

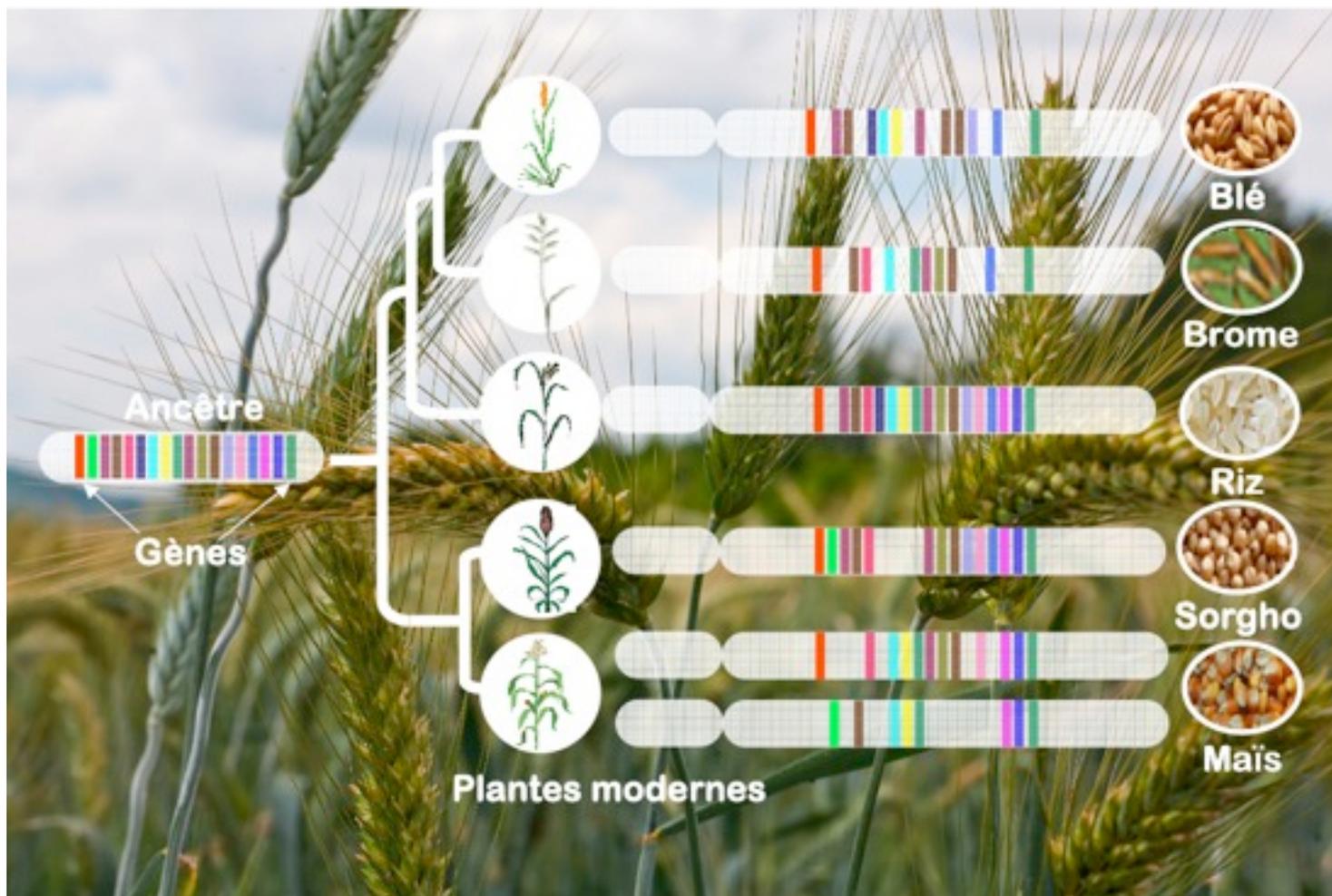


La paléogénomique des plantes pour l'amélioration variétale

Jérôme Salse

Directeur de Recherche INRA - Clermont-Ferrand Theix

Manuscrit révisé le 9 novembre 2012 - Publié le 28 octobre 2013



Résumé : L'accès croissant à la séquence des génomes durant la dernière décennie a permis de reconstruire l'histoire des plantes au cours de 150 millions d'années d'évolution. La comparaison de ces génomes a mis en évidence leurs ancêtres communs ainsi que les mécanismes ayant donné naissance aux plantes modernes. Il a ainsi pu être démontré que les plantes modernes sont issues d'ancêtres constitués de 5 à 7 chromosomes ancestraux porteurs de 10.000 gènes servant de fonctions biologiques fondatrices du règne végétal. Un mécanisme évolutif majeur, consistant en l'apparition récurrente de copies de gènes dupliqués, a permis la genèse de nouvelles espèces végétales, ainsi que leur adaptation à leur environnement. Ces résultats fournissent de nouvelles connaissances fondamentales permettant de mieux comprendre l'évolution des espèces, mais délivrent aussi des outils appliqués pour améliorer certains caractères agronomiques importants dans les programmes de sélection variétale. Le modèle évolutif proposé, qui définit précisément les régions chromosomiques des génomes de plantes modernes porteuses de gènes communs, permet désormais d'utiliser les connaissances acquises sur les génomes séquencés pour améliorer, chez des espèces plus complexes, des caractéristiques agronomiques aussi importantes que le rendement, la qualité ou la tolérance aux stress.

Glossaire

- (1) **Gène** : unité d'hérédité contrôlant un caractère particulier. Cet élément génétique correspond à un ou des segments d'ADN.
- (2) **Marqueur** : fragments spécifiques d'ADN pouvant être identifiés et positionnés au sein du génome complet.
- (3) **Génome** : ensemble du matériel génétique d'un individu. Patrimoine héréditaire d'un individu structuré en chromosomes.
- (4) **Carte génétique** : agencement de marqueurs moléculaires (balises) le long des chromosomes.
- (5) **Synténie** : conservation de groupes de gènes entre deux génomes.
- (6) **Polypléidie** : duplication totale d'un génome par la présence dans le même noyau de plusieurs jeux de chromosomes homologues.
- (7) **Paléogénomique** : étude de la structure et de la fonction des génomes ancêtres des espèces vivantes actuelles.
- (8) **Orthologie** : propriété de gènes homologues issus d'un gène ancestral unique par mutations.
- (9) **Paralogie** : propriété de gènes homologues issus d'événements de recombinaisons résultants d'insertions, duplications ou de polypléidie.
- (10) **QTL (Quantitative Trait Locus)** : locus à effet quantitatif, c'est-à-dire région du génome dont la variation allélique est associée à la variation d'un caractère phénotypique quantitatif, tel que la taille, le rendement...
- (11) **Génomique translationnelle** : transfert de connaissances acquises sur une espèce en termes de structure et fonction des gènes à une autre espèce apparentée grâce aux relations de synténie entre ces génomes.

PRÉAMBULE

Les plantes constituent une composante fondamentale de notre environnement et génèrent l'essentiel de la biomasse présente à la surface de la terre. Elles représentent aussi la plus grande source de nourriture du règne animal et leur domestication par l'homme, pour assurer son alimentation, a été déterminante dans sa capacité à peupler la planète. L'accès à la séquence des génomes d'espèces modernes ces dix dernières années a permis d'améliorer la compréhension de l'organisation des génomes des plantes. À ce jour, les génomes de 13 espèces ont été totalement ou partiellement séquencés : le riz, le maïs, le sorgho, le brome, la vigne, l'arabette, le peuplier, le soja, le papayer, le fraisier, le cacaoyer, le lotier, le pommier (Tableau 1). D'autres espèces d'intérêt agronomique majeur (pêcher, pois, blé, orge par exemple, cf Tableau 1) ne sont pas encore séquencées, mais nous disposons dans ces cas de cartes génétiques ou une fraction des gènes a été positionnée le long des chromosomes. Ces différents génomes contiennent plus de 20.000 et jusqu'à près de 50.000 gènes pour un nombre de chromosomes allant de 5 à plus de 20.

Les recherches en paléogénomique visent à comprendre comment les génomes de ces espèces se sont constitués au cours du temps à partir d'ancêtres fondateurs dont la restructure génomique peut être modélisée en termes de chromosomes et gènes ancestraux.

