons contributes to the haplotype variability of maize. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 9068-9073.
- Lal, S., Oetjens, M., and Hannah, L.C.** (2009). Helitrons: Enigmatic abductors and mobilizers of host genome sequences. *Plant Science* **176**, 181-186.

- Lal, S.K., and Hannah, L.C.** (2005). Plant genomes: massive changes of the maize genome are caused by Helitrons. *Heredity* **95**, 421-422.
- Lal, S.K., Giroux, M.J., Brendel, V., Vallejos, C.E., and Hannah, L.C.** (2003). The Maize Genome Contains a Helitron Insertion. *Society* **15**, 381-391.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., and al., e.** (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.
- Law, C., Worland, A., and Giorgi, B.** (1975). The genetic control of ear-emergence time by chromosomes 5A and 5D of wheat. *Heredity* **36**, 49-58.
- Le, Q., Wright, S., Yu, Z., and Bureau, T.** (2000). Transposon diversity in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **97**, 7376-7381.
- Leitch, A.R., and Leitch, I.J.** (2008). Genomic plasticity and the diversity of polyploid plants. *Science* **320**, 481-483.
- Leitch, I., and Bennett, M.** (1997). Polyploidy in angiosperms. *Trends in Plant Science* **2**, 470-476.
- Lev-Yadun, S., Gopher, A., and Abbo, S.** (2000). Archaeology. The cradle of agriculture. *Science* **288**, 1602-1603.
- Lévi-Strauss, C.** (1962). *La Pensée sauvage*. (Plon).
- Levis, R., Ganesan, R., Houtchens, K., Tolar, L., and Sheen, F.** (1993). Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* **75**, 1083-1093.
- Li, W., Huang, L., and Gill, B.S.** (2008). Recurrent deletions of puroindoline genes at the grain hardness locus in four independent lineages of polyploid wheat. *Plant Physiology* **146**, 200-212.
- Li, Y., Dooner, H.K., and Brunswick, N.** (2009). Excision of Helitron transposons in maize. *Science* **4684**.
- Li, X., Liu, Y.** (2010) The conversion of spring wheat into winter wheat and vice versa: false claim or Lamarckian inheritance? *Journal of biosciences* **35**, 321-325.
- Lijavetsky, D., Muzzi, G., Wicker, T., Keller, B., Wing, R., and Dubcovsky, J.** (1999). Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. *Genome* **42**, 1176-1182.
- Lim, K.B., Ramanna, M.S., De Jong, J.H., Jacobsen, E., and Van Tuyl, J.M.** (2001). Indeterminate restitution (IMR): a novel type of meiotic nuclear restitution mechanism detected in interspecific lily hybrids by GISH. *Theor. Appl. Genet.* **103**, 219-230.
- Liu, B., Vega, J.M., and Feldman, M.** (1998). Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. *Genome* **41**, 535-542.
- Luo, M.C., Yang, Z.L., and Dvorak, J.** (2000). The wheat. *Theor Appl Genet* **100**, 602-606.
- Lynch, C., and Tristem, M.** (2003). A co-opted gypsy-type LTR-retro-transposon is conserved in the genomes of humans, sheep, mice and rats. *Curr Biol* **13**, 1518-1523.
- Lynch, M., and Force, A.** (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459-473.
- Lysak, M.A., Berr, A., Pecinka, A., Schmidt, R., McBreen, K., and Schubert, I.** (2006). Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc Natl Acad Sci U S A* **103**, 5224-5229.

- Ma, J., and Bennetzen, J.L.** (2004a). Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12404-12410.
- MacKey, J.** (1954). Neutron and X-ray experiments in wheat and revision of the speltoid problem. *Hereditas* **40**, 65-180.
- Madlung, A., Tyagi, A.P., Watson, B., Jiang, H., Kagochi, T., Doerge, R.W., Martienssen, R., and Comai, L.** (2005). Genomic changes in synthetic *Arabidopsis* polyploids. *Plant J* **41**, 221-230.
- Malinsky, S., Bucheton, A., and Busseau, I.** (2000). New insights on homology-dependent silencing of I factor activity by transgenes containing ORF1 in *Drosophila melanogaster*. *Genetics* **156**, 1147-1155.
- Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C., and Paszkowski, J.** (2007). Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG methylation. *Cell* **130**, 851-862.
- Matsuoka, Y., and Nasuda, S.** (2004). Durum wheat as a candidate for the unknown female progenitor of bread wheat: an empirical study with a highly fertile F1 hybrid with *Aegilops tauschii* Coss. *Theor Appl Genet* **109**, 1710-1717.
