UCLouvain

Faculté des Bioingénieurs

Louvain Institute of Biomolecular Science and Technology (LIBST)

Groupe de Physiologie Moléculaire (FYMO)

Mapping and characterisation of regulatory variants controlling the expression of aquaporin genes in the mature zone of maize leaves

Thèse présentée par Maxime LAURENT en vue de l'obtention du grade de Docteur en Sciences Agronomiques et Ingénierie Biologique

Promoteur :	Prof. François Chaumont
Président du jury :	Prof. Michel Ghislain
Membres du jury :	Prof. Charles Hachez
	Prof. Xavier Draye
	Prof. Hilde Nelissen
	Dr. Clémentine Vitte

Février 2025

Acknowledgment

Il y a des projets qui, même s'ils semblent être le fruit d'un travail personnel, ne peuvent pas réellement être accomplis seul. Une thèse de doctorat en fait partie. C'est pourquoi je souhaite remercier toutes les personnes qui m'ont accompagné et soutenu tout au long de cette période de ma vie.

J'aimerais commencer par les personnes sans qui je n'aurais pas commencé mon mémoire au FYMO. D'abord François, je t'avais contacté pour en savoir plus à propos des projets de mémoire, et tu m'as vite convaincu de rejoindre ton équipe. Merci également pour ta confiance et de m'avoir proposé de continuer l'analyse des données de DROPS, connaissant mon intérêt pour la bioinformatique. Ensuite, Lei qui m'a encadré pendant mon mémoire et qui m'a appris les bases de la biologie moléculaire et du phénotypage. Il est ensuite devenu un fantastique collègue de bureau toujours prêt à répondre à mes nombreuses questions. Finalement, parce que l'opération recrutement du FYMO ne se fait pas uniquement via les promoteurs mais aussi via les TPs de biologie moléculaire, je tiens à remercier Joseph… Grand séducteur, il m'a aussi convaincu que le FYMO était l'endroit idéal pour mon mémoire. Toujours prêt à mettre son inestimable expérience au service des doctorants du labo, il a été une autre référence importante dans mon travail.

I would like to thank the members of my thesis committee, Clémentine Vitte, Hilde Nelissen, Xavier Draye, Charles Hachez and Michel Ghislain for accepting to carefully review my manuscript and giving me detailed feedback about how to improve it.

Pour la partie génétique quantitative de mon projet, je voudrais remercier Stéphane et Renaud pour leur encadrement. N'ayant aucune connaissance dans ce domaine avant ma thèse, votre expertise a été essentielle pour démarrer mes analyses.

Je tiens bien sûr remercier les anciens et actuels membres de l'équipe AQP : Laurie pour m'avoir aidé au début de ma thèse, Kim pour son aide lors de mon retour à la paillasse après un an sans manips, Estelle avec qui j'ai pu partager les galères quotidiennes de la dual-luciferase et du DAP-seq, Romina qui était là pour nous partager son expertise, et finalement les derniers arrivés, Jean et Joséphine pour avoir amené du sang frais et des nouvelles idées dans l'équipe notamment avec le fameux concours de memes, et Mannon qui découvre encore le labo. Jules qui n'est pas totalement dans l'équipe AQP, mais c'est tout comme. J'aimerais aussi remercier les deux mémorants que j'ai pu encadrer, Joséphine et Alexandre, qui m'ont obligé à être responsable et à faire semblant de savoir ce que je faisais…

Je souhaite aussi remercier les membres du FYMO en général, car s'il y a bien une chose que je vais retenir de cette expérience, c'est l'entraide. En effet, j'ai toujours eu quelqu'un pour m'aider à chaque fois que j'en ai eu besoin. En particulier, lorsque je devais faire des western blots où c'est presque toute l'équipe Pharm qui s'est mobilisée pour un résultat « EXCELLENT ». Un tout grand merci à vous tous. En plus de l'entraide, je retiendrai aussi la bonne ambiance générale qui règne au labo, notamment via les nombreuses activités de teambuilding, merci aux participants et surtout tous ceux impliqués dans leur organisation. Par exemple, les fameuses FYMOlympiades, soupers de Noël, cheese and wine, sorties quinzaines, et les escapes room dans le labo et en-dehors, resteront gravés dans ma mémoire.

Je n'oublie pas Ben, que j'ai souvent ennuyé avec de nouvelles idées pour le serveur. Encore désolé de l'avoir bloqué en mettant mes données au mauvais endroit. Marie-Christine, merci pour ton encadrement au niveau du confocal.

Je pense également aux personnes qui ont commencé leur thèse en même temps que moi après leur mémoire au FYMO : Estelle, Quentin, Louche, Nicolas et Guillemette. Déjà bravo à Nicolas pour cette fin de thèse mais aussi courage aux autres, vous y êtes presque !

Un tout grand merci à mes proches qui m'ont soutenu dans ma décision de faire un doctorat et qui ont toujours été à l'écoute lorsque j'en avais besoin. Je tiens particulièrement à remercier mes parents qui m'ont donné l'opportunité de suivre les études que je voulais et qui m'ont fourni l'environnement nécessaire pour les mener à bien. Je tiens également à remercier Shriya pour son support quotidien malgré la distance et qui commence à en avoir marre d'entendre parler d'aquaporines. Merci également à Guillaume, Thomas et mes colocs pour m'avoir aidé à me changer les idées après les journées de travail au labo. Enfin, merci à Julien et Charlie de m'avoir donné la force de continuer dans la dernière ligne droite.

Merci à vous tous pour ces années formidables !

Abstract

Aquaporins (AQP) are channels that modulate the permeability of cell membranes to water and small solutes. In plants, they are involved in numerous physiological and cellular processes, including the hydraulic conductivity of cells, tissues and organs, cell expansion and homeostasis. Given their involvement in such essential roles, they are tightly regulated at multiple levels according to developmental stages, organs, daytime, and in response to environmental stimuli. In this PhD thesis, we were mainly interested in the transcriptional regulation of *AQP* genes, for which expression profiles under various conditions have been reported. However, little data was available on the specific molecular components involved in their transcription.

Our goal was to discover mechanisms that regulate the transcription of *AQP* genes in maize leaves under well-watered (WW) and water deficit (WD) conditions. Based on 3' RNA-seq data from the mature zone of leaves from 254 maize lines grown under both watering conditions (DROPS panel), we identified *AQP* genes expressed in this region and analysed how their expression varied with soil water availability and genetic background. In addition, we performed a genome-wide association study (GWAS) and mapped expression quantitative trait loci (eQTL) for *AQP* genes to reveal regions where potential regulators of their expression might be located. All genes present in these eQTLs were retrieved, and those annotated as transcription factors (TFs) were selected for further analyses. We then conducted an enrichment analysis using 45 published gene regulatory networks generated on data from various conditions. Fifteen TFs were found in both eQTLs and enrichment analyses and were selected for further wet lab characterisation using dual-fluorescence and dual-luciferase reporter gene assays. We demonstrated that *ZmHSF13* and *ZmNAC40* influenced the promoter activities of *ZmTIP1;1* and *ZmPIP1;3*, respectively. Using DAP-seq, we confirmed that *ZmHSF13* directly binds to the *ZmTIP1;1* promoter. These results contribute to a better understanding of how TFs regulate *AQP* gene expression and provide important information for further investigation of the *AQP* gene regulatory network.

Contribution to publications

Ali, B., Huguenin-Bizot, B., Laurent, M., Chaumont, F., Maistriaux, L. C., Nicolas, S., Duborjal, H., Welcker, C., Tardieu, F., Mary-Huard, T., Moreau, L., Charcosset, A., Runcie, D., & Rincent, R. (2024). High-dimensional multi-omics measured in controlled conditions are useful for maize platform and field trait predictions. *Theoretical and Applied Genetics*, *137*(7), 175. https://doi.org/10.1007/s00122-024-04679-w

Contribution:

- Development of the preprocessing pipeline for the 3'RNA-seq data
- Analysis of the different preprocessing scenarios
- Reviewing the final version

Maistriaux, L. C., Laurent, M. J., Jeanguenin, L., Prado, S. A., Nader, J., Welcker, C., Charcosset, A., Tardieu, F., Nicolas, S. D., & Chaumont, F. (2024). Genetic variability of aquaporin expression in maize : From eQTLs to a MITE insertion regulating *PIP2;5* expression. *Plant Physiology*, kiae326. https://doi.org/10.1093/plphys/kiae326

Contribution :

- Collaboration in compiling the results into the manuscript
- Reviewing collaborators' comments and editing of the manuscript
- Conversion of the eQTL positions from B73 reference genome v2 to v4
- Update of Figures 3, 4, 5 to the v4 positions
- Drafting of the initial responses to reviewers and editing of the manuscript accordingly

Ding, L., Laurent, M.J., Milhiet, T., Aesaert, S., Van Lijsebettens, M., Pauwels, L., Nelissen, H., Inzé, D., & Chaumont, F. (2024). The maize aquaporin ZmPIP1;6 enhances stomatal opening and CO_2^- and ABA-induced stomatal closure. *Journal of Experimental Botany*, erae500. https://doi.org/10.1093/jxb/erae500

Contribution:

- Amplification of the *ZmPIP1;6* promoter from B104 germplasm
- Generation and validation of the genetic construct to express the *Zm*PIP1;6-mYFP fusion protein

Table of contents

ACKNOWLEDGMENT ..... **I**

ABSTRACT ..... **III**

CONTRIBUTION TO PUBLICATIONS ..... **V**

TABLE OF CONTENTS ..... **VII**

ABBREVIATIONS ..... **XI**

LIST OF FIGURES ..... **XIII**

LIST OF TABLES ..... **XVII**

INTRODUCTION **1**

1. Gene expression **2**
1.1 Gene expression key components 2
1.2 Measuring gene expression 4
1.3 Gene regulation networks and genetic variations 6
1.4 Identifying expression regulators via genetic variations 7
1.5 Methods to characterise expression regulators 11

2. Aquaporins **16**
2.1 Water transport in plants 16
2.2 Structure & phylogeny 17
2.3 Role of AQPs in leaf responses to drought stress 19
2.4 Regulators of AQP expression 23

RESEARCH OBJECTIVES **27**

RESULTS & DISCUSSION **29**

1. Genome-wide identification of *AQPs* in maize **29**
1.1 Identification of new AQPs and phylogenetic analysis 29
1.2 AQP families in several maize cultivars 32
1.3 Comparison with previous genome-wide AQP identification 33

2. 3'RNA-seq data analysis **34**
2.1 Diversity panel and experimental design 34
2.2 Identification of biases in the data and preprocessing 35
2.3 Differential expression analysis between WW and WD 40
2.4 Choice of preprocessing methods 42

3. eQTL mapping for *AQP* gene expression **48**
3.1 eQTL mapping 48
3.2 Markers quality control 51
3.3 eQTLs quality control 51
3.4 eQTLs analysis 55

4. Selection of candidates **57**
4.1 Enrichment analysis 57
4.2 Network on the GWAS 58
4.3 *AQP* gene promoter analysis 59

4.4 Analysis of *AQP* GRNs 60

5. Effect of candidate TFs on *AQP* gene expression 66

5.1 *ZmTIP1;1* eQTLs 66

5.2 *ZmPIP1;3* & *ZmPIP2;3* eQTLs 74

GENERAL DISCUSSION & PERSPECTIVES 81

MATERIALS & METHODS 85

Identification of novel maize *AQP* genes 85

Phylogenetic analysis 85

Intergenomic comparison of maize *AQP* genes 86

Plant material 86

RNA extraction 86

Library Preparation & Sequencing 87

Exploratory data analysis 87

3'RNA-seq preprocessing 87

Differential expression analysis 88

Gene ontology enrichment 88

3'RNA-seq spatial correction 89

Genotyping 89

eQTL identification 90

Marker annotation 91

Enrichment analysis of TF-AQP target gene associations 91

TF binding motifs in AQP promoters 92

Forward-backward analysis 92

Genetic constructs 92

TF CDS 93

AQP promoter amplification 95

Dual-luciferase plasmids 95

Dual-fluorescence plasmids 96

Maxiprep 96

Protoplast preparation and transfection 97

Dual-luciferase analysis 98

Dual-fluorescence analysis 99

DAP-sequencing experiment 99

Library preparation 99

DAP-sequencing 100

Read mapping, filtering and peak calling 101

General Packages used in this project **101**

REFERENCES **103**

APPENDIX **133**

Supplemental information **133**
Kinship 133
Confocal images intensity measurement 133

Supplemental Figures **134**

Supplemental Tables **144**

Papers **154**

Abbreviations

ABA	Abscisic acid
ABRE	ABA-responsive element
AHA	aromatic and large hydrophobic amino acid residues embedded in an acidic surrounding
AQP	Aquaporin
ar/R	aromatic/Arg
ASE	allele-specific expression
ATAF	Arabidopsis Transcription Activation Factor
BLUEs	best linear unbiased estimators
bp	base pairs
cDNA	complementary DNA
CDS	coding DNA sequences
ChIP-seq	Chromatin Immunoprecipitation sequencing
CO-like	CONSTANS-like
CREs	*cis*-regulatory elements
CRMs	*cis*-regulatory modules
CCT	CO, CO-like, TOC1
CUC	Cup-shaped Cotyledon
DAP-seq	DNA affinity purification sequencing
DEGs	differentially expressed genes
DRE	dehydration responsive elements
EB	elution buffer
EMSA	electromobility-shift assay
eQTLs	expression QTLs
FPKM	fragment per kilobase million
GCNs	Gene co-expression networks
gDNA	genomic DNA
GLK	Golden2-like
GRN	gene regulatory networks
GW	Gateway
GWAS	genome-wide association studies
HD-Zip	Homeodomain-leucine zipper
HSE	Heat shock element
InDels	insertions/deletions
LD	linkage disequilibrium
lncRNAs	long-noncoding RNAs
LOCO	leave-one-chromosome-out
LogFC	log-fold change
Lpr	root hydraulic conductivity
MAF	minor allele frequency
MBS	MYB-binding site
MIPs	major intrinsic protein
miRNAs	micro RNAs
MITE	miniature-inverted repeat transposable element
MLM	mixed linear model
mRNA	messenger RNA
NAC	NAM, ATAF1, 2 and CUC2
NAM	No Apical Meristem
NIPs	nodulin-26-like proteins
NPA	asparagine-proline-alanine
PAV	presence/absence variations
PCA	principal component analysis
PCs	principal components
PIPs	plasma membrane intrinsic proteins
PLB	passive lysis buffer

PVE	phenotypic variance explained
QTLs	quantitative trait loci
R2	blister stage
R4	dough stage
REML	restricted maximum log-likelihood
RIN	RNA Integrity Number
RLE	relative log expression normalisation
RLU	relative light units
RNA-seq	RNA sequencing
ROI	region of interest
ROS	reactive oxygen species
RPKM	read per kilobase million
RT-qPCR	reverse transcription quantitative PCR
scRNAseq	single-cell RNAseq
SIPs	small basic intrinsic proteins
siRNAs	small interfering RNAs
SNPs	single-nucleotide polymorphisms
TALE	Three Amino acid Loop Extension
TB1	TEOSINTE BRANCHED 1
Tes	Transposable elements
TF	transcription factor
TIPs	tonoplast intrinsic proteins
TMM	trimmed mean of M-values normalisation
TSSs	transcription start sites
TTS	transcription termination sites
TUM	Technical University of Munich
v	Volume
VT	tassel stage
WD	water deficit
WT	Wild type
WW	well-watered
XIPs	X-intrinsic proteins
Y1H	Yeast-One hybrid
ZF-HD	Zinc finger homeodomain

List of figures

Figure 1: Transcriptional activation (right) and repression (left). 2
Figure 2: Effect of chromatin compaction on transcription binding sites (TFBS) accessibility. 4
Figure 3: Comparison between RNA-seq and 3'RNA-seq. 5
Figure 4: Comparison of linkage mapping (left) and association mapping (right). 7
Figure 5: Theoretical situation where markers are in LD. 8
Figure 6: Representation of the classic GWAS workflow to detect markers associated with the height phenotype. 10
Figure 7: Example of methods used to validate a TF-target interaction 13
Figure 8: Schematic representation of ChIP-seq and DAP-seq experiments. 15
Figure 9: Pathways used by water to cross cell layers and reach xylem vessels. 17
Figure 10: AQP structure. 18
Figure 11: Schematic representation of the regulation of PIPs in cells. 23
Figure 12: Phylogenetic tree of the AQP family in maize B73 cultivar. 32
Figure 13: Genome comparison of the *AQP* gene families from 30 maize cultivars. 33
Figure 14: Schematic of the DROPS experimental design. 35
Figure 15: Library size across sequencing plates. 36
Figure 16: Distribution of the $\log_2$(cpm) of all genes for five sample chosen randomly in the panel. 37
Figure 17: Gene expression distribution for all the studied samples. 38
Figure 18: Principal component analysis of the normalised and transformed expression data in $\log_2$(cpm). 40
Figure 19: Heatmap showing the $\log_{10}$FC between WD and WW conditions for the 115 DEGs. 42
Figure 20: Venn Diagram of genes expressed in the dataset. 43
Figure 21: Benchmark plot for WW (left) and WD (right) data. 44
Figure 22: Plot of the heritabilities of the expression of all the *AQP* genes in both watering conditions WD in orange and WW in green. 46
Figure 23: *ZmPIP* genes expression in log(cpm) in WW and WD. 47
Figure 24: *ZmPIP* eQTLs. 49
Figure 25: Q-Q plot of the -$\log_{10}$(p-value) for all the associations tested for *ZmPIP1;1* expression (A) and *ZmTIP2;4* expression (B). 51
Figure 26: Close-up Manhattan plot on chromosome 4 for the *ZmPIP1;5* gene in WW (left) and WD (right). 52
Figure 27: Manhattan plots of *ZmPIP2;1* under both watering conditions, WW (top) and WD (bottom). 53
Figure 28: Manhattan plots of *ZmPIP1;6* under both watering conditions, WW (top) and WD (bottom). 54

Figure 29: Scatter plot between the eQTLs size (-$\log_{10}$(bp)) and the median adjacent LD (measured as r^2K) before grouping with LD_{win} (A) and after grouping with LD_{win} (B). 55
Figure 30: Marker analysis. 56
Figure 31: Enrichment plot for TF-target associations. 58
Figure 32: Network of TF-AQP associations. 59
Figure 33: Full and close-up Manhattan plots for *ZmTIP1;1* gene expression in WW (A and B) and WD (C and D). 68
Figure 34: Expression in $\log_2$(cpm) of *ZmHSF13* and *ZmTIP1;1* in WW (green) and WD (orange). 69
Figure 35: Activity of *ZmTIP1;1* promoter upon ZmHSF13 expression determined by dual-luciferase assays in maize B104 mesophyll protoplasts. 71
Figure 36: *Zm*HSF13 DAP-seq results. 73
Figure 37: GO analysis of ZmHSF13 targets. 74
Figure 38: Full and close-up Manhattan plots for *ZmPIP1;3* gene expression in WW (A and B) and WD (C and D). 75
Figure 39: Full and close-up Manhattan plots for *ZmPIP2;3* gene expression in WW (A and B) and WD (C and D). 76
Figure 40: Expression in $\log_2$(cpm) of *ZmNAC40*, *ZmPIP1;3*, and *ZmPIP2;3* in WW (green) and WD (orange). 77
Figure 41: Activity of *pPIP1;3* promoter upon ZmNAC40 TF expression determined by dual-fluorescence experiments in maize B73 mesophyll protoplasts. 79

Supplemental Figure 1: Principal component analysis of the normalised and transformed expression data in $\log_2$(cpm). 134
Supplemental Figure 2: Principal component analysis of the normalised and transformed expression data in $\log_2$(cpm). 135
Supplemental Figure 3: *ZmTIP* genes expression in log(cpm) in WW and WD.136
Supplemental Figure 4: *ZmNIP* and *ZmSIP* genes expression in log(cpm) in WW and WD. 136
Supplemental Figure 5: *ZmNIP* eQTLs. 137
Supplemental Figure 6: *ZmSIP* eQTLs. 138
Supplemental Figure 7: *ZmTIP* eQTLs. 139
Supplemental Figure 8: Enriched and in eQTLs TF genes expression in log(cpm) in WW and WD. 140
Supplemental Figure 9: *ZmTIP1;1* expression between WW (green) and WD (orange), and between lines carrying either the reference (like B73) and alternative alleles. 141
Supplemental Figure 10: Alignment between a 700 bp fragment from the 3'end of *ZmTIP1;1* and an 890 bp fragment from CNVMAIZE_INS_102207. 142

Supplemental Figure 11: Western blot to verify the production and the correct binding of the Halo-tagged proteins on the magnetic beads for the DAP-seq experiment. 143

List of tables

Table 1: Updated list of *AQP* genes in maize. 30
Table 2: Pearson's correlation between samples belonging to B73. 39
Table 3: Set of benchmarked scenarios. 44
Table 4: TF-AQP interaction including the TF family and whether the binding motif of the TF family was located in the *AQP* promoter region. 60
Table 5: Models tested in the forward-backward analysis. 92
Table 6: Primer's list used to generate the genetic constructs described in this work. 94

Supplemental Table 1: List of the 254 varieties used in the DROPS panel. . 144
Supplemental Table 2: Between 3'RNA-seq samples Pearson correlation coefficient. 145
Supplemental Table 3: Raw data of the dual-luciferase experiment performed in B104 mesophyll cells. 147
Supplemental Table 4: Information about the sequencing results of the DAP-seq experiment 148
Supplemental Table 5: Top 20 enriched GO terms in DE genes from the 3'RNA-seq experiment. 148
Supplemental Table 6: Raw data of the dual-fluorescence experiment performed in B73 mesophyll cells protoplasts. 148
Supplemental Table 7: List of the genes used in the benchmarking of the different preprocessing methods 151
Supplemental Table 8: List of TFs in the network. IDs of the entry vectors from the TFome collection are given in the plasmid ID column (Burdo et al. 2014). . 152
Supplemental Table 9: List of the plasmids used in this work. 153

Introduction

Maize (*Zea Mays* L.) is the most cultivated cereal crop worldwide (*FAOSTAT*, 2022). It has a wide variety of uses, ranging from food products (food and feed) to industrial uses (biofuels, plastics, etc.). In 2021, its production reached 1.2 billion tons. However, maize production is also affected by weather events. For example, in 2022, it was reported that drought events in Europe triggered a 4% drop in yearly production (Lombardi et al., 2023). Such events have already been observed in the US, with decreases in yield of up to 29% due to droughts and floods (Z. Yang et al., 2023). Actually, global warming is predicted to rise to 1.5°C by 2040 (Calvin et al., 2023). This will lead to an increase in the frequency and severity of dramatic climatic events, putting food security at risk in several regions of the world, as crop yields are predicted to decrease (Calvin et al., 2023; A. Gupta et al., 2020; Pörtner et al., 2023).

As sessile organisms, plants must cope with the environmental conditions they face. Drought is one of the most dramatic abiotic cues that affect yield and is defined as a period of dry weather marked by a significant lack of precipitation, which results in a water deficit (*WMO*, 2023). Plants have developed various drought resistance strategies, such as drought tolerance, avoidance, and escape (Aslam et al., 2015; A. Gupta et al., 2020; Levitt, 1980). Drought tolerance refers to the strategy by which plants can undergo a period of low internal water content while protecting themselves from potential damage caused by the stress. This occurs through reprogramming of the transcriptome and metabolome. In the drought avoidance strategy, plants increase their water uptake and thus their internal water content, while reducing their water loss by modifying their morphology and physiology (Agurla et al., 2018). Finally, by accelerating their life cycle and producing viable seeds quicker, they can escape periods of water shortage. These strategies are beneficial for plant survival, but they usually result in decreased yields. In a world with a growing population, the yield is crucial for food security (Burgess et al., 2023). Studying drought stress responses could lead to a better understanding of the underlying mechanisms and the development of drought-tolerant plants (Fang & Xiong, 2015). Drought stress responses involve complex mechanisms at the physiological, morphological, and molecular levels. These responses vary among plant species, genotypes of given species, and developmental stages. Describing this diversity is not part of this work, but information can be found in Fang & Xiang (2015). These phenotypes are the result of many underlying molecular interactions, and evidence suggests that genetic variations in gene expression are crucial for their diversity (S. Liu, Li, et al., 2020).

Key actors involved in the uptake and distribution of water in plants, and therefore in drought tolerance are aquaporins (AQPs). AQPs are small channels

that facilitate the transport of water and other small solutes across cell membranes. They are involved in many physiological processes. In the context of this thesis, we were interested in their role in response to drought and chose to focus on the regulation of their expression. The following sections of this introduction will first provide a basic understanding of gene expression, how it is measured and the methods used to identify gene expression regulators. Next, we will review what is currently known about AQPs, their involvement in drought stress, and their known expression regulators.

1. Gene expression

Gene expression is a dynamic process by which the genetic information encoded in the genome is used to synthesise a functional gene product. For instance, protein-coding genes are used to produce proteins through the transcription of genomic DNA (gDNA) into messenger RNA (mRNA), followed by its translation into amino acids (Fromson, 1977).

1.1 Gene expression key components

Transcription is the first step in gene expression, in which a region of gDNA is copied into RNA. It is performed by proteins belonging to the RNA polymerase family. In plant, five RNA polymerases can be observed, Pol I, Pol II, Pol III, Pol IV and Pol V (reviewed in Matzke et al. (2015)). Pol II interacts with other factors and is responsible for mRNA production for protein synthesis (reviewed in Al-Husini et al. (2020)). Transcription is a tightly regulated process driven by several interacting components, including *cis*-regulatory elements (CREs), *trans*-acting factors, and chromatin architecture (Figure 1).

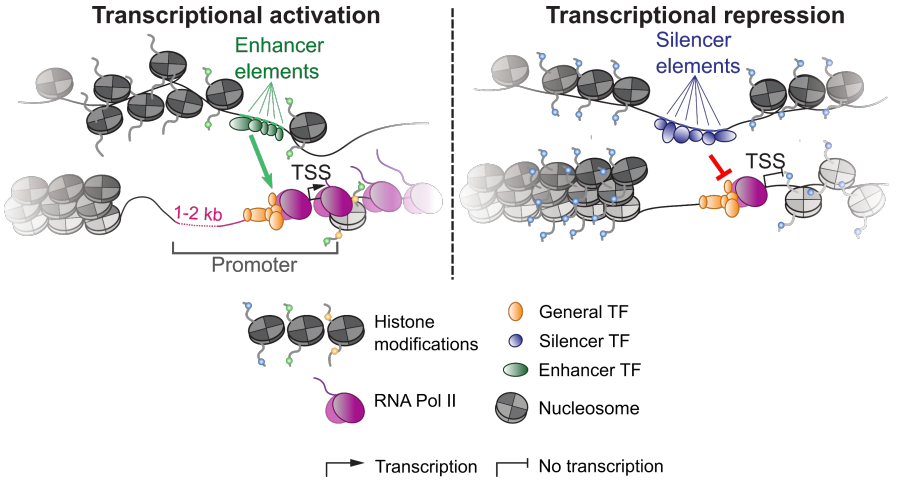


Figure 1: Transcriptional activation (right) and repression (left). *The core and proximal promoters are grouped into one region, called the promoter, which contains the CREs necessary to trigger gene expression. CREs that enhance or silence gene expression are grouped as enhancer and silencer elements, respectively. Different histone modifications as well as nucleosome occupancy (in grey)*

are also depicted. The transparency gradient of RNA Pol II indicates its movement on gDNA and mRNA production. Different TFs are coloured according to their roles (Adapted from Marand et al. (2023)).

CREs are specific gDNA sequences that are physically linked to their target genes and are therefore allele-specific. In plants, they are often assembled into *cis*-regulatory modules (CRMs) to form core promoters, enhancers, and silencers (Schmitz et al., 2022). The core promoter is a region of approximately 100 base pairs (bp) spanning around the transcription start sites (TSSs) that is essential for direct transcription initiation (Mejia-Guerra et al., 2015; Molina & Grotewold, 2005; C.-P. Yu et al., 2016). Several CREs can be found within 3k bp upstream of the core promoter, and together, they form the proximal promoter (core promoter and proximal promoter are hereafter referred to as promoter (Springer et al., 2019)). Enhancers and silencers are CRMs that are further away from the TSS of their target gene depending on the genome size, from a few hundred bp to several thousand bp (Z. Lu et al., 2019). They can, respectively, increase or decrease gene expression when bound by specific *trans*-acting factors, but not initiate their target gene transcription alone, unlike promoters.

Trans-acting factors are diffusible products that are not physically linked to their target genes; in diploids, they can target both alleles (Signor & Nuzhdin, 2018). They include DNA-binding proteins, long-noncoding RNAs (lncRNAs), small interfering RNAs (siRNAs), and micro RNAs (miRNAs). In this study, we were interested in transcription factors (TFs), which are DNA-binding proteins that are fundamental for gene expression (Spitz & Furlong, 2012). They are located at the end of signalling pathways and are responsible for the activation or repression of other genes (target genes) (O. P. Gupta et al., 2022). They specifically recognise CREs located near or distant from their target genes and form protein DNA-interactions guiding the transcription machinery (Schmitz et al., 2022).

Interactions between CREs and TFs dictate spatiotemporal expression patterns (Swift & Coruzzi, 2017). To prevent ectopic transcription that could lead to potentially harmful consequences, CRM activity is regulated by chromatin architecture (Marand et al., 2023). Chromatin has two roles: (i) condensing gDNA into the nucleus and (ii) regulating gene expression. In the nucleus, gDNA is wrapped around octamers of proteins called histones, forming nucleosomes. This packaging is not homogenous, and two distinct regions are observed, both having specific consequences on CRM activity. Each chromatin state is characterised by specific histone post-translational modifications, such as acetylation or methylation (Ricci et al., 2019). In condensed regions of chromatin (heterochromatin), where nucleosome occupancy is high, CREs are inaccessible by TFs and remain inactive. In contrast, less condensed regions (euchromatin) are depleted in nucleosomes, making CREs accessible and active (Figure 2). Specific TFs, called pioneer TFs, can bind to nucleosomal DNA and

recruit histone-modifying enzymes to modify chromatin conformation by altering these histone marks (Cirillo et al., 2002; Pajoro et al., 2014).

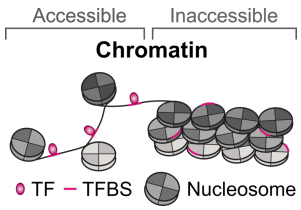


Figure 2: Effect of chromatin compaction on transcription binding sites (TFBS) accessibility. *TFBS are represented by pink lines and TFs by pink dots. When chromatin is compacted, CREs are inaccessible (Adapted from Marand et al. (2023)).*

1.2 Measuring gene expression

Gene expression of protein-coding genes is the quantitative measurement of the gene transcript, or mRNA, in a specific organ/tissue/cell at a given moment. This is commonly performed using reverse transcription quantitative PCR (RT-qPCR) or RNA sequencing (RNA-seq). In both cases, the extraction of the whole RNA from cells is followed by a reverse transcription to obtain complementary DNA (cDNA). Then, further steps are performed to achieve cDNA quantification depending on the method. RT-qPCR allows the quantification of a few genes through the design of specific primers and measurement by qPCR, while RNA-seq is usually performed to measure the transcriptome, the transcript abundance of all the genes, but require additional steps for the cDNA library preparation.

In this work, we used a method derived from RNA-seq called 3'RNA-seq (Eveland et al., 2008; Lohman et al., 2016). Both methods were compared in Ma et al. (2019) (Figure 3A). Compared to RNA-seq, which relies on the sequencing of cDNA fragments from the whole transcript, 3'RNA-seq only uses the 3'-end of the transcript. One major drawback of RNA-seq library preparation is that longer transcripts produce more cDNA fragments at the fragmentation step (Oshlack & Wakefield, 2009). Thus, in the differential expression analysis, the statistical power of longer transcripts is increased compared to that of shorter ones, resulting in an enrichment of longer genes being detected as differentially expressed. In 3'RNA-seq, only one cDNA molecule is produced per transcript, as there is no fragmentation step. Therefore, the number of reads directly reflects the number of transcripts. Moreover, 3'RNA-seq is more scalable and therefore powerful for performing analyses requiring many samples (Kremling et al., 2018). 3'RNA-seq has been used in several organisms, including plants. For instance, gene expression in seven tissues was measured across a panel of 299 maize lines, allowing the study of genetic diversity of gene expression (Kremling et al., 2018). However, when studying genetic variation, the use of the classical RNA-seq method can be beneficial because it is possible to detect genetic variants

within the sequencing reads and, therefore, directly in gene coding regions (Liang et al., 2022). The other RNA-seq biases due to the sequencing depth and RNA composition are also present in 3'RNA-seq and need to be considered in the data analysis (Figure 3B). It is worth noting that in 3' RNA-Seq, reads are concentrated at the 3'-end of transcripts. If these regions contain repetitive sequences, reads may map to multiple genomic locations; such multi-mapped reads are often discarded, which may result in underrepresentation of absence of certain genes in the analyses (Deschamps-Francoeur et al., 2020). Other powerful methods such as single-cell RNA-seq (scRNA-seq) and spatial transcriptomics allow gene expression studies at single-cell resolution, but their description is beyond the scope of this work.

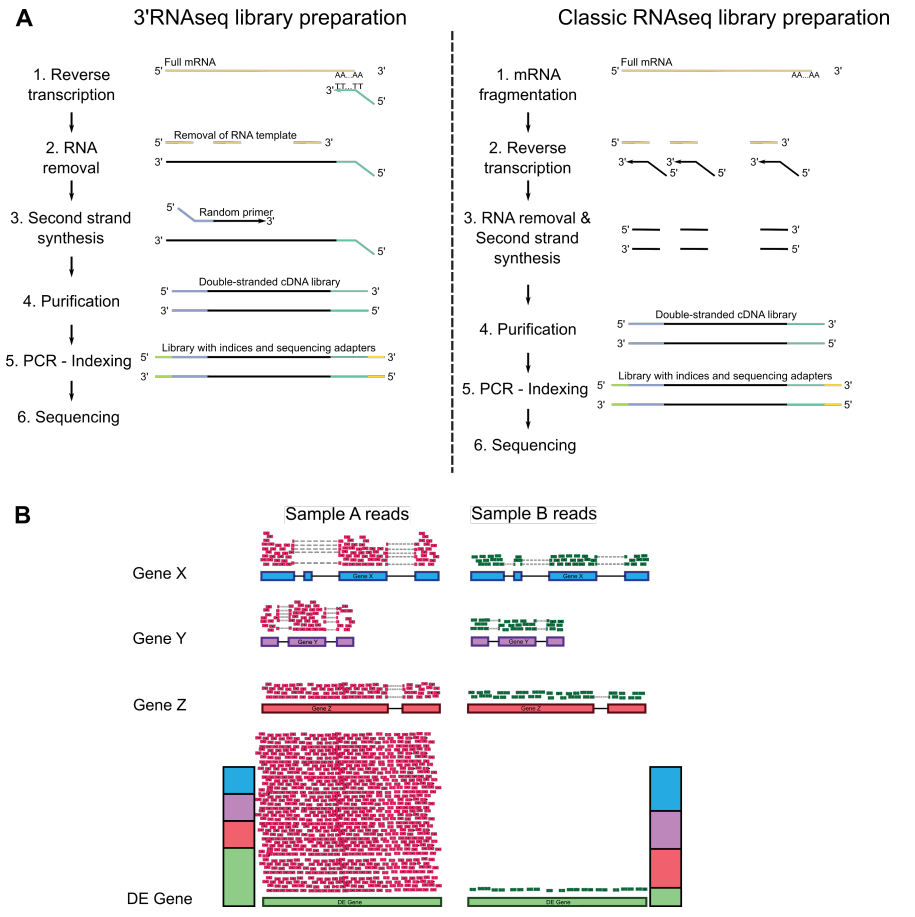


Figure 3: Comparison between RNA-seq and 3'RNA-seq. *(A) Library preparation workflow in 3'RNA-seq (left) and classic RNA-seq (right) (adapted from Ma et al. (2019)). (B) RNA-seq biases: sequencing depth; more reads were obtained for sample A than for sample B. Gene length; gene X is longer than gene Y and therefore has naturally more reads aligned to it. RNA composition; the high*

number of reads aligning to gene DE in sample A decreases the proportion of reads of the other genes compared to sample B (Adapted from hbctraining/DGE_workshop (2022)).

1.3 Gene regulation networks and genetic variations

Protein-DNA interactions between TFs and their target genes constitute complex gene regulatory networks (GRNs). These GRNs regulate phenotypes. Understanding how these components interact, how genetic variations affect their interactions, and the subsequent GRNs is crucial for crop improvement and trait selection (Springer et al., 2019).

Genetic variation refers to the differences in genotypes between individuals within a population or between different populations. It is observable through single-nucleotide polymorphisms (SNPs) and small insertions or deletions (InDels). These can trigger GRN rewiring, resulting in expression modification and phenotypic changes. Modifications in CRE DNA sequences can affect promoter, enhancer, or silencer activity, resulting in *cis*-regulatory variations. For example, the deletion of a region in the *ZmPHYB2* promoter, a red-light photoreceptor, decreased its activity, and individuals with that haplotype had an earlier flowering time (C. Li et al., 2023). An insertion in the *ZmPYL10* promoter, an abscisic acid (ABA) receptor, decreased its activity, which was positively correlated with plant and ear height (C. Li et al., 2023). Other genetic variations occurring in *trans*-acting factors also modify targeted gene expression and are named *trans*-regulatory variations. For example, a modification of a TF promoter can affect its abundance and, therefore, the expression of its target genes. Mutations occurring in a TF coding sequence can also affect its activity, either by impairing its DNA-binding capacity or by impairing its ability to bind to its cofactors (Springer et al., 2019). Transposable elements (TEs) are another major source of genetic variation. In maize, they account for 85% of the genome and have been shown to carry CREs (Schnable et al., 2009). For instance, maize domestication is the result of the insertion of the *Hopscotch* retrotransposon carrying an enhancer ~60k bp upstream of the promoter of the *TEOSINTE BRANCHED 1* (*TB1*) gene (Studer et al., 2011). This enhanced its activity, resulting in a higher apical dominance. Another example is the miniature-inverted repeat transposable element (MITE) insertion into the *ZmNAC111* TF promoter, which repressed its expression and induced drought tolerance (Mao et al., 2015). Similar studies have been performed in other species than maize, and together, they show that genetic *cis*- and *trans*-regulatory variations are major contributors to phenotypic diversity. TE movements can also trigger the relocation of enhancers carrying TF binding sites, thereby rewiring GRNs (Fagny et al., 2021).

1.4 Identifying expression regulators via genetic variations

The genetic architecture of complex phenotypes can be dissected by mapping quantitative trait loci (QTLs). These are regions in the genome associated with phenotypic variation and, therefore, contain potential regulators. When the phenotype is gene expression, these are called expression QTLs (eQTLs). This method has been applied in many plant species like Arabidopsis (*Arabidopsis thaliana* L.) (Keurentjes et al., 2007; Lowry et al., 2013), maize (Fu et al., 2013; Kremling et al., 2018; C. Li et al., 2023; Liang et al., 2022; S. Liu, Li, et al., 2020; Maistriaux et al., 2024), wheat (*Triticum aestivum* L.) (Samad-Zamini et al., 2017) or rice (*Oryza sativa* L.) (J. Wang et al., 2014). Two major methods can be used to map eQTLs: linkage mapping and genome-wide association mapping (Figure 4).

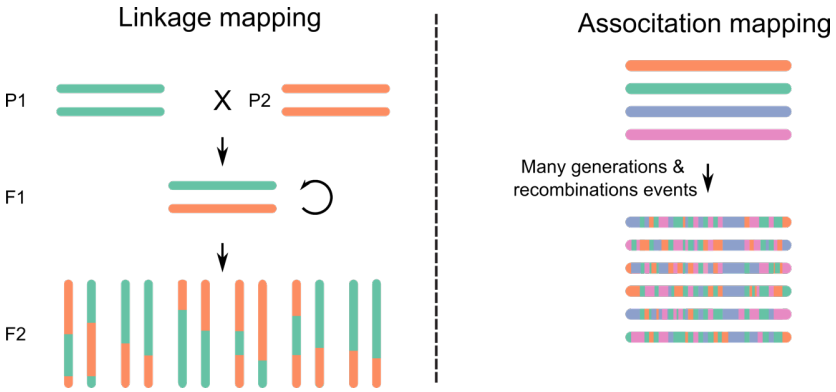


Figure 4: Comparison of linkage mapping (left) and association mapping (right). *Linkage mapping relies on the crosses between two parents (P1 and P2), and recombination events measured in the progeny can then be associated with the phenotype of interest. Association mapping relies on the study of a large natural panel containing individuals from several generations that underwent many recombination events. The LD in some regions allows association with the phenotype of interest at a higher resolution than linkage mapping.*

Linkage mapping relies on the analysis of manually generated populations of related lines. During their development, recombination events accumulate and produce genetic diversity. The measurement of this diversity, combined with the analysis of the phenotype of interest across the population, allows the mapping of regions that trigger phenotypic variations (Doerge, 2002). However, this method lacks resolution, as only large regions can be associated due to the limited number of recombination events that can happen during the population development (Y. Xu et al., 2017). Currently, the most prevalent method to dissect genetic architecture is association mapping using genome-wide association studies (GWAS), which rely on a principle similar to linkage mapping. However, they harness the power of natural genetic diversity accumulated through recombination events across many generations and use populations of *a priori* unrelated individuals (Dell'Acqua et al., 2015; McMullen et al., 2009; J. Yu & Buckler, 2006; C. Zhu et al., 2008). This results in a higher resolution for causal

variant mapping and higher allelic diversity. Moreover, there is no need to generate a whole population of individuals, dramatically reducing the research time.

1.4.1 Association mapping

Association mapping is based on linkage disequilibrium (LD), which measures the degree of non-random association between alleles at different loci in the studied population (Figure 5). For example, two biallelic loci should be randomly segregated by recombination in each generation. However, if these loci are close to each other, they will most likely be segregated together, meaning that they are in LD. Thus, causal variants responsible for phenotypic variation can be captured by markers detected in the population when they are in LD. This makes it possible to detect regions where potential regulators of the phenotype can be found (Figure 5). LD can be measured between all markers, and its decay with physical distance can be observed. LD decay is specific to each species and determines the resolution of the association mapping. In the case of slow LD decay, larger genomic segments segregate together, and the resolution is low. Fewer markers are needed to capture the genetic diversity in this situation. However, when LD decay is rapid, the resolution is higher, and a higher marker density is required to capture genetic diversity. In maize, LD decay is rapid, which makes it ideal for association mapping, as it allows single-gene resolution (H. Li et al., 2013). Moreover, maize has significant genetic and genomic resources simplifying trials and analysis (Hufford et al., 2021; Jiao et al., 2017; Schnable et al., 2009).

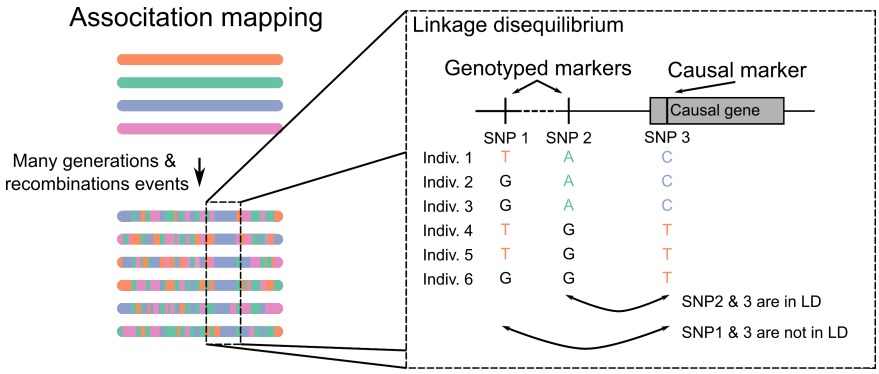


Figure 5: Theoretical situation where markers are in LD. *In this graph, SNP1 and SNP2 are markers that were observed in the panel of interest. SNP3 was not mapped but is the true causal variant. Because SNP2 is in LD with SNP3, it is possible to capture the causal variant. SNP1 is an example of a marker that is not in LD with the causal variant.*

Markers used in association mapping can come from several sources and a marker dataset is typically generated for the studied population. These are usually SNPs or InDels detected and mapped using techniques such as SNP-

arrays, genotyping-by-sequencing, and RNA-seq with high sequencing depth (S. Liu, Li, et al., 2020; Mabire et al., 2019; Negro et al., 2019).

GWAS can be performed using a wide variety of computational tools, such as FaST-LMM (Lippert et al., 2011; Listgarten et al., 2012), GEMMA (X. Zhou & Stephens, 2012), EMMAX (H. M. Kang et al., 2010) and P3D (Z. Zhang et al., 2010). While these tools all fit the same mixed linear model (MLM), they differ in their fitting algorithms. Describing these is beyond the scope of this work, but some of the concepts behind the use of MLMs in GWAS are relevant for the rest of this work, as they were used for model selection. MLMs are a specific type of linear model that combines fixed and random effects. To fit the MLM, two datasets are required: a vector containing the phenotypic values of the individuals in the population and a genotyping matrix containing their alleles for the markers measured for this panel. The general MLM equation is as follows:

$$Y = \mu + X\beta + Zu + e$$

where μ is the overall mean, X is the design matrix for the fixed effects, representing the genotypes, β is the vector of fixed effects, Z is the design matrix for the random effects, u is the vector of random effects, representing individual-specific genetic contributions, and e is the residual error term accounting for unexplained variation. The purpose of using an MLM in GWAS is to test the significance of the association between a given marker ($X\beta$; fixed effect) and variations in the phenotype of interest (Y), while controlling for false positives due to genetic relatedness (an example of phenotype marker association is displayed in Figure 6). Association mapping is theoretically based on unrelated individuals. However, in large populations, individuals inevitably show some degree of genetic relatedness, known as population structure. MLMs account for this by incorporating random effects (Zu), where u follows a normal distribution ($u \sim N(0, \sigma_g K)$) with a mean of zero and variance equal to the genetic variance (σ_g) multiplied by the kinship matrix (K), which measures the genetic similarity between individuals (Sul et al., 2018). Sometimes, more covariates are added to the model as fixed effects in the X matrix. Indeed, other methods to consider population structure include either adding a given number of principal components (PCs) to the genotyping matrix (Price et al., 2006) or calculating a structure matrix (Alexander et al., 2009; Pritchard et al., 2000). Moreover, unmeasured variables can affect phenotypic measurements, resulting in biased tests, these effects are known as batch effects. These can be taken into account by calculating hidden factors using tools such as PEER (Kremling et al., 2018) or by preprocessing the phenotyping matrix with a spatial model (Millet et al., 2016; Prado et al., 2018). To detect false-positive associations in a GWAS, a second GWAS can be performed with the most significant marker added as a covariate (Segura et al., 2012). This was done by Maistriaux et al. (2024), as some markers were in strong LD with the most significant SNP, which resulted in large eQTLs.

Genomic inflation factor is a useful metric for controlling false positives. It measures the extent to which the distribution of the observed test statistics deviates from that of the expected test statistics. A higher genomic inflation factor indicates the presence of false positives due to uncontrolled population structure, strong LD between markers, and systematic bias (Van Den Berg et al., 2019).

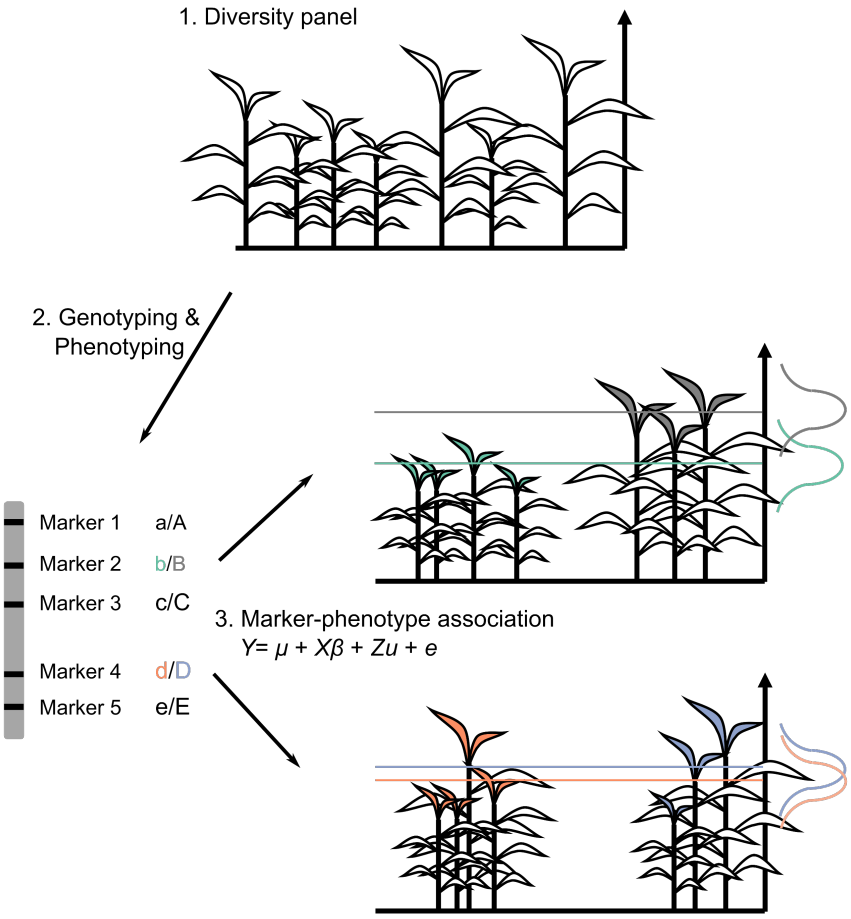


Figure 6: Representation of the classic GWAS workflow to detect markers associated with the height phenotype. *To begin, a diversity panel is constituted, then this panel is genotyped, and the phenotype is measured (the height of each individual is measured). Finally, the markers are associated with variation in the phenotype using an MLM. In this example, marker 2 is associated with, while marker 4 is not.*

1.4.2 Grouping significant markers in eQTLs

Once significant markers are detected, they can be grouped together into QTLs with the assumption that they capture the same causal variant. This can be performed using LD between markers. Many grouping methods can be found in

the literature. Some of them are based on genetic windows around markers. For instance, all the consecutive significant markers within 0.1 cM are grouped together in a QTL (Castelletti et al., 2020). Others are based solely on the physical distance between two consecutive significant markers. If two markers are close enough, they are clustered together, and the threshold can be large as 1M bp (G. Sun et al., 2023). Methods based on LD are also available. For example, if the pairwise LD between two significant markers is higher than a specific threshold, they can be considered to capture the same causal variant. These two methods are often combined (Fu et al., 2013; H. Liu et al., 2017). Negro et al. (2019) developed another method based on LD windows around each marker using the local recombination rate. For example, markers located in regions with high recombination rates would have a smaller window than markers located in regions with low recombination rates. Markers with overlapping windows are assumed to belong to the same QTL.

Subsequently, a list of all QTLs for the trait of interest is obtained. Expression QTLs provide insights into how genetic variation influences gene regulation and can be used in the construction of GRNs. TFs may appear in the list of candidate regulators, and TF-target gene interactions can then be characterised and functionally validated.

1.5 Methods to characterise expression regulators

Several methods have been developed to elucidate the molecular mechanisms involved in TF-target gene interactions. These are classified into three categories: gene-centred approaches, which identify TFs targeting a specific gene, TF-centred approaches, which identify target genes of specific TFs, gene co-expression networks, which establish expression correlation between genes (Mejia-Guerra et al., 2012). These categories often rely on generation of transgenic lines in which the expression of a given gene of interest is modified. These modifications can be achieved using many different techniques, for example, CRISPR/Cas9 technology can trigger precise modification of the genome. This technique is efficient and reliable in maize (Svitashev et al., 2016).

1.5.1 Gene-centred approaches

Gene-centred approaches include binding assays such as Yeast-One hybrid (Y1H), electromobility-shift assays (EMSA), and *trans*-activation assays such as dual-luciferase, dual-fluorescence gene reporter assays, and GUS assays.

Y1H and EMSA assess the direct binding of a TF to the DNA sequence of interest (i.e. a promoter). In Y1H, the sequence of interest is inserted into the genome of a yeast strain and used as a DNA bait. The yeast strain is then transformed with a construct encoding a TF (prey), and if the TF interacts physically with the DNA bait, the strain can grow on a selective medium. The yeast strain containing the DNA bait is usually transformed using a library of TF ORFs, allowing

simultaneous testing of many interactions at the same time. The surviving strains are then analysed to identify which TFs bound to the DNA-bait (Arda & Walhout, 2010; Walhout, 2006). This was performed to build a GRN for phenolic metabolism in maize (F. Yang et al., 2017). EMSA is based on the difference in migration speed between free DNA and TF-bound DNA during electrophoresis on polyacrylamide or agarose gels (Hellman & Fried, 2007). The sequence of interest is mixed with a specific TF, and the resulting mixture migrates on a gel. If binding occurs, a shift is observed because TF-bound DNA migrates more slowly than free DNA (example shown on Figure 7A). This method has been widely used to characterise protein-DNA interactions (Liao et al., 2017; M. Liu et al., 2023; Sakuma et al., 2002). However, these methods can result in interactions that do not occur *in vivo*, as the chromatin context is different from the organism from which the DNA fragment was cloned in Y1H and is non-existent in EMSA. In addition, they can give rise to false negatives when TFs need to be combined with other TFs or cofactors to bind to their targets (Walhout, 2011).

Trans-activation assays rely on the same principle as Y1H but are usually performed *in vivo*. Their principle is illustrated on Figure 7B. The promoter of the gene of interest is inserted upstream of a reporter gene. The reporter gene encodes the *firefly* luciferase in dual-luciferase reporter gene assay (McNabb et al., 2005), for a fluorescent protein in dual-fluorescence reporter gene assay, and the β-glucuronidase in the GUS assay (Jefferson et al., 1987). The promoter activity is then quantified by fluorescence in dual-fluorescence reporter gene assay or by enzyme activity in dual-luciferase reporter gene assay and GUS in presence and in absence of the TF. Genes encoding either *Renilla* luciferase in dual-luciferase reporter gene assay or a fluorescent protein in dual-fluorescence reporter gene assay, both under the control of a constitutive promoter, are used as internal controls to assess promoter activity in different conditions. For example, in the presence of a TF, if the promoter activity of the target gene increases or decreases, this suggests that they are likely part of the same GRN. These assays have been widely used to functionally validate regulators in GRNs (Studer et al., 2011; F. Yang et al., 2017; D. Zhu et al., 2014). However, biases can arise as the cell transformation technique can affect plasmid transfection efficiency or trigger GRN disruption, and TF expression can rewire the underlying GRNs or trigger variations in the internal control gene.

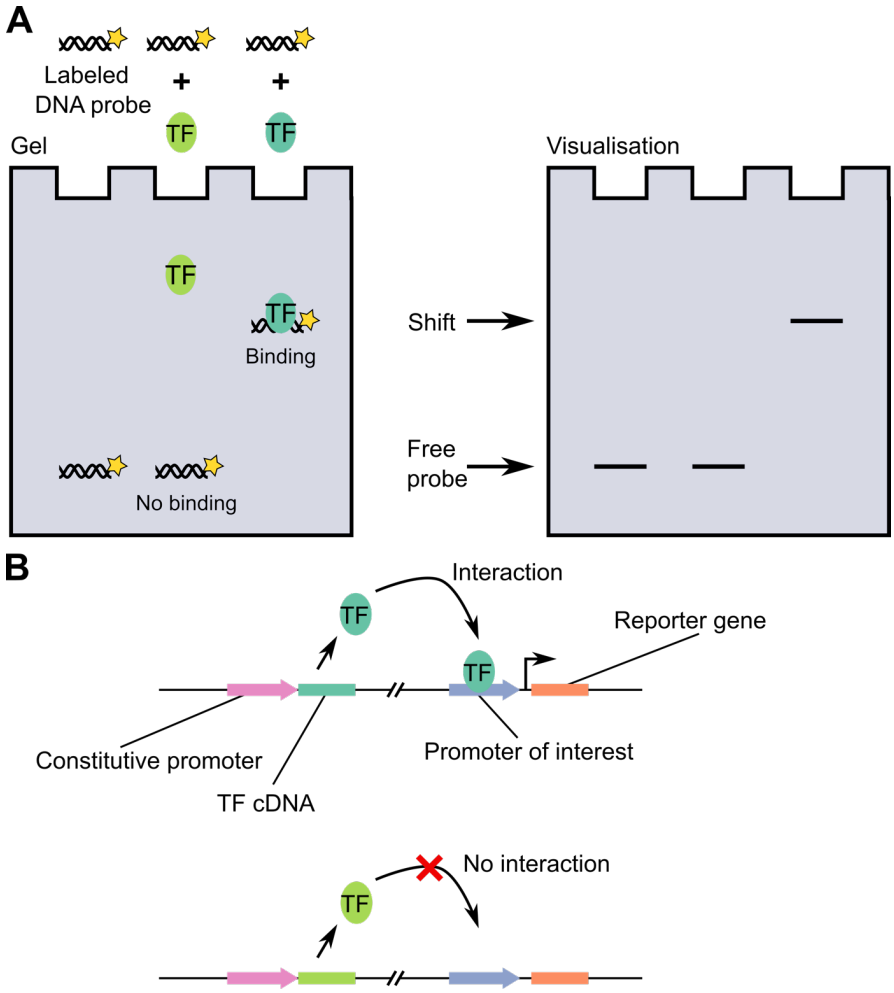


Figure 7: Example of methods used to validate a TF-target interaction *(A) In electromobility shift assays, fragments of the promoter of a gene of interest are labelled (with for example, 32P or cyanine) and then mixed with the TF of interest (light and dark green). If a binding occurs, the DNA-protein complex migrates more slowly than the unbound DNA, resulting in a shift. In the example, lane 1 is the negative control, where no protein is added, in lane 2, the TF does not bind the DNA and in lane 3, a binding occurs. The right panel shows the resulting gel after revealing the labelled DNA. (B) In trans-activation assays, the TF of interest (in dark and light green) are expressed under the control of a strong constitutive promoter (purple arrow). The promoter of the gene of interest (blue arrow), the target of the TF, is placed upstream of a reporter gene (in orange). When the TF interacts with the promoter of the gene of interest (dark green TF in upper panel), the activity of the reporter gene is modified. When there is no interaction (light green TF in lower panel), there is no change in promoter activity. The activity is measured through the reporter gene and compared to a condition in which no TF is expressed. This measurement is sometimes compared to an internal control (not represented here) and allows the validation of a TF-target gene promoter interaction.*

1.5.2 TFs-centred approaches

TF-centred methods aim to characterise the binding landscape of a specific TF by identifying all its target genes and binding motifs. Chromatin

Immunoprecipitation sequencing (ChIP-seq) is the most popular *in vivo* method (explained in Figure 8, left panel). Several ChIP-seq experiments were performed to identify maize target genes of TFs. For instance, these experiments helped to draw hypotheses about the biological functions of the KN1 TF (Bolduc et al., 2012) or contributed to the construction of the leaf regulatory network through ChIP-seq analyses of 104 TFs (Tu et al., 2020). Recently, the ChIP-seq protocol has been improved to overcome issues with crosslinking, chromatin solubilisation, and starting material, with the development of techniques such as cleavage under targets and release using nuclease (CUT&RUN (Skene & Henikoff, 2017)) and cleavage under targets and tagmentation (CUT&TAG (Kaya-Okur et al., 2020)). Zheng and Gehring (2019) showed that CUT&RUN could be performed in Arabidopsis. However, the need for specific antibodies raised against the targeted TFs constitutes the bottleneck of these techniques and reduces their scalability. Even though they could be performed using a transgenic line expressing a tagged-TF, its overexpression could disrupt genomic binding profiles (Kidder et al., 2011). Therefore, *in vitro* techniques such as systematic evolution of ligands by exponential enrichment (SELEX) (Jolma et al., 2010), protein binding microarrays (PBM) (Berger & Bulyk, 2009), and DNA affinity purification sequencing (DAP-seq, this technique is explained in Figure 8, right panel) (Bartlett et al., 2017) have been developed. These techniques are easier to set up for a large number of TFs because they do not rely on transgenic lines or TF-specific antibodies. O'Malley et al. (2016) performed a DAP-seq experiment involving 529 TFs to map all Arabidopsis CREs. Galli et al. (2018) investigated the binding landscape of the Auxin response factors family and assessed the binding motif redundancy between isoforms. The lipid metabolism regulatory network in *Camelina sativa* was also studied using DAP-seq (Gomez-Cano et al., 2022). Finally, the binding motifs of 63 Arabidopsis TFs were deciphered with PBM (Franco-Zorrilla et al., 2014). However, these methods can lead to erroneous binding due to the lack of chromatin architecture, the absence of cofactors, and the presence of a tag on the TF. Therefore, a combination of several characterisation methods is required.

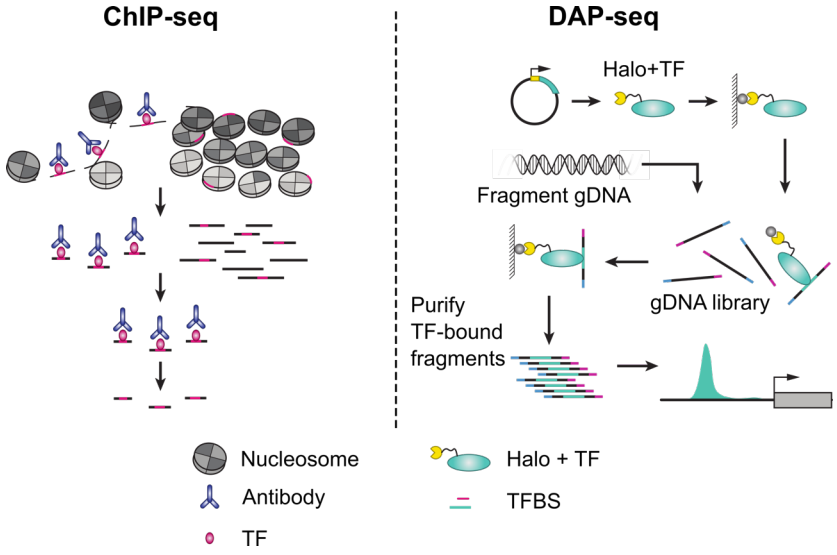


Figure 8: Schematic representation of ChIP-seq and DAP-seq experiments. *In the ChIP-seq experiment (left), the TF is crosslinked when bound to its targets* in vivo*, it is then recognised by a specific antibody, and the complex antibody-TF-DNA is purified. DNA fragments are then eluted and sequenced. The DAP-seq experiment (right) relies on a similar principle but is performed* in vitro*. The TF is tagged with a Halo-Tag that enables it to bind to magnetic beads. Once bound to the magnetic beads, a gDNA library is added to the mixture. The sites recognised by the TF will be captured at this step. After several washes, unbound DNA fragments are discarded. Finally, the captured gDNA fragments are released and sequenced. Both methods can be used to detect TF binding sites. (Adapted from Marand et al. (2023))*

1.5.3 Gene co-expression networks

Gene co-expression networks (GCNs) are another powerful tool to get information about underlying GRNs and help prioritise the characterisation of TF-target interactions (Schaefer et al., 2018). These methods are based on statistical inference and machine-learning techniques using gene expression data. For example, by measuring the transcriptome under several conditions, it is possible to detect hubs of genes with correlated expression variation and related biological function. This has been used in maize to uncover dynamic transcriptome changes during leaf development. The analysis showed that genes expressed at the leaf base were involved in basic cellular metabolism, whereas those towards the tip were related to photosynthesis (P. Li et al., 2010). The analysis of TFs within these networks not only indicates their potential function but also their potential targets (Haynes et al., 2013; P. Zhou et al., 2020). However, compared to GRNs, GCNs are based on expression correlations rather than validated interactions; therefore, they do not provide information about the validity of an interaction or the hierarchy of the genes in each biological pathway. Another notable limitation is that GCNs assume that a TF and its target genes are transcribed together, which is not always the case.

2. Aquaporins

2.1 Water transport in plants

Droughts create adverse conditions for plants due to the lack of water. This first results in osmotic stress, followed by oxidative stress, damage to cellular components, and molecular dysfunction. Understanding how water moves through plants is essential for identifying key factors that could enhance drought tolerance. As in other living organisms, the water content in plants is high and essential for their proper development (Boyer, 1982). Water plays various roles in physiological processes and metabolic reactions (e.g., as a solvent, reagent, carrier for nutrients, turgor maintaining, and thermal regulator). It is therefore important that water is well distributed throughout the plant body. Water moves passively through plant tissues along the water potential gradient ($\Delta\psi_w$), from a region of high-potential to a region of low-potential. Four components make up this potential: osmotic pressure (ψ_s), hydrostatic pressure (ψ_p), matrix potential (ψ_m), and gravitational potential (ψ_g) (Chaumont & Tyerman, 2014).

Long-distance water movement occurs during the day through transpiration when the stomata (small pores at the leaf surface) are open. The mechanism explaining this phenomenon is the cohesion-tension theory (Steudle, 2001). Opened stomata are gates for water evaporation because the water potential (ψ_w) is much lower in the atmosphere. Therefore, the water content at the leaf surface decreases along with ψ_w, triggering water movement from adjacent cells. This cell-to-cell water transfer reaches the tracheid walls, and the tension created causes an elevation of the water column inside the xylem vessels following $\Delta\psi_w$. At the root level, the xylem ψ_w is also decreased, and water moves from the root cells. As the ψ_w of the plant is generally smaller than that of the soil, there is an uptake of water at the root level in the maturation region of the root. Thus, there is a continuous hydraulic flow from the evaporating surfaces to the absorption surfaces facing the soil. When the stomata close, transpiration stops.

Within the xylem, water follows an axial flow and does not generally meet hydraulic resistance. However, to reach or exit xylem vessels in roots and leaves, water must move radially through several cell layers, a process that is subject to different hydraulic resistances (Steudle & Peterson, 1998). Water follows three different pathways, as described by the water composite transport model (Steudle, 2000): the symplastic path (through plasmodesmata), apoplastic path (through the cell walls), and transcellular path (through the plasma membrane) (Figure 9). The symplastic and transcellular paths are combined and named as the cell-to-cell pathway because it is impossible to measure them independently. Depending on the species, developmental stage, and environmental conditions, these three pathways contribute differently to water

movement. Evidence indicates that AQPs play a crucial role in facilitating water movement through the transcellular pathway in both roots and leaves (Chaumont & Tyerman, 2014; Ding et al., 2020, 2022; Heinen et al., 2009). Beside the involvement of AQPs in long-distance water transport, AQPs are also important for water homeostasis at the cellular level and regulation of cell expansion, stomatal movement, leaf epinasty, petal movement, etc (Azad et al., 2013; Heinen et al., 2009; N. Ma et al., 2008; Uehlein & Kaldenhoff, 2007; Xue et al., 2009).

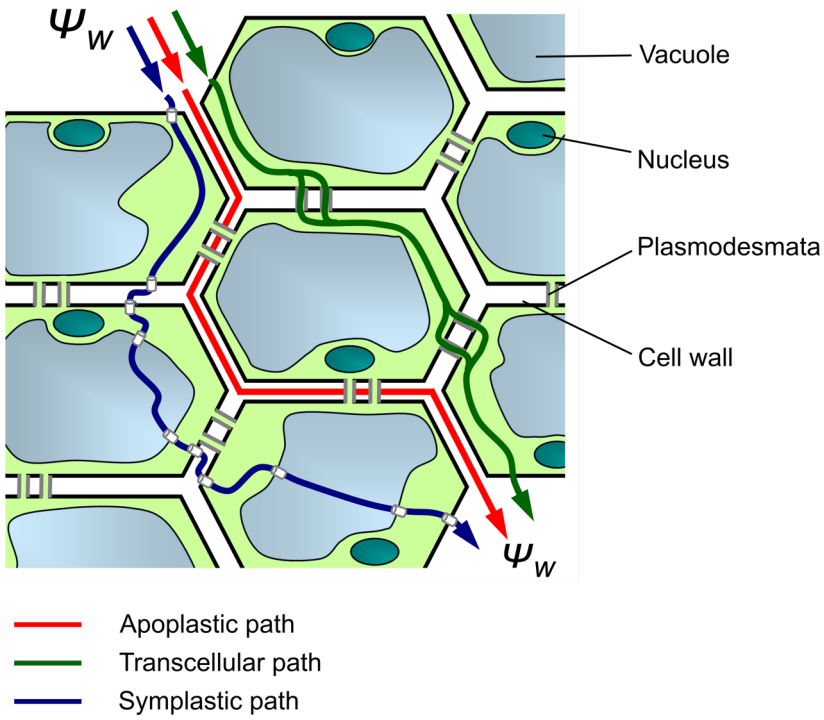


Figure 9: Pathways used by water to cross cell layers and reach xylem vessels. *The red arrow shows the apoplastic path, the green arrow the symplastic path, and the blue arrow the transcellular path. Water flow from high water potential to low water potential (ψ_w) (Figure prepared by R. Heinen, personal communication)*

2.2 Structure & phylogeny

AQPs are small (21–34 kDa) selective channel proteins present in most living organisms that facilitate the diffusion of water and small solutes, such as CO_2 (Uehlein et al., 2008), H_2O_2 (Bienert et al., 2007) or O_2 (Zwiazek et al., 2017), through the cellular membranes (Chaumont & Tyerman, 2014).

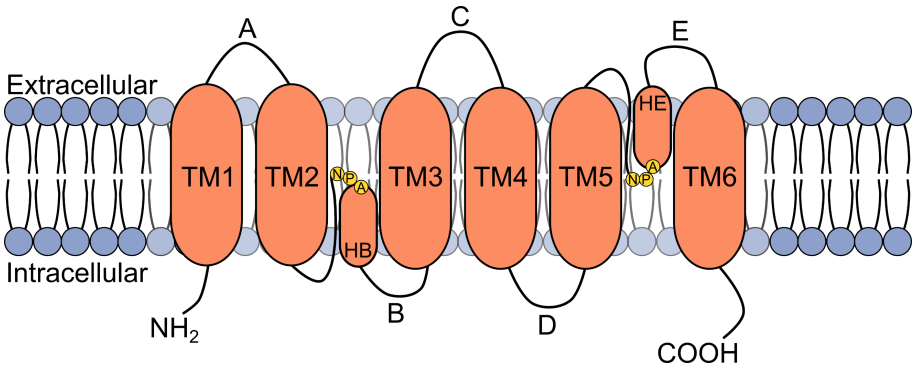


Figure 10: AQP structure. *Six transmembrane domains (TM1 to TM6) connected by five loops (A to E), and the N- and C-termini are represented. Part of loops E and B contain NPA motifs and forms short α-helices embedded in the membrane are noted HE and HB, respectively.*

AQPs consists of six α-helices crossing the membrane (called transmembrane domains 1 to 6 (TM1-6)) connected by five loops (A to E), with both the N- and C-termini facing the cytosol. Parts of loops B and E, made of hydrophobic amino acids, form short α-helices embedded halfway into the membrane (HB and HE, respectively) (Figure 10). These α-helices both contain asparagine-proline-alanine (NPA) domains, which, in addition to facing each other, create a size-exclusion filter (Mitsuoka et al., 1999; Murata et al., 2000). Another filter called aromatic/Arg (ar/R) composed of four amino acid residues is located at the external entrance of the pore. It plays a size-exclusion role and forms hydrogen bonds with the substrate owing to its hydrophilic interface. Along with the NPA domains, they constitute two constriction domains that control channel selectivity through charge, size, and polar exclusion mechanisms. These characteristics create a highly specific hourglass-shaped pore inside the membrane through which water molecules move in a single file. AQPs assemble as homo- or heterotetramers, and each monomer forms a separate pore channel (Fetter et al., 2004; Murata et al., 2000). Membrane pores can be gated through conformational changes. Depending on the phosphorylation status of the AQP, the D-loop can bind to the N-terminal end through polar interactions and occludes the pore (Törnroth-Horsefield et al., 2006).

An increase in reference genome sequencing has enabled the identification of AQPs in many species, indicating that plant AQPs constitute large gene families. Recently published AQP gene numbers are 41 for maize (Chaumont et al., 2001; Su et al., 2022), 84 for tobacco (*Nicothiana tabacum*) (Groszmann et al., 2021), 35 for Arabidopsis (Johanson et al., 2001), 65 for wheat (Madrid-Espinoza et al., 2018), 33 for rice (Sakurai et al., 2005), 41 for sorghum (*Sorghum bicolor* L.) (Reddy et al., 2015), 45 for cassava (*Manihot esculenta* Crantz) (Putpeerawit et al., 2017), and 42 for common bean (*Phaseolus vulgaris* L.)(Ariani & Gepts, 2015). AQP families have also been identified in many other species, which have

not been reported in this study. In vascular plants, AQPs are classified into five subfamilies based on amino acid sequence similarities and subcellular localisation, i.e., plasma membrane intrinsic proteins (PIPs), tonoplast intrinsic proteins (TIPs), small basic intrinsic proteins (SIPs), nodulin-26-like proteins (NIPs) and X-intrinsic proteins (XIPs). PIPs and TIPs are mostly located in the plasma membrane and vacuole membrane, respectively. SIPs are found in the endoplasmic reticulum. NIPs are present in the symbiosome membrane in legumes, plasma membrane, and endoplasmic reticulum. XIPs are in the plasma membrane but are not present in monocotyledons and Brassicaceae (Chaumont et al., 2001; Danielson & Johanson, 2008; Johanson et al., 2001).

2.3 Role of AQPs in leaf responses to drought stress

Leaves are vital organs for maintaining plant water homeostasis, as they are the primary sites for key processes such as photosynthesis and transpiration, both essential for plant growth and development. In monocots like maize, leaves present a specific type of stomata known as stomatal complexes, which consist of guard cells and subsidiary cells (Nunes et al., 2020). During the day, stomatal complexes open to allow the uptake of CO_2 for photosynthesis against the loss of water by evaporation. During the night or in response to a stress, they close to minimise water loss by reducing transpiration, thereby helping the plant conserve water.

The expression of PIP isoforms has been studied in many organs in maize, including roots, leaves, and stomatal complexes. Most *ZmPIP* genes are expressed in all organs, but their expression levels differ. In leaves, *ZmPIP* expression varied according to the developmental stage. Their expression was low at the base of the leaf, but then gradually increased in the elongation zone, peaking at the point of leaf emergence, where the evaporative demand increases as the leaf becomes exposed to the atmosphere (Hachez et al., 2008). In the mature zone, the expression of some *AQPs*, like *ZmPIP1;2*, *ZmPIP2;1, ZmPIP2;3, ZmPIP2;4, and ZmPIP2;5*, decreased, while the expression of others, like *ZmPIP1;1*, *ZmPIP1;3*, and *ZmPIP2;2*, stabilised, suggesting their involvement in photosynthesis-related processes or leaf movements (Hachez et al., 2008; Heinen et al., 2009). This pattern has also been observed for many other genes in GCNs (P. Li et al., 2010). *AQP*s expression also followed a diurnal cycle (Hachez et al., 2008; Heinen et al., 2014). This was also observed in the roots (Caldeira et al., 2014). The expression of some *AQP* isoforms is also tissue-specific. For example, *ZmPIP1;6* is specifically expressed in stomatal complexes, in contrast with other *ZmPIP* genes that are also expressed in other tissues of the leaf mature zone (Heinen et al., 2014).

These expression patterns suggest specific roles of AQPs in leaf physiology that have been intensively studied (for reviews, see Heinen et al. (2009); Chaumont

and Tyerman (2014); Li et al. (2014); Maurel et al. (2016); Shivaraj et al. (2021); Ding et al. (2024)). Owing to their ability to transport water and other small solutes, they are involved in many physiological processes, such as stomatal conductance (g_s; (Ding et al., 2020, 2022; Grondin et al., 2015; Uehlein et al., 2003), transpiration rate (Cui et al., 2008; Ding et al., 2020), leaf hydraulic conductance (Ding et al., 2020; Sade et al., 2014), CO_2 conductance (Heinen et al., 2014; Uehlein et al., 2003; C. Wang et al., 2016), detoxification of reactive oxygen species (ROS)(Bienert et al., 2007, 2014), ROS signalling (Boursiac et al., 2005, 2008; Ding et al., 2022; Grondin et al., 2015), and cell water homeostasis (Hachez et al., 2006; L. Wang et al., 2017).

2.3.1 Transgenic lines analysis

AQP involvement in such physiological processes makes them important players in drought stress tolerance (J.-K. Zhu, 2016). Several studies in transgenic lines have shown that *AQP* overexpression either improves their drought stress tolerance or increases their sensitivity. Cui et al. (2008) showed that Arabidopsis plants overexpressing *Vicia faba PIP1* had a greater drought stress tolerance by having faster stomatal closure, faster growth rate and lower transpiration rate. Similarly, overexpression of *ZmPIP2;5* in maize led to increased transpiration under well-watered conditions. However, under mild water deficit, transpiration and stomatal conductance decreased more rapidly compared to wild type plants. This was due to fact that *ZmPIP2;5* overexpressing plants responded more rapidly to ABA, with a quicker reduction in stomatal conductance, compared with wild type (WT) plants. In contrast, knocking down *ZmPIP2;5* resulted in slower stomatal conductance reduction, and transpiration remained higher than in WTs under mild water deficit, indicating reduced sensitivity to water stress (Ding et al., 2020, 2022). In tobacco (*Nicotiana tabacum* L.), a decrease in *NtAQP1* (*NtPIP1;5*) expression in antisense plants resulted in reduced root hydraulic conductivity (L_{pr}) and lowered water stress resistance (Siefritz et al., 2002). Similarly, Uehlein et al. (2003) showed that stomatal conductance was correlated with *NtAQP1* expression. Overexpression of common bean *PvXIP1;2* in Arabidopsis plants increased their drought stress. This was hypothesised to be due to reduced membrane damage and better osmotic adjustment in transgenic lines (Wu et al., 2022). The same hypothesis was drawn for the overexpression of bamboo (*Phyllostachys edulis* (Carrière) J.Houz.) *PePIP2;7 AQP* in Arabidopsis plants (H. Sun et al., 2021). *StPIP1* overexpression in potato (*Solanum tuberosum* L.) plants enhanced their drought tolerance by increasing their water use efficiency (L. Wang et al., 2017). In some cases, overexpression led to an increase in drought stress sensitivity. Overexpression of *Glycine soja TIP2;1* in Arabidopsis resulted in faster dehydration in transgenic lines during dehydration stress, indicating a decrease in drought stress tolerance (X. Wang et al., 2011). Aharon et al. (2003) highlighted that the overexpression of *AtPIP1;2* in tobacco, despite increasing plant vigour

under favourable condition, had a deleterious effect under drought stress. Wheat *TaPIP2;2* overexpression also compromised drought stress tolerance in Arabidopsis (C. Xu et al., 2013). However, studies demonstrating a decrease in drought tolerance have been performed by overexpressing *AQPs* in heterologous plant species. Therefore, the observed phenotype could not be fully attributed to the tested *AQP*. As mentioned in Li et al. (2014), the stress pathway of the host plant might not be able to recognise the AQP. Although these results are debatable, the overall observation is that *AQPs* can play a role in the drought stress response.

2.3.2 Drought-related traits GWAS

*AQP*s have been detected by GWAS in QTLs associated with physiological traits that are commonly measured to study drought stress tolerance in plants. Such analyses have been performed in several species and on traits such as survival rate after stress, root parameters (length, surface area, and number), and leaf parameters (leaf rolling, transpiration rate, and stomatal conductance) (S. Chen et al., 2023). Therefore, AQPs may be involved in this regulation.

In soybean (*Glycine max* L.), *GmTIP3;2* was found in a QTL involved in variations in the yield stability index, a metric that indicates the yield response of a specific genotype to water stress (Quero et al., 2021). In soybean, *GmTIP1;3* was also observed in a QTL associated with canopy wilting under long-duration drought stress (Saleem et al., 2022). Slow canopy wilting results in higher plant water use efficiency, a trait associated with drought tolerance (Fletcher et al., 2007). The authors suggested that this slow wilting resulted from a decrease in stomatal conductance and transpiration rate. In common bean, a marker located in an exon of *PvSIP1;3* was also found to be associated with wilting. Wu et al. (2022) identified *PvXIP1;2* in a QTL associated with survival rate after drought stress in common bean, and this was further validated by transgenic expression in Arabidopsis. The authors also highlighted that its expression was induced by drought, ABA, NaCl, and mannitol treatments. In *Medicago truncatula*, a homologue of *AtPIP1;2*, was associated with plant biomass, and its expression correlated with a decrease in plant growth under drought stress (Y. Kang et al., 2015). Although numerous GWAS and linkage analyses have been performed on drought-related traits in maize, the association between these traits and AQPs is less direct. For example, Rma1, a protein involved in AQP proteolysis (Lee et al., 2009), was negatively associated with drought tolerance (S. Liu, Li, et al., 2020), suggesting an indirect involvement in the drought stress response. Similarly, Chen et al. (2023) suggested that an association between a gene involved in autophagy and seedling plant height under drought conditions could trigger AQP degradation.

2.3.3 Drought responsive expression profiles

AQP expression profiles were responsive to drought stress. By analysing the transcriptome response in several species, it was shown that the *AQP* expression response was very diverse. Some *AQPs* were upregulated after drought stress in maize root meristematic and elongation regions (Opitz et al., 2016). Another study in maize showed that *ZmNIP2;2* was upregulated under drought stress in the roots, stems, and shoots (Y. Liu et al., 2015). In contrast, transcriptomic analysis of maize leaves revealed a decrease in *ZmPIP2;3*, *ZmPIP2;4*, and *ZmPIP2;5* gene expression from the end of the division zone through the elongation zone under drought stress, potentially to maintain leaf turgor (Nelissen et al., 2018). However, Cheng et al. (2024) identified more diverse *AQP* expression patterns. In an RNA-seq experiment performed on sweet potato (*Ipomoea batatas* L.) roots and leaves, they showed that some isoforms were upregulated and others down-regulated. Liu et al. (2020) did a similar observation when comparing the leaf transcriptome of drought tolerant and drought sensitive maize cultivars. Their transcriptome was analysed at four growth stages. In the drought-sensitive genotype, *ZmPIP2;5* was upregulated at the tassel stage (VT) and *ZmPIP2;3* was down-regulated at the kernel dough stage (R4, characterised by starch accumulation and a decrease in moisture content to approximately 70%). In the drought-tolerant genotype, *ZmPIP2;4* was upregulated at the VT stage and *ZmNIP1;1* was down regulated at the blister stage of the kernels (R2, characterised by kernels appearance on the ear with a moisture content around 85%). By comparing the transcriptomes of barley genotypes, Cantalapiedra et al. (2017) also showed that several *AQPs* were repressed in leaves after water stress. Eight *AQPs* were repressed in castor bean (*Ricinus communis* L.) leaves upon drought treatment (Jardim-Messeder et al., 2023). Some studies have focused specifically on *AQP* expression. The expression of *PvPIP1;2* in common bean leaves increased after drought stress, which was correlated with a decrease in transpiration rate (Aroca et al., 2006). In Arabidopsis, Alexandersson et al. (2005) monitored the expression of all 35 *AQP* homologues in response to drought stress using qRT-PCR and microarray. Among the *AQPs* expressed in Arabidopsis rosette leaves, the expression of all *PIPs* was downregulated, with the exception of*AtPIP1;4* and *AtPIP2;5*, which were upregulated, and *AtPIP2;6*, which remained stable. All expressed *TIPs* were also downregulated, and *AtNIP6;1*, *AtSIP2;1*, and *AtSIP1;1* remained stable. This analysis was also conducted on five Arabidopsis genotypes, and most PIP responses were consistent across genotypes, except for *AtPIP2;2*, *AtPIP2;5* and *AtPIP2;6*, which exhibited distinct responses to drought stress (Alexandersson et al., 2010). Additional analyses of small samples of genotypes displaying contrasting phenotypes related to water transport in roots and leaves also revealed significant *AQP* expression diversity between genotypes (Devi et al., 2016; Vandeleur et al., 2009). Interestingly, this natural variation in *AQP* gene expression has also been observed in larger panels, both in maize leaves

(Maistriaux et al., 2024) and roots (Rishmawi et al., 2023). Maistriaux et al. (2024) studied the expression of four *ZmPIP* genes in the elongation and mature zone of the leaf across a panel of 252 maize hybrids. They showed that the expression of *ZmPIP1;1*, *ZmPIP1;3*, *ZmPIP2;2* was stable across genotypes while the expression of *ZmPIP2;5* had a high variability.

It is difficult to assume that the observed physiological changes were solely due to AQPs, as many other genes were also differentially expressed. However, it is clear that AQPs are involved in drought stress response and that *AQP* gene expression patterns are isoform specific.

2.4 Regulators of AQP expression

Given the involvement of AQPs in crucial processes and their diverse expression patterns, they are interesting targets for improving water-related traits and drought tolerance in plants. Therefore, a number of steps have been taken towards understanding AQP regulatory networks. They are tightly regulated at multiple levels, as detailed by Chaumont and Tyerman (2014) (illustrated for PIPs in Figure 11). Regulation can occur at the transcriptional or post-translational level; however, in this thesis, we only focused on the regulation of their transcription.

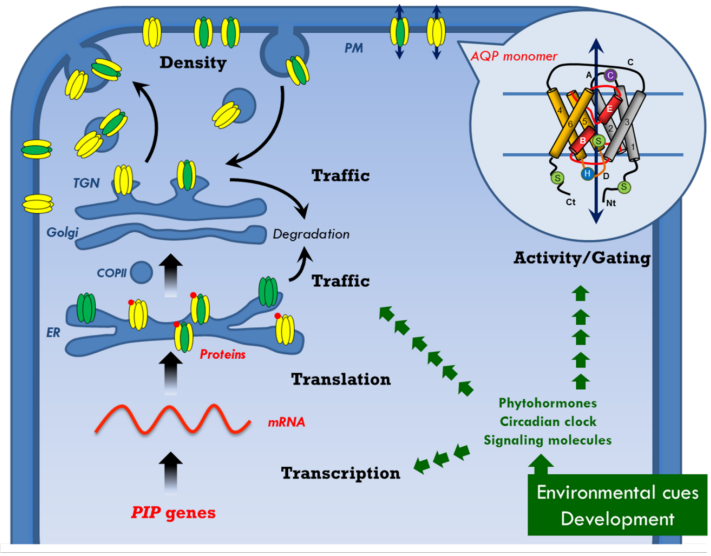


Figure 11: Schematic representation of the regulation of PIPs in cells. *The synthesis of PIPs is illustrated from the gene to the protein level, with PIP1s in green and PIP2s in yellow. Environmental cues can trigger modifications at the transcriptional, trafficking, and gating levels (Chaumont and Tyerman (2014)).*

In silico analyses of *AQP* promoter regions revealed the presence of CREs, such as ABA-responsive elements (ABRE), heat shock elements (HSE), MYB binding

sites (MBS), or dehydration responsive elements (DRE) (Putpeerawit et al., 2017; Reddy et al., 2015; Su et al., 2022). These are binding sites for TFs involved in abiotic stress responses, which potentially explains the drought responsive expression patterns observed in previous studies (Alexandersson et al., 2005; Cantalapiedra et al., 2017; Opitz et al., 2016).

In Arabidopsis, repression of a bZIP-type TF, *AtbZIP24*, reduced *AtPIP2;1* expression (O. Yang et al., 2009). Also in Arabidopsis, Rae et al. (2011) identified that *At*RAP2;4 and *At*RAP2;4B, two TFs belonging to the dehydration-responsive element binding (DREB) subgroup of the APETALA2/Ethylene Responsive Factor (AP2/ERF) family, were involved in the regulation of multiple *AQP*s by showing correlated expression in transgenic lines. The expression of other TFs belonging to this family also correlated with *AQP* expression in different species. *SlPIP1;2* and *SlTIP2-1-like* were significantly upregulated in *SlDREB2* overexpressing tomato (*S. lycopersicum* L.) plants, and their tolerance to salinity was enhanced (Hichri et al., 2016). Gu et al. (2017) expressed rice *OsDREB1B* in commercial strawberry and investigated the transcriptome changes between control and drought growth conditions. They found that *FvPIP2;1 like 1* was upregulated in the transgenic line and that its promoter contained a DRE cis-acting element. Even though these studies point out potential regulators of *AQP* gene expression, they do not indicate a direct regulation, but rather an involvement of *AQP* genes in the pathway regulated by the TF, as TF overexpression can disrupt the entire pathway in which it is involved. However, other characterisation methods have provided some evidence for direct regulation of *AQP* expression.

