ivia
Institut Valencià
d'Investigacions Agràries

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Universitat Politècnica de València
Programa de Doctorado en Biotecnología

Ph. D. Thesis

A multi-omic study of agronomic traits of emerging interest for rice breeding using genetic tools and population resources

Concha Domingo Carrasco	**Belén Picó Sirvent**	**Julia García-Romeral Pérez**
Director	Tutor	Ph. D. Candidate

Valencia, Junio, 2025

Agradecimientos

Mi primer y mayor agradecimiento es para la dra. Concha Domingo, por tantos años respaldándome y enseñándome en cada paso dado dentro del IVIA. Gracias por confiar en mí, dándome la oportunidad de seguir trabajando juntas hasta el último momento, y por la paciencia y honestidad con la que siempre me has tratado.

A Belén Picó, por aceptar a ser mi tutora de tesis de la Universitat Politècnica de València.

Quiero agradecer también a todas las personas del departamento de Genómica en Sueca que me han acompañado a lo largo de estos años, ya sea arremangándose conmigo, con un café y una tertulia, o con un buen consejo. Gracias Rosa por ser como una madre. Ruth, por ser la compañera que me introdujo en el mundo del arroz. Manuel, per tantes hores de treball junts, però sempre amb un somriure. A tot el gran equip de camp: Fran, Mauro, Barto, Javier i Jorge, treballar amb vosaltres, encara que fora plena de tarquim, sempre va ser divertit. Gràcies pel vostre esforç. I a Isabel, Amparo, Sofia, Ferran, Toni i Empar.

Por supuesto, gracias también a todas las personas del departamento de Genómica de Moncada. Mati, has sido siempre luz y una mano amable con la que poder contar. Toni, por todas las bromas y esa capacidad de alegrar cualquier día. Javier por siempre ser una persona con la que poder hablar de todo. Y a Quico, Juan, Carles, Estela y Amparo, muchos almuerzos y cafés compartidos que no olvidaré.

Imposible no recordar con cariño a mis compañeras de despacho. Mireia, por ser la primera en llenar ese espacio de compañía. Paloma, mi gallega-valenciana, gràcies amb tot el cor per ser tot el riure compartit Y Cris, mi apoyo incondicional, mi confidente, donde nació una gran amistad.

A Jaume, l'últim regal que m'emporte del IV IA. Gràcies per ser alé i diversió. Per tot.

De mi estancia en el CRAG, a Raúl y Pep, que no solo me invitaron y acogieron con los brazos abiertos, sino que estuvieron cada día atentos, ayudándome en cada paso

e inspirándome. Y al maraviloso grupo de personas que conformaban el laboratorio, Noemia, Carlos, Lucia, Marc y Nacho, gracias por incluirme como a una más. Y a, Albert, gràcies per fer de ta casa la meua a Barna, per ser com un pare, els sopars amb jazz i grans converses.

A mi familia y a mis amigos, que tantas horas me han tenido que escuchar hablar de todo esto. Papá gracias por empujarme siempre a llegar más lejos. Mamá por querer ayudarme con todo. Mono porque sé que siempre estarás. Y a la familia que he creado contigo Juan, tú ya eras mi hogar, y ahora eres también el mejor padre que podía desear para nuestro pequeño. Gracias por hacer magia entre nosotros.

Abbreviations

DNA Deoxyribonucleic Acid
cDNA Complementary DNA
RNA Ribonucleic Acid
mRNA Messenger RNA
EDTA Ethyl-diamino-tetraacetate
FDR False Discovery Rate
GLM General Linear Model
GO Gene Ontology
GWAS Genome-Wide Association Studies
Ha Hectares
IRRI International Rice Research Institute
IVIA Instituto Valenciano de Investigaciones Agrarias
LD Linkage Disequilibrium
MAF Minimum Allele Frequency
MLM Mixed Linear Model
MAPAMA Ministry of Agriculture, Fisheries and Food
NGS Next-Generation Sequencing
PAL Phenylalanine Ammonia Lyase
PC Principal Component
PCA Principal Component Analysis
PCR Polymerase Chain Reaction
PN Panicle Number
QTL Quantitative tait Locus
RT-qPCR Reverse Transcriptase Quantitative Polymerase Chain Reaction
SDS Sodium Dodecyl Sulphate
SNP Single Nucleotide Polymorphism
Tris 2-amino-2-hydroxymethyl-1,3-propanediol

Table of Contents

Abstract

Rice is one of the world's most important crops, serving as the primary source of calories for a large portion of the global population. Continuous shifts in the demands of the rice sector and the emergence of new climatic conditions require the development of new improved varieties adapted to local environments. Additionally, rice cultivation must now address the dual and pressing challenge of reducing agrochemical inputs, particularly herbicides and nitrogen fertilizers, which are key contributors to environmental degradation in sensitive ecosystems, while still ensuring high productivity. Rice breeding is carried out locally due to the high sensitivity of rice plants to environmental factors such as photoperiod and climate. In Spain, temperate *japonica* rice is cultivated which is adapted to the Mediterranean climate.

Recent advances in genomic tools have modernized rice breeding, making it faster and more targeted. This progress has been further supported by the increasing availability of diverse genetic resources and mapping populations adapted to local conditions.

This Ph.D. thesis addresses these challenges by integrating whole-genome sequencing, RNA sequencing (RNA-seq), and field phenotyping using two complementary genetic resources: a collection of 171 temperate *japonica* accessions and a MAGIC (Multi-parent Advanced Generation Inter-Cross) population of 192 lines. The work focuses on deciphering the genetic and molecular basis of (i) allelopathic weed suppression, (ii) agronomic traits related to plant architecture and phenology, and (iii) nitrogen use efficiency (NUE).

Allelopathy, the ability of some plants to suppress weed growth through root-released compounds, was evaluated through a genome-wide association study (GWAS) on the variety collection, using 75,998 high-quality SNPs. Significant natural variation in allelopathic potential against *Echinochloa crus-galli* was observed, and four associated QTLs were identified. Candidate genes within these loci included *OsPAL* and transcription factors *OsMYB30* and *OsMYB52*. Complementary RNA-seq analysis of root tissue in two contrasting varieties under co-cultivation with *E. crus-galli* revealed that the allelopathic genotype activated

phenylpropanoid biosynthetic pathways, while genes associated with diterpene phytoalexins were not induced.

The MAGIC population was genotyped with 71,710 SNPs and used to generate a high-resolution genetic map for multiple traits. GWAS identified 35 QTLs for plant architecture and phenology, including canonical genes such as *SD1*, *RAE2*, *TAC1*, and *Ghd8*, as well as pleiotropic loci qPL9 and qH9 on chromosome 9. Evaluation of the same lines grown under contrasting nitrogen conditions improved phenotypic accuracy and enabled the identification of 21 NUE-related QTLs. Candidate genes included *ERF142/NGR5*, *OsAAP12A*, and *OsGATA8*, involved in nitrogen uptake, assimilation, and tillering regulation.

Altogether, this work demonstrates the strength of integrating diverse genetic resources, GWAS, RNA-seq, and robust field phenotyping to study complex traits and support sustainable breeding in temperate *japonica* rice.

Resumen

El arroz es uno de los cultivos más importantes del mundo, ya que constituye la principal fuente de calorías para una gran parte de la población mundial. Los continuos cambios en la demanda del sector arrocero y la aparición de nuevas condiciones climáticas exigen el desarrollo de nuevas variedades mejoradas adaptadas a los entornos locales. Además, el cultivo del arroz debe hacer frente al doble y apremiante reto de reducir los insumos agroquímicos, en particular los herbicidas y los fertilizantes nitrogenados, que son factores clave de la degradación medioambiental en ecosistemas sensibles, sin dejar de garantizar una alta productividad. El mejoramiento genético del arroz se lleva a cabo a nivel local debido a la alta sensibilidad de las plantas de arroz a factores ambientales como el fotoperíodo y el clima. En España se cultiva arroz *japonica* templado, adaptado al clima mediterráneo.

Los recientes avances en las herramientas genómicas han modernizado el cultivo del arroz, haciéndolo más rápido y específico. Este progreso se ha visto respaldado por la creciente disponibilidad de poblaciones genéticas y nuevas variedades.

Esta tesis doctoral aborda estos retos mediante la integración de la secuenciación del genoma completo, la secuenciación del ARN (RNA-seq) y el fenotipado de campo utilizando dos recursos genéticos complementarios: una colección de 171 accesiones *japonica* templada y una población MAGIC (Multi-parent Advanced Generation Inter-Cross) de 192 líneas. El trabajo se centra en descifrar las bases genéticas y moleculares de (i) la capacidad alelopática del arroz para supresión de malas hierbas, (ii) los caracteres agronómicos relacionados con la arquitectura y la fenología de las plantas, y (iii) la eficiencia en el uso del nitrógeno (NUE).

La alelopatía, la capacidad de algunas plantas para suprimir el crecimiento de las malas hierbas a través de compuestos liberados por las raíces, se evaluó mediante un estudio de asociación del genoma completo (GWAS) en la colección de variedades, utilizando 75,998 SNPs de alta calidad. Se observó una variación natural significativa en el potencial alelopático contra *Echinochloa crus-galli* y se identificaron cuatro QTLs asociados. Entre los genes candidatos dentro de estos loci se encontraban *OsPAL* y los factores de transcripción *OsMYB30* y *OsMYB52*. El

análisis de RNA-seq de las raices en dos variedades contrastantes en cocultivo con *E. crus-galli* reveló que el genotipo alelopático activaba la vía biosintética de los fenilpropanoides, mientras que los genes asociados con la via de los diterpenos no se inducían.

La población MAGIC se genotipó obteniendo 71,710 SNPs, y se utilizó para generar un mapa genético de alta resolución para múltiples caracteres. El GWAS identificó 35 QTL para la arquitectura y la fenología de la planta, incluidos genes canónicos como *SD1*, *RAE2*, *TAC1* y *Ghd8*, así como los loci pleiotrópicos qPL9 y qH9 en el cromosoma 9. La evaluación de las mismas líneas en condiciones de nitrógeno contrastantes mejoró la precisión fenotípica y permitió la identificación de 21 QTL relacionados con la NUE. Los genes candidatos incluyeron *ERF142/NGR5*, *OsAAP12A* y *OsGATA8*, implicados en la absorción de nitrógeno, la asimilación y la regulación de las panículas.

En conjunto, este trabajo demuestra la solidez de la integración de diversos recursos genéticos, GWAS, RNA-seq y fenotipado robusto en campo para estudiar rasgos complejos y apoyar el mejoramiento sostenible del arroz *japonica* templado.

Resum

L'arròs és un dels cultius més importants del món, ja que constituïx la principal font de calories per a una gran part de la població mundial. Els continus canvis en la demanda del sector arrosser i l'aparició de noves condicions climàtiques exigixen el desenrotllament de noves varietats millorades adaptades als entorns locals. A més, el cultiu de l'arròs ha de fer front al doble i urgent repte de reduir els inputs agroquímics, en particular els herbicides i els fertilitzants nitrogenats, que són factors clau de la degradació mediambiental en ecosistemes sensibles, sense deixar de garantir una alta productivitat. El millorament genètic de l'arròs es du a terme a nivell local a causa de l'alta sensibilitat de les plantes d'arròs a factors ambientals com el fotoperíode i el clima. A Espanya es cultiva arròs *japonica* temperat, adaptat al clima mediterrani.

Els recents avanços en les ferramentes genòmiques han modernitzat el cultiu de l'arròs, fent-lo més ràpid i específic. Este progrés s'ha vist recolzat per la creixent disponibilitat de poblacions genètiques i noves varietats.

Esta tesi doctoral aborda estos reptes mitjançant la integració de la seqüenciació del genoma complet, la seqüenciació de l'ARN (RNA-seq) i el fenotipat en camp durant diversos anys utilitzant dos recursos genètics complementaris: una col·lecció de 171 varietats *japonica* temperada i una població MAGIC (Multi-parent Advanced Generation Inter-Cross) de 192 línies. El treball se centra en desxifrar les bases genètiques i moleculars de (i) la capacitat alelopàtica de l'arròs per a la supressió de males herbes, (ii) els caràcters agronòmics relacionats amb l'arquitectura i la fenología de les plantes, i (iii) l'eficiència en l'ús del nitrogen (NUE).

La alelopatía, la capacitat d'algunes plantes per a suprimir el creixement de les males herbes a través de compostos alliberats per les arrels, es va avaluar mitjançant un estudi d'associació del genoma complet (GWAS) en la col·lecció de accessions, utilitzant 75,998 SNPs d'alta qualitat. Es va observar una variació natural significativa en el potencial alelopàtic contra *Echinochloa crus-galli* i es van identificar quatre QTLs associats. Entre els gens candidats dins d'estos loci es trobaven *OsPAL* i els factors de transcripció *OsMYB30* i *OsMYB52*. L'anàlisi transcriptòmic de les arrels en dos varietats contrastants en cocultiu amb *E. crus-*

galli va revelar que el genotip alelopàtic activava la via biosintètica dels fenilpropanoids, mentres que els gens associats amb la via dels diterpenoids no s'induïen.

La població MAGIC es va genotipar obtenint 71,710 SNPs, i es va utilitzar per a generar un mapa genètic d'alta resolució per a múltiples caràcters. El GWAS va identificar 35 QTLs per a l'arquitectura i la fenología de la planta, inclosos gens canònics com *SD1*, *RAE2*, *TAC1* i *Ghd8*, així com els loci pleiotròpics qPL9 i qH9 en el cromosoma 9. L'avaluació de les mateixes línies en condicions de nitrogen contrastants va millorar la precisió fenotípica i va permetre la identificació de 21 QTLs relacionats amb la NUE. Els gens candidats van incloure *ERF142/NGR5*, *OsAAP12A* i *OsGATA8*, implicats en l'absorció de nitrogen, l'assimilació i la regulació de les panícules.

En conjunt, este treball demostra la solidesa de la integració de diversos recursos genètics, GWAS, RNA-seq i fenotipat robust en camp per a estudiar trets complexos i donar suport al millorament sostenible de l'arròs *japonica* temperat.

General Introduction

1. Rice

1.1. Evolutionary history and domestication: origin, subpopulations and global expansion.

Rice, one of the most important crops worldwide, belongs to the Poaceae family of grasses, genus Oryza, which includes 21 recognised species, of which only two are cultivated: Oryza sativa L. (Asian rice) and O. glaberrima Steud. (African rice) (Khush, 1997; Vaughan et al., 2003) (Figure 1).

Both species are diploid (2n = 24), although they differ significantly in their evolutionary and geographical history. While O. glaberrima was independently domesticated in West Africa from O. barthii, O. sativa was domesticated from the wild species Oryza rufipogon in a process that began approximately 9,000 years ago in the Yangtze River basin in southern China (Fornasiero et al., 2022; Fuller, 2011). The first domesticated was the *japonica* genetic population, selected for traits such as loss of seed dehiscence, increased grain size, and reduced sensitivity to photoperiod. Subsequently, domesticated *japonica* populations expanded southward and eastward into Asia, where they hybridised with local populations of O. rufipogon and O. nivara. This process of recurrent introgression, followed by selection, gave rise to the indica and aus genetic groups, adapted to the tropical and monsoon conditions of South Asia (Choi & Purugganan, 2018; Huang et al., 2012). This model of staggered domestication with local hybridisation replaces previous hypotheses that proposed completely independent domestications.

At the genetic and ecological level, O. sativa is structured into five main subpopulations: indica, aus, tropical *japonica*, temperate *japonica* and aromatic (Garris et al., 2005a; Sweeney & McCouch, 2007). These differ in morphological, agronomic and adaptive characteristics, particularly the sensitivity to photoperiod, reflecting both their evolutionary history and the environmental contexts in which they have developed.

Indica varieties dominate in tropical regions of Asia such as India, Sri Lanka, Thailand and Malaysia, and show great genetic diversity, while *japonica* varieties, which include the subtropical and temperate groups, show reduced genetic diversity (Glaszmann, 1987; Q. Zhang et al., 1992). The temperate varieties, which appear to

derive from the tropical varieties, are cultivated in colder climates and spread later from Southeast Asia (Huang et al., 2012). These are widely grown in China, Korea, Japan, as well as in Europe, North America and Australia. Aus varieties, adapted to more northern conditions such as Bangladesh, are characterised by their flowering in long days and a certain reproductive incompatibility from indica (Parsons et al., 1999). The aromatic group has been described as an intermediate lineage between indica and *japonica*, although it has also been proposed as a separate subpopulation, with reduced genetic diversity pointing to recent and intense selection (Garris et al., 2005b; Nagaraju et al., 2002).

In parallel, a completely independent process of domestication occurred in West Africa around 3,000 years ago, where populations of the wild species Oryza barthii were selected, giving rise to African rice (O. glaberrima) in the Niger River delta region (Fornasiero et al., 2022). Despite its sparsely cultivated today, O. glaberrima is a valuable genetic resource, especially for its tolerance to biotic and abiotic stresses.

Finally, it is important to mention that the domestication process not only involved changes in grain morphology and plant architecture, but also a series of ecological and physiological adaptations that have allowed rice to expand from tropical areas to temperate environments at high latitudes and altitudes, making it a crop with global reach (Fuller et al., 2010).

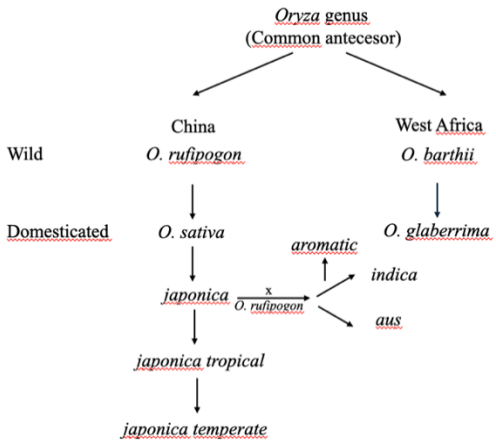


Figure 1. Domestication pathways from the *Oryza* genus (according to Huang et al. 2012; Choi & Purugganan 2017 and Fornasiero et al. 2022).

1.2. Economic and social relevance of the crop

Rice is one of the world's most important crops, not only from a food perspective, but also economically, socially and culturally. It is the staple food for more than half of the world's population and provides approximately 20% of the total calories consumed globally (Fukagawa & Ziska, 2019), a proportion that can reach between 50% and 80% in many Asian countries, depending on the level of dietary dependence (FAO, 2003; Muthayya et al., 2014). The growing global demand, driven by population growth, especially in Asia, suggests a continued increase in rice consumption at an estimated rate of 0.9% per year over the next decade (OECD/FAO, 2023). This trend is already reflected in the available data: global consumption rose from 437 million tonnes in the 2008/2009 season to an estimated 523 million tonnes in 2024/2025 (Abigail, 2024) (**Figure 2**).

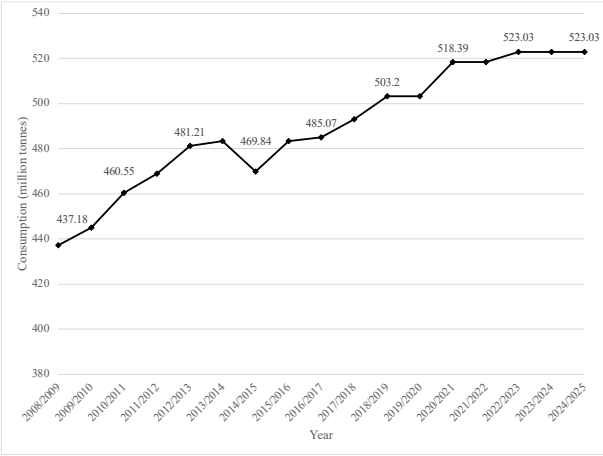


Figure 2. Global rice consumption from 2008/2009 to 2024/2025 in million tonnes (from www.statista.com).

In 2023, global production was estimated at 776 million tonnes, with more than 90% concentrated in Asia, with China and India as the main producers. In contrast, Europe accounts for only 1% of the total, and its production has stagnated in recent decades. Italy and Spain account for 80% of European rice, although in 2023, continental production fell below 2.2 million tonnes, one of the lowest figures recorded recently (FAO, 2024) (**Figure 3**).

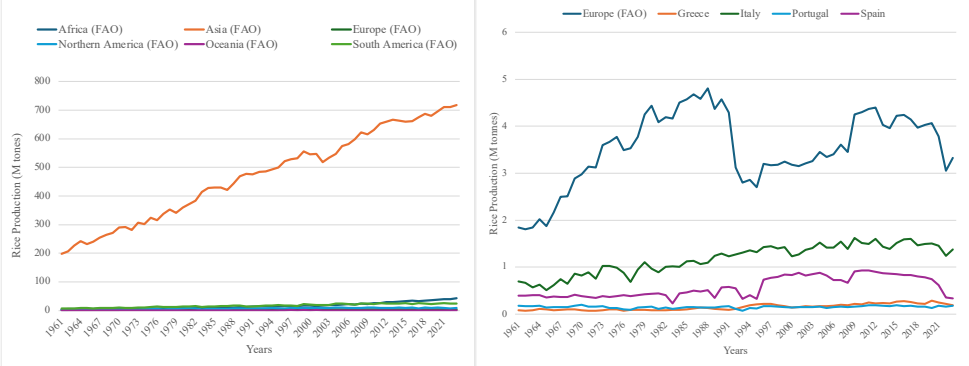


Figure 3. Rice production from 1961 to 2023, measured in million tonnes. Left panel: by continent. Rigth panel: Top 4 rice countries in Europe (from FAO, 2024).

Spain, as the second largest European producer, has suffered a sharp decline in both area and volume. In 2023, the area under cultivation fell to 55,178 hectares, 39% less than the average of the previous five seasons, with a total production of 327,257 tonnes. National production stood at 327,257 tonnes, far below the historical levels of around 800,000 tonnes in previous years (MAPA, 2023). This decline is related to water constraints due to severe droughts in 2023 and 2024, the increase in the cost of inputs and the loss of crop profitability. As a result, Spain has become a net importer, exceeding 600,000 tonnes in 2022, mostly from Myanmar, Pakistan and Uruguay (OECD/FAO, 2023) (**Table 1**).

Year	Surface (kha)	Yield (t/ha)	Production (kt)	Price (€/100kg)
2012	112.6	7.97	897.3	27.66
2013	112	7.97	872.7	27.51
2014	109.9	7.72	848	28.33
2015	109.7	7.76	847	27.68
2016	109.3	7.64	835.2	30.06
2017	107.6	7.76	835.2	29.01
2018	105	7.68	806.6	29.19
2019	103.4	7.62	787.8	30.51
2020	102.1	7.38	754.7	31.7
2021	84.7	7.37	624.8	38.77
2022	56	6.33	354.5	47.26

Table 1. Annual rice cultivation data in Spain from 2012 to 2022, including harvested surface (in thousand hectares), yield (tonnes per hectare), total production (in thousand tonnes), and average price received by farmers (euros per 100 kg). (from OECD/FAO, 2023)

Despite this decline, Spain maintains high production efficiency. The estimated average yield for 2024 stands at 7,758 kg/ha, representing a 20% increase over the

previous year and placing the country among the most productive in the world, on a par with Oceania and above Asia, America and Africa (FAO, 2025; MAPA, 2023). This high productivity is the result of a combination of specialised agricultural practices, a strong technical tradition and good use of varietal resource (**Figure 4**).

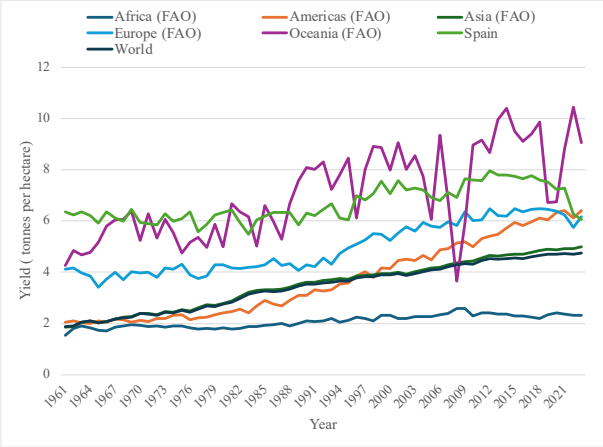


Figure 4. Evolution of rice yield (tonnes per hectare) from 1961 to 2021 across continents and in Spain (from FAO, 2024).

In this context, it is important to note that the rice grown in Spain belongs to the temperate *japonica* group, which is widely adapted to the agroclimatic conditions of the Mediterranean coast. The most widely grown variety in Spain is medium-grain pearl rice, which is preferred by the domestic industry for its culinary characteristics. Medium grain is particularly prevalent in regions such as Catalonia and the Valencian Community, while long-grain rice, mainly grown in Andalusia and Extremadura, is mainly destined for export. At the regional level, Andalusia and Extremadura led national production, but droughts in recent years have prevented sowing in these two regions, so that Catalonia now leads rice production in Spain with more than 132,000 tonnes in 2022, followed by the Valencian Community, which reached approximately 91,000 tonnes, mostly from the Albufera Natural Park (MAPA, 2023) (**Table 2**). This area became environmentally protected in 1986 and represents a farming system conditioned by restrictions on the use of chemical inputs, a predominantly small-scale farming structure and a declining area, which poses particular challenges for sustainability and varietal innovation.

Regional Comunity	Total Production (t)			Yield (t/ha)	
	Large	Medium/round	Total	Large	Medium/round
Andalucia	53,316	34,468	87,784	6.3	8.2
Cataluña	21,946	110,292	132,238	6.2	6.3
C. Valenciana	2,942	87,906	90,848	6.3	6.2
Aragon	185	17,720	17,905	5.6	5.6
Extremadura	5,127	7,755	12,882	6.0	7.0
R. de Murcia	-	1,968	1,968	-	4.6
Navarra	-	10,313	10,313	-	6.1
Castilla-La Mancha	-	514	514	-	5.0
Baleares	-	38	38	-	1.5

Table 2. Rice production and yield by grain type and regional community in Spain (from MAPA, 2023) .

In this context, regulatory pressure, rising costs and competition from imported rice are compromising the crop's profitability. Genetic improvement aimed at more efficient and resilient varieties is an essential way to sustain productivity and reduce dependence on inputs in agro-environmentally sensitive environments.

1.3. Ecology, cycle and morphology of the crop

Rice is an annual herbaceous monocotyledonous plant which, like many other cereals, is adapted to warm, humid conditions. It is characterised by remarkable ecological plasticity, allowing it to grow from sea level to altitudes of over 3,000 metres, and in environments ranging from tropical ecosystems to temperate zones. This adaptability is explained by intraspecific genetic diversity and the development of varieties with different sensitivities to photoperiod and tolerance to extreme abiotic conditions (Hori et al., 2015).

Rice is mainly grown in flooded fields, which account for more than 75% of the world's rice-growing area (Bouman et al., 2005). This practice facilitates the absorption of nutrients, especially nitrogen and phosphorus, reduces weed growth and acts as a physical barrier against soil pathogens. However, there is also a segment of the crop grown in systems where irrigation is provided by rainfall. This is also known as ‘upland rice’, particularly in tropical or semi-tropical areas of Asia and

Africa, although with low yields and great variability in productivity (Gupta & O'toole, 1986).

From a phenological point of view, the rice cycle comprises three distinct phases: the vegetative phase, which includes germination to panicle initiation; the reproductive phase, which includes panicle initiation to flowering; and finally the ripening phase, which includes flowering to mature grain. The total duration of the cycle varies widely depending on the variety and agroclimatic conditions, ranging from 90 to 180 days from germination to maturity (Yoshida, 1981) (Figure 5).

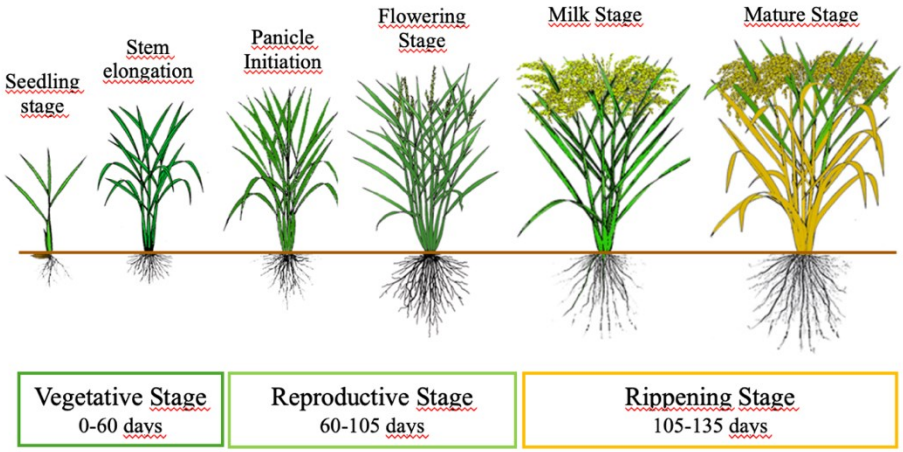


Figure 5. Growth stages of the rice plant: Vegetative, Reproductive and Rippening.

From a morphological perspective, the plant has a fasciculated root system, comprising primary and adventitious roots, which enables efficient water and nutrient absorption in saturated environments. The stem, also called culm, is cylindrical, hollow and divided into nodes and internodes, from which one leaf emerges per node. The leaves are elongated and linear, and are composed of a basal sheath that surrounds the stem, a blade or limb, and two specialised structures in the insertion zone, the ligule and the auricles, which are fundamental in studies of leaf development and varietal differentiation (Olmos, 2007) (**Figure 6**).

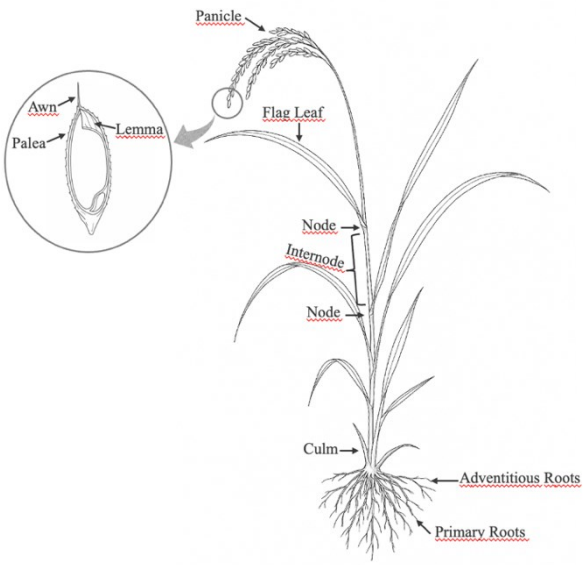


Figure 6. Diagram of morphological structures of an adult rice plant. Close-up of grain structures.

During the vegetative phase, the plant produces lateral shoots called tillers, which can generate additional panicles in the reproductive stage. The panicle is the inflorescence of rice and is made up of primary and secondary axes bearing spikelets; each of these contains a single fertile flower, protected by structures called lemma and palea, which may have an awn in some varieties. The number of spikelets per panicle, as well as their fertility, are closely linked to the genotype and growing conditions. After fertilisation, the flower gives rise to the grain, which is structured into glumes (husks), starch-rich endosperm and an embryo, whose quality and composition are key aspects in genetic improvement programs due to their influence on both the productivity and nutritional value of rice (Olmos, 2007).

1.4. Agronomic characteristics. Adaptation to sensitive environments: NUE and allelopathy

The yield and productive stability of rice depend largely on a set of agronomic traits that have historically been prioritised in breeding programs. These include plant height, number and length of panicles, resistance to lodging, length of the vegetative cycle, plant architecture such as erect leaf posture, which supports high plant density, and the fact that the basal nodes of the plant receive more solar radiation (Olmos,

2007). However, in recent decades, genetic improvement objectives have diversified to address new challenges arising from climate change and environmental protection restrictions. These include tolerance to abiotic stresses such as drought, salinity and submergence, and biotic stresses such as resistance to diseases such as blast, caused by the fungus *Magnaporthe oryzae*. Most of these traits are quantitatively inherited, controlled by multiple genes with small effects, which makes them difficult to identify, characterise and improve using conventional phenotypic selection methods. Therefore, their study requires integrated approaches that combine molecular genetic tools, such as QTL identification, marker-assisted selection (MAS) or functional genomics, which allow for more precise and efficient improvement of these complex traits (Siddiq & Vemireddy, 2021).

In addition to these widely addressed objectives, there are other traits of growing agronomic interest that respond to local production contexts and specific sustainability requirements. In this regard, this study focuses on two strategic traits for rice cultivation in environmentally sensitive conditions: nitrogen use efficiency (NUE) and allelopathic potential against weeds, particularly *Echinochloa crus-galli*.

Efficient nitrogen use is extremely important in situations where fertiliser use must be limited for environmental reasons, such as in protected areas or crops grown under agroecological standards. Improving NUE allows high yields to be maintained with less nitrogen fertiliser, which not only reduces production costs but also mitigates aquifer contamination from excessive fertilisation. From a physiological point of view, NUE is defined as the ability of a plant to absorb nitrogen from the soil—mainly as nitrate (NO_3^-) and ammonium (NH_4^+)—transport it to the aerial organs, assimilate it into organic compounds (such as amino acids and proteins), and mobilise it efficiently to the reproductive organs (**Figure 7**). In recent years, numerous genes involved in these processes have been identified, such as *NRT1.1B* (nitrate transport), *OsNR2* and *OsNiR* (reduction and assimilation), *OsNAC42*, *OsNLP4* and *OsWRKY23* (transcriptional regulation), as well as *DEP1* (nitrogen architecture and partitioning). Although many of these studies were conducted on *indica* varieties, their functional validation in *japonica* varieties has progressed significantly; however, there is still considerable room for improvement in the identification and application of genes specifically adapted to *japonica*, especially under real growing conditions and with limited inputs (Hu et al., 2015; Tang et al., 2019; Yu et al., 2021; Y. Zhang et al., 2021).

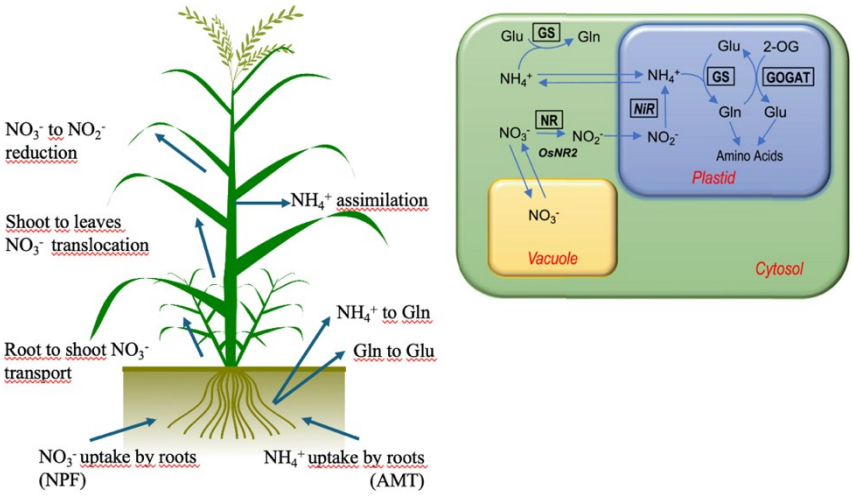


Figure 7. Nitrogen uptake, transport, assimilation, and remobilization in the plant nitrogen cycle (adapted from Wang et al. 2022).

At the phenotypic level, the assessment of NUE can be achieved using yield traits in which the number of panicles and grain weight reflect the efficiency with which the absorbed nitrogen is converted into yield (Koutroubas & Ntanos, 2003; Wei et al., 2012; Sandhu et al., 2021).

On the other hand, weed control without the use of chemical herbicides is a priority challenge in regions where their application is increasingly restricted. In this context, allelopathy, understood as the ability of a plant to release bioactive compounds that affect the development of competing species, is emerging as a sustainable management alternative. In rice, this interaction occurs mainly through root exudates, crop residues and volatile compounds, which interfere with key weed processes such as germination, root elongation and cell division (**Figure 8**) (Olofsdotter M, 1998).

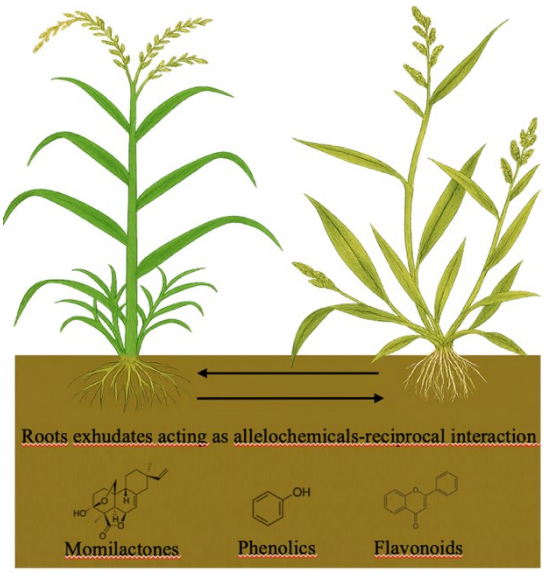


Figure 8. Allelopathic interaction between rice (left) and barnyardgrass (right) mediated by root exudates (from Olofsdotter M, 1998)

Among the most relevant compounds involved in allelopathy are phenolic acids (such as p-coumaric acid), flavonoids such as tricin and momilactones A and B, and diterpenoids exclusive to rice with phytotoxic effects. These compounds have been shown to effectively inhibit the growth of some of the main weeds in rice fields (Kato-Noguchi & Peters, 2013; Kong et al., 2006). Most functional studies have focused on the *japonica* variety PI312777 (Fang et al., 2020; Song et al., 2008), widely recognised for its allelopathic potential at the molecular level. Key genes involved in the biosynthesis and regulation of these compounds have been identified, such as *OsPAL*, *OsCPS4*, *OsMAS* and the transcription factor *OsMYB57*, whose activation significantly improves the allelopathic response (Rahaman et al., 2022). However, many varieties remain to be characterised and it is expected to find alternative mechanisms of action and different allelochemicals that have yet to be identified.

The development of rice varieties with allelopathic potential against invasive species such as *Echinochloa crus-galli*, which is one of the most aggressive weeds in the Mediterranean, capable of causing losses of up to ~70%, represents a promising strategy for reducing dependence on synthetic herbicides.

2. Targeted genetic improvement

2.1. Genetic basis of rice. Its value as a model plant.

Rice has traditionally been one of the plant species of greatest interest in genetic and genomic studies, not only because of its agricultural importance, but also because of its role as a model in the study of monocotyledons. This status is due, among other factors, to its small genome (373 Mb in *japonica* and 389–391 Mb in *indica*), the most compact among the major cereals, as well as its relatively low number of chromosomes (2n = 24), its short life cycle, ease of genetic manipulation, and the existence of extensive genetic and genomic resources.

In 2002, the International Rice Genome Sequencing Project (IRGSP) published the reference sequence of the *japonica* subspecies (cv. Nipponbare) for the first time, making it the first cereal with a completely sequenced genome (Goff et al., 1999; Matsumoto et al., 2005). Subsequently, the sequence of the *indica* subspecies (cultivar 93-11) was published using an independent assembly strategy (Yu et al., 2002), allowing the two subspecies to be compared at the genomic level. Since then, the genome of multiple rice varieties has been sequenced and a pangenome of rice has been generated (Qin et al. 2021). Moreover, the novo assembly and comparison of genomic data of an accession panel comprising both cultivated and wild species, including *O. glaberrima* and *O. rufipogon*, has led to the generation of a pangenome reference of wild and cultivated rice (Shang et al 2022).

The functional information derived from the sequencing is collected in databases such as the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu), which includes more than 66,000 annotated transcripts, of which 46.8% correspond to putative genes and 26% are associated with transposable elements. At the same time, the Japanese database RAP-DB (https://rapdb.dna.affrc.go.jp/) offers complementary annotations focused on the Nipponbare genome and its functionality. In addition, genome annotations are also available at the Gramene Pan Oryza database (https://oryza.gramene.org/) and the genome sequences and assemblies at the Rice Population Reference Panel (RPRP, https://yongzhou2019.github.io/Rice-Population-Reference-Panel/data/).

This wealth of information has driven the development of mutant collections, stable genetic transformation protocols and introgressed lines, consolidating rice as a reference model system for functional genomics, adaptation and genetic improvement studies in grasses.

2.2. Structured genetic resources. Populations and collections.

The genetic study of agronomic traits in rice relies on a wide variety of genetic resources, including both variety collections and experimental populations designed for mapping loci of interest. International germplasm banks conserve more than 780,000 rice accessions, including traditional and modern varieties and wild species. The International Rice Genebank (IRRI) houses 128,000 accessions, while resources such as the USDA Rice Genetic Stock Centre and OryzaBase complement this diversity with mutant, introgressed and experimental lines.

A key milestone in the characterisation of this diversity was the launch of the 3K Rice Genome Project in 2014, which sequenced 3,024 rice accessions from 89 countries, identifying more than 18.9 million SNPs. This information was integrated into the SNP-Seek database (http://snp-seek.irri.org), which allows comparative exploration of the genetic diversity and phenotypic characteristics of these varieties, constituting an essential platform for association studies and assisted selection.

The use of natural variety panels, which include both traditional accessions and modern cultivars and breeding lines, allows the high genetic diversity present in the species to be directly exploited, enabling the identification of multiple functional variants for the same trait. While their analysis may involve certain challenges, such as population structure and the low frequency of some alleles, which can make difficult to detect genetic associations, these panels are nonetheless ideal for the discovery of novel phenotypes that would be unlikely to emerge in pre-designed experimental populations.

On the other hand, biparental populations, such as recombinant inbred lines (RILs), near-isogenic lines (NILs) or F2 populations, have been fundamental in QTL mapping thanks to their simple and controlled genetic structure. However, they have low allelic diversity and limited resolution. More advanced designs such as NAM (Nested Association Mapping) or BILs (Backcross Inbred Lines) expand the available genetic range, but their use in rice remains more restricted due to crossing barriers between subspecies.

In this context, MAGIC (Multiparent Advanced Generation InterCross) populations have established themselves as a reference tool for the genetic study of complex traits in rice. These combine multiple parental lines (usually between 4 and 16) to maximise recombination and capture greater genetic diversity within a structured and controlled population. In rice, MAGIC populations have proven particularly useful for decomposing complex quantitative traits, thanks to their balanced design and their potential to detect both common and rare alleles (Bandillo et al., 2013)(**Figure 9**).

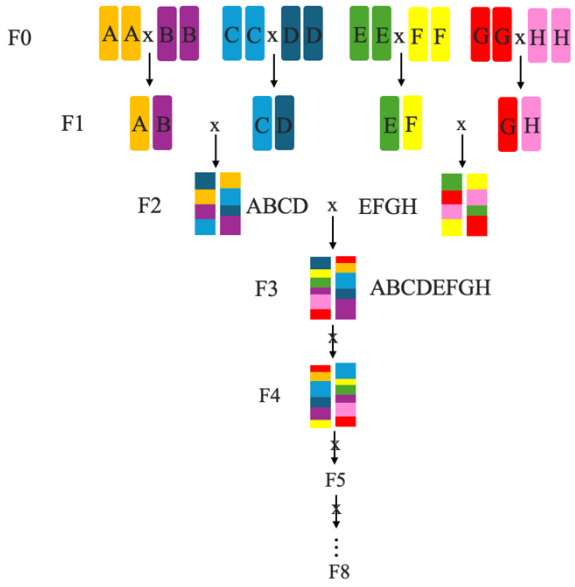


Figure 9. Crossing scheme used to generate a MAGIC population through sequential intercrossing of eight founder lines.

By combining multiple parental lines (usually between 4 and 16), MAGIC maximises recombination, increases allelic diversity and allows high-resolution genetic analysis. Their balanced design overcomes the diversity constraints of biparental populations and the statistical limitations of natural panels, facilitating the detection of both common and rare alleles. Thanks to their versatility, these populations are used in both QTL mapping and genomic selection.

2.3. Tools for targeted genetic improvement

2.3.1 Massive sequencing (WGS) and GWAS

The emergence of next-generation sequencing (NGS) technologies has transformed the genomic research in rice, enabling rapid, accurate and low-cost whole genome sequencing (WGS). WGS sequencing is based on the fragmentation of genomic DNA, its subsequent amplification, and the massive reading of millions of short fragments using platforms such as Illumina, which allow them to be assembled according to their overlap. This approach provides a comprehensive view of the genomic architecture, identifies single nucleotide polymorphisms (SNPs), insertions/deletions and structural variations, thus facilitating the genetic characterisation of individuals and populations.

Thanks to this accessibility, large-scale genetic variation maps have been generated, and new strategies for studying complex traits have emerged, such as genome-wide association studies (GWAS). This type of analysis allows phenotypic variation to be correlated with genetic polymorphisms, mainly SNPs, in large panels of genetically diverse accessions. Unlike classical mapping with biparental populations, GWAS allows multiple natural variants affecting the same agronomic trait to be identified, taking advantage of the diversity present in global collections (Ingvarsson & Street, 2011; Yu et al., 2006).

Among the resources available for this type of analysis, the SNP-Seek database (https://snp-seek.irri.org) stands out, which centralises data from the 3K Rice Genome Project, offering access to more than 18 million SNPs genotyped in more

than 3,000 rice accessions. On the other hand, the QTARO database (https://qtaro.abr.affrc.go.jp) compiles curated information on loci and QTLs associated with relevant agronomic traits in rice, integrating data from genetic maps and functional studies, and allowing the exploration of the location, function and validation of QTLs reported in the literature.

Once the associated SNPs or genomic regions have been identified, they can be used in breeding programs through marker-assisted selection (MAS), without the need for complete phenotyping. This approach shortens selection cycles and improves the efficiency of conventional programs (Lam et al., 2010). In addition, these strategies have been refined through integration with complementary tools, such as transcriptomic analysis, mutant lines and functional annotation, which allow candidate genes to be validated with greater accuracy (Wang & Han, 2022).

2.3.2. RNA-seq

Among these tools, RNA sequencing (RNA-seq) stands out as one of the most powerful techniques to emerge from the development of NGS, largely replacing microarrays as the gold standard for transcriptome analysis. Unlike the genome, which remains relatively stable, the transcriptome, the set of all RNAs expressed by a cell or tissue at a given time, varies according to cell type, stage of development or environmental conditions.

RNA-seq allows the expression of all transcripts present in a sample to be quantified without the need for a set of predefined probes, as in microarrays. Its methodology is based on the extraction of RNA (total or enriched in mRNA), conversion to cDNA, and subsequent sequencing using platforms such as Illumina. The short reads obtained (30–400 bp) are aligned against a reference genome or transcriptome, or assembled de novo, allowing the transcriptional structure to be reconstructed and the expression levels of each gene to be estimated. Furthermore, RNA-seq does not present problems of saturation or background noise, which are common in previous methodologies.