LA PALEOGÉNOMIQUE REVISITÉE PAR LES PROGRÈS DU SÉQUENÇAGE.

La paléogénomique, étude de la structure et de la fonction des génomes ancêtres des espèces vivant à l'heure actuelle, permet l'identification et la caractérisation des mécanismes (duplications, inversions et délétions...) qui ont façonné le génome des espèces modernes au cours de leur évolution. Deux approches scientifiques permettent d'aborder cette étude. La première approche consiste à séquencer au moins partiellement l'ADN d'espèces ancestrales fossiles. Cette approche a été appliquée aux mammifères par l'accès à la séquence du génome du mammoth (Poinar *et al.* 2006) ou de l'homme de Néandertal (Green *et al.* 2008) par exemple. La connaissance du génome des ancêtres fossiles donne des informations immédiates sur les événements et mécanismes ayant façonné les espèces modernes qui en ont dérivé.

La deuxième approche réside dans la modélisation du génome ancêtre, à la fois de sa structure chromosomique et de son contenu en gènes. Cette modélisation consiste à identifier la structure génomique minimale considérée comme ancestrale, à partir de la comparaison des génomes des espèces actuelles, *i.e.* analyse de synténie ou de génomique comparée. Chez les vertébrés, les travaux de paléogénomique réalisés par Nakatani *et al.* (2007) ont permis d'établir un modèle évolutif sur 130 millions d'années (ci-après abrégé MA) à partir d'un ancêtre fondateur constitué de 10 à 12 protochromosomes (Murat *et al.* 2012). Dans ce modèle tous les mammifères modernes sont issus de cet ancêtre fondateur. Les réarrangements génomiques (fusion, fission, délétion, duplication) subis au cours du temps expliquent alors comment l'Homme, la Souris, le Cheval, le Chimpanzé par exemple, ont un nombre de chromosomes et de gènes différents à l'heure actuelle. Toutefois, malgré des structures de génomes distincts, ils partagent toujours près de 80% de gènes communs.

Contrairement aux mammifères, la paléogénomique est encore peu étudiée chez les plantes et plus précisément chez les Angiospermes (plantes à fleurs) qui ont subi un grand nombre de duplications segmentaires voire globales de leur génome (polyploïdisation) ainsi que de nombreux réarrangements locaux (inversion, délétion de gènes) ayant longtemps rendu difficiles les travaux de génomique comparée au sein des monocotylédones (principalement les Graminées, dont notamment les céréales) et des dicotylédones (principalement les Rosidées, dont notamment les plantes légumières et les arbres fruitiers), comme il sera discuté dans la suite de l'article. Toutefois, les avancées majeures des techniques de séquençage au cours des dix dernières années ont permis d'avoir accès à l'information héréditaire portée par les différents chromosomes de nombreuses espèces végétales. Ainsi, les travaux de paléogénomique chez les plantes, par la comparaison de ces génomes, furent rendus possibles sur cette base par la stratégie illustrée dans la Figure 1.

En conclusion, la paléogénomique visant à reconstruire l'histoire évolutive des espèces vivant à l'heure actuelle peut être abordée par le séquençage d'ADN fossiles ou par la comparaison des génomes d'espèces modernes. Chez les plantes, l'accès récent à la séquence de nombreux génomes a permis de retracer leur histoire évolutive sur les derniers 200 millions d'années.

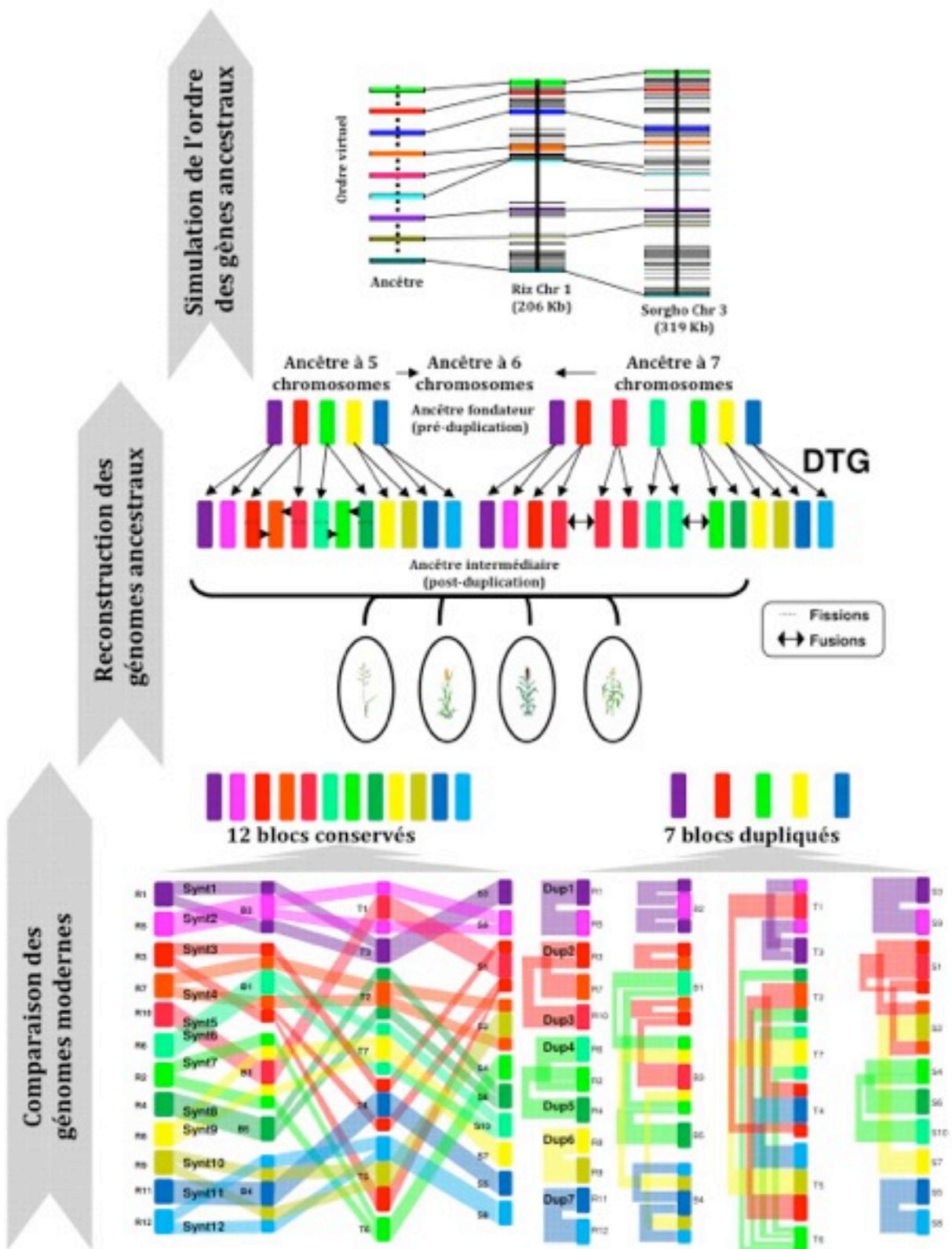


Figure 1 : Méthode de reconstruction des génomes ancestraux (adapté de Salse *et al.* 2009a)
 La reconstruction des génomes ancestraux se déroule en trois phases. La première phase (bas) consiste à comparer les génomes modernes afin d'identifier les blocs de synténie ou de conservation (à gauche, notés Synt1 à Synt 12) et les blocs de duplication (à droite, notés Dup1 à Dup7). Les espèces de céréales considérées à titre d'exemple sont le riz (12 chromosome notés R1 à R12), le brome (5 chromosome notés B1 à B5), le blé (7 chromosome notés T1 à T7, T pour *Triticeae*) et le sorgho (10 chromosome notés S1 à S10). La deuxième phase (milieu) réside dans l'intégration des régions conservées et dupliquées afin de mettre en évidence le nombre de blocs (ou chromosome) fondateurs et ainsi proposer un génome ancestral pré- et post-duplication. Enfin, la dernière phase (en haut) consiste pour chaque chromosome ancestral à reconstruire l'ordre des gènes en se basant sur le contenu/ordre en gène de ce même chromosome au sein des espèces modernes. DTG : Duplication Totale du Génome.

