



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

Claire Guérin

► To cite this version:

Claire Guérin. 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 [2017-2020], 2019. Français. NNT : 2019CLFAC014 . tel-02284845

HAL Id: tel-02284845

<https://theses.hal.science/tel-02284845>

Submitted on 12 Sep 2019

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

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



**ECOLE DOCTORALE
SCIENCES DE LA VIE, SANTE,
AGRONOMIE, ENVIRONNEMENT**

Thèse

Présentée à l'Université Clermont Auvergne

Pour l'obtention du grade de

DOCTEUR D'UNIVERSITE

Spécialité : Biologie cellulaire et moléculaire végétale

Soutenue le 29 Avril 2019

Claire GUERIN

**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**

Président : Mme Huguette Sallanon, Professeur à l'Université d'Avignon

Rapporteurs : Mme Béatrice Teulat, Maître de Conférence-HDR à Agrocampus Ouest CFR d'Angers
M. Soulaïman Sakr, Professeur, IRHS, Angers

Membres : Mme Jane Roche, Maître de Conférence à l'Université Clermont-Auvergne (Examinatrice)
M. Saïd Mouzeyar, Professeur à l'Université Clermont-Auvergne (Directeur de thèse)
M. M-Fouad Bouzidi, Professeur à l'Université Clermont-Auvergne (Co-Directeur de thèse)



UMR 1095 GDEC INRA-Université Clermont-Auvergne, 1 impasse Amélie Murat 63178 Aubière

Remerciements

Tout d'abord, je remercie les membres du jury, Huguette Sallanon, Soulaïman Sakr et Béatrice Teulat, pour avoir accepté de juger mes travaux. Merci à vous, j'espère que la lecture de ce manuscrit vous sera agréable.

Un grand merci aussi Catherine Ravel et Stéphane Lafarge, les deux membres de mes comités de thèse pour leurs conseils et leur aide, notamment concernant la génétique d'association. Catherine, j'aimerais te remercier tout particulièrement pour le temps que tu as consacré à ces recherches malgré ton emploi du temps très chargé.

Maintenant, je souhaite adresser des énormes remerciements à mes encadrants : Saïd Mouzeyar, Fouad Bouzidi et Jane Roche ; vous êtes un trio formidable et votre double casquette de chercheurs et d'enseignants fait de vous des pédagogues extraordinaires. Grâce à votre complémentarité et votre disponibilité, j'ai énormément appris à votre contact, professionnellement et humainement. Saïd et Fouad, merci d'avoir su guider l'étudiante inexpérimentée que j'étais pour en faire une chercheuse, vous avez su apporter les conseils et le recul nécessaire pour faire avancer cette thèse. Jane, ton énergie et ton optimisme au quotidien sont une source de fascination (et d'épuisement parfois) ! Quel est ton secret ? Est-ce le Fameux Calendrier accroché au-dessus de ton bureau qui te donne cette force ? Et pourquoi renommes-tu tous tes stagiaires ? Il reste tellement de questions en suspens...

J'aimerais ensuite remercier Monique Sibaud. Momo, à notre premier contact, ta rigueur as fait peur à la petite étudiante que j'étais. Heureusement, tu n'es pas bien haute non plus et j'ai pu vite découvrir que sous la blouse blanche (que tu ne quittais jamais) se cache une excellente pédagogue. Merci de m'avoir appris tout ce que je sais à la paillasse, je te souhaite une super retraite !

Merci à toi Céline, tu es la personne la plus douce que je connaisse. Merci pour toutes les manips et toutes les discussions que l'on a pu partager. Et merci d'avoir partagé ta fascination pour le mot Slip !

D'abord Steph puis Clémence, merci les filles pour le redoutable travail qu'on a abattu en serre ! Et pour les fous-rire qu'on a pu y partager. Céline, Clémence, leçon de vie : enfermer 3 filles dans une serre pour ensacher ou décortiquer des épis toute une journée est dangereux pour la santé mentale mais bon pour les zygomatiques !

Merci à Dominique pour les belles photos de grain et le traitement des images ; et merci à Sandrine pour son efficacité et sa disponibilité à toute épreuve.

Merci aux collègues de l'INRA de Crouël avec qui j'ai eu le plaisir d'interagir : Philippe Leroy, les membres de la plateforme ValFon, surtout Caro, Stéphane et Lucas, ce n'était pas de tout repos mais on l'a fait ! Et merci à Annie qui a si bien pris soin de la première génération de transgénique ainsi que Sibille et David qui ont assuré pour le phénotypage de la dernière génération et tous les membres de la plateforme Gentyane, qui m'ont permis l'accès à leurs supers robots.

Merci aux collègues de bureau : Osama, grâce à qui je sais dire « l'âne rouge » en arabe et ça c'est un véritable atout ; Manel ; Romain-Martine. Une mention spéciale pour Elodie-Marion Caillou qui, malgré mes remarques persistantes sur son alimentation 100% gras de poivron, s'est bel et bien portée volontaire pour relire toute ma prose, encore merci meuf, tu gères ! Et bon courage à Alice, à qui je passe le flambeau pour prendre soin des plantes et de Roberte.

Les repas à la bonne franquette vont terriblement me manquer. Jane, Momo, Sylvaine, JP, Lolo, Jeannette, puis Juliette, Aurélie, Romain-Martine, Clémence, Enrico-Martial, Elodie-Marion, Antonin & Aurélia, Amélie, Pauline, Saïd et tous les gens de passage ; les pauses repas ont été des formidables moyens de décompresser. Je remercie au passage le verbicruciste de la Montagne, qui nous aura appris pléthores de mots utiles comme « raout » ou « Ernst ». Et puisqu'on est rendu au chapitre bouffe, merci aussi à la boulangerie-du-coin-de-ma-rue et ses délicieuses pâtisseries, dont la farine « a été élaborée en partenariat avec l'INRA ». Leurs gâteaux ont toujours été là pour moi en cas de coup dur !

Et parce que réussir une thèse demande aussi de décompresser, merci à la meute angevine : Daph, Lapelle, Jojo, Juju, Paaf et au Pandours d'être aussi fous et aussi forts. Et un merci spécial à Estelle qui a poussé le soutien jusqu'à venir à Clermont et trouver un stage dans mon bâtiment ! ... Comment ça, je n'étais pas la seule motivation ?

Malice, merci pour les soirées mojitos, les soirées bouffe, les soirées jeux, et pour les belles promenades dans ta région. Vive les volcans ! Selver, tu es une grande folle, douce et sage. Un jour, je ferai un crochet par Istanbul et on se partagera des tchoureks. Harsh, thanks for your enthusiasm and our long talks-walks by night.

Je dédie cette thèse :

- À l'odeur de l'huile de lin - essence de térébenthine,
- Au bruit d'une bague qui frotte contre une table en chêne,
- À la douceur moelli-rugueuse d'une cramousse dans la force de l'âge,
- Au goût du puits d'amour quand il cache une petite fée blonde.

Vous avez toujours été là dans les bons moments comme dans les coups durs, merci de m'avoir menée jusqu'ici.

Et ces remerciements n'auraient aucun sens si je n'y citais pas le Lord Docteur-presque-au-carré de Glencoe qui existe pour de vrai et qui me subit depuis bientôt 7 ans. Merci pour nos week-ends et nos fous-rires, merci d'avoir été à mes côtés malgré la distance. On va enfin pouvoir s'inventer notre nid !

Et pour finir, merci à l'*Homo Sapiens* qui a eu un beau matin l'idée saugrenue de domestiquer le blé. Je t'ai parfois maudit ces 3 dernières années mais ça en valait le coup !

RÉSUMÉ

Le blé tendre, *Triticum aestivum*, est une des céréales les plus cultivées dans le monde. Le changement climatique qui se développe actuellement contraint fortement les cultures et altère leur rendement. La compréhension des mécanismes de réponse du blé tendre aux stress abiotiques est donc une problématique d'actualité. Plusieurs grandes familles de facteurs de transcription, dont la famille NAC, interviennent dans le développement de la plante et dans sa réponse aux stress environnementaux. Cette thèse, structurée en 3 volets, est ciblée sur l'étude de la famille NAC chez le blé tendre : les TaNAC.

Dans un premier temps, nous avons étudié la structuration génomique et phylogénétique des 488 membres de la famille TaNAC, recensés à partir de la base de données la plus récente du blé tendre. Nous avons aussi étudié l'histoire évolutive de cette famille, qui a été marquée par des événements de duplication et de rétroposition. Enfin, une analyse de sa diversité allélique a permis d'identifier des gènes qui présentent des SNP montrant une forte association avec des paramètres d'accumulation des protéines de réserve dans le grain.

Le deuxième chapitre de cette thèse a porté sur l'étude de l'expression de ces 488 gènes TaNAC dans plusieurs organes et en réponse aux stress thermique et sécheresse. Une analyse globale a été réalisée à partir de données bio-informatiques, suivie d'une étude *in planta* de l'expression d'une sélection de 23 gènes. Les profils d'expression obtenus ont révélé l'existence de 4 gènes TaNAC, encore jamais décrits dans la littérature et qui interviennent dans le développement du grain de blé tendre mais aussi dans sa réponse adaptative à plusieurs stress abiotiques.

Le troisième volet de cette thèse a donc porté sur la caractérisation génétique, moléculaire et physiologique de ces 4 facteurs de transcription TaNAC. Ils appartiennent à un clade rassemblant des séquences présentant des similitudes génomique et structurale. De plus, ils sont localisés dans le noyau et leurs profils d'expression sont similaires, avec toutefois un niveau variable entre gènes et entre homéologues pour chaque gène. En réponse à un stress thermique modéré, ce profil d'expression est accéléré au cours du développement du grain ; le stade 120°Cj étant le stade clé qui montre la plus grande différence d'expression de ces gènes entre les conditions contrôle et stressée. Pour des raisons techniques, la production de plantes transgéniques sur- et sous-exprimant ces gènes n'a pas permis de valider l'implication de ces 4 TaNAC dans le développement du grain et en réponse à la température. Une analyse de génétique d'association a toutefois permis de mettre en évidence un lien entre des marqueurs moléculaires situés dans ces gènes et l'accumulation des protéines de réserve.

Globalement, les résultats obtenus ont montré que des membres de la famille TaNAC sont impliqués dans le développement du blé tendre et dans sa réponse aux stress abiotiques. Plus particulièrement, 4 facteurs de transcription TaNAC semblent jouer un rôle clé dans l'accumulation des protéines dans le grain en réponse à un stress thermique modéré.

Mots clés : blé tendre, facteurs de transcription, famille NAC, expression de gènes, stress abiotiques.

ABSTRACT

Bread wheat, *Triticum aestivum*, is one of the most cultivated cereal in the world. The climate change that is currently developing strongly constrains crops and impairs their yield. Understanding the wheat response mechanisms to abiotic stresses is therefore a current issue. Several major families of transcription factors, including the NAC family, are involved in the plant development and its response to environmental stresses. This thesis, structured in three parts, is focused on the study of the NAC family in bread wheat (TaNAC).

First, we studied the genomic and phylogenetic structure of the 488 members of the TaNAC family identified from the latest database of bread wheat. We also studied the evolutionary history of this family, which was marked by duplication and retroposition events. Finally, an analysis of its allelic diversity allows us to identify genes with SNP showing a strong association with storage protein accumulation parameters in the grain.

In a second part, we studied the expression of these 488 TaNAC genes in several organs and in response to heat and drought. An overall analysis was performed using bioinformatic data, followed by an *in planta* study of the expression of a selection of 23 genes. The expression profiles revealed that four TaNAC genes, never described in the literature, are involved in the wheat grain development but also in its adaptive response to several abiotic stresses.

In a third part, we focused on the genetic, molecular and physiological characterization of these four TaNAC transcription factors. They belong to a clade gathering sequences with genomic and structural similarities. Moreover, they are localized in the nucleus and their expression profiles are similar, with a variable level between genes and between homeologs for each gene. In response to moderate heat stress, this expression profile is accelerated during grain development and a key stage at 120°Cj was identified, it shows the greatest difference in genes expression level between control and stressed conditions. For technical reasons, the production of transgenic plants over- and under-expressing these genes did not validate the involvement of these 4 TaNAC in grain development and in its temperature response. An association genetic analysis, however, showed a link between molecular markers located in these genes and the storage proteins accumulation.

Overall, the results showed that members of the TaNAC family are involved in the bread wheat development and its response to abiotic stresses. In particular, four TaNAC transcription factors appear to play a key role in grain protein accumulation in response to a moderate heat stress.

Key words: bread wheat, transcription factors, NAC family, genes expression, abiotic stresses.

TABLE DES MATIÈRES

Abréviations et Glossaire.....	1
---------------------------------------	----------

SYNTHESE BIBLIOGRAPHIQUE

1. Le blé tendre	4
1.1. Origines des blés cultivés	4
1.2. Utilisation du blé tendre, importances économiques et sociétales	5
1.3. Anatomie du blé tendre.....	5
1.4. Développement du blé tendre	6
1.4.1. La période végétative	7
1.4.2. La période reproductive	7
2. Le grain de blé.....	9
2.1. Anatomie du grain de blé.....	9
2.1.1. Les enveloppes	9
2.1.2. L'albumen	10
2.1.3. Le germe.....	10
2.1.4. Morphologie standard d'un grain de blé	11
2.2. Composition biochimique du grain de blé.....	11
2.2.1. L'amidon	11
2.2.2. Les protéines	12
2.3. Phases de développement du grain de blé	13
2.3.1. Phase de divisions cellulaires	13
2.3.2. Phase de remplissage.....	14
2.3.3. Phase de maturation	15
2.4. Statut hydrique du grain au cours de son développement	16
3. Effets des principaux stress abiotiques de l'environnement sur le développement et la croissance du blé.....	16
3.1. Effets des stress sur le développement des céréales à l'échelle du peuplement et de la plante 18	
3.1.1. Stress thermique	19
3.1.2. Stress hydrique	21
3.1.3. Stress salin.....	22
3.1.4. Multi-stress.....	22

3.2.	Réponses de la plante aux stress hydrique et thermique aux niveaux moléculaire et cellulaire	23
3.2.1.	Réponse à un stress hydrique	25
3.2.2.	Réponse à un stress thermique	25
3.3.	Rôles des facteurs de transcription dans la réponse au choc thermique	26
4.	Les facteurs de transcription NAC	28
4.1.	Historique	28
4.2.	Structure protéique et interaction avec l'ADN	29
4.3.	Rôles des NAC dans l'évolution des plantes	30
4.4.	Rôles des NAC dans le développement de la plante	31
4.5.	Rôles des NAC dans la réponse de la plante aux stress biotiques et abiotiques.....	33
4.5.1.	Stress biotiques.....	34
4.5.1.	Stress abiotiques.....	35
5.	Description du projet BreedWheat, contextualisation du sujet de thèse	40
6.	Objectifs de la thèse	42

MATERIELS & METHODES

1.	Analyse <i>in silico</i> de la famille des gènes NAC chez le blé tendre (TaNAC)	44
1.1.	Recensement des membres de la famille TaNAC	44
1.1.1.	Récupération et filtration des données	44
1.1.2.	Constitution d'un set consensus	44
1.2.	Structure et organisation de la famille des gènes TaNAC	45
1.2.1.	Séquences génomiques, localisation chromosomique et identification des groupes d'homéologie	45
1.2.2.	Phylogénie et orthologie.....	46
1.2.3.	Analyse des séquences protéiques de la famille TaNAC	46
1.2.4.	Analyse des duplications	46
1.3.	Méta-analyse de l'expression <i>in silico</i> des gènes appartenant à la famille TaNAC..	47
1.3.1.	Sur la plateforme Genevestigator	47
1.3.2.	Sur la plateforme WheatExpression	48
1.4.	Evaluation de la diversité allélique au sein de la famille TaNAC et association avec des caractères phénotypiques.....	49
1.4.1.	Etude de la diversité allélique de la famille TaNAC.....	49
1.4.2.	Association entre SNP et caractères phénotypiques	50
2.	Expérimentations sur blé tendre (<i>Triticum aestivum</i>).....	51

2.1.	Stress thermique modéré dans le cadre du projet BreedWheat	52
2.1.1.	Dispositif expérimental	52
2.1.2.	Etudes transcriptomique, phénotypique et réseau de co-expression	53
2.1.3.	Obtention de données d'expression pour 4 gènes TaNAC	53
2.2.	Expérimentation sécheresse en plein champ	54
2.2.1.	Génotypes utilisés	54
2.2.2.	Dispositif expérimental (Phéno3C)	54
2.2.3.	Prélèvements et suivi de l'expression de 23 gènes TaNAC	55
2.2.4.	Analyse statistique globale des données d'expression en Analyse en Composantes Principales (ACP)	55
2.3.	Choc thermique.....	56
3.	Méthodes de suivi de l'expression des gènes.....	56
3.1.	Extraction des ARN totaux.....	56
3.1.1.	Sur feuille	56
3.1.2.	Sur grain	56
3.2.	Analyses d'expression	57
3.2.1.	Transcription inverse des ARN en ADN complémentaires	57
3.2.2.	Dessin de couples d'amorces gène complet, homéologues-spécifique et généraliste	57
3.2.3.	PCR	58
3.2.4.	PCR quantitative en temps réel (RT-qPCR)	58
3.3.	Vérification des produits d'amplification PCR par séquençage.....	59
3.3.1.	Purification des produits d'amplification	59
3.3.2.	Sous-clonage dans le plasmide pGEM-T Easy : TA cloning.....	59
3.3.3.	Transformation de la souche bactérienne <i>E. coli</i> DH5α compétente, vérification de la présence de l'insert/Criblage des colonies	60
3.3.4.	Extraction d'ADN plasmidique par lyse alcaline (mini prep).....	61
3.3.5.	Séquençage.....	61
4.	Transformation transitoire de feuilles de tabac.....	61
4.1.	Clonage du gène dans le vecteur pEarleyGate104 (pEG104)	61
4.1.1.	Le plasmide pEG104	61
4.1.2.	Clonage de type Gateway	62
4.2.	Transformation d' <i>Agrobacterium tumefaciens</i>	62
4.2.1.	Souche d' <i>A. tumefaciens</i> utilisée.....	62
4.2.2.	Mise en compétence et transformation de la souche.....	62
4.3.	Infiltration de l'épiderme inférieur de feuilles de tabac (<i>Nicotiana tabacum</i>)	63
4.4.	Observations en microscopie à épifluorescence	63

5. Expérimentations sur petit épeautre (<i>Triticum monococcum</i>).....	64
5.1. Génotype utilisé.....	64
5.2. Cinétique de développement du grain	64
5.2.1. Dispositif expérimental	65
5.2.2. Notation de l’anthèse.....	65
5.2.3. Récolte et analyse des grains.....	66
5.3. Application d’un stress thermique modéré durant les stades précoces de développement du grain de petit épeautre	66
5.3.1. Dispositif expérimental	67
5.3.2. Détermination des stades de développement pour le suivi de l’expression	67
5.3.3. Récolte des épis	67
5.4. Obtention d’embryons immatures pour la transgénése	68
5.4.1. Matériel végétal et dispositif expérimental	68
5.4.2. Récolte des embryons immatures.....	68
5.5. Transformation génétique du petit épeautre (<i>T. monococcum</i>) par biolistique	69
5.5.1. Surexpression <i>in planta</i> de TmNAC1, TmNAC2, TmNAC3 et TmNAC4	69
5.5.2. Sous-expression <i>in planta</i> des 4 TmNAC par ARN interférence (RNAi)	75
6. Significativité des différences statistiques observées dans les lignées transgéniques et dans les analyses transcriptionnelles	78

CHAPITRE I : RECENSEMENT DES MEMBRES DE LA FAMILLE TaNAC CHEZ LE BLE TENDRE

1. Recensement et description de la famille de facteurs de transcription TaNAC	80
2. Structuration phylogénétique et motifs conservés chez les gènes TaNAC	82
3. Distribution génomique et duplication de la famille TaNAC	84
3.1. Distribution génomique	84
3.2. Duplications au sein de la famille TaNAC – Exemple de la région 4B-5A	84
3.3. Devenir des gènes TaNAC dupliqués.....	86
4. Diversité allélique au sein de la famille TaNAC.....	87
4.1. Selon les données de la puce BW_420K	87
4.2. Selon les données de la base de données Whealbi	87
4.3. Compilation des données de variabilité obtenues avec BW_420K et avec Whealbi.....	88
5. Association polymorphisme x phénotypes	88

5.1.	SNP et composition protéique du grain	88
5.2.	SNP et caractères technologiques	90
5.3.	SNP et caractères agronomiques	90
6.	Les gènes TaNAC, une grande famille	91
7.	Expansion de la famille TaNAC	94
7.1.	Les duplications à petite échelle, un moteur de l'expansion de la famille TaNAC... 94	
7.2.	Les rétropositions, un autre moteur de l'expansion de la famille TaNAC	96
7.3.	Duplication et fonction des gènes.....	97
8.	Répartition du polymorphisme au sein de la famille TaNAC.....	98
9.	Génétique d'association avec les caractères phénotypiques d'intérêt	99
10.	Les idées clés	101

CHAPITRE II : REPONSE PHYSIOLOGIQUE DES MEMBRES DE LA FAMILLE TaNAC AU COURS DU DEVELOPPEMENT ET EN REPONSE À DES STRESS ABIOTIQUES

1.	Expression des gènes TaNAC dans cinq organes à trois stades de développement	103
2.	Effet des stress abiotiques sur l'expression des membres de la famille TaNAC	104
2.1.	Choc thermique et/ou sécheresse sur feuilles au stade plantule	104
2.1.1.	Analyse des données RNAseq disponibles publiquement	104
2.1.2.	Sélection de 23 gènes représentatifs parmi ceux différentiellement exprimés dans un organe ou en réponse à un stress.....	105
2.1.3.	Suivi de l'expression <i>in planta</i> de 23 gènes TaNAC en réponse à un choc thermique	105
2.2.	Suivi de l'expression <i>in planta</i> de 23 gènes TaNAC après application d'un stress sécheresse en plein champ	107
2.2.1.	Mesures des paramètres éco-physiologiques en réponse à la sécheresse.....	107
2.2.2.	Identification des gènes TaNAC exprimés de manière différentielle chez les plantes cultivées au champ.....	108
2.3.	Effet du stress thermique modéré sur l'expression des gènes TaNAC.....	109
3.	L'expression des gènes TaNAC évolue au cours du développement du blé tendre	111
4.	Un choc thermique induit une modification de l'expression d'au moins 10 gènes TaNAC dans les plantules de blé tendre	112

5. Comportement des gènes TaNAC au cours du développement du blé en plein champ et lors de sa réponse à la sécheresse.....	113
5.1. Avant-propos	113
5.2. Des gènes TaNAC montrent une expression préférentielle dans le grain ou la feuille	113
5.3. La réponse des TaNAC à la sécheresse a lieu principalement dans les feuilles, durant la phase de remplissage du grain	114
5.4. Quatre gènes TaNAC exprimés dans le grain ont une réponse génotype-dépendante à la sécheresse et à la température.....	117
6. Les idées clés	119

CHAPITRE III : CARACTERISATION DE 4 GENES TaNAC CHEZ LE BLE et transfert au petit épeautre

1. Etude des gènes TaNAC exprimés préférentiellement dans le grain	121
1.1. Structure génomique.....	121
1.2. Recherche d'un motif conservé	122
2. Clonage des homéologues de TaNAC1, TaNAC2, TaNAC3 et TaNAC4.....	122
3. Variabilité de la séquence des 4 TaNAC et association avec un caractère phénotypique.....	123
3.1. Variabilité de la séquence des 4 TaNAC étudiés	123
3.2. Association SNP x phénotypes.....	124
4. Localisation subcellulaire des 4 TaNAC	124
5. Expression des 4 TaNAC au cours du développement du blé tendre	125
5.1. Expression dans le grain	125
5.2. Implication des homéologues des 4 TaNAC au cours du développement du grain	125
6. Effet des stress abiotiques sur l'expression des 4 TaNAC chez le blé	126
6.1. Implication de chaque TaNAC en réponse au stress thermique.....	127
6.2. Implication des homéologues des 4 TaNAC en réponse au stress thermique	127
6.3. Réponse à d'autres stress abiotiques	128
7. Validation fonctionnelle.....	128
7.1. Transfert de l'étude des 4 TaNAC chez le petit épeautre.....	129
7.1.1. Développement du grain de petit épeautre	129
7.1.2. Séquençage des 4 TmNAC chez le petit épeautre.....	130

7.1.3.	Cinétique de développement en réponse à un stress thermique chez le petit épeautre	130
7.2.	Validation fonctionnelle	131
7.2.1.	Intégration de la cassette et suivi de l'expression des 5 gènes NAC	132
7.2.2.	Phénotypage	132
8.	Un clade de l'arbre phylogénétique porteur de gènes aux caractéristiques proches	134
9.	Les 4 TaNAC sont regroupés dans un clade phylogénétique	137
9.1.	Comparaison entre les séquences de l'IWGS et les séquences clonées	137
9.2.	Régulation de l'activité des 4 TaNAC étudiés	139
9.2.1.	Régulation transcriptionnelle	139
9.2.2.	Régulation post-transcriptionnelle	140
9.2.3.	Régulation post-traductionnelle	140
9.3.	Rôle et mécanisme de fonctionnement supposés des 4 TaNAC en condition de température optimale	141
9.3.1.	À l'échelle du grain	141
9.3.2.	À l'échelle tissulaire	144
9.4.	Rôle et mécanisme de fonctionnement supposés des 4 TaNAC en réponse aux stress abiotiques	145
9.4.1.	Lors d'une élévation de température	145
9.4.2.	Une réponse génotype-dépendante à la sécheresse	147
10.	Validation fonctionnelle	148
11.	Les idées clés	150
<u>DISCUSSION GENERALE, CONCLUSION et PERSPECTIVES</u>		151
Références bibliographiques		157
Annexes		185

ABRÉVIATIONS

°Cj	degrés Celsius-jour
aa	acide aminé
ABA	Acide ABscissique
ACP	Analyse en Composante Principale
ADNc	Acide DésoxyriboNucléique codant
ARN(m)	Acide Ribonucléique (messenger)
ATAF	Arabidopsis Transcription Activation Factor
BLAST	Basic Local Alignment Search Tool
CUC2	CUp-shaped Cotyledon 2
DEGs	Differentially Expressed Genes
DREB	Dehydration-Responsive Element Binding protein
ERD1	Early Responsive to Dehydration stress 1
FAO(stat)	Food and Agriculture Organization of the united nations (statistics)
GDEC	Génétique Diversité et Ecophysiologie des Céréales
H ₂ O ₂	Peroxyde d'hydrogène
HMW-GS	High-Molecular-Weight Glutenin Subunit
HSF	Heat Shock Factor
(s)HSP	(small) Heat Shock Protein
JAA	Jour Après Anthèse
kb	kilo-bases, soit 1000 paires de bases
LIMMA	Linear Models for Microarray data
LMW-GS	Low-Molecular-Weight Glutenin Subunit
MB	Maître Brin
miARN	microARN (noté miR)
NAC	NAM-ATAF-CUC2
NACRS	NAC Recognition Sequence
NAM	No Apical Meristem
NAP	NAC-like activated by AP3/PI
NS	Null-Ségrégant
NTL	NAC with Transmembrane motif 1 – Like
pb	paire de bases
PCR	Polymerase Chain Reaction
pEG	plasmide Earl Gate
PEG	PolyEthylène Glycol
PMG	Poids de Mille Grains
PMI	PhosphoMano Isomérase
QTL	Quantitative Trait Locus
RNAi	ARN interférence
RNAseq	RiboNucleic Acid sequencing

ROS	Reactive Oxygen Species
RT-qPCR	Real-Time quantitative Polymerase Chain Reaction
SNAC	Stress-related NAC
SNBE	Secondary wall NAC Binding Element
SNP	Single Nucleotide Polymorphism
SOD	SuperOxyde Dismutase
T1	Talle 1
Ta	<i>Triticum aestivum</i>
TermNOS	Termineur de la NOpaline Synthase
Tm	<i>Triticum monococcum</i>
UTR	Untranslated Transcribed Region
YFP	Yellow Fluorescent Protein

GLOSSAIRE

La définition des mots suivant peut varier en fonction des domaines d'étude et être sujet à débat. C'est dans cette optique que le sens qui leur est donné dans ce rapport est clarifié.

Acclimatation : modification phénotypique rapide et réversible d'un individu exposé à un changement environnemental, permise par la plasticité phénotypique (Gerken *et al.* 2015).

Adaptation : résultat de la sélection naturelle, sélection d'une population dotée de caractéristiques génétiques et héréditaires permettant un taux de survie et/ou de reproduction optimisé dans un environnement donné. Les gènes déterminant ces traits sont sélectionnés d'une génération à l'autre, formant un processus graduel et irréversible qui a lieu sur de longues périodes et à l'échelle d'une population (Gienapp *et al.* 2008).

Sécheresse : ce mot étant largement utilisé dans la littérature comme synonyme de « stress hydrique » (drought), il est ici aussi employé comme tel.

Tolérance : capacité à maintenir un niveau de production (et/ou de qualité) élevé en présence de stress. Ce concept est à opposer à la résistance (capacité d'éliminer ou de limiter la contrainte par des mécanismes génétiques et moléculaires) et l'échappement (capacité d'échapper à la contrainte) (Ney *et al.* 2013).

SYNTHÈSE BIBLIOGRAPHIQUE

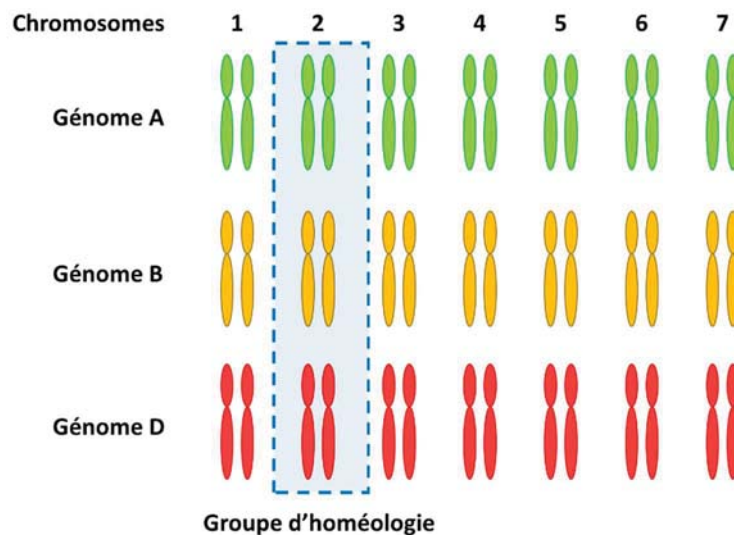


Figure 1 : Organisation du génome hexaploïde du blé tendre. Ce génome comprend 3 sous-génomes homéologues A, B et D, issus de chaque ancêtre du blé tendre ; et 7 chromosomes par sous-génome. Un groupe d'homéologie correspond aux 3 séquences homéologues portées respectivement par les 3 sous-génomes.

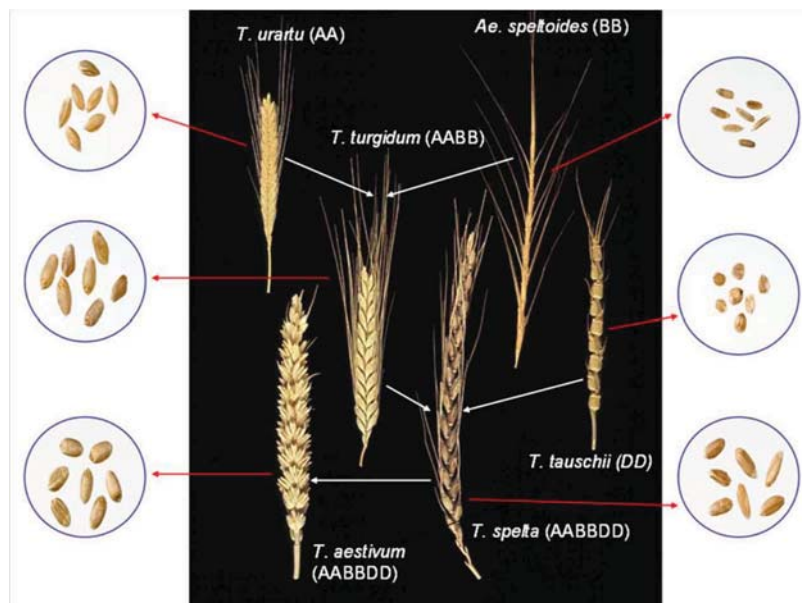


Figure 2 : Relations phylogénétiques entre le blé tendre (*T. aestivum*) et ses ancêtres sauvages (d'après Shewry 2009). Un exemple d'épi et de grains est donné pour chaque espèce. Le blé tendre est la résultante de 2 hybridations successives. Le croisement de *Triticum urartu* avec une espèce proche d'*Aegilops speltoides* a donné *Triticum turgidum*, à l'origine du blé dur, *Triticum turgidum* ssp *durum*. En se croisant avec *Triticum tauschii*, ce dernier a mené à l'apparition du blé tendre.

1. Le blé tendre

1.1. Origines des blés cultivés

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 devenu 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.

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. bœoticum*, porteurs du génome A^bA^b (l'indice b correspond à *bœoticum*) 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 (Figure 1). Le blé possède donc un génome de très grande taille mais avec une forte proportion de séquences répétées (80% du génome) (Paux *et al.* 2008). 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* spp. *dicoccoides*) au génome AABB par croisement naturel de *T. monococcum* (AA) avec une espèce proche d'*Aegilops speltoides* (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* spp. *dicoccoides* 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 2).

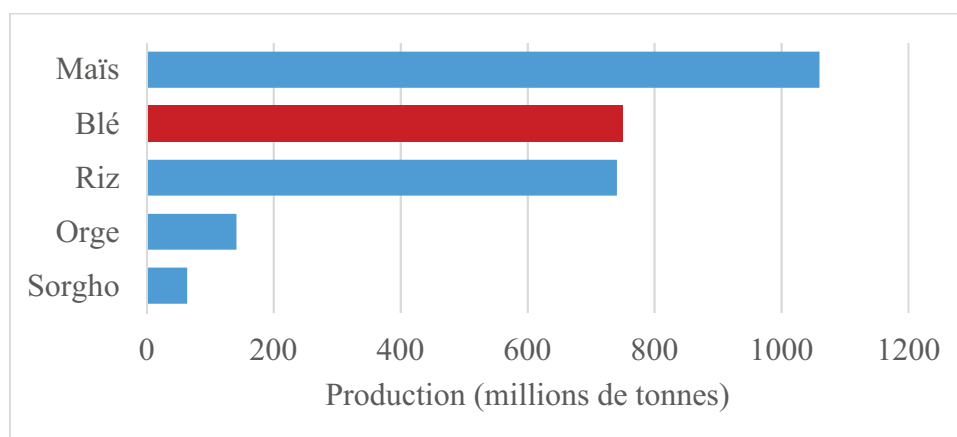


Figure 3 : Les 5 premières productions céréalières mondiales, en 2016. Le blé est la deuxième céréale la plus produite dans le monde et le blé tendre en représente 95%. Source : FAOstat, <http://faostat.fao.org>.

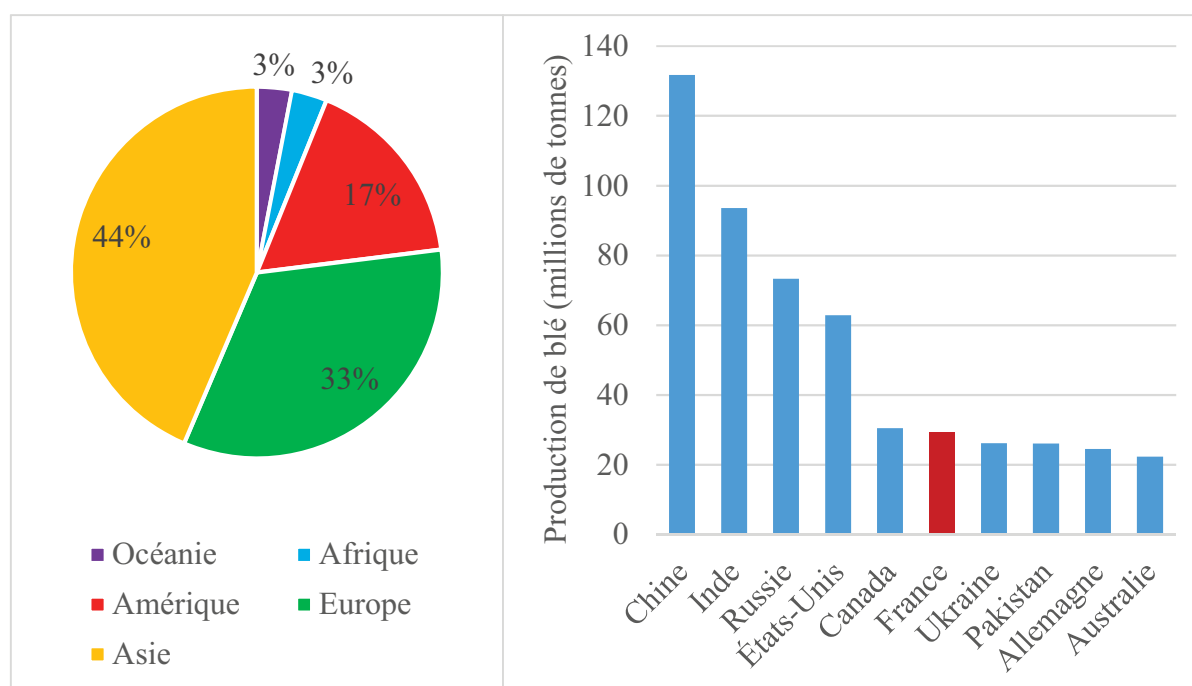


Figure 4 : Production mondiale de blé tendre en 2016. À gauche, la répartition de la production par continent. Avec 250 millions de tonnes récoltées en 2016, l'Europe est le 2ème producteur mondial de blé. À droite, les 10 premiers pays producteurs de blé. La France est le 6ème producteur mondial, avec 29,5 millions de tonnes de blé en 2016. Source : FAOstat, <http://faostat.fao.org>.

1.2. 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^{ème} culture à l'échelle mondiale derrière le maïs (Figure 3). L'Asie à elle seule représente 43,6% de cette production, suivie de l'Europe à 33,4% (FAOstat 2018, Food and Agriculture Organization of the United Nations – statistics, <http://faostat.fao.org>) (Figure 4). Le blé tendre représente 95% de la production mondiale de blés cultivés (Shewry 2009), tandis que la majorité des 5% 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 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 (Source : <https://stats.agriculture.gouv.fr/>). 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).

1.3. Anatomie 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 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. Développement du blé tendre

Deux types de blés tendres existent : les blés de printemps et les blés d'hiver. Ces derniers nécessitent une période de vernalisation (0 à 7°C, pendant 4 à 8 semaines) pour acquérir la capacité de fleurir, ce qui n'est pas le cas des blés de printemps (Acevedo *et al.* 2002). L'existence de 2 types de blé est liée à leurs aires de culture : les blés tendres d'hiver sont

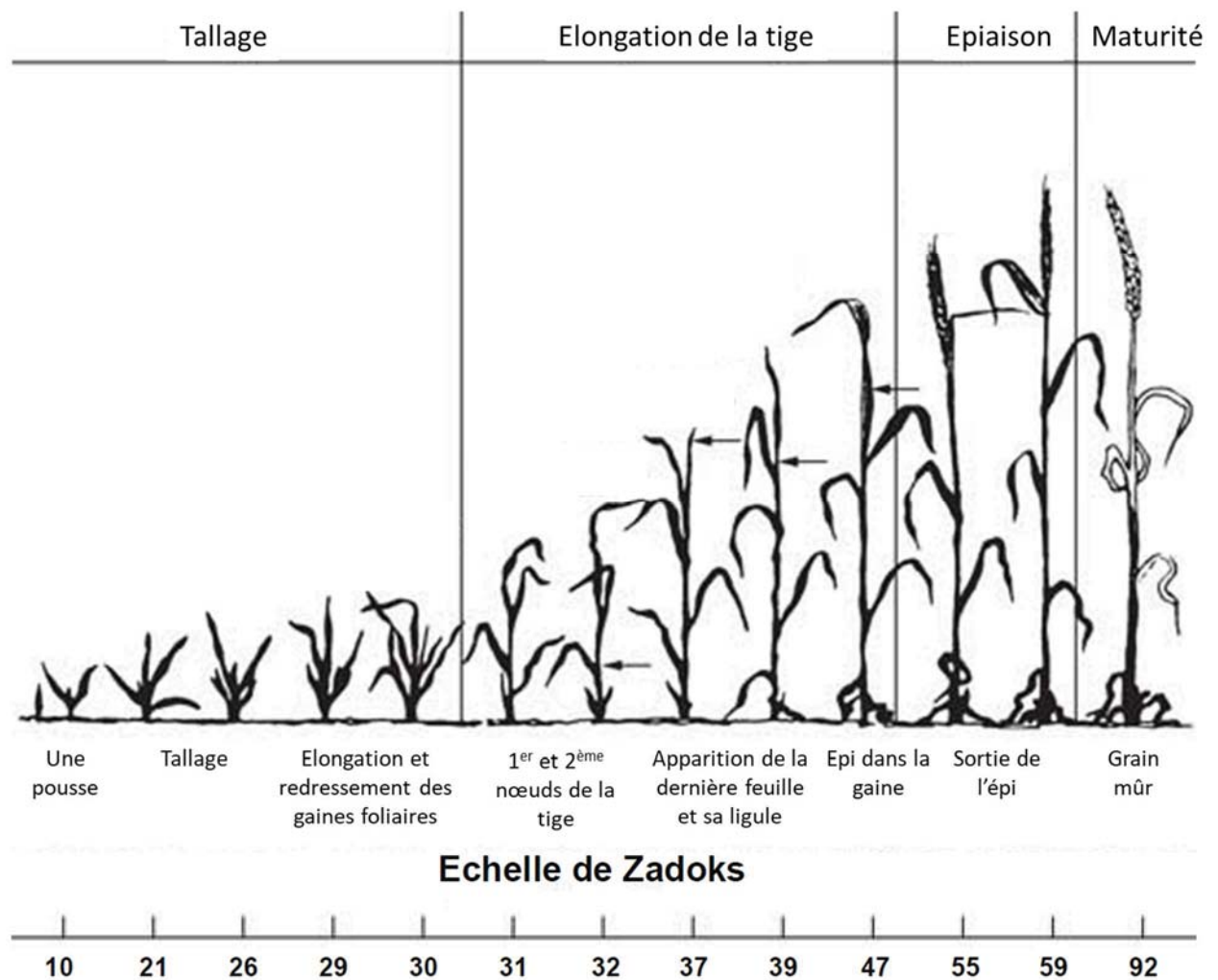


Figure 5 : Stades de développement du blé selon Zadoks et al. (1974). La période végétative se déroule du tallage à l'épiaison, puis la période reproductive a lieu jusqu'à la maturité du grain. Chaque stade est désigné par un nombre allant de 0 à 99.

caractéristiques des régions méditerranéennes et tempérées ; semés à l'automne, ils tolèrent bien le froid. Les blés de printemps sont semés au printemps dans des pays à hiver très rude.

Le développement du blé se compose de 2 périodes : la période végétative et la période reproductive, structurées en plusieurs phases clés. Plusieurs échelles de classification visuelle du développement du blé existent, basées sur l'existence de ces phases ; la plus simple d'utilisation est l'échelle de Zadoks (Zadoks *et al.* 1974) (Figure 5). Durant le développement végétatif, 4 phases physiologiques peuvent être dissociées : germination, levée, tallage et montaison ; suivies de l'épiaison et la maturation qui vont constituer la période reproductive. À l'échelle du grain, cette dernière est subdivisée en 3 grandes phases : divisions cellulaires, remplissage en réserves et maturation du grain.

1.4.1. La période végétative

Cette phase s'étend de la germination jusqu'à la montaison et permet à la plante de développer feuilles et racines. Pour germer, le grain de blé doit présenter un contenu en eau de 35 à 45% de son poids et être à une température comprise entre 4 et 37°C, avec un optimum entre 12 et 25°C. Le potentiel germinatif n'est pas impacté par la taille du grain mais celle-ci peut jouer un rôle sur la vitesse de croissance de la plantule, le nombre de talles fertiles par plante et le rendement, notamment dans le cas d'un stress hydrique (Acevedo *et al.* 2002).

Le tallage correspond au stade de mise en place et de développement des talles, qui se déroule en 2 étapes. Tout d'abord, un méristème axillaire est initié sur le plateau de tallage, à l'aisselle des premières feuilles de la plantule de blé ; puis cette talle s'allonge et croît. Le nombre potentiel de talles varie en fonction du génotype et de facteurs environnementaux. Par exemple, un plus grand nombre de talles est observé sur les blés d'hiver et les variétés semi-naines (Acevedo *et al.* 2002). Cependant, beaucoup de talles avortent avant l'anthèse, ne produisant donc pas d'épi.

1.4.2. La période reproductive

La période reproductive démarre lors de la formation de l'épi, elle se poursuit avec la formation du grain puis sa maturation.



Figure 6 : Epi de blé tendre en floraison. La floraison (ou anthèse) s'observe à partir du moment où les étamines jaunes sortent des glumelles (Crédit photo : CC By SA).

1.4.2.1. Initiation florale

Cette phase se met en place lorsque le plant de blé a établi entre 4 et 8 feuilles. L'apex cesse de former des ébauches de feuilles, il s'allonge jusqu'à environ 0,5 mm, se segmente et forme des ébauches d'épi. Au centre de chaque épi, les épillets possèdent 8 à 12 fleurs et seulement 6 à 8 fleurs sont contenues dans les épillets distaux et basaux. Plus de la moitié de ces fleurs avortera avant anthèse. Le nombre d'épillets par épi est déterminé à ce stade, qui est particulièrement sensible à des stress azoté et hydrique (Acevedo *et al.* 2002).

1.4.2.2. Montaison

Cette phase s'étend sur une durée relativement peu variable, entre 28 et 30 jours. La croissance des talles cesse tandis que les entrenœuds s'allongent très rapidement. Le développement de l'épi se met en place dès l'apparition de l'avant-dernière feuille et se poursuit jusqu'à 10 JAA (Jour Après Anthèse). Le dernier nœud s'élève rapidement et la formation de l'épi s'accélère lorsque la ligule de la dernière feuille devient visible. Acevedo *et al.* (2002) rapportent que l'avortement des fleurs excédentaires qui a lieu à ce stade de développement serait probablement dû à une compétition pour le carbone. Une autre compétition s'instaure entre les talles portant des épis en formation et ceux qui n'ont pu monter faute de nutriments, ces derniers vont régresser puis mourir. Cette phase prend fin avec le gonflement de la gaine, provoqué par l'épi prêt à en émerger, qui coïncide avec la différenciation des stigmates des fleurs.

1.4.2.3. Epiaison

L'épiaison correspond à l'émergence de l'épi de la gaine de la dernière feuille et dure environ 32 jours. Durant l'épiaison, les organes floraux achèvent leur formation par une méiose qui donnera le pollen dans les anthères et le sac embryonnaire dans le carpelle. Cette méiose commence au centre de l'épi pour s'étendre de part et d'autre de cette zone (Zadoks *et al.* 1974).

1.4.2.4. Anthèse

L'anthèse, ou floraison, s'observe lorsque les anthères sortent des glumelles de la fleur. D'abord jaunes, les anthères blanchissent ensuite rapidement (Figure 6). Ce stade est pris comme point de référence pour la fécondation. Chaque épillet contient potentiellement entre 3 et 6 fleurs qui vont s'auto-polliniser dans 96 % des cas, ce qui fait du blé une plante autogame.

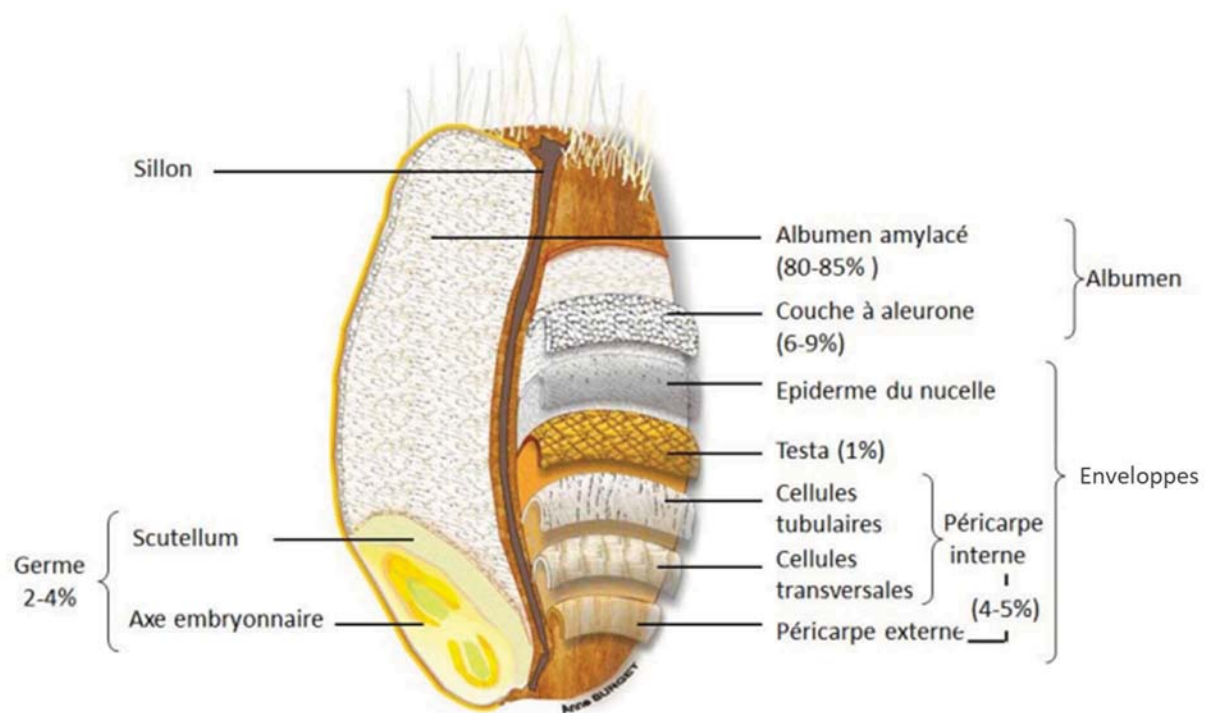


Figure 7 : Anatomie schématique du grain de blé tendre et proportions relatives de ses principaux tissus (d'après Surget et Barron 2005). Le grain de blé est composé de 3 parties principales : les enveloppes (nucellus, testa, péricarpe interne et externe), l'albumen et le germe (scutellum et axe embryonnaire).

Chez le blé tendre, l'anthèse débute dans la partie centrale de l'épi pour s'étendre à ses extrémités, sur une période de 3 à 5 jours. La fécondation des fleurs de la partie centrale de l'épi a donc lieu plus tôt que celle des fleurs distales, ce qui explique le fait que les grains centraux aient en général un poids plus élevé que les grains distaux. Après fertilisation, le développement du grain va passer par 3 phases : d'abord une phase de divisions cellulaires, puis le remplissage des cellules en amidon et protéines pour finir par une phase de maturation ou dessiccation du grain (Acevedo *et al.* 2002).

2. Le grain de blé

2.1. Anatomie du grain de blé

Comme c'est le cas des autres graminées, le blé produit un fruit sec indéhiscant, aussi appelé caryopse, dont les téguments de la graine et le péricarpe du fruit sont soudés. Le grain de blé est composé de 3 parties : les enveloppes, l'albumen et le germe (Figure 7).

2.1.1. Les enveloppes

Les enveloppes sont composées de 4 couches différenciées, de l'intérieur du grain vers l'extérieur :

- **Le nucellus (appelé aussi épiderme du nucelle, couche nucellaire ou bande hyaline)** est une assise cellulaire entourant directement l'embryon et l'albumen. Il est relié à la couche à aleurone par une fine cuticule et jouerait un rôle majeur dans la circulation de l'eau entre l'intérieur et l'extérieur du grain,
- **Le testa (ou tégument séminal)** est composé de 2 couches de cellules aux axes cellulaires perpendiculaire entre eux et riches en lipides chez le blé (une seule couche chez le riz) (Evers et Bechtel 1988). Les cuticules plus épaisses que dans le nucellus en font un tissu très hydrophobe et rendent le grain peu perméable à l'eau,
- **Le péricarpe interne** est composé des **cellules tubulaires** puis des **cellules transversales** (propres aux graminées), respectivement parallèles et perpendiculaires à l'axe longitudinal de l'embryon. Le péricarpe interne contient des chloroplastes qui donnent leur couleur verte aux grains immatures (Evers et Millar 2002),
- **Le péricarpe externe** est constitué de cellules mortes ayant la capacité de retenir l'eau et d'augmenter le poids du grain de 4 à 5% après seulement quelques minutes

d'imbibition (Hinton 1955). En effet, pendant le développement du grain, les péricarpes externe et interne évitent les pertes en eau mais n'empêchent pas sa pénétration (Evers *et al.* 1999).

2.1.2. L'albumen

L'albumen, aussi appelé amande, est le compartiment le plus important du grain : il représente environ 80% de son poids (Evers et Bechtel 1988) et joue le rôle de tissu de réserves. Il est constitué de 2 parties : l'albumen amylicé et la couche à aleurone.

L'albumen amylicé, principalement constitué de granules d'amidon enchâssées dans une matrice protéique, forme une masse solide au centre du grain. Ses cellules aux parois fines peuvent être classées en 3 types (Evers et Millar 2002) :

- les cellules périphériques, sous la couche à aleurone,
- les cellules prismatiques, agencées en colonne,
- les cellules centrales, de forme arrondie ou polygonale.

La couche à aleurone est une monocouche de cellules aux parois épaisses et au noyau volumineux, jouant un rôle nourricier et protecteur en entourant l'albumen amylicé du grain de blé. La taille de ces cellules, riches en protéines, lipides, vitamines et minéraux, est indépendante de la taille du grain. Chez certaines céréales comme le maïs, les pigments de types anthocyanes contenus dans la couche à aleurone peuvent colorer le grain mature de teintes bleues, rouges voire noires. Dans le grain mature, les cellules de la couche à aleurone sont les seules cellules vivantes restantes de l'albumen. Après imbibition du grain, des enzymes protéolytiques et hydrolytiques sont produites dans ces cellules pour digérer les parois cellulaires de l'albumen amylicé et permettre la mobilisation des réserves pour la croissance de l'embryon (Sabelli et Larkins 2009).

2.1.3. Le germe

Le germe est composé de 2 parties : l'embryon ou axe embryonnaire, qui donnera ensuite la tigelle, le mésocotyle et la radicule de la future plantule ; et le scutellum qui donnera le cotylédon, séparé de l'albumen par une assise cellulaire responsable de la digestion de l'albumen lors du développement de la plantule (Evers et Millar 2002 ; Surget et Barron 2005). Le germe est la partie du grain où le taux d'humidité et la concentration en lipides sont les plus

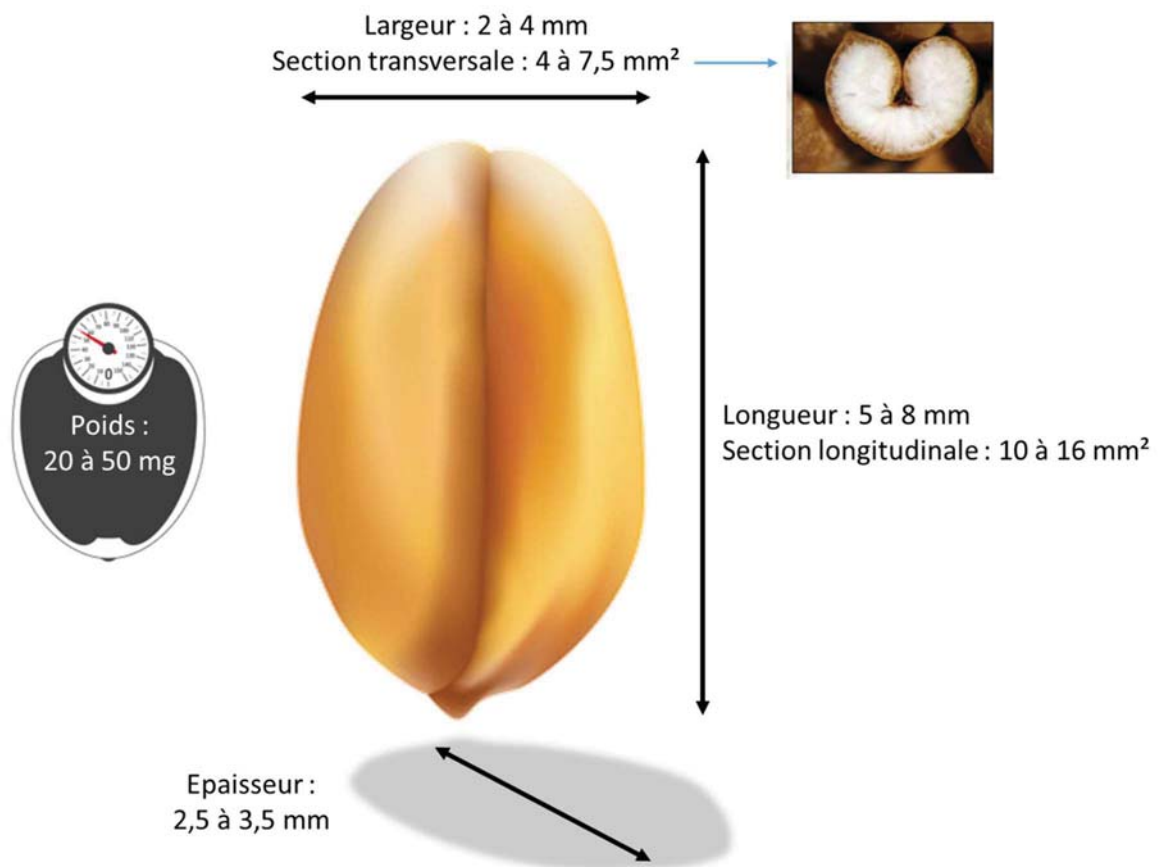


Figure 8 : Dimensions moyennes d'un grain de blé. Crédit images : Fotofolia, Arvalis.

importants (Evers et Bechtel 1988). Une fois le grain mature, le germe représente 3% de sa masse sèche totale (Barron *et al.* 2007).

2.1.4. Morphologie standard d'un grain de blé

Comme les autres membres de la famille des *Triticae*, le grain de blé tendre possède un sillon sur toute sa longueur, du côté du germe. Ce sillon est la résultante d'une invagination des téguments vers l'intérieur du grain. Les faisceaux nourriciers du grain en cours de développement sont localisés au fond de ce sillon (Soltner 1988). Les dimensions d'un grain de blé standard sont présentées en Figure 8. Néanmoins, la taille et le poids du grain sont dépendants, entre autres, de sa position dans l'épi (Feillet 2000 ; Calderini et Ortiz-Monasterio 2003).

2.2. Composition biochimique du grain de blé

Le grain de blé est composé d'amidon (70% de la matière sèche), de protéines (10 à 15% de la matière sèche), d'eau (12 à 14% de la matière fraîche), de pentosanes (2 à 3% de la matière sèche) et en proportions plus minoritaires, de lipides, cellulose, sucres libres, minéraux et vitamines.

2.2.1. L'amidon

La synthèse des granules d'amidon de blé a lieu à partir de la conversion du saccharose en ADP-glucose. Ces granules sont des très grands polysaccharides composés de 2 types de polymères de glucose, l'amylose et l'amylopectine, arrangés en structure semi-cristalline (Sabelli et Larkins 2009). La qualité de l'amidon dépend de sa teneur relative entre amylopectine et amylose, qui varie en fonction des génotypes. Les granules d'amidon sont classées en 3 types selon leur taille : granules de classe A (15-30 μm), de classe B (<10 μm) et de classe C (< 5 μm) (Bechtel *et al.* 1990 ; Shewry 2009). La distribution de ces granules est affectée par les conditions environnementales ; par exemple un stress hydrique diminue le nombre de granules B tandis que des fortes températures diminuent la taille des granules A (Hurkman *et al.* 2003 ; Lu *et al.* 2014).

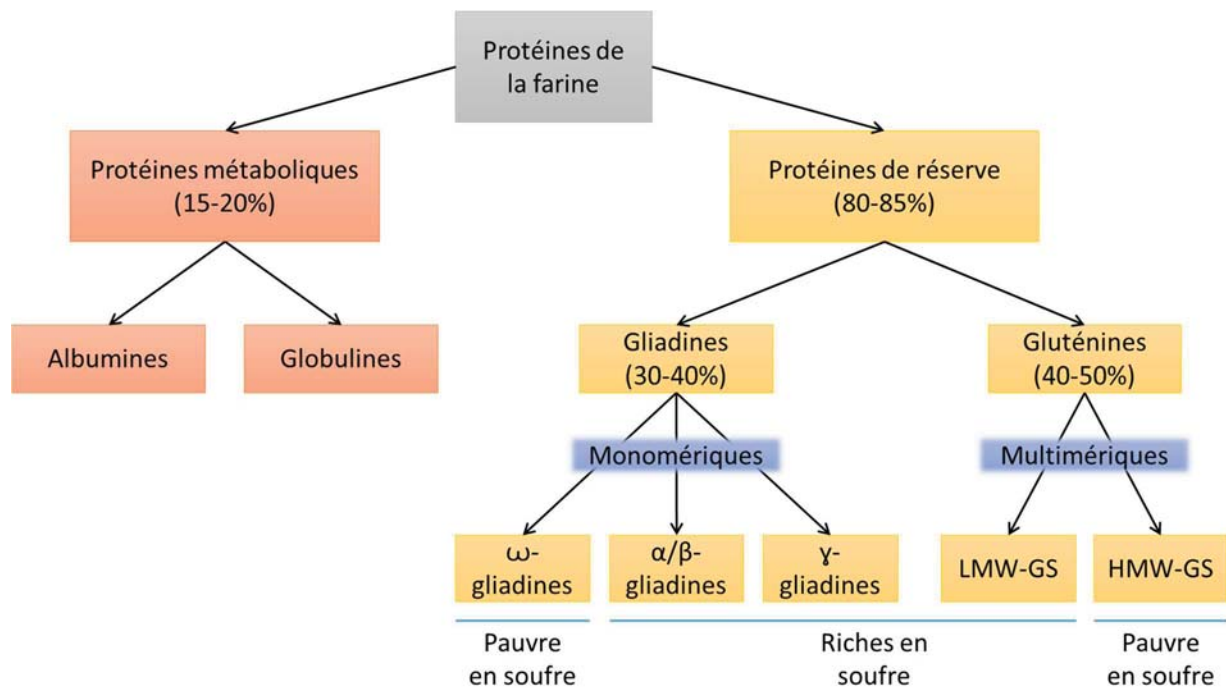


Figure 9 : Composition des protéines de la farine de blé, rapprochement entre les classifications d'Osborne et de Shewry (d'après Feillet 2000). La part de chaque classe protéique est exprimée en pourcentage des protéines totales.

Durant la maturation du grain, les amyloplastes accumulent cet amidon qui sera la principale ressource énergétique permettant le démarrage de la germination de la future plantule. Il s'accumule exclusivement dans l'albumen amidonné, tandis que les protéines vont se répartir inégalement entre les différents tissus. La couche à aleurone et le germe sont d'ailleurs les fractions les plus riches en protéines.

2.2.2. Les protéines

Le blé est une céréale riche en protéines (12-15%) en comparaison du riz ou du maïs, qui contiennent de 2 à 3% (Abhinandan *et al.* 2018). Les protéines du grain de blé sont généralement classées selon leur solubilité (Osborne 1924), elles se différencient en :

- albumines (solubles dans l'eau),
- globulines (solubles dans une solution saline neutre),
- gliadines (solubles dans une solution d'alcool à 70%),
- gluténines (insolubles dans les solvants précités mais solubles dans des solutions acides ou alcalines) (Figure 9).

Shewry *et al.* (1986) ont proposé une révision de cette classification pour former 2 grandes classes : les protéines métaboliques ou cytoplasmiques (albumines, globulines et petites protéines riches en cystéine) et les protéines de réserve (gliadines et gluténines). Elles peuvent aussi être classées selon leur degré de polymérisation : protéines polymériques (gluténines, albumines de haut poids moléculaire et triticines) ou monomériques (gliadines, globulines et les autres albumines).

Les albumines et globulines représentent 15 à 20% de la farine et sont parfois réunies sous le nom de protéines solubles. Ces protéines interviennent entre autres dans le métabolisme cellulaire ou les voies de régulation.

Les gliadines et les gluténines sont parfois réunies sous le nom de prolamines, du fait de leur richesse en proline. Elles sont synthétisées dans l'albumen et constituent les protéines de réserve dans lesquelles la jeune plantule puisera les acides aminés dont elle a besoin au moment de sa germination. Les gliadines sont des protéines monomériques de 30 à 80 kDa, réparties en 4 classes : α -, β -, γ - et ω -gliadines. Les gluténines sont des protéines multimériques et les sous-unités qui les composent sont réparties en 2 groupes : les sous-unités gluténines de haut poids moléculaire (HMW-GS, 67-88 kDa) et les sous-unités gluténines de faible poids moléculaire

	Stades				
	I	II	III	IV	V
Jours après anthèse [équivalent en Degrés Celsius Jours]*	0-7 [0-121]	7-14 [121-242]	14-21 [242-362]	21-31 [362-535]	31-50 [535-863]
Couleur du grain	Blanc	Blanc / Vert menthe	Vert	Vert	Brun
Longueur du grain (mm)	< 5,5	5,5-6,5	6,5-7,0	6,5-7,0	6,5
Poids frais du grain	< 20	20-35	35-50	50-60	40-50
Caractéristique de l'albumen	Aqueux	Laiteux, début de pâteux	Pâteux mou	Pâteux dur	Dur
Caractéristique du péricarpe	Charnu	Contracté	Réduit et flexible	Sec et fin	Sec et fin
Longueur de l'embryon, scutellum inclus (mm)	< 0,4	0,4-1,5	1,5-2,5	2,5-2,7	2,5-2,7

Tableau 1 : Caractéristiques des différents stades de développement du grain de blé tendre (d'après Rogers et Quatrano 1983). Les phases I et II sont les phases précoces et tardives de la division cellulaire, les phases III et IV sont les phases précoces et tardives du remplissage et la phase V est la phase de maturation. *: photopériode 18h/6h à 19°C/12°C.

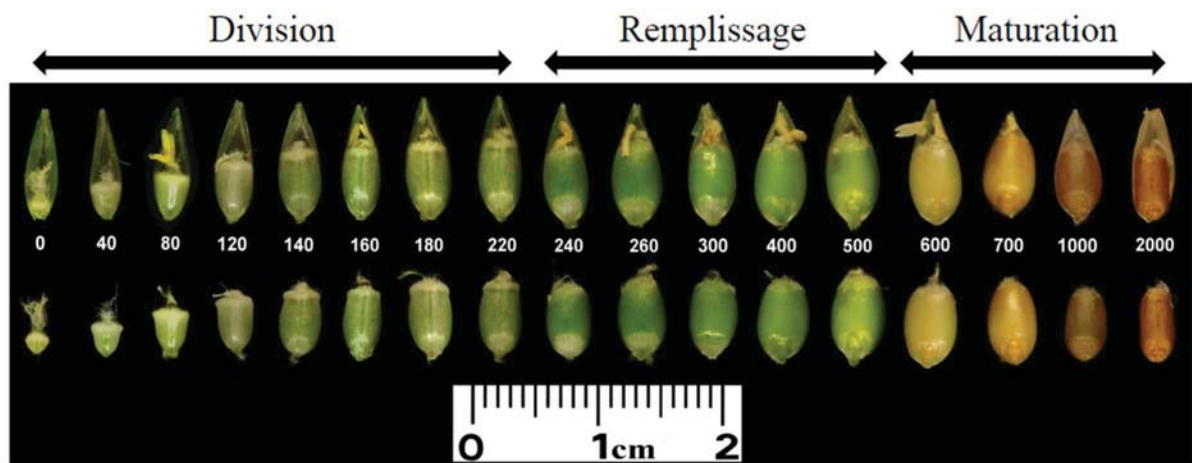


Figure 10 : Développement du grain de blé (d'après Capron et al. 2012). Le développement du grain de blé peut être divisé en 3 phases : la division cellulaire durant laquelle le grain acquiert sa longueur finale ; le remplissage des cellules en réserve, provoquant l'élargissement du grain ; et la maturation-dessiccation du grain, ce qui entraîne une diminution de sa taille. Le grain est représenté avec et sans son enveloppement (respectivement, ligne supérieure et inférieure). Les chiffres indiquent les stades de développement exprimés en °Cj (degrés Celsius jour).

(LMW-GS, 36-44 kDa). Les gliadines sont responsables de la composante visqueuse du comportement rhéologique des pâtes et les gluténines de leur composante élastique (Naeem *et al.* 2012 ; Barak *et al.* 2013). Elles représentent respectivement 30-40% et 40-50% des protéines de la farine et sont les principaux constituants du gluten, réseau protéique complexe dont les propriétés viscoélastiques sont capitales en panification. Le gluten est un mélange de protéines (75-85%), lipides (5-7%), amidon (5-10%) et d'eau (5-8%) (Feillet 2000). Le rapport gliadines/gluténines influence ainsi fortement les qualités rhéologiques de la pâte.

2.3. Phases de développement du grain de blé

Le développement du grain de blé débute avec la double fécondation de l'ovule. L'albumen ($3n$) est formé par la fusion de l'un des 2 gamètes mâles avec les 2 noyaux polaires de la cellule centrale du sac embryonnaire tandis que l'embryon ($2n$) est formé grâce à la fécondation de l'oosphère par le second gamète mâle. Classiquement, le développement du grain est ensuite structuré en 3 phases basées sur le développement de l'albumen : les phases de divisions cellulaires, de remplissage et de maturation (Sabelli et Larkins 2009). En 1983, Rogers et Quatrano (1983) ont proposé une subdivision du développement du grain de blé en 5 phases, décrites dans le Tableau 1. Des paramètres tels que le génotype, la température, la longueur du jour et la date de semis vont influencer sur la durée de chacune des phases. Par ailleurs, d'autres paramètres peuvent raccourcir la durée de développement du blé, comme une température élevée, un stress hydrique ou salin.

Au cours de son développement, l'albumen subit une succession de processus biologiques : la formation d'un syncytium, sa cellularisation, ses différenciations et divisions cellulaires, puis l'accumulation des réserves et pour finir la maturation du tissu comprenant la mort cellulaire programmée, la dormance et la dessiccation (Sabelli et Larkins 2009) (Figure 10). Néanmoins, la transition entre ces étapes a lieu de manière graduelle et les processus physiologiques qui caractérisent chacune d'elles peuvent se superposer au cours du développement du grain.

2.3.1. Phase de divisions cellulaires

Après la double fécondation et jusqu'à environ 3-4 JAA, la cellule triploïde connaît une succession de divisions sans cytokinèse, formant un cœnocyte (ou syncytium) contenant une large vacuole centrale et plus de 2000 noyaux répartis en périphérie (Mares *et al.* 1975). De 3-

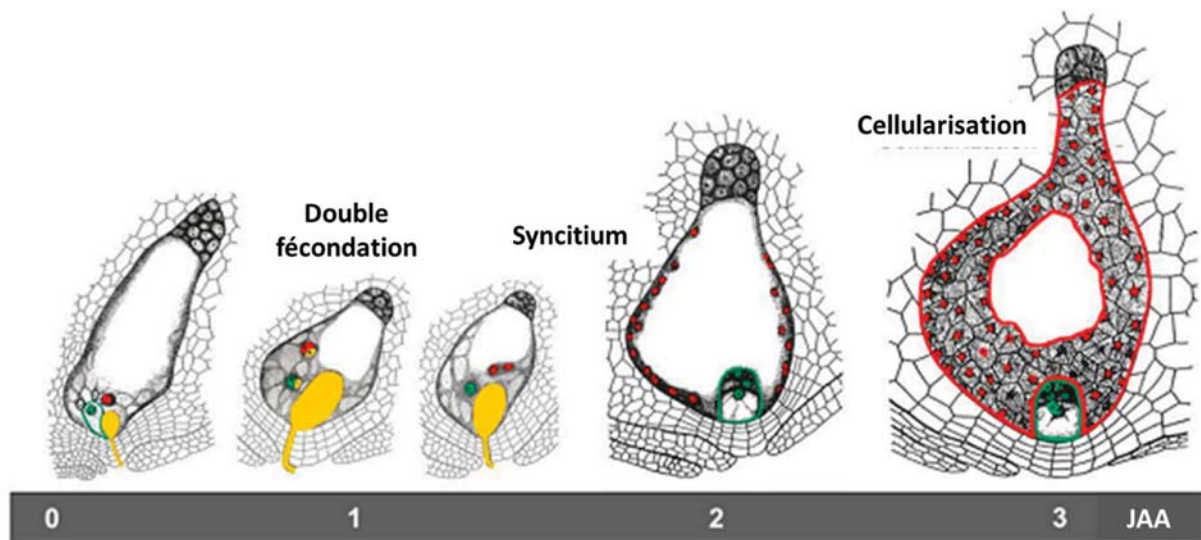


Figure 11 : Phases de développement de l'albumen de maïs de 0 à 3 jours après anthèse (Sabelli et Larkins 2009). Suite à l'anthèse, une double fécondation mène à la formation d'un syncytium puis d'une cellularisation de l'albumen. Le tube pollinique et le noyau du pollen sont représentés en jaune, les noyaux des cellules de l'albumen ainsi que le contour de celles-ci sont indiqués en rouge. Enfin, le noyau de l'embryon et le contour de celui-ci est coloré en vert. JAA = Jours Après Anthèse.

4 JAA jusqu'à environ 11 JAA, une étape d'isolation des noyaux par formation de parois cellulaires (cellularisation) est suivie de divisions et d'expansions cellulaires qui aboutissent à la formation d'un albumen possédant jusqu'à 300 000 cellules (Figure 11). Durant 10 jours, les principaux types cellulaires de l'albumen se différencient : cellules de transfert, de la couche à aleurone, de l'albumen amylicé et de la région entourant l'embryon.

Au début de cette phase précoce, l'albumen et l'embryon sont entourés par le nucelle, tissu maternel nourricier nécessaire à la croissance rapide de l'albumen. Au cours des divisions cellulaires, ce nucelle est rapidement dégradé afin de libérer de l'espace pour l'expansion des cellules de l'albumen. Dans le grain mature, il ne restera que l'épiderme nucellaire (Drea *et al.* 2005). Le cytoplasme dense des cellules de transfert est riche en mitochondries sphériques de petite taille. Elles facilitent l'accumulation rapide de nutriments (principalement saccharose et acides aminés) par l'albumen (Sabelli et Larkins 2009). La couche à aleurone est composée d'une seule couche de cellules, différenciée à la périphérie de l'albumen du grain de blé. Les cellules de l'albumen amylicé accueillent des granules d'amidon dont la formation débute à la fin de la cellularisation à 3-4 JAA et s'arrête brusquement à 7 JAA. Plusieurs couches de cellules composent la région entourant l'embryon. Elles se différencient dès la fin de la cellularisation et semblent impliquées dans l'apport de sucres à l'embryon par une voie apoplastique, dans la défense contre des pathogènes et dans la signalisation à l'interface embryon-albumen (Cosségal *et al.* 2007). Cette région disparaît avec la croissance de l'embryon et il n'en reste que des vestiges à la base de l'albumen à 12 JAA. Les divisions cellulaires de l'albumen et de l'embryon cessent après 11 JAA. La longueur du grain est fixée à la fin de cette phase, qui joue un rôle clé dans la détermination du poids du grain mature car celui-ci est proportionnel au nombre de cellules de l'albumen (Rogers et Quatrano 1983 ; Chojecki *et al.* 1986).

2.3.2. Phase de remplissage

Une fois que le nombre de cellules disponibles dans le grain est établi, vient le temps de leur remplissage en produits de réserves, essentiellement de l'amidon et des protéines. Le profil d'accumulation de l'amidon est similaire dans le temps à celui des protéines (Shewry *et al.* 2012). Cette phase débute vers 12 JAA, est maximale de 14 à 21 JAA puis ralentit jusqu'à 42 JAA. À ce moment, la concentration en eau du grain est proche de 45 g pour 100 g de matière fraîche (Schnyder et Baum 1992).

Les cellules sous la couche à aleurone accumulent beaucoup plus de protéines que les cellules centrales de l'albumen, certaines d'entre-elles ne se remplissent d'ailleurs que de protéines (Shewry *et al.* 2012). Parmi ces protéines, on retrouve principalement des gliadines et gluténines de faible poids moléculaires (Tosi *et al.* 2009, 2011), tandis que la proportion des protéines métaboliques (albumines et globulines) décline durant toute la phase de remplissage (Hurkman *et al.* 2009). Ces protéines de réserve fourniront les nutriments nécessaires à la germination (Soltner 1988).

À l'inverse des protéines, l'amidon s'accumule préférentiellement dans les cellules centrales de l'albumen (Shewry *et al.* 2012). Cette accumulation de réserves dans l'albumen est influencée par le ratio acide abscissique/auxine (ABA/Aux) (Yang *et al.* 2001 ; Xu *et al.* 2007). À la fin de la phase de remplissage, la largeur et l'épaisseur du grain atteignent leur maximum.

2.3.3. Phase de maturation

La dernière phase du développement du grain est la phase de maturation – dessiccation. Durant cette phase, on observe :

- l'arrêt de l'accumulation des réserves et la mort cellulaire programmée de l'albumen, qui débute lentement vers 16 JAA pour culminer vers 30 JAA (Young et Gallie 2000 ; Sabelli 2012), elle concerne toutes les cellules de l'albumen exceptées celles de la couche à aleurone. Cette mort cellulaire programmée permettrait de faciliter l'hydrolyse des nutriments et donc leur absorption par l'embryon lors de la germination (Nguyen *et al.* 2007),
- la mise en dormance du grain, correspondant à une inhibition de sa germination qui ne sera de nouveau possible que lorsque les conditions environnementales seront favorables. Les processus de mort cellulaire et de dormance du grain impliquent des voies de régulation faisant intervenir l'ABA (Sabelli et Larkins 2009),
- la dessiccation, une perte très rapide en eau, remplacée au sein des cellules par l'amidon et les protéines. Le grain atteint la maturité physiologique à 42 JAA, à ce stade son taux de matière sèche a atteint son niveau maximal (Altenbach *et al.* 2003 ; Wan *et al.* 2008). La masse fraîche du grain chute très rapidement à partir de 42 JAA, grâce à la dessiccation (Shewry *et al.* 2012).

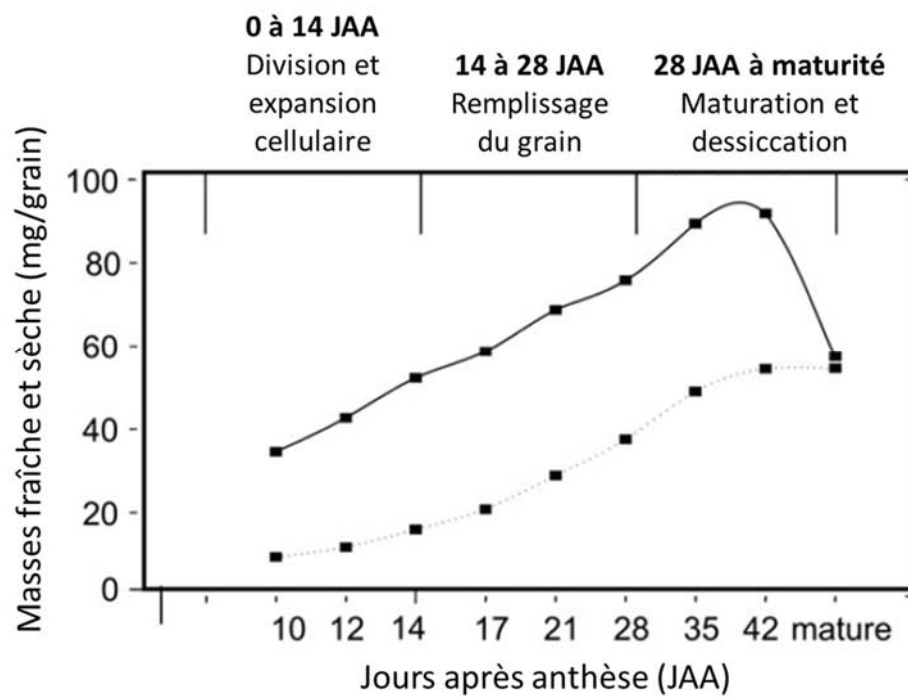


Figure 12 : Profil d'évolution des masses fraîche (trait continu) et sèche (trait pointillé) d'un grain de blé tendre (d'après Shewry et al. 2012).

2.4. Statut hydrique du grain au cours de son développement

Le contenu du grain de blé en eau est le reflet de son évolution au cours des 3 grandes phases de développement (Figure 12). La phase de divisions cellulaires est couplée à une accumulation importante d'eau durant les premiers jours pour atteindre son maximum vers 10-14 JAA. Lors de la phase de remplissage, pendant 2 à 3 semaines, le contenu en eau se stabilise, c'est le plateau hydrique (Schnyder et Baum 1992). Il chute ensuite rapidement à environ 30 JAA, lorsque le remplissage des cellules décroît et la dessiccation commence. La teneur en eau est alors de l'ordre de 40% du poids frais (Feillet 2000). La croissance initiale du grain dans les phases précoces de son développement est en grande partie due à l'accumulation d'eau (Schnyder et Baum 1992). Par ailleurs, l'arrêt de la division cellulaire a lieu à peu près au même moment que l'arrêt de cette accumulation rapide d'eau dans le grain (Nadaud *et al.* 2010). Schnyder et Baum (1992) ont proposé que le contenu en eau détermine le volume de croissance du grain au cours de la phase de division cellulaire de l'albumen et la taille finale des cellules serait influencée en partie par l'accumulation en eau (Egli 1998). Au cours de la phase de remplissage, la dynamique d'accumulation de la matière sèche est étroitement liée au statut hydrique de l'albumen. D'ailleurs, la fin d'accumulation de la matière sèche et de l'eau dans l'albumen est conjointe (Schnyder et Baum 1992 ; Calderini et Reynolds 2000). Au cours de la dessiccation du grain, son statut hydrique chute à 10-15%. La réduction des activités métaboliques durant la phase de maturation semble notamment causée par cette perte massive en eau (Young et Gallie 2000).

3. Effets des principaux stress abiotiques de l'environnement sur le développement et la croissance du blé

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 ; 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 <http://www.fao.org>). 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.

Or, 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'ores et déjà un impact négatif du changement climatique sur le rendement de grandes cultures céréalières (maïs, blé, 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. Les composantes du rendement sont définies par les critères suivants :

- le nombre de plantes par m²,
- le nombre de talles et d'épis par plante,
- le nombre d'épillets et de grains par épi,
- le Poids de Mille Grains (PMG).

Parmi ces composantes, les plus importantes sont le nombre de grains par plante et le PMG. Le PMG est déterminé par les longueur x largeur x épaisseur du grain et est dépendant du nombre de cellules et de leur remplissage, qui sont établis durant les 2 premières phases de son développement (divisions cellulaires et remplissage) (Campbell *et al.* 1999 ; Dholakia *et al.* 2003). La largeur du grain est la composante la plus importante du PMG car elle n'est pas exclusivement reliée au rendement en grains mais présente également des effets sur la vigueur et la précocité de croissance de la plantule de blé.

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). La réponse de la culture aux stress abiotiques est génotype dépendante ; l'amélioration variétale passe donc 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).

3.1. Effets des stress sur le développement des céréales à l'échelle du peuplement et de la plante

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 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 ; Girousse *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.

3.1.1. Stress thermique

De manière générale, une augmentation modérée de la température accélère le développement de la plante, ce qui conduit à une réduction de sa phase photosynthétique et donc à une diminution du potentiel maximal d'apport des organes sources. Des plants de riz cultivés sous contrainte thermique (+9°C) ont montré une diminution de la croissance des structures aériennes et souterraines, que ce soit chez un géotype sensible ou tolérant (Kilasi *et al.* 2018).

Le stress thermique influence généralement l'activité photosynthétique, et la réduction de la teneur en eau causée par la chaleur impacte négativement la division et la croissance cellulaire (Hasanuzzaman *et al.* 2013). Cependant, l'effet du stress thermique sur le développement et la physiologie de la plante dépend de l'intensité, la durée et l'amplitude de variation de la température (Wahid *et al.* 2007 ; Zinn *et al.* 2010). En outre, la perturbation des processus fondamentaux (comme l'assimilation du carbone, la respiration ou la transpiration) peut

entraîner des défauts de développement végétatif tels que des organes moins nombreux, plus petits et/ou malformés (Takeoka *et al.* 1991 ; Maestri *et al.* 2002).

La température élevée peut également affecter négativement la reproduction sexuelle et par conséquent le rendement en fruits ou en grains (Peet *et al.* 1997 ; Erickson et Markhart 2002 ; Zinn *et al.* 2010). Durant la période reproductive, Saini *et al.* (1984) ont montré qu'une température de 30°C pendant la méiose perturbe la division cellulaire et conduit à un développement anormal du pollen, engendrant une réduction du nombre de grains sans compensation au niveau du PMG des grains restants.

Un stress thermique altère la durée et la vitesse du remplissage du grain. Au-delà d'une température moyenne journalière de 20°C, la durée du remplissage du grain est réduite (Wiegand et Cuellar 1981). La température peut également plus ou moins impacter la vitesse de remplissage en réserves en fonction de la variété considérée (Sofield *et al.* 1977 ; Tashiro et Wardlaw 1990 ; Wardlaw et Moncur 1995). Les fortes températures réduisent également le nombre final de cellules de l'albumen (Acevedo *et al.* 2002). Hors de la gamme de température optimale pour la croissance (entre 12°C et 25°C), Acevedo *et al.* (2002) ont estimé que le poids du grain réduit de 4 à 8% par °C supérieur à la gamme de température de croissance optimale. Il semble cependant moins impacté par cette augmentation de température que ne l'est le nombre de grains.

En raison d'une diminution de la durée et de la vitesse de remplissage du grain en protéines, la teneur en protéines totales du grain est plus faible chez un blé soumis à de fortes températures par rapport au contrôle (Wiegand et Cuellar 1981). Cependant, Gooding *et al.* (2003) ont montré une augmentation du taux de protéines en réponse à une contrainte thermique modérée. Ceci est dû au fait que l'accumulation d'amidon dans le grain est plus sévèrement affectée par la température que l'accumulation d'azote, qui est corrélée au taux final de protéines. En effet, il existe un mécanisme de compensation entre vitesse et durée de remplissage pour les protéines du grain, ce qui n'est pas le cas pour l'amidon (Triboi et Triboi-Blondel 2002). De manière générale, les facteurs agro-climatiques entraînent des variations de teneur protéiques plus importantes que les facteurs génétiques (Cooper *et al.* 2001). Girousse *et al.* (2018) ont proposé l'hypothèse que les sucres et les protéines accumulées ne joueraient pas uniquement le rôle de nutriments mais qu'ils interviendraient en tant que signaux déclencheurs des réponses développementales adaptatives.

3.1.2. Stress hydrique

Au moment de la germination, les stress abiotiques peuvent impacter le pourcentage de germination des semis et donc la densité de culture. Un manque d'eau ou une salinité élevée empêchent l'imbibition du grain, nécessaire à sa germination (Dodd et Donovan 1999). De plus, un retard de germination peut exposer la plantule à des températures de plus en plus élevées avec l'avancée de la saison.

Durant le tallage, l'application d'un stress impacte négativement le nombre d'épis, mais aussi les structures végétatives comme la surface foliaire ou encore l'accumulation de matières sèches (Ihsan *et al.* 2016). Le stress hydrique provoque une réduction de l'activité photosynthétique, ralentissant la synthèse et l'accumulation des sucres qui seront ensuite nécessaires au développement des grains (Shangguan *et al.* 1999 ; Subrahmanyam *et al.* 2006).

La méiose mâle (microsporogénèse) et la déhiscence des anthères sont 2 stades particulièrement sensibles (Barnabás *et al.* 2008 ; Craufurd *et al.* 2013). Un stress hydrique au stade jeune microspore du développement du pollen conduit à la formation d'un pollen anormalement stérile et donc la réduction du nombre de grains (Ji *et al.* 2010). Un stress hydrique lors de l'anthèse impacte le développement du méristème floral, conduisant à des avortements et affectant la taille des grains restants (Ji *et al.* 2010 ; Dolferus *et al.* 2011).

Le stress hydrique diminue la taille du grain en raccourcissant sa période de remplissage sans modifier la vitesse d'accumulation des réserves (Brooks *et al.* 1982 ; Nicolas *et al.* 1984 ; Kobata *et al.* 1992 ; Altenbach *et al.* 2003). Par ailleurs, Ahmadi et Baker (2001) ont mis en évidence l'impact d'un stress hydrique sur la qualité du grain. En effet, ils ont observé une réduction du contenu des grains en saccharose et en amidon en réponse à un stress hydrique appliqué en fin de phase de divisions cellulaires-début de remplissage (180°Cj ou 270°Cj jours après anthèse). Cet effet est corrélé avec une altération de l'activité des enzymes de biosynthèse de l'amidon comme l'ADP-Glucose-Pyrophosphorylase ou la Granule-Bound Starch Synthase. Celles-ci sont sensibles à un stress hydrique et la modification de leur activité explique l'obtention de grains de blé pauvres en amidon (Ahmadi et Baker 2001). Cependant, il a été remarqué que lors d'une exposition prolongée à un déficit hydrique, leur activité reste relativement constante (Caley *et al.* 1990 ; Ahmadi et Baker 2001). À cause des modifications osmotiques, l'activité ADP-glucose pyrophosphorylase est aussi fortement ralentie par un déficit hydrique.

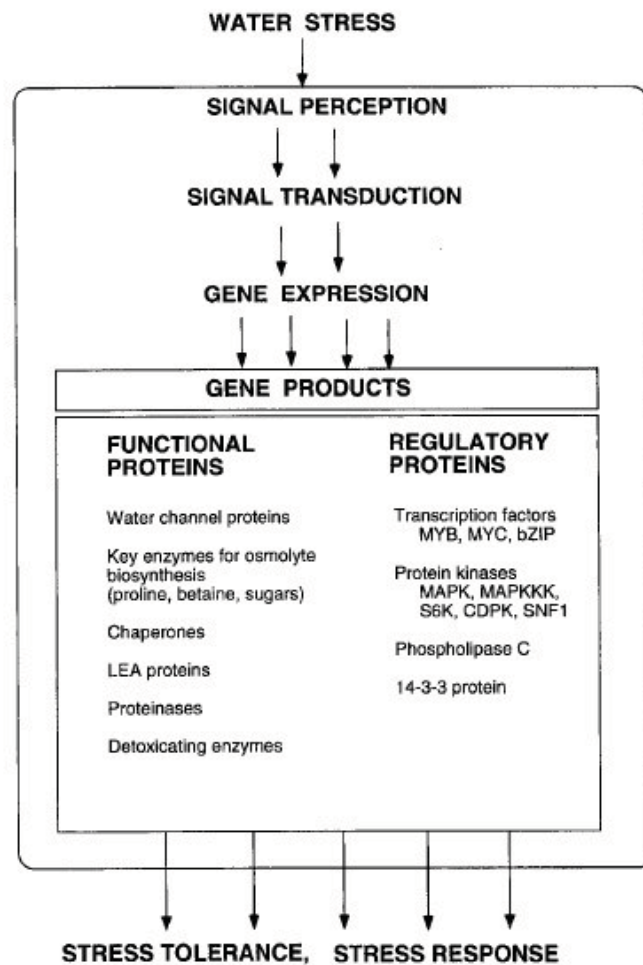


Figure 13. Fonction des gènes dont l'expression est induite par un stress hydrique, dans la réponse et la tolérance à ce stress. Globalement, ces gènes sont classés en 2 groupes : les protéines fonctionnelles impliquées dans la tolérance au stress hydrique et l'adaptation cellulaire, et les protéines régulatrices pouvant intervenir dans l'expression des gènes et la transduction du signal en réponse au stress (tiré de Shinozaki et Yamaguchi-Shinozaki 1997).

3.1.3. Stress salin

Les salinités élevées sont un stress qui affecte aux alentours de 20% des terres arables, et il va concerner jusqu'à la moitié des sols cultivés en 2050 (Shrivastava et Kumar 2015). Un stress salin appliqué au cours du développement des grains induit une diminution de leur nombre et de leur poids (Francois *et al.* 1984). Zheng *et al.* (2008) ont noté que l'application d'un stress salin durant la croissance du blé induit une réduction de l'activité photosynthétique, la conductance stomatique, la quantité de pigments et d'ions, la surface, le contenu en eau et la durée de vie des feuilles, et de l'accumulation de matières sèches dans les épis.

3.1.4. Multi-stress

En 2013, Rasmussen *et al.* (2013) ont montré que, lorsque des plants d'*Arabidopsis* sont soumis à des multi-stress, 61% des modifications transcriptomiques observées en réponse à ces multi-stress ne sont pas prédites par la réponse des plantes aux traitements mono-stress.

En environnement naturel, les contraintes sont souvent multiples. De nombreuses publications étudient le stress hydrique en serre, en conditions contrôlées, mais les réponses des plantes à ce mono-stress ne reflètent que partiellement leur réponse à un stress sécheresse en plein champs qui sera très probablement couplé à une contrainte thermique par exemple. En effet, l'interaction entre 2 stress ne représente pas uniquement l'addition des effets de chaque stress séparément mais crée une synergie. Shah et Paulsen (2003) ont montré que les effets d'un stress hydrique (diminution de la photosynthèse, la conductance stomatique, la surface foliaire, le poids des grains et le taux de sucres solubles contenus dans les épis) se trouvent amplifiés lorsque l'on ajoute un stress thermique à la culture. *A contrario*, l'efficacité de l'utilisation de l'eau par la plante, qui est améliorée par le stress hydrique, se trouve diminuée par le stress thermique. Selon les auteurs, la synergie qui se met en place entre les 2 stress réduit considérablement la productivité du blé, principalement en impactant les processus photosynthétiques. Blum (1996) a montré que la sécheresse en plein champs affecte la croissance des structures aériennes et souterraines ainsi que l'accumulation de matières sèches, la photosynthèse et la conductance stomatique. De plus, elle induit un évitement de la part de la plante qui va accélérer sa floraison afin d'échapper à ce stress, ce qui a pour effet d'impacter négativement son rendement (Kottmann *et al.* 2016). Par ailleurs, sa perception par la plante en début de période reproductive conduit à une réduction de la demande en carbone en diminuant le nombre de puits vs les sources ; cela passe par une dégénérescence des talles, la mort du

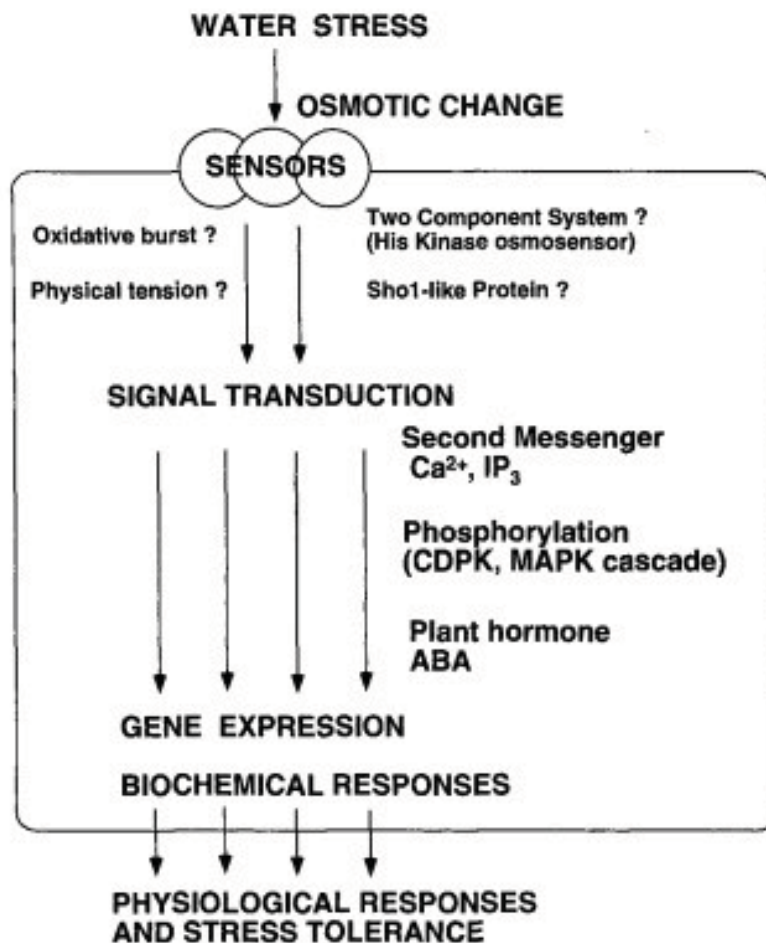


Figure 14 : Messagers secondaires et facteurs impliqués dans la perception et la transduction du signal dans la réponse au stress hydrique. La kinase His est supposée fonctionner comme osmosenseur chez les plantes. Le calcium (Ca^{2+}) et l'inositol 1,4,5-triphosphate (IP_3) sont les messagers secondaires les plus probables du signal stress hydrique. Le processus de phosphorylation fonctionne via des voies de transduction du signal de stress hydrique et des voies ABA. L'ABA joue un rôle important dans la régulation de l'expression des gènes et dans les réponses physiologiques en cas de stress hydrique. (Tiré de Shinozaki et Yamaguchi-Shinozaki 1997).

pollen ou encore un avortement des ovules (Blum 1996). Le développement du grain est aussi très sensible à la sécheresse en plein champ qui peut, par exemple, conduire à une réduction de la taille du grain en affectant la phase de remplissage (Sanjari Pireivatlou et Yazdansepas 2010).

Gouache *et al.* (2014) ont proposé une méthode de mesure de l'impact des stress par le suivi de l'« indice foliaire vert intégré », révélateur d'une sénescence prématurée des surfaces vertes de la plante. L'intérêt principal de cet indice est sa capacité à appréhender des stress multiples et serait selon son auteur un moyen efficace de caractériser la tolérance de nouvelles variétés à des nombreux stress.

Comment alors les signaux de stress sont-ils perçus et intégrés par la plante ? Quels sont les mécanismes moléculaires de la mise en place d'une réponse aux stress abiotiques ?

3.2. Réponses de la plante aux stress hydrique et thermique aux niveaux moléculaire et cellulaire

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) (Figure 13).

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 14). Les gènes régulateurs du signal stress sont des gènes impliqués dans la transduction

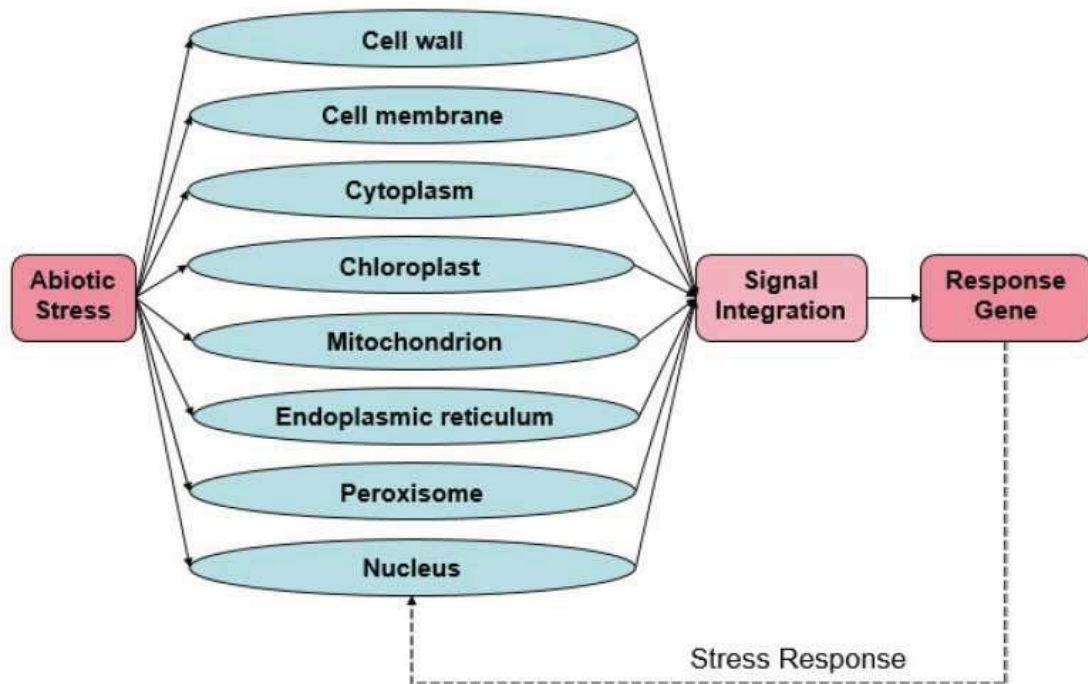


Figure 15 : Détection du stress et signalisation ; un modèle de réponse aux stress abiotiques dans différents organites (tiré de Zhang et al. 2018c). Le signal généré par les organites peut provoquer l'expression des gènes et des activités cellulaires, ce qui peut restaurer l'homéostasie cellulaire sous stress abiotique.

du signal : les récepteurs ancrés à la membrane, les senseurs de calcium, les protéines kinases, les protéines phosphatases et les facteurs de transcription.

La détection du signal stress a souvent lieu à la surface des membranes cellulaires, pour être ensuite relayé vers plusieurs organites (Figure 15). 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 de ROS tels que les anion superoxydes, le peroxyde d'hydrogène (H_2O_2), les radicaux hydroxyle et l'oxygène singulet affecte le fonctionnement des chloroplastes et impacte la photosynthèse. H_2O_2 et O_2^- 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 *catalase2* 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).

3.2.1. Réponse à un stress hydrique

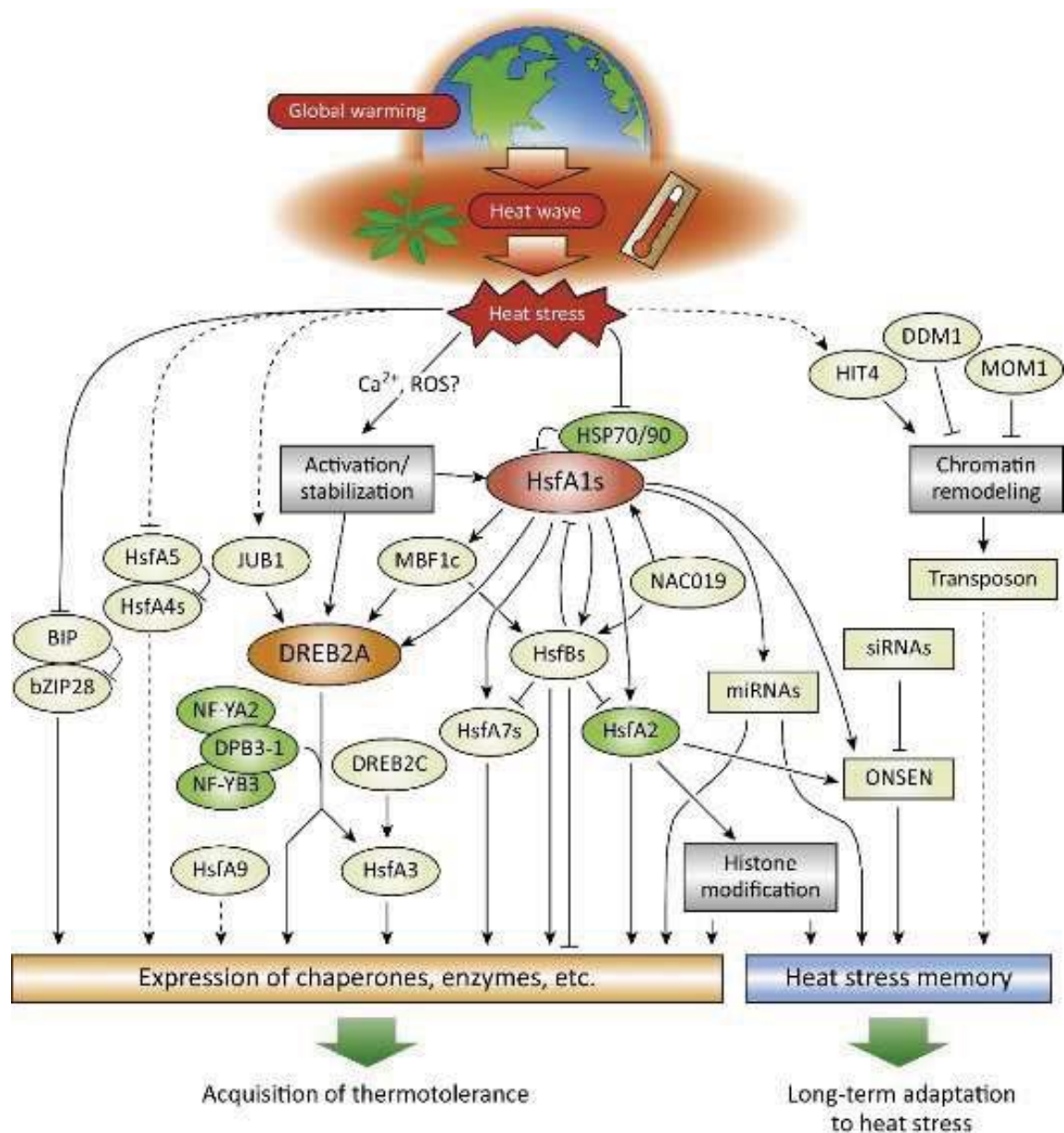
Les protéines kinases et phosphatases interviennent dans la perception et la réponse aux stress environnementaux. La surexpression de plusieurs TaSnRK2 (Sucrose non fermenting 1-Related protein Kinase 2) de blé améliore la tolérance des Arabidopsis transgéniques face à de nombreux stress abiotiques (Xu *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 protein 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'H₂O₂ 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).

3.2.2. Réponse à un stress thermique

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



Trends in Plant Science

Figure 16 : Réseau de régulation transcriptionnelle impliqué dans la réponse au stress thermique (tiré de Ohama et al., 2017). Les flèches continues présentent les relations observées par la communauté scientifique. Les lignes pointillées indiquent les liens à confirmer.

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. 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.

3.3. Rôles des facteurs de transcription dans la réponse au choc thermique

Les facteurs de transcription 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 (Figure 16) ; 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 1 500 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 transcriptionnel complexes comprenant de nombreux régulateurs de la transcription sont maintenant connus en réponse au choc thermique.

Les Heat shock transcription factor A1 (HsfA1) sont des facteurs de transcription jouant un rôle de régulateur indispensable dans les voies de réponses au choc thermique. La diminution (*knock-down*) ou l'extinction (*knock-out*) de leur expression conduit à une réduction de l'expression des gènes liés à la réponse au choc thermique et engendre des plantes sensibles à ce stress (Mishra *et al.* 2002 ; Liu *et al.* 2011 ; Yoshida *et al.* 2011). Yoshida *et al.* (2011) ont montré qu'en réponse à un choc thermique, les protéines HsfA1 sont relocalisées dans le noyau et induisent l'expression de gènes codant pour les facteurs de transcription sensibles au choc thermique, comme la protéine DREB2A (Dehydration-Responsive Element Binding protein 2A). Les protéines HSP90, en interagissant avec les protéines HsfA1, régulent négativement leur accumulation dans le noyau. En condition contrôle, les protéines DREB2A sont ciblées par les protéines DREB2a-Interacting Protein 1 et 2 (DRIP1 et DRIP2, de la famille des E3 ligases) qui les inactivent et permet leur dégradation par le protéasome 26S (Qin *et al.* 2008b). En réponse à un stress thermique, DREB2A et 3 autres facteurs de transcription forment un complexe activateur de la transcription qui se fixe sur le promoteur de HsfA3 et induit son expression (Sato *et al.* 2014). DREB2C est un homologue de DREB2A chez Arabidopsis, et il active aussi l'expression de HsfA3 en réponse à un stress thermique (Chen *et al.* 2010). *A contrario*, HsfB1 et HsfB2b sont des répresseurs transcriptionnels qui régulent négativement l'expression des HSP et des HsfA2, HsfA7a, HsfB1, et HsfB2b en conditions non stressantes ou lors de la phase d'atténuation après l'exposition au stress. Leur activité est aussi requise pour l'acquisition de la thermotolérance (Ikeda *et al.* 2011).