In rice, this technique has been widely applied to identify differentially expressed genes between contrasting genotypes or in response to biotic and abiotic stresses, providing key functional information for genetic improvement. More recently, RNA-seq has begun to be applied at the single-cell level (scRNA-seq), allowing high-resolution analysis of transcriptomic dynamics in complex tissues, such as roots or reproductive organs, and the study of cell differentiation trajectories (Wang & Han, 2022).

To facilitate access to analysis of transcriptomic data, there are specialised databases such as Gramene (www.gramene.org), which integrates genomic and expression information on *O. sativa* through Ensembl Plants, allowing expression profiles to be visualised in different tissues and omic data to be superimposed on metabolic pathways. Also noteworthy is CARMO (http://bioinfo.sibs.ac.cn/carmo), a platform designed specifically for the functional and statistical analysis of RNA-seq data in rice, which offers tools for the detection of differentially expressed genes, functional enrichment (GO/KEGG), and the construction of co-expression networks.

Bibliography

Abigail, O. (2024). Consumo mundial de arroz entre 2008 y 2025. Statista. https://es.statista.com/estadisticas/598941/consumo-total-de-arroz-en-el-mundo-2008/#statisticContainer

Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C. J., Tung, C. W., McCouch, S., Thomson, M., Mauleon, R., Singh, R. K., Gregorio, G., Redoña, E., & Leung, H. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: Progress and potential for genetics research and breeding. *Rice*, *6*(1), 1–15. https://doi.org/10.1186/1939-8433-6-1

Bouman, B. A. M., Peng, S., Castañeda, A. R., & Visperas, R. M. (2005). Yield and water use of irrigated tropical aerobic rice systems. *Agricultural Water Management*, *74*(2), 87–105. https://doi.org/10.1016/j.agwat.2004.11.007

Choi, J. Y., & Purugganan, M. D. (2018). Multiple origin but single domestication led to Oryza sativa. *G3: Genes, Genomes, Genetics*, *8*(3), 797–803. https://doi.org/10.1534/g3.117.300334

Fang, C., Yang, L., Chen, W., Li, L., Zhang, P., Li, Y., He, H., & Lin, W. (2020). MYB57 transcriptionally regulates MAPK11 to interact with PAL2;3 and modulate rice allelopathy. *Journal of Experimental Botany*, *71*(6), 2127–2141. https://doi.org/10.1093/jxb/erz540

FAO. (2003). *International Rice Commission Newsletter*. *52*. https://www.fao.org/3/Y5167E/Y5167E00.htm

FAO. (2024). Food Outlook – Biannual report on global food markets. In *Food Outlook – Biannual report on global food markets*. FAO. https://doi.org/10.4060/cd1158en

FAO. (2025). *FAOSTAT*. https://www.fao.org/faostat/en/#data/QCL

Fornasiero, A., Wing, R. A., & Ronald, P. (2022). Rice domestication. In *R20 Current Biology* (Vol. 32). http://www.sustainablerice.org/

Fukagawa, N. K., & Ziska, L. H. (2019). Rice: Importance for Global Nutrition. In *J Nutr Sci Vitaminol* (Vol. 65). http://www.riceassociation.org.uk/content/1/18/types-

Fuller, D. Q. (2011). Pathways to Asian civilizations: Tracing the origins and spread of rice and rice cultures. *Rice*, *4*(3–4), 78–92. https://doi.org/10.1007/s12284-011-9078-7

Fuller, D. Q., Sato, Y. I., Castillo, C., Qin, L., Weisskopf, A. R., Kingwell-Banham, E. J., Song, J., Ahn, S. M., & van Etten, J. (2010). Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological and Anthropological Sciences*, *2*(2), 115–131. https://doi.org/10.1007/s12520-010-0035-y

Garris, A. J., Tai, T. H., Coburn, J., Kresovich, S., & McCouch, S. (2005a). Genetic structure and diversity in Oryza sativa L. *Genetics*, *169*(3), 1631–1638. https://doi.org/10.1534/genetics.104.035642

Garris, A. J., Tai, T. H., Coburn, J., Kresovich, S., & McCouch, S. (2005b). Genetic structure and diversity in Oryza sativa L. *Genetics*, *169*(3), 1631–1638. https://doi.org/10.1534/genetics.104.035642

Glaszmann, J. C. (1987). Isozymes and classification of Asian rice varieties. *Theorical and Applied Genetics*, *74*, 21–30.

Goff, S. A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., … Briggs, † Steven. (1999). A Draft Sequence of the Rice Genome (Oryza sativa L. ssp. *japonica*). In *Curr. Opin. Cell Biol* (Vol. 23, Issue 111). www.tmri.

Gupta, P. C., & O'toole, J. C. (1986). *UPLAND RICE A Global Perspective*.

Ho, T. L., Nguyen, T. T. C., Vu, D. C., Nguyen, N. Y., Nguyen, T. T. T., Phong, T. N. H., Nguyen, C. T., Lin, C. H., Lei, Z., Sumner, L. W., & Le, V. V. (2020). Allelopathic potential of rice and identification of published allelochemicals by cloud-based metabolomics platform. *Metabolites*, *10*(6), 1–20. https://doi.org/10.3390/metabo10060244

Hori, K., Nonoue, Y., Ono, N., Shibaya, T., Ebana, K., Matsubara, K., Ogiso-Tanaka, E., Tanabata, T., Sugimoto, K., Taguchi-Shiobara, F., Yonemaru, J. ichi, Mizobuchi, R., Uga, Y., Fukuda, A., Ueda, T., Yamamoto, S. ichi, Yamanouchi, U., Takai, T., Ikka, T., … Yano, M. (2015). Genetic architecture

of variation in heading date among Asian rice accessions. *BMC Plant Biology*, *15*(1). https://doi.org/10.1186/s12870-015-0501-x

Hu, B., Wang, W., Ou, S., Tang, J., Li, H., Che, R., Zhang, Z., Chai, X., Wang, H., Wang, Y., Liang, C., Liu, L., Piao, Z., Deng, Q., Deng, K., Xu, C., Liang, Y., Zhang, L., Li, L., & Chu, C. (2015). Variation in NRT1.1B contributes to nitrate-use divergence between rice subspecies. *Nature Genetics*, *47*(7), 834–838. https://doi.org/10.1038/ng.3337

Huang, X., Kurata, N., Wei, X., Wang, Z. X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., Guo, Y., Lu, Y., Zhou, C., Fan, D., Weng, Q., Zhu, C., Huang, T., Zhang, L., Wang, Y., … Han, B. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature*, *490*(7421), 497–501. https://doi.org/10.1038/nature11532

Ingvarsson, P. K., & Street, N. R. (2011). Association genetics of complex traits in plants. In *New Phytologist* (Vol. 189, Issue 4, pp. 909–922). https://doi.org/10.1111/j.1469-8137.2010.03593.x

Kato-Noguchi, H., & Peters, R. J. (2013). The Role of Momilactones in Rice Allelopathy. *Journal of Chemical Ecology*, *39*(2), 175–185. https://doi.org/10.1007/s10886-013-0236-9

Khush, G. S. (1997). Origin, dispersal, cultivation and variation of rice. In *Ri%gr* (Vol. 35).

Kong, C. H., Li, H. B., Hu, F., Xu, X. H., & Wang, P. (2006). Allelochemicals released by rice roots and residues in soil. *Plant and Soil*, *288*(1–2), 47–56. https://doi.org/10.1007/s11104-006-9033-3

Li, J. Y., Wang, J., & Zeigler, R. S. (2014). The 3,000 rice genomes project: New opportunities and challenges for future rice research. In *GigaScience* (Vol. 3, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/2047-217X-3-8

MAPA. (2023). *Anuario de estadística 2023*. https://cpage.mpr.gob.es/

Matsumoto, T., Wu, J., Kanamori, H., Katayose, Y., Fujisawa, M., Namiki, N., Mizuno, H., Yamamoto, K., Antonio, B. A., Baba, T., Sakata, K., Nagamura, Y., Aoki, H., Arikawa, K., Arita, K., Bito, T., Chiden, Y., Fujitsuka, N., Fukunaka, R., … Burr, B. (2005). The map-based sequence of the rice genome. *Nature*, *436*(7052), 793–800. https://doi.org/10.1038/nature03895

Muthayya, S., Sugimoto, J. D., Montgomery, S., & Maberly, G. F. (2014). An overview of global rice production, supply, trade, and consumption. *Annals of the New York Academy of Sciences*, *1324*(1), 7–14. https://doi.org/10.1111/nyas.12540

Nagaraju, J., Kathirvel, M., Ramesh Kumar, R., Siddiq, E. A., & Hasnain, S. E. (2002). Genetic analysis of traditional and evolved Basmati and non-Basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. *PNAS*, *99*(9), 5836–5841. www.pnas.orgcgidoi10.1073pnas.042099099

OECD/FAO. (2023). *OCDE-FAO Perspectivas Agrícolas 2023-2032*. OECD. https://doi.org/10.1787/2ad6c3ab-es

Olmos, S. (2007). *Apunte de morfología, fenología, ecofisiología y mejoramiento genético del arroz*.

Olofsdotter M. (1998). *Allelopathy in rice*.

Parsons, B. J., Newbury, H. J., Jackson, M. T., & Ford-Lloyd, B. V. (1999). The genetic structure and conservation of aus, aman and boro rices from Bangladesh. In *Genetic Resources and Crop Evolution* (Vol. 46).

Rahaman, F., Shukor Juraimi, A., Rafii, M. Y., Uddin, K., Hassan, L., Chowdhury, A. K., Karim, S. M. R., Yusuf Rini, B., Yusuff, O., Bashar, H. M. K., & Hossain, A. (2022). Allelopathic potential in rice - a biochemical tool for plant defence against weeds. *Frontiers in Plant Science*, *13*(December), 1–14. https://doi.org/10.3389/fpls.2022.1072723

Siddiq, E. A., & Vemireddy, L. R. (2021). Advances in genetics and breeding of rice: An overview. In *Rice Improvement: Physiological, Molecular Breeding and Genetic Perspectives* (pp. 1–29). Springer International Publishing. https://doi.org/10.1007/978-3-030-66530-2_1

Song, B., Xiong, J., Fang, C., Qiu, L., Lin, R., Liang, Y., & Lin, W. (2008). Allelopathic enhancement and differential gene expression in rice under low nitrogen treatment. *Journal of Chemical Ecology*, *34*(5), 688–695. https://doi.org/10.1007/s10886-008-9455-x

Sweeney, M., & McCouch, S. (2007). The complex history of the domestication of rice. In *Annals of Botany* (Vol. 100, Issue 5, pp. 951–957). https://doi.org/10.1093/aob/mcm128

Tang, W., Ye, J., Yao, X., Zhao, P., Xuan, W., Tian, Y., Zhang, Y., Xu, S., An, H., Chen, G., Yu, J., Wu, W., Ge, Y., Liu, X., Li, J., Zhang, H., Zhao, Y., Yang, B., Jiang, X., … Wan, J. (2019). Genome-wide associated study identifies NAC42-activated nitrate transporter conferring high nitrogen use efficiency in rice. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-13187-1

Vaughan, D. A., Morishima, H., & Kadowaki, K. (2003). Diversity in the Oryza genus. In *Current Opinion in Plant Biology* (Vol. 6, Issue 2, pp. 139–146). Elsevier Ltd. https://doi.org/10.1016/S1369-5266(03)00009-8

Wang, C., & Han, B. (2022). Twenty years of rice genomics research: From sequencing and functional genomics to quantitative genomics. In *Molecular Plant* (Vol. 15, Issue 4, pp. 593–619). Cell Press. https://doi.org/10.1016/j.molp.2022.03.009

Yoshida, S. (1981). *Fundamentals of rice crop science*.

Yu, J., Hu, S., Wang, J., Ka-Shu Wong, G., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., … Yang, H. (2002). A Draft Sequence of the Rice Genome (Oryza sativa L. ssp. indica). *Science*, *296*(79). http://btn.

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (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

Yu, J., Xuan, W., Tian, Y., Fan, L., Sun, J., Tang, W., Chen, G., Wang, B., Liu, Y., Wu, W., Liu, X., Jiang, X., Zhou, C., Dai, Z., Xu, D., Wang, C., & Wan, J. (2021). Enhanced OsNLP4-OsNiR cascade confers nitrogen use efficiency by promoting tiller number in rice. *Plant Biotechnology Journal*, *19*(1), 167–176. https://doi.org/10.1111/pbi.13450

Zhang, Q., Maroof, M. A. S., Lil, T. Y., & Shen, B. Z. (1992). Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis. *Theorical and Applied Genetics*, *83*, 495–499.

Zhang, Y., Wang, F., Huang, L.-Y., Li, X.-X., Zhang, Y.-B., Fahad, S., & Fei, W. (2021). dep1 improves rice grain yield and nitrogen use efficiency simultaneously by enhancing nitrogen and dry matter translocation. *Article in Journal of Integrative Agriculture*. https://doi.org/10.1016/S2095-3119(21)63795-4

Objetives

The **overall objective** of this doctoral thesis is to decipher the genetic and molecular basis of agronomically important traits in temperate *japonica* rice (*O. sativa*) with the ultimate goal of contributing to the development of more sustainable and competitive varieties. The research focuses on identifying functional variants and mechanisms associated with allelopathy, to combat the weed *E. crus-galli*, and with nitrogen use efficiency (NUE), to obtain high yields with low fertilizer inputs. This will be achieved using two genetic resources, consisting in a collection of cultivated *japonica* varieties and a MAGIC population, along with multi-omic approaches. This general objective is broken down into the following specific objectives:

1.- To identify the genetic basis of **allelopathic potential** in *japonica* rice cultivated in temperate climates through a genomic association study (**GWAS**) of a diverse **collection** of 171 accessions. This study considers allelopathy as the ability to inhibit the root growth of *E. crus-galli* under controlled conditions and aims to detect functional variants and candidate genes involved in weed growth inhibition.

2.- To investigate the molecular mechanisms involved in **allelopathy** by conducting a transcriptomic analysis (**RNA-seq**) of the roots of two contrasting varieties, Ganao (allelopathic) and Bomba (non-allelopathic), during co-cultivation with *E. crus-galli*, aiming to identify differentially expressed genes, associated functional pathways and potential regulators involved in the allelopathic response.

3.- To characterize a **MAGIC** rice **population** as a resource for the genetic analysis of traits of agronomic interest by sequencing the genome of the lines, analyzing their genetic structure, and evaluating their phenotype related to plant architecture and phenology, with the aim of identifying genomic regions associated with these traits.

4.- To investigate **NUE** and **yield** by conducting field trials under two fertilization levels with the characterized MAGIC population, and identifying associated loci and candidate genes through **GWAS**.

Chapters

Chapter I

Deciphering the Genetic Basis of Allelopathy in *japonica* Rice Cultivated in Temperate Regions Using a Genome-Wide Association Study

The content of this chapter is published in:
Rice 17:22 (2024)
DOI: https://doi.org/10.1186/s12284-024-00701-3
Julia García-Romeral[1], Raúl Castanera[2], Josep Casacuberta[2] and Concha Domingo[1]

1 Departamento del Arroz, Centro de Genómica, Instituto Valenciano de Investigaciones Agrarias (IVIA)
2 Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB

Abstract

Allelopathy has been considered as a natural method of weed control. Despite the nature of allelochemical compounds has been studied, little is known about the genetic basis underlying allelopathy. However, it is known that rice exhibits diverse allelopathic potentials across varieties, and breeding for rice plants exhibiting this potential would be highly desirable as they would confer an advantage against weeds in paddy fields. Knowledge of the gene factors and the identification of the genomic regions responsible for allelopathy would facilitate breeding programs. Taking advantage of the existing genetic diversity in rice, particularly in temperate *japonica* rice, we conducted a comprehensive investigation into the genetic determinants that contribute to rice allelopathy. Employing Genome-Wide Association Study, we identified four Quantitative Trait Loci, with the most promising loci situated on chromosome 2 and 5. Subsequent inspection of the genes located within these QTLs revealed genes associated with the biosynthesis of secondary metabolites such as Phenylalanine Ammonia Lyase (PAL), a key enzyme in the synthesis of phenolic compounds, and two genes coding for R2R3-type MYB transcription factors. The identification of these two QTLs associated to allelopathy in rice provides a useful tool for further exploration and targeted breeding strategies.

Introduction

Rice is a global staple food and high yields are needed to sustain a large population. It is also important to facilitate a sustainable crop and to increase the incomes of farmers and the economic profitability of the crop. Weeds, in their competition for nutrients, moisture, and light, diminish rice yields. The management of weed control involves the application of herbicides, which is environmentally unsustainable, and requires a high amount of manual labor, which is costly and profitless. The quest for an economically and environmentally sustainable approach to weed control remains a challenge. In this context, allelopathy can be considered as an environmentally friendly practice for weed control. Allelopathy has been described as the ability of an organism to affect growth, survival, and reproduction of other organisms through the secretion of allelochemicals or their release into the environment (Rice 1984). Allelopathy has been observed in numerous plant species, and its potential impact is well-documented. In fact, the efficiency of using straw and hulls from certain crops, including rice, as mulch for weed control, has been successfully demonstrated (Khamare et al. 2022). Among others, barnyardgrass (*Echinochloa crus-galli*) is a globally extended harmful weed, frequently found in paddy fields, that produces field losses (Oerke and Dehne 2004) and, worriedly, the appearance of resistance to herbicides has been frequently observed (Amaro-Blanco et al. 2021).

In rice, allelopathic effects have been described in certain varieties (Dilday et al. 1994). Some varieties can synthesize and release chemical compounds into the environment that can affect the growth and development of neighbouring plants by interfering with their physiological processes, such as seed germination, root elongation, and nutrient uptake. Their inhibitory effect on weeds, such as ducksalad (*Heteranthera limosa*) or barnyardgrass, has been reported (Dilday et al. 1994). In the late 1980s, hundreds of accessions from germplasm collections were examined in field experiments for their allelopathic potential towards ducksalad, redstem (*Ammannia coccinea* Rottb.), lettuce (*Lactuca sativa* L.) and barnyard grass among others (Dilday et al. 1994; Hassan et al. 1998). More recently, bioassays were developed to reduce environmental factors, and hundreds of accessions were also screened in the laboratory (Kohli et al. 2008). As a result, several accessions with strong allelopathic potential have been identified, as is the case of PI213777 (Dilday et al. 1994). Several allelochemicals have been isolated from plants and root

exudates of different plant species and many studies have been performed to find the nature and the effect of such compounds in the rice with barnyardgrass interaction (Macías et al. 2007). Most reported allelochemicals in rice are secondary plant metabolites, including fatty acids, indoles, momilactones, phenolics and terpenes (Seal et al. 2004; Khanh et al. 2007; Kato-Noguchi and Peters 2013). Transcriptomic studies have also been conducted in the interaction of rice with barnyardgrass and differential expressed genes associated with momilactone and phenolic acid biosynthesis were found to be up-regulated in a quick allelopathic response (Song et al. 2008; Sultana et al. 2023). But no candidate genes have already been reported.

Secondary metabolites are mainly derived from the phenylpropanoid metabolic and isoprenoid pathways and their associated branches (Fang et al. 2020). Phenylalanine ammonia lyase (PAL) is an initial key enzyme in the biosynthesis of phenolic compounds, catalysing the conversion of phenylalanine to cinnamic acid, which is a precursor for the synthesis of allelopathic phenolic compounds such as ferulic acid and p-coumaric acid. The role of PAL in the regulation of chemical induction in allelopathy has been reported previously and the up-regulated expression of PAL is associated with the enhanced inhibitory effect of allelopathy grown with barnyard (Zhang et al. 2019b).

Momilactones A and B were first isolated from rice husk, as growth inhibitors (Takahashi et al. 1976) and more recently they have been found in root exudates of allelopathic cultivars as PI312777 and Koshihikari (Kong et al. 2004; Kato-Noguchi et al. 2008). Momilactones have been considered phytoalexins in rice protection against fungi (Cartwright et al. 1977; Cartwright et al. 1981). Later on, their ability to inhibit weed growth was demonstrated (Kato-Noguchi et al. 2010). The biosynthetic pathway of momilactones is well known (Kato-Noguchi and Peters 2013), and knock-outs of relevant genes participating in the diterpene synthesis pathway such as copalyl diphosphate synthase 4 (OsCPS4) or kaurene synthase-like 4 (OsKSL4) result in reduced allelopathic activity of rice in barnyard (Xu et al. 2012). Two gene clusters have been identified in the synthesis of momilactones (Xu et al. 2012; Kato-Noguchi and Peters 2013).

The role of phenolic compounds in allelopathy has been questioned on several occasions. The main concern is that the levels of phytotoxic compounds in the rice soil are not sufficient to cause phytotoxicity (Olofsdotter et al. 2002). In a previous

study, the action of phenolic acids in rice roots exudates was investigated individually or combined finding that, despite rice showed more phytotoxic tolerance to these compounds that arrowhead (*Sagittaria monotevidensis* Cma. & Schltdl), the concentration required for growth inhibition was much higher than the levels found in root exudates (Olofsdotter et al. 2002; Seal et al. 2004). But their participation in allelopathy is still unclear and it has been suggested that phenolic acids might act synergistically with other chemicals to function as allelochemicals. On the contrary, momilactone B has been proposed to be a major allelochemical in rice (Kato-Noguchi et al. 2010).

Previous studies indicated that rice allelopathy is a quantitative genetic trait, which is influenced by environmental conditions. The genetic control of allelopathy has been investigated previously in biparental populations. Quantitative trait loci (QTLs) analysis in a population of PI312777 and Rexmont, using a 215 RFLP markers, revealed seven loci associated with inhibition of the root length of lettuce plants distributed in chromosomes 1, 3, 5, 6, 7, 11, and 12 (Ebana et al. 2001). Additionally, four main QTLs associated to allelopathic activity against barnyardgrass under laboratory conditions were mapped, using 140 DNA markers, in chromosomes 2, 3 and 8 in a population derived from a cross between cultivar IAC 165 (*japonica* variety) and cultivar CO-39 (*indica* variety) (Jensen et al. 2001). In the same way, two QTLS were identified in chromosomes 4 and 7 in a biparental population derived from AC1423, an *indica* variety with high allelopathic potential and Aus 196, an Aus variety with low allelopathic potential (Jensen et al. 2008). These studies were carried out before the development of the new-generation sequencing techniques. The availability of whole genome sequencing and its low cost, allows the development of high-density SNP-type marker panels that provide much greater precision in assays than previous ones. In addition, GWAS in a wide diverse population has become an extensive mapping tool to dissect complex agronomic traits. Biparental populations limit the analysis to a relatively low number of genes and, in contrast, GWAS enables to search in a broader gene pool as it is conducted using a wide diversity. Till now, GWAS has not been conducted in rice allelopathy.

Most of the genetic and molecular analyses of the allelopathic potential have been carried out on line PI312777, with strong weed suppressing ability. It is accepted that momilactones (A and B) and phenolics are responsible for the allelopathic potential of this line (Kong et al. 2004, 2006; He et al. 2012). Despite the ability of

rice plants to produce diterpenoids that are released into the rhizosphere was recognized decades ago, as well as their function as phytoalexins, the allelopathic activity of these compounds as a natural strategy to prevent weed growth remains relatively unexplored. Particularly there is limited information about allelopathic potential in modern and European varieties and it is not known whether other mechanisms responsible for the allelopathic effect remain to be discovered in other varieties.

In this sense, the search for cultivars adapted to European agroclimatic conditions with the potential to inhibit barnyardgrass, and the elucidation of genes responsible for allelopathy in temperate *japonica* rice remain to be done. In this manuscript, we have investigated allelopathy in a temperate *japonica* rice panel, conducted a GWAS using a high-density panel of SNPs, and assessed the feasibility of initiating a breeding program to incorporate allelopathy in rice while maintaining grain yield and quality, which will ultimately benefit both farmers and the environment.

Material and Methods

Plant Material

A set of 171 accessions showing a wide genetic diversity was generated by selecting a set of 157 accessions from a previously generated rice collection (*Oryza sativa*) (Reig-Valiente et al. 2018) and adding 14 recently released accessions. The set comprises 163 *japonica* and 8 *indica* type accessions.

Barnyard grass seeds were collected from two plants from experimental fields at Tancat de Malta (39° 18' N 0° 20' W, Valencia, Spain) and were identified as *Echinochloa crus-galli* and *Echinochloa hispidula*.

Phenotypic Bioassay

The allelopathic potential of 171 accessions was evaluated using a modified relay seeding technique (Navarez and Olofsdotter 1996). The bioassay was performed in separate batches and replicated twice.

On the first day (**Figure 1**, day 1) 12 rice seeds of each accession were sterilized by immersion in a 2.5% sodium hypochlorite solution for 30 min, followed by rinsing with distilled water thrice. The seeds were placed equidistantly in a 9 cm diameter Petri dish containing two filter papers at the base, a thin layer of perlite at top, watered with 20 ml of MES 1 mM at pH 6.0, and the lid was closed. The seeds were germinated and cultured in a growth chamber with a constant temperature of 28 °C in darkness for three days. On the fourth day (**Figure 1**, day 4) the lids were removed and batches comprising 16 accessions were placed in a plastic box shielded with a transparent plastic cover, to avoid evaporation. Plants were incubated in a growth chamber at 26 °C under a 16 h/8 h (day/night) photoperiod and the plates were watered each day with 3 ml of MES 1 mM pH 6.0. On the same day, sterilized barnyardgrass seeds were germinated in Petri dishes and incubated in a growth chamber at 28 °C under dark conditions for three days. On the seventh day (**Figure 1**, day 7), 9 barnyardgrass seedlings showing root length between 6 and 8 mm were equidistantly placed in small holes in the perlite layer using a forceps in order to keep the roots of the barnyardgrass submerged, avoiding the growth above the surface. As

a control, 20 barnyardgrass seedlings were transplanted into an additional Petri dish with perlite, in the absence of rice. All plates were incubated for an additional week.

On the fifteenth day (**Figure 1**, day 15), rice and barnyardgrass (from co-cultured and control plates) root length was measured with a vernier caliper after washing off the perlite. The percentage of inhibition of barnyardgrass root growth was recorded as the difference in root length between barnyardgrass growing in the presence or absence of rice, as shown:

$$\%\ of\ inhibition\ = \frac{RLB\_control\ (mm) - RLB\_\ cocultured\ (mm)}{RLB\_control\ (mm)} \times\ 100$$

Where RLB_control is the root length of the barnyardagrass under control conditions (absence of rice) and RLB_cocultured is the root length of the barnyardgrass under co-cultivation conditions.

Data filtering, processing, and plotting of phenotype data were conducted by using R statistical software (ver. 4.2.0), employing the packages stats and ggplot2.

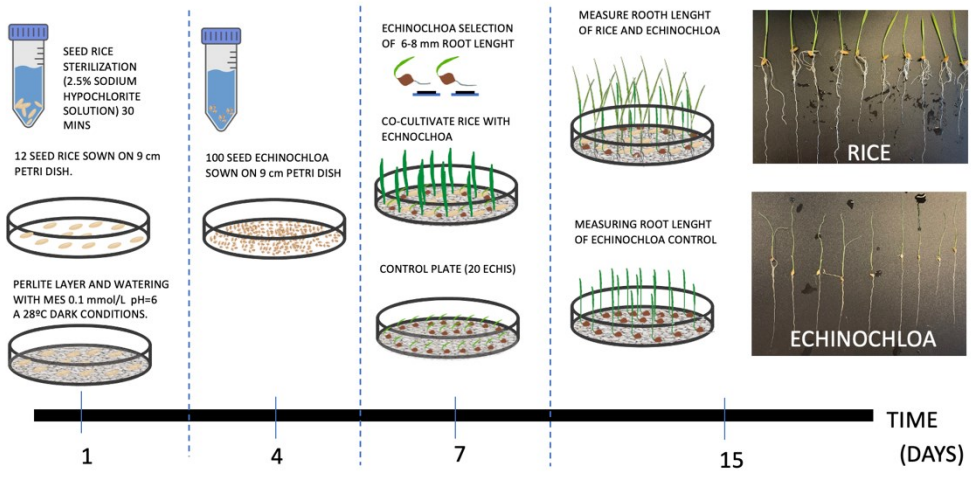


Figure 1. Timeline of the rice–barnyardgrass co-culture bioassay to quantify the percentage of root growth inhibition of *Echinochloa* induced by each of the rice varieties evaluated.

Whole Genome Sequencing and Data Analysis

DNA was extracted from rice fresh leaf tissue using the cetyltrimethylammonium bromide (CTAB) method with slight modifications, as previously described (Murray and Thompson 1980). DNA was quantified using a NanoDrop 1000 Spectrophotometer (Thermo Scientific, Waltham, MA, USA) and Qubit™ dsDNA Quantification Assay Kit using the Qubit 2.0 Fluorometer (Life Technologies, EEUU). Quality was checked by electrophoresis. Whole Genome Sequencing was performed at Novogene UK Company Limited (Cambridge, United Kingdom).

Genomic DNA was randomly sheared into short fragments that were end-repaired, adenylated, and further ligated to Illumina specific indexed paired-end adaptors. The fragments with adapters were PCR amplified, size selected and purified. The final DNA libraries were quantified using Qubit and real-time PCR and a bioanalyzer was used for size distribution detection. The libraries were sequenced using an Illumina NovaSeq 6000 system PE150, according to a 350 bp library type and to standard Illumina operation procedures with a yield of > 8 Gb per lane and median phred quality score of Q36. BBDuck function, from BBMap software (https:// sourceforge. net/ proje cts/ bbmap/) was used for adaptor and quality trimming (Forward: 5′-AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AGA TCT CGG TGG TCG CCG TAT CATT-3′; Reverse: 5′-GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC GGA TGA CTA TCT CGT ATG CCG TCT TCT GCTTG-3′).

Clean reads were aligned to the Nipponbare reference genome (IRGSP-1.0) using BWA-MEM (Li 2013). The Samtools suite (Danecek et al. 2021) was used for downstream processing of alignment files. Specifically, “samtools flagstat” was used to obtain the mapping statistics from each alignment and “samtools view” was deployed to remove unmapped reads and very low-quality alignments (Q < 10).

Multi-sample variant calling was performed with BCFTOOLS “mpileup” and “call” functions to obtain SNPs and INDELs. For each variant, we retained its allelic depth (AD), number of high-quality bases (DP), phred-scaled strand bias P-value (SP), phred-scaled genotype quality (GQ) and posterior genotype probability in Phred scale (GP). BCFTOOLS “view” was used to transform from the bcf format file to a gvcf format and “norm” function for outputting only the first instance when a record is present in multiple lines.

The analysis yielded 351.4 million variants, that were filtered with vcftools using the parameters: --remove-indels, --maf 0.05, --max-missing 0.9, --minQ 30, --minDP 10, --maxDP 30, --max-meanDP 30, --min-meanDP 10, and only 1.0 million SNPs remained.

Additionally, the SNP data matrix was pruned with PLINK software (Purcell et al. 2007) for high levels of pairwise linkage disequilibrium (LD) using "--indep-pairwise 50 5 0.9" (window of 50 SNPs with a shift of 5 SNPs and 0.9 R2 cutoff). At this step multiallelic SNPs were also removed. Finally, a 124.019 SNP data array was obtained for the 171 varieties.

Linkage Disequilibrium (LD) Estimation

The linkage disequilibrium (LD) decay rate was calculated using TASSEL v.5 software (Bradbury et al. 2007) with a sliding window of 50 SNPs, as the chromosomal distance where the Pearson's correlation coefficient (r^2) between SNP pairs dropped to half its maximum estimated value. Data Plotting was performed with R software v.4.3.0 (R Core Team 2023).

Principal Component Analysis (PCA)

Principal Component Analysis was performed using "SNPRelate" package (Zheng et al. 2012) from R software. The plot was drawn with the "rgl" package (Murdoch and Adler 2023).

Population Structure Analysis

The population genetic structure of the collection was assessed using the Bayesian clustering method implemented in STRUCTURE 2.3.4 (Pritchard et al. 2000). For that aim a selection of 18,728 coding SNPs from the 171 varieties was used. To obtain the pool, we intersected the 124,019 filtered SNPs with coding regions (CDS) of the 21,418 IRGSP Nipponbare genes conserved in the whole Rice Gene Index (RGI) collection (core genes) (Yu et al. 2023). The RGI collection encompasses 16 platinum standard reference genomes of rice (Zhou et al. 2020). The 18,728 SNP

dataset was generated through this intersection process using the bedtools intersect function.

The software estimated the optimal number of genetic clusters (K) and calculated each variety's membership proportion. Analyses were based on the admixture ancestral model for a range of K values from 1 to 7. We performed 10 runs for each K. Each run was implemented with a burn-in period of 10,000 steps followed by 100,000 Monte Carlo Markov Chain iterations. The optimal number of K clusters was estimated with the ad hoc parameter (ΔK) of Evanno (Evanno et al. 2005) in Structure Harvester (Earl and VonHoldt 2012).

Phylogenetic analysis was performed by converting the 124,019 SNP dataset in VCF format into a PHYLIP file using the vcf2phylip.py script (Ortiz 2019). To calculate pairwise genetic distances between rice cultivars a maximum likelihood (ML) method was applied by IQ-TREE v.2.1.2. (Minh et al. 2020) with 1000 bootstrap replicates and the GTR + ASC model, specific for SNP data. Data representation in tree format was displayed with iTOL v.6.8.1. (Letunic and Bork 2021).

Marker Trait Association Analysis

The relationship between phenotype and the panel of 124,019 SNPs was examined through GWAS using TASSEL v.5 software (Bradbury et al. 2007). A kinship matrix (K) was created from the genotype data utilizing the Centered IBS (Identity-By-State) method, which calculates the probability that alleles randomly drawn from two individuals at the same locus are identical. Loci may consist of one or more nucleotides. In this approach, genotypes are encoded as 2, 1, or 0, corresponding to the count of one of the alleles at that particular locus. As the mean of the percentage of missing data was very low after filtering, 0.08%, then the missing genotype values were replaced with the average genotypic score at that locus before estimating a relationship matrix.

In addition, a distance matrix (Q) of the genotype data was calculated as 1—IBS (Identity-By-State) similarity. From this matrix, a Multidimensional Scaling (MDS) analysis, also known as Principal Co-ordinate Analysis (PcoA), was conducted. The axes produced by MDS were used as covariates to correct the population structure.

In addition, batches due to differences in the temporal analysis of the phenotypic analysis were used as covariates.

The MDS results, in conjunction with the genotype (VCF file) and phenotype data (percentage of inhibition), were integrated, and the kinship matrix was incorporated to perform the phenotype-genotype association analysis. For comparison, two statistical models, Mixed Linear Model (MLM) and General Linear Model (GLM), were employed. In the case of MLM, K and Q matrix were used as kinship and population structure controls respectively. Given that MLM offers an effective control of false positives but with potential occasional false negatives, we opted to stablish a threshold value calculated using a non-stringent method, the minimum Bayesian Factor (mBF) (Goodman 2001; Wakefield 2009; Zhang et al. 2019a). The threshold P-value for significant association was set at $P \leq 2.57 \times 10{-}4$, corresponding to a $-\log10(P) = 3.59$ calculated using the following formula: mBF = −e*P*ln(P).

For GLM, the association analysis was performed using a least squares fixed effects linear model. As GLM may cause potential false positives, the Bonferroni method was used to determine the threshold, given its strength. The threshold for significance was set at $P \leq 8.1 \times 10{-}6$, corresponding to $-\log10(P) = 5.09$, calculated by the rough Bonferroni correction.

SNPs situated within a region with a distance similar to the LD (considering the calculated LD decay distance of 157 kb) were considered as a single QTL. SNPs displaying the lowest p-value in a QTL were regarded as the leading SNPs. The Manhattan plot was drawn using the R package”qqman” (Turner 2018). Phenotype effect of the leading SNPs was plotted with “ggstatsplot” package (Patil 2021) from the same software.

Results

Allelopathic Potential Assay

The ability of inhibition of barnyardgrass root growth was evaluated in plants from a panel of 163 *japonica* varieties adapted to temperate climate (**Figure 2**; **Table 1**). The collection comprised modern and old varieties as well as some landraces from different geographical origins to cover a wide genetic diversity. The European accessions were the most abundant group, with 105 from four different countries. A set of 8 *indica* cultivars was also included in the collection as a reference for genetic divergence.

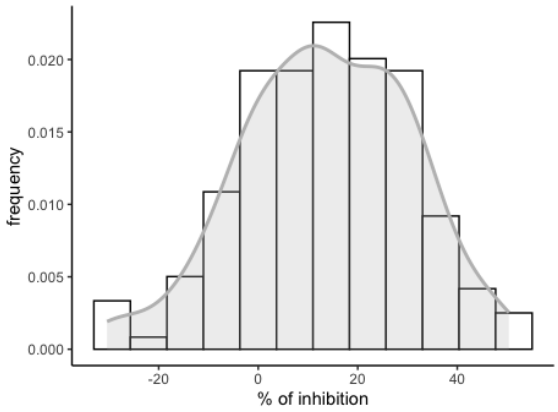


Figure 2. Distribution of barnyardgrass root growth inhibition frequencies. Results obtained from the allelopathic potential bioassay evaluated under coculture conditions with 171 different commercial rice varieties.

Number	Accesion	Origin	% of Inhibition	Number	Accesion	Origin	% of Inhibition
1	Adriano	France	5.79	87	Copsemar-7	Spain	23.57
2	Agami	Egypt	0.00	88	Olesa	Spain	5.41
3	Akihikari	Korea	17.70	90	Albada	Spain	12.82
4	Albufera	Spain	-0.37	91	Betis	Spain	15.23
5	Apolo	Italy	0.00	92	Clot	Spain	-10.17
6	Arpa	Italy	14.76	93	Jucar	Spain	34.20
7	Arroz da Terra	Portugal	46.24	94	Mareny	Spain	31.55

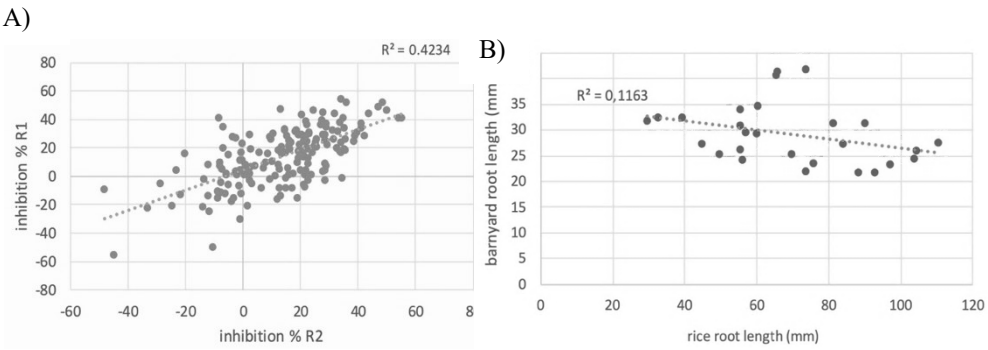
8	Arsenal	Italy	1.70	95	Niva	Spain	25.82
9	Bahia	Spain	6.30	96	Balilla	Italy	18.68
10	Baixet	Spain	5.13	97	Colina	Spain	-12.08
11	Bendret	Spain	-1.07	98	Bombón	Spain	-28.91
12	Benisants	Spain	16.32	99	Liso	Spain	-26.23
13	Bomba	Spain	4.77	100	Sollana	Spain	-6.68
14	Calmochi-101	USA	10.55	101	Dosel	Spain	25.05
15	Campino	Portugal	-4.68	102	Clavel	Spain	-16.84
16	Carmen	Spain	-4.54	103	Susi	Spain	-9.17
17	Carnice precoce	Italy	16.95	104	Montsianell	Spain	4.62
18	Cesare	Italy	10.82	105	Carnaroli	Italy	31.31
19	CNA-6159	Brasil	18.41	106	Icaro	Italy	3.60
20	Cormorán	Spain	0.72	107	Deneb	Italy	-12.80
21	CT-13432	Colombia	23.42	108	Fragrance	Italy	25.59
22	CT-18	Colombia	27.25	109	Giano	Italy	18.05
23	Drago	Italy	18.60	110	Poseidone	Italy	11.37
24	Ducato	Italy	-2.45	111	Genio	Italy	18.93
25	Eridano	Italy	20.13	112	Maratelli	Italy	38.72
26	Fado	Spain	0.00	113	Lido	Italy	35.62
27	Fonsa	Spain	20.90	114	Antares	Italy	44.49
28	Gavina	Spain	41.90	115	Cerere	Italy	3.33
29	Gema	Puerto Rico	29.08	116	Giglio	Italy	-0.31
30	Gigante Vercelli	Italy	35.88	117	Millin	Australia	-0.52
31	Giovanni Marchetti	Italy	25.95	118	Langhi	Australia	26.38
32	Gladio	Italy	12.74	119	Kyeema	Australia	25.59
33	Gleva	Spain	-0.96	120	Illabong	Australia	30.08
34	Guadiamar	Spain	-8.56	121	Amaroo	Australia	16.05
35	Habataki	Japan	12.44	122	M-101	USA	11.59
36	Harra	Australia	30.42	123	M-103	USA	23.74
37	Hidalgo	Spain	0.83	124	M-203	USA	20.49
38	Hwayoung	Korea	5.10	125	M-205	USA	11.45
39	IR-58	Philippines	15.64	126	M-206	USA	31.90
40	Irat 289	French Guiana	14.12	127	M-208	USA	23.67
41	Irat 307	French Guiana	13.28	128	L-203	USA	1.21
42	Irat 320	French Guiana	10.13	129	L-206	USA	16.50
43	Irat 350	French Guiana	9.08	130	Calmati-202	USA	12.52
44	Jsendra	Spain	-2.00	131	S-101	USA	16.17
45	Kalao	France	12.51	132	Newbonnet	USA	23.59
46	Kinmaze	Japan	-17.40	133	Bluebonnet	USA	25.32
47	Koshihikari	Japan	-30.37	134	Calbelle	USA	32.43
48	L-202	USA	-2.12	135	Tebonnet	USA	31.04
49	Leda	Spain	10.24	136	Labelle	USA	30.99
50	Lemont	USA	8.02	137	Maybelle	USA	35.56
51	Litchikiang	China	12.81	138	Mars	USA	29.41

52	Italica Livorno	Italy	6.41	139	Northrose	USA	44.03
53	Loto	Italy	-7.01	140	Nova 66	USA	32.76
54	M-201	USA	8.81	141	Katy	USA	48.40
55	M-202	USA	12.19	142	A-301	USA	39.03
56	Marisma	Spain	9.08	143	Ganao	France	50.37
57	Marjal	Spain	9.67	144	Adret	France	20.35
58	Moroberekan	Guinea	4.56	145	Matusaska	Japan	-9.57
59	Nano	Italy	-9.20	146	YR 71011	Australia	-0.24
60	Nipponbare	Japan	-50.45	147	Capataz	Spain	-0.99
62	Nuevo Maratelli	Italy	25.56	148	Frances	Spain	-8.15
63	Opale	Italy	26.74	149	Fuzisaka	Japan	11.38
64	Pavlovsky	Russia	32.04	150	Mangetsumochi	Japan	-15.65
65	Pegonil	Spain	-5.59	151	Precoz-6	Italy	27.20
66	PUITA	Argentina	19.16	152	Sigikari	Japan	7.80
67	Puntal	Spain	23.13	153	Centa A-6	El Salvador	-4.11
68	Ricastello	Spain	38.12	154	Lady Wright	USA	40.83
69	Rivera	Spain	20.58	155	Daekwanbyeo	Korea	13.14
70	Romolo	Italy	32.94	156	Giza 181	Egypt	20.96
71	Sarcet	Spain	34.83	157	Giza 178	Egypt	-2.05
72	Savio	Italy	27.00	158	Nam Yang	Korea	21.76
73	Senia	Spain	47.74	159	Alena	Spain	14.62
74	Sequial	Spain	28.50	160	Copsemar-9	Spain	25.04
75	SIS R215	Italy	24.85	161	Estany	Spain	27.12
76	Sivert	Spain	5.69	162	Gageron	France	33.45
77	Tainung 67	Taiwan	-2.14	163	Gines	France	24.87
78	Ullal	Spain	8.92	164	Guadiagran	Spain	21.03
79	Urano	Italy	8.97	165	Guara	Spain	26.68
80	Venere	Italy	23.03	166	Hispagran	Spain	16.56
81	Zhonghua 15	China	2.12	167	Hispalong	Spain	11.11
82	Benlloch	Spain	-17.56	168	Hispamar	Spain	11.36
83	Colusa	Spain	-22.47	170	Soto	Spain	35.30
84	Insen x Tremesino	Spain	-2.19	171	Teti	Portugal	17.31
85	Tebre	Spain	6.44	172	Regina	Spain	34.28
86	Argila	Spain	-8.76	173	Muga	Portugal	2.15

Table 1. List of varieties included in the bioassay, their origin and percentage of inhibition of *Echinochloa* roots in the presence of each one.

Rice seedlings were incubated in the presence of barnyardgrass plants under controlled laboratory conditions for 7 days and then, the length of barnyardgrass roots was measured. We found marked differences in the inhibitory activity among the varieties, with a coefficient of variation of 130.5% indicating the suitability of this collection for the association study. The assay was performed in duplicate and the replicability was evaluated showing a correlation of 0.42 between both replicates (**Figure 3A**). Most of the cultivars inhibited the growth of barnyardgrass, among

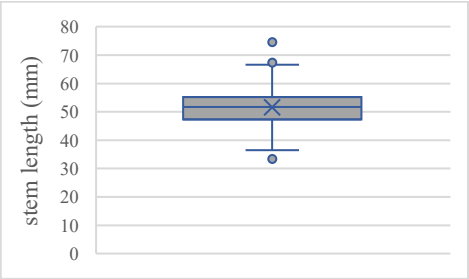
them, 8 varieties suppressed barnyardgrass root growth in 40–50%, representing a 4.7% of the collection. The highest allelopathic potential was exhibited by Ganao, Katy and Senia, inhibiting the barnyardgrass root growth 50.4%, 48.4% and 47.7% respectively compared to plants grown without the presence of rice plants (**Table 1**). Some varieties showed no inhibition or even stimulation of barnyardgrass root growth (**Table 1**). This is the case of Koshihikari, a known variety to exudate momilactone (Kato-Noguchi et al. 2008), that didn't promote barnyard grass growth stimulation in our conditions. Distribution of root growth inhibition frequencies is shown in **Figure 1**.



In addition, the rice root length was measured to find out if the co-cultured could

Figure 3. A) Correlation between root inhibition observed between replicates; B) Correlation between the length of the rice root and *Echinochloa*.

affect the rice root growth. We didn't find any correlation between the length of barnyardgrass and rice root length, indicating that the inhibition wasn't caused by space competition (**Figure 3B**). Barnyardgrass stem length was also recorded, and no differences in growth were found when cultured in the presence or absence of rice (**Figure 4**).



Count	181
Average	51.56
Standard Deviation	6.82
Coefficient of Variation	13.24%
Minimum	33.41
Maximum	74.59
Range	41.18
Standardised Bias	1.82
Standardised Kurtosis	2.28

Figure 4. Stem length of *Echinochloa* plants co-cultivated with different rice varieties. Statistical analysis is shown.