In several species, AP2/ERF TFs directly bind to *AQP* promoters and regulate their expression. In Arabidopsis, *At*TG enhances drought tolerance in transgenic plants and controls water balance by directly activating the expression of *AQP* genes, such as *AtTIP1;1*, *AtTIP2;3* and *AtPIP2;2* (D. Zhu et al., 2014). Heterologous expression of wild apple (*Malus sieversii* Ledeb.) *MsDREB6.2*, enhanced abiotic stress tolerance in Arabidopsis and drought tolerance in apples (*Malus domestica* Borkh.). Overexpression affected stomatal apertures, densities, and L_{pr}. In fact, AQPs were involved in the *MsDREB6.2*-mediated increase in L_{pr} that occurred under drought stress. In *MsDREB6.2* OE lines, *MdPIP1;3* and *Mdγ-TIP* were upregulated and *Mdβ-TIP* was downregulated, whereas in knockdown lines, *AQPs* were downregulated. EMSA showed that *MsDREB6.2* bound directly to the *MdPIP1;3* promoter (Liao et al., 2017). Zhang et al. (2024) explained the drought tolerance conferred by *Pb*ERF3 in pears (*Pyrus betulaefolia* Bunge) through the regulation of *PbPIP1;4*, which is involved in H_2O_2 transport, and its interaction with *Pb*HsfC1a, which induces the activation of *PbNCED4*, a TF involved in ABA biosynthesis. Its overexpression in Arabidopsis plants also enhanced their drought tolerance. It was shown that *Pb*ERF3 binds to the *PbPIP1;4* promoter by Y1H, EMSA, and activates *PbPIP1;4* transcription by

dual-luciferase reporter gene assay, indicating that *Pb*ERF3-*Pb*HsfC1a functions together as a transcriptional activator of *PbPIP1;4* and *PbNCED4*.

MYB TFs have also been found to regulate *AQP* gene expression. Jiang et al. (2020) performed a genome-wide identification of the *PIP* subfamily in poplar (*Populus trichocarpa* Torr. & A.Gray ex Hook). They performed *an in silico* analysis of their promoter and observed various CREs. One of these is an MBS, which is involved in drought inducibility and is very frequent within *Pt*PIP promoters. They generated *Pag*MYB216 overexpressing plants and observed that most *PtPIP*s were upregulated. They then performed a Y1H assay to test the binding of *Pag*MYB216 to the *PtPIP1;1* promoter and a dual-luciferase reporter gene assay to test its functional activity. They not only found that *Pag*MYB216 bound and regulated *PtPIP1;1* expression, but also that *PtPIP1; 1* upregulation had a negative effect on drought tolerance in poplar transgenic plants. Zhang et al. (2022) also showed that *Cs*MYB96 reduced water loss in citrus fruits and Arabidopsis and targeted several *AQP* genes. The regulation of *CsPIP1;1* and *CsPIP1;4* expression by *Cs*MYB96 was validated using EMSA and dual-luciferase reporter gene assay. The results showed that *Cs*MYB96 negatively regulated their expression, resulting in decreased water loss under drought stress. In bamboo, a co-expression pattern involving the *SnRK2* gene, *PeSAPK4*, MYB TF *PeMYB99*, and AQP *PeTIP4;3* was discovered (C. Zhu et al., 2024). The direct regulation of *PeTIP4;3* expression by *Pe*MYB99 was characterised using Y1H, EMSA, dual-luciferase reporter assay, and GUS activity assay. *Pe*MYB99 binds directly to MYB motifs located in *PeTIP4;3* promoter and activates its expression. It was shown that *Pe*SAPK4 interacted with and phosphorylated *Pe*MYB99, resulting in an increase in *PeTIP4;3* expression. Furthermore, overexpression of both these genes enhanced drought and salt tolerance in transgenic rice (C. Zhu et al., 2024).

TFs from other families were also observed to affect *AQP* gene expression. Liu et al. (2023) showed that *At*PLATZ4, an ABA responsive gene, can confer drought stress tolerance in Arabidopsis as *AtPLATZ4* overexpressing plants survived better than WT and knocked down plants. They observed that *AtPIP2;8* was downregulated in *AtPLATZ4* overexpressing plants. They confirmed *At*PLATZ4 binding to the *AtPIP2;8* promoter using EMSA, GUS, and dual-luciferase assays. Interestingly, Arabidopsis transgenic lines overexpressing *AtPIP2;8* exhibited impaired stomatal closure under drought conditions, which decreased their drought tolerance. The authors concluded that interaction between *At*PLATZ4 and *At*PIP2;8 modulates ABA-dependent stomatal closure.

TF binding to CREs located in *AQP* promoters were also observed via ChIP-seq experiments. These experiments made it possible to identify most of the TF target genes. In a ChIP-seq experiment performed on maize leaf protoplasts, Tu et al. (2020) showed that *AQP* genes were among the targets of 99 of the 104 TFs

tested. These TFs belong to families such as AP2/ERFs, NAC, MYB, and bHLH, which are involved in stress responses. Another ChIP-seq experiment performed on tomato leaves subjected to drought stress showed that AQPs are targeted by ASR1, another TF involved in drought stress responses (Ricardi et al., 2014).

AQP GRNs have not only been tested for *trans*-acting factors. Recently, Maistriaux et al. (2024) identified a CRE insertion in the promoter of *Zm*PIP2;5 in maize using GWAS. Significant markers indicated the presence of MITE in its promoter in the Ph207 maize cultivar, which was absent in the B73 cultivar. The comparison of both alleles in a DLF assay performed in epidermal cells of maize leaves showed that MITE insertion negatively regulated *Zm*PIP2;5 expression.

Research objectives

Genetic modifications that affect gene expression are responsible for phenotypic diversity. AQPs display a wide range of expression patterns and are involved in the regulation of drought-related traits such as stomatal conductance or transpiration rate, making them interesting targets for crop improvement. However, although some studies have identified *cis* and *trans* regulators of *AQP* gene expression, their understanding remains incomplete and requires further study. We will focus on *trans*-regulators, particularly TFs, because they sit at the top of biological pathways and control many downstream genes, making them ideal targets for modulating polygenic traits such as drought tolerance. Studies have identified *AQPs* as TFs targets using transgenic lines and characterised their functional regulation. However, most of these studies used TF-centred approaches and only focus on one interaction at a time. We hypothesise that studying TFs regulating *AQP* gene expression could reveal a more comprehensive regulatory network, including TFs that influence drought response and additional downstream pathways.

In this project, we aim to decipher the *trans*-regulatory variations in the transcription of *AQP* genes in the maize leaf mature zone in well-watered (WW) and water deficit (WD) conditions using an *AQP*-centred approach. To achieve this objective, we will use 3'RNA-seq data from the leaf mature zone of 254 maize lines obtained previously as part of a collaboration between our laboratory, the INRAE GQE team and Limagrain. We assume that genetic and bioinformatic approaches will enable the identification of the landscape of leaf TFs that modulates *AQP* expression in response to drought. This work will be divided into five parts.

First, we will perform genome-wide identification of *AQP* genes in maize to obtain an updated list of *AQP* genes and will also use the NCBI Blast tool to identify *AQPs* in newly published maize genomes. Secondly, the 3'RNA-seq data will be analysed and preprocessed to perform a GWAS. Several preprocessing methods will be tested and evaluated. This step will enable us to identify the genes that are differentially expressed between WW and WD and detect which *AQPs* respond to mild water deficit, allowing us to focus on the most interesting *AQPs*. In the third part, the expression data will be used in a GWAS to localise the eQTLs of *AQP* genes expressed in the leaf mature zone. These regions will indicate where potential *cis* and *trans* regulators are located. In the fourth part, we will build an *AQP* regulatory network by selecting the TFs located within the eQTLs and expressed in the 3'RNA-seq to analyse which TF target *AQPs*. Finally, the last part will aim to validate the observed statistical connections between the *AQP* genes and their potential regulators by *in vitro* and *in vivo* techniques and to functionally characterise the interactions.

This work will enable us to build the regulatory network of *AQP* genes in the maize leaf and to highlight the potential molecular mechanisms that regulate *AQP* gene transcription in various genotypes by exploiting their genetic diversity.

Results & Discussion

1. Genome-wide identification of *AQPs* in maize

The aim of this project was to update the *AQP* gene list for maize. With the publication of the B73 v4 and v5 reference genomes, we anticipated that we would be able to generate a comprehensive list of *AQP* genes in maize. In addition to the B73 v5 release, 25 other maize genomes were published in the NAM project (Hufford et al., 2021), and four were published by the Technical University of Munich (TUM) (Haberer et al., 2020), allowing us to expand our analysis to other maize lines.

An update on the *AQP* genes found in maize was described by Bari et al. (2018) at the beginning of this thesis. They updated the list from Chaumont et al. (2001) using the B73 v4 genome; however, there were many discrepancies in naming, duplications, and NCBI accessions. Therefore, we decided to perform a new analysis starting from the NCBI *AQP* accession list published by Chaumont et al. (2001) and another *AQP* characterised by Lawrence et al. (2013). Meanwhile, another update was published in 2022 by Su et al. (2022).

1.1 Identification of new AQPs and phylogenetic analysis

Our work identified 43 *AQP*s genes, representing 10 additional *AQP* genes than in the original list from Chaumont et al. (2001) and Lawrence et al. (2013). Several characteristics were used to categorise them as AQPs. First, the predicted TM domains were examined. Because AQPs are known to have six TM domains, we used DeepTMHMM to verify this feature. Next, we located double NPA motifs in these sequences. All had these conserved motifs; however, the newly identified *Zm*NIP4;1 had one NPA and one NPI motif. As it was already observed that substitutions could occur in the motifs, for example, *Zm*NIP3;1 has an NPS/NPV pair, *Zm*NIP4;1 was considered as an AQP. The last step consisted of searching the conserved domains using the CD-search tool from NCBI, which associated the sequence with a gene family based on these domains. In our case, the 10 new sequences were associated with the MIP superfamily to which all AQPs belong. We also concluded that the *ZmPIP1;3* and *ZmPIP1;4* genes described previously (Chaumont et al., 2001) were the same gene after several BLAST analyses, and *ZmPIP1;4* was therefore removed from our list.

Phylogenetic analysis was then performed using all AQP protein sequences (Figure 12). The sequences were grouped into four clusters according to four AQP subfamilies (NIPs, PIPs, SIPs, and TIPs), confirming that no XIPs were present in the maize genomes. Two new AQPs were grouped with PIPs, five with NIPs, and three with TIPs.

In general, PIPs showed the highest level of conservation. Newly identified PIPs were assigned to the PIP2 subgroup and named *Zm*PIP2;8 and *Zm*PIP2;9. The protein sequence of *Zm*PIP2;9 was very similar to that of *Zm*PIP2;6, which explains why the two are close to each other in the phylogenetic tree. On the other hand, *Zm*PIP2;8 was the most divergent PIP2s. The high level of conservation between PIPs could be the result of recent duplication events and/or selection limiting the appearance of new mutations within their sequence. Interestingly, this conservation is illustrated by their transport specificity as they display a more restricted range than the other AQP subfamilies (Fox et al., 2017).

NIPs were less conserved, and four subgroups could be distinguished instead of the three previously described ones. Two new NIPs were assigned to the NIP1 subgroup and named *Zm*NIP1;3 and *Zm*NIP1;4. Two additional NIPs were added to pre-existing subgroups: one to the NIP2s and another to the NIP3s. The last new NIP had a less conserved sequence, and we considered that it would represent a separate subgroup and therefore named it *Zm*NIP4;1.

The TIP subfamily consisted of five subgroups. One of the new AQPs clustered with TIP2s and was very similar to *Zm*TIP2;3. It was named *Zm*TIP2;4. The other two TIPs were grouped in the TIP3 subgroup and named *Zm*TIP3;3 and *Zm*TIP3;4, according to their proximity to the other TIP3s. The final list of AQPs is presented in Table 1.

Notably, the *Zm*PIP1;6 gene ID was absent from the B73 v5 annotation and was not detected in our BLAST analysis. We decided to BLAST its coding sequence on the genome file and found that it aligned in both B73 v4 and B73 v5. We also searched the NAM project files and found that the gene was located in B73, but no gene ID was attributed yet, whereas for other lines of the NAM project, a gene ID was attributed. Therefore, we decided to consider this gene as an *AQP*. Furthermore, physiological data in the laboratory have demonstrated that this gene is indeed an *AQP* (Heinen et al., 2014; Ding et al. in prep). The B75 v5 gene IDs have now been updated, and *Zm*PIP1;6 can now be found with the Zm00001eb372820 gene ID.

Table 1: Updated list of AQP genes in maize. *Positions are based on the NCBI annotation of version 5 of the B73 genome.*

Gene Name	B73 v4 ID	B73 v5 ID	NCBI Symbol	PanGeneID	Position (chr:start..stop)
ZmNIP1;1	Zm00001d016237	Zm00001eb239170	LOC542741	pan_gene_18108	chr5:151136346..151138494
ZmNIP1;3	Zm00001d037733	Zm00001eb283960	LOC100285584	pan_gene_21345	chr6:142368151..142369878
ZmNIP1;4	Zm00001d046303	Zm00001eb384790, Zm00001eb384800	LOC103640036	pan_gene_29074	chr9:81614032..81615316
ZmNIP2;1	Zm00001d018037	Zm00001eb254540	LOC542643	pan_gene_19254	chr5:215155268..215158776
ZmNIP2;2	Zm00001d037228	Zm00001eb279660	LOC541884	pan_gene_21003	chr6:123754366..123758140, complement
ZmNIP2;3	Zm00001d044826	Zm00001eb372380	LOC542497	pan_gene_28188	chr9:5793224..5796676
ZmNIP2;4	Zm00001d052261	Zm00001eb194530	LOC100192638	pan_gene_14688	chr4:187161627..187164292, complement
ZmNIP3;1	Zm00001d032461	Zm00001eb043650	LOC541885	pan_gene_3308	chr1:229359385..229362637

ZmNIP3;2	Zm00001d024634	Zm00001eb416440	LOC103642560	pan_gene_31468	chr10:83208610..83209940
ZmNIP3;3	Zm00001d049578	Zm00001eb172850	LOC100284417	pan_gene_9475	chr4:36302228..36303658
ZmNIP4;1	Zm00001d040108	Zm00001eb126240	LOC100282312	pan_gene_13031	chr3:28353194..28355004
ZmPIP1;1	Zm00001d002690	Zm00001eb074220, Zm00001eb074210	LOC542434	pan_gene_5590	chr2:19549332..19551311, complement
ZmPIP1;2	Zm00001d017526	Zm00001eb249940	LOC541779	pan_gene_18886	chr5:197669038..197672294
ZmPIP1;3	Zm00001d051403	Zm00001eb186900	LOC541886	pan_gene_14096	chr4:159005634..159009024
ZmPIP1;5	Zm00001d051872	Zm00001eb190950	LOC542014	pan_gene_99417	chr4:175098463..175099887, complement
ZmPIP1;6	Zm00001d020383	Zm00001eb372820	LOC541887	pan_gene_102421	chr9:7776028..7777383
ZmPIP2;1	Zm00001d019563	Zm00001eb306380	LOC541888	pan_gene_23030	chr7:42614459..42617478, complement
ZmPIP2;2	Zm00001d005421	Zm00001eb096680	LOC542644	pan_gene_7222	chr2:173064245..173067554
ZmPIP2;3	Zm00001d051174	Zm00001eb185300	LOC541889	pan_gene_13978	chr4:149013956..149016499
ZmPIP2;4	Zm00001d017288	Zm00001eb247760	LOC541890	pan_gene_18730	chr5:190826445..190829130
ZmPIP2;5	Zm00001d003006	Zm00001eb077130	LOC542619	pan_gene_5815	chr2:29372604..29374823, complement
ZmPIP2;6	Zm00001d019565	Zm00001eb306400	LOC541891	pan_gene_23030	chr7:42734583..42736042, complement
ZmPIP2;7	Zm00001d022608	Zm00001eb331830	LOC109941001	pan_gene_16958	chr7:184445254..184446669, complement
ZmPIP2;8	Zm00001d014285	Zm00001eb223900, Zm00001eb223910	LOC107648857	pan_gene_7222	chr5:39636026..39637196, complement
ZmPIP2;9	Zm00001d005410	Zm00001eb096620	LOC103647293	pan_gene_24982	chr2:172798793..172800184
ZmSIP1;1	Zm00001d048959	Zm00001eb167790	LOC542494	pan_gene_12649	chr4:10937043..10943018, complement
ZmSIP1;2	Zm00001d010325	Zm00001eb349840	LOC541892	pan_gene_26393	chr8:109600466..109605073, complement
ZmSIP2;1	Zm00001d028951	Zm00001eb015230	LOC542646	pan_gene_1185	chr1:52104364..52106929, complement
ZmTIP1;1	Zm00001d027652	Zm00001eb003730	LOC541675	pan_gene_287	chr1:10086567..10088583, complement
ZmTIP1;2	Zm00001d011778	Zm00001eb362170	LOC541893	pan_gene_27392	chr8:162759676..162760771, complement
ZmTIP2;1	Zm00001d051362	Zm00001eb186560, Zm00001eb186570	LOC541894	pan_gene_14069	chr4:157787227..157788499, complement
ZmTIP2;2	Zm00001d017485	Zm00001eb249500	LOC541895	pan_gene_100377	chr5:196242534..196243720, complement
ZmTIP2;3	Zm00001d026177	Zm00001eb429750	LOC541687	pan_gene_32489	chr10:141833690..141835020
ZmTIP2;4	Zm00001d002738	Zm00001eb074660	LOC103646110	pan_gene_5627	chr2:21258450..21259691, complement
ZmTIP3;1	Zm00001d013956	Zm00001eb221090	LOC541896	pan_gene_16741	chr5:26527062..26528165
ZmTIP3;2	Zm00001d032610	Zm00001eb044880	LOC541912	pan_gene_3413	chr1:233908519..233909803, complement
ZmTIP3;3	Zm00001d048520	Zm00001eb404170	LOC100280539	pan_gene_30580	chr9:160940705..160941981, complement
ZmTIP3;4	Zm00001d002952	Zm00001eb076690	LOC103646186	pan_gene_5778	chr2:28116472..28118444, complement
ZmTIP4;1	Zm00001d037779	Zm00001eb284350	LOC541897	pan_gene_21380	chr6:144263001..144265338
ZmTIP4;2	Zm00001d010410	Zm00001eb350360	LOC103635714	pan_gene_102082	chr8:113245402..113247720, complement
ZmTIP4;3	Zm00001d039713	Zm00001eb123230	LOC109280157	pan_gene_9263	chr3:13496267..13499135
ZmTIP4;4	Zm00001d039293	Zm00001eb119200	LOC542647	pan_gene_8949	chr3:1973580..1974598, complement
ZmTIP5;1	Zm00001d026178	Zm00001eb429760	LOC541900	pan_gene_32489	chr10:141835217..141836453

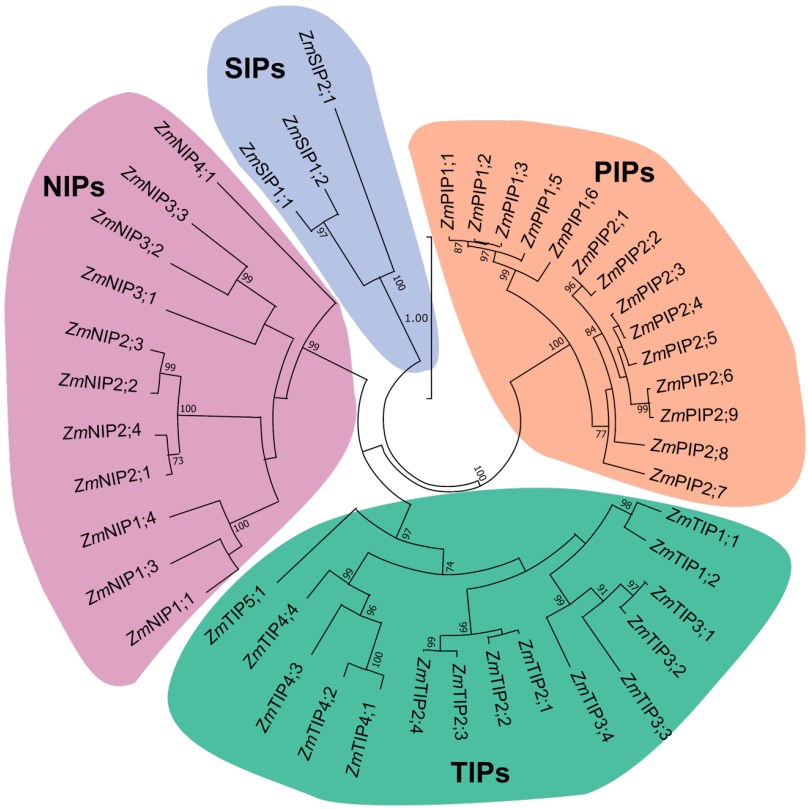


Figure 12: Phylogenetic tree of the AQP family in maize B73 cultivar. *Four subfamilies are observed with different background colours. The AQP protein sequences were aligned using the Muscle algorithm, and a tree was constructed using the maximum likelihood method based on the JTT matrix-based model with 1000 bootstrap replicates. These methods were implemented using MEGA X software.*

1.2 AQP families in several maize cultivars

Given the available resources, such as the 25 NAM genomes and four TUM genomes, we decided to constitute the *AQP* gene family in all these lines. Comparison of these *AQP* families in the different maize genomes showed that *Zm*PIP2;5 and *Zm*TIP2;2 were tandem duplicated in five and one genotypes, respectively (Figure 13). This raised interesting evolutionary questions, although addressing them was outside the scope of this work. These duplications may have arisen recently, as they are not restricted to individuals coming from one subpopulation. In addition, a comparison of the *ZmPIP2;5* and *ZmPIP2;5b* promoters in B97 showed a high level of conservation.

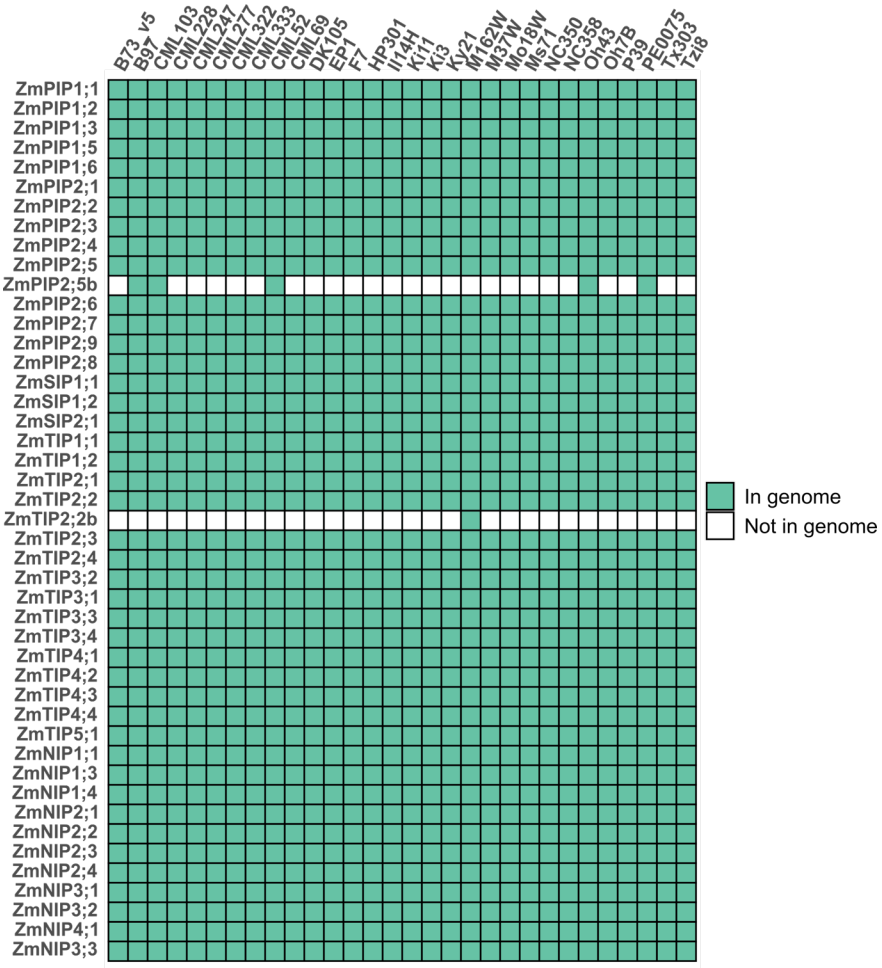


Figure 13: Genome comparison of the* AQP *gene families from 30 maize cultivars. *AQP protein sequences were obtained for the 25 NAM founder lines and four lines sequenced in the TUM project, by aligning B73 AQP sequences on their genome using BLASTp (Haberer et al., 2020; Hufford et al., 2021). Sequence from alignment hits with at least 70% query coverage, the highest percentage of identity, and the lowest score (like a p-value) were considered as AQP protein sequences. Each square represents the presence or absence of AQP (x-axis) in the reference genome (y-axis). When coloured green, AQP is present in the genome annotations, and when coloured red, AQP is absent.*

1.3 Comparison with previous genome-wide AQP identification

A comparison of our data with the most recent updates of the *AQP* gene family (Su et al., 2022) highlighted several differences.

One gene belonging to the NIP family, named *Zm*NIP3;3, was identified in our study, but was missing in Su et al. (2022). It was predicted to have six transmembrane domains and two NPA motifs (NPA/NPA), which is relevant for considering it an AQP.

In the PIP family, Su et al. (2022) discarded *ZmPIP1;4* and *ZmPIP1;6* genes. We agree with this decision for *ZmPIP1;4* because it is identical to *ZmPIP1;3*. However, *ZmPIP1;6* should still be considered a maize *AQP* gene. This gene was not initially annotated in the publicly available B73 genome v5. However, it has now been included despite being classified as a non-coding gene. In the framework of the NAM genomes project (Hufford et al., 2021), annotations for 25 other maize genotypes were published. Analysis of *AQP* genes in these genotypes revealed the presence of *ZmPIP1;6*. Moreover, *ZmPIP1;6* was not annotated as absent in the B73 annotation file, but as a gmap ID, indicating that the gene was present in the GMAP alignment at the time of mapping, but some criteria to classify it as a pan-gene were missing. It is now considered a pan-gene in the recently updated maize gdb database. As *ZmPIP1;6* is present in the B73 genotype and other maize genotypes, it must be considered an *AQP* gene. In addition, its expression was detected by RT-qPCR in maize roots, leaves, and stomatal complexes, and transport assay experiments in *Xenopus* oocytes and yeast cells showed that it acts as a water and CO_2 channel (Hachez et al., 2006, 2008; Heinen et al., 2014).

2. 3'RNA-seq data analysis

2.1 Diversity panel and experimental design

In the framework of the European DROPS (DROught-tolerant yielding PlantS) project, a diversity panel of maize hybrids was produced by crossing a common flint parent (UH007) with 254 dent inbred lines with a 10-day flowering window. These lines were previously grouped into six heterotic groups using ADMIXTURE software (Negro et al., 2019). The groups were named following their key founder, B73, Ph207, B14a, Oh43, Mo17, and W117-Type, and contained 37, 29, 38, 16, 36, and 98 hybrid lines, respectively (Supplemental Table 1). This panel has been used in several other experiments, some of which have been published (Maistriaux et al., 2024; Millet et al., 2016; Negro et al., 2019; Prado et al., 2018).

Before diving into more details about the 3'RNA-seq experiment, it is worth reporting some information about the experimental design, as its impact was assessed during the preprocessing of the data. The panel was grown on the phenotyping platform PhenoArch (LEPSE, INRAE, Montpellier (Cabrera-Bosquet et al., 2016)) hosted at the M3P (Montpellier Plant Phenotyping Platforms). The sowing took place in May 2013. The 254 maize hybrids were grown under two watering conditions: WW and WD. Each hybrid was replicated three times under both watering conditions. This gave a total number of 1,524 plants. The design is shown in Figure 14. The hybrid positions within replicates were randomised, but not between treatments. Mature zones of the 10th leaf at the 14-leaf stage of each hybrid were sampled within a specific time frame (from 7:20 to 10:30 AM) over three days to avoid biases due to the circadian rhythm. The sampling period

was divided into eight bins, and each sample was assigned to a bin according to the sampling rate and order. On day one, replicates one and two from the WW condition were harvested; on day 2, WW replicate three and WD replicate one were harvested; and on day 3, replicates two and three of the WD plants were harvested. Three replicates of 20 hybrids and two replicates of 234 hybrids were selected for the 3'RNA-seq experiment. Thus, of the 1,524 plants initially grown, 1,056 were used in this experiment. Selecting only two replicates for most of the samples in the 3'RNAseq experiment could have resulted in a loss of power, but this decision had to be taken given cost constraints.

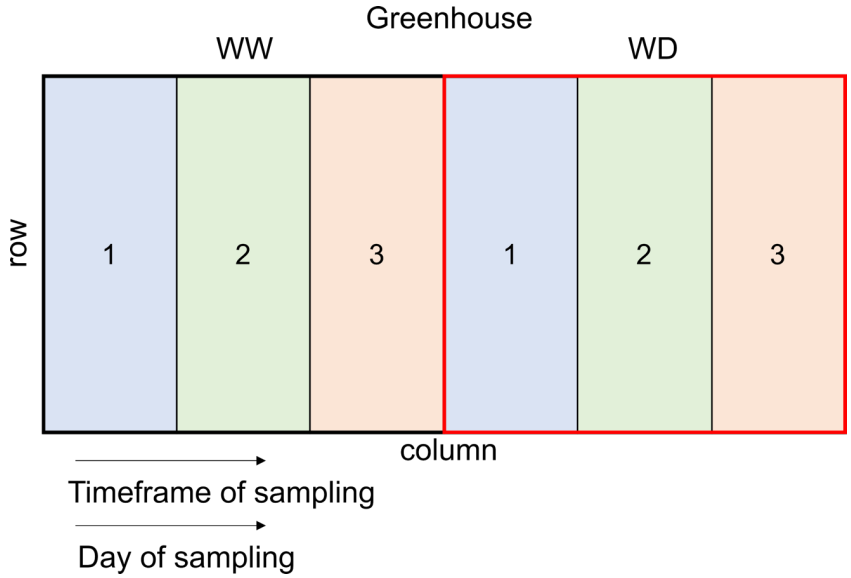


Figure 14: Schematic of the DROPS experimental design. *Plants cultivated in WW are shown in the left panel, and WD plants are shown in the right panel, circled in red. Numbers and colours indicate replicates. The rows and columns represent the positions of the plant in a Cartesian coordinate system. Plants were harvested from left to right, as indicated by the arrows.*

2.2 Identification of biases in the data and preprocessing

The purpose of this section is to describe in detail the various preprocessing steps and identify variables that have undesirable effects on gene expression, also known as batch effects. The first step was to analyse the homogeneity of library sizes (the total number of reads in a sample) (Figure 15). There was a pooling issue with libraries on sequencing plates 10 and 11 that resulted in over- and under-sequencing; thus, the library sizes varied drastically between samples compared to the other plates. Moreover, one sample (I211-WW-3) had a library with a very low number of reads compared to the others (4,788 reads while the average library size was 1,570,093 reads) and one sample was lost during sequencing due to technical problems (I267-WW-2). Therefore, the I211-WW-3 and I267-WW-2 samples, including their replicates (I211-WW-2 and I267-

WW-1), were discarded, resulting in a total of 1,052 samples. Although these samples were not retained in the final dataset, they were retained in the library size, normalisation, and transformation analyses for visualisation purposes.

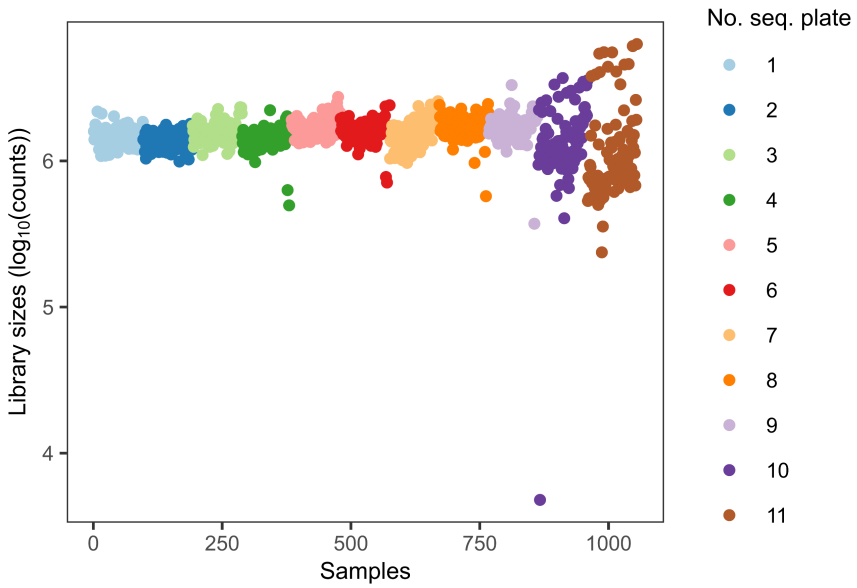


Figure 15: Library size across sequencing plates. *Each sample library size obtained by summing the total read count per sample is displayed by its $\log_{10}$ for visualisation purposes. The samples were coloured according to their sequencing plate to observe the variability arising from the plate. Plate are numbered from 1 to 11 and samples from 1 to 1056.*

2.2.1 Filtering low counts

A filtering step was then performed to remove genes that were not considered to be expressed in the dataset. In other words, genes with no or very low number of reads were removed from the count matrix. Conservation of these genes was unnecessary, as there was little evidence of activity, as a minimum level of expression is required for a gene before it is translated into a protein and has biological relevance (Love et al., 2014; Robinson et al., 2010). Furthermore, removing unexpressed genes saved computer memory and therefore sped up the following statistical analyses (Love et al., 2014; Robinson et al., 2010). Finally, in the case of DE analysis, retaining unexpressed genes could lead to an increase in false positives, since their number could be higher than the negative binomial law could assume.

Several filtering methods have been used in the literature. However, a nonstringent filter was chosen. All genes with less than one count per million in 10% of the samples were discarded. The count per million (cpm), before normalisation, is defined as the number of reads of a gene in a sample divided by the library size. This definition changes after normalisation because the

normalisation factors must be considered. In this step, where the purpose was to visualise and identify biases in the data, the filter was applied to the entire dataset. Approximately 20,500 expressed genes and 32 *AQP* genes were identified in the maize leaf mature zone under both conditions. Before filtering (Figure 16A), we observed that the low-expressed genes accounted for the majority, but after filtering (Figure 16B), their proportion drastically decreased.

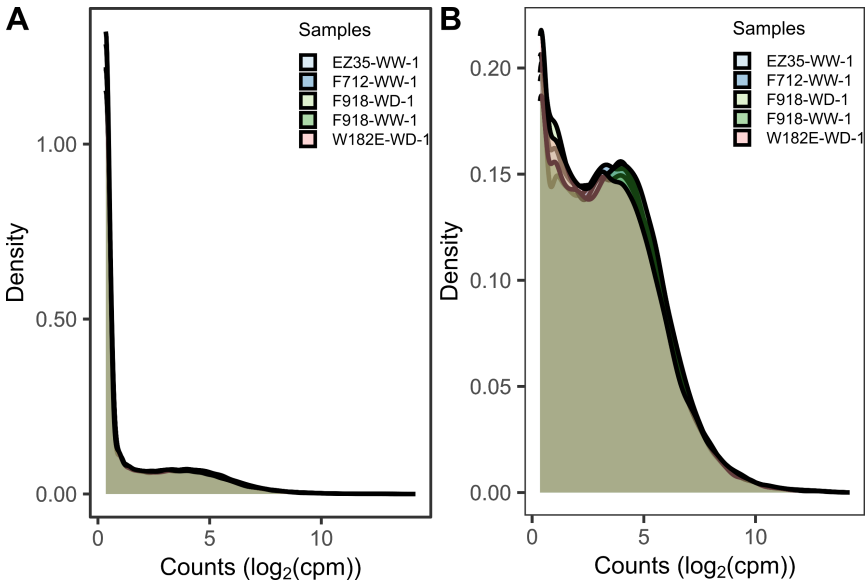


Figure 16: Distribution of the log_2(cpm) of all genes for five sample chosen randomly in the panel. *The distribution for genes with less than 1cpm in 10% of the samples before filtering (A) and after filtering (B) is shown. The colours differentiate the selected samples.*

2.2.2 Normalisation

Normalisation of the RNA-seq data was required to remove biases due to RNA composition, gene length, and sequencing depth (Evans et al., 2018). Two widely used methods were identified: relative log expression normalisation (RLE) implemented in the DESeq2 package (Anders & Huber, 2010; Love et al., 2014) and trimmed mean of M-values normalisation (TMM) implemented in the edgeR package (Robinson et al., 2010; Robinson & Oshlack, 2010).

The gene expression distribution per sample before normalisation is shown in Figure 17. There was a clear technical bias for the samples that were sequenced on the last sequencing plates, as their distribution varied significantly compared to the other samples (Figure 17A). After normalisation, the bias was considerably reduced (Figure 17B). We could still observe the I211-WW-3 sample with a very low library size by its biased distribution.

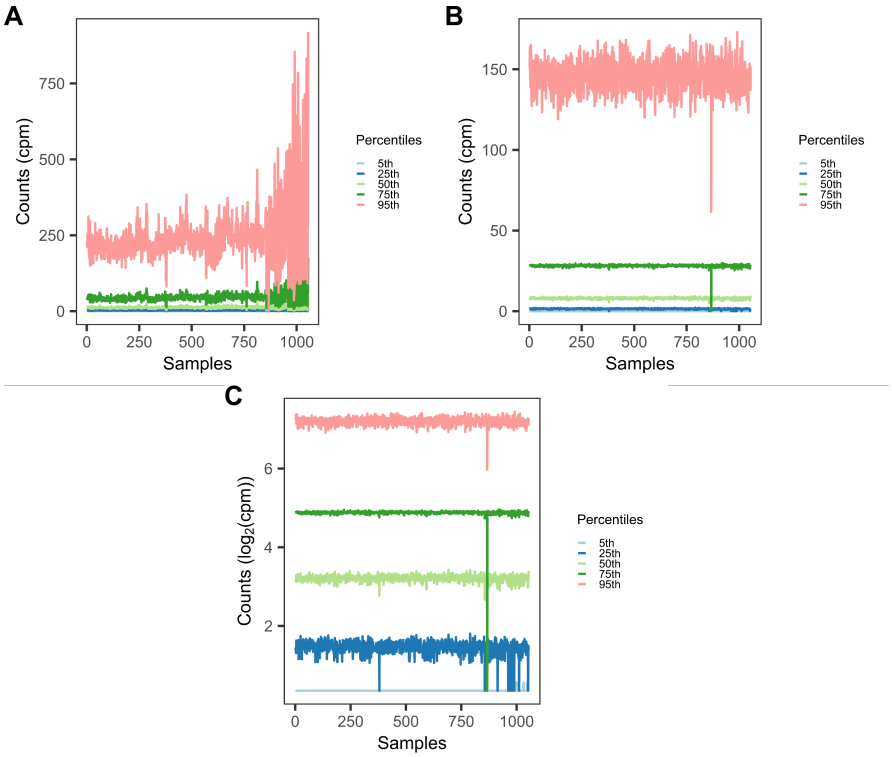


Figure 17: Gene expression distribution for all the studied samples. *The distributions are displayed using their 5, 25, 50, 75, and 90th percentiles values on the y-axis. The x-axis represents the sample numbered from 1 to 1056. Distributions are shown at different preprocessing steps: before normalisation (A), after normalisation but not transformation (B), and after normalisation and transformation (C). Units vary according to the preprocessing step from cpm when no transformation was performed to log$_2$(cpm) after transformation.*

2.2.3 Transformation

RNA-seq data (and 3'RNA-seq) is known to follow a negative binomial distribution, with many genes having a low number of counts and some genes having a very high number of counts. The transformation step was required for visualising the data or principal component analysis (PCA). Three methods (vst, log-cpm, and voom) from different R packages were tested in the preprocessing (Love et al., 2014; Ritchie et al., 2015; Robinson et al., 2010), each having advantages and disadvantages. At this step, normalisation and transformation were conducted using the edgeR package. The effects of normalisation and transformation on the distribution of gene expression per sample are shown in Figure 17C. After transformation, reproducibility was assessed by calculating the Pearson correlation coefficients of the B73 samples (Table 2 and Supplemental Table 2). The average expression correlation coefficient of the biological replicates was 93%, indicating high repeatability during sample collection.

Table 2: Pearson's correlation between samples belonging to B73. *The first two columns show the sample ID between which the correlation was calculated.*

Sample 1	Sample2	Pearson correlation coefficient
B73-WW-1	B73-WW-2	0.90
B73-WW-1	B73-WW-3	0.94
B73-WW-2	B73-WW-3	0.89
B73-WD-1	B73-WD-2	0.93
B73-WD-1	B73-WD-3	0.95
B73-WD-2	B73-WD-3	0.93

2.2.4 Principal component analysis (PCA)

To verify the presence of biases due to the position of the plants in the greenhouse, a PCA was performed. The 3'RNA-seq was normalised using the TMM method and transformed using the log-cpm method, and the 500 genes with the highest variance among the samples were selected. The PCA results for the first two PCs are shown in Figure 18. Together, these two PCs explained 24% of the data variance. Each point and colour on the PCA plots represent a sample and the group to which the sample belongs depending on the variable of interest, respectively. When the samples were coloured following the day of sampling (Figure 18A), their heterotic group (Figure 18B), no clear group was observed. When coloured following the treatment, a slight division was observable on the right part of Figure 18C, but no clear group could be observed. This means that neither of the first two PCs was related to these variables; hence, the variance in the data originated from a different variable. The fact that the sampling day had little or no effect on the data was reassuring. The treatment was correlated with the third PC, explaining 5% of the variance. This was unexpected, as the treatment was our tested variable. We further investigated the pattern observed in Figure 18C by colouring the samples with the combination of the day, the treatment and the replicate. We observed that the two groups in Figure 18C were replicate 1 in WW and replicate 1 and 3 in WD (Supplemental Figure 1). These replicates were the last replicates to be collected each day, and therefore, the division occurring between the two groups could be related to the time of sampling. When the samples were coloured according to the time bins, a clear pattern was observed, and there was a strong correlation between PC1 and the bins (Figure 18D). This supposedly resulted from the sampling step of the experiment. We hypothesise that the time bins also explain the two groups between WW and WD located on the right side of Figure 18C. This suggests that later in the day, there could be a greater difference between the two conditions. This could happen because during the day plants transpire and therefore lose water, as the day continues the stress condition can increase, which is reflected in the gene expression data. As mentioned earlier, the time frame of sampling was kept as short as possible, but the influence of circadian rhythm on gene expression may still have had a significant impact. Hopefully, several methods exist to take these batch effects into account and recover the variance explained by the treatment. These will be mentioned in the following analyses.

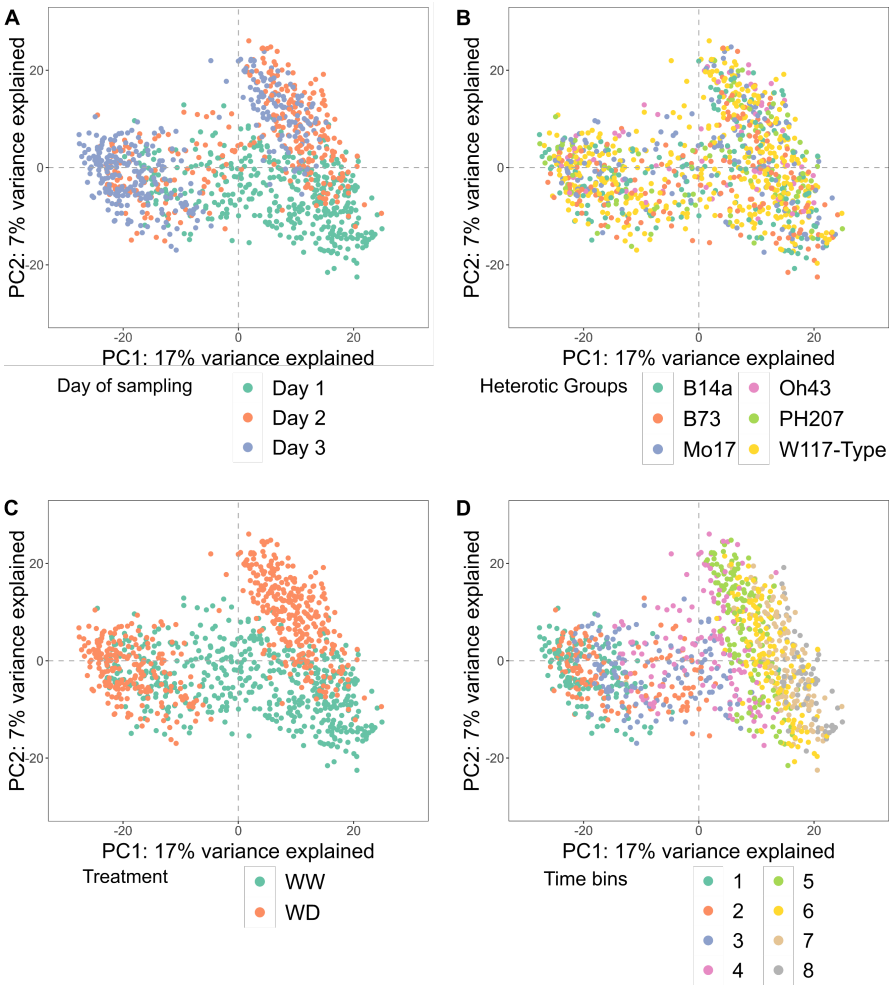


Figure 18: Principal component analysis of the normalised and transformed expression data in log_2(cpm). *The PCA analysis was performed using the 500 genes presenting the highest variance in expression between the samples. The results for the first two PCs are shown with the percentage of variance explained by each of them. Each dot represents a sample. The samples are coloured following the day of sampling (A), their heterotic group determined by ADMIXTURE (B), treatment applied to the plants (C), and time bins representing the time samples were collected (D).*

2.3 Differential expression analysis between WW and WD

To evaluate the transcriptional response to drought stress in the panel, we performed a differential analysis between WW and WD. Hybrids with missing samples in one condition were excluded, thereby reducing the number of lines in the panel to 252. Batch effects observed in the PCA were corrected using the Combat-seq method from the sva package, and then the gene expression dataset was separated by line. The filtering was then performed to keep genes with at least 1 cpm in 10% of the samples, resulting in different numbers of

expressed genes per hybrid. Raw counts were normalised using the TMM method, and a quasi-likelihood F-test was performed to identify differentially expressed genes (DEGs), as recommended by the edgeR guidelines for experiments with a low number of replicates (Lun et al., 2016). No transformation was required as the package was designed to handle data following a negative binomial distribution. Results for each line were extracted and compared across the whole panel. Gene IDs of genes with FDR-corrected p-value < 0.05 in at least 15% of the hybrids were considered as differentially expressed. This threshold of 15% was chosen to represent approximately one heterotic group (~40 hybrids). This method enabled the identification of 115 DEGs. This is a low number compared to other DE analyses, but it could be explained by two reasons: (1) the water deficit applied was mild, and the response might be too, and (2) the number of samples per hybrid is low, which can reduce the power to detect DE genes, and we used a conservative threshold to control for false positives. Gene ontology enrichment analysis was performed with GOATOOLS using maize-Gamer annotations. It revealed several enriched GO terms associated with biological processes involved in drought stress (i.e. GO:0009414: "response to water deprivation", GO:0009628: "response to abiotic stimulus", GO:0006970: "response to osmotic stress") (Supplemental Table 5). This provided evidence of a response to water stress in the panel. Among the DEGs, one gene (Zm00001d027652) attracted our attention because it encodes *Zm*TIP1;1 AQP. It was observed as a DEG in 24% of the lines and was mainly downregulated in WD. Other AQPs were differentially expressed but in fewer hybrids. For example, *ZmPIP2;5* and *ZmTIP4;1* were differentially expressed in 11% of the hybrids and *ZmPIP2;6* in 10%. In general, *AQPs* were downregulated in WD condition. This downregulation of *AQP* gene expression in leaf upon drought stress has already been observed in other studies as highlighted in the introduction.

Figure 19 shows a heatmap of the log-fold change (LogFC) between the WD and WW conditions for the 115 DEGs observed in our dataset. Regarding the gene clusters (1 vs. 2), we observed two major groups: genes globally upregulated in WD (A1, B1, C1), and genes globally downregulated in WD (A2, B2, C2). Some genes from the second group were considered to be more stable across the panel. The hybrid clusters did not follow the heterotic groups (A vs. B vs. C); therefore, we could not observe any response due to specific families. However, some diversity was observed among these lines. A group of lines located in the middle of the heatmap (B) showed less response to WD as the LogFC compared to the other two groups. Lines belonging to A and C groups showed a similar pattern with genes strongly upregulated and other downregulated, but these genes were different in both groups, some of them being more upregulated in group C than in group A.

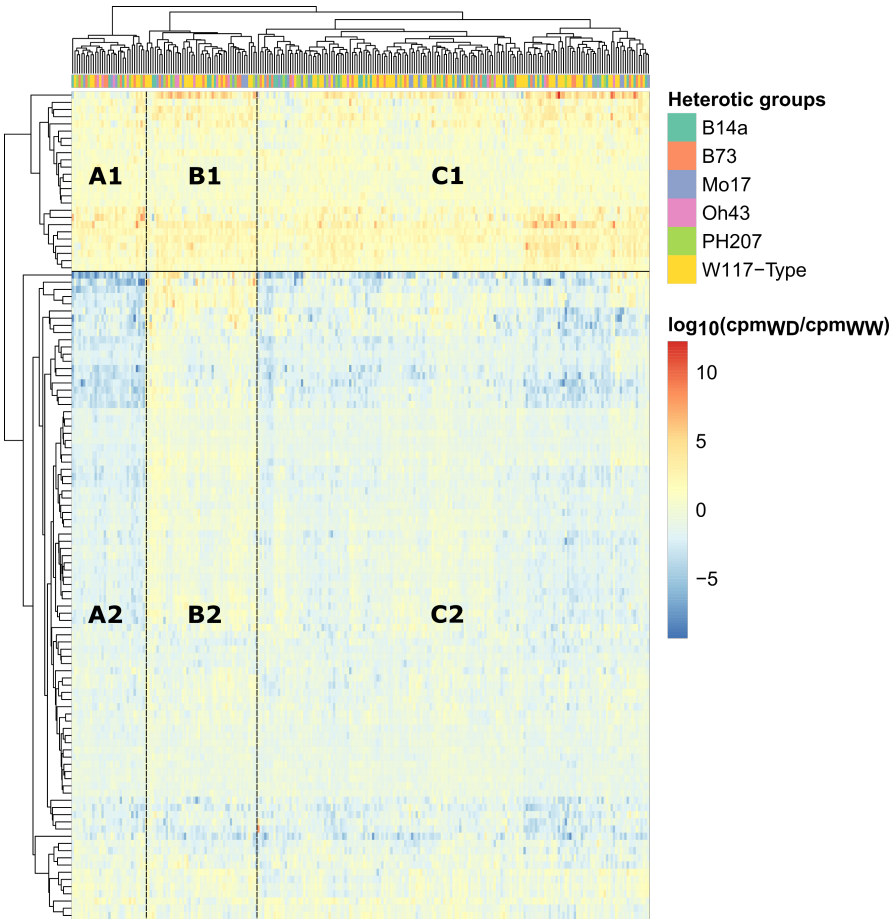


Figure 19: Heatmap showing the log_{10}FC between WD and WW conditions for the 115 DEGs. *Hierarchical clustering was performed between genes and hybrids using the complete linkage method based on between-sample Euclidean distance. The different groups are shown using the letter and number combinations. The log_{10}FC is represented by the blue-red gradient and the six heterotic groups determined by ADMIXTURE are associated with one colour.*

A PCA analysis was also performed on the batch-corrected data, and the treatment explained 8% of the variance and was correlated to the first PC (Supplemental Figure 2). The effect of the time bin also disappeared. We believed that adjusting the data increased the power of the experiment, as when the same analysis was performed on the original data, the same top genes were identified, but their FDR corrected p-value was just above the threshold, and they were not considered significant.

2.4 Choice of preprocessing methods

Another purpose of generating this dataset was to conduct a GWAS to map the regulatory variants involved in gene expression responses to drought. For this part of the project, all lines were maintained regardless of the number of

replicates, resulting in 252 and 254 hybrids in the WW and WD treatments, respectively. As previously mentioned, the 3'RNA-seq data followed a negative binomial distribution and were therefore not suitable for entry into the mixed-linear models used in GWAS. Moreover, to merge the replicates, a spatial model was used instead of Combat-seq. The R package SpATS was chosen to perform this, which also relies on mixed-linear models. Therefore, the 3'RNA-seq data needed to be transformed before correction for spatial effects and GWAS.

We used a pipeline similar to that used for the DE analysis, which included steps such as filtering, normalisation, and transformation. These steps were followed by spatial correction to remove batch effects. Here, the filtering was performed separately for each watering condition, resulting in 20,475 and 20,642 genes with 1 cpm in at least 10% of the WD and WW samples, respectively. Of these genes, 20,221 were expressed under both conditions, while 254 and 421 were specific to WD and WW, respectively (Figure 20). A GO analysis was carried out on the gene set specific to each analysis; however, after p-value correction, no terms were significantly enriched. Compared to the DE analysis, several normalisation and transformation methods were assessed to determine if some performed better than others.

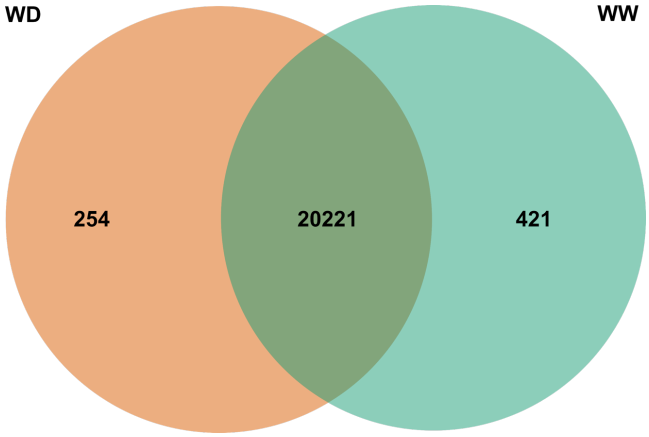


Figure 20: Venn Diagram of genes expressed in the dataset. *The numbers of genes expressed in WD and WW are represented by orange and green circles, respectively. The number of genes expressed under both conditions is shown at the intersection of the circles.*

Given the number of methods to be tested at each step, the number of combinations to be compared was large; however, this work was necessary to optimally prepare the data for the GWAS. Two normalisation methods (RLE from DESeq2 and TMM from edgeR) and three transformation methods (vst from DESeq2, log-cpm and voom from edgeR) were selected. For the spatial analysis, one model was chosen as the most intuitive, given the observations previously made in the PCA. All scenarios are listed in Table 3. The location of each plant was expressed as a row number (rowID) and a column number (colID), as it was

the position in a Cartesian coordinate system. These effects were considered random effects. Genotype was the variable containing the hybrid accessions, and repID was the replicate number of the sample. Their effects were considered fixed. The sampling time bin (time bin) was also included in the model as a fixed effect.

Table 3: Set of benchmarked scenarios. *Nine combinations of normalisation, transformation, and spatial models were tested. Genes with 1 cpm in 10% of the samples were retained, and the samples were split following treatment before the filtering step. In the spatial model, the fixed effects are written in italics, and the random effects are in bold.*

Scenario	Normalisation	Transformation	Spatial Model
1	none	none	Trait = genotype + *repID* + **rowID** + **colID** + *TimeBin* + ε
2	TMM	none (cpm)	
3	RLE	none (cpm)	
4	TMM	log-cpm	
5	TMM	voom	
6	RLE	vst	

The benchmark was performed using a dataset containing 45 genes. The expression of 13 of these genes was previously determined using RT-qPCR (five AQPs studied in Laurie Maistriaux's PhD thesis, four flowering genes studied by Castelletti et al. (2020), and four housekeeping genes used in qPCR). We also selected one structural variant (usually found in some lines, but absent in others), 24 genes that correlated with PCs (eight per PC for the first three PCs), and four genes known to be involved in the stress response. Finally, three genes were manually selected based on their expression patterns (Supplemental Table 7). The broad-sense heritability value (hereafter referred to as heritability) of the genes was used to assess the quality of the correction and to compare the scenarios. Heritability is part of the phenotypic variability explained by the genotype. It was expected that the heritability would increase with the quality of the correction, as the proportion of the variance resulting from confounding variables would be reduced.

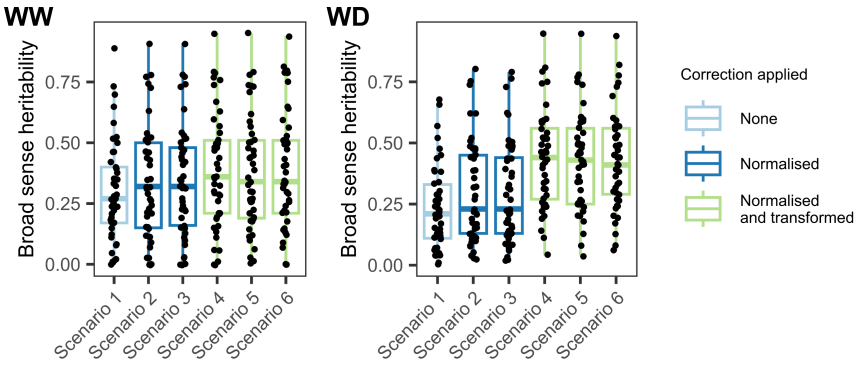


Figure 21: Benchmark plot for WW (left) and WD (right) data. *Boxplot of the heritabilities of 45 genes from six scenarios. Heritabilities increased with normalisation and transformation. Heritabilities were calculated using SpATS with a model including the genotype, plant positions in the*

greenhouse coded with row ID (rowID) and column ID (colID) as random effects and the sampling time bin as fixed effect. The residuals were fitted as random effects.

The heritabilities for the WD and WW data are shown in Figure 21. Normalisation improved heritability, but the improvement was greater when it was followed by transformation. However, no clear conclusion could be drawn regarding these methods, as they all performed equally well. Scenario #4 was chosen for data preprocessing, as it had already been used in a similar analysis (Chateigner et al., 2020) and was closest to that used for the DE analysis. The expression data was then normalised using the TMM method and transformed using the log-cpm method. Finally, the best linear unbiased estimators (BLUEs) of the model were extracted and used as the adjusted mean for each gene from each genotype, and all replicates for each genotype were collapsed into a single value. This yielded adjusted means for all genes expressed in the dataset suitable for use in the GWAS.

The AQP heritabilities were also studied. They differed greatly between isoforms and treatments (Figure 22). Globally, *Zm*PIPs showed homogeneously higher heritabilities under both watering conditions than the other AQPs. The most divergent family was the *Zm*NIPs, in which *ZmNIP3;1* had a heritability of 0.89 in WW and 0.85 in WD, whereas the heritabilities of *ZmNIP1;3* were 0.25 in WW and 0.03 in WD. *Zm*SIPs showed heritabilities below 0.50 but were similar between treatments. Finally, *Zm*TIPs also had divergent heritabilities, with some isoforms with high values and others with very low values.

These results showed that the variation in the expression of some isoforms was mostly due to the genotypes, while for others with lower values, there might be more noise or uncorrected batch effects in the data. Lower heritability values could limit the power of GWAS to detect markers associated with causal polymorphisms.

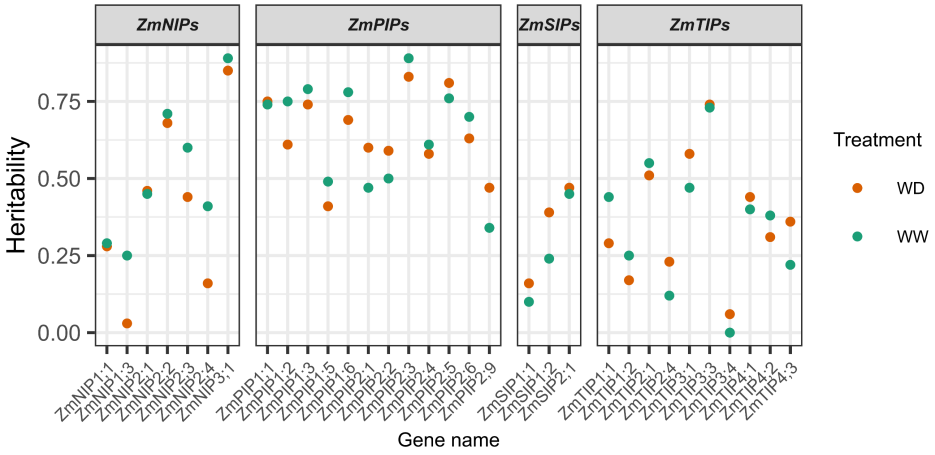


Figure 22: Plot of the heritabilities of the expression of all the* AQP *genes in both watering conditions WD in orange and WW in green. *Each panel represents one AQP subfamily. Heritabilities were calculated using SpATS with a model including the genotype, plant positions in the greenhouse coded with row ID (rowID) and column ID (colID) as random effects and the sampling time bin as fixed effect. The residuals were fitted as random effects.*

The expression of *ZmPIPs* classified by heterotic groups in the mature zone under the two watering conditions is displayed in Figure 23 (Supplemental Figure 3-4 for other families). Twelve of the 14 *ZmPIPs* were expressed in the leaf mature zone. The expression of some of them varied more than others following the heterotic group, but there were no significant differences between the two watering conditions based on DE analysis. These results are similar to what was already observed in the analysis performed by Maistriaux et al. (2024), where interestingly no spatial correction was performed.

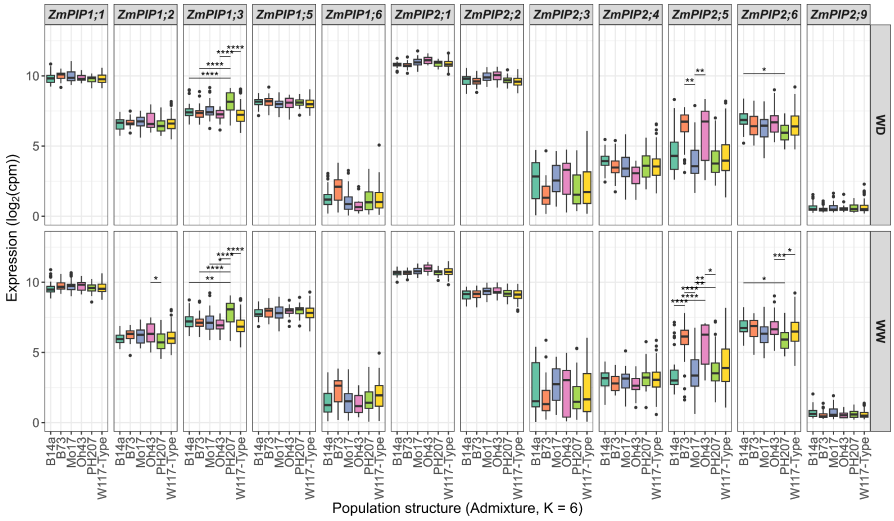


Figure 23:* ZmPIP *genes expression in log(cpm) in WW and WD. *The hybrids were grouped according to their heterotic groups. Differential expression was tested between the groups, and significance is shown at the top of the boxplots (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 10^{-4}).*

Our method to preprocess the 3'RNA-seq data was relatively similar to what is now commonly used in other publications, with some differences in the package used, but based on the same hypothesis (Kremling et al., 2018; Liang et al., 2022). Interestingly, several studies (Fu et al., 2013; C. Li et al., 2023; Pang et al., 2019; G. Sun et al., 2023) still used fragment per kilobase million (FPKM) or read per kilobase million (RPKM) transformations, even though they are not recommended as they do not take into account the RNA composition in their count modelling (Robinson & Oshlack, 2010). We then took batch effects into account in the data. We observed that the sampling time had a strong effect on our data due to the sampling method. To remove these effects before GWAS, we used a spatial effect model (Rodríguez-Álvarez et al., 2018). We considered it relevant to do so, as the time of sampling correlated with the position of the plants in the greenhouse. Furthermore, this type of model was specifically designed to take these effects into account and generate accurate BLUE and BLUP in other QTLs studies (Millet et al., 2016; Prado et al., 2018). However, this was not common with RNA-seq or 3'RNA-seq data. In other studies, batch effects were estimated using PEER factors or PCs and imputed in the GWAS model, as in Kremling et al. (2018) or in Miculan et al. (2021). As our method was almost uncommon, we paid attention in the results for potential false positives related to circadian rhythm or light response. We were also able to perform this spatial correction because more than one replicate for each line was available at different positions, and we discarded lines with only one replicate.

3. eQTL mapping for *AQP* gene expression

The expression data processed in the previous chapter was used in a GWAS to map variants associated with changes in *AQP* gene expression in the panel. This method enabled us to map regions known as eQTLs, where potential regulators could be found. This analysis required the phenotyping dataset containing the expression data, a genotyping dataset containing information about the alleles of the maize lines within the panel, and a kinship matrix containing information about the relatedness between individuals. In our case, 32 *AQPs* were expressed in two watering conditions leading to 64 different situations (defined as Gene Expression x Environment).

The final MLM used in this study is described in the Materials and Methods section; however, several methods were tested before obtaining this final set of eQTLs. Briefly, two methods for calculating the kinship matrix were compared. One was based on the leave-one-chromosome-out (LOCO) principle developed by Listgarten et al. (2012) and computed for our panel in Rincent et al. (2014), and the other used a kinship matrix calculated on the entire set of markers (more details in the Appendix). We decided to use the latter because it gave a lower genomic inflation factor (1.11 for the LOCO method vs 1.03 for the non-LOCO method). This could lead to a loss of power to detect significant markers due to proximal contamination, but we were less likely to have false positives. We also tested various mixed-linear models. We changed the number of covariates by adding the most significant marker as a fixed effect, as performed for *ZmPIP2;5* by Maistriaux et al. (2024). However, due to the complexity of choosing the right number of covariates to add to the model, we kept our model at its simplest. Another nomenclature for eQTL categories was also introduced. We classified eQTLs as “local” when the gene with chich they were associated was located within their genomic region, and as “distal” when the gene was located outside their genomic region. These terms were chosen instead of the conventional *cis*- and *trans*- terms to avoid making assumptions about the underlying biological mechanisms that drive their effects.

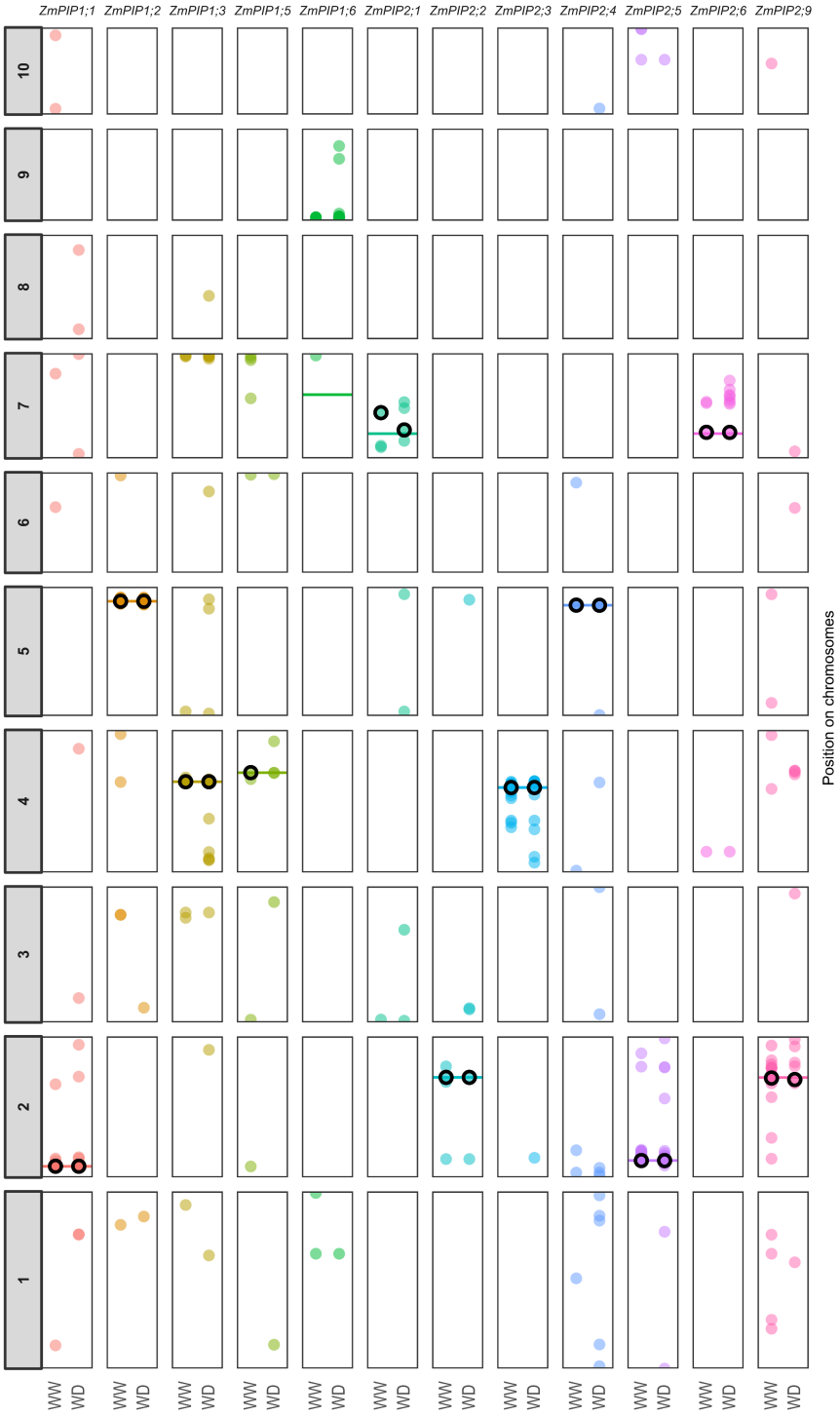
3.1 eQTL mapping

From the GWAS results, 15,736 markers had a $-\log_{10}$(p-value) above 5 and were therefore considered significantly associated with variation in their target gene (the gene for which the markers were tested). The significant markers were then grouped to form eQTLs using the LD_{ajd} and LD_{win} methods described in Materials and Methods. The most significant marker of each eQTL was defined as the lead marker, and its effect was used to reflect the eQTL effect. LD_{adj} grouping used alone yielded 2,961 eQTLs, but when both grouping methods were combined, a total of 850 eQTLs were observed for all phenotypes (Figure 24 for the PIP subfamily; Supplemental Figure 5-7 for the other subfamilies.). Several

inconsistencies can be observed in this figure for *ZmPIP1;5*, *ZmPIP1;6*, and *ZmPIP2;1*. For *ZmPIP1;5*, there was a local eQTL in WW, but not in WD. For *ZmPIP1;6*, there was no local eQTL, and *ZmPIP2;1* local eQTL position was very different in WW and WD. These issues will be addressed in more detail in the next section.

Figure 24: ZmPIP eQTLs. *Each dot represents a physical eQTL position (B73 RefGen v4), and the transparency of the dot indicates the significance of the lead marker. Columns represent chromosomes, while rows represent genes. Data is shown for both WW and WD conditions. ZmPIP positions are indicated with vertical-coloured lines, and local eQTL are circled in black.*

(Figure on the next page)


ZmPIP1;1
ZmPIP1;2
ZmPIP1;3
ZmPIP1;5
ZmPIP1;6
ZmPIP2;1
ZmPIP2;2
ZmPIP2;3
ZmPIP2;4
ZmPIP2;5
ZmPIP2;6
ZmPIP2;9
10
9
8
7
6
5
4
3
2
1
WW
WD
Position on chromosomes

3.2 Markers quality control

To control the quality of the data and any potential inflation of false positives, the Q-Q plots of the p-values were analysed. These represent the observed p-values (y-axis) compared to the expected p-values (x-axis) under the null hypothesis that follow a uniform distribution. A straight line indicates that all the tested markers are not significant, as they would follow the null hypothesis. In a GWAS, it is expected that most markers have no effect on the phenotype and therefore follow the null hypothesis, while a small number of markers are significant. This results in a plot showing a straight line with a deviation on the right side. Interestingly, two patterns were observed in our data.

In most cases, we observed the expected Q-Q plots, indicating that only a few SNPs were associated with the causal polymorphism (Figure 25A). To detect any potential inflation in false positives, we used the genomic inflation factor, and as all values were between 0.94 and 1.11, with a median value of 1.03, we considered that there was no evidence of false positive inflation.

For some genes, the p-values followed an almost perfectly uniform distribution (Figure 25B). Significant markers observed for these genes could therefore be false positives because they result from the distribution rather from a real association with a causal polymorphism. This could be due to the lack of markers in LD with causal polymorphism or a lack of power to detect it. It could also mean that the observed traits were highly polygenic, resulting in many causal polymorphisms with very small effects and less significant p-values. When we selected the candidates to characterise, we took care to select genes presenting expected Q-Q plots.

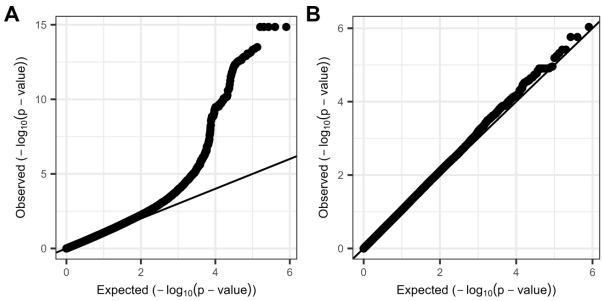


Figure 25: Q-Q plot of the -log$_{10}$(p-value) for all the associations tested for ZmPIP1;1 expression (A) and ZmTIP2;4 expression (B). *The black line indicates the complete correlation between the observed and expected -log$_{10}$(p-value). The expected p-values were generated from a uniform distribution that contained the same number of samples.*

3.3 eQTLs quality control

The number of hybrids under both conditions differed by two, with more data available for WD (254) than for WW (252). That difference had an impact on the

marker minor allele frequency (MAF) and thus on the marker selection in the eQTL analysis, as only markers with a MAF higher than 0.05 were selected. We verified whether significant markers shared between both watering conditions were lost due to this filtering step.

Two significant markers in the entire dataset were discarded for low MAF. They were associated with a variation in *ZmPIP1;5* expression (Zm00001d051872) in both WW and WD conditions but had a MAF lower than 0.05 in WW (Figure 26). However, the eQTL itself was not lost as it was captured by other markers, and therefore, this eQTL was still present in both conditions. In WW, it was considered as a local eQTL because there were other significant markers in this region, and this resulted in a larger eQTL containing the *AQP* gene. In WD, only two markers of this eQTL were significant, and therefore the eQTL was not large enough to contain the gene. This explained why there was a local eQTL in WW, but not in WD. It is important to keep in mind that when, working on the full GWAS dataset, more markers could be lost simply because of the MAF difference between the two conditions. Their p-values could have been lower if the phenotypic data was available for all hybrids.

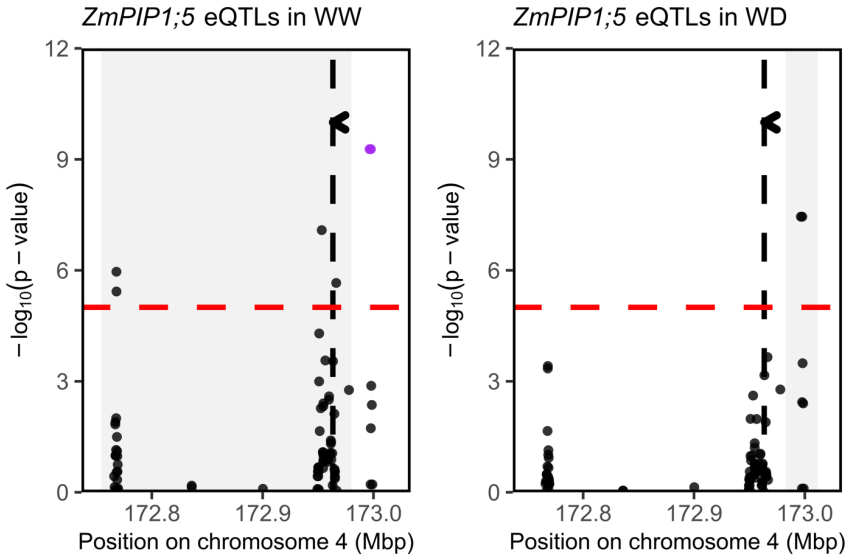


Figure 26: Close-up Manhattan plot on chromosome 4 for the ZmPIP1;5 gene in WW (left) and WD (right). *The x-axis represents the physical positions of the markers along chromosome 4 (B73 RefGen v4). The y-axis represents the -log₁₀(p-value) obtained from the GWAS performed using FASTLMM. The red horizontal dashed line indicates the significance threshold of 5. The black vertical dashed line and arrow show the position and direction of the ZmPIP1;5 gene, respectively. Markers in purple are those with a low MAF and were therefore not kept in the analysis. The gray background indicates the eQTL borders.*

We also observed some very large eQTLs of several millions of bp. For example, *ZmPIP2;1* local eQTL on chromosome 7 was approximately 125M bp, which is

much larger than the average gene length of ~3,000 bp. All markers belonging to this eQTL were in high LD with the lead marker (Figure 27). The position of the local eQTL between WW and WD differed on Figure 24 because the lead marker of the eQTL was different in both conditions. Such a pattern could arise from assembly error or chromosomal rearrangement (Negro et al., 2019).

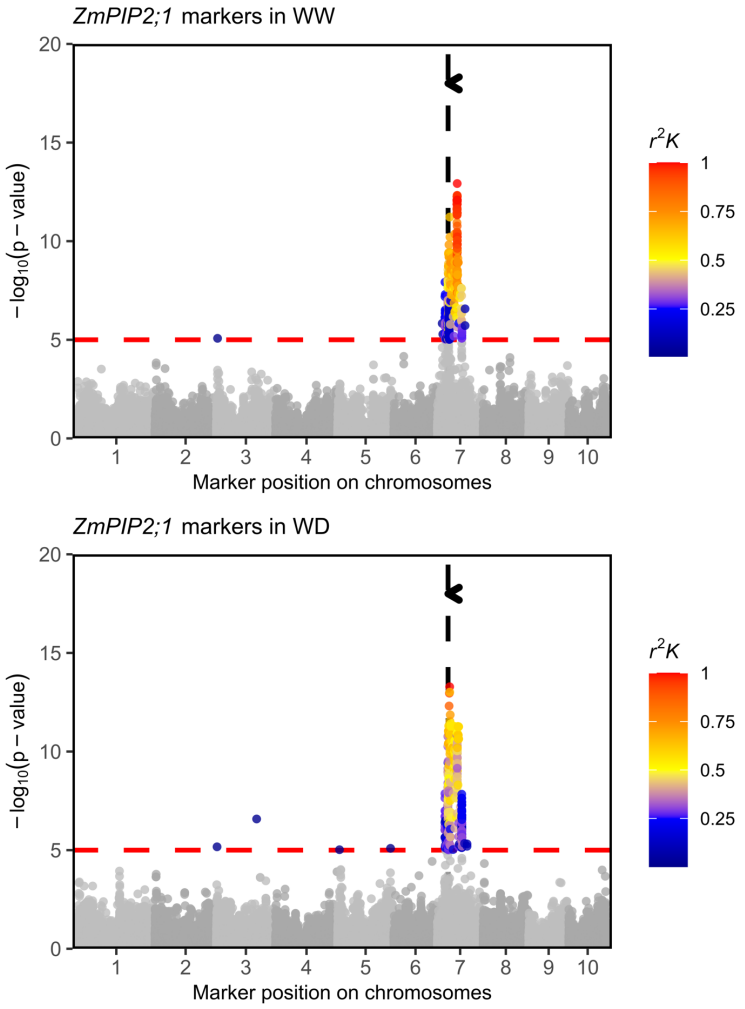


Figure 27: Manhattan plots of ZmPIP2;1 under both watering conditions, WW (top) and WD (bottom). *Each dot represents a marker. The x-axis represents their physical positions along the chromosomes (B73 RefGen v4). The y-axis represents the -log₁₀(p-value). The red horizontal dashed line indicates the significance threshold of 5. Significant markers are coloured following their LD (measured as r²K) with the most significant marker. The black vertical dashed line and arrow show the position and direction of the gene, respectively.*

It is worth noting that another error was observed on the same chromosome for *ZmPIP1;6* (Figure 28). Actually, *ZmPIP1;6* is located on chromosome 9 in other

maize genome versions and, here, we observed that the most significant markers were located on the ninth chromosome instead of the seventh were *ZmPIP1;6* was mapped in v4, confirming its true position.

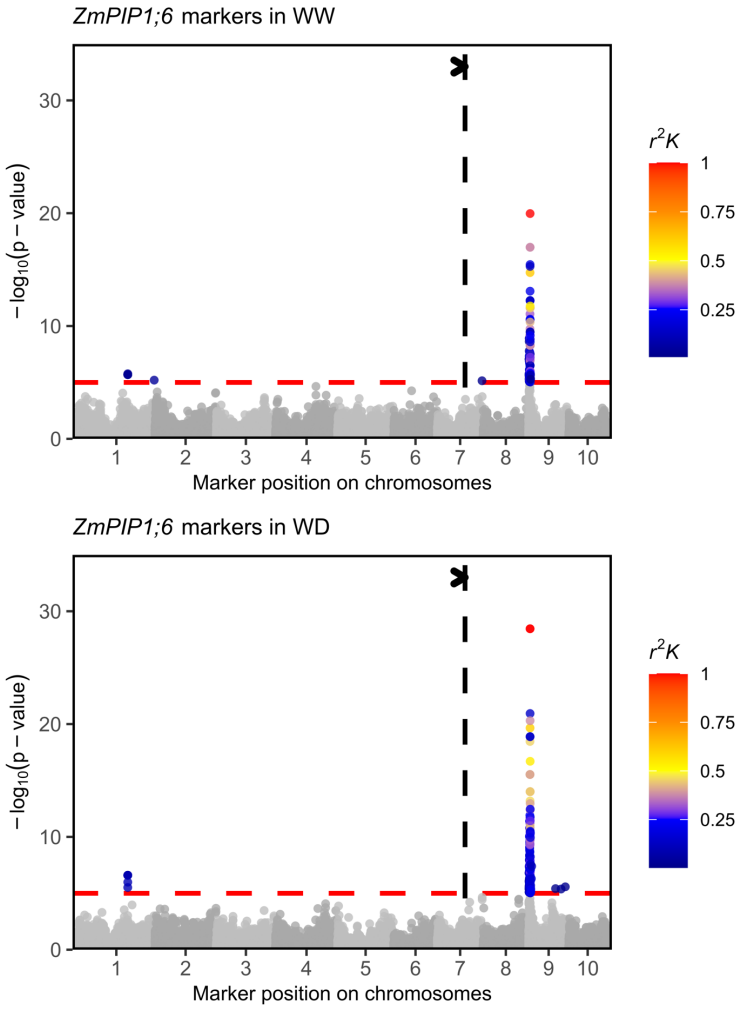


Figure 28: Manhattan plots of ZmPIP1;6 under both watering conditions, WW (top) and WD (bottom). *Each dot represents a marker. The x-axis represents their physical positions along the chromosomes (B73 RefGen v4). The y-axis represents the -log₁₀(p-value). The red horizontal dashed line indicates the significance threshold of 5. Significant markers are coloured following their LD (measured as r²K) with the most significant marker. The black vertical dashed line and arrow show the position and direction of the gene, respectively.*

We also used the median adjacent LD to verify that the markers were grouped consistently together in eQTLs instead of randomly. A high LD would indicate that markers are highly correlated and should be in the same eQTL; on the contrary, a low LD would indicate that the markers are less likely to belong to the

same eQTL. We compared the median LD of the eQTLs with their sizes and observed that they were not associated (Figure 29A, grouping based only on LD_{adj}). However, after grouping markers using both the LD_{adj} and LD_{win} methods, we observed that some eQTLs had a lower median LD (Figure 29B). The LD_{win} method is based on the estimated windows of LD around each marker (Negro et al., 2019). The threshold chosen to merge two eQTLs together, based on the overlap between two windows, was low, which could explain the decrease in LD. However, we decided to keep this grouping procedure after comparing it with methods solely based on physical distances (G. Sun et al., 2023) as it was based on LD around markers.

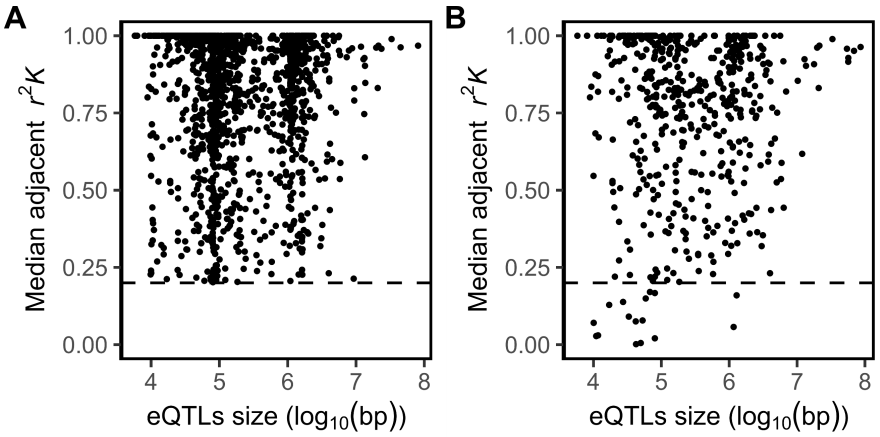


Figure 29: Scatter plot between the eQTLs size (-$\log_{10}$(bp)) and the median adjacent LD (measured as r^2K) before grouping with LD_{win} (A) and after grouping with LD_{win} (B). *The black horizontal dashed line represents the LD threshold used to group markers using the LD_{adj} method.*

3.4 eQTLs analysis

The eQTLs were annotated using the v4 annotations. Two types of eQTLs were defined in this step: local eQTLs that contained the gene they were associated with and distal eQTLs that did not contain this gene. We observed that the lead markers of local eQTLs had larger effects than the lead markers of distal eQTLs (Figure 30A) and explained more phenotypic variance (Figure 30B), even though this observation should be tested on all the expressed genes. Lead markers from local eQTLs were usually highly significant (-$\log_{10}$(p-value) > 20).

