Harnessing multi-omics data to unravel the genetic and molecular basis of complex traits : A systems genetics study of maize drought response

Yacine Djabali

université PARIS-SACLAY

Harnessing multi-omics data to unravel the genetic and molecular basis of complex traits: A systems genetics study of maize drought response

Exploitation des données multi-omiques pour élucider les bases génétiques et moléculaires de caractères complexes : Une étude de génétique des systèmes de la réponse du maïs à la sécheresse

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 567, Sciences du végétal : du gène à l'écosystème (SEVE)
Spécialité de doctorat : Génétique
Graduate School : Biosphera

Thèse préparée dans l'unité de recherche **GQE** (Université Paris-Saclay, INRAE, CNRS, AgroParisTech), sous la direction de **Mélisande BLEIN-NICOLAS**, Ingénieure de Recherche, et la co-direction de **Marie-Laure MARTIN**, Directrice de Recherche

Thèse soutenue à Paris-Saclay, le 13 décembre 2023, par

Yacine DJABALI

Composition du jury

Membres du jury avec voix délibérative

Andrea RAU	Présidente
Directrice de recherche, Université Paris-Saclay	
Nicolas LANGLADE	Rapporteur & Examinateur
Directeur de recherche, Université de Toulouse	
Vincent SEGURA	Rapporteur & Examinateur
Chargé de recherche (HDR), Université de Montpellier	
Mikaël LUCAS	Examinateur
Chargé de recherche, Université de Montpellier	
Élisabeth PETIT-TEIXEIRA	Examinatrice
Professeure, Université d'Evry Val d'Essonne	

THESE DE DOCTORAT
NNT : 2023UPASB086

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences du végétal : du gène à l'écosystème (SEVE)

Titre: Exploitation des données multi-omiques pour élucider les bases génétiques et moléculaires de caractères complexes : Une étude de génétique des systèmes de la réponse du maïs à la sécheresse

Mots clés: biologie des systèmes, génétique d'association, intégration de données multi-omiques, réseaux moléculaires, relation génotype-phénotype, interaction génotype-environnement

Résumé: Bien que les interactions entre la recherche académique et les entreprises semencières au cours du siècle dernier aient permis une avancée significative dans l'amélioration des plantes cultivées , cela fait plusieurs années que les rendements des grandes cultures n'ont pas augmenté de manière significative. De plus, selon le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), les émissions de gaz à effet de serre ont déclenché une hausse irréversible des températures qui rendra les terres agricoles plus sèches et augmentera considérablement les mauvaises récoltes au cours des 30 prochaines années. Nourrir près de dix milliards de personnes dans ce contexte de changement climatique est donc l'un des plus grands défis de ce siècle. Parmi les cultures touchées par la sécheresse, le maïs est au centre des recherches visant à améliorer les variétés pour les rendre plus résistantes au stress hydrique. Cependant, la tolérance à la sécheresse est un caractère polygénique (*i.e.*, un caractère contrôlé par plusieurs gènes) qui dépend fortement de l'environnement, ce qui fait de la compréhension de son déterminisme génétique une tâche considérable. En effet, après avoir perçu un stress hydrique, les plantes déclenchent de multiples voies moléculaires affectant leur développement et leur rendement. Grâce aux avancées biotechnologiques, il est possible de générer des ensembles de données multi-omiques permettant d'appliquer des approches de génétique des systèmes dans l'étude de la tolérance à la sécheresse. Ma thèse visait à mieux comprendre les bases génétiques et moléculaires de la réponse du maïs à la sécheresse en réalisant une analyse intégrative de données multi-omiques (génomique : un million de SNP, protéomique : 2000 protéines, métabolomique : 1500 métabolites, et phénomique : 6 traits écophysiologiques liés à la sécheresse) mesurées pour 254 hybrides de maïs cultivés sous deux régimes hydriques contrastés. Dans la première partie de la thèse, je me suis concentré sur l'analyse des données phénomiques afin de quantifier la pertinence de l'intégration des indices de plasticité dans les études de génétique d'association pour détecter des QTLs impliqué dans l'interaction génotype-disponibilité en eau (GxW). Le principal résultat de cette partie est que les QTL de plasticité ne se chevauchent pas avec les QTL détectés sur les moyennes phénotypiques, et qu'ils capturent exclusivement une partie importante de la variance GxW (10-70% en fonction des caractères). Outre l'identification de nouvelles régions génétiques potentiellement impliquées dans la réponse à la sécheresse, mes résultats soutiennent le postulat selon lequel la plasticité phénotypique est un caractère indépendant avec son propre déterminisme génétique. Dans la deuxième partie de la thèse, j'ai mené une approche de génétique des systèmes en intégrant les données de génomique, de protéomique et de phénomique pour i) inférer un réseau multi-échelles révélant les bases génétiques et moléculaires de la réponse au stress hydrique, ii) évaluer les apports des données de protéomique pour expliquer la variance GxW, et iii) fournir une annotation fonctionnelle des QTLs maximisant la proportion de variance GxW capturée. Tout d'abord, j'ai pu identifier des régions génomiques enrichies en pQTLs et les traduire en réseaux d'interactions protéine-protéine. Cela m'a permis de montrer que les protéines associées à des pQTLs situés dans ces régions pouvaient interagir physiquement avec des protéines codées par des gènes couverts par ces régions. Deuxièmement, j'ai identifié un ensemble de QTL et de pQTL qui, ensemble, capturent 84% de la variance GxW. Troisièmement, j'ai inféré un réseau multi-échelle comprenant 531 loci, 63 protéines et 6 caractères répondant à la sécheresse. Ces résultats mettent en avant le potentielle des données omiques afin de révéler les bases génétiques et moléculaires de caractères complexes tels que la tolérance à la sécheresse.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences du végétal : du gène à l'écosystème (SEVE)

Title: Harnessing multi-omics data to unravel the genetic and molecular basis of complex traits: A systems genetics study of maize drought response

Keywords: system biology, genome-wide association study, multi-omics data integration, molecular networks, genotype-phenotype relationship, genotype-by-environment interaction

Abstract: Although interactions between academic research and plant breeding companies over the last century have enabled significant progress in crop improvement by producing high-yielding varieties, it has been several years since major crop yields have increased significantly. In addition, according to the Intergovernmental Panel on Climate Change (IPCC), anthropogenic greenhouse gas emissions have triggered an irreversible rise in temperatures that will make agricultural land drier and significantly increase crop failures over the next 30 years. Feeding nearly ten billion people in the face of climate change is, therefore, one of the greatest challenges of this century. Among crops affected by drought, maize is at the center of research into improving varieties to make them more resilient to water stress. However, drought tolerance is a polygenic trait (*i.e.*, a trait under the control of multiple genes) that is highly dependent on the environment, which makes understanding its genetic determinism a tremendous task. After perceiving water stress, plants trigger multiple molecular pathways that moderate root water uptake and leaf evapotranspiration, which can affect their development and reduce yield. Thanks to advances in biotechnology, it is now possible to generate multi-omics datasets that can be used to conduct systems genetic approaches to address the complexity of drought tolerance. My PhD thesis aimed to gain insight into the genetic and molecular basis of maize drought response by performing an integrative analysis of multi-omics data (genomics: one million SNPs, proteomics: 2,000 proteins, metabolomics: 1,500 metabolites, and phenomics: 6 drought-related ecophysiological traits) measured for 254 maize hybrids grown in a greenhouse under two contrasting water regimes. In the first part of the thesis, I focused on the analysis of phenomics data to quantify the relevance of integrating plasticity indices in genome-wide association studies to detect QTLs involved in the genotype-by-water availability interaction (GxW). The main result of this part was that plasticity QTLs do not overlap with QTLs detected on phenotypic means, and they exclusively capture an important part of the GxW variance (10-70% depending on the traits). Besides identifying novel genetic regions potentially involved in drought response, my results support the postulate that phenotypic plasticity is an independent trait with its own genetic determinisms. The latter could be advantageous in breeding in order to design high-yielding varieties that optimize their water management through plastic ecophysiological traits. In the second part of the thesis, I conducted a systems genetics approach by integrating genomics, proteomics, and phenomics to i) infer a multiscale network revealing the genetic and molecular basis of the response to water stress, ii) assess the contribution of proteomics data in explaining GxW variance, and iii) provide a functional annotation of QTLs maximizing the proportion of GxW variance captured. Firstly, I was able to identify genomic regions enriched in pQTLs and translate them into protein-protein interaction networks. This enabled me to show that proteins associated with pQTLs located in these regions could physically interact with proteins encoded by genes covered by these regions. Secondly, I identified a set of QTLs and pQTLs that together capture 84% of the GxW variance. Thirdly, I inferred a multi-scale network comprising 531 loci, 63 proteins, and 6 drought-responsive traits. These results highlight the potential of omics data to reveal the genetic and molecular basis of complex traits such as drought tolerance. Overall, the results of my thesis encourage the consideration of phenotypic plasticity and the use of omics data to facilitate the design of drought-tolerant varieties in molecular-assisted selection.

Remerciements (French)

Vous voici à la page de remerciement d'un manuscrit de thèse de doctorat. Une production qui souvent nécessite de faire preuve d'un grand dévouement en nos idées, afin de les défendre avec la plus haute authenticité. Oui, ces 3 années ou plus, sont synonyme d'un parcours éprouvant, sur lequel il est nécessaire de faire preuve de courage et d'une immense détermination. Bien que l'accomplissement d'une thèse de doctorat est, pour le peu qu'on puisse dire, rayonnante, les années qui ont précédé, ont pu être à quelques reprises, traversées par des épisodes de brumes. Des périodes où l'on est débordé par un état d'esprit défaitiste, qui fatalement, nous emporte dans un ouragan de questions. Je me demande, si cela peut aussi être une raison, pour laquelle la Recherche, est avant tout une discipline communautaire. Car, en effet, c'est bien dans ma communauté que j'ai trouvée refuge. Encadrants, collaborateurs, collègues, amis, ou encore famille, un groupe d'individus moteur de ma motivation, qui m'a permis de sortir de cet ouragan de questions, sous les rayons d'un victorieux développement personnel. Il m'en a fallu du temps pour admettre enfin, que derrière un verre à moité vide, se cache aussi un verre à moitié plein.

Les lignes qui vont suivre, ont donc été dédiées à toutes les personnes, avec qui j'ai eu la chance d'interagir au cours de ces années de thèse, qui aussi peuvent se résumer en une incroyable et époustouflante période de ma vie.

J'aimerais, dans un premier temps, remercier Mélisande Blein-Nicolas, ma directrice de thèse. Je suis ravi d'avoir pu me former au travail de recherche auprès d'elle, car outre son soutien scientifique, j'ai pu ressentir un instinct maternel empli de bienveillance à l'égard de ma personne. Je tiens à exprimer toute ma gratitude envers ma co-directrice de thèse, Marie-Laure Martin. C'est bien durant ces très précieux moments partagés à boire du café ou à coder des scripts R, que nos discussions, ont permis de nouer un lien apprenti - mentor très fort, avec de l'admiration d'un côté et de la réjouissance de l'autre. Je ne pourrai oublier ce soutien inébranlable dans la réalisation de mes idées que Marie-Laure a su fournir au long de ces trois années de thèse. Je remercie également Renaud Rincent pour sa participation scientifique, ainsi que le temps qu'il a consacré à ma recherche. Cette proximité qu'il a tenu garder envers mon sujet de thèse, nous a été très bénéfique.

Je remercie tout les mouloniens qui m'ont offert un accueil aussi chaleureux que bienveillant. Bien que mon séjour dans l'extraordinaire et très prestigieuse ferme du Moulon se soit écourté au profit d'un déménagement vers le gros bâtiment blanc que forme à présent l'IDEEV, l'esprit tant sympathique de cette UMR a rendu ces années de thèse très agréable. Merci Sophie Pin, Nathalie Galic et Harry Belcram veillant toujours à faire répandre leur joie de vivre dans nos bureaux. Merci aux membres des équipes PAPPSO et GQMS pour votre sympathie et de m'avoir fait transparaître ce sentiment d'être toujours considéré comme l'un des votres malgrès ma réaffectation d'équipe en thèse. Un grand merci aux membres de l'équipe BASE, "La meilleure équipe du Moulon" pour reprendre les mots de mon frère d'arme Sacha Revillon. Le simple fait de tous vous revoir en réunion d'équipe, et sous l'animation de Judith Legrand, était suffisant pour égailler ma journée. Merci à Christine Dillmann, qui malgré la charge de travail gargantuesque que lui fait subir son statut de directrice d'unité, n'a pas manqué une occasion pour m'avoir fait ressentir son intérêt pour mes travaux de thèse. Merci à Dominique de Vienne pour sa transmission de savoir sur l'histoire de la génétique et son aide sur le début de l'introduction de mon manuscrit de thèse. Merci à l'ensemble des doctorants et post-doctorants, qui sans eux mes séjours au Moulon aurait eux une tout autre saveurs : Agustin Galaretto, Baber Ali, Michel Turbet-Deloff, Annaïg De Walsche, Thomas-Sylvestre Michau, Cécile Moulin, Maëlla Semery, Arthur Wojcik, Noa Vazeux-Blumental et enfin mention spéciale pour mes deux précieux mousquetaires Romain Benoist et Sacha Revillon.

Comment pourrais-je oublier les séjours passés à l'IPS2 égaillé par les Ginettes (membres de l'équipe GNET) pendant les pauses cafés. La ponctualité de Christine Paysant-Le roux, les chocolats-pâtisseries-compotes faites maison de Jean Phillipe Tamby, les rires de Véronique Bruneaux, le professionalisme (ou psychorigidité selon Véro) de Cécile Guichard, l'humour subtils de Guillem Rigaill, la gentillesse de Nathalie Boudet, et la sympathie de Marie-Hélène Mucchielli-Giorgi. Merci à vous tous pour ces moments partagés en équipe à l'IPS2. Je remercie mes compères, doctorants affiliés à l'équipe GNET pour l'entraide qu'il y a

pu avoir entre nous: Arnaud Lierhmann, Simon Gosset et Margot Correra. Je remercie également Etienne Delannoy et Jean Colcombet pour leur inssatiable curiosité envers mon sujet de thèse qui a souvent permis d'être moteur de mes avancées.

Un grand merci à Sophie Donnet et Hugo Gangloff, pour leur motivation à superviser des étudiants sur des axes de recherches "annexes" à cette thèse, et direct sortie de mon imagination farfelue. Ces expériences m'ont permis de donner goût à l'encadrement.

Merci à Nicolas Langlade et Vincent Segura pour avoir rapporté cette thèse et avoir fait ressorti des points de discussions très intéressants durant la soutenance. Merci également aux examinateurs Andrea Rau, Mikaël Lucas, et Elisabeth Petit-Teixeira pour leurs questions plus que pertinente, et permettant d'ouvrir mon travail vers d'autres horizons.

Pour terminer, je remercie ma famille, mes amis et bien sûr sans t'oublier, Asma, pour m'avoir supporté et soutenue tout au long de cette belle aventure.

Résumé (French)

Bien que les interactions entre la recherche académique et les entreprises semencières au cours du siècle dernier aient permis une avancée significative dans l'amélioration des plantes cultivées en produisant des variétés à haut rendement, cela fait plusieurs années que les rendements des grandes cultures n'ont pas augmenté de manière significative. De plus, selon le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), les émissions anthropiques de gaz à effet de serre ont déclenché une hausse irréversible des températures qui rendra les terres agricoles plus sèches et augmentera considérablement les mauvaises récoltes au cours des 30 prochaines années. Nourrir près de dix milliards de personnes dans ce contexte de changement climatique est donc l'un des plus grands défis de ce siècle. Parmi les cultures touchées par la sécheresse, le maïs est au centre des recherches visant à améliorer les variétés pour les rendre plus résistantes au stress hydrique. Cependant, la tolérance à la sécheresse est un caractère polygénique (*i.e.*, un caractère contrôlé par plusieurs gènes) qui dépend fortement de l'environnement, ce qui fait de la compréhension de son déterminisme génétique une tâche considérable. En effet, après avoir perçu un stress hydrique, les plantes déclenchent de multiples voies moléculaires qui modèrent l'absorption d'eau par les racines et l'évapotranspiration des feuilles, et qui peuvent affecter leur développement et leur rendement. Grâce aux avancées biotechnologiques, il est maintenant possible de générer des ensembles de données multi-omiques permettant d'appliquer des approches de génétique des systèmes dans l'étude des caractères complexes tels que la tolérance à la sécheresse. Ma thèse de doctorat visait à mieux comprendre les bases génétiques et moléculaires de la réponse du maïs à la sécheresse en réalisant une analyse intégrative de données multi-omiques (génomique : un million de SNP, protéomique : 2000 protéines, métabolomique : 1500 métabolites, et phénomique : 6 traits écophysiologiques liés à la sécheresse) mesurées pour 254 hybrides de maïs cultivés dans une serre sous deux régimes hydriques contrastés. Dans la première partie de la thèse, je me suis concentré sur l'analyse des données phénomiques afin de quantifier la pertinence de l'intégration des indices de plasticité dans les études de génétique d'association pour détecter des QTLs impliqué dans l'interaction génotype-disponibilité en eau (GxW). Le principal résultat de cette partie est que les QTL de plasticité ne se chevauchent pas avec les QTL détectés sur les moyennes phénotypiques, et qu'ils capturent exclusivement une partie importante de la variance GxW (10-70% en fonction des caractères). Outre l'identification de nouvelles régions génétiques potentiellement impliquées dans la réponse à la sécheresse, mes résultats soutiennent le postulat selon lequel la plasticité phénotypique est un caractère indépendant avec son propre déterminisme génétique. Ce dernier point pourrait être avantageux pour la sélection afin de concevoir des variétés à haut rendement qui optimisent leur gestion de l'eau grâce à des caractéristiques écophysiologiques plastiques. Dans la deuxième partie de la thèse, j'ai mené une approche de génétique des systèmes en intégrant les données de génomique, de protéomique et de phénomique pour i) inférer un réseau multi-échelles révélant les bases génétiques et moléculaires de la réponse au stress hydrique, ii) évaluer les apports des données de protéomique pour expliquer la variance GxW, et iii) fournir une annotation fonctionnelle des QTLs maximisant la proportion de variance GxW capturée. Tout d'abord, j'ai pu identifier des régions génomiques enrichies en pQTLs et les traduire en réseaux d'interactions protéine-protéine. Cela m'a permis de montrer que les protéines associées à des pQTLs situés dans ces régions pouvaient interagir physiquement avec des protéines codées par des gènes couverts par ces régions. Deuxièmement, j'ai identifié un ensemble de QTL et de pQTL qui, ensemble, capturent 84% de la variance GxW. Troisièmement, j'ai inféré un réseau multi-échelle comprenant 531 loci, 63 protéines et 6 caractères répondant à la sécheresse. Ces résultats mettent en avant le potentielle des données omiques afin de révéler les bases génétiques et moléculaires de caractères complexes tels que la tolérance à la sécheresse. Dans l'ensemble, les résultats de ma thèse encouragent la prise en compte de la plasticité phénotypique et l'utilisation de données omiques pour faciliter la conception de variétés tolérantes à la sécheresse dans le cadre de la sélection assistée par marqueurs.

Contents

Chapter 1

General introduction

1.1 . Historical view of the science of heredity and trait variation

1.1.1 . Establishment of the mathematical formalism to study heredity and variation of traits

Genetics was initiated by Gregor Johann Mendel in 1865 with his experimental work on pea (*Pisum sativum*) hybridization (Mendel, 1865; Gayon, 2016). Although this work represented a major scientific contribution, it was not recognized until thirty years later. Indeed, in 1900, Mendel's experimental work on plant hybridization was rediscovered by three independent researchers Hugo de Vries, Carl Correns, and Erich von Tschermak. In 1902, Willam Bateson, a biologist who was convinced by Mendel's experiments on heredity, wrote a book,*Mendel's Principles of Heredity* (Bateson and Mendel, 1902), in which he summarized Mendel's discoveries in three laws (Box 1 and Fig. 1.1) and defined the terms allele, zygote, homozygote, and heterozygote (Box 2). He was also the one who, in 1905, coined the word "Genetics" (Box 2) to refer to the discipline of studying the heredity and variation of traits.

	Allele	
Color	A : Green	a : Yellow
Shape	B : Round	b : Wrinkled

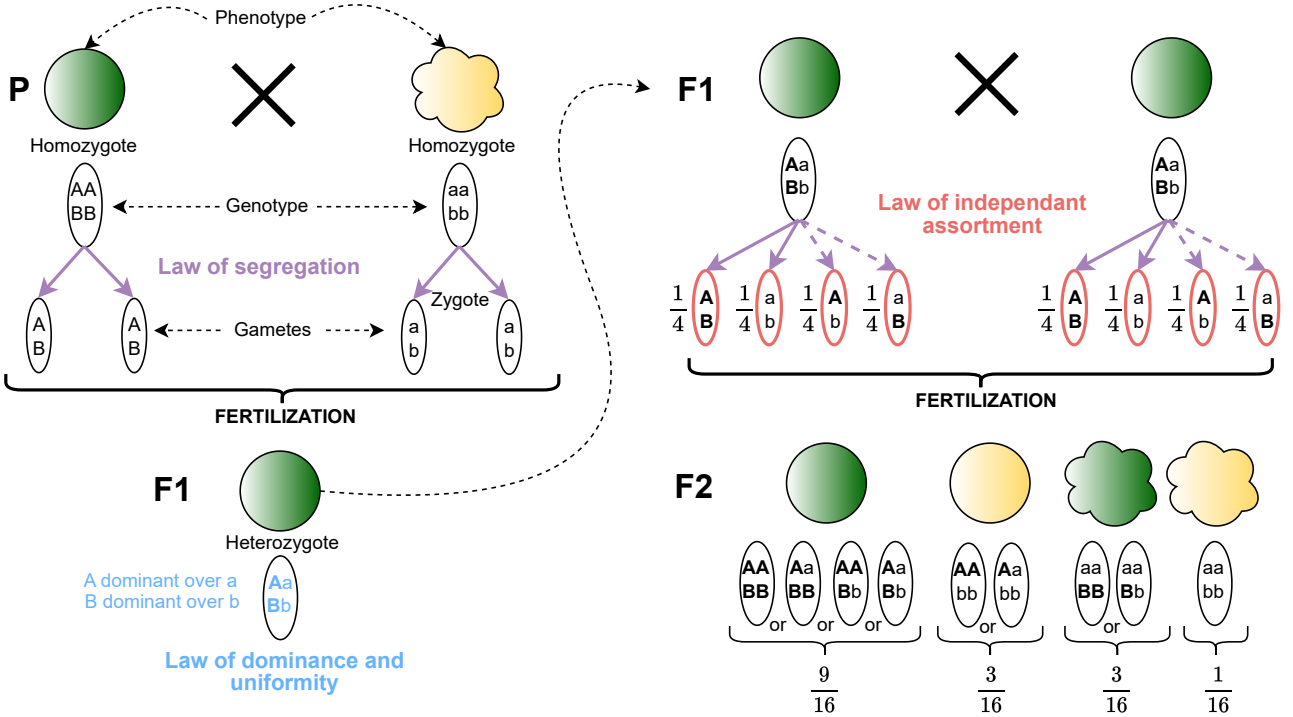


Figure 1.1: **Mendel's *Pisum sativum* heredity experiment.** The pea's color and shape were represented as two traits controlled by two distinct loci. Two versions (alleles) for the locus determining the color of the pea are set as "A" for green and "a" for yellow. Two versions (alleles) for the locus determining the shape of the pea are set as "B" for round and "b" for wrinkled. The F1 individuals were obtained by crossing two homozygotes parents P with extreme phenotypes. The F2 individuals were obtained by crossing two F1 individuals.

Box 1: The Mendelian's Laws (see also illustration Fig. 1.1)

- **The law of dominance and uniformity**: if only one parental trait is passed on to the progeny, then one of the transmitted parental alleles is dominant over the other.
- **The law of segregation**: gametogenesis causes the segregation of the two copies of the parental heredity information to produce two gametes with one copy each.
- **The law of independent segregation**: alleles segregate independently. A heterozygous individual could produce gametes with different combinations of alleles.

In 1909, Willhelm Johannsen, by studying the inheritance of bean size, found that the phenotypic variability could be explained by both genetic and non-genetic factors, such as the position of the bean on the pod (Johannsen, 1911) and proposed to distinguish the genotype from the phenotype (Roll-Hansen 2014; de Vienne 2022, Box 2). However, despite the success of the *Mendel's Principles of Heredity* (Bateson and Mendel, 1902), a group of scientists called the biometricians (Box 2), who were developing statistics to provide a mathematical description of biology, largely criticized Mendel's laws for their lack of statistical validity in describing the inheritance of traits with a continuous range of variation. This confrontation between Mendelians and biometricians was at the origin of several tensions between William Bateson and Karl Pearson, an important figure of the biometric school (Olby, 1989) until the validity of Mendelian genetics with phenotypes showing continuous variation was conciliated by Ronald Aylmer Fisher(Fisher, 1918; Morrison, 2002).

Box 2: Definition of the basic concepts

Allele: a version of a gene.
Zygote: the product of gamete fertilization.
Heterozygote: an individual with different alleles at a given locus.
Homozygote: an individual with identical alleles at a given locus.
Genetics: the science of trait inheritance and variation.
Genotype: the complete set of an individual's genetic material.
Gene: the basic unit of inheritance.
Phenotype: the result of the expression of a genotype in an environment.
Biometry: the statistical analysis of biological observations and phenomena.

Fisher published in 1918, *The Correlation between Relatives on the Supposition of Mendelian Inheritance* (Fisher, 1918), where he postulated that the continuous range of variation of complex traits could be the effect of many genes combined with the environment. He named such traits quantitative traits, and for the first time, he gave a mathematical relationship between the genotype and the phenotype.

Fisher's proposal was to build a statistical model to characterize a population's continuous range of phenotypic values as a random variable P following a normal distribution. Thus, the distribution of P can be defined by two parameters: the phenotypic mean μ (or the expectation of P, $E(P)$) and the phenotypic variance $V(P) = E((P - E(P))^2$. He started to decompose the phenotypic value P as a sum of genotypic and environment random effects:

$$P = G + E \tag{1.1}$$

As the phenotypic value P, the genotypic effect G is a random variable defined by a genotypic mean $E(G)$ and a genotypic variance $V(G)$. Then, he decomposes the genotypic effect G as a sum of independent random effects defined by their transmission mode ($G = A + D + I$):

$$P = A + D + I + E \tag{1.2}$$

with:

- A: The additive effect, *i.e.*, the effect related to the sum of the effect of each allele at each locus.
- D: The dominance effect, *i.e.*, the allele interactions in the same locus.

- I: The epistasis effect, *i.e.*, the allele interactions at different loci.
- E: The environment effect.

We illustrate the additive and dominance effects in the phenotypic mean with the simple case of a single locus (no epistasis, i.e., $I = 0$) under the biallelic model. Let P be the set of phenotypic values of a studied population, p the frequency of the dominant allele A_2, and q the frequency of the recessive allele A_1 with $p + q = 1$. Assuming panmixia and the Hardy-Weinberg equilibrium (Box 3), the frequencies of the genotypes as p^2 for A_2A_2, q^2 for A_1A_1, and $2pq$ for A_2A_1.

Box 3: Genotype frequencies under panmixia and the Hardy-Weinberg equilibrium

Genotype frequencies under panmixia

Assuming random mating between individuals (and gamete random fertilization as well) in a population, the proportions of different genotypes in the biallelic case are:

- Genotype A_1A_1 (G_{11}): to produce a zygote A_1A_1, two gametes bearing the allele A_1 of frequency q have to mate. The probability of random occurrence of two alleles A_1 is $q \times q = q^2$
- Genotype A_2A_2 (G_{22}): to produce a zygote A_2A_2, two gametes bearing the allele A_2 of frequency p have to mate. The probability of random occurrence of two alleles A_2 is $p \times p = p^2$
- Genotype A_2A_1 (G_{21}): to produce a zygote A_2A_1, A gamete carrying the allele A_1 and with frequency q has to mate a gamete carrying the allele A_2 and with frequency p. Since there are two ways to produce a zygote A_2A_1, depending on which parents the gametes carrying the alleles A_1 and A_2 come from, their probability random occurrence is equal to $2 \times p \times q = 2pq$

The Hardy-Weinberg equilibrium

The Hardy-Weinberg frequencies are stable over the generations if the size of the population is large and if there is no evolutionary constraint (*i.e.* no migration, no sexual selection, no natural selection, and no mutation).

The genotypic values G_{11}, G_{22}, and G_{21} can be estimated as the respective phenotypic means $E(P_{A_1A_1})$, $E(P_{A_2A_2})$, and $E(P_{A_2A_1})$. Thus, the phenotypic midpoint can be calculated as

$$m = \frac{G_{11} + G_{22}}{2}$$

The additive effect $|a|$ is equal to the deviation of G_{11} (or G_{22}) to m (Fig. 1.2). Assuming that the allele A_2 is dominant over A_1, $a > 0$ for genotype A_2A_2 and $a < 0$ for genotype A_1A_1. Similarly, the dominance effect $|d|$ equals the deviation of G_{21} to m. When $0 < |d| < |a|$, the dominance is partial, $|d| = |a|$ indicates full dominance, and $|d| > |a|$ indicates overdominance.

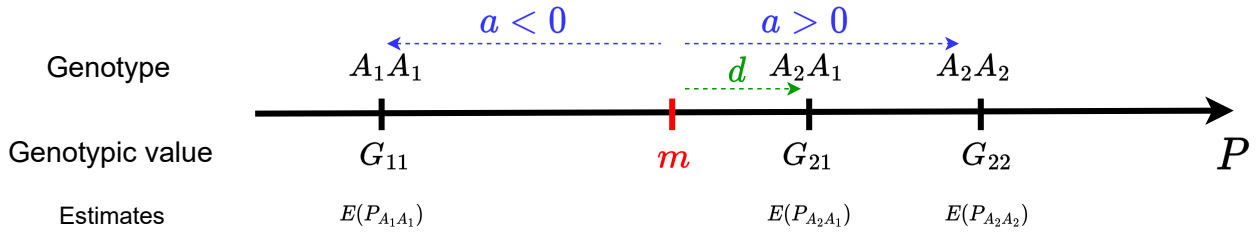


Figure 1.2: **Additive and dominance effects.** As the allele A_2 is dominant over A_1 for the phenotype P, the homozygotes individuals A_2A_2 have, on average, higher phenotypic values than the homozygotes individuals A_1A_1. The midpoint m is represented in red, the additive effect $|a|$ in blue, and the dominance effect $|d|$ in green.

We can express the genotypic effect G as the sum of genotype frequencies weighted by their respective

effects:

$$\begin{aligned} G &= ap^2 + 2pqd - aq^2 \\ &= a(p^2 - q^2) + 2pqd \\ &= a\underbrace{(p+q)}_{1}(p-q) + 2pqd \\ &= a(p-q) + 2pqd \end{aligned} \tag{1.3}$$

The phenotypic mean $\mu = E(P)$ can be expressed as the deviation caused by the addition of the genotypic effect to m:

$$\begin{aligned} \mu &= m + G \\ &= m + a(p-q) + 2pqd \end{aligned}$$

Since parents pass half of their genetic makeup to their offspring and not their entire genome, we can express the effect of each allele rather than genotypes. The average effect of allele A_2 (α_2) is the deviation between the mean of the offspring receiving allele A_2 and the mean of all offspring (Fig. 1.3a). It can be calculated as the sum of the probability that a gamete with allele A_2 will mate with another gamete with allele A_2 (p) multiplied by G_{22} ($m + a$) and the probability that a gamete with allele A_2 will mate with a gamete of allele A_1 (q) multiplied by G_{21} ($m + d$):

$$\begin{aligned} \alpha_2 &= p(m+a) + q(m+d) \\ &= m + pa + qd \end{aligned}$$

Similarly for allele A_1 we obtain $\alpha_1 = m - qa + pd$ (Fig. 1.3a). The difference $\alpha_2 - \alpha_1$ is equal to the effect of allele substitution α *i.e.*, the change in phenotypic mean when an allele A_2 is substituted by an allele A_1 (Fig. 1.3a). Expressing the genotypic values according alleles i and j with their respective additive (α_i and α_j) and dominance (δ_{ij}) effects, we obtain:

$$G_{ij} = \mu + \underbrace{(\alpha_i + \alpha_j)}_{A} + \underbrace{\delta_{ij}}_{D} \tag{1.4}$$

with the dominance effect δ_{ij} corresponding to interaction between alleles i and j at the same locus.

For a given locus and given alleles ($i = 1$, $j = 2$), genotype A_1A_1 can be assumed to be the reference, and Equation (1.4) can be rewritten as:

$$G_{ij} = \mu + 2\alpha_1 + \underbrace{\alpha}_{(\alpha_2 - \alpha_1)} N + \delta_{ij} \tag{1.5}$$

Where α is the effect of substituting the allele α_1 for α_2, and N is the number of alleles A_2 carried by the individuals. The last equation is the most common regression formula for expressing phenotypic values as a function of the allelic dose of the A_2 allele (Fig. 1.3b).

In 1930, Fisher was interested in studying phenotypic and genotypic variance to formulate changes in population fitness attributed to allele frequency (Fisher, 1930). Hence, as expressed in Equation (1.2), the phenotypic variance can be formulated as:

$$V(P) = V(G) + V(E) + 2Cov(G, E) \tag{1.6}$$

The proportion of phenotypic variance explained by the genetic variance is the broad-sense heritability H^2, also known as the linear regression coefficient of the genotype over the phenotype:

$$H^2 = \frac{V(G)}{V(P)}$$

As the genotypic variance can be decomposed by the additive and the dominance variance $V(G) = V(A) + V(D) + 2Cov(A, D)$, we can define the narrow sense heritability h^2 as the part of phenotypic variance explained by the additive fraction of the genetic variance:

$$h^2 = \frac{V(A)}{V(P)}$$

The narrow-sense heritability is the part of the phenotypic variance transmissible from the parents to their progeny. Being able to quantify the contribution of genetics to the population mean and variance allowed further development in animal and plant breeding, especially by giving the possibility to make predictions.

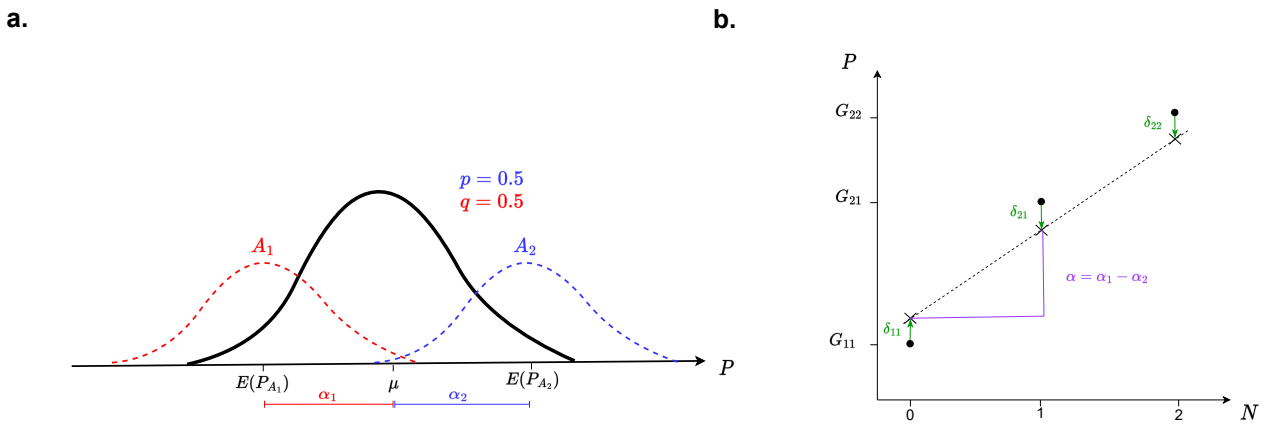


Figure 1.3: **Schematic of the alleles' contributions on the phenotypic mean. a.**, Representation of the density of the phenotypic values according to the proportion of alleles A_1 (red) and A_2 (blue) in the population. The deviation of the phenotypic mean of the individuals carrying the allele A_1 ($E(P_{A_1})$) is the allelic effect α_1 (red). The deviation of the phenotypic mean of the individuals carrying the allele A_2 ($E(P_{A_2})$) is the allelic effect α_2 (blue). **b.**, Regression on genotypic values according to the number of alleles A_2 possessed. The crosses on the regression line are the values we can predict for one individual according to the number of alleles A_1 possessed. The slope of the regression line depicted in purple is the effect of allele substitution α. The black circles are the observed genotypic values. The green deviations δ_{11}, δ_{12}, and δ_{22} are the deviations caused by dominance. With $\delta_{11} = -2q^2d$, $\delta_{12} = 2pqd$, and $\delta_{22} = -2p^2d$.

1.1.2 . The genotype-by-environment interaction

Referring to Equation (1.1), phenotypic values comprise a genetic and environmental component. Unlike the additive component of G, E is not predictable, yet Yates and Cochran (1938) showed in plants that micro and macro-environmental changes can affect the phenotypic value of individuals differently according to their genotypes. This interaction between G and E is important for varietal improvement, especially for plant adaptation (Finlay and Wilkinson, 1963).

$$P = G + E + G \times E$$

This so-called phenotypic plasticity (Bradshaw, 1965) is often represented as reaction norms, with the phenotypic values in the y-axis and the different environments in the x-axis (Fig. 1.4).

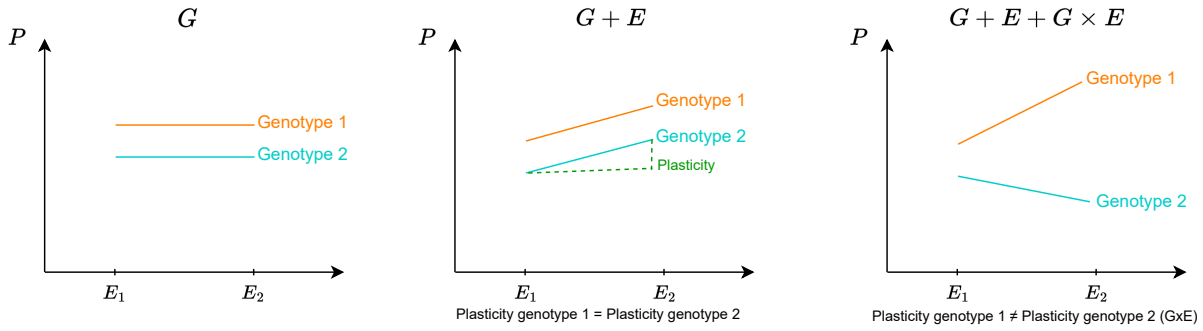


Figure 1.4: **Consideration of the environment on phenotypic variation.** Three different cases of reaction norms displayed by two genotypes in two environments (E_1 and E_2).

Phenotypic plasticity variation between genotypes leads to $G \times E$. Phenotypic plasticity contributes to adaptation and can consequently improve individual fitness in response to environmental changes. Conferring an advantage on individual fitness, phenotypic plasticity can be subjected to evolutionary pressure and then be selected. To better understand the mechanistic phenomenon underlying phenotypic plasticity, three genetics models explaining the expression of plasticity were hypothesized (Scheiner, 1993; Via et al, 1995):

- The overdominance model: phenotypic plasticity is related to the number of heterozygous loci (Gillespie and Turelli, 1989).
- The allele sensitivity model: the environment affects the allelic effect of the genetic factors determining a trait.

- The gene-regulatory model: phenotypic plasticity results from epistatic interactions between structural and regulatory alleles.

The three models can co-exist and are not mutually exclusive. Nonetheless, the gene-regulatory model supported by Bradshaw (1965) statement considering phenotypic plasticity as an independent trait with its own genetic determinism is the most advantageous for plant breeding.

Democratization of genetics in the 20th century, thanks to the rediscovery of Mendel's studies by de Vries, Correns, and von Tschermak, intensified scientific debates on Darwin's evolution theory very anchored at this epoch. The mathematical formalization of the genotype-phenotype relationship (GP) relationship has led to various applications, from medicine to animal and plant breeding. However, these scientific advances in genetics were made before there was any real understanding of the physical objects that carry the genetic material. The following section focuses on the molecular biological discoveries that contributed to the progress of genetics.

1.1.3 . Molecular foundations of genetics supporting the understanding of the genotype-phenotype relationship

At the beginning of the 20th century, elucidating the mitosis and meiosis processes drove scientific attention to chromosomes. In 1902, two cytologists, Walter Sutton and Theodor Boveri, postulated that chromosomes are the component carrying the "Mendelian's heredity unit" (Sutton, 1903; Gayon, 2016). They were authors of a book entitled *The Chromosomes in Heredity*, where they conceptually demonstrated the relevance of linking the process of meiosis with the Mendelian laws of segregation. In 1915, Thomas Hunt Morgan provided much of the mechanistic evidence for combining Mendelian genetics with the chromosomal theory of heredity (Morgan, 1920).

Using this chromosome property of making crossing-overs (*i.e.*, the process by which genetic material is exchanged between two homologous chromosomes during meiosis) (Fig. 1.5), Alfred Henry Sturtevant, a Thomas Morgan apprentice, had the idea to relatively locate on one chromosome genes associated with different Drosophila (*Drosophila Melanogaster*) traits (color of the body, color of eyes, and wings size) by retracing the recombinations events over generations (Fig. 1.6a). Thus, the first genetic map in history could be edited, allowing to spatially locate functional genetic regions on a chromosome from their proportion of recombination, called the centiMorgan (cM) (Fig. 1.6b).

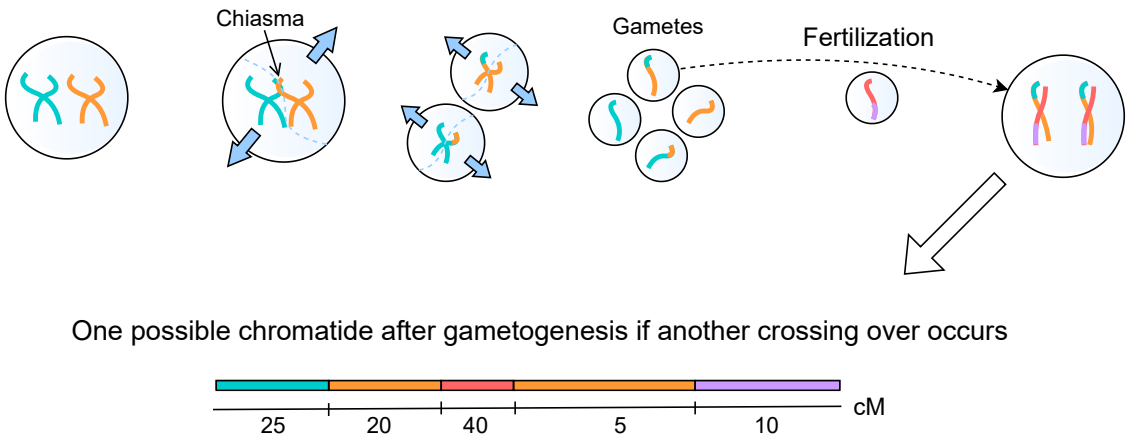


Figure 1.5: **Process of crossing over during meiosis**, Simple case of gametogenesis during meiosis on a diploid cell with one chromosome. Two different colors (blue and orange) for the two homologous chromosomes indicate two different genetic materials provided by parents. Fertilization occurred with another gamete produced by another individual having two genetic materials (red and purple).

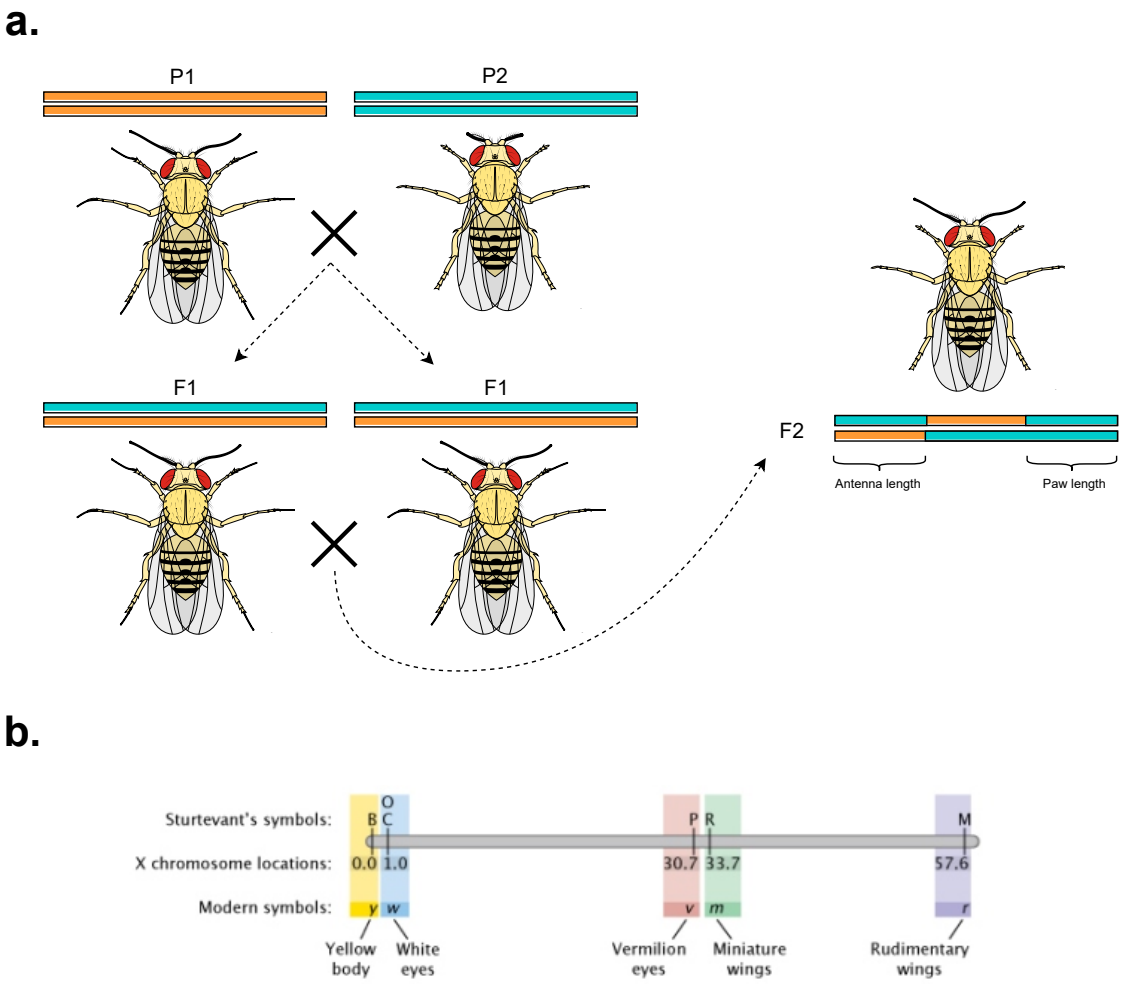


Figure 1.6: **Genetic map. a**, Gene mapping approach based on recombination events. **b**, Sturtevant's gene map, six traits are arranged along a linear chromosome according to the relative distance of each trait. Adapted from Pierce (2005) provided by Nature Education 2013.

Before the 1940s, the physical nature of genes remained unknown, and the scientific community believed that proteins carried heredity. This belief changed in 1940 when Oswald Avery, Colin MacLeod, and Maclyn McCarty demonstrated through genetic transformations (Griffith, 1928) that the deoxyribonucleic acid (DNA) was the molecule carrying heredity (Avery et al, 1944). In 1953, after the discovery of the renowned double helix structure of DNA by Watson and Crick (Watson and Crick, 1953), a sequence of molecular biology discoveries was initiated (Box 4).

Box 4: Molecular biology discoveries between Watson and Crick (1953) and Sanger et al (1977)

- 1955 - 1958: creation of the central dogma of molecular biology by Crick (1958). DNA can be replicated by the DNA polymerase to give the same molecule of DNA. DNA can also be transcribed into a ribonucleic acid (RNA) thanks to the RNA polymerase. Finally, RNA is translated by ribosomes into protein.
- 1955 - 1960: elucidation of the genetic code in 1955 by Har Gobind Khorana. A sequence of three nucleotides is called a codon, and each codon corresponds to one amino acid. A short sequence of amino acids is called a peptide, and a polypeptide having a functional role is called a protein.
- 1977: the discovery of RNA splicing and alternative splicing by (Berget et al, 1977). The DNA is transcribed into messenger RNA (mRNA), which comprises coding and non-coding sequences. Among the non-coding sequences, there are introns, and the coding regions are exons. During RNA splicing, the spliceosome, a protein complex, removes introns and clipped exons to give an RNA ready to be translated into protein. The process of alternative splicing allows the composition of exons to vary, resulting in multiple proteins with one mRNA.
- 1977: development of DNA sequencing by (Sanger et al, 1977). The method is based on the determination of the chain-terminating nucleotide incorporated by a DNA polymerase during the replication of a DNA fragment.

In the 1980s, began the development of molecular markers, highly supported by one of the major progress in biotechnologies, the polymerase chain reaction (PCR) of Kary Mullis (Mullis et al, 1986). The PCR allowed the amplification of a section of DNA in large quantities. That technology has enabled cheaper and automated protocols for designing molecular markers. Following the definition of Amiteye (2021), a molecular marker can be defined as the difference in DNA nucleotide sequence between individual organisms or species (called a genetic polymorphism) in the proximity of a target gene. Usually, a molecular marker does not refer only to the polymorphisms identified but to the complete procedure allowing its detection.

Few non-PCR-based markers exist, such as the random fragment length polymorphism (RFLP), identifying polymorphisms located on enzyme restriction sites. Probed fragments of DNA belonging to different individuals are compared after enzyme digestion. If the DNA fragments revealed by the probe on electrophoresis gel have different weights, it means that one mutation is present in an enzyme restriction site in one individual. However, this method is time-consuming and can't be automated compared to molecular markers based on PCR, which are also cheaper. The most common molecular markers based on PCR are:

- Randomly amplified polymorphic DNA (RAPD): identification of polymorphism at random genome locations by amplifying DNA fragments with short arbitrary oligonucleotide primers. Thus, polymorphism on primers' hybridization sites is revealed if, for one sample, the amplicons resulting from the hybridization of flanking primers are not observed on electrophoresis gel.
- Amplified fragment length polymorphism (AFLP): techniques integrating RFLP and PCR to identify polymorphisms located on enzyme restriction sites. DNA fragments are first obtained by enzyme digestion, which are then amplified with primers recognizing enzyme restriction sites in DNA fragments. If, for one sample, the amplicons resulting from the hybridization of flanking primers are not observed on the electrophoresis gel, it means that a polymorphism is located in one enzyme restriction site.
- Simple sequence repeat (SSR: SSRs are tandemly repeated motifs of one to six nucleotides (also called microsatellites). The nature of polymorphisms based on SSRs depends on the number of repeated motifs. It was confirmed that this kind of repeated sequence is present in protein-coding genes. To reveal polymorphisms in SSR, primers are designed to hybridize with microsatellites' left and right borders. Then, if the weight of the SSR fragment amplified is different between individuals on electrophoresis gel, it means that there is a polymorphism in the number of repeated motifs on the genome location studied.

A decade after the development of PCR, a new method of DNA hybridization arose, the DNA microarrays. This technique is founded on an arrangement of thousands of microscopic DNA spots on a solid

substrate. Each DNA spot comprises a specific DNA sequence, usually a small gene segment. The principal benefit of DNA microarray is its potential for high throughput genotyping, enabling the analysis of thousands of polymorphisms simultaneously. The technology uses single-nucleotide polymorphism (SNP) as a molecular marker, a type of marker with a higher level of polymorphism. The last-mentioned technology has widespread usage and generates large-scale genotyping data for various organisms.

Identifying genetic polymorphism between individuals allowed quantitative genetics studies to pinpoint genetic regions that could be causally linked with the expression of traits. Nevertheless, despite the employment of the cutting-edge technologies mentioned earlier, their utility for understanding the genetic determinism of quantitative traits is not sufficient on its own. In contrast to the monogenic determinism of Mendelian traits, quantitative traits exhibit polygenic determinism, whereby numerous genetic regions are involved in trait elaboration. To address this, statistical models have been designed to evaluate whether a relationship exists between the variations of quantitative traits and the displayed polymorphisms of molecular markers. This method, known as association mapping, facilitates the detection of loci linked with the variation of quantitative traits. These corresponding loci are referred to as quantitative trait loci (QTLs), and their identification is elaborated on in the subsequent section.

1.2 . Handling the complexity of the genotype-phenotype relationship

1.2.1 . Linking genotype to phenotype through association mapping

The use of molecular markers allows to pinpoint the genetic regions where polymorphism exists between a group of individuals. Consequently, the molecular markers give the genotypes of the studied individuals at each marked position. If the same group of individuals has also been phenotyped, a linear model based on the allelic dose regression presented in Equation (1.5) can be applied for each marked position.

$$Y_{ij} = \mu + M_j + \varepsilon_{ij} \tag{1.7}$$

Y_{ij} is the phenotypic value of the individual i with a dose j of the dominant allele ($j \in \{0; 1, 2\}$); μ is the phenotypic mean; M_j is the effect of the dose j of dominant allele possessed by the individual i at the marked genetic position; ε_{ij} is the residual error. If M_j is significantly different from zero, we can suppose that a statistical relationship exists between the polymorphism at the molecular marker and the variations observed on phenotypic values. However, we can not affirm that the polymorphism identified is causal. Indeed, it is important to consider the genetic linkage. Genetic linkage refers to the observation of a higher probability that two adjacent DNA sequences located on the same chromosome will be conserved during recombination events. Thus, applying this notion to population genetics, a non-random association between alleles located in different loci in a given population can exist. These loci are defined as being in linkage disequilibrium (LD). Hence, because a molecular marker can be in linkage disequilibrium with the causal polymorphism, we can not affirm that the marked genetic polymorphism is one of the causes of the observed phenotypic variation.

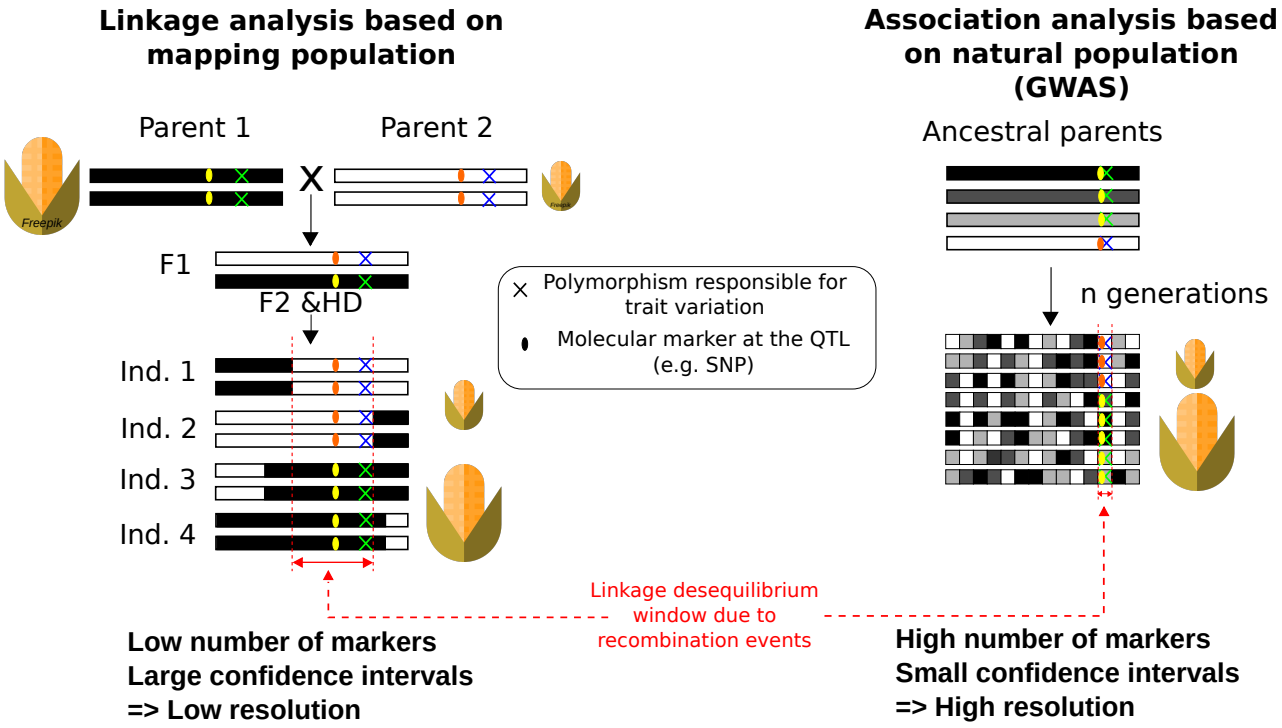


Figure 1.7: **Linkage disequilibrium usage for QTL detection (Adapted from Cardon and Bell (2001) made by S.D. Nicolas and M. Blein-Nicolas).** On the left is represented the linkage analysis based on a mapping population. Generation of a F2 population obtained from two contrasted inbred lines. In this population, the linkage disequilibrium windows are wide because of the limited events of recombination that occurred. Thus, it requires a low number of markers to capture the effect of a causal polymorphism, but the resolution on the position of this polymorphism is weak. On the right is represented the principle of genome-wide association study (GWAS) based on a natural population. The population studied comes from the random mating of an ancestral population. As a consequence, a high number of recombinations occurred, resulting in small linkage disequilibrium windows. The strength of this approach resides in its high resolution for identifying the position of a causal polymorphism using a dense set of molecular markers.

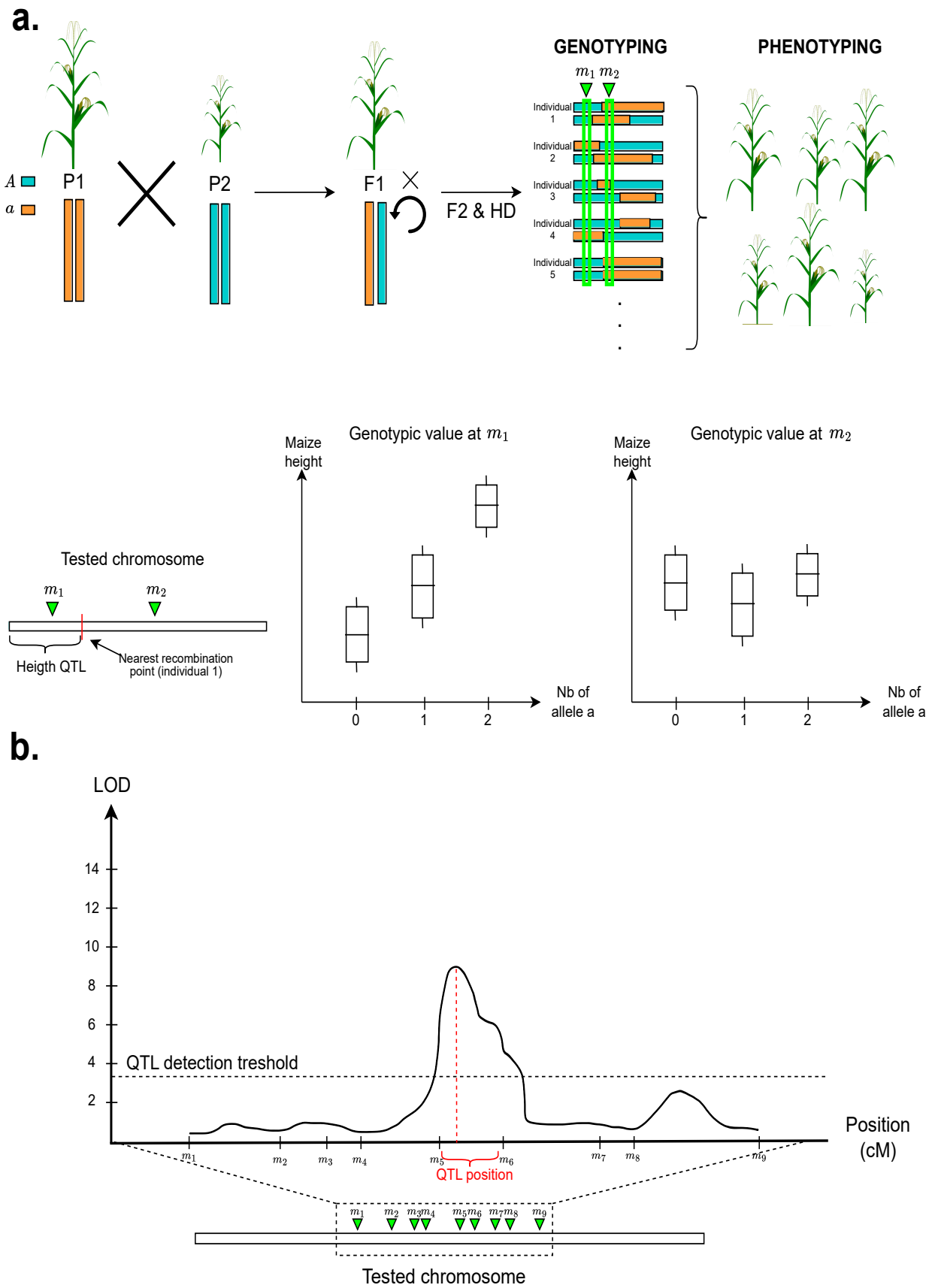


Figure 1.8: **Schematic of association study analysis based on mapping population. a**, QTLs detection from a maize mapping population generated by two parent inbred lines contrasted on height. The mapping population is genotyped with two molecular markers m_1 and m_2, and phenotyped on plant height. The markers regressions shows that m_1 has a significant effect and a height QTL could be located in the tested chromosome. **b**, LOD curves representation showing QTLs positioning from an interval mapping study.

The approach aiming to find associations between genetic polymorphisms and trait variation is called association mapping. This method is based on the LD to estimate genetic regions from the significant markers called quantitative trait loci (QTLs). Two approaches exist to conduct association mapping: the linkage analysis and the genome-wide association study (GWAS). Thus, the approach used to detect a QTL will determine the precision and accuracy of its location. The choice of the method is mainly driven by the population studied and the number of molecular markers used.

Box 5: Consideration of the genetic stratification (population structuration and cryptic relatedness) in GWAS (Astle and Balding, 2009)

Genetic stratification is important to consider in GWAS to prevent an excess of spurious associations due to:

- Population structuration: Even though the genotyping data were generated on a natural population, it is possible that high fragments of DNA (called haplotype blocks) are shared between individuals, and distant markers (that can even be located on different chromosomes) can be in LD (Fig. 1.9a). This phenomenon is called genetic structuration, and it is mostly explained by the fact that natural populations are not in panmixia and can share a recent common ancestry. Thus, individuals in the same geographical zone are more susceptible to sharing the same alleles because their ancestors were submitted to the same selection pressures.
- Cryptic relatedness: Contrary to population structuration, which describes patterns of relatedness in a large group of individuals, cryptic relatedness refers to the genetic distance among a smaller group of individuals regarding a common ancestry.

The first approach of association mapping, called linkage analysis, was based on the use of a limited population of individuals whose parental genotypes were known (mapping population) to be able to follow the recombination events (Fig. 1.7). Each individual of the population is phenotyped for the trait of interest and genotyped using a modest collection of molecular markers ($\sim$ 100-500 random fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), or simple sequence repeat (SSR) markers for the first rigorous applications). Using Equation (1.7), single marker analyses were made to identify polymorphisms significantly associated with the phenotypic variations of the trait (Fig. 1.8a). The QTLs identified with this approach could be very large (Fig. 1.7). Other approaches consisting of considering several markers to gain precision in the location of QTLs were developed, such as the interval mapping (Lander and Botstein, 1989) and the composite interval mapping (Zeng, 1994). One major advantage of interval mapping approaches is that the presence of QTLs is tested with a likelihood ratio test (Morton, 1996) across a genetic region (a portion of chromosome marked with molecular markers). Thus, the continuous profile of the likelihood ratio test allows drawing results as LOD curves, which gives a comprehensive view of QTLs position in the genetic region studied (Fig. 1.8b).

The development of DNA arrays considerably increased the number of available molecular markers, which opened the way to a second approach of association mapping called genome-wide association study (GWAS), enables the carrying of association mapping studies on natural populations (Fig. 1.7) with a dense set of single nucleotide polymorphism (SNP) markers ($> 1M.$). These kinds of studies are called Genome-wide association studies (GWAS) and became the reference of quantitative genetics for identifying QTLs (Visscher and Goddard, 2019). As the studied population is natural, we assumed that the proportion of recombination events between individuals is sufficiently large to drastically reduce the genetic linkage across the genome (Fig. 1.7). Thus, using a denser set of markers increases genome coverage and, therefore, resolution. The most common statistical method used in GWAS is the single marker analysis mentioned in Equation (1.7), but by considering the relatedness and the genetic structuration of the population (Box 5 and Fig. 1.9a-b). One example of a robust GWAS model used in plant genetics is the linear mixed model of Yu et al (2006), where the genetic structuration S is considered as a fixed effect and the relatedness R as a random effect:

$$Y_{ijk} = \mu + M_j + S_k + \underline{R_i} + \underline{\varepsilon_{ijk}} \tag{1.8}$$

Y_{ijk} is the phenotypic value of the individual i in the group k with a dose j of the dominant allele ($j \in \{0; 1, 2\}$); μ is the phenotypic mean; M_j is the effect of the dose j of dominant allele possessed by the individual i at the marked genetic position; S_k is the effect of the group of individuals k; $R_i \sim \mathcal{N}(0, \sigma_g^2.K)$

the random genetic effect of the individual i, considering its relatedness with the other individuals of the population through the kinship matrix K; $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$ is the residual error. Practically, S is often neglected because R allows the capture of a significant part of S.

In practice, GWAS tests a large number of markers individually, so it is necessary to control the false positive rate using a correction for multiple testing. A very popular representation of GWAS results is the Manhattan plot where the $-log(p - value)$ of SNPs tested are represented according to their position on the genome (Fig. 1.9c).

From candidate genes discovery to clarification of biological functions, association mapping facilitated discoveries in the understanding of quantitative traits (Visscher et al, 2017). Further application in the medical field allows the identification of novel genes underlying complex diseases such as diabetes (Duggirala et al, 1999). In breeding, association mapping opened new avenues for animal and plant breeding, especially with the development of marker-assisted selection (Tanksley and Nelson, 1996).

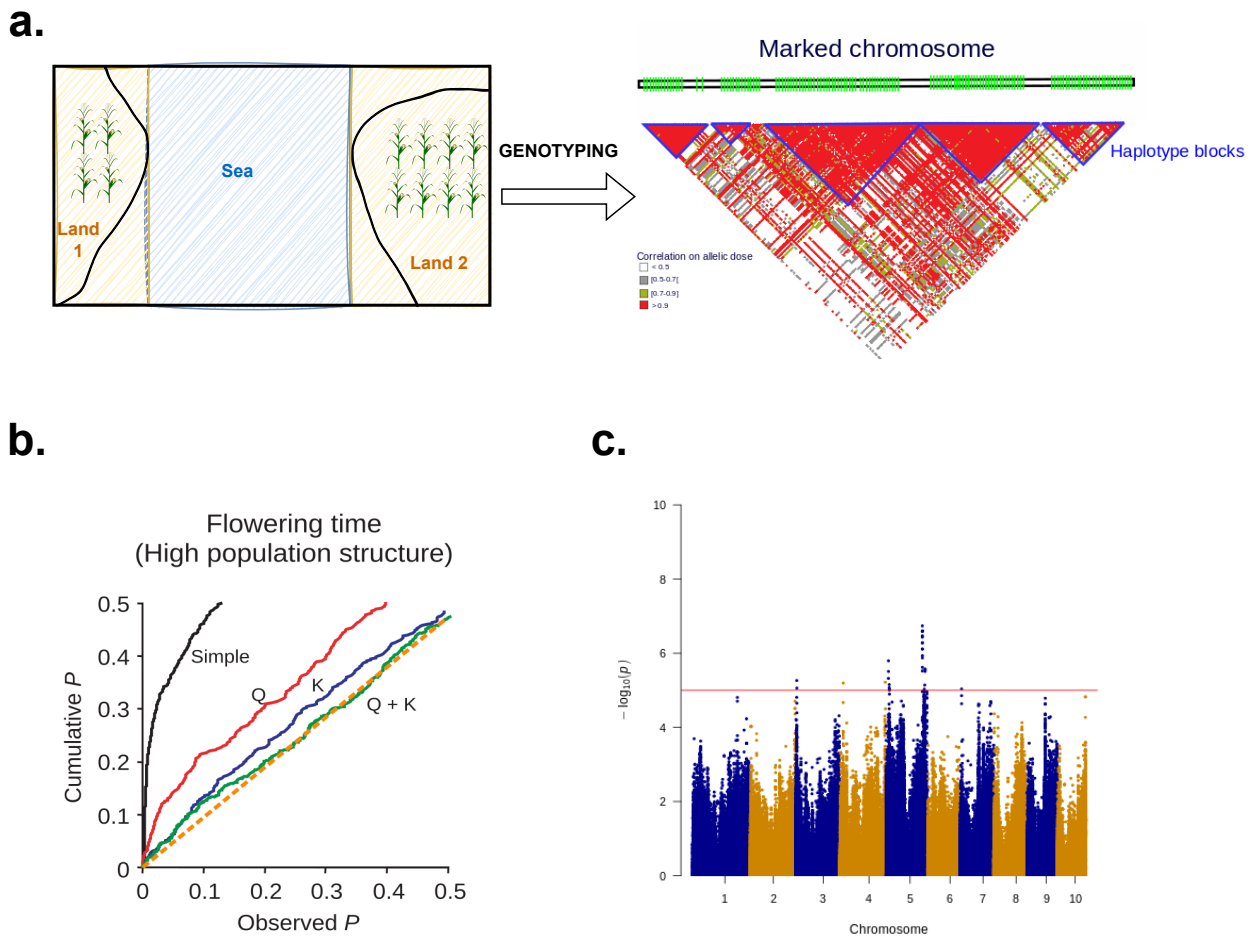


Figure 1.9: **Schematic of Genome-wide association study analysis based on a natural population. a.**, Genotyped natural population presenting genetic structuration. The squared Pearson's correlation r^2 on the individual allelic dose is usually used as an indicator of LD between molecular markers (Adapted from Walton et al (2005)). Usually, a $r^2 > 0.9$ means that the alleles displayed between the two tested markers are in LD. Thus, we can observe clusters of adjacent SNPs with high r^2 characterizing the haplotype blocks, a consequence of genetic structuration. **b.**, Effect of GWAS model specification on p-value distribution (Adapted from Yu et al (2006)). Under the null hypothesis of the GWAS model, we expect that the p-values of all tests follow a uniform distribution between 0 and 1. This expectation on p-values distribution is represented by the dashed orange line where the observed P-values are plotted according to their cumulative sum. The other lines correspond to results obtained with different GWAS models conducted in maize flowering time. Simple (black) corresponds to a GWAS model with no correction, Q (red) is a GWAS model corrected on population structuration, K (blue) is a GWAS model corrected on cryptic relatedness, and Q+K (green) is a GWAS model corrected on population structuration and cryptic relatedness. **c.**, Example of Manhattan plot obtained by conducting GWAS with 1M SNPs and leaf area measurements on 254 maize genotypes. The significance threshold is set at $-log(p - value) > 5$ (dashed red line). The SNPs significantly associated with trait variations correspond to dots above the significance threshold.

1.2.2 . A new model of inheritance for complex traits: The omnigenic model

Despite the success of GWAS in identifying several hundreds of QTLs and providing valuable insights into complex traits' genetic architecture, the QTLs identified generally have small effects and explain a small fraction of trait heritability. For instance, in Visscher (2008), it was demonstrated for population height, which is one of the most heritable traits with a heritability estimated at 80%, that 40 QTLs identified on 10,000 individuals explained only 5% of phenotypic variance. These observations on the small part of genetic variance explained by QTLs identified on GWAS lead to further thinking about unraveling this missing part of heritability (Manolio et al, 2009). Following Manolio et al (2009), this missing part of phenotypic variance could be hidden in:

- Many undiscovered variants with small effects.
- Rare variants that are not well adapted to being detected in a GWAS.
- Structural variants, which are genetic polymorphisms including several base pairs (e.g., Insertion/Deletions, Copy number variants, or transposable elements) not widely used yet as molecular markers.
- Gene-gene interactions, which are computationally complicated to integrate in GWAS models.
- Gene-by-environment interactions

Many studies show results favoring the two first points as Yang et al (2010) and Shi et al (2016). Indeed, their studies demonstrated that by using a whole set of SNPs, which could be composed of several hundred thousand markers, they could unravel almost half part of heritability. Thus, it might be that most part of the missing heritability is hidden in several common polymorphisms having small effects that do not reach the GWAS significance threshold. These observations give further insight into the genetic basis of complex traits, in particular into the establishment of the omnigenic model proposed in 2017 by Boyle et al (2017).

The omnigenic model proposes that almost all genes present in a genome can have an influence on complex trait variations. Synthesizing results obtained in the study of complex traits genetic architecture Boyle et al (2017) concluded that complex traits are mainly controlled by central genes that play a relevant role in the biological pathways underlying the traits, but the regulatory networks of these genes can involve all genes expressed in the cell. Thus, to a lesser extent, all genes expressed in a cell can influence complex trait variation through the regulation of central genes. In such cell regulatory networks, the genes having a direct effect on the trait are referred to as "core", whereas the others which have an indirect effect on the trait by altering the function or the regulation of core genes as "peripheral" (Fig. 1.10). The high number of peripheral genes compared to core genes implies that the peripheral genes explain most of the genetic variance contribution of complex traits (Fig. 1.10).