Genome Sequencing and Identification of Polymorphisms

Genome resequencing of the varieties generated a mean of 51 × 106 short reads per cultivar which correspond to 9.53 GB per sample with a mean coverage of 20×. The clean reads were mapped to the Nipponbare reference genome (IRGSP-1.0) and 98% of them were mapped and properly paired. We detected 351 million raw variants in total, which were filtered according to different criteria as call rate > 90% and read coverage between 10× and 30× and quality (see methods). Indels were removed as well as SNPs with a minimum allele frequency (MAF) below 5% were removed. The extent of linkage disequilibrium (LD) in the population was estimated in 157 kbp, when r^2 drops to 0.23 (**Figure 5**). Finally, after SNP pruning by LD, a panel of 124,019 SNPs was obtained, with an average value of 10,334.9 markers per chromosome. Assuming from the newest genome assembly a size of 373 Mpb (Kawahara et al. 2013), the average density was 3.1 kbp/marker, ranging from 2.2 for chromosome 11 to 4.1 for chromosome 4.

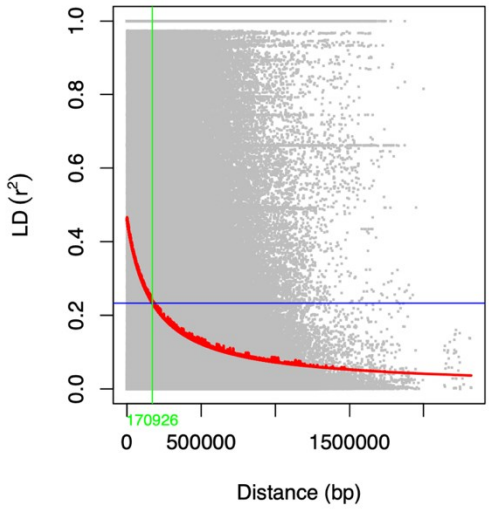


Figure 5. Decrease in linkage disequilibrium (LD) in the collection, represented as the decay of r^2 values as a function of the physical distance between pairs of SNPs. The red curve indicates the non-linear fit of the decay of r^2. The blue line represents the threshold of $r^2 = 0.23$. The green line marks the physical distance (170,926 bp) at which r^2 falls below this threshold.

Genetic Structure of the Collection

Population structure of the collection was estimated using STRUCTURE software. According to this analysis, ΔK showed maximum values for K = 4 (ΔK=1484.25) indicating that the optimum number of subpopulations was 4 (**Figure 6**). Differences in these four subpopulations were corroborated by the Fst values obtained for each group (**Figure 6**). The collection showed a strong structure with most varieties belonging to two main genetic groups (**Fig 7A**). While subgroup 1 had limited weight, subgroup 2 included mostly medium grain type cultivars with different geographical origins from Europe, Asia, America and Australia. Subgroup 3 comprises a small group of *indica* accessions included in the collection. Finally, long grain type accessions conformed subgroup 4 and displayed also different geographical origins, mainly from Europe and America. To deepen the patterns of population structure, we performed principal components analysis (PCA) that supported two main groups (**Figure 7B**). The first principal component (PC1) explained the 15.43% of the variance, and the second component (PC2) the 10.97%

while the third component (PC3) explained the 4.11% of the genetic variance. These two main PCs separated varieties in accordance with the four subpopulations displayed using STRUCTURE.

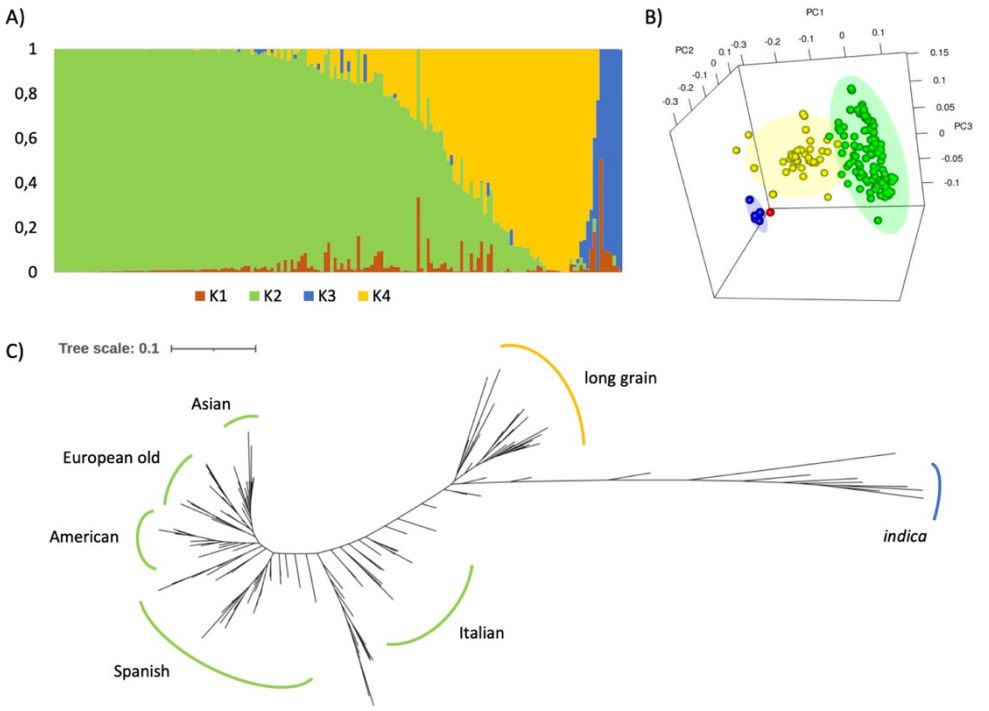


Figure 7. Population structure of the collection based on genotype data. **A** The estimated membership probability of assigning cultivars to 4 groups. Bar length represents the probability of each variety belonging to different subgroups. **B.** 3D Principal Component Analysis (PCA) plots. Three Principal Components are represented as PC1, PC2 and PC3 respectively. **C**. Unrooted phylogenetic tree inferred by the maximum likelihood (ML) method. For all nodes, bootstrap percentage was 100. Length branches show the genetic divergence between varieties

The relationship of accessions in the collection and the genetic distances among them were estimated. The distribution of cultivars obtained was roughly in accordance with their origin and grain type in agreement with the population genetic analysis and with previous results (Reig-Valiente et al. 2016). The *indica* accessions, included in subgroup 3 in the population structure analysis, were grouped into a highly supported cluster, and closely, *japonica* long grain type varieties, included in subgroup 4, diverge into an additional cluster (**Figure 7; Figure 8**). Most of the

japonica varieties displaying medium grain, included in subgroup 2, were grouped in 5 different clusters with a distribution that clearly matches their origin and, even from some occasions, the release date of the cultivars. Asian and American/Australian varieties were clustered in two different clades, while European accessions were fragmented into several small subgroups that included, mostly, European ancient varieties, Italian and Spanish varieties.

Association Analysis

Both Mixed Linear Model (MLM) and General Linear Model (GLM) were used to detect associations between SNP markers and variations in the inhibitory potential of barnyardgrass root growth. It has been suggested that MLM is a stringent method that provides fewer spurious associations than other methods but in contrast increases the number of false negatives (Kaler et al. 2020). Therefore, we used GLM as an additional approach for comparison and verification of the results. P-values and q-values were calculated. In the case of MLM, we set a p-value < 10−4 threshold, according to the minimum Bayesian Factor, to consider a SNP as significantly associated with the allelopathic potential. In the case of GLM, as it is a low restrictive method, we set a threshold of p-value < 10−5, calculated by the Bonferroni correction test. The q-values were also estimated. The quantile–quantile (QQ) plot and Manhattan plot for the analysed trait obtained using the PCA Qmatrix is shown in Figure 7. QQ plot indicated that the model was well fitted to the data; the observed p-values were uniformly distributed with some deviation at high values from the expected p-values (**Figure 9**).

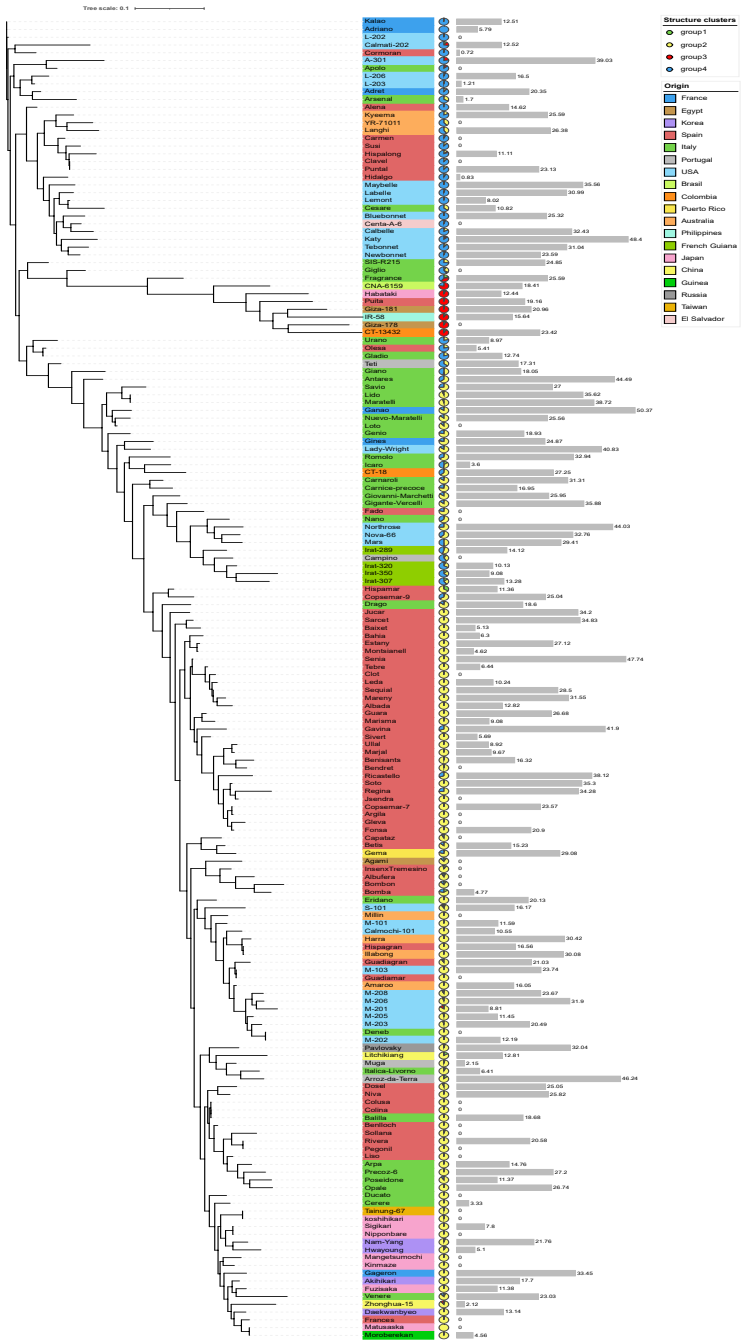


Figure 8. Neighbour-Joining tree of 171 rice accessions. Growth inhibition capacity is represented by grey bars. Membership of genetic structure groups is indicated by circles for each variety, with colours reflecting the proportion of assignment to each group. The country of origin of each accession is also shown.

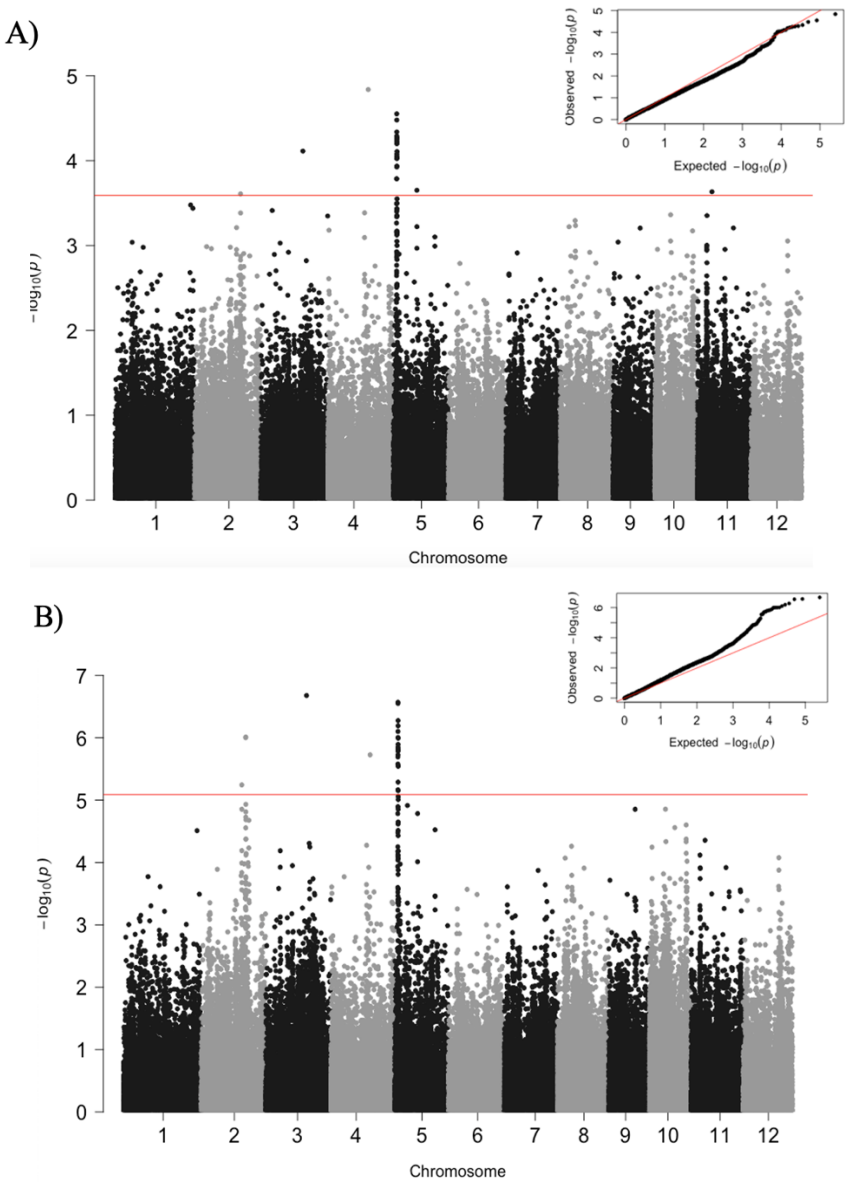


Figure 9. Genome-wide association mapping of the barnyardgrass root growth inhibition. Manhattan and quantile–quantile plots are shown using (A) MLM and (B) GLM. The red horizontal lines indicate genome-wide significant thresholds

The statistical analysis revealed several SNPs significantly associated with barnyardgrass root growth inhibition (**Table 2**; **Table 3**), which corresponded to 4 QTLs, considering an association locus as a chromosomal region in which the distance between the adjacent pairs of associated SNPs is given by the estimated LD. qAll-2a and qAll-2b, only detected using GLM, were located in chromosome 2 at positions 22,671,912 and 24,842,411, respectively. qAll-2b consisted in three SNPs located in a region of 64.1 Kb. qAll-3 was located in chromosome 3 at position 22,755,757. qAll-5, in chromosome 5, comprises a region of 141.1 Kb or 194.6 Kb depending on the applied methodology displaying a peak value at position 1,843,023, given by the lowest p-value (**Table 2**). An additional single SNP above the threshold using both methods was found on chromosome 4 at position 21,850,604, which was excluded because there were no significant SNPs located nearby and to the undetermined allele content in the accessions showing differential root growth inhibition (**Figure 10E**).

QTL	Chr	Lead SNP position	MLM			GLM		
			p-value	Sign. SNPs	Genetic Interval	p-value	Sign SNPs	Genetic Interval
qAll-2a	2	22,671,912				5.69E-06	1	
qAll-2b	2	24,870,799				9.71E-07	3	64.115
qAll-3	3	22,755,757	7.71E-05	1		2.10E-07	1	
qAll-5	5	1,843,023	2.80E-05	16	141.090	2.69E-07	20	194.608

Table 2. List of lead-SNPs significantly associated with the barnyardgrass root growth inhibition (MLM, p-value < 10^{-4}; GLM, p-value < 10^{-5}).

Chr	Position	F marker	p-value	QTL
2	22,671,912	13.10	5.69E-06	qAll-1
2	24,842,411	26.02	9.94E-07	qAll-2
2	24,870,799	15.24	9.71E-07	qAll-2
2	24,906,566		1.17E-05	qAll-2
3	22,755,757	29.60	2.10E-07	qAll-3
4	21,850,604	14.41	1.88E-06	
5	1,786,831	13.98	2.64E-06	qAll-4
5	1,796,832	24.39	2.03E-06	qAll-4
5	1,840,349	29.02	2.81E-07	qAll-4

5	1,843,015	26.58	7.91E-07	qAll-4
5	1,843,023	29.02	2.69E-07	qAll-4
5	1,846,070	26.06	9.91E-07	qAll-4
5	1,852,744	12.91	6.83E-06	qAll-4
5	1,857,320	25.15	1.50E-06	qAll-4
5	1,860,636	24.58	1.88E-06	qAll-4
5	1,881,798	24.90	1.64E-06	qAll-4
5	1,886,598	25.27	1.42E-06	qAll-4
5	1,904,731	13.96	2.76E-06	qAll-4
5	1,905,516	22.40	5.13E-06	qAll-4
5	1,914,216	27.48	5.29E-07	qAll-4
5	1,919,059	27.06	6.42E-07	qAll-4
5	1,923,245	12.89	7.03E-06	qAll-4
5	1,923,466	25.43	1.28E-06	qAll-4
5	1,931,963	25.04	1.57E-06	qAll-4
5	1,967,152	23.65	2.87E-06	qAll-4
5	1,981,439	26.05	1.01E-06	qAll-4

Table 3. List of SNPs significantly associated with the inhibition of *Echinochloa* root growth.

We analyzed the allele content in the lead SNP and found that they displayed promising allele distribution among the cultivars as significative differences in the growth inhibition ability according to allele present in the accessions were found in all identified QTLs (**Figure 10A-D**). Accessions carrying different alleles of SNP qAll-2 showed a different barnyardgrass root growth inhibition value, those carrying T displayed 24.86% in contrast to those carrying G which displayed 9.08%. Similarly, qAll-2b, accessions carrying the minor allele C showed a root growth inhibition of 25.59%, meanwhile, those accessions carrying the major allele T showed 11.11% inhibition. While accessions carrying allele T in qAll-3 lead SNP didn't show barnyardgrass root inhibition, plants with the most common allele A showed a root growth inhibition of 16.4%. The mean values for barnyardgrass root growth inhibition were lower in the accession containing C, 9.08% of inhibition, than those carrying T, 24.95% of inhibition, at the lead SNP in qAll-5. The analysis of Ganao allele content in the lead marker of the four QTLs revealed that they carried the favourable allelopathy allele.

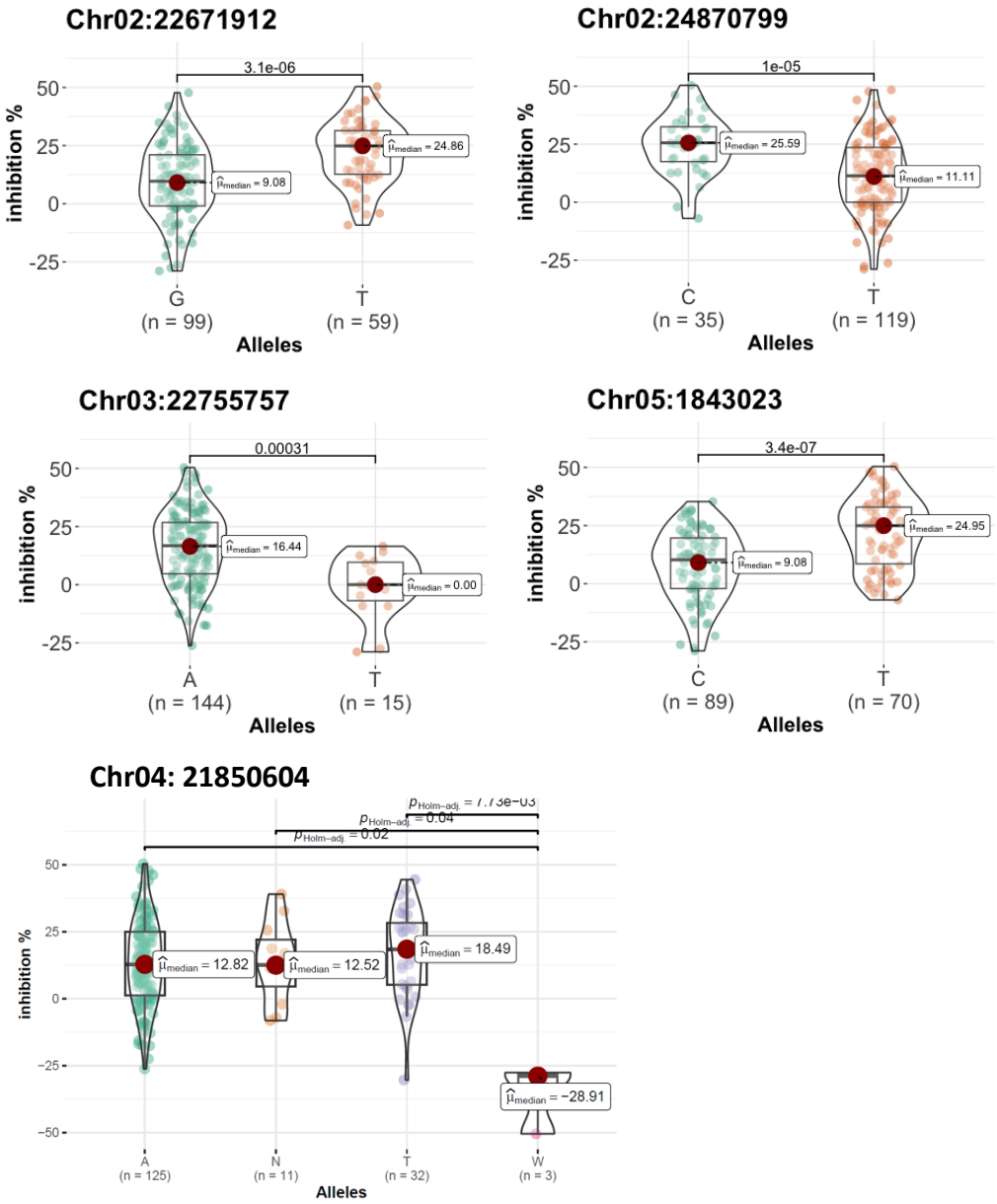


Figure 10. Violin plots. Comparison of the allele content at the lead SNP A qAll-2 B qAll-2b C qAll-3 and D qAll-5 in the population.

Mining Candidate Genes for Allelopathy

We looked for annotated genes that could be involved in allelopathic activity within the four QTL regions in an extended the interval of 150 Kb, using the annotated reference genome Nipponbare IRGSP-1.0 (RAP-DB) (**Table 4**).

Locus ID	Description	Gen Symbol	QTL	Referencia
Os05g0132700	R2R3-MYB TRANSCRIPTION FACTOR 52	*2R_MYB52*	4	
Os02g0623400	G1 LIKE PROTEIN 3	*G1L3*	2	Sultana et al 2023; Fang et al 2009
Os02g0624300	R2R3-type MYB transcription factor	*MYB30*	2	
Os02g0626100	Phenylalanine ammonia-lyase 1	*OsPAL1*	2	
Os02g0626400	Phenylalanine ammonia-lyase	*OsPAL*	2	
Os02g0626600	phenylalanine ammonia-lyase 3	*OsPAL3*	2	Sultana et al 2023
Os02g0627100	Phenylalanine ammonia-lyase 4	*OsPAL4*	2	
Os02g0589000	Lecithin:cholesterol acyltransferase family protein		1	Sultana et al 2023
Os02g0589700	Lecithin:cholesterol/phospholipid:diacylglycerol acyltransferase domain containing protein		1	
Os02g0590000	Lecithin:cholesterol/phospholipid:diacylglycerol acyltransferase domain containing protein		1	
Os02g0590200	Lecithin:cholesterol/phospholipid:diacylglycerol acyltransferase domain containing protein		1	
Os02g0590400	Lecithin:cholesterol acyltransferase family protein		1	

Table 4. List of known genes potentially related to allelopathy within 0.15 Mb distance interval from each QTL lead marker.

Os02g0589000, coding for a lecithin:cholesterol acyltransferase, was localised in qAll-1. Lecithin:colesterol acyltransferases, also named as phospholipid:diacylglycerol acyltransferases, are involved in the synthesis of triacylglycerol (Dahlqvist et al. 2000). The up-regulation of this gene has been observed in previous allelopathic assays (Sultana et al. 2023). Four additional genes coding for proteins of the lecithin:cholesterol acyltransferases family were localised in the vicinity in qAll-2. Another gene involved in lipid metabolism, *Os05g0133401*, coding for an esterase, is localised in qAll-5 (**Supplementary Table 1**).

PAL comprises a small family of proteins that are key enzymes in the phenylpropanoid synthesis pathway and, it has been related to allelopathic activity on barnyard grass on several occasions (Fang et al. 2013). Four genes coding for

PAL could be found in qAll-2b. One of these genes *Os02g0626600*, coding for PAL3, was previously found to be up-regulated in response to the presence of barnyard grass (Song et al. 2008; Sultana et al. 2023). *G1 LIKE PROTEIN 3* (*G1L3*) is also localized next to qAll-2b SNP peak. In the context of allelopathy, the up-regulation of *G1L3* has been reported previously in transcriptomic analysis of rice in the interaction with barnyardgrass (Fang et al. 2009; Sultana et al. 2023).

Two genes coding for R2R3-type MYB transcription factors, *MYB30* and *MYB52*, were found located in qAll-2b and qAll-5. R2R3-MYB factors play a positive role in the regulation of genes involved in secondary metabolism (Fang et al. 2020). It has been suggested that *OsMYB30* is a regulator of root cell elongation under ROS signals (Mabuchi et al. 2018) and root hair elongation under stress in Arabidopsis (Xiao et al. 2021). More interestingly, *OsMYB30* has been implicated in the biosynthesis of the phenolic compound stilbene in grapevine (Mu et al. 2023).

qAll-5 covers a genomic area that comprises 21 annotated genes (**Supplementary Table 1**). *Os05g0131500,* coding for a ferroportin was localised next to the leader marker position, but the role of this protein in allelopathy has not been reported previously (**Supplementary Table 1**).

No genes identified previously as associated with allelopathy have been detected in the region near qAll-3 (**Supplementary Table 1**).

Discussion

The management of damaging weeds in rice cultivation poses a significant challenge across various regions. Efforts to control these weeds while minimizing herbicide use are highly desirable, as this approach can help prevent the development of herbicide resistance and contribute to cost reduction in crop production. Some rice varieties can exhibit allelopathy and it has emerged as a potential solution for weed management in rice fields. When barnyardgrass seedlings were co-cultivated with specific rice varieties, they exhibited a reduction in root growth, highlighting the potential of allelopathy as a promising avenue for weed control in rice cultivation.

In this study, we examined the allelopathic potential of 171 *japonica* rice varieties by assessing their ability to inhibit the growth of barnyardgrass roots when co-cultured. Allelopathy can be induced by several factors as stress or nutrient deficiency (Song et al. 2008). It has been demonstrated that plant defense signalling hormones, such as jasmonic acid or salicylic acid, can induce allelopathy (Fang et al. 2009). To avoid inductor factors affecting the allelopathic activity, the bioassays were performed under controlled laboratory conditions. We performed analysis at early developmental stages as our goal was to find rice varieties with allelopathic potential, and the genes responsible for this potential, to improve the ability of rice seedlings to resist paddy weeds as early as possible in the field to provide a competitive advantage for rice by inhibiting the growth of competing weeds. Our findings revealed a substantial variation in the inhibition of barnyardgrass root growth when co-cultured with the different varieties. Within this collection, we identified novel allelopathic potential in varieties, as Ganao.

We employed GWAS to uncover genetic variants linked to the interaction between rice and barnyardgrass. The varieties we investigated were part of a previously established *japonica* collection, which encompasses those typically grown in temperate climates (Reig-Valiente et al. 2018). This collection has previously been utilized successfully, using a panel of 1,713 SNPs, to identify genetic polymorphisms associated with various agronomic traits and serves as a valuable resource for pinpointing genetic factors responsible for the variations in agronomic traits within these regions (Reig-Valiente et al. 2018). In this case, to facilitate a more precise phenotypic-genotypic association analysis, we conducted whole-genome

resequencing of the selected rice varieties, enabling the creation of a high-density SNP panel. As a result, we used a high-density SNP panel and identified four QTLs as candidate to play a role in the interaction between barnyardgrass and rice in *japonica* temperate rice.

Among these QTLs, qAll-2b stands out as it encompasses a cluster of four PAL coding genes associated with phenylpropanoid synthesis, which are compounds known to be involved in allelopathy. The first steps in the synthesis of phenylpropanoid-derived compounds are catalyzed by PAL, cinnamate 4-hydroxylase (C4H), and p-coumaroyl coenzyme A ligase (4CL). In this context PAL plays a pivotal role in the phenylpropanoid pathway, responsible for synthesizing various secondary metabolites. PAL facilitates the deamination of phenylalanine, an essential amino acid, to produce trans-cinnamic acid, a crucial precursor for a diverse array of phenolic compounds, including flavonoids, lignin, and other phenylpropanoid derivatives. These compounds are integral to the allelopathic properties of rice. Earlier studies have also highlighted the positive regulatory role of PAL in rice's allelopathic potential (Fang et al. 2013; Zhang et al. 2019b). PAL exhibits up-regulation in hydroponic systems in response to barnyard grass roots (Zhang et al. 2019b; Sultana et al. 2023). Additionally, when the expression of OsPAL is silenced, it results in reduced allelopathic activity from donor rice to barnyard grass (Fang et al. 2020). Notably, *PAL* is part of a multi-gene family in plants, with four members located on chromosome 2, and these were identified in qAll-2b.

In our association analysis, we detected two R2R3-MYB transcription factors within qAll-2b and 5. Prior studies have demonstrated the role of MYB transcription factors in regulating genes involved in the plant phenylpropanoid metabolic pathway. Specifically, R2R3-MYB factors are known to govern genes related to secondary metabolism (Fang et al. 2020). For instance, MYB57, an R2R3-MYB transcription factor, has been shown to transcriptionally regulate *MAPK11*, which interacts with PAL2;3 and modulates rice allelopathy (Fang et al. 2020). MYB30, on the other hand, is implicated in root hair development and functions within a MYB30-EIN3 antagonistic module in Arabidopsis (Xiao et al. 2021).

qAll-5 represents a particularly promising locus, housing a genomic region containing 21 genes. While some of these genes may have a role in allelopathy as

2R_MYB52, others, such as ferroportin found next to the lead marker position, have not been previously associated with this context. Additional investigation is required to elucidate their involvement in allelopathy.

Contents of allelochemicals in rice have been known to be varietal dependent (Kato-Noguchi et al. 2010). Several secondary metabolites have been found in root exudates that include phenolic acids, terpenes, fatty acids and indoles that have been considered as potential allelochemicals (Macías et al. 2007). Among them, momilactone B is present in roots exudates from different varieties and it has been proposed that the allelopathic activity of rice may depend primarily on the secretion level of momilactone B (Kato-Noguchi et al. 2010). The highly allelopathic potential line PI312777 also secretes momilactone B in the root exudates when cocultured with barnyadgrass. The GWAS analysis of the *japonica* temperate collection identified candidate genes for allelopathy that were related to phenolic acid and lipid metabolism, but not genes directly involved in the synthesis of momilactones were found. Our findings point to the participation of phenolic acids in allelopathy. Whether they are sufficient or they need to be accompanied by other chemicals to produce allelopathy, needs to be investigated.

Previous QTL analysis was conducted using RFLP markers in a bi-parental population resulting from the cross between PI31277 and Rexmont (Ebana et al. 2001). This analysis identified seven QTLs, with one of them located on chromosome 5, approximately 0.65 Mb away from qAll-5. Our use of a diverse panel of rice varieties, combined with genome sequencing techniques, significantly expands the potential to pinpoint QTLs associated with allelopathy with a higher degree of precision. Given the lower precision of the previously utilized technique compared to the use of a high-density panel of SNPs for GWAS analysis, it is highly likely that both QTLs are the same. The fact that this QTL has been identified in two studies, one of which employed a wide genetic diversity, strengthens the case for the significant role of this genomic region in allelopathy. Moreover, a detailed examination of the allelic content in the leader marker of qAll-5 revealed that 70 varieties carrying one allele inhibited *Echinochloa* root growth by 24.9%, in contrast to 9.1% for the 89 varieties carrying the other allele. qAll-5 emerges as a promising candidate for genetic breeding focused on enhancing allelopathy in rice.

Conclusions

We have assessed the allelopathic potential of a collection of *japonica* rice adapted to temperate regions and identified varieties that can inhibit the root growth of barnyardgrass when co-cultured. The association analysis using a high-density panel of markers mapped 4 QTLs that may help to understand allelopathy mechanism of action. These QTLs indicate that the phenylpropanoid synthesis pathway is taken part in the allelopathic potential of varieties within this collection. Among the identified QTLs, qAll-2b and qAll-5 are robust and constitute candidates to initiate a breeding program. The discovery of rice genotypes possessing allelopathic potential against weeds, along with the corresponding identification of associated QTLs, opens the opportunity to initiate a breeding program to incorporate allelopathy in rice while maintaining grain yield and quality that, ultimately, will benefit farmers as well as the environment as a powerful tool to combat parasitic weeds providing new weed management strategies in agriculture.

References

Amaro-Blanco I, Romano Y, Palmerin JA et al (2021) Different mutations providing target site resistance to ALS- and ACCase-inhibiting herbicides in Echinochloa spp. from rice fields. Agriculture 11:382. https://doi.org/10.3390/agriculture1105 0382

Bradbury PJ, Zhang Z, Kroon DE et al (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Cartwright D, Langcake P, Pryce RJ, Leworthy DP, Pryce RJ (1977) Chemical activation of host defense mechanisms as a basis for crop protection. Nature 267:511–513

Cartwright DW, Langcake P, Pryce RJ et al (1981) Isolation and characterization of two phytoalexins from rice as momilactones A and B. Phytochemistry 20:535–537. https://doi.org/10.1016/S0031-9422(00)84189-8

Dahlqvist A, Ståhl U, Lenman M et al (2000) Phospholipid:diacylglycerol acyltransferase: an enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. Proc Natl Acad Sci 97:6487–6492. https://doi.org/10.1073/pnas.120067297

Danecek P, Bonfield JK, Liddle J et al (2021) Twelve years of SAMtools and BCFtools. Gigascience. https://doi.org/10.1093/gigascience/giab008

Dilday RH, Lin J, Yan W (1994) Identification of allelopathy in the USDA-ARS rice germplasm collection. Aust J Exp Agric 34:907–910

Earl DA, VonHoldt BM (2012) STRU CTU RE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361. https://doi.org/10.1007/s12686-011-9548-7

Ebana K, Yan W, Dilday RH et al (2001) Analysis of QTL associated with the allelopathic effect of rice using water-soluble extracts. Breed Sci 51:47–51

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fang C-X, Xiong J, Qiu L et al (2009) Analysis of gene expressions associated with increased allelopathy in rice (Oryza sativa L.) induced by exogenous salicylic acid. Plant Growth Regul 57:163–172. https://doi.org/10.1007/s10725-008-9333-0

Fang C, Zhuang Y, Xu T et al (2013) Changes in rice allelopathy and rhizosphere microflora by inhibiting rice phenylalanine ammonia-lyase gene expression. J Chem Ecol 39:204–212. https://doi.org/10.1007/s10886-013-0249-4

Fang C, Yang L, Chen W et al (2020) MYB57 transcriptionally regulates MAPK11 to interact with PAL2;3 and modulate rice allelopathy. J Exp Bot 71:2127–2141. https://doi.org/10.1093/jxb/erz540

Goodman SN (2001) Of p-values and bayes: a modest proposal. Epidemiology 12:295–297. https://doi.org/10.1097/00001648-200105000-00006

Hassan SM, Aidy IR, Bastawisi AO, Draz AE (1998) Weed management using allelopathic rice varieties in Egypt. In: MO (ed) Allelopathy in rice. International Rice Research Institute, Manila, pp 27–37

He H, Wang H, Fang C et al (2012) Barnyard grass stress up regulates the biosynthesis of phenolic compounds in allelopathic rice. J Plant Physiol 169:1747–1753. https://doi.org/10.1016/j.jplph.2012.06.018

Jensen LB, Courtois B, Shen L et al (2001) Locating genes controlling allelopathic effects against barnyardgrass in upland rice. Agron J 93:21–26. https://doi.org/10.2134/agronj2001.93121x

Jensen LB, Courtois B, Olofsdotter M (2008) Quantitative trait loci analysis of allelopathy in rice. Crop Sci 48:1459–1469. https://doi.org/10.2135/cropsci2007.09.0533

Kaler AS, Gillman JD, Beissinger T, Purcell LC (2020) Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. Front Plant Sci 10:1–13. https://doi.org/10.3389/fpls.2019.01794
Kato-Noguchi H, Peters RJ (2013) The role of momilactones in rice allelopathy. J Chem Ecol 39:175–185. https://doi.org/10.1007/s10886-013-0236-9
Kato-Noguchi H, Ino T, Ota K (2008) Secretion of momilactone A from rice roots to the rhizosphere. J Plant Physiol 165:691–696. https://doi.org/10.1016/j. jplph.2007.07.018
Kato-Noguchi H, Hasegawa M, Ino T et al (2010) Contribution of momilactone A and B to rice allelopathy. J Plant Physiol 167:787–791. https://doi.org/10.1016/j. jplph.2010.01.014
Kawahara Y, de la Bastide M, Hamilton JP et al (2013) Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. Rice 6:4. https://doi.org/10.1186/1939-8433-6-4
Khamare Y, Chen J, Marble SC (2022) Allelopathy and its application as a weed management tool: a review. Front Plant Sci. https://doi.org/10.3389/fpls.2022.10346 49
Khanh TD, Xuan TD, Chung IM (2007) Rice allelopathy and the possibility for weed management. Ann Appl Biol 151:325–339. https://doi. org/10.1111/j.1744-7348.2007.00183x
Kohli R, Singh HP, Batish DR (2008) Allelopathic potential in rice germplasm against ducksalad, redstem and barnyard grass. J Crop Prot 4:287–301. https://doi.org/10.1300/ J144v 04n02
Kong L, Xu X, Hu F, Wang P, Jiang Y (2004) Release and activity of allelochemicals from allelopathic rice seedlings. J Agric Food Chem 52:2861–2865. https://doi.org/10.1021/jf035467i
Kong CH, Li HB, Hu F et al (2006) Allelochemicals released by rice roots and residues in soil. Plant Soil 288:47–56. https://doi.org/10.1007/s11104- 006- 9033-3
Letunic I, Bork P (2021) Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301
Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM
Mabuchi K, Maki H, Itaya T et al (2018) MYB30 links ROS signaling, root cell elongation, and plant immune responses. Proc Natl Acad Sci. https://doi.org/10.1073/pnas.1804233115
Macías FA, Molinillo JM, Varela RM, Galindo JC (2007) Allelopathy—a natural alternative for weed control. Pest Manag Sci 63:327–348. https://doi.org/10.1002/ps.1342
Minh BQ, Schmidt HA, Chernomor O et al (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015
Mu H, Li Y, Yuan L et al (2023) MYB30 and MYB14 form a repressor–activato module with WRKY8 that controls stilbene biosynthesis in grapevine. Plant Cell 35:552–573. https://doi.org/10.1093/plcell/koac308
Murdoch D, Adler D (2023) rgl: 3D Visualization Using OpenGL. R package version 1.2.1
Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4326. https://doi.org/10.1093/nar/8.19.4321
Navarez D, Olofsdotter M (1996) Relay seeding technique for screening allelopathic rice (Oryza sativa). In: Brown H, et al (eds) Proceeding of 2nd international weed control congress. Copenhagen, Denmark, pp 1285–1290
Oerke E-C, Dehne H-W (2004) Safeguarding production—losses in major crops and the role of crop protection. Crop Prot 23:275–285. https://doi.org/10.1016/j.cropro.2003.10.001
Olofsdotter M, Rebulanan M, Madrid A et al (2002) Why phenolic acids are unlikely primary allelochemicals in rice. J Chem Ecol 28:229–242
Ortiz EM (2019) vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis
Patil I (2021) Visualizations with statistical details: The “ggstatsplot” approach. J Open Source Softw 6:3167. https://doi.org/10.21105/joss.03167

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. https://doi.org/10.1093/genetics/155.2.945
Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. https://doi.org/10.1086/519795
R Core Team (2023) R: a language and environment for statistical computing
Reig-Valiente JL, Viruel J, Sales E et al (2016) Genetic diversity and population structure of rice varieties cultivated in temperate regions. Rice 9:58. https://doi.org/10.1186/s12284-016-0130-5
Reig-Valiente JL, Marqués L, Talón M, Domingo C (2018) Genome-wide association study of agronomic traits in rice cultivated in temperate regions. BMC Genomics 19:706. https://doi.org/10.1186/s12864-018-5086-y
Rice E (1984) Allelopathy, 2nd ed
Seal AN, Pratley JE, Haig T, An M (2004) Identification and quantitation of compounds in a series of allelopathic and non-allelopathic rice root exudates. J Chem Ecol 30:1647–1662. https://doi.org/10.1023/B:JOEC.0000042074.96036.14
Song B, Xiong J, Fang C et al (2008) Allelopathic enhancement and differential gene expression in rice under low nitrogen treatment. J Chem Ecol 34:688–695. https://doi.org/10.1007/s10886-008-9455-x
Sultana MH, Alamin M, Qiu J et al (2023) Transcriptomic profiling reveals candidate allelopathic genes in rice responsible for interactions with barnyardgrass. Front Plant Sci. https://doi.org/10.3389/fpls.2023.1104951
Takahashi N, Kato T, Tsunagawa M et al (1976) Mechanisms of dormancy in rice seeds: II. New growth inhibitors, Momilactone-A and-B isolated from the hulls of rice seeds. Jpn J Breed 26:91–98. https://doi.org/10.1270/jsbbs1951.26.91
Turner SD (2018) qqman: an R package for visualizing GWAS results using Q–Q and manhattan plots. J Open Source Softw 3:731. https://doi.org/10.21105/joss.00731
Wakefield J (2009) Bayes factors for Genome-wide association studies: comparison with P-values. Genet Epidemiol 33:79–86. https://doi.org/10.1002/gepi.20359
Xiao F, Gong Q, Zhao S et al (2021) MYB30 and ETHYLENE INSENSITIVE3 antagonistically modulate root hair growth in Arabidopsis. Plant J 106:480–492. https://doi.org/10.1111/tpj.15180
Xu M, Galhano R, Wiemann P et al (2012) Genetic evidence for natural product-mediated plant–plant allelopathy in rice (Oryza sativa). New Phytol 193:570–575. https://doi.org/10.1111/j.1469- 8137.2011.04005.x
Yu Z, Chen Y, Zhou Y et al (2023) Rice Gene Index: a comprehensive pangenome database for comparative and functional genomics of Asian rice. Mol Plant 16:798–801. https://doi.org/10.1016/j.molp.2023.03.012
Zhang P, Zhong K, Zhong Z, Tong H (2019a) Genome-wide association study of important agronomic traits within a core collection of rice (Oryza sativa L.). BMC Plant Biol 19:1–12. https://doi.org/10.1186/s12870-019-1842-7
Zhang Q, Zheng X-Y, Lin S-X et al (2019b) Transcriptome analysis reveals that barnyard grass exudates increase the allelopathic potential of allelopathic and non-allelopathic rice (Oryza sativa) accessions. Rice 12:30. https://doi.org/10.1186/s12284-019-0290-1
Zheng X, Levine D, Shen J et al (2012) A high-performance computing toolsetfor relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328. https://doi.org/10.1093/bioinformatics/bts606
Zhou Y, Chebotarov D, Kudrna D et al (2020) A platinum standard pan-genome resource that represents the population structure of Asian rice. Sci Data 7:113. https://doi.org/10.1038/s41597-020-0438-2

Supplemental Data

Tables

qAll-2a

Locus ID	Description	Gene Symbol
Os02g0584700	Heavy metal transport/detoxification protein domain containing protein.	*OsHPP3*
Os02g0584800	Copper metallochaperone heavy metal-associated plant protein	*OsHPP04*
Os02g0584900	Conserved hypothetical protein	
Os02g0585100	Similar to ATFP4	
Os02g0585200	Cd-binding protein, Metal chaperone, Cadmium detoxification by regulating Cd accumulation	*OsHIPP16*
Os02g0585566	Conserved hypothetical protein	
Os02g0585650	Hypothetical gene	
Os02g0585700	Quinonprotein alcohol dehydrogenase-like domain containing protein	
Os02g0586000	Quinonprotein alcohol dehydrogenase-like domain containing protein	
Os02g0586400	Similar to Small GTP-binding protein	*OsRab2A2*
Os02g0586500	Similar to OSIGBa0124N08.5 protein	
Os02g0586600	Conserved hypothetical protein	
Os02g0586700	Conserved hypothetical protein	
Os02g0586800	Homolog of fluorescent (FLU) in Arabidopsis thaliana, Tetratricopeptide repeat (TPR)-containing protein	*OsFLU2*
Os02g0586900	Similar to Glycine rich protein	*OsEnS-40*
Os02g0587000	Similar to Glycine rich protein	
Os02g0587300	Conserved hypothetical protein	
Os02g0587600	Conserved hypothetical protein	
Os02g0587800	Virulence factor, pectin lyase fold family protein	
Os02g0588250	Hypothetical gene	
Os02g0588300	Similar to Fimbrin 1 (AtFIM1)	
Os02g0588500	Hypothetical conserved gene	
Os02g0588550	Hypothetical gene	
Os02g0588600	Conserved hypothetical protein	
Os02g0588700	Conserved hypothetical protein	
Os02g0588775	Non-protein coding transcript	
Os02g0589000	Lecithin:cholesterol acyltransferase family protein	
Os02g0589100	Hypothetical conserved gene	
Os02g0589400	UDP-glucuronosyl/UDP-glucosyltransferase family protein	*UGT*
Os02g0589450	Hypothetical protein	
Os02g0589500	Hypothetical gene	
Os02g0589600	Hypothetical conserved gene	
Os02g0589700	Lecithin:cholesterol/phospholipid:diacylglycerol acyltransferase domain containing protein	
Os02g0589900	Hypothetical conserved gene	
Os02g0590000	Lecithin:cholesterol/phospholipid:diacylglycerol acyltransferase domain containing protein	
Os02g0590200	Lecithin:cholesterol/phospholipid:diacylglycerol acyltransferase domain containing protein	
Os02g0590400	Lecithin:cholesterol acyltransferase family protein	

qAll-2b

Locus ID	Description	Gene Symbol
Os02g0621800	Glucose-methanol-choline (GMC) oxidoreductase family protein, Hybrid breakdown (HB)	*HWH1*
Os02g0621850	Hypothetical gene	
Os02g0622100	Casein kinase I, Low temperature tolerance, Root development, Hormone sensitivity	*HBD2*
Os02g0622150	Non-protein coding transcript	
Os02g0622200	Phosphatidylinositol N-acetylglucosaminyltransferase family protein	
Os02g0622300	Ribosomal protein L14 family protein	*OsRPL14*
Os02g0622350	Similar to Root abundant protein	
Os02g0622400	Glycolipid transfer protein, GLTP domain containing protein	
Os02g0622500	Conserved hypothetical protein	
Os02g0623300	Conserved hypothetical protein	*ONAC051*
Os02g0623400	Similar to G1-like protein	*OsG1L3*
Os02g0623500	Lysyl-tRNA synthetase, class II domain containing protein	
Os02g0623550	Non-protein coding transcript	
Os02g0623600	Similar to OSIGBa0145M07.8 protein	*OsWAK12*
Os02g0623766	Hypothetical protein	
Os02g0623932	Similar to WAK80 - OsWAK receptor-like protein kinase	*OsWAK13*
Os02g0624100	EGF-type aspartate/asparagine hydroxylation site domain containing protein	*OsWAK13*
Os02g0624300	R2R3-type MYB transcription factor, Negative regulation of cold tolerance, Maintenance of Pi homeostasis	*OsMYB30*
Os02g0624350	Hypothetical protein	
Os02g0624400	Glycosyl transferase, family 8 protein	*OsPGSIP-B1*
Os02g0624700	Hypothetical protein	
Os02g0625000	Cryptochrome 2, blue light photoreceptor, Promotion of flowering time	*OsCRY2*
Os02g0625100	Hypothetical conserved gene	
Os02g0625300	Similar to PEPC kinase	*OsPPCK2*
Os02g0625500	Similar to Adenosine kinase-like protein	
Os02g0625900	Conserved hypothetical protein	
Os02g0626100	Similar to Phenylalanine ammonia-lyase	*PAL1*
Os02g0626200	Non-protein coding transcript	
Os02g0626400	Phenylalanine ammonia-lyase (EC 4.3.1.5)	*PAL2*
Os02g0626532	Hypothetical gene	
Os02g0626600	Phenylalanine ammonia-lyase, Salicylic acid biosynthesis	*PAL3*
Os02g0627100	Phenylalanine ammonia-lyase, Broad spectrum disease resistance	*OsPAL4*

qAll-3

Locus ID	Description	Gene Symbol
Os03g0603800	Peptidase A22A, presenilin family protein	
Os03g0604200	Similar to UDP-glucose dehydrogenase	
Os03g0604401	Similar to CBL-interacting protein kinase 14	
Os03g0604432	Hypothetical gene	
Os03g0604500	Hypothetical gene	
Os03g0604566	Conserved hypothetical protein	
Os03g0604600	Myosin-2 heavy chain-like protein	
Os03g0604700	Hypothetical protein	

Os03g0604950	Non-protein coding transcript	
Os03g0605300	Similar to Subtilisin-like protease	*OsSub30*
Os03g0605400	Non-protein coding transcript	
Os03g0606100	Similar to predicted protein	
Os03g0606200	Mitochondrial ATP synthase 6 kDa subunit, Response to salt and osmotic stress	*RMtATP6*
Os03g0606333	Hypothetical gene	
Os03g0606466	Conserved hypothetical protein	
Os03g0606600	Hypothetical conserved gene	
Os03g0607100	Hypothetical conserved gene	
Os03g0607200	Gibberellin regulated protein family protein	
Os03g0607400	Similar to Metal transporter Nramp6	
Os03g0607500	Similar to Seed maturation protein PM23 (Fragment)	
Os03g0607600	Similar to Cyclin-A3-1	*CycA3;1*
Os03g0607700	Zinc finger, C2H2-like domain containing protein	
Os03g0607800	Conserved hypothetical protein	
Os03g0608000	Hypothetical protein	
Os03g0608200	Non-protein coding transcript	
Os03g0608600	Peptidase, trypsin-like serine and cysteine domain containing protein	*OsDeg-like 5*

qAll-5

Locus ID	Description	Gene Symbol
Os05g0129700	KNOX (Knotted1-like) class homeobox protein	*Oskn2*
Os05g0129800	DNA-binding WRKY domain containing protein	*OsWRKY109*
Os05g0129900	Tetratricopeptide-like helical domain containing protein	
Os05g0130100	Hypothetical conserved gene	
Os05g0130200	Conserved hypothetical protein	
Os05g0130300	Conserved hypothetical protein	
Os05g0130400	Protein of unknown function DUF247, plant family protei	
Os05g0130500	Hypothetical protein	
Os05g0130600	Conserved hypothetical protein	
Os05g0131000	Protein of unknown function DUF247, plant family protein	
Os05g0131100	Conserved hypothetical protein	
Os05g0131401	Conserved hypothetical protein	
Os05g0131500	Ferroporti-1 domain containing protein	
Os05g0131900	Hypothetical conserved gene	
Os05g0132000	Pentatricopeptide repeat domain containing protein	
Os05g0132100	AMP-dependent synthetase and ligase domain containing protein	*OsLACS1*
Os05g0132300	Hypothetical protein	
Os05g0132400	Agenet domain containing protein	
Os05g0132500	Protein of unknown function DUF246, plant family protein	
Os05g0132601	Non-protein coding transcript	
Os05g0132650	Trans-acting small interfering RNA precursor, Putative precursor TASiRNA 3-A2	*OsTAS3e*
Os05g0132700	Similar to R2R3 Myb-like protein	*Os2R_MYB52*
Os05g0132800	Conserved hypothetical protein	
Os05g0133100	Similar to PII protein (Fragment)	*OsGlnB*
Os05g0133401	Esterase, SGNH hydrolase-type domain containing protein	*OsGELP60*

Os05g0133900	Similar to Shaggy-related protein kinase eta (EC 2.7.1.-)	*OsDRM3*
Os05g0134000	Similar to Glycogen synthase kinase	*OsSK13*

Supplementary Table 1. List of annotated genes within 150 kb distance interval from the SNP marker peak.