- Matzke, M.A., Mette, M.F., Kanno, T., and Matzke, A.J.** (2003). Does the intrinsic instability of aneuploid genomes have a causal role in cancer? . *Trends Genet* **19**, 253-256.
- McClintock, B.** (1984). The significance of responses of the genome to challenge. *Science* **226**, 792-801.
- McFadden, E.S., and Sears, E.R.** (1944). The artificial synthesis of *Triticum spelta*. . *Rec. Soc. Genet. Am.* **13**, 26-27.
- McFadden, E.S., and Sears, E.R.** (1946). The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *J. Hered.* **37**, 81-89, 107-116.
- Mendiola, M., Bernales, I., and Cruz, F.** (1994). Differential roles of the transposon termini in IS91 transposition. *Proc Natl Acad Sci USA* **91**, 1922-1926.
- Mestiri, I., Chague, V., Tanguy, A.M., Huneau, C., Huteau, V., Belcram, H., Coriton, O., Chalhoub, B., and Jahier, J.** (2010). Newly synthesized wheat allohexaploids display progenitor-dependent meiotic stability and aneuploidy but structural genomic additivity. *New Phytol* **186**, 86-101.
- Miller, T., and Reader, S.** (1982). A major deletion of part of chromosome 5A of *Triticum aestivum*. *Wheat Inf Serv* **88**, 10-12.
- Mitrofanova, O.** (1997). The inheritance and effect of Cp (Compact plant) mutation induced in common wheat. *Russian J. Genet* **33**, 393-398.
- Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A., and Rafalski, A.** (2005). Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nature genetics* **37**, 997-1002.
- Morrell, P.L., and Clegg, M.T.** (2007). Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 3289-3294.
- Morris, R., and Sears, E.R.** (1967). The cytogenetics of wheat and its relatives. In *Wheat and wheat improvement.*, K.S. Quisenberry and L.P. Reitz, eds (Madison, WI, USA: American Society of Agronomy), pp. 19-87.
- Mukai, Y., Endo, T.R., and Gill, B.S.** (1991). Physical mapping of the 18s-26s rRNA multigene family in common wheat: identification of a new locus. . *Chromosoma* **100**, 71-78.
- Muntzing, A.** (1936). The evolutionary significance of autopolyploidy. *Hereditas* **21**, 263-378.

- Murad, L., and al., e.** (2004). The origin and evolution of geminivirus-related DNA sequences in *Nicotiana*. *Heredity* **92**, 352–358.
- Muramatsu, M.** (1963). Dosage effect of the *spelta* gene *q* of hexaploid wheats. *Genetics* **48**, 469-482.
- Nair, S.K., Wang, N., Turuspekov, Y., Pourkheirandish, M., Sinsuwongwat, S., Chen, G., Sameri, M., Tagiri, A., Honda, I., Watanabe, Y., Kanamori, H., Wicker, T., Stein, N., Nagamura, Y., Matsumoto, T., and Komatsuda, T.** (2010). Cleistogamous flowering in barley arises from the suppression of microRNA-guided HvAP2 mRNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 490-495.
- Nalam, V., Vales, M., Watson, C., Kianian, S., and Riera-Lizarazu, O.** (2006). Map based analysis of genes affecting the brittle rachis character in tetraploid wheat (*Triticum turgidum*). *Theor Appl Genet* **112**, 373-381.
- Napoli, C., Lemieux, C., and Jorgensen, R.** (1990). Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell* **2**, 279–289.
- Nesbitt, M.** (2004). Can We identify a center, a region, or a supra-region for Near Eastern Plant Domestication ? Neo-lithic 1/04, 38-40.
- Newton, W.C.F., and Pellew, C.** (1929). *Primula kewensis* and its derivatives. . *J. Genet.* **20**, 405-467.
- Nicolas, S.D., Le Mignon, G., Eber, F., Coriton, O., Monod, H., Clouet, V., Huteau, V., Lostanlen, A., Delourme, R., Chalhou, B., Ryder, C.D., Chevre, A.M., and Jenczewski, E.** (2007). Homeologous recombination plays a major role in chromosome rearrangements that occur during meiosis of *Brassica napus* haploids. *Genetics* **175**, 487-503.
- Nilsson-Ehle, H.** (1911). Kreuzungsuntersuchungen an Hafer und Weizen. II. *Lunds Univ. AD rsskr NF Afd. 2*, N6: 84 ss.