The identification of peripheral genes altering the regulation and the functions of core genes could allow to better understand the genetic and molecular basis of complex traits. However, those peripheral genes are difficult to identify by conducting GWAS due to their weak effect on complex traits. A solution to tackle this lack of power detection is by studying intermediate molecular traits. Indeed, progress in biotechnologies opens new avenues in the generation of so-called "omics" data composed with measurements obtained from the metabolome, the proteome, or the transcriptome. Harnessing data measured from various biological complexity levels allows one to follow the flow of information from the genotype to the phenotype and identify the molecular networks underlying complex traits.

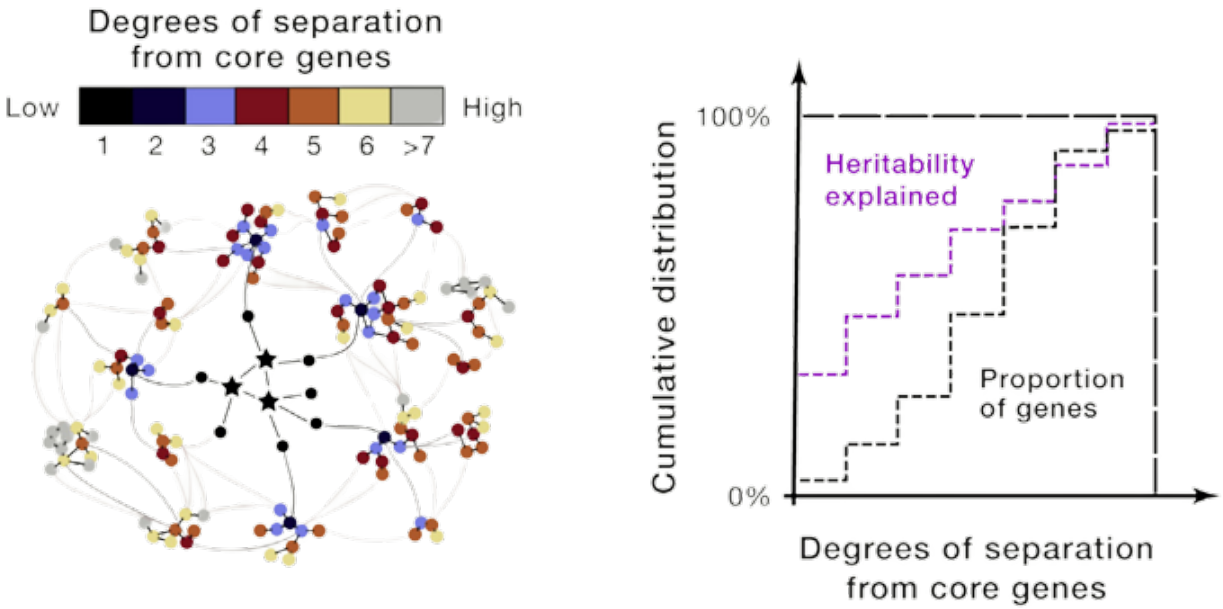


Figure 1.10: **The omnigenic model adapted from Boyle et al (2017).** An example of a cell regulatory network is represented on the left. The stars located in the center of the network represent the core genes, which are outnumbered by the other nodes representing the peripheral genes. The degree of separation of peripheral genes from core genes is illustrated with different colors. On the right represented the degrees of separation from core genes according to the cumulative distribution of heritability explained and the proportion of genes. As we can see, the further we move away from the core genes in the network, the more we integrate genes, and the more it explains heritability.

1.2.3 . Bridging the gap between the genotype and the phenotype through omics data integration

Nowadays, the study of the different biological complexity levels (*e.g.*, the epigenome, the transcriptome, the proteome, and the metabolome) has taken a major step forward with the development of high-throughput technologies. That enables large-scale studies on different infra-cellular levels through the generation of omics data. Among the omics technologies most commonly used, we may cite:

- Epigenomics (Wang and Chang, 2018): the study of the chromatin structure and its effects on gene expression. Several high-throughput technologies exist to analyze the epigenome, such as High Chromosome Contact map (Hi-C) (*i.e.*, captures the site of chromatin conformation through the measure of the frequency of physical interactions between two DNA fragments), ChIP-seq (*i.e.*, captures the site of interactions between a DNA fragment and proteins), and ATAC-seq (*i.e.*, identification of chromatin accessibility sites).

- Transcriptomics (Lowe et al, 2017): the study of the whole set of expressed genes through the quantification of RNA transcripts. The contemporary techniques to quantify gene expression are microarrays (*i.e.*, measurements of the abundance of a defined set of transcripts using an array containing complementary probes. The hybridization between a transcript and its probes gives a fluorescence the intensity of which indicates the abundance of transcripts), and RNA-Seq (*i.e.*, quantification of the whole set of transcripts in an RNA extract by counting the number of times that a same transcript was sequenced).

- Proteomics (Aslam et al, 2017): the study of the protein content of a cell, a tissue, an organ, or an organism. This includes the identification and quantification of the expressed proteins and their post-translational modifications. Currently, proteomics relies on the use of liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS).

- Metabolomics (Ashrafian et al, 2020): the study of the metabolite content of a cell, tissue, organ or organism. Metabolomics is based on mass spectrometry and nuclear magnetic resonance

The various information brought by the different types of omics data displayed by the different types of omics data can be viewed as an advantage by providing different types of analysis such as the identification of differentially expressed genes, proteins, or metabolites (Hu et al, 2017; Porcu et al, 2021) to molecular networks inference (Langfelder and Horvath, 2008; Hawe et al, 2019). However, from a data integration perspective, linking the different molecules and traits to each other can become very complex. Indeed, depending on the technologies, omics data can differ in the type (binary, count, or continuous) or in the scale. In addition, each omics data has its proper steps of data preprocessing composed of control quality filters and normalization (López de Maturana et al, 2019). The approach aiming at integrating different biological entities from different complexity levels is called system biology and is based on vertical integration of omics data (Fig. 1.11a).

The integration of multi-omics data requires sophisticated statistical methods to handle high-dimensioned and heterogeneous data (Picard et al, 2021). It is important to know which methods are the most appropriate depending on the objectives (predict or explain the phenotype) and the type of data. Two major frameworks of multi-omics data integration were reported in Ritchie et al (2015): the multi-staged integration and the multi-dimensional integration. The multi-staged integration combines results obtained from each omics in an asynchronous way. For instance, the relationship between a phenotypic and a molecular trait can be found by conducting two independent GWAS. Indeed, an SNP significantly associated with both phenotypic and molecular trait variations can be used as an anchor to assume a potential relationship. An advantage of this framework is that it allows to easily interpret the results. It is also highly modulable in the way that at each two-by-two analysis, any methods can be used to integrate, such as linear relationship, non-linear, predictive ability, and so on. Also, it can be the advantage that for data with high dimensionality, each step consists of reducing the dimension by keeping only the important features. The multi-dimensional integration consists of integrating each omics data in a synchronous way. Three approaches allow to conduct a multidimensional integration: concatenation-based, transformation-based, and model-based (Fig. 1.11b).

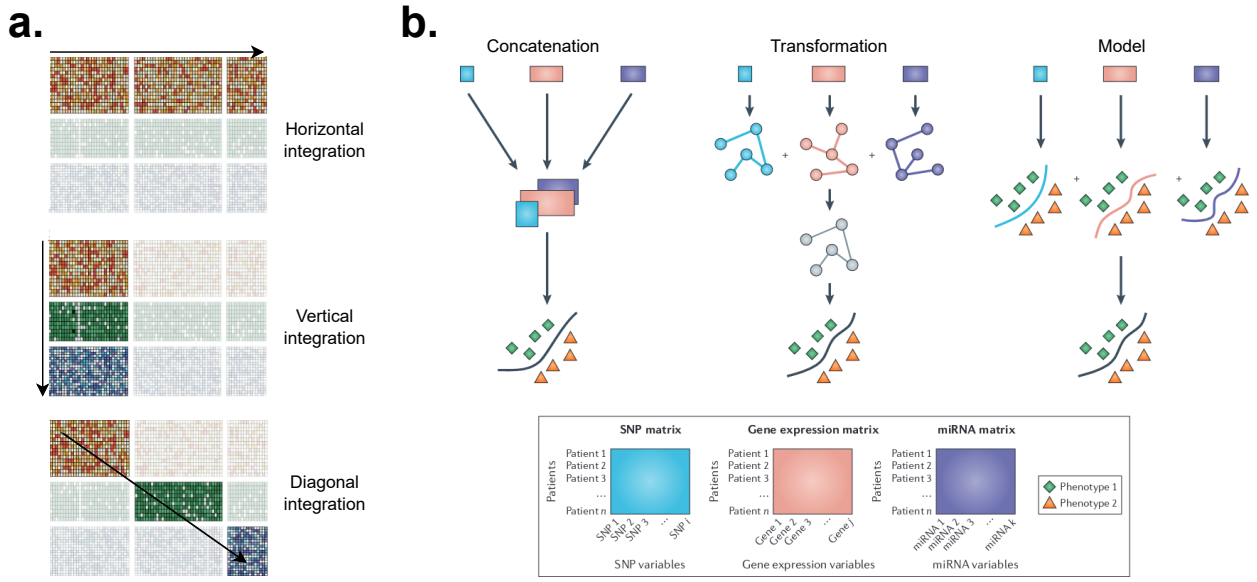


Figure 1.11: **Multi-omics data integration. a.**, The different types of omics data integration adapted from Argelaguet et al (2021). The horizontal integration consists of increased observations of the same measured features. This type of integration does not lead to the integration of different omics data types. The vertical integration aims to increase the number of measured features on the same observation. The features added can come from different omics data types. The diagonal integration integrates datasets comprising measurements obtained on new observations and not for the same biological features. **b.**, Multi-dimensional integration adapted from Ritchie et al (2015). Three approaches allow integrated multi-omics data in a synchronous way. The concatenation-based: all omics data are merged into a matrix to apply statistical methods such as multifactorial analysis (Argelaguet et al, 2018; Singh et al, 2019). The transformation-based: Approach consisting of transforming omics data into an intermediate format allowing to describe the relation between individuals such as individual networks (Kim et al, 2012) or kernels (Lanckriet et al, 2004). The model-based: Approach mostly used for predicting the phenotypes. For each omics data, the best predictive models is selected, and then these models are merged to generate the final one (Holzinger et al, 2014).

System genetics (Civelek and Lusis, 2014; van der Sijde et al, 2014), a sub-branch of system biology, is a powerful approach to deciphering the genetic and molecular basis of complex traits. This approach is based on the combination of multi-omics data integration and association analysis to identify the molecular networks underlying complex traits with their genetic determinants. Indeed, conducting GWAS on the intermediate molecular traits (Wu et al 2018, Fig. 1.12a-b), leads to the detection of molecular QTLs (*e.g.*, methylation-QTL, expression-QTL, protein-QTLs and metabolite-QTLs) giving more statistical power to detect relevant genetic variants. Indeed, this approach can tackle the missing heritability phenomenon by revealing genetic variations observed only by studying the molecular levels because the effect of the genetic variant on the trait does not reach the GWAS significance threshold (Fig. 1.12a). This phenomenon called phenotypic buffering (or robustness) (Félix and Barkoulas, 2015) is explained by the non-linear processes existing across the various biological complexity levels (Kimura, 1983; de Vienne, 2022) making a trait robust in the face of genetic or environmental perturbations. The incorporation of such molecular QTLs into molecular networks underlying complex traits enables an in-depth genetic characterization in line with the omnigenic model of Boyle et al (2017). Thus, system genetics could fill the gap between the genotype and the phenotype in a comprehensive way through molecular networks inference and identification of genetic variants explaining part of the missing heritability (Fig. 1.12b).

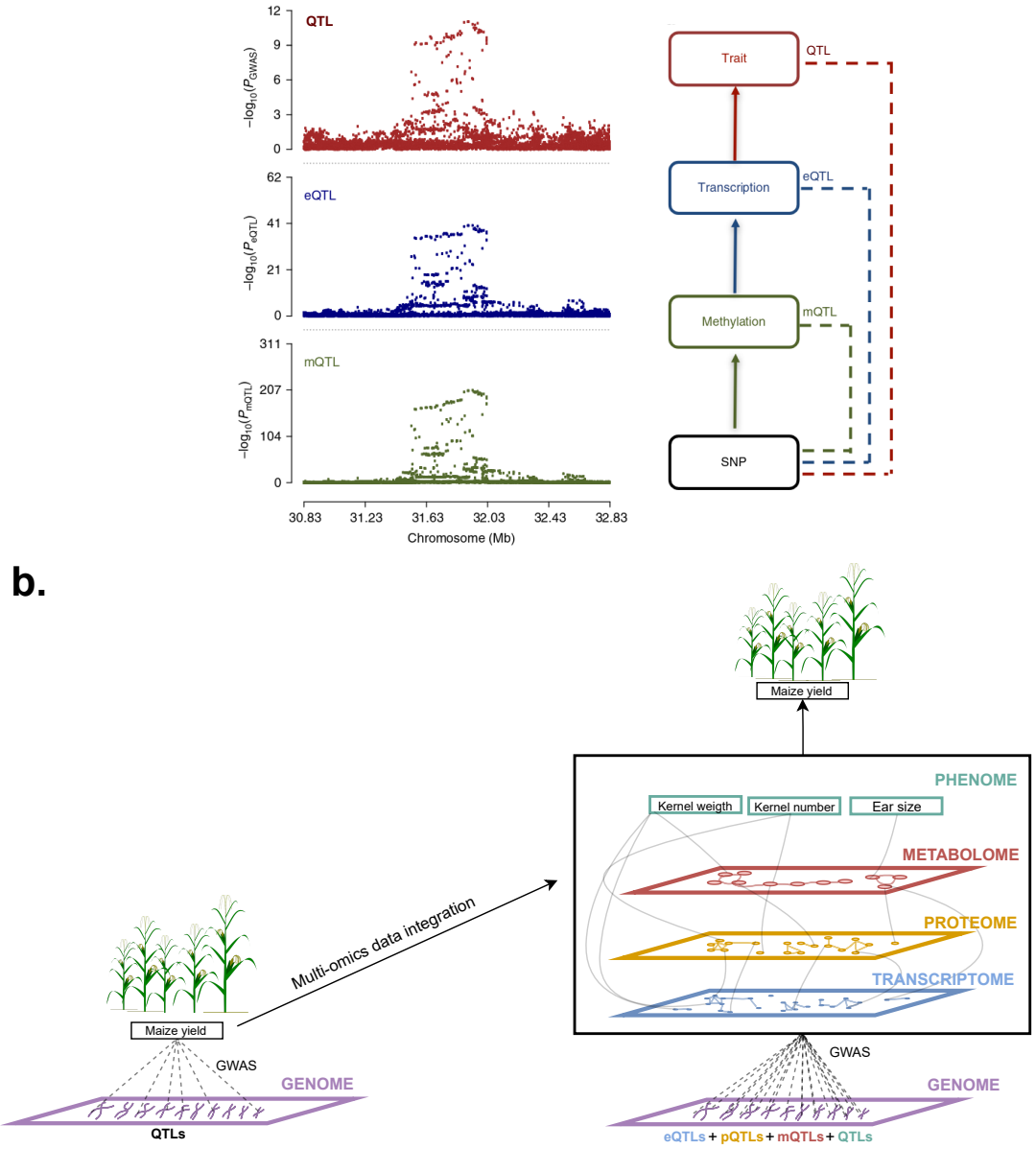


Figure 1.12: **Bridging the gap between the genotype and the phenotype. a.**, Locus comprising QTL, expression-QTL, and methylation-QTL identified from GWAS conducted on trait, transcript and methylation-site (adapted from Wu et al (2018)). The effect of the different SNPs defining the locus decreases from the methylation site to the trait. **b.**, Schematic on system genetic conducted on maize yields. On the left is represented a study on maize yield genetic determinisms based on GWAS performed on a direct measurement of yield. On the right is represented the same study on maize yield genetic determinisms but by following a system genetic approach. This approach leads to the detection of a limited number of QTLs insufficient to elucidate the genetic determinism of such a complex trait. Multi-omics data integration enables inference of interactions between the different molecular and phenotypic traits associated with maize yield. Moreover, GWAS conducted on all added biological entities allows the detection of a richer set of significantly associated loci composed of eQTLs, pQTLs, mQTLs, and QTLs. This approach provides information on both molecular and genetic basis underlying complex traits.

During this Ph.D. I addressed the study of the genetic and molecular basis underlying maize drought response. Indeed, regarding concern about climate change and its repercussions on crop production, the creation of drought-tolerant varieties is becoming more and more urgent. Knowing that maize is nowadays the cereal most cultivated in the world, allowing it to feed both the human population and livestock animals makes it a good model for study. Furthermore, maize is also a model organism extensively researched, which contributes to the generation of a large amount of multi-omics data for integrative analyses. In the next section, I give more details on this biological model and on its related socio-economic and scientific contexts.

1.3 . Elucidating maize drought response

1.3.1 . Maize's rise and stagnation: a production limit to break with global warming

Zea mays originated from the domestication of its wild ancestor teosinte (*Zea mays subsp. parviglumis*) in Mexico approximately 9,000 years ago (Matsuoka et al, 2002). Its domestication was spread along the American continent, resulting in the so-called maize "landraces" varieties adapted for various environments. Maize was introduced in Europe from three principal independent events (Tenaillon and Charcosset, 2011) (Fig. 1.13a). The first introduction occurred in southern Spain in 1493: Colombus brought a Caribbean variety (called tropical) adapted to hot temperatures. The second introduction concerns a northern American variety (the Nothern Flint) in northern Europe during the 16th century (Dubreuil et al, 2006; Camus-Kulandaivelu et al, 2006). The last introduction was from South America to Italy (Tenaillon and Charcosset, 2011). Maize introduction in Europe allowed its cultivation to spread into the rest of the world (Mir et al, 2013). It also allowed the diversification of genetic groups adapted for a wide range of environments. Despite their ability to adapt across Europe, the first maize varieties did not produce high yields (Tenaillon and Charcosset, 2011; Swarts et al, 2017). Because it is an allogamous plant, maize was first cultivated in heterogenous populations called open-pollinated varieties that included heterozygotes individuals. These varieties represented the optimal material for the breeding program before the 1900s.

In 1914, the discovery of hybrid vigor, called heterosis by Shull (1914), revolutionized maize breeding with the implementation of new breeding schemes based on the crossing of two genetically distant individuals. This discovery skyrocketed maize yields between 1937 and 1955, with the use of double hybrid methods where two F1 hybrids were crossed. A second surge in maize yield from 1955 to today was made by the application of a simple hybrid method where two homozygous lines are crossed. Since 1960, maize yields have kept increasing to reach a worldwide yearly production of 1.2 billion tons in 2021 (FAOSTAT), thus making maize the most cultivated cereal in the world (Fig. 1.13b). However, in temperate European regions such as France, maize production has plateaued at 15 million tonnes per year since the 1990s (Fig. 1.13c). Predictions of the future intense drought scenarios caused by global warming (Büntgen et al, 2021; Williams et al, 2022; Jiang and Zhou, 2023; Intergovernmental Panel on Climate Change, 2023) (Fig. 1.13d), combined with maize's susceptibility to drought, which can reduce yields by 20 to 50% (Sah et al, 2020), are raising growing concerns about global food security(Shiferaw et al, 2011; Grote et al, 2021). Developing more drought-tolerant varieties could therefore break the yield plateau observed in developed countries in temperate regions and mitigate future yield losses in a warmer world.

The conception of drought-tolerant maize varieties is far from trivial due to the complexity of drought tolerance. Indeed, drought tolerance is a highly integrated trait resulting from the expression of several morphological, physiological, and molecular traits depending on the phenological stage, the drought scenario, the environment, and the genotype of the plants (Fig. 1.14). Significant improvements in maize drought tolerance will require a better understanding of the genetic and molecular determinants of drought response.

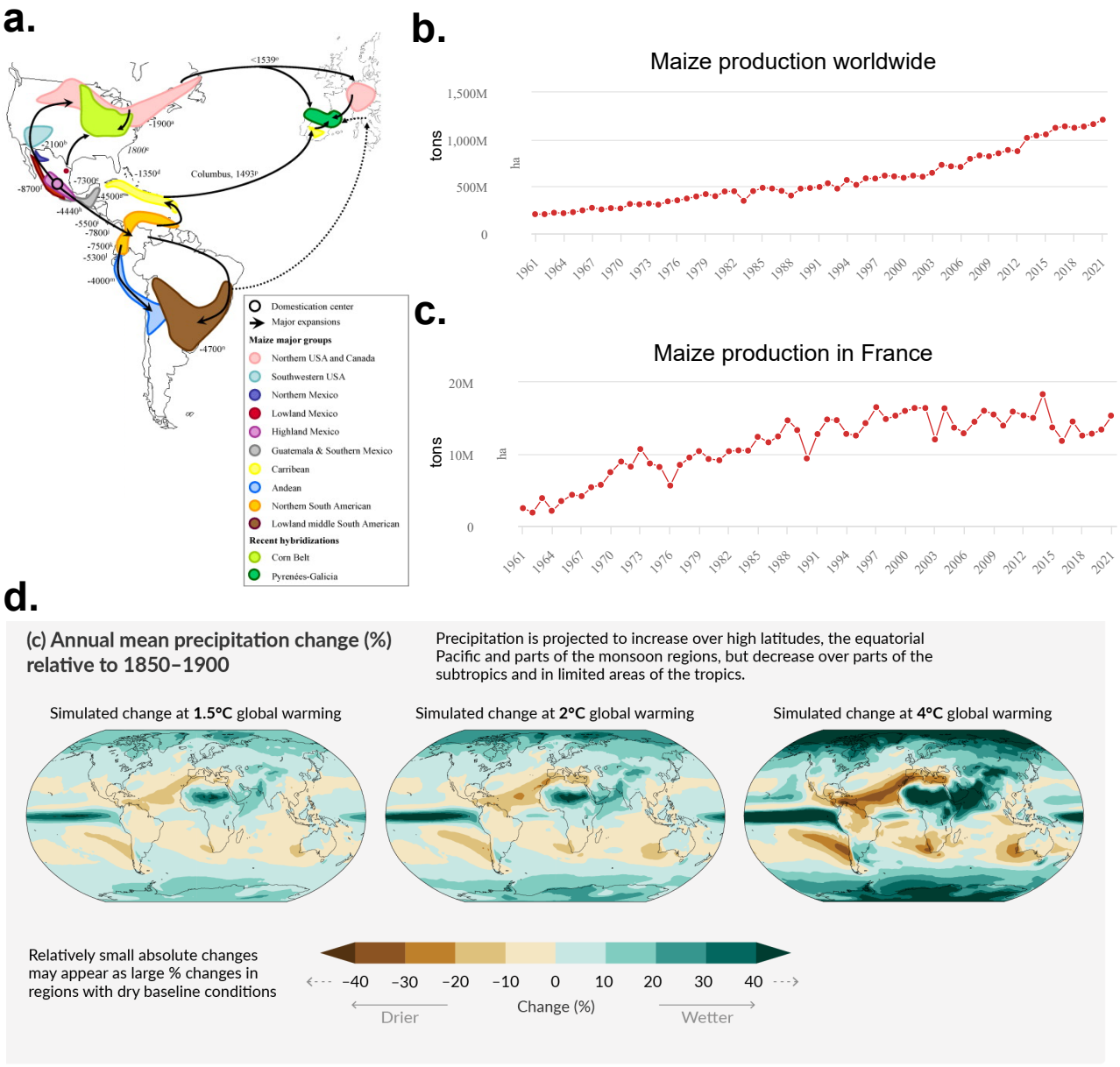


Figure 1.13: **Context and socio-economics issues of maize. a.**, Map of maize introgression in America and Europe (adapted from Tenaillon and Charcosset (2011)). **b.**, Worldwide maize production (FAOSTAT). **c.**, French maize production (FAOSTAT). **d.**, Prediction of precipitation changes relative to 1850-1900 according to temperature increases caused by global warming (Figure from Intergovernmental Panel on Climate Change (2023))

1.3.2 . The complexity of maize drought response

Water scarcity can be very detrimental to maize grain yields, especially at certain stages of development (Sah et al, 2020). The most critical stage is flowering (Fig. 1.15a), which includes the development of silks, *i.e.*, the female reproductive organ that is essential for grain production. A lack of water at flowering can cause up to 50% yield loss (Denmead and Shaw, 1960). At the vegetative and grain-filling stages, water deficit can also lead to significant yield losses (25% and 21%, respectively (Denmead and Shaw, 1960)). Yield losses are related to phenotypic changes due to water stress. The traits whose phenotypes are altered by water deficit include green-leaf duration, plant performance, ear length, seed weight, plant height, number of grains per ear, leaf number, ear per plant, kernel row per ear, kernels per row, and early leaf senescence. The phenotypic changes observed for these traits under water deficit can depend on the genotype. Three mechanisms of adaptation to drought were selected during the evolution of maize (Pamungkas et al 2022 and Fig. 1.15b):

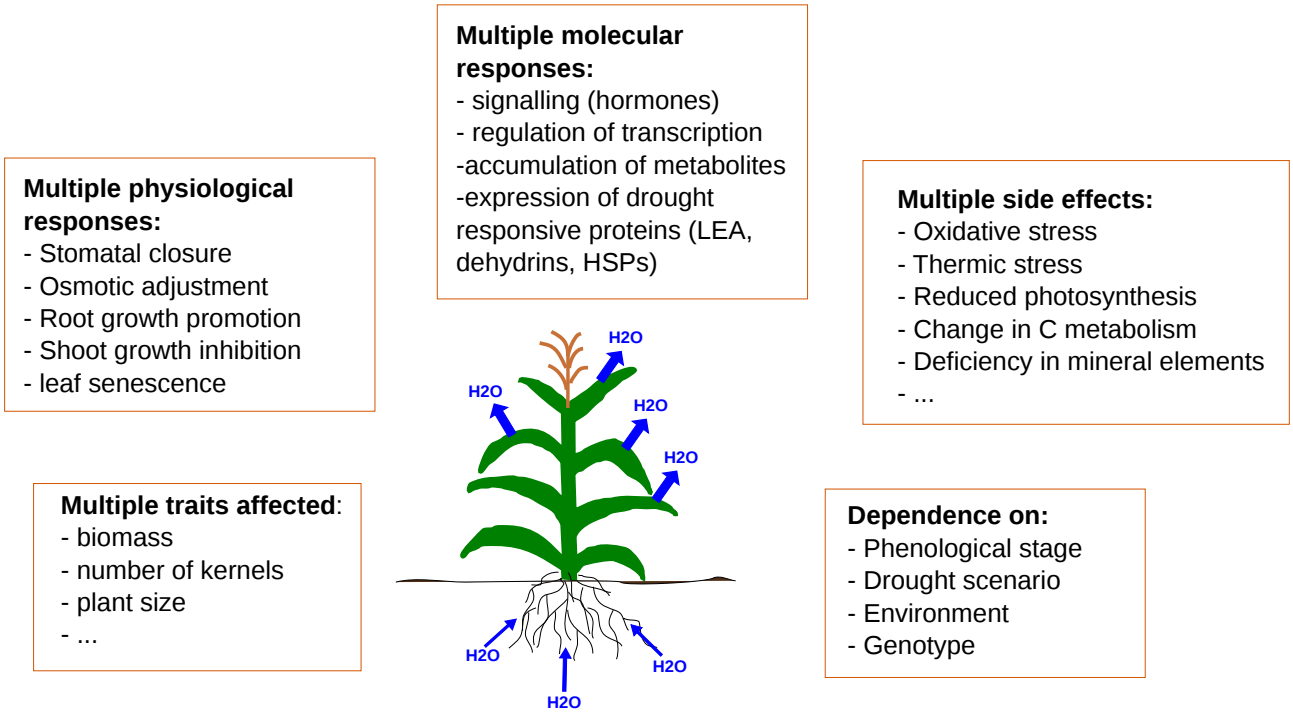


Figure 1.14: **The different factors associated with maize drought response.**

- Drought escape: This mechanism is based on stopping growth (vegetative phase) as soon as possible in order to produce seeds or fruits (generative phase).
- Drought avoidance: Plants that follow this mechanism induce a high water potential in their cells by developing a deep root system and maintaining water uptake.
- Drought tolerance: In this mechanism, plants are able to survive and withstand a water deficit.

These mechanisms include several morpho-physiological changes triggered by complex signaling pathways (Gupta et al, 2020; Dietz et al, 2021; Santini et al, 2022) that involve several types of molecules, such as mitogen-activated protein kinase (MAPK), abscisic acid (ABA), calcium, reactive oxygen species (ROS), and transcription factors (TFs) (Fig. 1.15c). The loss of turgor in plant cells caused by water shortage allows the activation of signaling mechanisms by ABA, calcium, MAPK, and ROS. These four types of molecules lead to the activation of regulatory genes such as TFs, allowing the regulation and accumulation of functional drought-responsive genes, such as dehydrins, aquaporins, late embryogenesis abundant (LEA), and heat shock proteins (HSP) (Valliyodan and Nguyen, 2006; Seki et al, 2007). These different pathways can lead to changes in carbohydrate metabolisms, osmoprotectant synthesis, and accumulation of ROS-scavenging species that influence plant morphology and physiology. These morpho-physiological changes can lead to drought tolerance. For example, ABA accumulation will induce stomatal closure on leaves and then improve water use efficiency by reducing transpiration (Mahmood et al, 2020). Also, the synthesis of osmoprotectants such as sugars or proline leads to osmotic adjustments and maintains turgor pressure in cells. However, as mentioned above, these mechanisms can lead to deleterious developmental effects (Mahmood et al, 2020). For example, stomatal closure reduces internal leaf CO_2, which reduces photosynthesis activity and affects growth, yield, and biomass production (Mahmood et al, 2020).

The negative impact of drought on cereal production has made drought tolerance a trait of interest to the agricultural industry. In fact, according to the FAO, drought is the most costly agricultural disaster, with losses to crop and livestock production estimated at $37 billion, far exceeding those caused by floods, with losses estimated at $21 billion, and storms, with losses estimated at $19 billion. Thus, a better understanding of the genotype-phenotype in the case of maize drought response could help breeders identify interesting loci for designing drought-tolerant varieties (Sheoran et al, 2022). In the following section, I will address the major advances in maize genetics and genomics that allow further steps in our knowledge related to maize drought response.

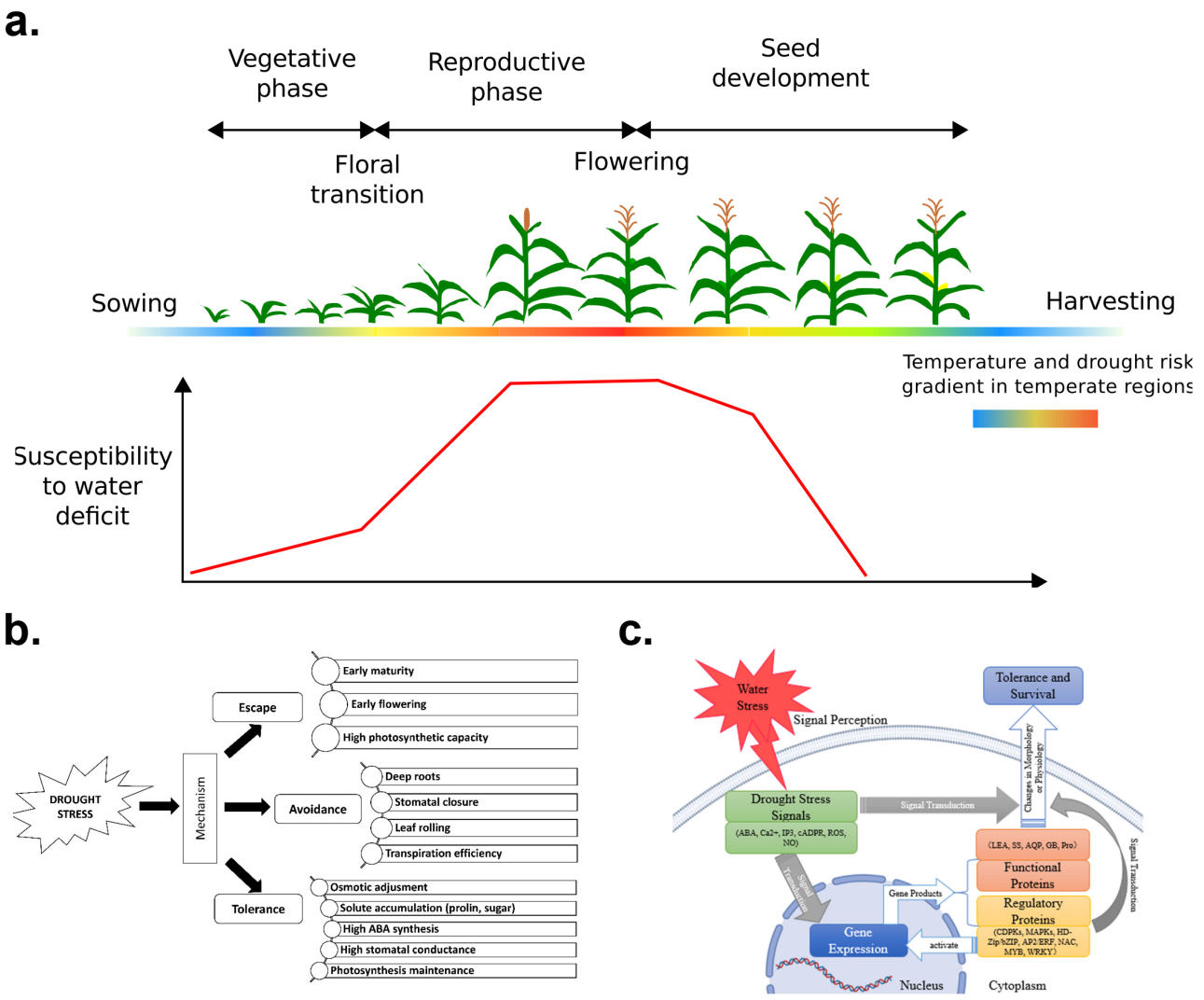


Figure 1.15: **Mechanism related to maize drought response. a.**, Maize susceptibility to water deficit according to its phenological stages. **b.**, The three main mechanisms of drought adaptation (adapted from Pamungkas et al (2022)). **c.**, Molecular signaling pathways related to water stress (adapted from Yang et al (2021)).

1.3.3 . The starting point in the study of the genetic determinism related to maize drought response

In 2009, Schnable et al (2009) published the first reference genome sequence for maize, based on the inbred line B73 (B73_RefGen_v1), enabling the development of extensive research in maize functional genomics (Liu et al, 2020). Nowadays, the B73 genome assembly is at the fifth version (B73 RefGen_v5), which includes 40,621 gene models (Hufford et al, 2021), and the genome of several other maize lines have been sequenced (Hirsch et al, 2016; Yang et al, 2017; Springer et al, 2018; Yang et al, 2019). In particular, high-quality assemblies have been achieved for Mo17, SK, and W22. The vast amount of genomic information generated by the maize community can be retrieved in the Maize Genetics and Genomics Database (MaizeGDB) (Portwood et al, 2019), which is the most valuable database for maize genomics. The comparison between B73 and the other inbred lines allowed the identification of millions of SNPs and structural variants (SV) (Springer et al, 2009; Lai et al, 2010; Jiao et al, 2012; Hirsch et al, 2014; Sun et al, 2018). This high density of molecular markers has expanded our knowledge of the genetic basis of important traits such as yield-related traits (Xiao et al, 2016; Liu et al, 2017), plant height (Pan et al, 2017), and flowering time (Buckler et al, 2009). However, direct selection for yield under drought stress proved ineffective in making maize more tolerant (Ziyomo and Bernardo, 2013). Indeed, drought tolerance integrates several secondary traits (*e.g.*, flowering time, leaf area, root architecture) that are highly dependent on the environment. The study of drought tolerance, therefore, requires phenotyping experiments that measure secondary traits in drought-controlled environments. For example, multi-trial experiments (METs) (Boer et al, 2007) help increase the precision of genotypic means and assess the genotype-by-environment (GxE) interactions that are crucial for drought tolerance breeding. In addition, the development of high-throughput phenotyping

platforms enables the generation of massive phenomics datasets in a more accurate and cost-effective way for complex morpho-physiological traits, including stomatal conductance, evapotranspiration, and water-use efficiency (Cabrera-Bosquet et al, 2016; Alvarez Prado et al, 2017). Together with the advancement in maize genetics and genomics, the dissection of the genetic architecture of trait related to drought becomes the solution of choice to support maize breeding.

The last decades have been relatively productive in detecting drought-responsive QTLs. For example, Almeida et al (2013) identified a total of 145 QTLs associated with grain yield and the anthesis-silking-interval (*i.e.*, the difference in days between the female and male flowering) across multiple environments and different genetic backgrounds. Similarly, Zhao et al (2018) conducted GWAS on plant height, ear height, anthesis-silking-interval, ear weight, cob weight, 100-kernel weight, and ear length under drought and control conditions, resulting in the detection of 69 QTLs explaining 4.0–17.2% of the phenotypic variation. However, the inclusion of such QTLs in maize breeding programs remains scarce due to a lack of validation. Indeed, most of the QTLs identified for drought tolerance are subject to epistasis and environmental interaction, making them unstable across genetic and environmental backgrounds. Nevertheless, this has not prevented the successful validation of a few genes that confer drought tolerance, such as *ZmVPP1* (Wang et al, 2016) and *ZmPP2C-A* (Xiang et al, 2017).

The dissection of maize drought tolerance through the use of omics data has provided a better insight into the complex mechanisms related to drought response (Budak et al, 2015). Up to now, omics data have been extensively used to identify genes or regulatory pathways. For example, Huang et al (2012) highlights differentially expressed proteins that could play a role in drought tolerance. Benešová et al (2012) get a closer look at the changes induced on maize leaf proteome. Several omics studies also reveal that proteins involved in maize drought response are mainly protective proteins (*e.g.*, heat shock proteins) (Benešová et al, 2012; Wang et al, 2019), late embryogenesis abundant proteins (Huang et al, 2012; Wang et al, 2019), and signaling proteins (*e.g.*, auxin repressed protein, serine/threonine protein kinase) (Bonhomme et al, 2012; Li et al, 2021). Among the studies that used omics data to decipher the complex mechanisms related to drought response Blein-Nicolas et al (2020) identified interesting genomics regions on chromosomes 5 and 7 by conducting a system genetics approach integrating genomics, proteomics, and phenomics data.

This Ph.D. is in line with the work of Blein-Nicolas et al (2020). It aimed to gain insight into the genetic and molecular bases of maize drought response using a systems genetics approach. To do this, I analyzed previously published genomics, proteomics, metabolomics, and phenomics data generated during the Amaizing project, a research project financed by the French National Research Agency (ANR). In the following, I will present my thesis objectives in the context of the Amaizing project.

1.4 . Objectives of the thesis

This Ph.D. project is a continuation of the Amaizing project (2011-2020), which was funded by the French National Research Agency (ANR) and supervised by Alain Charcosset, an INRAE senior researcher at the Quantitative genetics and evolution - Le Moulon (GQE) research unit. This project aimed to develop knowledge of selection methods and agricultural practices for high-yielding maize varieties with better environmental values. Twenty-five partners, including fourteen INRAE laboratories, nine private breeding companies, and two technological platforms, worked together to improve our understanding of maize genome organization, adaptation, and plasticity mechanisms. The various studies carried out during this project have led to the generation of a large amount of publicly available genomic and phenomic datasets (Recherche Data Gouv).

In Rincent et al (2014), a diversity panel of 254 maize dent lines from Europe and America was identified and selected for the study of drought tolerance. These lines were selected within a restricted flowering window to avoid confounding the drought escape effect due to variation in flowering time with the expression of genomic regions involved in drought tolerance. Negro et al (2019) identified four major genetic groups within the diversity panel by analyzing and assembling a large genomic data set of 977,459 SNPs obtained from three technologies: 50K Infinium HD Illumina array (Ganal et al, 2011), 600K Axiom Affymetrix array (Unterseer et al, 2014) and genotyping-by-sequencing. Genomic data production allowed Millet et al (2016) to conduct GWAS on this panel. They considered field phenotyping data composed of grain-yield component

traits measured in 25 sites located across Europe. The meta-analysis of the GWAS results obtained for each trait in each location allowed the identification of drought-responsive QTLs. These authors also quantified the contribution of QTLs to the genetic variance of the traits using linear mixed models. Alvarez Prado et al (2017) used the same approach as Millet et al (2016) to study the genetic architecture of drought-responsive traits measured in the high-throughput phenotyping platform Phenoarch (Cabrera-Bosquet et al, 2016). They produced phenomics data that included leaf area, biomass, transportation rate, stomatal conductance, water uptake, and water use efficiency measurements made on hybrids obtained by crossing the 254 dent lines with a flint tester line. These hybrids were grown under well-watered (WW) and water deficit (WD) conditions during four independent trials conducted in spring 2012, spring 2013, winter 2013, and spring 2016. During the spring 2012 trial, leaf samples were collected for proteomic analyses. By using a mass-spectrometry-based proteomics approach, Blein-Nicolas et al (2020) thus quantified 2,055 proteins in collaboration with the PAPPSO proteomics platform. They used the proteomics data to infer protein co-expression networks for each watering condition and to detect nearly 20,000 pQTLs. By integrating genomics, proteomics, and phenomics data, they identified colocalizations between the pQTLs detected in their study and the QTLs previously identified by Alvarez Prado et al (2017). Finally, during the spring 2013 experiment, leaf samples were collected for metabolomic analyses. Metabolomics data were generated by the Bordeaux Metabolome platform using a mass-spectrometry-based approach and used for phenotypic predictions (Prigent et al., in prep). The unique multi-omic dataset generated during the Amaizing project (Fig. 1.16) provides an opportunity to take a holistic view of maize drought response. Indeed, an integrative analysis of the four biological complexity levels available, genomics, proteomics, metabolomics, and phenomics, could give further insight in plant drought tolerance mechanisms. Thus, the aim of my thesis was to gain insight into the genetic and molecular bases of maize drought response using an integrative approach based on the multi-omics data acquired during the Amazing project. To achieve this goal, I addressed the two following issues:

- **Deciphering the genetic determinism of maize drought response**: Drought tolerance results from adaptative changes that occur in molecular and phenotypic traits in response to water deficit. Thus, the identification of genetic regions involved in the genotype-by-water availability interaction (GxW) would be particularly relevant to deciphering the genetic bases of drought response. The identification of such genetic regions is commonly made by conducting GWAS on phenotypic measurements made on plants grown in well-watered (WW) and water deficit (WD) conditions and selecting the condition-specific QTLs. However, as plasticity is defined as the ability of a genotype to produce different phenotypes in response to environmental changes (Bradshaw, 1965), we asked whether performing GWAS on plasticity indices would improve the detection of QTL involved in GxW. To address this question, I calculated the WD/WW ratio for each trait and for each genotype and performed GWAS on these plasticity indices. Then, I compared plasticity QTLs *vs* QTLs detected on WW and WD phenotypic means in terms of genomic position and their ability to capture the GxW variance of the traits. I found that plasticity QTLs were located in different genetic regions than QTLs, and that plasticity QTLs explained exclusively from 60 to 100% of the GxW traits' variance. These results are presented in the first section of the Chapter 2, corresponding to a peer-reviewed article Djabali et al (2023), accepted in the journal *Theoretical and Applied Genetics*. The results obtained in this first section give me intuition on the genetic control of the studied traits' plasticity. Thus, the second section addressed a functional analysis of plasticity QTLs to identify further evidence on the gene-regulatory model of plasticity.

- **Integration of proteomics and metabolomics data to provide an in-depth genetic and molecular characterization of maize response to water deficit**: The major goal of this chapter was to assess the benefits of omics data integration to better characterize complex traits genetic determinism such as drought response. Thus, I conducted a system genetics approach on maize water stress response by integrating molecular data. In the first section, I was able to harness proteomics data to i) translate genetic regions where are located QTLs, explaining one-third of the GxW variance, into protein-protein networks that linked proteins from proteomics data and proteins encoded by genes located in the genetic regions; ii) identify pQTLs capturing missing heritability for drought response by inferring multi-scale networks integrating genomics, proteomics, and phenomics; iii) prioritize genetics and molecular targets, which could explain nearly 10% of the GxW variance. The second section addressed the ability of mQTLs to unravel missing heritability by conducting the same integrative approach as the one used in the first section.

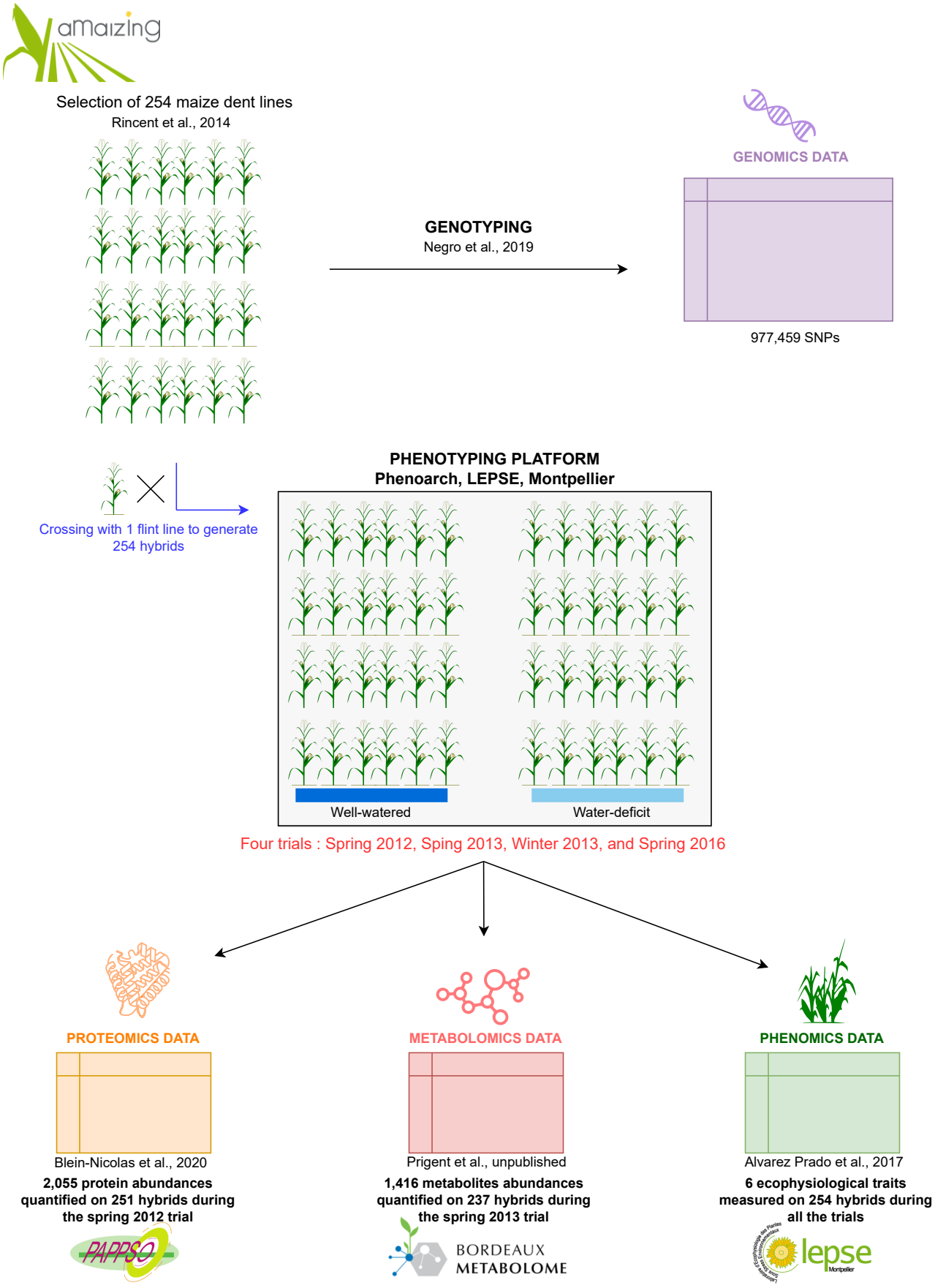


Figure 1.16: **Omics data generated during the Amaizing project and exploited during this thesis**

Bibliography

Almeida GD, Makumbi D, Magorokosho C, et al (2013) QTL mapping in three tropical maize populations reveals a set of constitutive and adaptive genomic regions for drought tolerance. Theoretical and Applied Genetics 126(3):583–600. https://doi.org/10.1007/s00122-012-2003-7

Alvarez Prado S, Cabrera-Bosquet L, Grau A, et al (2017) 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

Amiteye S (2021) Basic concepts and methodologies of DNA marker systems in plant molecular breeding. Heliyon 7(10):e08,093. https://doi.org/10.1016/j.heliyon.2021.e08093

Argelaguet R, Velten B, Arnol D, et al (2018) Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. Molecular Systems Biology 14(6):e8124. https://doi.org/10.15252/msb.20178124

Argelaguet R, Cuomo ASE, Stegle O, et al (2021) Computational principles and challenges in single-cell data integration. Nature Biotechnology 39(10):1202–1215. https://doi.org/10.1038/s41587-021-00895-7

Ashrafian H, Sounderajah V, Glen R, et al (2020) Metabolomics: The Stethoscope for the Twenty-First Century. Medical Principles and Practice 30(4):301–310. https://doi.org/10.1159/000513545

Aslam B, Basit M, Nisar MA, et al (2017) Proteomics: Technologies and Their Applications. Journal of Chromatographic Science 55(2):182–196. https://doi.org/10.1093/chromsci/bmw167

Astle W, Balding DJ (2009) Population Structure and Cryptic Relatedness in Genetic Association Studies. Statistical Science 24(4):451–471. https://doi.org/10.1214/09-STS307

Avery OT, MacLeod CM, McCarty M (1944) STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. Journal of Experimental Medicine 79(2):137–158. https://doi.org/10.1084/jem.79.2.137

Bateson W, Mendel G (1902) Mendel's Principles of Heredity: A Defence, with a Translation of Mendel's Original Papers on Hybridisation

Benešová M, Holá D, Fischer L, et al (2012) The Physiology and Proteomics of Drought Tolerance in Maize: Early Stomatal Closure as a Cause of Lower Tolerance to Short-Term Dehydration? PLOS ONE 7(6):e38,017. https://doi.org/10.1371/journal.pone.0038017

Berget SM, Moore C, Sharp PA (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proceedings of the National Academy of Sciences of the United States of America 74(8):3171–3175

Blein-Nicolas M, Negro SS, Balliau T, et al (2020) A systems genetics approach reveals environment-dependent associations between SNPs, protein coexpression, and drought-related traits in maize. Genome Research 30(11):1593–1604. https://doi.org/10.1101/gr.255224.119

Boer MP, Wright D, Feng L, et al (2007) A Mixed-Model Quantitative Trait Loci (QTL) Analysis for Multiple-Environment Trial Data Using Environmental Covariables for QTL-by-Environment Interactions, With an Example in Maize. Genetics 177(3):1801–1813. https://doi.org/10.1534/genetics.107.071068

Bonhomme L, Valot B, Tardieu F, et al (2012) Phosphoproteome Dynamics Upon Changes in Plant Water Status Reveal Early Events Associated With Rapid Growth Adjustment in Maize Leaves*. Molecular & Cellular Proteomics 11(10):957–972. https://doi.org/10.1074/mcp.M111.015867

Boyle EA, Li YI, Pritchard JK (2017) An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169(7):1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Bradshaw AD (1965) Evolutionary Significance of Phenotypic Plasticity in Plants. In: Caspari EW, Thoday JM (eds) Advances in Genetics, vol 13. Academic Press, p 115–155, https://doi.org/10.1016/S0065-2660(08)60048-6

Buckler ES, Holland JB, Bradbury PJ, et al (2009) The Genetic Architecture of Maize Flowering Time. Science 325(5941):714–718. https://doi.org/10.1126/science.1174276

Budak H, Hussain B, Khan Z, et al (2015) From Genetics to Functional Genomics: Improvement in Drought Signaling and Tolerance in Wheat. Frontiers in Plant Science 6. https://doi.org/10.3389/fpls.2015.01012

Büntgen U, Urban O, Krusic PJ, et al (2021) Recent European drought extremes beyond Common Era background variability. Nature Geoscience 14(4):190–196. https://doi.org/10.1038/s41561-021-00698-0

Cabrera-Bosquet L, Fournier C, Brichet N, et al (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

Camus-Kulandaivelu L, Veyrieras JB, Madur D, et al (2006) Maize Adaptation to Temperate Climate: Relationship Between Population Structure and Polymorphism in the Dwarf8 Gene. Genetics 172(4):2449–2463. https://doi.org/10.1534/genetics.105.048603

Cardon LR, Bell JI (2001) Association study designs for complex diseases. Nature Reviews Genetics 2(2):91–99. https://doi.org/10.1038/35052543

Civelek M, Lusis AJ (2014) Systems genetics approaches to understand complex traits. Nature Reviews Genetics 15(1):34–48. https://doi.org/10.1038/nrg3575

Crick FH (1958) On protein synthesis. Symposia of the Society for Experimental Biology 12:138–163

de Vienne D (2022) What is a phenotype? History and new developments of the concept. Genetica 150(3):153–158. https://doi.org/10.1007/s10709-021-00134-6

Denmead OT, Shaw RH (1960) The Effects of Soil Moisture Stress at Different Stages of Growth on the Development and Yield of Corn1. Agronomy Journal 52(5):272–274. https://doi.org/10.2134/agronj1960.00021962005200050010x

Dietz KJ, Zörb C, Geilfus CM (2021) Drought and crop yield. Plant Biology 23(6):881–893. https://doi.org/10.1111/plb.13304

Djabali Y, Rincent R, Martin ML, et al (2023) Drought plasticity QTLs specifically contribute to the genotype x water availability interaction in maize

Dubreuil P, Warburton M, Chastanet M, et al (2006) More on the introduction of temperate maize into Europe: Large-scale bulk SSR genotyping and new historical elements. Maydica 51:281–291

Duggirala R, Blangero J, Almasy L, et al (1999) Linkage of Type 2 Diabetes Mellitus and of Age at Onset to a Genetic Location on Chromosome 10q in Mexican Americans. The American Journal of Human Genetics 64(4):1127–1140. https://doi.org/10.1086/302316

Félix MA, Barkoulas M (2015) Pervasive robustness in biological systems. Nature Reviews Genetics 16(8):483–496. https://doi.org/10.1038/nrg3949

Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant-breeding programme. Australian Journal of Agricultural Research 14(6):742–754. https://doi.org/10.1071/ar9630742

Fisher RA (1918) XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. Earth and Environmental Science Transactions of The Royal Society of Edinburgh 52(2):399–433. https://doi.org/10.1017/S0080456800012163

Fisher RA (1930) The Genetical Theory of Natural Selection. The Genetical Theory of Natural Selection, Clarendon Press, Oxford, England, https://doi.org/10.5962/bhl.title.27468

Ganal MW, Durstewitz G, Polley A, et al (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(12):e28,334. https://doi.org/10.1371/journal.pone.0028334

Gayon J (2016) From Mendel to epigenetics: History of genetics. Comptes Rendus Biologies 339(7):225–230. https://doi.org/10.1016/j.crvi.2016.05.009

Gillespie JH, Turelli M (1989) Genotype-environment interactions and the maintenance of polygenic variation. Genetics 121(1):129–138. https://doi.org/10.1093/genetics/121.1.129

Griffith F (1928) The Significance of Pneumococcal Types. Epidemiology & Infection 27(2):113–159. https://doi.org/10.1017/S0022172400031879

Grote U, Fasse A, Nguyen TT, et al (2021) Food Security and the Dynamics of Wheat and Maize Value Chains in Africa and Asia. Frontiers in Sustainable Food Systems 4. https://doi.org/10.3389/fsufs.2020.617009

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

Hawe JS, Theis FJ, Heinig M (2019) Inferring Interaction Networks From Multi-Omics Data. Frontiers in Genetics 10. https://doi.org/10.3389/fgene.2019.00535

Hirsch CN, Foerster JM, Johnson JM, et al (2014) Insights into the Maize Pan-Genome and Pan-Transcriptome. The Plant Cell 26(1):121–135. https://doi.org/10.1105/tpc.113.119982

Hirsch CN, Hirsch CD, Brohammer AB, et al (2016) Draft Assembly of Elite Inbred Line PH207 Provides Insights into Genomic and Transcriptome Diversity in Maize. The Plant Cell 28(11):2700–2714. https://doi.org/10.1105/tpc.16.00353

Holzinger ER, Dudek SM, Frase AT, et al (2014) ATHENA: The analysis tool for heritable and environmental network associations. Bioinformatics 30(5):698–705. https://doi.org/10.1093/bioinformatics/btt572

Hu X, Wang H, Li K, et al (2017) Genome-wide proteomic profiling reveals the role of dominance protein expression in heterosis in immature maize ears. Scientific Reports 7(1):16,130. https://doi.org/10.1038/s41598-017-15985-3

Huang H, Møller IM, Song SQ (2012) Proteomics of desiccation tolerance during development and germination of maize embryos. Journal of Proteomics 75(4):1247–1262. https://doi.org/10.1016/j.jprot.2011.10.036

Hufford MB, Seetharam AS, Woodhouse MR, et al (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

Intergovernmental Panel on Climate Change (2023) Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 1st edn. Cambridge University Press, https://doi.org/10.1017/9781009157896

Jiang J, Zhou T (2023) Agricultural drought over water-scarce Central Asia aggravated by internal climate variability. Nature Geoscience 16(2):154–161. https://doi.org/10.1038/s41561-022-01111-0

Jiao Y, Zhao H, Ren L, et al (2012) Genome-wide genetic changes during modern breeding of maize. Nature Genetics 44(7):812–815. https://doi.org/10.1038/ng.2312

Johannsen W (1911) The Genotype Conception of Heredity. The American Naturalist 45(531):129–159. https://doi.org/10.1086/279202

Kim D, Shin H, Song YS, et al (2012) Synergistic effect of different levels of genomic data for cancer clinical outcome prediction. Journal of Biomedical Informatics 45(6):1191–1198. https://doi.org/10.1016/j.jbi.2012.07.008

Kimura M (1983) The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9780511623486

Lai J, Li R, Xu X, et al (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. Nature Genetics 42(11):1027–1030. https://doi.org/10.1038/ng.684

Lanckriet GRG, De Bie T, Cristianini N, et al (2004) A statistical framework for genomic data fusion. Bioinformatics (Oxford, England) 20(16):2626–2635. https://doi.org/10.1093/bioinformatics/bth294

Lander ES, Botstein D (1989) Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps. Genetics 121(1):185–199

Langfelder P, Horvath S (2008) WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics 9(1):559. https://doi.org/10.1186/1471-2105-9-559

Li H, Yue H, Xie J, et al (2021) Transcriptomic profiling of the high-vigour maize (Zea mays L.) hybrid variety response to cold and drought stresses during seed germination. Scientific Reports 11(1):19,345. https://doi.org/10.1038/s41598-021-98907-8

Liu J, Huang J, Guo H, et al (2017) The Conserved and Unique Genetic Architecture of Kernel Size and Weight in Maize and Rice. Plant Physiology 175(2):774–785. https://doi.org/10.1104/pp.17.00708

Liu J, Fernie AR, Yan J (2020) The Past, Present, and Future of Maize Improvement: Domestication, Genomics, and Functional Genomic Routes toward Crop Enhancement. Plant Communications 1(1):100,010. https://doi.org/10.1016/j.xplc.2019.100010

López de Maturana E, Alonso L, Alarcón P, et al (2019) Challenges in the Integration of Omics and Non-Omics Data. Genes 10(3):238. https://doi.org/10.3390/genes10030238

Lowe R, Shirley N, Bleackley M, et al (2017) Transcriptomics technologies. PLOS Computational Biology 13(5):e1005,457. https://doi.org/10.1371/journal.pcbi.1005457

Mahmood T, Khalid S, Abdullah M, et al (2020) Insights into Drought Stress Signaling in Plants and the Molecular Genetic Basis of Cotton Drought Tolerance. Cells 9(1):105. https://doi.org/10.3390/cells9010105

Manolio TA, Collins FS, Cox NJ, et al (2009) Finding the missing heritability of complex diseases. Nature 461(7265):747–753. https://doi.org/10.1038/nature08494

Matsuoka Y, Vigouroux Y, Goodman MM, et al (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proceedings of the National Academy of Sciences of the United States of America 99(9):6080–6084. https://doi.org/10.1073/pnas.052125199

Mendel G (1865) EXPERIMENTS IN PLANT HYBRIDIZATION (1865)

Millet EJ, Welcker C, Kruijer W, et al (2016) Genome-Wide Analysis of Yield in Europe: Allelic Effects Vary with Drought and Heat Scenarios. Plant Physiology 172(2):749–764. https://doi.org/10.1104/pp.16.00621

Mir C, Zerjal T, Combes V, et al (2013) Out of America: Tracing the genetic footprints of the global diffusion of maize. Theoretical and Applied Genetics 126(11):2671–2682. https://doi.org/10.1007/s00122-013-2164-z

Morgan TH (1920) The Physical Basis of Heredity. https://philpapers.org/rec/MORTPB

Morrison M (2002) Modelling Populations: Pearson and Fisher on Mendelism and Biometry. The British Journal for the Philosophy of Science 53(1):39–68. https://doi.org/10.1093/bjps/53.1.39, https://arxiv.org/abs/3541640

Morton NE (1996) Logarithm of odds (lods) for linkage in complex inheritance. Proceedings of the National Academy of Sciences of the United States of America 93(8):3471–3476

Mullis K, Faloona F, Scharf S, et al (1986) Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. 1986. Biotechnology (Reading, Mass) 24:17–27

Negro SS, Millet EJ, Madur D, et al (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

Olby R (1989) The Dimensions of Scientific Controversy: The Biometric-Mendelian Debate. The British Journal for the History of Science 22(3):299–320. https://doi.org/10.1017/S0007087400026170, https://arxiv.org/abs/4026898

Pamungkas SST, Suwarto, Suprayogi, et al (2022) Drought Stress: Responses and Mechanism in Plants. Reviews in Agricultural Science 10:168–185. https://doi.org/10.7831/ras.10.0_168

Pan Q, Xu Y, Li K, et al (2017) The Genetic Basis of Plant Architecture in 10 Maize Recombinant Inbred Line Populations. Plant Physiology 175(2):858–873. https://doi.org/10.1104/pp.17.00709

Picard M, Scott-Boyer MP, Bodein A, et al (2021) Integration strategies of multi-omics data for machine learning analysis. Computational and Structural Biotechnology Journal 19:3735–3746. https://doi.org/10.1016/j.csbj.2021.06.030

Pierce BA (2005) Genetics: A Conceptual Approach, 2nd edn. W.H. Freeman, New York

Porcu E, Sadler MC, Lepik K, et al (2021) Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. Nature Communications 12(1):5647. https://doi.org/10.1038/s41467-021-25805-y

Portwood JLII, Woodhouse MR, Cannon EK, et al (2019) MaizeGDB 2018: The maize multi-genome genetics and genomics database. Nucleic Acids Research 47(D1):D1146–D1154. https://doi.org/10.1093/nar/gky1046

Rincent R, Nicolas S, Bouchet S, et al (2014) Dent and Flint maize diversity panels reveal important genetic potential for increasing biomass production. Theoretical and Applied Genetics 127(11):2313–2331. https://doi.org/10.1007/s00122-014-2379-7

Ritchie MD, Holzinger ER, Li R, et al (2015) Methods of integrating data to uncover genotype–phenotype interactions. Nature Reviews Genetics 16(2):85–97. https://doi.org/10.1038/nrg3868

Roll-Hansen N (2014) The holist tradition in twentieth century genetics. Wilhelm Johannsen's genotype concept. The Journal of Physiology 592(Pt 11):2431–2438. https://doi.org/10.1113/jphysiol.2014.272120

Sah RP, Chakraborty M, Prasad K, et al (2020) Impact of water deficit stress in maize: Phenology and yield components. Scientific Reports 10(1):2944. https://doi.org/10.1038/s41598-020-59689-7

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74(12):5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Santini M, Noce S, Antonelli M, et al (2022) Complex drought patterns robustly explain global yield loss for major crops. Scientific Reports 12(1):5792. https://doi.org/10.1038/s41598-022-09611-0

Scheiner SM (1993) Genetics and Evolution of Phenotypic Plasticity. Annual Review of Ecology and Systematics 24:35–68. https://doi.org/10.1146/annurev.es.24.110193.000343, https://arxiv.org/abs/2097172

Schnable P, Ware D, Fulton R, et al (2009) The B73 maize genome: Complexity, diversity, and dynamics. Science 326:1112–5

Seki M, Umezawa T, Urano K, et al (2007) Regulatory metabolic networks in drought stress responses. Current Opinion in Plant Biology 10(3):296–302. https://doi.org/10.1016/j.pbi.2007.04.014

Sheoran S, Kaur Y, Kumar S, et al (2022) Recent Advances for Drought Stress Tolerance in Maize (Zea mays L.): Present Status and Future Prospects. Frontiers in Plant Science 13. https://doi.org/10.3389/fpls.2022.872566

Shi H, Kichaev G, Pasaniuc B (2016) Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. The American Journal of Human Genetics 99(1):139–153. https://doi.org/10.1016/j.ajhg.2016.05.013

Shiferaw B, Prasanna B, Hellin J, et al (2011) Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. Food Secur 3:307

Shull GH (1914) Duplicate genes for capsule-form inBursa bursa-pastoris. Zeitschrift für induktive Abstammungs- und Vererbungslehre 12(1):97–149. https://doi.org/10.1007/BF01837282

Singh A, Shannon CP, Gautier B, et al (2019) DIABLO: An integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics 35(17):3055–3062. https://doi.org/10.1093/bioinformatics/bty1054

Springer NM, Ying K, Fu Y, et al (2009) Maize Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/Absence Variation (PAV) in Genome Content. PLOS Genetics 5(11):e1000,734. https://doi.org/10.1371/journal.pgen.1000734

Springer NM, Anderson SN, Andorf CM, et al (2018) The maize W22 genome provides a foundation for functional genomics and transposon biology. Nature Genetics 50(9):1282–1288. https://doi.org/10.1038/s41588-018-0158-0

Sun S, Zhou Y, Chen J, et al (2018) Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. Nature Genetics 50(9):1289–1295. https://doi.org/10.1038/s41588-018-0182-0

Sutton WS (1903) The Chromosomes in Heredity. Biological Bulletin 4(5):231–251. https://doi.org/10.2307/1535741, https://arxiv.org/abs/1535741

Swarts K, Gutaker RM, Benz B, et al (2017) Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. Science 357(6350):512–515. https://doi.org/10.1126/science.aam9425

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theoretical and Applied Genetics 92(2):191–203. https://doi.org/10.1007/BF00223376

Tenaillon MI, Charcosset A (2011) A European perspective on maize history. Comptes Rendus Biologies 334(3):221–228. https://doi.org/10.1016/j.crvi.2010.12.015

Unterseer S, Bauer E, Haberer G, et al (2014) A powerful tool for genome analysis in maize: Development and evaluation of the high density 600 k SNP genotyping array. BMC Genomics 15(1):823. https://doi.org/10.1186/1471-2164-15-823

Valliyodan B, Nguyen HT (2006) Understanding regulatory networks and engineering for enhanced drought tolerance in plants. Current Opinion in Plant Biology 9(2):189–195. https://doi.org/10.1016/j.pbi.2006.01.019

van der Sijde MR, Ng A, Fu J (2014) Systems genetics: From GWAS to disease pathways. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1842(10):1903–1909. https://doi.org/10.1016/j.bbadis.2014.04.025

Via S, Gomulkiewicz R, De Jong G, et al (1995) Adaptive phenotypic plasticity: Consensus and controversy. Trends in Ecology & Evolution 10(5):212–217. https://doi.org/10.1016/S0169-5347(00)89061-8

Visscher PM (2008) Sizing up human height variation. Nature Genetics 40(5):489–490. https://doi.org/10.1038/ng0508-489

Visscher PM, Goddard ME (2019) From R.A. Fisher's 1918 Paper to GWAS a Century Later. Genetics 211(4):1125–1130. https://doi.org/10.1534/genetics.118.301594

Visscher PM, Wray NR, Zhang Q, et al (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation. American Journal of Human Genetics 101(1):5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Walton R, Kimber M, Rockett K, et al (2005) Haplotype block structure of the cytochrome P450 CYP2C gene cluster on chromosome 10. Nature Genetics 37(9):915–916; author reply 916. https://doi.org/10.1038/ng0905-915

Wang KC, Chang HY (2018) Epigenomics—Technologies and Applications. Circulation research 122(9):1191–1199. https://doi.org/10.1161/CIRCRESAHA.118.310998

Wang X, Wang H, Liu S, et al (2016) Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. Nature Genetics 48(10):1233–1241. https://doi.org/10.1038/ng.3636

Wang X, Zenda T, Liu S, et al (2019) Comparative Proteomics and Physiological Analyses Reveal Important Maize Filling-Kernel Drought-Responsive Genes and Metabolic Pathways. International Journal of Molecular Sciences 20(15):3743. https://doi.org/10.3390/ijms20153743

Watson JD, Crick FHC (1953) Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature 171(4356):737–738. https://doi.org/10.1038/171737a0

Williams AP, Cook BI, Smerdon JE (2022) Rapid intensification of the emerging southwestern North American megadrought in 2020–2021. Nature Climate Change 12(3):232–234. https://doi.org/10.1038/s41558-022-01290-z

Wu Y, Zeng J, Zhang F, et al (2018) Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Nature Communications 9(1):918. https://doi.org/10.1038/s41467-018-03371-0

Xiang Y, Sun X, Gao S, et al (2017) Deletion of an Endoplasmic Reticulum Stress Response Element in a ZmPP2C-A Gene Facilitates Drought Tolerance of Maize Seedlings. Molecular Plant 10(3):456–469. https://doi.org/10.1016/j.molp.2016.10.003

Xiao Y, Tong H, Yang X, et al (2016) Genome-wide dissection of the maize ear genetic architecture using multiple populations. New Phytologist 210(3):1095–1106. https://doi.org/10.1111/nph.13814

Yang J, Benyamin B, McEvoy BP, et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nature Genetics 42(7):565–569. https://doi.org/10.1038/ng.608

Yang N, Xu XW, Wang RR, et al (2017) Contributions of Zea mays subspecies mexicana haplotypes to modern maize. Nature Communications 8(1). https://doi.org/10.1038/s41467-017-02063-5

Yang N, Liu J, Gao Q, et al (2019) Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. Nature Genetics 51(6):1052–1059. https://doi.org/10.1038/s41588-019-0427-6

Yang X, Lu M, Wang Y, et al (2021) Response Mechanism of Plants to Drought Stress. Horticulturae 7(3):50. https://doi.org/10.3390/horticulturae7030050

Yates F, Cochran WG (1938) The analysis of groups of experiments. The Journal of Agricultural Science 28(4):556–580. https://doi.org/10.1017/S0021859600050978

Yu J, Pressoir G, Briggs WH, et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38(2):203–208. https://doi.org/10.1038/ng1702

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136(4):1457–1468. https://doi.org/10.1093/genetics/136.4.1457

Zhao X, Peng Y, Zhang J, et al (2018) Identification of QTLs and Meta-QTLs for Seven Agronomic Traits in Multiple Maize Populations under Well-Watered and Water-Stressed Conditions. Crop Science 58(2):507–520. https://doi.org/10.2135/cropsci2016.12.0991

Ziyomo C, Bernardo R (2013) Drought Tolerance in Maize: Indirect Selection through Secondary Traits versus Genomewide Selection. Crop Science 53(4):1269–1275. https://doi.org/10.2135/cropsci2012.11.0651

Chapter 2

Improve the detection of genetic determinants involved in the water stress response by considering inter-trial variation and plasticity indices

2.1 . Standfirst

The following section addresses the genetic basis of maize drought response by studying phenomics data previously published by Alvarez Prado et al (2017). This dataset was composed of six drought-responsive traits measured on 254 maize hybrids grown under well-watered (WW) and water-deficit (WD) conditions in four repeated platform trials in spring 2012, spring 2013, winter 2013, and spring 2016. Alvarez Prado et al (2017) performed single-trial genome-wide association studies (GWAS) on each trait, measured in each watering condition in each trial, leading to the identification of 531 QTLs.

The principal goal of this study was to assess the relevance of detecting plasticity QTLs to explain the genotype by water availability (GxW) interaction observed for ecophysiological traits. As we were interested in studying the genetic basis of maize drought response in general, we also decided to evaluate to what extent other non-controlled environmental factors present during phenotyping experiments and not considered in single-trial GWAS could affect QTL detection. Indeed, in platform experiments, plants are subject to uncontrolled environmental variations, for example, in incident light or vapor pressure deficit. Thus, for the experiments aiming to study the effect of one specific component of the environment, it is necessary to be able to separate the effect of this component from the other residual components of the environment. To our knowledge, the contribution of plasticity QTLs in genetic variance and the impact of dissociating watering treatment effects from the other environmental factor effects in QTL detection were never studied.

To deal with these two points, I address the following questions:

- What are the consequences of taking into account the residual environmental effects of the trials in GWAS for the detection of QTLs?
- What is the benefit of conducting GWAS on phenotypic plasticity for studying the genetic architecture of maize water stress response?