Chapter II

Comparative RNA-Seq of an Allelopathic and Non-Allelopathic Temperate Rice Varieties Reveals Candidate Genes Driving Early Weed-Suppressive Responses to *Echinochloa crus-galli*

Julia García-Romeral and Concha Domingo

Departamento del Arroz, Centro de Genómica, Instituto Valenciano de Investigaciones Agrarias (IVIA)

Manuscript prepared for submission

Abstract

Weed infestation is one of the major constraints on rice cultivation and the increasing presence of weeds in Europe means that sustainable alternatives to chemical control are needed. Allelopathy, the plant's ability to release bioactive compounds that affect neighbouring species, offers a promising strategy for weed suppression. Many aspects of the interaction between weeds and rice plants remain poorly understood, including the mechanism by which weeds can trigger the release of allelochemicals from rice plants. To better understand the effect of barnyardgrass (*Echinochloa crus-galli*) on rice, we explored the transcriptomic response of two temperate *japonica* rice varieties with contrasting allelopathic potential, Ganao (allelopathic) and Bomba (non-allelopathic), during co-cultivation. Our goal is to uncover the genes and mechanisms underlying allelopathic potential, with the aim of enhancing the ability of rice seedlings to suppress weed growth, providing a competitive advantage. RNA-seq analysis was performed on root samples collected at 0, 6, and 36 hours after the start of co-cultivation. The results revealed a progressive activation of secondary metabolism in Ganao, including strong upregulation of phenylpropanoid biosynthetic genes at 36 hours, such as those encoding PAL and 4CL enzymes, and transcription factors like *OsMYB30* and *OsMYB52*, previously mapped in a GWAS. In contrast, genes involved in the biosynthesis of momilactones and other diterpenoid phytoalexins were down-regulated, suggesting a mechanism distinct from that of previously studied allelopathic varieties like PI312777. The transcriptomic response of Ganao pointed to a phenylpropanoid-mediated allelopathic strategy, likely supported by coordinated activation of detoxification and structural reinforcement pathways rather than the synthesis of diterpenoid route. The identification of candidate genes and pathways involved in the plant-weed interaction provide useful for breeding varieties with allelopathic potential.

Introduction

Rice (*Oryza sativa* L.) remains one of the world's most essential crop, with approximately 165 million hectares dedicated to its cultivation, constituting about 11% of the world's arable land. Field losses due to infestations from weeds, pests, and diseases cause significant and irreparable damage to the crop at a global scale. Among these factors, weed infestation stands out as the leading cause of yield loss, exceeding the combined losses caused by pests and diseases (Asaduzzaman et al. 2010). To reduce the yield loss of rice due to weed infestation, different strategies such as herbicides application and manual labor are incorporated in rice fields that are costly and profitless. Besides, the use of herbicides raises substantial environmental concerns, including soil and water contamination, as well as negative effects on non-target species and biodiversity. Furthermore, herbicides can become ineffective due to the development of resistance to their active compounds (Amaro-Blanco et al. 2021).

One of the most world-wide extended weeds in rice paddy fields is barnyardgrass (*Echinochloa* spp.), a highly competitive and persistent genus leading to yield reductions from 21% to 79% (Bajwa et al. 2015). Its mimetic and adaptative capabilities challenge the traditional weed management practices (Turra et al. 2023; Zhang et al. 2021). These challenges underscore the urgent need for novel and sustainable weed control strategies. Allelopathy has emerged as a promising approach. It's defined as the biological phenomenon by which some organisms are able to release compounds into the environment to affect the growth and development of neighboring species (Rice 1984).

Since the concept of allelopathy was first introduced in rice (Olofsdotter M 1998), extensive research has aimed the isolation and identification of allelochemicals from plants and roots exudates with inhibitory potential (Chung et al. 2002; Dilday et al. 1994; Olofsdotter et al. 1995). In this context, numerous studies have found a range of phytotoxic compounds in rice, including fatty acids, benzoxazinoids, indoles (Belz 2007), phenolic acids (Rimando et al. 2001), phenyl alkanoic acids such as alkyl resorcinol (Bouillant et al. 1994), flavones (Kong et al. 2004) and diterpenoids.

The allelopathic potential differs between rice varieties and efforts have been made to identify those with the ability to suppress weed growth. Early studies were focused on screening a wide range of rice accessions and revealed significant variation, particularly between rice genetic population (Dilday et al. 1989; Olofsdotter et al. 1997). *Japonica* varieties have been found to exhibit stronger allelopathic effects compared to *indica* varieties (Khanh et al. 2007). These findings stablish allelopathy in *japonica* rice as an optimal model for investigating the molecular mechanisms underlying natural weed suppression. The genetic basis of allelopathy has been investigated studying biparental populations with several QTL being identified (Jensen et al. 2001; Okuno et al. 2003). Most research efforts have concentrated on the PI312777 rice line (Ebana et al. 2001; Fang et al. 2013; Xiao et al. 2020; H. Zhang et al. 2024), revealing that its major allelochemicals include phenolic compounds, organic acids, and diterpenoids, such as tricin and momilactone B (Kato-Noguchi et al. 2003). Momilactone B, has been found in roots exudates of allelopathic cultivars, as PI31277 and Koshihikari, and is considered as a main allelochemical responsible for weed suppression in rice (Kato-Noguchi 2011; Kato-Noguchi et al. 2005; Kong et al. 2004; Serra Serra et al. 2021). Interestingly, momilactone B is also involved in the modulation of soil microbial communities (Kong et al. 2004). The biosynthetic pathway of momilactones is well understood (De la Peña et al. 2021). It has been proven that the lack of function of enzymes involved in synthesis of diterpertenes such as the *copalyl diphosphate synthase 4* (*OsCPS4*) and *kaurene synthase-like 4* (*OsKSL4*), results in a reduction in the allelopathic activity of Zhonghua 11 and Hwayoung respectively against barnyardgrass (Xu et al. 2012). The transcriptomics of line PI312777 in interaction with plants or exudates of barnyardgrass have been studied using RNA sequencing (RNA-seq) (Sultana et al. 2023; Zhang et al. 2019), providing a list of potential candidate genes that may be responsible. However, despite these advances, focusing solely on the allelopathic mechanisms of a single variety provides a limited view of the phenomenon, leaving other potential mechanisms and allelochemicals unexplored.

In the previous chapter, a Genome-Wide Association Study (GWAS) was conducted using a temperate *japonica* rice collection and four QTLs were mapped (García-Romeral et al. 2024). The most promising loci, qAll2b and qAll5, included genes involved in the biosynthesis of secondary metabolites, such as four genes encoding

Phenylalanine Ammonia Lyase (PAL), a key enzyme in phenolic compound synthesis, as well as *OsMYB30* and *OsMYB52*, two genes encoding R2R3-type MYB transcription factors. R2R3-MYB factors play a positive role in the regulation of genes involved in secondary metabolism (Cao et al. 2020). It has been suggested that *VvMYB30* is implicated in the biosynthesis of the phenolic compound stilbene in grapevine (Mu et al. 2023). Employing HiSeq and chromatin immunoprecipitation, Fang et al. 2020 suggested that *OsMYB57* positively regulates rice allelopathy through the transcriptionally regulation of MAPK11, a mitogen-activated protein kinase that interacts with phenylalanine ammonia-lyase PAL2;3. Li et al. (2020) demonstrated that *OsPAL2-1* was crucial for regulating rhizospheric microbial communities that aid allelopathic effects.

The mechanism of allelopathy remains unclear. The role of phenolics in allelopathy has been questioned on several occasions and it has been suggested that they might act synergistically with other compounds to function as allelochemicals (Olofsdotter et al. 2002). In the same sense, despite a main role in allelopathy has been given to momilactone B in PI312777, their participation in other root inhibition system has been questioned as momilactones A and B have very low correlation with weed resistance compared to drought or salt tolerance (Xuan et al 2016). In addition, no QTLs linked to momilactones were detected in previous biparental populations or GWAS analysis in *japonica* rice (Jensen et al 2001; García-Romeral et al 2024).

RNA-seq technology enables in-depth analysis of gene expression through transcriptome quantification, thereby providing valuable insights into the genes and pathways activated in allelopathic interactions. In this study, RNA-seq was employed to investigate the differences in the transcriptome between the allelopathic variety Ganao and the none-allelopathic Bomba during co-cultivation with barnyardgrass, with the aim of identifying the genes and pathways that play a role in the allelopathic response. Ganao was selected from 163 *japonica* varieties adapted to temperate climates, based on previous bioassays conducted in García-Romeral et al. (2024), where their ability to inhibit barnyardgrass root growth was evaluated. The highest allelopathic potential was exhibited by Ganao. Bomba variety showed no inhibition, as the baryarngrass roots growth was equal in the presence or in the absence of the variety.

Material and Methods

Plant material and bioassay

The allelopathic potential of Ganao and Bomba was assayed in co-cultured experiments with barnyardgrass, *Echinochloa crus-galli* and *Echinochloa hispidula* as previously described (García-Romeral et al, 2024). Briefly, on the first day, 12 rice seeds from both varieties were sown separately in a thin perlite layer on 9 cm diameter Petri dishes and 20 ml of MES 1 mM at pH 6.0, with a total of 12 plates per variety and time-point. Seeds were germinated and grown in a growth chamber for six additional days at 26 °C under 16h light/8h dark photoperiod conditions. On the fourth day, barnyardgrass seeds were sown in Petri dishes at 28 °C in dark conditions and cultured for three additional days. On the seventh day, rice roots from each variety were collected and immediately frozen in liquid nitrogen to stablish time 0 (T0). At the same time, 9 barnyardgrass seedlings with root lengths between 6 and 8 mm were placed equidistant from each other in small holes in the perlite layer. The seedlings were grown in co-culture with each rice variety for 6 hours (T1) or 36h (T2), after which rice roots were collected and frozen. All samples collected at the three time points were stored at -80 °C prior to RNA isolation. The tests were performed 4 times independently, with 2 to 3 replicates in each case.

The percentage of inhibition of barnyardgrass root growth was recorded as the difference in root length when barnyardgrass was grown in the presence or absence of rice, as shown:

$$\%\ of\ inhibition\ = \frac{RLB\_control\ (mm) - RLB\_\ cocultured\ (mm)}{RLB\_control\ (mm)} \times\ 100$$

Where RLB_control is the root length of the barnyardagrass under control conditions (absence of rice) and RLB_cocultured is the root length of the barnyardgrass under co-cultivation conditions.

RNA isolation and sequencing

Total RNA was extracted from the rice roots using a CTAB-LiCl extraction method. Frozen root samples were ground with a mortar and pestle, and RNA was extracted from three replicates of 800 mg each using extraction buffer (0.1M LiCl; 0.1M Tris pH8; 1% SDS; 0.01M EDTA) and a mixture of phenol:chloroform:isoamyl (25:24:1) at pH=4.5, and was then precipitated with LiCl at a final concentration of 2M LiCl and resuspended in TE. RNA was quantified with the NanoDrop 1000 Spectrophotometer (Thermo Scientific, Waltham, MA, USA) and the Qubit™ RNA BR Quantification Assay Kit using the Qubit 2.0 Fluorometer (Life Technologies,EEUU).

RNA-seq was performed at Novogene UK Company Limited (Cambridge, United Kingdom). mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. After fragmentation, the first strand of cDNA was synthesized using random hexamer primers followed by the second strand cDNA synthesis. The library was ready after end repair, A-tailing, adapter ligation, size selection, amplification, and purification. Quantified libraries were pooled and sequenced on Illumina NovaSeq 6000 system PE150 according to a 350 bp library type and to standard Illumina operation procedures with a yield of > 12 Gb per lane and median phred quality score of Q36.

The raw reads containing adapters were removed, as well as those containing more than a 10% of undetermined or low quality bases (Qscore<= 5).

Clean reads were aligned to the Nipponbare reference annotated genome (IRGSP-1.0 2024-01-11) from The Rice Annotation Project Database (rap-db, https://rapdb.dna.affrc.go.jp/index.html) with STAR aligner and specifying a maximum intron length of 3000 and a maximum genomic distance between mates of 1000. Read counts from the gene expression was quantified from the alignment using Feature Counts software (Subread package 2.0.8) with default parameters based on rice RAP-DB annotation gtf file from 2024-01-11.

Genes with very low expression were removed, remining those with a minimal count per million (CPM) of 0.33 in at least 1. The data was normalized by the regularized

log (Love et al. 2014), and the missing values were treated as the gene median. After filtering, from the 337,840 genes, 28,619 genes remained.

K-means clustering analysis was performed, in first place, ranking the genes by their standard deviation across all samples (those with the highest variability). The expression data was centered by subtracting the mean expression level of each gene. A distance matrix was calculated using the formula 1- r, where r is Pearson's correlation coefficient. Clustering was carried out using the average-linkage method with the hclust function from the stats R package (v4.3.0).

PCA was calculated with the prcomp function in R stadistical software (v.4.3.0).

DEGS and enrichment analysis

DESeq2 (V 1.20.0) package of the R program was employed to estimate the log2 fold change (FC), the adjusted pvalue for multiple testing and the differentially expressed genes (DEGs) from the read count data matrix normalized of the expression level of the genes. The threshold level was set at FDR ≤ 0.1, and log2 FC $\geq \pm 1$. All comparisons between time points and varieties were applied.

Functional annotation according to Gene Ontology (GO) was performed through the AmiGO2 database (https://amigo.geneontology.org/amigo/landing). DEGs were considered significantly enriched in GO terms at P-value ≤ 0.05.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was performed to retrieve the enriched significant pathways. Pathview (https://pathview.uncc.edu) was used for data integration and visualization of gene expression across the pathways. The GO terms with FDR < 0.05 were considered enriched.

To provide a comprehensive overview of the results, we utilized the MapMan software (https://mapman.gabipd.org/mapmanstore) for gene expression visualization within the context of metabolic and regulatory pathways.

RT-qPCR

Real-time PCR (RT-qPCR) in one step was performed using an LightCycler 2.0. (Roche Diagnostics GmbH, Mannheim, Germany) using the LightCycler 4.0 software for data registration, as melting temperature (Tm), and threshold cycle (Ct). Primers were designed using Primer3Plus (Table S1). Ubiquitin was used as control to normalize expression between different samples.

The reaction mixtures were prepared following the manufacturer's instructions of the LightCycler® FastStart DNA MasterPLUS SYBR Green I (Roche, Ref: 03515885001), adding MultiScribe™ Reverse Transcriptase (Invitrogen™, Ref: 4311235), using 100 ng of RNA. After 30 min of reverse transcription at 48°C, followed by 10 min of hot start at 95°C, the amplification consisted of 40 cycles, which cycle including 10s of denaturation at 95°C, 10 s of annealing at 55 to 60°C, depending on the primer used, and 10 s of extension at 72°C. Additionally, a melting curve program was set as follows: 1 cycle of 15 s at 95 °C, 1 min at 42 °C with a temperature increase slope of 0.1 °C/s from 42°C to 95°C.

The relative mRNA abundance was calculated by the threshold cycle (Ct). Standard curve for each selected gene and ubiquitin reference gene were generated to evaluate the Efficiency of amplification (E) which was equal to 10-1/slope of standard curve. The relative mRNA abundance was determined by the log2 fold change as described in Pfaffl, M (2001), given by the formula:

$$log2FC = -log_2\left(\frac{EGanao^{ct}/Eubi^{ct}}{EBomba^{ct}/Eubi^{ct}}\right)$$

The real-time PCR reactions for the tested genes and the internal reference gene were respectively subjected to three replicates for each sample.

Results

Allelopathic effects of japonica rice on barnyardgrass

The inhibition of barnyardgrass root growth due to the presence of two *japonica* rice varieties, Ganao and Bomba, was evaluated in four independent bioassays according to previous analysis (García-Romeral et al 2024). Differences in the barnyardgrass root length could be observed after 15 days of culture when rice root length was recorded (**Table 1**, **Figure 1**). Barnyardgrass co-cultured with Ganao showed shorter roots, 17.8 mm length, than when co-cultured with Bomba or independently, 33.2 and 33.5 mm respectively. Ganao exhibited a high inhibitory rate of 49.3%, whereas Bomba showed an almost negligible inhibitory effect -1.1% compared to barnyardgrass plants without rice plants.

	Root Length (mm)	Root Growth Inhibition (%)
Barnyardgrass + Ganao	17.8 ± 4.4	49.3 ± 7.2
Barnyardgrass + Bomba	33.2 ± 2.2	-1.1 ± 5.4
Barnyardgrass	33.5 ± 4.2	

Table 1. Effect of allelopathic and non-allelopathic rice lines on the growth of *E. crus-galli*. Root length in co-culture with Ganao, allelopathic, and Bomba, non-allelopathic, and independently (Control). The barnyard root length and root growth inhibition are indicated.

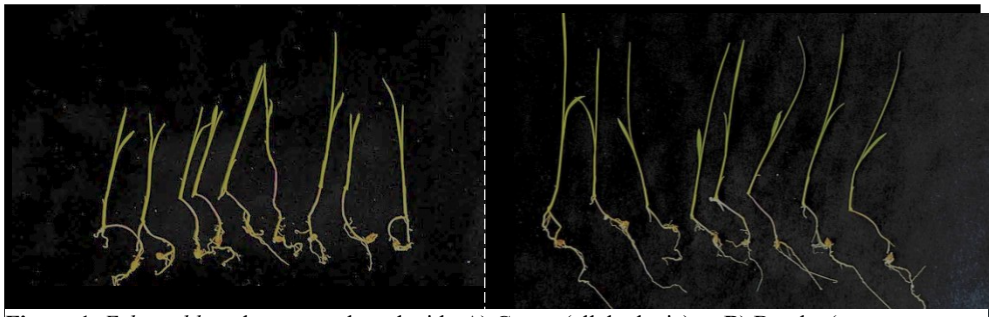


Figure 1. *Echinochloa* plants co-cultured with **A)** Ganao (allelophatic) or **B)** Bomba (non-alellophatic)

Transcriptional changes in rice roots in response to the presence of barnyardgrass

Rice plants were grown in the presence of barnyardgrass plants and roots were collected at three different time points, 0, 6 and 36 h, and RNA was isolated and sequenced. Reads obtained by RNA sequencing were filtered according to quality parameters and mapped against the Nipponbare reference genome. The statistic summary of the transcriptomic data is shown in **Table 2**. The total clean reads generated varied from 40.47 to 98.21 million. The Q30, equivalent to a 99.9% accuracy rate for a specific base call, and GC content mean was 79.5 and 50.3 %, respectively. And the overall mean mapping rate was 78.6% with 37,840 genes mapped where 28,619 passed the minimum CPM of 0.33 filtering (**Figure 2**). After data normalization, using the regularized log method, numerical analysis of the differential gene expression was performed.

Variety	Time point	Replicate	Raw Bases	Uniquely Mapped Reads	Multi-Mapping Reads	Raw Read Counts
Ganao	**T0**	1	49536476	41226505 (83.2%)	1028077 (2.1%)	33761728
		2	84066899	63050764 (75.0%)	1508941 (1.8%)	51277168
		3	44443085	34723322 (78.1%)	872537 (2.0%)	28045161
	T1	1	42663080	34634997 (81.2%)	1140913 (2.7%)	28747026
		2	44577169	34567781 (77.5%)	1407061 (3.2%)	28423680
		3	48955483	41483435 (84.7%)	1337248 (2.7%)	34354615
	T2	1	43863166	35190619 (80.2%)	1674979 (3.8%)	29322994
		2	40472336	31409864 (77.6%)	1647427 (4.1%)	26192779
		3	42411087	31770415 (74.9%)	1249941 (2.9%)	26646825
Bomba	**T0**	1	45352074	38221568 (84.3%)	1129699 (2.5%)	31357027
		2	46835792	38783934 (82.8%)	1222646 (2.6%)	32317999
		3	44554050	35914337 (80.6%)	941650 (2.1%)	29456002
	T1	1	45663163	39427998 (86.3%)	1175150 (2.6%)	32583968
		2	46821905	40795660 (87.1%)	1229490 (2.6%)	33865775
		3	98217802	84690700 (86.2%)	2634867 (2.7%)	70457264
	T2	1	49029010	31432184 (64.1%)	1354310 (2.8%)	26632963
		2	43600453	28510078 (65.4%)	1150712 (2.6%)	24079189
		3	46648786	30707888 (65.8%)	1738271 (3.7%)	25681743

Table 2. Alignment results from RNA sequencing reads

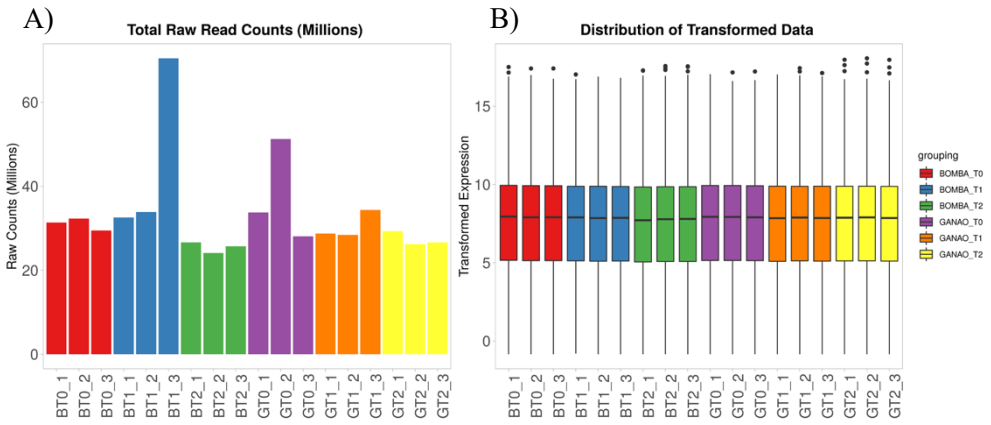


Figure 2. A) Barplot of total Raw Read Counts (Millions) across experimental conditions for both the Bomba and Ganao rice varieties at three time points (T0, T1, and T2). Each bar represents the total number of reads for a given sample, with color-coding indicating the sample grouping (Bomba or Ganao) and the respective time point. **B)** Boxplot of transformed expression data across different varieties and time points. The median expression level is marked by the bold horizontal line within each box, and the whiskers extend to 1.5 times the interquartile range, with outliers shown as points.

Principal Component Analysis (PCA) (**Figure 3**) showed that the three biological replicates were tightly grouped, and that the samples were distributed according to time on the x-axis and variety on the y-axis. PC1 was found to be correlated with time (p=2.10E-11) and PC2 with variety (p=3.27E-07) (**Figure 3**). The separation of the samples across both axes reflects the influence of both temporal progression and genetic background on the overall variation in gene expression.

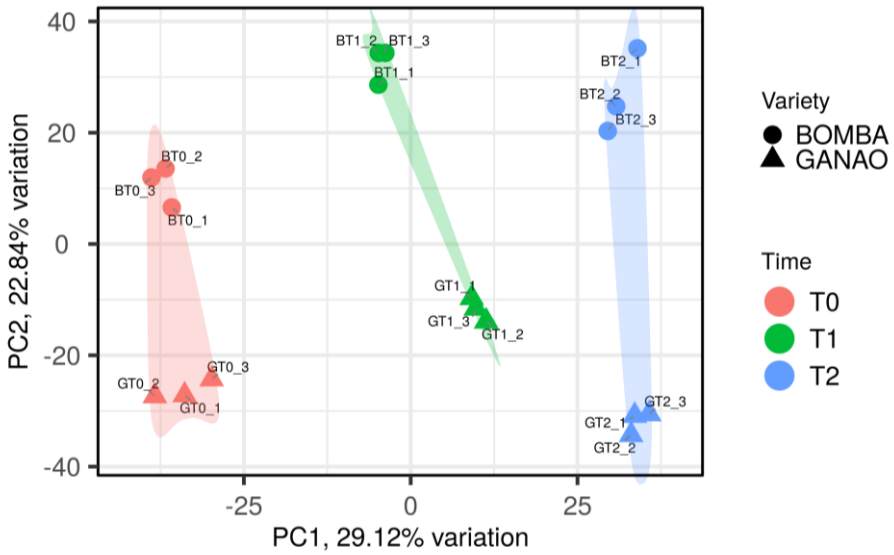
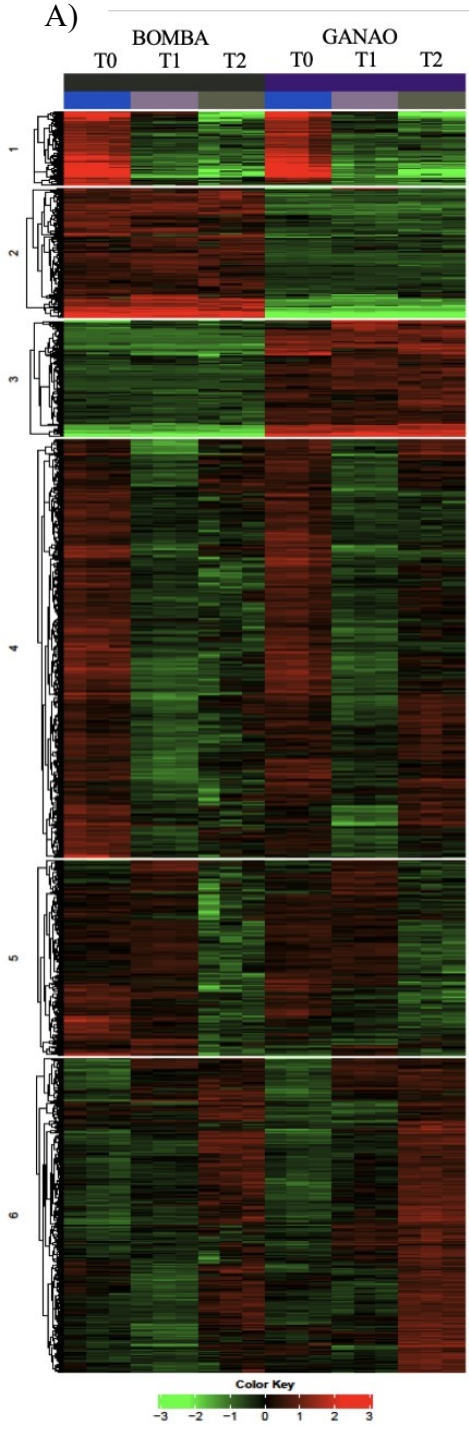


Figure 3. Principal Component Analysis (PCA) of transformed gene expression data from Bomba and Ganao at three different time points. The plot shows three biological replicates for each variety and time point. Samples are grouped along the y-axis according to variety and along the x-axis according to time.

K-means clustering of the 2000 most expressed genes (based on their standard deviation) generated 6 expression patterns groups (**Figure 4**). Within each group, genes were classified according to the GO Biological Process subontology. The p-value, calculated from the one-sided hypergeometric test and then adjusted for multiple tests using the Benjamini-Hochberg procedure, is converted to FDR (false discovery rate). The FDR indicates the probability of observing this level of enrichment by chance. Fold enrichment is defined as the percentage of genes in the list that belong to a given group divided by the corresponding percentage of background genes. As a measure of effect size, fold enrichment indicates how significantly genes in each group are overrepresented.

Clusters 1, 5 and 6 displayed similar expression patterns in both cultivars. Clusters 1 and 5 displayed a high number of transcripts before the presence of barnyardgrass and decreasing over time. On the contrary, cluster 6 included genes with low expression at T0 and T1 but increasing at T2 in both cultivars. The abundance of transcripts in cluster 4 decreased at T1 but showed different levels of expression at T2. Clusters 2 and 3 included genes with a clear differential expression between

A)



B)

Group	FDR	Fold	Pathway
1	2.55E-09	114.85	Cold acclimation
	1.33E-06	14.36	Response to abscisic acid
	1.33E-06	14.08	Response to alcohol
	3.42E-06	20.52	Response to water deprivation
	3.42E-06	20.16	Response to water
	3.92E-06	19.14	Response to acid chemical
	3.92E-06	18.83	Response to cold
	3.16E-05	8.67	Response to lipid
	3.80E-05	10.17	Response to temperature stimulus
	2.62E-04	12.60	Seed development
2	1.69E-02	9.35	Lipid transport
	1.69E-02	8.60	Lipid localization
	1.85E-02	98.95	Reg. of protein export from nucleus
	1.85E-02	29.68	Reg. of protein serine/threonine phosphatase activity
	2.45E-02	79.16	Reg. of nucleocytoplasmic transport
	3.38E-02	2.90	Defense response
	3.89E-02	18.55	Reg. of phosphoprotein phosphatase activity
	4.25E-02	17.90	Reg. of phosphatase activity
	4.29E-02	47.80	Ribosomal small subunit export from nucleus
	4.41E-02	35.98	Ribosomal large subunit export from nucleus
3	1.82E-02	14.40	Phenylpropanoid biosynthetic process
	1.82E-02	12.40	Secondary metabolite biosynthetic process
	1.82E-02	88.75	Reg. of DNA-binding TF activity
	1.82E-02	111.20	Positive reg. of DNA-binding TF activity
	4.33E-02	18.50	Iron ion homeostasis
	6.47E-02	6.50	Secondary metabolic process
	9.41E-02	7.60	Phenylpropanoid metabolic process
	9.41E-02	12.10	Transitionn metal ion homeostasis
4	7.22E-06	2.00	Response to chemical
	1.13E-05	5.20	Cellular modified amino acid metabolic process
	1.13E-05	6.50	Glutathione metabolic process
	4.96E-05	3.80	Secondary metabolic process
	5.28E-05	3.30	Sulfur compound metabolic process
	2.13E-04	5.20	Response to fungus
	2.52E-04	7.90	Response to salicylic acid
	1.69E-03	5.00	Defense response to fungus
	4.23E-03	4.20	Secondary metabolite biosynthetic process
5	6.36E-03	3.80	RRNA processing
	6.36E-03	25.03	Nitrate transport
	6.36E-03	3.65	RRNA metabolic process
	1.63E-02	3.00	Ribosome biogenesis
	1.63E-02	14.90	Endonucleolytic cleavage involved in rRNA processing
	1.63E-02	14.90	Endonucleolytic cleavage of tricistronic rRNA transcript
	1.63E-02	6.20	Glutamine family amino acid metabolic process
	1.63E-02	2.70	Ribonucleoprotein complex biogenesis
	1.63E-02	7.20	Maturation of SSU-rRNA
	2.18E-02	7.90	Maturation of SSU-rRNA from tricistronic rRNA transcript
6	2.04E-10	38.20	Photosynthesis light harvesting in photosystem I
	2.35E-08	23.88	Photosynthesis light harvesting
	2.20E-07	10.00	Phenylpropanoid biosynthetic process
	6.54E-07	17.50	Protein-chromophore linkage
	8.73E-07	7.46	Secondary metabolite biosynthetic process
	3.40E-06	6.01	Phenylpropanoid metabolic process
	9.90E-06	17.36	Photosynthesis
	1.22E-05	17.36	Secondary metabolic process
	1.43E-04	4.48	Water transport

Figure 4. **A)** K-means clustering heatmap of the top 2,000 differentially expressed genes between *Ganao* and *Bomba* at three time points (T0, T1, T2) under co-cultivation with *E. crus-galli*. Red and green colors indicate up and down-regulated transcripts respectively **B)** Significantly enriched Gene Ontology (GO) biological process terms for each cluster, including False Discovery Rate (FDR) values and fold enrichment scores.

Ganao and Bomba. Cluster 2 included terms with higher number of transcripts in Bomba than in Ganao. Interestingly, GO term annotation revealed that "defense response", as well as "lipid metabolism and transport" were included in cluster 2. Genes in cluster 3 were more highly expressed in Ganao than in Bomba. The enriched terms in cluster 3 included terms related to secondary metabolic processes and phenylpropanoid biosynthesis, as well as terms such as "regulation of DNA-binding transcription factor activity" or "metal ion homeostasis".

Comparison of the expression profiles of the allelopathic Ganao and non-allelopathic Bomba

The expression profiles of the allelopathic Ganao were compared with non-allelopathic Bomba. A multi-factorial linear model test was performed on the entire dataset, using DESeq2, to analyze transcriptomic data taking into account both varieties and the different time points. A threshold of 2-fold-change and FDR <0.1 was set for evaluating differential gene expression and a total of 7,576 differentially expressed genes (DEGs) were identified, including the three time points (**Figure 5A**). At the initial point, T0, immediately before barnyardargrass was added to the culture, a total of 1,167 DEGs were detected of which 664 were down-regulated in Ganao, which was higher than the 503 up-regulated DEGs. At T1, 1,690 DEGs were found. Of these, 948 were up-regulated and 742 were down-regulated. The largest number of DEGs was observed at T2, with a considerable increase in the number of transcripts, reaching a total of 4,719l DEGs. Of these, 3,123 corresponded to up-regulated genes and 1,596 to down-regulated genes. Among the upregulated genes, 104, 391 and 2,560 transcripts were specific to T0, T1 and T2 respectively, meanwhile 263, 336 and 1,171 downregulated were specific to T0, T1 and T2 respectively (**Figure 5B**).

From the total significant DEGs, only those with a p-value < 0.05 were classified using AmiGO2 enrichment analysis tool according to their annotation and the assignment of the functional categories based on their GO into Biological Processes (BP), Cellular Component (CC) and Molecular Function (MF) (**Figure 6**).

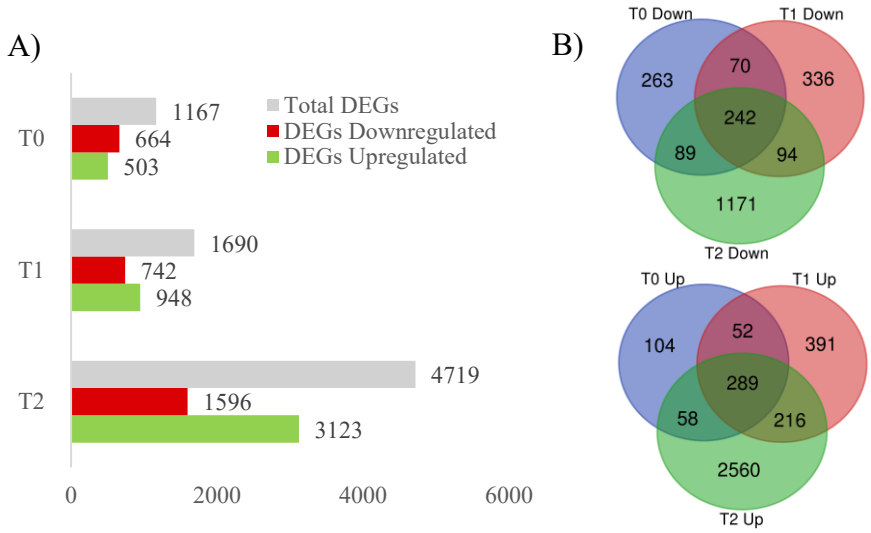


Figure 5. Differentially Expressed Genes (DEGs) in Ganao roots compared to Bomba at T0, T1, and T2. **A)** Barplot of total, down and up-regulated DEGs. **B)** Venn diagrams of up-regulated (down) and down-regulated (up) genes for Ganao showing the overlapping number of genes obtained across the three time points.

The number of GO terms showed an increase at T1 and T2 with more than 50 and 46 up-regulated GO terms respectively, encompassing a broad array of biological processes (**Figure 7**). At T1, the pathways associated with nitrate assimilation and nitrate transport were notably enriched (FDR = 1.06E-05), suggesting an activation of nitrogen metabolism during this phase, caused by the presence of barnyardgrass seedlings in the co-culture. Additionally, "hydrogen peroxide catabolic process," were also up-regulated. Notably, the defense response in Ganao was transiently down-regulated compared to Bomba at T1, as indicated by the enrichment of different GO terms such as “defense response”, “response to stress” or “response to biotic stimulus”. At T1 (**Figure 7**), the biosynthesis and metabolic processes of diterpenoids were downregulated. Consequently, GO terms analysis suggests that Bomba elicits an earlier response to the stress induced by co-cultivation with barnyardgrass, while simultaneously activating the secondary metabolite pathway responsible for phytoalexin production

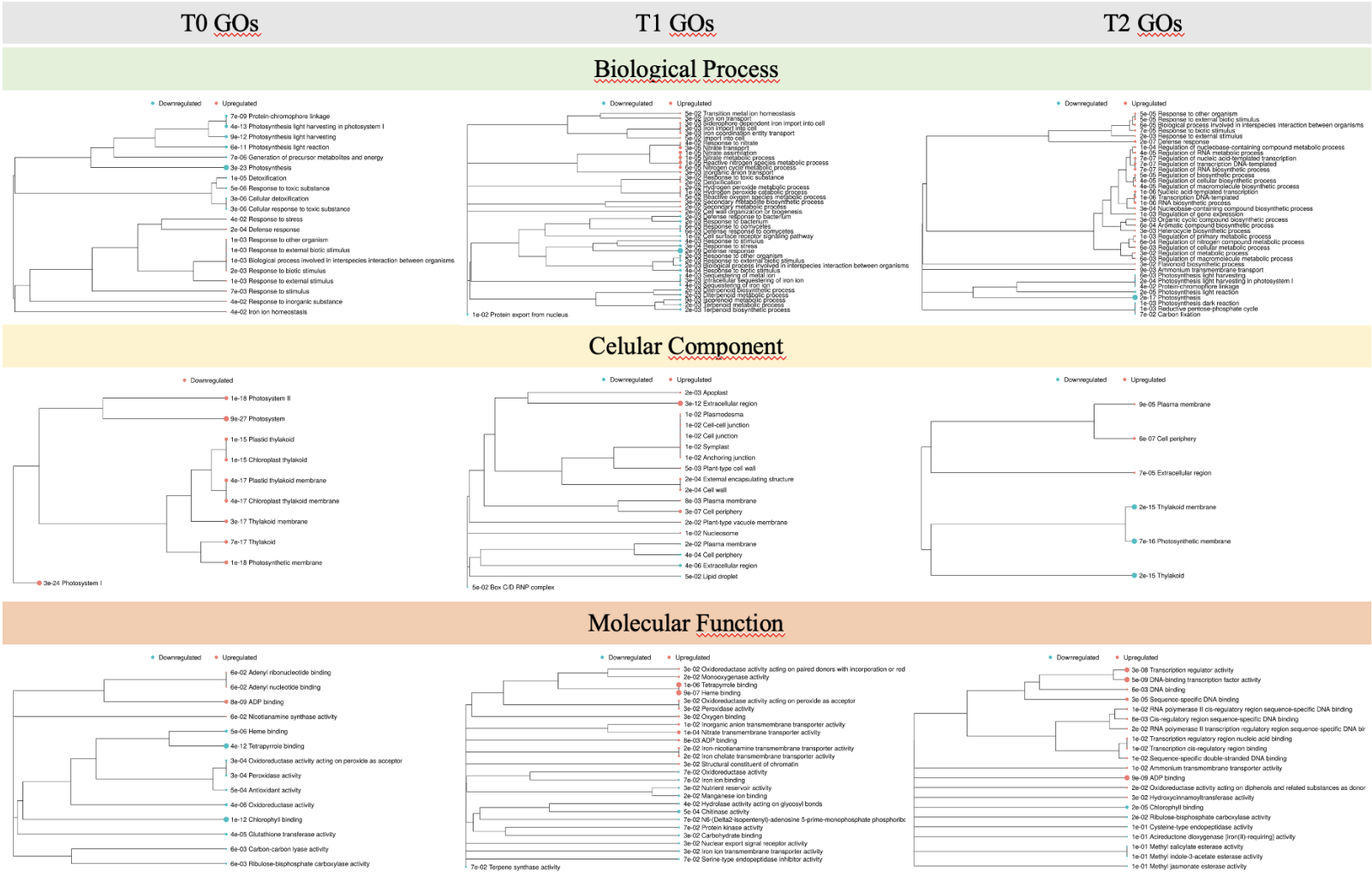


Figure 6. GO term enrichment displaying the up-regulated (red) and down-regulated (green) according to the categories of Biological Processes (BP) Cellular Component (CC), and Molecular Function (MF) across three time points: T0, T1, and T2.

Figure 7. GO term enrichment displaying the up-regulated (red) and down-regulated (green) Biological Processes (BP) within Gene Ontology (GO) terms across three time points: A) T0, B) T1, and C) T2. Terms are ordered from the lowest to the highest false discovery rate (FDR). Each bar displays the number of genes associated with the respective GO terms, color-coded to reflect the corresponding Fold Change values.

At T2 (**Figure 7**), photosynthesis term is down-regulated, reaching the initial level, at T0. Additionally, pathways associated with cellular structure, such as cell wall organization, highlight the organism's adaptive strategies to ensure structural integrity in response to barnyardgrass co-cultivation.