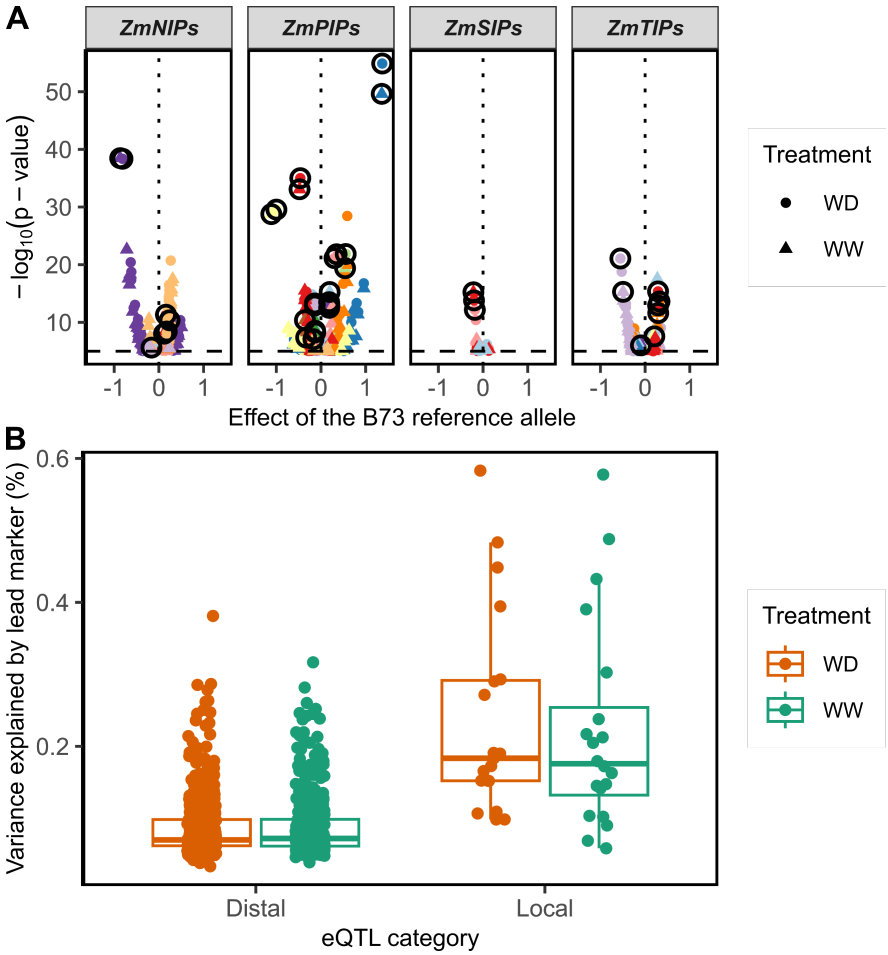


Figure 30: Marker analysis. *(A) p-value of the eQTL lead markers compared to the effect of the B73 reference allele on the phenotype for all AQP families, each isoform is represented by a different colour. Dots circled in black indicate lead markers belonging to local eQTLs, and the shape of the dots indicates the treatment. (B) Boxplot of the phenotypic variance explained by the lead marker (R^2_{LR}) of each eQTL in both watering conditions (WD in orange and WW in green). The eQTLs were sorted by category (local or distant).*

We observed that out of the 32 *AQP* genes expressed in our dataset, 22 had a local eQTL in both watering conditions, while 5 *AQP*s presented a local eQTL in only one condition (including *ZmPIP1;5* for which the WD local eQTL was lost due to the MAF threshold). When local eQTLs were present in both conditions, they overlapped on average by 72%, and their effect was in the same direction (Figure 30A), suggesting that they likely captured the same causal variant independently of the watering condition. This also shows that markers located close to the gene with which they are associated have a greater effect on its expression and could therefore be selected as a priority.

4. Selection of candidates

Since eQTLs were available for 32 *AQP* genes expressed in the mature zone of maize leaves, we performed a fine mapping to identify their potential regulators. As described in the Introduction, there are two types of regulators: *cis*- and *trans*-regulators. In this thesis, we decided to focus on *trans*-regulators and more specifically on TFs. This chapter describes the process by which TFs were selected for wet-lab validation. As there were many genes located in the eQTLs, we only selected genes that were expressed in the 3'RNA-seq as these genes were expressed at the same time than their potential target *AQP*. However, it is important to note that genes not passing the expression filter, due to being expressed in too few lines, may have been overlooked. Additionally, genes located in a distal eQTL that biologically capture a CRE could be mistakenly classified as *trans*-regulatory factors. Ultimately, this approach drastically decreased the total of potential regulators. Then, we compared the resulting dataset with data gathered from the literature.

4.1 Enrichment analysis

To refine the list of potential TFs that regulate *AQP* expression, we first used data from another study and compared the results with our eQTLs. TF-target gene associations from 45 GCNs were downloaded from the study of Zhou et al. (2020). These GCNs were built using expression data coming from many different genotypes, organs and environmental conditions. A total of 1,000,000 TF-target gene associations were published per network involving 2,048 distinct TFs associated with 35,781 distinct target genes, and 41 *AQP* genes were observed in the targets of this dataset. We focused on the 100,000 strongest associations in the networks, reducing the number of target genes to 35,627. In this study, 56 TF families were represented. An enrichment analysis based on the 45 networks altogether yielded 105 TFs enriched for *AQP*s (Figure 31). The decision to perform this analysis without taking into account the different sampling conditions was explained by the fact that we wanted to capture a broad picture of *AQP* expression regulation. This could have an effect of lowering the power to detect enriched TFs with the Fisher's exact test as we increased the total number of targets per TF. To verify that we were not losing TFs due to the power decrease, we performed an enrichment analysis per network, but the yield of enriched TF was even lower. Therefore, we decided to continue with the aggregated data. The median number of *AQP*s in the enriched TFs targets was 10. These TFs belonged to 30 families, including some that are known to be related to drought resistance like NAC, ERF, or bZIP TFs, and others that have not been characterised yet. Of these 105 TFs, 17 were also present in the *AQP* eQTLs (Figure 31) and represented 12 TF families.

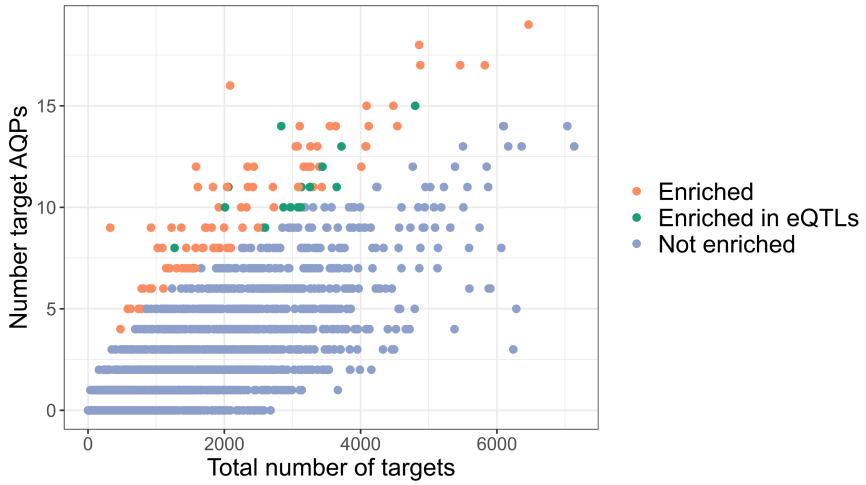


Figure 31: Enrichment plot for TF-target associations. *Each dot represents a TF, the x-axis indicates its total number of targets and the y-axis the number of AQPs in its targets. Three categories of TFs can be overserved; TFs significantly enriched for AQP genes (in orange), enriched for AQPs and in AQPs eQTLs (in green), and not enriched (in blue).*

4.2 Network on the GWAS

At this step, we represented the suspected interactions as a network (Figure 32). We decided to select AQPs with a good Q-Q plot profile showing a low inflation rate and significant markers. We observed that most TF-AQP interactions were specific to one watering condition, but five of them were detected in both. Some *AQP*s were targeted by several TFs, such as *ZmTIP3;1*, *ZmNIP2;2,* and *ZmPIP1;3*. Some TFs also targeted multiple *AQP*s, like *ZmGLK24* and *ZmNAC40*. Thank to this information, we reduced the total number of interactions to 17.

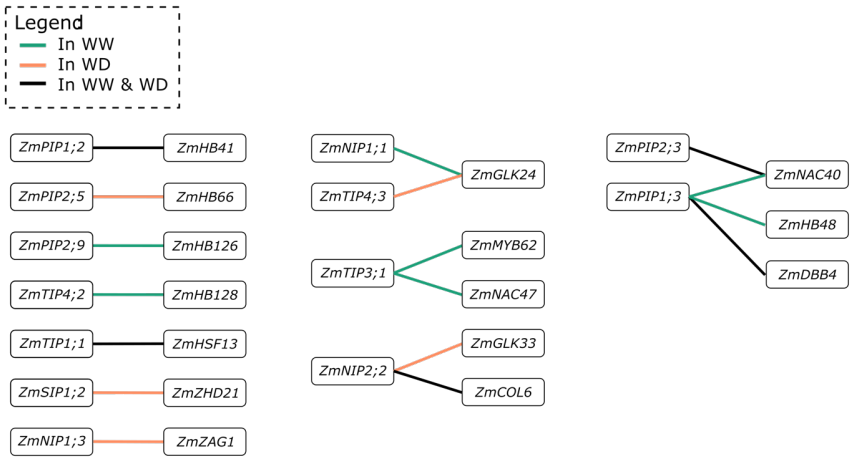


Figure 32: Network of TF-AQP associations. *The names of AQP and TFs are given. The colour of the edges indicates in which watering condition the interaction was detected: black when detected in both, green in WW and orange in WD.*

4.3 *AQP* gene promoter analysis

To gather more information and draw hypotheses about the mechanism of action of the TFs on the *AQP* promoters, we predicted TF binding sites in 2-3k bp regions upstream of the *AQP* genes, defined as promoters. This length range was chosen because, as we saw in the introduction, several CREs responsible for gene expression can be located within 3k bp upstream of the core promoter. This distance varied because for some *AQP* genes, the whole 3k bp promoter region could not be amplified experimentally, and a smaller region had to be selected. This explains why promoter size varies between *AQPs*. The families of the TF binding sites were then analysed to determine if direct binding was possible. The absence of a binding site in a promoter does not exclude an interaction, as it could be an indirect interaction or it would require the recruitment of other TFs to bind (Long et al., 2016).

The tool we used to predict the binding motifs in the promoter yielded results for TFs for which the motif had already been characterised. We therefore had information about which motifs could be found in the promoters, which TF could bind to it, and the TF family. This allowed us to check whether an enriched TF located in the eQTLs had a binding motif in the *AQP* promoter. This was the case for *ZmHB128*, which potentially interacts with *ZmTIP4;2* promoter. Interestingly, this motif was found in other *AQP* promoters. We also analysed the TF families because members of the same family can recognise similar motifs (O'Malley et al., 2016). We therefore listed enriched TFs for which family members could bind to the *AQP* promoter. This led to the results presented in Table 4. After obtaining these results, we decided to focus on *AQPs* for which an enriched TF had been

detected and that family motif was found in the promoter. The direct interaction could be validated using EMSA, but here we validated the effect of the TF overexpression on promoter activity.

Table 4: TF-AQP interaction including the TF family and whether the binding motif of the TF family was located in the AQP promoter region.

TF accession	Name	TF family	AQP name	Binding motif in promoter
Zm00001d002799	*Zm*HB66	HD-ZIP	*ZmPIP2;5*	YES
Zm00001d006195	*Zm*HB126	HD-ZIP	*ZmPIP2;9*	YES
Zm00001d008817	*Zm*NAC47	NAC	*ZmTIP3;1*	NO
Zm00001d015521	*Zm*GLK24	Golden2-like	*ZmTIP4;3*	NO
Zm00001d015521	*Zm*GLK24	Golden2-like	*ZmNIP1;1*	YES
Zm00001d017422	*Zm*HB41	HD-ZIP	*ZmPIP1;2*	YES
Zm00001d021268	*Zm*HB128	HD-ZIP	*ZmTIP4;2*	YES
Zm00001d027757	*Zm*HSF13	HSF	*ZmTIP1;1*	YES
Zm00001d035604	*Zm*MYB62	MYB	*ZmTIP3;1*	YES
Zm00001d037015	*Zm*GLK33	Golden2-like	*ZmNIP2;2*	YES
Zm00001d037737	*Zm*ZAG1	MIKC-type MADS	*ZmNIP1;3*	YES
Zm00001d041780	*Zm*ZHD21	ZF-HD	*ZmSIP1;2*	NO
Zm00001d042864	*Zm*HB48	TALE	*ZmPIP1;3*	YES
Zm00001d045804	*Zm*COL6	CO-like	*ZmNIP2;2*	YES
Zm00001d050893	*Zm*NAC40	NAC	*ZmPIP2;3*	YES
Zm00001d050893	*Zm*NAC40	NAC	*ZmPIP1;3*	YES
Zm00001d051018	*Zm*DBB4	DBB	*ZmPIP1;3*	NO

4.4 Analysis of *AQP* GRNs

The 15 TFs identified as potential regulators of *AQP* gene expression belong to 10 TF families out of the 56 TF families found in maize (Jin et al., 2015, 2017). Homeodomain-leucine zipper (HD-ZIP) was the most represented family with four members, followed by NAC (NAM, ATAF1, 2 and CUC2), and Golden-like (G2-like) TFs with two members. All other families had one member associated with *AQP* expression. We analysed the expression of these TFs in the DROPS panel between the different heterotic groups, but no clear expression pattern could be detected (Supplemental Figure 8), suggesting low variability between the lines.

Homeobox superfamily

HD-ZIP belongs to the homeobox protein superfamily. The zinc finger homeodomain (ZF-HD) and three amino acid loop extension (TALE) TF families that we observed in the network also belong to this superfamily. They are characterised by the presence of a conserved 60-amino-acid DNA-binding motif, known as the homeodomain (Ariel et al., 2007). Families belonging to this superfamily differ in the sequence encoding the homeodomain, its size and location, and in the presence of other conserved domains (Ariel et al., 2007).

HD-ZIP proteins have a plant-specific leucine zipper motif immediately downstream of the homeodomain (Schena & Davis, 1992). The homeodomain binds DNA, while the leucine zipper motif acts as a dimerisation motif. HD-ZIP TFs are divided into four subfamilies (I-IV) based on their physiological function, protein structure, and conserved domains. HD-ZIP members are involved in plant development and responses to environmental conditions (Ariel et al., 2007). Qiu et al. (2022) identified 42 HD-ZIP drought-responsive members in maize, using the expression data of Cao et al. (2019). These include *ZmHB41*, *ZmHB66*, *ZmHB126*, and *ZmHB128*, all of which belong to subfamily I and are associated with the expression of AQP genes, which contain homeodomain-binding motifs in their promoter sequences. Interestingly, members of this family are mainly involved in responses to abiotic stress. For example, in Arabidopsis, RNA blot analysis showed that the expression of *AtHB7* and *AtHB12* was induced by ABA and drought (Olsson et al., 2004), and deregulation of their expression in transgenic lines disrupted water loss and uptake compared to the WT through changes in stomatal conductance and aperture (Ré et al., 2014). The associated AQPs are *ZmPIP2;5*, *ZmPIP2;9*, *ZmPIP1;2*, and *ZmTIP4;2*, all of which containing binding sites for HD-ZIP members in their promoter. As highlighted in the introduction, *ZmPIP2;5* influences stomatal conductance and ABA responsiveness (Ding et al., 2022). This reinforces the hypothesis that HD-ZIPs regulate *AQP* expression. Therefore, it can be assumed that direct binding is possible.

ZF-HD proteins contain a zinc finger domain composed of cysteine and histidine residues, coordinating a zinc ion in their N-terminal region and a homeodomain in their C-terminal region. The homeodomain is responsible for DNA binding, whereas the zinc finger domain enhances protein-DNA binding and enables protein-protein interactions (Ciftci-Yilmaz & Mittler, 2008). In maize, the ZF-HD family contains 21 members divided into eight categories (I-VIII) (Islam et al., 2022). In our data, *ZmZHD21* was associated with variation in expression of *ZmSIP1;2*, whose promoter does not contain a ZF-HD binding motif. The formation of heterodimers between ZF-HD TFs has already been observed in *Flaveria trinervia* and rice by Yeast two-hybrid assay. We could hypothesise indirect regulation by interaction with ZF-HD with different DNA-binding motifs,

although this is unlikely as the DNA-binding specificity in ZF-HD does not seem high, as observed in Y1H assays (Figueiredo et al., 2012; Windhövel et al., 2001). Islam et al. (2022) reported that *ZmHD21* has a greater expression in floral tissue where AQPs from other subfamilies are also active (Azad et al., 2013; W. Chen et al., 2013; N. Ma et al., 2008). In addition, ZF-HDs are involved in the regulation of floral development in Arabidopsis (Tan & Irish, 2006). ZF-HD may also have a role in abiotic stress responses. Sun et al. (2021) showed that silencing of *Nt*ZHD21 in tobacco reduced drought resistance.

TALE gene family is characterised by a modified homeodomain. Indeed, their homeodomain contains a three-amino acid extension that creates a loop structure (Hamant & Pautot, 2010). This structure enhances DNA-binding specificity and functional capabilities. Initially, TALE TFs were shown to play an important role in growth and developmental processes (Hamant & Pautot, 2010). However, recent studies have highlighted their potential involvement in responses to abiotic stress by analysis *cis*-elements in their promoters (Y. He et al., 2022; Qian et al., 2024; Q. Yang et al., 2022). Qian et al. (2024) reported 40 members of this family in maize, among which *ZmHB48*, also named *ZmTALE19*, belongs to the BELL1-like subfamily

MYB family

MYB TFs are another family found in the AQP regulatory network. Members of this family have already been identified as regulators of AQP expression in poplar, citrus, and bamboo, where they influence drought tolerance (Jiang et al., 2020; P.-Y. Zhang et al., 2021; C. Zhu et al., 2024). Interestingly, Zhang et al. (2021) reanalysed Cao et al. (2019) transcriptomic data and identified 54 drought-responsive MYB genes, including *ZmMYB62*, which we found in our network. MYB TFs constitute a large family characterised by one, two, or three imperfect 51-53 amino acids sequence repeats in their N-terminal part responsible for DNA-binding specificity (Dubos et al., 2010). MYB are key players in abiotic stress signalling (G. Li et al., 2014). In Arabidopsis, *AtMYB60* regulates stomatal movement, and the overexpression of one of its splice variants decreased the transcript levels of several *AQP*s in microarray analysis under drought stress (Oh et al., 2011). *ZmMYB62* was associated with *ZmTIP3;1* expression, which contains a MYB binding site, making it a potential direct target. In our data, *ZmTIP3;1* was lowly expressed in both conditions, as also reported by Su et al. (2022). In fact, it is principally expressed in seeds and involved in seed vigour. As lowly expressed genes present a higher expression noise, this association could be a false positive.

GARP (Golden2-like, ARR-B, Psr1) family

Golden2-like (GLK) TFs belong to the GARP TF family, which is related to the MYB superfamily (Riechmann et al., 2000). GLK TFs are grouped into seven groups (I-

VII). They contain two conserved domains at their C-terminal end, an MYB-like DNA-binding domain, and a GCT-box specific to land plants (Rossini et al., 2001). They are involved in cellular differentiation of photosynthetic tissues (M. Chen et al., 2016). *ZmGLK33* belongs to the VII group and is associated with the expression of *ZmNIP2;2*, which contains a GLK-binding motif. Because some AQPs facilitate the diffusion of CO_2 into the chloroplasts (Uehlein et al., 2008), we suggest that they are regulated by members of the GLK family. However, there is no evidence of CO_2 transport by NIPs (Fox et al., 2017).

MADS family

ZmZAG1 belongs to the MIKCS-type MADS, which is one of the five subfamilies of the MADS-box TF family. They are plant-specific and are involved in flower formation (Nam, 2003; Pařenicová et al., 2003). The presence of a MIKCS-type MADS-binding domain in the *ZmNIP1;3* promoter and the fact that some AQPs are involved in petal development suggest that it could be directly regulated by MIKCS-type MADS TFs (Azad et al., 2013).

B-box family

B-box TFs contain one to two b-box conserved domains (including a zinc finger domain) at their N-terminus and a region of 41-43 amino acids at the C-terminus, known as the CCT domain (CO, CO-like, TOC1) (Griffiths et al., 2003). The b-box regulates protein-protein interactions, while the CCT domain is responsible for the correct localisation of the protein, DNA-binding, and its function (Gendron et al., 2012). Currently, there is no consensus on the division of members of this family into smaller families and their naming, but differences occur in the distance between the two b-box domains, their conservation, and the presence/absence of the CCT domain (Khanna et al., 2009). B-box proteins are generally associated with crucial roles in various aspects of plant growth, development, responses to both biotic and abiotic stresses, as well as phytohormone and light signalling (Datta et al., 2007, 2008; Gangappa & Botto, 2014).

ZmCOL6, associated with *ZmNIP2;2* expression, is classified as a class A *CONSTANS-like* (*CO-like*) gene belonging to B-box protein by Song et al. (2018). In maize, they identified 19 *CO-like* genes divided into three classes (A, B, and C) involved in several processes related to flower development and circadian clock regulation (Song et al., 2018). Interestingly, finding such a gene in our network could be due to the remaining effect of the time bins described in part 2, as some lines could have been sampled later the same day, resulting in expression differences due to the circadian and diurnal rhythms. *AQP* expression also follows these rhythms (Caldeira et al., 2014; Hachez et al., 2008; Heinen et al., 2009); therefore, we may have identified a potential regulator of this pattern.

Additionally, *ZmDBB4* was observed in *ZmPIP1;3* eQTLs. ZmDBB4 is a double b-box protein that does not contain a CCT domain (W. Li et al., 2017). Phylogenetic analysis revealed that *ZmDBB4* and *ZmDBB6* are closely related. Their genetic structure and motif composition show similarities, indicating that they might share functional roles. In fact, their diurnal expression patterns and sensitivity to light are also similar (W. Li et al., 2017). However, ZmDBB6 was not detected in our network. The *ZmPIP1;3* promoter did not have a DBB binding motif, but we can hypothesise an indirect binding as b-box proteins can form homo/hetero-dimers (Datta et al., 2007, 2008; Gangappa & Botto, 2014; Qi et al., 2012).

Heat shock factors (HSF) family

The members of the HSF family are characterised by a highly conserved DNA-binding domain, which can accurately identify and bind the heat shock *cis*-elements (HSEs; nGAAnnTTCn or nTTCnnGAAn) in their N-terminal part (Nelson & Littlefield, 1999). A linker region, composed of 15–80 amino acids, connects the DNA-binding domain to a conserved region responsible for oligomerisation, known as the HR-A/B region. Based on the diversity of this specific region, HSF TFs can be divided into three groups: A, B, and C (Nover et al., 2001; Scharf et al., 2012). Additionally, all HSFs contain a nuclear localisation signal in the C-terminal part of the oligomerisation domain. Class A HSFs possess two specific motifs on their C-terminal end: a nuclear export signal and two adjacent activator modules, which consist of aromatic and large hydrophobic amino acid residues embedded in an acidic surrounding (AHA). The AHA domains are responsible for activating transcription (Kotak et al., 2004). Interestingly, class B HSF possess a repression motif in their C-terminal part (Czarnecka-verner et al., 2004). Although HSFs are primarily known to mediate the heat stress response, there is increasing evidence for their involvement in other molecular responses such as salt, drought, osmotic stress, and infection (Bechtold et al., 2013; Ogawa et al., 2007; Yokotani et al., 2008). Recently, Wang et al. (2020) showed that in apple, the drought-responsive *MdHSFA8a* promotes the accumulation of flavonoids, the scavenging of reactive oxygen species, and plant survival under drought conditions. They also demonstrated that *MdHSFA8a* participates in ABA-induced stomatal closure and promotes the expression of ABA signalling-related genes, highlighting its broad role in stress adaptation. In maize, 31 HSF TFs have been identified (Myers et al., 2023), among which *ZmHSF13* is associated with the variation of *ZmTIP1;1* expression in our network. An HSE was found in *ZmTIP1;1* promoter suggesting a direct binding. *ZmHSF13* is a class A HSF and therefore contains an AHA motif allowing transactivation. It is closely related to *At*HSF6b in Arabidopsis, which is involved in ABA response and thermotolerance (Huang et al., 2016; Z. Li & Howell, 2021). The overexpression of *ZmHSF13* in Arabidopsis plants conferred thermotolerance and a higher chlorophyll content under heat stress (H. Zhang et al., 2020). Moreover, as

highlighted in the introduction, in pear, there is evidence for the regulation of *PbPIP1;4* expression by an interaction between *Pb*ERF3 and *Pb*HSF1Ca.

NAC family

The last TFs are *Zm*NAC40 and *Zm*NAC47, both of which belong to the NAC superfamily, which is unique to plants (Olsen et al., 2005). The NAC name is derived from TFs families sharing a conserved domain: No Apical Meristem (NAM) from petunia, Arabidopsis Transcription Activation Factor (ATAF), and Cup-shaped Cotyledon (CUC) from Arabidopsis (Aida et al., 1997; Souer et al., 1996). This highly conserved domain, now called the NAC domain, is located at the N-terminus and is involved in DNA binding (Ernst et al., 2004). The C-terminus is more divergent among members and is responsible for activating transcription (Ooka et al., 2003). Fan et al. (2014) identified 124 NAC TFs in maize and divided the family into 13 distinct subgroups. NAC TFs play crucial roles in growth and developmental processes, including floral development (Sablowski & Meyerowitz, 1998), leaf senescence (Guo & Gan, 2006), and responses to abiotic and biotic stresses, such as salt and drought or pathogen infection (P.-L. Lu et al., 2006; Taji et al., 2004; X. Wang et al., 2009). In our network, *ZmNAC47* was identified in *ZmTIP3;1* eQTLs, although no NAC binding motif was observed. In contrast, *ZmNAC40* was linked to *ZmPIP1;3* and *ZmPIP2;3* eQTLs and both *PIP* genes contain NAC binding motifs in their promoter. The link between *AQP* expression and NAC TFs has already been established in two previous studies. For example, in *Rosa hybrida*, silencing *RhNAC100* led to increased expression of *RhPIP1;1* and *RhPIP1;2* in petals. Furthermore, *Rh*NAC100 binds directly to their promoters both *in vitro* (via EMSA) and *in vivo* (via ChIP assays in Arabidopsis protoplasts) (Pei et al., 2013). Similarly, in *Tamarix hispida* (Willd.), a direct binding of *Th*NAC12 to the *ThPIP2;5* promoter was validated by ChIP, Y1H, and dual-luciferase reporter gene assays. This interaction enhanced salt stress tolerance by regulating *ThPIP2;5* expression, which itself improves salt stress tolerance by boosting ROS-scavenging capacity and antioxidant enzyme activity in transgenic plants (L. Wang et al., 2018; R. Wang et al., 2021). These examples demonstrate that *AQP* genes are potential targets for NAC TFs.

We avoided drawing conclusions regarding the condition specificity of the identified associations. This decision is based on the fact that nearly all TF families analysed contain members involved in drought or osmotic stress responses. As a result, attributing a specific regulatory interaction to a particular treatment (e.g., WW or WD conditions) is challenging. Without testing these interactions in controlled environments, it is difficult to determine whether the associations observed are truly specific to a given condition or reflect a broader role in stress response mechanisms.

We decided to characterise three associations detected in the network. The first one is the interaction between *Zm*HSF13 and *ZmTIP1;1*, both of which were differentially expressed between WW and WD. This provides additional further evidence of an expression correlation in addition to what was observed in the GWAS. The second and third associations are between *Zm*NAC40 and *ZmPIP1;3* and *ZmPIP2;3*, based on the established role of NAC TFs in regulating *AQPs* and ongoing research in the lab on *Zm*NAC24-mediated regulation of *ZmPIP2;5*. By investigating these associations, we aimed to better understand of how specific TFs modulate *AQP* gene expression under varying environmental conditions.

5. Effect of candidate TFs on *AQP* gene expression

In this section, we will now focus on some associations previously described, from the eQTL data to the wet lab validation. We performed transactivation assays to validate the most promising potential regulators of *AQP* gene expression.

5.1 *ZmTIP1;1* eQTLs

5.1.1 Local eQTLs

We first focused on *ZmTIP1;1* local eQTLs. We observed that there was a local eQTL in WD, but it was not detected in WW, probably because the boundaries of the closest eQTL stopped upstream of the gene (Figure 33A and C). This can be explained by the grouping method used. More markers were significant in this region in WW and, when calculating the LD between adjacent markers, there were more chances for the eQTL to be "broken" due to one marker pair having an LD value below the threshold, resulting in more dispersed eQTLs. This is easily comparable with the WD situation where fewer markers were significant, and their adjacent LD was higher than the threshold required to form an eQTL. This explained why the local eQTL was lost in WW. However, it was worth studying the markers within these eQTLs. The two markers constituting the local eQTL in WD were also significant in WW. These markers were CNVMAIZE_INS_102207, which was the lead marker in WW ($-\log_{10}$(p-value) of 17.56), and CNVMAIZE_INS_89412, which was the lead marker in WD ($-\log_{10}$(p-value) of 7.57). When comparing *ZmTIP1;1* expression in lines carrying the reference allele (B73 allele) and the alternative allele, we observed that the alternative allele of both markers decreased *ZmTIP1;1* expression (Supplemental Figure 9). These markers are a specific type of InDels called presence/absence variants (PAVs), they were identified for our panel by Mabire et al. (2019). These are InDels of at least 100 bp that are either completely absent in the reference genome (B73) but present in at least one of the other lines (F2, PH207, or C103), or present in the reference genome but missing in another line, categorised as PAV insertions and PAV deletions, respectively. Their sequences were first BLASTed against the NCBI database. Surprisingly, the results for

CNVMAIZE_INS_102207 showed the presence of a 425 bp fragment from the 3'end of *ZmTIP1;1* inside its sequence. This was unexpected as the PAV was presumed absent in B73, with the estimated insertion locus corresponding to a position approximately 100k bp away from this gene's location in the B73 genome (Supplemental Figure 10). We then BLASTed the sequence of CNVMAIZE_INS_102207 against the B73 RefGen_v4 on maizeGDB and observed that it aligned well in B73, although at a different position than the one obtained through the conversion from v2 to v4. We assumed that these discrepancies were likely due to two key factors: the reference used to identify the PAVs and the position conversion method. The identification of the PAVs was performed using B73 RefGen_v2, where *ZmTIP1;1* was not annotated, rather than B73 RefGen_v4, where *ZmTIP1;1* was introduced. Therefore, finding *ZmTIP1;1* in CNVMAIZE_INS_102207 sequence may be due to the improvement in sequencing technologies that allowed the detection of more gDNA regions. In addition, this could highlight a potential mistake in the CrossMap conversion from v2 to v4, as the PAV position was far from the gene. This conversion method could indeed introduce errors (Pan et al., 2019) and explain the distance between the location of the marker in the B73 genome and *ZmTIP1;1*. There are several methods for performing conversions, liftovers and re-alignments. The method we used belongs to liftover methods, and many tools exist to perform this (Luu et al., 2020). They are based on algorithms that use a chain file to lift the coordinate from one reference to another. Liftovers are faster and cost-effective, but they can introduce mistakes. The ideal solution would be to re-align the original sequence data of the markers on the new reference. However, this is time-consuming (Luu et al., 2020), and we were unable to perform this in the framework of this thesis. Therefore, the liftover approach was used due to technical and time constraints. This does not mean that the positions of the PAVs are wrong, they are simply less accurate. As only PAVs positions were converted by liftover, most of the eQTLs were correct thanks to the SNPs positions. In conclusion, this indicated that there was indeed a local eQTL in both WW and WD. Regarding the CNVMAIZE_INS_89412 sequence, no gene was detected by BLAST analysis.

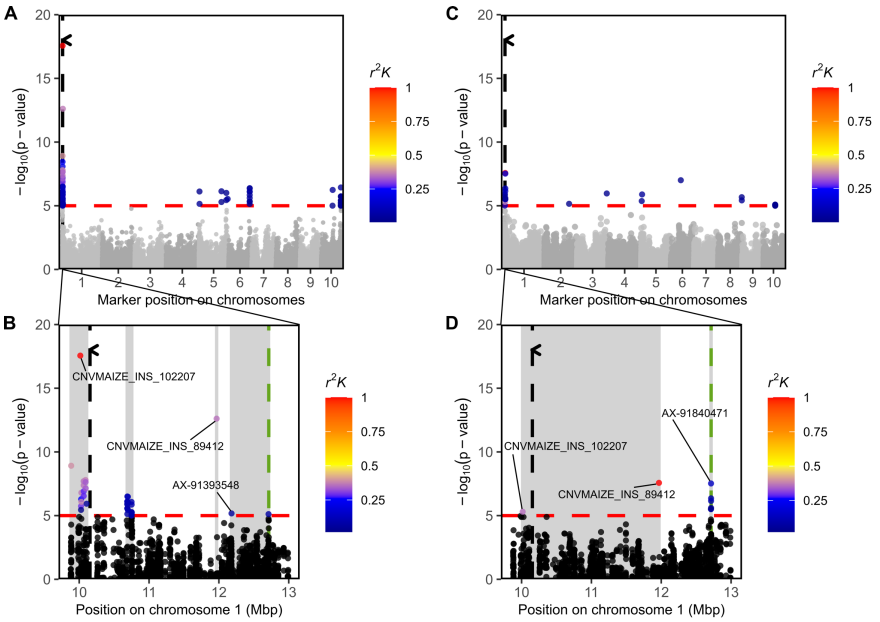


Figure 33: Full and close-up Manhattan plots for ZmTIP1;1 gene expression in WW (A and B) and WD (C and D). *The x-axis represents the physical positions of the markers along the chromosome (B73 RefGen v4). The y-axis represents the -log$_{10}$(p-value). The red horizontal dashed line indicates the significance threshold of 5. The black vertical dashed line and arrow show the position and direction of the ZmTIP1;1 gene, respectively. In the close-up Manhattan plots, the ZmHSF13 position is indicated by the green dashed vertical line. The eQTLs are highlighted by grey backgrounds. The names of the lead markers and markers analysed in the forward-backward analysis are indicated in the plots. Significant markers are coloured following their LD (measured as r²K) with the most significant marker.*

5.1.2 ZmTIP1;1 & ZmHSF13 interaction – distal eQTL

We then focused on the distal eQTL that highlighted the interaction reported in the regulatory network (Figure 32) between *ZmTIP1;1* and *ZmHSF13*. Interestingly, *ZmTIP1;1* and *ZmHSF13* were DE between WW and WD. Their expression was higher in WW than in WD (Figure 34). *Zm*HSF13 is a class A HSF and contains an AHA domain responsible for transcriptional activation, so it could influence *ZmTIP1;1* expression.

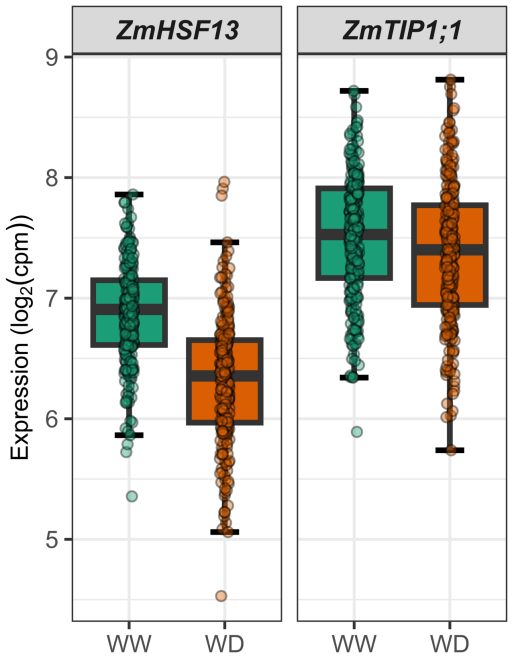


Figure 34: Expression in log_2(cpm) of ZmHSF13 and ZmTIP1;1 in WW (green) and WD (orange). *Each dot represents the adjusted mean expression value for a given hybrid. Boxplots indicate the distribution of expression among the population. Although for* ZmTIP1;1 *it is less clear, both genes were differentially expressed in at least 15% of the hybrids. The differential analysis was performed using a quasi-likelihood F test per variety and gene with an FDR-adjusted p-value below 0.05 were considered differentially expressed.*

This eQTL was present in both watering conditions but was captured by different markers. The lead markers were AX-91840471 in WD (-log_{10}(p-value) of 7.52) and AX-91393548 in WW (-log_{10}(p-value) of 5.17, Figure 33B and D). More markers were significant in WD than in WW. We did not know what the exact causal variant was. However, we observed that some diversity was present in the vicinity of *ZmHSF13*. For example, the lead marker in WD (AX-91840471) and AX-91203280, another marker of the eQTL containing the gene in WW, were annotated as upstream gene variants located 1,110 and 3,138 bp from the gene. Yet, to verify if the lead markers of the distal eQTL captured different variants, we performed a forward-backward analysis with MM4LMM (similar to the covariate GWAS performed in Maistriaux et al. (2024)). We tested several models by adding different covariates following the forward-backward procedure described in Segura et al. (2012). In short, we first tested the most significant marker alone and then added the lead marker of the distal eQTL containing *ZmHSF13*. Then, the markers were removed one by one, starting from the first added maker. We observed that the other markers were not significant when adding CNVMAIZE_INS_102207 in WW and CNVMAIZE_INS_89412 in WD as covariates.

This meant that the causal variant was probably captured by one of these markers.

Given the enrichment analysis, the presence of HSF-binding motifs in *ZmTIP1;1* promoter (*pTIP1;1*), the DEG analysis, and the fact that *Zm*HSF13 was in an eQTL, we decided to simply test whether *Zm*HSF13 influenced *pTIP1;1* activity, despite not knowing the exact causal polymorphism responsible for the expression changes. We conducted a dual-luciferase gene reporter assay in maize mesophyll protoplasts, where the activity of *pTIP1;1* was assessed by measuring the *firefly* luciferase activity. We had a control condition where there was no expression of *ZmHSF13* and a test condition where *ZmHSF13* was expressed under the control of *p35S* promoter (*p35S::ZmHSF13*). For each condition, we used a total of three plasmids: one to express the Renilla luciferase as an internal control, another to express the *firefly* luciferase under the control of the *AQP* promoter (reporter plasmid), and finally to express the TF (effector plasmid). Expression of the TF under the control of the *p35S* promoter led to a statistically significant increase in *pTIP1;1* activity (Figure 35A and Supplemental Table 3). However, we also observed that this overexpression also increased the activity of *Renilla* luciferase (internal control), possibly minimising the observed effect on *pTIP1;1* (Figure 35B). This was not an ideal situation, and we believe that this was *Zm*HSF13-dependant as other members from the laboratory working with tobacco protoplasts did not observe such change. Our hypothesis was that the *Zm*HSF13 could modify the activity of *p35S* promoter even though no known HSF binding site was found in its sequence. This experiment could therefore be tested using another constitutive promoter, like the maize *Ubi1* promoter (Christensen & Quail, 1996) in front of the *Renilla* luciferase cDNA, as well as the TF CDS. For the latter, the use of its endogenous promoter would be a possibility as far as this promoter is active in the leaf protoplasts.

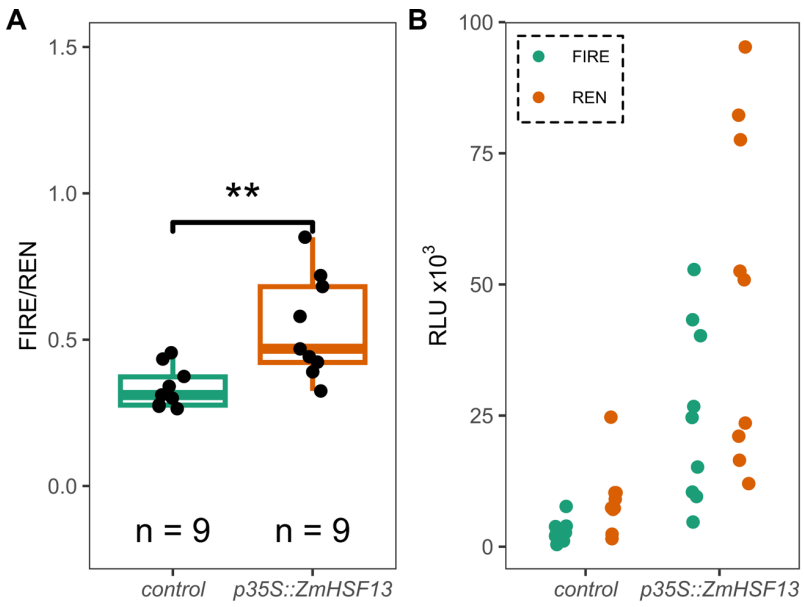


Figure 35: Activity of ZmTIP1;1 promoter upon ZmHSF13 expression determined by dual-luciferase assays in maize B104 mesophyll protoplasts. *In both conditions, control and* p35S::ZmHSF13*, three plasmids were used: one to express the Renilla luciferase as an internal control, one to express the firefly luciferase under the control of the* AQP *promoter as a reporter plasmid, and one to express the TF as an effector plasmid. In the* p35S::ZmHSF13 *plasmid,* ZmHSF13 *CDS was added downstream of the* p35S *promoter, while in the control condition, no CDS was added. (A) Plot of the ratio between firefly and Renilla intensities. The ratio difference was statistically tested with a mixed-linear model, and the significance is shown at the top of the boxplots (** p-value < 0.01), n is the total number of replicates (B) Plot of the measured relative light units (RLU x 10^3) for both the firefly and Renilla luciferase in each condition.*

5.1.3 Establishing ZmHSF13 binding landscape

A DAP-seq experiment was performed to identify the binding sites and targets of *Zm*HSF13 (O'Malley et al., 2016). We introduced *Zm*HSF13 CDS in the pIX-Halo vector that allowed its expression fused at the N-terminus to a Halo tag (Bartlett et al., 2017). The Halo-GFP fusion protein was used as a negative control. We generated three gDNA libraries from the mature zones of the 10th leaves at the 14-leaf stage. We performed DAP-seq on each library for both *Zm*HSF13 and the negative control. We confirmed that the fusion proteins were expressed and bound correctly to the magnetic beads by performing a western blot before and after the binding step. We observed a decrease in the protein quantity after the binding (Supplemental Figure 11).

We obtained an average of 22.25 million reads per sample. Read quality was assessed by the proportion of reads with a Q30 score, indicating a 1 in 1,000 chance of sequencing error per base. The sample with the lowest quality was GFP_LIB2, the negative control for the library n°2, where only 76% of the reads had a quality score of Q30. Low-quality reads and adapters were trimmed and

mapped onto the chromosomal DNA sequences of the B73 v4 reference genome (Jiao et al., 2017). The mapping quality was different for each sample (Supplemental Table 4), but, on average, 42% reads mapped to a single location in the genome (uniquely mapped).

Peaks were identified using the MACS2 algorithm, 32,511 peaks were present in all replicates, and six peaks were further discarded based on a low enrichment score compared to the GFP samples (logFC ≤ 0.5). Most of the peaks were mapped to the intergenic regions (86.7%) meaning that they did not map within +/- 3k bp from any annotation (Figure 36A). Peaks mapped in genic regions (4,457 peaks) were divided into four categories: 26.7% in promoters (3k bp upstream of the TSS), 17.4% in transcription termination sites (TTS, 3k bp downstream of the gene), 28.4% in exons (containing the untranslated regions), and 27.5% in introns (Figure 36B). We then identified the consensus motif using MEME (Bailey & Elkan, 1994) in the peaks from the genic regions. The top predicted motif was then compared with motifs published for Arabidopsis (O'Malley et al., 2016) using TomTom (S. Gupta et al., 2007). We observed that the most similar motif was the one detected for the *At*HSF3 TF, which also belongs to class A HSFs (Figure 36C). The motif contained an HSE (Xiao & Lis, 1988; P. Zhao et al., 2020), which is commonly recognised by HSFs (Y.-X. Lin et al., 2011).

We then verified whether *Zm*HSF13 binds to regions close to *AQP* genes. We observed peaks in the *ZmTIP1;1* promoter region and close to its TTS (Figure 36D). This confirmed that HSF could play a role in *ZmTIP1;1* transcription regulation. In addition to *ZmTIP1;1*, we observed it also binds to the *ZmPIP1;3* promoter (Figure 36E).

These results led us to conclude that the interaction between *ZmHSF13* and *ZmTIP1;1* detected in our network was relevant. Firstly, the presence of an HSE in *ZmTIP1;1* promoter suggested that an HSF TF could directly bind to it. Secondly, overexpression of *Zm*HSF13 in maize leaf protoplasts increased *ZmTIP1;1* promoter activity (Figure 35). Thirdly, direct binding of *Zm*HSF13 to *ZmTIP1;1* promoter was confirmed by DAP-seq (Figure 36). In addition, this experiment showed that *Zm*HSF13 could bind to the *ZmPIP1;3* promoter and identified HSE as the most enriched motif in MEME analysis. This observation made us confident about the DAP-seq data, which is an *in vitro* experiment, sometimes leading to false negative/positive results due to the absence of native chromatin (O'Malley et al., 2016). HSF are also known to form trimers *in vivo* upon stress sensing after release from heat shock proteins (Scharf et al., 2012). The observation that *Zm*HSF13 binds in *ZmTIP1;1* and *ZmPIP1;3* genes *in vitro* means that the trimers are either not required for DNA binding in these conditions or that the presence of HaloTag fused to *Zm*HSF13 did not compromise its oligomerisation.

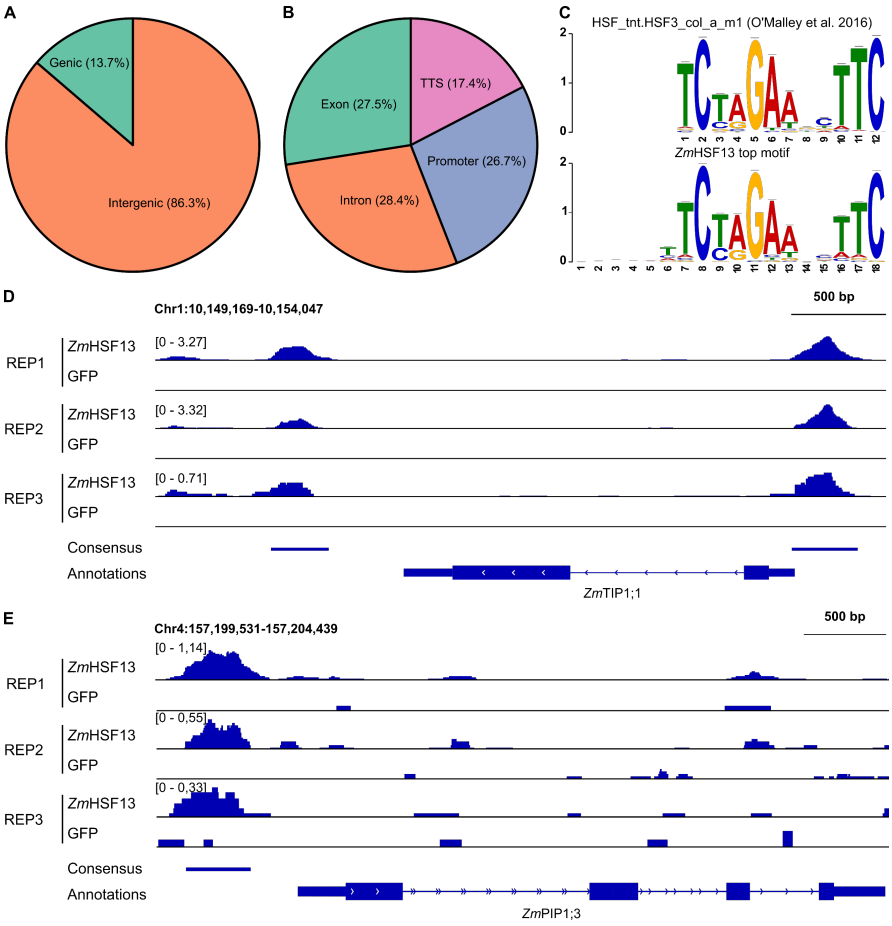


Figure 36: ZmHSF13 DAP-seq results. *(A) Pie chart showing the proportion of peaks falling in genic and intergenic regions. (B) Pie chart showing peaks in genic regions falling in exons, introns, promoter, and TTS. (C) Aligned AtHSF3 and ZmHSF13 motifs. (D-E) Visualisation of binding profiles (peak) generated from the DAP-seq experiments of ZmHSF13 and GFP (negative control) highlighting the binding to ZmTIP1;1 promoter and TTS (D) and ZmPIP1;3 promoter (E). The gene models are from B73 v4 reference genome (Jiao et al. 2017) at the bottom. Peak heights correspond to the number of reads normalised for sequencing depth.*

In order to assess the relevance of this biological interaction and gain more knowledge on the biological processes in which *Zm*HSF13 is involved, we performed a gene ontology (GO) analysis on its target genes (Figure 37 and Supplemental Table 5). The most significant terms were the response to heat (GO:0009408), protein folding (GO:0006457), response to temperature stimulus (GO:0009266), response to high light intensity (GO:0009644), and response to hydrogen peroxide (GO:0042542). This highlights its involvement in abiotic stress responses. As AQPs are involved in ROS signalling, the enrichment of other GO terms like response to ROS (GO:0006979) and response to oxidative stress

(GO:0000302) is of interest. These processes are somewhat related to heat stress. *Zm*TIP1;1 has shown to transports H_2O_2, highlighting its potential involvement in heat stress responses (Bárzana et al., 2014). Interestingly, responses to drought and heat stress share some similarities. Both can lead to stomatal closure to reduce transpiration but also limits CO_2 uptake, thus affecting photosynthesis. These stresses also induce oxidative stress due to the increase of ROS production. This can lead to cellular damage and affect overall plant health. Therefore, we hypothesise that *Zm*HSF13 activates *ZmTIP1;1* gene expression in either of these scenarios, resulting in more *Zm*TIP1;1 in the tonoplast to facilitate the ROS detoxification. However, further experiments are required to validate the physiological relevance of their interaction.

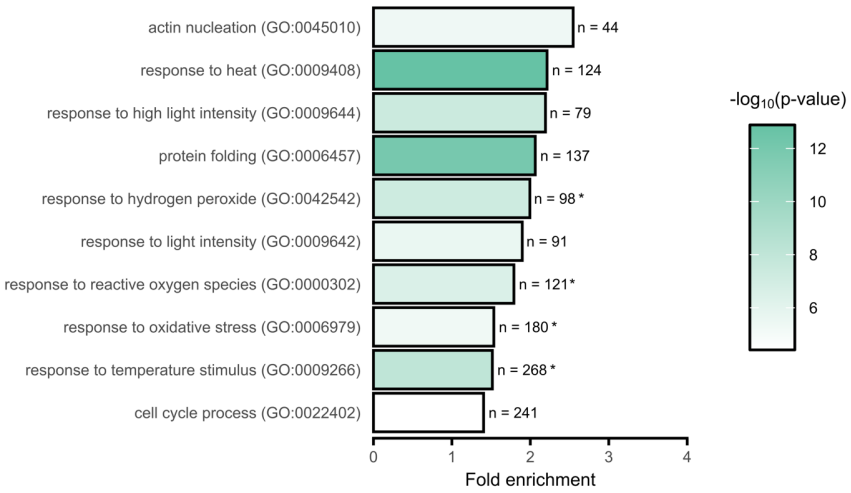


Figure 37: GO analysis of ZmHSF13 targets. *Targets were defined as genes to which ZmHSF13 bound within their promoter, intron, exon, or TTS. Their accession was retrieved and used in GOATOOLs to perform GO analysis with all the other B73 v4 genes used as the background. Significant GO terms were ordered following fold enrichment, and the top 10 most significant GO terms are displayed. The p-value significance is represented by the colour gradient. The number of genes involved in each biological process is indicated at the top of the bars. Biological processes containing* ZmTIP1;1 *are indicated by asterisks (*).*

5.2 *ZmPIP1;3* & *ZmPIP2;3* eQTLs

5.2.1 Local eQTLs

In both WW and WD, *ZmPIP1;3* showed a strong local eQTL captured by the lead marker CNVMAIZE_INS_69044 (-$\log_{10}$(p-value) of 33.08 in WW and 35.01 in WD, Figure 38A and C). Many small eQTLs were also located around this strong local eQTL. A similar pattern was observed by Maistriaux et al. (2024), where it was hypothesised that the lead marker created noise around itself. This suggested that the association of the local eQTL with the causal polymorphism was so strong that the ability to detect other associated markers was lost. Moreover,

many markers belonging to other eQTLs were not grouped in the local eQTL despite having an LD higher than 0.2 (the threshold used in the LD_{adj} method) (Figure 38B). We probably faced a similar scenario to that for *ZmTIP1;1* where an eQTL was "broken" because one marker pair had a lower LD than the threshold used in the LD_{adj} method and the LD_{win} method was unable to group these two eQTLs together. *Zm*NAC40 was located in a large distal eQTL on chromosome 4 at 28M bp of *ZmPIP1;3*. Many markers constituting the distal eQTL had an LD of approximately 0.3, with CNVMAIZE_INS_69044. These markers could capture the same causal polymorphism. Therefore, we performed the same analysis as for *ZmTIP1;1* with MM4LMM and observed that when the lead markers of the local eQTL and distal eQTL were in the same model, only the marker representing the local eQTL was significant. Thus, they may have captured the same causal polymorphism.

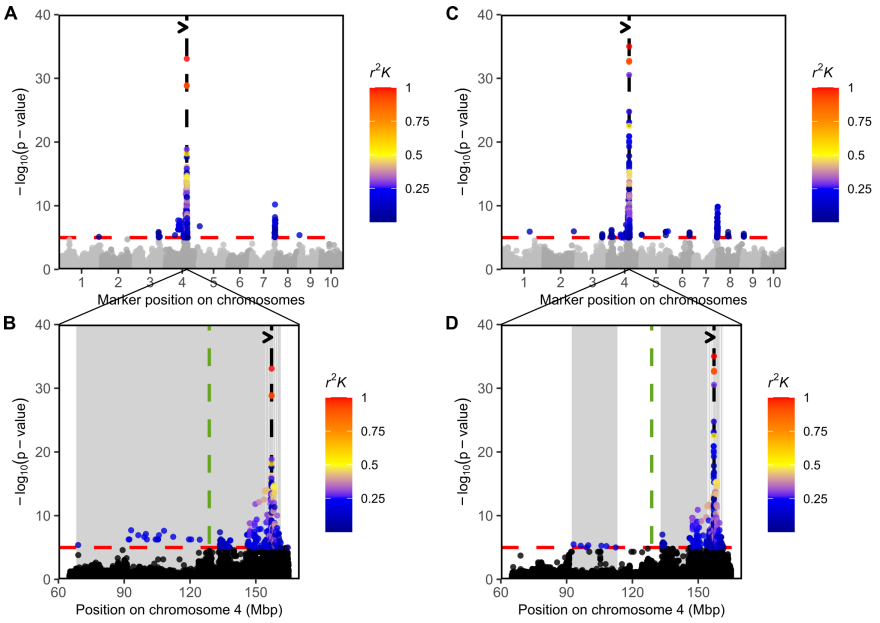


Figure 38: Full and close-up Manhattan plots for ZmPIP1;3 gene expression in WW (A and B) and WD (C and D). *The x-axis represents the physical positions of the markers along the chromosome (B73 RefGen v4). The y-axis represents the -log₁₀(p-value). The red horizontal dashed line indicates the significance threshold of 5. The black vertical dashed line and arrow show the position and direction of the ZmPIP1;3 gene, respectively. In the close-up Manhattan plots, the ZmNAC40 position is indicated by the green dashed vertical line. The eQTLs are highlighted by grey backgrounds. The names of the lead markers and markers analysed in the forward-backward analysis are indicated in the plots. Significant markers are coloured following their LD (measured as r²K) with the most significant marker.*

We faced the same scenario for *ZmPIP2;3*, which had strong local eQTLs that could impair the detection of other causal polymorphisms (Figure 39A and C). The lead marker representing the local eQTL was CNVMAIZE_INS_33292 with -

log_{10}(p-values) of 29.6 and 27.8 in WD and WW, respectively. *Zm*NAC40 was located 18M bp away from *ZmPIP2;3* in a distal eQTL represented by the lead marker AX-91347480 in WD (-log_{10}(p-value) of 6.49) and AX-91620487 in WW (-log_{10}(p-value) of 6.27) (Figure 39B and D). The forward-backward analysis showed similar results to those for *ZmPIP1;3* and *ZmTIP1;1*, where one of the lead markers of the local eQTL remained significant when both lead markers were in the model.

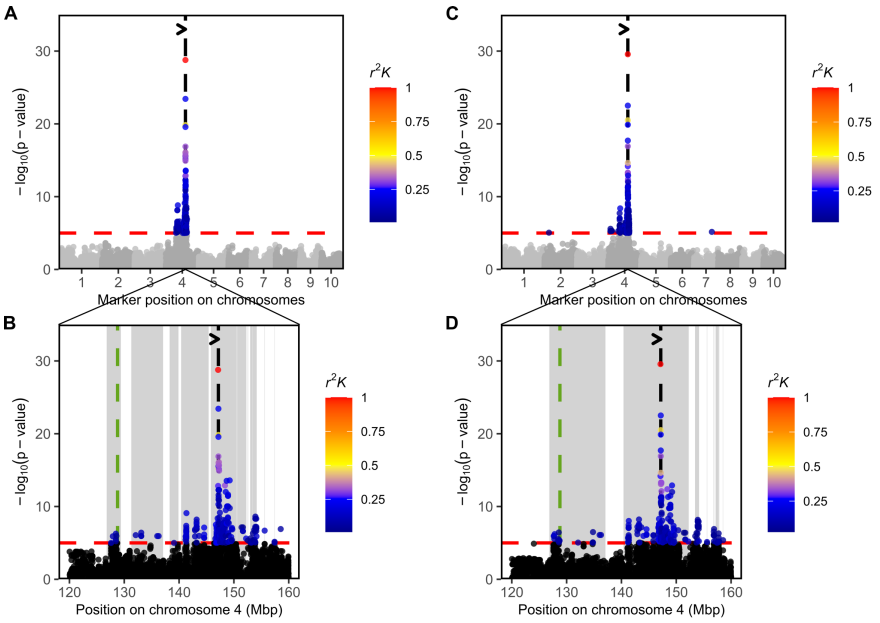


Figure 39: Full and close-up Manhattan plots for ZmPIP2;3 gene expression in WW (A and B) and WD (C and D). *The x-axis represents the physical positions of the markers along the chromosome (B73 RefGen v4). The y-axis represents the -log₁₀(p-value). The red horizontal dashed line indicates the significance threshold of 5. The black vertical dashed line and arrow show the position and direction of the ZmPIP2;3 gene, respectively. In the close-up Manhattan plots, the ZmNAC40 position is indicated by the green dashed vertical line. The eQTLs are highlighted by grey backgrounds. The names of the important markers are indicated in the plots. Significant markers are coloured following their LD (measured as r²K) with the most significant marker.*

5.2.2 Distal eQTLs

ZmNAC40 was found in *ZmPIP1;3* and *ZmPIP2;3* eQTLs in WW and both watering conditions, respectively. None of these were differentially expressed even though the trend showed that *ZmNAC40*, *ZmNAC47*, and *ZmPIP1;3* expression increased slightly in WD (Figure 40).

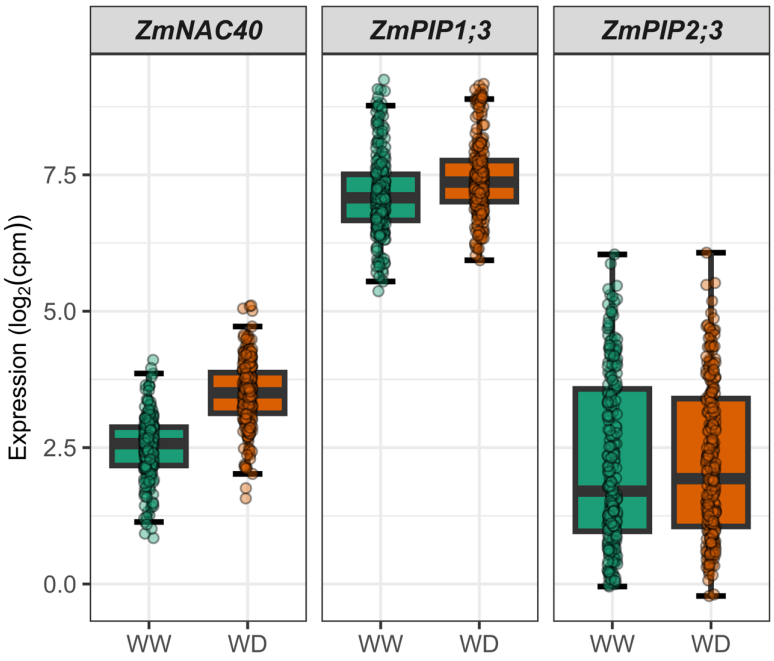


Figure 40: Expression in log₂(cpm) of ZmNAC40, ZmPIP1;3, and ZmPIP2;3 in WW (green) and WD (orange). *Each dot represents the adjusted mean expression value for a given hybrid. Boxplots indicate the expression distribution among the population. A differential analysis was performed using a quasi-likelihood F test per variety and gene with an FDR-adjusted p-value below 0.05 were considered differentially expressed. However, none of these genes was differentially expressed in at least 15% of the hybrids and could not be considered as differentially expressed.*

We decided to test the associations of *Zm*NAC40 with *ZmPIP1;3* and *ZmPIP2;3* expressions based on the results of the enrichment analysis, the presence of NAC TFBS in both promoters (*pPIP1;3* and *pPIP2;3*), and the physiological roles of NAC TFs. Unfortunately, *pPIP1;3* and *pPIP2;3* showed no activity in the dual-luciferase assay. The values for the firefly signal were equal to the background when no plasmid was added to the transfection mix, whereas the *Renilla* signal was clearly detected. These observations were different from what we observed for *pTIP1;1*. Several parameters were tested to resolve this problem, such as the quality and quantity of the plasmids added to the transfection mixture, the protoplast preparation procedure, and the inbreds used to prepare the protoplasts, but none of these modifications led to an improvement in the signal. These experiments were also tested in tobacco leaves protoplasts, but the promoter showed no activity in tobacco either. The viability of the protoplasts was also assayed through FDA staining after the isolation step and their integrity was verified by microscopy after transfection. The FDA viability test could be performed at more steps during the experiment. Similar problems were encountered by other members of our laboratory and other research groups (Dr Stam, personal communications) during their experiments, in which the

promoter of interest did not show any activity. One solution could be to use a *p35S* minimal promoter between the promoter of interest and the reporter gene cassette to obtain a basal level of expression. In parallel, the genetic constructs required for the dual-fluorescence reporter gene assay to test the *pPIP1;3*-*Zm*NAC40 interaction were obtained. In this assay, the promoter activity was quantified using the ratio between two fluorescent proteins, mVenus and mCherry. Two conditions were tested: a control condition, with no expression of *ZmNAC40*, and a test condition, with *ZmNAC40* expressed under the control of the *p35S* promoter (*p35S::ZmNAC40*). This assay relied on the use a one plasmid containing three cassettes with distinctive purposes. One enabled the expression of mCherry and served as an internal control, another one to express mVenus under the control of the *AQP* promoter which was the reporter cassette, and finally one to express the TF which was the effector cassette. In the control condition, no promoter was included upstream of *ZmNAC40* CDS. In this case, *pPIP1;3* showed a basal signal, and the expression of *Zm*NAC40 significantly increased its transcriptional activity (Figure 41 and Supplemental Table 6). This is interesting as we expected the dual-luciferase experiment to be more sensitive than the dual-fluorescence. These results highlight the relevance of the detected interaction as in addition to containing a NAC-binding site, *ZmPIP1;3* promoter's activity increased when *Zm*NAC40 was overexpressed. These promising results are similar to observations made in the framework of E. Teirlinckx's thesis on the interaction between *Zm*NAC24 and *ZmPIP2;5* which was detected in the PIP eQTLs generated in Maistriaux et al. (2024). Moreover, their involvement in water-related processes and the fact that there is evidence of NAC regulation of *AQP* expression is also encouraging (Pei et al., 2013; R. Wang et al., 2021).

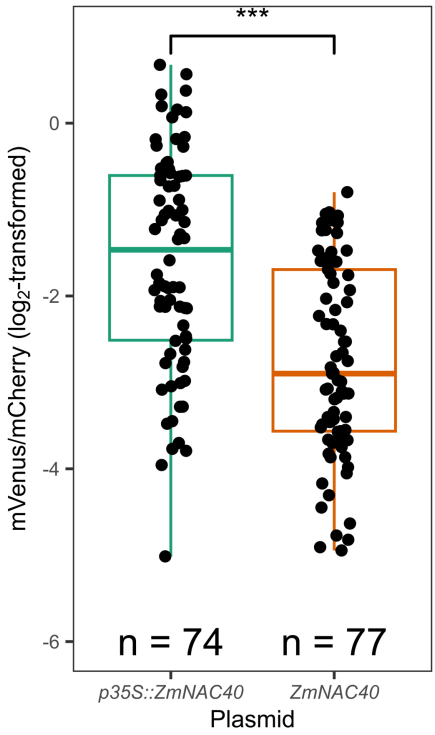


Figure 41: Activity of pPIP1;3 promoter upon ZmNAC40 TF expression determined by dual-fluorescence experiments in maize B73 mesophyll protoplasts. *In both conditions,* p35S::ZmNAC40 *and* ZmNAC40 *(control condition), one plasmid was used. It was designed to express a TF under the control of the* p35S *promoter, express mCherry fluorescent protein under the control of the* p35S *promoter as an internal control, and express mVenus fluorescent protein under the control of an* AQP *promoter. In the* ZmNAC40 *plasmid, there was no promoter upstream of the* ZmNAC40 *CDS. Plot of the log_2-transformed ratio between the mVenus and mCherry signals. The difference in ratio was statistically tested with a mixed-linear model, and the significance is shown at the top of the boxplots (*** p-value < 0.001), where n is the number of images taken per condition.*

General discussion & Perspectives

In this thesis, we aimed at building the *AQP* gene regulatory. We analysed 3'RNA-seq data from maize leaves in WW and WD conditions, identified 115 drought-responsive genes, and performed a GWAS to map eQTLs of *AQP* genes. We then combined this data with an enrichment analysis based on 45 gene regulatory networks to detect potential *AQP* expression *trans*-regulators. This led to the identification of 15 TFs potentially involved in the regulation of *AQP* expression. Two of these were selected for further validation on the basis of the biological relevance of their interactions with *AQPs* through gene reporter assays and DAP-seq. These analyses have set the foundation for exciting research. Two directions can now be followed, continuing the analysis of the GWAS results to detect new targets, or continuing to explore the biological functions of the detected interactions.

Before continuing with the analysis of the GWAS results, it would be interesting to convert the marker positions in v5 and verify that similar results are obtained as we observed some discrepancies between v4 and v5. The marker reads should be realigned to the new reference genome as conversion tools tend to make mistakes (Luu et al., 2020). This would allow the data to be updated and the latest update of the maize B73 reference genome and its related multi-omics datasets to be used (Hufford et al., 2021). Studying eQTLs for the entire gene dataset could also lead to more global conclusions about the diversity of gene expression in the panel, as we have only used *AQP* eQTLs. For example, we could identify regulatory hotspots shared among the eQTLs of multiple genes or build a regulatory network by establishing a hierarchy where one gene resides within the eQTL of another, which in turn is located within the eQTL of a third gene (S. Liu, Li, et al., 2020).

Identifying the exact causal polymorphisms responsible for expression variations could also be highly beneficial, as we obtained regions where potential regulators were located, including a high number of sequence variants, which could be either in the intergenic region or in the coding DNA (Bhat et al., 2021). Here, our approach was to focus on TFs that have biologically relevant functions for AQPs. Analysing the sequences of these genes across multiple genotypes may help identify causal polymorphisms. This can be done either using published genomes, as in Li et al. (2023), by targeted sequencing (Muraya et al., 2015), or by PCR amplifying and sequencing the region of interest in all the varieties, but this process would be highly time consuming. Other layers of information can also be added to eQTLs by comparing them with other high-throughput datasets. Generating Assay for Transposase-Accessible Chromatin (ATAC-seq) or MNase-defined cistrome-Occupancy Analysis (MOA-seq) data for

the DROPS diversity panel would enable the identification of accessible chromatin regions enriched for CREs in eQTLs. Liang et al. (2022) combined their eQTL data with MOA-seq from heat-stressed and control plants. They used this combination to prioritise candidates for characterisation. Integrating the data generated in other studies on maize leaf mature zone samples could also be an option, as these experiments are costly. Gomez-Cano et al. (2024) showed how the combination of multi-omics data such as eQTLs, DAP-seq, ChIP-seq can empower the building of GRNs. The prioritisation of candidate genes can be strengthened by linking our expression data to physiological measurements (stomatal conductance, transpiration or yield) performed on the same panel (Millet et al., 2016; Prado et al., 2018). The expression of *AQP* and the detected *TFs* could be analysed in lines showing contrasting phenotype. Transcriptome wide association study (TWAS) could be performed to see if some genes regulate both *AQPs* expression and a water-related phenotype, which would help to link AQPs to phenotypes, as done by Kremling et al. (2019), Liu et al. (2020) or Lin et al. (2022). Other methods using phenotyping and transcriptomic data have also been developed to discover gene functions and phenotypic causal variants (Fu et al., 2013; H. Liu et al., 2017; Miculan et al., 2021). Finally, studying the genetic variability of intergenic sequences could lead to the characterisation of CREs located in transposons that relocated in specific varieties, such as in Maistriaux et al. (2024), where a MITE insertion was detected in the *ZmPIP2;5* promoter in PH207. This could lead to interesting experiments with the aim to assess the effect of specific CREs on *AQP* promoters, such as a dual-luciferase experiments where several CRE versions are inserted upstream of the reporter gene to assess their effect on gene expression. The activity of the detected CREs could also be confirmed by analysing self-transcribing active regulatory region sequencing data from Jores et al. (2020) or Ricci et al. (2019). The eQTL data could also be compared with GRNs linking enhancer and target genes (Fagny et al., 2021) or GRNs inferred using the MINI-AC tool (Manosalva Pérez et al., 2024).

In order to better characterise the network interactions, the dual-luciferase assay could be adapted by inserting the promoter of the other *AQPs* upstream of the luciferase reporter gene. However, we could face the same problem as with *pPIP1;3* and *pPIP2;3*. Therefore, we could rethink the genetic constructs used in the assay. For example, in the laboratory, a genetic construct aimed at placing all the cassettes on the same plasmid was recently obtained by Golden Gate cloning, and it would be interesting to consider testing the interaction with these constructs. The dual-luciferase assay can be performed using another constitutive promoter such as the maize *Ubi1* promoter (Christensen & Quail, 1996) upstream of the internal control and effector genes. The use of a minimal *p35S* promoter between the tested promoter and reporter gene would also allow basal expression of the reporter gene, which would solve the problem we faced when the activity of our tested promoter was too low. EMSAs would be another

method for validating *in vitro* the interaction between a given TF and the promoter of its associated *AQP*. However, the binding should not be tested with the whole promoter but instead on smaller fragments. In Liu et al. (2023), the exact binding site of *At*PLATZ4 was identified through an iterative process. . They started with ~400 bp fragments, then progressively reduced the binding region by dividing the fragments that showed binding activity into smaller ~200 bp sections, repeating this until they reached fragments of ~30 bp. Moreover, the EMSA has already been set up in our laboratory. It could also be used to validate the peaks observed in the DAP-seq results. A subcellular localisation assay could be performed by fusing the TF coding sequences with a fluorescent protein, such as mVenus or mYFP, to verify that they localise in the nucleus and could therefore affect gene expression (Pei et al., 2013). However, some TFs are in the cytoplasm under normal conditions and translocate to the nucleus upon stress sensing, which must be taken in account when performing the experiment (Pireyre & Burow, 2015).

Arabidopsis or maize transgenic lines overexpressing or silencing a target *TF* gene can be generated to investigate its regulatory roles. These lines can be subjected to RNA-seq analysis to identify DEGs associated with the TF activity. A co-expression network could then be constructed, and a GO analysis of the different co-expression clusters could reveal the biological processes regulated by the TF. *AQP* genes may appear among the DEGs, potentially confirming their regulation by this TF. To validate the involvement of TFs in specific biological processes, physiological traits should be measured. Transgenic plants could be exposed to various stress conditions, such as drought or heat stress, to determine whether deregulation of *TF* expression has a beneficial or detrimental effect on plant survival. Traits related to water transport, such as stomatal conductance, ABA sensitivity, transpiration rate, or root hydraulic conductance, could be measured to assess the role of TFs in the drought stress response through water movement regulation and to formulate hypotheses about the regulatory pathway upstream of *AQP* genes. In addition, we highlighted that *ZmHSF13* may be involved in the oxidative stress response, and it could be interesting to investigate the tolerance of plants to this stress, which often results from drought or heat waves. Oxidative stress is triggered by excessive ROS production, which can have a negative impact on membranes and proteins and trigger cell death (Bienert et al., 2007). Measuring the activity of antioxidant enzymes such as superoxide dismutase, peroxidase, and catalase gives insights into the plant ability to scavenge ROS, while measuring the malondialdehyde content indicates the level of oxidative damage caused to cell membranes. These parameters are commonly used to measure a plant's stress response to oxidative stress (Z. He et al., 2019; L. Wang et al., 2018). Another interesting parameter to measure would be the proline content. In plants, proline can protect cells by serving as an osmotic agent and radical scavenger under biotic

or abiotic stress. For example, overexpression of *ThPIP2;5* enhanced proline biosynthesis, leading to increased plant tolerance to salt and osmotic stress (L. Wang et al., 2018). In our lab, transgenic Arabidopsis plants overexpressing *ZmHSF13* are currently being generated. Their generation will take weeks, but it will enable us to measure water-related physiological traits and demonstrate the potential joint effect of *ZmHSF13* and *ZmTIP1;1*.

One aspect that was not addressed in this project due to time constraints was the comparison of the eQTLs between WW and WD conditions. After identifying DEGs between WW and WD, each condition was treated separately. It would be interesting to analyse the difference in eQTLs between both conditions. One way to do this would be to perform a GWAS on the *AQP* expression ratio between WW and WD. This would enable the identification of the regions directly involved in the expression response. Starting with *ZmTIP1;1* could be relevant, as it was also differentially expressed.

In the long term, these data provide valuable information for the study of *AQP* gene expression and resources for linking traits measured in the framework of the DROPS project to their underlying molecular mechanisms. Once linked to phenotypes, these results will highlight potential targets which breeders will need to focus on and will make it possible to select genotypes that are better adapted to adverse conditions.

In conclusion, we identified DEGs in the mature zone of maize leaves under two watering conditions: WW and WD. We also detected potential regulatory variants controlling *AQP* gene expression under both conditions on a genome-wide scale. We then focused on candidate TFs within the eQTLs that we prioritised using a gene regulatory network and analysed the biological relevance of the detected associations. Finally, we selected two associations for functional validation using the dual-luciferase gene reporter assay, DAP-seq, and dual-fluorescence gene reporter assay. Further experiments are required to assess the physiological relevance of the validated associations. However, this work provides valuable information for further research into the transcriptional regulation of *AQP* genes.

Materials & Methods

Identification of novel maize *AQP* genes

Protein sequences of maize AQPs reported by Chaumont et al. (2001) and Lawrence et al. (2013) were retrieved from the NCBI database using EDirect. These were then used as queries in a BLASTp search against the NCBI database to identify novel *AQP* genes. Accessions of sequences with more than 30% identity were downloaded, duplicates were removed, and each accession was analysed individually. Particular attention was paid to the length of the protein sequence, number of predicted transmembrane domains, presence of NPA motifs, and CD-search predicted family (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). TM domains were localised using DeepTMHMM (Hallgren et al., 2022).

The NPA motifs were searched with an R script using the str_locate_all function implemented in the stringr v1.5.0 package (Wickham, 2023). Various combinations of NPA, such as NPI, NPL, NPS, NPV, and NPT, were searched. Sequences with more than two motifs were examined manually and the positions were aligned using UniProt prediction. Motifs consistent with the prediction were retained as the main motifs.

Accessions with six transmembrane domains, two NP(A) motifs, and those predicted to belong to the major intrinsic protein (MIPs) superfamily were retained and defined as AQPs. Using the protein accession number, the gene model IDs were obtained for B73 reference genome v4 (Zm-B73-REFERENCE-GRAMENE-4.0 downloaded from https://www.maizegdb.org/ [accessed on 14-06-2021]) and v5 (Zm-B73-REFERENCE-NAM-5.0 downloaded from https://www.maizegdb.org/ [accessed on 14-06-2021]), each giving access to the physical location in these genome versions. The positions provided by NCBI were also collected and used as references in this project when sequence data needed to be downloaded. The pan-gene IDs of *AQP* genes were also included (Hufford et al., 2021). They were retrieved from the pangene files downloaded from the NAM genomes project website (14-11-2021).

Phylogenetic analysis

AQP protein sequences were aligned using the Molecular Evolution Genetic Analysis v10 (MEGA) software (Kumar et al., 2018). Alignment was performed using the muscle method, and phylogenetic analysis was performed using the maximum likelihood method based on the Jones-Taylor-Thornton (JTT) matrix model with 1000 bootstraps. The new AQPs were named according to the old nomenclature and their proximity to previously identified AQPs.

Intergenomic comparison of maize *AQP* genes

AQP families were constituted for 30 recently published maize genomes (25 from the NAM found lines project (Hufford et al., 2021) and four from the Technical University Munich (Haberer et al., 2020)). For this purpose, the sequences obtained from the B73 v5 genome were used as queries in a BLASTp search against databases containing the protein sequences of all genotypes (downloaded from https://www.maizegdb.org/ [accessed on 14-06-2021]). Hits with at least 70% query coverage, the highest percentage of identity, and the minimum score were considered AQPs, as few mutations were expected in their coding sequences between maize cultivars. The positions of the *AQP* genes in each genome were determined using the gene model ID in the gff3 annotation files (downloaded from https://www.maizegdb.org/ [accessed on 14-06-2021]).

Plant material

For the 3'RNA-seq experiment, plants were grown on the PhenoArch platform (LEPSE, INRAE, Montpellier) under two watering regimes: WW (soil water potential of -0.05MPa) and WD (soil water potential of -0.3 MPa). These conditions were maintained constant by watering the pots three times a day with watering stations combining ST-Ex weighing terminals with 1 g accuracy (Bizerba, Balingen, Germany) and high-precision pump-watering 520 U stations (Watson Marlow, Wilmington, MA, USA). The genotypes were replicated three times under both conditions for a total of 1,524 samples. Within each replicate block, the genotypes were randomised, but no randomisation occurred between the two conditions. Mature zones from the 10th leaf were sampled at the 14-leaf developmental stage. Once harvested, the samples were immediately frozen in liquid nitrogen and stored at -80°C. The entire sampling process lasted for three successive mornings (~3-h ranges occurring between 7:30 and 10:30 am).

RNA extraction

RNA extraction was performed during a previous PhD project (Maistriaux et al., 2024) using a Mixer Mill MM 400 grinder (Retsch, Haan, Germany) to grind the tissues with metallic beads, and the Spectrum™ Plant Total RNA kit (Sigma-Aldrich, St. Louis, MO, USA) to perform the rest of the extraction. The total RNA concentration was then measured in ng/µl using a Nanodrop ND-1000 spectrophotometer (Isogen Life Science, Utrecht, The Netherlands), and RNA quality was determined by analysing the RNA Integrity Number (RIN) obtained with a 2100 Bioanalyzer (Agilent technologies, Santa Clara, CA, USA). All replicates of 20 genotypes were retained, while for 234, only two replicates were selected based on their quality. When the quality of the first and second replicates was good, they were chosen; however, when one of the two had a bad quality, the third replicate was used to replace the one with poor quality. In the

framework of the Amaizing Consortium, the 1,056 samples were sent to Limagrain (Saint Beauzire, France) for library preparation and 3'RNA-seq.

Library Preparation & Sequencing

These steps were performed by H. Duborjal from Limagrain. The sample concentrations were normalised to 100 ng/µl using RNAse-free water. Sequencing libraries were then synthesised using the QuantSeq 3' mRNA-Seq Library Prep Kit FWD (Lexogen, Vienna, Austria). After library preparation, the quality and quantity of each library were assessed using LabChip GX microfluidic electrophoresis (Perkin Elmer, Waltham, MA, USA).

Sequencing was then performed on 11 high-output flow cells on a Next Seq 550 sequencer (Illumina, San Diego, CA, USA), with each flow cell containing 96 libraries equimolarly pooled. Single reads of 75 bp were generated according to the manufacturer's instructions.

The resulting sequences were trimmed using Cutadapt v1.8 in single-read mode. After removal of the adapter sequence and quality trimming (q10), trimmed reads of at least 50 bp were kept for alignment.

STAR v2.6.0b was used to map reads on the B73 v4 reference genome (Zea_mays.B73_RefGen_v4.dna.toplevel.fa.gz) and the file from AGPv4.38 (Zea_mays.AGPv4.38.chr.gtf) was used for annotation. These files were downloaded from the Ensembl Gramene website (release 38; http://ftp.ensemblgenomes.org/pub/plants/release-38/). Reads per gene were quantified using the --quantMode option in STAR. The number of reads aligned to the 1st read strand for each gene was obtained from the STAR output. This resulted in a count matrix containing the raw number of reads aligned to each gene in each sample.

Exploratory data analysis

PCA was performed using the prcomp function with default settings in the stats v4.2.3 R package. For the analysis, the data were normalised using the TMM method from edgeR v3.38.4 (Robinson et al., 2010; Robinson & Oshlack, 2010) and transformed to $\log_2$(cpm). The first two PCs were then plotted against each other, and the samples were coloured according to the group they belonged to for different cofactors.

3'RNA-seq preprocessing

Two types of preprocessing were performed on the raw expression data to remove unwanted batch effects highlighted in the PCA.

For the differential expression analysis, the ComBat_seq function from the sva v3.44.0 package (Leek et al., 2012) was used to remove the time bin effect on the data before any normalisation or transformation, a common procedure in the literature. The time bin was set as the batch effect in the function argument. Variety ID and treatment were set as covariates to maintain their effect in the data.

For GWAS data preparation, different R packages designed for RNA-seq analysis were compared to verify whether some of them had a better performance. An R pipeline was developed, and the code can be found on GitHub (max-laurent/amazing-rnaseq). Almost all functions used come from two packages, edgeR v3.38.4 and DESeq2 v1.36.0 (Love et al., 2014). One package designed solely for gene filtering, HTSFilter v1.36.0 (Rau et al., 2013), was also included.

Differential expression analysis

DE analysis was performed using edgeR v3.38.4. The corrected expression data were separated by maize line, and lines with missing data in any of the two watering conditions were discarded. This resulted in a total of 252 subgroups. In this step, all subgroups were treated independently. Genes with at least 1 cpm in 10% of the samples were kept. The data was normalised using the TMM normalisation. A comparison between WW and WD expression levels was performed for each subgroup using a quasi-likelihood F-test. This test was implemented with two edgeR functions: first, the glmQLFit function with the robust argument set to TRUE, and then the function glmQLFTest. P-values were corrected for multiple testing using the Benjamini-Hochberg procedure applied with the p.adjust function from the stats v4.2.3 package. P-values of the 252 subgroups were extracted, and genes were considered differentially expressed when their FDR-adjusted p-value was below 0.05 in at least 15% of the hybrids.

Gene ontology enrichment

The find_enrichment.py function from the Python package GOATOOLS v0.4.7 (Klopfenstein et al., 2018) was used to perform a pathway enrichment. The gene IDs of the DEGs were tested against all the expressed genes in the expression data. The maize-GAMER GO annotation was retrieved from maize-gdb and used as the association file in the function that maps a gene ID to a GO category. The .obo file named go-basic.obo containing the latest GO category annotation was downloaded from the website: http://current.geneontology.org/ontology. Biological pathways were considered enriched when their FDR-adjusted p-value was below 0.05.

3’RNA-seq spatial correction

MLMs were used to obtain adjusted mean and broad-sense heritability values (referred as heritability) for all genes expressed in the dataset. Gene expression was adjusted for the effects observed in PCA. The R package statgenSTA v1.0.11 (Rossum, 2024) was used to format the data and fit MLMs. The method running inside this package came from the SpATS package v1.0-18 (Rodríguez-Álvarez et al., 2018), which estimates the variance components with restricted maximum log-likelihood (REML) and performs a P-spline adjustment of the spatial effects. The number of segments used for spline adjustment was set to 20.

The genotype was first fitted as a random effect, and the value of the heritability was obtained. This value represents the percentage of variance in gene expression explained by the genotype. The genotype was then fitted as a fixed effect to obtain the BLUEs of the genotypic mean, which was the adjusted mean for the gene. These BLUEs were used to perform a GWAS. The covariate used as the fixed effect was the sampling time bin. The covariates used as random effects were the plant positions in the greenhouse coded with row ID (rowID) and column ID (colID). The residuals were always fitted as random effects.

Genotyping

Two marker datasets were generated for this panel. One dataset included SNPs generated by three genotyping technologies (Genotyping-By-Sequencing, Illumina Infinium 50 K, and Affymetrix Axiom 600 K arrays), as described by Negro et al. (2019). The other dataset included PAVs generated using the genotyping tool developed by Mabire et al. (2019). The data was supplied privately by Dr Stéphane Nicolas (INRAE, Le Moulon).

The 3’RNA-seq reads were aligned on v4 of the B73 reference genome (Jiao et al., 2017), which was the standard at the time of the 3’RNA-seq experiment. To maintain consistency between the marker physical positions and the gene positions, all the data needed to be mapped on the maize B73 v4 reference. In Negro et al. (2019), the SNP positions were mapped on v2 but were converted to v4 by Delphine Madur, who shared the updated physical positions privately. However, PAVs were only mapped on v2 and needed to be converted. The Python package CrossMap v0.5.4 (H. Zhao et al., 2014) was used for this purpose. The main function of the CrossMap needed two inputs: the v2 InDels positions in a bed file and a chain file that contains the pairwise alignment between v2 and v4 B73 reference assemblies (downloaded on: https://ftp.gramene.org/release-68/assembly_chain/zea_mays/). During this process, some markers could not be mapped in v4 and were classified as unmapped; therefore, they were not used in the eQTL analysis. Some markers were mapped to contigs of v4; these were regions that were not located on any chromosome. These markers were also

discarded for eQTL analysis. In total, the genotyping dataset contained 1,004,516 markers, including 934,530 SNPs and 69,986 PAVs.

In general, when an RNA-seq experiment is performed, it is possible to detect polymorphisms using the aligned reads. These polymorphisms can then be used as markers in GWAS. In our case, we had 3'RNA-seq data, which did not yield enough sequencing depth, and therefore the reads did not help us to detect new polymorphisms (Quinn et al., 2013). Moreover, there would be a clear bias for markers located at the 3' end.

eQTL identification

The software FaST-LMM v2.07 (Lippert et al., 2011) was used to perform marker associations with individual *AQP* gene expression for each watering condition. The first time, the R pipeline used was the same as that developed by Negro et al. (2019); however, the latter was translated into another programming language called NextFlow. This translation halved the execution time, and the pipeline is now available on GitHub (max-laurent/nf_pipeline_gwas).

To detect marker-phenotype associations, we used the following MLM:

$$Y = \mu + X\beta + G + E$$

Where Y is the vector of BLUE expression values, μ is the overall mean, X is the vector of marker values, β is the vector of coefficients of the marker reflecting the effect of alleles on the expression value, G and E represent random genetic effects and residual effects, respectively. The variance-covariance matrix of G was calculated using the kinship matrix as in Rincent et al. (2014) and Negro et al. (2019), using all SNPs, except those located on chromosome 11.

The percentage of phenotypic variance explained by each marker was calculated using the formula described by Magee (1990) and was considered a good estimate of R^2 for MLMs in Sun et al. (2010):

$$R^2_{LR} = 1 - \exp\left(-\frac{2}{n}(logL_M - logL_0)\right)$$

where n is the number of individuals, $logL_M$ and $logL_0$ are the logarithms of the maximum likelihood of the marker-containing model and the intercept-only model, respectively.

The MAF of each marker was calculated in both conditions using the plink v1.07 --freq function. MAF differed under both conditions because the number of individuals was different. Markers with a MAF > 0.05 were retained for the analyses. The allele of reference was set to the B73 allele for all markers. A $-\log_{10}$(p-value) > 5 was chosen as the significance threshold. The LD between the most significant marker and all other significant markers was calculated with the

r^2K formula described by Negro et al. (2019), using the same kinship matrix as in the GWAS. This is the only LD formula used in this study.

Significant markers were then grouped into eQTLs using the LD_{adj} method with a threshold of 0.2 and the LD_{win} method with a threshold of 0. In the LD_{adj} method, the LD between two adjacent markers was calculated, and if it was higher than the threshold, the markers were grouped in the same eQTL. In the LD_{win} method, LD windows calculated for each marker in Negro et al. (2019) and converted to v4 positions were used. The overlap between LD windows of markers at the boundaries of consecutive eQTLs was calculated, and, when the overlap was above the threshold, the eQTLs were merged. An overlap value of 0 meant that the window boundaries were located at the same position, a negative value meant that there was no overlap, and a positive value indicated that there was an overlap of at least 1 bp between the windows. A lead marker was then defined for each eQTL as the most significant marker, and its effect was used to explain the eQTL effect.

Marker annotation

To detect the effect of the SNPs on the coding DNA sequences (CDS) of the genes in maize B73 v4, we used the software SnpEff v5.2c (Cingolani et al., 2012). FASTA sequence and gff3 annotation files were required to perform this analysis and were downloaded from maizegdb. The marker file also had to be converted to the vcf format. This analysis yielded an annotated vcf, which could then be used to study the effect of significant markers.