Le gène bZIP28 est un facteur de transcription qui régule la réponse à un stress thermique. Sa protéine possède un domaine transmembranaire et, sous l'action de la protéine BiP (Binding Protein), elle est ancrée à la membrane du réticulum endoplasmique en condition contrôle. En réponse à un choc thermique ou une modification des conditions redox, la protéine bZIP28 est libérée et relocalisée dans le noyau où elle régule l'expression de gènes cibles (Srivastava *et al.* 2014). Kataoka *et al.* (2017) ont montré que, lorsque la voie de régulation de bZIP28 est déficiente, d'autres voies sont activées pour compenser et éviter l'altération de la tolérance d'Arabidopsis à un stress thermique. Parmi elles, on retrouve les voies dépendantes de l'ascorbate peroxidase ou des HSP, ainsi qu'une augmentation de l'expression de HsfA2 et de l'accumulation d'H₂O₂ en réponse à un choc thermique. La protéine BiP intervient aussi dans la régulation des voies de réponse au stress hydrique. Sa surexpression chez le soja conduit à une accumulation des transcrits de plusieurs gènes induits par le stress hydrique comme le gène NAC2, une glutathione-S-transferase ou une protein disulphide isomerase (Valente *et al.* 2009).

Tableau 2 : Recensement des membres de la famille NAC chez un large panel d'espèces végétales.

Espèce	Nombre de gènes	Référence bibliographique
Abricotier du Japon	113	(Zhuo <i>et al.</i> 2018)
<i>Arabidopsis thaliana</i>	115 / 117 / 106	(Shen <i>et al.</i> 2009; Nuruzzaman <i>et al.</i> 2010; Fan <i>et al.</i> 2018)
Banane	172	(Cenci <i>et al.</i> 2014)
Blé dur	168	(Saidi <i>et al.</i> 2017)
<i>Brachypodium distachyon</i>	118 / 101	(You <i>et al.</i> 2015; Zhu <i>et al.</i> 2015)
Cacaoyer	100	(Fan <i>et al.</i> 2018)
Canne à sucre	85	(Ramaswamy <i>et al.</i> 2017)
Colza	60	(Wang <i>et al.</i> 2015)
Cotonnier en arbre	147 / 142	(Fan <i>et al.</i> 2018; Sun <i>et al.</i> 2018)
Cotonnier mexicain	283 / 211	(Fan <i>et al.</i> 2018; Sun <i>et al.</i> 2018)
Eucalyptus	189	(Hussey <i>et al.</i> 2015)
Fraisier cultivé	112	(Moyano <i>et al.</i> 2018)
Fraisier des bois	37	(Zhang <i>et al.</i> 2018a)
Haricot	86	(Wu <i>et al.</i> 2016)
Luzerne tronquée	93	(Shen <i>et al.</i> 2009)
Maïs	177 / 124 / 152 / 157 / 148	(Shen <i>et al.</i> 2009; Fan <i>et al.</i> 2014; Shiriga <i>et al.</i> 2014; Lu <i>et al.</i> 2015; Peng <i>et al.</i> 2015)
Manioc	96	(Hu <i>et al.</i> 2015)
Oignon	39	(Zheng <i>et al.</i> 2016)
Orge	48	(Christiansen <i>et al.</i> 2011)
Pastèque	80	(Lv <i>et al.</i> 2016)
<i>Physcomitrella patens</i>	35	(Shen <i>et al.</i> 2009)
Peuplier de l'Ouest	148 / 170	(Shen <i>et al.</i> 2009; Yao <i>et al.</i> 2018)
Poivre	104	(Diao <i>et al.</i> 2018)
Radis	172	(Karanja <i>et al.</i> 2017)
Riz	140 / 144 / 151	(Fang <i>et al.</i> 2008; Shen <i>et al.</i> 2009; Nuruzzaman <i>et al.</i> 2010)
<i>Selaginella mœllendorffii</i>	42	(Shen <i>et al.</i> 2009)
Sarrasin de Tartarie	80	(Liu <i>et al.</i> 2019)
Sésame	87	(Zhang <i>et al.</i> 2018b)
Soja	101 / 177 / 139	(Pinheiro <i>et al.</i> 2009; Shen <i>et al.</i> 2009; Hussain <i>et al.</i> 2017)
Sorgo	113/131	(Shen <i>et al.</i> 2009; Sanjari <i>et al.</i> 2019)
Théier	45	(Wang <i>et al.</i> 2016)
Tomate	74	(Kou <i>et al.</i> 2014)
Vigne	75 / 68	(Shen <i>et al.</i> 2009; Fan <i>et al.</i> 2018)

Chez le blé, TaWRKY1 et TaWRKY33 sont 2 facteurs de transcription nucléaires dont l'expression est induite par le stress hydrique, le choc thermique et l'ABA. Les basses températures induisent l'expression de TaWRKY33, tandis qu'elles répriment celles de TaWRKY1 (He *et al.* 2016). Six gènes de la famille des facteurs de transcription MYB de blé sont induits par des températures élevées. Parmi eux, TaMYB80 améliore la tolérance aux fortes températures et au stress hydrique par une voie ABA-dépendante lorsqu'il est surexprimé chez *Arabidopsis* (Zhao *et al.* 2017).

Parmi ces facteurs de transcription, il existe aussi la famille NAC qui est de plus en plus étudiée par la communauté scientifique. De nombreux NAC ont été décrits comme intervenant dans le développement de multiples espèces et leur réponse à un large panel de stress.

4. Les facteurs de transcription NAC

4.1. Historique

La famille NAC est l'une des plus grandes familles de facteurs de transcription spécifiques aux plantes, et l'expression de ses membres est modulée lors du développement de la plante et dans sa réponse aux stress biotiques et abiotiques (Olsen *et al.* 2005). L'acronyme **NAC** est tiré de **NAM** (No Apical Meristem) – ATAF1 et 2 (*Arabidopsis* Transcription Activation Factor) – **CUC2** (Cup-shaped Cotyledons), des gènes caractérisés indépendamment et qui contiennent tous un domaine NAC. Souer *et al.* (1996) ont décrit le premier gène NAC comme lié au développement du méristème apical des pousses et à la détermination de la position des méristèmes et des primordia (premières feuilles après la différenciation des méristèmes) chez le pétunia.

Depuis lors, les NAC ont été largement étudiés par la communauté scientifique et avec le développement des technologies de séquençage à haut débit, leurs membres ont été recensés chez de nombreuses espèces végétales d'intérêt (Tableau 2). Des gènes NAC ont été identifiés chez tous les groupes de plantes terrestres, les halophytes mais aussi chez les algues vertes streptophytes (Maugarny-Calès *et al.* 2016 ; Khedia *et al.* 2018). Cependant, ils ne semblent être présents que chez les végétaux car aucun homologue n'a été identifié chez les autres eucaryotes (Riechmann *et al.* 2000).

En 2000, Kikuchi *et al.* (2000) ont proposé une classification de la famille NAC du riz en 3 sous-familles. Les études comparatives entre génomes ont ensuite permis d'élaborer des sous-familles plus nombreuses et plus conservées entre espèces (Ooka *et al.* 2003 ; Fang *et al.* 2008

; Nuruzzaman *et al.* 2010). Rushton *et al.* (2008) ont estimé que la famille NAC pouvait être subdivisée en au moins 7 sous-familles fonction-spécifiques. La même année, Fang *et al.* (2008) ont proposé une classification en 5 groupes, le groupe I contenant les NAC en lien avec le développement de la plante tandis que les gènes liés au stress (appelés SNAC ou Stress-related NAC) sont réunis dans le groupe III. En 2009, Shen *et al.* (2009) ont recensé les NAC présents chez 11 organismes et répartis les séquences obtenues en 8 sous-familles nommées NAC-a à NAC-h, à partir de l'homologie entre les domaines NAC. Toutes ces sous-familles tendent à contenir des séquences issues d'espèces monocotylédones et dicotylédones. Shen *et al.* (2009) ont aussi constaté un regroupement par sous-famille des NAC ayant une fonction proche. Par exemple, des NAC impliqués dans la réponse aux stress biotiques et abiotiques tendent à être rattachés à la sous-famille NAC-a tandis que 89% des NAC liés à la membrane appartiennent à la sous-famille NAC-b ; et cette sous-famille pourrait aussi être associée avec la division cellulaire et les réponses au stress. Zhu *et al.* (2012) ont proposé une classification en 21 sous-familles, reflétant selon eux les similitudes de séquence mais aussi les relations évolutives entre les protéines.

4.2. Structure protéique et interaction avec l'ADN

Une protéine NAC typique contient un domaine hautement conservé de liaison à l'ADN dans sa région N-terminale, appelé domaine NAC. D'environ 150 aa (acides aminés), il se lie à la séquence core CACG dans les promoteurs de certains gènes cibles (Wan *et al.* 2008). Ce domaine est subdivisé en 5 sous-domaines nommés A à E (Ooka *et al.* 2003). Les sous-domaines A, C et D sont hautement conservés (Mathew *et al.* 2016). Du fait de leur rôle de facteurs de transcription, les protéines NAC possèdent la capacité de se fixer à l'ADN pour réguler l'expression de gènes cibles, seules ou en formant des complexes de régulation avec d'autres protéines. Par exemple, Greve *et al.* (2003) ont mis en évidence une interaction entre la protéine ANAC et 3 protéines possédant un domaine RING-H2 dans le noyau chez *Arabidopsis*. Le motif de reconnaissance ADN, WKATGTDK, est présent sur le sous-domaine C et hautement conservé entre les facteurs NAC (Welner *et al.* 2012). Ce motif est assez similaire au motif WRKYGQK présent chez les membres de la famille WRKY (Yamasaki *et al.* 2008).

Dans le promoteur des gènes cibles, les facteurs NAC se lient aux core-séquences NACBS (NAC Binding Site) ou NACRS (NAC Recognition Sequence). Par exemple, l'expression du

gène ERD1 (Early Responsive To Dehydration stress 1) est médiée par la fixation de NAC inductibles par le stress sur la séquence CACG de son promoteur (Tran *et al.* 2004). Chez le maïs, plusieurs membres de la famille NAC ont la capacité à se fixer au motif SNBE (Secondary wall NAC Binding Element) du promoteur de leurs gènes cibles (Xiao *et al.* 2018). Chez Arabidopsis, ce motif est une séquence palindromique imparfaite de 19 paires de bases : (T/A)NN(C/T) (T/C/G)TNNNNNNA(A/C)GN(A/C/T)(A/T) (Zhong *et al.* 2010). Chez la patate douce, IbNAC1 se fixe au motif TACAATATC qui se trouve dans le Sporamin Wound-Response cis-Element, une région de 53 pb (paires de bases) dans le promoteur du gène de la sporamine (Chen *et al.* 2016b).

Dans la région C-terminale, les gènes NAC possèdent aussi une région régulatrice moins étudiée, peu conservée et de longueur variable. Celle-ci agirait comme un activateur ou un répresseur transcriptionnel des gènes cibles (Xie *et al.* 2000 ; Duval *et al.* 2002 ; Hao *et al.* 2011 ; Wang *et al.* 2015b). Cette région régulatrice C-terminale contient souvent des répétitions simples d'acides aminés et des régions riches en sérine et thréonine, proline et glutamine, ou des résidus acides (Olsen *et al.* 2005). Le domaine NAC et le domaine C-terminal sont tous les 2 impliqués dans la dimérisation des protéines NAC, formant des homodimères ou des hétérodimères (Zhu *et al.* 2012 ; Mathew et Agarwal 2018). Ainsi, Xie *et al.* (2000) ont montré que le domaine NAC de la protéine NAC1 chez Arabidopsis lui confère la capacité de se dimériser.

Alors que la majorité des facteurs de transcription NAC ont été trouvés sous forme de protéines solubles dans le noyau, certains d'entre eux possèdent un domaine de liaison aux membranes nucléaires, plasmiques ou du réticulum endoplasmique (Kim *et al.* 2007b; Bhattacharjee *et al.* 2017). Ces facteurs sont classiquement regroupés dans une sous-famille appelée MTF (Membrane-associated Transcription Factor) ou NTL (NAC with Transmembrane motif 1 – Like). Par exemple, 9 NTL ont été identifiés chez le blé dur (Saidi *et al.* 2017), au moins 85 et 45 chez Arabidopsis et le riz respectivement (Kim *et al.* 2010) et 17 chez le radis (Karanja *et al.* 2017).

4.3. Rôles des NAC dans l'évolution des plantes

En 2012, Zhu *et al.* ont avancé l'hypothèse de l'apparition des NAC il y a plus de 400 millions d'années, lors de l'émergence des plantes terrestres et avant la divergence des Angiospermes et des Lycophytes. Les NAC se seraient ensuite diversifiés et leurs rôles se seraient complexifiés

en même temps que les organismes qui les portaient. Cette hypothèse a été ajustée par Maugarny-Calès *et al.* (2016) qui ont montré l'existence de NAC dans la classe des *Klebsormidiales* (algues vertes streptophytes). Leur origine serait donc antérieure à l'apparition des plantes terrestres, soit il y a 725–1200 millions d'années. Par ailleurs, ces auteurs ont suggéré que l'élargissement de la famille NAC s'est effectué par petits à-coups plutôt que selon un processus continu. Selon eux, le premier à-coup aurait eu lieu lors de l'apparition des plantes terrestres (soit il y a environ 470 millions d'années), et le second lors de l'apparition des Angiospermes, durant le Crétacé. Il est d'ailleurs possible que les NAC aient facilité le passage de l'environnement aquatique à terrestre en aidant à la coordination des réponses aux stress induits par cette transition. L'hypothèse du premier à-coup est appuyée par la forte conservation des NAC impliqués dans la conduction de l'eau chez la mousse *Physcomitrella patens* et *Arabidopsis*, fonction nécessaire chez les plantes terrestres (Xu *et al.* 2014).

L'histoire des NAC est aussi liée à celle de l'Homme. Par exemple, la texture des variétés de sorgho varie en fonction de leur utilisation (tige sèche pour le sorgho-fibre ou juteuse pour le sorgho doux). Le gène *Dry* contrôlant la quantité de jus de la tige est un NAC. La délétion ou la perte de fonction de ce gène est observée chez le sorgho doux, ce qui induit un collapsus cellulaire et une modification de la composition des parois cellulaires secondaires dans la tige. La sélection de plantes possédant une mutation du gène *Dry* a été une étape majeure pour l'apparition du sorgho doux qui est maintenant largement utilisé en agriculture. Ce gène, conservé chez les céréales, pourrait être un levier pour modifier la texture de la tige des plants cultivés (Zhang *et al.* 2018).

4.4. Rôles des NAC dans le développement de la plante

Les NAC forment une des plus larges familles de facteurs de transcription, aux rôles pléiotropiques dans le développement de la plante et dans sa réponse aux stress biotiques et abiotiques (Ernst *et al.* 2004). Ils interviennent dans de nombreuses étapes clés du développement des plantes, selon des voies hormones-dépendantes ou non. De par leur action ubiquitaire, l'expression des NAC est finement régulée de manière transcriptionnelle et post-transcriptionnelle. Selon des analyses réalisées chez *Arabidopsis* par Jensen *et al.* (2010), l'expression d'un tiers des NAC est modifiée en réponse à l'application d'un précurseur de l'éthylène. Dans le prochain paragraphe, l'implication des facteurs NAC au cours du développement sera décrite en prenant en compte des étapes clés du développement de la plante.

Chez *Arabidopsis*, les gènes NAC25 et NAC1L interviendraient dans la **germination** des grains en tant que régulateurs de l'expression du gène EXPA2 (Expansin 2), responsable du contrôle spatiotemporel de l'expansion cellulaire de l'albumen médiée par la gibbérelline (Sánchez-Montesino *et al.* 2019). De la même manière, les gènes NAM, CUC1, CUC2 et CUC3 interviennent dans le développement du **méristème apical** (Souer *et al.* 1996 ; Vroemen *et al.* 2003), mais aussi dans la séparation des cotylédons et des organes floraux (Aida *et al.* 1997). NTM1 (NAC with TransMembrane motif 1), est un facteur de transcription NAC dont le précurseur inactif est ancré à la membrane. Son clivage de la membrane donne une forme active qui peut être relocalisée dans le noyau et induire ou réprimer l'expression de gènes en réponse au stimuli ayant provoqué son activation/relocalisation. NTM1 est un régulateur négatif de la **division cellulaire** en lien avec la cytokinine (Kim *et al.* 2006). Le gène NAP (NAC-like activated by AP3/PI) est ciblé par la protéine APETALA3/PISTILLATA et joue un rôle dans la transition entre **division et expansion cellulaire** chez *Arabidopsis* (Sablowski et Meyerowitz 1998). Ce même gène NAP est un activateur de la sénescence foliaire chez *Arabidopsis* et probablement aussi chez le riz et le haricot rouge (Guo et Gan 2006). En effet, il a été montré que, chez le riz et *Arabidopsis*, ORE1/ANAC092/AtNAC2, ORS1, AtNAP sont des régulateurs positifs de la **sénescence foliaire** (Guo et Gan 2006 ; Balazadeh *et al.* 2008, 2011 ; Kim *et al.* 2009). Chez *Arabidopsis*, ANAC019, AtNAP, ANAC047, ANAC055, ORS1 et ORE1 sont associés à la sénescence par la voie éthylène, en aval des gènes EIN2 et EIN3 (Ethylene-Insensitive 2 et 3). EIN3 se lie aux promoteurs des gènes AtNAP et ORE1 pour induire leur expression, ceux-ci induisent ensuite l'expression d'autres gènes dont des NAC (Kim *et al.* 2014). ANAC019, ANAC055 et ANAC072 induisent, sous le contrôle de MYC2, MYC3 et MYC4, l'expression de gènes de catabolisme de la chlorophylle durant la sénescence foliaire. Des triples mutants *anac019-anac055-anac072* présentent un phénotype stay-green après traitement au méthyl-jasmonate (Zhu *et al.* 2015b). NAC1 est un autre gène NAC qui intervient positivement dans l'initiation des **racines latérales** en réponse à un signal auxine chez *Arabidopsis*, en activant l'expression des gènes DBP et AIR3 (Xie *et al.* 2000). L'accumulation des transcrits de NAC1 est régulée par le miR164, un microARN qui est aussi induit par l'auxine. Ce mécanisme permet de réguler le signal auxine au sein de la plante (Guo *et al.* 2005). Dans une autre étude, Pei *et al.* (2013) ont montré que l'accumulation de la protéine RhNAC100 chez le rosier est modulée par la voie de l'éthylène *via* une régulation post-transcriptionnelle par le miR164. Ils ont également montré que RhNAC100 diminue la taille des pétales de rose en réduisant **l'expansion cellulaire**, ceci en régulant l'expression d'au moins 22 gènes impliqués dans le développement. Parmi ceux-ci, une cellulose synthase et 2 aquaporines

Tableau 3 : Expression de plusieurs gènes NAC chez différentes espèces, en réponse à divers stress biotiques et abiotiques (tiré de Christianson et al. 2010). Les carrés verts symbolisent l'induction de la transcription du gène en réponse au stress, les carrés hachurés représentent sa répression. Les carrés blancs indiquent un résultat inconnu ou une absence de modification de l'expression du gène par le stress.

		ABA	Wounding	Drought	Salt	Cold	Pathogen
Arabidopsis	ATAF1						
	ATAF2						
	ANAC032						
	ANAC102						
Rice	OsNAC5						
	OsNAC6						
Sugarcane	SsNAC23						
Cotton	GhNAC1						
	GhNAC2						
	GhNAC3						
Barley	HvNAC6						
Cayenne pepper	CaNAC1						
Potato	SINAC						
Tomato	SINAC						
Rapeseed	BnNAC1-1						
	BnNAC3						
	BnNAC5-1						
	BnNAC5-7						
	BnNAC5-8						
	BnNAC5-11						
	BnNAC14						
	BnNAC18						
Soybean	GmNAC2						

pourraient être directement régulés par la fixation de RhNAC100 sur leurs promoteurs. MdNAC1 a été identifié par Jia *et al.* (2018) comme un acteur du **nanisme** chez le pommier. Sa surexpression engendre des arbustes aux tiges et aux racines plus courtes et plus fines, et présentant une surface foliaire diminuée. Le nombre de cellules dans les tiges est aussi diminué, de même que les niveaux d'ABA, de brassinostéroïdes, et des gènes impliqués dans leur synthèse. Chez le blé, GPC (Grain Protein Content) est un gène NAC dont la copie sur le chromosome 1B est absente ou non-fonctionnelle chez la plupart des variétés commercialisées, tandis que les copies 1A et 1D sont existantes. Le gène GPC est associé à une accélération de la **sénescence et une accumulation de protéines, de fer et de zinc dans les grains**, démontrant son rôle dans la remobilisation des nutriments (Avni *et al.* 2014). ZmNST3 et ZmNST4 sont 2 gènes NAC exprimés spécifiquement dans les cellules formant des parois secondaires chez le maïs. Ces 2 gènes se fixent sur le motif SNBE présent dans le promoteur des gènes ZmMYB109, ZmMYB128 et ZmMYB149 pour induire leur expression. La surexpression de ZmNST3 et ZmNST4 conduit à un épaississement des **parois secondaires** de la tige chez le maïs et Arabidopsis (Xiao *et al.* 2018). Chez Arabidopsis, SND1 et VND7 sont aussi des NAC régulant la biosynthèse et la modification de la **paroi secondaire** dans les fibres et les vaisseaux respectivement, ainsi que la **mort cellulaire** programmée *via* la régulation d'une cohorte de gènes (facteurs de transcription ou non) dont les promoteurs contiennent tous des motifs SNBE permettant la liaison des protéines NAC (Zhong *et al.* 2010). La famille NAC est également liée au **rendement et à la qualité des grains**. Par exemple, une expression réduite d'OsNAP chez le riz augmente le rendement en retardant la sénescence des feuilles (Liang *et al.* 2014). Chez le blé, Zhao *et al.* (2015) ont constaté une augmentation de la concentration en azote dans des grains issus de lignées sur-exprimant la protéine TaNAC-S, qui est un facteur NAC exprimé principalement dans les feuilles. Il permettrait une sénescence foliaire post-anthèse retardée (phénotype stay-green) ainsi qu'une meilleure remobilisation des nutriments durant cette sénescence tout en conservant un rendement identique aux lignées contrôles.

4.5. Rôles des NAC dans la réponse de la plante aux stress biotiques et abiotiques

Les facteurs de transcription NAC semblent donc faire le lien entre les signaux internes de la plante et sa réponse en terme de développement, mais ils jouent aussi un rôle dans la réponse aux signaux externes et tout particulièrement en réponse aux stress environnementaux biotiques et abiotiques (Tableau 3).

4.5.1. Stress biotiques

Chez le riz, Sun *et al.* (2015) ont montré que l'expression de 65 gènes NAC répond à au moins 2 stress biotiques parmi une infection fongique, bactérienne, virale ou des plantes parasites. Parmi ces gènes, ONAC026, ONAC069, ONAC095, ONAC108 et ONAC117 répondent à tous les stress biotiques testés ainsi qu'à des stress abiotiques tels que les stress salin, hydrique et froid. Ceci met en évidence le rôle pléiotropique de nombreux facteurs NAC chez le riz. C'est également le cas du gène OsNAC6 qui est un facteur de transcription nucléaire dont l'expression est induite par des stress abiotiques (stress hydrique, salin et froid) mais aussi par des blessures et par la pyriculariose. Il améliore la tolérance de la plante à ces stress, en activant notamment l'expression d'un gène codant pour une peroxidase (Nakashima *et al.* 2007). ONAC122 et ONAC131 codent également pour des protéines nucléaires dont l'expression est induite après une infection par *Magnaporthe grisea*, l'agent responsable de la pyriculariose; mais aussi en réponse à un traitement à l'acide salicylique, le méthyl-jasmonate ou un précurseur de l'éthylène. L'extinction de ces gènes provoque une sensibilité accrue à l'agent pathogène et une diminution de l'expression des gènes OsLOX, OsPR1a, OsWRKY45 et OsNH1, qui sont des gènes de défense et de signalisation (Sun *et al.* 2013). En 2015, Nuruzzaman *et al.* ont montré que l'expression de 75 gènes NAC chez le riz est impactée par au moins une infection par l'un des 4 virus testés (rice dwarf-virus, rice black-streaked dwarf virus, rice grassy stunt virus, rice ragged stunt virus et rice transitory yellowing virus).

Chez le blé, certains facteurs NAC sont aussi exprimés en réponse à un stress biotique. C'est le cas de TaNAC8, homologue de OsNAC8 du riz et fortement exprimé au cours du développement du grain de blé; et de TaNAC4, homologue de OsNAC4 du riz et préférentiellement exprimé dans les racines de la plantule de blé. L'expression de ces 2 gènes est induite dans les feuilles en réponse à une infection par *Puccinia striiformis* (responsable de la rouille jaune du blé) ainsi qu'à des traitements au méthyl-jasmonate ou à l'éthylène (mais pas à l'acide salicylique et seul TaNAC4 répond à l'ABA); mais aussi aux stress salin et froid. TaNAC8 est aussi surexprimé en réponse au stress hydrique, et TaNAC4 en réponse à une blessure (Xia *et al.* 2010b, 2010a). L'expression de TaNAC30, un activateur transcriptionnel nucléaire, est également induite en réponse à une infection par *P. striiformis* et régule négativement la résistance de la plante en inhibant l'accumulation d'H₂O₂ (Wang *et al.* 2018a). TaNAC1 est également un régulateur négatif de la résistance aux maladies fongiques (*P. striiformis*) et bactériennes (*Pseudomonas syringae*) chez le blé, agissant en amont des voies de signalisation des acides jasmonique et salicylique (Wang *et al.* 2015b). Un autre NAC,

TaNAC21/22, est un activateur transcriptionnel localisé dans le noyau et dont l'abondance est régulée de manière post-transcriptionnelle par le miR164. C'est également un régulateur négatif de la tolérance à la rouille jaune du blé (Feng *et al.* 2014).

Chez Arabidopsis, XND1 (Xylem NAC Domain 1) régule la conductivité hydraulique des racines d'Arabidopsis en agissant *via* la formation du xylème. Il apporte une résistance à la plante contre le pathogène racinaire *Ralstonia solanacearum* en levant l'inhibition de la formation de xylème induite par ce pathogène (Tang *et al.* 2018). Induites par une infection pathogène chez Arabidopsis, les protéines NAC4 se fixent sur les promoteurs des gènes LURP1, WRKY40 et WRKY54 qui sont des régulateurs négatifs de la mort cellulaire, et réprime leur expression. NAC4 promeut donc la mort cellulaire, sous le fin contrôle de miR164 qui le cible pour induire sa dégradation (Lee *et al.* 2017).

Le rôle des gènes NAC en réponse aux stress biotiques a aussi été étudié chez d'autres espèces. GbNAC1 est une protéine nucléaire exprimée dans les organes végétatifs du coton et plus particulièrement dans ses faisceaux vasculaires. Sa surexpression chez Arabidopsis améliore la résistance à la verticilliose, ainsi qu'aux stress salin, mannitol et ABA (Wang *et al.* 2016c). Chez la patate douce, IbNAC1 se lie au motif TACAATATC présent dans le promoteur du gène de la sporamine, une protéine de défense inhibitrice de la trypsine et conférant une résistance aux insectes, pour activer son expression dans les feuilles après une blessure. Lorsqu'il est surexprimé, IbNAC1 améliore la résistance contre *Spodoptera litura*. Par la voie de l'acide jasmonique, il intervient dans l'inhibition de la formation de racine, l'accumulation d'anthocyanes, la réduction des stress abiotiques et l'induction de l'expression des ROS (Chen *et al.* 2016b).

4.5.1. Stress abiotiques

De nombreux NAC ont été décrits en réponse aux stress abiotiques, et beaucoup présentent une modification de leur expression en réponse à plusieurs stress. Nous avons focalisé cette partie sur les stress salin, hydrique et thermique car ce sont les stress abiotiques les plus rencontrés.

4.5.1.1. Stress salin

Chez Arabidopsis, ANAC092/AtNAC2/ORE1 et ATAF1 sont des facteurs NAC dont l'expression est induite par une forte salinité et par l'ABA. Leur surexpression engendre une

sensibilité au stress salin chez *Arabidopsis*, tandis que la surexpression de ATAF1 chez le riz le rend tolérant à ce même stress (Balazadeh *et al.* 2010 ; Liu *et al.* 2016). GmNAC11 et GmNAC20 sont des protéines nucléaires exprimées préférentiellement dans les racines et les cotylédons du soja et capables de se fixer à la séquence ADN core CGT(G/A). GmNAC11 est un activateur transcriptionnel qui améliore la tolérance au stress salin ; tandis que GmNAC20 jouerait un rôle positif dans la réponse aux stress salin et froid par la voie DREB/CBF-COR et dans la formation des racines latérales par la voie auxine (Hao *et al.* 2011). Chez le blé, Xia *et al.* (2010a, 2010b) ont montré que les 2 gènes TaNAC4 et TaNAC8 sont impliqués dans la réponse du blé aux stress salin, basse température et à une blessure de la plante.

NTL8 est un facteur NAC possédant une forme dormante associée à la membrane chez *Arabidopsis* et chez le riz (Kim *et al.* 2007a, 2008). Elle peut être transformée en une forme nucléaire transcriptionnellement active en réponse à un stimulus et interviendrait dans le processus de floraison en condition de stress salin *via* la régulation de l'expression du gène FT (Flowering locus T). Ce facteur jouerait aussi un rôle dans l'inhibition de la germination en réponse au stress salin *via* la voie de l'acide gibbérellique et indépendamment de l'ABA (Kim *et al.* 2008). Kim *et al.* (2007b) suggèrent que la translocation des NAC ancrés à la membrane vers le noyau permettrait une réponse transcriptionnelle rapide de la plante aux stimuli externes et plus particulièrement aux stress abiotiques.

4.5.1.2. Stress hydrique

L'expression des NAC est régulée par la fixation de facteurs de transcription sur leurs promoteurs, notamment en réponse au stress hydrique. Par exemple chez le soja, les protéines GmMYB174 et GmWRKY27 interagissent pour se fixer sur le promoteur de GmNAC29 et réprimer l'expression de ce régulateur négatif de la tolérance au stress hydrique (Wang *et al.* 2015a). De même, l'expression de GsNAC019 est induite par la fixation du facteur de transcription Gshdz4 sur le cis-élément (CAATA/TA) de son promoteur (Cao *et al.* 2017).

De par leur fonction de facteur de transcription, les NAC peuvent également se fixer sur le promoteur de différents gènes de réponse au stress et induire ou réprimer leur expression, grâce à la présence de motifs spécifiques dans le promoteur de ces gènes. En se liant au cis-motif CATGTG présent dans le promoteur du gène ERD1, les 3 facteurs de transcription ANAC019, ANAC055 et ANAC072 jouent un rôle positif dans la tolérance d'*Arabidopsis* aux stress hydrique et salin (Tran *et al.* 2004). Au contraire, ATAF1 et ATAF2 en sont des régulateurs

négatifs (Delessert *et al.* 2005 ; Lu *et al.* 2007) mais ils jouent un rôle dans la résistance d'*Arabidopsis* aux pathogènes (Jensen *et al.* 2008).

NAC016 est un autre régulateur négatif de la réponse au stress hydrique chez *Arabidopsis*. Il se fixe au motif GATTGGAT(A/T)CA du promoteur des gènes cibles tels que les gènes AREB1 (ABA-Responsive Element Binding protein1) dont il réprime l'expression et NAP dont il active l'expression et qui, lui aussi, réprime l'expression de AREB1. Ce trio de gènes permet de réguler finement la réponse au stress hydrique et joue sur la sénescence foliaire d'*Arabidopsis* (Sakuraba *et al.* 2015).

Chez le blé, TaNAC47 est une protéine nucléaire capable de se lier au cis-élément ABRE (ABA-Responsive Element) et dont l'expression est induite par les stress hydrique, salin et froid selon une voie liée à l'ABA (Zhang *et al.* 2016). TaNAC29, ZmNAC55 et CarNAC4 sont aussi des protéines nucléaires impliquées dans la réponse aux stress hydrique et salin chez le blé, le maïs et le pois chiche respectivement ; de même que VaNAC26 chez la vigne de l'amour (vigne semi-sauvage, très tolérante aux stress environnementaux), ou encore MINAC9 et MINAC12 chez *Miscanthus*. Surexprimés chez *Arabidopsis*, ces gènes améliorent la tolérance aux stress hydrique et salin en réduisant l'accumulation de ROS et les dommages membranaires, *via* l'expression d'enzymes anti-oxydantes comme la SOD, la catalase et la peroxidase (Huang *et al.* 2015 ; Xu *et al.* 2015 ; Fang *et al.* 2016 ; Mao *et al.* 2016 ; Yu *et al.* 2016 ; Zhao *et al.* 2016 ; Yang *et al.* 2018). MINAC12 et VaNAC26 se fixent au motif NACRS présent dans le promoteur de leurs gènes cibles, tandis que CarNAC4 se lie au motif CGT(G/A) pour activer la transcription de gènes de réponse aux stress comme RD29A, ERD10, COR15A, COR47, KIN1 et DREB2A. PbeNAC1 chez *Pyrus betulifolia* et CsATAF1 chez le concombre sont des régulateurs positifs de la tolérance au stress hydrique, en diminuant l'accumulation des ROS (Jin *et al.* 2017a; Wang *et al.* 2018c). PbeNAC1 peut interagir avec PbeDREB1 et PbeDREB2A tandis que CsATAF1 régule directement l'expression de 3 gènes antioxydants : CsCu-ZnSOD, CsABI5, et CsDREB2C. En conditions non-stressantes, la protéine MfNACsa est associée à la membrane plasmique cellulaire de la luzerne sauvage *via* un ancrage lipidique. Sa relocalisation dans le noyau et sa fixation au promoteur de la glyoxalase I, qui maintient le pool de glutathion à l'état réduit, améliore la tolérance de la plante à un stress hydrique (Duan *et al.* 2017). L'expression de SNAC1 et OsNAP est induite par les stress hydrique et salin, dont ils sont des régulateurs positifs de la tolérance chez le riz. Le stress hydrique induit l'expression de SNAC1 dans les cellules de garde plus particulièrement. Leur surexpression chez le riz améliore sa tolérance à ces stress au stade végétatif (par une augmentation de la quantité d'eau et de

chlorophylles contenues dans les feuilles pour SNAC1) comme au stade reproductif où le rendement est moins impacté. OsNAP entraîne l'induction de l'expression de nombreux gènes de réponse dont des facteurs de transcription liés aux stress, comme OsDREB1A, OsMYB2, OsAP37 et OsAP59 (Hu *et al.* 2006 ; Saad *et al.* 2013 ; Chen *et al.* 2014). Jeong *et al.* (2010) ont aussi constaté une amélioration de la tolérance au stade végétatif (*via* un élargissement des racines) et une augmentation du rendement en condition de stress hydrique lors de la surexpression de OsNAC10 dans les racines de riz. En effet, OsNAC10 possède une expression préférentielle dans la racine et la panicule. Chez le blé, TaRNAC1 est aussi un facteur NAC exprimé dans les racines de manière constitutive, mais son expression n'est pas activée en réponse à l'application d'un stress. Néanmoins, lorsqu'il est surexprimé de manière artificielle dans les racines, il conduit à l'amélioration du rendement en condition de stress hydrique en augmentant la biomasse et la longueur racinaire (Chen *et al.* 2018).

TaNAC2a, TaNAC2D, TaNAC4a, TaNAC6, TaNAC7, TaNAC13 et TaNTL5 sont induits par le stress hydrique, une salinité élevée et/ou une température basse (Tang *et al.* 2012 ; Huang et Wang 2016). Ces protéines sont localisées dans le noyau ou liées à la membrane et 5 d'entre elles possèdent une expression organe-spécifique. L'expression du gène ONAC022 chez le riz est également induite par le stress hydrique et une salinité élevée et améliore la tolérance du riz à ces stress selon une voie de régulation liée à l'ABA. Sa surexpression provoque une diminution de la transpiration en réponse au stress hydrique, et une diminution de l'accumulation des ions Na⁺ dans les tiges et les racines en réponse à la salinité (Hong *et al.* 2016). De la même manière, l'expression du gène ONAC095 est activée par l'ABA, mais il semble jouer des rôles opposés dans la tolérance au stress hydrique (régulateur négatif) et au froid (régulateur positif) chez le riz (Huang *et al.* 2016). Chez le blé, les gènes TaNAC67 et TaNAC2 sont des régulateurs positifs de la réponse aux stress hydrique, salin, froid *via* une voie impliquant l'ABA (Mao *et al.* 2012, 2014). La surexpression de TaNAC67 chez *Arabidopsis* induit une meilleure stabilité membranaire, un taux de chlorophylle et un potentiel photosynthétique augmentés ainsi qu'une meilleure capacité de rétention de l'eau.

Chez la banane, des lignées sur-exprimant MusaNAC042 présentent un plus haut niveau de chlorophylle en réponse à un stress hydrique et un stress salin en modulant l'expression de gènes de réponse comme CBF/DREB, LEA et WRKY (Tak *et al.* 2017). En réponse à des stress hydrique et salin, des plants de tomate sous-exprimant SINAC4 présentent une réduction du développement racinaire et aérien, ainsi qu'une perte en eau plus élevée et une diminution du taux de chlorophylle ; le tout lié à une diminution de l'induction de l'expression de gènes de

réponse. L'expression de SINAC4 est induite par le méthyl-jasmonate mais pas par l'ABA (Zhu *et al.* 2014a). Chez la même espèce, l'expression de SINAC35 est induite par les stress hydrique, salin mais aussi par des pathogènes bactériens. Sa surexpression chez le tabac soumis à des stress hydrique et salin stimule son développement et sa croissance racinaire ainsi que la résistance aux pathogènes bactériens (Wang *et al.* 2016b). L'expression de SINAM1 est induite dans la tomate en réponse aux stress hydrique, salin, froid et aux traitements ABA et méthyl-jasmonate. Sa surexpression chez le tabac améliore son taux de germination et son activité photosynthétique en réponse à un stress froid, tandis que son accumulation de ROS diminue (Li *et al.* 2016b).

4.5.1.3. Stress thermique

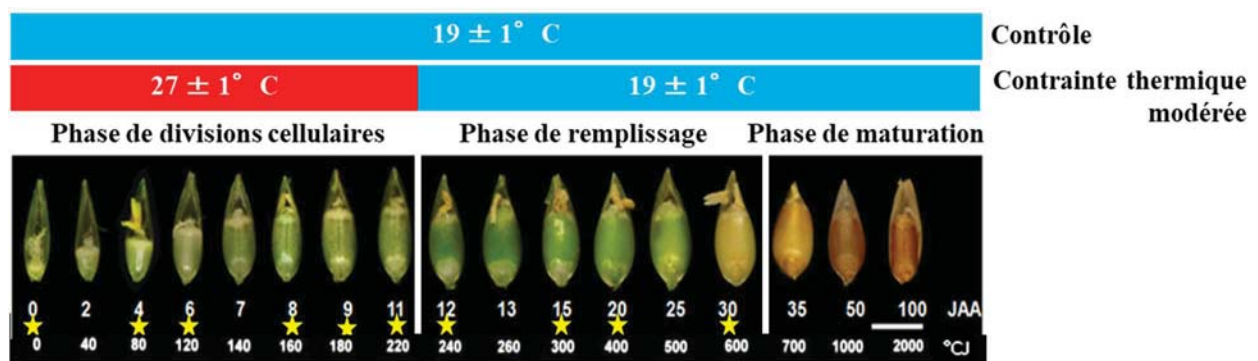
JUB1/ANAC042 (JUnGBrunnen1) est un régulateur négatif de la sénescence et un régulateur positif de la tolérance aux stress thermique, salin et hydrique chez Arabidopsis. L'expression de JUB1 est induite par le stress oxydatif, et sa protéine active l'expression du gène de réponse aux stress DREB2A en se fixant sur son promoteur. AtHB13 est un facteur de transcription induit en réponse au stress hydrique, qui se lie à l'élément CAATAAATG du promoteur de JUB1 et induit son expression chez Arabidopsis (Wu *et al.* 2012 ; Ebrahimian-Motlagh *et al.* 2017). En aval, JUB1 régule l'expression d'autres facteurs de transcription comme DREB1 et DREB2A, et réduit l'accumulation des ROS (Shahnejat-Bushehri *et al.* 2012 ; Wu *et al.* 2012). Chez la tomate SIJUB1 régule aussi positivement la tolérance au stress hydrique en diminuant les pertes en eau et l'accumulation des ROS. De plus, il contrôle aussi l'expression des gènes homologues SIDREB1, SIDREB2 et SIDELLA en se liant directement à leurs promoteurs (Thirumalaikumar *et al.* 2018). ANAC069 est une autre protéine qui régule l'expression de gènes cibles (dont des gènes antioxydants comme la SOD ou la peroxidase) en se liant à la séquence génomique C(A/G)CG(T/G) présente dans leurs promoteurs (He *et al.* 2017). AtNAC019 est un facteur de transcription qui intervient dans la médiation du signal température chez Arabidopsis. Après avoir été déphosphorylée par RCF2, sa protéine se fixe aux promoteurs des gènes HSFA1b, HSFA6b, HSFA7a et HSFC1 (Guan *et al.* 2014). AtNAC019 est un intégrateur du signal forte température qui permet l'activation du maillage des facteurs de transcription HSF et sa surexpression améliore la thermo-tolérance.

L'accumulation des protéines NAC dans la cellule peut être contrôlée de manière post-traductionnelle. TaNAC2L est un gène NAC chez le blé dont l'expression est induite par les fortes températures et dont l'accumulation post-traductionnelle est probablement régulée par

une voie liée au protéasome 26S. Lorsqu'il est surexprimé chez *Arabidopsis*, il mène à un phénotype plus tolérant à la température sans provoquer d'altération phénotypiques visibles en condition contrôle. À l'échelle moléculaire, cette surexpression active l'expression de gènes liés à la température (DREB2A, DREB2B, AtCYP181, RD17, HSP26.5, LEA, AtGolS1, HSP70, AtHsfA3 et RD29A) (Guo *et al.* 2015). Chez le riz, OsNTL3 et OsNTL4 sont 2 facteurs de transcription NAC liés à la membrane, qui interviennent dans la réponse de la plante au choc thermique (6h à 37°C) (Kim *et al.* 2010). NTL4 est aussi lié à la membrane chez *Arabidopsis*, il facilite la production de ROS (notamment H₂O₂) et participe à une boucle de régulation positive modulant la mort cellulaire programmée en réponse à un stress hydrique et un choc thermique (Lee *et al.* 2012, 2014). Chez le maïs, parmi les 7 ZmNTL identifiés, 6 sont associés à la membrane plasmique et le dernier à la membrane du réticulum endoplasmique. En réponse à un stress oxydatif (application d'H₂O₂), l'expression de ces 7 gènes est activée dans les racines et la tige mais réprimée dans les feuilles, probablement selon une voie ABA-dépendante (Wang *et al.* 2016a). L'expression ubiquiste de SNAC3 chez le riz est aussi induite par l'ABA et les stress salin, hydrique et fortes températures. SNAC3 confère une tolérance au choc thermique, aux stress hydrique et oxydatif en modulant l'accumulation des ROS (Fang *et al.* 2015). Chez *Caragana intermedia*, CiNAC3 et CiNAC4 sont 2 gènes induits par plusieurs stress abiotiques (choc thermique, déshydratation, salinité et froid). Lorsqu'elles sont surexprimées chez *Arabidopsis*, ces protéines nucléaires diminuent la sensibilité à l'ABA durant la germination et améliore sa tolérance à la forte salinité. CiNAC4 promeut aussi la formation des racines latérales (Han *et al.* 2015). Morishita *et al.* (2009) ont montré qu'en réponse à un stress de fortes luminosité et température, la surexpression du gène ANAC078 induit une activation de l'expression de plus de 166 gènes chez *Arabidopsis*, dont des facteurs de transcription liés à la voie de biosynthèse des flavonoïdes. ANAC078 interviendrait donc en amont de la voie d'accumulation des flavonoïdes et des anthocyanes en réponse à un cumul de stress lumineux et choc thermique.

5. Description du projet BreedWheat, contextualisation du sujet de thèse

Comme nous l'avons évoqué en introduction de ce mémoire, l'impact du réchauffement climatique sur les cultures est une préoccupation majeure de la communauté scientifique et de certains gouvernements. Dans ce contexte et dans le cadre du programme Investissements d'Avenir, l'Agence Nationale de la Recherche a financé le projet BreedWheat (<http://www.breedwheat.fr/fr/>) sur 9 ans (2011 à 2019). Ce projet rassemble 27 partenaires dont



Capron et al. 2012

Figure 17 : Développement du grain de blé exprimé en jours après anthèse (JAA) et en degrés Celsius jour après anthèse (°Cj) et conditions thermiques appliquées dans le cadre du projet ANR BreedWheat. La partie supérieure de la figure présente le dispositif expérimental utilisé dans le cadre du projet BreedWheat : des plants de blé sont mis en culture à $19^\circ\text{C} \pm 1^\circ\text{C}$ (condition contrôle) et d'autres sont exposés à une contrainte thermique de $27^\circ\text{C} \pm 1^\circ\text{C}$ durant la phase de division cellulaire. Les étoiles correspondent aux stades prélevés pour l'analyse transcriptomique.

15 laboratoires de recherche publique et 10 entreprises privées et s'articule en 7 axes de recherche. Il a pour vocation d'identifier les facteurs génétiques impliqués dans les caractères d'intérêt agronomique tels que le rendement, la qualité et la tolérance aux stress biotiques et abiotiques chez le blé. Son objectif à long terme est d'identifier et de combiner des allèles d'intérêts pour de nouvelles variétés plus performantes, dans des conditions de culture respectueuses de l'environnement et adaptées au changement climatique.

Dans le cadre de ce projet, l'effet d'une contrainte thermique modérée sur le développement du grain de blé tendre a été étudié. Pour cela, 2 lignées de blé ont été soumises à 2 conditions : la condition contrôle à 19°C, qui est la température moyenne optimale de croissance du blé tendre, et la condition stressée. Dans ce cas, les plants de blé subissaient une contrainte thermique de 27°C durant 10 jours pendant la phase de division cellulaire, qui est connue pour être une phase sensible dans la réponse du développement du grain de blé à une contrainte thermique (Figure 17). Les premières analyses éco-physiologiques ont montré un raccourcissement de la phase de division cellulaire conduisant à une diminution de 19,6 % du nombre de cellules dans l'albumen des grains stressés. Elles ont aussi montré un raccourcissement de la phase de remplissage, engendrant une diminution de la taille des cellules des grains stressés (avec une perte de 17,5% de la masse sèche). En parallèle, des analyses transcriptomiques par puce micro-array ont permis de suivre l'expression de 40 000 gènes au cours du développement du grain. Parmi ces 40 000 gènes, 500 gènes présentent un profil d'expression dépendant du stade de développement du grain et de la condition thermique. La majorité de ces 500 gènes codent pour des enzymes impliquées dans l'accumulation et la protection des composés de réserves (accumulation d'amidon, synthèse de protéines de réserve). Cependant, 4 facteurs de transcription de la famille des NAC ont aussi été identifiés comme présentant un profil d'expression modifié par cette augmentation de la température.

6. Objectifs de la thèse

J'ai réalisé cette thèse au sein de l'équipe BIANCA (Biologie Intégrative de l'Adaptation du blé aux Contraintes Abiotiques) de l'UMR 1095 GDEC (Génétique Diversité et Ecophysiologie des Céréales). Nous nous sommes intéressés au développement du grain de blé tendre en réponse à l'application de stress abiotique tels que la sécheresse ou un stress thermique modéré (tout en restant dans des conditions physiologiques). Ces types de stress conduisent à la mise en place de mécanismes d'adaptation aux niveaux cellulaire et moléculaire, dont la compréhension présente un intérêt pour les programmes d'amélioration variétale. Afin de contribuer à l'identification et la description de ces mécanismes, nous avons décidé d'analyser la famille des gènes NAC chez le blé tendre et d'évaluer leurs rôles dans le développement de cette espèce et sa réponse à des stress abiotiques.

Pour cela, nous avons subdivisé le travail en plusieurs parties :

- d'abord, recenser les membres de la famille NAC chez le blé tendre, étudier sa structuration et son expansion,
- ensuite, identifier les gènes TaNAC présentant une expression spécifique au grain de blé ainsi que les gènes répondant aux stress thermique et sécheresse, car ces gènes sont potentiellement d'un grand intérêt pour l'amélioration du rendement sous contraintes environnementales,
- pour finir, réaliser une validation fonctionnelle de 4 NAC exprimés dans le grain et répondant à un stress thermique, identifiés dans le projet BreedWheat. Pour cela, nous avons cherché à générer des plantes sur- et sous-exprimant ces 4 membres de la famille NAC.

MATERIELS & METHODES

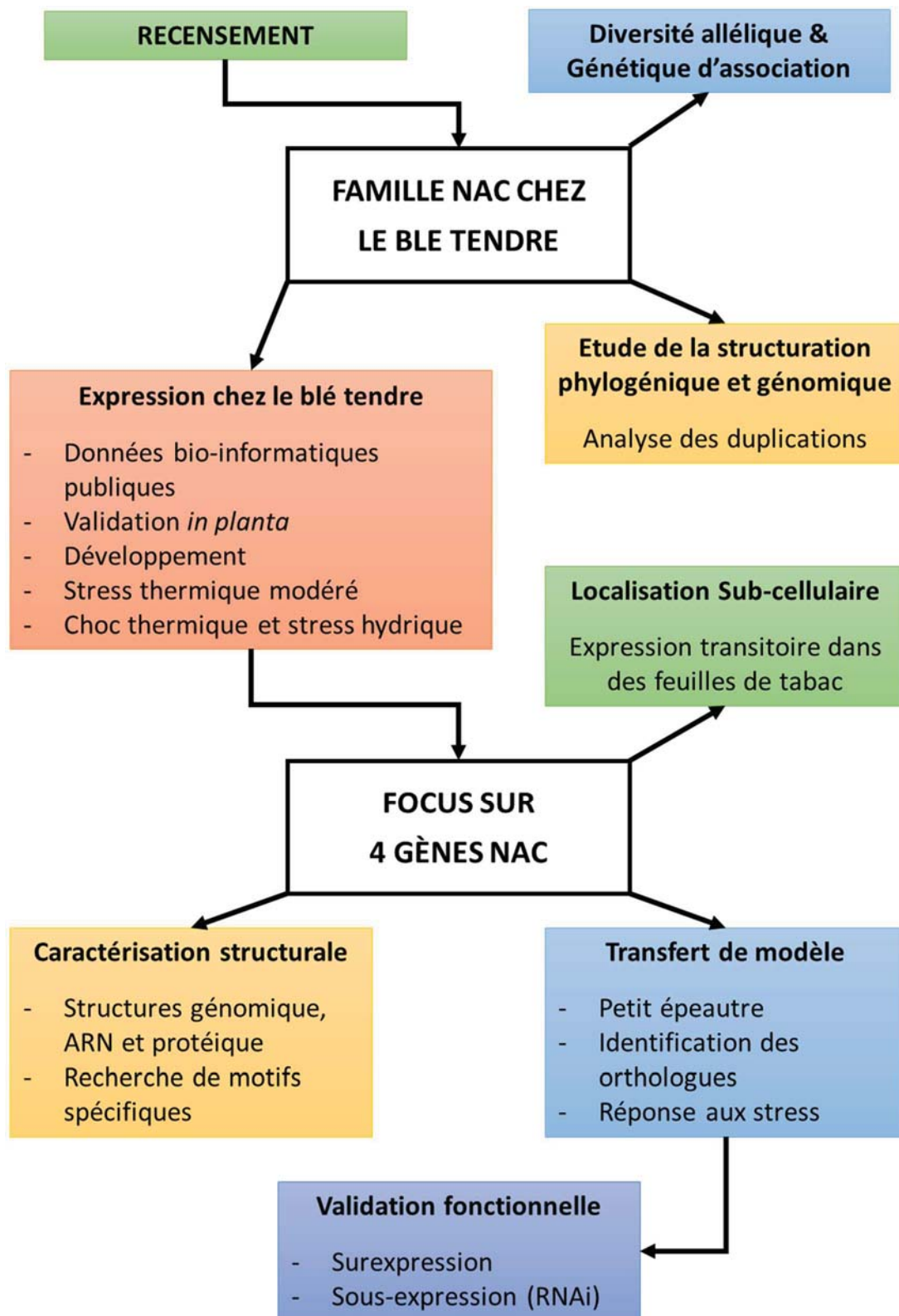


Figure 18 : Schéma bilan général de la thèse. Les 2 premiers volets portent sur la caractérisation de la famille NAC chez le blé tendre (les TaNAC), le recensement et le profil d'expression de ses membres. Le troisième volet caractérise 4 membres de cette famille de manière plus approfondie.

1. Analyse *in silico* de la famille des gènes NAC chez le blé tendre (TaNAC)

1.1. Recensement des membres de la famille TaNAC

1.1.1. Récupération et filtration des données

Le Consortium International de Séquençage du Génome du Blé (International Wheat Genome Sequencing Consortium, IWGSC) a rendu public, le 4 juillet 2018, la dernière version de l'assemblage du génome du blé : IWGS RefSeq v1.0 (<https://www.wheatgenome.org/>). Cette dernière version donne accès à des bases de données annotées des séquences génomiques, codantes (ADNc) et protéiques du cultivar modèle de blé Chinese Spring.

La base de données protéique fournie par l'IWGS a été analysée à l'aide du motif NAM (PF02365, généré par les alignements de 587 core-séquences sur le site Pfam) en utilisant le programme HMMER2 implémenté dans le logiciel Unipro UGENE 1.25 (Okonechnikov *et al.* 2012). Cette analyse a généré une liste de 498 séquences pleine longueur potentielles, contenant le motif NAM avec une e-value seuil de e^{-1} . Leur appartenance à la superfamille des « séquences contenant un domaine NAC » (NAC domain superfamily, IPR036093) a été vérifiée et confirmée en utilisant l'outil InterProScan (<http://www.ebi.ac.uk/interpro/search/sequence-search> ; Jones *et al.* 2014) implémenté dans le logiciel Blast2Go (Götz *et al.* 2008). Cette seconde analyse a abouti à un set final de 488 séquences protéiques entières identifiées comme appartenant à la famille des NAC chez le blé *Triticum aestivum* : les TaNAC (Figure 18).

1.1.2. Constitution d'un set consensus

Une fois le recensement de la famille TaNAC effectué au sein de la base de données IWGS RefSeq v1.0, nous avons voulu croiser ces résultats avec ceux publiés par Borrill *et al.* (2017), qui ont recensé cette famille à partir de la base de données blé tendre d'une autre version de la séquence du génome du blé (TGAC ; Clavijo *et al.* 2017). Pour cela, une recherche d'homologie réciproque a été effectuée entre les bases de données d'ADN codant TGACv1 et IWGS RefSeq v1.0 (incluant les séquences High et Low Confidence), en utilisant le logiciel Usearch (Edgar 2010). Ainsi, la totalité des séquences d'une base de données a été alignée sur l'autre et inversement. Les résultats ont ensuite été filtrés pour ne récupérer que les alignements concernant le set des 488 gènes TaNAC précédemment recensés ; et parmi eux les séquences

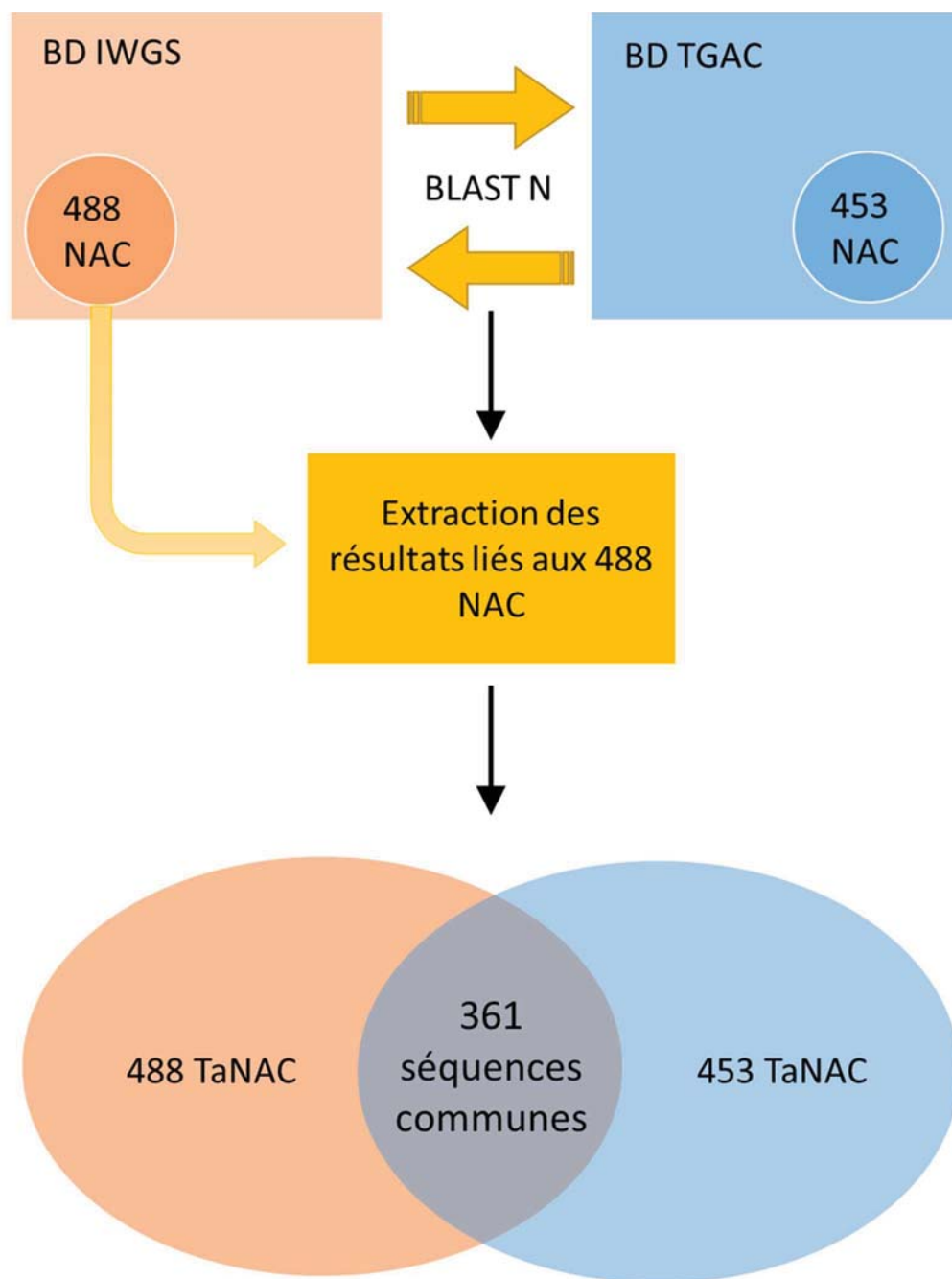


Figure 19 : Schéma de création d'un set de 361 gènes TaNAC consensus entre les 2 bases de données (IWGS et TGAC).

pour lesquelles la recherche d'homologie réciproque a identifié la même paire de gènes TaNAC dans les 2 sens d'alignement. Seules les séquences pleine longueur identiques à 100% entre les 2 bases de données ont été conservées, menant à la création d'un sous-ensemble (appelé set consensus) de 361 séquences (Figure 19). Ces séquences, identiques entre 2 programmes de recherche indépendants ayant utilisés des outils différents pour identifier et annoter les gènes, peuvent être considérées comme de très haut niveau de confiance concernant leur existence et leur séquence réelle. Elles seront donc utilisées dans le cadre de l'analyse phylogénétique de la famille TaNAC mais c'est le set complet de 488 séquences qui sera exploité lors de son étude transcriptomique.

1.2. Structure et organisation de la famille des gènes TaNAC

1.2.1. Séquences génomiques, localisation chromosomique et identification des groupes d'homéologie

Les données de séquençage du génome complet récemment publiées par l'IWGS ont permis de collecter la séquence génomique de tous les membres de la famille TaNAC ainsi que leur découpage exon-intron et leur position sur les chromosomes. Pour cartographier les gènes TaNAC sur la pseudomolécule de blé tendre, les coordonnées chromosomiques de chaque TaNAC ont été extraites et visualisées à l'aide de MapGene2Chromosome V2 (<http://mg2c.iask.in/mg2cv2.0/>). L'identification des groupes d'homéologie a été effectuée selon 2 critères :

- leur localisation chromosomique. Nous avons uniquement pris en compte les groupes de 3 gènes appartenant au même chromosome (n°1 à 7).
- leur apparentement phylogénétique. Seuls les groupes rattachés à un clade ne contenant que 3 gènes issus chacun d'un des 3 génomes ont été pris en compte.

Les clades ne présentant que 2 membres sur les 3 attendus ont été considérés comme groupes d'homéologie incomplets.

1.2.2. Phylogénie et orthologie

Pour étudier la structuration de la famille TaNAC, un arbre phylogénétique a été construit à partir de la totalité des 488 séquences TaNAC. La méthode du maximum de vraisemblance a été utilisée à l'aide du logiciel MEGA7 (Kumar *et al.* 2016) pour construire cet arbre phylogénétique non-raciné basé sur l'alignement des séquences protéiques des TaNAC. L'alignement a été réalisé en utilisant l'algorithme MUSCLE (Edgar 2004). Les paramètres utilisés pour générer l'arbre sont les suivants : test de phylogénie: bootstrap (500 réplicats), traitement des données manquantes : délétion partielle, modèle/méthode : Modèle LG, taux parmi les sites: gamma distribué avec des sites invariants (G).

Afin d'identifier les homologues des 488 séquences TaNAC chez différentes espèces de graminées, les génomes de *Triticum urartu*, du riz et du sorgho ont été récupérés de la base de données EnsemblPlants (<http://plants.ensembl.org/index.html>). Les 488 séquences TaNAC protéiques ont été alignées sur la totalité du génome de ces 3 espèces à l'aide du logiciel BioEdit et le meilleur alignement a été conservé pour chaque séquence. Un seuil de e^{-10} a été appliqué pour considérer 2 séquences comme meilleur alignement.

1.2.3. Analyse des séquences protéiques de la famille TaNAC

La présence de motifs conservés entre les 488 séquences protéiques de la famille TaNAC a été étudiée avec la suite MEME (<http://meme-suite.org/>). Les réglages par défaut ont été utilisés, excepté le nombre de motifs à chercher : 10, et les tailles minimale et maximale du motif : 10 à 150 aa. La présence de peptides signal et de domaines transmembranaires ont été recherchés sur ces mêmes séquences en utilisant le site Phobius (<https://www.ebi.ac.uk/Tools/pfa/phobius/>) avec les réglages par défaut.

1.2.4. Analyse des duplications

Pour l'analyse de duplications, seules les séquences contenues dans le set consensus ont été utilisées. La localisation chromosomique et l'analyse phylogénétique de la famille TaNAC ont suggéré qu'une proportion importante des gènes TaNAC pouvait être apparue suite à des événements de duplication. Afin d'explorer cette hypothèse, nous avons d'abord cherché à obtenir une vue d'ensemble en étudiant la présence de zones dupliquées dans les 3 sous-

génomomes du blé tendre. Nous nous sommes ensuite concentrés sur 2 clades de l'arbre phylogénétique comme exemples au sein de la famille TaNAC. Pour explorer la contribution de la duplication des gènes à l'expansion de la famille NAC chez le blé tendre, nous avons pris en compte le gène et les séquences flanquantes de chaque membre TaNAC contenu dans ces clades. Pour chaque locus contenant un gène TaNAC, 25 kb ont été collectés de part et d'autre du gène. Lorsque plusieurs gènes TaNAC étaient séparés par moins de 25 kb, la séquence extraite a été étendue aux 25 kb au-delà des gènes TaNAC les plus à gauche et les plus à droite de la région considérée. Le degré d'homologie entre séquences nucléotidiques a été obtenu en utilisant YASS, un outil de type Blast (Basic Local Alignment Search Tool ; Noé et Kucherov, 2005) avec une e-value seuil de e^{-30} . Les séquences présentant au moins 80% d'identité sur plus de 2 kb ont été considérées comme dupliquées et retenues pour la suite de l'étude. Les figures montrant ces duplications ont été générées à l'aide de CIRCA (OMGenomics, <http://omgenomics.com/circa/>).

1.3. Méta-analyse de l'expression *in silico* des gènes appartenant à la famille TaNAC

Les bases de données bio-informatiques publiques recensant notamment des données issues d'études RNAseq ou de puce à ADN nous ont permis d'orienter les expérimentations qui ont suivi et d'en corroborer les résultats obtenus.

1.3.1. Sur la plateforme Genevestigator

La plateforme Genevestigator (<https://genevestigator.com/gv/>) répertorie les données d'expression issues de plusieurs centaines d'études utilisant les puces à ADN de type Affymétrie. Chez le blé, ces puces à ADN étaient au nombre de 1800 au moment de la consultation de la plateforme (2015). Un BlastN de la séquence des TaNAC sur la base de données Affymétrie a permis l'identification des sondes correspondant aux gènes et la collecte de leurs données d'expression dans diverses conditions. Les graphiques issus de cette analyse ont été générés par la plateforme.

1.3.2. Sur la plateforme WheatExpression

1.3.2.1. Collecte des données d'expression brutes

La plateforme expVIP www.wheat-expression.com (Borrill *et al.* 2016) héberge les données d'expression de 850 échantillons RNAseq (RiboNucleic Acid sequencing) issus de 32 tissus à différents stades de développement et/ou ayant subi différents stress. Ces données d'expression ont été reliées à des gènes cartographiés sur l'annotation IWGS RefSeq v1.0. Les profils d'expression brute des gènes TaNAC dans différents tissus (racine, feuille, tige, épi et grain ; tirés de l'étude de Choulet *et al.* (2014)) ou en réponse à différents stress abiotiques dans les feuilles (choc thermique et sécheresse à un stade précoce ; tirés de l'étude de Liu *et al.* 2015) ont été extraites de cette plateforme. Dans ces publications, les auteurs avaient identifié les gènes différentiellement exprimés à partir de l'ancienne annotation du génome du blé et selon des critères différents entre les 2 études. Nous avons donc recommencé ce travail de filtration des données pour obtenir un résultat homogène sur la base de l'annotation la plus récente disponible.

1.3.2.2. Traitement statistique des données RNAseq et identification des gènes différentiellement exprimés

Les données d'expression brutes correspondant à l'ensemble du génome du blé ont été téléchargées et filtrées pour éliminer les gènes faiblement exprimés (les gènes sont conservés s'ils sont exprimés à un Log_count per million $> 0,5$ dans au moins 2 échantillons). À l'aide du package R Limma (Linear Models for Microarray data ; Ritchie *et al.* 2015), l'expression des gènes conservés a ensuite été normalisée à l'aide des méthodes Trimmed Mean of M values (TMM) et Voom (Robinson et Oshlack 2010 ; Law *et al.* 2014). Les expressions ont été comparées entre les échantillons et les gènes qui affichaient un Log(ratio) absolu > 1 avec une p-value ajustée $< 0,05$ ont été considérés comme Gènes Différemment Exprimés (DEG). Enfin, les gènes TaNAC ont été extraits de cet ensemble de DEG. Le clustering hiérarchique des TaNAC DEG a été réalisé sur les données d'expression normalisées par Voom, à l'aide du logiciel MeV (Howe *et al.* 2011) avec une distance euclidienne et une liaison moyenne (average linkage).

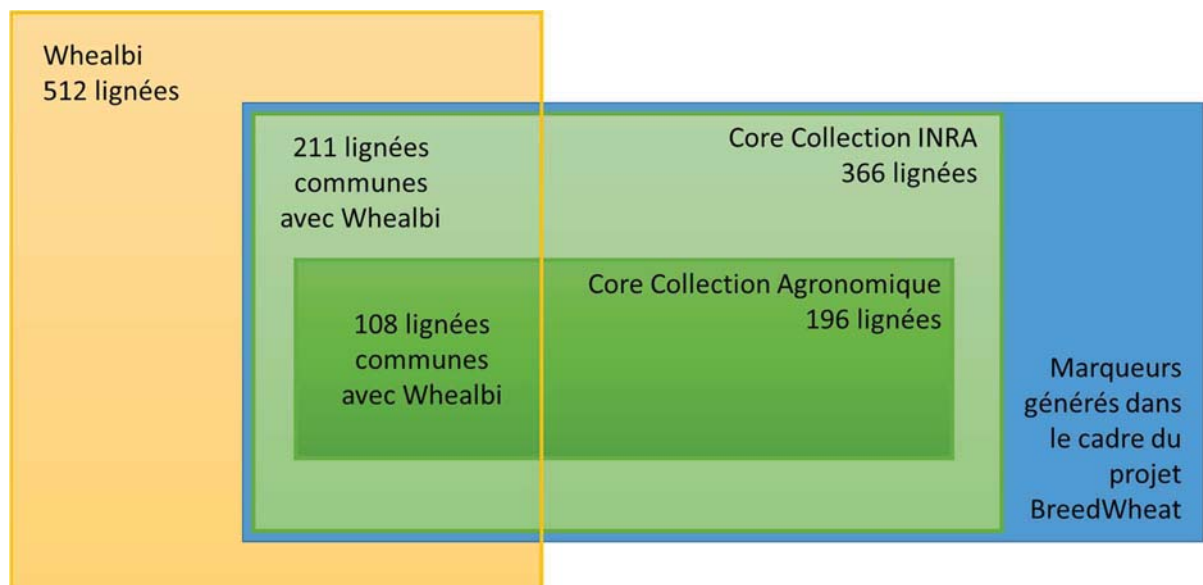


Figure 20 : Structuration du matériel et origine des marqueurs utilisés pour l'étude de la variabilité des membres de la famille TaNAC (bases de données Whealbi et BW_420K, en jaune et bleu respectivement) et la recherche d'association avec les données phénotypiques (Core Collection INRA et Agronomique en vert ; et Whealbi en jaune).

1.4. Evaluation de la diversité allélique au sein de la famille TaNAC et association avec des caractères phénotypiques

Afin d'estimer le polymorphisme présent dans la famille NAC de blé tendre et son éventuel impact sur différents caractères phénotypiques de la plante en général et du grain en particulier, nous avons dans un premier temps réalisé une étude de la diversité allélique présente au sein de la famille TaNAC à partir des données disponibles, puis nous avons étudié les éventuelles associations entre variations alléliques et caractères phénotypiques d'intérêt.

1.4.1. Etude de la diversité allélique de la famille TaNAC

Les marqueurs de polymorphisme ont été recensés à partir de données générées dans le cadre des projets BreedWheat (<https://breedwheat.fr/>) et Whealbi (<https://www.whealbi.eu/fr/>) (Figure 20).

- Le projet investissement d'avenir BreedWheat a permis la création d'une puce de génotypage qui contient 423 385 marqueurs SNP (Single Nucleotide Polymorphism) génomiques génotypés sur environ 7 800 accessions de blé (Rimbert *et al.* 2018), dont les 366 lignées de la Core Collection INRA (Balfourier *et al.* 2007). Par la suite, les données de génotypage issues de cette puce pour ces 366 lignées seront notées « BW_420K ».
- Le projet européen Whealbi a généré des données de polymorphisme par capture d'exome de 512 lignées (454 lignées de blé tendre et 58 lignées d'espèces apparentées). La capture d'exome a consisté en une hybridation de l'ADN génomique total avec des sondes spécifiques des exons, selon la technique décrite par Mascher *et al.* (2013). Environ 50% des gènes existants chez le blé ont été capturés. Leur séquençage a produit un fichier contenant l'ensemble des variations observées sur la totalité de la collection (512 lignées). Seules les 454 lignées de blé tendre ont été utilisées pour la recherche de polymorphisme génétique dans la famille TaNAC. Par la suite, les données de génotypage issues de ce projet seront notées « Whealbi ».

Nous avons pour objectif de rechercher le polymorphisme présent dans les 488 gènes de la famille TaNAC, et dans leurs 5' et 3' UTRs (Untranslated Transcribed Regions) qui pourraient influencer leur niveau d'expression en modifiant par exemple le site de fixation de protéines

régulatrices. Pour cela, nous avons extrait les données de génotypage des fichiers BW_420K et Whealbi en utilisant les coordonnées génomiques sur la pseudomolécule du blé de chaque gène d'intérêt bordés de 1500 pb en amont et en aval ; de sorte à tenir compte des UTR. Les données obtenues ont été filtrées : les marqueurs ayant un allèle rare (dont la fréquence est inférieure à 5% au sein de la base de données considérée) ou trop de données manquantes ont été écartés de l'analyse.

1.4.2. Association entre SNP et caractères phénotypiques

Les données de phénotypage utilisées dans le cadre de cette étude d'association sont d'origines multiples.

- Au sein de l'équipe BIANCA, les mesures de taille du grain (surface, longueur et largeur moyennes) et de PMG ont été réalisées à l'optomachine pour les 512 lignées de Whealbi ainsi que pour les 366 lignées de la core collection INRA (incluant les lignées de la Core Collection Agronomique).
- La dureté et la concentration en protéines des grains ont été mesurées au NIRS sur les 366 lignées de la core collection INRA comme décrit par Bordes *et al.* (2011).
- Après avoir été cultivés dans trois environnements, Clermont-Ferrand et Le Moulon avec une forte fertilisation azotée et Le Moulon avec un faible apport d'azote (Clermont-Ferrand High Nitrogen, Le Moulon High Nitrogen et Le Moulon Low Nitrogen), les 196 lignées de la Core Collection Agronomique ont été phénotypées (Bordes *et al.* (2011) donnent les détails de l'essai agronomique). Le PMG ainsi que la composition protéique fine du grain ont été acquis (Bordes *et al.* 2013 ; Plessis *et al.* 2013).

Les études d'association entre le polymorphisme moléculaire identifié et les phénotypes pour les caractères mesurés sur les différentes collections ont été réalisées avec un modèle comprenant une matrice d'appariement (matrice kinship, K), qui permet d'évaluer la ressemblance génétique entre les individus, disponible dans le package rrBLUP (Endelman 2011) sous R (<http://www.R-project.org> ; R development Core Team, 2008). Ce type de modèle permet de limiter le nombre d'associations faussement positives dues au déséquilibre de liaison de longue portée généré par la structuration (l'apparentement) de la population utilisée.

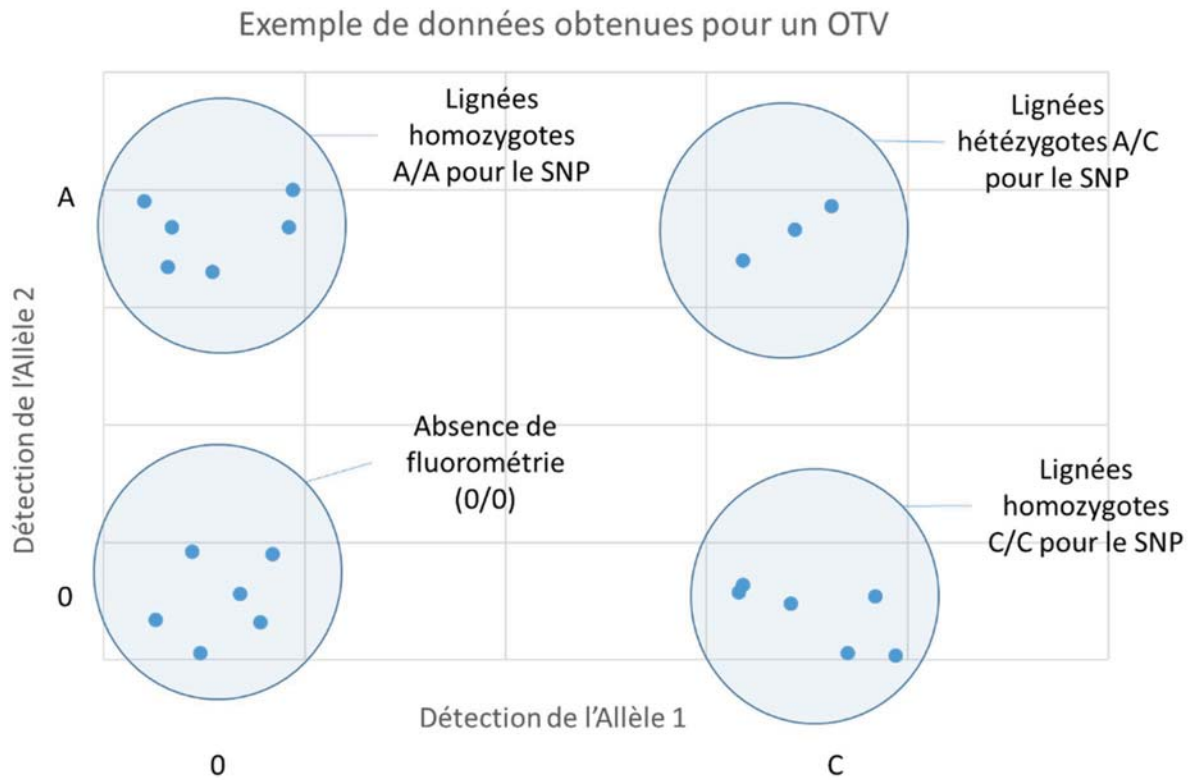


Figure 21 : Exemple de l'existence d'un Out of Target Variant (OTV). Un OTV est un site de variation possédant plus de 2 variations, ici les allèles A ou C ou une absence de signal, notée 0. Cette variation allélique à 3 variables a été recodée en un SNP biallélique (allèle1/allèle2) et un marqueur de type Présence/Absence. L'existence d'un OTV peut être due, par exemple, à la délétion de ses séquences flanquantes dans certaines lignées, les sondes complémentaires à ces séquences ne peuvent donc pas se fixer. Biologiquement, cela peut donc traduire une délétion partielle ou totale du gène ciblé.

Pour augmenter la précision des associations, nous avons utilisé la méthode K-loco pour la recherche d'association sur la Core Collection Agronomique. Pour rechercher les associations avec les marqueurs d'un chromosome donné, cette méthode demande d'estimer une matrice K sans ces derniers marqueurs. Il faut donc calculer une matrice K par chromosome. Dans toutes les autres conditions, c'est une matrice kinship calculée à partir de tous les marqueurs qui a été utilisée. La matrice kinship utilisée avec les données de génotypage issues de Whealbi a été construite par Gilles Charmet, en utilisant la totalité des marqueurs qui n'étaient pas en déséquilibre de liaison sur les 512 lignées de blé tendre. La matrice kinship utilisée avec les données de génotypage issues des Core Collection INRA et Agronomique a été construite par Catherine Ravel à partir de 3449 marqueurs qui ne présentaient pas de déséquilibre de liaison parmi les 420 000 marqueurs de la puce BW_420K.

Pour la recherche d'association à partir des marqueurs issus de la puce BW_420K, la totalité des données de phénotypage obtenue sur les Core Collection INRA et Agronomique a pu être exploitée (366 et 196 lignées respectivement). Cependant, pour la variabilité estimée à partir des données Whealbi, seules les données de phénotypage des 211 lignées de la Core Collection INRA et des 108 lignées de la Core Collection Agronomique ont pu être exploitées. Un score seuil, robuste sans être trop strict, a été fixé à $-\text{Log}(p\text{-value}) = 3$, donc $p\text{-value} = 0,001$.

Parmi les données de polymorphismes disponibles issues de la puce BW_420K, on distingue SNP (Single Nucleotide Polymorphism) et OTV (Out of Target Variant) (Figure 21). Les OTV sont des sites pour lesquels, en plus des 2 allèles, un signal nul a été détecté pour plusieurs lignées, ce sont donc des SNP tri-alléliques (allèle1/allèle2/nul). Ce type de variation peut s'expliquer de différentes façons. Par exemple, le SNP étudié est proche d'un ISBP (Insertion Site-Based Polymorphism) zone très polymorphe, ou il est ancré dans un gène qui est absent dans plusieurs lignées de la collection étudiée et présent dans d'autres. Afin d'étudier le polymorphisme de ces SNP non bi-alléliques, ils ont été recodés en un SNP bi-allélique « allèle1/allèle2 » en considérant l'absence de signal comme une donnée manquante et un marqueur avec un polymorphisme « Présence de SNP / Absence de SNP ».

2. Expérimentations sur blé tendre (*Triticum aestivum*)

Dans le cadre de cette thèse, nous avons étudié des gènes présentant un profil d'expression qui évolue au cours du développement du grain mais aussi en réponse à la température. Estimer le

stade de développement du grain en comptant le nombre de jours post-anthèse induirait donc un biais entre les plantes croissant en température contrôlée et celles en stress thermique puisque ces dernières se développeront plus rapidement. Afin de gommer ce biais, les stades de développement ont été calculés en degrés Celsius-jours après anthèse (°Cj), une unité qui correspond à la somme des températures moyennes journalières cumulées à partir du jour de la première anthèse de l'épi. Cette mesure est facile à calculer et permet d'estimer le stade de développement des grains de blé de manière homogène entre les conditions.

2.1. Stress thermique modéré dans le cadre du projet BreedWheat

Cette thèse est née suite aux résultats obtenus dans le cadre du projet BreedWheat, publié par Girousse *et al.* (2018). Dans le cadre de ce projet, 2 lignées recombinantes quasi-isogéniques ont été utilisées, toutes 2 issues d'un croisement entre 2 variétés élités de blé tendre (Seri M82 et Babax) présentant une réponse contrastée à la sécheresse et un fort potentiel de rendement (Pinto *et al.* 2010). Nous avons travaillé à partir des échantillons obtenus pour l'une de ces 2 lignées recombinantes utilisées dans l'article de Girousse *et al.* (2018) : SerixBabax 49.

Au sein de ce projet, une culture de blé tendre a été réalisée sous contrainte thermique modérée appliquée durant les 10 jours suivant l'anthèse. Les échantillons issus de cette expérimentation ont fait l'objet d'une étude transcriptomique de type microarray et d'un phénotypage qui ont permis, par analyse de réseaux de co-expression, d'identifier des gènes intervenant dans le développement du grain et dont l'expression est impactée par une contrainte thermique modérée.

2.1.1. Dispositif expérimental

Les plants de blé tendre SerixBabax 49 ont été soumis à des conditions de culture naturelles jusqu'à anthèse puis ont été transférés dans des enceintes de culture fermées alimentées par la lumière naturelle. La température journalière moyenne était maintenue à 19°C ($\pm 1^\circ\text{C}$) avec une humidité relative de 83%. Le Maître Brin de chaque plante (premier épi fleuri, noté MB) a été étiqueté et sa date d'anthèse notée.

Tableau 4 : Niveau de stress hydrique observé dans les parcelles irriguées et soumise à un stress sécheresse, à l'anthèse et aux 2 stades de prélèvement. Ce niveau de stress est mesuré à partir de la valeur du facteur de stress hydrique (K_s) calculé comme décrit par Allen et al. (1998). Un K_s inférieur à 1 est considéré comme un indicateur du stress hydrique, dont la gravité augmente lorsque la valeur de K_s diminue.

	Parcelles contrôles		Parcelles stressées	
	K_s	Interprétation	K_s	Interprétation
Anthèse	0,85	Stress très modéré	0,36	Stress sévère
250°Cj après anthèse	0,91	Stress très modéré	0,27	Stress très sévère
450°Cj après anthèse	0,6	Stress modéré	0,2	Stress très sévère

Le jour de leur anthèse, 2 régimes de températures ont été appliqués. Les plantes en condition contrôle ont été conservées à une température basse de 19°C ($\pm 1^\circ\text{C}$) durant tout leur développement tandis que les autres ont été soumises à un stress thermique modéré de 27°C ($\pm 1^\circ\text{C}$) de J+2 à J+13 après anthèse, soit durant 10 jours. Ces 10 jours correspondent chez le blé à la phase de divisions cellulaires du grain (Figure 17).

2.1.2. Etudes transcriptomique, phénotypique et réseau de co-expression

Pour l'analyse transcriptomique, des épis ont été récoltés à 0, 80, 120, 160, 180, 220, 240, 300 et 400°Cj. Les ARN totaux des grains de l'épillet central ont été extraits et une analyse microarray a été réalisée sur ces 9 stades. En parallèle, une étude phénotypique (masse fraîche et sèche, volume, longueur, largeur, épaisseur, comptage du nombre de cellules de l'albumen) a été réalisée sur les grains basaux d'épis à 300°Cj et à maturité. Une analyse pondérée du réseau de co-expression de gènes (appelée aussi WGCNA ou Weighted Gene Coexpression Network Analysis ; Zhang et Horvath 2005) a été réalisée pour constituer des groupes de gènes possédant le même profil de réponse au cours du développement en condition contrôle et au stress thermique. Elle a aussi permis de relier les groupes de gènes avec les caractères phénotypiques observés (cf. (Girousse *et al.* 2018) pour plus de détails).

2.1.3. Obtention de données d'expression pour 4 gènes TaNAC

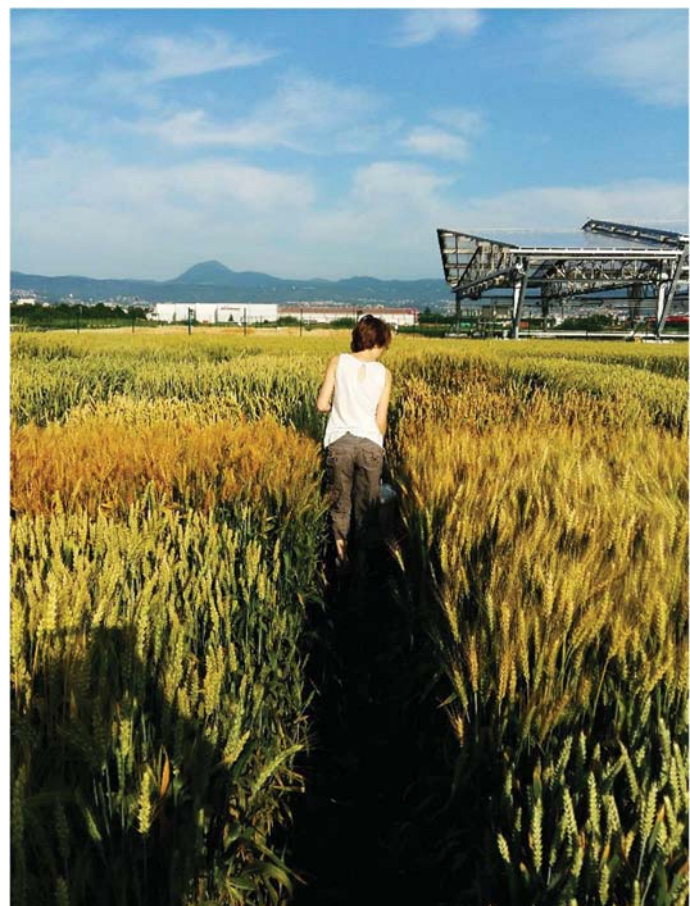
Les résultats portant sur l'effet des températures modérées issus du projet Breedwheat ont conduit à l'identification de 4 facteurs de transcription TaNAC présentant des profils d'expression évoluant durant le développement du grain et en fonction du stress thermique. L'expression des 3 homéologues de ces 4 gènes a été suivie par RT-qPCR (Real-Time quantitative Polymerase Chain Reaction) sur le génotype SerixBabax 49 aux stades 0°Cj, 80°Cj, 120°Cj, 220°Cj (4 stades de la phase division cellulaire) et 300 °Cj (un stade de la phase remplissage du grain). La méthode de suivi de l'expression des gènes est donnée en Matériel & Méthode – Partie 3. Les prélèvements ont été centrés sur la phase de division cellulaire du grain car c'est une période clé qui conditionne ensuite le bon déroulement de la phase de remplissage du grain.



Figure 22 :
Photographie de la parcelle soumise à un stress sécheresse sur la plateforme Phéno3C, le 12 juin 2017. Chaque micro-plot de cette parcelle contient un génotype différent ; à cette période, ils sont approximativement au stade de remplissage du grain. Le stress

hydrique a entraîné la disparition précoce des surfaces photosynthétiques par rapport aux parcelles contrôles (absentes de la photo). De grosses fissures sont ouvertes au sol. Une station météo est visible au milieu de la parcelle. En arrière-plan, un abri est visible. Grâce à des rails, il glissera au-dessus de la parcelle à la moindre pluie.

Figure 23 : *Photographie de la parcelle irriguée sur la plateforme Phéno3C, le 15 juin 2017. Chaque micro-plot de cette parcelle contient un génotype différent ; à cette période, ils sont approximativement au stade de remplissage du grain. L'irrigation mise en place sur cette parcelle a permis aux plants de conserver des surfaces vertes. En arrière-plan, le Puy de Dôme ainsi que l'abri d'une parcelle stressée sont visibles. L'échelle placée au centre de la photo mesure 1m60.*



2.2. Expérimentation sécheresse en plein champ

2.2.1. Génotypes utilisés

L'expérimentation mise en place par Vincent Allard sur la plateforme Phéno3C avait pour objectif global d'étudier la réponse d'un grand nombre de génotypes de blé tendre à la sécheresse en plein champ. Dans ce contexte, nous avons eu effectué des prélèvements sur 2 génotypes : la variété Allez-Y (2011, Limagrain) et la variété Altigo (2007, Limagrain). Ces 2 variétés ont été sélectionnées sur la base de l'impact d'un stress sécheresse sur leur rendement, respectivement en tant que cultivar sensible et tolérant au stress.

2.2.2. Dispositif expérimental (Phéno3C)

L'expérimentation sécheresse en plein champ sur blé tendre s'est déroulée sur la plateforme de phénotypage Phéno3C (INRA de Crouël, France). Cette installation, composée de 4 abris automatiques protégeant le sol de la pluie, permet de générer des régimes d'eau contrastés en protégeant 4 parcelles « stressées » des précipitations naturelles, tandis que 4 parcelles « témoins » sont maintenues sous des conditions météorologiques naturelles et irriguées si nécessaire pour éviter tout stress hydrique. Afin de ne pas impacter sur l'apport de lumière naturelle aux plantes sous les abris, ceux-ci sont placés sur rail et se déplacent au-dessus des parcelles « stressées » uniquement en cas de pluie.

Les parcelles de 22 m x 25 m sont subdivisées en micro-parcelles (1,9 m x 1,2 m). Un génotype est semé par micro-parcelle. Toutes les parcelles ont été semées le 4 novembre 2016 et gérées selon les pratiques agronomiques locales. Le stress a été appliqué à partir du 24 mars 2017 en supprimant toute pluie sur les parcelles stressées par la mise en place des abris automatiques. La teneur en eau du sol a été mesurée par des capteurs placés à différentes profondeurs. La capacité de rétention moyenne du sol et sa capacité en eau disponible sont respectivement de 570 et 280 mm. À partir de ces valeurs, un facteur de stress hydrique (K_s) a été calculé comme décrit par Allen *et al.* (1998) à plusieurs stades clés (Tableau 4). Lors de l'anthèse (20 mai 2018), les parcelles témoins avaient reçu 130 mm de pluie (environ 50% de la capacité de rétention maximale du sol) en plus que les parcelles stressées (Figure 22 et Figure 23).

2.2.3. Prélèvements et suivi de l'expression de 23 gènes TaNAC

Dans le cadre de cette thèse, 3 micro-parcelles par génotype et par traitement (irrigué et sec) ont été échantillonnées aux stades 250°Cj et 450°Cj après anthèse. Ces stades ont été sélectionnés parce qu'ils appartiennent respectivement à la fin de la phase de divisions cellulaires où s'établit la taille potentielle maximale du grain, et à la phase médiane de remplissage du grain en protéines et amidon, qui est une phase importante pour la qualité du grain. À 250°Cj, le déficit en eau des parcelles en condition sécheresse était de 190 mm par rapport aux parcelles contrôle, ce déficit atteignait 210 mm à 450°Cj. Les dates de prélèvement ont été calculées à partir des données de température fournies par les stations météorologiques disposées régulièrement sur les parcelles de culture. Afin d'obtenir des échantillons les plus homogènes possibles, les prélèvements ont été effectués à heure fixe en début de journée. Dans chacune des 3 micro-parcelles (triplicats biologiques), 3 épis et 3 feuilles ont été prélevés au hasard et congelés immédiatement dans de l'azote liquide.

Les composantes de rendement de chaque parcelle ont été évaluées à la fin de la culture :

- le rendement en grains a été mesuré avec une moissonneuse-batteuse expérimentale,
- le nombre d'épis par m² a été déterminé manuellement par comptage. Ce paramètre est déterminé avant anthèse (Slafer *et al.* 2014) lors du tallage de la plante, il s'agit donc d'un bon indicateur de la réponse au stress de la plante avant anthèse.

Sur les échantillons de grain/feuille des génotypes Altigo/Allez-Y à 220°Cj/450°Cj en condition contrôle/sécheresse, l'expression de 23 gènes (identifiés comme présentant une expression différentielle entre organes ou en réponse à un stress abiotique) a été suivie. Ceci a été réalisé par RT-qPCR sur ARN totaux comme décrit en Matériel & Méthode – Partie 3, en concevant des amorces spécifiques aux homéologues étudiés (Annexe I).

2.2.4. Analyse statistique globale des données d'expression en Analyse en Composantes Principales (ACP)

Une analyse en composantes principales a été réalisée à l'aide des données d'expression normalisées des 23 gènes dans tous les échantillons et du package R ggfortify (Tang *et al.* 2016). Pour chaque gène, un test non paramétrique de Kruskal-Wallis (homoscédasticité non respectée) a permis de mettre en évidence les différences entre les moyennes des échantillons.

2.3. Choc thermique

La variété de blé tendre TAM107, publiée par l'Université Texas A & M (Porter *et al.* 1987) a été très utilisée à l'ouest du Kansas entre les années 1980 et 1990 du fait de sa tolérance à la chaleur et la sécheresse (Qin *et al.* 2008a). Liu *et al.* (2015) ont étudié la réponse de ce génotype à un choc thermique au stade plantule. Les données de RNAseq issues de ces expériences sont disponibles publiquement sur la plateforme expVIP.

Afin d'aller plus loin dans la connaissance de l'expression des membres de la famille TaNAC au cours du développement et en réponse aux stress abiotiques, nous avons réalisé une culture de blé tendre (cultivar Chinese Spring) dans les mêmes conditions que celles décrites par Liu *et al.* (2015). Au stade 3 feuilles, 12 plantules ont été placées dans un incubateur réglé à température ambiante (24°C) ou à 40°C pour un choc thermique. La totalité des feuilles des plantules a été prélevée au bout de 1h ou 6h de ces traitements et immédiatement congelée à l'azote liquide. Au cours de cette expérimentation, 2 plantes ont été prélevées à 1h ou 6h en condition contrôle, et 4 plantes l'ont été après 1h ou 6h de choc thermique. Sur ces échantillons, l'expression de 23 gènes a été suivie par RT-qPCR sur ARN totaux, comme décrit en Matériel & Méthode – Partie 3.

3. Méthodes de suivi de l'expression des gènes

3.1. Extraction des ARN totaux

3.1.1. Sur feuille

Après un broyage au mortier et pilon dans de l'azote liquide, les ARN totaux issus des différents prélèvements de feuilles ont été extraits au TRIzol Reagent© selon les instructions du fabricant (ThermoFisher Scientific).

3.1.2. Sur grain

Les grains ont été broyés dans de l'azote liquide au mortier et pilon. Du fait de la très forte concentration en amidon et protéines du grain de blé et de petit épeautre aux stades tardifs de leurs développements, la méthode d'extraction des ARNs totaux au TRIzol n'est pas adaptée.

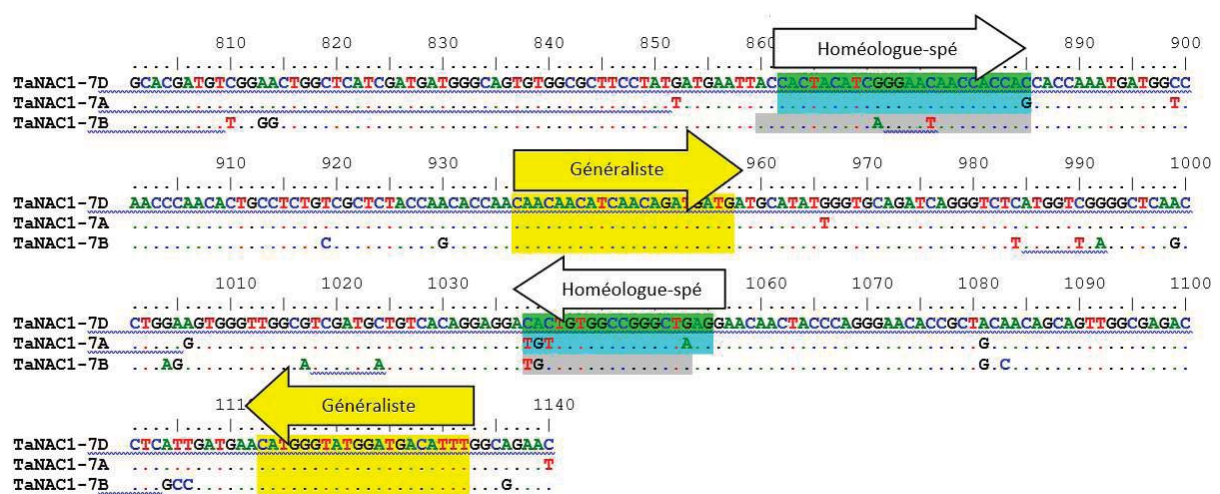


Figure 24 : Dessin des amorces généralistes et homéologue-spécifiques sur les gènes NAC du blé tendre, exemple du gène TaNAC1. Les séquences des 3 homéologues de chaque gène sont alignées à l'aide du logiciel ClustalX. Les amorces généralistes (en jaune) sont dessinées dans des zones conservées entre les 3 séquences, tandis que les amorces homéologue-spécifiques (en blanc) sont dessinées dans des zones polymorphes. La combinaison d'une amorce sens et anti-sens permet la spécificité de l'amplification d'un homéologue donné.

Un protocole spécifique grain adapté de Bogorad *et al.* (1983) a donc été utilisé pour les extractions d'ARN totaux de grain, tout stade confondu, à partir de 250 mg de poudre (Annexe II). Un traitement à la DNase I a ensuite permis d'éliminer les éventuelles contaminations en ADN génomique. Après purification, la concentration et la pureté des ARN totaux ont été évaluées au spectrophotomètre (Nanodrop®), et leur intégrité (non-dégradation) a été vérifiée par électrophorèse sur gel d'agarose 2%. Les échantillons ont ensuite été conservés à -80°C.

3.2. Analyses d'expression

3.2.1. Transcription inverse des ARN en ADN complémentaires

Les ARNs messagers (ARNm) sont d'abord reverse-transcrits en ADN complémentaire simple brin pour pouvoir être amplifiés et détectés par PCR. La synthèse d'ADN complémentaire a été réalisée à partir de 250 ng d'ARN totaux, à l'aide du kit Thermo Scientific Maxima First Strand cDNA Synthesis selon les instructions du fabricant. Ce kit inclut un traitement à la DNase en début de protocole pour dégrader d'éventuelles traces d'ADN génomique. Un traitement par RNase H (Thermo Scientific 18021-071) a été effectué après la transcription inverse afin d'éliminer les ARN résiduels, puis les ADN complémentaires ont ensuite été conservés à -20°C.

3.2.2. Dessin de couples d'amorces gène complet, homéologues-spécifique et généraliste

Des amorces amplifiant de manière spécifique entre 150 et 300 pb des gènes considérés ont été développées pour étudier les profils d'expression des gènes chez le petit épeautre (TM35821) par RT-qPCR. Concernant le blé (génotype SerixBabax 49), des amorces généralistes (c'est-à-dire amplifiant les 3 copies homéologues du gène concerné) et homéologues-spécifiques (n'amplifiant qu'une seule des 3 copies du gène) ont été utilisées (Figure 24).

Toutes ces amorces ont été dessinées à l'aide du logiciel PrimerQuest (<https://eu.idtdna.com/PrimerQuest>). Leur spécificité d'amplification a été vérifiée en amont de la RT-qPCR, par PCR classique puis purification du produit PCR, clonage dans un plasmide, extraction de l'ADN plasmidique et séquençage de l'insert. Plusieurs clones bactériens ont été séquencés pour chaque couple d'amorce, qui a été conservé si tous les clones possédaient

Tableau 5 : Volume réactionnel utilisé pour la RT-qPCR.

Mix qPCR kit Roche	1x
H2O qsp 10µl	1
Amorce sens (10µM)	0.8
Amorce anti-sens (10µM)	0.8
LightCycler® 480 SYBR Green I Master (Roche Diagnostics # 04887352001)	7.5
ADN à 5ng/µl	5
Volume réactionnel	15

Tableau 6 : Programme utilisé pour la RT-qPCR.

Cycle	T°	Durée
Pré-incubation	95°C	10 min
Amplification (45 cycles)	Dénaturation 95°C	10 sec
	Hybridation Tm	15 sec
	Elongation 72°C	15 sec
Courbe de fusion	95°C	15 sec
	62	15 sec
	95	- 0,1°C/sec
Refroidissement	40	∞

l'insert attendu. Le protocole détaillé est décrit dans le Matériel & Méthode – Partie 3.3. Un gradient de température a été utilisé afin de déterminer la température optimale d'amplification de chaque couple d'amorces. Leur spécificité a été vérifiée par la présence d'un seul pic d'amplification sur la courbe de fusion lors de la RT-qPCR. Pour finir, l'efficacité d'amplification de chaque couple d'amorces a été estimée grâce à une amplification RT-qPCR sur des ADNc issus d'une gamme de dilution au demi réalisée sur 6 points. Seuls les couples d'amorces présentant une efficacité comprise entre 90 et 110% ont été conservés.

3.2.3. PCR

Les réactions d'amplification par Réaction de Polymérisation en Chaîne (PCR) ont été réalisées sur ADNc selon le protocole Advantage®2PCR Enzyme System (Clontech Laboratories, Inc). Les conditions d'amplification, de température d'hybridation et de temps d'élongation ont été optimisées pour chaque amplification.

3.2.4. PCR quantitative en temps réel (RT-qPCR)

Les réactions de PCR quantitative en temps réel (RT-qPCR) ont été réalisées sur la plateforme de génotypage à haut débit GENTYANE (UMR 1095 GDEC, <http://gentyane.clermont.inra.fr>). La distribution du mix de réaction (Tableau 5) dans les plaques qPCR 384 puits a été effectuée à l'aide du robot Microlab Hamilton Star, puis la réaction de RT-qPCR a été réalisée dans un thermocycleur LightCycler Roche LC480 selon le programme donné dans le Tableau 6. Toutes les réactions ont été réalisées en duplicata.

Trois gènes de ménage ont été testés (Ta2776, Ta2291 et Ta54227 tirés de la publication de Paolacci *et al.* (2009)). Parmi eux, le gène Ta2776 (protéine analogue à un inhibiteur de RNase L) a été choisi pour normaliser les données de RT-qPCR entre plaques car son niveau moyen d'expression était similaire à ceux des gènes testés et qu'il n'est pas impacté par l'application d'un stress abiotique (Paolacci *et al.* 2009). L'expression relative des gènes d'intérêt a été déterminée selon la méthode $2^{-\Delta\Delta Ct}$ (Livak et Schmittgen 2001).

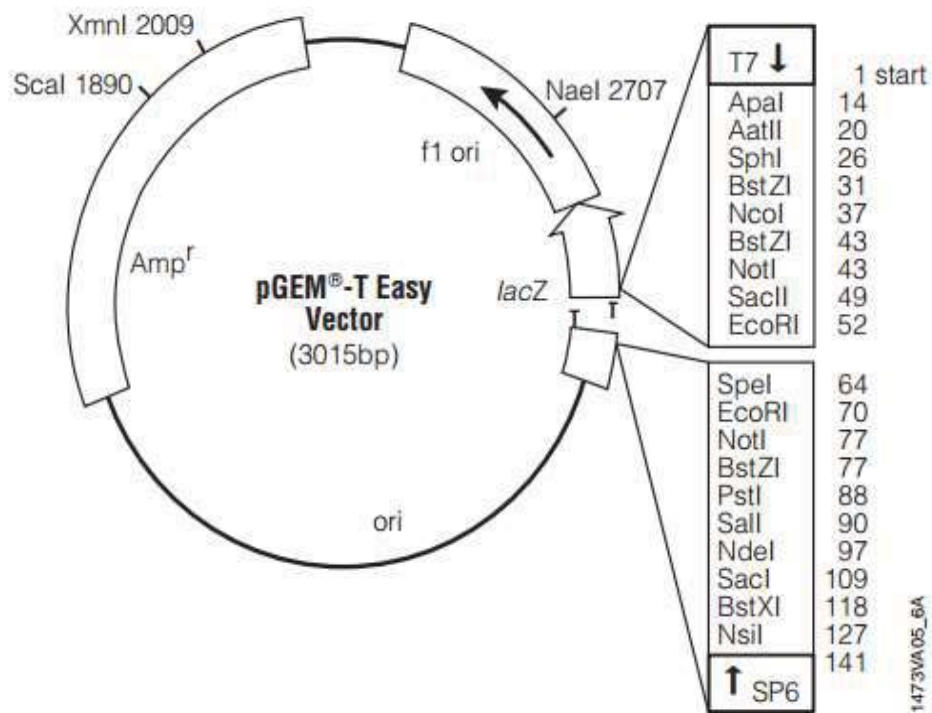


Figure 25 : Le vecteur pGEM-T Easy (Promega). Ce plasmide linéarisé de 3015 pb, permet l'insertion d'un amplicon par TA-cloning grâce aux thymidines ajoutées à ses extrémités 3'. Il porte un gène de résistance à l'ampicilline et 2 sites *EcoRI* qui flanquent le site d'insertion et permettent l'excision de l'insert. Cet insert est positionné à l'intérieur du gène rapporteur *lacZ*, ce qui permet une sélection rapide des clones positifs qui apparaissent blancs.

Enfin, des amorces permettant l'amplification d'une séquence intronique du gène Protein Disulfide Isomerase (PDI) chez le blé (Paolacci *et al.* 2009) ont été utilisées afin de vérifier l'absence de contamination par de l'ADN génomique.

3.3. Vérification des produits d'amplification PCR par séquençage

Afin d'obtenir la séquence des différents gènes étudiés chez le blé (génotype SerixBabax 49) et le petit épeautre (TM35821), des amorces gène-complet ont été dessinées sur la séquence théorique de ces gènes, puis utilisées pour l'amplifier par PCR. L'amorce sens commençait donc par l'ATG de la séquence et l'amorce anti-sens finissait par le codon stop. Les produits d'amplification PCR ont été purifiés et clonés dans un plasmide utilisé pour transformer des bactéries compétentes ; qui ont ensuite été mises à croître sur milieu sélectif puis repiquées en culture liquide. Pour finir, l'ADN plasmidique en a été extrait et l'insert a été séquencé.

3.3.1. Purification des produits d'amplification

Les fragments d'ADN obtenus après amplification par PCR ont été séparés par électrophorèse sur gel d'agarose et visualisés grâce à la fluorescence émise par le BET (bromure d'éthidium) contenu dans le gel. La bande à la taille attendue a été purifiée par le kit Nucleospin extract II (Macherey-Nagel) puis l'ADN a été élué dans 30 µl de tampon d'élution. La concentration d'ADN plasmidique purifié a ensuite été mesurée en déposant 3 µl de l'échantillon sur gel d'agarose grâce à une échelle de poids moléculaire quantitative (DNA ladder 1 kb, Euromedex).