Additionally, the transcriptomic analysis revealed significant differences in the activation of genes from large enzyme families between the rice varieties Ganao and Bomba. At T0, prior to co-cultivation, both varieties exhibited no differences in the expression with no major stress-related responses (**Figure 8**). However, after 6 hours (T1) and particularly at 36 hours (T2), Ganao showed a strong upregulation of critical enzyme families, including cytochrome P450s, UDP-glucosyltransferases (UDPs), glutathione S-transferases (GSTs), and peroxidases (**Figure 8**). These enzyme families play vital roles in the detoxification, oxidative stress management, and the biosynthesis of secondary metabolites, which seems to be essential in the allelopathic interactions.

KEGG pathway and MAPMAN analysis of DEGS

KEGG database was used to investigate the implication of the DEGs in common biological processes during the interaction of rice and barnyardgrass. The KEGG pathway enrichment highlighted significant up-regulation of the biosynthesis of plant secondary metabolites in Ganao plants compared to Bomba after 6h, T1, and 36h, T2, of co-culture (**Table 3**). The biosynthesis of secondary metabolite pathway was further studied using MAPMAN to investigate dynamic changes over the three time points (**Figure 9**). During the interaction between Ganao and barnyardgrass plants, the upregulation of several key secondary metabolite biosynthetic pathways was evident at T1 and increased at T2 (**Figure 9**). This included a significant increase in phenylpropanoid biosynthesis, simple phenols, and flavonoids. In contrast, Bomba showed minimal activation of these pathways at the same time point, suggesting a lower metabolic response interaction with barnyardgrass. These findings highlight the role of secondary metabolites in the allelopathic mechanism.

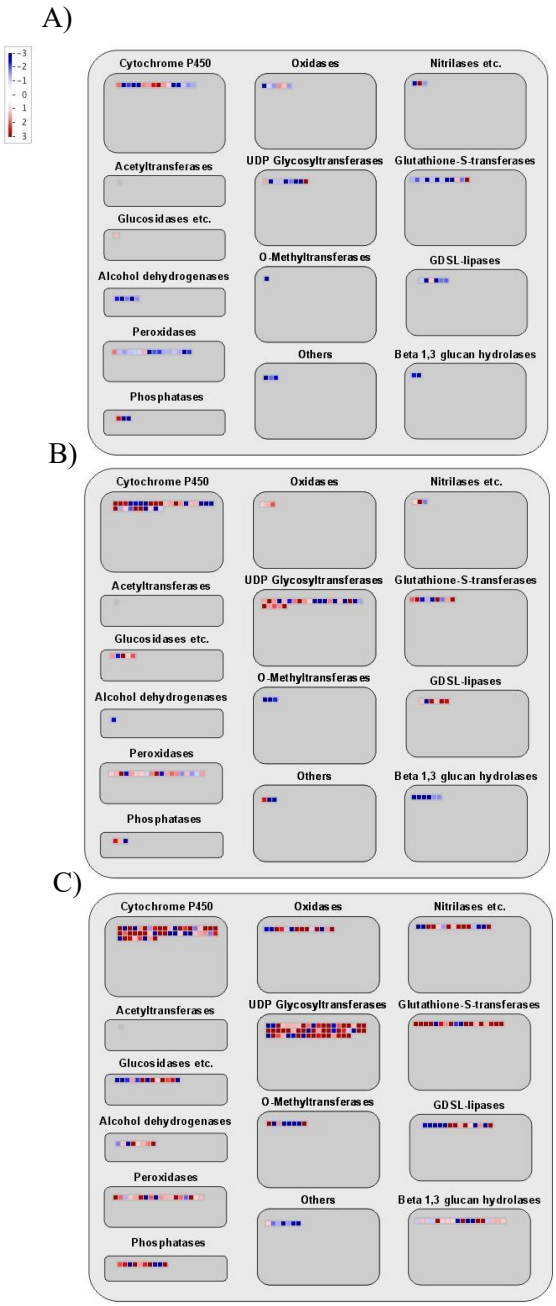


Figure 8. Expression of large enzyme families in rice varieties Ganao and Bomba during co-cultivation with *E. crus-galli*. A) T0: initial time, before co-cultivation. B) T1: After 6 hours. C) T2: After 36 hours. Red and blue boxes indicate up-regulated and down-regulated genes respectively.

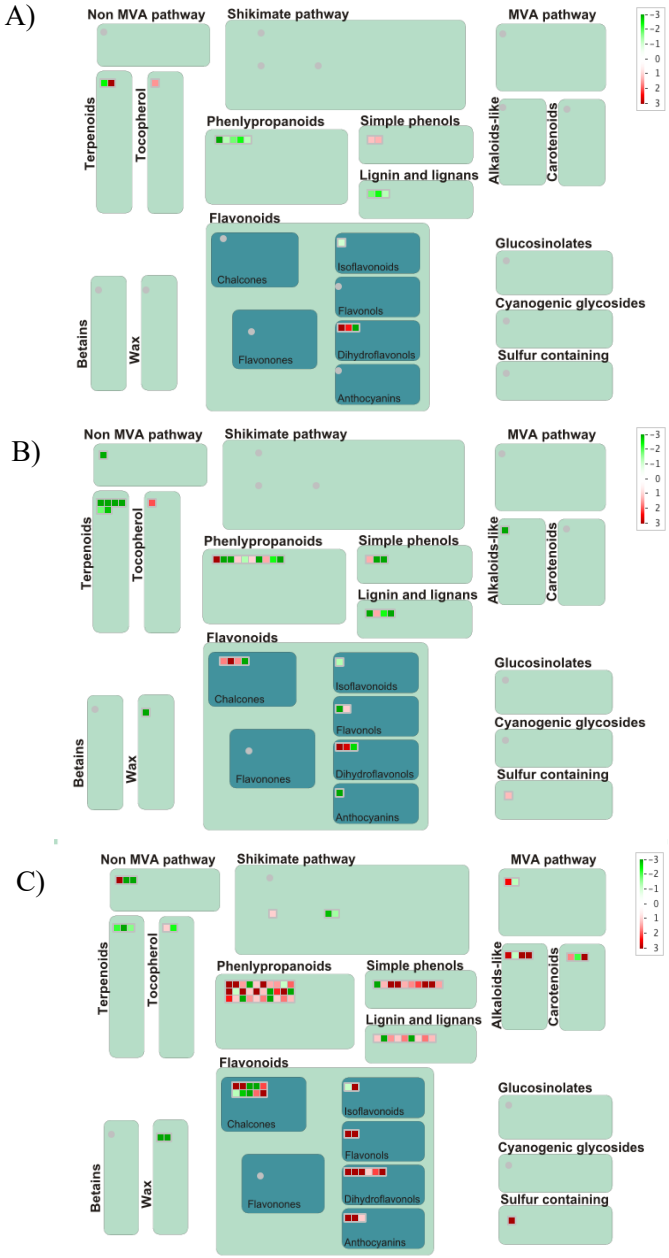


Figure 9. Differential regulation of secondary metabolite biosynthesis pathways in Ganao and Bomba at three time points A) T0, B) T1, C) T2. Red and green boxes indicate up-regulated and down-regulated genes respectively given by its FoldChange value. The analysis was performed using MAPMAN

	T0			T1			T2	
Pathway		**Adj.Pval**	**FC**		**Adj.Pval**	**FC**		**Adj.Pval**	**FC**
Biosynthesis of secondary metabolites				up	9.36E-3	1.6	up	1.53E-3	1.5
Biosynthesis of various plant secondary metabolites	up	2.76E-2	9.3	up	1.67E-3	7.4			
Phenylpropanoid biosynthesis	down	9.29E-4	3.3	up	1.06E-3	4.2	up	9.27E-5	3.1
Diterpenoid biosynthesis				down	2.13E-4	10.1			
Cysteine and methionine metabolism	up	3.03E-2	5.6						
Carbon fixation in photosynthetic organisms	down	1.17E-2	3.8				down	4.30E-2	3
Glycolysis/Gluconeogenesis	down	3.35E-2	2.7						
Metabolic pathways	down	1.78E-9	1.6	up	1.67E-3	1.5	up	1.09E-2	1.2
Nitrogen metabolism				up	1.06E-3	12.7			
Photosynthesis	down	5.26E-15	14.2				down	7.38E-8	7.5
Photosynthesis-antenna proteins	down	8.23E-10	20.3				down	6.67E-5	10.1
Alpha-Linolenic acid metabolism	down	5.51E-2	4.1						

Table 3. Significant KEGG pathways identified through the comparison of Ganao vs Bomba at three time points (T0, T1, and T2). The table lists the pathways, their respective adjusted p-values (Adj. Pval), and fold changes (FC) indicating the direction and magnitude of regulation.

The further analysis of KEGG pathways revealed that the phenylpropanoid pathway showed distinct gene expression changes over the three time points (T0, T1, and T2) during the co-cultivation of rice with barnyardgrass. At T0 (**Figure 10A**), genes in this pathway were predominantly down-regulated. However, at T1 (**Figure 10B**) (after 6 hours of co-cultivation), a clear shift in transcription occurred with the activation of several genes. Interestingly, while T1 showed significant upregulation compared to T0, not all components of the phenylpropanoid pathway were fully activated at this stage. This indicates that the 6-hour co-cultivation period was sufficient to initiate the early activation of the pathway, but not to maximize the expression of most of its components, as observed at T2 (**Figure 10C**).

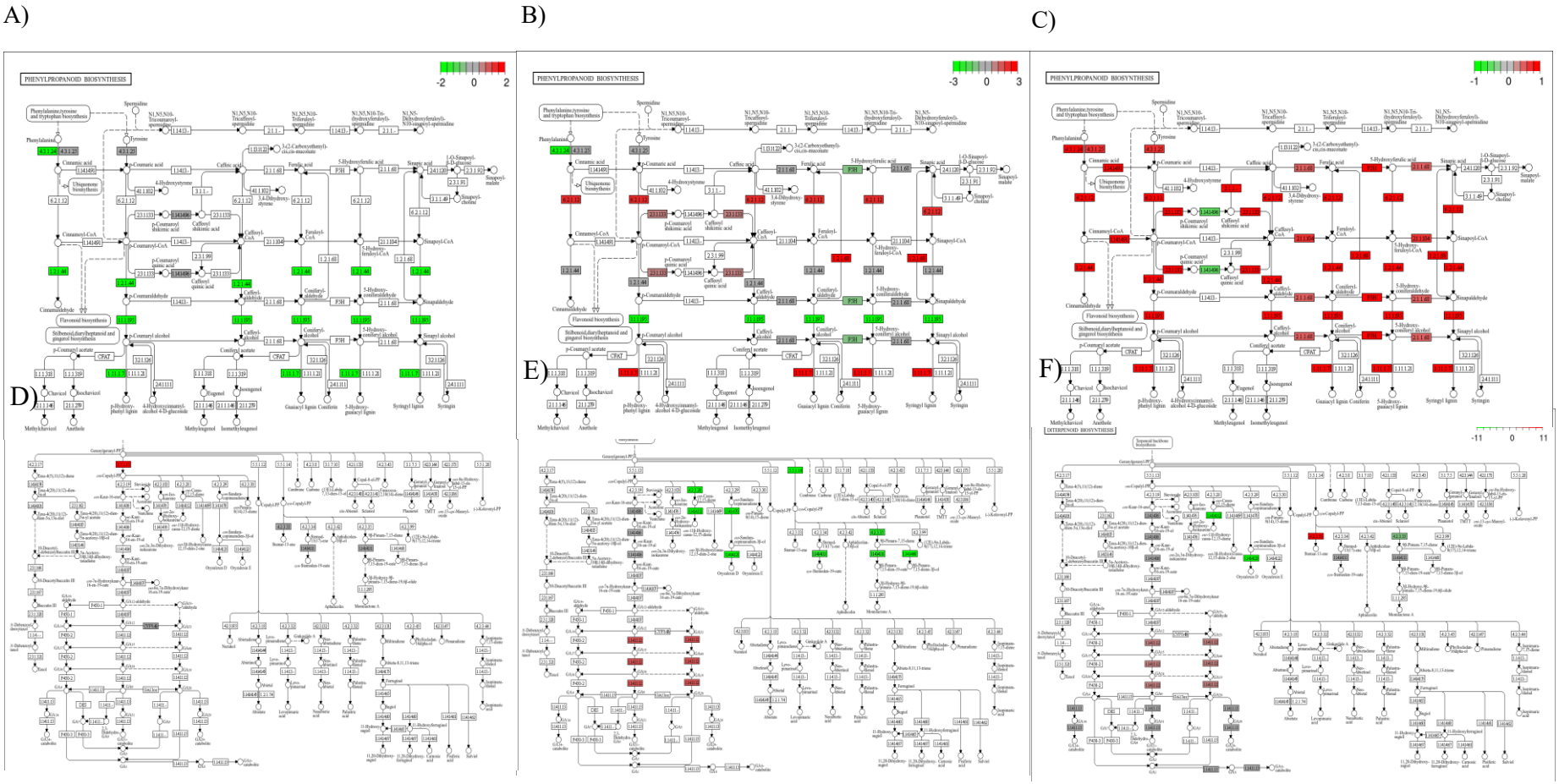


Figure 10. Gene expression patterns in the phenylpropanoid (a, b and c) and diterpenoid biosynthetic (d, e and f) pathways from KEGG under co-cultivation with Echinochloa. The three time points (T0, T1, T2) represent the rice plants before co-cultivation (T0), after 6 hours of co-cultivation (T1), and after extended co-cultivation (T2). The top panel shows the phenylpropanoid pathway, where genes initially down-regulated at T0 **A)** become up-regulated at T1 **B)** and full pathway activation is observed at T2 **C)**. The bottom panel illustrates the diterpenoid pathway, highlighting a consistent downregulation of genes involved in momilactone and oryzoalexins biosynthesis across all time points, T0 **D)**, T1 **E)**, and T2 **F)**

Remarkably, the diterpenoid biosynthesis pathway was downregulated at T1, with a Fold Change = 10.1 (**Table 3**). Focusing on the production of key metabolites such as momilactones and oryzalexins, the pathway exhibited a contrasting pattern of regulation at different time points. At T0 (**Figure 10D**), there were no significant differences in the expression levels of genes involved in the biosynthesis of diterpenoids between Ganao and Bomba. However, at T1 (**Figure 10E**), there was a high decrease in the fold change of genes involved in the synthesis of momilactones and oryzalexins D and E. At T2 (**Figure 10F**), their expression levels recovered to nearly reach the initial point observed at T0. This finding is noteworthy, as while diterpenoids are widely reported in the literature to be associated with allelopathic activity (Kato-Noguchi et al. 2013), they are not participating in the interaction with barnyardgrass with Ganao. On the other hand, the step involving the conversion of stemar-13-ene to oryzalexin S showed a distinct upregulation, particularly at T2.

Cysteine and methionine metabolism was also up-regulated at T0 but no significant differences could be observed at T1 and T2. Nitrogen metabolism was up-regulated at T1, but no significant differences were found at T0 orT2. On the other hand, photosynthesis and carbon fixation in photosynthetic organism was down-regulated at T0 and T2. Glycolisis/gluconeogenesis and the metabolism alpha-linolenic acid were down-regulated at T0.

Transcripts associated with Phenylpropanoid and Lignin Biosynthesis

The analysis of individual genes in the phenylpropanoid biosynthesis pathway enabled the identification of distinct expression trends involved in various steps, providing insight into how gene regulation evolves throughout the metabolic process. The dynamics of gene activity at each step of the pathway were carefully assessed over time (**Table 4**).

Activity	RAPDB ID	Annotation	GT0-BT0_log2FC	GT0-BT0_adjPval	GT1-BT1_log2FC	GT1-BT1_adjPval	GT2-BT2_log2FC	GT2-BT2_adjPval
PAL	*Os04g0518400*	Phenylalanine ammonia-lyase	-1.80	9.18E-03	-3.72	1.28E-10	0.08	9.86E-01
	Os02g0627100	Phenylalanine ammonia-lyase (OsPAL4)	0.20	8.25E-01	-0.52	2.89E-02	1.12	1.17E-09
Acid → CoA	*Os08g0245200*	4-coumarate:coenzyme A ligase	0.27	9.91E-01	1.40	7.77E-04	1.57	4.96E-05
	Os02g0177600	4-coumarate:coenzyme A ligase	0.04	1.00E+00	0.63	4.09E-07	1.10	2.48E-21
	Os08g0143300	AMP-dependent synthetase and ligase domain containing protein	-0.49	1.00E+00	0.08	9.98E-01	6.35	8.77E-03
CoA → Aldehyde	*Os02g0808800*	Cinnamoyl-CoA reductase 1	-6.31	4.67E-05	-1.70	5.89E-01	-6.75	9.17E-04
	Os01g0283600	Similar to Cinnamoyl CoA reductase	4.82	5.94E-01	3.11	7.65E-01	7.13	8.19E-02
	Os09g0419200	Cinnamoyl-CoA reductase 19	0.62	4.09E-02	0.60	4.45E-02	1.76	1.07E-15
Aldehyde → CoA	*Os01g0591300*	Similar to Cytosolic aldehyde dehydrogenase RF2D	-0.57	1.31E-01	1.36	8.11E-08	0.29	3.97E-01
	Os01g0591000	Cytosolic aldehyde dehydrogenase	0.50	6.69E-01	1.82	1.15E-05	1.06	1.31E-02
CoA → shikimic/quinic acid	*Os02g0611800*	Similar to Hydroxyanthranilate hydroxycinnamoyltransferase 3	0.36	3.73E-01	1.01	3.63E-06	1.19	8.06E-09
	Os09g0422000	Transferase family protein	0.05	1.00E+00	1.06	1.87E-04	1.50	4.93E-09
Aldehyde → Alcohol	*Os09g0399800*	Cinnamyl alcohol dehydrogenase	-1.27	1.98E-05	0.26	7.31E-01	0.43	2.16E-01
	Os09g0400400	Cinnamyl alcohol dehydrogenase	-2.36	4.90E-06	-0.71	4.64E-01	1.24	2.73E-02
	Os04g0229100	Similar to Sinapyl alcohol dehydrogenase	-1.31	4.07E-01	-2.34	1.60E-02	1.78	4.67E-02
Hydroxylase	*Os10g0512400*	Coniferaldehyde 5-hydroxylase	0.26	8.69E-01	-0.81	9.55E-03	1.24	3.05E-06
Alcohol → Lignin	*Os02g0240100*	Class III peroxidase	0.04	1.00E+00	2.92	3.25E-02	0.50	7.99E-01
	Os03g0234900	Similar to Peroxidase	-1.17	3.42E-03	0.49	4.59E-01	0.14	8.42E-01
	Os03g0368900	Haem peroxidase family protein	-1.22	2.38E-07	0.13	8.87E-01	0.74	2.65E-03
	Os04g0688600	Peroxidase	-0.87	1.00E+00	-0.40	9.70E-01	6.99	2.65E-03
	Os05g0162000	Similar to Peroxidase	-1.37	8.74E-02	0.37	8.53E-01	1.87	2.04E-03
	Os05g0135500	Haem peroxidase family protein	0.14	1.00E+00	1.01	5.26E-02	1.46	4.96E-04
	Os06g0695300	Haem peroxidase	-0.26	1.00E+00	0.74	8.63E-01	9.08	2.24E-07
	Os07g0104500	Haem peroxidase	-0.41	9.30E-01	1.23	4.38E-02	0.23	7.89E-01
	Os07g0677300	Peroxidase	0.00	1.00E+00	1.08	2.15E-08	1.13	1.91E-09
	Os07g0638300	1 Cys-peroxiredoxin A	-0.47	1.00E+00	-0.30	9.41E-01	4.36	1.50E-02
	Os07g0639400	Similar to Peroxidase 1	-0.53	6.46E-01	0.73	2.58E-01	1.99	7.96E-07
	Os08g0205000	Transferase domain containing protein	0.81	1.00E+00	6.44	5.82E-02	8.06	3.34E-03
	Os10g0566800	Haem peroxidase	0.29	1.00E+00	1.67	9.94E-03	0.85	2.14E-01
	Os11g0112400	Peroxidase	-1.03	8.85E-01	-1.09	7.02E-01	2.87	2.04E-02
R2-R3 MYB	*Os02g0624300*	R2R3-type MYB transcription factor (OsMYB30)	0.1	1.00E+00	-3.4	4.50E-01	1.39	5.60E-01
	Os05g0132700	R2R3-MYB Transcription Factor 52 (OsMYB52)	0.9	1.00E+00	-1.9	6.30E-01	6.65	3.60E-03

Table 4. Log2 fold change (log2FC) and adjusted p-values (adjPval) for gene expression changes across three time points (T0, T1, T2) in the phenylpropanoid biosynthesis pathway. Each gene is grouped based on its role in the pathway. Genes shown displayed log2 FC $\geq \pm 1$ in at least one of the time points.

The first steps in the phenylpropanoid biosynthesis pathway involve the deamination of phenylalanine to form cinnamate and ammonia, a reaction catalyzed by phenylalanine ammonia-lyase (PAL). We found two PAL DEGs, *Os04g0518400* and *Os02g0627100*, which exhibit different expression patterns. *Os04g0518400* was significantly down-regulated at both T0 and T1, with a log2FC = -1.80 and -3.71, respectively (**Table 4**). In contrast, *Os02g0627100* showed low expression at T0 with an increase of expression at T2, with a log2FC = 1.12 compared with non-allelopathic Bomba. In the following steps of the phenylpropanoid biosynthesis pathway, two DEGs for 4-coumarate:coenzyme A ligases (4CL), *Os08g0245200* and *Os02g0177600*, showed marked upregulation over time. Both genes displayed moderate increases at T1, with further significant upregulation at T2. Additionally, *Os08g0143300*, encoding an AMP-dependent synthetase and ligase domain containing protein, undergoes a sharp increase in expression at T2 (log2FC = 6.35). Additionally, three cinnamoyl-CoA reductases DEGs were found with different patterns of expression. While *Os02g0808800* was down-regulated and *Os01g0283600* up-regulated over the whole period of interaction, *Os09g0419200* was up-regulated at T2 (**Table 4**).

Genes responsible for the conversion of shikimic and quinic acids, such as *Os02g0611800*, encoding a protein similar to Hydroxyanthranilate hydroxycinnamoyltransferase 3, and *Os09g0422000*, encoding a transferase, exhibited similar temporal expression patterns, with consistent upregulation from T0 to T2. At T2, *Os02g0611800* reaches a log2FC of 1.18, while *Os09g0422000* achieves a log2FC of 1.50, suggesting increased activity through these branches downwards along the pathway.

The expression of downstream genes involved in the transition from aldehydes to alcohols and CoA intermediates varies. *Os09g0400400* and *Os04g0229100*, both encoding cinnamyl alcohol dehydrogenases implicated in aldehyde conversion, showed significant downregulation at T0 and T1 (log2FC = -2.35 and -1.31, respectively), followed by notable recovery at T2 (log2FC = 1.23 and 1.78, respectively). Interestingly, *Os09g0399800*, encoding also a cinnamyl alcohol dehydrogenase and involved in the aldehyde-to-alcohol transition, displays a contrasting pattern of expression, remaining down-regulated at T0 and showing moderate recovery by T1, but without significant upregulation at T2.

A notable result was the strong upregulation of the R2R3-MYB transcription factor genes *Os02g0624300* (*MYB30*) and *Os05g0132700* (*MYB52*) at T2. As described in chapter I of this thesis, these two genes were previously identified as candidates in the GWAS of allelopathy in the collection of rice varieties (García-Romeral et al. 2024). Their expression profiles revealed minimal signal at T0, followed by downregulation at T1, and a marked increase at T2, reaching Log2FC values of 1.39 and 6.65, respectively.

Peroxidases are enzymes that participate in the polymerization of mognolignols in the synthesis of lignin. Among DEGs, 11 peroxidase-related were up-regulated in Ganao compared to Bomba at T1. Among them, *Os02g0240100* displayed notably upregulation with log2FC value of 2.92. Notably, the expression of all the detected peroxidase genes increased their expression levels at T2, with a remarkable increase of *Os04g0688600* and *Os06g0695300* reaching log2FC values of 6.99 and 9.08 respectively. In addition, *Os08g0205000*, a gene encoding a transferase domain-containing protein involved in the synthesis of p-hydroxyphenyl lignin, displayed was highly expressed through the process with a log2FC of 6.44 at T1 and 8.06 at T2.

Transcripts associated with Diterpenoid Pathway Regulation

The expression of genes involved in the diterpenoid biosynthesis pathway revealed differential expression patterns in genes responsible for key steps in the pathway. Notably, a strong down-regulation of several genes was observed in Ganao at T1, suggesting an early response to barnyardgrass exposure, particularly in the non-allelopathic variety Bomba. The gene expression data are presented in **Table 5**, and the corresponding metabolic pathway, including key enzymatic reactions, is illustrated in **Figure 10**. Oryzalexins and momilactones have previously been associated with allelopathy in rice (Kato-Noguchi and Peters 2013; Serra Serra et al. 2021; Zhao et al. 2018). We found DEGs involved in the oryzalexin D and E biosynthesis that were down-regulated in Ganao compared to Bomba during the barnyard interaction. This downregulation was already evident at T1 (**Table 5**) as it is the case of *Os02g0571100*, encoding ent-copalyl diphosphate synthase, showing

log2FC = -5.60, *Os02g0570400*, responsible for ent-cassa-12,15-diene production, with a strong downregulation, with log2FC = -7.51, and *Os02g0569400*, involved in oryzalexin D biosynthesis, displaying a log2FC = -5.54, and *Os06g0569500*, associated with oryzalexin E synthesis, with marked downregulation log2FC = -7.27. The down-regulation of these genes was also observed at T2 (**Table5**). DEGs involved in the synthesis of momilactones exhibited of lower levels of transcripts in Ganao, with a similar pattern of expression, particularly at T1. *Os04g0179700*, which encodes 9β-pimara-7,15-diene synthase, was strongly down-regulated at T1 (log2FC = -4.90), *Os04g0179200*, a key gene for momilactone A synthesis, also showed significant downregulation in Ganao at T1 (log2FC = -5.61).

An exception to this trend was observed in *Os11g0474800*, responsible for stemar-13-ene synthesis or Oryzalexin S. This gene was notably up-regulated in Ganao at T2 (log2FC = 10.28), suggesting that Ganao may shift towards producing other diterpenoid compounds, such as stemar-13-ene, later in the interaction with barndyardgrass, potentially as part of a late response.

Pathway Role	RAPDB ID	Activity	GT0-BT0 log2FC	GT0-BT0 adjPval	GT1-BT1 log2FC	GT1-BT1 adjPval	GT2-BT2 log2FC	GT2-BT2 adjPval
GGPP -> ent-Copapyl-PP	*Os02g0571100*	Ent-copalyl diphosphate synthase	19.09	3.89E-08	-5.60	3.19E-01	-1.46	8.12E-01
ent-Copapyl-PP -> ent-Cassa-12,15-diene	*Os02g0570400*	Diterpene cyclase, Ent-cassa-12,15-diene synthase	1.80	8.69E-01	-7.51	9.28E-04	-0.43	9.40E-01
ent-Cassa-12,15-diene -> ent-11B-hydroxy-cassa-12,15-diene ent-sandara-copimaradiene-3B-ol -> Oryzalexin D	*Os02g0569400*	Oryzalexin D synthase-like	1.07	1.00E+00	-5.54	3.19E-02	-8.86	6.12E-05
ent-sandara-copimaradiene -> ent-sandara-copimaradiene-3B-ol 9B-pimara-7,15-diene -> 9B-pimara-7,15-diene-3B-ol	*Os06g0569500*	Ent-kaurene oxidase	1.75	7.99E-01	-7.27	2.30E-04	-0.33	9.53E-01
GGPP -> syn-Copapyl-PP	*Os04g0178300 (OsCPS4)*	Syn-copalyl diphosphate synthase-like	0.86	5.64E-01	-9.36	2.87E-14	0.15	9.25E-01
syn-Copapyl-PP -> 9B-pimara-7,15-diene	*Os04g0179700 (OsKSL4)*	9-beta-pimara-7,15-diene synthase	-0.15	1.00E+00	-4.90	2.00E-02	-3.68	6.93E-02
syn-Copapyl-PP -> syn-stemoden-19-oate 9B-pimara-7,15-diene -> 9B-pimara-7,15-19-oate	*Os04g0178400 (CYP99A2)*	Cytochrome P450 (CYP) mono-oxygenase	-3.15	9.57E-05	-5.42	1.51E-13	-1.80	2.95E-02
3B-Hydroxy-9B-pimara-7,15d-iene-19,6B-olide -> Momilactone A	*Os04g0179200 (OsMAS)*	Similar to Stem secoisolariciresinol dehydrogenase	-0.99	1.00E+00	-5.61	1.99E-01	-0.94	8.84E-01
syn-Copapyl-PP -> stemar-13-ene	*Os11g0474800*	Stemar-13-ene synthase. Biosynthesis of oryzalexin S	3.98	8.42E-02	1.40	7.63E-01	10.28	6.75E-07

Table 5. Log2 fold change (log2FC) and adjusted p-values (adjPval) for gene expression changes across three time points (T0, T1, T2) in the diterpenoid biosynthesis pathway. Positive log2FC values indicate upregulation in Ganao relative to Bomba, while negative values reflect downregulation

RT-qPCR for RNA-SEQ Data Validation

We selected 7 genes to confirm RNA-Seq data by quantitative real-time PCR (qRT-PCR) analysis (**Table 6**). These genes were involved in various biosynthetic pathways, particularly in phenylpropanoid, flavonoid, and diterpene biosynthesis. The level of expression of these genes were assessed at three time points (T0, T1, and T2). Similar expression patterns were observed in the qRT-PCR results and the RNA-Seq data (**Figure 11**).

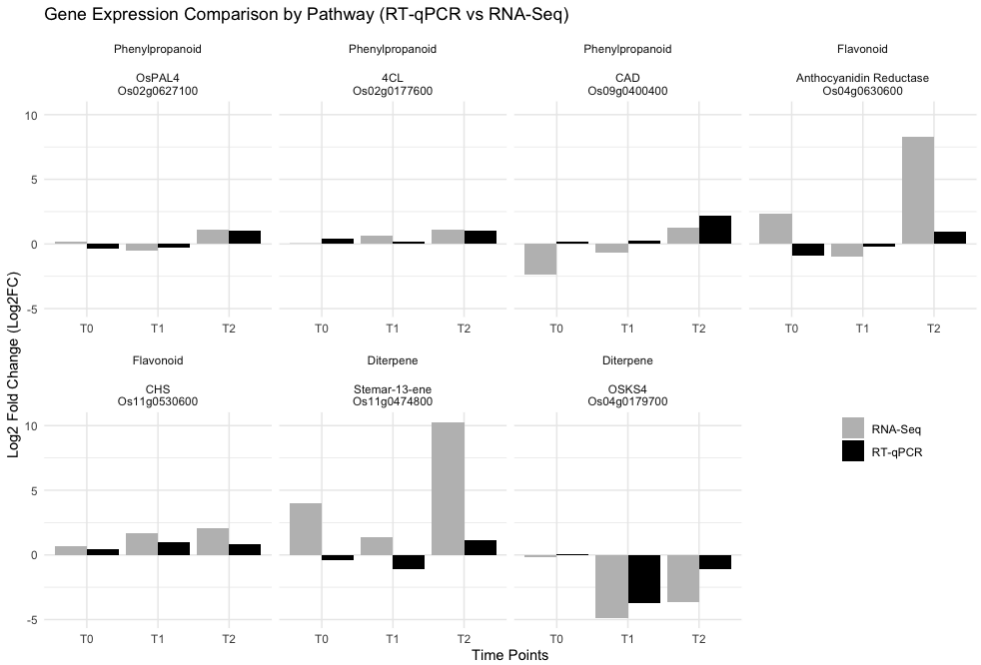


Figure 4. Log2 Fold Change in gene expression over three time points (T0, T1, T2) comparing two methods: RNA-Seq (grey) and RT-qPCR (black). The analysis covers genes involved in different metabolic pathways, including the phenylpropanoid pathway (OsPAL4, 4CL, CAD), flavonoid pathway (Anthocyanidin Reductase, CHS), and diterpene biosynthesis (Stemar-13-ene, OSK4).

Gene ID	Locus	Description	Pathway	Primer Sequence F/R (5'-3')	Tm (ºC)	Efficiency
OsPAL4	*Os02g0627100*	Phenylalanine ammonia-lyase	Phenylpropanoid biosynthesis	CATCGACCGTGCTCTTTGAG/TTCCCTCCAAGATGTGCTCC	56	2.150
4CL3	*Os02g0177600*	4-coumarate:coenzyme A ligase	Phenylpropanoid biosynthesis	TGGATGGAGAGAACCCTAACC/GCCAAGATCAAACTTCCGCA	56	2.332
CAD8D	*Os09g0400400*	Cinnamyl alcohol dehydrogenase	Phenylpropanoid biosynthesis	CCGACCTGCACATCATCAAG/TCGGGCAGTAGTTCTCGTAC	56	2.304
OsKSL8	*Os11g0474800*	Stemar-13-ene synthase	Diterpenoid, Oryzoalexin S biosynthesis	CGGGTGATAGTGAGGAAGCA/ATGATCCGTCCTCCTGTTGG	57	1.956
OsKS4	*Os04g0179700*	9-beta-pimara-7,15-diene synthase	Diterpenoid, Momilactone biosynthesis	ATGCCACGCTGGAACCTCTA/GATAAGCCCTGGCATCAAAATT	56	1.880
OsSTA	*Os04g0630600*	Anthocyanidin reductase ((2S)-flavan-3-ol-forming)	Flavonoid biosynthesis	CGACGTTGATGGTGGCAGAA/GGCTTCTCCGTTCTCAGGTAG	58	2.111
CHS	*Os11g0530600*	Chalcone synthase	Flavonoid biosynthesis	GGGCTCATCTCGAAGAACAT/CCTCATCCTCTCCTTGTCCA	55	2.174
UFD	*Os03g0234200*	Ubiquitin fusion protein	Constitutive gene	GCTCCGTGGCGGTATCAT/CGGCAGTTGACAGCCCTAG	57	1.831

Table 6. List of rice genes selected for qRT-PCR validation, including gene ID, locus, functional description, associated metabolic pathway, primer sequences (forward/reverse, 5'-3'), annealing temperature (Tm, °C), and amplification efficiency. Pathways include key branches of secondary metabolism (phenylpropanoids, flavonoids, and diterpenoids), as well as a constitutive gene used as reference (UFD).

Discussion

The importance of finding an effective and sustainable method of controlling weeds in rice fields has led to an increased interest in allelopathy. The nature of allelochemical compounds has been extensively investigated, and several metabolites have been reported as allelochemicals to act as natural herbicides by inhibiting the growth of neighboring plants. In this study, the transcriptomic profiles of an allelopathic rice variety, Ganao, and non-allelopathic, Bomba, were compared to understand the molecular mechanisms underlying the allelopathic interaction between rice and barnyardgrass. The allelopathic potential of Ganao was previously described in a search for varieties with the ability to inhibit barnyardgrass roots at the seedling stage within a temperate *japonica* rice collection (García-Romeral et al 2024). Here, we compared the transcriptomic changes occurring between both Ganao and Bomba in the interaction with barnyard after 6h and 36h of co-cultivation with barnyardgrass under controlled conditions in growth chambers. Ganao displayed a slow but strong response, with the number of DEGs reaching a maximum at 36 hours. A total of 4,719 DEGs were identified, similarly to other studies on allelopathic plants (Sultana et al. 2023). Transcriptomic analysis revealed significant differences in gene expression and pathways activation during co-cultivation with barnyardgrass, particularly with regard to secondary metabolism, defense-related processes, and oxidative stress responses.

The phenylpropanoid pathway entails the synthesis of numerous secondary metabolites as phenolics, flavonoids, coumarins, and lignans that have been reported as allelochemicals (Kostina-Bednarz et al. 2023; Li et al. 2010). In our system, we could observe that, 6h after being in co-culture, the barnyardgrass triggers a response in Ganao that includes the activation of genes involved in the phenylpropanoid biosynthetic pathway, among others. This response continued with a higher increase in transcripts after 36h, suggesting that prolonged exposure beyond 6 hours is required for the complete activation of this metabolic pathway. This supports the hypothesis that allelopathic responses in rice are dynamic and cumulative (Kato-Noguchi 2011). The KEGG transcriptomic analysis revealed the up-regulation of *PAL* at T2 in the allelopathic Ganao compared to non-alleopathic Bomba. This is in agreement with previous association analysis in a *japonica* rice collection where QTL, qAll2b, that included four genes from the multigenic family of *PAL*s (García-

Romeral et al. 2024) and reinforces the idea that *PAL* is a key enzyme in rice allelopathy. qAll2b also included *OsMYB30*, a R2R3-MYB transcription factor, that is up-regulated at T0 and T2 in the RNAseq analysis. The implication of some MYB transcription factors in the plant phenylpropanoid metabolic pathway has been reported previously (Cao et al. 2020). More specifically, R2R3-MYB factors are known to govern genes related to secondary metabolism, as this is the case of *OsMYB57*, that modulates rice allelopathy by the regulation of the transcription of *MAPK11*, which interacts with *PAL2;3* (Fang et al 2020). Another R2R3-MYB transcription factor, *OsMYB52*, was located in qAll5 (García-Romeral et al 2024) and it is also up-regulated at T2 in the allelopathic variety compared to non-allelopathic variety.

We could observe the increase in DEG implicated in lignin biosynthesis in Ganao at T2. Lignin is an essential component of the plant cell wall that gives shape and strength to the cell. Lignin is synthesized during plant growth and, also, as a defense mechanism in response to biotic stress reinforcing the cell wall to protect the plant from a biotic threaten (Xie et al. 2018). We cannot discard the possibility that the production of lignin in Ganao may be part of the chemical inhibition of allelopathic mechanism as a structural fortification to withstand the presence of barnyardgrass. The expression of lignin biosynthesis genes, particularly at T2, matches with an increasing demand for lignin production in response to developmental cues or biotic stress. This is essential for forming structural and defensive polymers in the cell wall. This dual strategy of chemical and structural defense highlights the complexity of allelopathy.

A transient upregulation of DEGS in the pathways involved in nitrate assimilation and nitrate transport in Ganao was observed after 6h of coculture with barnyardgrass. This contrasts with previous results in which the nitrogen metabolism pathway was inhibited in PI312777 after 3 days of growing in the presence of barnyardgrass, but no difference was seen with control at 3h (Sultana et al. 2023).

Interestingly, the comparison between both varieties indicates that Ganao, the allelopathic variety, has the defence mechanism activated during the interaction but transiently depleted at T1, in the presence of barnyardgrass. This suggests that Bomba may perceive barnyardgrass as a threat, while it is not perceived as such by Ganao. By T2, the continued activation of detoxification and cell wall organization

pathways shows a sustained defense response, crucial for long-term adaptation to competition.

The enzyme activation patterns (**Suppl. Figure 2**) show baseline expression at T0, with Ganao exhibiting significant upregulation of key enzyme families at T1 and T2. This is the case of cytochrome P450 enzymes, which are essential for producing secondary metabolites. CYP73, CYP84 and CYP98, among others, are known to participate in the phenylpropanoid pathway (Chakraborty et al. 2023; Mizutani 2012). CYPs are also involved in detoxifying xenobiotics, synthesizing hormones for growth and stress responses, and modulating nutrient uptake (Chakraborty et al. 2023; Pandian et al. 2020). Other group of enzymes function as UDP-glucosyltransferases (UDPs), which enhance metabolite solubility and bioactivity through glycosylation and contribute to cell wall integrity (Dong, Nai Qian et al. 2020). Glutathione S-transferases (GSTs) manage oxidative stress by conjugating glutathione to various compounds, aiding in resilience under stress (Edwards and Dixon 2005).

KEGG pathway analyses provided a deeper knowledge of which processes are activated in response to barnyardgrass. At T1, after 6h of co-culture with barnyardgrass, several pathways were significantly up-regulated in Ganao, such as nitrogen metabolism, with a Fold Change = 12.7, reflecting an enhanced capacity to assimilate nitrogen in the presence of barnyardgrass, and phenylpropanoid biosynthesis, with a Fold Change = 4.2, reinforcing its role in allelopathy. A significant increase in the biosynthesis of secondary metabolites could be observed. These findings suggest that Ganao responds to the presence of barnyardgrass by up-regulating key metabolic pathways, particularly those involved in nitrogen metabolism and secondary metabolite production. In contrast, the only significant downregulation observed at this time point was in diterpenoid biosynthesis with a Fold Change = 10.1. This unexpected trend suggests a divergence from the commonly accepted role of diterpenoids in allelopathy, indicating that their involvement in the interactions may be more complex than previously understood.

Momilactones and other diterpenoid phytoalexins have been identified in previous studies as key allelopathic compounds that inhibit the growth of different species of weeds (Kato-Noguchi and Peters 2013; Serra Serra et al. 2021; Zhao et al. 2018). Momilactone A has been identified as the main compound responsible for the

allelopathic potential in the PI312777 accession, widely studied as a reference of allelopathic variety (Kong et al. 2004). Nonetheless momilactone A was not detected in the previous GWAS, using the rice collection, as part of the response in temperate *japonica* rice. This observation is in agreement with the transcriptomic analysis where, interestingly, the diterpenoid biosynthesis pathway branch leading to the synthesis of momilactones was not activated in Ganao compared to the non-allelopathic Bomba, even with a transient down-regulation at T1. A similar assertion was observed by Xuan et al., (2016) where the research found low correlation between the momilactones and weed tolerance in both *indica* and *japonica* rice.

Similarly, genes implicated in the synthesis of the terpenoid phytoalexins Oryzalexin D and E were not activated in Ganao compared to Bomba. This is not the case of oryzalexin S, synthesized via *Diterpene Cyclase 2* (*OsDTC2*), displaying stemar-13-ene synthase function. It has been shown that in the presence of chitin as an elicitor of a defense response in suspension-cultured rice cells, the level of *OsDTC2* mRNA began to increase 3h after the initiation of the treatment with the elicitor, reaching a maximum after 8h (Nemoto et al. 2004). In our system, the expression of *OsDTC2* was increased 3d after the barnyardgrass was co-cultured with allelopathic rice. The differential regulation suggests that, in Ganao, this alternative diterpenoid compound, rather than momilactones, may play a role in its prolonged allelopathic response.

The transcriptomic analysis suggests that the allelopathic potential of Ganao is based on a different mechanism than the synthesis of diterpene under the conditions tested. As an alternative, the phenylpropanoid pathway appears to adopt a main role in the inhibition of barnyardgrass. These outcomes reveal the potential for varied response strategies among different rice varieties.

The results of this study suggest new directions for future research. A comprehensive metabolomic analysis of key phenolic compounds, and the Oryzalexin S, is needed to clarify their roles in plant defense mechanisms against barnyardgrass. Furthermore, performing bioassays to evaluate the allelopathic effects of these compounds on *E. crus-galli* will provide valuable insights into their potential as tools for weed control. Additionally, exploring gene expression patterns in later developmental stages, beyond the seedling phase, could yield significant data regarding their regulatory functions throughout the plant life cycle. Finally,

extending these bioassays to field conditions will be crucial for assessing the practical applications of these findings in agricultural systems.

These findings not only enhance our understanding of the genetic basis of allelopathy in rice but also highlight potential metabolic pathways that could be leveraged in breeding programs to improve weed suppression.

Conclusions

This comparative transcriptomic analysis demonstrates that the strong allelopathic potential observed in the rice variety Ganao is driven by a coordinated and temporally progressive response, focused on the activation of the phenylpropanoid biosynthetic pathway, which becomes fully induced by 36 hours of co-cultivation. Key candidate genes previously identified by GWAS and described in Chapter I (García-Romeral et al. 2024), including *PAL* and the R2R3-type MYB transcription factors *OsMYB30* and *OsMYB52*, were strongly up-regulated. This response was accompanied by the activation of detoxification-related enzyme families such as cytochrome P450s, UGTs, and GSTs, as well as genes involved in cell wall reinforcement. In contrast to previously studied allelopathic lines, diterpenoid phytoalexin genes, including those for momilactone biosynthesis, were transiently or consistently down-regulated, highlighting a mechanistic divergence from the momilactone-centric model typified by PI312777. These findings indicate that Ganao suppresses barnyardgrass primarily through phenylpropanoid-derived metabolites, potentially strengthened by lignification. The candidate genes and pathways identified here provide a molecular framework for breeding or engineering rice cultivars with enhanced, sustainable weed-suppressive ability.

References

Asaduzzaman, M., Islam M. M., & Sultana S. (2010). Allelopathy and Allelochemicals in Rice Weed Management. Bangladesh Research Publications Journal 4(1), 1–14. http://www.bdresearchpublications.com/admin/journal/upload/09146/09146.pdf.

Bajwa, A., Jabran, K., Shahid, M., Ali, H. H., Chauhan, B. S., & Ehsanullah, E. (2015). Eco-biology and management of Echinochloa crus-galli. Elsevier In Crop Protection 75, 151–162. https://doi.org/10.1016/j.cropro.2015.06.001

Belz, R. G. (2007). Allelopathy in Crop/Weed Interactions — an Update. Pest Management Science 63(4): 308–326. https://doi.org/10.1002/ps.1320

Bouillant M.I., Jacoud C., Zanella I., Favre-Bonvin J., & Bally R. (1994). Identification of 5-(12-Heptadecenyl)-Resorcinol in Rice Root Exudates. Phytochemistry 35(3), 769–771. https://doi.org/10.1016/S0031-9422(00)90602-2

Cao, Y., Li, K., Li, Y., Zhao, X., & Wang, L. (2020). MYB transcription factors as regulators of secondary metabolism in plants. MDPI AG. In Biology 9(3), 61. https://doi.org/10.3390/biology9030061

Chakraborty, P., Ghosh, S., Chakraborty, S., & Singh, N. (2023). Cytochrome P450 Gene Families: Role in Plant Secondary Metabolites Production and Plant Defense. Journal of Xenobiotics 13(3): 402–23. https://doi.org/10.3390/jox13030026

Childs, N., & Jarrell, P. (2024). Rice outlook: September 2024 (ERS Situation and Outlook RCS-24H). USDA-ERS.

Chung I. M., Kim K. H., Ahn J. K., Chun S. C., Kim C. S., Kim J. T., & Kim S. H. (2002). Screening of Allelochemicals on Barnyardgrass (Echinochloa Crus-Galli) and Identification of Potentially Allelopathic Compounds from Rice (Oryza Sativa) Variety Hull Extracts. Crop Protection 21(10): 913–20. https://doi.org/10.1016/S0261-2194(02)00063-7

Dilday, R. H., Lin, J., & Yan, W. (1994). Identification of Allelopathy in the USDA-ARS Rice Germplasm Collection. Australian Journal of Experimental Agriculture 34(7): 907–910. https://doi.org/10.1071/EA9940907

Dilday, R H, P Nastasi, and R J jr. Smith. (1989). Allelopathic Observations in Rice (Oryza Sativa L.) to Ducksalad (Heteranthera Limosa). Proceedings Arkansas Academy of Science 43: 21–22.