During this work, I was able to demonstrate, on the one hand, that QTLs detected with multi-trial GWAS models captured more of the GxW variance than the genotype by trial (GxT) interaction variance compared to the QTLs detected in Alvarez Prado et al (2017). On the other hand, I showed that 90% of the plasticity QTLs identified do not overlap with QTLs detected on phenotypic means, and these plasticity QTLs captured exclusively from 60 to 100% of GxW variance depending on traits. The results of this study give further evidence of the importance of considering phenotypic plasticity to identify genetic polymorphisms involved in stress response. The high proportion of non-overlapping plasticity QTLs is consistent with the results previously found by Kusmec et al (2017) and Diouf et al (2020). It supports that phenotypic plasticity and phenotypic means are independent, as stipulated by Bradshaw (1965). It also suggests that the drought response is mainly driven by the gene regulatory model, where plasticity results from epistasis between regulatory and structural genes. Following the gene regulatory model implying that plasticity results

from epistasis between regulatory and structural genes, my interpretation is that the genetic factors under plasticity QTLs could regulate the genetic factors under QTLs.

To verify this assumption, we decided to continue this work by doing a functional analysis of the QTLs and the plasticity QTLs identified. This work was realized by Amal Ksontini, an M1 student intern under the supervision of Marie-Laure and me. Her work is summarized in section 2.3, and her internship report is attached in the appendix.

2.2 . Plasticity QTLs specifically contribute to the genotype x water availability interaction in maize

This section corresponds to a peer-reviewed publication accepted at *Theoretical and Applied Genetics (TAG)* DOI: 10.1007/s00122-023-04458-z

Plasticity QTLs specifically contribute to the genotype x water availability interaction in maize

Yacine Djabali[1,2,3], Renaud Rincent[3], Marie-Laure Martin[1,2,4*] and Mélisande Blein-Nicolas[3*]

[1]Université Paris-Saclay, CNRS, INRAE, Université Evry, Institute of Plant Sciences Paris-Saclay (IPS2), 91190, Gif sur Yvette, France.
[2]Université de Paris Cité, Institute of Plant Sciences Paris-Saclay (IPS2), 91190, Gif sur Yvette, France.
[3]Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE-Le Moulon, 91190, Gif-Sur-Yvette, France.
[4]Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA Paris-Saclay, 91120, Palaiseau, France.

*Corresponding author(s). E-mail(s): marie-laure.martin@inrae.fr; melisande.blein-nicolas@inrae.fr;
Contributing authors: yacine.djabali@inrae.fr; renaud.rincent@inrae.fr;

Abstract

Key message **Multi-trial genome wide association study of plasticity indices allow to detect QTLs specifically involved in the genotype x water availability interaction.**

Abstract Concerns regarding high maize yield losses due to increasing occurrences of drought events are growing, and breeders are still looking for molecular markers for drought tolerance. However, the genetic determinism of traits in response to drought is highly complex and identification of causal regions is a tremendous task. Here, we exploit the phenotypic data obtained from four trials carried out on a phenotyping platform, where a diversity panel of 254 maize hybrids was grown under well-watered and water deficit conditions, to investigate the genetic bases of the drought response in maize. To dissociate drought effect from other environmental factors, we performed multi-trial genome-wide association study on well-watered and water deficit phenotypic means, and on phenotypic plasticity indices computed from measurements made for six ecophysiological traits. We identify 102 QTLs and 40 plasticity QTLs. Most of them were new compared to those obtained from a previous study on the same dataset. Our results show that plasticity QTLs cover genetic regions not identified by QTLs. Furthermore, for all ecophysiological traits, except one, plasticity QTLs are specifically involved in the genotype by water availability interaction, for which they explain between 60% and 100% of the variance. Altogether, QTLs and plasticity QTLs captured more than 75% of the genotype by water availability interaction variance, and allowed to find new genetic regions. Overall, our results demonstrate the importance of considering phenotypic plasticity to decipher the genetic architecture of trait response to stress.

Keywords: drought response, phenotypic plasticity, multi-trial GWAS, QTL, maize

Introduction

Maize, currently the leading cereal crop ahead of rice and wheat (FAOSTAT, 2022), is massively produced, traded and exported worldwide (Wu and Guclu, 2013; Erenstein et al, 2022). The major maize-producing countries are the United States, China and Brazil, with production exceeding seven hundred million tons per year since 2019 (FAOSTAT, 2022). This success is due not only to advances in agronomic practices and breeding, which have improved maize agronomic performance (Mazur et al, 1999; Balconi et al, 2007; Kelliher et al, 2019; Simmons et al, 2021), but also to the capacity of maize to adapt to a wide range of environments (Lanari, 1979; Rotili et al, 2019). Besides being the most widely produced crop in the world, maize also has a strong social and economic impact. As an easy and cheap source of calories and micro-nutrients, maize has become a staple food for many people, especially in Sub-Saharan countries. In developed countries, maize is also widely used in the starch industry and

-

in the production of livestock products (Shiferaw et al, 2011; Ranum et al, 2014; Ekpa et al, 2018).

One of the principal threats to maize production is drought (Zipper et al, 2016; Daryanto et al, 2016; Song et al, 2020). Despite having a C4 metabolism, which ensures good water use efficiency (Crafts-Brandner and Salvucci, 2002), maize can be severely affected by water deficits (Salehi-Lisar and Bakhshayeshan-Agdam, 2016). The defense mechanisms that decrease water losses also reduce plant growth (Tardieu et al, 2017). For instance, the water loss / CO_2 absorption trade-off associated with stomata closure leads to a decrease in photosynthetic activity and biomass production, indirectly affecting grain yield (Efeoğlu et al, 2009; Wang et al, 2019; Song et al, 2019). Water deficit can also directly induce severe yield loss if it occurs during flowering and prevents silk development, an essential step for grain production (Sah et al, 2020). With climate change, drought scenarios are expected to occur more frequently in maize-producing regions (Seager et al, 2014; Cook et al, 2014; Gudmundsson and Seneviratne, 2016). Together with the increase in human population, this is a major concern for global food security (Harrison et al, 2014; Lobell et al, 2014; Meng et al, 2016). Drought tolerance is a highly integrated trait resulting from the combination of many genetically variable traits, such as water uptake, leaf growth and transpiration rate (Tardieu et al, 2014). Consequently, developing drought-tolerant maize varieties through breeding programs is a solution of choice to mitigate yield losses (Campos et al, 2004, 2006; Cooper et al, 2014). Aided by advances in genomics, genome-wide association studies (GWAS) are one of the most popular and powerful approaches for identifying genetic polymorphisms associated with inter-individual variations of traits of interest (Zhang et al, 2020; Zhao et al, 2022). These so-called quantitative trait loci (QTLs) are further used in breeding programs. However, detecting QTLs related to the genotype x water availability interaction is complex since drought tolerance strongly depends on the environmental conditions faced by the plants. When only a single trial is conducted to study drought response, it is impossible to dissociate the effect of water conditions from the effect of other environmental factors that may fluctuate between trials. To tackle these confounding effects, it is necessary to carry out multi-environment trials (METs) (Boer et al, 2007; Rodrigues, 2018).

METs are experiments where a trait of interest is measured in several trials. For example, in Millet et al (2016), yield was measured in 29 different fields representing multiple trials with contrasting conditions. More complex METs can include paired conditions in each trial. For example, in Alvarez Prado et al (2017), stomatal conductance in maize was measured in four trials, each with two different watering conditions (well-watered and water deficit). To detect QTLs of interest from METs, GWAS can be carried out for each trial separately. Then, among all identified QTLs, some are selected for their contribution to the genetic effects using a modeling approach (Diouf et al, 2020) or because they are present in a specific set of trials (Millet et al, 2016; Alvarez Prado et al, 2017; Touzy et al, 2019; Hu et al, 2021). Another approach in METs with paired conditions is based on phenotypic plasticity, *i.e.* the variation in phenotype for a given genotype in response to different conditions (Bradshaw, 1965). Plasticity indices can be computed for each trait with regression models as proposed by Finlay and Wilkinson (1963) or by computing the relative difference or ratio between two studied conditions (Peleg et al, 2009; Zhai et al, 2014; dos Santos Silva et al, 2021). Performing GWAS on plasticity indices then allows the detection of condition-responsive QTLs, hereafter called plasticity QTLs (Wang et al, 2013; Zhai et al, 2014; Ye et al, 2019). Recent studies in rice and wheat have shown that plasticity QTLs tend to be positioned near stress-responsive genes (Mai et al, 2021; Fatiukha et al, 2021) and that many of them do not overlap with the QTLs detected separately in each studied condition (Kusmec et al, 2017; Diouf et al, 2020).

In this study, we investigated the genetic bases of the drought response in maize using a MET approach. To this end, we analyzed previously published phenotypic data acquired for six drought-related ecophysiological traits and 254 maize genotypes grown under two watering conditions repeated in four independent trials (Alvarez Prado et al, 2017). Because the four trials were conducted over three years and two seasons, high variations in vapor pressure deficit and light were observed (Alvarez Prado et al, 2017) and were considered as trial effects. We first assessed the effect of the trial on QTL detection by performing multi-trial GWAS on well-watered and water deficit phenotypic means. Then, considering plasticity indices as different traits, as postulated by Bradshaw (1965), we investigated the relative contribution of QTLs and plasticity QTLs to the genotype by water availability interaction for the six studied ecophysiological traits.

Methods

Description of the phenotypic and genomic data used

The phenotypic dataset used in this study was previously published by Alvarez Prado et al (2017). This dataset consisted of six ecophysiological traits, namely biomass (Biom), leaf area (LA), transpiration rate (Transp), stomatal conductance (gs_max), water uptake (WU) and water use efficiency (WUE), measured on a diversity panel of maize hybrids obtained by crossing 254 dent lines selected for their restricted flowering window with a standard flint line (UH007). Three replicates of each hybrid were grown under two watering conditions (well-watered, WW, and water deficit, WD). This experimental design was replicated in four different trials during three different years and two different seasons: spring 2012, spring 2013, spring 2016 and winter 2013. Plants were grown at the INRAE PhenoArch phenotyping platform located in France at Montpellier (Cabrera-Bosquet et al, 2016; Alvarez Prado et al, 2017), with applied soil water potentials equal to -0.05 MPa for the WW condition and ranging from -0.3 to -0.6 MPa for the WD condition. For each trait, watering condition and trial, the average of the three replicates (hereafter called the phenotypic mean) was adjusted by taking into account the greenhouse spatial effect , as described in Alvarez Prado et al (2017).

The genomic dataset contained 977,459 SNPs obtained using a combination of a 50K Infinium HD Illumina array (Ganal et al, 2011), a 600K Axiom Affymetrix array (Unterseer et al, 2014) and 500K markers obtained by genotyping by sequencing (Negro et al, 2019). SNPs with a minor allele frequency (MAF) below 0.05 or a heterozygosity rate above 0.15 were filtered. Missing values were imputed by Beagles 3.1 (Browning and Browning, 2007). SNPs were mapped on the Zm00001d.2 gene models annotation of the B73 reference assembly (ZmB73_RefGen_v4) of the maize genome obtained from MaizeGDB (https://www.maizegdb.org/assembly#downloads)

Calculation of plasticity indices

Plasticity indices were calculated as ecophysiological trait-related stability indices (Bouslama and Schapaugh Jr., 1984). They are defined as ratios between the phenotypic mean of a trait for a genotype under water deficit to the phenotypic mean of the same trait for the same genotype under the well-watered condition. These plasticity indices were computed for each ecophysiological trait in each trial.

Multi-trial GWAS

Multi-trial GWAS was performed by adding a fixed effect of the trial in the single locus mixed model of Yu et al (2006) :

$$Y_{gt} = \mu + T_t + \alpha.X_g + \underline{G_g} + \underline{\varepsilon_{gt}} \quad (1)$$

where Y_{gt} is a plasticity index or a phenotypic mean under the WW or the WD condition of genotype g in the trial t; μ is the overall mean; T_t is the fixed effect of trial t; α is the fixed effect of the SNP allelic dose X_g (coded as 0,1 and 2) for the genotype g; $G_g \sim \mathcal{N}(0, \sigma^2_g.K)$ is the random effect of genotype g, with K the kinship matrix; $\varepsilon_{gt} \sim \mathcal{N}(0, \sigma^2.I_n)$ is the residual error. The kinship matrix K was computed with the whole set of SNPs except those located on the same chromosome as the tested SNP (Rincent et al, 2014), following the approach published by Astle and Balding (2009) and implemented in the R package *statgenGWAS*. Model (1) was run using the function *GWAS* of the R package *rrBLUP* (Endelman, 2011).

SNPs were considered to be significantly associated if their p-values were below 10^{-5}. SNPs less than 0.1 cM apart were clumped into QTLs identified by the ecophysiological trait to which they are related and the most significant SNP.

QTLs associated with plasticity indices were defined as plasticity QTLs while those associated with phenotypic means each of the two watering conditions were defined as QTLs.

QTL colocalization is defined by the overlap of the linkage disequilibrium (LD) windows of QTLs as described in Negro et al (2019). Genes associated with QTLs, *i.e.* genes located in QTL LD windows, were retrieved from the Zm00001d.2 gene models annotation .gff3 of the B73 reference assembly (https://www.maizegdb.org/assembly#downloads)

-

Estimation of the relevance of QTLs and plasticity QTLs

To assess the biological relevance of the detected QTLs and plasticity QTLs, we followed the approach previously described by van Eeuwijk et al (2010). A multi-environment mixed model with random effects for genotype (G), genotype by water availability interaction ($G \times W$), genotype by trial ($G \times T$) was first fitted following Alvarez Prado et al (2017) in order to estimate the variance components of random effects by the restricted maximum likelihood (REML):

$$Y_{gwt} = \mu + E_{wt} + PC_g + (PC \times E)_{gwt} + \underline{G_g} + \underline{(G \times W)_{gw}} + \underline{(G \times T)_{gt}} + \underline{\varepsilon_{gwt}} \quad (2)$$

where : Y_{gwt} is the phenotypic mean of genotype g in the watering condition w and the trial t; μ is the overall mean; E_{wt} is the fixed effect of the environment defined as the combination between the watering condition w and the trial t; PC_g are coordinates of genotype g projected onto principal component analysis axes built with the kinship matrix K. The number of axes used was chosen following the Kaiser criterion; $(PC \times E)_{gwt}$ are the fixed interaction effects between the genetic structure PC_g and the environment defined as the combination between the watering condition w and the trial t; $\underline{\varepsilon_{gwt}}$ is the residual of the model.

For each ecophysiological trait, the significance of $\underline{G}$, $\underline{G \times W}$, $\underline{G \times T}$ random effects were tested by comparing the model defined Equation (2) and the same model without the tested random effect. These random effects were considered to be significant if their p-values were below 0.05.

Then, a multi-locus multi-environment mixed model (van Eeuwijk et al, 2010; Alvarez Prado et al, 2017) was fitted by adding the fixed effects of the QTLs and QTLs by environment interaction in (2):

$$Y_{gwt} = \mu + E_{wt} + PC_g + PCQ_g + (PC \times E)_{gwt} + (PCQ \times E)_{gwt} + \underline{G_g} + \underline{(G \times W)_{gw}} + \underline{(G \times T)_{gt}} + \underline{\varepsilon_{gwt}} \quad (3)$$

where PCQ_g is the fixed effect of a given set of QTLs or plasticity QTLs. PCQ_g are coordinates of the genotype g projected onto principal component analysis axes built with a kinship matrix computed with a set of significant SNPs that describe the QTLs or plasticity QTLs. The number of axes used was chosen following the Kaiser criterion.

Let r be one of the three random effects : G, GxT, and GxW. The proportion of variance γ_{qr} explained by a given set of SNPs q for the random effect r is defined by:

$$\gamma_{qr} = \frac{\Gamma_r - \Gamma_r^*}{\Gamma_r} \quad (4)$$

where: Γ_r is the variance component of the random effect r in (2) and Γ_r^* is the variance component of the random effect r in (3).

Results

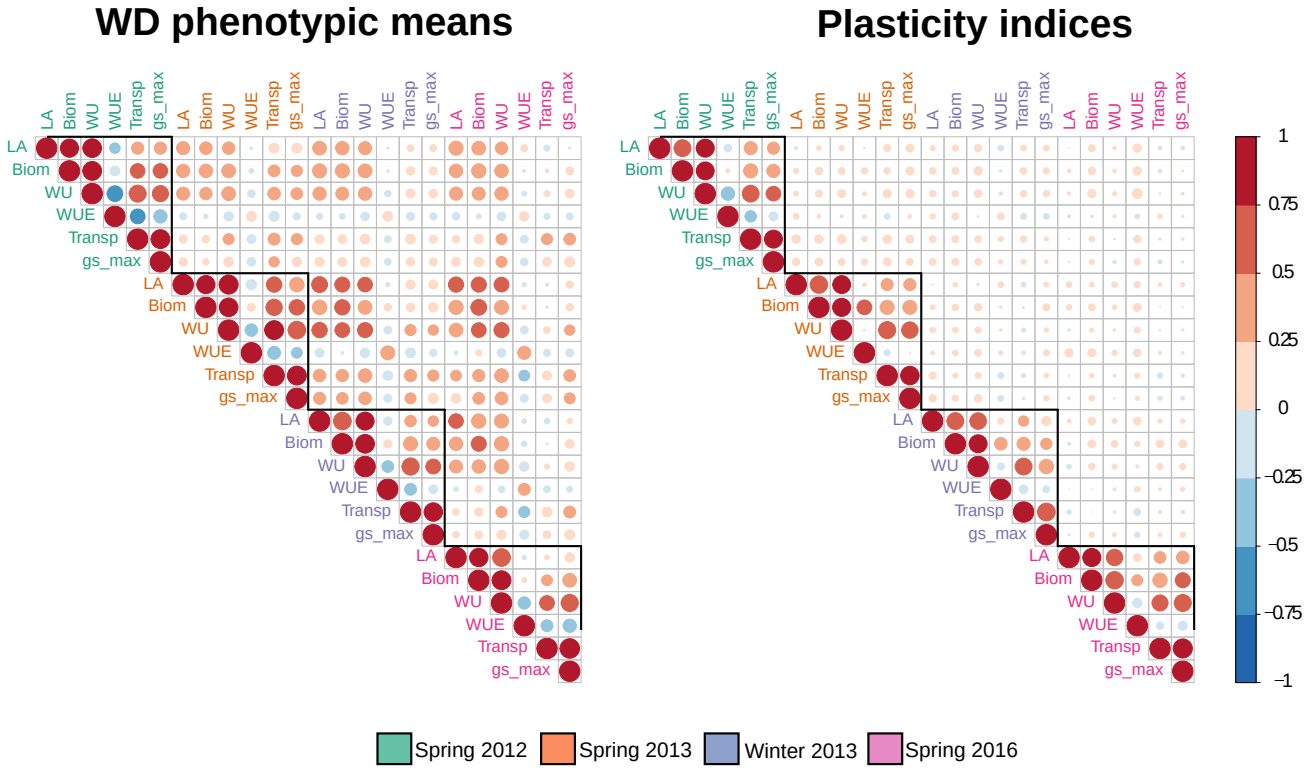
GxE interactions are driven more by trial effects than by water availability

To evaluate the trial effects, we calculated the Pearson correlation coefficients between the four trials for each phenotypic mean and plasticity index (Fig. 1). This showed that the phenotypic means or plasticity indices computed for a same ecophysiological trait in two different trials are more distant from each other than the phenotypic means or plasticity indices computed for different ecophysiological traits in the same trial. This is even more apparent for plasticity indices (Fig. 1B) than for phenotypic means (Fig. 1A and Fig. S1). This result highlights the importance of the effects of the trials on the phenotypes and response phenotypes of drought-related ecophysiological traits.

Table 1: Variance components of each random effect calculated from (2) and their contribution in percentage to the total ecophysiological trait variance. The p-values in brackets indicate the significance of the associated effect.

Trait	G		$G \times W$		$G \times T$		ε	
	Γ_G	%	$\Gamma_{G\times W}$	%	$\Gamma_{G\times T}$	%	Γ_ε	%
LA	$5e^{-4}$ ($<1e^{-16}$)	44	$2{,}7e^{-5}$ (0,051)	3	$1e^{-4}$ ($6{,}7e^{-7}$)	9	$5e^{-4}$	44
Biom	484 ($<1e^{-16}$)	40	104,2 ($3e^{-8}$)	9	31,9 (0,15)	3	582	48
Transp	0,23 ($<1e^{-16}$)	28	0,007 (0,6)	1	0,05 (0,005)	6	0,54	65
gs_max	9,02 ($<1e^{-16}$)	22	0,61 (0,4)	1	2,8 (0,015)	7	28,9	70
WU	0,08 ($<1e^{-16}$)	39	0,009 (0,001)	4	0,017 ($5{,}5e^{-5}$)	8	0,1	49
WUE	6,5 ($1e^{-13}$)	19	1,66 (0,007)	5	4,96 ($2{,}6e^{-7}$)	14	21,6	62

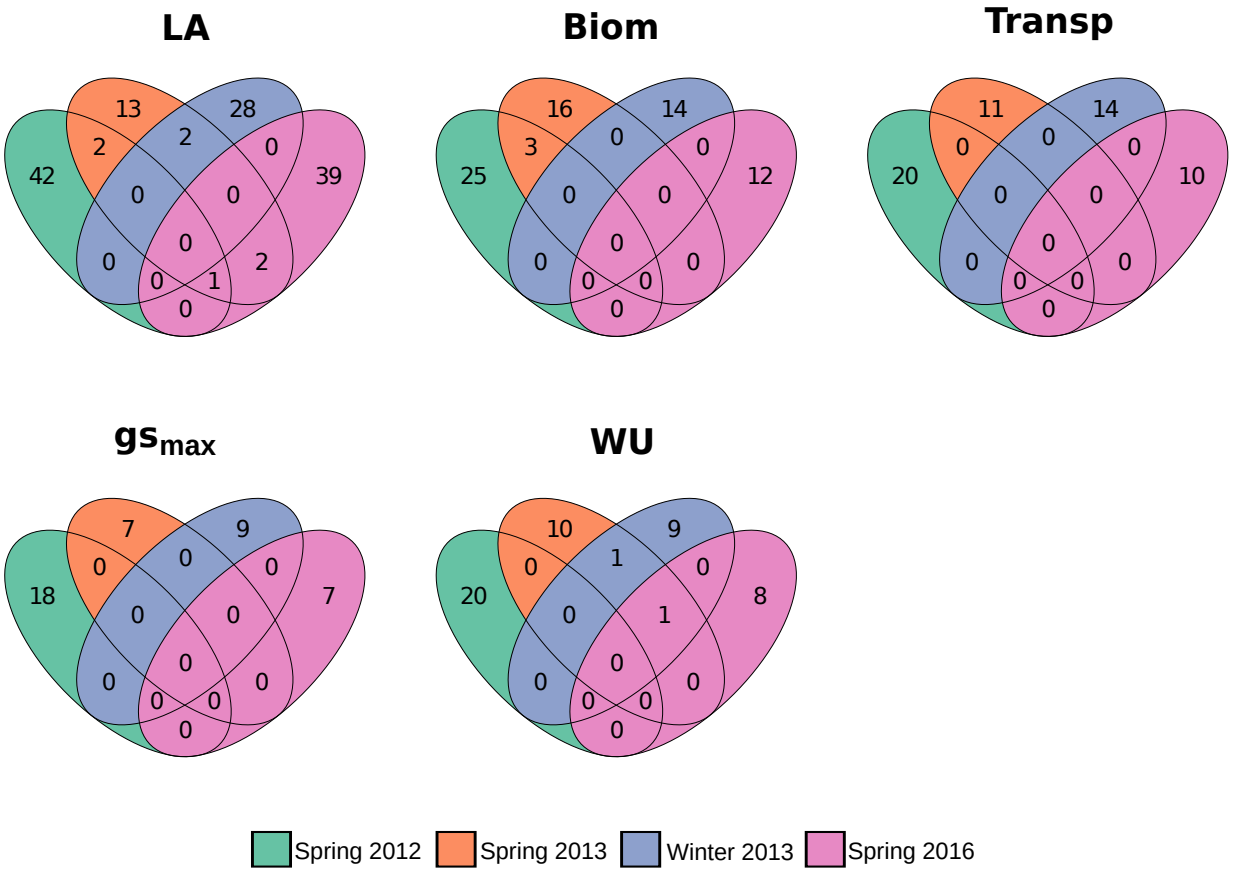
Fig. 1: Correlogram of Pearson's correlations obtained for WD phenotypic means and plasticity indices between each pair of trial x trait combinations. The correlogram obtained for WW phenotypic means is shown in Fig. S1.



To further quantify the contribution of the trials to genetic variance, we computed, for each ecophysiological trait, the significance of the G, GxT and GxW effects and their variance components (Table 1). We observed that G effect was significant for all ecophysiological traits, and it was the most important effect explaining 19-44% of the trait variances. Concerning the GxW effect, the contributions in the trait variance was low (1 - 9%) and this effect was not significant for LA, Transp, and gs_max. The contribution on the GxT effect on the trait variance was 2 to 7 times higher than the variance explained by GxW, and the GxT was significant except for Biom. These results indicate that by conducting single-trial GWAS, the probability of detecting a trial-responsive QTL is higher than that of detecting a water availability-responsive QTL. This result is supported by the fact that there is almost no overlap between the sets of QTLs detected by Alvarez Prado et al (2017) in the four trials (Fig. 2).

-

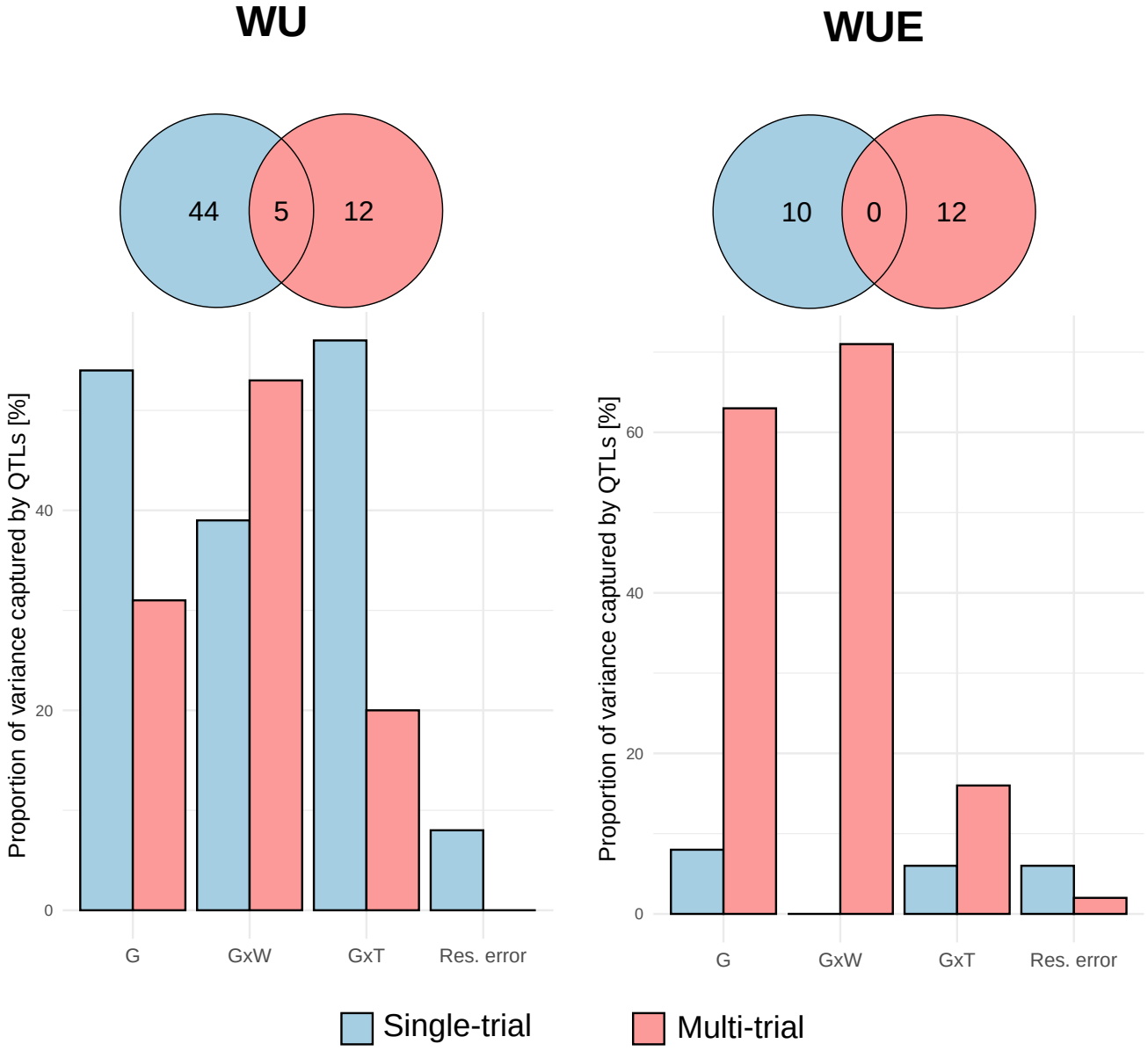
Fig. 2: Venn diagrams showing the overlap between the QTL sets detected by Alvarez Prado et al (2017) in four trials by using single-trial GWAS.



Multi-trial GWAS enhances the detection of water availability-responsive QTLs

To assess the effect of the trials on QTL detection, we performed multi-trial GWAS on phenotypic means. This allowed us to detect 102 QTLs, (60 in the WD condition and 42 in the WW condition) (Table S1), spanning 395 genes in total (Table S4). Compared to the QTLs previously obtained by single-trial GWAS with a p-value threshold of 10^{-5} (Alvarez Prado et al, 2017), QTLs obtained by multi-trial GWAS were less numerous (2.9 to 7.9 times less for all traits except WUE) and were mainly new QTLs (see the Venn diagrams in Fig. 3, Fig. S2). We then compared the contribution of all QTLs detected in the four single-trial GWAS *vs* those detected by multi-trial GWAS to the variations due to G, GxT and GxW (see bar plots in Fig. 3, Fig. S2, Table S2). For three traits, (Biom, gs_max and Transp), the QTLs detected only by multi-trial GWAS contributed less to the variations due to G, GxT and GxW than the QTLs detected only by single-trial GWAS. For gs_max and Transp, this may be explained by the fact that the number of multi-trial QTLs considered was very low compared to that of single-trial GWAS (2 and 5 *vs* 45 and 53, respectively). For the two traits LA and WU, the QTLs detected only by multi-trial GWAS contributed less to the variations due to G and GxT but much more to the variations due to GxW. Finally, for WUE, the QTLs detected only by multi-trial GWAS contributed more to the variations due to G, GxT and GxW. Overall, these results show that for half of the traits considered, multi-trial GWAS allowed to decrease the noise from the trials and detect new QTLs that were more responsive to water availability than to the trials.

Fig. 3: Comparison between QTLs detected by single-trial and multi-trial GWAS for WU and WUE. The Venn diagrams show the overlap between the two QTL sets. The barplots show the proportions of variance of the G, GxT interaction and GxW interaction effects that were captured by a given QTL set, as computed from (2) and (3). Results for the others ecophysiological traits are shown in Fig. S2.



Plasticity QTLs specifically contribute to the GxW interaction

Using multi-trial GWAS, we identified 40 plasticity QTLs (Table S1), none of which overlapped with the 102 QTLs (see the Venn diagrams in Fig. 4 and Fig.S3). These plasticity QTLs are highlighted on the karyoplot shown in Fig. 5, and cover 240 genes (Table S4).

-

Fig. 4: Comparison between QTLs and plasticity QTLs detected from multi-trial GWAS for WU and WUE. The barplots show the proportions of variance of the G, GxT interaction and GxW interaction effects that were captured by a given QTL set, as computed from (2) and (3). For each trait, 1000 control sets were constituted with the QTLs and n SNPs randomly selected among the total number of SNPs available (977,459) in order to control over-fitting. With n the number of plasticity QTLs detected for the trait. Results for the others traits are shown in Fig.S3.

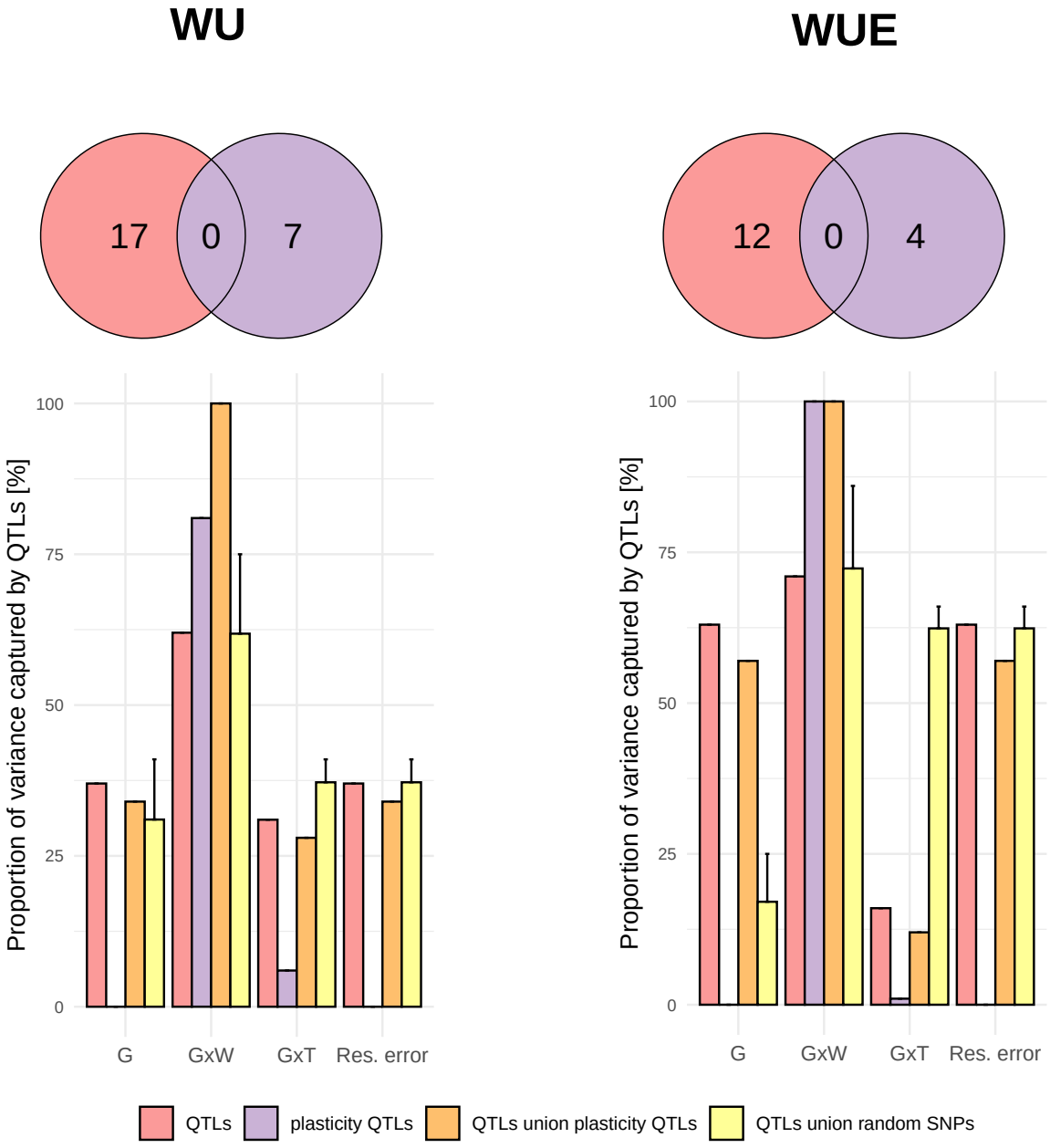
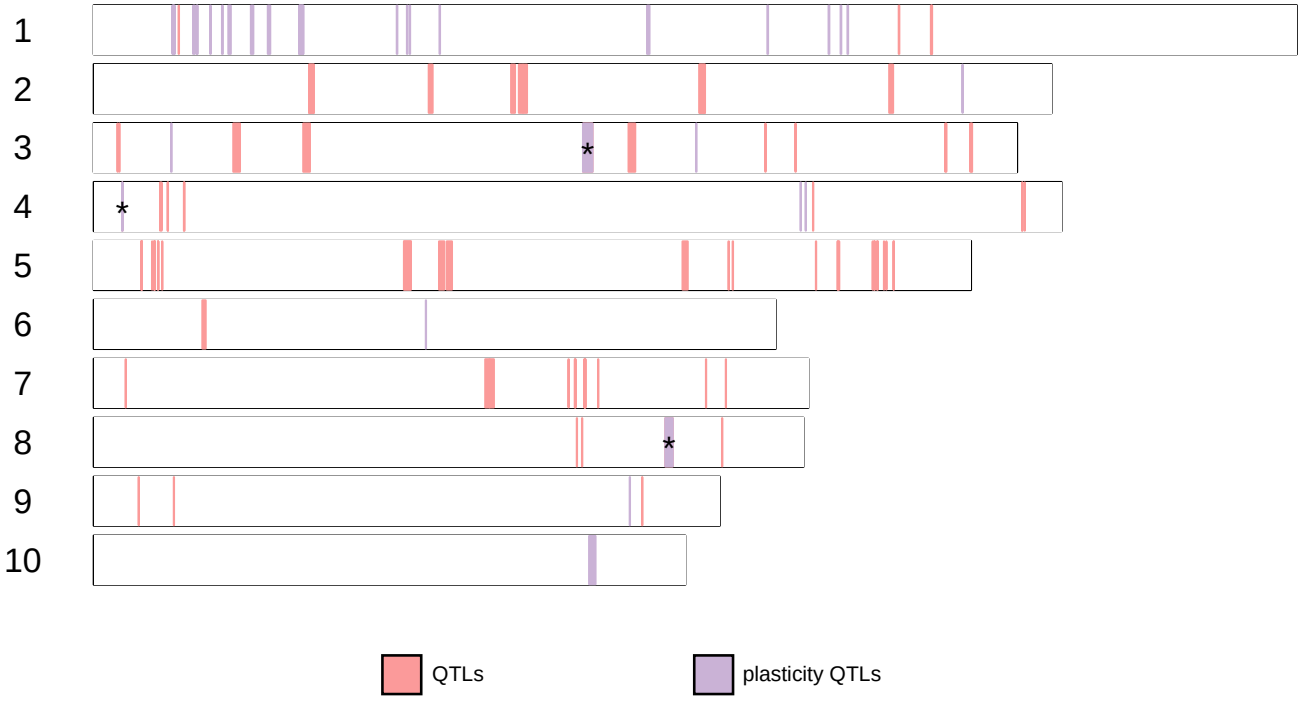


Fig. 5: Karyoplot showing the results of the multi-trial GWAS performed on phenotypic means and plasticity indices. The QTLs identified on phenotypic means are depicted in red and those identified on plasticity indices in purple. Asterix correspond to a colocalization between QTLs and plasticity QTLs.



Plasticity QTLs were 2 to 7 times less numerous than QTLs and contributed weakly to the variance due to G and GxT effects (Fig. 4, Fig.S3 and Table S3). However, excepted for gs_max, plasticity QTLs strongly contributed to the variance due to GxW effects (60 - 100%). QTLs mainly contributed to the variance due to GxW effects in four traits (LA, gs_max, WU and WUE), as well as to the effects of G and GxT. Plasticity QTLs contributed more to the GxW effect than QTLs. To evaluate to what extent plasticity QTLs were complementary to QTLs, we considered jointly plasticity QTLs and QTLs and compared their biological relevance to a control set comprising QTLs and randomly selected SNPs (Fig. 4 and Fig.S3). Excepted for gs_max, QTLs and plasticity QTLs jointly contributed more to the GxW effect than the union of the QTLs and randomly selected SNPs. Altogether, these results show that plasticity QTLs are good candidates for understanding the genetic architecture of the GxW interaction.

Discussion

Plants being sessile, they are continuously exposed to variable and potentially harsh environmental conditions that can cause biotic or abiotic stress. In crops, the ability to resist or tolerate such stresses is of great importance to maintain productivity without resorting to inputs. Here, we investigated the genetic bases of the drought response in maize using a MET approach. The objectives of our study were i) to assess the effect of the trials on QTL detection, and ii) to estimate the extent to which plasticity QTLs contribute to the GxW effect on drought-related ecophysiological traits in maize compared to QTLs.

Even though the trials were carried out in a greenhouse with well-controlled watering conditions, the meteorological conditions outside the greenhouse were different from one trial to the other. We took these different meteorological conditions into account, which allowed us to decompose the GxE interaction into GxW and GxT. We showed that the GxE interactions observed in the data were driven more by trial effects than by water availability. This may explain the low overlap between the sets of QTLs identified from each trial by Alvarez Prado et al (2017).

Compared to single-trial GWAS, multi-trial GWAS allowed to better fit the ecophysiological traits related to drought response: residual errors and GxT interactions were both smaller in multi-trial GWAS than in single-trial GWAS. In addition, the newly detected QTLs captured a larger part of the GxW variability for LA, WU and WUE. For the three other traits (Biom, Transp and gs_max), results are more mitigated: the number of newly detected QTLs was small compared to the several dozens of QTLs detected by the single-trial GWAS, which already explained a large part of the GxW interaction. Overall, by performing multi-trial GWAS, we multiplied

-

individual observations and thereby increased the power to detect QTLs across trials (Cantor et al, 2010; Thomas, 2010). These results are consistent with those of Benaouda et al (2022), who showed that four multi-environment QTLs explained 20.6% of the heading time variance in wheat compared to 9.5% for the six single-environment QTLs detected by Langer et al (2014).

In this study, we also explored the gain provided by performing GWAS on six plasticity indices for dissecting the genetic architecture of ecophysiological trait response to drought. We identified 40 plasticity QTLs and highlighted 38 genetic regions that differed from those associated with the QTLs. The results obtained with the multi-locus multi-environment model showed that plasticity QTLs specifically captured the variance of the GxW interaction. By comparison, QTLs not only captured the variance of the GxW interaction, but also a large part of the G variance and, to a lesser extent, part of the variance of the GxT interaction. These results indicate that, for the ecophysiological traits studied, the genetic control of phenotypic plasticity in response to drought does not completely overlap with that of the genetic control of the phenotypic mean. Similar results were previously observed in maize (Kusmec et al, 2017), tomato (Diouf et al, 2020) and cassava (dos Santos Silva et al, 2021).

All together, our results help dissecting the genetic basis of response to water deficit. Indeed, three mutually non exclusive genetic models have been proposed to explain phenotypic plasticity (Scheiner, 1993; Via et al, 1995). First, the over-dominance model assumes that phenotypic plasticity is related to the number of heterozygous loci (Gillespie and Turelli, 1989). Second, the allelic-sensitivity model considers that the environment affects the allelic effect of the genetic factors that determine a trait. Third, the gene-regulatory model assumes that phenotypic plasticity results from epistatic interactions between structural and regulatory alleles. Our results favor this last model, in agreement with the hypothesis of Bradshaw (1965) stating that the plasticity of a trait is an independent property of that trait and is under its own specific genetic control.

In conclusion, considering phenotypic means and plasticity as different traits and taking trial effects into account allow us to gain a more precise understanding of how ecophysiological traits respond to water availability. In the short term, our perspective is to go deeper into the functional annotation of the genes associated with the QTLs and plasticity QTLs. By comparing the two gene lists against gene regulatory databases, a strong result in favor of the gene-regulatory model would be to find that genes associated with the plasticity QTLs are regulators of genes associated with the QTLs. These findings may shed further light on the genetic regulatory system underlying the response of plants to stress.

Acknowledgments. We thank Claude Welcker for the time he dedicated to helpful discussions about this work. We thank Santiago Alvarez Prado for giving us further information on the multi-environment multi-locus mixed models.

Fundings. This work and IPS2 have benefited from the support of the LabEx Saclay Plant Sciences-SPS (reference n° ANR-17-EUR-0007, EUR SPS-GSR). This work benefits also from two French State grants from INRAE and C-Land (reference n° ANR-16-CONV-0003) managed by the French National Research Agency under the "Investissements d'avenir" program integrated into France 2030 (reference n° ANR-11-IDEX-0003-02). The data analyzed in this work were produced during the Amaizing project (reference n° ANR-10-BTBR-01) managed by the French National Research Agency under the "Investissements d'avenir".

Declarations

Competing interests. The authors have no relevant financial or non-financial interests to disclose.

Consent to participate. Not applicable

Consent for publication. Not applicable

Data availability. The genotyping datasets analyzed during the current study are published in Negro et al (2019) The phenomics dataset analyzed during the current study is published in Alvarez Prado et al (2017)

Code availability. https://forgemia.inra.fr/gqe-base/djabali-drought-plasticity-qtls-article1

Authors' contributions. MBN and MLM defined the research project. YD carried out the analyses. RR, MBN, and MLM supervised the work. YD wrote the manuscript and MNB, RR, and MLM revised and approved it.

References

Alvarez Prado S, Cabrera-Bosquet L, Grau A, et al (2017) 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

Astle W, Balding DJ (2009) Population Structure and Cryptic Relatedness in Genetic Association Studies. Statistical Science 24(4):451–471. https://doi.org/10.1214/09-STS307

Balconi C, Hartings H, Lauria M, et al (2007) Gene discovery to improve maize grain quality traits. Maydica 52:357–373

Benaouda S, Dadshani S, Koua P, et al (2022) Identification of QTLs for wheat heading time across multiple-environments. Theoretical and Applied Genetics 135(8):2833–2848. https://doi.org/10.1007/s00122-022-04152-6

Boer MP, Wright D, Feng L, et al (2007) A Mixed-Model Quantitative Trait Loci (QTL) Analysis for Multiple-Environment Trial Data Using Environmental Covariables for QTL-by-Environment Interactions, With an Example in Maize. Genetics 177(3):1801–1813. https://doi.org/10.1534/genetics.107.071068

Bouslama M, Schapaugh Jr. WT (1984) Stress Tolerance in Soybeans. I. Evaluation of Three Screening Techniques for Heat and Drought Tolerance1. Crop Science 24(5):cropsci1984.0011,183X002400050,026x. https://doi.org/10.2135/cropsci1984.0011183X002400050026x

Bradshaw AD (1965) Evolutionary Significance of Phenotypic Plasticity in Plants. In: Caspari EW, Thoday JM (eds) Advances in Genetics, vol 13. Academic Press, p 115–155, https://doi.org/10.1016/S0065-2660(08)60048-6

Browning SR, Browning BL (2007) Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. The American Journal of Human Genetics 81(5):1084–1097. https://doi.org/10.1086/521987

Cabrera-Bosquet L, Fournier C, Brichet N, et al (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

Campos H, Cooper M, Habben JE, et al (2004) Improving drought tolerance in maize: A view from industry - ScienceDirect. Field Crops Research 90:19–34. https://doi.org/10.1016/j.fcr.2004.07.003

Campos H, Cooper M, Edmeades GO, et al (2006) Changes in drought tolerance in maize associated with fifty years of breeding for yield in the US Corn Belt. Maydica 51:369–381

Cantor RM, Lange K, Sinsheimer JS (2010) Prioritizing GWAS Results: A Review of Statistical Methods and Recommendations for Their Application. American Journal of Human Genetics 86(1):6–22. https://doi.org/10.1016/j.ajhg.2009.11.017

Cook BI, Smerdon JE, Seager R, et al (2014) Global warming and 21st century drying. Climate Dynamics 43(9-10):2607–2627. https://doi.org/10.1007/s00382-014-2075-y

Cooper M, Gho C, Leafgren R, et al (2014) Breeding drought-tolerant maize hybrids for the US corn-belt: Discovery to product. Journal of Experimental Botany 65(21):6191–6204. https://doi.org/10.1093/jxb/eru064

Crafts-Brandner SJ, Salvucci ME (2002) Sensitivity of Photosynthesis in a C4 Plant, Maize, to Heat Stress. Plant Physiology 129(4):1773–1780. https://doi.org/10.1104/pp.002170

Daryanto S, Wang L, Jacinthe PA (2016) Global Synthesis of Drought Effects on Maize and Wheat Production. PLoS ONE 11(5). https://doi.org/10.1371/journal.pone.0156362

Diouf I, Derivot L, Koussevitzky S, et al (2020) Genetic basis of phenotypic plasticity and genotype × environment interactions in a multi-parental tomato population. Journal of Experimental Botany 71(18):5365–5376. https://doi.org/10.1093/jxb/eraa265

dos Santos Silva PP, Sousa MBe, de Oliveira EJ, et al (2021) Genome-wide association study of drought tolerance in cassava. Euphytica 217(4):60. https://doi.org/10.1007/s10681-021-02800-4

-

Efeoğlu B, Ekmekçi Y, Çiçek N (2009) Physiological responses of three maize cultivars to drought stress and recovery. South African Journal of Botany 75(1):34–42. https://doi.org/10.1016/j.sajb.2008.06.005

Ekpa O, Palacios-Rojas N, Kruseman G, et al (2018) Sub-Saharan African maize-based foods: Technological perspectives to increase the food and nutrition security impacts of maize breeding programmes. Global Food Security-Agriculture Policy Economics and Environment 17:48–56. https://doi.org/10.1016/j.gfs.2018.03.007

Endelman JB (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. The Plant Genome 4(3). https://doi.org/10.3835/plantgenome2011.08.0024

Erenstein O, Jaleta M, Sonder K, et al (2022) Global maize production, consumption and trade: Trends and R&D implications. Food Security https://doi.org/10.1007/s12571-022-01288-7

Fatiukha A, Deblieck M, Klymiuk V, et al (2021) Genomic Architecture of Phenotypic Plasticity in Response to Water Stress in Tetraploid Wheat. International Journal of Molecular Sciences 22(4):1723. https://doi.org/10.3390/ijms22041723

Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant-breeding programme. Australian Journal of Agricultural Research 14(6):742–754. https://doi.org/10.1071/ar9630742

Ganal MW, Durstewitz G, Polley A, et al (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(12):e28,334. https://doi.org/10.1371/journal.pone.0028334

Gillespie JH, Turelli M (1989) Genotype-environment interactions and the maintenance of polygenic variation. Genetics 121(1):129–138. https://doi.org/10.1093/genetics/121.1.129

Gudmundsson L, Seneviratne SI (2016) Anthropogenic climate change affects meteorological drought risk in Europe. Environmental Research Letters 11(4):044,005. https://doi.org/10.1088/1748-9326/11/4/044005

Harrison MT, Tardieu F, Dong Z, et al (2014) Characterizing drought stress and trait influence on maize yield under current and future conditions. Global Change Biology 20(3):867–878. https://doi.org/10.1111/gcb.12381

Hu X, Wang G, Du X, et al (2021) QTL analysis across multiple environments reveals promising chromosome regions associated with yield-related traits in maize under drought conditions. The Crop Journal 9(4):759–766. https://doi.org/10.1016/j.cj.2020.10.004

Kelliher T, Starr D, Su X, et al (2019) One-step genome editing of elite crop germplasm during haploid induction. Nature Biotechnology 37(3):287–292. https://doi.org/10.1038/s41587-019-0038-x

Kusmec A, Srinivasan S, Nettleton D, et al (2017) Distinct genetic architectures for phenotype means and plasticities in Zea mays. Nature Plants 3(9):715–723. https://doi.org/10.1038/s41477-017-0007-7

Lanari D (1979) Expansion of the Area of the Maize Crop. In: Bowman JC, Susmel P (eds) The Future of Beef Production in the European Community. Current Topics in Veterinary Medicine and Animal Science, Springer Netherlands, Dordrecht, p 360–379, https://doi.org/10.1007/978-94-009-9329-7_24

Langer SM, Longin CFH, Würschum T (2014) Flowering time control in European winter wheat. Frontiers in Plant Science 5. https://doi.org/10.3389/fpls.2014.00537

Lobell DB, Roberts MJ, Schlenker W, et al (2014) Greater Sensitivity to Drought Accompanies Maize Yield Increase in the U.S. Midwest. Science 344(6183):516–519. https://doi.org/10.1126/science.1251423

Mai NTP, Mai CD, Nguyen HV, et al (2021) Discovery of new genetic determinants of morphological plasticity in rice roots and shoots under phosphate starvation using GWAS. Journal of Plant Physiology 257:153,340. https://doi.org/10.1016/j.jplph.2020.153340

Mazur B, Krebbers E, Tingey S (1999) Gene Discovery and Product Development for Grain Quality Traits. Science 285(5426):372–375. https://doi.org/10.1126/science.285.5426.372

Meng Q, Chen X, Lobell DB, et al (2016) Growing sensitivity of maize to water scarcity under climate change. Scientific Reports 6:19,605. https://doi.org/10.1038/srep19605

Millet EJ, Welcker C, Kruijer W, et al (2016) Genome-Wide Analysis of Yield in Europe: Allelic Effects Vary with Drought and Heat Scenarios. Plant Physiology 172(2):749–764. https://doi.org/10.1104/pp.16.00621

Negro SS, Millet EJ, Madur D, et al (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

Peleg Z, Fahima T, Krugman T, et al (2009) Genomic dissection of drought resistance in durum wheat × wild emmer wheat recombinant inbreed line population. Plant, Cell & Environment 32(7):758–779. https://doi.org/10.1111/j.1365-3040.2009.01956.x

Ranum P, Peña-Rosas JP, Garcia-Casal MN (2014) Global maize production, utilization, and consumption. Annals of the New York Academy of Sciences 1312:105–112. https://doi.org/10.1111/nyas.12396

Rincent R, Moreau L, Monod H, et al (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

Rodrigues PC (2018) An overview of statistical methods to detect and understand genotype-by-environment interaction and QTL-by-environment interaction. Biometrical Letters 55(2):123–138. https://doi.org/10.2478/bile-2018-0009

Rotili DH, Giorno A, Tognetti PM, et al (2019) Expansion of maize production in a semi-arid region of Argentina: Climatic and edaphic constraints and their implications on crop management. Agricultural Water Management 226:105,761. https://doi.org/10.1016/j.agwat.2019.105761

Sah RP, Chakraborty M, Prasad K, et al (2020) Impact of water deficit stress in maize: Phenology and yield components. Scientific Reports 10(1):2944. https://doi.org/10.1038/s41598-020-59689-7

Salehi-Lisar SY, Bakhshayeshan-Agdam H (2016) Drought Stress in Plants: Causes, Consequences, and Tolerance. In: Hossain MA, Wani SH, Bhattacharjee S, et al (eds) Drought Stress Tolerance in Plants, Vol 1: Physiology and Biochemistry. Springer International Publishing, Cham, p 1–16, https://doi.org/10.1007/978-3-319-28899-4_1

Scheiner SM (1993) Genetics and Evolution of Phenotypic Plasticity. Annual Review of Ecology and Systematics 24:35–68. https://doi.org/10.1146/annurev.es.24.110193.000343

Seager R, Liu H, Henderson N, et al (2014) Causes of Increasing Aridification of the Mediterranean Region in Response to Rising Greenhouse Gases. Journal of Climate 27(12):4655–4676. https://doi.org/10.1175/JCLI-D-13-00446.1

Shiferaw B, Prasanna BM, Hellin J, et al (2011) Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. Food Security 3(3):307. https://doi.org/10.1007/s12571-011-0140-5

Simmons CR, Lafitte HR, Reimann KS, et al (2021) Successes and insights of an industry biotech program to enhance maize agronomic traits. Plant Science 307:110,899. https://doi.org/10.1016/j.plantsci.2021.110899

Song L, Jin J, He J (2019) Effects of Severe Water Stress on Maize Growth Processes in the Field. Sustainability 11(18):5086. https://doi.org/10.3390/su11185086

Song Y, Linderholm HW, Luo Y, et al (2020) Climatic Causes of Maize Production Loss under Global Warming in Northeast China. Sustainability 12(18):7829. https://doi.org/10.3390/su12187829

Tardieu F, Parent B, Caldeira CF, et al (2014) Genetic and Physiological Controls of Growth under Water Deficit. Plant Physiology 164(4):1628–1635. https://doi.org/10.1104/pp.113.233353

Tardieu F, Varshney RK, Tuberosa R (2017) Improving crop performance under drought – cross-fertilization of disciplines. Journal of Experimental Botany 68(7):1393–1398. https://doi.org/10.1093/jxb/erx042

Thomas D (2010) Gene–environment-wide association studies: Emerging approaches. Nature Reviews Genetics 11(4):259–272. https://doi.org/10.1038/nrg2764

-

Touzy G, Rincent R, Bogard M, et al (2019) Using environmental clustering to identify specific drought tolerance QTLs in bread wheat (T. aestivum L.). Theoretical and Applied Genetics 132(10):2859–2880. https://doi.org/10.1007/s00122-019-03393-2

Unterseer S, Bauer E, Haberer G, et al (2014) A powerful tool for genome analysis in maize: Development and evaluation of the high density 600 k SNP genotyping array. BMC Genomics 15(1):823. https://doi.org/10.1186/1471-2164-15-823

van Eeuwijk FA, Bink MC, Chenu K, et al (2010) Detection and use of QTL for complex traits in multiple environments. Current Opinion in Plant Biology 13(2):193–205. https://doi.org/10/cd33h3

Via S, Gomulkiewicz R, De Jong G, et al (1995) Adaptive phenotypic plasticity: Consensus and controversy. Trends in Ecology & Evolution 10(5):212–217. https://doi.org/10.1016/S0169-5347(00)89061-8

Wang B, Liu C, Zhang D, et al (2019) Effects of maize organ-specific drought stress response on yields from transcriptome analysis. BMC Plant Biology 19:335. https://doi.org/10.1186/s12870-019-1941-5

Wang Z, Pang X, Lv Y, et al (2013) A dynamic framework for quantifying the genetic architecture of phenotypic plasticity. Briefings in Bioinformatics 14(1):82–95. https://doi.org/10.1093/bib/bbs009

Wu F, Guclu H (2013) Global Maize Trade and Food Security: Implications from a Social Network Model: Global Maize Trade and Food Security. Risk Analysis 33(12):2168–2178. https://doi.org/10.1111/risa.12064

Ye M, Jiang L, Chen C, et al (2019) Np2 QTL: Networking phenotypic plasticity quantitative trait loci across heterogeneous environments. The Plant Journal: For Cell and Molecular Biology 99(4):796–806. https://doi.org/10.1111/tpj.14355

Yu J, Pressoir G, Briggs W, et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhai Y, Lv Y, Li X, et al (2014) A synthetic framework for modeling the genetic basis of phenotypic plasticity and its costs. The New Phytologist 201(1):357–365. https://doi.org/10.1111/nph.12458

Zhang H, Zhang J, Xu Q, et al (2020) Identification of candidate tolerance genes to low-temperature during maize germination by GWAS and RNA-seq approaches. BMC Plant Biology 20(1):333. https://doi.org/10.1186/s12870-020-02543-9

Zhao M, Liu S, Pei Y, et al (2022) Identification of genetic loci associated with rough dwarf disease resistance in maize by integrating GWAS and linkage mapping. Plant Science 315:111,100. https://doi.org/10.1016/j.plantsci.2021.111100

Zipper SC, Qiu J, Kucharik CJ (2016) Drought effects on US maize and soybean production: Spatiotemporal patterns and historical changes. Environmental Research Letters 11(9):094,021. https://doi.org/10.1088/1748-9326/11/9/094021

2.3 . Functional analysis of plasticity quantitative trait loci related to water stress response in maize

Considering phenotypic plasticity as an independent and heritable trait (Bradshaw, 1965) leads to the elaboration of two genetic control assumptions: the allele sensitivity and gene regulatory hypothesis.

The allele sensitivity hypothesis implies that genes underlying phenotypic means and phenotypic plasticity are the same, and as a consequence, the environment directly influences the expression of these genes. In opposition, the gene regulatory hypothesis affirms that genes underlying phenotypic means and phenotypic plasticity are different, and as a consequence, plasticity genes are regulators of phenotypic means genes (Scheiner, 1993; Via et al, 1995).

The results presented in section 2.2 suggest that the plasticity of the ecophysiological traits studied follows the gene regulatory model. Thus, to verify this hypothesis, Amal Ksontini conducted a comparative functional annotation analysis of the genes located in QTLs and plasticity QTLs identified in section 2.2. Her work, which is attached as an annex, dealt with the following two questions:

- What are the biological functions of the genes located in QTLs and plasticity QTLs?
- Can we find any further evidence of the gene regulatory model of plasticity of ecophysiological traits by functionally analyzing QTLs and plasticity QTLs?

Briefly, Amal found no functional enrichment among the genes covered either by QTLs or plasticity QTLs. However, by analyzing the proximal regions of genes covered by QTLs, she was able to find enriched patterns (Bernard et al, 2010) recognized by a class of transcription factors encoded by genes located in plasticity QTLs. By contrast, no pattern enrichment was found in the proximal regions of genes covered by plasticity QTL recognized by a class of transcription factors encoded by genes located in QTLs. These two last results support our intuitions on the regulatory role of genes located in plasticity QTLs.

NB: As Amal's report was written before receiving reviewers' comments on the article presented in Chapter 2, some jargon terms are present in the report :

- steady-state QTLs $\rightarrow$ QTLs detect on phenotypic means
- Multi-environment GWAS $\rightarrow$ Multi-trial GWAS

Bibliography

Alvarez Prado S, Cabrera-Bosquet L, Grau A, et al (2017) 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

Bernard V, Brunaud V, Lecharny A (2010) TC-motifs at the TATA-box expected position in plant genes: A novel class of motifs involved in the transcription regulation. BMC Genomics 11:166. https://doi.org/10.1186/1471-2164-11-166

Bradshaw AD (1965) Evolutionary Significance of Phenotypic Plasticity in Plants. In: Caspari EW, Thoday JM (eds) Advances in Genetics, vol 13. Academic Press, p 115–155, https://doi.org/10.1016/S0065-2660(08)60048-6

Diouf I, Derivot L, Koussevitzky S, et al (2020) Genetic basis of phenotypic plasticity and genotype × environment interactions in a multi-parental tomato population. Journal of Experimental Botany 71(18):5365–5376. https://doi.org/10.1093/jxb/eraa265

Kusmec A, Srinivasan S, Nettleton D, et al (2017) Distinct genetic architectures for phenotype means and plasticities in Zea mays. Nature Plants 3(9):715–723. https://doi.org/10.1038/s41477-017-0007-7

Scheiner SM (1993) Genetics and Evolution of Phenotypic Plasticity. Annual Review of Ecology and Systematics 24:35–68. https://doi.org/10.1146/annurev.es.24.110193.000343, https://arxiv.org/abs/2097172

Via S, Gomulkiewicz R, De Jong G, et al (1995) Adaptive phenotypic plasticity: Consensus and controversy. Trends in Ecology & Evolution 10(5):212–217. https://doi.org/10.1016/S0169-5347(00)89061-8

Chapter 3

Systems genetics provide an in-depth genetic and molecular characterization of maize response to water deficit

3.1 . Standfirst

In the following section, I conducted a systems genetics approach integrating genomics, proteomics, and phenomics data to provide an in-depth genetic characterization of maize response to water deficit. The genomics and phenomics data presented in Chapter 2 were integrated with proteomics data generated by Blein-Nicolas et al (2020). Briefly, proteomics data consisted of 2,055 protein abundances quantified from leaf samples taken on 251 hybrids grown in WW and WD conditions during the spring 2012 trial. Blein-Nicolas et al (2020) conducted single-trial GWAS on each protein abundance obtained in each condition, leading to the identification of more than 22,000 pQTLs. They showed that several of the pQTLs colocalized with the 531 QTLs detected by Alvarez Prado et al (2017) and give further intuition on the uses of proteomics to better characterize the genetic determinism of ecophysiological traits.

Here, based on the previous results and those obtained in Chapter 2, we decided to go further by better estimating an overall genotype x watering availability interaction (GxW) variance by fitting the whole set of phenomics data into a multi-trait, multi-environment linear mixed model. Then, using the QTLs and plasticity QTLs identified in Chapter 2, I investigated which set of QTLs captured the maximum part of GxW variance with multi-locus multi-trait multi-environment linear mixed models. From here, I addressed these questions:

- How the integration of proteomics data, pQTLs, and plasticity (PL) pQTLs with QTLs explaining the highest part of GxW variance can provide a comprehensive molecular insight into to genetic determinism of water deficit response?
- Do the pQTLs that do not colocalize with QTLs capture an additional part of the GxW variance? How can we select pQTLs that maximize the part of the GxW variance captured?
- What are the genetic and molecular bases of maize response to water deficit?

The major result of this study was the inference of a multi-scaled network with an original method combining Gaussian graphical models and multi-locus multi-trait multi-environment linear mixed models. This method allowed a simultaneous selection of proteins linked with ecophysiological traits and pQTLs that captured an additional part of the GxW variance. Thus, we inferred a multi-scaled network capturing 84% of the GxW variance comprising 531 associated loci, 63 proteins, and the 6 ecophysiological traits. Among the 531 associated loci, 48 were located in genomics regions enriched in pQTLs (hotspots). Three hotspots were considered as the most important as a loss of 7 points in the part of the GxW variance was observed if we removed the QTLs and pQTLs that constituted them. Using StringDB, I integrated proteins coded by genes located in these hotspots in protein-protein interaction networks (PPI) involved in the biosynthesis of amino acids and aminoacyl-tRNA, RNA degradation, and oxidative phosphorylation. I also identified a PPI network integrating 29 of the 63 proteins present in the multi-scaled network, highlighting groups of proteins highly connected and involved in response to stress, protein folding, and the oxidation-reduction process.

Together, our results first show the benefit of integrating intermediate molecular traits, such as protein abundances, to unravel the missing heritability of a combination of ecophysiological traits related to

drought response. Second, our results provide an in-depth genetic characterization of water stress response. Third, they give further insight into the involvement of the genetic and molecular factors driving phenotypic changes from WW condition to the WD condition.

In addition to proteomics data, we applied our system genetics approach on metabolomics data produced by the Metabolomics Bordeaux platform (Prigent et al., unpublished). This work was realized by Romain Poupon, an M1 student intern under the supervision of Marie-Laure and me. His work is summarized in section 3.3, and his internship report is attached in the appendix.

3.2 . Integration of phenomics, proteomics, and genomics data into a multi-scale network unravels missing heritability for maize response to water deficit

3.2.1 . Introduction

Advances in high-throughput technologies now make it possible to genotype several hundred individuals with marker densities easily approaching several million single nucleotide polymorphisms. This increase in the amount of genomic data is dramatically improving the study of the genetic determinism of phenotypic traits through genome-wide association studies (GWAS) (Cano-Gamez and Trynka, 2020; Uffelmann et al, 2021). However, despite the success of GWAS in identifying and fine-mapping quantitative trait loci (QTLs), the genetic determinism for complex traits is generally not fully unraveled since a part of the heritability remains unexplained by QTLs. This missing heritability phenomenon, as defined by Manolio et al (2009), is mainly explained by the lack of statistical power (Yang et al, 2010; Shi et al, 2016).

To address this lack of heritability, Boyle et al (2017) proposed a new model of heritability for complex traits, called the omnigenic model, which suggests that most genes influence complex traits through highly connected regulatory networks. The authors distinguish between two types of genes. Core genes, which are limited in number, are thought to have direct effects on complex traits but make a small contribution to total heritability. Peripheral genes do not directly affect complex traits, but because they are much more numerous than core genes, they can explain a large proportion of total heritability. The development of omics technologies, such as transcriptomics, proteomics, or metabolomics, now enables large-scale phenotyping of intermediate molecular traits. This provides a route to identifying QTLs with indirect effects on complex phenotypic traits by establishing relationships between phenotypic traits, intermediate molecular traits, and genetic polymorphisms.

Bridging the genotype-phenotype gap by linking different levels of biological complexity is far from trivial. Systems genetics is an approach that aims to improve our understanding of this relationship by unraveling the genetic variants and the molecular networks that underlie complex traits (Civelek and Lusis, 2014; van der Sijde et al, 2014). Until recently, this approach was based on the comparison between the position of QTLs underlying phenotypic trait variation to that of QTLs underlying the variation of intermediate molecular traits such as transcript expression (eQTLs) or protein abundance (pQTLs) (Christie et al, 2017; Blein-Nicolas et al, 2020). However, although the inclusion of such QTLs directly into inferred molecular networks showed its effectiveness in providing an in-depth genetic characterization of complex traits (Fagny et al, 2017; Capriotti et al, 2019), to our knowledge, this approach is still rare in the study of plant drought response and has not been used to unravel missing heritability. Several methods based on molecular network modeling have been proposed to link phenotypic and intermediate molecular traits (Hawe et al, 2019). Among these, Gaussian graphical models (GGMs) are powerful tools that allow the inference of conditional dependencies between phenotypic and molecular traits by generating a conditional dependency network (Shutta et al, 2022). The ability of GGMs to build sparse networks that consist only of direct effects enhances the relevance and interpretability of this approach for incorporating QTLs, eQTL, and pQTLs compared to the other networks based on classical correlation measures.

Here, we aimed to harness the potential of systems genetics approaches to gain insight into the genetic and molecular basis of complex traits to unravel the missing heritability of maize (*Zea mays*) drought re-

sponse. Maize is one of the most widely cultivated cereals in the world, with a production of more than 1.2 billion tonnes in 2021 (FAOSTATS, 2023). However, water deficit can lead to significant yield losses, ranging from 20% to 50%, depending on the developmental stage (Sah et al, 2020). The development of more drought-tolerant maize varieties is therefore becoming urgent to address the increasing frequency and intensity of droughts caused by climate change (Cook et al, 2018; Intergovernmental Panel on Climate Change, 2023).

To carry out our study, we used data and results previously generated from the same maize diversity panel composed of hybrids obtained after crossing 254 dent lines with a single flint tester. These data are composed of i) nearly one million SNPs obtained using a combination of high-throughput genotyping approaches (Negro et al, 2019), ii) six drought-adaptive ecophysiological traits measured under well-watered (WW) and water deficit (WD) conditions in four trials conducted on a high-throughput phenotyping platform (Alvarez Prado et al, 2017), and iii) 2,055 proteins quantified by mass spectrometry-based proteomics in leaves from plants grown in one of the four aforementioned trials (Blein-Nicolas et al, 2020). In a first study, Alvarez Prado et al (2017) associated the genomics and phenomics data through single-trial GWAS to analyze the genetic architecture of environmental effects on stomatal conductance. They could detect 531 QTLs spread across the trials and the watering conditions. Then, Blein-Nicolas et al (2020) associated the genomics and proteomics data, also through single-trial GWAS, to decipher the molecular mechanisms associated with the genetic polymorphisms underlying the variations of ecophysiological traits. They detected a set of 10,906 pQTLs in the WW condition (referred to as the pQ_w set) and a set of 11,930 pQTLs in the WD condition (referred to as the pQ_d set). Several of the pQTLs colocalized with QTLs previously detected by Alvarez Prado et al (2017). Altogether, these first two studies allowed to establish links between SNPs, proteins, and ecophysiological traits. Lastly, Djabali et al (2023) used the genomics and phenomics data to perform multi-trial GWAS, which allowed the detection of 42 QTLs for the ecophysiological traits in the WW condition (the Q_w set), 60 QTLs for the ecophysiological traits in the WD condition (the Qd set), and 40 QTLs for PL indices calculated from the phenotypic measurements made in the WW and WD conditions (the Qp set). Their results show that plasticity QTLs (PL QTLs) specifically contribute to the genotype x water availability (GxW) interaction.

In the present study, we go one step further by integrating phenomics, proteomics, and genomics data into a multi-scale network. To this end, we have developed an original systems genetics method that combines GGMs with the genetic information present in proteins. This allowed us to explain a part of the GxW variance of ecophysiological traits that was previously unraveled if only QTLs were considered.

3.2.2 . Results

One-third of the heritability for maize drought response is not explained by QTLs

In this first part, we aimed to estimate the proportion of missing heritability for maize drought response using the three QTLs sets previously reported by Djabali et al (2023) (*i.e.,* Q_w, Q_d, and Q_p). To this end, we sought to identify which set of QTLs, or combination thereof, best explained the GxW variance of the drought-related ecophysiological traits. We first analyzed the degree of overlap between Q_w, Q_d, and Q_p by looking at the colocalization between QTLs. This showed that the 142 QTLs actually covered 109 unique loci, of which five were common between Q_w and Q_d, two were common between the Q_d and Q_p, and one was common between Q_w and Q_p (Fig. 3.1a). This result shows that there is little redundancy between Q_w, Q_d, and Q_p.

We then assessed how the phenotypic variance in drought response was captured by Q_w, Q_d, Q_p, and their combinations by using a modeling approach similar to that described in Djabali et al (2023). This approach compares the variance components of a reference multi-environment mixed model, which does not include fixed effects of QTLs, to those of a multi-locus multi-environment mixed model which includes QTLs as fixed effects. Assuming that drought response is a highly integrated trait resulting from the combination of many genetically variable traits, we considered ecophysiological traits altogether rather than separately by fitting a multi-locus multi-trait multi-environment linear mixed model instead of single-trait multi-locus multi-trait multi-environment linear mixed models as in Djabali et al (2023). In this model, the total phenotypic variance was decomposed into five components: genotype (G), genotype by water

availability interaction ($G \times W$), genotype by ecophysiological trait interaction ($G \times P$), genotype by trial interaction ($G \times T$), and the residuals (ε). For the model of reference M_0, which does not include fixed effects of QTLs, the genotype contributed to 13% of the total phenotypic variance, the $G \times W$ interaction to 7%, the $G \times P$ interaction to 21%, the $G \times T$ interaction to 9%, and the residuals to 50% (Fig. 3.1b).