- Nilsson-Ehle, H.** (1917). Untersuchungen über speltoid mutationen beim weizen. *Botan Notiser* **305**, 3020.
- Nilsson, B.** (1973). A bibliography of literature concerning chromosome identification with special reference to fluorescence and Giemsa staining techniques. *Hereditas* **73**, 259-270.
- Nobuta, K., Lu, C., Shrivastava, R., Pillay, M., DePaoli, E., and al., e.** (2008). Distinct size distribution of endogenous siRNAs in maize: evidence from deep sequencing in the *mop1-1* mutant. *Proc. Natl. Acad. Sci. USA* **105**, 14958–14963.
- Novina, C.D., and Sharp, P.a.T.** (2004). The RNAi revolution. *Nature* **430**, 161–164.
- Ogihara, Y., Hasegawa, K., and Tsujimoto, H.** (1994). High-resolution cytological mapping of the long arm of chromosome 5A in common wheat using a series of deletion lines induced by gametocidal (*Gc*) genes of *Aegilops speltoides*. *Gen Genet* **244**, 253-259.
- Orgel, L.E., and Crick, H.C.** (1980). SelfishDNA: The ultimate parasite. . *Nature* **284**, 604–607.
- Osborn, T.C., Butrille, D.V., Sharpe, A.G., Pickering, K.J., Parkin, I.A., Parker, J.S., and Lydiate, D.J.** (2003a). Detection and effects of a homeologous reciprocal transposition in *Brassica napus*. *Genetics* **165**, 1569-1577.
- Osborn, T.C., Pires, J.C., Birchler, J.A., Auger, D.L., Chen, Z.J., Lee, H.S., Comai, L., Madlung, A., Doerge, R.W., Colot, V., and Martienssen, R.A.** (2003b). Understanding mechanisms of novel gene expression in polyploids. *Trends Genet* **19**, 141-147.
- Otto, S.P., and Whitton, J.** (2000). Polyploid incidence and evolution. *Annu Rev Genet* **34**, 401-437.

- Ozkan, H., Levy, A.A., and Feldman, M.** (2001). Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* **13**, 1735-1747.
- Ozkan, 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.
- Özkan, H., Willcox, G., Graner, A., Salamini, F., and Kilian, B.** (2010). Geographic distribution and domestication of wild emmer wheat (*Triticum dicoccoides*). *Genetic Resources and Crop Evolution* **58**, 11-53.
- Pang, M., Woodward, A.W., Agarwal, V., Guan, X., Ha, M., Ramachandran, V., Chen, X., Triplett, B.A., Stelly, D.M., and Chen, Z.J.** (2009). Genome-wide analysis reveals rapid and dynamic changes in miRNA and siRNA sequence and expression during ovule and fiber development in allotetraploid cotton (*Gossypium hirsutum* L.). *Genome biology* **10**, R122.
- Paterson, A., and al., e.** (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature* **457**, 551-556.
- Paux, E., Roger, D., Badaeva, E., Gay, G., Bernard, M., Sourdille, P., and Feuillet, C.** (2006). Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. *The Plant journal : for cell and molecular biology* **48**, 463-474.
- Peleg, Z., Fahima, T., Korol, A.B., Abbo, S., and Saranga, Y.** (2011). Genetic analysis of wheat domestication and evolution under domestication. *J Exp Bot* **62**, 5051-5061.
- Peng, J., Ronin, Y., Fahima, T., Röder, M., Li, Y., Nevo, E., and Korol, A.** (2003). Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat. *Proc Natl Acad Sci USA* **100**, 2489-2494.
- Peng, J.H., Sun, D., and Nevo, E.** (2011). Domestication evolution, genetics and genomics in wheat. *Molecular Breeding* **28**, 281-301.
- Petrov, D.A.** (2002a). Mutational equilibrium model of genome size evolution. *Theor. Popul. Biol.* **61**, 531-544.
- Petrov, D.A., A., S.T., Johnston, J.S., Hartl, D.L., and Shaw, K.L.** (2000). Evidence for DNA loss as a determinant of genome size. *Science* **287**, 1060-1062.
- Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., and al., e.** (2006). Doubling genome size without polyploidization: dynamics of retrotransposon-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res.* **16**, 1262-1269.
- Pritham, E.** (2009). Transposable elements and factors influencing their success in eukaryotes. *J. Hered.* **100**, 648-655.