3.3.2. Sous-clonage dans le plasmide pGEM-T Easy : TA cloning

Le plasmide pGEM-T Easy, commercialisé par la société Promega, est un vecteur linéarisé de 3015 pb, utilisé pour le clonage de produits PCR. Ce plasmide haute-copie présente une origine de répllication dans un système bactérien, un gène de résistance à l'ampicilline (100µg/ml) et un site de clonage multiple flanqué par les promoteurs de l'ARN polymérase SP6 et T7 (Figure 25). Ce site de clonage multiple est situé dans le gène LacZ, ce qui permet un crible blanc/bleu des transformants. En effet, le gène LacZ code pour l'enzyme β -galactosidase capable de cliver

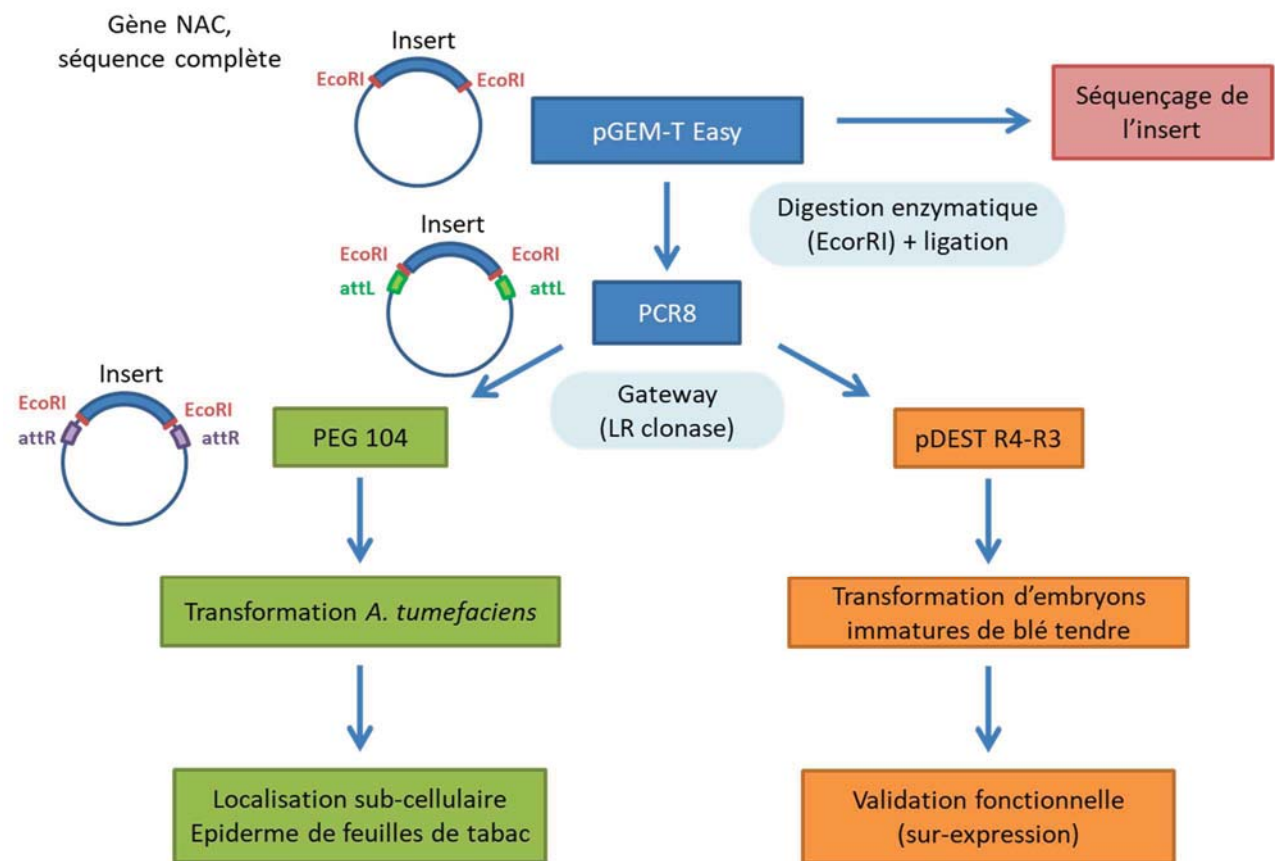


Figure 26 : Schéma global des clonages successifs de la séquence complète dans des plasmides variés pour différentes utilisations. L'étude de la localisation subcellulaire et la validation fonctionnelle de 4 gènes TaNAC ont nécessité un clonage par la technologie Gateway.

le substrat chromogène X-gal présent dans le milieu, ce qui produit une coloration bleue de la colonie. L'inactivation de la β -galactosidase par clonage TA d'un produit PCR au milieu de ce gène produit une coloration blanche de la colonie bactérienne transformée. Le clonage des produits PCR purifiés dans ce plasmide a été réalisé par la T4 DNA ligase, à l'aide du kit pGEM®-T Vector System II (Promega), selon les recommandations du fournisseur. La Figure 26 récapitule les clonages réalisés dans les différents plasmides et leur objectif.

3.3.3. Transformation de la souche bactérienne *E. coli* DH5 α compétente, vérification de la présence de l'insert/Criblage des colonies

La souche bactérienne d'*Escherichia coli* utilisée pour les clonages est la souche DH5 α Chemically Competent (Invitrogen). Cette souche présente une efficacité de transformation importante, amplifie fortement les plasmides et permet un criblage visuel des transformants par ajout du substrat de la β -galactosidase (XGal), ainsi que de l'IPTG.

Cinq μ l de produit de ligation ont été transférés dans 100 μ L d'une suspension de bactéries compétentes. Le mélange a été laissé 30 min sur glace, un choc thermique a ensuite été réalisé 30 sec à 42°C, puis le mélange a été mis à reposer 2 min dans la glace. Cinq cent μ L de milieu SOC (Super Optimal broth with Catabolite repression ; 2 g Bacto®-tryptone, 0,5 g Bacto®-yeast extract, 1 ml NaCl 1M, 0,25 ml KCl 1M, 1 ml Mg²⁺ 2M filtré et stérilisé, 1 ml glucose 2M filtré et stérilisé, H₂O qsp 100 ml, pH7) ont été ajoutés à la suspension, qui a ensuite été placée 1h à 37°C sous agitation (250 rotations par minutes). Les bactéries transformées ont ensuite été étalées sur boîte de Petri contenant du milieu Luria-Bertani (ou milieu LB ; 10g de Bacto-tryptone, 5g de Bacto-yeast extract, 10g de NaCl, H₂O Qsp 1L) supplémenté en antibiotiques adaptés pour permettre la sélection des colonies transformées. Les boîtes ont été placées une nuit dans l'obscurité à 37°C, puis les colonies présentant une coloration blanche ont été repiquées dans 5 ml de milieu LB liquide supplémenté en antibiotique. Après une nuit d'incubation à 37°C sous agitation, 10 μ l de chaque culture bactérienne liquide a été centrifugé 30 sec à 14000 g. Le culot a été ensuite suspendu dans 10 μ l d'eau et chauffé 5 min à 95°C afin de lyser les bactéries. La présence de l'insert a été vérifiée par PCR sur ce lysat bactérien, avec des amorces spécifiques à l'insert et selon les mêmes conditions que celles décrites en Matériel & Méthode – Partie 3.2.3.

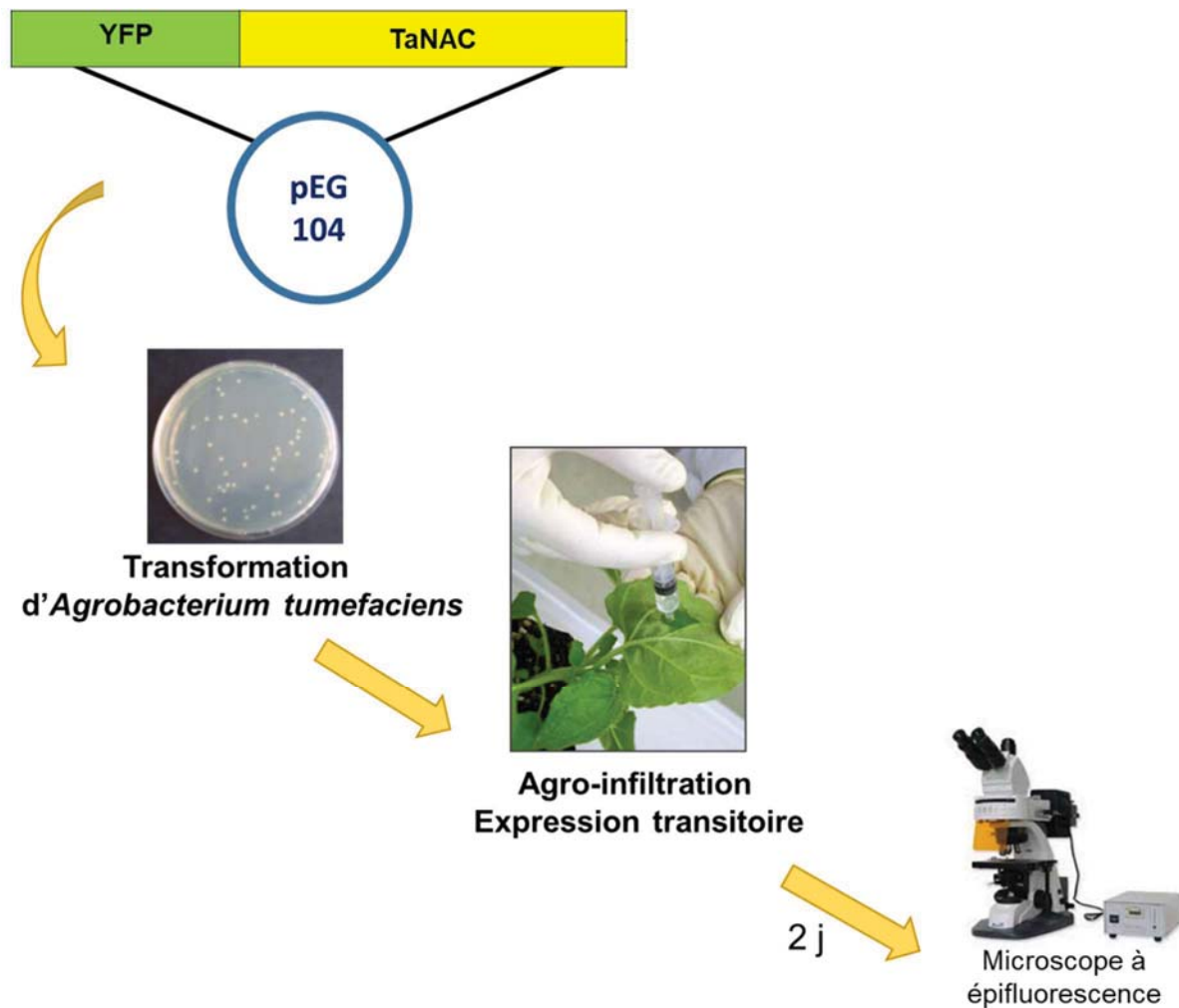


Figure 27 : Synthèse du protocole expérimental de la transformation de l'épiderme inférieur de feuille de tabac pour l'étude de l'expression transitoire d'une protéine TaNAC couplée à la protéine fluorescente YFP. Cette expérimentation permet d'identifier la localisation subcellulaire de la protéine TaNAC in planta.

3.3.4. Extraction d'ADN plasmidique par lyse alcaline (mini prep)

L'ADN plasmidique a été extrait à partir de cultures bactériennes liquides par la méthode de lyse alcaline (Sambrook *et al.* 2001), qui a pour avantage d'éliminer l'ADN du chromosome bactérien (Protocole détaillé en Annexe III). Une fois extrait, l'ADN plasmidique a été dosé au spectrophotomètre (Nanodrop®) et sur gel d'agarose 1% par évaluation visuelle de l'intensité des bandes en comparaison avec une échelle quantitative (1 kb DNA Ladder).

3.3.5. Séquençage

La séquence exacte de l'insert a été obtenue par séquençage de type Sanger sur les ADN plasmidiques extraits par lyse alcaline. Ce séquençage a été réalisé par l'entreprise Beckman Coulter Genomics (Takeley, Royaume Uni), sur un minimum de 3 clones différents pour chaque couple d'amorce.

4. Transformation transitoire de feuilles de tabac

Cette expérimentation a pour but d'étudier la localisation des protéines TaNAC au sein de la cellule végétale. La Figure 27 en synthétise les grandes étapes.

4.1. Clonage du gène dans le vecteur pEarleyGate104 (pEG104)

4.1.1. Le plasmide pEG104

Les séquences entières des gènes d'intérêt ont été clonées par une ligation Gateway de type LR dans un vecteur de destination, ici le vecteur binaire pEarleyGate 104 (pEG104). Ce plasmide, décrit par Earley *et al.* (2006), est utilisé pour la transformation des cellules végétales par le système *Agrobacterium tumefaciens*. Son site de clonage de type Gateway permet l'insertion du gène d'intérêt en position C-terminale de la protéine fluorescente YFP (Yellow Fluorescent Protein). La séquence de la protéine de fusion YFP + ADNc ainsi créée est sous le contrôle du promoteur 35S du virus de la mosaïque du chou-fleur. Cette cassette d'expression est bordée par les régions de transfert d'ADN-T, reconnues par le plasmide Ti d'*A. tumefaciens* qui permet

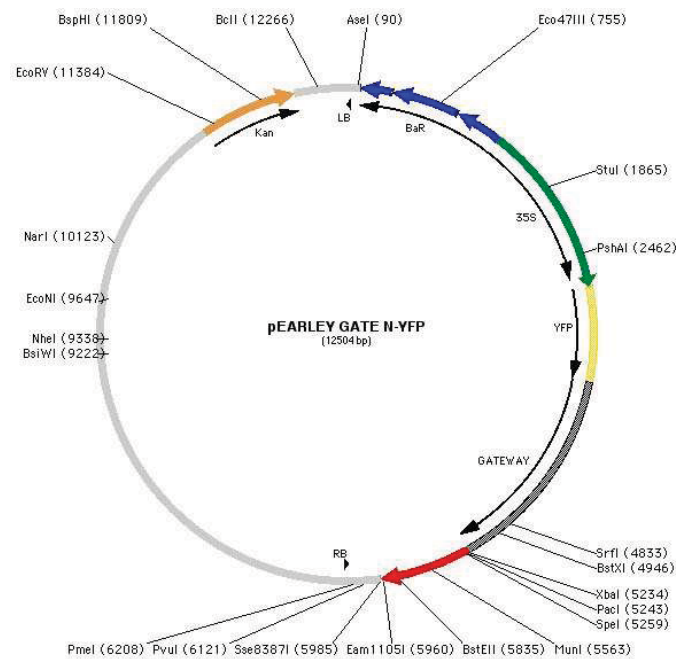


Figure 28: Le vecteur binaire pEG104 (pEarley Gate 104). Ce plasmide permet de produire une protéine chimère constituée d'un fluorochrome accolé à la partie N-terminale de la protéine étudiée, sous le contrôle du promoteur 35S. Cette cassette d'expression est encadrée par des zones de bordure (LB à gauche et RB à droite) qui permettent son intégration dans le génome des cellules de tabac. Le vecteur possède aussi un site de clonage de type Gateway, un gène de résistance à la kanamycine, et 2 origines de répllication, l'une pour *E. coli* et l'autre pour *A. tumefaciens*.

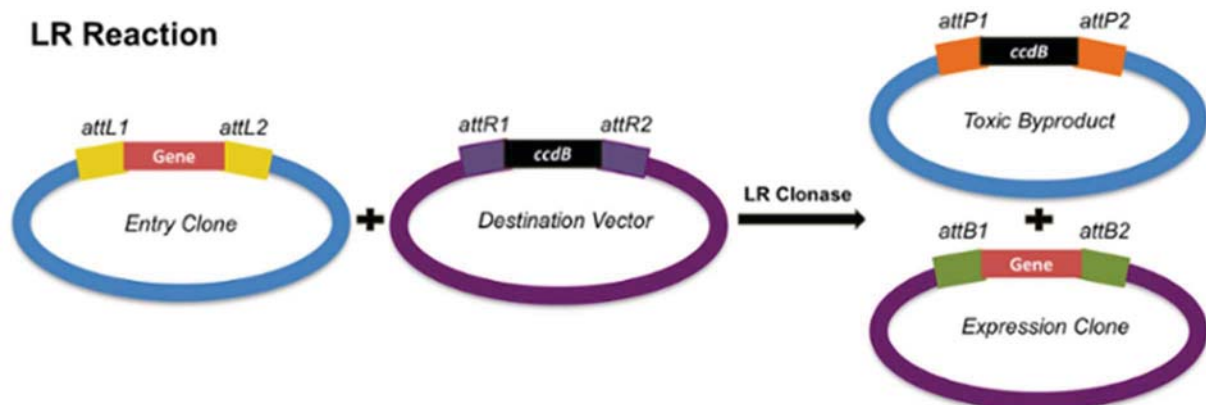


Figure 29 : Principe du clonage d'un gène d'intérêt par recombinaison homologue (Technologie Gateway, Invitrogen). Le clonage de la séquence dans un vecteur de destination contenant les sites attL1 et attL2 par une recombinaison LR permet l'obtention d'un clone d'expression. Le gène *ccdB* est toxique pour les bactéries *E. coli* et permet de sélectionner uniquement les colonies bactériennes ayant intégré un vecteur recombinant.

son excision du plasmide. La région est bornée par les bordures LB (Left Border) et RB (Right Border) responsables de l'intégration d'ADN exogène chez les plantes. Ce plasmide comporte aussi 2 origines de réplication, l'une fonctionnelle dans *E. coli* et l'autre dans *A. tumefaciens*, et un gène de résistance à la kanamycine (Figure 28).

4.1.2. Clonage de type Gateway

Le clonage de type Gateway (Figure 29) est un clonage par recombinaison homologue entre le vecteur donneur et le vecteur de destination. La ligation Gateway de type LR est réalisée à l'aide d'une enzyme LR Clonase (Invitrogen) selon les recommandations du fournisseur. Elle permet une recombinaison entre des sites attL1 et attL2 d'un clone d'entrée et attR1 et attR2 d'un vecteur de destination, permettant l'échange de 2 inserts (la cassette d'intérêt vs le gène létal ccdB). À la fin de la ligation, le vecteur d'expression contient des sites attB et le gène d'intérêt qui est prêt à être traduit.

4.2. Transformation d'*Agrobacterium tumefaciens*

4.2.1. Souche d'*A. tumefaciens* utilisée

La souche d'*A. tumefaciens* GV3101 (Flückiger *et al.* 2003) a été utilisée pour la transformation transitoire de feuilles de tabac et l'analyse de la localisation *in planta* de protéines chimères fusionnées au chromophore YFP. Cette souche porte un plasmide Ti qui code pour les gènes de virulence permettant le transfert et l'intégration de l'ADN-T dans le génome végétal, et un gène de résistance à la gentamycine.

4.2.2. Mise en compétence et transformation de la souche

Une pré-culture d'*A. tumefaciens* est diluée dans 150 ml de milieu LB. Après une incubation de 3-4 heures en agitation à 28°C (DO600 = 0,5), les bactéries en phase exponentielle de croissance sont centrifugées 20 min à 3000g et 4°C. Le culot bactérien est ensuite lavé dans 10 ml de tampon TE froid (Tris pH8 100 mM ; EDTA 1 mM) puis ressuspendu dans 20 ml de

milieu LB. Des aliquots de 500 µl sont utilisés directement pour la transformation ou congelés dans l'azote liquide puis stockés à -80°C.

La méthode de transformation utilisée est une adaptation du protocole décrit par Höfgen et Willmitzer (1988). Pour la transformation, 200 µl d'agrobactéries compétentes sont mélangées avec 1 à 5 µg d'ADN plasmidique, puis le choc thermique suivant est réalisé : 5 min dans la glace, 5 min dans l'azote liquide et 5 min à 37°C. La suspension bactérienne est ensuite diluée dans 1 ml de milieu LB puis placée à 28°C sous agitation (180 rotations par minute) durant 3-4 heures. Un aliquot de 200 µl est alors étalé sur une boîte de milieu LB gélosé complétement en kanamycine (100 µg/ml, dont la résistance est introduite par le vecteur pEG104) et en gentamycine (25 µg/ml, dont la résistance est portée par le plasmide Ti présent dans les souches GV301). Après 2 jours d'incubation à 28°C, des colonies isolées sont prélevées sur la boîte de Petri et mises en culture liquide pour la transformation de cellules végétales.

4.3. Infiltration de l'épiderme inférieur de feuilles de tabac (*Nicotiana tabacum*)

Afin d'étudier l'expression transitoire de protéines chimères et déterminer la localisation subcellulaire de ces protéines, des plants de tabac *Nicotiana tabacum* cv. Samsun-NN ont été cultivés en chambre de culture pendant 4-5 semaines à 24°C (16h de lumière / 8h d'obscurité) pour être transformés par agro-infiltration au stade 5 feuilles.

L'infiltration par l'agrobactérie GV3101 s'effectue sur l'épiderme inférieur de feuilles de tabac. Les colonies d'agrobactéries transformées sont mises en culture dans du LB à 28°C pendant 16h, puis 1 ml de culture est centrifugé 3 min à 5500 g. Le culot est lavé 2 fois dans 1 ml d'eau et centrifugé à 5500 g pendant 20 min. Les agrobactéries sont ensuite reprises dans de l'eau pour une densité optique comprise entre 0,5 et 0,8 à 600 nm. La suspension est laissée 3h à température ambiante, puis infiltrée dans les feuilles de tabac par simple pression d'une seringue de 1 ml sans aiguille sur l'épiderme inférieur légèrement griffé des feuilles.

4.4. Observations en microscopie à épifluorescence

Après 2-3 jours, la fluorescence est observée dans les feuilles exprimant les protéines chimères en utilisant un microscope optique à épifluorescence (Carl Zeiss Axioplan 2 imaging



Figure 30 : Cinétique de développement du grain de petit épeautre. Comme pour le blé, le développement du grain de petit épeautre peut être divisé en 3 phases : division cellulaire, remplissage des cellules en réserves et maturation-dessiccation du grain. Les chiffres indiquent les stades de développement, exprimés en °Cj (degrés Celsius jour). Les étoiles représentent les 8 stades de prélèvement (0, 40, 80, 120, 180, 220, 300 et 700°Cj après anthèse) réalisés pour le suivi de l'expression des gènes *TmNAC* (photos : D. Marcon).

fluorescence microscope). Celui-ci est équipé d'un système d'acquisition (caméra AxiocamMR) couplé au logiciel de traitement d'images *Axiovision*. Une lampe UV et un bloc filtre permettent la détection des fluorochromes. Le filtre FITC permet de visualiser l'YFP qui a un pic d'excitation à 514 nm et un pic d'émission à 527 nm.

5. Expérimentations sur petit épeautre (*Triticum monococcum*)

5.1. Génotype utilisé

Le blé est une plante d'intérêt agronomique mais son étude dans la cadre de la recherche fondamentale n'est pas chose aisée pour plusieurs raisons : son cycle de développement est long (environ 6 mois) et la complexité de son génome (constitué de 3 sous-génomes) oblige, lors de l'étude d'un gène, à devoir prendre en compte ses 3 homéologues qui peuvent présenter de légères différences de séquences et d'expression et rendre les résultats difficiles à analyser. Du fait de ces particularités, nous avons donc choisi de réaliser la transgénèse non pas sur des plants de blé tendre mais de petit épeautre ou *Triticum monococcum*, aussi appelé engrain. Il possède un génome diploïde AA, très similaire au sous-génome AA du blé. De plus, son cycle de développement est plus court (environ 4-5 mois). Nous avons travaillé avec le génotype TM35821, qui est une variété ancienne de petit épeautre de printemps fournie par le CRA-SCV. Il a été sélectionné parmi une vingtaine de variétés anciennes et de lignées pour usage scientifique au sein de l'équipe BIANCA, parce qu'il était facile à battre et qu'il montrait les meilleurs taux de régénération et de transformation après transgénèse.

Afin de pouvoir transférer notre étude du blé tendre au petit épeautre, le développement et la réponse au stress thermique de cette espèce ainsi que le profil d'expression de ses NAC homologues ont été observés. Les NAC étudiés chez le petit épeautre *Triticum monococcum* sont notés les TmNAC.

5.2. Cinétique de développement du grain

Une première culture de petit épeautre a permis de multiplier le lot de grains fourni par l'INRA de Crouël et de réaliser une cinétique du développement du grain de petit épeautre (Figure 30)



Figure 31 : Photo d'un épi de petit épeautre et numérotation des 5 premiers épillets sur l'épi.

Les jeunes fleurs sont d'un jaune vif tandis que les fleurs sorties quelques jours plus tôt sont blanchies. Le jour de l'anthèse, le numéro des épillets fleuris correspondant à leur position sur l'épi, est relevé. Pour cela, le comptage s'effectue à partir de l'épillet basal et continue en montant en zigzag le long du rachis. Ici, l'exemple est donné pour les 5 premiers épillets.

Tableau 7 : Liste des 40 stades pour lesquels un épi a été prélevé, décortiqué, le nombre d'épillets compté et dont les grains ont été pesés et photographiés. Les temps en jours après anthèse (JAA) et en degré-jour après anthèse (°Cj) sont linéaires puisque les températures étaient contrôlées et uniformes tout au long de la culture.

JAA	1	2	3	4	5	6	7	8	9	10
°Cj	21	42	63	84	105	126	147	168	189	210
JAA	11	12	13	14	15	16	17	19	20	21
°Cj	231	252	273	294	315	336	357	399	420	441
JAA	22	23	24	25	26	27	28	29	30	34
°Cj	462	483	504	525	546	567	588	609	630	714
JAA	36	38	40	42	44	48	50	54	58	61
°Cj	756	798	840	882	924	1008	1050	1134	1218	1281

qui a ensuite été mis en parallèle avec la cinétique du grain de blé réalisée au sein du laboratoire et publiée par Capron *et al.* (2012).

5.2.1. Dispositif expérimental

La culture s'est déroulée dans une serre de confinement S2, sur le campus des Cézeaux (Clermont-Ferrand, France). Le 17 novembre 2015, 50 grains ont été semés dans un terreau spécial semis et placés en serre à 16 h de jour à 22°C et 30 000 lux, 8h de nuit à 19°C et une hygrométrie constante de 55%. Le pourcentage de germination a été de 86%. Les plantes au stade 3 feuilles ont été placées en armoire de vernalisation à 4°C pendant un mois puis rempotées en pots individuels de 3 litres et replacées dans des conditions de culture identiques à celles précédant leur vernalisation.

5.2.2. Notation de l'anthèse

Au sein de chaque plante, le développement des épis n'est pas équivalent. Nous n'avons effectué un suivi que pour le 1^{er} et le 2^{ème} épis fleuris par plante, nommés respectivement Maître Brin (MB) et Talle 1 (T1), et étiquetés afin d'être aisément identifiables.

De même que pour le blé, il n'est pas possible de déterminer le moment exact de la fécondation de la fleur de petit épeautre par le pollen mais cette espèce étant autogame (c'est-à-dire qu'elle s'autoféconde, comme le blé), il est admis que le moment de l'anthèse coïncide avec celui de la fécondation. La date d'anthèse est donc l'élément central de la détermination du stade de développement du grain, et la surveillance de l'anthèse a été effectuée chaque matin. L'anthèse des épillets au sein de l'épi s'étale sur environ 2-3 jours, le jour de la sortie de la première fleur est noté comme le jour de l'anthèse de l'épi entier. Le jour de l'anthèse, la fleur d'une couleur jaune vif devient apparente et pend de l'épillet. Elle blanchit dans les jours suivants, pour ensuite sécher et tomber (Figure 31). L'épi est alors étiqueté MB ou T1, le numéro de la plante et la date, ainsi que le nombre total d'épillets de l'épi et le numéro des épillets fleuris au sein de l'épi sont notés. Le numéro de chaque épillet est compté comme suit : l'épillet basal est le n° 1, puis les suivants sont comptés en montant en alternance de part et d'autre de la tige (Figure 31).

Tableau 8 : Nombre d'épis MB et T1 de petit épeautre, récoltés aux 8 stades de prélèvement (en temps thermique °Cj, et en jours après anthèse JAA) en condition contrôle et contrainte thermique modérée. Du fait de la petite taille des grains, les stades les plus précoces ont nécessité le prélèvement d'un nombre d'épis plus important pour avoir suffisamment de matériel biologique.

Stade de récolte (°Cj)	Stade de récolte (JAA)	Condition - Répétition	Nombre d'épis récoltés
0	0	Contrôle - 1	13
0	0	Contrôle - 2	13
0	0	Contrôle - 3	16
0	0	Contrôle - 4	9
40	2	Contrôle - 1	12
40	2	Contrôle - 2	12
40	1	Stress thermique - 1	12
40	1	Stress thermique - 2	12
80	4	Contrôle - 1	6
80	4	Contrôle - 2	6
80	3	Stress thermique - 1	6
80	3	Stress thermique - 2	6
120	6	Contrôle - 1	6
120	6	Contrôle - 2	6
120	4	Stress thermique - 1	6
120	4	Stress thermique - 2	7
180	9	Contrôle - 1	6
180	9	Contrôle - 2	6
180	7	Stress thermique - 1	6
180	7	Stress thermique - 2	6
220	12	Contrôle - 1	6
220	12	Contrôle - 2	6
220	8	Stress thermique - 1	6
220	8	Stress thermique - 2	6
300	16	Contrôle - 1	6
300	16	Contrôle - 2	6
300	12	Stress thermique - 1	6
300	12	Stress thermique - 2	6
700	37	Contrôle - 1	6
700	37	Contrôle - 2	6
700	33	Stress thermique - 1	6
700	33	Stress thermique - 2	6
TOTAL			244

Chez le blé, les épillets fleurissent en premier dans la partie centrale de l'épi. Chez le petit épeautre, le départ de floraison semble plutôt se faire dans la partie basale. De plus, chez le blé, qui a été particulièrement sélectionné par l'Homme au cours de son évolution, la date d'anthèse est synchronisée entre plants ; mais les plants de petit épeautre possèdent une anthèse beaucoup plus étalée dans le temps. Pour les 50 plantes de cette culture, les anthèses ont débuté le 2 mars 2016, soit 106 jours après le semis, et se sont étalées sur 19 jours.

5.2.3. Récolte et analyse des grains

Pour chacun des 40 stades de développement étudiés (présentés dans le Tableau 7), un épi MB ou T1 a été récolté et placé sur glace dans un sachet saturé en humidité. Les plantes ayant eu un épi prélevé n'ont pas été prélevées une deuxième fois afin d'éviter un biais de mesure par modification de l'équilibre source/puits. Les épis ont été pesés, le nombre d'épillets total et le nombre d'épillets fertiles (pour les stades tardifs, après 17 JAA = 357°Cj) ont été comptés. Les 5 épillets les plus basaux parmi ceux fleuris le jour de l'anthèse de l'épi ont été décortiqués, le nombre de grains par épillet a été noté ainsi que le poids frais des grains. Les 2 grains les plus représentatifs ont ensuite été photographiés par Dominique Marcon, l'un avec le sillon visible l'autre non, puis placé 48h dans une étuve à 60°C afin de déterminer leur masse sèche. Les grains destinés à la multiplication du lot de grains ont été récoltés à maturité.

5.3. Application d'un stress thermique modéré durant les stades précoces de développement du grain de petit épeautre

Nous avons pour objectif d'étudier son comportement en réponse à un stress thermique modéré de même ordre que celui appliqué sur le blé dans le cadre du projet BreedWheat. La présence d'un profil d'expression similaire des NAC étudiés chez ces 2 espèces permettra d'appuyer le transfert de plante modèle du blé au petit épeautre pour la suite des expérimentations, notamment la transgénèse.

Tableau 9 : Nombre d'embryons ayant subi un bombardement par biolistique pour chaque construction réalisée. *NAC1-2-3-4-5 = bombardement avec une cassette portant les 5 gènes ; NAC1+2+3+4+5 = bombardement avec 5 cassettes portant chacune un gène.*

Construction	Répétition	Nombre d'embryons bombardés
Surexpression TmNAC 1	1	426
	2	487
Surexpression TmNAC 2	1	425
	2	312
	3	544
Surexpression TmNAC 3	1	660
	2	546
Surexpression TmNAC 4	1	734
	2	705
RNAi NAC1-2-3-4-5	1	465
	2	654
RNAi NAC1+2+3+4+5	1	378
	2	399

5.3.1. Dispositif expérimental

Au stade 3 feuilles, les plantules ont été vernalisées à 4°C pendant un mois avant d'être repotées en pots de 4 L supplémenté avec du fertilisant azoté à libération lente, et placées en serre de confinement S2. Les conditions de culture sont identiques à celles appliquées dans le projet BreedWheat (cf. Matériel & Méthode – Partie 2.1), soit une culture à une température moyenne de 19°C avec l'application d'un stress thermique de 27°C durant 10 jours après anthèse du maître brin.

Parmi les 288 plantes suivies, 133 ont fait l'objet d'un prélèvement après une culture en condition contrôle et 92 ont été prélevées après l'application du stress thermique. Les plantes restantes constituaient les bordures et n'ont pas été incluses dans les prélèvements. Du fait de l'absence de synchronisation dans leurs dates d'anthèse, l'application du stress s'est effectuée par transfert manuel des plantes dans un autre compartiment réglé à 27°C. Après 10 jours, ces plantes étaient replacées dans le compartiment en condition contrôle.

5.3.2. Détermination des stades de développement pour le suivi de l'expression

La notation de l'anthèse et la détermination des stades de développement ont été réalisées de la même manière que pour la culture décrite en Matériel & Méthode – Partie 5.2. Cependant, pour cette culture, des sachets de cellulose transparents ont été apposés sur les épis étudiés avant leur anthèse, afin de sécuriser l'autofécondation de chaque fleur.

5.3.3. Récolte des épis

Les épis maître brins issus des plantes en conditions contrôle et stress thermique ont été prélevés à 8 stades de développement (Figure 30). Sept de ces stades sont centrés sur les phases de divisions cellulaires et de remplissage qui conditionnent les paramètres de rendement, tandis que le point 700°Cj donne un aperçu de la phase de maturation. Pour chaque stade, 2 répétitions ont été réalisées, dans lesquelles plusieurs plantes ont été prélevées, le Tableau 8 donne le nombre de plantes allouées à chaque échantillon.

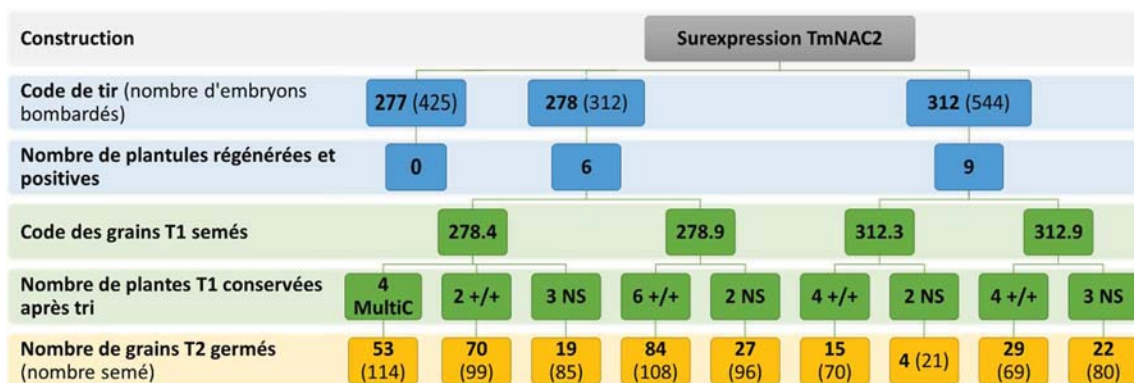


Figure 32 : Schéma du protocole expérimental de génération des lignées transgéniques sur-exprimant TmNAC2 chez le petit épeautre. MultiC = plantes potentiellement multicopies (ayant intégré plusieurs fois la cassette de transformation) ; +/+ = plantes homozygotes pour le transgène ; NS = plantes Null-Ségrégant (-/-).

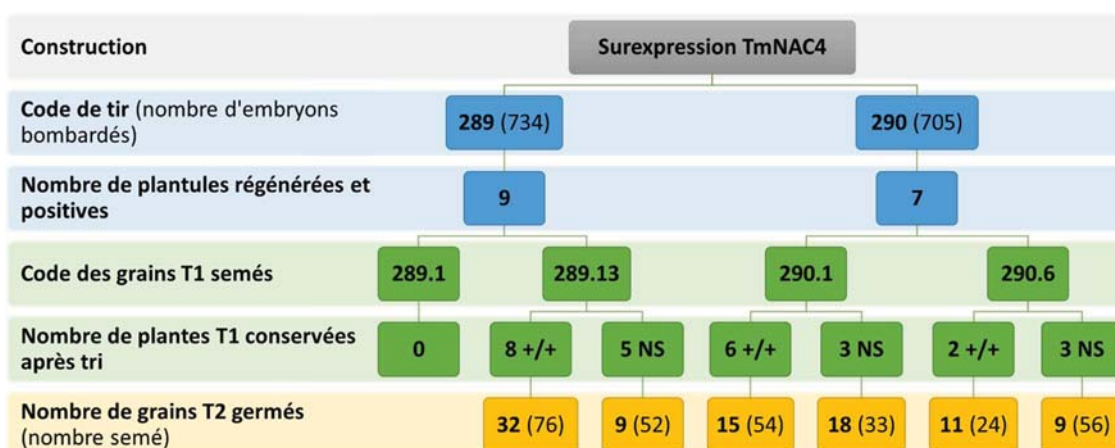


Figure 33 : Schéma du protocole expérimental de génération des lignées transgéniques sur-exprimant TmNAC4 chez le petit épeautre. +/+ = plantes homozygotes pour le transgène ; NS = plantes Null-Ségrégant (-/-).

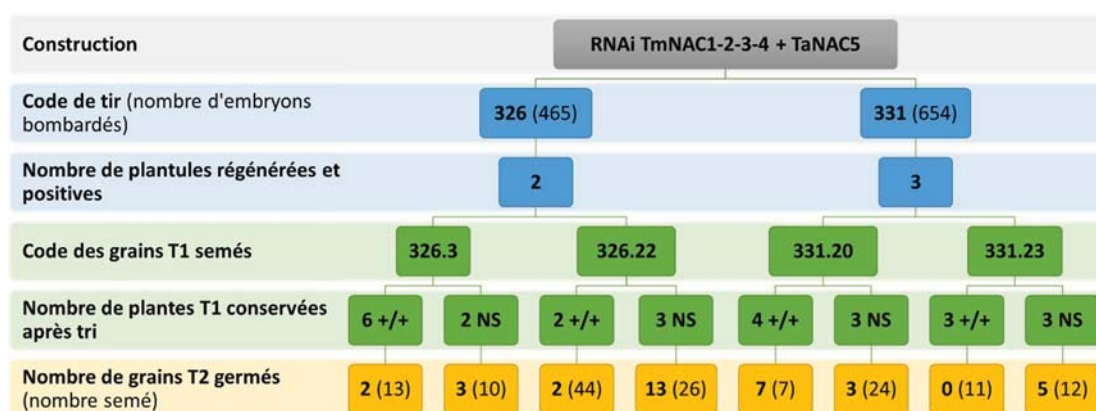


Figure 34 : Schéma du protocole expérimental de génération des lignées transgéniques sous-exprimant TmNAC1, TmNAC2, TmNAC3, TmNAC4 et TaNAC5 chez le petit épeautre. +/+ = plantes homozygotes pour le transgène ; NS = plantes Null-Ségrégant (-/-).

Des études antérieures conduites chez le blé ne prenaient en considération au sein du maître brin que les épillets fleuris le jour de la 1^{ère} anthèse. Cependant, vu la petite taille des grains et le fort taux de stérilité des plants de *T. monococcum* dans la partie apicale de l'épi, nous avons choisi de considérer tous les épillets de la plante. Pour chaque échantillon, les épis ont donc été prélevés, placés dans l'azote liquide puis décortiqués, réunis, broyés et les ARN totaux ont été extraits pour suivre l'expression de 4 gènes TmNAC par RT-qPCR comme décrit en Matériel & Méthode – Partie 3.

5.4. Obtention d'embryons immatures pour la transgénèse

5.4.1. Matériel végétal et dispositif expérimental

L'obtention de plants de petit épeautre transgéniques s'est déroulée sur le site de l'INRA de Crouël (Clermont-Ferrand, France), avec le matériel et l'encadrement du personnel de la plateforme Validation Fonctionnelle de l'UMR GDEC. Des semis sont réalisés toutes les semaines afin d'avoir en permanence des plantes portant des grains au stade adéquat. Les plantes ont été semées puis cultivées en serre en température contrôlée (15°C la nuit et un minimum de 24°C le jour, suivant les conditions d'ensoleillement et la saison) et sub-irrigation manuelle. Une fertilisation a été apportée tous les 15 jours et le développement des pucerons a été limité par l'utilisation du savon noir sur les zones infestées. La réussite de la transformation est conditionnée en partie par la bonne santé des plantes mères.

5.4.2. Récolte des embryons immatures

La transformation des embryons immatures par biolistique a été réalisée le jour de la récolte des grains, qui a lieu entre 12 et 14 jours après l'anthèse. La taille des embryons, estimée à l'œil, doit être d'environ 1mm. Après décortilage et stérilisation des grains immatures, les embryons ont été disséqués et placés à l'obscurité sur milieu plasmolysant (Annexe IV) pour limiter leur destruction lors de la transformation. Environ 500 grains sont prélevés pour chaque bombardement, le nombre d'embryons exact utilisé par bombardement est présenté dans le Tableau 9.

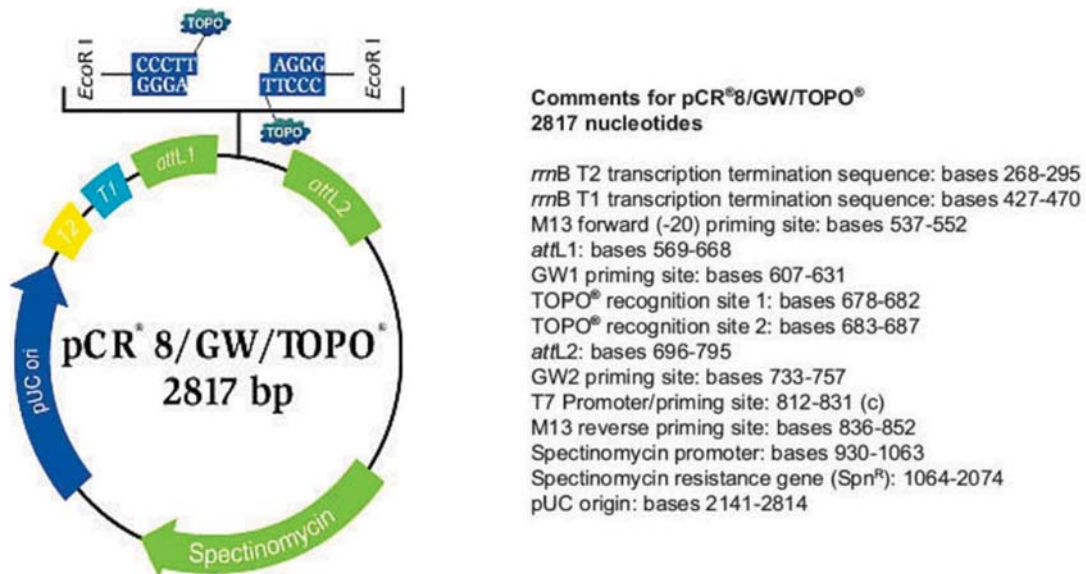


Figure 35 : Le vecteur pCR8/GW/TOPO (Invitrogen). Ce plasmide haute-copie linéarisé de 2817 pb est utilisé comme vecteur donneur de type Gateway[®]. Il possède des sites *attL1* et *attL2* flanquant le site de clonage et nécessaires à la recombinaison site-spécifique, 2 sites *EcoRI* flanquant le site d'insertion du gène d'intérêt et permettant son excision, et des sites d'hybridation des couples d'amorces M13 et GW1/GW2 permettant le séquençage de l'insert. Il possède aussi un gène de résistance à la spectinomycine ainsi qu'un promoteur T7 permettant une transcription in vitro de l'insert.

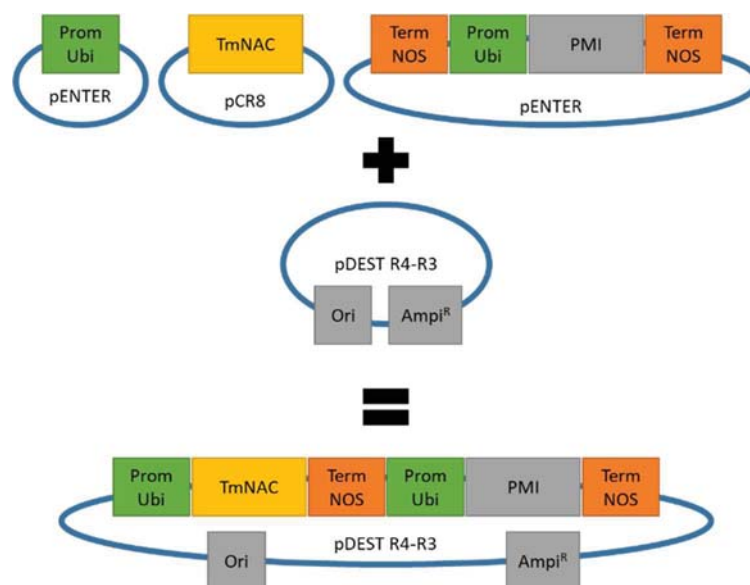


Figure 36 : Schéma de la LR réaction donnant la cassette de surexpression dans le vecteur receveur pDEST R4-R3.

5.5. Transformation génétique du petit épeautre (*T. monococcum*) par biolistique

La construction des cassettes de transformation et le bombardement d'embryons immatures de petit épeautre par biolistique pour l'obtention de lignées transgéniques T0 ont été réalisés en collaboration avec la plateforme de Validation Fonctionnelle de l'UMR. Une fois régénérées, la culture des plantes en serre transgénique jusqu'à l'obtention des grains T1 issus des plantes T0 s'est déroulée dans les serres de confinement S2 de l'INRA de Crouël (Clermont-Ferrand, France). La suite du protocole, du semis des grains T1 à la récolte des grains T3, s'est déroulée dans des serres de confinement S2, sur le campus des Cézeaux (Clermont-Ferrand, France). Les Figure 32, Figure 33 et Figure 34 rappellent les grandes étapes des expérimentations.

5.5.1. Surexpression *in planta* de TmNAC1, TmNAC2, TmNAC3 et TmNAC4

5.5.1.1. Clonage dans le vecteur donneur pCR8/GW/TOPO

Le vecteur donneur pCR8/GW/TOPO est un plasmide haute-copie de 2817 pb qui possède des sites attL1 et attL2 permettant un clonage Gateway, des sites *EcoRI* et des sites d'hybridation des couples d'amorces M13 et GW1/GW2 (Figure 35). Il possède aussi un gène de résistance à la spectinomycine (100 µg/ml), ainsi qu'un promoteur T7 permettant une transcription *in vitro* de l'insert. Les séquences des gènes TmNAC1, TmNAC2, TmNAC3 et TmNAC4 du petit épeautre ont été clonées dans ce vecteur selon un clonage de type TA, par digestion du pGEM-T Easy-TaNAC et du pCR8/GW/TOPO par l'enzyme *EcoRI*. Ce clonage n'est pas directionnel et le sens de l'insertion doit ensuite être vérifié par digestion enzymatique.

La ligation est réalisée dans un volume de 5 µL contenant 1 µL de tampon, 1 µL de vecteur et 3 µL de produit PCR (de 50 à 100 ng/µL), puis le mélange est laissé 30 minutes à température ambiante.

5.5.1.2. Construction et multiplication des cassettes de surexpression

Les cassettes de surexpression ont été générées par recombinaison LR (technologie Gateway) dans un vecteur de destination à entrées multiples, le plasmide pDESTR4-R3. Les 3 plasmides donneurs sont :

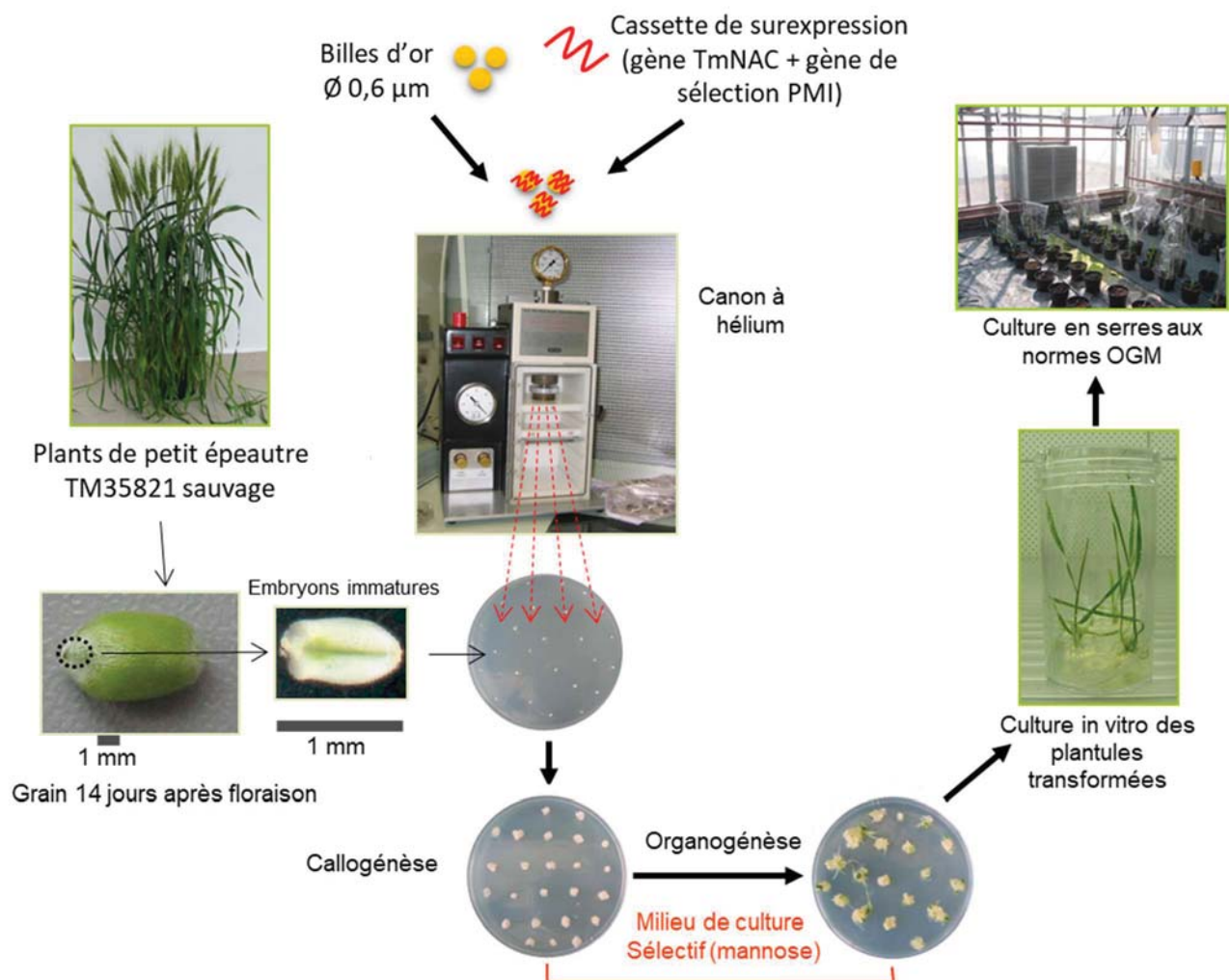


Figure 37 : Schéma récapitulatif de la transformation de plants de petit épeautre par biolistique. A l'aide d'un canon à hélium, des embryons immatures sont bombardés avec des billes d'or recouvertes de la cassette de sur-expression TmNAC + PMI (fixée de manière non-covalente sur les billes grâce à une préparation en présence de spermidine et de CaCl_2). Le gène de sélection PMI permet une sélection des plantes transformées sur un milieu possédant le mannose (sucre non métabolisable par les plantes sauvages) comme seule source de carbone. Après induction de la callogénèse et de l'organogénèse (sur milieu sélectif), les plantules transformées ont été vernalisées 2 mois en chambre froide et placées en culture à la serre.

- ➔ Un plasmide pENTER 5'element contenant le promoteur constitutif fort de l'ubiquitine de maïs (PromUbi),
- ➔ Le plasmide pCR8 dans lequel la séquence ADNc complète de l'un des 4 TmNAC a été clonée,
- ➔ Un plasmide pENTER 3'element contenant le terminateur du gène de la NOpaline Synthase (TermNOS), suivi du gène de sélection Phospho-Mannose Isomérase (PMI) encadré par un second PromUbi et un TermNOS. Ce gène permet à l'embryon transformé d'utiliser le mannose comme source carbonée (Figure 36).

Chacune des cassettes de surexpression permet de regrouper physiquement la cassette sur-exprimant le gène considéré et celle sur-exprimant le gène PMI, toutes les 2 sous le contrôle d'un promoteur de forte activité. Les embryons régénérant sur milieu sélectif ont donc forcément intégré les 2 cassettes, ce qui limite fortement le nombre de faux positifs (c'est-à-dire les plantes se développant sur milieu sélectif sans pour autant sur-exprimer le gène cible).

Le plasmide recombinant obtenu est ensuite digéré par l'enzyme de restriction NotI qui permet d'en extraire les cassettes de surexpression. Celles-ci sont ensuite purifiées par migration sur gel et lavages au butanol et à l'éther (protocole confidentiel de l'INRA). Ceci permet de réaliser la transformation uniquement avec la cassette de surexpression et d'éviter l'insertion de fragments de plasmide inutiles dans les embryons de blé transformés.

5.5.1.3. Transformation d'embryons immatures et régénération de plantules T0

Les protocoles de transformation et de régénération des embryons immatures ont été adaptés de ceux décrits par Wright *et al.* (2001) et Pellegrineschi *et al.* (2002) (Figure 37).

Brièvement, 100 ng de cassette de surexpression ont été mis en contact avec des billes d'or de 0,6 µm de diamètre (BioRad). Des embryons immatures de petit épeautre placés sur milieu plasmolysant (Annexe IV) ont été bombardés avec ce mélange à l'aide d'un canon à particules (BioRad PDS-1000/He biolistic particle system, Du Pont). Après 24h, les embryons bombardés ont été transférés sur un milieu de callogénèse non-sélectif (Annexe IV) et conservés 15 jours à l'obscurité. Régulièrement, les départs de différenciation ont été enlevés des cals pour éviter

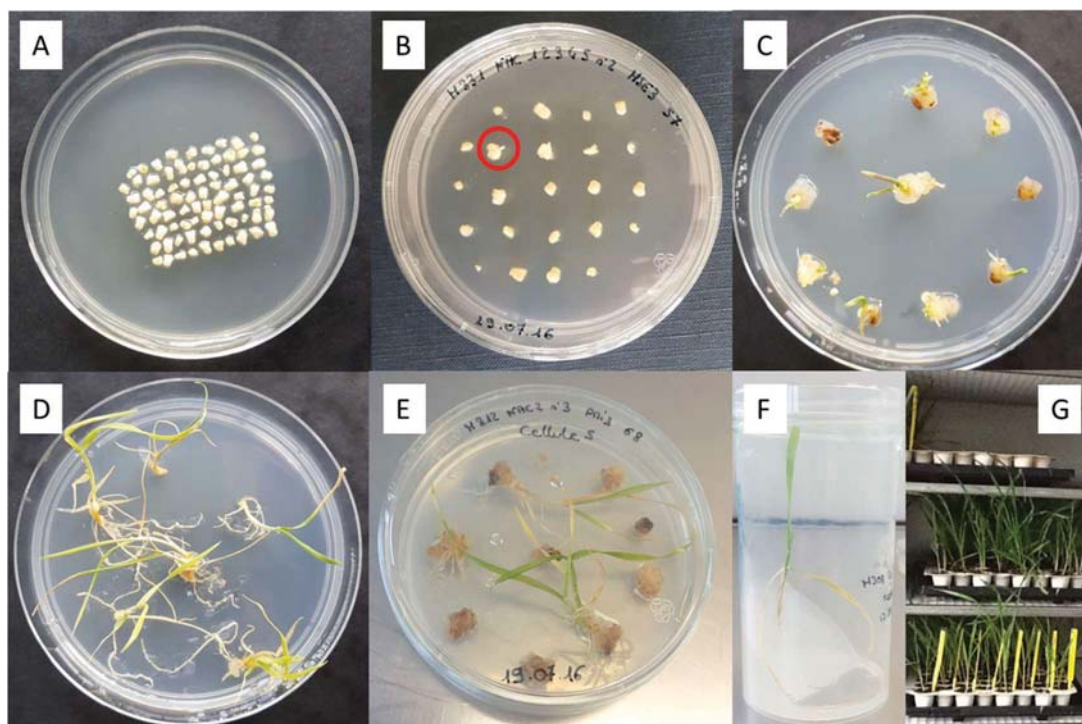


Figure 38 : Différentes étapes de la régénération de plantules transformées in vitro. Les embryons sont extraits des grains immatures puis placés sur milieu plasmolysant pour être bombardés (A). Après bombardement, ils sont transférés sur milieu de callogénèse non-sélectif (B) où les départs de différenciation trop précoces sont retirés régulièrement (cercle rouge). Une fois les cals suffisamment développés, ils sont transférés sur milieu sélectif, contenant une part de mannose de plus en plus importante comme source carbonée (C - E). Les cals n'ayant pas intégré la cassette transgénique portant le gène PMI se nécrosent, les autres se différencient en racines et feuilles (E). Les plantules survivantes sont ensuite repiquées dans des pots individuels (F) puis dans des godets de terre (G).

Tableau 10 : Nombre de plantes obtenues pour les générations transgéniques T1.

Construction	Nombre de plantes T1 obtenues
Sur-Expression-NAC1	120
Sur-Expression-NAC2	120
Sur-Expression-NAC3	104
Sur-Expression-NAC4	120
RNAi TmNAC1-TmNAC2-TmNAC3-TmNAC4-TaNAC5	92
RNAi TmNAC1+TmNAC2+TmNAC3+TmNAC4-TaNAC5	97
TOTAL	653

qu'ils se différencient avant d'avoir un cal d'une taille suffisante pour assurer leur survie. Les cals ont ensuite été repiqués sur un milieu de régénération sélectif (Annexe IV) et disposés dans une chambre de culture à 24°C, sous lumière tamisée puis à la lumière vive. Ils ont été repiqués 3 fois successivement sur un milieu neuf. Les plantules développant des départs de différenciation en feuilles (identifiables grâce à la coloration verte due à la présence de chlorophylle) ont été repiquées sur un milieu d'enracinement (Annexe IV) dans des pots individuels et laissés à la lumière. Leurs systèmes foliaire et racinaire assez développés, ces plantules ont été repiquées en terre dans des godets placés en chambre transgénique puis repotées en pots de 3 litres dans les serres transgéniques de Crouël. Les photos de ces différentes étapes sont présentées en Figure 38). Des sachets d'autofécondation ont été posés sur les épis dès leur apparition et les grains T1 ont ensuite été récoltés à maturité.

Deux séries de tirs par construction ont été réalisées en réplicats techniques. Deux lignées par tir ont été utilisées pour la suite des expérimentations, ceci représente un total de 4 lignées issues d'événements indépendants (c'est-à-dire issues d'embryons différents ayant subi une insertion à des endroits différents dans leur génome).

5.5.1.4. Génotypage et tri à la génération T1

Les grains T1 issus des plantes T0 ayant subi le tir et la régénération *in vitro* ont été semés en godet (2 grains par godet) dans les serres transgéniques des Cézeaux. Du fait de leur très faible taux de germination, les semis ont été complétés si nécessaires pour un total de 653 plantes qui ont ensuite été cultivées à 19°C comme décrit dans le Matériel & Méthode – Partie 5.2. (Tableau 10).

5.5.1.5. Extraction d'ADN génomique

L'extraction d'ADN génomique de feuilles de plantules T1 a été réalisée en série sur la plateforme Gentyane de l'INRA de Crouël. Des échantillons de 50-80 mg de feuilles ont été prélevés, congelés et broyés à l'aide de billes métalliques. L'ADN génomique a ensuite été extrait à l'aide du robot oKtopure (LGC Group), selon le protocole fourni par le fabricant, qui consiste en une lyse des tissus cellulaires, une fixation de l'ADN sur des billes métalliques tandis que le surnageant est retiré, des nettoyages successifs de l'ADN, qui est ensuite

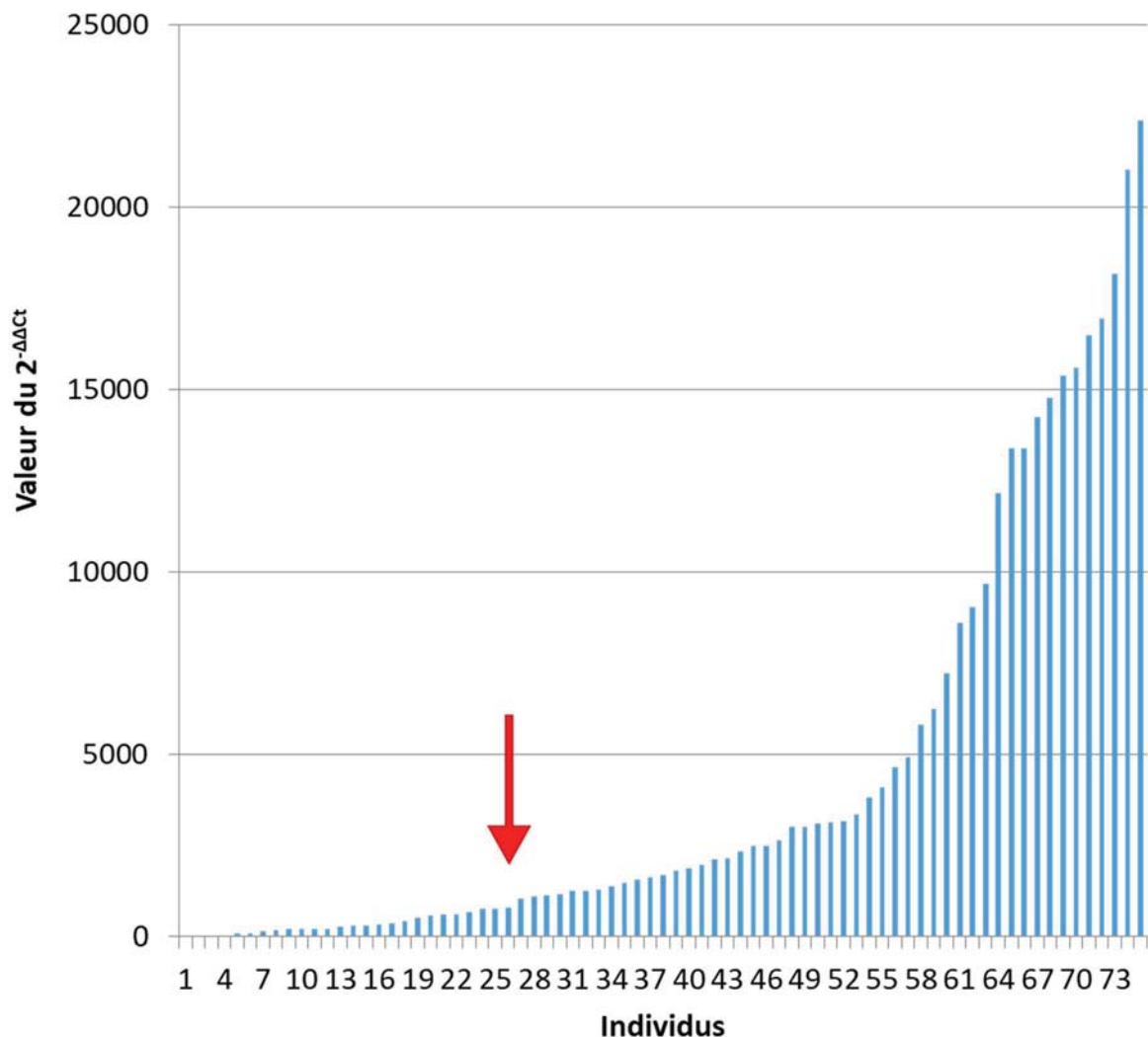


Figure 39 : Valeurs des $2^{-\Delta\Delta C_t}$ obtenus pour toutes les plantes de la construction *SurExpression TmNAC2*. Les 4 individus avec le $2^{-\Delta\Delta C_t}$ le plus faible ont été considérés comme des plantes ayant intégré plusieurs copies du transgène. La flèche rouge indique le décrochement qui nous sert de repère : les individus avant ce décrochement sont considérés comme homozygotes pour le transgène (environ $\frac{1}{4}$ de la population) tandis que les individus après le décrochement sont considérés comme hémizygotes (environ $\frac{1}{2}$ de la population). Les individus avec les valeurs de $2^{-\Delta\Delta C_t}$ les plus élevées sont considérés comme des Null-Ségrégant, des plantes n'ayant pas intégré le transgène.

resuspendu dans un tampon d'éluion. Pour finir, la concentration des échantillons est estimée au NanoQuant (Tecan).

5.5.1.6. Identification des plantes homozygotes pour le transgène

Les plantes de la génération T1 sont supposées ségréger, c'est-à-dire que $\frac{1}{4}$ de ces plantes présenteraient un génotype homozygote pour le transgène (+/+), $\frac{1}{4}$ un génotype homozygote pour l'absence de transgène (Null-Ségrégant) et $\frac{1}{2}$ seraient hémizygotes (ne possédant qu'une seule copie du transgène). La présence de la cassette de surexpression dans l'ADN génomique de feuilles de chacune des plantes T0 a été vérifiée par RT-qPCR en utilisant les amorces listées en Annexe I, qui ciblent spécifiquement le gène TmNAC de la cassette et n'amplifient pas la sa copie endogène, selon les mêmes conditions que précédemment (Matériel & Méthode – Partie 3.2.4). Seules les plantes présentant un profil homozygote ont été sélectionnées.

Parmi les lignées issues d'un tir, la sélection de 8 plantes homozygotes pour le transgène s'est déroulée selon 2 critères :

- Une valeur de Ct inférieure à 35,
- Sur la courbe représentant les $2^{-\Delta\Delta Ct}$ de toutes les plantes d'une construction, classées par ordre croissant, un décrochement a été observé entre environ $\frac{1}{4}$ de la population et les $\frac{3}{4}$ restant. Nous avons considéré que ce $\frac{1}{4}$ correspondait aux homozygotes pour le transgène et avons conservé uniquement les plantes concernées (Figure 39).

Par ailleurs, les 5 plantes par tir présentant les valeurs de Ct les plus élevées possibles, voire supérieure à la limite de détection du thermocycleur (supérieur à 45) ont été considérées comme Null-Ségrégant, c'est-à-dire dont les parents ont subi le tir et la régénération *in vitro*, mais ayant perdu le transgène par ségrégation. Quatre plantes de surexpression TmNAC2 présentant un nombre de copie du gène TmNAC2 nettement plus élevé que les autres ont également été conservées en supposant qu'elles pouvaient être des plantes ayant intégré plusieurs copies du transgène (Figure 39).

Au total, 173 plantes ont été repotées dans des pots de 4 litres pour continuer la culture jusqu'au stade récolte des grains T2. Des problèmes au cours de la culture (acariens phytophages, repotage extrêmement tardif dû aux délais d'analyses et sur-fertilisation) ont conduit à un taux de stérilité très élevé de cette génération.

5.5.1.7. Application d'un stress thermique modéré sur la génération T2 et analyse des grains T3

Les plantes de la deuxième génération après transformation ont été semées fin janvier 2018 puis soumises à un stress thermique similaire à celui mis en place dans l'étude de BreedWheat (cf. Matériel & Méthode – Partie 2.1.1) pour enfin être génotypées et phénotypées en juin 2018. Dû à des limites techniques (temps, nombre de grains disponibles et espace de culture), seules les constructions de surexpression TmNAC2 et TmNAC4 ont été étudiées.

5.5.1.8. Dispositif expérimental

Après 12 jours de levée de dormance à 4°C, 1184 grains ont été mis à germer sur papier Wattman humidifié en boîte de Petri, à l'obscurité et température ambiante. Ces 1184 grains correspondent à la totalité des grains obtenus de la génération précédente pour les 2 constructions de surexpression et la construction de sous-expression (présentée par la suite). Après 3 jours, ceux-ci ont été repotés en godet puis en pots de 4L dans la serre transgénique des Cégeaux et arrosés en hydroponie. Au total, 464 plantes se sont développées sur 1184 semées, soit un taux de germination de 39,2 %.

La température a été abaissée à 15°C/13°C (jour/nuit) pendant 2 semaines au stade végétatif pour stimuler le tallage puis remontée à 21°C/15°C (jour/nuit, 19°C de moyenne). Le jour de l'anthèse de leur maître brin, la moitié des plantes a été transférée dans un compartiment de stress thermique (16h à 29°C jour/ 8h à 23°C nuit) pendant 10 jours, avant d'être transférée à nouveau dans le compartiment contrôle. Des sachets d'autofécondation ont été apposés sur la majorité des épis au fur et à mesure de leur épiaison afin d'assurer une descendance homozygote.

5.5.1.9. Génotypages des plantes (présence de la cassette dans le génome)

Au cours de leur développement, des fragments de feuilles ont été prélevés sur quelques plantes. La présence de la cassette dans leur ADN génomique a été vérifiée par PCR selon les mêmes conditions que décrites dans le Matériel & Méthode – Partie 3.2.3.

5.5.1.10. Suivi de l'expression de gènes par qPCR (expression de la cassette)

Au cours du développement des plantes, des prélèvements d'épis maître brins ont été réalisés à 180°Cj et 400°Cj après anthèse en conditions contrôle et stress thermique. Le stade 180°Cj a été choisi en fonction des résultats obtenus lors de la cinétique de développement du petit épeautre en stress thermique. En effet les gènes présentent la différence maximale de réponse entre les conditions contrôle et stressée à ce stade. Le stade 400°Cj a été choisi parce qu'il correspond à la phase de remplissage, durant laquelle l'expression des TmNAC a normalement atteint son expression maximale et forme un plateau.

Après extraction des ARN selon le protocole décrit dans le Matériel & Méthode – Partie 3.1, l'expression des gènes TmNAC1, TmNAC2, TmNAC3 et TmNAC4 a été suivie par qPCR dans des plantes +/+ et Null-Ségrégant issues de différentes lignées dans les mêmes conditions que décrit en Matériel & Méthode – Partie 3.2.4. L'expression des 4 gènes TmNAC a été suivie dans les plantes sur-exprimant les gènes TmNAC2 ou TmNAC4 afin de voir si une éventuelle diminution de leur expression aurait pu se mettre en place par compensation de la surexpression. Les amorces utilisées sont présentées en Annexe I.

5.5.1.11. Phénotypage des plantes transformées (effet de la cassette)

Mesure de paramètres agronomiques : Au cours du développement de chaque plante, son homozygotie (+/+ ou Null-Ségrégant), les dates d'épiaison (sortie de l'épi de sa feuille drapeau) et d'anthèse du maître brin ont été notées, ainsi que le numéro des épillets fleuris le jour de l'anthèse du maître brin. Des prélèvements d'épis maître brins ont été réalisés à maturité physiologique (stade 900°Cj), sur lesquels nous avons mesuré le nombre d'épis par plante ainsi que le nombre d'épillets et d'épillets fertiles par maître brin. Pour finir, l'utilisation d'un système photographique assisté par ordinateur a permis d'obtenir :

- le nombre de grains par maître brin,
- le poids de la totalité des grains de l'épi,
- son poids de mille grains (PMG),
- la surface, longueur et largeur minimum, maximum, moyenne des grains de l'épi et l'écart-type associé.

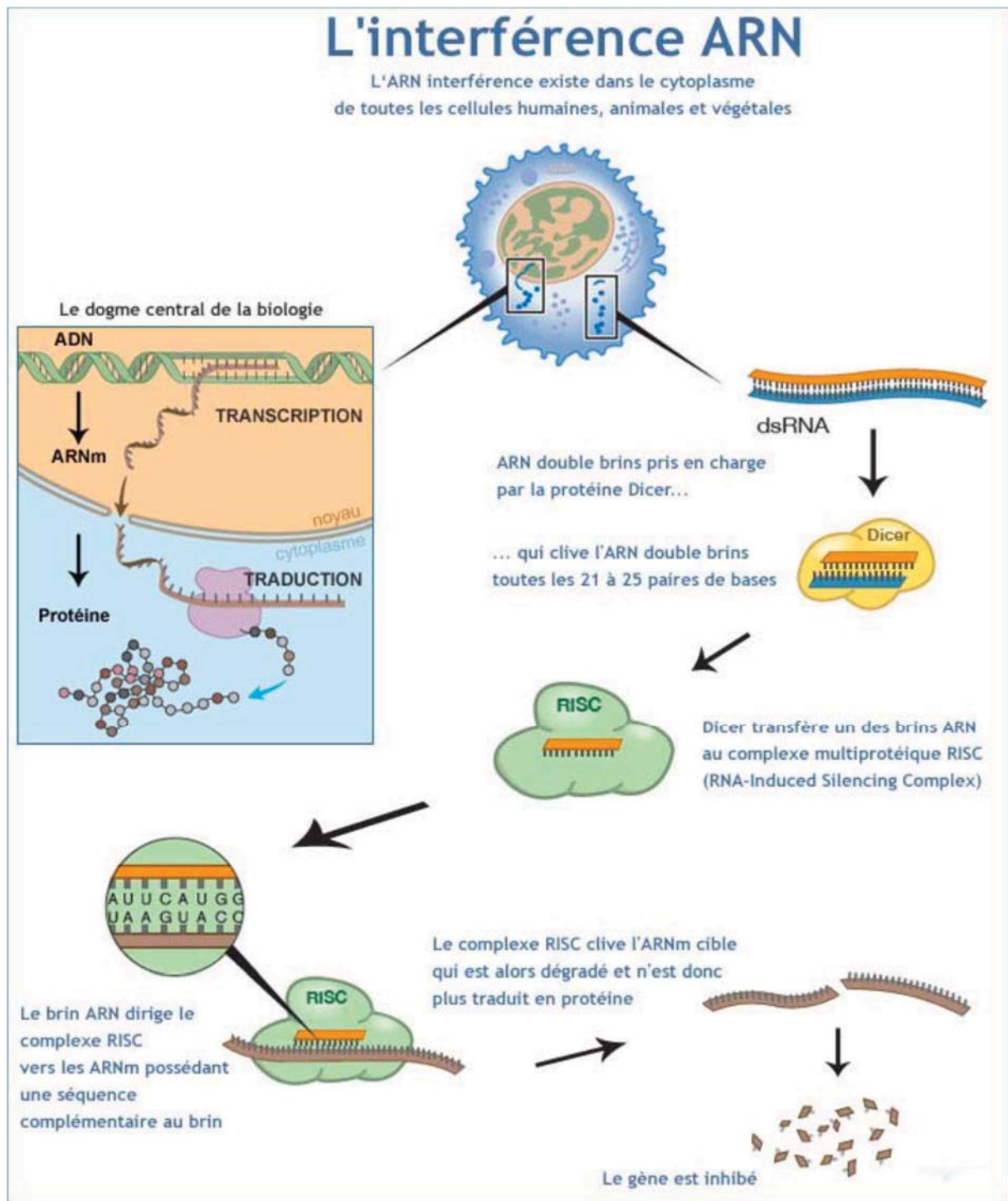


Figure 40 : Principe général du processus d'extinction post-transcriptionnel de l'ARN interférence. Le mécanisme d'entretien de l'extinction au niveau transcriptionnel n'est pas représenté ici (source : Sciences et Avenir).

Détermination de la teneur en protéines : Les grains restants de chaque maître brin ont été broyés en farine complète à l'aide d'un broyeur et de billes métalliques. Le contenu total en azote (exprimé en pourcentage de masse de farine) a été mesuré pour 5 mg de chaque échantillon selon la méthode de combustion de Dumas modifiée (méthode AOAC n°992.23) à l'aide d'un analyseur FlashEA 1112N/Protein analyzer (Thermo Scientific). Chaque mouture a été analysée en duplicat. Le contenu total en protéines a été estimé en multipliant le contenu total en azote par 5,7 (Sosulski et Imafidon 1990). La valeur en pourcentage peut être ramenée à une masse de protéines en la multipliant par le poids du grain considéré, ce qui a permis de calculer :

- le pourcentage de protéines du grain,
- la quantité de protéines du grain,
- la quantité de protéines du maître brin entier.

5.5.2. Sous-expression *in planta* des 4 TmNAC par ARN interférence (RNAi)

Pour compléter les informations obtenues par le phénotypage de plantes sur-exprimant les 4 TmNAC, nous avons produit en parallèle des plantes transgéniques sous-exprimant ces gènes.

5.5.2.1. Principe

L'ARN interférence est une technique basée sur la dégradation post-transcriptionnelle séquence-spécifique d'ARN messagers dans le cytoplasme, suite à l'introduction d'ARN double brin dans l'organisme transformé.

Pour éteindre l'expression d'un gène de manière spécifique, une construction est créée à partir d'une portion de sa séquence codante insérée 2 fois, tête-bêche, dans un plasmide qui est ensuite inséré de manière non-ciblée dans le génome de l'organisme. La transcription de cette séquence par la cellule produit des ARN double-brin avec une structure en épingle à cheveux et homologues à la séquence du gène ciblé. L'organisme identifie les ARN double brin formés comme des produits aberrants et les dégrade en séquences de 21 à 24 nucléotides, ainsi que tous les ARN du gène ciblé possédant la même séquence (Figure 40).

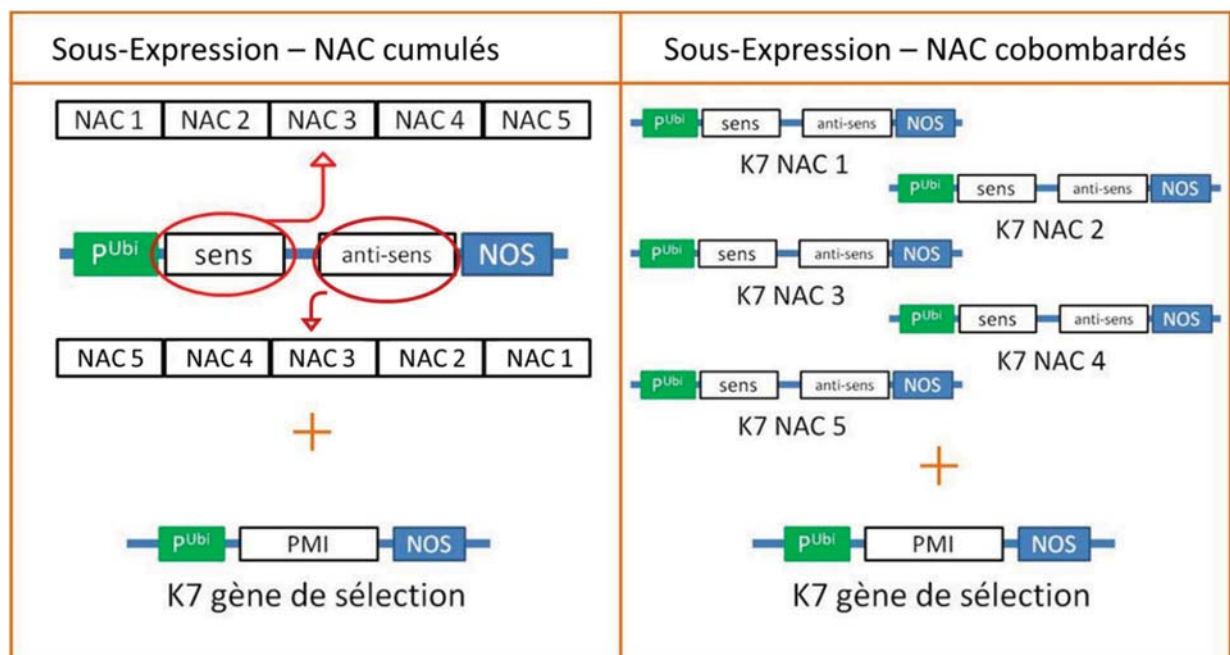


Figure 41 : Structure des cassettes utilisées pour l'ARN interférence. Une portion de 156 pb de la séquence codante de *TmNAC1*, *TmNAC2*, *TmNAC3*, *TmNAC4* et *TaNAC5* est clonée en orientation sens et anti-sens dans un plasmide *pSTARGATE*. Dans un cas, ces portions sont cumulées les unes à la suite des autres dans le but d'obtenir des plantes sous-exprimant tous les gènes et d'autre part, ces portions sont co-bombardées pour obtenir des embryons sous-exprimant des combinaisons différentes de gènes, à des niveaux variables. Une cassette de sélection est co-bombardée avec cette cassette RNAi, à un ratio molaire de 2:1 en faveur de la cassette RNAi pour en favoriser l'insertion.

Cependant, cette dégradation n'est pas forcément totale, et l'ARN interférence permet de générer des plantes transgéniques présentant un taux d'extinction variable, très utile lorsque l'extinction complète du gène est létale pour l'organisme (Travella *et al.* 2006 ; Fu *et al.* 2007). De plus, la diminution de l'expression du gène est persistante à travers les divisions cellulaires et peut être transmise aux générations suivantes.

Le point clé pour la réussite d'une validation fonctionnelle par ARN interférence chez le blé tendre est que la portion du gène insérée dans la cassette RNAi doit produire des séquences 21 nucléotides spécifiques du gène ciblé et uniquement de celui-ci, afin de ne pas entraîner l'extinction d'autres gènes (off-target) (Fu *et al.* 2007). Ceci peut poser problème dans le cas de grandes familles de gènes aux séquences très proches.

5.5.2.2. Construction des cassettes de transformation RNAi

Lors de la construction des cassettes RNAi (2015), les données disponibles sur la plateforme Genevestigator ont révélé l'existence d'un gène chez le blé tendre présentant un profil d'expression spécifique au grain et identique aux 4 NAC suivis. Nous l'avons baptisé TaNAC5. L'amplification PCR de ce gène chez le petit épeautre a été un échec. Toutefois, afin qu'aucun phénomène de compensation de fonction physiologique lié à la présence d'un gène similaire aux 4 TmNAC n'opère, la séquence du blé tendre de ce gène a été ajoutée pour la création des constructions d'extinction.

Ainsi, 2 types de plantes RNAi ont donc été générés (Figure 41) :

- une première série d'embryons immatures a été bombardée avec un mélange des 5 cassettes contenant la construction « Promoteur Ubiquitine – Séquence RNAi de TmNAC sens – Intron espaceur - Séquence RNAi de TmNAC anti-sens – Termineur NOS » construites pour TmNAC1, TmNAC2, TmNAC3, TmNAC4 et TaNAC5, ainsi que la cassette de sélection contenant le gène PMI. Les lignées régénérées de ce co-bombardement d'embryons auront intégré entre une et 5 des cassettes RNAi en plus de la cassette de sélection, et par ségrégation, devraient donner un panel de plantes transgéniques ayant une combinaison variée de gènes TmNAC sous-exprimés,
- une deuxième série d'embryons immatures a été bombardée par une cassette contenant les séquences RNAi des 4 TmNAC et de TaNAC5 à la suite, tête-bêche, encadrées par le

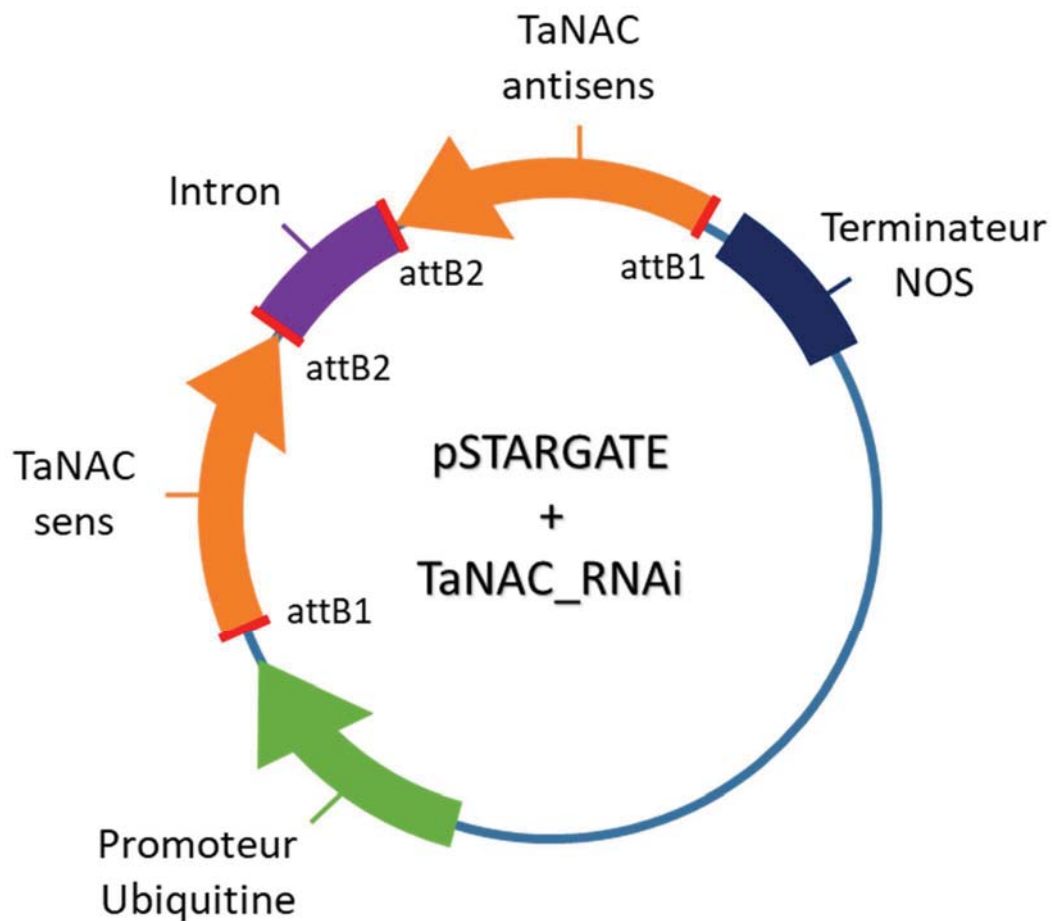


Figure 42 : Le vecteur pSTARGATE. Le promoteur Ubiquitine permet une forte expression de la cassette chez les monocotylédones transformées. L'insertion de la séquence d'intérêt en sens et anti-sens a lieu via une recombinaison LR aux sites de recombinaisons attR1 et attR2. Le terminateur du gène NOS (NOpaline Synthase) marque la fin de la cassette d'expression. La séquence intronique permet un repliement du transcrit en épingle à cheveux, puis elle est excisée. L'enzyme Dicer reconnaît ces ARN double brin et les clive en petits ARN de 21 nucléotides. Ceux-ci vont permettre au complexe RISC de reconnaître et dégrader les transcrits homologues de la séquence d'intérêt dans la cellule végétale.

même promoteur ubiquitine et le terminateur NOS ; ainsi que par la cassette contenant le gène de sélection PromUbi – PMI. Les lignées régénérées de ce bombardement d'embryons devraient donc sous-exprimer les 5 NAC et éviter tout phénomène de compensation entre eux au cours du développement du grain de blé.

Contrairement à la surexpression, les cassettes RNAi contenant les portions RNAi des gènes NAC et la cassette contenant le gène de sélection PMI ne sont pas physiquement liées. Ceci engendre un nombre plus important de faux positifs (des plantes croissant sur le milieu sélectif sans posséder la cassette RNAi). Afin de contrebalancer cela, la cassette ou le mix des cassettes RNAi et la cassette de sélection ont été mélangées selon un ratio molaire de 2:1 avant d'être bombardé, pour favoriser l'insertion de la/des séquence(s) RNAi.

Des fragments de 156 pb de la séquence codante de chacun des 5 NAC ont été choisis sur la base de la spécificité de tous les tronçons de 21 nucléotides, encadrés par des AttL1 et AttL2 permettant le clonage Gateway, et synthétisé par la société Proteogenix dans le vecteur donneur pUC57.

5.5.2.3. Le vecteur receveur pSTARGATE

Un clonage par réaction LR (technologie Gateway) a été réalisé dans le plasmide pSTARGATE, qui possède 2 sites d'insertion de la cassette, en orientation sens et anti-sens situés entre un promoteur ubiquitine et un terminateur NOS. Entre ces 2 sites d'insertion se trouve une séquence intronique issue du génome d'*Arabidopsis*, qui aide à la formation de la structure ARN double-brin en épingle à cheveu (Smith *et al.* 2000).

La recombinaison s'effectue entre le pUC57 et un vecteur de destination Gateway, le pSTARGATE adapté, fourni par le CSIRO selon la technologie Gateway (LR) (Australie ; <http://www.pi.csiro.au/rnai/vectors.htm>; Figure 42, Figure 43), qui a ensuite été utilisé pour transformer des cellules compétentes *E. coli* DH5α comme décrit en Matériel & Méthode – Partie 3.3.3. Le vecteur recombinant a ensuite été digéré par l'enzyme de restriction NotI et la cassette a été purifiée par migration sur gel suivie de lavages au butanol et à l'éther (protocole confidentiel – propriété de l'INRA). Des embryons immatures de petit épeautre ont été transformés par biolistique avec un mélange de ces cassettes et les plantules T0 ont été régénérées, comme précédemment mentionné en Matériel & Méthode – Partie 5.5.1.

TaNAC, séquence RNAi

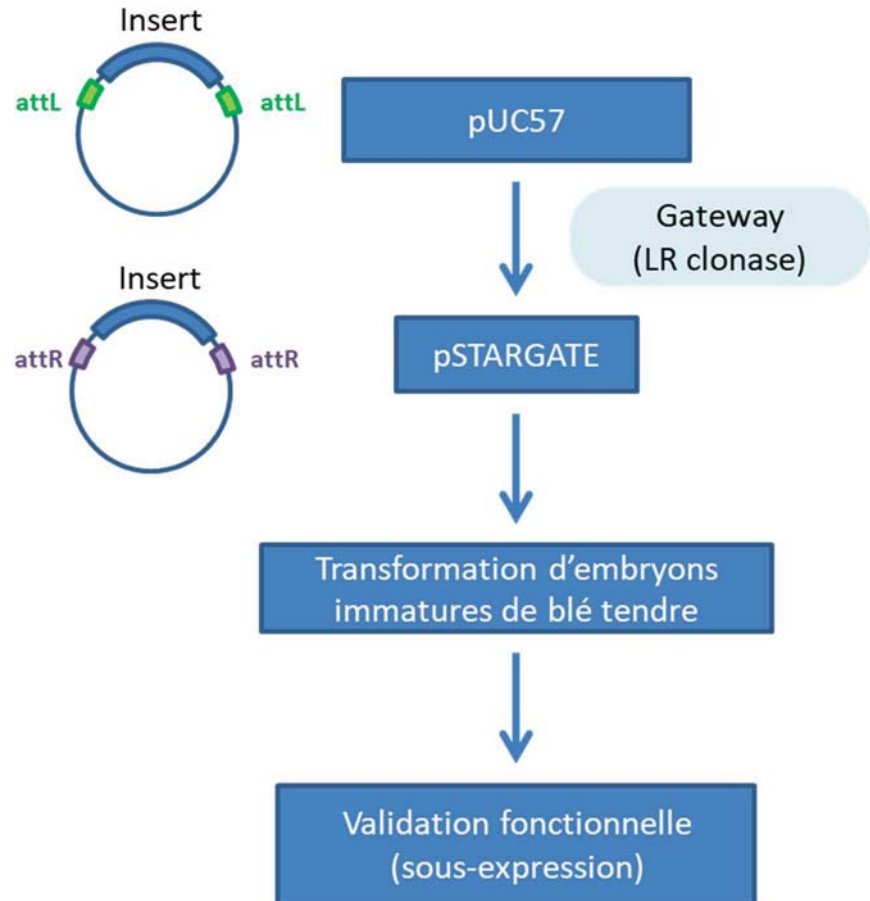


Figure 43 : Schéma global des étapes successives de la validation fonctionnelle par sous-expression. La séquence RNAi de chacun des 4 TaNAC a été clonée dans le vecteur receveur pSTARGATE par la technologie Gateway.

5.5.2.4. Génotypage et suivi de l'expression des plantes transformées

Les plantes transformées T0 ont été génotypées pour la présence de la cassette par PCR sur ADN génomique, à l'aide de 2 couples d'amorces amplifiant une partie de la séquence intronique d'Arabidopsis. Dans les plantes T1, la présence d'une ou plusieurs cassettes d'expression et l'identification de ces cassettes ont été étudiées par RT-qPCR à l'aide de couples d'amorces ciblant spécifiquement chaque cassette utilisée, dans les mêmes conditions que décrit en Matériel & Méthode – Partie 3.2.4. Les amorces sont décrites dans l'Annexe I.

Seules les plantes transformées avec la cassette cumulant les 5 TmNAC ont été étudiées en génération T2. Elles étaient peu nombreuses et ont donc été étudiées uniquement en conditions thermique contrôle (19°C). Le niveau d'expression des gènes dans les plantes transgéniques et Null-Ségrégant a été suivi par RT-qPCR sur ARN totaux de grain (cf. Matériel & Méthode – Partie 3), en utilisant les mêmes amorces que pour l'étude des plantes de surexpression. Leur phénotypage a été identique à celui réalisé sur les plantes sur-exprimant TmNAC2 ou TmNAC4.

6. Significativité des différences statistiques observées dans les lignées transgéniques et dans les analyses transcriptionnelles

Le test statistique utilisé est précisé dans la légende de chacun des graphiques qui ont nécessité une analyse statistique.

Dans le cas d'une analyse 2 à 2, le test de Fisher a été utilisé pour étudier l'égalité des variances 2 à 2, puis le test de Student pour mettre en évidence une différence significative entre les moyennes, notée * au seuil 0,05 et ** au seuil 0,01.

Lors de la recherche de différences significatives sur plus de 2 mesures, la normalité de la distribution a été vérifiée par un test de Shapiro (seuil à 0,05), et l'homoscédasticité des variances par un test de Bartlett (seuil à 0,05). Ces 2 conditions n'étant jamais réunies, le test non-paramétrique de Kruskal-Wallis a été utilisé pour l'analyse de variances, à l'aide du logiciel R Studio (<http://www.R-project.org>, R development Core Team, 2008). Cela a notamment été utilisé pour l'étude des données de génotypage (accumulation des transcrits) et de phénotypage des lignées transgéniques vs les lignées Null-Ségrégant en conditions contrôle et stress thermique.

Chapitre I

RECENSEMENT DES MEMBRES DE LA FAMILLE TaNAC CHEZ LE BLE TENDRE

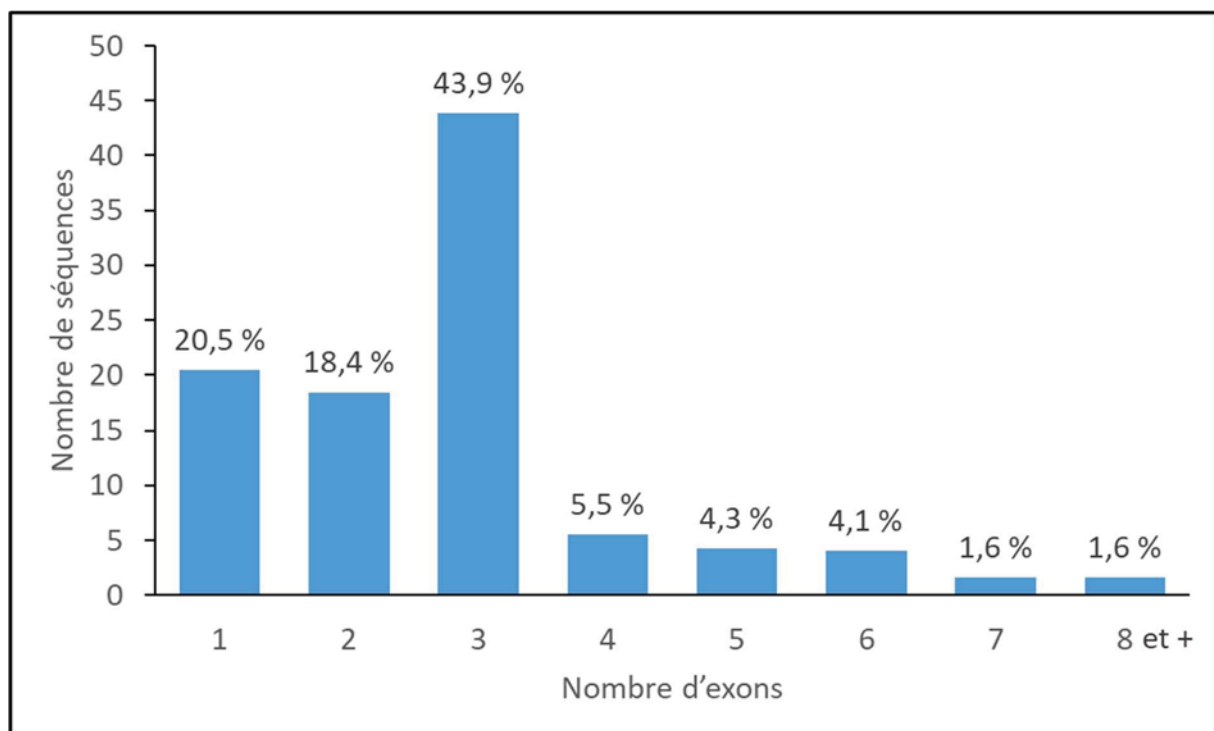


Figure 44 : Distribution des gènes TaNAC selon le nombre d'exons identifié dans leurs séquences génomiques. La majorité des séquences génomiques des TaNAC portent 3 exons et 2 introns, le gène *TraesCS5B01G550800* contient un maximum de 16 exons. Les chiffres au-dessus des histogrammes représentent leur pourcentage par rapport aux 488 séquences.

Avant-Propos :

Les NAC sont une large famille de facteurs de transcription aux rôles pléiotroiques, dont les membres ont été recensés chez de nombreuses espèces. Au commencement de ma thèse, cette famille n'avait pas été décrite chez le blé tendre ni le blé dur. L'objectif de ce chapitre a été de réaliser le recensement de la famille NAC chez le blé tendre, en utilisant la dernière base de données publique disponible.

1. Recensement et description de la famille de facteurs de transcription TaNAC

L'IWGS RefSeq v1.0 est l'annotation la plus récente disponible du génome de blé tendre. Elle a été publiée par l'International Wheat Genome Sequence Consortium le 17 août 2018 (Appels *et al.* 2018), et contient le dernier assemblage de la pseudomolécule du cultivar Chinese Spring. Cette base de données assure la couverture de 94% du génome du blé et recense 107 891 gènes de haute confiance (high-confidence).

Les gènes NAC sont caractérisés par la présence d'un domaine NAC dans la partie N-terminale de leur séquence. Le profil Pfam HMM (Hidden Markov Model) de ce domaine, appelé PF02365 et équivalent au profil IPR036093 d'Interproscan, a été construit à partir de 539 protéines NAC représentatives de diverses espèces végétales. Pour identifier les séquences qui possèdent le domaine NAC, ce profil a été utilisé sous le logiciel UGENE, ce qui a permis de procéder à un premier crible de la totalité de la base de données de l'IWGS RefSeq v1.0. Cette première analyse a généré une liste de 501 séquences contenant potentiellement le motif NAM avec une valeur seuil de e-value égale à e^{-1} . La présence du domaine NAC (IPR036093) sur ces séquences a ensuite été vérifiée avec InterProScan implémenté sous le logiciel Blast2Go. Seules les séquences entières ont été conservées, ce qui a abouti à un set final de 488 séquences protéiques identifiées comme appartenant à la famille des NAC chez le blé *Triticum aestivum* : les TaNAC. La liste de ces 488 séquences est disponible dans la Supplemental Table S3 de la publication Guérin *et al.* (2019). Parmi celles-ci, 460 et 28 séquences sont respectivement High Confidence (HC, haut niveau de confiance) et Low Confidence (LC, faible niveau de confiance). Dans ce dernier cas, « LC » est ajouté à la fin de l'identifiant du gène. Par exemple si l'on décompose l'identifiant TraesCS2B01G556700LC.1, cela nous apprend que c'est un gène de *Triticum aestivum*, cultivar Chinese Spring, ancré sur le chromosome 2B, issu de l'annotation 01 ; s'en suit un code d'ancrage sur la pseudomolécule (ses gènes voisins portent

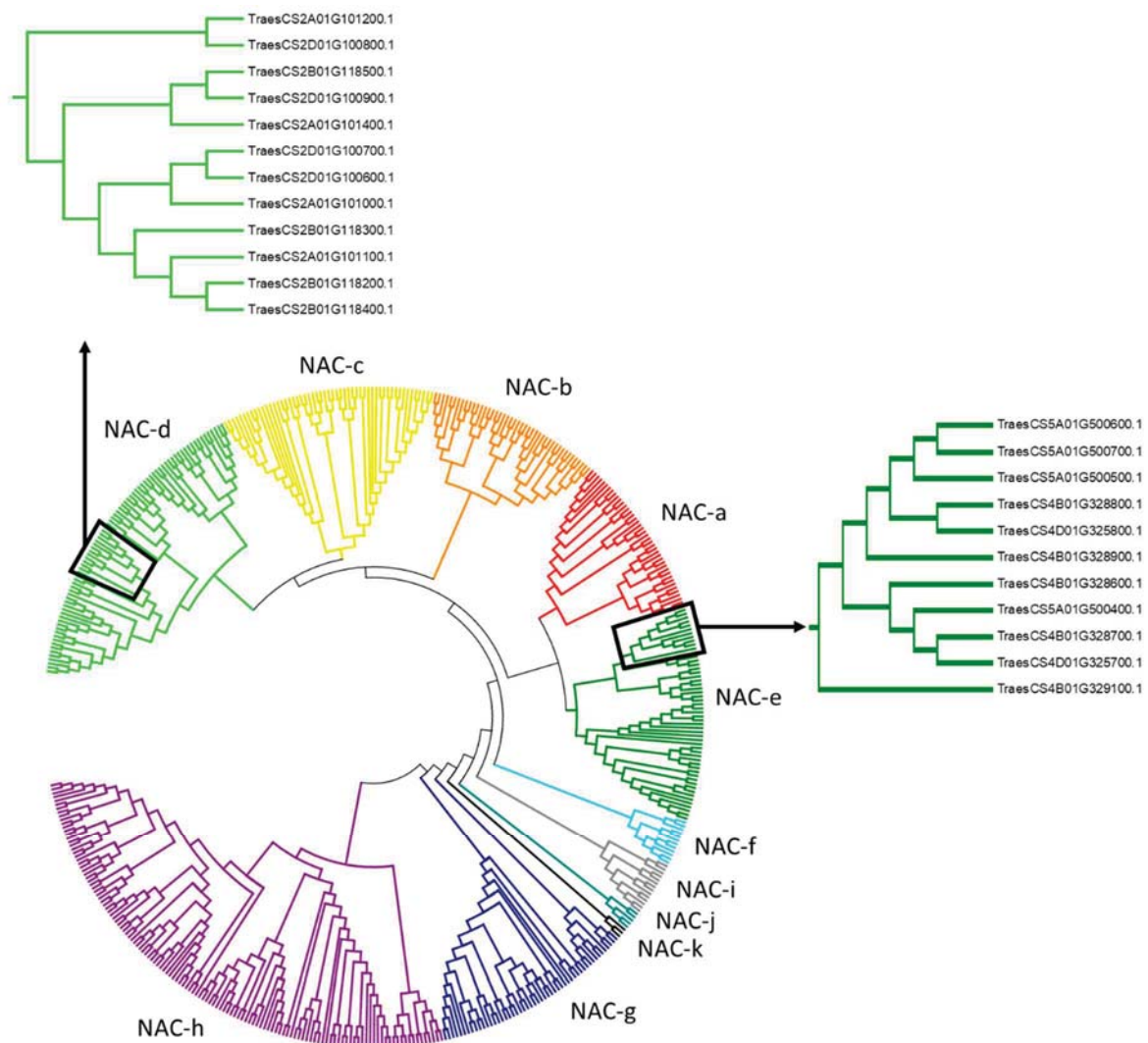


Figure 45 : *Arbre phylogénétique réalisé à partir de l'alignement des séquences protéiques complètes des 488 TaNAC. L'alignement a été réalisé selon la méthode du maximum de vraisemblance. Les 11 sous-familles sont représentées par des couleurs différentes. Les 2 clades étudiés comme exemple de la duplication en tant que moteur de l'expansion de la famille TaNAC sont détaillés à en haut et à droite de la figure.*

les codes 556600 et 556800), la mention LC indique que c'est un gène Low Confidence et pour finir « .1 » correspond au variant ARN utilisé.

En plus des séquences en elles-mêmes, l'assemblage de la pseudomolécule fournit des informations sur la structure exon-intron des gènes TaNAC. La majorité (214 séquences, soit 44%) contiennent 3 exons, tandis qu'un cinquième des séquences (c'est-à-dire 100 séquences sur 488) ne possèdent pas d'intron. À l'opposé, une séquence (TraesCS5B01G550800) contient 16 exons (Figure 44).

La longueur de la séquence génomique des 488 séquences varie de 273 à 15074 pb, correspondant à des séquences protéiques de 91 (TraesCS2B01G556700LC) à 1030 aa (TraesCS5B01G550800), et d'une longueur moyenne de 370 aa. Tous les domaines NAC ont approximativement la même longueur (environ 150 pb) et la même position dans la protéine. Cependant, 8 séquences, représentant 3 groupes d'homéologie, contiennent 2 domaines NAC (TraesCS2A01G363700, TraesCS2B01G381700, TraesCS2D01G361500 ; TraesCS3A01G269900, TraesCS3B01G303800, TraesCS3D01G269600 ; et TraesCS2A01G328100, TraesCS2D01G334800) (Supplemental Table S1 de la publication Guérin *et al.* (2019)).

Une première liste répertoriant les membres de la famille TaNAC a déjà été réalisée par Borrill *et al.* (2017) à partir de la base de données TGAC publiée par Clavijo *et al.* (2017). Nous avons évalué le nombre de séquences TaNAC 100% identiques entre ces 2 listes. Ceci nous a permis de constituer un sous-ensemble de 361 séquences que nous considérons comme de très haute qualité, puisque présentes dans 2 bases de données construites de manière indépendante. Ce sous-ensemble a été utilisé pour les analyses phylogénétiques, tandis que la totalité des 488 séquences recensées a été utilisée pour le reste des analyses, notamment les analyses transcriptomiques. Par ailleurs, la combinaison des 2 annotations en un jeu de données de 361 séquences a permis d'améliorer les informations disponibles pour certains gènes TaNAC, notamment en ce qui concerne leur localisation chromosomique. Par exemple, sur les 2 gènes non ancrés sur un chromosome dans la base de données IWGS RefSeq v1.0, le gène TraesCSU01G137200 s'est avéré 100% identique au gène TRIAE_CS42_2AL_TGACv1_093618_AA0283710, cartographié sur le chromosome 2A dans la base de données TGAC. De la même manière, parmi les 13 séquences non cartographiées dans la base de données TGAC, 12 étaient identiques à des séquences ancrées sur un chromosome de l'IWGS RefSeq v1.0.

Tableau 11 : Homologie observée entre la séquence TuG1812G0700002220.01.P02 de T. urartu et 15 séquences TaNAC chez le blé tendre. Pour ces 15 gènes TaNAC, la meilleure séquence homologue chez T. urartu est la même. Ce résultat est un argument pour avancer l'hypothèse de nombreux évènements de duplication locaux comme moteurs de l'expansion importante de la famille NAC chez le blé tendre par rapport aux autres céréales. Le seuil de e-value minimal pour désigner 2 séquences comme « meilleures homologues » a été fixé à e^{-10} .

Séquence TaNAC de blé tendre	Meilleur hit chez <i>T. urartu</i>	e-value
TraesCSU01G174800.1	TuG1812G0700002220.01.P02	2.70E-63
TraesCS7A01G209100.1	TuG1812G0700002220.01.P02	6.90E-62
TraesCS6D01G135400.1	TuG1812G0700002220.01.P02	3.80E-60
TraesCS7D01G211300.1	TuG1812G0700002220.01.P02	8.70E-60
TraesCS7A01G209500.1	TuG1812G0700002220.01.P02	2.40E-59
TraesCS6A01G146100.1	TuG1812G0700002220.01.P02	4.20E-59
TraesCS7B01G116200.1	TuG1812G0700002220.01.P02	2.90E-51
TraesCS7D01G210900.1	TuG1812G0700002220.01.P02	1.70E-38
TraesCS7B01G116300.1	TuG1812G0700002220.01.P02	5.30E-32
TraesCS7A01G209200.1	TuG1812G0700002220.01.P02	1.70E-18
TraesCS7A01G209400.1	TuG1812G0700002220.01.P02	1.70E-18
TraesCS7A01G209300.1	TuG1812G0700002220.01.P02	4.80E-18
TraesCSU01G230600.1	TuG1812G0700002220.01.P02	5.00E-18
TraesCS7D01G211400.1	TuG1812G0700002220.01.P02	1.50E-17
TraesCS7D01G205200.1	TuG1812G0700002220.01.P02	1.10E-16

L'analyse des 488 séquences TaNAC protéiques par le prédicteur Phobius a montré que 18 protéines possèdent une hélice transmembranaire d'environ ~21 aa située dans leur partie C-terminale. Seize de ces séquences appartiennent à la sous-famille NAC-b, les 2 séquences restantes appartiennent à la sous-famille NAC-i. Une autre protéine (TraesCS3D01G329300.1) possède un peptide signal de 19 aa en position N-terminale. Nous n'avons cependant pas d'information concernant le compartiment auquel ce peptide pourrait adresser la protéine.

2. Structuration phylogénétique et motifs conservés chez les gènes TaNAC

Pour mieux comprendre les relations phylogénétiques entre les gènes TaNAC, nous avons généré un arbre phylogénétique basé sur l'alignement des 488 séquences TaNAC protéiques entières en utilisant la méthode du maximum de vraisemblance (Figure 45). Cet arbre phylogénétique a mis en évidence des similarités entre séquences et a pu apporter des renseignements sur les mécanismes évolutifs ayant permis le développement de cette famille.

Tout d'abord, cet arbre a montré un regroupement primaire des séquences sous forme de groupes d'homéologie. Sur la base de l'alignement des séquences, 231 gènes TaNAC appartenant aux génomes A, B et D ont été regroupés en 77 groupes d'homéologie complets dans l'arbre phylogénétique. Ce résultat représente 47% du total des gènes TaNAC. Quarante-huit autres groupes d'homéologie incomplets (ne comptant que 2 membres) ont été identifiés et représentent 20% du total des gènes TaNAC. Enfin, 161 séquences n'ont pu être rattachées à aucune séquence homéologue au sein de l'arbre.