We defined then seven multi-locus multi-trait multi-environment linear mixed models (M_1-M_7), each one including one of the three Q_w, Q_d, or Q_p sets, or a combination of them (Fig. 3.1c). The genetic variances captured by these sets of QTLs are presented in Table 3.1, and the proportion of captured variance is illustrated in Fig. 3.1d. Depending on the model, QTLs captured 5%-80% of the G variance, 2%-65% for the $G \times W$ variance, 5%-78% for the $G \times P$ variance, 5%-20% for the $G \times T$ variance, and 1%-5% of the ε variance. The combination of Q_w and Q_p in M_5 captured the highest proportion of the GxW variance (65%). By comparison, the combination of Q_w, Q_d, and Q_p in M_7 captured 60% of the GxW variance. This result indicates that WD QTLs are not required to explain the GxW interaction. This is confirmed by M_2, where Q_d alone captured only 7% of the GxW variance. Together, Q_w and Q_p gathered 82 QTLs spanning 351 genes with no functional enrichment (Table S1).

Overall, this work shows that the combination of Q_w and Q_p, denoted to as Q_{ref} in the following, best explains the phenotypic variations in maize drought response, with a baseline at 65% of explained GxW variance. Considering these QTLs, the proportion of missing heritability for maize drought response is, therefore 35%.

Model	σ^2_G	$\sigma^2_{G\times W}$	$\sigma^2_{G\times T}$	$\sigma^2_{G\times F}$	σ^2
M_0	0.026	0.015	0.042	0.019	0.097
M_1	0.010	0.089	0.022	0.018	0.095
M_2	0.007	0.014	0.022	0.018	0.095
M_3	0.005	0.008	0.015	0.017	0.093
M_4	0.025	0.007	0.040	0.018	0.095
M_5	0.012	0.005	0.022	0.017	0.094
M_6	0.009	0.007	0.021	0.016	0.094
M_7	0.006	0.006	0.014	0.015	0.092

Table 3.1: Random effects variances captured by the different set of QTLs included in the models M_1 - M_7. The model M_0 is the reference model without QTLs.

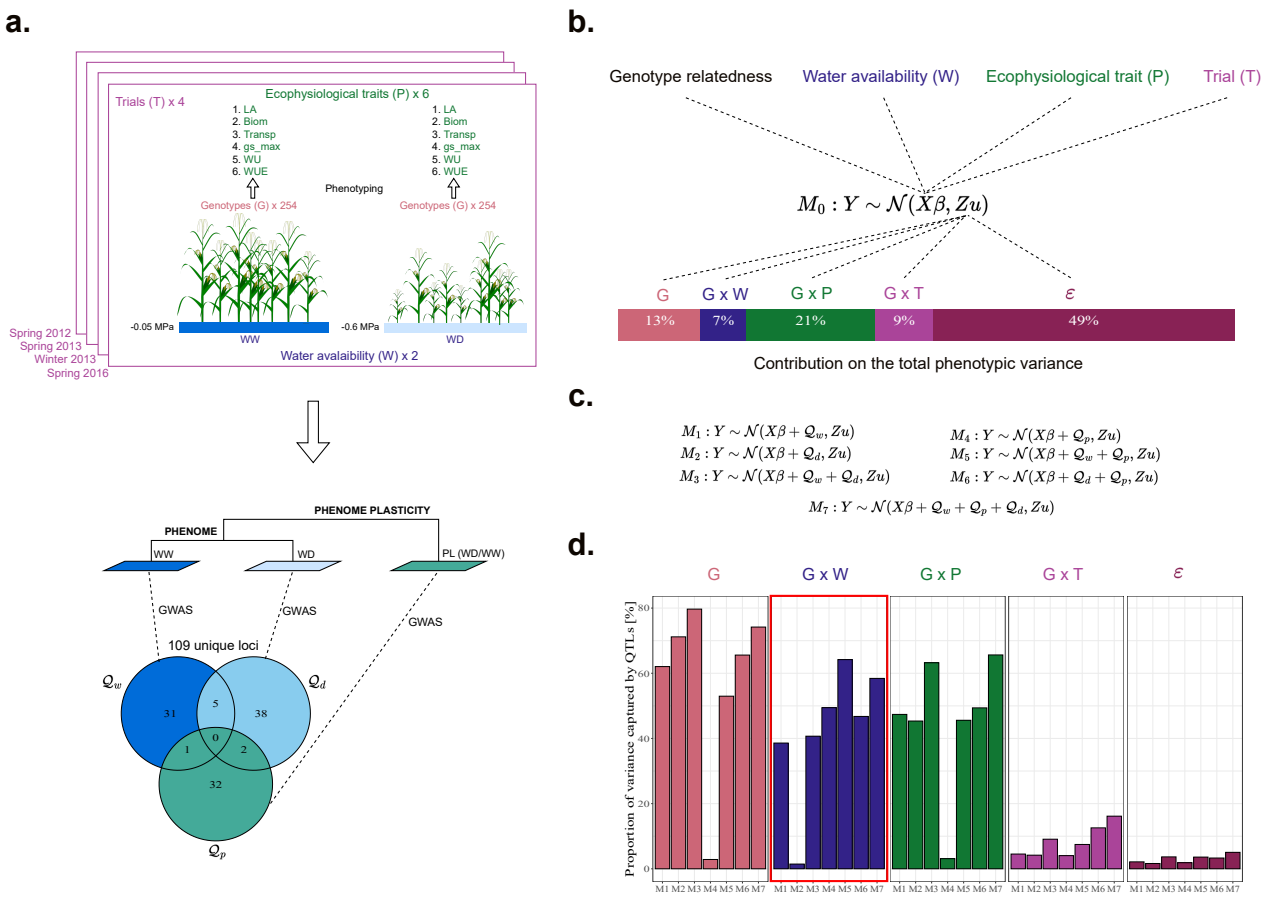


Figure 3.1: **Genome-wide association studies exhibit a genetic architecture related to water deficit response. a.**, Schematic representation of the study design and summary of the results of the GWAS performed by Djabali et al (2023). A panel of 254 maize hybrids was grown in well-watered (WW) and water-deficit (WD) conditions in a phenotyping platform during four trials: spring 2012, spring 2013, winter 2013, and spring 2016. From the four trials, six drought-related ecophysiological traits were measured: leaf area (LA), biomass (Biom), transpiration rate (Transp), stomatal conductance (gs_max), water uptake (WU), and water use efficiency (WUE). Phenotypic plasticities were calculated as the ratio between WD and WW phenomics data. A multi-trial GWAS was conducted on WW, WD and PL phenomics data. The Venn diagram displays the colocalization between the QTLs in Q_w, Q_d, and Q_p. **b.**, Schematic representation of the multi-trait multi-environment linear mixed model M_0. All phenotypic means were scaled and concatenated together in one vector Y. $X\beta$ corresponds to fixed effects: the genotype relatedness, the water availability (W), the ecophysiological trait (P), and the trial (T). Zu corresponds to random effects: the genotype (G), the genotype by the water availability interaction (GxW), the genotype by the ecophysiological trait interaction (GxP), the genotype by the trial (GxT), and the residual error. **c.**, Description of the multi-locus multi-trait multi-environment linear mixed models $M_1 - M_7$. **d.**, Proportion of random effect variances captured by QTLs in the M_1 to M_7 model

pQTLs represent an important reservoir of genetic polymorphisms for unraveling missing heritability

Previous work by Blein-Nicolas et al (2020) has shown that pQTL/QTL colocalization can occur for co-expressed proteins whose abundance is correlated with ecophysiological traits, suggesting that the same genetic polymorphism may be responsible for both protein co-expression and phenotypic trait variation. Based on these results, for those proteins whose abundance is correlated with phenotypic traits, we hypothesize that pQTLs not colocalized with QTLs may also influence the traits and thus be part of the missing heritability. Under such a hypothesis, we expect the number of pQTLs that do not colocalize with QTLs to be high. To test this, we compared the genetic landscape of the proteome with that of the drought response as determined by Q_{ref}.

We first complemented the two pQTL sets previously detected by Blein-Nicolas et al (2020), pQ_w and pQ_d, with a set of 10,073 plasticity pQTLs, denoted pQ_p, thus bringing the total number of available pQTLs to 32,909. We then performed a pQTLs colocalization analysis for each protein and observed that the intersection between pQ_w, pQ_d, and pQ_p was always empty (Fig. S1a). Furthermore, the intersections between

pQ_w and pQ_p and between pQ_d and pQ_p were also empty for more than 90% of the proteins having at least one pQTL in each pQTL set. The intersection between pQ_w and pQ_d was also empty for more than 75% of the proteins. However, when we analyzed the colocalization of pQTLs for all proteins simultaneously, we showed that 5,006 (59%) of the 8,476 unique loci spanned by the 32,909 pQTLs were shared between pQ_w, pQ_d, and pQ_p (Fig. 3.2a). This means that the same locus can be shared by several pQTLs sets while being associated with different proteins according to the plant water status and its plasticity. We, therefore, examined how the 8,476 unique loci represented by the three pQTL sets were distributed across the genome. This allowed us to identify 100 loci associated with more than twenty different proteins, which we call hereafter hotspots (Table S2).

We then analyzed the position of the QTLs in Q_{ref} relative to that of the pQTLs. In total, Q_{ref}, pQ_w, pQ_d, and pQ_p represented 8,488 unique loci (Fig. 3.2a), of which 8,416 (99.2%) were spanned by pQTLs only, 12 (0.1%) were spanned by QTLs only, and 60 (0.7%) were spanned by both pQTLs and QTLs. Of these 60 loci, five correspond to hotspots of pQTLs located on chromosomes 1, 8, and 10 (Fig. 3.2b). These five hotspots included 156 proteins that were associated with a total of 2,645 pQTLs. Among these pQTLs, 93% were located outside the five hotspots that colocalized with QTLs.

For each of the five hotspots that included a QTL, we performed a functional analysis of the pQTL-associated proteins and we constructed a protein-protein interaction (PPI) network taking into account both the pQTL-associated proteins and the genes covered by the hotspot. The results of these analyses are summarized in Table 3.2 (Hotspots c242, c267, c268, c6902, and c8174). They show that, i) regardless of the hotspot considered the pQTL-associated proteins show a significant functional enrichment; ii) many pQTL-associated proteins (48% to 57%) and only a few genes covered by the hotspot (3% to 17%) are involved in PPI network; iii) many of the interactions in the PPIs (25% to 53%) are experimentally validated, iv) depending on the hotspots, many of the experimentally validated interactions in the PPIs (17% to 66%) link a gene covered by the hotspot to a pQTL-associated protein. For example, for hotspot c242, the gene encoding the unannotated protein GRMZM2G087275_P01 was linked to the pQTL-associated protein GRMZM2G017110_P01 (glutamate decarboxylase). For hotspot c267, the gene encoding the unannotated protein GRMZM2G309363_P01 was linked to three pQTL-associated proteins: GRMZM2G154595_P01 (malate dehydrogenase 2), GRMZM2G415359_P02 (malate dehydrogenase 4) and GRMZM2G033208_P01 (transketolase 1). These three proteins were exclusively linked by experimentally validated interactions (Fig. 3.2c). For hotspot c268, the gene encoding the protein GRMZM2G102365_P02 (thioredoxin-like 1-1 chloroplastic) was linked to the pQTL-associated proteins GRMZM2G577677_P02 (peptide methionine sulfoxide reductase msrB) and GRMZM2G011025_P01 (methionine sulfoxide reductase 6). For hotspot c6902, the gene encoding the protein GRMZM2G085577_P01 (alpha-1,4-glucan phosphorylase) was linked to the pQTL-associated protein GRMZM2G155253_P02 (fructose-bisphosphate aldolase). Finally, for hotspot c8174, the gene encoding the protein GRMZM2G124411_P04 (60S ribosomal protein) was linked to the pQTL-associated proteins GRMZM2G448151_P01 (30S ribosomal protein) and GRMZM2G448142_P02 (NAD(P)H-quinone oxidoreductase subunit K) (Fig. 3.2c). These two proteins were linked by a non-experimentally validated interaction and are encoded by genes with overlapping coding DNA sequences.

Hotspot	Chr.	Start-End (Mb.)	# G[1]	# P[2]	PPIs Comp.[3]	Exp./Tot.[4]	# G/Exp.[5]	Functional enrichment on pQTL-associated proteins (p-value)
c242	1	34-35	74	37	6-19	$17/32$	$3/17$	Glutathione metabolism (1.7e-05)
c267	1	52-53	63	39	4-19	$23/47$	$8/23$	Fructose and mannose metabolism (1.7e-03)
c268	1	52-54	57	28	2-14	$3/12$	$2/3$	Alpha-linolenic acid metabolism (2.7e-02)
c6902	8	145-147	150	27	26-14	$22/51$	$7/22$	Biosynthesis of secondary metabolites (1.5e-02)
c8174	10	126-128	179	33	21-19	$26/53$	$9/26$	Aminoacyl-tRNA biosynthesis (1.3e-03)
c7403	9	87-91	151	41	14-27	$21/64$	$10/21$	Aminoacyl-tRNA biosynthesis (3.9e-02)
c8166	10	112-115	187	31	10-12	$2/22$	$1/2$	-

[1] Number of genes located in the hotspot.
[2] Number of proteins associated with pQTLs located in the hotpot.
[3] PPIs composition: Number proteins encoded by genes located in hotspot - Number of proteins associated with pQTLs located in the hotpot.
[4] Number of experimentally verified interactions / Total number of the interactions present in the PPIs
[5] Number of experimentally verified interactions linked to proteins encoded by genes located in hotspot / Total number of experimentally verified interactions present in the PPIs

Table 3.2: Composition and PPI associated with selected pQTLs hotspots

Overall, the PPI analysis of five hotspots of pQTLs shows that the proteins associated with pQTLs located in the same hotspot can physically interact with each other. These interactions can occur between proteins associated with the three types of pQTLs. This result thus confirms that it is highly relevant to jointly analyze the genetic landscapes of the proteome obtained from different datasets. Furthermore, we showed that physical interactions can also occur between the pQTL-associated proteins and the proteins encoded by the genes covered by the hotspot. This provides clues to identify the candidate genes and the molecular mechanisms underlying pQTLs. Finally, we identified a large number of pQTL hotspots distributed across the genome, only a few of which also included QTLs. For those that contained loci shared between pQTLs and QTLs, the pQTL-associated proteins were also associated with many other pQTLs located elsewhere in the genome. These results suggest that pQTLs represent an important potential reservoir of genetic polymorphisms for unraveling missing heritability.

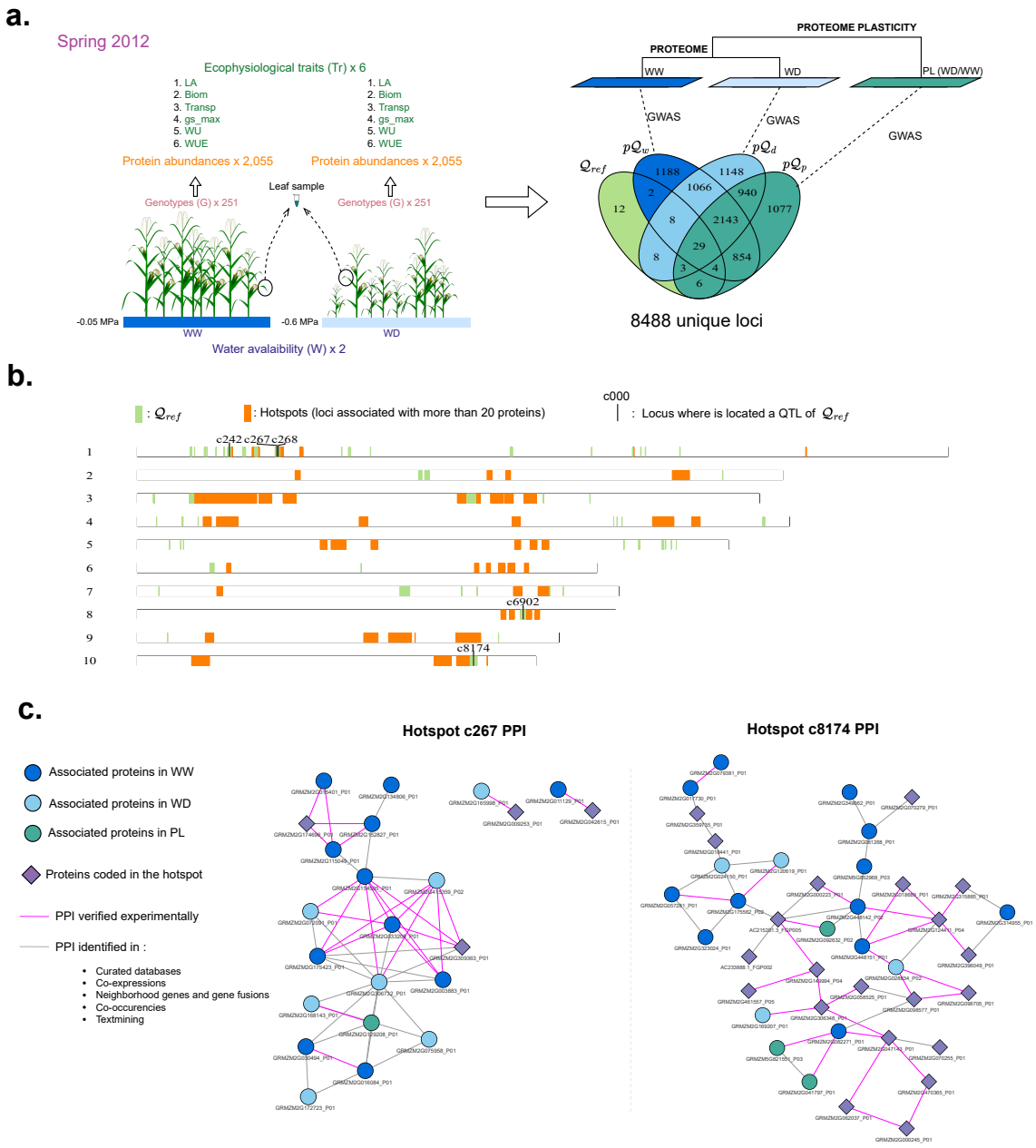


Figure 3.2: **Harnessing proteomics data to identify and characterize genomic regions enriched in pQTLs a.**, Schematic representation of the proteomics study design and GWAS related to protein abundances. Protein abundances were quantified by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) from leaf samples taken on plants grown during the spring 2012 trial. Phenotypic plasticities were calculated as the ratio between WD and WW proteomics data. Then, single-environment GWASs were performed on the WW, WD, and PL proteomics data. The Venn diagram displays the colocalization between Q_{ref}, pQ_w, pQ_d, and pQ_p. **b.**, Physical positions of the QTLs in Q_{ref} and pQTL hotspots on the maize genome. The position of hotspots colocalizing with QTLs in Q_{ref} are represented in black. **c.**,PPI networks obtained for hotspots c267 and c8174.

pQTLs that capture missing heritability for drought response can be identified using systems genetics

In order to unravel the 35% of missing heritability observed for maize drought response, we implemented an approach of systems genetics, which aims to identify pQTLs that could capture part of this missing heritability. This approach is based on the inference of trait-protein co-expression networks using Gaussian graphical models (GGMs), which have the advantage of estimating the strength of the relationship between two variables while taking into account the effects of the other variables. We inferred trait-protein networks from the three following datasets: i) the phenotypic means in the WW condition (WW dataset); ii) the phenotypic means in the WD condition (WD dataset); iii) the plasticity indices (PL dataset). To avoid the ultra-high dimension phenomenon, we reduced the number of proteins used as input in the GGMs by focusing on the 951 proteins that had at least one pQTL in each of the pQ_w, pQ_d, and pQ_p sets. These 951 proteins represented a pool of 7,385 unique loci out of the 8,437 total loci spanned by pQ_w, pQ_d, and pQ_p.

For each of the three datasets considered, we first estimated a GGM by using the Glasso procedure. To examine how the networks evolved, we considered the same regularization grid of length 100 (the lower the regularization parameter, the higher the number of nodes and edges). This allowed us to examine how the networks evolved with the regularization parameter λ (Fig. 3.3a). We observed that, regardless of the dataset, trait-trait edges appeared before trait-protein edges. Furthermore, the number of trait-protein edges was much higher for WU, Biom, and LA than for gs_max, WUE, and Transp. This difference between the ecophysiological traits was more important in the WD dataset than in the WW and PL datasets.

Several choices of mathematical criteria exist to select the most relevant regularization parameter and get the coexpression network. These criteria are based on mathematical considerations to be sure that the estimated network has good statistical properties. Here, we decided to select the regularization parameter that maximizes the proportion of GxW variance explained by adding the pQTLs of the proteins in the neighborhood of the traits in the M_5 model. We limited the research to the first 70 proteins present in the neighborhood of the traits. Thus, for the WW and PL datasets, we could select a regularization parameter λ equals to 0.40 and 0.25, respectively, but none could be selected for the WD dataset (Fig. 3.3b).

The network obtained for the WW dataset contained 7 proteins associated with 54 pQTLs, eight of which were located in pQTL hotspots (one of these hotspots was hotspot c8174 described above). The network representation is shown in Fig. 3.3c. When added to Q_{ref} in M_5, the pQTLs associated with the proteins in the network increased the proportion of GxW variance explained from 64% to 70%. To better illustrate the influence of these pQTLs on ecophysiological traits, we focused on SNP S7_70368465 which was significantly associated with protein GRMZM2G079256_P01 (p-value $<$ 1e-16) but not with WU (p-value $=$ 0.085) in the WW condition. By considering the abundance of GRMZM2G079256_P01 as a fixed covariate in the GWAS model used for QTL detection, S7_70368465 became significantly associated with WU (p-value $=$ 0.012, Fig. 3.3d). This result confirms that S7_70368465 is part of the missing heritability for WU.

The network obtained for the PL dataset, whose representation is shown in Fig. S2, contained 45 proteins, three of which were shared with those of the network obtained for the WW dataset. These 45 proteins were associated with 258 pQTLs, 31 of which were located in pQTL hotspots, including hotspots c267, c242, and c268. When added to M_5, these pQTLs increased the proportion of GxW variance explained from 64% to 69%. Therefore, despite its proteins being associated with the highest number of pQTLs, the network obtained for the PL dataset did not allow to improve the proportion of GxW variance explained as compared to the network obtained for the WW dataset.

Altogether, these results show that pQTLs can actually capture missing heritability for drought response. These genetic polymorphisms were not detected directly from the traits by GWAS because of a lack of statistical power and an effect of genetic buffering, as illustrated by SNP S7_70368465. As the networks were obtained from the WW and PL datasets on the basis of maximizing the proportion of GxW variance explained, they are expected to reflect molecular mechanisms that occur in response to drought. The fact that they share proteins suggests that they may not be independent of each other.

An incremental systems genetics approach allows to increase the part of missing heritability captured by pQTLs

To take into account that networks inferred from the WW, WD, and PL datasets may not be independent, we developed a variant of our systems genetic approach to unravel missing heritability called the incremental approach. In practice, this consists of selecting the values of λ to infer networks from successive datasets using multi-locus multi-trait multi-environment linear mixed models that incrementally take into account the pQTLs associated with the proteins of the successively inferred networks. Since no network could be selected from the WD dataset (Fig. 3.3b), we explored all the possible incremental paths starting from the networks inferred either from the WW or the PL dataset. This led us to define ten multi-locus multi-trait multi-environment linear mixed models (M_8 to M_{17}), which are presented in Table 3.3. These models all included as fixed effects the set Q_{ref} supplemented by one, two or three sets of pQTLs, depending on the incremental path. Models M_8 and M_{13} correspond to the one-step incremental paths that gave rise to the two networks previously inferred from the WW and PL datasets.

Starting from the network inferred from the WW dataset, two two-step incremental paths were possible: $WW \rightarrow PL$ (model M_9) and $WW \rightarrow WD$ (model M_{11}). Using model M_9, a network was inferred from the PL dataset. It contained 48 proteins associated with 271 pQTLs, allowing to gain 8 points in GxW variance as compared to M_8 (78% *vs* 70%). Model M_{11} allowed to infer a network from the WD dataset. This network contained 30 proteins associated with 272 pQTLs, allowing a gain of only 1 point in GxW variance as compared to M_8.

Starting from the network inferred from the PL dataset, two two-step incremental paths were possible: $PL \rightarrow WW$ (model M_{14}) and $PL \rightarrow WD$ (model M_{16}). Using model M_{14}, a network was inferred from the WW dataset. It contained 8 proteins associated with 56 pQTLs that allowed to gain only 1 point in GxW variance as compared to model M_{13} (70% *vs* 69%). No network could be inferred from the WD dataset using model M_{16}.

Three three-step incremental paths were then possible: $WW \rightarrow PL \rightarrow WD$ (model M_{10}), $WW \rightarrow WD \rightarrow PL$ (model M_{12}), and $PL \rightarrow WW \rightarrow WD$ (model M_{15}). Models M_{10} and M_{15} both allowed the inference of networks from the WD dataset. The network obtained using M_{10} contained 23 proteins associated with 228 pQTLs, allowing a gain of 6 points in GxW variance as compared to M_9 (84% *vs* 78%, respectively). The network obtained using M_{15} also contained 23 proteins, but these were associated with 228 pQTLs that allowed a gain of 11 points in GxW variance as compared to M_{14} (81% *vs* 70%, respectively). Finally, model M_{12} allowed the inference of a network from the PL dataset. This network contained 11 proteins associated with 52 pQTLs that increased the proportion of GxW variance explained to 82%, a gain of 11 points as compared to M_{11}.

Altogether, these results show that the protein composition of networks inferred from the same dataset varies according to the incremental path. For example, the networks inferred from the PL dataset using M_9, M_{12}, and M_{13}, respectively, contain 48 proteins associated with 271 pQTLs, 11 proteins associated to 52 pQTLs, and 45 proteins associated to 258 pQTLs. This indicates that the networks inferred from two different datasets are not independent. Our results also show that the proportion of GxW variance explained depends on the incremental order. This is illustrated by M_{10}, M_{12}, and M_{15}: these three models correspond to different three-step paths, but they explain 84%, 82% and 81% of GxW variance, respectively. Interestingly, the incremental path that gave the best result was $WW \rightarrow PL \rightarrow WD$ (M_{12}, 84% of GxW variance explained). This path is consistent with the watering changes undergone by plants cultivated in the WD condition. Indeed, these plants were first grown under optimal soil water content before being submitted to progressive mild water deficit.

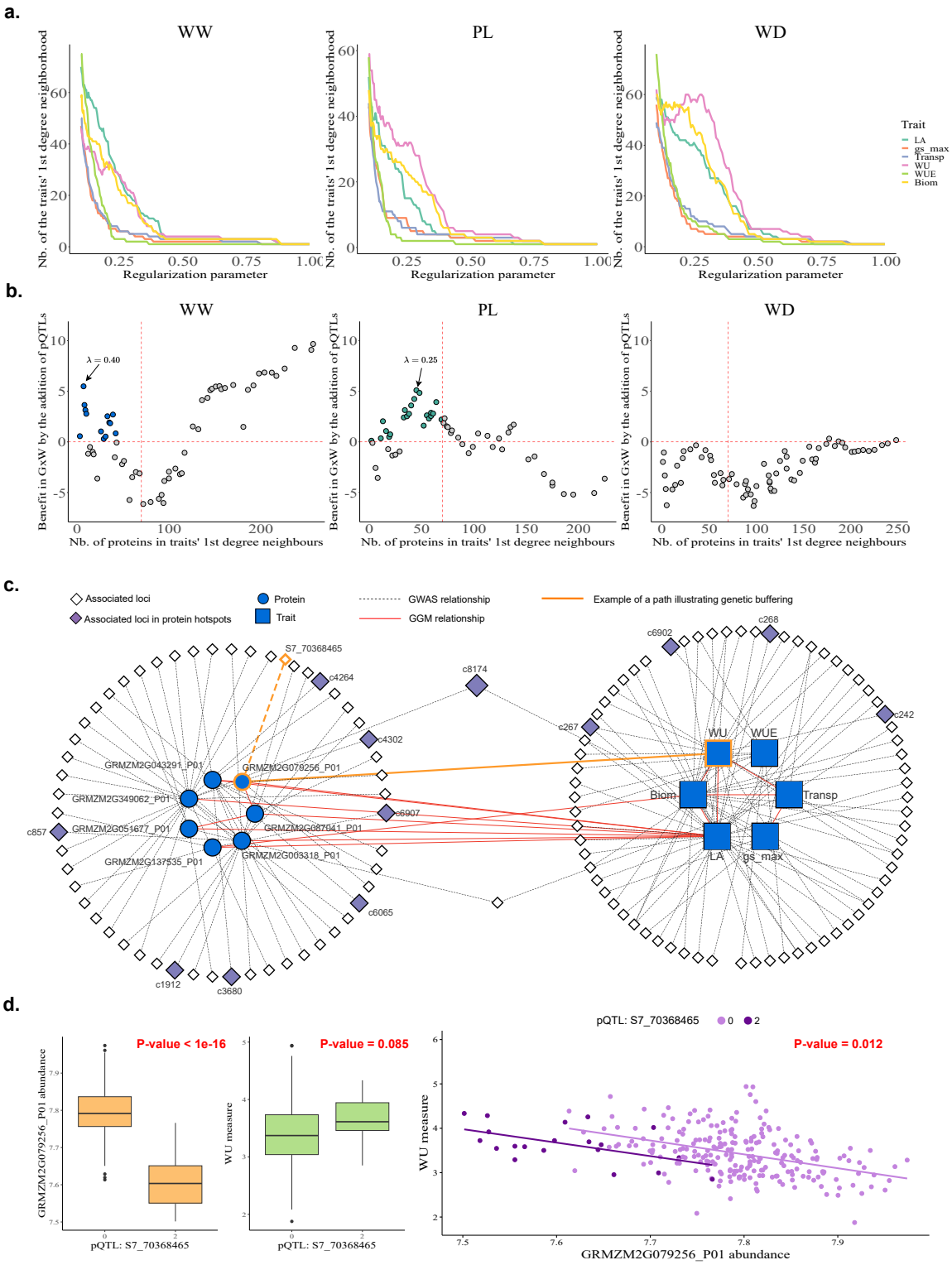


Figure 3.3: **Identification of pQTLs capturing additional part of GxW variance. a.**, Number of the traits' first-degree neighbors in coexpression networks according to the regularization parameter. **b.**, Gain in the part of GxW variance captured by adding pQTLs according to the number of proteins present in the traits' first-degree neighbors at each λ. **c.**, Representation of the multi-scaled network selected in the WW condition.**d.**, Illustration of the buffering effect of pQTLs "S7_70368465" on WU through the associated proteins GRMZM2G079256_P01. Orange and green boxplots represent maize genotypes at marker S7_70368465 according to the abundance of GRMZM2G079256_P01 and the measurement of WU, respectively. The scatter plot shows the abundance of GRMZM2G079256_P01 in function of WU measurement in the WW condition. Each point is colored according to the genotype at marker S7_70368465.

Model	Incremental path	$\sigma^2_{G\times W}$	γ^1	# P[2]	# pQTLs[3]
M_8	WW	0.0044	70	7	54
M_9	$WW \rightarrow PL$	0.0032	78	48	271
M_{10}	$WW \rightarrow PL \rightarrow WD$	0.0024	84	23	228
M_{11}	$WW \rightarrow WD$	0.0043	71	30	272
M_{12}	$WW \rightarrow WD \rightarrow PL$	0.0026	82	11	52
M_{13}	PL	0.0045	69	45	258
M_{14}	$PL \rightarrow WW$	0.0044	70	8	56
M_{15}	$PL \rightarrow WW \rightarrow WD$	0.0028	81	23	228
M_{16}	$PL \rightarrow WD$	NA	NA	NA	NA
M_{17}	$PL \rightarrow WD \rightarrow WW$	NA	NA	NA	NA

[1] Part of GxW variance captured
[2] Number of proteins in the inferred network.
[3] Number of pQTLs associated with the proteins in the inferred network.

Table 3.3: Description of ten multi-trait multi-environment linear mixed models quantifying the GxW variance captured by the associated multi-scaled network.

Modeling a multiscale network of maize drought response

Since the genomics landscape related to the networks obtained from the three WW, WD, and PL datasets following the incremental path $WW \rightarrow PL \rightarrow WD$ were not independent (Fig. 3.4), we represented them as a single multiscale network. To do that, we collapsed their 553 (54+271+228) pQTLs and QTLs in Q_{ref}, into 531 unique loci (Fig. 3.4). If a protein or a trait appeared in several networks, it was represented as a single node. The resulting network, shown in Fig. 3.5, included 600 nodes and 996 edges. Sixty-three (10%) of the nodes were proteins. Of them, three were linked to traits only in the network inferred for the WW dataset, 35 were linked to traits only in the network inferred for the PL dataset, and 11 were linked to traits only in the network inferred for the WD dataset. This result shows that for capturing the highest proportion of GxW variance, it was necessary to combine information from the three datasets. We further investigated these proteins by constructing a PPI network (Fig. 3.5b). The major component of the PPI network was composed of 29 nodes and 52 edges, 37 of which were experimentally verified and linked proteins selected in different networks. This component included a strongly connected clique mainly composed of proteins involved in response to stress and protein folding all linked by experimentally verified interactions: four heat-shock proteins (GRMZM2G063676_P01, GRMZM2G002220_P01, GRMZM2G153815_P01, and GRMZM5G813217_P01), two chaperons (GRMZM2G127609_P06, GRMZM5G856084_P01), one endoplasmin-like protein (GRMZM2G141931_P01), and an unannotated protein (GRMZM2G130121_P01). This clique is connected with two translationally controlled tumor proteins (TCTP) known to be involved in drought stress signalization (Kim et al, 2012) and with another group of highly connected proteins involved in the oxidation-reduction process. This group is composed of an NADH-cytochrome b5 reductase (GRMZM2G061830_P02), a putative citrate synthase (GRMZM2G063909_P01), an enoyl-(acyl-carrier-protein) reductase (GRMZM2G079256_P01), a ferredoxin-1 (GRMZM2G043162_P01), an aldose reductase (GRMZM2G479423_P01), and a pheophorbide a oxygenase (GRMZM2G349062_P01).

Among the 531 unique loci involved in the network and representing 89% of the nodes, 48 were located in pQTL hotspots. To evaluate the importance of these hotspots, we calculated the loss in GxW variance induced when model M_{10} was run without the pQTLs and QTLs determined by the loci located in the same hotspot (Table S2). The top three most important hotspots were c8166, c7403, and c8174, causing a loss of GxW variance of -7.35, -7.29, and -7.13 points, respectively. Hotspots c7403 and c8166 globally had the same characteristics as the hotspots colocalizing with QTLs previously described and including c8174 (Table 3.2). Briefly, hotspot c7403 contained pQTLs associated with proteins enriched in proteins involved in the biosynthesis of amino acids, RNA degradation, and oxidative phosphorylation. Eight experimentally verified PPI linked pQTL-associated proteins to genes covered by hotspot c7403. One of these linked the gene encoding the unannotated protein GRMZM2G309363_P01 to four highly interconnected pQTL-associated proteins: GRMZM2G027451_P04 (60S ribosomal protein), GRMZM2G099352_P03 (ribosomal protein S3), GRMZM2G072315_P01 (mouse transplantation antigen homolog), and GRMZM2G315088_P01 (unannotated). Hotpot c8166 gathered pQTLs associated with proteins showing no functional enrichment. One of these protein, GRMZM2G140288_P01 (histone acetyltransferase) was linked by an experimentally verified

PPI to a gene covered by the hotspot that encode the protein GRMZM2G179005_P01 (histone H3.2).

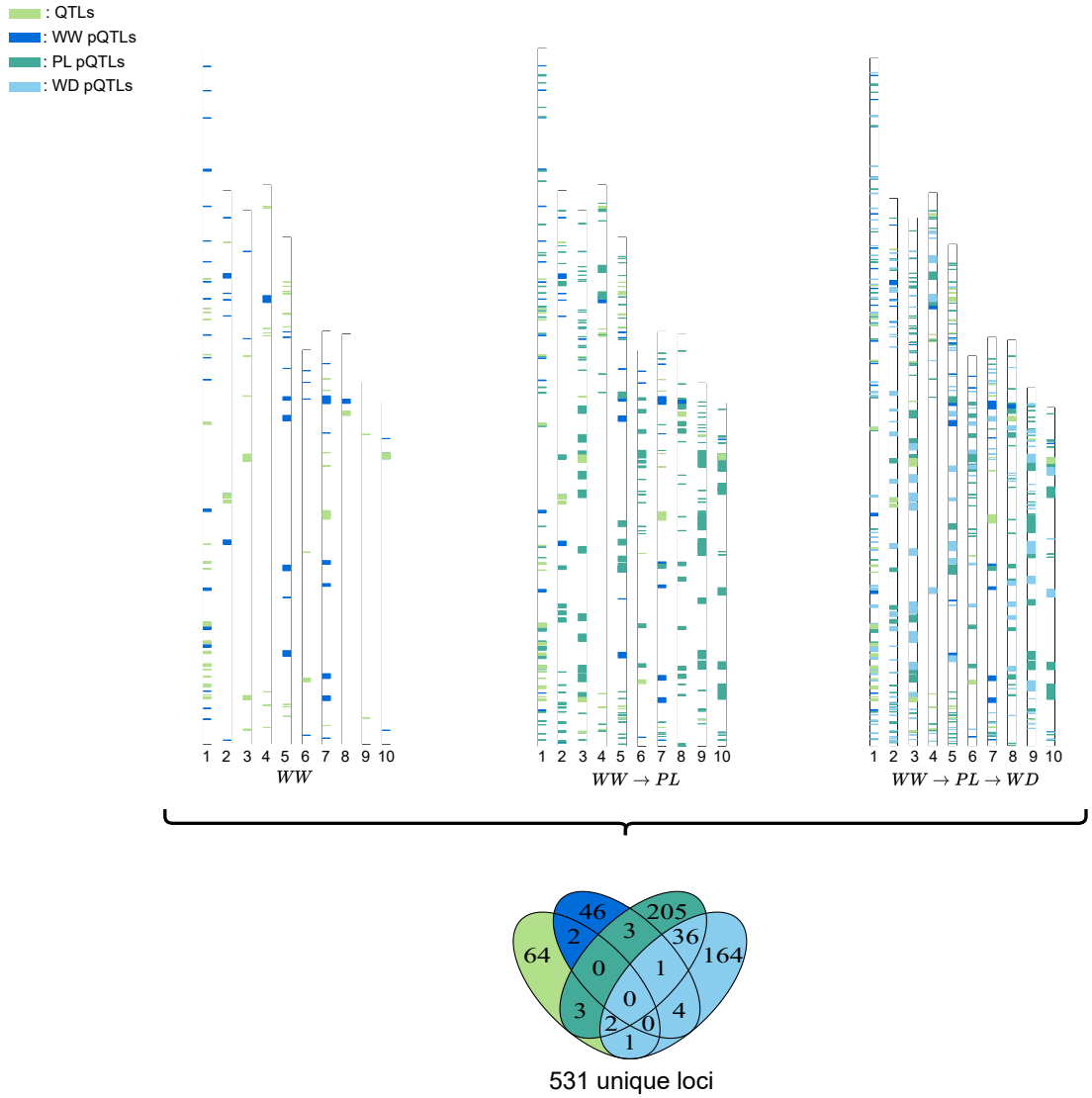


Figure 3.4: **Incremental path conducting to the genomics landscape associated to** M_{10}.

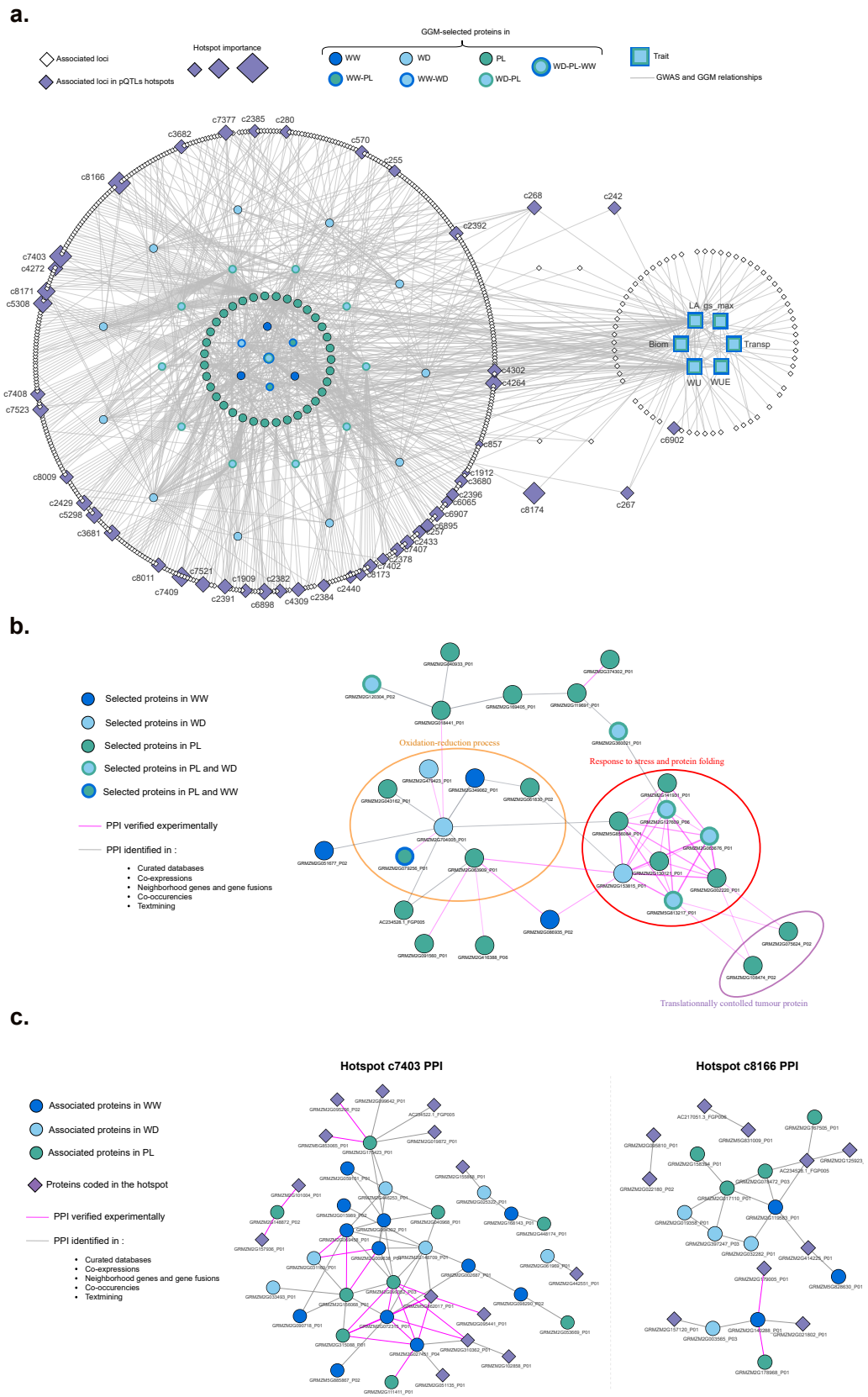


Figure 3.5: **Functional annotation of the multi-scaled network explaining the highest part of the GxW variance a.**, Representation of the multi-scaled network inferred following the incremental path $WW \rightarrow PL \rightarrow WD$. **b.**, PPI network obtained on the final set of proteins directly linked to ecophysiological traits in the multi-scaled network, explaining the highest part of the GxW variance. **c.**, PPI networks obtained for hotspots c7403 and c8166.

3.2.3 . Discussion

To better characterize the genetic determinism of maize drought response, we conducted an integrative analysis of phenomics, proteomics, and genomics data obtained on a diversity panel of 254 maize genotypes. These analyses were based on detecting and evaluating QTLs and pQTLs from the joint use of GWAS and multi-environment multi-locus mixed linear models. Multi-environment multi-locus mixed linear models (van Eeuwijk et al, 2010) are a powerful approach to decompose the GxE variance into more precise secondary interactions and evaluate to which random variances a set of QTLs contributes (Millet et al, 2016; Alvarez Prado et al, 2017). However, the extension of such models to incorporate measurements obtained in several traits is limited to GWAS (Korte et al, 2012) or genomic predictions (Runcie et al, 2021). Here, we used the set of drought adaptative traits present in phenomics data to better take into account the drought response complexity and estimate an overall GxW variance. Thus, we showed that the major part of GxW variance could be explained by the complementarity between QTLs detected in the WW condition and plasticity QTLs, excluding the QTLs detected in the WD condition.

The functional analysis of the global set of QTLs maximizing the part of GxW captured showed as not able to highlight molecular insights. This supports one attribute of the omnigenic model of (Boyle et al, 2017) that stipulates that the enrichment of signal in relevant genes is weak for complex traits overall. Thus, we decided to conduct a more targeted functional analysis on genetic regions presenting pleiotropic effects on protein abundances. We identified pQTLs hotspots (Porth et al, 2012; Blein-Nicolas et al, 2020; Acharjee et al, 2018) associated with protein plasticity and protein abundances quantified in the two watering conditions. The discovery of pQTL hotspots that also included QTLs gave us the possibility to transcribe genetic variations, capturing GxW variance into biological pathways. Indeed, the identification of PPI between proteins encoded by genes located in hotspots with proteins associated with the pQTLs that defined the hotspots allowed us to place genetic variations into glutathione, fructose, and alpha-linolenic acid metabolism pathways known to be directly related to drought response and drought tolerance (Hasanuzzaman et al, 2017; Guo et al, 2018; Zi et al, 2022). Our results showed the potential of QTL integration in PPI to provide comprehensive functional support by incorporating candidate genes into known biological networks.

Previous work by Blein-Nicolas et al (2020) has shown that pQTL/QTL colocalization can occur for co-expressed proteins whose abundances are correlated to ecophysiological traits, suggesting that the same genetic polymorphism can be responsible for both the co-expression of proteins and the variations of ecophysiological traits. This hypothesis is supported by the results presented above. Based on these results, we hypothesized that, for those proteins whose abundance is correlated with a phenotypic trait, the pQTLs that do not colocalize with a QTL for the trait in question may also influence the trait. To test this hypothesis and further exploit the proteome, we implemented an original approach of systems genetics combining network modeling with multi-locus multi-trait multi-environment linear mixed models. This approach allowed us to jointly identify a protein set directly related to ecophysiological traits with their attributed pQTLs, capturing an additional part of the GxW variance. Relations between ecophysiological traits, the selected proteins, and the associated loci were modeled as multi-scale networks, highlighting the indirect relation of pQTLs with ecophysiological traits. Indeed, an explanation of the non-detection of these QTLs on the ecophysiological traits could be explained by the lack of statistical power in the detection of all genetic polymorphisms involved in trait variation (Yang et al, 2010; Shi et al, 2016). These undetected QTLs include rare variants but also variations affecting molecules that are not translated into variation on the complex trait due to genetic buffering (Gibson and Dworkin, 2004; Fu et al, 2009; Félix and Barkoulas, 2015). Thus, we showed that by exploiting the intermediate molecular traits, we could enrich our analysis with molecular QTLs that were not detected on the complex traits. Together, we identify 473 additional loci, bringing a gain of 20 points on the captured GxW variance. Our findings are consistent with the omnigenic model (Boyle et al, 2017) that stipulates that part of the missing heritability (Manolio et al, 2009) can be unraveled by considering peripheral genes.

GWAS provides a fixed picture of the genetic landscape of the proteome, obtained from a dataset representing a particular context (WW, WD or PL). However, these pictures are not independent of each other, as the response to drought is a dynamic process that allows the plant to modify its metabolism and cellular function according to the availability of water in the soil. Given this observation, we therefore hypothesised that the previously inferred WW, PL and WD networks are not independent of each other, and that taking into account a dependency structure between the WW, WD and PL networks could provide a better explanation for the GxW variance. Thus, by applying our integrative approach incrementally, we are able to verified our hypothesis by showing that the maximum part of GxW captured is obtained by following an incremental path consistent with the watering changes undergone by the plants cultivated in the WD condition ($WW \rightarrow PL \rightarrow WD$). Thus, the representation of these results as a multi-scale network gives rich support to genetic and molecular annotations related to drought response. The proteins that are directly related to the traits in the network highlight the process of signalization, protein folding, and oxidation-reduction processes. Finally, we identify two hotspots spanned exclusively detected by pQTLs contributing near to 10 points of the total part of the GxW variance captured. These two hotspots provide a list of candidate genes interacting with functionally known proteins using the PPIs identified.

3.2.4 . Methods

Description of the plant material and the platform experiments

The diversity panel includes 254 maize dent lines originating from Europe and the Americas (Rincent et al, 2014b). It can be subdivided into four genetic subgroups: Iodent, Lancaster, Stiff-stalk, and F252-like. Lines were selected within a restricted flowering window to avoid confounding drought escape effect due to variation in flowering time with the expression of genomic regions involved in drought response (Millet et al, 2016; Alvarez Prado et al, 2017; Negro et al, 2019). The panel was genotyped using a 50K Infinium HD Illumina array (Ganal et al, 2011), a 600K Axiom Affymetrix array (Unterseer et al, 2014) and 500K markers obtained by genotyping by sequencing (Negro et al, 2019). The genomic dataset thus generated consisted 977,459 SNPs mapped to the maize reference genome B73_AGP_v2 (release 5a) (Schnable et al, 2009; Negro et al, 2019). Only SNPs with a minor allele frequency (MAF) below 0.05 and a heterozygosity rate above 0.15 were considered. Missing values were imputed using Beagles 3.1 (Browning and Browning, 2007).

Hybrids have been produced by crossing each of the 254 dent lines with a tester flint line (UH007). Hybrids were sowed in four trials carried out at the INRAE PhenoArch phenotyping platform (Montpellier, France) (Cabrera-Bosquet et al, 2016; Alvarez Prado et al, 2017; Welcker et al, 2022) in spring 2012, spring 2013, winter 2013, and spring 2016. The four trials followed an alpha-lattice design, including two watering conditions and three replicates. The applied soil water potentials were -0.05 MPa for the well-watered (WW) conditions and ranged from -0.3 to -0.6 MPa, for the water deficit condition (WD).

Phenomics data description

The phenotypic dataset used in this study was previously published by Alvarez Prado et al (2017). It comes from phenotypic measurements taken during plant growth in each of the four above-mentioned trials. This dataset consists of six ecophysiological traits, namely biomass (Biom), leaf area (LA), transpiration rate (Transp), stomatal conductance (gs_max), water uptake (WU), and water use efficiency (WUE). For each trait, watering condition, trial, and hybrid, the average of the three replicates was adjusted by taking into account the spatial greenhouse effect as described in Alvarez Prado et al (2017). For each trial, two phenomics datasets of dimension 254 x 6 were generated, hereafter called WW and WD phenomics datasets. A third dataset, called PL phenomics dataset, was produced from the first two by computing plasticity indices for each hybrid x trial combination, as described in Djabali et al (2023).

Proteomics data description

The proteomics dataset used in this study was previously published by Blein-Nicolas et al (2020). It comes from the proteomics analysis of leaf samples collected on plants of the spring 2012 trial at the pre-flowering stage. This dataset contains the abundance values for 2,055 proteins in 502 hybrids x watering condition combinations (251 out of 254 hybrids could be properly analysed by proteomics). For details about proteomics data acquisition and processing, see Blein-Nicolas et al (2020). Two proteomics datasets of dimension 251 x 2,055 were generated, hereafter called WW and WD proteomics datasets. A third proteomics dataset, called PL proteomics dataset, was produced from the first two by computing plasticity

indices as abundance log^2(fold-change) between the WD and WW conditions for each protein and each hybrid.

QTLs and pQTLs detection

The QTLs used in this study were previously published by Djabali et al (2023). These are 42, 60, and 40 QTLs identified from the WW, WD, and PL phenomics datasets and called WW QTLs, WD QTLs, and PL QTLs, respectively. These QTLs were identified by adapting the single locus mixed model of Yu et al (2006) to take into account trial effects as mentioned in Djabali et al (2023).

Part of the pQTLs used in this study were previously published by Blein-Nicolas et al (2020). The are 10,906 and 11,930 pQTLs identified from the WW and WD proteomics datasets and hereafter called WW pQTLs and WD pQTL, respectively. In this study, we completed this first pQTLs set by identifying PL pQTLs from the PL proteomics dataset. PL pQTLs were identified following the same approach as Blein-Nicolas et al (2020). Briefly, SNPs significantly associated with the variation of the plasticity indices of proteins were first detected by using the single locus mixed model of Yu et al (2006) :

$$Y_g = \mu + \alpha.X_g + \underline{G_g} + \underline{\varepsilon_g} \tag{3.1}$$

where: Y_g is the plasticity index of a protein quantified in genotype g; μ is the overall mean; α is the fixed effect of the SNP allelic dose X_g (coded as 0,1 and 2) for the genotype g; $G_g \sim \mathcal{N}(0, \sigma^2_g.K)$ is the random effect of genotype g, with K the kinship matrix computed with the whole set of SNPs except those located on the same chromosome as the tested SNP (Rincent et al, 2014a), following the approach published by Astle and Balding (2009) and implemented in the R package *statgenGWAS*; $\varepsilon_g \sim \mathcal{N}(0, \sigma^2.I_n)$ is the residual error. This model was run using the Fast-LMM algorithm (Lippert et al, 2011).

Then, the pQTLs were identified from the associated SNPs following the geometric method described in (Blein-Nicolas et al, 2020). After ordering, for each chromosome, SNPs according to their physical position, the $-log(p-value)$ signal was smoothed by computing the maximum $-log(p-value)$ in a sliding window containing N consecutive SNPs. An association peak was detected when the smoothed $-log(p-value)$ signal exceeded a max threshold M. Two consecutive peaks were considered two different pQTLs when the $-log(p-value)$ signal separating them dropped below a minimum threshold m. The parameters for pQTL detection were fixed empirically at $N = 500$, $M = 5$, and $m = 4$.

A colocalization between two QTLs/pQTLs is defined by the overlap of the linkage disequilibrium (LD) windows of these QTLs/pQTLs as described in Negro et al (2019). A QTLs/pQTLs hotspot is characterized as a colocalization regions comprises pQTLs associated with more than twenty different proteins. The start of the hotspot region is defined with the upper bound of the first marker's LD windows at the beginning of the colocalization regions, and the end of the hotspot region is defined with the lower bound of the last marker's LD windows at the end of the colocalization regions.

Multi-trait multi-environment modelling

A multi-trait multi-environment mixed model (van Eeuwijk et al, 2010) with random effects for genotype (G), genotype by water availability interaction $(G \times W)$, genotype by ecophysiological trait interaction $(G \times P)$, genotype by trial interaction $(G \times T)$ was first fitted to estimate the variance components of random effects by the restricted maximum likelihood (REML):

$$\begin{aligned} Y_{gwpt} = \mu + E_{wt} + P_p + (E \times P)_{wtp} + PC_g + (PC \times E)_{gwt} + (PC \times P)_{gp} \\ + \underline{G_g} + \underline{(G \times W)_{gw}} + \underline{(G \times P)_{gp}} + \underline{(G \times T)_{gt}} + \underline{\varepsilon_{gwpt}} \end{aligned} \tag{3.2}$$

where: Y_{gwtf} is the phenotypic value of genotype g in the watering condition w for the ecophysiological trait p in the trial t; μ is the overall mean; E_{wt} is the fixed effect of the environment defined as the combination between the watering condition w and the trial t; P_p is the fixed effect of the ecophysiological trait p; $(E \times P)_{wtp}$ the fixed interaction effects between the environment wt and the ecophysiological trait p; PC_g are coordinates of genotype g projected onto principal component analysis axes built with the kinship matrix K. The number of axes used was chosen following the Kaiser criterion; $(PC \times E)_{gwt}$ are the fixed interaction effects between the genetic structure PC_g and the environment wt; $(PC \times P)_{gp}$ are the fixed interaction effects between the genetic structure PC_g and the ecophysiological trait t; $\underline{\varepsilon_{gwpt}}$ are the residuals of the model.

To simplify mathematical notations, equation 3.2 was referred as the reference model M_0 and rewritten under matrices form:

$$M_0 : Y \sim \mathcal{N}(X\beta, Zu) \tag{3.3}$$

where: $X\beta$ is all fixed effects described in Equation (3.2); Zu is all random effects described in Equation (3.2).

The significance of $\underline{G}$, $\underline{G \times W}$, $\underline{G \times P}$, $\underline{G \times T}$ random effects was tested by comparing the model defined in Equation (3.2) with the same model without random effects.

Multi-trait multi-environment multi-locus modelling

To assess the biological relevance of the detected QTLs and pQTLs, we first fitted a multi-trait, multi-environment multi-locus mixed model (van Eeuwijk et al, 2010) by adding fixed effects of a QTL set $\mathcal{Q}$ (with $\mathcal{Q}$ containing QTLs and/or pQTLs) as in Equation (3.3):

$$Y \sim \mathcal{N}(X\beta + \mathcal{Q}, Zu) \tag{3.4}$$

with $\mathcal{Q} = PC\mathcal{Q}_g + (PC\mathcal{Q} \times E)_{gwt} + (PC\mathcal{Q} \times P)_{gp}$. $PC\mathcal{Q}_g$ is the fixed effect of the QTL set $\mathcal{Q}$. $PC\mathcal{Q}_g$ are coordinates of the genotype g projected onto principal component analysis axes built with the kinship matrices computed with the sets of significant SNPs that describe each QTL of $\mathcal{Q}$. $(PC\mathcal{Q} \times E)_{gwt}$ is the interaction between the genotype g at the QTLs of set $\mathcal{Q}$ and environment wt. $(PC\mathcal{Q} \times P)_{gp}$ is the interaction between the genotype g at the QTLs of set $\mathcal{Q}$ and trait p.

Let $M_{\mathcal{Q}}$ a multi-trait multi-environment multi-locus mixed model defined by the addition of a QTL set $\mathcal{Q}$. The proportion of $G \times W$ variance explained by $\mathcal{Q}$, $\gamma_{\mathcal{Q}}$, was calculated as:

$$\gamma_{\mathcal{Q}} = \frac{\Gamma - \Gamma_{\mathcal{Q}}}{\Gamma}$$

where: Γ is the variance component of the $G \times W$ random effect in Equation (3.2) and $\Gamma_{\mathcal{Q}}$ is the variance component of the $G \times W$ random effect in Equation (3.4) with the QTL set $\mathcal{Q}$.

Gaussian graphical models estimation and coexpression network inference

Gaussian graphical models (GGMs) were used to build networks between proteins and ecophysiological traits. Only the 951 proteins for which WW, WD, and PL pQTLs were detected were kept to avoid the ultra-high dimension phenomenon. Spring 2012 phenomics and proteomics datasets were first merged on genotype identifiers after data scaling. This resulted in three multi-omics datasets (WW, WD, and PL) of dimensions 248 x 957. Then, each multi-omics dataset was transformed with a non-paranormal Gaussianization (Liu et al, 2009) with the function npn() of the *huge* package in R (Team, 2022; Zhao et al, 2012) using the shrunken Empirical Cumulative Distribution Functions (ECDF) method.

Finally, GGMs were estimated by following the graphical lasso (Glasso) approach (Friedman et al, 2008; Mazumder and Hastie, 2012) proposed in the function huge() of the *huge* package in R (Team, 2022; Zhao et al, 2012). A unique, irregular grid Λ composed of a hundred values of regularization parameters λ ranging from 0.09 to 0.99 was generated for the three multi-omics datasets.

The coexpression network inference consisted of choosing one λ from Λ that does not exceed 70 proteins in the traits' 1st-degree neighborhood and maximizes the proportion of GxW variance explained by adding the pQTLs of the proteins in relation to the traits. Let $\gamma(\mathcal{Q})$ and $\gamma(\mathcal{Q}_\lambda)$, the proportions of $G \times W$ variances explained by a set of QTLs $\mathcal{Q}$ which does not contain pQTLs, and a set of QTLs $\mathcal{Q}_\lambda$ which contains pQTLs of proteins in relation to the trait for the regularization parameters λ. The formalized selection criterion can be written as :

$$\underset{\lambda \in \Lambda}{\operatorname{argmax}}\, f = \gamma(\mathcal{Q}_\lambda) - \gamma(\mathcal{Q}) \geq 0 \tag{3.5}$$

Functional classification and analysis of proteins and genes identified

The genes underlying QTLs and hotspots were those located in QTLs LD windows and hotspots interval. They were retrieved from the B73_AGP_v2 gene models annotation .gff3 of the B73 reference assembly. The v3, v4, and v5 gene identifiers were translated sequentially from v2 genes using the Translate Gene Model IDs tools provided by MaizeGDB (https://www.maizegdb.org/gene_center/gene).

Functional annotation of V5 genes underlying QTLs was based on Gene Ontology (GO) from MaizeMine (Shamimuzzaman et al, 2020). Functional enrichment analysis of genes located under a set of QTLs was performed by comparing the relative occurrence of each term to its relative occurrence in the list of genes present in the entire B73 reference genome RefGen_v5 Zm00001eb.1 (Woodhouse et al, 2021) using a hypergeometric test with the R function phyper. P-values were adjusted by the Benjamini-Hochberg (BH) procedure to control the False Discovery Rate (FDR). An enriched term had its adjusted P-value lower than 0.05.

The protein-protein interaction networks (PPI) were retrieved by using Stringdb v11.5 Szklarczyk et al (2021) from the V3 genes identifiers, which code the set of proteins. Functional annotation of proteins was based on KEGG (Kanehisa and Goto, 2000) enrichment analysis provided by Stringdb using the entire maize genome as a reference.

Hotpost importance on the GxW variance captured is equal to the difference between the maximum part of GxW variance captured with the full set of QTLs and pQTLs minus the part of GxW variance captured without the set of QTLs and pQTLs located in a given hotspot.

Network representation and analysis

All network representations and analysis were conducted using Cytoscape v3.10(Shannon et al, 2003). Multi-scaled networks were first built in R with the package *iGraph* and then transferred into a Cytoscape session with the R package *Ryc3* and the function createNetworkFromIgraph(). PPI networks were transferred from Stringdb v11.5 to a Cytoscape session using the stringApp v2.0.1 Cytoscape's app and the "send to Cytoscape" functionality present in the Stringdb v11.5 website platform.

3.3 . Application of our multi-scaled networks inference approach on metabolomics data

In the preceding section, we highlighted how proteomics data has the potential to offer extensive functional evidence for QTLs that capture the majority of the genotype by watering availability variance (GxW) via protein-protein interaction networks. Additionally, we showed that pQTLs can be used as an allele reservoir to explain an additional proportion of the GxW variance. In the subsequent study, Romain Poupon applied the same multi-omics approach but using metabolomics instead of proteomics.

Although metabolomics data can really provide functional clues through metabolism pathways, this particularity of metabolomics data was not been possible to be fully exploited in this thesis because of time constraints, and this data has not yet been published. Thus, we decided to limit the study to the detection of metabolites QTLs (mQTLs) by conducting GWAS on metabolomics data.

The metabolomics data comprised 1,416 secondary metabolite abundances quantified from leaf samples taken on 237 hybrids grown in WW and WD during the spring 2013 trial (Fig. 3.6a). During the first year of my thesis, I performed GWAS on each metabolite's abundance for the two contrasting watering conditions but also on metabolite plasticity indices. In total, 17 174, 16 885, and 4 731 WW, WD, and PL mQTLs were detected, respectively, which spanned 8,422 unique loci (Fig. 3.6b). As proteomics, more than 99% of loci do not overlap with QTLs in $\mathcal{Q}_{ref}$, and the goal of Romain was to apply the multi-scaled networks inference approach presented in the previous section to quantify the contribution of mQTLs on the GxW variance.

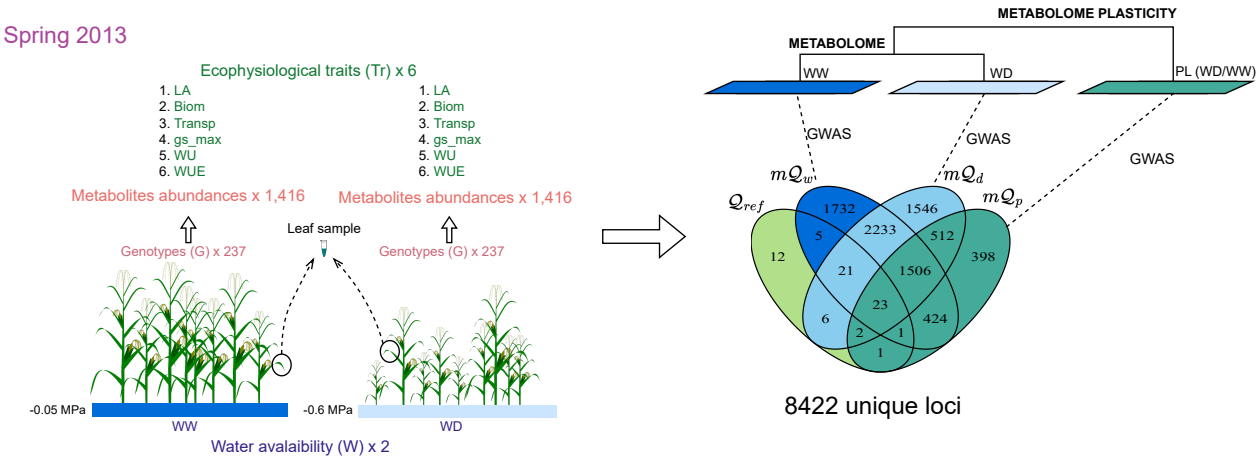


Figure 3.6: **Schematic representation of the metabolomics study design and GWAS related to metabolites abundances.** Metabolite abundances were quantified by gas chromatography coupled to tandem mass spectrometry (GC-MS/MS) from leaf samples taken on plants grown during the spring 2013 trial. Phenotypic plasticities were calculated as the ratio between WD and WW metabolomics data. Then, single-environment GWASs were performed on the WW, WD, and PL metabolomics data. The Venn diagram displays the colocalization between Q_{ref}, mQ_w, mQ_d, and mQ_p.

Romain's work, which is attached as an annex, dealt with the following questions: Do mQTLs that do not overlap with QTLs capture an additional part of GxW? To what extent could multi-scale network inference applied to proteomics data be applied to metabolomics data?

He first showed that mQTLs, as pQTLs, can capture an additional part of the GxW variance. Then, he showed that selecting mQTLs of metabolites most correlated with ecophysiological traits allowed him to capture a better part of the GxW variance than using the full set of mQTLs and plasticity mQTLs detected on all metabolites.

NB: As Romain's report was written before receiving reviewers' comments on the article presented in Chapter 2 and before I finished the study related to the previous section, some jargon terms and slightly methodological differences are present in the report :

- steady-state QTLs $\rightarrow$ QTLs detect on phenotypic means
- Multi-environment GWAS $\rightarrow$ Multi-trial GWAS
- $G \times F$ (Genotype by fluctuating environment interaction) $\rightarrow$ $G \times T$ (Genotype by trial interaction)
- $G \times T$ (Genotype by ecophysiological trait interaction) $\rightarrow$ $G \times P$
- The trait WUE was excluded from the analysis first but re-integrated in section 3.2
- As M5 QTLs was not yet identified as those capturing the maximum part of GxW variance, Romain integrate also the WD QTLs.
- An arbitrary method for choosing λ was first used. It was based on the 5th, 10th, and 20th first metabolites linked to ecophysiological traits in conditional networks.

Bibliography

Acharjee A, Chibon PY, Kloosterman B, et al (2018) Genetical genomics of quality related traits in potato tubers using proteomics. BMC Plant Biology 18(1):20. https://doi.org/10.1186/s12870-018-1229-1

Alvarez Prado S, Cabrera-Bosquet L, Grau A, et al (2017) 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

Astle W, Balding DJ (2009) Population Structure and Cryptic Relatedness in Genetic Association Studies. Statistical Science 24(4):451–471. https://doi.org/10.1214/09-STS307

Blein-Nicolas M, Negro SS, Balliau T, et al (2020) A systems genetics approach reveals environment-dependent associations between SNPs, protein coexpression, and drought-related traits in maize. Genome Research 30(11):1593–1604. https://doi.org/10.1101/gr.255224.119

Boyle EA, Li YI, Pritchard JK (2017) An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169(7):1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Browning SR, Browning BL (2007) Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. The American Journal of Human Genetics 81(5):1084–1097. https://doi.org/10.1086/521987

Cabrera-Bosquet L, Fournier C, Brichet N, et al (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

Cano-Gamez E, Trynka G (2020) From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Frontiers in Genetics 11. https://doi.org/10.3389/fgene.2020.00424

Capriotti E, Ozturk K, Carter H (2019) Integrating molecular networks with genetic variant interpretation for precision medicine. Wiley Interdisciplinary Reviews Systems Biology and Medicine 11(3):e1443. https://doi.org/10.1002/wsbm.1443

Christie N, Myburg AA, Joubert F, et al (2017) Systems genetics reveals a transcriptional network associated with susceptibility in the maize–grey leaf spot pathosystem. The Plant Journal 89(4):746–763. https://doi.org/10.1111/tpj.13419

Civelek M, Lusis AJ (2014) Systems genetics approaches to understand complex traits. Nature Reviews Genetics 15(1):34–48. https://doi.org/10.1038/nrg3575

Cook BI, Mankin JS, Anchukaitis KJ (2018) Climate Change and Drought: From Past to Future. Current Climate Change Reports 4(2):164–179. https://doi.org/10.1007/s40641-018-0093-2

Djabali Y, Rincent R, Martin ML, et al (2023) Drought plasticity QTLs specifically contribute to the genotype x water availability interaction in maize

Fagny M, Paulson JN, Kuijjer ML, et al (2017) Exploring regulation in tissues with eQTL networks. Proceedings of the National Academy of Sciences 114(37):E7841–E7850. https://doi.org/10.1073/pnas.1707375114

Félix MA, Barkoulas M (2015) Pervasive robustness in biological systems. Nature Reviews Genetics 16(8):483–496. https://doi.org/10.1038/nrg3949

Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9(3):432–441. https://doi.org/10.1093/biostatistics/kxm045

Fu J, Keurentjes JJB, Bouwmeester H, et al (2009) System-wide molecular evidence for phenotypic buffering in Arabidopsis. Nature Genetics 41(2):166–167. https://doi.org/10.1038/ng.308

Ganal MW, Durstewitz G, Polley A, et al (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(12):e28,334. https://doi.org/10.1371/journal.pone.0028334

Gibson G, Dworkin I (2004) Uncovering cryptic genetic variation. Nature Reviews Genetics 5(9):681–690. https://doi.org/10.1038/nrg1426

Guo R, Shi L, Jiao Y, et al (2018) Metabolic responses to drought stress in the tissues of drought-tolerant and drought-sensitive wheat genotype seedlings. AoB Plants 10(2):ply016. https://doi.org/10.1093/aobpla/ply016

Hasanuzzaman M, Nahar K, Anee TI, et al (2017) Glutathione in plants: Biosynthesis and physiological role in environmental stress tolerance. Physiology and Molecular Biology of Plants 23(2):249–268. https://doi.org/10.1007/s12298-017-0422-2

Hawe JS, Theis FJ, Heinig M (2019) Inferring Interaction Networks From Multi-Omics Data. Frontiers in Genetics 10. https://doi.org/10.3389/fgene.2019.00535

Intergovernmental Panel on Climate Change (2023) Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 1st edn. Cambridge University Press, https://doi.org/10.1017/9781009157896

Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 28(1):27–30. https://doi.org/10.1093/nar/28.1.27

Kim YM, Han YJ, Hwang OJ, et al (2012) Overexpression of Arabidopsis Translationally Controlled Tumor Protein Gene AtTCTP Enhances Drought Tolerance with Rapid ABA-Induced Stomatal Closure. Molecules and Cells 33(6):617–626. https://doi.org/10.1007/s10059-012-0080-8

Korte A, Vilhjálmsson BJ, Segura V, et al (2012) A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nature Genetics 44(9):1066–1071. https://doi.org/10.1038/ng.2376

Lippert C, Listgarten J, Liu Y, et al (2011) FaST linear mixed models for genomewide association studies. Nat Methods 8:833–835

Liu H, Lafferty J, Wasserman L (2009) The Nonparanormal: Semiparametric Estimation of High Dimensional Undirected Graphs. https://doi.org/10.48550/arXiv.0903.0649, 0903.0649

Manolio TA, Collins FS, Cox NJ, et al (2009) Finding the missing heritability of complex diseases. Nature 461(7265):747–753. https://doi.org/10.1038/nature08494

Mazumder R, Hastie T (2012) The graphical lasso: New insights and alternatives. Electronic Journal of Statistics 6(none). https://doi.org/10.1214/12-EJS740

Millet EJ, Welcker C, Kruijer W, et al (2016) Genome-Wide Analysis of Yield in Europe: Allelic Effects Vary with Drought and Heat Scenarios. Plant Physiology 172(2):749–764. https://doi.org/10.1104/pp.16.00621

Negro SS, Millet EJ, Madur D, et al (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

Porth I, White R, Jaquish B, et al (2012) Genetical Genomics Identifies the Genetic Architecture for Growth and Weevil Resistance in Spruce. PLOS ONE 7(9):e44,397. https://doi.org/10.1371/journal.pone.0044397

Rincent R, Moreau L, Monod H, et al (2014a) 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

Rincent R, Nicolas S, Bouchet S, et al (2014b) Dent and Flint maize diversity panels reveal important genetic potential for increasing biomass production. Theoretical and Applied Genetics 127(11):2313–2331. https://doi.org/10.1007/s00122-014-2379-7

Runcie DE, Qu J, Cheng H, et al (2021) MegaLMM: Mega-scale linear mixed models for genomic predictions with thousands of traits. Genome Biology 22(1):213. https://doi.org/10.1186/s13059-021-02416-w

Sah RP, Chakraborty M, Prasad K, et al (2020) Impact of water deficit stress in maize: Phenology and yield components. Scientific Reports 10(1):2944. https://doi.org/10.1038/s41598-020-59689-7

Schnable P, Ware D, Fulton R, et al (2009) The B73 maize genome: Complexity, diversity, and dynamics. Science 326:1112–5

Shamimuzzaman M, Gardiner JM, Walsh AT, et al (2020) MaizeMine: A Data Mining Warehouse for the Maize Genetics and Genomics Database. Frontiers in Plant Science 11. https://doi.org/10.3389/fpls.2020.592730

Shannon P, Markiel A, Ozier O, et al (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research 13(11):2498–2504. https://doi.org/10.1101/gr.1239303

Shi H, Kichaev G, Pasaniuc B (2016) Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. The American Journal of Human Genetics 99(1):139–153. https://doi.org/10.1016/j.ajhg.2016.05.013

Shutta KH, De Vito R, Scholtens DM, et al (2022) Gaussian graphical models with applications to omics analyses. Statistics in Medicine 41(25):5150–5187. https://doi.org/10.1002/sim.9546

Szklarczyk D, Gable AL, Nastou KC, et al (2021) The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Research 49(D1):D605–D612. https://doi.org/10.1093/nar/gkaa1074

Team RC (2022) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing

Uffelmann E, Huang QQ, Munung NS, et al (2021) Genome-wide association studies. Nature Reviews Methods Primers 1(1):1–21. https://doi.org/10.1038/s43586-021-00056-9

Unterseer S, Bauer E, Haberer G, et al (2014) A powerful tool for genome analysis in maize: Development and evaluation of the high density 600 k SNP genotyping array. BMC Genomics 15(1):823. https://doi.org/10.1186/1471-2164-15-823

van der Sijde MR, Ng A, Fu J (2014) Systems genetics: From GWAS to disease pathways. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1842(10):1903–1909. https://doi.org/10.1016/j.bbadis.2014.04.025

van Eeuwijk FA, Bink MC, Chenu K, et al (2010) Detection and use of QTL for complex traits in multiple environments. Current Opinion in Plant Biology 13(2):193–205. https://doi.org/10/cd33h3

Welcker C, Spencer NA, Turc O, et al (2022) Physiological adaptive traits are a potential allele reservoir for maize genetic progress under challenging conditions. Nature Communications 13(1):3225. https://doi.org/10.1038/s41467-022-30872-w

Woodhouse MR, Cannon EK, Portwood JL, et al (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

Yang J, Benyamin B, McEvoy BP, et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nature Genetics 42(7):565–569. https://doi.org/10.1038/ng.608

Yu J, Pressoir G, Briggs WH, et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38(2):203–208. https://doi.org/10.1038/ng1702

Zhao T, Liu H, Roeder K, et al (2012) The huge Package for High-dimensional Undirected Graph Estimation in R. Journal of machine learning research : JMLR 13:1059–1062

Zi X, Zhou S, Wu B (2022) Alpha-Linolenic Acid Mediates Diverse Drought Responses in Maize (Zea mays L.) at Seedling and Flowering Stages. Molecules 27(3):771. https://doi.org/10.3390/molecules27030771

Chapter 4

General discussion

4.1 . Standfirst

During these three years, the overall goal of my PhD was to gain a better understanding of the genetic and molecular basis of drought tolerance in maize. To achieve this goal, I had to develop interdisciplinary approaches at the interface of quantitative genetics, biostatistics and data science. Thanks to the Amazing project, I had a unique multi-omics dataset, including genomics, proteomics, metabolomics, and phenomics data measured in 254 maize genotypes, which allowed me to address my objectives holistically. However, the study of the genotype-phenotype relationship (GP) for a trait as complex as drought tolerance was very challenging for me. Therefore, to cope with the complexity of studying drought tolerance, I decided to divide my research into two main axes, corresponding to Chapters 2 and 3.