Enrichment analysis of TF-AQP target gene associations

TF-target gene associations were retrieved for 45 co-expression networks built by Zhou et al. (2020). The networks were downloaded from their publication, and the top 100k edges (strongest TF-target gene associations) from the networks were extracted. The TF-target gene associations were then merged into a single dataset. To perform the enrichment analysis, the number of distinct targets for each TFs was counted, and within these targets, the number of *AQP*s was also counted. A Fisher's exact test was performed with the fisher.test function from the stats v4.2.3 package to check whether *AQP*s were overrepresented in the targets compared to the other genes. The p-values were corrected for multiple testing using the Benjamini-Hochberg procedure implemented in the p.adjust function from the stats v4.2.3 package. Adjusted p-values lower than 0.05 were considered significant. TFs with a significant test were considered enriched for *AQP*s.

To add information about the TF families, the database from PlantTFDB4.0 (Jin et al., 2015, 2017) was downloaded. However, geneIDs in the latter were in v3 and thus needed to be converted to v4 to be usable in this project. Two different

conversion tools were tested: the Grassius (Yilmaz et al., 2009) and maizeGDB (Woodhouse et al., 2021) translation tools. Their output was compared and a new database containing the geneIDs in v4 was created.

TF binding motifs in AQP promoters

The promoter regions (2-3k bp upstream of the gene) of the *AQP* genes for which enriched TFs were detected within the eQTLs were extracted from the B73 v5 reference genome FASTA file with homemade scripts. These were then uploaded in the PlantRegMap tool from PlantTFDB4.0, and predictions of the TF binding sites located in the *AQP* promoter regions were obtained.

Forward-backward analysis

To perform the forward backward analysis, we used the MM4LMM R package v3.0.2. The MMEst function was used to fit the three different models shown in Table 5. Models 1 and 3 were similar as the one used for the GWAS. The p-value of each covariate was obtained using the AnovaTest function with the Type 1 parameter. The entire code is available on GitHub (max-laurent/fw_bw). This method enabled us to test the independence of the most significant markers of the eQTLs of interest.

Table 5: Models tested in the forward-backward analysis. *$\boldsymbol{Y}$ is the vector of BLUE expression values, M_{local} is the vector of the lead marker values from the local eQTL, M_{distal} is the vector of the lead marker values from the distal eQTL, K is the kinship matrix, and E is the error term.*

Name	Mathematical representation
Model 1	$\boldsymbol{Y = M_{local} + K + E}$
Model 2	$\boldsymbol{Y = M_{local} + M_{distal} + K + E}$
Model 3	$\boldsymbol{Y = M_{distal} + K + E}$

Genetic constructs

To characterise the effects of the TFs selected in the eQTLs on the *AQP* promoter, we performed two types of transactivation assays: a dual-luciferase reporter gene assay and a dual-fluorescence reporter gene assay. Both assays are based on the presence of a reporter gene downstream of the promoter of interest and expression of the associated TF. In the dual-luciferase assay, three plasmids were used: an effector plasmid that allowed the expression of the TF, a reporter plasmid that allowed the assessment of the effect of the TF on the promoter activity, and an internal control that was stable in all conditions. Two conditions were tested: one in which the TF was expressed (test) and another in which the TF was not expressed (control). The dual-fluorescence assay is based on the same principle. However, all cassettes were located on the same plasmid. To

create these plasmids, two types of cassettes were required: the TF CDS and the promoter region of the *AQP* of interest.

TF CDS

The TFs CDS were ordered from the maize TFome collection (Burdo et al., 2014) on the ABRC website. The collection contained the CDS of many TFs in L1-L2 Entry clones ready for Gateway (GW) cloning (Invitrogen, Waltham, MA, USA). We ordered the CDS of 11 TFs, and all plasmids are listed in Supplemental Table 8. Upon arrival, the plasmids were sequenced for verification using M13 forward (GTAAAACGACGGCCAGT) and reverse (GGAAACAGCTATGACCATG) sequencing primers. This allowed the detection of two mutations in the *ZmNAC40* CDS (pUT6157) that modified its protein sequence compared to the sequence published in Burdo et al. (2014). These were corrected by triple PCR, and the resulting sequence was used to build another L1-L2 Entry clone. The primers used for triple PCR are listed in Table 6.

Table 6: Primer's list used to generate the genetic constructs described in this work. *Gateway cloning sites are underlined and were added at the boundaries of certain cassettes to create entry vectors.*

Primer name	Sequence (5'-3')	Use
NAC40_GW_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGCACGACGTGCACCA	Gateway forward primer (attB1) – triple PCR
NAC40_GW_R	GGGGACCACTTTGTACAAGAAAGCTGGGTTGAAGCCCCAGAAGGAGGCCG	Gateway reverse primer (attB2) – triple PCR
NAC40_corr1_F	CCGCTGATGGACGTGTCCGGCG	Primer used to correct the D219G – triple PCR
NAC40_corr1_R	CGCCGGACACGTCCATCAGCGG	Primer used to correct the D219G – triple PCR
NAC40_corr2_F	AGGACACCGGCGTCACCTCGGA	Primer used to correct the V319I – triple PCR
NAC40_corr2_R	TCCGAGGTGACGCCGGTGTCCT	Primer used to correct the V319I – triple PCR
p35S_HpaI_F	GGTGGTGTTAACACTAGAGCCAAGCTGATCTC	Primer used to clone the p35S-ZmNAC40-t35S cassette in pGB0110m
t35S_XbaI_R	GGTGGTTCTAGAAGGTCACTGGATTTTGGTTT	Primer used to clone the p35S-ZmNAC40-t35S cassette in pGB0110m
NAC40_HpaI_F	GGTGGTGTTAACATGGAGCACGACGTGCACCA	Primer used to clone the ZmNAC40-t35S cassette in pGB0110m
pPIP1;3_USER_F	GGTCTTAAUCAAAAGAATTTTGCGATATC	Primers used to insert pPIP1;3 in pGB0110m
pPIP1;3_USER_R	GGGGTTAAUCTTTGCTCCCCGGCCCCACGCTACT	Primers used to insert pPIP1;3 in pGB0110m
pPIP1;3_GW_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCAAAAGAATTTTGCGAT	Gateway forward primer (attB1)
pPIP1;3_GW_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTTGCTCCCCGGCCCCACGCTACT	Gateway reverse primer (attB2)
pPIP2;3_GW_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGGCTAGTTTGGTGGTTGGG	Gateway forward primer (attB1)
pPIP2;3_GW_R	GGGGACCACTTTGTACAAGAAAGCTGGGTGTCAGGAGACCTGCAGCAGC	Gateway reverse primer (attB2)
pTIP1;1_GW_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATCGGTTCGATTCAGCGT	Gateway forward primer (attB1)
pTIP1;1_GW_R	GGGGACCACTTTGTACAAGAAAGCTGGGTTTTCACGGCCCTCTAGCTA	Gateway reverse primer (attB2)

AQP promoter amplification

For *ZmPIP2;3*, a region spanning from the start codon to 2k bp upstream of the gene was selected as the promoter (*pPIP2;3*). In the design, B4-B1R Gateway cloning sites were added around the promoter, and the whole cassette was flanked by additional EcoRI and BamHI restriction sites. The cassette was then ordered from GenScript (Piscataway, NJ, USA). They synthesised and inserted it into a pUC57 plasmid. For *ZmPIP1;3* promoter (*pPIP1;3*), a 5k bp region was first amplified by PCR from B73 gDNA and inserted in pGEM-T Easy (Promega, Madison, WI, USA) then a 3k bp region was reamplified with primers containing B4-B1R Gateway cloning sites. Both *pPIP1;3* and *pPIP2;3* were designed to be inserted in L4-R1 entry clones but the final plasmids used for the dual-luciferase reporter gene assay required L1-L2 recombination sites. Therefore, new primers were designed with the desired B1-B2 sites, and the fragments were amplified by PCR. In the case of *ZmTIP1;1* promoter (*pTIP1;1*), a 2k bp region upstream of the start codon was directly amplified with primers containing B1-B2 recombination sites from gDNA extracted from maize B73 leaves with the Wizard® genomic DNA purification kit (Promega, Madison, WI, USA). After amplification, the PCR fragments were purified from a 1% agarose gel using the innuPREP DOUBLEpure kit (Analytik Jena, Jena, Germany) and used in a BP GW reaction (Invitrogen, Wlatham, MA, USA) with pDONR221 P1-P2 to create L1-L2 entry clones. All PCR primers are listed in Table 6. The other promoters listed were amplified from the same gDNA as *ZmTIP1;1*, but with regular primers, and then inserted by ligation with T4 DNA ligase (Promega, Madison, WI, USA) in pGEM®-T Easy after undergoing an A-tailing (NEB, Ipswich, MA, USA). Inserted fragments were sequenced and reamplified with primers containing B1-B2 recombination sites to create L1-L2 entry clones.

Dual-luciferase plasmids

Backbone plasmids for the dual-luciferase reporter gene assay were ordered from the VIB website. p2GW7 was chosen as the effector plasmid because it contained a R1-R2 GW cassette between the *p35S* promoter and *t35S* terminator, thus allowing TF expression. In the control condition, empty p2GW7 was used to keep the amount of DNA constant between the conditions. pGLW7 was chosen as the reporter plasmid to express *firefly* luciferase under the control of the *AQP* promoters. It contained a R1-R2 GW cassette upstream of the *firefly* luciferase CDS. Both pGWL7 and p2GW7 plasmids were released by Karimi et al. (2002). This was convenient, as all final plasmids could be obtained through one LR reaction (Invitrogen, Waltham, MA, USA) using the entry clones previously described. We used the plasmid pGB0109 as the internal control because it allowed the expression of *Renilla* luciferase under the control of the *p35S*

promoter (Sarrion-Perdigones et al., 2013). All the plasmids and their antibiotic resistance profiles are listed in Supplemental Table 9.

Dual-fluorescence plasmids

The pGB0110m plasmid built by Maistriaux et al. (2024) was used to perform the dual-fluorescence reporter gene assay. It contained three major cassettes: a USER cassette to insert a promoter upstream of the *mVenus* CDS, the *mCherry* CDS under the control of the *p35S* promoter as an internal control, and an MCS where the TF CDS could be inserted. Only the interaction between *pPIP1;3* and *Zm*NAC40 was tested in this assay. For promoter insertion, we reamplified *pPIP1;3* using primers containing the USER restriction sites and then performed a USER reaction (NEB, Ipswich, MA, USA). After promoter insertion, the *ZmNAC40* CDS was added. Two plasmids were built for this experiment, one in which *Zm*NAC40 expression was driven by the *p35S* promoter (test) and one in which no promoter was present (control). These cassettes were amplified from the p2GW7 plasmid constructed for the dual-luciferase experiment. Primers used for the PCR contained HpaI and XbaI restriction sites to allow the cassette insertion by restriction-ligation in the *pPIP1;3*-pGB0110m plasmid. All primers used are listed in Table 6.

Maxiprep

As the reporter gene assays were performed in maize leaf protoplasts, we had to prepare a large amount of plasmid DNA because each transfection required a high quantity of plasmid. A more efficient protocol was developed because the previous protocol based on the Nucleobond® AX kit (Macherey-Nagel, Düren, Germany) was time-consuming and had low throughput.

We adapted the SmartPure kit (Eurogentec, Liège, Belgium) that is normally used for miniprep of 5 ml of bacterial culture to perform maxiprep from 10 ml of bacterial culture. The culture was grown overnight in 10 ml of LB containing the appropriate antibiotic. The following day, cells were harvested by centrifugation at 10,000 x g for 1 min and resuspended in 1 ml of resuspension buffer by pipetting. Then, they were lysed with the addition of 1 ml of lysis buffer. The lysis was interrupted by adding 1.4 ml of neutralisation buffer, and then the tube was 20,238 x g for 10 min at room temperature. The supernatant was transferred to another tube, and 1 volume (v) of isopropanol was added. The supernatant-isopropanol mix was applied on two SmartPure columns per bacterial culture. The loaded columns were centrifuged at 6,000 x g for 30 s, and the flow-through was discarded. This step was repeated until all the supernatant-isopropanol mix passed through the column. Afterwards, 650 µl of wash buffer was added to each column, followed by centrifugation at 12,000 x g for 1 min. This step was repeated twice. Then, the columns were centrifuged at 20,238 x g for 2 min to remove any residual ethanol from the wash buffer. The plasmid DNA was eluted

using 50 µl of water, and elutions from the same initial culture were pooled together. This protocol yielded the same amount of plasmid DNA as the Macherey-Nagel kit, but more samples could be processed simultaneously and in a shorter period (~1h).

Protoplast preparation and transfection

B104 seeds (accession no. PI 594047; Hallauer et al. (1997)) were used for the dual-luciferase reporter gene assay, and B73 seeds (accession no. PI 677128) was used for the dual fluorescence assay.

Seeds were sterilised with a 20% (v/v) NaClO solution for 5 min. After five washes with ddH_2O, they were kept between two water-saturated Tork® hand towels inside a square Petri dish (Greiner Bio-One, Vilvoorde, Belgium) wrapped in aluminium foil. They were kept in a 30°C dark room for germination for two days before being transferred into Jiffypots® (6 cm diameter) filled with 80% (v/v) potting soil (DCM, Grobbendonk, Belgium) and 20% (v/v) Agra-vermiculite (Pull Rhenen, Rhenen, The Netherlands). The plants were grown until the second leaf was developed in phytotrons with a 16/8 h light/dark cycle (25/22°C) and a daytime light intensity of 200 $mmol.m^{-2}.s^{-1}$. They were then transferred to a dark box located in the phytotron for ~7-10 days until the third leaf was fully developed. This process took approximately two weeks.

The third leaf was cut and the abaxial epidermis was removed with glass paper (P1200). The midrib was removed, and the leaves were cut into 5 mm sections. They were transferred to 2 ml of digestion buffer (650 mM sorbitol, 10 mM MES, 10 mM KCl, 1 mM $CaCl_2$, pH 5.5) with the abaxial side facing the buffer in a six-well plate (Greiner Bio-One, Vilvoorde, Belgium). The digestion buffer contained 0.6% (w/v) cellulase RS (Duchefa, Haarlem, The Netherlands), 0.5% (w/v) BSA (Sigma-Aldrich, Saint-Louis, MO, US), 0.5% (w/v) PVP K30 (Fluka, Charlotte, NC, US), and 0.01% (w/v) Pectolyase Y-23 (Duchefa, Haarlem, The Netherlands). The digestion mixture was incubated in the dark at 25°C on a shaker at 60 rpm. After 3 h, the leaf fragments were separated into small filaments with forceps, the solution was separated into two wells, and 1 ml of digestion buffer without enzyme was added to each well. The protoplasts were released after another 15 min of incubation on a shaker at 120 rpm. Then, the non-digested leaf parts were removed, and the liquid containing the protoplasts was filtered through a 100 µm nylon mesh (Fisherbrand, Waltham, MA, US). The protoplast solution was then transferred to a 2 ml Eppendorf tube and centrifuged at 100 x g for 3 min. The supernatant was removed, and the protoplasts were washed twice with the digestion buffer without enzymes. Finally, they were resuspended in 1 ml of digestion buffer. The concentration of protoplasts was determined with a Fuchs-Rosenthal cell-counting plate (Thermo Fisher Scientific, Waltham, MA, USA) where 10 µl of the solution were spread and the number was counted under a

fluorescence microscope (Axio observer 7, Zeiss, Oberkochen, Germany). By counting the number of protoplasts in five large squares of the cell-counting plate, the concentration in one µl could be determined.

A transfection buffer was prepared by mixing 1.5 ml of MaMES buffer (600 mM mannitol, 5 mM MES, pH 5.7) with 12.5 µl of 1 M CaCl2. In 2 ml Eppendorf tubes, between 70,000 – 100,000 protoplasts were added, and 100 µl of the transfection buffer was added without mixing. In the case of the dual-fluorescence reporter gene assay, 10 µg of plasmids was added. In the case of the dual-luciferase reporter gene assay, 10 µg of the reporter, effector, and control plasmids were added. After the addition of DNA, the protoplast solution was mixed and incubated at room temperature until 110 µl of PEG4000 (PEG4000 40% (m/v) with 300 mM MES and 100 mM $Ca(NO_3)_2.4H_2O$) was added. The mixture was incubated at room temperature for 2 min. Then, 440 µl of W5 buffer (154 mM NaCl, 166 mM $CaCl_2$, 5 mM KCl, 2 mM MES, pH 5.7) were mixed into the solution. The solution was centrifuged for 1 min at 100 x g. The supernatant was removed and 1 ml of culture buffer (600 mM mannitol, 4 mM MES, 40 mM KCl, pH 5.7) containing 100 µg/ml cefotaxime was added. Transfected protoplasts were incubated for 24 h in the dark at 20°C.

Dual-luciferase analysis

The dual-luciferase assay was performed using the Dual-Luciferase® Reporter Assay System kit (Promega, Madison, WI, USA). After overnight incubation, protoplasts were pelleted by centrifugation at 100 x g for 3 min. Lysis was performed by adding 50 µl of passive lysis buffer (PLB) and placing the tubes for 15 min on an agitator at 120 rpm at 25°C. The firefly and *Renilla* luciferases signals measurements were performed with a GloMax® 20/20 luminometer (Promega, Madison, WI, USA) using the DLR-O-Inj preloaded protocol. In a clean Eppendorf tube, 100 µl of firefly substrate (LARII reagent) were added and mixed with 20 µl of lysed protoplasts Firefly luciferase activity was measured immediately after the addition of lysed protoplasts. The reaction was then quenched by adding 100 µl of 10X Stop & Glo buffer while the *Renilla* reaction was initiated. The solution was quickly vortexed to avoid bubbles remaining on the sides of the tube, and the *Renilla* intensity was measured. Activities were measured in relative light units (RLU), and the ratio was calculated by dividing the firefly signal by the *Renilla* signal. For each interaction, two transformations were performed using the same batch of protoplasts. This experiment was repeated using three batches of protoplasts prepared on different days.

The statistical analysis was performed using a MLM implemented with the R package lmerTest v3.1-3. This model was designed to consider the plasmid mix used for transfection as a fixed effect and the day of analysis as a random effect.

Significance was assessed with the p-value of the plasmid mix variable; if it was lower than 0.05, it was considered significant.

Dual-fluorescence analysis

Two transfections were performed for each plasmid, using the same batch of protoplasts. For each transfection, a total of 60 µl of protoplasts was observed by confocal microscopy at 20x magnification (Leica Stellaris 8 Falcon) after overnight incubation in a 10 well plate (Greiner Bio-One, Kremsmünster, Austria). The observation range was determined using control plasmids. The negative control did not contain any promoter upstream of the mVenus CDS, and the positive control contained a *p35S* upstream of the mVenus CDS. The mCherry fluorophore was excited at 587 nm, and its emission signal was collected within the range of 597–651 nm. The mVenus fluorophore was excited at 515 nm, and its emission intensity was measured in the range of 523–581 nm. Laser intensities of 0.5% and 5% were used for mCherry and mVenus excitation, respectively. In the test wells, transfected protoplasts were selected based on mCherry fluorescence and 8-15 images were taken per well. Images were saved in LIF format. The mVenus and mCherry intensities were measured using FIJI software (v2.15.0 (Schindelin et al., 2012)) according to the protocol described in the Appendix. This protocol allowed the removal of both background noise and saturated pixels, and the measurement of the same area in both mCherry and mVenus channels. The intensities were used to calculate the mVenus/mCherry ratio, and log2 transformation was applied before statistical analysis. This experiment was repeated three times using protoplasts prepared on different days. The procedure was performed blindly.

The statistical analysis was performed using a MLM implemented with the R package lmerTest v3.1-3. This model was designed to consider the plasmid mix used for transfection as a fixed effect and the day of analysis as a random effect. Significance was assessed with the p-value of the plasmid mix variable. If it was lower than 0.05, it was considered significant.

DAP-sequencing experiment

The DAP-seq experiments were performed as described by Bartlett et al. (2017), with the following modifications.

Library preparation

Parts of the 10th leaf at the 14-leaf developmental stage were harvested from B73 plants (~ 100 mg). The gDNA was extracted using the Wizard® genomic DNA purification kit (Promega, Madison, WI, USA) and resuspended in the elution buffer (EB) of the MinElute Reaction Cleanup kit (Qiagen, Hilden, Germany). Then, 10 µg of extracted gDNA was fragmented to a size range of 100–200 bp in

an Elmasonic S10H ultrasonic bath (Elma, Singen, Germany) for 20 min. The fragment size was verified on a 1% agarose gel. The fragmented DNA was further end-repaired using a Fast DNA End Repair Kit (Thermo Fisher Scientific, Waltham, MA, USA), and TruSeq adapters were ligated with T4 DNA ligase (Promega, Madison, WI, USA). Between sonication, end-repair, and adapter ligation, gDNA was precipitated and cleaned as described by Bartlett et al. (2017). The gDNA library concentration was measured using a Qubit™ dsDNA Quantification broad-range kit (Invitrogen, Waltham, MA, USA). Afterwards, control PCR was performed with primers from the NEBNext® Multiplex Oligos kit (NEB, Ipswich, MA, USA) to confirm that the adapter ligation worked correctly. Three libraries were prepared following this protocol, each containing enough gDNA to perform four DAP reactions.

DAP-sequencing

Genetic constructs containing a cassette to express the Halo-tagged TFs were obtained using the LR Gateway reaction (Invitrogen, Waltham, MA, USA) between the previously described L1-L2 entry plasmids and the pIX-Halo plasmid (ABRC, stock no. CD3-1742). Halo-tagged GFP was used as a negative control. The Halo-tagged proteins (3 TFs and 1 GFP) were subsequently expressed in the wheat (*Triticum aestivum*) germ *in vitro* transcription/translation SP6 promoter system (Promega, Madison, WI, USA). The protein expression and binding to the Magne® HaloTag® Beads (Promega, Madison, WI, USA) was validated by western blot with the 1:1,000 Anti-HaloTag® Polyclonal Rabbit Antibody (G9281 - Promega, Madison, WI, USA). The secondary antibody used to reveal the western blots was the 1:20,000 Anti-Rabbit antibody (A3687 – Sigma-Aldrich, St-Louis, MI, USA). All buffers and procedures for protein-DNA binding were as published by Bartlett et al. (2017), except for the input gDNA library amount. Approximately 1,000 ng of the gDNA library was added as an input to mix with each Halo-tagged protein. The bound fragments were eluted from the beads by adding 30 µl of MinElute EB (Qiagen, Hilden, Germany) and heating at 98°C for 10 min. This was followed by a PCR of 22 cycles to add sequencing indices. The PCR mix contained: 1 µl of Phusion DNA polymerase (NEB, Ipswich, MA, USA), 2.5 µl of NEBNext® Multiplex primers, 1.5 µl of DMSO, 2.5 µl of 10 mM dNTPs (Promega, Madison, WI, USA), 10 µl of 5x Phusion HF Buffer and 7.5 µl of Ambion™ Nuclease-Free Water (Invitrogen, Waltham, MA, USA). Afterwards, the PCR product was cleaned with the MinElute Reaction Cleanup kit (Qiagen, Hilden, Germany) and the elution was performed in a volume of 30 µl. Finally, the concentration of eluted DNA was measured using a Qubit HS dsDNA assay kit, and final concentrations of approximately 2–30 $ng.\mu l^{-1}$ were obtained. The libraries were sent to Macrogen (Amsterdam, The Netherlands) for quality control by TapeStation (Agilent Technologies, Santa Clara, CA, USA) and sequencing by the Illumina NovaSeq X (Illumina, San Diego, CA, USA) platform with paired-end 150-bp reads.

Read mapping, filtering and peak calling

Sequencing reads were aligned using version 4 of the B73 maize genome (Jiao et al., 2017). Both the reference FASTA file and the gff3 annotation file were retrieved from maizegdb (https://www.maizegdb.org [Accessed on 23-07-2024]). Alignment was performed using the nf-core/chipseq v2.0.0 analysis pipeline with nextflow v23.10.0. Read quality was assessed with FASTQC v0.11.9 (Andrews, 2010). Adapters and low-quality reads were trimmed with trimgalore v0.6.7 (Krueger et al., 2021). The alignment was performed using bowtie2 v2.4.4 (--aligner bowtie2; (Langmead & Salzberg, 2012)) only on chromosomal DNA sequences. Contigs, plastid, and mitochondrial sequences were manually removed from the FASTA and gff3 files. Its quality was assessed with Picard v2.27.4 ('Picard Toolkit', 2019) and duplicated reads were discarded (--keep_multi_map false). MACS2 v2.2.7.1 (Y. Zhang et al., 2008) was used for the peak calling step (--narrow_peak true). Peaks identified in at least three replicates were retained (--min_reps_consensus 3). DNA sequences of the peaks located in the promoters, TSS, TTS, exons, and introns were retrieved using the getfasta function from bedtools v2.26.0 (Quinlan & Hall, 2010). Sequences were then used as input in MEME v5.5.5 to predict the enriched motifs (Bailey & Elkan, 1994). The most frequent motif was then used in a TomTom search (S. Gupta et al., 2007) to find similar motifs in the O'Malley et al. (2016) published motif library. The most significant motifs were used in FIMO to determine their frequency in the sequence data (Grant et al., 2011). Only peaks with the top motifs were retained for the rest of the analysis.

General Packages used in this project

Some packages were regularly used in all subprojects of the thesis. Figures were created using the ggplot2 v3.4.3 package combined with Inkscape. Data management was performed using tidyverse v2.0.0. During the writing process of this thesis, Paperpal and ChatGPT were used occasionally for assistance with grammar and sentence clarity.

References

Agurla, S., Gahir, S., Munemasa, S., Murata, Y., & Raghavendra, A. S. (2018). Mechanism of Stomatal Closure in Plants Exposed to Drought and Cold Stress. In M. Iwaya-Inoue, M. Sakurai, & M. Uemura (Eds.), *Survival Strategies in Extreme Cold and Desiccation* (Vol. 1081, pp. 215–232). Springer Singapore. https://doi.org/10.1007/978-981-13-1244-1_12

Aharon, R., Shahak, Y., Wininger, S., Bendov, R., Kapulnik, Y., & Galili, G. (2003). Overexpression of a Plasma Membrane Aquaporin in Transgenic Tobacco Improves Plant Vigor under Favorable Growth Conditions but Not under Drought or Salt Stress. *The Plant Cell*, *15*(2), 439–447. https://doi.org/10.1105/tpc.009225

Aida, M., Ishida, T., Fukaki, H., Fujisawa, H., & Tasaka, M. (1997). Genes involved in organ separation in Arabidopsis: An analysis of the cup-shaped cotyledon mutant. *The Plant Cell*, *9*(6), 841–857. https://doi.org/10.1105/tpc.9.6.841

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Alexandersson, E., Danielson, J. Å. H., Råde, J., Moparthi, V. K., Fontes, M., Kjellbom, P., & Johanson, U. (2010). Transcriptional regulation of aquaporins in accessions of Arabidopsis in response to drought stress. *The Plant Journal*, *61*(4), 650–660. https://doi.org/10.1111/j.1365-313X.2009.04087.x

Alexandersson, E., Fraysse, L., Sjövall-Larsen, S., Gustavsson, S., Fellert, M., Karlsson, M., Johanson, U., & Kjellbom, P. (2005). Whole Gene Family Expression and Drought Stress Regulation of Aquaporins. *Plant Molecular Biology*, *59*(3), 469–484. https://doi.org/10.1007/s11103-005-0352-1

Al-Husini, N., Medler, S., & Ansari, A. (2020). Crosstalk of promoter and terminator during RNA polymerase II transcription cycle. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1863*(12), 194657. https://doi.org/10.1016/j.bbagrm.2020.194657

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), R106. https://doi.org/10.1186/gb-2010-11-10-r106

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data* [Computer software]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arda, H. E., & Walhout, A. J. M. (2010). Gene-centered regulatory networks. *Briefings in Functional Genomics*, *9*(1), 4–12. https://doi.org/10.1093/bfgp/elp049

Ariani, A., & Gepts, P. (2015). Genome-wide identification and characterization of aquaporin gene family in common bean (Phaseolus vulgaris L.).

Molecular Genetics and Genomics, *290*(5), 1771–1785. https://doi.org/10.1007/s00438-015-1038-2

Ariel, F. D., Manavella, P. A., Dezar, C. A., & Chan, R. L. (2007). The true story of the HD-Zip family. *Trends in Plant Science*, *12*(9), 419–426. https://doi.org/10.1016/j.tplants.2007.08.003

Aroca, R., Ferrante, A., Vernieri, P., & Chrispeels, M. J. (2006). Drought, Abscisic Acid and Transpiration Rate Effects on the Regulation of PIP Aquaporin Gene Expression and Abundance in Phaseolus vulgaris Plants. *Annals of Botany*, *98*(6), 1301–1310. https://doi.org/10.1093/aob/mcl219

Aslam, M., Maqbool, M. A., & Cengiz, R. (2015). *Drought Stress in Maize (Zea mays L.)*. Springer International Publishing. https://doi.org/10.1007/978-3-319-25442-5

Astle, W., & Balding, D. J. (2009). Population Structure and Cryptic Relatedness in Genetic Association Studies. *Statistical Science*, *24*(4). https://doi.org/10.1214/09-STS307

Azad, A. K., Hanawa, R., Ishikawa, T., Sawa, Y., & Shibata, H. (2013). Expression profiles of aquaporin homologues and petal movement during petal development in *Tulipa gesneriana*. *Physiologia Plantarum*, *148*(3), 397–407. https://doi.org/10.1111/j.1399-3054.2012.01717.x

Bailey, T. L., & Elkan, C. (1994). Fitting a Mixture Model By Expectation Maximization To Discover Motifs In Biopolymer. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, *2*, 28–36.

Bari, A., Farooq, M., Hussain, A., Tahir Ul Qamar, M., Abbas, M. W., Mustafa, G., Karim, A., Ahmed, I., & Hussain, T. (2018). Genome-Wide Bioinformatics Analysis of Aquaporin Gene Family in Maize (Zea mays L.). *Journal of Phylogenetics & Evolutionary Biology*, *06*(02). https://doi.org/10.4172/2329-9002.1000197

Bartlett, A., O'Malley, R. C., Huang, S. C., Galli, M., Nery, J. R., Gallavotti, A., & Ecker, J. R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nature Protocols*, *12*(8), 1659–1672. https://doi.org/10.1038/nprot.2017.055

Bárzana, G., Aroca, R., Bienert, G. P., Chaumont, F., & Ruiz-Lozano, J. M. (2014). New Insights into the Regulation of Aquaporins by the Arbuscular Mycorrhizal Symbiosis in Maize Plants Under Drought Stress and Possible Implications for Plant Performance. *Molecular Plant-Microbe Interactions®*, *27*(4), 349–363. https://doi.org/10.1094/MPMI-09-13-0268-R

Bechtold, U., Albihlal, W. S., Lawson, T., Fryer, M. J., Sparrow, P. A. C., Richard, F., Persad, R., Bowden, L., Hickman, R., Martin, C., Beynon, J. L., Buchanan-Wollaston, V., Baker, N. R., Morison, J. I. L., Schöffl, F., Ott, S., & Mullineaux, P. M. (2013). Arabidopsis HEAT SHOCK TRANSCRIPTION FACTORA1b overexpression enhances water productivity, resistance to drought, and infection. *Journal of Experimental Botany*, *64*(11), 3467–3481. https://doi.org/10.1093/jxb/ert185

Berger, M. F., & Bulyk, M. L. (2009). Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature Protocols*, *4*(3), 393–411. https://doi.org/10.1038/nprot.2008.195

Bhat, J. A., Yu, D., Bohra, A., Ganie, S. A., & Varshney, R. K. (2021). Features and applications of haplotypes in crop breeding. *Communications Biology*, *4*(1), 1266. https://doi.org/10.1038/s42003-021-02782-y

Bienert, G. P., Heinen, R. B., Berny, M. C., & Chaumont, F. (2014). Maize plasma membrane aquaporin ZmPIP2;5, but not ZmPIP1;2, facilitates transmembrane diffusion of hydrogen peroxide. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1838*(1), 216–222. https://doi.org/10.1016/j.bbamem.2013.08.011

Bienert, G. P., Møller, A. L. B., Kristiansen, K. A., Schulz, A., Møller, I. M., Schjoerring, J. K., & Jahn, T. P. (2007). Specific Aquaporins Facilitate the Diffusion of Hydrogen Peroxide across Membranes. *Journal of Biological Chemistry*, *282*(2), 1183–1192. https://doi.org/10.1074/jbc.M603761200

Bolduc, N., Yilmaz, A., Mejia-Guerra, M. K., Morohashi, K., O'Connor, D., Grotewold, E., & Hake, S. (2012). Unraveling the KNOTTED1 regulatory network in maize meristems. *Genes & Development*, *26*(15), 1685–1690. https://doi.org/10.1101/gad.193433.112

Boursiac, Y., Boudet, J., Postaire, O., Luu, D., Tournaire-Roux, C., & Maurel, C. (2008). Stimulus-induced downregulation of root water transport involves reactive oxygen species-activated cell signalling and plasma membrane intrinsic protein internalization. *The Plant Journal*, *56*(2), 207–218. https://doi.org/10.1111/j.1365-313X.2008.03594.x

Boursiac, Y., Chen, S., Luu, D.-T., Sorieul, M., Van Den Dries, N., & Maurel, C. (2005). Early Effects of Salinity on Water Transport in Arabidopsis Roots. Molecular and Cellular Features of Aquaporin Expression. *Plant Physiology*, *139*(2), 790–805. https://doi.org/10.1104/pp.105.065029

Boyer, J. S. (1982). Plant Productivity and Environment. *Science*, *218*(4571), 443–448. https://doi.org/10.1126/science.218.4571.443

Burdo, B., Gray, J., Goetting-Minesky, M. P., Wittler, B., Hunt, M., Li, T., Velliquette, D., Thomas, J., Gentzel, I., Brito, M. D. S., Mejía-Guerra, M. K., Connolly, L. N., Qaisi, D., Li, W., Casas, M. I., Doseff, A. I., & Grotewold, E. (2014). The Maize TF ome – development of a transcription factor open reading frame collection for functional genomics. *The Plant Journal*, *80*(2), 356–366. https://doi.org/10.1111/tpj.12623

Burgess, A. J., Masclaux-Daubresse, C., Strittmatter, G., Weber, A. P. M., Taylor, S. H., Harbinson, J., Yin, X., Long, S., Paul, M. J., Westhoff, P., Loreto, F., Ceriotti, A., Saltenis, V. L. R., Pribil, M., Nacry, P., Scharff, L. B., Jensen, P. E., Muller, B., Cohan, J., … Baekelandt, A. (2023). Improving crop yield potential: Underlying biological processes and future prospects. *Food and Energy Security*, *12*(1), e435. https://doi.org/10.1002/fes3.435

Cabrera-Bosquet, L., Fournier, C., Brichet, N., Welcker, C., Suard, B., & Tardieu, F. (2016). High-throughput estimation of incident light, light interception

and radiation-use efficiency of thousands of plants in a phenotyping platform. *New Phytologist*, *212*(1), 269–281. https://doi.org/10.1111/nph.14027

Caldeira, C. F., Jeanguenin, L., Chaumont, F., & Tardieu, F. (2014). Circadian rhythms of hydraulic conductance and growth are enhanced by drought and improve plant performance. *Nature Communications*, *5*(1), 5365. https://doi.org/10.1038/ncomms6365

Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P. W., Trisos, C., Romero, J., Aldunce, P., Barrett, K., Blanco, G., Cheung, W. W. L., Connors, S., Denton, F., Diongue-Niang, A., Dodman, D., Garschagen, M., Geden, O., Hayward, B., Jones, C., … Péan, C. (2023). *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland*. (First). Intergovernmental Panel on Climate Change (IPCC). https://doi.org/10.59327/IPCC/AR6-9789291691647

Cantalapiedra, C. P., García-Pereira, M. J., Gracia, M. P., Igartua, E., Casas, A. M., & Contreras-Moreira, B. (2017). Large Differences in Gene Expression Responses to Drought and Heat Stress between Elite Barley Cultivar Scarlett and a Spanish Landrace. *Frontiers in Plant Science*, *8*, 647. https://doi.org/10.3389/fpls.2017.00647

Cao, L., Lu, X., Zhang, P., Wang, G., Wei, L., & Wang, T. (2019). Systematic Analysis of Differentially Expressed Maize ZmbZIP Genes between Drought and Rewatering Transcriptome Reveals bZIP Family Members Involved in Abiotic Stress Responses. *International Journal of Molecular Sciences*, *20*(17), 4103. https://doi.org/10.3390/ijms20174103

Castelletti, S., Coupel-Ledru, A., Granato, I., Palaffre, C., Cabrera-Bosquet, L., Tonelli, C., Nicolas, S. D., Tardieu, F., Welcker, C., & Conti, L. (2020). Maize adaptation across temperate climates was obtained via expression of two florigen genes. *PLOS Genetics*, *16*(7), e1008882. https://doi.org/10.1371/journal.pgen.1008882

Chateigner, A., Lesage-Descauses, M.-C., Rogier, O., Jorge, V., Leplé, J.-C., Brunaud, V., Roux, C. P.-L., Soubigou-Taconnat, L., Martin-Magniette, M.-L., Sanchez, L., & Segura, V. (2020). Gene expression predictions and networks in natural populations supports the omnigenic theory. *BMC Genomics*, *21*(1), 416. https://doi.org/10.1186/s12864-020-06809-2

Chaumont, F., Barrieu, F., Wojcik, E., Chrispeels, M. J., & Jung, R. (2001). Aquaporins Constitute a Large and Highly Divergent Protein Family in Maize. *Plant Physiology*, *125*(3), 1206–1215. https://doi.org/10.1104/pp.125.3.1206

Chaumont, F., & Tyerman, S. D. (2014). Aquaporins: Highly Regulated Channels Controlling Plant Water Relations. *Plant Physiology*, *164*(4), 1600–1618. https://doi.org/10.1104/pp.113.233791

Chen, M., Ji, M., Wen, B., Liu, L., Li, S., Chen, X., Gao, D., & Li, L. (2016). GOLDEN 2-LIKE Transcription Factors of Plants. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.01509

Chen, S., Dang, D., Liu, Y., Ji, S., Zheng, H., Zhao, C., Dong, X., Li, C., Guan, Y., Zhang, A., & Ruan, Y. (2023). Genome-wide association study presents insights into the genetic architecture of drought tolerance in maize seedlings under field water-deficit conditions. *Frontiers in Plant Science*, *14*, 1165582. https://doi.org/10.3389/fpls.2023.1165582

Chen, W., Yin, X., Wang, L., Tian, J., Yang, R., Liu, D., Yu, Z., Ma, N., & Gao, J. (2013). Involvement of rose aquaporin RhPIP1;1 in ethylene-regulated petal expansion through interaction with RhPIP2;1. *Plant Molecular Biology*, *83*(3), 219–233. https://doi.org/10.1007/s11103-013-0084-6

Cheng, P., Kong, F., Han, Y., Liu, X., & Xia, J. (2024). Transcriptome Analysis Reveals Genes and Pathways Associated with Drought Tolerance of Early Stages in Sweet Potato (Ipomoea batatas (L.) Lam.). *Genes*, *15*(7), 948. https://doi.org/10.3390/genes15070948

Christensen, A. H., & Quail, P. H. (1996). Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Research*, *5*(3), 213–218. https://doi.org/10.1007/BF01969712

Ciftci-Yilmaz, S., & Mittler, R. (2008). The zinc finger network of plants. *Cellular and Molecular Life Sciences*, *65*(7–8), 1150–1160. https://doi.org/10.1007/s00018-007-7473-4

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w ^{1118} ; iso-2; iso-3. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M., & Zaret, K. S. (2002). Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. *Molecular Cell*, *9*(2), 279–289. https://doi.org/10.1016/S1097-2765(02)00459-8

Cui, X.-H., Hao, F.-S., Chen, H., Chen, J., & Wang, X.-C. (2008). Expression of the Vicia faba VfPIP1 gene in Arabidopsis thaliana plants improves their drought resistance. *Journal of Plant Research*, *121*(2), 207–214. https://doi.org/10.1007/s10265-007-0130-z

Czarnecka-verner, E., Pan, S., Salem, T., & Gurley, W. B. (2004). Plant class B HSFs inhibit transcription and exhibit affinity for TFIIB and TBP. *Plant Molecular Biology*, *56*(1), 57–75. https://doi.org/10.1007/s11103-004-2307-3

Danielson, J. Å., & Johanson, U. (2008). Unexpected complexity of the Aquaporin gene family in the moss Physcomitrella patens. *BMC Plant Biology*, *8*(1), 45. https://doi.org/10.1186/1471-2229-8-45

Datta, S., Hettiarachchi, C., Johansson, H., & Holm, M. (2007). SALT TOLERANCE HOMOLOG2, a B-Box Protein in *Arabidopsis* That Activates Transcription

and Positively Regulates Light-Mediated Development. *The Plant Cell*, *19*(10), 3242–3255. https://doi.org/10.1105/tpc.107.054791

Datta, S., Johansson, H., Hettiarachchi, C., Irigoyen, M. L., Desai, M., Rubio, V., & Holm, M. (2008). LZF1/SALT TOLERANCE HOMOLOG3, an *Arabidopsis* B-Box Protein Involved in Light-Dependent Development and Gene Expression, Undergoes COP1-Mediated Ubiquitination. *The Plant Cell*, *20*(9), 2324–2338. https://doi.org/10.1105/tpc.108.061747

Dell'Acqua, M., Gatti, D. M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., Hlaing, A. L., Aung, H. H., Nelissen, H., Baute, J., Frascaroli, E., Churchill, G. A., Inzé, D., Morgante, M., & Pè, M. E. (2015). Genetic properties of the MAGIC maize population: A new platform for high definition QTL mapping in Zea mays. *Genome Biology*, *16*(1), 167. https://doi.org/10.1186/s13059-015-0716-z

Deschamps-Francoeur, G., Simoneau, J., & Scott, M. S. (2020). Handling multi-mapped reads in RNA-seq. *Computational and Structural Biotechnology Journal*, *18*, 1569–1576. https://doi.org/10.1016/j.csbj.2020.06.014

Devi, M. J., Sinclair, T. R., Jain, M., & Gallo, M. (2016). Leaf aquaporin transcript abundance in peanut genotypes diverging in expression of the limited-transpiration trait when subjected to differing vapor pressure deficits and aquaporin inhibitors. *Physiologia Plantarum*, *156*(4), 387–396. https://doi.org/10.1111/ppl.12378

Ding, L., Fox, A. R., & Chaumont, F. (2024). Multifaceted role and regulation of aquaporins for efficient stomatal movements. *Plant, Cell & Environment*, *47*(9), 3330–3343. https://doi.org/10.1111/pce.14942

Ding, L., Milhiet, T., Couvreur, V., Nelissen, H., Meziane, A., Parent, B., Aesaert, S., Van Lijsebettens, M., Inzé, D., Tardieu, F., Draye, X., & Chaumont, F. (2020). Modification of the expression of the aquaporin ZmPIP2;5 affects water relations and plant growth. *Plant Physiology*, pp.01183.2019. https://doi.org/10.1104/pp.19.01183

Ding, L., Milhiet, T., Parent, B., Meziane, A., Tardieu, F., & Chaumont, F. (2022). The plasma membrane aquaporin ZmPIP2;5 enhances the sensitivity of stomatal closure to water deficit. *Plant, Cell & Environment*, *45*(4), 1146–1156. https://doi.org/10.1111/pce.14276

Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics*, *3*(1), 43–52. https://doi.org/10.1038/nrg703

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., & Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. *Trends in Plant Science*, *15*(10), 573–581. https://doi.org/10.1016/j.tplants.2010.06.005

Ernst, H. A., Nina Olsen, A., Skriver, K., Larsen, S., & Lo Leggio, L. (2004). Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. *EMBO Reports*, *5*(3), 297–303. https://doi.org/10.1038/sj.embor.7400093

Evans, C., Hardin, J., & Stoebel, D. M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Briefings in Bioinformatics*, *19*(5), 776–792. https://doi.org/10.1093/bib/bbx008

Eveland, A. L., McCarty, D. R., & Koch, K. E. (2008). Transcript Profiling by 3'-Untranslated Region Sequencing Resolves Expression of Gene Families. *Plant Physiology*, *146*(1), 32–44. https://doi.org/10.1104/pp.107.108597

Fagny, M., Kuijjer, M. L., Stam, M., Joets, J., Turc, O., Rozière, J., Pateyron, S., Venon, A., & Vitte, C. (2021). Identification of Key Tissue-Specific, Biological Processes by Integrating Enhancer Information in Maize Gene Regulatory Networks. *Frontiers in Genetics*, *11*, 606285. https://doi.org/10.3389/fgene.2020.606285

Fan, K., Wang, M., Miao, Y., Ni, M., Bibi, N., Yuan, S., Li, F., & Wang, X. (2014). Molecular Evolution and Expansion Analysis of the NAC Transcription Factor in Zea mays. *PLoS ONE*, *9*(11), e111837. https://doi.org/10.1371/journal.pone.0111837

Fang, Y., & Xiong, L. (2015). General mechanisms of drought response and their application in drought resistance improvement in plants. *Cellular and Molecular Life Sciences*, *72*(4), 673–689. https://doi.org/10.1007/s00018-014-1767-0

FAOSTAT. (2022). Food and Agriculture Organization of the United Nations. https://www.fao.org/faostat/en/#home

Fetter, K., Van Wilder, V., Moshelion, M., & Chaumont, F. (2004). Interactions between Plasma Membrane Aquaporins Modulate Their Water Channel Activity. *The Plant Cell*, *16*(1), 215–228. https://doi.org/10.1105/tpc.017194

Figueiredo, D. D., Barros, P. M., Cordeiro, A. M., Serra, T. S., Lourenço, T., Chander, S., Oliveira, M. M., & Saibo, N. J. M. (2012). Seven zinc-finger transcription factors are novel regulators of the stress responsive gene OsDREB1B. *Journal of Experimental Botany*, *63*(10), 3643–3656. https://doi.org/10.1093/jxb/ers035

Fletcher, A. L., Sinclair, T. R., & Allen, L. H. (2007). Transpiration responses to vapor pressure deficit in well watered 'slow-wilting' and commercial soybean. *Environmental and Experimental Botany*, *61*(2), 145–151. https://doi.org/10.1016/j.envexpbot.2007.05.004

Fox, A. R., Maistriaux, L. C., & Chaumont, F. (2017). Toward understanding of the high number of plant aquaporin isoforms and multiple regulation mechanisms. *Plant Science*, *264*, 179–187. https://doi.org/10.1016/j.plantsci.2017.07.021

Franco-Zorrilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., & Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences*, *111*(6), 2367–2372. https://doi.org/10.1073/pnas.1316278111

Fromson, D. (1977). Gene Expression and Its Regulation. In J. G. Wilson & F. C. Fraser (Eds.), *Handbook of Teratology* (pp. 3–17). Springer US. https://doi.org/10.1007/978-1-4615-8933-4_1

Fu, J., Cheng, Y., Linghu, J., Yang, X., Kang, L., Zhang, Z., Zhang, J., He, C., Du, X., Peng, Z., Wang, B., Zhai, L., Dai, C., Xu, J., Wang, W., Li, X., Zheng, J., Chen, L., Luo, L., … Wang, G. (2013). RNA sequencing reveals the complex regulatory network in the maize kernel. *Nature Communications*, *4*(1), 2832. https://doi.org/10.1038/ncomms3832

Galli, M., Khakhar, A., Lu, Z., Chen, Z., Sen, S., Joshi, T., Nemhauser, J. L., Schmitz, R. J., & Gallavotti, A. (2018). The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family. *Nature Communications*, *9*(1), 4526. https://doi.org/10.1038/s41467-018-06977-6

Gangappa, S. N., & Botto, J. F. (2014). The BBX family of plant transcription factors. *Trends in Plant Science*, *19*(7), 460–470. https://doi.org/10.1016/j.tplants.2014.01.010

Gendron, J. M., Pruneda-Paz, J. L., Doherty, C. J., Gross, A. M., Kang, S. E., & Kay, S. A. (2012). *Arabidopsis* circadian clock protein, TOC1, is a DNA-binding transcription factor. *Proceedings of the National Academy of Sciences*, *109*(8), 3167–3172. https://doi.org/10.1073/pnas.1200355109

Gomez-Cano, F., Chu, Y., Cruz-Gomez, M., Abdullah, H. M., Lee, Y. S., Schnell, D. J., & Grotewold, E. (2022). Exploring *Camelina sativa* lipid metabolism regulation by combining gene co-expression and DNA affinity purification analyses. *The Plant Journal*, *110*(2), 589–606. https://doi.org/10.1111/tpj.15682

Gomez-Cano, F., Rodriguez, J., Zhou, P., Chu, Y.-H., Magnusson, E., Gomez-Cano, L., Krishnan, A., Springer, N. M., De Leon, N., & Grotewold, E. (2024). *Prioritizing Maize Metabolic Gene Regulators through Multi-Omic Network Integration*. https://doi.org/10.1101/2024.02.26.582075

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics*, *27*(7), 1017–1018. https://doi.org/10.1093/bioinformatics/btr064

Griffiths, S., Dunford, R. P., Coupland, G., & Laurie, D. A. (2003). The Evolution of *CONSTANS* -Like Gene Families in Barley, Rice, and Arabidopsis. *Plant Physiology*, *131*(4), 1855–1867. https://doi.org/10.1104/pp.102.016188

Grondin, A., Rodrigues, O., Verdoucq, L., Merlot, S., Leonhardt, N., & Maurel, C. (2015). Aquaporins Contribute to ABA-Triggered Stomatal Closure through OST1-Mediated Phosphorylation. *The Plant Cell*, *27*(7), 1945–1954. https://doi.org/10.1105/tpc.15.00421

Groszmann, M., De Rosa, A., Ahmed, J., Chaumont, F., & Evans, J. R. (2021). A consensus on the Aquaporin Gene Family in the Allotetraploid Plant, *Nicotiana tabacum*. *Plant Direct*, *5*(5), e00321. https://doi.org/10.1002/pld3.321

Gu, X., Gao, Z., Yan, Y., Wang, X., Qiao, Y., & Chen, Y. (2017). RdreB1BI enhances drought tolerance by activating AQP-related genes in transgenic

strawberry. *Plant Physiology and Biochemistry*, *119*, 33–42. https://doi.org/10.1016/j.plaphy.2017.08.013

Guo, Y., & Gan, S. (2006). AtNAP, a NAC family transcription factor, has an important role in leaf senescence. *The Plant Journal*, *46*(4), 601–612. https://doi.org/10.1111/j.1365-313X.2006.02723.x

Gupta, A., Rico-Medina, A., & Caño-Delgado, A. I. (2020). The physiology of plant responses to drought. *Science*, *368*(6488), 266–269. https://doi.org/10.1126/science.aaz7614

Gupta, O. P., Deshmukh, R., Kumar, A., Singh, S. K., Sharma, P., Ram, S., & Singh, G. P. (2022). From gene to biomolecular networks: A review of evidences for understanding complex biological function in plants. *Current Opinion in Biotechnology*, *74*, 66–74. https://doi.org/10.1016/j.copbio.2021.10.023

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., & Noble, W. (2007). Quantifying similarity between motifs. *Genome Biology*, *8*(2), R24. https://doi.org/10.1186/gb-2007-8-2-r24

Haberer, G., Kamal, N., Bauer, E., Gundlach, H., Fischer, I., Seidel, M. A., Spannagl, M., Marcon, C., Ruban, A., Urbany, C., Nemri, A., Hochholdinger, F., Ouzunova, M., Houben, A., Schön, C.-C., & Mayer, K. F. X. (2020). European maize genomes highlight intraspecies variation in repeat and gene content. *Nature Genetics*, *52*(9), 950–957. https://doi.org/10.1038/s41588-020-0671-9

Hachez, C., Heinen, R. B., Draye, X., & Chaumont, F. (2008). The expression pattern of plasma membrane aquaporins in maize leaf highlights their role in hydraulic regulation. *Plant Molecular Biology*, *68*(4–5), 337–353. https://doi.org/10.1007/s11103-008-9373-x

Hachez, C., Zelazny, E., & Chaumont, F. (2006). Modulating the expression of aquaporin genes in planta: A key to understand their physiological functions? *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1758*(8), 1142–1156. https://doi.org/10.1016/j.bbamem.2006.02.017

Hallauer, A. R., Lamkey, K. R., & White, P. R. (1997). Registration of Five Inbred Lines of Maize: B102, B103, B104, B105, and B106. *Crop Science*, *37*(4), 1405–1406. https://doi.org/10.2135/cropsci1997.0011183X003700040094x

Hallgren, J., Tsirigos, K. D., Pedersen, M. D., Almagro Armenteros, J. J., Marcatili, P., Nielsen, H., Krogh, A., & Winther, O. (2022). *DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks*. https://doi.org/10.1101/2022.04.08.487609

Hamant, O., & Pautot, V. (2010). Plant development: A TALE story. *Comptes Rendus Biologies*, *333*(4), 371–381. https://doi.org/10.1016/j.crvi.2010.01.015

Haynes, B. C., Maier, E. J., Kramer, M. H., Wang, P. I., Brown, H., & Brent, M. R. (2013). Mapping functional transcription factor networks from gene expression data. *Genome Research*, *23*(8), 1319–1328. https://doi.org/10.1101/gr.150904.112

Hbctraining/DGE_workshop. (2022). https://github.com/hbctraining/DGE_workshop/tree/master

He, Y., Yang, T., Yan, S., Niu, S., & Zhang, Y. (2022). Identification and characterization of the BEL1-like genes reveal their potential roles in plant growth and abiotic stress response in tomato. *International Journal of Biological Macromolecules*, *200*, 193–205. https://doi.org/10.1016/j.ijbiomac.2021.12.175

He, Z., Li, Z., Lu, H., Huo, L., Wang, Z., Wang, Y., & Ji, X. (2019). *The NAC Protein from Tamarix hispida, ThNAC7, Confers Salt and Osmotic Stress Tolerance by Increasing Reactive Oxygen Species Scavenging Capability*. https://doi.org/10.3390/plants8070221

Heinen, R. B., Bienert, G. P., Cohen, D., Chevalier, A. S., Uehlein, N., Hachez, C., Kaldenhoff, R., Le Thiec, D., & Chaumont, F. (2014). Expression and characterization of plasma membrane aquaporins in stomatal complexes of Zea mays. *Plant Molecular Biology*, *86*(3), 335–350. https://doi.org/10.1007/s11103-014-0232-7

Heinen, R. B., Ye, Q., & Chaumont, F. (2009). Role of aquaporins in leaf physiology. *Journal of Experimental Botany*, *60*(11), 2971–2985. https://doi.org/10.1093/jxb/erp171

Hellman, L. M., & Fried, M. G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nature Protocols*, *2*(8), 1849–1861. https://doi.org/10.1038/nprot.2007.249

Hichri, I., Muhovski, Y., Clippe, A., Žižková, E., Dobrev, P. I., Motyka, V., & Lutts, S. (2016). SlDREB2 , a tomato dehydration-responsive element-binding 2 transcription factor, mediates salt stress tolerance in tomato and A rabidopsis. *Plant, Cell & Environment*, *39*(1), 62–79. https://doi.org/10.1111/pce.12591

Huang, Y.-C., Niu, C.-Y., Yang, C.-R., & Jinn, T.-L. (2016). The heat-stress factor HSFA6b connects ABA signaling and ABA-mediated heat responses. *Plant Physiology*, pp.00860.2016. https://doi.org/10.1104/pp.16.00860

Hufford, M. B., Seetharam, A. S., Woodhouse, M. R., Chougule, K. M., Ou, S., Liu, J., Ricci, W. A., Guo, T., Olson, A., Qiu, Y., Della Coletta, R., Tittes, S., Hudson, A. I., Marand, A. P., Wei, S., Lu, Z., Wang, B., Tello-Ruiz, M. K., Piri, R. D., … Dawe, R. K. (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science*, *373*(6555), 655–662. https://doi.org/10.1126/science.abg5289

Islam, Md. A. U., Nupur, J. A., Khalid, M. H. B., Din, A. M. U., Shafiq, M., Alshegaihi, R. M., Ali, Q., Ali, Q., Kamran, Z., Manzoor, M., Haider, M. S., Shahid, M. A., & Manghwar, H. (2022). Genome-Wide Identification and In Silico Analysis of ZF-HD Transcription Factor Genes in Zea mays L. *Genes*, *13*(11), 2112. https://doi.org/10.3390/genes13112112

Jardim-Messeder, D., Cassol, D., Souza-Vieira, Y., Ehlers Loureiro, M., Girke, T., Boroni, M., Lopes Corrêa, R., Coelho, A., & Sachetto-Martins, G. (2023). Genome-wide identification of core components of ABA signaling and transcriptome analysis reveals gene circuits involved in castor bean

(Ricinus communis L.) response to drought. *Gene*, *883*, 147668. https://doi.org/10.1016/j.gene.2023.147668

Jefferson, R. A., Kavanagh, T. A., & Bevan, M. W. (1987). GUS fusions: Beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *The EMBO Journal*, *6*(13), 3901–3907. https://doi.org/10.1002/j.1460-2075.1987.tb02730.x

Jiang, C., Song, X., He, H., Chu, L., Zhou, H., Zhao, Y., Xu, Y., Zeng, W., Lin, X., & Lu, M.-Z. (2020). Genome-wide identification of plasma membrane aquaporin gene family in Populus and functional identification of PIP1;1 involved in osmotic stress. *Environmental and Experimental Botany*, *179*, 104200. https://doi.org/10.1016/j.envexpbot.2020.104200

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., Campbell, M. S., Stein, J. C., Wei, X., Chin, C.-S., Guill, K., Regulski, M., Kumari, S., Olson, A., Gent, J., Schneider, K. L., Wolfgruber, T. K., May, M. R., Springer, N. M., … Ware, D. (2017). Improved maize reference genome with single-molecule technologies. *Nature*, *546*(7659), 524–527. https://doi.org/10.1038/nature22971

Jin, J., He, K., Tang, X., Li, Z., Lv, L., Zhao, Y., Luo, J., & Gao, G. (2015). An *Arabidopsis* Transcriptional Regulatory Map Reveals Distinct Functional and Evolutionary Features of Novel Transcription Factors. *Molecular Biology and Evolution*, *32*(7), 1767–1773. https://doi.org/10.1093/molbev/msv058

Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., & Gao, G. (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, *45*(D1), D1040–D1045. https://doi.org/10.1093/nar/gkw982

Johanson, U., Karlsson, M., Johansson, I., Gustavsson, S., Sjövall, S., Fraysse, L., Weig, A. R., & Kjellbom, P. (2001). The Complete Set of Genes Encoding Major Intrinsic Proteins in Arabidopsis Provides a Framework for a New Nomenclature for Major Intrinsic Proteins in Plants. *Plant Physiology*, *126*(4), 1358–1369. https://doi.org/10.1104/pp.126.4.1358

Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J. M., Yan, J., Sillanpää, M. J., Bonke, M., Palin, K., Talukder, S., Hughes, T. R., Luscombe, N. M., Ukkonen, E., & Taipale, J. (2010). Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Research*, *20*(6), 861–873. https://doi.org/10.1101/gr.100552.109

Jores, T., Tonnies, J., Dorrity, M. W., Cuperus, J. T., Fields, S., & Queitsch, C. (2020). Identification of Plant Enhancers and Their Constituent Elements by STARR-seq in Tobacco Leaves. *The Plant Cell*, *32*(7), 2120–2131. https://doi.org/10.1105/tpc.20.00155

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S., Freimer, N. B., Sabatti, C., & Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, *42*(4), 348–354. https://doi.org/10.1038/ng.548

Kang, Y., Sakiroglu, M., Krom, N., Stanton-Geddes, J., Wang, M., Lee, Y., Young, N. D., & Udvardi, M. (2015). Genome-wide association of drought-related and biomass traits with HapMap SNPs in *Medicago truncatula*. *Plant, Cell & Environment*, *38*(10), 1997–2011. https://doi.org/10.1111/pce.12520

Karimi, M., Inzé, D., & Depicker, A. (2002). GATEWAY™ vectors for Agrobacterium-mediated plant transformation. *Trends in Plant Science*, *7*(5), 193–195. https://doi.org/10.1016/S1360-1385(02)02251-3

Kaya-Okur, H. S., Janssens, D. H., Henikoff, J. G., Ahmad, K., & Henikoff, S. (2020). Efficient low-cost chromatin profiling with CUT&Tag. *Nature Protocols*, *15*(10), 3264–3283. https://doi.org/10.1038/s41596-020-0373-x

Keurentjes, J. J. B., Fu, J., Terpstra, I. R., Garcia, J. M., Van Den Ackerveken, G., Snoek, L. B., Peeters, A. J. M., Vreugdenhil, D., Koornneef, M., & Jansen, R. C. (2007). Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. *Proceedings of the National Academy of Sciences*, *104*(5), 1708–1713. https://doi.org/10.1073/pnas.0610429104

Khanna, R., Kronmiller, B., Maszle, D. R., Coupland, G., Holm, M., Mizuno, T., & Wu, S.-H. (2009). The *Arabidopsis* B-Box Zinc Finger Family. *The Plant Cell*, *21*(11), 3416–3420. https://doi.org/10.1105/tpc.109.069088

Kidder, B. L., Hu, G., & Zhao, K. (2011). ChIP-Seq: Technical considerations for obtaining high-quality data. *Nature Immunology*, *12*(10), 918–922. https://doi.org/10.1038/ni.2117

Klopfenstein, D. V., Zhang, L., Pedersen, B. S., Ramírez, F., Warwick Vesztrocy, A., Naldi, A., Mungall, C. J., Yunes, J. M., Botvinnik, O., Weigel, M., Dampier, W., Dessimoz, C., Flick, P., & Tang, H. (2018). GOATOOLS: A Python library for Gene Ontology analyses. *Scientific Reports*, *8*(1), 10872. https://doi.org/10.1038/s41598-018-28948-z

Kotak, S., Port, M., Ganguli, A., Bicker, F., & Von Koskull-Döring, P. (2004). Characterization of C-terminal domains of *Arabidopsis* heat stress transcription factors (Hsfs) and identification of a new signature combination of plant class A Hsfs with AHA and NES motifs essential for activator function and intracellular localization. *The Plant Journal*, *39*(1), 98–112. https://doi.org/10.1111/j.1365-313X.2004.02111.x

Kremling, K. A. G., Chen, S.-Y., Su, M.-H., Lepak, N. K., Romay, M. C., Swarts, K. L., Lu, F., Lorant, A., Bradbury, P. J., & Buckler, E. S. (2018). Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature*, *555*(7697), 520–523. https://doi.org/10.1038/nature25966

Kremling, K. A. G., Diepenbrock, C. H., Gore, M. A., Buckler, E. S., & Bandillo, N. B. (2019). Transcriptome-Wide Association Supplements Genome-Wide Association in *Zea mays*. *G3 Genes|Genomes|Genetics*, *9*(9), 3023–3033. https://doi.org/10.1534/g3.119.400549

Krueger, F., James, F., Ewels, P., Afyounian, E., Weinstein, M., Schuster-Boeckler, B., Hulselmans, G., & sclamons. (2021). *TrimGalore* (Version 0.6.7) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.7598955

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lawrence, S., Novak, N., Xu, H., & Cooke, J. (2013). Herbivory of maize by southern corn rootworm induces expression of the major intrinsic protein ZmNIP1;1 and leads to the discovery of a novel aquaporin ZmPIP2;8 . *Plant Signaling & Behavior*, *8*(8), e24937. https://doi.org/10.4161/psb.24937

Lee, H. K., Cho, S. K., Son, O., Xu, Z., Hwang, I., & Kim, W. T. (2009). Drought Stress-Induced Rma1H1, a RING Membrane-Anchor E3 Ubiquitin Ligase Homolog, Regulates Aquaporin Levels via Ubiquitination in Transgenic *Arabidopsis* Plants. *The Plant Cell*, *21*(2), 622–641. https://doi.org/10.1105/tpc.108.061994

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, *28*(6), 882–883. https://doi.org/10.1093/bioinformatics/bts034

Levitt, J. (1980). *Responses of plants to environmental stresses* (2nd ed). Academic press.

Li, C., Li, Y., Song, G., Yang, D., Xia, Z., Sun, C., Zhao, Y., Hou, M., Zhang, M., Qi, Z., Wang, B., & Wang, H. (2023). Gene expression and expression quantitative trait loci analyses uncover natural variations underlying the improvement of important agronomic traits during modern maize breeding. *The Plant Journal*, *115*(3), 772–787. https://doi.org/10.1111/tpj.16260

Li, G., Santoni, V., & Maurel, C. (2014). Plant aquaporins: Roles in plant physiology. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1840*(5), 1574–1582. https://doi.org/10.1016/j.bbagen.2013.11.004

Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., Han, Y., Chai, Y., Guo, T., Yang, N., Liu, J., Warburton, M. L., Cheng, Y., Hao, X., Zhang, P., Zhao, J., Liu, Y., Wang, G., Li, J., & Yan, J. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nature Genetics*, *45*(1), 43–50. https://doi.org/10.1038/ng.2484

Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S. L., Kebrom, T. H., Provart, N., Patel, R., Myers, C. R., Reidel, E. J., Turgeon, R., Liu, P., Sun, Q., Nelson, T., & Brutnell, T. P. (2010). The developmental dynamics of the maize leaf transcriptome. *Nature Genetics*, *42*(12), 1060–1067. https://doi.org/10.1038/ng.703

Li, W., Wang, J., Sun, Q., Li, W., Yu, Y., Zhao, M., & Meng, Z. (2017). Expression analysis of genes encoding double B-box zinc finger proteins in maize.

Functional & Integrative Genomics, *17*(6), 653–666. https://doi.org/10.1007/s10142-017-0562-z

Li, Z., & Howell, S. H. (2021). Heat Stress Responses and Thermotolerance in Maize. *International Journal of Molecular Sciences*, *22*(2), 948. https://doi.org/10.3390/ijms22020948

Liang, Z., Myers, Z. A., Petrella, D., Engelhorn, J., Hartwig, T., & Springer, N. M. (2022). Mapping responsive genomic elements to heat stress in a maize diversity panel. *Genome Biology*, *23*(1), 234. https://doi.org/10.1186/s13059-022-02807-7

Liao, X., Guo, X., Wang, Q., Wang, Y., Zhao, D., Yao, L., Wang, S., Liu, G., & Li, T. (2017). Overexpression of *Ms DREB 6.2* results in cytokinin-deficient developmental phenotypes and enhances drought tolerance in transgenic apple plants. *The Plant Journal*, *89*(3), 510–526. https://doi.org/10.1111/tpj.13401

Lin, M., Qiao, P., Matschi, S., Vasquez, M., Ramstein, G. P., Bourgault, R., Mohammadi, M., Scanlon, M. J., Molina, I., Smith, L. G., & Gore, M. A. (2022). Integrating GWAS and TWAS to elucidate the genetic architecture of maize leaf cuticular conductance. *Plant Physiology*, *189*(4), 2144–2158. https://doi.org/10.1093/plphys/kiac198

Lin, Y.-X., Jiang, H.-Y., Chu, Z.-X., Tang, X.-L., Zhu, S.-W., & Cheng, B.-J. (2011). Genome-wide identification, classification and analysis of heat shock transcription factor family in maize. *BMC Genomics*, *12*(1), 76. https://doi.org/10.1186/1471-2164-12-76

Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., & Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. *Nature Methods*, *8*(10), 833–835. https://doi.org/10.1038/nmeth.1681

Listgarten, J., Lippert, C., Kadie, C. M., Davidson, R. I., Eskin, E., & Heckerman, D. (2012). Improved linear mixed models for genome-wide association studies. *Nature Methods*, *9*(6), 525–526. https://doi.org/10.1038/nmeth.2037

Liu, H., Luo, X., Niu, L., Xiao, Y., Chen, L., Liu, J., Wang, X., Jin, M., Li, W., Zhang, Q., & Yan, J. (2017). Distant eQTLs and Non-coding Sequences Play Critical Roles in Regulating Gene Expression and Quantitative Trait Variation in Maize. *Molecular Plant*, *10*(3), 414–426. https://doi.org/10.1016/j.molp.2016.06.016

Liu, M., Wang, C., Ji, Z., Lu, J., Zhang, L., Li, C., Huang, J., Yang, G., Yan, K., Zhang, S., Zheng, C., & Wu, C. (2023). Regulation of drought tolerance in *Arabidopsis* involves the PLATZ4 -mediated transcriptional repression of plasma membrane aquaporin *PIP2;8*. *The Plant Journal*, *115*(2), 434–451. https://doi.org/10.1111/tpj.16235

Liu, S., Li, C., Wang, H., Wang, S., Yang, S., Liu, X., Yan, J., Li, B., Beatty, M., Zastrow-Hayes, G., Song, S., & Qin, F. (2020). Mapping regulatory variants controlling gene expression in drought response and tolerance in maize. *Genome Biology*, *21*(1), 163. https://doi.org/10.1186/s13059-020-02069-1

Liu, S., Zenda, T., Li, J., Wang, Y., Liu, X., & Duan, H. (2020). Comparative transcriptomic analysis of contrasting hybrid cultivars reveal key drought-responsive genes and metabolic pathways regulating drought stress tolerance in maize at various stages. *PLOS ONE*, *15*(10), e0240468. https://doi.org/10.1371/journal.pone.0240468

Liu, Y., Zhou, M., Gao, Z., Ren, W., Yang, F., He, H., & Zhao, J. (2015). RNA-Seq Analysis Reveals MAPKKK Family Members Related to Drought Tolerance in Maize. *PLOS ONE*, *10*(11), e0143128. https://doi.org/10.1371/journal.pone.0143128

Lohman, B. K., Weber, J. N., & Bolnick, D. I. (2016). Evaluation of TagSeq, a reliable low-cost alternative for RNA seq. *Molecular Ecology Resources*, *16*(6), 1315–1321. https://doi.org/10.1111/1755-0998.12529

Lombardi, L., Conforti, P., Valdivia, C., Kovrova, I., & d'Ortigue, O. L. (2023). *Agricultural production statistics 2000–2022*. FAO. https://doi.org/10.4060/cc9205en

Long, H. K., Prescott, S. L., & Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell*, *167*(5), 1170–1187. https://doi.org/10.1016/j.cell.2016.09.018

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lowry, D. B., Logan, T. L., Santuari, L., Hardtke, C. S., Richards, J. H., DeRose-Wilson, L. J., McKay, J. K., Sen, S., & Juenger, T. E. (2013). Expression Quantitative Trait Locus Mapping across Water Availability Environments Reveals Contrasting Associations with Genomic Features in *Arabidopsis*. *The Plant Cell*, *25*(9), 3266–3279. https://doi.org/10.1105/tpc.113.115352

Lu, P.-L., Chen, N.-Z., An, R., Su, Z., Qi, B.-S., Ren, F., Chen, J., & Wang, X.-C. (2006). A novel drought-inducible gene, ATAF1, encodes a NAC family protein that negatively regulates the expression of stress-responsive genes in Arabidopsis. *Plant Molecular Biology*, *63*(2), 289–305. https://doi.org/10.1007/s11103-006-9089-8

Lu, Z., Marand, A. P., Ricci, W. A., Ethridge, C. L., Zhang, X., & Schmitz, R. J. (2019). The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nature Plants*, *5*(12), 1250–1259. https://doi.org/10.1038/s41477-019-0548-z

Lun, A. T. L., Chen, Y., & Smyth, G. K. (2016). It's DE-licious: A Recipe for Differential Expression Analyses of RNA-seq Experiments Using Quasi-Likelihood Methods in edgeR. In E. Mathé & S. Davis (Eds.), *Statistical Genomics* (Vol. 1418, pp. 391–416). Springer New York. https://doi.org/10.1007/978-1-4939-3578-9_19

Luu, P.-L., Ong, P.-T., Dinh, T.-P., & Clark, S. J. (2020). Benchmark study comparing liftover tools for genome conversion of epigenome sequencing data. *NAR Genomics and Bioinformatics*, *2*(3), lqaa054. https://doi.org/10.1093/nargab/lqaa054

Ma, F., Fuqua, B. K., Hasin, Y., Yukhtman, C., Vulpe, C. D., Lusis, A. J., & Pellegrini, M. (2019). A comparison between whole transcript and 3' RNA sequencing methods using Kapa and Lexogen library preparation methods. *BMC Genomics*, *20*(1), 9. https://doi.org/10.1186/s12864-018-5393-3

Ma, N., Xue, J., Li, Y., Liu, X., Dai, F., Jia, W., Luo, Y., & Gao, J. (2008). *Rh-PIP2;1* , a Rose Aquaporin Gene, Is Involved in Ethylene-Regulated Petal Expansion. *Plant Physiology*, *148*(2), 894–907. https://doi.org/10.1104/pp.108.120154

Mabire, C., Duarte, J., Darracq, A., Pirani, A., Rimbert, H., Madur, D., Combes, V., Vitte, C., Praud, S., Rivière, N., Joets, J., Pichon, J.-P., & Nicolas, S. D. (2019). High throughput genotyping of structural variations in a complex plant genome using an original Affymetrix® axiom® array. *BMC Genomics*, *20*(1), 848. https://doi.org/10.1186/s12864-019-6136-9

Madrid-Espinoza, J., Brunel-Saldias, N., Guerra, F. P., Gutiérrez, A., & Del Pozo, A. (2018). Genome-Wide Identification and Transcriptional Regulation of Aquaporin Genes in Bread Wheat (Triticum aestivum L.) under Water Stress. *Genes*, *9*(10), 497. https://doi.org/10.3390/genes9100497

Magee, L. (1990). R 2 Measures Based on Wald and Likelihood Ratio Joint Significance Tests. *The American Statistician*, *44*(3), 250. https://doi.org/10.2307/2685352

Maistriaux, L. C., Laurent, M. J., Jeanguenin, L., Prado, S. A., Nader, J., Welcker, C., Charcosset, A., Tardieu, F., Nicolas, S. D., & Chaumont, F. (2024). Genetic variability of aquaporin expression in maize: From eQTLs to a MITE insertion regulating *PIP2;5* expression. *Plant Physiology*, kiae326. https://doi.org/10.1093/plphys/kiae326

Manosalva Pérez, N., Ferrari, C., Engelhorn, J., Depuydt, T., Nelissen, H., Hartwig, T., & Vandepoele, K. (2024). MINI-AC: Inference of plant gene regulatory networks using bulk or single-cell accessible chromatin profiles. *The Plant Journal*, *117*(1), 280–301. https://doi.org/10.1111/tpj.16483

Mao, H., Wang, H., Liu, S., Li, Z., Yang, X., Yan, J., Li, J., Tran, L.-S. P., & Qin, F. (2015). A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. *Nature Communications*, *6*(1), 8326. https://doi.org/10.1038/ncomms9326

Marand, A. P., Eveland, A. L., Kaufmann, K., & Springer, N. M. (2023). *Cis* -Regulatory Elements in Plant Development, Adaptation, and Evolution. *Annual Review of Plant Biology*, *74*(1), 111–137. https://doi.org/10.1146/annurev-arplant-070122-030236

Matzke, M. A., Kanno, T., & Matzke, A. J. M. (2015). RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. *Annual Review of Plant Biology*, *66*(1), 243–267. https://doi.org/10.1146/annurev-arplant-043014-114633

Maurel, C., Verdoucq, L., & Rodrigues, O. (2016). Aquaporins and plant transpiration: Aquaporins and plant transpiration. *Plant, Cell & Environment*, *39*(11), 2580–2587. https://doi.org/10.1111/pce.12814

McMullen, M. D., Kresovich, S., Villeda, H. S., Bradbury, P., Li, H., Sun, Q., Flint-Garcia, S., Thornsberry, J., Acharya, C., Bottoms, C., Brown, P., Browne, C., Eller, M., Guill, K., Harjes, C., Kroon, D., Lepak, N., Mitchell, S. E., Peterson, B., … Buckler, E. S. (2009). Genetic Properties of the Maize Nested Association Mapping Population. *Science*, *325*(5941), 737–740. https://doi.org/10.1126/science.1174320

McNabb, D. S., Reed, R., & Marciniak, R. A. (2005). Dual Luciferase Assay System for Rapid Assessment of Gene Expression in *Saccharomyces cerevisiae*. *Eukaryotic Cell*, *4*(9), 1539–1549. https://doi.org/10.1128/EC.4.9.1539-1549.2005

Mejia-Guerra, M. K., Li, W., Galeano, N. F., Vidal, M., Gray, J., Doseff, A. I., & Grotewold, E. (2015). Core Promoter Plasticity Between Maize Tissues and Genotypes Contrasts with Predominance of Sharp Transcription Initiation Sites. *The Plant Cell*, *27*(12), 3309–3320. https://doi.org/10.1105/tpc.15.00630

Mejia-Guerra, M. K., Pomeranz, M., Morohashi, K., & Grotewold, E. (2012). From plant gene regulatory grids to network dynamics. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1819*(5), 454–465. https://doi.org/10.1016/j.bbagrm.2012.02.016

Miculan, M., Nelissen, H., Ben Hassen, M., Marroni, F., Inzé, D., Pè, M. E., & Dell'Acqua, M. (2021). A forward genetics approach integrating genome-wide association study and expression quantitative trait locus mapping to dissect leaf development in maize (*Zea mays*). *The Plant Journal*, *107*(4), 1056–1071. https://doi.org/10.1111/tpj.15364

Millet, E., Welcker, C., Kruijer, W., Negro, S., Nicolas, S., Praud, S., Ranc, N., Presterl, T., Tuberosa, R., Bedo, Z., Draye, X., Usadel, B., Charcosset, A., Van Eeuwijk, F., Tardieu, F., Coupel-Ledru, A., & Bauland, C. (2016). Genome-wide analysis of yield in Europe: Allelic effects as functions of drought and heat scenarios. *Plant Physiology*, pp.00621.2016. https://doi.org/10.1104/pp.16.00621

Mitsuoka, K., Murata, K., Walz, T., Hirai, T., Agre, P., Heymann, J. B., Engel, A., & Fujiyoshi, Y. (1999). The Structure of Aquaporin-1 at 4.5-Å Resolution Reveals Short α-Helices in the Center of the Monomer. *Journal of Structural Biology*, *128*(1), 34–43. https://doi.org/10.1006/jsbi.1999.4177

Molina, C., & Grotewold, E. (2005). Genome wide analysis of Arabidopsis core promoters. *BMC Genomics*, *6*(1), 25. https://doi.org/10.1186/1471-2164-6-25

Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J. B., Engel, A., & Fujiyoshi, Y. (2000). Structural determinants of water permeation through aquaporin-1. *Nature*, *407*(6804), 599–605. https://doi.org/10.1038/35036519

Muraya, M. M., Schmutzer, T., Ulpinnis, C., Scholz, U., & Altmann, T. (2015). Targeted Sequencing Reveals Large-Scale Sequence Polymorphism in Maize Candidate Genes for Biomass Production and Composition.

