

الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne Démocratique et Populaire  
وزارة التعليم العالي والبحث العلمي  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

كلية علوم الطبيعة والحياة  
Faculté des Sciences de  
la Nature et de la Vie



جامعة الإخوة منتوري قسنطينة 1  
Université Frères Mentouri  
Constantine 1

## MEMOIRE DE FIN D'ÉTUDE

Réalisé en vue de l'obtention du diplôme de MASTER

Filière : Sciences Biologiques

Spécialité : Bioinformatique

Intitulé

**Analyse *in silico* de quatre facteurs de transcription exprimés chez *Triticum aestivum* sous conditions de stress biotique et abiotique.**

Soutenu le : 23/09/2021

Présenté par :

- TALBI HOURIA
- SENOUCI SOMIA

Jury d'évaluation :

Président : Dr. CHEHILI H. M.C. A. UFM Constantine 1

Encadreur : Dr. KELLOU K. M.A. A. UFM Constantine 1

Examineur : Dr. TEMAGHOULT M. M.A. A. UFM Constantine 1

Année universitaire 2020/2021

INTRODUCTION GENERALE .....	Erreur ! Signet non défini.
SYNTHESE BIBLIOGRAPHIQUE.....	2
Chapitre 1 : Le blé tendre.....	2
1.1. Histoire du blé.....	2
1.3. Morphologie du blé tendre .....	4
1.4. Utilisation du blé tendre, importances économiques et sociétales .....	5
1.5. Stress abiotique et développement des céréales.....	6
1.6. Effets des principaux stress abiotiques sur le développement et la croissance du blé .....	7
1.7. Réponses moléculaires et cellulaires des plantes aux stress abiotiques.....	8
1.8. Les fonctions régulatrices des facteurs de transcriptions en réponse aux stress abiotiques et biotiques.....	13
Chapitre 2: La recherche <i>in silico</i> .....	16
1.1. Définitions .....	16
2.1.1. Recherche <i>in silico</i> .....	16
2.1.2. La bioinformatique .....	17
2.1.3. La génomique.....	17
2.1.4. La phylogénie.....	18
2.1.5. L'annotation .....	19
1.2. Les différentes facettes de la bioinformatique.....	19
2.2.1. Compilation et organisation des données.....	19
2.2.2. Traitements systématiques des séquences (l'annotation des séquences).....	20
2.2.3. Bioinformatique et logiciels .....	20
1.3. Domaines de la bioinformatique .....	21
1.4. Etapes de travail en bioinformatique .....	22
Matériel et Méthodes .....	23
1. Méthode de création de la base de données .....	23
2. Méthode de recherche des données biologiques (gènes /protéine).....	25
3. Méthode de création le script python .....	28
4. Méthode de faire Alignement multiple.....	34
5. Méthode de création de site web .....	36
Résultat et discussion .....	39
Base de données complète.....	39
1. Les résultats du scripte .....	39
2. Les résultats de l'alignement.....	40
Interface du site web .....	47
Discussion .....	48
1-Facteurs de transcription NAC.....	50
2-Facteurs de transcription WRKY .....	50
3-Facteurs de transcription MYB .....	51
4-Facteurs de transcription bZIP .....	51
Conclusion.....	52
Référence bibliographique .....	54
Résumé .....	55

## LISTE DES FIGURES

<b>Figure 1</b> : Phylogénie du blé tendre (Salamani <i>et al.</i> , 2002).....	4
<b>Figure 2</b> : Relations phylogénétiques entre le blé tendre ( <i>T. aestivum</i> ) et ses ancêtres sauvages..	4
<b>Figure 3</b> : Phylogénie des blés (d’après Feldman, 2001).....	5
<b>Figure 4</b> : Détection du stress et signalisation ; un modèle de réponse aux stress abiotiques dans différents organites (tiré de Zhang <i>et al.</i> , 2018) .....	11
<b>Figure 5</b> : Réseau de régulation transcriptionnelle impliqué dans la réponse au stress thermique (tiré de Ohama <i>et al.</i> , 2017).....	14
<b>Figure 6</b> : Fonction des gènes dont l’expression est induite par un stress abiotique, dans la réponse et la tolérance à ce stress.....	15
<b>Figure 7</b> : Messagers secondaires et facteurs impliqués dans la perception et la transduction du signal dans la réponse au stress hydrique.....	16
<b>Figure 8</b> : Modèle schématique des voies de signalisation impliquées dans les réponses aux stress abiotiques et biotiques.....	16
<b>Figure 9</b> : Schéma général de la recherche <i>in silico</i> .....	18
<b>Figure 10</b> : Phylogénie du règne végétal.....	20
<b>Figure 11</b> : Exemple d’alignement de séquences de récepteurs nucléaires.....	21
<b>Figure 12</b> : Ressources en bioinformatique. (Modifiée à partir de Pevsner, 2015).....	22
<b>Figure 13</b> : Représentation schématique des procédés utilisés par la bioinformatique.....	23
<b>Figure 14</b> : Représentation schématique de l’importance des réseaux en bioinformatique.....	24
<b>Figure 15</b> : Créer une table de base de données “fafa”.....	27
<b>Figure 16</b> : Choix du type de données dans la table de la base de données "fafa".....	27
<b>Figure 17</b> : Recherche sur Google a banque de données genbank.....	28
<b>Figure 18</b> : Le gène NAM_B1 du facteur de transcription NAC de <i>Triticum aestivum</i> .....	28
<b>Figure 19</b> : Format fasta du gène NAM_B1 du facteur de transcription NAC de <i>Triticum aestivum</i> . .....	29
<b>Figure 20</b> : La protéine nam_b1 de <i>Triticum aestivum</i> .....	29
<b>Figure 21</b> : La séquence protéique nam_b1 de <i>Triticum aestivum</i> format fasta.....	30
<b>Figure 22</b> : La recherche de protéine par ID dans uniprot .....	30
<b>Figure 23</b> : Téléchargement de Format Fasta de la protéine.....	30
<b>Figure 24</b> : Séquence de la protéine WRKY d' <i>Oryza sativa</i> .....	31
<b>Figure 25</b> : Base de données complète.....	31
<b>Figure 26</b> : Fonction de démarrage d’écran .....	32
<b>Figure 27</b> : Les 4 variables forme lettre du couleur.....	32
<b>Figure 28</b> : Le code python de script base de données stade 1 .....	34
<b>Figure 29</b> : Le code python de script base de données stade 2.....	35
<b>Figure 30</b> : Le code python de script base de données stade 3.....	35
<b>Figure 31</b> : Le code python de script base de données stade 4.....	36

<b>Figure 32</b> : La recherche dans Google du programme Clustal omega.....	36
<b>Figure 33</b> : Interface du programme Clustal omega.....	37
<b>Figure 34</b> : La 1 <sup>ère</sup> étape de l'alignement multiple des séquences nucléaires.....	37
<b>Figure 35</b> : Résultats d'alignement multiple.....	38
<b>Figure 36</b> : Output format d'alignement multiple.....	38
<b>Figure 37</b> : Préparation des fichiers du projet.....	39
<b>Figure 38</b> : Le code HTML du site web de projet pour la structure.....	40
<b>Figure 39</b> : Le style globale (Code css+js ) de projet pour le DESSIN.....	40
<b>Figure 40</b> : Le style globale (Code css+js ) de projet pour le DESSIN.....	41
<b>Figure 41</b> : Base de données complète de 4 facteurs de transcription NAC, WRKY, bzip et MYB.....	43
<b>Figure 42</b> : Interface du script de la base de données.....	43
<b>Figure 43</b> : Interface du script de la base de données, famille de facteur transcription NAC..	43
<b>Figure 44</b> : Interface du script de la base de données du <i>Brachypodium</i> .....	44
<b>Figure 45</b> : Interface du script de la base de données commande non trouvé.....	44
<b>Figure 46</b> : Interface du script de la base de données commande non trouvé et retour au premier composé.....	44
<b>Figure 47</b> : Alignement multiple des gènes des familles NAC/WRKY/myb/b-zip chez <i>Triticum aestivum</i> , <i>Oryza sativa</i> et <i>Brachypodium</i> .....	45
<b>Figure 48</b> : Arbre phylogénétique.....	47
<b>Figure 49</b> : Alignement multiple des gènes des familles NAC/WRKY/myb/b-zip chez <i>Triticum aestivum</i> , <i>Oryza sativa</i> et <i>Brachypodium</i> .....	47
<b>Figure 50</b> : Arbre phylogénétique des facteurs de transcription NAC, WRKY, b-zip, myb.....	48
<b>Figure 51</b> : Alignement multiple protéique des gènes des familles NAC/WRKY chez <i>Triticum aestivum</i> et <i>Oryza sativa</i> .....	48
<b>Figure 52</b> : Cladogramme des facteurs de transcription NAC, WRKY.....	49
<b>Figure 53</b> : Alignement multiple des gènes des familles NAC/WRKY/myb du <i>Triticum aestivum</i> et d' <i>Oryza sativa</i> .....	49
<b>Figure 54</b> : Cladogramme des facteurs de transcription NAC, WRKY.....	50
<b>Figure 55</b> : Alignement multiple des gènes des familles NAC/WRKY/myb/b-zip entre <i>Triticum aestivum</i> et <i>Oryza sativa</i> .....	50
<b>Figure 56</b> : Cladogramme des facteurs b-zip, myb et WRKY.....	50
<b>Figure 57</b> : Interface du site web.....	51

## *REMERCIEMENTS*

En tout premier lieu, Nous remercions ALLAH, tout puissant, de nous avoir donné la force pour survivre, ainsi que l'audace pour dépasser toutes les difficultés, puis Nous présentons nos profonds remerciements à notre encadreur Mr. **Kamel Kellou**, de nous avoir encadré et orienté dans nos travaux de recherche.

Nous sommes profondément reconnaissantes au docteur **Kamza Chehili**, d'accepter de présider notre jury et d'examiner notre mémoire pour ses orientations, ses précieux conseils et de nous avoir encadrées durant notre master.

Nous remercions également le docteur **Mahmoud Temagoult** d'avoir accepté d'examiner et de juger ce modeste travail.

Les mots sont insuffisants pour décrire notre reconnaissance et notre gratitude au professeur **Mohamed Abdelhafid Kamidechi**. Nous tenons à lui remercier également pour sa gentillesse, sa disponibilité, sa patience, ses judicieux conseils, ses orientations, son aide à faire cette spécialité et son soutien pour mener à terme notre master. Merci encore.

Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'évaluer notre mémoire de fin d'études et leur contribution qui l'enrichit .

Nous remercions également tous nos enseignants qui ont assuré notre formation tout au long de ces années. Nous soulignons que leurs efforts qui ont été bénéfiques pour effectuer ce travail.

Nous voudrions que ce petit bout de chemin que nous avons accompli ensemble en étant étudiants de master donne ses fruits et que nous puissions élaborer et accomplir notre chemin par un projet de recherche de Doctorat.

Un grand merci à nos amis pour leur sympathie et leurs soutiens. Nous remercions nos parents, nos frères et sœurs, sans eux, ce travail n'aurait jamais pu être réalisé. Un énorme merci à nos familles pour leur soutien indéfectible et leurs encouragements tout au long de l'élaboration de notre travail.

**Dédicace**

*A mes parents avant tout et pour tout*

*C'est avec un grand honneur et une grande humilité que nous dédions cet humble travail à  
celui qui m'a donné la vie*

*Ma mère, que Dieu lui fasse miséricorde et mon père*

*A mes frères et mes sœurs ;*

*les familles s'encoui et khenchouche*

*A tous ceux que nous aime ;*

*A tous mes amis sans exception ;*

*A tous ceux qui sont proches et chères ;*

*Nous te demande de bien vouloir trouver ici ma reconnaissance pour ta patience et ton  
encouragement, et ton grand soutien*

*pour réaliser ce travail.*

*SOMIA*

### *Dédicace*

*Dédié à mes parents : Pour tout l'amour qu'ils m'ont donné et les sacrifices qu'ils ont faits pour la réussite de mes études et de ma formation. Aucune dévotion ne peut exprimer, à sa juste valeur, ma gratitude et mon amour. A ma sœur Maryam à mes amies Samia frahtia et benaida Samaha et Dr: Dardour fathi Mr: khlola Abi taher : la famille saad el-ddin + talbi Témoin des liens de fraternité qui nous unissent, permettez-moi de leur exprimer ma profonde gratitude et mes vœux de bonheur et de réussite.*

*Houria Aya*

## ACRONYMES

- ABA : Acide ABscissique
- ADN: Acide désoxyribonucléotide.
- AP2/EREBP : ethylene-responsive element binding protein
- ARN: acide ribonucléique.
- bZIP :Basic Leucine Zipper Domain <https://www.biomedicalabbreviations.com>
- Ca<sup>2+</sup>: ions calcium .
- CDPK : protéine kinase dépendante du calcium.
- CDS : coding DNA sequence.
- CSS :Cascading Style Sheets.
- EBI : L'Institut européen de bioinformatique.
- ET : l'éthylène.
- genbank : National Genetic Sequence Data Base.
- HGP : programme de séquençage complet du génome humain.
- HTML : Le Hyper Text Markup Language.
- HSF : Heat Shock Factors .
- HSPs : Heat Shock Proteins .
- id :idem .
- JA : l'acide jasmonique .
- MAPK : la protéine kinase activée par les mitogènes .
- MB: Maitre Brin .
- MICADO : MICRobial Advanced Database Organ
- MSA : L'alignement de séquences multiples.
- NAC NAM-ATAF-CUC2
- NCBI : Centre national d'information sur la biotechnologie .
- ROS :Reactive oxygen species .
- SA : l'acide salicylique .
- TFBS :Transcription Factor Binding Sites.
- TF :Les facteurs de transcription .
- TIC : technologies de l'information et de la communication .
- UCSC : l'Université de Californie à Santa Cruz .
- WRKY: transcription factors



# **INTRODUCTION**

Le blé est l'une des trois principales céréales avec le maïs et le riz. C'est, avec environ 700 millions de tonnes par an, le troisième en termes de volume de récolte mondiale, et le riz le plus consommé par l'homme. Le blé, dans la civilisation occidentale, en Afrique du Nord, au Moyen-Orient et dans le nord de la Chine, est un élément central de l'alimentation humaine. Sa consommation remonte à l'Antiquité.

La demande du blé devrait augmenter fortement avec l'augmentation de la population. On estime que la population mondiale atteindra 9 milliards d'ici 2050 (Hubert et *al.*, 2010). Ainsi, la production du blé aura un impact décisif sur la sécurité alimentaire et l'économie mondiale dans les décennies à venir. Cette culture est très sensible au changement climatique et aux maladies provoquées principalement par des champignons parasites. L'amélioration de la production de blé passe donc par une meilleure maîtrise de ces problèmes. Cependant, ces besoins classiques s'ajoutent aux exigences nées de la grande conscience sociale dans la protection de l'environnement, résultant en une réduction des traitements de fertilisation phytosanitaire et chimique. En effet, développer des variétés à haut rendement et de haute qualité reste l'objectif principal de l'obteneur.

Néanmoins, les marqueurs moléculaires apparaissent comme des outils essentiels pour soutenir les nouvelles technologies. Or, la pression de sélection exercée par les taxons avec la résistance à un seul gène favorise le développement de nouveaux biotypes capables de contourner ces résistances et/ou s'adapter aux aléas climatiques, ce qui nécessite une recherche continue et rapide de nouvelles sources de résistance et de tolérance. Les recherches de ces dernières années ont été fructueuses pour identifier les facteurs de transcription (TF) pour réguler la résistance aux stress biotiques et abiotiques du blé et d'autres espèces végétales. De plus en plus, les études ont démontré que les facteurs de transcription sont des régulateurs forts, positifs, négatifs, cis et trans qui activent un gène fonctionnel.

Dans notre étude, nous mettons en évidence la relation possible de 4 familles de facteurs de transcription TF (dont NAC, WRKY, MYB et bZIP) avec la régulation de la défense du blé contre les agents pathogènes et le changement climatique. En se basant sur l'analyse des 4 familles de facteurs de transcription chez les monocotylédones modèles (*Oryza sativa* et *Brachypodium*) et leur comparaison avec les gènes orthologues du blé tendre (*Triticum aestivum*) une étude *in silico* a servira à la conception d'une base de données et la mettre en disponibilité de la communauté scientifique.

**SYNTHESE  
BIBLIOGRAPHIQUE**

**Chapitre 1 : Le blé  
tendre**

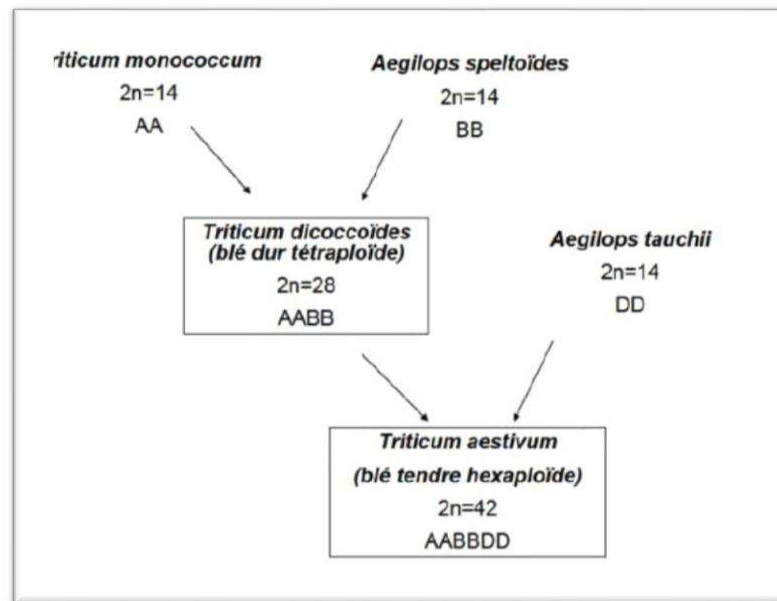
## 1.1. Histoire du blé

L'histoire du blé s'est trouvée liée à celle de l'homme par la naissance de l'agriculture durant la révolution néolithique il y a environ 10 000 ans, en Mésopotamie (Shewry, 2009). Depuis, le blé est devenue l'une des principales ressources alimentaires de l'humanité. De nos jours, les blés cultivés appartiennent au genre *Triticum*, qui regroupe plusieurs espèces présentant des niveaux de ploïdie variables, témoignage de leurs histoires évolutives. Il existe des blés diploïdes (tels que le petit épeautre *T. monococcum*), tétraploïdes (comme l'amidonnier, *T. turgidum*) et hexaploïdes (notamment le blé tendre, *T. aestivum*). L'évolution a permis l'apparition de plusieurs sous-espèces. Le blé dur (*T. turgidum* spp *durum*) et l'amidonnier domestiqué (*T. turgidum* spp *dicoccoïdes*) sont des exemples de la différenciation qui s'est effectuée au sein des espèces tétraploïdes.

L'étude des 3 sous-génomes du blé tendre a permis d'identifier les 3 étapes majeures de son histoire phylogénétique. Premièrement, l'apparition d'un génome ancestral qui a divergé des espèces relatives de la tribu des Triticeae il y a environ 15 millions d'années. Ce génome ancestral est à l'origine des génomes A, B et D (Langridge et al. 2017). Deuxièmement, l'apparition d'un blé dur sauvage (*T. turgidum* ssp. *dicoccoïdes*) au génome AABB par croisement naturel de *T. monococcum* (AA) avec une espèce proche d'*Aegilops spelta* (BB), il y a environ 0,8 millions d'années. Ce croisement est à l'origine du blé dur actuel *T. durum*. Pour finir, le génome DD a ensuite été apporté il y a environ 0,4 millions d'années par croisement de *T. turgidum* ssp. *dicoccoïdes* avec l'espèce diploïde *A. tauschii*, conduisant à l'apparition de l'ancêtre hexaploïde du blé tendre actuel, *T. aestivum* (Feuillet et al., 2008 ; Marcussen et al., 2014) (Figure 1).

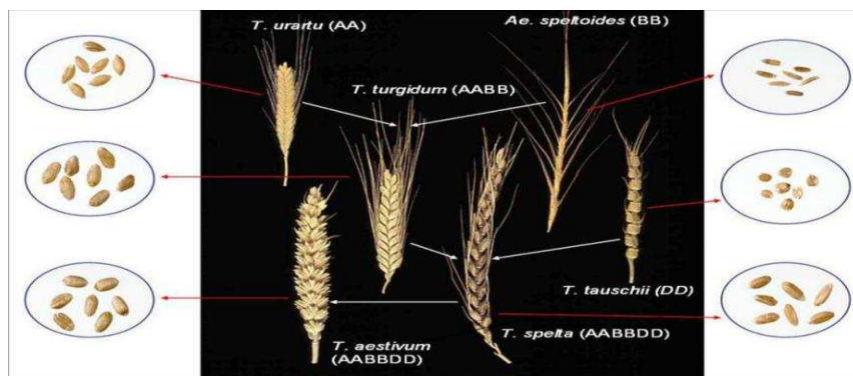
Les blés diploïdes sont les membres les plus anciens du genre *Triticum*. Trois espèces sont classiquement reconnues : l'espèce cultivée de petit épeautre *T. monococcum* et son ancêtre sauvage *T. boeoticum*, porteurs du génome  $A^bA^b$  (l'indice b correspond à *boeoticum*) et *T. urartu* porteur du génome  $A^uA^u$ . Ces 2 génomes auraient divergé il y a approximativement un million d'années (Huang et al., 2002 ; Adonina et al., 2015).

Concernant les blés hexaploïdes, 2 sous-espèces sont cultivées aujourd'hui : le grand épeautre (*T. aestivum* spp *spelta*) et le blé tendre (*T. aestivum* spp *aestivum*). Les blés hexaploïdes possèdent 3 sous-génomes AABBDD constitués chacun de 7 paires de chromosomes, soit 2n = 42 chromosomes pour un total estimé à 17 milliards de paires de bases. Le blé possède un génome de très grande taille mais avec une forte proportion de séquences répétées (80% du génome) (Paux et al. 2008). (Figures 1, 2 et 3).



**Figure 1** : Phylogénie du blé tendre (Salamani *et al.*, 2002). Grine, 2015.

La figure 2 expose l'état actuel de nos connaissances. Ainsi, c'est seulement dans la seconde moitié du XX<sup>e</sup> siècle qu'il est devenu clair pour la communauté scientifique qu'il n'existait pas d'ancêtres sauvages des blés hexaploïdes et que les blés hexaploïdes cultivés résultaient d'hybridation spontanée entre blés tétraploïdes cultivés et espèces sauvages diploïdes.



**Figure 2** : Relations phylogénétiques entre le blé tendre (*T. aestivum*) et ses ancêtres sauvages. « Un exemple d'épi et de grains est donné pour chaque espèce » (d'après Shewry, 2009). Bogard. Published 2011.

## 1.2. Systématique du blé

Le blé est une plante herbacée monocotylédone qui appartient au genre *Triticum*. Un genre de la famille des *Poaceae* (graminées), sous-famille des *Pooideae*, originaire d'Eurasie, qui comprend d'une dizaine à une trentaine d'espèces selon les auteurs. Ce sont des plantes herbacées annuelles aux tiges (chaumes) dressées et aux inflorescences en épis linéaires dont le grain est un fruit sec indéhiscent appelé caryopse constitué d'une graine et de téguments (Feillet, 2000).

Ce genre comprend les diverses espèces de blé cultivées, notamment le blé tendre (*Triticum aestivum*) qui est l'une des céréales les plus importantes. Il contribue pour l'essentiel au génome des blés modernes, polyploïdes, qui intègre aussi des gènes provenant du genre voisin *Aegilops* (Figure 3).

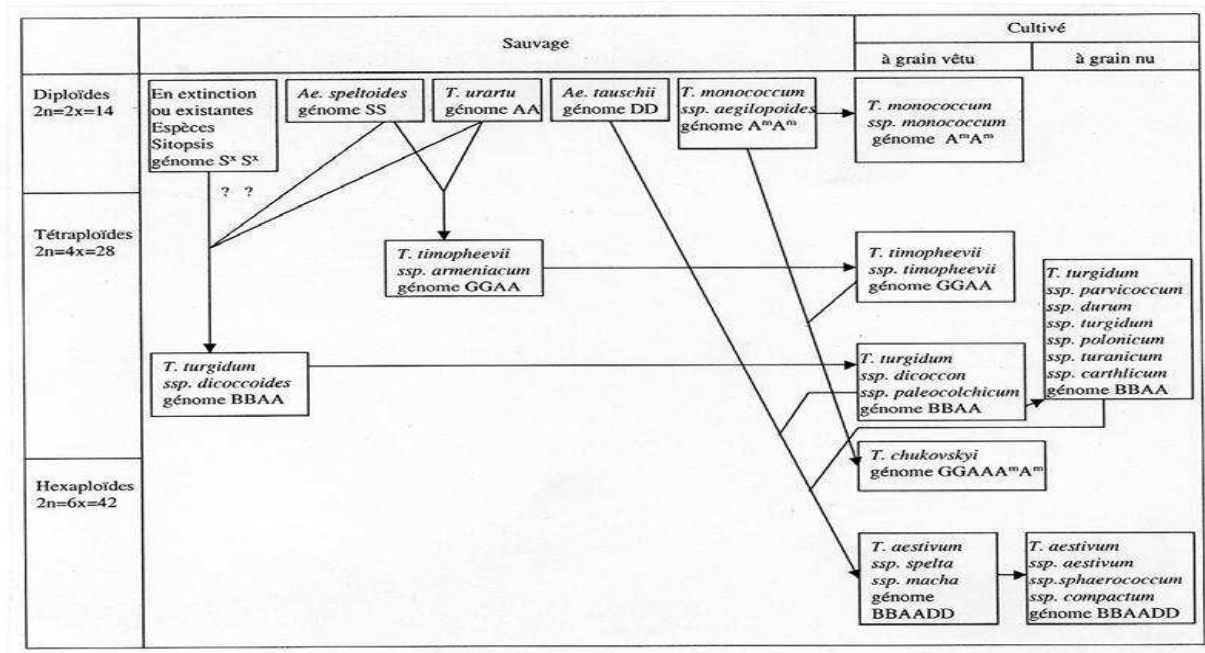


Figure 3 : Phylogénie des blés (d'après Feldman, 2001) Haddad, 2010

### 1.3. Morphologie du blé tendre

Le genre *Triticum* regroupe des plantes angiospermes (ou plantes à fleur) monocotylédones de la famille des Graminées (appelée aussi Poacées) et possédant des fleurs portées par des épillets arrangés en épi. Ce sont des herbacées annuelles à croissance définie, possédant une inflorescence en épi terminal dense et un fruit sec appelé aussi caryopse. Le blé est une céréale à paille d'une hauteur de 0,5 à 1,5 m et cultivée pour ses grains. Il est constitué d'un appareil végétatif (système aérien et système racinaire) et d'un appareil reproducteur (inflorescence).