La structuration de l'arbre phylogénétique en clades a aussi permis la création de plusieurs sous-familles au sein de la famille TaNAC. Par alignement avec les sous-familles NAC décrites par Christiansen *et al.* (2011) chez l'orge et par Shen *et al.* (2009) chez le riz, Borrill *et al.* (2017) avaient structuré la famille NAC de blé tendre en 8 sous-familles en fonction du regroupement des séquences entre les clades de l'arbre phylogénétique. Cette classification a été adaptée aux clades observés dans notre arbre phylogénétique. Les 488 séquences TaNAC ont donc été réparties en 11 sous-familles, de NAC-a à NAC-k. Les sous-familles NAC-i (14 membres), NAC-j (5 membres) et NAC-k (3 membres) ont été créées pour regrouper les séquences non rattachées aux clades principaux.

Tableau 12 : Séquences TaNAC et leurs homologues chez *T. urartu*, le riz et le sorgho. Les 21 séquences possèdent toutes le même homologue chez le riz et 18 d'entre elles montrent aussi un homologue identique chez le sorgho. De plus, certains de ces TaNAC sont ancrés côte à côte sur la pseudomolécule de blé tendre (un exemple de 5 gènes dont les identifiants se suivent, ancrés sur le chromosome 2B et surlignés en gras dans le tableau). Ces résultats appuient l'hypothèse d'un élargissement de la famille NAC chez le blé tendre par rapport aux autres céréales, par des mécanismes de duplications locales.

Séquences TaNAC	Meilleur Hit		
	<i>T. urartu</i>	Sorgho	Riz
TraesCS2A01G326500.1	TuG1812G0200003836.01.P01	SB04G024780	OS02G38130
TraesCS2B01G355400.1			
TraesCS2B01G355500.1			
TraesCS2D01G336400.1	TuG1812G0200003837.01.P01		
TraesCS2A01G326400.2			
TraesCS2B01G355700.1			
TraesCS2B01G355600.1			
TraesCS2D01G336500.1	TuG1812G0700000325.01.P01		
TraesCS2A01G325900.2			
TraesCS2A01G326600.1			
TraesCS2B01G355300.2			
TraesCS2D01G336300.1	TuG1812G0600002466.01.P02		
TraesCS6A01G208900.3			
TraesCS6B01G237900.3			
TraesCS6D01G192400.1			
TraesCS2D01G435800LC.2	TuG1812G0200003841.01.P01		
TraesCS2D01G436300LC.1			
TraesCS2A01G326300.2	TuG1812G0200003833.01.P01		
TraesCS7A01G208200.1	TuG1812G0700002206.01.P01	SB10G009670	
TraesCS7B01G115400.1			
TraesCS7D01G210600.1			

Au sein de cet arbre phylogénétique, nous avons constaté que certains clades regroupent des séquences présentant des similitudes de structure exon-intron. De plus, ces séquences sont ancrées en clusters sur la pseudomolécule du blé. Ce constat nous a conduit à considérer la possibilité que des phénomènes de duplication au sein de la famille TaNAC puissent être à l'origine de l'apparition de ces séquences proches physiquement (localisation chromosomique) et phylogénétiquement.

Par ailleurs, nous avons identifié 488 membres de la famille NAC chez le blé tendre, ce qui est un nombre particulièrement élevé par rapport à ce qui a déjà été observé chez d'autres céréales comme le riz (151 membres ; Nuruzzaman *et al.* 2010) ou le blé dur (168 membres ; Saidi *et al.* 2017). Cette grande différence entre le blé et les autres céréales est partiellement expliquée par la présence de 3 sous-génomes chez le blé tendre, mais pas uniquement. En effet, nous avons conduit une comparaison entre le génome du blé tendre et celui de *T. urartu*, un proche parent porteur du génome A^uA^u. Nous avons constaté, pour 51 séquences de *T. urartu*, qu'elles ressortent chacune comme le meilleur homologue de 3 séquences TaNAC chez le blé, ce qui est cohérent avec la ploïdie de ces 2 espèces (diploïde vs hexaploïde). Cependant, 70 séquences TaNAC ne possèdent pas d'homologue chez *T. urartu* et dans de nombreux cas, une séquence unique chez *T. urartu* est le meilleur homologue pour plus de 3 gènes chez le blé tendre. Par exemple, la séquence TuG1812G0700002220.01.P02 de *T. urartu* est l'homologue de 15 TaNAC chez le blé tendre (Tableau 11). Ces 15 séquences appartiennent à un clade commun, qui ne montre pas une structuration nette en groupes d'homéologie mais plutôt un regroupement en sous-génomes. Nous retrouvons notamment 5 gènes TaNAC physiquement voisins sur le chromosome 7A (TraesCS7A01G209100, TraesCS7A01G209200, TraesCS7A01G209300, TraesCS7A01G209400 et TraesCS7A01G209500). Ces 5 séquences pourraient donc être issues de duplications en tandem d'une séquence de départ, homologue au gène TuG1812G0700002220.01.P02 chez *T. urartu*. Du fait de cette redondance, au total seules 110 séquences différentes de *T. urartu* ont été identifiées comme meilleurs homologues des 488 séquences TaNAC. Le blé tendre étant composé de 3 sous-génomes, une famille composée de 330 gènes serait attendue chez cette espèce si on se réfère aux séquences obtenues pour *T. urartu*. Or, $488 - 330 = 158$ séquences TaNAC restent excédentaires.

Le nombre de TaNAC surnuméraires est encore plus élevé si l'on réalise ce même travail avec d'autres céréales comme le riz ou le sorgho (Tableau 12). En effet, dû au même phénomène de redondance, seules 85 et 83 séquences uniques de ces 2 espèces sont identifiées comme meilleurs hits des 488 séquences TaNAC. En conséquence, nous estimons que le nombre de

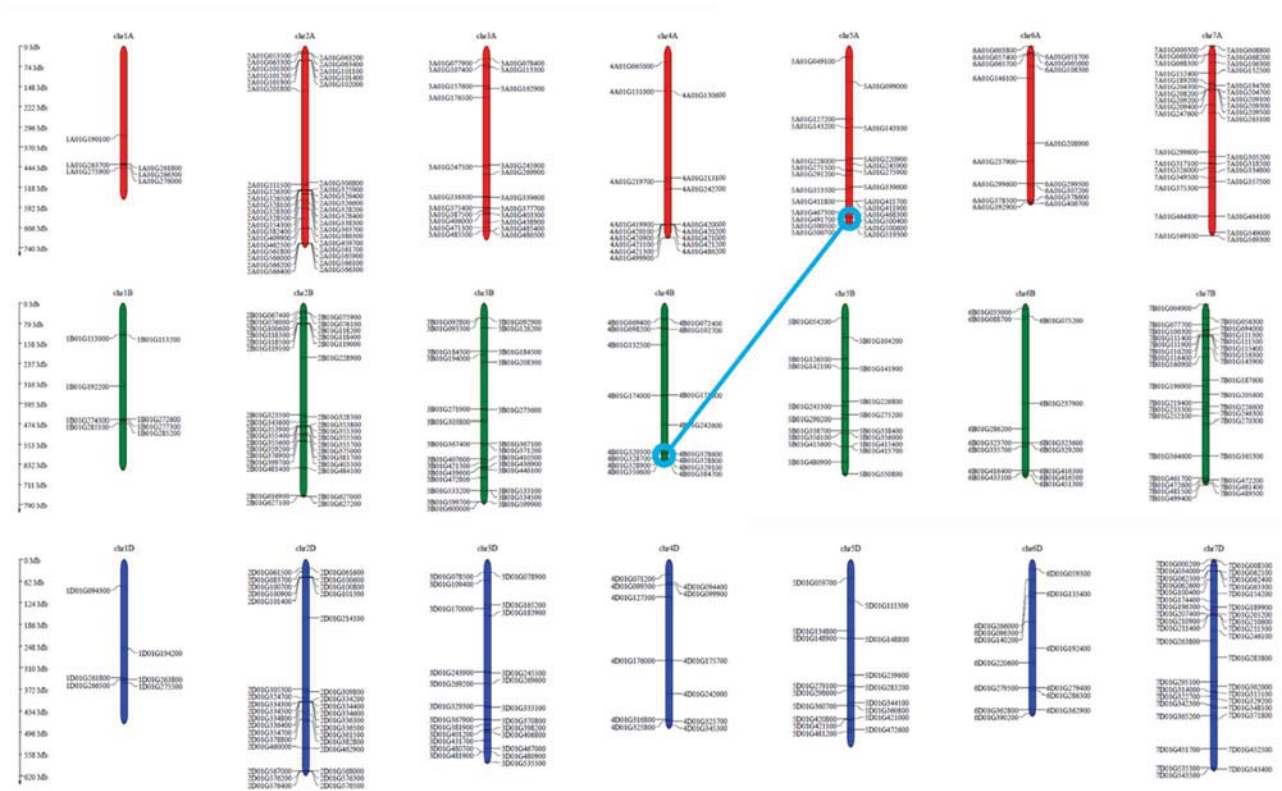


Figure 46 : Distribution des membres de la famille TaNAC le long des 21 chromosomes du blé tendre. En utilisant leurs informations de localisation chromosomique, 451 séquences ont été cartographiées sur les 7 chromosomes des 3 sous-génomes A (rouge), B (vert) et D (bleu). Les 7 premières lettres des noms de séquence ont été supprimées pour plus de lisibilité ; par exemple, 4B01G328600 fait référence à *TraesCS4B01G328600*, situé sur le chromosome 4B conformément à l'annotation IWGSC. La position de chaque TaNAC peut être estimée à l'aide des échelles à gauche. Les 2 cercles montrent les régions chromosomiques étudiées comme exemple de duplication.

gènes TaNAC chez le blé tendre est élevé par rapport aux autres espèces végétales proches, et que ceci serait dû en partie à la nature hexaploïde du blé mais serait aussi le résultat d'importants événements de duplication chez cette espèce.

3. Distribution génomique et duplication de la famille TaNAC

3.1. Distribution génomique

Parmi les 488 séquences TaNAC, 478 gènes ont été ancrés sur les 21 chromosomes du blé tendre ; les 10 dernières séquences ne présentent pas d'information de localisation chromosomique sur la base de données IWGS RefSeq v1.0. Les chromosomes 2 et 7 contiennent le plus grand nombre de gènes TaNAC avec respectivement 120 et 105 membres, ce qui représente 24,6% et 21,5% de la famille. Au contraire, le chromosome 1 est celui portant le moins de séquences TaNAC, avec seulement 21 gènes (4,3%) (Figure 46). Les gènes TaNAC sont répartis de manière homogène entre les sous-génomes A, B et D, qui portent 33,6%, 32,8% et 31,6% des membres de la famille respectivement, les séquences restantes ne sont pas ancrées sur un chromosome.

Les gènes TaNAC sont distribués de manière inégale le long de chacun des chromosomes. En effet, ils se situent majoritairement dans leurs régions distales (Figure 46). En outre, certains d'entre eux apparaissent comme structurés en clusters. Par exemple, TraesCS4B01G328600, TraesCS4B01G328700, TraesCS4B01G328800, TraesCS4B01G328900 et TraesCS4B01G329100 sont côte-à-côte sur 347,276 kb du chromosome 4B. De même, TraesCS4A01G419900, TraesCS4A01G420000, TraesCS4A01G420100, TraesCS4A01G420200, TraesCS4A01G420900, TraesCS4A01G421000, TraesCS4A01G421100, TraesCS4A01G421200 et TraesCS4A01G421300 sont aussi étroitement liés sur 796,023 kb du chromosome 4A, comme ils le sont sur l'arbre phylogénétique.

3.2. Duplications au sein de la famille TaNAC – Exemple de la région 4B-5A

Plusieurs gènes TaNAC appartenant à des chromosomes différents forment des clusters sur les chromosomes du blé tendre et sont également regroupés dans un même clade au sein de l'arbre

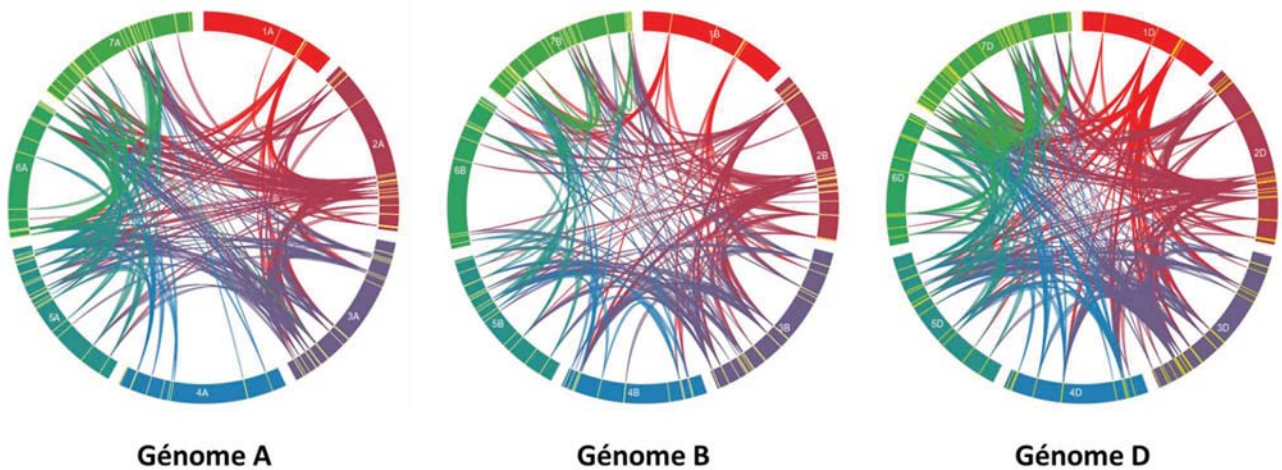


Figure 47 : Diagramme Circa des 7 chromosomes des génomes A, B et D. Chacun des 7 chromosomes est représenté par une portion d'une couleur différente sur le cercle. Les traits jaunes sur ces chromosomes indiquent la présence d'un gène TaNAC. Les bandes au centre relient 2 régions chromosomiques qui sont d'une longueur minimale de 3 kb et présentant plus de 90% d'identité.

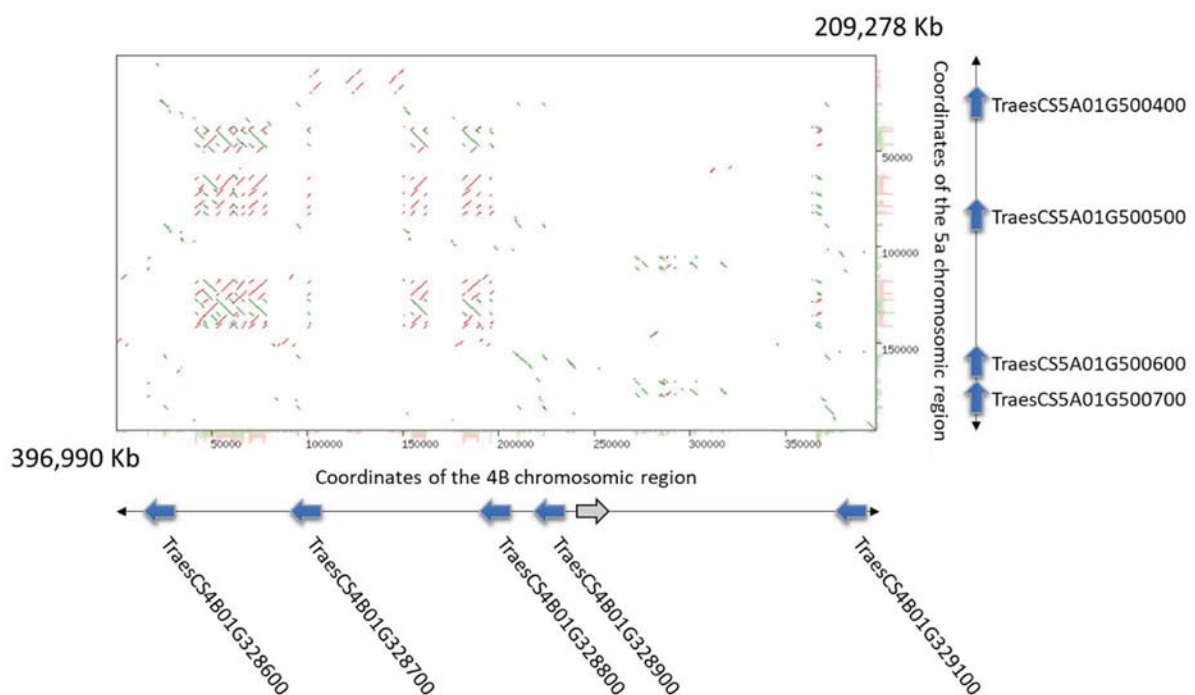


Figure 48 : Conservation structurelle entre les 2 régions génomiques des chromosomes 4B et 5A, révélée par une analyse DotPlot. La localisation, le sens et l'identifiant de chaque gène TaNAC ancré sur ces régions génomiques sont représentés sur les axes x et y par une flèche bleue. La flèche grise indique la présence d'un gène autre qu'un TaNAC. Les points verts et rouges symbolisent les homologies sens et anti-sens, respectivement.

phylogénétique. Ce constat suggère que l'expansion de la famille TaNAC pourrait être le résultat d'événements de duplication locaux et interchromosomiques. Pour tester cette hypothèse, nous avons souhaité en premier lieu avoir un aperçu des duplications existantes à l'échelle du génome entier. Pour cela, chacun des 3 sous-génomes a été aligné sur lui-même et les fragments présentant une identité d'au moins 90% sur un minimum de 3 kb ont été conservés. En utilisant ces critères,

- 318 fragments dupliqués ont été identifiés dans le sous-génome A, avec une longueur moyenne de 5,75 kb et une longueur médiane de 5,255 kb. Le plus gros fragment est de 13,093 kb,
- 207 fragments dupliqués ont été détectés dans le sous-génome B, avec une longueur moyenne égale à 5,546 kb, et une longueur médiane de 5,146 kb. Le plus grand fragment dupliqué est de 10,009 kb,
- 428 fragments dupliqués ont été détectés dans le sous-génome D, avec une longueur moyenne de 5,639 kb et une longueur médiane de 5,332 kb. Le plus grand fragment dupliqué est de 9,947 kb (Figure 47, le détail des données utilisées pour générer la figure est fourni en Supplemental Table S4 de la publication Guérin *et al.* (2019)).

Afin d'avoir des indications sur le schéma d'expansion de la famille TaNAC, nous nous sommes ensuite concentrés sur un clade en particulier, détaillé à droite de l'arbre dans la Figure 45. Ce clade a été choisi car il contient des gènes TaNAC portés par 3 régions génomiques ancrées sur chacun des 3 sous-génomes, et que les gènes étudiés sont tous voisins sur ces 3 régions génomiques. Ainsi, il nous permet d'étudier la présence de duplications au sein de chaque sous-génome et entre les sous-génomes.

Ces régions contiennent 11 gènes TaNAC ancrés sur les chromosomes 4B, 4D et 5A et physiquement étroitement liés (voisins ou séparés par un gène au maximum). Leurs séquences codantes sont structurellement proches puisqu'ils sont regroupés sur un clade de l'arbre phylogénétique, mais leur structure génomique est aussi similaire car ils possèdent tous 2 exons et 1 intron. Pour éviter les relations d'homéologie entre les sous-génomes pouvant masquer les duplications, le sous-génome 4D contenant 2 gènes TaNAC a été écarté de l'analyse. En utilisant une fenêtre de +/- 25 kb autour des gènes TaNAC les plus distaux de chacun des 2 locus restants, nous avons retenu un premier locus de 396,990 kb sur le sous-génome 4B (coordonnées sur la pseudomolécule : 619563787 à 619960777) et contenant 5 gènes TaNAC ;

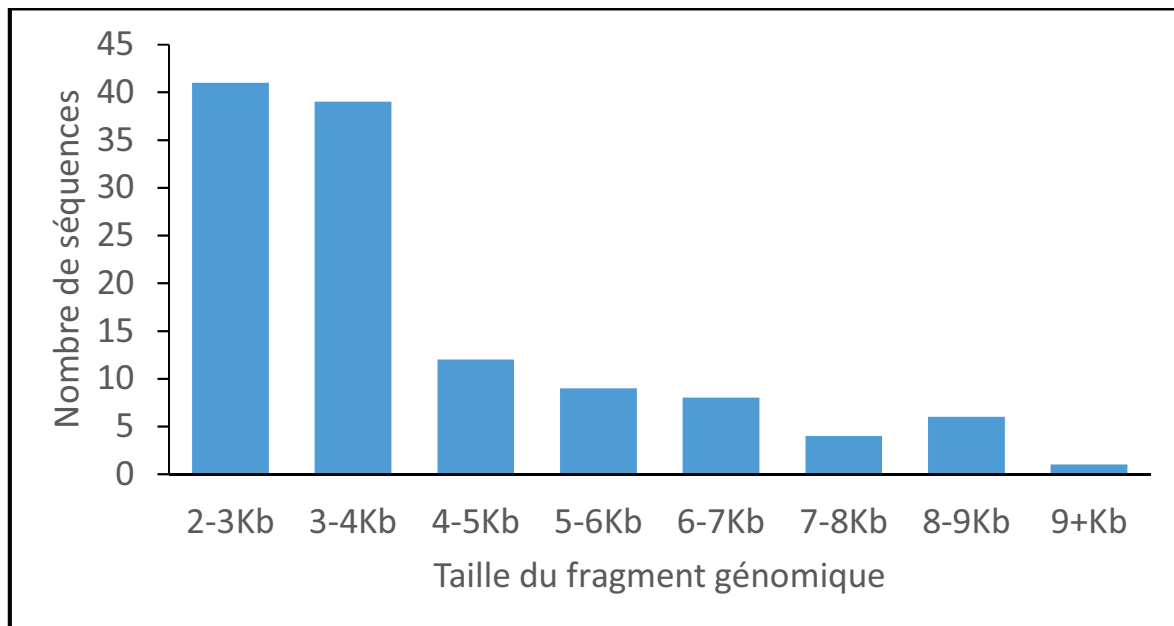


Figure 49 : Longueur des fragments génomiques (seuil minimum à 2 kb) présentant au moins 80% d'homologie entre eux, résultant du blast entre les régions chromosomiques 4B et 5A. La taille maximale détectée est de 9006 pb.

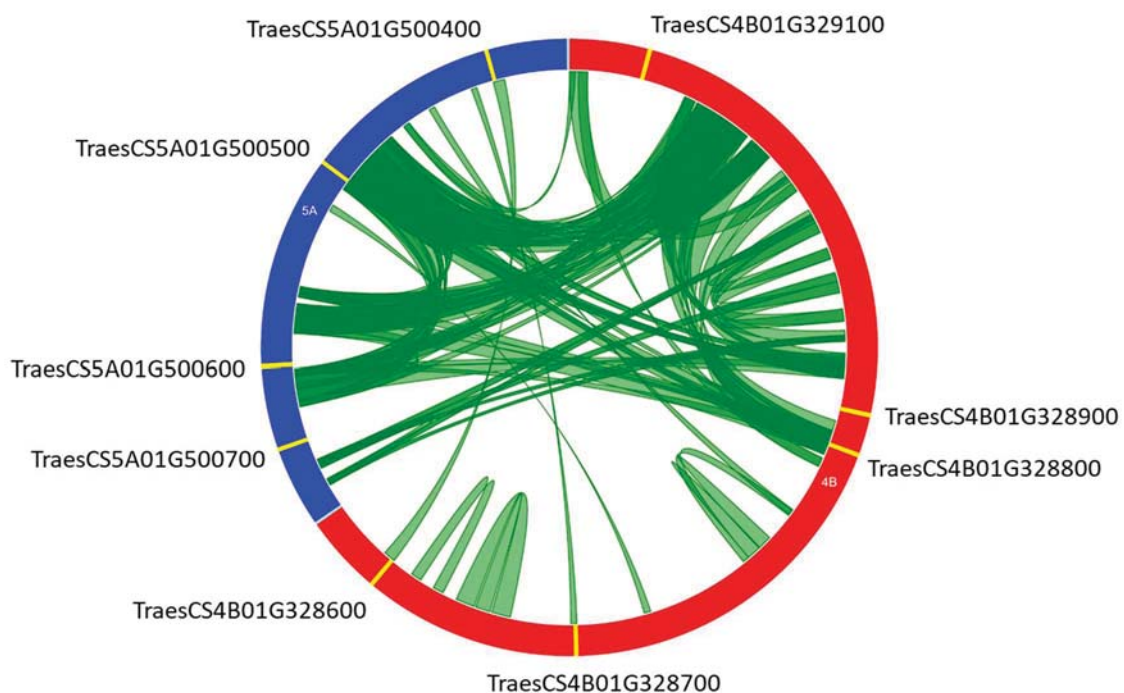


Figure 50 : Diagramme Circa des régions génomiques des chromosomes 4B (en rouge ; 396,990 kb) et 5A (en bleu ; 209,278 kb) contenant 9 gènes TaNAC (traits jaunes). Chaque bande verte relie 2 régions chromosomiques présentant plus de 80% d'identité sur une longueur minimale de 2 kb.

et un second locus de 209,278 kb sur le sous-génome 5A (coordonnées sur la pseudomolécule : 666499425 à 666708703) et contenant 4 gènes TaNAC.

Un dot-plot de ces 2 régions génomiques (Figure 48) a mis en évidence de nombreux fragments dupliqués en orientation sens (vert) et anti-sens (rouge). Pour clarifier la nature et l'amplitude des duplications, nous avons aligné ces séquences par paires en utilisant l'algorithme YASS (Noé et Kucherov 2005), ce qui nous a permis d'identifier tous les fragments dupliqués de ces régions, d'une longueur minimale de 2 kb et présentant au moins 80% d'identité entre eux. Un total de 119 fragments dupliqués non-chevauchants ont été identifiés, parmi lesquels 54 fragments correspondent à des duplications intra-chromosomiques (31 sur le chromosome 4B et 23 sur le chromosome 5A). Les 65 fragments dupliqués restants correspondent à des duplications interchromosomiques. La majorité de ces duplications (66,7%) possèdent une longueur inférieure à 4 kb, mais un fragment possède la taille maximale de 9006 pb (Figure 49). Les fragments dupliqués entre les loci de ces régions génomiques sont présentés sur la Figure 50. 41% de ces fragments (c'est-à-dire 49) sont dupliqués dans l'orientation sens et les fragments restants (c'est-à-dire 70) sont dupliqués en orientation anti-sens (Supplemental Table S5 de la publication Guérin *et al.* (2019)).

3.3. Devenir des gènes TaNAC dupliqués

Depuis la publication du livre " Evolution by Gene Duplication " par Ohno (1970), de nombreuses études ont montré le rôle des duplications dans la création de variants génétiques pouvant conférer des capacités d'adaptation à l'espèce. Pour tester si des gènes issus d'une duplication peuvent évoluer différemment, nous avons comparé les profils d'expression normalisés des gènes dupliqués choisis en exemple, et ce dans 5 organes de blé au cours de 3 stades de développement (racine, tige, feuille, épi et grain ; Choulet *et al.* 2014). En parallèle, les profils d'expression de ces mêmes gènes ont été comparés en réponse à des stress abiotiques (choc thermique, sécheresse et une combinaison de ces 2 stress ; Liu *et al.* 2015). Seuls les gènes significativement exprimés dans les 2 études ont été retenus pour comparaison. Afin d'évaluer la corrélation entre les profils d'expression *in silico* de ces gènes TaNAC dupliqués, le coefficient de corrélation de Pearson a été calculé entre chaque paire de ces gènes à l'aide du logiciel R (<http://www.R-project.org>, R development Core Team, 2008). Au cours du développement, ces gènes présentent un coefficient de Pearson allant de 0,41 (entre

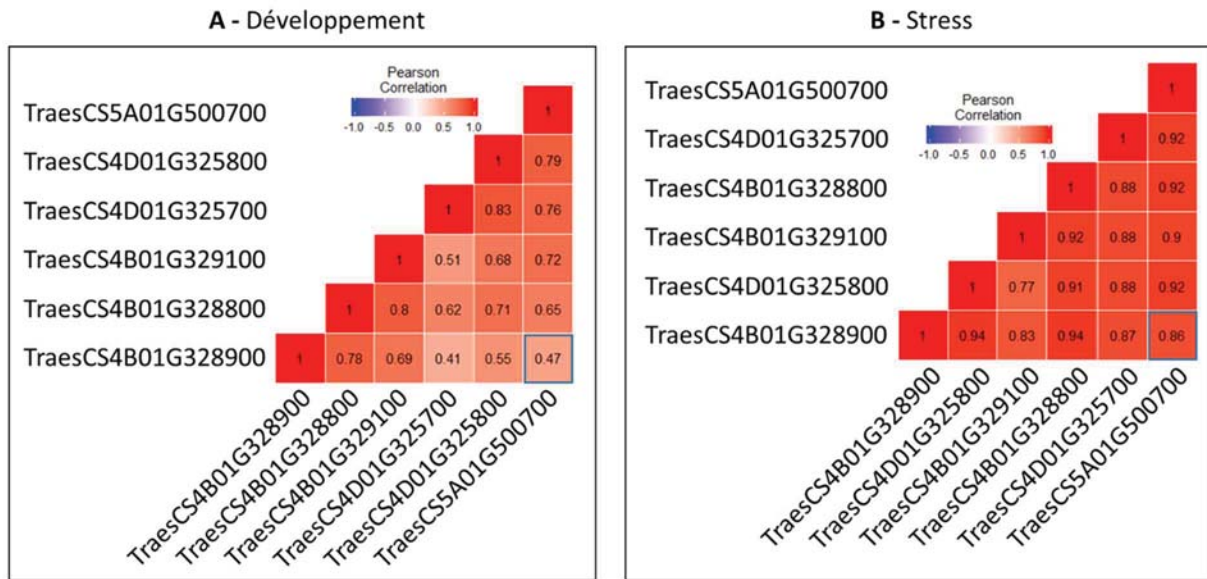


Figure 51 : Coefficient de corrélation de Pearson calculé entre les gènes dupliqués des régions chromosomiques 4B, 4D et 5A durant le développement (A) et en réponse aux stress abiotiques (B). Le comportement des 2 paralogues TraesCS5A01G500700 et TraesCS4B01G328900 est illustré en exemple (cadre bleu).

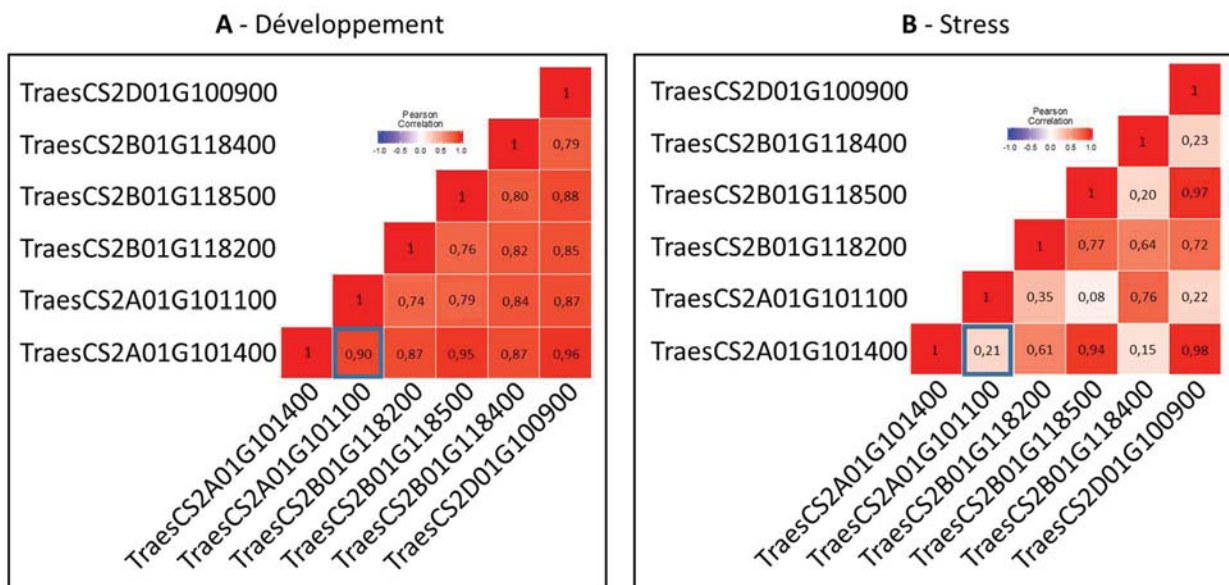


Figure 52 : Coefficient de corrélation de Pearson calculé entre les gènes dupliqués des régions chromosomiques 2A, 2B et 2D durant le développement (A) et en réponse aux stress abiotiques (B). Le comportement des 2 paralogues TraesCS2A01G101400 et TraesCS2A01G101100 est illustré en exemple (cadre bleu).

TraesCS4D01G325700 et TraesCS4B01G328900) à 0,83 (entre TraesCS4D01G325800 et TraesCS4D01G325700). En revanche, les corrélations sont plus élevées en réponse aux stress, leurs valeurs variant de 0,77 à 0,94 (Figure 51). Collectivement, ces résultats indiquent que les membres de cette duplication présentent des profils d'expression similaires en réponse aux stress abiotiques testés, mais cette corrélation entre gènes est moins importante pendant le développement. Ce constat peut être illustré par le comportement des 2 paralogues TraesCS5A01G500700 et TraesCS4B01G328900 (encadrés sur la Figure 51).

De manière similaire, l'examen d'une seconde duplication (en haut dans la Figure 45) est présenté dans la Figure 52. Dans ce cas, les profils d'expression des paralogues sont fortement corrélés pendant le développement, alors qu'ils le sont beaucoup moins en réponse aux stress abiotiques. Pour exemple, la valeur de corrélation minimale en réponse aux stress abiotiques est de 0,08 entre les gènes TraesCS2B01G118500 et TraesCS2A01G101100.

4. Diversité allélique au sein de la famille TaNAC

4.1. Selon les données de la puce BW_420K

La localisation sur la pseudomolécule des 420 000 marqueurs identifiés avec la puce BW_420K est connue. Par ce biais, seuls les marqueurs contenus dans la séquence génomique des 488 TaNAC et des 1500 pb en amont et en aval de chaque gène ont été conservés. Ceci a permis de mettre en évidence 494 marqueurs de polymorphisme, répartis sur 158 gènes TaNAC soit 32,4% des membres de cette famille. Ce polymorphisme est présent dans la séquence génomique et/ou dans les 1500 pb de part et d'autre de cette séquence. Entre 1 et 14 marqueurs ont été identifiés par gène pour ces 158 séquences (Tableau 13).

4.2. Selon les données de la base de données Whealbi

Cette analyse a révélé 1358 marqueurs polymorphes répartis sur 174 séquences TaNAC (soit 35,7% des membres de la famille), qui contiennent entre 1 et 79 marqueurs par gène (Tableau 14). Seule la moitié des gènes existants a été capturée pour la recherche de polymorphisme. Par conséquent, les gènes qui ne présentent pas de variabilité lors de cette analyse peuvent être monomorphes pour les lignées étudiées, ou simplement non-capturés.

Tableau 13 : Nombre de SNP identifiés, à partir des données de la puce BW 420K, dans les exons et les introns de la séquence génomique des 478 TaNAC, et dans les 1500 pb en amont (5' UTR) et en aval (3' UTR). Tous les gènes n'ont pas montré de polymorphisme, la dernière ligne du tableau recense donc les gènes concernés parmi les 488 gènes testés.

BW_420K	5'UTR	Exon	Intron	3'UTR	TOTAL
Nombre de SNP	99 (20%)	108 (22%)	130 (26%)	157 (32%)	494
Nombre min - max de SNP par gène	1 à 5	1 à 6	1 à 6	1 à 5	1 à 14
Nombre de gènes TaNAC concernés	57	53	66	91	158

Tableau 14 : Nombre de SNP identifiés, à partir des données de Whealbi, dans les exons et les introns de la séquence génomique des 478 TaNAC, et dans les 1500 pb en amont (5' UTR) et en aval (3' UTR).. Tous les gènes n'ont pas montré de polymorphisme, la dernière ligne du tableau recense donc les gènes concernés parmi les 488 gènes testés. Une séquence présente 64 SNP dans sa séquence intronique, ce qui est un nombre particulièrement élevé sachant que la seconde n'en possède que 25.

Whealbi	5'UTR	Exon	Intron	3'UTR	TOTAL
Nombre de SNP	305 (23%)	380 (28%)	332 (24%)	341 (25%)	1358
Nombre min - max de SNP par gène	1 à 32	1 à 29	1 à 64	1 à 26	1 à 79
Nombre de gènes TaNAC concernés	65	98	80	81	174

Tableau 15 : Nombre de SNP identifiés, à partir des données de BW 420K et Whealbi, sans doublon, dans les exons et les introns de la séquence génomique des 478 TaNAC, et dans les 1500 pb en amont (5' UTR) et en aval (3' UTR). La longueur concernée correspond au cumul des longueurs d'UTR, d'exon ou d'intron de tous les gènes (par exemple, pour les 5'UTR, comme 1500 pb sont prises en compte pour chacun des 478 gènes, la longueur concernée = 1500 pb * 478 gènes = 717 000 pb au total). La dernière ligne du tableau recense le nombre de gènes ayant montré du polymorphisme parmi les 488 gènes testés.

BW_420K + Whealbi sans doublon	5'UTR	Exon	Intron	3'UTR	TOTAL
Nombre de SNP	382 (22%)	461 (27%)	422 (24%)	460 (27%)	1725
Longueur concernée (pb)	717 000	528 864	428 103	717 000	2390967
Nombre min - max de SNP par gène	1 à 32	1 à 30	1 à 64	1 à 26	1 à 79
Nombre de gènes TaNAC concernés	105	125	111	136	238

Les gènes TraesCS5D01G360700 et TraesCS5D01G360800 ne sont séparés que de 2449 pb, le SNP identifié dans les 1500 pb respectivement en position 3' et 5' de chacun de ces 2 gènes est donc le même, à la base 440369017 de la pseudomolécule. Les 2 identifications de ce SNP unique ont été conservées pour l'analyse d'association car il n'est pas possible de déterminer dans quelle mesure il participe à la régulation de l'expression de l'un ou l'autre des 2 gènes.

4.3. Compilation des données de variabilité obtenues avec BW_420K et avec Whealbi.

Certains marqueurs sont présents dans les 2 bases de données, additionner les résultats obtenus pour BW_420K et Whealbi mènerait donc à une surestimation du polymorphisme au sein de cette famille. Pour éviter ce problème, une compilation sans doublon des marqueurs identifiés a été réalisée, ce qui a permis d'analyser la variabilité « totale » identifiée pour cette famille (Tableau 15). Nous pouvons constater que BW_420K a fourni un nombre plus restreint de SNP que Whealbi, ce qui était attendu compte tenu des différentes technologies utilisées.

La distribution des SNP entre les exons, les introns, les 5' et 3' UTRs, est relativement constante et équitable : environ 25% des SNP sont identifiés dans chacune des régions génomiques susnommées. Cependant, leur longueur n'est pas équivalente. Le Tableau 16 présente le nombre de SNP identifiés par 1000 pb de chacune des 4 régions étudiées. Nous pouvons constater que les introns concentrent le maximum de variabilité avec 0,99 SNP / 1000 pb, tandis que les UTRs ne présentent que 0,53 à 0,64 SNP / 1000 pb. Lorsque cette valeur est calculée pour chacun des 478 gènes ancrés sur un chromosome, le gène TraesCS7D01G543400 présente la plus forte variabilité dans les exons, avec 24,4 SNP / 1000 pb codantes.

5. Association polymorphisme x phénotypes

5.1. SNP et composition protéique du grain

L'association des SNP identifiés précédemment avec les données de phénotypage de la Core Collection Agronomique cultivée en 3 environnements différents (Annexe V et Tableau 17) a été jugée significative lorsque le score était ≥ 3 . Plusieurs marqueurs polymorphes génotypés avec la puce BW_420K, dont 10 marqueurs provenant d'OTV et notés en Présence/Absence, montrent une association significative avec des caractères de composition protéique. Six

Tableau 16 : Récapitulatif du nombre de SNP et de la longueur totale des régions 5'UTR, exon, intron et 3'UTR des 478 TaNAC ancrés sur la pseudomolécule de blé tendre. La dernière colonne présente le nombre moyen de SNP identifié dans 1000 pb de chacune des régions susdites.

	Nombre de SNP	Longueur totale (pb)	Nombre de SNP par 1000 pb
5' UTR	382	717000	0,53
Exons	461	528864	0,87
Introns	422	428103	0,99
3' UTR	460	717000	0,64

Tableau 17 : Résultats d'association présentant un score ≥ 3 entre les SNP identifiés par la capture d'exome de Whealbi et génotypés par séquençage, et le phénotypage de la Core Collection Agronomique cultivée en 3 environnements différents. Les mesures ont été réalisées dans 3 environnements : Clermont-ferrand (CF), Le Moulon fort azote (LMHN) et Le Moulon faible azote (LMLN). Le gène portant le plus de marqueurs associés à un phénotype est surligné en jaune. Les autres couleurs sont des repères visuels pour les différents caractères phénotypiques identifiés.

Phénotype	Environnement	Gène	Chromosome	Position	Score
Pourcentage de gluténines dans l'azote total	LMLN	TraesCS2B01G100600	chr2B	61270118	4.817
Pourcentage de gluténines dans l'azote total	LMHN	TraesCS6A01G378600	chr6A	599716001	4.000
Quantité d' α/β -gliadines par grain	LMHN	TraesCS6A01G051700	chr6A	26934781	4.258
Quantité d' α/β -gliadines par grain	LMHN	TraesCS6A01G051700	chr6A	26934650	4.600
Quantité de γ -gliadines par grain	LMHN	TraesCS6A01G051700	chr6A	26935047	4.189
Quantité de γ -gliadines par grain	LMHN	TraesCS6A01G051700	chr6A	26934781	4.399
Quantité de γ -gliadines par grain	LMHN	TraesCS6A01G051700	chr6A	26934650	4.244
Quantité de protéines de réserve par grain	LMHN	TraesCS6A01G051700	chr6A	26935047	4.366
Quantité de protéines de réserve par grain	LMHN	TraesCS6A01G051700	chr6A	26934781	4.584
Quantité de protéines de réserve par grain	LMHN	TraesCS6A01G051700	chr6A	26934650	4.376
Quantité de gliadines par grain	LMHN	TraesCS6A01G051700	chr6A	26935047	4.096
Quantité de gliadines par grain	LMHN	TraesCS6A01G051700	chr6A	26934781	4.504
Quantité de gliadines par grain	LMHN	TraesCS6A01G051700	chr6A	26934650	4.674
Quantité de gliadines par grain	LMHN	TraesCS7B01G111300	chr7B	127651143	4.233

marqueurs génotypés dans le cadre du projet Whealbi montrent également une association significative avec des caractères de composition protéique. Toutes les associations significatives entre caractères de composition protéique et marqueurs Whealbi ont été obtenues sur le site du Moulon et presque toutes (sauf une) ont été obtenues en condition de fort apport en azote (Tableau 17). Parmi ceux-ci, un SNP et les marqueurs notés en Présence/Absence de 3 OTV dans le gène TraesCS2A01G561800 sont significativement associés avec de nombreuses variables de composition en protéines de réserve telles que :

- le pourcentage d' α/β -gliadine parmi les gliadines, à Clermont Ferrand et le Moulon en condition fort azote,
- la quantité d' α/β -gliadine par grain au Moulon en condition fort azote,
- le pourcentage de γ -gliadine parmi les gliadines, à Clermont Ferrand et le Moulon en condition fort azote,
- le ratio gliadine/gluténine, intéressant pour la technologie,
- le pourcentage de gliadine parmi les protéines de réserve, exprimé en pourcentage de l'azote totale,
- le pourcentage d' α/β -gliadine parmi les protéines de réserve, à Clermont-Ferrand et le Moulon en condition fort ou faible azote,
- le pourcentage des gluténines de faible poids moléculaire parmi les protéines de réserve, exprimé en pourcentage de l'azote totale, au Moulon en condition fort azote.

Ce gène semble donc intervenir dans l'accumulation des formes α/β et γ -gliadines en condition de fort azote, ce qui impacte le ratio gliadine/gluténine et le pourcentage de gliadine dans les protéines de réserve. Cette association semble extrêmement forte, le score atteignant 6,15 pour le SNP en position 763314614 avec le pourcentage d' α/β -gliadines parmi les protéines de réserve au Moulon en condition fort azote.

De manière similaire, le gène TraesCS6A01G051700 porte 5 SNP génotypés par la puce BW_420K et 3 SNP identifiés par Whealbi. Sept parmi ces 8 SNP présentent une association significative avec des paramètres d'accumulation des α/β et γ -gliadines dans le grain en condition fort azote, avec une répercussion sur la quantité de protéines de réserve dans le grain (Figure 53).

La limite du faible nombre de lignées pour les résultats observés sur les 106 lignées de l'intersection Core Collection Agronomique x Whealbi, est levée par la concordance de ces



Figure 53 : Structure et SNP de la séquence génomique de TraesCS6A01G051700. Cette séquence génomique ne possède pas d'intron. Cinq SNP ont été identifiés avec les données de la puce BW_420K, dans les 1500 pb en amont de la séquence codante, et un SNP dans la séquence codante (traits jaunes ou gris). Trois SNP (en orange) ont été identifiés par les données Whealbi. Les SNP jaunes et oranges présentent une association significative avec des paramètres d'accumulation d' α/β -gliadines dans le grain de blé, le SNP gris ne présente pas d'association.

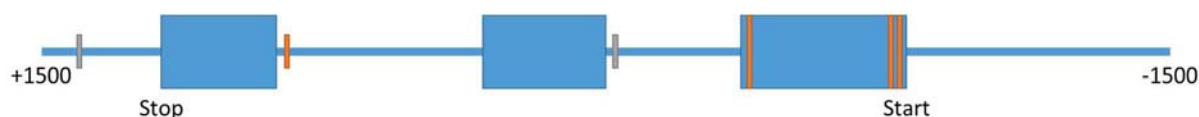


Figure 54 : Structure et SNP de la séquence génomique de TraesCS6A01G307200. Cette séquence génomique possède 2 introns, 6 SNP ont été identifiés sur cette séquences avec les données Whealbi (traits oranges ou gris). Quatre SNP montrent une association significative avec la dureté du grain (traits oranges). Les SNP en gris ne montrent pas d'association

Tableau 18 : Données d'association entre les SNP identifiés à partir des données Whealbi et les caractères technologiques mesurés par Bordes et al. (2011) sur la Core Collection Agronomique de 196 lignées. Seule la dureté du grain a montré une association significative avec des SNP, tous réunis sur le gène TraesCS6A01G307200.

Phénotype	Gène	Chromosome	Position du SNP	Score
Dureté	TraesCS6A01G307200	6A	541237368	4.423
Dureté	TraesCS6A01G307200	6A	541237404	4.423
Dureté	TraesCS6A01G307200	6A	541237945	4.423
Dureté	TraesCS6A01G307200	6A	541239513	4.423

résultats avec ceux obtenus de la puce BW_420K, sur les 196 lignées de la Core Collection Agronomique complète.

5.2. SNP et caractères technologiques

Seul un gène porteur de marqueurs polymorphes identifiés dans la base de donnée Whealbi est significativement associé à la dureté du grain. Ce gène, TraesCS6A01G307200, est localisé sur le chromosome 6A et structuré en 3 exons. Il regroupe 6 marqueurs dont 4 (en positions 541237368, 541237404, 541237945 et 541239513 sur la pseudomolécule) sont liés à la dureté du grain (Figure 54 ; Tableau 18).

Certains SNP identifiés par la puce BW_420K ont montré une association avec la concentration en protéines du grain (Tableau 19). Nous pouvons remarquer parmi eux que les gènes TraesCS4A01G419900, TraesCS4A01G420000 et TraesCS4A01G420100 sont ancrés côte à côte sur le chromosome 4A. Ces gènes pourraient appartenir à un QTL (Quantitative Trait Locus) de régulation de l'accumulation de protéines dans le grain.

5.3. SNP et caractères agronomiques

Les données de polymorphismes obtenues avec la base de données Whealbi n'ont pas montré d'association avec les caractères agronomiques mesurés. Cependant, les SNP identifiés à partir des données de la puce BW_420K ont montré une association entre 7 gènes TaNAC et le PMG mesuré par Bordes *et al.* (2013) (Tableau 20). Parmi eux, un SNP ancré dans la région 3'UTR du gène TraesCS6B01G237900 montre une association très forte avec le PMG (score de 6,9). Ce SNP a aussi été identifié comme associé à tous les paramètres de taille du grain mesurés (PMG, surface, longueur, largeur) (Tableau 21). C'est d'ailleurs le seul à présenter une association supérieure au seuil limite pour tous les paramètres mesurés, et c'est celui qui montre l'association la plus forte dans tous les cas. La séquence génomique de ce gène est porteuse d'un autre SNP, identifié dans une région intronique. Celui-ci ne présente pas d'association avec des caractères d'intérêt. Nous pouvons donc supposer que ce gène pourrait jouer un rôle majeur dans l'établissement de la taille et du poids du grain, par des mécanismes de régulation de son expression qui se joueraient dans sa partie 3'UTR (qui joueraient par exemple sur la stabilité de l'ARNm).

Tableau 19 : Données d'association entre les SNP identifiés à partir de la puce BW 420K et les caractères technologiques mesurés par Bordes et al. (2011) sur la Core Collection Agronomique de 196 lignées. Seule la concentration du grain en protéines a montré une association significative avec des SNP.

Phénotype	Gène	Chromosome	Position	Score
Protéines	TraesCS3A01G107400	3A	72293906	3.372
Protéines	TraesCS3A01G107400	3A	72294740	3.276
Protéines	TraesCS4A01G419900	4A	690689901	3.117
Protéines	TraesCS4A01G420000	4A	690878437	3.385
Protéines	TraesCS4A01G420100	4A	690886482	4.028

Tableau 20 : Données d'association entre les SNP identifiés à partir de la puce BW 420K et les caractères agronomiques mesurés par Bordes et al. (2013) sur la Core Collection Agronomique de 196 lignées.

Phénotype	Gène	Chromosome	Position	Score
PMG	TraesCS6B01G237900	6B	413939874	6,896
PMG	TraesCS1A01G266300	1A	461778087	5,672
PMG	TraesCS7B01G365300	7B	628734488	4,767
PMG	TraesCS3B01G421300	3B	658464386	4,520
PMG	TraesCS6A01G378600	6A	599715266	3,573
PMG	TraesCS2A01G566400	2A	765548774	3,492
PMG	TraesCS2A01G566100	2A	765392038	3,363
PMG	TraesCS2A01G566400	2A	765548771	3,218
PMG	TraesCS2A01G566400	2A	765548648	3,176

Tableau 21 : Données d'association entre les SNP identifiés à partir de la puce BW 420K et les paramètres de taille du grain (surface, longueur, largeur) et poids de milles grains (PMG) mesurés au sein de l'équipe BIANCA sur la Core Collection Agronomique et sur les lignées de Whealbi. Les valeurs en vert sont supérieures au score seuil d'association.

Gène	Chromosome	Position	Score			
			PMG	Surface	Longueur	Largeur
TraesCS6B01G237900	6B	413939874	7.224	5.258	4.560	5.459
TraesCS3B01G421300	3B	658464386	4.128	3.085	1.369	4.182
TraesCS7A01G247600	7A	228262191	3.991	3.581	2.704	3.557
TraesCS2A01G566400	2A	765548774	3.302	3.675	2.755	2.959
TraesCS1A01G266300	1A	461778087	3.285	2.072	0.788	3.532
TraesCS2A01G566400	2A	765548771	2.672	3.172	2.856	1.949
TraesCS2A01G561800	2A	763314613	2.560	2.375	0.971	3.477
TraesCS2D01G567000	2D	635782907	1.838	2.007	0.450	3.365
TraesCS2A01G566400	2A	765548648	2.869	2.649	1.319	3.159

DISCUSSION du CHAPITRE I

Avant-propos :

Récemment, le séquençage du génome complet du blé tendre (IWGS RefSeq v1.0) ainsi que des données d'expression concernant différents génotypes, organes et stades de développement du blé soumis à différentes conditions de culture (expVIP) ont été rendus publics. Ces ressources sont une mine d'informations pour le recensement des gènes et l'analyse de leur régulation transcriptionnelle, et elles sont utiles pour l'identification de gènes candidats impliqués dans le développement du blé tendre ou sa réponse à différents stress.

Parmi les facteurs de transcription, la famille NAC est l'une des plus grandes familles, et elle est étudiée dans pléthore de céréales (Lu *et al.* 2015 ; Zhu *et al.* 2015a ; Saidi *et al.* 2017), au cours du développement de la plante (Distelfeld *et al.* 2012 ; Fan *et al.* 2018b ; Guo *et al.* 2018) et lors de la réponse à différents types de stress abiotiques (Baloglu *et al.* 2012, 2014 ; Mao *et al.* 2014). Cependant, peu d'informations sont disponibles sur cette famille chez le blé tendre.

Récemment, Borrill *et al.* (2017) ont identifié 453 gènes TaNAC avec un haut niveau de confiance, collectés à partir de la base de données TGAC (Clavijo *et al.* 2017). Cette version de l'assemblage du génome du blé tendre peut être considérée comme préliminaire. Par ailleurs, des données RNAseq ont été rendues publiques ; et plus particulièrement des données d'expression issues de différents organes du blé tendre (Choulet *et al.* 2014) et en réponse à plusieurs stress abiotiques (comme le choc thermique et la sécheresse ; Liu *et al.* 2015). Sur la base de ces données, nous avons étudié l'organisation de la famille NAC chez le blé tendre et apporté des hypothèses sur les mécanismes potentiels de l'expansion de cette famille au cours de son évolution, notamment *via* des événements de duplication de petite taille.

6. Les gènes TaNAC, une grande famille

Dans cette thèse, 488 membres de la famille NAC ont été identifiés chez le blé tendre en utilisant la base de données IWGS RefSeq v1.0 qui est la plus récente disponible. La confrontation de ces gènes TaNAC avec les séquences identifiées par Borrill *et al.* (2017) et issues de la base de données TGAC a permis la création d'un sous-ensemble consensus

contenant 361 séquences strictement identiques entre les 2 bases de données. Parmi les 127 séquences restantes, la majorité existe dans les 2 bases de données mais présente un découpage exon/intron différent qui modifie le CDS et les exclue de fait du sous-ensemble consensus.

Les séquences TaNAC contiennent entre 1 et 16 exons. Nuruzzaman *et al.* (2010) ont recensé les membres de la famille NAC chez le riz et ont montré qu'ils possèdent entre 1 et 17 exons. Le nombre maximal d'exons semble cependant être plus faible chez d'autres espèces comme le soja dont les NAC possèdent entre 2 et 8 exons (Hussain *et al.* 2017) ou la banane dans laquelle ils possèdent 1 à 7 exons (Cenci *et al.* 2014). La majorité des TaNAC possèdent 3 exons, ce résultat a déjà été observé chez la famille NAC du sarrasin de Tartarie (Liu *et al.* 2019).

Comme le blé tendre est une plante hexaploïde, une copie de chaque gène est théoriquement portée par chacun des 3 sous-génomes (A, B et D), créant des triplets d'homéologie. Parmi les 488 TaNAC, 47% des gènes ont pu être regroupés dans des groupes d'homéologie complets (un clade distinct de l'arbre phylogénétique, contenant 3 gènes localisés sur les sous-génomes A, B et D du même chromosome), 20% appartiennent à des groupes d'homéologie incomplets (ne contenant que 2 membres sur les 3 attendus) et 33% sont des séquences seules. La présence de groupes d'homéologie incomplets et de séquences seules au sein de la famille TaNAC peut être expliquée par une annotation encore imparfaite du génome qui n'a pas permis d'identifier tous les gènes présents chez le blé tendre. Une autre raison possible est la divergence des séquences d'un même groupe d'homéologie au cours de l'évolution, par exemple par accumulation de mutations. Ceci a déjà été observé pour d'autres familles de facteurs de transcription, comme la famille bZIP (Li *et al.* 2016a). Enfin, une troisième possibilité serait l'absence de certains homéologues suite à des délétions.

Shen *et al.* (2009) ont proposé une classification des NAC en 8 sous-familles nommées de NAC-a à NAC-h, et ils ont suggéré que ces sous-familles regroupaient des gènes aux fonctions différentes. La création de ces sous-familles a été réalisée sur la base d'un arbre phylogénétique issu de l'alignement des domaines NAC uniquement, et Borrill *et al.* (2017) ont utilisé le même protocole pour l'identification des sous-familles chez le blé. Notre arbre phylogénétique n'est pas basé sur le domaine NAC uniquement mais sur la longueur totale des séquences protéiques. Les sous-familles TaNAC identifiées par Borrill *et al.* (2017) et construites à partir de celles dessinées par Shen *et al.* (2009) ont été adaptées à notre arbre phylogénétique. Ceci a mené à la création de 3 sous-familles supplémentaires correspondant à des clades de séquences non référencées par ces auteurs. Cette adaptation n'a que très peu modifié les sous-familles NAC

Tableau 22 : Nombre de séquences constituant chacune des 11 sous-familles TaNAC, ainsi que la longueur moyenne des CDS contenus dans ces sous-familles.

Sous-famille	Nombre de membres	Longueur moyenne du CDS (pb)
NAC-a	42	1022
NAC-b	44	1384
NAC-c	54	1057
NAC-d	79	988
NAC-e	54	1097
NAC-f	12	707
NAC-g	51	1267
NAC-h	130	1127
NAC-i	14	967
NAC-j	5	1339
NAC-k	3	1179

de blé tendre dessinées par Borrill *et al.* (2017), des divergences ont été observées uniquement pour 21 gènes, assignés par Borrill *et al.* (2017) à une sous-famille qui n'était pas rattachée au clade adéquat dans notre arbre. Dix-huit de ces gènes ont été assignés à l'une des 3 nouvelles sous-familles (NAC-i, NAC-j ou NAC-k) dans notre classification. Les 3 autres ont été changés de sous-famille. La répartition des TaNAC dans les 11 sous-familles n'est pas équilibrée (Tableau 22). La sous-famille NAC-h regroupe le plus grand nombre de séquences (130). Les séquences CDS de la sous-famille NAC-b sont en moyenne les plus longues (1384 pb) tandis que NAC-f regroupe les séquences les plus courtes (707 pb en moyenne).

Les gènes TaNAC ne sont pas également distribués entre les 7 chromosomes, ce qui concorde avec des observations réalisées chez d'autres espèces végétales comme le maïs (Peng *et al.* 2015) et le riz (Nuruzzaman *et al.* 2010). Plus récemment, Murozuka *et al.* (2018), en recensant la famille NAC chez l'orge, ont observé une accumulation préférentielle des gènes NAC sur les chromosomes 2 et 7, comme pour le blé tendre. Nous avons aussi observé que les 3 sous-génomes du blé tendre portent la même proportion de gènes TaNAC, ce qui est cohérent avec la distribution observée chez le blé dur tétraploïde (Saidi *et al.* 2017). Ce résultat suggère que les gènes TaNAC auraient été présents dans le génome ancestral des sous-génomes A, B et D.

Les 8 gènes présentant 2 domaines NAC correspondent à 2 groupes d'homéologie complets et 1 groupe incomplet. Ils ne sont pas regroupés sur un clade à part dans l'arbre phylogénétique et leur séquence génomique ne comporte aucun intron, ce qui écarte l'hypothèse d'une erreur de découpage exon/intron. Une autre possibilité est que ces TaNAC pourraient avoir été créés par rétroposition d'un gène TaNAC rétro-transcrit au milieu de la séquence génomique d'un autre TaNAC, menant à la création d'un gène sans intron mais avec 2 domaines NAC. Chez *Arabidopsis*, Olsen *et al.* (2005) ont montré l'existence de 5 gènes présentant 2 domaines NAC. Cependant ils ne semblent pas homologues aux gènes que nous avons identifiés.

Dix-huit gènes TaNAC ont été identifiés avec un domaine transmembranaire dans la partie C-terminale de leur séquence protéique (maximum 12 aa avant le stop). Parmi ces séquences, 16 appartiennent à la sous-famille NAC-b et 2 à la sous-famille NAC-i. Les 3 séquences TraesCS7B01G196900, TraesCS7D01G283800 et TraesCSU01G135000 de la sous-famille NAC-b sont des homologues de NTL5, une protéine associée à la membrane et située dans le noyau chez le blé et le riz (Tang *et al.* 2012 ; Kim *et al.* 2010). Cependant, selon Liang *et al.* (2015), la protéine NTL5 d'*Arabidopsis* n'est pas ancrée à la membrane, car elle n'a pas de domaine transmembranaire.

7. Expansion de la famille TaNAC

7.1. Les duplications à petite échelle, un moteur de l'expansion de la famille TaNAC

Nous avons identifié 488 membres de la famille NAC chez le blé tendre. À titre de comparaison, seulement 117 membres ont été identifiés chez *Arabidopsis*, 151 chez le riz (Nuruzzaman *et al.* 2010) ou 168 chez le blé dur (Saidi *et al.* 2017). La présence de 3 sous-génomes chez le blé tendre explique en partie cette différence mais nous avançons l'hypothèse qu'elle serait aussi due à un nombre élevé de duplications apparues au cours de l'évolution de son génome.

Les schémas évolutifs des familles de gènes varient entre les taxons végétaux. Jin *et al.* (2017b) ont observé 2 scénarios d'évolution distincts des facteurs de transcription NAC entre dicotylédones et graminées. Selon ces auteurs, les lignées de dicotylédones montrent une acquisition d'un plus grand nombre de gènes que les lignées de graminées, tandis que ces dernières ont montré moins de pertes génétiques. Au sein des graminées, l'expansion des familles de facteurs de transcription varie entre espèces. Nous avons montré que beaucoup de séquences de la famille NAC de blé tendre sont homologues d'un même gène chez *T. urartu*, le riz ou encore le sorgho. Ce résultat est cohérent avec l'observation de Murozuka *et al.* (2018), qui a constaté l'expansion d'un clade particulier au sein des *Triticeae* par rapport à d'autres espèces de Poacées (riz et maïs). Ce clade intervient dans le développement du grain et selon les auteurs, son expansion serait corrélée avec un développement des *Triticeae* dans un environnement méditerranéen qui a nécessité la mise en place de mécanismes de maturation rapide du grain. Ce clade particulier sera discuté plus amplement dans le Chapitre II de cette thèse.

Différents mécanismes ont été proposés pour expliquer l'expansion des familles de gènes chez les plantes, par exemple les duplications segmentaires, la rétroposition, la transposition d'ADN dupliqué, les cross-over inégaux ou encore la polyploïdisation (Salse *et al.* 2008 ; Ning *et al.* 2017 ; Huo *et al.* 2018). En plus de cela, Wicker *et al.* (2010) puis Glover *et al.* (2015) ont proposé un autre mécanisme conduisant à l'évolution et à l'expansion des familles de gènes. En particulier, Glover *et al.* (2015) ont montré que des duplications inter-chromosomiques à petite échelle se produisent dans le génome du blé tendre après réparation d'une cassure d'ADN double brin. Ce mécanisme peut conduire à l'apparition de gènes aux séquences très similaires mais qui sont pourtant non-synténiques (c'est-à-dire ancrés sur des chromosomes différents). Quels que soient les mécanismes impliqués dans la duplication des gènes, celle-ci peut conduire à une pseudogénisation (accumulation de mutations conduisant à un gène non-fonctionnel), une

conservation de la fonction du gène, une sous-fonctionnalisation ou une néo-fonctionnalisation des gènes dupliqués (Magadum *et al.* 2013).

L'inférence d'un arbre phylogénétique a permis de mettre en évidence l'existence de séquences TaNAC très proches phylogénétiquement (regroupées dans un clade) et physiquement voisines sur le même chromosome. Ces séquences ne présentent pas toujours la structuration phylogénétique classique en groupes d'homéologie. *A contrario*, d'autres clades regroupent des séquences qui ressemblent à des groupes d'homéologie, sans pour autant être localisées sur les 3 sous-génomes d'un même chromosome. Ces 2 constats suggèrent l'existence de duplications intra- et inter-chromosomiques au sein de la famille TaNAC.

Pour tester l'hypothèse des duplications à petite échelle comme moteur de l'expansion de la famille TaNAC, nous nous sommes concentrés sur 3 régions génomiques portées par les chromosomes 4B, 4D et 5A, qui englobent 11 gènes TaNAC présentant des similarités de séquence (regroupement sur un clade de l'arbre phylogénétique) et de structure génomique (2 exons) ainsi qu'un regroupement physique sur ces 3 régions chromosomiques (séquences voisines). En récupérant les 3 régions génomiques portant ces gènes et en les alignant, nous avons mis en évidence une homologie élevée entre elles (Figure 50). Cette constatation appuie l'hypothèse évoquée par Glover *et al.* (2015) qui avance que l'expansion des familles de gènes est en partie due à des duplications à petite échelle entraînées par des séquences répétitives. Murozuka *et al.* (2018) ont d'ailleurs identifié 10 duplications en tandem et 8 duplications segmentaires liées à des gènes NAC chez l'orge.

La forte homologie entre ces 3 régions génomiques abonde dans le sens de l'hypothèse évoquée par Salse *et al.* (2018), proposant que de nombreux événements de duplication se soient produits au sein de la famille TaNAC lors de l'évolution du blé. En supposant que ces régions soient liées par un événement de duplication, cette duplication pourrait avoir eu lieu après la polyploïdisation du blé, car les séquences TaNAC portées sur ces régions ne sont pas structurées en groupe d'homéologie. Les gènes TaNAC liés à des portions dupliquées des régions génomiques 5A, 4B et 4D appartiennent à la sous-famille NAC-e. Compte tenu du fait que les gènes TaNAC aux fonctions homologues semblent être regroupés dans les mêmes sous-familles (Shen *et al.* 2009), nous pouvons supposer qu'un élargissement spécifique d'une sous-famille *via* des duplications pourrait permettre l'évolution des mécanismes d'adaptation des plantes en multipliant les gènes liés à des voies spécifiques. C'est d'ailleurs l'hypothèse avancée par Murozuka *et al.* (2018) pour expliquer l'apparition du clade porteur de gènes intervenant dans

le développement du grain. Cependant, cette hypothèse nécessite des validations expérimentales pour déterminer sa pertinence biologique chez le blé tendre.

De plus, les gènes TaNAC sont majoritairement identifiés dans les régions distales des chromosomes, où de nombreux événements de duplication sont observés. Cette répartition inégale n'est pas surprenante, étant donné que la densité de gènes augmente avec la distance au centromère (Choulet *et al.* 2014). Or Akhunov *et al.* (2003) ont montré que chez le blé, les loci non dupliqués sont fréquents dans les régions chromosomiques proches du centromère, tandis que les duplications ont le plus souvent lieu dans les régions distales.

7.2. Les rétropositions, un autre moteur de l'expansion de la famille TaNAC

La rétroposition d'un gène a lieu lorsque son ARN est reverse-transcrit en ADNc et réintégré dans le génome (Brosius 1991). Au sein de la famille TaNAC, 100 séquences (soit 20%) ne possèdent pas d'intron, ce qui est une caractéristique des rétroposons (Brosius 1991). Mis à part 2 groupes d'homéologie (6 séquences), toutes ces séquences appartiennent à la sous-famille NAC-h. Cependant, en plus de l'absence d'intron, il existe d'autres signatures présentes dans les régions flanquantes d'un gène issu de rétroposition, comme les marques Terminal Side Duplication ou les queues poly-A aux extrémités 3' (Brosius 1991). Chez le peuplier, Wang *et al.* (2018b) ont pu identifier des queues poly-A aux extrémités de séquences NAC possédant des introns. Cependant, aucune de ces signatures n'a pu être identifiée dans nos séquences TaNAC sans intron. Il est possible que les signatures de la rétroposition aient été modifiées par de multiples mutations au cours de l'évolution à tel point qu'elles ne soient plus reconnaissables. De plus, plusieurs de ces gènes TaNAC sans intron sont structurés en groupes d'homéologie. Ce constat suggère que les événements de rétroposition/duplication supposés se seraient produits avant la polyploïdisation du blé.

Il est intéressant de noter que la distribution des gènes TaNAC sans intron sur les 7 chromosomes du blé tendre n'est pas équivalente à celle observée pour l'ensemble de la famille TaNAC. Par exemple, les proportions de gènes de la famille TaNAC entière sont proches sur les chromosomes 3 et 5 (13 et 16% respectivement), mais le chromosome 3 porte 29% des gènes sans intron tandis que le chromosome 5 n'en porte aucun. Les gènes nés d'une rétroposition peuvent devenir des pseudogènes non-exprimés à cause de l'absence de séquences régulatrices de leur expression (Brosius 1991). Wang *et al.* (2006) ont montré que les

pseudogènes issus d'une rétroposition sont distribués équitablement le long des chromosomes car leur localisation n'est pas soumise à une pression de sélection, tandis que les gènes fonctionnels issus d'une rétroposition sont préférentiellement ancrés dans les régions distales des chromosomes. Or, certains gènes sans intron portés par le chromosome 3 sont localisés dans sa région centromérique et pourraient donc être non-fonctionnels. Une vérification de leurs niveaux d'expression sur le site expVIP (<http://www.wheat-expression.com>) a montré que les gènes ancrés dans les parties distales de ce chromosome ont tendance à présenter un niveau d'expression plus important que les gènes ancrés dans ses régions centromériques.

À une autre échelle, l'acquisition de gènes par rétroposition n'est pas observée de manière équitable entre les sous-familles de TaNAC. En effet, les gènes ne possédant pas d'intron, donc potentiellement issus d'une rétroposition, sont majoritairement regroupés dans la sous-famille NAC-h. Chez le coton, Fan *et al.* (2018a) ont constaté de la même manière que les événements de duplication ont une forte préférence pour certaines sous-familles de NAC.

Ces résultats indiquent que des phénomènes de rétroposition et de duplication à petite échelle ont très probablement participé à l'expansion de cette famille de gènes. En outre, ces mécanismes pourraient agir de manière combinée : la rétroposition d'un gène suivie de duplications à petite échelle (entraînées, par exemple, par des séquences répétées encadrant ce gène) vont conduire à une multiplication de ce gène en plusieurs copies dans le génome du blé. Par exemple, TraesCS2A01G328100, TraesCS2A01G328200, TraesCS2A01G328300 et TraesCS2A01G328500 sont des gènes sans intron, côte à côte sur le chromosome 2 et appartenant au même clade de l'arbre phylogénétique. Ils pourraient constituer un exemple de rétroposition/duplication ayant eu lieu au cours de l'évolution de la famille TaNAC.

7.3. Duplication et fonction des gènes

Quelle que soit la nature du mécanisme qui conduit à la duplication de gènes, ces duplications joueraient un rôle majeur dans la création de variants génétiques, qui peuvent ensuite conférer des capacités d'adaptation à l'espèce. La divergence entre gènes peut être estimée de plusieurs manières, par exemple par le calcul du ratio Ka/Ks, c'est-à-dire le ratio entre mutations non-synonymes par site non-synonyme et mutations synonymes par site synonyme (Hurst 2002). Cette approche a d'ailleurs été utilisée par Peng *et al.* (2015) pour estimer la divergence fonctionnelle existant dans la famille NAC chez le maïs. Cependant, dans notre travail, nous

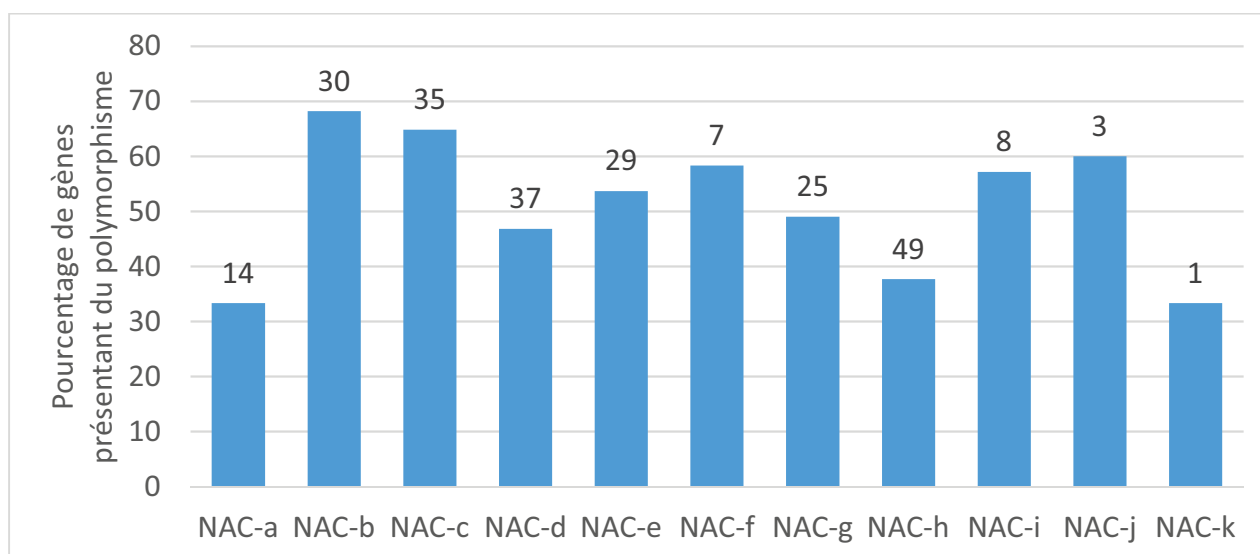


Figure 55 : Pourcentage des gènes TaNAC présentant du polymorphisme dans chaque sous-famille des NAC du blé tendre, selon les données issues de la compilation des données de la puce BW 420K et de Whealbi. En moyenne, 49% des gènes sont porteurs d'au moins un SNP, toute sous-famille confondue. L'étiquette au-dessus de chaque histogramme correspond au nombre de séquences concernées par sous-famille.

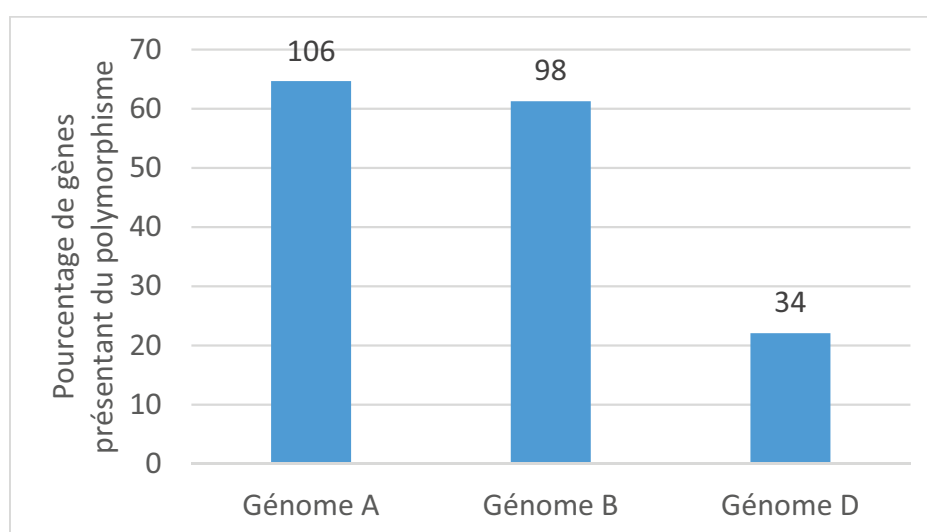


Figure 56 : Pourcentage de gènes TaNAC présentant du polymorphisme dans chacun des 3 sous-génomes du blé tendre, selon les données issues de la compilation des données de la puce BW 420K et de Whealbi. En moyenne, 49% des gènes sont porteurs d'au moins un SNP, toute sous-famille confondue. Tandis que les sous-génomes A et B portent respectivement 65% et 61% de de gènes avec du polymorphisme, seulement 22% des gènes TaNAC du sous-génome D possèdent un SNP. L'étiquette au-dessus de chaque histogramme correspond au nombre de séquences concernées par sous-génome.

avons cherché si une telle divergence entre séquences pouvait se manifester par des profils d'expressions différents. Par exemple, les TaNAC portés par les 3 régions génomiques 4B, 4D et 5A étudiées comme exemple de duplication présentent des profils d'expression fortement corrélés en réponse à des stress abiotiques (thermique et sécheresse), mais ces profils apparaissent faiblement corrélés au cours du développement des différents organes du blé. L'étude d'une seconde duplication concernant des gènes portés par les chromosomes 2A, 2B et 2D a révélé le cas de figure inverse : les profils d'expression de ces paralogues sont fortement corrélés durant le développement, mais beaucoup moins en réponse aux stress abiotiques. La divergence d'expression entre gènes NAC dupliqués a été observée chez d'autres espèces comme le sarrasin de Tartarie (Liu *et al.* 2019).

Bien que nous n'ayons pas conduit cette analyse particulière sur la totalité des TaNAC et sur l'ensemble des données d'expressions disponibles, ces premiers exemples indiquent que, suite à l'expansion de la famille TaNAC, certains membres ont certainement acquis des propriétés divergentes, notamment en terme de profils d'expression, et que ces variations peuvent être des arguments adaptatifs face à des stress par exemple.

8. Répartition du polymorphisme au sein de la famille TaNAC

Les gènes TaNAC porteurs d'au moins un marqueur de polymorphisme, ainsi que leur sous-famille et leur chromosome ont été identifiés. Plus de 60% des gènes présents dans les sous-famille NAC-b, NAC-c et NAC-j possèdent au moins un site de polymorphisme, tandis que cela représente moins de 40% des membres des familles NAC-a, NAC-h et NAC-k (Figure 55).

La répartition du polymorphisme entre les sous-génomes est largement inégale. En effet, 65% et 61% des gènes portés par les sous-génomes A et B respectivement présentent du polymorphisme tandis qu'ils ne sont que 22% sur le sous-génome D (Figure 56). Cette variation est due à la différence de polymorphisme entre les 3 sous-génomes, connue depuis longtemps et confirmée par Rimbert *et al.* (2018). Ces auteurs ont identifié 3,3 million de SNP, dont 41% portés par le sous-génome A, 49% par le sous-génome B et seulement 10% par le sous-génome D. Or, il est connu que la plasticité, et donc la capacité à accumuler des SNP, est variable entre les 3 sous-génomes du blé tendre. En effet, Pont *et al.* (2013) ont décrit le sous-génome D comme « essentiel » c'est-à-dire beaucoup moins plastique que les sous-génomes A et B.

Tableau 23 : Nombre de SNP, longueur des séquences codantes et nombre de SNP/1000 pb de séquences codante pour les gènes porteurs de polymorphisme au sein des 3 sous-génomes du blé tendre. Nous pouvons constater que, si le sous-génome D présente un nombre plus faible de SNP, ceux-ci sont répartis de manière plus dense.

	Nombre de SNP dans la séquence codante (exon)	Somme de la longueur des séquences codantes (pb)	Nombre de SNP / 1000 pb codant
Génome A	204	123291	1,65
Génome B	146	113904	1,28
Génome D	111	40950	2,71

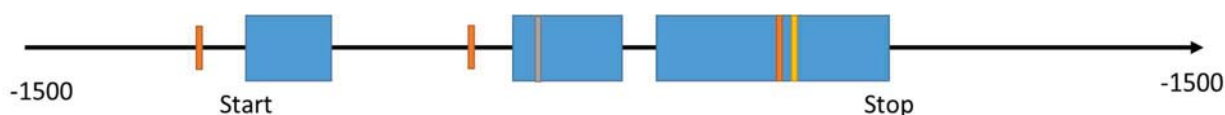


Figure 57 : Structure et SNP de la séquence génomique de TraesCS2A01G561800. Cette séquence génomique possède 2 introns, et 5 SNP ont été identifiés dessus. Quatre SNP (3 OTV en orange et un SNP bi-allélique en jaune) ont été identifiés avec les données de la puce BW_420K, et montrent une association significative avec l'accumulation des gliadines dans le grain de blé tendre soumis à une condition de forte azote. Le SNP noté en gris n'a pas montré d'association significative avec des caractères d'intérêt.

Par ailleurs, nous avons observé que les régions introniques sont les plus polymorphes, avec 0,99 SNP/ 1000 pb. Ceci s'explique par une pression de sélection beaucoup plus faible dans les régions non codantes (intron) par rapport aux régions codantes (exon). Bien que les UTRs ne soient pas des régions codantes, ils montrent un niveau de variabilité plus faible. Nous pouvons supposer que ceci est dû à la présence de régions régulatrices qui subissent une pression de sélection empêchant l'apparition de variants. Cependant les méthodes d'identification des SNP ciblées préférentiellement sur les parties transcrites n'ont certainement pas permis de détecter tous les SNP des UTR, ce qui provoque donc un biais.

Si le sous-génome D porte peu de gènes présentant du polymorphisme, celui-ci est plus dense que chez les 2 autres sous-génomes (Tableau 23). D'ailleurs, TraesCS4A01G499900 (sur le sous-génome A), TraesCS7D01G543400 et TraesCS2D01G567000 (sur le sous-génome D) sont les gènes qui présentent le plus grand nombre de SNP dans la partie codante, respectivement 30, 25 et 23 SNP. TraesCS4A01G499900 et un second gène (TraesCS7B01G461700) portent le plus grand nombre de SNP au total : 79 SNP chacun.

9. Génétique d'association avec les caractères phénotypiques d'intérêt

Les associations observées entre la composition protéique du grain et les marqueurs issus de Whealbi ont été calculées sur 108 lignées. Ce nombre assez faible est une limite certaine, et il serait intéressant de vérifier l'existence de cette association sur des bases de données plus larges, afin de valider la robustesse de l'association.

L'association de 3 variables Présence/Absence du gène TraesCS2A01G561800 avec ces paramètres d'accumulation des formes α/β et γ -gliadines en condition de fort azote nous permet d'avancer l'hypothèse qu'il pourrait y avoir une délétion partielle ou complète de ce gène dans certains génotypes, et que celle-ci serait corrélée avec l'accumulation des protéines de réserve (et notamment de gliadines) dans le grain. Cette hypothèse est cependant contradictoire avec le SNP bi-allélique (en jaune sur la Figure 57) qui est distant de 20 pb du SNP OTV voisin et qui présente un profil classique et non pas un profil Présence/Absence. Elle est aussi contradictoire avec le SNP qui n'a montré aucune association significative avec un caractère d'intérêt (en gris sur la Figure 57) et qui est placé au milieu de la séquence codante de ce gène.

Nous avons identifié le gène TraesCS6A01G051700, localisé sur le chromosome 6A, comme associé avec des paramètres d'accumulation des protéines des groupes α/β et γ -gliadine dans le grain en condition « fort azote ». Sur le chromosome 6B, le gène NAM-B1 a été très étudié dans la tribu des *Triticeae*, car les espèces sauvages (comme l'amidonnier) en possèdent un allèle fonctionnel tandis que les espèces domestiquées comme le blé dur ou le blé tendre possèdent, pour la grande majorité, un allèle avec une insertion d'une paire de bases qui le rend non-fonctionnel. Il est donc considéré tantôt comme un gène de domestication et tantôt comme un gène de diversification (Lundström *et al.* 2017). NAM-B1 est localisé dans un QTL sur le chromosome 6B. Il a été trouvé lié aux protéines de réserve du grain, en accélérant la sénescence et en augmentant la remobilisation des nutriments des feuilles aux grains en développement (Uauy *et al.* 2006). Cependant, son homéologue a déjà été identifié sur le sous-génome A par Cormier *et al.* (2015) et ce n'est pas le gène TraesCS6A01G051700 que nous avons identifié. Plessis *et al.* (2013) avaient déjà trouvé, sur le chromosome 6A, des marqueurs associés à la concentration du grain en protéines et au pourcentage de l'azote total alloué aux protéines de réserve. Ces mêmes auteurs avaient identifié, sur le chromosome 6B, 2 marqueurs associés à la composition du grain en gliadine qui n'étaient pas reliés au NAM-B1. Nous pouvons donc supposer que, en plus de celui qui porte le NAM-B1, un second QTL lié à l'accumulation des protéines de réserve et plus particulièrement des gliadines serait présent sur les chromosomes du groupe 6 et qu'il inclurait le gène TaNAC TraesCS6A01G051700. Il serait intéressant d'étudier l'éventuelle colocalisation de ce gène et de ses homéologues avec un QTL d'accumulation des protéines de réserve car la caractérisation d'un facteur de transcription jouant un rôle sur la composition protéique du grain peut présenter un fort intérêt pour les programmes d'amélioration variétale.

10. Les idées clés

- En utilisant la base de données IWGS RefSeq v1.0, 488 séquences TaNAC ont été recensées chez le blé tendre et structurées en 11 sous-familles,
- Parmi elles, 361 séquences sont strictement identiques à celles identifiées par Borrill *et al.* (2017) à partir de la base de données TGAC, ce qui appuie d'autant plus leur existence biologique,
- Les analyses génomiques ont fourni des arguments pour soutenir l'hypothèse de duplications et de rétropositions comme moteurs de l'expansion de la famille TaNAC :
 - Des séquences sont regroupées sur l'arbre phylogénétique ; certaines sont voisines sur les chromosomes, d'autres non. Ces séquences pourraient être issues de duplications intra- et inter-chromosomiques à petite échelle,
 - 20% des gènes TaNAC ne possèdent pas d'intron. Ces gènes pourraient être issus de rétroposition (intégration d'un ARN rétro-transcrit dans le génome).
- Les profils d'expression des gènes TaNAC contenus dans des régions dupliquées présentent une corrélation variable au cours du développement ou en réponse à des stress abiotiques, ce qui suggère une diversification fonctionnelle parmi ces gènes,
- Les TaNAC portés par les sous-génomes A et B présentent beaucoup plus de diversité allélique que les TaNAC ancrés sur le sous-génome D. Cette diversité varie aussi entre les sous-familles de TaNAC. Les régions introniques sont les plus denses en SNP,
- Deux membres de la famille TaNAC possèdent des marqueurs alléliques associés avec des caractères d'accumulation de protéines de réserve dans le grain, notamment les gliadines ; d'autres gènes sont porteurs de SNP associés avec des caractères de PMG et de taille du grain. Ils pourraient être des cibles de programmes de sélection.

Ce chapitre a fait l'objet d'une valorisation sous forme de publication dans la revue PlosOne (Guérin et al. (2019), Annexe VI).

Chapitre II

REPONSE PHYSIOLOGIQUE DES MEMBRES DE LA FAMILLE TaNAC AU COURS DU DEVELOPPEMENT ET EN REPONSE À DES STRESS ABIOTIQUES

Tableau 24 : Recensement des 3 stades de développement utilisés pour les analyses RNAseq de chacun des 5 organes étudiés : feuille, racine, tige, épi et grain. Les stades prélevés pour les analyses RNAseq sont marqués par une croix (tiré de Choulet et al. 2014).

Stade	Stade de développement du blé	Echelle de Feekes	Echelle de Zadoks	Feuille	Racine	Tige	Epi	Grain
Semis	1ère feuille du coléoptile	1	10	X	X			
3 feuilles	3 feuilles déployées		13		X			
3 talles	Pousse principale et 3 talles		23	X				
Epis à 1 cm	Erection de la tige	5	30			X		
2 nœuds	2ème nœuds détectable	7	32			X	X	
Méiose	Ligule de la feuille drapeau visible	9	39		X		X	
Anthèse	1/2 de la floraison complète		65			X	X	
2 JAA (50°Cj)	Grains au stade aqueux		71	X				X
14 JAA (350°Cj)	Grains au stade laiteux		75					X
30 JAA (700°Cj)	Grain au stade pâteux		85					X

Tableau 25 : Parmi la totalité des gènes suivis, nombre de gènes significativement sur-exprimés ($\text{Log}(\text{ratio}) \geq 1$) ou sous-exprimés ($\text{Log}(\text{ratio}) \leq -1$) dans un organe par rapport aux 4 autres, avec pour seuil une p-value ajustée de 0,05.

	Grain	Feuille	Epi	Tige	Racine
Nombre de gènes sous-exprimés	6505	5815	6533	6356	6362
Nombre de gènes surexprimés	6862	5508	4468	5117	8447

Avant-propos :

Suite au recensement des 488 membres de la famille TaNAC, le second objectif de ma thèse vise à étudier leurs profils d'expression au cours du développement de la plante et lors de sa réponse à différents types de stress abiotiques. Pour cela, des données d'expression publiques de gènes cartographiés sur la pseudomolécule du blé ont été utilisées. Ces données de type RNAseq correspondent à l'expression de gènes dans différents organes au cours de leur développement ainsi qu'en réponse à certains stress abiotiques (choc thermique et sécheresse).

1. Expression des gènes TaNAC dans cinq organes à trois stades de développement

Les gènes NAC sont impliqués dans le développement des plantes, leurs profils d'expression peuvent donc varier entre organes et, pour un organe considéré, au cours de son développement. Afin d'identifier les gènes TaNAC présentant une expression préférentielle dans l'un ou l'autre des organes du blé, nous avons utilisé les données d'expression brutes RNAseq générées dans le cadre de l'étude de Choulet *et al.* (2014) et disponibles sur la plateforme www.wheat-expression.com (Borrill *et al.* 2016). Ces données correspondent au suivi de l'expression des gènes totaux dans 5 organes (racines, feuille, tige, épi et grain) mesurés à 3 stades de développement (recensés dans le Tableau 24) du cultivar modèle de blé tendre Chinese Spring.

Après normalisation à l'aide du package R Limma (Ritchie *et al.* 2015), nous avons pu identifier les Gènes Différentiellement Exprimés (DEG) entre organes, avec pour seuils un Log(ratio) absolu de 1 et une p-value ajustée de 0,05. Le nombre de DEG identifiés dans la totalité du transcriptome est présenté dans le Tableau 25.

Cette étude a ensuite été centrée sur les membres de la famille TaNAC présentant une expression fortement préférentielle pour un organe. Pour cela, il a été choisi de ne conserver que les gènes TaNAC présentant une expression plus forte pour un organe par rapport aux 4 autres testés, avec pour seuil un Log(ratio) fixé à +12. En utilisant ce seuil assez strict, 52 gènes appartenant à la famille TaNAC ont été identifiés comme présentant une expression préférentielle à un organe. Ces gènes sont répartis comme suit : 19 gènes sont préférentiellement exprimés dans le grain, 7 dans la feuille, 20 dans la racine, 6 dans l'épi et aucun dans la tige ; ils sont recensés dans le Tableau 26 et la Figure 58 présente une heatmap de leur expression

Tableau 26 : Liste des 52 gènes appartenant à la famille TaNAC et présentant une expression préférentielle à un organe parmi les 5 testés (racine, feuille, tige, épi, grain). L'expression est considérée comme préférentielle à un organe lorsque son Log(ratio) est $\geq$ à 12 et la p-value ajustée $\leq$ à 0,05. Le Log(ratio) et la p-value ajustée associés sont précisés pour chaque gène.

Transcrit	Organe dans lequel le gène est préférentiellement exprimé	Log(ratio)	p-value
TraesCS2D01G576500.2	Feuille	18,1776	0,0009
TraesCS4B01G329100.1	Feuille	17,5725	0,0121
TraesCS2D01G101300.1	Feuille	17,4715	0,0231
TraesCS4B01G328900.1	Feuille	17,2993	0,0298
TraesCS7A01G318500.1	Feuille	16,7010	0,0074
TraesCS7D01G315100.1	Feuille	16,4001	0,0356
TraesCS2A01G566000.1	Feuille	16,2103	0,0377
TraesCS7B01G489500.2	Grain	24,7698	0,0026
TraesCS7D01G543500.2	Grain	24,3941	0,0070
TraesCS7A01G569100.1	Grain	24,2687	0,0038
TraesCS2A01G326600.1	Grain	24,0249	0,0000
TraesCS7B01G100300.1	Grain	23,8023	0,0025
TraesCS7A01G189200.1	Grain	23,7296	0,0002
TraesCS7B01G056300.1	Grain	23,7225	0,0050
TraesCS7D01G154200.1	Grain	23,7186	0,0070
TraesCS7A01G194700.1	Grain	23,6626	0,0059
TraesCS7D01G196300.1	Grain	23,4319	0,0031
TraesCS7B01G094000.1	Grain	23,3344	0,0005
TraesCS7A01G152400.1	Grain	23,2576	0,0009
TraesCS3B01G092800.1	Grain	23,0765	0,0007
TraesCS4B01G132500.1	Grain	22,8857	0,0000
TraesCS7D01G543400.1	Grain	22,8792	0,0004
TraesCS5A01G245900.2	Grain	22,6697	0,0147
TraesCS7A01G569300.1	Grain	21,7286	0,0011
TraesCS3A01G486500.1	Grain	20,5655	0,0000
TraesCS3A01G077900.1	Grain	20,5196	0,0329
TraesCS2D01G309800.1	Racine	23,5958	0,0000
TraesCS2A01G462500.1	Racine	23,1702	0,0000
TraesCS2A01G311500.1	Racine	22,2661	0,0000
TraesCS1D01G275500.1	Racine	21,7449	0,0000
TraesCS2D01G462900.1	Racine	21,0843	0,0002
TraesCS6A01G307200.1	Racine	20,6267	0,0000
TraesCS2B01G328300.1	Racine	20,4833	0,0000
TraesCS2B01G484100.1	Racine	20,3648	0,0002
TraesCS6D01G286300.1	Racine	20,3249	0,0000
TraesCS2D01G698900LC.1	Racine	19,9227	0,0000

dans les organes considérés. Il est intéressant de souligner que, parmi ces gènes, beaucoup se regroupent en triplets d'homéologie.

2. Effet des stress abiotiques sur l'expression des membres de la famille TaNAC

2.1. Choc thermique et/ou sécheresse sur feuilles au stade plantule

2.1.1. Analyse des données RNAseq disponibles publiquement

De la même manière que pour le suivi de l'expression des gènes dans plusieurs organes au cours du développement en conditions contrôlées, des valeurs brutes d'expression en réponse à des stress abiotiques sont disponibles sur la plateforme www.wheat-expression.com (Borrill *et al.* 2016). Dans le cadre de cette thèse, nous avons utilisé les données RNAseq publiées par Liu *et al.* (2015), qui ont soumis des plantules de blé de la variété TAM107, connue comme étant tolérante à la température et la sécheresse, à un choc thermique de 40°C et/ou un stress sécheresse dans une chambre de culture. Ce dernier stress était provoqué par addition de PEG (PolyEthylène Glycol) dans le milieu de culture des plantules. Les prélèvements réalisés 1h et 6h après l'application du/des stress ainsi qu'en condition contrôle ont été étudiés par les auteurs par RNAseq, dont les résultats bruts sont disponibles sur la plateforme www.wheat-expression.com. Après analyse de ces données selon les mêmes critères que précédemment, les DEG en réponse à ces stress abiotiques ont été identifiés avec un seuil de Log(ratio) absolu = 1 et une p-valeur à 0,05. Le Tableau 27-A montre, parmi les 107 891 gènes identifiés dans la base de données IWGS RefSeq v1.0, le nombre de gènes surexprimés ou réprimés dans la condition stressée par rapport au contrôle.

Parmi l'ensemble des DEG, 162 gènes sont des membres de la famille TaNAC, ce qui représente un tiers (33,2 %) des 488 séquences de cette famille chez le blé. Ces gènes sont répartis comme suit : 109 gènes TaNAC sont différentiellement exprimés en réponse à un choc thermique, 100 le sont en réponse à la sécheresse et 114 le sont en réponse à la combinaison des 2 stress (choc thermique x sécheresse). Parmi eux, certains gènes sont différentiellement exprimés dans plusieurs conditions de stress, et à une seule ou aux 2 durées d'application du stress testées (1h ou 6h). Le Tableau 27-B présente les TaNAC DEG identifiés dans chacune des conditions testées. Les niveaux d'expression obtenus à 1h et 6h ont ensuite été moyennés pour constituer une liste de TaNAC différentiellement exprimés dans chaque condition, ce qui

TraesCS7D01G329200.1	Racine	19,2614	0,0007
TraesCS7B01G233300.1	Racine	19,2588	0,0001
TraesCS7A01G349500.1	Racine	19,0006	0,0009
TraesCS1A01G275900.1	Racine	18,5284	0,0001
TraesCS4A01G242700	Racine	18,3197	0,0008
TraesCS4B01G072400.1	Racine	18,2910	0,0003
TraesCS5B01G275200.1	Racine	18,0402	0,0021
TraesCS5D01G283200.1	Racine	17,9848	0,0045
TraesCS2D01G382800.1	Racine	17,4006	0,0022
TraesCS4D01G071200.1	Racine	16,8766	0,0136
TraesCS7D01G371800.1	Epi	19,5257	0,0002
TraesCS7B01G143900.1	Epi	18,4879	0,0025
TraesCS7A01G334800.1	Epi	18,3718	0,0042
TraesCS7B01G246300.1	Epi	18,2189	0,0053
TraesCS2B01G582900LC.1	Epi	17,8617	0,0050
TraesCS7D01G342300.1	Epi	17,1981	0,0151

(Fin Tableau 26)

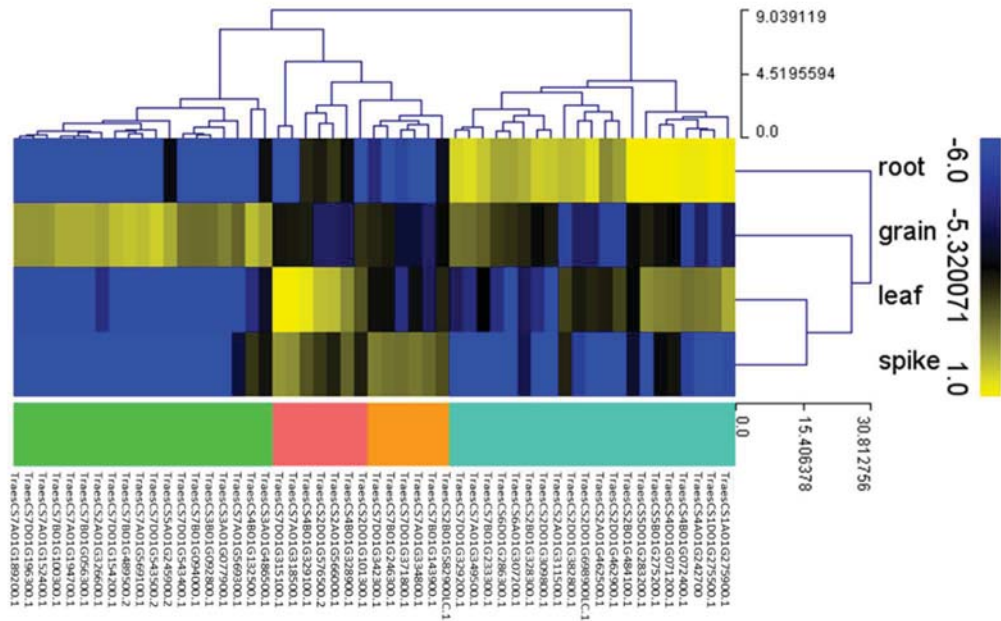


Figure 58 : Heatmap des profils d'expression de 52 gènes TaNAC dans 4 organes (grain, épi, feuille et racine), générée à partir des données RNAseq de Choulet et al. (2014). Ces 52 gènes TaNAC présentent chacun une expression préférentielle à un organe : 19 dans le grain (en vert), 7 dans les feuilles (en rouge), 20 dans les racines (en bleu) et 6 dans l'épi (en orange). Aucun gène exprimé préférentiellement dans la tige n'a été identifié, cet organe n'est donc pas représenté. Cette figure a été générée à l'aide du logiciel MeV. Un regroupement hiérarchique (hierarchical clustering) a été effectué pour l'abondance des transcrits dans toutes les conditions.

a permis de générer la Figure 59. Les gènes dont l'expression est impactée par un seul type de stress (9 gènes pour le choc thermique, 29 gènes pour la sécheresse et 12 gènes pour le choc thermique x sécheresse) ainsi que leur Log(ratio) et leur p-value associés à 1h et/ou 6h (lorsqu'elle est significative) sont listés dans la Supplemental Table S6 de la publication Guérin *et al.* (2019). Certains gènes sont donc différentiellement exprimés dans plusieurs conditions stressées tandis que d'autres ne voient leur expression évoluer qu'en réponse à un seul type de stress. Il est d'ailleurs intéressant de constater que l'expression de 12 gènes est impactée uniquement en réponse à l'application d'un multi-stress choc thermique x sécheresse.

2.1.2. Sélection de 23 gènes représentatifs parmi ceux différentiellement exprimés dans un organe ou en réponse à un stress

À partir des résultats des analyses transcriptomiques précédentes, les gènes présentant une expression préférentielle dans les feuilles ou dans les grains, ainsi que ceux surexprimés en réponse au choc thermique et/ou à la sécheresse ont été sélectionnés. Parmi ces gènes, beaucoup formaient des groupes d'homéologie ; dans ce cas, et afin de réduire le nombre de gènes à suivre, nous avons fait l'hypothèse que le comportement des membres d'un même groupe d'homéologie était similaire et nous avons décidé de sélectionner l'homéologue le plus surexprimé des 3. Ceci a mené à une liste de 50 gènes pour lesquels des couples d'amorces RT-qPCR ont été dessinées et testées. Seules les amorces fonctionnelles et spécifiques (amplification d'une seule bande à la taille attendue) ont été conservées, ce qui a représenté 23 couples d'amorces permettant le suivi de 23 gènes chez des plants de blé soumis à un stress sécheresse ou un choc thermique. Les gènes et les amorces correspondantes sont listés dans la Supplemental Table S6 de la publication Guérin *et al.* (2019).

2.1.3. Suivi de l'expression *in planta* de 23 gènes TaNAC en réponse à un choc thermique

Les résultats de choc thermique et de stress sécheresse obtenus *in silico* à partir des données de l'étude de Liu *et al.* (2015) ont été générés sur le génotype TAM107 qui est un génotype peu utilisé en laboratoire. Les mêmes conditions de choc thermique que celles décrites par cette étude (40°C pendant 1h et 6h) ont été appliquées à des plantules du cultivar de blé tendre

Tableau 27 : Nombre de gènes surexprimés ou réprimés en réponse à un stress par rapport à l'échantillon en conditions contrôles. Les 3 stress testés sont un choc thermique de 40°C, un stress sécheresse par application de PEG dans le milieu de culture, et une combinaison des 2 ; appliqués pendant 1h ou 6h. Au total, parmi ces gènes DEG, 162 ont été identifiés comme gènes TaNAC uniques.

A	Choc thermique (40°C) 1h	Choc thermique (40°C) 6h	Sécheresse 1h	Sécheresse 6h	Choc thermique x Sécheresse 1h	Choc thermique x Sécheresse 6h
Nombre de gènes réprimés par le stress	6763	11420	339	6426	7018	11288
Nombre de gènes surexprimés par le stress	9837	11195	3117	6352	11796	11629

B	Choc thermique (40°C) 1h	Choc thermique (40°C) 6h	Sécheresse 1h	Sécheresse 6h	Choc thermique x Sécheresse 1h	Choc thermique x Sécheresse 6h
Nombre de gènes TaNAC réprimés par le stress	24	39	5	28	28	41
Nombre de gènes TaNAC surexprimés par le stress	43	31	40	54	43	43

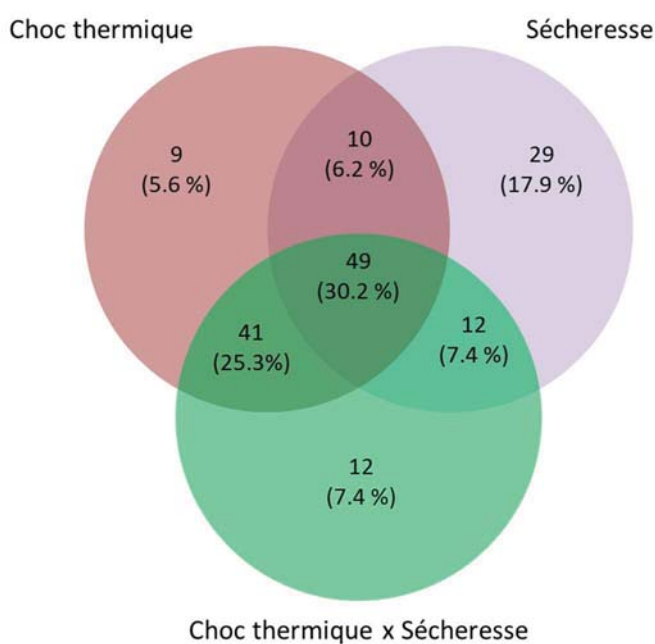


Figure 59 : Diagramme de Venn des 162 gènes TaNAC présentant une expression différentielle en réponse à un choc thermique, à un stress sécheresse et à la combinaison des 2 stress. Les chiffres correspondent au nombre de gènes différentiellement exprimés pour chaque condition, le pourcentage entre crochets correspond à la proportion que cela représente parmi les 162 gènes concernés.

Chinese Spring afin de comparer la réponse des gènes TaNAC entre ce génotype modèle du blé tendre et le génotype TAM107. Des prélèvements des feuilles en conditions contrôle et stressée ont été réalisés après 1h et 6h et l'expression des 23 gènes sélectionnés précédemment a été quantifiée par RT-qPCR. Les résultats obtenus ont été présentés sous 2 formes complémentaires nommées par la suite **données normalisées** et **ratio**.

- Les **données normalisées** permettent de comparer les niveaux d'expression normalisés entre gènes, afin d'identifier par exemple des groupes de gènes au comportement similaire, mais elles rendent difficile la lecture de la réponse de ces gènes au stress.
- Les **ratios** de l'expression du gène en condition stressée sur son expression en condition contrôle permettent d'identifier l'intensité de réponse du gène considéré face au stress, et ce sans considération de son niveau d'expression de base au sein de la feuille.

La Figure 60, présentant les données normalisées, montre une structuration des gènes en 3 grands groupes : des gènes présentant un niveau d'expression plutôt bas dans toutes les conditions (en vert sur la figure), d'autres fortement exprimés (en rouge) et enfin les derniers présentant un niveau d'expression bas qui augmente après 6h de choc thermique.

La Figure 61, illustrant les ratios, révèle une très forte structuration des gènes en 2 groupes : un premier groupe de 7 gènes dont l'expression est réprimée par le choc thermique à 6h et un deuxième contenant 16 gènes dont l'expression augmente après 6h de stress. Dans ce deuxième groupe, 2 tendances se dégagent, d'une part certains gènes réprimés à 1h ne le sont plus à 6h et d'autre part, il y a des gènes qui ne montrent pas de modification de leur expression entre la condition contrôle et stress thermique à 1h mais dont l'expression augmente en réponse au stress à 6h.

On peut noter que les 5 gènes préférentiellement exprimés dans le grain (identifiés dans le Chapitre II – Partie 1) sont logiquement ceux qui présentent le niveau d'expression le plus bas dans les feuilles, indépendamment du stress. Cependant, en étudiant leur ratio, nous pouvons constater qu'ils présentent malgré tout un profil de réponse au choc thermique dans les feuilles. Cette réponse est du type « augmentation de l'expression après 1h de choc thermique, puis répression après 6h » significative pour 3 des 5 gènes préférentiellement exprimés dans le grain.

Lorsque le gène répond au choc thermique, cette réponse se met en place dans la grande majorité des cas dès 1h de stress et est maintenue après 6h de stress. Aucun gène n'a été observé comme DEG à 1h de stress et non-DEG à 6h de stress.

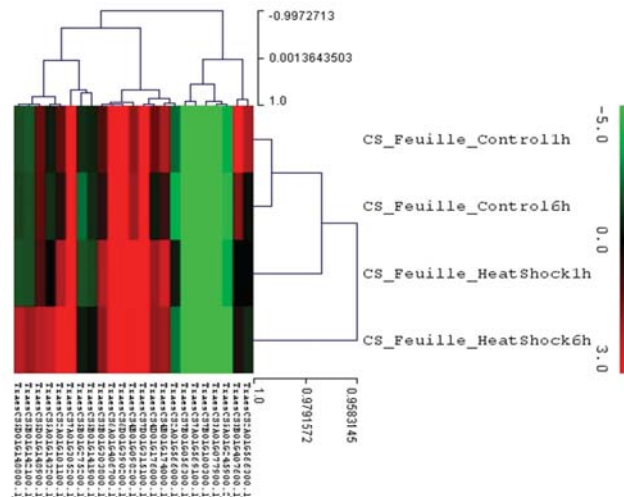


Figure 60 : Heatmap des profils d'expression des 23 gènes TaNAC suivis in planta dans les feuilles de plantules de blé tendre (cultivar *Chinese Spring*) en condition contrôle (24h) et après 1h ou 6h de choc thermique (40°C). Les gènes sont faiblement (en vert) ou fortement (en rouge) exprimés. Les 5 gènes identifiés précédemment comme étant préférentiellement exprimés dans le grain sont, de manière cohérente, regroupés et les plus faiblement exprimés (vert clair).

