



الجمهورية الجزائرية الديمقراطية الشعبية  
RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي و البحث العلمي  
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE



Université des Frères Mentouri Constantine 1  
Faculté des Sciences de la Nature et de la Vie

جامعة الاخوة منتوري قسنطينة 1  
كلية علوم الطبيعة و الحياة

Département : Biologie et Ecologie Végétale

قسم : البيولوجيا و علم البيئة النباتية

Mémoire présenté en vue de l'obtention du Diplôme de Master

Domaine : Sciences de la Nature et de la Vie

Filière : Biotechnologies

Spécialité : *Biotechnologie et Génomique Végétale*

Intitulé :

---

# Séquençage : Intérêt et impact en amélioration des plantes -cas du blé-

---

Présenté et soutenu par : *KAREK Soulaf*

Le : 17/07/2019

Jury d'évaluation :

Président(e) du jury : *YKHLF NADIA* (Professeur - UFM Constantine).

Encadrant : *BOUSBA RATIBA* (Docteur - UFM Constantine).

Examineur (rice) : *HAMLA CHOUROUK* (Docteur - UFM Constantine).

*Année universitaire  
2018 - 2019*

# Remerciements

*Je tiens à remercier et rendre grâce à Dieu le tout puissant de nous avoir donné le courage et la volonté de mener à bon terme ce modeste travail.*

*À travers ce mémoire de fin d'étude, je rends hommage à toutes les personnes qui ont fait que l'initiation, la réalisation et la finalisation de cette étude soit possible.*

*Mes plus vifs remerciements à :*

*Mon promoteur Mme BOUSBA Ratiba . A pour avoir proposé ce thème et m'avoir formée tout au long de ce travail. Me lui disons merci encor pour sa totale disponibilité et sa modestie à mon égard.*

*C'est avec sincérité que j'exprime ma gratitude et mon profond respect à Melle Ykhlef. N, notre responsable pédagogique du master. Pr. Université des Frères Mentouri qui a accepté et bien voulu nous faire l'honneur de présider le jury.*

*À Mlle. HAMLA Chourouk d'avoir accepté d'examiner ce modeste travail et d'être membre de ce jury.*

*En fin, et profondes reconnaissances à tout ceux qui ont participé de près ou de loin à la réalisation de ce travail mais qui ne sont pas cités ici, nous les remercions tous chaleureusement.*

# *Dédicaces*

*À la personne qui est toujours avec moi, qui a sacrifiée jour et nuit pour mon éducation et mon bien être, à mon cher père.*

*À ma chère maman, qui est toujours pré de moi, m'encourage, me conseille, avec tous les moyens, aucun mot, ne peut exprimer ce que tu mérites pour tous les sacrifices*

*que tu n'as pas cessée de me donner depuis ma naissance, merci maman que dieu te garde et te protège.*

*Je tiens à dédier mes très chers frères et sœurs qui m'ont aidée beaucoup plus moralement*

*À tout membre de ma famille, chacun par son nom.*

*À toute mes amies et mes collègues.*

*À tous mes enseignants et surtout de Biotechnologie et Genomique Végétale de Constantine.*

*Enfin, à tous ceux qui me connaissent.*

## Résumé :

Le blé, une espèce polyploïde, à trois grands génomes A, B et D constitue un génome complexe. Ce travail présente une recherche bibliographique sur la mise au point de la stratégie originale de séquençage du blé et les résultats obtenus dans le cadre du projet de cartographie du génome de blé, en utilisant les marqueurs moléculaires, les QTL à l'aide de marqueurs SNP à haut débit, banque des mutants, et de banque des EST.

Le projet de séquençage du génome de blé qui a été commencé en 2005 par la création de *Consortium International de Séquençage du Génome du Blé*. Une séquence de référence annotée représentant le génome du blé panifiable hexaploïde 21 pseudomolécules ont été analysées pour identifier la distribution et le contexte génomique des éléments codants et non codants dans les sous-génomes A, B et D.

Avec une estimation de couverture de 94% du génome et contenant 107 891 modèles de gènes à haut niveau de confiance, cet assemblage a permis la découverte de tissus et de stades de développement réseaux de co-expression en fournissant un atlas des transcriptomes représentant les principales étapes du développement du blé.

Le dynamique des familles de gènes complexes impliqués dans l'adaptation environnementale et la qualité d'utilisation finale ont été révélés à sous-génome résolution et contextualisée connus locus agronomiques à un seul gène ou à caractère quantitatif.

Cette ressource communautaire établit la Fondation pour accélérer la recherche sur le blé et l'application de la compréhension du blé de la biologie et de génomique de la sélection assistée.

L'IWGSC a obtenu ce résultat en combinant toutes les ressources qu'il a générées au cours des treize dernières années en utilisant les méthodes classiques de construction de cartes physiques, ainsi que les technologies de séquençages d'ADN les plus récentes. Les séquences d'ADN ont été assemblées et ordonnées le long des 21 chromosomes en utilisant des algorithmes performants et les gènes ont été identifiés grâce à des programmes bio-informatiques dédiés.

**Mots clés :** blé, génome, séquençage, séquence, référence, chromosome,

## **Abstract**

Wheat, a polyploid species with three large A, B, and D genomes, is a complex genome. This work presents a literature search on the development of the original wheat sequencing strategy and the results obtained in the wheat genome mapping project, using molecular markers, QTL using high throughput SNP markers, mutant Bank, and EST bank.

The wheat genome sequencing project started in 2005 with the creation of an International Wheat Genome Sequencing Consortium . An annotated reference sequence representing the genome of hexaploid bread wheat 21 pseudomolecules was analyzed to identify the distribution and genomic context of coding and noncoding elements in subgenomes A, B, and D.

With an estimated coverage of 94% of the genome and containing 107,891 high-confidence gene models, this assembly led to the discovery of co-expression network tissues and developmental stages by providing an atlas of transcriptomas representing the major stages of Wheat Development.

The dynamics of complex gene families involved in environmental adaptation and end-use quality were revealed at known single-gene or quantitative agronomic loci.

This community resource establishes the foundation for accelerating wheat research and the application of wheat understanding to biology and genomics of marker-assisted selection.

The IWGSC has achieved this by combining all the resources it has generated over the past 13 years using traditional map-building methods, as well as the latest DNA sequencing technologies. DNA sequences were assembled and sequenced along the 21 chromosomes using efficient algorithms and genes were identified through dedicated bioinformatics programs.

**Key words:** wheat, genome, sequencing, sequence, reference, chromosome

# ABRÉVIATIONS

<b>GB</b>	Giga basses
<b>Inra</b>	Institut national de la recherche agronomique
<b>MB</b>	Méga basses
<b>QTL</b>	Quantitative Trait Locus.
<b>ABA</b>	Acide abscicssique
<b>OGM</b>	Organisme génétiquement modifiée
<b>RFLP</b>	Polymorphisme de longueur des fragments de restriction de l'ADN
<b>BAC</b>	Bacterial Artificial Chromosome
<b>PCR</b>	Polymerase Chain Reaction
<b>DMS</b>	diméthyle sulfate
<b>NGS</b>	Next Generation Sequencing
<b>ZMW</b>	Zero-Mode Waveguide
<b>IWGSC</b>	Consortium international de séquençage du génome du blé
<b>URGI</b>	l'Unité de recherche en Génomique-Info
<b>EST</b>	Expressed Sequence Tags

# *LISTE DES FIGURES*

<b>Figure 1 :</b>	Evolution de la production mondiale de blé.	<b>6</b>
<b>Figure 2 :</b>	Evolution de la surface mondiale et des rendements mondiaux de blé.	<b>7</b>
<b>Figure 3 :</b>	: Diagramme schématique des relations entre les génomes du blé avec l'histoire et la généalogie de la polyploïdisation	<b>9</b>
<b>Figure 4 :</b>	: Organisation du génome du blé hexaploïde.	<b>10</b>
<b>Figure 5 :</b>	Quelques gènes localisés sur les chromosomes de blé. <i>Gli</i> : gliadine, <i>Glu</i> : gluténine (2 composants du gluten), <i>Ha</i> : gène qui contrôle la dureté du grain, <i>Q</i> : gène qui contrôle la structure de l'épi et <i>Ph1</i> : gène qui limite les appariements homéologues.	<b>12</b>
<b>Figure 6 :</b>	Les trois génomes de blé (génome polyploïde)	<b>13</b>
<b>Figure 7 :</b>	Schéma représentatif de l'évolution du blé ( <i>Triticum et Aegilops</i> ).	<b>14</b>
<b>Figure 8 :</b>	Différentes étapes du séquençage.	<b>16</b>
<b>Figure 9 :</b>	Dénaturation de l'ADN à la chaleur.	<b>18</b>
<b>Figure 10 :</b>	Marquage de L'ADN.	<b>18</b>
<b>Figure 11 :</b>	Altération de l'ADN par l'hydralazine	<b>19</b>
<b>Figure 12 :</b>	Brins d'ADN de tailles différentes dans les quatre tubes.	<b>19</b>
<b>Figure 13 :</b>	Autoradiogramme après traitement chimique des fragments.	<b>20</b>
<b>Figure 14 :</b>	Technique de Maxam et Gilbert.	<b>20</b>
<b>Figure 15 :</b>	Amplification et dénaturation de l'ADN.	<b>21</b>
<b>Figure 16 :</b>	Ajout d'un ADN amorcé dans les quatre récipients.	<b>21</b>
<b>Figure 17 :</b>	Ajout des ddNTPs.	<b>22</b>
<b>Figure 18 :</b>	<b>A</b> - Electrophorèse en gel standard. <b>B</b> - Séquençage à l'aide de fluorophores.	<b>23</b>

<b>Figure 19 :</b>	Aperçu de la technologie de séquençage 454	<b>26</b>
<b>Figure 20 :</b>	Ajout des adaptateurs spécifiques aux extrémités d'ADN	<b>27</b>
<b>Figure 21 :</b>	Flow Cell.	<b>28</b>
<b>Figure 22 :</b>	Amplification en pont de la technologie Illumina	<b>29</b>
<b>Figure 23 :</b>	Aperçu de la technologie de séquençage Illumina.	<b>30</b>
<b>Figure 24 :</b>	Aperçu de la technologie de séquençage SOLiD	<b>31</b>
<b>Figure 25 :</b>	Alignement entre une partie du chromosome 6 de l'homme ( <i>Hsa6</i> ) et du chromosome 4 de la souris ( <i>Mmu4</i> ).	<b>39</b>
<b>Figure 26 :</b>	Une carte comparative consensus des graminées.	<b>40</b>
<b>Figure 27 :</b>	Segments dupliqués d' <i>Arabidopsis</i>	<b>43</b>
<b>Figure 28 :</b>	génomme du blé est cinq fois plus grand que le génome humaine .	<b>46</b>
<b>Figure 29 :</b>	Génomme du blé déchiffré, assemblé et commandé.	<b>49</b>
<b>Figure 30 :</b>	Atlas de transcription du blé	<b>53</b>
<b>Figure 31 :</b>	Distribution génomique de familles des gènes .	<b>55</b>



# ***LISTE DES TABLEAUX***

<b>Tableau 1 :</b>	Consommation de blé, en kg, par an	<b>8</b>
<b>Tableau 2 :</b>	Avantages et faiblesses du NGS par rapport aux méthodes de séquençage «classique »	<b>33</b>
<b>Tableau 3 :</b>	Statistiques d'assemblage d'IWGSC RefSeq v 1.0.	<b>50</b>
<b>Tableau 4 :</b>	Proportion relative des principaux éléments du génome du blé. Proportions de TE sont exprimés en pourcentage de séquences attribuées à chaque superfamille par rapport à la taille du génome.	<b>51</b>
<b>Tableau 5 :</b>	<b>Groupes de gènes homéologues chez le blé. Les gènes homologues sont «Orthologues du sous-génome» et ont été déduits du rapprochement des espèces dans la famille de gènes respective.</b>	<b>52</b>

# ***SOMMAIRE***

<b>Introduction</b>	<b>1</b>
<b>Chapitre 1. Le blé</b>	<b>3</b>
<b>Partie 1. Origine et classification</b>	<b>3</b>
1.. Définition et origine	3
1.1. Origine génétique	3
1.2. Origine géographique	4
2. Classification botanique	4
2.1. Systématique	5
2.2. Caractère botanique	5
3. L'importance du blé	5
3.1 Le blé dans le contexte international	5
3.1.1. Production	6
3.2. Le blé dans le contexte national	7
3.2.1. Production	7
3.2.2. Consommation	8
<b>Partie 2. la génomique du blé</b>	<b>8</b>
1. Le génome du blé	9
2. Evolution de génome polyploïde du <i>Triticum</i>	11
1.2. Homéologie des Génomes	12
3. Cytogénétique du blé	12
3.1. Domestication du blé	13
<b>Chapitre 2. Séquençage chez les plantes</b>	<b>15</b>
<b>Partie 1 : Historique et généralités</b>	<b>15</b>
1 Historique	15
1.1 Séquençage de l'ADN	15

1.2Etapes du séquençage	15
2. Intérêt du séquençage	17
3. automatisation de séquençage	17
<b>Partie 2 : Séquençage de la première génération</b>	<b>17</b>
1. Technique de Maxam et Gilbert	21
2. Technique de Sanger	21
2.1. Comment fonctionne-t-elle?	23
<b>Partie 3 : Séquençage de nouvelle génération</b>	<b>23</b>
1. Développement et caractéristiques du séquençage de nouvelle génération (NGS)	23
2. Séquençage de seconde génération	24
2.1. Trois grandes technologies du séquençage NGS	25
2.1.1. 454 de Roche	25
2.1.2 Solexa/Illumina	26
2.1.3 SOLiD	32
3. Séquençage de troisième génération	32
3.1. Hélioscope	32
3.2. Technologies des nanopores :	32
3.3. Pacific Biosciences	33
3.4. Starlight	34
4. Avantages et inconvénients du NGS	34
5. application des NGS en agriculture	35
6.Amélioration variétale du blé et sa culture	35
<b>Chapitre 3 Apport des biotechnologies dans la connaissance des bases génétiques des caractères d'intérêt.</b>	<b>36</b>
1. Déterminer la fonction des gènes.	36

2. Identifier les régions du génome favorables pour un caractère donné	36
3. La sélection assistée par marqueurs	38
4. cartographie et comparaison entre espèces .	39
4.1. L’histoire des études comparatives et intérêt des cartes génétiques	39
4.2. Evolution de la structure des génomes	40
4.3. Connaissance sur l’évolution des génomes	41
5 .Amélioration de la cartographie génétique et physique	41
6. Clonage d’un gène par utilisation de l’orthologie	42
6.1 De nouvelles plantes modèles pour la synténie	44
6.2 Les défis de la génomique comparative	44
<b>Chapitre4 séquençage du génome du blé( défis et véritable exploit scientifique).</b>	<b>45</b>
Introduction	45
Méthodes	47
Résultats	48
1. Assemblage à l’échelle chromosomique du génome du blé	50
2. La composition du génome du blé	52
3. Atlas de transcription révèle trait associée de réseaux de Co régulation de gènes	56
4. Amélioration rapide des caractères en utilisant des marqueurs résolus physiquement et édition du génome.	56
Conclusion	58
<b>Conclusion général et perspectives</b>	
<b>Références bibliographiques</b>	
<b>Résumé</b>	
<b>Abstract</b>	



## Introduction

Le blé est une source d'alimentation essentielle pour un tiers de la population mondiale, dont 95% correspond à du blé hexaploïde (*T. aestivum*) est cultivé dans plus de 120 pays, sur environ 215 millions d'ha (17% des surfaces cultivées) pour une production de plus de 700 Mt/an. (Maryland, 2014).

En Algérie le blé occupe annuellement plus d'un million d'hectares et la production Algérienne reste instable et faible d'une année à une autre, 5 134 006 tonnes pour la saison 2011- 2012 elle ne couvre que 20 à 25% des besoins du pays. (Benbelkacem, 2012) Cependant, durant les dix dernières années, à sept reprises, la production annuelle n'a pas permis de satisfaire une demande en constante augmentation.

Le blé est une plante monocotylédone polyploïde qui possède six jeux de chromosomes par cellule associant trois sous-génomes, et dont la taille (17 000 Mb) représente cinq fois celle du génome humain et 40 fois le génome du riz. Le blé tendre (*Triticum aestivum*), son génome contient 21 chromosomes et plus de 120.000 gènes. (Berges, 2010).

Bien que nos connaissances sur cette espèce soient moins avancées que sur les plantes modèles (arabette et riz), principalement à cause de la taille de son génome, de nombreuses populations ont été cartographiées. Ainsi, le séquençage de génome de blé ou le décryptage de l'information génétique contenu dans les séquences génomiques de cette céréale implique l'établissement de la carte de référence, «*ITMI map*», qui comporte plus de 1500 marqueurs et la carte composite avec plus de 3700, la diversification et la variabilité génétique en sont donc accrues et le séquençage du génome en est plus complexe. (Anonyme, 2012).

Cette complexité en fait un bon modèle pour l'analyse des grands génomes polyploïdes, le décryptage de sa séquence a représenté un véritable défi technique et méthodologique, qui a longtemps retardé le développement des programmes de génomique et de séquençage en particulier.

Le séquençage est le processus utilisé pour la connaissance de la séquence d'ADN qui est le génome contenu dans chacune de nos cellules, il contient l'information nécessaire à l'expression des gènes, une information importante aux yeux des biologistes.

Plusieurs méthodes existent et s'améliorent toujours plus avec le temps, nous en sommes aujourd'hui à la troisième génération de ces outils de séquençage [4] .en effet l'évolution rapide des techniques de séquençages a permis de produire une séquence de référence du blé tendre basée sur une carte physique et dans la qualité et comparable à celle du génome du riz .

on peut prévoir que l'existence de sequence de ce genome complexe va révolutionner plusieurs domaines de recherche .

L'objectif de ce travail est de montrer l'intérêt du séquençage et de cette ressource scientifique comme outil précieux qui ouvre la voie a des innovations dans le domaine de l'agriculture et de la nutrition

# Chapitre 01. Le blé

## Partie 1. Origine et classification:

Le blé est une plante herbacée monocotylédone qui appartient au genre *Triticum* des graminées. C'est une céréale dont le grain est un fruit sec indéhiscant appelé caryopse constitué d'une graine et de téguments (Feillet., 2000).

### 1. Définition et origine

Le blé est une monocotylédone de la famille des *Poaceae* appartenant au genre *Triticum*. Cette plante annuelle produit un fruit sec indéhiscant, le caryopse. Le blé tendre (*Triticum aestivum*) et le blé dur (*Triticum durum*) sont les deux espèces les plus cultivées dans le monde. Le blé tendre est constitué de trois génomes possédant chacun 7 paires de chromosomes homéologues, soit 42 chromosomes au total. Il possède une structure génomique hexaploïde (**AA BB DD**) et le blé dur une structure tétraploïde (**AA BB**).

La culture de blé est très ancienne, on trouve ses traces dès le néolithique. Le blé a été cultivé 2700 ans avant notre ère en Chine. Les civilisations babyloniennes et égyptiennes se sont développées autour du blé (Moule, 1980).

### 1.1 Origine génétique

L'allopoloïdie a joué un rôle fondamental dans l'évolution des plantes en permettant l'apparition de types nouveaux qui n'ont souvent que de lointains rapports avec les espèces qui leurs ont donné naissance (Prevost, 1976).

De part leur constitution chromosomique, Boyeldieu (1980) ; Simon *et al.* (1989) distinguent l'existence de trois sous groupes de céréales :

**Le groupe diploïde** ( $2n = 14$  chromosomes) ou engrain.

- *Triticum beoticum* ;
- *Triticum monococcum*.

**Le groupe tétraploïdes** ( $2n = 28$  chromosomes) ou groupe de *Triticum dicoccum* (amidonnier) ; on distingue :

- *Triticum diccocoïdes* ou amidonnier sauvage ;
- *Triticum turgidum* ou blé poulard ;
- *Triticum polonicum* ou blé de Pologne ;
- *Triticum durum* ou blé dur.

**Le groupe hexaploïdes** ( $2n = 42$ ) ou groupe de *Triticum spelta* (épeautre) ; on distingue :

- *Triticum vulgare* ou blé tendre ;



· *Triticum compactum* ou blé hérisson.

Selon Prevost (1976), les blés à 28 chromosomes sont des allotétraploïdes possédant les génomes A et B.

Prevost (1976) et Grignac (1978), soulignent l'origine hybride des tétraploïdes dont le blé dur (*Triticum durum*), ceux-ci proviendraient du croisement suivi du doublement des chromosomes entre *Triticum monococcum*, apportant le génome A, et *Egilops spelta* apportant le génome B.

Une telle hybridation aurait donné naissance au *Triticum diccocoïdes* qui serait diversifié en *Triticum dicocum* et *Triticum durum* (Moule, 1980).

## 1.2. Origine géographique

Vavilov cité par Erroux et Laumont (1961) situe l'origine du blé dur en Abyssinie, ce dernier considérait trois centres d'origine distincts pour les trois groupes d'espèces du genre *Triticum* :

- Le foyer Syrien et nord Palestinien pour le groupe diploïde ;
- Le foyer Abyssinien pour la diversification des blé tétraploïdes ;
- Le foyer Afghano-Indien pour la diversification des blés hexaploïdes.

L'Algérie se trouvant à proximité de ce centre primaire d'origine, la diversification et le polymorphisme considérable de l'espèce blé dur dans nos région ont invité Vavilov à considérer l'Afrique du nord comme centre secondaire d'origine du *Triticum durum*.

Grignac (1978), désigne, quant à lui, le Moyen-Orient comme l'origine géographique ou coexisteraient les espèces parentales. Selon le même auteur, c'est à partir de cette zone d'origine que l'espèce s'est différenciée vers trois différentes régions : le bassin occidental de la méditerranée, le sud de l'ex URSS et le proche orient, chaque centre de différenciation donne des caractères morphologiques et physiologiques particuliers.

## 2. Classification botanique

### 2.1. Systématique

Le blé est une plante autogame appartenant au groupe des céréales à paille qui sont caractérisées par des critères morphologiques particuliers.

D'après Dekhil (1998) qui a indiqué que Dahlgren et Clifford (1985) ont proposé la classification suivante :

EMB	Spermaphytes
S / EMB	Angiospermes
CL	Monocotylédones
O	Poales

S / O	Commeliniflorales
F	Graminaceae OU Poaceae
G	<i>Triticum sp</i>

## 2.2. Caractère botanique

Deux espèces de *Triticum*, dominant actuellement les cultures de blé, le « Blé tendre » *T. aestivum* L. Thell et le « Blé dur » *T. durum* (Bousnane., 2003). Mais un certain nombre d'autres espèces ont été cultivées autrefois, et subsistent dans les collections, ou vivent à l'état sauvage au Moyen-Orient (Messiaen., 1981). Ces espèces forment une série polyploïde où l'on rencontre :

**Le groupe diploïde (2n = 14 Chromosomes) :** Représenté par l'engrain *Triticum beoticum* (sauvage) et *Triticum monococcum* (cultivé), ce groupe possède le génome **A (AA)**, c'est une espèce à faible rendement, très anciennement cultivée et qui est à l'origine des blés actuels.

**Le groupe tétraploïde (2n = 28 Chromosomes) :** Le blé dur actuel; très riche en gluten et utilisé pour produire les semoules et les pâtes alimentaires, ce groupe possède les génomes **A et B/G** est cultivé dans les zones sèches et chaudes. Ils ont représenté par : (AABB) pour : *T. dicoccoides* (sauvage), *T. dicoccum* (cultivé), *T. durum* (cultivé- le blé dur) et (AAGG) pour : *T. timopheevi* (cultivé) (Messiaen., 1981).

**Le groupe hexaploïde (2n = 42 Chromosomes) :** Tous cultivé, il possède trois génomes **A, B et D** et est d'origine Afgano-indienne; il est utilisé pour faire la farine panifiable pour le pain. On trouve également dans ce groupe le *Triticum aestivum ssp spelta* ou épeautre, qui est une sous espèce du blé tendre d'origine iranienne. Ils ont représenté par : (AABBDD) pour *T. spelta*, *T. aestivum*, le blé tendre et (AAAAGG) pour *T. zhukouskyi* (Messiaen., 1981).

## 3. L'importance du blé

### 3.1 Le blé dans le contexte international

La production mondiale de blé dur en 2015/16 au cours de mois de février 2016 est en hausse de 1,7 million de tonnes par rapport au mois de janvier de la même année, atteignant 39,7 millions de tonnes, un bond de 15 % par rapport au résultat de l'année précédente.

La production mondiale de blé tendre en 2015/16 est de 731,8 millions de tonnes, soit une augmentation de 5% par rapport à la campagne 2014/15.

Quant à la consommation et aux échanges, ils ont augmenté respectivement en 2015/16 (soit 719,6 Mt et 152,3 Mt) (ONFAA, 2016).

Le classement de l'année 2015 des principaux premiers producteurs du blé indique que l'UE est toujours en première position. Et la Chine en deuxième position. Par contre les Etats unis se situent en quatrième position après l'Inde (FAO, 2015).

L'UE et le continent américain sont excédentaires en blé, ce qui leur confère un avantage économique et géopolitique indéniable. Au contraire, l'Asie et l'Afrique apparaissent déficitaires, ce qui renforce leur dépendance à l'égard des grands pays exportateurs.