De La Peña, R., & Sattely, E. S. (2021). Rerouting plant terpene biosynthesis enables momilactone pathway elucidation. Nature Chemical Biology, 17(2), 205–212. https://doi.org/10.1038/s41589-020-00669-3

Dong, NQ., Sun, Y., Guo, T. et al. (2020). UDP-glucosyltransferase regulates grain size and abiotic stress tolerance associated with metabolic flux redirection in rice. Nat Commun 11, 2629. https://doi.org/10.1038/s41467-020-16403-5

Ebana, K., Yan, W., Dilday, R. H., Namai, H., & Okuno, K. (2001). Analysis of QTL Associated with the Allelopathic Effect of Rice Using Water-Soluble Extracts. Breeding Science 51: 47–51. https://doi.org/10.1270/jsbbs.51.47

Edwards, R., & Dixon, D. P. (2005). Plant Glutathione Transferases. Methods in Enzymology 401: 169–86. https://doi.org/10.1016/S0076-6879(05)01011-6

Fang, C., Zhang, P., Li, Y., Li, L., Yang, L., & Lin, W. (2013). Changes in Rice Allelopathy and Rhizosphere Microflora by Inhibiting Rice Phenylalanine Ammonia-Lyase Gene Expression. Journal of Chemical Ecology 39(2): 204–12. https://doi.org/10.1007/s10886-013-0249-4

Fang, C., Li, Y., Li, C., Li, B., Ren, Y., Zheng, H., & Lin, W. (2015). Identification and comparative analysis of microRNAs in barnyardgrass (Echinochloa crus-galli) in response to rice allelopathy. Plant Cell Environ. 38, 1368–1381. https://doi.org/10.1111/pce.12492

Fang, C., Yang, L., Chen, W., Li, L., Zhang, P., Li, Y., He, H., & Lin, W. (2020). MYB57 transcriptionally regulates MAPK11 to interact with PAL2;3 and modulate rice allelopathy. Journal of Experimental Botany, 71(6), 2127–2141. https://doi.org/10.1093/jxb/erz540

Freitas, C. D. T., Gonçalves, J. D. O., Costa, K. A. D., & Orellana, S. C. (2024). Class III Plant Peroxidases: From Classification to Physiological Functions. International Journal of Biological Macromolecules 263, 130306. https://doi.org/10.1016/j.ijbiomac.2024.130306

García-Romeral, J., Castanera, R., Casacuberta, J., & Domingo, C. (2024). Deciphering the Genetic Basis of Allelopathy in *Japonica* Rice Cultivated in Temperate Regions Using a Genome-Wide Association Study. Rice 17(1). https://doi.org/10.1186/s12284-024-00701-3.

Jensen, L. B., Courtois, B., Shen, L., Li, Z., & Dilday, R. H. (2001). Locating Genes Controlling Allelopathic Effects against Barnyardgrass in Upland Rice. Agronomy Journal 93(1): 21–26. https://doi.org/10.2134/agronj2001.93121x

Kato-Noguchi, H. & Ino, T. (2003). Rice Seedlings Release Momilactone b into the Environment. Phytochemistry 63(5): 551–554. https://doi.org/10.1016/S0031-9422(03)00194-8

Kato-Noguchi, H. & Ino, T. (2005). Possible Involvement of Momilactone B in Rice Allelopathy. Journal of Plant Physiology 162(6): 718–21. https://doi.org/10.1016/j.jplph.2004.11.009

Kato-Noguchi, H. (2011). Barnyard Grass-Induced Rice Allelopathy and Momilactone B. Journal of plant physiology 168(10): 1016–20. https://doi.org/10.1016/j.jplph.2010.12.021

Kato-Noguchi, H., & Peters, R. J. (2013). The Role of Momilactones in Rice Allelopathy. Journal of Chemical Ecology 39(2): 175–85. https://doi.org/10.1007/s10886-013-0236-9

Khanh, T. D., Xuan, T. D., & Chung, I. M. (2007). Rice Allelopathy and the Possibility for Weed Management. Annals of Applied Biology 151(3): 325–39. https://doi.org/10.1111/j.1744-7348.2007.00183.x

Kong, C., Hu, F., Zhang, C., Shen, X., & Zhang, P. (2004). Release and Activity of Allelochemicals from Allelopathic Rice Seedlings. Journal of Agricultural and Food Chemistry 52(10): 2861–65. https://doi.org/10.1021/jf035467i

Kostina-Bednarz, M., Płonka, J., & Barchanska, H. (2023). 22 Reviews in Environmental Science and Biotechnology Allelopathy as a Source of Bioherbicides: Challenges and Prospects for Sustainable Agriculture. Springer Netherlands 22, 471–504. https://doi.org/10.1007/s11157-023-09656-1

Li, Z. H., Wang, Q., Ruan, X., Pan, C. D., & Jiang, D. A. (2010). Phenolics and Plant Allelopathy. Molecules 15(12): 8933–52. https://doi.org/10.3390/molecules15128933

Li, Y., Fang, C., Li, C., Ren, Y., & Lin, W. (2020). OsPAL2-1 Mediates Allelopathic Interactions Between Rice and Specific Microorganisms in the Rhizosphere Ecosystem. Frontiers in Microbiology 11(July): 1–16. https://doi.org/10.3389/fmicb.2020.01117

Mizutani, M. (2012). Impacts of Diversification of Cytochrome P450 on Plant Metabolism. Biological and Pharmaceutical Bulletin 35(6): 824–32. https://doi.org/10.1248/bpb.35.824

Mu, H., Li, Y., Yuan, L., Jiang, J., Wei, Y., Duan, W., Fan, P., Li, S., Liang, Z., & Wang, L. (2023). MYB30 and MYB14 form a repressor-activator module with WRKY8 that controls stilbene biosynthesis in grapevine. Plant Cell, 35(1), 552-573. https://doi.org/10.1093/plcell/koac308

Nemoto, T., Nishikiori, T., Yamashita, T., Yamada, Y., & Minami, E. (2004). Stemar-13-Ene Synthase, a Diterpene Cyclase Involved in the Biosynthesis of the Phytoalexin Oryzalexin S in Rice. FEBS Letters 571(1–3): 182–86. https://doi.org/10.1016/j.febslet.2004.07.002

Okuno, K., & Ebana, K. (2003). Identification of QTL Controlling Allelopathic Effects in Rice: Genetic Approaches to Biological Control of Weeds. Jarq-japan Agricultural Research Quarterly 37: 77–81. https://api.semanticscholar.org/CorpusID:84216589.

Olofsdotter, M., Navarez, D. & Moody, K. (1995). Allelopathy Potential in Rice (Oryza Sativa L.) Germplasm. Ann Appl Biol 127(3): 543–60. https://doi.org/10.1111/j.1744-7348.1995.tb07611.x

Olofsdotter, M., Navarez, D., & Rebulanan, M. (1997). Rice Allelopathy: Where Are We and How Far Can We Get? In Brighton Crop Protection Conference., Brighton, UK., 99–104.

Olofsdotter M. (1998). Allelopathy in Rice. Manila, Philippines.

Olofsdotter, M., Rebulanan, M., Madrid, A., Dali, W., Navarez, D., & Olk, D. C. (2002). Why Phenolic Acids are Unlikely Primary Allelochemicals in Rice. In Journal of Chemical Ecology 28(1), 229–242. https://doi.org/10.1023/A:1013531306670

Pandian, B. A., Sathishraj, R., Djanaguiraman, M., Prasad, P. V. V., & Pandian, S. R. K. C. (2020). Role of Cytochrome P450 Enzymes in Plant Stress Response. Antioxidants 9(5): 1–15. https://doi.org/10.3390/antiox9050454

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. In Nucleic Acids Research 29(9), e45. https://doi.org/10.1093/nar/29.9.e45

Rice, E. L. (1984). Allelopathy. 2nd editio. ed. Orlando: Academic.

Serra Serra, N., N., Shanmuganathan, R., & Becker, C. (2021). Allelopathy in Rice: A Story of Momilactones, Kin Recognition, and Weed Management. Journal of Experimental Botany 72(11): 4022–37. https://doi.org/10.1093/jxb/erab084

Sultana, M. H., Fang, C., Zhang, P., Li, L., Yang, L., Li, Y., & Lin, W. (2023). Transcriptomic Profiling Reveals Candidate Allelopathic Genes in Rice Responsible for Interactions with Barnyardgrass. Frontiers in Plant Science 14(February): 1–15. https://doi.org/10.3389/fpls.2023.1104951

Turra, G. M., Pinto, M. E., Panozzo, L. E., Agostinetto, D., & Rizzardi, M. A. (2023). Variability to Flooding Tolerance in Barnyardgrass and Early Flooding Benefits on Weed Management and Rice Grain Yield. Field Crops Research 300, 108999. https://doi.org/10.1016/j.fcr.2023.108999

Xiao, Z., Zou, T., Lu, S., & Xu, Z. (2020). Soil Microorganisms Interacting with Residue-Derived Allelochemicals Effects on Seed Germination. Saudi Journal of Biological Sciences 27(4): 1057–65. https://doi.org/10.1016/j.sjbs.2020.01.013.

Xie, M., Zhang, L., Wang, X., Zhou, Y., & Zhang, M. (2018). Regulation of Lignin Biosynthesis and Its Role in Growth-Defense Tradeoffs." Frontiers in Plant Science 9, 1427. https://doi.org/10.3389/fpls.2018.01427

Xu, M., Galhano, R., Wiemann, P., Bueno, E., Tiernan, M., Wu, W., Chung, I. M., Gershenzon, J., Tudzynski, B., Sesma, A., & Peters, R. J. (2012). Genetic evidence for natural product-mediated plant-plant allelopathy in rice (Oryza sativa). New Phytologist, 193(3), 570–575. https://doi.org/10.1111/j.1469-8137.2011.04005.x

Xuan, T. D., Minh, T. N., Anh, L. H., & Khanh, T. D. (2016). Allelopathic Momilactones A and B Are Implied in Rice Drought and Salinity Tolerance, Not Weed Resistance. Agronomy for Sustainable Development 36(3). http://dx.doi.org/10.1007/s13593-016-0383-9.

Zhang, Q., Zheng, X. Y., Lin, S. X., Gu, C. Z., Li, L., Li, J. Y., Fang, C. X., & He, H. B. (2019). Transcriptome analysis reveals that barnyard grass exudates increase the allelopathic potential of allelopathic and non-allelopathic rice (Oryza sativa) accessions. Rice, 12(1), 1–14. https://doi.org/10.1186/s12284-019-0290-1

Zhang, Z., Wu, H., Chen, J., Wu, W., Wang, Y., & Liu, K. (2021). Co-Planted Barnyardgrass Reduces Rice Yield by Inhibiting Plant above- and Belowground-Growth during Post-Heading Stages. Crop Journal 9(5): 1198–1207. https://doi.org/10.1016/j.cj.2020.10.011.

Zhang, H., Li, Y., Wang, J., Fang, C., Yang, L., Li, Y., & Lin, W. (2024). Glutathione S-Transferase Activity Facilitates Rice Tolerance to the Barnyard Grass Root Exudate DIMBOA. BMC Plant Biology 24(1): 1–16. https://doi.org/10.1186/s12870-024-04802-5

Zhao, M., Reddy, G. V. P., Ridgway, R. L., & Wang, Q. (2018). Momilactone and Related Diterpenoids as Potential Agricultural Chemicals. Journal of Agricultural and Food Chemistry 66(30): 7859–72. https://doi.org/10.1021/acs.jafc.8b02602

Chapter III

Development of an Eight-Parent MAGIC Rice Population Identifies QTLs Controlling Plant Architecture and Phenology

Julia García-Romeral and Concha Domingo

Departamento del Arroz, Centro de Genómica, Instituto Valenciano de Investigaciones Agrarias (IVIA)

Manuscript prepared for submission

Abstract

Multi-parent Advanced Generation Inter-Cross (MAGIC) populations represent a powerful genetic resource to identify the genomic basis of complex agronomic traits. Rice breeding for enhanced plant architecture and optimal phenology traits is crucial to improve yield and adaptability in temperate *japonica* varieties. We characterized an eight-parent MAGIC rice population at genomic and phenotypic level. Employing Genome-Wide Association Studies (GWAS) with 70,289 SNPs across 192 MAGIC lines, we identified 35 significant Quantitative Trait Loci (QTLs) linked to plant height, panicle length, awning, lemma pubescence, tiller and flag leaf angle, heading date, and grain-filling duration. Prominent genomic regions included well-characterized candidate genes such as *Semidwarf 1* (*SD1*) and *Semidwarf 5* (*SDF5*), associated with plant height regulation; *Regulator of Awn Elongation 2* (*RAE2*) and *Glabrous Leaf and Hull 1* (*GL1*), governing awning and lemma pubescence; and genes controlling architectural traits such as *Tiller Angle Control 1* (*TAC1*) and *Transport Inhibitor Response 1* (*TIR1*). Phenological QTLs harbored crucial flowering genes (*EHD4, MADS50, ELF4B*, and *Ghd8*) and grain-filling regulators (*OsGW10, OsAUX5*, and *OsXLG2*). Notably, strong signals were detected for panicle length (qPL9) and plant height (qH9) on chromosome 9, each explaining over 20% of phenotypic variance. The identification of these QTLs provides a robust foundation for breeding strategies aiming to optimize rice architecture and phenology in temperate *japonica* rice.

Introduction

Rice breeding dates back to the origins of agriculture, when early farmers selected plants based on visible traits. Over time, this process led to crops that were better adapted to local agro-climatic conditions and produced higher yields. Since those early days of visual selection, breeding methods have evolved significantly. Today, rice breeding relies on genomic and molecular techniques to guide improvements more precisely. The sequencing of the rice genome and the emergence of new genomic and bioinformatic tools have dramatically expanded our understanding of the genetic basis of agronomic traits and enable the identification of QTLs and genes, and their favourable haplotypes, associated with important agronomic traits.

Breeding relies on available genetic resources, which offers genetic variability. Traditionally, researchers generated populations from crosses between two parental lines that differed in a specific trait. These bi-parental populations, such as recombinant inbred lines (RILs), produce only two segregating haplotypes at each locus. While they have proven useful for detecting causal loci, their design limits the scope to analysing one or a few traits at a time. Germplasm collections provide access to much broader genetic diversity, but they often exhibit strong population structure. In GWAS using such panels, population structure must be carefully controlled to minimise false positives while maintaining sufficient statistical power to detect true associations. As an alternative to germplasm collections, MAGIC populations have emerged as a useful genetic resource. By combining multiple parental lines and increasing recombination, MAGIC populations enable high-resolution mapping across multiple alleles and help reduce confounding population structure. First developed in mice (Churchill et al., 2004), MAGIC populations were later introduced in plants such as *Arabidopsis thaliana* (Kover et al., 2009), and subsequently adapted for crops, beginning with four-way crosses in wheat (Huang et al., 2012). In rice, more than a dozen MAGIC populations have been developed to date (Bandillo et al., 2013; Li et al., 2013; Meng et al., 2016; Ogawa et al., 2018; Bossa-Castro et al., 2018; Descalsota-Empleo et al., 2022). The selection of founders is also crucial. Parental lines should carry desirable agronomic traits, such as resistance to biotic and abiotic stresses, favourable grain quality, optimal phenology, high yield potential, or ideal plant architecture.

One of the most widely used approaches to identify QTLs is genome-wide association analysis (GWAS). Achieving high mapping resolution is essential for effective GWAS and depends on a dense, genome-wide distribution of SNP markers. This has become feasible due to next-generation sequencing (NGS) technologies, which now offer cost-effective, high-throughput, and near-complete genotyping solutions across the genome. GWAS on temperate *japonica* rice have been previously performed (Reig-Valiente et al. 2018).

Yield is a multifactorial trait determined by its components and modulated by genetic and environmental interactions. Plant architecture determines in a large extent yield in rice. An appropriate plant height (Ht) in rice enhances yield, harvest uniformity, lodging resistance and tolerance to dense planting (F. Liu, et al. 2018; Cheng et al. 2022). Ht is primarily determined by internode elongation, which is regulated by gibberellins through the stimulation of cell division and expansion (Nagai et al., 2020). Mutations affecting gibberellins biosynthesis result in a range of dwarf phenotypes. Mutations in *Semidwarf 1* (*SD1*), which encodes a gibberellin 20 oxidase, produce semi-dwarf plants while retain fertility (Sakamoto et al., 2004). The identification of the *SD1* enabled the development of the high-yielding variety IR8, which played a central role in the Green Revolution. SD1 remains the only GA-related gene widely used in breeding programs (Cheng et al. 2022; Sha, H. et al. 2022). However, the exclusive dependence on a single sd1 allele may compromise genetic diversity and increase susceptibility to biotic stress (Kadambari et al., 2018). Therefore, identifying novel alleles or combinations that contribute to Ht variation is crucial for sustainable breeding.

Bristle-like structures at the tip of the lemma, called awns, are common in wild rice, where they promote seed dispersal and avoid predation. In cultivated rice, however, awnless plants have been widely selected for its practical benefits, as awns complicate manual harvesting and reduce post-harvest efficiency. Awns have also been linked to reduced yield (Toriba et al., 2014). Consequently, most modern cultivars are awnless, and this trait is considered a key feature of domestication and agronomic optimisation (Furuta et al., 2015; Amarasinghe et al., 2020; Luong et al., 2022). Molecular studies have revealed that the transition to awnless rice has occurred through the selection on several key genes, including *An-1/RAE1*, corresponding to the transcription factor, *LABA1/An-2*, encoding a cytokinin-activating enzyme, and *RAE2/GAD1/GLA*, coding for the Epidermal Patterning

Factor-Like protein 1 (Luo et al., 2013; Gu et al., 2015; Hua et al., 2015; Yano et al., 2016). Of them, *RAE2/GAD1/GLA* is exhibited in many modern awnless varieties (Bessho-Uehara, K. et al. 2016).

In rice, trichomes may appear on both leaves and grains. These structures, derived from aerial epidermal cells, serve various protective functions, such as resistance to herbivorous insects, tolerance to freezing, and protection from UV radiation (Ishida et al., 2008). However, the molecular mechanisms underlying trichome development have been extensively studied only in Arabidopsis, where the corresponding to a homeodomain leucine zipper protein (*GL2*), and an R3MYB protein (*TRY*) play essential roles in trichome initiation and the differentiation of hairless cells (Rerie et al. 1994; Schellmann, S. et al. (2002). The expression of these genes is further regulated by a complex that includes *TTG1, GL3, EGL3*, and *GL1*, resulting in glabrous or pubescent grains (Payne et al. 2000). In rice, however, only one gene, the *Glabrous Rice 1* (*GLR1*), also known as *WUSCHEL-like homeobox* (*OsWOX3B*) has been reported as essential for trichome development (Li et al. 2012).

Erectness is a key component of ideal plant architecture that is desired for leaf flags, culms and tillers. These traits have been a breeding target for decades. Flag leaf angle is determined by the structure of the lamina joint, where factors such as the relative elongation of abaxial and adaxial cells, the development of mechanical tissues, and the composition of the cell wall all play essential roles (Ning et al., 2011). Leaf angle refers to the inclination between the leaf blade and the stem. The flag leaf, which is the uppermost leaf closest to the panicle, has the highest photosynthetic capacity and plays a major role in grain productivity, as it is the main source of carbohydrates that accumulate in the grains (Chen et al., 2007; Dong et al., 2018). Erect flag leaves can perceive light more efficiently than other plants with leaves with an angle and plants with erect leaves can be grown at higher densities without reducing photosynthesis per leaf, which ultimately increases grain yield. Leaf angle results from an imbalance in cell division and expansion on the two sides of the lamina. This process is regulated by phytohormones, particularly brassinosteroids, gibberellins, and auxins, where brassinosteroids signalling plays a central role in promoting cell division and elongation on the adaxial side of the lamina joint.

Culm and tiller angle refers to the inclination formed between the main stem and the vertical axis. The erect phenotype improves the plant's ability to acquire light, water,

and nutrients. In addition, spreading tillers occupy more space and are more prone to lodging, resulting in yield losses. Several key genes have been implicated in the transition from prostrate to erect growth during rice domestication. Among them, the *Rice Plant Architecture Domestication* (*RPAD*) locus, which includes *Prostate Growth 1* (*PROG1*) and *Semi-Prostrate Growth 1* (*SPROG1*), as well as *Tiller Angle Control 1* (*TAC1*) and *Tiller Inclined Growth 1* (*TIG1*), played central roles in regulating tiller angle. Notably, *TAC1* orthologues have been identified in several other plant species, where they also regulate lateral branch angle (Ku et al., 2011; Dardick et al., 2013; Zhao et al., 2014; Kangben, F. et al. 2023).

An optimal heading date (Hd) is necessary for the adaptation of cultivars to specific agro-ecological conditions and to reach a high yield. Hd is governed mainly by the local interaction of temperature and day-length characteristics of each climate zone. Flowering time marks the transition from vegetative to reproductive growth and it is defined as the number of days from sowing to the emergence of the panicle from the flag leaf. This period is critical, as it determines the duration available for photosynthate accumulation, influencing the conditions for grain development, and ultimately affecting both yield and grain quality. Winter cold temperatures in temperate climate regions constrain rice cultivation. But shortening the growth duration in a few days is desirable by farmers as it increases crop security by reducing the risk of pathogen attacks or weather adverse conditions like storms at the end of the growing season.

The genetic regulation of flowering has been extensively studied, with the main regulatory genes being characterized (Matsubara & Yano, 2018; Zhou et al., 2021). Flowering is regulated by photoperiod through a fine-tuning mechanism with *Heading date 3a* (*Hd3a*) and *Rice Flowering Locus T 1* (*RFT1*) acting as master genes to promote flowering under short or long day conditions. *Hd3a* is the primary floral activator under short day conditions, while RFT1 promotes flowering under long day conditions. The expression of *Hd3a* and *RFT1* is mainly modulated by two genes that represent two regulatory pathways: *Early heading date 1* (*Ehd1*), which acts as an integrator of different signals and *Heading date 1* (*Hd1*), which is regulated by the circadian cycle. *Hd1* inhibits flowering under LD conditions. Studies of diversity in *HD1* have shown that non-functional alleles are prevalent among varieties cultivated in northern latitudes suggesting that the lack of *HD1* function is one of the mechanisms by which rice has been adapted to regions with temperate

climate (Naranjo et al 2014). Other key regulatory genes include Ghd7, *DTH8*, *DTH3* (*OsMADS50*) and *PRR37*, that participate in the modulation of *Hd3a* and *RFT1* (Xue et al., 2008; Itoh et al., 2010; Zhou et al., 2021; Chen et al., 2022).

Grain filling followed by maturation is a critical period in which grain weight, yield, and quality is determined. An extended grain filling period (GF) allows for greater starch accumulation, leading to heavier grains and increased yield, whereas a shorter duration may reduce grain weight and increase chalkiness. This trait is genetically complex, controlled by multiple loci involved in starch metabolism, hormone signalling, and stress responses (Liu et al., 2025; Ruan et al., 2021). Grain size also influences GF, with genes such *TGW6* (Akabane et al., 2025), *GFD1* (Sun et al., 2023), *OsAUX5* (Shi et al., 2023), *OsXLG2* (Biswal et al., 2022), and *OsGW10* (Zhan et al., 2021) affecting both grain development and the duration of the filling phase. Additionally, flowering time regulators such as Ghd7 and Ghd8 (Xue et al. 2008; Yan et al. 2011) exhibit pleiotropic effects, delaying Hd while extending the GF period.

Panicle length (PL) influences yield and plant architecture, as it determines the number of spikelets and, consequently, the number of grains that a panicle can produce (Bai et al. 2021). Generally, longer panicles accommodate more spikelets, leading to higher grain yield per panicle, desirable for farmers. However, excessively long panicles may produce incomplete GF or increased risk of lodging. Therefore, breeders aim to optimize PL that maximizes grain number without compromising plant stability. Notable genes associated with PL include *OsPRA2*, involved in panicle architecture; *LP1* (Liu et al. 2016), which promotes cell elongation in the rachis; and *Gn1a* (*OsCKX2*) (Ashikari et al., 2005), which increases cytokinin accumulation and enhances panicle branching. Additionally, genes like *OsSPL18* (Yuan, H. et al., 2019) and *OsRA2* (Lu et al., 2017) have been shown to affect rachis development and panicle elongation.

To support breeding, it is essential to dissect the genetic basis of these traits adapted to our agroclimate conditions so they can be effectively incorporated into crop improvement programs.

In this study, we developed a MAGIC population derived from eight elite founders to investigate classical agronomic traits related to plant architecture including plant

Ht, flag leaf angle and culm angle (An), and PL; and phenology such as Hd and GF. To this end, we conducted GWAS using SNPs obtained through whole-genome sequencing of 192 MAGIC lines, to identify associated QTLs and candidate genes.

Materials and Methods

Plant materials. The MAGIC population development.

The MAGIC population was developed by inter-crossing eight temperate *japonica* varieties. The founders comprised 8 varieties known to exhibit different levels of yield potential, grain quality, cycle length and tolerance to a range of biotic and abiotic stresses (Reig-Valiente et al. 2016) (**Table 1**).

Variety	Origin	Agronomic relevance
Bomba	Spain	C + GR + PR + ST/D
Fonsa	Spain	C
Regina	Spain	GR + ST/D
Irat-307	French Guiana	GR
Loto	Italy	C
Lido	Italy	CT
Giza-181	Egypt	ST/D
Alena	Spain	PR

Table 1. Rice varieties, including their subspecies, geographic origin, and main agronomic traits of interest. Agronomic relevance codes: C = cycle duration, GR = grain quality, CT = cold tolerance, ST/D = salt and drought tolerance, PR = resistance to blast.

The population was developed following a similar scheme to that described by Bandillo et al (2013) but on a smaller scale and with some differences. Briefly, the first step involved inter-mating the eight founder varieties and resulting in 28 biparental crosses. In the second step, only 45 out of 70 possible 4-way crosses were made by inter-crossing the resulting 28 F1s. These 45 crosses were selected so that no parent variety was represented more than once in one 4-way cross. The final step involved the intercrossing of the 45 4-way crosses to derive 8-way crosses. Only 12 out of the 35 possible 8-way crosses were made with no parent represented more than once. From each of the 12 8-way crosses 35 seeds were advanced by selfing. Single plant selections were then made to advance to the next generation. The targeted population size was therefore 420 lines. The crossing program and early

generation advancement (up to the initial stages of selfing) were carried out by the Rice Genomics Department at IVIA.

Each 8-way cross was advanced through successive selfing generations, culminating in F6-8 progeny. This stabilization of the population, including the last two selfing generations, was completed by the author. Among the 420 recombinant inbred lines, 192 (including the parentals) were randomly selected for phenotypic and genotypic analysis.

Phenotyping for agronomical traits

The purpose of the phenotyping experiments was to evaluate the population for multiple desirable traits at a relatively stable generation (F6–F8) in plants grown in open-air pools. The traits analyzed were selected based on their agronomic relevance, and data collection was performed following the guidelines suggested by the Standard Evaluation System for Rice (SES) (IBPGR & IRRI, 1980) and the IRRI-descriptors for rice (IRRI, 1980).

Over three consecutive years (2021, 2022, and 2023), five plants of each line were grown in pools containing soil from the local fields consisting of 28.7% clay, 51.3% sand, and 20.0% silt. Data was collected from the three inner plants. During cultivation, measurements were taken for Plant Height (Ht), Flag Leaf Angle (FLA), Culm Angle (CMa), Days to Heading (Hd) and Grain Filling Period (GF). After harvesting, additional traits were recorded, including Panicle Length (PL), Awning (An), and Lemma and Palea Pubescence (LmPb).

Ht was measured from the soil surface to the tip of the tallest panicle (excluding awns). FLA was measured near the collar, defined as the angle of attachment between the flag leaf blade and the main panicle axis. This measurement was taken using a protractor and classified numerically by the degree of opening: 1 (erect, ~90°), 3 (intermediate, ~45°), 5 (horizontal, ~0°), and 7 (descending, ~-45°). Culm Angle (CMa) was classified using a similar scale, where 1 represents erect (<30°), 3

intermediate (~45°), 5 open (~60°), 7 spreading (>60°), and 9 procumbent (where the culm or its lower part rests on the ground surface).

Days to Hd were recorded as the number of days from the effective seeding date to the point when 50% of the plants exhibited heading. Maturity was noted as the number of days from seeding to 85% grain ripening. Additionally, the interval between Hd and maturity, the GF was calculated. PL was measured from the internode to the last grain, excluding any awns.

Boxplots for each trait were generated to evaluate differences in trait values across years, with statistical significance determined by Student's t-test using ggplot2 and geom_signif for numeric data, and Chi-square test was conducted using the chisq.test package for binary data.

Broad-sense heritability (H^2) was estimated for numerical, binary, and ordinal traits using mixed-effects models. For numerical traits, a linear mixed-effects model was applied with year as a fixed effect and varieties as a random effect. For binary traits, a generalized linear mixed-effects model with a logit link function was used. For ordinal traits, a cumulative link mixed model with a probit link was fitted. Genetic variance ($\sigma g2$) and residual variance ($\sigma e2$) were extracted from each model. Residual variance was approximated as $\pi 2/3$ for binary traits and set to 1 for ordinal traits. Total phenotypic variance ($\sigma p2$) was calculated as $\sigma g2 + \sigma e2$, and heritability was estimated as $H2 = \sigma g2 / \sigma e2$. All analyses were performed in R using the lme4, ordinal, and tidyverse packages.

Correlation between years for each trait under study was computed using the built-in COEF.DE.CORREL function in Excel.

Pairwise Pearson correlations between all traits were calculated using the cor function in R, with missing values handled via pairwise.complete.obs. The correlation matrix was plotted using the ggcorrplot package.

Histograms were generated using the ggplot2 package in R to visualize the distribution of each trait's mean across years.

Genotypic Analysis

DNA Extraction and Whole Genome Sequencing

The procedures for DNA extraction, library preparation, sequencing, alignment to the reference genome, and multi-sample variant calling were performed identically to those described in Chapter I. This process yielded 10.563 million variants, from which indels were removed, resulting in 9.702 million SNPs. After filtering using the same parameters as previously described, 70,289 high-quality SNPs were retained. Samples with more than 50% missing data (Ns) were excluded, leaving a total of 179 individuals for downstream analyses.

Population Genomic Analysis

For LD estimation the SNP data matrix was pruned with PLINK software (Purcell et al. 2007) for high levels of pairwise linkage disequilibrium (LD) using "--indep-pairwise 50 5 0.9" (window of 50 SNPs with a shift of 5 SNPs and 0.9 R2 cutoff). Finally, a 3.842 SNP data array was obtained for the 179 varieties.

The linkage disequilibrium (LD) decay rate was calculated using TASSEL v.5 software (Bradbury et al. 2007) with a sliding window of 50 SNPs, as the chromosomal distance where the Pearson's correlation coefficient (r2) between SNP pairs dropped to half its maximum estimated value. Data Plotting was performed with R software v.4.3.0 (R Core Team 2023).

The population genetic structure of the collection was assessed using the Bayesian clustering method implemented in STRUCTURE 2.3.4 (Pritchard et al. 2000). For that aim a pruned SNP data matrix with 3.842 SNPs from the 179 varieties was used. The software estimated the optimal number of genetic clusters (K) and calculated each variety's membership proportion. Analyses were based on the admixture ancestral model for a range of K values from 1 to 10. We performed 10 runs for each K. Each run was implemented with a burn-in period of 10,000 steps followed by 100,000 Monte Carlo Markov Chain iterations. The optimal number of K clusters was estimated with the ad hoc parameter (ΔK) of Evanno (Evanno et al. 2005) in StructureSelector (Li YL, Liu JX 2018).

Phylogenetic analysis was performed by converting the 70.289 SNP dataset in VCF format into a PHYLIP file using the vcf2phylip.py script (Ortiz 2019). To calculate pairwise genetic distances between rice cultivars a maximum likelihood (ML) method was applied by IQ-TREE v.2.1.2. (Minh et al. 2020) with 1000 bootstrap replicates and the GTR + ASC model, specific for SNP data. Data representation in tree format was displayed with iTOL v.6.8.1. (Letunic and Bork 2021).

Multidimensional Scaling Analysis was performed using TASSEL v.5 software (Bradbury et al. 2007). The plot was drawn with the "rgl" package (Murdoch and Adler 2023).

Genome Wide Association Study GWAS

Genome-wide association analyses were performed using TASSEL v.5 (Bradbury et al., 2007) based on a set of 70,289 high-quality SNPs and phenotypic data. A mixed linear model (MLM) was implemented to test for associations between genotypes and traits, incorporating both a kinship matrix (K), calculated using the centered identity-by-state (IBS) method to account for genetic relatedness, and a population structure matrix (Q), derived from Multidimensional Scaling (MDS) of the IBS distance matrix.

The GWAS included nine agronomic traits: Height, Awn, Heading, Maturity, Glabrous, Flag Leaf Angle, CmA, GF, and Panicle Length. Panicle Weight and Panicle Number were excluded due to significant phenotypic differences between years, which prevented the use of mean values across environments and could compromise the robustness of association results.

Significance thresholds were empirically determined based on the distribution and strength of association signals observed for each trait. A threshold of $-\log_{10}(p) = 3$ ($p \leq 0.001$) was used for Awn, Flag Leaf Angle, Heading, Maturity, and GF, while a more stringent threshold of $-\log_{10}(p) = 4$ ($p \leq 0.0001$) was applied for Height, Glabrous, CmA, and Panicle Length.

Manhattan plots were generated using the qqmant R package (Turner 2018), and phenotypic effects of the top-associated SNPs were visualized using ggstatsplot (Patil, 2021).

Candidate Gene Mapping

Significantly associated SNPs, those exceeding the defined threshold for each trait ($-\log_{10}(p) \geq 3$ or ≥ 4), were examined to identify putative candidate genes. Genomic regions surrounding each lead SNP were defined based on the estimated linkage disequilibrium (LD) decay distance (879 kb), and genes within these intervals were annotated using in-house R scripts with data retrieved from the Rice Annotation Project Database (RAP-DB). SNPs located within the same LD-defined interval were grouped as a single QTL, and the SNP with the lowest p-value within each region was designated as the lead SNP. Functional relevance of nearby genes was assessed based on annotated roles and previously reported functions in rice.

Results

Phenotypic Characterization of the MAGIC Population

Population Development and Overview of Agronomic Trait Evaluation

A MAGIC rice population was developed through the intercrossing of eight elite inbred founder lines, adapted to temperate climate (**Table 1**). Following the crossing methodology proposed by Bandillo et al. (2013), 420 recombinant inbred lines (RILs) were generated, from which 192 (parentals included) were randomly selected.

These lines were evaluated for key agronomic in pools during the crop season to analyze population variability and assess environmental stability across different years.

Traits studied included Ht, FLA, CMa, Hd and GF. After harvesting, additional traits were recorded, such as PL, An, and Lemma and LmPb.

The phenotypic analysis of the population lines revealed notable variation across all evaluated traits (**Table 2**).

Numerical traits showed considerable differences between maximum and minimum values, and all mean values remained highly consistent across years. However, the standard deviations were relatively high for most traits, reflecting significant variability within the population. Phenological variables were the steadiest between seasons. Heading occurred, on average, 85 days after sowing and ripening at 120 days, with coefficients of variation (CV) below 10 %. Structural traits, as plant Ht and PL were similarly consistent, with stable mean values, and moderate coefficients of variation (~ 15%).

A)

	2021					2022					2023					21-22-23				
Numeral Traits	**min**	**max**	**mean**	**ds**	**cv**	**min**	**max**	**mean**	**ds**	**cv**	**min**	**max**	**mean**	**ds**	**cv**	**min**	**max**	**mean**	**ds**	**cv**
Height (cm)	60.0	155.0	101.8	15.9	15.6	60.0	150.0	101.9	16.6	16.3	63.0	165.0	100.9	16.8	16.7	65.0	151.7	101.2	16.1	15.9
Heading (Days)	68.0	106.0	84.3	7.2	8.6	64.0	123.0	85.2	8.5	10.0	73.0	106.0	85.2	5.8	6.8	69.3	109.3	84.9	6.4	7.6
Grain Filling (Days)	17.0	58.0	36.5	6.7	18.3	19.0	54.0	35.6	6.8	19.1	22.0	48.0	35.4	5.6	15.8	19.3	53.3	35.9	6.4	17.7
Panicle Lenght (cm)	13.0	27.2	19.0	3.2	16.9	13.0	27.4	19.5	3.0	15.6	13.1	27.2	18.9	2.8	15.0	13.4	26.9	19.1	2.7	13.9

B)

	2021		2022		2023		21-22-23	
Binary Traits	**Presence**	**Absence**	**Presence**	**Absence**	**Presence**	**Absence**	**Presence**	**Absence**
Awn	74	118	69	123	63	129	74	120
Glabrous	67	125	66	126	67	125	67	125

C)

	2021		2022		2023		21-22-23	
Ordinal Traits	**mean**	**dv**	**mean**	**dv**	**mean**	**dv**	**mean**	**dv**
Flag Leaf Angle	2.61	1.42	3.04	1.71	2.89	1.41	2.84	1.28
Culm Angle	2.52	1.89	2.25	1.77	2.51	1.79	2.43	1.82

Table 2. Descriptive statistics for agronomic traits of the MAGIC population during three consecutive seasons (2021–2023). **A)** Numerical traits (Ht, Hd, GF and PL) **B)** Binary traits (Awn and LmPb) reported as the number of accessions showing presence or absence of the carácter. **C)** Ordinal traits (FLA and CMa) scored on a 1–5 visual scale.

The presence of awn and pubescence, two binary traits, demonstrated stable presence and absence proportions across the three years evaluated. When considering the pooled data from all years, 38% accessions displayed awns, while 62% were awnless. While the glabrous trait resulted in 35% classified as glabrous and 65% as non-glabrous.

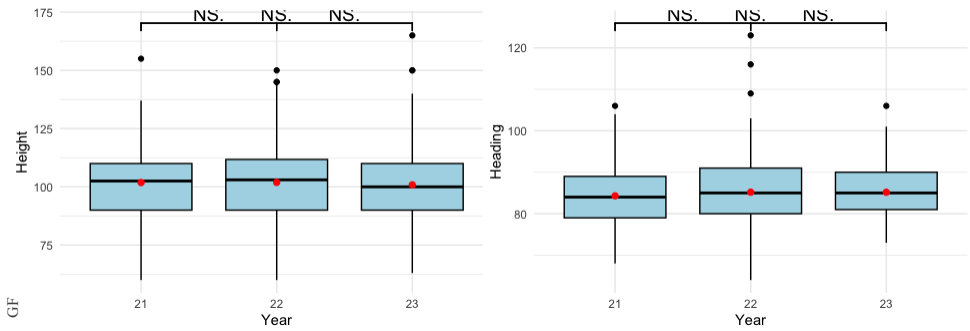
Culm and flag leaf angles, two ordinal traits, showed moderate variability within the population but remained relatively stable across seasons. The coefficients of variation of CmA were slightly higher than those for flag leaf angle. But they were relatively uniform across years.

Year-to-Year Stability of Quantitative Traits

To further investigate year-to-year variability, boxplot visualizations and statistical tests were conducted for each agronomic trait measured, aiming to evaluate the influence of year-specific environmental factors (**Table 3; Figure. 1**). Student's t-tests and Chi-square tests indicated that most traits did not show significant differences between years, supporting a stable phenotypic expression over time.

	Ht	**Hd**	**GF**	**PL**	**An**	**LmPb**	**FLA**	**CMa**
P-value	0.777	0.396	0.265	0.103	0.503	0.992	0.178	0.289

Table 3. P-values obtained from year-to-year comparisons of the agronomic traits in the MAGIC population (2021 – 2023). Numerical traits were analysed with two-sample Student's *t*-tests. Binary and Ordinal traits were evaluated with Chi-square tests.



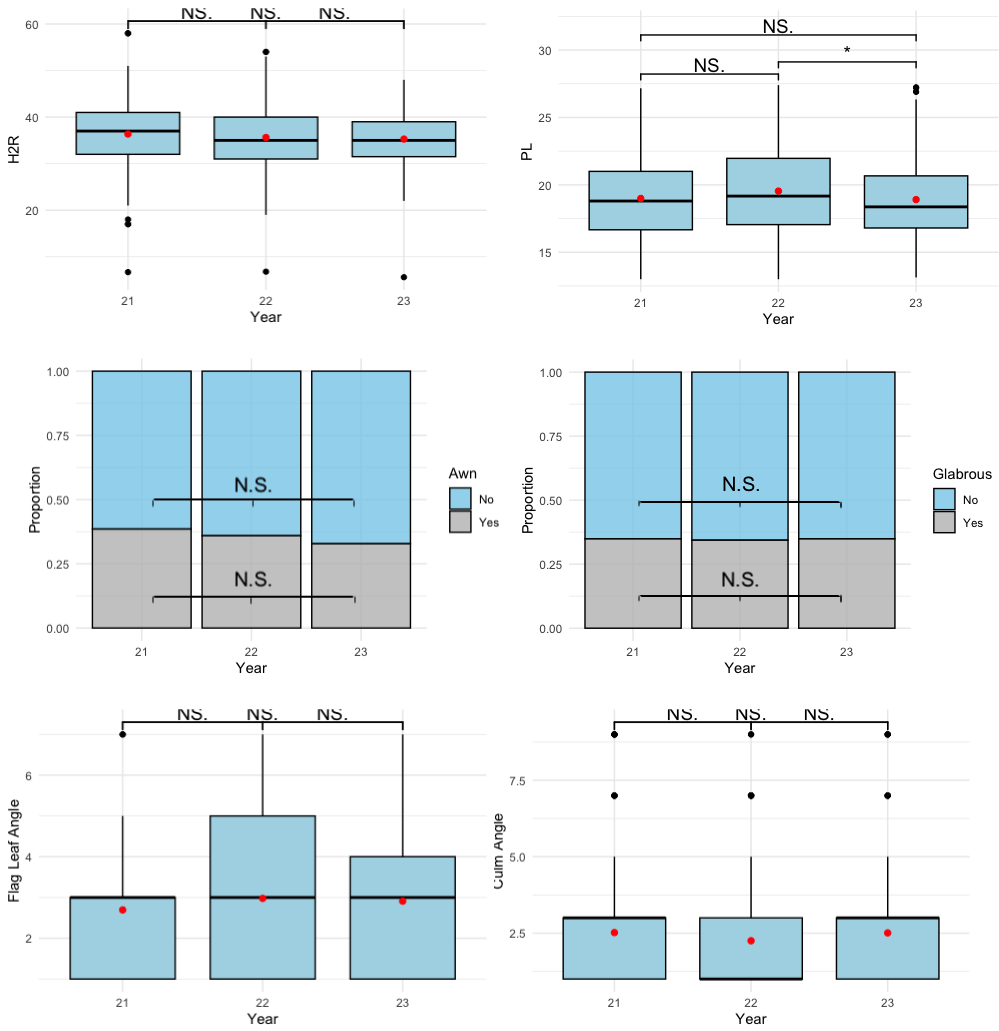


Figure 1. Boxplot comparison of the eight agronomic trait year to year. Medians are shown by thick horizontal lines and means by red dots. Stacked bar plots summarize the two binary traits (An and LmPb). Horizontal brackets mark pairwise statistical tests between years; * indicates a significant difference at $P < 0.05$, whereas "N.S." denotes non-significance (Student's *t*-tests for quantitative/ordinal traits, χ^2 tests for binary traits).

Heritability of Agronomic Traits

The broad-sense heritability (H^2) was also calculated for each agronomic trait to estimate the proportion of phenotypic variance attributable to genetic factors (**Table 4**) The results revealed a wide range of heritability values across the traits studied. Traits such as Awning and Lemma and Palea Pubescence displayed exceptionally high heritability ($H^2 = 0.998$), indicating that their phenotypic expression is almost entirely genetically determined. Similarly, Ht ($H^2 = 0.832$) and PL ($H^2 = 0.778$) showed high heritability, suggesting strong genetic control.

Other traits, such as FLA ($H^2 = 0.638$), and Hd ($H^2 = 0.667$), showed moderate heritability, reflecting a balanced contribution of genetic and environmental components. In contrast, GF ($H^2 = 0.381$) and CmA ($H^2 = 0.53$) exhibited lower heritability, emphasizing a greater influence of environmental factors or genotype-by-environment interactions.

	Ht	Hd	GF	PL	An	LmPb	FLA	CMa
Heritability (H2)	0.832	0.667	0.381	0.778	0.998	0.998	0.638	0.530
Total phenotypic variance (σp2)	271.9	52.8	42.2	9.5	1492.7	2196.5	2.8	2.1

Table 4. Broad-sense heritability (H^2) and total phenotypic variance (σ^2_p) for the eight agronomic traits.

Phenotypic Distributions and Parental Contribution

The phenotypic distributions of the MAGIC population for all studied agronomic traits seems to follow a normal pattern, as shown by the histograms with density curves (**Figure 2**). The eight parental lines contribute to the observed variability, displaying distinct values across most traits. Some traits show slight deviations, such as CmA, which is left-skewed with a predominance of smaller angles, and Flag Leaf Angle, where density increases at lower values, indicating a concentration of individuals with narrower angles. Additionally, for lemma and palea pubescence, most individuals exhibit pubescence rather than being glabrous, indicating a higher prevalence of hairy structures. In contrast, for awn presence, most individuals lack awns, with a smaller proportion displaying this trait.