PLOS ONE, *10*(7), e0132120. https://doi.org/10.1371/journal.pone.0132120

Myers, Z. A., Wootan, C. M., Liang, Z., Zhou, P., Engelhorn, J., Hartwig, T., & Nathan, S. M. (2023). Conserved and variable heat stress responses of the Heat Shock Factor transcription factor family in maize and *SETARIA VIRIDIS* . *Plant Direct*, *7*(4), e489. https://doi.org/10.1002/pld3.489

Nam, J. (2003). Antiquity and Evolution of the MADS-Box Gene Family Controlling Flower Development in Plants. *Molecular Biology and Evolution*, *20*(9), 1435–1447. https://doi.org/10.1093/molbev/msg152

Negro, S. S., Millet, E. J., Madur, D., Bauland, C., Combes, V., Welcker, C., Tardieu, F., Charcosset, A., & Nicolas, S. D. (2019). Genotyping-by-sequencing and SNP-arrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. *BMC Plant Biology*, *19*(1), 318. https://doi.org/10.1186/s12870-019-1926-4

Nelissen, H., Sun, X., Rymen, B., Jikumaru, Y., Kojima, M., Takebayashi, Y., Abbeloos, R., Demuynck, K., Storme, V., Vuylsteke, M., De Block, J., Herman, D., Coppens, F., Maere, S., Kamiya, Y., Sakakibara, H., Beemster, G. T. S., & Inzé, D. (2018). The reduction in maize leaf growth under mild drought affects the transition between cell division and cell expansion and cannot be restored by elevated gibberellic acid levels. *Plant Biotechnology Journal*, *16*(2), 615–627. https://doi.org/10.1111/pbi.12801

Nelson, H. C. M., & Littlefield, O. (1999). A new use for the 'wing' of the 'winged' helix-turn-helix motif in the HSF–DNA cocrystal. *Nature Structural Biology*, *6*(5), 464–470. https://doi.org/10.1038/8269

Nover, L., Bharti, K., Döring, P., Mishra, S. K., Ganguli, A., & Scharf, K.-D. (2001). Arabidopsis and the heat stress transcription factor world: How many heat stress transcription factors do we need? *Cell Stress & Chaperones*, *6*(3), 177. https://doi.org/10.1379/1466-1268(2001)006<0177:AATHST>2.0.CO;2

Nunes, T. D. G., Zhang, D., & Raissig, M. T. (2020). Form, development and function of grass stomata. *The Plant Journal*, *101*(4), 780–799. https://doi.org/10.1111/tpj.14552

Ogawa, D., Yamaguchi, K., & Nishiuchi, T. (2007). High-level overexpression of the Arabidopsis HsfA2 gene confers not only increased themotolerance but also salt/osmotic stress tolerance and enhanced callus growth. *Journal of Experimental Botany*, *58*(12), 3373–3383. https://doi.org/10.1093/jxb/erm184

Oh, J. E., Kwon, Y., Kim, J. H., Noh, H., Hong, S.-W., & Lee, H. (2011). A dual role for MYB60 in stomatal regulation and root growth of Arabidopsis thaliana under drought stress. *Plant Molecular Biology*, *77*(1–2), 91–103. https://doi.org/10.1007/s11103-011-9796-7

Olsen, A. N., Ernst, H. A., Leggio, L. L., & Skriver, K. (2005). NAC transcription factors: Structurally distinct, functionally diverse. *Trends in Plant Science*, *10*(2), 79–87. https://doi.org/10.1016/j.tplants.2004.12.010

Olsson, A., Engstr□m, P., & S□derman, E. (2004). The homeobox genes ATHB12 and ATHB7encode potential regulators of growth in response to water deficit in Arabidopsis. *Plant Molecular Biology*, *55*(5), 663–677. https://doi.org/10.1007/s11103-004-1581-4

O'Malley, R. C., Huang, S. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., Galli, M., Gallavotti, A., & Ecker, J. R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell*, *165*(5), 1280–1292. https://doi.org/10.1016/j.cell.2016.04.038

Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., Matsubara, K., Osato, N., Kawai, J., Carninci, P., Hayashizaki, Y., Suzuki, K., Kojima, K., Takahara, Y., Yamamoto, K., & Kikuchi, S. (2003). Comprehensive Analysis of NAC Family Genes in Oryza sativa and Arabidopsis thaliana. *DNA Research*, *10*(6), 239–247. https://doi.org/10.1093/dnares/10.6.239

Opitz, N., Marcon, C., Paschold, A., Malik, W. A., Lithio, A., Brandt, R., Piepho, H.-P., Nettleton, D., & Hochholdinger, F. (2016). Extensive tissue-specific transcriptomic plasticity in maize primary roots upon water deficit. *Journal of Experimental Botany*, *67*(4), 1095–1107. https://doi.org/10.1093/jxb/erv453

Oshlack, A., & Wakefield, M. J. (2009). Transcript length bias in RNA-seq data confounds systems biology. *Biology Direct*, *4*(1), 14. https://doi.org/10.1186/1745-6150-4-14

Pajoro, A., Madrigal, P., Muiño, J. M., Matus, J., Jin, J., Mecchia, M. A., Debernardi, J. M., Palatnik, J. F., Balazadeh, S., Arif, M., Ó'Maoiléidigh, D. S., Wellmer, F., Krajewski, P., Riechmann, J.-L., Angenent, G. C., & Kaufmann, K. (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biology*, *15*(3), R41. https://doi.org/10.1186/gb-2014-15-3-r41

Pan, B., Kusko, R., Xiao, W., Zheng, Y., Liu, Z., Xiao, C., Sakkiah, S., Guo, W., Gong, P., Zhang, C., Ge, W., Shi, L., Tong, W., & Hong, H. (2019). *Additional file 2: Of Similarities and differences between variants called with human reference genome HG19 or HG38* (p. 648520 Bytes) [Dataset]. figshare. https://doi.org/10.6084/M9.FIGSHARE.7849988

Pang, J., Fu, J., Zong, N., Wang, J., Song, D., Zhang, X., He, C., Fang, T., Zhang, H., Fan, Y., Wang, G., & Zhao, J. (2019). Kernel size-related genes revealed by an integrated eQTL analysis during early maize kernel development. *The Plant Journal*, *98*(1), 19–32. https://doi.org/10.1111/tpj.14193

Pařenicová, L., De Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., Cook, H. E., Ingram, R. M., Kater, M. M., Davies, B., Angenent, G. C., & Colombo, L. (2003). Molecular and Phylogenetic Analyses of the Complete MADS-Box Transcription Factor Family in Arabidopsis. *The Plant Cell*, *15*(7), 1538–1551. https://doi.org/10.1105/tpc.011544

Pei, H., Ma, N., Tian, J., Luo, J., Chen, J., Li, J., Zheng, Y., Chen, X., Fei, Z., & Gao, J. (2013). An NAC Transcription Factor Controls Ethylene-Regulated Cell Expansion in Flower Petals. *PLANT PHYSIOLOGY*, *163*(2), 775–791. https://doi.org/10.1104/pp.113.223388

Picard Toolkit. (2019). In *Broad Institute, GitHub repository* [Computer software]. Broad Institute. https://broadinstitute.github.io/picard/

Pireyre, M., & Burow, M. (2015). Regulation of MYB and bHLH Transcription Factors: A Glance at the Protein Level. *Molecular Plant*, *8*(3), 378–388. https://doi.org/10.1016/j.molp.2014.11.022

Pörtner, H., Roberts, D. C., Adams, H., Adelekan, I., Adler, C., Adrian, R., Aldunce, P., Ali, E., Begum, R. A., Bednar-Friedl, B., Kerr, R. B., Biesbroek, R., Birkmann, J., Bowen, K., Angela Caretta, M., Carnicer, J., Castellanos, E., & Sung Cheong, T. (2023). *Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009325844

Prado, S. A., Cabrera-Bosquet, L., Grau, A., Coupel-Ledru, A., Millet, E. J., Welcker, C., & Tardieu, F. (2018). Phenomics allows identification of genomic regions affecting maize stomatal conductance with conditional effects of water deficit and evaporative demand. *Plant, Cell & Environment*, *41*(2), 314–326. https://doi.org/10.1111/pce.13083

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904–909. https://doi.org/10.1038/ng1847

Pritchard, J. K., Stephens, M., Rosenberg, N. A., & Donnelly, P. (2000). Association Mapping in Structured Populations. *The American Journal of Human Genetics*, *67*(1), 170–181. https://doi.org/10.1086/302959

Putpeerawit, P., Sojikul, P., Thitamadee, S., & Narangajavana, J. (2017). Genome-wide analysis of aquaporin gene family and their responses to water-deficit stress conditions in cassava. *Plant Physiology and Biochemistry*, *121*, 118–127. https://doi.org/10.1016/j.plaphy.2017.10.025

Qi, Q., Gibson, A., Fu, X., Zheng, M., Kuehn, R., Wang, Y., Wang, Y., Navarro, S., Morrell, J. A., Jiang, D., Simmons, G., Bell, E., Ivleva, N. B., McClerren, A. L., Loida, P., Ruff, T. G., Petracek, M. E., & Preuss, S. B. (2012). Involvement of the N-terminal B-box Domain of Arabidopsis BBX32 Protein in Interaction with Soybean BBX62 Protein. *Journal of Biological Chemistry*, *287*(37), 31482–31493. https://doi.org/10.1074/jbc.M112.346940

Qian, B., Wang, Q., Zhang, C., Guo, J., Yu, Z., Han, J., Xia, H., Zhao, R., & Yin, Y. (2024). Exploring the Roles of TALE Gene Family in Maize Drought Stress Responses. *Agronomy*, *14*(6), 1267. https://doi.org/10.3390/agronomy14061267

Qiu, X., Wang, G., Abou-Elwafa, S. F., Fu, J., Liu, Z., Zhang, P., Xie, X., Ku, L., Ma, Y., Guan, X., & Wei, L. (2022). Genome-wide identification of HD-ZIP transcription factors in maize and their regulatory roles in promoting drought tolerance. *Physiology and Molecular Biology of Plants*, *28*(2), 425–437. https://doi.org/10.1007/s12298-022-01147-x

Quero, G., Simondi, S., Ceretta, S., Otero, Á., Garaycochea, S., Fernández, S., Borsani, O., & Bonnecarrère, V. (2021). An integrative analysis of yield stability for a GWAS in a small soybean breeding population. *Crop Science*, *61*(3), 1903–1914. https://doi.org/10.1002/csc2.20490

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Quinn, E. M., Cormican, P., Kenny, E. M., Hill, M., Anney, R., Gill, M., Corvin, A. P., & Morris, D. W. (2013). Development of Strategies for SNP Detection in RNA-Seq Data: Application to Lymphoblastoid Cell Lines and Evaluation Using 1000 Genomes Data. *PLoS ONE*, *8*(3), e58815. https://doi.org/10.1371/journal.pone.0058815

Rae, L., Lao, N. T., & Kavanagh, T. A. (2011). Regulation of multiple aquaporin genes in Arabidopsis by a pair of recently duplicated DREB transcription factors. *Planta*, *234*(3), 429–444. https://doi.org/10.1007/s00425-011-1414-z

Rau, A., Gallopin, M., Celeux, G., & Jaffrézic, F. (2013). Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics*, *29*(17), 2146–2152. https://doi.org/10.1093/bioinformatics/btt350

Ré, D. A., Capella, M., Bonaventure, G., & Chan, R. L. (2014). Arabidopsis AtHB7 and AtHB12evolved divergently to fine tune processes associated with growth and responses to water stress. *BMC Plant Biology*, *14*(1), 150. https://doi.org/10.1186/1471-2229-14-150

Reddy, P. S., Rao, T. S. R. B., Sharma, K. K., & Vadez, V. (2015). Genome-wide identification and characterization of the aquaporin gene family in Sorghum bicolor (L.). *Plant Gene*, *1*, 18–28. https://doi.org/10.1016/j.plgene.2014.12.002

Ricardi, M. M., González, R. M., Zhong, S., Domínguez, P. G., Duffy, T., Turjanski, P. G., Salgado Salter, J. D., Alleva, K., Carrari, F., Giovannoni, J. J., Estévez, J. M., & Iusem, N. D. (2014). Genome-wide data (ChIP-seq) enabled identification of cell wall-related and aquaporin genes as targets of tomato ASR1, a drought stress-responsive transcription factor. *BMC Plant Biology*, *14*(1), 29. https://doi.org/10.1186/1471-2229-14-29

Ricci, W. A., Lu, Z., Ji, L., Marand, A. P., Ethridge, C. L., Murphy, N. G., Noshay, J. M., Galli, M., Mejía-Guerra, M. K., Colomé-Tatché, M., Johannes, F., Rowley, M. J., Corces, V. G., Zhai, J., Scanlon, M. J., Buckler, E. S., Gallavotti, A., Springer, N. M., Schmitz, R. J., & Zhang, X. (2019). Widespread long-range cis-regulatory elements in the maize genome. *Nature Plants*, *5*(12), 1237–1249. https://doi.org/10.1038/s41477-019-0547-0

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C.-Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J. Z., Ghandehari, D., Sherman, B. K., & -L. Yu, G. (2000). *Arabidopsis* Transcription Factors: Genome-Wide Comparative

Analysis Among Eukaryotes. *Science*, *290*(5499), 2105–2110. https://doi.org/10.1126/science.290.5499.2105

Rincent, R., Moreau, L., Monod, H., Kuhn, E., Melchinger, A. E., Malvar, R. A., Moreno-Gonzalez, J., Nicolas, S., Madur, D., Combes, V., Dumas, F., Altmann, T., Brunel, D., Ouzunova, M., Flament, P., Dubreuil, P., Charcosset, A., & Mary-Huard, T. (2014). Recovering Power in Association Mapping Panels with Variable Levels of Linkage Disequilibrium. *Genetics*, *197*(1), 375–387. https://doi.org/10.1534/genetics.113.159731

Rishmawi, L., Bauget, F., Protto, V., Bauland, C., Nacry, P., & Maurel, C. (2023). Natural variation of maize root hydraulic architecture underlies highly diverse water uptake capacities. *Plant Physiology*, *192*(3), 2404–2418. https://doi.org/10.1093/plphys/kiad213

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. https://doi.org/10.1093/nar/gkv007

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, *11*(3), R25. https://doi.org/10.1186/gb-2010-11-3-r25

Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A., & Eilers, P. H. C. (2018). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Statistics*, *23*, 52–71. https://doi.org/10.1016/j.spasta.2017.10.003

Rossini, L., Cribb, L., Martin, D. J., & Langdale, J. A. (2001). The Maize *Golden2* Gene Defines a Novel Class of Transcriptional Regulators in Plants. *The Plant Cell*, *13*(5), 1231–1244. https://doi.org/10.1105/tpc.13.5.1231

Rossum, B.-J. van. (2024). *statgenSTA: Single Trial Analysis (STA) of Field Trials* [Computer software]. https://biometris.github.io/statgenSTA/index.html

Sablowski, R. W. M., & Meyerowitz, E. M. (1998). A Homolog of NO APICAL MERISTEM Is an Immediate Target of the Floral Homeotic Genes APETALA3/PISTILLATA. *Cell*, *92*(1), 93–103. https://doi.org/10.1016/S0092-8674(00)80902-2

Sade, N., Shatil-Cohen, A., Attia, Z., Maurel, C., Boursiac, Y., Kelly, G., Granot, D., Yaaran, A., Lerner, S., & Moshelion, M. (2014). The Role of Plasma Membrane Aquaporins in Regulating the Bundle Sheath-Mesophyll Continuum and Leaf Hydraulics. *Plant Physiology*, *166*(3), 1609–1620. https://doi.org/10.1104/pp.114.248633

Sakuma, Y., Liu, Q., Dubouzet, J. G., Abe, H., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2002). DNA-Binding Specificity of the ERF/AP2 Domain of Arabidopsis DREBs, Transcription Factors Involved in Dehydration- and

Cold-Inducible Gene Expression. *Biochemical and Biophysical Research Communications*, *290*(3), 998–1009. https://doi.org/10.1006/bbrc.2001.6299

Sakurai, J., Ishikawa, F., Yamaguchi, T., Uemura, M., & Maeshima, M. (2005). Identification of 33 Rice Aquaporin Genes and Analysis of Their Expression and Function. *Plant and Cell Physiology*, *46*(9), 1568–1577. https://doi.org/10.1093/pcp/pci172

Saleem, A., Roldán-Ruiz, I., Aper, J., & Muylle, H. (2022). Genetic control of tolerance to drought stress in soybean. *BMC Plant Biology*, *22*(1), 615. https://doi.org/10.1186/s12870-022-03996-w

Samad-Zamini, M., Schweiger, W., Nussbaumer, T., Mayer, K. F. X., & Buerstmayr, H. (2017). Time-course expression QTL-atlas of the global transcriptional response of wheat to *Fusarium graminearum*. *Plant Biotechnology Journal*, *15*(11), 1453–1464. https://doi.org/10.1111/pbi.12729

Sarrion-Perdigones, A., Vazquez-Vilar, M., Palaci, J., Castelijns, B., Forment, J., Ziarsolo, P., Blanca, J., Granell, A., & Orzaez, D. (2013). GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. *PLANT PHYSIOLOGY*, *162*(3), 1618–1631. https://doi.org/10.1104/pp.113.217661

Schaefer, R. J., Michno, J.-M., Jeffers, J., Hoekenga, O., Dilkes, B., Baxter, I., & Myers, C. L. (2018). Integrating Coexpression Networks with GWAS to Prioritize Causal Genes in Maize. *The Plant Cell*, *30*(12), 2922–2942. https://doi.org/10.1105/tpc.18.00299

Scharf, K.-D., Berberich, T., Ebersberger, I., & Nover, L. (2012). The plant heat stress transcription factor (Hsf) family: Structure, function and evolution. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1819*(2), 104–119. https://doi.org/10.1016/j.bbagrm.2011.10.002

Schena, M., & Davis, R. W. (1992). HD-Zip proteins: Members of an Arabidopsis homeodomain protein superfamily. *Proceedings of the National Academy of Sciences*, *89*(9), 3894–3898. https://doi.org/10.1073/pnas.89.9.3894

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. https://doi.org/10.1038/nmeth.2019

Schmitz, R. J., Grotewold, E., & Stam, M. (2022). Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. *The Plant Cell*, *34*(2), 718–741. https://doi.org/10.1093/plcell/koab281

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T. A., Minx, P., Reily, A. D., Courtney, L., Kruchowski, S. S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., … Wilson, R. K. (2009). The B73 Maize Genome:

Complexity, Diversity, and Dynamics. *Science*, *326*(5956), 1112–1115. https://doi.org/10.1126/science.1178534

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., & Nordborg, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics*, *44*(7), 825–830. https://doi.org/10.1038/ng.2314

Shivaraj, S. M., Sharma, Y., Chaudhary, J., Rajora, N., Sharma, S., Thakral, V., Ram, H., Sonah, H., Singla-Pareek, S. L., Sharma, T. R., & Deshmukh, R. (2021). Dynamic role of aquaporin transport system under drought stress in plants. *Environmental and Experimental Botany*, *184*, 104367. https://doi.org/10.1016/j.envexpbot.2020.104367

Siefritz, F., Tyree, M. T., Lovisolo, C., Schubert, A., & Kaldenhoff, R. (2002). PIP1 Plasma Membrane Aquaporins in Tobacco: From Cellular Effects to Function in Plants. *The Plant Cell*, *14*(4), 869–876. https://doi.org/10.1105/tpc.000901

Signor, S. A., & Nuzhdin, S. V. (2018). The Evolution of Gene Expression in cis and trans. *Trends in Genetics*, *34*(7), 532–544. https://doi.org/10.1016/j.tig.2018.03.007

Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*, *6*, e21856. https://doi.org/10.7554/eLife.21856

Song, N., Xu, Z., Wang, J., Qin, Q., Jiang, H., Si, W., & Li, X. (2018). Genome-wide analysis of maize CONSTANS-LIKE gene family and expression profiling under light/dark and abscisic acid treatment. *Gene*, *673*, 1–11. https://doi.org/10.1016/j.gene.2018.06.032

Souer, E., Houwelingen, A. van, Kloos, D., Mol, J., & Koes, R. (1996). 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*, *85*(2), 159–170. https://doi.org/10.1016/S0092-8674(00)81093-4

Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: From enhancer binding to developmental control. *Nature Reviews Genetics*, *13*(9), 613–626. https://doi.org/10.1038/nrg3207

Springer, N., De León, N., & Grotewold, E. (2019). Challenges of Translating Gene Regulatory Information into Agronomic Improvements. *Trends in Plant Science*, *24*(12), 1075–1082. https://doi.org/10.1016/j.tplants.2019.07.004

Steudle, E. (2000). Water uptake by plant roots: An integration of views. In O. Gašparíková, M. Čiamporová, I. Mistrík, & F. Baluška (Eds.), *Recent Advances of Plant Root Structure and Function* (pp. 71–82). Springer Netherlands. https://doi.org/10.1007/978-94-017-2858-4_9

Steudle, E. (2001). The Cohesion-Tension Mechanism and the Acquisition of Water by Plant Roots. *Annual Review of Plant Physiology and Plant Molecular Biology*, *52*(1), 847–875. https://doi.org/10.1146/annurev.arplant.52.1.847

Steudle, E., & Peterson, C. A. (1998). How does water get through roots? *Journal of Experimental Botany*, *49*(322), 775–788. https://doi.org/10.1093/jxb/49.322.775

Studer, A., Zhao, Q., Ross-Ibarra, J., & Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene tb1. *Nature Genetics*, *43*(11), 1160–1163. https://doi.org/10.1038/ng.942

Su, Y., Liu, Z., Sun, J., Wu, C., Li, Y., Zhang, C., & Zhao, L. (2022). Genome-Wide Identification of Maize Aquaporin and Functional Analysis During Seed Germination and Seedling Establishment. *Frontiers in Plant Science*, *13*, 831916. https://doi.org/10.3389/fpls.2022.831916

Sul, J. H., Martin, L. S., & Eskin, E. (2018). Population structure in genetic studies: Confounding factors and mixed models. *PLOS Genetics*, *14*(12), e1007309. https://doi.org/10.1371/journal.pgen.1007309

Sun, G., Yu, H., Wang, P., Lopez-Guerrero, M., Mural, R. V., Mizero, O. N., Grzybowski, M., Song, B., van Dijk, K., Schachtman, D. P., Zhang, C., & Schnable, J. C. (2023). A role for heritable transcriptomic variation in maize adaptation to temperate environments. *Genome Biology*, *24*(1), 55. https://doi.org/10.1186/s13059-023-02891-3

Sun, G., Zhu, C., Kramer, M. H., Yang, S.-S., Song, W., Piepho, H.-P., & Yu, J. (2010). Variation explained in mixed-model association mapping. *Heredity*, *105*(4), 333–340. https://doi.org/10.1038/hdy.2010.11

Sun, H., Wang, S., Lou, Y., Zhu, C., Zhao, H., Li, Y., Li, X., & Gao, Z. (2021). A bamboo leaf-specific aquaporin gene PePIP2;7 is involved in abiotic stress response. *Plant Cell Reports*, *40*(7), 1101–1114. https://doi.org/10.1007/s00299-021-02673-w

Sun, J., Xie, M., Li, X., Li, Z., Wang, Q., Ding, A., Wang, W., & Sun, Y. (2021). Systematic Investigations of the ZF-HD Gene Family in Tobacco Reveal Their Multiple Roles in Abiotic Stresses. *Agronomy*, *11*(3), 406. https://doi.org/10.3390/agronomy11030406

Svitashev, S., Schwartz, C., Lenderts, B., Young, J. K., & Mark Cigan, A. (2016). Genome editing in maize directed by CRISPR–Cas9 ribonucleoprotein complexes. *Nature Communications*, *7*(1), 13274. https://doi.org/10.1038/ncomms13274

Swift, J., & Coruzzi, G. M. (2017). A matter of time—How transient transcription factor interactions create dynamic gene regulatory networks. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1860*(1), 75–83. https://doi.org/10.1016/j.bbagrm.2016.08.007

Taji, T., Seki, M., Satou, M., Sakurai, T., Kobayashi, M., Ishiyama, K., Narusaka, Y., Narusaka, M., Zhu, J.-K., & Shinozaki, K. (2004). Comparative Genomics in Salt Tolerance between Arabidopsis and Arabidopsis-Related Halophyte Salt Cress Using Arabidopsis Microarray. *Plant Physiology*, *135*(3), 1697–1709. https://doi.org/10.1104/pp.104.039909

Tan, Q. K.-G., & Irish, V. F. (2006). The Arabidopsis Zinc Finger-Homeodomain Genes Encode Proteins with Unique Biochemical Properties That Are Coordinately Expressed during Floral Development. *Plant Physiology*, *140*(3), 1095–1108. https://doi.org/10.1104/pp.105.070565

Törnroth-Horsefield, S., Wang, Y., Hedfalk, K., Johanson, U., Karlsson, M., Tajkhorshid, E., Neutze, R., & Kjellbom, P. (2006). Structural mechanism of plant aquaporin gating. *Nature*, *439*(7077), 688–694. https://doi.org/10.1038/nature04316

Tu, X., Mejía-Guerra, M. K., Valdes Franco, J. A., Tzeng, D., Chu, P.-Y., Shen, W., Wei, Y., Dai, X., Li, P., Buckler, E. S., & Zhong, S. (2020). Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. *Nature Communications*, *11*(1), 5089. https://doi.org/10.1038/s41467-020-18832-8

Uehlein, N., & Kaldenhoff, R. (2007). Aquaporins and Plant Leaf Movements. *Annals of Botany*, *101*(1), 1–4. https://doi.org/10.1093/aob/mcm278

Uehlein, N., Lovisolo, C., Siefritz, F., & Kaldenhoff, R. (2003). The tobacco aquaporin NtAQP1 is a membrane CO2 pore with physiological functions. *Nature*, *425*(6959), 734–737. https://doi.org/10.1038/nature02027

Uehlein, N., Otto, B., Hanson, D. T., Fischer, M., McDowell, N., & Kaldenhoff, R. (2008). Function of *Nicotiana tabacum* Aquaporins as Chloroplast Gas Pores Challenges the Concept of Membrane CO_2 Permeability. *The Plant Cell*, *20*(3), 648–657. https://doi.org/10.1105/tpc.107.054023

Van Den Berg, S., Vandenplas, J., Van Eeuwijk, F. A., Lopes, M. S., & Veerkamp, R. F. (2019). Significance testing and genomic inflation factor using high-density genotypes or whole-genome sequence data. *Journal of Animal Breeding and Genetics*, *136*(6), 418–429. https://doi.org/10.1111/jbg.12419

Vandeleur, R. K., Mayo, G., Shelden, M. C., Gilliham, M., Kaiser, B. N., & Tyerman, S. D. (2009). The Role of Plasma Membrane Intrinsic Protein Aquaporins in Water Transport through Roots: Diurnal and Drought Stress Responses Reveal Different Strategies between Isohydric and Anisohydric Cultivars of Grapevine. *Plant Physiology*, *149*(1), 445–460. https://doi.org/10.1104/pp.108.128645

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Walhout, A. J. M. (2006). Unraveling transcription regulatory networks by protein–DNA and protein–protein interaction mapping. *Genome Research*, *16*(12), 1445–1454. https://doi.org/10.1101/gr.5321506

Walhout, A. J. M. (2011). What does biologically meaningful mean? A perspective on gene regulatory network validation. *Genome Biology*, *12*(4), 109. https://doi.org/10.1186/gb-2011-12-4-109

Wang, C., Hu, H., Qin, X., Zeise, B., Xu, D., Rappel, W.-J., Boron, W. F., & Schroeder, J. I. (2016). Reconstitution of CO_2 Regulation of SLAC1 Anion Channel and Function of CO_2 -Permeable PIP2;1 Aquaporin as CARBONIC ANHYDRASE4 Interactor. *The Plant Cell*, *28*(2), 568–582. https://doi.org/10.1105/tpc.15.00637

Wang, J., Yu, H., Weng, X., Xie, W., Xu, C., Li, X., Xiao, J., & Zhang, Q. (2014). An expression quantitative trait loci-guided co-expression analysis for

constructing regulatory network using a rice recombinant inbred line population. *Journal of Experimental Botany*, *65*(4), 1069–1079. https://doi.org/10.1093/jxb/ert464

Wang, L., Liu, Y., Feng, S., Yang, J., Li, D., & Zhang, J. (2017). Roles of Plasmalemma Aquaporin Gene StPIP1 in Enhancing Drought Tolerance in Potato. *Frontiers in Plant Science*, *8*, 616. https://doi.org/10.3389/fpls.2017.00616

Wang, L., Zhang, C., Wang, Y., Wang, Y., Yang, C., Lu, M., & Wang, C. (2018). *Tamarix hispida* aquaporin *ThPIP2;5* confers salt and osmotic stress tolerance to transgenic *Tamarix* and *Arabidopsis*. *Environmental and Experimental Botany*, *152*, 158–166. https://doi.org/10.1016/j.envexpbot.2017.05.018

Wang, N., Liu, W., Yu, L., Guo, Z., Chen, Z., Jiang, S., Xu, H., Fang, H., Wang, Y., Zhang, Z., & Chen, X. (2020). HEAT SHOCK FACTOR A8a Modulates Flavonoid Synthesis and Drought Tolerance. *Plant Physiology*, *184*(3), 1273–1290. https://doi.org/10.1104/pp.20.01106

Wang, R., Zhang, Y., Wang, C., Wang, Y.-C., & Wang, L.-Q. (2021). ThNAC12 from Tamarix hispida directly regulates ThPIP2;5 to enhance salt tolerance by modulating reactive oxygen species. *Plant Physiology and Biochemistry*, *163*, 27–35. https://doi.org/10.1016/j.plaphy.2021.03.042

Wang, X., Basnayake, B. M. V. S., Zhang, H., Li, G., Li, W., Virk, N., Mengiste, T., & Song, F. (2009). The Arabidopsis ATAF1, a NAC Transcription Factor, Is a Negative Regulator of Defense Responses Against Necrotrophic Fungal and Bacterial Pathogens. *Molecular Plant-Microbe Interactions®*, *22*(10), 1227–1238. https://doi.org/10.1094/MPMI-22-10-1227

Wang, X., Li, Y., Ji, W., Bai, X., Cai, H., Zhu, D., Sun, X.-L., Chen, L.-J., & Zhu, Y.-M. (2011). A novel Glycine soja tonoplast intrinsic protein gene responds to abiotic stress and depresses salt and dehydration tolerance in transgenic Arabidopsis thaliana. *Journal of Plant Physiology*, *168*(11), 1241–1248. https://doi.org/10.1016/j.jplph.2011.01.016

Wickham, H. (2023). *stringr: Simple, Consistent Wrappers for Common String Operations* [Computer software]. https://stringr.tidyverse.org

Windhövel, A., Hein, I., Dabrowa, R., & Stockhaus, J. (2001). Characterization of a novel class of plant homeodomain proteins that bind to the C4 phosphoenolpyruvate carboxylase gene of Flaveria trinervia. *Plant Molecular Biology*, *45*(2), 201–214. https://doi.org/10.1023/A:1006450005648

WMO. (2023, March 21). World Meteorological Organization. https://wmo.int/about-us/world-meteorological-day/wmd-2020/drought

Woodhouse, M. R., Cannon, E. K., Portwood, J. L., Harper, L. C., Gardiner, J. M., Schaeffer, M. L., & Andorf, C. M. (2021). A pan-genomic approach to genome databases using maize as a model system. *BMC Plant Biology*, *21*(1), 385. https://doi.org/10.1186/s12870-021-03173-5

Wu, L., Chang, Y., Wang, L., Wang, S., & Wu, J. (2022). The aquaporin gene PvXIP1;2 conferring drought resistance identified by GWAS at seedling stage in common bean. *Theoretical and Applied Genetics*, *135*(2), 485–500. https://doi.org/10.1007/s00122-021-03978-w

Xiao, H., & Lis, J. T. (1988). Germline Transformation Used to Define Key Features of Heat-Shock Response Elements. *Science*, *239*(4844), 1139–1142. https://doi.org/10.1126/science.3125608

Xu, C., Wang, M., Zhou, L., Quan, T., & Xia, G. (2013). Heterologous Expression of the Wheat Aquaporin Gene TaTIP2;2 Compromises the Abiotic Stress Tolerance of Arabidopsis thaliana. *PLoS ONE*, *8*(11), e79618. https://doi.org/10.1371/journal.pone.0079618

Xu, Y., Li, P., Yang, Z., & Xu, C. (2017). Genetic mapping of quantitative trait loci in crops. *The Crop Journal*, *5*(2), 175–184. https://doi.org/10.1016/j.cj.2016.06.003

Xue, J., Yang, F., & Gao, J. (2009). Isolation of Rh-TIP1;1, an aquaporin gene and its expression in rose flowers in response to ethylene and water deficit. *Postharvest Biology and Technology*, *51*(3), 407–413. https://doi.org/10.1016/j.postharvbio.2008.08.011

Yang, F., Li, W., Jiang, N., Yu, H., Morohashi, K., Ouma, W. Z., Morales-Mantilla, D. E., Gomez-Cano, F. A., Mukundi, E., Prada-Salcedo, L. D., Velazquez, R. A., Valentin, J., Mejía-Guerra, M. K., Gray, J., Doseff, A. I., & Grotewold, E. (2017). A Maize Gene Regulatory Network for Phenolic Metabolism. *Molecular Plant*, *10*(3), 498–515. https://doi.org/10.1016/j.molp.2016.10.020

Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., & Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics*, *46*(2), 100–106. https://doi.org/10.1038/ng.2876

Yang, O., Popova, O. V., Süthoff, U., Lüking, I., Dietz, K.-J., & Golldack, D. (2009). The Arabidopsis basic leucine zipper transcription factor AtbZIP24 regulates complex transcriptional networks involved in abiotic stress resistance. *Gene*, *436*(1–2), 45–55. https://doi.org/10.1016/j.gene.2009.02.010

Yang, Q., Yuan, C., Cong, T., Wang, J., & Zhang, Q. (2022). Genome-wide identification of three-amino-acid-loop-extension gene family and their expression profile under hormone and abiotic stress treatments during stem development of Prunus mume. *Frontiers in Plant Science*, *13*, 1006360. https://doi.org/10.3389/fpls.2022.1006360

Yang, Z., Cao, Y., Shi, Y., Qin, F., Jiang, C., & Yang, S. (2023). Genetic and molecular exploration of maize environmental stress resilience: Toward sustainable agriculture. *Molecular Plant*, *16*(10), 1496–1517. https://doi.org/10.1016/j.molp.2023.07.005

Yilmaz, A., Nishiyama, M. Y., Fuentes, B. G., Souza, G. M., Janies, D., Gray, J., & Grotewold, E. (2009). GRASSIUS: A Platform for Comparative Regulatory Genomics across the Grasses. *Plant Physiology*, *149*(1), 171–180. https://doi.org/10.1104/pp.108.128579

Yokotani, N., Ichikawa, T., Kondou, Y., Matsui, M., Hirochika, H., Iwabuchi, M., & Oda, K. (2008). Expression of rice heat stress transcription factor OsHsfA2e enhances tolerance to environmental stresses in transgenic Arabidopsis. *Planta*, *227*(5), 957–967. https://doi.org/10.1007/s00425-007-0670-4

Yu, C.-P., Lin, J.-J., & Li, W.-H. (2016). Positional distribution of transcription factor binding sites in Arabidopsis thaliana. *Scientific Reports*, *6*(1), 25164. https://doi.org/10.1038/srep25164

Yu, J., & Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology*, *17*(2), 155–160. https://doi.org/10.1016/j.copbio.2006.02.003

Zhang, F., Pan, Z., Han, C., Dong, H., Lin, L., Qiao, Q., Zhao, K., Wu, J., Tao, S., Zhang, S., & Huang, X. (2024). *Pyrus betulaefolia* ERF3 interacts with HsfC1a to coordinately regulate aquaporin *PIP1;4* and *NCED4* for drought tolerance. *Horticulture Research*, *11*(5), uhae090. https://doi.org/10.1093/hr/uhae090

Zhang, H., Li, G., Hu, D., Zhang, Y., Zhang, Y., Shao, H., Zhao, L., Yang, R., & Guo, X. (2020). Functional characterization of maize heat shock transcription factor gene ZmHsf01 in thermotolerance. *PeerJ*, *8*, e8926. https://doi.org/10.7717/peerj.8926

Zhang, M., Wang, J., Liu, R., Liu, H., Yang, H., Zhu, Z., Xu, R., Wang, P., Deng, X., Xue, S., Zhu, F., & Cheng, Y. (2022). CsMYB96 confers resistance to water loss in citrus fruit by simultaneous regulation of water transport and wax biosynthesis. *Journal of Experimental Botany*, *73*(3), 953–966. https://doi.org/10.1093/jxb/erab420

Zhang, P.-Y., Qiu, X., Fu, J.-X., Wang, G.-R., Wei, L., & Wang, T.-C. (2021). Systematic analysis of differentially expressed ZmMYB genes related to drought stress in maize. *Physiology and Molecular Biology of Plants*, *27*(6), 1295–1309. https://doi.org/10.1007/s12298-021-01013-2

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M., & Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, *42*(4), 355–360. https://doi.org/10.1038/ng.546

Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., & Wang, L. (2014). CrossMap: A versatile tool for coordinate conversion between genome assemblies. *Bioinformatics*, *30*(7), 1006–1007. https://doi.org/10.1093/bioinformatics/btt730

Zhao, P., Javed, S., Shi, X., Wu, B., Zhang, D., Xu, S., & Wang, X. (2020). Varying Architecture of Heat Shock Elements Contributes to Distinct Magnitudes of Target Gene Expression and Diverged Biological

Pathways in Heat Stress Response of Bread Wheat. *Frontiers in Genetics*, *11*, 30. https://doi.org/10.3389/fgene.2020.00030

Zheng, X., & Gehring, M. (2019). Low-input chromatin profiling in Arabidopsis endosperm using CUT&RUN. *Plant Reproduction*, *32*(1), 63–75. https://doi.org/10.1007/s00497-018-00358-1

Zhou, P., Li, Z., Magnusson, E., Gomez Cano, F., Crisp, P. A., Noshay, J. M., Grotewold, E., Hirsch, C. N., Briggs, S. P., & Springer, N. M. (2020). Meta Gene Regulatory Networks in Maize Highlight Functionally Relevant Regulatory Interactions. *The Plant Cell*, *32*(5), 1377–1396. https://doi.org/10.1105/tpc.20.00080

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, *44*(7), 821–824. https://doi.org/10.1038/ng.2310

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and Prospects of Association Mapping in Plants. *The Plant Genome*, *1*(1), plantgenome2008.02.0089. https://doi.org/10.3835/plantgenome2008.02.0089

Zhu, C., Lin, Z., Yang, K., Lou, Y., Liu, Y., Li, T., Li, H., Di, X., Wang, J., Sun, H., Li, Y., Li, X., & Gao, Z. (2024). A bamboo ' PESAPK4-PEMYB99- *PeTIP4-3* ' regulatory model involved in water transport. *New Phytologist*, *243*(1), 195–212. https://doi.org/10.1111/nph.19787

Zhu, D., Wu, Z., Cao, G., Li, J., Wei, J., Tsuge, T., Gu, H., Aoyama, T., & Qu, L.-J. (2014). TRANSLUCENT GREEN, an ERF Family Transcription Factor, Controls Water Balance in Arabidopsis by Activating the Expression of Aquaporin Genes. *Molecular Plant*, *7*(4), 601–615. https://doi.org/10.1093/mp/sst152

Zhu, J.-K. (2016). Abiotic Stress Signaling and Responses in Plants. *Cell*, *167*(2), 313–324. https://doi.org/10.1016/j.cell.2016.08.029

Zwiazek, J. J., Xu, H., Tan, X., Navarro-Ródenas, A., & Morte, A. (2017). Significance of oxygen transport through aquaporins. *Scientific Reports*, *7*(1), 40411. https://doi.org/10.1038/srep40411

Appendix

Supplemental information

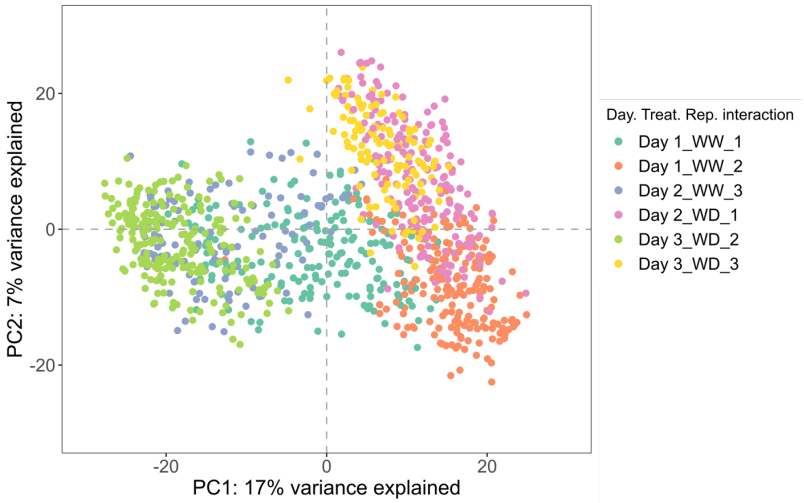
Kinship

We mentioned that the population structure was accounted for using the kinship matrix. Several methods also exist to compute it, but the most common are described in VanRaden et al. (2008) and Astle & Balding (2009). However, these methods for calculating the kinship matrix are expected to lower the power of some markers, owing to a phenomenon called proximal contamination. This refers to the fact that a marker is included twice in the model: once as a fixed effect and once as a random effect in the kinship matrix (Lippert et al., 2011; Listgarten et al., 2012; J. Yang et al., 2014). To overcome this issue, (Listgarten et al., 2012) proposed a method in which kinship is calculated using all markers except a subset comprising the tested marker and markers in its proximity, which are more likely to be in LD. It was further adapted using all markers except those physically linked to the tested markers on the same chromosome. This method is known as the leave-one-chromosome-out (LOCO) method. This method has been proven to be efficient in maize panels (Rincent et al., 2014). However, recovering the power with the LOCO method can sometimes lead to an increase in false positives, as discussed by van den Berg et al. (2019). This can be assessed by comparing the genomic inflation factors obtained using different kinship methods.

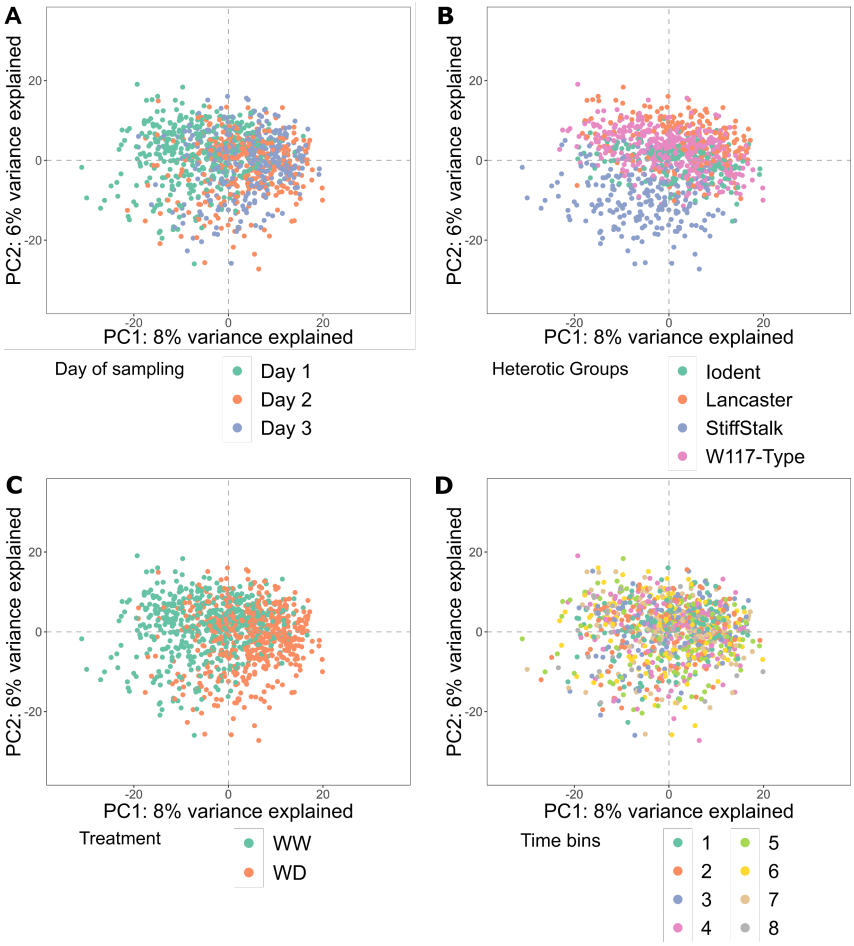
Confocal images intensity measurement

All image analyses were performed using FIJI (Fiji Is Just ImageJ, v2.15.0). Image files were first opened using the File > Open command with the colorized colour mode. Thresholding was then applied by opening the Threshold window (Shift + Cmd/Ctrl + T). The Auto function was used to set the initial minimum threshold and remove the background noise. The maximum threshold value was then manually decreased to 254 to avoid overexposed areas. To define the region of interest (ROI), the Create Selection function was used (Edit > Selection > Create Selection). Signal intensities for both mCherry and mVenus channels were measured by using the shortcut Cmd/Ctrl + M, which generated quantitative data for the selected regions. The integrated density was then used as the total signal output.

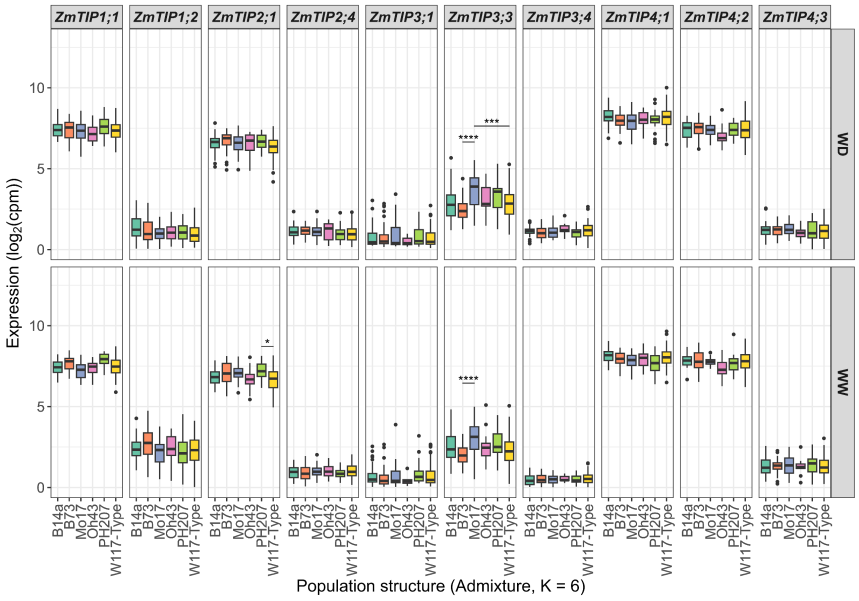
Supplemental Figures



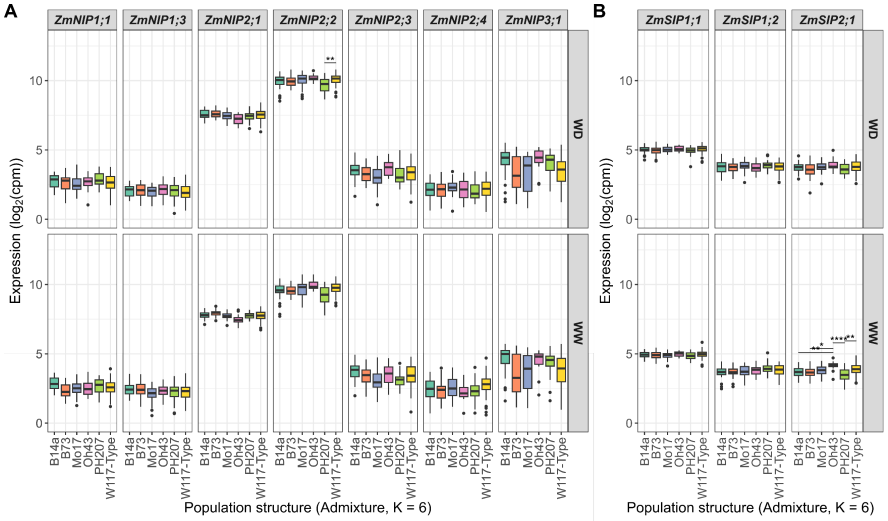
Supplemental Figure 1: Principal component analysis of the normalised and transformed expression data in log_2(cpm). *The PCA analysis was performed using the 500 genes presenting the highest variance in expression between the samples. The results for the first two PCs are shown with the percentage of variance explained by each of them. Each dot represents a sample. The samples are coloured following a combination of the day of sampling, the treatment and the replicate number.*



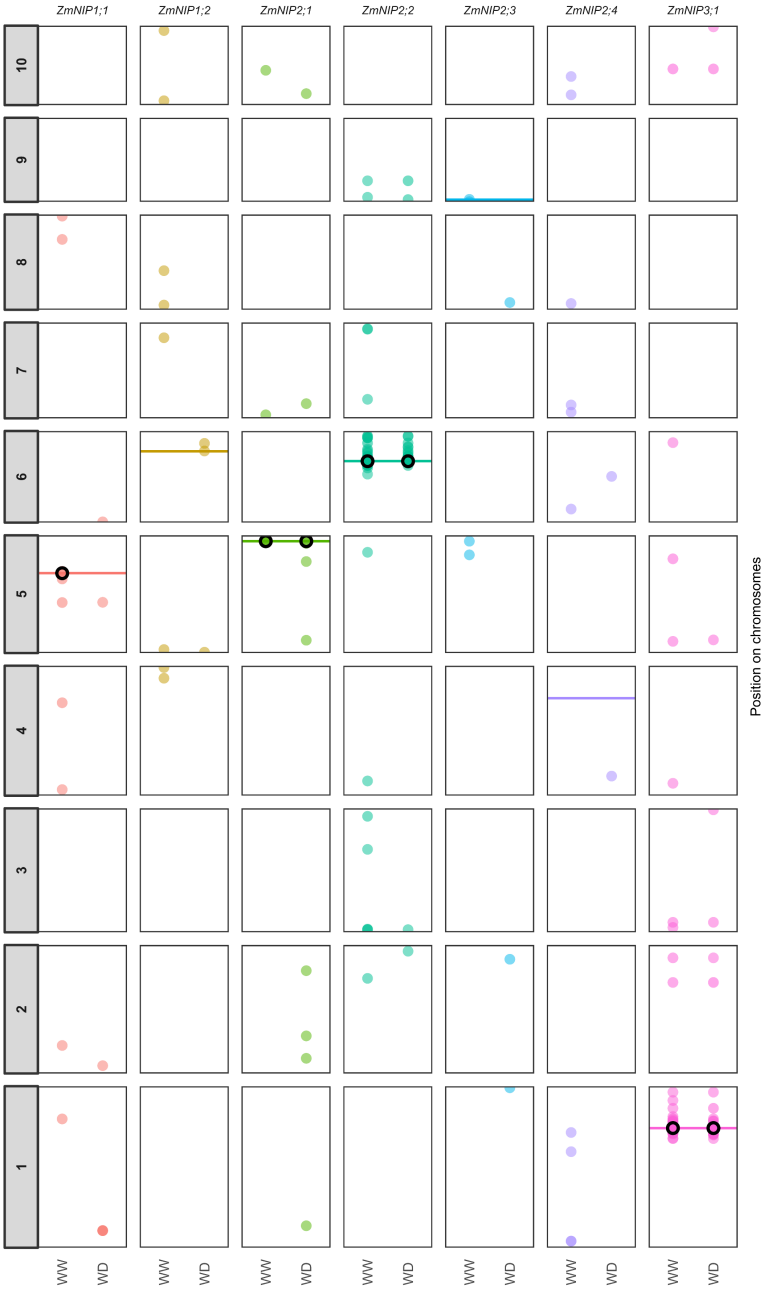
Supplemental Figure 2: Principal component analysis of the normalised and transformed expression data in log_2(cpm). *The PCA analysis was performed using the 500 genes presenting the highest variance in expression between the samples. The results for the first two PCs are shown with the percentage of variance explained by each of them. Each dot represents a sample. The samples are coloured following the day of sampling (A), their heterotic group determined by ADMIXTURE (B), treatment applied to the plants (C), and time bins representing the time samples were collected (D).*



Supplemental Figure 3:* ZmTIP *genes expression in log(cpm) in WW and WD. *The hybrids were grouped according to their heterotic groups. Differential expression was tested between the groups, and significance is shown at the top of the boxplots (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 10^{-4}).*

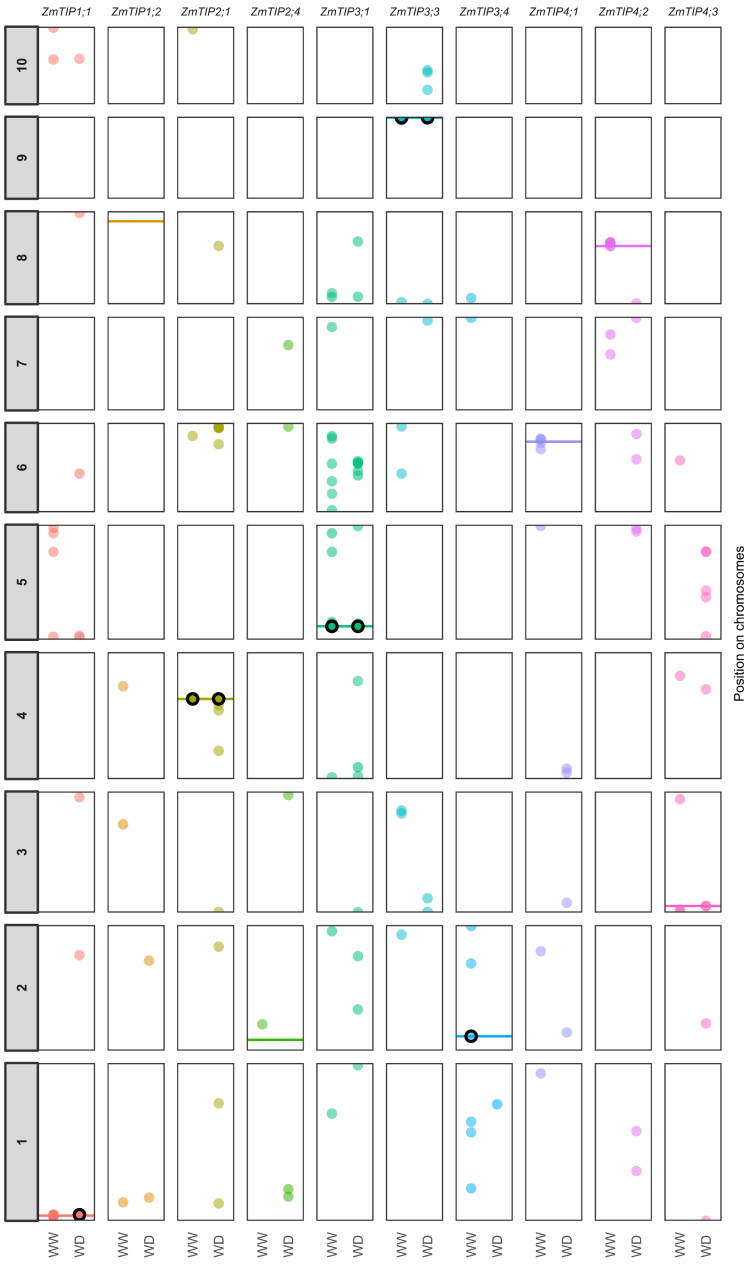


Supplemental Figure 4:* ZmNIP *and ZmSIP genes expression in log(cpm) in WW and WD. *The hybrids were grouped according to their heterotic groups. Differential expression was tested between the groups, and significance is shown at the top of the boxplots (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 10^{-4}).*

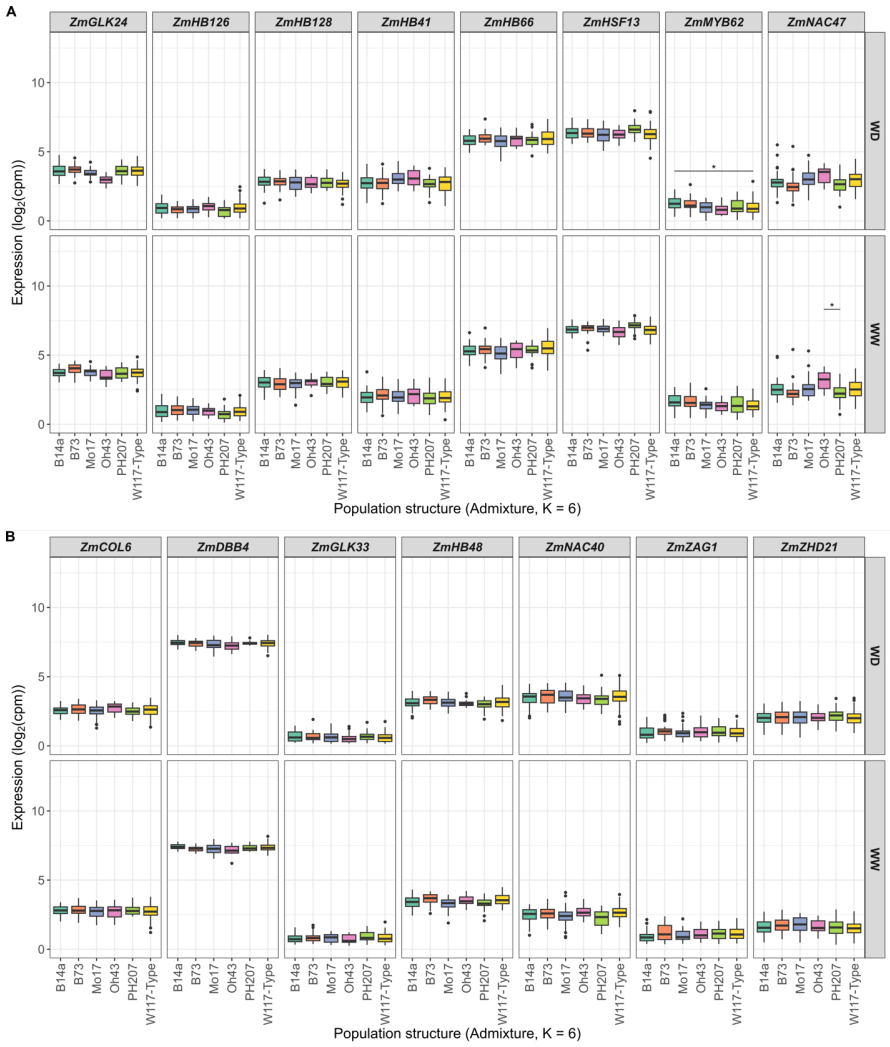


***Supplemental Figure 5:* ZmNIP eQTLs.** *Each dot represents a physical eQTL position (B73 RefGen v4), and the transparency of the dot indicates the significance of the lead marker. Columns represent chromosomes, while rows represent genes. Data is shown for both WW and WD conditions. ZmNIP positions are indicated with vertical-coloured lines, and local eQTLs are cicled in black.*

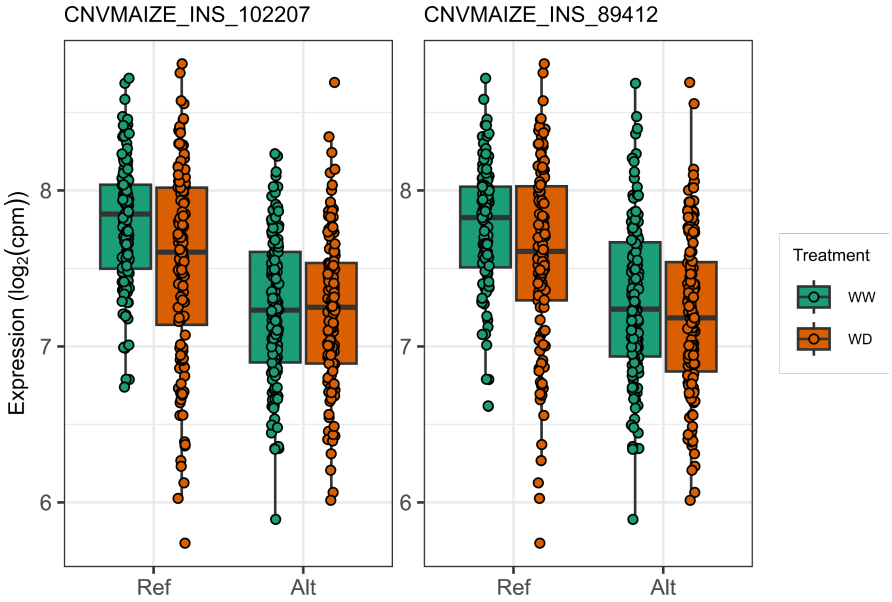
Supplemental Figure 6:* ZmSIP *eQTLs. *Each dot represents a physical eQTL position (B73 RefGen v4), and the transparency of the dot indicates the significance of the lead marker. Columns represent chromosomes, while rows represent genes. Data is shown for both WW and WD conditions. ZmSIP positions are indicated with vertical-coloured lines, and local eQTL are circled in black.*



Supplemental Figure 7: **ZmTIP eQTLs.** *Each dot represents a physical eQTL position (B73 RefGen v4), and the transparency of the dot indicates the significance of the lead marker. Columns represent chromosomes, while rows represent genes. Data is shown for both WW and WD conditions. ZmTIP positions are indicated with vertical-coloured lines, and local eQTL are circled in black.*



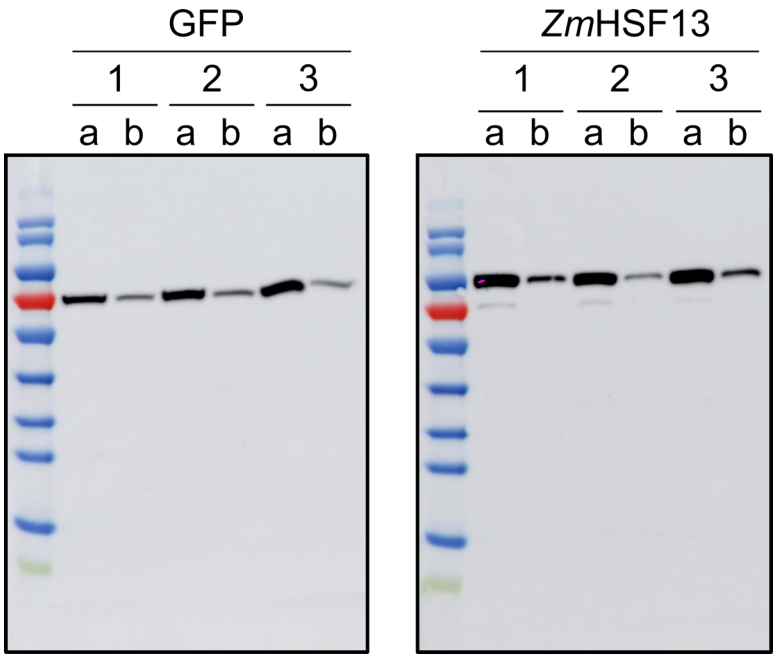
Supplemental Figure 8: Enriched and in eQTLs* TF *genes expression in log(cpm) in WW and WD. *The hybrids were grouped according to their heterotic groups. Differential expression was tested between the groups, and significance is shown at the top of the boxplots (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 10^{-4}).*



Supplemental Figure 9:* ZmTIP1;1 *expression between WW (green) and WD (orange), and between lines carrying either the reference (like B73) and alternative alleles. *The expression is expression in log_2(cpm). The expression is compared for two markers, CNVMAIZE_INS_102207 (left) and CNVMAIZE_INS_89412 (right). Each dot represents the BLUE of a given line.*

```
CLUSTAL O(1.2.4) multiple sequence alignment

Chr1:10151536..10150836      ggctcctgtactgggtggcgcagctgctggggtccaccgtggcgtgcttcctgctccgct      60
CNVMAIZE_INS_102207          ------------------------------------------------------------      0

Chr1:10151536..10150836      tctcgacgggcgggcaggccacgggcaccttcgggctgacgggcgtgtcggtgtgggagg      120
CNVMAIZE_INS_102207          ------------------------------------------------------------      0

Chr1:10151536..10150836      cgctggtgctggagatcgtgatgaccttcgggctggtgtacacggtgtacgcgacggcgg      180
CNVMAIZE_INS_102207          ------------------------------------------------------------      0

Chr1:10151536..10150836      tggacccgaagaagggcagcctgggcaccatcgcccccattgccatcggcttcatcgtgg      240
CNVMAIZE_INS_102207          ------------------------------------------------------------      0

Chr1:10151536..10150836      gggccaacatcctggtgggcggcgccttcgacggcgcgtccatgaaccccgccgtgtcct      300
CNVMAIZE_INS_102207          ----------------------------GACGGCGCGTCCATGAACCCCGCCGTGTCCT      31
                                                         ******************************

Chr1:10151536..10150836      tcggccccgccctcgtcagctgggagtggggctaccagtgggtgtactgggtcggccccc      360
CNVMAIZE_INS_102207          TCGGCCCCGCCCTCGTCAGCTGGGAGTGGGGCTACCAGTGGGTGTACTGGGTCGGCCCCC      91
                             ************************************************************

Chr1:10151536..10150836      tcatcggcggcggcctcgccggcgtcatctacgagctgctcttcatctcccacacccacg      420
CNVMAIZE_INS_102207          TCATCGGCGGCGGCCTCGCCGGCGTCATCTACGAGCTGCTCTTCATCTCCCACACCCACG      151
                             ************************************************************

Chr1:10151536..10150836      agcagctcccctccaccgactactaaaagccgaagccgaccgacccaggggtgcgcagcg      480
CNVMAIZE_INS_102207          AGCAGCTCCCCTCCACCGACTACTAAAAGCCGAAGCCGACCGACCCAGGGGTGCGCAGCG      211
                             ************************************************************

Chr1:10151536..10150836      cagcgcagctagcatccagcgctctctccc-----cgctcgccgccgtttagtttctctc      535
CNVMAIZE_INS_102207          CAGCGCAGCTAGCATCCAGCTCTCTCTCCCCGGCCCGCTCGCCGCCGTTTAGTTTCTCTC      271
                             ******************** *********     ************************

Chr1:10151536..10150836      gtgcacgtgcatgcatcgtagactgtctcaggccgccgctgcgccggcgccgctc-----      590
CNVMAIZE_INS_102207          GTGCACGTGCATGCATCGTAGACTGTCTCAGGCCGCCGCTGCGNNNNNNNNNNNNNNNNN      331
                             ******************************************

Chr1:10151536..10150836      -------gcatcgtttgcttccttggaccagcagtcgtcgccgtcaggctgaactgcttg      643
CNVMAIZE_INS_102207          CGCTCGCATCGTTTGCTTTTCCTTGGACCAGCAGTCGTCGCCGTCAGGTTGAGCTGCTTG      391
                                    *    ***************************** *** *******

Chr1:10151536..10150836      taaccgtgaattccggacccgtttgcttctgtaagatttgtgcatttgcctgagtaca--      701
CNVMAIZE_INS_102207          TAACCGTTAATTCCGGACCCGTTTGCTTCTGTAAGATTTGTGCATTTGCCTGTGAGTATT      451
                             ******* ********************************************* *   *

Chr1:10151536..10150836      ------------------------------------------------------------      701
CNVMAIZE_INS_102207          CACCGGCCCGCCCGACTCCCCACGTAGTAGCTCGGAGTACACCTCGATCTGCAGCATGCT      511

Chr1:10151536..10150836      ------------------------------------------------------------      701
CNVMAIZE_INS_102207          GCATGCGTGCGCTTGAAAATCTGGGATGCTTCTGCCTCCGATCGCGATGATTTCAACGGG      571

Chr1:10151536..10150836      ------------------------------------------------------------      701
CNVMAIZE_INS_102207          GGAATATAATCGTCCGTTGTCTCGTCGTCGACCCGCCTCCGGCGTGCAGCAGGCCGCGAT      631

Chr1:10151536..10150836      ------------------------------------------------------------      701
CNVMAIZE_INS_102207          AATCATGCAGCGCCATGTGCGGGATGGTACGCGCGCCACGTCGGAGTGACGGGGCATAGC      691

Chr1:10151536..10150836      ------------------------------------------------------------      701
CNVMAIZE_INS_102207          GATAAGGCTGACATGCAGATCTGCGCCATGTCGCTGTCGGATGGACCTGGANNNNNNNNN      751

Chr1:10151536..10150836      ------------------------------------------------------------      701
CNVMAIZE_INS_102207          ACCCAATATTCACCCCGGGCGCGCGAGTGGTGCCAATAGCCGCATAGCAGTAAGCCAGTA      811

Chr1:10151536..10150836      ------------------------------------------------------------      701
CNVMAIZE_INS_102207          AACGCGCACTGGACACCGATGGTCCTCCTCCCCCTCCATGGATGGGCAAGGCAACGGCAT      871

Chr1:10151536..10150836      ------------------          701
CNVMAIZE_INS_102207          CTACTCCGAGCAGTAGCTA         890
```

Supplemental Figure 10: Alignment between a 700 bp fragment from the 3'end of ZmTIP1;1 and an 890 bp fragment from CNVMAIZE_INS_102207. *The alignment was performed using clustal omega. The sequence for the ZmTIP1;1 fragment was downloaded from maizeGDB from the B73 v4 sequence and reverse complemented.*



Supplemental Figure 11: Western blot to verify the production and the correct binding of the Halo-tagged proteins on the magnetic beads for the DAP-seq experiment. *Halo-tagged proteins were produced using the wheat (Triticum aestivum) germ in vitro transcription/translation SP6 promoter system (Promega, Madison, WI, USA). The protein expression and binding to the Magne® HaloTag® Beads (Promega, Madison, WI, USA) was validated by western blot with the 1:1,000 Anti-HaloTag® Polyclonal Rabbit Antibody (G9281 - Promega, Madison, WI, USA). The secondary antibody used to reveal the western blots was the 1:20,000 Anti-Rabbit antibody (A3687 – Sigma-Aldrich, St-Louis, MI, USA). These results show that the production of GFP (on the left) and ZmHSF13 (on the right) was successful in the three replicates (one to three) and that the protein quantity was higher before the binding to the Magne® HaloTag® Beads (a) than after (b).*

Supplemental Tables

Supplemental Table 1: List of the 254 varieties used in the DROPS panel. *Varieties are organised in two columns, one with the variety name and the other with the ADMIXTURE group they were associated to by Negro et al. (2019).*

Variety ID	Heterotic Group	Variety ID	Heterotic Group	Variety ID	Heterotic Group	Variety ID	Heterotic Group	Variety ID	Heterotic Group
11430	PH207	EZ34	Oh43	I233	W117-Type	N25	W117-Type	UH_S025	W117-Type
A3	W117-Type	EZ35	W117-Type	I238	W117-Type	N6	W117-Type	UH250	W117-Type
A310	W117-Type	EZ36	W117-Type	I242	W117-Type	NC290	W117-Type	UH304	PH207
A347	W117-Type	EZ37	B14a	I261	W117-Type	NC358	W117-Type	Va26	Oh43
A374	W117-Type	EZ38	B73	I267	W117-Type	NDB8	W117-Type	W117	W117-Type
A375	W117-Type	EZ40	B14a	IDT	W117-Type	NK764	B73	W153R	W117-Type
A554	W117-Type	EZ42	B14a	LAN496	Mo17	NK807	B14a	W182E	W117-Type
AS5707	Mo17	EZ47	W117-Type	LH123Ht	Oh43	NQ508	PH207	W23	W117-Type
B100	Oh43	EZ48	B14a	LH145	B14a	NS501	PH207	W602S	Mo17
B104	B73	F04401	PH207	LH38	Oh43	NS701	B73	W604S	Mo17
B105	W117-Type	F04402	PH207	LH39	Oh43	Oh02	W117-Type	W64A	W117-Type
B106	W117-Type	F04701	Mo17	LH59	Mo17	Oh33	W117-Type	W9	W117-Type
B107	PH207	F04702	Mo17	LH60	Mo17	Oh40B	Oh43	W95115	W117-Type
B108	W117-Type	F05101	B73	LH65	Mo17	Oh43	Oh43	Wf9	W117-Type
B109	B73	F05404	B73	LH74	B73	Os426	W117-Type		
B110	B73	F1808	B14a	LH82	W117-Type	P465P	W117-Type		
B113	W117-Type	F1890	PH207	LH93	W117-Type	Pa35	W117-Type		
B14a	B14a	F218	W117-Type	Lo1016	B73	Pa36	W117-Type		
B37	B14a	F252	W117-Type	Lo1026	B73	Pa405	W117-Type		
B73	B73	F353	PH207	Lo1035	Mo17	Pa91	Oh43		
B84	B73	F354	W117-Type	Lo1038	W117-Type	PB116	W117-Type		
B89	W117-Type	F608	W117-Type	Lo1056	Mo17	PB98TR	W117-Type		
B97	W117-Type	F618	B14a	Lo1063	B14a	PH207	PH207		
B98	W117-Type	F7001	W117-Type	Lo1087	B73	PHB09	B14a		
C103	Mo17	F7019	B14a	Lo1094	B73	PHG35	Mo17		
CO109	W117-Type	F7025	B73	Lo1095	Mo17	PHG39	B14a		
CR1Ht	Mo17	F7028	W117-Type	Lo1101	B73	PHG47	Oh43		
D09	W117-Type	F7057	W117-Type	Lo1106	B73	PHG50	PH207		
DE811	B14a	F7058	W117-Type	Lo1123	B14a	PHG71	PH207		
DK4676A	B14a	F7081	W117-Type	Lo1124	Mo17	PHG80	B14a		
DK78010	B73	F7082	W117-Type	Lo1172	Mo17	PHG83	PH207		
DK78371A	Oh43	F712	W117-Type	Lo1180	PH207	PHG84	Mo17		
DKFAPW	B14a	F748	B14a	Lo1187	B73	PHG86	B73		
DKFBHJ	B14a	F752	B14a	Lo1199	B73	PHH93	PH207		
DKIBO2	PH207	F816	Mo17	Lo1203	Mo17	PHJ40	Mo17		
DKMBST	Oh43	F838	W117-Type	Lo1223	B73	PHK29	B73		
EA1027	W117-Type	F874	W117-Type	Lo1242	Mo17	PHK76	Mo17		
EA1163	W117-Type	F888	W117-Type	Lo1251	B73	PHR36	Mo17		
EA3076	W117-Type	F894	B14a	Lo1253	B73	PHT77	Mo17		
EC136	B73	F908	W117-Type	Lo1261	B73	PHV63	B14a		
EC140	B73	F912	PH207	Lo1266	B14a	PHW65	Mo17		
EC151	B14a	F918	B14a	Lo1270	Mo17	PHZ51	Mo17		
EC169	B14a	F922	W117-Type	Lo1273	B14a	PP147	W117-Type		
EC175	B14a	F924	B73	Lo1274	Mo17	SC-Malawi	W117-Type		
EC232	W117-Type	F98902	B73	Lo1280	Mo17	SGI061	PH207		
EC242C	PH207	FP1	W117-Type	Lo1282	Mo17	UH_2500	W117-Type		
EC334	B14a	FR19	B14a	Lo1284	Mo17	UH_6102	PH207		
EP10	W117-Type	FR697	Mo17	Lo1288	Mo17	UH_6179	PH207		
EP2008-18	Oh43	H99	Oh43	Lo1290	Mo17	UH_P024	PH207		
EP2008-22	Oh43	HMV5301	PH207	Lo1301	B73	UH_P060	W117-Type		
EP29	W117-Type	HMV5325	B73	Lo904	B73	UH_P064	PH207		
EP51	W117-Type	HMV5343	B14a	Lp5	B73	UH_P074	PH207		
EP52	W117-Type	HMV5347	B14a	ML606	Oh43	UH_P087	W117-Type		
EP55	W117-Type	HMV5405	B14a	Mo15W	W117-Type	UH_P089	PH207		
EP67	W117-Type	HMV5409	B14a	Mo17	Mo17	UH_P104	W117-Type		
EP72	W117-Type	HMV5422	B73	MS153	W117-Type	UH_P115	PH207		
EP77	W117-Type	HMV5502	B14a	MS71	W117-Type	UH_P128	PH207		
EZ11A	W117-Type	I198	W117-Type	N16	W117-Type	UH_P148	PH207		
EZ18	W117-Type	I205	W117-Type	N192	B73	UH_S018	W117-Type		
EZ31	B14a	I211	W117-Type	N22	W117-Type	UH_S020	W117-Type		

Supplemental Table 2: Between 3'RNA-seq samples Pearson correlation coefficient. *The coefficient was calculated only for varieties having three replicates.*

Variety_ID	Code_Sample_1	Code_Sample_2	Pearson coefficient
A374	A374-WW-1	A374-WW-3	0.93
A374	A374-WW-1	A374-WW-2	0.95
A374	A374-WW-2	A374-WW-1	0.95
A374	A374-WW-2	A374-WW-3	0.92
A374	A374-WD-1	A374-WD-2	0.92
A374	A374-WD-1	A374-WD-3	0.96
A374	A374-WD-2	A374-WD-1	0.92
A374	A374-WD-2	A374-WD-3	0.91
B37	B37-WW-1	B37-WW-2	0.94
B37	B37-WW-1	B37-WW-3	0.95
B37	B37-WW-2	B37-WW-1	0.94
B37	B37-WW-2	B37-WW-3	0.95
B37	B37-WD-1	B37-WD-2	0.95
B37	B37-WD-1	B37-WD-3	0.94
B37	B37-WD-2	B37-WD-1	0.95
B37	B37-WD-2	B37-WD-3	0.93
DKMBST	DKMBST-WW-1	DKMBST-WW-2	0.96
DKMBST	DKMBST-WW-1	DKMBST-WW-3	0.94
DKMBST	DKMBST-WW-2	DKMBST-WW-1	0.96
DKMBST	DKMBST-WW-2	DKMBST-WW-3	0.93
DKMBST	DKMBST-WD-1	DKMBST-WD-3	0.92
DKMBST	DKMBST-WD-1	DKMBST-WD-2	0.91
DKMBST	DKMBST-WD-2	DKMBST-WD-1	0.91
DKMBST	DKMBST-WD-2	DKMBST-WD-3	0.89
EZ31	EZ31-WW-1	EZ31-WW-2	0.94
EZ31	EZ31-WW-1	EZ31-WW-3	0.94
EZ31	EZ31-WW-2	EZ31-WW-3	0.91
EZ31	EZ31-WW-2	EZ31-WW-1	0.94
EZ31	EZ31-WD-1	EZ31-WD-2	0.92
EZ31	EZ31-WD-1	EZ31-WD-3	0.94
EZ31	EZ31-WD-2	EZ31-WD-1	0.92
EZ31	EZ31-WD-2	EZ31-WD-3	0.92
EZ47	EZ47-WW-1	EZ47-WW-2	0.95
EZ47	EZ47-WW-1	EZ47-WW-3	0.95
EZ47	EZ47-WW-2	EZ47-WW-3	0.93
EZ47	EZ47-WW-2	EZ47-WW-1	0.95
EZ47	EZ47-WD-1	EZ47-WD-2	0.92
EZ47	EZ47-WD-1	EZ47-WD-3	0.95
EZ47	EZ47-WD-2	EZ47-WD-3	0.93
EZ47	EZ47-WD-2	EZ47-WD-1	0.92
F252	F252-WW-1	F252-WW-2	0.90
F252	F252-WW-1	F252-WW-3	0.91
F252	F252-WW-2	F252-WW-3	0.91
F252	F252-WW-2	F252-WW-1	0.90
F252	F252-WD-1	F252-WD-3	0.93
F252	F252-WD-1	F252-WD-2	0.93
F252	F252-WD-2	F252-WD-3	0.93
F252	F252-WD-2	F252-WD-1	0.93
F7001	F7001-WW-1	F7001-WW-3	0.94
F7001	F7001-WW-1	F7001-WW-2	0.96
F7001	F7001-WW-2	F7001-WW-3	0.92
F7001	F7001-WW-2	F7001-WW-1	0.96
F7001	F7001-WD-1	F7001-WD-2	0.93
F7001	F7001-WD-1	F7001-WD-3	0.95
F7001	F7001-WD-2	F7001-WD-1	0.93
F7001	F7001-WD-2	F7001-WD-3	0.93

F7019	F7019-WW-1	F7019-WW-2	0.92
F7019	F7019-WW-1	F7019-WW-3	0.95
F7019	F7019-WW-2	F7019-WW-3	0.92
F7019	F7019-WW-2	F7019-WW-1	0.92
F7019	F7019-WD-1	F7019-WD-2	0.91
F7019	F7019-WD-1	F7019-WD-3	0.95
F7019	F7019-WD-2	F7019-WD-3	0.93
F7019	F7019-WD-2	F7019-WD-1	0.91
F874	F874-WW-1	F874-WW-2	0.94
F874	F874-WW-1	F874-WW-3	0.93
F874	F874-WW-2	F874-WW-3	0.92
F874	F874-WW-2	F874-WW-1	0.94
F874	F874-WD-1	F874-WD-3	0.95
F874	F874-WD-1	F874-WD-2	0.91
F874	F874-WD-2	F874-WD-3	0.93
F874	F874-WD-2	F874-WD-1	0.91
H99	H99-WW-1	H99-WW-3	0.94
H99	H99-WW-1	H99-WW-2	0.95
H99	H99-WW-2	H99-WW-3	0.93
H99	H99-WW-2	H99-WW-1	0.95
H99	H99-WD-1	H99-WD-2	0.92
H99	H99-WD-1	H99-WD-3	0.93
H99	H99-WD-2	H99-WD-1	0.92
H99	H99-WD-2	H99-WD-3	0.93
HMV5409	HMV5409-WW-1	HMV5409-WW-2	0.95
HMV5409	HMV5409-WW-1	HMV5409-WW-3	0.92
HMV5409	HMV5409-WW-2	HMV5409-WW-3	0.91
HMV5409	HMV5409-WW-2	HMV5409-WW-1	0.95
HMV5409	HMV5409-WD-1	HMV5409-WD-2	0.92
HMV5409	HMV5409-WD-1	HMV5409-WD-3	0.95
HMV5409	HMV5409-WD-2	HMV5409-WD-3	0.94
HMV5409	HMV5409-WD-2	HMV5409-WD-1	0.92
Lo1094	Lo1094-WW-1	Lo1094-WW-3	0.95
Lo1094	Lo1094-WW-1	Lo1094-WW-2	0.93
Lo1094	Lo1094-WW-2	Lo1094-WW-3	0.90
Lo1094	Lo1094-WW-2	Lo1094-WW-1	0.93
Lo1094	Lo1094-WD-1	Lo1094-WD-3	0.93
Lo1094	Lo1094-WD-1	Lo1094-WD-2	0.89
Lo1094	Lo1094-WD-2	Lo1094-WD-3	0.92
Lo1094	Lo1094-WD-2	Lo1094-WD-1	0.89
Lo1123	Lo1123-WW-1	Lo1123-WW-2	0.96
Lo1123	Lo1123-WW-1	Lo1123-WW-3	0.96
Lo1123	Lo1123-WW-2	Lo1123-WW-3	0.95
Lo1123	Lo1123-WW-2	Lo1123-WW-1	0.96
Lo1123	Lo1123-WD-1	Lo1123-WD-3	0.94
Lo1123	Lo1123-WD-1	Lo1123-WD-2	0.91
Lo1123	Lo1123-WD-2	Lo1123-WD-3	0.93
Lo1123	Lo1123-WD-2	Lo1123-WD-1	0.91
Lo1203	Lo1203-WW-1	Lo1203-WW-2	0.95
Lo1203	Lo1203-WW-1	Lo1203-WW-3	0.94
Lo1203	Lo1203-WW-2	Lo1203-WW-1	0.95
Lo1203	Lo1203-WW-2	Lo1203-WW-3	0.94
Lo1203	Lo1203-WD-1	Lo1203-WD-2	0.94
Lo1203	Lo1203-WD-1	Lo1203-WD-3	0.94
Lo1203	Lo1203-WD-2	Lo1203-WD-3	0.95
Lo1203	Lo1203-WD-2	Lo1203-WD-1	0.94
Lo1290	Lo1290-WW-1	Lo1290-WW-3	0.96
Lo1290	Lo1290-WW-1	Lo1290-WW-2	0.96
Lo1290	Lo1290-WW-2	Lo1290-WW-3	0.95
Lo1290	Lo1290-WW-2	Lo1290-WW-1	0.96
Lo1290	Lo1290-WD-1	Lo1290-WD-2	0.94
Lo1290	Lo1290-WD-1	Lo1290-WD-3	0.95

Lo1290	Lo1290-WD-2	Lo1290-WD-1	0.94
Lo1290	Lo1290-WD-2	Lo1290-WD-3	0.92
Mo17	Mo17-WW-1	Mo17-WW-2	0.92
Mo17	Mo17-WW-1	Mo17-WW-3	0.96
Mo17	Mo17-WW-2	Mo17-WW-1	0.92
Mo17	Mo17-WW-2	Mo17-WW-3	0.93
Mo17	Mo17-WD-1	Mo17-WD-2	0.92
Mo17	Mo17-WD-1	Mo17-WD-3	0.93
Mo17	Mo17-WD-2	Mo17-WD-1	0.92
Mo17	Mo17-WD-2	Mo17-WD-3	0.94
PH207	PH207-WW-1	PH207-WW-3	0.95
PH207	PH207-WW-1	PH207-WW-2	0.93
PH207	PH207-WW-2	PH207-WW-3	0.92
PH207	PH207-WW-2	PH207-WW-1	0.93
PH207	PH207-WD-1	PH207-WD-2	0.93
PH207	PH207-WD-1	PH207-WD-3	0.95
PH207	PH207-WD-2	PH207-WD-1	0.93
PH207	PH207-WD-2	PH207-WD-3	0.93
UH_S020	UH-S020-WW-1	UH-S020-WW-3	0.95
UH_S020	UH-S020-WW-1	UH-S020-WW-2	0.95
UH_S020	UH-S020-WW-2	UH-S020-WW-3	0.93
UH_S020	UH-S020-WW-2	UH-S020-WW-1	0.95
UH_S020	UH-S020-WD-1	UH-S020-WD-2	0.93
UH_S020	UH-S020-WD-1	UH-S020-WD-3	0.95
UH_S020	UH-S020-WD-2	UH-S020-WD-3	0.93
UH_S020	UH-S020-WD-2	UH-S020-WD-1	0.93
Wf9	Wf9-WW-1	Wf9-WW-3	0.96
Wf9	Wf9-WW-1	Wf9-WW-2	0.95
Wf9	Wf9-WW-2	Wf9-WW-3	0.95
Wf9	Wf9-WW-2	Wf9-WW-1	0.95
Wf9	Wf9-WD-1	Wf9-WD-2	0.94
Wf9	Wf9-WD-1	Wf9-WD-3	0.95
Wf9	Wf9-WD-2	Wf9-WD-1	0.94
Wf9	Wf9-WD-2	Wf9-WD-3	0.93

Supplemental Table 3: Raw data of the dual-luciferase experiment performed in B104 mesophyll cells. *Replicate column indicate the number of the protoplast transfection for a given TF cassette on the day of the experiment. The firefly and* Renilla *luciferases intensity were measured with a luminometer, and the ratio was calculated (FIRE/REN ratio column).*

Replicate	Date	firefly Luc.	*renilla* Luc.	FIRE/REN ratio	TF cassette	Promoter cassette	Cultivar
1	10-05-24	7680	24699	0.31	*control*	*pTIP1;1*	B104
2	10-05-24	3934	9061	0.43	*control*	*pTIP1;1*	B104
3	10-05-24	2425	7116	0.34	*control*	*pTIP1;1*	B104
1	10-05-24	43269	50882	0.85	*p35S::HSF13*	*pTIP1;1*	B104
2	10-05-24	52838	77582	0.68	*p35S::HSF13*	*pTIP1;1*	B104
3	10-05-24	15185	21072	0.72	*p35S::HSF13*	*pTIP1;1*	B104
1	13-05-24	2220	7406	0.30	*control*	*pTIP1;1*	B104
2	13-05-24	3838	10291	0.37	*control*	*pTIP1;1*	B104
3	13-05-24	2708	10302	0.26	*control*	*pTIP1;1*	B104
1	13-05-24	26735	82263	0.32	*p35S::HSF13*	*pTIP1;1*	B104
2	13-05-24	40226	95262	0.42	*p35S::HSF13*	*pTIP1;1*	B104
3	13-05-24	24636	52532	0.47	*p35S::HSF13*	*pTIP1;1*	B104
2	15-05-24	1094	2396	0.46	*control*	*pTIP1;1*	B104
3	15-05-24	415	1525	0.27	*control*	*pTIP1;1*	B104
1	15-05-24	2026	7335	0.28	*control*	*pTIP1;1*	B104

1	15-05-24	9563	16485	0.58	*p35S::HSF13*	*pTIP1;1*	B104
3	15-05-24	10420	23567	0.44	*p35S::HSF13*	*pTIP1;1*	B104
2	15-05-24	4704	12034	0.39	*p35S::HSF13*	*pTIP1;1*	B104

Supplemental Table 4: Information about the sequencing results of the DAP-seq experiment

Sample	Number of read pairs	Q30 (% of total reads)	Number of reads after trimming	Uniquely mapped reads (% of reads)
GFP_LIB1	23,398,389	80.3	23,366,281	11.9
GFP_LIB2	21,169,573	76.6	21,158,970	54.7
GFP_LIB3	21,794,299	77.2	21,772,041	27.3
***Zm*HSF13_LIB1**	23,463,172	92.9	23,460,649	59.3
***Zm*HSF13_LIB2**	20,882,026	90.4	20,855,063	54.9
***Zm*HSF13_LIB3**	22,614,815	83	22,589,975	46.8

Supplemental Table 5: Top 20 enriched GO terms in DE genes from the 3'RNA-seq experiment.