LES GÉNOMES DE PLANTES SONT CONSERVÉS DANS LEUR CONTENU EN GÈNES.

La génomique comparative, fondement de la paléogénomique, chez les plantes, a fait l'objet d'intenses recherches au cours des deux dernières décennies avec l'avènement des cartes génétiques porteuses de marqueurs moléculaires communs sans toutefois aboutir à des conclusions précises de paléogénomique, c'est-à-dire de caractérisation de génomes ancestraux ou de scénario évolutif précis. Nos travaux de paléogénomique chez les plantes menés depuis 2005 ont mis en œuvre la comparaison de 17 génomes référencés dans le Tableau ci-après. Ces génomes sont différents dans leur structure, que ce soit leur nombre de chromosomes (de 5 pour l'arabette (*Arabidopsis thaliana*) ou le brome jusqu'à 21 pour le blé), leur taille physique (de 119 millions de bases pour l'arabette à 17 milliards de bases pour le blé) corrélée avec le pourcentage de séquences non-géniques (ADN ne codant pas des protéines) et répétées (de moins de 20% pour l'arabette jusqu'à plus de 80% pour le blé) et leur nombre de gènes (de 27.228 gènes pour l'arabette jusqu'à 57.386 pour le pommier).

ESPÈCES	NOMS	STRUCTURE GÉNOMES ¹			EVOLUTION GÉNOMES ²			
		NOMBRE DE CHROMOSOMES	TAILLE (MPB) ET (% GÉNOME)	NOMBRE DE GÈNES	SYNTÉNIE NBR GÈNE / NBR BLOC / % GÉNOME	DUPLICATION NBR GÈNE / NBR BLOC / % GÉNOME	EQUATION CHROMOSOMIQUE ³	NOMBRE DE DTG
DICOTYLEDONES								
<i>Vitis vinifera</i>	Vigne	19	302 / 22	33514	GR	543 - 23 - 71	21+2-4	1
<i>Arabidopsis thaliana</i>	Arabette	5	119 / 19	27228	2389 - 80 - 99	1630 - 55 - 83	[21+10-22]x2-13	3
<i>Populus trichocarpa</i>	Peuplier	19	294 / 35	45654	4555 - 87 - 92	4164 - 46 - 73	[21+6-15]x2+4-9	2
<i>Glycine max</i>	Soja	20	949 / 50	46430	4013 - 164 - 97	9533 - 89 - 55	[21+1-16]x2x2+13-17	3
<i>Carica papaya</i>	Papayer	9	234 / 31	28629	3199 - 65 - 75	215 - 36 - 55	21+6-18	1
<i>Fragaria</i>	Fraisier	7	208 / 24	34809	3289 - 94 - 70	114 - 27 - 19	[21+3-15]-2	1
<i>Theobroma cacao</i>	Cacaoyer	10	218 / 26	28798	4472 - 21 - 81	370 - 19 - 66	21+2-13	1
<i>Lotus japonicus</i>	Lotier	6	462 / 40	30799	1720 - 80 - 61	145 - 32 - 35	[21+1-16]x2-6	2
<i>Malus x domestica</i>	Pommier	17	528 / 42	57386	3498 - 104 - 70	2845 - 69 - 59	[21+3-15]x2+4-5	2
<i>Prunus persica</i>	Pêcher	8	~218 / >20	784*	307-132-34	Nd**	[21+3-15]+1-2	1
<i>Pisum sativum</i>	Pois	7	~5000 / >50	214*	149-33-37	Nd**	[21+1-16]x2+4-9	2
Eudicot Total					27591 - 860 - 72	19559-396-57		
MONOCOTYLEDONES								
<i>Oryza sativa</i>	Riz	12	372 / 39	40577	GR	448-10-73	12	1
<i>Sorghum bicolor</i>	Sorgho	10	659 / 62	34496	6147-12-99	409-10-84	12+2-4	1
<i>Zea mays</i>	Mais	10	2365 / 84	32540	4454-30-82	3454-17-99	[12+2-4]x2+7-17	2
<i>Brachypodium distachyon</i>	Brome	5	271 / 28	25532	8533-12-99	642-13-79	12+7-14	1
<i>Triticum aestivum</i> ^f	Blé	21	~17Gb / >80	5003*	827-13-91	102-10-75	12+5-10 [x3 blé tendre]	1
<i>Hordeum vulgare</i> ^f	Orge	7	~5000 / >80	3423*	309-13-84	38-9-75	12+5-10	1
Monocot Total					20270-80-91	5093-69-81		
PLANTE TOTAL					47407-775-86	24652-465-69		

¹ Les paramètres structuraux des génomes de plantes (nombre chromosomes/gènes et taille) sont mentionnés comme référencés dans Salse 2012.

² Les paramètres évolutifs des génomes de plantes (synténie, duplication, équation) sont mentionnés comme référencés dans Salse 2012.

³ L'Équation Chromosomique décompose le nombre de chromosomes d'une espèce moderne par : le nombre de chromosome ancêtre (X), le nombre de duplication (éventuellement répétées n fois, x2n), le nombre de fission (+Y) et fusion (-Z) subies au cours de l'évolution; i.e. Nombre de chromosomes modernes = X x2n + Y - Z.

GR pour Génome de Référence indique que le riz (*Oryza sativa*) et la vigne (*Vitis vinifera*) ont été considérés comme génomes de référence pour les analyses de paléogénomique respectivement au sein des monocotylédones et des dicotylédones.

* Gènes cartographiés et en absence de séquence complète du génome.

**Nd : non déterminé, due au nombre limité de gènes cartographiés et l'absence de séquence du génome.

Abréviations : CHR = Chromosomes - DTG = Duplication Totale du Génome - ET = Eléments Transposables - MPB = Millions de paires de bases - NBR = nombre.

Le développement de nouvelles méthodes de comparaison des génomes a été nécessaire pour identifier les régions génomiques communes (c'est-à-dire porteuses de gènes conservés) et celles spécifiques des espèces étudiées. La Figure 1 illustre cette stratégie composée de 3 étapes et appliquée aux Graminées (riz, brome, blé, sorgho). La première phase a consisté à comparer les génomes modernes ayant abouti à l'identification de 12 blocs de conservation (Synt1 à Synt12) et 7 blocs de duplication (Dup1 à Dup7). La deuxième phase est réalisée par l'intégration des régions conservées et dupliquées ayant mis en évidence 5 à 7 blocs (ou chromosomes) fondateurs. Ces structures ancestrales alternatives sont toutes plausibles suivant le type de réarrangement considéré (fusion ou fission des chromosomes ancestraux) permettant d'expliquer la transition entre les ancêtres pré-duplication (définis par le nombre de blocs dupliqués, *i.e.* Dup) et post-duplication (défini par le nombre de blocs conservés, *i.e.* Synt). Enfin, la dernière phase a consisté pour chaque chromosome ancestral à reconstruire l'ordre des gènes en se basant sur le contenu/ordre en gènes de ce même chromosome au sein des espèces modernes, riz, brome, blé et sorgho.

Cette méthode de paléogénomique a permis d'établir que chez les monocotylédones, essentiellement les céréales dans notre analyse, 20.270 gènes sont conservés définissant 80 blocs et couvrant 91% des génomes considérés. De la même manière chez les dicotylédones, 27.591 gènes conservés définissant 860 blocs et couvrant 72% des génomes considérés ont été mis en évidence (*cf* Tableau). Cette analyse établit désormais clairement qu'entre génomes de plantes, plus de 80% des gènes/génomes sont conservés lorsque des relations sont établies uniquement sur la base d'une conservation de séquence de gènes (c'est-à-dire de fonctions biologiques conservées), alors que moins de 20% sont conservés en temps qu'orthologues stricts (c'est-à-dire conservation de séquence mais aussi de l'orientation et de l'ordre des gènes) ; Salse *et al.* 2002, 2004, 2008a, 2009a-b, 2011, 2012, Abrouk *et al.* 2010, Murat *et al.* 2010.

En conclusion, tout en présentant des structures de génome très différentes en moyenne, 86% des 17 génomes de plantes étudiés sont porteurs de gènes communs et ceci malgré une grande hétérogénéité en termes de nombre et de taille de chromosomes. Les plantes présentent une conservation de synténie élevée.