Le marché mondial du blé est segmenté en différents groupes de pays qui ont diverses capacités de production et de consommation de blé, ce qui rend ce marché plus propice à la volatilité des prix. Seulement 19% de la production mondiale du blé est échangée et il s'agit d'un marché de surplus et d'excédent (Charvet, 2012).

### 3.1.1. Production

La production mondiale a fortement progressé depuis les années 1960 (environ 2,4% par an) (CNUCED., 2011). Elle a presque triplé en l'espace de 50 ans, passant ainsi de 222 millions de tonnes en 1961 à de 691.5 millions de tonnes en 2011. Ce très net décollage peut être attribué principalement à la progression des rendements mondiaux (Figure 1).

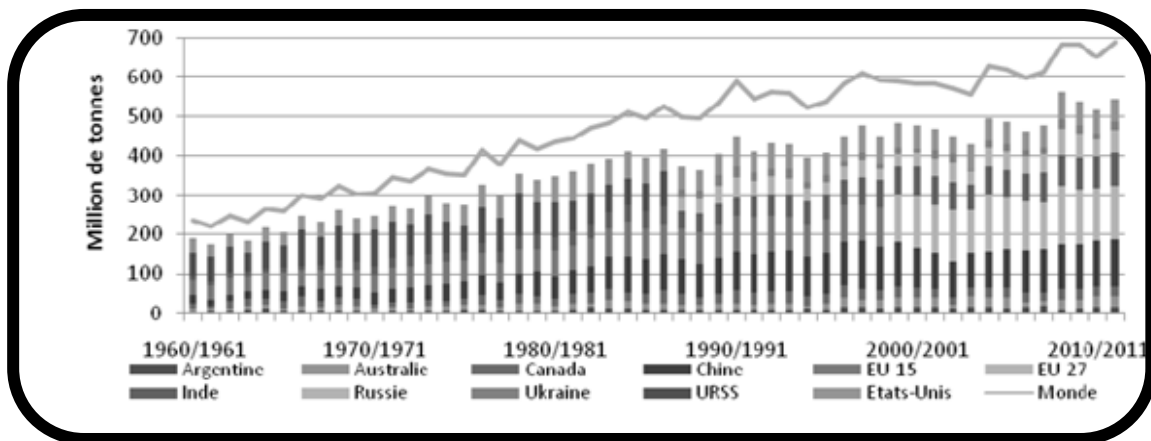
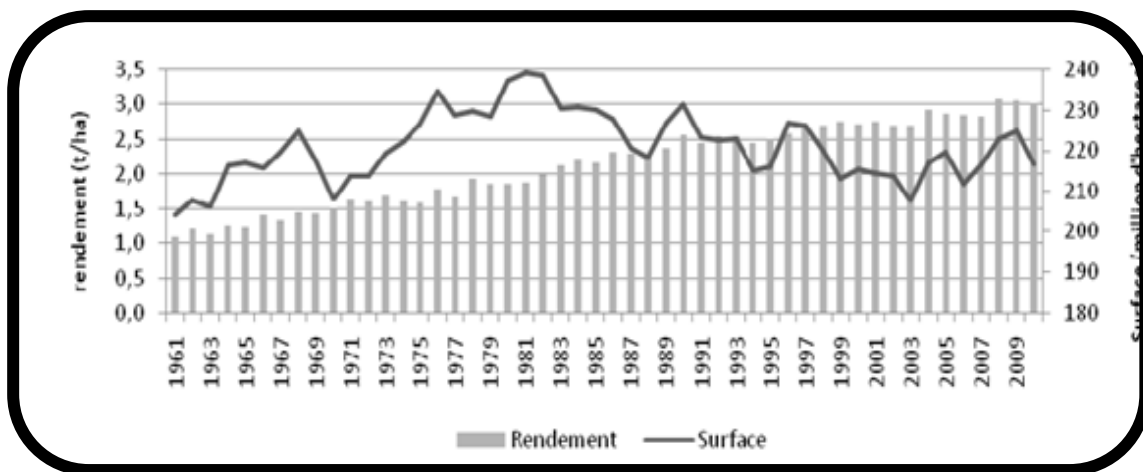


Figure 1 : Evolution de la production mondiale de blé.

Source : USDA (2011)

Cette hausse de la production de blé est principalement due à une augmentation constante des rendements à l'hectare (multiplié par 2,8 sur les cinquante dernières années) plutôt qu'une augmentation des surfaces mondiales cultivées en blé (Terrones Gavira et Burny., 2012). En effet, le nombre d'hectares cultivés en blé, après avoir connu augmentation jusqu'en 1981 (239,2 millions d'hectares de blé), n'a pas cessé de diminuer pour atteindre 216,8 millions d'hectares en 2010 (Figure 2).



**Figure 2 :** évolution de la surface mondiale et des rendements mondiaux de blé.

Source : Terrones Gavira et Burny (2012).

### 3.2. Le blé dans le contexte national

Chaque année, environ 3,3 millions d'hectares sont consacrés à des cultures céréalières dont environ 1,5 million d'hectares sont plantés de blé dur, 600 000 hectares de blé tendre, la récolte de céréales a atteint 4 MMT dont le blé panifié représentait 1% de la production totale. Le blé étant le produit de consommation de base, les habitants des pays magrébins sont les plus gros consommateurs de cette denrée au monde notamment l’Algérie avec près de 600 grammes par personne et par jour (**Abis, 2012**).

Cette consommation de blé a légèrement augmenté ces dernières années en raison de l'urbanisation accrue, de la croissance de la population et de l'augmentation de la capacité de broyage, mais devrait rester plus ou moins stagnante (**Hales et Rush, 2016**).

Selon la FAO durant l’année 2014 l’Algérie est classée en quatrième position au niveaux Africaines et à la dix-septième position au niveaux mondial avec une production du blé de 2.4 millions de tonnes, colletée est constituée en moyenne de blé dur 58,7%, blé tendre 33% (**FAO, 2014**).

#### 3.2.1. Production

La production de blé sera plutôt bonne, selon les prévisions du ministère de l’Agriculture. Benaïssa en 2013 dit que la production de blé pour cette saison devra être légèrement supérieure à celle de la saison de 2012.

Pour rappel la saison dernière (2012) la production céréalière algérienne a été estimée à près de 55 millions quintaux (Benaïssa., 2013) et cette saison (2013) la quantité record était de 61 millions de quintaux. En termes de performance, l’Algérie a produit un volume record de céréale, en 2009, dépassant le seuil de 61 millions de quintaux.

Le rendement moyen des cultures sur 135 ans (1876-2011), caractérise un BD dont la moyenne est de 6,3 q/ha. Ce du BT est 7,6, q/ha en relation avec des productions plus faibles.

#### 3.2.2. Consommation

Selon une étude menée par le bureau Algérie Consultations internationales, un algérien consomme, en moyenne 240 kg de blé par an.

**Tableau 1 : Consommation de blé, en kg, par an**

L'année	05/06	06/07	07/08	08/09	09/10	10/11	11/12	12/13	13/14
<b>Consommation (kg)</b>	232	238	239	240	241	242	243	243	244

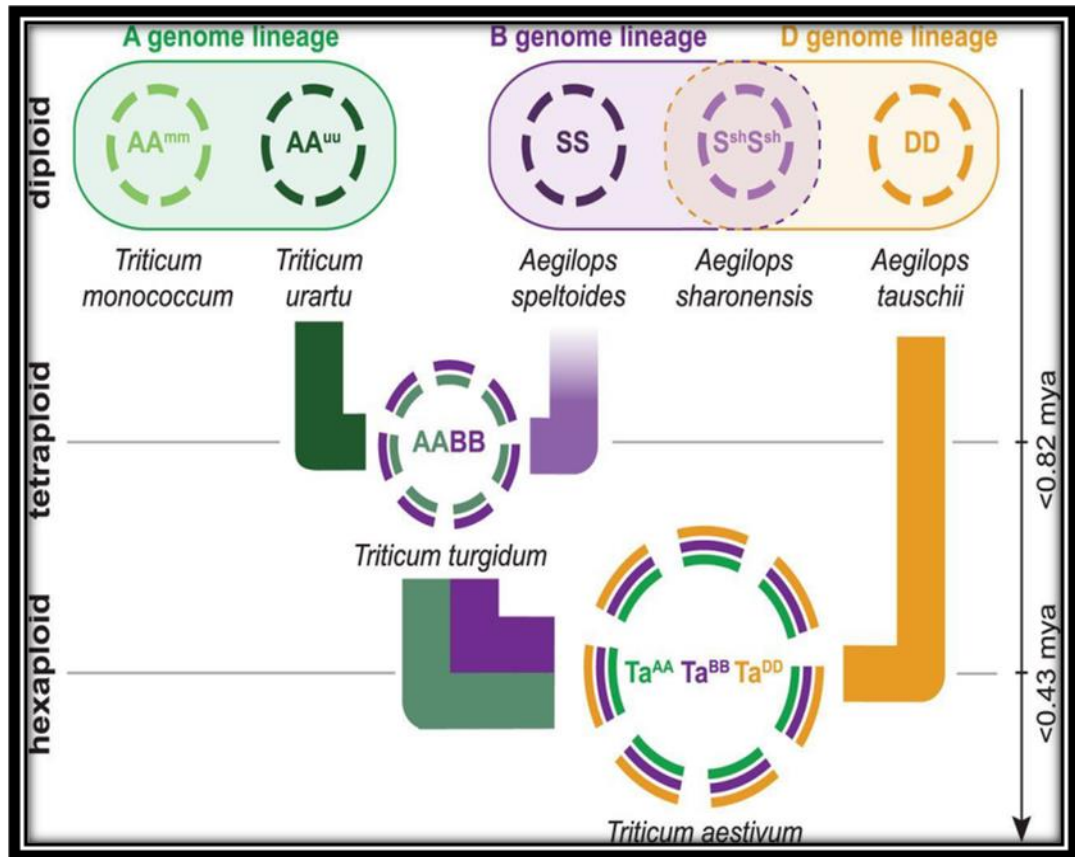
Source : Anonyme a (2014).

## Partie 2.la génomique du blé

### 1. Le génome du blé

Avec 16,7 Gb pour le génome nucléaire, le blé possède l'un des génomes les plus complexes parmi les céréales et, au-delà, dans l'ensemble du monde vivant.

De plus, le blé comporte non pas un « simple » génome nucléaire mais un génome nucléaire composite, une association de trois génomes de trois espèces différentes, regroupés dans la même cellule et formant par là même une nouvelle espèce (fig. 3).



**Figure 3 :** Diagramme schématisant des relations entre les génomes du blé avec l’histoire et la généalogie de la polyploïdisation. (Mayer et al., 2014).

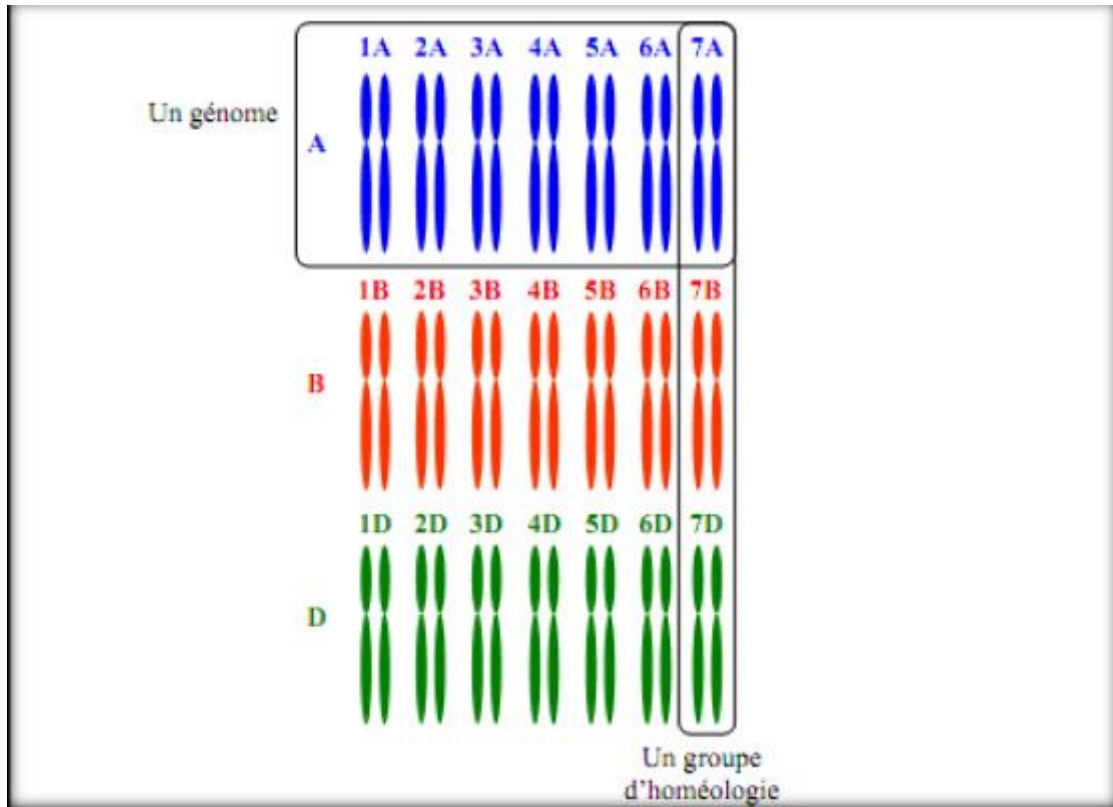
Les noms et la nomenclature des génomes sont indiqués par des cercles qui montrent une représentation schématique de la complémentarité chromosomale pour chaque espèce. Les temps sont estimés selon Marcussen et al., 2014 ; ‘mya : million years ago’.

**2. Evolution de génome polyploïde du *Triticum***

Le Blé, une plante domestiquée au génome polyploïde complexe. Le génome du blé tendre est constitué de 17 milliards de paires de base, dont plus de 80% de séquences répétées.

La taille ainsi que la forte proportion de séquences répétées constituent des obstacles importants pour le séquençage du génome du blé (Paux et al., 2008).

Outre un génome nucléaire, le blé possède comme tout végétal un génome mitochondrial et un génome chloroplastique.



**Figure 4 :** Organisation du génome du blé hexaploïde. (Bogard.,2011)

Ce génome comprend 21 chromosomes intégrant les trois génomes homéologues (A, B et D) des espèces ayant successivement formé la série d’allopolyploïdes du genre *Triticum*. L’ensemble formé par trois chromosomes homéologues se nomme “groupe d’homéologie” (fig.2). L’événement de polyploïdisation a résulté du croisement entre deux espèces diploïdes ( $2n = 14$ ) *Triticum urartu* (génome A) et une espèce proche d’*Aegilops spelta* (génome B) disparue aujourd’hui. Ce croisement a eu lieu avant la domestication par voie de sélection naturelle il y a environ 0.5 million d’années. Les descendants sont les blés tétraploïdes comme le blé dur (*Triticum turgidum ssp durum*) et l’amidonnier (*Triticum dicoccum*).

Un second croisement entre une variété domestiquée d’un blé tétraploïde et *Triticum tauschii* porteur du génome D a conduit, il y a environ 10 000 ans, à l’obtention de blés hexaploïdes tels que le blé tendre (*Triticum aestivum*) ou l’épautre (*Triticum spelta*) (Griffiths et al., 2006).

Cette allopolyploïdisation a conduit à la formation de génomes de très grande taille. Le génome du blé tendre est structuré en 21 paires de chromosomes regroupées en sept groupes homéologues représentant les génomes de chaque ancêtre.

Du fait de l’apparentement des espèces constituant le génome du blé tendre, les gènes

homéologues ont des fonctions proches mais ont pu, éventuellement au cours de l'évolution, diverger suffisamment pour acquérir des fonctions biologiques propres. Le gène *ph1* "Pairing homeologous 1", présent sur le bras long du chromosome 5B, supprime les appariements homéologues à la méiose, ce qui a pour conséquence un comportement diploïde strict à la méiose (Griffiths et al., 2006).

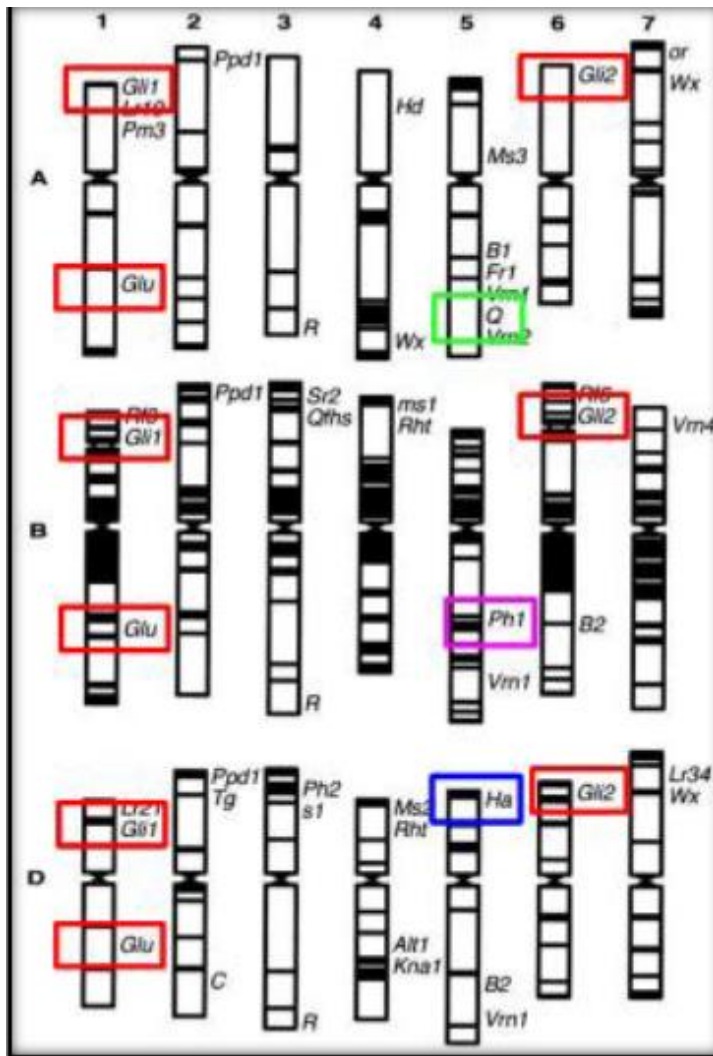
## 1.2. Homéologie des Génomes

Des cartes plus précises permettent d'argumenter les homéologies.

Les chromosomes 1 des trois groupes A, B et D sont homologues, donc portent les mêmes gènes (et de même pour les chromosomes 2, 3, et 4 .... 7). La localisation de quelques gènes sur les chromosomes A1, B1 et D1 (*Glu1* et *Gli1*). Sur le chromosome 2 le gène *Ppd1* et ainsi de suite (fig. 3). Ceci confirme qu'il s'agit de chromosomes homologues.

Toutefois, et ce n'est pas uniquement dû au caractère partiel de la carte génétique utilisée, tous les gènes ne sont pas tripliqués. L'homologie des chromosomes A, B et D d'un groupe n'est donc pas identique à celle des chromosomes homologues au sens classique du terme. On dit qu'ils sont homéologues. L'homologie ne se traduit donc pas par une identité totale des gènes portés par les chromosomes.





**Figure 5:** Quelques gènes localisés sur les chromosomes de blé. (Levesque et Madre, 2013)  
*Gli* : gliadine, *Glu* : gluténine (2 composants du gluten), *Ha* : gène qui contrôle la dureté du grain, *Q* : gène qui contrôle la structure de l'épi et *Ph1* : gène qui limite les appariements homéologues.

### 3. Cytogénétique du blé

Le genre *Triticum* comprend les espèces sauvages et domestiques généralement considérées comme du blé. La filiation génétique des blés est complexe et incomplètement élucidée. Il est acquis que le génome A provient de *Triticum monococcum*, le génome B d'un *Aegilops* (*bicornis*, *speltoides*, *longissima* ou *searsii*) et le génome D d'*Aegilops squarrosa* (également dénommé *Triticum tauschii*). (Hamel, 2010)

Le croisement naturel *T. monococcum* x *Aegilops* (porteur du génome B) a permis l'apparition d'un blé dur sauvage de type AA BB (*Triticum turgidum ssp. dicoccoides*) qui a ensuite progressivement évolué vers *L. turgidum ssp. dicoccum* puis vers *T. durum* (blé dur

cultivé). Les blés tendres cultivés (AA BB DD) seraient issus d'un croisement, également naturel, entre *T turgidum ssp. dicoccum* (AA BB) et *Aegilops squarrosa* (DD). (Hamel, 2010)

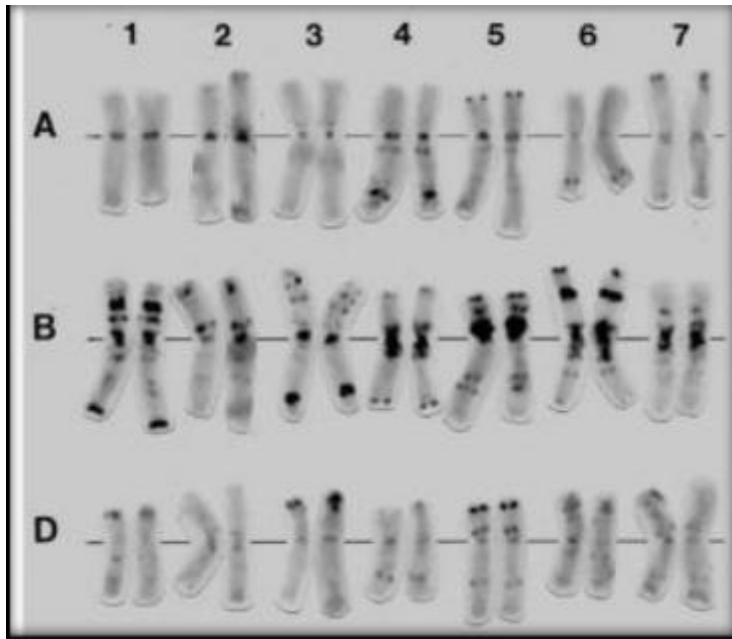
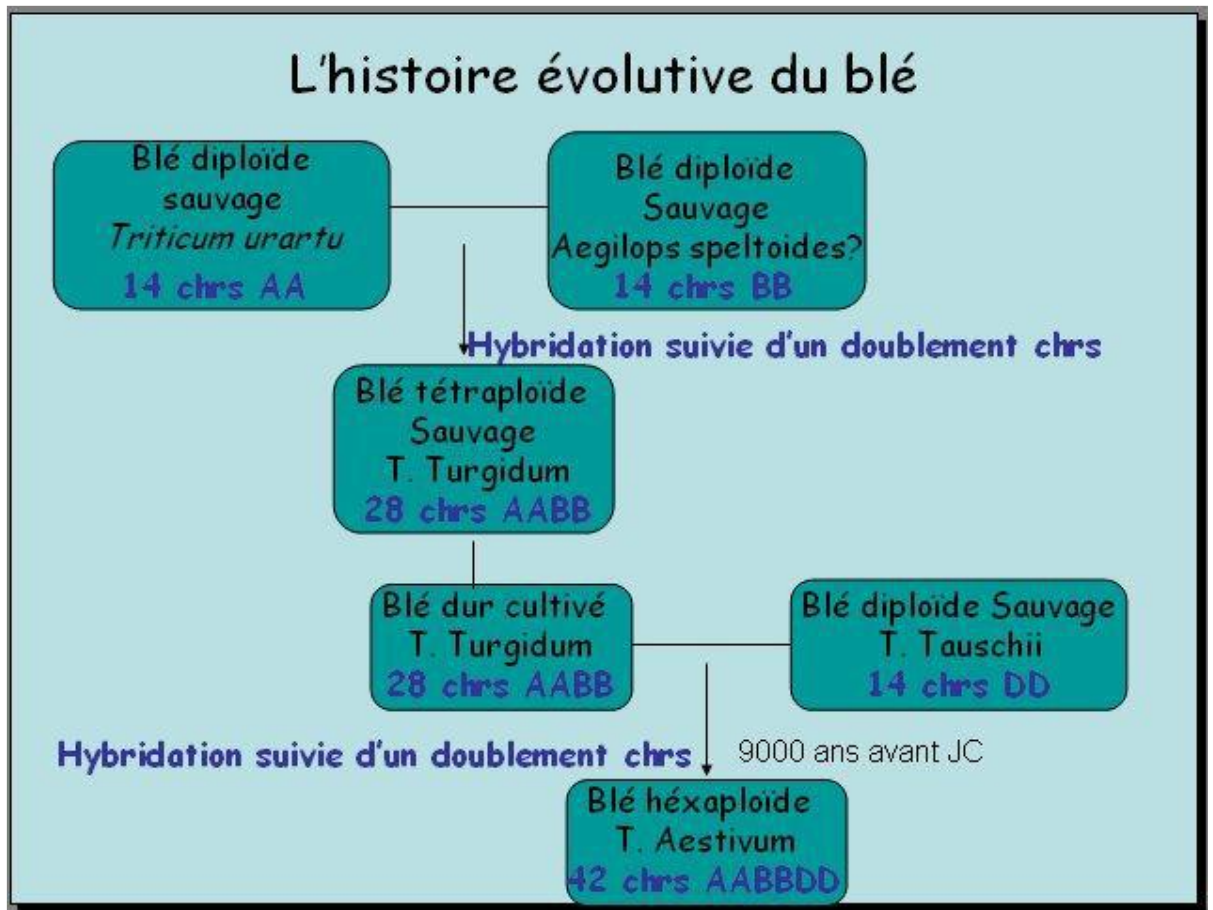


Figure6 : Les trois génomes de blé (génome polyploïde) (Zhang et al.,2011)

### 3.1.Domestication du blé

Trois espèces de blé sont actuellement cultivées dans le monde depuis 10 000 ans: le blé tendre (*Triticum aestivum*), le blé dur (*Triticum turgidum*) et l'engrain (*Triticum monococcum* ou *urartu*). La farine obtenue à partir des grains de blé tendre sert à fabriquer le pain alors que celle obtenue à partir du blé dur est utilisée dans l'industrie des pâtes et des semoules. L'engrain, premier blé cultivé, l'est encore aujourd'hui sur de petites surfaces dans certaines régions montagneuses de Turquie et de Grèce où il sert surtout à l'alimentation du bétail. Dans certaines régions on trouve encore le *Triticum turgidum* et le *Triticum monococcum* à l'état sauvage. En revanche le blé tendre n'a jamais été trouvé à l'état sauvage (fig. 5). (Levesque et Madre 2013) Par exemple, l'allèle Q du blé actuel est récent, les espèces sauvages ont toutes l'allèle q qui donne un épi fin et allongé. Le rôle du gène Q dans l'évolution de la plante au cours de la domestication apparaît essentiel. Il contrôle la forme de l'épi et des épillets. La comparaison des versions homéologues de ce gène permet d'argumenter l'histoire évolutive du blé. (Zhang et al.,2011 in : Levesque et Madre 2013).