- Purugganan, M.D., and Fuller, D.Q.** (2009). The nature of selection during plant domestication. *Nature* **457**, 843-848.
- Ragupathy, R., and Cloutier, S.** (2008). Genome organisation and retrotransposon driven molecular evolution of the endosperm Hardness (Ha) locus in *Triticum aestivum* cv Glenlea. *Molecular genetics and genomics : MGG* **280**, 467-481.
- Ramsey, J., and Schemske, D.W.** (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* **29**, 467-501.
- Ramsey, J., and Schemske, D.W.** (2002). Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics* **33**, 589-639.

- Rao, M.** (1973). Mapping of the compactum gene C on chromosome 2D of wheat. *Wheat Information Service* **35**.
- Rapp, R.A., Udall, J.A., and Wendel, J.F.** (2009). Genomic expression dominance in allopolyploids. *BMC Biol* **7**, 18.
- Riddle, N.C., and Birchler, J.A.** (2003). Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybrids. *Trends Genet* **19**, 597-600.
- Riley, R., and Kimber, G.** (1961). Aneuploids and the cytogenetic structure of wheat varietal populations. *Heredity* **16**, 275-290.
- Sabot, F.o., Simon, D., and Bernard, M.** (2004). Plant transposable elements, with an emphasis on grass species. *Euphytica* **139**, 227-247.
- Sakamura, T.** (1918). Kurze Mitteilung uber die Chromosomenzahlen und die Verwandtschaftsverhältnisse der Triticum-Arten. *Bot. Mag.* **32**, 151-154.
- Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P.** (2011). A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language? *Cell* **146**, 353-358.
- Salse, J., Abrouk, M., Bolot, S., Guilhot, N., Courcelle, E., Faraut, T., Waugh, R., Close, T.J., Messing, J., and Feuillet, C.** (2009). Reconstruction of monocotyledonous protochromosomes reveals faster evolution in plants than in animals. *Proc Natl Acad Sci USA* **106**, 14908-14913.
- Salse, J.r., ChaguÃ©, V.r., Bolot, S.p., Magdelenat, G., Huneau, C.c., Pont, C., Belcram, H., Couloux, A., Gardais, S., Evrard, A.l., Segurens, B.a., Charles, M., Ravel, C., Samain, S., Charmet, G., Boudet, N., and Chalhou, B.** (2008a). 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* **9**, 555-555.
- Salse, J.r.m., Bolot, S.p., Throude, M.l., Jouffe, V., Piegu, B.t., Quraishi, U.M., Calcagno, T., Cooke, R., Delseny, M., and Feuillet, C.** (2008b). Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *The Plant cell* **20**, 11-24.
- SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., and Bennetzen, J.L.** (1998). The paleontology of intergene retrotransposons of maize. *Nat. Genet.* **20**, 43-45.
- SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., and al., e.** (1996). Nested retrotransposons in the inter-genic regions of the maize genome. *Science* **274**, 765-768.
- SanMiguel, P.J., Ramakrishna, W., Bennetzen, J.L., Busso, C.S., and Dubcovsky, J.** (2002). Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A(m). *Functional & integrative genomics* **2**, 70-80.
- Sax, K.** (1922). Sterility in wheat hybrids. II. Chromosome behavior in partially sterile hybrids. *Genetics* **7**, 513-550.
- Schwarz, D., Hutvágner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.** (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199-208.
- Sears, E.** (1952). Misdivision of univalents in common wheat. *Chromosoma* **4**, 535- 550.
- Sears, E.** (1966). Nullisomic-tetrasomic combinations in hexaploid wheat. In *Chromosome manipulations and plant genetics*, R. Rilly and K. Lewis, eds (Edinburgh: Oliver and Boyd), pp. 29-45.
- Sears, E., and MS., S.** (1978). The telocentric chromosomes of common wheat. In *Proceeding of the 6th International wheat genetics Symposium*, R. S, ed (New Dehli: Indian Society of Genetics and Plant Breeding).

- Severinghaus, J.P., Sowers, T., Brook, E.J., Alley, R.B., and Bender, M.L.** (1998). Timing of abrupt climate change at the end of the Younger Dryas interval from thermally fractionated gases in polar ice. *Nature* **391**, 141–146.
- Sharma, H., and Waynes, J.** (1980). Inheritance of tough rachis in crosses of *Triticum monococcum* and *Triticum boeoticum*. *J Hered* **7**, 214–216.