In Chapter 2, my research focused on the identification of genetic determinants involved in the genotype-by-environment (GxE) interaction and, in particular, the genotype-by-water availability (GxW) interaction. Indeed, since drought tolerance is expressed until the plant perceives water stress, the identification of QTLs contributing to the GxW is of particular interest, especially if they are associated with drought-responsive morphophysiological traits. Therefore, I decided to investigate the genetic architecture associated with the plasticity of six drought-responsive traits by performing GWAS on plasticity indices. The results obtained showed that the detected plasticity QTL exclusively captured 60 to 100% of the GxW variance for almost all the traits studied. Moreover, the plasticity QTLs did not overlap with the QTLs detected on the phenotypic means of the traits obtained in the two irrigation conditions. My results encourage the consideration of plasticity measurements in association studies to detect consistent QTLs involved in GxE that cannot be identified by studying phenotypic means alone. Furthermore, the low overlap between plasticity QTLs and QTLs associated with the phenotypic means of traits supports the gene regulatory model as the main model for the genetic control of phenotypic plasticity. This regulatory model proposes that plasticity results from the interaction between regulatory and structural genes (Bradshaw, 1965; Kusmec et al, 2017; Diouf et al, 2020).

In Chapter 3, I proceed through the study of the GP in regard to drought tolerance, taking into account the results obtained in Chapter 2 and using the proteomics data. I conducted a systems genetics approach to decipher the genetic and molecular basis of the drought response. My approach was based on the inference of molecular networks as a function of plant water status but also plant plasticity, integrating genomics, proteomics and phenomics data. Using an approach that combined network inference and estimation of the contribution of pQTLs to GxW variance, I was able to construct a multi-scale network that included 63 proteins directly associated with traits and trait plasticity, and a total of 531 loci, 86% of which contained only pQTLs. The addition of proteomics data was beneficial for i) providing a comprehensive annotation of genetic regions enriched in pQTLs through the hotspots PPIs, and ii) unraveling the missing heritability of drought response by considering pQTLs in the multi-locus multi-trait multi-environment linear mixed models. Indeed, characterizing hotspots as PPIs highlighted experimentally verified physical interactions between proteins encoded by genes located in the hotspots and proteins associated with pQTLs located in the hotspots. This approach is relatively powerful and demonstrates how proteomics can be used to link genetic variants in molecular pathways. Moreover, the addition of pQTLs in the multi-locus, multi-environment mixed models allows the proportion of GxW variance captured to be increased to 20 points, demonstrating the power of using intermediate molecular traits in association studies to highlight specific genetic variants and unravel the missing heritability of complex traits. This last result was also validated with the integration of metabolomics. Overall, the results presented in this last chapter strongly support the consideration of omics data to bridge the gap between genotype and phenotype for complex traits. Although the omics data used were measured in only one trial, I have shown that they can contribute to a detailed genetic

characterization of the drought response and open up new avenues for their use in breeding for the selection of new drought-tolerant varieties.

Although systems genetics studies are still scarce for understanding the genetic determinism of plants related to drought response, they represent one of the most methodological advances for unraveling the omnigenic determinism of complex traits (Wu, 2021). Conducting such studies is complex and requires interdisciplinary skills. At present, systems genetics is an emerging approach, and there is no "gold standard" method. In Chapter 3, I proposed a novel approach to performs system genetics and the first part of my discussion will be about the limitations, advantages, and disadvantages of the method. Furthermore, this research was motivated by the increasing concern about climate change and its impact on maize production. To address this issue, the aim of this thesis was to provide insight into the genetic basis of drought tolerance for varietal improvement. In the second part of my discussion, I will ask how the results presented in this thesis could be used or exploited to progress in the identification of drought-tolerant maize varieties. Finally, I close this herein manuscript by enumerating the perspectives of my PhD work and giving a personal conclusion.

4.2 . Original methods for conducting systems genetics studies

Understanding the flow of biological information that underlies complex traits is the main goal of systems genetics (Civelek and Lusis, 2014). This approach aims to identify all molecular interactions and their underlying genetic polymorphisms that may influence a phenotypic trait. Systems genetics requires the ability to interpret results from the integration of different types of omics data. However, multi-omics data integration is still a challenging task with many possible methodological approaches (Ritchie et al, 2015). In Chapter 3, I present a systems genetics study by following two original steps of data integration that allow the functional and quantitative exploitation of omics data. The first one is based on the identification of colocalization between QTLs and pQTLs, and the second one is based on the combination of multi-scaled network inference and estimation of QTLs contribution on the genetic variance of traits.

Identification of colocalization between QTLs and molecular QTLs is the most straightforward approach to identifying molecular actors involved in complex trait variations. Indeed, this approach is based on the identification of genetic polymorphisms associated with both complex trait variation and molecular trait variation (Giambartolomei et al, 2014). Thus, to identify such shared genetic polymorphisms between molecular and complex phenotypic traits, several approaches can be used. Indeed, Nica et al (2010) proposed an empirical approach based on the consideration of the linkage disequilibrium (LD) structure around the most associated SNPs resulting in "colocalized" genomic regions. Whereas, Giambartolomei et al (2014), was more interested in detecting "the" SNP significantly associated with both complex and molecular trait variations through hypothesis testing. The approach conducted in Chapter 3, is analogous to the method developed by Nica et al (2010), because a co-localization between a QTL and pQTL was defined by the overlapping of their physical LD windows. Although this approach is not as specific as the one proposed by Giambartolomei et al (2014), it allowed us to be less conservative and consider genetic regions containing several polymorphisms that could interact with each other or be in LD. Moreover, by following this approach, as Blein-Nicolas et al (2020), I was able to pinpoint genomic regions enriched in pQTLs called hotspots. These hotspots are of particular interest because we can hypothesize that the polymorphisms located in the hotspot could have an effect on the abundance of several proteins involved in a common biological pathway. To be in line with this hypothesis in Chapter 3, I proposed to take advantage that the molecular traits were proteins to investigate protein-protein interaction networks (PPIs) on the basis of their colocalized genomic regions. This original manner to functionally analyze pQTL hotspots allowed us to highlight several experimentally verified interactions between proteins encoded by genes located in the hotspots and proteins having pQTLs located in the hotspot. By conducting this approach on hotpots comprising QTLs detected on ecophysiological traits, we were able to simultaneously provide a functional annotation and to identify potential molecular networks underlying these traits. However, one of the major drawbacks of integrative analysis based on colocalization between QTLs and molecular QTLs is that it is restricted to genetic regions

detectable with GWAS conducted on the traits. Indeed, GWAS are based on statistical tests and may suffer from a lack of power despite the number of genotypes considered (Manolio et al, 2009). To tackle this problem, we investigated directly the associations between the complex phenotypic traits and the molecular traits by using Gaussian graphical models (GGMs).

Gaussian graphical models (GGMs) allow the inference of conditional dependencies between a set of variables, assuming that their distribution is a multivariate Gaussian distribution (Altenbuchinger et al, 2020). The major advantage of GGMs is the generation of a conditional dependency network where the edges between variables correspond to partial correlations (Baba et al, 2004). In contrast to classical pairwise association measures such as Pearson's correlation, partial correlation discriminates between direct and indirect effects by estimating the strength of the relationship between two variables while accounting for the effects of the other variables (Hawe et al, 2019). Thus, conditional dependency networks generated by GGMs are sparser than correlation-based networks and provide a straightforward interpretation in terms of conditional dependence given the state of the measured biological system (Shutta et al, 2022). In Chapter 3, we applied GGMs to infer multi-scale networks, including both proteins and ecophysiological traits as nodes. In our case, we restricted the powerfulness of GGMs on the ability to find trait-protein interactions, which is a tool that overcomes our goal. Knowing the complexity of applying GGMs on datasets presenting high dimensionality, perhaps a more straightforward method, such as proteome-wide association study (PWAS) (Brandes et al, 2020), would be simpler and more relevant. Indeed, one must pay attention to the problem of high dimensionality when using GGMs. The parameter of the GGMs is estimated through the optimization of a regularized function. As a consequence, we have to follow the GLasso procedure (Mazumder and Hastie, 2012), which requires choosing a regularization parameter that does not conduct to a network with too many edges. Several choices of mathematical criteria exist to select the most relevant regularization parameter, such as the extended Bayesian information criteria (eBIC), and the stability approach for regularization selection (StARS) (Liu et al, 2010). These criteria are based on mathematical considerations to be sure that the estimated network has good statistical properties. In Chapter 3, we used an original method of selection based on the combination of empirical and biological prior-driven criteria. Indeed, we selected the regularization parameter that did not exceed 70 proteins in the traits' 1st-degree neighborhood and maximized the proportion of GxW interaction variance explained by adding the pQTLs of the proteins in relation to the traits. To be able to apply this criterion, I combined GGMs with the multi-locus multi-trait multi-environment linear mixed models (van Eeuwijk et al, 2010). The great strength of this method is the inference of multi-scale networks, including both the genetic and molecular actors, explaining specifically the maximum part of a targeted component of the phenotypic variance.

4.3 . Putting my results into a plant breeding perspective

Optimizing plasticity in selection could be a solution for exploiting GxE and keeping yields high and stable even in case of environmental fluctuations (Kusmec et al, 2018; Monforte, 2020). As Kusmec et al (2017) and Diouf et al (2020), our results presented in Chapter 2 support the gene-regulatory model (Scheiner, 1993; Via et al, 1995) as the genetic control underlying phenotypic plasticity. In this model, phenotypic plasticity results from epistasis between regulator genes and structural genes. As a consequence, as Bradshaw (1965) stipulated, plasticity can be considered an an independent trait with its own genetic determinism. The validity of this hypothesis through association studies (Kusmec et al, 2017; Diouf et al, 2020; Djabali et al, 2023) suggests that phenotypic means and phenotypic plasticity of traits could be selected independently. Two types of phenotypic plasticity exist: active and passive plasticity (Brooker et al, 2022). Active plasticity is the phenotypic changes induced by the activation of physiological or molecular mechanisms in response to environmental changes. Passive plasticity corresponds to phenotypic changes induced by the environment but not linked to physiological or molecular response mechanisms (*e.g.*, resource shortage limiting plant growth). Improving plant adaptability in response to environmental change requires being able to make a selection on active plasticity. However, as GxE is displayed by both passive and active plasticity, the first challenge is to be able to differentiate the part of GxE variance induced by the two types of plasticity (Brooker et al, 2022). To tackle this, the experimental design and the methodological approach described in Chapter 2 could be viewed as a way to remove the most passive component of plasticity. Indeed, first, the study of plasticity was carried out on ecophysiological traits known to be responsive to drought. Sec-

ond, the plasticity indices used were specific to changes induced by drought (WD/WW ratios), and third, plasticity QTLs were identified by considering the residual environmental trial effect on GWAS model. All these steps contribute to being more focused on the active part of plasticity induced by water deficit. The second challenge is to select plasticity conferring an agronomic advantage regarding environmental changes. Indeed, the goal of the breeding programs is to keep high and stable yields. It will be interesting to understand how the QTLs identified affect the plasticity of the traits studied and how these modifications could affect yields and yield stability. To do that, it will be necessary to cross the results presented in Chapter 2 with yield measurements made on the same diversity panel. Finally, Welcker et al (2022) addressed an interesting question about the margin of progress existing in the exploitation of QTLs detected on the ecophysiological traits studied in Chapter 2 for selection. Indeed, we can wonder if the selection of adaptive traits is still possible regarding the past breeding programs. Indeed, we can imagine that in the current elite varieties obtained from the selection on yield, indirectly, the selection on adaptive was already made. Thus, Welcker et al (2022) showed in their study evidence that past breeding programs on yields did not affect adaptive traits, and a margin of progress exists by exploiting alleles affecting the adaptive plasticity.

In line with the study of Welcker et al (2022), it will be possible that a margin of progress in varietal improvement exists not only by exploiting QTLs in favor of adaptive traits but also on molecular traits. Thus, an omic-driven selection could give the possibility to make a fine-tuned varietal improvement by selecting in the scale of biological pathways.

According to Shen et al (2022), the recent progress in the generation of multi-omics datasets will lead to interdisciplinary studies that could be favorable for another breakthrough innovation in plant breedings. Indeed, where the first complex traits genetic dissection with GWAS showed a limit (Manolio et al, 2009; Yang et al, 2010; Shi et al, 2016), the addition of several omics layers could give enough statistical power to reveal new candidates genes and molecular pathways underlying traits. The different omics layers can be treated as causal factors for complex traits but also as intermediate traits to identify additional genetic variants. Estimation of interactions between all actors, as we did in Chapter 3 with our systems genetics study, leads to modeled complex traits of genetic determinism as molecular networks. Such models could support the identification of candidate genes for selection by identifying genetic and molecular targets in the networks. For instance, in Chapter 3, we were able to weight hotspots according to their importance for capturing GxW interaction variance. Thus, based on these results, we can provide a list of candidate genes by exploiting the PPIs of the most important hotspots. The importance of genetic and molecular targets can also be measured through their capacity to improve phenotypic prediction. However, the interdisciplinary of several scientific fields, such as mathematics, biology, chemistry, computer science, or even physics, is crucial to making this type of breeding practicable.

4.4 . Perspectives

My Ph.D. thesis showed the benefit of using omics data to gain insight into the genetic and molecular basis underlying complex traits. Indeed, the addition of proteomics data allowed us to go deeper into the genetic determinism related to maize water stress response by highlighting important genetic regions and providing a comprehensive functional annotation. Considering the results obtained with proteomics, we can imagine that adding other types of omics data could be beneficial for the understanding of the genetic determinism of maize water stress response. To this end, the analysis made on metabolomics data can be integrated with those obtained with proteomics to provide a more complete view of the molecular pathways underlying water stress response. Moreover, transcriptomics data (about 20,000 quantified transcripts) were also generated on the same diversity panel. The addition of the transcriptomics data could be beneficial by considerably increasing the number of potential molecular actors through the identification of eQTLs. Finally, we could further enrich the multi-omics data by estimating metabolic fluxes. Indeed, the study by Petrizzelli (2019) in yeast showed that protein abundances are good predictors of metabolic fluxes when integrated into a genome-scale metabolic model. Thus, metabolic fluxes could be estimated by integrating the leaf metabolic model of maize Seaver et al (2015) with proteomics data. Similarly to the other omics data, flux QTLs could be detected and used to provide information on the potential metabolic reactions linked to water stress response.

The studies I made during my Ph.D. has exclusively an explanation goal. I identified significant associations between biological features and ecophysiological traits. A very popular front of research in plant breeding is genomic selection (Crossa et al, 2017). Genomic selection is based on the ability of genotypic data to predict trait values. One of the most popular genomics selection models is Genomic Best Linear Unbiased Prediction (GBLUP) (Kärkkäinen and Sillanpää, 2013; Clark and van der Werf, 2013). This method uses genomic relationships to estimate an individual's genetic contribution to a trait. This is done using a genomic relationship matrix estimated from DNA marker information (also called kinship matrix). The kinship matrix defines the covariance between individuals based on observed similarity at the genomic level, resulting in accurate predictions of the genetic contributions. As this method includes all the information contained in the whole set of DNA markers, this method does not use knowledge of the biological mechanisms underlying trait variation. Thus, Sarup et al (2016) developed Genomic Feature Best Linear Unbiased Prediction (GFBLUP), an extension of GBLUP, that can assess the collective effects of sets of DNA markers. Our results presented in Chapter 3 allowed us to highlight a set of QTLs, pQTLs and proteins involved in GxW variance. It will be interesting to see to what extent these biological features can improve phenotype prediction for traits measured under water-deficit conditions.

Finally, it would be interesting to functionally validate candidate genes identified from the results obtained in Chapter 3, for example, by starting on genes located in the most important hotspots. This would give more confidence and legitimacy to our *in silico* approaches and perhaps interest more collaborators.

Bibliography

Altenbuchinger M, Weihs A, Quackenbush J, et al (2020) Gaussian and Mixed Graphical Models as (multi-)omics data analysis tools. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1863(6):194,418. https://doi.org/10.1016/j.bbagrm.2019.194418

Baba K, Shibata R, Sibuya M (2004) Partial Correlation and Conditional Correlation as Measures of Conditional Independence. Australian & New Zealand Journal of Statistics 46(4):657–664. https://doi.org/10.1111/j.1467-842X.2004.00360.x

Blein-Nicolas M, Negro SS, Balliau T, et al (2020) A systems genetics approach reveals environment-dependent associations between SNPs, protein coexpression, and drought-related traits in maize. Genome Research 30(11):1593–1604. https://doi.org/10.1101/gr.255224.119

Bradshaw AD (1965) Evolutionary Significance of Phenotypic Plasticity in Plants. In: Caspari EW, Thoday JM (eds) Advances in Genetics, vol 13. Academic Press, p 115–155, https://doi.org/10.1016/S0065-2660(08)60048-6

Brandes N, Linial N, Linial M (2020) PWAS: Proteome-wide association study—linking genes and phenotypes by functional variation in proteins. Genome Biology 21(1):173. https://doi.org/10.1186/s13059-020-02089-x

Brooker R, Brown LK, George TS, et al (2022) Active and adaptive plasticity in a changing climate. Trends in Plant Science 27(7):717–728. https://doi.org/10.1016/j.tplants.2022.02.004

Civelek M, Lusis AJ (2014) Systems genetics approaches to understand complex traits. Nature Reviews Genetics 15(1):34–48. https://doi.org/10.1038/nrg3575

Clark SA, van der Werf J (2013) Genomic Best Linear Unbiased Prediction (gBLUP) for the Estimation of Genomic Breeding Values. In: Gondro C, van der Werf J, Hayes B (eds) Genome-Wide Association Studies and Genomic Prediction, vol 1019. Humana Press, Totowa, NJ, p 321–330, https://doi.org/10.1007/978-1-62703-447-0_13

Crossa J, Pérez-Rodríguez P, Cuevas J, et al (2017) Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends in Plant Science 22(11):961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Diouf I, Derivot L, Koussevitzky S, et al (2020) Genetic basis of phenotypic plasticity and genotype × environment interactions in a multi-parental tomato population. Journal of Experimental Botany 71(18):5365–5376. https://doi.org/10.1093/jxb/eraa265

Djabali Y, Rincent R, Martin ML, et al (2023) Drought plasticity QTLs specifically contribute to the genotype x water availability interaction in maize

Giambartolomei C, Vukcevic D, Schadt EE, et al (2014) Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLOS Genetics 10(5):e1004,383. https://doi.org/10.1371/journal.pgen.1004383

Hawe JS, Theis FJ, Heinig M (2019) Inferring Interaction Networks From Multi-Omics Data. Frontiers in Genetics 10. https://doi.org/10.3389/fgene.2019.00535

Kärkkäinen HP, Sillanpää MJ (2013) Fast Genomic Predictions via Bayesian G-BLUP and Multilocus Models of Threshold Traits Including Censored Gaussian Data. G3: Genes|Genomes|Genetics 3(9):1511–1523. https://doi.org/10.1534/g3.113.007096

Kusmec A, Srinivasan S, Nettleton D, et al (2017) Distinct genetic architectures for phenotype means and plasticities in Zea mays. Nature Plants 3(9):715–723. https://doi.org/10.1038/s41477-017-0007-7

Kusmec A, de Leon N, Schnable PS (2018) Harnessing Phenotypic Plasticity to Improve Maize Yields. Frontiers in Plant Science 9. https://doi.org/10.3389/fpls.2018.01377

Liu H, Roeder K, Wasserman L (2010) Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models. Advances in neural information processing systems 24(2):1432–1440

Manolio TA, Collins FS, Cox NJ, et al (2009) Finding the missing heritability of complex diseases. Nature 461(7265):747–753. https://doi.org/10.1038/nature08494

Mazumder R, Hastie T (2012) The graphical lasso: New insights and alternatives. Electronic Journal of Statistics 6(none). https://doi.org/10.1214/12-EJS740

Monforte AJ (2020) Time to exploit phenotypic plasticity. Journal of Experimental Botany 71(18):5295–5297. https://doi.org/10.1093/jxb/eraa268

Nica AC, Montgomery SB, Dimas AS, et al (2010) Candidate Causal Regulatory Effects by Integration of Expression QTLs with Complex Trait Genetic Associations. PLOS Genetics 6(4):e1000,895. https://doi.org/10.1371/journal.pgen.1000895

Petrizzelli M (2019) Mathematical modelling and integration of complex biological data : Analysis of the heterosis phenomenon in yeast. PhD thesis, Université Paris Saclay (COmUE)

Ritchie MD, Holzinger ER, Li R, et al (2015) Methods of integrating data to uncover genotype–phenotype interactions. Nature Reviews Genetics 16(2):85–97. https://doi.org/10.1038/nrg3868

Sarup P, Jensen J, Ostersen T, et al (2016) Increased prediction accuracy using a genomic feature model including prior information on quantitative trait locus regions in purebred Danish Duroc pigs. BMC Genetics 17:11. https://doi.org/10.1186/s12863-015-0322-9

Scheiner SM (1993) Genetics and Evolution of Phenotypic Plasticity. Annual Review of Ecology and Systematics 24:35–68. https://doi.org/10.1146/annurev.es.24.110193.000343, https://arxiv.org/abs/2097172

Seaver SMD, Bradbury LMT, Frelin O, et al (2015) Improved evidence-based genome-scale metabolic models for maize leaf, embryo, and endosperm. Frontiers in Plant Science 6. https://doi.org/10.3389/fpls.2015.00142

Shen Y, Zhou G, Liang C, et al (2022) Omics-based interdisciplinarity is accelerating plant breeding. Current Opinion in Plant Biology 66:102,167. https://doi.org/10.1016/j.pbi.2021.102167

Shi H, Kichaev G, Pasaniuc B (2016) Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. The American Journal of Human Genetics 99(1):139–153. https://doi.org/10.1016/j.ajhg.2016.05.013

Shutta KH, De Vito R, Scholtens DM, et al (2022) Gaussian graphical models with applications to omics analyses. Statistics in Medicine 41(25):5150–5187. https://doi.org/10.1002/sim.9546

van Eeuwijk FA, Bink MC, Chenu K, et al (2010) Detection and use of QTL for complex traits in multiple environments. Current Opinion in Plant Biology 13(2):193–205. https://doi.org/10/cd33h3

Via S, Gomulkiewicz R, De Jong G, et al (1995) Adaptive phenotypic plasticity: Consensus and controversy. Trends in Ecology & Evolution 10(5):212–217. https://doi.org/10.1016/S0169-5347(00)89061-8

Welcker C, Spencer NA, Turc O, et al (2022) Physiological adaptive traits are a potential allele reservoir for maize genetic progress under challenging conditions. Nature Communications 13(1):3225. https://doi.org/10.1038/s41467-022-30872-w

Wu R (2021) Specialty Grand Challenge: Systems Genetics. Frontiers in Systems Biology 1. https://doi.org/10.3389/fsysb.2021.738155

Yang J, Benyamin B, McEvoy BP, et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nature Genetics 42(7):565–569. https://doi.org/10.1038/ng.608

Appendices

Suplementary material Chapter 2

Supplementary figures and tables related to section 2.2

Fig. S1: Correlogram of Pearson's correlations obtained for WW phenotypic means

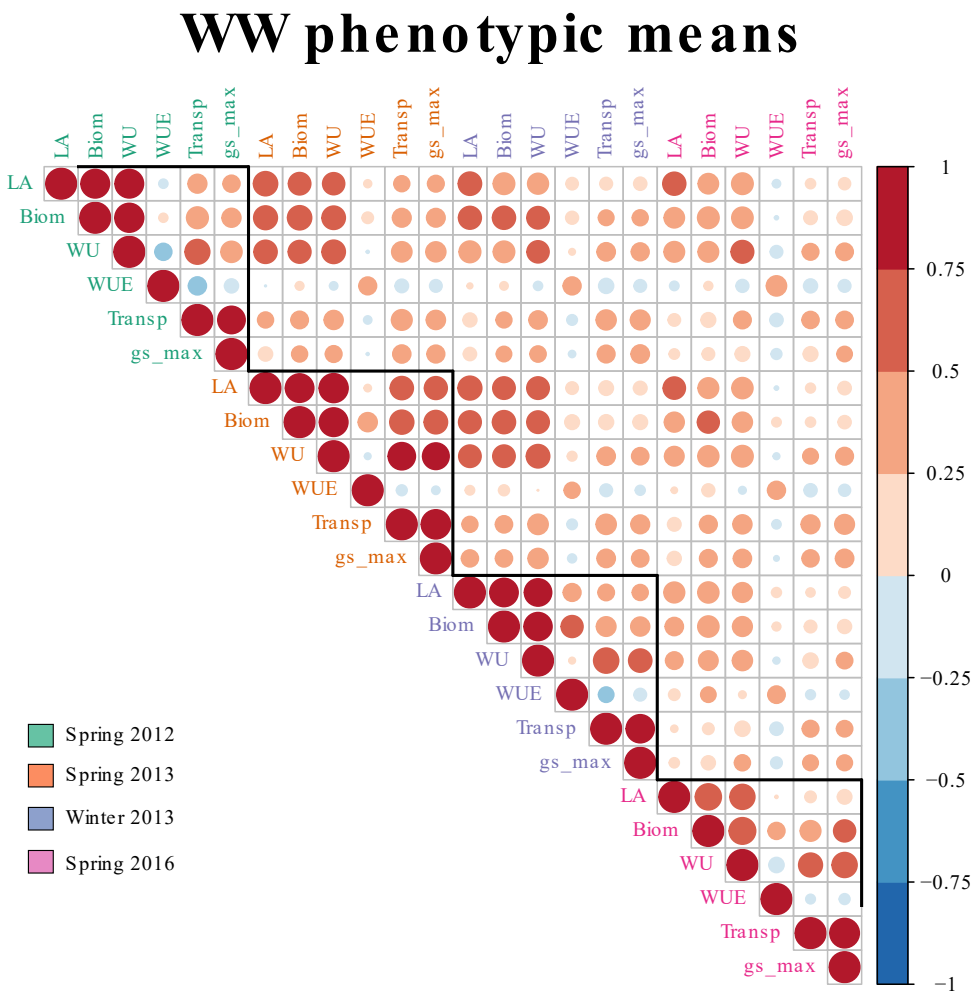


Fig. S2 : Comparison between QTLs detected by single-trial and multi-trial GWAS for LA, Biom, Transp and gs_max.

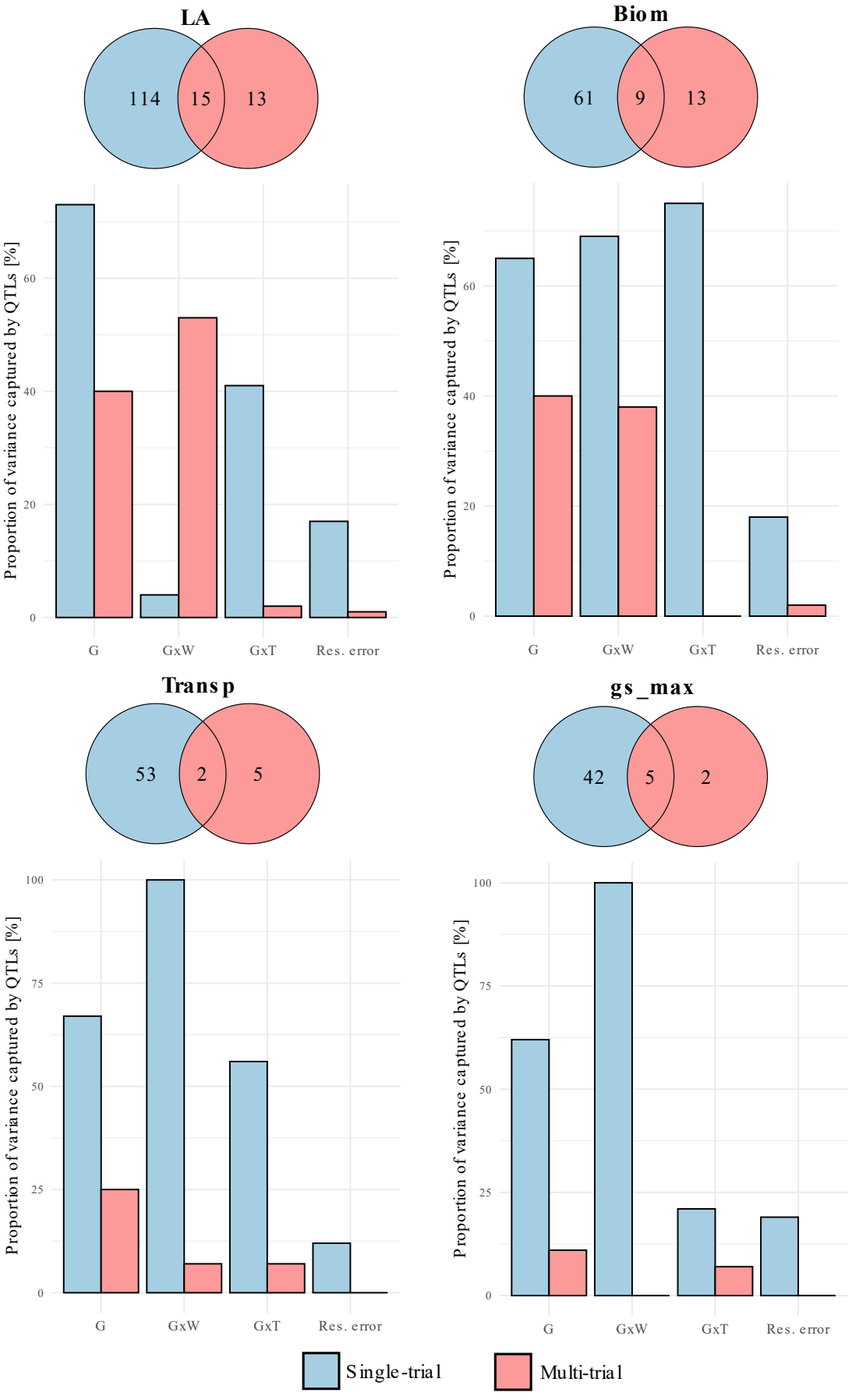


Fig. S3: Comparison between QTLs and plasticity QTLs detected from multi-trial GWAS for LA, Biom, Transp and gs_max.

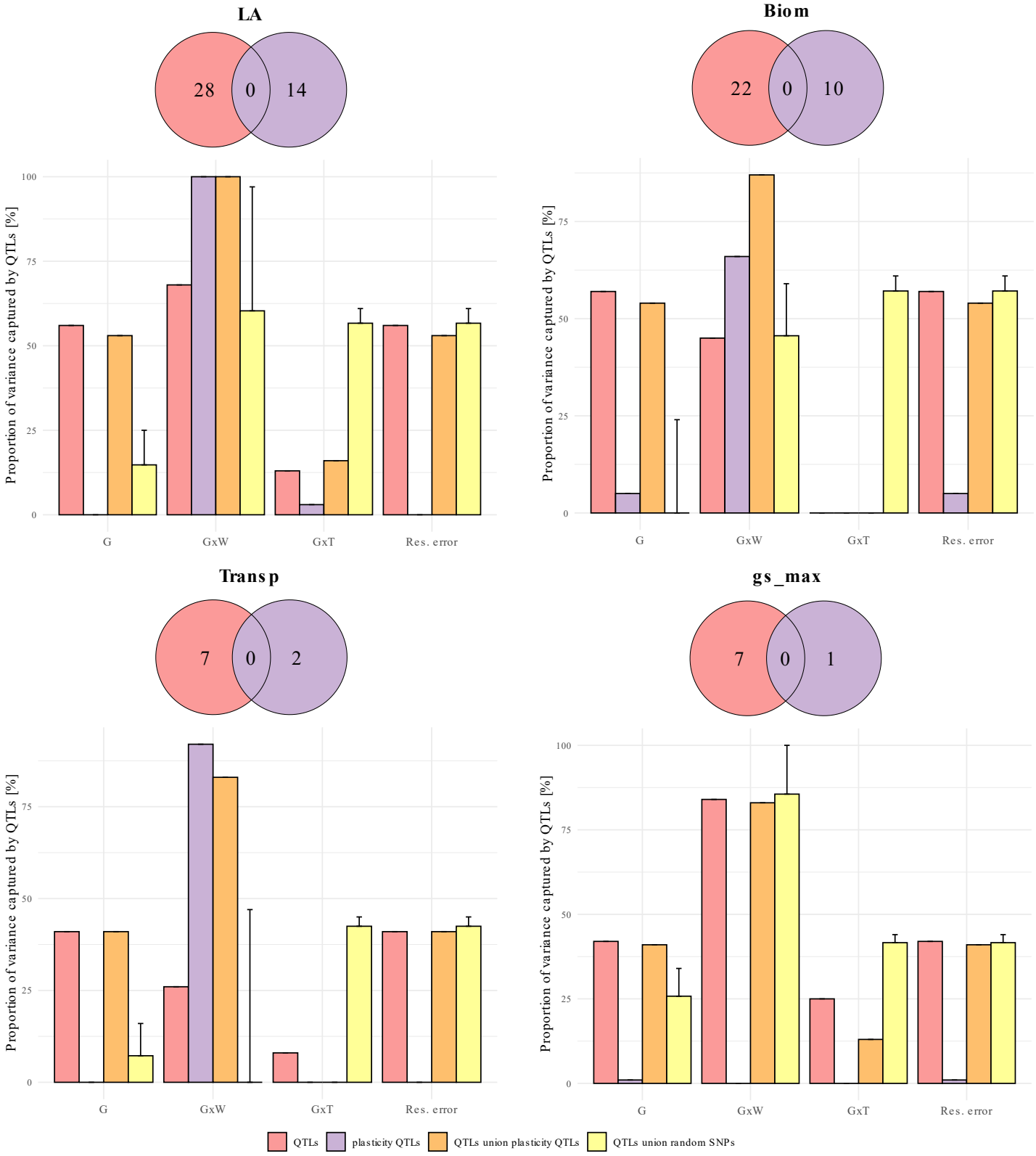


Table S1 : List of WD, WW plasticity QTLs detected on the studied traits

SNP.name	Chromosome	Position	Trait	LogPvalue	AllelicEffect	Type
AX-90609193	3	223747793	LA	5.30447111857145	0.028	Phenotypic mean (WD)
AX-90617893	5	204172495	LA	5.36227729649731	-0.025	Phenotypic mean (WD)
AX-90620016	7	125405619	WU	5.35049094378828	-0.127	Phenotypic mean (WD)
AX-90620016	7	125405619	Biom	5.1717593112174	-5.066	Phenotypic mean (WD)
AX-90643100	5	12223981	Biom	5.96783602943874	15.274	Phenotypic mean (WD)
AX-90705520	1	213766976	WUE	5.62776146703696	4.705	Phenotypic mean (WD)
AX-90705525	1	213738640	WUE	5.62776146703696	4.705	Phenotypic mean (WD)
AX-90829472	3	137350920	Biom	5.42229427079208	13.085	Phenotypic mean (WD)
AX-91054365	7	125235348	Transp	5.10262684824561	-0.204	Phenotypic mean (WD)
AX-91213807	3	36468765	gs_max	5.16323521990897	-0.966	Phenotypic mean (WD)
AX-91223642	5	90761008	WU	5.28526446140019	-0.057	Phenotypic mean (WD)
AX-91375072	2	55679719	WUE	5.55637896245588	4.573	Phenotypic mean (WD)
AX-91436202	5	12215772	LA	6.96189083792789	0.009	Phenotypic mean (WD)
AX-91436202	5	12215772	WU	5.57771496779547	0.112	Phenotypic mean (WD)
AX-91563636	3	54324627	gs_max	5.30767541041884	1.45	Phenotypic mean (WD)
AX-91594151	3	217358140	LA	5.50716939702046	0.022	Phenotypic mean (WD)
AX-91594185	3	217565112	LA	5.00535614499821	-0.023	Phenotypic mean (WD)
AX-91644584	5	14902572	Biom	5.06295415465214	16.223	Phenotypic mean (WD)
AX-91644700	5	15544264	WU	6.42986532588306	0.218	Phenotypic mean (WD)
AX-91644700	5	15544264	Biom	6.23575920209394	14.818	Phenotypic mean (WD)
AX-91644700	5	15544264	LA	6.05463994077937	0.018	Phenotypic mean (WD)
AX-91676735	5	189909534	LA	5.31760194480248	0.026	Phenotypic mean (WD)
AX-91775135	8	160381339	LA	5.06386789078191	-0.014	Phenotypic mean (WD)
AX-91839077	7	125474566	WU	5.04331129822773	-0.123	Phenotypic mean (WD)
PZE-102095227	2	108996680	gs_max	5.30710689543462	-2.021	Phenotypic mean (WD)
PZE-102116673	2	155082333	Transp	5.75871436661091	-0.068	Phenotypic mean (WD)
PZE-105073586	5	79851377	WU	5.01246877513412	-0.349	Phenotypic mean (WD)
S1_213917744	1	213917744	WUE	5.3366641162674	4.587	Phenotypic mean (WD)
S2_155504254	2	155504254	Transp	5.48249161735794	-0.533	Phenotypic mean (WD)
S2_203521616	2	203521616	WUE	5.25547569252917	0.24	Phenotypic mean (WD)
S2_55446463	2	55446463	WUE	5.65716925052053	4.564	Phenotypic mean (WD)
S2_85996553	2	85996553	Transp	5.12489775885214	0.643	Phenotypic mean (WD)
S3_179124720	3	179124720	WUE	5.22118149537481	4.51	Phenotypic mean (WD)
S3_223987133	3	223987133	Biom	5.14719467967165	11.634	Phenotypic mean (WD)
S3_224153021	3	224153021	LA	5.26509965487187	0.031	Phenotypic mean (WD)
S4_18916660	4	18916660	LA	5.35143241987487	-0.017	Phenotypic mean (WD)
S4_18916736	4	18916736	Biom	5.9530294672701	-14.196	Phenotypic mean (WD)
S4_23073199	4	23073199	WUE	5.02123318254539	2.682	Phenotypic mean (WD)
S4_7304050	4	7304050	Biom	5.6182216304582	-8.648	Phenotypic mean (WD)
S5_150934135	5	150934135	LA	5.04067380323106	-0.018	Phenotypic mean (WD)
S5_15455288	5	15455288	Biom	6.38041022406441	15.548	Phenotypic mean (WD)
S5_15455288	5	15455288	WU	5.59902990196014	0.208	Phenotypic mean (WD)
S5_162058939	5	162058939	LA	5.00825515686147	-0.021	Phenotypic mean (WD)
S5_163120356	5	163120356	LA	5.32225216478473	-0.018	Phenotypic mean (WD)
S5_16428195	5	16428195	WU	5.40578642247247	0.015	Phenotypic mean (WD)
S5_16428195	5	16428195	Biom	5.83627616591926	5.21	Phenotypic mean (WD)
S5_199288741	5	199288741	LA	6.21260876116658	0.014	Phenotypic mean (WD)
S5_199288744	5	199288744	WU	5.60401825898215	0.082	Phenotypic mean (WD)
S5_201691474	5	201691474	Biom	5.03122399610096	-0.173	Phenotypic mean (WD)
S5_80280650	5	80280650	LA	5.83621900995694	-0.047	Phenotypic mean (WD)
S5_88791868	5	88791868	Transp	6.08945535276797	-0.592	Phenotypic mean (WD)

S7_121192950	7	121192950	Biom	5.26986109582027	-5.433	Phenotypic mean (WD)
S7_125348551	7	125348551	Transp	5.34597680112097	-0.225	Phenotypic mean (WD)
S7_125348551	7	125348551	gs_max	5.17538408646068	-2.612	Phenotypic mean (WD)
S7_125348551	7	125348551	WU	5.52994308431201	-0.132	Phenotypic mean (WD)
S8_123323084	8	123323084	LA	5.27991838387333	0.031	Phenotypic mean (WD)
S8_124644939	8	124644939	Biom	5.10617729247593	22.442	Phenotypic mean (WD)
S8_147276130	8	147276130	WU	5.00379074917294	0.092	Phenotypic mean (WD)
S9_139985370	9	139985370	LA	5.07084051243007	0.022	Phenotypic mean (WD)
S9_20467522	9	20467522	WUE	5.44355860655464	0.558	Phenotypic mean (WD)
AX-90591657	9	11498756	WUE	5.03634440153841	-0.046	Phenotypic mean (WW)
AX-90617893	5	204172495	LA	5.57542313187026	-0.033	Phenotypic mean (WW)
AX-90703311	1	205531946	gs_max	5.21738448697405	-3.662	Phenotypic mean (WW)
AX-90795567	3	6520929	LA	5.06127109847071	-0.038	Phenotypic mean (WW)
AX-90904002	4	183600566	WUE	5.64088159461552	3.389	Phenotypic mean (WW)
AX-90970356	5	190229458	LA	5.87351584110113	0.038	Phenotypic mean (WW)
AX-90984392	6	28147127	gs_max	5.18235386497565	-3.321	Phenotypic mean (WW)
AX-91048276	7	101578020	WU	5.08476133911149	0.158	Phenotypic mean (WW)
AX-91217929	4	17282972	LA	5.19479735012346	-0.018	Phenotypic mean (WW)
AX-91436202	5	12215772	LA	5.79588422197806	0.001	Phenotypic mean (WW)
AX-91436202	5	12215772	WU	5.30394556457377	-0.027	Phenotypic mean (WW)
AX-91436202	5	12215772	Biom	5.64641195781851	9.817	Phenotypic mean (WW)
AX-91529215	2	107020947	gs_max	5.72290353598346	-6.212	Phenotypic mean (WW)
AX-91577233	3	125732326	WUE	5.16866756435054	2.334	Phenotypic mean (WW)
AX-91645061	5	17478458	LA	5.01161606219548	-0.004	Phenotypic mean (WW)
AX-91676735	5	189909534	LA	6.73855395017276	0.045	Phenotypic mean (WW)
AX-91678360	5	198818515	Biom	5.15928694517604	30.612	Phenotypic mean (WW)
AX-91735439	7	122811242	WU	5.13273234883376	0.268	Phenotypic mean (WW)
S1_21713786	1	21713786	WUE	5.25973490034444	-1.556	Phenotypic mean (WW)
S2_110086525	2	110086525	Transp	5.12286706015051	-0.758	Phenotypic mean (WW)
S3_171437400	3	171437400	WUE	6.88896255175746	7.014	Phenotypic mean (WW)
S3_6098394	3	6098394	LA	5.26454617353703	-0.032	Phenotypic mean (WW)
S4_16979538	4	16979538	WU	5.2908890289459	0.286	Phenotypic mean (WW)
S4_23073199	4	23073199	WUE	5.57584186144312	3.129	Phenotypic mean (WW)
S4_236976065	4	236976065	LA	5.21829845206766	-0.038	Phenotypic mean (WW)
S4_236976065	4	236976065	Biom	5.66746475498211	-42.081	Phenotypic mean (WW)
S4_237611655	4	237611655	Biom	5.00334774780448	-7.883	Phenotypic mean (WW)
S5_16428195	5	16428195	LA	5.15267497029268	0.005	Phenotypic mean (WW)
S5_184347132	5	184347132	WU	5.04301241957428	-0.228	Phenotypic mean (WW)
S5_190042338	5	190042338	LA	5.93598336653756	0.026	Phenotypic mean (WW)
S5_199208460	5	199208460	Biom	5.18980955712004	9.823	Phenotypic mean (WW)
S5_199976003	5	199976003	Biom	5.11661064145599	-21.258	Phenotypic mean (WW)
S5_199978118	5	199978118	LA	5.03092129298598	-0.021	Phenotypic mean (WW)
S5_200117524	5	200117524	Biom	5.21413085462961	-4.148	Phenotypic mean (WW)
S5_202298957	5	202298957	LA	5.14085777735796	-0.032	Phenotypic mean (WW)
S7_100476901	7	100476901	WU	5.28906251079118	-0.02	Phenotypic mean (WW)
S7_122921991	7	122921991	Biom	5.14753166836813	-38.527	Phenotypic mean (WW)
S7_122921991	7	122921991	WU	5.34514716018255	-0.195	Phenotypic mean (WW)
S7_128740287	7	128740287	Transp	5.20499467447904	0.184	Phenotypic mean (WW)
S7_156270246	7	156270246	Biom	5.42563378954195	40.023	Phenotypic mean (WW)
S7_161348225	7	161348225	Biom	5.06088811246865	28.112	Phenotypic mean (WW)
S7_8177978	7	8177978	LA	5.03816703948701	0.055	Phenotypic mean (WW)
AX-90600662	1	79997930	LA plasticity	5.13724869419477	0.028	Plasticity
AX-90655484	1	20115983	Biom plasticity	5.12373079190998	0.038	Plasticity
AX-90655531	1	20528160	Biom plasticity	5.14368635887928	0.038	Plasticity
AX-90655539	1	20634797	Biom plasticity	5.18699866714987	-0.005	Plasticity

AX-90658062	1	29790479	Transp plasticity	5.05968638908001	0.009	Plasticity
AX-90659369	1	34663629	WU plasticity	5.3955205503559	0.038	Plasticity
AX-90660998	1	40421472	LA plasticity	5.4784060084063	-0.014	Plasticity
AX-90673821	1	88324138	LA plasticity	5.2460071241887	0.01	Plasticity
AX-90686976	1	141532437	LA plasticity	5.05033250573513	-0.007	Plasticity
AX-90698746	1	187643875	LA plasticity	5.26379587077178	0.007	Plasticity
AX-90798967	3	19796584	Biom plasticity	5.63240870000244	-0.041	Plasticity
AX-90833610	3	153768601	WU plasticity	5.39214991244562	0.071	Plasticity
AX-90857499	4	7305553	gs_max plasticity	5.52017842153297	0.086	Plasticity
AX-90857499	4	7305553	Transp plasticity	5.89390290456927	0.028	Plasticity
AX-90998624	6	84786567	WU plasticity	5.17150245273568	0.036	Plasticity
AX-91380622	1	20218168	Biom plasticity	5.14368635887928	0.038	Plasticity
AX-91411976	1	53151661	LA plasticity	5.69227045129072	-0.05	Plasticity
AX-91447668	1	172020935	LA plasticity	5.08015828766336	0.053	Plasticity
AX-91457267	1	20280276	Biom plasticity	5.31438779799967	-0.048	Plasticity
AX-91457270	1	20382187	Biom plasticity	5.14368635887928	0.038	Plasticity
AX-91458327	1	26333618	WU plasticity	5.08702686502756	-0.051	Plasticity
AX-91458327	1	26333618	Biom plasticity	5.82375518394043	-0.052	Plasticity
AX-91461738	1	44735660	LA plasticity	5.43123254068514	0	Plasticity
AX-91577322	3	126445639	Biom plasticity	6.17386324371565	0.059	Plasticity
PZE-101089042	1	80680115	LA plasticity	5.97338998007367	-0.022	Plasticity
PZE-108089331	8	146380069	WUE plasticity	5.18967259403665	-0.045	Plasticity
PZE-110070241	10	127247580	Biom plasticity	5.11940757332394	0.025	Plasticity
S1_190702732	1	190702732	LA plasticity	6.12549564007344	-0.043	Plasticity
S1_192453061	1	192453061	LA plasticity	5.08269539001494	-0.047	Plasticity
S1_25414680	1	25414680	Biom plasticity	5.45895030379211	-0.055	Plasticity
S1_25717036	1	25717036	WU plasticity	5.33377582252341	-0.054	Plasticity
S1_26437971	1	26437971	WU plasticity	5.66406731026335	-0.063	Plasticity
S1_32839020	1	32839020	LA plasticity	5.22305002714836	-0.038	Plasticity
S1_52765910	1	52765910	WU plasticity	5.24502360267145	-0.049	Plasticity
S1_52922759	1	52922759	LA plasticity	5.04809081814444	-0.018	Plasticity
S1_77411465	1	77411465	LA plasticity	5.59622163656739	-0.055	Plasticity
S2_221753995	2	221753995	WUE plasticity	5.17253779102185	0.023	Plasticity
S4_180444282	4	180444282	WUE plasticity	5.92861201754642	0.044	Plasticity
S9_136840702	9	136840702	LA plasticity	5.57138776978068	0.033	Plasticity
SYN5704	4	181726408	WUE plasticity	5.13759734752962	-0.031	Plasticity

Table S2 : Trait variance components captured by single-trial and multi-trial GWAS according each random effect

Model	Trait	G	GxW	GxT	Residual error
Control (no QTLs)	LA	0.000539977032044924	2.68765961589647e-05	0.000108153237154901	0.000490398087550201
Single-trial	LA	0.000143503703233023	2.57871971390649e-05	6.36230516555747e-05	0.000407640374801496
Multi-trial	LA	0.000326458674696746	1.27551287473551e-05	0.000106207146214463	0.000484577151895041
Control (no QTLs)	Biom	484.045110680815	104.181418819259	31.8939326255071	581.942890271104
Single-trial	Biom	170.342920922118	31.7894620773865	7.98417259757167	474.973451101123
Multi-trial	Biom	292.471992200434	65.0252772931579	32.7259349125706	572.853108500744
Control (no QTLs)	WU	0.0846528958941729	0.00943251329467889	0.0170260277369968	0.0997286722145494
Single-trial	WU	0.0387619751214422	0.00579235610812599	0.00729447298868563	0.0916932982609004
Multi-trial	WU	0.058774832644887	0.00447412824885227	0.0136795837783194	0.100766843948894
Control (no QTLs)	WUE	6.48108106199427	1.66247148787375	4.96197652104105	21.6412359282072
Single-trial	WUE	5.96012005446033	1.77244371741201	4.65413989321912	20.4116196566939
Multi-trial	WUE	2.42549546232121	0.477987002818191	4.15417293152019	21.2973645802754
Control (no QTLs)	Transp	0.234289573175583	0.00715361788886021	0.0546680090972899	0.49567477118405
Single-trial	Transp	0.077210666144033	2.02180755745129e-09	0.0242972201615929	0.436403828442915
Multi-trial	Transp	0.174981178983491	0.00662864490812369	0.0507658615075576	0.495902476360516
Control (no QTLs)	gs_max	9.0257057730995	0.610324495960822	2.79801938173239	28.8834441943793
Single-trial	gs_max	3.43897760810407	0	2.22384670646459	23.4962861002188
Multi-trial	gs_max	8.06213014429586	0.641728713885699	2.60014432014139	28.8509420970091

Table S3 : Trait variance components captured by QTLs and plasticity QTLs according each random effect

Model	Trait	G	GxW	GxT	Residual error
Control (no QTLs)	LA	0.000539977032217229	2.68765960019925e-05	0.000108153237270391	0.000539977032217229
QTLs	LA	0.000239551884594056	8.51462388703905e-06	9.44640273868594e-05	0.000239551884594056
plasticity QTLs	LA	0.000563553143509639	0	0.000104592422091411	0.000563553143509639
QTLs union plasticity QTLs	LA	0.000251784326637073	0	9.13465116019307e-05	0.000251784326637073
Control (no QTLs)	Biom	484.045109890096	104.181419186009	31.8939330617905	484.045109890096
QTLs	Biom	206.095951770734	56.7875360792459	37.870742696219	206.095951770734
plasticity QTLs	Biom	460.007222162509	35.3039177328937	32.8339461340679	460.007222162509
QTLs union plasticity QTLs	Biom	221.006979304091	13.3214126800512	35.2215145293967	221.006979304091
Control (no QTLs)	WU	0.0846528959561735	0.00943251324484503	0.0170260277018061	0.0846528959561735
QTLs	WU	0.0537533051977796	0.00356302853548683	0.01172705212541	0.0537533051977796
plasticity QTLs	WU	0.0868822947329495	0.00178979108262136	0.0159256719025715	0.0868822947329495
QTLs union plasticity QTLs	WU	0.0561378869780664	0	0.0121824385416397	0.0561378869780664
Control (no QTLs)	WUE	6.48108107672992	1.66247148684198	4.96197649541997	6.48108107672992
QTLs	WUE	2.42549545563629	0.477986990870019	4.15417291701369	2.42549545563629
plasticity QTLs	WUE	7.31533167626109	4.22619000904989e-09	4.92510653852224	7.31533167626109
QTLs union plasticity QTLs	WUE	2.76285944857558	0	4.38857503375215	2.76285944857558
Control (no QTLs)	Transp	0.234289573195733	0.00715361785570571	0.0546680088814755	0.234289573195733
QTLs	Transp	0.137218042732734	0.00529237510308876	0.0503153297139926	0.137218042732734
plasticity QTLs	Transp	0.235098901925917	0.000548197861639742	0.0597454933944119	0.235098901925917
QTLs union plasticity QTLs	Transp	0.137968117581682	0.00122693008433319	0.0558456814122731	0.137968117581682
Control (no QTLs)	gs_max	9.02570577137586	0.610324499192428	2.79801937352893	9.02570577137586
QTLs	gs_max	5.26719673512929	0.0962822636919747	2.08438386570734	5.26719673512929
plasticity QTLs	gs_max	8.93765024665016	0.610660194895303	2.79621178372672	8.93765024665016
QTLs union plasticity QTLs	gs_max	5.31836105353881	0.103698841841343	2.43280698167131	5.31836105353881

Table S4 : Genes located into QTL and plasticity QTL disequilibrium windows

Chr	Database	type	Start	Stop	Strand	Sens	Frame	ID
1	gramene	plasticity QTLs	79984693	79986558	.	+	.	Zm00001d029620
1	gramene	plasticity QTLs	20105976	20131237	.	-	.	Zm00001d028007
1	gramene	plasticity QTLs	20133912	20136948	.	-	.	Zm00001d028008
1	gramene	plasticity QTLs	20533271	20539413	.	+	.	Zm00001d028020
1	gramene	plasticity QTLs	20540461	20546803	.	-	.	Zm00001d028021
1	gramene	plasticity QTLs	20604487	20614451	.	+	.	Zm00001d028022
1	gramene	plasticity QTLs	29770419	29781683	.	+	.	Zm00001d028302
1	gramene	plasticity QTLs	29785411	29788053	.	-	.	Zm00001d028303
1	gramene	plasticity QTLs	29830012	29832203	.	-	.	Zm00001d028304
1	gramene	plasticity QTLs	29833586	29835464	.	-	.	Zm00001d028305
1	gramene	plasticity QTLs	34508764	34510317	.	+	.	Zm00001d028422
1	gramene	plasticity QTLs	34540138	34541979	.	+	.	Zm00001d028423
1	gramene	plasticity QTLs	34542493	34546543	.	+	.	Zm00001d028424
1	gramene	plasticity QTLs	34589914	34593828	.	+	.	Zm00001d028425
1	gramene	plasticity QTLs	34649020	34655072	.	+	.	Zm00001d028426
1	gramene	plasticity QTLs	34657783	34659222	.	+	.	Zm00001d028427
1	gramene	plasticity QTLs	34690471	34704954	.	-	.	Zm00001d028428
1	gramene	plasticity QTLs	34709436	34713166	.	-	.	Zm00001d028429
1	gramene	plasticity QTLs	34818314	34833657	.	+	.	Zm00001d028431
1	gramene	plasticity QTLs	34899450	34900187	.	-	.	Zm00001d028432
1	gramene	plasticity QTLs	34919688	34921312	.	-	.	Zm00001d028433
1	gramene	plasticity QTLs	34926403	34927063	.	+	.	Zm00001d028434
1	gramene	plasticity QTLs	34959814	34966593	.	+	.	Zm00001d028436
1	gramene	plasticity QTLs	40117200	40118564	.	+	.	Zm00001d028591
1	gramene	plasticity QTLs	40123313	40123759	.	+	.	Zm00001d028592
1	gramene	plasticity QTLs	40162847	40168804	.	+	.	Zm00001d028593
1	gramene	plasticity QTLs	40221978	40223031	.	+	.	Zm00001d028594
1	gramene	plasticity QTLs	40238228	40239306	.	+	.	Zm00001d028595
1	gramene	plasticity QTLs	40240031	40250055	.	+	.	Zm00001d028596
1	gramene	plasticity QTLs	40253344	40260838	.	+	.	Zm00001d028597
1	gramene	plasticity QTLs	40268286	40268585	.	-	.	Zm00001d028598
1	gramene	plasticity QTLs	40283742	40287068	.	-	.	Zm00001d028599
1	gramene	plasticity QTLs	40428759	40431019	.	+	.	Zm00001d028601
1	gramene	plasticity QTLs	40510526	40516015	.	-	.	Zm00001d028603
1	gramene	plasticity QTLs	40561826	40563132	.	+	.	Zm00001d028604
1	gramene	plasticity QTLs	40564067	40565906	.	+	.	Zm00001d028605
1	gramene	plasticity QTLs	40594616	40599175	.	+	.	Zm00001d028606
1	gramene	plasticity QTLs	40713008	40718506	.	+	.	Zm00001d028608
1	gramene	plasticity QTLs	88322356	88323479	.	+	.	Zm00001d029816
1	gramene	plasticity QTLs	141705734	141709496	.	+	.	Zm00001d030526
1	gramene	plasticity QTLs	187563113	187564209	.	+	.	Zm00001d031370
1	gramene	plasticity QTLs	187621387	187624780	.	-	.	Zm00001d031372
1	gramene	plasticity QTLs	187625250	187625966	.	-	.	Zm00001d031373
1	gramene	plasticity QTLs	187631492	187632082	.	+	.	Zm00001d031375
1	gramene	plasticity QTLs	187718812	187721184	.	-	.	Zm00001d031377
3	gramene	plasticity QTLs	19801735	19803372	.	+	.	Zm00001d039927
3	gramene	plasticity QTLs	19804533	19805393	.	+	.	Zm00001d039928
3	gramene	plasticity QTLs	19809738	19810514	.	+	.	Zm00001d039929
3	gramene	plasticity QTLs	19815106	19816391	.	+	.	Zm00001d039930
3	gramene	plasticity QTLs	19818598	19829515	.	+	.	Zm00001d039931
3	gramene	plasticity QTLs	19837328	19837744	.	+	.	Zm00001d039932
3	gramene	plasticity QTLs	19846081	19846545	.	-	.	Zm00001d039933

6	gramene	plasticity QTLs	84773604	84774833	.	-	.	Zm00001d036339
6	gramene	plasticity QTLs	84776706	84778799	.	+	.	Zm00001d036340
1	gramene	plasticity QTLs	20219977	20230817	.	-	.	Zm00001d028009
1	gramene	plasticity QTLs	52850182	52853358	.	+	.	Zm00001d028955
1	gramene	plasticity QTLs	52860772	52861856	.	+	.	Zm00001d028957
1	gramene	plasticity QTLs	52873169	52879776	.	-	.	Zm00001d028958
1	gramene	plasticity QTLs	52938630	52945326	.	-	.	Zm00001d028960
1	gramene	plasticity QTLs	53030837	53032126	.	+	.	Zm00001d028962
1	gramene	plasticity QTLs	53142425	53144911	.	+	.	Zm00001d028963
1	gramene	plasticity QTLs	53145602	53154854	.	+	.	Zm00001d028964
1	gramene	plasticity QTLs	53169397	53176921	.	-	.	Zm00001d028965
1	gramene	plasticity QTLs	53177235	53181012	.	+	.	Zm00001d028966
1	gramene	plasticity QTLs	53183379	53188438	.	+	.	Zm00001d028967
1	gramene	plasticity QTLs	53190337	53190834	.	-	.	Zm00001d028968
1	gramene	plasticity QTLs	53292747	53300706	.	+	.	Zm00001d028971
1	gramene	plasticity QTLs	53302218	53303096	.	-	.	Zm00001d028972
1	gramene	plasticity QTLs	53309064	53309600	.	+	.	Zm00001d028973
1	gramene	plasticity QTLs	53368438	53370381	.	+	.	Zm00001d028974
1	gramene	plasticity QTLs	53393407	53405238	.	-	.	Zm00001d028975
1	gramene	plasticity QTLs	53415528	53416657	.	-	.	Zm00001d028976
1	gramene	plasticity QTLs	53417301	53424648	.	-	.	Zm00001d028977
1	gramene	plasticity QTLs	53451859	53454849	.	+	.	Zm00001d028980
1	gramene	plasticity QTLs	172056749	172058475	.	-	.	Zm00001d030984
1	gramene	plasticity QTLs	172065913	172071656	.	-	.	Zm00001d030985
1	gramene	plasticity QTLs	20263677	20268913	.	-	.	Zm00001d028010
1	gramene	plasticity QTLs	20291775	20293058	.	-	.	Zm00001d028011
1	gramene	plasticity QTLs	20300349	20303168	.	+	.	Zm00001d028012
1	gramene	plasticity QTLs	20304479	20310840	.	-	.	Zm00001d028013
1	gramene	plasticity QTLs	20351878	20352705	.	+	.	Zm00001d028015
1	gramene	plasticity QTLs	20354036	20360460	.	+	.	Zm00001d028016
1	gramene	plasticity QTLs	20410911	20411813	.	-	.	Zm00001d028017
1	gramene	plasticity QTLs	26334264	26346936	.	+	.	Zm00001d028207
1	gramene	plasticity QTLs	44512777	44518048	.	+	.	Zm00001d028727
1	gramene	plasticity QTLs	44521225	44532226	.	+	.	Zm00001d028728
1	gramene	plasticity QTLs	44532354	44536303	.	+	.	Zm00001d028730
1	gramene	plasticity QTLs	44536542	44539529	.	-	.	Zm00001d028731
1	gramene	plasticity QTLs	44548793	44552517	.	-	.	Zm00001d028732
1	gramene	plasticity QTLs	44554395	44560803	.	+	.	Zm00001d028733
1	gramene	plasticity QTLs	44718501	44719536	.	+	.	Zm00001d028735
1	gramene	plasticity QTLs	44721407	44722183	.	+	.	Zm00001d028736
1	gramene	plasticity QTLs	44733389	44733898	.	-	.	Zm00001d028737
1	gramene	plasticity QTLs	44765885	44768411	.	+	.	Zm00001d028738
1	gramene	plasticity QTLs	44836475	44837783	.	+	.	Zm00001d028739
1	gramene	plasticity QTLs	44843630	44844552	.	-	.	Zm00001d028740
1	gramene	plasticity QTLs	44865682	44866604	.	-	.	Zm00001d028741
1	gramene	plasticity QTLs	44975070	44978277	.	+	.	Zm00001d028742
3	gramene	plasticity QTLs	125608671	125609400	.	+	.	Zm00001d041520
3	gramene	plasticity QTLs	125708192	125710214	.	-	.	Zm00001d041521
3	gramene	plasticity QTLs	125843356	125846615	.	-	.	Zm00001d041522
3	gramene	plasticity QTLs	125849221	125851417	.	+	.	Zm00001d041523
3	gramene	plasticity QTLs	125901674	125905287	.	-	.	Zm00001d041525
3	gramene	plasticity QTLs	125967179	125969813	.	-	.	Zm00001d041526
3	gramene	plasticity QTLs	125971928	125972875	.	-	.	Zm00001d041528
3	gramene	plasticity QTLs	126000149	126004313	.	-	.	Zm00001d041529
3	gramene	plasticity QTLs	126035560	126036633	.	+	.	Zm00001d041530

3	gramene	plasticity QTLs	126088939	126092644	.	+	.	Zm00001d041531
3	gramene	plasticity QTLs	126104284	126105794	.	-	.	Zm00001d041533
3	gramene	plasticity QTLs	126106707	126109981	.	+	.	Zm00001d041534
3	gramene	plasticity QTLs	126110008	126110211	.	-	.	Zm00001d041535
3	gramene	plasticity QTLs	126130758	126132890	.	-	.	Zm00001d041536
3	gramene	plasticity QTLs	126166316	126167305	.	-	.	Zm00001d041537
3	gramene	plasticity QTLs	126304770	126305192	.	-	.	Zm00001d041538
3	gramene	plasticity QTLs	126342753	126344912	.	+	.	Zm00001d041539
3	gramene	plasticity QTLs	126429213	126430962	.	+	.	Zm00001d041541
3	gramene	plasticity QTLs	126522185	126523196	.	-	.	Zm00001d041544
3	gramene	plasticity QTLs	126535818	126539525	.	-	.	Zm00001d041545
3	gramene	plasticity QTLs	126536801	126539525	.	+	.	Zm00001d041546
3	gramene	plasticity QTLs	126563754	126564497	.	+	.	Zm00001d041547
3	gramene	plasticity QTLs	126577794	126583834	.	+	.	Zm00001d041548
3	gramene	plasticity QTLs	126665476	126667691	.	-	.	Zm00001d041549
3	gramene	plasticity QTLs	126862145	126865908	.	-	.	Zm00001d041550
3	gramene	plasticity QTLs	126876629	126878600	.	+	.	Zm00001d041551
3	gramene	plasticity QTLs	126926038	126926802	.	-	.	Zm00001d041553
3	gramene	plasticity QTLs	127112005	127118515	.	-	.	Zm00001d041556
3	gramene	plasticity QTLs	127251972	127252855	.	+	.	Zm00001d041558
3	gramene	plasticity QTLs	127255197	127260823	.	+	.	Zm00001d041559
3	gramene	plasticity QTLs	127262273	127263420	.	-	.	Zm00001d041560
3	gramene	plasticity QTLs	127265528	127267097	.	+	.	Zm00001d041561
3	gramene	plasticity QTLs	127274051	127280589	.	+	.	Zm00001d041563
3	gramene	plasticity QTLs	127280924	127282074	.	-	.	Zm00001d041564
3	gramene	plasticity QTLs	127284443	127285976	.	+	.	Zm00001d041565
1	gramene	plasticity QTLs	80665246	80665596	.	-	.	Zm00001d029630
1	gramene	plasticity QTLs	80672733	80673116	.	-	.	Zm00001d029631
8	gramene	plasticity QTLs	145974288	145978521	.	+	.	Zm00001d011287
8	gramene	plasticity QTLs	145979013	145985394	.	+	.	Zm00001d011288
8	gramene	plasticity QTLs	145986281	145987233	.	-	.	Zm00001d011289
8	gramene	plasticity QTLs	145988735	145990140	.	-	.	Zm00001d011290
8	gramene	plasticity QTLs	146006899	146007294	.	-	.	Zm00001d011291
8	gramene	plasticity QTLs	146006924	146007241	.	+	.	Zm00001d011292
8	gramene	plasticity QTLs	146062014	146075281	.	-	.	Zm00001d011294
8	gramene	plasticity QTLs	146093892	146095667	.	+	.	Zm00001d011297
8	gramene	plasticity QTLs	146108765	146113773	.	+	.	Zm00001d011298
8	gramene	plasticity QTLs	146128641	146131503	.	+	.	Zm00001d011299
8	gramene	plasticity QTLs	146136810	146145088	.	-	.	Zm00001d011300
8	gramene	plasticity QTLs	146174056	146176755	.	+	.	Zm00001d011301
8	gramene	plasticity QTLs	146180242	146181231	.	-	.	Zm00001d011302
8	gramene	plasticity QTLs	146189282	146189740	.	+	.	Zm00001d011303
8	gramene	plasticity QTLs	146192052	146196828	.	+	.	Zm00001d011304
8	gramene	plasticity QTLs	146317656	146322089	.	+	.	Zm00001d011308
8	gramene	plasticity QTLs	146353906	146364466	.	-	.	Zm00001d011309
8	gramene	plasticity QTLs	146409219	146411054	.	+	.	Zm00001d011311
8	gramene	plasticity QTLs	146414508	146420244	.	-	.	Zm00001d011312
8	gramene	plasticity QTLs	146490050	146490819	.	+	.	Zm00001d011314
8	gramene	plasticity QTLs	146513037	146520523	.	-	.	Zm00001d011315
8	gramene	plasticity QTLs	146523359	146524518	.	-	.	Zm00001d011316
8	gramene	plasticity QTLs	146546991	146554204	.	-	.	Zm00001d011319
8	gramene	plasticity QTLs	146619391	146622999	.	+	.	Zm00001d011321
8	gramene	plasticity QTLs	146669804	146671210	.	+	.	Zm00001d011323
8	gramene	plasticity QTLs	146675491	146676993	.	-	.	Zm00001d011324
8	gramene	plasticity QTLs	146690622	146694044	.	-	.	Zm00001d011325