Le système aérien est structuré en talles partant du plateau de tallage et formées d'une tige feuillée portant une inflorescence à son extrémité. La première talle à se développer est appelée Maître Brin (MB). La tige de chaque talle contient des zones méristématiques, les nœuds, à partir desquelles les feuilles et les entrenœuds se différencient. Une feuille se développe à partir de chaque nœud, en alternance sur 2 rangs le long de la tige. Les feuilles sont structurées en 2 parties : la gaine attachée au niveau du nœud et qui enveloppe l'entrenœud, et le limbe ; ces 2 parties étant reliées par une petite membrane dentelée, la ligule (Soltner, 1988).

Le système racinaire est en réalité composé de 2 systèmes qui se succèdent : Le système de racines primaires ou séminales, qui se met en place à la levée et perdure jusqu'au début du tallage. Il est composé d'une racine principale de laquelle partent 2 paires de racines latérales, Le système de racines secondaires, coronales ou de tallage, qui remplace progressivement le système primaire lors du développement des talles. Il est de type fasciculé et constitué de nombreuses racines adventives d'importance équivalente (Soltner, 1988).

L'inflorescence du blé, de type d'épi, est structurée en épillets qui forment l'unité morphologique de base. L'axe de l'épi, autrement appelé rachis, porte entre 15 et 25 épillets répartis en 2 rangs de part et d'autre de l'axe. Chaque épillet contient 2 à 4 fleurs fertiles dépourvues de pétale mais enveloppées d'une glumelle inférieure et d'une glumelle supérieure, incluses dans 2 bractées ou glumes. La majorité du temps, la fleur de blé s'autoféconde à l'intérieur des glumelles avant que les étamines n'apparaissent à l'extérieur, ce qui en fait une fleur cléistogame (c'est-à-dire autogame avec fécondation avant l'ouverture de la fleur). Contrairement à d'autres graminées comme l'orge ou l'avoine, les glumelles se détachent du grain lors de sa dissémination. Le battage réalisé après la récolte permet de séparer le grain de ses glumes et glumelles. Une fois isolé, le grain de blé présente une section arrondie ou ovale et un poids moyen d'environ 45 mg qui varie selon les génotypes (Soltner, 1988).

#### **1.4. Utilisation du blé tendre, importances économiques et sociétales**

Au sein du genre *Triticum*, le blé tendre et le blé dur sont les 2 espèces les plus cultivées (Feillet, 2000). En 2016, 220 millions d'hectares de surfaces agricoles ont été dévolus à la culture du blé dans le monde, conduisant à la production de 749 millions de tonnes de blé, ce qui en fait la 2<sup>ème</sup> culture à l'échelle mondiale derrière le maïs. L'Asie à elle seule représente 43,6% de cette production, suivie de l'Europe à 33,4% (FAOstat, 2018).

Le blé tendre représente 90% de la production mondiale de blés cultivés, tandis que la majorité des 10% restants est dévolue au blé dur. Cette prédominance peut s'expliquer par la grande plasticité génomique du blé tendre, qui permet de le cultiver dans la majorité des zones agricoles mondiales, et par son potentiel de rendement élevé (Shewry, 2009).

En 2017, il a été récolté 772 millions de tonnes de blé cultivés sur 218 millions d'hectares, soit 38,8% des surfaces agricoles cultivées (FAOstat, 2018 ; Abhinandan et al., 2018). Environ 50% des calories consommées par l'homme seraient dérivées de l'amidon, et cela peut atteindre 80% dans les pays les moins développés (Burrell, 2003). Selon les prévisions, l'humanité atteindra les 9 milliards de personnes d'ici 2050, ce qui nécessiterait une augmentation de 70% de la production alimentaire générale pour garantir la sécurité

alimentaire (FAO, 2009). Cette augmentation de la production pourrait passer par plusieurs voies telles qu'une augmentation des surfaces cultivables ou une amélioration du rendement des cultures en place.

En effet, sous réserve d'un apport suffisant en eau et nutriments et en l'absence de phytopathogènes, son rendement peut atteindre, voire excéder, les 100 quintaux par hectare. En France, le rendement moyen en 2017 a été de 73,7 quintaux par hectare pour le blé tendre contre 57,3 pour le blé dur. À l'échelle mondiale, le rendement moyen était plutôt de 34 quintaux/hectare en 2016 (FAOstat, 2018).

De plus, récolté à maturité (avec un contenu en eau inférieur à 15% du poids sec) et stocké à l'abri des pathogènes, il possède une très longue capacité de conservation. Mais la raison principale de la suprématie de sa culture est la capacité unique de panification de sa farine due au réseau viscoélastique que forment les protéines de réserve au contact de l'eau, réseau plus connu sous le nom de gluten. Cette propriété permet d'en tirer une large gamme de produits transformés tels que les pains, biscuits, gâteaux, etc. De ce fait, le pain est un aliment à la base de nombreux régimes alimentaires à travers le monde, il représente par exemple 11% des apports énergétiques et protéiques des britanniques (Shewry et Hey, 2015).

Les qualités du blé n'ont pas cessé de s'accroître depuis les débuts de sa culture. Les caractéristiques agronomiques des nouvelles variétés (rendement, résistance aux stress) mais également technologiques (aptitude à la panification) s'améliorent de plus en plus. L'amélioration génétique et le progrès des techniques culturales a conduit à une augmentation importante des rendements qui sont passés de 8,6 q/ha en 1900 à 11,5 en 1950 au niveau mondial pour atteindre aujourd'hui près de 29 q/ha dans le monde. Dans les mêmes années en France les rendements sont passés de 10 à 20 q/ha de 1900 à 1950 pour atteindre aujourd'hui près de 70 q/ha (Shewry et Hey, 2015).

### **1.5. Stress abiotique et développement des céréales**

En 1984, Jones et Qualset ont défini les stress abiotiques comme toutes les conditions environnementales qui affectent négativement l'expression du potentiel génétique de croissance, de développement et de reproduction. La compréhension des mécanismes physiologiques, moléculaires et génétiques de tolérance aux stress abiotiques est l'un des domaines les plus actifs de recherche en biologie végétale, de même que l'amélioration de la tolérance et de l'acclimatation des végétaux à ces stress environnementaux.

De manière générale, l'application d'un stress peut réduire la durée de développement du grain, ce qui conduit à une diminution du nombre de cellules (due à la réduction de la phase



de divisions cellulaires) et une diminution de la taille de ces cellules (due à la réduction de la phase de remplissage). Par exemple, un déficit en eau ou un stress thermique fort ou modéré durant la phase de division cellulaire entraînent une diminution du nombre de cellules maximales de l'albumen ; ce qui engendre une limitation de la taille potentielle maximale du grain (Nicolas *et al.*, 1984 ; Ober *et al.*, 1991 ; Gooding *et al.*, 2003 ; Semenov *et al.*, 2014 ; Grousse *et al.*, 2018). Ce stress peut aussi avoir un effet direct sur l'activité enzymatique liée à l'accumulation des réserves, conduisant à un grain de qualité altérée (Altenbach *et al.*, 2003 ; Thitisaksakul *et al.*, 2012).

L'adaptation des plantes à un stress abiotique passe par l'acquisition d'une tolérance vis-à-vis de ce stress, contrairement aux stress biotiques pour lesquels on parle d'acquisition de gènes de résistance. La tolérance à un stress abiotique consiste en la mise en place de mécanismes permettant de diminuer l'impact du stress sur le rendement et la qualité de la culture (Ney *et al.*, 2013). De plus, la résistance aux stress biotiques est souvent décrite selon un mécanisme gène pour gène tandis que la mise en place de mécanismes de tolérance à un stress abiotique présenterait un caractère multigénique, beaucoup plus complexe à étudier.

### **1.6. Effets des principaux stress abiotiques sur le développement et la croissance du blé**

Les contraintes environnementales tendent aujourd'hui à réduire les surfaces arables (indisponibilité en eau, salinisation des terres...) et impactent négativement les rendements.

En effet, le changement climatique subi par la terre depuis plusieurs années provoque un bouleversement des écosystèmes, principalement dû à une hausse des températures.

Selon Hansen *et al.* (2016), les trois dernières décennies seraient les plus chaudes que la terre ait connu depuis 1400 ans dans l'hémisphère Nord et selon les modèles climatiques, la température moyenne globale continuera d'augmenter de 1 à 4°C par rapport aux niveaux de la fin du vingtième siècle (Pachauri *et al.*, 2015). Cette augmentation de la température moyenne dans de nombreuses régions du monde est couplée avec une augmentation de la fréquence et de la durée des épisodes de fortes chaleurs (chocs thermiques) qui tendent à modifier la répartition des aires de production du blé (Olesen et Bindi, 2002).

Avec l'augmentation des températures, une modification de la disponibilité en eau induit un stress hydrique pour un grand nombre de cultures en zones non-irriguées, et un stress salin provoqué par l'évaporation de l'eau et la concentration des sols en sels (Abhinandan *et al.*, 2018). Un stress hydrique ou salin entraîne dans tous les cas un stress osmotique qui peut occasionner des dommages cellulaires à tous les stades de développement, plus ou moins importants en fonction de l'intensité et de la durée du stress, et se traduit par un impact négatif important sur le développement du blé et donc sur son rendement.

Face à ces contraintes, le blé est capable de mettre en place des réponses d'acclimatation qui ne permettent toutefois pas de compenser entièrement l'impact négatif de ces conditions sur le rendement en grains. Les pertes de production en blé les plus importantes sont causées par les stress abiotiques comme le stress hydrique, le stress salin et les hautes températures, plus que par les stress biotiques (Abhinandan et *al.*, 2018). La présence d'une contrainte thermique modérée deviendra, dans les années à venir, une problématique omniprésente dans toutes les cultures agricoles.

De nombreuses études couvrant un large éventail de régions montrent d'abord un impact négatif du changement climatique sur le rendement de grandes cultures céréalières (maïs, blé et riz) (Lobell et *al.*, 2011). Plus particulièrement, Asseng et *al.*, (2015) ont estimé une perte de 6% de la production mondiale en blé (soit 44 millions de tonnes) par 1°C supplémentaire de la température moyenne. Cette diminution du rendement couplée à l'augmentation de la population humaine est à l'origine des inquiétudes vis-à-vis de la sécurité alimentaire mondiale (Lobell et *al.*, 2011). De manière plus précise, les stress abiotiques impactent les composantes du rendement et la qualité des grains des grandes cultures céréalières.

Chez les céréales, le rendement en grain est exprimé en masse à 15% d'humidité par hectare et correspond à la quantité de grains produits. Avec la qualité, il présente un intérêt agronomique majeur et l'augmentation de la demande (due à l'augmentation de la population mondiale et au développement des biocarburants) couplée à l'impact négatif du changement climatique font que le rendement est actuellement la principale cible des programmes d'amélioration variétale (Sreenivasulu et Schnurbusch, 2012).

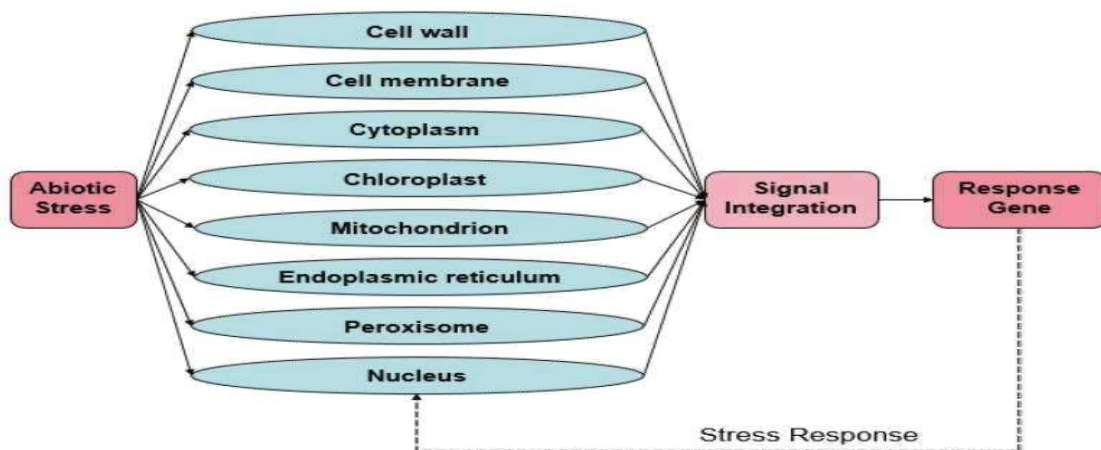
L'amélioration variétale passe par la compréhension des mécanismes de tolérance aux stress abiotiques retrouvés chez des génotypes tolérants, puis par l'exploitation de la diversité allélique existante dans le pool de gènes et l'identification des caractères d'intérêt pour améliorer la tolérance aux stress (Abhinandan et *al.*, 2018).

### **1.7. Réponses moléculaires et cellulaires des plantes aux stress abiotiques**

En biologie végétale, la compréhension des réponses moléculaires, physiologiques et génétiques des plantes aux conditions de stress environnemental est un domaine de recherche très actif, dont de nombreuses approches visent à améliorer la tolérance et l'acclimatation des plantes cultivées. La réponse de la plante aux stress comme les températures extrêmes et le stress hydrique ou salin implique des gènes catégorisés classiquement en 2 groupes (Shinozaki et Yamaguchi-Shinozaki, 1997).

Les gènes de fonction codent pour des protéines assurant une protection directe de la cellule végétale face au stress, et les gènes de régulation permettent la perception et la transduction du signal stress au sein des cellules végétales par des voies de régulation-signalisation générales et spécifiques. Sous leur contrôle, l'activation de réseaux de régulation complexes permet de modifier l'expression de gènes de réponse afin de contrer les effets négatifs et de rétablir l'homéostasie cellulaire (Reguera *et al.*, 2012 ; Suzuki *et al.*, 2012).

De manière générale, les senseurs détectent le signal stress et initient des cascades de signalisation qui permettent la réponse de la plante. L'initiation et la régulation de ces processus de réponse aux stress impliquent principalement des molécules de signalisation comme les phytohormones (Pieterse *et al.*, 2012 ; Bartoli *et al.* 2013), les ions calcium ( $Ca^{2+}$ ) (Stael *et al.*, 2015), les espèces réactives de l'azote et de l'oxygène (Kocsy *et al.*, 2013 ; Baxter *et al.*, 2014) (Figure 4). Les gènes régulateurs du signal stress sont des gènes impliqués dans la transduction du signal : les récepteurs ancrés à la membrane, les senseurs de calcium, les protéines kinases, les protéines phosphatases et les facteurs de transcription.



**Figure 4 :** Détection du stress et signalisation ; un modèle de réponse aux stress abiotiques dans différents organites (tiré de Zhang *et al.*, 2018). Claire, 2019

Le signal généré par les organites peut provoquer l'expression des gènes et des activités cellulaires, ce qui peut restaurer l'homéostasie cellulaire sous stress abiotique.

La détection du signal stress a souvent lieu à la surface des membranes cellulaires, pour être ensuite relayé vers plusieurs organites (Figure 4). Chez le blé, Leucci *et al.*, (2008) ont montré une modification de la proportion en composés pectiniques en réponse à un stress hydrique chez le génotype tolérant Capeiti, qui était absente chez le génotype sensible Creso. Les auteurs ont supposé que cette modification structurelle de la paroi agirait à la fois comme gélifiant et comme anti-dessicant, limitant les pertes hydriques.

Parmi les mécanismes communs qui se mettent en place en réponse aux stress abiotiques, la production des espèces réactives de l'oxygène (ROS) est particulièrement étudiée. Les ROS jouent un double rôle dans la réponse des plantes aux stress abiotiques. En effet, ce sont à la fois des sous-produits toxiques du métabolisme du stress mais aussi des molécules importantes dans la transduction des signaux. Ce sont des initiateurs et intégrateurs des réseaux régulant un large éventail de processus tels que les réponses de croissance, de développement, de défense et d'acclimatation aux stress biotiques et abiotiques (Gapper et Dolan, 2006 ; Baxter et *al.*, 2014).

La signalisation via les ROS est étroitement liée à la réponse aux hormones végétales, notamment l'acide salicylique, l'acide jasmonique, l'éthylène, l'ABA et l'auxine (Blomster et *al.*, 2011 ; Pieterse et *al.*, 2012 ; Bartoli et *al.*, 2013). La production des ROS tels que les anion superoxydes, le peroxyde d'hydrogène (H<sub>2</sub>O<sub>2</sub>), les radicaux hydroxyle et l'oxygène singulet affecte le fonctionnement des chloroplastes et impacte la photosynthèse. H<sub>2</sub>O<sub>2</sub> et O<sub>2</sub><sup>-</sup> sont des exemples de ROS générés en réponse à de nombreuses conditions de stress comme le choc thermique, les stress hydrique ou salin (Apel et Hirt, 2004 ; Suzuki et Mittler, 2006 ; Baxter et *al.*, 2014).

Les stress abiotiques induisent aussi l'expression de gènes codant des enzymes du système anti-oxydant permettant de piéger les ROS, comme l'ascorbate peroxydase ou la catalase. Ces enzymes servent à détoxifier les cellules faisant face à une surproduction de ROS qui, en trop grandes quantités, sont des inducteurs de mort cellulaire. En effet, même si ce sont des médiateurs secondaires qui améliorent les voies sensibles au choc thermique (Suzuki et Mittler 2006 ; Baxter et *al.*, 2014), une surproduction de ROS en condition de stress devient délétère pour la plante. Par exemple, les mutants knockout catalase 2 montrent un phénotype sensible au stress thermique chez *Arabidopsis* (Vanderauwera et *al.*, 2011).

D'autres gènes intervenant dans les modifications des histones et la méthylation de l'ADN montrent une expression impactée par les stress hydrique, salin et température (Luo et *al.*, 2012).

Les protéines kinases et phosphatases interviennent dans la perception et la réponse aux stress environnementaux. La surexpression de plusieurs TaSnRK2 (Sucrose non fermenting 1-Related protein Kinase 2) de blé améliore la tolérance des *Arabidopsis* transgéniques face à de nombreux stress abiotiques (Xu et *al.*, 2009 ; Mao et *al.*, 2010 ; Zhang et *al.*, 2010, 2011 ; Tian et *al.*, 2013). De la même manière, le gène TaSK5 de blé code pour une kinase qui est

induite constitutivement par les stress abiotiques et améliore la tolérance aux stress salin et hydrique chez *Arabidopsis* (Christov et al., 2014).

Lorsqu'il est surexprimé, TaABC1 améliore aussi la tolérance aux stress hydrique, salin et froid chez *Arabidopsis* (Wang et al., 2011). Les phosphatases aussi répondent à un stress hydrique. Par exemple, TaPP2A, une sérine/ thréonine protéine phosphatase du blé, confère une meilleure tolérance au stress hydrique chez le tabac lorsque qu'il surexprime l'une de ses sous-unités. Certains microARN impactent la tolérance au stress thermique. TamiR159 du blé induit une plus grande sensibilité au stress thermique lorsqu'il est surexprimé chez le riz (Wang et al., 2012).

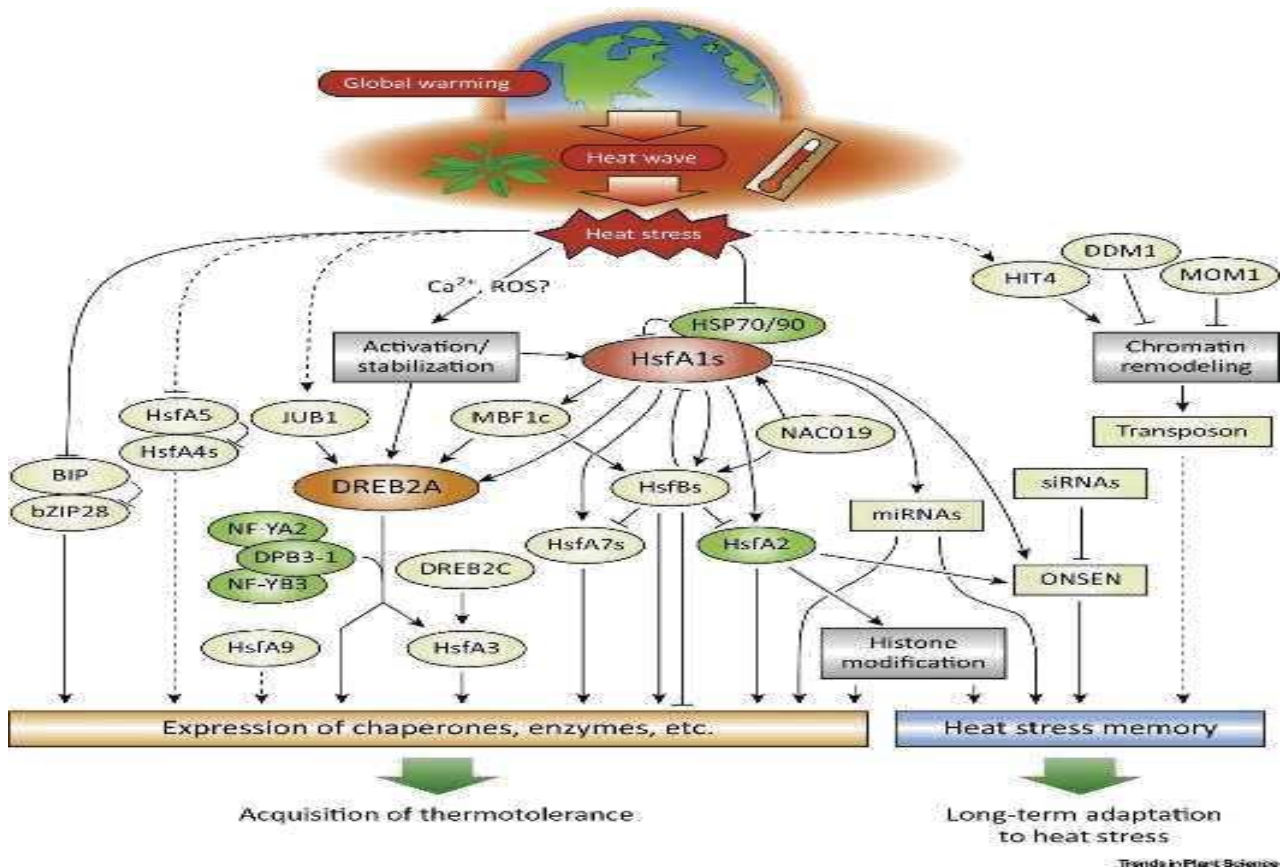
Deux aquaporines ont été décrites comme présentant un rôle opposé en réponse à un stress hydrique chez le blé. La surexpression de TaAQP7 induit une augmentation de la tolérance au stress hydrique chez le tabac entre autre en réduisant l'accumulation d'H<sub>2</sub>O<sub>2</sub> et en augmentant l'activité de la SuperOxide Dismutase (SOD) et de la catalase (Zhou et al., 2012). À l'inverse, la surexpression de TaTIP2;2 chez *Arabidopsis* augmente sa sensibilité aux stress hydrique et salin (Xu et al., 2013). D'autres gènes comme TaLEA3 (Late Embryogenesis-Abundant), TaHPS (inhibiteur de l'alpha-amylase) ou encore 3 gènes FBE (enzymes de biosynthèses des carbohydrates) sont des régulateurs positifs de la tolérance au stress hydrique (Wang et al., 2009 ; Bie et al., 2012 ; Xiao et al., 2013).

La détection d'un stress thermique se fait en premier lieu au niveau de la membrane plasmique cellulaire. Il altère la fluidité membranaire, la structure du cytosquelette, modifie la conformation des protéines et impacte les processus métaboliques (Chakraborty et Chakraborty, 2015).

En réponse à un stress abiotique, et spécialement un choc thermique (température en dehors des températures physiologiques), on observe la production de HSPs (Heat Shock Proteins), des chaperons moléculaires (Kotak et al., 2007 ; Qu et al., 2013 ; Muthusamy et al., 2017) qui protègent les protéines d'une dénaturation et leur permet de conserver un fonctionnement efficace au sein de la cellule. Cinq familles de HSP sont reconnues : HSP100, HSP90, HSP70, les chaperonines (HSP60) et les small HSP (ou sHSP, de 12 à 40 kDa) (Qu et al., 2013).