- Shewry, P.R.** (2009). Wheat. *Journal of experimental botany* **60**, 1537-1553.
- Shitsukawa, N., Tahira, C., Kassai, K., Hirabayashi, C., Shimizu, T., Takumi, S., Mochida, K., Kawaura, K., Ogihara, Y., and Murai, K.** (2007). Genetic and epigenetic alteration among three homoeologous genes of a class E MADS box gene in hexaploid wheat. *Plant Cell* **19**, 1723-1737.
- Simons, K.J., Fellers, J.P., Trick, H.N., Zhang, Z., Tai, Y.-S., Gill, B.S., and Faris, J.D.** (2006). Molecular characterization of the major wheat domestication gene Q. *Genetics* **172**, 547-555.
- Slotkin, R.K., and Martienssen, R.** (2007). Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**, 272-285.
- Soltis, D.E., and Soltis, P.S.** (1995). The dynamic nature of polyploid genomes. *Proc Natl Acad Sci U S A* **92**, 8089-8091.
- Soltis, D.E., Albert, V.a., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., Depamphilis, C.W., Wall, P.K., and Soltis, P.S.** (2009). Polyploidy and angiosperm diversification. *American journal of botany* **96**, 336-348.
- Soltis, P.S., and Soltis, D.E.** (2004). The origin and diversification of angiosperms. *Am J Bot* **91**, 1614-1626.
- Soltis, P.S., and Soltis, D.E.** (2009). The role of hybridization in plant speciation. *Annu Rev Plant Biol* **60**, 561-588.
- Song, K., Lu, P., Tang, K., and Osborn, T.C.** (1995). Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proc Natl Acad Sci U S A* **92**, 7719-7723.
- Stebbins, G.** (1971). *Chromosomal Evolution in Higher Plants*. (Ltd., London: Edward, A.).
- Stebbins, G.L.** (1950). *Variation and evolution in plants*. (New York.: Columbia University Press).
- Storlie, E.W., and Talbert, L.E.** (1993). Cause of Tall Off-Types in a Semidwarf Spring Wheat. *Crop Science* **33**, 1131-1135.
- Swaminathan, M.** (1963). Induced mutations at the Q locus in relation to the phylogeny of hexaploid *Triticum* species.
- Swaminathan, M.** (1966). Mutational analysis of the hexaploid *Triticum* complex // *Proc. 2nd Int Wheat Genet. Symp. Hereditas Suppl.* **2**, 418–438.
- Syncelli, G.** *Chronogr.*, fol. 1652, p. 28.
- Talbert, L.E., Smith, L.Y., and Blake, N.K.** (1998). More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. *Genome* **41**, 402-407.
- Tang, H., Bowers, J.E., Wang, X., and Paterson, A.H.** (2010). Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 472-477.
- Tanno, K.-I., and Willcox, G.** (2006). How fast was wild wheat domesticated? *Science (New York, N.Y.)* **311**, 1886-1886.

- Tsujimoto, H., and Noda, K.** (1989). Structure of chromosome 5A of *Aegilops speltoides* wheat speltoid mutants induced by the gametocidal genes of *Aegilops speltoides*. *Genome* **32**, 1085-1090.
- Tsunewaki, K., and Jenkins, B.** (1961). Monosomic and conventional analyses of genes in common wheat. II. *Jpn J Genet* **36**, 428-443.
- Udall, J.A., Quijada, P.A., and Osborn, T.C.** (2005). Detection of chromosomal rearrangements derived from homologous recombination in four mapping populations of *Brassica napus* L. *Genetics* **169**, 967-979.
- Van de Peer, Y., Maere, S., and Meyer, A.** (2009b). The evolutionary significance of ancient genome duplications. *Nat Rev Genet* **10**, 725-732.
- Van de Peer, Y., Fawcett, J.A., Proost, S., Sterck, L., and Vandepoele, K.** (2009a). The flowering world: a tale of duplications. *Trends Plant Sci* **14**, 680-688.
- Wang, J., Tian, L., Lee, H.-S., Wei, N.E., Jiang, H., Watson, B., Madlung, A., Osborn, T.C., Doerge, R.W., Comai, L., and Chen, Z.J.** (2006). Genomewide nonadditive gene regulation in *Arabidopsis* allotraploids. *Genetics* **172**, 507-517.
- Wang, Q., and Dooner, H.K.** (2006b). Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 17644-17649.