GO	name	p-value (FDR)
GO:0009414	response to water deprivation	1.13E-07
GO:0009415	response to water	1.52E-07
GO:0001101	response to acid chemical	1.52E-07
GO:0009628	response to abiotic stimulus	3.06E-07
GO:0050896	response to stimulus	5.26E-06
GO:0010035	response to inorganic substance	7.44E-06
GO:1901700	response to oxygen-containing compound	5.44E-05
GO:0006950	response to stress	1.16E-04
GO:0009611	response to wounding	1.80E-04
GO:0009804	coumarin metabolic process	4.59E-04
GO:0009805	coumarin biosynthetic process	4.59E-04
GO:0043436	oxoacid metabolic process	7.23E-04
GO:0006970	response to osmotic stress	7.95E-04
GO:0006082	organic acid metabolic process	7.95E-04
GO:0009651	response to salt stress	9.35E-04
GO:0042398	cellular modified amino acid biosynthetic process	1.08E-03
GO:0006598	polyamine catabolic process	1.43E-03
GO:0071941	nitrogen cycle metabolic process	1.70E-03
GO:0042402	cellular biogenic amine catabolic process	1.87E-03
GO:0009737	response to abscisic acid	2.30E-03

Supplemental Table 6: Raw data of the dual-fluorescence experiment performed in B73 mesophyll cells protoplasts. *Well column indicate the letter of the protoplast transfection for a given plasmid on the day of the experiment and the replicate column indicate the number of the number of the protoplast. The mCherry and mVenus intensity were quantified with ImageJ software as explained in material and methods, and the ratio and its log_2 were calculated.*

Date	Well	Rep	Plasmid	mCherry	mVenus	ratio	log_2ratio
11-08-23	a	1	p35S::Nac40	14745.37	2087.68	0.14	-2.82
11-08-23	a	2	p35S::Nac40	18866.19	1692.54	0.09	-3.48
11-08-23	a	3	p35S::Nac40	21697.27	4925.44	0.23	-2.14
11-08-23	a	4	p35S::Nac40	15855.95	2334.14	0.15	-2.76
11-08-23	a	5	p35S::Nac40	6426.25	4663.15	0.73	-0.46
11-08-23	a	6	p35S::Nac40	15816.73	2307.79	0.15	-2.78
11-08-23	a	7	p35S::Nac40	17923.18	8871.38	0.49	-1.01
11-08-23	a	8	p35S::Nac40	23551.34	3699.97	0.16	-2.67
11-08-23	a	9	p35S::Nac40	45026.17	5596.46	0.12	-3.01
11-08-23	a	10	p35S::Nac40	13722.61	1411.58	0.10	-3.28

11-08-23	a	11	p35S::Nac40	20965.96	13264.70	0.63	-0.66
11-08-23	a	12	p35S::Nac40	14976.28	4019.70	0.27	-1.90
11-08-23	a	1	Nac40	10022.47	4518.76	0.45	-1.15
11-08-23	a	2	Nac40	22154.70	7315.39	0.33	-1.60
11-08-23	a	3	Nac40	13499.16	3016.12	0.22	-2.16
11-08-23	a	4	Nac40	18257.33	735.74	0.04	-4.63
11-08-23	a	5	Nac40	4796.19	2283.73	0.48	-1.07
11-08-23	a	6	Nac40	19238.15	2286.68	0.12	-3.07
11-08-23	a	7	Nac40	20882.58	2653.34	0.13	-2.98
11-08-23	a	8	Nac40	13704.58	1297.71	0.09	-3.40
11-08-23	a	9	Nac40	17044.30	1169.27	0.07	-3.87
11-08-23	a	10	Nac40	12707.63	1181.39	0.09	-3.43
11-08-23	a	11	Nac40	17176.35	6112.87	0.36	-1.49
11-08-23	a	12	Nac40	25290.80	4377.65	0.17	-2.53
11-08-23	a	13	Nac40	26431.67	5002.30	0.19	-2.40
11-08-23	a	14	Nac40	25855.14	1558.42	0.06	-4.05
11-08-23	b	1	Nac40	32734.74	2071.48	0.06	-3.98
11-08-23	b	2	Nac40	23246.84	2290.69	0.10	-3.34
11-08-23	b	3	Nac40	11068.54	3746.78	0.34	-1.56
11-08-23	b	4	Nac40	8142.27	573.96	0.07	-3.83
11-08-23	b	5	Nac40	31585.53	3606.48	0.11	-3.13
11-08-23	b	6	Nac40	25977.73	3854.69	0.15	-2.75
11-08-23	b	7	Nac40	19083.20	964.60	0.05	-4.31
11-08-23	b	8	Nac40	29018.75	3380.47	0.12	-3.10
11-08-23	b	9	Nac40	29944.75	3307.04	0.11	-3.18
11-08-23	b	10	Nac40	21628.39	7160.25	0.33	-1.59
11-08-23	b	11	Nac40	9161.25	2400.85	0.26	-1.93
11-08-23	b	12	Nac40	23053.72	10570.55	0.46	-1.12
11-08-23	b	13	Nac40	28225.53	5632.98	0.20	-2.33
11-08-23	b	1	p35S::Nac40	21756.17	3536.22	0.16	-2.62
11-08-23	b	2	p35S::Nac40	14736.69	2618.98	0.18	-2.49
11-08-23	b	3	p35S::Nac40	20947.42	1349.48	0.06	-3.96
11-08-23	b	4	p35S::Nac40	16849.46	4414.99	0.26	-1.93
11-08-23	b	5	p35S::Nac40	13112.48	2969.34	0.23	-2.14
11-08-23	b	6	p35S::Nac40	18554.44	8520.51	0.46	-1.12
11-08-23	b	7	p35S::Nac40	20281.31	2564.89	0.13	-2.98
11-08-23	b	8	p35S::Nac40	5690.21	4750.17	0.83	-0.26
11-08-23	b	9	p35S::Nac40	19809.27	4539.78	0.23	-2.13
11-08-23	b	10	p35S::Nac40	11223.48	7306.22	0.65	-0.62
11-08-23	b	11	p35S::Nac40	26790.97	2059.15	0.08	-3.70
11-08-23	b	12	p35S::Nac40	27601.73	18137.74	0.66	-0.61
11-08-23	b	13	p35S::Nac40	4965.00	3433.21	0.69	-0.53
11-08-23	b	14	p35S::Nac40	31477.96	7223.26	0.23	-2.12
11-08-23	b	15	p35S::Nac40	6483.73	4508.22	0.70	-0.52
02-08-23	a	1	Nac40	11525.08	3444.14	0.30	-1.74
02-08-23	a	2	Nac40	10943.64	3936.62	0.36	-1.48
02-08-23	a	3	Nac40	10984.29	866.87	0.08	-3.66
02-08-23	a	4	Nac40	20065.32	2189.96	0.11	-3.20
02-08-23	a	5	Nac40	25183.60	3402.69	0.14	-2.89
02-08-23	a	6	Nac40	14329.61	1355.18	0.09	-3.40
02-08-23	a	7	Nac40	12247.72	433.25	0.04	-4.82
02-08-23	a	8	Nac40	5970.82	198.99	0.03	-4.91
02-08-23	a	9	Nac40	9267.35	843.63	0.09	-3.46
02-08-23	a	10	Nac40	12524.68	2668.15	0.21	-2.23
02-08-23	a	11	Nac40	17780.18	1398.09	0.08	-3.67
02-08-23	a	12	Nac40	15823.11	6715.20	0.42	-1.24
02-08-23	a	13	Nac40	19857.54	644.61	0.03	-4.95
02-08-23	a	14	Nac40	13770.15	1731.34	0.13	-2.99
02-08-23	a	15	Nac40	38194.36	3406.56	0.09	-3.49
02-08-23	b	1	Nac40	8998.78	1061.83	0.12	-3.08
02-08-23	b	2	Nac40	16272.24	1278.42	0.08	-3.67
02-08-23	b	3	Nac40	27306.90	2330.70	0.09	-3.55

02-08-23	b	4	Nac40	16305.89	2514.10	0.15	-2.70
02-08-23	b	5	Nac40	10140.82	4292.13	0.42	-1.24
02-08-23	b	6	Nac40	14814.35	1693.25	0.11	-3.13
02-08-23	b	7	Nac40	8865.93	659.43	0.07	-3.75
02-08-23	b	8	Nac40	44042.95	10776.05	0.24	-2.03
02-08-23	b	9	Nac40	36317.16	7233.92	0.20	-2.33
02-08-23	b	10	Nac40	33416.41	16342.53	0.49	-1.03
02-08-23	b	11	Nac40	6252.76	228.79	0.04	-4.77
02-08-23	b	12	Nac40	18444.88	6068.65	0.33	-1.60
02-08-23	b	13	Nac40	5238.68	2529.48	0.48	-1.05
02-08-23	b	14	Nac40	20356.18	2729.45	0.13	-2.90
02-08-23	a	1	p35S::Nac40	37654.16	6813.07	0.18	-2.47
02-08-23	a	2	p35S::Nac40	38816.37	15434.35	0.40	-1.33
02-08-23	a	3	p35S::Nac40	11025.18	5295.13	0.48	-1.06
02-08-23	a	4	p35S::Nac40	11327.90	6853.52	0.61	-0.72
02-08-23	a	5	p35S::Nac40	25557.94	6808.06	0.27	-1.91
02-08-23	a	6	p35S::Nac40	20512.58	4925.65	0.24	-2.06
02-08-23	a	7	p35S::Nac40	10130.68	4833.56	0.48	-1.07
02-08-23	a	8	p35S::Nac40	17165.19	4636.19	0.27	-1.89
02-08-23	a	9	p35S::Nac40	7107.65	513.03	0.07	-3.79
02-08-23	a	10	p35S::Nac40	12832.96	8397.66	0.65	-0.61
02-08-23	a	11	p35S::Nac40	14906.32	18749.60	1.26	0.33
02-08-23	a	12	p35S::Nac40	16831.75	4074.53	0.24	-2.05
02-08-23	a	13	p35S::Nac40	4039.76	4233.22	1.05	0.07
02-08-23	a	14	p35S::Nac40	11847.19	2721.35	0.23	-2.12
02-08-23	a	15	p35S::Nac40	21623.37	11625.24	0.54	-0.90
02-08-23	a	16	p35S::Nac40	44657.49	13246.69	0.30	-1.75
02-08-23	b	1	p35S::Nac40	19356.85	1989.69	0.10	-3.28
02-08-23	b	2	p35S::Nac40	9295.30	1834.72	0.20	-2.34
02-08-23	b	3	p35S::Nac40	13263.83	973.00	0.07	-3.77
02-08-23	b	4	p35S::Nac40	3640.21	3974.33	1.09	0.13
02-08-23	b	5	p35S::Nac40	17336.15	4809.19	0.28	-1.85
02-08-23	b	6	p35S::Nac40	19658.00	17265.75	0.88	-0.19
02-08-23	b	7	p35S::Nac40	9583.56	1129.98	0.12	-3.08
02-08-23	b	8	p35S::Nac40	29224.71	11973.38	0.41	-1.29
02-08-23	b	9	p35S::Nac40	21195.78	3698.34	0.17	-2.52
02-08-23	b	10	p35S::Nac40	5814.90	180.00	0.03	-5.01
11-07-23	a	1	Nac40	10239.32	4739.56	0.46	-1.11
11-07-23	a	2	Nac40	20703.06	8574.37	0.41	-1.27
11-07-23	a	3	Nac40	3031.68	233.06	0.08	-3.70
11-07-23	a	4	Nac40	4392.42	201.15	0.05	-4.45
11-07-23	a	5	Nac40	2234.82	1004.56	0.45	-1.15
11-07-23	a	6	Nac40	13837.17	4089.20	0.30	-1.76
11-07-23	a	7	Nac40	8817.14	745.17	0.08	-3.56
11-07-23	a	8	Nac40	2595.92	365.91	0.14	-2.83
11-07-23	a	9	Nac40	6687.76	608.16	0.09	-3.46
11-07-23	a	10	Nac40	4772.80	401.44	0.08	-3.57
11-07-23	a	11	Nac40	12486.02	3468.64	0.28	-1.85
11-07-23	a	12	Nac40	8081.43	449.05	0.06	-4.17
11-07-23	b	1	Nac40	19743.50	11349.44	0.57	-0.80
11-07-23	b	2	Nac40	10236.08	3457.40	0.34	-1.57
11-07-23	b	3	Nac40	34761.36	2383.07	0.07	-3.87
11-07-23	b	4	Nac40	15709.16	3735.72	0.24	-2.07
11-07-23	b	5	Nac40	18722.09	3245.88	0.17	-2.53
11-07-23	b	6	Nac40	14305.18	4420.75	0.31	-1.69
11-07-23	b	7	Nac40	23624.07	2057.95	0.09	-3.52
11-07-23	b	8	Nac40	19305.75	6946.42	0.36	-1.47
11-07-23	b	9	Nac40	24231.94	3843.41	0.16	-2.66
11-07-23	a	1	p35S::Nac40	9170.35	839.58	0.09	-3.45
11-07-23	a	2	p35S::Nac40	12908.88	6428.78	0.50	-1.01
11-07-23	a	3	p35S::Nac40	8188.82	9128.83	1.11	0.16
11-07-23	a	4	p35S::Nac40	4490.70	7169.58	1.60	0.67

11-07-23	a	5	p35S::Nac40	6590.68	7553.17	1.15	0.20
11-07-23	a	6	p35S::Nac40	5469.14	2956.72	0.54	-0.89
11-07-23	a	7	p35S::Nac40	10717.91	7180.73	0.67	-0.58
11-07-23	a	8	p35S::Nac40	20933.93	8247.99	0.39	-1.34
11-07-23	a	9	p35S::Nac40	9617.28	14244.34	1.48	0.57
11-07-23	a	10	p35S::Nac40	8854.88	3785.72	0.43	-1.23
11-07-23	a	11	p35S::Nac40	8805.52	2931.02	0.33	-1.59
11-07-23	a	12	p35S::Nac40	14322.30	3836.78	0.27	-1.90
11-07-23	b	1	p35S::Nac40	8797.06	5295.68	0.60	-0.73
11-07-23	b	2	p35S::Nac40	8903.57	1078.54	0.12	-3.05
11-07-23	b	3	p35S::Nac40	8951.28	8007.40	0.89	-0.16
11-07-23	b	4	p35S::Nac40	4859.92	3556.94	0.73	-0.45
11-07-23	b	5	p35S::Nac40	5483.40	7116.34	1.30	0.38
11-07-23	b	6	p35S::Nac40	7987.31	6612.00	0.83	-0.27
11-07-23	b	7	p35S::Nac40	2900.87	1312.47	0.45	-1.14
11-07-23	b	8	p35S::Nac40	4679.59	3073.29	0.66	-0.61
11-07-23	b	9	p35S::Nac40	6575.19	5790.88	0.88	-0.18

Supplemental Table 7: List of the genes used in the benchmarking of the different preprocessing methods

B73 v4 ID	Selection
Zm00001d010159	housekeeping gene RT-qPCR
Zm00001d053838	housekeeping gene RT-qPCR
Zm00001d046449	housekeeping gene RT-qPCR
Zm00001d002690	RT-qPCR performed on the same panel
Zm00001d051403	RT-qPCR performed on the same panel
Zm00001d019563	RT-qPCR performed on the same panel
Zm00001d005421	RT-qPCR performed on the same panel
Zm00001d003006	RT-qPCR performed on the same panel
Zm00001d011183	correlated with PC1
Zm00001d031332	correlated with PC1
Zm00001d030995	correlated with PC1
Zm00001d022561	correlated with PC1
Zm00001d031275	correlated with PC1
Zm00001d007445	correlated with PC1
Zm00001d052827	correlated with PC1
Zm00001d054044	correlated with PC1
Zm00001d044529	correlation hard to determine
Zm00001d017762	correlation hard to determine
Zm00001d031222	correlation hard to determine
Zm00001d012391	correlation hard to determine
Zm00001d048709	correlation hard to determine
Zm00001d017152	correlation hard to determine
Zm00001d038891	correlation hard to determine
Zm00001d024522	correlation hard to determine
Zm00001d037894	correlated with PC3
Zm00001d014083	correlated with PC3
Zm00001d025055	correlated with PC3
Zm00001d052063	correlated with PC3
Zm00001d044254	correlated with PC3
Zm00001d042922	correlated with PC3
Zm00001d018298	correlated with PC3
Zm00001d037232	correlated with PC3
Zm00001d040888	low expression stable
Zm00001d035409	Structural variant (Dr. Stéphane Nicolas)
Zm00001d020804	high expression stable
Zm00001d008397	high expression unstable
Zm00001d010048	Involved in stress response (Qin et al. 2007)
Zm00001d028514	involved in stomatal closure - stress
Zm00001d014773	involved in stomatal closure - stress

Zm00001d005884	Drought and salt tolerance (Ying et al. 2012)
Zm00001d010752	(Castelletti et al. 2020)
Zm00001d038725	(Castelletti et al. 2020)
Zm00001d043461	(Castelletti et al. 2020)
Zm00001d045735	(Castelletti et al. 2020)
Zm00001d001913	(Castelletti et al. 2020)

Supplemental Table 8: List of TFs in the network. IDs of the entry vectors from the TFome collection are given in the plasmid ID column (Burdo et al. 2014).

B73 v4 ID	Plasmid ID	Name
Zm00001d002799		***ZmHB66***
Zm00001d006195		***ZmHB126***
Zm00001d008817		***ZmNAC47***
Zm00001d015521	**pUT1502**	***ZmGLK24***
Zm00001d017422	**pUT6833**	***ZmHB41***
Zm00001d021268	**pUT6749**	***ZmHB128***
Zm00001d027757	**pUT3138**	***ZmHSF13***
Zm00001d035604	**pUT6642**	***ZmMYB62***
Zm00001d037015		***ZmGLK33***
Zm00001d037737	**pUT6776**	***ZmZAG1***
Zm00001d041780	**pUT3098**	***ZmZHD21***
Zm00001d042864	**pUT6661**	***ZmHB48***
Zm00001d045804	**pUT5895**	***ZmCOL6***
Zm00001d050893	**pUT6157**	***ZmNAC40***
Zm00001d051018	**pUT6018**	***ZmDBB4***

Supplemental Table 9: List of the plasmids used in this work.

Plasmid	Source	Use	Antibiotic
pGB0110m	Maistriaux et al. (2024) derived from pGB0110 (Sarrion-Perdigones et al., 2013)	Backbone for the insertion of promoter upstream of mVenus CDS, insertion of TF in mcs and control with mCherry	Kanamycin
pGB0109	Sarrion-Perdigones et al. (2013)	Internal control for dual-luciferase experiment	Kanamycin
pDONR221™ P1P2	Thermo Fisher (Invitrogen)	Used to create pENTRY vectors through BP reaction	Kanamycin
pGWL7	Karimi et al. (2002)	Used to express firefly luciferase under the control of a chosen promoter	Ampicillin
p2GW7	Karimi et al. (2002)	Used to overexpress TFs with p35S	Ampicillin
pIX-Halo	O'Malley et al. (2016)	Express TF fused to HALO-tag (N-terminal) and 6x His (C-terminal)	Ampicillin

Papers

Theoretical and Applied Genetics (2024) 137:175
https://doi.org/10.1007/s00122-024-04679-w

ORIGINAL ARTICLE

High-dimensional multi-omics measured in controlled conditions are useful for maize platform and field trait predictions

Baber Ali[1] · Bertrand Huguenin-Bizot[2] · Maxime Laurent[3] · François Chaumont[3] · Laurie C. Maistriaux[3] · Stéphane Nicolas[1] · Hervé Duborjal[4] · Claude Welcker[5] · François Tardieu[5] · Tristan Mary-Huard[1] · Laurence Moreau[1] · Alain Charcosset[1] · Daniel Runcie[6] · Renaud Rincent[1]

Received: 20 March 2024 / Accepted: 15 June 2024 / Published online: 3 July 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Key message **Transcriptomics and proteomics information collected on a platform can predict additive and non-additive effects for platform traits and additive effects for field traits.**

Abstract The effects of climate change in the form of drought, heat stress, and irregular seasonal changes threaten global crop production. The ability of multi-omics data, such as transcripts and proteins, to reflect a plant's response to such climatic factors can be capitalized in prediction models to maximize crop improvement. Implementing multi-omics characterization in field evaluations is challenging due to high costs. It is, however, possible to do it on reference genotypes in controlled conditions. Using omics measured on a platform, we tested different multi-omics-based prediction approaches, using a high dimensional linear mixed model (*MegaLMM*) to predict genotypes for platform traits and agronomic field traits in a panel of 244 maize hybrids. We considered two prediction scenarios: in the first one, new hybrids are predicted (CV-NH), and in the second one, partially observed hybrids are predicted (CV-POH). For both scenarios, all hybrids were characterized for omics on the platform. We observed that omics can predict both additive and non-additive genetic effects for the platform traits, resulting in much higher predictive abilities than *GBLUP*. It highlights their efficiency in capturing regulatory processes in relation to growth conditions. For the field traits, we observed that the additive components of omics only slightly improved predictive abilities for predicting new hybrids (CV-NH, model *MegaGAO*) and for predicting partially observed hybrids (CV-POH, model *GAOxW-BLUP*) in comparison to *GBLUP*. We conclude that measuring the omics in the fields would be of considerable interest in predicting productivity if the costs of omics drop significantly.

Introduction

Climate change affects crop production in several ways, such as significant changes in rainfall and weather patterns, increasing temperature, and drought stress (Lee et al. 2023). Such factors threaten global food security, requiring immediate action to better adapt crops to these challenging conditions. The ever-decreasing cost of high-throughput phenotyping and genotyping technologies offers unprecedented opportunities to accelerate genetic gain by enhancing the efficiency of crop breeding programs. Especially in plant breeding, the interest in multi-environmental genomic predictions is constantly increasing thanks to pedo-climatic data availability (Westhues et al. 2022), which can help to model genotype by environment (GxE) interactions leading to greater predictive abilities (Crossa et al. 2022; Heslot et al. 2014; Jarquín et al. 2014; Rincent et al. 2019). Such environmental data, also known as environmental covariates (ECs),

Communicated by Daniela Bustos-Korts.

✉ Renaud Rincent
renaud.rincent@inrae.fr

1 INRAE, CNRS, AgroParisTech, GQE-Le Moulon, Université Paris-Saclay, 91190 Gif-Sur-Yvette, France

2 Laboratoire Reproduction Et Développement Des Plantes, CNRS, ENS de Lyon-46, Allée d'Italie, 69364 Lyon, France

3 Louvain Institute of Biomolecular Science and Technology, UCLouvain, Louvain-La-Neuve, Belgium

4 Limagrain, Limagrain Fields Seeds, Research Centre, 63720 Chappes, France

5 LEPSE, Univ Montpellier, INRAE, Montpellier, France

6 Department of Plant Sciences, University of California Davis, Davis, CA, USA

can be included in different ways in multi-environment prediction models. For instance, the reaction norm approach by Jarquín et al. (2014) models markers and environmental covariates using covariance structures, where their interactions are modeled through a Kronecker product between the genomic and the environmental covariance matrices. This model relies on a linear statistical relationship between genomic markers and the phenotype.

The advances in multi-omics technologies now make it possible to develop high-dimensional transcriptomic, proteomic, and metabolomic profiles that can efficiently capture the influence of environmental conditions on plants. Most studies that used metabolic data from field plants have proposed several approaches to integrate these in prediction models and have exhibited intermediate to high predictive abilities (de Abreu e Lima et al. 2017; Riedelsheimer et al. 2012; Ward et al. 2015), generally competitive with those obtained when using only genomic markers. Multi-omics data (Hu et al. 2021) are still expensive and cannot be made available for each trial of a breeding program. Breeders generally rely on phenotyping platforms to accurately simulate reproducible stressing conditions to study the multi-omics response of plants. Such data are quite helpful in identifying genes and QTLs related to stress responses (Blein-Nicolas et al. 2020; Jiang et al. 2024; Kilasi et al. 2018; Wen et al. 2019). However, their integration in the genomic prediction models of field traits is still quite challenging due to differences in environmental conditions in the field and platform.

Hu et al. (2020) illustrated that metabolomics and transcriptomics collected from the platform and the field are highly correlated between the same tissue types. This evidence suggests that platform omics data can be helpful for field predictions. In Fu et al. (2012), platform transcriptomic data resulted in promising predictive abilities across genetically distant groups. Other studies have combined different platform omics, such as mRNA and sRNA transcriptomics and metabolomics data measured at an early stage, to improve the predictive abilities of field traits in maize (Schrag et al. 2018). Westhues et al. (2017) collected transcriptomic data from plants in highly controlled conditions and reported improving their prediction models with transcriptomic data and genomic information while predicting maize field traits such as dry matter yield and dry matter content.

As omics abundances are influenced by a plant's response to the growth conditions, we can expect them to be particularly efficient in predicting traits measured on the same plants or plants grown in similar conditions. Omics are intermediary phenotypes, so we hypothesize that they may efficiently capture non-additive genetic effects for predicting growth conditions like those in which the omics were measured. However, in contrast to DNA markers such as SNP, the heritability of omics measurements is lower than one. Moreover, Christensen et al. (2021) stressed that omics can capture genetic and non-genetic effects, which can be an issue when the same individuals are characterized for omics and are also predicted. The model proposed by Christensen et al. (2021), GOBLUP, can trim environmental noise and residuals from the omics expressions. Using this approach to remove the non-genetic part of the omics may be a way to increase the predictive ability of field traits with platform omics at the cost of removing the non-additive genetic part (i.e., the genetic part that is not captured by the additive kinship) of the omics.

Integrating multi-faceted data can also be done by fitting a multivariate linear mixed model (MvLMM). MvLMMs utilize correlations between traits to improve predictions. However, these correlations are not robust if there are more predictors than the number of genotypes ($p >> n$) (Bernardo 2010). Further, the phenotypic correlations between traits are generally a poor representation of genetic correlations. Fitting such MvLMMs is quite computationally inefficient because of the number of iteratively estimated covariance components and their inverses, whose burden increases manifold with the increasing number of traits (Zhou and Stephens 2014). Runcie et al. (2021) proposed the Mega-Scale Linear Mixed Model (*MegaLMM*) as one possible solution that decreases the computational burden of MvLMM by re-parameterizing them as Bayesian sparse factor models. Like Bayesian Sparse Factor Analysis of Genetic Covariance Matrices (*BSFG*) (Runcie and Mukherjee 2013), MegaLMM also considers that a few latent factors can explain covariance among many traits. Following such an approach, MegaLMM has exhibited remarkable predictive abilities and runtime improvements over well-established Bayesian methods like Phenix (Dahl et al. 2016) and *MTG2* (Lee and Van der Werf 2016) in wheat by using hyperspectral reflectance data and in *Arabidopsis thaliana* by using gene expression profiles (Runcie et al. 2021). *MegaLMM* is certainly a model of choice for predicting agronomic traits with multi-omics data.

In the present study, we worked on a panel of 244 maize Dent lines crossed to a Flint tester, characterized for ecophysiological traits, transcriptomics, and proteomics on a high-throughput phenotyping platform in two contrasted water treatments. In addition, grain yield was measured for this same panel of hybrids in 25 trials in different locations, years, and treatment combinations. Our primary objective was to assess the ability of multi-omics data obtained from platform experiments to predict both platform and field traits compared to genomic selection models. We worked both on platform and field traits to evaluate the potential of platform omics to predict non-additive effects for traits measured in similar conditions (platform) or in completely different conditions (fields).

Additionally, we aim to evaluate the effectiveness of platform omics in capturing GxE interactions in the fields. It involved a comparative analysis between models incorporating multi-omics data and conventional genomic prediction models relying solely on genomic information. Our particular focus lies in evaluating the potential advantages of employing the high-dimensional multivariate *MegaLMM* model in conjunction with omics data. Furthermore, we also assessed models combining both multi-omics and ECs.

Material and methods

Plant material and experiments

Maize panel and platform experiments

The platform experiments were conducted on the PhenoArch platform (Cabrera-Bosquet et al. 2016). The platform experiments comprised a hybrid panel of 254 maize dent lines crossed to a common flint parent (UH007) and grown under two contrasting water conditions, i.e., well-watered (WW) and water deficient (WD). The detailed experimental and growing conditions are available in Alvarez-Prado et al. (2018) and Blein-Nicolas et al. (2020). Experiments with two contrasting conditions were carried out in Spring 2012 and 2013 and Winter 2013. Plants grown in WW conditions faced a soil water potential of − 0.05 MPa, while the ones grown under WD conditions faced a range from − 0.3 to − 0.6 MPa. Eight ecophysiological traits related to growth and transpiration were measured during these experiments. The traits included early biomass (*Be*), late biomass (*Bl*), early leaf area (*LAe*), late leaf area (*LAl*), water use efficiency (*WUE*), water use (*WU*), stomatal conductance (*gs*), and transpiration rate (*Trate*). For each of these traits, we have up to three experiments within each treatment (Table S1).

Drought-tolerant yielding plants (DROPS) multi-environment trial

Phenology and production traits, including grain yield, were measured on 254 hybrids of the previously mentioned panel at 25 locations from Eastern to Western Europe and an external site in Graneros, Chile. For each experiment, genotypic means were computed for each hybrid using a mixed model with hybrids and replicates as fixed effects and spatial effects and spatially correlated errors as random effects (Millet et al. 2016). The model was fitted using ASReml-R (Butler et al. 2009; R Core Team 2013), and the best linear unbiased estimators (BLUEs) were used for the subsequent analyzes. Like the platform experiments, DROPS (https://cordis.europa.eu/project/id/244374/reporting/en) trials were also conducted in two a priori contrasting water regimes, i.e., WW and WD. ECs expected to affect maize growth at different phenological stages were measured in these trials, while some of them were computed using the APSIM crop growth model (Hammer et al. 2010; Millet et al. 2019). The ECs include longitude and latitude, soil water potential, maximum temperature, night temperature, radiation intercepted, and vapor pressure deficit during early flowering and grain-filling stages. A detailed overview of the environmental covariates is provided by Millet et al. (2019).

Multi-omics data

Genomic data

All lines were genotyped using the 50 K Infinium HD Illumina array (Ganal et al. 2011), a 600 K Axiom Affymetrix array (Unterseer et al. 2014), and more than 500 K markers obtained through genotype by sequencing method (Negro et al. 2019). As a result, we had access to a dense panel of 978,134 markers spread across the maize genome. In addition, we applied a filter on minor allele frequency (MAF) and removed markers with an MAF lower than 0.02, resulting in 842,165 SNP markers to work with.

Proteins

We had access to data on the abundances of 1982 proteins from 254 hybrids grown under two contrasting watering conditions on the PhenoArch platform in the Spring 2012 platform experiment (Blein-Nicolas et al. 2020; Prado et al. 2018). For each hybrid, there were two replicates per condition, resulting in four leaf samples (one per plant) obtained at the pre-flowering stage. Proteins were extracted from the frozen leaves through standard trichloroacetic and acetone solutions-based methods and digested into peptides before mass spectrometry analysis. The proteins were quantified based on either extracted ion currents (XIC) or spectral count (SC). We used normalized and spatially corrected protein abundances obtained from Blein-Nicolas et al. (2020). The broad sense heritabilities of the protein abundance adjusted means over replicates were computed independently for WD and WW treatments (Blein-Nicolas et al. 2020). For the analyzes in this study, we removed proteins with a heritability lower than 0.4 and kept the ones above this threshold. It resulted in the selection of 546 WD proteins and 475 WW proteins for trait predictions. Further details on protein extraction, quantification, spatial correction, and normalization can be found in Blein-Nicolas et al. (2020).

3′RNA seq gene expression analysis

RNA was extracted from mature leaf tissues on the 1524 plants from 254 different genotypes (254 hybrids × 2 treatments × 2 to 3 repetitions) of the spring 2013 platform experiment. The quantity of RNA was determined by spectrophotometry (Nanodrop®, Ozyme), and the quality was evaluated by capillary electrophoresis (Agilent®) using the RNA Integrity Number (RIN). The concentration of RNA varied between 160 and 1000 ng/ml, and the RIN between 6.5 and 7. The next generation sequencing (NGS) libraries were synthesized using the QuantSeq-3' mRNA-Seq Library Prep Kit FWD (Lexogen, Vienna). The quality and the quantity of each library were assessed using microfluidic electrophoresis on a LabChip GX (Perkin Elmer, Waltham, MA). 1,056 samples were selected among these 1,524 samples based on RNA quality (> 200 ng). For each sequencing batch, 1,056 samples were randomized between and within 11 plates of 96 wells according to four factors: genotype, treatment, genetic group origin, sampling date and hour, and line of sampling in a glasshouse using the "designRandomize" R package (Brien, 2018). Accordingly, treatments, genetic structure, and sampling hour/date were well distributed between and within each plate. Then, 96 libraries of each sample in each plate were equimolarly pooled and sequenced on a Next Seq 550 (Illumina, San Diego, USA) using a High-output flow cell according to the manufacturer's recommendations. The resulting reads were 75 base pairs long. The sequences were first trimmed with Cutadapt v1.8 on single read mode, removing adapter sequences and quality trimming (q 10). The minimum length to keep a read was 50. All samples were individually barcoded during the library process. 80% of the samples had more than 4 million reads, and 92% had more than 3 million reads. Alignment of the sequences was done with the software STAR, using reference genome version 4 of B73 maize. The alignment was improved using an annotation file of the genome, allowing us to consider the splicing junctions. Only reads aligning to a single gene were considered. Moreover, only genes with at least one count per million reads in at least 10% of the samples were considered, which resulted in 20,475 and 20,642 transcripts for WD and WW conditions, respectively. In each condition, the transcripts were normalized with the approach of Trimmed Mean of M-values (Robinson et al. 2009) and transformed with the log2-cpm approach with package edgeR (McCarthy et al. 2012; Robinson et al. 2009).

For a given transcript, the last step was to fit a spatial model to compute adjusted means and estimate heritabilities. For this, we used the following model for each treatment (WW and WD):

$$y_{iklmp} = \mu + h_m + R_p + r_k + c_l + g_i + \epsilon_{iklmp}, \quad (1)$$

where y_{iklmp} is the observed value for transcript expression of hybrid i sown in row k, column l, repetition p. h_m is the fixed effect of the hour of sampling m, R_p is the fixed effect of repetition p, r_k and c_l are the random effects of row k and column l, respectively, g_i is the fixed genotypic effect of hybrid i, and ϵ_{iklmp} is the random error effect: $\epsilon_{iklmp} \sim N(0, \boldsymbol{I}\sigma_\epsilon^2)$. Here, $\boldsymbol{I}$ is the identity matrix and σ_ϵ^2 is the random error variance.

The SPaTS package (Rodriguez-Alvarez et al. 2018) within package statgenSTA (van Rossum et al. 2023) was used for this purpose, allowing for P-spline adjustment of the spatial effects. The number of segments used for adjusting the splines was 20 for rows and columns. The model was first fitted with a fixed genotype effect to compute marginal genetic means and then with a random genotype effect (identity matrix as variance/covariance matrix) to estimate broad sense heritabilities, as in Oakey et al. (2006). After the normalization procedure and merging all the phenotypic and omics data, we obtained a list of 244 genotypes with transcript quantitative expressions. For all further analyzes, we filtered the transcripts based on their broad sense heritabilities to keep only those with heritabilities higher than 0.4. The filter resulted in 7051 WD transcripts and 7043 WW transcripts.

Models

Univariate prediction models

We used the following univariate models for platform traits and field trait predictions.

GBLUP: As a reference, we applied a univariate *GBLUP* model for single trait or environment analysis, defined as follows:

$$y_i = \mu + g_i + \epsilon_i, \quad (2)$$

where y_i is the phenotypic value of a trait for hybrid i. μ is the overall mean. g_i is the random genetic value of hybrid i. Random genetic effects are considered to follow a normal distribution, i.e., $\boldsymbol{g} \sim N\left(0, \boldsymbol{K_G}\sigma_g^2\right)$. Here, $\boldsymbol{K_G}$ is the genomic kinship matrix computed as $\boldsymbol{K_G} = \boldsymbol{MM'}/2\sum p_s(1-p_s)$ with $\boldsymbol{M}$ as the matrix of centered marker genotypes and p_s as the frequency of reference allele for SNP s (VanRaden 2008). σ_g^2 is the genetic variance. ϵ_i is the random error effect. Residuals are considered to be independent from $\boldsymbol{g}$ and follow a normal distribution, i.e., $\boldsymbol{\epsilon} \sim N\left(0, \boldsymbol{I}\sigma_\epsilon^2\right)$.

OBLUP and G+OBLUP: Kernel-based methods generally offer an easy solution for multi-omics data integration. We used the following formula to calculate omics-based kernels:

$$\boldsymbol{M}_{f_t} = \frac{\boldsymbol{O}_{ft}\boldsymbol{O}'_{ft}}{n_{ft}}, \tag{3}$$

where $\boldsymbol{M}_{f_t}$ is the resulting similarity matrix for omics type f in treatment t. $\boldsymbol{O}_{ft}$ is the centered and scaled matrix of pretreated abundances of omics type f in treatment t, and n_{ft} is the number of features of the concerned type (proteins or transcripts). All kernel matrices, including the genomic kinship matrix, were scaled to have a sample variance of 1 (average of diagonal close to 1) to avoid biased parameter estimations due to different scales (Forni et al. 2011; Kang et al. 2010).

Firstly, we applied a univariate prediction model, hereafter referred to as *OBLUP*, with multiple random effects corresponding to different transcriptomic and proteomic effects as follows:

$$y_i = \mu + t_{r_i} + t_{w_i} + p_{r_i} + p_{w_i} + \varepsilon_i, \tag{4}$$

where $\boldsymbol{y_i}$, μ, and ε_i are the same as defined earlier for Eq. (2). t_{r_i} is the random WD standard-transcriptomic value for hybrid i; $\boldsymbol{t_r} \sim N\left(0, \boldsymbol{M_{t_r}}\sigma^2_{t_r}\right)$. Similarly, t_{w_i} is the random WW standard-transcriptomic value for hybrid i; $\boldsymbol{t_w} \sim N\left(0, \boldsymbol{M_{t_w}}\sigma^2_{t_w}\right)$. p_{r_i} is the random WD standard-proteomic value for hybrid i; $\boldsymbol{p_r} \sim N\left(0, \boldsymbol{M_{p_r}}\sigma^2_{p_r}\right)$. p_{w_i} is the random WW standard-proteomic value for hybrid i; $\boldsymbol{p_w} \sim N\left(0, \boldsymbol{M_{p_w}}\sigma^2_{p_w}\right)$. $\boldsymbol{M_{t_r}}$, $\boldsymbol{M_{t_w}}$, $\boldsymbol{M_{p_r}}$, and $\boldsymbol{M_{p_w}}$ are the standard-omics-based kernel matrices. All random effects are assumed to be independent.

Afterward, we also extended the model to include a random genetic effect (genomic) as defined in Eq. (2). The extended model hereafter will be referred to as *G+OBLUP* and can be represented as follows:

$$y_i = \mu + g_i + t_{r_i} + t_{w_i} + p_{r_i} + p_{w_i} + \varepsilon_i \tag{5}$$

where $\boldsymbol{y_i}$, μ, g_i, and ε_i are the same as defined in Eq. (2). t_{r_i}, t_{w_i}, p_{r_i}, and p_{w_i} are the same as defined in Eq. (4). Like *GBLUP*, we applied both *OBLUP* and *G+OBLUP* on each platform and field trait one by one. We applied these models using the R package Sommer (Covarrubias-Pazaran 2016).

AOBLUP and G+AOBLUP: We proposed *AOBLUP* and *G+AOBLUP* as extensions of the two previous models, *OBLUP* and *G+OBLUP*. In the first step, we extracted the additive component (BLUPs) of omics features by applying an independent *GBLUP* model to each feature (transcript or protein):

$$O_{m,i} = \mu + u_{m,i} + \varepsilon_{m,i}, \tag{6}$$

where $O_{m,i}$ is the adjusted mean of omics feature m for hybrid i, which can be either a single protein or transcript. $u_{m,i}$ is the random genetic value of genome-wide markers on the omics feature m for hybrid i. The random genetic effects are considered to follow a normal distribution, i.e., $\boldsymbol{u_m} \sim N\left(0, \boldsymbol{K_G}\sigma^2_{u_m}\right)$ with $\boldsymbol{K_G}$ as the genomic kinship matrix and $\sigma^2_{u_m}$ as the genetic variance. $\varepsilon_{m,i}$ is the random error effect independent from $u_{m,i}$ and follows a normal distribution as $\boldsymbol{\varepsilon_m} \sim N\left(0, \boldsymbol{I}\sigma^2_{\varepsilon_m}\right)$ with $\boldsymbol{I}$ as the identity matrix and $\sigma^2_{\varepsilon_m}$ as the residual variance. We also computed Pearson correlations between omics and phenotypic traits and between additive components of omics [predicted by model Eq. (6)] and phenotypic traits to evaluate which omics (standard or additive) are more useful for predicting a trait of interest.

In the second step, we estimated omics-based kernel matrices following the same formula as Eq. (3) but with the genomic BLUPs of omics features (i.e., the additive part of omics features: $\hat{\boldsymbol{u}}_{\boldsymbol{m}}$) obtained in step one. In the third step, we fitted the *AOBLUP* and *G+AOBLUP* models separately by using additive-omics-based covariance matrices, i.e., $\boldsymbol{M}_{\tilde{f}_t} = \hat{\boldsymbol{O}}_{ft}\hat{\boldsymbol{O}}_{ft}\boldsymbol{\prime}/n_{ft}$, instead of standard-omics-based kernel matrices. $\boldsymbol{M}_{\tilde{f}_t}$ is the resulting relationship kernel of additive omic type $\tilde{f}$ (proteins or transcripts) from treatment t (WD or WW). From now tilde (~) will be used to represent effects associated or modeled with additive components of omics. $\hat{\boldsymbol{O}}_{ft}$ is the matrix of genomic BLUPs (referred to as additive) of omics obtained in the first step. The *AOBLUP* model can be represented as follows:

$$y_i = \mu + \tilde{t}_{r_i} + \tilde{t}_{w_i} + \tilde{p}_{r_i} + \tilde{p}_{w_i} + \varepsilon_i, \tag{7}$$

where y_i, μ, and ε_i are the same as defined earlier for Eq. (2). $\tilde{t}_{r_i}$ is the random WD additive-transcriptomic value; $\tilde{\boldsymbol{t}}_{\boldsymbol{r}} \sim N\left(0, \boldsymbol{M_{\tilde{t}_r}}\sigma^2_{\tilde{t}_r}\right)$. Similarly, $\tilde{t}_{w_i}$ is the random WW additive-transcriptomic value; $\tilde{\boldsymbol{t}}_{\boldsymbol{w}} \sim N\left(0, \boldsymbol{M_{\tilde{t}_w}}\sigma^2_{\tilde{t}_w}\right)$. $\tilde{p}_{r_i}$ is the random WD additive-proteomic value; $\tilde{\boldsymbol{p}}_{\boldsymbol{r}} \sim N\left(0, \boldsymbol{M_{\tilde{p}_r}}\sigma^2_{\tilde{p}_r}\right)$. $\tilde{p}_{w_i}$ is the random WW additive-proteomic value; $\tilde{\boldsymbol{p}}_{\boldsymbol{w}} \sim N\left(0, \boldsymbol{M_{\tilde{p}_w}}\sigma^2_{\tilde{p}_w}\right)$. Here, $\boldsymbol{M_{\tilde{t}_r}}$, $\boldsymbol{M_{\tilde{t}_w}}$, $\boldsymbol{M_{\tilde{p}_r}}$, and $\boldsymbol{M_{\tilde{p}_w}}$ are additive-omics-based kernel matrices.

While the *G+AOBLUP* can be represented as follows:

$$y_i = \mu + g_i + \tilde{t}_{r_i} + \tilde{t}_{w_i} + \tilde{p}_{r_i} + \tilde{p}_{w_i} + \varepsilon_i, \tag{8}$$

where y_i, μ, g_i, and ε_i are the same as defined in Eq. (2). $\tilde{t}_{r_i}$, $\tilde{t}_{w_i}$, $\tilde{p}_{r_i}$, and $\tilde{p}_{w_i}$ are the same as defined for the Eq. (7). Like *GBLUP*, we applied these models to all platform and field traits one by one. We applied these models using the R package Sommer (Covarrubias-Pazaran 2016).

Multi-environment prediction models with interactions for field traits prediction

Multi-environment prediction models enable information sharing between different environments, which can help in improving predictive abilities. These models were only used for field trait predictions.

GxE-BLUP, GOxE-BLUP, and GAOxE-BLUP: The model, hereafter referred to as *GxE-BLUP*, integrates a GxE effect as follows:

$$y_{ij} = \mu_j + g_{ij} + \varepsilon_{ij}, \tag{9}$$

with y_{ij} being the adjusted mean of hybrid i in environment j. μ_j is the mean effect of environment j. g_{ij} is the random genetic effect of hybrid i in environment j. ε_{ij} is the residual of hybrid i in environment j. The random effect g_{ij} follows a normal distribution, i.e., $g_{ij} \sim MVR(0, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} = \boldsymbol{K_G} \otimes \boldsymbol{\Sigma_E}$. Here, $\boldsymbol{K_G}$ is the genomic kinship matrix, $\otimes$ represents the Kronecker product, and $\boldsymbol{\Sigma_E}$ is the between-environment genetic covariance matrix with compound symmetry parametrization:

$$\boldsymbol{\Sigma_E} = \begin{bmatrix} \sigma_1^2 & \sigma_2^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_1^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_2^2 & \sigma_1^2 \end{bmatrix}, \tag{10}$$

with σ_1^2 as the main genetic effect variance plus the GxE variance and σ_2^2 is the covariance between the environments. The compound symmetry model considers the same within environment variance and between environments covariance.

We introduce *GOxE-BLUP* and *GAOxE-BLUP* as extensions of *GxE-BLUP* to integrate the omics data. In these two models, we included both the main omic effect and omic x environment interaction effects of omics. The first extension can be written as:

$$y_{ij} = \mu_j + g_{ij} + t_{w_{ij}} + t_{r_{ij}} + p_{w_{ij}} + p_{r_{ij}} + \varepsilon_{ij}, \tag{11}$$

similar to g_{ij}, $t_{w_{ij}}$, $t_{r_{ij}}$, $p_{w_{ij}}$, and $p_{r_{ij}}$ are the standard-omics-based genetic effects of hybrid i in environment j. Like g_{ij}, these effects also follow a multivariate normal distribution as; $t_{r_{ij}} \sim MVR(0, \boldsymbol{\Sigma_{t_r}})$, with $\boldsymbol{\Sigma_{t_r}} = \boldsymbol{M_{t_r}} \otimes \boldsymbol{\Sigma_{E_{tr}}}$, $t_{w_{ij}} \sim MVR(0, \boldsymbol{\Sigma_{t_w}})$, with $\boldsymbol{\Sigma_{t_w}} = \boldsymbol{M_{t_w}} \otimes \boldsymbol{\Sigma_{E_{tw}}}$, $p_{r_{ij}} \sim MVR(0, \boldsymbol{\Sigma_{p_r}})$, with $\boldsymbol{\Sigma_{p_r}} = \boldsymbol{M_{p_r}} \otimes \boldsymbol{\Sigma_{E_{pr}}}$, $p_{w_{ij}} \sim MVR(0, \boldsymbol{\Sigma_{p_w}})$, with $\boldsymbol{\Sigma_{p_w}} = \boldsymbol{M_{p_w}} \otimes \boldsymbol{\Sigma_{E_{pw}}}$. Here, $\boldsymbol{M_{t_r}}$, $\boldsymbol{M_{t_w}}$, $\boldsymbol{M_{p_r}}$, $\boldsymbol{M_{p_w}}$ are the standard-omics-based kinship matrices. $\boldsymbol{\Sigma_{E_{tr}}}$, $\boldsymbol{\Sigma_{E_{tw}}}$, $\boldsymbol{\Sigma_{E_{pr}}}$, and $\boldsymbol{\Sigma_{E_{pw}}}$ are the respective between environment genetic covariance matrices following compound symmetry parameterizations like $\boldsymbol{\Sigma_E}$.

The second extension can be written as:

$$y_{ij} = \mu_j + g_{ij} + \tilde{t}_{r_{ij}} + \tilde{t}_{w_{ij}} + \tilde{p}_{r_{ij}} + \tilde{p}_{w_{ij}} + \varepsilon_{ij}, \tag{12}$$

here, $\tilde{t}_{r_{ij}}$, $\tilde{t}_{w_{ij}}$, $\tilde{p}_{r_{ij}}$, and $\tilde{p}_{w_{ij}}$ are the additive-omics-based genetic effects of hybrid i in environment j. Like before in Eq. (11), these effects also follow a multivariate normal distribution as; $\tilde{t}_{r_{ij}} \sim MVR(0, \boldsymbol{\Sigma_{\tilde{t}_r}})$, with $\boldsymbol{\Sigma_{\tilde{t}_r}} = \boldsymbol{M_{\tilde{t}_r}} \otimes \boldsymbol{\Sigma_{E_{\tilde{t}r}}}$, $\tilde{t}_{w_{ij}} \sim MVR(0, \boldsymbol{\Sigma_{\tilde{t}_w}})$, with $\boldsymbol{\Sigma_{\tilde{t}_w}} = \boldsymbol{M_{\tilde{t}_w}} \otimes \boldsymbol{\Sigma_{E_{\tilde{t}w}}}$, $\tilde{p}_{r_{ij}} \sim MVR(0, \boldsymbol{\Sigma_{\tilde{p}_r}})$, with $\boldsymbol{\Sigma_{\tilde{p}_r}} = \boldsymbol{M_{\tilde{p}_r}} \otimes \boldsymbol{\Sigma_{E_{\tilde{p}r}}}$, $\tilde{p}_{w_{ij}} \sim MVR(0, \boldsymbol{\Sigma_{\tilde{p}_w}})$, with $\boldsymbol{\Sigma_{\tilde{p}_w}} = \boldsymbol{M_{\tilde{p}_w}} \otimes \boldsymbol{\Sigma_{E_{\tilde{p}w}}}$. Here, $\boldsymbol{M_{\tilde{t}_r}}$, $\boldsymbol{M_{\tilde{t}_w}}$, $\boldsymbol{M_{\tilde{p}_r}}$, $\boldsymbol{M_{\tilde{p}_w}}$ are the additive-omics-based kernel matrices. $\boldsymbol{\Sigma_{E_{\tilde{t}r}}}$, $\boldsymbol{\Sigma_{E_{\tilde{t}w}}}$, $\boldsymbol{\Sigma_{E_{\tilde{p}r}}}$, and $\boldsymbol{\Sigma_{E_{\tilde{p}w}}}$ are the respective between environment genetic covariance matrices following compound symmetry parameterizations similar to $\boldsymbol{\Sigma_E}$. We applied these models using the R package Sommer (Covarrubias-Pazaran 2016).

GxW-BLUP, GOxW-BLUP, and GAOxW-BLUP: ECs can also be used to describe environmental conditions that can impact and interact with the genotypes (Jarquín et al. 2014). We fitted the following model to include the effect of standardized ECs on traits (Jarquín et al. 2014):

$$y_{ij} = \mu + g_i + E_j + gw_{ij} + \varepsilon_{ij}, \tag{13}$$

where μ is the overall mean, g_i is the effect of hybrid genotype i and E_j is the main effect of environment j. The term gw_{ij} is the genotype x ECs interaction term, which can be defined as $gw \sim N\left(0, \boldsymbol{K_G} \otimes \boldsymbol{W}\sigma^2_{gw}\right)$, where $\boldsymbol{W}$ is the inter-environmental covariance matrix defined as $\boldsymbol{W} = 1 - \frac{ED}{\max(ED)}$. ED here represents the Euclidean distance between different environments calculated as $ED_{jj'} = \sqrt{\sum_{q=1}^{n} \left(EC_{jq} - EC_{j'q}\right)^2}$, EC_{jq} is the qth EC in the environment j. We followed the same approach as Rincent et al. (2019) for estimating $\boldsymbol{W}$.

Similar to *GOxE-BLUP* and *GAOxE-BLUP*, we extended the *GxW-BLUP* model into *GOxW-BLUP* [Eq. (14)] and *GAOxW-BLUP* [Eq. (15)] to account for standard and additive-omics data, respectively.

$$\begin{aligned} y_{ij} = \mu + g_i + E_j + gw_{ij} + t_{r_i} + t_r w_{ij} + t_{w_i} \\ + t_w w_{ij} + p_{r_i} + p_r w_{ij} + p_{w_i} + p_w w_{ij} + \varepsilon_{ij}, \end{aligned} \tag{14}$$

and

$$\begin{aligned} y_{ij} = \mu + g_i + E_j + gw_{ij} + \tilde{t}_{r_j} + \tilde{t}_r w_{ij} + \tilde{t}_{w_j} \\ + \tilde{t}_w w_{ij} + \tilde{p}_{r_i} + \tilde{p}_r w_{ij} + \tilde{p}_{w_i} + \tilde{p}_w w_{ij} + \varepsilon_{ij}, \end{aligned} \tag{15}$$

where $t_r w_{ij}$, $t_w w_{ij}$, $p_r w_{ij}$, and $p_{wr} w_{ij}$ are the interactions between different standard-omics-based values and ECs, which can be defined as $t_r w \sim N\left(0, \left[\boldsymbol{M_{t_r}}\right] \otimes \boldsymbol{W}\sigma^2_{t_r w}\right)$, $t_w w \sim N\left(0, \left[\boldsymbol{M_{t_w}}\right] \otimes \boldsymbol{W}\sigma^2_{t_w w}\right)$, $p_r w \sim N\left(0, \left[\boldsymbol{M_{p_r}}\right] \otimes \boldsymbol{W}\sigma^2_{p_r w}\right)$

and $p_w w \sim N\left(0, \left[\boldsymbol{M}_{\boldsymbol{p}_w}\right] \otimes \mathbf{W}\sigma^2_{p_w w}\right)$, respectively. While, $\tilde{t}_w w_{ij}$, $\tilde{t}_r w_{ij}$, $\tilde{p}_w w_{ij}$, and $\tilde{p}_r w_{ij}$ are the interactions between different additive-omics-based values and ECs, which can be defined as $\tilde{t}_r w \sim N\left(0, \left[\boldsymbol{M}_{\boldsymbol{t}_r}\right] \otimes \boldsymbol{W}\sigma^2_{\tilde{t}_r w}\right)$, $\tilde{t}_w w \sim N\left(0, \left[\boldsymbol{M}_{\boldsymbol{t}_w}\right] \otimes \boldsymbol{W}\sigma^2_{\tilde{t}_w w}\right)$, $\tilde{p}_r w \sim N\left(0, \left[\boldsymbol{M}_{\tilde{\boldsymbol{p}}_r}\right] \otimes \boldsymbol{W}\sigma^2_{\tilde{p}_r w}\right)$ and $\tilde{p}_w w \sim N\left(0, \left[\boldsymbol{M}_{\tilde{\boldsymbol{p}}_w}\right] \otimes \mathbf{W}\sigma^2_{\tilde{p}_w w}\right)$, respectively. We fitted these models using the R package BGLR (Pérez and de los Campos 2014).

MegaLMM: With MegaLMM, all traits (platform and yield) and the omics are considered jointly in the **Y** matrix, each column of **Y** being a trait or an omic feature. *MegaLMM* utilizes latent factors to model joint effects of all predictors, including fixed, random, and residuals (Runcie et al. 2021). *MegaLMM* decomposes trait variation into two components, the first component is dependent on the *K* latent factors, which model correlations between the *t* traits, and the second component is trait-specific variation uncorrelated between the traits. In other words, *MegaLMM* decomposes the variation of *t* traits into correlated and uncorrelated components. Further details about *MegaLMM* can be found in Runcie et al. (2021). In our case the traits used as input of MegaLMM were grain yield in the different environments, or platform traits, and potentially omics data, depending on the prediction scenario.

The final model implemented in this study can be written as follows:

$$\mathbf{Y} = \mathbf{F}\boldsymbol{\Lambda} + \mathbf{E}, \tag{16}$$

With

$$\forall \boldsymbol{v}, \boldsymbol{F}_{\boldsymbol{v}} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{g} + \boldsymbol{\varepsilon}, \tag{17}$$

where **Y** is the *n x t* matrix for *n* hybrids and *t* traits (can be field traits, platform traits, and omics). **F** is a *n x K* matrix of *n* hybrids and *K* latent factors (*K* is defined by the user). **Λ** is the *K x t* loading matrix whose elements, i.e. λ_{kj}, determines the relation between the trait *k* and the corresponding factor *j*. **E** is a matrix of independent residuals for each trait. The *K* latent factors present in **F**, i.e., F_v for latent factor *v*, are further individually decomposed into fixed effect $\boldsymbol{\beta}$ and random effect $\boldsymbol{g}$, with $\boldsymbol{X}$ and $\mathbf{Z}$, respectively, as their incidence matrices. $\boldsymbol{g}$ is modeled as follows: $\boldsymbol{g} \sim N\left(0, \mathbf{K}_{\mathbf{G}}\sigma^2_g\right)$. $\boldsymbol{\varepsilon}$ is the residual of each independent factor modeled as $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma^2_{\varepsilon})$.

We implemented three versions of *MegaLMM*. The first version contained only platform traits or field grain yield in the **Y** matrix. From now on, we will refer to this model as MegaLMM. The second version contained traits, i.e., platform or field grain yield and standard-omics expressions and abundances. From now on, we will refer to this version as MegaGO. The third version contained traits, i.e., platform or field grain yield and additive-omics expressions and abundances. From now on, we will refer to this version as MegaGAO. We specified 10 latent factors (*K*) for *MegaLMM* models and 150 latent factors (*K*) for models including omics, i.e., *MegaGO* and *MegaGAO*. *MegaLMM*-based models can estimate genetic correlations between all the traits (phenotypic traits and platform omics). These genetic correlations allow *MegaLMM* to use more information than univariate models, that consider each trait independently. The *MegaLMM* model uses Bayesian priors to regularize different effects. We used Automatic Relevance Determination (ARD) (Mbuvha et al. 2020) for the factor loadings matrix. In high-dimensional multi-omics datasets with many features, ARD priors can help identify and emphasize important features, improving predictive ability and model interpretability. They introduce sparsity to the model by setting many predictor coefficients close to zero, aiding in feature selection and preventing overfitting. ARD priors are particularly useful when dealing with collinearity and limited sample sizes, as they enhance data efficiency and model stability. We fitted the *MegaLMM* models using the MegaLMM R-package (Runcie et al. 2021).

Prediction scenarios

We compared two prediction scenarios that plant breeders classically face to evaluate the efficiency of the different models and omics data. The first one corresponds to a classical situation in which breeders predict the performance of hybrids unobserved for the target trait (CV-NH). The second one predicts the performance of partially observed hybrids (CV-POH), similar to sparse testing. For both scenarios, all hybrids (training and test) had omics characterization available (Fig. 1). Moreover, we used the same five-fold and five-repeat divisions of hybrids for all platform traits and grain yield trials for both CV-NH and CV-POH. Comparing the predictive abilities of the different models for platform traits and field traits allows us to evaluate the efficiency of omics data to predict non-additive effects when predicting traits measured in similar conditions (platform traits in a different experiment) or in completely different conditions (field) than those in which the omics were measured (the same platform experiment).

CV-NH for platform traits

In CV-NH, we performed fivefold with 5-repeats cross-validations for all the platform traits (Fig. 1). Meaning that for each iteration of the cross-validation, 20% of the hybrids (test set) were masked for all the platform traits. This scheme was applied with *MegaLMM*, *MegaGO*, and *MegaGAO*. The

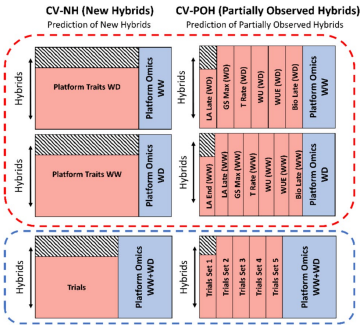


Fig. 1 Description of the fivefold and 5-repeat cross-validation schemes. CV-NH is the scenario where we predict hybrids having no phenotypic observations but only platform omics data. CV-POH is the scenario where we predict hybrids with partially observed phenotypes and complete platform omics data. The red (dashed) square represents the cross-validation settings of platform traits, while the blue (dashed) square represents the cross-validation settings of field grain yield. The pink color represents phenotypes available for training the models (platform traits or field grain yield), and the blue represents platform omics used as predictors. The hatched rectangle represents the hybrids/traits that are to be predicted (colour figure online)

omics information of both the training and test hybrids were used in models capitalizing on omics.

CV-POH for platform traits

In the CV-POH scheme, we first categorized platform traits into trait sets. Each set contained different combinations of years and seasons for the same trait and the same water treatment. We formed six distinct trait sets for the WD platform traits, corresponding to traits *LAe*, *gs*, *Trate*, *WU*, *WUE*, and *Bl*, with three year-season combinations (three experiments) each. We created seven distinct trait sets for the WW platform traits, where the first two sets comprised one and two-year + season combinations of *LAe* and *LAl*, respectively. The remaining 5 comprised three-year + season combinations of 5 platform traits: gs, Trate, *WU*, *WUE*, and *Bl* (Fig. 1). We performed fivefold and 5-repeat cross-validation for each trait set and pooled all predictive abilities. It means that for each iteration of the cross-validation, 20% of the hybrids (test set) were masked for the trait set (one at a time) that had to be predicted. We applied this scheme to *MegaLMM*, *MegaGO*, and *MegaGAO*, with the last two also using omics information as predictors.

CV-NH for field traits (grain yield)

In CV-NH, we applied fivefold cross-validation with 5 repeats, in which we masked the phenotypes of the test hybrids across all the field trials (Fig. 1). We applied this validation scheme with the multienvironment and multivariate models defined above, i.e., *GxE-BLUP*, *GxW-BLUP*, *MegaLMM*, *GOxE-BLUP*, *GOxW-BLUP*, *MegaGO*, *GAOxE-BLUP*, *GAOxW-BLUP*, and *MegaGAO*. We also included WW and WD omics information for both training and test hybrids for the models capitalizing on omics.

CV-POH for field traits (grain yield)

In CV-POH, we first randomly assigned the 25 yield trials in five sets of five trials each, referred to as Trial Sets 1 to 5. We performed five-fold and five-repeat cross-validation for each trial set (Fig. 1). For each cross-validation iteration, 20% of the hybrids were masked for the trial set that had to be predicted. We applied this validation scheme with the same multivariate models mentioned in the previous section. We also included omics information for both training and test hybrids for the models capitalizing on omics.

As univariate models (*GBLUP*, *OBLUP*, *G* + *OBLUP*, *AOBLUP*, and *G* + *AOBLUP*) consider a single platform trait or a grain yield trial at a time, they resulted in the same predictive abilities in CV-NH and in CV-POH. In addition, models that included a GxE component or ECs were only applied to grain yield field trials.

An important point to note is that, unlike grain yield, we independently predicted the two platform trait treatments (WW and WD). We predicted WW platform traits with omics originating from WD treatment and predicted WD platform traits with omics originating from WW treatment (Fig. 1). We implemented this precaution to ensure that the predictive capabilities based on omics solely depended on genetic effects rather than any potential non-genetic effects that might arise from obtaining omics data and traits from the same plants.

Predictive ability was defined as the Pearson correlation between the predictions and the adjusted means of the test hybrids. As this study relies on model comparison, we used gain compared to the reference model (*GBLUP*) as a measure of model performance. We defined gain as the predictive ability of each model minus the ability of univariate *GBLUP*, which only uses information from within the single predicted platform trait or grain yield trial, as was the case for all univariate models.

Table 1 Predictive abilities of different genomic and omics models for predicting WD & WW platform traits and field traits (grain yield) in CV-NH and CV-POH cross-validation schemes

Model type	Model	Platform traits				Field traits (GY)	
		WD		WW		CV-NH	CV-POH
		CV-NH	CV-POH	CV-NH	CV-POH		
Genomic	GBLUP	0.26 ± 0.11		0.24 ± 0.10		0.57 ± 0.09	
	GxE-BLUP	–	–	–	–	0.58 ± 0.09	0.72 ± 0.06
	GxW-BLUP	–	–	–	–	0.58 ± 0.05	0.72 ± 0.08
	MegaLMM	0.25 ± 0.11	0.88 ± 0.03	0.25 ± 0.10	0.89 ± 0.03	0.59 ± 0.09	0.77 ± 0.05
Standard Multi-omics	OBLUP	0.44 ± 0.09		0.42 ± 0.09		0.55 ± 0.10	
	G + OBLUP	0.43 ± 0.09		0.42 ± 0.09		0.54 ± 0.10	
	GOxE-BLUP	–	–	–	–	0.57 ± 0.10	0.72 ± 0.06
	GOxW-BLUP	–	–	–	–	0.59 ± 0.10	0.72 ± 0.06
	MegaGO	0.15 ± 0.14	0.58 ± 0.12	0.13 ± 0.15	0.59 ± 0.09	0.58 ± 0.09	0.61 ± 0.10
Additive Multi-omics	AOBLUP	0.40 ± 0.11		0.40 ± 0.11		0.58 ± 0.10	
	GAOBLUP	0.38 ± 0.12		0.36 ± 0.12		0.58 ± 0.10	
	GAOxE-BLUP	–	–	–	–	0.58 ± 0.09	0.72 ± 0.06
	GAOxW-BLUP	–	–	–	–	0.59 ± 0.09	0.74 ± 0.06
	MegaGAO	0.30 ± 0.13	0.35 ± 0.12	0.30 ± 0.15	0.39 ± 0.12	0.60 ± 0.09	0.60 ± 0.10

Model type represents whether the model has genomic, standard multi-omics, or additive multi-omics information. Univariate models presented were tested only trait by trait for platform and environment by environment for grain yield under five-fold and five-repeat CV, resulting in the same predictive abilities for CV-NH and CV-POH. Standard deviations are computed individually for each platform trait and grain yield, followed by a mean over platform traits or field trials

Results

Univariate and multivariate models for platform trait prediction without transcript and protein information

CV-NH predictions

We predicted platform traits with different univariate and multivariate models under CV-NH (Table 1; Fig. 2). Firstly, in the case of WD traits, the reference univariate *GBLUP* model resulted in a mean predictive ability of 0.26 over all the traits. The multivariate *MegaLMM* model resulted in a predictive ability of 0.25 with negligible difference compared to the reference *GBLUP* model. The results of WW platform traits are similar to the WD platform traits. For instance, the reference *GBLUP* model resulted in a mean predictive ability of 0.24, while the multivariate *MegaLMM* model resulted in 0.25.

CV-POH predictions

In the case of CV-POH (Table 1; Fig. 2) with WD traits, the *MegaLMM* model resulted in a predictive ability of 0.88, 0.62 units higher than the reference *GBLUP*. Similar results were also observed with WW platform traits in CV-POH with *MegaLMM*, resulting in a mean predictive ability of 0.89. In the CV-POH scenario, *MegaLMM* was the best model for predicting platform traits.

Univariate and multivariate models for platform trait prediction with transcript and protein information

CV-NH predictions

We used WW omics as predictors for WD platform traits (Table 1; Fig. 2). The *OBLUP* model applied to WD traits resulted in a mean predictive ability of 0.44, which is 0.18 units higher than the benchmark *GBLUP* model. Adding genomic information in addition to transcripts and proteins, i.e., in the $G + OBLUP$ model, we observed a similar mean predictive ability of 0.43. We observed a slight decrease in predictive abilities when we shifted to additive-omics. The model solely with transcripts and proteins as predictors, i.e., *AOBLUP*, resulted in a mean predictive ability of 0.40. At the same time, the model that also used genomic information ($G + AOBLUP$) resulted in a mean predictive ability of 0.38.

The predictive abilities of models predicting WW platform traits were similar to those of WD platform traits. We used WD omics as predictors for WW platform traits (Table 1; Fig. 2). The OBLUP model resulted in a mean predictive ability of 0.42. Adding genomic information

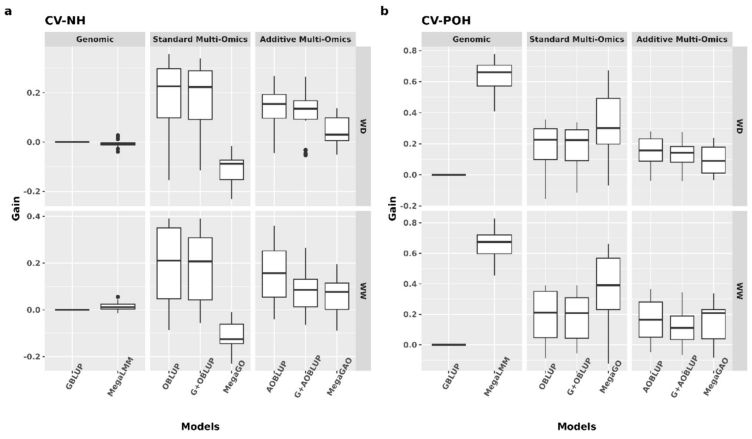


Fig. 2 Boxplots of predictive gains of different models compared to GBLUP in different cross-validation schemes for platform traits. **a** CV-NH Scheme for predicting new hybrids, **b** CV-POH Scheme for predicting partially observed hybrids. Here, WD represents WD platform traits predicted with WW omics, and WW represents WW traits predicted with WD omics. "Genomic" represents the scenarios without transcripts and proteins, "Standard Multi-omics" represents the case where standard transcripts and proteins are used, and "Additive Multi-omics" represents the case where the additive component of transcripts and proteins was used

($G + OBLUP$) did not change the predictive ability. As before, we observed a decrease in mean predictive abilities with additive-omics. For example, the *AOBLUP* model resulted in a predictive ability of 0.40, while $G + AOBLUP$ resulted in an even lower predictive ability of 0.36.

Now, we will compare different multivariate prediction models with omics information for predicting platform traits (Table 1; Fig. 2). In the case of WD traits, *MegaGO*, the model with standard-omics, resulted in a mean predictive ability of 0.15, lower than any other prediction model for the same traits. Shifting to additive-omics with *MegaGAO* resulted in a mean predictive ability of 0.30, higher than *MegaGO* but quite lower than all other univariate models for the WD platform traits.

The performances of multivariate *MegaGO* and *MegaGAO* for predicting WW platform traits (Table 1; Fig. 2) were similar to their respective cases with WD platform traits. For example, both resulted in mean predictive abilities of 0.13 and 0.30, respectively, again falling behind univariate models for the same traits.

CV-POH predictions

Under CV-POH, we tested two models capitalizing on omics information, *MegaGO* and *MegaGAO* (Table 1; Fig. 2), with the former utilizing standard-omics while the latter utilizing additive-omics. The *MegaGO* model applied to WD platform traits resulted in a mean predictive ability of 0.58, indicating a gain of 0.32 units over the reference *GBLUP* model but a loss of 0.30 units compared to *MegaLMM*, the model without omics. Shifting to additive-omics made the case worse. *MegaGAO* resulted in a mean predictive ability of 0.35, comparable to the reference *GBLUP* and lower than *MegaLMM* and *MegaGO*. We also applied *MegaGO* and *MegaGAO* models

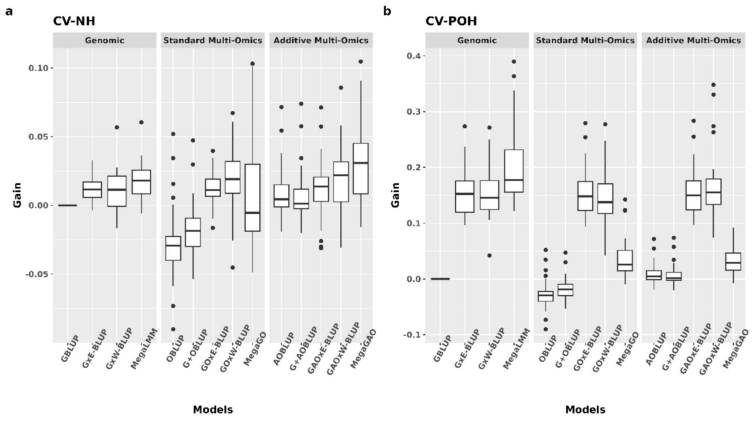


Fig. 3 Boxplots of predictive gains of different models compared to GBLUP in different cross-validation schemes for field grain yield. **a** CV-NH Scheme for predicting new hybrids, **b** CV-POH Scheme for predicting partially observed hybrids, "Genomic" represents the scenarios without transcripts and proteins, "Standard Multi-omics" represents the case where standard transcripts and proteins are used, and "Additive Multi-omics" represents the case where the additive components of transcripts and proteins were used

to WW platform predictions (Table 1; Fig. 2). Both models performed similarly as in the predictions of WD platform traits, resulting in mean predictive abilities of 0.59 and 0.39, respectively.

Univariate and multivariate models for field trait predictions without transcript and protein information

CV-NH predictions

At first, we applied both univariate and multivariate models to predict grain yield without adding multi-omics information (Table 1; Fig. 3). The reference *GBLUP* model resulted in a mean predictive ability of 0.57 over all the folds and repeats of all 25 field trials. The two interactions-based models, i.e., *GxE-BLUP* and *GxW-BLUP*, resulted in similar mean predictive abilities of 0.59 and 0.58, respectively. Similarly, the multivariate *MegaLMM* model also resulted in a mean predictive ability of 0.59. Even though the average predictive abilities of *MegaLMM* appear to be closer to the reference model, there are several trials where important gains were observed, for instance, 0.06 points in the case of *Karlsruhe.2012.OPT* in contrast to the minimum gain equal to 0 in *Campagnola.2012.WD* (Fig. S1).

CV-POH predictions

We tested *GxE-BLUP*, *GxW-BLUP*, and *MegaLMM* in a less challenging CV-POH situation for predicting grain yield (Table 1; Fig. 3). In this cross-validation scenario, *GxE-BLUP* and *GxW-BLUP* models resulted in a similar predictive ability of 0.72, which is 0.15 units higher than *GBLUP*. *MegaLMM* performed better than the reference *GBLUP* and the two interaction-based models, with a mean predictive ability of 0.77. We also observed considerable gains with *MegaLMM* for many environments. For instance, *Craiova.2012.WD* improved by 0.39 units in comparison to GBLUP (Fig. S2).

Univariate and multivariate models for field trait predictions with transcript and protein information

CV-NH predictions

Firstly, we tested the predictive ability of the omics data to predict field traits one by one using univariate models (Table 1; Fig. 3). We observed that adding standard-omics information in the univariate models decreased the predictive abilities compared to *GBLUP* (Table 1; Fig. 4). For example, *OBLUP* and $G + OBLUP$ resulted in mean

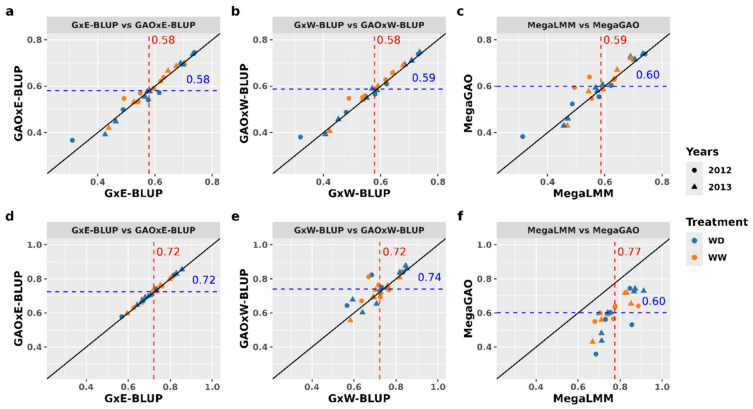


Fig. 4 Comparison of predictions of different additive-omics-based models with their versions with only genomics in CV-NH **a–c** and CV-POH **d–f** for grain yield. The X-axis represents models with genomics, and the Y-axis represents models involving additive-omics. The red dashed line represents the mean predictive ability of the model without omics (x-axis), while the blue dashed line represents the mean predictive ability of the model involving additive-omics (y-axis). The black line represents the diagonal. Colors and the shapes of points refer to the treatment and years of field trials

predictive abilities of 0.55 and 0.54, respectively, lower than the *GBLUP* benchmark model (0.57). Shifting to additive-omics with *AOBLUP* and *G + AOBLUP* models, we observed the same predictive ability of 0.58 with both models. Even though their predictive ability is somewhat higher than *OBLUP* and *G + OBLUP*, it is similar to that of the reference *GBLUP*.

The two interaction-based models with standard-omics, *GOxE-BLUP* and *GOxW-BLUP*, resulted in predictive abilities of 0.57 and 0.59, respectively, similar to their predecessors, *GxE-BLUP* and *GxW-BLUP*. The two interaction-based models with additive-omics, *GAOxE-BLUP* and *GAOxW-BLUP*, resulted in comparable predictive abilities of 0.58 and 0.59, respectively. Even though the performance is similar, *GAOxW-BLUP* resulted in important predictive gains in some trials, as high as 0.09 (*Craiova.2012.WD*), with a maximum loss of –0.03 (*Campagnola.2013.OPT*), as observed in Fig. 4.

The predictive ability of multivariate *MegaGO* (0.58) was also quite similar to the reference *GBLUP* and *MegaLMM* models. However, using additive-omics information, *MegaGAO* showed some improvement over *GBLUP* and *MegaLMM* with a mean predictive ability of 0.60. Even though we could not observe overall important improvements by using additive-omics data in *MegaGAO*, some environments, such as *Craoiva.2012.OPT* and *Campagnola.2012.OPT*, indicated gains as high as 0.10 compared to *MegaLMM*, while the maximum loss was –0.04 in *Campagnola.2013.OPT* (Fig. 4). Overall, in the challenging situation of CV-NH, all the models, including the additive-omics information, exhibited gains relative to their predecessors with standard-omics information, although these gains were small.

CV-POH predictions

The interaction-based models involving standard-omics, *GOxE-BLUP* and *GOxW-BLUP*, resulted in the same predictive abilities (0.72) as *GxE-BLUP* and *GxW-BLUP*, suggesting that omics information was not beneficial in the CV-POH scenario. The additive-omics-based models, *GAOxE-BLUP* and *GAOxW-BLUP*, resulted in mean predictive abilities of 0.72 and 0.74, respectively. Even though the gain with *GAOxW-BLUP* compared to *GxW-BLUP* is small, there are several trials where we observed gains as high as 0.09 (*Craiova.2012.WD*) and a maximum loss of 0.03 (*Campagnola.2013.OPT*) (Fig. 4). Multivariate *MegaGO* with standard-omics information resulted in a mean predictive ability of 0.61, indicating a sharp decrease compared to 0.77 of *MegaLMM* (Table 1; Fig. 3). It suggests that adding

standard-omics in the model was worsening the situation. This situation is also true with the multivariate *MegaGAO* model containing additive-omics information.

Discussion

The interest in utilizing omics data in trait predictions is increasing due to their ability to capture interactions and regulation processes in response to the environment. However, due to the high costs of multi-omics characterization, it is currently impossible for a breeder to collect such data for all field plots each year. It is possible to measure omics in controlled conditions on phenotyping platforms. Even though omics data originating from the platform help detect stress QTLs (Blein-Nicolas et al. 2020), their capability to predict platform and field traits needs further evaluation. Therefore, we evaluated different univariate and multivariate models with or without omics data. Our main objectives were to assess the efficiency of those omics and of a high dimensional multivariate model, *MegaLMM* (Runcie et al. 2021), to predict traits measured in similar conditions (platform), or to predict productivity traits measured in completely different conditions (the fields), than those in which the omics were measured (a platform experiment). For this, we have considered standard univariate models such as *GBLUP*, *OBLUP*, *AOBLUP*, *G* + *OBLUP*, and *G* + *AOBLUP*, predicting each platform trait or each grain yield trial one by one. We also fitted *GxE-BLUP* and *GxW-BLUP* models that were developed to capture GxE in multi-environment settings. We extended these models to include standard-omics, *GOxE-BLUP*, and *GOxW-BLUP*, or additive-omics, *GAOxE-BLUP* and *GAOxW-BLUP*. We have compared the efficiency of the multi-trait models in two different prediction schemes, CV-NH and CV-POH. CV-NH focuses on predicting new hybrids that are not phenotyped for platform or field traits but have multi-omics characterization. CV-POH predicts hybrids partially observed for some platform traits or field trials. It is generally easier to predict based on CV-POH than on CV-NH due to the fact that test hybrids are partially observed for platform traits or field trials in CV-POH.

We first tested and compared models without using the omics information to predict platform traits. In the case of CV-NH, *MegaLMM* performs similarly to *GBLUP* for both WW and WD platform traits. It is consistent with studies showing that multi-trait models are not much better than univariate models if none of the traits is phenotyped for the test set. In the case of CV-POH, *MegaLMM* performed extremely well (0.77) compared to *GBLUP* (0.57, Table 1). In this case, the *MegaLMM* model benefits from the genetic correlations available for the test hybrids because of unmasked platform traits. *MegaLMM* trait loadings matrix (*K x t*) captures the correlated variation in the phenotypic data (Runcie et al. 2021). Genetically correlated traits tend to load onto the same latent factors (*K*) (Fig. S3). *MegaLMM* can effectively use these genetic correlations between traits, especially when information about the test set is available for unmasked trait sets as in the case of CV-POH. We observed high genetic correlations between platform traits as modeled by *MegaLMM* (Fig. S3). This scenario, similar to trait-assisted prediction, has been shown to result in highly accurate predictions when predicted and observed traits are correlated (Calus and Veerkamp 2011; Fernandes et al. 2018; Henderson and Quaas 1976; Jia and Jannink 2012; Pszczola et al. 2013; Robert et al. 2020).

Platform Omics are very efficient in predicting additive and non-additive genetic effects for platform traits

We also tested different standard-omics and additive-omics-based prediction models on platform traits. *OBLUP* and *G* + *OBLUP*, the two univariate prediction models with standard-omics information, were much better than the reference GBLUP model (Table 1; Fig. 2), with a gain in predictive ability as high as 0.36 (*Spring_2013.WD.WU*) and 0.34 (*Spring_2012.WD.WU*), respectively. Meaning that omics are very efficient in making predictions in environmental conditions similar to those in which they were measured. This is of particular interest for prediction approaches based on physiological traits measured in controlled conditions (Bouidghaghen et al. 2023). Note that we ensured that the improved predictive abilities of models *OBLUP* and *G* + *OBLUP* were only due to genetics (and not because of some shared environmental factors) by predicting traits measured on different plants than those used for omics measurements. More precisely, we used omics measured in the WD treatment to predict plants grown under WW treatment and the reverse. It means that the only non-genetic effect that could be involved is the quality of the seed lots, as the same seed lots were used for both treatments.

One key result is that upon fitting the models with additive-omics, we observed an important decrease in predictive abilities with *AOBLUP* and *G* + *AOBLUP* compared to *OBLUP* and *G* + *OBLUP*. It means that the non-additive genetic part of the omics was very useful in predicting platform traits. Another important result to note is that omics measured in similar conditions as the predicted trait (but different water treatments) can capture some non-additive genetic variation like epistasis and responses to growth conditions. We can indeed hypothesize that responses to local conditions are driven by regulatory processes, with epistasis consequently. These findings are also in line with the suggestions of Li et al. (2019) regarding the importance of omics expressions in capturing non-additive genetic variation. The potential ability of *OBLUP* to capture non-additive genetics

works when conditions are similar because the epistatic variation is likely to be more correlated among the platform traits than between platform and field traits.

GxE-BLUP and *GxW-BLUP* models are useful for predicting grain yield

The benchmark *GBLUP* model (Table 1; Fig. 3) resulted in a mean predictive ability of 0.57 for grain yield (25 trials). In both CV-NH and CV-POH validation schemes, multi-environment models *GxE-BLUP* and *GxW-BLUP* models performed better than *GBLUP*. As expected, the gains with models in CV-NH are not as prominent as in CV-POH. These results are comparable to Burgueño et al. (2012), where authors observed a significant improvement in CV-POH but not in CV-NH. Even though our *GxE-BLUP* model performs extremely well, it has one disadvantage due to its underlying assumption of the same variance and covariance among different trials. Furthermore, we expected the model with envirotyping data, i.e., *GxW-BLUP*, to perform better than *GxE-BLUP*, but this was not the case. The potential advantage of using *GxW-BLUP* is to give more weight to trials with similar environmental conditions. Literature indicates that including environmental covariates is not always advantageous. For instance, Monteverde et al. (2019); Nguyen et al. (2023) reported that adding environmental covariates resulted in worse predictions than *GBLUP* in rice while predicting new environments, which could also be attributed to higher inter-environmental differences. In the studies predicting new genotypes, Jarquín et al. (2014); Malosetti et al. (2016), and Rincent et al. (2019) observed little to no improvement with a GxW model.

MegaLMM is an efficient model for predicting field traits in multi-environment scenarios

We also implemented a two-level hierarchical model, *MegaLMM* (Runcie et al. 2021). *MegaLMM* was the best model for both CV-NH and CV-POH prediction schemes (before multi-omics inclusion). The gain was marginal for CV-NH, but *MegaLMM* seemed to be very efficient at dissecting genetic correlations (Fig. S4) between trials and benefit from those, particularly in the CV-POH scenario, with a gain of 0.21 units in comparison to *GBLUP* and 0.05 compared to *GxE-BLUP* and *GxW-BLUP*. In previous studies, *MegaLMM* was shown to result in a high predictive gain compared to *GBLUP* in sparse testing scenarios (Runcie et al. 2021). However, it has not been tested in the more challenging situation of CV-NH before. These results confirm that multi-trait models are very helpful when some information on the predicted set is available, for example, in the case of trait-assisted predictions or sparse trial networks.

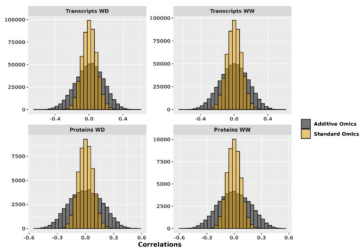


Fig. 5 The distribution of the correlations between omics and grain yield from all 25 trials. The gray color refers to the additive component of omics, and the yellow color refers to the standard-omics. WD represents omics measured in water-deficit and WW represents omics measured in well-watered conditions. The top row corresponds to transcripts and the bottom row corresponds to proteins

The inclusion of platform-based omics can only slightly improve the predictive ability of field traits

We also fitted different prediction models with standard-omics or with the additive part of the omics. We hypothesized that extracting the additive genetic component of omics might be more beneficial for field predictions because it is a way of removing part of the error (Christensen et al. 2021) and because non-additive genetic effects are probably different on the platform than in the field, and so not useful for field trait predictions. It is clearly visible in Fig. 5 that the additive-omics expressions have a higher Pearson correlation with grain yield than the standard-omics, which was not the case for platform traits (Fig. S5). In CV-NH (Fig. 3), none of the models sufficiently benefited from adding the standard-omics. However, with the addition of additive-omics instead of standard-omics, most of the models exhibited small to intermediate gains. For instance, the *AO-BLUP* model appeared to be 0.03 units better than *OBLUP* but similar to *GBLUP*, meaning that the additive part of omics alone can predict similarly to genomic markers. It is likely that additive-omics-based kernel matrices get closer to the genomic kinship matrix, which is evident from the higher correlations between additive-omics kernels and genomic kinship matrix (0.68, 0.61, 0.62, and 0.60 for $\boldsymbol{M}_{\hat{t}_1}$, $\boldsymbol{M}_{\hat{t}_2}$, $\boldsymbol{M}_{\hat{p}_1}$, and $\boldsymbol{M}_{\hat{p}_2}$, respectively) compared to standard-omics kernels and genomic kinship matrix (0.59, 0.56, 0.58, and 0.57 for $\boldsymbol{M}_{t_1}$, $\boldsymbol{M}_{t_2}$, $\boldsymbol{M}_{p_1}$, and $\boldsymbol{M}_{p_2}$, respectively). The *AO-BLUP* model does not include genomic markers in the regression step on field traits but includes them in the first step, where we remove non-additive genetic and residual noise from

the omics. Even though gains with omics information were low on average, predictions of some environments showed important improvements, whereas just a few environments exhibited small losses (Fig. 4).

As the omics capture a response to the environmental conditions, we also hypothesized that transcriptomics and proteomics could be useful to model GxE interactions, which are important in field trials as they can account for a significant part of phenotypic variance, sometimes even more than the genetic variance (Rogers et al. 2021). In our dataset, the amount of variance explained by GxE is similar to the variance explained by G alone (Millet et al. 2016). We proposed two multi-environment models with omics information, and according to our knowledge, our approach to model interactions between omics and environmental covariates in a multi-environment setting is the first implementation in the literature. In the CV-NH scheme, the *GAOxW-BLUP* model gains slightly compared to the same model without omics (Table 1). The gain compared to *GxW-BLUP* is higher in CV-POH, where *GxW-BLUP* predictive ability stands at 0.72 and *GAOxW-BLUP* at 0.74, with some environments gaining as high as 0.14 units (Fig. 4). If we also compare these results to *GxE-BLUP* and *GAOxE-BLUP*, the improvement with *GAOxW-BLUP* might likely have occurred due to the interactions between additive-omics and ECs.

For CV-NH, the best model was the one based on the additive part of omics and *MegaLMM*, i.e., *MegaGAO*, which was expected since trimming off residual variance and noise from the omics increased correlations with the grain yield (Fig. 5). Upon removing the non-additive part from the omics data, we observed a strong increase in the number of features with an absolute correlation above 0.4. Overall, additive-omics are useful in CV-NH because they offer some phenotypic information about test hybrids, which is better than having no information at all. The performance of *MegaGAO* in CV-POH was lower than expected. *MegaGAO* loses in CV-POH despite unmasked field trials probably because their information gets diluted within the high dimensions of omics information, and the genetic correlations between grain yield trials and omics are lower than among grain yield trials. This case of *MegaGAO* in CV-POH was also true for *MegaGO*.

The scientific community's interest in implementing biological information in genomic prediction models to boost genetic gain is consistently increasing (MacLeod et al. 2016). Our work indicates that omics data can be considerably useful for making predictions in conditions similar to those in which they were measured. In contrast, when predicting field traits with platform omics, the gain was only marginal, and only the additive part of the omics was useful. In comparison to genomic selection, the inclusion of the multi-omics profiles in the breeding process could result in a delay due to the necessity to grow plants to measure omics abundances. However, it is possible to collect high-dimensional omics profiles from the nurseries when double haploid (DH) lines undergo rapid seed multiplication through selfing. Our work lays hypothetical grounds that these high-dimensional omics data could potentially be helpful for field trait predictions. We conclude from our study, that measuring the omics in field conditions is necessary to capture non-additive effects and boost predictive ability. It was for instance proposed in a similar way, that near-infrared data measured in the fields could be useful for predicting non additive effects and in particular GxE interaction (Robert et al. 2022a, 2022b).This may also be possible with molecular omics in the near future thanks to the important decrease of cost that are ongoing.

Conclusion

In this study, we integrated omics information collected from the platform to predict field and platform traits using different univariate and multi-environmental models and a mega-scale multivariate model, i.e., *MegaLMM*. We also introduced a novel multi-environmental model to integrate omics and ECs simultaneously (*GOxW-BLUP*). Platform omics data could dramatically increase predictive ability compared to *GBLUP* for traits measured in similar conditions (platform). Still, the gain was marginal for predicting traits in completely different conditions (fields). The main reason was that platform omics were able to predict non-additive effects only for plants grown in similar conditions. In this case, the non-additive component of omics becomes important due to its ability to capture non-additive genetic variations like epistasis related to the response to environmental conditions. We observed that trimming omics data off their non-additive genetic and residual components for field traits can improve field predictions (especially while predicting new hybrids) because such components cannot explain the field phenotypic variation. We also observed that *MegaLMM* could extract information from additive-omics in predicting field traits, particularly for predicting new hybrids, which were only evaluated for omics information. Finally, in the absence of omics, we confirm the very high predictive abilities of the multi-trait model *MegaLMM* in the sparse testing scenario (CV-POH) due to its ability to capitalize on genetic correlations between traits.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-024-04679-w.

Acknowledgements We thank all the collaborators and funding of the PIA Amaizing (ANR-10-BTBR-01) and the DROPS (FP7-244374) projects for the production of these very rich datasets. We are grateful to the INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale Bioinformatics Facility, https://doi.org/10.15454/1.55723906553432 93E12) for providing help and/or computing and/or storage

resources. We thank Melisande Blein-Nicolas, Michel Zivy, and the platform PAPPSO for introducing us to the world of proteomics.

Author contribution BA performed the statistical modeling and analyzes under the supervision of RR, DR, AC, TM-H, and LM. BA wrote the manuscript. RR edited the manuscript. DR supported BA for the MegaLMM analyzes. FC, LCM, SN, FT, CW and HD designed the 3'RNAseq experiment. HD pretreated and mapped the 3' RNAseq data, BH-B and ML treated and analyzed the 3' RNAseq data under the supervision of SN, RR, and FC. All authors approved the submission of the final version of the manuscript.

Funding We thank Ecole Doctorale Agriculture, alimentation, biologie, environnement et santé (ED ABIES) for the funding of B. Ali's PhD. We also thank SPS – Sciences des Plantes de Saclay and ED ABIES for providing mobility funding for B. Ali's research internship in D. Runcie's laboratory at the University of California, Davis. D. Runcie was supported by Agriculture and Food Research Initiative Grants Nos. 2020-67013-30904 and 2018-67015-27957 from the USDA National Institute of Food and Agriculture and by the United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA), Hatch Project 1010469. The 3'RNAseq experiment was funded by INRAE within the Selgen metaprogram. We thank all the collaborators and funding of the PIA Amaizing (ANR-10-BTBR-01) and the DROPS (FP7-244374) projects for the production of these very rich datasets.

Data availability Field grain yield and environmental covariates can be found at: https://doi.org/10.1104/pp.16.00621. Platform phenotypic data are available online using the PHIS information system (Neveu et al. 2019) at http://www.phis.inra.fr/openphis/web/index.php?r=site%2Flogin-as-guest. From this webpage, data are accessible by clicking on Experimental organization > Project > Systems genetics for maize drought tolerance (Amaizing project). For proteomics data, The raw MS output files were submitted to PROTICdb (http://moulon.inra.fr/protic/amaizing) (Ferry-Dumazet et al. 2005; Langella et al. 2007, 2013), the MassIVE database (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) under accession number MSV000085594, and the ProteomeXchange database (http://www.proteomexchange.org) under accession number PXD019804. Detailed information on all peptides and proteins identified in the LC–MS/MS runs, as well as peptide intensities and protein abundances obtained for each sample, are also freely available on PROTICdb at the same URL. The genotyping data can be found at: https://doi.org/10.15454/AEC4BN. Transcriptomics data are available to anyone upon reasonable request. R scripts to run the analysis are available at: https://github.com/rrincent/MegaGO.

Declarations

Conflict of interest The authors declare no conflict of interest or personal relationships that could appear to influence the work reported in this paper.