Certaines HSP sont exprimées constitutivement et servent à la conformation, l'assemblage, le transport et la dégradation des protéines dans la cellule. D'autres, comme la sHSP26, sont exprimées constitutivement à certains stades de développement mais sont aussi fortement induites par un choc thermique (Chauhan et al., 2012). La transcription des gènes HSP est

contrôlée par des Heat Shock Factors (HSF), des protéines régulatrices exprimées en réponse aux stress thermique, hydrique et salin (Figure 5). Certains HSF sont aussi exprimés constitutivement et les protéines sont complexées dans un état inactif dans le cytoplasme (Xue et *al.*, 2014). Les HSF existent chez tous les eucaryotes, mais ils sont particulièrement nombreux chez les plantes. Par exemple, Xue et *al.*, (2014) ont recensé 56 membres de la famille des HSF chez le blé, tandis que l'homme en possède 7 membres.



**Figure 5 :** Réseau de régulation transcriptionnelle impliqué dans la réponse au stress thermique (tiré de Ohama et *al.*, 2017). Claire, 2019.

Les flèches continues présentent les relations observées par la communauté scientifique. Les lignes pointillées indiquent les liens à confirmer.

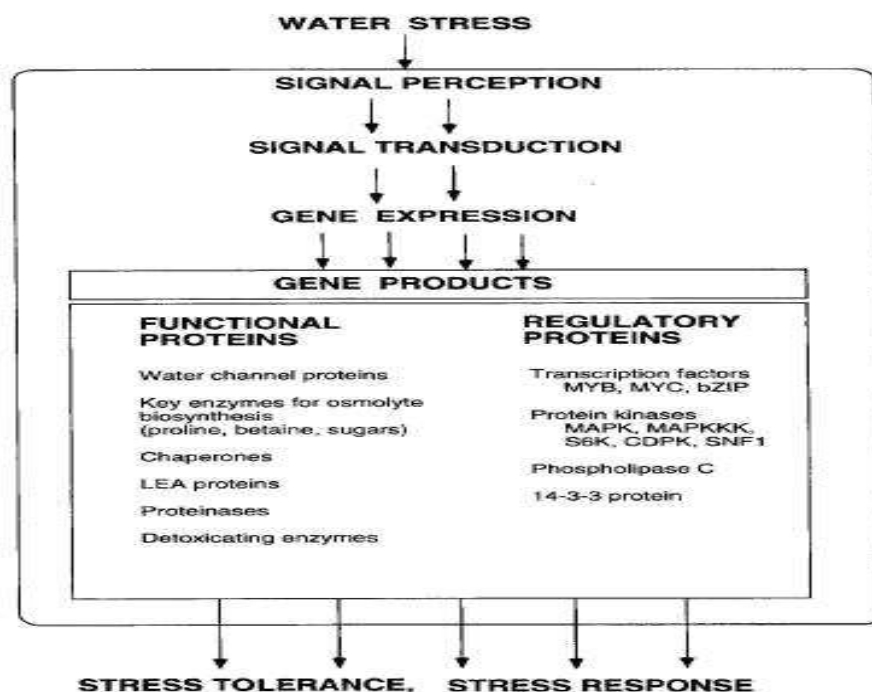
Les facteurs de transcription (TF) jouent un rôle essentiel dans la signalisation, la régulation et la transduction du signal stress, à l'interface entre sa perception et l'expression des gènes de réponse. Ce sont des intégrateurs du signal stress permettant sa transmission et la réponse de la plante via la modification du transcriptome de la cellule ; pour cela, ils interviennent dans la régulation du niveau d'expression de gènes cibles en recrutant ou bloquant la machinerie transcriptionnelle. Environ 6 à 8% du génome d'*Arabidopsis* code pour plus de

1500 facteurs de transcription (Franco-Zorrilla *et al.*, 2014), dont 45% appartiennent à des familles de gènes spécifiques aux plantes (Riechmann *et al.*, 2000). Parmi plus de 80 familles de facteurs de transcription, les familles AP2/EREBP, bZIP, MYB/MYC, NAC et WRKY sont connues pour être fortement impliquées dans la réponse aux stress biotiques et abiotiques chez les plantes (Riechmann et Meyerowitz 1998 ; Uno *et al.*, 2000 ; Jakoby *et al.*, 2002 ; Abe *et al.*, 2003 ; Fujita *et al.*, 2004 ; Tran *et al.*, 2004, 2007 ; Yanhui *et al.*, 2006 ; Lata et Prasad 2011 ; Zhu *et al.*, 2013 ; Huang *et al.*, 2015 ; Gahlaut *et al.*, 2016). Leur importance dans le contrôle en amont d'un grand nombre de voies métaboliques en fait des cibles de choix pour les études portant sur la caractérisation des réponses aux stress.

Chez *Arabidopsis*, des portions de réseaux transcriptionnels complexes comprenant de nombreux régulateurs de la transcription sont maintenant connus en réponse choc thermique.

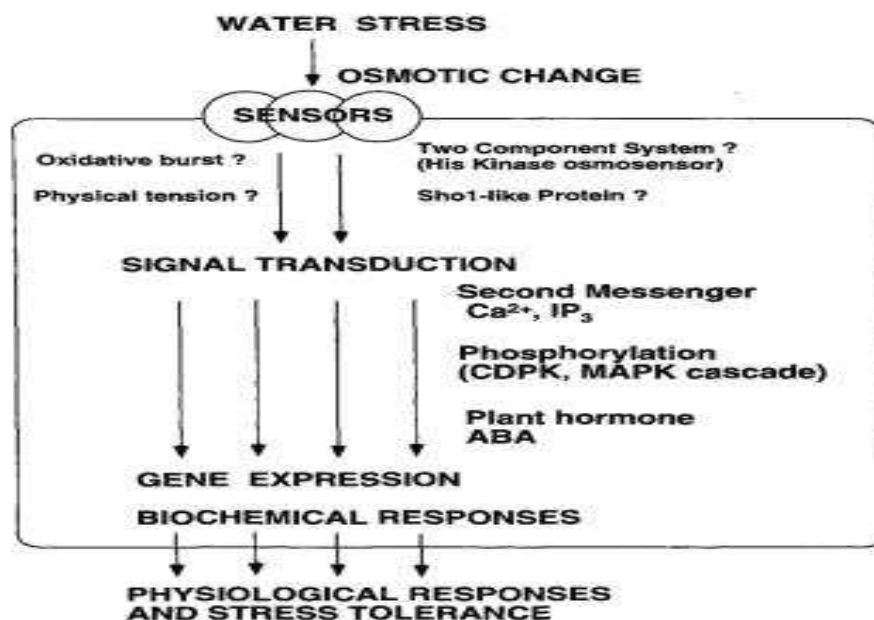
### 1.8. Les fonctions régulatrices des facteurs de transcriptions en réponse aux stress abiotiques et biotiques

Les plantes ont développé des stratégies de réponse rapide aux conditions défavorables, et celles-ci impliquent des réseaux interconnectés au niveau moléculaire contrôlés par des cascades de signaux. Les différents composants des réponses au stress sont la perception du signal, la transduction du signal et l'expression de gènes sensibles au stress (Figure 6).



**Figure 6 :** Fonction des gènes dont l'expression est induite par un stress abiotique, dans la réponse et la tolérance à ce stress. Claire, 2019

Lorsque les cellules végétales perçoivent un signal de stress, des récepteurs ou des capteurs dans la paroi cellulaire ou la membrane détectent le stimulus de stress, suivi d'une réponse rapide qui traduit le signal externe en signaux intracellulaires. Des cascades de signaux impliquant des molécules ou des ions intracellulaires sont activées avec des cascades de kinases, qui sont généralement cytoplasmiques. Des cascades majeures sont associées aux espèces réactives de l'oxygène (ROS) et aux ions calcium ( $\text{Ca}^{2+}$ ). Les phytohormones, y compris l'acide abscissique (ABA), l'acide jasmonique (JA), l'acide salicylique (SA) et l'éthylène (ET), sont de puissants seconds messagers qui coordonnent les voies de transduction du signal pendant les réponses au stress. Ces signaux activent plusieurs voies de transduction parallèles, qui impliquent souvent des phosphatases et des protéines kinases (Figure 7).



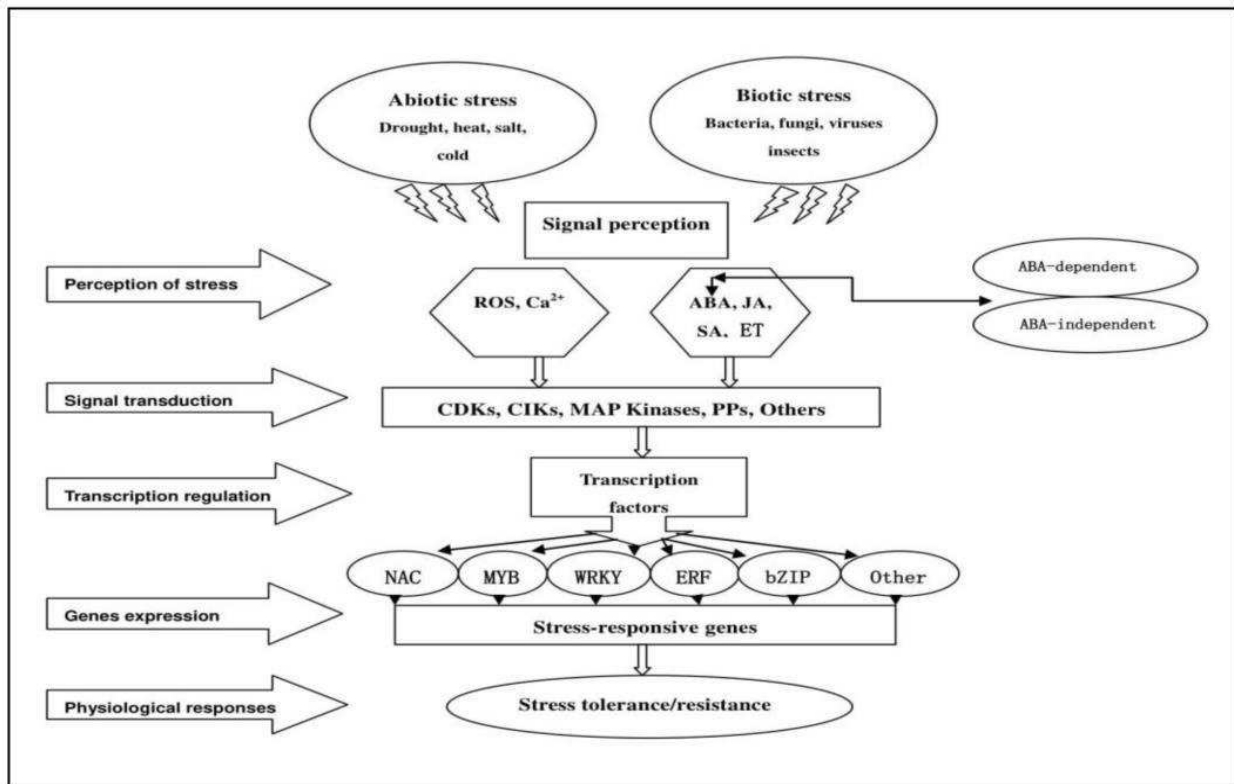
**Figure 7 :** Messagers secondaires et facteurs impliqués dans la perception et la transduction du signal dans la réponse au stress hydrique. Claire, 2019

Après l'étape initiale de perception du signal, les plantes activent deux cascades de signaux majeures : les voies de la protéine kinase activée par les mitogènes (MAPK) et de la protéine kinase dépendante du calcium (CDPK). Enfin, des TF spécifiques sont régulés à la hausse ou la baisse par des protéines kinases ou des phosphatases, et les TF se lient aux éléments cis des gènes liés au stress pour améliorer ou supprimer leur transcription.

La figure 8, résume globalement l'activation et l'activité générales des TF. Étant donné que de nombreuses interactions établissent les réseaux de régulation qui modulent l'expression des gènes sensibles au stress, les fonctions des TF dans les plantes sont d'une portée



considérable. Fondamentalement, cependant, les TF modulent l'expression de gènes clés en aval. Les sections suivantes résument les découvertes récentes sur les principales familles de TF.



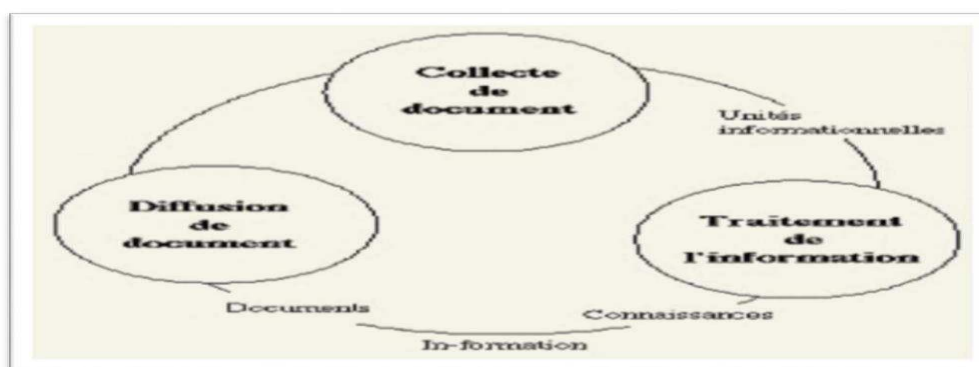
**Figure 8 :** Modèle schématique des voies de signalisation impliquées dans les réponses aux stress abiotiques et biotiques.

**Chapitre 2 :**  
**La recherche**  
*in silico*

## 2.1. Définitions

### 2.1.1. Recherche *in silico*

Ce vocable indique le début et l'ampleur d'un phénomène en biologie moléculaire ; les recherches ne sont plus seulement *in vivo* ou *in vitro*, mais ont un recours de plus en plus essentiel aux analyses informatiques. Il souligne ainsi l'importance des technologies de l'information et de la communication (TIC) dans le développement de cette discipline et en désigne surtout deux champs spécifiques : la génomique et la bioinformatique (Gabriel, 2002). L'activité informationnelle des chercheurs est circonscrite par le modèle schématisé ci-dessous (Figure 9).



**Figure 9** : Schéma général de la recherche *in silico*. Gabriel G.2002

La collecte peut être réalisée à partir de banques de données, de sites web, d'expériences dans les laboratoires, de "butinage" (browsing) dans les rayonnages d'une bibliothèque... Le traitement correspond à l'activité cognitive des chercheurs ou à des manipulations par des outils informatiques. La diffusion est définie comme l'ensemble des opérations nécessaires à la propagation des connaissances. La transformation de l'information s'inscrit dans ce processus où la connaissance est la formation des idées, l'information est la mise en forme des connaissances (information) et l'information inscrite sur un support constitue un document (Gabriel. 2002).

En biologie «*in silico*», traitement automatisé d'informations biologiques, "*in silico*" (dans la silice, composant des «puces» des ordinateurs). Terme apparu dans «A. Danchin, C. Médigue, O. Gascuel, H. Soldano, A. Hénaut - *From data banks to data bases - Research in Microbiology* (1991) 142: 913-916».

### **2.1.2. La bioinformatique**

Association de la biologie et de l'informatique afin d'analyser, modéliser ou prédire les informations issues de données biologiques expérimentales et développer les outils nécessaires. Concevoir des méthodes de traitement de l'information inspirées des systèmes biologiques permettant de résoudre des problèmes algorithmiques complexes, (algorithmes génétiques, réseaux de neurones formels, etc...)

À l'interface entre biologie, informatique et mathématiques, la bioinformatique analyse et interprète, au moyen de méthodes informatiques, les données biologiques que sont les séquences des gènes et des protéines cellulaires, et apporte ainsi de nouvelles connaissances sur le fonctionnement des cellules et des organismes vivants.

La bioinformatique est définie comme l'utilisation de bases de données et d'algorithmes informatiques pour analyser les gènes, les protéines, et la collection complète d'acide désoxyribonucléique (ADN) d'un organisme vivant (le génome).

Un défi majeur en biologie consiste à comprendre les énormes quantités de données de séquence et de données structurales générées par les expériences biologiques (ex : les projets de séquençage).

Elle est basée sur l'Analyse de la séquence, et de l'expression, la prédiction de structure et la Phylogénie pour comprendre les systèmes de régulation, étudier les interactions moléculaires et reconstituer ainsi l'histoire de la «vie».

### **2.1.3. La génomique**

En génétique, la recherche d'exhaustivité, liée au fait que toute l'information génétique nécessaire à un organisme est contenue dans son ADN, a propulsé la biologie moléculaire dans l'ère de la génomique. Pour faciliter l'accès et le traitement des séquences biologiques, il y a eu nécessité de les enregistrer dans les banques, désormais en ligne sur Internet. L'impulsion du programme de séquençage complet du génome humain (le HGP), puis de celui d'autres génomes, et la généralisation et l'évolution concomitante des séquenceurs, ont fait croître de manière exponentielle la production des séquences d'ADN. Le HGP aura un effet mobilisateur, tant sur le plan des techniques, que sur l'impulsion de projets de séquençage d'autres génomes. Du fait de la taille réduite de leurs chromosomes, les microbes sont les premiers organismes vivants et autonomes qui ont fait l'objet d'une lecture complète de leur information génétique (Gabriel, 2002).

Pourquoi comparer les séquences ?

➤ Une ressemblance entre séquences peut indiquer :

- une fonction biologique proche,
- une structure 2D ou 3D semblable,
- une origine et/ou histoire évolutive commune.

➤ La comparaison de séquence permet aussi

- d'assembler des fragments de séquences,
- de mettre en évidence les différences de séquençage entre différents laboratoires.

### 2.1.4. La phylogénie

La phylogénie correspond à l'étude des liens existant entre espèces apparentées. Grâce à elle, il est possible de retracer les principales étapes de l'évolution des organismes depuis un ancêtre commun et ainsi de classer plus précisément les relations de parentés entre les êtres vivants. <https://www.futura-sciences.com/>

La classification phylogénétique est un système de classification des êtres vivants qui a pour objectif de rendre compte des degrés de parenté et des proximités évolutives entre les espèces. Elle ne reconnaît pas les groupes para-phylétiques. Seuls les groupes monophylétiques (ou clades) sont admis, c.-à-d. qu'ils doivent inclure tous et seulement les descendants d'un ancêtre commun (hypothétique). <https://biologievegetale.be/>

La phylogénie est représentée sous la forme d'un arbre phylogénétique (Figure 10). Celui-ci peut se présenter sous deux formes : un cladogramme, qui considère en chaque nœud l'apparition ou la perte d'un caractère ; un dendrogramme, se focalisant davantage sur la distance génétique entre les populations étudiées. <https://www.futura-sciences.com/>

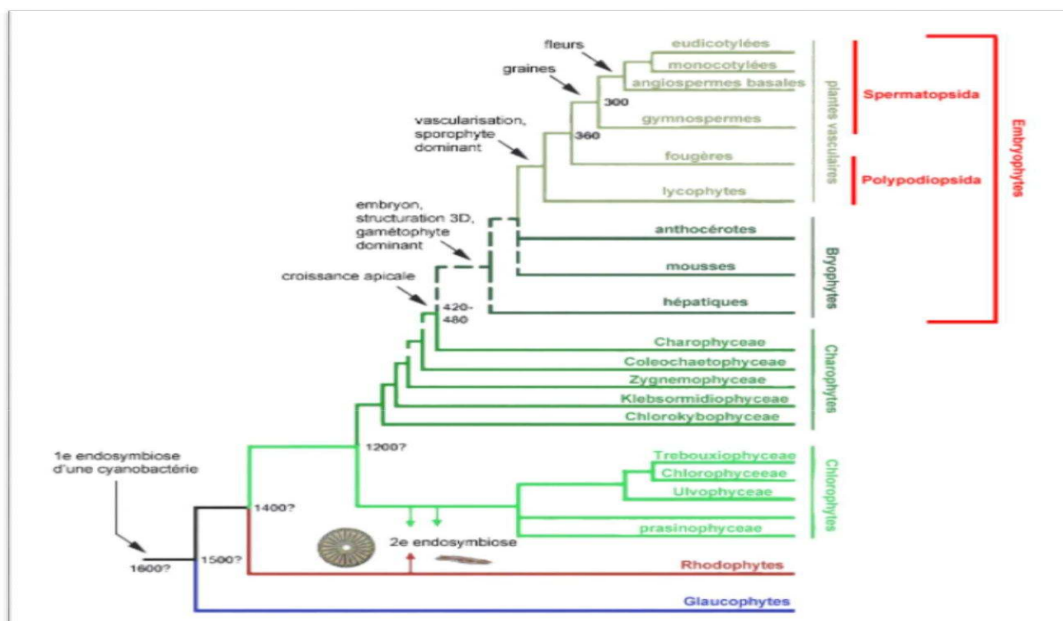


Figure 10 : Phylogénie du règne végétal <https://biologievegetale.be/>

### 2.1.5. L'annotation

L'annotation des génomes est une analyse informatique des séquences obtenues lors du séquençage permettant d'identifier les séquences informatives des génomes. Ces séquences sont principalement les gènes (on parle alors de prédiction de gènes).

L'Annotation des génomes des plantes en bioinformatique est en deux étapes :

#### - Production

La production est l'action d'un sujet qui transforme une matière première pour faire exister un nouvel objet. On rencontre ce phénomène de production dans la société, mais aussi bien dans la nature. C'est pourquoi on peut l'étudier soit sous l'angle économique et sociologique, soit sous l'angle biologique.

#### - Alignement

En bioinformatique, l'alignement de séquences (ou alignement séquentiel) est une manière de représenter deux ou plusieurs séquences de macromolécules biologiques (ADN, ARN ou protéines) les unes sous les autres, de manière à en faire ressortir les régions homologues ou similaires. L'objectif de l'alignement est de disposer les composants (nucléotides ou acides aminés) pour identifier les zones de concordance. Ces alignements sont réalisés par des programmes informatiques dont l'objectif est de maximiser le nombre de coïncidences entre nucléotides ou acides aminés dans les différentes séquences. Ceci nécessite en général l'introduction de « trous » à certaines positions dans les séquences, de manière à aligner les caractères communs sur des colonnes successives. Ces trous correspondent à des insertions ou des délétions (appelés indel) de nucléotides ou d'acides aminés dans les séquences biologiques. Le résultat final est traditionnellement représenté comme des lignes d'une matrice.

```
-----D-PGDF--DRNVPRICVCGDRATGPHFNAMTCEGCKGFFRRSMRKA--LFTCP-FNGDCRIITKDNRRHCOACRLKRCVDIGMMKEFILTD
IRPQKRK-KGPAP-KMLGNEICSVCGDKASGPHYNVLSCEGCKGFFRRSVIKGA--HYICH-SGGHCPMDTYMRKCOECLRKCROAGMREECVLSE
SVPGKPS-VNADE-EVGGPQICRVCGDKATGYHFNVMTCCEGCKGFFRRAMKRNA--RLRCPFRKGACEITRKTTRQCOACRLKRCLES GMKEMIMSD
EPEKRRK-KGPAP-KMLGHELCSVCGDKASGPHYNVLSCEGCKGFFRRSVVRRGGARRYACR-GGGTCOMDAFMRKCOOCLRKCCKEAGMREOCVLSE
PVTKKPRMGASAG-RIKGDDELCSVCGDRASGYHYNALTCCEGCKGFFRRSITRNA--VYKCK-NGGNCVMDMYMRKCOECLRKCCKEMGMLAECMYTG
QTEEKCK-KGYIPSYLDKDELCSVCGDKATGYHYRCITCEGCKGFFRRSITQKNLHPSYSCK-YEGKCVIDKVTNRQCOECLRKCCKEAGMREOCVLSE
----SPS-PPPPP---RVYKPCFVCCNDKSSGYHYGVSAACEGCKGFFRRSIOKNM--VYTCR-RDKNCIINKVTRNRQCOECLRKCCKEAGMREOCVLSE
----PPS-PLPPP---RVYKPCFVCCNDKSSGYHYGVSAACEGCKGFFRRSIOKNM--VYTCR-RDKNCIINKVTRNRQCOECLRKCCKEAGMREOCVLSE
----PPS-PPPLP---RIVKPCFVCCNDKSSGYHYGVSAACEGCKGFFRRSIOKNM--VYTCR-RDKNCIINKVTRNRQCOECLRKCCKEAGMREOCVLSE
```

**Figure 11** : Exemple d'alignement de séquences de récepteurs nucléaires.

Les acides aminés conservés sont surlignés en bleu et en vert. Par endroits, on a inséré des trous, symbolisés par des tirets « - », pour permettre un alignement optimal.

## 1.2. Les différentes facettes de la bioinformatique

Pour l'analyse des données expérimentales que représentent les séquences biologiques, cet apport informatique concerne principalement trois aspects :

### 2.2.1. Compilation et organisation des données

Cet aspect concerne essentiellement la création de bases de données.

Certaines ont pour vocation de réunir le plus d'informations possible (**bases de données généralistes**) sans expertise particulière de l'information déposée alors que d'autres sont spécialisées dans un domaine considéré avec l'intervention d'experts (**bases de données spécialisées**).

Les banques de données spécialisées sont généralement construites autour de thèmes précis comme l'ensemble des séquences d'une même espèce ou les facteurs de transcription. Incontestablement, toutes ces banques de données constituent une source de connaissance d'une grande richesse que l'on peut exploiter dans le développement de méthodes d'analyse ou de prédiction.