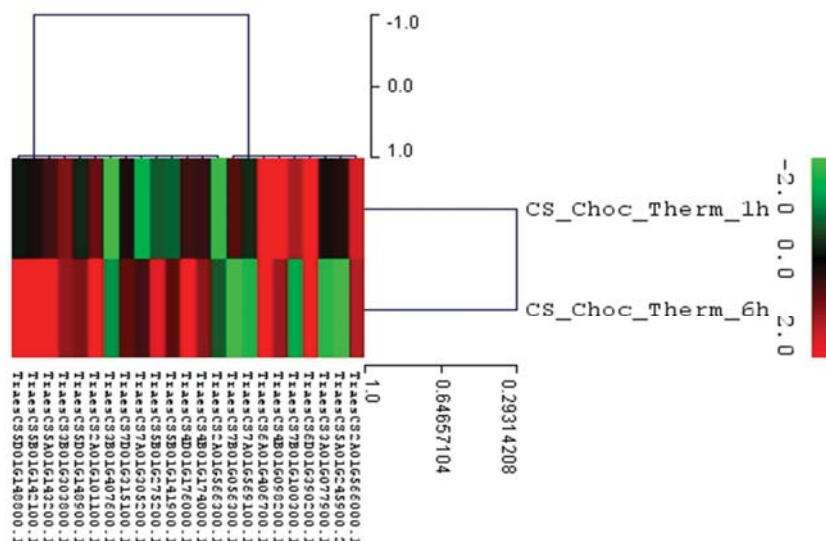


Figure 61 : Heatmap du ratio de l'expression en réponse à un choc thermique sur l'expression en condition contrôle, à 1h et 6h, pour les 23 gènes TaNAC dans les feuilles de plantules de blé tendre (cultivar *Chinese Spring*). La couleur rouge traduit une augmentation importante de l'expression en réponse au choc thermique, tandis que la couleur verte correspond à une diminution de l'expression en réponse au même stress. L'arbre à gauche de la heatmap traduit très clairement le regroupement des gènes en 2 groupes très distincts.

2.2. Suivi de l'expression *in planta* de 23 gènes TaNAC après application d'un stress sécheresse en plein champ

Dans la plupart des publications scientifiques, la sécheresse correspond en fait à un mono-stress hydrique mis en place en conditions contrôlées, simulé par un arrêt de l'arrosage ou l'ajout de PEG dans le milieu de culture. Cependant, dans un environnement non contrôlé comme c'est le cas des cultures en plein champ, ce stress est très souvent combiné à d'autres contraintes environnementales, dont notamment la température. Or, comme les résultats obtenus à partir des analyses *in silico* combinant les stress thermique et sécheresse l'ont montré (cf. Chapitre II – Partie 2.1.1), il existe des voies de réponse faisant intervenir des gènes différents entre les stress étudiés séparément et leur combinaison. L'étude de la sécheresse en conditions naturelles revêt donc une importance capitale pour observer des mécanismes ayant une réelle pertinence biologique dans les cultures en champ.

C'est dans ce cadre de réflexion qu'une expérimentation de sécheresse sur blés tendres en plein champ a été mise en place sur la plateforme Phéno3C (V. Allard, INRA). Sur cette plateforme, différents génotypes de blé tendre ont été soumis à un stress sécheresse grâce à un auvent coulissant automatiquement, qui a permis de protéger la parcelle des précipitations ; tandis que des parcelles témoins ont été maintenues sous des conditions météorologiques naturelles et irriguées si nécessaire pour éviter tout stress hydrique. Nous avons réalisé des prélèvements sur 2 génotypes de blé soumis à ce stress sécheresse : les variétés Allez-Y et Altigo, respectivement sensible et tolérante au stress sécheresse, sur la base de l'impact de la sécheresse sur leur rendement.

2.2.1. Mesures des paramètres éco-physiologiques en réponse à la sécheresse

Le rendement et le nombre d'épis par m² ont été utilisés comme marqueurs physiologiques de l'effet du stress hydrique généré lors de l'expérimentation sécheresse. Une perte de rendement de 45,2% et 45,8% a été observée pour les génotypes Altigo et Allez-y respectivement, en condition sécheresse en comparaison avec les plantes irriguées. De même, une diminution du nombre d'épis par m² a été observée (-26,6% pour Altigo et -39,1% pour Allez-y) (Figure 62). La modification de ce paramètre, qui est déterminé avant l'anthèse, traduit un niveau élevé de stress hydrique qui a affecté la croissance des plantes dès la phase de tallage (Slafer, 2014).

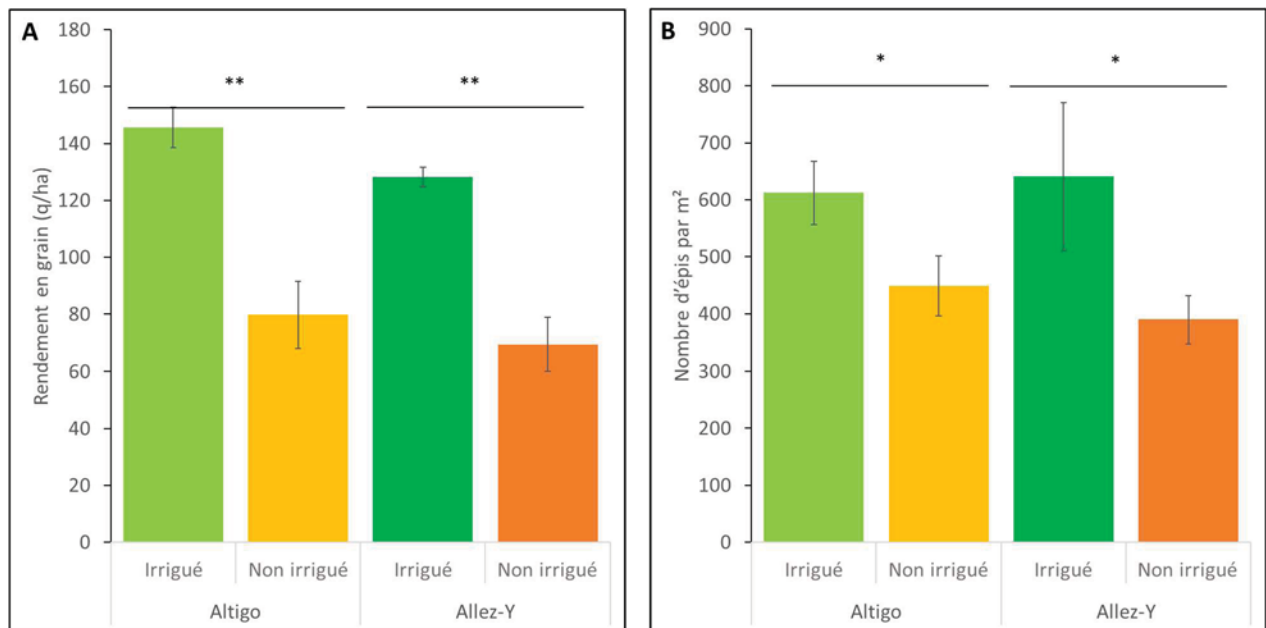


Figure 62 : Rendement en grain (A) et nombre d'épis par m² (B) obtenus pour les génotypes Allez-Y et Altigo en conditions contrôle (irrigué) et sécheresse (non-irrigué). * : différence significative à p -value = 0,05 ; ** : différence significative à p -value = 0,05.

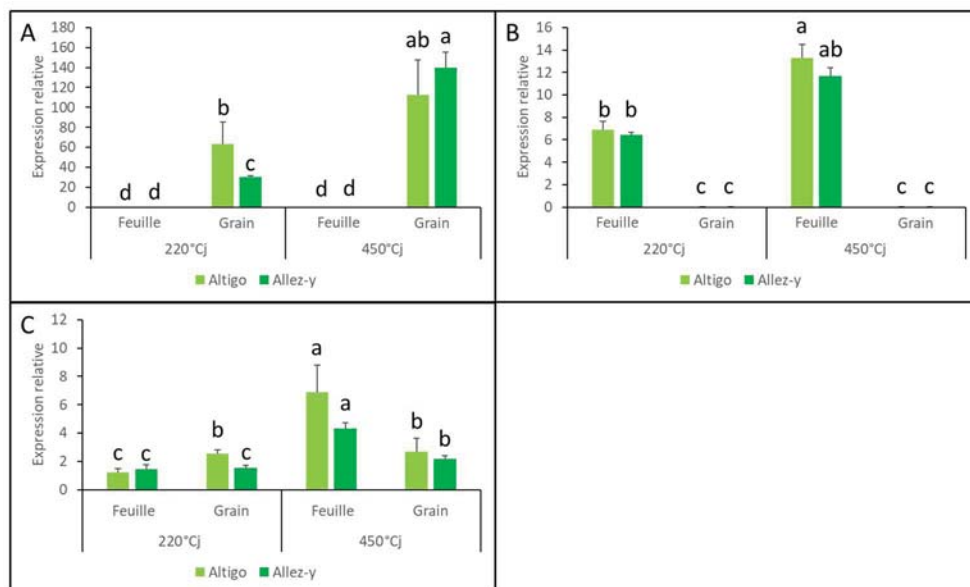


Figure 63 : Expression relative de 3 gènes TaNAC (TraesCS5A01G245900.2 (A), TraesCS7D01G315100.1 (B) et TraesCS2A01G101100.1 (C)) chez les génotypes Altigo et Allez-Y, mesurée dans les feuilles et les grains à 220°Cj et 450°Cj en conditions irriguées. Les données d'expression ont été obtenues par RT-qPCR. Trois répétitions ont permis de calculer les valeurs moyennes et les écarts-types. En utilisant le test de Kruskal-Wallis, les groupes ne présentant pas de différence significative portent la même lettre.

2.2.2. Identification des gènes TaNAC exprimés de manière différentielle chez les plantes cultivées au champ

L'expression des 23 gènes TaNAC sélectionnés en Chapitre II – Partie 2.1.2 a été suivie dans 2 organes (feuilles et grains), à 2 stades représentatifs du développement des grains (division cellulaire (220°Cj) et remplissage (450°Cj)), chez 2 génotypes de blé (Altigo et Allez-y) dans des conditions irriguées et non-irriguées. Afin d'étudier leur comportement au cours du développement de la plante, les profils d'expression obtenus pour les 23 gènes en condition contrôle (irrigué) ont d'abord été analysés. Nous avons ensuite étudié l'impact du stress sécheresse sur les profils de ces gènes pour connaître leur éventuelle implication dans la réponse de la plante à cette contrainte abiotique.

2.2.2.1. Profil d'expression des gènes TaNAC en conditions irriguées

Des différences entre organes et entre stades de développement ont été observées parmi les plantes en condition irriguée. Les 23 gènes TaNAC présentent des taux de transcrits statistiquement différents entre les organes (feuilles et grains), quels que soient le génotype et le stade concernés. Sept gènes sont préférentiellement exprimés dans le grain et 16 dans la feuille. Le gène TraesCS5A01G245900.2 présente la plus forte expression dans le grain par rapport à la feuille (10 018 fois plus), alors que TraesCS7D01G315100.1 est le gène le plus exprimé dans la feuille (332 fois plus que le grain) (Figure 63-A et 63-B).

En outre, l'expression de certains gènes TaNAC varie en fonction du stade de développement (220°Cj vs 450°Cj). Un gène est exprimé préférentiellement à 220°Cj et 15 gènes le sont à 450°Cj. Par exemple, le niveau d'expression du gène TraesCS2A01G101100.1 est 4,19 fois plus élevé dans les feuilles à 450°Cj qu'à 220°Cj (Figure 63-C).

2.2.2.2. Profil d'expression des gènes TaNAC en réponse à la sécheresse

La totalité des données d'expression RT-qPCR obtenues en réponse à la sécheresse ont été présentées dans une Analyse en Composantes Principales (ACP) (Figure 64). Les 2 premières composantes principales (CP1 et CP2) expliquent respectivement 83,5% et 10,8% de la variation totale. Cumulativement, ces 2 composantes principales ont contribué à 94,3% de la

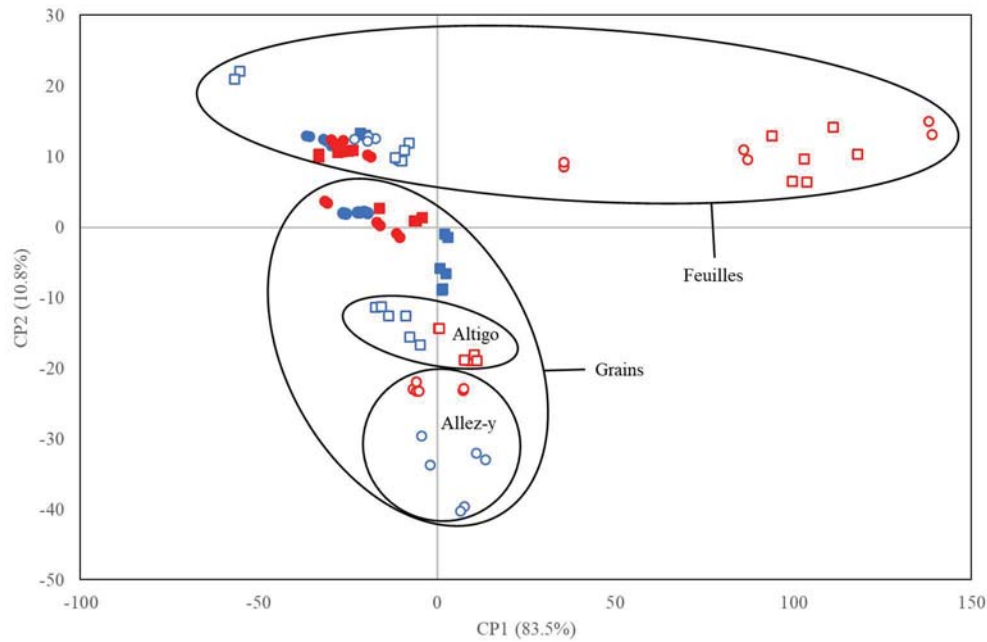


Figure 64 : Biplot montrant les 2 premières variables (CP1 et CP2) de l'analyse en composantes principales. Ce graphique représente l'analyse des variables influant l'expression des 23 gènes TaNAC dans les feuilles et les grains des génotypes Altigo (carré) et Allez-y (cercle) à 220°Cj (forme pleine) et 450°Cj (forme vide) dans des conditions irriguées (bleu) ou en réponse à un stress sécheresse (rouge).

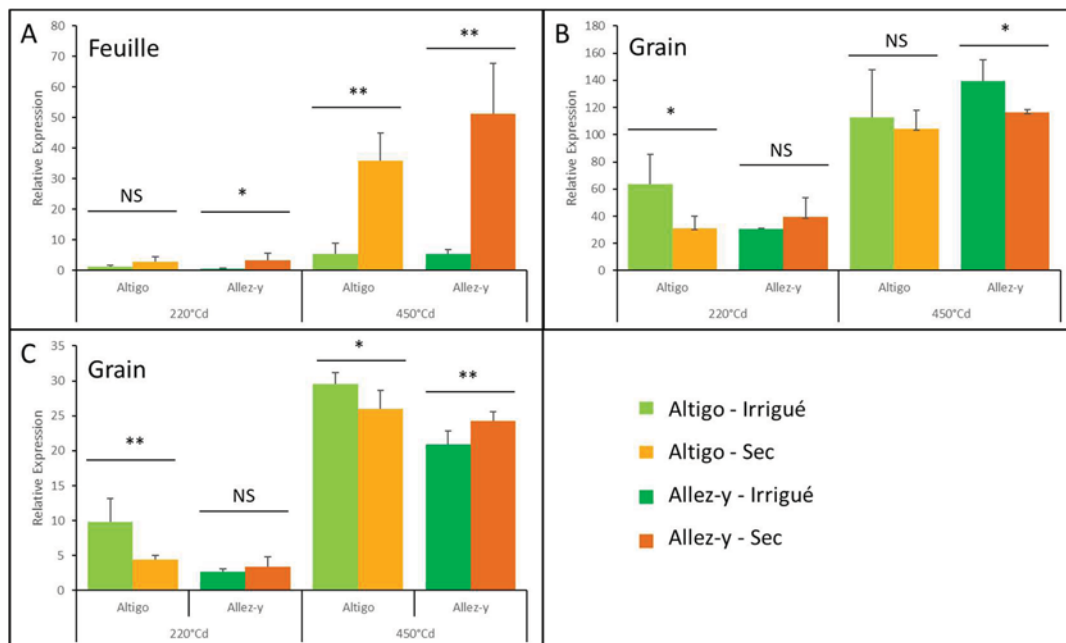


Figure 65 : Analyse de l'expression RT-qPCR de *TraesCS5A01G143200.1* (A), *TraesCS5A01G245900.2* (B) et *TraesCS7A01G569100.1* (C). Les valeurs moyennes et l'écart type ont été obtenus à partir de 3 répétitions (* indique une $p < 0,05$ et ** une $p < 0,01$ selon le test T de Student).

variation totale de l'expression des gènes en réponse aux conditions de sécheresse, d'organe, de génotype et de stade.

Cette analyse met en évidence une combinaison particulière d'expression des gènes dans les feuilles soumises à la sécheresse à un stade tardif (450°Cj) qui explique principalement la variabilité des échantillons (CP1) (les formes vides en haut à droite sur la Figure 64). En effet, dans les feuilles, la réponse la plus forte à la sécheresse comparativement au contrôle est observée à 450°Cj par rapport à 220°Cj, quel que soit le génotype (Figure 65-A).

Le second axe (axe vertical) s'explique principalement par un effet organe. En effet, les échantillons de grains sont regroupés dans la partie inférieure de la Figure 64. Dans le sous-ensemble grains, les échantillons correspondant aux stades de développement précoce et tardif des grains (forme pleine : 220°Cj, et forme vide : 450°Cj) constituent des groupes distincts le long du second axe (CP2). Et enfin, parmi les grains à 450°Cj, les 2 génotypes sont regroupés en 2 groupes différents. En effet, en réponse à la sécheresse, une différence d'expression des gènes entre les 2 génotypes a été observée dans les grains. Cet effet concerne plus précisément les 5 gènes possédant une expression préférentielle au grain : celle-ci est affectée par la sécheresse à un stade différent selon le génotype. Pour le génotype Altigo, une répression de ces 5 gènes en réponse à la sécheresse a été observée au stade 220°Cj ; alors que ceux-ci sont significativement réprimés (TraesCS3A01G077900.1, TraesCS5A01G245900.2, TraesCS7B01G056300.1 et TraesCS7B01G100100.1) ou surexprimés (TraesCS7A01G569100.1) seulement au stade 450°Cj dans le génotype Allez-y (Figure 65-B et 65-C).

2.3. Effet du stress thermique modéré sur l'expression des gènes TaNAC

Dans la nature, ce que l'on nomme réchauffement climatique peut être décomposé en 2 phénomènes : une contrainte thermique modérée présente en continu, couplée à une augmentation de la fréquence des chocs thermiques. La réponse des végétaux au choc thermique a été largement décrite dans la littérature parce que ce stress est simple à mettre en place et les mécanismes de survie qu'il engendre chez la plante sont facilement détectables. En revanche, la réponse à un stress thermique modéré est différente car elle fait intervenir des mécanismes de régulation et d'adaptation à une température élevée mais physiologique. Ce type de contrainte, beaucoup moins étudié, nécessite cependant d'être pris en compte pour la création

de nouvelles variétés de blé tendre aptes à faire face au changement climatique. C'est dans ce contexte que le projet BreedWheat, décrit dans la synthèse bibliographique de cette thèse, a été construit. Les résultats présentés dans cette partie s'inscrivent dans le cadre de ce projet et ont donné lieu à une publication (Girousse, Roche, Guérin *et al.* 2018).

Des plants de blé tendre de 2 génotypes (SerixBabax 49 et SerixBabax 139) ont été soumis à un stress thermique modéré (27°C vs 19°C) durant 10 jours après anthèse, et prélevés à différents stades pour une analyse transcriptomique et phénotypique. Cette expérimentation a montré une diminution de 13,7% du PMG chez les plantes soumises à ce stress thermique modéré. Les masses fraîches et sèches, ainsi que les volume, longueur, largeur épaisseur et nombre de cellules de l'albumen ont été négativement impactés par le stress thermique. *A contrario*, la teneur en azote par grain a, elle, été augmentée par le stress, probablement du fait d'une diminution de la quantité d'amidon dans le grain. Les études transcriptomiques ont mis en évidence 6258 et 5220 gènes dont l'expression est affectée par la température chez les génotypes 139 et 49 respectivement. La majorité de ces gènes sont activés pendant le stade 120°Cj, identifié comme le stade montrant la plus grande différence au niveau transcriptomique entre la condition contrôle et la contrainte thermique. Parmi les facteurs de transcription dont l'expression est impactée par la température dans le génotype SerixBabax 49, 4 sont des TaNAC. Parmi ces 4 TaNAC, 2 ont été détectés avec le même profil de réponse dans le génotype SerixBabax 139.

Ces 4 facteurs TaNAC font partie des 5 gènes s'exprimant préférentiellement dans le grain. Pour rappel, lors d'un choc thermique, leur expression dans les feuilles de plantule de blé tendre reste minimale, mais la sécheresse impacte leur profil d'expression de manière génotype-dépendante. Sur la base de ce résultat, nous avons décidé d'étudier plus spécifiquement ces 4 gènes TaNAC qui n'ont encore jamais été décrits dans la littérature chez le blé.

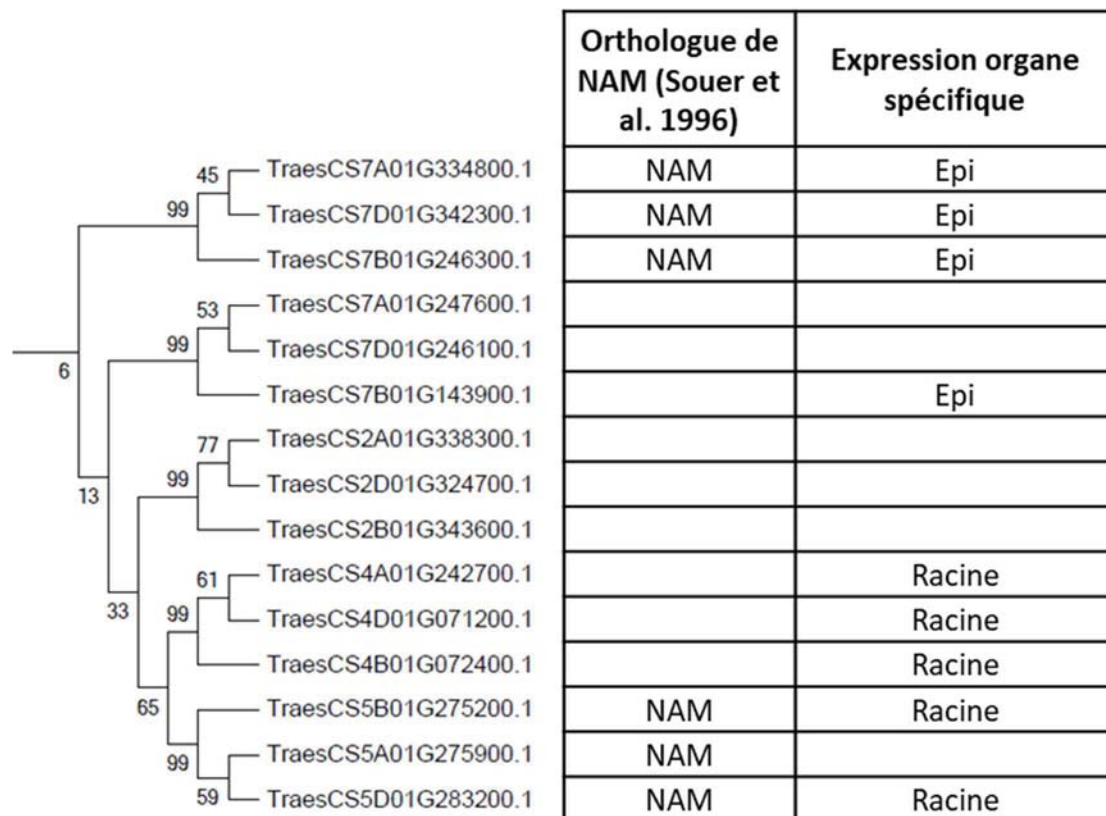


Figure 66 : Présentation d'un clade de l'arbre phylogénétique, porteur de 5 groupes d'homéologie dont 2 présentent une forte homologie avec le gène NAM décrit chez le pétunia (Souer et al. 1996). Ces 2 groupes présentent cependant une expression préférentielle dans des organes différents (épi ou racine).

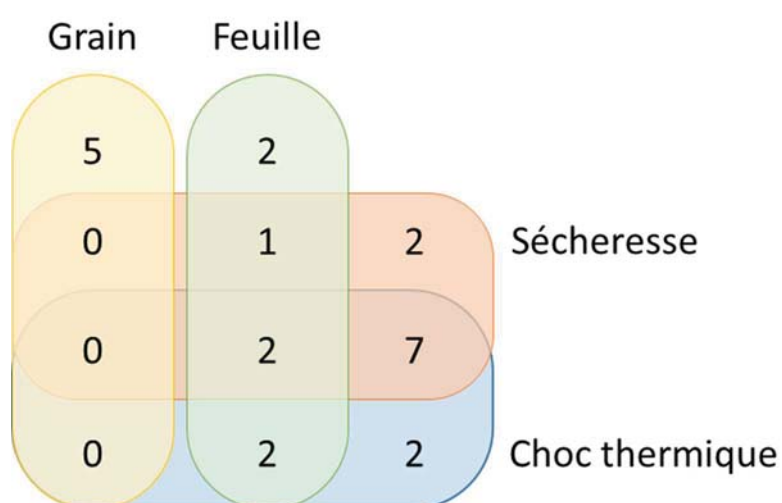


Figure 67 : Gènes préférentiellement exprimés dans un organe et/ou en réponse à une contrainte environnementale.

DISCUSSION du CHAPITRE II

Avant-propos :

Dans le premier chapitre de cette thèse, l'étude des 2 exemples de duplication a montré que l'expression de membres d'un groupe de duplication était plus ou moins corrélée au cours du développement et en réponse à des stress abiotiques. Ceci indique que les gènes de la famille TaNAC peuvent exhiber une grande variation d'expression et il est possible que cette variation se traduise par une spécialisation au niveau des tissus/organes et/ou réponse à des stimuli de l'environnement. Ce second chapitre a donc permis de mieux appréhender le comportement des membres de la famille au sein de la plante et lors de sa réponse à différents types de stress abiotiques, comme la sécheresse, le choc thermique mais aussi le stress thermique modéré. Plusieurs gènes exprimés de manière différentielle chez des blés cultivés dans des conditions de sécheresse en plein champ ont été identifiés. Les résultats obtenus mettent en évidence l'impact potentiel du génotype sur la variabilité de l'expression des membres de cette famille.

3. L'expression des gènes TaNAC évolue au cours du développement du blé tendre

L'utilisation de données transcriptomiques RNAseq disponibles publiquement a permis d'identifier 52 gènes présentant une expression préférentielle à un organe pour 4 organes parmi les 5 analysés. Parmi eux, beaucoup de gènes correspondent à des groupes d'homéologie complets ou incomplets.

Le gène NAM du pétunia est le premier NAC à avoir été caractérisé (Souer *et al.* 1996). Les groupes d'homéologie TraesCS7A01G334800.1 / TraesCS7D01G342300.1 / TraesCS7B01G246300.1 et TraesCS5B01G275200.1 / TraesCS5D01G283200.1 sont les meilleurs homologues pour ce gène chez le blé (respectivement e-63 en TBlastN et e-19 en BlastN). Ce gène NAM joue un rôle dans l'établissement des méristèmes et des *primordia* chez le pétunia, notamment des méristèmes apicaux. Or, les 2 groupes d'homéologie homologues à ce NAM présentent une expression préférentielle dans l'épi pour l'un (TraesCS7A01G334800.1 / TraesCS7D01G342300.1 / TraesCS7B01G246300.1) et dans la racine pour le second (TraesCS5B01G275200.1 / TraesCS5D01G283200.1). Ces résultats suggèrent que ces gènes NAC pourraient jouer un rôle dans l'établissement des méristèmes et

Tableau 28 : Récapitulatif de l'effet d'un choc thermique sur l'expression de 23 gènes TaNAC. Les premières colonnes présentent les résultats obtenus à partir des données RNAseq de la plateforme expVIP, sur des feuilles de plantules de blé génotype TAM107. Les 2 dernières colonnes présentent les résultats obtenus par RT-qPCR sur des feuilles de plantules de blé du cultivar Chinese Spring cultivées dans les mêmes conditions. P-value seuil = 0,05. Les gènes TaNAC différentiellement exprimés (DEG) après 1h et 6h de stress thermique, à une p-value seuil de 0,05, sont notés « Oui », les autres sont notés « Non ».

Gènes	Données bio-informatiques		Données <i>in planta</i>	
	DEG après 1h de choc thermique	DEG après 6h de choc thermique	DEG après 1h de choc thermique	DEG après 6h de choc thermique
TraesCS3B01G303800.1	Oui	Oui	Oui	Oui
TraesCS4B01G098200.1	Oui	Oui	Oui	Oui
TraesCS4D01G176000.1	Oui	Oui	Oui	Oui
TraesCS5A01G143200.1	Oui	Oui	Oui	Oui
TraesCS5B01G141900.1	Oui	Oui	Oui	Oui
TraesCS5D01G148900.1	Oui	Oui	Oui	Oui
TraesCS6A01G406700.1	Oui	Oui	Oui	Oui
TraesCS6D01G390200.1	Oui	Oui	Oui	Oui
TraesCS5B01G275200.1	Non	Oui	Oui	Oui
TraesCS2A01G101100.1	Oui	Non	Oui	Oui
TraesCS2B01G627000.1	Non	Non	Oui	Oui
TraesCS3B01G407600.1	Non	Non	Oui	Oui
TraesCS7D01G315100.1	Non	Non	Oui	Oui
TraesCS4B01G174000.1	Oui	Non	Non	Oui
TraesCS2A01G566300.1	Non	Non	Non	Oui
TraesCS3A01G077900.1	Non	Non	Non	Oui
TraesCS5A01G245900.2	Non	Non	Non	Oui
TraesCS5B01G142100.1	Non	Non	Non	Oui
TraesCS5D01G148800.1	Non	Non	Non	Oui
TraesCS7A01G305200.1	Non	Non	Non	Non
TraesCS7A01G569100.1	Non	Non	Non	Non
TraesCS7B01G056300.1	Non	Non	Non	Non
TraesCS7B01G100300.1	Non	Non	Non	Non

primordia apicaux des organes reproducteurs et racinaires respectivement. Ils appartiennent à un même clade, qui porte 5 groupes d'homéologie entiers, dont 2 montrent une expression préférentielle dans l'épi et 2 autres dans la racine (Figure 66).

Parmi les 20 gènes exprimés préférentiellement dans la racine, TraesCS6A01G307200.1 présente la meilleure identité avec FtNAC43 (e-value 8E-22) et FtNAC46 (e-value 8E-16) ; et TraesCS5B01G275200.1 est aussi similaire avec FtNAC58 (e-value 3E-12), qui sont des NAC identifiés chez le sarrasin de Tartarie et aussi exprimés préférentiellement dans la racine (Liu *et al.* 2019).

Parmi les 52 gènes présentant une expression préférentielle à un organe, d'autres sont homologues avec des gènes décrits chez plusieurs espèces végétales. Cependant, ces gènes font partie du set des 23 gènes dont l'expression en réponse à la sécheresse et au choc thermique a été suivie en RT-qPCR ; nous les traiterons donc ultérieurement dans cette discussion.

4. Un choc thermique induit une modification de l'expression d'au moins 10 gènes TaNAC dans les plantules de blé tendre

Les données d'expression obtenues par RT-qPCR sur les feuilles de blé du cultivar Chinese Spring au stade plantule en réponse à un choc thermique sont globalement cohérentes avec celles obtenues sur le génotype TAM107 étudié par Liu *et al.* (2015) dans les mêmes conditions et quantifiées par RNAseq. En comparant les résultats de ces 2 génotypes, nous avons cependant identifié plus de gènes différentiellement exprimés en réponse au choc thermique chez Chinese Spring (13 gènes après 1h de stress et 19 après 6h) que chez TAM107 (10 gènes après 1h de stress et 9 après 6h) (Tableau 28). Quasiment tous les gènes différentiellement exprimés dans le génotype TAM107 le sont aussi dans le cultivar Chinese Spring. Les variations observées peuvent être dues à la différence de technique utilisée (RNAseq vs RT-qPCR) entre les 2 expérimentations ; même si, selon Everaert *et al.* (2017) les 2 méthodes présentent des résultats très proches sur des échantillons humains. Elles peuvent aussi être imputables au changement de génotype. De la même manière, les 5 gènes qui ont été précédemment identifiés comme s'exprimant préférentiellement dans le grain présentent *in planta* un niveau d'expression très bas dans les feuilles des plantules testées, en condition contrôle comme en réponse au choc thermique (Figure 67).

Tableau 29 : Expression des 23 gènes TaNAC, à partir des données RNAseq de la plateforme expVIP, et à partir des résultats de l'analyse RT-qPCR sur des échantillons de feuille et de grains, à 2 stade de développement, en condition irriguée ou sécheresse. "x" = pas d'expression préférentielle.

Gènes TaNAC	Données RNAseq de la plateforme expVIP Expression augmentée par...	Expression mesurée lors de l'expérimentation sécheresse (RT-qPCR)		
		Préférentielle à un organe (≥ 10 x supérieure)	Préférentielle à un stade	Préférentielle à une condition environnementale
TraesCS2A01G101100.1	Choc thermique 1h	x	450°Cj	Sécheresse
TraesCS2A01G566000.1	Feuille	Feuille	x	Condition irriguée
TraesCS2A01G566300.1	Feuille	Feuille	220°Cj	x
TraesCS3A01G077900.1	Grain	Grain	450°Cj	x
TraesCS3B01G303800.1	Choc thermique 1h + 6h	x	450°Cj	Sécheresse
TraesCS3B01G407600.1	Feuille + sécheresse 1h	Feuille	x	Sécheresse
TraesCS4B01G098200.1	Choc thermique 1h + 6h	x	x	x
TraesCS4B01G174000.1	Choc thermique 1h	x	450°Cj	Sécheresse
TraesCS4D01G176000.1	Choc thermique 1h + 6h	Feuille	450°Cj	Sécheresse
TraesCS5A01G143200.1	Choc thermique 1h + 6h	x	450°Cj	Sécheresse
TraesCS5A01G245900.2	Grain	Grain	450°Cj	x
TraesCS5B01G141900.1	Choc thermique 1h + 6h	x	x	Sécheresse
TraesCS5B01G142100.1	Sécheresse 1h + 6h	Feuille	450°Cj	Sécheresse
TraesCS5B01G275200.1	Choc thermique 6h + Racine	x	x	x
TraesCS5D01G148800.1	Sécheresse 1h + 6h	x	450°Cj	Sécheresse
TraesCS5D01G148900.1	Choc thermique 1h + 6h	x	x	Sécheresse
TraesCS6A01G406700.1	Choc thermique 1h + 6h	x	450°Cj	Sécheresse
TraesCS6D01G390200.1	Choc thermique 1h + 6h	x	450°Cj	Sécheresse
TraesCS7A01G305200.1	Feuille	Feuille	x	Condition irriguée
TraesCS7A01G569100.1	Grain	Grain	450°Cj	x
TraesCS7B01G056300.1	Grain	Grain	450°Cj	x
TraesCS7B01G100300.1	Grain	Grain	450°Cj	x
TraesCS7D01G315100.1	Feuille	Feuille	450°Cj	x

5. Comportement des gènes TaNAC au cours du développement du blé en plein champ et lors de sa réponse à la sécheresse

5.1. Avant-propos

En tant que stress environnemental majeur largement répandu, la sécheresse a une incidence considérable sur les rendements de blé dans le monde. Les interactions entre le développement des plantes et leur réponse aux stress sont complexes, en particulier en plein champ où les cultures sont soumises à une combinaison de variables environnementales qui interagissent les unes avec les autres (par exemple, la température et la disponibilité en eau). Bien que peu de facteurs de transcription NAC aient été étudiés chez le blé par rapport à d'autres plantes modèles comme *Arabidopsis* ou le riz, il a été démontré que certains membres des TaNAC sont impliqués dans la réponse aux conditions environnementales en lien avec le développement de la plante (Gahlaut *et al.* 2016). Par exemple, l'expression de TaNAC2a, TaNAC2D, TaNAC4a, TaNAC6, TaNAC7, TaNAC13 et TaNTL5 est induite à des intensités différentes par le stress hydrique, salin et/ou les basses températures (Tang *et al.* 2012 ; Huang et Wang 2016). Récemment, il a été constaté que TaRNAC1, exprimé constitutivement dans les racines, jouerait un rôle positif dans la tolérance au stress hydrique (Chen *et al.* 2018). Sa surexpression chez le blé entraîne une biomasse, un rendement en grain et une longueur de racine plus élevés. Au cours de cette thèse, les profils d'expression de 23 gènes TaNAC ont été étudiés afin d'identifier de nouveaux gènes impliqués dans le développement et/ou la réponse à la sécheresse chez le blé tendre.

5.2. Des gènes TaNAC montrent une expression préférentielle dans le grain ou la feuille

Les gènes NAC sont largement impliqués dans de nombreux processus de croissance et développement, tels que la différenciation des principaux organes (Mathew et Agarwal 2018). Notre étude de 23 gènes TaNAC au sein de blé en plein champ a mis en évidence des gènes dont l'expression était au moins 10 fois supérieure dans la feuille par rapport au grain (7 gènes) ou au moins 10 fois supérieure dans le grain par rapport à la feuille (5 gènes), quels que soient le génotype, le stade de développement ou les conditions de culture appliquées (Tableau 29). Ce constat suggère une implication spécifique de ces gènes dans le développement de la feuille ou du grain. Ainsi, le gène TraesCS7A01G305200.1, qui est exprimé préférentiellement dans les feuilles correspond au gène TaNAC-S décrit par Zhao *et al.* (2015). Dans ce travail, les

auteurs ont mis en évidence un phénotype stay-green et une augmentation de la concentration en azote dans les grains de blé provenant de lignées sur-exprimant la protéine TaNAC-S. Les auteurs ont conclu que ce facteur de transcription, principalement exprimé dans les feuilles, pouvait contribuer à retarder la sénescence foliaire post-anthèse (phénotype stay-green) grâce à une meilleure remobilisation des nutriments au cours de la sénescence. Un autre gène exprimé principalement dans les feuilles dans notre expérimentation (TraesCS4D01G176000.1) est l'homologue de RhNAC100 chez la rose, caractérisé comme répresseur de l'expansion cellulaire des pétales de rose (Pei *et al.* 2013), ce qui suggère son implication dans la détermination de taille des cellules.

Parmi les 5 gènes TaNAC exprimés préférentiellement dans le grain, les gènes TraesCS7A01G569100.1 et TraesCS5A01G245900.2 sont homologues à ONAC20 et ONAC026 décrits chez le riz par Mathew *et al.* (2016). Dans cette étude, les auteurs ont montré que les 3 gènes ONAC020, ONAC026 et ONAC023 sont exprimés spécifiquement et très fortement lors du développement du grain, suggérant un rôle potentiel de ces gènes, *via* leur capacité d'hétéromérisation, dans la régulation de la taille et du poids du grain de riz. Un autre gène, TraesCS3A01G077900.1, s'est avéré être l'homologue de HvNAC019 chez l'orge, lui aussi préférentiellement exprimé dans le grain et qui pourrait en réguler le développement en association avec 2 autres gènes (HvNAC017 et HvNAC018) (Christiansen *et al.* 2011).

La caractérisation de ces gènes à l'expression fortement liée à un organe pourrait améliorer notre compréhension des mécanismes de régulation du développement chez le blé tendre. La production de plantes transgéniques présentant une expression modifiée du gène candidat est une technique de validation fonctionnelle largement utilisée.

5.3. La réponse des TaNAC à la sécheresse a lieu principalement dans les feuilles, durant la phase de remplissage du grain

L'analyse des données transcriptomiques de Liu *et al.* (2015) ont montré que l'expression de 12 gènes est impactée par l'application du multi-stress « choc thermique x sécheresse », mais pas par un stress thermique ou sécheresse seul. Chez Arabidopsis, Rasmussen *et al.* (2013) ont montré que 61% des modifications transcriptomiques à un multi-stress étaient différentes de celles observées lors des traitements mono-stress. La réponse à un stress thermique et sécheresse ne serait donc pas qu'une addition des réponses aux stress thermique d'une part et

hydrique d'autre part, mais créerait une synergie qui ferait intervenir des mécanismes de régulation spécifiques. Cependant cette étude pourrait être complétée et confortée :

- Les résultats ont été obtenus sur la seule variété TAM107. Il serait intéressant alors de vérifier si des réponses similaires pourraient être observées chez d'autres variétés.
- Cette étude a eu lieu sur du blé au stade plantule et n'a donc pas permis d'étudier la réponse aux stress thermique et sécheresse des gènes exprimés dans d'autres organes et tissus comme l'épi ou le grain,
- Et pour finir, les plantules ont été cultivées en chambre de culture, ce qui permet un contrôle fin des paramètres environnementaux mais ce qui peut conduire à des résultats différents d'une étude en plein champ du fait, justement, d'une exposition différente des plantes aux stress biotiques et abiotiques environnementaux.

Dans cette optique, nous avons étudié l'expression de 23 gènes chez un autre génotype en réponse à un stress sécheresse en plein champs. Il est très intéressant de noter que, parmi les 23 gènes sélectionnés, 11 sont surexprimés en réponse à un choc thermique de 40°C selon les données transcriptomiques de Liu *et al.* (2015). Et parmi ces 11 gènes, l'expression de 9 gènes est augmentée par la sécheresse en plein champ dans les plants de blé au stade reproductif alors qu'elle ne l'est pas significativement par le stress hydrique au stade plantule (Supplemental Table S6 de la publication Guérin *et al.* (2019)). Nous pouvons considérer ce résultat comme un argument appuyant l'idée que les plantes soumises à un stress sécheresse en plein champs répondent aussi à une contrainte thermique. Une autre possibilité serait que la réponse de la plante à un même stress soit dépendante de son stade de développement.

Seize gènes ont été trouvés différentiellement exprimés dans les feuilles et les grains entre le stade de développement précoce du grain (220°Cj, division cellulaire) et son stade de remplissage (450°Cj), la plupart voyant leur expression augmentée à 450°Cj (15 gènes). Il est connu que le stress induit par la sécheresse pendant le remplissage des grains stimule la sénescence des feuilles et améliore la remobilisation des glucides hydrosolubles dans les grains (Zhang *et al.* 2014). Or, nous avons trouvé un ensemble de 8 gènes TaNAC qui, en réponse à la sécheresse, sont fortement induits dans les feuilles au stade de remplissage du grain. Ces gènes pourraient participer à la remobilisation des ressources des feuilles aux grains. Par exemple, les 3 membres du même groupe homéologue situés sur le chromosome 5 (TraesCS5A01G143200.1 (Figure 65), TraesCS5B01G142100.1 et TraesCS5D01G148800.1) correspondent au gène TaNAC69. Ce gène a bien été caractérisé en tant que régulateur du stress hydrique, et il a été démontré qu'il améliorerait le rendement en grain (Xue *et al.* 2011).

Cependant, chez le blé, l'expression de TaNAC69 a été trouvée principalement dans les racines par rapport aux feuilles (Xue *et al.* 2011 ; Baloglu *et al.* 2014 ; Chen *et al.* 2016a). Il est intéressant de noter que, dans notre étude, la copie A du gène a été trouvée exprimée à un niveau plus élevé dans les feuilles que les copies B et D homéologues. Cette observation suggère que les copies B et D pourraient s'exprimer dans la racine, expression qui n'a pas été quantifiée dans notre étude. En effet, des variations dans le profil d'expression des membres d'un groupe d'homéologie ont déjà été observées chez le blé (Leach *et al.* 2014 ; Liu *et al.* 2015). Le gène homologue de TraesCS5A01G143200.1, TraesCS5B01G142100.1 et TraesCS5D01G148800.1 chez le riz (OsNAC10) est également impliqué dans l'adaptation à la sécheresse, principalement en augmentant les niveaux d'expression des gènes stress-protecteurs, comme la chitinase, les gènes ZIM et glyoxalase I. La surexpression d'OsNAC10 augmente considérablement le rendement en grain dans des conditions de sécheresse en plein champ (Jeong *et al.* 2010). Nous pouvons supposer que les gènes homologues de OsNAC10 chez le blé joueraient un rôle similaire, et il serait donc intéressant de les y caractériser.

Parmi les 23 gènes suivis en réponse à la sécheresse en plein champ, 3 gènes (TraesCS5D01G148800.1, TraesCS5B01G142100.1, TraesCS3B01G407600.1) présentent une augmentation de leur expression en réponse à un stress hydrique selon les données transcriptomiques de Liu *et al.* (2015). Rappelons que ce stress hydrique est induit par ajout de PEG dans le milieu de culture de plantules du génotype TAM107 de blé tendre. Or, sur les 2 génotypes (Altigo et Allez-Y) dont la réponse à la sécheresse en plein champ a été testée en phase reproductive, ces 3 gènes sont également surexprimés. Ainsi, le profil d'expression de ces gènes est similaire entre un mono-stress hydrique contrôlé en serre et un stress sécheresse en plein champs, quel que soit le stade d'application testé. Ils présentent donc probablement un rôle important dans la réponse de la plante à ce type de stress.

Parmi les 23 gènes suivis, 2 membres du même groupe homéologue (TraesCS4D01G176000.1, TraesCS4B01G174000.1) sont des homologues du gène de riz ONAC39 (OSCUC1/ONAC039), dont l'expression est induite par un stress hydrique et un choc thermique (Sun *et al.* 2015 ; Ahn *et al.* 2017). Les membres de ce groupe d'homéologie chez le blé pourraient présenter une réponse similaire et être aussi associés aux mécanismes de réponse aux stress hydrique et thermique.

Un autre gène (TraesCS2A01G101100.1, Figure 63) est régulé positivement en réponse à la sécheresse en plein champ dans les feuilles de blé à 450°Cj. Il a déjà été décrit sous le nom de

TaGRAB2 (Xie *et al.* 1999). Les auteurs, qui l'ont trouvé impliqué dans la réponse au stress biotique (un virus à ADN phytopathogène, Geminiviridae), ont constaté en revanche qu'il présentait un niveau d'expression à peine détectable dans les feuilles. Ces résultats suggèrent que le même gène pourrait être exprimé ou non dans les feuilles, en fonction du stress appliqué.

Enfin, aucun homologue caractérisé n'a été trouvé dans la littérature pour TraesCS3B01G303800.1 et TraesCS6D01G390200.1. Ces gènes peuvent être des candidats intéressants pour la tolérance à la sécheresse sur le terrain et nécessitent des investigations supplémentaires.

5.4. Quatre gènes TaNAC exprimés dans le grain ont une réponse génotype-dépendante à la sécheresse et à la température

Cinq gènes TaNAC (TraesCS3A01G077900.1, TraesCS5A01G245900.2, TraesCS7A01G569100.1, TraesCS7B01G056300.1 et TraesCS7B01G100300.1) présentent une expression préférentielle dans le grain de blés cultivés en plein champ. Fait intéressant, leur comportement en réponse à la sécheresse varie en fonction du génotype. Leur expression diminue en réponse au traitement sécheresse au stade précoce de développement du grain (division cellulaire) dans le génotype qui présente le plus grand nombre d'épis/m² (Altigo, Figure 62) ; alors que dans l'autre génotype (Allez-y), elle est diminuée (pour 4 gènes) ou augmentée (pour TraesCS7A01G569100.1) plus tardivement au cours de la phase de remplissage du grain (Figure 65-B et 65-C). Ces résultats suggèrent que, selon le patrimoine génétique, ces gènes peuvent se présenter sous différents allèles pouvant être liés au niveau de tolérance de la plante à la sécheresse.

L'expression de ces 5 gènes en réponse à la sécheresse variant entre les génotypes, il serait extrêmement intéressant :

- De séquencer ces gènes dans les 2 génotypes étudiés, afin de constater une éventuelle variation allélique de leurs séquences codantes et régulatrices, qui pourrait ensuite être reliée à la variabilité de leur expression,
- De faire de même pour un panel d'autres génotypes.

En effet, si la variabilité allélique de ces gènes est corrélée avec la réponse du blé tendre à la sécheresse, alors ils pourront être très intéressants à prendre en compte dans les programmes de sélection des futures variétés élites de blé tendre.

De plus, Girousse *et al.* (2018) ont montré que 4 de ces 5 gènes (TraesCS7A01G569100.1, TraesCS5A01G245900.2, TraesCS7B01G056300.1 et TraesCS7B01G100300.1) sont exprimés différemment dans les grains du génotype SerixBabax 49 en réponse à un stress thermique modéré de + 8°C appliqué pendant la phase de division cellulaire du grain. Ces résultats suggèrent que ces 4 gènes TaNAC pourraient être impliqués dans la réponse à plusieurs types de stress abiotiques (augmentation modérée de la température et sécheresse), comme c'est le cas pour d'autres gènes NAC tels que MINAC9 chez *Miscanthus*, SINAC3 chez la tomate, ONAC022 chez le riz, GmNAC3 et GmNAC4 chez le soja, impliqués dans plusieurs stress environnementaux (Pinheiro *et al.* 2009 ; Han *et al.* 2012 ; Hong *et al.* 2016 ; Zhao *et al.* 2016). Ces gènes sont rapportés pour la première fois dans la littérature chez le blé tendre. Sachant qu'ils sont préférentiellement exprimés dans le grain et que leur expression est modulée en réponse à au moins 2 types de stress abiotiques (température modérée et sécheresse en plein champ) nécessitant la mise en place de mécanismes d'adaptation, nous avançons l'hypothèse qu'ils interviendraient dans ces mécanismes. Leur caractérisation pourrait présenter un intérêt certain pour les programmes de sélection visant l'amélioration de ces processus adaptatifs aux stress environnementaux.

Le dernier chapitre de ma thèse est consacré à la caractérisation fonctionnelle de ces 4 gènes NAC au sein du genre *Triticum*.

6. Les idées clés

- L'analyse des données transcriptomiques disponibles publiquement ont permis de montrer que :
 - parmi les 488 membres de la famille TaNAC, 52 gènes présentent une expression préférentielle à un organe. Par exemple, 19 gènes sont exprimés préférentiellement dans le grain,
 - le profil d'expression d'un tiers des gènes TaNAC est modifié par un choc thermique et/ou un stress sécheresse au stade plantule. Certains sont exprimés spécifiquement en réponse aux 2 stress cumulés, ce qui appuie l'hypothèse d'une synergie entre les 2 mécanismes de réponse.
- L'expression de 23 gènes suivie *in planta* chez 2 génotypes en réponse à une sécheresse en plein champ a montré que :
 - la majorité sont exprimés préférentiellement dans les feuilles (16 gènes) et/ou à 450°Cj (15 gènes),
 - les gènes exprimés préférentiellement dans les feuilles répondent fortement à la sécheresse à 450°Cj dans les 2 génotypes testés,
 - les gènes exprimés préférentiellement dans les grains montrent un profil de réponse moins marqué à la sécheresse mais différent entre les 2 génotypes.
- Parmi les 23 gènes suivis *in planta* chez le blé tendre en réponse à différents stress abiotiques, 5 gènes sont exprimés préférentiellement dans le grain,
 - leur expression, très faible dans la feuille, diminue encore après 6h de choc thermique,
 - leur réponse à un stress sécheresse en plein champ est positive et génotype-dépendante,
 - un stress thermique modéré appliqué à partir de l'anthèse impacte le profil d'expression de 4 de ces 5 gènes exprimés préférentiellement dans le grain pour le génotype SerixBabax 49 dont 2 ont également été trouvés impactés dans un deuxième génotype, SerixBabax 139.

Chapitre III

CARACTERISATION DE 4 GENES

TaNAC CHEZ LE BLE

et transfert au petit épeautre

Tableau 30 : Correspondance entre la nomenclature de l'IWGS RefSeq v1.0 et celle utilisée dans ce chapitre de thèse. *TaNAC1, TaNAC2 et TaNAC4 sont situés sur le chromosome 7 tandis que TaNAC3 est ancré sur le chromosome 5. Par convention et par exemple, la copie du gène TaNAC1 portée par le génome A du blé tendre *Triticum aestivum* sera notée « TaNAC1-7A ». Les copies B et D du TaNAC3 n'ont pas pu être identifiées.*

	Génome A	Génome B	Génome D
TaNAC1	TraesCS7A01G569100	TraesCS7B01G489500	TraesCS7D01G543500
TaNAC2	TraesCS7A01G152500	TraesCS7B01G056300	TraesCS7D01G154200
TaNAC3	TraesCS5A01G245900	Absent	Absent
TaNAC4	TraesCS7A01G194700	TraesCS7B01G100300	TraesCS7D01G196300

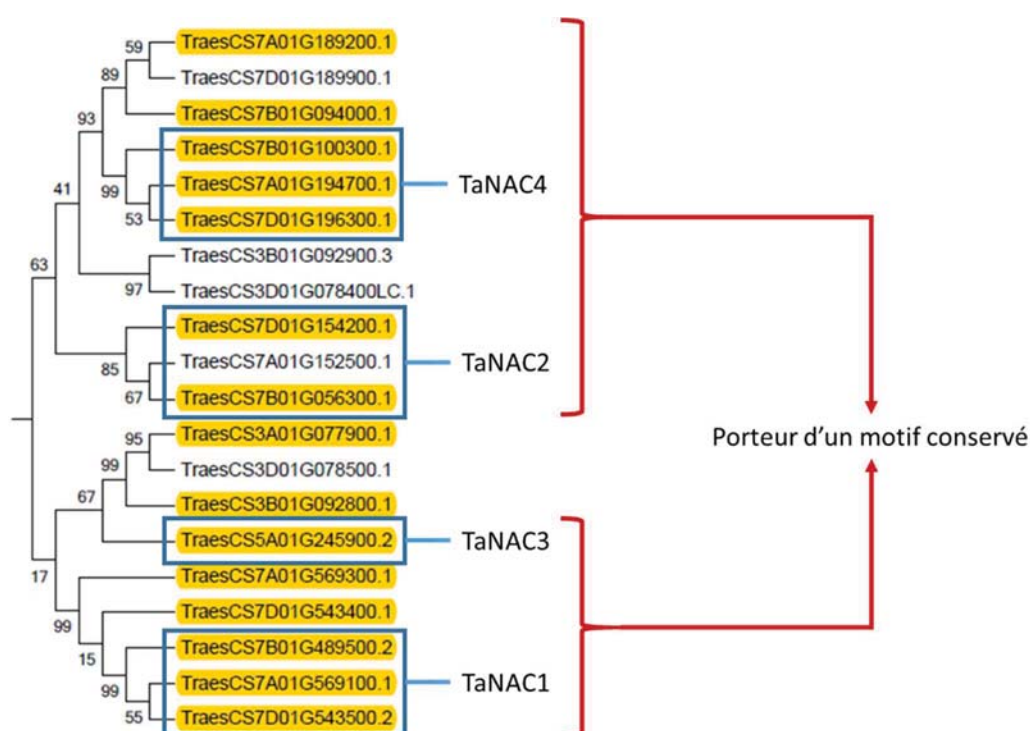


Figure 68 : Un clade extrait de l'arbre phylogénétique portant les 488 séquences de la famille TaNAC du blé tendre. *Les homéologues de TaNAC1, TaNAC2, TaNAC3 et TaNAC4 existant dans cette base de données sont encadrés en bleu. Parmi les 19 séquences exprimées préférentiellement dans le grain, 15 sont portées par ce clade (surlignées en jaune). Parmi les 18 séquences porteuses d'un motif conservé spécifique, 17 sont présentes dans ce clade (accolade rouge).*

Avant-propos :

Au sein de la famille TaNAC, un clade spécifique est porteur de gènes qui s'expriment principalement dans le grain de blé. Parmi ces gènes, 4 membres possèdent une expression modulée au cours du développement du grain, et impacté par un stress thermique modéré et par un stress sécheresse. Ils n'ont encore jamais été étudiés chez le blé tendre, nous avons donc choisi de centrer le dernier chapitre de cette thèse sur leur étude ; afin d'avancer des hypothèses quant à leurs rôles au cours du développement normal du grain de blé et dans sa réponse aux stress. Dans la suite de ce chapitre et pour plus de clarté, ces 4 gènes ont été conventionnellement nommés TaNAC1, TaNAC2, TaNAC3 et TaNAC4 chez le blé tendre (Tableau 30).

1. Etude des gènes TaNAC exprimés préférentiellement dans le grain

1.1. Structure génomique

Parmi les gènes identifiés dans le Chapitre II – Partie 1 comme ayant une expression préférentielle à un organe, 19 gènes sont exprimés préférentiellement dans le grain. Quinze de ces gènes sont regroupés dans un clade classé dans la sous-famille NAC-d et contenant 20 séquences au total (Figure 68). Les 4 gènes restants sont répartis dans les sous-familles NAC-g (TraesCS2A01G326600.1), NAC-h (TraesCS4B01G132500.1 et TraesCS3A01G486500.1) et NAC-j (TraesCS7A01G152400.1) de l'arbre phylogénétique.

Le clade porteur de la majorité des gènes NAC exprimés préférentiellement dans le grain a été étudié plus particulièrement. Les 20 gènes qui le constituent sont en fait structurés en 9 groupes d'homéologie, dont 5 sont complets et 4 incomplets (seulement 1 ou 2 membres) (Figure 68). Il porte notamment les 4 gènes TaNAC1, TaNAC2, TaNAC3 et TaNAC4 identifiés dans le Chapitre II. Evolutivement, il est possible que TraesCS7B01G094000.1 / TraesCS7D01G189900.1 / TraesCS7A01G189200.1 soit un groupe d'homéologie lié à TaNAC4 par une relation de duplication puisque leurs homéologues présentent des séquences très proches. De la même manière, la séquence de TraesCS3A01G077900.1 / TraesCS3D01G078500.1 / TraesCS3B01G092800.1 semble être très proche de TaNAC3.

Les séquences TaNAC génomiques de ce clade ont été collectées sur le site de l'URGI (<http://wheat-urgi.versailles.inra.fr/>). Elles présentent une structure exon-intron similaire : les

20 gènes possèdent tous 2 exons et 1 intron. Cette configuration ne représente pourtant que 19% des membres de la famille TaNAC, la majorité (44%) des séquences possédant 3 exons et 2 introns. La taille de l'intron est variable entre les séquences, elle mesure entre 107 et 131 pb pour la majorité des séquences, mais 5 d'entre elles présentent un intron mesurant entre 563 et 567 pb.

1.2. Recherche d'un motif conservé

Un domaine de régulation de l'activité transcriptionnelle est situé dans la partie C-terminale des protéines NAC, il est cependant beaucoup moins conservé que le domaine NAC et, par conséquent, peu décrit. Afin de rechercher des similarités de séquence entre les membres de la famille TaNAC pouvant traduire la présence de ce domaine, des motifs conservés ont été recherchés parmi les 488 séquences protéiques de cette famille à l'aide de la plateforme MEME suite (<http://meme-suite.org/tools/meme>). Ceci a permis d'identifier un motif de 70 aa conservé entre 18 séquences avec une e-value de $1,8e^{-491}$, dont 17 membres appartiennent à ce clade contenant des gènes exprimés préférentiellement dans le grain (Figure 68, Figure 69). La dernière séquence n'est pas dans le même clade mais elle fait partie des gènes présentant une expression préférentielle dans le grain (TraesCS7A01G152400). Tous les homéologues connus de TaNAC1, TaNAC2, TaNAC3 et TaNAC4 sont porteurs de ce motif.

2. Clonage des homéologues de TaNAC1, TaNAC2, TaNAC3 et TaNAC4

La taille de l'intron est conservée entre les homéologues de TaNAC2 (110 pb), TaNAC3 (110 pb) et TaNAC4 (113 pb), mais ce n'est pas le cas pour TaNAC1 dont les introns des 3 homéologues mesurent entre 565 et 567 pb (Tableau 31). En théorie, chaque gène du blé tendre possède 3 homéologues. En pratique, il est courant que les 3 homéologues ne soient pas identifiés pour tous les gènes, du fait de l'évolution constante des génomes qui conduit parfois à des insertions/délétions ou de fortes divergences entre les séquences d'un même groupe d'homéologie. Dans la base de données IWGS RefSeq v1.0, les 3 homéologues de TaNAC1, TaNAC2 et TaNAC4 ont pu être identifiés (Tableau 30). Concernant TaNAC3, seule la copie portée par le génome A possède un identifiant au sein de cette base de données (Figure 70). Une base de données plus ancienne générée par l'IWGS

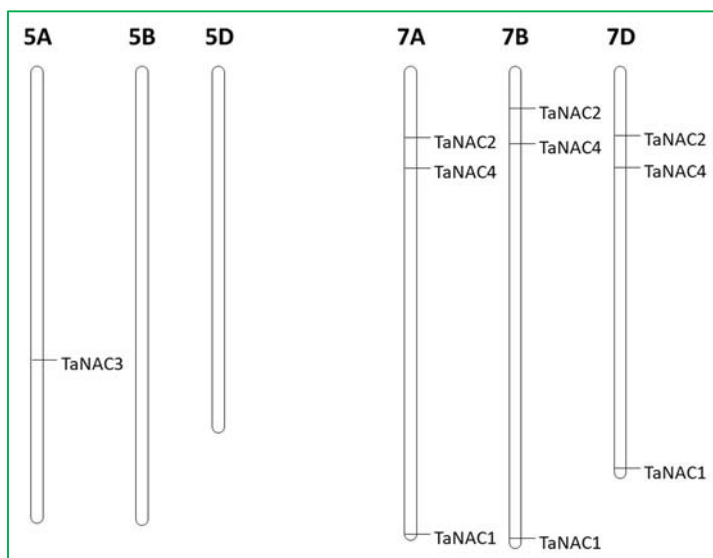


Figure 70 : Ancrage des homéologues de TaNAC3 sur le chromosome 5 et TaNAC1, TaNAC2 et TaNAC4 sur le chromosome 7.

Tableau 32 : Taille des séquences génomiques et codantes (ADNc) récupérées de la base de données IWGS RefSeq 1.0, et des séquences amplifiées in planta. La dernière colonne présente le pourcentage d'identité entre les séquences issues de la base de données IWGS RefSeq v1.0 et les séquences obtenues par séquençage in planta. Celui-ci est plus faible pour les séquences TaNAC2 car un décalage du cadre de lecture conduit à des protéines de taille très différentes mais l'identité de la partie commune entre ces 2 séquences est de 99%.

	Séquence génomique IWGS (pb)	ADNc IWGSC (pb)	ADNc in planta (pb)	Identité entre ADNc IWGS et ADNc in planta
TaNAC1				
7A	1708	1143	1107	1106/1143 (96.8%)
7B	1610	1044	Pas d'amplification	
7D	1710	1143	1101	1100/1143 (96.2%)
TaNAC2				
7A	1181	1071	879	877/1071 (81.9%)
7B	1184	1074	939	936/1074 (87.2%)
7D	1181	1071	936	934/1071 (87.2%)
TaNAC3				
5A	1094	987	939	937/987 (94.9%)
5B	Pas d'identifiant	Pas d'identifiant	360	
5D	Pas d'identifiant	Pas d'identifiant	Pas d'amplification	
TaNAC4				
7A	1163	1050	1050	1041/1050 (99.1%)
7B	1085	972	Pas d'amplification	
7D	1163	1050	Pas d'amplification	

(*Triticum_aestivum*.IWGSC1.0+popseq.30) présente un identifiant pour le TaNAC3-5B : Traes_5BL_59AFA4B01. Cette séquence n'étant pas correctement annotée (pas de codon start et stop en début et fin de séquence), nous l'avons alignée avec la séquence amplifiée *in planta* et avec la séquence génomique de la région du chromosome 5B que l'on supposait porteur de ce gène non-annoté. Par comparaison, nous avons pu reconstruire les séquences génomique et codante de ce gène.

Des amorces amplifiant les gènes complets ont été dessinées à partir de ces séquences « putatives », dans le but d'amplifier *in planta* les 3 homéologues de chacun des 4 TaNAC étudiés chez le génotype SerixBabax 49 de blé tendre, utilisé par Girousse *et al.* (2018). Après séquençage des produits d'amplification, la séquence codante de 8 des 12 homéologues recherchés a pu être obtenue (Tableau 32). Les séquences amplifiées sont plus courtes que les séquences inférées par analyse bioinformatique, mais elles restent très similaires (minimum 81,9% d'identité). Aucun identifiant ni aucune séquence biologique n'ont pu être trouvés pour TaNAC3-5D. Il est possible que le gène TaNAC3 ait, au cours de son évolution, perdu l'un de ses homéologues, ou que la séquence de celui-ci ait tellement divergé des copies A et B qu'elle n'ait pas pu être amplifiée. De la même manière, la copie TaNAC3-5B apparaît largement plus courte (360 pb) que son homéologue TaNAC3-5A (939 pb).

3. Variabilité de la séquence des 4 TaNAC et association avec un caractère phénotypique

3.1. Variabilité de la séquence des 4 TaNAC étudiés

Aucune variabilité de séquence n'a été observée dans la base de données **Whealbi** pour les 10 homéologues des 4 gènes TaNAC dont nous possédons la position sur la pseudomolécule de blé tendre. Cependant, à partir des données de variabilité génétique obtenues avec la puce **BW_420K**, des informations concernant leur diversité allélique ont pu être collectées. Les SNP obtenus sont majoritairement localisés dans les séquences codantes (les exons) des homéologues de ces 4 TaNAC (Tableau 33). L'homéologue TaNAC4-7A présente la plus grande variabilité (12 SNP) tandis qu'aucune variabilité n'a été identifiée sur les séquences TaNAC2-7A et TaNAC4-7D. Par une contrainte de temps, nous n'avons pas pu obtenir d'information concernant la nature des SNP identifiés, nous ne pouvons donc pas prédire si ces SNP sont silencieux ou s'ils ont un impact sur la conformation ou le fonctionnement de la protéine produite.

Tableau 33 : Nombre de SNP identifiés dans la partie 5'UTR (1500pb en amont de l'ATG sur la séquence génomique), les 2 exons, l'intron, et la partie 3'UTR (1500pb en aval du codon stop) pour les homéologues disponibles des 4 gènes TaNAC étudiés. Les données sont issues de la puce BW_420K.

Nombre de SNP	5'UTR (1500 pb en amont du gène)	1er Exon	Intron	2ème Exon	3'UTR (1500 pb en aval du gène)	Total
NAC1						
TraesCS7A01G569100	0	2	1	3	2	8
TraesCS7B01G489500	1	2	5	2	1	11
TraesCS7D01G543500	3	1	0	1	0	5
NAC2						
TraesCS7A01G152500	0	0	0	0	0	0
TraesCS7B01G056300	3	1	0	1	2	7
TraesCS7D01G154200	1	2	2	0	0	5
NAC3						
TraesCS5A01G245900	1	1	0	1	0	3
NAC4						
TraesCS7A01G194700	3	2	2	4	1	12
TraesCS7B01G100300	4	1	1	2	1	9
TraesCS7D01G196300	0	0	0	0	0	0

Tableau 34 : Extrait du tableau SNP_BW_CompoProt présenté en Chapitre II, ne présentant que les données d'association obtenues en rapport avec les 4 TaNAC étudiés.

Phénotype	Gène concerné	Position	Score
Pourcentage de gluténine dans l'azote total (Le Moulon faible azote)	TaNAC1-7B	743457269	3.900
Quantité de w1,2-gliadines par grain (Le Moulon fort azote)	TaNAC2-7B	59180076	3.080
	TaNAC2-7B	59180596	3.018

3.2. Association SNP x phénotypes

Parmi les SNP identifiés chez les différents homéologues des 4 TaNAC, seuls 3 ont montré une association significative avec un caractère phénotypique.

Un SNP chez TaNAC1-7B montre une association avec le pourcentage d'azote alloué aux gluténines dans la culture réalisée sur le site du Moulon en condition faible azote. Ce SNP est localisé au milieu de l'intron central ; et son association avec le phénotype peut être considérée comme robuste puisque son score de 3,9 passe largement le seuil fixé à 3 (Tableau 34). Deux SNP chez TaNAC2-7B montrent une association avec la quantité d' Ω -gliadine 1-2 dans la culture réalisée au Moulon en condition fort azote. Le SNP en position 59180076 est localisé dans le premier exon de la séquence tandis que le SNP en position 59180596 se trouve dans la partie 5'UTR du gène. Leur association avec le phénotype est juste au-dessus du seuil de détection (3,02 et 3,08 respectivement). Nous pouvons supposer que la mutation présente en 5' UTR pourrait altérer la fixation des régulateurs de la transcription du gène. Étonnamment, aucun des autres SNP contenus dans ces 2 séquences n'a montré d'association significative avec ces mêmes caractères phénotypiques. Il est possible que ces autres SNP soient des mutations silencieuses ou à faible impact, qui ne modifient pas la séquence protéique au point d'altérer significativement la fonction du gène touché. Cependant, parmi les autres SNP identifiés sur la séquence TaNAC2-7B, 5 SNP (localisés en 5'UTR, dans le 2^{ème} exon et en 3'UTR) ont montré une association pour le même paramètre, avec un score compris entre 2,50 et 2,95 donc sous le seuil de significativité. Nous pourrions néanmoins les qualifier de sub-significatifs.

4. Localisation subcellulaire des 4 TaNAC

Des cassettes exprimant une protéine de fusion du gène avec le gène rapporteur YFP ont été produites pour chacun des 4 TaNAC étudiés, puis clonées dans le plasmide pEG104 sous le contrôle d'un promoteur de forte expression. Des souches d'*A. tumefaciens* ont ensuite été transformées avec ce plasmide et infiltrées dans l'épiderme inférieur de feuilles de jeunes plants de tabacs. Après infiltration, ceux-ci présentent une expression transitoire des protéines TaNAC1, TaNAC2, TaNAC3 ou TaNAC4 dont il est possible d'observer la localisation subcellulaire par microscope optique à épifluorescence. L'homéologue D des gènes TaNAC1 et TaNAC2 et l'homéologue A des gènes TaNAC3 et TaNAC4 ont été utilisés car c'étaient les

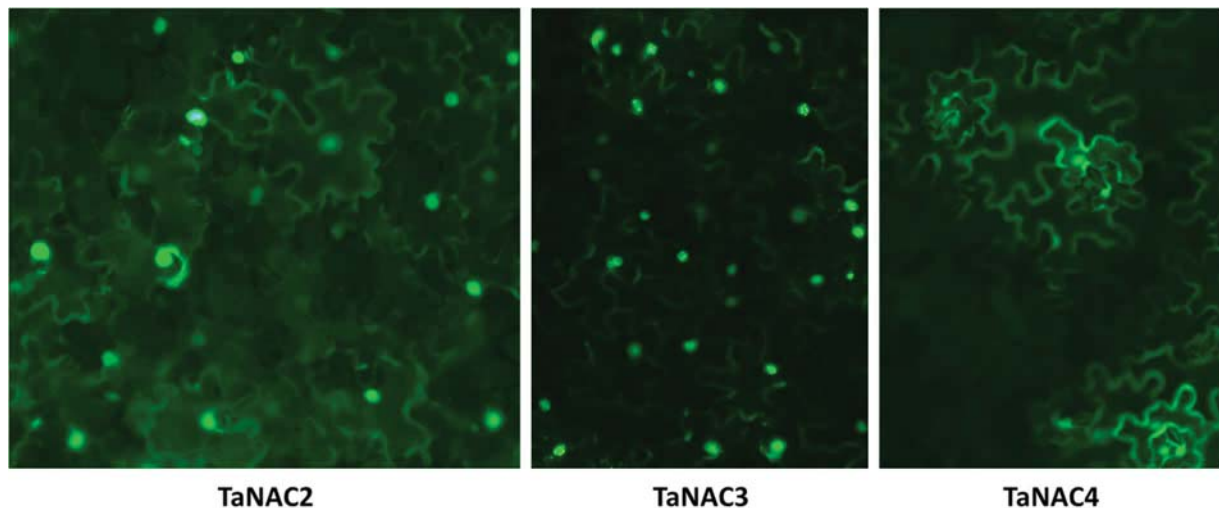


Figure 71 : Agro-infiltration de feuilles de tabac avec des bactéries *Agrobacterium tumefaciens* transformées par des plasmides surexprimant des protéines chimères TaNAC-YFP. Les cellules de feuille de tabac possèdent une forme polylobée caractéristique. TaNAC2 et TaNAC3 montrent une localisation nucléaire tandis que TaNAC4 présente une localisation nucléaire et cytoplasmique (la vacuole, très volumineuse, colle le cytoplasme à la membrane cellulaire). Les expérimentations menées sur TaNAC1 n'ont pas permis d'obtenir des images concluantes quant à la localisation des protéines.

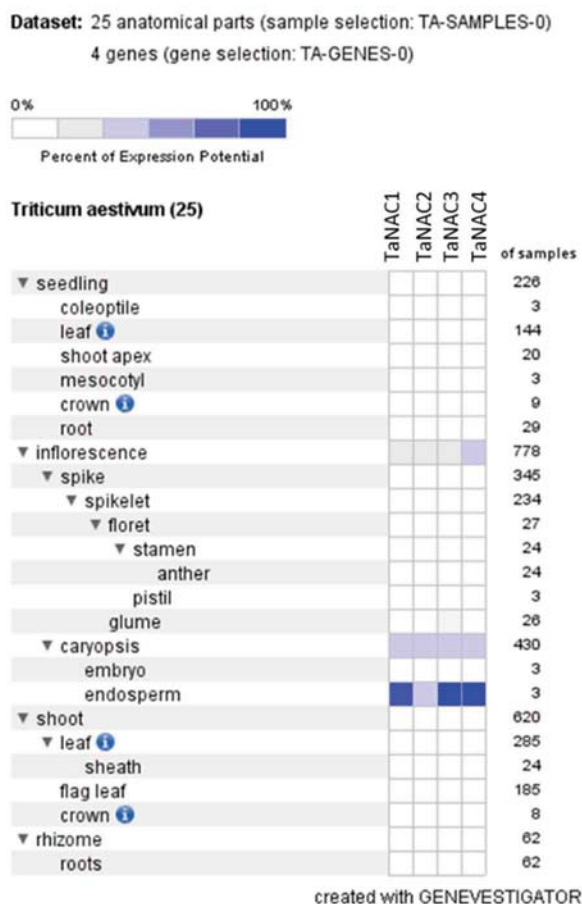


Figure 72 : Expression des gènes TaNAC1, TaNAC2, TaNAC3 et TaNAC4 dans 25 organes du blé en utilisant le site **Genevestigator** (<https://genevestigator.com/gv/>). Pour un gène et un tissu donnés, le niveau d'expression est représenté en % du niveau maximal trouvé parmi l'ensemble des résultats microarray pour ce gène. Les 4 gènes TaNAC semblent exprimés dans le grain (caryopse) et préférentiellement dans l'albumen (noté endosperm).

seuls que nous avons amplifié au moment de l'expérimentation. Les résultats n'ont pas été concluants pour TaNAC1. TaNAC2 et TaNAC3 présentent une localisation nucléaire, tandis que TaNAC4 montre une localisation nucléaire et cytoplasmique (Figure 71).

5. Expression des 4 TaNAC au cours du développement du blé tendre

L'expression des 4 TaNAC a été suivie au cours des différents stades de développement de la plante et du grain dans le cadre de plusieurs expérimentations. De plus, les données bioinformatiques disponibles offrent des renseignements quant à la réponse de ces TaNAC à un large panel de conditions environnementales.

5.1. Expression dans le grain

Les données bioinformatiques issues de la plateforme expVIP (cf. Chapitre II – Partie 1) ont montré que, mis à part TaNAC2-7A, les homéologues des 4 TaNAC font partie des 19 gènes ayant une expression préférentielle dans le grain. Ce résultat a été corroboré par les données issues de la plateforme Genevestigator, qui montrent une expression des 4 TaNAC localisée dans le grain (caryopse) et plus précisément dans le compartiment albumen (Figure 72).

Nous avons ensuite confirmé cette information *in planta* lors de l'expérimentation sécheresse en plein champ, dans laquelle l'expression d'un homéologue de chacun de ces 4 TaNAC a été suivie dans les grains et les feuilles de blé tendre. Ils ont toujours montré un niveau d'expression beaucoup plus élevé dans les grains par rapport aux feuilles, quelle que soit la condition environnementale testée (Figure 73). De la même manière, dans les échantillons de blé prélevés en réponse au choc thermique, nous avons constaté une expression très basse des 4 TaNAC étudiés dans les feuilles.

5.2. Implication des homéologues des 4 TaNAC au cours du développement du grain

Les échantillons obtenus dans le cadre du projet BreedWheat ont permis de suivre l'expression des 4 gènes TaNAC de manière plus précise au cours du développement et en réponse à un stress thermique modéré. Les profils d'expression globaux des 4 TaNAC au cours du

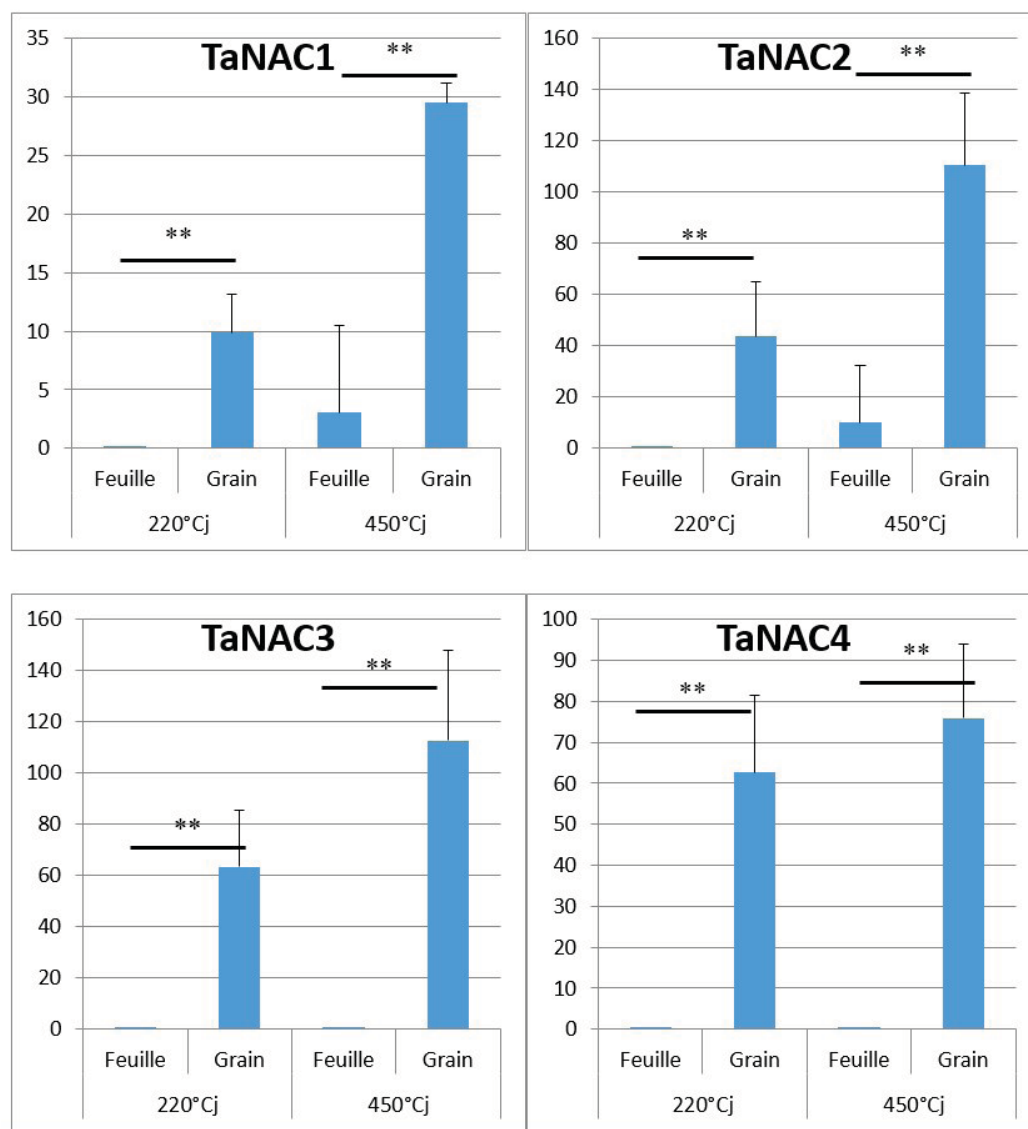


Figure 73 : Niveau d'expression de TaNAC1, TaNAC2, TaNAC3 et TaNAC4 dans des feuilles et des grains du génotype Altigo de blé tendre, à 2 stades en condition contrôle et en réponse à un stress sécheresse. Les valeurs d'expression relatives ont été calculées selon la méthode du $2^{-\Delta\Delta Ct}$, en utilisant le gène Ta2776 (Paolacci et al., 2009) comme gène de référence. Pour les 4 gènes étudiés, le profil d'expression est minimal dans les feuilles par rapport aux grains, quelques soient le stade étudié. **: association avec une p-value inférieure à 0,01 selon le test T de Student.

développement ont été obtenus par RT-qPCR grâce à des amorces généralistes (sans discrimination des homéologues) utilisées sur des échantillons de grain à différents stades en condition contrôle. Ces profils sont similaires : une première phase basale est observée de 0°Cj à 120°Cj, puis le niveau d'expression relatif des TaNAC par rapport au gène de référence exprimé constitutivement (Ta2776) augmente progressivement pour atteindre sa valeur maximale à 300°Cj, qui est le dernier stade mesuré (Figure 74).

Cependant, si le profil d'expression général est le même, le taux d'expression relative diffère entre les 4 TaNAC. Tous homéologues confondus, le gène TaNAC4 est le plus exprimé (valeur d'expression relative = $2^{\Delta\Delta Ct} = 68,6$), suivi par TaNAC2 (49,2), TaNAC1 (37,9) puis TaNAC3 dont l'expression est plus faible au cours du développement du grain (2,7). Ces profils sont cohérents avec l'analyse *in silico* réalisée à l'aide de la plateforme Genevestigator (<https://genevestigator.com/gv/>), montrant une expression basse durant la partie végétative du développement de la plante, qui augmente lorsque qu'elle entre en période reproductive (Figure 75).

De la même manière, le suivi indépendant de l'expression de chaque homéologue des 4 TaNAC a montré qu'ils présentaient une implication différente dans la réponse globale du gène considéré. Pour TaNAC1, TaNAC2 et TaNAC3 la copie B du gène est plus exprimée que les 2 autres. TaNAC4, lui, montre une plus forte expression de sa copie D ($2^{\Delta\Delta Ct} = 23,3$ pour la copie D contre 5,9 et 2,2 pour la copie B et A respectivement) (Figure 76).

6. Effet des stress abiotiques sur l'expression des 4 TaNAC chez le blé

Parmi les 23 gènes TaNAC dont l'expression a été suivie en réponse à la sécheresse au champ, 5 gènes ont été sélectionnés parce qu'ils s'exprimaient préférentiellement dans le grain. TaNAC1, TaNAC2, TaNAC3 et TaNAC4 appartiennent à ce groupe de 5 gènes. Les résultats ont montré que leur expression n'est pas autant impactée par la sécheresse que ne le sont les gènes exprimés préférentiellement dans les feuilles ; mais contrairement aux feuilles, cette réponse est génotype-dépendante (cf. Chapitre II – Partie 2.2.2.2). En outre, dans le cadre du projet BreedWheat, les résultats d'expression ont montré que les 4 gènes TaNAC étudiés répondent à une contrainte thermique modérée. Nous nous sommes donc concentrés sur l'étude plus fine de chacun de ces 4 gènes TaNAC en réponse au stress thermique modéré.

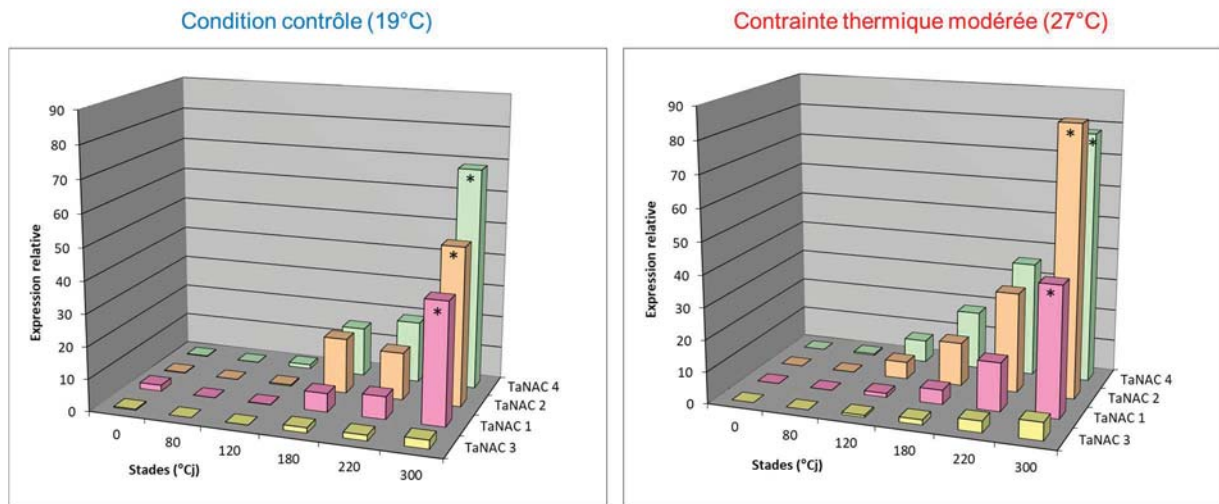


Figure 74 : Profil d'expression des 4 facteurs de transcription TaNAC sur 5 stades de développement du grain en condition contrôle (19°C, à gauche) et en réponse à une contrainte thermique modérée appliquée durant la phase de division cellulaire du grain de blé tendre (27°C à droite). Les valeurs d'expression relatives ont été calculées selon la méthode du $2^{-\Delta\Delta C_t}$, en utilisant le gène Ta2776 (Paolacci et al., 2009) comme gène de référence. Les étoiles indiquent une expression significativement différente entre stades (ANOVA suivi d'un test de Scheffé). Une cinétique de réponse identique est observée mais avec un niveau d'expression augmenté en réponse à la contrainte thermique.

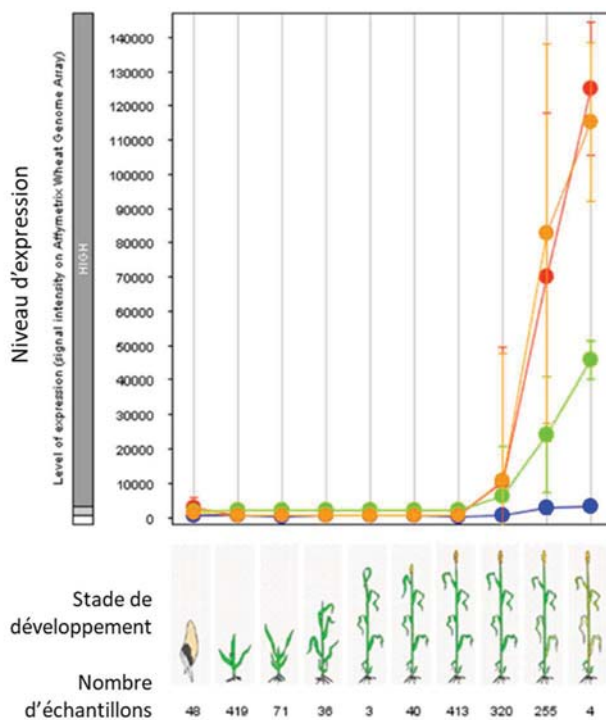


Figure 75 : Niveau d'expression des 4 gènes TaNAC au cours du développement du blé, selon les données bio-informatiques fournies par Genevestigator. Le graphique a été généré automatiquement par la plateforme Genevestigator. L'expression des 4 gènes, minimale dans la période végétative de la plante, augmente fortement lors de la période reproductive.

6.1. Implication de chaque TaNAC en réponse au stress thermique

De manière globale, le profil d'expression des 4 TaNAC est similaire en condition contrôle et en réponse au stress thermique modéré ; c'est-à-dire que l'expression est minimale dans les phases précoces, puis elle augmente durant la phase de division cellulaire du développement du grain pour finir par atteindre un plateau vers les 300°Cj (Figure 77) (voir Girousse *et al.* 2018). Mais la phase d'augmentation de l'expression semble plus précoce en réponse au stress thermique. Cependant, que ce soit en condition contrôle ou en réponse à la température, les expressions relatives entre chaque TaNAC restent similaires ; c'est-à-dire que TaNAC3 présente l'expression la plus faible, suivi de TaNAC1 tandis que TaNAC2 et TaNAC4 possèdent la plus forte expression (Figure 77).

Afin d'estimer plus finement l'intensité de réponse au stress, le ratio de l'expression en condition stressée sur l'expression en condition contrôle a été réalisé pour chaque stade et chaque gène (Figure 61). Cela a permis de mettre en évidence :

- un stade clé de réponse au stress à 120°Cj. Parmi les 5 stades testés, c'est à ce moment qu'apparaît la différence maximale entre conditions contrôle et stressée,
- une intensité de réponse au stress à 120°Cj qui est variable entre les 4 TaNAC. En effet, la réponse de TaNAC2 est la plus forte avec un ratio d'expression stressé/contrôle de 14,1, suivi de TaNAC3 (12,3) et TaNAC1 (8,9), tandis que TaNAC4 est le gène qui répond le moins au stress avec un ratio de 4,5.

Certains gènes montrent donc une forte expression au cours du développement mais peu de réponse au stress (comme TaNAC4) tandis que d'autres sont plus faiblement exprimés au cours du développement mais répondent au stress de manière importante (comme TaNAC3).

6.2. Implication des homéologues des 4 TaNAC en réponse au stress thermique

Le ratio d'expression stressé/contrôle a été réalisé pour chaque homéologue de chacun des 4 TaNAC. Les copies homéologues des gènes TaNAC1, TaNAC2 et TaNAC3 présentent un comportement similaire : l'homéologue appartenant au génome B est le plus exprimé, suivi par l'homéologue D et enfin l'homéologue du génome A. En revanche, l'homéologue D du gène TaNAC4 est largement plus exprimé que les homéologues B et A (Figure 78).

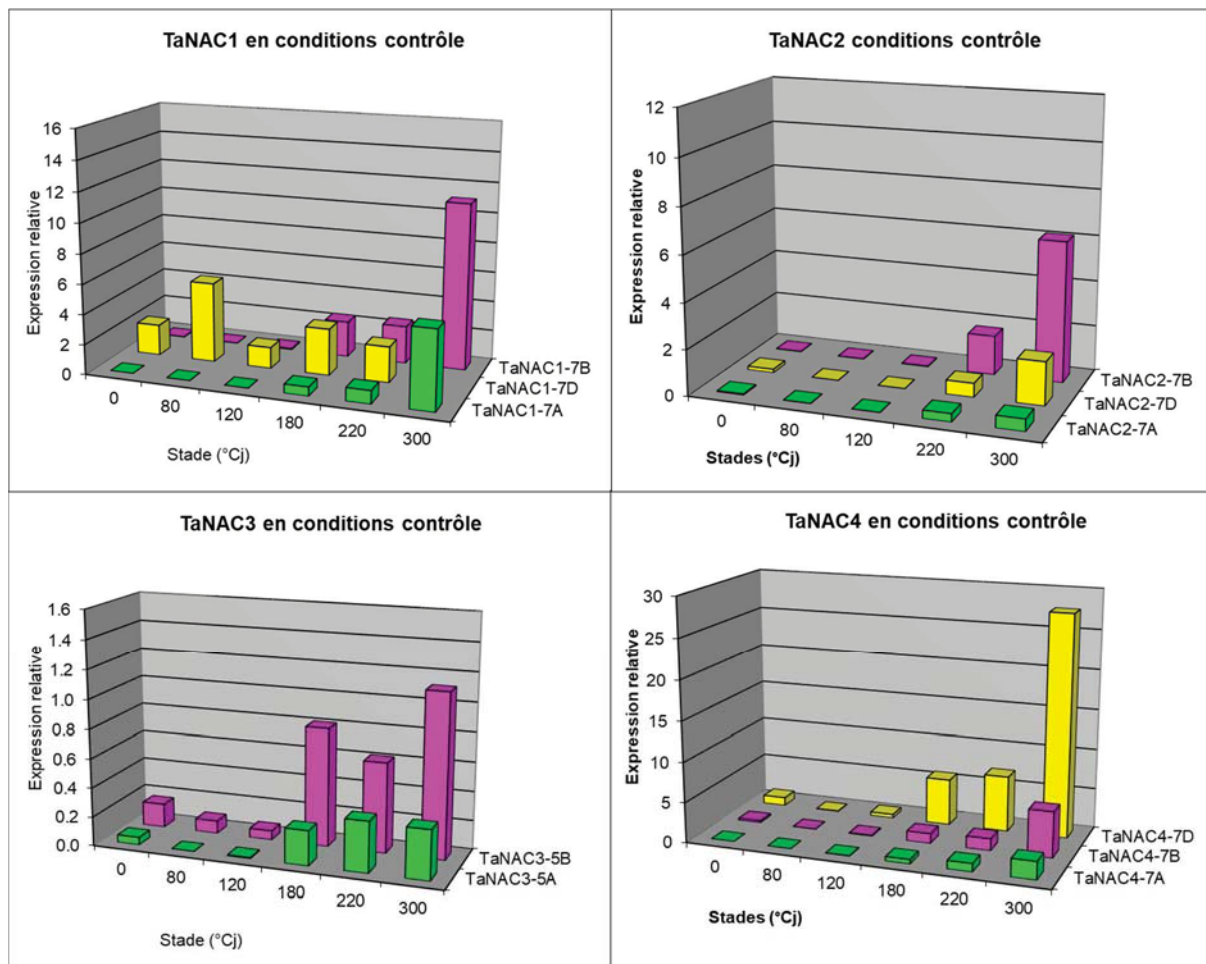


Figure 76 : Profil d'expression des homéologues de TaNAC1, TaNAC2, TaNAC3 et TaNAC4 au cours du développement du grain en condition contrôle (19°C). L'homéologue de TaNAC3-5D n'a pas pu être identifié. Pour TaNAC1, TaNAC2 et TaNAC3, c'est la copie B qui est la plus exprimée, tandis que pour TaNAC4, c'est la copie D. La copie A montre toujours un niveau d'expression faible par rapport aux 2 autres. Les valeurs d'expression relatives ont été calculées selon la méthode du $2^{-\Delta\Delta C_t}$, en utilisant le gène Ta2776 (Paolacci et al., 2009) comme gène de référence.

La réponse de tous les homéologues suivis à la température est la plus forte à 120°Cj, mis à part pour TaNAC1-7D (Figure 79). Malgré sa basse expression par rapport aux copies B et D, la copie A de TaNAC1, TaNAC2 et TaNAC4 montre la meilleure réponse à la température, tandis que pour TaNAC2, cela concerne la copie D. De manière cohérente avec le suivi des 3 homéologues confondus, les 3 copies de TaNAC4 ne montrent pas une réponse très forte à la température à 120°Cj par rapport aux autres gènes.

6.3. Réponse à d'autres stress abiotiques

Les données *in silico* disponibles publiquement sur la plateforme Genevestigator ont apporté des éléments quant à la réponse de nos 4 gènes à d'autres types de stress et de conditions culturales (Qin *et al.* 2008a; Wan *et al.* 2008) (Figure 80). Comme le stress thermique modéré, la fertilisation azotée semble induire une augmentation importante de l'expression des 4 gènes. À l'inverse, le choc thermique sur feuille, tout comme la déshydratation ou un stress froid ne semblent pas induire une modification significative de leur expression. Ce choc thermique sur feuille a été répété en laboratoire et présenté en Chapitre II – Partie 2.1.3. Il a montré que les 4 TaNAC présentent un niveau d'expression très bas dans les feuilles, qui est réprimé après 6h de stress pour TaNAC2 et TaNAC3 uniquement (Figure 81).

7. Validation fonctionnelle

Les expérimentations précédentes ont permis d'élaborer des hypothèses quant au rôle éventuel des 4 TaNAC dans le grain, de par leur localisation subcellulaire et leurs profils d'expression au cours du développement du blé tendre et lors de sa réponse à différents types de stress abiotiques. Cependant, et bien qu'ils constituent une bonne indication, ces résultats ne sont que des observations de corrélations sans lien de causalité. Une validation fonctionnelle étant nécessaire pour mettre en évidence un lien de cause à effet entre l'expression des 4 gènes TaNAC et une potentielle modification du phénotype, nous avons donc décidé de générer des plantes transgéniques sur- et sous-exprimant ces gènes.

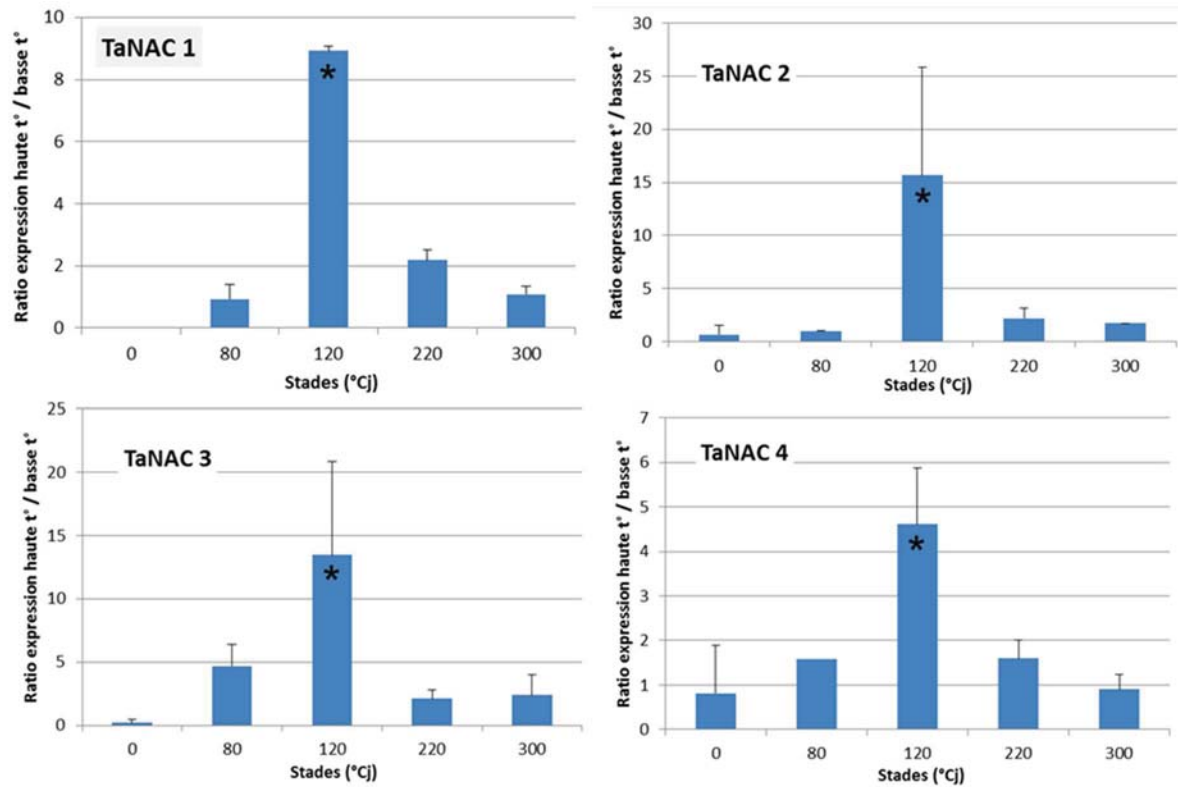


Figure 77 : Ratio d'expression des gènes TaNAC1, TaNAC2, TaNAC3 et TaNAC4 en réponse à une contrainte thermique sur leur expression en condition contrôle au cours de 5 stades du développement. La quantification a été réalisée par PCR quantitative sur les ARN totaux extraits de grains de blé en condition contrôle et fortes températures. L'écart type a été calculé à partir de 2 réplicats biologiques comprenant chacun 2 réplicats techniques. Les valeurs d'expression relatives ont été calculées selon la méthode du $2^{-\Delta\Delta C_t}$, en utilisant le gène Ta2776 (Paolacci et al., 2009) comme gène de référence. L'étoile représente une différence significative d'expression du gène (p -value < 0.05) au stade indiqué (ANOVA suivie d'un test de Scheffé). Le stade clé de la réponse à une contrainte thermique appliqué durant le stade de division cellulaire chez le blé apparaît à 120°Cj (stade de division cellulaire).

7.1. Transfert de l'étude des 4 TaNAC chez le petit épeautre

Les phénomènes de compensation de l'expression entre gènes aux fonctions similaires sont un risque existant lorsque l'on modifie l'expression d'un gène issu d'une famille nombreuse. De plus, le blé est une plante possédant 3 sous-génomes, ce qui contraint à gérer les 3 homéologues de chaque gène afin d'éviter toute compensation entre homéologues (Zhang *et al.* 2017, 2019 ; Rey *et al.* 2018). Par ailleurs, sa culture est assez longue (6 mois) et l'étude de plantes transgéniques en général nécessite de travailler sur un minimum de 3 générations pour éviter d'observer des effets liés au stress de la transformation. À l'échelle d'une thèse, c'est un facteur extrêmement limitant d'un point de vue technique. Pour ces différentes raisons, nous avons donc transféré cette validation fonctionnelle sur une plante du même genre, diploïde (porteuse du génome A) et dont le cycle de développement est beaucoup plus court : *Triticum monococcum*, ou petit épeautre, dont la transformation est maîtrisée au sein de notre unité. En effet, le centre INRA de Crouël avait déjà étudié le taux de réussite de transformation du petit épeautre et sélectionné un génotype montrant de bons résultats en terme de taux d'intégration du transgène et de régénération *in vitro*. Ainsi, en partenariat avec la plateforme Validation Fonctionnelle de l'UMR GDEC, nous avons décidé de générer des plants de petit épeautre transgéniques, sur-exprimant ou sous-exprimant chacun des 4 gènes (cf. matériel et méthode).

7.1.1. Développement du grain de petit épeautre

Afin de pouvoir comparer le développement du grain de petit épeautre avec celui du blé, une cinétique a été produite sur le modèle de celle réalisée par Capron *et al.* (2012) sur le blé tendre (Figure 82). Les 3 grandes phases de développement du grain sont similaires entre le blé tendre et le petit épeautre mais, une fois mature, le grain de petit épeautre présente une forme plus allongée. Le raccourcissement de la durée de culture est bien visible, puisque l'on peut constater qu'à 700°Cj le grain de blé n'est pas mature tandis que celui de petit épeautre a fini la phase de dessiccation à 750°Cj. L'hétérogénéité naturelle présente chez le petit épeautre (dont les lignées sont beaucoup moins conservées que le blé) a été un frein à l'établissement d'une cinétique homogène entre les stades.

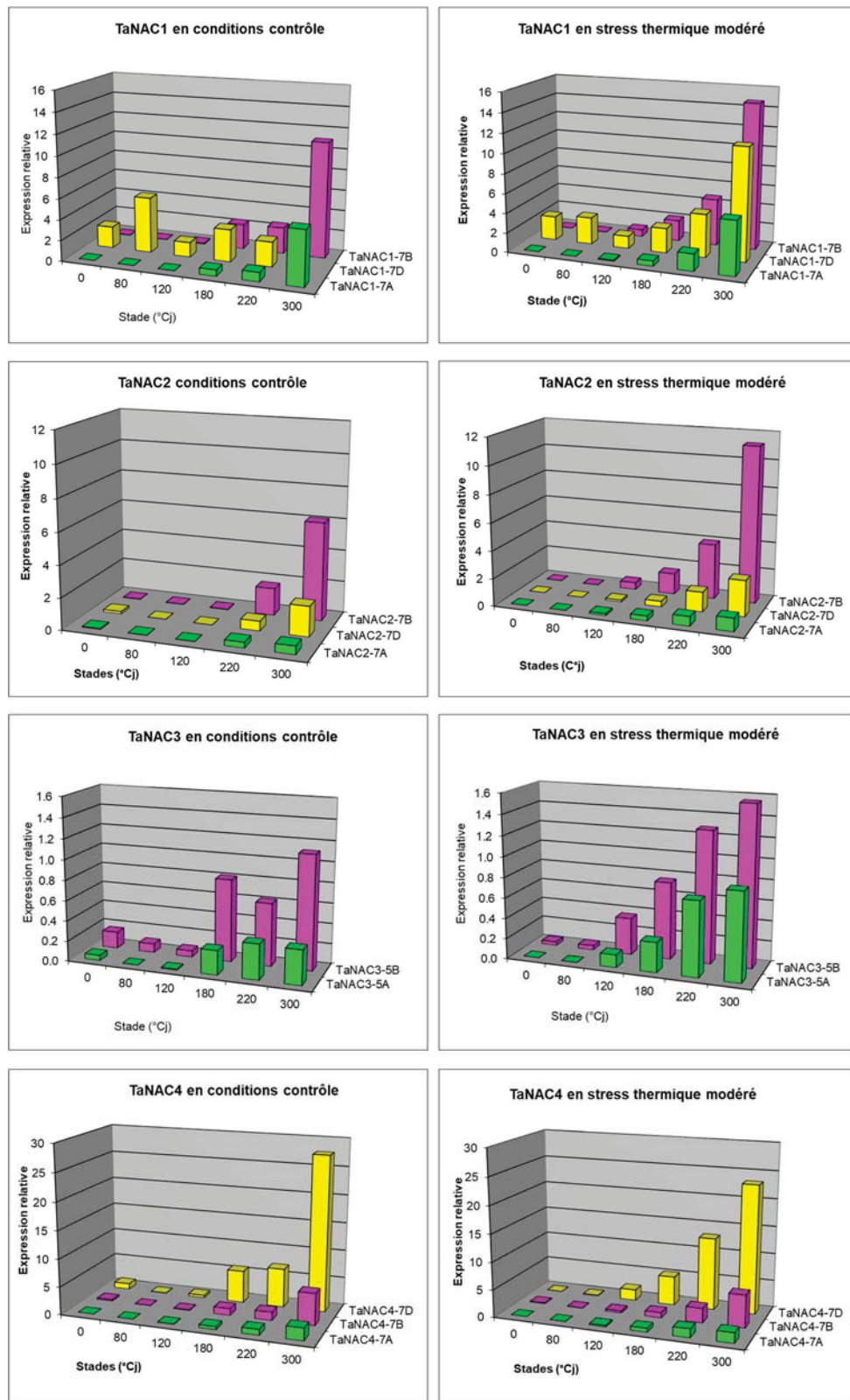


Figure 78 : Expression des 3 homéologues de TaNAC1, TaNAC2, TaNAC3 et TaNAC4 au cours du développement du grain en condition contrôle (19°C, à gauche) et en réponse à un stress thermique modéré (27°C, à droite) appliqué durant 10 jours après l'anthesis. Ces résultats ont été obtenus à partir des échantillons de l'expérimentation BreedWheat.

7.1.2. Séquençage des 4 TmNAC chez le petit épeautre

Les homologues des TaNAC de blé tendre chez le petit épeautre sont notés des TmNAC : des NAC de *Triticum monococcum*. De la même manière que cela a été réalisé pour les 4 TaNAC chez le blé tendre (cf. Chapitre 3 – Partie 2), la séquence des 4 TmNAC a été amplifiée chez le génotype TM35821 de petit épeautre. Les séquences obtenues présentent des tailles similaires à la copie A de leur homologue chez le blé tendre (Tableau 35), sauf TmNAC3 dont la fin de la séquence apparaît tronquée par rapport au TaNAC3-5A. En effet, la base 633 à la base 911, soit 279 pb de la séquence TaNAC3 sont absentes chez TmNAC3. Le pourcentage d'identité entre les séquences de la copie A du blé tendre et leur homologue du petit épeautre est présenté pour les 4 NAC dans le Tableau 35. Les homologues NAC1 et NAC4 sont très proches (98% d'homologie dans les 2 cas), tandis que les homologues NAC2 semblent plus divergents. D'ailleurs, TmNAC2 présente plus d'identité avec TaNAC4 (1019/1050, soit 97%) qu'avec TaNAC2 (687/1048 soit 66%). De plus, TmNAC4 et TmNAC2 présentent 1038/1050 (99%) d'identité entre eux.

7.1.3. Cinétique de développement en réponse à un stress thermique chez le petit épeautre

Afin de valider le transfert des données du blé tendre au petit épeautre, il était important de vérifier que les 4 gènes NAC suivis présentaient un profil d'expression similaire au cours du développement du grain de ces deux espèces, mais aussi lors de leur réponse à la température. Pour cette raison, une culture de petit épeautre a été conduite dans les mêmes conditions que celles appliquées sur le blé tendre lors du projet BreedWheat ; à savoir une condition contrôle de 19°C et l'application d'un stress thermique modéré de 27°C durant 10 jours après l'anthèse.

Le profil d'expression des 4 TmNAC est effectivement similaire à celui observé pour leurs homologues de blé tendre, c'est-à-dire que leur expression est minimale durant les phases précoces de développement du grain, puis elle augmente aux alentours de 120°Cj pour atteindre une phase maximale à 300°Cj. Nous pouvons cependant constater que le point de mesure 700°Cj, absent pour le blé tendre, montre une diminution de l'expression dans les phases tardives de développement du grain de petit épeautre (Figure 83).