8	gramene	plasticity QTLs	146725502	146726798	.	-	.	Zm00001d011326
8	gramene	plasticity QTLs	146781789	146783051	.	+	.	Zm00001d011327
8	gramene	plasticity QTLs	146855836	146859841	.	-	.	Zm00001d011328
8	gramene	plasticity QTLs	146863198	146868117	.	-	.	Zm00001d011329
10	gramene	plasticity QTLs	126527371	126529992	.	+	.	Zm00001d025694
10	gramene	plasticity QTLs	126530158	126532242	.	-	.	Zm00001d025695
10	gramene	plasticity QTLs	126532005	126537113	.	+	.	Zm00001d025696
10	gramene	plasticity QTLs	126601226	126609111	.	-	.	Zm00001d025697
10	gramene	plasticity QTLs	126604796	126605766	.	+	.	Zm00001d025698
10	gramene	plasticity QTLs	126708804	126710162	.	-	.	Zm00001d025699
10	gramene	plasticity QTLs	126772727	126775435	.	-	.	Zm00001d025700
10	gramene	plasticity QTLs	126776580	126778346	.	-	.	Zm00001d025701
10	gramene	plasticity QTLs	126789244	126793900	.	+	.	Zm00001d025703
10	gramene	plasticity QTLs	126794136	126795161	.	-	.	Zm00001d025704
10	gramene	plasticity QTLs	126856805	126878792	.	-	.	Zm00001d025705
10	gramene	plasticity QTLs	126880549	126888843	.	-	.	Zm00001d025706
10	gramene	plasticity QTLs	126903970	126906191	.	+	.	Zm00001d025707
10	gramene	plasticity QTLs	126927635	126928758	.	-	.	Zm00001d025709
10	gramene	plasticity QTLs	126946639	126951605	.	+	.	Zm00001d025710
10	gramene	plasticity QTLs	126996928	126997912	.	-	.	Zm00001d025711
10	gramene	plasticity QTLs	126998464	127006041	.	-	.	Zm00001d025712
10	gramene	plasticity QTLs	127154778	127159945	.	-	.	Zm00001d025713
10	gramene	plasticity QTLs	127160289	127165558	.	-	.	Zm00001d025714
10	gramene	plasticity QTLs	127294441	127299500	.	+	.	Zm00001d025715
10	gramene	plasticity QTLs	127318068	127321356	.	+	.	Zm00001d025716
10	gramene	plasticity QTLs	127360555	127364032	.	-	.	Zm00001d025717
10	gramene	plasticity QTLs	127367089	127369910	.	-	.	Zm00001d025718
10	gramene	plasticity QTLs	127374340	127375910	.	-	.	Zm00001d025719
10	gramene	plasticity QTLs	127515419	127518791	.	+	.	Zm00001d025720
10	gramene	plasticity QTLs	127518803	127520726	.	-	.	Zm00001d025721
10	gramene	plasticity QTLs	127556445	127565804	.	+	.	Zm00001d025722
10	gramene	plasticity QTLs	127571715	127575954	.	-	.	Zm00001d025723
10	gramene	plasticity QTLs	127578165	127583188	.	+	.	Zm00001d025724
10	gramene	plasticity QTLs	127600573	127605016	.	-	.	Zm00001d025725
10	gramene	plasticity QTLs	127609718	127611616	.	+	.	Zm00001d025726
10	gramene	plasticity QTLs	127668399	127678365	.	-	.	Zm00001d025727
10	gramene	plasticity QTLs	127881516	127887702	.	-	.	Zm00001d025733
10	gramene	plasticity QTLs	127905113	127929830	.	+	.	Zm00001d025734
10	gramene	plasticity QTLs	127934025	127936164	.	+	.	Zm00001d025735
10	gramene	plasticity QTLs	127951568	127954961	.	+	.	Zm00001d025737
10	gramene	plasticity QTLs	127957093	127963022	.	-	.	Zm00001d025738
10	gramene	plasticity QTLs	128021608	128027534	.	-	.	Zm00001d025739
1	gramene	plasticity QTLs	190649919	190653721	.	-	.	Zm00001d031454
1	gramene	plasticity QTLs	192424400	192424708	.	-	.	Zm00001d031513
1	gramene	plasticity QTLs	192436661	192437457	.	-	.	Zm00001d031514
1	gramene	plasticity QTLs	192483919	192484951	.	+	.	Zm00001d031515
1	gramene	plasticity QTLs	25715085	25717227	.	+	.	Zm00001d028181
1	gramene	plasticity QTLs	25720877	25723883	.	+	.	Zm00001d028182
1	gramene	plasticity QTLs	25725534	25732327	.	-	.	Zm00001d028183
1	gramene	plasticity QTLs	25734022	25737544	.	+	.	Zm00001d028184
1	gramene	plasticity QTLs	25738664	25741208	.	-	.	Zm00001d028185
1	gramene	plasticity QTLs	25742203	25745382	.	+	.	Zm00001d028186
1	gramene	plasticity QTLs	25755280	25759748	.	+	.	Zm00001d028187
1	gramene	plasticity QTLs	26428405	26434744	.	+	.	Zm00001d028208
1	gramene	plasticity QTLs	32771811	32783919	.	-	.	Zm00001d028382

1	gramene	plasticity QTLs	32801911	32809405	.	+	.	Zm00001d028384
1	gramene	plasticity QTLs	32816668	32820422	.	+	.	Zm00001d028385
1	gramene	plasticity QTLs	32857622	32861798	.	-	.	Zm00001d028386
1	gramene	plasticity QTLs	52499430	52509398	.	+	.	Zm00001d028946
1	gramene	plasticity QTLs	52515222	52517115	.	+	.	Zm00001d028947
1	gramene	plasticity QTLs	52518690	52523101	.	+	.	Zm00001d028948
1	gramene	plasticity QTLs	52582858	52587987	.	-	.	Zm00001d028949
1	gramene	plasticity QTLs	52644689	52645454	.	+	.	Zm00001d028950
1	gramene	plasticity QTLs	52645733	52648243	.	-	.	Zm00001d028951
1	gramene	plasticity QTLs	52672158	52680876	.	-	.	Zm00001d028952
1	gramene	plasticity QTLs	52749160	52749765	.	-	.	Zm00001d028953
1	gramene	plasticity QTLs	52836664	52841244	.	+	.	Zm00001d028954
1	gramene	plasticity QTLs	77398644	77401233	.	+	.	Zm00001d029577
1	gramene	plasticity QTLs	77449634	77455875	.	-	.	Zm00001d029578
2	gramene	plasticity QTLs	221755305	221756948	.	+	.	Zm00001d007076
2	gramene	plasticity QTLs	221760272	221762560	.	+	.	Zm00001d007077
4	gramene	plasticity QTLs	180435545	180436378	.	-	.	Zm00001d052129
4	gramene	plasticity QTLs	180439840	180442555	.	+	.	Zm00001d052130
4	gramene	plasticity QTLs	180443235	180444935	.	-	.	Zm00001d052131
4	gramene	plasticity QTLs	181715802	181717531	.	-	.	Zm00001d052174
4	gramene	plasticity QTLs	181724426	181727817	.	+	.	Zm00001d052176
4	gramene	plasticity QTLs	181728119	181728577	.	+	.	Zm00001d052177
4	gramene	plasticity QTLs	181729433	181731132	.	-	.	Zm00001d052178
5	gramene	QTLs	204181533	204182126	.	+	.	Zm00001d017684
5	gramene	QTLs	12228329	12230206	.	-	.	Zm00001d013465
1	gramene	QTLs	213778926	213779774	.	+	.	Zm00001d032120
1	gramene	QTLs	213709256	213712753	.	-	.	Zm00001d032118
3	gramene	QTLs	136545200	136546312	.	-	.	Zm00001d041766
3	gramene	QTLs	136598676	136603856	.	-	.	Zm00001d041767
3	gramene	QTLs	136774212	136776802	.	+	.	Zm00001d041768
3	gramene	QTLs	136872860	136877613	.	-	.	Zm00001d041769
3	gramene	QTLs	136898355	136898755	.	+	.	Zm00001d041770
3	gramene	QTLs	136996989	137007669	.	+	.	Zm00001d041772
3	gramene	QTLs	137017318	137022601	.	+	.	Zm00001d041773
3	gramene	QTLs	137063373	137063753	.	-	.	Zm00001d041774
3	gramene	QTLs	137091203	137104652	.	+	.	Zm00001d041775
3	gramene	QTLs	137250657	137254412	.	-	.	Zm00001d041776
3	gramene	QTLs	137266061	137269605	.	+	.	Zm00001d041777
3	gramene	QTLs	137460338	137470304	.	+	.	Zm00001d041778
3	gramene	QTLs	137588464	137590201	.	+	.	Zm00001d041780
3	gramene	QTLs	137903169	137908918	.	-	.	Zm00001d041781
3	gramene	QTLs	137921659	137922555	.	+	.	Zm00001d041782
3	gramene	QTLs	137968493	137970915	.	+	.	Zm00001d041784
3	gramene	QTLs	138022243	138030035	.	+	.	Zm00001d041785
3	gramene	QTLs	138047384	138065736	.	-	.	Zm00001d041786
3	gramene	QTLs	138087163	138091017	.	+	.	Zm00001d041787
3	gramene	QTLs	138093726	138100085	.	-	.	Zm00001d041788
4	gramene	QTLs	183587060	183587659	.	-	.	Zm00001d052220
4	gramene	QTLs	183593619	183599554	.	+	.	Zm00001d052221
5	gramene	QTLs	190228612	190230421	.	+	.	Zm00001d017246
5	gramene	QTLs	190232645	190237410	.	-	.	Zm00001d017247
6	gramene	QTLs	27973065	27977307	.	+	.	Zm00001d035462
6	gramene	QTLs	28055121	28055806	.	-	.	Zm00001d035463
6	gramene	QTLs	28058154	28062105	.	-	.	Zm00001d035465
6	gramene	QTLs	28062742	28064462	.	+	.	Zm00001d035466

6	gramene	QTLs	28070797	28073319	.	-	.	Zm00001d035467
6	gramene	QTLs	28113842	28124886	.	-	.	Zm00001d035468
6	gramene	QTLs	28271626	28278463	.	+	.	Zm00001d035470
6	gramene	QTLs	28314084	28314467	.	-	.	Zm00001d035471
6	gramene	QTLs	28367590	28369394	.	+	.	Zm00001d035473
6	gramene	QTLs	28423443	28424771	.	-	.	Zm00001d035474
6	gramene	QTLs	28512487	28520162	.	+	.	Zm00001d035475
6	gramene	QTLs	28516750	28525006	.	-	.	Zm00001d035476
7	gramene	QTLs	101174606	101177305	.	+	.	Zm00001d020228
7	gramene	QTLs	101183127	101183962	.	+	.	Zm00001d020229
7	gramene	QTLs	101295370	101299977	.	+	.	Zm00001d020230
7	gramene	QTLs	101354950	101357248	.	-	.	Zm00001d020231
7	gramene	QTLs	101449550	101452600	.	+	.	Zm00001d020232
7	gramene	QTLs	101487360	101490008	.	+	.	Zm00001d020233
7	gramene	QTLs	101492829	101504550	.	-	.	Zm00001d020234
7	gramene	QTLs	101645190	101646950	.	-	.	Zm00001d020236
7	gramene	QTLs	101648949	101652728	.	+	.	Zm00001d020237
7	gramene	QTLs	101652755	101657949	.	-	.	Zm00001d020238
7	gramene	QTLs	101925496	101931975	.	+	.	Zm00001d020242
7	gramene	QTLs	101933697	101935610	.	+	.	Zm00001d020243
7	gramene	QTLs	101950149	101950695	.	+	.	Zm00001d020245
7	gramene	QTLs	102017704	102020268	.	-	.	Zm00001d020246
7	gramene	QTLs	102055809	102057946	.	-	.	Zm00001d020248
7	gramene	QTLs	102059093	102060260	.	+	.	Zm00001d020249
3	gramene	QTLs	35652118	35654852	.	-	.	Zm00001d040281
3	gramene	QTLs	35692155	35692739	.	-	.	Zm00001d040285
3	gramene	QTLs	35721411	35728565	.	-	.	Zm00001d040286
3	gramene	QTLs	35732184	35735342	.	-	.	Zm00001d040287
3	gramene	QTLs	35908042	35908935	.	-	.	Zm00001d040289
3	gramene	QTLs	35941532	35951651	.	+	.	Zm00001d040290
3	gramene	QTLs	36013414	36017372	.	+	.	Zm00001d040291
3	gramene	QTLs	36020981	36021268	.	-	.	Zm00001d040292
3	gramene	QTLs	36027929	36029668	.	-	.	Zm00001d040293
3	gramene	QTLs	36194689	36199305	.	-	.	Zm00001d040294
3	gramene	QTLs	36255672	36256522	.	-	.	Zm00001d040295
3	gramene	QTLs	36259400	36260123	.	+	.	Zm00001d040296
3	gramene	QTLs	36319753	36320169	.	-	.	Zm00001d040297
3	gramene	QTLs	36417931	36418257	.	-	.	Zm00001d040298
3	gramene	QTLs	36490987	36491586	.	+	.	Zm00001d040299
3	gramene	QTLs	36498228	36498710	.	+	.	Zm00001d040300
3	gramene	QTLs	36557966	36563190	.	-	.	Zm00001d040301
3	gramene	QTLs	36669164	36675206	.	-	.	Zm00001d040302
3	gramene	QTLs	36817316	36823842	.	+	.	Zm00001d040303
3	gramene	QTLs	36892545	36892988	.	+	.	Zm00001d040304
3	gramene	QTLs	36893536	36902030	.	-	.	Zm00001d040305
3	gramene	QTLs	36998433	36999059	.	-	.	Zm00001d040306
3	gramene	QTLs	36999882	37000898	.	-	.	Zm00001d040307
3	gramene	QTLs	37004302	37006317	.	-	.	Zm00001d040308
3	gramene	QTLs	37107695	37114751	.	+	.	Zm00001d040309
3	gramene	QTLs	37114976	37116953	.	-	.	Zm00001d040310
4	gramene	QTLs	17281947	17287115	.	+	.	Zm00001d049135
5	gramene	QTLs	90212611	90212916	.	-	.	Zm00001d015437
5	gramene	QTLs	90267647	90277773	.	-	.	Zm00001d015440
5	gramene	QTLs	90343794	90344970	.	-	.	Zm00001d015442
5	gramene	QTLs	90418616	90437459	.	+	.	Zm00001d015443

5	gramene	QTLs	90487714	90495272	.	-	.	Zm00001d015444
5	gramene	QTLs	90494901	90498102	.	+	.	Zm00001d015445
5	gramene	QTLs	90502234	90559505	.	-	.	Zm00001d015446
5	gramene	QTLs	90859128	90860638	.	+	.	Zm00001d015448
5	gramene	QTLs	90918110	90920511	.	+	.	Zm00001d015449
5	gramene	QTLs	90962311	90968591	.	-	.	Zm00001d015450
5	gramene	QTLs	91106167	91109100	.	+	.	Zm00001d015451
5	gramene	QTLs	91146435	91148057	.	+	.	Zm00001d015452
5	gramene	QTLs	91150513	91151598	.	-	.	Zm00001d015453
5	gramene	QTLs	91154171	91158469	.	+	.	Zm00001d015454
5	gramene	QTLs	91162670	91169361	.	+	.	Zm00001d015455
5	gramene	QTLs	91172191	91173713	.	+	.	Zm00001d015456
5	gramene	QTLs	91310895	91315659	.	-	.	Zm00001d015457
5	gramene	QTLs	91323709	91324631	.	+	.	Zm00001d015458
2	gramene	QTLs	55227577	55229673	.	-	.	Zm00001d003701
2	gramene	QTLs	55231147	55233488	.	-	.	Zm00001d003702
2	gramene	QTLs	55390825	55395981	.	+	.	Zm00001d003703
2	gramene	QTLs	55527083	55529059	.	+	.	Zm00001d003705
2	gramene	QTLs	55636062	55638024	.	+	.	Zm00001d003706
2	gramene	QTLs	55659655	55661242	.	-	.	Zm00001d003707
2	gramene	QTLs	55723782	55724426	.	+	.	Zm00001d003709
2	gramene	QTLs	55847268	55852437	.	+	.	Zm00001d003710
2	gramene	QTLs	55894819	55896267	.	+	.	Zm00001d003712
2	gramene	QTLs	56082381	56097514	.	+	.	Zm00001d003713
2	gramene	QTLs	56092761	56093555	.	-	.	Zm00001d003714
2	gramene	QTLs	56141056	56143346	.	+	.	Zm00001d003715
2	gramene	QTLs	56144508	56147710	.	+	.	Zm00001d003716
2	gramene	QTLs	106617018	106618030	.	+	.	Zm00001d004364
2	gramene	QTLs	106625633	106628492	.	+	.	Zm00001d004365
2	gramene	QTLs	106628660	106630400	.	-	.	Zm00001d004366
2	gramene	QTLs	106726023	106726880	.	+	.	Zm00001d004370
2	gramene	QTLs	106774658	106775819	.	-	.	Zm00001d004371
2	gramene	QTLs	106904147	106908784	.	+	.	Zm00001d004372
2	gramene	QTLs	107040785	107042488	.	-	.	Zm00001d004375
2	gramene	QTLs	107046240	107049098	.	+	.	Zm00001d004376
2	gramene	QTLs	107365020	107365884	.	-	.	Zm00001d004379
2	gramene	QTLs	107469703	107470980	.	-	.	Zm00001d004380
3	gramene	QTLs	53767845	53777713	.	+	.	Zm00001d040611
3	gramene	QTLs	53863310	53874662	.	+	.	Zm00001d040612
3	gramene	QTLs	53996519	53997786	.	-	.	Zm00001d040613
3	gramene	QTLs	54002070	54005318	.	-	.	Zm00001d040614
3	gramene	QTLs	54060892	54062596	.	-	.	Zm00001d040617
3	gramene	QTLs	54065171	54067102	.	+	.	Zm00001d040618
3	gramene	QTLs	54139362	54139841	.	+	.	Zm00001d040619
3	gramene	QTLs	54202177	54203490	.	-	.	Zm00001d040620
3	gramene	QTLs	54361997	54365806	.	+	.	Zm00001d040621
3	gramene	QTLs	54366428	54368879	.	-	.	Zm00001d040622
3	gramene	QTLs	54606634	54607098	.	-	.	Zm00001d040623
3	gramene	QTLs	54717613	54718843	.	+	.	Zm00001d040624
3	gramene	QTLs	54768172	54773642	.	-	.	Zm00001d040625
3	gramene	QTLs	54948300	54950704	.	-	.	Zm00001d040627
3	gramene	QTLs	55023821	55026070	.	+	.	Zm00001d040628
3	gramene	QTLs	55075888	55088174	.	-	.	Zm00001d040629
3	gramene	QTLs	55088527	55091474	.	+	.	Zm00001d040630
3	gramene	QTLs	124919974	124923682	.	+	.	Zm00001d041508

3	gramene	QTLs	124926387	124927241	.	-	.	Zm00001d041509
3	gramene	QTLs	125020876	125028056	.	+	.	Zm00001d041510
3	gramene	QTLs	125028657	125033309	.	-	.	Zm00001d041511
3	gramene	QTLs	125119438	125121733	.	-	.	Zm00001d041513
3	gramene	QTLs	125122808	125126410	.	+	.	Zm00001d041514
3	gramene	QTLs	125289850	125330088	.	-	.	Zm00001d041515
3	gramene	QTLs	125552392	125557077	.	+	.	Zm00001d041516
3	gramene	QTLs	125560575	125561600	.	-	.	Zm00001d041518
3	gramene	QTLs	125563618	125564283	.	+	.	Zm00001d041519
3	gramene	QTLs	125608671	125609400	.	+	.	Zm00001d041520
3	gramene	QTLs	125708192	125710214	.	-	.	Zm00001d041521
3	gramene	QTLs	125843356	125846615	.	-	.	Zm00001d041522
3	gramene	QTLs	125849221	125851417	.	+	.	Zm00001d041523
3	gramene	QTLs	125901674	125905287	.	-	.	Zm00001d041525
3	gramene	QTLs	125967179	125969813	.	-	.	Zm00001d041526
3	gramene	QTLs	125971928	125972875	.	-	.	Zm00001d041528
3	gramene	QTLs	126000149	126004313	.	-	.	Zm00001d041529
3	gramene	QTLs	126035560	126036633	.	+	.	Zm00001d041530
3	gramene	QTLs	126088939	126092644	.	+	.	Zm00001d041531
3	gramene	QTLs	126104284	126105794	.	-	.	Zm00001d041533
3	gramene	QTLs	126106707	126109981	.	+	.	Zm00001d041534
3	gramene	QTLs	126110008	126110211	.	-	.	Zm00001d041535
3	gramene	QTLs	126130758	126132890	.	-	.	Zm00001d041536
3	gramene	QTLs	126166316	126167305	.	-	.	Zm00001d041537
3	gramene	QTLs	126304770	126305192	.	-	.	Zm00001d041538
3	gramene	QTLs	126342753	126344912	.	+	.	Zm00001d041539
3	gramene	QTLs	126429213	126430962	.	+	.	Zm00001d041541
3	gramene	QTLs	126522185	126523196	.	-	.	Zm00001d041544
3	gramene	QTLs	126535818	126539525	.	-	.	Zm00001d041545
3	gramene	QTLs	126536801	126539525	.	+	.	Zm00001d041546
3	gramene	QTLs	126563754	126564497	.	+	.	Zm00001d041547
3	gramene	QTLs	217573164	217581039	.	+	.	Zm00001d044042
5	gramene	QTLs	17465790	17467595	.	-	.	Zm00001d013689
8	gramene	QTLs	160377553	160379112	.	-	.	Zm00001d011737
8	gramene	QTLs	160386965	160392640	.	+	.	Zm00001d011738
2	gramene	QTLs	108518434	108535572	.	-	.	Zm00001d004392
2	gramene	QTLs	108539540	108542092	.	+	.	Zm00001d004394
2	gramene	QTLs	108736242	108762263	.	-	.	Zm00001d004396
2	gramene	QTLs	108913222	108913905	.	+	.	Zm00001d004397
2	gramene	QTLs	109230347	109237741	.	-	.	Zm00001d004400
2	gramene	QTLs	109327004	109328229	.	+	.	Zm00001d004401
2	gramene	QTLs	154614760	154616824	.	+	.	Zm00001d005026
2	gramene	QTLs	154617731	154621976	.	+	.	Zm00001d005027
2	gramene	QTLs	154629678	154631157	.	-	.	Zm00001d005028
2	gramene	QTLs	154743078	154746493	.	-	.	Zm00001d005029
2	gramene	QTLs	154831004	154835182	.	-	.	Zm00001d005030
2	gramene	QTLs	154983726	154985286	.	-	.	Zm00001d005031
2	gramene	QTLs	155039785	155042337	.	-	.	Zm00001d005032
2	gramene	QTLs	155214141	155229333	.	+	.	Zm00001d005035
2	gramene	QTLs	155301263	155306254	.	+	.	Zm00001d005036
2	gramene	QTLs	155349292	155350224	.	+	.	Zm00001d005037
2	gramene	QTLs	155403486	155405109	.	+	.	Zm00001d005038
2	gramene	QTLs	155435155	155436114	.	+	.	Zm00001d005039
2	gramene	QTLs	155438279	155441199	.	-	.	Zm00001d005040
5	gramene	QTLs	79214291	79242188	.	+	.	Zm00001d015203

5	gramene	QTLs	79242210	79243503	.	-	.	Zm00001d015204
5	gramene	QTLs	79250420	79253647	.	+	.	Zm00001d015205
5	gramene	QTLs	79254798	79255229	.	-	.	Zm00001d015206
5	gramene	QTLs	79389803	79395312	.	+	.	Zm00001d015208
5	gramene	QTLs	79498782	79511605	.	+	.	Zm00001d015209
5	gramene	QTLs	79546867	79550125	.	+	.	Zm00001d015210
5	gramene	QTLs	79551939	79555712	.	+	.	Zm00001d015211
5	gramene	QTLs	79554503	79570576	.	-	.	Zm00001d015212
5	gramene	QTLs	79757206	79759176	.	-	.	Zm00001d015213
5	gramene	QTLs	79824018	79825123	.	+	.	Zm00001d015215
5	gramene	QTLs	79871178	79872762	.	+	.	Zm00001d015216
5	gramene	QTLs	79954321	79956357	.	+	.	Zm00001d015217
5	gramene	QTLs	79960119	79963951	.	-	.	Zm00001d015219
5	gramene	QTLs	79981188	79994280	.	+	.	Zm00001d015220
5	gramene	QTLs	80009524	80019739	.	-	.	Zm00001d015221
5	gramene	QTLs	80021026	80021571	.	-	.	Zm00001d015223
5	gramene	QTLs	80045134	80047032	.	+	.	Zm00001d015224
5	gramene	QTLs	80063827	80067249	.	+	.	Zm00001d015225
5	gramene	QTLs	80066943	80069666	.	-	.	Zm00001d015226
5	gramene	QTLs	80071631	80074957	.	-	.	Zm00001d015227
5	gramene	QTLs	80071631	80077888	.	-	.	Zm00001d015228
5	gramene	QTLs	80099849	80104698	.	-	.	Zm00001d015229
5	gramene	QTLs	80210099	80215267	.	+	.	Zm00001d015231
5	gramene	QTLs	80234208	80239340	.	+	.	Zm00001d015233
5	gramene	QTLs	80281071	80289642	.	-	.	Zm00001d015234
5	gramene	QTLs	80463268	80472312	.	-	.	Zm00001d015238
1	gramene	QTLs	213876669	213877625	.	+	.	Zm00001d032125
1	gramene	QTLs	213878646	213879350	.	-	.	Zm00001d032126
1	gramene	QTLs	213895961	213901572	.	+	.	Zm00001d032127
1	gramene	QTLs	213904619	213907090	.	+	.	Zm00001d032128
1	gramene	QTLs	213957721	213959642	.	+	.	Zm00001d032130
1	gramene	QTLs	21680697	21686575	.	+	.	Zm00001d028050
1	gramene	QTLs	21690752	21692638	.	-	.	Zm00001d028051
2	gramene	QTLs	110107388	110108146	.	-	.	Zm00001d004405
2	gramene	QTLs	110120314	110125179	.	+	.	Zm00001d004407
2	gramene	QTLs	110207245	110210898	.	+	.	Zm00001d004409
2	gramene	QTLs	110216975	110222616	.	-	.	Zm00001d004410
2	gramene	QTLs	110288909	110290031	.	-	.	Zm00001d004411
2	gramene	QTLs	110310942	110312623	.	-	.	Zm00001d004412
2	gramene	QTLs	110314499	110316622	.	+	.	Zm00001d004413
2	gramene	QTLs	110343635	110345001	.	-	.	Zm00001d004414
2	gramene	QTLs	110345638	110358087	.	-	.	Zm00001d004415
2	gramene	QTLs	110383916	110417211	.	-	.	Zm00001d004416
2	gramene	QTLs	110448343	110449312	.	-	.	Zm00001d004417
2	gramene	QTLs	110450100	110451204	.	-	.	Zm00001d004418
2	gramene	QTLs	110481721	110482872	.	-	.	Zm00001d004419
2	gramene	QTLs	155778352	155779569	.	+	.	Zm00001d005043
2	gramene	QTLs	155847589	155848806	.	+	.	Zm00001d005044
2	gramene	QTLs	155849179	155854203	.	+	.	Zm00001d005045
2	gramene	QTLs	203038721	203039941	.	+	.	Zm00001d006260
2	gramene	QTLs	203134987	203135343	.	+	.	Zm00001d006261
2	gramene	QTLs	203183372	203184196	.	+	.	Zm00001d006262
2	gramene	QTLs	203194535	203195885	.	+	.	Zm00001d006263
2	gramene	QTLs	203274963	203284292	.	+	.	Zm00001d006267
2	gramene	QTLs	203285298	203286451	.	-	.	Zm00001d006268

2	gramene	QTLs	203287605	203291085	.	-	.	Zm00001d006269
2	gramene	QTLs	203313532	203322278	.	-	.	Zm00001d006270
2	gramene	QTLs	203376662	203391214	.	+	.	Zm00001d006271
2	gramene	QTLs	203379297	203381647	.	-	.	Zm00001d006272
2	gramene	QTLs	203395420	203402757	.	-	.	Zm00001d006274
2	gramene	QTLs	203397990	203399055	.	+	.	Zm00001d006275
2	gramene	QTLs	203401394	203401852	.	+	.	Zm00001d006276
2	gramene	QTLs	203403246	203403638	.	-	.	Zm00001d006277
2	gramene	QTLs	203403256	203403672	.	+	.	Zm00001d006278
2	gramene	QTLs	203405094	203405564	.	+	.	Zm00001d006279
2	gramene	QTLs	203487434	203487826	.	-	.	Zm00001d006280
2	gramene	QTLs	203487444	203487860	.	+	.	Zm00001d006281
2	gramene	QTLs	203489297	203489769	.	+	.	Zm00001d006282
2	gramene	QTLs	203490871	203491346	.	-	.	Zm00001d006283
2	gramene	QTLs	203553961	203554453	.	-	.	Zm00001d006285
2	gramene	QTLs	203633940	203638997	.	+	.	Zm00001d006286
2	gramene	QTLs	203642768	203644284	.	+	.	Zm00001d006287
2	gramene	QTLs	203647462	203651595	.	+	.	Zm00001d006288
2	gramene	QTLs	203652583	203653658	.	-	.	Zm00001d006289
2	gramene	QTLs	203787929	203790186	.	+	.	Zm00001d006291
2	gramene	QTLs	203858974	203864535	.	+	.	Zm00001d006293
2	gramene	QTLs	203866427	203867230	.	-	.	Zm00001d006294
2	gramene	QTLs	203933323	203937922	.	-	.	Zm00001d006295
2	gramene	QTLs	203995835	204009490	.	+	.	Zm00001d006296
2	gramene	QTLs	54981065	54984287	.	-	.	Zm00001d003696
2	gramene	QTLs	55078599	55079495	.	+	.	Zm00001d003698
2	gramene	QTLs	55135625	55136584	.	-	.	Zm00001d003699
2	gramene	QTLs	85517707	85522558	.	+	.	Zm00001d004135
2	gramene	QTLs	85695276	85704125	.	+	.	Zm00001d004136
2	gramene	QTLs	85731223	85732587	.	+	.	Zm00001d004137
2	gramene	QTLs	85736310	85737686	.	-	.	Zm00001d004138
2	gramene	QTLs	85783948	85803502	.	-	.	Zm00001d004139
2	gramene	QTLs	86026385	86044778	.	-	.	Zm00001d004140
2	gramene	QTLs	86093715	86098066	.	-	.	Zm00001d004143
2	gramene	QTLs	86233344	86236604	.	-	.	Zm00001d004146
2	gramene	QTLs	86359248	86367245	.	+	.	Zm00001d004147
3	gramene	QTLs	171421717	171425401	.	+	.	Zm00001d042541
3	gramene	QTLs	171440599	171442476	.	-	.	Zm00001d042542
3	gramene	QTLs	179131107	179131481	.	+	.	Zm00001d042765
3	gramene	QTLs	179135191	179136021	.	+	.	Zm00001d042766
3	gramene	QTLs	6094711	6099557	.	-	.	Zm00001d039505
4	gramene	QTLs	16983481	16984955	.	-	.	Zm00001d049129
4	gramene	QTLs	18913325	18914911	.	-	.	Zm00001d049169
4	gramene	QTLs	23084615	23087441	.	-	.	Zm00001d049256
5	gramene	QTLs	150295060	150299829	.	+	.	Zm00001d016197
5	gramene	QTLs	150300976	150305724	.	-	.	Zm00001d016198
5	gramene	QTLs	150352205	150353890	.	-	.	Zm00001d016199
5	gramene	QTLs	150355001	150356922	.	-	.	Zm00001d016200
5	gramene	QTLs	150380818	150382317	.	+	.	Zm00001d016202
5	gramene	QTLs	150386027	150389534	.	-	.	Zm00001d016203
5	gramene	QTLs	150391274	150394468	.	-	.	Zm00001d016204
5	gramene	QTLs	150429683	150431718	.	-	.	Zm00001d016206
5	gramene	QTLs	150432936	150435131	.	-	.	Zm00001d016207
5	gramene	QTLs	150578014	150578814	.	-	.	Zm00001d016208
5	gramene	QTLs	150582724	150584481	.	+	.	Zm00001d016211

5	gramene	QTLs	150587103	150594630	.	+	.	Zm00001d016212
5	gramene	QTLs	150600170	150603876	.	+	.	Zm00001d016214
5	gramene	QTLs	150673357	150674317	.	+	.	Zm00001d016216
5	gramene	QTLs	150753673	150756808	.	-	.	Zm00001d016217
5	gramene	QTLs	150960524	150964808	.	-	.	Zm00001d016219
5	gramene	QTLs	150963096	150964787	.	-	.	Zm00001d016220
5	gramene	QTLs	150977946	150978437	.	+	.	Zm00001d016221
5	gramene	QTLs	151001655	151004102	.	-	.	Zm00001d016223
5	gramene	QTLs	151007142	151007609	.	-	.	Zm00001d016224
5	gramene	QTLs	151044249	151046438	.	-	.	Zm00001d016225
5	gramene	QTLs	151100013	151100423	.	+	.	Zm00001d016227
5	gramene	QTLs	151137109	151142706	.	-	.	Zm00001d016228
5	gramene	QTLs	151210417	151211022	.	+	.	Zm00001d016229
5	gramene	QTLs	151213035	151216982	.	-	.	Zm00001d016230
5	gramene	QTLs	151405284	151407104	.	-	.	Zm00001d016231
5	gramene	QTLs	151417504	151418350	.	-	.	Zm00001d016232
5	gramene	QTLs	151507764	151512142	.	-	.	Zm00001d016234
5	gramene	QTLs	151552379	151552993	.	+	.	Zm00001d016235
5	gramene	QTLs	162045948	162047120	.	+	.	Zm00001d016427
5	gramene	QTLs	162048202	162048648	.	-	.	Zm00001d016428
5	gramene	QTLs	163090196	163091701	.	-	.	Zm00001d016460
5	gramene	QTLs	163098850	163102287	.	+	.	Zm00001d016461
5	gramene	QTLs	184324784	184338621	.	+	.	Zm00001d017068
5	gramene	QTLs	190041422	190045024	.	-	.	Zm00001d017238
5	gramene	QTLs	190049317	190050888	.	+	.	Zm00001d017239
5	gramene	QTLs	190054668	190058017	.	+	.	Zm00001d017240
5	gramene	QTLs	190060998	190063941	.	+	.	Zm00001d017241
5	gramene	QTLs	199282391	199283780	.	-	.	Zm00001d017548
5	gramene	QTLs	201699854	201703271	.	+	.	Zm00001d017613
5	gramene	QTLs	80579817	80605889	.	-	.	Zm00001d015239
5	gramene	QTLs	80675439	80680775	.	+	.	Zm00001d015242
5	gramene	QTLs	80676960	80685239	.	-	.	Zm00001d015243
5	gramene	QTLs	80936638	80937557	.	+	.	Zm00001d015244
5	gramene	QTLs	88146164	88150303	.	+	.	Zm00001d015394
5	gramene	QTLs	88158162	88161170	.	-	.	Zm00001d015395
5	gramene	QTLs	88228087	88230594	.	-	.	Zm00001d015396
5	gramene	QTLs	88306260	88306964	.	+	.	Zm00001d015397
5	gramene	QTLs	88328874	88330634	.	-	.	Zm00001d015398
5	gramene	QTLs	88372291	88376274	.	+	.	Zm00001d015399
5	gramene	QTLs	88443306	88448870	.	+	.	Zm00001d015400
5	gramene	QTLs	88448697	88451738	.	-	.	Zm00001d015401
5	gramene	QTLs	88484657	88490422	.	+	.	Zm00001d015404
5	gramene	QTLs	88491212	88497833	.	+	.	Zm00001d015405
5	gramene	QTLs	88518024	88520798	.	+	.	Zm00001d015406
5	gramene	QTLs	88561896	88565000	.	-	.	Zm00001d015407
5	gramene	QTLs	88962266	88972559	.	+	.	Zm00001d015409
5	gramene	QTLs	89049594	89053097	.	+	.	Zm00001d015410
5	gramene	QTLs	89059774	89060808	.	-	.	Zm00001d015411
5	gramene	QTLs	89090290	89091195	.	+	.	Zm00001d015412
5	gramene	QTLs	89126963	89129557	.	+	.	Zm00001d015414
5	gramene	QTLs	89132049	89135337	.	-	.	Zm00001d015415
5	gramene	QTLs	89200500	89201452	.	+	.	Zm00001d015417
5	gramene	QTLs	89294499	89295622	.	-	.	Zm00001d015419
5	gramene	QTLs	89297902	89299723	.	+	.	Zm00001d015420
5	gramene	QTLs	89376383	89379344	.	+	.	Zm00001d015421

5	gramene	QTLs	89382778	89386161	.	-	.	Zm00001d015423
7	gramene	QTLs	100204794	100211183	.	+	.	Zm00001d020208
7	gramene	QTLs	100243410	100247563	.	+	.	Zm00001d020209
7	gramene	QTLs	100351627	100351913	.	-	.	Zm00001d020210
7	gramene	QTLs	100353278	100354591	.	-	.	Zm00001d020211
7	gramene	QTLs	100359175	100364348	.	+	.	Zm00001d020213
7	gramene	QTLs	100369403	100370862	.	-	.	Zm00001d020214
7	gramene	QTLs	100656060	100660194	.	+	.	Zm00001d020219
7	gramene	QTLs	100663432	100666790	.	+	.	Zm00001d020220
7	gramene	QTLs	100722127	100724203	.	-	.	Zm00001d020222
7	gramene	QTLs	100770949	100773059	.	+	.	Zm00001d020223
7	gramene	QTLs	100781683	100784608	.	-	.	Zm00001d020224
7	gramene	QTLs	100911743	100932200	.	+	.	Zm00001d020225
7	gramene	QTLs	101004708	101009836	.	-	.	Zm00001d020227
8	gramene	QTLs	146781789	146783051	.	+	.	Zm00001d011327
8	gramene	QTLs	146855836	146859841	.	-	.	Zm00001d011328
8	gramene	QTLs	146863198	146868117	.	-	.	Zm00001d011329
8	gramene	QTLs	146957682	146961482	.	+	.	Zm00001d011330
8	gramene	QTLs	147057186	147058121	.	-	.	Zm00001d011331
8	gramene	QTLs	147150138	147150711	.	-	.	Zm00001d011333
8	gramene	QTLs	147199565	147204907	.	+	.	Zm00001d011334
8	gramene	QTLs	147223917	147226682	.	+	.	Zm00001d011335
8	gramene	QTLs	147228068	147231731	.	-	.	Zm00001d011336
8	gramene	QTLs	147322981	147323332	.	+	.	Zm00001d011340
8	gramene	QTLs	147353740	147354088	.	+	.	Zm00001d011342
8	gramene	QTLs	147384452	147384742	.	+	.	Zm00001d011345
8	gramene	QTLs	147385385	147385869	.	-	.	Zm00001d011346
8	gramene	QTLs	147387271	147396132	.	-	.	Zm00001d011347
8	gramene	QTLs	147533284	147535094	.	-	.	Zm00001d011348
8	gramene	QTLs	147535376	147539282	.	-	.	Zm00001d011349
8	gramene	QTLs	147564557	147566080	.	-	.	Zm00001d011350
8	gramene	QTLs	147718573	147727814	.	+	.	Zm00001d011351
9	gramene	QTLs	20480509	20481751	.	+	.	Zm00001d045392

Amal KSONTINI's internship report

RAPPORT DE STAGE

du 2 Mai au 2 Août 2023

Sujet : Analyses fonctionnelle de loci de caractères quantitatifs (QTLs) associés à la plasticité au stress hydrique chez le maïs

Amal KSONTINI

Master 1 GENIOMHE

2022-2023

Enseignant référent : Marie-Helene MUCCHIELLI GIORGI
Encadrants de stage : Yacine DJABALI et Marie-Laure MARTIN

Équipe Genomic Networks (GNet)
Institut de Sciences des Plantes Paris-Saclay (IPS2)
Bâtiment 630, rue de Noetzlin
91190 - Gif-sur-Yvette

Remerciements

Je tiens tout d'abord à remercier toute l'équipe GNet pour son accueil pendant ce stage. Je souhaite tout particulièrement remercier mes encadrants, Marie-Laure MARTIN et Yacine DJABALI, pour leur expertise, leurs précieux conseils et leur accompagnement tout au long de mon stage. Je tiens également à remercier Mélisande BLEIN-NICOLAS pour les dialogues et les échanges enrichissants pendant ce projet.

Sommaire

Abréviations

PGM : Poids de millle grains
GWAS : Genome Wide association studies
QTL : Quantitative trait locus = locus de trait quantitatif
WW : Well-Watered = Bien irrigué
WD : Water-Deficit = Déficit en eau
GxE : Interaction Génome-Environnement
BAC : Bacterial artificial chromosome
SMRT : Single Molecule Real Time
SNP : Single nucleotid polymorphism

1. Introduction

1.1. Présentation de l'IPS2 et de l'équipe GNet

L'Institut des Plantes Paris-Saclay est un laboratoire de recherche spécialisé dans les sciences du végétal. L'établissement est dédié à approfondir la compréhension des mécanismes génétiques et moléculaires régissant les plantes ainsi que leur régulation par des signaux endogènes et exogènes. Dans le but de favoriser la collaboration et le partage des connaissances, l'IPS2 dispose de 5 plateformes comprenant la biologie translationnelle , la transcriptomique , la métabolomique, l'interactomique et l'épigénomique.

J'ai effectué mon stage au sein de l'équipe *Réseaux Génomiques* Gnet qui travaille sur le développement de méthodes statistiques et d'approches bioinformatiques pour répondre à des questions d'intérêt agronomique.

1.2. Contexte

Les changements atmosphériques causés par les activités humaines sont la source d'une augmentation rapide des températures dans différentes régions du monde. Selon les estimations du GIEC pour le court terme (2021-2040), le réchauffement climatique atteindra 1.5 C° par rapport à 1850-1900 même dans le cas d'une émission très diminuée de gaz à effet de serre.

Les phénomènes climatiques sont la cause principale d'insécurité alimentaire pour 56.8 millions de personnes dans 12 pays différents (FAO, 2022). En effet, les augmentations de température contribuent aux scénarios de sécheresse impactant les rendements des cultures en raison du stress hydrique. La perte de rendement pour certaines cultures comme le blé, le riz ou encore le maïs provoquerait un accroissement de l'insécurité alimentaire (FAO, 2023). L'environnement qui nous attend dans les prochaines décennies intensifiera ces défis. Il est donc important de trouver des solutions pour rendre les espèces d'intérêt agronomiques plus résistantes au stress hydrique et améliorer leur rendement.

Parmi les cultures impactées par le stress hydrique il y a le maïs, la céréale la plus cultivée dans le monde avec une production de 1.151 million de tonnes en 2022 (Statista, 2022- 2023). Cette céréale est sollicitée dans l'approvisionnement alimentaire au niveau mondial, elle est utilisée majoritairement pour nourrir le bétail mais est également utilisée comme source de nourriture pour certaines populations de pays en voie de développement. Les plus grands producteurs de maïs sont les États-Unis, la Chine et le Brésil avec respectivement 383.943.000, 272.552.000 et 88.461.943 tonnes de maïs produit pour l'année 2021 (FAOSTAT, 2021). Ces trois grands producteurs localisés dans des différentes régions du monde montrent à quel point cette espèce possède une grande capacité d'adaptation à son environnement. Cela fait de lui un bon candidat dans un éventuel futur où les échanges intercontinentaux seraient limités.

Cependant, cette espèce reste très sensible à la sécheresse malgré un métabolisme C4 lui conférant une efficacité d'utilisation de l'eau. Les carences en eau notamment

au moment de la floraison déclenchent des mécanismes qui réduisent les pertes d'eau mais cela entrave la croissance des plantes. Le rendement en grains peut également être compromis indirectement sous l'effet de la fermeture des stomates induisant une diminution de l'activité photosynthétique et de la production de biomasse. De plus, il a été montré que la sécheresse a un effet significatif sur la phénologie du maïs altérant les étapes de croissance végétative et raccourcissant les étapes de reproduction (Qi et al., 2022). Le remplissage des grains est également affecté leur poids peut être réduit de 20 à 40 % en cas de stress hydrique aigu. Ainsi la physiologie du maïs est considérablement affectée en réponse au stress hydrique.

La diversité phénotypique présente chez cette céréale a fait d'elle une espèce très étudiée des biologistes et est devenue un modèle biologique sur lequel plusieurs études génétiques ont été réalisées. En effet, l'exploration de son génome constitué de 2,4 Giga pb et de près de 56000 gènes répartis sur 10 chromosomes (Schnable et al., 2009) pourrait amener à mieux comprendre ses mécanismes de tolérance à la sécheresse dans une optique d'amélioration variétale. L'approche consiste alors à utiliser des marqueurs moléculaires pour identifier des gènes candidats impliqués dans les mécanismes de réponse à la sécheresse. Mon travail se place sur ce front de recherche en identifiant et en caractérisant ces gènes candidats.

1.3. Etat de l'art

Les études d'association génétique permettent d'identifier des polymorphismes génétiques (ex : SNPs single nucleotide polymorphisms) en étudiant la variabilité inter-individuelle d'un caractère observé sur un ensemble d'individus. Des régions du génome sont cartographiées par l'intermédiaire des polymorphismes retenus et sont appelés locis de traits quantitatifs (QTLs). Un QTL est une région du génome associée à une variation quantitative d'un trait phénotypique. Il s'agit d'une séquence d'ADN qui est corrélée de manière statistiquement significative avec des variations observées dans les mesures phénotypiques d'un trait donné. Les QTLs peuvent être identifiés à l'aide de techniques telles que les études de liaison génétique ou les analyses d'association pangénomique. Ils jouent un rôle clé dans la compréhension de la base génétique des caractères complexes et dans la cartographie des gènes responsables de ces traits.

Dans le cas du maïs, où de nombreux efforts ont été mis en place grâce au projet investissement d'avenir Amaizing, il existe un jeu de données sur un ensemble de 254 hybrides de maïs. Ils on été cultivés dans deux conditions dépendant de l'apport en eau des plantes : Bien irrigué (WW) et en déficit en eau (WD). Des GWAS single-environnement sur 4 expériences conduites pour trois années et deux saisons différentes ont permis d'identifier 402 QTLs à partir de 6 caractères phénotypiques (Prado et al., 2018). Cependant il y avait peu de reproductibilité d'une expérience à une autre dans les QTLs détectés. Ainsi (Djabali et al.,in revision) a réalisé un GWAS multi-environnement pour dissocier l'effet du déficit hydrique des autres effets liés à l'environnement de culture. De plus, il a considéré dans son analyse une étude de la plasticité qui est une

mesure de réponse au stress. Elle représente la façon dont un génotype peut donner différents phénotypes en fonction de l'environnement (Via et al., 1995). Pour cela il a réalisé une étude d'association sur le ratio entre la condition de déficit hydrique et la condition d'irrigation normale des 6 caractères. Ce travail qui améliore la modélisation du jeu de données a permis d'identifier 102 QTLs d'état stationnaire associés à la condition d'irrigation et 40 QTLs associés à la réponse au déficit hydrique Les QTLs détectés par ce modèle sont pour la plupart nouveaux par rapport à ceux qui ont identifié dans l'étude de Prado et al., 2018.

(Kusmec et al., 2017, Djabali et al.,in revision) ont montré que les QTLs d'état stationnaire et les QTLs de plasticité ne se superposent pas et sont situés dans des régions génétiques différentes. Cela implique que le contrôle génétique sous-jacents à la variation phénotypique dans des conditions stables (état stationnaire) diffère de celui impliqué dans la réponse phénotypique en réponse à des variations environnementales (plasticité). Cela souligne l'importance d'étudier à la fois les gènes candidats qui sont à l'origine des deux types de QTLs pour comprendre le mécanisme de régulation en réponse au stress hydrique. Une hypothèse présente dans la littérature indique que les gènes associés à la plasticité seraient des régulateurs de gènes associés aux états stationnaires Via et al., 1995. Ce modèle postule que la plasticité phénotypique est le résultat de l'interaction épistatique entre les éléments structuraux (gènes qui seraient associés aux QTLs d'états stationnaires) et les éléments régulateurs (gènes qui seraient associés aux QTLs de plasticité).

1.4. Objectifs du stage

Dans ce contexte à partir des résultats du travail de Yacine Djabali sur l'identification de QTLs de plasticité et des deux états stationnaires des 6 caractères écophysiologiques, l'objectif de mon stage était de faire la caractérisation fonctionnelle des gènes situés sous les QTLs et de vérifier l'hypothèse de régulation génique formulée dans (Via et al., 1995). Pour cela, mon stage a consisté à

- Caractériser les QTLs en fonction des traits écophysiologiques auxquels ils sont associés
- Identifier les gènes sous les QTLs associés à la plasticité et aux 2 conditions d'irrigation
- Caractériser fonctionnellement les gènes identifiés
- Identifier des facteurs de transcription ou des protéines régulatrices dans les listes des gènes identifiés

2. Matériel & méthodes

2.1. Matériel Disponible

2.1.1. Matériel végétal

Un panel de diversité a été obtenu en croisant 254 individus de lignée dentée avec une lignée flint standard (UH007). Trois plantes de chaque hybride ont été cultivées dans deux conditions d'irrigation. Dans la première condition, les plantes étaient arrosées suffisamment (WW) et dans l'autre condition, elles souffraient d'un déficit hydrique (WD).

Ces plantes ont été cultivées à la plateforme de phénotypage PhenoArch d'INRAE située à Montpellier. Six caractères écophysiologiques ont été mesurés : la Biomasse (Biol), la surface foliaire (LAl), le taux de transpiration (Trate), la conductance stomatique (gs_max), l'utilisation de l'eau (WU) et l'efficacité d'utilisation de l'eau (WUE). La moyenne génotypique, correspondant à la moyenne des trois réplicats, a été calculée et ajustée en prenant en compte l'effet spatial de la serre pour chaque trait et condition d'arrosage.

2.1.2. Données des QTLs

Au début de mon stage, j'ai reçu un tableau contenant pour chaque trait la liste des QTLs d'états stationnaires (WW ou WD) et de la plasticité. Chaque QTL est caractérisé par l'identifiant du SNP le plus significatif. Sur ce fichier, figure également la position des QTLs ainsi que le chromosome sur lequel chacun des QTLs est localisé. Ces informations sont indispensables pour identifier les gènes localisés sous les QTLs.

2.2. Annotation structurale

2.2.1. La base de données MaizeGDB

MaizeGDB [1] est une base de données organisant toutes les connaissances disponibles sur l'organisme modèle Zea mays. C'est un membre fondateur d'AgBioData, un consortium de ressources en ligne liées à l'agriculture qui s'engage à rendre les données de recherche en agriculture accessibles, interopérables et réutilisables. Cette plateforme facilite la navigation dans le génome du maïs, la recherche de gènes, l'analyse de QTL et l'annotation fonctionnelle.

2.2.2. Les versions du génome de Zea mays

La première version du génome de référence du maïs B73 a été publié en 2009 (Schnable et al., 2009), il s'agissait d'un assemblage BAC by BAC, qui consiste à cloner de larges fragments d'ADN dans des bactéries artificielles (BACs) afin de séquencer et assembler le génome par morceaux distincts. Depuis sa publication initiale, le génome de référence a été révisé 4 fois.

1. https://www.maizegdb.org/

La version actuelle de l'assemblage B73, ZmB73-REFERENCE-NAM-5.0 [2], a été publié en janvier 2021 (Hufford et al., 2021) . Elle a été séquencée et assemblée à partir d'un ensemble de 25 lignées consanguines par le Consortium NAM en utilisant une stratégie mate-pair sur de longs reads PacBio. Les Scaffolds ont été validés par cartographie optique BioNano, puis ordonnés à l'aide des données de liaison et des marqueurs pangénomes. De plus, des données RNA-seq mesurant l'expression des gènes dans différents tissus ont été intégrées afin d'améliorer l'annotation structurale du génome.

Les trois premiers assemblages B73 RefGen_v1, B73 RefGen_v2, et B73 RefGen_v3 étaient tous fondés sur une stratégie de séquençage par BAC et les changements entre les versions concernent l'assemblage. La version v3 a permis de compléter des régions manquantes entre les Bac de la V2 ainsi que de compléter l'annotation structurale par rapport aux gènes de la version 2 en utilisant des reads Roche/454 produits par WGS (Whole Genome Shotgun). Ainsi plusieurs contigs ont été déplacés ou inversés.

Pour l'assemblage B73 RefGen_v4, une nouvelle approche a été utilisée. C'est un assemblage *de novo* complet utilisant la technologie PacBio sur l'ADN extrait d'un descendant de l'accession utilisé pour les assemblages v1-v3. Le séquençage PacBio SMRT (Single Molecule Real Time) a été effectué au laboratoire Cold Spring Harbor avec une couverture de 60X, les scaffolds ont été créés à l'aide de cartographie optique du génome entier. Cependant comme les reads PacBio ont typiquement un taux d'erreur relativement élevé 15%, une correction des séquences PacBio a été effectuée en utilisant du séquençage long read Illumina qui a un taux d'erreur autour de 0.1%. L'annotation structurale a été réalisée en utilisant environ 111,000 long reads de transcrits PacBio provenant de six tissus de mais.

Pour aider la communauté à suivre ces évolutions dans leur travail de recherche, ces informations ainsi qu'une nomenclature des différents assemblages sont disponibles sur MaizeGDB [3]

2.2.3. Les gènes de la V2 à la V5

Pour ce projet, on utilise au départ les gènes identifiés sur la version B73 RefGen_v2 du génome de Zea mays car les QTLs de plasticité et des états stationnaires ont été cartographiés sur cette version. Étant donné qu'une analyse fonctionnelle des gènes candidats va être conduite, les gènes en v2 doivent être traduits la version V5 qui dispose d'une annotation fonctionnelle plus récente. Pour passer de la V2 à la V5, nous avons utilisé l'outil de traduction d'identifiants disponible sur MaizeGDB [4] en réalisant 3 étapes de traduction. Au lieu de demander la traduction directement en V5, j'ai réalisé la traduction de manière incrémentale en passant de la V2 à la V3, puis de la V3 à la V4 pour terminer par la traduction de la V4 en V5.

2. https://nam-genomes.github.io/
3. https://www.maizegdb.org/assembly
4. https://www.maizegdb.org/gene_center/gene#translate

2.2.4. Données génomiques

J'ai téléchargé un fichier d'annotation de référence des données du génome du maïs intitulé `ZmB73_5a.59_WGS.gff` sur MaizeGDB. Chaque ligne de ce fichier représente un élément génétique (CDS, element transposable,exon,intron, gene etc...) et permet d'extraire les informations structurales nécessaires pour la réalisation d'une analyse.

Pour obtenir un tableau avec tous les gènes identifiés pour l'espèce *Zea mays* selon l'assemblage B73 RefGen_v2, j'ai traité le fichier en utilisant des fonctions codées sur le logiciel R. J'ai chargé les données sous la forme d'un tableau, j'ai renommé les colonnes et n'ai gardé que les lignes correspondant à des gènes.

Dans ce fichier, les identifiants de gènes sont dans des lignes qui contiennent d'autres attributs. J'ai donc réaménagé ces informations pour ne garder que les identifiants des gènes. Ainsi le tableau final contient pour chaque gène sur la version v2 leur identifiant, la localisation chromosomique ainsi que la position de début et de fin.

Au début de mon stage, on m'a fourni un tableau contenant une liste de 977,459 SNPs. Ces SNPs ont été obtenus en combinant une puce Illumina Infinium HD 50K (Ganal et al., 2011), une puce Affymetrix Axiom 600K (Unterseer et al., 2014) et 500.0000 marqueurs obtenus par génotypage et par séquençage (Negro et al., 2019). Seuls les SNPs avec une fréquence allélique mineure (MAF) supérieure à 0,05 ou un taux d'hétérozygotie inférieur à 0,15 ont été gardés. Les SNPs ont été cartographiés sur la version v2 du génome de Zea mays. Dans ce tableau, les colonnes importantes sont :

— "SNP.name" l'identifiant des SNPs
— "Chromosome" qui est le chromosome où sont localisés les SNPs
— "Position" qui indique la position fixe du SNP
— "BeginPhys_Interval_DL01" est la borne inférieur de l'intervalle de confiance [5] sur la position du SNP
— "EndPhys_Interval_DL01" est la borne supérieur de l'intervalle de confiance sur la position du SNPs

2.3. Analyse fonctionnelle

On note que l'analyse fonctionnelle a été conduite sur les gènes identifiés sur l'assemblage ZmB73-REFERENCE-NAM-5.0 dont les identifiants ont été obtenu en suivant l'approche mentionnée dans la partie 2.2.3.

2.3.1. Outil MaizeMine

MaizeMine [6] est une ressource scientifique basée sur la plateforme InterMine, offrant un accès convivial aux données génomiques, protéomiques, d'interactions et de littérature pour la culture du maïs. Elle a été créee pour MaizeGDB. Cet outil offre la possibilité d'explorer et analyser les informations disponibles sur les gènes, les transcrits

5. L'intervalle de confiance correspond à la fenêtre de déséquilibre de liaison des QTLs dont la longueur a été estimé par un modèle statistique (Negro et al., 2019)
6. http://maizemine.maizegdb.org

et les voies biologiques spécifiques au maïs. L'interface utilisateur donne accès à un nombre d'outils permettant la recherche, le filtrage et la récupération de données ainsi que l'utilisation d'outils d'analyse et de visualisation des données.

Pour l'utiliser, il faut fournir une liste de gènes en v5, l'outil lance les analyses dont un test d'enrichissement fonctionnel. C'est un test hypergéométrique permettant de rechercher les termes GO enrichis. Le principe est de déterminer s'il existe des termes GO dont la proportion dans la liste de gènes d'intérêt est supérieure à la proportion dans les gènes annotés dans la v5. Puisque qu'un test est réalisé par terme GO, nous sommes dans un contexte de test multiple et pour contrôler la proportion de faux-positifs, j'ai choisi d'appliquer la correction de Benjamin-Hochberg.

2.3.2. Recherche de sites de fixation de facteurs de transcription (TFBS)

Les recherches de motifs ont été faites en utilisant l'outil PlantPLMView [7]. Le principe de l'outil Plant-PLMview est d'analyser les génomes de plantes afin de détecter des motifs d'ADN spécifiques qui sont préférentiellement situés dans les régions proximales des gènes. Ces motifs, appelés PLM (Preferentially Located Motifs), sont considérés comme des candidats prometteurs pour des sites de liaison de facteurs de transcription (TFBS). L'outil permet aux utilisateurs d'accéder aux régions proximales des gènes de 20 génomes de plantes, de rechercher des motifs d'ADN, d'utiliser un outil de correspondance pour caractériser les motifs sur-représentés, et de visualiser les résultats sous forme de graphiques. L'objectif est de faciliter l'identification de motifs d'ADN associés à des processus biologiques spécifiques dans les plantes.

Pour l'analyse, j'ai testé les 840 motifs d'ADN issus de JASPAR, AGRIS et PLACE et mis à disposition dans PlantPLMView. Pour *Zea mays*, la liste des gènes doit être fournie dans la version v4.

2.4. Le logiciel R

L'identification des gènes et la caractérisation ont été réalisé sur le logiciel R version 4.3.0 sur une plateforme x86_64-pc-linux-gnu.

3. Résultats

3.1. Méta-Analyse des QTLs

3.1.1. Pléiotropie des locus associés aux QTLs

J'ai observé que les 142 QTLs sont localisés sur 123 loci uniques. Parmi ces locis, 110 sont associés à un seul trait écophysiologique, 9 sont associés à 2 traits écophysiologiques différents et 4 sont associés à 3 traits différents. Donc la majorité des loci identifiés ne sont pas responsable de la variation de plusieurs traits écophysiologiques.

7. http://plmview.ips2.universite-paris-saclay.fr

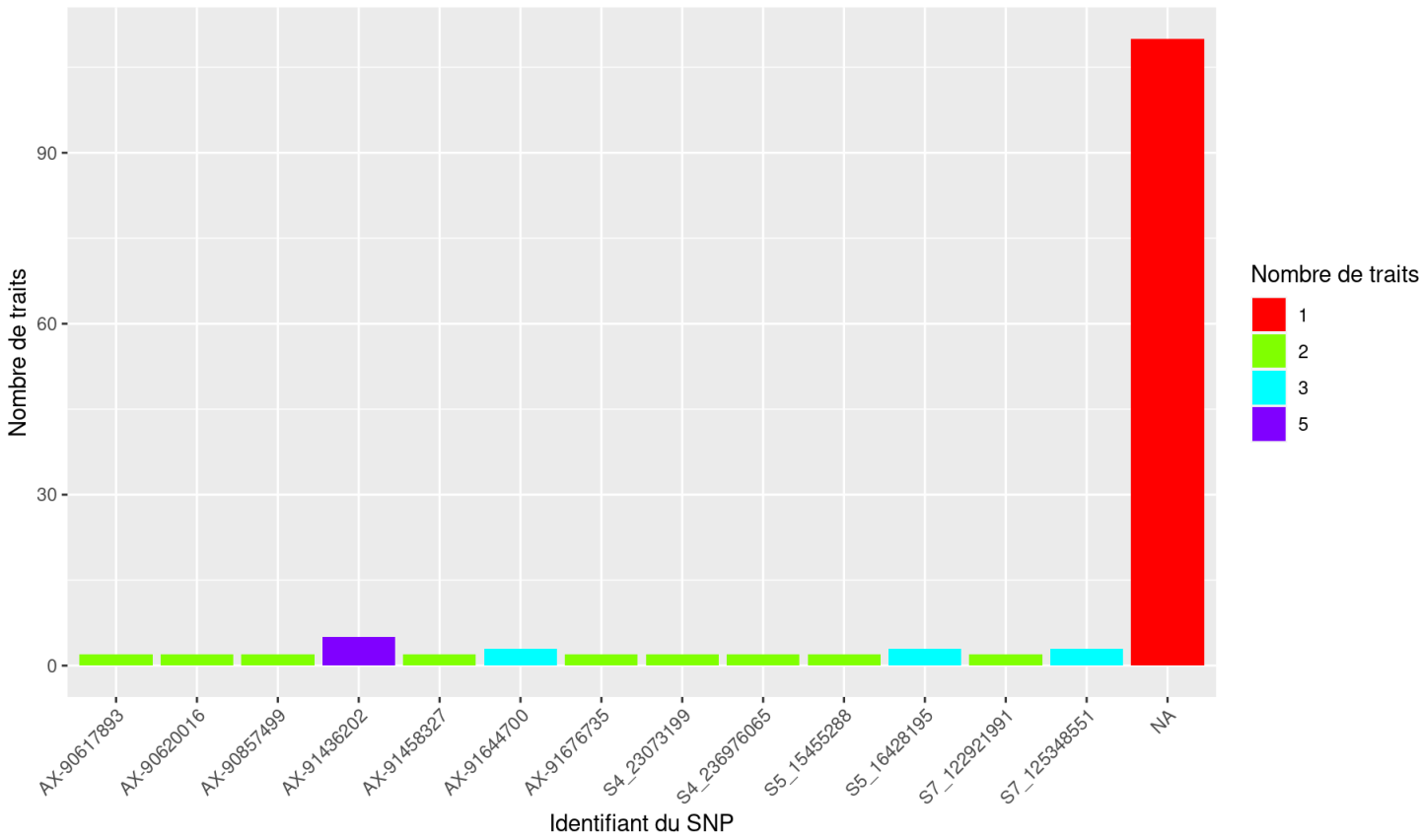


FIGURE 1 – Schéma représentant les loci en fonction du nombre de Traits écophysiologiques associés

3.1.2. Caractérisation des loci associés aux QTL en fonction du nombre de traits écophysiologiques

En se concentrant sur les 123 loci (figure1), j'ai observé que 4 loci d'état stationnaire (que j'ai décidé de les nommer par l'identifiant du SNP) sont associés à plus de 2 traits.

— Le locus AX-91436202 qui est localisé le chromosome 5 fait l'objet de 5 associations. Trois correspondent à des associations en WD pour la biomasse, la surface foliaire et l'utilisation de l'eau. Deux correspondent à des associations en WW pour la surface foliaire et l'utilisation de l'eau. Comme les conséquences du stress hydrique chez les plante se traduisent par une diminution de la croissance, une diminution de la photosynthèse et une diminution de la surface foliaire (KRAMER & BOYER, 1995), cette région identifiée par le locus "AX-91436202" semble alors jouer un rôle dans la réponse à un stress hydrique.

— Le locus AX-91644700 qui est localisé sur le chromosome 5 fait l'objet de 3 associations. Les trois correspondent à des associations en WD pour la biomasse, l'utilisation de l'eau et la surface foliaire.

— Le locus S5_16428195 qui est localisé sur le chromosome 5 fait l'objet de 3 associations : Deux correspondent à des association en WD pour la biomasse et l'utilisation de l'eau. Une correspond à une association en WW pour la surface foliaire.

— Le locus S7_125348551 qui est localisé sur le chromosome 7 fait l'objet de 3 associations. Les trois correspondent à des associations en WD pour le taux de

transpiration, l'utilisation de l'eau et la conductance stomatique.

Bien qu'il n'y a pas de loci de plasticité associé à 3 traits différents, j'ai identifié 2 loci de plasticité associés à 2 traits différents et je les ai étudiés plus précisément :

— Le locus AX-90857499 qui est localisé sur le chromosome 4 fait l'objet d'une association à la conductance stomatique et une association au taux de transpiration.

— Le locus AX-91458327 qui est localisé sur le chromosome 1 fait l'objet d'une association à la variation d'utilisation d'eau et une association à la biomasse.

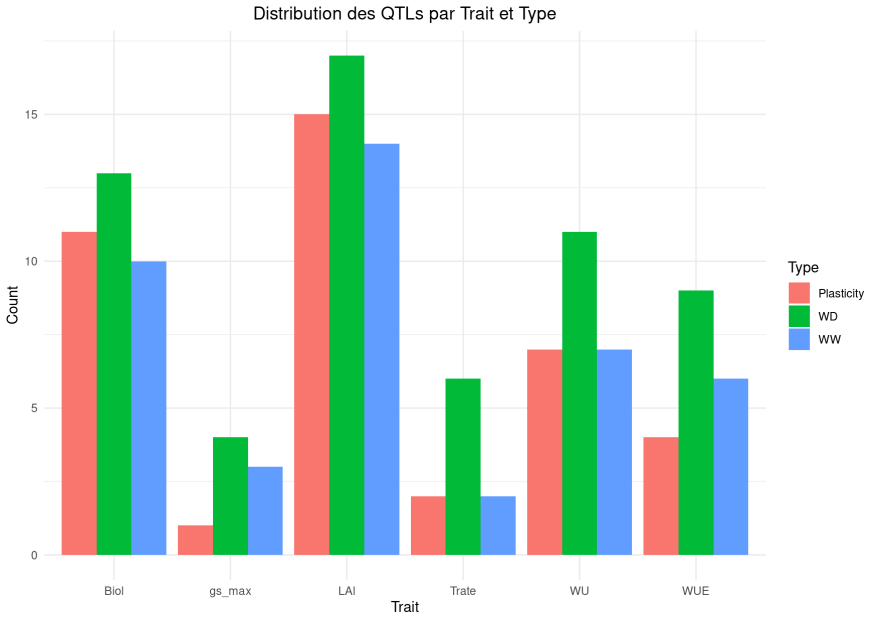


FIGURE 2 – Distribution des QTL par trait et type

SNP.name	Chromosome	Position	Trait	LogPvalue	Type
AX-91839077	7	125474566	WU	5.04	WD
AX-91645061	5	17478458	LAI	5.01	WW
AX-91735439	7	122811242	WU	5.13	WW

FIGURE 3 – Schéma représentant les QTLs ne contenant pas de gènes

Donc les loci définissant les QTL de plasticité associés à deux traits écophysiologiques différents n'ont pas de traits communs auxquels ils sont associés. Chacun est associé à un caractère écophysiologique sensible à un déficit hydrique mais le fait qu'ils soient associé à une combinaison de 2 traits différents suggère une possibilité de réponse complémentaire pour réguler les processus biologiques impliqués dans la réponse au stress hydrique.

3.1.3. Les QTLs par caractère écophysiologique et par type

La surface foliaire présente le plus grand nombre de QTLs associés parmi les différents traits étudiés. La majorité des QTLs de chaque type (plasticité, déficit en eau et bien irrigué) sont associés à ce trait en particulier. Pour chaque trait étudié, on retrouve au moins un QTL de chaque type qui est associé à la variation phénotypique de ce trait. Cette observation souligne la grande variabilité génétique de ce caractère. En condition

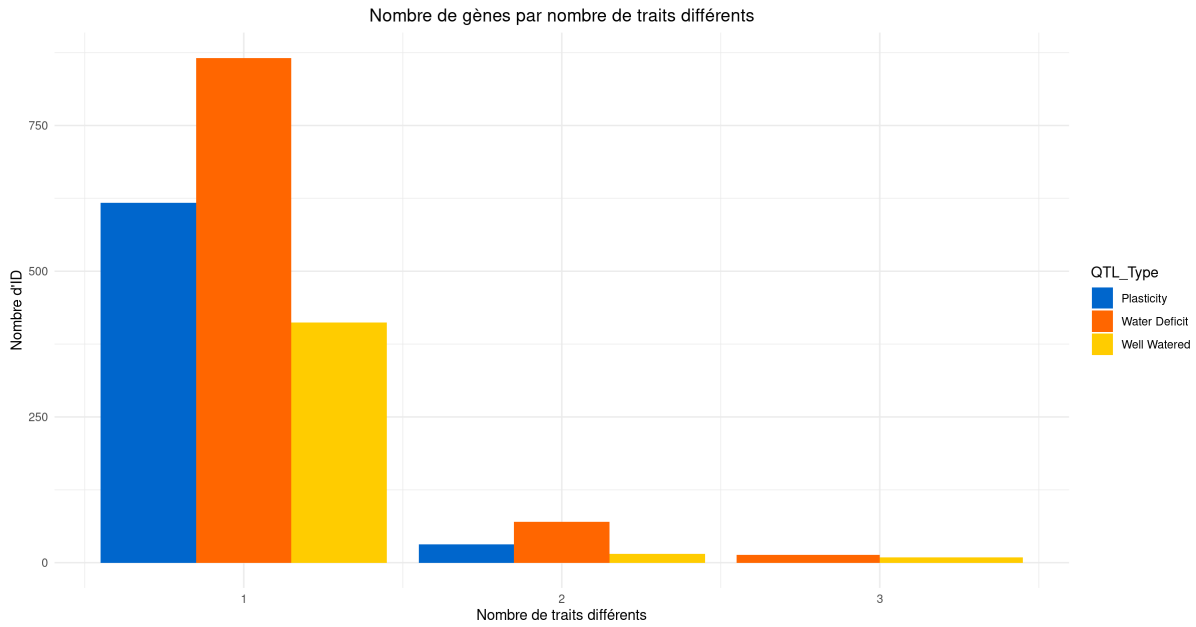


FIGURE 4 – Schéma représentant les gènes candidats en fonction du nombre de traits écophysiologiques associés en discernant le type des QTLs associés

de déficit d'eau, on observe un plus grand nombre de QTLs associés, suggérant que la réponse à un déficit en eau mobilise plus de régions génomiques (Figure 2)

3.2. Identification des gènes sous les QTLs

A partir des 142 QTLs identifiés, j'ai analysé les gènes qui se trouvent à l'intérieur de la fenêtre de déséquilibre de liaison de chacun des QTLs. Initialement, nous avons identifié 142 QTLs associés aux différents traits écophysiologiques regroupés sur 123 loci. Parmi ces 123 loci, 120 sont des régions génomiques qui contiennent des gènes (qui correspondait à 40 QTLs de plasticité et 102 QTLs d'état stationnaire). Les 3 loci restants sont situés dans des régions intergéniques. Néanmoins, cela n'écarte pas la possibilité qu'ils aient un rôle dans le processus de régulation de la réponse au stress hydrique. Ils pourraient par exemple faire partie d'un motif de fixation de facteurs de transcription ou d'un site de fixation d'une protéine régulatrice.

A partir des 120 loci, j'ai identifié 1348 gènes en v2 associés à l'état stationnaire et 648 gènes en v2 associés à la plasticité. J'ai remarqué que 62 gènes présents dans la fenêtre de déséquilibre de liaison des QTLs associés à l'état stationnaire tombent également dans la fenêtre de déséquilibre de liaison des QTLs associés à la plasticité (exemple :GRMZM2G395153). Pour trois positions sur les chromosomes 3, 4 et 8, les fenêtres de déséquilibre de liaison de certains QTLs de plasticité et d'état stationnaire se recouvrent, ainsi les gènes localisés sous ces QTLs se recouvrent également. Donc ces gènes ne respectent pas l'hypothèse Gene-regulatory.

3.3. Description des gènes sous les QTLs

Pour les gènes sous les QTLs d'état stationnaire, j'ai identifié 1250 associés à un seul trait écophysiologique, 78 gènes associés à deux traits différents et 20 gènes associés à 3 traits différents. Pour les gènes sous les QTLs de plasticité, j'ai identifié 617 gènes associés à un seul trait écophysiologique et 31 gènes associés à 2 traits différents. Ainsi il y a plus de gènes sous les QTLs d'état stationnaire que sous les QTLs de plasticité.