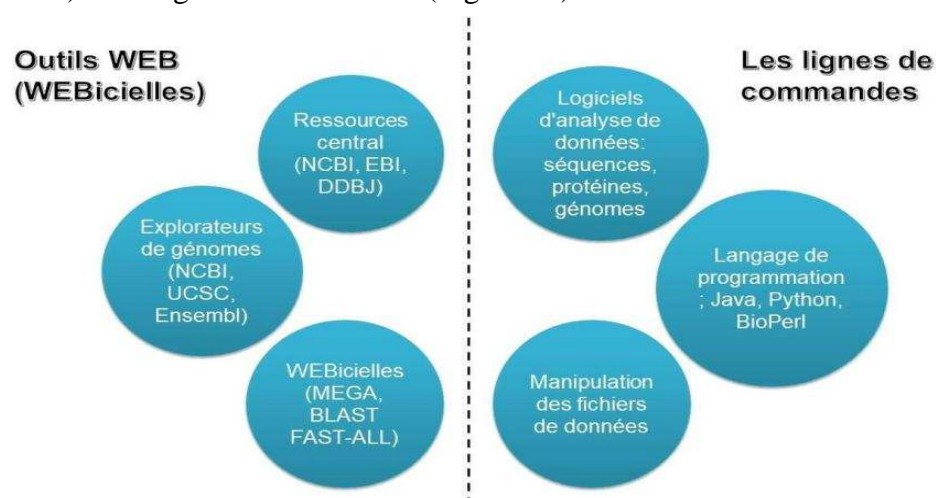
### 2.2.2. Traitements systématiques des séquences (l'annotation des séquences)

L'objectif principal est de **repérer ou de caractériser une fonctionnalité** ou **un élément biologique intéressant**. Ces programmes représentent les traitements couramment utilisés dans l'analyse des séquences comme l'identification de phases codantes (CDS) sur une molécule d'ADN ou la recherche de similitudes d'une séquence avec l'ensemble des séquences d'une base de données.

### 2.2.3. Bioinformatique et logiciels

Il est maintenant facile et courant d'effectuer certaines opérations plus ou moins complexes à **l'aide de logiciels** plutôt que manuellement. Pourtant, ces pratiques ne sont pas toujours systématiques car il est souvent difficile pour certains utilisateurs de savoir quel programme utiliser en fonction d'une situation biologique déterminée ou d'exploiter les résultats fournis par une méthode.

Il existe deux approches totalement différentes de la bioinformatique : l'utilisation d'outils Web (WEBicielles) et de lignes de commande (Figure 12).



**Figure 12** : Ressources en bioinformatique. (Modifiée à partir de Pevsner, 2015). Fondrait

Granger Brunet et al . 2017

## - Les outils lignes de commandes

Ces outils peuvent être difficiles à utiliser pour la plupart des biologistes, mais offrent presque toujours plus d'options pour l'exécution des programmes. Ils sont plus appropriés pour analyser des ensembles de données à grande échelle qui sont rencontrés actuellement en bioinformatique.

## - Les Outils Web (Web-Based Software)

Les outils Web, parfois appelés « point-and-click », ne nécessitent pas de connaissances en programmation et sont immédiatement accessibles à la communauté scientifique.

### 1.3. Domaines de la bioinformatique

Les domaines de la bioinformatique s'appuient fortement sur Internet pour accéder aux données de séquence, aux logiciels utiles pour analyser les données moléculaires et pour intégrer différents types de ressources et d'informations relatives à la biologie. Nous allons décrire une variété de sites Web. Dans un premier temps, nous nous concentrerons sur les principales bases de données accessibles au public qui servent de référentiels pour les données sur l'ADN et les protéines. Ceux-ci comprennent :

- Le Centre national d'information sur la biotechnologie (**NCBI**), qui héberge la GenBank et d'autres ressources;
- L'Institut européen de bioinformatique (**EBI**);
- Ensemble, qui comprend un navigateur génomique et des ressources pour étudier des dizaines de génomes;
- Le site de bioinformatique du génome de l'Université de Californie à Santa Cruz (**UCSC**), comprenant un navigateur Web et un navigateur de tableaux pour diverses espèces.

Les principaux avantages offerts par les sites Web sont un accès facile, des mises à jour rapides, une bonne visibilité pour la communauté scientifique et une facilité d'utilisation (étant donné que les compétences de d'algorithmique et de programmation ne sont pas nécessaires)

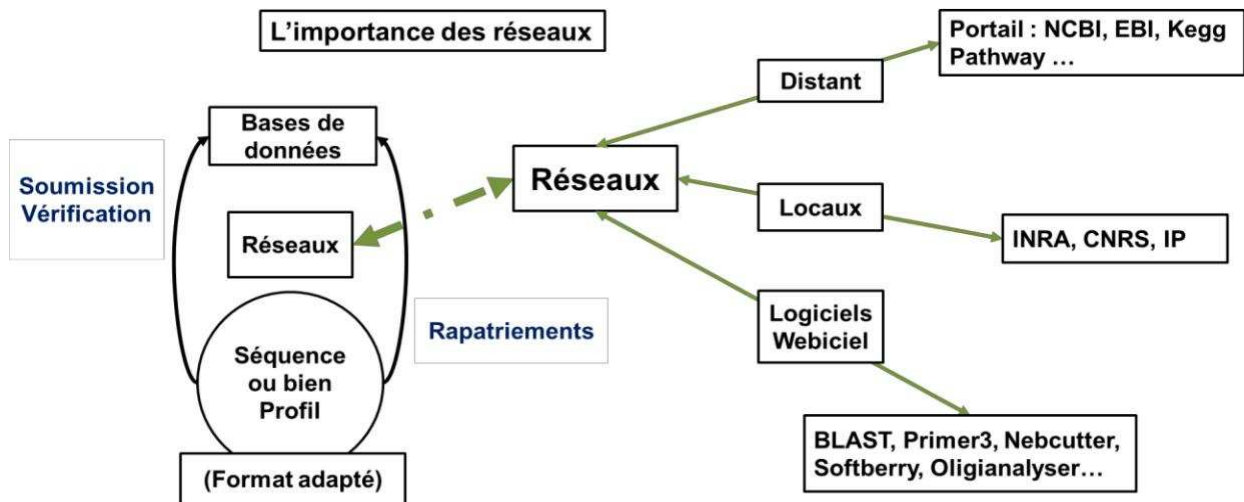
- **Stockage et Gestion des données** : Banques de données généralistes et spécialisées.
- **Structures moléculaires** : Visualisation, analyse, classification, prédiction.
- **Analyse de séquences** : Alignements, recherches de similarités, détection de motifs.
- **Génomique structurale** : Annotation des génomes, génomique comparative.
- **Génomique fonctionnelle** : Transcriptome, protéome, interactome.



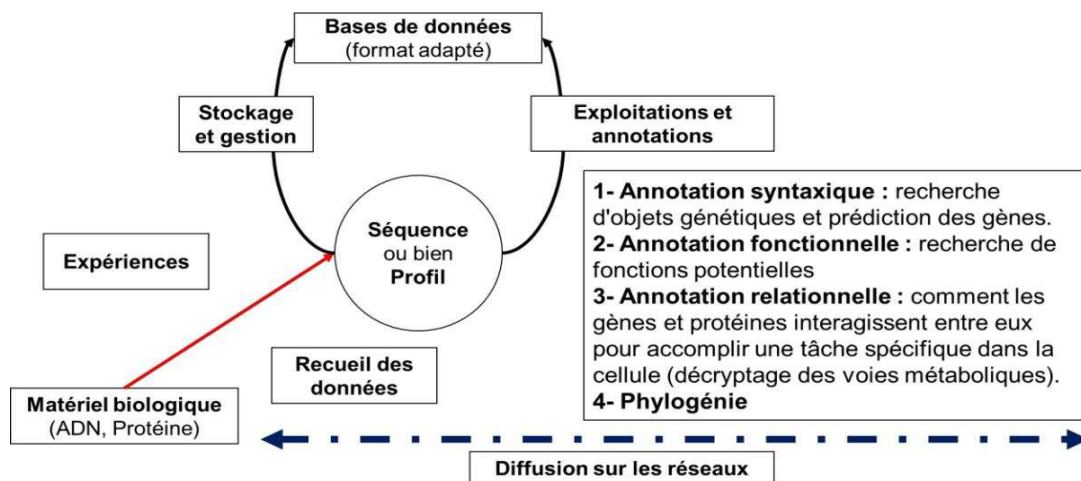
- **Phylogénie** : Relations évolutives entre gènes, entre génomes, entre organismes ; Inférence de scénarios évolutifs.
- **Analyse des réseaux biomoléculaires** : Réseaux métaboliques, d'interactions protéiques, de régulation génétique,...

#### 1.4. Étapes de travail en bioinformatique

Les principales étapes de travail en bioinformatique sont schématisées sur les figures 13 & 14.



**Figure 13** : Représentation schématique des procédés utilisés par la bioinformatique Fondrait Granger Brunet et al . 2017



**Figure 14** : Représentation schématique de l'importance des réseaux en bioinformatique. Fondrait Granger Brunet et al . 2017

# **Matériel et Méthodes**

## Matériel

Notre analyse *in silico* comprend trois activités principales :

- 1) Acquisition et organisation des données biologiques,
- 2) Conception de base de données après leur analyse, comparaison et modélisation,
- 3) Analyse des résultats obtenus dans la base de données lors de sa création.

### \*Accès aux banques de données d'abrie, ncbi, genbank

-Séquences nucléiques -Séquences protéiques

### \*Analyse de séquences

-Nucléiques -Protéiques

### \*Modélisation moléculaire

-Homologiques (similarités de séquences)

-Analogiques (similarités fonctionnelles)

-Contraintes (JohnWiley et *al.*, 2004)

La Base De Données est un ensemble de données organisé en vue de son utilisation par des programmes correspondant à des applications distinctes et de manière à faciliter l'évolution indépendante des données et des programmes.

Par exemple, on peut considérer la banque GenBank comme un énorme fichier contenant une suite d'enregistrement et pour chacun des champs spécifiques définis, avec une seule clé d'index comme entrée.

Par exemple, MICADO (MICRobial Advanced Database Organization) est une base de données relationnelle (système de gestion PostgreSQL), dédiée aux génomes microbiens.

Elle intègre notamment l'ensemble des séquences primaires microbiennes issues de Genbank, les génomes complets microbiens réannotés dans la banque Emglib et les données d'analyse fonctionnelle de la bactérie modèle *B. subtilis*.

Il existe un grand nombre de banques ou bases de données d'intérêt biologique. Cette introduction sera limitée à une présentation des principales banques de données publiques, basées sur la structure primaire des séquences.

### 1. Méthode de création de la base de données

Tout d'abord, le programme utilisé DB Browser SQLite a été utilisée pour créer notre base de données. Il a été utilisé pour ne pas avoir besoin d'une connexion à un serveur, car c'est simple et rapide, et il peut être placé dans un appareil sans avoir besoin d'un serveur hôte interne ou externe. (SQLite) Ce programme nous permet de créer, éditer et modifier des bases de données de ce type. Créer des tables de colonnes pour ajouter des données à l'intérieur des colonnes.

1/ Création d'une donnée nommée xx Puis une table a été créée "fafa" puis créé 13 colonnes  
 Chaque colonne a un nom. La première colonne est nommée Nom id de gene/  
 Loucus+ID/..... à la dernière colonne nommée ID l'identifiant contient des nombres de 1-60.  
 Voici le nombre de lignes dans la base de données. Afin d'utiliser son numéro d'identification  
 et voici comment l'utiliser ; Dans un script Python pour appeler n'importe quelle ligne, nous  
 voulons que ses données commencent uniquement à partir du numéro d'identification.

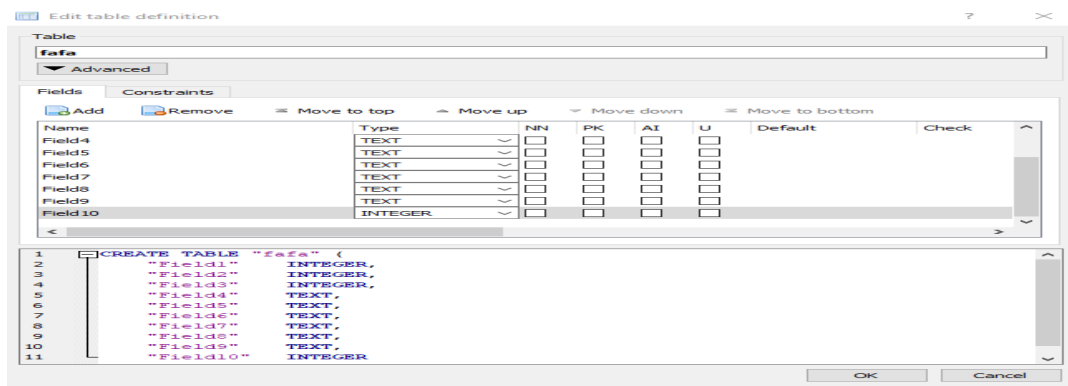


Figure 15 : Créer une table de base de données "fafa"

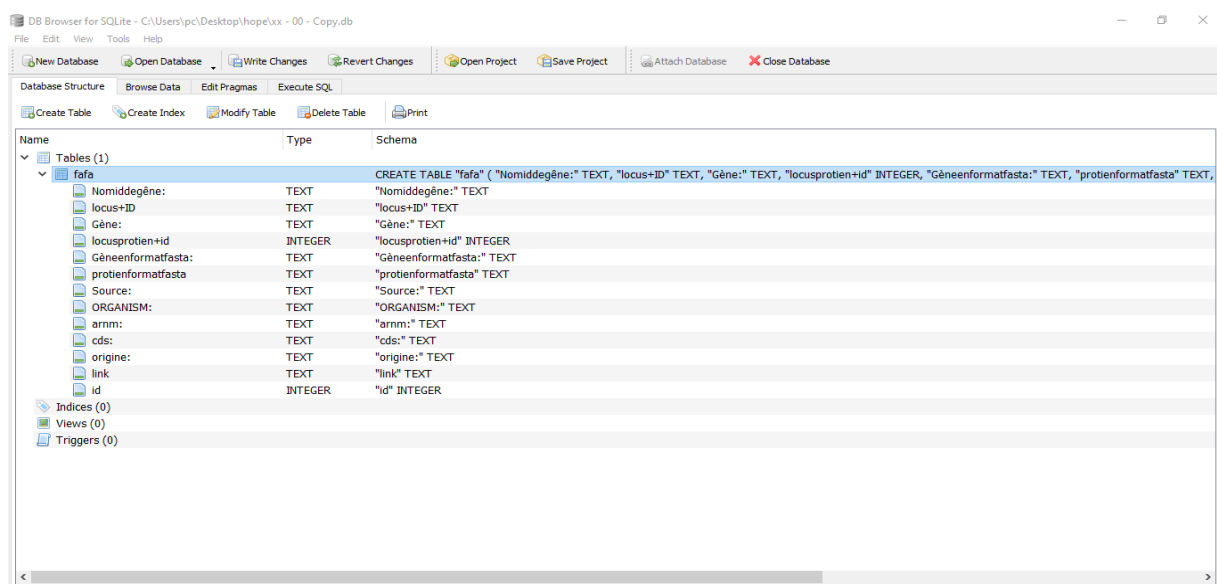


Figure 16 : Choix du type de données dans la table de la base de données "fafa"

## 2. Méthode de recherche des données biologiques (gènes /protéine)

Les données ont été placées dans la base de données extraite de la banque de données NCBI, genbank.



**Figure 17** : Recherche sur Google la banque de données genbank

Sur le portail NCBI/GENBANK :

<https://www.ncbi.nlm.nih.gov/gen...> Traduire cette page  
**GenBank Overview - NCBI - NIH**  
 26 avr. 2021 — GenBank<sup>®</sup> is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (Nucleic Acids Research, ...  
[Sequence Identifiers](#) · [Sample GenBank Record](#) · [Submitting data](#) · [PubMed/NCBI](#)

<https://www.ncbi.nlm.nih.gov> Traduire cette page  
**National Center for Biotechnology Information**  
 GenBank release 244.0 (6/26/2021) is now available on the NCBI FTP site. This release has 14.78 trillion bases and 2.46 billion records. The current release ...

et choisir dans le menu sélectionner :

GenBank	Nucleotide
GenBank	Gene
GenBank	Protein
GenBank	Genome

et dans la ligne de recherche saisir :

Gene	Gene	nac triticum aestivum	Search
Nucleotide	Nucleotide	nac triticum aestivum	Search
Protein	Protein	nac Triticum aestivum	Search
Genome	Genome	triticum aestivum[orgn]	Search

Nous avons utilisé aussi la banque de données uniprot pour trouver des protéines aussi et choisir

NCBI Resources How To Sign in to NCBI

Nucleotide [Nucleotide] | nac Triticum aestivum [Advanced] Search Help

**COVID-19 Information**  
[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

GenBank: MG587696.1  
 Send to: Change region shown  
 Customize view

**Triticum aestivum cultivar INRA8097 NAC transcription factor NAM-B1 (NAM-B1) gene, partial cds**  
 GenBank: MG587696.1  
[FASTA](#) [Graphics](#) [PopSet](#)

Go to: [v]

LOCUS MG587696 1595 bp DNA linear PLN 08-JAN-2018  
 DEFINITION *Triticum aestivum* cultivar INRA8097 NAC transcription factor NAM-B1 (NAM-B1) gene, partial cds.  
 ACCESSION MG587696  
 VERSION MG587696.1  
 KEYWORDS .  
 SOURCE *Triticum aestivum* (bread wheat)  
 ORGANISM *Triticum aestivum*  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Analyze this sequence  
 Run BLAST  
 Pick Primers  
 Highlight Sequence Features  
 Find in this Sequence

Related information  
 Protein

Figure 18: le gène NAM\_B1 de facteur de transcription NAC de *Triticum aestivum*

Sequence format fasta :

NCBI Resources How To Sign in to NCBI

Nucleotide [Nucleotide] Search Help

**COVID-19 Information**  
[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

FASTA: MG587696.1  
 Send to: Change region shown  
 Customize view

**Triticum aestivum cultivar INRA8097 NAC transcription factor NAM-B1 (NAM-B1) gene, partial cds**  
 GenBank: MG587696.1  
[GenBank](#) [Graphics](#) [PopSet](#)

>MG587696.1 *Triticum aestivum* cultivar INRA8097 NAC transcription factor NAM-B1 (NAM-B1) gene, partial cds  
 AGCAGATCACCTGCCCGCTCTCTCTCTCTCTCCCAACCGTCCCTGTAAACAATCCCACTCCGTTCCCT  
 TCCTTCACACTACCTAGAAGCTTTGGCAGTTGAGTTAGGTGCCACACAGGGGTCTCTGTGGATCA  
 TCCTGTGTGTTGTTGTTAGTACCTAGCTAGCTAGGGGAAGAAGATCTGATGAGGTCCATGGGCAGCTCCGAC  
 TCATCTTCGGGCTCGGCAAAAAGCAACGGGTATCACCATCAGCATCAGCCGGCCCTCCGACGGGG  
 GCTCGGCGCCGGAGCTCCCGCGGCTCCGGTTCACACCAACCTGATCATGACAGTACCGCTCAC  
 CAAGAAGAAGCGCCCAAGGGCCGCTCCCGCTCAACATCATCGCCGAGGTGGATCTCTACAAGTTCGAC  
 CCATGGGAGCTCCCGGTATGTTATGTCTATCTCTCGCCGCGCCGTCTTACTTTATCAAGCGCCGAA  
 ATTTTCGGTGAATTAATAATCGAATAATCATCCATCTCATGCTTATACTCTGTGACAGTATGAT  
 TTTTATATCTTCCAGTACACATGTGTAGATGGTTATGTATGTATCTGTCTGTCTTGTCTATCGG  
 CTCGGATCCGGATCCATCAGAAGAAGCGACCATCCGGGAGCAGGATGTACTTCTCAGCCCGCGGA  
 CCGCAAGTACCCCAACGGCGCGGCCAACCAGGGCGGCGACGTCCGGGTACTGGAAGCCACCGGACG  
 GACAAGCCTATCTGGCTTCGGGACCGGGTCCGGCTGGTCCGGGAGAGCTCCGGGTCAGAAGGGCGC  
 TCGTGTCTACCGGGGAGCGCCGAGGGCTCAAAACCACTGATCATGACAGTACCGCTCAC  
 CGAGGATCTGGCTCCACACCGCCACCAACCGACCGCCCGGTGACCGGGGAGCAGGCGCTGCTCC  
 TCTCTCAGGGTACGTACAGTGTGGTCCGACGGTCTAGCAGTATTAATTGCTCTCCAGCTAAATTAGG  
 GTATTGTTGATGTTGATGAAGTAAATATGTACGTCTCTCATCTGGACGACTGGGTGCTGTGCCG  
 CATCTACAGAAAGATCAACAAGGCCCGGCGGCGATCAGCAGAGGAGCAGGAGTCCGAGGACTCCCTG  
 GAGGACGGGTCAGCGGTACCGCTATGCTCAGCGGGGAGTGGCGGTTGAGCTGGCATGGAGCA  
 ACTACGCTTACCTTCACTGCTCATCATCAGGACAGCCATTTCCTGGACGGCTGTTCCAGCAGACGA  
 TGGCGGCTCTGGCGGGCCACCTCCCTGAGACACTAGCAGCGGCGGAGGGCGAGCCCGCTCCG  
 ACCAACAAGTTTCTCGCCCTCTGCTTCAACCCATTCAACTGGCTCGATGGTCAACCGGTCCGCTCC  
 TCCACAGGAAAGAAATTTCTCGGTTTACACGGAGCAGAAAGCTTGGCAATATGCTGCTCATGAG  
 GGCCGACATGGCTGGCGAGTGGACAAGGTTGGAGCAATGGCGTGAACGCCATG

Related information  
 Protein  
 PubMed  
 Taxonomy  
 Full text in PMC  
 PopSet

Recent activity  
 Turn Off Clear  
 NAC transcription factor NAM-B1, partial [Triticum aestivum] Protein

Figure 19: format fasta du gène NAM\_B1 de facteur de transcription NAC de *Triticum aestivum*

La protéine du même gène

NCBI Resources How To Sign in to NCBI

Protein [Protein] Search Help

**COVID-19 Information**  
[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

GenPept: AUJ88691.1  
 Send to: Change region shown  
 Customize view

**NAC transcription factor NAM-B1, partial [Triticum aestivum]**  
 GenBank: AUJ88691.1  
[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to: [v]

LOCUS AUJ88691 359 aa linear PLN 08-JAN-2018  
 DEFINITION NAC transcription factor NAM-B1, partial [Triticum aestivum].  
 ACCESSION AUJ88691  
 VERSION AUJ88691.1  
 DBSOURCE accession MG587696.1  
 KEYWORDS .  
 SOURCE *Triticum aestivum* (bread wheat)  
 ORGANISM *Triticum aestivum*  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Analyze this sequence  
 Run BLAST  
 Identify Conserved Domains  
 Highlight Sequence Features  
 Find in this Sequence

Related information

**Figure 20:** la protéine nam\_b1 du *Triticum aestivum*

Séquence format fasta :

**NAC transcription factor NAM-B1, partial [Triticum aestivum]**

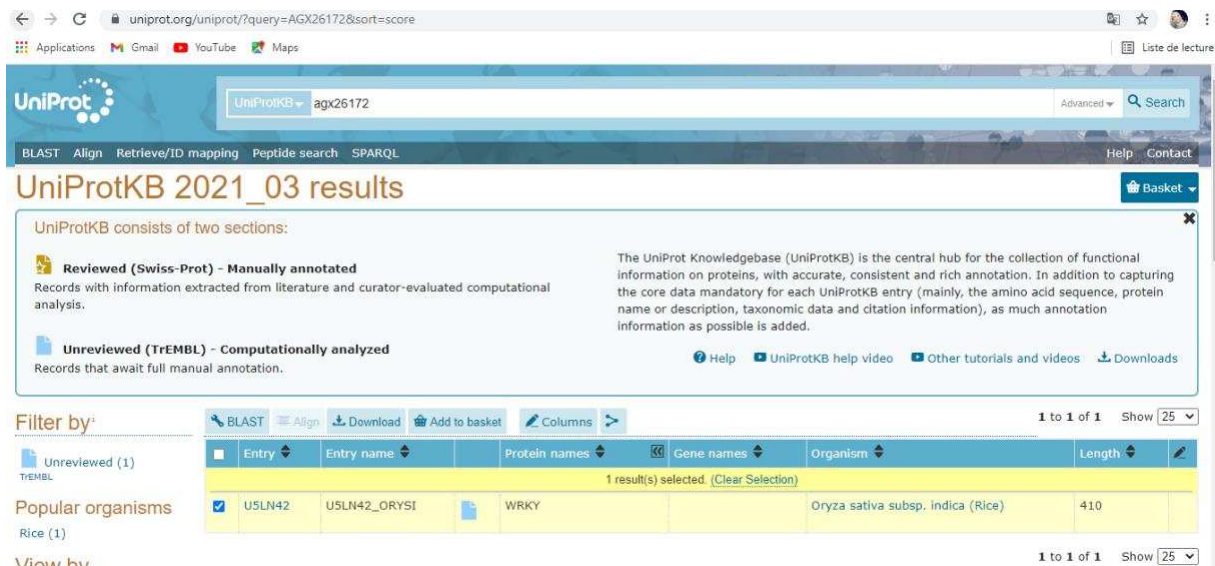
GenBank: AUJ88691.1

[GenPept](#) [Identical Proteins](#) [Graphics](#)

```
>AUJ88691.1 NAC transcription factor NAM-B1, partial [Triticum aestivum]
MGSSDSSSGSAQKATRYHHQHQP PPPQRGSAPELPPGFRFHPTDEELVVHYLKKKADKAPLPVNIIEVD
LYKFDPWELPEKATIGEQEWYFFSPDRKYPNGARPNRAATSGYWKATGDKPILASGTGCGLVREKLGV
KKALVFYRGKPKKGLKTNWIMHEYRLTDASGTTATNRPPTGGVRAAASLRDDWVLCRIYKKINKAA
AGDQQRSTECEDSVEDAVTAYPLYATAGMAGAAAAGS NYASPSLLHHQDSHF LDGLFTADDAGLSAGATS
LRHLAAAARASPAPTKQFLAPSSSTPFNWLDASPVGILPQARNFPGFNRSRNVGNMSSLSTADMAGAVDN
GGGNAVNAM
```

**Figure 21:** la séquence protéique nam\_b1 du *Triticum aestivum* format fasta.

La recherche de protéine par ID dans uniprot pour les protéines que nous n'avons pas trouvées



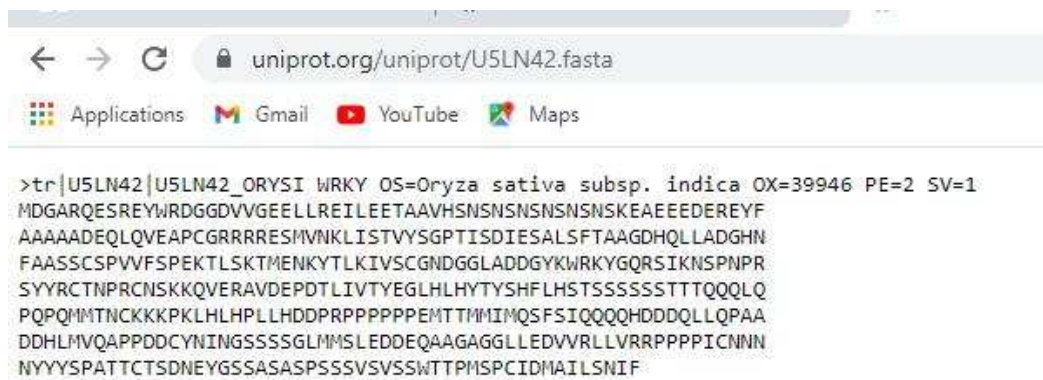
**Figure 22 :** La recherche de la protéine par ID dans uniprot pour les protéines que nous n'avons pas trouvées.