**Figure 7:** Schéma représentatif de l'évolution du blé (*Triticum et Aegilops*). (Salamé, 2012)

# Chapitre 2. Séquençage chez les plantes

## Partie 1 : Historique et généralités

### 1 Historique

#### 1.1 Séquençage de l'ADN

En **1977** deux méthodes de séquençage ont été développées indépendamment, l'une par l'équipe de Walter Gilbert aux Etats-Unis, et l'autre par Frederick Sanger au Royaume-Uni. L'approche de Sanger est une méthode par **synthèse enzymatique** sélective, tandis que celle de Maxam et Gilbert est une méthode par **dégradation chimique** sélective. Pour cette découverte, Gilbert et Sanger ont été récompensés par le prix Nobel de chimie en **1980**.

Le **virus bactériophage  $\phi$ X174** de la famille des *Microviridae* était le premier organisme séquencé par la méthode de Sanger en **1977 (Sanger F et al, 1977)**.

L'amélioration de la technique de production des amorces, l'amplification des brins et la généralisation des traceurs fluorescents ont permis d'améliorer considérablement les techniques de séquençage. L'apparition des séquenceurs automatiques a notamment permis l'automatisation de ces technologies et a contribué à la finalisation du génotype des plantes en **2003**. Actuellement, le séquençage du génome d'un individu se fait en quelques jours pour un coût d'environ 500 euros (**Piton N et al, 2017**).

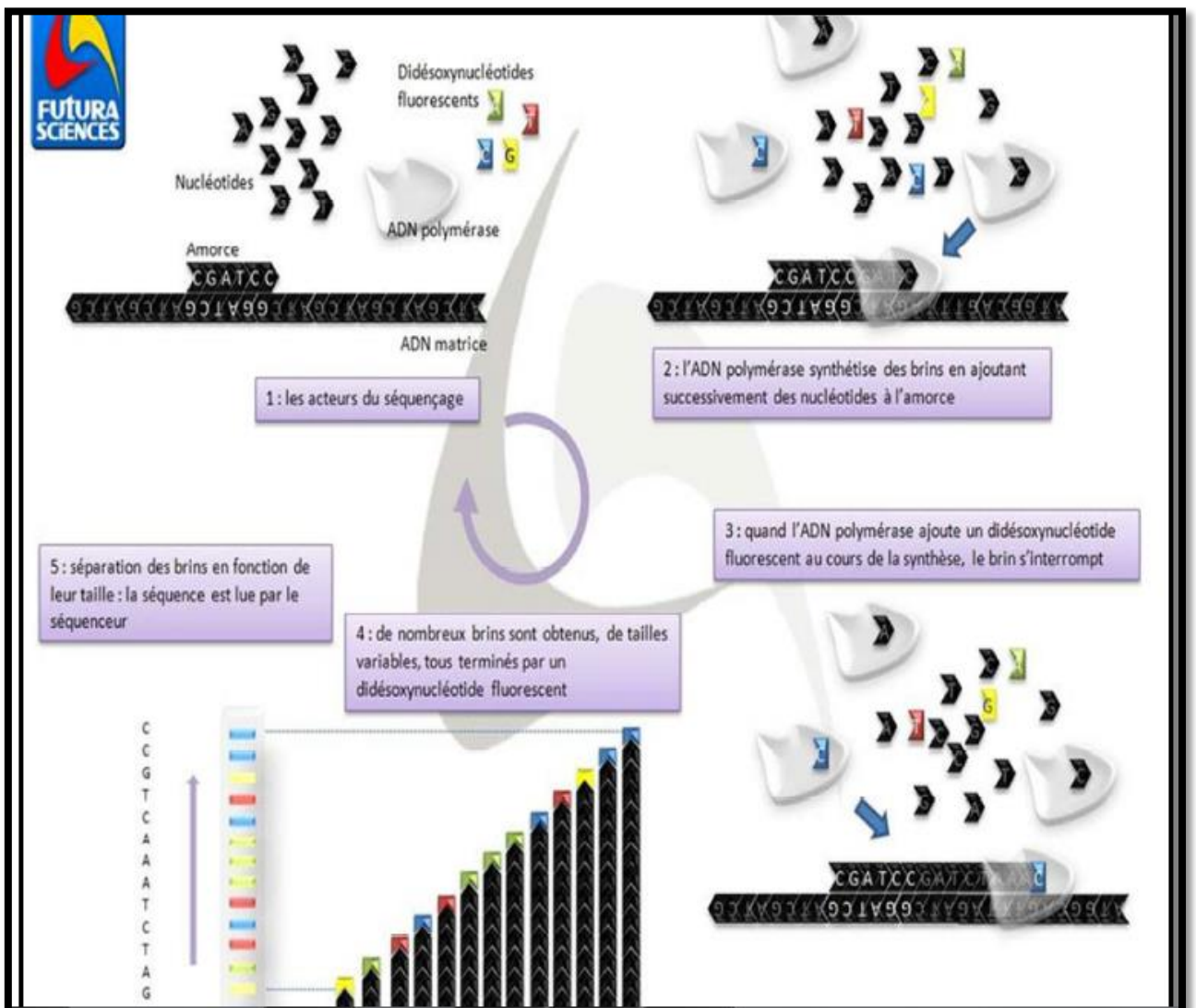
Le séquençage de l'ADN consiste à déterminer la séquence nucléotidique d'un gène ou d'un fragment de gène (**Schneider V, et al**), ce procédé repose sur la synthèse d'un brin d'ADN par une ADN polymérase, il a été mis au point pour la première fois en **1977 (Sanger F, 1977)**.

#### 1.2. Etapes du séquençage :

1. le séquençage se passe dans un tube à essai en présence des acteurs de la synthèse d'ADN: -ADN à séquencer, Nucléotides, Amorce, ,ADN polymérase.
2. L'ADN polymérase utilise aléatoirement les nucléotides présents dans le milieu pour copier le brin matrice en synthétisant un ADN de séquence complémentaire.
3. Lorsque l'ADN polymérase choisit par hasard un didésoxynucléotide (ce qui est rare puisqu'il y en a moins que des nucléotides) et qu'elle l'incorpore dans la chaîne en synthèse, celle-ci s'interrompt prématurément. Chaque didésoxynucléotide est marqué par un fluorochrome différent (**A** vert, **T** rouge, **G** jaune et **C** bleu), une chaîne qui se termine par exemple par un A sera verte.

4. Puisqu'un grand nombre de réactions de synthèse ont lieu dans le tube, il existe statistiquement des chaînes de toutes les tailles (correspondant à un arrêt de la synthèse à chaque nucléotide) et beaucoup de fragments d'une même taille. Ces chaînes commencent toutes au même endroit sur l'ADN matrice, toutes celles qui possèdent la même longueur se terminent par le même didésoxynucléotide marqué.

5. Il est alors possible de séparer les chaînes d'ADN obtenues en fonction de leur taille sur un gel d'acrylamide en présence d'un courant électrique. Plus les chaînes sont courtes, plus elles migrent loin et tous les fragments d'une même taille migrent à la même distance. On obtient alors une succession de bandes colorées, chacune correspondant au dernier nucléotide incorporé. Il suffit de lire la succession des couleurs pour connaître l'ordre des nucléotides.(figure8)



**Figure8 : Différentes étapes du séquençage(Peltier C.2011)**

## **2. Intérêt du séquençage**

En **2001**, 95 % de l'ADN humain était séquencé et en **2006** il a été finalisé (**Lander ES et al ,2001**).(**Venter JC et al,2001**). Par ailleurs, les génomes de nombreux agents infectieux de mammifères et de plantes ont également été séquencés dans leur totalité, leur connaissance a modifié considérablement les recherches biomédicales et biologiques en ouvrant de vastes panoramas dans le domaine de la médecine (diagnostic, thérapeutique, prédiction, pronostic, prévention. . .) et dans de nombreuses autres disciplines biologiques (anthropologie, agronomie, environnement. . .).

Les techniques de séquençage évoluent et leurs applications s'élargissent. Par ailleurs, cette technique a pu se démocratiser dans de nombreux laboratoires, en partie depuis la description de la Polymerase Chain Reaction (PCR) en **1985**, suivie de sa diffusion très large dans les laboratoires de biologie moléculaire. Depuis **2000**, outre la PCR que nous ne décrivons pas, de nouvelles techniques de séquençage se sont développées. Elles constituent un progrès technologique révolutionnaire (**Lamoril J. et al,2008**).

## **3. Automatisation du séquençage :**

Grace à cette automatisation on peut séquencer jusqu'à 1000 bases avec les meilleurs séquenceurs contre 200 à 300 via une méthode manuelle. En **2010**, le nombre de génomes séquencés approche les 3000 (**Human genome ,2010**).

Aujourd'hui, la plupart des séquençages sont réalisés par des séquenceurs industriels entièrement automatisés. Les fragments d'ADN sont marqués par des marqueurs fluorescents; leur taille est ensuite déterminée par chromatographie ou électrophorèse assistée par ordinateur. En effet, lors de l'électrophorèse manuelle le nombre de bases est limité afin de ne pas rendre le chromatogramme illisible par la surcharge des bandes et ainsi ne plus permettre une lecture horizontale. C'est notamment grâce à la rapidité de ces appareils que le séquençage du génome humain fut réalisé en un temps record par rapport aux prévisions effectuées lors du démarrage du projet.(**Aghrouh N ;2018**).

## **Partie 2 : Séquençage de première génération**

### **1 Technique de Maxam et Gilbert :**

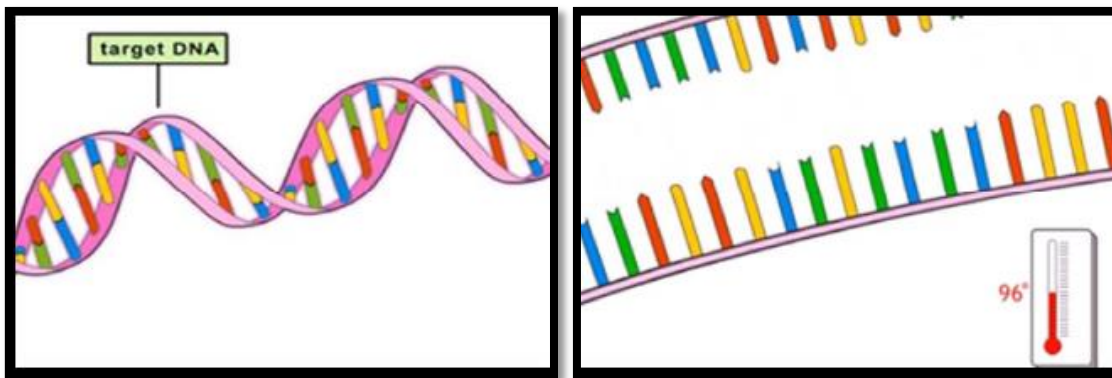


Le séquençage débute lorsque Maxam et Gilbert développent une technique basée sur le marquage radioactif de fragment et leur coupure sélective par dégradation chimique (**Maxam *et al* ,1977**).

La méthode de Maxam et Gilbert publiée parallèlement à celle de Sanger, par son caractère révolutionnaire a grandement contribué à l'histoire de la biologie moléculaire. Cette technique permettait d'analyser des fragments allant jusqu'à 500 Pb.

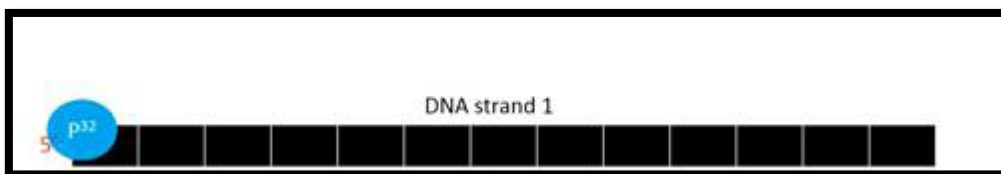
**Procédure :**

**1.** Dénaturation d'un ADN double brin en simple brin en augmentant la température , Rupture des liaisons hydrogène entre les bases nucléiques des deux chaînes complémentaires de l'ADN , Obtention d'un ADN simple brin.



**Figure 9:** Dénaturation de l'ADN à la chaleur.(anonyme,2018)

**2.** Marquage radioactif d'une extrémité 5' du fragment d'ADN à séquencer par le phosphore radioactif (P32).



**Figure 10:** Marquage de l'ADN.

**3.** Clivage du brin d'ADN à des positions spécifiques en utilisant des réactions chimiques : Cette technique est basée sur la propriété de certains agents chimiques : l'hydrazine, le diméthyle sulfate (DMS) et l'acide formique, de modifier les bases de l'ADN.

- Le DMS agit au niveau des bases «G».
- L'acide formique agit au niveau des bases «A+G».
- L'hydrazine agit au niveau des bases «C+T» (en milieu alcalin, l'hydrazine agit

uniquement au niveau des « C »).

On utilise l'un des produits chimiques suivi de la pipéridine, cette dernière casse les brins d'ADN au niveau des bases modifiées.

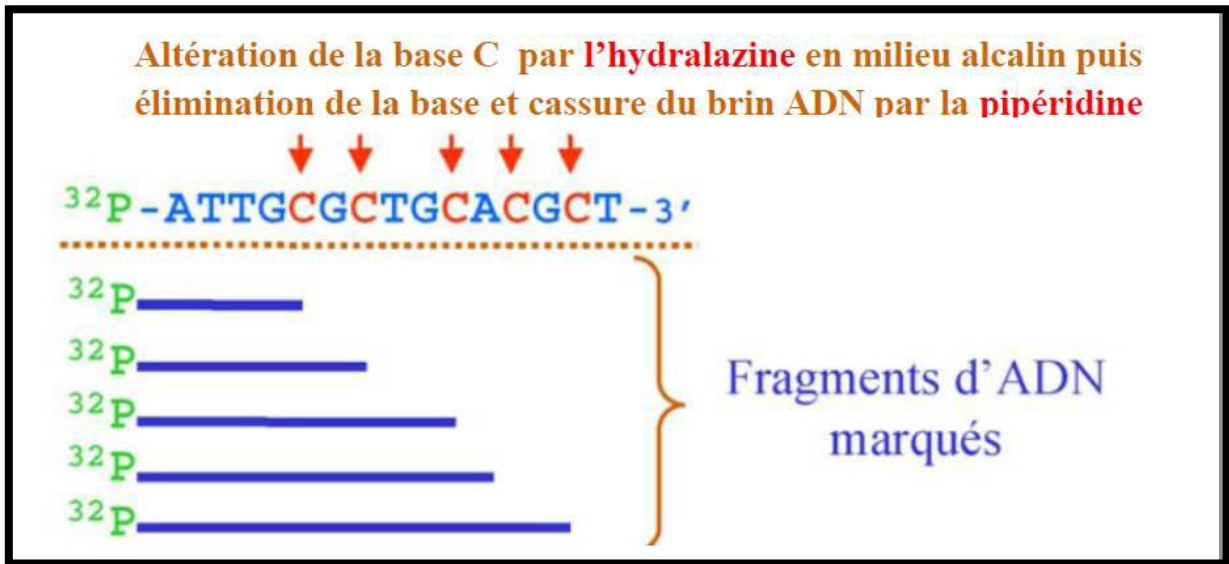


Figure 11: Altération de l'ADN par l'hydralazine

4. Obtention de plusieurs brins d'ADN de tailles différentes dans quatre tubes de réaction.

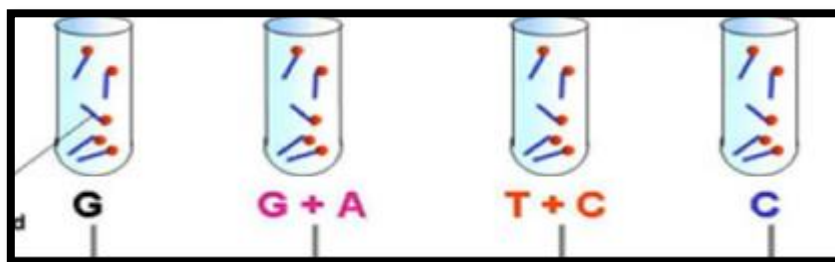


Figure 12: Brins d'ADN de tailles différentes dans les quatre tubes.

5. Les fragments sont soumis à une électrophorèse dans des gels d'acrylamide à haute résolution pour la séparation des tailles.

Ces gels sont placés sous un film radiographique qui donne une série de bandes foncées et montrent l'emplacement des molécules d'ADN radiomarquées. Les fragments sont classés par taille et nous pouvons donc déduire la séquence de la molécule d'ADN dans le sens 5' → 3' du bas en haut.



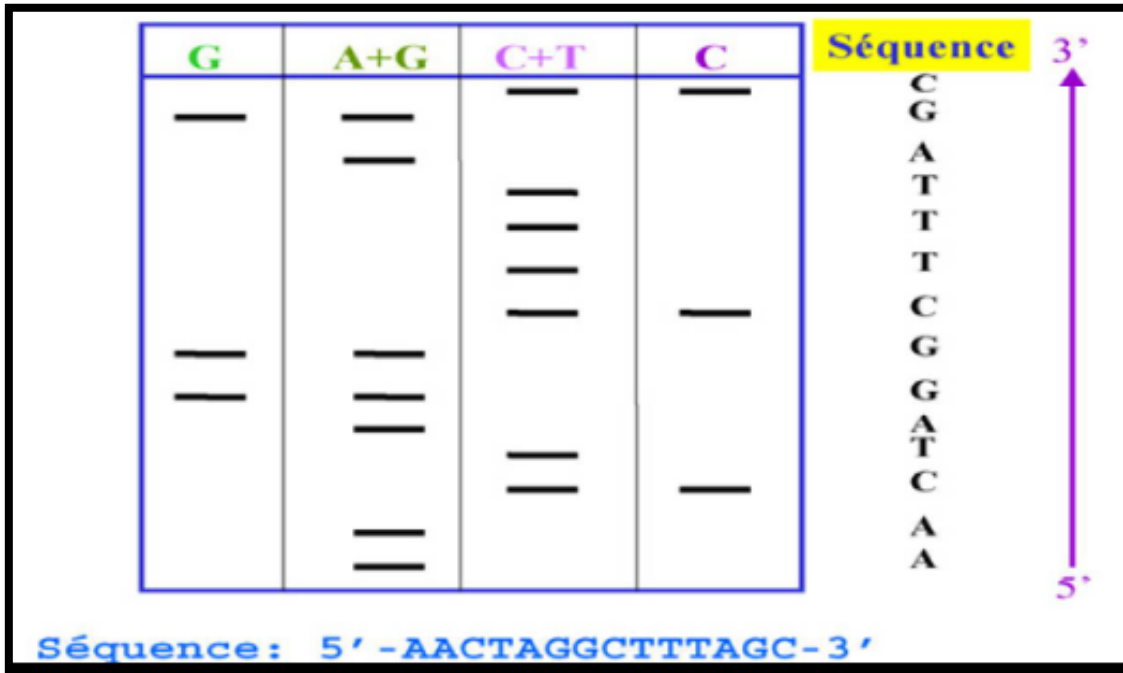
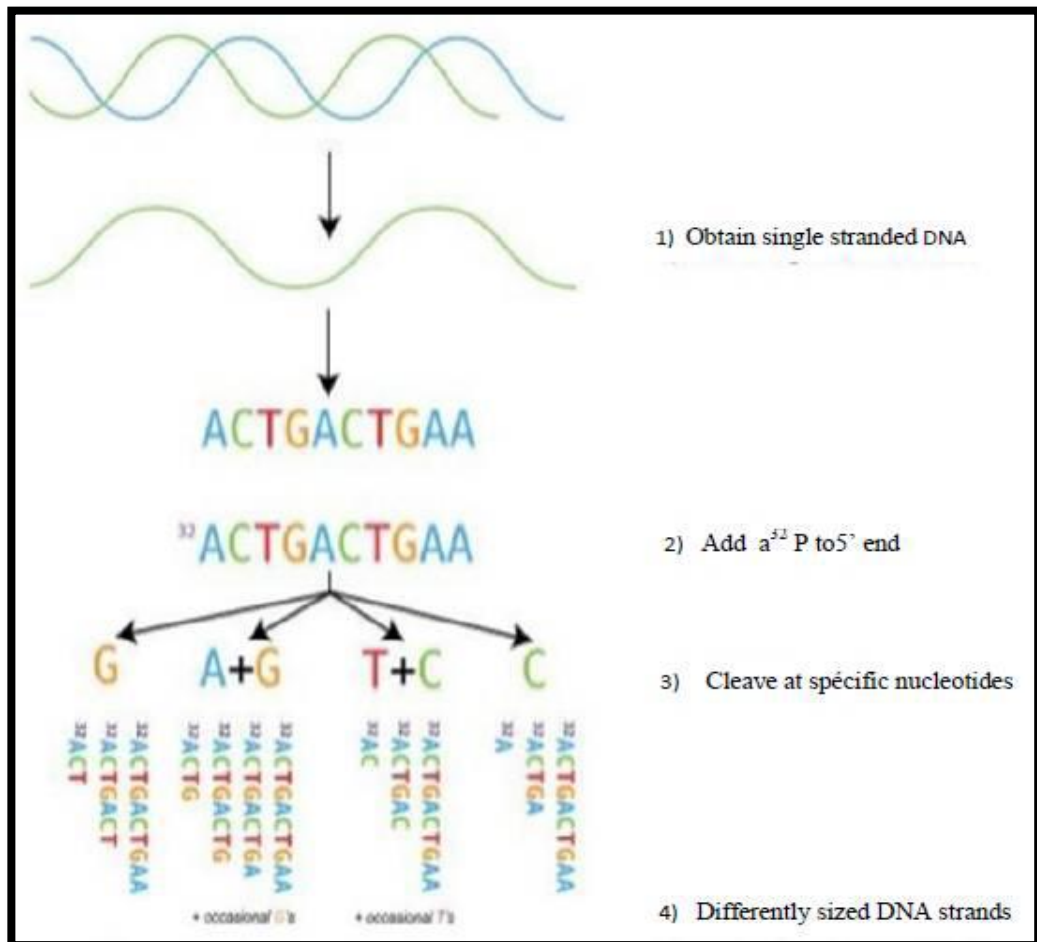


Figure 13: Autoradiogramme après traitement chimique des fragments.



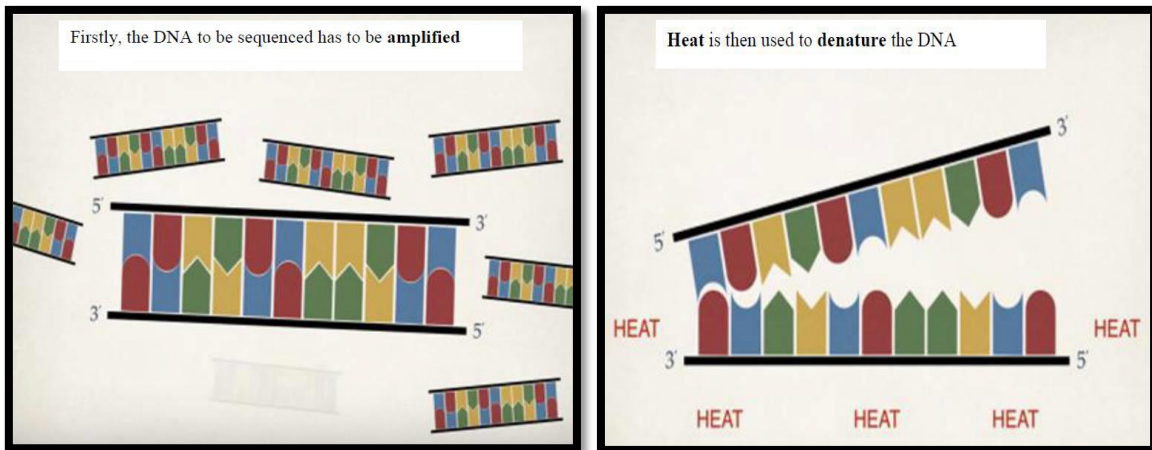
**Figure 14:** Technique de Maxam et Gilbert.