LES DUPLICATIONS GÉNOMIQUES ONT JOUÉ UN RÔLE MAJEUR DANS L'ÉVOLUTION DES PLANTES.

Les études cytologiques ont suggéré, il y a déjà fort longtemps, que les diploïdes actuels étaient d'ancien polyplœïdes (McClintock 1930). Le doublement du contenu chromosomique joue un rôle majeur dans l'évolution des plantes, car il constitue un mécanisme important de diversification et de variabilité des gènes, permettant notamment l'adaptation des plantes à leur environnement. La majorité des plantes, y compris celles cultivées à l'heure actuelle, ont connu des événements de duplication de leur génome relativement récemment (c'est-à-dire quelques millions d'années), appelés néo-polyplœïdie, ou conservent des "vestiges" d'événements de duplication plus anciens (c'est-à-dire plusieurs dizaines de millions d'années), appelés paléo-polyplœïdie.

L'utilisation des nouvelles méthodes de comparaison des séquences génomiques, mentionnées dans la section précédente, nous a permis de faire l'inventaire précis des duplications chromosomiques chez les plantes. Au total, 69 duplications chromosomiques (couvrant 81% des

génomés concernés) ont été caractérisées individuellement au sein des six génomes de monocotylédones étudiés et 396 blocs couvrant 57% des 11 génomes de dicotylédones étudiés (voir Tableau). L'identification expertisée et indépendante des relations de synténie et de duplication au sein des 17 génomes étudiés permet d'envisager la caractérisation d'événements de duplications ancestrales, par essence communs aux génomes étudiés, mais aussi les événements spécifiques de chacun de ceux-ci. L'intégration des duplications couvrant en moyenne 69% des génomes étudiés et des relations de synténie couvrant en moyenne 86% de ces mêmes génomes, a permis pour la première fois la caractérisation de 7 paléoduplications présentes au sein des 17 génomes étudiés et identifiées en position d'orthologie (c'est-à-dire conservées) entre ceux-ci (Figure 1), Salse *et al.* 2002, 2004, 2008a, 2009a-b, 2011, 2012, Abrouk *et al.* 2010, Murat *et al.* 2010.

En conclusion, nos études ont permis de mettre en évidence qu'en moyenne 69% de chaque génome des plantes étudiées est dupliqué. Ces gènes dupliqués sont issus de 7 blocs de duplication communs provenant d'un événement de polyploidie ancestral datant de plus de 60 millions d'années.

TOUTES LES PLANTES SONT ISSUES D'UN ANCÊTRE À 5 OU 7 CHROMOSOMES

Ces données confirment les données déjà publiées attestant de l'existence de duplications dans les génomes d'espèces végétales actuellement diploïdes (Bolot *et al.* 2009), mais surtout délivrent un inventaire exhaustif et précis du nombre ainsi que de la nature de ces duplications en discriminant 7 blocs de duplications ancestraux et de nombreuses duplications spécifiques de chacun des 17 génomes de plantes étudiés. Ces 7 paléoduplications permettent d'identifier ainsi 5 ou 7 groupes de chromosomes ancestraux qui ne partagent aucune paléoduplication entre eux. Les monocotylédones dérivent d'un ancêtre constitué de 5 à 7 chromosomes et les dicotylédones sont issues d'un ancêtre à 7 chromosomes (Figure 2). Nous avons montré que sept duplications sont communes à ces différentes espèces et la datation de ces duplications ancestrales suggère qu'elles ont pu avoir lieu au moment à la dernière phase connue d'extinction majeure des espèces vivantes lors de la transition tertiaire/crétacé, il y a 65 millions d'années (Murat *et al.* 2010, Salse 2012) (Figure 2). Ces duplications auraient offert une plasticité de génome ayant permis aux plantes de pouvoir s'adapter contrairement à certaines espèces de mammifères n'ayant pas pu évoluer par ce processus de polyploidie.

A partir de la représentation précédente, il est possible d'appréhender finement la structure (gènes fondateurs) et la fonction (fonctions biologiques fondatrices) des génomes ancêtres des monocotylédones constitués de 5 ou 7 protochromosomes et des dicotylédones constitués de 7 protochromosomes. Pour ce faire nous avons identifié les gènes potentiellement présents sur les chromosomes ancestraux reconstitués. En effet, à partir de la comparaison des chromosomes ou régions chromosomiques des 17 génomes, le génome ancestral apparaît constitué de 5 à 7 protochromosomes et est porteur de près de 10.000 protogènes représentant une taille physique minimale de moins de 50 millions de bases (Murat *et al.* 2010, Salse 2012). Cette taille physique du génome ancestral est un cumul de l'espace génique (*i.e.* somme des gènes) ancestral et ne prend aucunement en considération la possible présence de séquences non-géniques répétées (c'est-à-dire de séquences intergéniques) dans ce génome. Nous ne

pouvons toutefois pas exclure la présence de ce type d'élément, mais il est impossible d'en estimer l'impact et la teneur par la comparaison des génomes de plantes qui sont au centre de notre étude. En effet, le taux, la rapidité ainsi que la dynamique de transposition ou de mobilité de ces éléments intergéniques les rendent indétectables en position d'orthologie (donc potentiellement ancestraux) entre génomes ayant divergé d'un ancêtre commun il y a 200 millions d'années (Salse 2012).