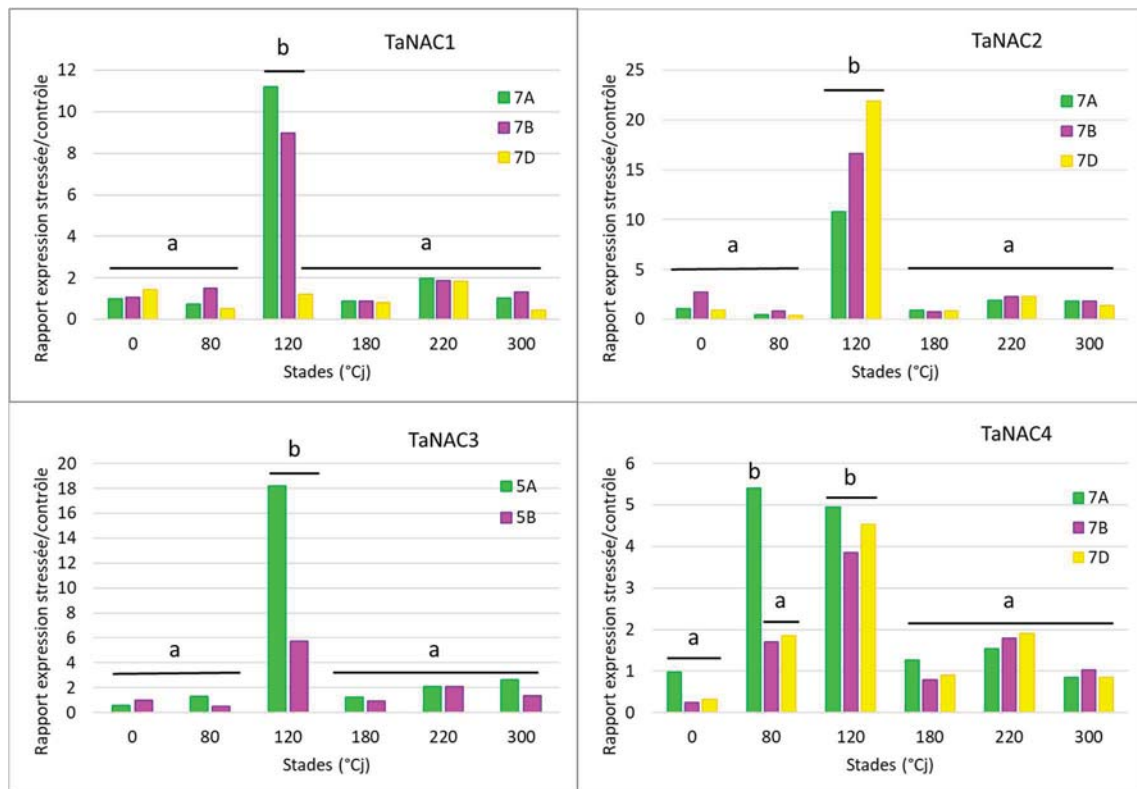


Figure 79 : Ratio de l'expression des 3 homéologues de *TaNAC1*, *TaNAC2*, *TaNAC3* et *TaNAC4* en réponse au stress thermique modéré sur leur expression en condition contrôle. Ces résultats ont été obtenus à partir des échantillons de l'expérimentation BreedWheat. Les tests statistiques ont été réalisés sur les données de chaque homéologue indépendamment.

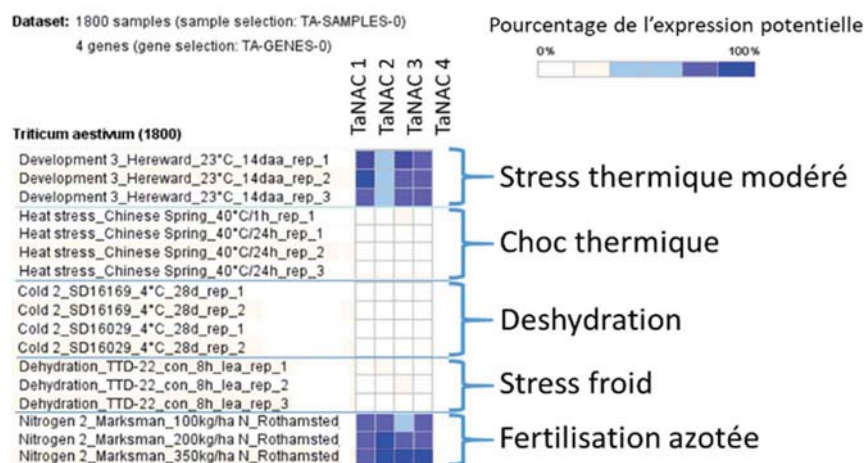


Figure 80 : Expression des 4 gènes *TaTAC1*, *TaTAC2*, *TaTAC3* et *TaTAC4* en réponse à différents stress abiotiques chez le blé en utilisant le site Genevestigator (<https://genevestigator.com/gv/>). Pour un gène et une condition donnés, le niveau d'expression est représenté en % du niveau maximal trouvé parmi l'ensemble des résultats microarray pour ce gène.

Par ailleurs, la plus forte différence entre la condition stressée et la condition contrôle, présentée dans la Figure 84, n'est pas à 120°Cj comme pour le blé tendre mais à 80°Cj. Ce résultat est concordant avec la durée de développement plus courte chez le petit épeautre (environ 4-5 mois) que chez le blé tendre (environ 6 mois). Cependant, à ce stade, les niveaux d'expression sont encore très bas. Notre choix du stade du prélèvement optimal pour la détection d'une modification du niveau d'expression pour les expérimentations à venir s'est donc porté sur le stade 180°Cj. En effet, après le stade 80°Cj, c'est ce stade qui présente le deuxième écart le plus important en réponse à la température et avec des niveaux d'expression plus élevés permettant une détection optimale d'une éventuelle variation.

Pour finir, nous pouvons constater que, comme pour le blé, TmNAC3 est très faiblement exprimé par rapport aux 3 autres gènes ; en revanche, contrairement au blé, il répond faiblement au stress thermique. TmNAC2, quant à lui, est fortement exprimé comme chez le blé mais sa réponse à la température est moins importante.

7.2. Validation fonctionnelle

Des plantes transgéniques ont été générées par bombardement d'embryons immatures de petit épeautre selon la technique de biolistique. Plusieurs lignées de surexpression ont donc pu être générées pour chacun des 4 TmNAC étudiés ; et, en parallèle, des lignées sous-exprimant un seul, plusieurs ou la totalité des TmNAC étudiés ont aussi été générées. Lors du dessin de la cassette de transformation, nous avons identifié, à l'aide des données bio-informatiques disponibles à ce moment, un gène exprimé dans le grain de blé tendre et dont le profil d'expression était identique aux 4 gènes étudiés. Craignant que l'expression de l'homologue de ce gène chez le petit épeautre ne compense la diminution de l'expression des 4 TmNAC sous-exprimés et n'en masque l'effet sur le phénotype ; nous avons choisi d'ajouter ce gène au set de gènes sous-exprimés. Pour la suite de ce chapitre, ce gène sera nommé NAC5. Nous avons essayé d'amplifier le gène **TmNAC5** chez le petit épeautre sans y parvenir, la séquence RNAi utilisée pour le cibler a donc été tirée de la séquence du gène **TaNAC5** de blé tendre. Ses identifiants chez le blé sont TraesCS7A01G569300.1 et TraesCS7D01G543400.1, l'homologue 7B n'a pas pu être identifié. Ces séquences sont très similaires à TaNAC1 (82% d'identité en moyenne).

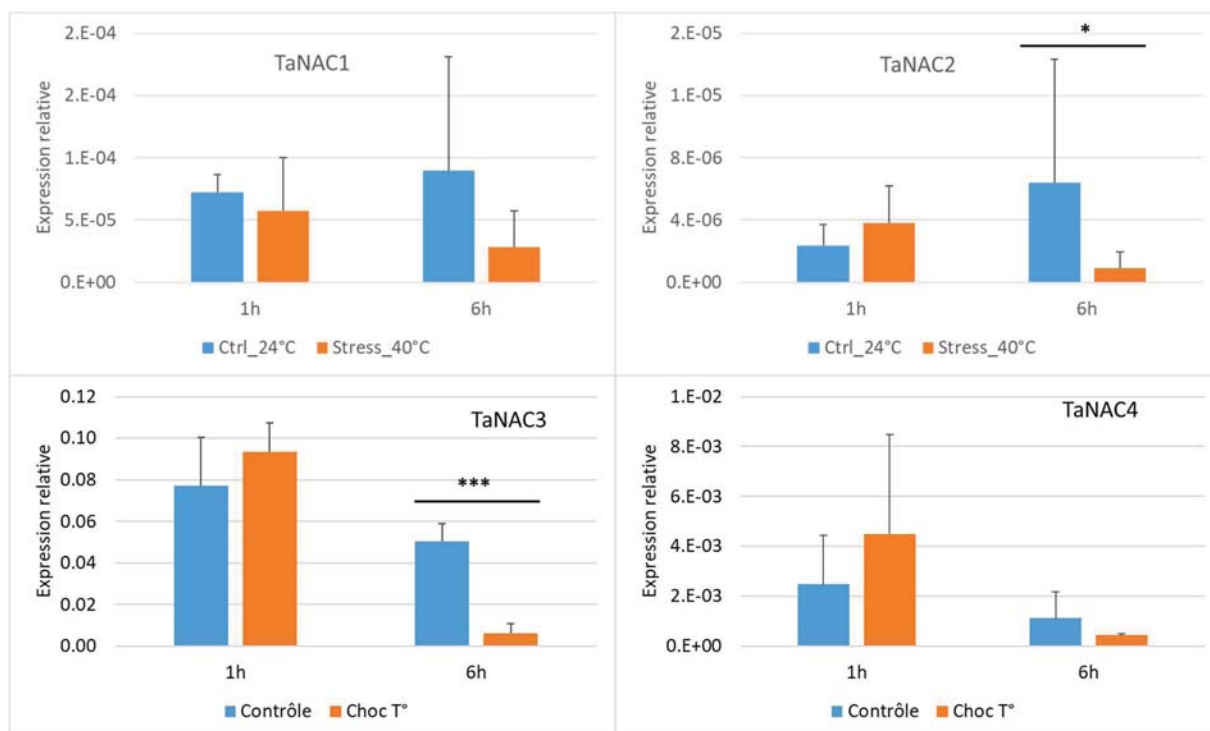


Figure 81 : Expression de TaNAC1, TaNAC2, TaNAC3 et TaNAC4 dans les plantules de blé tendre en condition contrôle et en réponse à l'application d'un choc thermique durant 1h ou 6h. Ces résultats ont été obtenus par qRT-PCR sur un pool de feuilles. L'écart type a été calculé à partir de 2 réplicats biologiques comprenant chacun 2 réplicats techniques. Les valeurs d'expression relatives ont été calculées selon la méthode du $2^{-\Delta\Delta C_t}$, en utilisant le gène Ta2776 (Paolacci et al., 2009) comme gène de référence. * : p-value inférieure à 0,05 et *** : p-value inférieure à 0.001 selon le test T de Student.

Bien que des transformants aient été obtenus pour l'ensemble des 4 TaNAC, nous avons choisi de mener à terme l'étude de 3 constructions uniquement :

- la surexpression de TmNAC2,
- la surexpression de TmNAC4,
- la sous-expression de TmNAC1-TmNAC2-TmNAC3-TmNAC4-TaNAC5 en même temps.

Pour les 2 constructions de surexpression, 4 lignées transgéniques sélectionnées ont été autofécondées jusqu'à la deuxième génération (T2). Cette génération a subi un stress thermique durant la phase précoce de développement des grains T3, qui ont été génotypés et phénotypés. Les lignées de sous-expression ont, quant à elles, été étudiées uniquement en condition contrôle du fait d'un nombre réduit de plantes.

7.2.1. Intégration de la cassette et suivi de l'expression des 5 gènes NAC

Pour le génotypage, le niveau d'expression des gènes TmNAC1, TmNAC2, TmNAC3, TmNAC4 et TaNAC5 a été suivi à 180°Cj et 400°Cj après anthèse, en condition contrôle et en réponse au stress thermique. Statistiquement, aucune différence d'expression significative n'a été observée entre les plantes transformées avec la cassette de surexpression TmNAC2 ou TmNAC4 et leurs témoins Null-Ségrégant respectifs, quels que soient le stade ou la condition thermique étudiés (Figure 85, Figure 86). Les résultats sont identiques pour les plantes RNAi, aucune ne sous-exprime significativement l'un des 5 gènes suivis par rapport au Null-Ségrégant (Figure 87).

7.2.2. Phénotypage

Un phénotype anormal a pu être observé sur les plantes Null-Ségrégant comme sur les lignées transformées des générations T0 et T1 (Figure 88). Nous avons notamment observé des épis malformés, un amincissement de la tige qui compliquait l'épiaison et n'était pas assez rigide pour maintenir l'épi à la verticale. Une autre modification morphologique évidente était la présence d'une languette à la base de l'épi, absente des plantes sauvages. Ces malformations sont probablement des stigmates du stress de la transformation et de la régénération *in vitro*

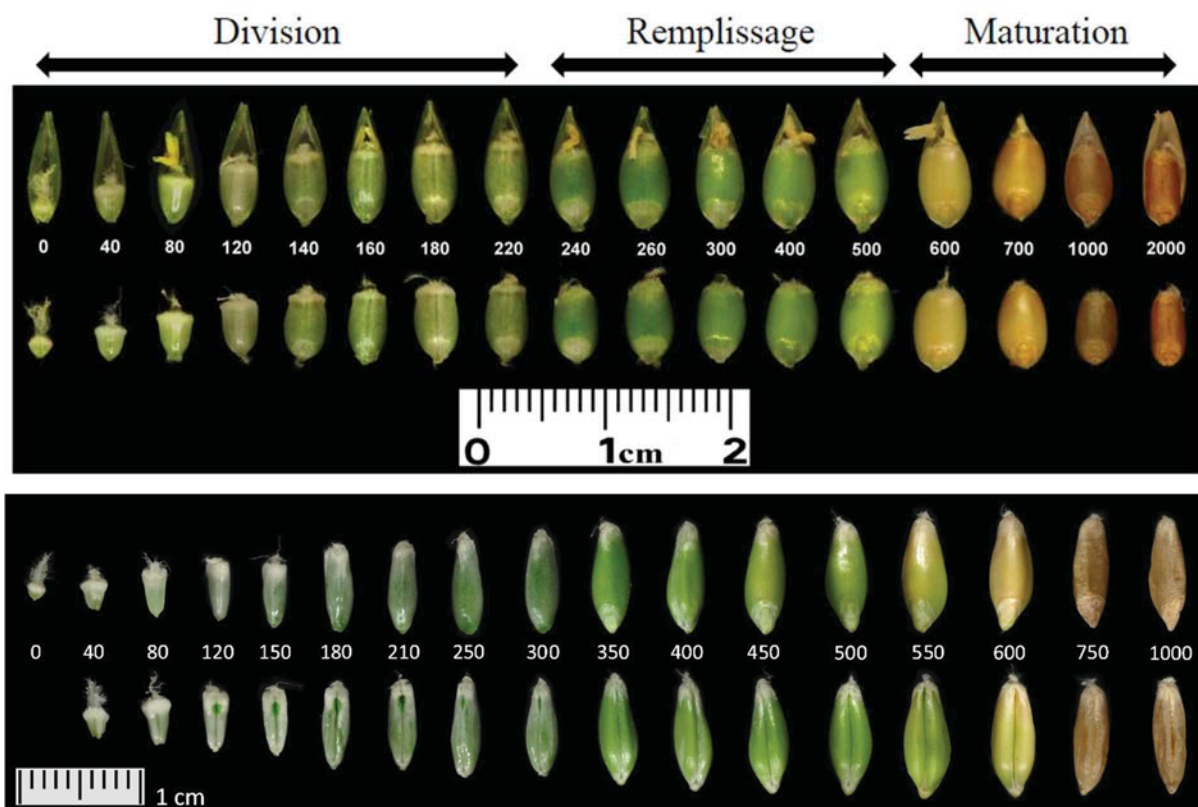


Figure 82 : Cinétique du développement du grain de petit épeautre (en bas), en comparaison avec la cinétique du grain de blé tendre réalisée par Capron et al. (2012) (en haut). Les stades de développement sont en °Cj après anthèse. Le grain de blé tendre a été photographié en vue dorsale, avec et sans enveloppe (ligne supérieure et inférieure respectivement). Les grains de petit épeautre sont sans enveloppe, en prise de vue dorsale (ligne supérieure) et ventrale (sillon visible, ligne inférieure).

Tableau 35 : Tableau récapitulatif de la taille des séquences NAC amplifiées (séquencées) chez le génotype TM35821 de petit épeautre et la copie A du génotype SeriXBabax 49 de blé tendre. La dernière colonne présente le pourcentage d'identité entre ces séquences homologues.

	Blé tendre	Petit épeautre	Identité
NAC1	1107 pb	1107 pb	1082/1107 (98%)
NAC2	879 pb	1047 pb	687/1048 (66%)
NAC3	939 pb	708 pb	675/939 (72%)
NAC4	1050 pb	1050 pb	1029/1050 (98%)

subies par les embryons puis les plantules T0. Seule la multiplication des générations peut permettre d'effacer les effets de ce stress. De plus, le pourcentage de germination de chaque génération a été impacté par la transformation puis la régénération *in vitro*. Le Tableau 36 présente l'impact de la transformation sur le taux de germination de la génération T2. Nous pouvons constater une très forte variabilité de germination entre lignées transformées et, même si les lignées Null-Ségrégant présentent un taux de germination relativement constant (de 34 à 38 %), celui-ci est largement plus faible que les plantes sauvages (73 %).

La mesure d'un large panel de caractères phénotypiques a été réalisée sur les plantes de la génération T2. En fonction de la disponibilité en matériel végétal pour chaque lignée de chaque construction testée, les mesures de phénotypage ont été effectuées sur un nombre variable et parfois restreint de plantes. Des effets phénotypiques ont été observés, que nous pouvons imputer à la variabilité importante du petit épeautre dont les variétés sont moins fixées que le blé tendre. En effet, aucun d'entre eux n'a pu être relié à une variation de l'expression des gènes d'intérêt puisque aucune variation n'a pu être observée au niveau du génotypage.

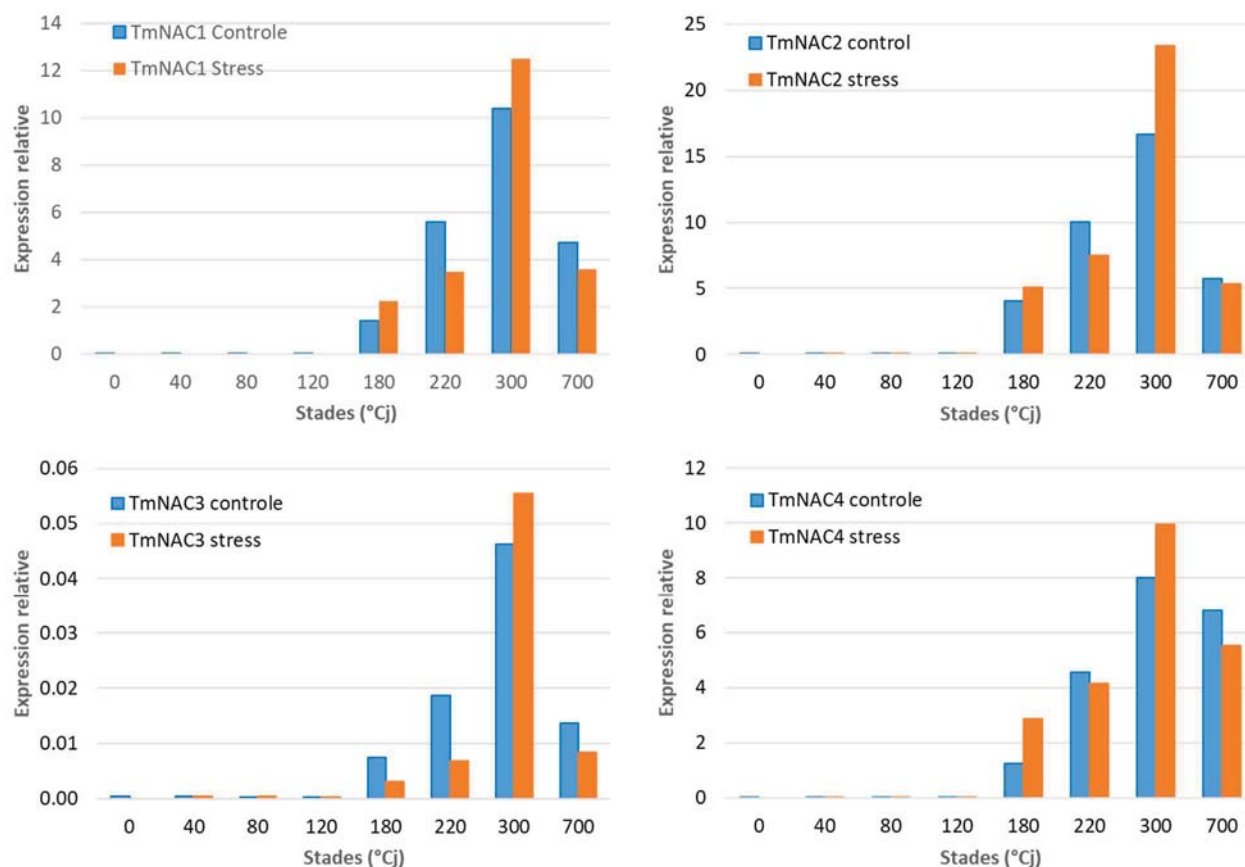


Figure 83 : Profil d'expression de TmNAC1, TmNAC2, TmNAC3 et TmNAC4 en condition contrôle (bleu) et en réponse à un stress thermique modéré appliqué durant 10 jours après l'anthèse (orange).

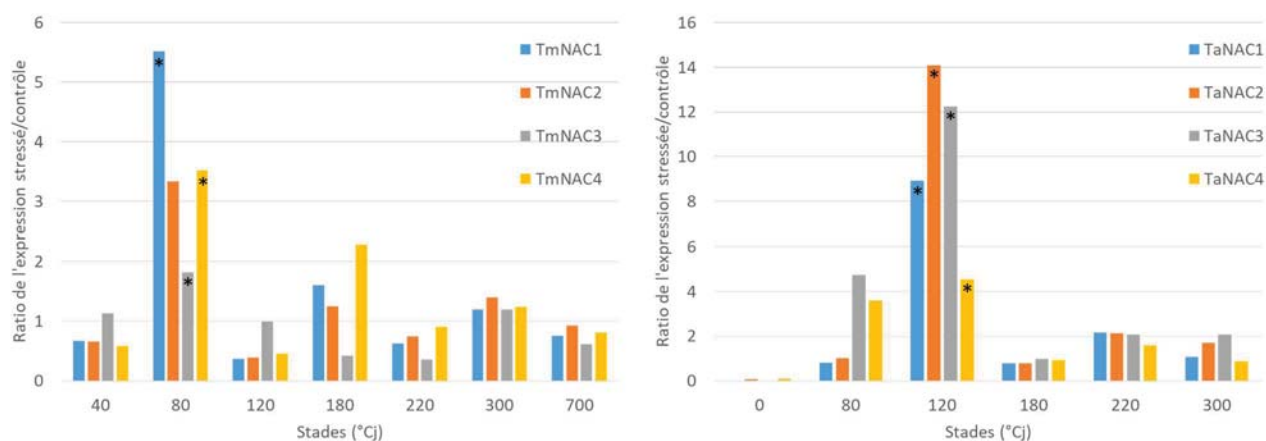


Figure 84 : Ratio de l'expression en réponse à une contrainte thermique sur l'expression en condition contrôle pour NAC1, NAC2, NAC3 et NAC4 chez le petit épeautre (en haut) et la copie A du blé tendre (en bas). Le stade avec une étoile présente une différence significative avec au moins 4 des 5 autres stades testés.

DISCUSSION CHAPITRE III

Avant-propos :

Dans le cadre du projet BreedWheat dont les résultats ont été publiés par Girousse *et al.* (2018), des analyses transcriptomiques chez le blé tendre ont permis de détecter des gènes exprimés dans le grain et dont l'expression est modifiée par l'application d'un stress thermique modéré. Quatre de ces gènes codent pour des facteurs de transcription de la famille NAC. Comme cela a été exposé dans les premières parties de cette thèse, la famille NAC est particulièrement étendue, tant en terme de structure génique que d'expression puisque ces gènes interviennent dans le développement et dans la réponse des plantes à divers stress. Ces facteurs de transcription sont particulièrement intéressants car, en tant que régulateurs de l'expression de gènes de réponse, ils appartiennent à une signalisation cellulaire globale médiatrice de divers messages de stress et de développement. Les particularités d'expression des 4 gènes NAC étudiés dans ce chapitre semblent également montrer qu'ils pourraient être spécialisés aux fonctions physiologiques du grain et particulièrement en réponse à un stress thermique modéré.

Dans cette optique et afin de valider leur rôle dans l'établissement des paramètres physiologiques liés au développement du grain, nous avons réalisé une caractérisation spécifique de ces gènes aux niveaux structural et moléculaire chez le blé tendre (*T. aestivum*, hexaploïde) ; puis nous avons mis en place une expérimentation de validation fonctionnelle chez une espèce proche du blé et au cycle de développement raccourci : le petit épeautre (*T. monococcum*, diploïde).

8. Un clade de l'arbre phylogénétique porteur de gènes aux caractéristiques proches

Au sein de l'arbre phylogénétique, nous avons concentré notre attention sur un clade en particulier, constitué de 20 gènes appartenant à la sous-famille NAC-d qui contient 79 membres au total. Selon Shen *et al.* (2009), les gènes NAC qui interviennent dans l'initiation et la différenciation florale semblent groupés dans la sous-famille NAC-d. Des études portant sur différents membres de cette sous-famille chez plusieurs espèces ont montré qu'elle incluait aussi des gènes jouant un rôle dans les processus de sénescence au cours du développement et en réponse à des stress (Souer *et al.* 1996 ; Xie *et al.* 2000 ; Kim *et al.* 2009 ; Zheng *et al.* 2009

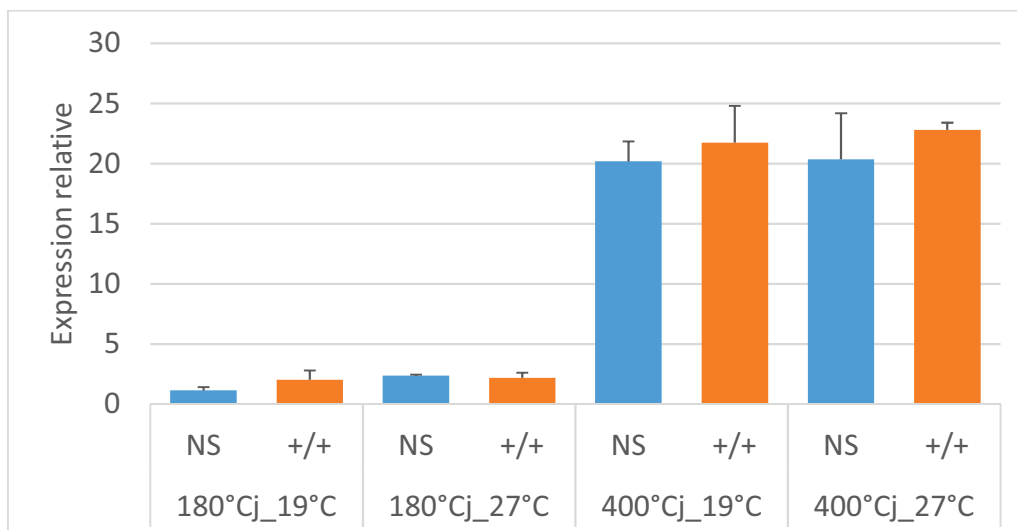


Figure 85 : Expression de TmNAC2 mesurée dans les grains de petit épeautre aux stades 180°Cj et 400°Cj, dans les lignées de sur-expression de TmNAC2 (+/+) et Null-Ségrégant (NS). Les mesures ont été réalisées en conditions contrôle (19°C) et en réponse à un stress thermique modéré (27°C) appliqué durant 10 jours après anthèse. Aucune différence significative au seuil 0,05 n'a pu être observée entre les lignées +/+ et les NS selon le test de Kruskal Wallis.

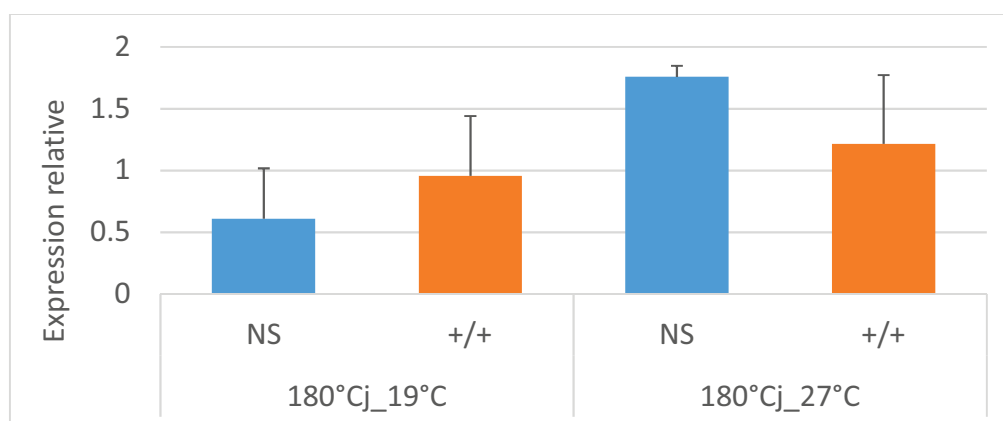


Figure 86 : Expression de TmNAC4 mesurée dans les grains de petit épeautre au stade 180°Cj, dans les lignées de sur-expression de TmNAC4 (+/+) et les lignées Null-Ségrégant (NS). Les mesures ont été réalisées en conditions contrôle (19°C) et en réponse à un stress thermique modéré (27°C) appliqué durant 10 jours après anthèse. Aucune différence significative au seuil 0,05 n'a pu être observée entre les lignées +/+ et les NS selon le test de Kruskal Wallis.

; Balazadeh *et al.* 2010, 2011 ; Oda-Yamamizo *et al.* 2016). L'analyse de la structure de la sous-famille NAC-d chez les monocotylédones par Murozuka *et al.* (2018) a montré que les espèces de *Triticeae* ont près de 2 fois plus de gènes NAC dans ce sous-groupe que les autres espèces monocotylédones. Les auteurs suggèrent l'existence d'une expansion des facteurs de transcription NAC dans ce sous-groupe spécifiquement au cours de son évolution.

Tous les gènes de ce clade possèdent 2 exons et 1 intron dont la taille est relativement conservée. Cet intron mesure entre 107 et 131 pb pour 15 séquences du clade et de 563 à 567 pb pour les 5 séquences restantes. La conservation du découpage exon-intron entre les séquences d'un clade a déjà été observée dans plusieurs études. Lors du recensement de la famille NAC chez *Brachypodium* (Zhu *et al.* 2015a), le maïs (Peng *et al.* 2015) ou encore le concombre (Liu *et al.* 2018), les auteurs ont constaté que les gènes NAC regroupés dans le même clade phylogénétique présentaient des structures exon – intron similaires, notamment en ce qui concerne le nombre d'introns par séquence. Si, comme l'avance (Murozuka *et al.* 2018), ce clade est issu de phénomènes de duplication, il est logique de retrouver une structure génomique similaire entre les séquences.

Quinze des 19 gènes présentant une expression préférentielle dans le grain sont regroupés dans ce clade. Chez l'orge, Murozuka *et al.* (2018) ont identifié 7 gènes (HORVU4Hr1G089450.1, HORVU7Hr1G039700.1, HORVU7Hr1G122680.1, HORVU3Hr1G014090.1, HORVU7Hr1G031260.1, HORVU3Hr1G014100.1 et HORVU2Hr1G082320.2) regroupés dans un clade et présentant une expression forte et quasi-exclusive dans le grain en développement. Ces auteurs ont proposé d'appeler ces gènes « Grain-NACs ». Six de ces 7 gènes appartiennent à la sous-famille NAC-d et sont des homologues des gènes exprimés préférentiellement dans le grain de blé et portés par notre clade d'intérêt (Tableau 37). La dernière séquence NAC d'orge (HORVU2Hr1G082320.2) appartient à la sous-famille NAC-h et son homologue chez le blé tendre (TraesCS2B01G353900.1 / TraesCS2D01G334300.1 / TraesCS2A01G328200.1) n'a pas été identifié comme exprimé préférentiellement dans le grain. De par leur expression préférentielle dans le grain, nous pouvons avancer que ces NAC jouent un rôle dans la formation, la différenciation et/ou le développement du grain de blé tendre, et potentiellement dans la mise en place des tissus de réserve qui permettent ce développement.

Selon Murozuka *et al.* (2018), l'expression des gènes « Grain-NACs » semble être initiée 7 à 10 jours après anthèse chez l'orge et le blé. Si l'on considère une culture en condition thermique

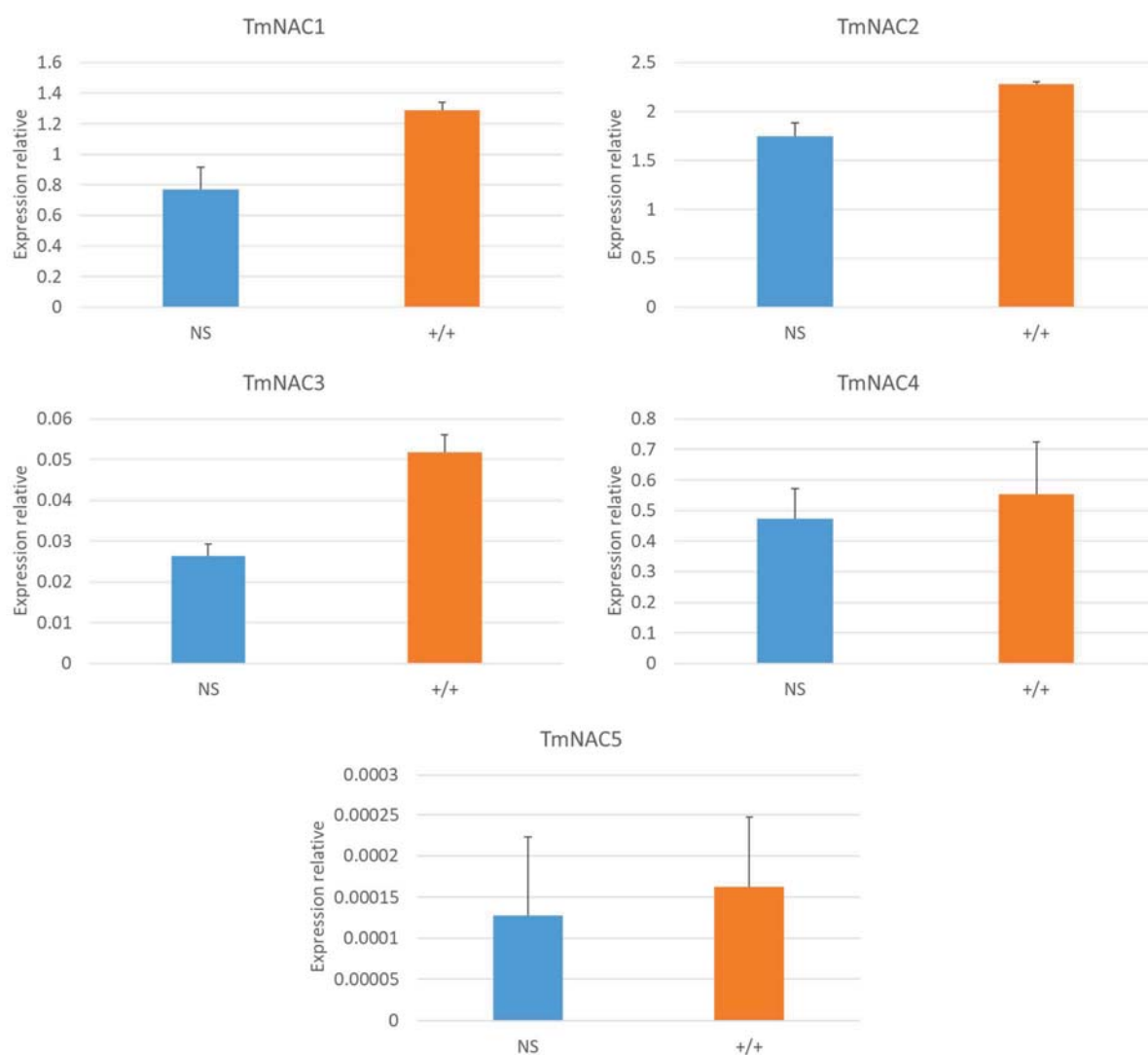


Figure 87 : Expression de TmNAC1, TmNAC2, TmNAC3, TmNAC4 et TaNAC5 mesurée dans les grains de petit épeautre au stade 180°Cj, dans les lignées de sous-expression TmNAC1-TmNAC2-TmNAC3-TmNAC4-TaNAC5 (+/+) et les lignées Null-Ségréant (NS). Les mesures ont été réalisées en conditions contrôle (19°C) uniquement. Aucune différence significative au seuil 0,05 n'a pu être observée entre les lignées +/+ et les NS selon le test de Kruskal Wallis.

non-stressante, soit à 19°C, cela correspond à 130-190°Cj après anthèse. Ce résultat est donc concordant avec les observations réalisées sur les gènes TaNAC1, TaNAC2, TaNAC3 et TaNAC4 qui montrent une augmentation visible de leur expression à partir de 120°Cj-220°Cj après anthèse. Chez le petit épeautre, leur expression observe un pic à 300°Cj, puis une diminution à 700°Cj. Murozuka *et al.* (2018) ont estimé que l'expression maximale des gènes « Grain-NACs » chez l'orge et le blé tendre a lieu entre 15 et 25 jours après anthèse, ce qui correspond à 285-475°Cj pour une culture à 19°C. Ceci nous permet donc de supposer que le niveau d'accumulation des transcrits observé à 300°Cj pour nos 4 gènes NAC serait approximativement le maximal observé dans les cellules du grain. Il serait intéressant de regarder quel profil d'expression est observé entre les stades 300°Cj et la maturité chez le blé et le petit épeautre.

Parmi les 8 groupes d'homéologie présents dans le clade porteur de gènes exprimés préférentiellement dans le grain, 6 groupes présentent une variation allélique. Même si ces SNP n'ont pas pu tous être reliés à une modification phénotypique, il pourrait être intéressant de mener une caractérisation plus large de ces gènes intervenant dans le développement du grain. Il est possible qu'ils interviennent, par exemple, dans sa tolérance à des stress non-étudiés dans cette thèse, notamment aux stress biotiques comme c'est le cas pour d'autres membres de cette famille (Christianson *et al.* 2010 ; Jensen *et al.* 2010 ; Nuruzzaman *et al.* 2013 ; Feng *et al.* 2014).

Au sein de la famille NAC chez le blé, nous avons identifié 18 gènes porteurs, dans leur partie C-terminale, d'un motif spécifique de 70 aa conservés. Parmi eux, 17 gènes appartiennent au clade d'intérêt étudié ici. La dernière séquence porteuse de ce motif (TraesCS7A01G152400) possède 2 exons et un intron comme les autres gènes du clade mais son expression n'est pas grain-préférentielle. Le rôle de ce motif n'est pas connu mais nous pouvons supposer qu'il appartient au domaine de régulation de l'activité transcriptionnelle, peu conservé au sein de la famille et localisé dans la partie C-terminale de la protéine. Ce travail de recensement n'a été réalisé que sur les 488 membres de la famille TaNAC, rechercher l'éventuelle présence de ce motif dans des gènes appartenant à d'autres familles pourrait nous donner des indications sur son rôle.

La présence de ce motif conservé dans la région C-terminale des TaNAC exprimés préférentiellement dans le grain est concordante avec les résultats observés par Murozuka *et al.* (2018), qui ont constaté que les 6 « Grain-NACs » de l'orge possèdent une partie C-terminale

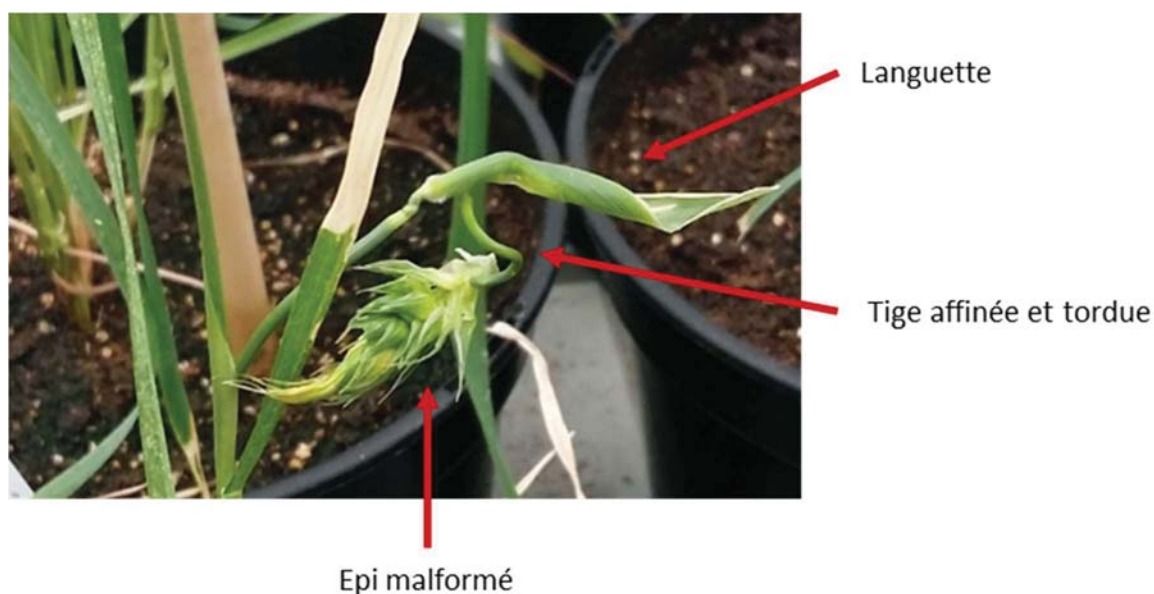


Figure 88 : Photo d'un épi de la génération T1, présentant un phénotype affecté par le stress de la transformation et de la régénération *in vitro*. Ce phénotype était observable sur toutes les lignées des plantes transgéniques comme chez les Null-Ségrégant.

Tableau 36 : Taux de germination observé lors du semis de la génération T2 des 2 lignées de sur- ou sous-expression étudiées ainsi que pour les plantes Null-Ségrégant affiliées. Le nombre de grains germés est précisé entre parenthèses. Le taux de germination de plantes sauvages semées en contrôle est aussi précisé en dernière colonne. Nous pouvons constater une grande disparité dans l'efficacité de germination, surtout dans les lignées de sous-expression.

	Sur-Expression TmNAC2	Sur-Expression TmNAC4	Sous-expression TmNAC 1-2-3-4-5	Sauvages
+/+ Tir 1	76 % (154)	36 % (32)	9 % (4)	73 % (22)
+/+ Tir 2	35 % (44)	37 % (26)	50 % (7)	
Null-Ségrégant	34 % (72)	38 % (36)	38 % (24)	
Multicopie Tir 1	47 % (53)	/	/	

de 175 aa conservée à 51,5-64,5% d'identité. Cette identité tombe à 25% maximum lorsque cette partie C-terminale est comparée avec l'ensemble des autres NAC que Murozuka *et al.* (2018) ont identifié chez l'orge. Ils ont proposé l'hypothèse que les « Grain-NACs » seraient issus d'événements récents de duplication (ayant eu lieu pendant ou peu de temps après la formation des Poacées, famille incluant les *Triticeae*) qui auraient permis l'expansion de ce sous-groupe spécifique et la multiplication des gènes porteurs de ce motif.

9. Les 4 TaNAC sont regroupés dans un clade phylogénétique

Parmi les gènes présents dans le clade étudié ci-avant, nous avons choisi de nous concentrer sur 4 gènes, totalisant 10 homéologues identifiés dans la base de données IWGS : TaNAC1, TaNAC2, TaNAC3 et TaNAC4. Ces gènes sont porteurs d'un motif conservé et, excepté le TaNAC2-7A, tous présentent une expression préférentielle dans le grain. L'expression de TaNAC2-7A n'a pas été reconnue comme préférentielle au grain car il ne possède pas d'identifiant dans la base de données expVIP, mais nous pouvons supposer qu'il se comporte comme ses homéologues B et D.

9.1. Comparaison entre les séquences de l'IWGS et les séquences clonées

Le clonage suivi du séquençage des gènes TaNAC présents dans le génotype SerixBabax 49 (utilisé par Girousse *et al.* (2018)) a révélé des différences de type SNP avec les séquences extraites de la base de données IWGS utilisant le cultivar Chinese Spring. Certains de ces SNP sont concordants avec ceux identifiés lors de notre recherche de variabilité génétique. Par ailleurs, Plessis *et al.* (2013) ont déjà identifié des SNP dans les séquences des copies A et B de TaNAC1 et TaNAC3, dont certains sont aussi communs avec ceux identifiés lors de notre recherche de variabilité génétique. Ceci n'est pas étonnant, dans la mesure où les bases de données de variabilité employées sont en partie identiques.

Par une contrainte de temps, les variations alléliques (exemple : A/C à une position donnée) des SNP identifiés dans les séquences génomiques des TaNAC porteurs de polymorphisme n'ont pas pu être récupérées. Cependant, Plessis *et al.* (2013) ont renseigné les variants alléliques de 5 et 3 SNP des séquences TaNAC1-7A et 7B, et 3 et 1 SNP de TaNAC3-5A et 5B respectivement. Parmi ces SNP, la majorité sont des mutations silencieuses, mais 4 induisent

Tableau 37 : Tableau présentant les relations d'homologie existant entre les 6 gènes « Grain-NACs » identifiés chez l'orge par Murozuka et al. (2018) et les gènes TaNAC appartenant au clade d'intérêt.

Gènes Grains-NAC présents chez l'orge	Séquences TaNAC du clade
	TraesCS7A01G189200
HORVU7Hr1G039700	TraesCS7D01G189900
	TraesCS7B01G094000
HORVU4Hr1G089450	TraesCS7B01G100300 (TaNAC4)
HORVU7Hr1G039700	TraesCS7A01G194700 (TaNAC4)
HORVU4Hr1G089450	TraesCS7D01G196300 (TaNAC4)
HORVU3Hr1G014100	TraesCS3B01G092900
	TraesCS3D01G078400LC
HORVU7Hr1G031260	TraesCS7D01G154200 (TaNAC2)
HORVU7Hr1G031260	TraesCS7A01G152500 (TaNAC2)
HORVU7Hr1G031260	TraesCS7B01G056300 (TaNAC2)
HORVU3Hr1G014090	TraesCS3A01G077900
HORVU3Hr1G014090	TraesCS3D01G078500
HORVU3Hr1G014090	TraesCS3B01G092800
	TraesCS5A01G245900 (TaNAC3)
HORVU7Hr1G122680	TraesCS7A01G569300
HORVU7Hr1G122680	TraesCS7D01G543400
	TraesCS7B01G489500 (TaNAC1)
	TraesCS7A01G569100 (TaNAC1)
	TraesCS7D01G543500 (TaNAC1)

une modification d'acide aminé (Tableau 38). Deux de ces SNP non-silencieux sont localisés dans le domaine NAC des protéines TaNAC3-5A et TaNAC3-5B. Il est donc possible qu'ils puissent altérer la capacité de liaison à l'ADN de ces protéines, puisque cette capacité est conférée par le domaine NAC. Il serait intéressant de réaliser ce travail de collecte de données pour les autres SNP identifiés, afin de savoir si la variation allélique est silencieuse ou si elle induit une modification de la séquence protéique. En effet, un SNP peut induire la génération d'une protéine au repliement anormal ou d'un codon stop prématuré qui peut modifier, voire abolir le fonctionnement de la protéine.

Le gène TaNAC3-5B n'est pas annoté dans la dernière base de données IGWS RefSeq v1.0. Il a pu cependant être séquencé dans le génotype SerixBabax 49, et le résultat obtenu concorde avec la séquence génomique du chromosome 5B correspondant à la région génomique où on s'attend à trouver ce gène. Ainsi, la séquence existe mais n'a pas été identifiée comme un gène par les « pipelines » utilisés par l'IWGS pour l'annotation automatique du génome du blé tendre. Ce gène code pour une protéine de 120 aa, ce qui est une des plus courtes protéines identifiées au sein de la famille TaNAC, la plus courte faisant 91 aa. Son homéologue TaNAC3-5A possède une longueur de 313 aa, et ces 2 gènes présentent des profils d'expression similaires, ce qui suggère qu'il est fonctionnel. Cependant, à notre connaissance, aucun NAC d'une taille similaire n'a été caractérisé fonctionnellement à ce jour dans la littérature. Une recherche du motif identifié dans les gènes exprimés préférentiellement dans le grain par l'outil MAST (<http://meme-suite.org/doc/mast.html>) a montré que ce motif n'était pas présent dans la séquence TaNAC3-5B.

La base de données IWGS RefSeq v1.0 a permis de collecter les séquences génomiques dont, entre autres, celles des 4 TaNAC étudiés. L'intron de TaNAC1 est beaucoup plus grand (566 pb en moyenne entre les homéologues) que les introns de TaNAC2, TaNAC3 et TaNAC4 (110 pb en moyenne). Or, lorsque les introns sont alignés, nous pouvons constater que les ~70 premières bases de l'intron de TaNAC2, TaNAC3 et TaNAC4 présentent la meilleure homologie avec le début de l'intron de TaNAC1 tandis que les ~40 bases restantes s'alignent préférentiellement avec les 40 dernières bases de l'intron de TaNAC1. Nous pouvons donc supposer qu'une mutation de type insertion aurait eu lieu dans la séquence intronique des homéologues de TaNAC1 ou une mutation de type délétion aurait pu advenir dans la séquence des 3 autres TaNAC étudiés.

Tableau 38 : Identification des 4 SNP décrits par Plessis et al. (2013) pour TaNAC1 et TaNAC3 copies A et B, induisant une modification de la séquence protéique. Les SNP induisant une modification silencieuse ne sont pas cités dans ce tableau. Deux SNP ont été localisés dans un domaine NAC.

Gène	Position SNP sur CDS	Allèles	Acide aminé
TaNAC1-7A	564	C/T	R/C
TaNAC1-7B	/	/	/
TaNAC3-5A	293 (domaine NAC)	G/C	W/S
	646	C/T	P/S
TaNAC3-5B	64 (domaine NAC)	A/G	T/A

9.2. Régulation de l'activité des 4 TaNAC étudiés

Les facteurs de transcription jouent un rôle clé dans le déroulement des mécanismes cellulaires. Leur activité d'activation ou de répression de la transcription des gènes cibles est importante et doit donc être finement régulée. Cette régulation peut se faire à plusieurs échelles au sein de la cellule végétale : transcriptionnelle, post-transcriptionnelle et post-traductionnelle. L'accumulation des transcrits des 4 TaNAC étudiés au sein des cellules est modulée au cours du développement du grain, selon des mécanismes encore inconnus.

9.2.1. Régulation transcriptionnelle

Murozuka *et al.* (2018) ont montré la présence d'éléments cis-régulateurs dans les régions promotrices des gènes « Grain-NAC » chez l'orge, homologues des gènes TaNAC exprimés préférentiellement dans le grain et recensés dans cette thèse. Ces cis-éléments sont retrouvés selon un ordre d'apparition conservé. Ils ont été recherchés et identifiés par les mêmes auteurs dans 6 gènes homologues du blé tendre, notamment TaNAC2 et TaNAC4. Ces motifs sont connus pour intervenir dans la régulation de l'expression des protéines de stockage du grain dans l'orge et le blé, pour permettre la liaison de différents facteurs de transcription comme les MYB ou les bZIP ou des régulateurs positifs de la signalisation liée à l'éthylène (Murozuka *et al.* 2018). Le motif permettant la liaison de ces derniers (RYREPEAT/EIN3 motifs) est présent en plusieurs copies chez les promoteurs étudiés. Les auteurs avancent que le nombre de copies dans le promoteur du gène pourrait être responsable du niveau d'expression du gène. Il serait intéressant de regarder si la présence de ce motif dans les régions promotrices des 4 TaNAC du blé tendre présente une corrélation avec les différences d'intensité d'expression observées entre ces gènes.

De plus, connaître les facteurs de transcription intervenant en amont de la voie de régulation des 4 TaNAC étudiés permettrait d'ouvrir la voie sur les mécanismes de régulation dans lesquels ils sont impliqués. Une mise en contact du protéome nucléaire du grain de blé avec les séquences promotrices des TaNAC (par une approche simple-hybride par exemple), et une identification des protéines fixées à ces séquences promotrices serait donc à prévoir pour comprendre la place de ces TaNAC dans les mécanismes moléculaires du grain.

9.2.2. Régulation post-transcriptionnelle

Plusieurs publications ont décrit le rôle du microARN miR164 dans la régulation de l'accumulation des transcrits NAC (Guo *et al.* 2005 ; Kim *et al.* 2009 ; Feng *et al.* 2014 ; Zhuo *et al.* 2018). Cependant, en utilisant l'outil psRNATarget (<http://plantgrn.noble.org/psRNATarget/>), nous n'avons pas pu mettre en évidence de correspondance entre un microARN et les séquences de nos 4 TaNAC chez le blé tendre. Bien que cette approche soit uniquement *in silico*, elle semble indiquer que l'accumulation des transcrits de ces 4 TaNAC n'est pas régulée à ce niveau par un microARN connu.

9.2.3. Régulation post-traductionnelle

Les gènes TaNAC2 et TaNAC3 présentent une localisation nucléaire tandis que TaNAC4 présente une localisation nucléaire et cytoplasmique. Certains NAC possèdent des motifs d'ancrage membranaires et plusieurs publications ont montré que ces formes inactives ancrées à une membrane peuvent être relocalisées dans le noyau sous l'effet d'un stimulus (Kim *et al.* 2006, 2007a; Chi *et al.* 2017). Cependant, aucune hélice transmembranaire susceptible de retenir une forme dormante de la protéine ancrée à une membrane ; ni aucun peptide signal (par exemple, un Nuclear Localization Signal qui pourrait adresser la protéine vers le noyau) n'a pu être identifié pour les 4 TaNAC étudiés. Malgré tout, la localisation nucléaire et cytoplasmique de TaNAC4 amène à proposer l'hypothèse qu'il pourrait présenter un mécanisme de régulation par relocalisation des protéines du cytoplasme au noyau selon un mécanisme non encore élucidé.

Par ailleurs, le motif conservé dans 14 gènes TaNAC qui présentent une expression préférentielle dans le grain (et dont font partie les 4 TaNAC étudiés) est situé dans la partie C-terminale qui est connue pour porter un domaine peu conservé de régulation de l'activité transcriptionnelle. Il pourrait donc intervenir dans l'interaction de la protéine avec la machinerie transcriptionnelle. Afin de valider cette hypothèse, il serait intéressant de tester *in vivo* la capacité d'activation de la transcription d'un gène rapporteur par les 4 gènes TaNAC avec et sans ce motif. Cela pourrait être réalisé en observant, chez la levure, la capacité des protéines TaNAC tronquées en C-terminale à activer l'expression d'un gène rapporteur. Par ailleurs, le criblage d'une banque d'ADNc pourrait être réalisé par la technique du double hybride chez la levure, cela permettrait d'identifier les interacteurs protéiques potentiels de ces gènes TaNAC

et observer une éventuelle altération de ces interactions lorsque le motif est délété. En complément, un éventuel changement de l'adressage des protéines vers le compartiment cellulaire nucléaire pourrait être détecté par fusion des protéines TaNAC tronquées avec une protéine fluorescente, et révélation de leur localisation subcellulaire dans des cellules de tabac transitoirement transformées ; comme nous l'avons fait pour les protéines complètes. De plus, connaître les variations alléliques de chaque SNP identifiés sur les séquences TaNAC pourrait permettre d'identifier les SNP induisant une modification importante de la séquence protéique et donc potentiellement de la conformation et de la fonction de la protéine produite (l'exemple extrême étant la formation d'un codon stop prématuré).

Enfin, la régulation de l'activité d'une protéine produite et fonctionnelle peut se faire selon une troisième voie : la disponibilité de ses interacteurs. En effet, plusieurs protéines NAC ont déjà été identifiées comme agissant en complexe avec d'autres NAC mais aussi d'autres facteurs de transcription en général, pour permettre le recrutement de la machinerie transcriptionnelle et l'expression de gènes cibles (Li *et al.* 2012). Par exemple, les gènes ONAC20 et ONAC26, décrits par Mathew *et al.* (2016), sont les homologues de TaNAC1 et TaNAC3 chez le riz. Ils se complexent en un hétérodimère, qui, lorsqu'il est formé, est relocalisé dans le noyau. ONAC26 interagit aussi avec ONAC23. Certaines protéines NAC peuvent donc former un complexe avec plusieurs interacteurs différents. Nous pouvons donc supposer que les TaNAC homologues pourraient posséder des propriétés de dimérisation similaires.

9.3. Rôle et mécanisme de fonctionnement supposés des 4 TaNAC en condition de température optimale

9.3.1. À l'échelle du grain

L'expression préférentielle de ces 4 TaNAC dans l'organe grain indique probablement une relation étroite entre leur profil d'expression et le développement du grain de blé tendre. Cette expression tissu-spécifique se retrouve pour de nombreux gènes NAC chez plusieurs espèces : Zhu *et al.* (2014b) ont montré une expression préférentielle dans les racines, feuilles et fleurs pour 6 NAC chez la tomate ; tandis que Tang *et al.* (2012) ont identifié 2 TaNAC possédant une expression préférentielle dans la feuille chez le blé tendre. Chez le riz, Fang *et al.* (2008) ont identifié 12 NAC possédant une expression préférentielle à un tissu (cal, racine, étamines, albumen immature), et plus récemment, Mathew *et al.* (2016) ont identifié 9 gènes chez le riz

Tableau 39 : Nombre d'élément RYREPEATBNNAPA identifiés dans les promoteurs (1500 pb) des homéologues des 4 TaNAC étudiés. La recherche a été réalisée à l'aide du prédicteur PLACE (<http://www.dna.affrc.go.jp/htdocs/PLACE/>), puis les résultats ont été filtrés pour ne retenir que le motif RYREPEATBNNAPA.

	Homéologue A	Homéologue B	Homéologue D
TaNAC1	8	10	2
TaNAC2	2	0	0
TaNAC3	2	/	/
TaNAC4	5	4	5

dont l'expression est préférentielle dans le grain, et parmi eux ONAC020, ONAC026 et ONAC023 qui sont 3 gènes qui présentent la plus forte expression. De plus, ONAC020 et ONAC026 semblent être homologues de TaNAC1 et TaNAC3 respectivement.

Le niveau d'expression de chacun des 4 TaNAC dans le grain n'est pas équivalent. Mathew *et al.* (2016) ont avancé l'hypothèse que la présence de multiples copies de l'élément RYREPEATBNNAPA dans le promoteur du gène considéré peut mener à une augmentation de son expression. Cet élément permet la fixation de facteurs tels que FUS3 ou ABI3, qui agissent comme des régulateurs importants de l'expression des protéines de réserve chez les dicotylédones (Kawakatsu et Takaiwa 2010). Une recherche avec l'outil PLACE database (Higo *et al.* 1999) dans les 1500 pb de la séquence génomique en amont des 4 TaNAC étudiés a révélé la présence d'éléments RYREPEATBNNAPA dans leurs promoteurs, sans qu'elle ne puisse être clairement liée au niveau d'expression de chacun des gènes (Tableau 39). Il serait intéressant d'étudier plus finement les cis-éléments présents dans les promoteurs des 4 gènes TaNAC étudiés afin de voir les grandes classes de cis-éléments présents et leur proportion.

Le niveau d'expression de chacun des homéologues des 4 TaNAC dans le grain n'est pas équivalent non plus. Pour le gène TaNAC4, l'homéologue D serait le plus exprimé en condition haute température comme en condition contrôle, tandis que pour les gènes TaNAC1, TaNAC2 et TaNAC3, l'homéologue B est le plus exprimé. Dans une étude de génétique d'association impliquant plusieurs familles de facteurs de transcription, Plessis *et al.* (2013) ont aussi constaté la même dominance d'expression de l'homéologue B de TaNAC1 dans le grain de blé, ce qui est cohérent avec les résultats obtenus par Leach *et al.* (2014), qui ont montré une prédominance de l'expression d'un homéologue pour la majorité des gènes étudiés chez le blé tendre.

En 2013, Plessis *et al.* (2013) avaient déjà identifié un fort déséquilibre de liaison entre TaNAC1-7B (nommé NAC22 dans leur publication) et un marqueur lié à la relation gluténine/teneur en azote totale. Ils avaient aussi constaté une association entre un marqueur en déséquilibre de liaison avec TaNAC3-7B (noté NAC27-B dans leur publication) et le même caractère phénotypique ; ce qui n'a pas été observé ici. Dans des conditions de faible teneur en azote sur le site Le Moulon, NAC22-B était également associé aux résidus de la relation gluténine - azote total et du ratio gliadine-gluténine. Ces 2 gènes ainsi que TaNAC4 (nommé NAC18 dans leur publication) avaient été précédemment identifiés par Vincent *et al.* (2011) comme hautement connectés au réseau de facteurs de transcription qui contrôle l'expression des gènes d'accumulation de protéines de stockage dans le grain. Ils n'ont cependant pas identifié

de marqueur sur le gène TaNAC2, en association avec un caractère phénotypique. Cela est sans doute dû à l'absence de marqueurs connus sur ce gène à cette époque. Il serait intéressant d'étudier la variabilité allélique de ces TaNAC plus en détail, nous pourrions notamment recenser les allèles existants aux SNP identifiés dans cette thèse, mais aussi rechercher d'autres bases de données de génotypage et surtout de phénotypage pour rechercher d'éventuelles associations avec des paramètres phénotypiques plus nombreux et variés.

De même, le niveau d'expression des 4 TaNAC varie au cours du développement du grain. L'expression relative des 4 gènes est faible dans les premières phases de développement et augmente à partir de 120 °Cj. Dans la publication de Girousse *et al.* (2018), un module de gènes nommé Turquoise regroupe 687 gènes présentant un profil similaire en condition contrôle. La plupart de ces gènes sont des gènes dits « de réponse » comme les gènes codant pour des protéines de réserve. De manière générale, Girousse *et al.* (2018) ont montré un enrichissement de ce module en gènes de type « nutrient reservoir activity » et en inhibiteurs de l'activité endopeptidase, mais on y retrouve aussi des intégrateurs de signaux tels que des facteurs de transcription et parmi eux ces 4 TaNAC. En conclusion, les résultats d'analyse de l'expression de ces gènes et des études d'association convergent pour indiquer que ces 4 facteurs de transcription joueraient un rôle dans le développement du grain de blé, *via* le contrôle de l'expression de gènes codant des protéines de réserve.

De par leur localisation nucléaire et leur fonction de facteurs de transcription, nous pouvons supposer qu'ils se fixent sur les promoteurs des gènes cibles afin de réguler leur expression. Cette régulation a lieu probablement en complexation avec d'autres facteurs de transcription. En effet, chez le riz, Mathew *et al.* (2016) ont montré que ONAC20 et ONAC26 sont les homologues de TaNAC1 et TaNAC3 respectivement. ONAC20 est localisé dans le réticulum endoplasmique tandis que ONAC26 est une protéine nucléaire. Ces 2 protéines présentent une faible capacité d'activation et une forte capacité de répression de la transcription, qui est abolie lorsque le gène présente une délétion dans la région C-terminale. ONAC26 interagit avec ONAC20 pour former un hétérodimère qui est relocalisé dans le noyau. Il peut aussi former un hétérodimère avec ONAC23. Ces 3 NAC chez le riz montrent un profil d'expression similaire aux 4 TaNAC de blé étudiés dans cette thèse : une expression basale dans les phases précoces du développement du grain, qui augmente par la suite. Cependant, les auteurs ont constaté que le niveau d'expression des 3 gènes en condition contrôle est génotype-dépendant. Par ailleurs, des SNP identifiés dans la séquence promotrice de ces 3 NAC de riz ont montré une association avec la taille et le poids du grain. Nos 4 TaNAC montrent, comme ces NAC chez le riz, un

profil de réponse identique au cours du développement du grain, ainsi que face à plusieurs stress abiotiques. Ces résultats suggèrent qu'ils seraient donc exprimés au même moment. De plus, de par leur localisation nucléaire, leur interaction serait donc possible comme cela a été observé pour ONAC20 et ONAC26 (Mathew *et al.* 2016). Cette interaction potentielle entre les 4 protéines TaNAC pourrait être visualisée *in planta* par la technique du Bimolecular Fluorescence Complementation.

Il aurait été intéressant de constater une éventuelle compensation de l'expression entre les gènes TaNAC lors du suivi de leur expression dans les plantes transgéniques. Ceci n'a pas pu être réalisé, dû au fait que les plantes transgéniques ne montrent finalement pas de modification de l'expression des 4 TaNAC.

9.3.2. À l'échelle tissulaire

L'analyse des données disponibles sur la plateforme Genevestigator a montré une augmentation de l'expression des 4 TaNAC en réponse à une fertilisation azotée. Ce résultat est particulièrement intéressant car il ouvre des possibilités sur le rôle éventuel de ces gènes dans l'accumulation des réserves protéiques, qui sont liées à la disponibilité en azote au cours de la phase de remplissage du grain (Shewry *et al.* 2000 ; Triboi *et al.* 2000). Ce résultat peut être corrélé avec l'association observée entre des SNP portés par TaNAC1-7B et TaNAC2-7B et l'accumulation de gluténines et de gliadines dans les grains (cf. Chapitre 3 – Partie 3.2), liée à la disponibilité en azote puisque ces associations n'ont été trouvées que dans une condition azotée (high nitrogen ou low nitrogen) et pas dans les deux. De plus, Plessis *et al.* (2013) ont observé que TaNAC1-7B était également associé aux résidus de la relation gluténine-azote total et au rapport gliadine-gluténine dans des conditions de faible teneur en azote. Pour finir, leur expression est induite lors de la transition de la phase de division cellulaire vers la phase de remplissage, nous pouvons supposer qu'ils jouent un rôle dans l'accumulation des réserves dans le grain, et donc influent sur le rendement final obtenu, ce qui pourrait en faire des gènes d'intérêt pour la sélection de nouvelles variétés.

La technique d'hybridation *in situ* (construction de sondes qui se fixent sur l'ARN et permettent de localiser une séquence de nucléotides monobrin (ARN ou ADN) connue sur une coupe histologique de tissu) permettrait de savoir dans quel compartiment histologique du grain de blé les TaNAC sont exprimés plus particulièrement, ce qui pourrait nous fournir des

renseignements quant à leur fonction dans le grain. Selon les données obtenues par les analyses Genevestigator, il semblerait que cette expression soit localisée dans l'albumen mais ces résultats bio-informatiques nécessitent une validation *in vivo*.

9.4. Rôle et mécanisme de fonctionnement supposés des 4 TaNAC en réponse aux stress abiotiques

De nombreuses études ont montré que des plantes de plusieurs espèces sur-exprimant des gènes NAC présentent un phénotype plus tolérant à un large panel de stress abiotiques. Par exemple, chez le blé, des Arabidopsis qui sur-expriment TaNAC2, TaNAC29, ou TaNAC67 montrent une tolérance améliorée à de multiple stress abiotiques (Mao *et al.* 2012, 2014 ; Huang *et al.* 2015). De la même manière, l'expression de TaNAC2a, TaNAC2D, TaNAC4a, TaNAC6, TaNAC7, TaNAC13 et TaNTL5 est induite par un stress hydrique, salin et/ou une température basse (Tang *et al.* 2012 ; Huang et Wang 2016). Récemment, il a été constaté que TaRNAC1 améliorait la tolérance au stress hydrique, entraînant une biomasse, un rendement en grains et une longueur de racines plus élevés (Chen *et al.* 2018). De plus, le blé transgénique sur-exprimant TaNAC47 a montré un phénotype résistant au stress hydrique pouvant être induit en réponse à l'ABA (Zhang *et al.* 2016). Nous avons pour objectif de conduire des expérimentations permettant de valider une corrélation entre la modification du niveau d'expression des gènes TaNAC et l'altération de la tolérance des plants au stress thermique. Cette expérimentation n'a pas pu fournir de résultats concluants. Nous avons cependant apporté des connaissances concernant l'expression de ces gènes dans des plants de blé tendre non transformé en réponse à plusieurs types de stress abiotiques.

9.4.1. Lors d'une élévation de température

Les résultats obtenus par RT-qPCR montrent que l'augmentation de l'expression des 4 TaNAC étudiés a eu lieu à un stade plus précoce en réponse à l'application d'une contrainte thermique (27°C) par rapport au contrôle (19°C). En effet, en condition de stress thermique modéré, l'augmentation significative de l'expression des TaNAC commence à 120°Cj tandis qu'elle n'apparaît qu'au stade 220°Cj en condition contrôle. Ces résultats sont cohérents avec ceux observés par Wan *et al.* (2008), qui ont étudié l'impact d'un stress thermique chaud modéré

Profil moyen d'expression des 500 gènes

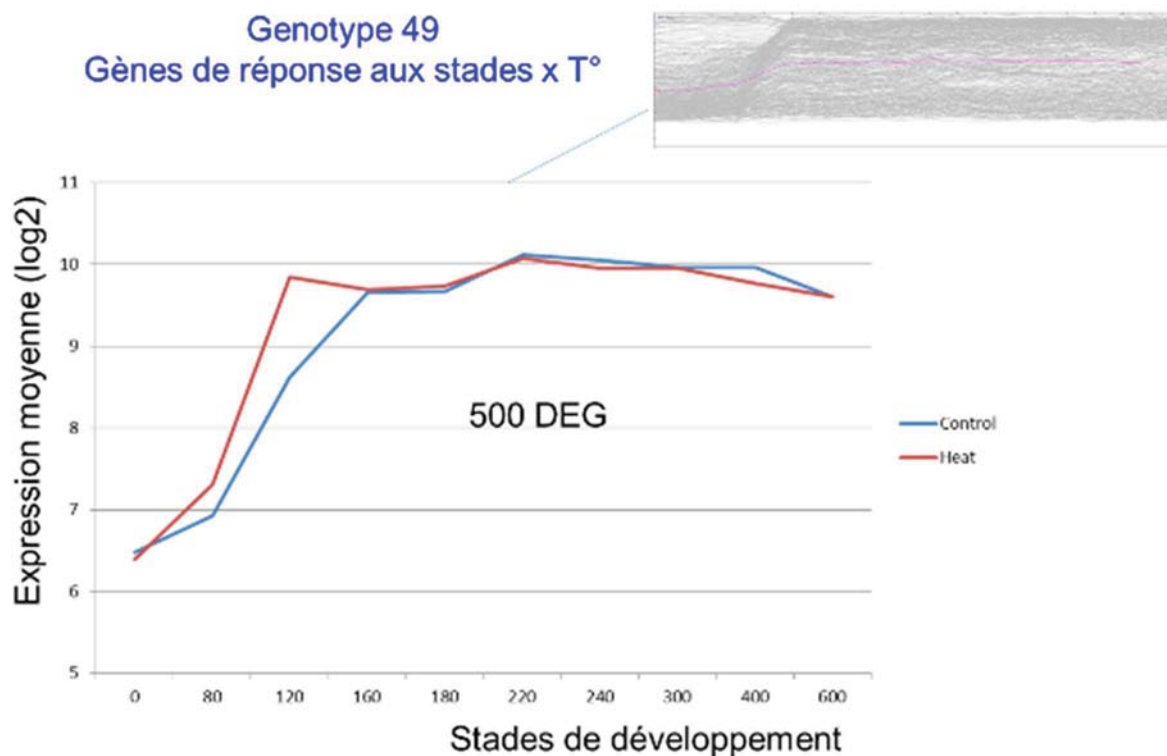


Figure 89 : Profil d'expression moyen de 500 gènes au cours du développement du grain. Ces 500 gènes ont été identifiés par Girousse et al. (2018) comme différentiellement exprimés au cours du développement du grain et en réponse à un stress thermique modéré mais dans des températures physiologiques ; avec un stade de réponse maximal à la température à 120°Cj. Cette réponse au développement et au stress thermique est similaire pour tous les gènes observés ; dont le cumul des profils est visible grisé, en haut à droite.

continu (19°C en condition contrôle / 24°C en condition stressée) sur l'expression de plusieurs facteurs de transcription du grain de blé en développement. Cette étude a montré une cinétique d'expression similaire à la nôtre pour TaNAC1 (noté Ta.25258 dans la publication). Ces résultats sont aussi cohérents avec ceux observés par Girousse *et al.* (2018) lors de l'analyse par puce microarray (Figure 89). Le stade à 120°Cj qui se situe dans la phase de divisions cellulaires serait donc un stade clé du développement dans la réponse à des températures élevées mais néanmoins physiologiques. Chez le rosier, le gène RhNAC100, qui présente de fortes homologies avec ANAC100 chez *Arabidopsis* et avec TaNAC2 chez le blé, a été identifié comme régulateur négatif de l'expansion cellulaire, dont la surexpression engendre une diminution de la taille des pétales de rose (Pei *et al.* 2013). Dans le cadre du projet BreedWheat, les données éco-physiologiques obtenues ont montré une réduction de 19,6% du nombre de cellules dans l'albumen et une diminution de 17,5% de la masse sèche des grains de blé en fin de développement en réponse à un stress thermique (Girousse *et al.* 2018). Ces résultats suggèrent un lien entre l'accélération du cycle qui conduit au raccourcissement des stades de division cellulaire et d'accumulation de la matière sèche ; et l'augmentation de l'expression de TaNAC au cours de la phase de divisions cellulaires en réponse à l'application d'un stress thermique. Cette hypothèse, qui aurait dû être vérifiée par l'analyse fonctionnelle sur plantes transgéniques, n'est pour l'instant pas validée.

Il est intéressant de noter que lors de l'analyse méta-génomique à l'aide du logiciel Genevestigator, l'expression relative des TaNAC étudiés ne semble pas affectée en condition de choc thermique (40°C) (Qin *et al.* 2008a), ni par d'autres stress abiotiques tels que la déshydratation (Ergen *et al.* 2009) ou encore un stress froid (Sutton *et al.* 2009). Cependant, nous avons observé une modification de leur expression (relativement faible par rapport à d'autres gènes mais néanmoins significative) en réponse à la sécheresse, et ce de manière génotype-dépendante. Ce seraient donc des protéines qui interviendraient dans l'acclimatation, l'adaptation de la plante à une température élevée mais physiologique et non pas des protéines exprimées en cas de stress thermique intense de courte durée mettant en place des processus de survie (comme l'expression de protéines chaperonnes).

En revanche, lors de la même analyse méta-génomique, il apparaît que les 4 TaNAC sont fortement exprimés en réponse à une modification du statut azoté du sol : leur expression relative est augmentée après un apport en azote (que ce soit 100, 200 ou 350 kg/ha) (Chope *et al.* 2014). De plus, il semblerait que les TaNAC jouent un rôle dans le remplissage du grain en protéines de réserve. En effet, Uauy *et al.* (2006) ont montré l'existence d'une corrélation

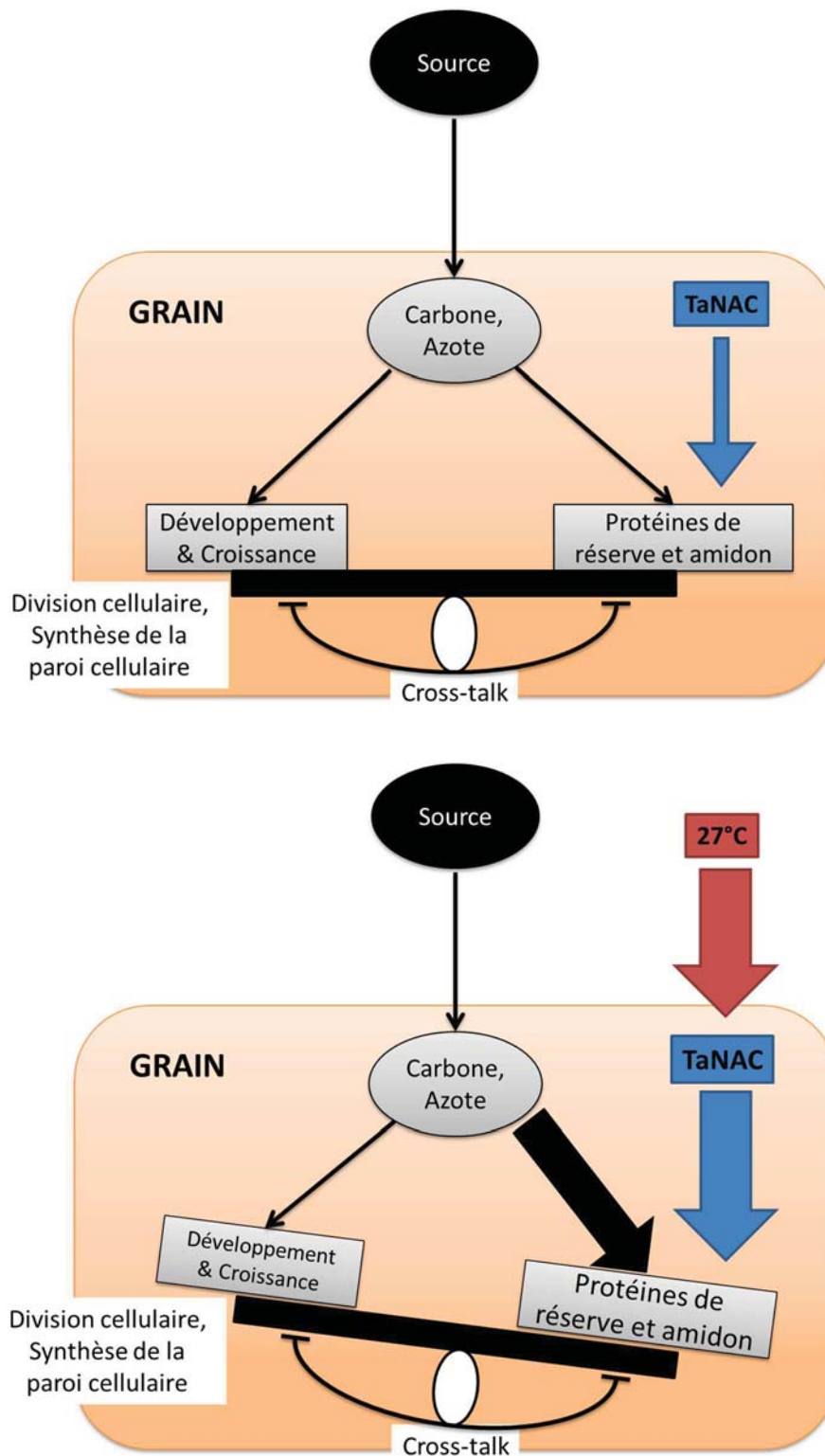


Figure 90 : Modèle hypothétique du rôle des 4 TaNAC au cours du développement du grain en conditions non-stressantes (en haut) et en réponse à un stress thermique modéré (en bas).

positive entre le niveau d'expression de facteurs TaNAM (facteurs appartenant à la famille TaNAC) et la quantité de protéines de réserve ainsi que la concentration en Zn et en Fe du grain. Ces gènes joueraient donc un rôle central dans la remobilisation des nutriments vers le grain en développement pendant la sénescence des feuilles. Gooding *et al.* (2003) ont également montré que le taux de protéines dans le grain augmentait proportionnellement avec la température moyenne. Ceci serait dû au fait que l'accumulation d'amidon dans le grain est plus sévèrement affectée par la température que l'accumulation d'azote dont la teneur est corrélée au taux de protéines final. Nadaud *et al.* (2010) ont observé une accumulation des protéines de réserve dès le début de la phase de division cellulaire, en même temps que l'expression des 4 TaNAC étudiés augmente. D'après Plessis *et al.* (2013), l'expression de 2 TaNAC est positivement et fortement associée à la proportion de gluténines (par rapport à l'azote total) contenue dans le grain. Or, ces 2 TaNAC, appelés NAC22 et NAC27 dans leur étude correspondent à TaNAC1 et TaNAC3. De plus, le facteur NAC27 (= TaNAC3) est hautement connecté au réseau de facteurs de transcription contrôlant l'expression des gènes impliqués dans l'accumulation des protéines de réserve (Plessis *et al.* 2013). L'ensemble de ces résultats conduit à émettre l'hypothèse d'une régulation directe ou indirecte de la composition du grain en protéines de réserve par les gènes TaNAC, laquelle régulation serait influencée par la température. Nous proposons ici un modèle hypothétique du rôle de ces 4 facteurs de transcription TaNAC au sein du blé tendre, en condition contrôle et en réponse à un stress thermique modéré (Figure 90).

Afin d'étudier cette hypothèse, nous avons pour objectif d'analyser les modifications physiologiques du grain liées au rendement et à la qualité (composition en protéines, en amidon et sucres solubles, nombre de cellules, poids de mille grains, etc.) en réponse à une modification de l'expression des 4 NAC étudiés par transgénèse. Cependant, l'expérimentation de validation fonctionnelle n'a pas permis de valider ces hypothèses.

9.4.2. Une réponse génotype-dépendante à la sécheresse

En réponse à la sécheresse, l'expression des 4 TaNAC dans le grain est génotype-dépendante. En effet, dans le génotype Altigo (qui présente le plus grand nombre d'épis/m²), l'expression des 4 gènes TaNAC est diminuée par la sécheresse durant les phases précoces du développement du grain. Au contraire, l'expression de TaNAC2, TaNAC3 et TaNAC4 diminue seulement au stade de remplissage du grain dans le génotype Allez-Y. TaNAC1, pour sa part,

voit son expression augmenter à 450°Cj dans ce même génotype, en réponse à la sécheresse. Ces résultats suggèrent que, selon le patrimoine génétique, ces gènes pourraient héberger différents allèles susceptibles d'affecter leur tolérance à la sécheresse. Ils suggèrent aussi que ces 4 gènes TaNAC pourraient être impliqués dans la réponse à des stress abiotiques proches (température élevée et sécheresse), comme d'autres gènes NAC tels que MINAC9 chez *Miscanthus*, SINAC3 chez la tomate, ONAC022 chez le riz, GmNAC3 and GmNAC4 chez le soja, qui ont été décrits comme impliqués dans plusieurs stress environnementaux (Pinheiro *et al.* 2009 ; Han *et al.* 2012 ; Hong *et al.* 2016 ; Zhao *et al.* 2016).

10. Validation fonctionnelle

Malgré les précautions prises, l'expérience de validation fonctionnelle n'a pas permis de générer des plantes présentant une modification de l'expression significative des gènes TmNAC visés. Les différences de phénotype observées ne sont donc pas imputables à une altération de l'expression d'un TmNAC mais plutôt à la variabilité intrinsèque de l'espèce utilisée pour la transformation. Plusieurs hypothèses peuvent être avancées pour expliquer l'absence de modification de l'expression des TmNAC ciblés. Pour les mutants de surexpression, l'utilisation de 2 promoteurs ubiquitine forts à quelques centaines de paires de bases de distance seulement peut avoir entraîné un mécanisme de silencing par la plante, qui aurait perçu cette forte activité de transcription comme une infection virale. Concernant le RNAi, les taux de sous-expression varient entre les lignées, dépendamment de la zone d'insertion de la cassette de transformation. Or, des plantes exprimant très faiblement plusieurs gènes jouant un rôle essentiel dans le développement du grain peuvent ne pas donner de grains, ou des grains qui ne germent pas (par exemple parce qu'ils n'ont pas été en mesure d'accumuler suffisamment de réserves pour permettre la survie de l'embryon). Nous pourrions donc éventuellement avoir, au cours des générations, sélectionné sans le vouloir des lignées ne présentant qu'une faible diminution de l'expression des TmNAC ciblés ; ce qui aurait eu un effet trop faible sur le phénotype pour être détecté, surtout sur un nombre de plantes aussi faible que celui que nous avions.

L'hypothèse d'une erreur dans la sélection des lignées transgénique étant la plus probable, de nouveaux grains T1 issus des plantes originellement transformées ont été semés en janvier 2019 pour recommencer le processus de sélection. De plus, nous utiliserons un nombre de plantes

plus élevées lors du phénotypage des lignées transgéniques afin de masquer l'effet de variabilité naturelle du petit épeautre.

Cependant, la cassette de sous-expression des 5 TmNAC cumulés a été dessinée en 2015. Entre temps, la publication d'une nouvelle base de données IWGS RefSeq v1.0 de blé tendre a mis en évidence l'existence de 3 gènes potentiels « off-target » chez le blé qui appartiennent eux aussi au groupe des TaNAC exprimés préférentiellement dans le grain. Si ces séquences sont suffisamment proches entre le blé tendre et le petit épeautre pour induire leur silencing, les éventuels effets de la sous-expression observés sur le phénotype ne seront pas imputables uniquement aux 5 gènes ciblés au départ mais aussi à ces off-target. Il faudrait donc identifier les homologues de ces 3 gènes chez le petit épeautre, vérifier si oui ou non ils sont touchés par l'activité de la cassette transgénique et, si c'est le cas, les inclure dans l'analyse génotypique et phénotypique. Si les mutants transgéniques de sous-expression ne sont définitivement pas exploitables, nous pourrions envisager de générer de nouveaux mutants avec la technologie CRISPR-Cas9, dont la maîtrise technique est acquise sur la plateforme de Validation Fonctionnelle, ce qui n'était pas le cas en 2015. Cette technique permettra d'éviter les « off-targets » supposés, mais les gènes NAC semblant intervenir dans le développement, il peut sembler risqué d'en éteindre complètement l'expression et de générer possiblement des plantes mutées mais stériles.

Une autre alternative serait d'utiliser des mutants TILLING (Target Induced Local Lesions In Genome), issus d'une banque de mutants pré-existante créée et caractérisée par Krasileva *et al.* (2017). La disponibilité de la dernière version de la séquence du génome de blé dont l'assemblage et l'annotation sont de meilleures qualités permettent de choisir de manière la plus optimale possible les lignées mutées dans les gènes cibles. Un criblage rapide de cette banque (<http://www.wheat-tilling.com/>) a montré la présence de mutations sur les scaffolds porteurs des 4 TaNAC étudiés. Si la présence de mutations dans les gènes est confirmée, il serait possible de rapatrier les plantes porteuses d'une mutation type codon-stop prématuré pour chacun des homéologues de chaque gène ciblé, d'en purifier les descendants par croisement avec des plantes non-mutées, pour obtenir une lignée ne portant que la mutation sur les 3 homéologues du gène considéré, et de phénotyper cette lignée. Les contraintes de cette approche sont le temps et les moyens nécessaires pour la purification.

11. Les idées clés

- Un clade regroupe 20 gènes dont 15 sont exprimés préférentiellement dans le grain et 17 présentent un motif spécifique,
- Dans ce clade, les homéologues de 4 TaNAC ont été caractérisés plus finement,
- Ces gènes interviennent probablement dans l'accumulation de protéines de réserve dans le grain, en lien avec la disponibilité en azote du milieu,
- Ces gènes participent probablement à la transition entre la phase de divisions cellulaires et la phase de remplissage. Ce sont des intégrateurs des signaux de stress abiotiques (notamment le stress thermique modéré) survenant pendant la phase de division cellulaire, qui conduisent à une altération des phases tardives du développement du grain.

DISCUSSION GENERALE, CONCLUSION et PERSPECTIVES

Le blé tendre est une céréale cultivée par l'Homme depuis 10 000 ans et les histoires évolutives de ces 2 espèces sont intimement liées. Depuis plusieurs décennies, l'activité humaine a contribué de manière négative au changement climatique, qui se traduit dans la majorité des régions du monde par une augmentation de la température moyenne couplée à des pics épisodiques de forte chaleur. L'augmentation de la température modifie la disponibilité en eau, créant un stress hydrique pour les cultures, et qui peut provoquer également un stress salin dans certaines régions. Afin de garantir la sécurité alimentaire, plusieurs perspectives d'évolution de l'agriculture sont possibles. Certaines approches s'appuient sur une adaptation des pratiques culturales. Par exemple, des études ont montré que, dans certaines régions des Etats-Unis, semer une variété de blé de printemps plus précocement permettrait de réduire l'impact négatif d'une élévation forte de la température au cours du printemps (Tubiello *et al.* 2002). D'autres approches proposent de changer régulièrement de variétés cultivées et d'utiliser les plus adaptées à l'environnement changeant des zones cultivées. En effet, une analyse de la production de maïs dans l'état de l'Iowa (U.S.A.) entre 1964 et 2013 a montré que l'épisode de sécheresse sévère et de fortes températures qui a eu lieu en 2012 a eu un impact très négatif sur le rendement. Cependant, ce rendement a tout de même été équivalent à celui que l'on pouvait obtenir en condition optimales dans les années 1980. Les auteurs ont conclu que sans l'utilisation de variétés adaptées, cet effet négatif aurait été beaucoup plus important (Atlin *et al.* 2017). Dans la mesure où les conditions de cultures varient fréquemment, ces mêmes auteurs recommandent ainsi d'adapter les potentiels génétiques par la voie de l'amélioration génétique. Si ces deux approches sont clairement différentes, elles sont en revanche complémentaires.

La création de nouvelles variétés de blé plus performantes en condition de stress abiotiques *via* l'amélioration génétique nécessite de mieux connaître les mécanismes cellulaires et moléculaires liés au développement des plantes ainsi que leurs réponses aux contraintes environnementales. Ainsi, la perception du signal par la plante et sa réponse impliquent la modulation de l'expression d'un large panel de gènes dont les facteurs de transcription sont des acteurs moléculaires importants. Plusieurs grandes familles de facteurs de transcription sont connues pour intervenir dans le développement de la plante et dans sa réponse aux stress, notamment les familles AP2/EREBP, bZIP, MYB/MYC, WRKY ou encore les NAC (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).

C'est dans ce contexte que cette thèse a été conduite, avec l'objectif de générer des connaissances supplémentaires concernant la famille NAC chez une espèce d'intérêt agronomique majeur : le blé tendre.

Nous avons tout d'abord effectué un recensement des membres de la famille TaNAC à partir de la base de données la plus récente existante du blé tendre. Ceci a permis de générer une liste de 488 membres TaNAC, dont la structuration génomique et phylogénétique a été étudiée. Nous avons aussi étudié l'histoire évolutive de cette famille *via* l'analyse des phénomènes de duplication et de rétroposition. Pour finir, nous avons pu réaliser une analyse de la diversité allélique présente au sein de cette famille TaNAC et identifier des gènes porteurs de SNP montrant une forte association avec les paramètres d'accumulation des protéines de réserve dans le grain.

Dans un second temps, nous avons observé le niveau d'expression de ces 488 gènes TaNAC dans différents organes et conditions de stress grâce à l'analyse de données transcriptomiques disponibles publiquement. Ces analyses ont permis d'identifier des gènes préférentiellement exprimés dans certains organes (grain, feuille, racine, tige) et/ou en réponse à certains stress (choc thermique, sécheresse et combinaison des deux stress). À partir de ces données, nous avons constitué un set de 23 gènes dont l'expression a été suivie *in planta*, en réponse à un choc thermique et un stress sécheresse en plein champs. Certains gènes parmi ces 23 ont montré une expression importante dans les feuilles, fortement impactée par un stress sécheresse durant la période de remplissage du grain. D'autres ont montré une expression préférentielle au grain, plus faiblement impactée par la sécheresse mais de manière génotype-dépendante. Parmi ces derniers, 4 gènes ont particulièrement retenu notre attention car leur expression est aussi impactée par un stress thermique modéré et n'avaient jamais été décrits dans la littérature.

La caractérisation génétique, moléculaire et physiologique de ces 4 facteurs de transcription TaNAC a fait l'objet du troisième volet de cette thèse. Ils appartiennent à un clade porteur d'une vingtaine de séquences dont la structure génomique est similaire, qui possèdent un motif conservé dans leur séquence codante, et qui sont exprimées préférentiellement dans le grain. De plus, les 4 gènes TaNAC présentent une localisation nucléaire et un profil d'expression similaire entre eux, avec toutefois un niveau variable entre gènes et entre homéologues de chaque gène. En réponse à un stress thermique, ce profil d'expression est accéléré au cours du développement du grain et nous avons identifié un stade clé montrant la plus grande différence d'expression de ces gènes entre les conditions contrôle et stressée : le stade 120°Cj. Il apparaît donc que, dans le cas d'un stress thermique modéré, l'expression des gènes s'accélère et le grain

Tableau 40 : Nombre moyen d'intron(s) dans la séquence génomique des gènes TaNAC, pour chacune des 11 sous-familles décrites chez le blé tendre.

Sous-famille	Nombre d'intron(s)
NAC-g	4,43
NAC-k	4
NAC-b	3,16
NAC-j	3
NAC-i	2,21
NAC-c	2
NAC-e	1,72
NAC-a	1,67
NAC-d	1,63
NAC-f	1,5
NAC-h	0,61

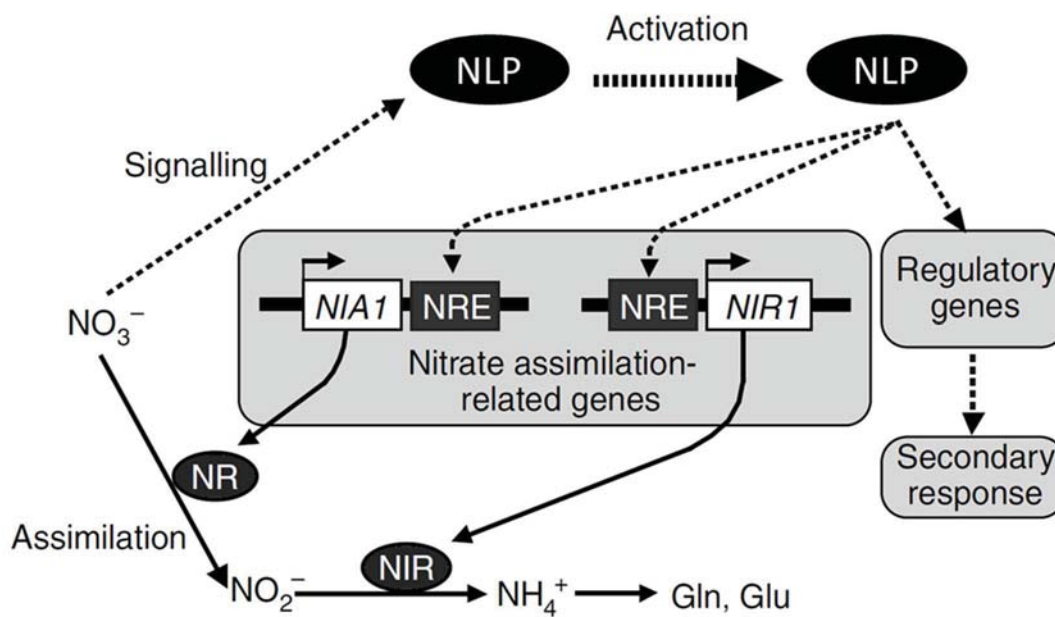


Figure 91 : Modèle de la fonction des gènes NLP dans la voie de réponse à l'azote (d'après Konishi & Yanagisawa 2013) . Le signal azote convertit la forme inactive de la protéine NLP en forme active. Cette protéine active interagit avec les nitrate-responsive elements (NRE) et induit l'expression de gènes azote-inductibles. Comme les NLP activent à la fois des gènes impliqués dans l'assimilation des nitrates et des gènes régulateurs (y compris des facteurs de transcription), l'activation des NLP induite par les nitrates peut produire un large panel de réponses secondaires.

quitte la phase de division cellulaire pour entrer en phase de remplissage plus rapidement. Pour pouvoir avancer des hypothèses plus précises quant à la fonction de ces gènes dans le grain et en réponse à l'élévation de température, nous avons tenté de produire des plantes transgéniques sur- et sous-exprimant ces gènes. Malheureusement, cette expérimentation n'a pas abouti.

Parmi les gènes TaNAC montrant une expression préférentielle à un organe et/ou dont l'expression était impactée par des stress abiotiques, la majorité appartiennent à la sous-famille NAC-d, qui est connue chez beaucoup d'espèces pour intervenir dans le développement de la plante et sa réponse à plusieurs stress. En observant la structure génomique de cette famille chez le riz, Liu *et al.* (2018) ont constaté que la duplication segmentaire dans la famille NAC se traduit par un taux de perte d'intron plus rapide que le taux de gain d'intron. Chez le blé, il apparaît que le nombre d'introns moyen varie au sein de chaque sous-famille (Tableau 40), ce qui permet d'avancer l'hypothèse que les sous-familles NAC-f, NAC-h ou NAC-d seraient les plus susceptibles d'être issues de duplication. Précédemment, nous avons présenté l'hypothèse de Murozuka *et al.* (2018) selon laquelle le groupe des NAC exprimé préférentiellement dans le grain chez l'orge serait issu d'une duplication de gènes NAC s'exprimant dans la feuille (sénescence foliaire). Les gènes dupliqués auraient subi une néo-fonctionnalisation suite à l'acquisition de promoteurs spécifiques du grain. Or, les TaNAC exprimés préférentiellement dans le grain chez le blé appartiennent à l'une des sous-familles possédant le moins d'introns (NAC-d), ce qui suggère l'existence d'un mécanisme similaire de duplication chez le blé.

Les résultats de ma thèse ont permis de montrer que certains gènes TaNAC sont porteurs d'un polymorphisme qui semble lié à l'accumulation des protéines de réserve dans le grain de blé tendre. Or, la modulation de l'accumulation de ces protéines de réserve est un objectif des programmes d'amélioration variétale, du fait des utilisations de plus en plus variées du blé dont certaines (boulangerie industrielle, alimentation animale par exemple) requièrent un taux de protéines plus élevé que d'autres (Morris et Meleard 2015). Une caractérisation plus fine (phénotypage d'un plus grand nombre de lignées) de la variabilité allélique des gènes concernés et de son lien avec la qualité du grain pourrait permettre de valider ces gènes candidats qui, après sélection d'allèles favorables, permettraient de réguler le pourcentage de protéines de réserve dans les grains.

Par ailleurs, nous avons montré qu'une augmentation de 8°C de la température moyenne pendant la phase de divisions cellulaires dans le grain de blé entraînait une diminution du nombre de cellules dans l'albumen et une diminution du rendement d'environ 17% (Girousse, Roche, Guérin *et al.* 2018). Au cours de cette étude, nous avons également montré que

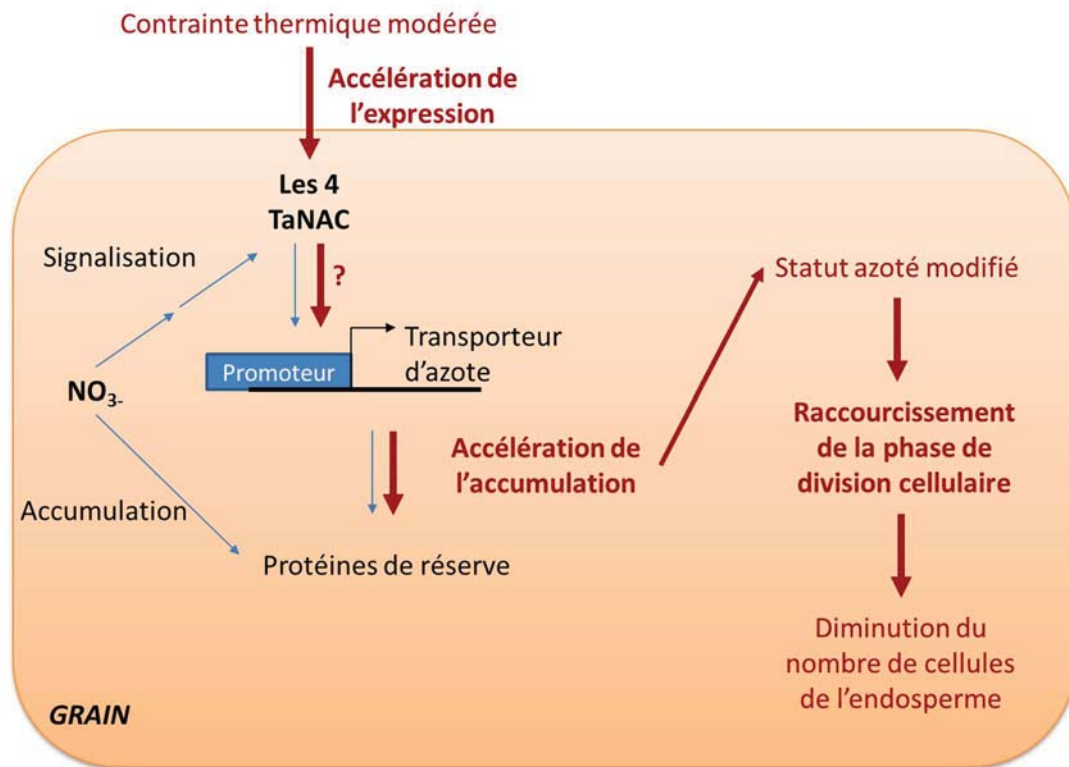


Figure 92 : Modèle hypothétique de l'impact des 4 TaNAC sur le statut azoté au cours du développement du grain, et en réponse à un stress thermique modéré. Une augmentation modérée de la température induit une accélération de l'expression des 4 TaNAC étudiés, ce qui pourrait influencer sur l'accumulation de protéines de réserve dans le grain de blé. La modification du statut azoté au profit de l'accumulation des protéines de réserve diminuerait l'azote disponible pour la signalisation et engendrerait des réponses développementales via la régulation des facteurs de transcription TaNAC, qui pourraient se fixer sur les promoteurs de gènes de réponse. Ces gènes interviendraient dans le transport et l'assimilation de l'azote, mais seraient aussi des facteurs de transcription pouvant induire par la suite des cascades de réponse régulant la synthèse d'ARN, de protéine et l'expansion cellulaire.

l'expression de 4 facteurs de transcription TaNAC est induite de manière plus précoce par l'élévation de la température. Si on suppose l'hypothèse selon laquelle les 4 gènes TaNAC seraient impliqués dans la régulation de la synthèse des protéines de réserve, nous pouvons supposer que leur action pourrait influencer la teneur en azote des cellules de l'albumen en réponse au stress thermique. Ceci pourrait constituer un signal conduisant à des réponses physiologiques et développementales telles qu'une réduction du nombre de cellules. En effet, certains gènes NAC ont déjà été décrits comme des intégrateurs du signal azoté chez plusieurs céréales, par exemple TaNAC2-5A (différent du TaNAC2 décrit dans notre publication) chez le blé, OsNAC19 chez le riz et ZmNAC67 chez le maïs dont l'expression est induite par un traitement azoté (He *et al.* 2015). TaNAC2-5A est exprimé préférentiellement dans les racines, où il se fixe sur le promoteur de gènes codant pour des transporteurs d'azote (TaNRT2.1-B1, TaNPF7.1-D1 et TaGS2-2A) et régule positivement leur expression. Lorsqu'il est surexprimé chez le blé, TaNAC2-5A engendre une augmentation du rendement et de l'accumulation d'azote dans les parties aériennes, et induit l'allocation d'une quantité plus importante d'azote dans les grains. D'autres gènes, tels que les facteurs de transcription NLP (Nodule INception-Like Proteins) ont été identifiés comme activateurs non seulement des réponses d'assimilation ou de réduction de l'azote cellulaire en induisant l'expression de transporteurs et des réducteurs d'azote, mais également des médiateurs de réponses développementales chez *Arabidopsis* (Konishi et Yanagisawa 2013). En effet, il a été montré qu'un apport en azote stimule l'expression des gènes de biosynthèse de nucléotides et d'acides aminés, de synthèse et maturation des ARN et des protéines, et de gènes intervenant dans les mécanismes d'expansion et de croissance cellulaire (Scheible *et al.* 2004). Ainsi, Konishi et Yanagisawa (2013) ont proposé un modèle moléculaire liant le statut azoté au développement végétal par les facteurs de transcription NLP (Figure 91).

Ici, nous avons proposé un modèle similaire incluant les 4 TaNAC étudiés et leur fonction potentielle dans le grain au cours du développement et en réponse à la température. En condition optimales, l'homéostasie de l'azote conduit aux réponses développementales normales du grain de blé. Lors d'une élévation de la température, le profil d'expression de nombreux gènes (dont les facteurs de transcription TaNAC) est avancé dans le temps, ce qui conduirait à une induction plus précoce de gènes codant pour des protéines de réserves et par conséquent à une modification du statut azoté (Figure 92). L'avancement de l'entrée en phase de remplissage conduirait logiquement à un raccourcissement de la phase précédente, la phase de divisions

cellulaires, ce qui provoquerait une diminution du nombre final de cellules dans le grain (compartiment albumen).

Ce modèle peut encore être généralisé en impliquant le dialogue azote-carbone et les réponses développementales du grain de blé. S'il est intéressant, ce modèle reste néanmoins hypothétique et doit être validé expérimentalement. Dans cet objectif, nous pourrions envisager :

- d'étudier les mécanismes de régulation de l'expression des 4 gènes TaNAC. Connaître les protéines induisant leur expression nous renseignerait sur les liens de ces 4 facteurs avec les voies de signalisation liées à l'azote et au stress thermique. Pour cela, je propose une analyse préliminaire des éléments de régulation présents dans les promoteurs qui donnera des interacteurs protéiques potentiels. Nous pourrions ensuite valider leur capacité de fixation sur les promoteurs par la technique du gel-retard.
- d'étudier la capacité des 4 protéines TaNAC à se fixer sur le promoteur de gènes codant pour des transporteurs d'azote dans le grain. Cette hypothèse pourrait être testée par la technique de simple hybride en système hétérologue avec les 4 TaNAC utilisées comme protéines appât. Dans le même temps, l'hypothèse éventuelle de l'interaction de ces 4 protéines (comme cela a été constaté chez le riz par Mathew *et al.* (2016)) pourra être observée.
- d'étudier le lien entre l'expression de ces 4 gènes TaNAC, le statut azoté et l'accumulation protéique. Pour cela, une deuxième sélection de plantes transgéniques (démarrée actuellement dans la continuité de ces travaux de thèse) devrait fournir un matériel végétal qui, soumis à des apports variés en azote, apportera des réponses quant à l'accumulation des protéines dans le grain en lien avec l'expression des 4 gènes TaNAC. Nous pouvons aussi envisager l'utilisation de mutants TILLING purifiés.

REFERENCES BIBLIOGRAPHIQUE

- Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) Function as Transcriptional Activators in Abscissic Acid Signaling. *The Plant Cell*. 2003;15(1):63-78.
- Abhinandan K, Skori L, Stanic M, Hickerson NM, Jamshed M, Samuel MA. Abiotic Stress Signaling in Wheat – An Inclusive Overview of Hormonal Interactions During Abiotic Stress Responses in Wheat. *Front Plant Sci*. 2018;9.
- Acevedo E, Silva P, Silva H. Wheat growth and physiology. In: Rajaram S, Curtis BC, Gómez Macpherson H, éditeurs. *BREAD WHEAT : Improvement and Production*. Italie; 2002.
- Adonina IG, Goncharov NP, Badaeva ED, Sergeeva EM, Petrash NV, Salina EA. (GAA)n microsatellite as an indicator of the A genome reorganization during wheat evolution and domestication. *Comparative Cytogenetics*. 2015;9(4):533-47.
- Ahmadi A, Baker DA. The effect of water stress on the activities of key regulatory enzymes of the sucrose to starch pathway in wheat. *Plant Growth Regulation*. 2001;35:81–91.
- Ahn H, Jung I, Shin S-J, Park J, Rhee S, Kim J, et al. Transcriptional Network Analysis Reveals Drought Resistance Mechanisms of AP2/ERF Transgenic Rice. *Front Plant Sci*. 2017;8.
- Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M. Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. *Plant Cell*. 1997;9(6):841-57.
- Akhunov ED, Goodyear AW, Geng S, Qi L-L, Echalié B, Gill BS, et al. The Organization and Rate of Evolution of Wheat Genomes Are Correlated With Recombination Rates Along Chromosome Arms. *Genome Res*. 2003;13(5):753-63.
- Allen RG, Pereira LS, Raes D, Smith M. Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56. 1998;15.
- Altenbach SB, DuPont FM, Kothari KM, Chan R, Johnson EL, Lieu D. Temperature, Water and Fertilizer Influence the Timing of Key Events During Grain Development in a US Spring Wheat. *Journal of Cereal Science*. 2003;37(1):9-20.
- Apel K, Hirt H. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu Rev Plant Biol*. 2004;55:373-99.
- Appels R, Consortium (IWGSC) TIWGS, Investigators IR principal, Eversole K, Feuillet C, Keller B, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*. 2018;361(6403):eaar7191.
- Asseng S, Ewert F, Martre P, Rötter RP, Lobell DB, Cammarano D, et al. Rising temperatures reduce global wheat production. *Nature Climate Change*. 2015;5(2):143-7.
- Atlin GN, Cairns JE, Das B. Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. *Global Food Security*. 2017;12:31-7.
- Avni R, Zhao R, Pearce S, Jun Y, Uauy C, Tabbita F, et al. Functional characterization of GPC-1 genes in hexaploid wheat. *Planta*. 2014;239(2):313-24.
- Balazadeh S, Kwasniewski M, Caldana C, Mehrnia M, Zanor MI, Xue G-P, et al. ORS1, an H₂O₂-Responsive NAC Transcription Factor, Controls Senescence in Arabidopsis thaliana. *Molecular Plant*. 2011;4(2):346-60.