**2 Technique de Sanger :**

Le séquençage génétique de Sanger est un moyen de déterminer l'ordre des quatre nucléotides dans un brin d'ADN. Ce séquençage a permis aux scientifiques de séquencer une vaste gamme d'organismes, il était la méthode de choix avant la découverte des séquenceurs de nouvelle génération. Cette méthode part de l'idée que, en copiant des brins d'ADN et en surveillant quels nucléotides sont ajoutés, un par un, on trouve la séquence de nucléotides.

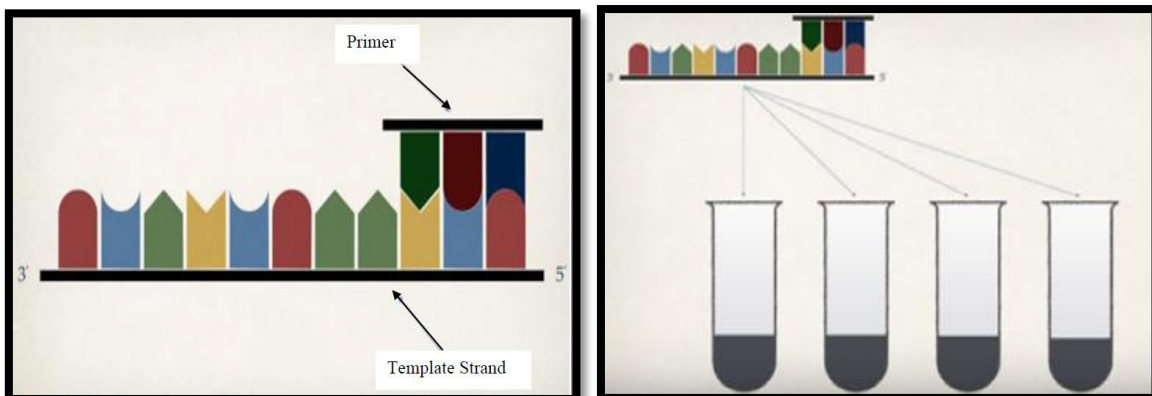
**2.1. Comment fonctionne-t-elle?**

1. Amplification de l'ADN à séquencer → Obtention d'un très grand nombre de copies d'une séquence d'ADN choisie. Puis, Dénaturation de l'ADN par la chaleur.



**Figure 15:** Amplification et dénaturation de l'ADN.

2. Un ADN amorcé est dispersé entre quatre récipients de réaction.



**Figure 16:** Ajout d'un ADN amorcé dans les quatre récipients.

3. L'ajout de l'ADN polymérase à tous les quatre récipients suivi par les quatre dNTPs (dATP, dCTP, dGTP, dTTP) marqués par des fluorophores qui dégagent une lumière colorée lors de l'incorporation d'un nucléotide ; la couleur émise indique la base incorporée.

4. Un seul type de ddNTPs est ajouté à chaque récipient de réaction.

Ces didésoxyribonucléotides une fois incorporés dans le nouveau brin synthétisé, empêchent la poursuite de l'élongation, car ils n'ont pas de groupe hydroxyle au niveau du carbone.

Attention : les ddntps sont ajoutés à faible quantité pour inhiber la synthèse d'une façon aléatoire (F. Sanger, *et al*, 1977).

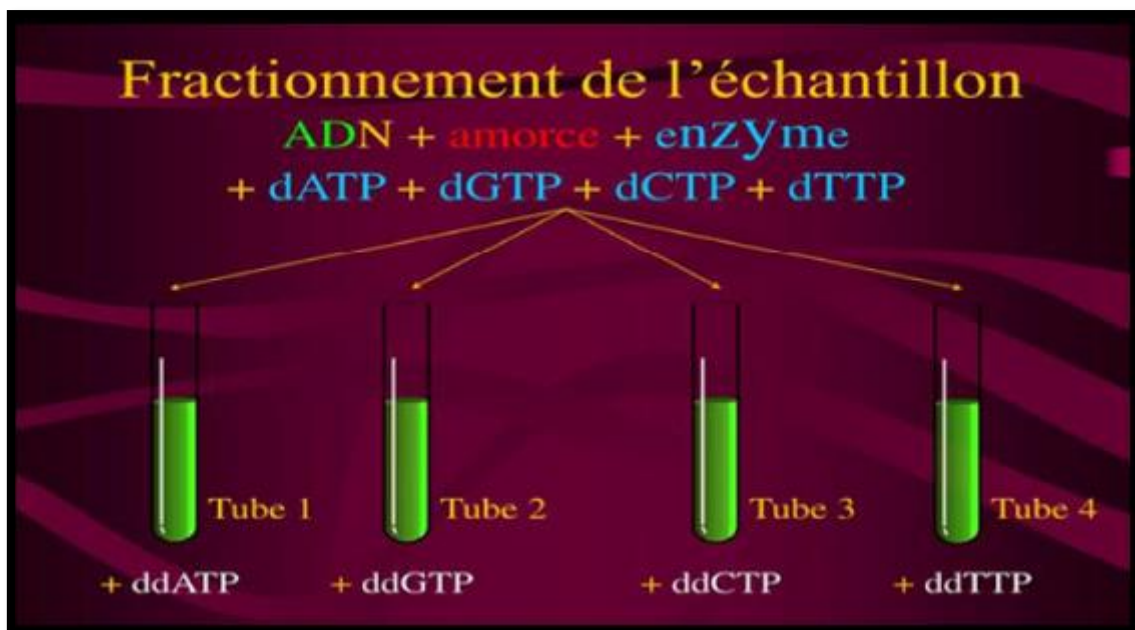
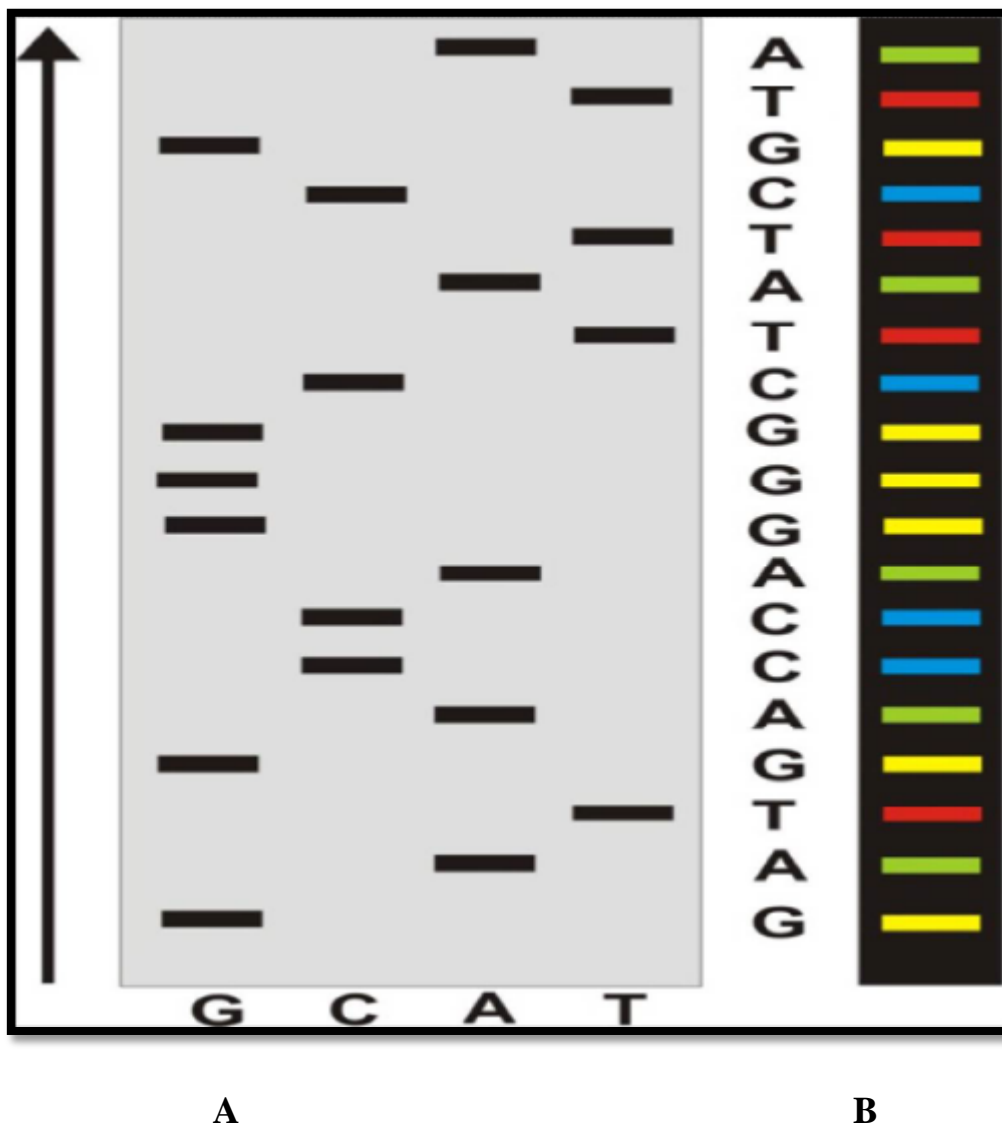


Figure 17: Ajout des ddNTPs.

5. On obtient un mélange de fragments d'ADN de tailles croissantes qui se terminent tous au niveau d'une des bases dans la séquence.

6. On sépare ensuite les fragments synthétisés par électrophorèse en gel (les ADN sont des anions, ils vont donc vers le pôle +), en fonction de leur longueur (les plus petits vont plus vite).

7. On lit ensuite les taches successives identifiées par leur couleur ce qui révèle la séquence des fragments synthétisés. Pour la lecture, on commence par le bas et on regarde les bandes (tirets noirs) afin de déterminer la séquence du fragment d'ADN (A Debourgogne, *et al*, 2017). L'innovation du marquage par les fluorophores a rendu possible le séquençage de beaucoup plus de brins d'ADN.



**Figure 18:** **A** - Electrophorèse en gel standard.  
**B** - Séquençage à l'aide de fluorophores.

### Partie 3 :Séquençage de nouvelle génération

A partir de **2006**, une convergence entre les techniques microfluidiques, les nanotechnologies et les progrès de l'informatique, a permis l'émergence de nouvelles machines capables de réaliser plusieurs millions de réactions de séquence en parallèle. Ces automates sont appelés séquenceurs à très haut débit ou séquenceurs de nouvelle génération (**Bai X, et al, 2005**).

C'est grâce notamment à la lecture de plusieurs millions de séquences en parallèle que ces nouveaux séquenceurs ont pu révolutionner les analyses en génomique (**Metzker M L ;2010**).

Le séquençage à haut débit a permis des gains de temps très importants par rapport aux méthodes traditionnelles (**Harris S R, et al ,2010**).

### **1. Développement et caractéristiques du séquençage de nouvelle génération (NGS) :**

Pendant la dernière décennie, de nombreuses innovations techniques ont permis de mettre de nouveaux outils à la disposition des biologistes. Parmi ces avancées, la plus importante est l'avènement des nouvelles techniques de séquençage de l'ADN (ou NGS : Next Generation Sequencing).

Le terme NGS regroupe un ensemble de techniques et de technologies développées dans le but d'obtenir un haut débit de séquençage d'ADN comparativement aux méthodes préexistantes telles que le séquençage de Sanger.

En effet, ces anciennes techniques, bien que fiables et éprouvées, ne permettent qu'une acquisition lente des données.

Aujourd'hui, une seule machine peut séquencer 1 Gigabase (Gb) en quelques heures (**Kingsmore SP et Saunders CJ, 2010**).

Par exemple, un séquenceur de Sanger permet de générer 19 kb par heure, soit 96 séquences de 600 Pb en 3 heures. A ce rythme, le séquençage des 3,2 Gb d'un génome humain à l'aide d'une unique machine prendrait plus de 19 ans. . De nombreux progrès dans des domaines indépendants de la biologie, tels que l'augmentation des puissances de calcul informatique, le développement de capteurs optiques à très haute résolution, l'amélioration des algorithmes de traitement d'image et de signal et la maîtrise de la microfluidique ont permis le développement de ces nouvelles technologies de séquençage.(**Aghrouh N ,2018**).

Ces technologies sont basées sur le séquençage parallèle, indépendant et simultané de centaines de milliers, voire de centaines de millions de brins d'ADN, pouvant ainsi générer plus de 50 Mb par heure par machine (plus de 2600 fois plus que le séquençage classique). .(**Aghrouh N ,2018**).

### **2. Séquençage de seconde génération :**

Depuis une dizaine d'année, les technologies de séquençage ont connu une évolution impressionnante, d'autant plus depuis **2005** avec l'arrivée des séquenceurs de seconde génération sur le marché en **2007**( **Mardis E R.,2011**).

Le séquençage de Sanger a été la référence pendant une vingtaine d'années mais actuellement il est dépassé par la nouvelle génération de séquenceurs permettant le séquençage à très grande échelle et de manière beaucoup plus rapide avec un coût à la base beaucoup plus faibleLe premier séquençage complet d'un génome humain a été réalisé en 15 ans alors qu'il peut désormais l'être en 10 jours sur une plateforme Illumina HiSeq 2000.

Toutes les méthodes de NGS reposent approximativement sur le même principe (**Shendure J, Ji H.2008**), la majorité de ces technologies sont propriétaires et sont commercialisées par des grands groupes industriels, le coût de développement et de fabrication des séquenceurs n'étant pas à la portée des laboratoires. Parmi les technologies disponibles, on citera: « 454 sequencing » de Roche Applied Science, « Solexa » d'Illumina, « SOLiD » d'Applied Biosystems, « Heliscope » d'Helicos, « Ion Torrent » de Life Technologies et « PacBio » de Pacific Bioscience. Le séquençage 454 a été la première technologie NGS commercialement disponible, suivi peu de temps après par le Solexa. Ces deux technologies sont les plus utilisées à l'heure actuelle du fait de leur ancienneté qui permet la disponibilité d'un grand parc de machines et surtout une meilleure compréhension de leurs avantages et inconvénients respectifs.

### **2.1. Trois grandes technologies du séquençage de nouvelle génération séquençage :**

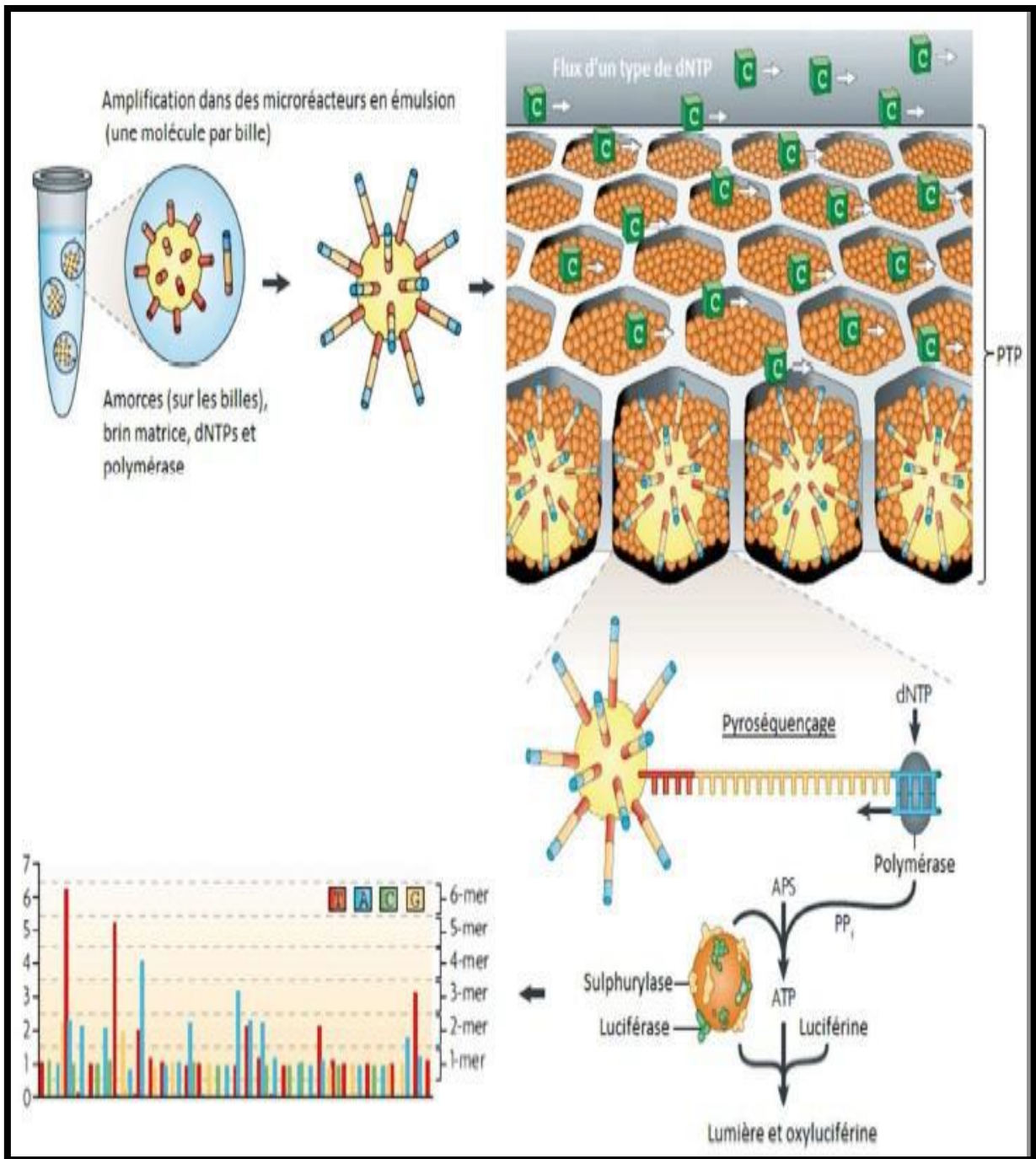
Trois technologies NGS sont apparues de façon quasi simultanée et dominent actuellement le marché: le pyroséquençage (454/Roche), le séquençage avec des terminateurs réversibles (Illumina/Solexa) et le séquençage par ligation (SOLiD™) (**Woollard P.M.2011**) et (**Desai A.N. A Jere,2012**).

Elles utilisent la détection de signaux chimiques ou lumineux lors de l'incorporation de bases nucléotidiques pendant la synthèse du brin complémentaire pour déterminer la séquence (**Metzker, M L.2010**). Ces technologies partent d'un simple brin d'ADN préalablement préparé mais utilisent par la suite des approches différentes (**Shendure J. H Ji.2008**).

#### **2.1.1. 454 de Roche :**

Le premier des séquenceurs de nouvelle génération a été commercialisé par 454 Life Sciences en 2005, depuis racheté par Roche, mais la plateforme est toujours connue sous le nom générique de 454. Jonathan M. Rothberg en a élaboré la technologie et a démontré sa robustesse avec le séquençage du génome de *Mycoplasma genitalium* (**A Debourgogne et al ,2017**).





**Figure19 : Aperçu de la technologie de séquençage 454**

La spécificité de cette technologie repose sur une PCR en émulsion pour l'amplification des fragments à séquencer : la PCR a lieu dans une microgoutte renfermant une microbille d'agarose en phase aqueuse, séparée des autres billes (plusieurs millions) par de l'huile ( **Figure 19**).

On obtient ainsi des copies d'un seul fragment d'ADN par bille. Chacune des billes est ensuite déposée dans un des 1,6 millions de puits d'un support solide appelé PTP (PicoTiter Plate) (**Bai X, 2005**).

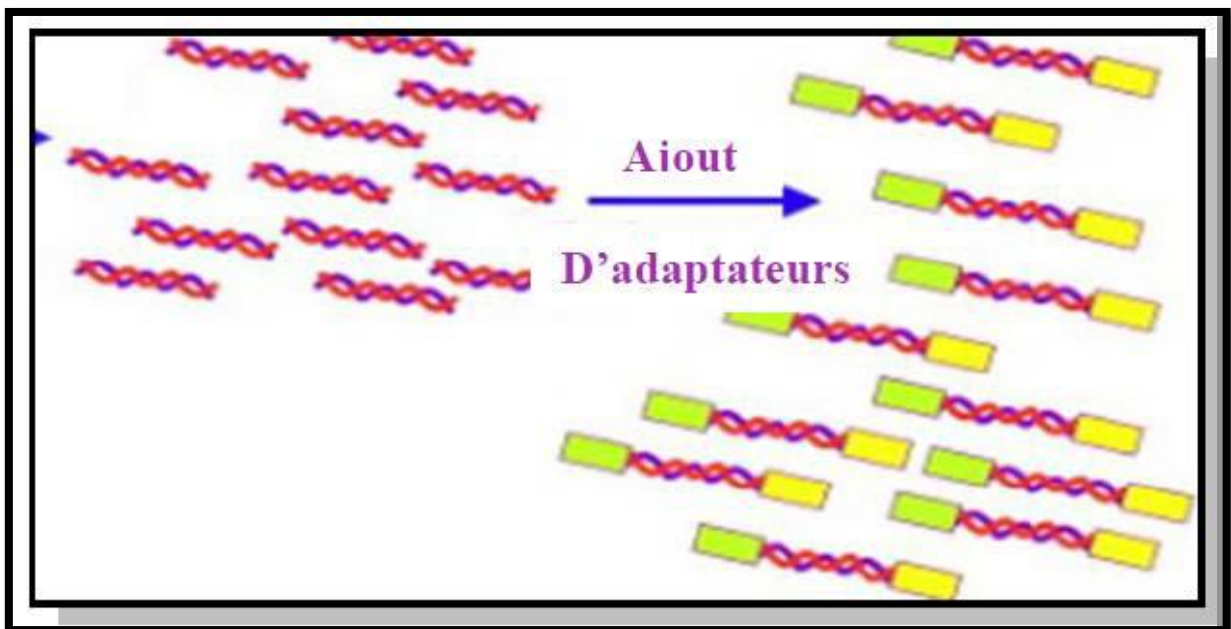
### 2.1.2 Solexa/Illumina

Cette méthode permet l'acquisition en parallèle de plus de 3 milliards de séquences de 150 paires de bases. Chaque position étant lue l'une après l'autre, les erreurs principales de cette technologie sont des erreurs de substitution d'une base par une autre (Desai A.N. A Jere.2012),( Loman , N J.2012).

Pour cette technique, l'amplification de l'échantillon à analyser ne s'effectue pas en solution mais sur un support solide. Cette technologie repose sur une amplification en pont des fragments à séquencer (bridge PCR : en phase solide) sur une surface de verre appelée flow cell (FC) similaire à une lame de microscope divisée en huit lignes (une ligne par échantillon).

#### Procédure :

1. Génération d'une banque d'ADN double brin à partir de l'échantillon à analyser avec l'ajout d'adaptateurs spécifiques aux extrémités des brins.