References

Bernardo R (2010) Breeding for Quantitative Traits in Plants, Vol. 2. Stemma Press, Woodbury, MN, p 390

Blein-Nicolas M, Negro SS, Balliau T, Welcker C, Cabrera-Bosquet L, Nicolas SD, Charcosset A, Zivy M (2020) A systems genetics approach reveals environment-dependent associations between SNPs, protein coexpression, and drought-related traits in maize. Genome Res 30:1593–1604

Bouidghaghen J, Moreau L, Beauchêne K, Chapuis R, Mangel N, Cabrera-Bosquet L, Welcker C, Bogard M, Tardieu F (2023) Robotized indoor phenotyping allows genomic prediction of adaptive traits in the field. Nat Commun 14:6603

Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype× environment interaction using pedigree and dense molecular markers. Crop Sci 52:707–719

Butler D, Cullis B, Gilmour A, Gogel B, Thompson R (2009) ASReml user guide, release 3.0. VSN International, Hemel Hempstead

Cabrera-Bosquet L, Fournier C, Brichet N, Welcker C, Suard B, Tardieu F (2016) High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. New Phytol 212:269–281

Calus MP, Veerkamp RF (2011) Accuracy of multi-trait genomic selection using different methods. Genet Sel Evol 43:1–14

Christensen OF, Börner V, Varona L, Legarra A (2021) Genetic evaluation including intermediate omics features. Genetics 219:iyab130

Covarrubias-Pazaran G (2016) Genome-assisted prediction of quantitative traits using the R package sommer. PLoS ONE 11:e0156744

Crossa J, Montesinos-López O, Pérez-Rodríguez P, Costa-Neto G, Fritsche-Neto R, Ortiz R, Martini JW, Lillemo M, Montesinos-Lopez A, Jarquin D (2022) Genome and environment based prediction models and methods of complex traits incorporating genotype× environment interaction. Methods Mol Biol (clifton, NJ) 2467:245–283

Dahl A, Iotchkova V, Baud A, Johansson Å, Gyllensten U, Soranzo N, Mott R, Kranis A, Marchini J (2016) A multiple-phenotype imputation method for genetic studies. Nat Genet 48:466–472

deAbreueLima F, Westhues M, Cuadros-Inostroza Á, Willmitzer L, Melchinger AE, Nikoloski Z (2017) Metabolic robustness in young roots underpins a predictive model of maize hybrid performance in the field. Plant J 90:319–329

Fernandes SB, Dias KO, Ferreira DF, Brown PJ (2018) Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. Theor Appl Genet 131:747–755

Forni S, Aguilar I, Misztal I (2011) Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet Sel Evol 43:1

Fu J, Falke KC, Thiemann A, Schrag TA, Melchinger AE, Scholten S, Frisch M (2012) Partial least squares regression, support vector machine regression, and transcriptome-based distances for prediction of maize hybrid performance with gene expression data. Theor Appl Genet 124:825–833

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner E-M, Hansen M, Joets J (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS ONE 6:e28334

Hammer GL, van Oosterom E, McLean G, Chapman SC, Broad I, Harland P, Muchow RC (2010) Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. J Exp Bot 61:2185–2202

Henderson C, Quaas R (1976) Multiple trait evaluation using relatives' records. J Anim Sci 43:1188–1197

Heslot N, Akdemir D, Sorrells ME, Jannink J-L (2014) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor Appl Genet 127:463–480

Hu H, Gutierrez-Gonzalez JJ, Liu X, Yeats TH, Garvin DF, Hoekenga OA, Sorrells ME, Gore MA, Jannink J-L (2020) Heritable temporal gene expression patterns correlate with metabolomic seed content in developing hexaploid oat seed. Plant Biotechnol J 18:1211–1222

Hu H, Campbell MT, Yeats TH, Zheng X, Runcie DE, Covarrubias-Pazaran G, Broeckling C, Yao L, Caffe-Treml M, La G (2021) Multi-omics prediction of oat agronomic and seed nutritional traits across environments and in distantly related populations. Theor Appl Genet 134:4043–4054

Jarquín D, Crossa J, Lacaze X, Du Cheyron P, Daucourt J, Lorgeou J, Piraux F, Guerreiro L, Pérez P, Calus M (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127:595–607

Jia Y, Jannink J-L (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192:1513–1522

Jiang W, Liu Y, Zhang C, Pan L, Wang W, Zhao C, Zhao T, Li Y (2024) Identification of major QTLs for drought tolerance in soybean, together with a novel candidate gene, GmUAA6. J Exp Bot 75:1852–1871

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-Y, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42:348–354

Kilasi NL, Singh J, Vallejos CE, Ye C, Jagadish SK, Kusolwa P, Rathinasabapathi B (2018) Heat stress tolerance in rice (*Oryza sativa* L.): Identification of quantitative trait loci and candidate genes for seedling growth under heat stress. Front Plant Sci 9:1578

Lee SH, Van der Werf JH (2016) MTG2: an efficient algorithm for multivariate linear mixed model analysis based on genomic information. Bioinformatics 32:1420–1422

Lee H, Calvin K, Dasgupta D, Krinner G, Mukherji A, Thorne P, Trisos C, Romero J, Aldunce P, Barret K (2023) IPCC, 2023: climate change 2023: synthesis report, summary for policymakers. In: Lee H, Romero J (eds.) Contribution of working groups I, II and III to the 6th assessment report of the intergovernmental panel on climate change [Core Writing Team]. IPCC, Geneva

Li Z, Gao N, Martini JW, Simianer H (2019) Integrating gene expression data into genomic prediction. Front Genet 10:126

MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE, Chamberlain AJ, Schrooten C, Hayes BJ, Goddard ME (2016) Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. BMC Genom 17:144

Malosetti M, Bustos-Korts D, Boer MP, van Eeuwijk FA (2016) Predicting responses in multiple environments: issues in relation to genotype× environment interactions. Crop Sci 56:2210–2222

Mbuvha R, Boulkaibet I, Marwala T (2020) An automatic relevance determination prior bayesian neural network for controlled variable selection. arXiv preprint arXiv:2001.01765

McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucl Acids Res 40:4288–4297

Millet EJ, Welcker C, Kruijer W, Negro S, Coupel-Ledru A, Nicolas SD, Laborde J, Bauland C, Praud S, Ranc N (2016) Genome-wide analysis of yield in Europe: allelic effects vary with drought and heat scenarios. Plant Physiol 172:749–764

Millet EJ, Kruijer W, Coupel-Ledru A, Alvarez Prado S, Cabrera-Bosquet L, Lacube S, Charcosset A, Welcker C, van Eeuwijk F, Tardieu F (2019) Genomic prediction of maize yield across European environmental conditions. Nat Genet 51:952–956

Monteverde E, Gutierrez L, Blanco P, Pérez de Vida F, Rosas JE, Bonnecarrère V, Quero G, McCouch S (2019) Integrating molecular markers and environmental covariates to interpret genotype by environment interaction in rice (*Oryza sativa* L.) grown in subtropical areas. G3: Genes. Genom, Genet 9:1519–1531

Negro SS, Millet EJ, Madur D, Bauland C, Combes V, Welcker C, Tardieu F, Charcosset A, Nicolas SD (2019) Genotyping-by-sequencing and SNP-arrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. BMC Plant Biol 19:1–22

Nguyen VH, Morantte RIZ, Lopena V, Verdeprado H, Murori R, Ndayiragije A, Katiyar SK, Islam MR, Juma RU, Flandez-Galvez H (2023) Multi-environment genomic selection in rice elite breeding lines. Rice 16:7

Oakey H, Verbyla A, Pitchford W, Cullis B, Kuchel H (2006) Joint modeling of additive and non-additive genetic line effects in single field trials. Theor Appl Genet 113:809–819

Pérez P, de los Campos G, (2014) BGLR: a statistical package for whole genome regression and prediction. Genetics 198:483–495

Prado SA, Cabrera-Bosquet L, Grau A, Coupel-Ledru A, Millet EJ, Welcker C, Tardieu F (2018) Phenomics allows identification of genomic regions affecting maize stomatal conductance with conditional effects of water deficit and evaporative demand. Plant, Cell Environ 41:314–326

Pszczola M, Veerkamp R, De Haas Y, Wall E, Strabel T, Calus M (2013) Effect of predictor traits on accuracy of genomic breeding values for feed intake based on a limited cow reference population. Animal 7:1759–1768

R Core Team R (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217–220

Rincent R, Malosetti M, Ababaei B, Touzy G, Mini A, Bogard M, Martre P, Le Gouis J, Van Eeuwijk F (2019) Using crop growth model stress covariates and AMMI decomposition to better predict genotype-by-environment interactions. Theor Appl Genet 132:3399–3411

Robert P, Brault C, Rincent R, Segura V (2022a) Phenomic Selection: A New and Efficient Alternative to Genomic Selection (GS). In: Ahmadi N, Bartholomé J (eds) Genomic Prediction of Complex Traits: Methods and Protocols. Springer US, New York, NY, pp 397–420

Robert P, Goudemand E, Auzanneau J, Oury F-X, Rolland B, Heumez E, Bouchet S, Caillebotte A, Mary-Huard T, Le Gouis J (2022b) Phenomic selection in wheat breeding: prediction of the genotype-by-environment interaction in multi-environment breeding trials. Theor Appl Genet 135:3337–3356

Robert P, Le Gouis J, Consortium B, Rincent R (2020) Combining crop growth modeling with trait-assisted prediction improved the prediction of genotype by environment interactions. Front Plant Sci 11:827

Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140

Rodriguez-Alvarez MX, Boer MP, van Eeuwijk FA, Eilers PH (2018) Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spat Stat 23:52–71

Rogers AR, Dunne JC, Romay C, Bohn M, Buckler ES, Ciampitti IA, Edwards J, Ertl D, Flint-Garcia S, Gore MA, Graham C, Hirsch CN, Hood E, Hooker DC, Knoll J, Lee EC, Lorenz A, Lynch JP, McKay J, Moose SP, Murray SC, Nelson R, Rocheford T, Schnable JC, Schnable PS, Sekhon R, Singh M, Smith M, Springer N, Thelen K, Thomison P, Thompson A, Tuinstra M, Wallace J, Wisser RJ, Xu W, Gilmour AR, Kaeppler SM, De Leon N, Holland JB (2021) The importance of dominance and genotype-by-environment interactions on grain yield variation in a large-scale public cooperative maize experiment. G3 Genes|Genomes|Genetics 11:jkaa050

Runcie DE, Mukherjee S (2013) Dissecting high-dimensional phenotypes with Bayesian sparse factor analysis of genetic covariance matrices. Genetics 194:753–767

Plant Physiology®

Plant Physiology, 2024, **196**, 368–384
https://doi.org/10.1093/plphys/kiae326
Advance access publication 5 June 2024
Research Article

Genetic variability of aquaporin expression in maize: From eQTLs to a MITE insertion regulating *PIP2;5* expression

Laurie C. Maistriaux,[1] Maxime J. Laurent,[1] Linda Jeanguenin,[1,†] Santiago Alvarez Prado,[2,‡] Joseph Nader,[1] Claude Welcker,[2] Alain Charcosset,[3] François Tardieu,[2] Stéphane D. Nicolas,[3,*] and François Chaumont[1,*]

[1]Louvain Institute of Biomolecular Science and Technology, UCLouvain, 1348 Louvain-la-Neuve, Belgium
[2]INRAE, LEPSE, Université de Montpellier, 34060 Montpellier, France
[3]INRAE, CNRS, AgroParisTech, GQE-Le Moulon, Université Paris-Saclay, 91190 Gif-sur-Yvette, France
*Author for correspondence: francois.chaumont@uclouvain.be (F.C.), stephane.nicolas@inrae.fr (S.D.N.)
[†]Present address: Department of Plant and Microbial Biology, Microbial and Plant Genomics Institute, University of Minnesota, St. Paul, MN 55108, USA
[‡]Present address: IICAR-CONICET, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Campo Experimental Villarino S/N, S2125ZAA, Zavalla, Santa Fe, Argentina
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) is François Chaumont (francois.chaumont@uclouvain.be).

Abstract
Plant aquaporins are involved in numerous physiological processes, such as cellular homeostasis, tissue hydraulics, transpiration, and nutrient supply, and are key players of the response to environmental cues. While varying expression patterns of aquaporin genes have been described across organs, developmental stages, and stress conditions, the underlying regulation mechanisms remain elusive. Hence, this work aimed to shed light on the expression variability of 4 plasma membrane intrinsic protein (*PIP*) genes in maize (*Zea mays*) leaves, and its genetic causes, through expression quantitative trait locus (eQTL) mapping across a 252-hybrid diversity panel. Significant genetic variability in *PIP* transcript abundance was observed to different extents depending on the isoforms. The genome-wide association study mapped numerous eQTLs, both local and distant, thus emphasizing the existing natural diversity of *PIP* gene expression across the studied panel and the potential to reveal regulatory actors and mechanisms. One eQTL associated with *PIP2;5* expression variation was characterized. Genomic sequence comparison and in vivo reporter assay attributed, at least partly, the local eQTL to a transposon-containing polymorphism in the *PIP2;5* promoter. This work paves the way to the molecular understanding of *PIP* gene regulation and its possible integration into larger networks regulating physiological and stress adaptation processes.

Introduction

The vital need for water homeostasis places water channels or aquaporins as essential life molecular components taking part in numerous processes at the cell to organism scales. Their particular diversity in plants emphasizes their multiple important roles in the sessile autotrophic lifestyle. Aquaporins are involved in the control of the root and leaf hydraulic conductivity, transpiration, photosynthesis, and osmoregulation. Over the years, a tight and complex regulation of aquaporins belonging to the plasma membrane intrinsic protein (PIP) subfamily acting at multiple levels (abundance, trafficking, gating) was revealed (Chaumont and Tyerman 2014; Maurel et al. 2015).

In maize (*Zea mays*), with a few rare exceptions, expression of most of the identified 13 *PIP* genes is detected both in leaves and roots, although to different extents, and is associated with protein abundance for isoforms for which specific antibodies were available (Hachez et al. 2006, 2008). In both organs, expression varies along the longitudinal development and displays diurnal or circadian variation (Lopez et al. 2003; Hachez et al. 2006, 2008; Caldeira et al. 2014). In leaves, *PIP* expression is generally low at the base and gradually increases in the elongation zone (EZ) to reach a peak where the leaf emerges from the sheath and the leaf surface becomes exposed to the atmosphere, increasing the evaporative demand from the xylem (Hachez et al. 2008). The expression of certain isoforms, such as *PIP1;1* and *PIP1;3*, plateaued or slightly decreases in the mature zone (MZ) of the leaf. In contrast, *PIP1;2* and *PIP2;1*, *PIP2;3*, *PIP2;4*, and *PIP2;5* exhibit a bell-shaped expression pattern, decreasing in the MZ. Transcript levels of *PIP1;4*, *PIP1;6*, *PIP2;2*, and *PIP2;6* remain relatively constant along the leaf longitudinal axis (Hachez et al. 2008). In addition, *PIP* expression in the EZ is quite stable with the time of the day, whereas in the MZ, almost all the *PIP* genes show a similar diurnal expression pattern with high expression in the morning, a drop in levels during the day, and a low level during the night. Together with immunocytochemistry data, these results suggest the importance of a tight regulation of *PIP* gene expression in specific leaf cells to account for water radial movement, in particular in vascular bundles and the mesophyll cells during evapotranspiration (Hachez et al. 2008; Heinen et al. 2009). Furthermore, a role of PIP aquaporins in maize stomatal dynamics was also demonstrated (Heinen et al. 2014; Ding et al. 2022). For instance, *PIP2;5* is expressed in guard cells of the stomatal complexes, and its over- and down-expression results in faster or slower stomatal closure upon abscisic acid treatment, respectively, compared with wild-type plants (Ding et al. 2022).

Received February 28, 2024. Accepted May 2, 2024.
© The Author(s) 2024. Published by Oxford University Press on behalf of American Society of Plant Biologists.
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 reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

While the spatial and dynamic expression patterns of aquaporin genes across organs, developmental stages, and environmental conditions have been extensively described, the existing natural diversity in aquaporin expression among different cultivars or lines within a given species has hardly been assessed and exploited so far, except for rare studies in which the comparison between only a few genetically divergent accessions was driven by peculiar phenotypes (Vandeleur et al. 2009; Sutka et al. 2011; Devi et al. 2016; Khan et al. 2019; Pou et al. 2022; Rishmawi et al. 2023; Vaziriyeganeh et al. 2023). In addition, the molecular determinants and mechanisms responsible for aquaporin gene expression regulation are scarce.

Here, we shed light on the natural genetic variability of *PIP* gene expression in maize leaves among lines selected over decades, via a panel representing the diversity of cultivated maize (a 252-hybrid panel, gathered in the framework of the European Drought-Tolerant Yielding Plants [DROPS] project; Millet et al. 2016). We also explored to what extent the variability is the result of a robust genetic architecture by carrying out a genome-wide association study (GWAS) of the expression of 4 *PIP* genes assessed by reverse transcription quantitative PCR (RT-qPCR). Samples were collected in both EZ and MZ of the leaf, because of their different physiological states and major processes at stake (water uptake–mediated cell growth vs. photosynthesis and transpiration) and the putative differential implications of *PIP* genes in these processes. The 4 *PIP* genes were chosen based on particular interests for their known expression patterns and functional data: *PIP1;1* because of its ubiquity and a priori water channel inactivity (Fetter et al. 2004); *PIP1;3* and *PIP2;2* because of their relative high expression in leaves, including in the stomatal complexes (Hachez et al. 2008; Heinen et al. 2014); and *PIP2;5* because its basal expression in the stomatal complexes affects their closure dynamics (Ding et al. 2022). The resulting mapped eQTLs were mined for molecular determinants regulating *PIP* gene expression, based on eQTL physiognomy, available genomic resources and annotations, and relevant literature. Ultimately, molecular validation of a putative *cis*-acting local eQTL of *PIP2;5*, consisting in a miniature inverted repeat transposable element (MITE)–containing insertion in its promoter, was conducted by transactivation assay in maize cells. This work constitutes a first step toward elucidating the regulatory elements affecting *PIP* gene expression in maize.

Results

PIP gene expression varies across maize genetic groups

PIP1;1, *PIP1;3*, *PIP2;2*, and *PIP2;5* mRNA transcript levels (i.e. referred to as "expression") were analyzed by RT-qPCR (primers in Supplementary Table S1), across a panel of 252 hybrids obtained by crossing 252 dent inbred lines with a common unrelated flint parent (UH007), a usual practice in hybrid crop genetics (Rincent et al. 2014a; Negro et al. 2019; Hu et al. 2022). As all hybrids share a common parent, the variation of mRNA abundance between hybrids depends on the variation of mRNA between lines, if we consider that there is no variation of specific interaction between lines and tester. We tested therefore at each locus if there was significant difference of expression between hybrids derived from lines, which shared a common allele with tester (homozygote), and hybrids from lines, which did not share alleles with tester (heterozygote). If the allele-specific expression (ASE) is defined as the imbalance of expression between 2 alleles at 1 locus (Cleary and Seoighe 2021), we tested if there was or not an ASE at each locus along the genome. In absence of ASE, we expected no difference between homozygote and heterozygote.

Table 1. "Heritability" of the *PIP* gene expression in both leaf zones

Leaf zone	PIP1;1	PIP1;3	PIP2;2	PIP2;5
EZ	0.567	0.821	0.936	0.419
MZ	0.848	0.798	0.887	0.728

It was calculated from a linear mixed model stating Expression ~ Genotype (random) and as $V_g/(V_g+V_e/n)$ where V_g is the variance attributed to the genotype, V_e is the residual variance, and n is the number of replicates.

Relative expression data showed significant weak to moderate correlations between isoforms and between zones for most isoforms (Supplementary Fig. S1). The relative range of expression of the different isoforms (*PIP1;1* > *PIP1;3* and *PIP2;2* > *PIP2;5*) coincided with published results in leaves (Supplementary Fig. S2; Hachez et al. 2008). All traits (combinations of *PIP* isoform × leaf zone, hereafter referred to as *PIPX;X_MZ/EZ*) showed moderate to high heritability values, supporting the relevance of the genetic factor in the observed variations (Table 1). The lower value for *PIP2;5_EZ* may be explained by higher experimental noise due to low expression close to the detection threshold.

Principal component (PC) analysis of *PIP* expression data (Fig. 1) revealed a partial isoform-specific correlation between the different leaf zones. However, the expressions between the different *PIP* genes were weakly correlated. Indeed, *PIP2;5* expression was strongly correlated with the second PC (capturing 19.4% of the variance), while other isoform expressions were generally more strongly associated with the first PC (27.9%), which could reflect the global expression of *PIP* isoforms. The second PC clearly discriminated PH207, B14a, and Mo17 genetic groups in the lower part from B73 and Oh43 groups in the upper part suggesting that *PIP* expression could depend on genetic groups. Furthermore, we revealed some significant differences in *PIP* expression according to genetic groups (Fig. 2). Among various group- and isoform-specific differences, the B73 and Oh43 groups showed remarkably higher *PIP2;5* expression in the MZ, with this trend being conserved with a smaller amplitude in the EZ (Fig. 2). In contrast, *PIP1;1* showed less global variation, especially in the EZ. The marked variability in *PIP2;5* expression confirmed what was highlighted in the PCA, in both the MZ and EZ.

eQTL mapping by GWAS reveals local and distant eQTLs controlling *PIP* expression

A GWAS with 737,202 single-nucleotide polymorphisms (SNPs) on expression of 4 *PIP* genes in EZ and MZ resulted in 3,490 significant associated SNPs (Fig. 3 and Table 2; Supplementary Table S2). We grouped these significant SNPs into 443 eQTLs based on the overlap of their estimated linkage disequilibrium (LD) windows (the so-called LD-win method; Negro et al. 2019). Local eQTLs were defined as containing the *PIP* gene of interest within their borders, while other eQTLs were termed distant eQTLs. Local and distant terms were preferred to *cis* and *trans* terms, which infer a biological mode of action. By using local and distant terms, we avoid making assumptions about their role (Blein-Nicolas et al. 2020). Local eQTLs were found for all traits but one, *PIP1;1_MZ* (circled in black in Figs. 3 and 4). In all but 1 case (*PIP1;1_EZ*), they appeared to be the most significant eQTLs and showed the largest effect (Fig. 4 and Table 2). Interestingly, local eQTLs associated with a given isoform in both leaf zones were not necessarily similar (*PIP1;1* and *PIP1;3*), suggesting that multiple regulatory elements

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

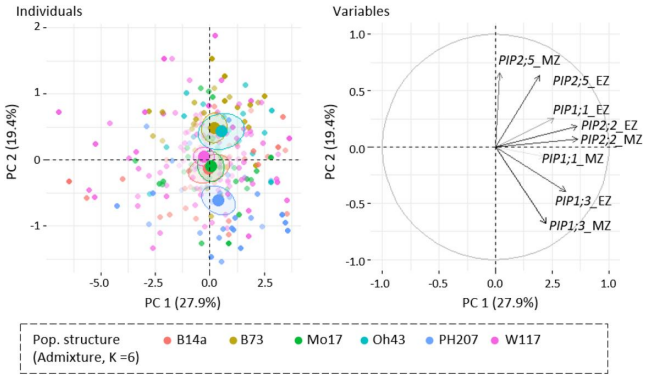


Figure 1. PCA of *PIP* expression data in the leaf MZ and EZ. On the left, the graph of individuals shows the different genotypes, colored according to their belonging to the defined genetic groups (admixture, K = 6), named according to their respective representative lines (B73, stiff stalk; B14a, stiff stalk early; Mo17, Lancaster; Oh43, Lancaster early; PH207, iodent; W117 admixture [W117, NC358, and C103]). Transparency of the points correlates with their quality of representation in the shown space (cos^2 parameter; i.e. plain points are better represented on this particular space than the more transparent ones). On the right, the graph of variables shows the relationships of the explanatory variables, i.e. the 4 *PIP* isoform expression data, with the first and second dimensions. Similarly, the arrow transparency reflects their representation quality. PCA was performed with the PCA function in R and plotted with the factoextra package functions.

with different impacts on *PIP* gene expression and spatial relevance are involved. In other cases, both EZ and MZ local eQTLs coincided either because they were described by the same lead SNP (*PIP2;2*) or because their respective lead SNPs showed the same direction effect, close proximity, and highly similar allele segregation across the panel (*PIP2;5*; Table 2), thus supporting the relevance of major regulation mechanisms acting regardless of the leaf zone. Besides local eQTLs, 436 distant eQTLs were also detected, distributed on all chromosomes (Fig. 3). Unlike local eQTLs, which were sometimes shared between the 2 leaf zones related to a unique *PIP* gene, 295 distant eQTLs were only found in a single leaf zone, suggesting a mostly isoform- and zone-specific regulation of *PIP* genes.

Covariate GWAS highlights new local eQTLs

For *PIP1;3* and *PIP2;5*, we observed many distant eQTLs clustered around their local eQTLs creating a large tail around them (Fig. 3). Since the effect of these local eQTLs was very high in the 2 leaf zones (e.g. $r^2_R = 0.50$ for *PIP2;5*; Table 2), large tails of significant associations around them certainly originated from the existence of moderate LD with the local ones and thus the same causal polymorphisms. Besides, significant associations were sometimes observed with distant eQTLs located far away on the same chromosome. It is noteworthy that such a phenomenon of moderate to high LD between very distant loci was already observed, even despite the correction for kinship, and may be explained by chromosomal rearrangement or strong epistatic interactions (Negro et al. 2019). Ultimately, the presence of eQTL tails around local eQTLs could hide the presence of eQTLs capturing the effect of other causal polymorphisms.

To solve this problem, we used a multilocus approach where we integrated the local eQTL lead SNP as a covariate in the GWAS to exclude all highly linked and redundant loci (referred to as "Cov"; Fig. 5A). This was implemented for traits with arbitrary large local eQTLs [with a $-\log_{10}$(*P*-value) > 12], i.e. *PIP1;3_EZ/MZ* and *PIP2;5_MZ*. As expected, this approach resulted in a large reduction in the number of associated SNP and QTL number (Table 3) and, in most cases, cleared the area around the local eQTL. All lost eQTLs may thus be considered with high caution as they were somehow linked to the local ones. On the other hand, this allowed the emergence of other eQTLs that could safely be considered as independent from the local ones. They showed globally lower $-\log_{10}$(*P*-value) than the initial ones. Interestingly, different local eQTLs were detected in some cases (*PIP1;3_EZ/MZ*), suggesting the presence of multiple causal polymorphisms in the close vicinity of the *PIP* genes (Fig. 5A).

Limitations of eQTL detection due to qPCR technical aspects

Due to the intrinsic influence of RT-qPCR endogenous control genes in the expression measurements, the presence of any important eQTL in the vicinity of these genes was controlled for all the traits. Attention was drawn by an eQTL associated with *PIP1;1* expression in the EZ, colocalizing with the *UBIQUITIN 2* gene (*Ubi2*) targeted by the *Ubi* primer pair. This so-called *ubi*-eQTL was only detected for *PIP1;1* and was the most significant [$-\log_{10}$(*P*-value) = 8.59, lead SNP = AX-91641079] for *PIP1;1_EZ* (Supplementary Fig. S3, A to C). A clear bias was observed for *PIP1;1_EZ* expression data (and, to a lower extent, for *PIP1;1_MZ*) in light of this *ubi*-eQTL (Supplementary Fig. S3D). We then calculated *PIP1;1* expression using only *ACTIN 1* (*Act1*) and *ELONGATION FACTOR 1α* (*Ef1α*) as reference genes, and this variation disappeared (Supplementary Fig. S3E), and, after performing a GWAS, the *ubi*-eQTL disappeared, suggesting that it was an artifact (Supplementary Fig. S3F). However, in

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

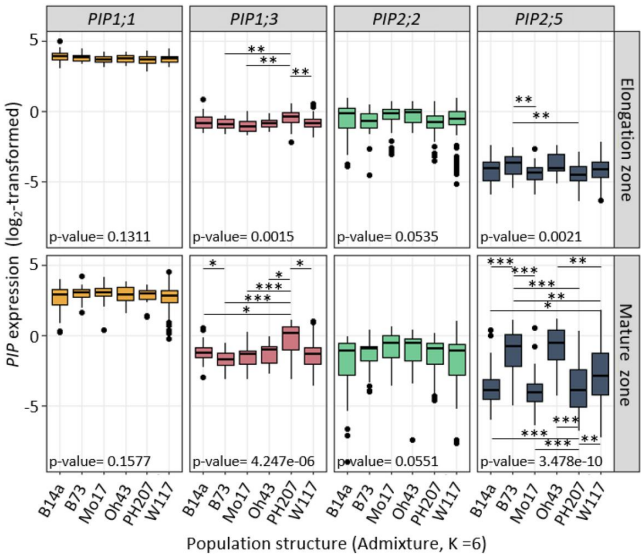


Figure 2. *PIP* gene expression data according to population structure groups. Relative expression data (log_2 transformed) are shown for the 4 *PIP* isoforms and the 2 leaf zones, the EZ (above) and the MZ (below). Data for all lines are classified according to their belonging to the defined population structure groups (admixture, $K = 6$), named according to their respective representative lines. The *P*-values come from the nonparametric one-way Kruskal–Wallis test (kruskal.test R function), while * indicates the significant pairwise comparisons (pairwise.wilcox.test R function, Holm adjustment for multitesting *0.1, **0.05, and ***0.01). Population groups are defined as B73, stiff stalk; B14a, stiff stalk early; Mo17, Lancaster; Oh43, Lancaster early; PH207, iodent; and W117 admixture (W117, NC358, and C103). The center line in the boxplot indicates the relative expression median, points are outliers, and the box upper and lower limits are the third and first quartiles, respectively. The whiskers represent the 1.5× interquartile range.

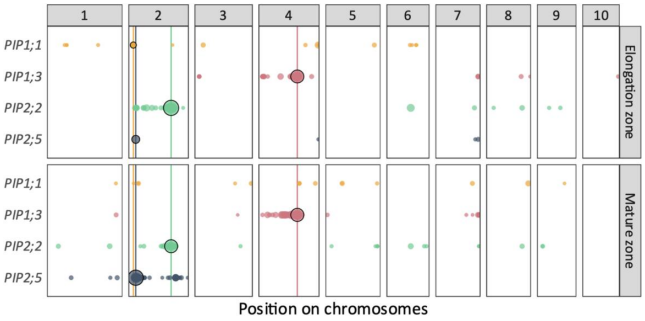


Figure 3. eQTLs detected for the *PIP* expression in the leaf EZ and MZ. Dot diameter is proportional to the $-\log_{10}$(*P*-value). This representation corresponds to a Manhattan plot viewed from the top. Chromosomes are indicated by numbers 1 to 10. Vertical lines indicate the position of the *PIP* genes of interest. The local eQTLs are circled in black. Physical positions of the markers are based on the B73 reference genome version 4. Significance threshold is set at 5.

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

Table 2. eQTLs for the *PIP* genes with the initial GWAS model

Gene	Zone	SNP no.	QTL no.	Local?	Local eQTL								
					Top?	SNP no.	Range[a]	Lead SNP	LogPval[b]	r^2_{LS}	Effect	Position[a]	MAF[d]
PIP1;1	EZ	50	18	Yes	No	11	chr2:19315776 to 19353266	AX-90735145	7.1	0.1	0.1	chr2:19340946	0.34
	MZ	38	15	No	…	…	…	…	…	…	…	…	…
PIP1;3	EZ	754	65	Yes	Yes	78	chr4:157020802 to 157249685	AX-90895950	28.8	0.4	−0.3	chr4:157203881	0.48
	MZ	890	97	Yes	Yes	70	chr4:157020309 to 157249685	S4_153627019	28.8	0.4	−0.6	chr4:157200014	0.33
PIP2;2	EZ	505	75	Yes	Yes	18	chr2:173109023 to 173430699	PZE-102122092	39.9	0.5	0.9	chr2:173382757	0.11
	MZ	167	48	Yes	Yes	9	chr2:173109023 to 173430699	PZE-102122092	28.4	0.4	1.1	chr2:173382757	0.11
PIP2;5	EZ	36	4	Yes	Yes	32	chr2:29138193 to 29445124	S2_27871636	10.5	0.1	0.4	chr2:29279736	0.33
	MZ	936	112	Yes	Yes	128	chr2:29137698 to 29507155	AX-91209412[e]	40.1	0.5	1.4	chr2:29280131	0.32

For each trait, the number of significant SNPs and defined eQTLs is indicated. When a local eQTL was detected, its range, lead SNP, and associated features are presented. It is also indicated whether or not it is the most significant peak detected for this trait ("Top?"). Ellipsis indicates the absence of information.
[a]Physical positions refer to the B73 reference genome version 4.
[b]LogPval stands for −log10(*P*-value).
[c]The r^2_{LS} estimates the percentage of variance explained by the lead SNP of the eQTL calculated as in Sun et al. (2010).
[d]MAF indicates the minor allele frequency of the given marker among the panel.
[e]Note that 3 redundant SNPs were detected as the most significant for the local *PIP2;5* eQTL (AX-91209412, AX-90737936, and AX-91514455). The first one, in terms of physical position (and relative position to *PIP2;5*), was chosen as the lead SNP.

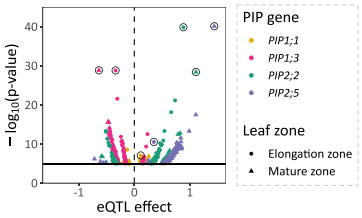


Figure 4. eQTL effect. Significance and effect of detected eQTLs, based on their representative lead SNP. Significance is represented as $-\log_{10}$(*P*-value), with a threshold set at 5. The eQTL effect was assessed using the B73 allele as reference. Local eQTLs are circled in black.

order to keep the expression data consistent for the other *PIP* genes, we took into account this experimental bias for both *PIP1;1* data sets by setting the lead SNP of this *ubi*-eQTL as covariate into the GWAS model leading to a new collection of eQTL (the so-called *ubi*-adjusted eQTLs; Fig. 5B). Interestingly, this new GWAS model allowed the recovering of a weak local eQTLs for *PIP1;1_MZ* that was previously the only data set not harboring one. The initial *PIP1;1_EZ* local eQTL remained, although slightly less significant [−log10(*P*-value) = 6.66 vs. 7.10]. It is noteworthy that no eQTL was detected for any other data sets around *ubiquitin* 2 and no noticeable eQTL was detected around the 2 other endogenous control genes (*Act1* and *Ef1α*; Supplementary Fig. S3).

Moreover, we investigated the presence of polymorphisms among the sequences targeted by the qPCR primers and evaluated the impact of such pinned polymorphisms on the eQTL detection (Supplementary Appendix S1). Three genotyped SNPs were identified in the primer pair targeting *PIP1;1*, but none of them seems to have an impact on the local eQTL detected (Supplementary Appendix S1). Regarding *PIP2;2*, genome and targeted sequence alignment revealed a very different sequence at the expected target site of the forward primer, and the local eQTL detected (identical lead SNP for MZ and EZ data sets) was strongly associated with the repartition of the alleles at this forward primer location. As the unfavorable allele concerned only 29 lines, we discarded them from the *PIP2;2* analysis, and only the 220 remaining ones, harboring the favorable allele, were used for the GWAS (3 line genotypes as heterozygous at this location were also discarded; Fig. 5C). For *PIP1;3*, 1 genotyped SNP within the forward primer was associated with the local eQTL obtained in the EZ, but not with the local eQTL detected in the MZ. However, multiple other causal polymorphisms seem to exist to explain the local eQTL in the EZ (Supplementary Appendix S1). Finally, no polymorphism was identified among the regions targeted by the RT-qPCR primers of the *PIP2;5* gene, either in the marker collection or when comparing genomes (Supplementary Appendix S1).

Final eQTL set identifies a *PIP2;5* local eQTL

We ended up with a final set of eQTLs (Table 3; Supplementary Tables S3 and S4). In the cases where covariate GWAS was implemented, eQTLs coming from the initial GWAS but lost during the covariate GWAS were still conserved because a part of them may still reveal biologically relevant linkage between distant loci (recent relocation, transposition, co-evolution, epistasis). They should, however, be considered with the greatest caution. As for the initial GWAS results, we did not notice any striking colocalization between distant eQTLs detected for different *PIP* isoforms that could have pointed out at potential shared regulators.

Both local *PIP2;5* eQTLs detected in EZ and MZ were considered as one, as their effects worked in the same direction and the lead SNPs showed close proximity (359 bp) and nearly identical allele segregation in the 252 lines of the panel. Therefore, the lead SNP of the *PIP2;5_MZ* data set (AX-91209412) was used as reference.

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

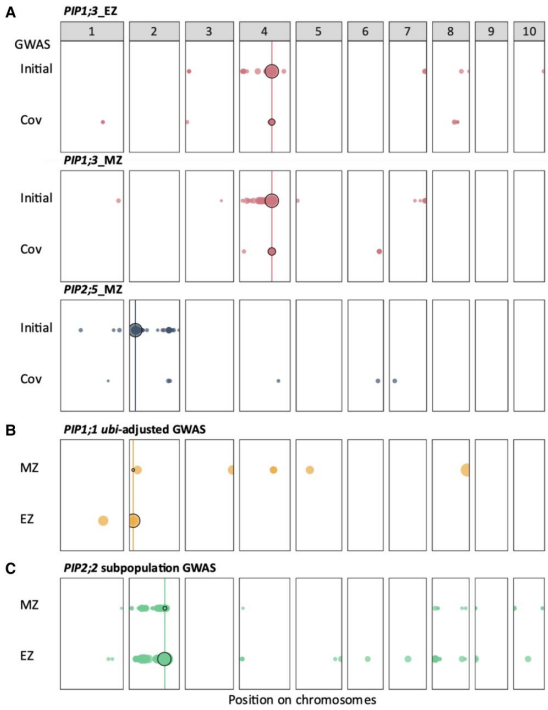


Figure 5. Covariate GWAS. **A)** When large local eQTL peaks were detected (arbitrary cutoff at $-\log_{10}$(P-value) > 12), covariate GWAS was performed by integrating the local lead SNP as additional covariate in the model. Both initial and covariate ("Cov") GWAS are shown for the concerned data sets. **B)** Ubi-adjusted GWAS for *PIP1;1* traits. **C)** GWAS on subpopulation for *PIP2;2* (220 lines), due to the detection of a low conservation at the location of the expected forward primer RT-qPCR hybridization. Dot diameter is proportional to the −log10(P-value). Vertical lines indicate the position of the *PIP* genes of interest, thus defining the local eQTLs circled in black. Physical positions of the markers are based on the B73 reference genome version 4. Significance threshold is set at 5.

Alleles at this position defined 2 distinct haplotypes in the population, extending over a large region (Fig. 6). Among available reference genomes at the time of the experiments (B73, PH207, and B104, matching the panel), B73 and PH207 were used as representatives of the first and second haplotypes, respectively. The B73 allele was associated with a positive effect on *PIP2;5* expression. Genomic sequence comparison showed a high identity of the *PIP2;5* coding sequence (98.72%). It is noteworthy that a point mutation close (7 bp upstream) to the 3′ splice site of the last intron was mapped and identified as putatively associated with splicing variation in the Ensembl Plants database. Collinearity between the genome sequences was observed up to ~30 kb upstream of *PIP2;5*, allowing the sequence alignment and identification of multiple indels, but was lost afterward (Supplementary Fig. S4). A large (~6 kb) indel polymorphism was present in the B73 sequence (absent from PH207), ~11.5 kb upstream of *PIP2;5*, within the next upstream gene (GRMZM2G178681–Zm00001d003007) annotated as belonging to the pentatricopeptide repeat (PPR) superfamily protein (or in the intergenic region, depending on the reference genome versions). Several indels, closer to the *PIP2;5* gene, were present only in the PH207 sequence and absent from B73 (Fig. 6E). One indel was also identified in the *PIP2;5* gene itself, within the last intron (present in PH207, absent from B73).

The presence of a 457 bp indel located ~430 bp upstream of the *PIP2;5* transcription starting site in PH207 was of particular interest. The major part of this indel (303 bp) was identified as a

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

Table 3. Final set of eQTLs detected for 4 AQP genes with different GWAS models with a focus on local eQTLs

Gene	Zone	GWAS	SNP no.	QTL no.	Local?	Top?	Local eQTL SNP no.	Range[a]	Lead SNP	LogPval[b]	r^2_{LK}	Effect	Position[a]	MAF[c]
PIP1;1	EZ	ubi-adj	13	3	Yes	Yes	5	chr2:19315776 to 19353213	AX-90735145	6.66	0.12	0.10	chr2:19340946	0.34
	MZ	ubi-adj	17	7	Yes	No	1	chr2:19329871 to 19354321	S2_18132982	5.02	0.10	0.23	chr2:19342082	0.39
PIP1;3	EZ	Initial	754	65	Yes	Yes	78	chr4:157020802 to 157249685	AX-90895950	28.83	0.39	−0.34	chr4:157203881	0.48
		Cov	31	14	Yes	Yes	5	chr4:157179459 to 157250577	AX-91439361	7.71	0.2	−0.17	chr4:157117582	0.38
	MZ	Initial	889	97	Yes	Yes	70	chr4:157020809 to 157249685	S4_153627019	28.84	0.39	−0.64	chr4:157200014	0.34
		Cov	148	33	Yes	No	29	chr4:157179459 to 157250577	S4_153627596	9.77	0.2	−0.35	chr4:157200591	0.13
PIP2;2	EZ	Subpop	1673	161	Yes	..	132	chr2:167619354 to 173427207	AX-90774142	28.46	0.4	0.52	chr2:170958393	0.06
	MZ	Subpop	475	82	Yes	No	1	chr2:173330947 to 173427207	S2_168551532	5.23	0.08	0.37	chr2:173379237	0.08
PIP2;5	EZ	Initial	36	4	Yes	Yes	32	chr2:29138193 to 29445124	S2_2787686	10.54	0.2	0.35	chr2:29279136	0.33
	MZ	Initial	935	112	Yes	Yes	128	chr2:29137698 to 29507155	AX-91209412	40.11	0.50	1.44	chr2:29280131	0.32
		Cov	22	6	No	...	...	...	...	...	...	...	...	...

For each trait, the number of significant SNPs and defined eQTLs is indicated including the GWAS model used. When detected, characteristics of the local eQTL are presented. It is indicated whether or not it is the most significant peak detected for this trait ("Top?"). Ellipsis indicates the absence of information.
[a]Physical positions refer to the B73 reference genome version 4.
[b]LogPval stands for $-\log_{10}$(*P*-value).
[c]The r^2_{LK} estimates the percentage of variance explained by the lead SNP of the eQTL calculated as in Sun et al. (2010).
[d]MAF indicates the minor allele frequency of the given marker among the panel.

putative MITE transposon using a MITE-finder blast tool (highest score *E*-value = e^{-166}; P-MITE; Chen et al. 2014). Blast results suggested that this MITE belongs to the maize *Heartbreaker* (*Hbr*) family, characterized by particular terminal inverted repeats (TIR), a conserved length (~310 bp), and variable target site duplications (Zhang et al. 2000; Supplementary Fig. S5).

The presence/absence of this particular MITE-containing indel was assessed in 10 lines (B73, Oh43, B104, F98902, and EP52 for the B73 haplotype and PH207, A374, B100, F894, and PHK76 for the PH207 haplotype) chosen for maximizing the admixture group diversity by specific amplifications of *PIP2;5* promoter region (Fig. 7). Bands of different sizes depending on the haplotypes were observed supporting a widespread conservation and a good correlation with the unfavorable, PH207-like, haplotype of the eQTL. Only B100 did not coincide, as it did not harbor this MITE insertion, but yet the second smaller, further upstream, indel included in the targeted promoter region, responsible for the intermediate size band in the upper panel. Sequencing revealed that the MITE-containing indel was perfectly identical in the PH207, A374, F894, and PHK76 genotypes. Table 4 gathers the conservation patterns for the polymorphisms judged as the most promising to explain this local *PIP2;5* eQTL, obtained from sequenced genomes or the manual sequencing from gDNA samples. We considered particularly the MITE-containing indel located the closest to *PIP2;5* and the larger ~6 kb indel located further away. We also looked at the conservation of the indel located in the last *PIP2;5* intron as well as the putative splice site variant SNP mapped in Ensembl Plants. Altogether, the MITE-containing indel (except for B100) and the ~6 kb indel (at least the ~3.5 kb part of it that was conserved in Oh43 and UH007 as well) showed the best coincidence with the eQTL haplotype.

MITE-containing indel affects the promoter activity

The MITE-containing indel stood as a promising causal candidate for the local eQTL, due to its strategic position in the promoter and the recurrence of demonstrated MITE-associated QTLs (Magalhaes et al. 2007; Salvi et al. 2007; Hou et al. 2012; Zerjal et al. 2012; Castelletti et al. 2014; Mao et al. 2015). In order to demonstrate its effect on the promoter activity, we compared the *PIP2;5* promoter (*pPIP2;5*) activity of both B73 and PH207 using a dual fluorescence reporter assay. The tested promoter sequence was defined from *PIP2;5* start codon (nonincluded) up to the next upstream annotated gene, meaning a 1,884 and 2,600 bp region for B73 and PH207, respectively (similar to the blue arrows-targeted region in Fig. 8). Promoter activity was monitored with mVenus (*pPIP2;5:mVenus*) as a reporter, mCherry (*CaMV p35S:mCherry*) being used as a normalizer (polycistronic vectors). Both fluorescence intensities were recorded from confocal microscopy images coming from maize leaf cells transiently transformed by biolistic particle delivery.

In order to remain as close as possible to the physiological reality, and because *PIP* gene expression is known to be sensitive to environmental conditions (diurnal cycle, light, humidity, etc.), the entire plants were used for the bombardment and kept complete and alive in their usual growing conditions until the signal observation (~40 h), unlike usual protocol using cut leaf pieces conserved on solid media in the dark (Chevalier et al. 2014). Thus, for practical reasons, only young plants were used for transformation (~12 d, usually 4th fully emerged leaf). Demonstration of the MITE-containing polymorphism effect was carried out by synthetically removing and adding the whole MITE-containing indel from PH207 and to B73 *PIP2;5* promoters, respectively, by triple PCR. This approach allowed

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

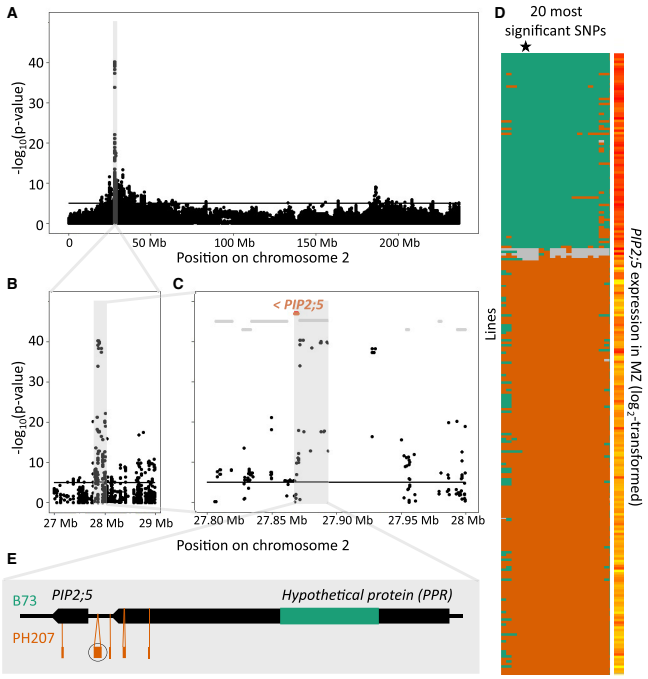


Figure 6. *PIP2;5* local eQTL. **A to C)** Manhattan plots showing the chromosome 2 **A)** and successive closer views of the local peak for *PIP2;5_MZ* (regardless of eQTL range definition) **B and C)**. Vertical axes represent the $-\log_{10}$(*P*-value). *PIP2;5* gene is highlighted in orange. Other genes are shown in gray. The black horizontal line represents the significance threshold of 5. **D)** Two haplotypes are defined by the alleles at the lead SNPs of the local eQTL. Lines are shown vertically, while the 20 most significant SNPs are displayed horizontally (labels are not visible due to the small scale), ordered according to their physical position (they cover a 235 kb range). The black star indicates the lead SNP (AX-91209412). Genotyping data are shown for the 20 most significant SNPs, colored in green for the B73 allele and orange for the other allele. The rare heterozygous alleles determined by the genotyping are colored in gray. *PIP2;5* expression data in the MZ are shown on the right, colored in a yellow (low) to red (high) gradient. **E)** Schematic representation of the locus containing *PIP2;5* and the next upstream annotated gene[4], covering a region of ~25 kb. B73- and PH207-specific large indel polymorphisms (>50 bp) are represented in green and orange depending on the lines they are present in, respectively. The PH207 indel circled in black encodes a MITE transposon. Arrowheads indicate the 5′-3′ gene orientation.

distinguishing the effect of the "background promoting sequence" (B73 or PH207) and the MITE-containing indel itself (presence or absence; Fig. 8). We observed that the Venus/mCherry ratio was lower when the MITE-containing indel sequence was present (either inserted in *pPIP2;5* from B73 [B73 p*PIP2;5* MITE+] or naturally present in PH207 p*PIP2;5*) compared with the *pPIP2;5* without the MITE-containing indel sequence from both backgrounds (Fig. 8B). Statistical analysis revealed the significance of the MITE-containing indel ($P = 0.0271^*$) but not the background sequence on the measured promoter activity. The interaction of MITE-containing indel presence and background sequence was tested as nonsignificant. This demonstrated the effectiveness of the MITE-containing indel on the *PIP2;5* promoter activity, independently of the background sequence, and the effect (without MITE-containing indel > with MITE-containing indel) agrees with the eQTL effect (B73 > PH207).

Discussion

In this work, we investigated the expression diversity of 4 *PIP* aquaporin genes at a large scale and mapped numerous eQTLs, both local and distant. They constitute valuable starting points for the identification and characterization of potential regulatory

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

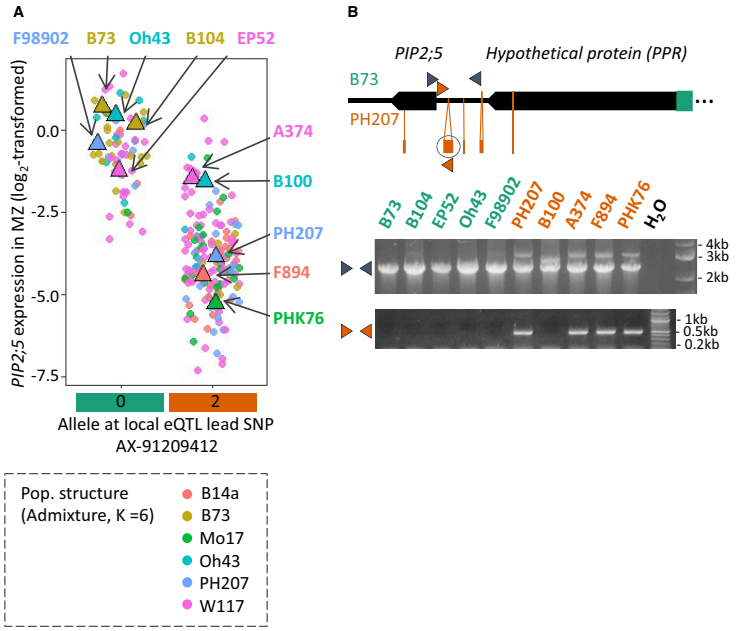


Figure 7. MITE-containing indel conservation. Five lines were selected among each subpopulation defined by the allele at the local eQTL to look for the conservation of the MITE-containing indel. **A)** Selected lines are highlighted on a graph representing *PIP2;5* expression on the MZ according to the allele at the lead SNP of the local eQTL. **B)** PCR amplifications using as template gDNA from the chosen hybrids (inbred line × UH007). Middle panel, amplification of the *PIP2;5* promoter region (primers shown by blue arrows—top panel) allowing the identification of 2 main alleles associated with the B73 and B104 (lower band) or PH207 (upper band) reference genomes, respectively. Lower panel, amplification figuring the presence of the MITE-containing indel (circled in black) thanks to both an internal and an external primer (shown by orange arrows).

elements and factors for *PIP* gene expression, illustrated here by the identification of a MITE transposon polymorphism in the *PIP2;5* promoter.

Maize is ideally suited for GWAS, thanks to its important genetic diversity and rapid LD decay (Buckler et al. 2006). However, the high significance of some local eQTLs was difficult to handle as it spread rapidly over neighboring regions as soon as some genetic linkage was still present. In some cases, we observed a large zone of pyramiding eQTLs spreading over a large portion of the chromosome. Since the strongest associated SNP explained a large part of the variation of *PIP* expression (Table 1), most additional local eQTLs probably originated from SNPs that were in LD with the main causal polymorphism. However, these eQTLs blurred the detection of some other local eQTLs that were linked to other causal polymorphisms. How to distinguish local eQTLs that capture different causal polymorphisms acting independently on gene expression and how to merge eQTLs that are in LD with each other were important questions. Indeed, the genetic determinism of the observed gene expression variation was likely associated with far fewer loci than implied by this high number of eQTLs, so it was necessary, for biological and practical reasons, to disentangle those problematic regions. Therefore, to handle those "local eQTL flooding zones," we implemented a multilocus model the so-called covariate approach (Segura et al. 2012). It allowed us to remove eQTLs that captured the same causal polymorphisms and bring to light new ones that are independent from the first one. Since genetic determinism of gene expression is likely oligogenic, we believed that the covariate approach may be valuable in larger-scale studies as well to point, for instance, at multiple local eQTLs, even if more stringent eQTL selection criteria are applied.

Variability of *PIP* gene regulation

The diversity panel and the eQTL mapping methodology applied in this work revealed that (i) a significant variability in *PIP* gene expression exists, even across economically and agronomically

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

Table 4. Structural polymorphisms around *PIP2;5* gene

Accession	PIP2;5 expression in MZ	Local eQTL lead SNP AX-91209412	MITE indel (457 bp)	Intron indel (235 bp)	Large indel (5,915 kb)	Splice variant (position: 2:27869387)	Source
PHK76	−5.31	0	+	+	–	+	Man
NC358	−5.16	0	+	–	–	+	Man
MS71	−4.66	0	+	+	–	+	Ref
F894	−4.39	0	+	+	–	+	Man
PH207	−3.89	0	+	+	–	+	Ref/man
Mo17	−3.51	0	+	+	–	+	Ref
B100	−1.59	0	–	–	–	+	Man
A374	−1.45	0	+	+	–	+	Man
B97	−1.20	1	+	+	–	+	Ref
EP52	−1.26	2	–	+	+	+	Man
F98902	−0.39	2	–	–	+	–	Man
B104	0.24	2	–	–	+	–	Ref/man
Oh43	0.43	2	–	–	Partial	+	Ref/man
B73	0.70	2	–	–	+	–	Ref
UH007	…	2	–	–	Partial	…	Man

Alleles at polymorphisms are indicated for 14 different lines of the panel and the common parent UH007 as well. Plus (+) means "presence," while minus (–) means "absence." Ellipsis indicates the absence of information. Source indicates whether the information comes from the sequenced genomes (ref) or the manual sequencing (man) from cDNA samples (hybrids). *PIP2;5* relative expression in the MZ is shown, and accessions have been ordered following their *PIP2;5* expression level in MZ.

relevant elite lines, and (ii) this variability is primarily associated with local eQTLs in the observed conditions. Interestingly, the patterns of variation were different among the studied isoforms: most of them showed a rather homogeneous variation irrespective of the population structure, while *PIP2;5* showed a clear interheterotic group variation (B73 and Oh43 > PH207 and B14a and Mo17). This major difference in expression, whose important genotypic foundation has led to the mapping of the cognate major local eQTL, was likely the result of an unbalanced allelic composition at the *PIP2;5* locus and its surroundings in the different heterotic groups. While this should be investigated further, this distinction is a clear heritage of past breeding, which would likely be confirmed by the investigation of specific selection signs in the SNP composition in this region. Whether *PIP2;5* expression (and any associated phenotype of interest) was the intended target of this past selection or, on the contrary, a side effect of a selection affecting a closely related locus is unclear so far. The investigation of these particular heterotic groups and their potentially contrasting phenotypes such as tolerance to specific climatic scenarios or hydraulic parameters might help answer this question. Regarding the prevalence of local eQTLs, specific functional characterization would be necessary to confirm the *cis* nature of the underlying regulatory mechanisms. The observation of 2 different leaf zones pointed to the different nature of most distant eQTLs, which could point to zone-specific regulation mechanisms. The consistency of the multiple local eQTLs identified for a unique isoform (2 zones and secondary local eQTLs identified by covariate GWAS) should be assessed more carefully by looking at local LD patterns and haplotypes among the population.

PIP1;1 expression varied in lower magnitude than other genes in the studied diversity panel. Such an implicit low variation in expression was interesting from a biological point of view. *PIP1;1* is a ubiquitous PIP isoform, expressed at high levels both in roots and aerial parts. Mechanistically, its channel properties are not yet resolved. Its permeability to water is ambiguous (Fetter et al. 2004). The difficulty of obtaining *pip1;1* knockout lines also supports its importance as an essential gene. So far, only partially silenced lines showing reduced but not null expression could be obtained (Ding and Chaumont, personal communication). The stable expression observed, here, across the diversity panel and the subsequent rather weak eQTLs detected are yet other arguments supporting the exceptional character of *PIP1;1* among the PIP subfamily. These observations may suggest a fundamental role for this specific isoform compared with a more adaptive role for the other isoforms, compatible with a higher degree of freedom and proneness to variations. A deeper investigation of the local and distant eQTLs reported in this work and other studies may bring valuable information to support or overturn this hypothesis.

In the future, effort should be made to benefit from, and connect, the different data collected for the same DROPS panel (Millet et al. 2016; Prado et al. 2018; Negro et al. 2019). In particular, phenotypical data related mainly to plant hydraulic parameters are available for the experiment from which our tissue samples were directly collected (Prado et al. 2018). Integrating all of them will be an opportunity to assess the involvement of *PIP* (and other aquaporin) genes on physiological and agronomical traits. The comparison of the (e)QTLs from our work and a hydraulic-oriented sister study (Prado et al. 2018) only revealed a few colocalizations, but they would still deserve closer attention. The absence of colocalization at the local *PIP* eQTL suggests that, if aquaporins are important in these processes, their involvement is probably more subtle and coordinated within the whole family rather than relying on 1 major isoform and a putatively nonadjustable local eQTL *cis*-acting underlying regulation.

The final eQTL list was compared with other eQTL studies based on whole-transcriptome RNA sequencing data (Li et al. 2013; Liu et al. 2017; Wang et al. 2018; Pang et al. 2019). It is noteworthy that they concern other types of samples: kernels (Liu et al. 2017; Pang et al. 2019), shoot apex on 14-d seedlings (Li et al. 2013), and aerial part of 11-d-old seedlings (Wang et al. 2018). They also worked on other diversity panels, including a maize–teosinte RIL population in the case of Wang et al. (2018). Therefore, it represented an opportunity to identify eQTLs that may be conserved across tissues and developmental stages and others that may be more specific to the particular physiological conditions. In addition, each study applied its own criteria, more or less stringent, for eQTL definition, characterization (local vs. distant), and selection. Supplementary Fig. S6 gathers the results from this work and these other studies. All local eQTLs were supported by several or all studies (in case of *PIP2;5* and *PIP1;3*). Colocalizations between distant eQTLs were detected as well, namely, for a *PIP1;1_EZ* eQTL (QTL_3) colocalizing with a *PIP1;1* distant eQTL from Liu et al. (2017)

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

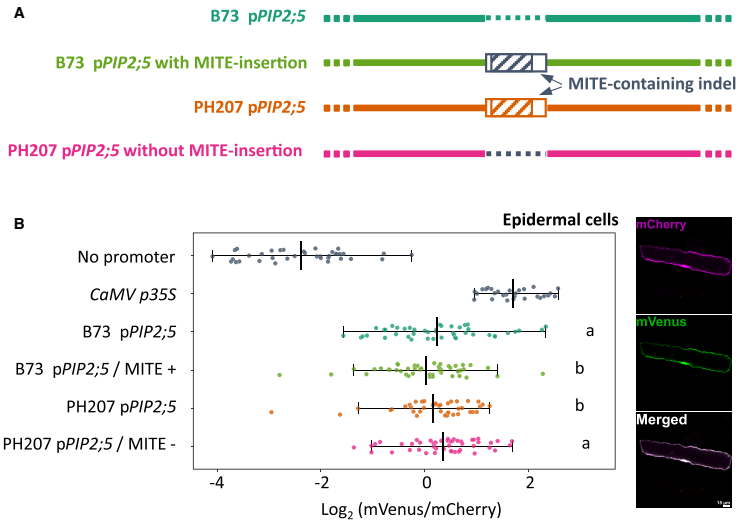


Figure 8. MITE-containing indel effect on the *PIP2;5* promoter activity. **A)** Constructs used (not drawn to scale). The green and orange/magenta backbones figure the B73 and PH207 *PIP2;5* promoter sequences. The rectangle indicates the presence of the MITE-containing indel (the hatched part representing the MITE itself), while the dashed line indicates its absence. Dark blue features indicate the artificial constructions. **B)** mVenus and mCherry fluorescent signal ratio in the leaf epidermal cells. The *mVenus* expression was driven by the tested promoter constructs as indicated in the vertical axis labels, while *mCherry* expression was driven by the double minimal *CaMV p35S* promoter, carried by the same vector and used as normalizer. Results are presented as $\log_2$(mVenus/mCherry) and group 3 independent experiment repetitions performed blind. Each dot represents an individual cell. Statistical tests were performed with a linear mixed model testing the background effect (B73 or PH207), the MITE-containing indel effect (presence or absence, referred to as "MITE+" and "MITE–," respectively, in the axis labels), and their interaction, with a random intercept taking the experience replication dependency into account (SAS or nlme function in R). Significance threshold was set to $P < 0.05$. Representative pictures of transformed epidermal were added, channels (top panel, mCherry; middle panel, mVenus; lower panel, both merged). The scale bar represents 15 μm.

(for kernel tissues) and a *PIP1;3_EZ* (QTL_9, covariate GWAS) distant eQTL colocalizing with a distant *PIP1;3* eQTL from Pang et al. (2019) (for kernel tissues). In both cases, they are located quite close to the local loci.

A MITE transposon regulates *PIP2;5* expression

A very significant eQTL peak located on the *PIP2;5* gene, both in leaf MZ and EZ. This large peak was composed of multiple pyramiding eQTLs. Patterns of local LD between SNPs from these different eQTLs suggested that these multiple eQTLs captured the same local causal polymorphisms regarding the large effect of main local eQTL (16% and 50% of the total variance of *PIP2;5* expression in EZ and MZ leaf zone, respectively [Table 3]). This was further confirmed by the covariate GWAS since the integration of local eQTL lead SNP wiped out most neighboring ones. As a whole, data suggest the presence of a sole local eQTL for *PIP2;5*.

Significant SNP segregation profiles and sequence comparison between lines diverging at the local lead SNP, B73, and PH207 allowed identifying 2 rather long haplotypes. Among the different polymorphisms identified, a MITE-containing indel, present in PH207 and located at ~430 bp from the start codon, stood out as a promising causative candidate, because of (i) its good apparent coincidence with the eQTL assessed on 14 lines, (ii) its strategic position in the promoter, and (iii) the data accumulating on MITEs, highlighting their role as genomic, genetic, and phenotypic diversity drivers. The MITE located in the PH207 *PIP2;5* promoter shares the features of the maize *Hbr* family (included in the *tourist* superfamily): conserved TIR, similar length (~310 bp), and diverse target site duplication. This family was reported to show a high degree of sequence conservation (usually ~90% sequence identity) suggesting their late origin and recent or current spreading activity (Zhang et al. 2000). Indeed, in rice, 3 elements (mPing, mGing, and mJing) belonging to different families of *tourist* MITE, with variable intrafamilial sequence similarity degrees—reflecting multiple historical amplification bursts (Lu et al. 2012)—were shown to still be active (Jiang et al. 2003; Naito et al. 2006; Dong et al. 2012; Tang et al. 2019). However, the lower sequence similarity of the identified MITE with the few other *Hbr* that were specifically observed, as well as the polymorphisms observed in one of the TIR, probably suggests a relatively ancient origin and lost transposition activity. In any case, *Hbr* MITEs were proposed as valuable

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

molecular markers for genotyping and genetic relationship prediction because of their stability, high degree of—presence/absence—polymorphism, and preference for genic regions (Casa et al. 2000, 2002, 2004; Zerjal et al. 2012; Venkatesh 2020). Therefore, like other transposons, but even more because of their smaller size and high copy number, MITEs shape genomes and create genomic diversity. They were proposed to have a major role in the formation of clusters, gathering related biosynthetic genes, for instance (Boutanaev and Osbourn 2018). However, beyond the genome structure, their integration and polymorphisms may potentially have consequences at the genetic, epigenetic, and phenotypic levels.

Genome-wide studies on 2 rice cultivars showed that genes with MITE insertions within the 500 bp upstream or downstream regions, or within introns, have significantly lower expression than genes located away from MITEs (Lu et al. 2012); however, no obvious effects of MITE insertions on individual gene expression were emphasized when crosschecking presence/absence variations and expression levels between cultivars in the same study. While most insertions are probably neutral (or have still undetected effects; Naito et al. 2009; Lu et al. 2012), in some cases, the local influence of MITEs on gene expression was concrete and attributed to several nonexclusive mechanisms. MITEs can modulate the nearby gene expression through the introduction of regulatory motifs, with positive, negative, or condition-specific effects (Yang et al. 2005; Naito et al. 2009), or via MITE-derived small RNAs (Kuang et al. 2009; Wei et al. 2014; Xin et al. 2019). MITE impacts on gene regulation may also come from RNA-independent DNA methylation (Yang et al. 2005), gene product modifications or truncations due to—yet rare—MITE insertion into exons (Guillet-Claude et al. 2004), or splicing alteration due to MITE insertion into introns (Balzan et al. 2018; Xin et al. 2019). Therefore, MITEs may have consequences at the phenotype level (Bureau and Wessler 1992; Tang et al. 2019, for instance) and were repeatedly proposed as the causal variations underlying QTLs for agronomic traits and/or historical evolution understanding. A MITE insertion into a peroxidase gene involved in lignification presumably explains a major QTL for maize cell wall digestibility, due to gene product truncation (Guillet-Claude et al. 2004). A sorghum aluminum tolerance QTL is also likely due to a MITE insertion upstream of a multidrug and toxic compound extrusion protein *MATE* gene, but the (de)regulatory mechanisms have not been characterized (Magalhaes et al. 2007). MITE insertions are also associated with several flowering-time genes and QTLs in maize, sorghum, and rapeseed (Salvi et al. 2007; Hou et al. 2012; Zerjal et al. 2012; Castelletti et al. 2014). The major flowering-time QTL *Vgt1* could be at least partly attributed to stable methylation spreading from a MITE, downregulating *ZmRap2.7* flowering-time gene transcription (located 70 kb downstream) in maize and sorghum (Salvi et al. 2002, 2007; Castelletti et al. 2014). The involvement of this particular locus was also supported by an open chromatin state coinciding with the MITE (Rodgers-Melnick et al. 2016) and long-range chromatin interaction with a nearby sequence (Peng et al. 2019), thus suggesting maybe the coincidence of multiple regulatory mechanisms. In addition, a *tourist* MITE (located ~60 kb from *TB1*) underlying the major QTL for apical dominance was shown to act as a repressor by dual luciferase assay. The QTL causative variation was, however, attributed to the close *Hopscotch* retrotransposon whose positive regulatory effect was consistent with the evolutionary data (Studer et al. 2011). Here again, it was further supported by genome-wide structural chromatin assay and long-range interaction detection (Rodgers-Melnick et al. 2016; Li et al. 2019). Finally, the cloning of a maize drought tolerance QTL revealed that a MITE inserted in the promoter of a *NAC* gene (*ZmNAC111*) reduced its expression via RNA-dependent DNA and histone methylation, thus lowering seedling drought tolerance (Mao et al. 2015). This particular MITE was particularly present in the temperate maize germplasm (Mao et al. 2015).

In our case, the negative effect of the MITE-containing indel on the *PIP2;5* promoter activity was demonstrated by a confocal microscopy–based reporter assay. The exact MITE-containing indel effect was highlighted by artificial insertion/deletion of the whole indel in and from the B73 and PH207 promoter sequences, respectively. Unfortunately, scales of such different experiments (reporter assay and RT-qPCR) did not easily match, so the magnitude of the effect relative to the expression could not be quantitatively estimated. Whether the observed effect was due to the MITE itself or the non-MITE portion of the indel was not determined. *Cis*-regulatory elements (including putative NAC, MYB, LAE, and other binding sites) were identified within the whole indel (including within the MITE) using the PlantPAN 2.0 TFBS promoter scanning tool (Chow et al. 2016) so that it might allow/increase the binding by *trans*-regulatory factors (see Supplementary Fig. S7). In addition, the MITE could be subjected to RNA-dependent DNA methylation, as small RNA database screening suggests its specific targeting by siRNAs (Supplementary Fig. S8, Johnson et al. 2007). Silencing methylation could then potentially spread across the *PIP2;5* promoter and gene region. However, mining available methylation profiles showed no particular difference in methylation at the *PIP2;5* gene and promoter location when comparing lines matching the panel and differing at this MITE-containing indel (including B73, B97, Mo17, MS71, and Oh43; Eichten et al. 2013), reducing the likelihood of this hypothesis.

In addition to the MITE-containing indel, other regulatory elements might also contribute to this very significant *PIP2;5* local eQTL. The compelling uncovering of the local regulatory elements could be achieved with a fine mapping in specifically bred nearly isogenic or backcrossed populations and by comparing and dissecting the effect of the local haplotypes in ideally identical genetic backgrounds. Another possibility of narrowing or defining the eQTLs lies in the integration of polymorphic indels of various sizes that have been genotyped in a large diversity panel including the one used in this work (Mabire et al. 2019).

In conclusion, such work on the dissection of the *PIP* gene expression variability is a valuable approach to ultimately elucidate the networks involved in the crop water homeostasis and their essential adaptation to various and ever-changing environmental scenarios.

Materials and methods

Plant material

The diversity panel was generated by crossing a common maize (*Z. mays*) flint parent (UH007) with dent lines having a restricted flowering-time variation (European DROPS project). The resulting panel comprised 252 hybrids with a 7-d flowering window and was used in several previously published studies (Millet et al. 2016; Prado et al. 2018; Negro et al. 2019; Blein-Nicolas et al. 2020). The panel was then divided into 6 genetic groups identified using the ADMIXTURE software (Negro et al. 2019). These groups were defined as follows: stiff stalk containing 32 lines, iodent containing 34 lines, Lancaster with 32 lines, Lancaster early containing 16 lines, stiff stalk early with 36 lines, and, finally, W117 admixture, which includes 102 lines that did not fit into the other groups, such as W177, D06, and NC358. One line of each group was used as a reference (B73, PH207, Mo17, Oh43, B14a, and W117 lines for stiff

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

stalk, iodent, Lancaster, Lancaster early, stiff stalk early, and W117 admixture, respectively).

In our case, the panel was grown in May 2013 in the phenotyping platform PhenoArch (LEPSE, INRAE, Montpellier) under semi-controlled conditions. Well-watered condition (soil water potential of −0.05 MPa, retention capacity) was maintained by automatic and personalized watering of the pots 3 times per day. The EZ samples (yellow cylinder at the stem base, covering green sheath removed) were collected after 10 d of culture, at 4-leaf stage, while the MZ samples (middle of the expanded leaf blade) were collected after 44 d, at 14-leaf stage, on leaf 10. Three successive days (3 h timeframe in the morning: 7:30 to 10:30 AM) was required to collect all samples while limiting the sampling time variation. Three biological replicates coming from different pots were harvested for each genotype. Samples were harvested and immediately frozen in liquid nitrogen and kept at −80 °C until RNA extraction.

For biolistic transient expression, B104 inbred seeds were surface disinfected with 50% (v/v) sodium hypochlorite for 5 min and rinsed 6 times with distilled water. Seeds were placed vertically between wet tissue papers in the dark for 2 d at 30 °C and transferred into soil in individual 6 cm diameter peat Jiffy pots filled with 80% (v/v) classic potting soil (DCM) and 20% (v/v) vermiculite (Agra-vermiculite) and grown under an 8/16 h dark/light regime (18 °C/25 °C; semicontrolled growth chamber conditions).

RNA extraction

Frozen tissues were ground with metallic beads using a Mixer Mill MM 400 grinder (Retsch), and RNA extraction was performed with the Spectrum Plant Total RNA Kit (Sigma-Aldrich). On-column DNA digestion was performed with DNAse I (Sigma-Aldrich) as recommended. Complementary DNA (cDNA) synthesis was performed from 2 μg of total RNA extract (measured with a Nanodrop ND-1000 [Isogen]), with M-MLV reverse transcriptase (Promega) in the presence of RNasin (Promega), according to the manufacturer's instructions. cDNAs were then purified with the NucleoSpin PCR Clean-up kit (Macherey-Nagel) and eluted in 50 μL of preheated (65 °C) sterile water.

RT-qPCR

The expression of 4 *PIP* genes (*PIP1;1*, Zm00001d002690; *PIP1;3*, Zm00001d051403; *PIP2;2*, Zm00001d005421; and *PIP2;5*, Zm00001d003006) was measured for the whole panel on the leaf EZ and MZ by RT-qPCR. RT-qPCR was carried out with primers (Supplementary Table S1) according to the protocol already described in Hachez et al. (2006) and Heinen et al. (2014), using the SYBR Green Master Mix 2× (Eurogentec) and StepOnePlus Real-Time PCR system (Applied Biosystems). They were performed in 96-well plates designed in a gene-optimized way: the 7 genes (4 *PIP* genes + 3 reference genes) were measured in a same plate for the 3 biological replicates of a same genotype (with 2 technical replicates). The results were normalized via the 2ΔCt method (Ct for cycle threshold) against the 3 reference genes (*Act1*, Zm00001d010159; *EF1α*, Zm00001d046449; and *Ubi*, Zm00001d053838), leading to relative expression data. The relative expression data were used as $\log_2$ transformed in all subsequent analyses. Analysis of the putative polymorphisms within the primer sequences was described in Supplementary Appendix S1.

For each trait (gene expression in a given zone), the linear mixed model was fitted:

$$Y = G + E$$

where Y is the vector of phenotypic observation, G is the random genotypic effect, and E is the residual effect. The heritability was then computed as $V_g/(V_g + V_e/n)$ where V_g is the variance attributed to the genotype, V_e is the residual variance, and n is the number of replicates.

GWAS

Lines were genotyped using 3 genotyping technologies: an Illumina HD 50K array developed by Ganal et al. (2011), an Affymetrix Axiom 600K array (Unterseer et al. 2014), and genotyping-by-sequencing, as described by Negro et al. (2019). The resulting collection gathered more than 960,000 SNPs, among which 737,202 with a minor allele frequency (MAF) > 0.05 were used for the GWAS. GWAS was implemented as described by Millet et al. (2016), Prado et al. (2018), and Negro et al. (2019). In brief, GWAS was performed on each trait (combination *PIP* gene × leaf zone, referred to as *PIPX; X_MZ/EZ*) individually, using the following single locus mixed model: $Y = \mu + X\beta + G + E$, where Y is the vector of phenotypic values, μ is the overall mean, X is the vector of SNP scores (the genotype value at the tested marker for all the lines), β is the additive effect of the tested marker (i.e. what we are testing), and G and E represent polygenic (approximately population structure effect) and residual effects, respectively. G modulates the genetic pairwise relatedness (or kinship) between individuals and was estimated from the 50K marker collection (Negro et al. 2019). In practice, it was derived from all SNPs except for those on the chromosome containing the marker being tested, according to the procedure of Rincent et al. (2014b). The marker effect β was estimated for each SNP by generalized least squares, and their significance was tested with an F-statistic. The $-\log_{10}$ of the P-values was retrieved and significance threshold was set to 5 as detailed in Negro et al. (2019). Analyses were performed with FaST-LMM v2.07 (Lippert et al. 2011). Inputted trait data were genotypic means (means of the 3 biological replicates), as $\log_2$-transformed relative expression data.

The grouping of the significant SNPs into QTLs and, hence, the eQTL delimitation was based on estimation of the LD windows for each SNP (*LD_win* approach; Negro et al. 2019). Significant SNPs with overlapping estimated LD windows (threshold of LD corrected for cryptic relatedness $r^2K = 0.1$) were considered as belonging to a single QTL. QTL intervals were delimited by the external limit of the LD window of the most extreme SNPs. Once collected and delineated, QTLs were defined by the characteristics of their lead SNP (= the most significant SNP). Colocations of QTLs from different analyses were defined by QTL interval overlapping.

Conditional GWAS were performed by introducing a SNP of interest as covariate in the GWAS model. It was generally implemented to integrate the powerful local eQTLs (defined as including the cognate *PIP* gene—delineated as the transcribed region, according to the B73 reference genome version 4 annotation—within its range), thus removing all redundant signals and shedding light on putatively hidden eQTLs of lower effects. It was also used to integrate a putative confounding effect arising from a reference gene used for the RT-qPCR, after the discovery of a QTL located on this gene (*Ubi*) for the *PIP1;1_EZ* data set. This so-called *ubi*-adjustment was implemented for both *PIP1;1* data sets.

Genetic constructs for promoter activity evaluation

All constructs used for transient expression were derived from the p*EGB 35S:Luciferase:Tnos* plasmid (GB0110, Addgene plasmid no. 68216; Sarrion-Perdigones et al. 2013). A USER cassette was

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

introduced by restriction and ligation upstream of the *Luc* coding sequence between *Nco*I and *Sfo*I sites resulting in a plasmid named GB0110m. The latter was then used as backbone to swap the *Luc* gene with a *mVenus* reporter coding sequence, to be able to observe and quantify the promoter activity by microscopy. This was obtained by digesting GB0110m with *Pac*I and *Afl*II and inserting the *mVenus* coding sequence (amplified and adapted for the insertion with 2 pairs of primers used successively, venus_GB110_F1-R1/F2-R2; Supplementary Table S5). The plasmid was named GB0110m-mVenus. In order to have a transfection control within the vector, a double minimal cauliflower mosaic virus 35S promoter *CaMV p35S:mCherry:tNOS* cassette was inserted at the *Aat*II restriction site (K7mCherry_AatII_F/R primers). The ultimate vector was named GB0110m-mVenus-K7mCherry.

MITE-containing indel detection and *PIP2;5* promoter amplification

The presence/absence of the MITE-containing indel into the *PIP2;5* promoter region of specific lines was assessed with different primer pairs (Supplementary Table S5; Fig. 8). The *PIP2;5* promoter region (~2 kb, up to the next upstream annotated gene) and the MITE sequence amplification resulted in either bands with different sizes or the presence/absence of the indel, respectively.

Both *PIP2;5* promoter versions were amplified by PCR from B73 and PH207 pure lines and cloned into pGEM-T Easy vector (Promega) for sequence validation. The promoter region was selected from the limit of the first upstream annotated gene (including a part of the presumed 3′ UTR) up to *PIP2;5* translation initiation ATG (not included), including the 5′ UTR, consisting of a 1,884 and 2,600 bp region for B73 and PH207, respectively. Validated promoter regions were then inserted into GB0110m_mVenus_K7mCherry in the USER cassette, with the appropriate cloning procedure (Nour-Eldin et al. 2006; F_user_GB0110m_prom1 and R_user_5UTR_G0110m_prom). To demonstrate the role of the MITE-containing indel present in the PH207 *PIP2;5* promoter allele, versions of the B73 and PH207 *PIP2;5* promoters were created by adding or removing, respectively, the whole MITE-containing indel sequence by triple PCR (primers in Supplementary Table S5).

Biolistic transient transfection of leaf epidermal cells

Bombardment of leaf tissues with plasmid DNA–coated microparticles was performed using the Biolistic PDS-1000/He Particle Delivery System (BioRad) according to the manufacturer's instructions and following a protocol adapted from Chevalier et al. (2014). For each bombardment, 0.5 mg of 0.6 µm gold microbeads (BioRad) was used, coated with ~2 µg of plasmid DNA. Twelve-day-old plants (B104 inbred line) were used for transfection. They were removed from soil, and roots were washed for practical purpose. The median portion of the last fully developed leaf (usually the 4th leaf) was attached on a Petri dish lid, adaxial face upward, and placed on the target shelf of the bombardment chamber, 3 cm below the macrocarrier launch assembly. The whole plant stood in the lower part of the chamber. The macrocarrier flight distance was set to 11 mm, while the gap distance was set to 10 mm. Vacuum was applied up to 29″ Hg (98 kPa) before firing (1,100 psi rupture disk—7,600 kPa). Plants were then carefully transferred back into soil, adding stakes if necessary, to keep them in the best possible shape. Plants were kept in the usual dark/light conditions for ~40 h before observation by microscopy.

Leaf epidermal cells were screened based on the transformation control (mCherry) and were imaged using a Zeiss LSM710 confocal microscope equipped with a spectral detector (C-Apochromat 40×/1.20 aqueous immersion objective). mCherry and mVenus were excited using 561 and 514 nm laser track, respectively. Emitted light was collected through a dichroic mirror on detectors 583 to 650 nm (561/633 MBS; master gain 700) for mCherry and 519 to 564 nm (485/514 MBS; master gain 799) for mVenus. Settings were kept identical along all experiments. Images were analyzed with the FIJI software (Schindelin et al. 2012). Signal was first smoothed, and then a rectangle was drawn to frame the epidermal cell of interest as closely as possible. Selected region was manually adjusted to exclude signal coming from potentially adjacent transformed cells and oversaturated nucleus region when present. The mVenus/mCherry ratio was calculated based on the total fluorescence obtained within the defined mask for each marker. Statistics were performed using SAS, fitting a linear mixed model, with a random intercept taking the experience dependency into account. The ratiometric data were log-transformed to fulfill the assumptions. Bombardment, image acquisition, and analysis were performed blind for 3 repetitions.

Accession numbers

Sequence data from this article can be found in the maize genome database (https://www.maizegdb.org/) under accession numbers Zm00001d010159 (*Act1*), Zm00001d046449 (*EF1α*), Zm00001d053838 (*Ubi2*), Zm00001d002690 (*PIP1;1*), Zm00001d051403 (*PIP1;3*), Zm00001d005421 (*PIP2;2*), and Zm00001d003006 (*PIP2;5*). Data regarding the primer sequences are presented in Supplementary Tables S1 and S5.

Acknowledgments

We thank the UCLouvain IMABIOL imaging platform and Marie-Christine Eloy for her confocal expertise.

Author contributions

L.M., L.J., S.D.N., A.C., C.W., F.T., and F.C. designed the experiments. L.M., L.J., S.A.P., and J.N. performed the experiments, and L.M., L.J., S.D.N., C.W., F.T., and F.C. analyzed and discussed the data. L.M. and F.C. wrote the manuscript. M.J.L., C.W., S.D.N., and F.T. reviewed and edited it. All authors read and approved the submitted version.

Supplementary data

The following materials are available in the online version of this article.

Supplementary Figure S1. Trait correlation.
Supplementary Figure S2. *PIP* gene expression data.
Supplementary Figure S3. Endogenous control gene location checking.
Supplementary Figure S4. B73–PH207 genome collinearity upstream of *PIP2;5*.
Supplementary Figure S5. The indel present in PH207 ~430 bp upstream of the *PIP2;5* start codon contains a MITE.
Supplementary Figure S6. *PIP* gene eQTLs.
Supplementary Figure S7. Predicted transcription factor binding sites within the MITE-containing indel sequence.
Supplementary Figure S8. Small RNA aligning on the MITE-containing indel sequence.
Supplementary Table S1. List of primers used for the measurement of relative expression of 5 PIP genes by RT-qPCR.

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

Supplementary Table S2. Initial eQTLs.
Supplementary Table S3. eQTLs detected with all models.
Supplementary Table S4. Summary of eQTLs detected with all models.
Supplementary Table S5. Primers used for molecular cloning.
Supplementary Appendix S1. Analysis of the putative polymorphisms within the primer sequences.

Funding

This work was supported by the European Union (project no. FP7-244374 [DROPS]), the Fonds De La Recherche Scientifique - FNRS (grant T.0038.18), the "Fédération Wallonie-Bruxelles-Actions de Recherches Concertées" (grant ARC16/21-075), and the Pierre and Colette Bauchau Award. L.C.M. was a research fellow at the FNRS (grant 1.A.366.17) and M.J.L. at the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture - FRIA (grant 1.E.086.21F).

Conflict of interest statement. None declared.

Data availability

The data underlying this article are available in the article and in its online supplementary material.