Cliquez sur Télécharger Format Fasta et copiez-le dans la base de données.



**Figure 23 :** Téléchargement Format Fasta la protéine dans la banque de données uniprot.

Séquence format fasta :

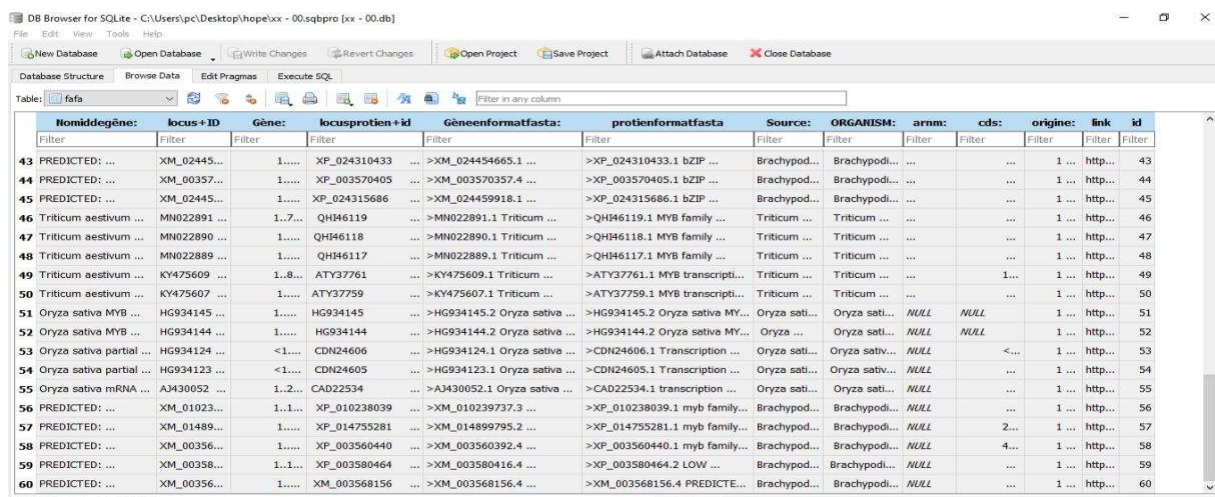


```
>tr|U5LN42|U5LN42_ORYSI WRKY OS=Oryza sativa subsp. indica OX=39946 PE=2 SV=1
MDGARQESREYWRDGGDVVGEELLREILEETAAVHSNSNSNSNSNSKEAEEDEREYF
AAAAAEQQLQVEAPCGRRRRESMVNKLISVYSGPTISDIESALSFTAAGDQQLADGHN
FAASSCSPVVFSPKTLTKMTENKYTLKIVSCNDGGLADDGYKWRKYGQRSIKNSPNPR
SYRRCNPRCNSKKQVERAVDEPDTLIVTYEGLHLHYTSHFLHSTSSSSSTTTQQQLQ
PDPQMMTNCKKKPKLHLHPLLHDDPRPPPPPEMTTMMIMQSFISIQQQHHDDQLLPAA
DDHLHVQAPPDDCYNINGSSSGLMMSLEDDEQAAGAGLLLEDVVRLLVRRPPPPICNNN
NYYSPATTCTSDNEYSSASAPSSSVSSSWTTPMSPCIDMAILSNIIF
```

**Figure 24:** séquence de la protéine WRKY d'*Oryza sativa*.

Et en copiant et collant nous mettons les données dont on a besoin dans la base de données qui a été créée

Nous obtenons cette base de données



nomiddegène:	locus+ID	Gène:	locusprotien+id	Gèneenformatfasta:	protienformatfasta	Source:	ORGANISM:	arm:	cds:	origine:	link	id
43 PREDICTED: ...	XM_02445...	1.....	XP_024310433 ...	>XM_024454665.1 ...	>XP_024310433.1 bZIP ...	Brachypod...	Brachypodi...			1 ...	http...	43
44 PREDICTED: ...	XM_00357...	1.....	XP_003570405 ...	>XM_003570357.4 ...	>XP_003570405.1 bZIP ...	Brachypod...	Brachypodi...			1 ...	http...	44
45 PREDICTED: ...	XM_02445...	1.....	XP_024315686 ...	>XM_024459918.1 ...	>XP_024315686.1 bZIP ...	Brachypod...	Brachypodi...			1 ...	http...	45
46 Triticum aestivum ...	MN022891 ...	1..7...	QH46119 ...	>MN022891.1 Triticum ...	>QH46119.1 MYB family ...	Triticum ...	Triticum ...			1 ...	http...	46
47 Triticum aestivum ...	MN022890 ...	1.....	QH46118 ...	>MN022890.1 Triticum ...	>QH46118.1 MYB family ...	Triticum ...	Triticum ...			1 ...	http...	47
48 Triticum aestivum ...	MN022889 ...	1.....	QH46117 ...	>MN022889.1 Triticum ...	>QH46117.1 MYB family ...	Triticum ...	Triticum ...			1 ...	http...	48
49 Triticum aestivum ...	KY475609 ...	1..8...	ATY37761 ...	>KY475609.1 Triticum ...	>ATY37761.1 MYB transcripti...	Triticum ...	Triticum ...		1...	1 ...	http...	49
50 Triticum aestivum ...	KY475607 ...	1.....	ATY37759 ...	>KY475607.1 Triticum ...	>ATY37759.1 MYB transcripti...	Triticum ...	Triticum ...			1 ...	http...	50
51 Oryza sativa MYB ...	HG934145 ...	1.....	HG934145 ...	>HG934145.2 Oryza sativa ...	>HG934145.2 Oryza sativa MY...	Oryza sati...	Oryza sati...	NULL	NULL	1 ...	http...	51
52 Oryza sativa MYB ...	HG934144 ...	1.....	HG934144 ...	>HG934144.2 Oryza sativa ...	>HG934144.2 Oryza sativa MY...	Oryza ...	Oryza sati...	NULL	NULL	1 ...	http...	52
53 Oryza sativa partial ...	HG934124 ...	<1...	CDN24606 ...	>HG934124.1 Oryza sativa ...	>CDN24606.1 Transcription ...	Oryza sati...	Oryza sativ...	NULL		<... 1 ...	http...	53
54 Oryza sativa partial ...	HG934123 ...	<1...	CDN24605 ...	>HG934123.1 Oryza sativa ...	>CDN24605.1 Transcription ...	Oryza sati...	Oryza sativ...	NULL		1 ...	http...	54
55 Oryza sativa mRNA ...	AJ430052 ...	1..2...	CAD22534 ...	>AJ430052.1 Oryza sativa ...	>CAD22534.1 transcription ...	Oryza sati...	Oryza sati...	NULL		1 ...	http...	55
56 PREDICTED: ...	XM_01023...	1..1...	XP_010238039 ...	>XM_010239737.3 ...	>XP_010238039.1 myb family...	Brachypod...	Brachypodi...	NULL		1 ...	http...	56
57 PREDICTED: ...	XM_01489...	1.....	XP_014755281 ...	>XP_014899795.2 ...	>XP_014755281.1 myb family...	Brachypod...	Brachypodi...	NULL	2...	1 ...	http...	57
58 PREDICTED: ...	XM_00356...	1.....	XP_003560440 ...	>XM_003560392.4 ...	>XP_003560440.1 myb family...	Brachypod...	Brachypodi...	NULL	4...	1 ...	http...	58
59 PREDICTED: ...	XM_00358...	1..1...	XP_003580464 ...	>XM_003580416.4 ...	>XP_003580464.2 LOW ...	Brachypod...	Brachypodi...	NULL		1 ...	http...	59
60 PREDICTED: ...	XM_00356...	1.....	XM_003568156 ...	>XM_003568156.4 ...	>XM_003568156.4 PREDICTE...	Brachypod...	Brachypodi...	NULL		1 ...	http...	60

**Figure 25 :** base de données complète.

### 3. Méthode de création du script python

1- Après avoir fini De la création de SQLite de base de données à l'aide du navigateur DB (SQLite) à la création d'une table appelée "fafa" et à l'insertion de données dans la base de données. nous avons passé à travailler sur le script Python. Nous devais d'abord travailler sur ce script. Installez la dernière version du programme Python ou Python interpréteur sur mon appareil [https://www. .python.org/downloads](https://www.python.org/downloads).

Dimension Installer la dernière version de python Utilisez également la bibliothèque coloram pour mettre en évidence les horodatages dans le script via `pip.py` <https://bootstrap.pypa.io/get-pip.py>.



Pour installer les bibliothèques Python nous avons utilisé Visual Studio Code et créé un fichier Python dans le même chemin que le fichier de données nous avons dû travailler dessus au début du fichier de script python on a appelé plusieurs bibliothèques dont nous allons besoin pour commencer à travailler sur le script.

1/ Comme on utilise une base de données SQLite, nous avons appelé la bibliothèque sqlite3 pour pouvoir gérer ce genre de base de données dans notre script python,

2/ Nous avons appelé la bibliothèque sys pour mieux contrôler la saisie d'un type dans un écran de script random. Cette bibliothèque a été appelée pour que l'heure apparaisse aléatoirement dans le script Depuis la bibliothèque aléatoire,

3/ Nous avons appelé "Choix" pour choisir l'heure au hasard on a appelé cette bibliothèque pour sélectionner,

4/ système d'exploitation pour Windows ou Linux time Cette bibliothèque est appelée pour contrôler,

5 / vitesse d'affichage d'impression dans le script,

6/ Nous avons utilisé la bibliothèque colorama pour les couleurs à l'intérieur de mon script et de l'intérieur nous avons utilisé la variable init pour contrôler plus d'impression XMA Header Image.

2- Tout d'abord, créez une fonction appelée logo() écran de démarrage ou bannière. Ensuite, créez un changement avec le nom clair Pour réinitialiser l'heure "\x1b[0m" Nous créons une deuxième variable qui est un tableau avec des nombres, chaque nombre représentant une couleur couleurs = [36, 32, 34, 35, 31, 37, ] Ces numéros seront récupérés sous une forme générale pendant un certain temps jusqu'à ce qu'ils nous apparaissent Nous créons une variable nommée x Contient des livres qui apparaîtront à l'écran pour N, ligne dans énumérer(x.split("\n")): nous utilisons pour et \n Pour qu'une impression sous l'autre apparaisse dans le bon sens Maintenant, nous pouvons utiliser la bibliothèque sys Il a son propre paramètre stdout À l'intérieur, nous trouvons random. choice (couleurs) Nous utilisons la bibliothèque aléatoire . Et sa propre variable de choix Et laissez-les être une variable matricielle qui contient pour nous un ensemble de nombres qui représentent le premier Chaque fois qu'un nombre est choisi au hasard, il se convertit en la forme générale de la couleur en Python, puis s'imprime.

```
def logo():
    clear = "\x1b[0m"
    colors = [36, 32, 34, 35, 31, 37 ]
```

**Figure 26:** fonction de démarrage d'écran logo ()

3-/ g = Avant.BLANC

r = Avant.ROUGE

b = Avant.ROUGE

w = Avant.JAUNE

Nous créons quatre variables sous forme de lettres, chaque lettre portant une couleur pour les copier plus tard dans bannière sous pour chaque famille.

```
g = Fore.WHITE
r = Fore.RED
b = Fore.RED
w = Fore.YELLOW
```

**Figure 27** : Les 4 variables forme lettre du couleur.

4-/ if os.name == "nt": os.system("cls") autre: os.system("effacer") par os . Bibliothèque  
Nous définissons le type de système qui s'exécutera dans notre script Au cas où ce n'était pas Windows La commande est appliquée cls Surtout en scannant un écran pour qu'on puisse voir les petits caractères Et s'il y avait un autre système ici, le manque de Linux, la commande pour effacer cet écran sera appliquée dégager

5-/ Nous créons quatre variables bannière bannière2 bannière3 bannière4 Ce sont des sous-fenêtres d'affichage pour chaque famille

6-/ Exécutez la première fonction dans le script de démarrage input\_message=logo() Puisque 4 familles, nous avons choisi dans chaque famille 15 Janes pour 3 créatures Nous créé 60 fonctions, chaque fonction intéressée m'a apporté des données spécifiques à une ligne de Database Montre le moi La première fonction que nous créons s'appelait NAC\_famille() En début de ligne, la variable de basse est monté conn sqlite3.connect("xx.db") où il y a une connexion avec ma base de données Qui est de type sqlite situé dans le même chemin que le fichier de script portant le nom xx.db Nous créons une deuxième variable nommée cur conn.cursor() Cela nous permet de manipuler la base de données du curseur pour créer un objet curseur () ici nous avons appelé la fonction Dans la troisième ligne de la fonction . Nous créons une autre variable nommée r\_set contient la valeur cur.execute Courons commande dans la base de données et il "SELECT \* FROM "fafa" WHERE id = '1'" Ici par cette commande et nous utilisons SÉLECTIONNER. Nous spécifions chaque donnée par \* De l'intérieur de la table "fafa" Celui à l'intérieur de la base de données est spécial pour nous O identifiant = '1' Ici, nous avons sélectionné les données les plus recherchées Où est ma colonne nommée id Il a une valeur de 1.

Et voici ce que signifie la première ligne de notre base de données Maintenant, nous créons une variable appelée record. Il contient toutes les données qui ont été récupérées wp utilise la fonction fetchall() C'est maintenant l'étape de l'impression des données qu'il a apportées de la base de données utilisant pour l'élément dans l'enregistrement : Une nouvelle variable nommée item Dans laquelle nous précisons le numéro de la colonne de mon père à partir de 0-11 Maintenant, nous faisons la première impression print (f"\u001b[31;1mNom id de gène \u001b[33;1m => {item[0]}") au début ligne d'impression f Nous l'ai utilisé ici car il y a des parenthèses à l'intérieur des parenthèses de la fonction d'impression Pour qu'aucune erreur ne se produise lors de l'impression Code couleur rouge \u001b[31;1m Symbole de la Lune jaune \u001b[33;1m Spécifiez le numéro de colonne à imprimer pour l'élément [0] Nous définissons plusieurs fonctions print . Sous l'autre chaque fois que nous mettons un nouveau numéro de ligne que nous voulons imprimer élément [1] élément [2] ..... Création de 60 fonctions avec la même forme à chaque fois qu'une nouvelle ligne de données est appelée depuis la base de données Change à chaque fois qu'un numéro de valeur de ligne WHERE id = '60' à WHERE id = '1'

```

147 input_message= logo()
148
149
150
151
152
153 #=====famille(NAC)=====
154 def NAC_famille():
155     conn = sqlite3.connect("xx.db")
156     cur = conn.cursor()
157     r_set=cur.execute("SELECT * FROM fafa WHERE id = '1'")
158     record = r_set.fetchall()
159     for item in record:
160         print(f"\u001b[31;1mNom id de gène \u001b[33;1m => {item[0]}")
161         print(f"\u001b[31;1mlocus+ID \u001b[33;1m => {item[1]}")
162         print(f"\u001b[31;1mGène \u001b[33;1m => {item[2]}")
163         print(f"\u001b[31;1mGène format fasta \u001b[33;1m => {item[3]}")
164         print(f"\u001b[31;1mlocus protien+ID \u001b[33;1m => {item[4]}")
165         print(f"\u001b[31;1m protien format fasta \u001b[33;1m => {item[5]}")
166         print(f"\u001b[31;1msource \u001b[33;1m => {item[6]}")
167         print(f"\u001b[31;1mORGANISM \u001b[33;1m => {item[7]}")
168         print(f"\u001b[31;1marnn \u001b[33;1m => {item[8]}")
169         print(f"\u001b[31;1mcds \u001b[33;1m => {item[9]}")
170         print(f"\u001b[31;1morigine \u001b[33;1m => {item[10]}")
171         print(f"\u001b[31;1mlinke \u001b[33;1m => {item[11]}")
172
173

```

**Figure 28** : Le code python du script base de données stade 1.

7-/ Puis nous avons créé quatre nouvelles fonctions sous le nom aaaa() bbbb() eeee() ffff() Chaque fonction est pour appeler et afficher la bannière Surtout pour chaque famille qui l'a créé aaaa() ----bannière+entrée()+if bbbb() ----bannière2+entrée()+if eeee() ----banner3+entrée()+if ffff() ----banner4+entrée()+if Nous expliquons le travail de la première fonction et le reste de celle-ci aaaa(). Au début de la ligne qui commence avec moi print