**Figure20:** Ajout des adaptateurs spécifiques aux extrémités d'ADN (Lamoril J,2008).

2. Dénaturation de l'ADN en simple brin.

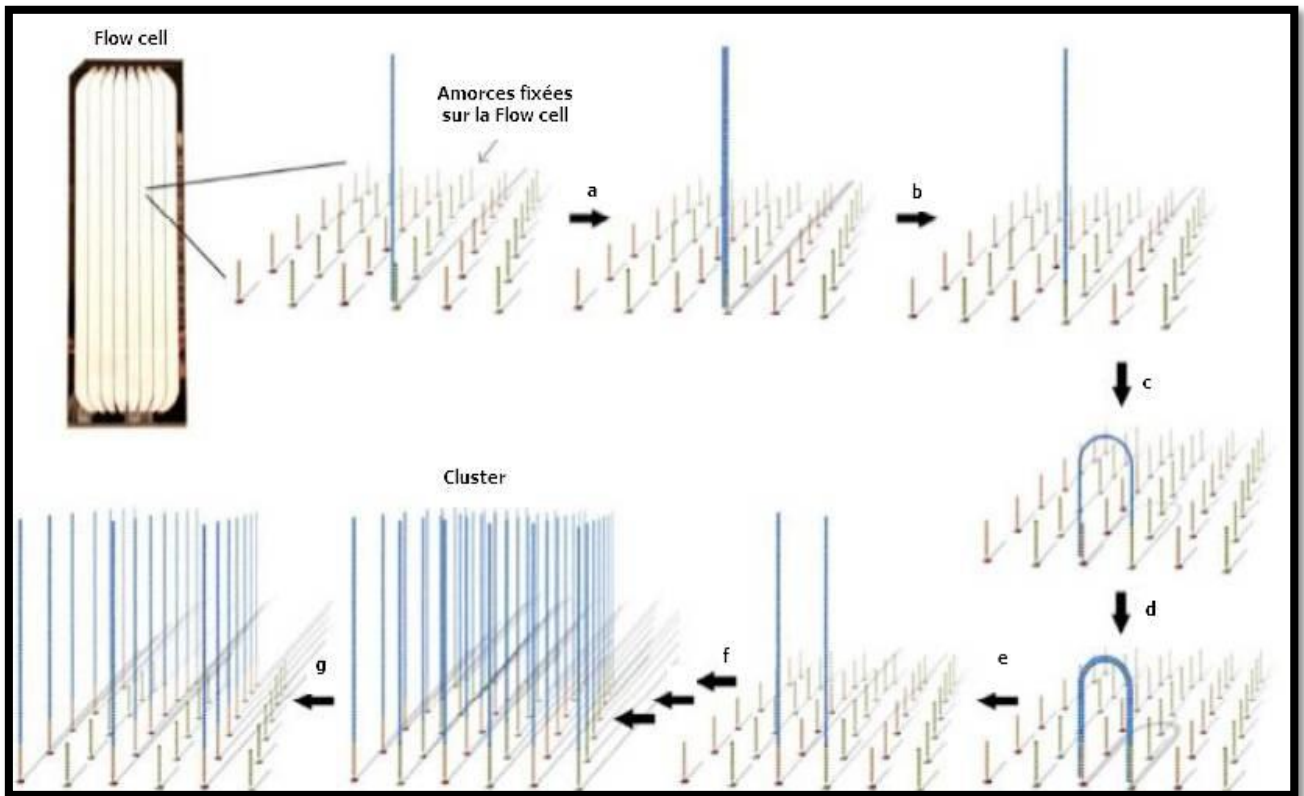
3. Fixation de l'extrémité des simples brins aléatoirement à la surface du flow cell





**Figure 21:** Flow Cell.

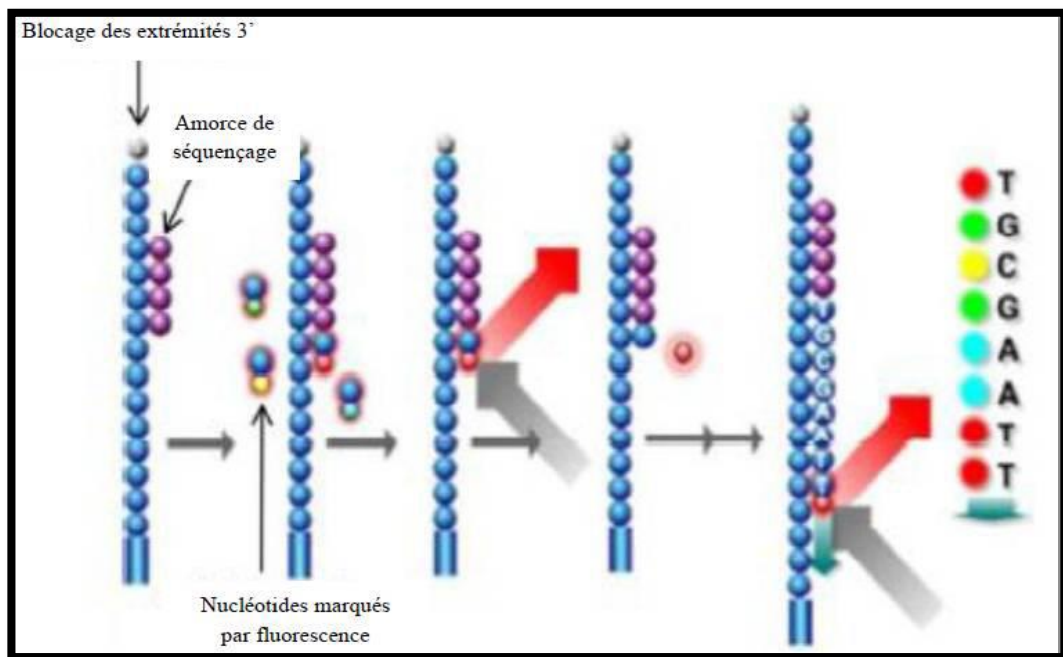
**4.** Un nouveau brin est alors synthétisé par une polymérase (**Figure 21 a**), il est fixé de façon covalente à la FC grâce à l'adaptateur spécifique. Le brin d'origine est alors éliminé par dénaturation (**b**) et l'extrémité libre du brin restant s'hybride à une amorce adjacente pour former un pont (**c**). La polymérase synthétise à nouveau le brin complémentaire pour former un pont d'ADN double brin (**d**) puis les deux copies sont libérées par dénaturation (**e**). Le cycle d'amplification en pont (étapes **c** à **e**) recommence pour former à terme un regroupement denses d'ADN clonés en une zone appelée cluster (**f**). Les brins anti-sens sont ensuite clivés (**g**) : c'est la linéarisation.(Figure21).



**Figure22** : Amplification en pont de la technologie Illumina

L'extrémité 3' libre des fragments d'ADN est bloquée et l'amorce de séquençage s'y hybride (Figure 22). Le séquençage s'effectue sur des centaines de millions de clusters simultanément grâce à une chimie de terminateurs réversibles : des nucléotides bloqués marqués par des fluorophores différents (sont excités lors de leur incorporation par l'ADN polymérase) sont ajoutés, l'un d'entre eux est incorporé, après excitation du fluorochrome porté par l'analogue régressif, le signal émis permet la détermination du nucléotide incorporé.

La fluorescence émise est relevée puis le fluorophore et le bloqueur sont clivés permettant l'ajout d'un nouveau nucléotide. Un seul nucléotide est rajouté à chaque cycle puisque les analogues nucléotidiques sont des terminateurs de chaîne. La lecture est effectuée cycle après cycle, elle est répétée jusqu'à la taille attendue du fragment.



**Figure23** : Aperçu de la technologie de séquençage Illumina.

Le séquençage Illumina a été utilisé pour la recherche des transcriptomes de la patate douce et le gymnosperme genre *Taxus* . .(Aghrouh N ,2018).

### 2.1.3 SOLiD

Le SOLiD (Sequencing by Oligonucleotide Ligation and Detection) a été la troisième plateforme de séquençage de nouvelle génération, commercialisée par Applied Biosystems (aujourd'hui Life Technologies) depuis 2007.



La technologie repose également sur une PCR en émulsion sur billes. Le séquençage ne s'effectue pas par synthèse comme sur les plateformes précédentes mais par ligation. Une amorce de séquençage universelle se fixe sur l'adaptateur puis des oligonucléotides dégénérés de 8 bases, marqués par fluorescence, sont ajoutés ( Figure 24).

### **3. Séquençage de troisième génération :**

Ces nouvelles méthodes utilisent pour la plupart la détection d'un signal chimioluminescent à chaque incorporation de nucléotide afin de séquencer directement les fragments d'ADN un par un ce qui rend l'amplification de l'ADN inutile. Ces innovations diminueraient la durée et simplifieraient considérablement les étapes de préparation des banques d'ADN. De plus, l'élimination de ces étapes permettrait aussi de diminuer les sources de biais possibles pouvant être introduits par les phases de PCR (erreurs de polymérisation ou recombinaisons par PCR).

Un autre avantage de ces techniques réside dans le fait que des ADN faiblement concentrés voir même dégradés pourraient ainsi tout de même devenir exploitables (**Orlando, L, Al 2011**).

Le principe de la troisième génération peut être symbolisé par le séquençage d'une molécule d'ADN sans étape de pré-amplification (contrairement à la génération actuelle type 454 Roche, SOLiD Life technologie, Ion Proton, HiSeq Illumina, ...) en conservant l'incorporation de nucléotides, par cycles ou non (dans ce dernier cas, le terme de « Séquençage d'ADN simple molécule en temps réel » est approprié). Les technologies « SMS » pour « Single Molecule Sequencing » peuvent être regroupées selon trois catégories:

- Technologies de séquençage en temps réel impliquant la synthèse du brin d'ADN complémentaire via une ADN polymérase.
- Technologies de séquençage par détection des bases successives d'une molécule d'ADN à travers des nanopores.
- Technologies de séquençage basées sur des techniques de microscopie.

#### **3.1. Hélioscope**

Hélicos Biosciences a développé le premier séquenceur d'une molécule unique : le Hélioscope Single Molécule Séquenceur. A la différence de la chimie utilisée sur la plateforme Illumina, ici les nucléotides sont marqués avec le même fluorophore. Hélicos ne vend désormais plus d'appareil et propose un service de séquençage. **(Aghrouch N ,2018)**.

### 3.2. Technologies des nanopores :

L'idée de séquencer des molécules d'ADN en leur faisant traverser des nanopores est apparue dès les années 1990, et la première démonstration de faisabilité a été publiée en 1996 (**Kasianowicz J,1996**).

Deux appareils (GridION et MinION) utilisant une technologie de séquençage par nanopore (Oxford Nanopore Technologies) ont été développés (**Schneider, et al,2012**),(**Eisenstein M,2012**)

L'entreprise a annoncé son premier produit, le MinION, en 2012),(**Eisenstein M,2012**) . sa commercialisation débute en 2014. Ces nanopores ont la capacité d'émettre un signal électrique spécifique lorsqu'ils se trouvent au contact de certaines molécules. L'ADN présenté à ces nanopores imbriqués dans des membranes hybrides polymèreslipides, les traversent et, à son passage, émettent un signal électrique spécifique analysé par un séquenceur.

La technique ne nécessite que peu ou pas de préparation de l'ADN et ne connaît virtuellement pas de limite de longueur de lecture (**Eisenstein M,2012**).

Les différentes bases peuvent être déterminés en temps réel grâce au courant qui traverse le pore, avec une grande précision 99,8% et à faible coût puisqu'aucun système optique n'est nécessaire ainsi peu de traitement des données (**Branton D,2008**). Ces systèmes sont de très petite taille : par exemple, le MinION est une unité de séquençage par nanopore à usage unique intégré dans une clef USB. Cette technique est rapide, ne nécessite pas de réactifs ni d'amplification de l'ADN.

### 3.3. Pacific Biosciences

Pacific Biosciences a développé le premier instrument capable de séquencer une molécule unique en temps réel, là encore par lecture directe et sans amplifier le matériel de départ : il s'agit du PacBio RS. Il utilise une structure composée de cellules SMRT (Single Molecule Real Time) (**Shendure J. H Ji ,2008**) (**Ahmadian,et al ,2006**).

Chacune de ces cellules contient 75000 nanostructures appelées détecteurs ZMW (Zero-Mode Waveguide) de 100 nm de diamètre, c'est-à-dire plus petits que les longueurs d'onde utilisées sur la plateforme (532 et 643 nm). La lumière ne peut donc pas s'y propager, d'où le terme de mode-zéro. Chacun de ces ZMW contient une polymérase qui y est immobilisée et qui incorpore des nucléotides liés à un fluorophore, libérant ainsi celui-ci à l'extérieur du ZMW et mettant fin au signal lumineux. L'intervalle de temps entre chaque pic ainsi que la durée de chaque pic de fluorescence sont propres à chaque nucléotide et permettent ainsi leur identification. Ces paramètres sont également différents pour des nucléotides méthylés et il est donc possible, sans traitement préalable, de différencier une cytosine

d'une méthylcytosine ou encore d'une hydroxyméthylcytosine (**Ronaghi *et al* ,1998**) (**Margulies M, *et al* ,2005**).

La longueur de lecture moyenne est de plus de 1000 pb, la préparation de l'échantillon dure environ 30 minutes et les données sont produites en quelques minutes seulement, ce qui en fait une des technologies les plus prometteuses des années à venir.

### 3.4. Starlight :

Life Technologies développe un séquenceur de molécule unique appelé Starlight qui utilise des points quantiques qui sont des particules nanométriques à base de semi-conducteur dont les propriétés de fluorescence peuvent être contrôlées par leur taille. Excitées par un laser, ces particules transfèrent leur énergie aux nucléotides marqués par fluorescence lorsque ceux-ci sont incorporés par la polymérase (**Pennisi E,2010**).

De plus, il est possible de remplacer la polymérase lorsqu'elle s'essouffle. On peut ainsi théoriquement séquencer la longueur désirée de fragment.

### 4. Avantages et inconvénients du NGS :

**Tableau 2** : Avantages et faiblesses du NGS par rapport aux méthodes de séquençage «classique» (**Ilie M, Al, 2014**)

Avantages
<b>Très grande sensibilité analytique.</b>
<b>Séquençage de l'exome/génome complet avec une large gamme d'altérations génomiques détectées (délétions, translocations, nombre de copies).</b>
<b>Adapté aux différents types de prélèvements (tissu tumoral fixe au formol, ADN libre circulant, cellules tumorales circulantes).</b>
<b>Découverte de nouvelles variantes génétiques et une meilleure compréhension de l'oncogenèse et de la progression tumorale.</b>
<b>Délai d'analyse et de réponse diminué.</b>
Inconvénients
<b>Nécessité d'un consentement spécifique incluant l'analyse mutationnelle somatique et germinale</b>
<b>Variabilité de performance des différentes plateformes de NGS.</b>
<b>Coût par test élevé. Coût des équipements modérément élevé.</b>
<b>Expertise informatique pour l'interprétation et le traitement des données.</b>

**Nécessité d'un fort investissement institutionnel en terme de capacité de stockage des données.**

**Pertinence des nouveaux polymorphismes découverts.**

**Veille éthique quant à l'utilisation des données constitutionnelles (mutation germinales).**

## **5. application des NGS en agriculture**

### **Amélioration variétale du blé et sa culture :**

Le Consortium international de séquençage du génome du blé (IWGSC), dans lequel l'Inra occupe une position de leader, annonce la publication de la première séquence de référence du génome du blé tendre *Triticum aestivum*. Des équipes françaises de l'Inra, duCEA, du CNRS et des universités de Clermont-Auvergne, d'Evry Val d'Essonne, de Paris- Sud et de Paris-Saclay, ont contribué à ce véritable exploit scientifique en raison de la taille et de la complexité de ce génome, cinq fois plus gros que le génome humain et 40 fois plus gros que celui du riz. Ce résultat permettra notamment l'identification de gènes d'intérêt agronomique,

Ouvrant ainsi de nouvelles perspectives pour l'amélioration des variétés de blé et de sa culture, face aux défis planétaires à relever. C'est également une étape majeure au plan fondamental, pour comprendre le fonctionnement et l'évolution de ce génome complexe.

En effet, l'analyse de cette séquence a conduit, entre autres, à la localisation précise de plus de 107 000 gènes, parmi lesquels des gènes potentiellement impliqués dans la qualité du grain, la résistance aux maladies ou la tolérance à la sécheresse. Elle a également permis de développer plus de quatre millions de marqueurs moléculaires dont certains sont déjà utilisés dans des programmes de sélection. **(IWGSC ,2018).**

Cette séquence permet également de mieux comprendre ce génome qui compte parmi les plus grands et les plus complexes du règne végétal. Il est ainsi possible d'étudier l'organisation des gènes et la régulation de leur expression ou encore d'élucider les mécanismes évolutifs ayant façonné ce génome depuis sa formation, il y a environ 10 000 ans**(IWGSC ,2018).**

Le séquenceur Ion Proton <sup>TM</sup> est le premier séquenceur de nouvelle génération à proposer un séquençage rapide et abordable du génome et de l'exome chez les plantes.



## **Chapitre 3. Apport des biotechnologies dans la connaissance des bases génétiques des caractères d'intérêt**

### **1. Déterminer la fonction des gènes.**

La possibilité de séquencer leur génome constitue un progrès fondamental pour la compréhension du fonctionnement des plantes, que ce soit en conditions normales de culture ou en conditions de stress. Le séquençage d'un génome consiste en la lecture complète des séquences d'ADN qui le constitue, puis en l'identification de tous les gènes présents le long de ces séquences. La première espèce végétale dont le génome a été entièrement séquencé, en 2000, est l'Arabidopsis des dames (*Arabidopsis thaliana*). Le riz l'a rejointe en 2005, puis le peuplier, la tomate, le maïs (2009), la pomme de terre, et enfin la vigne. *Arabidopsis thaliana*, petite crucifère parente du colza, est ce que l'on appelle une espèce modèle. Elle est devenue le centre d'intérêt de nombreux laboratoires de recherche qui l'utilisent pour la compréhension des mécanismes à la base du fonctionnement des plantes, de leur métabolisme, ou encore de leurs réponses aux contraintes biotiques ou abiotiques. Suite au séquençage et à l'identification des séquences correspondant aux gènes dans un génome, les scientifiques ont maintenant pour objectif d'attribuer à chacun d'entre eux une fonction biologique, c'est-à-dire de déterminer à quoi sert la protéine qu'il code dans le fonctionnement de l'organisme. ( **Laure G et al,2010**).

L'intérêt de travailler sur des plantes modèles est d'avoir à disposition les outils génétiques et moléculaires les plus performants pour comprendre les mécanismes cellulaires et physiologiques à la base du fonctionnement des plantes. À plus long terme, les connaissances acquises peuvent être transférées à d'autres espèces végétales proches. *Arabidopsis thaliana* est, par exemple, un excellent modèle pour le colza, alors que le riz constitue un modèle pour la plupart des autres céréales. ( **Laure G et al,2010**).

### **2. Identifier les régions du génome favorables pour un caractère donné**

Grâce au séquençage, il est possible d'étiqueter les régions du génome des plantes leur conférant les caractères agronomiques souhaités. Cette technique repose sur l'identification de séquences d'ADN, appelées marqueurs moléculaires, qui ont des positions connues dans le génome, grâce à l'établissement préalable de cartes chromosomiques. Les marqueurs moléculaires permettent de déterminer la forme du gène, c'est-à-dire l'allèle, présente dans les individus étudiés. En couplant, chez un ensemble d'individus, une analyse génétique des marqueurs moléculaires et une analyse de leur phénotype (caractères observables et/ou mesurables), on va chercher à détecter les relations statistiquement fiables entre la présence d'un marqueur et la valeur du caractère mesuré (par exemple, le nombre de grains). Il sera alors possible de conclure à la proximité de certains

marqueurs moléculaires avec de petites régions du génome codant pour certains des caractères étudiés et appelés QTL (Quantitative Trait Locus).

Cette analyse reposant sur les différences entre individus pour les caractères étudiés (taille du grain, par exemple), elle va également permettre de déterminer si ces individus portent une forme du QTL plus ou moins favorable au caractère étudié.

Les QTL peuvent être identifiés grâce à l'analyse génétique d'individus issus de croisements choisis, mais aussi grâce à une approche appelée génétique d'association. La génétique d'association consiste à rechercher des corrélations entre des marqueurs moléculaires et le caractère d'intérêt dans une large collection de plantes représentant au mieux la diversité existant dans l'espèce (**Thornsberry et al., 2001**).

Il faut noter que l'identification et la localisation des QTL ne renseigne pas sur la fonction du ou des gènes présents à ces QTL. Mais elle peut être une étape préalable à l'identification de gènes candidats et aux études de génomique fonctionnelle. Comme nous le verrons plus loin, ces informations pourront d'autre part être utilisées en amélioration variétale grâce à la sélection assistée par marqueurs.

La détection de QTL pour l'amélioration d'un caractère donné est aujourd'hui une voie de recherche privilégiée par de nombreux laboratoires. C'est notamment le cas pour l'amélioration de la tolérance à la sécheresse. Les premiers travaux ont commencé chez le maïs il y a une vingtaine d'années mais que ce soit chez la plante modèle *Arabidopsis thaliana*, le riz, ou encore le blé, de nombreuses études sont en cours pour identifier les régions du génome et les allèles favorables à ce caractère. Pour exemple, citons l'identification, chez le maïs d'un QTL de teneur en ABA associé au développement des racines (**Tuberosa et al. 1998**), chez *Arabidopsis thaliana*, d'un QTL lié à une forte accumulation d'ABA dans les feuilles et associée à une augmentation de rendement en conditions de stress hydrique (**Landi et al., 2001**), ou encore chez le maïs, l'identification de QTL favorisant le maintien de la croissance foliaire même en cas de déficit hydrique (**Reymond et al., 2003**).

Alors que le séquençage du génome d'un seul individu d'une espèce donnée était, il y a quelques années encore, un exploit technologique, l'explosion des technologies de séquençage permet aujourd'hui d'avoir accès à la séquence du génome de nombreux individus par espèce, rapidement et à relativement faible coût. Accompagnée par le développement des méthodes de traitement de données par la bioinformatique, cette avancée renouvelle notre capacité d'exploration de la diversité génétique du vivant et notre compréhension de son fonctionnement.

Les gènes qui ont été sélectionnés au cours de la domestication des plantes cultivées et de leur adaptation à de nouveaux milieux présentent une diversité organisée différemment de celle des autres gènes. Le génome des organismes vivants garde ainsi une empreinte de leur histoire évolutive. En étudiant la diversité des plantes cultivées au niveau très fin qu'est celui des séquences d'ADN, il est possible de détecter cette empreinte.

La détection de zones du génome sélectionnées au cours de l'adaptation à des conditions environnementales particulières, et qui sont donc impliquées dans le déterminisme génétique de caractéristiques particulières, est une approche différente et complémentaire des approches de recherche de QTL, pour se rapprocher de l'identification des gènes réellement impliqués dans ces adaptations. .( **Laure G et al,2010**).

### **3. La sélection assistée par marqueurs permet de cibler et d'accélérer le travail de sélection**

L'identification de QTL favorables pour un caractère donné, et celle de gènes impliqués dans des réponses d'adaptation des plantes à une contrainte environnementale donnée sont autant d'outils biotechnologiques permettant aux sélectionneurs de cibler et d'accélérer les processus de sélection pour l'amélioration génétique des plantes. .( **Laure G et al,2010**).

La sélection assistée par marqueurs ajoute une évaluation génétique à l'évaluation phénotypique à la base du processus de sélection classique des plantes.

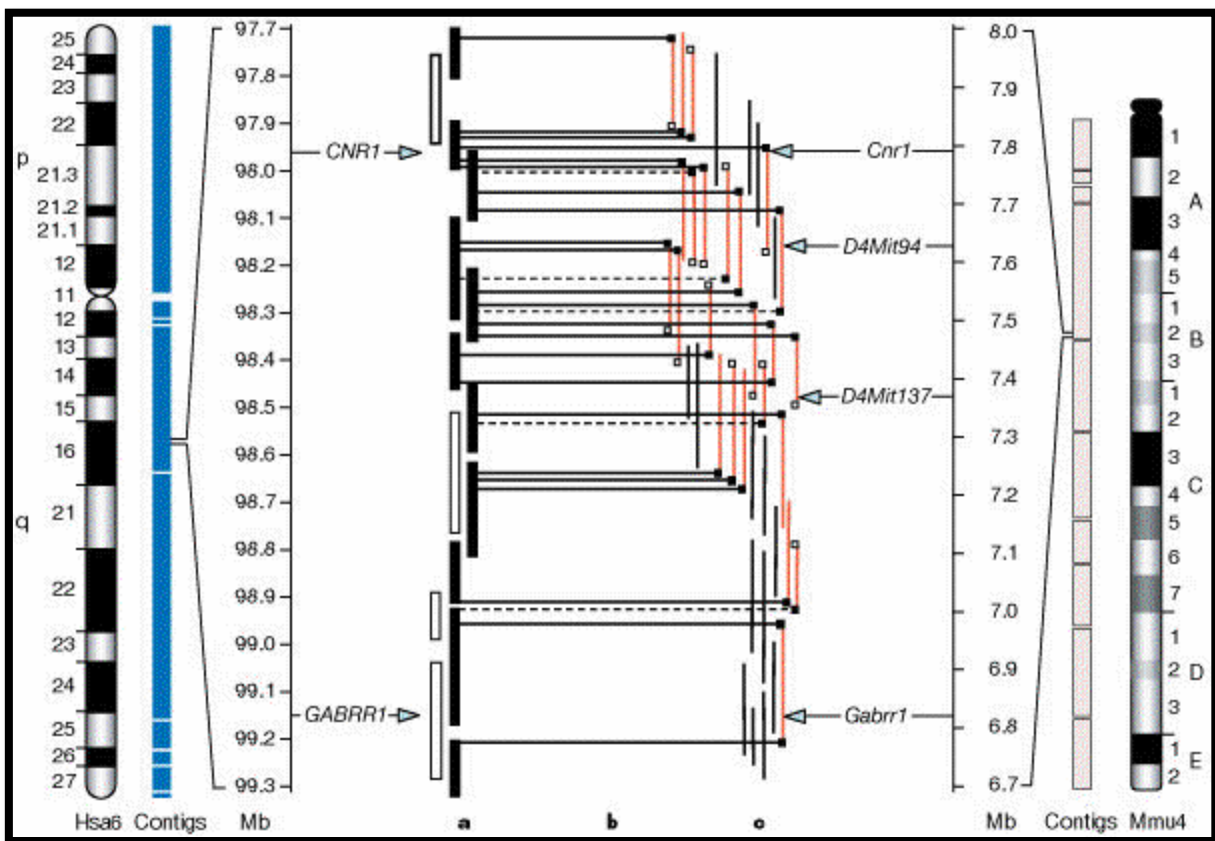
Elle s'appuie sur la variabilité génétique de l'espèce ciblée, et les travaux de détection de QTL, pour accélérer le processus de sélection en orientant de façon raisonnée le choix des génotypes qui combinent le maximum de caractères favorables. Lorsqu'un QTL a été détecté, l'allèle favorable pour le caractère étudié peut être transféré dans le génome d'une plante élite grâce à une succession de croisements.