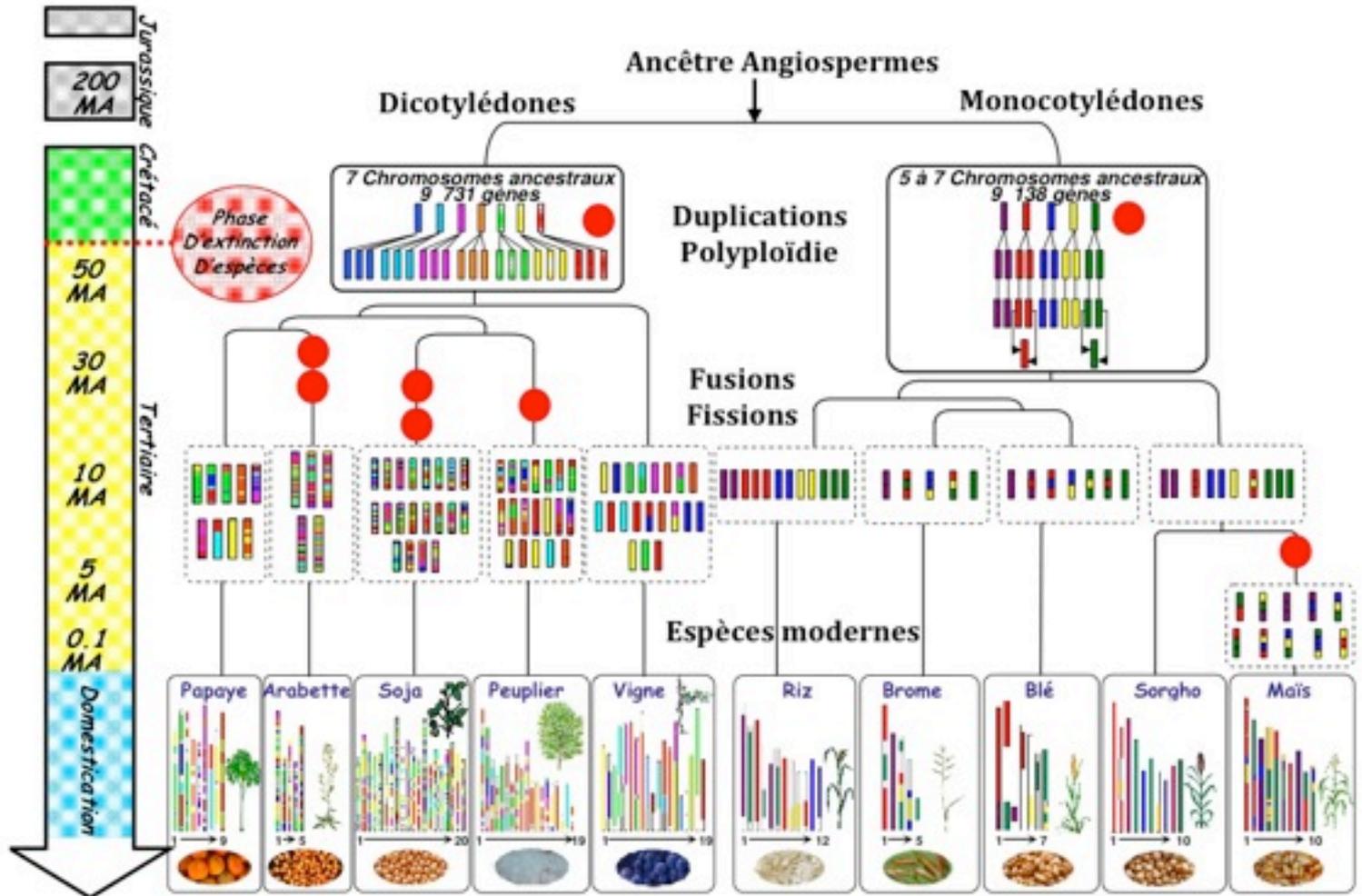


Figure 2 : Modèle évolutif des génomes de plantes (adapté de Salse 2012) - La figure représente le scénario évolutif des plantes modernes reconstruit à partir de la modélisation des ancêtres (en haut de la figure) de dicotylédones (7 et 21 chromosomes, à gauche) et de monocotylédones (5 à 7 et 12 chromosomes, à droite). Les génomes d'espèces modernes (en bas de la figure) ont été élaborés à partir des précédents ancêtres fondateurs par des événements distincts de fusion et fission chromosomiques (FC) au cours du temps (échelle à gauche de la figure), de telle sorte qu'ils sont matérialisés comme des mosaïques de couleurs représentant les chromosomes ancestraux. Enfin, les événements récurrents de duplication totale du génome (DTG) sont matérialisés par un point rouge sur la figure.

En conclusion, l'inventaire des conservations et duplications chez les plantes a permis de modéliser des ancêtres fondateurs constitués de 5/7 chromosomes porteurs de 10.000 gènes et datant de plus de 65 millions d'années. Ces ancêtres ont subi des événements de duplication il y a 65 millions d'années (transition Crétacé/Tertiaire) qui sont partagés par la totalité des plantes modernes.

LA FUSION DE CHROMOSOMES ANCESTRAUX EST À L'ORIGINE DES DIFFÉRENTES ESPÈCES VÉGÉTALES MODERNES.

Les duplications de chromosomes ont été suivies de fusions entre ceux-ci et de réarrangements. Les génomes de plantes modernes apparaissent ainsi comme une mosaïque des chromosomes ancestraux (Figure 2, bas). Nos travaux ont permis de démontrer que l'apparition de nouvelles espèces de plantes était liée à des fusions/fissions de chromosomes et que chaque espèce pouvait être associée à un nombre de fusions/fissions chromosomiques distinct. Ces connaissances permettent d'identifier avec précision les régions qui portent des gènes ayant une origine commune, au sein des génomes de riz, maïs, sorgho, brome, vigne, arabette, peuplier, soja, papayer, fraisier, cacaoyer, lotier, pommier. Elles permettent de rechercher de façon plus efficace les gènes impliqués dans les caractères d'intérêt agronomique dans les espèces dont le génome n'est pas encore séquencé, et d'en étudier la fonction permettant ainsi une meilleure exploitation en amélioration variétale comme il sera discuté dans la suite du document.

A partir de l'identification exhaustive des relations de synténie et de duplications ayant abouti à l'identification de 7 paléoduplications permettant de modéliser un ancêtre à 5/7 protochromosomes, il devient possible de reconstruire une histoire évolutive des plantes à partir de ces ancêtres communs. Ce scénario évolutif est structuré en trois étapes et illustré dans la Figure 2. La première étape est centrée sur les génomes ancestraux reconstruits de 5 chromosomes pour les monocotylédones (un scénario alternatif à 7 chromosomes a également été proposé, cf Abrouk *et al.* 2010, Salse 2012) et de 7 chromosomes pour les dicotylédones. Ces génomes fondateurs ont subi des duplications totales de leurs génomes (illustrées par des points rouges sur la Figure 2) pour aboutir à des ancêtres intermédiaires à 12 (paléo-tétraploïde) et 21 (paléo-hexaploïde) chromosomes respectivement pour les monocotylédones et dicotylédones. Tous les génomes de plantes à fleur sont issus de ces deux ancêtres intermédiaires ayant subi des fusions/fissions chromosomiques distinctes pour donner naissance aux espèces modernes. Ainsi, les génomes modernes peuvent être matérialisés par une équation génomique faisant intervenir le nombre de chromosomes ancêtres (X), le nombre de duplications (éventuellement répétées n fois, $x2n$) et le nombre de fissions (+Y) et fusions (-Z) subies au cours de l'évolution ; *i.e.* Nombre de chromosomes modernes = $X x2n - Y + Z$. À titre d'exemple pour les dicotylédones, le génome de la vigne peut être matérialisé par l'équation $21+2-4$ montrant que les 19 chromosomes modernes de la vigne sont issus de l'ancêtre à 21 chromosomes, qui n'a pas subi de duplication mais 2 fissions et 4 fusions chromosomiques. De la même manière pour les monocotylédones, le blé est décrit par l'équation $12+5-10$, montrant l'origine du blé à partir de l'ancêtre à 12 chromosomes ayant subi 5 fissions et 10 fusions chromosomiques qui lui sont spécifiques (Murat *et al.* 2010, Salse 2012). Le tableau précédemment mentionné fournit les équations chromosomiques des 11 dicotylédones et 6 monocotylédones étudiées.

Nous avons comparé, à l'échelle des chromosomes, l'ancêtre des monocotylédones (5 protochromosomes, 9.138 protogènes) et celui des dicotylédones (7 protochromosomes, 9.731 protogènes). Les résultats ne montrent aucune relation de conservation globale chromosome à chromosome (voire bras chromosomique et même fragments de chromosomes). Par conséquent, la colinéarité ou synténie a été érodée sur une échelle de temps de 150 à 300 MA, de telle sorte qu'il n'est possible à l'heure actuelle d'identifier qu'une centaine de régions génomiques très faiblement conservées ne couvrant chacune qu'un nombre très réduit de gènes (de l'ordre de la dizaine) comme en attestent les travaux qui ont été menés depuis 2002 en génomique comparée

entre le riz et l'arabette (*Arabidopsis thaliana*) spécifiquement (Salse *et al.* 2002, 2004). Nous avons, pour la première fois, pu modéliser l'histoire évolutive des monocotylédones et des dicotylédones à partir d'ancêtres respectivement paléotétraploïdes (structure génomique minimale de 5 à 7 chromosomes) et paléohexaploïdes (structure génomique minimale de 7 chromosomes) porteurs de moins de 10.000 gènes pour une taille physique minimale inférieure à 50 millions de bases. Nous avons également reconstruit précisément la nature et le nombre de fusions/fissions chromosomiques (FC) permettant d'expliquer le nombre différent de chromosomes observé au sein des génomes actuels de plantes (Figure 2), (Salse 2012).

En conclusion, la comparaison des génomes modernes a permis de modéliser un scénario évolutif dans lequel les espèces modernes proviennent d'ancêtres à nombre réduit de chromosomes (de 5 à 7 protochromosomes) ayant subi des événements ancestraux de duplication ou polyploïdie (de 12 à 21 protochromosomes) suivis de fusions et fissions chromosomiques spécifiques d'espèces (de 5 à 21 chromosomes modernes).

LES GENES DUPLIQUÉS SONT DÉRÉGULÉS POUR ÉLIMINER LEUR REDONDANCE.