commande `\">{tool}\` Exécutez ensuite la fonction `aaaa()`. Ce qui le ramène tout droit devant `print(f"Désolé cette commande \">{user_input}\` N'est pas trouvé") `aya ()`. Au début de la fonction est exécutée `logo()`. Qui contient mes données sur un écran vertical en script Sur la deuxième ligne la variable `user_input` qui détient la valeur d'une fonction `input()`. Il donne à l'utilisateur quatre options.

```

aya2.py
kampp > htdocs > 2021 > pro1 > aya2.py
1 print(
2
3 print("=====")
4 print("\u001b[32mMenu [1] pour les familles (MYB) ")
5 print("\u001b[32mMenu [2] pour les gène de famille(MYB)")
6 tool = input("Choisissez [1-2] => ").strip()
7 if tool == "1":
8     ayaaya()
9 elif tool == "2":
10    ffff()
11 else:
12    print(f"Désolé cette commande \">{tool}\ N'est pas trouvé")
13    oooo4()
14 #=====  

15 #=====  

16 #=====  

17 #=====  

18 #=====  

19 #=====  

20 #=====  

21 #=====  

22 #=====  

23 #=====  

24 #=====  

25 #=====  

26 #=====  

27 #=====  

28 #=====  

29 #=====  

30 #=====  

31 #=====  

32 #=====  

33 #=====  

34 #=====  

35 #=====  

36 #=====  

37 #=====  

38 #=====  

39 #=====  

40 #=====  

41 #=====  

42 #=====  

43 #=====  

44 #=====  

45 #=====  

46 #=====  

47 #=====  

48 #=====  

49 #=====  

50 #=====  

51 #=====  

52 #=====  

53 #=====  

54 #=====  

55 #=====  

56 #=====  

57 #=====  

58 #=====  

59 #=====  

60 #=====  

61 #=====  

62 #=====  

63 #=====  

64 #=====  

65 #=====  

66 #=====  

67 #=====  

68 #=====  

69 #=====  

70 #=====  

71 #=====  

72 #=====  

73 #=====  

74 #=====  

75 #=====  

76 #=====  

77 #=====  

78 #=====  

79 #=====  

80 #=====  

81 #=====  

82 #=====  

83 #=====  

84 #=====  

85 #=====  

86 #=====  

87 #=====  

88 #=====  

89 #=====  

90 #=====  

91 #=====  

92 #=====  

93 #=====  

94 #=====  

95 #=====  

96 #=====  

97 #=====  

98 #=====  

99 #=====  

100 #=====  

101 #=====  

102 #=====  

103 #=====  

104 #=====  

105 #=====  

106 #=====  

107 #=====  

108 #=====  

109 #=====  

110 #=====  

111 #=====  

112 #=====  

113 #=====  

114 #=====  

115 #=====  

116 #=====  

117 #=====  

118 #=====  

119 #=====  

120 #=====  

121 #=====  

122 #=====  

123 #=====  

124 #=====  

125 #=====  

126 #=====  

127 #=====  

128 #=====  

129 #=====  

130 #=====  

131 #=====  

132 #=====  

133 #=====  

134 #=====  

135 #=====  

136 #=====  

137 #=====  

138 #=====  

139 #=====  

140 #=====  

141 #=====  

142 #=====  

143 #=====  

144 #=====  

145 #=====  

146 #=====  

147 #=====  

148 #=====  

149 #=====  

150 #=====  

151 #=====  

152 #=====  

153 #=====  

154 #=====  

155 #=====  

156 #=====  

157 #=====  

158 #=====  

159 #=====  

160 #=====  

161 #=====  

162 #=====  

163 #=====  

164 #=====  

165 #=====  

166 #=====  

167 #=====  

168 #=====  

169 #=====  

170 #=====  

171 #=====  

172 #=====  

173 #=====  

174 #=====  

175 #=====  

176 #=====  

177 #=====  

178 #=====  

179 #=====  

180 #=====  

181 #=====  

182 #=====  

183 #=====  

184 #=====  

185 #=====  

186 #=====  

187 #=====  

188 #=====  

189 #=====  

190 #=====  

191 #=====  

192 #=====  

193 #=====  

194 #=====  

195 #=====  

196 #=====  

197 #=====  

198 #=====  

199 #=====  

200 #=====  

201 #=====  

202 #=====  

203 #=====  

204 #=====  

205 #=====  

206 #=====  

207 #=====  

208 #=====  

209 #=====  

210 #=====  

211 #=====  

212 #=====  

213 #=====  

214 #=====  

215 #=====  

216 #=====  

217 #=====  

218 #=====  

219 #=====  

220 #=====  

221 #=====  

222 #=====  

223 #=====  

224 #=====  

225 #=====  

226 #=====  

227 #=====  

228 #=====  

229 #=====  

230 #=====  

231 #=====  

232 #=====  

233 #=====  

234 #=====  

235 #=====  

236 #=====  

237 #=====  

238 #=====  

239 #=====  

240 #=====  

241 #=====  

242 #=====  

243 #=====  

244 #=====  

245 #=====  

246 #=====  

247 #=====  

248 #=====  

249 #=====  

250 #=====  

251 #=====  

252 #=====  

253 #=====  

254 #=====  

255 #=====  

256 #=====  

257 #=====  

258 #=====  

259 #=====  

260 #=====  

261 #=====  

262 #=====  

263 #=====  

264 #=====  

265 #=====  

266 #=====  

267 #=====  

268 #=====  

269 #=====  

270 #=====  

271 #=====  

272 #=====  

273 #=====  

274 #=====  

275 #=====  

276 #=====  

277 #=====  

278 #=====  

279 #=====  

280 #=====  

281 #=====  

282 #=====  

283 #=====  

284 #=====  

285 #=====  

286 #=====  

287 #=====  

288 #=====  

289 #=====  

290 #=====  

291 #=====  

292 #=====  

293 #=====  

294 #=====  

295 #=====  

296 #=====  

297 #=====  

298 #=====  

299 #=====  

300 #=====  

301 #=====  

302 #=====  

303 #=====  

304 #=====  

305 #=====  

306 #=====  

307 #=====  

308 #=====  

309 #=====  

310 #=====  

311 #=====  

312 #=====  

313 #=====  

314 #=====  

315 #=====  

316 #=====  

317 #=====  

318 #=====  

319 #=====  

320 #=====  

321 #=====  

322 #=====  

323 #=====  

324 #=====  

325 #=====  

326 #=====  

327 #=====  

328 #=====  

329 #=====  

330 #=====  

331 #=====  

332 #=====  

333 #=====  

334 #=====  

335 #=====  

336 #=====  

337 #=====  

338 #=====  

339 #=====  

340 #=====  

341 #=====  

342 #=====  

343 #=====  

344 #=====  

345 #=====  

346 #=====  

347 #=====  

348 #=====  

349 #=====  

350 #=====  

351 #=====  

352 #=====  

353 #=====  

354 #=====  

355 #=====  

356 #=====  

357 #=====  

358 #=====  

359 #=====  

360 #=====  

361 #=====  

362 #=====  

363 #=====  

364 #=====  

365 #=====  

366 #=====  

367 #=====  

368 #=====  

369 #=====  

370 #=====  

371 #=====  

372 #=====  

373 #=====  

374 #=====  

375 #=====  

376 #=====  

377 #=====  

378 #=====  

379 #=====  

380 #=====  

381 #=====  

382 #=====  

383 #=====  

384 #=====  

385 #=====  

386 #=====  

387 #=====  

388 #=====  

389 #=====  

390 #=====  

391 #=====  

392 #=====  

393 #=====  

394 #=====  

395 #=====  

396 #=====  

397 #=====  

398 #=====  

399 #=====  

400 #=====  

401 #=====  

402 #=====  

403 #=====  

404 #=====  

405 #=====  

406 #=====  

407 #=====  

408 #=====  

409 #=====  

410 #=====  

411 #=====  

412 #=====  

413 #=====  

414 #=====  

415 #=====  

416 #=====  

417 #=====  

418 #=====  

419 #=====  

420 #=====  

421 #=====  

422 #=====  

423 #=====  

424 #=====  

425 #=====  

426 #=====  

427 #=====  

428 #=====  

429 #=====  

430 #=====  

431 #=====  

432 #=====  

433 #=====  

434 #=====  

435 #=====  

436 #=====  

437 #=====  

438 #=====  

439 #=====  

440 #=====  

441 #=====  

442 #=====  

443 #=====  

444 #=====  

445 #=====  

446 #=====  

447 #=====  

448 #=====  

449 #=====  

450 #=====  

451 #=====  

452 #=====  

453 #=====  

454 #=====  

455 #=====  

456 #=====  

457 #=====  

458 #=====  

459 #=====  

460 #=====  

461 #=====  

462 #=====  

463 #=====  

464 #=====  

465 #=====  

466 #=====  

467 #=====  

468 #=====  

469 #=====  

470 #=====  

471 #=====  

472 #=====  

473 #=====  

474 #=====  

475 #=====  

476 #=====  

477 #=====  

478 #=====  

479 #=====  

480 #=====  

481 #=====  

482 #=====  

483 #=====  

484 #=====  

485 #=====  

486 #=====  

487 #=====  

488 #=====  

489 #=====  

490 #=====  

491 #=====  

492 #=====  

493 #=====  

494 #=====  

495 #=====  

496 #=====  

497 #=====  

498 #=====  

499 #=====  

500 #=====  

501 #=====  

502 #=====  

503 #=====  

504 #=====  

505 #=====  

506 #=====  

507 #=====  

508 #=====  

509 #=====  

510 #=====  

511 #=====  

512 #=====  

513 #=====  

514 #=====  

515 #=====  

516 #=====  

517 #=====  

518 #=====  

519 #=====  

520 #=====  

521 #=====  

522 #=====  

523 #=====  

524 #=====  

525 #=====  

526 #=====  

527 #=====  

528 #=====  

529 #=====  

530 #=====  

531 #=====  

532 #=====  

533 #=====  

534 #=====  

535 #=====  

536 #=====  

537 #=====  

538 #=====  

539 #=====  

540 #=====  

541 #=====  

542 #=====  

543 #=====  

544 #=====  

545 #=====  

546 #=====  

547 #=====  

548 #=====  

549 #=====  

550 #=====  

551 #=====  

552 #=====  

553 #=====  

554 #=====  

555 #=====  

556 #=====  

557 #=====  

558 #=====  

559 #=====  

560 #=====  

561 #=====  

562 #=====  

563 #=====  

564 #=====  

565 #=====  

566 #=====  

567 #=====  

568 #=====  

569 #=====  

570 #=====  

571 #=====  

572 #=====  

573 #=====  

574 #=====  

575 #=====  

576 #=====  

577 #=====  

578 #=====  

579 #=====  

580 #=====  

581 #=====  

582 #=====  

583 #=====  

584 #=====  

585 #=====  

586 #=====  

587 #=====  

588 #=====  

589 #=====  

590 #=====  

591 #=====  

592 #=====  

593 #=====  

594 #=====  

595 #=====  

596 #=====  

597 #=====  

598 #=====  

599 #=====  

600 #=====  

601 #=====  

602 #=====  

603 #=====  

604 #=====  

605 #=====  

606 #=====  

607 #=====  

608 #=====  

609 #=====  

610 #=====  

611 #=====  

612 #=====  

613 #=====  

614 #=====  

615 #=====  

616 #=====  

617 #=====  

618 #=====  

619 #=====  

620 #=====  

621 #=====  

622 #=====  

623 #=====  

624 #=====  

625 #=====  

626 #=====  

627 #=====  

628 #=====  

629 #=====  

630 #=====  

631 #=====  

632 #=====  

633 #=====  

634 #=====  

635 #=====  

636 #=====  

637 #=====  

638 #=====  

639 #=====  

640 #=====  

641 #=====  

642 #=====  

643 #=====  

644 #=====  

645 #=====  

646 #=====  

647 #=====  

648 #=====  

649 #=====  

650 #=====  

651 #=====  

652 #=====  

653 #=====  

654 #=====  

655 #=====  

656 #=====  

657 #=====  

658 #=====  

659 #=====  

660 #=====  

661 #=====  

662 #=====  

663 #=====  

664 #=====  

665 #=====  

666 #=====  

667 #=====  

668 #=====  

669 #=====  

670 #=====  

671 #=====  

672 #=====  

673 #=====  

674 #=====  

675 #=====  

676 #=====  

677 #=====  

678 #=====  

679 #=====  

680 #=====  

681 #=====  

682 #=====  

683 #=====  

684 #=====  

685 #=====  

686 #=====  

687 #=====  

688 #=====  

689 #=====  

690 #=====  

691 #=====  

692 #=====  

693 #=====  

694 #=====  

695 #=====  

696 #=====  

697 #=====  

698 #=====  

699 #=====  

700 #=====  

701 #=====  

702 #=====  

703 #=====  

704 #=====  

705 #=====  

706 #=====  

707 #=====  

708 #=====  

709 #=====  

710 #=====  

711 #=====  

712 #=====  

713 #=====  

714 #=====  

715 #=====  

716 #=====  

717 #=====  

718 #=====  

719 #=====  

720 #=====  

721 #=====  

722 #=====  

723 #=====  

724 #=====  

725 #=====  

726 #=====  

727 #=====  

728 #=====  

729 #=====  

730 #=====  

731 #=====  

732 #=====  

733 #=====  

734 #=====  

735 #=====  

736 #=====  

737 #=====  

738 #=====  

739 #=====  

740 #=====  

741 #=====  

742 #=====  

743 #=====  

744 #=====  

745 #=====  

746 #=====  

747 #=====  

748 #=====  

749 #=====  

750 #=====  

751 #=====  

752 #=====  

753 #=====  

754 #=====  

755 #=====  

756 #=====  

757 #=====  

758 #=====  

759 #=====  

760 #=====  

761 #=====  

762 #=====  

763 #=====  

764 #=====  

765 #=====  

766 #=====  

767 #=====  

768 #=====  

769 #=====  

770 #=====  

771 #=====  

772 #=====  

773 #=====  

774 #=====  

775 #=====  

776 #=====  

777 #=====  

778 #=====  

779 #=====  

780 #=====  

781 #=====  

782 #=====  

783 #=====  

784 #=====  

785 #=====  

786 #=====  

787 #=====  

788 #=====  

789 #=====  

790 #=====  

791 #=====  

792 #=====  

793 #=====  

794 #=====  

795 #=====  

796 #=====  

797 #=====  

798 #=====  

799 #=====  

800 #=====  

801 #=====  

802 #=====  

803 #=====  

804 #=====  

805 #=====  

806 #=====  

807 #=====  

808 #=====  

809 #=====  

810 #=====  

811 #=====  

812 #=====  

813 #=====  

814 #=====  

815 #=====  

816 #=====  

817 #=====  

818 #=====  

819 #=====  

820 #=====  

821 #=====  

822 #=====  

823 #=====  

824 #=====  

825 #=====  

826 #=====  

827 #=====  

828 #=====  

829 #=====  

830 #=====  

831 #=====  

832 #=====  

833 #=====  

834 #=====  

835 #=====  

836 #=====  

837 #=====  

838 #=====  

839 #=====  

840 #=====  

841 #=====  

842 #=====  

843 #=====  

844 #=====  

845 #=====  

846 #=====  

847 #=====  

848 #=====  

849 #=====  

850 #=====  

851 #=====  

852 #=====  

853 #=====  

854 #=====  

855 #=====  

856 #=====  

857 #=====  

858 #=====  

859 #=====  

860 #=====  

861 #=====  

862 #=====  

863 #=====  

864 #=====  

865 #=====  

866 #=====  

867 #=====  

868 #=====  

869 #=====  

870 #=====  

871 #=====  

872 #=====  

873 #=====  

874 #=====  

875 #=====  

876 #=====  

877 #=====  

878 #=====  

879 #=====  

880 #=====  

881 #=====  

882 #=====  

883 #=====  

884 #=====  

885 #=====  

886 #=====  

887 #=====  

888 #=====  

889 #=====  

890 #=====  

891 #=====  

892 #=====  

893 #=====  

894 #=====  

895 #=====  

896 #=====  

897 #=====  

898 #=====  

899 #=====  

900 #=====  

901 #=====  

902 #=====  

903 #=====  

904 #=====  

905 #=====  

906 #=====  

907 #=====  

908 #=====  

909 #=====  

910 #=====  

911 #=====  

912 #=====  

913 #=====  

914 #=====  

915 #=====  

916 #=====  

917 #=====  

918 #=====  

919 #=====  

920 #=====  

921 #=====  

922 #=====  

923 #=====  

924 #=====  

925 #=====  

926 #=====  

927 #=====  

928 #=====  

929 #=====  

930 #=====  

931 #=====  

932 #=====  

933 #=====  

934 #=====  

935 #=====  

936 #=====  

937 #=====  

938 #=====  

939 #=====  

940 #=====  

941 #=====  

942 #=====  

943 #=====  

944 #=====  

945 #=====  

946 #=====  

947 #=====  

948 #=====  

949 #=====  

950 #=====  

951 #=====  

952 #=====  

953 #=====  

954 #=====  

955 #=====  

956 #=====  

957 #=====  

958 #=====  

959 #=====  

960 #=====  

961 #=====  

962 #=====  

963 #=====  

964 #=====  

965 #=====  

966 #=====  

967 #=====  

968 #=====  

969 #=====  

970 #=====  

971 #=====  

972 #=====  

973 #=====  

974 #=====  

975 #=====  

976 #=====  

977 #=====  

978 #=====  

979 #=====  

980 #=====  

981 #=====  

982 #=====  

983 #=====  

984 #=====  

985 #=====  

986 #=====  

987 #=====  

988 #=====  

989 #=====  

990 #=====  

991 #=====  

992 #=====  

993 #=====  

994 #=====  

995 #=====  

996 #=====  

997 #=====  

998 #=====  

999 #=====  

1000 #=====  


```

**Figure 30** : Code python de script base de données stade 3.

1-4 Dans la troisième ligne, nous avons une condition qui commence par si si `user_input == "1"`: Ainsi, si la valeur d'une variable est 1, La fonction sera exécutée `aaaa()`. Et c'est pour moi, nous lance ma propre interface, la première Fami Nousan, dans un script spécial pour mes données À la fin de la condition jusqu'à ce que nous arrivions à autre :

Dans le cas où l'utilisateur définit un numéro qui n'existe pas dans les conditions, ce message lui sera affiché indiquant que le numéro de mon père a été mal défini `print(f"Désolé cette commande \">{user_input}\` N'est pas trouvé"). En même temps, une fonction s'exécute `aya ()`. Ce qui redémarre une mauvaise page.

```

C: > Users > pc > Desktop > work2021 > work.py
1658 #====input=zzzzzz=aya=====
1659 # Check If Command Is Exists
1660 user_input = input('Choisissez [1-4] => ').strip().lower()
1661 # Command List
1662 commands_list = ["1", "2", "3", "4", "5"]
1663 if user_input in commands_list:
1664
1665     if user_input == "1":
1666         aaaa()
1667     elif user_input == "2":
1668         bbbb()
1669     elif user_input == "3":
1670         eeee()
1671     elif user_input == "4":
1672         ffff()
1673     elif user_input == "5":
1674
1675         print(f"Désolé cette commande \"{user_input}\" N'est pas trouvé")
1676         ayaaya()
1677
1678     else:
1679
1680         print(f"Désolé cette commande \"{user_input}\" N'est pas trouvé")
1681
1682         ayaaya()
1683
1684 else:
1685
1686     print(f"Désolé cette commande \"{user_input}\" N'est pas trouvé")
1687     ayaaya()
1688
1689

```

Figure 31: Code python de script base de données stade 4.

#### 4. Méthode d'Alignement multiple

Nous utilisons le programme Clustal Omega.

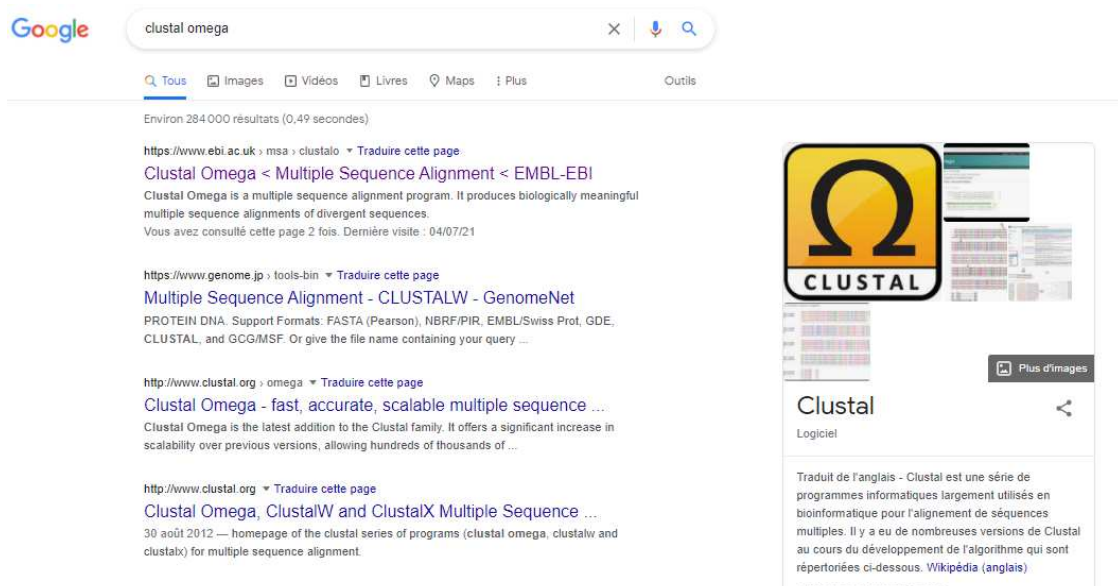


Figure 32 : Recherche dans Google a le programme Clustal omega.

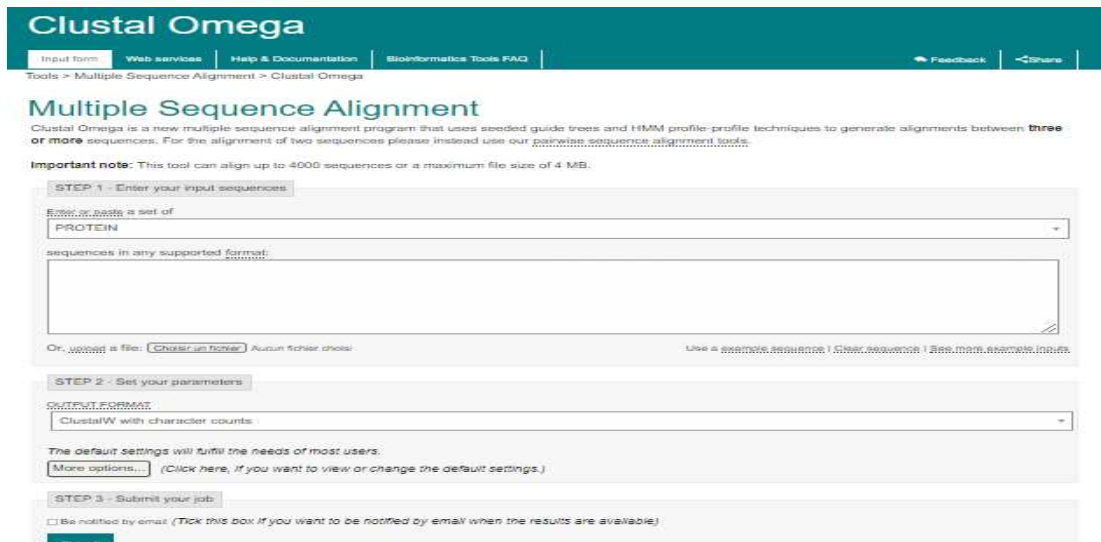


Figure 33 : Interface de programme Clustal omega.

Alignement multiple de séquences de protéines et d'ADN Nous choisissons à l'étape 1 :

Protéine, ADN, Ensuite on choisit le fichier contenant les séquences protéiques et géniques au format FASTA ou on copie toutes les séquences protéiques au format FASTA que l'on nomme en fonction des organismes et de la famille TF et on les colle dans la deuxième case.

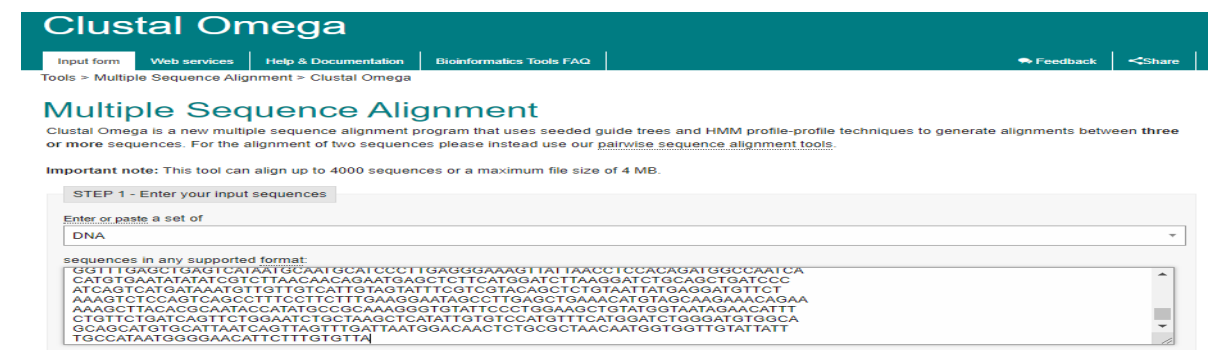
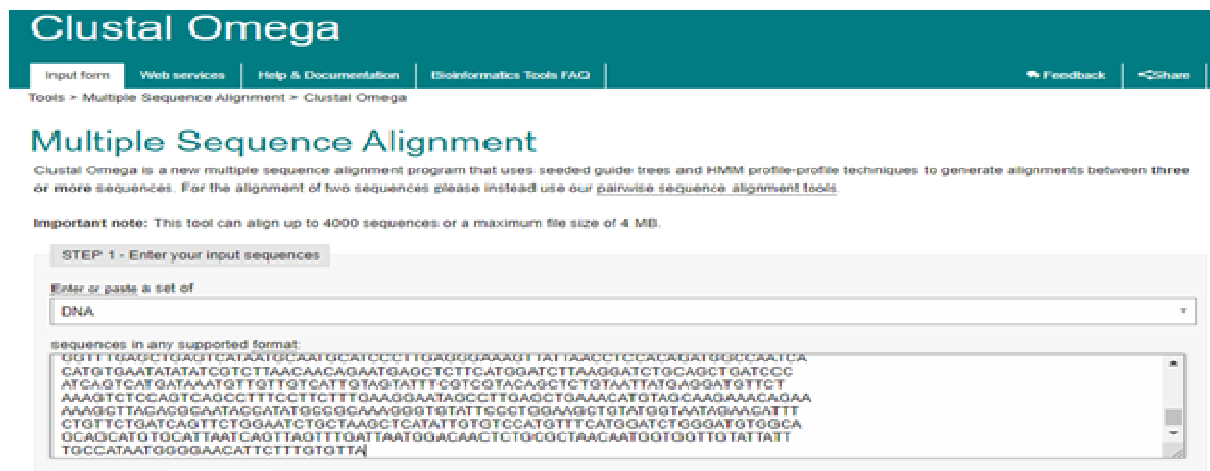


Figure 34 : 1<sup>ère</sup> étape de l'alignement multiple des séquences nucléaires.

Input form | Web services | Help & Documentation | Bioinformatics Tools FAQ | Feedback

Tools > Multiple Sequence Alignment > Clustal Omega

## Job Successfully Submitted

Your job has been successfully submitted. You will receive an email when the results are available...

If you don't receive any email, please check the status of your job by following this link: [toolresult.ebi?jobId=clustalo-E20210704-215952-0682-60341258-p1m](http://toolresult.ebi?jobId=clustalo-E20210704-215952-0682-60341258-p1m)

The job results will be available for 7 days.

**Figure 35 :** Prises des résultats d'alignement multiple.

A l'étape 2, nous choisissons le format de sortie que nous voulons, nous choisissons crystal wave le nombre de caractères dans les deux 3<sup>ème</sup> étapes : cliquez sur submit ;

STEP 2 - Set your parameters

OUTPUT FORMAT  
ClustalW with character counts

DEALIGN INPUT SEQUENCES	MBED-LIKE CLUSTERING GUIDE-TREE	MBED-LIKE CLUSTERING ITERATION	NUMBER of COMBINED ITERATIONS
no	yes	yes	default(0)
MAX GUIDE TREE ITERATIONS	MAX HMM ITERATIONS	ORDER	
default	default	aligned	

STEP 3 - Submit your job

Be notified by email (Tick this box if you want to be notified by email when the results are available)

Submit

**Figure 36:** Output format d'alignement multiple.

On attend le résultat ;

### \* Arbre phylogénétique

On l'obtient avec le même programme lorsque les résultats de l'alignement multiple apparaissent

## 5. Méthode de création du site web

- 1/ Dans un premier temps, nous avons préparé les fichiers du projet, images, textes, etc... ;
- 2/ Créer une infrastructure de site Web en HTML ;
- 3/ Ajouter un design de site Web en css ;
- 4 / Utiliser js ;
- 5/ Deuxième étape/ nous créons un dossier pour nos fichiers et y mettons le fichier html, css et js, les images que nous allons mettre sur le site, le fichier index2.py byteton et le fichier de base de données xx.db que nous utilisons xampp et placez les fichiers htdocs dans un dossier appelé dossier ayahouria#2/ nous travaillons sur du code HTML, nous



précisons le nom des cours et des scripts que nous voulons faire, ouvrons mon travail sur le code Visual Studio, J'écris du code html et écrivons le titre de mon projet ((Analyse in silico de quatre facteurs de transcription exprimés chez *Triticum aestivum* sous conditions de stress biotique et abiotique.)) puis nous lié les fichiers css en 3#/ nous sommes allés exécuter le schéma que nous voulions et avons mis les données pour mon projet puis enregistré le travail et vu à quoi ressemble le site sur le navigateur après avoir exécuté l'application xampp et cliqué sur démarrer pour l'hôte, tapé localhost /ayahouria 4/ travaillons sur csss, sur google fonts pour choisir la police et les couleurs et nous ajoutons dans le fichier css toute la liste statique que nous avons utiliser et coller et mettre il y avait un motif à tout et nous avons vu à quoi ressemblait le site 4#/. Ensuite nous avons utilisé js. Dans un fichier de script java et mettez-y une classe pour que l'animation fonctionne pour le déplacement et la compression.

Name	Date modified	Type	Size
rzt1	17/09/2021 14:47	File folder	
rzt2	18/09/2021 19:38	File folder	
rzt3	18/09/2021 20:12	File folder	
aya	17/09/2021 13:11	PNG File	89 KB
aya2	17/09/2021 13:14	Python File	85 KB
ayafinal	17/09/2021 13:14	Python File	85 KB
index	22/09/2021 13:41	Microsoft Edge H...	5 KB
index2	17/09/2021 13:37	Microsoft Edge H...	8 KB
Style_WheatBP	17/09/2021 13:11	CSS File	10 KB
xx	17/09/2021 13:14	DB File	1,060 KB

Figure 37 : Préparation des fichiers du projet.

```

1  <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
2  <html xmlns="http://www.w3.org/1999/xhtml"><head><meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
3  <title>Welcome to wheat: The Big Picture.</title>
4  <link rel="stylesheet" type="text/css" href="Style_WheatBP.css" media="screen">
5  </head>
6  <body>
7  <div id="Container">
8  <div id="Header">
9  <header>
10 <div>
11 </div>
12 </header>
13 <div>
14 <div id="Menu">
15 <a href="index2.html" class="menu">partie biologique</a>
16 <a href="xx.db" class="menu">base de donne</a>
17 <a href="aya2.py" class="menu">script base de donne de famille de gènes</a>
18 <a href="//rzt1" class="menu">Alignement multiple des gènes de famille MAC/WRCV/myb/biip triticum aestivum .oryza sativa .brachybo
19 <a href="//rzt1" class="menu">Arbre phylogénétique</a>
20 <a href="//rzt3" class="menu">Alignement multiple portique des gènes de famille mac/wexy triticum aestivum .oryza sativa </a>
21 <a href="//rzt1" class="menu">Arbre phylogénétique</a>
22 <a href="//rzt2" class="menu">Alignement multiple des gènes de famille MAC/WRCV/myb/biip triticum aestivum .oryza sativa</a>
23 <a href="//rzt2" class="menu">Arbre phylogénétique</a>
24 <a href="//rzt2" class="menu">Alignement multiple des gènes de famille MAC/WRCV/myb/biip triticum aestivum .oryza sativa</a>
25 <a href="//rzt2" class="menu">Arbre phylogénétique</a>
26 <a href="https://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/Documents/DOC_CerealsDB.php" class="menu">CerealsDB </a>
27 </div>
28 </div>
29 <div id="Content">
30 <h3>bienvenue a le mémoire de master 1 bioinformatique Analyse in silico de quatre facteurs de transcription exprimés chez Triticum aestivum sous

```

Figure 38 : Code HTML du site web du projet pour la structure.