References

Balzan S, Carraro N, Salleres B, Cortivo CD, Tuinstra MR, Johal G, Varotto S. Genetic and phenotypic characterization of a novel *brachytic2* allele of maize. *Plant Growth Regul*. 2018:86(1):81–92. https://doi.org/10.1007/s10725-018-0412-6

Blein-Nicolas M, Negro SS, Balliau T, Welcker C, Cabrera-Bosquet L, Nicolas SD, Charcosset A, Zivy M. A systems genetics approach reveals environment-dependent associations between SNPs, protein coexpression, and drought-related traits in maize. *Genome Res*. 2020:30(11):1593–1604. https://doi.org/10.1101/gr.255224.119

Boutanaev AM, Osbourn AE. Multigenome analysis implicates miniature inverted-repeat transposable elements (MITEs) in metabolic diversification in eudicots. *Proc Natl Acad Sci*. 2018:115(28):E6650–E6658. https://doi.org/10.1073/pnas.1721318115

Buckler ES, Gaut BS, McMullen MD. Molecular and functional diversity of maize. *Curr Opin Plant Biol*. 2006:9(2):172–176. https://doi.org/10.1016/j.pbi.2006.01.013

Bureau TE, Wessler SR. Tourist: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell*. 1992:4(10):1283–1294. https://doi.org/10.1105/tpc.4.10.1283

Caldeira CF, Jeanguenin L, Chaumont F, Tardieu F. Circadian rhythms of hydraulic conductance and growth are enhanced by drought and improve plant performance. *Nat Commun*. 2014:5(1):5365. https://doi.org/10.1038/ncomms6365

Casa AM, Brouwer C, Nagel A, Wang L, Zhang Q, Kresovich S, Wessler SR. The MITE family *Heartbreaker* (*Hbr*): molecular markers in maize. *Proc Natl Acad Sci*. 2000:97(18):10083–10089. https://doi.org/10.1073/pnas.97.18.10083

Casa AM, Mitchell SE, Smith OS, Register JC, Wessler SR, Kresovich S. Evaluation of Hbr (MITE) markers for assessment of genetic relationships among maize (*Zea mays* L.) inbred lines. *Theor Appl Genet*. 2002:104(1):104–110. https://doi.org/10.1007/s001220200012

Casa AM, Nagel A, Wessler SR. MITE display. *Mob Genet Elements*. 2004:260:175–188. https://doi.org/10.1385/1-59259-755-6:175

Castelletti S, Tuberosa R, Pindo M, Salvi S. A MITE transposon insertion is associated with differential methylation at the maize flowering time QTL *Vgt1*. *G3 (Bethesda)*. 2014:4(5):805–812. https://doi.org/10.1534/g3.114.010686

Chaumont F, Tyerman SD. Aquaporins: highly regulated channels controlling plant water relations. *Plant Physiol*. 2014:164(4):1600–1618. https://doi.org/10.1104/pp.113.233791

Chen J, Hu Q, Zhang Y, Lu C, Kuang H. P-MITE: a database for plant miniature inverted-repeat transposable elements. *Nucleic Acids Res*. 2014:42(D1):D1176–D1181. https://doi.org/10.1093/nar/gkt1000

Chevalier AS, Bienert GP, Chaumont F. A new LxxxA motif in the transmembrane Helix3 of maize aquaporins belonging to the plasma membrane intrinsic protein PIP2 group is required for their trafficking to the plasma membrane. *Plant Physiol*. 2014:166(1):125–138. https://doi.org/10.1104/pp.114.240945

Chow C-N, Zheng H-Q, Wu N-Y, Chien C-H, Huang H-D, Lee T-Y, Chiang-Hsieh Y-F, Hou P-F, Yang T-Y, Chang W-C. PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res*. 2016:44(D1):D1154–D1160. https://doi.org/10.1093/nar/gkv1035

Cleary S, Seoighe C. Perspectives on allele-specific expression. *Annu Rev Biomed Data Sci*. 2021:4:101–122. https://doi.org/10.1146/annurev-biodatasci-021621-122219

Devi MJ, Sinclair TR, Jain M, Gallo M. Leaf aquaporin transcript abundance in peanut genotypes diverging in expression of the limited-transpiration trait when subjected to differing vapor pressure deficits and aquaporin inhibitors. *Physiol Plant*. 2016:156(4):387–396. https://doi.org/10.1111/ppl.12378

Ding L, Milhiet T, Parent B, Meziane A, Tardieu F, Chaumont F. The plasma membrane aquaporin ZmPIP2;5 enhances the sensitivity of stomatal closure to water deficit. *Plant Cell Environ*. 2022:45(4):1146–1156. https://doi.org/10.1111/pce.14276

Dong H-T, Zhang L, Zheng K-L, Yao H-G, Chen J, Yu F-C, Yu X-X, Mao B-Z, Zhao D, Yao J, *et al.* A Gaijin-like miniature inverted repeat transposable element is mobilized in rice during cell differentiation. *BMC Genomics* 2012:13(1):135. https://doi.org/10.1186/1471-2164-13-135

Eichten SR, Briskine R, Song J, Li Q, Swanson-Wagner R, Hermanson PJ, Waters AJ, Starr E, West PT, Tiffin P, *et al.* Epigenetic and genetic influences on DNA methylation variation in maize populations. *Plant Cell* 2013:25(8):2783–2797. https://doi.org/10.1105/tpc.113.114793

Fetter K, Van Wilder V, Moshelion M, Chaumont F. Interactions between plasma membrane aquaporins modulate their water channel activity. *Plant Cell* 2004:16(1):215–228. https://doi.org/10.1105/tpc.017194

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner E-M, Hansen M, Joets J, *et al.* A large maize (*Zea mays L.*) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS One* 2011:6(12):e28334. https://doi.org/10.1371/journal.pone.0028334

Guillet-Claude C, Birolleau-Touchard C, Manicacci D, Rogowsky PM, Rigau J, Murigneux A, Martinant J-P, Barrière Y. Nucleotide diversity of the ZmPox3 maize peroxidase gene: relationships between a MITE insertion in exon 2 and variation in forage maize digestibility. *BMC Genet*. 2004:5(1):19. https://doi.org/10.1186/1471-2156-5-19

Hachez C, Heinen RB, Draye X, Chaumont F. The expression pattern of plasma membrane aquaporins in maize leaf highlights their role in hydraulic regulation. *Plant Mol Biol*. 2008:68(4–5):337–353. https://doi.org/10.1007/s11103-008-9373-x

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

Hachez C, Moshelion M, Zelazny E, Cavez D, Chaumont F. Localization and quantification of plasma membrane aquaporin expression in maize primary root: a clue to understanding their role as cellular plumbers. *Plant Mol Biol*. 2006:62(1–2):305–323. https://doi.org/10.1007/s11103-006-9022-1

Heinen RB, Bienert GP, Cohen D, Chevalier AS, Uehlein N, Hachez C, Kaldenhoff R, Le Thiec D, Chaumont F. Expression and characterization of plasma membrane aquaporins in stomatal complexes of *Zea mays*. *Plant Mol Biol*. 2014:86(3):335–350. https://doi.org/10.1007/s11103-014-0232-7

Heinen RB, Ye Q, Chaumont F. Role of aquaporins in leaf physiology. *J Exp Bot*. 2009:60(11):2971–2985. https://doi.org/10.1093/jxb/erp171

Hou J, Long Y, Raman H, Zou X, Wang J, Dai S, Xiao Q, Li C, Fan L, Liu B, *et al*. A Tourist-like MITE insertion in the upstream region of the BnFLC.A10 gene is associated with vernalization requirement in rapeseed (*Brassica napus* L.). *BMC Plant Biol*. 2012:12(1):238 https://doi.org/10.1186/1471-2229-12-238

Hu H, Crow T, Nojoomi S, Schulz AJ, Estévez-Palmas JM, Hufford MB, Flint-Garcia S, Sawers R, Rellán-Álvarez R, Ross-Ibarra J, *et al*. Allele-specific expression reveals multiple paths to highland adaptation in maize. *Mol Biol Evol*. 2022:39(11):msac239. https://doi.org/10.1093/molbev/msac239

Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR. An active DNA transposon family in rice. *Nature* 2003:421-(6919):163–167. https://doi.org/10.1038/nature01214

Johnson C, Bowman L, Adai AT, Vance V, Sundaresan V. CSRDB: a small RNA integrated database and browser resource for cereals. *Nucleic Acids Res* 2007:35:D829–D833. https://doi.org/10.1093/nar/gkl991

Khan S, Thomas BR, De la Mata R, Randall MJ, Zhang W, Zwiazek JJ. Variation in aquaporin and physiological responses among *Pinus contorta* families under different moisture conditions. *Plants* 2019:8(1):13. https://doi.org/10.3390/plants8010013

Kuang H, Padmanabhan C, Li F, Kamei A, Bhaskar PB, Ouyang S, Jiang J, Buell CR, Baker B. Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: new functional implications for MITEs. *Genome Res*. 2009:19(1):42–56. https://doi.org/10.1101/gr.078196.108

Li E, Liu H, Huang L, Zhang X, Dong X, Song W, Zhao H, Lai J. Long-range interactions between proximal and distal regulatory regions in maize. *Nat Commun*. 2019:10(1):2633. https://doi.org/10.1038/s41467-019-10603-4

Li L, Petsch K, Shimizu R, Liu S, Xu WW, Ying K, Yu J, Scanlon MJ, Schnable PS, Timmermans MCP, *et al*. Mendelian and non-Mendelian regulation of gene expression in maize. *PLoS Genet*. 2013:9(1):e1003202. https://doi.org/10.1371/journal.pgen.1003202

Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. *Nat Methods*. 2011:8(10):833–835. https://doi.org/10.1038/nmeth.1681

Liu H, Luo X, Niu L, Xiao Y, Chen L, Liu J, Wang X, Jin M, Li W, Zhang Q, *et al*. Distant eQTLs and non-coding sequences play critical roles in regulating gene expression and quantitative trait variation in maize. *Mol Plant*. 2017:10(3):414–426. https://doi.org/10.1016/j.molp.2016.06.016

Lopez F, Bousser A, Sissoëff I, Gaspar M, Lachaise B, Hoarau J, Mahé A. Diurnal regulation of water transport and aquaporin gene expression in maize roots: contribution of PIP2 proteins. *Plant Cell Physiol*. 2003:44(12):1384–1395. https://doi.org/10.1093/pcp/pcg168

Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H. Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. *Mol Biol Evol*. 2012:29(3):1005–1017. https://doi.org/10.1093/molbev/msr282

Mabire C, Duarte J, Darracq A, Pirani A, Rimbert H, Madur D, Combes V, Vitte C, Praud S, Rivière N, *et al*. High throughput genotyping of structural variations in a complex plant genome using an original Affymetrix Axiom array. *BMC Genomics* 2019:20(1):848. https://doi.org/10.1186/s12864-019-6136-9

Magalhaes JV, Liu J, Guimarães CT, Lana UGP, Alves VMC, Wang Y-H, Schaffert RE, Hoekenga OA, Piñeros MA, Shaff JE, *et al*. A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. *Nat Genet*. 2007:39(9):1156–1161. https://doi.org/10.1038/ng2074

Mao H, Wang H, Liu S, Li Z, Yang X, Yan J, Li J, Tran L-SP, Qin F. A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. *Nat Commun*. 2015:6(1):8326. https://doi.org/10.1038/ncomms9326

Maurel C, Boursiac Y, Luu DT, Santoni V, Shahzad Z, Verdoucq L. Aquaporins in plants. *Physiol Rev*. 2015:95(4):1321–1358. https://doi.org/10.1152/physrev.00008.2015

Millet EJ, Welcker C, Kruijer W, Negro S, Coupel-Ledru A, Nicolas SD, Laborde J, Bauland C, Praud S, Ranc N, *et al*. Genome-wide analysis of yield in Europe: allelic effects vary with drought and heat scenarios. *Plant Physiol*. 2016:172(2):749–764. https://doi.org/10.1104/pp.16.00621

Naito K, Cho E, Yang G, Campbell MA, Yano K, Okumoto Y, Tanisaka T, Wessler SR. Dramatic amplification of a rice transposable element during recent domestication. *Proc Natl Acad Sci*. 2006:103-(47):17620–17625. https://doi.org/10.1073/pnas.0605421103

Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* 2009:461(7267):1130–1134. https://doi.org/10.1038/nature08479

Negro SS, Millet EJ, Madur D, Bauland C, Combes V, Welcker C, Tardieu F, Charcosset A, Nicolas SD. Genotyping-by-sequencing and SNP-arrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. *BMC Plant Biol*. 2019:19(1):318. https://doi.org/10.1186/s12870-019-1926-4

Nour-Eldin HH, Hansen BG, Norholm MH, Jensen JK, Halkier BA. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res*. 2006:34:e122.

Pang J, Fu J, Zong N, Wang J, Song D, Zhang X, He C, Fang T, Zhang H, Fan Y, *et al*. Kernel size-related genes revealed by an integrated eQTL analysis during early maize kernel development. *Plant J*. 2019:98(1):19–32. https://doi.org/10.1111/tpj.14193

Peng Y, Xiong D, Zhao L, Ouyang W, Wang S, Sun J, Zhang Q, Guan P, Xie L, Li W, *et al*. Chromatin interaction maps reveal genetic regulation for quantitative traits in maize. *Nat Commun*. 2019:10(1):2632. https://doi.org/10.1038/s41467-019-10602-5

Pou A, Hachez C, Couvreur V, Maistriaux L, Ismail A, Chaumont F. Exposure to high nitrogen triggered a genotype-dependent modulation of cell and root hydraulics, which can involve aquaporin regulation. *Physiol Plant*. 2022:174(1):e13640. https://doi.org/10.1111/ppl.13640

Prado SA, Cabrera-Bosquet L, Grau A, Coupel-Ledru A, Millet EJ, Welcker C, Tardieu F. Phenomics allows identification of genomic regions affecting maize stomatal conductance with conditional effects of water deficit and evaporative demand. *Plant Cell Environ*. 2018:41(2):314–326. https://doi.org/10.1111/pce.13083

Rincent R, Moreau L, Monod H, Kuhn E, Melchinger AE, Malvar RA, Moreno-Gonzalez J, Nicolas S, Madur D, Combes V, *et al*. Recovering power in association mapping panels with variable

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

levels of linkage disequilibrium. *Genetics* 2014b:197(1):375–387. https://doi.org/10.1534/genetics.113.159731

Rincent R, Nicolas S, Bouchet S, Atlmann T, Brunel D, Revilla P, Malvar RA, Moreno-Gonzalez J, Campo L, Melchinger AE, *et al.* Dent and flint maize diversity panels reveal important genetic potential for increasing biomass production. *Theor Appl Genet.* 2014a:127-(11):2313–2331. https://doi.org/10.1007/s00122-014-2379-7

Rishmawi L, Bauget F, Protto V, Bauland C, Nacry P, Maurel C. Natural variation of maize root hydraulic architecture underlies highly diverse water uptake capacities. *Plant Physiol.* 2023:192(3): 2404–2418. https://doi.org/10.1093/plphys/kiad213

Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES. Open chromatin reveals the functional maize genome. *Proc Natl Acad Sci.* 2016:113-(22):E3177–E3184. https://doi.org/10.1073/pnas.1525244113

Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashev S, Bruggemann E, *et al.* Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc Natl Acad Sci.* 2007:104(27): 11376–11381. https://doi.org/10.1073/pnas.0704145104

Salvi S, Tuberosa R, Chiapparino E, Maccaferri M, Veillet S, van Beuningen L, Isaac P, Edwards K, Phillips RL. Toward positional cloning of vgt1, a QTL controlling the transition from the vegetative to the reproductive phase in maize. *Plant Mol Biol.* 2002:48-(5/6):601–613. https://doi.org/10.1023/A:1014838024509

Sarrion-Perdigones A, Vazquez-Vilar M, Palací J, Castelijns B, Forment J, Ziarsolo P, Blanca J, Granell A, Orzaez D. GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol.* 2013:162(3):1618–1631. https://doi.org/10.1104/pp.113.217661

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods.* 2012:9(7):676–682. https://doi.org/10.1038/nmeth.2019

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet.* 2012:44(7):825–830. https://doi.org/10.1038/ng.2314

Studer A, Zhao Q, Ross-Ibarra J, Doebley J. Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat Genet.* 2011:43(11):1160–1163. https://doi.org/10.1038/ng.942

Sun G, Zhu C, Kramer MH, Yang SS, Song W, Piepho HP, Yu J. Variation explained in mixed-model association mapping. *Heredity (Edinb)* 2010:105(4):333–340. https://doi.org/10.1038/hdy.2010.11

Sutka M, Li G, Boudet J, Boursiac Y, Doumas P, Maurel C. Natural variation of root hydraulics in Arabidopsis grown in normal and salt-stressed conditions. *Plant Physiol.* 2011:155(3):1264–1276. https://doi.org/10.1104/pp.110.163113

Tang Y, Ma X, Zhao S, Xue W, Zheng X, Sun H, Gu P, Zhu Z, Sun C, Liu F, *et al.* Identification of an active miniature inverted-repeat transposable element *mJing* in rice. *Plant J.* 2019:98(4):639–653. https://doi.org/10.1111/tpj.14260

Unterseer S, Bauer E, Haberer G, Seidel M, Knaak C, Ouzunova M, Meitinger T, Strom TM, Fries R, Pausch H, *et al.* A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics* 2014:15(1):823. https://doi.org/10.1186/1471-2164-15-823

Vandeleur RK, Mayo G, Shelden MC, Gilliham M, Kaiser BN, Tyerman SD. The role of plasma membrane intrinsic protein aquaporins in water transport through roots: diurnal and drought stress responses reveal different strategies between isohydric and anisohydric cultivars of grapevine. *Plant Physiol.* 2009:149(1):445–460. https://doi.org/10.1104/pp.108.128645

Vaziriyeganeh M, Khan S, Zwiazek JJ. Analysis of aquaporins in northern grasses reveal functional importance of *Puccinellia nuttalliana* PIP2;2 in salt tolerance. *Plant Cell Environ.* 2023:46(7): 2159–2173. https://doi.org/10.1111/pce.14589

Venkatesh NB. Miniature inverted-repeat transposable elements (MITEs), derived insertional polymorphism as a tool of marker systems for molecular plant breeding. *Mol Biol Rep.* 2020:47(4): 3155–3167. https://doi.org/10.1007/s11033-020-05365-y

Wang X, Chen Q, Wu Y, Lemmon ZH, Xu G, Huang C, Liang Y, Xu D, Li D, Doebley JF, *et al.* Genome-wide analysis of transcriptional variability in a large maize-teosinte population. *Mol Plant.* 2018:11(3): 443–459. https://doi.org/10.1016/j.molp.2017.12.011

Wei L, Gu L, Song X, Cui X, Lu Z, Zhou M, Wang L, Hu F, Zhai J, Meyers BC, *et al.* Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. *Proc Natl Acad Sci.* 2014:111(10):3877–3882. https://doi.org/10.1073/pnas.1318131111

Xin Y, Ma B, Xiang Z, He N. Amplification of miniature inverted-repeat transposable elements and the associated impact on gene regulation and alternative splicing in mulberry (*Morus notabilis*). *Mob DNA.* 2019:10(1):27. https://doi.org/10.1186/s13100-019-0169-0

Yang G, Lee Y-H, Jiang Y, Shi X, Kertbundit S, Hall TC. A two-edged role for the transposable element *Kiddo* in the *rice ubiquitin2* promoter. *Plant Cell* 2005:17(5):1559–1568. https://doi.org/10.1105/tpc.104.030528

Zerjal T, Rousselet A, Mhiri C, Combes V, Madur D, Grandbastien M-A, Charcosset A, Tenaillon MI. Maize genetic diversity and association mapping using transposable element insertion polymorphisms. *Theor Appl Genet.* 2012:124(8):1521–1537. https://doi.org/10.1007/s00122-012-1807-9

Zhang Q, Arbuckle J, Wessler SR. Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker* into genic regions of maize. *Proc Natl Acad Sci.* 2000:97(3):1160–1165. https://doi.org/10.1073/pnas.97.3.1160

Downloaded from https://academic.oup.com/plphys/article/196/1/368/7688441 by guest on 26 January 2025

Runcie DE, Qu J, Cheng H, Crawford L (2021) MegaLMM: mega-scale linear mixed models for genomic predictions with thousands of traits. Genom Biol 22:1–25

Schrag TA, Westhues M, Schipprack W, Seifert F, Thiemann A, Scholten S, Melchinger AE (2018) Beyond genomic prediction: combining different types of omics data can improve prediction of hybrid performance in maize. Genetics 208:1373–1385

Unterseer S, Bauer E, Haberer G, Seidel M, Knaak C, Ouzunova M, Meitinger T, Strom TM, Fries R, Pausch H (2014) A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. BMC Genom 15:1–15

van Rossum B-J, van Eeuwijk F, Boer M (2023) statgenSTA: single trial analysis (STA) of Field Trials. R Package version 1.0.12

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Ward J, Rakszegi M, Bedő Z, Shewry PR, Mackay I (2015) Differentially penalized regression to predict agronomic traits from metabolites and markers in wheat. BMC Genet 16:1–7

Wen J, Jiang F, Weng Y, Sun M, Shi X, Zhou Y, Yu L, Wu Z (2019) Identification of heat-tolerance QTLs and high-temperature stress-responsive genes through conventional QTL mapping, QTL-seq and RNA-seq in tomato. BMC Plant Biol 19:1–17

Westhues M, Schrag TA, Heuer C, Thaller G, Utz HF, Schipprack W, Thiemann A, Seifert F, Ehret A, Schlereth A (2017) Omics-based hybrid prediction in maize. Theor Appl Genet 130:1927–1939

Westhues CC, Simianer H, Beissinger TM (2022) learnMET: an R package to apply machine learning methods for genomic prediction using multi-environment trial data. G3 12:226

Zhou X, Stephens M (2014) Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods 11:407–409

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Journal of Experimental Botany
https://doi.org/10.1093/jxb/erae500 Advance Access Publication 19 December 2024

RESEARCH PAPER

The maize aquaporin ZmPIP1;6 enhances stomatal opening and CO_2- and ABA-induced stomatal closure

Lei Ding[1,*], Maxime J. Laurent[1], Thomas Milhiet[1], Stijn Aesaert[2,3], Mieke Van Lijsbettens[2,3], Laurens Pauwels[2,3,†], Hilde Nelissen[2,3], Dirk Inzé[2,3], and François Chaumont[1,*]

[1] Louvain Institute of Biomolecular Science and Technology, UCLouvain, 1348 Louvain-la-Neuve, Belgium
[2] Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium
[3] VIB Center for Plant Systems Biology, Technologiepark 71, 9052 Ghent, Belgium
[†] Present address: Department of Biotechnology, Faculty of Bioscience Engineering, Ghent University, Proeftuinstraat 86, 9000 Ghent, Belgium.

* Correspondence: francois.chaumont@uclouvain.be or lei.ding@uclouvain.be

Received 1 August 2024; Editorial decision 10 December 2024; Accepted 18 December 2024

Editor: Tracy Lawson, University of Essex, UK

Abstract

The plasma membrane aquaporin ZmPIP1;6 is expressed in maize stomatal complexes, with higher expression during the day than at night. To elucidate the role of ZmPIP1;6 in gas exchange and stomatal movement, it was expressed in maize (inbred line B104) under the control of the *p35S* promoter (overexpression) or its native promoter fused with *mYFP* (monomeric yellow fluorescent protein) cDNA (*mYFP-ZmPIP1;6*). In stomatal complexes of the mature zone of the leaf, mYFP–ZmPIP1;6 showed higher expression in subsidiary cells than in guard cells, and light and dark treatments influenced its subcellular localization. Notably, ZmPIP1;6 internalization increased in dark conditions compared with light. Stomatal opening was greater in *ZmPIP1;6* overexpression plants than in the wild type, while closure exhibited greater sensitivity to elevated CO_2 concentration or abscisic acid (ABA) treatment. Our findings revealed that reactive oxygen species (H_2O_2) were involved in ABA-induced stomatal closure, while ZmPIP1;6 was unable to facilitate H_2O_2 diffusion when expressed in yeast. Finally, *ZmPIP1;6* overexpression and *mYFP-ZmPIP1;6* transgenic plants exhibited higher abaxial stomatal density than the wild type. Overall, these results indicate that ZmPIP1;6 plays important roles in stomatal opening and CO_2- and ABA-induced stomatal closure.

Keywords: ABA signaling, aquaporin, CO_2 signaling, H_2O_2, maize stomata, stomatal movement.

Introduction

Gas exchange of CO_2 and water vapor between plants and the atmosphere takes place via stomata, which are micropores in the epidermis of the aerial parts of land plants (Chater *et al.*, 2017). Stomata are composed of dumbbell-shaped guard cells (GCs) surrounded by two subsidiary cells (SCs) in grasses such as maize, and of two kidney-shaped GCs in other species such as Arabidopsis. The opening and closing of stomata depend on variations in the turgor pressure of the GCs and, consequently, on the flow of water and ions across their plasma membrane. In this process, GCs can adjust their volume by ~40% in a few minutes (Franks *et al.*, 2001), based on a mechanism of facilitated water diffusion in the membrane.

© The Author(s) 2024. Published by Oxford University Press on behalf of the Society for Experimental Biology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erae500/7928862 by guest on 26 January 2025

Aquaporins play important roles in osmotic adjustment during stomatal movement (Ding *et al.*, 2024). Aquaporins are small membrane proteins, which facilitate the diffusion of water and other small solutes, and are involved in many physiological processes at different cell, organ, and tissue levels (Chaumont and Tyerman, 2014). Plant aquaporins are grouped into different subfamilies, including plasma membrane intrinsic proteins (PIPs) and tonoplast intrinsic proteins (TIPs). Furthermore, PIPs are phylogenetically divided into two groups, PIP1 and PIP2, that exhibit different subcellular localization and activities when expressed alone in heterologous systems or in maize protoplasts (Chaumont *et al.*, 2000; Fetter *et al.*, 2004; Zelazny *et al.*, 2007). Whereas PIP1s are retained in the endoplasmic reticulum, PIP2s are found in the plasma membrane, but PIP1s and PIP2s can assemble as heterotetramers, resulting in co-localization of both proteins in the plasma membrane and an increase in the cell membrane water permeability (Zelazny *et al.*, 2007; Berny *et al.*, 2016). Many *PIP* genes and several *TIP* genes are expressed in stomatal complexes/GCs, such as *AtPIP1;2* and *AtPIP2;1* in *Arabidopsis thaliana* (Kaldenhoff *et al.*, 1995; Grondin *et al.*, 2015; Ding *et al.*, 2024), *HvPIP1;6* in barley (Wei *et al.*, 2007), *NtAQP1* in tobacco (Uehlein *et al.*, 2008), *SoPIP1;1* in spinach (Frayse *et al.*, 2005), *VfPIP1* in broad bean (Cui *et al.*, 2008), and *SunTIP7* and *SunTIP20* in sunflower (Sarda *et al.*, 1997). In maize, ZmPIP1;2 and ZmPIP2;1/2;2 were detected in stomatal complexes using immunocytochemistry (Hachez *et al.*, 2008). Furthermore, a transcriptomic study of *ZmPIP* gene expression in maize laser-microdissected stomatal complexes revealed that almost all *ZmPIP* genes except *ZmPIP2;7* were expressed (Heinen *et al.*, 2014). Additionally, reverse genetic approaches showed that stomatal density, aperture, and movement were altered by the deregulation of *PIP* expression in tobacco (Aharon *et al.*, 2003; Ding *et al.*, 2004), *A. thaliana* (Cui *et al.*, 2008; Li *et al.*, 2015), *Populus×canescens* (Bi *et al.*, 2015), soybean (Zhou *et al.*, 2014), and rice (Hanba *et al.*, 2004; for a review, see Hachez *et al.*, 2017).

Stomatal movement is regulated by various factors including circadian rhythm (Webb, 2003), light (Shimazaki *et al.*, 2007), plant water status, pathogens (Melotto *et al.*, 2006; McLachlan *et al.*, 2014), and several signaling components, such as abscisic acid (ABA) (Pei *et al.*, 2000), H_2O_2 (Bright *et al.*, 2006; Hua *et al.*, 2012), and CO_2 (Zhang *et al.*, 2018). Drought stress induces stomatal closure through ABA signaling (Zhu, 2002), a process that regulates aquaporin activity in GCs (Grondin *et al.*, 2015; Rodrigues *et al.*, 2017; Ding *et al.*, 2022). In ABA-induced stomatal closure, ABA binds to RCAR/PYR/PYL receptors, to capture type 2C protein phosphatases, consequently releasing the inhibition of SnRK2-type protein kinases (SnRK2.6 and OST1) (Raghavendra *et al.*, 2010; Wang *et al.*, 2016). OST1 protein kinase phosphorylates AtPIP2;1 at Ser121, enhancing the osmotic water permeability of GCs and stomatal closure (Grondin *et al.*, 2015). ABA-induced stomatal closure also results from inward accumulation of H_2O_2 in GCs facilitated by AtPIP2;1 or ZmPIP2;5 (Rodrigues *et al.*, 2017; Ding *et al.*, 2022). Indeed, H_2O_2 acts as an important secondary messenger in inducing stomatal closure in interaction with ABA and CO_2 signaling pathways (Sussmilch *et al.*, 2019). Moreover, when expressed in *Xenopus* oocytes, AtPIP2;1 facilitates CO_2 diffusion and interacts with carbonic anhydrase, which catalyzes CO_2 to form bicarbonate. Consequently, the accumulation of CO_2 in GCs induces stomatal closure by activating downstream signaling and membrane transport systems (Chater *et al.*, 2015; Zhang *et al.*, 2018). However, no difference in high $[CO_2]$-induced stomatal closure is found between *atpip2;1* mutant and wild-type (WT) plants (Wang *et al.*, 2016).

Most *PIP* genes are expressed in maize stomatal complexes, with seven of them, *ZmPIP1;1*, *ZmPIP1;2*, *ZmPIP1;3*, *ZmPIP1;5*, *ZmPIP1;6*, *ZmPIP2;1*, and *ZmPIP2;2*, accounting for >98% of the total *PIP* transcripts (Heinen *et al.*, 2014). Notably, the expression of most *PIP* genes, including *ZmPIP1;6*, follows a diurnal pattern, with the highest expression during the day. Although ZmPIP1;6 was barely detectable in the leaf during the day in the dividing, elongating, or mature zones (Hachez *et al.*, 2008), *ZmPIP1;6* appeared to be specifically expressed in stomatal complexes (Heinen *et al.*, 2014). Functional analysis revealed that ZmPIP1;6 facilitates transmembrane water diffusion after its co-expression with the ZmPIP2;1 isoform in *Xenopus* oocytes, and also facilitates CO_2 diffusion across membranes when expressed alone in yeast. However, direct evidence is still lacking to ascertain whether ABA- and CO_2-induced stomatal movement depends on ZmPIP1s in maize. To address this question, we overexpressed *ZmPIP1;6* in the maize B104 genotype and investigated how this overexpression affects stomatal movement. Our results demonstrated that ZmPIP1;6 overexpression enhanced stomatal opening, and stomata closed more rapidly in response to ABA and CO_2 treatments compared with WT plants.

Materials and methods

Transgenic maize plants

To generate *ZmPIP1;6*-overexpressing (*ZmPIP1;6* OE) maize lines, *ZmPIP1;6* (Zm00001d020383) was amplified (for the primers, see Supplementary Table S1) and inserted into the pBb7m34GW backbone vector (Karimi *et al.*, 2013; Ding *et al.*, 2020) (https://vectorvault.vib.be/) employing USER™ (Hebelstrup *et al.*, 2010) and Gateway techniques (Invitrogen, Waltham, MA, USA). *ZmPIP1;6* was constitutively overexpressed under the control of the *p35S* promoter.

The 3 kb upstream of the *ZmPIP1;6* 5′-untranslated region (UTR) was considered as containing the *ZmPIP1;6* promoter. A 3449 fragment (including 3 kb upstream sequence and the 5′UTR) was first amplified from B73 gDNA with the forward primer, 5′AGCTTTGGCTTCATCTATGT3′ and the reverse primer, 5′CTGCAGCGTGCCTCCTGCCAT3′. This fragment was cloned into the pGEM-T Easy vector (Promega) and verified by sequencing. New primers (Supplementary Table S1) were used to amplify a 3428 bp upstream 5′UTR sequence, and the amplified DNA was cloned into pDONR221 P4-P1r. *mYFP* (monomeric yellow fluorescent protein)-*ZmPIP1;6* was

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erae500/7928862 by guest on 26 January 2025

amplified from pCAMBIA-*35S-EYFP-ZmPIP1;6* (Zelazny *et al.*, 2007) (for the primers, see Supplementary Table S1). This fragment was inserted into pDONR221 P1-P2. Finally, the *ZmPIP1;6* promoter and *mYFP-ZmPIP1;6* were assembled into the pBb7m24GW expression vector (*ZmPIP1;6::mYFP-ZmPIP1;6*) (https://vectorvault.vib.be/collection/pbb7m24gw).

The B104 inbred maize line was transformed according to previously established protocols (Coussens *et al.*, 2012; Ding *et al.*, 2020; Aesaert *et al.*, 2022), at the Crop Genome Engineering Facility of the VIB-UGent Center for Plant Systems Biology, Belgium. The presence of the transgene insertion in the gDNA was checked by PCR (for the primers see Supplementary Table S1). For *ZmPIP1;6* OE maize lines, seven T_0 independent lines were obtained and seeds were generated by backcrossing with B104 inbred lines. Two independent lines were selected in this study according to the transgene expression, one line with a relatively low level of *ZmPIP1;6* expression (OE-L) and the other with a high level (OE-H). T_1 and T_2 hemizygous plants segregating for the *ZmPIP1;6* OE and non-transgenic sibling (WT) were used in this study. For *ZmPIP1;6::mYFP-ZmPIP1;6* maize lines, six T_0 independent lines were obtained and seeds were generated by backcrossing with B104 inbred lines. Two independent lines were selected in this study, namely mYFP–PIP1;6-1 and mYFP–PIP1;6-2. T_1 transgenic and non-transgenic sibling plants were used in this study. The plants were grown in a phytotron or a greenhouse as described in Ding *et al.* (2020, 2022).

Confocal microscopy

The fluorescence of mYFP in leaves was checked using an LSM 710 confocal microscope equipped with an Airyscan module (Carl Zeiss, Oberkochen, Germany). Excitation was at 514 nm and emission light was collected with a detector at 520–560 nm.

RNA extraction, real-time quantitative PCR, protein extraction, and western blot analysis

Plants were grown in a phytotron with hydroponic culture (Ding *et al.*, 2020). Mature zone of leaves were collected and flash-frozen in liquid N_2 from the second leaf of 1-week-old maize seedlings. RNA was extracted by using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA). cDNA synthesis was performed with the M-MLV RT kit (Promega). Real-time quantitative PCR (RT-qPCR) was performed with SYBR Green I (Eurogentec, Liège, Belgium) and the signal was detected with the StepOnePlus™ real-time PCR System (Life Technologies, Foster City, CA, USA). The RT-qPCR program was the same as that described in Ding *et al.* (2020). The $2^{-\Delta Ct}$ method was used to analyze the data, and *EF1α* (Zm00001d046449) and *ACT1* (Zm00001d010159) were used as housekeeping genes. The mean Ct values of the two housekeeping genes was used for normalization using the $2^{-\Delta Ct}$ method. Three plants were analyzed for each WT, OE-L, and OE-H line, with two technical replicates for each plant. The primers are listed in Supplementary Table S1. To amplify the transgene, the reverse primer was hybridizing the *t35S* terminator.

To prepare plasma membrane fractions, 50 g of fresh leaves were ground in a cold room with a blender containing 200 ml of extraction buffer (250 mM sorbitol, 50 mM Tris–HCl pH 8.0, 2 mM EDTA) supplemented with 0.6% (w/v) polyvinylpyrrolidene (PVP) K30, 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.5 mM DTT, and 2 μg ml^{-1} of protease inhibitors (leupeptin, aprotinin, antipain, pepstatin, and chymostatin). A series of centrifugations at 4 °C (770 *g* for 10 min, 10 000 *g* for 10 min, and 54 000 *g* for 30 min) was performed to obtain the microsomal fraction pellet. This pellet was resuspended in suspension buffer (330 mM sucrose, 5 mM KH_2PO_4, 3 mM KCl, pH 7.8), and the plasma membrane fraction was isolated by partitioning in an aqueous two-phase system as described by Larsson *et al.* (1987). Western blot analysis was performed according to Hachez *et al.* (2006) and Heinen *et al.* (2014).

Gas exchange

Plants were grown in a greenhouse for 4 weeks, and gas exchange measurements were performed with a Li-6400XT portable photosynthesis system (LI-COR, Lincoln, NE, USA). The photosynthetic rate and the stomatal conductance (g_s) in response to light intensity were measured at eight light intensities (50, 100, 200, 400, 800, 1200, 1500, and 1800 μmol m^{-2} s^{-1}). During the measurement, the CO_2 concentration was maintained at 400 ppm. The light-saturated photosynthetic rate of newly expanded leaves was measured at a light intensity of 1500 μmol m^{-2} s^{-1}. To determine g_s in response to light intensity and CO_2 concentration, the data were automatically recorded every 20 s. g_s in response to ABA treatment was measured according to Ding *et al.* (2022).

Stomatal density and aperture

Stomatal density was determined by nail polish imprint with newly expanded leaves of 4-week-old maize plants grown in the greenhouse. The number of stomata was manually counted from photographs taken using a microscope (DMR, Leica, Germany) equipped with a CCD camera (ORCA-ER C4742-80, Hamamatsu, Japan).

Stomatal aperture was analyzed using 2-week-old maize plants grown in a phytotron according to Ding *et al.* (2022). In light-induced stomatal opening measurements, plants were placed in the dark for 1 h, leaves were then excised, and stomatal aperture was measured at the indicated times. For ABA-induced stomatal closure measurements, 5 μM ABA was applied. Stomatal aperture under different CO_2 treatments was measured according to Hu *et al.* (2010). To obtain stomatal opening buffer containing 400 ppm or 800 ppm CO_2, the buffer was aerated with 400 ppm or 800 ppm CO_2 for 20 min controlled by the Li-6400XT. To detect reactive oxygen species (H_2O_2) during stomatal closure induced by ABA and high CO_2, the peeled epidermis was incubated in the buffer containing 10 μM ABA or different CO_2 concentrations. The peels were stained by 10 μM H_2DCFDA according to Ding *et al.* (2022). The aperture size and fluorescence intensity were measured with ImageJ software (http://rsb.info.nih.gov/ij/).

Leaf temperature

Leaf temperature was measured using a thermal camera (FLIR, USA). Newly expanded leaves were excised from 2-week-old maize plants and incubated with 10 μM ABA (diluted with demineralized water) in a glass tube. At the indicated time, the leaves were imaged with a thermal camera. The average temperature at 3–6 cm near the leaf tip was measured by FLIR Tools (FLIR).

Expression in yeast

The optimized coding sequence of *ZmPIP1;6* (Heinen *et al.*, 2014) was amplified and expressed in *Saccharomyces cerevisiae* strain 31019b (*Δmep1-3*) with auxotrophy for uracil and leucine, according to Bienert and Chaumont (2014) and Bienert *et al.* (2018). The fragment was subcloned into the USER™ yeast expression vector pYeDP60u with a uracil selection marker and a galactose-inducible promoter. The constructs containing *ZmPIP2;5* (with a leucine selection marker and constitutive promoter), *ZmPIP2;5 H199K* mutant (with a leucine selection marker and constitutive promoter), and the empty vectors (with either a uracil or a leucine selection marker) were obtained from Bienert and Chaumont (2014). Empty vectors and the above-mentioned *PIP* constructs were co-transformed into the *Δmep1-3* strain by the one-step transformation method (Chen *et al.*, 1992). Colonies were selected on synthetic medium

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erae500/7928862 by guest on 26 January 2025

[2% (w/v) glucose, 50 mM succinic acid/Tris base, pH 6.5, 0.7% (w/v) yeast nitrogen base (YNB) without amino acids, and 0.2% (w/v) amino acid dropout mix without uracil and leucine (Reece-Hoyes and Walhout, 2018), 2% agar]. To induce PIP protein expression, 2% galactose was used to replace glucose. For H_2O_2 toxicity assays, different H_2O_2 concentrations were added to synthetic medium [2% (w/v) galactose, 50 mM succinic acid/Tris base, pH 6.5, 0.7% (w/v) YNB without amino acids, 0.3% (w/v) methionine, 0.3% (w/v) histidine, and 2% (w/v) agar], and an OD of 0.01 yeast culture was dropped on the plate. Growth was assessed after 7 d at 30 °C.

Phenotyping

The phenotyping of maize plants was performed at the WIWAM platform (https://www.wiwam.be/phenotyping-systems/wiwam-line/) (Van Hautegem *et al.*, 2024). The plants were grown in a chamber with a 16 h light/8 h dark cycle at a temperature of 26 °C (day)/21 °C (night) and a humidity of 55%, with an approximate light intensity of 170 µmol m^{-2} s^{-1}. This platform was equipped with automated weighing and irrigation capabilities, enabling the implementation of a pre-established irrigation schedule and the imposition of soil water deficit treatments. The well-watered scenario corresponded to no water deficit, with a soil water potential of approximately −0.01 MPa. The water deficit scenario corresponded to a soil water potential of −0.6 MPa. The plants were imaged automatically every day, and the leaf elongation rate (LER) was determined by measuring the daily length of the fourth leaf from the base of the plant to the tip of the fourth leaf until elongation ceased. The mean LER was calculated during the entire period of leaf elongation. The final fourth leaf length, width, and area, as well as the shoot dry weight were measured.

Statistical analysis

Statistical analyses were performed with GraphPad Prism 5 (GraphPad Software). Significant differences were determined either using the Student's *t*-test, where they are indicated by $*P<0.05$, $**P<0.01$, and $***P<0.001$; or by one-way ANOVA with Tukey's post-hoc test where they are indicated by different letters at the level of $P<0.05$.

Results

Characterization of transgenic maize lines and expression of *ZmPIP1;6* in stomata

ZmPIP1;6 is expressed in maize stomatal complexes, with higher expression during the day than at night (Heinen *et al.*, 2014). To further elucidate the role of ZmPIP1;6 in gas exchange and stomatal movement, *ZmPIP1;6* was overexpressed in the B104 maize line under the control of the *p35S* promoter, active in most plant tissues (Fig. 1A). The mRNA level of the transgene *ZmPIP1;6* in leaves was assessed by RT-qPCR, and two transgenic lines were selected for this study based on their expression levels in the mature zone of the leaf: one line with a relatively low level of *ZmPIP1;6* expression (OE-L) and the other with a high level (OE-H) (Fig. 1B). To examine the protein accumulation/abundance difference of ZmPIP1;6 between the WT and *ZmPIP1;6* OE, plasma membrane proteins were isolated from whole leaves and ZmPIP1;6 was immunodetected using antibodies raised against ZmPIP1;6 (Heinen *et al.*, 2014). A more intense signal was observed in both ZmPIP1;6 OE lines compared with the WT (Fig. 1C, D). The accumulation of ZmPIP1;6 increased ~1.6-fold in OE-L and ~3-fold in OE-H lines compared with WT plants. No significant difference in signal intensities of PIP2 proteins between WT and ZmPIP1;6 OE plants was observed (Supplementary Fig. S1).

To determine the cell expression pattern of *ZmPIP1;6* in maize leaves, a 3 kb DNA fragment located upstream of the *ZmPIP1;6* CDS, considered as the promoter, was cloned and used to drive the expression of *mYFP-ZmPIP1;6* (Fig. 1E). In stable transformed maize plants, the mYFP–ZmPIP1;6 fluorescent signal was detected by confocal microscopy in the leaf elongation zone and mature zone (Fig. 1F). In the elongation zone, the signal in immature stomatal complexes was much weaker compared with the signal in epidermal cells. In the mature zone, the mYFP signal was more intense in SCs than in GCs. The localization of mYFP–ZmPIP1;6 was affected by light conditions, with a more internalized signal notably around the nucleus probably corresponding to the endoplasmic reticulum, observed under dark conditions compared with stomatal complexes exposed to light (Fig. 1F).

Stomatal opening is enhanced by *ZmPIP1;6* overexpression

To determine the role of ZmPIP1;6 in photosynthesis, gas exchange was measured in *ZmPIP1;6* OE and WT plants. Firstly, the photosynthetic rate and the stomatal conductance (g_s) were measured under different light intensities with OE-H and WT plants. A significant increase in both parameters was observed in OE-H compared with the WT under a light intensity of 1500 µmol m^{-2} s^{-1} and 1800 µmol m^{-2} s^{-1}, respectively (Supplementary Fig. S2). For the following gas exchange measurements in this study, we used a light intensity of ~1500 µmol m^{-2} s^{-1}. g_s increased by 7% in OE-L and by 27% in OE-H plants in comparison with WT plants, although the increase in OE-H was not significant. Photosynthetic rate and transpiration were also higher in both *ZmPIP1;6* OE plants than in the WT (Fig. 2A–F). However, a significant difference was only observed between OE-L and WT plants, not in OE-H plants.

Stomatal aperture under light induction was compared between *ZmPIP1;6* OE and WT plants. Stomatal aperture was greater in both *ZmPIP1;6* OE lines than in WT plants after 30 min of light induction (Fig. 2G, H). The kinetics of g_s closure were also compared between WT and *ZmPIP1;6* OE plants, and similar curves were observed in response to light or dark (Supplementary Fig. S3).

CO_2- and ABA-induced stomatal closure is altered in *ZmPIP1;6* OE lines

The kinetics of stomatal movements induced by different [CO_2] might be affected by the accumulation of ZmPIP1;6,

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erae500/7928862 by guest on 26 January 2025

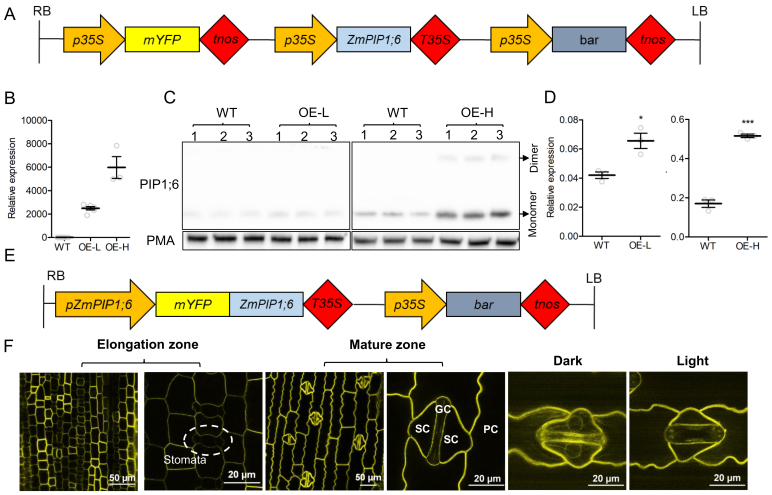


Fig. 1. Genetic constructs and expression of ZmPIP1;6 in the transgenic plants. (A and E) Schematic representation of the T-DNA used to overexpress *ZmPIP1;6* (A) and express *mYFP-ZmPIP1;6* (E) in the B104 maize line. RB, right border; *p*35S, CaMV 35S promoter; mYFP, sequence encoding the yellow fluorescent proteins; *tnos*, terminator of the *A. tumefaciens* nopaline synthase gene; t35S, CaMV 35S terminator; *bar*, the gene conferring resistance to bialaphos; LB, left border. (B) Relative expression of *ZmPIP1;6* transgene in leaves determined by RT-qPCR. (C) ZmPIP1;6 protein levels in WT and *ZmPIP1;6* OE lines (OE-L and OE-H). Plasma membrane protein was extracted from whole leaves (leaves were pooled from *n*>3 plants growing in a phytotron) and western blots were carried out using antibodies raised against ZmPIP1;6 and H$^+$-ATPase (PMA, loading control) (Heinen *et al.*, 2014). The upper and lower bands correspond to the dimers and monomers of ZmPIP1;6, respectively. (D) Western blot signal quantification. Data are expressed as the mean ±SE. Student's *t*-test was applied to compare the significant difference of protein signal between WT and OE plants at the level of **P*<0.05 and ****P*<0.001. (F) mYFP–ZmPIP1;6 fluorescence in different leaf zones of transgenic T_1 maize plants and in stomatal complexes in light and dark, detected by confocal microscopy. GC, guard cell; SC, subsidiary cell; PC, pavement cell.

which was shown to facilitate CO_2 diffusion across the plasma membrane when expressed in yeast (Heinen *et al.*, 2014). To test this hypothesis, the continuous recording of g_s was compared between WT, *ZmPIP1;6* OE, and *mYFP-ZmPIP1;6* plants under changing [CO_2] (Fig. 3A–F). At 400 ppm [CO_2], the g_s of OE-L, OE-H, and *mYFP-ZmPIP1;6* was higher than in WT plants (Fig. 3A, C, E). Upon switching [CO_2] from 100 ppm to 800 ppm, a faster decrease in g_s was observed in OE-H compared with the WT (Fig. 3C). To better visualize the change in g_s, the relative decrease of g_s was calculated (Fig. 3B, D, F), and to quantify this change in g_s, the time taken for g_s to decrease by half ($T_{1/2}$) was calculated (Fig. 3G). $T_{1/2}$ was significantly lower in OE-H plants than in the WT, while this difference was not significant in OE-L and *mYFP-ZmPIP1;6* plants (Fig. 3G).

To further analyze the response of single stomata to [CO_2], epidermal peels were incubated in buffers with different [CO_2] after the stomata were fully opened by light stimulation. Following a 15 min incubation with 800 ppm [CO_2], the stomatal aperture of both *ZmPIP1;6* OE lines exhibited a significant decrease compared with the 400 ppm [CO_2] treatment (Fig. 3H, I), while no significant difference was observed in the WT.

While we previously showed that ZmPIP2;5 played roles in ABA-induced stomatal closure (Ding *et al.*, 2022), no data were available regarding PIP1s. We therefore analyzed the stomatal closure under ABA treatments in *ZmPIP1;6* OE and WT lines (Fig. 4A, OE-L; Fig. 4B, OE-H). After 30 min of 5 μM ABA treatment, stomatal aperture decreased significantly in both *ZmPIP1;6* OE lines, while this reduction was significant in WT plants only after 60 min treatment (Fig. 4).

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erae500/7928862 by guest on 26 January 2025

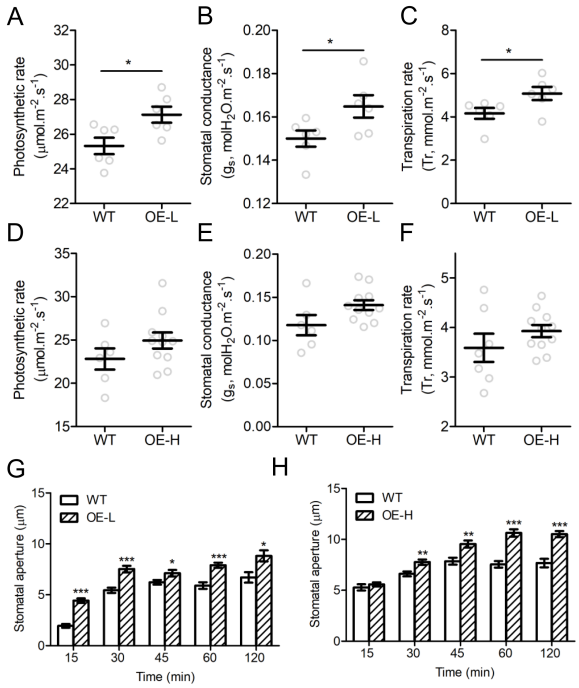


Fig. 2. Gas exchange parameters in WT and *ZmPIP1;6* OE plants. (A–F) Photosynthetic rate, g_s, and transpiration rate of plants growing in the greenhouse. Data are expressed as the mean ±SE with $n \geq 6$ plants. (G and H) Stomatal aperture after different times of light induction. Stomatal aperture was analyzed from $n > 50$ stomata from leaves taken from three plants growing in a phytotron. Student's *t*-test was applied to compare the significant difference of photosynthetic rate, g_s, and transpiration rate between WT and OE plants at the level of *$P<0.05$, **$P<0.01$, and ***$P<0.001$.

Detached leaves were incubated with 10 μM ABA, and leaf temperature and g_s were recorded in OE-H and WT plants. Thermal images were taken under ABA treatment for 0 min and 30 min (Fig. 5A). In the WT, leaf temperature increased significantly after 10 min of ABA treatment compared with the control treatment (Fig. 5B). Under control conditions, leaf temperature was significantly lower in OE-H than in the WT, while no difference was recorded between the WT and the previously characterized *ZmPIP2;5 OE* plants (Ding *et al.*, 2020, 2022) (Fig. 5C). Following ABA treatment for 30 min, a significant increase in leaf temperature was observed in both OE-H and *PIP2;5* OE plants when compared with WT plants. Furthermore, a more rapid decrease in g_s was observed in OE-H plants compared with the WT under ABA treatment (Fig. 5D, E). Additionally, the decrease in g_s was slightly faster in *mYFP-ZmPIP1;6* than in the WT after ABA treatment (Supplementary Fig. S4).

The above-mentioned faster stomatal closure may be beneficial due to the high density of stomata (Faralli *et al.*, 2019). To test this hypothesis, we compared stomatal density in *ZmPIP1;6* OE and WT plants. The abaxial stomatal density significantly increased by

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erae500/7928862 by guest on 26 January 2025

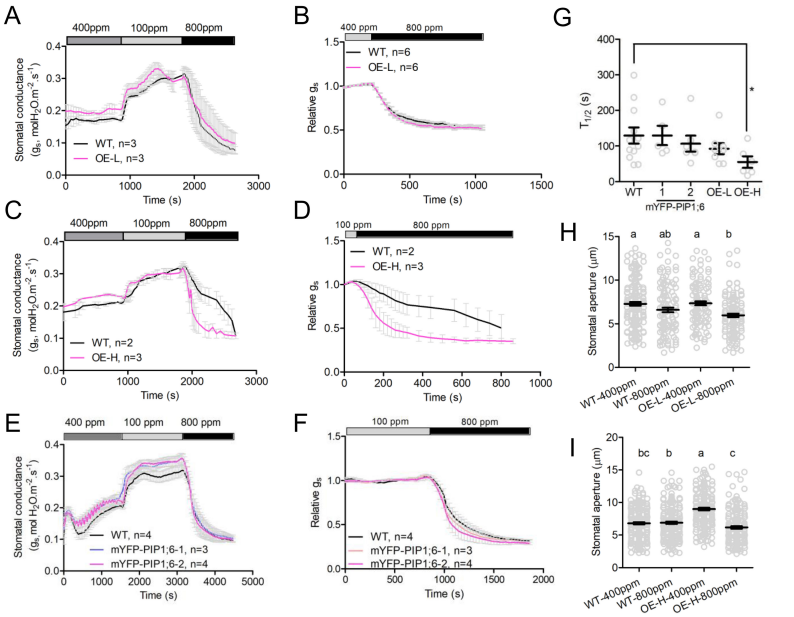


Fig. 3. Kinetics of g_s in response to changing $[CO_2]$ in WT, *ZmPIP1;6* OE (OE-L and OE-H), and *mYFP-ZmPIP1;6* (two lines of *mYFP-ZmPIP1;6-1* and *mYFP-ZmPIP1;6-2*) maize lines. (A, C, and E) Recording of g_s under 400, 100, and 800 ppm $[CO_2]$. (B, D, and F) Relative decrease in g_s when $[CO_2]$ increased from 100 ppm to 800 ppm. (G) Half-time ($T_{1/2}$) of the g_s decrease induced by an increase in $[CO_2]$. Data are expressed as the mean ±SE from three independent experiments with plants growing in the greenhouse. Student's *t*-test was applied to compare the significant difference of stomatal aperture between WT and OE plants at the level of **P*<0.05. (H and I) Comparison of stomatal aperture after 15 min incubation with 400 ppm or 800 ppm $[CO_2]$. Data are expressed as the mean ±SE with $n \geq 110$ stomata from three plants growing in a phytotron. One-way ANOVA with Tukey's post-test was applied to detect the significant difference at the *P*<0.05 level between genotypes and treatments. Different letters indicate significant differences.

10% and 7% in *ZmPIP1;6* OE-L and OE-H, respectively, compared with the WT, while no difference in adaxial stomatal density was observed (Fig. 6A). Similar higher abaxial stomatal density was measured in both *mYFP-ZmPIP1;6*-expressing lines (Fig. 6B).

H_2O_2 accumulates more rapidly in *ZmPIP1;6* OE stomata upon ABA treatments, but not by high $[CO_2]$ treatments

It was reported that ABA-induced stomatal closure also depends on H_2O_2 entry into GCs via its facilitated diffusion through AtPIP2;1 or ZmPIP2;5 (Rodrigues *et al.* 2017; Ding *et al.*, 2022). To determine whether the faster stomatal closure in *ZmPIP1;6* OE plants was related to the adjustment of cellular H_2O_2 levels, H_2O_2 levels in GCs were assessed by staining with H_2DCFDA under ABA and CO_2 treatments. Under 10 μM ABA treatment for 5 min, a significant increase in fluorescence signal was measured in OE-H GCs compared with the control treatment (Fig. 7A, B, E, F, J), whereas the signal decreased significantly in WT GCs (Fig. 7I). This decrease in H_2O_2 signal in WT GCs after 5 min of ABA treatment was previously observed in *A. thaliana* and maize (Rodrigues *et al.*,

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erae500/7928862 by guest on 26 January 2025

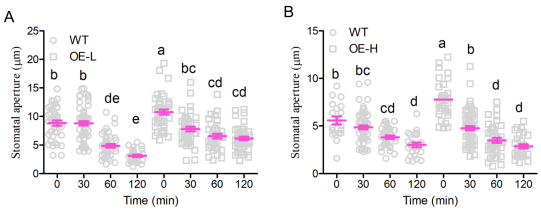


Fig. 4. Stomatal aperture of WT and *ZmPIP1;6* OE plants. Stomatal aperture of WT and (A) OE-L and (B) OE-H lines upon 10 μM ABA treatment. Data are expressed as the mean ±SE, with $n \geq 19$ stomata from three plants growing in a phytotron. One-way ANOVA with Tukey's post-hoc test was applied to detect significant differences at the $P<0.05$ level between genotypes and treatments. Different letters indicate significant differences.

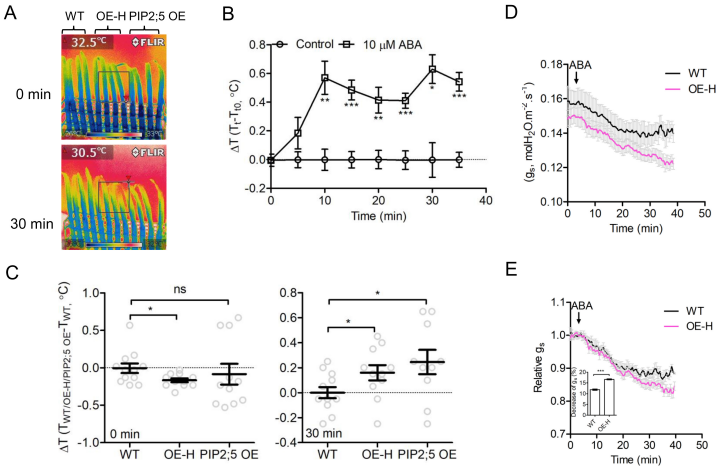


Fig. 5. Leaf temperature and g_s of WT, *ZmPIP1;6* OE, and *ZmPIP2;5* OE plants upon ABA treatment. (A) Thermal images under ABA treatment for 0 min and 30 min. (B) Change in leaf temperature after adding ABA in WT plants. (C) Difference in temperature between *ZmPIP1;6* OE, *ZmPIP2;5* OE, and WT plants after ABA treatment for 0 min and 30 min. (D and E) Change of g_s upon ABA treatment in WT and OE-H plants. Data are expressed as the mean ±SE. Three leaves were measured for each treatment in (B) and 10–12 leaves were measured for each genotype in (C). Six WT and nine OE-H leaves were measured in (D) and (E). Plants were grown in a phytotron. Student's *t*-test was applied to compare the significant difference of temperature and g_s between WT and OE plants at the level of $^*P<0.05$, $^{**}P<0.01$, and $^{***}P<0.001$. ns, not significant.

2017; Ding *et al.*, 2022). After 15 min of ABA treatment, an increase in signal was observed in both OE-H and WT GCs compared with the control treatment (Fig. 7C, D, G–J).

Following a 2 min treatment with 800 ppm [CO_2], a higher fluorescence signal was observed in WT stomata compared with those treated with 400 ppm [CO_2], whereas a lower signal was observed in OE-H stomata than in the WT (Fig. 7K). After 5 min treatment with 800 ppm [CO_2], no significant difference in signal intensity was observed between WT and OE-H plants compared with the 400 ppm [CO_2] treatment (Fig. 7L).

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erae500/7928862 by guest on 26 January 2025

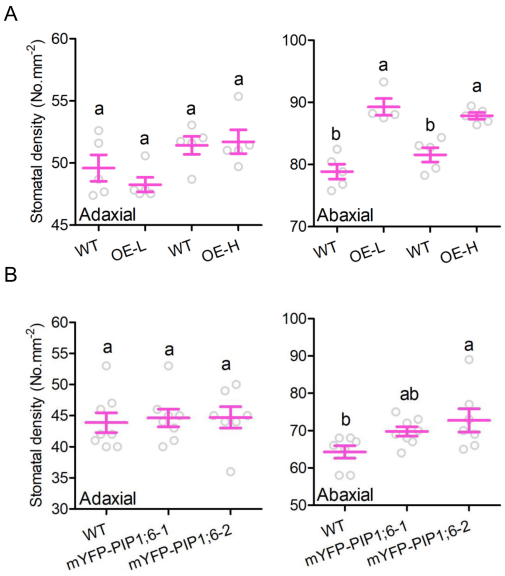


Fig. 6. Stomatal density in adaxial and abaxial sides of WT, *ZmPIP1;6* OE (OE-L and OE-H), and *mYFP-ZmPIP1;6* (two lines of *mYFP-ZmPIP1;6-1* and *mYFP-ZmPIP1;6-2*) new fully expanded leaves. (A) WT and *ZmPIP1;6* OE plants. (B) WT and *mYFP-ZmPIP1;6* plants. Data are expressed as the mean ±SE of $n \geq 5$ plants growing in the greenhouse. One-way ANOVA with Tukey's post-hoc test was applied to detect significant differences at the $P<0.05$ level. Different letters indicate significant differences.

Furthermore, we investigated whether ZmPIP1;6 facilitated the H_2O_2 diffusion in yeast. We previously demonstrated that PIP1s must be co-expressed with an inactive PIP2 isoform (i.e. ZmPIP2;5H199K) to reach the yeast plasma membrane and assess its H_2O_2 transport capacity (Bienert *et al.*, 2014). Accordingly, we co-expressed ZmPIP1;6 with ZmPIP2;5H199K and observed minimal inhibition of yeast growth on plates containing 2 mM H_2O_2, whereas yeast expressing WT ZmPIP2;5 exhibited pronounced growth inhibition (Supplementary Fig. S5). The growth of the yeast strains containing the empty vectors or the empty vector in addition to the vector expressing ZmPIP2;5H199K or ZmPIP1;6 was not impacted by the presence of 2 mM H_2O_2. The results indicate that ZmPIP1;6 is not an active channel for H_2O_2 transport when expressed in yeast.

Plant growth is affected in *ZmPIP1;6* OE plants under water deficit conditions

The opening and closure of stomata are affected in *ZmPIP1;6* OE plants, a phenotype that may influence plant behavior in response to water deficit. To determine the difference in plant growth between *ZmPIP1;6* OE and WT plants, we compared the LER of the fourth leaf, its length, width, and area, and the shoot dry weight under both control and water deficit conditions (–0.6 MPa). In WT plants, the LER, final length, width, and area, and the shoot dry weight decreased significantly under water deficit conditions in comparison with the control (Fig. 8A–E). Interestingly, this decrease was less pronounced and not statistically significant in both *ZmPIP1;6* OE lines, except for width and area in OE-L plants (Fig. 8C, D) and length in OE-H plants (Fig. 8B). These data indicate that

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erae500/7928862 by guest on 26 January 2025

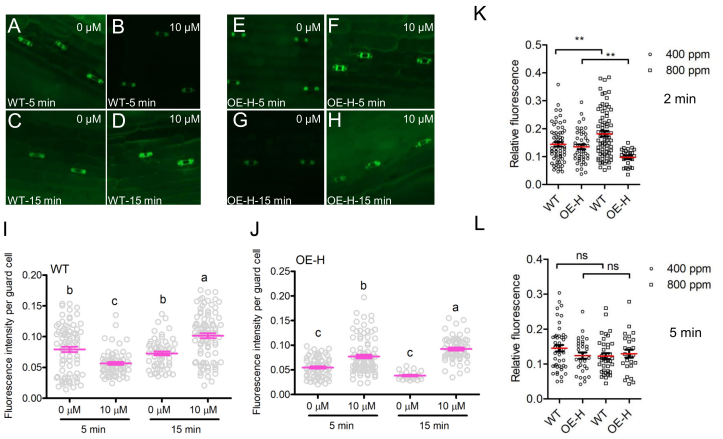


Fig. 7. Reactive oxygen species (ROS) in stomata under ABA treatment and under 400 ppm and 800 ppm CO_2 treatment. (A–H) The fluorescence signal of ROS in WT and OE-H stomata after ABA treatment for 5 min and 15 min. (I and J) The comparison of fluorescence intensity in GCs. (K and L) Relative fluorescence under 400 ppm and 800 ppm CO_2 treatment after 2 min (K) and 5 min (L). Data are expressed as the mean ±SE with $n \geq 20$ stomata from three plants growing in a phytotron, and individual data are shown in (H) and (I). One-way ANOVA with Tukey's post-hoc test was applied to detect significant differences at the $P<0.05$ level between treatments and time. Different letters indicate significant differences.

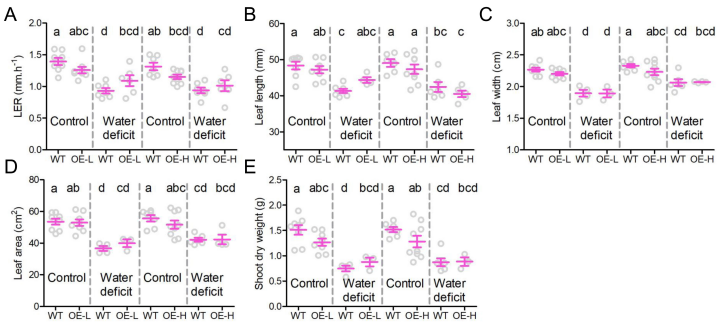


Fig. 8. The difference in plant growth between *ZmPIP1;6* OE and WT plants. Leaf elongation rate (A) of the fourth leaf, its final length (B), width (C), area (D), and shoot dry weight (E) under control and water-deficit treatments in *ZmPIP1;6* OE (OE-L and OE-H) and their respective WT plants. The values indicate means ±SE (n=3–9 plants). One-way ANOVA with Tukey's post-hoc test was applied to detect significant differences at the $P<0.05$ level. Different letters indicate significant differences.

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erae500/7928862 by guest on 26 January 2025

ZmPIP1;6 OE could mitigate the reduction in maize plant growth in water deficit conditions by a better regulation of stomata closure.

Discussion

ZmPIP1;6 expression affects gas exchange

In C_3 species, PIP aquaporins have been shown to play important roles in regulating photosynthesis, by affecting CO_2 transport conductance; that is, stomatal and mesophyll conductance (Uehlein *et al.*, 2003; Flexas *et al.*, 2006; Evans *et al.*, 2009; Kaldenhoff, 2012). In contrast, in C_4 species such as maize, the photosynthetic rate is less restricted by the CO_2 diffusion resistance, due to the mechanism of CO_2 concentration in bundle sheath cells. In this study, we showed that the photosynthetic rate, g_s, stomatal aperture, and stomatal density increased in *ZmPIP1;6* OE lines (Figs 2, 6). The high photosynthetic rate and g_s might result from increased stomatal aperture and density in *ZmPIP1;6* OE plants compared with the WT. Indeed, gas exchange was shown to be affected by deregulation of PIP expression in plants, by regulating g_s, stomatal aperture, and density (Maurel *et al.*, 2016; Ding and Chaumont, 2020; Ding *et al.*, 2024). In maize, *ZmPIP2;5* OE plants transpired more than WT plants, which is due to the increase in cell and tissue hydraulic conductivity (Ding *et al.*, 2020, 2022). In fact, gas exchange is enhanced by maintaining plant hydraulic conductivity. The water channel activity of ZmPIP1;6 was demonstrated after its co-expression with ZmPIP2;1 in *Xenopus* oocytes and swelling assays (Heinen *et al.*, 2014). In *ZmPIP1;6* OE maize plants, the hydraulic conductivity of cells and tissues might increase, further affecting gas exchange and stomatal behavior. In this study, stomatal aperture was greater in *ZmPIP1;6* OE plants than in the WT (Fig. 2). We hypothesized that turgor pressure may be higher in *ZmPIP1;6* OE GCs than in WT GCs. GC turgor pressure is regulated by a coupled inflow of ions and water coming from the SCs or other neighboring cells (Ding *et al.*, 2024). How *ZmPIP1;6* OE would increase the turgor pressure remains to be elucidated. Several PIPs have been shown to transport ions such as Na^+ (Byrt *et al.*, 2017; Tyerman *et al.*, 2021; Ahmed *et al.*, 2024) and, therefore, this channel activity may contribute to regulate GC turgor pressure, even though the cation fluxes associated with aquaporins are several orders of magnitude lower than those required and provided via other transporters (Ding *et al.*, 2022, 2024).

Expression of *mYFP-ZmPIP1;6* under the control of its own promoter (~3 kb fragment) showed that the protein is expressed in the epidermis of leaf elongation and mature zones (Fig. 1). Interestingly, the mYFP–ZmPIP1;6 abundance was very low in developing stomatal complexes but high in mature SCs. In addition, internalization of the protein was observed under dark conditions. This result was in accordance with *ZmPIP1;6* mRNA levels that were shown to be high during the day and low at night (Heinen *et al.*, 2014). However, little is known about how SCs regulate stomatal movement, but it was proposed that SCs serve as a source–sink of osmolytes to regulate the turgor pressure of GCs (Nunes *et al.*, 2020). Here, the high expression of ZmPIP1;6 in SCs suggests that aquaporin-regulated substrate transport in these specific cells plays pivotal roles in influencing stomatal movements. In Arabidopsis, the potassium channel (KC1) expressed in pavement cells tuned the distribution of K^+ in GCs, further affecting stomatal movements (Nieves-Cordones *et al.*, 2022). This result supports the hypothesis that SCs or pavement cells present as a source–sink of osmolytes to regulate stomatal movements.

Role of ZmPIP1;6 in CO_2- and ABA-induced stomatal closure

CO_2-triggered stomatal movement is a consequence of the regulation in both GCs and mesophyll cells (Zhang *et al.*, 2018). In GCs, CO_2 enters the cells by either membrane simple diffusion or facilitated diffusion through PIPs, such as AtPIP2;1 in Arabidopsis (Wang *et al.*, 2016). In the presence of a high CO_2 concentration, AtPIP2;1 interacts with carbonic anhydrase (βCA4), which catalyzes the production of bicarbonate, and further activates the downstream signaling cascade to induce stomatal closure. However, the kinetics of stomatal movement were not different in the Arabidopsis *pip2;1* mutant from that in WT plants, which was thought to be due to the redundancy of aquaporin genes in stomata (Wang *et al.*, 2016). ZmPIP1;5 and ZmPIP1;6 were also shown to facilitate the transport of CO_2 when they were expressed in yeast (Heinen *et al.*, 2014). However, direct evidence is still missing that CO_2-induced stomatal movement was disrupted by the deregulation of PIPs aquaporin *in planta*. Here, we found that *ZmPIP1;6* OE plants exhibited rapid stomatal closure when the CO_2 concentration was increased (Fig. 3). This was not observed in *mYFP-ZmPIP1;6*-expressing plants driven by the *ZmPIP1;6* promoter. In this case, the transgene was mostly expressed in SCs, while in *ZmPIP1;6* OE, the transgene was expressed under the control of the 35S promoter that is active in both GCs and SCs (Ding *et al.*, 2022). Together, these observations suggest that the *ZmPIP1;6* expression level and/or location was important to regulate CO_2-triggered stomatal movement.

PIPs also play important roles in regulating ABA-induced stomatal closure. Our previous data demonstrated that ZmPIP2;5 enhanced the sensitivity of stomatal closure to water deficit by regulating leaf and/or stomatal hydraulics and H_2O_2 transport in GCs (Ding *et al.*, 2022). Here, a similar phenotype was found in *ZmPIP1;6* OE plants. H_2O_2 accumulated faster in *ZmPIP1;6* OE GCs than in those of the WT, suggesting that ZmPIP1;6 might facilitate the transport of H_2O_2 (Fig. 7). However, ZmPIP1;6 did not seem to be a very efficient H_2O_2 channel when it was expressed in yeast (Supplementary Fig. S5). Aquaporins form homo- or heterotetramers on the plasma

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erae500/7928862 by guest on 26 January 2025

membrane, and PIP1s, including ZmPIP1;6, need to physically interact with PIP2s to reach the plasma membrane (Zelazny *et al.*, 2007). A faster H_2O_2 accumulation in *ZmPIP1;6* OE stomata could be due to specific membrane enrichment of PIP2s, which are known to be efficient H_2O_2 channels (Bienert *et al.*, 2014). In Arabidopsis, ABA activates OST1 kinase that phosphorylates Ser121 of AtPIP2;1, resulting in the activation of water and H_2O_2 channeling to contribute to stomatal closure (Grondin *et al.*, 2015; Rodrigues *et al.*, 2017).

CO_2-triggered stomatal closure was reported to be both ABA dependent and independent (Chater *et al.*, 2015; Hsu *et al.*, 2018). These contrasting results might be explained by difference in the methods of g_s and stomatal aperture measurements using the related genetic mutants. However, in Hsu *et al.* (2018), ABA signaling amplified elevated CO_2-induced stomatal closure. Here, we showed that both CO_2- and ABA-induced stomatal closure was enhanced by *ZmPIP1;6* OE in maize stomata, indicating that common regulation mechanisms might be shared by both inducers of stomatal closure. We checked if H_2O_2 was involved in CO_2-induced stomatal closure and observed that the H_2O_2 concentration in *ZmPIP1;6* OE GCs was not increased by 800 ppm CO_2 treatment for 2 min, contrary to what was observed in WT GCs (Fig. 7). This result suggests that high [CO_2]-induced rapid stomatal closure in *ZmPIP1;6* OE plants may be independent of H_2O_2 regulation.

Stomatal density was higher in *ZmPIP1;6* OE and *mYFP-ZmPIP1;6*-expressing maize plants than in the WT (Fig. 6), and it was proposed that high stomatal density resulted in a faster change in g_s (Faralli *et al.*, 2019). Therefore, faster CO_2- and ABA-induced stomatal movement might also be explained by higher stomatal density in *ZmPIP1;6* OE plants than in the WT. We hypothesized that *ZmPIP1;6* OE maize plants increased their stomatal density to compensate the high dynamics of the stomatal movement that could be detrimental in certain conditions and/or specific developmental stages. In fact, stomatal density development is complex and regulated by multiple genetic factors (Hepworth *et al.*, 2015) and environmental stimuli (Chater *et al.*, 2015). Altered leaf hydraulics or carbon fixation by deregulating aquaporin expression can possibly affect stomatal development.

We have previously shown that LER was higher in *ZmPIP2;5* OE maize lines than in WT plants under mild water deficit conditions, whereas this difference was not observed under well- watered conditions (Ding *et al.*, 2020). We proposed that the changes in root and leaf hydraulic conductance observed by deregulating *ZmPIP2;5* at the cellular level resulted in an overall change in whole-plant hydraulic fluxes (and leaf water potential) that affects the LER in stress conditions (Parent *et al.*, 2009; Caldeira *et al.*, 2014; Ding *et al.*, 2020). Here, LER was lower, but not significantly different, in both *ZmPIP1;6* OE lines than in the WT under control conditions, whereas this trend was opposite under water deficit conditions (Fig. 8A). Under control conditions, the greater stomatal opening in *ZmPIP1;6* OE plants than in WT plants may induce a more negative leaf water potential leading to a decrease in LER. With rapid stomatal closure induced by ABA (drought) treatments, *ZmPIP1;6* OE plants may prevent excessive water loss and maintain a higher leaf water potential and LER compared with WT plants, a phenotype similar to what was observed in *ZmPIP2;5* OE, where stomata closed more rapidly than in WT plants under ABA and drought treatments, leading to the maintenance of a high LER under water deficit conditions (Ding *et al.*, 2022). Finally, plant growth was therefore improved by this stomatal regulation under water deficit conditions in *ZmPIP1;6* OE plants (Fig. 8).

Supplementary data

The following supplementary data are available at *JXB* online.

Table S1. List of primers used for the genetic constructs and RT-qPCR.

Fig. S1. PIP2 protein expression in WT and *ZmPIP1;6* OE plants.

Fig. S2. Photosynthetic rate and g_s in response to changing light intensity in WT and OE-H plants.

Fig. S3. Change of g_s in response to lights on/off.

Fig. S4. Relative decrease in g_s in WT and mYFP–ZmPIP1;6 plants.

Fig. S5. H_2O_2 toxicity assay in yeast expressing ZmPIP1;6, ZmPIP2;5, and the ZmPIP2;5 H199K mutant.

Author contributions

LD, TM, and FC: designed the experiments; LD, ML, TM, and SA: performed the experiments; LP, MVL, and D I: supervised the maize transformation; HN: was in charge of the drought and growth assays; LD, ML, TM, and FC: data analysis; LD and FC: wrote the manuscript. All the authors contributed to the discussion and revision of the manuscript.

Conflict of interest

No conflict of interest declared.

Funding

This work was supported by the Belgian National Fund for Scientific Research (FNRS), the Interuniversity Attraction Poles Programme-Belgian Science Policy (grant IAP7/29), and the 'Communauté française de Belgique-Actions de Recherches Concertées' (grants ARC11/16-036 and ARC16/21-075). LD was a post-doctoral researcher of the FRS-FNRS.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erae500/7928862 by guest on 26 January 2025

References

Aesaert S, Impens L, Coussens G, *et al*. 2022. Optimized transformation and gene editing of the B104 public maize inbred by improved tissue culture and use of morphogenic regulators. Frontiers in Plant Science **13**, 883847.

Aharon R, Shahak Y, Wininger S, Bendov R, Kapulnik Y, Galili G. 2003. Overexpression of a plasma membrane aquaporin in transgenic tobacco improves plant vigor under favorable growth conditions but not under drought or salt stress. The Plant Cell **15**, 439–447.

Ahmed J, Ismail A, Ding L, Yool AJ, Chaumont F. 2024. A new method to measure aquaporin-facilitated membrane diffusion of hydrogen peroxide and cations in plant suspension cells. Plant, Cell & Environment **47**, 527–539.

Berny MC, Gilis D, Rooman M, Chaumont F. 2016. Single mutations in the transmembrane domains of maize plasma membrane aquaporins affect the activity of monomers within a heterotetramer. Molecular Plant **9**, 986–1003.

Bi Z, Merl-Pham J, Uehlein N, *et al*. 2015. RNAi-mediated downregulation of poplar plasma membrane intrinsic proteins (PIPs) changes plasma membrane proteome composition and affects leaf physiology. Journal of Proteomics **128**, 321–332.

Bienert GP, Chaumont F. 2014. Aquaporin-facilitated transmembrane diffusion of hydrogen peroxide. Biochimica et Biophysica Acta **1840**, 1596–1604.

Bienert GP, Heinen RB, Berny MC, Chaumont F. 2014. Maize plasma membrane aquaporin ZmPIP2;5, but not ZmPIP1;2, facilitates transmembrane diffusion of hydrogen peroxide. Biochimica et Biophysica Acta **1838**, 216–222.

Bienert MD, Diehn TA, Richet N, Chaumont F, Bienert GP. 2018. Heterotetramerization of plant PIP1 and PIP2 aquaporins is an evolutionary ancient feature to guide PIP1 plasma membrane localization and function. Frontiers in Plant Science **9**, 382.

Bright J, Desikan R, Hancock JT, Weir IS, Neill SJ. 2006. ABA-induced NO generation and stomatal closure in Arabidopsis are dependent on H_2O_2 synthesis. The Plant Journal **45**, 113–122.

Byrt CS, Zhao M, Kourghi M, *et al*. 2017. Non-selective cation channel activity of aquaporin AtPIP2;1 regulated by Ca^{2+} and pH. Plant, Cell & Environment **40**, 802–815.

Caldeira CF, Bosio M, Parent B, Jeanguenin L, Chaumont F, Tardieu F. 2014. A hydraulic model is compatible with rapid changes in leaf elongation under fluctuating evaporative demand and soil water status. Plant Physiology **164**, 1718–1730.

Chater C, Peng K, Movahedi M, *et al*. 2015. Elevated CO_2-induced responses in stomata require ABA and ABA signaling. Current Biology **25**, 2709–2716.

Chater CC, Caine RS, Fleming AJ, Gray JE. 2017. Origins and evolution of stomatal development. Plant Physiology **174**, 624–638.

Chaumont F, Barrieu F, Jung R, Chrispeels MJ. 2000. Plasma membrane intrinsic proteins from maize cluster in two sequence subgroups with differential aquaporin activity. Plant Physiology **122**, 1025–1034.

Chaumont F, Tyerman SD. 2014. Aquaporins: highly regulated channels controlling plant water relations. Plant Physiology **164**, 1600–1618.

Chen DC, Yang BC, Kuo TT. 1992. One-step transformation of yeast in stationary phase. Current Genetics **21**, 83–84.

Coussens G, Aesaert S, Verelst W, Demeulenaere M, De Buck S, Njuguna E, Inze D, Van Lijsebettens M. 2012. *Brachypodium distachyon* promoters as efficient building blocks for transgenic research in maize. Journal of Experimental Botany **63**, 4263–4273.

Cui XH, Hao FS, Chen H, Chen J, Wang XC. 2008. Expression of the *Vicia faba* VfPIP1 gene in *Arabidopsis thaliana* plants improves their drought resistance. Journal of Plant Research **121**, 207–214.

Ding L, Chaumont F. 2020. Are aquaporins expressed in stomatal complexes promising targets to enhance stomatal dynamics? Frontiers in Plant Science **11**, 458.

Ding L, Fox AR, Chaumont F. 2024. Multifaceted role and regulation of aquaporins for efficient stomatal movements. Plant, Cell & Environment **47**, 3330–3343.

Ding L, Milhiet T, Couvreur V, *et al*. 2020. Modification of the expression of the aquaporin ZmPIP2;5 affects water relations and plant growth. Plant Physiology **182**, 2154–2165.

Ding L, Milhiet T, Parent B, Meziane A, Tardieu F, Chaumont F. 2022. The plasma membrane aquaporin ZmPIP2;5 enhances the sensitivity of stomatal closure to water deficit. Plant, Cell & Environment **45**, 1146–1156.

Ding X, Iwasaki I, Kitagawa Y. 2004. Overexpression of a lily PIP1 gene in tobacco increased the osmotic water permeability of leaf cells. Plant, Cell & Environment **27**, 177–186.

Evans JR, Kaldenhoff R, Genty B, Terashima I. 2009. Resistances along the CO_2 diffusion pathway inside leaves. Journal of Experimental Botany **60**, 2235–2248.

Faralli M, Matthews J, Lawson T. 2019. Exploiting natural variation and genetic manipulation of stomatal conductance for crop improvement. Current Opinion in Plant Biology **49**, 1–7.

Fetter K, Van Wilder V, Moshelion M, Chaumont F. 2004. Interactions between plasma membrane aquaporins modulate their water channel activity. The Plant Cell **16**, 215–228.

Flexas J, Ribas-Carbó M, Hanson DT, Bota J, Otto B, Cifre J, McDowell N, Medrano H, Kaldenhoff R. 2006. Tobacco aquaporin NtAQP1 is involved in mesophyll conductance to CO_2 *in vivo*. The Plant Journal **48**, 427–439.

Franks PJ, Buckley TN, Shope JC, Mott KA. 2001. Guard cell volume and pressure measured concurrently by confocal microscopy and the cell pressure probe. Plant Physiology **125**, 1577–1584.

Fraysse LC, Wells B, McCann MC, Kjellbom P. 2005. Specific plasma membrane aquaporins of the PIP1 subfamily are expressed in sieve elements and guard cells. Biology of the Cell **97**, 519–534.

Grondin A, Rodrigues O, Verdoucq L, Merlot S, Leonhardt N, Maurel C. 2015. Aquaporins contribute to ABA-triggered stomatal closure through OST1-mediated phosphorylation. The Plant Cell **27**, 1945–1954.

Hachez C, Heinen RB, Draye X, Chaumont F. 2008. The expression pattern of plasma membrane aquaporins in maize leaf highlights their role in hydraulic regulation. Plant Molecular Biology **68**, 337–353.

Hachez C, Milhiet T, Heinen RB, Chaumont F. 2017. Roles of aquaporins in stomata. In: Chaumont F, Tyerman SD, eds. Plant aquaporins. Cham: Springer, 167–183.

Hachez C, Moshelion M, Zelazny E, Cavez D, Chaumont F. 2006. Localization and quantification of plasma membrane aquaporin expression in maize primary root: a clue to understanding their role as cellular plumbers. Plant Molecular Biology **62**, 305–323.

Hanba YT, Shibasaka M, Hayashi Y, Hayakawa T, Kasamo K, Terashima I, Katsuhara M. 2004. Overexpression of the barley aquaporin HvPIP2;1 increases internal CO_2 conductance and CO_2 assimilation in the leaves of transgenic rice plants. Plant and Cell Physiology **45**, 521–529.

Hebelstrup KH, Christiansen MW, Carciofi M, Tauris B, Brinch-Pedersen H, Holm PB. 2010. UCE: a uracil excision (USER™)-based toolbox for transformation of cereals. Plant Methods **6**, 15.

Heinen RB, Bienert GP, Cohen D, Chevalier AS, Uehlein N, Hachez C, Kaldenhoff R, Le Thiec D, Chaumont F. 2014. Expression and characterization of plasma membrane aquaporins in stomatal complexes of *Zea mays*. Plant Molecular Biology **86**, 335–350.

Hepworth C, Doheny-Adams T, Hunt L, Cameron DD, Gray JE. 2015. Manipulating stomatal density enhances drought tolerance without deleterious effect on nutrient uptake. New Phytologist **208**, 336–341.

Hsu PK, Takahashi Y, Munemasa S, Merilo E, Laanemets K, Waadt R, Pater D, Kollist H, Schroeder JI. 2018. Abscisic acid-independent stomatal CO_2 signal transduction pathway and convergence of CO_2 and ABA signaling downstream of OST1 kinase. Proceedings of the National Academy of Sciences, USA **115**, E9971–E9980.

Hu H, Boisson-Dernier A, Israelsson-Nordström M, Böhmer M, Xue S, Ries A, Godoski J, Kuhn JM, Schroeder JI. 2010. Carbonic

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erae500/7928862 by guest on 26 January 2025

anhydrases are upstream regulators of CO_2-controlled stomatal movements in guard cells. Nature Cell Biology **12**, 87–93.

Hua D, Wang C, He J, Liao H, Duan Y, Zhu Z, Guo Y, Chen Z, Gong Z. 2012. A plasma membrane receptor kinase, GHR1, mediates abscisic acid- and hydrogen peroxide-regulated stomatal movement in Arabidopsis. The Plant Cell **24**, 2546–2561.

Kaldenhoff R. 2012. Mechanisms underlying CO_2 diffusion in leaves. Current Opinion in Plant Biology **15**, 276–281.

Kaldenhoff R, Kölling A, Meyers J, Karmann U, Ruppel G, Richter G. 1995. The blue light-responsive AthH2 gene of *Arabidopsis thaliana* is primarily expressed in expanding as well as in differentiating cells and encodes a putative channel protein of the plasmalemma. The Plant Journal **7**, 87–95.

Karimi M, Inzé D, Van Lijsebettens M, Hilson P. 2013. Gateway vectors for transformation of cereals. Trends in Plant Science **18**, 1–4.

Larsson C, Widell S, Kjellbom P. 1987. Preparation of high-purity plasma membranes. Methods in Enzymology **148**, 558–568.

Li J, Yu G, Sun X, Liu Y, Liu J, Zhang X, Jia C, Pan H. 2015. AcPIP2, a plasma membrane intrinsic protein from halophyte *Atriplex canescens*, enhances plant growth rate and abiotic stress tolerance when overexpressed in *Arabidopsis thaliana*. Plant Cell Reports **34**, 1401–1415.

Maurel C, Verdoucq L, Rodrigues O. 2016. Aquaporins and plant transpiration. Plant, Cell & Environment **39**, 2580–2587.

McLachlan DH, Kopischke M, Robatzek S. 2014. Gate control: guard cell regulation by microbial stress. New Phytologist **203**, 1049–1063.

Melotto M, Underwood W, Koczan J, Nomura K, He SY. 2006. Plant stomata function in innate immunity against bacterial invasion. Cell **126**, 969–980.

Nieves-Cordones M, Azeem F, Long Y, *et al*. 2022. Non-autonomous stomatal control by pavement cell turgor via the K^+ channel subunit AtKC1. The Plant Cell **34**, 2019–2037.

Nunes TD, Zhang D, Raissig MT. 2020. Form, development and function of grass stomata. The Plant Journal **101**, 780–799.

Parent B, Hachez C, Redondo E, Simonneau T, Chaumont F, Tardieu F. 2009. Drought and abscisic acid effects on aquaporin content translate into changes in hydraulic conductivity and leaf growth rate: a trans-scale approach. Plant Physiology **149**, 2000–2012.

Pei ZM, Murata Y, Benning G, Thomine S, Klüsener B, Allen GJ, Grill E, Schroeder JI. 2000. Calcium channels activated by hydrogen peroxide mediate abscisic acid signalling in guard cells. Nature **406**, 731–734.

Raghavendra AS, Gonugunta VK, Christmann A, Grill E. 2010. ABA perception and signalling. Trends in Plant Science **15**, 395–401.

Reece-Hoyes JS, Walhout AJM. 2018. Generating yeast one-hybrid DNA-bait strains. Cold Spring Harbor Protocols **2018**, doi: 10.1101/pdb.prot094961.

Rodrigues O, Reshetnyak G, Grondin A, Saijo Y, Leonhardt N, Maurel C, Verdoucq L. 2017. Aquaporins facilitate hydrogen peroxide entry into guard cells to mediate ABA- and pathogen-triggered stomatal closure. Proceedings of the National Academy of Sciences, USA **114**, 9200–9205.

Sarda X, Tousch D, Ferrare K, Legrand E, Dupuis J, Casse-Delbart F, Lamaze T. 1997. Two TIP-like genes encoding aquaporins are expressed in sunflower guard cells. The Plant Journal **12**, 1103–1111.

Shimazaki KI, Doi M, Assmann SM, Kinoshita T. 2007. Light regulation of stomatal movement. Annual Review of Plant Biology **58**, 219–247.

Sussmilch FC, Schultz J, Hedrich R, Roelfsema MRG. 2019. Acquiring control: the evolution of stomatal signalling pathways. Trends in Plant Science **24**, 342–351.

Tyerman SD, McGaughey SA, Qiu J, Yool AJ, Byrt CS. 2021. Adaptable and multifunctional ion-conducting aquaporins. Annual Review of Plant Biology **72**, 703–736.

Uehlein N, Lovisolo C, Siefritz F, Kaldenhoff R. 2003. The tobacco aquaporin NtAQP1 is a membrane CO_2 pore with physiological functions. Nature **425**, 734–737.

Uehlein N, Otto B, Hanson DT, Fischer M, McDowell N, Kaldenhoff R. 2008. Function of *Nicotiana tabacum* aquaporins as chloroplast gas pores challenges the concept of membrane CO_2 permeability. The Plant Cell **20**, 648–657.

Van Hautegem T, Takasaki H, Lorenzo CD, *et al*. 2024. Division zone activity determines the potential of drought-stressed maize leaves to resume growth after rehydration. Plant, Cell & Environment doi: 10.1111/pce.15227.

Wang C, Hu H, Qin X, Zeise B, Xu D, Rappel WJ, Boron WF, Schroeder JI. 2016. Reconstitution of CO_2 regulation of SLAC1 anion channel and function of CO_2-permeable PIP2;1 aquaporin as CARBONIC ANHYDRASE4 interactor. The Plant Cell **28**, 568–582.

Webb AA. 2003. The physiology of circadian rhythms in plants. New Phytologist **160**, 281–303.

Wei W, Alexandersson E, Golldack D, Miller AJ, Kjellbom PO, Fricke W. 2007. HvPIP1;6, a barley (*Hordeum vulgare* L.) plasma membrane water channel particularly expressed in growing compared with non-growing leaf tissues. Plant and Cell Physiology **48**, 1132–1147.

Zelazny E, Borst JW, Muylaert M, Batoko H, Hemminga MA, Chaumont F. 2007. FRET imaging in living maize cells reveals that plasma membrane aquaporins interact to regulate their subcellular localization. Proceedings of the National Academy of Sciences, USA **104**, 12359–12364.

Zhang J, De-oliveira-Ceciliato P, Takahashi Y, *et al*. 2018. Insights into the molecular mechanisms of CO_2-mediated regulation of stomatal movements. Current Biology **28**, R1356–R1363.

Zhou L, Wang C, Liu R, Han Q, Vandeleur RK, Du J, Tyerman S, Shou H. 2014. Constitutive overexpression of soybean plasma membrane intrinsic protein GmPIP1;6 confers salt tolerance. BMC Plant Biology **14**, 181.

Zhu JK. 2002. Salt and drought stress signal transduction in plants. Annual Review of Plant Biology **53**, 247–273.

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erae500/7928862 by guest on 26 January 2025