L'accès à la structure des génomes ancestraux (5/7 protochromosomes et 10.000 gènes) nous permet de comparer les fonctions biologiques et identifier celles qui ont été conservées voire amplifiées préférentiellement au cours de l'évolution. Nous avons ainsi démontré que 4 classes sont statistiquement sous-représentées dans l'ancêtre par rapport au génome de plantes modernes. Ces 4 classes correspondent aux facteurs de transcription, ou toute fonction intervenant dans le processus de la transcription/régulation génique. Ce résultat suggère que ces fonctions biologiques ont été préférentiellement affectées par les mécanismes évolutifs dont notamment les duplications, au cours de l'évolution des génomes des plantes (Murat *et al.* 2010). En d'autres termes, les copies surnuméraires de ces fonctions biologiques acquises par duplication totale des génomes ont été conservées dans leur structure suggérant leur rôle dans l'adaptation favorable des plantes à leur environnement. Il y aurait donc un avantage sélectif majeur à la conservation de copies surnuméraires de telles fonctions biologiques (Murat *et al.* 2010).

Mis à part ces fonctions biologiques particulières, la majorité des gènes issus de duplications est perdue au cours de l'évolution pour revenir à un état diploïde initial. Ceci est appelé le processus de diploïdisation (Figure 3). Nous avons pu établir que, sur une échelle de 50 MA suivant un événement de duplication totale du génome, celui-ci retournait à un état diploïde complet par la perte de la redondance des gènes dupliqués. Cette perte de redondance est obtenue par trois phénomènes : la délétion ou la modification d'expression ou la modification de fonction des gènes dupliqués (Figure 3), (Throude *et al.* 2009, Pont *et al.* 2011). Ce processus de dérégulation des copies surnuméraires de gènes issus de duplications constitue une plasticité importante des génomes de plantes, de telle sorte que l'ajustement ou la spécialisation d'un jeu complet de gènes est possible après duplication pour faire face aux conditions voire aux contraintes environnementales changeantes.

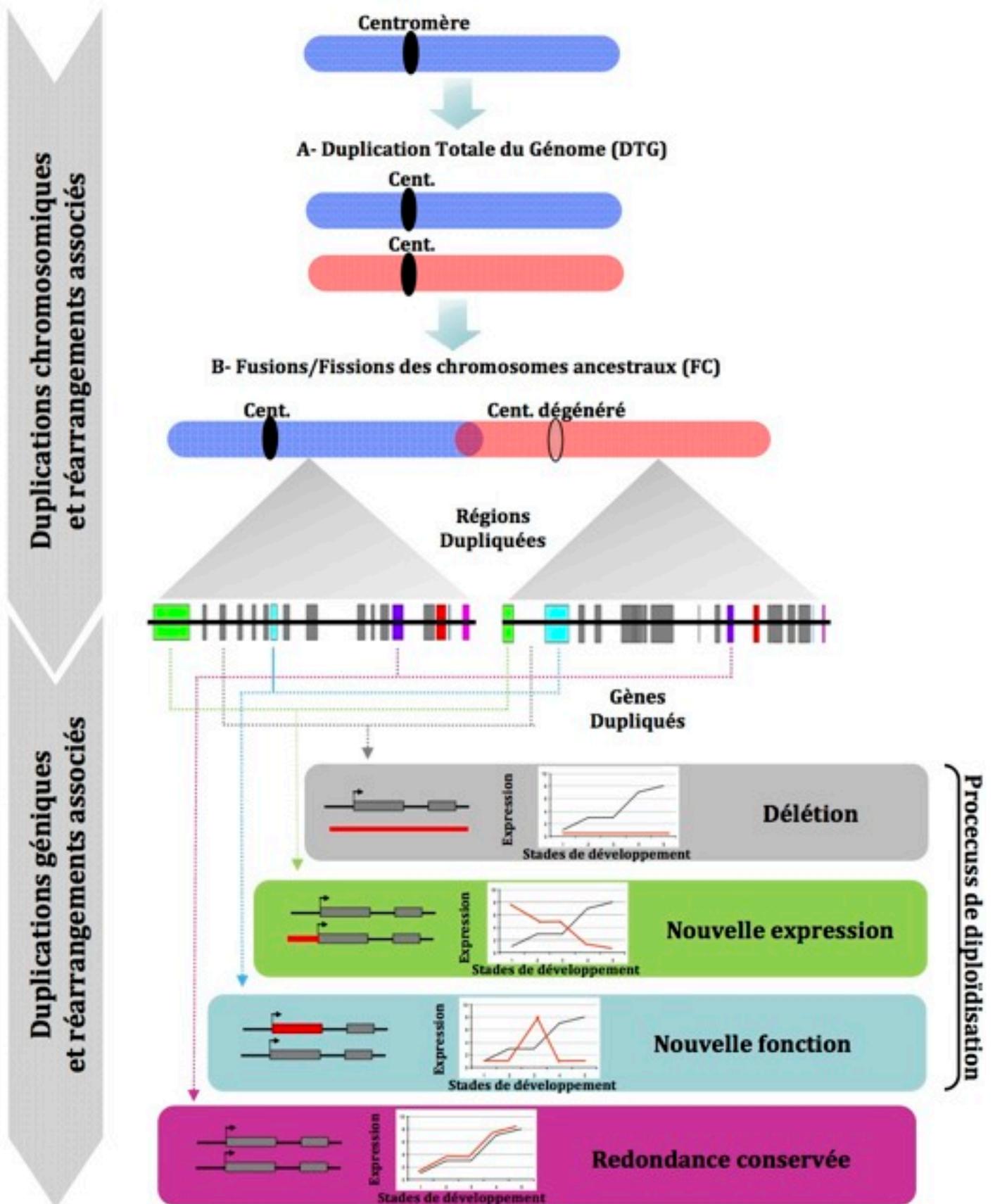


Figure 3 : Réarrangements génomiques et géniques (adapté de Salse 2012) - La figure illustre l'impact des duplications génomiques (en haut) et géniques (en bas) sur respectivement les fusions/fissions des chromosomes ancestraux ainsi que sur la dérégulation des gènes dupliqués. La duplication d'un chromosome ancestral induit un retour à l'état génomique diploïde par des processus de fusion et fission chromosomique. Le génome porteur de copies surnuméraires de gènes subit au niveau génique une élimination de la redondance fonctionnelle de ceux-ci par des délétions ou des modifications de l'expression ou de la fonction des copies dupliquées. Ces réarrangements au niveau génomique et génique, mis en place après un événement de duplication, participent à la diploïdisation des génomes polyploïdes.

En conclusion, la polyploidie des génomes constitue un choc génomique dû à la présence de copies surnuméraires de gènes. Cette redondance est éliminée au cours du temps à la fois au niveau génomique (par des processus de fusion et fission chromosomique) ainsi qu'au niveau génique (par des délétions ou des modifications de l'expression ou de la fonction des copies dupliquées). Cette spécialisation des gènes dupliqués offre une plasticité adaptative aux génomes de plantes.

LE TRANSFERT DE CONNAISSANCES ENTRE ESPÈCES EST AMÉLIORÉ.

La connaissance de la structure de l'ancêtre commun des génomes de plantes et des relations entre certaines régions chromosomiques des espèces cultivées nous permet d'identifier avec précision les régions au sein, par exemple, des génomes du riz, du blé, du sorgho et du maïs qui portent des gènes ayant une origine commune. Grâce à ce résultat, les informations obtenues sur la fonction biologique d'un gène chez une de ces espèces (dureté, teneur en protéines et teneur en fibres du grain par exemple) ont permis d'interpréter plus facilement sa fonction dans les autres espèces apparentées (Chantret *et al.* 2005 ; Gu *et al.* 2006 ; Salse *et al.* 2008b ; Quraishi *et al.* 2009, 2010, 2011). Ce concept est intitulé 'génomique translationnelle'. Notre modèle évolutif définit précisément les régions chromosomiques du riz, du blé, du sorgho et du maïs qui portent des gènes en commun. Il constitue un outil performant pour améliorer, par exemple chez le blé, des caractéristiques agronomiques aussi importantes que la qualité, le rendement ou la tolérance aux stress, ceci en transférant les connaissances acquises pour ces caractères chez les espèces modèles (c'est-à-dire séquencées).