- Watanabe, N.** (2005). The occurrence and inheritance of a brittle rachis phenotype in Italian durum wheat cultivars. *Euphytica* **142**, 247-251.
- Watkins, A.** (1927). Genetic and cytological studies in wheat. III. *J Genet Plant Breed* **18**, 375-396.
- Weiss, E., Kislev, M., and Hartmann, A.** (2006). Anthropology. Autonomous cultivation before domestication. *Science* **312**, 1608-1610.
- Wendel, J., and Doyle, J.** (2005). Polyploidy and evolution in plants. In *Plant Diversity and Evolution.*, R. Henry, ed (Wallingford: CABI Publishing
- Wendel, J.F.** (2000). Genome evolution in polyploids. *Plant molecular biology* **42**, 225-249.
- Wendel, J.F., Schnabel, A., and Seelanan, T.** (1995). An unusual ribosomal DNA sequence from *Gossypium gossypoides* reveals ancient, cryptic, intergenomic introgression. *Mol Phylogenet Evol* **4**, 298-313.
- Wendel, J.F., Cronn, R.C., Johnston, J.S., and Price, H.J.** (2002). Feast and famine in plant genomes. *Genetica* **115**.
- Werk, D., and al., e.** (2005). Developing an effective RNA interference strategy against a plus-strand RNA virus: silencing of coxsackievirus B3 and its cognate coxsackievirus-adenovirus receptor. *Biol. Chem* **386**, 857-863.
- Wicker, T., and Keller, B.** (2007b). Genome-wide comparative analysis of copia retrotransposons in Triticeae, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. *Genome research* **17**, 1072-1081.
- Wicker, T., Yahiaoui, N., Guyot, R., Schlagenhauf, E., Liu, Z.D., and al., e.** (2003b). Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and Am genomes of wheat. *Plant Cell* **15**, 1186-1197.
- Wicker, T., Zimmermann, W., Perovic, D., Paterson, A.H., Ganai, M., Graner, A., and Stein, N.** (2005). A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley Hv-eIF4E locus: recombination, rearrangements and repeats. *The Plant journal : for cell and molecular biology* **41**, 184-194.

- Wicker, T., Sabot, F.o., Hua-Van, A.I., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., and Schulman, A.H.** (2007a). A unified classification system for eukaryotic transposable elements. *Nature reviews. Genetics* **8**, 973-982.
- Willcox, G.** (2012a). The Beginnings of Cereal Cultivation and Domestication in Southwest Asia (
- Winge.** (1929). Zytologische untersuchungen uber speltoids und andere mutantenabnliche aberranten beim weizen. *Hereditas* **5**, 241-286.
- Wolfe, K.H.** (2001). Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet* **2**, 333-341.
- Wolfe, K.H., and Shields, D.C.** (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708-713.
- Worland, A.J., and Law, C.N.** (1985). Aneuploidy in semi dwarf wheat varieties. *Euphytica* **34**, 317-327.
- Yamamori, M., Nakaraura, T., Endo, T., and Nagamine, T.** (1994). Waxy protein deficiency and chromosomal location of coding genes In common wheat. . *Theor Appl Genet* **89**, 179-184.
- Yang, L., and Bennetzen, J.L.** (2009). Structure-based discovery and description of plant and animal Helitrons. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12832-12837.
- 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. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 3447-3452.
- Zhang, Z., Belcram, H., Gornicki, P., Charles, M., Just, J.r.m., Huneau, C.c., Magdelenat, G., Couloux, A., Samain, S., Gill, B.S., Rasmussen, J.B., Barbe, V., Faris, J.D., and Chalhoub, B.** (2011). Duplication and partitioning in evolution and function of homoeologous Q loci governing domestication characters in polyploid wheat. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 18737-18742.
- Zohary, D.** (1996). Monophyletic vs. polyphyletic origin of the crops on wich agriculture was founded in the Near East. *genetic Resources and Crop Evolution* **46**, 133-142.
- Zohary, D., and Feldman, M.** (1962). Hybridization between amphidiploids and the evolution of polyploids in the wheat (*Ægilops-Triticum*) group. . *Evolution* **16**, 4461.
- Zohary, D., and Hopf, M.** (2000). In *Domestication of plants in the Old World*, C. Press, ed (Oxford.
- Zuccolo, A., Sebastian, A., Talag, J., Yu, Y., Kim, H., and al., e.** (2007). Transposable element distribution, abundance and role in genome size variation in the genus *Oryza*. *BMC Evol. Biol.* **7**.