À chaque étape, les individus ayant hérité de l'allèle favorable sont identifiés et sélectionnés grâce aux marqueurs moléculaires qui lui sont associés. Cette approche permet de réduire le nombre de croisements (de six à huit) à quatre. .( **Laure G et al,2010**).

#### 4. cartographie et comparaison entre espèces .

##### 4.1. L'histoire des études comparatives et intérêt des cartes génétiques

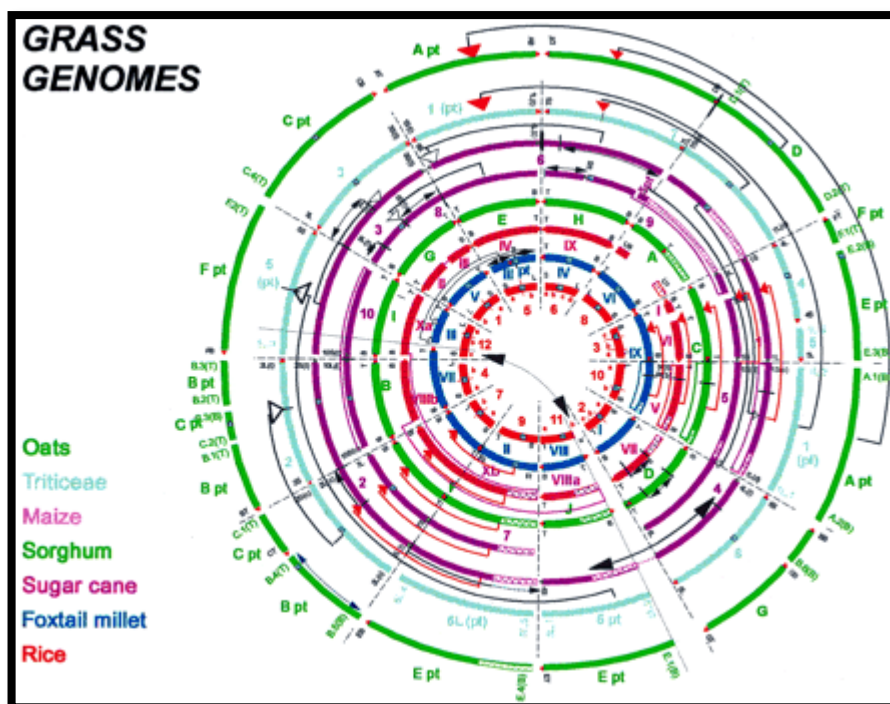
Les premières études de comparaison de génomes ont été faites chez les animaux en comparant des marqueurs moléculaires de cartes génétiques, de cartes d'hybrides de radiation et par hybridation sur chromosome (chromosome painting) entre le génome humain et des espèces proches, toutes étant des mammifères comme les rongeurs (Nadeau et Taylor 1984; Cheng *et al.* 1988), les ruminants (Echard *et al.* 1994; Gautier *et al.* 2002) ou le porc (Robic *et al.* 1999; 2001; Rink *et al.* 2002). Ces études ont révélé des conservations de blocs de marqueurs (macrosynténie) entre des espèces animales (Chowdhary *et al.* 1998; Gregory *et al.* 2002; Figure 29).



**Figure 25.** Aligment entre une partie du chromosome 6 de l'homme (*Hsa6*) et du chromosome 4 de la souris (*Mmu4*). (Gregory *et al.*, 2002)

Chez les plantes, les cartes génétiques d'un grand nombre d'espèces végétales ont été construites, ce qui a permis le développement de la cartographie génétique comparée à partir de la fin des années 1980 (Schmidt 2000). Les premières expériences de comparaison de cartes génétiques ont été faites sur des espèces de la famille des *Solanaceae* (*L. esculentum*, *S. tuberosum*) en utilisant des marqueurs RFLP dérivés de la tomate (Bonierbale *et al.* 1988; Gebhardt *et al.* 1991). L'ensemble de ces études de comparaison de cartes génétiques a montré qu'il existait une bonne macrosynténie entre les différentes espèces de *Solanaceae*, même si quelques remaniements ont lieu au sein des génomes.

Depuis, d'autres études ont été faites sur d'autres familles notamment chez les *Brassicaceae* dont fait partie la plante modèle *A. thaliana*. La comparaison entre *A. thaliana* et des espèces de *Brassica* a montré que les macro et microcolinéarités sont très conservées entre les *Brassicaceae* et avec *Arabidopsis* (Cavel *et al.* 1998; Acarkan *et al.* 2000). Chez le blé, Chao *et al.* (1989) ont montré la colinéarité entre les 3 génomes A, B et D du blé hexaploïde. Hulbert *et al.* (1990) ont ensuite démontré que les sondes développées chez le maïs pouvaient s'hybrider avec le sorgho, le millet et la canne à sucre. Ils ont ainsi produit le premier alignement des groupes de liaison entre maïs et sorgho mettant en évidence la similarité des 2 génomes. Cette conservation de l'ordre des gènes dans certaines régions chromosomiques ressort également des comparaisons entre blé et riz (Kurata *et al.* 1994), blé, orge et seigle (Devos *et al.* 1993), sorgho et maïs (Pereira *et al.* 1994), sorgho et canne à sucre (Ming *et al.* 1998), riz et sorgho



**Figure 26.** Une carte comparative consensus des graminées de Gale et Devos (1998)

(Ventelon *et al.* 2001). Gale et Devos (1998) ont pu ainsi aligner les génomes de plusieurs espèces de monocotylédones; le blé, le maïs, le riz, l'avoine, le sorgho, la canne à sucre et le millet (Figure 29).

#### 4.2. Evolution de la structure des génomes

La première application de la cartographie comparée est l'étude de l'évolution de la structure des chromosomes. Ainsi la comparaison des cartes génétiques de la tomate et la pomme de terre a mis en évidence l'homologie des 12 groupes de liaison et la conservation de l'ordre des marqueurs.

Cinq inversions paracentriques différencient les 2 espèces ont été mises en évidence (**Bonierbale et al. 1988**). De même, la comparaison entre les cartes génétiques du poivron, de la tomate, de la pomme de terre et de l'aubergine a permis d'identifier des mécanismes d'évolution du génome de cette famille (**Livingstone et al. 1999; Doganlar et al. 2002; Figure 9**). Une trentaine de cassures, comprenant des translocations, des inversions paracentriques et péricentriques, serait suffisantes pour expliquer les différences entre les 4 génomes.

Aussi la comparaison des cartes génétiques entre la tomate et le poivron a mis en évidence outre la conservation de larges régions colinéaires, des réarrangements chromosomiques. Ainsi des inversions paracentriques apparaissent comme le mécanisme premier d'évolution des chromosomes dans la famille des *Solanaceae* et que ces changements structurels ne sont pas constants entre les différentes paires de *Solanaceae* étudiées (**Doganlar et al. 2002**).

La comparaison entre *Arabidopsis* et des espèces de la famille des *Brassicaceae* a permis de mettre en évidence une conservation de la microsyténie avec quelques remaniements, insertion / perte de gène ou duplication de gènes (**Acarkan et al. 2000**). Environ 90 réarrangements ont été révélés entre *A. thaliana* et *B. nigra* avec des fragments conservés qui ont une longueur moyenne de 8 cM (**Lagercrantz 1998**).

#### **4.3. Connaissance sur l'évolution des génomes**

La cartographie génétique apporte des éclaircissements importants sur la structure, l'organisation et sur le mode et la vitesse d'évolution des génomes de plantes (**Schmidt, 2000**), il devient de plus en plus clair que ce n'est que la comparaison directe des séquences d'ADN de génomes entiers qui pourra élucider les processus de l'évolution de ceux-ci et le rôle joué par les réarrangements chromosomiques dans la spéciation (**M Sankoff et Nadeau 2003**). L'étude de la microsyténie requiert les séquences génomiques et leur annotation pour comparer les différents génomes. De nombreux programmes de séquençage de génomes ont été réalisés ou sont en cours afin de permettre les comparaisons de la microsyténie entre espèces. Le séquençage du génome de la levure (**Goffieau et al. 1996**), de la drosophile (**Adams et al. 2000**), de *Caenorhabditis elegans* (**The C. elegans Sequencing Consortium, 1998**), du fugu (**Taylor et al. 2002**), de l'homme (**International Human Genome Sequencing Consortium 2001**) et d'*A. thaliana* (**Lin et al. 1999**) a été achevé, mais d'autres organismes sont en cours de séquençage (*M. truncatula*, le riz, le maïs, la tomate, le caféier etc.) ainsi qu'un nombre croissant de programmes de génomique d'autres espèces animales et végétales (séquençages d'ARNm, de séquences EST, de fragments d'ADN génomique). Le séquençage de génomes complets et de fragments de génomes (notamment de clones BAC) a permis la comparaison de cartes physiques entre différentes espèces aussi bien animales que végétales.

## 5 .Amélioration de la cartographie génétique et physique

L'exploitation la plus fréquente aujourd'hui des cartes alignées est l'identification de marqueurs d'une variété d'espèces pour une région génomique donnée. C'est particulièrement utile pour des expériences de cartographie fine ou de clonage positionnel. Ainsi par exemple, la relation de colinéarité entre *Lotus japonicus*, *Pisum sativum* et *Arabidopsis thaliana* a été établie et exploitée pour améliorer la cartographie physique d'un gène impliqué dans la formation de nodules de fixation d'azote chez les légumineuses. Après détermination de la région synténique, un marqueur AUX-1 d'*Arabidopsis* a pu être cartographié proche de la région cible chez *Lotus japonicus* (**Stracke et al. 2004**). De même la microsyténie établie entre la tomate et *Arabidopsis* a permis des améliorations des cartes génétiques et physiques. (**Ku et al. 2000**), décrit le premier travail rapportant l'exploitation de microsyténie avec *Arabidopsis* pour des fins de clonage positionnel au sein d'une famille différente au sein du règne végétal. Cette étude a permis de construire une carte génétique à haute résolution pour la région du chromosome 2 de la tomate, co-segregeant avec la mutation ovate. Toujours entre

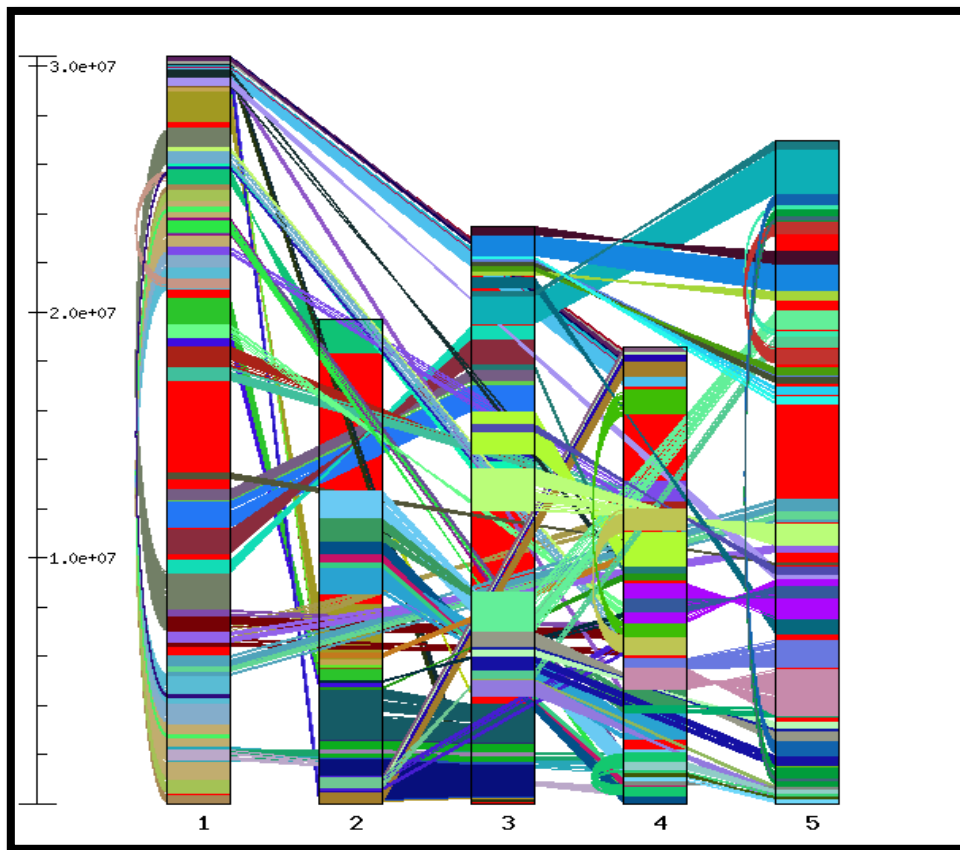
*Arabidopsis* et la tomate, la microsyténie a permis de réduire la distance entre les marqueurs encadrant le gène cible *Diageotropia (Dgt)* de 0.8 à 0.15 cM (**Oh et al. 2002**). Pareillement une importante zone synténique entre le chromosome 1 du riz et le chromosome 3 de l'orge (*Hordeum vulgare L.*), a permis de réaliser la saturation de la carte génétique contenant le gène de résistance de la rouille des feuilles de la région chromosomique de l'orge. Cette étude montre aussi clairement l'utilité des ressources génomiques du riz pour la recherche efficace d'EST orthologue de l'orge à des fins de saturation de la région génomique cible chez l'orge (**Mammadov et al. 2005**).

## 6. Clonage d'un gène par utilisation de l'orthologie

L'identification d'orthologues putatifs pour des traits monogéniques ou quantitatifs couvrant différentes espèces est facilitée par les comparaisons de cartes (**Paterson et al. 1995**). Ceci a été documenté d'une manière élégante par la cartographie et le clonage des homologues du gène d'*Arabidopsis GAI* qui codent pour un modulateur de la réponse à la gibbérelline chez le maïs, le riz et le blé. Les gènes du blé et du maïs correspondant aux locus *Dwarf* (**Peng et al. 1999**). De même, le gène récessif de résistance contre les potyvirus chez la tomate (*pot-1*) a pu être cartographié et isolé grâce à la colinéarité établie entre la tomate (*L. esculentum*) et le piment (*Capsicum spp.*) et l'utilisation de l'orthologue du facteur d'initiation eukaryotique 4E (eIF4E) du Tabac. Cette démarche éclaire l'utilité de comparaison des cartes pour l'identification de la séquence orthologue du gène récessif de résistance qui a débouché sur une approche gène-candidat. Ce résultat va à contre courant de l'idée d'une absence de gène



de résistance dans les régions synténiques. Ceci peut être expliqué par une évolution plus lente des gènes de résistance contre les potyvirus par rapport à la majorité des gènes de résistance clonés jusqu'à aujourd'hui, et peut donc être considéré comme une classe à part de gènes de résistance (Parrella *et al.* 2002).



**Figure 27.** Segments dupliqués d'*Arabidopsis*

[http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis\\_genome\\_duplication.shtml](http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis_genome_duplication.shtml)

### 6.1 De nouvelles plantes modèles pour la synténie

La disponibilité de la séquence complète du génome d'*Arabidopsis* a permis de comparer le génome de cette plante aux séquences de fragments d'ADN génomiques partiels de différentes espèces végétales. Par exemple entre *Arabidopsis* et des espèces de la famille des *Poaceae* (Mayer *et al.* 2001) ou des espèces appartenant à d'autres familles comme avec les *Fabaceae* (Grant *et al.* 2000) et les *Solanaceae* (Ku *et al.* 2000). La macro- et microsyténie sont plus ou moins fortement conservées en fonction des temps de divergence entre les espèces comparées, et même si quelques réarrangements peuvent avoir eu lieu, le transfert d'information d'un organisme à un autre est toujours possible (Schmidt 2000; Bancroft 2001; Barnes 2002). Ainsi les paires les plus étudiées sont *Arabidopsis*-pomme-de-terre (Dominguez *et al.* 2003), *Arabidopsis*-betterave (Dominguez *et*



*al. 2003*), *Arabidopsis*-tournesol (**Dominguez et al. 2003**), *Arabidopsis-Prunus* (**Dominguez et al. 2003**), *Arabidopsis*-coton (**Bowers et al. 2003**) et *Arabidopsis*-tomate (**Ku et al. 2000**)

## **6.2 Les défis de la génomique comparative**

Un des défis de la génomique comparative est de distinguer les orthologues (c.-à-d. les gènes homologues avec un ancêtre commun qui ont été séparés par un événement de spéciation) des paralogues (c.-à-d. homologues dérivant d'un événement de duplication de gène); ce qui est particulièrement problématique dans de grandes familles de gènes telles que les gènes de résistance. Un deuxième problème concerne la possibilité d'identifier des segments conservés à l'aide de la cartographie génétique. (**MAHÉ L,2007**)

Les familles multigéniques et les pseudogènes, par exemple, peuvent parfois rendre difficile l'identification du bon gène orthologue, surtout si les espèces comparées sont éloignées phylogénétiquement. En effet dans ce cas, il est difficile de distinguer l'érosion de l'identité de séquence due à la spéciation de celle due à la duplication. De plus, la distance entre deux marqueurs sur une carte génétique correspond à un pourcentage de recombinaison et ne reflète pas du tout le contenu en gènes et la séquence d'ADN de cette région du génome. Entre 2 marqueurs peuvent se cacher des centaines de gènes, et la macrosynténie entre deux espèces n'implique pas forcément l'existence de conservation des gènes et de leur ordre (microsynténie). (**MAHÉ L,2007**)

## **Chapitre 4 séquençage du génome du blé( défis et véritable exploit scientifique).**

### **Introduction :**

Le Consortium international de séquençage du génome du blé (« International Wheat Genome Sequencing Consortium » – IWGSC) a publié aujourd’hui dans le journal scientifique international *Science* une description détaillée de la séquence du génome du blé tendre.

Ce travail va permettre d’accélérer le développement de variétés de blé plus durables, mieux adaptées aux changements climatiques, ayant de meilleurs rendements et qualités nutritionnelles. L’article scientifique qui a plus de 200 auteurs travaillant dans 73 entités de recherche dans 20 pays présente la séquence de référence du blé tendre de la variété *Chinese Spring*. La séquence d’ADN de chacun des 21 chromosomes est de haute qualité, la meilleure produite à ce jour pour le blé.

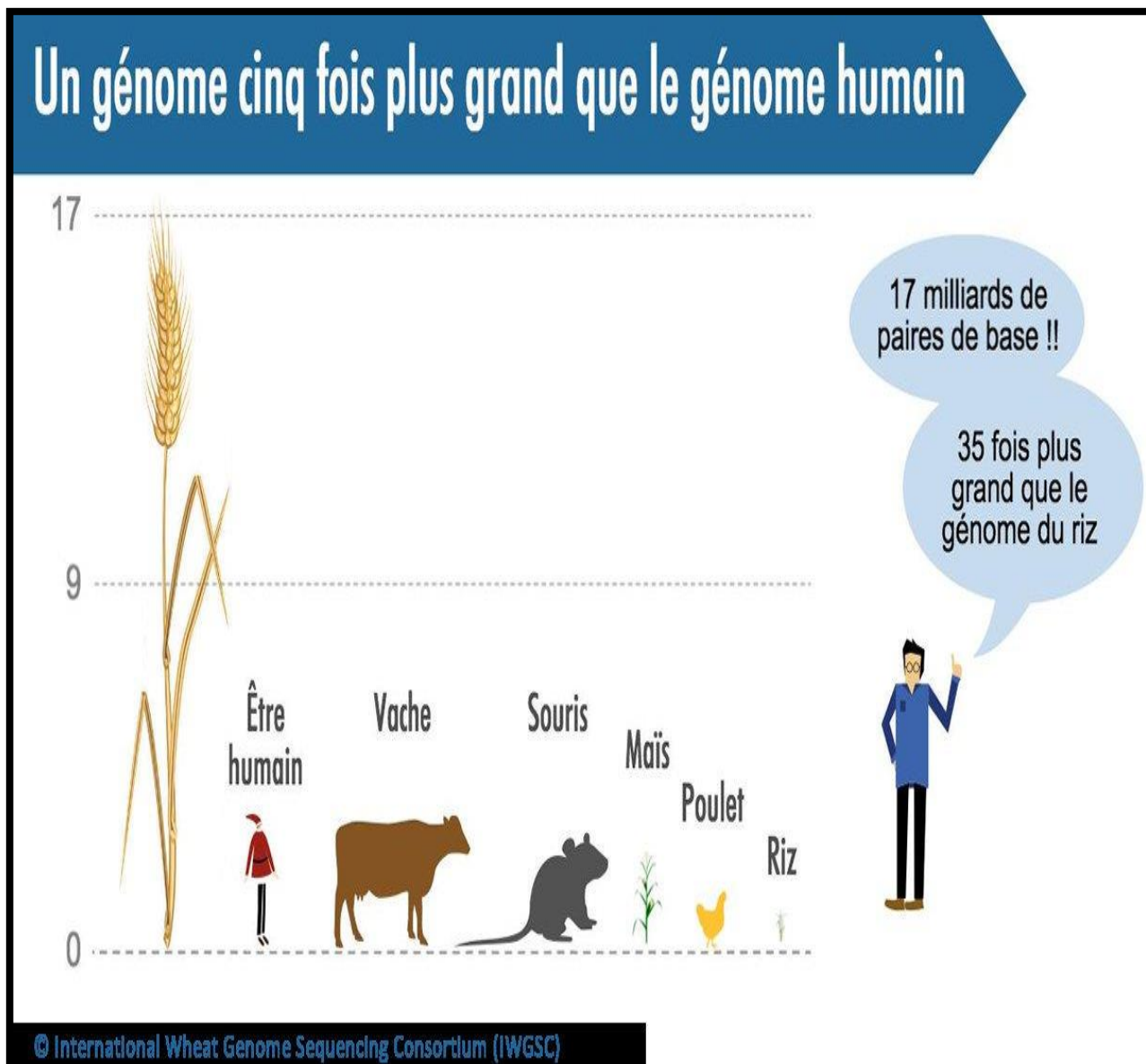
Le blé tendre est l’aliment de base de plus d’un tiers de la population mondiale et apporte près de 20% des calories et protéines consommées. Il est aussi une source importante de vitamines et de minéraux.

Pour couvrir les besoins de la population mondiale, qui atteindra of 9.6 millions en 2050, la production de blé doit augmenter de 1,6% par an, tout en préservant la biodiversité, et les ressources en eau et en nutriments. La majorité de cette augmentation doit donc être réalisée par amélioration génétique, sans augmentation des surfaces cultivées. Maintenant que la séquence de référence est connue, les sélectionneurs ont de nouveaux outils à leur disposition pour répondre à ces défis. Ils vont pouvoir identifier plus rapidement les gènes ou les éléments régulateurs responsables de caractères agronomiques complexes tels que le rendement, la qualité des grains, la résistance aux maladies fongiques, et la tolérance aux stress biotiques et abiotiques afin de produire des variétés de blé plus robustes et résistantes.

« La séquence du génome du blé nous permet d’avoir accès au moteur du blé » explique Rudi Appels, professeur à l’Université de Melbourne et chercheur a AgriBio en Australie. « Ce que nous voyons est extrêmement bien fait, permettant des variations et des adaptations à différents environnements, mais est aussi suffisamment stable pour maintenir des structures de base afin que la plante puisse survivre sous différentes conditions climatiques. »

Comme cela a été le cas lorsque les séquences de référence du maïs et du riz ont été produites, il est prévu que la disponibilité de la séquence de référence du blé va permettre d’accélérer l’amélioration variétale au cours des prochaines décennies.

Séquencer le génome du blé tendre a longtemps été considéré une tâche impossible, à cause de sa taille cinq fois plus grand que le génome humain et de sa complexité le blé tendre est constitué de trois sous-génomes et 85% sont des éléments répétés.(figure 28).



**Figure 28** génome du blé est cinq fois plus grand que le génome humaine .

« La publication de la séquence de référence du génome du blé tendre est le résultat du travail de nombreux chercheurs qui se sont regroupés sous la houlette de l’IWGSC pour s’atteler à ce qui était considéré impossible à faire », explique Kellye Eversole, la directrice exécutive de l’IWGSC.

« La méthode utilisée pour produire la séquence de référence et les principes et pratiques du consortium constituent un modèle pour séquencer des génomes de plantes complexes et de large taille et réaffirment l’importance des collaborations internationales pour contribuer à la sécurité alimentaire ».

L’impact de la séquence de référence du blé tendre est déjà important dans la communauté scientifique, comme l’atteste la publication le même jour de six autres articles scientifiques décrivant ou utilisant la séquence : un article paraissant dans le même numéro de *Science*, un autre

dans *Science Advances* et quatre autres dans *Genome Biology*. De plus, plus de 100 articles scientifiques citant la séquence de référence ont été publiés depuis janvier 2007, date à laquelle les données ont été mise à disposition de la communauté scientifique.

En plus de la séquence des 21 chromosomes, l'article de *Science* décrit la position précise de 107,891 gènes et de plus de 4 millions marqueurs moléculaires, ainsi que les séquences d'ADN entre ces gènes et marqueurs qui contiennent des éléments régulateurs agissant sur l'expression des gènes.