À ce titre, les connaissances sur la colinéarité entre les génomes des céréales ainsi que sur les duplications chromosomiques du blé ont été mises en œuvre pour l'utilisation en routine de marqueurs COS (Conserved Orthologous Set), gènes conservés entre espèces et utilisés comme marqueurs moléculaires. Ainsi, sur la base de l'analyse de la synténie entre les génomes de brome-sorgho-maïs-riz et du blé, nous pouvons désormais utiliser comme marqueurs les 10.000 gènes ancestraux pour lesquels nous avons identifié un orthologue chez les céréales (Quraishi *et al.* 2009). Ces marqueurs moléculaires sont d'une extrême utilité en absence de séquence pour le génome du blé. Nous avons développé une interface web (www.clermont.inra.fr/umr1095/PlantSynteny) à partir des résultats de synténie et de duplications décrits dans la première partie de cet article pour, à partir d'une région d'intérêt agronomique chez le blé, disposer automatiquement de marqueurs moléculaires à des fins appliquées d'amélioration variétale.

En conclusion, les travaux de paléogénomique permettent une caractérisation précise des éléments génomiques communs aux espèces modernes (éléments par définition ancestraux) ou spécifiques de chacune de celles-ci. Ils permettent d'exploiter de la manière la plus efficace les connaissances (par exemple les fonctions des gènes) et outils (par exemple les marqueurs moléculaires) acquis sur les espèces modèles pour des espèces plus complexes dans leur physiologie, biologie et structure génomique.

L'ÉVOLUTION ET LES DUPLICATIONS ONT MODELÉS LES CARACTÈRES AGRONOMIQUES.

Il est clairement établi que les génomes de plantes, de manière générale, et des céréales plus particulièrement, ont évolué notamment par de multiples séries de duplications totales ou partielles de leur génome. Nous avons clairement établi, à l'instar des connaissances récentes acquises chez les mammifères, que ces duplications doublant par définition le contenu génique, ont influencé très largement la structure et la régulation des génomes par le processus de diploïdisation précédemment décrit. Ces modifications structurales et fonctionnelles seraient à l'origine d'une grande partie de la variabilité phénotypique de caractères dont l'amélioration, au travers notamment de la caractérisation de QTL ('*Quantitative Trait Locus*'), est recherchée en sélection variétale. Plusieurs études récentes ont rapporté précisément l'impact des événements évolutifs des plantes en général et de duplication de génome en particulier sur l'élaboration des caractères agronomiques travaillés en amélioration variétale. Plusieurs études ont fait la preuve du rôle majeur des duplications génomiques présentées en détail dans cet article sur des caractères agronomiques majeurs (tels que la réponse de floraison, l'étiollement, la pigmentation, la photosynthèse, les caractères de qualité, de rendement, d'incompatibilité) et ceci pour de multiples espèces végétales (telles que le tournesol, le coton, l'orge, les Brassicaceae, le soja, le riz, le blé, l'arabette et le maïs).

Récemment, la modélisation du génome ancêtre chez les céréales ainsi que la conservation chez le blé, le brome, le riz, le sorgho et le maïs d'une région associée au caractère d'assimilation de l'azote a permis de caractériser une glutamate synthase (GoGAT) comme gène candidat pour améliorer ce caractère (Figure 4). Nous avons ainsi montré chez le blé (variété Arche), à l'instar de ce qui a été établi chez le riz, que l'expression de ce gène GoGAT est positivement corrélée à l'absorption de l'azote au cours de développement, de la floraison au remplissage du grain. Ce gène constitue désormais un candidat majeur chez les céréales pour l'optimisation de l'absorption de l'azote au cours du remplissage du grain, facteur important de la taille finale du grain, c'est-à-dire du rendement (Quraishi *et al.* 2011).

En conclusion, les duplications génomiques anciennes ou récentes ont joué un rôle majeur dans l'élaboration des caractères agronomiques. Les données de paléogénomique permettent de dégager des hypothèses voire des gènes candidats robustes pour l'amélioration des ces caractères chez les espèces étudiées en sélection variétale.

CONCLUSION ET PERSPECTIVES

L'accès récent à la séquence de nombreux génomes de plantes a permis de reconstruire leur histoire évolutive sur les dernières 200 millions d'années et de modéliser leurs ancêtres fondateurs disparus. L'histoire évolutive ainsi reconstruite à partir d'ancêtres modélisés et de nombreux réarrangements identifiés prouve que les génomes de plantes sont plastiques dans leur structure et organisation. Cette extrême plasticité notamment issue des duplications génomiques nombreuses a constitué, et constitue toujours à l'heure actuelle pour les plantes, un atout majeur, permettant aux plantes de s'adapter en permanence aux conditions environnementales changeantes. Cette plasticité pourra ainsi être mise en œuvre dans le cadre du réchauffement climatique global attendu dans les prochaines décennies. Ces connaissances

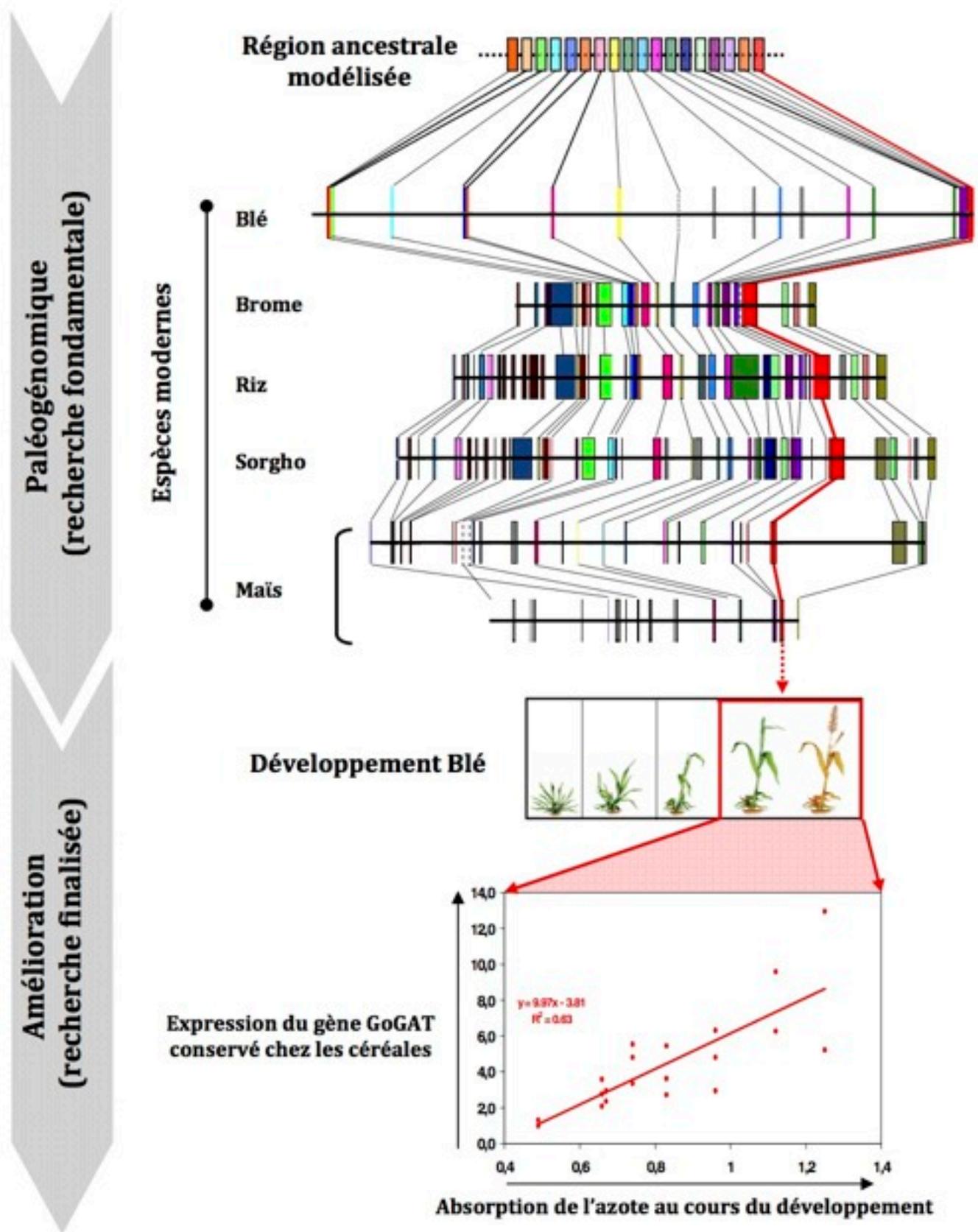


Figure 4 : La génomique translationnelle chez les plantes (adapté de Quraishi *et al.* 2011) - La figure illustre un exemple de génomique translationnelle chez les céréales pour le caractère d'assimilation de l'azote. La modélisation du génome ancêtre chez les céréales (haut) ainsi que la conservation d'une région associée au caractère d'assimilation de l'azote chez le blé, le brome, le riz, le sorgho et le maïs, ont permis de caractériser une enzyme conservée, la glutamate synthase (GoGAT) (en rouge), chez ces espèces. Chez le blé (variété Arche) l'expression de GoGAT (bas) est positivement corrélée (droite de régression en rouge avec en abscisses les phases de développement et en ordonnées la valeur d'expression du gène) avec l'absorption de l'azote au cours de développement, de la floraison jusqu'au remplissage du grain.

fondamentales permettent, d'un point de vue appliqué, de naviguer entre les génomes afin de pouvoir transférer les connaissances d'une espèce à l'autre. Ce transfert de connaissances et d'outils (les marqueurs moléculaires par exemple) constitue un atout majeur pour les programmes actuels d'amélioration variétale, notamment la sélection assistée par marqueurs moléculaires ou la transgénèse.