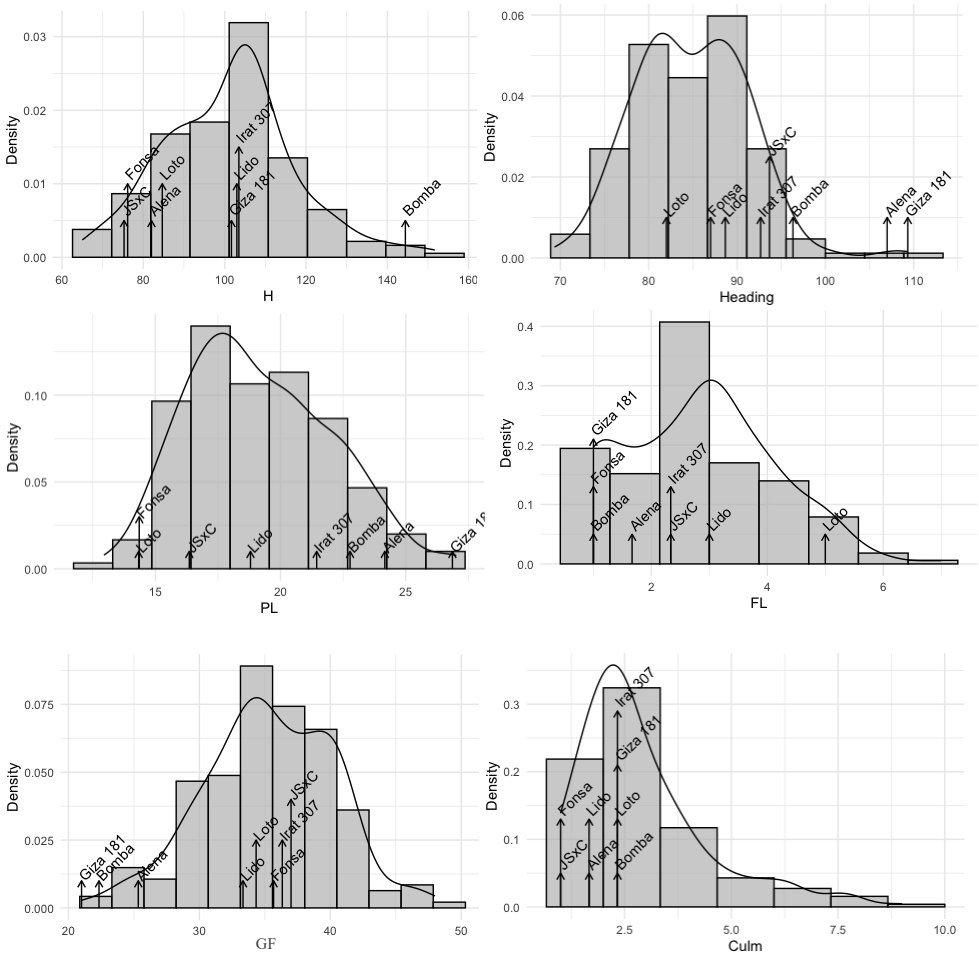


Figure 2. Frequency histograms with overlaid kernel-density curves illustrating the phenotypic distributions of six quantitative agronomic traits. Arrows mark the positions of the eight parental cultivars, highlighting their contribution to the overall range of variation.

Trait Correlations and Trade-offs Among Agronomic Traits

To explore potential relationships among the studied traits, a correlation matrix was generated (**Figure 3**). The results reveal significant positive correlations between certain traits. For example, a positive correlation was observed between Ht and PL ($r = 0.49$), suggesting that taller plants tend to develop longer panicles. Similarly, Hd

was positively correlated with PL (r = 0.34). Another positive correlation was observed between Ht and flag leaf angle (r = 0.33), which might reflect that taller plants often display wider flag leaf angles. In contrast, several negative correlations also provided insightful trends. Hd showed a negative correlation with GF (r = -0.39), indicating that earlier heading is associated with a longer period from heading to ripening.

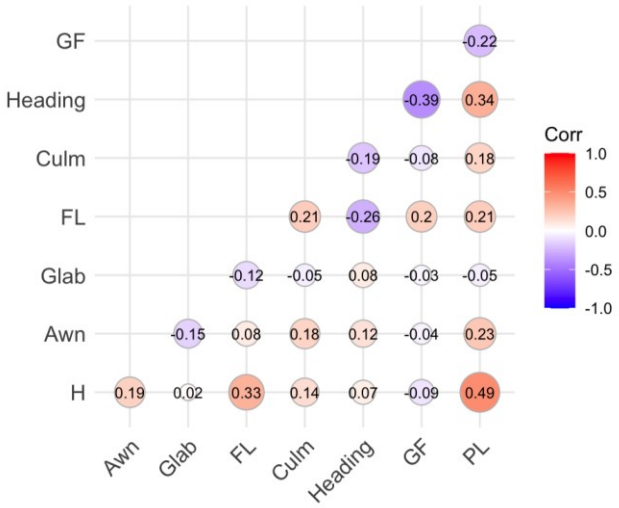


Figure 3. Pair-wise Pearson correlation matrix for the eight traits measured in the rice MAGIC population (mean values across 2021–2023). Circle colour indicates the direction and strength of the correlation (red = positive, blue = negative, while circle radius is proportional to |r|. Numeric labels give the exact correlation coefficients.

Genomic Characterization and Population Structure

SNP Discovery and Genomic Distribution

Genome sequencing of the MAGIC population generated an average of 63.5 million short reads per cultivar, corresponding to 9.58 GB per sample with a mean coverage of 47.6x. The clean reads were mapped to the *Nipponbare* reference genome (IRGSP-1.0), achieving a mapping rate of 98%, with properly paired reads. A total of 10.56 million raw variants were identified and subsequently filtered based on a call rate > 90%, read coverage between 10x and 30x, and quality thresholds (see

Methods). Indels, heterozygous SNPs, and those with a minor allele frequency (MAF) below 5% were removed, resulting in a final dataset of 70,289 high-quality SNPs.

The genomic distribution of SNP density across the 12 chromosomes, assessed within 1 Mb window sizes, revealed considerable variation (**Figure 4**). High-density regions (≥200 SNPs) and low-density regions (≤50 SNPs) were observed across all chromosomes, suggesting the presence of genomic regions under differential selective pressures or structural constraints. Based on the most recent genome assembly, with a size of 373 Mb (Kawahara et al., 2013), the average marker density was 5.29 Kb/SNP, ranging from 4.2 Kb on chromosome 7 to 13.31 Kb on chromosome 4.

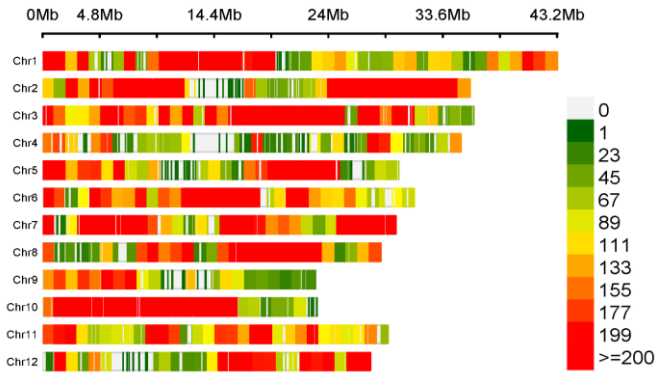


Figure 4. Chromosomal distribution of the 70,289 high-quality SNPs detected after filtering. Each horizontal bar represents one of the 12 Nipponbare reference chromosomes (IRGSP-1.0). The bar is split into 1-Mb blocks and coloured coloured segments indicate SNP density

Linkage Disequilibrium

Linkage disequilibrium (LD) decay analysis estimated a half decay distance of 879 kbp, where r2 declined to 0.23 (Figure 5), after pruning to 3,842 SNPs.

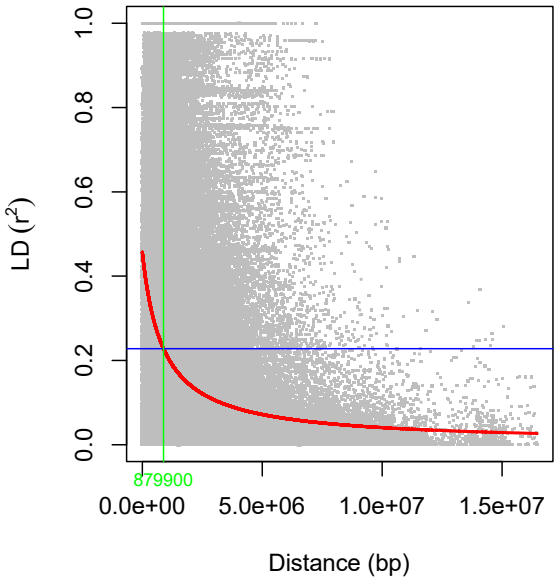


Figure 5. Linkage-disequilibrium (LD) decay in the rice MAGIC population. Each black dot shows LD (r^2) between a pair of SNPs plotted against the physical distance that separates them. The red curve is the average trend: LD falls quickly as distance grows. The blue horizontal line marks $r^2 = 0.23$, and the green vertical line (≈ 879 kb) is where the curve crosses this value, about the point at which LD has dropped to half of its starting level.

Population Structure

The population structure was assessed using STRUCTURE software, with ΔK identifying optimal subpopulations at K = 3 and K = 8 (ΔK = 29.23; ΔK = 17.62) (**Figure 6; Table 5**). Fst values confirmed genetic differentiation, ranging from 0.1211 to 0.83 at K = 3 and 0.3326 to 0.8354 at K = 8, indicating stronger structuring in the latter (**Supplementary Table 1; 2 and 3**). STRUCTURE bar plots (**Figure 7**) show clear clusters at K = 3 with some admixture, while K = 8 reflects higher complexity, with individuals displaying ancestry from multiple groups. The lack of exclusive parental clustering at K = 8 suggests that differentiation is shaped by extensive genetic mixing rather than strict lineage-based grouping. The neighbor-joining tree follows this trend, with three distinct branches at K = 3, while K = 8 reveals more interconnected subpopulations, consistent with the expected genetic diversification in MAGIC populations.

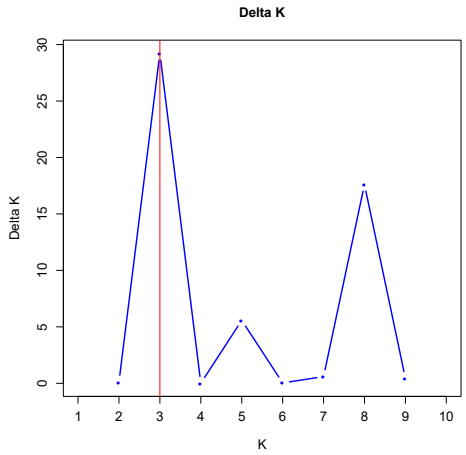


Figure 6. Delta-K plot from STRUCTURE analysis of the MAGIC population. Peaks mark the most likely numbers of genetic sub-populations according to the Evanno method.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	Ln''(K)	Delta K
1	10	-300634.87	3.30	NA	NA	NA
2	10	-284048.17	1767.62	16586.70	187.467	0.11
3	10	-267274.00	242.06	16774.17	7073.77	29.22
4	10	-257573.60	1445.31	9700.40	9.133	0.01
5	10	-247882.33	460.27	9691.27	2564.97	5.57
6	10	-240756.03	1460.06	7126.30	51.367	0.04
7	10	-233578.37	862.99	7177.67	508.467	0.59
8	10	-226909.17	169.38	6669.20	2983.57	17.62
9	10	-223223.53	3259.55	3685.63	1402.47	0.43
10	10	-218135.43	560.20	5088.10	NA	NA

Table 5. Summary of STRUCTURE run statistics for K = 1–10. Columns list the mean log probability of the data, its standard deviation, the first and second derivatives of Ln P(K), and the Evanno ΔK metric used to pinpoint the best-supported clustering levels (highest ΔK at K = 3; secondary peak at K = 8).

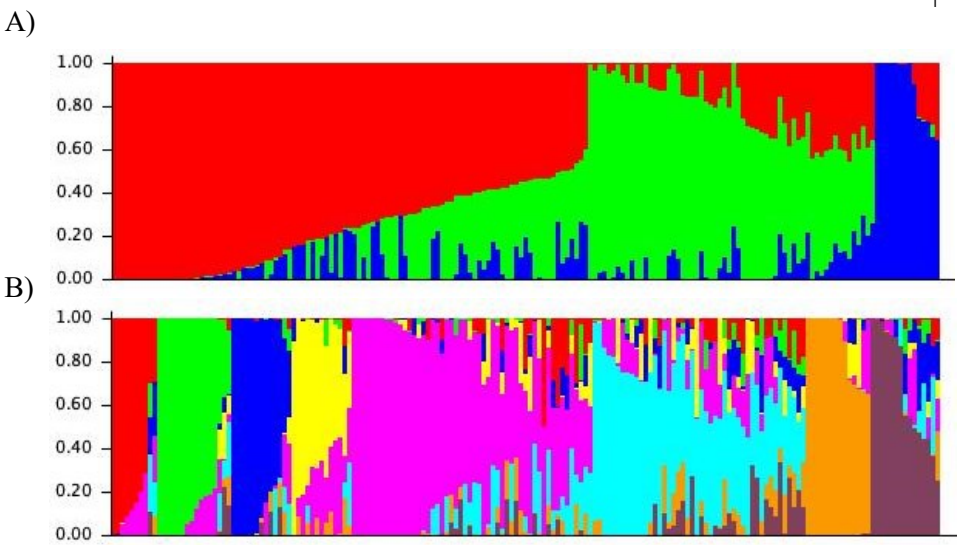


Figure 7. STRUCTURE bar plots profiles of the 192 MAGIC lines sorted by their membership coefficient (Q-value). **A)** with K = 3; **B)** with K = 8.

Principal Component Analysis (PCA)

Principal component analysis (PCA) further supports these patterns (**Figure 8**). PC1, PC2, and PC3 explain 31.47%, 24.63%, and 16.58% of genetic variance, respectively. In **Figure 8A** (K = 3), three distinct but slightly overlapping clusters suggest moderate differentiation. **Figure 8B** (K = 8) shows a more dispersed pattern, indicating increased complexity and admixture, reflecting the broader genetic variability within the population.

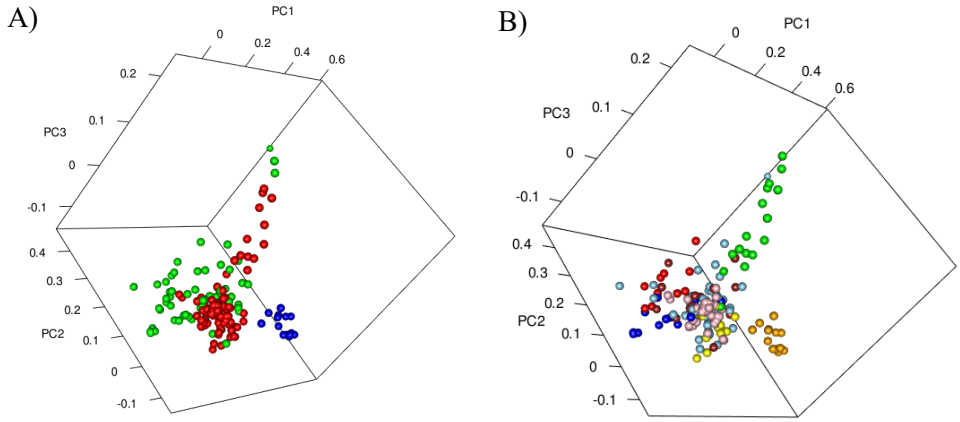


Figure 8. Three-dimensional principal-component analysis (PCA) of the 192 MAGIC rice lines based on 70,289 SNPs. **A)** For K = 3*:* points are coloured according to the three STRUCTURE clusters. B) K = 8*:* points are coloured with eight different colours for the eight STRUCTURE clusters.

Phylogenetic Analysis

Consistent with these observations, the Maximum Likelihood phylogenetic tree of the 8-way MAGIC population (**Figure 9**) shows that parental lines are interspersed throughout the tree rather than forming discrete clusters, except in the case of Lido and Loto which are genetically close. As expected in MAGIC populations, progeny lines exhibit mosaic genomes, inheriting alleles from multiple founders, this extensive recombination results in a randomized distribution of founder haplotypes, reinforcing the complex genetic structure generated by multiple intercrossing events during population development.

The population structure analysis together with the PCA and NJ analyses confirmed the absence of strong sub-structure, reducing the risk of false positives due to population stratification.

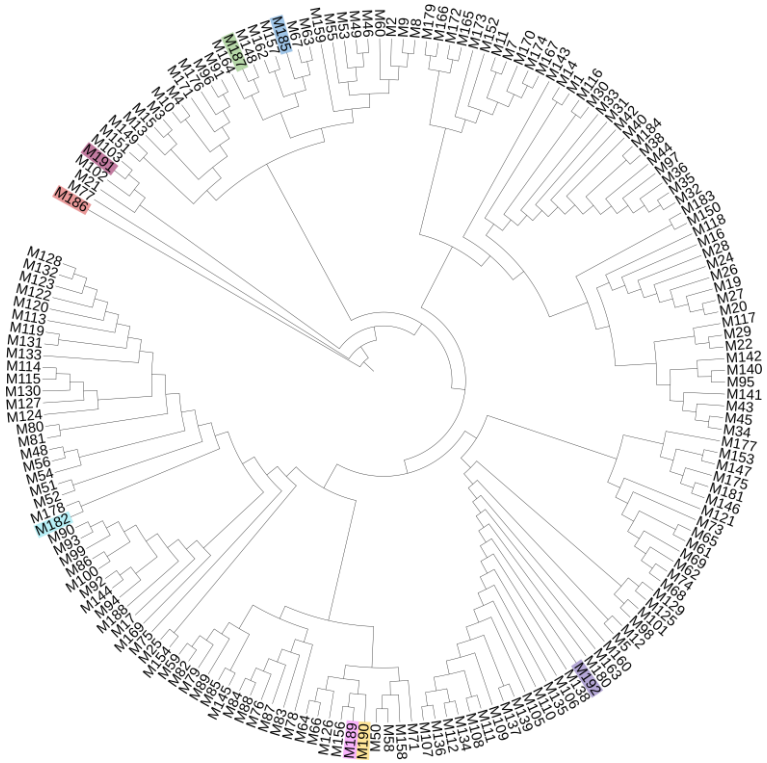


Figure 9. Unrooted circular phylogenetic tree of the 192 MAGIC lines based on 70,289 SNPs. The tree shows the genetic relationships among accessions, with branches reflecting relative clustering but not proportional to genetic distance. Parental lines are colour-labelled to illustrate their position.

GWAS and Identification of Candidate Genes

We conducted GWAS for nine agronomic traits related to plant architecture and phenology development (Height, Awn, Heading, GF, Glabrous, Flag Leaf Angle, CmA and Panicle Length) using phenotypic data collected over three plants grown in open-air pools and genotypic from 70,289 high-quality SNPs. A Mixed Linear Model (MLM) implemented in TASSEL v.5 was used to account for population structure and kinship. Trait-specific significance thresholds ($-\log_{10}(p) \geq 3$ or 4) were applied depending on the strength and consistency of the associations. Results were visualized through Manhattan and Q-Q plots (**Figure 10**).

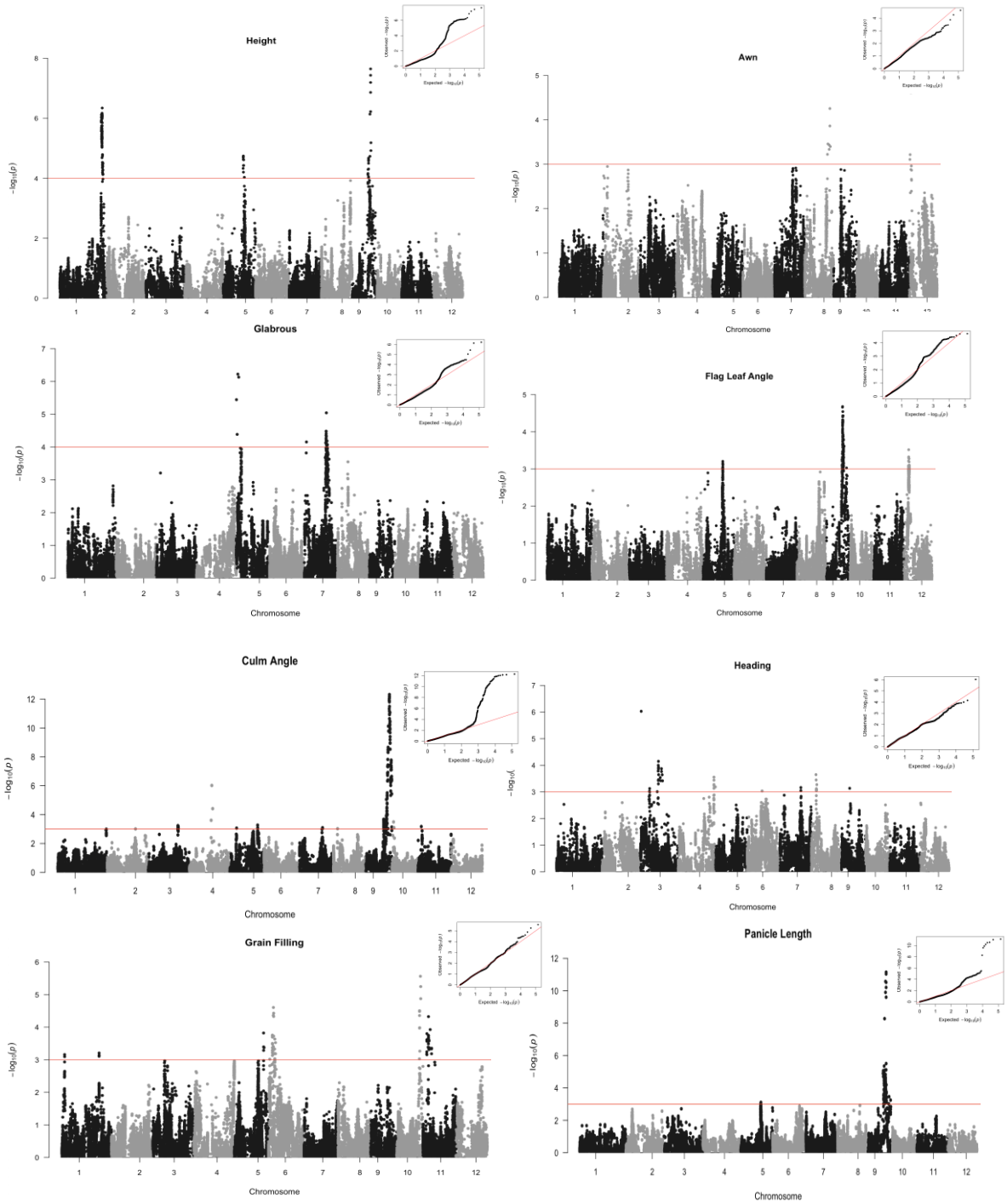


Figure 10. Manhattan and Q-Q plots of GWAS results. Each Manhattan plot displays the $-\log_{10}(p\text{-value})$ for SNP-trait associations across the 12 rice chromosomes. The red horizontal line indicates the significance threshold.

A total of 35 QTLs were detected across the nine traits (**Table 6**) Only associations supported by significant SNPs and well-defined peaks were retained. Traits related to phenology (Heading and GF) showed a higher number of QTLs, suggesting complex genetic control involving multiple loci distributed across the genome. In contrast, traits like PL displayed strong, well-differentiated signals, associated with a single major region.

QTL	Chr	Lead SNP pos	P-value	PVE (%)	Allele	Effect Size	Pop (%)	QTL Interval
Height								
qH1	1	39,035,622	4.53E-07	16.13	C	18.51	49.72	38,246,820 - 39,695,518
qH5	5	18,617,899	1.79E-05	11.42	C	-17.41	82.68	18,408,140 - 19,518,444
qH9	9	17,115,951	2.25E-08	20.39	A	-16.34	41.90	14,659,472 - 17,424,519
Awn								
qAwn8	8	24,857,921	5.55E-05	11.01	C	-0.51	84.36	22,480,705 - 25,297,626
qAwn12	12	1,835,386	2.42E-05	11.26	A	-0.34	35.75	808,650 - 1,835,386
Glabrous								
qGlb5	5	1,440,403	5.95E-07	16.17	C	-0.45	75.42	212,132 - 2,240,956
qGlb7	7	19,633,513	9.03E-06	13.30	C	0.44	45.25	19,027,731 - 20,610,304
Flaf Leaf Angle								
qFLA5	5	18,971,084	6.24E-04	7.28	A	1.30	21.79	18,806,626 - 19,108,286
qFLA9	9	15,777,388	2.09E-05	11.38	A	-1.26	64.25	15,234,847 - 17,281,984
qFLA12	12	5,130,750	3.04E-04	8.05	A	1.93	89.94	5,012,523 - 5,726,573
Culm Angle								
qCMa4	4	19,546,961	9.60E-07	15.16	C	-1.49	81.56	19,546,961 - 20,315,786
qCMa9	9	20,756,479	4.87E-13	36.88	A	-2.65	77.09	18,053,862 - 22,802,876
Heading								
qHd3.1	3	1,094,461	9.34E-07	15.20	A	6.30	18.44	-
qHd3.2	3	17,409,272	7.09E-05	9.34	A	6.56	10.06	16,289,975 - 21,288,080
qH4	4	33,787,498	2.79E-04	7.76	C	-4.38	66.48	33,773,923 - 33,799,486
qH6	6	14,170,873	9.20E-04	7.55	C	9.16	81.56	-
qH7	7	19,654,269	6.85E-04	7.04	G	3.96	52.51	19,654,246 - 19,654,269
qH8	8	4,351,455	2.24E-04	8.48	A	-4.90	16.76	4,351,455 - 5,133,187
qH9	9	7,974,566	7.30E-04	7.12	C	-4.10	45.81	-
Grain Filling								
qGF1.1	1	2,345,577	6.94E-04	7.37	A	-3.31	21.23	2,345,577 - 2,374,391
qGF1.2	1	32,481,234	6.22E-04	7.22	A	-7.05	5.03	32,450,786 - 32,498,903
qGF5	5	25,541,657	1.51E-04	8.75	A	3.45	29.61	25,541,657 - 25,928,622
qGF6	6	4,330,339	2.49E-05	11.00	A	3.85	41.34	3,457,608 - 6,165,836

qGF10	10	20,860,950	2.74E-06	14.13	A	5.01	73.18	19,994,100 - 21,136,910
qGF11	11	4,906,190	4.74E-05	10.11	A	-7.17	5.59	3,380,807 - 7,620,880
Panicle Length								
qPL9	9	17,115,954	7.15E-12	32.96	C	-3.44	41.34	14,198,097 - 17,281,984

Table 6. List of QTLs significantly associated with agronomic traits in the rice MAGIC population. Each QTL is identified by its name and chromosomal location (*Chr*), with the position of the lead SNP, associated p-value, percentage of phenotypic variance explained (PVE), favourable allele, effect size, population frequency (Pop %), and physical interval.

Q-Q plots confirmed good model fit and marker–trait association quality for most traits, with observed p-values closely following the expected null distribution. However, moderate deviation from the expected line was observed for Ht, CmA, and PL. This could reflect strong genetic effects or residual confounding due to population structure.

For plant height, 128 significant SNPs ($-\log_{10}(p) \geq 4$) were detected, corresponding to three QTLs on chromosomes 1, 5, and 9 (**Suppl. Table 5**). The strongest signal was found at qH9 (chr. 9), with a $-\log_{10}(p)$ close to 8, with the highest PVE (20.39%). The lead SNP's A variant, present in under half of the population, was linked to a marked height reduction, as shown by its effect size and by the violin plot (**Figure 11**).

Only 10 significant SNPs, corresponding to two QTLs located on chromosomes 8 and 12, were identified for the presence of Awn, when the threshold was set at $-\log_{10}(p) \geq 3$ (**Suppl. Table 6**). Both QTLs explained over 11% of the phenotypic variance. While the SNP leader C at qAwn8 was widespread in the population (84%), the A variant at qAwn12 showed moderate frequency (36%). Both alleles were associated with reduced awn development or the absence of awn.

In the case of the glabrous hull trait, 29 significant SNPs ($-\log_{10}(p) \geq 4$) defined two QTLs: qGlb5 and 7 qGlb7 (**Suppl. Table 7**). These loci explained 16.2% and 13.3% of the PVE, respectively. The lead SNP at qGlb5 (allele C) was found in 75% of the lines and was linked to absence of pubescence.

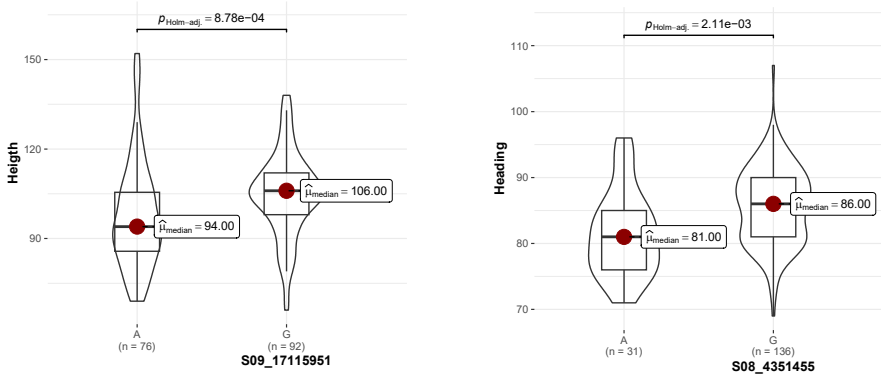
The highest number of significant SNPs ($-\log_{10}(p) \geq 3$), 183, was detected for flag leaf angle (**Suppl. Table 8**). Grouped into three QTLs on chromosomes 5, 9, and 12. These loci explained moderate phenotypic variance. Although the association at

qFLA12 was slight, the G allele of the lead SNP was linked to a fully erect flag leaf (**Figure 11**), where this genotype displayed a median angle of 90°. This variant was present in only about 10% of the population, providing considerable breeding potential.

Among phenological traits, significance levels were generally lower, and a threshold of $-\log_{10}(p) \geq 3$ was applied. For Hd, 40 significant SNPs mapped to seven QTLs with associations appearing more dispersed (**Suppl. Table 9**). Several loci, including qHd3.1, qH6, and qH7, were defined by only a single SNP. qHd3.1 showed the strongest signal despite its minimal physical span. Some lead SNPs, such as those at qHd3.2 and qH4, pointed to early-flowering alleles prevalent in the population. In contrast, qHd8 revealed a less common variant (allele A, present in 17%) that shortened the time to heading in 81 days, standing out as a strong candidate for breeding earlier varieties. Meanwhile, associations on qH6 and qH7 showed little phenotypic contrast between alleles. (**Figure 11**).

In the case of GF, 75 significant SNPs were identified that clustered into six QTLs across chromosomes 1, 5, 6, 10, and 11, of which two were detected on chromosome 1 (**Suppl. Table 10**). According to the allele content, varieties carrying one or the other allele in all QTLs, except qGF1.2, displayed significant differences in GF. The strongest association was at qGF10, where the minor T allele markedly reduced GF. At qGF11, the minor A allele shortened it to 30.5 days (**Figure 11**).

Finally, PL was associated with a single and highly robust QTL on chromosome 9, qPL9, where the leader peak showed Pvalue of 7.15E-12 (**Suppl. Table 11**). A total of 86 significant SNPs were detected within qPL9, spanning 3.08 Mb. With a PVE of 32.96%, this locus represents a major-effect region. The T allele at the leading SNP, linked with increased panicle length, was present in over half of the population (**Figure 11**).



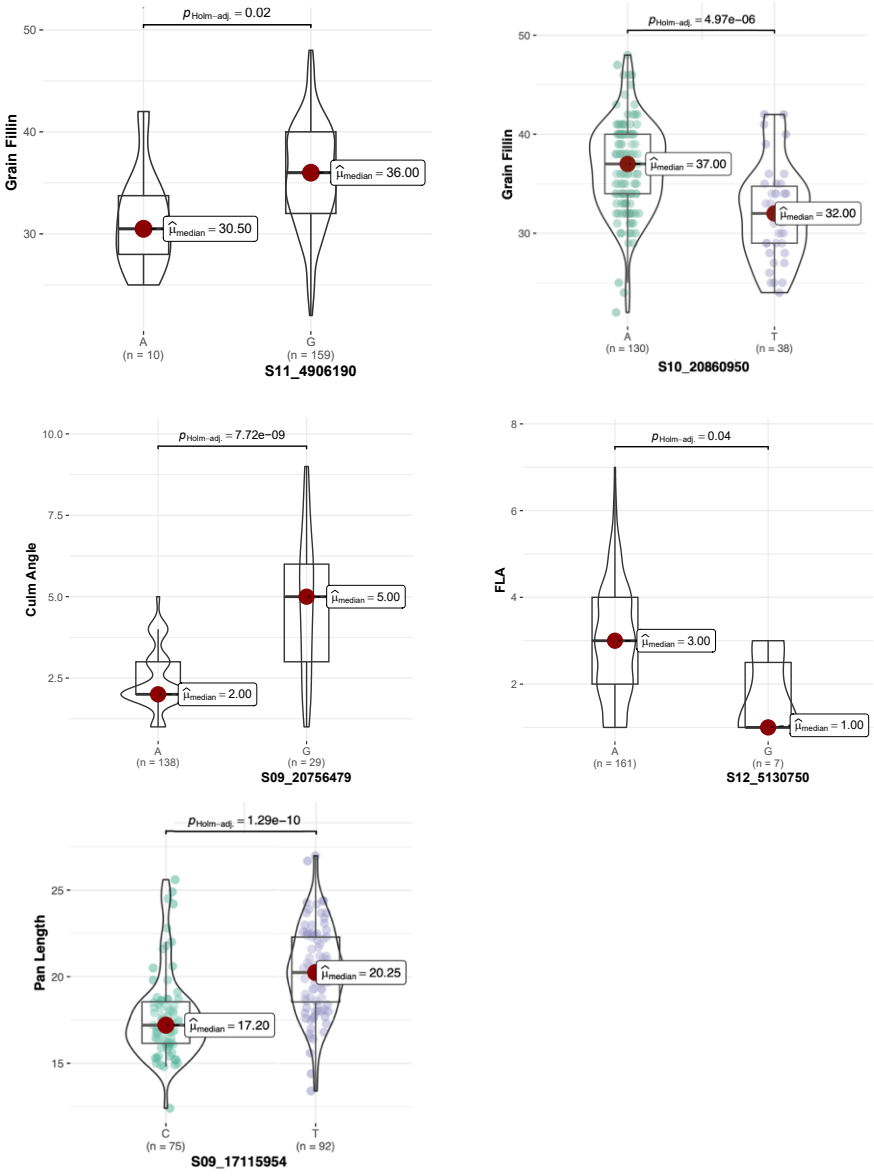


Figure 11. Violin plots showing phenotypic effects of lead SNPs at selected QTLs. Each panel displays the distribution of trait values grouped by SNP allele, with the median value highlighted by a red dot. Significant differences between genotypes ($p<0.05$) are indicated above each plot.

Candidate genes

Using the reference genome annotation IRGSP-1.0 (RAP-DB) we identified a total of 29 genes functionally consistent with the associated traits, located within the QTL regions extended by ±0.9 Mb, based on linkage disequilibrium (**Table 7**).

A larger set of candidate genes was found for Height, Heading, and GF, indicating that these traits are likely under more complex polygenic control. By contrast, only a single functionally relevant gene was detected for each of the remaining traits.

Classic GA-pathway genes, such as *Semidwarf 1* (*SD1*, *Os01g0883800*) and *Semidwarf 5* (*SDF5*, *Os05g0415800*), are present within height QTL regions. *SD1*, which encodes a GA20 oxidase, is involved in the reduction of stem elongation by limiting the gibberellin biosynthesis. *SDF5*, which encodes a cytochrome P450 protein, has been involved in the plant height as its dominant mutant reduces active GA_1/GA_4 levels and the internode length. Other genes involved in height regulation through hormonal and chromatin-related mechanisms were also detected. *Plant Height 1* (*PH1*, *Os01g0881500*), a chitin-induced gibberellins-responsive protein from the GRAS family transcription factor, which its elevated expression promotes stem elongation, and the chromatin regulator *Plant Height 9* (*PH9*, *Os09g0433600*), encoding Histone H4, were found located in qH1 and qH9. While *Ethylene Response Factor 102* (*ERF102*, *Os09g0457900*), an AP2/ERF transcription factor, restricts internode elongation by down-regulating a GA-biosynthetic gene, linking ethylene signalling to height control (Qi, W. et al. 2011).

The QTL, qAwn8.2, associated with the presence of awn, colocalized with *Regulator of Awn Elongation 2* (*RAE2*, *Os08g0485500*), a well-known causal gene encoding an epidermal patterning factor-like 1 protein that is cleaved by the subtilase *SLP1* in the spikelet and the mature peptide promotes awn elongation. It is known that independent loss-of-function mutations in *RAE2* discontinue the peptide activity, causing the awnless phenotype. Similarly, the QTLs associated with LmPb, qGlb5, contains a single well-known causal gene, *Glabrous Leaf and Hull 1* (*GL1*, *Os05g0118700*), which encodes a WOX3-class homeobox transcription factor, specifies marginal cell identity and trichome initiation on leaves and hulls. Null alleles of GL1 display glabrous grain surface. No candidate genes were found in qAwn12 nor qGlb7.

Architectural QTLs for flag-leaf and CmA include auxin-responsive genes. At qFLA5, we map the Transport Inhibitor 1 gene (*TIR1*, *Os05g0150500*), an F-box auxin receptor that tags Aux/IAA repressors for destruction when it senses auxin. Rice plants where *TIR1* (or its partner *AFB2*) is knocked down develop floppy flag leaves, showing that normal *TIR1* activity is needed to keep the flag leaf close to the culm. At qCMa9 interval we found the *Tiller Angle Control 1* gene (*TAC1*, *Os09g0529300*) an IGT/LAZY family gene member that guides auxin transport at the tiller base. A single AGGA→GGGA change at a 3′ splice site lowers *TAC1* mRNA, and plants carrying this tac1 allele form vertical tillers.

Within heading-date QTLs, we uncovered flowering regulators that act at several control layers. On chromosome 3 the gene *Early Heading Data 4* (*EHD4*, *Os03g0112700*) a zinc finger CCCH domain-containing protein is required for long day flowering. Plants lacking EHD4 do not head when day length is extended (Gao et al., 2013). The long non-coding RNA *Early Flowering-Completely Dominant* gene (*EF-CD*, *Os03g0122500*) accelerates heading by 7–20 days by boosting the transcript level of its neighbour, the floral activator MADS-box transcription factor 50 (*MADS50/DTH3*, *Os03g0122600*), which triggers the *Ehd1–Hd3a/RFT1* cascade (Fang et al., 2019; Lee et al., 2004). The clock component Early Flowering 4b *ELF4b* (*Os03g0410300*) also promotes timely heading; knockout lines flower markedly later (Wang X. et al., 2021).

In the chromosome region of qHd4 we identified three flowering regulators: the *RING-finger ligase Hal3-Interacting Protein 1* (*HIP1/HAF1*, *Os04g0648800*) that ubiquitinates the *Hd1*and promotes its degradation bringing forward the flowering period by two to three weeks (Yang Y. et al., 2015). The *DLN-repressor 127 motif transcription factor* (*DLN127*, *Os04g0676600*) which acts by slowing down flowering. And the *Late-Flowering 2* (*LFL2*, *Os04g0676650*) a LEAFY-like repressor, whose silencing by the PRC2 component *OsVIL2* releases *Ehd1* expression, making plants heading earlier (Yang J. et al., 2013).

On qHd8 we detected another transcription factor, the *Heading Date 5* gene (*Ghd8/DTH8/HD5*, *Os08g0174500*) that delays heading while boosting yield potential (Wei et al., 2020).

Locus	Gene Symbol	Gene Name	Description	QTL	Chr	Reference
Height						
Os01g0881500	*PH1, SCL1*	*Plant Height 1*	Chitin-induced gibberellins-responsive (GRAS family) protein	qH1	1	Kovi MR et al. 2011
Os01g0883800	*SD1*	*Semidwarf 1*	Similar to GA C20oxidase2	qH1	1	Sasaki A et al. 2002
Os05g0415800	*SDF5*	*Semidwarf 5*	14-alpha-demethylase cytochrome P450 (CYP51) protein	qH5	5	Yang Yachun et al. 2021
Os09g0433600	*PH9*	*Plant Height 9*	Histone H4	qH9	9	Xin W et al. 2022
Os09g0455900		*Gibberellin insensitive dwarf1 - like*	Alpha/beta hydrolase fold-3 domain containing protein..Similar to *Zea maize* gibberellin receptor GID1L2	qH9	9	RAP-DB
Os09g0455500		*Gibberellin insensitive dwarf1 - like*	Alpha/beta hydrolase fold-3 domain containing protein..Similar to *Zea maize* gibberellin receptor GID1L2	qH9	9	RAP-DB
Os09g0457900	*ERF102*	*Ethylene Response Factor 102*	AP2/ERF transcription factor	qH9	9	Qi W et al. 2011
Awn						
Os08g0485500	*RAE2*	*Regulator of Awn Alongation 2*	Epidermal patterning factor-like 1 (EPFL1) protein	qAwn8	8	Bessho-Uehara K et al. 2016 | Jin J et al. 2016
Glabrous						
Os05g0118700	*GL1*	*Glabrous Leaf and Hull 1*	A member of the Wuschel-Related Homebox gene family	qGlb5	5	Nardmann J et al. 2007 | Honda E et al. 2018
Flag Leaf Angle						
Os05g0150500	*TIR1*	*Transport Inhibitor Response 1*	F-Box auxin receptor protein	qFLA5	5	Bian H et al. 2012
Culm Angle						
Os09g0529300	*SPK, TAC1*	*Spreading stub, Tiller angle control 1*	Similar to Leaf angle-associated protein	qCMa9	9	He J et al. 2017; T. Komori et al. 2009; Yu B et al. 2007
Heading						
Os03g0112700	*EHD4*	*Early Heading Date 4*	Zinc finger CCCH domain-containing protein	qHd3.1	3	Wang D et al. 2008; Gao H et al. 2013
Os03g0122500	*EF-CD*	*Early Flowering-Completely Dominant*	Long noncoding RNA (lncRNA)	qHd3.1	3	Fang J et al. 2019

Os03g0122600	*MADS50/DTH3*	*MADS-box Transcription Factor 50*	MIKC-type MADS-box protein	qHd3.1	3	Lee S et al. 2004; Shinozuka Y et al. 1999
Os03g0410300	*ELF4B*	*Early Flowering 4b*	Ortholog of Arabidopsis ELF4	qHd3.2	3	Wang X et al. 2021
Os04g0648800	*HIP1/HAF3*	*Hal3-Interacting Protein 1*	C3HC4 RING domain-containing E3 ubiquitin ligase	qHd4	4	Yang Y et al. 2015; Sun SY et al. 2009
Os04g0676600	*DLN127*	*DLN Repressor 127*	B3 transcription factor	qHd4	4	Li L et al. 2024
Os04g0676650	*LFL2*	*Late-Flowering 2*	B3 domain-containing protein	qHd4	4	Li L et al. 2024
Os08g0174500	*HD5, DTH8, Ghd8*	*Heading date (QTL)-5(t)*	Putative HAP3 subunit of the CCAAT box-binding TF	qHd8	8	Fujino K et al. 2013; Wei X et al. 2010
Os09g0306700	*PFT1*	*Phytochrome and Flowering Time 1*	Mediator subunit 25	qHd9	9	Ren Y et al. 2020
Os09g0306800	*EMF2B*	*Embryonic Flower 2B*	Polycomb group protein	qHd9	9	Wang J et al. 2013; Yang J et al. 2013; Jeong HJ et al. 2016; Luo M et al. 2009; Wen Y et al. 2024
Os09g0307800	*LVP1*	*Long Vegetative Phase 1*	SET domain-containing histone methyltransferase	qHd9	9	Sun C et al. 2012
Grain Filling						
Os05g0509700	*TA1*	*Thick aleurone 1*	Mitochondrion-targeted single-stranded DNA binding protein	qGR5	5	Li DQ et al. 2021; Teng X et al. 2024
Os10g0508100	*OsGFR1*	*Grain fillinf rate 1*	DUF641 domain containing protein	qGF10	10	Liu E et al. 2019
Os10g0515400	*OsGW10*	*Grain Width 10*	Cytochrome P450 subfamily 89A2 homology protein	qGF10	10	Zhan P eta l. 2021
Os10g0522601	*OsGS10*	*Grain Size 10*	Armadillo (ARM) repeat protein	qGF10	10	Chen E et al. 2023
Os10g0525200	*OsCPq10*	-	Cytochrome P450 family protein	qGF10	10	Park JR et al. 2023
Os11g0169200	*OsAUX5*	*Auxin Transporter 5*	Amino acid transporter	qGR11	11	Shi Y et al. 2023
Os11g0195601	*AAP11C*	*Amino Acid Permease 11*	Amino acid permease 11	qGR11	11	Yang Y et al. 2023
Os11g0206700	*OsXLG2*	*Extra-Large G Protein 2*	Extra-large GTP-binding protein 2	qGR11	11	Biswal AK et al. 2021
PL						
Os09g0441900	*DEP1*	*Dense and Erect Panicle 1*	Phosphatidylethanolamine-binding protein (PEBP)	qPL9	9	Huang et al. 2009
Os09g0456100	*LP1*	*Long Panicle 1*	Remorin_C-containing protein	qPL9	9	Liu E et al. 2016

Table 7. List of known genes related to the traits within 0.9 Mb distance interval

Within the qHd9 region, three genes were found. The Mediator subunit *Phytochrome and Flowering Time 1* (*PFT1*, *Os09g0306700*), links a brassinosteroid factor and light-responsive phytochrome signals. Loss of *OsMED25* reduces *Hd3a/RFT1* and *Ehd1* RNA levels, delaying heading by one to two weeks (Ren et al., 2020). The Embryonic Flower 2B (EMF2B, Os09g0306800), a polycomb group protein. It adds the H3K27me3 mark to flowering repressors like *OsLF*. Mutants lack this silencing, causing misexpression of floral genes and strong early flowering (Luo et al., 2009). The *Long Vegetative Phase 1* (*LVP1*, *Os09g0307800*) an H3K36 methyl-transferase, which up-regulates *MADS50* and *RFT1* resulting in earlier flowering (Sun et al., 2012).

Grain-filling QTLs encompass multiple layers of control that range from mitochondrial energy supply to hormone and nutrient transport. In qGR5 we found the *Thick aleurone 1* (*TA1*, *Os05g0509700*), which encodes a mitochondrion-targeted single-stranded DNA-binding protein that safeguards mitochondrial DNA replication and RNA splicing. Loss of function in *TA1* results in deformed mitochondria, slowing GF, but doubles the aleurone layers, boosting protein, lipid and micronutrient density, improving nutritional quality (Li et al., 2021; Teng et al., 2024).

The qGF10 cluster contains a suite of regulators that modulate carbohydrate flow and brassinosteroid (BR) signalling. The *Grain filling rate 1* gene (*GFR1*, *Os10g0508100*), which encodes a membrane DUF641 protein, accelerates GF by interacting with the Rubisco small subunit (Liu et al., 2019). Immediately downstream, *Grain Width 10* (*GW10*, *Os10g0515400*), a CYP89A2 P450, and the *Grain Size 10* (*GS10*, *Os10g0522601*), an armadillo-repeat protein, form a BR module with opposite effects: *GW10* promotes BR response to increase the spikelet hulls, whereas *GS10* reduces the activity of the *TUD1–OsGSK2* pathway involved in BR signaling, giving narrower grains; both genes therefore help balance grain size and number (Zhan et al., 2021; Chen et al., 2023). In the same interval, *OsCPq10* (*Os10g0525200*), another cytochrome P450, was QTL mapped as a positive contributor to grain width (Park et al., 2023).