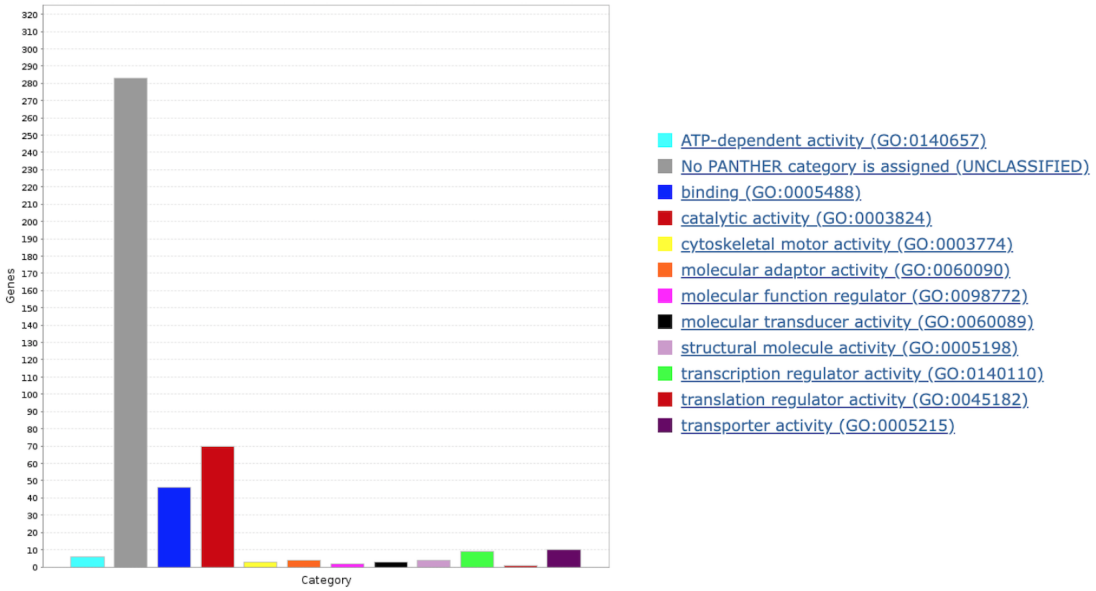


FIGURE 5 – Classification fonctionnelle des gènes associés à l'état stationnaire

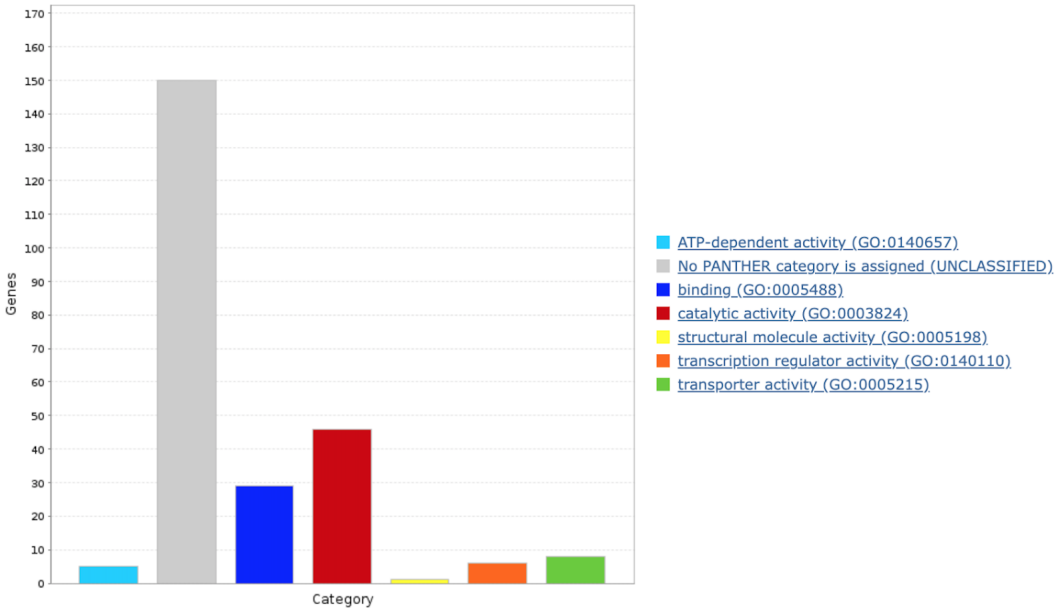


FIGURE 6 – Classification fonctionnelle des gènes associés à la plasticité

3.4. Caractérisation fonctionnelle des gènes

3.4.1. Test d'enrichissement fonctionnel

Comme cité dans la partie Versions du génome de la section Matériel et Méthode, l'analyse fonctionnelle est faite sur les gènes en version V5. Lors du passage de V2 à la V5 certains gènes fusionnent, d'autres sont séparés en deux ou plusieurs identifiants. Ainsi le nombre de gènes est différent entre l'assemblage B73 RefGen_v2 et ZmB73-REFERENCE-NAM-5.0.

En conséquence, pour les gène associés à l'état stationnaire, on passe de 1348 à 390 gènes et pour les gènes associés à la plasticité on passe de 648 à 231 gènes. Le test d'enrichissement fonctionnel (termes GO) avec une correction du FDR à 5% n'a donné aucun résultat significatif.

3.4.2. Classification fonctionnelle

Pour la plasticité et l'état stationnaire, la majorité des gènes n'ont pas été associés à une catégorie fonctionnelle Panther connue. Cependant, certains gènes des deux types appartiennent à la classification "binding" et "Catalytic Activity" (Figures 5, 6).

La fonction binding fait référence à liaison spécifique d'une protéine à une autre molécule (ADN, ARN, ligand, une autre protéine). Cette liaison joue un rôle clé dans dans divers processus biologiques tel que la signalisation cellulaire, le transport des molécules, la régulation génique, etc. Le terme "catalytic activity" fait référence à la capacité d'une protéine (enzyme) à catalyser une réaction chimique spécifique. Ainsi, retrouver ces deux catégories fonctionnelles pour les deux listes de gènes suggère que les protéines codées par ces gènes sont impliqués dans des processus métaboliques variés ainsi que des processus cellulaires essentiels.

Des gènes associés à la plasticité et à l'état stationnaire sont classifiés dans la catégorie fonctionnelle "transcription regulator activity" (Figure 5, 6). Cela signifie qu'ils codent pour des protéines impliqués dans la régulation de la transcription de gènes.

3.4.3. Identification et Analyse des régulateurs

Pour chercher des motifs reconnus par des facteurs de transcription, j'ai lancé la détection de motifs sur les gènes associés à l'état stationnaire, ensuite j'ai cherché si parmi les facteurs de transcription codés par les gènes de plasticité, il existe certains reconnaissant les motifs mentionnés précédemment. Ensuite j'ai fait la même analyse dans le sens inverse , c'est-à-dire la recherche de motifs sur les gènes de plasticité et la recherche de facteurs de transcription parmi les gènes d'état stationnaire.

3.4.3.1 Identification des facteurs de Transcription

J'ai identifié 17 facteurs de transcription parmi les 390 gènes associées aux deux conditions d'irrigation, soit 4.3% (Figure 9 en annexe) et 12 facteurs de transcription parmi les 231 gènes associés à la plasticité, soit 5.2% (Figure 10 en annexe).

annexe(10). Selon les résultats d'un test de Chi-deux réalisé à 5%, la présence de facteurs de transcription ne dépend pas du type des gènes.

3.4.3.2 Résultats PLM View

En plus des informations fournies par PlantPLMView sur les facteurs de transcription, j'ai récupéré certaines descriptions fonctionnelles des facteurs de transcriptions sur la base de données PlantTFBS [8] qui utilise les annotations fonctionnelles disponibles sur Uniprot.

Ainsi, pour 252 des 370 gènes associés à l'état stationnaire fournis en entrée à PlantPLMView, 58 TFBS uniques ont été détectés (Figure 7 en annexe). Les 118 gènes qui n'ont pas été inclus dans l'analyse sont des gènes dont le TSS n'est pas bien localisé. Parmi l'ensemble des motifs détectés pour ces gènes, j'ai identifié 4 TFBS qui sont de la même classe que les facteurs de transcription codés par 3 gènes de plasticité.

J'ai donc pu associer les motifs GAGAC et TCTCTCTC à deux facteurs de transcription codés par le gène Zm00001eb036170 qui est sous un QTL de plasticité. Ces facteurs de transcription reconnaissent et se fixent sur le motif GAGA pour activer la transcription. Ils sont également de la même famille BBR-BPC et présentent une description fonctionnelle identique. Ils sont des régulateurs de la transcription qui se fixent sur des éléments riches en répétitions GA et présents dans les régions régulatrices des gènes impliqués dans les processus de développement.

Les motifs GAAAAA et GRWAAW qui sont des motifs consensus reconnus par les protéines de la famille GT-1, j'ai trouvé un facteur de transcription codé par le gène Zm00001eb425230 qui est sous un QTl de plasticité. Ce facteur de transcription est de la famille protéique Trihélix et reconnait des motifs de la famille GT-1. Il est impliqué dans les processus de régulation de la transcription et peut agir comme un "molecular switch" en réponse à un signal lumineux.

Pour le motif GRSCCCAC, j'ai trouvé un facteur de transcription (BHLH153) codé par le gène Zm00001eb425590 qui est sous un QTL de plasticité. Ce facteur de transcription reconnaît des motifs de la classe BHLH, il se lie à une séquence d'ADN spécifique se situant en *cis* par rapport au TSS d'un gène transcrit par l'ARN polymérase II [9]. J'ai également remarqué la présence du motif CANNTG qui est enrichi dans certains gènes d'état stationnaire. Il correspond à une séquence consensus connu sous le nom de boite E à laquelle se lie les facteurs de transcription de la famille BHLH. Donc c'est également un motif auxquels le facteur de transcription codé par le gène Zm00001eb425590 peut probablement se lier.

Concernant la plasticité, j'ai trouvé 36 TFBS sur 159 gènes de plasticité (52 gènes sont écartés de l'analyse)(Figure 8 en annexe) et les gènes d'état stationnaire ne semblent pas coder pour des facteurs de transcription qui reconnaissent ces motifs. Pour les deux analyses réalisées, j'ai remarqué qu'il existe des ensemble de TFBS qui sont présents dans plusieurs gènes simultanément. Ceci peut suggérer que les gènes qui portent

8. http://planttfdb.gao-lab.org/
9. https://www.uniprot.org/uniprotkb/Q84RD0/entry

ces TFBS dans leur promoteur sont régulés de manière coordonnée. Les facteurs de transcription se lient aux TFBS et activent ou répriment simultanément l'expression de ces gènes dans le cadre d'une réponse spécifique ou un processus biologique commun. Ceci peut également suggérer l'existence d'un réseau de régulation génétique ou les facteurs de transcription agissent entre eux pour réguler l'expression des gènes.

4. Discussion - Perspectives

Dans ce projet notre point de départ était des régions génomiques associées à des caractères phénotypiques liés à la sécheresse. Ces régions génomiques ont été classées par Yacine Djabali pendant sa thèse en deux catégories : celles qui sont associées à la condition d'irrigation (WW ou WD) et celles qui sont associées à la plasticité.

Dans un premier temps, la méta-Analyse que j'ai effectuée sur ces régions génomiques (QTL) a permis d'identifier certaines régions qui sont responsables de la variation de deux ou trois traits écophysiologiques différents liés à la réponse à la sécheresse. Ceci suggère une corrélation génétique entre ces caractères. Pour les loci définissant les QTLs de plasticité qui sont associés à deux traits écophysiologiques, j'ai pu constaté qu'ils n'étaient pas associés à un même caractère écophysiologique. De plus comme ils sont sensibles au déficit hydrique, cela suggère une réponse complémentaire pour réguler la réponse au stress hydrique.

Par ailleurs, nous avons identifié la surface foliaire comme présentant une grande variabilité génétique puisqu'elle regroupe le plus grand nombre de QTLs associés. Nous avons également identifié la condition de déficit d'eau comme mobilisant le plus grand nombre de régions génomiques. Ces deux observations attestent de la grande variabilité génotypique ainsi que la complexité du mécanisme de réponse au stress hydrique.

Pour étudier la corrélation génétique entre les caractères écophysiologiques étudiés ainsi que les caractéristiques géniques des deux type de QTLs, nous avons identifié les gènes sous chaque QTL. A partir des 142 QTLs regroupés sur 123 loci, 3 QTL ne contiennent pas de gènes, c'est-à-dire que les SNPs représentatifs de ces loci sont dans des régions intergéniques. Pour les autres loci nous avons identifié en v2, 1348 gènes associés à l'état stationnaire et 648 gènes associés à la plasticité. Nous avons également remarqué qu'il existe un recouvrement de 62 gènes associés à la plasticité et à un état stationnaire. Ces gène ne respectent pas l'hypothèse gene-regulatory.

Concernant la description des gènes et des QTLs, il existe des gènes qui sont responsables de la variation de plus d'un caractère écophysiologique. L'état stationnaire présente une proportion plus élevé de gènes associés à plusieurs traits que la plasticité.

Pour réaliser la caractérisation fonctionnelle des gènes, il était nécessaire d'utiliser la version v5 des gènes. Pour le passage de la v2 à la v5, nous avons choisi la méthode de traduction d'Id des gènes en 3 étape. Or, en suivant cette méthode, on passe de 648 gènes de plasticité en v2 à 231 en v5 et de 1348 gène d'état stationnaire en v2 à 390 en v5. Ceci est dû aux différentes modifications qu'il y eu en terme de séquençage et d'assemblage entre la v2 et la v5. Nous aurions pu adopter une autre stratégie en

utilisant l'outil disponible sur EnsemblPlant permettant de donner des correspondances de positions dans le génome mais je n'ai pas pu mettre cette alternative en place par manque de temps.

Pour les gènes de plasticité comme pour ceux associés à un état stationnaire, aucun enrichissement fonctionnel significatif en terme GO n'a été trouvé. Ceci peut être dû à un problème de puissance du test statistique. Toutefois, les gènes ont pu être classifié dans des catégories fonctionnelle en utilisant leurs identifiants Uniprot.

Les principales catégories fonctionnelles sont communes entre les deux types de gènes à savoir la fonction liée à la fixation à l'ADN, l'activité catalytique et une activité lié à la régulation de la transcription.

Par ailleurs, nous avons tenu compte des régulateurs dans l'analyse fonctionnelle. Parmi les 58 TFBS identifiés sur les gènes d'état stationnaire, 4 motifs sont de la même classe que 3 facteurs de transcription codés par des gènes de plasticité. Tandis que les gènes d'état stationnaire ne semblent pas coder pour des facteurs de transcription reconnaissant les TFBS présents sur les gènes de plasticité. Dans les deux types de gènes, il existe des groupes de TFBS qui sont présents ensemble sur les gènes. Ceci suggère que ces gènes sont régulés de la même manière. Il serait intéressant de regarder si les TFBS identifiés sur les gènes d'état stationnaire et reconnus par des facteurs de transcription codés par les gènes de plasticité correspondent à des gènes associés à plusieurs traits écophysiologiques. Si c'est le cas cela indiquerait que les gènes de plasticité sont impliqués dans des voies de régulation des gènes d'état stationnaire affectant plusieurs traits.

Il existe un gène associé à l'état stationnaire qui code pour un facteur de transcription qui se fixe sur les promoteurs de gènes impliqués dans la réponse à un choc thermique [10]. Il serait intéressant d'explorer la piste selon laquelle ce facteur de transcription serait lui-même régulé par un facteur de transcription codé par un gène de plasticité.

Pour le mois qui me reste en stage, j'aimerais explorer et caractériser plus en profondeur les TFBS identifiés sur les deux types de gènes notamment en testant l'enrichissement des TFBS dans la fenêtre fonctionnelle des gènes par rapport à tous le génome de Zea mays pour savoir s'ils sont vraiment spécifiques aux conditions d'irrigation et à la réponse d'un déficit hydrique.

5. Conclusion

Au cours de ce stage, j'ai eu l'opportunité d'acquérir une bonne maîtrise de plusieurs outils bioinformatiques utilisés pour l'analyse de données biologiques. Ayant précédemment effectué un stage axé sur le développement d'un outil bioinformatique, cette expérience m'a permis de comprendre la distinction entre l'utilisation d'outils existants pour répondre à des questions biologiques et le processus de développement d'un nouvel outil répondant à des besoins spécifiques d'analyse de données biologiques.

10. https://www.ebi.ac.uk/ena/browser/view/ONM40079.1

De plus, ce stage m'a sensibilisé aux contraintes auxquelles on peut être confronté lors de l'utilisation de données génomiques provenant de différents assemblages et versions. J'ai réalisé l'importance de comprendre les modifications qui peuvent survenir entre les anciennes et les nouvelles versions d'un génome, ainsi que les contraintes que cela implique en termes de recherche de correspondance et de temps d'analyse.

Ces apprentissages et découvertes ont été précieux pour mon développement professionnel et me seront certainement bénéfiques dans mes futurs projets.

Bibliographie

Ganal, M. W., Durstewitz, G., Polley, A., Bérard, A., Buckler, E. S., Charcosset, A., Clarke, J. D., Graner, E.-M., Hansen, M., Joets, J., Le Paslier, M.-C., McMullen, M. D., Montalent, P., Rose, M., Schön, C.-C., Sun, Q., Walter, H., Martin, O. C., & Falque, M. (2011). A large maize (zea mays l.) SNP genotyping array : development and germplasm genotyping, and genetic mapping to compare with the b73 reference genome (L. Lukens, Éd.). *PLoS ONE*, *6*(12), e28334. https://doi.org/10.1371/journal.pone.0028334

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., Coletta, R. D., 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, janvier 16). *De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes* (preprint). Genomics. https://doi.org/10.1101/2021.01.14.426684

Kramer, P. J., & Boyer, J. S. (1995). *Water relations of plants and soils*. Academic Press.

Kusmec, A., Srinivasan, S., Nettleton, D., & Schnable, P. S. (2017). Distinct genetic architectures for phenotype means and plasticities in zea mays. *Nature Plants*, *3*(9), 715-723. https://doi.org/10.1038/s41477-017-0007-7

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

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

Qi, Y., Zhang, Q., Hu, S., Wang, R., Wang, H., Zhang, K., Zhao, H., Ren, S., Yang, Y., Zhao, F., Chen, F., & Yang, Y. (2022). Effects of high temperature and drought stresses on growth and yield of summer maize during grain filling in north china. *Agriculture*, *12*(11), 1948. https://doi.org/10.3390/agriculture12111948

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

Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., Meitinger, T., Strom, T. M., Fries, R., Pausch, H., Bertani, C., Davassi, A., Mayer, K. F., & Schön, C.-C. (2014). A powerful tool for genome analysis in maize : development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics*, *15*(1), 823. https://doi.org/10.1186/1471-2164-15-823

Via, S., Gomulkiewicz, R., De Jong, G., Scheiner, S. M., Schlichting, C. D., & Van Tienderen, P. H. (1995). Adaptive phenotypic plasticity : consensus and controversy. *Trends in Ecology & Evolution*, *10*(5), 212-217. https://doi.org/10.1016/S0169-5347(00)89061-8

Annexes

FIGURE 7 – schéma représentatif des TFBS détecté sur les gènes d'état Stationnaire

FIGURE 8 – schéma représentatif des TFBS détecté sur les gènes de Plasticité

Gene > Gene ID	Gene > Biotype	Gene > Description	Gene > Length	Gene > Chromosome ID
Zm00001eb007490	protein_coding	GNAT transcription factor	2122	chr1
Zm00001eb083590	protein_coding	Transcription factor E2FB	9645	chr2
Zm00001eb083640	protein_coding	Transcription repressor	1501	chr2
Zm00001eb130300	protein_coding	Putative mediator of RNA polymerase II transcription subunit 15 isoform X2	5543	chr3
Zm00001eb136300	protein_coding	Transcription initiation factor TFIID subunit 8	4694	chr3
Zm00001eb136360	protein_coding	Transcription termination factor MTERF15	19161	chr3
Zm00001eb136490	protein_coding	probable mediator of RNA polymerase II transcription subunit 37e	4098	chr3
Zm00001eb138450	protein_coding	CCR4-NOT transcription complex subunit 11	9941	chr3
Zm00001eb145660	protein_coding	PISTILLATA-like MADS-box transcription factor (Fragment)	2624	chr3
Zm00001eb159390	protein_coding	Heat stress transcription factor C-1a	2484	chr3
Zm00001eb239430	protein_coding	Ethylene-responsive transcription factor ERF043	1464	chr5
Zm00001eb248900	protein_coding	Putative homeodomain-like transcription factor superfamily protein isoform 1	3944	chr5
Zm00001eb251230	protein_coding	PLATZ transcription factor	2801	chr5
Zm00001eb251390	protein_coding	Putative transcription factor KAN2	5309	chr5
Zm00001eb251650	protein_coding	transcription factor bHLH91-like	3767	chr5
Zm00001eb251670	protein_coding	Myb family transcription factor PHL11	3141	chr5
Zm00001eb300910	protein_coding	Putative transcriptional regulator SLK2	7099	chr7

FIGURE 9 – Tableau des gènes sous les QTLs d'état stationnaire codant pour des facteurs de transcription

Gene > Gene ID	Gene > Biotype	Gene > Description	Gene > Length	Gene > Chromosome ID
Zm00001eb010770	protein_coding	Nuclear transcription factor Y subunit C-4	2350	chr1
Zm00001eb015440	protein_coding	Ethylene signal transcription factor	4074	chr1
Zm00001eb036170	protein_coding	GAGA-binding transcriptional activator	1953	chr1
Zm00001eb124790	protein_coding	MADS domain class transcription factor	16124	chr3
Zm00001eb136300	protein_coding	Transcription initiation factor TFIID subunit 8	4694	chr3
Zm00001eb136360	protein_coding	Transcription termination factor MTERF15	19161	chr3
Zm00001eb136490	protein_coding	probable mediator of RNA polymerase II transcription subunit 37e	4098	chr3
Zm00001eb272490	protein_coding	MADS-box transcription factor 16	4751	chr6
Zm00001eb425230	protein_coding	Trihelix transcription factor GT-1	3759	chr10
Zm00001eb425590	protein_coding	Transcription factor bHLH153	4370	chr10
Zm00001eb425660	protein_coding	NLP transcription factor (Fragment)	5718	chr10
Zm00001eb425720	protein_coding	Transcription activator factor	6240	chr10

FIGURE 10 – Tableau des gènes associés à la plasticité codant pour des facteurs de transcription

Résumé

Ce projet vise à caractériser des gènes impliqués dans la réponse du maïs au stress hydrique. Le maïs est largement cultivé dans le monde, mais il est sensible à la sécheresse, ce qui peut réduire considérablement les rendements dans un contexte mondial de réchauffement climatique.

Le projet s'appuie sur des études d'association génétique pour identifier les régions génomiques (QTLs) associées à des caractères phénotypiques liés à la sécheresse. Lors de sa thèse, Yacine Djabali a identifié des régions génomiques associées à la condition d'irrigation et des régions génomiques associées à la réponse du maïs à un déficit hydrique. L'objectif de ce stage a été de caractériser les régions génomiques ainsi que les gènes associés à ces régions.

Pendant ce stage, les gènes couverts par les QTLs ont été identifiés en utilisant des intervalles de confiance et les positions des gènes du maïs. Ensuite, les annotations fonctionnelles des gènes ont été recherchées dans différentes bases de données afin de réaliser une analyse d'enrichissement fonctionnel. La dernière partie du stage a consisté à étudier les motifs cis-régulateurs pour mettre en évidence des mécanismes de régulation impliqués dans la réponse au déficit hydrique.

Abstract

This project aims to characterize genes involved in maize response to water stress. Maize is a widely cultivated crop worldwide, but it is susceptible to drought, which can significantly reduce yields in the global context of climate change.

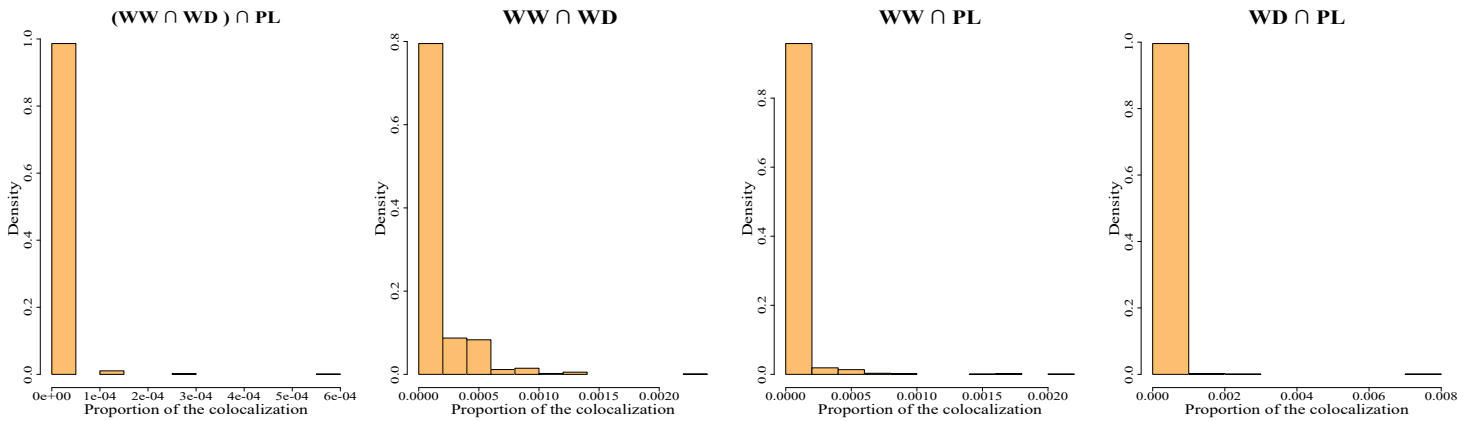
The project relies on genetic association studies to identify genomic regions (QTLs) associated with drought-related phenotypic traits. During his thesis, Yacine Djabali identified genomic regions associated with irrigated conditions and genomic regions associated with maize response to water deficit. The objective of this internship was to characterize these genomic regions and the genes associated with them.

During the internship, genes covered by the QTLs were identified using confidence intervals and maize gene positions. Functional annotations of the genes were then searched in various databases to perform functional enrichment analysis. The final part of the internship involved studying cis-regulatory motifs to highlight regulatory mechanisms involved in the response to water deficit.

Suplementary material Chapter 3

Supplementary figures and tables related to section 3.2

a.



b.

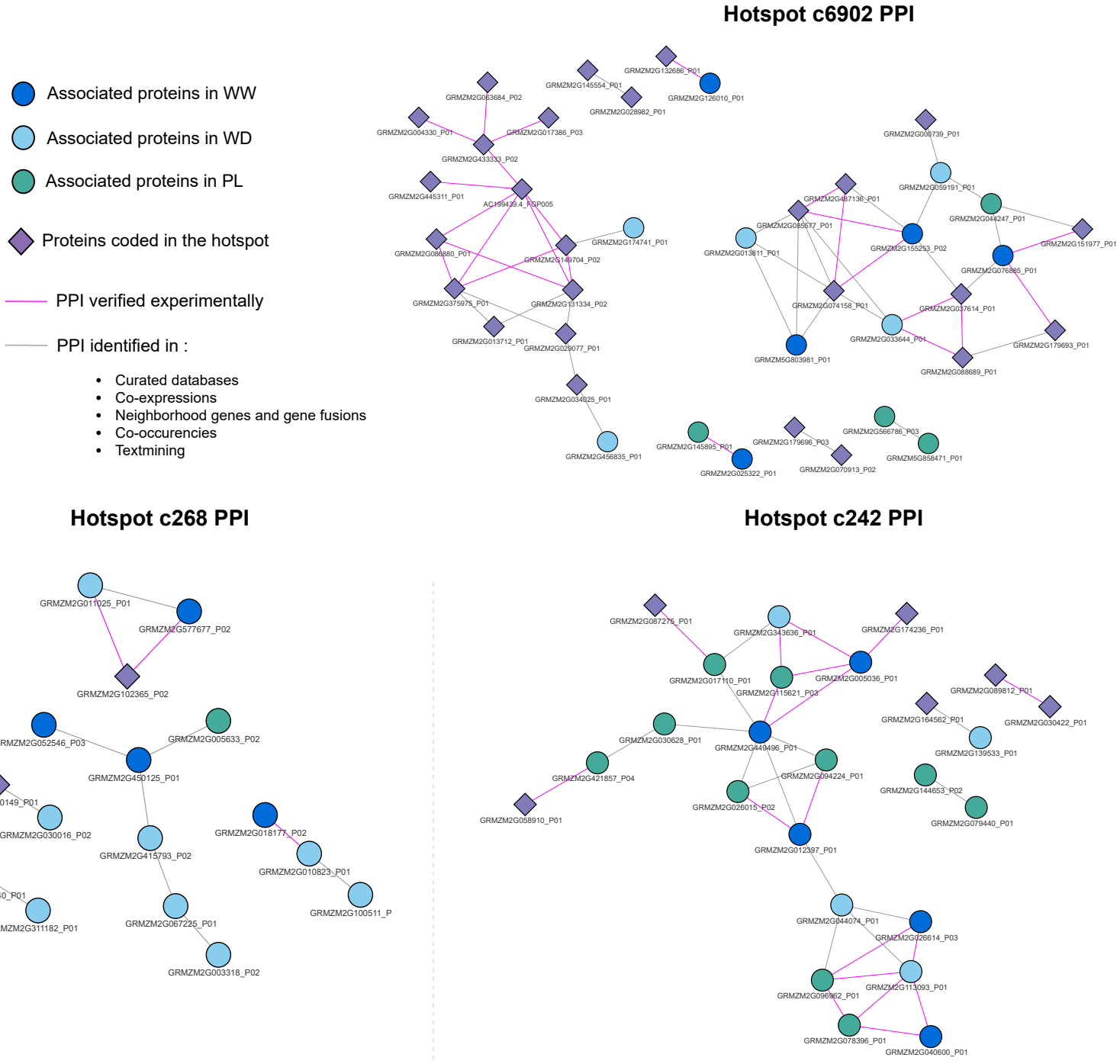


Fig. S1: a, Proportion of proteins having at least one localization between WW, WD, and/or PL pQTLs. **b**, c6902, c242, and c268 hotspots PPIs. The deep blue, light blue, and green nodes represent WW, WD, and PL proteins associated with the hotspots, and the purple diamond nodes represent proteins coded by genes located in the hotspot. The size of nodes represents their betweenness. The edges in purple inform that the interaction between two proteins was experimentally verified, and edges in grey correspond to functional interactions..

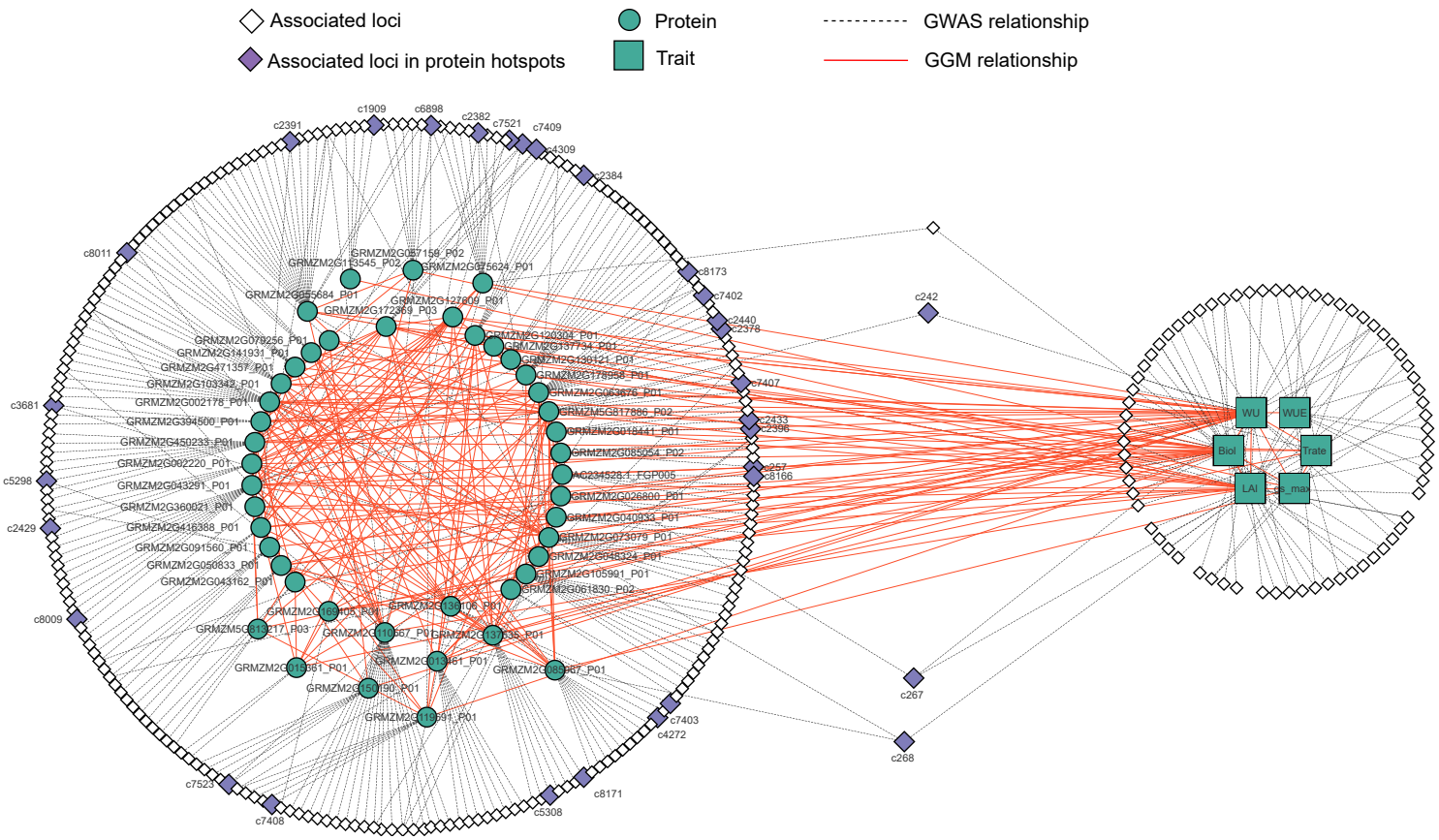


Fig. S2: Representation of the multi-scaled network selected in PL. The green square represents the traits, the green circles the proteins, the purple diamond the associated loci located in hotspots, and the white diamonds the associated loci. The edges colorized in red represent GGM relationship between traits and proteins, and those colorized in dashed gray represent the GWAS relationship between associated loci and proteins or traits.

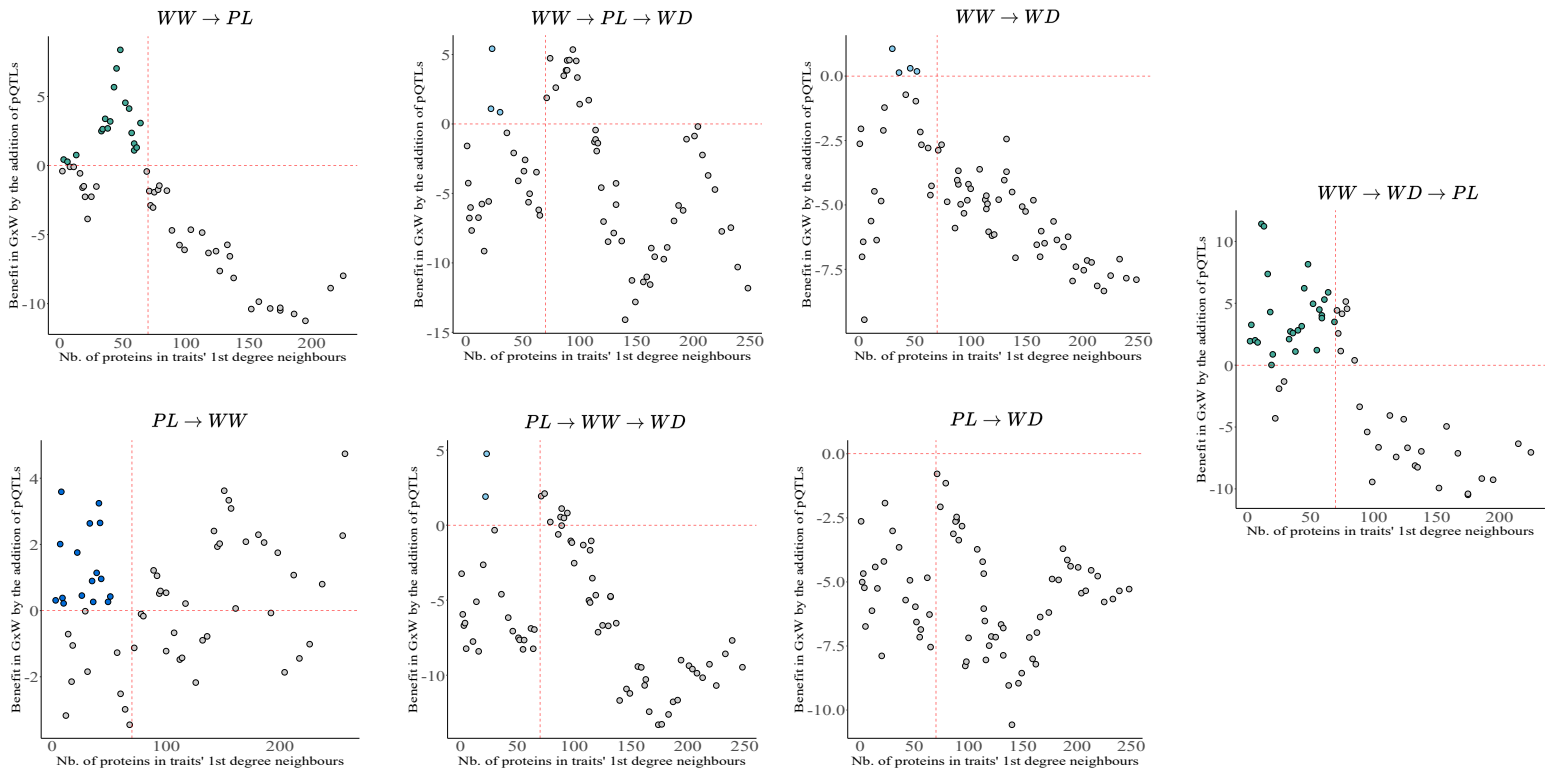


Fig. S3: Gain in GxW variance captured by adding pQTLs according to the number of proteins present in the traits' first-degree neighbors at each lambda for the incremental paths giving the models M9, M10, M11, M12, M14, M15 and M16

Table S1 : Genes located into M5 and M2 QTLs disequilibrium windows.

Gene_ID_V3	Gene_ID_V5	Gene_Descr	Chr	QTLs_type
GRMZM2G003466	Zm00001eb007000	AP domain DRE binding factor	chr1	M5_QTLs
GRMZM2G028136	Zm00001eb007010	POLIIIAc domain-containing protein	chr1	M5_QTLs
GRMZM2G028247	Zm00001eb007020	Carboxypeptidase	chr1	M5_QTLs
GRMZM2G432566	Zm00001eb007030	E3 ubiquitin-protein ligase hel2	chr1	M5_QTLs
NA	Zm00001eb007050	DNA polymerase sigma subunit	chr1	M5_QTLs
GRMZM2G155086	Zm00001eb007070	Pentatricopeptide repeat-containing family protein	chr1	M5_QTLs
GRMZM5G808775	Zm00001eb007090	Tetratricopeptide repeat (TPR)-like superfamily protein	chr1	M5_QTLs
GRMZM2G114234	Zm00001eb007440	Sulfhydryl oxidase	chr1	M5_QTLs
NA	Zm00001eb007450		chr1	M5_QTLs
GRMZM2G114232	Zm00001eb007460	Sulfhydryl oxidase	chr1	M5_QTLs
GRMZM2G114190	Zm00001eb007470	Protein SNOWY COTYLEDON 3	chr1	M5_QTLs
GRMZM2G114184	Zm00001eb007490	GNAT transcription factor	chr1	M5_QTLs
NA	Zm00001eb007500		chr1	M5_QTLs
GRMZM2G070716	Zm00001eb008370	NADH-ubiquinone oxidoreductase B16.6 subunit	chr1	M5_QTLs
GRMZM2G070442	Zm00001eb008380	Protein UPSTREAM OF FLC	chr1	M5_QTLs
GRMZM2G069765	Zm00001eb008400	T-complex protein 1 subunit gamma	chr1	M5_QTLs
NA	Zm00001eb008410		chr1	M5_QTLs
GRMZM2G107774	Zm00001eb008420		chr1	M5_QTLs
GRMZM2G107805	Zm00001eb008430	Pentatricopeptide repeat-containing protein	chr1	M5_QTLs
GRMZM2G107815	Zm00001eb008440	Heparanase-like protein 1	chr1	M5_QTLs
GRMZM2G107854	Zm00001eb008450	Hexosyltransferase	chr1	M5_QTLs
GRMZM2G107872	Zm00001eb008460	probable leucine-rich repeat receptor-like protein kinase At1g35710	chr1	M5_QTLs
NA	Zm00001eb008470	plant intracellular Ras-group-related LRR protein 7	chr1	M5_QTLs
GRMZM2G015804	Zm00001eb008530	4-hydroxy-7-methoxy-3-oxo-3	chr1	M5_QTLs
GRMZM2G130987	Zm00001eb008550	Protein transport protein Sec61 subunit alpha	chr1	M5_QTLs
GRMZM2G098714	Zm00001eb008630	Replication protein A subunit	chr1	M5_QTLs
GRMZM2G098667	Zm00001eb008640	Digalactosyldiacylglycerol synthase	chr1	M5_QTLs
GRMZM2G472693	Zm00001eb008650	Zinc finger CCCH domain-containing protein 55	chr1	M5_QTLs
GRMZM2G173067	Zm00001eb008660	TPT domain-containing protein	chr1	M5_QTLs
NA	Zm00001eb009580	Maf family protein	chr1	M5_QTLs
GRMZM2G135877	Zm00001eb009600	Protein DETOXIFICATION	chr1	M5_QTLs
GRMZM2G360234	Zm00001eb010340		chr1	M5_QTLs
GRMZM2G059260	Zm00001eb010350		chr1	M5_QTLs
GRMZM2G147942	Zm00001eb010360	subtilisin-like protease SBT1.6	chr1	M5_QTLs
GRMZM2G077942	Zm00001eb010370	Actin-depolymerizing factor 5	chr1	M5_QTLs
GRMZM2G102483	Zm00001eb010730	EF-Hand containing protein	chr1	M5_QTLs
GRMZM2G102347	Zm00001eb010740		chr1	M5_QTLs
NA	Zm00001eb010750		chr1	M5_QTLs
GRMZM2G089812	Zm00001eb010770	Nuclear transcription factor Y subunit C-4	chr1	M5_QTLs
GRMZM2G171736	Zm00001eb010780		chr1	M5_QTLs
GRMZM2G058910	Zm00001eb010790	V-type proton ATPase subunit a	chr1	M5_QTLs
GRMZM2G359879	Zm00001eb010800	Synaptotagmin-4 isoform B	chr1	M5_QTLs
GRMZM2G058870	Zm00001eb010810	Guanine-nucleotide-exchange protein	chr1	M5_QTLs
GRMZM2G355450	Zm00001eb010820	Protein kinase domain-containing protein	chr1	M5_QTLs
GRMZM2G087275	Zm00001eb010860	Long chain base biosynthesis protein 1	chr1	M5_QTLs
GRMZM2G097788	Zm00001eb010890	SANT domain-containing protein	chr1	M5_QTLs
GRMZM2G019567	Zm00001eb010910	Serine/threonine-protein kinase WAG1	chr1	M5_QTLs
GRMZM2G049159	Zm00001eb012280	Scarecrow-like protein 9	chr1	M5_QTLs
GRMZM2G350001	Zm00001eb012290		chr1	M5_QTLs
GRMZM2G174644	Zm00001eb012300	Oxysterol-binding protein-related protein 3C	chr1	M5_QTLs

GRMZM2G110309	Zm00001eb012310	Initiator binding protein1	chr1	M5_QTLs
GRMZM2G087600	Zm00001eb012330	Adipocyte plasma membrane-associated protein-like	chr1	M5_QTLs
GRMZM2G087459	Zm00001eb012340	Protein kinase APK1A chloroplastic	chr1	M5_QTLs
GRMZM2G112925	Zm00001eb012350	Protein-serine/threonine phosphatase	chr1	M5_QTLs
GRMZM2G113002	Zm00001eb012360	E3 ubiquitin-protein ligase XB3	chr1	M5_QTLs
GRMZM2G043030	Zm00001eb012380	S-acyltransferase	chr1	M5_QTLs
GRMZM2G173251	Zm00001eb013280		chr1	M5_QTLs
GRMZM2G024159	Zm00001eb013300		chr1	M5_QTLs
NA	Zm00001eb013320		chr1	M5_QTLs
NA	Zm00001eb013330		chr1	M5_QTLs
GRMZM2G074401	Zm00001eb013340	Plastidial delta 15 desaturase	chr1	M5_QTLs
GRMZM2G305267	Zm00001eb013350		chr1	M5_QTLs
AC212219.3_FG005	Zm00001eb013360	Jasmonate-regulated gene 21	chr1	M5_QTLs
GRMZM5G838147	Zm00001eb013370	rhodanese-like domain-containing protein 10	chr1	M5_QTLs
GRMZM2G122740	Zm00001eb015310	protein ALTERED PHOSPHATE STARVATION RESPONSE 1	chr1	M5_QTLs
GRMZM2G030272	Zm00001eb015320	WRKY domain-containing protein	chr1	M5_QTLs
GRMZM2G018728	Zm00001eb015330	14 kDa zinc-binding protein	chr1	M5_QTLs
GRMZM2G018678	Zm00001eb015340	NF-kappa-B inhibitor-like protein 2 isoform 1	chr1	M5_QTLs
GRMZM2G145065	Zm00001eb015350	MADS box interactor-like	chr1	M5_QTLs
GRMZM2G145065	Zm00001eb015370		chr1	M5_QTLs
GRMZM2G145024	Zm00001eb015380	ELMO domain-containing protein	chr1	M5_QTLs
GRMZM2G145008	Zm00001eb015390	Arf-GAP domain-containing protein	chr1	M5_QTLs
GRMZM2G144997	Zm00001eb015400	J domain-containing protein	chr1	M5_QTLs
GRMZM2G174696	Zm00001eb015420	Mitochondrial import receptor subunit TOM40-1	chr1	M5_QTLs
GRMZM2G174680	Zm00001eb015430	AAI domain-containing protein	chr1	M5_QTLs
GRMZM2G033570	Zm00001eb015440	Ethylene signal transcription factor	chr1	M5_QTLs
NA	Zm00001eb015450	Putative disease resistance protein RGA3 (Fragment)	chr1	M5_QTLs
GRMZM2G009690	Zm00001eb015460	Exostosin domain-containing protein	chr1	M5_QTLs
GRMZM2G407825	Zm00001eb015510	P-type phospholipid transporter	chr1	M5_QTLs
GRMZM2G107481	Zm00001eb015530	Phospholipid-transporting ATPase	chr1	M5_QTLs
GRMZM2G107463	Zm00001eb015540	RING/U-box superfamily protein	chr1	M5_QTLs
GRMZM2G102365	Zm00001eb015550	thioredoxin-like 1-1	chr1	M5_QTLs
GRMZM2G042615	Zm00001eb015560	U11/U12 small nuclear ribonucleoprotein 65 kDa protein	chr1	M5_QTLs
GRMZM2G031001	Zm00001eb015570	NAC domain-containing protein	chr1	M5_QTLs
GRMZM2G162690	Zm00001eb021270	Trehalase	chr1	M5_QTLs
GRMZM2G087243	Zm00001eb021400	Armadillo/beta-catenin-like repeat family protein	chr1	M5_QTLs
NA	Zm00001eb021420		chr1	M5_QTLs
GRMZM2G053742	Zm00001eb022970	PPC domain-containing protein	chr1	M5_QTLs
GRMZM2G053720	Zm00001eb022980	Proline dehydrogenase	chr1	M5_QTLs
GRMZM2G449779	Zm00001eb028000	Pumilio RNA-binding repeat	chr1	M5_QTLs
GRMZM2G303909	Zm00001eb028010	G-type lectin S-receptor-like serine/threonine-protein kinase LECRK2	chr1	M5_QTLs
GRMZM5G811095	Zm00001eb028020	MFS domain-containing protein	chr1	M5_QTLs
GRMZM5G828630	Zm00001eb028030	Alanine aminotransferase 2	chr1	M5_QTLs
NA	Zm00001eb028040		chr1	M5_QTLs
GRMZM2G117240	Zm00001eb031360	Hcy-binding domain-containing protein	chr1	M5_QTLs
GRMZM2G132763	Zm00001eb034800	Leucine-rich repeat receptor-like protein kinase PEPR1	chr1	M5_QTLs
GRMZM2G141320	Zm00001eb034810	Monogalactosyldiacylglycerol synthase	chr1	M5_QTLs
GRMZM2G132623	Zm00001eb034820	PfkB domain-containing protein	chr1	M5_QTLs
GRMZM2G132704	Zm00001eb034840	Nucleolar protein gar2-related isoform 2	chr1	M5_QTLs
GRMZM2G132748	Zm00001eb034850	Complex I subunit	chr1	M5_QTLs
GRMZM2G166355	Zm00001eb035660	Oligouridylate-binding protein 1B	chr1	M5_QTLs
GRMZM2G166407	Zm00001eb035670	Tubby-like F-box protein 10	chr1	M5_QTLs
GRMZM2G149480	Zm00001eb035680	probable mitochondrial adenine nucleotide transporter BTL1	chr1	M5_QTLs

GRMZM2G149474	Zm00001eb035690	Exocyst subunit Exo70 family protein (Fragment)	chr1	M5_QTLs
GRMZM2G449123	Zm00001eb035700	Pumilio homolog 5	chr1	M5_QTLs
GRMZM2G096585	Zm00001eb036130	Peptidylprolyl isomerase	chr1	M5_QTLs
GRMZM2G096806	Zm00001eb036140	FK506-binding protein 2-1	chr1	M5_QTLs
GRMZM2G179354	Zm00001eb036160	Glucan endo-1	chr1	M5_QTLs
GRMZM2G179366	Zm00001eb036170	GAGA-binding transcriptional activator	chr1	M5_QTLs
GRMZM2G318803	Zm00001eb038540	Histone-lysine N-methyltransferase SUVR4	chr1	M5_QTLs
GRMZM2G449177	Zm00001eb039660	Phosphatidylinositol-3	chr1	M5_QTLs
GRMZM2G149414	Zm00001eb039670	NADH dehydrogenase [ubiquinone] iron-sulfur protein 5-B	chr1	M5_QTLs
GRMZM2G149442	Zm00001eb039710	S-acyltransferase	chr1	M5_QTLs
GRMZM2G015588	Zm00001eb039720	7	chr1	M5_QTLs
NA	Zm00001eb039730		chr1	M5_QTLs
GRMZM2G171830	Zm00001eb087540		chr2	M5_QTLs
GRMZM2G154235	Zm00001eb088440		chr2	M5_QTLs
GRMZM2G154156	Zm00001eb088450	Zinc finger CCCH domain-containing protein 59	chr2	M5_QTLs
GRMZM2G023028	Zm00001eb088460	DNA-directed RNA polymerase subunit	chr2	M5_QTLs
GRMZM2G325749	Zm00001eb088470	50S ribosomal protein L18	chr2	M5_QTLs
GRMZM2G036708	Zm00001eb088500	Cysteine synthase	chr2	M5_QTLs
GRMZM2G029141	Zm00001eb088510	Mitochondrial inner membrane protease subunit 2	chr2	M5_QTLs
NA	Zm00001eb088520	RNA-dependent RNA polymerase	chr2	M5_QTLs
AC217553.3_FG008	Zm00001eb088530	RNA-dependent RNA polymerase	chr2	M5_QTLs
NA	Zm00001eb088540		chr2	M5_QTLs
GRMZM2G058597	Zm00001eb088560	Mitochondrial inner membrane protease subunit 2	chr2	M5_QTLs
GRMZM2G058591	Zm00001eb088570	RNA-dependent RNA polymerase	chr2	M5_QTLs
AC217553.3_FG001	Zm00001eb088580		chr2	M5_QTLs
GRMZM2G072388	Zm00001eb088840	Glyoxylate/hydroxypyruvate reductase HPR3	chr2	M5_QTLs
GRMZM2G074083	Zm00001eb088850	CBS domain-containing protein	chr2	M5_QTLs
GRMZM2G367367	Zm00001eb088860	Putative inactive tRNA-specific adenosine deaminase-like protein 3	chr2	M5_QTLs
GRMZM2G067096	Zm00001eb088870	Peroxidase	chr2	M5_QTLs
GRMZM2G131597	Zm00001eb088880	AAR2 protein family	chr2	M5_QTLs
GRMZM2G120132	Zm00001eb088900		chr2	M5_QTLs
NA	Zm00001eb088910	Putative transferase transferring glycosyl groups	chr2	M5_QTLs
GRMZM2G108686	Zm00001eb088930	Coa1 domain-containing protein	chr2	M5_QTLs
GRMZM2G081745	Zm00001eb088940	Sec23/Sec24 protein transport family protein isoform 1	chr2	M5_QTLs
GRMZM2G141383	Zm00001eb088950	Flavin-containing monooxygenase	chr2	M5_QTLs
GRMZM2G116634	Zm00001eb112830	Putative Non-intrinsic ABC protein 14	chr2	M5_QTLs
GRMZM2G106560	Zm00001eb112840	WRKY domain-containing protein	chr2	M5_QTLs
GRMZM2G413857	Zm00001eb121200	Pollen-specific protein SF21	chr3	M5_QTLs
GRMZM2G115834	Zm00001eb121210	MAR-binding filament-like protein 1-1	chr3	M5_QTLs
GRMZM2G022897	Zm00001eb124780	Di-glucose binding protein with Leucine-rich repeat domain	chr3	M5_QTLs
GRMZM2G320549	Zm00001eb124790	MADS domain class transcription factor	chr3	M5_QTLs
GRMZM2G161154	Zm00001eb124800		chr3	M5_QTLs
GRMZM2G410951	Zm00001eb136230	Formin-like protein 12	chr3	M5_QTLs
GRMZM2G385428	Zm00001eb136250		chr3	M5_QTLs
GRMZM2G332703	Zm00001eb136270	TPX2 domain-containing protein	chr3	M5_QTLs
GRMZM2G010628	Zm00001eb136280		chr3	M5_QTLs
NA	Zm00001eb136290	RGG repeats nuclear RNA binding protein B isoform A	chr3	M5_QTLs
GRMZM2G151717	Zm00001eb136300	Transcription initiation factor TFIID subunit 8	chr3	M5_QTLs
GRMZM2G092223	Zm00001eb136330	Vacuolar amino acid transporter 1	chr3	M5_QTLs
GRMZM2G159475	Zm00001eb136350	GRAS domain-containing protein	chr3	M5_QTLs
GRMZM2G177019	Zm00001eb136360	Transcription termination factor MTERF15	chr3	M5_QTLs
NA	Zm00001eb136370		chr3	M5_QTLs
GRMZM2G454189	Zm00001eb136380		chr3	M5_QTLs

GRMZM2G151720	Zm00001eb136390		chr3	M5_QTLs
GRMZM2G076499	Zm00001eb136430		chr3	M5_QTLs
GRMZM2G072435	Zm00001eb136440	Pollen-specific protein SF21	chr3	M5_QTLs
NA	Zm00001eb136450	RING-type E3 ubiquitin transferase	chr3	M5_QTLs
GRMZM2G176063	Zm00001eb136460	Dof zinc finger protein DOF2.2	chr3	M5_QTLs
NA	Zm00001eb136470		chr3	M5_QTLs
GRMZM2G340251	Zm00001eb136490	probable mediator of RNA polymerase II transcription subunit 37e	chr3	M5_QTLs
AC210168.4_FG003	Zm00001eb136500		chr3	M5_QTLs
GRMZM2G066650	Zm00001eb136530	Importin subunit beta-1	chr3	M5_QTLs
GRMZM2G066650	Zm00001eb136540	Importin N-terminal domain-containing protein	chr3	M5_QTLs
GRMZM2G016586	Zm00001eb136610		chr3	M5_QTLs
GRMZM2G067456	Zm00001eb136620	Putative Nucleic acid-binding	chr3	M5_QTLs
GRMZM2G067476	Zm00001eb136630	Translocon-associated protein subunit beta	chr3	M5_QTLs
GRMZM2G052093	Zm00001eb136640	Aspartic proteinase nepenthesin-1	chr3	M5_QTLs
GRMZM2G060680	Zm00001eb136650	Aspartic proteinase nepenthesin-1	chr3	M5_QTLs
NA	Zm00001eb136680		chr3	M5_QTLs
NA	Zm00001eb136690	Protein kinase domain-containing protein	chr3	M5_QTLs
NA	Zm00001eb136700	Protein kinase domain-containing protein	chr3	M5_QTLs
AC209381.3_FG007	Zm00001eb141850	RING-H2 finger protein ATL3C	chr3	M5_QTLs
GRMZM2G132506	Zm00001eb141870	Photosystem II reaction center Psb28 protein	chr3	M5_QTLs
GRMZM2G132509	Zm00001eb141880	Xylanase inhibitor TAXI-IV	chr3	M5_QTLs
NA	Zm00001eb141890	MSP domain-containing protein	chr3	M5_QTLs
NA	Zm00001eb145660	PISTILLATA-like MADS-box transcription factor (Fragment)	chr3	M5_QTLs
GRMZM2G110195	Zm00001eb145670	Amino acid permease 3	chr3	M5_QTLs
GRMZM2G110277	Zm00001eb145680	NAD(P)H-quinone oxidoreductase subunit N	chr3	M5_QTLs
GRMZM2G110398	Zm00001eb145690	Probable magnesium transporter	chr3	M5_QTLs
GRMZM2G146887	Zm00001eb156100	Glutathione-S-transferase Cla47	chr3	M5_QTLs
GRMZM2G426415	Zm00001eb167530	Glycosyltransferase	chr4	M5_QTLs
GRMZM2G117627	Zm00001eb169300	Galactolipase DONGLE	chr4	M5_QTLs
GRMZM2G161521	Zm00001eb170470		chr4	M5_QTLs
GRMZM2G470524	Zm00001eb194080	Glycosyltransferase	chr4	M5_QTLs
NA	Zm00001eb194100	RRM domain-containing protein	chr4	M5_QTLs
GRMZM2G161553	Zm00001eb194380	Exocyst complex component Sec6	chr4	M5_QTLs
GRMZM2G161566	Zm00001eb194410	Shikimate kinase	chr4	M5_QTLs
GRMZM2G325575	Zm00001eb195010	Ferritin	chr4	M5_QTLs
GRMZM2G092167	Zm00001eb195020	Pentatricopeptide repeat-containing protein At5g12100	chr4	M5_QTLs
GRMZM2G155384	Zm00001eb208050	26S proteasome non-ATPase regulatory subunit 2 homolog	chr4	M5_QTLs
GRMZM2G102475	Zm00001eb216800	Pinin/SDK/memA/ protein conserved region containing protein	chr5	M5_QTLs
GRMZM2G102601	Zm00001eb216810	Ethylene receptor	chr5	M5_QTLs
GRMZM2G102616	Zm00001eb216820		chr5	M5_QTLs
GRMZM2G102692	Zm00001eb216830	ATP-dependent DNA helicase HFM1/MER3	chr5	M5_QTLs
AC211276.4_FG008	Zm00001eb218590	glycosyltransferase BC10	chr5	M5_QTLs
GRMZM2G166109	Zm00001eb218600	CSL zinc finger domain-containing protein	chr5	M5_QTLs
GRMZM2G097015	Zm00001eb247090		chr5	M5_QTLs
NA	Zm00001eb247100		chr5	M5_QTLs
GRMZM2G422510	Zm00001eb247110	Filament-like plant protein 7	chr5	M5_QTLs
GRMZM2G422499	Zm00001eb247120	SPRY domain-containing protein	chr5	M5_QTLs
GRMZM2G409372	Zm00001eb248760	RING-H2 finger protein ATL79	chr5	M5_QTLs
GRMZM2G110541	Zm00001eb248800	B box-type domain-containing protein	chr5	M5_QTLs
GRMZM2G169458	Zm00001eb248870	Aldehyde dehydrogenase	chr5	M5_QTLs
GRMZM2G169558	Zm00001eb248880	salt tolerance receptor-like cytoplasmic kinase 1	chr5	M5_QTLs
NA	Zm00001eb248890		chr5	M5_QTLs
GRMZM2G169580	Zm00001eb248900	Putative homeodomain-like transcription factor superfamily protein isoform 1	chr5	M5_QTLs

NA	Zm00001eb251220		chr5	M5_QTLs
GRMZM2G004548	Zm00001eb251230	PLATZ transcription factor	chr5	M5_QTLs
GRMZM2G060837	Zm00001eb251370	Xyloglucan endotransglucosylase/hydrolase	chr5	M5_QTLs
AC233960.1_FG005	Zm00001eb251650	transcription factor bHLH91-like	chr5	M5_QTLs
GRMZM5G861093	Zm00001eb251660		chr5	M5_QTLs
AC233960.1_FG003	Zm00001eb251670	Myb family transcription factor PHL11	chr5	M5_QTLs
AC233960.1_FG002	Zm00001eb251680	Ubiquinol oxidase	chr5	M5_QTLs
GRMZM2G108996	Zm00001eb251710	Brevis radix	chr5	M5_QTLs
GRMZM2G466298	Zm00001eb252540	Protein kinase domain-containing protein	chr5	M5_QTLs
GRMZM2G466309	Zm00001eb252550	Cytochrome b561/ferric reductase transmembrane	chr5	M5_QTLs
NA	Zm00001eb252560		chr5	M5_QTLs
GRMZM2G075562	Zm00001eb253150	Zinc finger protein CONSTANS-LIKE 9	chr5	M5_QTLs
GRMZM2G075372	Zm00001eb253160	TPT domain-containing protein	chr5	M5_QTLs
GRMZM5G837841	Zm00001eb265420	Exonuclease V	chr6	M5_QTLs
GRMZM5G826979	Zm00001eb265430	IQ-domain 3	chr6	M5_QTLs
NA	Zm00001eb265460		chr6	M5_QTLs
GRMZM2G465617	Zm00001eb265470		chr6	M5_QTLs
GRMZM2G165805	Zm00001eb265490	LOB domain-containing protein	chr6	M5_QTLs
GRMZM2G165815	Zm00001eb265500	Bis(5'-adenosyl)-triphosphatase	chr6	M5_QTLs
NA	Zm00001eb265510	FeS cluster assembly protein sufB	chr6	M5_QTLs
GRMZM2G353967	Zm00001eb265520	Cycloeucalenol cycloisomerase	chr6	M5_QTLs
NA	Zm00001eb265530		chr6	M5_QTLs
NA	Zm00001eb265540	interactor of constitutive active ROPs 2	chr6	M5_QTLs
NA	Zm00001eb265550		chr6	M5_QTLs
NA	Zm00001eb265570	PAN domain-containing protein	chr6	M5_QTLs
NA	Zm00001eb265580	E3 ubiquitin-protein ligase UPL4	chr6	M5_QTLs
NA	Zm00001eb265590		chr6	M5_QTLs
GRMZM2G139073	Zm00001eb272490	MADS-box transcription factor 16	chr6	M5_QTLs
GRMZM2G076630	Zm00001eb300910	Putative transcriptional regulator SLK2	chr7	M5_QTLs
NA	Zm00001eb306080		chr7	M5_QTLs
GRMZM2G167580	Zm00001eb311160	Chaperone protein dnaJ 20	chr7	M5_QTLs
GRMZM2G425373	Zm00001eb311170	J domain-containing protein	chr7	M5_QTLs
GRMZM2G141679	Zm00001eb311180	AP domain DRE binding factor	chr7	M5_QTLs
GRMZM2G136663	Zm00001eb311190		chr7	M5_QTLs
GRMZM2G436649	Zm00001eb311200		chr7	M5_QTLs
GRMZM2G174396	Zm00001eb311230	Trehalose 6-phosphate phosphatase	chr7	M5_QTLs
AC212754.3_FG004	Zm00001eb311240	troponin T	chr7	M5_QTLs
GRMZM2G126413	Zm00001eb311250	VQ domain-containing protein	chr7	M5_QTLs
GRMZM2G176780	Zm00001eb311260	Carbohydrate transporter/ sugar porter/ transporter	chr7	M5_QTLs
GRMZM2G176759	Zm00001eb311270	MFS domain-containing protein	chr7	M5_QTLs
GRMZM2G176721	Zm00001eb311280	MFS domain-containing protein	chr7	M5_QTLs
GRMZM2G134496	Zm00001eb311300	MFS domain-containing protein	chr7	M5_QTLs
GRMZM2G434388	Zm00001eb311310	Carbohydrate transporter	chr7	M5_QTLs
GRMZM2G053514	Zm00001eb311320	OCRE domain-containing protein	chr7	M5_QTLs
GRMZM2G155806	Zm00001eb311340	Pumilio homolog 12	chr7	M5_QTLs
GRMZM2G307470	Zm00001eb311360	dTMP kinase	chr7	M5_QTLs
GRMZM2G108546	Zm00001eb311380	MND1-interacting protein 1	chr7	M5_QTLs
GRMZM2G016551	Zm00001eb311390	LAG1 longevity assurance homolog 3	chr7	M5_QTLs
GRMZM2G016660	Zm00001eb311400	Putative transferase	chr7	M5_QTLs
NA	Zm00001eb311410		chr7	M5_QTLs
GRMZM2G049364	Zm00001eb311420	F-box domain-containing protein	chr7	M5_QTLs
NA	Zm00001eb311430	Strictosidine synthase 3	chr7	M5_QTLs
GRMZM2G357371	Zm00001eb311440	Strictosidine synthase 1	chr7	M5_QTLs

GRMZM2G461279	Zm00001eb311440	Strictosidine synthase 1	chr7	M5_QTLs
GRMZM2G009030	Zm00001eb311450	Phosphopyruvate hydratase	chr7	M5_QTLs
GRMZM2G308503	Zm00001eb314350	Oxidoreductase/ transition metal ion binding protein	chr7	M5_QTLs
GRMZM2G037627	Zm00001eb315620	RING-type E3 ubiquitin transferase	chr7	M5_QTLs
GRMZM2G054155	Zm00001eb323740	Pentatricopeptide repeat-containing protein	chr7	M5_QTLs
GRMZM2G023346	Zm00001eb325790	BTB/POZ domain-containing protein At5g48130	chr7	M5_QTLs
GRMZM2G000739	Zm00001eb358930	Urophorphyrin methylase 1	chr8	M5_QTLs
GRMZM2G177150	Zm00001eb358940	PlsC domain-containing protein	chr8	M5_QTLs
GRMZM2G177076	Zm00001eb358960	X8 domain-containing protein	chr8	M5_QTLs
GRMZM2G017386	Zm00001eb358970	Receptor-like kinase lip2	chr8	M5_QTLs
GRMZM2G408537	Zm00001eb358990		chr8	M5_QTLs
GRMZM2G109144	Zm00001eb359000		chr8	M5_QTLs
GRMZM2G109152	Zm00001eb359020		chr8	M5_QTLs
GRMZM2G013726	Zm00001eb359030	Dihydroflavonol-4-reductase	chr8	M5_QTLs
GRMZM2G070913	Zm00001eb359060	Pectinesterase	chr8	M5_QTLs
GRMZM2G070837	Zm00001eb359070	Putative fatty-acid--CoA ligase fadD25	chr8	M5_QTLs
GRMZM2G179696	Zm00001eb359080	Polygalacturonase	chr8	M5_QTLs
GRMZM2G179693	Zm00001eb359100	Peptide deformylase	chr8	M5_QTLs
GRMZM2G179685	Zm00001eb359110	cinnamoyl-CoA reductase-like SNL6	chr8	M5_QTLs
GRMZM2G034025	Zm00001eb359130	Sec39 domain-containing protein	chr8	M5_QTLs
GRMZM2G302195	Zm00001eb359140	Rubisco LS methyltransferase	chr8	M5_QTLs
GRMZM2G004330	Zm00001eb359150	Proline-rich receptor-like protein kinase PERK13	chr8	M5_QTLs
GRMZM2G176182	Zm00001eb359160		chr8	M5_QTLs
GRMZM2G094165	Zm00001eb359190	Carbonic anhydrase	chr8	M5_QTLs
GRMZM2G094356	Zm00001eb359200		chr8	M5_QTLs
GRMZM2G063729	Zm00001eb359210	Ribosomal RNA large subunit methyltransferase N	chr8	M5_QTLs
GRMZM2G063684	Zm00001eb359220	serine/threonine-protein kinase STY13	chr8	M5_QTLs
GRMZM2G063672	Zm00001eb359230	AAA domain-containing protein	chr8	M5_QTLs
GRMZM2G063431	Zm00001eb359240		chr8	M5_QTLs
GRMZM2G043117	Zm00001eb359250	Plant-specific domain TIGR01570 family protein	chr8	M5_QTLs
GRMZM2G375975	Zm00001eb359260	Mitogen-activated protein kinase	chr8	M5_QTLs
GRMZM2G077553	Zm00001eb359270		chr8	M5_QTLs
GRMZM2G154508	Zm00001eb359280		chr8	M5_QTLs
GRMZM5G894582	Zm00001eb359290	4Fe-4S ferredoxin-type domain-containing protein	chr8	M5_QTLs
GRMZM5G830949	Zm00001eb359300	Zinc finger CCCH domain-containing protein 9	chr8	M5_QTLs
GRMZM2G050003	Zm00001eb359310	ADP	chr8	M5_QTLs
GRMZM2G171279	Zm00001eb359320	Potassium channel AKT1	chr8	M5_QTLs
GRMZM2G157334	Zm00001eb359330	GTP-binding nuclear protein	chr8	M5_QTLs
GRMZM2G004947	Zm00001eb359340	GDSL esterase/lipase	chr8	M5_QTLs
GRMZM2G020008	Zm00001eb359350	Pentatricopeptide repeat-containing protein	chr8	M5_QTLs
GRMZM2G321290	Zm00001eb359360	Phospholipase A1	chr8	M5_QTLs
NA	Zm00001eb359370	Proteasome assembly chaperone 4	chr8	M5_QTLs
GRMZM2G000177	Zm00001eb359400	Tset complex member tstf	chr8	M5_QTLs
GRMZM2G132547	Zm00001eb359410	CTP synthase	chr8	M5_QTLs
GRMZM2G432128	Zm00001eb359420	Isocitrate dehydrogenase [NADP]	chr8	M5_QTLs
GRMZM2G179814	Zm00001eb359430	Histone-lysine N-methyltransferase	chr8	M5_QTLs
GRMZM2G104047	Zm00001eb359440	Eukaryotic translation initiation factor 3 subunit 10	chr8	M5_QTLs
GRMZM2G089713	Zm00001eb374090	Sucrose synthase	chr9	M5_QTLs
GRMZM2G090980	Zm00001eb374100	Putative cinnamyl alcohol dehydrogenase 1	chr9	M5_QTLs
GRMZM2G148930	Zm00001eb396640	Senescence associated gene 20	chr9	M5_QTLs
GRMZM2G047370	Zm00001eb425230	Trihelix transcription factor GT-1	chr10	M5_QTLs
GRMZM2G047143	Zm00001eb425240	Mitotic spindle checkpoint protein MAD2	chr10	M5_QTLs
GRMZM2G306348	Zm00001eb425250	ATP-dependent DNA helicase Q-like 5	chr10	M5_QTLs

GRMZM2G702229	Zm00001eb425270	Acid phosphatase/vanadium-dependent haloperoxidase-related protein	chr10	M5_QTLs
GRMZM2G167295	Zm00001eb425290		chr10	M5_QTLs
GRMZM2G167280	Zm00001eb425300	Putative inactive receptor kinase	chr10	M5_QTLs
GRMZM2G437481	Zm00001eb425310	RING-type E3 ubiquitin transferase	chr10	M5_QTLs
GRMZM5G869501	Zm00001eb425380	HGWP repeat containing protein-like	chr10	M5_QTLs
GRMZM5G897749	Zm00001eb425390		chr10	M5_QTLs
NA	Zm00001eb425410		chr10	M5_QTLs
GRMZM2G062052	Zm00001eb425430	Protein TWIN SISTER of FT (Fragment)	chr10	M5_QTLs
GRMZM2G470365	Zm00001eb425440	Sister chromatid cohesion 1 protein 2	chr10	M5_QTLs
GRMZM2G179097	Zm00001eb425450	Protein kinase PITSLRE	chr10	M5_QTLs
GRMZM2G175313	Zm00001eb425460	2-oxoglutarate-dependent dioxygenase DIN11	chr10	M5_QTLs
GRMZM2G473607	Zm00001eb425480	Putative RING zinc finger domain superfamily protein	chr10	M5_QTLs
GRMZM2G018689	Zm00001eb425520	Ribosome-recycling factor	chr10	M5_QTLs
AC219061.2_FG007	Zm00001eb425530	Chaperone protein ClpB 2	chr10	M5_QTLs
GRMZM2G018441	Zm00001eb425540	Phosphoglycolate phosphatase	chr10	M5_QTLs
GRMZM2G018353	Zm00001eb425550	Nitrate transporter 1.2	chr10	M5_QTLs
NA	Zm00001eb425560	Inositol transporter 1	chr10	M5_QTLs
NA	Zm00001eb425570	Putative E3 ubiquitin-protein ligase ARI8	chr10	M5_QTLs
GRMZM2G006216	Zm00001eb425580	histidine protein methyltransferase 1 homolog	chr10	M5_QTLs
GRMZM2G005939	Zm00001eb425590	Transcription factor bHLH153	chr10	M5_QTLs
GRMZM2G005633	Zm00001eb425600	Chitinase	chr10	M5_QTLs
NA	Zm00001eb425610		chr10	M5_QTLs
GRMZM2G105004	Zm00001eb425660	NLP transcription factor (Fragment)	chr10	M5_QTLs
NA	Zm00001eb425670		chr10	M5_QTLs
GRMZM2G070279	Zm00001eb425700	( )-neomenthol dehydrogenase	chr10	M5_QTLs
GRMZM2G070255	Zm00001eb425710	Adenosylhomocysteinase	chr10	M5_QTLs
GRMZM2G070111	Zm00001eb425720	Transcription activator factor	chr10	M5_QTLs
GRMZM2G066162	Zm00001eb425730	Endoglucanase	chr10	M5_QTLs
GRMZM2G066059	Zm00001eb425740	Autophagy 10 variant 1	chr10	M5_QTLs
GRMZM2G096705	Zm00001eb425750	La protein 1	chr10	M5_QTLs
GRMZM2G137920	Zm00001eb425770		chr10	M5_QTLs
AC233888.1_FG002	Zm00001eb425790	COLD	chr10	M5_QTLs
AC233888.1_FG001	Zm00001eb425800	VAN3-binding protein-like	chr10	M5_QTLs
GRMZM2G125452	Zm00001eb425810	ATP synthase subunit e	chr10	M5_QTLs
GRMZM2G125635	Zm00001eb425830	S-adenosylmethionine decarboxylase proenzyme	chr10	M5_QTLs
NA	Zm00001eb425840	Peroxidase	chr10	M5_QTLs
GRMZM2G149994	Zm00001eb425850	Cell division control protein 48 homolog B	chr10	M5_QTLs
GRMZM2G082257	Zm00001eb041480	Putative endoplasmic reticulum membrane C16E8.02	chr1	M2_QTLs
GRMZM2G069656	Zm00001eb041520	Putative glycosyl hydrolase	chr1	M2_QTLs
GRMZM2G079123	Zm00001eb041540	Myb domain protein 106	chr1	M2_QTLs
GRMZM2G079298	Zm00001eb083440	Elicitor-inducible cytochrome P450	chr2	M2_QTLs
GRMZM2G079323	Zm00001eb083450	WD repeat-containing protein 44-like	chr2	M2_QTLs
NA	Zm00001eb083460		chr2	M2_QTLs
GRMZM2G369742	Zm00001eb083470	VQ domain-containing protein	chr2	M2_QTLs
GRMZM5G820287	Zm00001eb083480	Acyl-coenzyme A oxidase	chr2	M2_QTLs
GRMZM2G098128	Zm00001eb083490	RING-type E3 ubiquitin transferase	chr2	M2_QTLs
GRMZM2G313451	Zm00001eb083500		chr2	M2_QTLs
AC233887.1_FG005	Zm00001eb083510	Aluminum-activated malate transporter 1	chr2	M2_QTLs
AC233887.1_FG006	Zm00001eb083520	Digalactosyldiacylglycerol synthase	chr2	M2_QTLs
NA	Zm00001eb083530		chr2	M2_QTLs
GRMZM5G800014	Zm00001eb083540	VAMP protein SEC22	chr2	M2_QTLs
GRMZM2G128322	Zm00001eb083570	Clp R domain-containing protein	chr2	M2_QTLs
NA	Zm00001eb083580		chr2	M2_QTLs