REMERCIEMENTS

Les recherches décrites dans cet article ont été financées par deux programmes de l'Agence Nationale pour la Recherche (Programme ANRjc-PaleoCereal, ref: ANR-09-JCJC-0058-01 et programme ANR Blanc-PAGE, ref: ANR-2011-BSV6-00801). L'auteur souhaite également remercier Caroline Pont et Florent Murat de l'équipe 'Paléogénomique des Plantes pour l'Amélioration Variétale' (INRA-UMR-1095, Clermont-Ferrand) pour leur contribution dans la préparation des figures.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Abrouk M, Murat F, Pont C, Messing J, Jackson S, Faraut T, Tannier E, Plomion C, Cooke R, Feuillet C, Salse J (2010) Paleogenomics of Plants: Modern Species Synteny-Based Modelling of Extinct Ancestors. *TRENDS IN PLANT SCIENCE*. 15(9): 479-87.
- Bolot S, Abrouk M, Masood-Quraishi U, Stein N, Messing J, Feuillet C, Salse J (2009) The 'inner circle' of the cereal genomes. *CURR OPIN PLANT BIOL*. 12(2): 119-25.
- Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P, Gautier MF, Cattolico L, Beckert M, Aubourg S, Weissenbach J, Caboche M, Bernard M, Leroy P, Chalhou B (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *PLANT CELL*. 17: 1033-45.
- Green RE, Malaspina AS, Krause J, Briggs AW, Johnson PL, Uhler C, Meyer M, Good JM, Maricic T, Stenzel U, Prüfer K, Siebauer M, Burbano HA, Ronan M, Rothberg JM, Egholm M, Rudan P, Brajković D, Kućan Z, Gusić I, Wikström M, Laakkonen L, Kelso J, Slatkin M, Pääbo S (2008) A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *CELL*. 134(3): 416-426.
- Gu Y, Salse J, Coleman-Derr D, Dupin A, Crossman C, Lazo GD, Ravel C, Charmet G, Anderson OD, Chalhou B (2006) Types and Rates of Sequence Evolution at the Glu-1 locus in Hexaploid Wheat and Its Ancestor Genomes. *GENETICS*. 174(3): 1493-504.
- Murat F, Xu JH, Tannier E, Abrouk M, Guilhot N, Pont C, Messing J, Salse J (2010) Ancestral Grass Karyotype Reconstruction Unravels New Mechanisms of Genome Shuffling as a Source of Plant Evolution. *GENOME RESEARCH*. 20(11): 1545-1557.
- Murat F, Van de Peer Y, Salse J (2012) Decoding plant and animal genome plasticity from differential paleo-evolutionary patterns and processes. *GENOME BIOLOGY & EVOLUTION*. 4(9): 917-28
- Nakatani Y, Takeda H, Kohara Y, Morishita S. (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *GENOME RESEARCH*. 17: 1254-1265.
- McClintock (1930) *A Cytological Demonstration of the Location of an Interchange between Two Non-homologous Chromosomes of Zea mays*. *PROC NATL ACAD SCI U S A*. 16(12): 791-6
- Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, Rapp M, Miller W, Schuster SC (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *SCIENCE*. 311: 392-394
- Pont C, Murat F, Confolent C, Balzergue S, Salse J. (2011) RNA-seq in grain unveils fate of neo- and paleopolyploidization events in bread wheat (*Triticum aestivum* L.). *GENOME BIOLOGY*. 12(12): R119.

- Quraishi UM, Abrouk M, Bolot S, Pont C, Throude M, Guilhot N, Confolent C, Bortolini F, Praud S, Murigneux M, Charmet G, Salse J (2009) Genomics in cereals: From genome-wide conserved orthologous set (COS) sequences to candidate genes for trait dissection. *FUNCTIONAL & INTEGRATIVE GENOMICS*. 9(4): 473-84.
- Quraishi UM, Murat F, Abrouk M, Pont C, Confolent C, Oury FX, Ward J, Boros D, Gebruers K, Courtin C, Bedos Z, Saulnier L, Guillon F, Balzergue S, Shewry P, Feuillet C, Charmet G, Salse J (2010) Combined meta-genomics analyses unravel candidate genes for the grain dietary fiber content in bread wheat (*Triticum aestivum* L.). *FUNCTIONAL & INTEGRATIVE GENOMICS*. 11(1): 71-83.
- Quraishi UM, Abrouk M, Murat F, Pont C, Foucrier S, Desmaizieres G, Confolent C, Rivière N, Charmet G, Paux E, Murigneux A, Guerreiro L, Lafarge S, Le Gouis J, Feuillet C and Salse J (2011) Cross-Genome Map Based Dissection of a Nitrogen Use Efficiency Ortho-metaQTL in Bread Wheat Unravels Concerted Cereal Genome Evolution. *PLANT JOURNAL*. 65(5): 745-56.
- Salse J, Piegu B, Cooke R, Delseny M (2002) Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. *NUCLEIC ACIDS RES*. 30(11): 2316-2328.
- Salse J, Piegu B, Cooke R, Delseny M (2004) New in silico insight into the synteny between rice (*Oryza sativa* L.) and maize (*Zea mays* L.) highlights reshuffling and identifies new duplications in the rice genome. *PLANT JOURNAL*. 38(3): 396-409.
- Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Masood U, Calcagno T, Cooke R, Delseny M, Feuillet C (2008a) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *PLANT CELL*. 20(1): 11-24.
- Salse J, Chagué V, Bolot S, Magdelenat G, Huneau C, Pont C, Belcram H, Couloux A, Gardais S, Evrard A, Segurens B, Charles M, Ravel C, Samain S, Charmet G, Boudet N, Chalhoub B (2008b) 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(1): 555.
- Salse J, Abrouk M, Murat F, Masood Quraishi U, Feuillet C (2009a) Improved standards and new comparative genomics tools provide new insights into grasses paleogenomics. *BRIEFINGS IN BIOINF*. 10(6): 619-30.
- Salse J, Abrouk M, Bolot S, Guilhot N, Courcelle E, Faraut T, Waugh R, Close TJ, Messing J, Feuillet C (2009b) Reconstruction of monocotyledonous proto-chromosomes reveals faster evolution in plants than in animals. *PROC NATL ACAD SCI U S A*. 106: 14908-14913.
- Salse J, Feuillet C (2011) Palaeogenomics in cereals: modeling of ancestors for modern species improvement. *COMPTE RENDUS DE L'ACADÉMIE DES SCIENCES*. 334: 205-211.
- Salse J (2012) *In silico* archeogenomics unveils modern plant genome organization, regulation and evolution. *CURRENT OPINION IN PLANT BIOLOGY*. 15(2): 122-30.
- Throude M, Bolot S, Bosio M, Pont C, Sarda X, Masood Quraishi U, Bourgis F, Lessard P, Rogowsky P, Ghesquiere A, Murigneux A, Charmet G, Perez P, Salse J (2009) Structure and Expression Analysis of Rice Paleo-Duplications. *NUCLEIC ACIDS RES*. 37(4): 1248-59.