- Balazadeh S, Riaño-Pachón DM, Mueller-Roeber B. Transcription factors regulating leaf senescence in *Arabidopsis thaliana*. *Plant Biology*. 2008;10(s1):63-75.
- Balazadeh S, Siddiqui H, Allu AD, Matallana-Ramirez LP, Caldana C, Mehrnia M, et al. A gene regulatory network controlled by the NAC transcription factor ANAC092/AtNAC2/ORE1 during salt-promoted senescence. *The Plant Journal*. 2010;62(2):250-64.
- Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P, Boutet G, et al. A worldwide bread wheat core collection arrayed in a 384-well plate. *Theor Appl Genet*. 2007;114(7):1265-75.
- Baloglu MC, Inal B, Kavas M, Unver T. Diverse expression pattern of wheat transcription factors against abiotic stresses in wheat species. *Gene*. 2014;550(1):117-22.
- Baloglu MC, Oz MT, Oktem HA, Yucel M. Expression Analysis of TaNAC69-1 and TtNAMB-2, Wheat NAC Family Transcription Factor Genes Under Abiotic Stress Conditions in Durum Wheat (*Triticum turgidum*). *Plant Mol Biol Rep*. 2012;30(5):1246-52.
- Barak S, Mudgil D, Khatkar BS. Relationship of gliadin and glutenin proteins with dough rheology, flour pasting and bread making performance of wheat varieties. *LWT - Food Science and Technology*. 2013;51(1):211-7.
- Barnabás B, Jäger K, Fehér A. The effect of drought and heat stress on reproductive processes in cereals. *Plant, Cell & Environment*. 2008;31(1):11-38.
- Barron C, Surget A, Rouau X. Relative amounts of tissues in mature wheat (*Triticum aestivum* L.) grain and their carbohydrate and phenolic acid composition. *Journal of Cereal Science*. 2007;45:88–96.
- Bartoli CG, Casalongué CA, Simontacchi M, Marquez-Garcia B, Foyer CH. Interactions between hormone and redox signalling pathways in the control of growth and cross tolerance to stress. *Environmental and Experimental Botany*. 2013;94:73-88.
- Baxter A, Mittler R, Suzuki N. ROS as key players in plant stress signalling. *J Exp Bot*. 2014;65(5):1229-40.
- Bechtel DB, Zayas I, Kaleikau L, Pomeranz Y. Size-distribution of wheat starch granules during endosperm development. *Cereal Chemistry*. 1990;67(1):59-63.
- Bhattacharjee P, Das R, Mandal A, Kundu P. Functional characterization of tomato membrane-bound NAC transcription factors. *Plant Molecular Biology*. 2017;93(4-5):511-32.
- Bie X, Wang K, She M, Du L, Zhang S, Li J, et al. Combinational transformation of three wheat genes encoding fructan biosynthesis enzymes confers increased fructan content and tolerance to abiotic stresses in tobacco. *Plant Cell Reports*. 2012;31(12):2229-38.
- Blomster T, Salojärvi J, Sipari N, Brosche M, Ahlfors R, Keinänen M, et al. Apoplastic Reactive Oxygen Species Transiently Decrease Auxin Signaling and Cause Stress-Induced Morphogenic Response in *Arabidopsis*. *PLANT PHYSIOLOGY*. 2011;157(4):1866-83.
- Blum A. Crop responses to drought and the interpretation of adaptation. *Plant Growth Regulation*. 1996;20(2):135-48.
- Bogorad L, Gubbins EJ, Krebbers E, Larrinua IM, Mulligan BJ. Cloning and physical mapping of maize plastid genes [*Zea mays*, corn]. *Methods in Enzymology*. 1983;97(1):524-554.

- Bordes J, Ravel C, Jaubertie JP, Duperrier B, Gardet O, Heumez E, et al. Genomic regions associated with the nitrogen limitation response revealed in a global wheat core collection. *Theoretical and Applied Genetics*. 2013;126(3):805-22.
- Bordes J, Ravel C, Le Gouis J, Lapiere A, Charmet G, Balfourier F. Use of a global wheat core collection for association analysis of flour and dough quality traits. *Journal of Cereal Science*. 2011;54(1):137-47.
- Borrill P, Harrington SA, Uauy C. Genome-Wide Sequence and Expression Analysis of the NAC Transcription Factor Family in Polyploid Wheat. *Genes|Genomes|Genetics*. 2017;7(9):3019-29.
- Borrill P, Ramirez-Gonzalez R, Uauy C. expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform. *Plant Physiol*. 2016;170(4):2172-86.
- Brooks A, Jenner CF, Aspinall D. Effects of Water Deficit on Endosperm Starch Granules and on Grain Physiology of Wheat and Barley. *Functional Plant Biol*. 1982;9(4):423-36.
- Brosius J. Retroposons--seeds of evolution. *Science*. 1991;251(4995):753.
- Burrell MM. Starch: the need for improved quality or quantity--an overview. *Journal of Experimental Botany*. 2003;54(382):451-6.
- Calderini DF, Ortiz-Monasterio I. Grain position affects grain macronutrient and micronutrient concentrations in wheat. *Crop Science*. 2003;43:141-51.
- Calderini DF, Reynolds MP. Changes in grain weight as a consequence of de-graining treatments at pre- and post-anthesis in synthetic hexaploid lines of wheat (*Triticum durum* x *T. tauschii*). *Functional Plant Biol*. 2000;27(3):183-91.
- Caley CY, Duffus CM, Jeffcoat B. Effects of Elevated Temperature and Reduced Water Uptake on Enzymes of Starch Synthesis in Developing Wheat Grains. *Functional Plant Biol*. 1990;17(4):431-9.
- Campbell KG, Bergman CJ, Gualberto DG, Anderson JA, Giroux MJ, Hareland G, et al. Quantitative Trait Loci Associated with Kernel Traits in a Soft x Hard Wheat Cross. *Crop Science*. 1999;39(4):1184.
- Cao L, Yu Y, Ding X, Zhu D, Yang F, Liu B, et al. The Glycine soja NAC transcription factor GsNAC019 mediates the regulation of plant alkaline tolerance and ABA sensitivity. *Plant Molecular Biology*. 2017;95(3):253-68.
- Capron D, Mouzeyar S, Boulaflous A, Girousse C, Rustenholz C, Laugier C, et al. Transcriptional profile analysis of E3 ligase and hormone-related genes expressed during wheat grain development. *BMC Plant Biol*. 2012;12:35.
- Cenci A, Guignon V, Roux N, Rouard M. Genomic analysis of NAC transcription factors in banana (*Musa acuminata*) and definition of NAC orthologous groups for monocots and dicots. *Plant Molecular Biology*. 2014;85(1-2):63-80.
- Chakraborty U, Chakraborty B. *Abiotic Stresses in Crop Plants*. CABI; 2015.
- Chauhan H, Khurana N, Nijhavan A, Khurana JP, Khurana P. The wheat chloroplastic small heat shock protein (sHSP26) is involved in seed maturation and germination and imparts tolerance to heat stress. *Plant, Cell & Environment*. 2012;35(11):1912-31.
- Chen D, Chai S, McIntyre CL, Xue G-P. Overexpression of a predominantly root-expressed NAC transcription factor in wheat roots enhances root length, biomass and drought tolerance. *Plant Cell Reports*. 2018;37(2):225-37.

- Chen D, Richardson T, Chai S, Lynne McIntyre C, Rae AL, Xue G-P. Drought-Up-Regulated TaNAC69-1 is a Transcriptional Repressor of TaSHY2 and TaIAA7, and Enhances Root Length and Biomass in Wheat. *Plant Cell Physiol.* 2016a;57(10):2076-90.
- Chen H, Hwang JE, Lim CJ, Kim DY, Lee SY, Lim CO. Arabidopsis DREB2C functions as a transcriptional activator of HsfA3 during the heat stress response. *Biochem Biophys Res Commun.* 2010;401(2):238-44.
- Chen S-P, Lin IW, Chen X, Huang Y-H, Chang S-C, Lo H-S, et al. Sweet potato NAC transcription factor, IbNAC1, upregulates *sporamin* gene expression by binding the SWRE motif against mechanical wounding and herbivore attack. *The Plant Journal.* 2016b;86(3):234-48.
- Chen X, Wang Y, Lv B, Li J, Luo L, Lu S, et al. The NAC Family Transcription Factor OsNAP Confers Abiotic Stress Response Through the ABA Pathway. *Plant and Cell Physiology.* 2014;55(3):604-19.
- Chi YH, Melencion SMB, Alinapon CV, Kim MJ, Lee ES, Paeng SK, et al. The membrane-tethered NAC transcription factor, AtNTL7, contributes to ER-stress resistance in Arabidopsis. *Biochemical and Biophysical Research Communications.* 2017;488(4):641-7.
- Chojcecki AJS, Bayliss MW, Gale MD. Cell Production and DNA Accumulation in the Wheat Endosperm, and their Association with Grain Weight. *Ann Bot.* 1986;58(6):809-17.
- Chope GA, Wan Y, Penson SP, Bhandari DG, Powers SJ, Shewry PR, et al. Effects of Genotype, Season, and Nitrogen Nutrition on Gene Expression and Protein Accumulation in Wheat Grain. *Journal of Agricultural and Food Chemistry.* 2014;62(19):4399-407.
- Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, et al. Structural and functional partitioning of bread wheat chromosome 3B. *Science.* 2014;345(6194):1249721.
- Christiansen MW, Holm PB, Gregersen PL. Characterization of barley (*Hordeum vulgare* L.) NAC transcription factors suggests conserved functions compared to both monocots and dicots. *BMC Res Notes.* 2011;4:302.
- Christianson JA, Dennis ES, Llewellyn DJ, Wilson IW. ATAF NAC transcription factors: Regulators of plant stress signaling. *Plant Signaling & Behavior.* 2010;5(4):428-32.
- Christov NK, Christova PK, Kato H, Liu Y, Sasaki K, Imai R. TaSK5, an abiotic stress-inducible GSK3/shaggy-like kinase from wheat, confers salt and drought tolerance in transgenic Arabidopsis. *Plant Physiology and Biochemistry.* 2014;84:251-60.
- Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, et al. An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Research.* 2017;27(5):885-96.
- Cooper M, Woodruff D., Phillips I., Basford K., Gilmour A. Genotype-by-management interactions for grain yield and grain protein concentration of wheat. *Field Crops Research.* 2001;69(1):47-67.
- Cormier F, Throude M, Ravel C, Gouis J, Leveugle M, Lafarge S, et al. Detection of NAM-A1 Natural Variants in Bread Wheat Reveals Differences in Haplotype Distribution between a Worldwide Core Collection and European Elite Germplasm. *Agronomy.* 2015;5(2):143-51.

- Cosségal M, Vernoud V, Depège N, Rogowsky PM. The Embryo Surrounding Region. In: Olsen O-A, éditeur. *Endosperm: Developmental and Molecular Biology*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 57-71. (Plant Cell Monographs).
- Craufurd PQ, Vadez V, Jagadish SVK, Prasad PVV, Zaman-Allah M. Crop science experiments designed to inform crop modeling. *Agricultural and Forest Meteorology*. 2013;170:8-18.
- Delessert C, Kazan K, Wilson IW, Van Der Straeten D, Manners J, Dennis ES, et al. The transcription factor ATAF2 represses the expression of pathogenesis-related genes in *Arabidopsis*. *Plant J*. 2005;43(5):745-57.
- Dholakia BB, Ammiraju JSS, Singh H, Lagu MD, Röder MS, Rao VS, et al. Molecular marker analysis of kernel size and shape in bread wheat. *Plant Breeding*. 2003;122(5):392-5.
- Distelfeld A, Pearce SP, Avni R, Scherer B, Uauy C, Piston F, et al. Divergent functions of orthologous NAC transcription factors in wheat and rice. *Plant Molecular Biology*. 2012;78(4-5):515-24.
- Dodd GL, Donovan LA. Water potential and ionic effects on germination and seedling growth of two cold desert shrubs. *American Journal of Botany*. 1999;86(8):1146-53.
- Dolferus R, Ji X, Richards RA. Abiotic stress and control of grain number in cereals. *Plant Science*. 2011;181(4):331-41.
- Drea S, Leader DJ, Arnold BC, Shaw P, Dolan L, Doonan JH. Systematic Spatial Analysis of Gene Expression during Wheat Caryopsis Development. *Plant Cell*. 2005;17(8):2172-85.
- Duan M, Zhang R, Zhu F, Zhang Z, Gou L, Wen J, et al. A Lipid-Anchored NAC Transcription Factor Is Translocated into the Nucleus and Activates *Glyoxalase I* Expression during Drought Stress. *The Plant Cell*. 2017;29(7):1748-72.
- Duval M, Hsieh T-F, Kim SY, Thomas TL. Molecular characterization of AtNAM: a member of the *Arabidopsis* NAC domain superfamily. *Plant Mol Biol*. 2002;50(2):237-48.
- Earley KW, Haag JR, Pontes O, Opper K, Juehne T, Song K, et al. Gateway-compatible vectors for plant functional genomics and proteomics. *The Plant Journal*. 2006;45(4):616-29.
- Ebrahimian-Motlagh S, Ribone PA, Thirumalaikumar VP, Allu AD, Chan RL, Mueller-Roeber B, et al. JUNGBRUNNEN1 Confers Drought Tolerance Downstream of the HD-Zip I Transcription Factor AtHB13. *Frontiers in Plant Science*. 2017;8.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792-7.
- Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26(19):2460-1.
- Egli DB. Seed biology and the yield of grain crops. Wallingford, Oxon ; New York : CAB International; 1998.
- Endelman JB. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome*. 2011;4(3):250-5.
- Ergen NZ, Thimmapuram J, Bohnert HJ, Budak H. Transcriptome pathways unique to dehydration tolerant relatives of modern wheat. *Functional & Integrative Genomics*. 2009;9(3):377-96.

- Erickson AN, Markhart AH. Flower developmental stage and organ sensitivity of bell pepper (*Capsicum annuum* L.) to elevated temperature. *Plant, Cell & Environment*. 2002;25(1):123-30.
- Ernst HA, Nina Olsen A, Skriver K, Larsen S, Lo Leggio L. Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. *EMBO reports*. 2004;5(3):297-303.
- Everaert C, Luypaert M, Maag JLV, Cheng QX, Dinger ME, Hellemans J, et al. Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Scientific Reports*. 2017;7(1).
- Evers AD, Bechtel DB. Microscopic structure of the wheat grain. *Wheat: chemistry and technology*. Vol. I. St. Paul, Minnesota, U.S.A.: Y. Pomeranz; 1988.
- Evers AD, Blakeney AB, O'Brien L. Cereal structure and composition. *Australian Journal of Agricultural Research*. 1999;50:629-50.
- Evers T, Millar S. Cereal Grain Structure and Development: Some Implications for Quality. *Journal of Cereal Science*. 2002;36(3):261-84.
- Fan K, Li F, Chen J, Li Z, Lin W, Cai S, et al. Asymmetric Evolution and Expansion of the NAC Transcription Factor in Polyploidized Cotton. *Frontiers in Plant Science*. 2018a;9.
- Fan Z, Tan X-L, Chen J, Liu Z, Kuang J, Lu W, et al. BrNAC055, a Novel Transcriptional Activator, Regulates Leaf Senescence in Chinese Flowering Cabbage by Modulating Reactive Oxygen Species Production and Chlorophyll Degradation. *Journal of Agricultural and Food Chemistry*. 2018b;66(36):9399-408.
- Fang L, Su L, Sun X, Li X, Sun M, Karungo SK, et al. Expression of *Vitis amurens* NAC26 in *Arabidopsis* enhances drought tolerance by modulating jasmonic acid synthesis. *Journal of Experimental Botany*. 2016;67(9):2829-45.
- Fang Y, Liao K, Du H, Xu Y, Song H, Li X, et al. A stress-responsive NAC transcription factor SNAC3 confers heat and drought tolerance through modulation of reactive oxygen species in rice. *Journal of Experimental Botany*. 2015;66(21):6803-17.
- Fang Y, You J, Xie K, Xie W, Xiong L. Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. *Molecular Genetics and Genomics*. 2008;280(6):547-63.
- Feillet P. Le grain de blé: composition et utilisation. Paris: Institut national de la recherche agronomique (INRA); 2000.
- Feng H, Duan X, Zhang Q, Li X, Wang B, Huang L, et al. The target gene of tae-miR164, a novel NAC transcription factor from the NAM subfamily, negatively regulates resistance of wheat to stripe rust: TaNAC21/22 regulates resistance of wheat to *P. st.* *Molecular Plant Pathology*. 2014;15(3):284-96.
- Feuillet C, Langridge P, Waugh R. Cereal breeding takes a walk on the wild side. *Trends in Genetics*. 2008;24(1):24-32.
- Flückiger R, De Caroli M, Piro G, Dalessandro G, Neuhaus J-M, Di Sansebastiano G-P. Vacuolar system distribution in *Arabidopsis* tissues, visualized using GFP fusion proteins. *J Exp Bot*. 2003;54(387):1577-84.
- Francois LE, Donovan T, Maas EV. Salinity Effects on Seed Yield, Growth, and Germination of Grain Sorghum 1. *Agronomy Journal*. 1984;76(5):741-4.

- Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc Natl Acad Sci USA*. 2014;111(6):2367-72.
- Fu D, Uauy C, Blechl A, Dubcovsky J. RNA interference for wheat functional gene analysis. *Transgenic Research*. 2007;16(6):689-701.
- Fujita M, Fujita Y, Maruyama K, Seki M, Hiratsu K, Ohme-Takagi M, et al. A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *Plant J*. 2004;39(6):863-76.
- Gahlaut V, Jaiswal V, Kumar A, Gupta PK. Transcription factors involved in drought tolerance and their possible role in developing drought tolerant cultivars with emphasis on wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*. 2016;129(11):2019-42.
- Gapper C, Dolan L. Control of plant development by reactive oxygen species. *Plant Physiol*. 2006;141(2):341-5.
- Gerken AR, Eller OC, Hahn DA, Morgan TJ. Constraints, independence, and evolution of thermal plasticity: probing genetic architecture of long- and short-term thermal acclimation. *Proc Natl Acad Sci USA*. 2015;112:4399–4404.
- Gienapp P, Teplitsky P, Alho JS, Mills JA, Merilä J. Climate change and evolution: disentangling environmental and genetic responses. *Mol. Ecol*. 2008;17:167–178.
- Girousse C, Roche J, Guérin C, Le Gouis J, Balzegue S, Mouzeyar S, et al. Coexpression network and phenotypic analysis identify metabolic pathways associated with the effect of warming on grain yield components in wheat. Zhang A, éditeur. *PLOS ONE*. 2018;13(6):e0199434.
- Glover NM, Daron J, Pingault L, Vandepoele K, Paux E, Feuillet C, et al. Small-scale gene duplications played a major role in the recent evolution of wheat chromosome 3B. *Genome Biol*. 2015;16:188.
- Gooding MJ, Ellis RH, Shewry PR, Schofield JD. Effects of Restricted Water Availability and Increased Temperature on the Grain Filling, Drying and Quality of Winter Wheat. *Journal of Cereal Science*. 2003;37(3):295-309.
- Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 2008;36(10):3420-35.
- Gouache D, Bancal M-O, Bancal P, De Solan B, Gate P. Tolérance du blé tendre aux stress biotiques et abiotiques. *Innovations Agronomiques*. 2014;35:75-87.
- Greve K, Cour TL, Jensen MK, Poulsen FM, Skriver K. Interactions between plant RING-H2 and plant-specific NAC (NAM/ATAF1/2/CUC2) proteins: RING-H2 molecular specificity and cellular localization. *Biochemical Journal*. 2003;371(1):97-108.
- Guan Q, Yue X, Zeng H, Zhu J. The Protein Phosphatase RCF2 and Its Interacting Partner NAC019 Are Critical for Heat Stress-Responsive Gene Regulation and Thermotolerance in Arabidopsis. *The Plant Cell*. 2014;26(1):438-53.
- Guérin C, Roche J, Allard V, Ravel C, Mouzeyar S, Bouzidi MF. Genome-wide analysis, expansion and expression of the NAC family under drought and heat stresses in bread wheat (*T. aestivum* L.). *PLoS One*. 2019;14(3):e0213390.

- Guo H-S, Xie Q, Fei J-F, Chua N-H. MicroRNA Directs mRNA Cleavage of the Transcription Factor NAC1 to Downregulate Auxin Signals for Arabidopsis Lateral Root Development. *THE PLANT CELL ONLINE*. 2005;17(5):1376-86.
- Guo S, Dai S, Singh PK, Wang H, Wang Y, Tan JLH, et al. A Membrane-Bound NAC-Like Transcription Factor OsNTL5 Represses the Flowering in *Oryza sativa*. *Frontiers in Plant Science*. 2018;9.
- Guo W, Zhang J, Zhang N, Xin M, Peng H, Hu Z, et al. The Wheat NAC Transcription Factor TaNAC2L Is Regulated at the Transcriptional and Post-Translational Levels and Promotes Heat Stress Tolerance in Transgenic Arabidopsis. Gaquerel E, éditeur. *PLOS ONE*. 2015;10(8):e0135667.
- Guo Y, Gan S. AtNAP, a NAC family transcription factor, has an important role in leaf senescence. *The Plant Journal*. 2006;46(4):601-12.
- Han Q, Zhang J, Li H, Luo Z, Ziaf K, Ouyang B, et al. Identification and expression pattern of one stress-responsive NAC gene from *Solanum lycopersicum*. *Mol Biol Rep*. 2012;39(2):1713-20.
- Han X, Feng Z, Xing D, Yang Q, Wang R, Qi L, et al. Two NAC transcription factors from *Caragana intermedia* altered salt tolerance of the transgenic Arabidopsis. *BMC Plant Biology*. 2015;15(1).
- Hansen J, Sato M, Hearty P, Ruedy R, Kelley M, Masson-Delmotte V, et al. Ice melt, sea level rise and superstorms: evidence from paleoclimate data, climate modeling, and modern observations that 2 °C global warming could be dangerous. *Atmospheric Chemistry and Physics*. 2016;16(6):3761-812.
- Hao Y-J, Wei W, Song Q-X, Chen H-W, Zhang Y-Q, Wang F, et al. Soybean NAC transcription factors promote abiotic stress tolerance and lateral root formation in transgenic plants. *Plant J*. 2011;68(2):302-13.
- Hasanuzzaman M, Nahar K, Alam MM, Roychowdhury R, Fujita M. Physiological, Biochemical, and Molecular Mechanisms of Heat Stress Tolerance in Plants. *Int J Mol Sci*. 2013;14(5):9643-84.
- He G-H, Xu J-Y, Wang Y-X, Liu J-M, Li P-S, Chen M, et al. Drought-responsive WRKY transcription factor genes TaWRKY1 and TaWRKY33 from wheat confer drought and/or heat resistance in Arabidopsis. *BMC Plant Biology*. 2016;16(1).
- He L, Shi X, Wang Y, Guo Y, Yang K, Wang Y. Arabidopsis ANAC069 binds to C[A/G]CG[T/G] sequences to negatively regulate salt and osmotic stress tolerance. *Plant Molecular Biology*. 2017;93(4-5):369-87.
- He X, Qu B, Li W, Zhao X, Teng W, Ma W, et al. The Nitrate-Inducible NAC Transcription Factor TaNAC2-5A Controls Nitrate Response and Increases Wheat Yield. *Plant Physiol*. 2015;169(3):1991-2005.
- Higo K, Ugawa Y, Iwamoto M, Korenaga T. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research*. 1999;27(1):297-300.
- Hinton JJC. Resistance of the testa to entry of water into the wheat kernel. *Cereal Chemistry*. 1955;32:296–306.
- Höfgen R, Willmitzer L. Storage of competent cells for *Agrobacterium* transformation. *Nucleic Acids Res*. 1988;16(20):9877.

- Hong Y, Zhang H, Huang L, Li D, Song F. Overexpression of a Stress-Responsive NAC Transcription Factor Gene ONAC022 Improves Drought and Salt Tolerance in Rice. *Front Plant Sci.* 2016;7:4.
- Howe EA, Sinha R, Schlauch D, Quackenbush J. RNA-Seq analysis in MeV. *Bioinformatics.* 2011;27(22):3209-10.
- Hu H, Dai M, Yao J, Xiao B, Li X, Zhang Q, et al. Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *Proceedings of the National Academy of Sciences.* 2006;103(35):12987-92.
- Huang L, Hong Y, Zhang H, Li D, Song F. Rice NAC transcription factor ONAC095 plays opposite roles in drought and cold stress tolerance. *BMC Plant Biology.* 2016;16(1).
- Huang Q, Wang Y. Overexpression of TaNAC2D Displays Opposite Responses to Abiotic Stresses between Seedling and Mature Stage of Transgenic Arabidopsis. *Front Plant Sci.* 2016;7.
- Huang Q, Wang Y, Li B, Chang J, Chen M, Li K, et al. TaNAC29, a NAC transcription factor from wheat, enhances salt and drought tolerance in transgenic Arabidopsis. *BMC Plant Biol.* 2015;15:268.
- Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, et al. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat. *Proc Natl Acad Sci USA.* 2002;99(12):8133-8.
- Huo N, Zhang S, Zhu T, Dong L, Wang Y, Mohr T, et al. Gene Duplication and Evolution Dynamics in the Homeologous Regions Harboring Multiple Prolamin and Resistance Gene Families in Hexaploid Wheat. *Front Plant Sci.* 2018;9.
- Hurkman WJ, McCue KF, Altenbach SB, Korn A, Tanaka CK, Kothari KM, et al. Effect of temperature on expression of genes encoding enzymes for starch biosynthesis in developing wheat endosperm. *Plant Science.* 2003;164(5):873-81.
- Hurkman WJ, Vensel WH, Tanaka CK, Whitehand L, Altenbach SB. Effect of high temperature on albumin and globulin accumulation in the endosperm proteome of the developing wheat grain. *Journal of Cereal Science.* 2009;49(1):12-23.
- Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends in Genetics.* 2002;18(9):486-7.
- Hussain RM, Ali M, Feng X, Li X. The essence of NAC gene family to the cultivation of drought-resistant soybean (*Glycine max* L. Merr.) cultivars. *BMC Plant Biology.* 2017;17(1).
- Ihsan MZ, El-Nakhlawy FS, Ismail SM, Fahad S, Daur I. Wheat Phenological Development and Growth Studies As Affected by Drought and Late Season High Temperature Stress under Arid Environment. *Front Plant Sci.* 2016;7:795.
- Ikeda M, Mitsuda N, Ohme-Takagi M. Arabidopsis HsfB1 and HsfB2b Act as Repressors of the Expression of Heat-Inducible Hsfs But Positively Regulate the Acquired Thermotolerance. *PLANT PHYSIOLOGY.* 2011;157(3):1243-54.
- Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, et al. bZIP transcription factors in Arabidopsis. *Trends Plant Sci.* 2002;7(3):106-11.
- Jensen MK, Hagedorn PH, de Torres-Zabala M, Grant MR, Rung JH, Collinge DB, et al. Transcriptional regulation by an NAC (NAM-ATAF1,2-CUC2) transcription factor

- attenuates ABA signalling for efficient basal defence towards *Blumeria graminis* f. sp. *hordei* in *Arabidopsis*. *Plant J.* 2008;56(6):867-80.
- Jensen MK, Kjaersgaard T, Nielsen MM, Galberg P, Petersen K, O'Shea C, et al. The *Arabidopsis thaliana* NAC transcription factor family: structure–function relationships and determinants of ANAC019 stress signalling. *Biochemical Journal.* 2010;426(2):183-96.
- Jeong JS, Kim YS, Baek KH, Jung H, Ha SH, Do Choi Y, et al. Root-Specific Expression of OsNAC10 Improves Drought Tolerance and Grain Yield in Rice under Field Drought Conditions. *PLANT PHYSIOLOGY.* 2010;153(1):185-97.
- Ji X, Shiran B, Wan J, Lewis DC, Jenkins CLD, Condon AG, et al. Importance of pre-anthesis anther sink strength for maintenance of grain number during reproductive stage water stress in wheat: Reproductive stage drought tolerance in wheat. *Plant, Cell & Environment.* 2010;33(6):926-42.
- Jia D, Gong X, Li M, Li C, Sun T, Ma F. Overexpression of a Novel Apple NAC Transcription Factor Gene, MdNAC1, Confers the Dwarf Phenotype in Transgenic Apple (*Malus domestica*). *Genes.* 2018;9(5):229.
- Jin C, Li K-Q, Xu X-Y, Zhang H-P, Chen H-X, Chen Y-H, et al. A Novel NAC Transcription Factor, PbeNAC1, of *Pyrus betulifolia* Confers Cold and Drought Tolerance via Interacting with PbeDREBs and Activating the Expression of Stress-Responsive Genes. *Frontiers in Plant Science.* 2017a;8.
- Jin X, Ren J, Nevo E, Yin X, Sun D, Peng J. Divergent Evolutionary Patterns of NAC Transcription Factors Are Associated with Diversification and Gene Duplications in Angiosperm. *Frontiers in Plant Science.* 2017b;8.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30(9):1236-40.
- Jones RA, Qualset CO. Breeding Crops for Environmental Stress Tolerance. In: Collins GB, Petolino JG, éditeurs. *Applications of Genetic Engineering to Crop Improvement.* Dordrecht: Springer Netherlands; 1984. p. 305-40. (Advances in Agricultural Biotechnology).
- Karanja BK, Xu L, Wang Y, Muleke EM, Jabir BM, Xie Y, et al. Genome-wide characterization and expression profiling of *NAC* transcription factor genes under abiotic stresses in radish (*Raphanus sativus* L.). *PeerJ.* 2017;5:e4172.
- Kataoka R, Takahashi M, Suzuki N. Coordination between bZIP28 and HSFA2 in the regulation of heat response signals in *Arabidopsis*. *Plant Signaling & Behavior.* 2017;12(11):e1376159.
- Kawakatsu T, Takaiwa F. Cereal seed storage protein synthesis: fundamental processes for recombinant protein production in cereal grains. *Plant Biotechnol J.* 2010;8(9):939-53.
- Khedra J, Agarwal P, Agarwal PK. AINAC4 Transcription Factor From Halophyte *Aeluropus lagopoides* Mitigates Oxidative Stress by Maintaining ROS Homeostasis in Transgenic Tobacco. *Frontiers in Plant Science.* 2018;9.
- Kikuchi K, Ueguchi-Tanaka M, Yoshida KT, Nagato Y, Matsusoka M, Hirano H-Y. Molecular analysis of the NAC gene family in rice. *MGG - Molecular and General Genetics.* 2000;262(6):1047-51.

- Kilasi NL, Singh J, Vallejos CE, Ye C, Jagadish SVK, Kusolwa P, et al. Heat Stress Tolerance in Rice (*Oryza sativa* L.): Identification of Quantitative Trait Loci and Candidate Genes for Seedling Growth Under Heat Stress. *Frontiers in Plant Science*. 2018;9.
- Kim HJ, Hong SH, Kim YW, Lee IH, Jun JH, Phee B-K, et al. Gene regulatory cascade of senescence-associated NAC transcription factors activated by ETHYLENE-INSENSITIVE2-mediated leaf senescence signalling in *Arabidopsis*. *Journal of Experimental Botany*. 2014;65(14):4023-36.
- Kim JH, Woo HR, Kim J, Lim PO, Lee IC, Choi SH, et al. Trifurcate Feed-Forward Regulation of Age-Dependent Cell Death Involving miR164 in *Arabidopsis*. *Science*. 2009;323(5917):1053-7.
- Kim S-G, Kim S-Y, Park C-M. A membrane-associated NAC transcription factor regulates salt-responsive flowering via FLOWERING LOCUS T in *Arabidopsis*. *Planta*. 2007a;226(3):647-54.
- Kim S-G, Lee A-K, Yoon H-K, Park C-M. A membrane-bound NAC transcription factor NTL8 regulates gibberellic acid-mediated salt signaling in *Arabidopsis* seed germination. *The Plant Journal*. 2008;55(1):77-88.
- Kim S-G, Lee S, Seo PJ, Kim S-K, Kim J-K, Park C-M. Genome-scale screening and molecular characterization of membrane-bound transcription factors in *Arabidopsis* and rice. *Genomics*. 2010;95(1):56-65.
- Kim S-Y, Kim S-G, Kim Y-S, Seo PJ, Bae M, Yoon H-K, et al. Exploring membrane-associated NAC transcription factors in *Arabidopsis*: implications for membrane biology in genome regulation. *Nucleic Acids Research*. 2007b;35(1):203-13.
- Kim Y-S, Kim S-G, Park J-E, Park H-Y, Lim M-H, Chua N-H, et al. A membrane-bound NAC transcription factor regulates cell division in *Arabidopsis*. *Plant Cell*. 2006;18(11):3132-44.
- Kobata T, Palta JA, Turner NC. Rate of Development of Postanthesis Water Deficits and Grain Filling of Spring Wheat. *Crop Science*. 1992;32(5):1238-42.
- Kocsy G, Tari I, Vanková R, Zechmann B, Gulyás Z, Poór P, et al. Redox control of plant growth and development. *Plant Sci*. 2013;211:77-91.
- Konishi M, Yanagisawa S. *Arabidopsis* NIN-like transcription factors have a central role in nitrate signalling. *Nat Commun*. 2013;4:1617.
- Kotak S, Larkindale J, Lee U, von Koskull-Döring P, Vierling E, Scharf K-D. Complexity of the heat stress response in plants. *Curr Opin Plant Biol*. 2007;10(3):310-6.
- Kottmann L, Wilde P, Schittenhelm S. How do timing, duration, and intensity of drought stress affect the agronomic performance of winter rye? *European Journal of Agronomy*. 2016;75:25-32.
- Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, Clissold L, et al. Uncovering hidden variation in polyploid wheat. *PNAS*. 2017;114(6):E913-21.
- Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 2016;33(7):1870-4.
- Langridge PP, Bramel DP, Eversole DK, Rogers J, Keller PB, Appels R, et al. Achieving sustainable cultivation of wheat Volume 1: Breeding, quality traits, pests and diseases. Vol. 1. 2017.

- Lata C, Prasad M. Role of DREBs in regulation of abiotic stress responses in plants. *J Exp Bot*. 2011;62(14):4731-48.
- Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*. 2014;15(2):R29.
- Leach LJ, Belfield EJ, Jiang C, Brown C, Mithani A, Harberd NP. Patterns of homoeologous gene expression shown by RNA sequencing in hexaploid bread wheat. *BMC Genomics*. 2014;15:276.
- Lee M-H, Jeon HS, Kim HG, Park OK. An Arabidopsis NAC transcription factor NAC4 promotes pathogen-induced cell death under negative regulation by microRNA164. *New Phytologist*. 2017;214(1):343-60.
- Lee S, Lee H-J, Huh SU, Paek K-H, Ha J-H, Park C-M. The Arabidopsis NAC transcription factor NTL4 participates in a positive feedback loop that induces programmed cell death under heat stress conditions. *Plant Science*. 2014;227:76-83.
- Lee S, Seo PJ, Lee H-J, Park C-M. A NAC transcription factor NTL4 promotes reactive oxygen species production during drought-induced leaf senescence in Arabidopsis: NTL4 in drought-induced leaf senescence. *The Plant Journal*. 2012;70(5):831-44.
- Leucci MR, Lenucci MS, Piro G, Dalessandro G. Water stress and cell wall polysaccharides in the apical root zone of wheat cultivars varying in drought tolerance. *Journal of Plant Physiology*. 2008;165(11):1168-80.
- Li Q, Lin Y-C, Sun Y-H, Song J, Chen H, Zhang X-H, et al. Splice variant of the SND1 transcription factor is a dominant negative of SND1 members and their regulation in *Populus trichocarpa*. *Proc Natl Acad Sci U S A*. 2012;109(36):14699-704.
- Li X, Feng B, Zhang F, Tang Y, Zhang L, Ma L, et al. Bioinformatic Analyses of Subgroup-A Members of the Wheat bZIP Transcription Factor Family and Functional Identification of TabZIP174 Involved in Drought Stress Response. *Frontiers in Plant Science*. 2016a;7.
- Li X-D, Zhuang K-Y, Liu Z-M, Yang D-Y, Ma N-N, Meng Q-W. Overexpression of a novel NAC-type tomato transcription factor, SINAM1, enhances the chilling stress tolerance of transgenic tobacco. *J Plant Physiol*. 2016b;204:54-65.
- Liang C, Wang Y, Zhu Y, Tang J, Hu B, Liu L, et al. OsNAP connects abscisic acid and leaf senescence by fine-tuning abscisic acid biosynthesis and directly targeting senescence-associated genes in rice. *Proceedings of the National Academy of Sciences*. 2014;111(27):10013-8.
- Liang M, Li H, Zhou F, Li H, Liu J, Hao Y, et al. Subcellular Distribution of NTL Transcription Factors in *Arabidopsis thaliana*: Subcellular Distribution of AtNTLs. *Traffic*. 2015;16(10):1062-74.
- Liu H-C, Liao H-T, Charng Y-Y. The role of class A1 heat shock factors (HSFA1s) in response to heat and other stresses in Arabidopsis. *Plant, Cell & Environment*. 2011;34(5):738-51.
- Liu M, Ma Z, Sun W, Huang L, Wu Q, Tang Z, et al. Genome-wide analysis of the NAC transcription factor family in Tartary buckwheat (*Fagopyrum tataricum*). *BMC Genomics*. 2019;20(1).

- Liu X, wang T, Bartholomew E, Black K, Dong M, Zhang Y, et al. Comprehensive analysis of NAC transcription factors and their expression during fruit spine development in cucumber (*Cucumis sativus* L.). *Horticulture Research*. 2018;5(1).
- Liu Y, Sun J, Wu Y. Arabidopsis ATAF1 enhances the tolerance to salt stress and ABA in transgenic rice. *Journal of Plant Research*. 2016;129(5):955-62.
- Liu Z, Xin M, Qin J, Peng H, Ni Z, Yao Y, et al. Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.). *BMC Plant Biology*. 2015;15(1).
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001;25(4):402-8.
- Lobell DB, Schlenker W, Costa-Roberts J. Climate trends and global crop production since 1980. *Science*. 2011;333(6042):616-20.
- Lu H, Wang C, Guo T, Xie Y, Feng W, Li S. Starch composition and its granules distribution in wheat grains in relation to post-anthesis high temperature and drought stress treatments. *Starch - Stärke*. 2014;66(5-6):419-28.
- Lu M, Sun Q-P, Zhang D-F, Wang T-Y, Pan J-B. Identification of 7 stress-related NAC transcription factor members in maize (*Zea mays* L.) and characterization of the expression pattern of these genes. *Biochem Biophys Res Commun*. 2015;462(2):144-50.
- Lu P-L, Chen N-Z, An R, Su Z, Qi B-S, Ren F, et al. A novel drought-inducible gene, ATAF1, encodes a NAC family protein that negatively regulates the expression of stress-responsive genes in Arabidopsis. *Plant Mol Biol*. 2007;63(2):289-305.
- Lundström M, Leino MW, Hagenblad J. Evolutionary history of the NAM-B1 gene in wild and domesticated tetraploid wheat. *BMC Genetics*. 2017;18(1).
- Luo M, Liu X, Singh P, Cui Y, Zimmerli L, Wu K. Chromatin modifications and remodeling in plant abiotic stress responses. *Biochim Biophys Acta*. 1819;2012:129–136.
- Maestri E, Klueva N, Perrotta C, Gulli M, Nguyen HT, Marmioli N. Molecular genetics of heat tolerance and heat shock proteins in cereals. *Plant Mol Biol*. 2002;48(5):667-81.
- Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. Gene duplication as a major force in evolution. *Journal of Genetics*. 2013;92(1):155-61.
- Mao H, Yu L, Han R, Li Z, Liu H. ZmNAC55, a maize stress-responsive NAC transcription factor, confers drought resistance in transgenic Arabidopsis. *Plant Physiology and Biochemistry*. 2016;105:55-66.
- Mao X, Chen S, Li A, Zhai C, Jing R. Novel NAC Transcription Factor TaNAC67 Confers Enhanced Multi-Abiotic Stress Tolerances in Arabidopsis. Unver T, éditeur. *PLoS ONE*. 2014;9(1):e84359.
- Mao X, Zhang H, Qian X, Li A, Zhao G, Jing R. TaNAC2, a NAC-type wheat transcription factor conferring enhanced multiple abiotic stress tolerances in Arabidopsis. *Journal of Experimental Botany*. 2012;63(8):2933-46.
- Mao X, Zhang H, Tian S, Chang X, Jing R. TaSnRK2.4, an SNF1-type serine/threonine protein kinase of wheat (*Triticum aestivum* L.), confers enhanced multistress tolerance in Arabidopsis. *J Exp Bot*. 2010;61(3):683-96.

- Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, The International Wheat Genome Sequencing Consortium, et al. Ancient hybridizations among the ancestral genomes of bread wheat. *Science*. 2014;345(6194):1250092.
- Mares DJ, Norstog K, Stone BA. Early Stages in the Development of Wheat Endosperm. I. The Change From Free Nuclear to Cellular Endosperm. *Aust J Bot*. 1975;23(2):311-26.
- Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D, et al. Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J*. 2013;76(3):494-505.
- Mathew IE, Agarwal P. May the Fittest Protein Evolve: Favoring the Plant-Specific Origin and Expansion of NAC Transcription Factors. *BioEssays*. 2018;40(8):1800018.
- Mathew IE, Das S, Mahto A, Agarwal P. Three Rice NAC Transcription Factors Heteromerize and Are Associated with Seed Size. *Frontiers in Plant Science*. 2016;7.
- Maugarny-Calès A, Gonçalves B, Jouannic S, Melkonian M, Ka-Shu Wong G, Laufs P. Apparition of the NAC Transcription Factors Predates the Emergence of Land Plants. *Molecular Plant*. 2016;9(9):1345-8.
- Mishra SK, Tripp J, Winkelhaus S, Tschiersch B, Theres K, Nover L, et al. In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermotolerance in tomato. *Genes Dev*. 2002;16(12):1555-67.
- Morishita T, Kojima Y, Maruta T, Nishizawa-Yokoi A, Yabuta Y, Shigeoka S. Arabidopsis NAC Transcription Factor, ANAC078, Regulates Flavonoid Biosynthesis under High-light. *Plant and Cell Physiology*. 2009;50(12):2210-22.
- Morris P, Meleard B. Production de blé tendre : des protéines pour assurer les débouchés. Perspectives agricoles n°418. Arvalis-Institut du végétal. 2015; 52-55.
- Murozuka E, Massange-Sánchez JA, Nielsen K, Gregersen PL, Braumann I. Genome wide characterization of barley NAC transcription factors enables the identification of grain-specific transcription factors exclusive for the Poaceae family of monocotyledonous plants. Kashkush K, éditeur. *PLOS ONE*. 2018;13(12):e0209769.
- Muthusamy SK, Dalal M, Chinnusamy V, Bansal KC. Genome-wide identification and analysis of biotic and abiotic stress regulation of small heat shock protein (HSP20) family genes in bread wheat. *Journal of Plant Physiology*. 2017;211:100-13.
- Nadaud I, Girousse C, Debiton C, Chambon C, Bouzidi MF, Martre P, et al. Proteomic and morphological analysis of early stages of wheat grain development. *PROTEOMICS*. 2010;10(16):2901-10.
- Naeem HA, Paulon D, Irmak S, MacRitchie F. Developmental and environmental effects on the assembly of glutenin polymers and the impact on grain quality of wheat. *Journal of Cereal Science*. 2012;56(1):51-7.
- Nakashima K, Tran L-SP, Van Nguyen D, Fujita M, Maruyama K, Todaka D, et al. Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice: Rice OsNAC6 functions in stress responses. *The Plant Journal*. 2007;51(4):617-30.
- Ney B, Bancal MO, Bancal P, Bingham IJ, Foulkes J, Gouache D, et al. Crop architecture and crop tolerance to fungal diseases and insect herbivory. Mechanisms to limit crop losses. *European Journal of Plant Pathology*. 2013;135(3):561-80.

- Nguyen HN, Sabelli PA, Larkins BA. Endoreduplication and Programmed Cell Death in the Cereal Endosperm. In: Olsen O-A, éditeur. Endosperm: Developmental and Molecular Biology. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 21-43. (Plant Cell Monographs).
- Nicolas M, Gleadow R, Dalling M. Effects of Drought and High Temperature on Grain Growth in Wheat. *Functional Plant Biology*. 1984;11(6):553.
- Ning P, Liu C, Kang J, Lv J. Genome-wide analysis of WRKY transcription factors in wheat (*Triticum aestivum* L.) and differential expression under water deficit condition. *PeerJ*. 2017;5:e3232.
- Noé L, Kucherov G. YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res*. 2005;33(Web Server issue):W540-543.
- Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H, et al. Genome-wide analysis of NAC transcription factor family in rice. *Gene*. 2010;465(1-2):30-44.
- Nuruzzaman M, Sharoni AM, Kikuchi S. Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. *Frontiers in Microbiology*. 2013;4.
- Nuruzzaman M, Sharoni AM, Satoh K, Karim MR, Harikrishna JA, Shimizu T, et al. NAC transcription factor family genes are differentially expressed in rice during infections with Rice dwarf virus, Rice black-streaked dwarf virus, Rice grassy stunt virus, Rice ragged stunt virus, and Rice transitory yellowing virus. *Frontiers in Plant Science*. 2015;6:676.
- Ober ES, Setter TL, Madison JT, Thompson JF, Shapiro PS. Enzyme Activities and RNA Transcripts of Starch and Zein Synthesis, Absciscic Acid, and Cell Division. 1991;97:11.
- Oda-Yamamizo C, Mitsuda N, Sakamoto S, Ogawa D, Ohme-Takagi M, Ohmiya A. The NAC transcription factor ANAC046 is a positive regulator of chlorophyll degradation and senescence in Arabidopsis leaves. *Scientific Reports*. 2016;6(1).
- Ohama N, Sato H, Shinozaki K, Yamaguchi-Shinozaki K. Transcriptional Regulatory Network of Plant Heat Stress Response. *Trends in Plant Science*. 2017;22(1):53-65.
- Ohno S. Evolution by Gene Duplication. Springer; 1970.
- Okonechnikov K, Golosova O, Fursov M, UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012;28(8):1166-7.
- Olesen JE, Bindi M. Consequences of climate change for European agricultural productivity, land use and policy. *European Journal of Agronomy*. 2002;16(4):239-62.
- Olsen AN, Ernst HA, Leggio LL, Skriver K. NAC transcription factors: structurally distinct, functionally diverse. *Trends in Plant Science*. 2005;10(2):79-87.
- Ooka H, Satoh K, Doi K, Nagata T, Otomo Y, Murakami K, et al. Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. *DNA Res*. 2003;10(6):239-47.
- Osborne T. The vegetables proteins. 2nd éd. London, England: Longmans, Green and Co Edition; 1924.
- Pachauri RK, Mayer L, Intergovernmental Panel on Climate Change. Climate change 2014: synthesis report. Geneva, Switzerland: Intergovernmental Panel on Climate Change; 2015.

- Paolacci AR, Tanzarella OA, Porceddu E, Ciaffi M. Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. *BMC Mol Biol.* 2009;10:11.
- Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, et al. A Physical Map of the 1-Gigabase Bread Wheat Chromosome 3B. *Science.* 2008;322(5898):101-4.
- Peet MM, Willits DH, Gardner R. Response of ovule development and post-pollen production processes in male-sterile tomatoes to chronic, sub-acute high temperature stress. *Journal of Experimental Botany.* 1997;48(1):101-11.
- Pei H, Ma N, Tian J, Luo J, Chen J, Li J, et al. An NAC transcription factor controls ethylene-regulated cell expansion in flower petals. *Plant Physiol.* 2013;163(2):775-91.
- Pellegrineschi A, Noguera LM, Skovmand B, Brito RM, Velazquez L, Salgado MM, et al. Identification of highly transformable wheat genotypes for mass production of fertile transgenic plants. *Genome.* 2002;45(2):421-30.
- Peng X, Zhao Y, Li X, Wu M, Chai W, Sheng L, et al. Genomewide identification, classification and analysis of NAC type gene family in maize. *Journal of Genetics.* 2015;94(3):377-90.
- Pieterse CMJ, Van der Does D, Zamioudis C, Leon-Reyes A, Van Wees SCM. Hormonal modulation of plant immunity. *Annu Rev Cell Dev Biol.* 2012;28:489-521.
- Pinheiro G, Marques C, Costa M, A.B. Reis P, Alves M, Carvalho C, et al. Complete inventory of soybean NAC transcription factors: Sequence conservation and expression analysis uncover their distinct roles in stress response. *Gene.* 2009;444:10-23.
- Pinto RS, Reynolds MP, Mathews KL, McIntyre CL, Olivares-Villegas J-J, Chapman SC. Heat and drought adaptive QTL in a wheat population designed to minimize confounding agronomic effects. *Theoretical and Applied Genetics.* 2010;121(6):1001-21.
- Plessis A, Ravel C, Bordes J, Balfourier F, Martre P. Association study of wheat grain protein composition reveals that gliadin and glutenin composition are trans-regulated by different chromosome regions. *Journal of Experimental Botany.* 2013;64(12):3627-44.
- Pont C, Murat F, Guizard S, Flores R, Foucrier S, Bidet Y, et al. Wheat syntenome unveils new evidences of contrasted evolutionary plasticity between paleo- and neoduplicated subgenomes. *The Plant Journal.* 2013;76(6):1030-44.
- Porter KB, Worrall WD, Gardenhire JH, Gilmore EC, McDaniel ME, Tuleen NA. Registration of 'Tam 107' Wheat. *Crop Science.* 1987;27(4):818-9.
- Qin D, Wu H, Peng H, Yao Y, Ni Z, Li Z, et al. Heat stress-responsive transcriptome analysis in heat susceptible and tolerant wheat (*Triticum aestivum* L.) by using Wheat Genome Array. *BMC Genomics.* 2008a;9(1):432.
- Qin F, Sakuma Y, Tran L-SP, Maruyama K, Kidokoro S, Fujita Y, et al. Arabidopsis DREB2A-Interacting Proteins Function as RING E3 Ligases and Negatively Regulate Plant Drought Stress-Responsive Gene Expression. *THE PLANT CELL ONLINE.* 2008b;20(6):1693-707.
- Qu A-L, Ding Y-F, Jiang Q, Zhu C. Molecular mechanisms of the plant heat stress response. *Biochemical and Biophysical Research Communications.* 2013;432(2):203-7.
- Rasmussen S, Barah P, Suarez-Rodriguez MC, Bressendorff S, Friis P, Costantino P, et al. Transcriptome Responses to Combinations of Stresses in Arabidopsis. *Plant Physiol.* 2013;161(4):1783-94.

- Reguera M, Peleg Z, Blumwald E. Targeting metabolic pathways for genetic engineering abiotic stress-tolerance in crops. *Biochim Biophys Acta*. 2012;1819(2):186-94.
- Rey E, Abrouk M, Keeble-Gagnère G, Karafiátová M, Vrána J, Balzergue S, et al. Transcriptome reprogramming due to the introduction of a barley telosome into bread wheat affects more barley genes than wheat. *Plant Biotechnology Journal*. 2018;16(10):1767-77.
- Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, et al. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science*. 2000;290(5499):2105-10.
- Riechmann JL, Meyerowitz EM. The AP2/EREBP family of plant transcription factors. *Biol Chem*. 1998;379(6):633-46.
- Rimbert H, Darrier B, Navarro J, Kitt J, Choulet F, Leveugle M, et al. High throughput SNP discovery and genotyping in hexaploid wheat. Zhang A, éditeur. *PLOS ONE*. 2018;13(1):e0186329.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
- Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*. 2010;11(3):R25.
- Rogers SO, Quatrano RS. Morphological Staging of Wheat Caryopsis Development. *American Journal of Botany*. 1983;70(2):308-11.
- Rushton PJ, Bokowiec MT, Han S, Zhang H, Brannock JF, Chen X, et al. Tobacco Transcription Factors: Novel Insights into Transcriptional Regulation in the Solanaceae. *PLANT PHYSIOLOGY*. 2008;147(1):280-95.
- Saad ASI, Li X, Li H-P, Huang T, Gao C-S, Guo M-W, et al. A rice stress-responsive NAC gene enhances tolerance of transgenic wheat to drought and salt stresses. *Plant Science*. 2013;203-204:33-40.
- Sabelli PA. Replicate and die for your own good: Endoreduplication and cell death in the cereal endosperm. *Journal of Cereal Science*. 2012;56(1):9-20.
- Sabelli PA, Larkins BA. The development of endosperm in grasses. *Plant Physiol*. 2009;149(1):14-26.
- Sablowski RW, Meyerowitz EM. A homolog of NO APICAL MERISTEM is an immediate target of the floral homeotic genes APETALA3/PISTILLATA. *Cell*. 1998;92(1):93-103.
- Saidi MN, Mergby D, Brini F. Identification and expression analysis of the NAC transcription factor family in durum wheat (*Triticum turgidum* L. ssp. durum). *Plant Physiology and Biochemistry*. 2017;112:117-28.
- Saini HS, Sedgley M, Aspinall D. Development Anatomy in Wheat of Male Sterility Induced by Heat Stress, Water Deficit or Absciscic Acid. *Functional Plant Biol*. 1984;11(4):243-53.
- Sakuraba Y, Kim Y-S, Han S-H, Lee B-D, Paek N-C. The Arabidopsis Transcription Factor NAC016 Promotes Drought Stress Responses by Repressing *AREB1* Transcription through a Trifurcate Feed-Forward Regulatory Loop Involving NAP. *The Plant Cell*. 2015;27(6):1771-87.

- Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, et al. Identification and Characterization of Shared Duplications between Rice and Wheat Provide New Insight into Grass Genome Evolution. *THE PLANT CELL ONLINE*. 2008;20(1):11-24.
- Sambrook J, Russell DW, Laboratory CSH. *Molecular cloning : a laboratory manual*. 3rd. ed. Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory; 2001.
- Sánchez-Montesino R, Bouza-Morcillo L, Marquez J, Ghita M, Duran-Nebreda S, Gómez L, et al. A Regulatory Module Controlling GA-Mediated Endosperm Cell Expansion Is Critical for Seed Germination in Arabidopsis. *Molecular Plant*. 2019;12(1):71-85.
- Sanjari Pireivatlou A, Yazdansepa A. Evaluation of Wheat (*Triticum aestivum* L.) Genotypes under Pre- and Post-anthesis Drought Stress Conditions. *Journal of Agricultural Science and Technology*. 2010;10(0):109-21.
- Sato H, Mizoi J, Tanaka H, Maruyama K, Qin F, Osakabe Y, et al. Arabidopsis DPB3-1, a DREB2A Interactor, Specifically Enhances Heat Stress-Induced Gene Expression by Forming a Heat Stress-Specific Transcriptional Complex with NF-Y Subunits[C][W]. *Plant Cell*. 2014;26(12):4954-73.
- Scheible W-R, Morcuende R, Czechowski T, Fritz C, Osuna D, Palacios-Rojas N, et al. Genome-Wide Reprogramming of Primary and Secondary Metabolism, Protein Synthesis, Cellular Growth Processes, and the Regulatory Infrastructure of Arabidopsis in Response to Nitrogen. *Plant Physiol*. 2004;136(1):2483-99.
- Schnyder H, Baum U. Growth of the grain of wheat (*Triticum aestivum* L.). The relationship between water content and dry matter accumulation. *European Journal of Agronomy*. 1992;1(2):51-7.
- Semenov MA, Stratonovitch P, Alghabari F, Gooding MJ. Adapting wheat in Europe for climate change. *Journal of Cereal Science*. 2014;59(3):245-56.
- Shah NH, Paulsen GM. Interaction of drought and high temperature on photosynthesis and grain-filling of wheat. *Plant and Soil*. 2003;257(1):219-26.
- Shahnejat-Bushehri S, Mueller-Roeber B, Balazadeh S. Arabidopsis NAC transcription factor JUNGBRUNNEN1 affects thermomemory-associated genes and enhances heat stress tolerance in primed and unprimed conditions. *Plant Signaling & Behavior*. 2012;7(12):1518-21.
- Shangguan Z, Shao M, Dyckmans J. Interaction of Osmotic Adjustment and Photosynthesis in Winter Wheat Under Soil Drought. *Journal of Plant Physiology*. 1999;154(5-6):753-8.
- Shen H, Yin Y, Chen F, Xu Y, Dixon RA. A Bioinformatic Analysis of NAC Genes for Plant Cell Wall Development in Relation to Lignocellulosic Bioenergy Production. *BioEnergy Research*. 2009;2(4):217-32.
- Shewry PR. Wheat. *J Exp Bot*. 2009;60(6):1537-53.
- Shewry PR, Hey SJ. The contribution of wheat to human diet and health. *Food and Energy Security*. 2015;4(3):178-202.
- Shewry PR, Mitchell RAC, Tosi P, Wan Y, Underwood C, Lovegrove A, et al. An integrated study of grain development of wheat (cv. Hereward). *Journal of Cereal Science*. 2012;56(1):21-30.
- Shewry PR, Popineau Y, Lafiandra D, Belton P. Wheat glutenin subunits and dough elasticity: findings of the EUROWHEAT project. *Trends in Food Science & Technology*. 2000;11(12):433-41.

- Shewry PR, Tatham AS, Forde J, Kreis M, Mifflin BJ. The classification and nomenclature of wheat gluten proteins: A reassessment. *Journal of Cereal Science*. 1986;4(2):97-106.
- Shinozaki K, Yamaguchi-Shinozaki K. Gene Expression and Signal Transduction in Water-Stress Response. *Plant Physiology*. 1997;115(2):327-34.
- Shrivastava P, Kumar R. Soil salinity: A serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation. *Saudi Journal of Biological Sciences*. 2015;22(2):123-31.
- Slafer GA, Savin R, Sadras VO. Coarse and fine regulation of wheat yield components in response to genotype and environment. *Field Crops Research*. 2014;157:71-83.
- Smith NA, Singh SP, Wang M-B, Stoutjesdijk PA, Green AG, Waterhouse PM. Gene expression: Total silencing by intron-spliced hairpin RNAs. *Nature*. 2000;407(6802):319-20.
- Sofield I, Evans LT, Cook MG, Wardlaw IF. Factors Influencing the Rate and Duration of Grain Filling in Wheat. *Functional Plant Biol*. 1977;4(5):785-97.
- Soltner D. Les grandes productions végétales. 1988. (Collection sciences et techniques agricoles).
- Sosulski FW, Imafidon GI. Amino acid composition and nitrogen-to-protein conversion factors for animal and plant foods. *Journal of Agricultural and Food Chemistry*. 1990;38(6):1351-6.
- Souer E, van Houwelingen A, Kloos D, Mol J, Koes R. The No Apical Meristem Gene of *Petunia* Is Required for Pattern Formation in Embryos and Flowers and Is Expressed at Meristem and Primordia Boundaries. *Cell*. 1996;85(2):159-70.
- Sreenivasulu N, Schnurbusch T. A genetic playground for enhancing grain number in cereals. *Trends in Plant Science*. 2012;17(2):91-101.
- Srivastava R, Deng Y, Howell SH. Stress sensing in plants by an ER stress sensor/transducer, bZIP28. *Frontiers in Plant Science*. 2014;5.
- Stael S, Kmiciek P, Willems P, Van Der Kelen K, Coll NS, Teige M, et al. Plant innate immunity – sunny side up? *Trends Plant Sci*. 2015;20(1):3-11.
- Subrahmanyam D, Subash N, Haris A, Sikka AK. Influence of water stress on leaf photosynthetic characteristics in wheat cultivars differing in their susceptibility to drought. *Photosynthetica*. 2006;44(1):125-9.
- Sun L, Huang L, Hong Y, Zhang H, Song F, Li D. Comprehensive Analysis Suggests Overlapping Expression of Rice ONAC Transcription Factors in Abiotic and Biotic Stress Responses. *International Journal of Molecular Sciences*. 2015;16(2):4306-26.
- Sun L, Zhang H, Li D, Huang L, Hong Y, Ding XS, et al. Functions of rice NAC transcriptional factors, ONAC122 and ONAC131, in defense responses against *Magnaporthe grisea*. *Plant Molecular Biology*. 2013;81(1-2):41-56.
- Surget A, Barron C. Histologie du grain de blé. *Industrie des Céréales*. 2005;145:4-7.
- Sutton F, Chen D-G, Ge X, Kenefick D. Cbf genes of the Fr-A2 allele are differentially regulated between long-term cold acclimated crown tissue of freeze-resistant and – susceptible, winter wheat mutant lines. *BMC Plant Biology*. 2009;9(1):34.
- Suzuki N, Koussevitzky S, Mittler R, Miller G. ROS and redox signalling in the response of plants to abiotic stress. *Plant Cell Environ*. 2012;35(2):259-70.

- Suzuki N, Mittler R. Reactive oxygen species and temperature stresses: A delicate balance between signaling and destruction. *Physiologia Plantarum*. 2006;126(1):45-51.
- Tak H, Negi S, Ganapathi TR. Banana NAC transcription factor MusaNAC042 is positively associated with drought and salinity tolerance. *Protoplasma*. 2017;254(2):803-16.
- Takeoka Y, Hiroi K, Kitano H, Wada T. Pistil hyperplasia in rice spikelets as affected by heat stress. *Sexual Plant Reprod*. 1991;4(1):39-43.
- Tang N, Shahzad Z, Lonjon F, Loudet O, Vailleau F, Maurel C. Natural variation at XND1 impacts root hydraulics and trade-off for stress responses in *Arabidopsis*. *Nature Communications*. 2018;9(1).
- Tang Y, Horikoshi M, Li W. ggfortify: Unified Interface to Visualize Statistical Results of Popular R Packages. 2016;8:12.
- Tang Y, Liu M, Gao S, Zhang Z, Zhao X, Zhao C, et al. Molecular characterization of novel TaNAC genes in wheat and overexpression of TaNAC2a confers drought tolerance in tobacco. *Physiol Plant*. 2012;144(3):210-24.
- Tashiroy T, Wardlaw IF. The Response to High Temperature Shock and Humidity Changes Prior to and During the Early Stages of Grain Development in Wheat. *Functional Plant Biol*. 1990;17(5):551-61.
- Thirumalaikumar VP, Devkar V, Mehterov N, Ali S, Ozgur R, Turkan I, et al. NAC transcription factor JUNGBRUNNEN1 enhances drought tolerance in tomato. *Plant Biotechnology Journal*. 2018;16(2):354-66.
- Thitisaksakul M, Jiménez RC, Arias MC, Beckles DM. Effects of environmental factors on cereal starch biosynthesis and composition. *Journal of Cereal Science*. 2012;56(1):67-80.
- Tian S, Mao X, Zhang H, Chen S, Zhai C, Yang S, et al. Cloning and characterization of TaSnRK2.3, a novel SnRK2 gene in common wheat. *J Exp Bot*. 2013;64(7):2063-80.
- Tosi P, Gritsch CS, He J, Shewry PR. Distribution of gluten proteins in bread wheat (*Triticum aestivum*) grain. *Ann Bot*. 2011;108(1):23-35.
- Tosi P, Parker M, Gritsch CS, Carzaniga R, Martin B, Shewry PR. Trafficking of storage proteins in developing grain of wheat. *J Exp Bot*. 2009;60(3):979-91.
- Tran L-SP, Nakashima K, Sakuma Y, Osakabe Y, Qin F, Simpson SD, et al. Co-expression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the ERD1 gene in *Arabidopsis*. *Plant J*. 2007;49(1):46-63.
- Tran L-SP, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, et al. Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. *Plant Cell*. 2004;16(9):2481-98.
- Travella S, Klimm TE, Keller B. RNA interference-based gene silencing as an efficient tool for functional genomics in hexaploid bread wheat. *Plant Physiol*. 2006;142(1):6-20.
- Triboi E, Abad A, Michelena A, Lloveras J, Ollier J., Daniel C. Environmental effects on the quality of two wheat genotypes: 1. quantitative and qualitative variation of storage proteins. *European Journal of Agronomy*. 2000;13(1):47-64.
- Triboi E, Triboi-Blondel A-M. Productivity and grain or seed composition: a new approach to an old problem—invited paper. *European Journal of Agronomy*. 2002;16(3):163-86.

- Tubiello F, Rosenzweig C, Goldberg R, Jagtap S, Jones J. Effects of climate change on US crop production: simulation results using two different GCM scenarios. Part I: Wheat, potato, maize, and citrus. *Climate Research*. 2002;20:259-70.
- Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J. A NAC Gene Regulating Senescence Improves Grain Protein, Zinc, and Iron Content in Wheat. *Science*. 2006;314(5803):1298-301.
- Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K. Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. *Proc Natl Acad Sci USA*. 2000;97(21):11632-7.
- Valente MAS, Faria JAQA, Soares-Ramos JRL, Reis PAB, Pinheiro GL, Piovesan ND, et al. The ER luminal binding protein (BiP) mediates an increase in drought tolerance in soybean and delays drought-induced leaf senescence in soybean and tobacco. *Journal of Experimental Botany*. 2009;60(2):533-46.
- Vanderauwera S, Suzuki N, Miller G, van de Cotte B, Morsa S, Ravanat J-L, et al. Extranuclear protection of chromosomal DNA from oxidative stress. *Proceedings of the National Academy of Sciences*. 2011;108(4):1711-6.
- Vincent J, Martre P, Ravel C, Baillif A, Agier M. A weboriented platform for gene regulatory network inference. Application to seed storage proteins in wheat. In Institut Pasteur, Paris: Barillot E, Froidevaux C, Rocha E; 2011. p. 293–294.
- Vroemen CW, Mordhorst AP, Albrecht C, Kwaaitaal MACJ, Vries SC de. The CUP-SHAPED COTYLEDON3 Gene Is Required for Boundary and Shoot Meristem Formation in Arabidopsis. *The Plant Cell*. 2003;15(7):1563-77.
- Wahid A, Gelani S, Ashraf M, Foolad M. Heat tolerance in plants: An overview. *Environmental and Experimental Botany*. 2007;61(3):199-223.
- Wan Y, Poole RL, Huttly AK, Toscano-Underwood C, Feeney K, Welham S, et al. Transcriptome analysis of grain development in hexaploid wheat. *BMC Genomics*. 2008;9(1):121.
- Wang B, Wei J, Song N, Wang N, Zhao J, Kang Z. A novel wheat NAC transcription factor, TaNAC30, negatively regulates resistance of wheat to stripe rust: TaNAC30 negatively regulates wheat resistant to Pst. *Journal of Integrative Plant Biology*. 2018a;60(5):432-43.
- Wang C, Jing R, Mao X, Chang X, Li A. TaABC1, a member of the activity of bcl complex protein kinase family from common wheat, confers enhanced tolerance to abiotic stresses in Arabidopsis. *J Exp Bot*. 2011;62(3):1299-311.
- Wang D, Yu Y, Liu Z, Li S, Wang Z, Xiang F. Membrane-bound NAC transcription factors in maize and their contribution to the oxidative stress response. *Plant Science*. 2016a;250:30-9.
- Wang F, Chen H-W, Li Q-T, Wei W, Li W, Zhang W-K, et al. GmWRKY27 interacts with GmMYB174 to reduce expression of *GmNAC29* for stress tolerance in soybean plants. *The Plant Journal*. 2015a;83(2):224-36.
- Wang F, Lin R, Feng J, Chen W, Qiu D, Xu S. TaNAC1 acts as a negative regulator of stripe rust resistance in wheat, enhances susceptibility to *Pseudomonas syringae*, and promotes lateral root development in transgenic Arabidopsis thaliana. *Front Plant Sci*. 2015b;6:108.

- Wang G, Zhang S, Ma X, Wang Y, Kong F, Meng Q. A stress-associated NAC transcription factor (SlNAC35) from tomato plays a positive role in biotic and abiotic stresses. *Physiol Plant*. 2016b;158(1):45-64.
- Wang H, Wang M, Cheng Q. Capturing the Alternative Cleavage and Polyadenylation Sites of 14 NAC Genes in *Populus* Using a Combination of 3'-RACE and High-Throughput Sequencing. *Molecules*. 2018b;23(3):608.
- Wang J, Zhang L, Cao Y, Qi C, Li S, Liu L, et al. CsATAF1 Positively Regulates Drought Stress Tolerance by an ABA-Dependent Pathway and by Promoting ROS Scavenging in Cucumber. *Plant Cell Physiol*. 2018c;59(5):930-45.
- Wang L, Li X, Chen S, Liu G. Enhanced drought tolerance in transgenic *Leymus chinensis* plants with constitutively expressed wheat TaLEA 3. *Biotechnology Letters*. 2009;31(2):313-9.
- Wang W, Yuan Y, Yang C, Geng S, Sun Q, Long L, et al. Characterization, Expression, and Functional Analysis of a Novel NAC Gene Associated with Resistance to Verticillium Wilt and Abiotic Stress in Cotton. *Genes|Genomes|Genetics*. 2016c;6(12):3951-61.
- Wang W, Zheng H, Fan C, Li J, Shi J, Cai Z, et al. High Rate of Chimeric Gene Origination by Retroposition in Plant Genomes. *Plant Cell*. 2006;18(8):1791-802.
- Wang Y, Sun F, Cao H, Peng H, Ni Z, Sun Q, et al. TamiR159 Directed Wheat TaGAMYB Cleavage and Its Involvement in Anther Development and Heat Response. *PLoS One*. 2012;7(11).
- Wardlaw IF, Moncur L. The Response of Wheat to High Temperature Following Anthesis. I. The Rate and Duration of Kernel Filling. *Functional Plant Biol*. 1995;22(3):391-7.
- Welner DH, Lindemose S, Grossmann JG, Møllegaard NE, Olsen AN, Helgstrand C, et al. DNA binding by the plant-specific NAC transcription factors in crystal and solution: a firm link to WRKY and GCM transcription factors. *Biochemical Journal*. 2012;444(3):395-404.
- Wicker T, Buchmann JP, Keller B. Patching gaps in plant genomes results in gene movement and erosion of colinearity. *Genome Res*. 2010;20(9):1229-37.
- Wiegand CL, Cuellar JA. Duration of Grain Filling and Kernel Weight of Wheat as Affected by Temperature 1. *Crop Science*. 1981;21(1):95-101.
- Wright M, Dawson J, Dunder E, Suttie J, Reed J, Kramer C, et al. Efficient biolistic transformation of maize (*Zea mays* L.) and wheat (*Triticum aestivum* L.) using the phosphomannose isomerase gene, *pmi*, as the selectable marker. *Plant Cell Reports*. 2001;20(5):429-36.
- Wu A, Allu AD, Garapati P, Siddiqui H, Dortay H, Zanol M-I, et al. JUNGBRUNNEN1, a Reactive Oxygen Species-Responsive NAC Transcription Factor, Regulates Longevity in Arabidopsis. *THE PLANT CELL ONLINE*. 2012;24(2):482-506.
- Xia N, Zhang G, Liu X-Y, Deng L, Cai G-L, Zhang Y, et al. Characterization of a novel wheat NAC transcription factor gene involved in defense response against stripe rust pathogen infection and abiotic stresses. *Molecular Biology Reports*. 2010a;37(8):3703-12.
- Xia N, Zhang G, Sun Y-F, Zhu L, Xu L-S, Chen X-M, et al. TaNAC8, a novel NAC transcription factor gene in wheat, responds to stripe rust pathogen infection and abiotic stresses. *Physiological and Molecular Plant Pathology*. 2010b;74(5-6):394-402.

- Xiao W, Yang Y, Yu J. ZmNST3 and ZmNST4 are master switches for secondary wall deposition in maize (*Zea mays* L.). *Plant Science*. 2018;266:83-94.
- Xiao Y, Huang X, Shen Y, Huang Z. A novel wheat α -amylase inhibitor gene, TaHPS, significantly improves the salt and drought tolerance of transgenic *Arabidopsis*. *Physiologia Plantarum*. 2013;148(2):273-83.
- Xie Q, Frugis G, Colgan D, Chua N-H. *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. *Genes Dev*. 2000;14(23):3024-36.
- Xie Q, Sanz-Burgos AP, Guo H, García JA, Gutiérrez C. GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein. *Plant Mol Biol*. 1999;39(4):647-56.
- Xu B, Ohtani M, Yamaguchi M, Toyooka K, Wakazaki M, Sato M, et al. Contribution of NAC Transcription Factors to Plant Adaptation to Land. *Science*. 2014;343(6178):1505-8.
- Xu C, Wang M, Zhou L, Quan T, Xia G. Heterologous Expression of the Wheat Aquaporin Gene TaTIP2;2 Compromises the Abiotic Stress Tolerance of *Arabidopsis thaliana*. Ali J, éditeur. *PLoS ONE*. 2013;8(11):e79618.
- Xu G, Zhang J, Lam HM, Wang Z, Yang J. Hormonal changes are related to the poor grain filling in the inferior spikelets of rice cultivated under non-flooded and mulched condition. *Field Crops Research*. 2007;101(1):53-61.
- Xu Z, Gongbuzhaxi, Wang C, Xue F, Zhang H, Ji W. Wheat NAC transcription factor TaNAC29 is involved in response to salt stress. *Plant Physiology and Biochemistry*. 2015;96:356-63.
- Xu Z-S, Liu L, Ni Z-Y, Liu P, Chen M, Li L-C, et al. W55a Encodes a Novel Protein Kinase That Is Involved in Multiple Stress Responses. *Journal of Integrative Plant Biology*. 2009;51(1):58-66.
- Xue G-P, Sadat S, Drenth J, McIntyre CL. The heat shock factor family from *Triticum aestivum* in response to heat and other major abiotic stresses and their role in regulation of heat shock protein genes. *J Exp Bot*. 2014;65(2):539-57.
- Xue G-P, Way HM, Richardson T, Drenth J, Joyce PA, McIntyre CL. Overexpression of TaNAC69 leads to enhanced transcript levels of stress up-regulated genes and dehydration tolerance in bread wheat. *Mol Plant*. 2011;4(4):697-712.
- Yamasaki K, Kigawa T, Inoue M, Watanabe S, Tateno M, Seki M, et al. Structures and evolutionary origins of plant-specific transcription factor DNA-binding domains. *Plant Physiology and Biochemistry*. 2008;46(3):394-401.
- Yang J, Zhang J, Wang Z, Zhu Q, Wang W. Hormonal Changes in the Grains of Rice Subjected to Water Stress during Grain Filling. *Plant Physiol*. 2001;127(1):315-23.
- Yang X, He K, Chi X, Chai G, Wang Y, Jia C, et al. *Miscanthus* NAC transcription factor MINAC12 positively mediates abiotic stress tolerance in transgenic *Arabidopsis*. *Plant Science*. 2018;277:229-41.
- Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, et al. The MYB Transcription Factor Superfamily of *Arabidopsis*: Expression Analysis and Phylogenetic Comparison with the Rice MYB Family. *Plant Mol Biol*. 2006;60(1):107-24.
- Yoshida T, Ohama N, Nakajima J, Kidokoro S, Mizoi J, Nakashima K, et al. *Arabidopsis* HsfA1 transcription factors function as the main positive regulators in heat shock-responsive gene expression. *Molecular Genetics and Genomics*. 2011;286(5-6):321-32.

- Young TE, Gallie DR. Programmed cell death during endosperm development. *Plant Mol Biol*. 2000;44(3):283-301.
- Yu X, Liu Y, Wang S, Tao Y, Wang Z, Shu Y, et al. CarNAC4, a NAC-type chickpea transcription factor conferring enhanced drought and salt stress tolerances in *Arabidopsis*. *Plant Cell Reports*. 2016;35(3):613-27.
- Zadoks JC, Chang TT, Konzak CF. A decimal code for the growth stages of cereals. *Weed Research*. 1974;14(6):415-21.
- Zhang A, Li N, Gong L, Gou X, Wang B, Deng X, et al. Global analysis of gene expression in response to whole-chromosome aneuploidy in hexaploid wheat. *Plant Physiology*. 2017;pp.00819.2017.
- Zhang B, Horvath S. A General Framework for Weighted Gene Co-Expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology*. 2005;4(1).
- Zhang B, Li W, Chang X, Li R, Jing R. Effects of Favorable Alleles for Water-Soluble Carbohydrates at Grain Filling on Grain Weight under Drought and Heat Stresses in Wheat. Ali J, éditeur. *PLoS ONE*. 2014;9(7):e102917.
- Zhang H, Mao X, Jing R, Chang X, Xie H. Characterization of a common wheat (*Triticum aestivum* L.) TaSnRK2.7 gene involved in abiotic stress responses. *J Exp Bot*. 2011;62(3):975-88.
- Zhang H, Mao X, Wang C, Jing R. Overexpression of a Common Wheat Gene TaSnRK2.8 Enhances Tolerance to Drought, Salt and Low Temperature in *Arabidopsis*. *PLOS ONE*. 2010;5(12):e16041.
- Zhang L, Zhang L, Xia C, Zhao G, Jia J, Kong X. The Novel Wheat Transcription Factor TaNAC47 Enhances Multiple Abiotic Stress Tolerances in Transgenic Plants. *Front Plant Sci*. 2016;6.
- Zhang L-M, Leng C-Y, Luo H, Wu X-Y, Liu Z-Q, Zhang Y-M, et al. Sweet Sorghum Originated through Selection of *Dry*, a Plant-Specific NAC Transcription Factor Gene. *The Plant Cell*. 2018;30(10):2286-307.
- Zhang R, Geng S, Qin Z, Tang Z, Liu C, Liu D, et al. The genome-wide transcriptional consequences of the nullisomic-tetrasomic stocks for homoeologous group 7 in bread wheat. *BMC Genomics*. 2019;20(1).
- Zhao D, Derkx AP, Liu D-C, Buchner P, Hawkesford MJ. Overexpression of a NAC transcription factor delays leaf senescence and increases grain nitrogen concentration in wheat. *Plant Biol (Stuttg)*. 2015;17(4):904-13.
- Zhao X, Yang X, Pei S, He G, Wang X, Tang Q, et al. The *Miscanthus* NAC transcription factor MINAC9 enhances abiotic stress tolerance in transgenic *Arabidopsis*. *Gene*. 2016;586(1):158-69.
- Zhao Y, Tian X, Wang F, Zhang L, Xin M, Hu Z, et al. Characterization of wheat MYB genes responsive to high temperatures. *BMC Plant Biology*. 2017;17(1).
- Zheng X, Chen B, Lu G, Han B. Overexpression of a NAC transcription factor enhances rice drought and salt tolerance. *Biochemical and Biophysical Research Communications*. 2009;379(4):985-9.
- Zheng Y, Wang Z, Sun X, Jia A, Jiang G, Li Z. Higher salinity tolerance cultivars of winter wheat relieved senescence at reproductive stage. *Environmental and Experimental Botany*. 2008;62(2):129-38.

- Zheng X, Tang S, Zhu S, Dai Q, Liu T. Identification of an NAC Transcription Factor Family by Deep Transcriptome Sequencing in Onion (*Allium cepa* L.). Min XJ, éditeur. PLOS ONE. 22 juin 2016;11(6):e0157871.
- Zhong R, Lee C, Ye Z-H. Global Analysis of Direct Targets of Secondary Wall NAC Master Switches in Arabidopsis. *Molecular Plant*. 2010;3(6):1087-103.
- Zhou S, Hu W, Deng X, Ma Z, Chen L, Huang C, et al. Overexpression of the Wheat Aquaporin Gene, TaAQP7, Enhances Drought Tolerance in Transgenic Tobacco. Tran L-SP, éditeur. PLoS ONE. 2012;7(12):e52439.
- Zhu G, Chen G, Zhu J, Zhu Y, Lu X, Li X, et al. Molecular Characterization and Expression Profiling of NAC Transcription Factors in *Brachypodium distachyon* L. PLOS ONE. 2015a;10(10):e0139794.
- Zhu M, Chen G, Zhou S, Tu Y, Wang Y, Dong T, et al. A new tomato NAC (NAM/ATAF1/2/CUC2) transcription factor, SINAC4, functions as a positive regulator of fruit ripening and carotenoid accumulation. *Plant Cell Physiol*. 2014a;55(1):119-35.
- Zhu M, Hu Z, Zhou S, Wang L, Dong T, Pan Y, et al. Molecular Characterization of Six Tissue-Specific or Stress-Inducible Genes of NAC Transcription Factor Family in Tomato (*Solanum lycopersicum*). *Journal of Plant Growth Regulation*. 2014b;33(4):730-44.
- Zhu T, Nevo E, Sun D, Peng J. Phylogenetic analyses unravel the evolutionary history of NAC proteins in plants. *Evolution*. 2012;66(6):1833-48.
- Zhu X, Chen J, Xie Z, Gao J, Ren G, Gao S, et al. Jasmonic acid promotes degreening via MYC2/3/4- and ANAC019/055/072-mediated regulation of major chlorophyll catabolic genes. *The Plant Journal*. 2015b;84(3):597-610.
- Zhu X, Liu S, Meng C, Qin L, Kong L, Xia G. WRKY Transcription Factors in Wheat and Their Induction by Biotic and Abiotic Stress. *Plant Molecular Biology Reporter*. 2013;31(5):1053-67.
- Zhuo X, Zheng T, Zhang Z, Zhang Y, Jiang L, Ahmad S, et al. Genome-Wide Analysis of the NAC Transcription Factor Gene Family Reveals Differential Expression Patterns and Cold-Stress Responses in the Woody Plant *Prunus mume*. *Genes*. 2018;9(10):494.
- Zinn KE, Tunc-Ozdemir M, Harper JF. Temperature stress and plant sexual reproduction: uncovering the weakest links. *J Exp Bot*. 2010;61(7):1959-68.

ANNEXES

Annexe I – Amorces utilisées dans le cadre de cette thèse (1/3).

Gène suivi	Utilisation	Amorce sens (5' -> 3')	Amorce anti-sens (5' -> 3')	Taille de l'amplificat (pb)	T° optimale d'hybridation (°C)
TaNAC1	Clonage du gène complet chez le blé, utilisé pour la localisation subcellulaire après agro-infiltration en feuille de tabac	ATGGCAGACCACCTTCAA	TCAGTTCTGCCAAATGTCA	1044	54
TaNAC2		TCATGTGTTGTTGTTGTTG	ATGGCAGACCACCTTCAA	880	56
TaNAC3		TTACCGGTCAACGCCCTCT	ATGGCAGACCACCTTCAAGTTC	939	58
TaNAC4		ATGGCAGCCACCTTCAT	TTAGTAGTTCACATGCCATCCA	1050	58
TmNAC1	Clonage du gène complet chez le petit épeautre, utilisé pour faire la K7 transgénèse de surexpression	ATGTCTATGGTTCGCTCCAATATTACC	TCAATTCTGCCAAATGTCATCCA	1107	64
TmNAC2		ATGGCAGACCACCTTCAAGTTCAG	TCAGTACTTCCACATGCCATCCG	1068	65
TmNAC3		ATGGCAGACCACCTTCAAGTTCAA	TTAACTGTCAATGTTGTTGTGGCG	708	65
TmNAC4		ATGGCAGCCACCTTCAT	TTAGTAGTTCACATGCCATCCA	1047	60
Séquence intronique 1	Validation de la présence d'une cassette de sous-expression dans le génome des plantes T0 transgéniques de petit épeautre (par RT-qPCR)	TGTTAGAAATTCAAATCTGCTTGT	TCCCAACTGTAAATCAATCCAAA	338	62
Séquence intronique 2		GCTGGGTTCGAAATCGATAA	AAGTGGTCTAGAGGATCCAAAGC	452	62
TmNAC1-5 cumulés	Identification de la cassette de sous-expression dans le génome des plantes T1 transgéniques de petit épeautre (par RT-qPCR)	GATGCTCACCCCTGTTGTTTG	TGGTTGTTCCCGATGTAGTG	250	64
TmNAC1			TATTCTGGTAGTGGTGGTTCC	261	64
TmNAC2			GTTGCCCTTGGACGTTGTT	245	64
TmNAC3			CATCTGGTGGTGTGATTCT	260	64
TmNAC4			TCCCGATGTAGTGGTCATTC	243	64
TaNAC5_7A	Suivi de l'expression (par RT-qPCR) dans les plantes T2 transgéniques de surexpression et de sous-expression , chez le petit épeautre	AACACGGCTTCCAAGGAACAGTG	TCCAGGAAACAGCGGTGCA	227	64
TmNAC1		CATTAATGCATCTTCCAAGGAGGATAT	GCACCAGCGATGATATGCAC	198	62
TmNAC2		CCATTAATGCATCTTCCAAGAAGGAGC	AATGGTTGTTGTGGCGCGT	214	62
TmNAC3		TCTTCCAAGGAGGAGTATGTCGTTTGT	GCAGTGGGGCTAATGACGATGG	193	64
TmNAC4		TCACCAACGCGTCTTCCAAGGAA	ATGAGCCAGCACCGATGTCGC	260	62

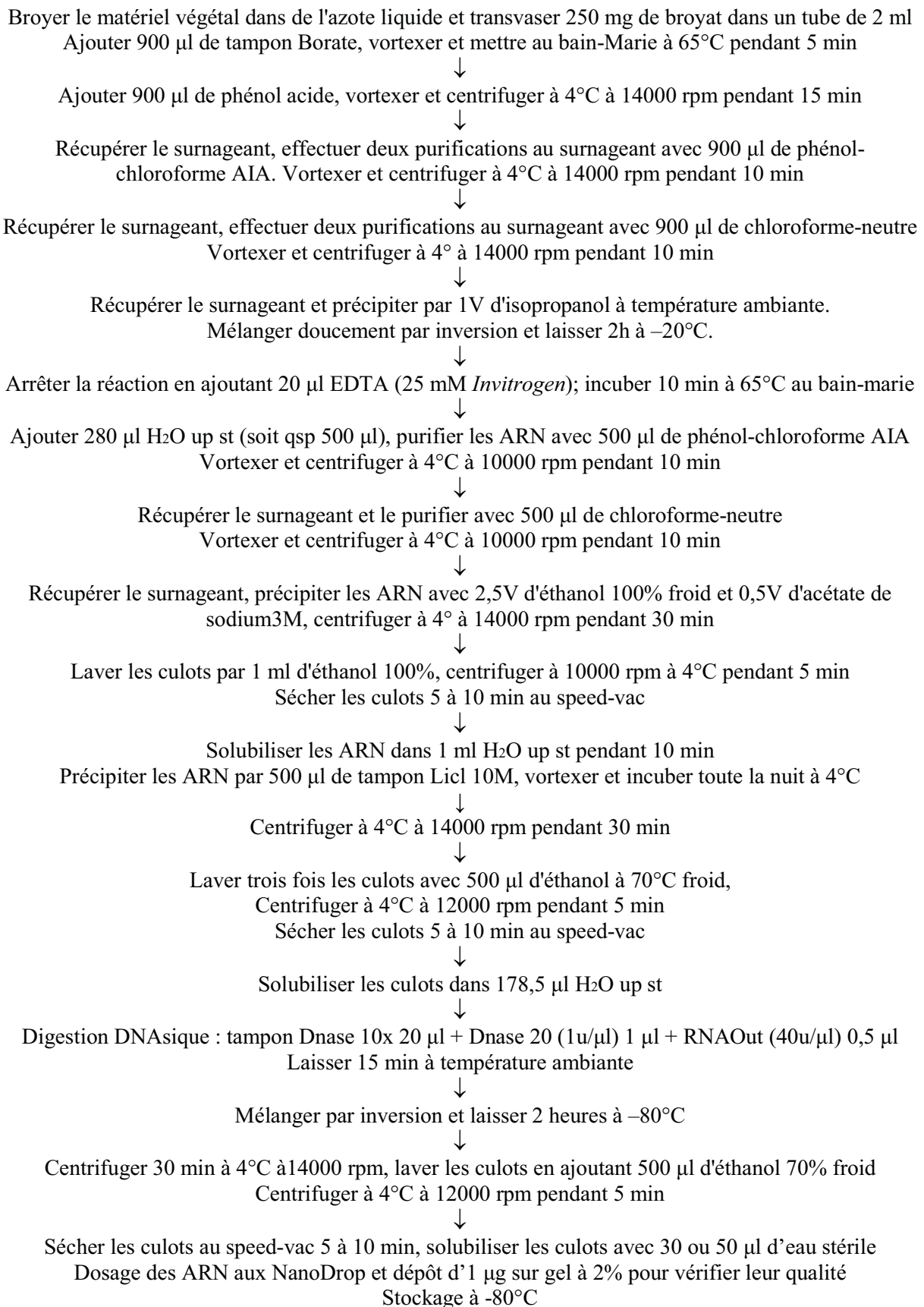
Annexe I – Amorces utilisées dans le cadre de cette thèse (2/3).

Gène suivi	Utilisation	Amorce sens (5' -> 3')	Amorce anti-sens (5' -> 3')	Taille de l'amplificat (pb)	T° optimale d'hybridation (°C)
Ta2991	Gènes de ménage pour RT-qPCR, testés mais non exploités.	GCTCTCCAACAACATTGCCAAC	GCTTCTGCCTGTCACATACGC	165	66
Ta54227		CAAATACGCCCATCAGGGAGAATC	CGCTGCCGAAACACGAGAC	227	66
Ta2776	Gène de ménage utilisé dans toutes les RT-qPCR	CGATTGAGAGCAGCGTATTGTTG	AGTTGGTCGGGTCTCTTCTAAATG	242	64
PDI intron	Contrôle de contamination génomique, utilisé dans toutes les RT-qPCR	CTTATGCAATCCAATGCTGG	TCTATGGTTCTGAAGAGGACC	250	56
TaNAC1 généraliste	Suivi de l'expression (par RT-qPCR) chez le blé en réponse à un stress thermique modéré durant les phases précoces de développement du grain (échantillons BreedWheat)	CAACAACATCAACAGATGATG	AAATGTCATCCATACCCATG	196	58
TaNAC1-7A		CACTACATCGGGAAACAACCACCAG	CCTTAGCCCGGCCACAACA	195	65
TaNAC1-7B		ACCACTACATCAGGAATAACCACCAC	CAGCCCGGCCACAGCA	194	65
TaNAC1-7D		CACTACATCGGGAAACAACCACCAC	CTCAGCCCGGCCACAGTG	194	65
TaNAC2 généraliste		TCATGGGCTATGGTGCATC	AACGACACCAACGGTCGT	193	65
TaNAC2-7A		GAGGAGAGGAGCAACATGGC	AGTGGTGGTTCCCGAAGTAGC	201	67
TaNAC2-7B		AGCAGAGGAGCAACATGCC	GGTAGTGGTGATTCTCTGAAATAGT	203	65
TaNAC2-7D		GAGCAGAGGAGCAACATGGA	GTAGTGGTGGTTCCCGAAGTAGT	203	67
TaNAC3 généraliste		TCACCACCTTCTTGACTCCAATG	AAGGGCAGGGCGCCTAG	190	67
TaNAC3-5A		AAGAACGACCCGTATGAGCTC	GCCGATGAGAGTGGGCAG	183	67
TaNAC3-5B		AAGAACGAGCCATATGAGCTT	CCGATGAGAGTGGCGCT	182	58
TaNAC4 généraliste		GTGCGGTGCTTCCAATGA	TGTTGCTCGTCCCAGTTAGAA	216	65
TaNAC4-7A		CGTCTTCCAAGGAGGAGTATGTCA	ACCGCACTACCCATTATAAATGACT	265	56
TaNAC4-7B		TAACCAATACITTTGAAAATCAT	CATGTGACACTATAGATGATGTG	167	58
TaNAC4-7D		CAACCAATACITTTGAGAATC	CATGTGACACTATAGATGATGTA	167	58
TaNAC5 géné		AACAAGTACGAGCCATGGG	CCTTACTCATGCATGATCCAGTT	264	62
TaNAC5_7A		TCACCAACGCGTCTTCCAAGGAA	ATGAGCCAGCACCGATGTCGC	260	62
TaNAC5_7D		GGCGGCTTCCAAGGAACAGTG	GAAGCGCCATGCTGCCAATC	271	56

Annexe I – Amorces utilisées dans le cadre de cette thèse (3/3).

Gène suivi	Utilisation	Amorce sens (5' -> 3')	Amorce anti-sens (5' -> 3')	Taille de l'amplificat (pb)	T° optimale d'hybridation (°C)
TmNAC1	Suivi de l'expression (par RT-qPCR) chez le petit épeautre en réponse à un stress thermique modéré durant les phases précoces de développement du grain	AACACGGCTTCCAAGGAACAGTG	TCCAGGAAACAGCGGTGCA	227	64
TmNAC2		CATTAATGCATCTTCAAGGAGGAGTAT	GCACGCGATGATGATGCAC	198	62
TmNAC3		CCATTAATGCATCTTCCAAGAAAGGAGC	AATGGTTGTTGTGGCGCCGT	214	62
TmNAC4		TCTTCCAAGGAGGAGTATGTCGTTTGT	GCAGTGGGGCTAATGACGATGG	193	64
TraesCS2A01G101100.1	Suivi de l'expression (par RT-qPCR) chez le blé en réponse à un stress sécheresse en plein champs ou un choc thermique sur feuilles	GGCTACGAAGAAGGGGTACTGG	TTTGTGAACACCCGGCAAACA	238	64
TraesCS2A01G566000.1		GCTTCAAGGACCGCAAGTA	GCTTGTCTGGAGCGGATAC	212	60
TraesCS2A01G566300.1		AACCAGAGGCAATCCTACAC	AGATGCCCTTCTTGCACTC	211	62
TraesCS3A01G077900.1		AACCAACCGAGCCACAAA	TGTTGATCTTGAGCCTGTACTC	182	62
TraesCS3B01G303800.1		CTCCGCGTCCACTTCAGGTAC	GGGATTCGGGGAAGATGAGCAG	165	64
TraesCS3B01G407600.1		CTTGTGTCGACTTTACACCAATC	GCAGCCAATGTGGCAATATC	219	62
TraesCS4B01G098200.1		AGATCCCGCGCAAGG	GAGGAGCTGGAGTCCGTGTG	187	64
TraesCS4B01G174000.1		CATCCTCGTCGGAATGAAGAA	AGTCAGTGATGAAGCCCAAG	248	62
TraesCS4D01G176000.1		GCTTCATCGGTGACTTCCTT	TACGGGCTCGAGTACATCAT	193	62
TraesCS5A01G143200.1		CTCACCGACCTGCTCAAC	GATGTTACGTAGCCATTGTTATT	195	64
TraesCS5A01G245900.2 (TaNAC3)		CGCTCCGTTGTTTCTATCT	ACCATGAAGCACTGGTCTG	202	60
TraesCS5B01G141900.1		CACCGACTCCATGTTCACTATC	GATTGAAGCTTCCAGCTGTAGTA	173	62
TraesCS5B01G142100.1		GACGACCACCTCCAGTTGCA	TTTCGGGAGGCAGCTGGTAG	169	64
TraesCS5B01G275200.1		AACAAACGCGCAACTAATG	ATTGGAGGCTGGCTTCAG	200	62
TraesCS5D01G148800.1		CCTCACCGACCTGCTCAACA	TCATCTACATGCGCCTCGGG	231	64
TraesCS5D01G148900.1		GGTGAGCCCGGAAAAGGGTCT	TGTTATCAGCAGCTGGGCGT	163	64
TraesCS6A01G406700.1		AGCAGGCCTAGTCAGATTCGTG	AAAGAAAGGATAGGCCGCCG	234	64
TraesCS6D01G390200.1	Suivi de l'expression (par RT-qPCR) chez le blé en réponse à un stress sécheresse en plein champs ou un choc thermique sur feuilles	TGCTTCCCGTGAGAACCTTTTAT	CCGTGACACGAATCTGACTAGG	191	62
TraesCS7A01G305200.1		TTTAGCGAGCTGGACCTTC	GCCTTCTTCTCAGGGTTGT	191	62
TraesCS7A01G569100.1 (TaNAC1)		CAGTGTGGCGCTTCCTATTA	CCTGGGTAGTTGTTCCCTTAGC	237	64
TraesCS7B01G056300.1 (TaNAC2)		GAGATCTTCCACCATGTGACTAC	GCAAAACCACATACTCCTCCTT	210	64
TraesCS7B01G100300.1 (TaNAC4)		AGAATCATCGCCACCAGATG	TGCACCACCAATTCCTCTAAA	203	62
TraesCS7D01G315100.1		CATCTGCTCTCAAGGCGTATAA	CTTGAGGTGGTGGTGGTAG	195	62

Annexe II : Protocole d'extraction des ARN totaux - Adapté de Bogorad *et al*, 1983.



Annexe III – Extraction d'ADN plasmidique par lyse alcaline (mini prep)

Une colonie bactérienne transformée est mise en culture liquide dans 5 ml de milieu LB additionné d'antibiotique (Ampicilline 100 µg/ml, Spectinomycine 100 µg/ml, Kanamycine 50 µg/ml ou Gentamycine 50 µg/ml selon le plasmide utilisé) et placée 12h à 37°C. Les bactéries qui se développent possèdent le plasmide portant le gène de résistance à l'antibiotique et le gène d'intérêt que l'on cherche à isoler. Un aliquot de 1,8 ml est centrifugé à 14000 rpm pendant 3 min. Le culot bactérien obtenu est resuspendu dans 300 µl de solution I additionnée de RNase A à 1 µg/ml. Ensuite, 300 µl de solution II (NaOH, 0,2M et SDS1%) sont ajoutés pour permettre la lyse des bactéries après 5 min d'incubation à température ambiante. Le mélange est neutralisé par ajout de 300 µl de solution III (acétate de potassium, 3M et acide acétique glacial, 40%). Afin d'éliminer les débris cellulaires, l'ensemble est centrifugé (14000g pendant 15 min) et l'ADN plasmidique présent dans le surnageant est précipité dans 800 µl d'isopropanol. Pour finir, le culot d'ADN est lavé à l'éthanol 70%, séché et repris dans 20 µl d'eau stérile.

Solution I (500 ml)

Glucose (50 mM) : 4,504g

TRIS HCl (25 mM) : 1,970g

EDTA (10 mM) : 1,861g

H₂O qsp 500 ml

Ajuster à pH 8, autoclaver 20 mn à 120°C

Solution II

NaOH (5M) 40 g pour 200 ml de H₂O

SDS (10 %) 20g 200 ml de H₂O

Tampon de LYSE

SDS 20% : 93,75 µl

NaOH 5M : 75 µl

H₂O qsp 1,6125 ml

Solution III tampon High – Salt (500 ml)

Acétate de Potassium (3M) : 147,21g

A dissoudre dans une petite quantité d'eau

Ajuster pH 5,5 avec acide acétique

Ajuster le volume et autoclaver 20 mn à 120°C

Annexe IV : Milieux de culture in vitro - Quantités pour 1 litre

I Milieu de calogénèse (MSE3)

MS micro+macro+vit	4.4 g
Sucrose	60 g
Asparagine	0.150 g
Thiamine HCl	0.04 g
Agar	8 g
2-4 D (1mg/ml)	2.5 ml
PH 5,7 avec KOH	

II Milieu plasmolysant

MS micro+macro+vit	4.4 g
Maltose	100 g
Asparagine	0.150 g
Thiamine HCl	0.04 g
Agar	8 g
2-4-D (1 mg/ml)	2.5 ml

III Milieu de régénération (MSR)

MS micro+macro+vit.	4.4 g
Sucrose	20 g
Asparagine	0.150 g
Thiamine HCl	0.04 g
IAA (1mg/ml)	0.5 ml
BAP (1mg/ml)	1 ml
Agar	8 g

IV Milieu d'enracinement (MS0 ou P0)

MS micro+macro+vit	2,2 g
Sucrose	20 g
IAA (1mg/ml)	1 ml
Agar	8 g

Sélection sur PMI

MSR PMI 1

MS micro+macro+vit. 4.4 g
 Sucrose 5 g
 Mannose 1% (10ml d'une solution a 1g/ml, filtrée sur filtre 0.2μ)
 Asparagine 0.150 g
 Thiamine HCl 0.04 g
 IAA (1mg/ml) 0.5 ml
 BAP (1mg/ml) 1 ml
 Agar 8 g
 PH 5.7 avec KOH

MSR PMI 3

MS micro+macro+vit. 4.4 g
 Mannose 1.5% (15ml d'une solution a 1g/ml, filtrée sur filtre 0.2μ)
 Asparagine 0.150 g
 Thiamine HCl 0.04 g
 IAA (1mg/ml) 0.5 ml
 BAP (1mg/ml) 1 ml
 Agar 8 g
 PH 5.7 avec KOH

MSR PMI 2

MS micro+macro+vit. 4.4 g
 Sucrose 20 g
 Mannose 0.5% (5ml d'une solution a 1g/ml, filtrée sur filtre 0.2μ)
 Asparagine 0.150 g
 Thiamine HCl 0.04 g
 IAA (1mg/ml) 0.5 ml
 BAP (1mg/ml) 1 ml
 Agar 8 g
 PH 5.7 avec KOH

MS0 PMI

MS micro+macro+vit 2,2 g
 Mannose 1.5 % (15ml d'une solution a 1g/ml, filtrée sur filtre 0.2μ)
 IAA (1mg/ml) 1 ml
 Agar 8 g
 PH 5.7 avec KOH

Hormones

Ces solutions mères sont filtrées sur filtre seringue 0.2μ, et aliquotées par 5 ml dans des tubes stériles. Elles peuvent être préparées à l'avance et stockées à -20°C (3 mois maximum).

IAA (indole-3acetic acid)

1mg/ml
 IAA 100 mg
 HCL 2 ml
 H2O QSP 100 ml

BAP (6-benzilaminopurine)

1mg/ml
 BAP 100 mg
 HCL 2 ml
 H2O QSP 100 ml

2-4 D (dichloro-phenoxyacetic acid)

2-4 D 100 mg
 NaOH 2 ml
 H2O QSP 100 ml

Annexe V : Résultats d'association présentant un score ≥ 3 entre les SNP génotypés avec la puce **BW 420K et le phénotypage la Core Collection Agronomique cultivée en 3 environnements différents.** La mention « P/A » signifie que ce marqueur est un OTV et que c'est la variable Présence/Absence qui a montré une association significative avec le caractère phénotypique. Les mesures ont été réalisées dans 3 environnements : Clermont-ferrand (CF), Le Moulon fort azote (LMHN) et Le Moulon faible azote (LMLN). Les 2 gènes portant le plus de marqueurs associés à un phénotype sont surlignés en vert et en jaune.

Phénotype	Environnement	Gène	Chromosome	Position	Score
Pourcentage de gliadines dans l'azote total	LMHN	TraesCS2A01G561800	P/A	763314094	3.049
Ratio gliadine/glutenine	LMHN	TraesCS2A01G561800	P/A	763314094	3.148
Quantité d' α/β -gliadines par grain	LMHN	TraesCS2A01G561800	P/A	763314094	3.359
Pourcentage d' α/β -gliadines dans les gliadines	CF	TraesCS2A01G561800	P/A	763314094	3.377
Pourcentage de γ -gliadines dans les gliadines	LMHN	TraesCS2A01G561800	P/A	763314094	3.684
Pourcentage d' α/β -gliadines dans les gliadines	LMHN	TraesCS2A01G561800	P/A	763314094	4.833
Pourcentage d' α/β -gliadines dans les protéines de réserve	LMHN	TraesCS2A01G561800	P/A	763314094	5.741
Quantité d' α/β -gliadines par grain	LMHN	TraesCS2A01G561800	P/A	763313260	3.041
Pourcentage de LMW-GS dans l'azote total	LMHN	TraesCS2A01G561800	P/A	763313260	3.102
Pourcentage d' α/β -gliadines dans les gliadines	LMHN	TraesCS2A01G561800	P/A	763313260	3.719
Pourcentage de γ -gliadines dans les gliadines	LMHN	TraesCS2A01G561800	P/A	763313260	3.879
Pourcentage de γ -gliadines dans les gliadines	CF	TraesCS2A01G561800	P/A	763313260	3.920
Pourcentage d' α/β -gliadines dans les protéines de réserve	LMHN	TraesCS2A01G561800	P/A	763313260	4.647
Pourcentage d' α/β -gliadines dans les protéines de réserve	CF	TraesCS2A01G561800		763313240	3.022
Pourcentage de γ -gliadines dans les gliadines	CF	TraesCS2A01G561800		763313240	3.340
Pourcentage de gliadines dans l'azote total	LMHN	TraesCS2A01G561800	P/A	763314613	3.169
Ratio gliadine/glutenine	LMHN	TraesCS2A01G561800	P/A	763314613	3.516

Phénotype	Environnement	Gène	Chromosome	Position	-Log(p-value)
Pourcentage d' α/β -gliadines dans les protéines de réserve	CF	TraesCS2A01G561800	2A	763314613	3.554
Quantité d' α/β -gliadines par grain	LMHN	TraesCS2A01G561800	2A	763314613	3.682
Pourcentage d' α/β -gliadines dans les protéines de réserve	LMNL	TraesCS2A01G561800	2A	763314613	3.716
Pourcentage d' α/β -gliadines dans les gliadines	CF	TraesCS2A01G561800	2A	763314613	3.897
Pourcentage de γ -gliadines dans les gliadines	CF	TraesCS2A01G561800	2A	763314613	4.059
Pourcentage de γ -gliadines dans les gliadines	LMHN	TraesCS2A01G561800	2A	763314613	4.066
Pourcentage d' α/β -gliadines dans les gliadines	LMHN	TraesCS2A01G561800	2A	763314613	5.038
Pourcentage d' α/β -gliadines dans les protéines de réserve	LMHN	TraesCS2A01G561800	2A	763314613	6.153
Ratio HMW-GS / LMW-GS	CF	TraesCS3B01G126200	3B	102595710	3.290
Ratio HMW-GS / LMW-GS	CF	TraesCS3B01G126200	3B	102594636	3.290
Ratio HMW-GS / LMW-GS	LMLN	TraesCS4B01G350600	4B	643311006	3.017
Quantité d'HMW-GS par grain	LMLN	TraesCS4B01G350600	4B	643311006	3.403
Ratio HMW-GS / LMW-GS	LMLN	TraesCS4B01G350600	4B	643311663	3.028
Quantité d'HMW-GS par grain	LMLN	TraesCS4B01G350600	4B	643311663	3.737
Pourcentage d' α/β -gliadines dans les gliadines	CF	TraesCS5B01G415700	5B	590000008	3.235
Pourcentage d' α/β -gliadines dans les gliadines	LMHN	TraesCS5B01G415700	5B	590000008	3.290
Pourcentage d' α/β -gliadines dans les protéines de réserve	CF	TraesCS6A01G051700	6A	26934518	3.538
Pourcentage d' α/β -gliadines dans les gliadines	CF	TraesCS6A01G051700	6A	26934518	4.178
Pourcentage d' α/β -gliadines dans les protéines de réserve	CF	TraesCS6A01G051700	6A	26933645	3.447
Pourcentage d' α/β -gliadines dans les gliadines	CF	TraesCS6A01G051700	6A	26933645	4.184
Pourcentage d' α/β -gliadines dans les protéines de réserve	CF	TraesCS6A01G051700	6A	26933577	3.450
Pourcentage d' α/β -gliadines dans les gliadines	CF	TraesCS6A01G051700	6A	26933577	4.193
Pourcentage d' α/β -gliadines dans les protéines de réserve	CF	TraesCS6A01G051700	6A	26935399	3.261
Pourcentage d' α/β -gliadines dans les gliadines	CF	TraesCS6A01G051700	6A	26935399	3.721

Phénotype	Environnement	Gène	Chromosome	Position	-Log(p-value)
Pourcentage d' α/β -gliadines dans les protéines de réserve	CF	TraesCS6A01G051700	6A	26933950	3.151
Pourcentage d' α/β -gliadines dans les gliadines	CF	TraesCS6A01G051700	6A	26933950	3.471
Ratio HMW-GS / LMW-GS	LMLN	TraesCS6A01G065700	6A	35181387	3.178
Pourcentage de γ -gliadines dans les gliadines	CF	TraesCS6B01G451300	6B	710101568	3.868
Concentration en protéines du grain	LMLN	TraesCS7A01G464800	7A	661279132	3.068
Concentration en protéines du grain	LMLN	TraesCS7A01G464800	7A	661279168	3.664
Quantité de w1,2-gliadines par grain	LMHN	TraesCS7B01G056300	7B	59180596	3.019
Quantité de w1,2-gliadines par grain	LMHN	TraesCS7B01G056300	7B	59180076	3.080
Pourcentage de γ -gliadines dans les protéines de réserve	LMLN	TraesCS7B01G364600	7B	627611636	3.215
Pourcentage de γ -gliadines dans les protéines de réserve	CF	TraesCS7B01G364600	7B	627611636	3.463
Pourcentage de protéines de réserve dans l'azote total	LMLN	TraesCS7B01G481500	7B	735811712	3.011
Pourcentage de gluténines dans l'azote total	LMLN	TraesCS7B01G481500	7B	735811712	3.934
Pourcentage de gluténines dans l'azote total	LMLN	TraesCS7B01G489500	7B	743457269	3.900
Pourcentage de gluténines dans l'azote total	LMLN	TraesCS7D01G543400	P/A	631921203	3.124

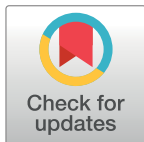
RESEARCH ARTICLE

Genome-wide analysis, expansion and expression of the NAC family under drought and heat stresses in bread wheat (*T. aestivum* L.)