AC214168.3_FG001	Zm00001eb083590	Transcription factor E2FB	chr2	M2_QTLs
GRMZM2G136306	Zm00001eb083600	DUF1279 domain-containing protein (Fragment)	chr2	M2_QTLs
GRMZM2G136364	Zm00001eb083610	AAI domain-containing protein	chr2	M2_QTLs
GRMZM2G055279	Zm00001eb083630	Copper transporter	chr2	M2_QTLs
GRMZM2G055257	Zm00001eb083640	Transcription repressor	chr2	M2_QTLs
GRMZM2G048793	Zm00001eb083650		chr2	M2_QTLs
NA	Zm00001eb083660		chr2	M2_QTLs
GRMZM2G369792	Zm00001eb083670	Pentatricopeptide repeat-containing protein	chr2	M2_QTLs
GRMZM2G001653	Zm00001eb083680	Photosystem I subunit O	chr2	M2_QTLs
NA	Zm00001eb083690		chr2	M2_QTLs
GRMZM2G052268	Zm00001eb086820	F-box domain containing protein	chr2	M2_QTLs
GRMZM2G013357	Zm00001eb086830	Putative long-chain-alcohol O-fatty-acyltransferase 4	chr2	M2_QTLs
GRMZM2G013357	Zm00001eb086840	Isoflavone 2'-hydroxylase	chr2	M2_QTLs
GRMZM2G583556	Zm00001eb086850		chr2	M2_QTLs
NA	Zm00001eb086860	S-acyltransferase	chr2	M2_QTLs
GRMZM2G100747	Zm00001eb088730	thaumatin-like protein 1b	chr2	M2_QTLs
NA	Zm00001eb088740	Pullulanase-type starch debranching enzyme1	chr2	M2_QTLs
GRMZM2G167658	Zm00001eb088790	Multidrug/pheromone exporter	chr2	M2_QTLs
GRMZM2G167613	Zm00001eb088800	Putative cinnamyl alcohol dehydrogenase 6	chr2	M2_QTLs
GRMZM2G010363	Zm00001eb094480	phosphatidylinositol/phosphatidylcholine transfer protein SFH1-like isoform X1	chr2	M2_QTLs
NA	Zm00001eb094490		chr2	M2_QTLs
GRMZM2G021299	Zm00001eb094520	HECT-type E3 ubiquitin transferase	chr2	M2_QTLs
AC217842.3_FG001	Zm00001eb094540	serine/threonine-protein kinase D6PKL2	chr2	M2_QTLs
GRMZM2G013936	Zm00001eb094550	Protein ENHANCED DOWNY MILDEW 2	chr2	M2_QTLs
GRMZM2G102242	Zm00001eb094560	Meiotic nuclear division protein 1 homolog	chr2	M2_QTLs
GRMZM2G127729	Zm00001eb105900	Serine/arginine-rich splicing factor SR45a	chr2	M2_QTLs
GRMZM2G127714	Zm00001eb105910	Sialidase domain-containing protein	chr2	M2_QTLs
GRMZM2G311187	Zm00001eb105950	PPM-type phosphatase domain-containing protein	chr2	M2_QTLs
GRMZM2G001820	Zm00001eb105960	Sialidase domain-containing protein	chr2	M2_QTLs
GRMZM2G479045	Zm00001eb105970		chr2	M2_QTLs
NA	Zm00001eb105990		chr2	M2_QTLs
GRMZM2G470075	Zm00001eb106010	Protein DETOXIFICATION	chr2	M2_QTLs
GRMZM2G169539	Zm00001eb106020	Protein DETOXIFICATION	chr2	M2_QTLs
GRMZM2G169615	Zm00001eb106040	Polyadenylate-binding protein RBP45C	chr2	M2_QTLs
GRMZM2G125935	Zm00001eb106050	DEAD-box ATP-dependent RNA helicase 32	chr2	M2_QTLs
GRMZM2G096944	Zm00001eb106060	Condensin-2 complex subunit H2	chr2	M2_QTLs
GRMZM2G026576	Zm00001eb106070	F-box domain-containing protein	chr2	M2_QTLs
GRMZM5G895534	Zm00001eb106080	Protein FATTY ACID EXPORT 6	chr2	M2_QTLs
GRMZM2G026532	Zm00001eb106090	Taxane 13-alpha-hydroxylase	chr2	M2_QTLs
GRMZM2G026309	Zm00001eb106100	Microtubule-associated protein TORTIFOLIA1	chr2	M2_QTLs
GRMZM2G426556	Zm00001eb106120	Patatin	chr2	M2_QTLs
GRMZM2G125495	Zm00001eb106130	Glutamate receptor	chr2	M2_QTLs
GRMZM2G125507	Zm00001eb106140	septum-promoting GTP-binding protein 1 isoform X1	chr2	M2_QTLs
GRMZM2G125527	Zm00001eb106150	60S ribosomal protein L44	chr2	M2_QTLs
GRMZM2G125850	Zm00001eb106160	Glucose-6-phosphate/phosphate translocator 2	chr2	M2_QTLs
NA	Zm00001eb106180	Thioredoxin domain-containing protein	chr2	M2_QTLs
GRMZM2G309048	Zm00001eb106190	Harpin-induced protein	chr2	M2_QTLs
NA	Zm00001eb106200		chr2	M2_QTLs
GRMZM2G010924	Zm00001eb106210	Harpin-induced protein	chr2	M2_QTLs
GRMZM2G010941	Zm00001eb106230	Rab-GAP TBC domain-containing protein	chr2	M2_QTLs
GRMZM2G131286	Zm00001eb106240	probable pyruvate	chr2	M2_QTLs
GRMZM2G046933	Zm00001eb106250		chr2	M2_QTLs
GRMZM2G175952	Zm00001eb106290	Harpin-induced protein	chr2	M2_QTLs

GRMZM2G475265	Zm00001eb106300	Aspartic proteinase nepenthesin-1	chr2	M2_QTLs
GRMZM2G175927	Zm00001eb106310	Aspartic proteinase nepenthesin-1	chr2	M2_QTLs
GRMZM2G175874	Zm00001eb106320	Putative receptor-like protein kinase 4	chr2	M2_QTLs
GRMZM2G365282	Zm00001eb106340		chr2	M2_QTLs
GRMZM2G101646	Zm00001eb123950		chr3	M2_QTLs
GRMZM2G091119	Zm00001eb127630	Importin subunit alpha	chr3	M2_QTLs
GRMZM2G091232	Zm00001eb127640		chr3	M2_QTLs
GRMZM2G122656	Zm00001eb127650	Pollen-specific protein like	chr3	M2_QTLs
GRMZM2G421742	Zm00001eb127660		chr3	M2_QTLs
GRMZM2G122666	Zm00001eb127670	GTP-binding protein SAR1A	chr3	M2_QTLs
GRMZM2G316967	Zm00001eb127680	Topless-related protein 2 isoform X1	chr3	M2_QTLs
NA	Zm00001eb127690		chr3	M2_QTLs
GRMZM2G402977	Zm00001eb127710	PGG domain-containing protein	chr3	M2_QTLs
GRMZM2G120115	Zm00001eb127720	t-SNARE coiled-coil homology domain-containing protein	chr3	M2_QTLs
GRMZM2G462433	Zm00001eb127740		chr3	M2_QTLs
GRMZM2G161658	Zm00001eb127750	AB hydrolase-1 domain-containing protein	chr3	M2_QTLs
GRMZM2G472231	Zm00001eb127760	Dynein light chain	chr3	M2_QTLs
GRMZM2G472248	Zm00001eb127770	Protein induced upon tuberization	chr3	M2_QTLs
GRMZM2G172956	Zm00001eb127780	PUA domain-containing protein	chr3	M2_QTLs
GRMZM2G038606	Zm00001eb127790	U6 snRNA-associated Sm-like protein LSm4	chr3	M2_QTLs
NA	Zm00001eb127810		chr3	M2_QTLs
GRMZM2G069009	Zm00001eb127840	Zinc finger CCCH domain-containing protein 58	chr3	M2_QTLs
GRMZM2G086994	Zm00001eb127850	Zinc finger CCCH domain-containing protein 58	chr3	M2_QTLs
NA	Zm00001eb127860	Serine/threonine-protein kinase PBL27	chr3	M2_QTLs
AC183319.3_FG001	Zm00001eb127870		chr3	M2_QTLs
GRMZM2G145772	Zm00001eb127880	Serine/threonine-protein kinase RUNKEL	chr3	M2_QTLs
GRMZM2G028010	Zm00001eb127890	Phosphoadenosine phosphosulfate (PAPS) reductase family protein	chr3	M2_QTLs
GRMZM2G007685	Zm00001eb130160	Group I pollen allergen	chr3	M2_QTLs
GRMZM2G087741	Zm00001eb130180	Homeobox protein liguleless 3	chr3	M2_QTLs
GRMZM2G092661	Zm00001eb130190		chr3	M2_QTLs
GRMZM2G103771	Zm00001eb130200	Outer envelope pore protein 16-2	chr3	M2_QTLs
NA	Zm00001eb130210	RING-CH-type domain-containing protein	chr3	M2_QTLs
GRMZM2G147791	Zm00001eb130220	60S ribosomal protein L29	chr3	M2_QTLs
GRMZM2G102014	Zm00001eb130250	Aspartic proteinase nepenthesin-1	chr3	M2_QTLs
GRMZM2G051256	Zm00001eb130260	Myb-related protein P	chr3	M2_QTLs
GRMZM2G051153	Zm00001eb130270		chr3	M2_QTLs
GRMZM2G368662	Zm00001eb130280		chr3	M2_QTLs
GRMZM2G070494	Zm00001eb130290	DOG1 domain-containing protein	chr3	M2_QTLs
NA	Zm00001eb130300	Putative mediator of RNA polymerase II transcription subunit 15 isoform X2	chr3	M2_QTLs
GRMZM2G047616	Zm00001eb130310	High-affinity potassium transporter protein (Fragment)	chr3	M2_QTLs
GRMZM2G396773	Zm00001eb130330		chr3	M2_QTLs
GRMZM2G096010	Zm00001eb130340	Lysophospholipid acyltransferase LPEAT1	chr3	M2_QTLs
GRMZM2G396825	Zm00001eb130350		chr3	M2_QTLs
GRMZM2G076946	Zm00001eb138340	Beta-glucosidase	chr3	M2_QTLs
GRMZM2G077015	Zm00001eb138350	Beta-glucosidase	chr3	M2_QTLs
GRMZM2G052713	Zm00001eb138360	AP complex subunit sigma	chr3	M2_QTLs
GRMZM5G821755	Zm00001eb138370	ZF-HD dimerization-type domain-containing protein	chr3	M2_QTLs
GRMZM2G160687	Zm00001eb138380	Floral homeotic protein AGAMOUS	chr3	M2_QTLs
GRMZM5G844894	Zm00001eb138390	Glucuronoxylan 4-O-methyltransferase 2	chr3	M2_QTLs
GRMZM2G081016	Zm00001eb138400		chr3	M2_QTLs
GRMZM2G143354	Zm00001eb138410	Coatomer subunit delta	chr3	M2_QTLs
GRMZM2G427713	Zm00001eb138420	NB-ARC domain containing protein	chr3	M2_QTLs
GRMZM2G163514	Zm00001eb138430	Suppressor of gene silencing 3	chr3	M2_QTLs

NA	Zm00001eb138450	CCR4-NOT transcription complex subunit 11	chr3	M2_QTLs
AC209377.3_FG003	Zm00001eb138460	Tetraacyldisaccharide 4'-kinase	chr3	M2_QTLs
GRMZM2G422175	Zm00001eb138480		chr3	M2_QTLs
NA	Zm00001eb138490		chr3	M2_QTLs
GRMZM2G338160	Zm00001eb138510		chr3	M2_QTLs
GRMZM2G348257	Zm00001eb138530		chr3	M2_QTLs
GRMZM2G138003	Zm00001eb138550	F13j11 PRLI-interacting factor G	chr3	M2_QTLs
GRMZM2G358076	Zm00001eb138570		chr3	M2_QTLs
GRMZM2G358050	Zm00001eb138590	Thioredoxin M1	chr3	M2_QTLs
NA	Zm00001eb138600	Thioredoxin	chr3	M2_QTLs
GRMZM2G400390	Zm00001eb147850	Laccase	chr3	M2_QTLs
GRMZM2G400390	Zm00001eb147860	Laccase	chr3	M2_QTLs
GRMZM2G089525	Zm00001eb159390	Heat stress transcription factor C-1a	chr3	M2_QTLs
GRMZM2G042421	Zm00001eb159430	WHy domain-containing protein	chr3	M2_QTLs
GRMZM2G180732	Zm00001eb159440	Neutral ceramidase	chr3	M2_QTLs
GRMZM2G480687	Zm00001eb159450		chr3	M2_QTLs
GRMZM2G458095	Zm00001eb161500	AAA domain-containing protein	chr3	M2_QTLs
GRMZM2G157574	Zm00001eb161510	DNL-type domain-containing protein	chr3	M2_QTLs
GRMZM2G157588	Zm00001eb161520	Iron-sulfur cluster co-chaperone protein HscB	chr3	M2_QTLs
GRMZM2G157605	Zm00001eb161530	UBC core domain-containing protein	chr3	M2_QTLs
GRMZM2G458159	Zm00001eb161540	Protein GAMETE EXPRESSED 3	chr3	M2_QTLs
GRMZM2G157616	Zm00001eb161550	Kinesin-like protein KIN-8A	chr3	M2_QTLs
GRMZM2G318530	Zm00001eb161620	Leucine-rich repeat extensin-like protein 1	chr3	M2_QTLs
GRMZM5G812270	Zm00001eb161640	Thioredoxin domain-containing protein	chr3	M2_QTLs
NA	Zm00001eb161710	Putative inactive receptor kinase	chr3	M2_QTLs
NA	Zm00001eb167540		chr4	M2_QTLs
GRMZM2G459532	Zm00001eb169670	Pentatricopeptide repeat-containing protein	chr4	M2_QTLs
GRMZM2G160013	Zm00001eb169680	ELMO domain-containing protein	chr4	M2_QTLs
NA	Zm00001eb169690		chr4	M2_QTLs
GRMZM2G340578	Zm00001eb218250	Cation/calcium exchanger 1	chr5	M2_QTLs
NA	Zm00001eb218260		chr5	M2_QTLs
GRMZM2G038801	Zm00001eb218270	DnaJ/Hsp40 cysteine-rich domain superfamily protein isoform 1	chr5	M2_QTLs
GRMZM2G112366	Zm00001eb218300	Putative signal peptidase complex subunit 1	chr5	M2_QTLs
AC191676.3_FG011	Zm00001eb218310	SHSP domain-containing protein	chr5	M2_QTLs
GRMZM2G112149	Zm00001eb218320	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase	chr5	M2_QTLs
GRMZM5G816609	Zm00001eb232220	26S proteasome non-ATPase regulatory subunit 2 homolog	chr5	M2_QTLs
GRMZM2G063942	Zm00001eb232230	DNA-binding protein BIN4	chr5	M2_QTLs
GRMZM2G033138	Zm00001eb232240	Pathogenesis-related homeodomain protein	chr5	M2_QTLs
GRMZM2G062885	Zm00001eb232250	CCT domain-containing protein	chr5	M2_QTLs
GRMZM2G062914	Zm00001eb232260	Mitogen-activated protein kinase	chr5	M2_QTLs
GRMZM5G852229	Zm00001eb232280		chr5	M2_QTLs
GRMZM2G024612	Zm00001eb232290		chr5	M2_QTLs
GRMZM2G031850	Zm00001eb232310	FMN-binding split barrel	chr5	M2_QTLs
NA	Zm00001eb232320		chr5	M2_QTLs
GRMZM2G367431	Zm00001eb232330	Cell number regulator like	chr5	M2_QTLs
GRMZM2G067313	Zm00001eb232340	Exocyst subunit Exo70 family protein (Fragment)	chr5	M2_QTLs
GRMZM5G829894	Zm00001eb232360	Protein-serine/threonine phosphatase	chr5	M2_QTLs
AC233959.1_FG003	Zm00001eb232380		chr5	M2_QTLs
AC233959.1_FG006	Zm00001eb232390	Homeobox protein ARX-like	chr5	M2_QTLs
GRMZM2G392320	Zm00001eb232400	E3 ubiquitin-protein ligase Arkadia	chr5	M2_QTLs
GRMZM2G147800	Zm00001eb232410	Adagio-like protein 1	chr5	M2_QTLs
GRMZM2G093490	Zm00001eb232420	Protein kinase domain-containing protein	chr5	M2_QTLs
GRMZM2G482256	Zm00001eb232430	Endoglucanase	chr5	M2_QTLs

NA	Zm00001eb232440		chr5	M2_QTLs
GRMZM2G425774	Zm00001eb232450	Fibronectin type-III domain-containing protein	chr5	M2_QTLs
NA	Zm00001eb233590	Gamma-aminobutyrate transaminase POP2 mitochondrial	chr5	M2_QTLs
NA	Zm00001eb233600	Gamma-aminobutyrate transaminase POP2 mitochondrial	chr5	M2_QTLs
AC195957.3_FG003	Zm00001eb233630		chr5	M2_QTLs
GRMZM2G136599	Zm00001eb233640	rRNA adenine N(6)-methyltransferase	chr5	M2_QTLs
GRMZM2G109818	Zm00001eb233650	SUN domain-containing protein	chr5	M2_QTLs
GRMZM2G111136	Zm00001eb233670	Squamosa promoter-binding-like protein 5	chr5	M2_QTLs
GRMZM2G051460	Zm00001eb233680		chr5	M2_QTLs
GRMZM2G417525	Zm00001eb233690		chr5	M2_QTLs
GRMZM2G075333	Zm00001eb233720	4-coumarate--CoA ligase 1	chr5	M2_QTLs
NA	Zm00001eb233730	Sm protein F	chr5	M2_QTLs
GRMZM2G134528	Zm00001eb233740	Fasciclin-like arabinogalactan protein 3	chr5	M2_QTLs
GRMZM2G150864	Zm00001eb233760		chr5	M2_QTLs
GRMZM2G420012	Zm00001eb233850	RING-H2 finger protein ATL2L	chr5	M2_QTLs
GRMZM2G083526	Zm00001eb233880	Aldehyde oxygenase (deformylating)	chr5	M2_QTLs
GRMZM2G469061	Zm00001eb233890		chr5	M2_QTLs
GRMZM2G046155	Zm00001eb233900		chr5	M2_QTLs
GRMZM2G081144	Zm00001eb233910	HD domain-containing protein 2	chr5	M2_QTLs
GRMZM2G472945	Zm00001eb233920	Tub domain-containing protein	chr5	M2_QTLs
NA	Zm00001eb233930	Heptahelical transmembrane protein 2	chr5	M2_QTLs
GRMZM2G119571	Zm00001eb233940	ATG11 domain-containing protein	chr5	M2_QTLs
GRMZM5G832138	Zm00001eb233960	U2 snRNP-associated SURP motif-containing protein-like isoform X1	chr5	M2_QTLs
GRMZM5G832138	Zm00001eb233980	O-fucosyltransferase family protein	chr5	M2_QTLs
GRMZM2G521353	Zm00001eb233990		chr5	M2_QTLs
GRMZM5G863198	Zm00001eb239400	BHLH domain-containing protein	chr5	M2_QTLs
GRMZM5G863198	Zm00001eb239410		chr5	M2_QTLs
GRMZM2G312827	Zm00001eb239420	Calmodulin-binding protein MPCBP	chr5	M2_QTLs
GRMZM2G005301	Zm00001eb239430	Ethylene-responsive transcription factor ERF043	chr5	M2_QTLs
GRMZM2G047918	Zm00001eb239450	Dehydration-responsive element-binding protein 3	chr5	M2_QTLs
GRMZM2G349103	Zm00001eb239470	Ribonuclease P protein subunit p25-like protein	chr5	M2_QTLs
GRMZM2G047818	Zm00001eb239480	Zinc finger CCCH domain-containing protein 18	chr5	M2_QTLs
GRMZM5G891967	Zm00001eb239500		chr5	M2_QTLs
GRMZM2G164950	Zm00001eb239510	Kinesin motor domain-containing protein	chr5	M2_QTLs
GRMZM2G126555	Zm00001eb239520	Monosaccharide-sensing protein 2	chr5	M2_QTLs
GRMZM2G149115	Zm00001eb241290	Glutamine amidotransferase subunit pdxT (Fragment)	chr5	M2_QTLs
GRMZM2G043279	Zm00001eb241550	60S ribosomal protein L35	chr5	M2_QTLs
GRMZM2G043602	Zm00001eb241560	3-oxoacyl-[acyl-carrier-protein] reductase	chr5	M2_QTLs
GRMZM2G043893	Zm00001eb241570	Leucine aminopeptidase 2	chr5	M2_QTLs
GRMZM2G060834	Zm00001eb251390	Putative transcription factor KAN2	chr5	M2_QTLs
GRMZM2G037595	Zm00001eb252340		chr5	M2_QTLs
GRMZM2G469409	Zm00001eb313980	Choline/ethanolamine kinase	chr7	M2_QTLs
GRMZM2G168020	Zm00001eb314000		chr7	M2_QTLs
GRMZM2G162598	Zm00001eb314840	Remorin family protein isoform 1	chr7	M2_QTLs
AC234149.1_FG002	Zm00001eb353210	LOB domain-containing protein	chr8	M2_QTLs
AC234149.1_FG001	Zm00001eb353220	DUF4283 domain-containing protein (Fragment)	chr8	M2_QTLs
GRMZM2G058690	Zm00001eb353550	IQ-domain 25	chr8	M2_QTLs
GRMZM2G058745	Zm00001eb353560	Phosphotransferase	chr8	M2_QTLs
NA	Zm00001eb359090		chr8	M2_QTLs
NA	Zm00001eb359170		chr8	M2_QTLs
GRMZM2G300375	Zm00001eb363860	Pre-mRNA-splicing factor ATP-dependent RNA helicase DEAH10	chr8	M2_QTLs
GRMZM2G134308	Zm00001eb363870	Exostosin domain-containing protein	chr8	M2_QTLs
GRMZM2G134256	Zm00001eb363880	Transaldolase	chr8	M2_QTLs

GRMZM2G134176	Zm00001eb363890	E2 ubiquitin-conjugating enzyme	chr8	M2_QTLs
GRMZM5G858610	Zm00001eb367340	Ubiquitin-conjugating enzyme E2 34	chr8	M2_QTLs
GRMZM2G173641	Zm00001eb377300	1-deoxy-D-xylulose-5-phosphate synthase	chr9	M2_QTLs
GRMZM2G173628	Zm00001eb377310	Superoxide dismutase (Fragment)	chr9	M2_QTLs
GRMZM2G173579	Zm00001eb377330	CRIB domain-containing protein	chr9	M2_QTLs
GRMZM5G888263	Zm00001eb397790	Omp85 domain-containing protein	chr9	M2_QTLs

Table S3 : Description of pQTLs hotspots.

Hotspot ID	Chromosome	Position	# proteins having a PQTL in the hotspot	# genes located In the hotspot	contain QTLs Or pQTLs capturing GxW variance	Effect on GxW var capturated If removed
c242	1	34277343-35496913	37	74	TRUE	-2.568
c243	1	34908510-36077195	27	61	FALSE	NA
c255	1	43450217-44642517	30	65	TRUE	-2.038
c257	1	45120508-46315918	26	69	TRUE	-2.563
c267	1	52232447-53457360	39	63	TRUE	-2.658
c268	1	52844991-54015558	28	57	TRUE	-3.274
c270	1	54229864-55390319	25	63	FALSE	NA
c280	1	61602335-62807213	25	82	TRUE	-2.6
c570	1	187888737-188203332	23	23	TRUE	-2.784
c857	1	253154603-253496349	35	29	TRUE	0.086
c1565	2	59782102-61739623	23	85	FALSE	NA
c1617	2	132461077-134415907	26	68	FALSE	NA
c1623	2	139563560-141377917	25	91	FALSE	NA
c1909	2	202644239-204572278	34	150	TRUE	-2.876
c1911	2	204710508-206545443	23	111	FALSE	NA
c1912	2	205926645-207876099	22	155	TRUE	1.39
c1913	2	206909982-208869763	28	162	FALSE	NA
c1914	2	207944803-209163811	23	96	FALSE	NA
c2378	3	20304510-23583807	40	198	TRUE	-1.844
c2379	3	22029811-25367590	41	192	FALSE	NA
c2380	3	23919689-27289114	32	131	FALSE	NA
c2381	3	26616331-29941329	36	156	FALSE	NA
c2382	3	28328232-31659496	46	159	TRUE	-2.966
c2383	3	30076386-33357326	28	148	FALSE	NA
c2384	3	32215197-35529469	45	148	TRUE	-2.315
c2385	3	33998406-37163422	35	125	TRUE	-2.661
c2386	3	35743966-39013200	33	117	FALSE	NA
c2387	3	38444811-41784384	29	149	FALSE	NA
c2389	3	42018999-45257770	31	131	FALSE	NA
c2391	3	46164777-49341014	36	149	TRUE	-3.486
c2392	3	47868118-50939720	21	133	TRUE	-3.057
c2396	3	55196453-58238374	21	142	TRUE	-2.382
c2397	3	56978667-60172816	35	135	FALSE	NA
c2426	3	121225152-124608710	21	123	FALSE	NA
c2429	3	126654050-129926190	39	171	TRUE	-4.008
c2433	3	133767803-137111237	31	167	TRUE	-3.185
c2434	3	135479073-138634079	26	161	FALSE	NA
c2436	3	139057522-142417510	22	171	FALSE	NA
c2440	3	146434190-149723400	21	179	TRUE	-2.47
c2441	3	148127980-151224275	25	167	FALSE	NA
c3326	4	24934516-28019294	49	164	FALSE	NA
c3329	4	29977653-33059061	25	167	FALSE	NA
c3331	4	33325206-36360362	23	142	FALSE	NA
c3332	4	35221084-38130811	23	168	FALSE	NA
c3357	4	84024893-87209245	27	117	FALSE	NA
c3384	4	141993614-145093144	33	163	FALSE	NA
c3680	4	195153252-198181212	23	195	TRUE	-2.645
c3681	4	196784094-199994526	47	216	TRUE	-4.154
c3682	4	198408543-201634690	29	202	TRUE	-3.095

c3683	4	200033462-203248404	30	218	FALSE	NA
c3689	4	210073676-213202264	29	158	FALSE	NA
c4259	5	69259467-71753742	23	141	FALSE	NA
c4262	5	73384717-75868228	35	118	FALSE	NA
c4263	5	75026189-77531896	24	122	FALSE	NA
c4264	5	76596679-79048876	24	113	TRUE	-4.543
c4272	5	88494422-91030708	27	108	TRUE	-3.452
c4302	5	142940177-145244034	24	108	TRUE	-3.071
c4306	5	148777765-151400300	23	103	FALSE	NA
c4309	5	153287947-155906535	29	145	TRUE	-3.09
c4960	6	33797656-35394334	24	74	FALSE	NA
c5287	6	127714264-129371290	26	108	FALSE	NA
c5292	6	132211709-133776640	25	74	FALSE	NA
c5297	6	136780716-138312984	22	99	FALSE	NA
c5298	6	137641402-139292215	25	109	TRUE	-4.417
c5301	6	140367304-142029302	24	120	FALSE	NA
c5302	6	141378875-143015485	23	128	FALSE	NA
c5308	6	146600151-148223730	42	143	TRUE	-5.597
c5799	7	30117301-32204215	23	93	FALSE	NA
c6058	7	142465640-144574939	33	142	FALSE	NA
c6059	7	143584006-145764511	25	147	FALSE	NA
c6065	7	151805614-153840687	40	135	TRUE	-1.783
c6066	7	153022139-155174561	26	153	FALSE	NA
c6067	7	154172628-156141400	31	156	FALSE	NA
c6895	8	137727122-139633413	29	101	TRUE	-2.943
c6898	8	140980870-142866934	46	112	TRUE	-4.17
c6902	8	145186420-147115613	27	150	TRUE	-2.668
c6904	8	147419852-149402832	22	114	FALSE	NA
c6907	8	150561545-152406417	21	126	TRUE	-3.67
c7377	9	25758656-28947203	40	166	TRUE	-3.731
c7402	9	85764297-89347369	21	179	TRUE	-3.048
c7403	9	87601471-91155296	41	151	TRUE	-7.295
c7407	9	95176072-98290154	21	147	TRUE	-2.902
c7408	9	97593360-101089026	23	207	TRUE	-3.303
c7409	9	99406813-102805659	43	196	TRUE	-5.922
c7410	9	101298814-103861093	28	117	FALSE	NA
c7423	9	105156866-105289572	21	9	FALSE	NA
c7520	9	120680970-124142043	21	196	FALSE	NA
c7521	9	122880520-126388152	29	257	TRUE	-3.905
c7522	9	124720985-128281744	29	233	FALSE	NA
c7523	9	126551237-130085226	36	196	TRUE	-4.011
c8009	10	20576128-23810736	31	111	TRUE	-2.858
c8010	10	22256817-25468294	26	154	FALSE	NA
c8011	10	24136598-27376204	35	146	TRUE	-3.314
c8166	10	112433409-115586575	31	187	TRUE	-7.355
c8168	10	115866116-118965117	32	181	FALSE	NA
c8171	10	120937439-124107507	23	194	TRUE	-5.078
c8172	10	122717175-125921700	45	199	FALSE	NA
c8173	10	124380590-127501534	34	224	TRUE	-2.14
c8174	10	126072213-128627355	33	179	TRUE	-7.136
c8209	10	132521471-132577177	26	14	FALSE	NA

Romain POUPON's internship report

RAPPORT DE STAGE

du 2 Mai au 5 Juillet 2023

Romain POUPON

Master 1

Bio-informatique et biostatistiques

Enseignant référent : Olivier LESPINET

Encadrant de stage : Marie-Laure MARTIN et Yacine DJABALI

Équipe Genomic Networks (GNet)

Institut de Sciences des Plantes Paris-Saclay (IPS2)

Bâtiment 630, rue de Noetzlin

91190 - Gif-sur-Yvette

Remerciements

Je remercie mes encadrants, Yacine Djabali et Marie-Laure Martin, ainsi que toute l'équipe GNet.

Sommaire

1 Introduction

1.1 Présentation de l'IPS2 et de l'équipe GNet

L'institut des sciences des plantes – Paris-Saclay (IPS2) est un laboratoire de recherche dont le but est de mieux comprendre les mécanismes génétiques et moléculaires qui contrôlent la croissance de la plante ainsi que la régulation de ces mécanismes en interaction avec son environnement et en particulier dans le cadre du changement climatique.

J'ai effectué mon stage au sein de l'équipe Gnet qui fait partie du département « Interaction plantes micro-organismes et réseaux ». L'objectif de ce département est d'étudier les plantes, au sein de leur environnement dans le but de mieux comprendre les interactions entre les plantes et l'environnement, et d'appliquer ces connaissances aux conditions réelles.

L'équipe GNet, animée par Marie-Laure Martin, développe des modèles statistiques et des approches de biologie computationnelle pour améliorer les connaissances biologiques et génétiques des mécanismes de réponse aux contraintes environnementales des plantes. Les principales études se font sur la plante modèle *Arabidopsis thaliana* mais l'équipe travaille également sur des espèces d'intérêt agronomiques comme le blé ou le maïs. Dans le cadre de mon stage, j'ai travaillé sur le maïs.

1.2 Contexte

Le maïs est la céréale la plus produite au monde en volume devant le riz et le blé. Il est indispensable dans le système agro-alimentaire mondial où il joue un rôle croissant dans l'alimentation humaine et animale (FAOSTAT, 2023). De plus, il est largement utilisé dans les produits industriels, y compris pour la production de biocarburants. Ainsi, au cours de la prochaine décennie, il est en passe de devenir la culture la plus répandue en termes de superficie (Erenstein et al., 2022). Il fournit au moins 30% des calories alimentaires à plus de 4,5 milliards de personnes dans 94 pays en développement. Il est donc important pour la sécurité alimentaire mondiale, notamment pour son rôle dans le régime alimentaire des populations pauvres en Afrique et en Amérique latine (Ranum et al., 2014 ; Shiferaw et al., 2011).

La culture du maïs demande beaucoup d'eau ce qui le rend très sensible à la sécheresse, en particulier durant la phase de reproduction (Daryanto et al., 2016). Ainsi, les changements de température ainsi que les périodes de sécheresse liés au réchauffement climatique risquent de faire baisser les rendements agricoles et menacent la sécurité alimentaire mondiale (Bezner Kerr et al., 2022). Pour répondre à la future demande en maïs, une solution est de faire de l'amélioration génétique afin de créer des variétés de maïs plus résistantes à la sécheresse (Campos et al., 2004).

Cependant, la conception de variété tolérante à un déficit hydrique est difficile, car la capacité à résister au manque d'eau est une caractéristique complexe qui dépend de l'interaction de plusieurs caractères comme l'utilisation de l'eau, la croissance des feuilles et le taux de transpiration (Tardieu et al., 2017). De plus, ces caractères sont le résultat d'interactions moléculaires au niveau des gènes, des protéines ou encore des métabolites.

1.3 Etat de l'art

Des études de génétiques d'associations (GWAS) on été faites afin d'identifier des gènes impliqués dans la résistance à la sécheresse (Prado et al., 2018). Elles consistent à identifier sur un ensemble d'hybrides de maïs les polymorphismes génétiques pouvant être responsables de la variation inter-hybride observée sur un caractère. Ce type d'étude permet de détecter des régions d'ADN associés a un caractère quantitatif. Ces régions sont appelées des QTLs (quantitative trait loci). Ainsi, sur le jeu de données auquel j'ai pu avoir accès, Prado et al., 2018, a déjà identifié 16 QTLs associé à la conductance stomatique, un caractère qui joue un rôle central dans la résistance au déficit d'eau chez le maïs. Ils ont également montré que ces QTLs se superposent à des QTLs de transpiration et de biomasse deux caractères impactés par un déficit hydrique. De plus, Blein-Nicolas et al., 2020 a identifié plus de 22,000 pQTLs associés à 2,055 abondances de protéines.

La recherche de QTLs réalisé par Prado et al., 2018 a été faite sur les caractères écophysiologiques et celle réalisé par Blein-Nicolas et al., 2020 a été faite sur les protéines. Cependant, comme dit précédemment, la réponse à un déficit hydrique implique des réponses moléculaires complexes. Ainsi, la thèse de Yacine Djabali qui a débuté en 2020 a pour objectif de développer des méthodes de biologie computationnelle pour intégrer dans un modèle de génétique d'association les données moléculaires aux données mesurant les caractères écophysiologiques. Dans une première partie de sa thèse, il a travaillé sur les caractères et il a montré qu'en analysant les 4 années dans un même modèle linéaire mixte, il identifiait 142 QTLs. Il montre que ces QTLs capturent plus de 75 % de la variance de l'interaction génotype x disponibilité de l'eau.

Dans une seconde partie de sa thèse, il a développé une méthode qui a permis d'identifier 28 protéines corrélées aux caractères écophysiologiques et il a montré que l'intégration des 237 pQTLS associées aux 28 protéines permet de mieux capturer la variance génétique et la variance de l'interaction génotype x disponibilité en eau des caractères écophysiologiques.

1.4 Objectifs du stage

Mon travail se situe dans la continuité du travail de la seconde partie de thèse de Yacine Djabali. En effet en plus des données protéomiques, des données métabolomiques sont également disponibles et l'objectif principal du stage a été d'appliquer l'approche développée pour les protéines pour quantifier l'apport d'information génétique contenue dans les données métabolomiques pour la modélisation des caractères écophysiologiques. Pour cela, j'ai d'abord mesuré le gain d'information apporté par l'ensemble des mQTLs. Ensuite, j'ai cherché à affiner l'intégration des métabolites dans le modèle en sélectionnant un sous-ensemble de métabolites pour faciliter l'interprétation des résultats.

2 Matériel & méthodes

2.1 Matériel végétal

Les données proviennent du projet Amaizing, financé par l'Agence Nationale de la Recherche. Elles ont été produites par la plateforme PhenoArch du M3P (Montpellier Plant Phenotyping Platforms) (Cabrera-Bosquet et al., 2016 ; Prado et al., 2018). Les plantes étudiées sont des hybrides provenant d'un croisement entre un parent d'une lignée corné (UH007) et 254 lignées dentées sélectionnées pour leur fenêtre de floraison restreinte . Les 254 lignées de maïs ont été génotypés à l'aide de trois technologies de génotypage (Genotyping-By-Sequencing, Illumina Infinium 50 K et Affymetrix Axiom 600 K arrays) ce qui a permis d'identifier 977,459 SNPs (Negro et al., 2019).

Les hybrides ont été cultivés sous deux conditions d'irrigations : bien irrigué (WW, potentiel hydrique du sol de -0.05 Mpa) et en déficit d'eau (WD, potentiel hydrique du sol moyen de -0.45 MPa) au printemps 2012, en hiver 2013, aux printemps 2013 et 2016. Durant ces 4 expériences, 6 caractères écophysiologiques on été mesuré : biomass (Biom), leaf area (LA), transpiration rate (Transp), stomatal conductance (gs_max), water uptake (WU) et water use efficiency (WUE).

En plus des caractères écophysiologiques, des données protéomiques ont été recueillies sur les individus du printemps 2012 et des analyses LC-MS ont permis de quantifier 2055 protéines (Blein-Nicolas et al., 2020). Au printemps 2013, ce sont des données métabolomiques qui ont été recueillies sur les hybrides et des analyses GC-MS, réalisées par la plateforme Bordeaux Métabolome [1], ont permis de quantifier 1,416 métabolites. Contrairement aux données protéomiques, ces données ne sont pas encore publiées, elles ont été mises à notre disposition dans le cadre d'une collaboration.

Des indices de plasticité (Bouslama & Schapaugh, 1984) ont été calculés pour chaque génotype en faisant le ratio entre la moyenne génotypique d'un caractère donné en condition WD, sur celle en condition WW. Les QTLs détectés sous les conditions d'irrigation WW et WD sont par la suite appelés

1. https://metabolome.cgfb.u-bordeaux.fr/

des QTLs d'état stationnaire et ceux associés indices plasticité sont par la suite appelés des QTLs de plasticité.

2.2 Décomposition de la variance génétique à l'aide d'un modèle linéaire mixte

Des modèles linéaires mixtes multi-environnements et multi-caractères ont été utilisés pour décomposer la variance. Par la suite, je présente les 4 modèles utilisés durant mon stage.

2.2.1 Modèle M1 : sans QTL

Ce premier modèle décompose le signal sans intégrer de QTLs dans la décomposition de la moyenne. Il sera le modèle de référence pour mesurer la part de variance expliquée pour un effet aléatoire donné par l'ensemble des QTL intégré dans les autres modèles où des QTLs sont ajoutés.

$$\begin{aligned} Y_{gwft} &= \mu + T_t + E_{wf} + K_g + (T \times E)_{wft} + (K \times E)_{gwf} + (K \times T)_{gt} \\ &+ \underline{G_g} + \underline{(G \times T)_{gt}} + \underline{(G \times W)_{gw}} + \underline{(G \times F)_{gf}} + \underline{\varepsilon_{gwf}} \end{aligned}$$

où :

— Y_{gwft} est la moyenne du caractère t, corrigée des effets spatiaux de la serre, des 3 réplicas mis en culture pour le génotype g dans la condition d'irrigation w pendant l'expérience f ;
— μ est la moyenne générale ;
— T_t est l'effet fixe lié au caractère t ;
— E_{wf} est l'effet fixe de l'environnement défini par la condition d'irrigation w et l'expérience f ;
— K_g les coordonnées du génotype g projeté sur les axes d'une analyse en composante principale à partir de la matrice d'apparentement. Le nombre d'axes utilisés a été choisi selon le critère de Kaiser ;
— $(T \times E)$ est l'effet fixe lié à l'interaction du caractère t avec l'environnement défini par la condition d'irrigation w et l'expérience f ;
— $(K \times E)_{gwf}$ est l'effet fixe liés à l'interaction de la structure génétique avec l'environnement ;
— $(K \times T)_{gt}$ est l'effet fixe liés à l'interaction de la structure génétique et du caractère ;
— $\underline{G_g}$ l'effet aléatoire de la variance du génotype ;
— $\underline{(G \times T)_{gt}}$ est l'effet aléatoire lié à l'interaction du génotype avec le caractère ;
— $\underline{(G \times W)_{gw}}$ est l'effet aléatoire lié à l'interaction du génotype avec la disponibilité en eau ;
— $\underline{(G \times F)_{gf}}$ est l'effet aléatoire lié à l'interaction du génotype avec l'environnement ;
— $\underline{\varepsilon_{gwf}}$ est le vecteur d'erreurs aléatoires.

Tous les effets aléatoires suivent une loi gaussienne centrée dont la variance sera estimée en même temps que les paramètres de la moyenne par restricted maximum likelihood (REML).

2.2.2 Modèle M2 : intégrant les QTLs de caractères (d'état stationnaire et de plasticité)

Ce deuxième modèle décompose le signal en intégrant les QTLs de caractères dans la décomposition de la moyenne. Il sera le modèle de référence pour mesurer le gain des autres modèles où des mQTLs sont ajoutés.

$$\begin{aligned} Y_{gwft} &= \mu + T_t + E_{wf} + K_g + QS_g + QP_g + (T \times E)_{wft} + (K \times E)_{gwf} \\ &+ (K \times T)_{gt} + (QS \times E)_{gwf} + (QS \times T)_{gt} + (QP \times E)_{gwf} \\ &+ (QP \times T)_{gt} + \underline{G_g} + \underline{(G \times T)_{gt}} + \underline{(G \times W)_{gw}} + \underline{(G \times F)_{gf}} + \underline{\varepsilon_{gwf}} \end{aligned}$$

Où :

— QS_g les coordonnées du génotype g, projeté sur les axes d'une analyse en composante principale à partir de la matrice d'apparentement, calculée avec un ensemble de SNP qui décrivent le set de QTL correspondant (le nombre d'axes utilisés a été choisi selon le critère de Kaiser) ;
— QP_g les coordonnées du génotype g projeté sur les axes d'une analyse en composante principale à partir de la matrice d'apparentement, calculée avec un ensemble de SNP qui décrivent le set de QTL correspondant (le nombre d'axes utilisés a été choisi selon le critère de Kaiser) ;
— $(QS \times E)_{gwf}$ est l'effet fixe lié à l'interaction des QTLs de caractère d'état stationnaire avec l'environnement ;
— $(QS \times T)_{gt}$ est l'effet fixe lié à l'interaction des QTLs de caractère d'état stationnaire avec le caractère ;
— $(QP \times E)_{gwf}$ est l'effet fixe lié à l'interaction des QTLs de caractère de plasticité avec l'environnement ;
— $(QP \times T)_{gt}$ est l'effet fixe lié à l'interaction des QTLs de caractère de plasticité avec le caractère.

2.2.3 Modèle M3 : intégrant les mQTLs d'état stationnaire

Ce troisième modèle décompose le signal en intégrant en plus des QTLs de caractère les mQTLs d'état stationnaire dans la décomposition de la moyenne. Il permettra de mesurer le gain apporté par les mQTLs d'état stationnaire.

$$\begin{aligned} Y_{gwft} &= \mu + T_t + E_{wf} + K_g + QS_g + QP_g + MQS_g + (T \times E)_{wft} + (K \times E)_{gwf} \\ &+ (K \times T)_{gt} + (QS \times E)_{gwf} + (QS \times T)_{gt} + (QP \times E)_{gwf} + (QP \times T)_{gt} \\ &+ (MQS \times E)_{gwf} + (MQS \times T)_{gt} + \underline{G_g} + \underline{(G \times T)_{gt}} + \underline{(G \times W)_{gw}} \\ &+ \underline{(G \times F)_{gf}} + \underline{\varepsilon_{gwf}} \end{aligned}$$

Où :
— MQS_g les coordonnées du génotype g projetés sur les axes de l'analyse en composante principale obtenu grâce une matrice de parenté, calculée avec un ensemble de SNP qui décrivent le set de QTL correspondant (le nombre d'axes utilisés a été choisi selon le critère de Kaiser) ;
— $(MQS \times E)_{gwf}$ est l'effet fixe lié à l'interaction des mQTLs d'état stationnaire avec l'environnement ;
— $(MQS \times T)_{gt}$ est l'effet fixe lié à l'interaction des mQTLs d'état stationnaire avec le caractère.

2.2.4 Modèle M4 : intégrant les mQTLs d'état stationnaire et de plasticité

Ce quatrième modèle décompose le signal en intégrant dans la décomposition de la moyenne, en plus des QTLs de caractère et des mQTLs d'état stationnaire, les mQTLs de plasticité. Il permettra de mesurer le gain apporté par les mQTLs d'état stationnaire et de plasticité.

$$\begin{aligned} Y_{gwft} &= \mu + T_t + E_{wf} + K_g + QS_g + QP_g + MQS_g + MQP_g + (T \times E)_{wft} \\ &+ (K \times E)_{gwf} + (K \times T)_{gt} + (QS \times E)_{gwf} + (QS \times T)_{gt} + (QP \times E)_{gwf} \\ &+ (QP \times T)_{gt} + (MQS \times E)_{gwf} + (MQS \times T)_{gt} + (MQP \times E)_{gwf} \\ &+ (MQP \times T)_{gt} + \underline{G_g} + \underline{(G \times T)_{gt}} + \underline{(G \times W)_{gw}} + \underline{(G \times F)_{gf}} + \underline{\varepsilon_{gwf}} \end{aligned}$$

Où :
— MQP_g les coordonnées du génotype g projetés sur les axes de l'analyse en composante principale obtenu grâce une matrice de parenté, calculée avec un ensemble de SNP qui décrivent le set de QTL correspondant (le nombre d'axes utilisés a été choisi selon le critère de Kaiser) ;
— $(MQP \times E)_{gwf}$ est l'effet fixe lié à l'interaction des mQTLs de plasticité avec l'environnement ;
— $(MQP \times T)_{gt}$ est l'effet fixe lié à l'interaction des mQTLs de plasticité avec le caractère.

2.3 Calcul du gain apporté par un ensemble donné de mQTLs

Posons r un des effets aléatoires. La part de variance expliquée pour cet effet par l'ensemble des QTL intégré dans le modèle est :

$$\gamma_{qr} = \frac{\Gamma_r^* - \Gamma_r}{\Gamma_r^*}$$

Avec Γ_r^* la variance de l'effet aléatoire r dans le modèle M1 et Γ_r la variance de l'effet aléatoire r dans le modèle intégrant des QTLs (M2, M3 ou M4).

2.4 Inférence du réseau incluant les caractères et les métabolites

Les modèles graphiques gaussiens (GGM) sont un cas particulier de modèle graphique non orienté qui représente un vecteur aléatoire multivarié $X = (X_1, ..., X_p)$ sous la forme d'un graphe $G(V, E)$, où V est l'ensemble des nœuds représentant les p composantes de X et E est un ensemble d'arêtes entre ces nœuds, représentant les dépendances conditionnelles entre les variables (Shutta et al., 2022).

A l'aide de cette modélisation, j'ai estimé pour les 2 conditions d'irrigation WW et WD ainsi que pour la réponse, les dépendances conditionnelles entre les 5 caractères et les 1421 métabolites.

L'estimation d'un GGM dépend d'un paramètre λ compris entre 0 et 1 qui contrôle la densité du réseau. Plus le paramètre λ est proche de 0, plus le réseau est dense. Tous les réseaux sont estimés sur une grille de taille 100, constituée de λ rangés en ordre décroissant entre 1 et 0. Cela permet d'ordonner en ordre croissant les dépendances conditionnelles et comme la grille est la même pour les trois réseaux, cela va permettre de les comparer (Shutta et al., 2022).

Comme les données ne suivent pas forcément une loi normale multivariée, j'ai procédé à une transformation qui cherche à trouver un ensemble de fonctions monotones $f_1, ..., f_p$ telles que la distribution de $f(X) = (f_1(X_1), ..., f_p(X_p))$ soit normale(Liu et al., 2012).

2.5 Le logiciel R

Les analyses ont été effectuées sur R (R Core Team 2022) version 4.2.1. (R Core Team, 2023). Les modèles linéaires mixtes ont été ajustés avec le package lme4 (Bates et al., 2015). Les GGM ont été estimés avec le package huge (Jiang et al., 2021) en utilisant la fonction huge.glasso après avoir procédé à la transformation nonparanormale des données avec la fonction huge.npn. Enfin, les réseaux ont été visualisés avec le package igraph (Csardi & Nepusz, 2006) .

Pour que les trois réseaux aient le même vecteur de λ, on le construit en suivant la méthode de la fonction huge.glasso mais en prenant pour lambda.max le maximum entre les lambda.max des 3 jeux de données et pour lambda.min le minimum des lambda.min des trois jeux de données (Jiang et al., 2021).

3 Résultats

J'ai calculé la composante de variance de chaque effets aléatoire et leur contribution en pourcentage à la variance totale, à partir du modèle M1 (Tab. 1). L'effet de la résiduel est le plus important, vient ensuite celui de la variance du génotype qui explique 23% de la variance totale. Les effets $G \times W$, $G \times T$ et $G \times F$ expliquent entre 10% et 14% de variance totale. Toutes les p-values sont très faibles donc tous les effets aléatoires sont significatifs.

G		$G \times W$		$G \times T$		$G \times F$		ε	
Γ_G	%	$\Gamma_{G\times W}$	%	$\Gamma_{G\times T}$	%	$\Gamma_{G\times F}$	%	Γ_ε	%
$4{,}6e^{-2}$ ($8e^{-19}$)	23	$2{,}1e^{-2}$ ($4e^{-84}$)	10	$2{,}7e^{-2}$ ($6e^{-108}$)	13	$2{,}8e^{-2}$ ($3e^{-127}$)	14	$7{,}9e^{-2}$	39

TABLE 1 – Composantes de variance de chaque effets aléatoires et leur contribution en pourcentage à la variance totale, calculés à partir du modèle M1. La p-value entre parenthèse indique la significativité de cet effet.

3.1 L'intégration de tout les mQTLs capture de la variance génétique des caractères

J'ai calculé les parts de variance capturée par chaque effet aléatoire et je les ai comparées entre le modèle M2, le modèle M3 intégrant l'ensemble des 27,629 mQTLs d'état stationnaire et le modèle M4 intégrant l'ensemble des mQTLS d'état stationnaire et de plasticité (Fig. 1).

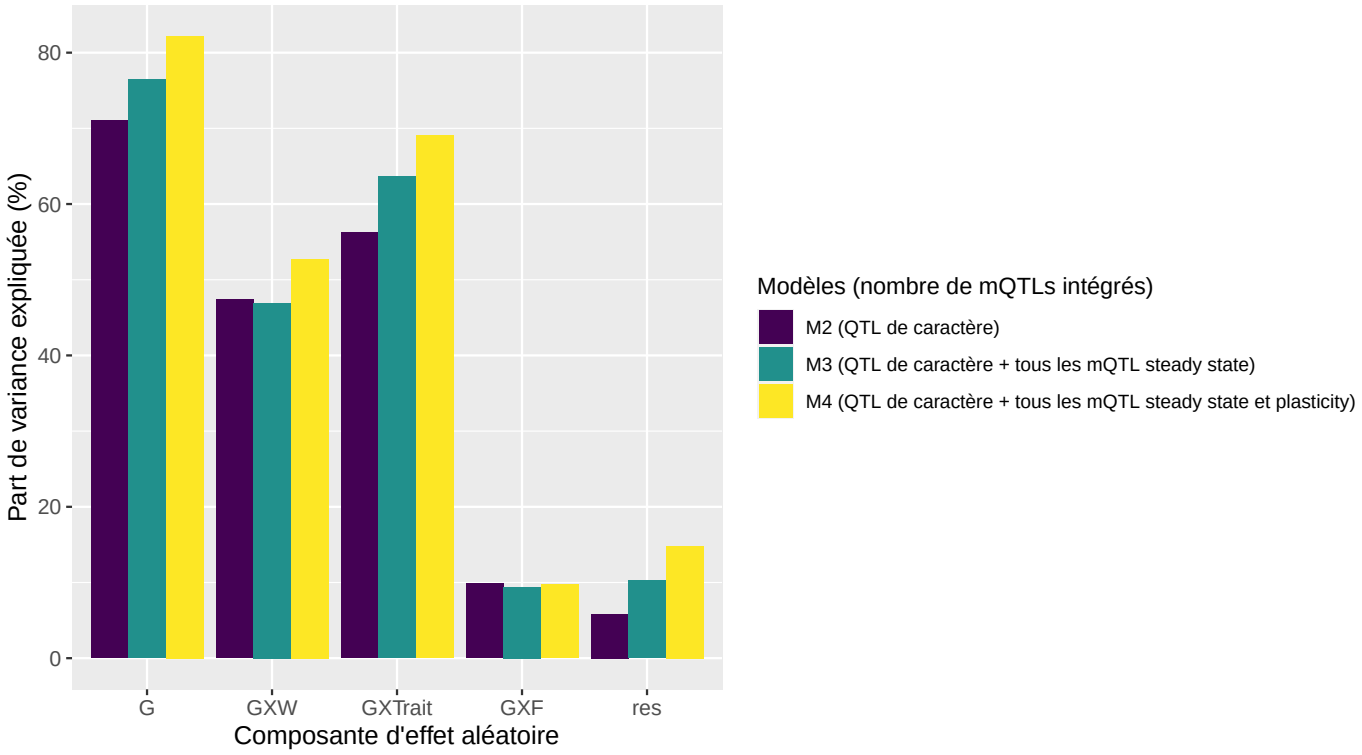


FIGURE 1 – Comparaison des parts de variance expliquées en fonction des composantes d'effets aléatoire entre le modèle M2 n'intégrant que les QTLs de caractère, le modèle M3 intégrant les QTLs de caractère et tous les mQTLs d'état stationnaire, et le modèle M4 intégrant les QTLs de caractère et tous les mQTLs d'état stationnaire et de plasticité.

J'observe une augmentation de la part de variance capturée pour les effets G, $G \times W$ et $G \times T$ chez le modèle M4 et le modèle M2. Pour l'effet $G \times F$, on observe une petite diminution de la part de variance expliquée et une augmentation pour les résidus.

Lorsque j'intègre les mQTLs de plasticité en plus des mQTLs d'état stationnaire, les parts de variance expliquées sont supérieures à celles observées lorsque l'on intègre que les mQTLs d'état stationnaire, et ce, pour chaque effet aléatoire. L'intégration des données métabolomiques permet donc de mieux capturer la variance génétique et l'intégration des mQTLs de plasticité est pertinent car il permet de mieux capturer cette variance génétique.

3.2 Construction du réseau dans les trois environnements

3.2.1 Évolution du voisinage des caractères en fonction du λ

Pour chacun des trois GGM, j'ai regardé le nombre de voisins de degré 1 de chaque caractère écophysiologique en fonction de la valeur de l'hyperparamètre λ (Fig. 2).

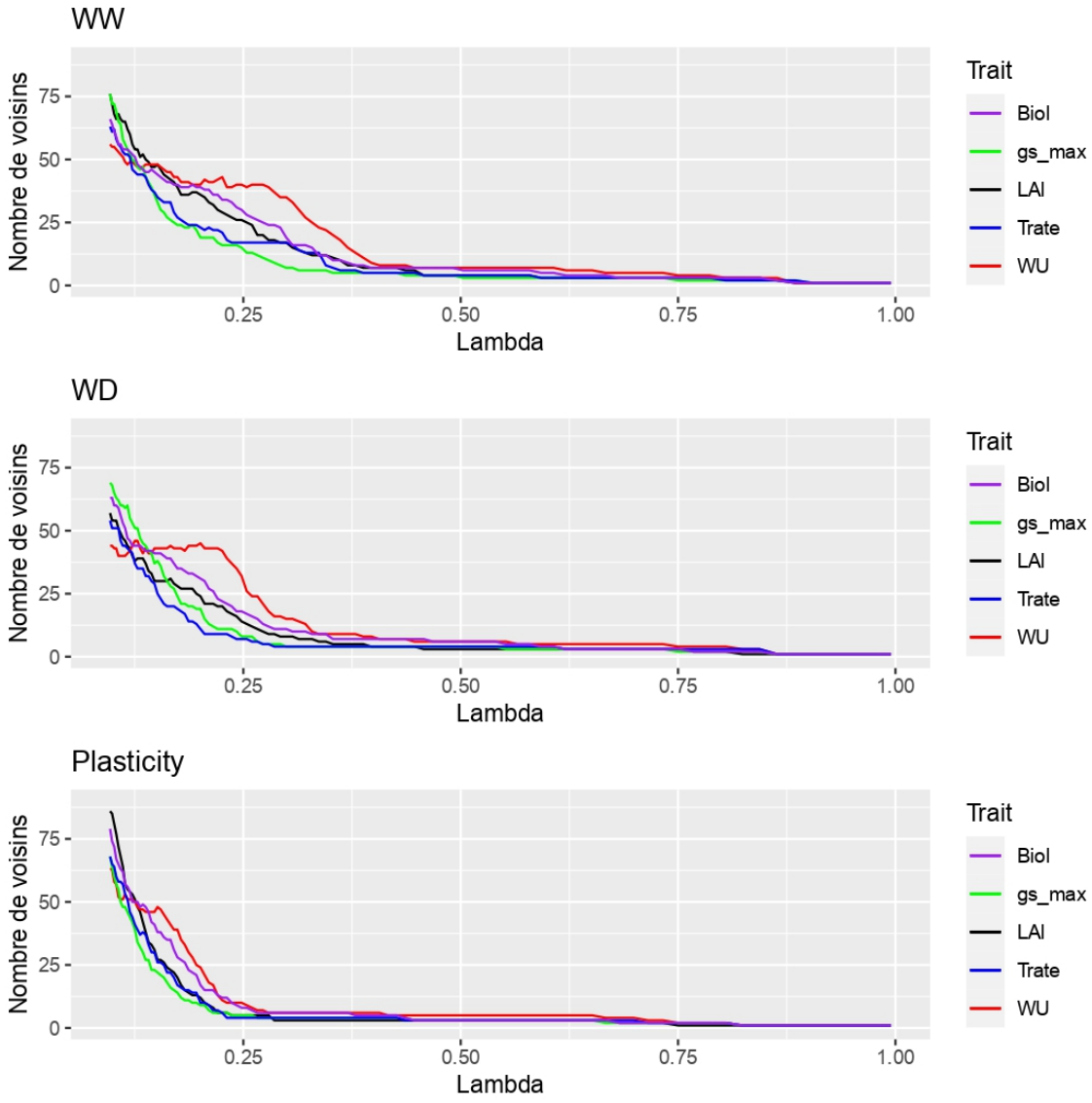


FIGURE 2 – Évolution du nombre de voisins des caractères biomass (Biol), leaf area (LAl), transpiration rate (Transp), stomatal conductance (gs_max) et water uptake (WU) en fonction de la valeur du paramètre λ, pour les trois graphes, estimés dans les deux conditions d'irrigation, WW et WD, et pour la plasticité (Plasticity).

J'observe qu'il y a des différences entre les trois GGM : il y a un décrochage de la courbe du nombre

de voisins du caractère "absorption d'eau" (WU) par rapport aux autres caractères : quand la valeur du λ diminue, le nombre de voisins du caractère "absorption d'eau" (WU) augmente plus vite que celui des autres caractères. Ce décrochage est moins marqué pour le GGM de plasticité. Le nombre de voisins des caractères écophysiologique augmente plus rapidement dans le GGM en condition d'irrigation "bien irrigué" (WW) que dans le réseau "en déficit d'eau" (WD). Dans le réseau "en déficit d'eau" (WD), il augmente plus vite que dans le réseau de plasticité. Les caractères qui ont un nombre de voisins qui augmente le plus vite sont les caractères "absorption d'eau" (WU), "biomasse" (Biom) et "surface foliaire" (Lal). Ceux qui augmentent le moins vite sont les caractères "conductance stomatique" (gs_max) et "taux de transpiration" (Trate).

3.3 Sélection d'un ensemble de mQTLs à intégrer

J'ai cherché dans chacun des réseaux à identifier les métabolites qui sont le plus lié aux caractères selon le critère de dépendance conditionnelle. Il s'agit des métabolites qui se lient en premiers aux caractères quand λ décroit. Ensuite, j'ai récupéré les mQTLs associés à ces métabolites pour les intégrer dans le modèle M4. J'ai essayé en intégrant les 5, 10 et 20 premiers métabolites à se lier aux caractères, dans les trois GGM.

J'ai calculé et comparé les gains de part de variance expliquée, par rapport au modèle M2 qui inclut uniquement les QTLs, des modèles M4 intégrant les mQTLs associés aux 5, 10 et 20 premiers métabolites (Fig. 3). Pour ce faire, j'ai soustrait les parts de variance expliquée dans le modèle M2 à celles des modèles M4, et ce, pour chaque effet aléatoire. Ces gains correspondent à un gain d'information apporté par un ensemble donné de mQTLs.

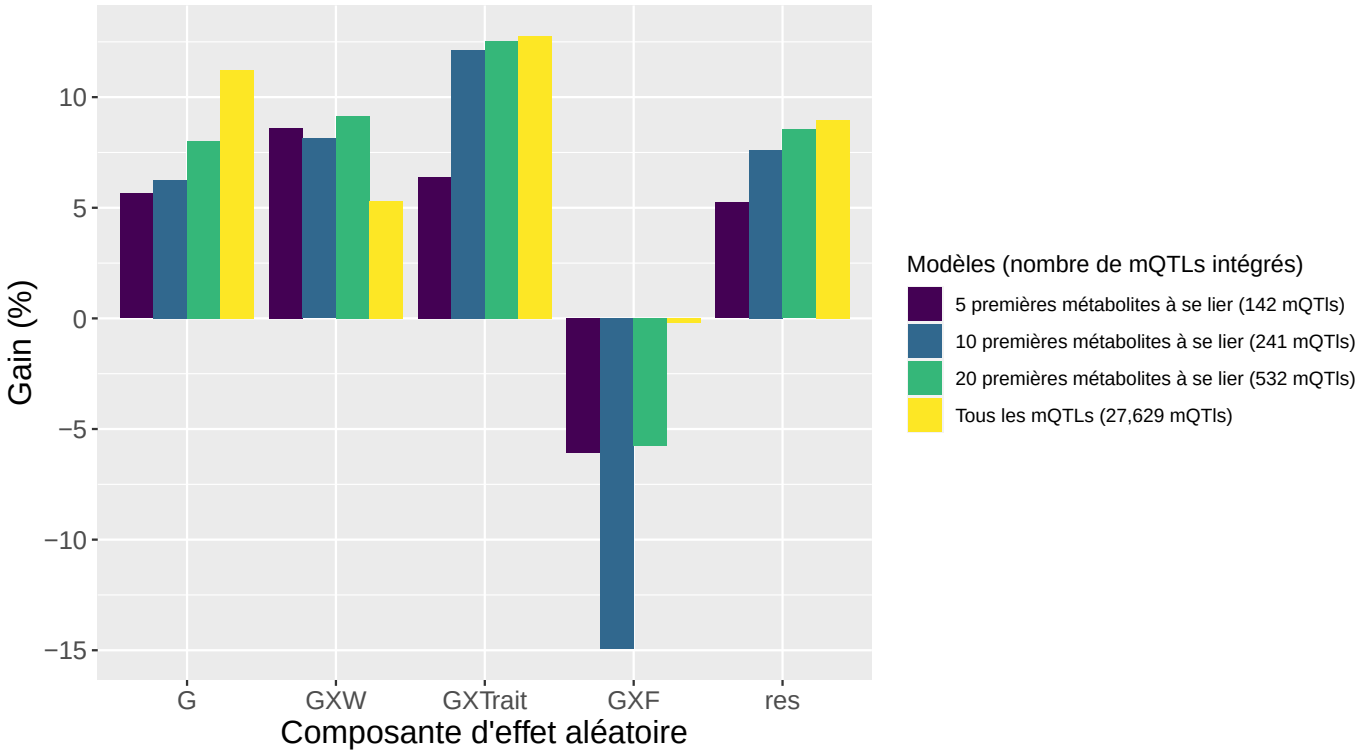


FIGURE 3 – Gains de part de variance expliquée par rapport au modèle M2, en fonction de la composante d'effet aléatoire, pour les modèles M4 intégrant les mQTLs associés aux 5,10 et 20 premier métabolites à se lier aux caractères dans trois GGM, et le modèle M4 intégrant tous les mQTLs.

J'ai observé un gain dans chacun des effets aléatoires sauf pour l'effet $G \times F$. Cette augmentation reste néanmoins inférieure à celle observée chez le modèle intégrant l'ensemble des mQTLs sauf pour $G \times W$ où elle est supérieure. Ainsi, lorsque l'on intègre un nombre restreint, mais bien choisis de mQTLs, on observe un gain comparable et parfois même supérieur au gain observé lorsque l'on intègre l'ensemble

des mQTLs. Parmi nos modèles avec un sous-ensemble de mQTLS, le modèle « 20 premiers » est celui qui maximise le gain pour les effets G et $G \times W$. Il permet des gains de part de variance expliqué pour chaque effet sauf $G \times F$ et a des résultats comparables, voir supérieur avec $G \times W$ au modèle intégrant l'ensemble des mQTLs.

4 Discussion

Bien que les données métabolomiques n'aient été produites que durant une seule année sur les quatre que comptent les expériences, l'intégration de l'ensemble des mQTLs capture de la variance génétique des caractères. Ce travail montre donc l'intérêt d'intégrer des données métabolomiques pour améliorer la compréhension du déterminisme génétique des caractères. De plus, l'intégration des mQTLs de plasticité permet un gain d'information plus important, il est donc pertinent de les intégrer aussi. Yacine Djabali avait mis au point la méthodologie à partir de données protéomiques et avait également vu un gain dans les parts variance expliquée par les effets aléatoires. Mon stage sur les données métabolomiques, suivant la même méthodologie, a permis de générer des résultats confirmant l'intérêt d'intégrer des données moléculaires dans des études d'association génétiques de caractères ecophysiologiques.

Les graphes nous ont permis de sélectionner les métabolites les plus fortement dépendantes des caractères écophysiologiques. Cependant, le critère de sélection est arbitraire et il faudrait trouver des moyens plus efficaces de sélectionner des métabolites. Yacine Djabali propose actuellement sur les protéines un critère du choix du λ qui maximise les gains de variance pour les effets aléatoires G et $G \times W$. Il serait intéressant de le faire également sur les métabolites.

Dans les trois graphes, les métabolites se lient en premier au caractère "absorption d'eau" (WU) et son nombre de voisins augmente plus vite que les autres caractères. Ainsi, dans les conditions de stress hydrique, les métabolites sont plus corrélées au caractère "absorption d'eau" qu'aux autres caractères.

Le meilleur modèle semble être le modèle intégrant les mQTLs associés aux 20 premiers métabolites qui obtient des résultats équivalents voir meilleurs que le modèle avec tout les mQTLs. Cette approche semble donc être la plus efficace. Ainsi, ce travail m'a permis de sélectionner 532 mQTLs qui pourrait permettre de trouver des gènes candidats pour de l'amélioration génétique et de créer des variétés de maïs plus résistantes à la sécheresse. De plus, ce travail a permis d'identifier 20 métabolites impliqués dans la gestion du déficit hydrique chez le maïs.

Nous n'avons pas accès aux noms des métabolites, mais si c'était le cas, on pourrait vérifier si les métabolites sélectionnés sont connues comme étant impliqués dans des voies en lien avec la résistance au déficit d'eau.

Les résultats obtenus, lorsqu'on prend les modèles intégrant les mQTLs associés aux 5,10 et 20 premiers métabolites, sont différents de ceux obtenus par Yacine Djabali sur ces mêmes modèles, avec les données protéomiques. Ici, on observe des gains plus importants sur les effets aléatoires $G \times W$ et $G \times T$, alors que pour les données protéomiques, on observe peu de gains, voir des pertes pour ces effets. A l'inverse, pour l'effet $G \times F$, on observe des gains avec les données protéomiques et des pertes pour les données métabolomiques.

Une perspective supplémentaire à un choix judicieux du serait d'intégrer dans le même modèle des données métabolomiques et des données protéomiques et ainsi voir si cela apporte un gain ou une perte d'information. On peut aussi faire le même travail, mais avec des données transcriptomiques, quantifier le gain d'information et comparer les résultats obtenus avec les données métabolomiques et protéomiques.

Bibliographie

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Bezner Kerr, R., Hasegawa, T., Lasco, R., Bhatt, I., Deryng, D., Farrell, A., Gurney-Smith, H., Ju, H., Lluch-Cota, S., Meza, F., Nelson, G., Neufeldt, H., & Thornton, P. (2022). Food, Fibre, and Other Ecosystem Products. In H. O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem & B. Rama (Éd.), *Climate Change 2022 : Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (p. 713-906). Cambridge University Press. https://doi.org/10.1017/9781009325844.007.714

Blein-Nicolas, M., Negro, S. S., Balliau, T., Welcker, C., Cabrera-Bosquet, L., Nicolas, S. D., Charcosset, A., & Zivy, M. (2020). A systems genetics approach reveals environment-dependent associations between SNPs, protein coexpression, and drought-related traits in maize. *Genome Research*, *30*(11), 1593-1604. https://doi.org/10.1101/gr.255224.119

Bouslama, M., & Schapaugh, W. T. (1984). Stress Tolerance in Soybeans. I. Evaluation of Three Screening Techniques for Heat and Drought Tolerance [1]. *Crop Science*, *24*(5), 933-937. https://doi.org/10.2135/cropsci1984.0011183X002400050026x

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

Campos, H., Cooper, M., Habben, J., Edmeades, G., & Schussler, J. (2004). Improving drought tolerance in maize : a view from industry. *Field Crops Research*, *90*(1), 19-34. https://doi.org/10.1016/j.fcr.2004.07.003

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Systems*, 1695. https://igraph.org

Daryanto, S., Wang, L., & Jacinthe, P.-A. (2016). Global Synthesis of Drought Effects on Maize and Wheat Production (D. Hui, Éd.). *PLOS ONE*, *11*(5), e0156362. https://doi.org/10.1371/journal.pone.0156362

Erenstein, O., Jaleta, M., Sonder, K., Mottaleb, K., & Prasanna, B. (2022). Global maize production, consumption and trade : trends and R&D implications. *Food Security*, *14*(5), 1295-1319. https://doi.org/10.1007/s12571-022-01288-7

Jiang, H., Fei, X., Liu, H., Roeder, K., Lafferty, J., Wasserman, L., Li, X., & Zhao, T. (2021). *huge : High-Dimensional Undirected Graph Estimation*. https://CRAN.R-project.org/package=huge R package version 1.3.5

Liu, H., Han, F., Yuan, M., Lafferty, J., & Wasserman, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *The Annals of Statistics*, *40*(4). https://doi.org/10.1214/12-AOS1037

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

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

R Core Team. (2023). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ranum, P., Peña-Rosas, J. P., & Garcia-Casal, M. N. (2014). Global maize production, utilization, and consumption. *Annals of the New York Academy of Sciences*, *1312*(1), 105-112. https://doi.org/10.1111/nyas.12396

Shiferaw, B., Prasanna, B. M., Hellin, J., & Bänziger, M. (2011). Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Security*, *3*(3), 307-327. https://doi.org/10.1007/s12571-011-0140-5

Shutta, K. H., De Vito, R., Scholtens, D. M., & Balasubramanian, R. (2022). Gaussian graphical models with applications to omics analyses. *Statistics in Medicine*, *41*(25), 5150-5187. https://doi.org/10.1002/sim.9546

Tardieu, F., Varshney, R. K., & Tuberosa, R. (2017). Improving crop performance under drought – cross-fertilization of disciplines. *Journal of Experimental Botany*. https://doi.org/10.1093/jxb/erx042