## \* Style global (Code css+js )

```

Explorer (Ctrl+Shift+E)  index.html  # Style_WheatBP.css  X
C: > xampp > htdocs > 2021 > pro1 > # Style_WheatBP.css > ...
1  /*****
2  **** Global Styles ****
3  *****/
4
5  html
6  {
7      background-color: white;
8  }
9
10 body
11 {
12     color: black;
13     background-color: transparent;
14 }
15
16 body,p,h1,h2,h3,h4,table,td,th,ul,ol,textarea,input
17 {
18     font-family: "Trebuchet MS", sans-serif;
19     font-size: 14px;
20     line-height: 160%;
21     margin: 0px;
22     padding: 0px;
23 }
24
25
26
27 h1
28 {
29     font-size: 350%;
30     font-weight: normal;
31     font-style: italic;

```

Figure 39 : Style global (Code css+js ) du projet pour le DESSIN.

```

div#Menu
{
    width: 18%;
    min-height: 450px;
    padding: 10px 10px 10px 10px;
    margin: 10px 0px 0px 0px;
    float: left;
}

a.menu
{
    display: block;
    border-style: solid;
    border-width: 2px;
    border-top-color: #DFDFDF;
    border-right-color: #666;
    border-bottom-color: #333;
    border-left-color: #858585;
    font-size: 80%;
    font-weight: bold;
    color: white;
    width: 165px;
    text-align: center;
    padding: 4px;
    text-decoration: none;
    text-transform: uppercase;
}

a.menu:link, a.menu:visited

```

Figure 40 : Style global (Code css+js ) de projet pour le DESSIN.

# Résultats et discussion

## 1. Base de données complète

Une base de données complète contenant toutes les informations sur les familles de gènes sélectionnées pour cette étude, NAC, WRKY, bZIP et MYB de trois organismes *Triticum aestivum*, *Oryza sativa* et *Brachypodium* qui contient 60 gènes.

nomiddegène	locus+id	gène	locusproten+id	gèneenformatfasta	protenformatfasta	Source	ORGANISM	arm	cds	origine	link	id
43 PREDICTED: ...	XM_02445...	1.....	XP_024310433 ...	>XM_024454665.1 ...	>XP_024310433.1 bZIP ...	Brachypod...	Brachypodi...	...	...	1 ...	http...	43
44 PREDICTED: ...	XM_00357...	1.....	XP_003570405 ...	>XM_003570357.4 ...	>XP_003570405.1 bZIP ...	Brachypod...	Brachypodi...	...	...	1 ...	http...	44
45 PREDICTED: ...	XM_02445...	1.....	XP_024315686 ...	>XM_024459918.1 ...	>XP_024315686.1 bZIP ...	Brachypod...	Brachypodi...	...	...	1 ...	http...	45
46 Triticum aestivum ...	MN022891 ...	1..7...	QH46119 ...	>MN022891.1 Triticum ...	>QH46119.1 MYB family ...	Triticum ...	Triticum ...	...	...	1 ...	http...	46
47 Triticum aestivum ...	MN022890 ...	1.....	QH46118 ...	>MN022890.1 Triticum ...	>QH46118.1 MYB family ...	Triticum ...	Triticum ...	...	...	1 ...	http...	47
48 Triticum aestivum ...	MN022889 ...	1.....	QH46117 ...	>MN022889.1 Triticum ...	>QH46117.1 MYB family ...	Triticum ...	Triticum ...	...	...	1 ...	http...	48
49 Triticum aestivum ...	KY475609 ...	1..8...	ATY37761 ...	>KY475609.1 Triticum ...	>ATY37761.1 MYB transcripti...	Triticum ...	Triticum ...	...	1..	1 ...	http...	49
50 Triticum aestivum ...	KY475607 ...	1.....	ATY37759 ...	>KY475607.1 Triticum ...	>ATY37759.1 MYB transcripti...	Triticum ...	Triticum ...	...	...	1 ...	http...	50
51 Oryza sativa MYB ...	HG934145 ...	1.....	HG934145 ...	>HG934145.2 Oryza sativa ...	>HG934145.2 Oryza sativa MY...	Oryza sati...	Oryza sati...	NULL	NULL	...	http...	51
52 Oryza sativa MYB ...	HG934144 ...	1.....	HG934144 ...	>HG934144.2 Oryza sativa ...	>HG934144.2 Oryza sativa MY...	Oryza sati...	Oryza sati...	NULL	NULL	...	http...	52
53 Oryza sativa partial ...	HG934124 ...	<1....	CN24606 ...	>HG934124.1 Oryza sativa ...	>CN24606.1 Transcription ...	Oryza sati...	Oryza sativ...	NULL	NULL	<...	http...	53
54 Oryza sativa partial ...	HG934123 ...	<1....	CN24605 ...	>HG934123.1 Oryza sativa ...	>CN24605.1 Transcription ...	Oryza sati...	Oryza sativ...	NULL	NULL	...	http...	54
55 Oryza sativa mRNA ...	A3430052 ...	1..2...	CAD22534 ...	>A3430052.1 Oryza sativa ...	>CAD22534.1 transcription ...	Oryza sati...	Oryza sati...	NULL	NULL	...	http...	55
56 PREDICTED: ...	XM_01023...	1.....	XP_010238039 ...	>XM_010239737.3 ...	>XP_010238039.1 myb family...	Brachypod...	Brachypodi...	NULL	NULL	...	http...	56
57 PREDICTED: ...	XM_01489...	1.....	XP_014755281 ...	>XM_014899795.2 ...	>XP_014755281.1 myb family...	Brachypod...	Brachypodi...	NULL	2...	...	http...	57
58 PREDICTED: ...	XM_00356...	1.....	XP_003560440 ...	>XM_003560392.4 ...	>XP_003560440.1 myb family...	Brachypod...	Brachypodi...	NULL	4...	...	http...	58
59 PREDICTED: ...	XM_00358...	1..1...	XP_003580464 ...	>XM_003580416.4 ...	>XP_003580464.2 LOW ...	Brachypod...	Brachypodi...	NULL	...	...	http...	59
60 PREDICTED: ...	XM_00356...	1.....	XM_003568156 ...	>XM_003568156.4 ...	>XM_003568156.4 PREDICTE...	Brachypod...	Brachypodi...	NULL	...	...	http...	60

Figure 41 : Base de données complète des 4 facteurs de transcription NAC, WRKY, bzip et MYB.

## 2. Résultats du scripte



Figure 42 : Interface du script de la base de données.



Figure 43 : Interface du script de la base de données, famille du facteur de transcription NAC.

2\*

```
Choisissez [1-15] => 12
Nom id de gène => Brachypodium distachyon NAC domain-containing protein 67 (LOC100837540), mRNA
locus+ID => XM_024456209 989 bp mRNA linear PLN 27-MAR-2018
Gène => 1..989
/gene="LOC100837540"
/Note="Derived by automated computational analysis using
gene prediction method: Gnomon. Supporting evidence
includes similarity to: 1 Protein, and 100% coverage of
the annotated genomic feature by RNAseq alignments"
/db_xref="GeneID:100837540"
61 cttcgatcat ggcaatggcg gtgcggaatc tgcgaccggg ctactgtttc agaccacaag
121 ccctcgagct gtcgaacac ttctctgccc ccaaggcctt cggccgagac gtccctcccg
181 gctctgctgc ggaggcgctg gacgtcttct cgcggagccc cgaagcgctg ctccccctca
241 gcagacaaca caggaagggg cacggggagg tctggggcta cttcttcgag ccccagcccg
301 cggcagggg ggccgtggg ggcgcggcg ggtgcgtgct cccgtacggc cccgtgaaag
361 cgtaccgggg cggggacggg gaggcggctc ctttcggcca gaggttcggg taccacatgg
421 tccgctcac catcgggag gtctcgccg cggccagggg ggcgtggcgg cagacgttgt
481 ggctgatgac ggagtaccgg ctcaacaagg gtgccggcgc cttcggcaac gcgcagccgg
541 gcgaccacaa ggtgaacatg gattgcgtgg tccgcaaggt cttcaccagg tccgaggtcg
601 ctctcgccga gcaactgccc tgcctgccc gctcagcga cgaaggagac acgtgctcca
661 actagccag cgggacgag gaagccgctt actccggcca ggaagggcgg cgaagccggc
721 ctgatttag ctgagacttg agacactga gacgagcga cggatcgtg atcggcggat
781 tccgttcgtg tctcgtgta attttgcga agcgcgaata gtcgagaaaa ctagtggaa
841 tttttctgc tcaacttttt ctgcaggccg tgtttctgag aaagaaagaa acacatcccg
901 aatttcagc tgtttgatcc atgatcagtt catgcacaaa tttcatctcc cgacagcaaa
961 gagcaatgca gtaaccgga ttttcctc
Gène format fasta => >>XM_024456209.1 PREDICTED: Brachypodium distachyon NAC domain-containing protein 67 (LOC100837540), mRNA
AAATCCAACAAGCCTCCCTCCCAACAAGTTCTAGATATACCTAGTTTCGGTTCGCAACTCTTCGATCAT
GGCAATGGCCGTGGGGAATCTGCGACCGGGCTACGTGTTGAGACCAANGCCCTCGAGCTCGTTCGAACAC
TTCTGTGTTCCCAAGGCCTGGGCGGAGGTCCTGCGCCGCTTGGTGGGAGAGGGCTGGAGGTCCTTCT
CCGCGAGCCCGACCGCTGCTCCCTTTCAGCAGCAACACAGGAAGGGGACAGGGGAGGTCGGGGCTA
CTTCTTCGCGCCAGCCCGCGGCGAGGGGGCCGTGGCGGCGCCGGGCGGCTGGCTCCCGTACGCG
CCCGTGAAGCGCTACCGCGCGGGGACGGGGAGGCGGTGCCTTCGCGCAGAGGTTCCGCTACCCATGG
TCCGCTCACCATCGGGAGGTCGCGCGCGCGGACGGGGGGCGTGGCGCAGACGTTGTGGCTGATGAC
GGAGTACCGGCTCAACAAGGTCGCGCGGCTTCCGCAAGCGGCGAGCGGGGACCAAGGTGAACATG
GATTGCGTGGTCCGAAGGTTTACCACAACTCCGAGGTCGCTCCGCGCAGCCACTGCCCTCGCTGCC
GCTCCAGCGACGAGGAGACGCTGCTCCAACACGCCAGCGCGGACGAGGAAGCCGGCTACTCCGGCCA
GGAGCAGCGCGGAGCGGCTCGATTTAGGTGAGACTTGAGACGACTGAGACGAGCGCACGGATCCGTT
ATCGCGGATTCGGTTCGTGCTCGCTGTAATTTTCGAAAGCGCAATTAGTCGCAAGAACTAGTAGGAA
```

Figure 44 : Interface du script de la base de données du *Brachypodium*.

```
901 aattttcagc tgtttgatcc atgatcagtt catgcacaaa tttcatc
961 gagcaatgca gtaaccgga ttttcctc
linke => https://www.ncbi.nlm.nih.gov/nuccore/XM_024456209
=====
Menu [1] pour les familles (NAC)
Menu [2] pour les gène de famille(NAC)
Choisissez [1-2] => 12
Désolé cette commande "12" N'est pas trouvé
=====
Menu [1] pour les familles (NAC)
Menu [2] pour les gène de famille(NAC)
Choisissez [1-2] =>
```

Figure 45 : Interface du script de la base de données commande n'est pas trouvé.

3\*

```
Choisissez [1-4] => 555555
Désolé cette commande "555555" N'est pas trouvé

S.O.B.F.C

+-----+
| Script de base de données de famille de gènes |
+-----+

1) Afficher tous les gènes d'une famille(NAC)
2) Afficher tous les gènes d'une famille(Wrky)
3) Afficher tous les gènes d'une famille(bZIP)
4) Afficher tous les gènes d'une famille(MYB)
```

Figure 46: Interface du script de la base de données commande n'est pas trouvé et retour au premier composé.

### 3. Résultats des alignements

L'Alignement multiple protéique et l'Arbre phylogénétique :

L'alignement multiple des familles de gènes NAC, WRKY, myb et bzip chez *Triticum aestivum*, *Oryza sativa* et *Brachyrodium*.



**Figure 47 :** Alignement multiple des gènes NAC, WRKY, myb et bzip chez *Triticum aestivum*, *Oryza sativa* et *Brachyrodium*.

On clique sur show colors et on obtient le résultat suivant :



12nacBrachypodium	-----LRPGYVFRPKALE-----LVEHFLVPK-----ALGRDVLPG-----	38
13nacBrachypodium	-----LPSGFKFRPSDKD-----LILELYYR-----SIDEPL-PW-----	38
11nacBrachypodium	-----VPPGFRFHPTDEE-----LVDYLRKK-----VASRRI-DL-----	39
3nacTriticum	-----LPPGFRFYPTDEE-----LIVHYLRK-----AAAAPC-PA-----	54
14nacBrachypodium	-----LPPGFQFHPTDEE-----LILHYLRN-----AASVPC-PV-----	69
10nacOryza	-----LPPGFRFHPTDEE-----LIVHYLRNQ-----AASVKC-PV-----	40
1nacTriticum	-----LPPGFRFHPTDEE-----LVVHYLKK-----ADKAPL-PV-----	63
4nacTriticum	-----LPPGFRFHPTDEE-----LVVHYLKK-----ADKAPL-PV-----	63
5nacTriticum	-----LPPGFRFHPTDEE-----LVVHYLKK-----ADKAPL-PV-----	63
6nacOryza	-----LPPGFRFHPTDEE-----LVEHYLCRK-----AAGQRL-PV-----	46
7nacOryza	-----	0
15nacBrachypodium	-----APPPYHIDC-----YSYFVP-----T-----	25
8nacOryza	-----LPPGFRFHPTDAE-----VILSYLLQK-----FLNPSF-TS-----	48
2nacTriticum	-----LPPGFRFYPSDQE-----LVCHYLK-----VTNERA-SQ-----	39
9nacOryza	-----LPPGFRFHPTDEE-----LITHYLAKK-----VADARF-AA-----	41
58mybBrachypodium	-----	0
59mybBrachypodium	-----	0
56mybBrachypodium	-----	0
57mybBrachypodium	-----HM--YHA--KKFS--V-PFA--PQRAQN	19
60mybBrachypodium	-----HM--YHA--KKFS--V-PFA--PQRAQN	19
51mybOryza	-----CSPIRSAGPCWSYACLGNEL-----WQTCQQQETFKSILC-LWL-----CGG	92
17wrkyTriticum	-----	0
19wrkyTriticum	-----	0
18wrkyTriticum	-----	0
22wrkyOryza	-----	0
55mybOryza	-----AAECFPYRTEV-----QCLHRWQKVL-----NPELIKG-PWT-----	35
50mybTriticum	TQRSSGATMPCSVKLDL-----DL-LHHEET-----AEEIRRQ-PWT-----	45
49mybTriticum	-----MG-RAPC-----CE-----KMGKRG-PWT-----	18
47mybTriticum	-----MG-RQPC-----CD-----KVGKKG-PWT-----	18
46mybTriticum	-----MG-RQPC-----CD-----KVGKKG-PWT-----	18
48mybTriticum	-----MG-RQPC-----CD-----KVGKKG-PWT-----	18
41bzpBrachypodium	-----	0
34bzpTriticum	-----	0
35bzpTriticum	-----	0
40bzpOryza	-----	0

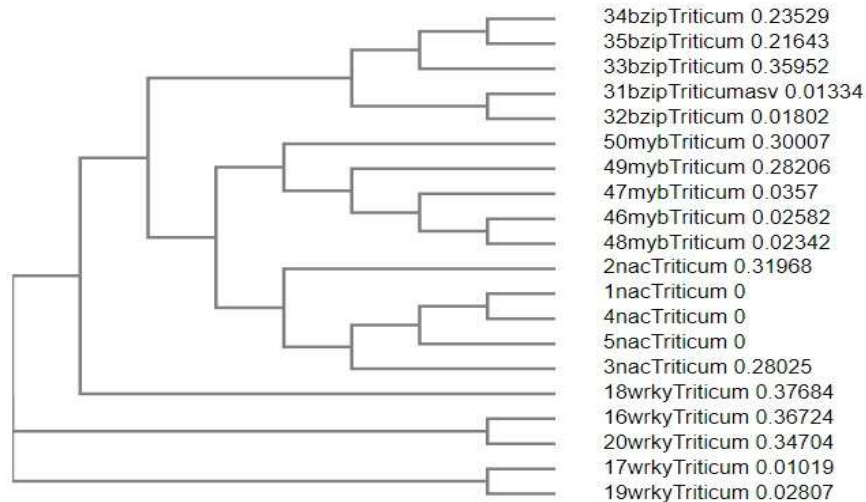
**Figure 48 :** Alignement multiple des gènes NAC, WRKY, myb et bzip chez *Triticum aetivum*, *Oryza sativa* et *Brachyodidium*.

Pour obtenir l'arbre phylogénétique cliquez sur [phylogentic tree](#) ;

## Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

Branch length:  Cladogram  Real



**Figure 49 :** Arbre phylogénétique des gènes NAC, WRKY, MYB et bZIP chez le *Triticum aetivum*

Alignement multiple nucléique :



Figure 50 : Alignement multiple des gènes NAC, WRKY, myb et bzip chez *Triticum aestivum*, *Oryza sativa* et *Brachypodium*.

## Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

Branch length:  Cladogram  Real

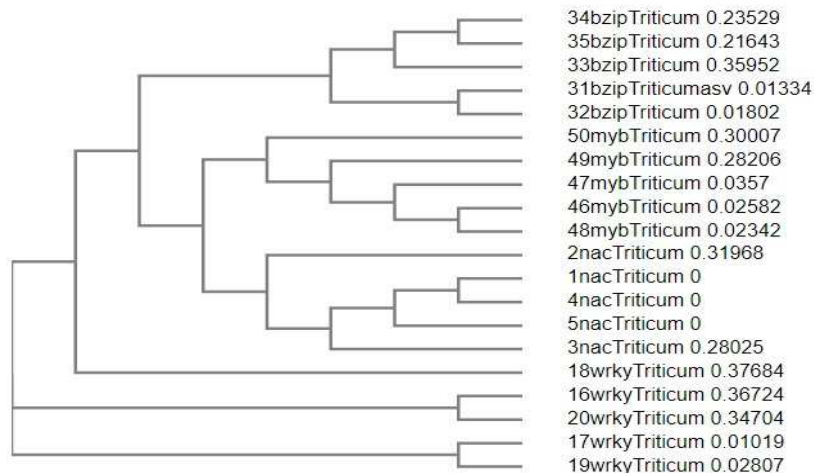
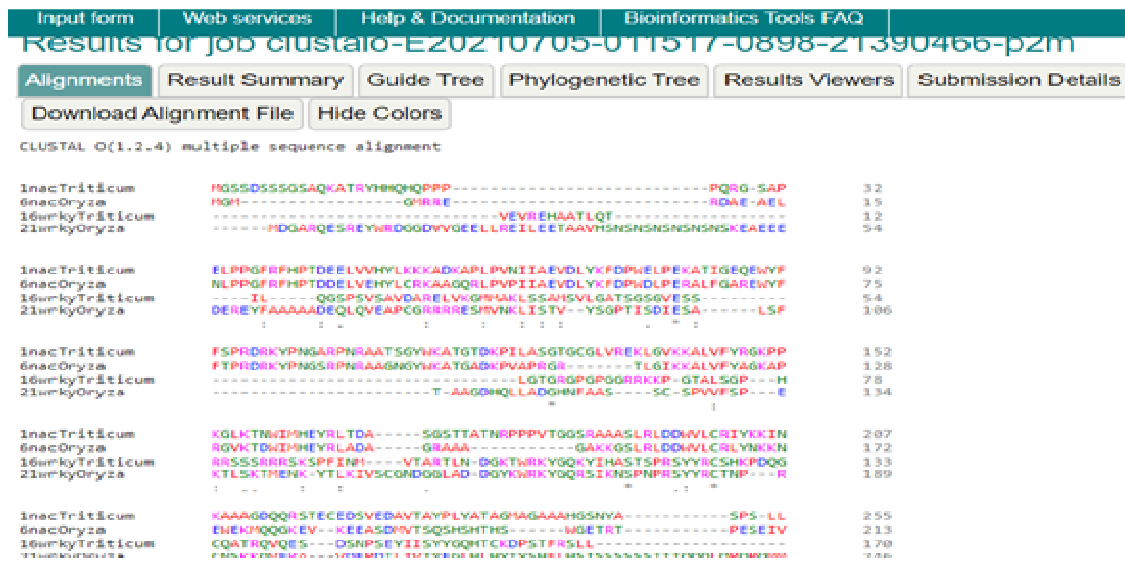


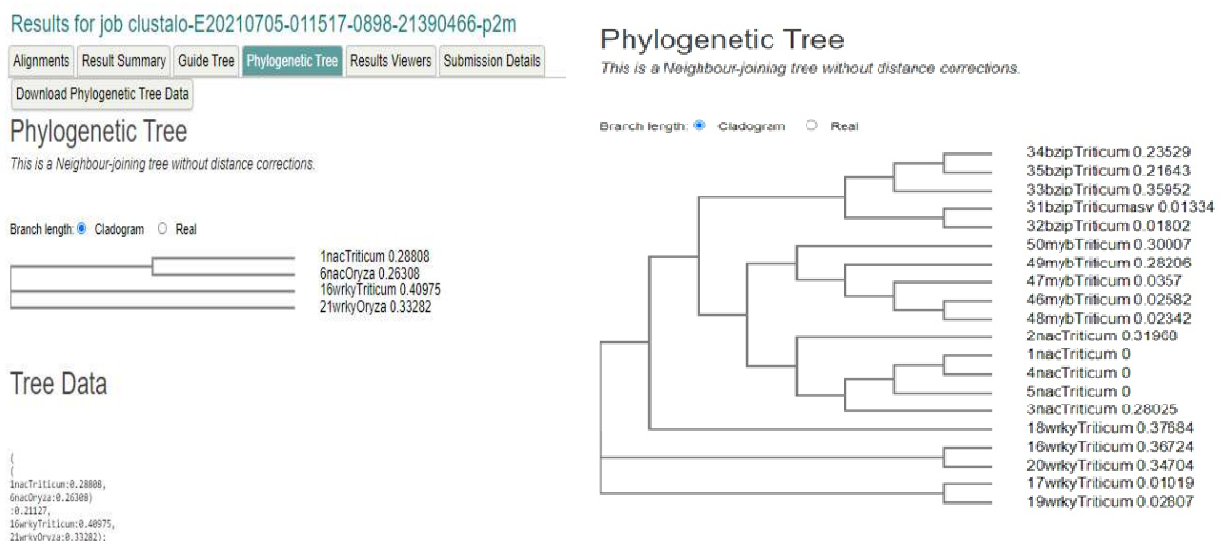
Figure 51: Arbre phylogénétique des facteurs de transcription NAC, WRKY, bzip et MYB chez le *Triticum aestivum*.



Alignement multiple protéique des gènes des familles NAC, WRKY chez *Triticum aestivum* et *Oryza sativa*.



**Figure 52 :** Alignement multiple protéique des gènes des familles NAC et WRKY chez *Triticum aestivum* et *Oryza sativa*.



**Figure 53 :** Cladogramme de facteur de transcription NAC et WRKY chez le *Triticum aestivum*

Alignement multiple des gènes des familles NAC, WRKY et myb du *Triticum aestivum* et d'*Oryza sativa*





Figure 56 : Alignement multiple des gènes NAC, WRKY, myb et bzip entre *Triticum aestivum* et *Oryza sativa*.

### Phylogenetic Tree

This is a Neighbour joining tree without distance corrections.

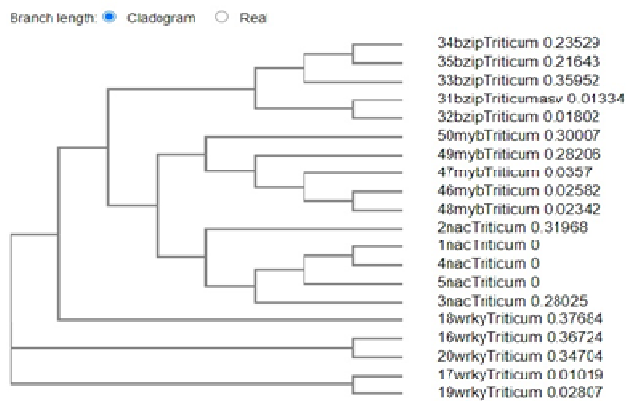


Figure 57 : Cladogramme de bzip, MYB et wrky chez le *Triticum aestivum*.