Claire Guérin, Jane Roche^{*}, Vincent Allard, Catherine Ravel, Said Mouzeyar, Mohamed Fouad Bouzidi

UMR 1095 Génétique, Diversité et Ecophysiologie des Céréales, Université Clermont Auvergne, INRA, Clermont-Ferrand, France

* jane.roche@uca.fr



Abstract

The NAC family is one of the largest plant-specific transcription factor families, and some of its members are known to play major roles in plant development and response to biotic and abiotic stresses. Here, we inventoried 488 NAC members in bread wheat (*Triticum aestivum*). Using the recent release of the wheat genome (IWGS RefSeq v1.0), we studied duplication events focusing on genomic regions from 4B-4D-5A chromosomes as an example of the family expansion and neofunctionalization of *TaNAC* members. Differentially expressed *TaNAC* genes in organs and in response to abiotic stresses were identified using publicly available RNAseq data. Expression profiling of 23 selected candidate *TaNAC* genes was studied in leaf and grain from two bread wheat genotypes at two developmental stages in field drought conditions and revealed insights into their specific and/or overlapping expression patterns. This study showed that, of the 23 *TaNAC* genes, seven have a leaf-specific expression and five have a grain-specific expression. In addition, the grain-specific genes profiles in response to drought depend on the genotype. These genes may be considered as potential candidates for further functional validation and could present an interest for crop improvement programs in response to climate change. Globally, the present study provides new insights into evolution, divergence and functional analysis of *NAC* gene family in bread wheat.

OPEN ACCESS

Citation: Guérin C, Roche J, Allard V, Ravel C, Mouzeyar S, Bouzidi MF (2019) Genome-wide analysis, expansion and expression of the NAC family under drought and heat stresses in bread wheat (*T. aestivum* L.). PLoS ONE 14(3): e0213390. <https://doi.org/10.1371/journal.pone.0213390>

Editor: Maoteng Li, Huazhong University of Science and Technology, CHINA

Received: October 15, 2018

Accepted: February 19, 2019

Published: March 6, 2019

Copyright: © 2019 Guérin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

As sessile organisms, plants have developed diverse strategies to modulate their development and growth according to environmental signals such as day/night alternation or daily temperature variations [1]. At the molecular and cellular levels, the early mechanisms of response to these stimuli include signal perception by receptors, followed by signalling cascades involving changes in membrane permeability, protein phosphorylations by protein kinases or regulations of the target protein expression by transcription factors (TFs) [2]. About 6–8% of the

plant genome is allocated to the coding of more than 1,500 TFs [3] of which 45% belong to plant-specific families [4]. Among more than 80 TF families, AP2/EREBP, bZIP, MYB/MYC, NAC and WRKY families are known to be strongly involved in the response to biotic or abiotic stresses in plants [5–16]. The NAC (NAM (No Apical Meristem)–ATAF (Arabidopsis Transcription Activation Factor)–CUC (Cup-shaped Cotyledons)) family is one of the largest groups of plant specific TFs, and has been described in several plant species. In the model plants, Arabidopsis and rice, this family includes 117 and 151 members, respectively [17]. NAC family members have also been reported in *Brachypodium distachyon* (118 members) [18], soybean (101 genes) [19], maize (157 members) [20], durum wheat (168 members) [21], and more recently in *Fragaria* × *ananas* fruits (112 members) [22].

A typical NAC protein contains a highly conserved DNA-binding domain, measuring ~150 amino acids, and located in the N-terminal region. This domain is subdivided into five subdomains named A to E [23]. Subdomains A, C, and D are highly conserved. Subdomains C and D are probably involved in the binding of the NAC protein to its DNA target [24]. An additional poorly conserved transcription regulatory region is present in the C-terminal region of the protein and acts as a transcriptional activator [25] or repressor [26] of target genes. The C-terminal regulatory region often contains simple amino acid repeats and regions rich in serine and threonine, proline and glutamine, or acidic residues [27]. While the majority of the NAC TFs were found as soluble proteins in the nucleus, some NAC have been identified as membrane-bound TFs. They are classified as membrane-related and named NTL (NTM1-like or ‘NAC with transmembrane motif 1’-Like) TFs. For example, nine NTL were identified in durum wheat [21] and at least 85 and 45 in Arabidopsis and rice respectively [28].

Expression of NAC family genes is modulated during plant development and in response to biotic and abiotic stresses [27]. NACs are one of the largest families of genes with very diverse functions. The first NAC gene was described as related to the shoot apical meristem development and the position determination of meristems and primordia [29]. The NAC family regulates many developmental and growth processes including flowering [30], root development [25], leaf senescence [31], secondary cell wall biosynthesis [32], cell division [33], and seed development [34]. The NAC family is also related to grain yield and quality. For example, a reduced expression of *OsNAP* in rice and *TaNAC-S* in wheat increases the grain yield by delaying leaf senescence [35–38]. Moreover, NAC proteins have also been shown to be involved in response to biotic stresses such as pathogen infection [39], hypersensitive response-like cell death [40], and abiotic stresses including salt, drought, and temperature [41–43]. Due to their key role in abiotic stress signalling in plants, NAC TFs could be used to improve crop plant tolerance to unfavourable environmental conditions. This is particularly important for bread wheat (*Triticum aestivum*) because the yield of this major crop, which constitutes the diet basis of millions of people, is constrained by abiotic stresses.

Bread wheat is a hexaploid species. Its genome is composed of three subgenomes AA (originated from *T. urartu*), BB (from *Aegilops speltoides*) and DD (from *Ae. tauschii*) [44]. In a recent publication, Borrill *et al.* [45] carried out an inventory of the TF coding genes in wheat using published TGAC gene models [46], with an emphasis on the NAC family. They identified 574 (453 high and 121 low-confidence) NAC genes belonging to eight subfamilies. More recently, a high-quality version of the wheat genome (cv Chinese Spring) assembly was released by the International Wheat Genome Sequencing Consortium (IWGSC) [47]. In this study, we used this last version to analyse the genomic organization and the expansion of the NAC family. We identified 488 (460 high and 28 low-confidence) NAC genes, named *TaNAC*, in the three subgenomes A, B, and D of bread wheat. We established the correspondence (one-to-one) between the two annotations (488 vs 574) yielding a consensus subset containing 361 identical sequences. The expression profiles of these 361 identical *TaNAC* genes were

monitored in different tissues (grains, spikes, leaves and roots) or in response to different abiotic stresses (heat, drought, and heat + drought combination), based on the data available on the www.wheat-expression.com platform [48]. In addition, 23 representative *TaNAC* genes were selected to investigate their response to drought stress using two contrasting wheat cultivars at two developmental stages (220 and 450 degree days ($^{\circ}\text{Cd}$) after anthesis). Our results provide valuable information about *TaNAC* candidate genes to improve drought stress tolerance in wheat.

Materials & methods

Meta-analysis of the NAC family in bread wheat

Survey of *TaNAC* in the wheat sequence. For *TaNAC* retrieval, the annotated proteins in the database of the International Wheat Genome Sequencing Consortium (IWGS RefSeq v1.0) [47] were searched using the HMMER2 program implemented in Unipro UGENE 1.25 [49] and by querying the NAM motif (PF02365, generated by alignment of 587 seed sequences) downloaded from Pfam.

The full-length sequences obtained with an E-value cut-off e^{-1} were scanned for NAC domain superfamily (IPR036093) using InterProScan (<http://www.ebi.ac.uk/interpro/search/sequence-search>, S1 Table) [50]. The combination of HMMER2 and InterProScan searches led to a final set of 488 *TaNAC* proteins, whose 460 are high-confidence and 28 are low-confidence. The *TaNAC* genes were mapped on the wheat pseudomolecule from the IWGS Consortium to determine their chromosomal coordinates and visualized using MapGen-e2Chromosome V2 (http://mg2c.iask.in/mg2c_v2.0/). Homoeologous genes were determined using their chromosomal coordinates.

To create a high quality consensus set of *TaNAC* genes, a reciprocal homology search was performed between the TGACv1 [46] and the IWGS RefSeq v1.0 [47] cDNA databases (including high and low-confidence sequences), using Usearch program [51]. This query led to the identification of common sequences between the *TaNAC* gene set of Borrill *et al.* [45] extracted from TGACv1 cDNA database and the set of 488 *TaNAC* sequences extracted from IWGS RefSeq V1.0. To minimize the fortuitous detection of homology, the final set of protein-coding sequences in both annotations was used as a query. When reciprocal homology search identified the same pair of *TaNAC* genes in both directions, they were considered as homologs.

The NAC family structuration in bread wheat. To study the structuration of the *TaNAC* family and confirm the homoeologous groups, a phylogenetic study was performed. The maximum likelihood method available in MEGA7 software [52] was used to construct an unrooted phylogenetic tree based on the alignment of the amino acid sequences of the common set of sequences established. Alignment was performed using the MUSCLE alignment algorithm [53]. The tree obtained was the consensus of 500 single trees provided by bootstraps, gaps/missing data treatment: partial deletion, model/method LG model, rates among sites: gamma distributed with invariant sites (G).

To explore the contribution of gene duplication to the expansion of the NAC family in bread wheat, we focused on flanking sequences of each *TaNAC* gene. For each locus containing a *TaNAC* gene, 25 Kb on either side of the gene were collected. When several *TaNAC* genes are separated by less than 25 Kb, the window for sequence retrieval was extended to the 25 Kb beyond the leftmost and rightmost *TaNAC* genes. Homologies between nucleotide sequences were obtained using YASS, a Blast-like tool [54] with an e-value threshold of e^{-30} . The homology search results were then parsed and sequences with at least 80% identity over

more than 2 Kb were considered duplicates. Figures showing these duplications were generated using CIRCA (OMGenomics, <http://omgenomics.com/circa/>).

To monitor the correlation between expression profiles of duplicated genes, Pearson correlation coefficients between *TaNAC* genes were calculated using the R software (<http://www.R-project.org>, R development Core Team, 2008).

Expression of *TaNAC* genes. The expression profiles of *TaNAC* genes were monitored in different tissues (grains, spikes, leaves, and roots) [55] or in response to different abiotic stresses in leaves (heat shock, polyethylene glycol-induced drought, and a combination of both at a young stage) [56], based on the RNA seq data available on the www.wheat-expression.com platform [48]. Briefly, raw expression data were downloaded and analysed using the R package limma [57]. Raw expression data corresponding to the entire wheat genome were first filtered to remove weakly expressed genes (genes are retained if they are expressed at a counts-per-million (CPM) > 0.5 in at least two samples). The expression of retained genes were then normalized using TMM and Voom methods. Finally, expressions were compared between samples and those genes that displayed an absolute log Fold Change (logFC) > 1 with an adjusted p-value < 0.05 were deemed Differentially Expressed and noted DEG. Finally, *TaNAC* genes were retrieved from this set of DEG genes. Hierarchical Clustering of *TaNAC* DEG was conducted using Mev software [58] with euclidean distance and average linkage.

On the basis of this meta-analysis, 23 *TaNAC* genes which displayed large fold changes in their expression were selected to assess their response in wheat plants cultivated in field conditions and submitted to drought.

Expression analysis of 23 *TaNAC* candidate genes in field-grown stressed wheat plants

To gain insights into the function of wheat *NAC* genes in response to abiotic stresses, we decided to assess the expression patterns of 23 selected *TaNAC* genes in plants subjected to drought stress in field.

Plant material and field drought treatment. Field drought treatment was performed on the Pheno3C plant phenotyping platform. This four automatic rain shelters facility is able to generate contrasted water regimes by protecting four “stressed” plots from rainfall, while four “control” plots are managed under natural weather and irrigated if necessary to prevent water stress. The eight plots are distributed over a homogeneous three hectares area and minimum distance between two plots is 30 m. Each plot (22 m x 25 m) is subdivided into 96 micro-plots (1.9 m x 1.2 m) where different genotypes were sown. Among those genotypes and based on local expert knowledge and previous trials, two modern bread wheat genotypes (Allez-y (2011, Limagrain) and Altigo (2007, Limagrain) were selected as stress sensitive and stress tolerant cultivars respectively on the basis of yield impact under dry conditions. Soil volumetric water content (SWC) was recorded hourly with CS 655 (Cambell Scientific, Logan, UT, USA) sensors installed at four depth (10, 35, 50 and 75 cm) at three locations per plot. Total soil water content was calculated by integrating sensor values up to 100 cm depth (average soil depth of the plots). Average soil field capacity and available water capacity are respectively 570 and 280 mm. From those values and daily measurements of SWC, a water stress factor (Ks) was calculated from Allen *et al.* (1998) [59]. A Ks value below 1 is considered as an indicator of water stress that increases in terms of severity when the Ks value decreases.

All plots were sown on 11/04/2016, and managed following local agronomic practices. The stress was applied starting on 03/24/2017 by sheltering the stressed plots. At anthesis (05/20/2017), control plots had received 130 mm more rain than stressed plots, leading to a Ks of 0.85 and 0.36 in the control and stressed plots respectively, defining a very moderate stress in the

control and a severe stress in the stress plots. Converted degree-days ($^{\circ}\text{Cd}$) were calculated as the sum of mean daily temperature (base 0) from anthesis. For the present experiment, three micro-plots per genotype and per hydric conditions (biological triplicates) located in two adjacent blocks were used at both sampling dates. The samplings occurred at 250°Cd after anthesis (end of grain cell division) and 450°Cd after anthesis (middle of the linear phase of grain filling). At 250°Cd , water stresses were very moderate ($K_s = 0.91$) and very severe ($K_s = 0.27$) in the control and stressed plots respectively. At 450°Cd , water stresses were moderate ($K_s = 0.6$) and very severe ($K_s = 0.2$) in the control and stressed plots respectively. For each sample, three spikes and three leaves were taken randomly and frozen immediately in liquid nitrogen prior to subsequent RT-qPCR analysis. Yield components of each plot were also assessed. Spike number per m^2 , which is determined before anthesis [60], is a good indicator of the plant stress response before flowering, reflecting the stress impact on the plant tillering. This variable was determined manually by counting all spikes on two 1.9 m rows per micro-plot. Grain yield was measured at maturity with an experimental plot combine harvester.

RNA isolation and relative quantification of candidate genes. To quantify the expression of the 23 candidate genes, a total RNA extraction was performed on the collected samples, using the protocol described by Capron *et al.* [61] for grain samples, and the TRIzol protocol according to the manufacturer's instructions (Invitrogen) for leaf samples. Reverse-transcription (RT) was performed on 2.5 μg of total RNA, using the Thermo Scientific Maxima First Strand cDNA Synthesis Kit according to the manufacturer's instructions. A RNase H (Thermo Scientific 18021–071) treatment was performed after the reverse transcription.

When selecting the 23 *Ta*NAC genes, only one member per homoeologous group was retained. Then, their specific expression profile was studied using Real Time-quantitative Polymerase Chain Reaction (RT-qPCR) and by designing primers specific to the coding sequence of the considered homoeolog (listed in S2 Table), with PrimerQuest (<https://eu.idtdna.com/PrimerQuest>). Amplification efficiencies of primer pairs designed were tested to keep only those with an efficiency comprised between 90 and 110% (S2 Table).

Real Time-quantitative Polymerase Chain Reactions (RT-qPCR) were carried out in a 15 μl volume containing 25 ng of cDNA, 7.5 nM gene-specific primers and 7.5 μl of LightCycler 480 SYBR Green I Master (Roche Diagnostics #04887352001), on a LightCycler 480 system (Roche) on the Gentyane platform, INRA de Crouël, France (<http://gentyane.clermont.inra.fr/>). The thermal cycle used was done as follows: a pre-incubation at 95°C for 10 min, 40 amplifications cycles of 95°C for 10s, optimal temperature of the primer for 15s, and 72°C for 15s.

Relative target gene expression was determined by the $2^{-\Delta\Delta\text{Ct}}$ method [62] using Ta2776 (RNase L inhibitor-like protein) for normalization [61]. A principal component analysis was carried out using the expression data of the whole set of 23 genes across all samples and the R package ggfortify [63]. For each gene, a Kruskal-Wallis test was performed to uncover differences between samples.

Results

Inventory and description of the NAC TF family in bread wheat using the IWGS RefSeq v1.0 database

The HMM profile of the Pfam NAC domain (PF02365) was used as a query to identify *Ta*NAC genes in the most recent version of the annotated wheat genome released by the IWGSC RefSeq v1.0, which contains the latest assembly of the wheat cv *Chinese Spring* pseudomolecule.

The initial HMM search identified 501 putative *Ta*NAC sequences that were manually curated for the presence of the NAC domain (IPR036093) signature using the Interproscan platform. Thirteen sequences were discarded, giving a whole set of 488 full-length sequences

(460 high-confidence and 28 low-confidence sequences) containing the NAC domain ([S1 Table](#)).

This set was compared with the sequences identified by Borrill *et al.* [45] using the TGAC database. This analysis provided a robust subset of 361 common sequences, showing a single hit between the two databases ([S3 Table](#)). The combination of the two annotations improved the knowledge on the *Ta*NAC genes present in this subset, particularly in terms of chromosomal location. For example, no information on the chromosomal location of the TraesC-SU01G137200 gene was available from the IWGS RefSeq v1.0 database, but it was found to be identical to the TRIAE_CS42_2AL_TGACv1_093618_AA0283710 gene in the TGAC database mapped on the chromosome 2A. Similarly, among the 13 non-mapped sequences of the TGAC database, 12 were identical to already mapped IWGS sequences. Finally, 359 *Ta*NAC genes among the 361 consensus sequences have been mapped on the 21 bread wheat chromosomes, the two remaining sequences did not have chromosomal location information. Chromosomes 2 and 7 contain the highest numbers of *Ta*NAC genes with 89 and 81 members respectively, which represent 24.8% and 22.6% of the total. On the contrary, chromosome 1 carries the lowest number of *Ta*NAC sequences, with only 17 genes (4.7%) ([S1 Fig](#)). The *Ta*NAC genes are evenly distributed among the subgenomes A, B, and D with a similar proportions (34.5%, 33.5%, and 32.0% respectively) indicating that the expansion of the NAC family predates the bread wheat appearance through interspecific hybridization.

*Ta*NAC genes are mostly distributed next to the telomeric regions vs centromeric one ([S1 Fig](#)). Furthermore, most of these genes are tightly linked and lay within clusters. For example, TraesCS4B01G328600, TraesCS4B01G328700, TraesCS4B01G328800, TraesCS4B01G328900, and TraesCS4B01G329100 are close on chromosome 4B within 347 Kb. Similarly, the *Ta*NAC genes denoted TraesCS4A01G419900, TraesCS4A01G420000, TraesCS4A01G420100, TraesCS4A01G420200, TraesCS4A01G420900, TraesCS4A01G421000, TraesCS4A01G421100, TraesCS4A01G421200, and TraesCS4A01G421300 are tightly linked and lay within 796 Kb.

The genomic length of the sequences within the whole set of 488 *Ta*NAC genes, ranged from 273 to 15074 base pairs, corresponding to protein sequences from 91 (TraesCS2B01G556700LC.1) to 1030 amino acids (TraesCS5B01G550800.1) with a mean length of 370 amino acids. One-fifth of the sequences (*i.e.* 100 sequences out of 488) are intronless. On the other hand, one sequence (TraesCS5B01G550800.1) contains 16 exons. The majority (214 sequences, 44%) contain three exons ([S2 Fig](#)). It is interesting to note that eight sequences, representing three homoeologous groups, contain two NAC domains (TraesCS2A01G363700, TraesCS2B01G381700, TraesCS2D01G361500; TraesCS3A01G269900, TraesCS3B01G303800, TraesCS3D01G269600; and TraesCS2A01G328100, TraesCS2D01G334800) ([S1 Table](#)).

Evolution by duplication of the *Ta*NAC family

The evolution of the *Ta*NAC family was evaluated by studying the phylogenetic relationships between *Ta*NAC genes. A phylogenetic tree based on multiple sequence alignment of the common set including 361 full-length proteins was constructed ([Fig 1](#)), using the Maximum Likelihood method.

The phylogenetic tree showed that, in general, sequences contained in the clades belong to the same homoeologous groups. Based on sequences alignment, 156 *Ta*NAC genes belonging to A, B, and D genomes were grouped into 52 complete homoeologous groups in the phylogenetic tree, representing 43% of the total *Ta*NAC genes. Another 56 incomplete homoeologous groups containing only two members were identified and represent 31% of the total *Ta*NAC genes. Finally, 93 sequences could not be linked to any homoeologous group. It is interesting

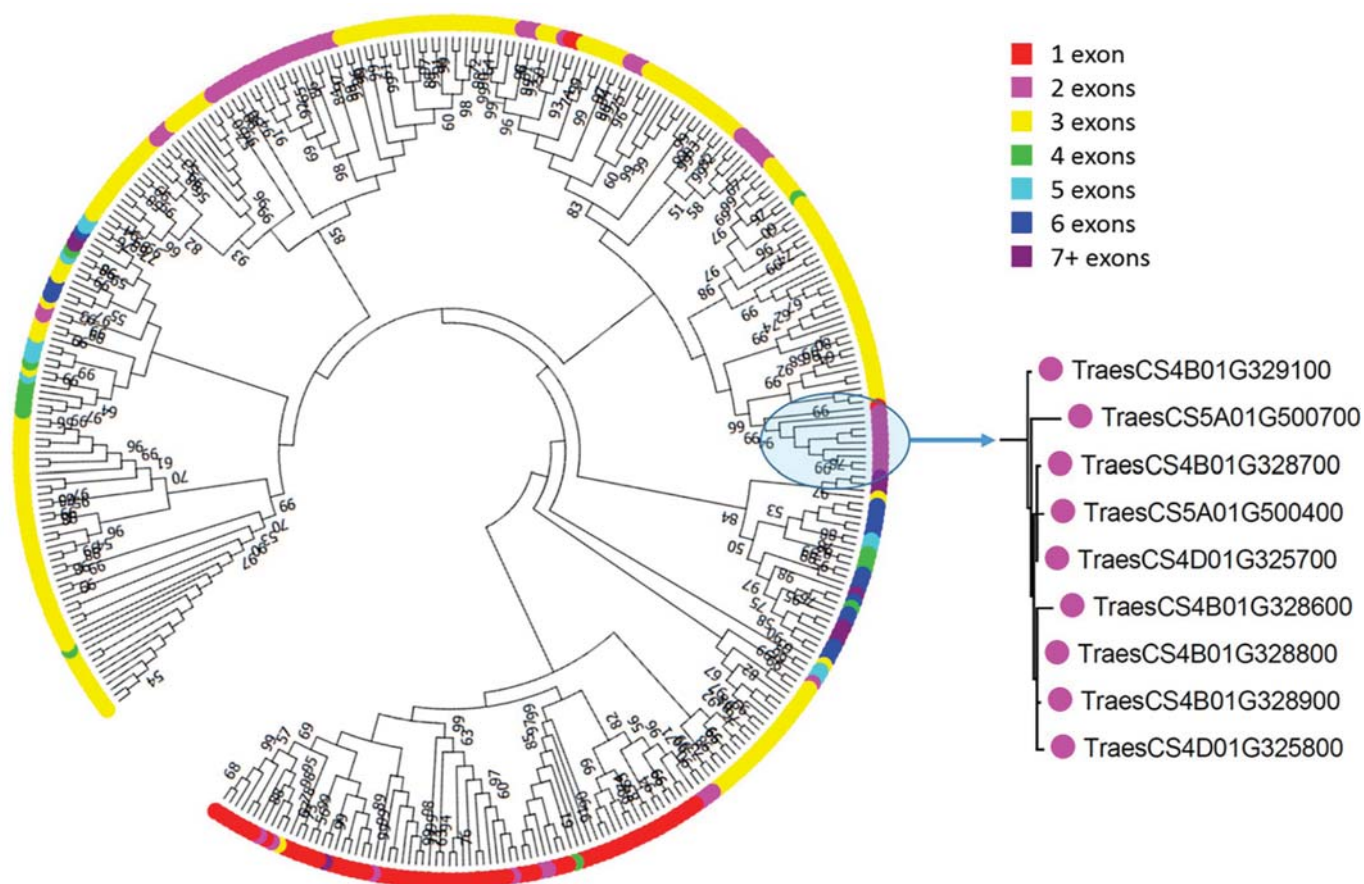


Fig 1. Maximum likelihood phylogeny of the 361 common complete protein sequences. The zoom focused on the clade studied as a case of duplication. Colors indicate the number of exons in each sequence.

<https://doi.org/10.1371/journal.pone.0213390.g001>

to note that several sequences with the same exon-intron structuration and clustered on the wheat genome are grouped together in the phylogenetic tree.

In addition, these clades contain genes originating from non-homoeologous chromosomes. These observations suggest that the expansion of the TaNAC family might be the result of both local and interchromosomal duplication events. To test this hypothesis, we retrieved the genomic sequences of 50 Kb loci spanning *TaNAC* genes and carried out pairwise alignments using the YASS algorithm [54].

For the purpose of simplicity, we first considered duplications within each subgenome. Fragments displaying at least 90% identity over at least 3 Kb were retained. Using these criteria, 318 duplicated fragments were identified within the subgenome A, with mean-length = 5.75 Kb and median-length = 5.255 Kb. The largest fragment was 13.093 Kb. Concerning the subgenome B, 207 duplicated fragments were detected, with mean-length = 5.546 Kb and median-length = 5.146. The largest duplicated fragment was 10.009 Kb. Concerning the subgenome D, 428 duplicated fragments were detected, with mean-length = 5.639 Kb and median-length = 5.332. The largest duplicated fragment was 9.947 Kb (Fig 2, S4 Table).

To complete that first analysis and provide also insight into duplications between subgenomes and consequently into the expansion pattern of the TaNAC family, we carefully surveyed one robust clade containing *TaNAC* genes originating from each wheat subgenome.

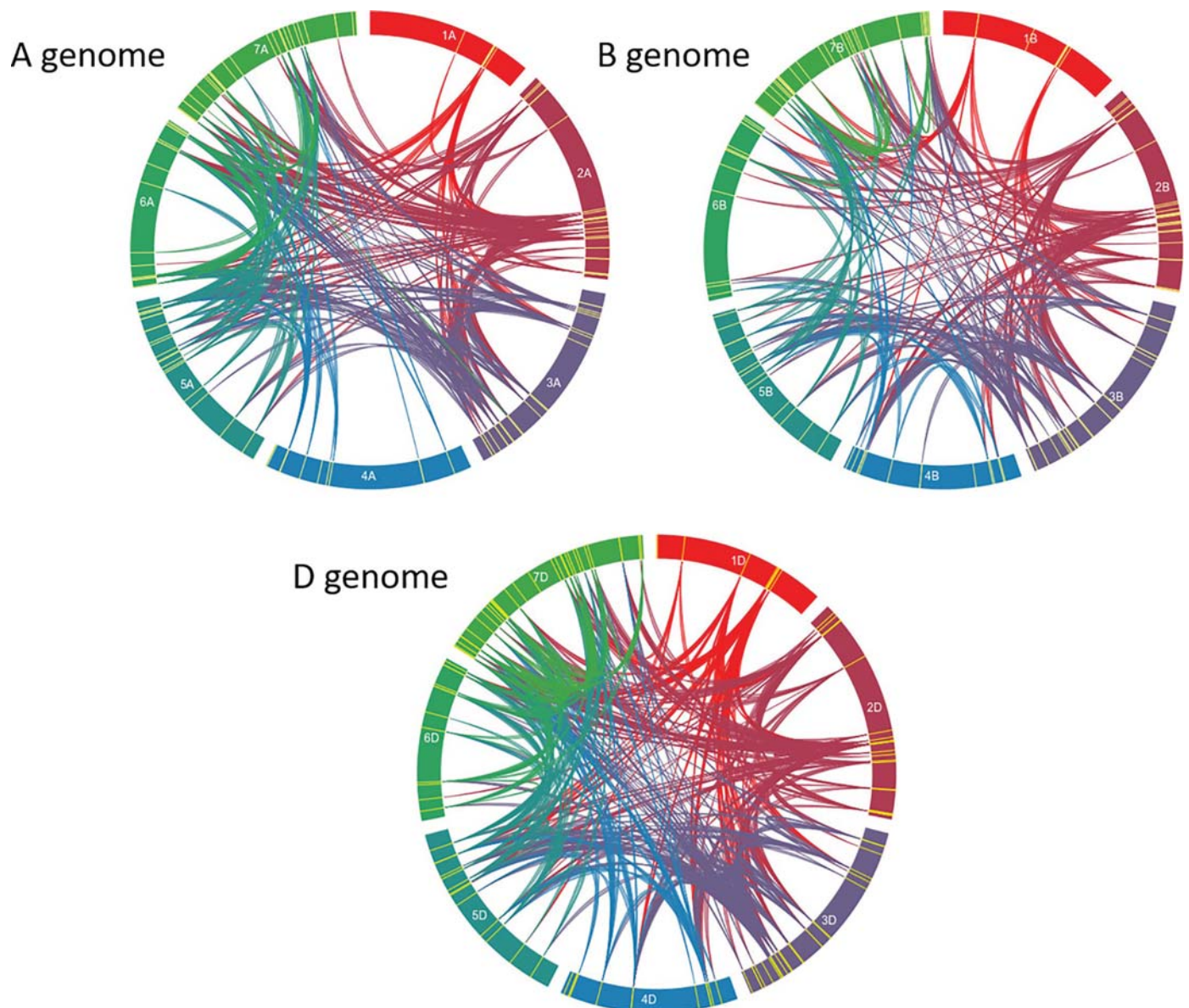


Fig 2. Circos diagram of the seven chromosomes of genomes A, B, and D. Yellow features indicate the presence of a TaNAC. Bands link two chromosomal regions with more than 90% of homology and a minimal length of 3 Kb.

<https://doi.org/10.1371/journal.pone.0213390.g002>

This clade (pointed out in Fig 1 and boxed in blue in S3 Fig) was chosen because it contained several *TaNAC* belonging to A (2 sequences from chromosome 5), B (5 sequences from chromosome 4), and D genomes (2 sequences from chromosome 4). Careful examination of regions containing the nine genes of this clade indicated that *TaNAC* genes from a given chromosome are physically tightly linked (separated by maximum one gene). To avoid homoeology relationships between subgenomes, which may obscure duplications, the subgenome 4D containing two *TaNAC* genes was not further considered. By using a 25 Kb window around the most distal *TaNAC* genes within each locus, we retained a first 396.990 Kb long (619563787 to 619960777 bp) locus from the chromosome 4B containing the five *TaNAC*

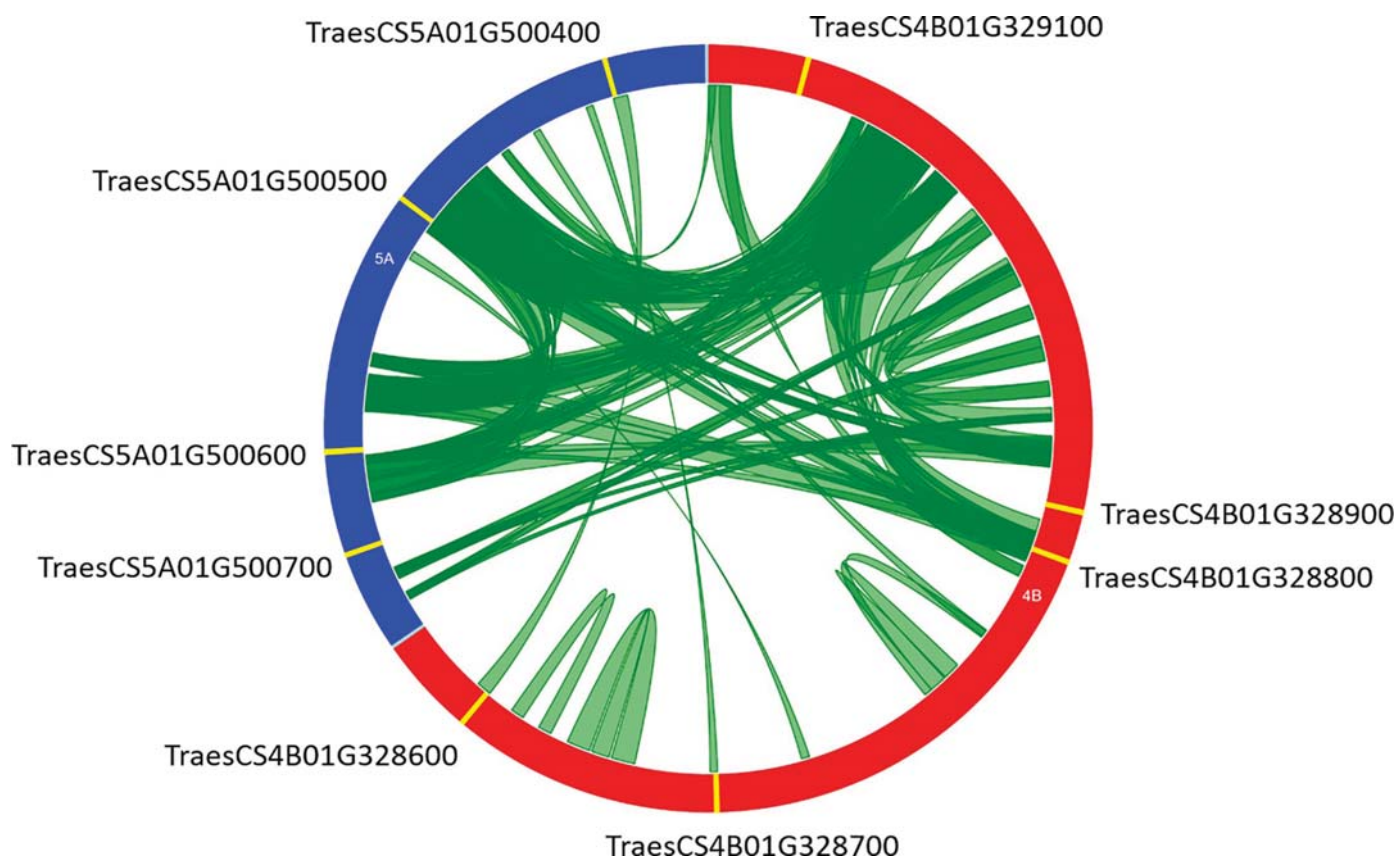


Fig 3. Circos diagram of chromosomes 4B (red, 396.990 Kb) and 5A (blue, 209.278 Kb) genomic regions containing nine *TaNAC* genes (yellow features). Green bands link two chromosomal regions with more than 80% of homology and a minimal length of 2 Kb.

<https://doi.org/10.1371/journal.pone.0213390.g003>

genes and a second 209.278 Kb long (666499425 to 666708703 bp) locus from the 5A containing four *TaNAC* genes. Over the four NAC sequences found on this chromosome, only TraesCS5A01G500400 and TraesCS5A01G500700 genes were present in the set of 361 sequences and studied for duplication. A dot plot of these two genomic regions (S4 Fig) highlighted numerous duplicated fragments in sense and antisense orientations. To clarify the nature and the extent of duplications, we identified all the duplicated fragments with a minimal length of 2 Kb and at least 80% identity. A total of 119 non-overlapping duplicated fragments were identified, of which 54 fragments correspond to intrachromosomal duplications (31 on chromosome 4B and 23 on 5A). The remaining 65 duplicated fragments correspond to interchromosomal duplications. The majority of these duplications (66.7%) represents a length less than 4 Kb, but a fragment has the maximum size of 9,006 bp. Duplications between loci of these genomic regions are presented in Fig 3. About 41% of these fragments (49) are duplicated in the sense orientation (S5 Table).

To test whether some duplicated genes could evolve differently, we compared the normalized expression patterns of duplicated genes in five wheat organs during three developmental stages: root (harvested at seedling, three leaves and meiosis), stem (harvested at spike at 1 cm, two nodes and anthesis), leaf (seedling, three tillers avec 50°Cd after anthesis), spike (harvested at two nodes, meiosis and anthesis), and grain (harvested at 50°Cd, 350°Cd and 700°Cd after anthesis [55]). In parallel, the expression profiles of the same genes were compared in response to abiotic stresses (heat, drought, and combined heat-drought) [56]. Only genes that were

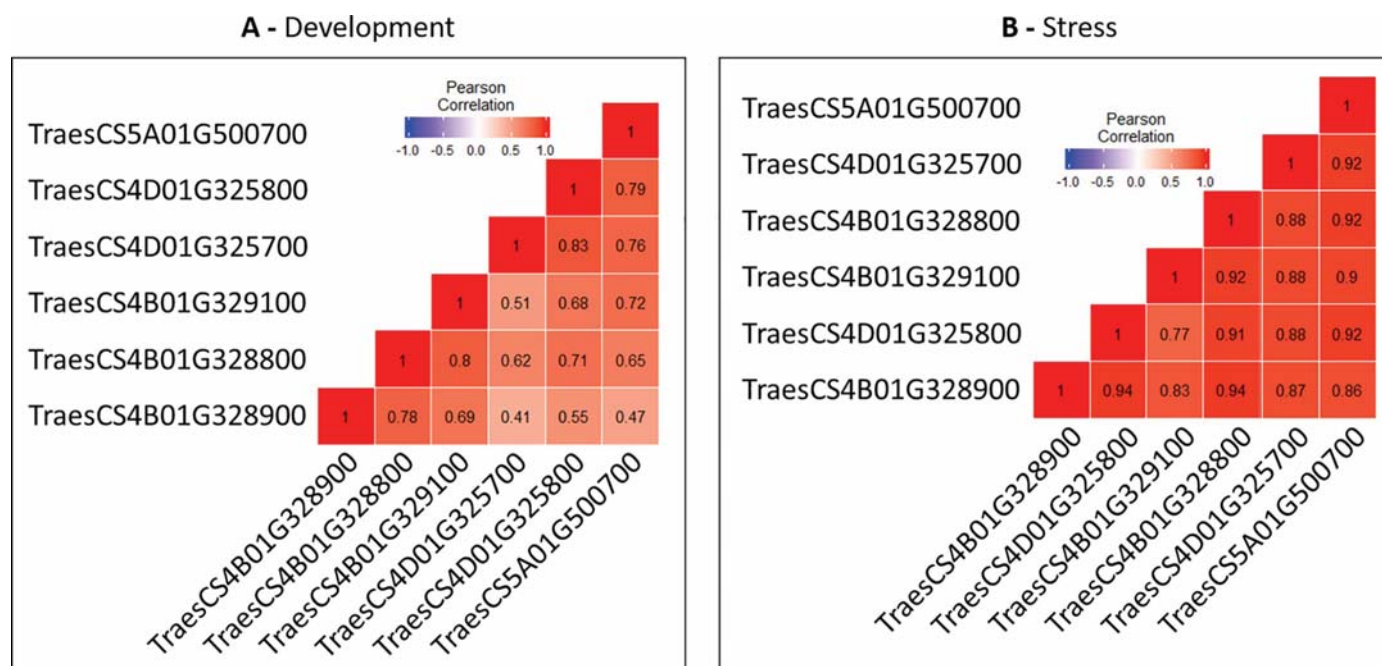


Fig 4. Pearson correlations between duplicated genes of the studied 4B-4D-7B-5A chromosomal regions during development (A) and stress (B).

<https://doi.org/10.1371/journal.pone.0213390.g004>

expressed in both studies were retained for comparison. The correlation (Pearson coefficient) was computed between pairs of genes (Fig 4).

During development, the Pearson coefficient ranged from 0.41 (between TraesCS4D01G325700 and TraesCS4B01G328900) to 0.83 (between TraesCS4D01G325800 and TraesCS4D01G325700). In contrast, correlations were higher during responses to stress and ranged from 0.77 to 0.94. These results indicate that the members of this duplication display similar expression profiles in response to the abiotic stresses tested but a substantial variation in expression profiles is observed during development between the same members as shown by the behavior of the two paralogs TraesCS4D01G325700 and TraesCS4B01G328900 (Fig 4).

In a similar way, an examination of a second duplication (boxed in red in the phylogenetic tree provided in S3 Fig) was performed. In this case, the expression profiles of the paralogs were highly correlated during development while they could be weakly correlated during responses to abiotic stresses (*i.e.* between TraesCS2B01G118500 and TraesCS2A01G101100) (S5 Fig).

***In silico* analysis of the expression of *Ta*NAC genes during development or in response to abiotic stresses**

In order to monitor the expression of wheat *Ta*NAC genes during development and in response to several abiotic stresses, we took advantage of valuable resources publicly available. In particular, we used a compendium of RNA-seq reads corresponding to various experiments and mapped to the wheat genome by Borrill *et al.* [45]. Raw counts corresponding to five organs at three developmental stages [55] or to abiotic stress treatments [56] were downloaded, normalized using TMM and voom methods, then Differentially Expressed Genes (DEG) were identified using an absolute logFC = 1 and an adjusted p = 0.05 as thresholds (R package limma) [57].

Table 1. Number of differentially expressed genes in response to two abiotic stresses.

	Heat stress (40°C) 1h	Heat stress (40°C) 6h	Drought 1h	Drought 6h	Heat + Drought 1h	Heat + Drought 6h
Number of down-regulated genes	6763	11420	339	6426	7018	11288
Number of down-regulated <i>Ta</i> NAC genes	24	39	5	28	28	41
Number of up-regulated genes	9837	11195	3117	6352	11796	11629
Number of up-regulated <i>Ta</i> NAC genes	43	31	40	54	43	43

<https://doi.org/10.1371/journal.pone.0213390.t001>

Responses of *Ta*NAC genes to abiotic stresses. We identified differentially expressed wheat genes in one or several comparisons. The Table 1 shows the details of this analysis where DEG are denoted up- or down-regulated genes in comparison with the non-treated control sample. Among the whole set of DEG genes, 162 were annotated as *Ta*NAC genes and dispatched as follows: 109 unique *Ta*NAC genes were identified in heat treatments, 100 in drought treatments and 114 in heat-and-drought treatments. Some of them are common in two or more conditions (Fig 5). Table 1 and S6 Table show the distribution of these differentially expressed *Ta*NAC genes alongside their logFC and the associated adjusted p-value.

Expression of *Ta*NAC genes during development. The *Ta*NAC genes have been shown to be involved in plant development with expression profiles that can vary between organs and

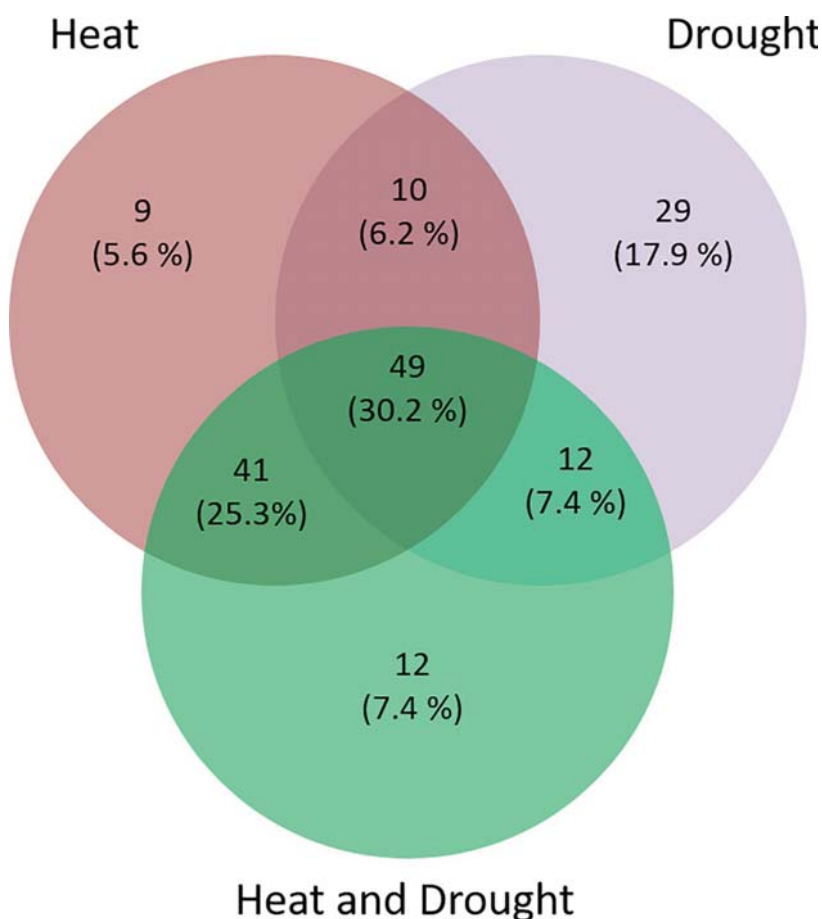


Fig 5. Venn diagram of the 162 differentially expressed *Ta*NAC genes in response to heat, drought, and heat + drought stress.

<https://doi.org/10.1371/journal.pone.0213390.g005>

during organ development. In order to detect preferentially expressed *TaNAC* genes in certain wheat organs and at certain stages, we used the RNA-seq data of Choulet *et al.* [55], comprising five organs and three developmental stages per organ. We then also selected the most differentially expressed *TaNAC* genes compared to other tissues using a logFC threshold of +12. Using these relatively stringent criteria, we identified 52 *TaNAC* genes with a highly preferred expression pattern distributed as follows: 19 in the grain, 7 in the leaf, 20 in the root, 6 in the spike and none in the stem (Table 2, Fig 6).

***TaNAC* expression analysis in field under drought stress**

Grain yield and spike number per m² were used to characterize the effect of the water stress generated in the drought trial. Under non-irrigated treatment, yield losses of 45.2% and 45.8% were observed for Altigo and Allez-y genotypes respectively. In the same way, a decrease of the spike number per m² was observed (-26.6% for Altigo and -39.1% for Allez-y) (Fig 7), which translates to a high level of water stress that affected plant growth [60].

***TaNAC* gene expression profile under irrigated conditions.** The *in silico* analysis resulted in a set of *TaNAC* genes tagged as DEG. A subset of 23 *TaNAC* displaying high fold changes in their expression were selected and analyzed in plants grown in the field under agronomic conditions. Their expression was monitored in two organs (leaf and grain) at two stages representative of grain development: cell division (220°Cd) and grain filling (450°Cd) in Altigo and Allez-y.

The 23 *TaNAC* genes tested in irrigated fields exhibited statistically differential transcript levels between organs (leaves and grains), regardless of the genotype or the stage. Seven genes appeared to be preferentially expressed in the grain and 16 in the leaf. The gene TraesCS5A01G245900 showed the highest expression in the grain compared to the leaf (10 018 times more) whereas TraesCS7D01G315100 was the most expressed gene in the leaf (332 times more than in the grain) (Fig 8A and 8B).

In addition, some *TaNAC* gene expressions varied depending on the developmental stage (220°Cd vs 450°Cd). One gene (TraesCS2A01G566300.1) appeared to be preferentially expressed at 220°Cd while 15 genes were preferentially expressed at 450°Cd. For example, the expression level of TraesCS2A01G101100 gene was found to be 4.19 times more in leaves at 450°Cd compared to 220°Cd (Fig 8C).

Expression of *TaNAC* genes under drought stress. The expression data of the 23 selected genes in drought field conditions were presented in a principal component analysis (S6 Fig). The first two Principal Components (PC1 and PC2) explained respectively 83.5% and 10.8% of the total variation. Cumulatively, these two PCs contributed to 94.3% of the total variation of gene expression in response to drought, organ, genotype, and stage conditions. The first axis is explained by gene expression in the leaves under drought conditions at the later stage, 450°Cd (S6 Fig). The second axis is mainly explained by an organ x stage effect. Indeed, grain samples at 450°Cd were plotted in the lowest part of the S6 Fig. In the grain subset, the samples corresponding to early and late grain developmental stages (full form– 220°Cd and empty form– 450°Cd) constitute distinct groups along the second axis (PC2). In addition, considering only the data from grain at 450°Cd, it can be noticed that the two genotypes are plotted into two different groups.

In leaves, the highest difference under drought stress was observed at 450°Cd compared to 220°Cd, regardless of the genotype (Fig 9A). A genotype effect was observed in response to drought only in grains compared to leaves. More precisely, grain-specific genes were differentially expressed at a different stage depending on the genotype. Indeed, a down-regulation under drought treatment was observed at 220°Cd in Altigo for five grain-specific genes. In

Table 2. Number of differentially expressed genes (A) and list of preferentially expressed *Ta*NAC genes (B) in five wheat organs.

A	Grain	Leaf	Spike	Stem	Root
Number of down-regulated genes	6505	5815	6533	6356	6362
Number of up-regulated genes	6862	5508	4468	5117	8447
B	Organ with a preferential expression of the gene			logFC	Adjusted P.Value
Transcript					
TraesCS2D01G576500.2	Leaf			18.178	0.0009
TraesCS4B01G329100.1	Leaf			17.573	0.0121
TraesCS2D01G101300.1	Leaf			17.472	0.0231
TraesCS4B01G328900.1	Leaf			17.299	0.0298
TraesCS7A01G318500.1	Leaf			16.701	0.0074
TraesCS7D01G315100.1	Leaf			16.4	0.0356
TraesCS2A01G566000.1	Leaf			16.21	0.0377
TraesCS7B01G489500.2	Grain			24.77	0.0026
TraesCS7D01G543500.2	Grain			24.394	0.007
TraesCS7A01G569100.1	Grain			24.269	0.0038
TraesCS2A01G326600.1	Grain			24.025	0
TraesCS7B01G100300.1	Grain			23.802	0.0025
TraesCS7A01G189200.1	Grain			23.73	0.0002
TraesCS7B01G056300.1	Grain			23.723	0.005
TraesCS7D01G154200.1	Grain			23.719	0.007
TraesCS7A01G194700.1	Grain			23.663	0.0059
TraesCS7D01G196300.1	Grain			23.432	0.0031
TraesCS7B01G094000.1	Grain			23.334	0.0005
TraesCS7A01G152400.1	Grain			23.258	0.0009
TraesCS3B01G092800.1	Grain			23.077	0.0007
TraesCS4B01G132500.1	Grain			22.886	0
TraesCS7D01G543400.1	Grain			22.879	0.0004
TraesCS5A01G245900.2	Grain			22.67	0.0147
TraesCS7A01G569300.1	Grain			21.729	0.0011
TraesCS3A01G486500.1	Grain			20.566	0
TraesCS3A01G077900.1	Grain			20.52	0.0329
TraesCS2D01G309800.1	Root			23.596	0
TraesCS2A01G462500.1	Root			23.17	0
TraesCS2A01G311500.1	Root			22.266	0
TraesCS1D01G275500.1	Root			21.745	0
TraesCS2D01G462900.1	Root			21.084	0.0002
TraesCS6A01G307200.1	Root			20.627	0
TraesCS2B01G328300.1	Root			20.483	0
TraesCS2B01G484100.1	Root			20.365	0.0002
TraesCS6D01G286300.1	Root			20.325	0
TraesCS2D01G698900LC.1	Root			19.923	0
TraesCS7D01G329200.1	Root			19.261	0.0007
TraesCS7B01G233300.1	Root			19.259	0.0001
TraesCS7A01G349500.1	Root			19.001	0.0009
TraesCS1A01G275900.1	Root			18.528	0.0001
TraesCS4A01G242700	Root			18.32	0.0008
TraesCS4B01G072400.1	Root			18.291	0.0003
TraesCS5B01G275200.1	Root			18.04	0.0021

(Continued)

Table 2. (Continued)

A	Grain	Leaf	Spike	Stem	Root
Number of down-regulated genes	6505	5815	6533	6356	6362
Number of up-regulated genes	6862	5508	4468	5117	8447
B	Organ with a preferential expression of the gene		logFC	Adjusted P.Value	
TraesCS5D01G283200.1	Root		17.985	0.0045	
TraesCS2D01G382800.1	Root		17.401	0.0022	
TraesCS4D01G071200.1	Root		16.877	0.0136	
TraesCS7D01G371800.1	Spike		19.526	0.0002	
TraesCS7B01G143900.1	Spike		18.488	0.0025	
TraesCS7A01G334800.1	Spike		18.372	0.0042	
TraesCS7B01G246300.1	Spike		18.219	0.0053	
TraesCS2B01G582900LC.1	Spike		17.862	0.005	
TraesCS7D01G342300.1	Spike		17.198	0.0151	

A *TaNAC* gene is preferentially expressed in a tissue if i) its logFC is the highest in this tissue vs the others and ii) the logFC threshold is ≥ 12 .

<https://doi.org/10.1371/journal.pone.0213390.t002>

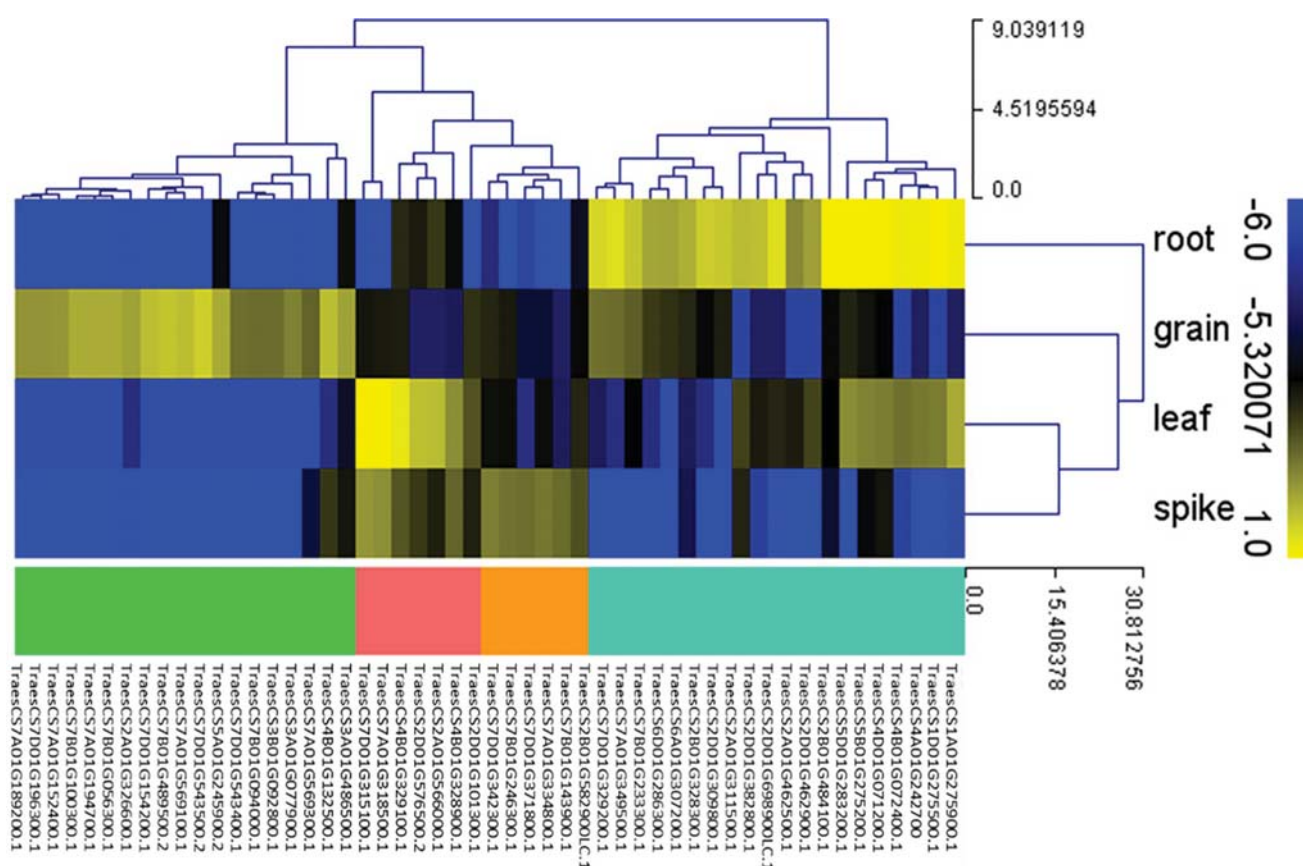


Fig 6. Heat map of 52 *TaNAC* genes expression profile in four organs (grain, spike, leaf, and root), using the RNAseq data of Choulet *et al.* [55]. These 52 *TaNAC* genes have a highly preferred expression pattern distributed as follows: 19 in the grain (green branches), 7 in the leaf (red branches), 20 in the root (blue branches), and 6 in the spike (orange branches). This figure was generated using MeV software. Hierarchical clustering was performed for the transcript abundance from all conditions.

<https://doi.org/10.1371/journal.pone.0213390.g006>

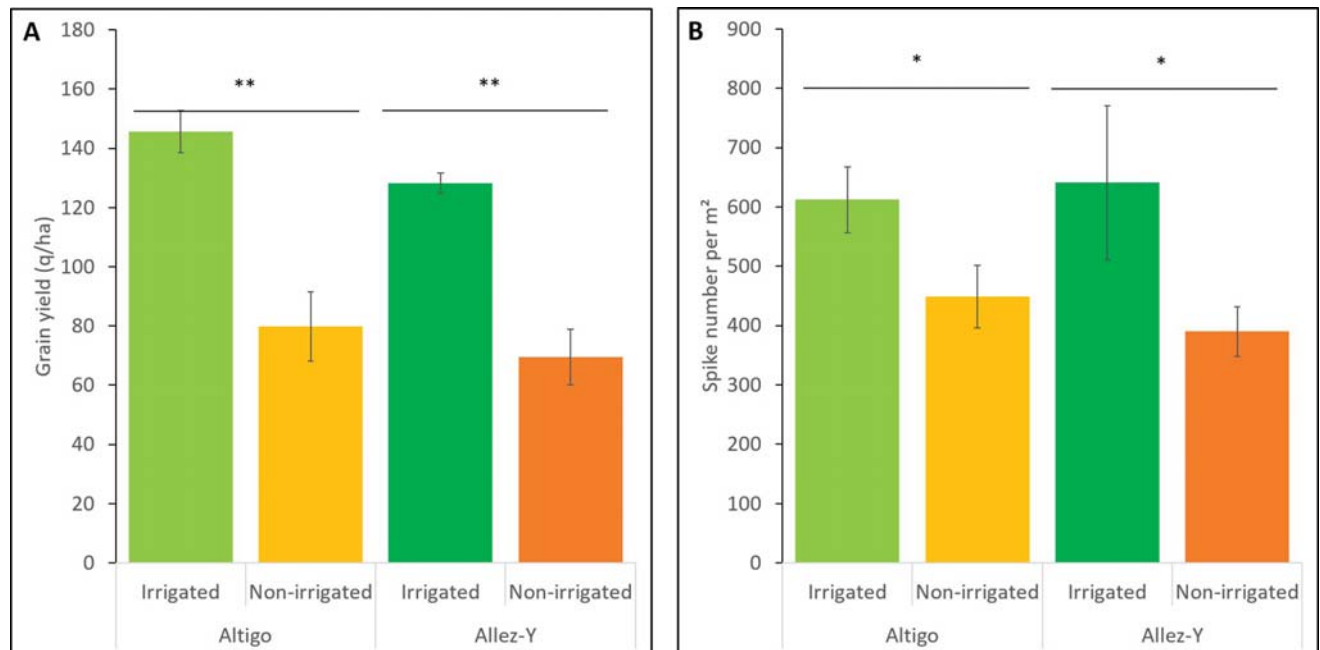


Fig 7. Drought effect on the grain yield (q/ha) (A) and spike number per m² (B) of two genotypes (Altigo and Allez-y). Mean values and standard deviation were obtained from three replicates (Student's T test, * indicates a P < 0.05 and ** indicates a P < 0.01).

<https://doi.org/10.1371/journal.pone.0213390.g007>

Allez-y, four of these transcripts (TraesCS3A01G077900.1, TraesCS5A01G245900.2, TraesCS7B01G056300.1 and TraesCS7B01G100300.1) were down-regulated and the last one was up-regulated (TraesCS7A01G569100.1) only at 450°Cd (Fig 9B, Fig 9C).

Discussion

The sequence availability of the whole bread wheat genome and expression data among different wheat genotypes, organs, and conditions are rich resources for systematic analysis of gene transcriptional regulation. Particularly, these data provide useful information for the identification of candidate genes involved in wheat development and its stress response.

Among transcription factors, the NAC family is one of the largest families and they are studied in a plethora of cereals [18,20,21] particularly in response to abiotic stresses [42,64,65], but only scarce information is available on bread wheat.

However, valuable resources and studies have recently been released. Particularly, Borrill *et al.* [45] identified 453 high confidence *TaNAC* retrieved from TGAC database [48]. In addition, large RNA-seq expression data exist on several wheat genotypes, in different organs [55] and abiotic stresses (heat shock or drought) [56] and are available to the scientific community. Based on these data, we studied the organisation of the NAC family in bread wheat and its expansion that could have occurred during evolution through potential duplication events. These resources were also used to identify differentially expressed *TaNAC* genes in wheat grown under drought conditions in field and to validate the potential specialization of some *TaNAC* members.

TaNAC genes: A large family

First, we identified 488 members of the NAC family in bread wheat using the IWGS RefSeq v1.0 database [47], whereas only 117 members were identified in Arabidopsis, 151 in rice [17]

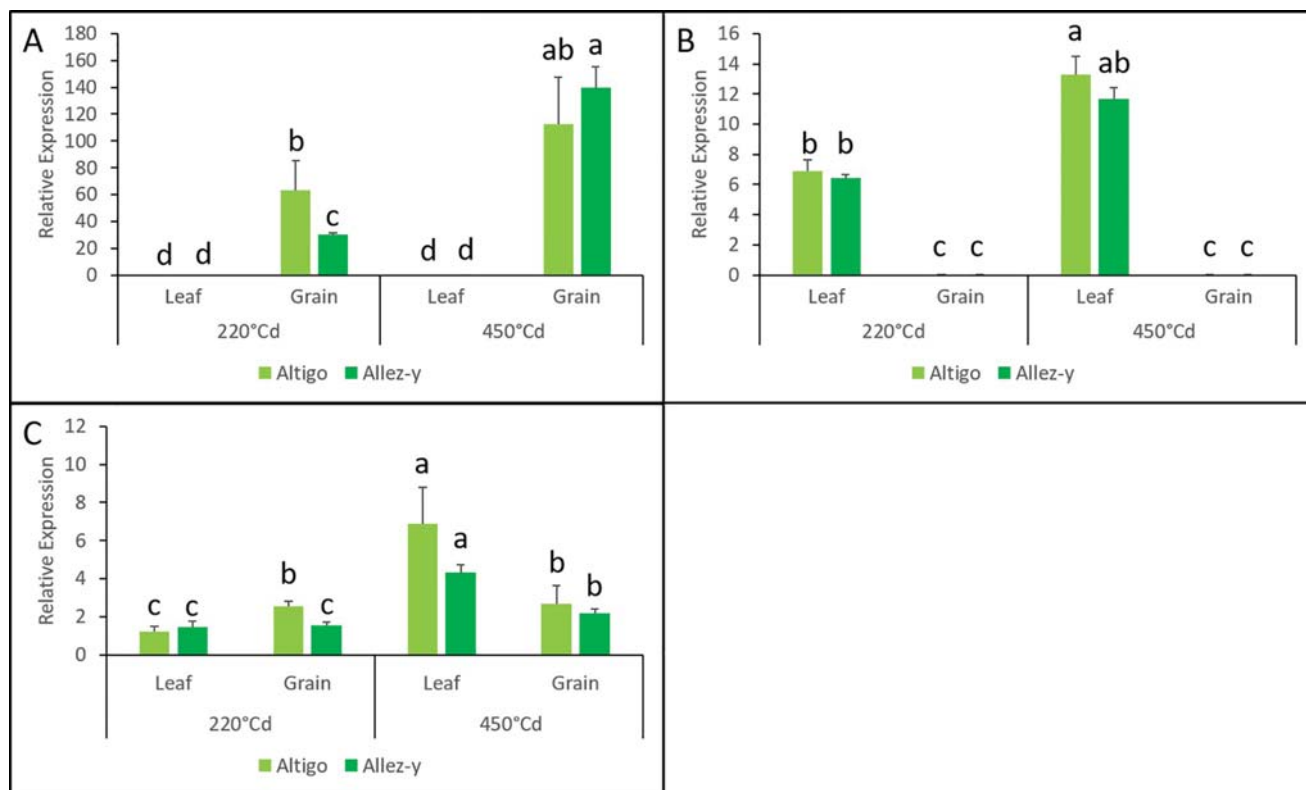


Fig 8. RT-qPCR analysis of three *TaNAC* genes (TraesCS5A01G245900.2 (A), TraesCS7D01G315100.1 (B), and TraesCS2A01G101100.1 (C)) in leaves and grains at 220°Cd and 450°Cd in irrigated conditions. Mean values and standard deviation were obtained from three replicates. According to a Kruskal-Wallis test, the same letter indicates no significant difference.

<https://doi.org/10.1371/journal.pone.0213390.g008>

or 168 in durum wheat [21]. This high number is due to the presence of three subgenomes in bread wheat. After improving our dataset with TGAC sequences retrieved by Borrill *et al.* [45], we ended up with 361 sequences representing a high quality consensus subset. Considering the hexaploidy of wheat, one copy of each gene is expected to be carried by each of the three genomes (A, B, and D) creating homoeologous groups among the consensus subset. In this family, 43% of the *TaNAC* members can be gathered in complete homoeologous groups, 31% form incomplete homoeologous groups and 26% are not linked to a group. A divergence of sequences due to several mutations or a lack of sequences present in the subset could explain the incomplete homoeologous groups. This result has already been observed in other TF families, such as the bZIP family [66].

Similar to what has been described in other plant species such as maize [67] and rice [17], the *TaNAC* gene distribution among the chromosomes was uneven. On the contrary, the three subgenomes carry the same proportion of *TaNAC* genes, which is consistent with the distribution observed in durum wheat, a diploid species [21]. This result suggests that *TaNAC* family expansion was likely prior to a polyploidization event in bread wheat.

Expansion of the wheat *TaNAC* genes family due to small-scale duplications

Several recent studies have proposed different mechanisms to explain the duplication events and gene expansion in plants. These proposals included, for instance, retrotransposition, duplicated DNA transposition, unequal crossing-over, and polyploidization [68–70]. In

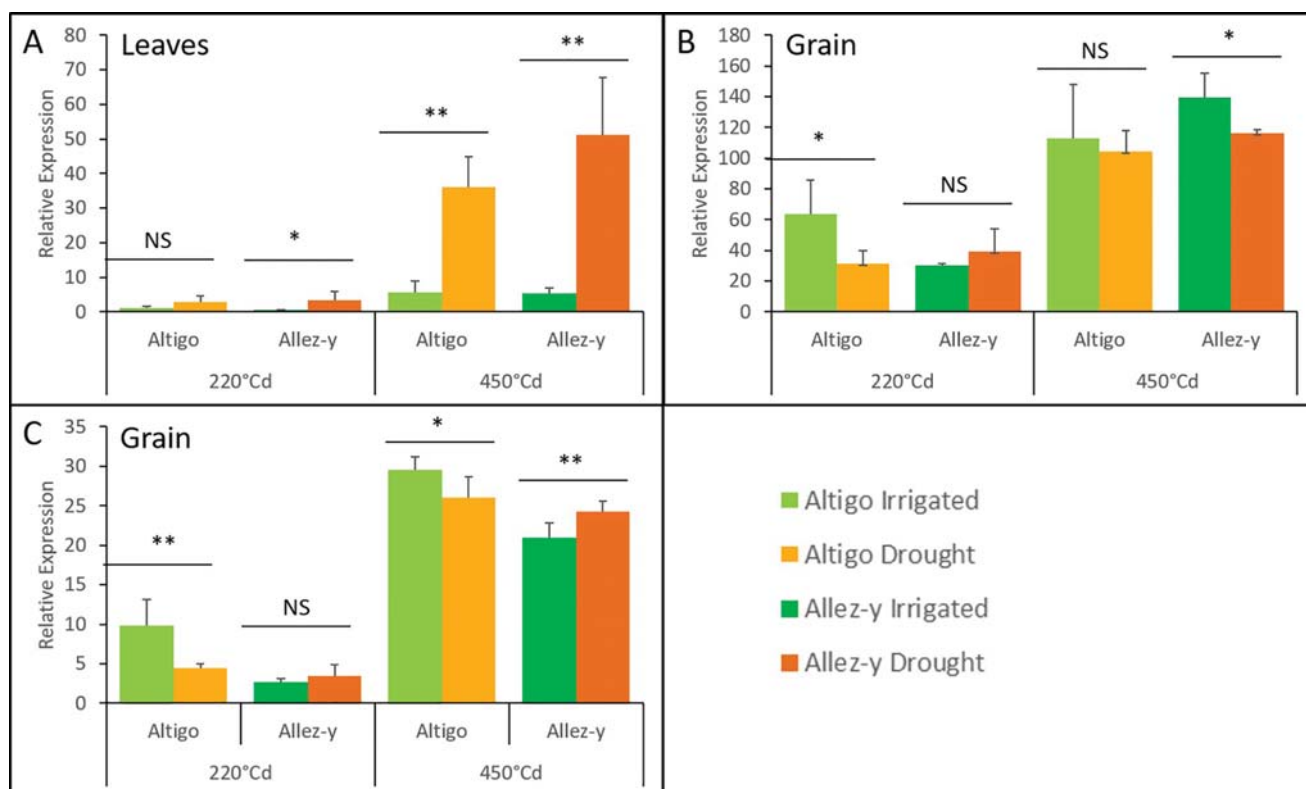


Fig 9. RT-qPCR expression analysis of *TraesCS5A01G143200.1* (A), *TraesCS5A01G245900.2* (B), and *TraesCS7A01G569100.1* (C). The $2^{-\Delta\Delta C_t}$ method was used to calculate the ratio of their relative expression levels in drought conditions with the one in irrigated conditions. Mean values and standard deviation were obtained from three replicates (Student's T test, * indicates a $P < 0.05$ and ** indicates a $P < 0.01$).

<https://doi.org/10.1371/journal.pone.0213390.g009>

addition to these mechanisms, Wicker *et al.* [71] then Glover *et al.* [72] proposed another mechanism leading to gene movement and expansion. In particular, Glover *et al.* [72] showed that massive small-scale interchromosomal gene duplications occurred in the bread wheat genome after double-strand DNA break repair. Consequently, many non-syntenic genes have high sequence similarity. Regardless of the mechanisms involved in gene duplication, this may eventually lead to a pseudogenization, a conservation of the gene function, a subfunctionalization or a neofunctionalization of duplicate genes [73].

In this study, we noticed that several clades in the phylogenetic tree contained sequences that were either physically tightly linked on the same chromosome or mapped to a non-homologous group, suggesting that both intra and inter-chromosomal duplication events occurred. To test the possibility of small-scale duplications, we focused on three genomic regions on the chromosomes 4B, 4D, and 5A that encompass 11 *TaNAC* genes, which display highly similar sequences and similar genomic structures (2 exons). By aligning the sequences at the vicinity of these genes, we uncovered a high homology between the genomic regions (Fig 3). This finding support the hypothesis evoked by Glover *et al.* [72] about the expansion of the multi-genes family that is partially due to small-scale duplication driven by repetitive sequences.

Expansion of the wheat *TaNAC* genes family due to retroposition

Of the 361 high-quality sequences, 68 *TaNAC* genes (19%) are intronless, which is a characteristic of retroposons [74]. The subfamily NAC-h includes 65 of these 68 genes that are grouped

in the same clade of the phylogenetic tree. However, in addition to an absence of intron, the presence of a poly(A) tract at the 3' end and a TSD (Target Site Duplication) in a gene sequence constitutes a hallmark of a retroposition event [74]. However, we failed to identify any of these signatures in these sequences. These signatures were possibly obscured by multiple mutations during evolution to such extent that they are no longer recognisable. Furthermore, several intronless *Ta*NAC genes of this clade are grouped into homoeologous groups suggesting that retroposition/duplication events probably occurred prior to wheat polyploidization.

Interestingly, the distribution of *Ta*NAC intronless genes among chromosomes was found not to be equivalent to the one observed for the whole *Ta*NAC family. For example, although the gene proportions of the whole *Ta*NAC family on the chromosome 3 and 5 are close (13 and 16% respectively), a higher proportion of intronless *Ta*NAC genes was noticed on chromosome 3 (29% of the total intronless genes) whereas no intronless genes were observed on chromosome 5.

Taken together, these results indicate that both small-scale duplication and retroposition phenomena led to the expansion of this family of genes. Furthermore, these mechanisms could act in combination, *i.e.* a gene retroposition could be followed by small-scale duplications driven by repetitive sequences at the very vicinity of the gene.

For example, TraesCS2A01G328100, TraesCS2A01G328200, TraesCS2A01G328300 and TraesCS2A01G328500 are intronless and tightly linked genes, and could be an example of the retroposition/duplication events that could have occurred during evolution.

Expression of *Ta*NAC genes under drought field conditions

As a major widely spread environmental stress, drought is known to significantly affect wheat yields worldwide. Interactions between plant development and stress response are complex, especially in the field, where crops are submitted to a combination of environmental variables (such as water availability and temperature). Although a few NAC TFs have been studied in wheat compared with the model plants *Arabidopsis* and rice, it has been shown that they may also be involved in response to environmental conditions [5]. Some *Ta*NAC genes have already been studied in bread wheat. For example, *Arabidopsis* overexpressing *Ta*NAC2, *Ta*NAC29, or *Ta*NAC67 display improved tolerance to multiple abiotic stresses [16,42,75]. In the same way, *Ta*NAC2a, *Ta*NAC2D, *Ta*NAC4a, *Ta*NAC6, *Ta*NAC7, *Ta*NAC13, and *Ta*NTL5 were induced by water stress, high salinity, and/or low temperature, at different intensities [76,77]. Recently, *Ta*RNAC1 was found to enhance drought tolerance leading to higher biomass, grain yield, and root length [78]. In addition, transgenic wheat overexpressing *Ta*NAC47 showed a drought-tolerant phenotype that may be induced in response to ABA [79].

Here, we have studied 23 *Ta*NAC genes that we identified as the most up-regulated (p -value < 0.05) in specific organs (leaf or grain) or in response to treatments (drought and heat) (S7 Table). Among the 23 selected genes, 20 belong to the subfamilies NAC-a to d, which have been described as Stress Related NAC containing subfamilies [80] (S3 Table).

Some *Ta*NAC genes exhibit an extreme level of transcript abundance specifically depending on the organ. NAC genes are recognized to be largely involved in many developmental and growth processes, such as the differentiation of major organs [81]. Our study highlighted genes that are preferentially expressed in the leaf (16) or in the grain (7) regardless of the genotype, the grain developmental stage, or the field conditions.

Interestingly, of those genes, seven leaf-specific and five grain-specific genes were found, with an expression level at least 10 times higher than in the other organ (S7 Table), suggesting their specific implication in the leaf or grain development. For example, TraesCS7A01G305200.1 gene up-regulated in leaves either in the earliest grain developmental stage (220°Cd) and the

grain filling stage (450°Cd), was found to be similar to TaNAC1/TaNACS already described by Zhao *et al.* [37]. In this work, the authors highlighted an increase in nitrogen concentration in wheat grains from lines over-expressing the TaNAC-S protein (TaNAC1) and a stay-green phenotype. These authors concluded that NAC TFs, expressed mainly in the leaves, may participate to delay post-anthesis foliar senescence (stay-green phenotype) through a better remobilization of the nutrients during senescence. Another gene responding mainly in leaf in our experiment (TraesCS4D01G176000.1) is homologous to a NAC gene in the rose, *RhNAC100*, that has been characterized as a repressor of cellular expansion of rose petals [82], suggesting its involvement in the determination of cell size.

Among the seven up-regulated TaNAC grain-specific genes, one gene (TraesCS7A01G569100.1) was found to be homologous to ONAC20 and ONAC026 described in rice [24]. In this latter study, the authors showed that three NAC genes (ONAC020, ONAC026 and ONAC023) are expressed specifically during seed development at a very high level, suggesting a potential role of these genes, *via* their heteromerization capability, in the regulation of the grain size/weight. Another gene, TraesCS3A01G077900.1, was found to be homologous to a barley *HvNAC019* gene which is preferentially expressed in grain, indicating that this gene could regulate seed development in association with two other genes (*HvNAC017* and *HvNAC018*) [83].

Leaf-specific TaNAC genes are up-regulated at the grain filling stage in response to drought. Sixteen genes were found differentially expressed in both leaves and grains between the early grain developmental stage (cell division) and the grain filling stage, and mainly up-regulated at 450°Cd (15 genes).

Drought stress during grain filling stimulates leaf senescence and enhances remobilization of water-soluble carbohydrates to the grains [84]. Interestingly, we found a set of eight TaNAC genes that were preferentially expressed in leaves during grain filling in response to drought. This finding may indicate that these genes could participate in the resource remobilization from the leaves to the grains. For example, the three members of the same homoeologous group, located on chromosome 5 (TraesCS5A01G143200.1 (Fig 9A), TraesCS5B01G142100.1 and TraesCS5D01G148800.1) were found similar to TaNAC69. This gene has been well-characterized as a drought stress regulator [85], and was shown to improve grain yield. However, its expression was found mainly in roots compared to leaves [65,85,86]. It is worthy to note that, in our study, the genome A copy-gene was expressed at a higher level than the two other homoeologous copies in leaves. This finding may suggest that the B and D copies could be related to the root-specific expression that has not been measured in this study. The homologous gene in rice (*OsNAC10*) has also been demonstrated to be involved in drought adaptation mainly by enhancing the level of stress-protection genes, such as chitinase, ZIM family gene, and glyoxalase I family gene; and increases grain yield significantly under field drought conditions [87]. Also, two members of the same homoeologous group (TraesCS4D01G176000, TraesCS4B01G174000) are homologs to the rice gene ONAC39 (*OSCUC1/ONAC039*) that has also been described by Jeong *et al.* [87] as a drought responsive gene (fold change: 9.44). The genes of this homoeologous group in wheat may have similar functions and indicates that these genes may be associated with drought stress.

Another wheat gene (TraesCS2A01G101100, Fig 8C) was up-regulated in leaves under drought at 450°Cd in our field conditions. In contrast, its homolog, TaGRAB2, which was involved in response to biotic stress (a plant DNA virus *Geminiviridae*), showed an expression pattern barely detectable in leaves [39]. These findings suggest that the same gene could be expressed or not in leaves, depending on the stress.

Finally, no characterized homolog has been found in the literature for TraesCS3B01G303800.1 and TraesCS6D01G390200.1. These genes may be interesting candidates for field drought tolerance and require further investigations.

Grain-specific *Ta*NAC genes respond differentially to drought according to the genotype. Five *Ta*NAC genes (TraesCS3A01G077900.1, TraesCS5A01G245900.2, TraesCS7A01G569100.1, TraesCS7B01G056300.1, and TraesCS7B01G100300.1) are grain-specific. Interestingly, their behavior in response to drought varies according to the genotype. They are down-regulated under drought treatment during the early grain developmental stage (cell division) in the genotype which exhibited the highest number of spikes/m² (Fig 7B) (Altigo); whereas they are down-regulated (4 genes) or up-regulated later during the grain filling stage (TraesCS7A01G569100.1), in the other genotype (Allez-y) (Fig 9B, Fig 9C). These results suggest that, depending on the genetic background, those genes may harbor different alleles that could affect their drought tolerance.

Four of these five genes (TraesCS7A01G569100.1, TraesCS5A01G245900.2, TraesCS7B01G056300.1, and TraesCS7B01G100300.1) have been found previously differentially expressed in grain in response to a moderate warming of +8°C applied during the grain cell division stage [88]. These results suggest that these four *Ta*NAC genes may be involved in response to close abiotic stresses (drought and high temperature), such as other NAC genes like *MINAC9*, *SINAC3*, *ONAC022*, *GmNAC3* and *GmNAC4*, that were described involved in several environmental stresses [19,89–91].

Conclusion

In conclusion, we performed a comprehensive study of NAC TFs in the bread wheat genome (*T. aestivum*), cultivar *Chinese Spring*. We identified a high quality set of 361 *Ta*NAC genes in common with those described in Borrill *et al.* [45]. We provided evidences as to how small scale duplications and retroposition contributed to the expansion of this family of genes. Finally, the expression patterns of 23 *Ta*NAC genes in two organs at two stages under drought in the field, combined with our study of duplication events, revealed a functional diversification of this gene family's members. Further functional characterizations of five grain specific *Ta*NAC genes are in process to identify their role in the wheat grain development and in abiotic stress responses.

Supporting information

S1 Fig. Chromosomal distribution of the *Ta*NAC gene family in bread wheat. Using their chromosomal location information, 451 sequences have been mapped on the seven chromosomes of the three wheat subgenomes A (red), B (green), and D (blue). The first seven letters of the sequence names have been removed (*e.g.* 4B01G328600 refers to TraesCS4B01G328600, located on the chromosome 4B according to the IWGSC annotation). The position of each *Ta*NAC can be estimated using the scale on the left. The two circles show the chromosomal regions studied as an example of duplication.
(TIF)

S2 Fig. Exon number in the genomic sequence of the 488 *Ta*NAC members. The percentage of sequences is indicated at the top of each histogram.
(TIF)

S3 Fig. Maximum likelihood phylogeny of the 361 consensus sequences using complete protein sequences. The blue and red boxes focused on the clades studied as case of duplication. The number of exons is represented by a red (1 exon), pink (2 exons), yellow (3 exons), green (4 exons), clear blue (5 exons), dark blue (6 exons), or violet (7 and more exons) circle.
(PDF)

S4 Fig. Structural conservation between the two regions of chromosomes 4B and 5A genomics sequences, revealed by dot plot analysis. Each gene location is indicated on the x- and y-axis by a blue arrow. Green and red dots indicate sense and antisense homologies respectively.

(TIF)

S5 Fig. Pearson correlations between duplicated genes of the studied 2A-2B-2D chromosomal regions during development (A) and stress (B).

(TIF)

S6 Fig. Biplot showing the first two variables of the principal component analysis (PC1 and PC2). Expression of the 23 *TaNAC* genes in leaves and grains of Altigo (square) and Allez-y (circle) genotypes at 220°Cd (full form) and 450°Cd (empty form) in irrigated (blue) and drought (red) conditions.

(TIF)

S1 Table. Blast results using Blast2Go software [92] on 488 *TaNAC* sequences.

(XLSX)

S2 Table. List of primers used in this study.

(XLSX)

S3 Table. *TaNAC* sequences and structural information extracted from the IWGS RefSeq v1.0 database and their TGAC correspondence.

(XLSX)

S4 Table. Duplicated fragments contained within the three genomes of the bread wheat, with a minimal length of 3 Kb and at least 90% of identity.

(XLSX)

S5 Table. Duplicated fragments contained between genomic regions of 4B and 5A chromosomes of the bread wheat, with a minimal length of 2 Kb and at least 80% of identity.

(XLSX)

S6 Table. Distribution of the differentially expressed *TaNAC* genes in response to heat and drought stresses, alongside their logFC and the associated adjusted p-value.

(XLSX)

S7 Table. Specific expression of 23 *TaNAC* genes studied by RT-qPCR and their orthologues in other species [24, 37, 39, 40, 65, 82, 83, 85, 87, 93–96]. The preferential expression have been verified by a Student Test, with an e-value threshold at 0.05.

(XLSX)

Acknowledgments

We would like to thank C. Dupuits and the Gentyane Plateform for their technical assistance.

Author Contributions

Conceptualization: Claire Guérin, Jane Roche, Said Mouzeyar, Mohamed Fouad Bouzidi.

Data curation: Jane Roche, Catherine Ravel.

Formal analysis: Claire Guérin, Jane Roche, Mohamed Fouad Bouzidi.

Methodology: Said Mouzeyar.

Resources: Vincent Allard.

Supervision: Mohamed Fouad Bouzidi.

Writing – original draft: Claire Guérin, Jane Roche, Vincent Allard, Said Mouzeyar, Mohamed Fouad Bouzidi.

Writing – review & editing: Claire Guérin, Catherine Ravel, Said Mouzeyar.

References

1. Harmer SL. The circadian system in higher plants. *Annu Rev Plant Biol.* 2009; 60:357–77. <https://doi.org/10.1146/annurev.arplant.043008.092054> PMID: 19575587
2. Khan S-A, Li M-Z, Wang S-M, Yin H-J. Revisiting the Role of Plant Transcription Factors in the Battle against Abiotic Stress. *Int J Mol Sci.* 31 mai 2018; 19(6).
3. Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc Natl Acad Sci USA.* 11 févr 2014; 111(6):2367–72. <https://doi.org/10.1073/pnas.1316278111> PMID: 24477691
4. Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, et al. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science.* 15 déc 2000; 290(5499):2105–10. PMID: 11118137
5. Gahlaut V, Jaiswal V, Kumar A, Gupta PK. Transcription factors involved in drought tolerance and their possible role in developing drought tolerant cultivars with emphasis on wheat (*Triticum aestivum* L.). *Theor Appl Genet.* nov 2016; 129(11):2019–42. <https://doi.org/10.1007/s00122-016-2794-z> PMID: 27738714
6. Riechmann JL, Meyerowitz EM. The AP2/EREBP family of plant transcription factors. *Biol Chem.* juin 1998; 379(6):633–46. PMID: 9687012
7. Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K. Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. *Proc Natl Acad Sci USA.* 10 oct 2000; 97(21):11632–7. <https://doi.org/10.1073/pnas.190309197> PMID: 11005831
8. Jakoby M, Weissshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, et al. bZIP transcription factors in Arabidopsis. *Trends Plant Sci.* mars 2002; 7(3):106–11. PMID: 11906833
9. Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) Function as Transcriptional Activators in Abscisic Acid Signaling. *The Plant Cell.* 1 janv 2003; 15(1):63–78. <https://doi.org/10.1105/tpc.006130> PMID: 12509522
10. Fujita M, Fujita Y, Maruyama K, Seki M, Hiratsu K, Ohme-Takagi M, et al. A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *Plant J.* sept 2004; 39(6):863–76.
11. Tran L-SP, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, et al. Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. *Plant Cell.* sept 2004; 16(9):2481–98.
12. Tran L-SP, Nakashima K, Sakuma Y, Osakabe Y, Qin F, Simpson SD, et al. Co-expression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the ERD1 gene in Arabidopsis. *Plant J.* janv 2007; 49(1):46–63. <https://doi.org/10.1111/j.1365-3113.2006.02932.x> PMID: 17233795
13. Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, et al. The MYB Transcription Factor Superfamily of Arabidopsis: Expression Analysis and Phylogenetic Comparison with the Rice MYB Family. *Plant Mol Biol.* 1 janv 2006; 60(1):107–24. <https://doi.org/10.1007/s11103-005-2910-y> PMID: 16463103
14. Lata C, Prasad M. Role of DREBs in regulation of abiotic stress responses in plants. *J Exp Bot.* oct 2011; 62(14):4731–48. <https://doi.org/10.1093/jxb/err210> PMID: 21737415
15. Zhu X, Liu S, Meng C, Qin L, Kong L, Xia G. WRKY Transcription Factors in Wheat and Their Induction by Biotic and Abiotic Stress. *Plant Molecular Biology Reporter.* oct 2013; 31(5):1053–67.
16. Huang Q, Wang Y, Li B, Chang J, Chen M, Li K, et al. TaNAC29, a NAC transcription factor from wheat, enhances salt and drought tolerance in transgenic Arabidopsis. *BMC Plant Biol.* 4 nov 2015; 15:268. <https://doi.org/10.1186/s12870-015-0644-9> PMID: 26536863
17. Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H, et al. Genome-wide analysis of NAC transcription factor family in rice. *Gene.* 1 oct 2010; 465(1–2):30–44. <https://doi.org/10.1016/j.gene.2010.06.008> PMID: 20600702

18. Zhu G, Chen G, Zhu J, Zhu Y, Lu X, Li X, et al. Molecular Characterization and Expression Profiling of NAC Transcription Factors in *Brachypodium distachyon* L. PLOS ONE. 7 oct 2015; 10(10):e0139794. <https://doi.org/10.1371/journal.pone.0139794> PMID: 26444425
19. Pinheiro G, Marques C, Costa M, A.B. Reis P, Alves M, Carvalho C, et al. Complete inventory of soybean NAC transcription factors: Sequence conservation and expression analysis uncover their distinct roles in stress response. Gene. 1 juill 2009; 444:10–23. <https://doi.org/10.1016/j.gene.2009.05.012> PMID: 19497355
20. Lu M, Sun Q-P, Zhang D-F, Wang T-Y, Pan J-B. Identification of 7 stress-related NAC transcription factor members in maize (*Zea mays* L.) and characterization of the expression pattern of these genes. Biochem Biophys Res Commun. 26 juin 2015; 462(2):144–50. <https://doi.org/10.1016/j.bbrc.2015.04.113> PMID: 25937463
21. Saidi MN, Mergby D, Brini F. Identification and expression analysis of the NAC transcription factor family in durum wheat (*Triticum turgidum* L. ssp. durum). Plant Physiol Biochem. mars 2017; 112:117–28. <https://doi.org/10.1016/j.plaphy.2016.12.028> PMID: 28064119
22. Moyano E, Martínez-Rivas FJ, Blanco-Portales R, Molina-Hidalgo FJ, Ric-Varas P, Matas-Arroyo AJ, et al. Genome-wide analysis of the NAC transcription factor family and their expression during the development and ripening of the *Fragaria × ananassa* fruits. PLOS ONE. 3 mai 2018; 13(5):e0196953. <https://doi.org/10.1371/journal.pone.0196953> PMID: 29723301
23. Ooka H, Satoh K, Doi K, Nagata T, Otomo Y, Murakami K, et al. Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. DNA Res. 31 déc 2003; 10(6):239–47. PMID: 15029955
24. Mathew IE, Das S, Mahto A, Agarwal P. Three Rice NAC Transcription Factors Heteromerize and Are Associated with Seed Size. Front Plant Sci. 2016; 7:1638. <https://doi.org/10.3389/fpls.2016.01638> PMID: 27872632
25. Xie Q, Frugis G, Colgan D, Chua NH. Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes Dev. 1 déc 2000; 14(23):3024–36. PMID: 11114891
26. Hao Y-J, Wei W, Song Q-X, Chen H-W, Zhang Y-Q, Wang F, et al. Soybean NAC transcription factors promote abiotic stress tolerance and lateral root formation in transgenic plants. Plant J. oct 2011; 68(2):302–13. <https://doi.org/10.1111/j.1365-3113X.2011.04687.x> PMID: 21707801
27. Olsen AN, Ernst HA, Leggio LL, Skriver K. NAC transcription factors: structurally distinct, functionally diverse. Trends Plant Sci. févr 2005; 10(2):79–87. <https://doi.org/10.1016/j.tplants.2004.12.010> PMID: 15708345
28. Kim S-G, Lee S, Seo PJ, Kim S-K, Kim J-K, Park C-M. Genome-scale screening and molecular characterization of membrane-bound transcription factors in *Arabidopsis* and rice. Genomics. janv 2010; 95(1):56–65. <https://doi.org/10.1016/j.ygeno.2009.09.003> PMID: 19766710
29. Souer E, van Houwelingen A, Kloos D, Mol J, Koes R. The no apical meristem gene of *Petunia* is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. Cell. 19 avr 1996; 85(2):159–70. PMID: 8612269
30. Sablowski RW, Meyerowitz EM. A homolog of NO APICAL MERISTEM is an immediate target of the floral homeotic genes APETALA3/PISTILLATA. Cell. 9 janv 1998; 92(1):93–103. PMID: 9489703
31. Guo Y, Gan S. AtNAP, a NAC family transcription factor, has an important role in leaf senescence. Plant J. mai 2006; 46(4):601–12. <https://doi.org/10.1111/j.1365-3113X.2006.02723.x> PMID: 16640597
32. Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M. The NAC transcription factors NST1 and NST2 of *Arabidopsis* regulate secondary wall thickenings and are required for anther dehiscence. Plant Cell. nov 2005; 17(11):2993–3006. <https://doi.org/10.1105/tpc.105.036004> PMID: 16214898
33. Kim Y-S, Kim S-G, Park J-E, Park H-Y, Lim M-H, Chua N-H, et al. A membrane-bound NAC transcription factor regulates cell division in *Arabidopsis*. Plant Cell. nov 2006; 18(11):3132–44. <https://doi.org/10.1105/tpc.106.043018> PMID: 17098812
34. Kim S-G, Lee A-K, Yoon H-K, Park C-M. A membrane-bound NAC transcription factor NTL8 regulates gibberellic acid-mediated salt signaling in *Arabidopsis* seed germination. Plant J. juill 2008; 55(1):77–88. <https://doi.org/10.1111/j.1365-3113X.2008.03493.x> PMID: 18363782
35. Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J. A NAC Gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science. 24 nov 2006; 314(5803):1298–301. <https://doi.org/10.1126/science.1133649> PMID: 17124321
36. Liang C, Wang Y, Zhu Y, Tang J, Hu B, Liu L, et al. OsNAP connects abscisic acid and leaf senescence by fine-tuning abscisic acid biosynthesis and directly targeting senescence-associated genes in rice. Proc Natl Acad Sci USA. 8 juill 2014; 111(27):10013–8. <https://doi.org/10.1073/pnas.1321568111> PMID: 24951508

37. Zhao D, Derkx AP, Liu D-C, Buchner P, Hawkesford MJ. Overexpression of a NAC transcription factor delays leaf senescence and increases grain nitrogen concentration in wheat. *Plant Biol (Stuttg)*. juill 2015; 17(4):904–13.
38. He X, Qu B, Li W, Zhao X, Teng W, Ma W, et al. The Nitrate-Inducible NAC Transcription Factor TaNAC2-5A Controls Nitrate Response and Increases Wheat Yield. *Plant Physiol*. nov 2015; 169(3):1991–2005. <https://doi.org/10.1104/pp.15.00568> PMID: 26371233
39. Xie Q, Sanz-Burgos AP, Guo H, García JA, Gutiérrez C. GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein. *Plant Mol Biol*. mars 1999; 39(4):647–56. PMID: 10350080
40. Wang F, Lin R, Feng J, Chen W, Qiu D, Xu S. TaNAC1 acts as a negative regulator of stripe rust resistance in wheat, enhances susceptibility to *Pseudomonas syringae*, and promotes lateral root development in transgenic *Arabidopsis thaliana*. *Front Plant Sci*. 2015; 6:108. <https://doi.org/10.3389/fpls.2015.00108> PMID: 25774162
41. Hu H, Dai M, Yao J, Xiao B, Li X, Zhang Q, et al. Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *Proc Natl Acad Sci USA*. 29 août 2006; 103(35):12987–92. <https://doi.org/10.1073/pnas.0604882103> PMID: 16924117
42. Mao X, Chen S, Li A, Zhai C, Jing R. Novel NAC transcription factor TaNAC67 confers enhanced multi-abiotic stress tolerances in *Arabidopsis*. *PLoS ONE*. 2014; 9(1):e84359. <https://doi.org/10.1371/journal.pone.0084359> PMID: 24427285
43. Fang Y, Liao K, Du H, Xu Y, Song H, Li X, et al. A stress-responsive NAC transcription factor SNAC3 confers heat and drought tolerance through modulation of reactive oxygen species in rice. *J Exp Bot*. nov 2015; 66(21):6803–17. <https://doi.org/10.1093/jxb/erv386> PMID: 26261267
44. Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, et al. Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol Biol Evol*. juill 2007; 24(7):1506–17. <https://doi.org/10.1093/molbev/msm077> PMID: 17443011
45. Borrill P, Harrington SA, Uauy C. Genome-Wide Sequence and Expression Analysis of the NAC Transcription Factor Family in Polyploid Wheat. *G3 (Bethesda)*. 07 2017; 7(9):3019–29.
46. Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, et al. An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res*. 2017; 27(5):885–96. <https://doi.org/10.1101/gr.217117.116> PMID: 28420692
47. Consortium (IWGSC) TIWGS, Investigators IR principal, Appels R, Eversole K, Feuillet C, Keller B, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*. 17 août 2018; 361(6403):eaar7191. <https://doi.org/10.1126/science.aar7191> PMID: 30115783
48. Borrill P, Ramirez-Gonzalez R, Uauy C. expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform. *Plant Physiol*. 2016; 170(4):2172–86. <https://doi.org/10.1104/pp.15.01667> PMID: 26869702
49. Okonechnikov K, Golosova O, Fursov M, UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 15 avr 2012; 28(8):1166–7. <https://doi.org/10.1093/bioinformatics/bts091> PMID: 22368248
50. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 1 mai 2014; 30(9):1236–40. <https://doi.org/10.1093/bioinformatics/btu031> PMID: 24451626
51. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 1 oct 2010; 26(19):2460–1. <https://doi.org/10.1093/bioinformatics/btq461> PMID: 20709691
52. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 2016; 33(7):1870–4. <https://doi.org/10.1093/molbev/msw054> PMID: 27004904
53. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004; 32(5):1792–7. <https://doi.org/10.1093/nar/gkh340> PMID: 15034147
54. Noé L, Kucherov G. YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res*. 1 juill 2005; 33(Web Server issue):W540–543. <https://doi.org/10.1093/nar/gki478> PMID: 15980530
55. Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, et al. Structural and functional partitioning of bread wheat chromosome 3B. *Science*. 18 juill 2014; 345(6194):1249721.
56. Liu Z, Xin M, Qin J, Peng H, Ni Z, Yao Y, et al. Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.). *BMC Plant Biol*. 20 juin 2015; 15:152. <https://doi.org/10.1186/s12870-015-0511-8> PMID: 26092253

57. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 20 avr 2015; 43(7):e47. <https://doi.org/10.1093/nar/gkv007> PMID: 25605792
58. Howe EA, Sinha R, Schlauch D, Quackenbush J. RNA-Seq analysis in MeV. *Bioinformatics.* 15 nov 2011; 27(22):3209–10. <https://doi.org/10.1093/bioinformatics/btr490> PMID: 21976420
59. Allen RG, Pereira LS, Raes D, Smith M. Crop evapotranspiration—Guidelines for computing crop water requirements—FAO Irrigation and drainage paper 56:15.
60. Slafer GA, Savin R, Sadras VO. Coarse and fine regulation of wheat yield components in response to genotype and environment. *Field Crops Research.* 15 févr 2014; 157:71–83.
61. Capron D, Mouzeyar S, Boulaflous A, Girousse C, Rustenholz C, Laugier C, et al. Transcriptional profile analysis of E3 ligase and hormone-related genes expressed during wheat grain development. *BMC Plant Biol.* 14 mars 2012; 12:35. <https://doi.org/10.1186/1471-2229-12-35> PMID: 22416807
62. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods.* déc 2001; 25(4):402–8. <https://doi.org/10.1006/meth.2001.1262> PMID: 11846609
63. Tang Y, Horikoshi M, Li W. ggfortify: Unified Interface to Visualize Statistical Results of Popular R Packages. 2016; 8:12.
64. Baloglu MC, Oz MT, Oktem HA, Yucel M. Expression Analysis of TaNAC69-1 and TtNAMB-2, Wheat NAC Family Transcription Factor Genes Under Abiotic Stress Conditions in Durum Wheat (*Triticum turgidum*). *Plant Mol Biol Rep.* 1 oct 2012; 30(5):1246–52.
65. Baloglu MC, Inal B, Kavas M, Unver T. Diverse expression pattern of wheat transcription factors against abiotic stresses in wheat species. *Gene.* 15 oct 2014; 550(1):117–22. <https://doi.org/10.1016/j.gene.2014.08.025> PMID: 25130909
66. Li X, Feng B, Zhang F, Tang Y, Zhang L, Ma L, et al. Bioinformatic Analyses of Subgroup-A Members of the Wheat bZIP Transcription Factor Family and Functional Identification of TabZIP174 Involved in Drought Stress Response. *Front Plant Sci.* 2016 Nov 16; 7:1643. <https://doi.org/10.3389/fpls.2016.01643> PMID: 27899926
67. Peng X, Zhao Y, Li X, Wu M, Chai W, Sheng L, et al. Genomewide identification, classification and analysis of NAC type gene family in maize. *J Genet.* sept 2015; 94(3):377–90.
68. Huo N, Zhang S, Zhu T, Dong L, Wang Y, Mohr T, et al. Gene Duplication and Evolution Dynamics in the Homeologous Regions Harboring Multiple Prolamin and Resistance Gene Families in Hexaploid Wheat. *Front Plant Sci.* 2018 May 23; 9:673. <https://doi.org/10.3389/fpls.2018.00673> PMID: 29875781
69. Ning P, Liu C, Kang J, Lv J. Genome-wide analysis of WRKY transcription factors in wheat (*Triticum aestivum* L.) and differential expression under water deficit condition. *PeerJ.* 4 mai 2017; 5:e3232. <https://doi.org/10.7717/peerj.3232> PMID: 28484671
70. Salse J, Bolot S, Throude M, Jouffe V, Piegue B, Quraishi UM, et al. Identification and Characterization of Shared Duplications between Rice and Wheat Provide New Insight into Grass Genome Evolution. *The Plant Cell.* 1 janv 2008; 20(1):11–24. <https://doi.org/10.1105/tpc.107.056309> PMID: 18178768
71. Wicker T, Buchmann JP, Keller B. Patching gaps in plant genomes results in gene movement and erosion of colinearity. *Genome Res.* sept 2010; 20(9):1229–37.
72. Glover NM, Daron J, Pingault L, Vandepoele K, Paux E, Feuillet C, et al. Small-scale gene duplications played a major role in the recent evolution of wheat chromosome 3B. *Genome Biol.* 9 sept 2015; 16:188. <https://doi.org/10.1186/s13059-015-0754-6> PMID: 26353816
73. Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. Gene duplication as a major force in evolution. *Journal of Genetics.* avr 2013; 92(1):155–61. PMID: 23640422
74. Brosius J. Retroposons—seeds of evolution. *Science.* 15 févr 1991; 251(4995):753. PMID: 1990437
75. Mao X, Zhang H, Qian X, Li A, Zhao G, Jing R. TaNAC2, a NAC-type wheat transcription factor conferring enhanced multiple abiotic stress tolerances in *Arabidopsis*. *J Exp Bot.* mai 2012; 63(8):2933–46. <https://doi.org/10.1093/jxb/err462> PMID: 22330896
76. Tang Y, Liu M, Gao S, Zhang Z, Zhao X, Zhao C, et al. Molecular characterization of novel TaNAC genes in wheat and overexpression of TaNAC2a confers drought tolerance in tobacco. *Physiol Plant.* mars 2012; 144(3):210–24. <https://doi.org/10.1111/j.1399-3054.2011.01539.x> PMID: 22082019
77. Huang Q, Wang Y. Overexpression of TaNAC2D displays opposite responses to abiotic stresses between seedling and mature stage of transgenic *Arabidopsis*. *Front. Plant Sci.* 2016 Nov 23; 7:1754. <https://doi.org/10.3389/fpls.2016.01754> PMID: 27933076
78. Chen D, Chai S, McIntyre CL, Xue G-P. Overexpression of a predominantly root-expressed NAC transcription factor in wheat roots enhances root length, biomass and drought tolerance. *Plant Cell Rep.* févr 2018; 37(2):225–37. <https://doi.org/10.1007/s00299-017-2224-y> PMID: 29079898

79. Zhang L, Zhang L, Xia C, Zhao G, Jia J, Kong X. The Novel Wheat Transcription Factor TaNAC47 Enhances Multiple Abiotic Stress Tolerances in Transgenic Plants. *Front Plant Sci.* 2016 Jan 18; 6:1174. <https://doi.org/10.3389/fpls.2015.01174> PMID: 26834757
80. Shen H, Yin Y, Chen F, Xu Y, Dixon RA. A Bioinformatic Analysis of NAC Genes for Plant Cell Wall Development in Relation to Lignocellulosic Bioenergy Production. *BioEnergy Research.* déc 2009; 2 (4):217–32.
81. Mathew IE, Agarwal P. May the Fittest Protein Evolve: Favoring the Plant-Specific Origin and Expansion of NAC Transcription Factors. *Bioessays.* 25 juin 2018;e1800018. <https://doi.org/10.1002/bies.201800018> PMID: 29938806
82. Pei H, Ma N, Tian J, Luo J, Chen J, Li J, et al. An NAC transcription factor controls ethylene-regulated cell expansion in flower petals. *Plant Physiol.* oct 2013; 163(2):775–91.
83. Christiansen MW, Holm PB, Gregersen PL. Characterization of barley (*Hordeum vulgare* L.) NAC transcription factors suggests conserved functions compared to both monocots and dicots. *BMC Res Notes.* 19 août 2011; 4:302. <https://doi.org/10.1186/1756-0500-4-302> PMID: 21851648
84. Zhang B, Li W, Chang X, Li R, Jing R. Effects of Favorable Alleles for Water-Soluble Carbohydrates at Grain Filling on Grain Weight under Drought and Heat Stresses in Wheat. Ali J, éditeur. *PLoS ONE.* 18 juill 2014; 9(7):e102917. <https://doi.org/10.1371/journal.pone.0102917> PMID: 25036550
85. Xue G-P, Way HM, Richardson T, Drenth J, Joyce PA, McIntyre CL. Overexpression of TaNAC69 leads to enhanced transcript levels of stress up-regulated genes and dehydration tolerance in bread wheat. *Mol Plant.* juill 2011; 4(4):697–712. <https://doi.org/10.1093/mp/ssr013> PMID: 21459832
86. Chen D, Richardson T, Chai S, Lynne McIntyre C, Rae AL, Xue G-P. Drought-Up-Regulated TaNAC69-1 is a Transcriptional Repressor of TaSHY2 and TaAA7, and Enhances Root Length and Biomass in Wheat. *Plant Cell Physiol.* oct 2016; 57(10):2076–90. <https://doi.org/10.1093/pcp/pcw126> PMID: 27440550
87. Jeong JS, Kim YS, Baek KH, Jung H, Ha S-H, Do Choi Y, et al. Root-specific expression of OsNAC10 improves drought tolerance and grain yield in rice under field drought conditions. *Plant Physiol.* mai 2010; 153(1):185–97. <https://doi.org/10.1104/pp.110.154773> PMID: 20335401
88. Girousse C, Roche J, Guerin C, Le Gouis J, Balzegue S, Mouzeyar S, et al. Coexpression network and phenotypic analysis identify metabolic pathways associated with the effect of warming on grain yield components in wheat. *PLoS ONE.* 2018; 13(6):e0199434. <https://doi.org/10.1371/journal.pone.0199434> PMID: 29940014
89. Zhao X, Yang X, Pei S, He G, Wang X, Tang Q, et al. The Miscanthus NAC transcription factor MINAC9 enhances abiotic stress tolerance in transgenic Arabidopsis. *Gene.* 15 juill 2016; 586(1):158–69. <https://doi.org/10.1016/j.gene.2016.04.028> PMID: 27085481
90. Han Q, Zhang J, Li H, Luo Z, Ziaf K, Ouyang B, et al. Identification and expression pattern of one stress-responsive NAC gene from *Solanum lycopersicum*. *Mol Biol Rep.* févr 2012; 39(2):1713–20. <https://doi.org/10.1007/s11033-011-0911-2> PMID: 21637957
91. Hong Y, Zhang H, Huang L, Li D, Song F. Overexpression of a Stress-Responsive NAC Transcription Factor Gene ONAC022 Improves Drought and Salt Tolerance in Rice. *Front Plant Sci.* 2016; 7:4. <https://doi.org/10.3389/fpls.2016.00004> PMID: 26834774
92. Conesa A, Götz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics.* 2008; 2008:619832. <https://doi.org/10.1155/2008/619832> PMID: 18483572
93. Xia N, Zhang G, Sun Y-F, Zhu L, Xu L-S, Chen X-M, et al. TaNAC8, a novel NAC transcription factor gene in wheat, responds to stripe rust pathogen infection and abiotic stresses. *Physiological and Molecular Plant Pathology.* 1 sept 2010; 74(5):394–402.
94. Macovei A, Tuteja N. microRNAs targeting DEAD-box helicases are involved in salinity stress response in rice (*Oryza sativa* L.). *BMC Plant Biol.* 8 oct 2012; 12:183. <https://doi.org/10.1186/1471-2229-12-183> PMID: 23043463
95. Takasaki H, Maruyama K, Kidokoro S, Ito Y, Fujita Y, Shinozaki K, et al. The abiotic stress-responsive NAC-type transcription factor OsNAC5 regulates stress-inducible genes and stress tolerance in rice. *Mol Genet Genomics.* sept 2010; 284(3):173–83. <https://doi.org/10.1007/s00438-010-0557-0> PMID: 20632034
96. Fang Y, Xie K, Xiong L. Conserved miR164-targeted NAC genes negatively regulate drought resistance in rice. *J Exp Bot.* mai 2014; 65(8):2119–35. <https://doi.org/10.1093/jxb/eru072> PMID: 24604734