L'IWGSC a obtenu ce résultat en combinant toutes les ressources qu'il a générées au cours des treize dernières années en utilisant les méthodes classiques de construction de cartes physiques, ainsi que les technologies de séquençages d'ADN les plus récentes. Les séquences d'ADN ont été assemblées et ordonnées le long des 21 chromosomes en utilisant des algorithmes performants et les gènes ont été identifiés grâce à des programmes bioinformatiques dédiés.

Toutes les ressources générées par l'IWGSC sont publiques, accessibles à tous, sur la plateforme de l'Unité de recherche en Génomique-Info (URGI) de l'INRA, ainsi que dans les autres banques de données scientifiques internationales telle GrainGenes and Ensembl Plant.

L'article de *Science* est intitulé « Shifting the limits in wheat research and breeding using a fully annotated reference genome » (« Dépasser les limites en recherche et sélection du blé en utilisant un génome de référence complètement annoté »).

L'IWGSC, avec 2,400 membres dans 68 pays, est un consortium international créé en 2005 à l'initiative de chercheurs en sciences végétales et de sélectionneurs publics et privés. L'objectif de l'IWGSC est de produire et de rendre publique une séquence complète du blé tendre d'excellente qualité, ressource indispensable pour la recherche fondamentale et pour l'amélioration variétale du blé.

#### **Méthodes :**

Séquençage du génome entier du cultivar Chinese Printemps par Courte lecture séquençage par synthèse fourni les données pour l'assemblage de génome de novo et d'échafaudage avec le logiciel DenovoMAGIC2. L'assemblée a été superscaffolded et ancré dans 21 pseudomolécules avec haute densité génétique (POPSEQ) et physique (Hi-C et 21 cartes physiques spécifiques aux chromosomes) les informations cartographiques et en intégrant des ressources génomiques.

Validation de l'assemblage utilisation génétique indépendante (génotypage de novo par- séquençage des cartes) et des données de cartographie physique (cartes hybrides de radiation, bionano-optiques " maps " pour les chromosomes homéologues du groupe 7).

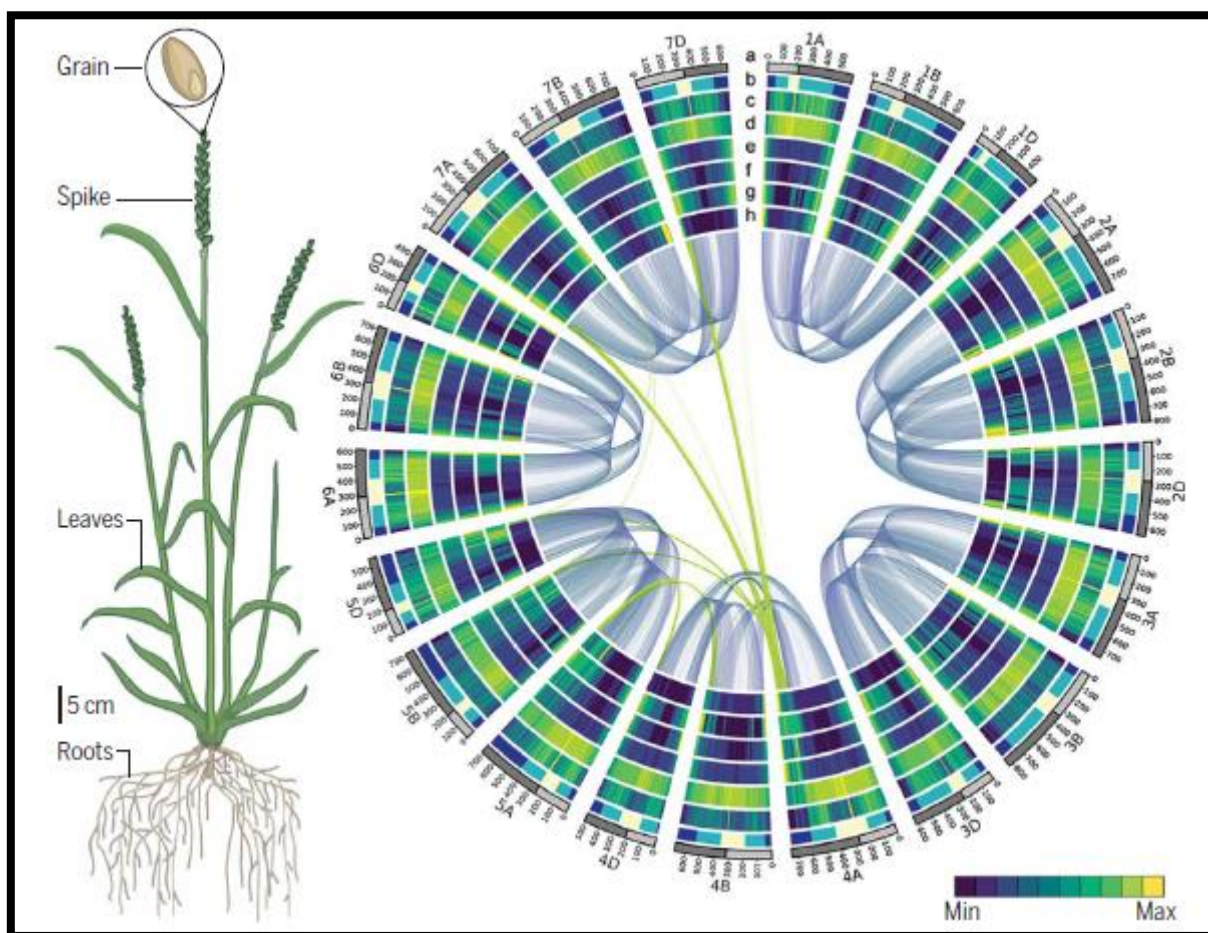
L'assemblage du génome a été annoté pour les gènes, d'ADN répétitif, Validation de l'assemblage utilisation génétique indépendante (génotypage de novo par- séquençage des cartes) et des données de cartographie physique (cartes hybrides de radiation, bionano-optiques " maps " pour les chromosomes homéologues du groupe 7).

L'assemblage du génome a été annoté pour les gènes, d'ADN répétitif, Un atlas de la transcription des gènes de blé a été construit à partir d'un vaste panneau de 850 indépendants transcriptome data set et était alors utilisé pour étudier les réseaux de co-expression génique. Outre, l'assemblée a été utilisé pour la dissection d'une tige importante-solidness QTL et concevoir des cibles pour l'édition du génome des gènes impliqué dans le contrôle du temps de floraison chez le blé.

### Résultats :

#### 1. Assemblage à l'échelle chromosomique du génome du blé

Des séquences de pseudomolécules représentant le 21 les chromosomes du génome du blé tendre étaient assemblé en intégrant un projet de tout le génome de novo assemblage (WGA), construit à partir d'Illumina shortread séquences utilisant NRGene de NovoMagic2 (Fig. 24 Tableau 3).



**Figure 29:** Génome du blé déchiffré, assemblé et commandé.

Les graines, ou grains, sont ce qui compte avec respect des rendements de blé (panneau de gauche), mais toutes les parties de la plante contribuent aux performances de la culture. Avec accès complet à la séquence ordonnée des 21 chromosomes du blé, le contexte de la réglementation séquences et le réseau d'interaction des gènes exprimés - tous présentés ici sous forme de diagramme circulaire (à droite). panneau) avec des pistes concentriques pour divers aspects de la composition du génome du blé. les sélectionneurs et les chercheurs ont maintenant la possibilité de réécrire l'histoire de l'amélioration des cultures de blé. (fig. 24).

dans l' assemblage du génome de 14,5 Gb résultant, contigs et échafaudages avec N50 de 52 kb et 7 Mo, respectivement, étaient liés en super échafaudages (N50 = 22,8 Mo), avec 97% (14,1 Go) des séquences attribué et ordonné le long des 21 chromosomes et presque tous les échafaudages de séquence assignés et orientés les uns par rapport aux autres (13,8 Go, 98%).

la nouvelle taille du génome des estimations de 15,4 à 15,8 Go peuvent être comptabilisées par des séquences effondrées ou non assemblées de grappes très répétées, telles que l'ARN ribosomal régions codantes et séquences télomériques. Une caractéristique clé qui distingue le RefSeq IWGSC v1.0 des assemblées de blé préliminaires précédentes (**International Wheat Genome Sequencing Consortium (IWGSC),,2014**).( **A. V. Zimin et al.,2017**) l'organisation à long terme, avec 90% des génome représenté dans les super-échafaudages plus grands que 4,1 Mo et avec chaque chromosome représenté, en moyenne, seulement 76 super-échafaudages (tableau 3).

**Tableau 3.** Statistiques d'assemblage d'IWGSC RefSeq v 1.0.

<b>Assembly characteristics</b>	<b>Values</b>
Assembly size	14.5 Gb
Number of scaffolds	138,665
Size of assembly in scaffolds $\geq$ 100 kb	14.2 Gb
Number of scaffolds $\geq$ 100 kb	4,443
N50 contig length	51.8 kb
Contig L50 number	81,427
N90 contig length	11.7 kb
Contig L90 number	294,934
Largest contig	580.5 kb
Ns in contigs	0
N50 scaffold length	7.0 Mb
Scaffold L50 number	571
N90 scaffold length	1.2 Mb
Scaffold L90 number	2,390
Largest scaffold	45.8 Mb
Ns in scaffolds	261.9 Mb
Gaps filled with BAC sequences	183 (1.7 Mb)
Average size of inserted BAC sequence	9.5 kb
N50 superscaffold length	22.8 Mb
Superscaffold L50 number	166
N90 superscaffold length	4.1 Mb
Superscaffold L90 number	718
Largest superscaffold	165.9 Mb
Sequence assigned to chromosomes	14.1 Gb (96.8%)
Sequence $\geq$ 100 kb assigned to chromosomes	14.1 Gb (99.1%)
Number of superscaffolds on chromosomes	1,601
Number of oriented superscaffolds	1,243
Length of oriented sequence	13.8 Gb (95%)
Length of oriented sequence $\geq$ 100 kb	13.8 Gb (97.3%)
Smallest number of superscaffolds per subgenome chromosome	35 (7A), 68 (2B), 36 (1D)
Largest number of superscaffolds per subgenome chromosome	111 (4A), 176 (3B), 90 (3D)
Average number of superscaffolds per chromosome	76

## 2.La composition du génome du blé

Analyses des composants de la séquence du génome révélé la distribution des éléments clés et a permis des comparaisons détaillées de l'homologue Sous-génomes A, B et D. Comptabilité pour 85% du génome, avec une distribution relativement égale dans les trois sous-génomes (tableau 4).



**Tableau4 :** Proportion relative des principaux éléments du génome du blé. Proportions de TE sont exprimés en pourcentage de séquences attribuées à chaque superfamille par rapport à la taille du génome.

Major elements	Wheat subgenome			
	AA	BB	DD	Total
Assembled sequence assigned to chromosomes (Gb)	4.935	5.180	3.951	14.066
Size of TE-related sequences (Gb)	4.240	4.388	3.285	11.913
TEs (%)	85.9	84.7	83.1	84.7
Class 1				
LTR-retrotransposons				
Gypsy (RLG)	50.8	46.8	41.4	46.7
Copia (RLC)	17.4	16.2	16.3	16.7
Unclassified LTR-retrotransposons (RLX)	2.6	3.5	3.7	3.2
Non-LTR-retrotransposons				
Long interspersed nuclear elements (RIX)	0.81	0.96	0.93	0.90
Short interspersed nuclear elements (SIX)	0.01	0.01	0.01	0.01
Class 2				
DNA transposons				
CACTA (DTC)	12.8	15.5	19.0	15.5
Mutator (DTM)	0.30	0.38	0.48	0.38
Unclassified with terminal inverted repeats	0.21	0.20	0.22	0.21
Harbinger (DTH)	0.15	0.16	0.18	0.16
Mariner (DTT)	0.14	0.16	0.17	0.16
Unclassified class 2	0.05	0.08	0.05	0.06
hAT (DTA)	0.01	0.01	0.01	0.01
Helitrons (DHH)	0.0046	0.0044	0.0036	0.0042
Unclassified repeats	0.55	0.85	0.63	0.68
Coding DNA	0.89	0.89	1.11	0.95
Unannotated DNA	13.2	14.4	15.7	14.4
(Pre)-microRNAs	0.039	0.057	0.046	0.047
tRNAs	0.0056	0.0050	0.0068	0.0057

Analyse d'un sous-ensemble de 181 036 gènes [“filtré ensemble de gènes » tableau 3] comprenant 103 757 Gènes HC et 77 279 LC identifiés 39 238 groupes homéologues, c'est-à-dire des clades de A, B et Orthologues du sous-génome déduits des arbres génétiques contenant un total de 113.653 gènes (63% des ensemble filtré). Pertes génétiques ou rétention et gène gains (duplications de gènes) ont été déterminés pour tous les locus homologues de IWGSC RefSeq v1.0 (Tableau 3).

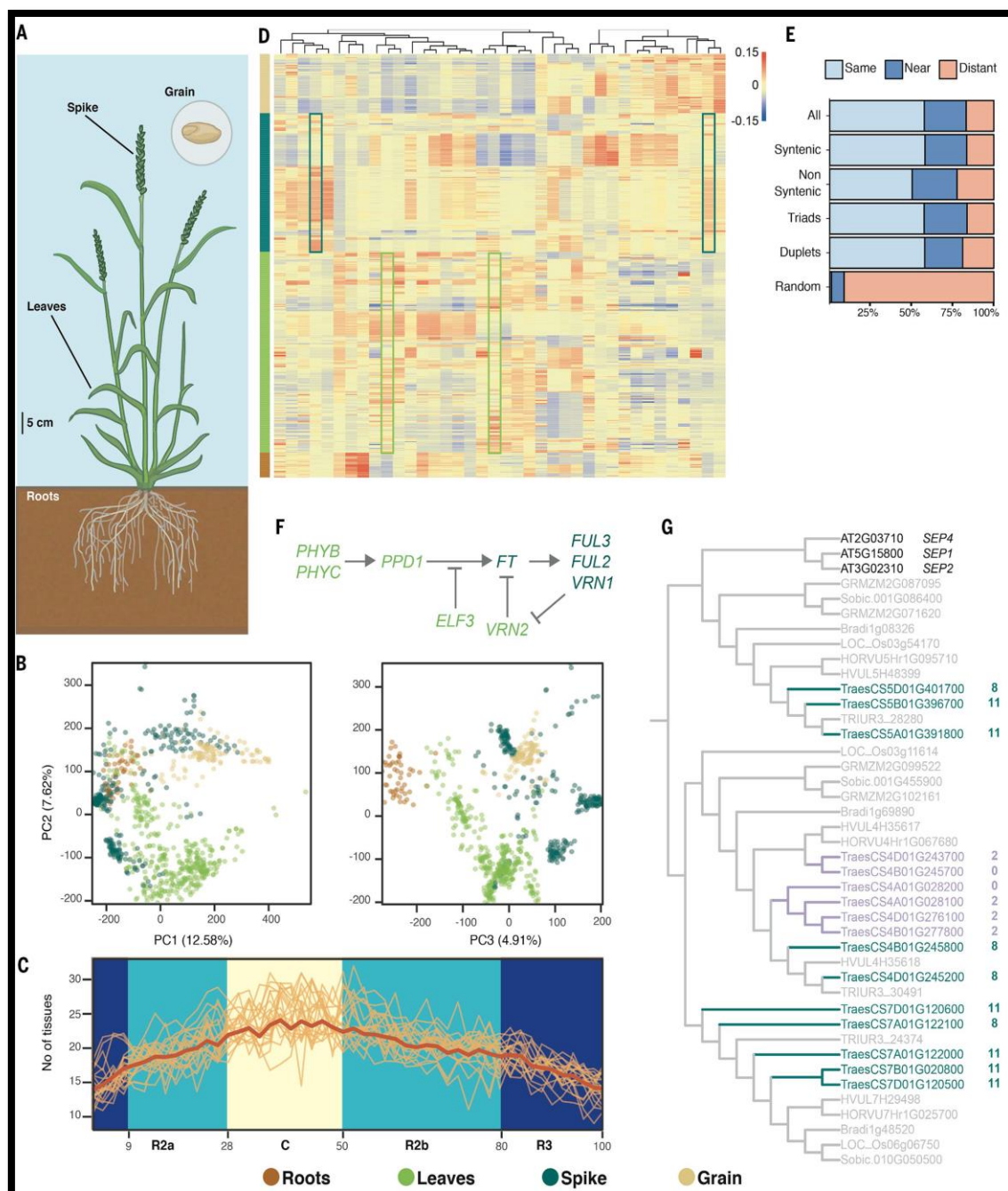


**Tableau 5:** Groupes de gènes homéologues chez le blé. Les gènes homéologues sont «Orthologues du sous-génome» et ont été déduits du rapprochement des espèces dans la famille de gènes respective.

Homeologous group (A:B:D)	Number in wheat genome	Composition of groups (%)	Number of genes in A	Number of genes in B	Number of genes in D	Total number of genes
1:1:1	21,603	55.1	21,603	21,603	21,603	64,809
1:1:N	644	1.6	644	644	1,482	2,770
1:N:1	998	2.5	998	2,396	998	4,392
N:1:1	761	1.9	1,752	761	761	3,274
1:1:0	3,708	9.5	3,708	3,708	0	7,416
1:0:1	4,057	10.3	4,057	0	4,057	8,114
0:1:1	4,197	10.7	0	4,197	4,197	8,394
Other ratios	3,270	8.3	4,999	5,371	4,114	14,484
1:1:1 in microsynteny	18,595	47.4	18,595	18,595	18,595	55,785
Total in microsynteny	30,339	77.3	27,240	27,063	28,005	82,308
1:1:1 in macrosynteny	19,701	50.2	19,701	19,701	19,701	59,103
Total in macrosynteny	32,591	83.1	29,064	30,615	30,553	90,232
<b>Total in homeologous groups</b>	<b>39,238</b>	<b>100.0</b>	<b>37,761</b>	<b>38,680</b>	<b>37,212</b>	<b>113,653</b>
Conserved subgenome orphans			12,412	12,987	10,844	36,243
Nonconserved subgenome singletons			10,084	12,185	8,679	30,948
Nonconserved subgenome duplicated orphans			71	83	38	192
<b>Total (filtered)</b>			<b>60,328</b>	<b>63,935</b>	<b>56,773</b>	<b>181,036</b>

en supposant la présence d'une seule copie du gène à chaque locus homéologue (appelé "triade"). Le pourcentage de gènes dans les groupes homéologues pour toutes les configurations (ratios) est très similaire, donc équilibré entre les trois sous-génomes: 63% (A), 61% (B) et 66% (D).

### 3. Atlas de transcription révèle trait associée à réseaux de Co régulation des gènes



**Figure 30.** Atlas de transcription du blé

- (A) Illustration schématique et définitions de la plante du blé et de tissu de haut niveau «racines», «feuilles», «épi» et «Grain» utilisé dans l'analyse ultérieure.
- (B) Analyse en composantes principales (PC) parcelles pour la similarité de la transcription globale, avec des échantillons colorés selon le tissu d'origine de niveau élevé [introduit dans (A)].

La clé de couleur pour le tissu est indiqué au bas de la figure sous (C).

(C) distribution chromosomique de largeur d'expression moyenne [nombre de tissus dans lesquels les gènes sont exprimés (nombre total de tissus,  $n = 32$ )]. Ligne orange) est calculée sur la base d'une position mise à l'échelle de chaque gène au sein de le compartiment génomique correspondant (bleu, aqua et jaune clair arrière-plan) sur les 21 chromosomes (lignes orange).

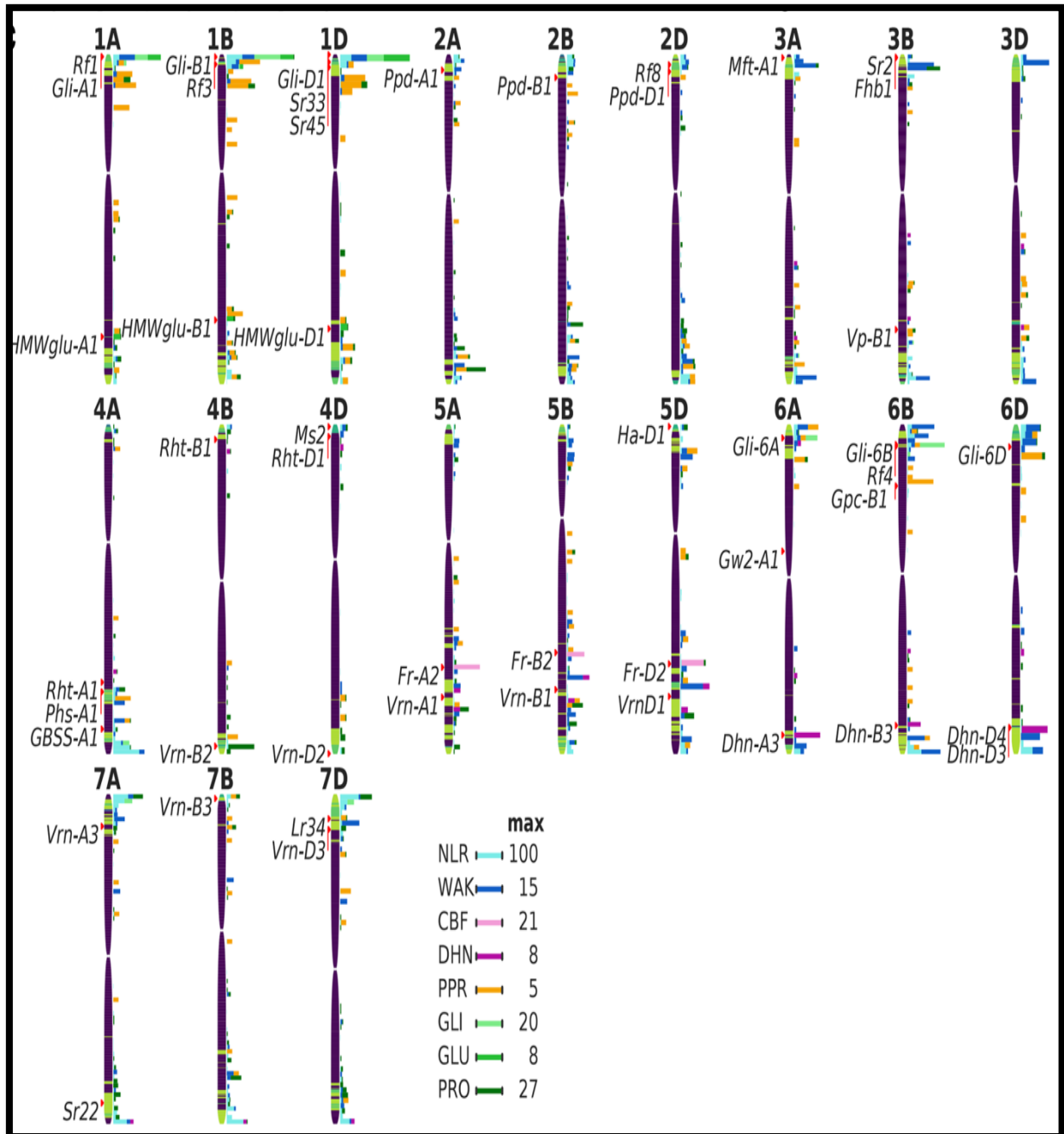
(D) Carte thermique illustrant l'expression d'un gène représentatif ( pour la 38 modules de coexpression définis par WGCNA. Les modules sont représentés comme colonnes, avec le dendrogramme illustrant la relation entre eigengene. Chaque rangée représente un échantillon. Les barres colorées à gauche indiquent le tissu de haut niveau d'origine; la clé de couleur est indiquée au bas de la figure sous

(C). DESeq2- les niveaux d'expression normalisés sont indiqués. Modules 1 et 5 (cases vert clair) étaient le plus en corrélation avec les tissus foliaires de haut niveau, alors que les modules 8 et 11 (boîtes vert foncé) étaient les plus en corrélation avec le pic.

(E) Barre graphique du module assignation (identique, proche ou distante) de triades homéologues et de duplication dans Réseau WGCNA.

(F) voie de floraison simplifiée chez le blé polyploïde. Les gènes sont colorés en fonction de leur affectation à la feuille (vert clair) - ou à l'épi (vert foncé) - modules corrélés.

(G) Extrait d'arbre phylogénétique pour Facteurs de transcription MADS, Les branches vertes représentent les orthologues de blé des modules 8 et 11, alors que les branches pourpres sont des orthologues de blé affectés à d'autres modules (0 et 2). Les branches grises indiquent des gènes non liés au blé.



**Figure31 :** Distribution génomique de familles de gènes associées à une adaptation au stress biotique (bleu clair et foncé) ou abiotique (rose clair et foncé), au métabolisme de l'ARN dans les organites et à la fertilité masculine (orange) ou de qualité finale (vert clair, moyen et vert foncé).

Les positions connues des gènes et des locus importants sur le plan agronomique sont indiquées par des flèches et des points flèches rouges à gauche des barres chromosomiques. Les taux de recombinaison sont affichés sous forme de cartes thermiques dans les barres chromosomiques [7,2 cM / Mb (vert clair) à 0 cM / Mb (noir)].