## 4. Interface du site web



localhost/ayahouria/ 70 % Rechercher

**bienvenue a Le mémoire de master 2 bioinformatique Analyse in silico de quatre facteurs de transcription exprimés chez *Triticum aestivum* sous conditions de stress biotique et abiotique.**

**Résumé de le mémoire**

Le blé tendre, *Triticum aestivum*, est l'une des céréales les plus cultivées au monde. C'est donc l'aliment de base pour de nombreuses personnes. Cependant, cette culture est très sensible aux maladies et aux changements climatiques. La maîtrise et la compréhension des mécanismes de réponse du blé tendre aux stress abiotiques est un problème actuel. Plusieurs familles de facteurs de transcription, dont NAC, WRKY, MYB et bZIP sont impliqués dans le développement de la plante et sa réponse aux pressions environnementales. Dans l'objectif d'analyser ces 4 familles de facteurs de transcription chez les plantes modèles des monocotylédones (*Oryza sativa* et *Brachypodium*) et les comparés avec leurs orthologues chez le blé tendre (*Triticum aestivum*) nous avons mené à une étude in silico pour la conception d'une base de données et la mettre en disponibilité de la communauté scientifique. Nous avons étudié les 4 familles sélectionné pour la réalisation de notre base de données (NAC, WRKY, MYB et bZIP) leur rôle dans l'évolution et le développement des plantes vis-à-vis les stress biotiques et abiotiques. Des alignements multiples ont été effectué entre l'ensemble des gènes collectés dans la base de données des quatre familles chez les trois espèces *Triticum aestivum*, *Oryza sativa* et *Brachypodium*. A travers les résultats obtenus nous avons pu montrer que les gènes des familles, WRKY, MYB, bZIP NAC sont impliqués dans le développement du blé tendre et sa réponse aux stress biotiques et abiotiques puisque le contrôle transcriptionnel de l'expression des gènes est important, en particulier dans la croissance des plantes suite au stress. En fait, les facteurs de transcription joue un rôle majeur dans l'accumulation des protéines dans le grains en réponse à un stress thermique modéré. De plus, les plantes n'ont la capacité d'échapper au parasite qu'en modifiant l'expression des gènes ou des ensembles de gènes pour améliorer leur accessibilité. L'apport de la bioinformatique par la conception d'une base de données avec l'étude phylogénétique peuvent-être contribué à une nouvelle façon d'améliorer la plasticité adaptative du blé dans des environnements très variables en raison de l'introduction d'une plus grande diversité dans le pool de gènes de résistance des systèmes de culture. Mots clés : Facteurs de transcription, base de données, stress abiotique et phylogénie.

- PARTIE BIOLOGIQUE
- BASE DE DONNEE
- SCRIPT BASE DE DONNE DE FAMILLE DE GÈNES
- ALIGNEMENT MULTIPLE DES GÈNES DE FAMILLE NAC/WRKY /MYB/BZIP TRITICUM ASETIVUM .ORYZA SATIVA .BRACHYBODIUM
- ARBRE PHYLOGÉNÉTIQUE
- ALIGNEMENT MULTIPLE PORTIQUE DES GÈNES DE FAMILLE NAC/WRKY TRITICUM ASETIVUM .ORYZA SATIVA
- ARBRE PHYLOGÉNÉTIQUE
- ALIGNEMENT MULTIPLE DES GÈNES DE FAMILLE NAC/WRKY /MYB/BZIP TRITICUM ASETIVUM .ORYZA SATIVA
- ARBRE PHYLOGÉNÉTIQUE
- CEREALSDB

Figure 58 : Interface du site web

## 5. Discussion

L'interprétation des alignements des séquences biologiques repose sur la théorie darwinienne de l'évolution. En général, les séquences alignées correspondant à des molécules remplissant des fonctions similaires, il peut s'agir par exemple de gènes orthologues de la même famille chez les différentes espèces, dont on suppose qu'elles dérivent d'un même ancêtre commun. Les divergences entre les séquences sont interprétées comme résultant de mutations. Les régions contenant des nucléotides ou des acides aminés conservés sont supposées correspondre à des zones où s'exerce une pression de sélection pour maintenir la fonction de la macromolécule.

L'alignement a plusieurs utilisations importantes en bioinformatique car il permet un certain nombre de prédictions. Il permet notamment d'identifier des sites fonctionnels (site catalytique, zone d'interaction...) qui correspondent en général aux régions les plus conservées, car ce sont elles sur lesquelles la pression de sélection est la plus grande. On peut aussi utiliser l'alignement de séquences pour prédire la ou les fonctions d'une protéine, si on détecte une homologie avec une protéine de fonction connue. Si la structure secondaire ou tertiaire de cette protéine de fonction homologue est connue, l'alignement peut être utilisé pour prédire la structure d'une protéine.

Enfin, en cas d'alignements multiples au sein d'une famille de protéines, ceux-ci peuvent permettre d'établir une phylogénie entre elles.

Des symboles sont utilisés pour indiquer les acides aminés alignés sur les sites marqués par le symbole.

« \* » indique un alignement parfait.

« : » indique un site appartenant au groupe présentant une forte similarité. "." indique un site appartenant à un groupe présentant une faible similarité.

Le critère pour distinguer une similarité forte d'une similarité faible est le suivant : Une similarité forte correspond à un score PAM250 MATRIX entre acides aminés supérieur à 0,5, tandis qu'une similarité faible correspond à un score de 0,5 ou moins.

Dans l'extrait README, les lignes horizontalement adjacentes aux phrases ;

STANEQK : Indique les groupes d'acides aminés dans les cas pour lesquels le symbole correspondant est présent (Ceux-ci sont écrits en utilisant une notation à un caractère pour les acides aminés).

Le fichier README inclus avec le package source ClustalW contient le texte suivant :

L'alignement de séquences multiples (MSA) est généralement l'alignement de trois séquences biologiques ou plus (protéine ou acide nucléique) de longueur similaire. A partir de la sortie, l'homologie peut être déduite et les relations évolutives entre les séquences étudiées.

La plupart des méthodes d'alignement de séquences multiples tentent de minimiser le nombre d'insertions/suppressions (lacunes) et, par conséquent, de produire des alignements compacts. Cela pose plusieurs problèmes si les séquences à aligner contiennent des régions non homologues, si les lacunes sont informatives dans une analyse de phylogénie. Ces problèmes sont courants dans les séquences nouvellement produites qui sont mal annotées et peuvent contenir des décalages de trame, des domaines erronés ou des exons épissés non homologues. Les régions d'ADN non codantes, en particulier les TFBS (Transcription Factor Binding Sites), sont plutôt plus conservées et pas nécessairement liées à l'évolution, et peuvent avoir convergé à partir d'ancêtres non communs. Ainsi, les hypothèses utilisées pour aligner les séquences protéiques et les régions codantes d'ADN sont intrinsèquement différentes de celles qui s'appliquent aux séquences TFBS. Bien qu'il soit significatif d'aligner les régions codant l'ADN pour les séquences homologues à l'aide d'opérateurs de mutation, l'alignement des séquences de sites de liaison pour le même facteur de transcription ne peut pas reposer sur des opérations de mutation liées à l'évolution.

De même, l'opérateur évolutif des mutations ponctuelles peut être utilisé pour définir une distance d'édition pour les séquences codantes, mais cela a peu de sens pour les séquences TFBS car toute variation de séquence doit maintenir un certain niveau de spécificité pour que le site de liaison fonctionne. Cela devient particulièrement important lorsque vous essayez d'aligner des séquences TFBS connues pour créer des modèles supervisés afin de prédire des emplacements inconnus du même TFBS.

### **Utilisation en phylogénétique**

Plusieurs alignements de séquences peuvent être utilisés pour créer un arbre phylogénétique.

Ceci est rendu possible par deux raisons.

La première est que les domaines fonctionnels connus dans les séquences annotées peuvent être utilisés pour l'alignement dans les séquences non annotées.

L'autre est que des régions conservées connues pour être fonctionnellement importantes peuvent être trouvées.

Cela permet d'utiliser plusieurs alignements de séquences pour analyser et trouver des relations évolutives par homologie entre les séquences. Les mutations ponctuelles et les événements d'insertion ou de suppression (appelés indels) peuvent être détectés.

Des alignements de séquences multiples peuvent également être utilisés pour identifier des sites fonctionnellement importants, tels que des sites de liaison, des sites actifs ou des sites correspondant à d'autres fonctions clés, en localisant des domaines conservés. Lorsque l'on examine plusieurs alignements de séquences, il est utile de considérer différents aspects des séquences lors de la comparaison des séquences. Ces aspects comprennent l'identité, la similitude

et l'homologie. L'identité signifie que les séquences ont des résidus identiques à leurs positions respectives.

D'autre part, la similitude à avoir avec les séquences comparées ayant des résidus similaires quantitativement. Par exemple, en termes de séquences nucléotidiques, les pyrimidines sont considérées comme similaires les unes aux autres, tout comme les purines. La similitude conduit finalement à l'homologie, en ce que plus les séquences sont similaires, plus elles sont proches d'être homologues. Cette similitude dans les séquences peut ensuite aider à trouver une ascendance commune.

L'arbre phylogénétique montre la succession des émergences des groupes d'organismes vivants au cours du temps, mais surtout leurs relations de parenté. Il est fondé sur l'analyse de nombreux caractères chez les espèces qu'il présente.

Une homologie désigne un lien évolutif entre deux traits observés chez deux espèces différentes, qui est dû au fait que toutes deux l'ont hérité d'un ancêtre commun. Ces traits sont alors dits homologues.

Une orthologie est un lien évolutif entre deux gènes présents chez deux espèces différentes.

Deux séquences génétiques homologues de deux espèces différentes sont orthologues si elles descendent d'une séquence unique présente dans le dernier ancêtre commun aux deux espèces.

### **1-Facteurs de transcription NAC**

Le domaine NAC a été caractérisé sur la base d'un consensus séquences de *Petunia NAM* et *Arabidopsis ATAF1/2* et Protéines CUC2. De nombreux NAC TF jouent un rôle important dans le développement et un stress biotique. De même, la majorité des rapports ont indiqué que les TF du CNA jouent un rôle central dans la reprogrammation transcriptionnelle associée la plante en système immunitaire naturel, défense basale et résistance systémique acquise résistance.

A ce jour, 134 gènes NAC ont été identifiés à partir de *T. aestivum* dans Plant TFDB, alors que seuls plusieurs d'entre eux sont impliqués dans plante répondant à un agent pathogène.

Ces fonctions semblent principalement jouer certains rôles dans la réponse des plantes au stress fongique. Le gène StNAC de la pomme de terre montre une expression induite dans les réponses à Infection à *Phytophthora infestans* et traitement des plaies.

### **2-Facteurs de transcription WRKY**

Un domaine WRKY d'environ 60 acides aminés est une caractéristique tic du domaine des protéines WRKY avec le domaine absolument conservé séquence WRKYGQK suivie d'un motif doigt de zinc. Le domaine WRKY se lie à la boîte W (C/TTGACT/C) du gène cible promoteurs pour moduler la transcription. Les WRKY TF de nombreuses cultures ont été identifiées à grande échelle et bien analysées dans les plantes cultivées, telles que *Oryza sativa*, *Brassica napus* et *T. aestivum*.

Bien que la plupart des WRKY TF se soient concentrés sur le stress abiotique et l'hormone exogène, le point de vue que WRKY les facteurs de transcription jouent un rôle important dans la régénération de l'immunité des plantes.

### **3-Facteurs de transcription MYB**

Les protéines MYB peuvent être classées en trois sous-familles en attente du nombre de répétitions adjacentes dans le domaine MYB (un, deux ou trois), et étaient appelés MYB1R, R2R3 et Facteurs MYB3R respectivement. Contrairement aux animaux, les plantes contiennent une sous-famille de protéines MYB qui est caractérisée par le R2R3- tapez le domaine MYB.

### **4-Facteurs de transcription bZIP**

Les TF bZIP sont caractérisés par deux motifs : une région basique pour une liaison spécifique à son ADN cible, et une fermeture à glissière leucine requis pour la dimérisation TF. La génétique moléculaire et les analyses biochimiques indiquent que les bZIP sont des régulateurs de nombreux processus végétaux importants tels que la différenciation, le développement, contrôle de l'équilibre des nutriments, signalisation hormonale et sucrée, le stress oxydatif et la défense contre les agents pathogènes. Les Protéines avec domaines bZIP se lient généralement aux séquences d'ADN avec un ACGT cœur. Les plantes bZIP se lient à la A-box (TACGTA), C-box (GACGTC) et G-box (CACGTG), bien que les sites de liaison non palindromiques sont également rapportés.



# Conclusion

## Conclusion

A la fin de ce mémoire nous savons que l'activité humaine a contribué négativement au changement climatique, qui dans la plupart des régions du monde a entraîné une augmentation de la température moyenne combinée à des pics occasionnels de température élevée. L'augmentation de la température altère la disponibilité en eau, ce qui conduit à un stress hydrique des cultures, qui peut également conduire à un stress salin dans certaines zones.

Ainsi, la perception du signal par la plante et sa réponse impliquent une transgression exprimer un grand groupe de gènes pour lesquels les facteurs de transcription sont des éléments moléculaires importants. Plusieurs grandes familles de facteurs de transcription sont connues pour interférer avec le développement de la plante et sa réponse au stress, notamment les familles bZIP, MYB et WRKY ou encore les NAC.

Les transcrits ont la capacité d'activer ou de réprimer des gènes par diverses stratégies de défense des plantes, y compris la régulation d'un autre promoteur de TF. Ces facteurs peuvent avoir des fonctions qui se chevauchent étroitement, entravant la fonction génétique de leurs rôles respectifs. Certains facteurs de transcription ont montré une fonction diversifiée à différents stress, cependant, la fonction des TF dans la plante modèle est réalisée dans le décodage de la régulation de l'expression des gènes de stress pathogène dans le blé panifiable hexaploïde.

De la perception des signaux de stress à l'expression de gènes sensibles au stress, les facteurs de transcription jouent un rôle essentiel et sont par la suite apparus comme des outils puissants pour améliorer la tolérance au stress grâce à la manipulation de voies métaboliques complexes chez le blé.

# **Références bibliographiques**

## Références bibliographiques

### A

Amara Korba R. 2020. Cours en bioinformatique. Site web. <https://fsnv.univ-bba.dz>

### B

Bourak K. Sélection assistée par marqueurs SNP et étude phénotypique du blé tendre (*Triticum aestivum* L.). Sciences et techniques gestion conservation de la biodiversité. Université Sidi Mohamed Ben Abdellah, 2018, 92.

### C

Claire G. Analyse des facteurs de transcription de la famille NAC chez le blé tendre (*Triticum aestivum* L.) et leur implication dans la réponse à des stress abiotiques. Génétique des plantes. Université Clermont Auvergne, 2019, 398.

Clauvain. U. 2019. Règne végétal. Formation biologie végétale. Site web <https://biologievegetale.be/regne-vegetal/systematique-des-plantes/phylogenie-du-regne-vegetal/>

### E

Element binding protein. <https://www.ncbi.nlm.nih.gov/>

### F

Fondrait C. Granger G. Brunet M et al. 2017. Cours d'autoformation en bioinformatique. Site web de l'Université René Descartes Paris 5 France. <http://www.dsi.univ-paris5.fr>

### H

Haddad L. Contribution à l'étude de la stabilité des rendements du blé dur (*Triticum durum* Desf.) sous climat méditerranéen. Agronomie. Mémoire de magistère. Université Farhat Abbas Sétif. 2010. p18.

### G

Gabriel G. 2002. La recherche in silico Editionsdu cercle de la librairie, pp.229-253.

Grine N. H. La sélection assistée par marqueurs moléculaires de type SSR pour l'amélioration du blé tendre (*Triticum aestivum*) vis-à-vis la résistance aux rouilles. Biologie et Génomique végétale. Université Frères Mentouri Constantine 1, 2015. 82.

### M

M. Bogard. 2011. Analyse génétique et écophysiological de l'écart à la relation teneur en protéines - rendement en grains chez le blé tendre (*Triticum aestivum* L.). Biology. 81252393, Figure 2.4

Michel .1992. Reactive d'oxygene species. Site web <https://www.medecinesciences.org/>  
<https://www.biologievegetale.be/>  
<https://www.biomedicalabbreviations.com>  
<https://www.larousse.fr>

**Résumé :**

Le blé tendre, *Triticum aestivum*, est l'une des céréales les plus cultivées au monde. C'est donc l'aliment de base pour de nombreuses personnes. Cependant, cette culture est très sensible aux maladies et aux changements climatiques. La maîtrise et la compréhension des mécanismes de réponse du blé tendre aux stress abiotiques est un problème actuel. Plusieurs familles de facteurs de transcription, dont NAC, WRKY, MYB et bZIP sont impliqués dans le développement de la plante et sa réponse aux pressions environnementales. Dans l'objectif d'analyser ces 4 familles de facteurs de transcription chez les plantes modèles des monocotylédones (*Oryza sativa* et *Brachypodium*) et les comparés avec leurs orthologues chez le blé tendre (*Triticum aestivum*) nous avons mené à une étude *in silico* pour la conception d'une base de données et la mettre en disponibilité de la communauté scientifique. Nous avons étudié les 4 familles sélectionné pour la réalisation de notre base de données (NAC, WRKY, MYB et bZIP) leur rôle dans l'évolution et le développement des plantes vis-à-vis les stress biotiques et abiotiques. Des alignements multiples ont été effectués entre l'ensemble des gènes collectés dans la base de données des quatre familles chez les trois espèces *Triticum aestivum*, *Oryza sativa* et *Brachypodium*. A travers les résultats obtenus nous avons pu montrer que les gènes des familles, WRKY, MYB, bZIP NAC sont impliqués dans le développement du blé tendre et sa réponse aux stress biotiques et abiotiques puisque le contrôle transcriptionnel de l'expression des gènes est important, en particulier dans la croissance des plantes suite au stress. En fait, les facteurs de transcription joue un rôle majeur dans l'accumulation des protéines dans le grains en réponse à un stress thermique modéré. De plus, les plantes n'ont la capacité d'échapper au parasite qu'en modifiant l'expression des gènes ou des ensembles de gènes pour améliorer leur accessibilité. L'apport de la bioinformatique par la conception d'une base de données avec l'étude phylogénétique peuvent-être contribué à une nouvelle façon d'améliorer la plasticité adaptative du blé dans des environnements très variables en raison de l'introduction d'une plus grande diversité dans le pool de gènes de résistance des systèmes de culture.

**Mots clés :** Facteurs de transcription, base de données, stress abiotique et phylogénie.

**Abstract:**

Bread wheat, *Triticum aestivum*, is one of the most widely cultivated grains in the world. It is therefore the staple food for many people. However, this crop is very sensitive to disease and climate change. The mastery and understanding of the response mechanisms of common wheat to abiotic stresses is a current problem. Several families of transcription factors, including NAC, WRKY, MYB and bZIP, are involved in plant development and its response to environmental pressures. In order to analyze these 4 families of transcription factors in model plants of monocotyledons (*Oryza sativa* and *Brachypodium*) and compare them with their orthologs in common wheat (*Triticum aestivum*), we carried out an *in silico* study for the design of a database and make it available to the scientific community. We have studied the 4 families selected for the realization of our database (NAC, WRKY, MYB and bZIP) their role in the evolution and development of plants vis-à-vis biotic and abiotic stresses. Multiple alignments were performed between the set of genes collected in the four family database from the three species *Triticum aestivum*, *Oryza sativa* and *Brachypodium*. Through the results obtained we were able to show that the genes of the WRKY, MYB, bZIP NAC families are involved in the development of common wheat and its response to biotic and abiotic stresses since the transcriptional control of gene expression is important, especially in the growth of plants as a result of stress. In fact, transcription factors play a major role in the accumulation of proteins in the grain in response to moderate heat stress. In addition, plants only have the ability to escape the parasite by modifying the expression of genes or sets of genes to improve their accessibility. The contribution of bioinformatics by the design of a database with the phylogenetic study may have contributed to a new way of improving the adaptive plasticity of wheat in highly variable environments due to the introduction of a greater diversity in the resistance gene pool of cropping systems.

**Keywords:** Transcription factors, database, abiotic stress and phylogeny.

## الملخص :

القمح اللين، هو أحد أكثر الحبوب المزروعة على نطاق واسع في العالم. لذلك فهو الغذاء الأساسي لكثير من الناس. ومع ذلك ، فإن هذا المحصول حساس للغاية للأمراض وتغير المناخ. يعتبر إتقان وفهم آليات استجابة القمح الشائع للضغوط الأحيائية مشكلة حالية. تشارك العديد من عائلات عوامل النسخ ، بما في ذلك NAC و WRKY و MYB و bZIP ، في تطوير المصنع واستجابته للضغوط البيئية. من أجل تحليل هذه العائلات الأربعة من عوامل النسخ في نباتات نموذجية أحادية الفلقة (*Brachypodium* و *Oryza sativa*) ومقارنتها بأخصائي تقويم العظام في القمح الشائع (*Triticum aestivum*) أجرينا دراسة في السيليكو لتصميم قاعدة بيانات وجعلها متاح للمجتمع العلمي. لقد درسنا العائلات الأربعة المختارة لتحقيق قاعدة البيانات الخاصة بنا (NAC و WRKY و MYB و bZIP) ودورها في تطور وتطور النباتات مقابل الضغوط الحيوية وغير الحيوية. تم إجراء محاكاة متعددة بين مجموعة الجينات التي تم جمعها في قاعدة بيانات الأسرة الأربعة من الأنواع الثلاثة *Brachypodium* و *Oryza sativa* و *Triticum aestivum*. من خلال النتائج التي تم الحصول عليها ، تمكنا من إظهار أن جينات عائلات WRKY و MYB و NAC و bZIP تشارك في تطوير القمح الشائع واستجابته للضغوط الحيوية وغير الحيوية نظرًا لأن التحكم النسخي للتعبير الجيني مهم ، خاصة في نمو النباتات نتيجة الإجهاد. في الواقع ، تلعب عوامل النسخ دورًا في تراكم البروتينات في الحبوب استجابةً للإجهاد الحراري المعتدل. بالإضافة إلى ذلك ، تمتلك النباتات فقط القدرة على الهروب من الطفيل عن طريق تعديل تعبير الجينات أو مجموعات الجينات لتحسين إمكانية الوصول إليها. قد تكون مساهمة المعلوماتية الحيوية من خلال تصميم قاعدة بيانات مع دراسة علم الوراثة النباتية قد ساهمت في طريقة جديدة لتحسين اللدونة التكيفية للقمح في بيئات شديدة التغير بسبب إدخال تنوع أكبر في مجموعة جينات المقاومة لأنظمة المحاصيل.

**الكلمات المفتاحية :** عوامل النسخ ، قاعدة البيانات ، الإجهاد اللاحيوي وعلم النبات.

Présenté par : *SENOUCI SOMIA*  
*TALBI HOURIA*

## Analyse *in silico* de quatre facteurs de transcription exprimés chez *Triticum aestivum* L. sous conditions de stress biotique et abiotique

Mémoire de fin de cycle pour l'obtention du diplôme de Master en Bioinformatique

### Résumé :

Le blé tendre, *Triticum aestivum*, est l'une des céréales les plus cultivées au monde. C'est donc l'aliment de base pour de nombreuses personnes. Cependant, cette culture est très sensible aux maladies et aux changements climatiques. La maîtrise et la compréhension des mécanismes de réponse du blé tendre aux stress abiotiques est un problème actuel. Plusieurs familles de facteurs de transcription, dont NAC, WRKY, MYB et bZIP sont impliqués dans le développement de la plante et sa réponse aux pressions environnementales. Dans l'objectif d'analyser ces 4 familles de facteurs de transcription chez les plantes modèles des monocotylédones (*Oryza sativa* et *Brachypodium*) et les comparés avec leurs orthologues chez le blé tendre (*Triticum aestivum*) nous avons mené à une étude *in silico* pour la conception d'une base de données et la mettre en disponibilité de la communauté scientifique. Nous avons étudié les 4 familles sélectionné pour la réalisation de notre base de données (NAC, WRKY, MYB et bZIP) leur rôle dans l'évolution et le développement des plantes vis-à-vis les stress biotiques et abiotiques. Des alignements multiples ont été effectué entre l'ensemble des gènes collectés dans la base de données des quatre familles chez les trois espèces *Triticum aestivum*, *Oryza sativa* et *Brachypodium*. A travers les résultats obtenus nous avons pu montrer que les gènes des familles, WRKY, MYB, bZIP NAC sont impliqués dans le développement du blé tendre et sa réponse aux stress biotiques et abiotiques puisque le contrôle transcriptionnel de l'expression des gènes est important, en particulier dans la croissance des plantes suite au stress. En fait, les facteurs de transcription joue un rôle majeur dans l'accumulation des protéines dans le grains en réponse à un stress thermique modéré. De plus, les plantes n'ont la capacité d'échapper au parasite qu'en modifiant l'expression des gènes ou des ensembles de gènes pour améliorer leur accessibilité. L'apport de la bioinformatique par la conception d'une base de données avec l'étude phyllogénétique peuvent-être contribué à une nouvelle façon d'améliorer la plasticité adaptative du blé dans des environnements très variables en raison de l'introduction d'une plus grande diversité dans le pool de gènes de résistance des systèmes de culture.

**Mots clés :** Facteurs de transcription, base de données, stress abiotique et phylogénie.

**Soutenu le : 23/09/2021**

Jury d'évaluation :

<b>Président :</b>	<b>Dr. CHEHILI H.</b>	M.C.A. UFM Constantine 1
<b>Encadreur :</b>	<b>Dr. KELLOU K.</b>	M.A.A. UFM Constantine 1
<b>Examineur :</b>	<b>Dr. TEMAGHOULT M.</b>	M.A.A. UFM Constantine 1

**Année universitaire : 2020/2021**