#### **4.Amélioration rapide des caractères en utilisant des marqueurs résolus physiquement et édition du génome.**

La sélection et la modification de la variation génétique traits agronomiques sous-jacents en sélection programmes est souvent compliqué si phénotypique la sélection dépend de l'expression de plusieurs locus avec des effets quantitatifs qui peuvent être fortement influencé par l'environnement.

Le potentiel de RefSeq IWGSC v1.0, avec l'annotation détaillée du génome, accélérer l'identification des gènes candidats sous-jacents importants agronomiques les traits ont été illustrés pour deux cibles. L'approche génétique a été utilisée pour résoudre complètement un QTL pour la solidité de la tige conférant une résistance à la sécheresse et aux dommages causés par les insectes **.(K. T. Nilsen *et al.*,2017)**

La connaissance des espèces modèles peut également être utilisé pour annoter les gènes et fournir une voie de l'amélioration des traits par la génétique inverse.

#### **Conclusion**

Travaux publics et services gouvernementaux Canada RefSeq v1.0 est une ressource qui a potentiel d'innovation perturbatrice dans l'amélioration du blé. Par nécessité, les éleveurs travaillent avec le génome au niveau du chromosome entier, tel que chaque nouvelle crois implique la modification de les réseaux géniques à l'échelle du génome qui contrôlent l'expression de traits complexes tels que le rendement avec le génome de référence annoté et ordonné séquence en place, les chercheurs et les éleveurs peuvent maintenant, accédez facilement à l'information au niveau de la séquence définir les changements dans les génomes des lignées programme. Bien que plusieurs centaines de QTL de blé ont été publiés, seul un petit nombre de les gènes ont été clonés et caractérisées.

Travaux publics et services gouvernementaux Canada RefSeq v1.0 souliers immédiats application en fournissant l'accès à la réglementation les régions, et il servira de base aux ancrer tous les QTL connus à un commun annoté référence.

Combiner ces connaissances avec la distribution de fréquence de recombinaison méiotique et la diversité génomique permettra aux éleveurs pour relever plus efficacement les défis imposés par la nécessité d'équilibrer le parallèle de sélection processus d'adaptation aux conditions biotiques et abiotiques stress, qualité à l'utilisation finale et amélioration du rendement.

Les stratégies peuvent maintenant être définies plus précisément amener les allèles souhaitables en phase de couplage, en particulier dans les régions moins-recombinantes de la génome du blé.

Chapitre4..... séquençage du génome du blé (défis et véritable exploit scientifique).

Ici, le potentiel de la l'information Nouvellement disponible sur le génome pourrait être réalisé à travers la mise en œuvre de DNAmarker plates-formes et technologies d'élevage ciblées, y compris l'édition du génome (**H. Puchta ; 2017**).

## CONCLUSION

Dix mille ans après que l'homme a commencé à cultiver le blé, le génome du blé tendre, ou froment (*Triticum aestivum*), vient d'être décrit et analysé en détail. Ce résultat a été obtenu par le Consortium international de séquençage du génome du blé (IWGSC). Il représente l'aboutissement d'un travail colossal commencé il y a 13 ans.

Le génome du blé est en effet d'une taille et d'une complexité considérables : avec environ 15,5 milliards de paires de bases et plus de 107 000 gènes identifiés, il est environ 40 fois plus grand que celui du riz et 5 fois plus que celui de l'homme (qui contient, lui, environ 3,2 milliards de paires de bases).

Cela s'explique en partie par le fait que le blé est une plante polyploïde : il contient trois génomes, conservés lors du croisement de deux espèces d'abord, puis avec une troisième ensuite.

Le génome du blé tendre est ainsi constitué de trois fois sept paires de chromosomes.

De nombreux gènes peuvent être trouvés en plus d'un exemplaire. Mais ce n'est pas tout : plus de 85 % du génome du blé est constitué de séquences répétées.

le génome du blé est l'une des dernières espèces d'intérêt agronomique majeur dont la séquence du génome n'est pas disponible.

Au terme de ce travail, nous avons obtenu une quantité importante d'informations sur les efforts fournies par le consortium de séquençage appelé *International Wheat Genome Sequencing Consortium* (IWGSC) impliquant des acteurs publics et privés de différentes parties du monde (<http://www.wheatgenome.org/>), afin de disposer de la première séquence du génome du blé tendre. En effet, l'amélioration des techniques de génomique de haut débit par l'utilisation de séquenceurs de nouvelle génération permettent de finaliser le séquençage du génome du blé.

Afin de disposer de la première séquence du génome du blé tendre, les scientifiques ont établi un séquençage de chacun des bras chromosomiques purifiés par cytométrie de flux et aujourd'hui tous les bras chromosomiques sont séquencés.

C'est l'aliment de base d'un tiers de la population mondiale. Grâce au décryptage du génome du blé, les chercheurs espèrent identifier des gènes potentiellement impliqués dans la qualité des grains, la résistance aux maladies ou la tolérance à la sécheresse.

## Références bibliographiques

**Moule C., 1980.** Les céréales. Ed. La maison rustique. Paris. 318 p.

**Prevost ph., 1976.** Génétique. Ed. Lavoisier. Paris, 299 p.

**Boyeldieu J., 1980.** Les cultures céréalières. Ed Hachette. France.

**Simon H., Codaccion P. et Lecoeur X., 1989.** produire des céréales à paille. Agriculture d'aujourd'hui. Ed. Lavoisier. Paris, 346 p.

**Grignac P., 1978.** Le blé dur : Techniques agricoles. Tome I, 6-10.

**Erroux J. et Laumont P., 1961.** Mémoire de la société d'histoire naturelle de l'Afrique du nord, 5, 95 p.

**Dekhil S., (1998) :** Les principales maladies cryptogamiques des feuilles du blé dans le Nord-Est de l'Algérie. 75 pp

**Bousnane M., (2003) :** Etude des effets de la mouche de Hesse sur le blé. Mémoire d'ingénieur. ISN. Univ. Constantine. 39 p.

**Messiaen C. M., (1981) :** les variétés résistantes ; Méthode de lutte contre les maladies et ennemis des plantes. INRA. Eds. Louis-Jean. ISBN 2-85340-355-6. 333 pp.

**FAO (2015)** Perspectives de récolte et situation alimentaire 1 : 7p.

**ONFAA (2016)** Observation national des filières Agricoles et Agroalimentaires. Le commerce international des céréales 7 : 1-2  
onfaa.inraa.dz.

**Charvet JP (2012)** Claire Levasseur. Atlas de l'agriculture: 14p.

**Terrones Gavira F et Burny P., (2012) :** Evolution du marché mondial du blé au cours des cinquante dernières années. In "Céréales". ULg Gembloux Agro-Bio Tech et CRA-W Gembloux. PP 9.

**Abis S (2012)** Le blé en Méditerranée sociétés, commerce et stratégies. Économie et territoire relations commerciales CIHEAM Paris : 241-247.

**FAO (2014)** Afrique classement des pays producteurs de matières premières : 2p.

**Hales N, Rush C (2016)** Algeria Grain and Feed Annual 9: 1-11.

**Bogard M., (2011) :** Analyse génétique et écophysiological de l'écart à la relation teneur en protéines - rendement en grains chez le blé tendre (*Triticum aestivum* L.)

**Terrones Gavira F et Burny P., (2012) :** Evolution du marché mondial du blé au cours des cinquante dernières années. In "Céréales". ULg Gembloux Agro-Bio Tech et CRA-W Gembloux. PP 9.



**Abis S (2012)** Le blé en Méditerranée sociétés, commerce et stratégies. Économie et territoire relations commerciales CIHEAM Paris : 241-247.

**Hales N, Rush C (2016)** Algeria Grain and Feed Annual 9: 1-11.

**FAO (2014)** Afrique classement des pays producteurs de matières premières : 2p.

**Bogard M., (2011)** : Analyse génétique et écophysologique de l'écart à la relation teneur en protéines - rendement en grains chez le blé tendre (*Triticum aestivum* L.)

**Hamel L., (2010)** : Appréciation de la variabilité génétique des blés durs et des blés apparentés par les marqueurs biochimiques, 08 JUIL 2010''.

**Levesque H., Madre J-F., (2013)** : Le blé, une plante domestiquée au génome polyploïde complexe, 31 OCT 2014''.

**Salamé N., (2012)** : Hybridation et polyplôidisation : l'exemple du blé , 18 AVR 2012''.

**Sanger F, Nicklen S , Coulson AR** . DNA sequencing with chain-terminating inhibitors . Proc Natl Acad Sci USA 1977: 5463-7.

**Piton N, Lamy A, Sabourin J C.** Séquençage des tumeurs : évolutions et révolutions. Cancer/Radiothérapie 2017 :580-3.

**Schneider, G F, C Dekker.** DNA sequencing with nanopores. Nat Biotechnol 2012:

**Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al.** Initial sequencing and analysis of the human genome. Nature 2001:860—921. 326-8.

**Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al.** The sequence of the human genome. Science 2001:1304—51.

**Lamoril J, N Ameziane, M Bogard.,** Les techniques de séquençage de l'ADN : une révolution en marche. Première partie. Immuno-analyse et Biologie Spécialisée 2008: 260-79.

**Aghrouch Nada( 2018)** séquençage de l'ADN principe ,indication et application

**IWGSC** . Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 2018.

**A Debourgogne, Delhaes S, Houzé F, Morio C.** Chapitre 15 - Techniques de séquençage. Parasitologie et Mycologie Médicales - Guide des Analyses et des Pratiques Diagnostiques 2017:209-11

**Bai X, Edwards J , Ju J.** Molecular engineering approaches for DNA sequencing and analysis. Expert Rev Mol Diagn 2005: 797-808.

**Metzker M L.** Sequencing technologies - the next generation. Nat Rev Genet 2010: 31-46.

**Harris S R, Feil E J, Holden M T, Quail M A, Nickerson E K, Chantratita N, Gardete S, Tavares A, Day N, Lindsay J A.** Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 2010 :469-74.

**Kingsmore SP , Saunders CJ.** Deep sequencing of patient genomes for disease diagnostic : When will it become routine. *Sci Transl Med* 2011: 23

**Mardis E R.** A decade's perspective on DNA sequencing technology . *Nature* 2011:198-203

**Shendure J. H Ji.**Next generation DNA sequencing. *Nat Biotechnol* 2008:1135-45.

**Woollard P.M.** The application of next generation sequencing technologies to drug discovery and development. *Drug Discov Today* 2011: 512-9.

**Desai A.N. A Jere.** Next generation sequencing: ready for the clinics?.*Clin Genet* 2012: 503-10.

**Human genome:**Genomes by the thousand. *Nature* 2010: 1026-7

**Maxam, A M,W Gilbert,** A new method for sequencing DNA. *Proc Natl Acad Sci U S A* 1977: 560-4.

**F. Sanger, S Nicklen et A R Coulson.** DNA sequencing with chain-terminating inhibitors , *Proc Nat Acad Sci USA* 1977 : 5463-7

**A Debourgogne, Delhaes S, Houzé F, Morio C.** Chapitre 15 - Techniques de séquençage. *Parasitologie et Mycologie Médicales - Guide des Analyses et des Pratiques Diagnostiques* 2017:209-11

**Bai X, Edwards J , Ju J.** Molecular engineering approaches for DNA seqencing and analysis. *Expert Rev Mol diagn* 2005: 797-808.

**Loman , N J.** Performance comparison of benchtop high throughput sequencing platforms. *Nat Biotechnol* 2012: 434-9.

**Orlando, L, Al .** True single molecule DNA sequencing of a pleistocene horse bone. *Genome Res* 2011: 1705-19.

**Kasianowicz J, Brandin E, Branton D, Deamer DW.** Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci USA* 1996: 13770-3.

**Eisenstein M .** Oxford Nanopore announcement sets sequencing sector abuzz. *Nat Biotechnol* 2012: 295-6.

**Eisenstein M .** Miniature wireless sensors presage smart phone medicine. *Nat Biotechnol* 2012: 1013-4.

**Branton D, Deamer D W, Marziali A , Bayley H , Benner S A , Butler T , Huang X.**the potential and challenges of nanopore sequencing. *Nat Biotechnol* 2008:1146-53.

**Ahmadian, M Ehn, S Hober.** Pyrosequencing: history, biochemistry and future. Clin Chim Acta 2006 : 83-94.

**Ronaghi , M Uhlen , P Nyren .** A sequencing method based on real time pyrophosphate. Science 1998 : 363-5.

**Margulies M, Egholm M, Altman W E, Attiya S, Bader J S, Berka J.** Genome sequencing in microfabricated high density picolitre reactors . Nature 2005 :376-80

**Pennisi E .** Genomics Semiconductors inspire new sequencing technologies. Science 2010: 1190.

**Ilie M, Al .** Les méthodes de séquençage de « nouvelle génération » (NGS) et le cancer bronchopulmonaire: principales technologies, applications et limites actuelles en pathologie. Revue Francophone des Laboratoires 2014: 51-8.

**Laure G et al,2010**

**Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D., Buckler E.S., 2001.** Dwarf polymorphisms associate with variation in flowering time. Nature Genetics, 28, 286-289.

**Tuberosa R., Sanguineti M.C., Landi P., Salvi S., Casarini E., Conti S., 1998.** RFLP mapping of quantitative trait loci controlling abscisic acid concentration in leaves of drought-stressed maize (*Zea mays* L.). Theor Appl Genet, 97, 744-755.

**Reymond M., Muller B., Leonardi A., Charcosset A., Tardieu F., 2003.** Combining quantitative trait loci analysis and an ecophysiological model to analyse the genetic variability of the responses of leaf growth to temperature and water deficit. Plant Physiology, 131, 664-675.

**Landi P., Sanguineti M.C., Conti S., Tuberosa R., 2001.** Direct and correlated responses to divergent selection for leaf abscisic acid concentration in two maize populations. Crop Science, 41, 335-344.

**Nadeau J.H. and Taylor B.A. (1984).** Lengths of chromosomal segments conserved since divergence of man and mouse. Proceedings of the National Academy of Sciences (USA), **81**: 814-818

**Cheng S.V., Nadeau J.H., Tanzi R.E., Watkins P.C., Jagadesh J., Taylor B.A., Haines J.L., Sacchi N. and Gusella J.F. (1988).** Comparative mapping of DNA markers from the familial Alzheimer disease and Down syndrome regions of human chromosome 21 to mouse chromosomes 16 and 17. Proceedings of the National Academy of Sciences (USA), **85**: 6032-6036.

**Echard G., Broad T.E., Hill D. and Pearce P. (1994).** Present status of ovine gene map (*Ovis aries*); comparison with the bovine map (*Bos Taurus*). Mamm. Genome, **5**: 324-332.

**Gautier M., Hayes H. and Eggen A. (2002).** An extensive and comprehensive radiation hybrid map of bovine Chromosome 15 : comparison with human Chromosome 11.

Mammalian Genome, **13**: 316-319.

**Robic A., Jeon J.T., Amarger V., Chardon P., Looft C., Andersson L., Gellin J. and Milan D.** (2001). Construction of a high-resolution RH map of the human 2q35 region on TNG panel and comparison with a physical map of the porcine homologous region 15q25. Mammalian Genome, **12**: 380-386.

**Rink A., Santschi M.E., Eyer K.M., Roelofs B., Hess M., Godfrey M., Karajusuf E.K., Yerle M., Milan D. and Beatie C.W.** (2002). A first-generation EST RH comparative map of the porcine and human genome. Mammalian Genome, **13**: 578-587.

**Chowdhary B.P., Raudsepp T., Frönicke L. and Scherthan H.** (1998). Emerging Patterns of Comparative Genome Organization in Some Mammalian Species as Revealed by Zoo-FISH. Genome Research, **8**: 577-589.

**Gregory S.G., Sekhon M., Schein J., Zhao S., Osoegawa K., Scott C.E., Evans R.S., Burrige P.W., Cox T.V., Fox C.A. et al.** (2002). A physical map of the mouse genome. Nature, **418**: 743-750.

**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.

**Schmidt T.** (2000). Synteny : recent advances and future prospects. Current Opinion in Plant Biology, **3**: 97-102.

**Bonierbale M.W., Plaisted R.L. and Tanksley S.D.** (1988). RFLP maps based on a common set clones reveal modes of chromosomal evolution in potato and tomato. Genetics, **120**: 1095-1103.

**Gebhardt C., Ritter E., Barone A., Debener T., Walkemeier B., Schachtschnabel U., Kaufmann H., Thompson R.D., Bonierbale M.W., Ganai M.W., et al.** (1991). RFLP maps of potato and their alignment with the homoeologous tomato genome. Theoretical and Applied Genetics, **83**: 49-57.

**Cavell A.C., Lydiate D.J., Parkin I.A.P., Dean C. and Trick M.** (1998) Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. Genome, **41**: 62-69.

**Acarkan A., Rossberg M., Koch M. and Schmidt R.** (2000). Comparative genome analysis reveals extensive conservation of genome organisation for *Arabidopsis thaliana* and *Capsella rubella*. The Plant Journal, **23**: 55-62.

**Kurata N., Moore G., Nagamura Y., Foote T., Yano M., Minobe Y. and Gale M.D.** (1994). Conservation of Genome Structure Between Rice and Wheat. Bio-Technology, **12**: 276-278.

**Devos K.M., Atkinson M.D., Chinoy C.N., Liu C.J. and Gale M.D.** (1993). RFLP-based genetic map of the homoeologous group 2 chromosomes of wheat, rye, and barley. Theoretical and Applied Genetics, **85**: 784-792.

**Ming R., Liu S.C., Lin Y.R., da Silva J., Wilson W., Braga D., van Deynze A., Wensla**

**T.F., Wu K.K., Moore P.H., Burnquist W., Sorrells M.E., Irvine J.E. and Paterson A.H.** (1998). Detailed alignment of saccharum and sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. *Genetics*, **150**: 1663-1682.

**Pereira M.G., Lee M., Bramel-Cox P., Woodman W., Doebley J. and Whitkus R.** (1994). Construction of an RFLP map in sorghum and comparative mapping in maize. *Genome*, **37**: 236-243.

**Ventelon M., Deu M., Garsmeur O., Doligez A., Ghesquière A., Lorieux M., Rami J.F., Glaszmann J.C. and Grivet L.** (2001). A direct comparison between the genetic maps of sorghum and rice. *Theoretical and Applied Genetics*, **102**: 379-386.

**Bonierbale M.W., Plaisted R.L. and Tanksley S.D.** (1988). RFLP maps based on a common set clones reveal modes of chromosomal evolution in potato and tomato. *Genetics*, **120**: 1095-1103.

**Doganlar S., Frary A., Daunay M.C., Lester R.N. and Tanksley S.D.** (2002). A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the solanaceae. *Genetics*, **161**: 1697-1711.

**Lagercrantz U.** (1998). Comparative mapping between *Arabidopsis thaliana* and *brassica nigra* indicates that brassica genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics*, **150**: 1217-1228.

**Schmidt T.** (2000). Synteny : recent advances and future prospects. *Current Opinion in Plant Biology*, **3**: 97-102.

**Goffeau A., Barrell B.G., Bussey H., Davis R. W., Dujon B., Feldmann H., Galibert F., Hoheisel J. D., Jacq C., Johnston M., et al.** (1996). Life with 6000 Genes. *Science*, **274**: 563-567.

**International Human Genome Sequencing Consortium** (2001). Initial sequencing and analysis of the human genome. *Nature*, **409**: 860-921.

**Lin X., Kaul S., Rounsley S., Shea T.P., Benito M.-I., et al.** (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature*, **402**: 761-772.

**Stracke S., Sato S., Sandal N., Koyama M., Kaneko T., Tabata S. and Parniske M.** (2004). Exploitation of colinear relationships between the genomes of *Lotus japonicus*, *Pisum sativum* and *Arabidopsis thaliana* , for positional cloning of a legume symbiosis gene. *Theoretical and Applied Genetics*, **108**: 442-449.

**Ku H-M., Vision T., Liu J. and Tanksley S.D.** (2000). Comparing sequenced segments of the tomato and *Arabidopsis* genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. *Proceedings of the National Academy of Sciences (USA)*, **97**: 9121-912.

**Oh KC., Hardeman K., Ivanchenko M.G., Ellard-Ivey M., Nebenführ A., White T.J. and Lomax T.L.** (2002). Fine mapping in tomato using microsynteny with the *Arabidopsis* genome: the Diageotropica (Dgt) locus. *Genome biology*, **3**: 1-11.

- Mammadov J.A., Steffenson B.J. and Saghai Maroof M.A.** (2005). High-resolution mapping of the barley leaf rust resistance gene Rph5 using barley expressed sequence. *Theoretical and Applied Genetics*, **111**: 1651-1660.
- Patocchi A., Vinatzer B.A., Gianfranceschi L., Tartarini S., Zhang H.B., Sansavini C. and Gessler C.** (1999). Construction of a 550 kb BAC contig spanning the genomic region containing the apple scab resistance gene *Vf*. *Molecular Genomics Genetics*, **262**: 884-891.
- Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M., Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F. et al.** (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature*, **400**: 256-261.
- Parrella G., Ruffel S., Moretti A., Morel C., Palloix A. and Caranta C.** (2002). Recessive resistance genes against potyviruses are localized in colinear genomic regions of the tomato (*Lycopersicon* spp.) and pepper (*Capsicum* spp.) genomes. *Theoretical and Applied Genetics*, **105**: 855-861.
- MAHÉ L, 2007**) Contribution à l'amélioration génétique de la résistance des caféiers (*Coffea arabica* L.) à la rouille (*Hemileia vastatrix*)
- Dominguez I., Graziano E., Gebhardt C., Baraka A., Berr S., Arus P., Delseny M. and Barnes S.** (2003). Plant genome archaeology: evidence for conserved ancestral chromosome segments in dicotyledonous plant species. *Plant Biotechnology Journal*, **1**: 91-100
- Bowers J.E., Chapman B.A., Rong J. and Paterson A.H.** (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, **422**: 433-438.
- Ku H-M., Vision T., Liu J. and Tanksley S.D.** (2000). Comparing sequenced segments of the tomato and *Arabidopsis* genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. *Proceedings of the National Academy of Sciences (USA)*, **97**: 9121-912.
- Grant D., Cregan P. and Shoemaker R.C.** (2000). Genome organization in dicots: Genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. *Proceedings of the National Academy of Sciences (USA)*, **97**: 4168-4173.
- Mayer K., Schüller C., Wambutt R., Murphy G., Volckaert G., et al.** (1999). Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature*, **402**: 769-780.
- K. T. Nilsen et al.**, High density mapping and haplotype analysis of the major stem-solidness locus SSt1 in durum and common wheat. *PLOS ONE* 12, e0175285 (2017).
- H. Puchta**, Applying CRISPR/Cas for genome engineering in plants: The best is yet to come. *Curr. Opin. Plant Biol.* 36, 1–8 (2017).



**Titre: séquençage : Intérêt et impact en amélioration des plantes-cas du blé-**

Mémoire de fin de cycle pour l'obtention du diplôme de Master en Biotechnologie et Génomique Végétale

**Résumé :**

Le blé, une espèce polyploïde, à trois grands génomes A, B et D constitue un génome complexe. Ce travail présente une recherche bibliographique sur la mise au point de la stratégie originale de séquençage du blé et les résultats obtenus dans le cadre du projet de cartographie du génome de blé, en utilisant les marqueurs moléculaires, les QTL à l'aide de marqueurs SNP à haut débit, banque des mutants, et de banque des EST, le projet de séquençage du génome de blé qui a été commencé en 2005 par la création de *Consortium International de Séquençage du Génome du Blé*. Une séquence de référence annotée représentant le génome du blé panifiable hexaploïde 21 pseudomolécules ont été analysées pour identifier la distribution et le contexte génomique des éléments codants et non codants dans les sous-génomes A, B et D. Avec une estimation de couverture de 94% du génome et contenant 107 891 modèles de gènes à haut niveau de confiance, cet assemblage a permis la découverte de tissus et de stades de développement réseaux de co-expression en fournissant un atlas des transcriptomes représentant les principales étapes du développement du blé, Le dynamique des familles de gènes complexes impliqués dans l'adaptation environnementale et la qualité d'utilisation finale ont été révélés à sous-génome résolution et contextualisée connus locus agronomiques à un seul gène ou à caractère quantitatif, Cette ressource communautaire établit la Fondation pour accélérer la recherche sur le blé et l'application de la compréhension du blé de la biologie et de génomique de la sélection assistée, L'IWGSC a obtenu ce résultat en combinant toutes les ressources qu'il a générées au cours des treize dernières années en utilisant les méthodes classiques de construction de cartes physiques, ainsi que les technologies de séquençages d'ADN les plus récentes, les séquences d'ADN ont été assemblées et ordonnées le long des 21 chromosomes en utilisant des algorithmes performants et les gènes ont été identifiés grâce à des programmes bio-informatiques dédiés.

**Mots clés :** blé, génome, séquençage, séquence, référence, chromosome.

**Jury d'évaluation :**

**Président(e) :** YKHLF .NADIA (professeur - UFM Constantine),  
**Encadrant :** BOUSBA .RATIBA (docteur - UFM Constantine),  
**Examinateur(rice) :** HAML A CHOROUK (docteur - UFM Constantine).

**Date de soutenance :** .17/07/2019