Three additional transport or signaling genes were identified at qGF11.The *Auxin Transporter 5* gene (*OsAUX5*, *Os11g0169200*) is an auxin influx carrier transporter that encodes an amino acid transporter whose favorable promoter haplotype

increases essential amino-acid levels in grain without impacting yield (Shi et al., 2023). The *Amino Acid Permease 11*, (*OsAAP11*, *Os11g0195601*), an endosperm-specific amino-acid permease that optimizes protein content in grains (Yang et al., 2023). Finally, the *Extra-Large G Protein 2* (*OsXLG2*, *Os11g0206700*), an extra-large G-protein that collaborates with its XLG paralogs to promote GF. Mutants displayed reduced grain weight and filled panicles (Biswal et al., 2021).

The search of genes related to PL within qPL9 led to the detection of *Long Panicle 1* (*LP1*, *Os09g0456100*) a remorin-C gene was the unique gene identified. Previous studies reported a *LP1* allele carrying two SNPs in the coding region, that lead to amino acid changes, that produced longer panicles and more filled grains than those produced by other alleles resulting in higher yield without altering other traits as heading date (Liu et al., 2016). *Dense and Erect Panicle 1* (*DEP1*, *Os09g0441900*), which was located in qPL9, is a well-known gene involved in the regulation of the length of the inflorescence internode, an increased number of grains per panicle and a consequent increase in grain yield (Huang et al 2009).

Discussion

The development of new populations represents a fundamental genetic and agronomic resource for modern breeding programs, as it enables the genetic dissection of complex traits. Among available strategies, multi-parent populations have become increasingly relevant. These populations are constructed by intercrossing multiple genetically diverse founders across several generations before generating inbred lines. This process enhances recombination and genetic variability, and thus MAGIC populations can overcome two main limitations of traditional mapping biparental populations: restricted recombination and low mapping resolution. Furthermore, biparental populations often require large numbers of lines to achieve acceptable mapping precision. By contrast, MAGIC populations result in stable, homozygous lines with high allelic diversity, allowing for superior mapping power and resolution, while also reducing confounding effects due to population structure.

A primary use of MAGIC populations is to map QTLs for agronomic traits across genetically diverse backgrounds with relatively high resolution. This has been facilitated by next-generation sequencing technologies, which now allow efficient, cost-effective, and near-complete genotyping across entire genomes (Auton et al., 2015; Mardis, 2017). The drop in sequencing costs has made high-density SNP data accessible even in non-model species, enabling the identification of candidate alleles within QTL intervals that may be putatively causal.

MAGIC populations are typically designed for long-term use, with broad utility rather than narrow focus on specific traits or variants. Once established, they serve as enduring platforms for dissecting complex traits, including those not initially targeted. Furthermore, when constructed with the aim of maximizing segregating genetic variation, MAGIC populations offer rich allelic diversity from multiple founders.

In this study, a MAGIC rice population derived from eight founders was characterized. The parents were selected based on their distinct agronomic relevance and adaptation to local temperate agro-climatic conditions. Of the 420 genetically

stable lines, 192 were randomly selected for whole-genome sequencing. This resulted in a high-quality panel of 70,289 SNPs, confirming the presence of considerable genetic variability. This is consistent with the high phenotypic variance observed for the recorded phenotypic traits related to plant architecture and phenology, which reflected substantial variation among the population. These traits exhibited high temporal stability across years, with no significant differences between years, and their distributions close to normal. The minor year effect on trait means, along with coefficients of variation below 20%, supports the strong genetic basis of these traits and low genotype × environment interaction. This fact allowed the three-year datasets to be pooled for the analysis. Furthermore, genetic variance accounted for a substantial proportion of the phenotypic variation, with broad-sense H^2 values near 1.0 for domestication traits such as awn presence and LmPb ($H^2 \approx 0.998$), and high values for Ht (0.832) and PL (0.778). In practical terms, for these traits, less than 15% of the observed phenotypic variation is attributable to environmental factors or error in data collection, with the remainder portion of phenotypic variation to genetic effects. High H^2 values indicates that the studied traits are genetically controlled and are therefore expected to respond efficiently to selection in breeding programs. Considering that plants were grown under natural conditions during summer with similar day-length cycle, the year-to-year consistency of phenological the traits Hd and GF, that could be affected by photoperiod, supports the genetic stability of these traits.

Positive and strong correlations were observed between some of the studied traits such as Pt and PL (r2 = 0.49). This high correlation is consistent with the QTL mapping results, as both traits shared a QTL on chromosome 9, qH9 and qPL9. This makes sense since the regulation of the final length of stem and panicle may have regulatory mechanisms in common, as those related to cell elongation among others. It is worth mentioning that genomic editing of *OsDEP1*, present in qLP9/qH9 and which regulates the meristematic activity in the panicle, not only produces shorter panicles but also shorter plants (Li et al. 2016). In addition, other traits as Hd and PL also showed moderate positive correlation, or, in contrast, Hd and GF displaying a negative correlation, suggesting that earlier heading is associated with a longer GF period.

The genetic analysis revealed LD decay to $r^2 = 0.23$ at 879 kb, comparable to values reported for other MAGIC populations (Meng et al., 2016; Raghavan et at., 2017).

The gentle LD decay compared to global diversity panels reflects the limited number of founders and controlled recombination. This greater conservation of haplotype blocks given by a large LD decay distance indicates a low recombination rate compared with natural populations, and genes remain in disequilibrium over longer genomic distances, as expected given the population's derivation from only 8 parental lines. As a result, this LD is sufficient and appropriate for detecting significant associations in the MAGIC population for which the SNP density was calculated to be of 5.29 kb.

The GWAS of the eight agronomic traits detected significant association, pointing to specific genomic regions and allowing the identification of 35 QTLs. Phenological traits showed several associations along the genome (7 QTLs for Hd and 6 for GF period), while traits such as An, LmPb, CmA, and PL each had one or two highly significant QTLs, suggesting that one or a few major loci are responsible for variation in this population. QTLs associated to Ht displayed the largest effect sizes, whereas PL and CmA showed the highest PVE (36.88% and 32.96%, respectively). The most significant SNPs were found for Ht (pValue=2.2E–08 at qH9), CmA (pValue=4.87E–13 at qCMa9), and PL (pValue=7.15E–12 at qPL9). Notably, qH9 and qPL9 overlap on chromosome 9, with their lead SNPs only three bases apart, pointing to a common genomic region involved in plant architecture.

The analysis of the allele distributions of the lead SNPs were consistent with phenotypic effects, allowing us to identify favourable alleles. Varieties carrying allele A at the leader peak in qH9 exhibited reduced height. Meanwhile, allele C at the leader peak in qAwn8 were present in varieties with grains without awn and allele G at qFLA12, present in only 10% of the population but associated with fully erect flag leaves. At qCMa9, allele A was present in varieties displaying erect culms. The favourable allele for early heading was common, whereas for qHd8, the favourable allele A appeared in only 16% of lines. For GF, both qGF10 and qGF11 had favourable minor alleles associated with shorter GF periods. In panicle length, the favourable allele T at qPL9, present in half of the lines, was associated with a median length of 20.25 cm.

Searching in the genome annotation databases, Rice Genome Annotation Project (RGAP) and the Rice Annotation Project Database (RAP-DB)/International Rice

Genome Sequencing Project, we could propose a total of 31 candidate genes in 35 QTLs.

Moderate plant Ht is a key target in breeding, as it reduces the risk of lodging and facilitates harvesting, both manual and mechanised. In addition, shorter plants improve the harvest index by reducing the proportion of biomass allocated to stem and leaf, thereby increasing the fraction dedicated to grain production. The gibberellin biosynthesis pathway and its regulatory transcription factors are central to controlling internode elongation and stem length. In this context, we identified several well-known genes associated with reduced plant height: *SD1* and *PH1* within qH1, *SD5* within qH5, and both *PH9* and *ERF102* within qH9. Two genes coding for proteins similar to maize gibberellin receptor GID1L2 co-localized also with qH9. These genes have been extensively reported in the literature for their role in producing semi-dwarf phenotypes that align with the breeding objectives for improved plant stature.

The genes responsible for the absence of awn in rice grains have been selected during domestication, but also into the early 20th century, coinciding with the onset of agricultural mechanisation. Over time, the awns of commercial cultivars have been shortened or eliminated as the absence of awns conferred multiple advantages to farmers: grain dehusking becomes faster, the sharp effect on the straw is eliminated, mechanical harvesters are less likely to jam, and storage in the silos is more efficient as grains takes less space. In our study, we identified within one of the detected QTLs, qAwn8, a gene encoding an *Epidermal Patterning Factor-Like 1* protein, *RAE2*, which has been previously associated with the reduction of awn length. Similarly, for the lemma pubescence, we identified within the qGlb5 locus the *GL1* gene, which is responsible for the presence of trichomes or hair-like structures on the grain surface. This feature is generally considered favourable by farmers, as it can slightly increase grain weight and may also provide protection in the field against pests and environmental stress. However, no previously described genes were found within the qAwn12 and qGBL12 intervals, which opens up new possibilities for gene discovery and for breeding, particularly in view of our local agroclimatic conditions.

The trait conferring plant erectness, both at the base of the culm and in the flag leaf, is also highly desirable in breeding programs. Erect plants are less likely to lodge, which reduces yield losses. Additionally, this architecture enables higher planting

densities, since erect plants occupy less horizontal space and intercept light more efficiently. In this study, we identified two classical genes associated with this trait. For flag leaf angle (FLA), the gene *TIR1*, a key component of the auxin signalling pathway, was detected. Its activity is essential for maintaining erect flag leaf architecture. For culm angle, another gene, *TAC1*, was identified within the qCMa9 region. Also related to the auxin pathway, *TAC1* plays a well-documented role in regulating tiller and culm angle, contributing to upright plant structure.

Hd is a long-standing priority trait in rice breeding. To reach maximum yield, it is necessary to choose an optimum heading date appropriate to the day length in a given region. Shorter reproductive phases help avoid yield losses, reduce water and energy use, mitigate climate and disease risks (e.g. late-season rains). They also help to accelerate breeding cycles. All lines of MAGIC population analysed were able to flower during the summer, just like the founder parents that are adapted to the Mediterranean climate. For this reason, it was expected to found no differences in the major photoperiod regulatory genes of flowering in temperate climate, with long days during the growing season, such as *RTF1*, *Hd3a* or *Ehd1* (Naranjo et al., 2014). However, there are other additional regulatory genes that may be involved in variations in Hd. In this study, we identified a total of 18 candidate genes associated with Hd. Among them, there are well-known genes involved in the regulation of flowering as *EHD4*, *MADS50/DTH3*, *ELF4B*, localized in qHd3, or *DTH8/Ghd8* in qHd8. Other genes encoding transcriptional regulators involved in flowering, *HIP1/HAF3*, *DLN127*, and *LFL2*, were detected in chromosome 4.

GF is a complex trait that contributes greatly to grain weight, and it is governed by several QTLs. Grain filling and endosperm development are critical for the accumulation of major storage compounds in rice grains. These processes depend on the extensive remobilization of carbon reserves from source to sink and the precise regulation of sucrose conversion into starch. The developmental progression of the panicle, along with environmental cues influence the carbon flow between the leaves, leaf sheath, stem, and spikelets during grain filling. These processes required effective coordination between different source and sink organs by precise and intricated regulatory mechanisms, which are mediated by genetic factors and phytohormones. The analysis mapped 6 QTLs associated to GF. Related to these processes, candidate genes involved in grain size, *GW10*, *GS10* and *CPq10* colocalized with qGF10. And *AUX5*, encoding an auxin and amino acid transport,

and *AAP11* which is involved in amino acid absorption and affects the protein content of rice grains are located in qGR11. Notably, despite identifying multiple candidate genes for some QTLs, no candidates were detected for qHd6 and qHd7, nor for qGF1.1, qGF1.2, qGF5, and qGF6.

PL may affect yield potential, as longer panicles allow for a higher number of spikelets per panicle. A major QTL was associated with PL in the MAGIC population. LP1, located within the qPL9 region. Certain LP1 allele has been shown to increase PL by up to 40%, which in turn can result in an 11% increase in grain number (Liu et al., 2016). *OsDEP1*, which encodes a protein with a phosphatidylethanolamine-binding protein-like domain and participates in the regulation of internode length in panicles. Diversity analysis of *OsDEP1* revealed a dominant allele that confers a gain of function. This results in increased meristematic activity and, consequently, a reduction in the length of the panicle internode and an increased density of grains in the panicle (Huang et al 2009).

Overall, these results confirm that MAGIC populations are ideally suited for studying a wide array of agronomic traits. The population examined here displayed high genetic diversity within each trait, consistent with strong genetic support and highly stable phenotypic performance. The GWAS analyses clearly identified discrete genomic regions associated with the traits of interest, enabling fine mapping and the identification of biologically meaningful candidate genes. In addition to recovering expected associations, our study revealed novel genomic regions not previously described in rice, highlighting QTLs that may be specific to this population and offering valuable targets for breeding improved varieties adapted to temperate agroclimatic conditions.

Conclusions

We characterized a MAGIC population derived from eight temperate *japonica* founders to establish a novel genetic resource for the study and improvement of complex agronomic traits in rice. The population was genotyped using over 70,000 high-quality SNPs and phenotyped for eight plant architecture and phenological traits over three consecutive seasons. The population exhibited broad phenotypic variation and high heritability for most traits, confirming its suitability as a high-resolution mapping population. GWAS analysis identified 35 QTLs associated with eight agronomic traits, including plant height, panicle length, awning, lemma pubescence, tiller and flag leaf angle, heading date, and grain-filling duration. Among these QTLs, we found robust genomic regions, some of which have been previously reported and some which were novel, including qH9 for plant height, qPL9 for panicle length, and qCMa9 for culm angle. This highlight candidate genes with critical roles in plant development and phenology, such as *SD1, DEP1, TAC1*, and *RAE2*. Overall, these results highlight the effectiveness of MAGIC populations for gene discovery and provide valuable targets for marker-assisted selection. The identified QTLs and candidate genes offer a solid basis for ideotype optimization and improved phenological adaptation under temperate growing conditions.

References

Akabane T. et al. (2025) Source ability is regulated by THOUSAND-GRAIN WEIGHT 6 in rice. Plant Physiol. Biochem. 222: 109760. https://doi.org/10.1016/j.plaphy.2025.109760

Amarasinghe Y.P.J., Kuwata R., Nishimura A. et al. (2020). Evaluation of domestication loci associated with awnlessness in cultivated rice, Oryza sativa. Rice 13, 26. https://doi.org/10.1186/s12284-020-00386-4

Ashikari M. et al. (2005) Cytokinin oxidase regulates rice grain production. Science 309: 741-745. https://doi.org/10.1126/science.1113373

Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., Abecasis, G. R., & The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. Nature, 526(7571), 68–74. https://doi.org/10.1038/nature15393

Bai S. et al. (2021) Dissection of the genetic basis of panicle architecture in rice. Rice 14: 77. https://doi.org/10.1186/s12284-021-00520-w

Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., et al. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice, 6, 11. https://doi.org/10.1186/1939-8433-6-11

Bessho-Uehara, K., Wang, D. R., Furuta, T., Minami, A., Nagai, K., Gamuyao, R., Asano, K., Angeles-Shim, R. B., Shimizu, Y., Ayano, M., Komeda, N., Doi, K., Miura, K., Toda, Y., Kinoshita, T., Okuda, S., Higashiyama, T., Nomoto, M., Tada, Y., Shinohara, H., Matsubayashi, Y., Greenberg, A., Wu, J., Yasui, H., Yoshimura, A., Mori, H., McCouch, S. R., & Ashikari, M. (2016). Loss of function at RAE2, a previously unidentified EPFL, is required for awnlessness in cultivated Asian rice. Proceedings of the National Academy of Sciences, 113(32), 8969-8974. https://doi.org/10.1073/pnas.1604849113

Bian H-W, Xie Y-K, Guo F, Han N, Ma S-Y, Zeng Z-H, Wang J-H, Yang Y-N, Zhu M-Y. 2012. Distinctive expression patterns and roles of the miRNA393/TIR1 homolog module in regulating flag-leaf inclination and primary and crown-root growth in rice. New Phytologist 196 (1): 149-161. https://doi.org/10.1111/j.1469-8137.2012.04248.x

Biswal AK, Wu T-Y, Urano D, Pelissier R, Morel J-B, Jones AM & Biswal AK (2022). Novel Mutant Alleles Reveal a Role of the Extra-Large G Protein in Rice Grain Filling, Panicle Architecture, Plant Growth, and Disease Resistance. Front. Plant Sci. 12:782960. https://doi.org/10.3389/fpls.2021.782960

Bossa-Castro, A. M., Tekete, C., Raghavan, C., Delorean, E. E., Dereeper, A., et al. (2018). Allelic variation for broad-spectrum resistance and susceptibility to bacterial pathogens identified in a rice MAGIC population. Plant Biotechnology Journal, 16(9), 1559-1568. https://doi.org/10.1111/pbi.12895

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, 23(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Chen, E., Hou, Q., Liu, K., Gu, Z., Dai, B., Wang, A., Feng, Q., Zhao, Y., Zhou, C., Zhu, J., Shangguan, Y., Wang, Y., Lv, D., Fan, D., Huang, T., Wang, Z., Huang, X., & Han, B. (2023). Armadillo-repeat-only protein GS10 negatively regulates brassinosteroid signalling to control rice grain size. Plant Physiology, 192(2), 967-981. https://doi.org/10.1093/plphys/kiad117

Chen Z.W., Li H.S., Wang R.Z., Xiao L.T. (2007). Relationship between grain yield and leaf photosynthetic rate in some super-hybrid rices. Acta Agronomica Sinica 33(6): 531-537.

Cheng, X., Huang, Y., Tan, Y., Tan, L., Yin, J., Zou, G. (2022). Potentially useful dwarfing or semi-dwarfing genes in rice breeding in addition to the sd1 gene. Rice, 15, 66. https://doi.org/10.1186/s12284-022-00615-y

Churchill, G. A., Airey, D. C., Allayee, H., Angel, J. M., Attie, A. D., et al. (2004). The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nature Genetics, 36(11), 1133-1137. https://doi.org/10.1038/ng1104-1133

Dardick, C., Callahan, A., Horn, R., et al. (2013). PpeTAC1 promotes horizontal branch growth in peach and belongs to a conserved gene family regulating shoot architecture. The Plant Journal, 75 (4): 618-630. https://doi.org/10.1111/tpj.12234

Descalsota-Empleo, G. I. L., Swamy, B. P. M., Zaw, H., Thomson, M. J., Ebron, L. A., Cruz, C. V., & Collard, B. C. Y. (2018). Genome-wide association mapping in a rice MAGIC Plus population detects QTLs and genes useful for biofortification. Frontiers in Plant Science, 9, 1347. https://doi.org/10.3389/fpls.2018.01347

Dong H., Zhao H., Li S., Han Z., Hu G., Liu C. et al. (2018). Genome-wide association studies reveal that members of bHLH subfamily 16 share a conserved function in regulating flag-leaf angle in rice (Oryza sativa). PLoS Genetics 14(4): e1007323. https://doi.org/10.1371/journal.pgen.1007323

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology, 14(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fang, J., Zhang, F., Wang, H., Wang, W., Zhao, F., Li, Z., Sun, C., Chen, F., Xu, F., Chang, S., Wu, L., Bu, Q., Wang, P., Xie, J., Chen, F., Huang, X., Zhang, Y., Zhu, X., Han, B., Deng, X., & Chu, C. (2019). Ef-cd locus shortens rice maturity duration without yield penalty. Proceedings of the National Academy of Sciences, 116(14), 18717–18722. https://doi.org/10.1073/pnas.1815030116

Furuta T., Komeda N., Asano K. et al. (2015). Convergent loss of awn in two cultivated rice species Oryza sativa and Oryza glaberrima is caused by mutations in different loci. G3 Genes|Genomes|Genetics, 5(11), 2267–2274. https://doi.org/10.1534/g3.115.020834

Gao, H., Zheng, X.-M., Fei, G., Chen, J., Jin, M., Ren, Y., Wu, W., Zhou, K., Sheng, P., Zhou, F., Wang, X., Dong, Z., Liu, S., Wang, J., Wang, H., & Wan, J. (2013). Ehd4 encodes a novel and Oryza-genus-specific regulator of photoperiodic flowering in rice. PLOS Genetics, 9(2), e1003281. https://doi.org/10.1371/journal.pgen.1003281

Gu B., Zhou T., Luo J. et al. (2015). An-2/LABA1 encodes a cytokinin-activating enzyme that regulates awn elongation and grain production in rice. Plant Cell 27, 1875-1888. https://doi.org/10.1016/j.molp.2015.08.001

He J, Shao G, Wei X, Huang F, Sheng Z, Tang S, Hu P. 2017. Fine mapping and candidate-gene analysis of qTAC8, a major QTL controlling tiller angle in rice (Oryza sativa L.). PLoS ONE 12 (5): e0178177. https://doi.org/10.1371/journal.pone.0178177

Honda E, Yew C-L, Yoshikawa T, Sato Y, Hibara K-I, Itoh J-I. 2018. LEAF LATERAL SYMMETRY1, a member of the WUSCHEL-RELATED HOMEOBOX3 gene family, differentially regulates lateral-organ development in rice. Plant & Cell Physiology 59 (2): 376-391. https://doi.org/10.1093/pcp/pcx196

Hua L., Wang D.R., Tan L. et al. (2015). LABA1, a domestication gene associated with long, barbed awns in wild rice. Plant Cell 27, 1875-1888. https://doi.org/10.1105/tpc.15.00260

Huang, B. E., George, A. W., Forrest, K. L., Kilian, A., Hayden, M. J., et al. (2012). A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotechnology Journal, 10(7), 826-839. https://doi.org/10.1111/j.1467-7652.2012.00702.x

Huang, X., Qian, Q., Liu, Z., Sun, H., He, S., Luo, D., Xia, G., Chu, C., Li, J., & Fu, X. (2009). Natural variation at the DEP1 locus enhances grain yield in rice. Nature Genetics, 41(4), 494-497. https://doi.org/10.1038/ng.352

International Board for Plant Genetic Resources (IBPGR) & International Rice Research Institute (IRRI). (1980). Descriptors for rice (Oryza sativa L.). IBPGR-IRRI Rice Advisory Committee. Manila, Philippines: International Rice Research Institute.

International Rice Research Institute (IRRI). (1980). Standard evaluation system for rice (SES) (2nd ed.). Los Baños, Philippines: International Rice Research Institute.

Ishida T., Kurata T., Okada K., Wada T. (2008). A genetic regulatory network in the development of trichomes and root hairs. Annual Review of Plant Biology 59: 365-386. https://doi.org/10.1146/annurev.arplant.59.032607.092949

Itoh, H., Nonoue, Y., Yano, M., & Izawa, T. (2010). A pair of floral regulators sets critical day length for Hd3a florigen expression in rice. Nature Genetics, 42 (7): 635-638. https://doi.org/10.1038/ng.606

Izawa, T., Oikawa, T., Sugiyama, N., et al. (2002). Phytochrome mediates the external light signal to repress FT orthologs in photoperiodic flowering of rice. Genes & Development, 16 (15): 2006-2020.

Jin J., Hua L., Zhu Z. et al. (2016). GAD1 encodes a secreted peptide that regulates grain number, grain length and awn development in rice domestication. Plant Cell 28, 2453-2463. https://doi.org/10.1105/tpc.16.00379

Kadambari G., Vemireddy L.R., Srividhya A., Nagireddy R. et al. (2018). QTL-Seq-based genetic analysis identifies a major genomic region governing dwarfness in rice (Oryza sativa L.). Plant Cell Reports 37, 677-687. https://doi.org/10.1007/s00299-018-2260-2

Kangben, F., Kumar, S., Li, Z., et al. (2024). Phylogenetic and functional analysis of TAC1 homeologs in allotetraploid cotton. Frontiers in Plant Science, 14: 1320638. https://doi.org/10.3389/fpls.2023.1320638

Komori T, Miyata M, Yamamoto T, Nitta N, Hiei Y, Yano M, Ueki J, Komari T. 2009. Isolation and functional analysis of the gene controlling the stub-spreading trait in rice (Oryza sativa L.). Plant Breeding 128 (6): 568-575. https://doi.org/10.1111/j.1439-0523.2009.01639.x

Kover, P. X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I. M., et al. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in Arabidopsis thaliana. PLoS Genetics, 5(7), e1000551. https://doi.org/10.1371/journal.pgen.1000551

Kovi, M. R., Bai, X., Mao, D., Xing, Y. (2011). Seasonal plant-height responses in diverse *indica* rice accessions. Euphytica, 179, 319-331. https://doi.org/10.1007/s10681-010-0332-7

Ku, L., Wei, X., Zhang, S., et al. (2011). Cloning and characterization of a putative TAC1 ortholog associated with leaf angle in maize (Zea mays L.). PLoS ONE, 6 (6): e20621. https://doi.org/10.1371/journal.pone.0020621

Lee, S., Kim, J., Han, J. J., Han, M. J., & An, G. (2004). Functional analyses of the flowering time gene OsMADS50, the putative SUPPRESSOR OF OVEREXPRESSION OF CO1/AGAMOUS-LIKE 20 (SOC1/AGL20) ortholog in rice. The Plant Journal, 38(5), 754–764. https://doi.org/10.1111/j.1365-313X.2004.02082.x

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. Nucleic Acids Research, 49(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Li, D.-Q., Wu, X.-B., Wang, H.-F., Feng, X., Yan, S.-J., Wu, S.-Y., Liu, J.-X., Yao, X.-F., Bai, A.-N., Zhao, H., Song, X.-F., Guo, L., Zhang, S.-Y., & Liu, C.-M. (2021). Defective mitochondrial function by mutation in THICK ALEURONE 1 encoding a mitochondrion-targeted single-stranded DNA-binding protein leads to increased aleurone cell layers and improved nutrition in rice. Molecular Plant, 14(8), 1343-1361. https://doi.org/10.1016/j.molp.2021.05.016

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.

Li, J., Yuan, Y., Lu, Z., Yang, L., Gao, R., Lu, J., Li, J., & Xiong, G. (2012). Glabrous Rice 1, encoding a homeodomain protein, regulates trichome development in rice. Rice, 5, 32. https://doi.org/10.1186/1939-8433-5-32

Li, M., Li, X., Zhou, Z., Wu, P., Fang, M., Pan, X., Lin, Q., Luo, W., Wu, G., & Li, H. (2016). Reassessment of the Four Yield-related Genes Gn1a, DEP1, GS3, and IPA1 in Rice Using a CRISPR/Cas9 System. Frontiers in Plant Science, 7, 377. https://doi.org/10.3389/fpls.2016.00377

Li, X.-F., Liu, Z.-X., Lu, D.-B., Liu, Y.-Z., Mao, X.-X., Li, Z.-X., & Li, H.-J. (2013). Development and evaluation of multigenotype varieties of rice derived from MAGIC lines. Euphytica, 192(1), 77-86. https://doi.org/10.1007/s10681-013-0879-1

Li, Y.-L., & Liu, J.-X. (2018). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. Molecular Ecology Resources, 18(1), 176–177. https://doi.org/10.1111/1755-0998.12719

Liu B. et al. (2025) Carbohydrate flow during grain filling: phytohormonal regulation and genetic control in rice. J. Integr. Plant Biol. 67: 1086-1104. https://doi.org/10.1111/jipb.13904

Liu, E., Liu, Y., Wu, G., Zeng, S., Tran Thi, T.G., Liang, L., Liang, Y., Dong, Z., She, D., Wang, H., Zaid, I.U. & Hong, D. (2016). Identification of a Candidate Gene for Panicle Length in Rice (Oryza sativa L.) Via Association and Linkage Analysis. Frontiers in Plant Science, 7: 596. https://doi.org/10.3389/fpls.2016.00596

Liu, E., Zeng, S., Zhu, S., Liu, Y., Wu, G., Zhao, K., Liu, X., Liu, Q., Dong, Z., Dang, X., Xie, H., Li, D., Hu, X., & Hong, D. (2019). Favourable alleles of GRAIN-FILLING RATE 1 increase grain-filling rate and yield of rice. Plant Physiology, 181(3), 1207-1222. https://doi.org/10.1104/pp.19.00413

Liu F., Wang P., Zhang X., Li X., Yan X., Fu D. & Wu G. (2018). The genetic and molecular basis of crop height based on a rice model. Planta 247, 1-26. https://doi.org/10.1007/s00425-017-2798-1

Liu, L., Wang, X., Wang, S., Wang, X., Song, J., Zhang, Y., Ma, X., Xu, B., Liu, G., Ren, Y., & Wang, X. (2016). Identification of a candidate gene for panicle length in rice (Oryza sativa L.). Frontiers in Plant Science, 7, 596. https://doi.org/10.3389/fpls.2016.00596

Lu H. et al. (2017) OsRAMOSA2 shapes panicle architecture through regulating pedicel length. Front. Plant Sci. 8: 1538. https://doi.org/10.3389/fpls.2017.01538

Luo J., Liu H., Zhou T., Gu B. et al. (2013). An-1 encodes a basic helix-loop-helix protein that regulates awn development, grain size, and grain number in rice. Plant Cell 25, 3360-3376. https://doi.org/10.1105/tpc.113.113589

Luo, M., Platten, D., Chaudhury, A., Peacock, W. J., & Dennis, E. S. (2009). Expression, imprinting and evolution of rice homologs of the Polycomb-group genes. Molecular Plant, 2(4), 711-723. https://doi.org/10.1093/mp/ssp036

Luong N.H., Balkunde S.G., Shim K.-C. et al. (2022). Characterization of domestication loci associated with awn development in rice. Rice 15, 61. https://doi.org/10.1186/s12284-022-00607-y

Mardis, E. R. (2008). Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics, 9, 387–402. https://doi.org/10.1146/annurev.genom.9.081307.164359

Matsubara, K. & Yano, M. (2018). Genetic and Molecular Dissection of Flowering Time Control in Rice. Rice Genomics, Genetics and Breeding (pp. 177-190). Springer, Singapore. https://doi.org/10.1007/978-981-10-7461-5_11

Meng, L., Guo, L., Ponce, K., Zhao, X., & Ye, G. (2016). Characterization of three *indica* rice multiparent advanced generation inter-cross (MAGIC) populations for quantitative trait loci identification. The Plant Genome, 9(2), eplantgenome2015.10.0109. https://doi.org/10.3835/plantgenome2015.10.0109

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution, 37(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Moyers, B. T., Henry, A., Raghavan, C., Zaw, H., Marrano, A., et al. (2022). Strong genotype by environment interactions in the rice Global MAGIC population across seedling-stage drought. bioRxiv. https://doi.org/10.1101/2022.02.07.479433

Murdoch, D., & Adler, D. (2023). rgl: 3D visualization using OpenGL (Versión 1.3-18) [Paquete R]. CRAN. https://CRAN.R-project.org/package=rgl

Nagai K., Mori Y., Ishikawa S. et al. (2020). Antagonistic regulation of the gibberellic acid response during stem growth in rice. Nature 584, 109-114. https://doi.org/10.1038/s41586-020-2501-8

Naranjo, L., Talón, M., & Domingo, C. (2014). Diversity of floral regulatory genes of *japonica* rice cultivated at northern latitudes. BMC Genomics, 15: 101. https://doi.org/10.1186/1471-2164-15-101

Ning J., Zhang B., Wang N., Zhou Y., Xiong L., Wang Z. et al. (2011). Increased Leaf Angle1, a Raf-Like MAPKKK that interacts with a nuclear-protein family, regulates mechanical-tissue formation in the lamina joint of rice. The Plant Cell 23(12): 4334-4349. https://doi.org/10.1105/tpc.111.093419

Ogawa, D., Yamamoto, E., Ohtani, T., Kanno, N., Tsunematsu, H., et al. (2018). Haplotype-based allele mining in the Japan-MAGIC rice population. Scientific Reports, 8, 4379. https://doi.org/10.1038/s41598-018-22657-3

Ortiz, E. M. (2019). vcf2phylip v2.0: Convert a VCF matrix into several matrix formats for phylogenetic analysis (Versión 2.0) [Software]. Zenodo. https://doi.org/10.5281/zenodo.2540861

Patil, I. (2021). Visualizations with statistical details: The 'ggstatsplot' approach. Journal of Open Source Software, 6(61), 3167. https://doi.org/10.21105/joss.03167

Park, J.-R., Seo, J., Park, S., Jin, M., Jeong, O.-Y., & Park, H.-S. (2023). Identification of potential QTLs related to grain size in rice. Plants, 12(9), 1766. https://doi.org/10.3390/plants12091766

Payne C. T., Zhang F. & Lloyd A. M. (2000). GL3 encodes a bHLH protein that regulates trichome development in Arabidopsis through interaction with GL1 and TTG1. Genetics 156, 1349-1362

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81(3), 559–575. https://doi.org/10.1086/519795

Qi, W., Sun, F., Wang, Q., Chen, M., Huang, Y., Feng, Y.-Q., et al. (2011). Rice ethylene-response AP2/ERF factor OsEATB restricts internode elongation by down-regulating a gibberellin biosynthetic gene. Plant Physiology, 157(1), 216–228. https://doi.org/10.1104/pp.111.179945

R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Raghavan, C., Mauleon, R., Lacorte, V., Jubay, M., Zaw, H., Bonifacio, J., Singh, R. K., Huang, B. E., & Leung, H. (2017). Approaches in Characterizing Genetic Structure and Mapping in a Rice Multiparental Population. G3: Genes, Genomes, Genetics, 7(6), 1721–1730. https://doi.org/10.1534/g3.117.042101

Reig-Valiente JL, Marqués L, Talón M, Domingo C (2018) Genome-wide association study of agronomic traits in rice cultivated in temperate regions. BMC Genomics 19:706. https://doi.org/10.1186/s12864-018-5086-y

Ren, Y., Tian, X., Li, S., Mei, E., He, M., Tang, J., Xu, M., Li, X., Wang, Z., Li, C., & Bu, Q. (2020). Oryza sativa mediator subunit OsMED25 interacts with OsBZR1 to regulate brassinosteroid signalling and plant architecture in rice. Journal of Integrative Plant Biology, 62(6), 793-811. https://doi.org/10.1111/jipb.12914

Rerie W.G., Feldmann K.A., Marks M.D. (1994). The GLABRA2 gene encodes a homeo-domain protein required for normal trichome development in Arabidopsis. Genes & Development 8(12): 1388-1399. https://doi.org/10.1101/gad.8.12.1388

Ruan Q. et al. (2021) Dynamics of dry-matter accumulation indicate source–sink relations during grain filling of *japonica* rice. Field Crops Res. 263: 108009. https://doi.org/10.1016/j.fcr.2020.108009

Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H, Matsuoka M. 2002. Green revolution: a mutant gibberellin-20-oxidase gene in rice. Nature 416: 701-702. https://doi.org/10.1038/416701a

Sakamoto T., Miura K., Itoh H. et al. (2004). An overview of gibberellin-metabolism enzyme genes and their related mutants in rice. Plant Physiology 134, 1642-1653. https://doi.org/10.1104/pp.103.033696

Schellmann S. et al. (2002). TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root-hair patterning in Arabidopsis. EMBO J 21, 5036-5046.

Sha, H., Liu, H., Zhao, G. et al. (2022). Elite sd1 alleles in *japonica* rice and their breeding applications in Northeast China. The Crop Journal, 10, 224-233. https://doi.org/10.1016/j.cj.2021.05.005

Shi, Y., Zhang, Y., Sun, Y., Xie, Z., Luo, Y., Long, Q., Wang, Z., Chen, G., Wang, H., Yu, C., Chen, Y., Zhang, H., Li, C., Zhang, W., Yang, Y., & Lin, Q. (2023). Natural variations of OsAUX5, a target of OsWRKY78, control neutral essential amino-acid content in rice grains. Molecular Plant, 16(2), 322-336. https://doi.org/10.1016/j.molp.2022.12.013

Sun, C., Fang, J., Zhao, T., Xu, B., Zhang, F., Liu, L., Tang, J., Zhang, G., Deng, X., Chen, F., Qian, Q., Cao, X., & Chu, C. (2012). The histone methyltransferase SDG724 mediates H3K36me2/3 deposition at MADS50 and RFT1 and promotes flowering in rice. The Plant Cell, 24(8), 3235-3247. https://doi.org/10.1105/tpc.112.101436

Sun C. et al. (2023) MATE transporter GFD1 cooperates with sugar transporters to control grain-filling duration and grain size in rice. Plant Biotechnol. J. 21: 621-634. https://doi.org/10.1111/pbi.13976

Teng, X., Wang, Y., Liu, L., Yang, H., Zhu, Y., Zhang, R., Jing, J., Lei, J., Li, X., Bao, J., Wan, J. Rice floury endosperm26 (mtSSB1.1) is essential for mitochondrial RNA splicing and endosperm development, increasing dorsal and ventral aleurone thickness. Plant Science 346: 112151 (2024). DOI: 10.1016/j.plantsci.2024.112151

Toriba T. & Hirano H.-Y. (2014). The DROOPING LEAF and OsETTIN2 genes promote awn development in rice. Plant Journal 77, 616-626. https://doi.org/10.1111/tpj.12411

Turner, S. D. (2018). qqman: An R package for visualizing GWAS results using Q-Q and Manhattan plots. Journal of Open Source Software, 3(25), 731. https://doi.org/10.21105/joss.00731

Xin W, Liu H, Yang L, Ma T, Wang J, Zheng H, Liu W, Zou D. 2022. BSA-Seq and fine linkage mapping identify qPH9, a novel locus for mature plant height in rice (Oryza sativa). Rice 15: 26. https://doi.org/10.1186/s12284-022-00576-2

Xue, W., Xing, Y., Weng, X., et al. (2008). Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nature Genetics, 40 (6): 761-767. https://doi.org/10.1038/ng.143

Wang, X., He, Y., Wei, H., & Wang, L. (2021). A clock regulatory module is required for salt tolerance and control of heading date in rice. Plant, Cell & Environment, 44(10), 3283–3301. https://doi.org/10.1111/pce.14167

Wei, H., Wang, X., Xu, H., & Wang, L. (2020). Molecular basis of heading date control in rice. aBIOTECH, 1(4), 219–232. https://doi.org/10.1007/s42994-020-00019-w

Yang, J., Lee, S., Hang, R., Kim, S. R., Lee, Y. S., Cao, X., Amasino, R., & An, G. (2013). OsVIL2 functions with PRC2 to induce flowering by repressing OsLFL1 in rice. The Plant Journal, 73(4), 566–578. https://doi.org/10.1111/tpj.12057

Yang, Y., Fu, D., Zhu, C., He, Y., Zhang, H., Liu, T., Li, X., & Wu, C. (2015). The RING-finger ubiquitin ligase HAF1 mediates Heading date 1 degradation during photoperiodic flowering in rice. The Plant Cell, 27(9), 2455–2468. https://doi.org/10.1105/tpc.15.00320

Yang, Y., Li, J., Li, H., Xu, Z., Qin, R., Wu, W., Wei, P., Ding, Y., & Yang, J. (2021). SDF5 encoding P450 protein is required for internode elongation in rice. Rice Science, 28 (4), 313-316. https://doi.org/10.1016/j.rsci.2021.05.001

Yang, Y., Zhang, Y., Sun, Z., Shen, Z., Li, Y., Guo, Y., Feng, Y., Sun, S., Guo, M., Hu, Z., Yan, C., Chen, X., Zhu, F., & Zhou, J. (2023). Knocking out OsAAP11 to improve rice grain quality using CRISPR/Cas9. International Journal of Molecular Sciences, 24(18), 14360. https://doi.org/10.3390/ijms241814360

Yu B, Lin Z, Li H, Li X, Li J, Wang Y, Zhang X, Zhu Z, Zhai W, Wang X, Xie D, Sun C. 2007. TAC1, a major quantitative trait locus controlling tiller angle in rice. The Plant Journal 52 (5): 891-898. https://doi.org/10.1111/j.1365-313X.2007.03284.x

Yuan H. et al. (2019) OsSPL18 controls grain weight and grain number in rice. J. Genet. Genomics 46: 41-51. https://doi.org/10.1016/j.jgg.2019.01.003

Zhan P. et al. (2021) GW10, a cytochrome P450, regulates grain size and grain number in rice. Theor. Appl. Genet. 134: 3941-3955. https://doi.org/10.1007/s00122-021-03939-3

Zhang, H., Wu, K., Wang, Y., Peng, Y., Hu, F., Wen, L., Han, B., Qian, Q., & Teng, S. (2012). A WUSCHEL-like homeobox gene, OsWOX3B, responds to the NUDA/GL-1 locus in rice. Rice, 5, 30. https://doi.org/10.1186/1939-8433-5-30

Zhao, H., Huai, Z., Xiao, Y., et al. (2014). Natural variation and genetic analysis of the tiller-angle gene MsTAC1 in Miscanthus sinensis. Planta, 240 (1): 161-175. https://doi.org/10.1007/s00425-014-2070-x

Zhou, S., Jiang, L., Duan, C., et al. (2021). Transcriptional and post-transcriptional regulation of heading date in rice. New Phytologist, 230 (3): 943-956. https://doi.org/10.1111/nph.17158

Supplemental Data

Tables

Label	Inferred clusters			Label	Inferred clusters		
	Cluster 1	Cluster 2	Cluster 3		Cluster 1	Cluster 2	Cluster 3
M1	0.323	0.45	0.227	M100	0.043	0.95	0.002
M2	0.584	0.326	0.09	M101	0.383	0.58	0.033
M3	0.608	0.275	0.117	M102	0.612	0.22	0.166
M4	0.452	0.455	0.093	M103	0.355	0.57	0.074
M5	0.224	0.559	0.217	M105	0.25	0	0.749
M7	0.254	0.581	0.165	M106	0.284	0.05	0.664
M8	0.391	0.409	0.2	M107	0.099	0	0.9
M9	0.156	0.633	0.212	M108	0	0	1
M10	0.547	0.349	0.103	M109	0	0	1
M11	0.523	0.472	0.004	M110	0.264	0	0.734
M12	0.354	0.383	0.263	M111	0	0	1
M13	0.701	0.191	0.108	M112	0	0	1
M14	0.489	0.322	0.19	M113	0.697	0.3	0.002
M15	0.639	0.359	0.002	M114	0.535	0.46	0.001
M16	0.144	0.855	0.001	M115	0.699	0.3	0.001
M17	0.576	0.401	0.023	M116	0.54	0.33	0.125
M19	0.255	0.743	0.002	M117	0.535	0.45	0.013
M20	0.326	0.67	0.003	M118	0.407	0.54	0.05
M21	0.591	0.365	0.045	M119	0.614	0.36	0.024
M22	0.338	0.565	0.097	M120	0.533	0.46	0.001
M24	0.087	0.912	0.002	M121	0.447	0.28	0.267
M25	0.787	0.183	0.03	M122	0.564	0.43	0.001
M26	0.034	0.953	0.013	M123	0.575	0.42	0.001
M27	0.303	0.695	0.001	M124	0.78	0.2	0.014
M28	0.205	0.794	0	M125	0.459	0.44	0.094
M29	0.608	0.36	0.033	M126	0.691	0.3	0.005
M30	0.593	0.319	0.088	M127	0.671	0.32	0.001
M31	0.178	0.814	0.008	M128	0.317	0.68	0.001
M32	0.11	0.889	0.001	M129	0.332	0.5	0.167
M33	0.395	0.48	0.125	M130	0.651	0.34	0.002
M34	0.004	0.891	0.106	M131	0.714	0.27	0.013
M35	0.119	0.881	0.001	M132	0.442	0.55	0.002
M36	0.092	0.907	0.001	M133	0.636	0.36	0
M38	0.004	0.961	0.036	M134	0.001	0	0.999
M40	0.003	0.994	0.003	M135	0.268	0	0.729
M42	0.151	0.848	0.001	M136	0	0	1
M43	0.124	0.875	0.001	M137	0.001	0	0.999
M44	0.157	0.842	0.001	M138	0.354	0	0.645
M45	0.028	0.859	0.113	M139	0	0	0.999
M46	0.393	0.443	0.164	M140	0.154	0.84	0.001
M48	0.863	0.017	0.12	M141	0.204	0.76	0.033
M49	0.499	0.425	0.076	M142	0.435	0.54	0.02
M50	0.808	0.08	0.112	M143	0.707	0.29	0.001

M51	0.996	0.002	0.002	M144	0.002	0.91	0.085
M52	0.999	0.001	0.001	M145	0.998	0	0
M53	0.407	0.463	0.13	M146	0.275	0.6	0.125
M54	0.861	0.135	0.004	M147	0.545	0.25	0.2
M55	0.558	0.291	0.151	M148	0.909	0.09	0.001
M56	0.93	0.008	0.062	M149	0.665	0.33	0.002
M58	0.893	0.006	0.102	M150	0.497	0.4	0.102
M59	0.999	0	0.001	M151	0.759	0.23	0.002
M60	0.747	0.153	0.1	M152	0.53	0.46	0.001
M61	0.764	0.001	0.235	M153	0.659	0.11	0.222
M62	0.759	0.03	0.212	M154	0.972	0	0.025
M63	0.786	0.001	0.213	M156	0.998	0	0.001
M64	0.942	0.002	0.057	M157	0.941	0	0.054
M65	0.763	0.012	0.226	M158	0.838	0	0.161
M66	0.971	0.002	0.027	M159	0.923	0	0.072
M67	0.973	0.008	0.02	M160	0.839	0	0.16
M68	0.819	0.008	0.173	M162	0.847	0	0.152
M69	0.731	0.02	0.249	M163	0.729	0	0.265
M71	0.946	0.001	0.053	M164	0.894	0.01	0.091
M73	0.805	0.001	0.194	M165	0.738	0.26	0.002
M74	0.66	0.153	0.188	M166	0.393	0.35	0.257
M75	0.999	0	0	M167	0.345	0.65	0.001
M76	0.999	0.001	0	M169	0.812	0.18	0.001
M77	0.961	0.001	0.038	M170	0.29	0.7	0.001
M78	0.999	0.001	0.001	M171	0.581	0.28	0.138
M79	0.996	0.003	0.001	M172	0.82	0.17	0.006
M80	0.979	0.003	0.017	M173	0.713	0.28	0.001
M81	0.986	0.013	0.001	M174	0.296	0.69	0.006
M82	0.999	0.001	0.001	M175	0.583	0.24	0.177
M83	0.998	0.001	0.001	M176	0.41	0.55	0.039
M84	0.999	0.001	0	M177	0.721	0.16	0.119
M85	0.999	0.001	0	M178	0.909	0.06	0.024
M86	0.033	0.965	0.002	M179	0.394	0.52	0.084
M87	0.999	0.001	0	M180	0.705	0	0.294
M88	0.998	0.001	0.001	M181	0.501	0.23	0.268
M89	0.999	0.001	0	M182	0.985	0	0.013
M90	0.001	0.859	0.14	M183	0.089	0.91	0.001
M91	0.113	0.781	0.106	M184	0.002	0.96	0.038
M92	0.046	0.858	0.097	M185	0.598	0.4	0
M93	0.003	0.938	0.059	M186	0.986	0	0.012
M94	0.056	0.914	0.03	M187	0.96	0.03	0.001
M95	0.158	0.788	0.055	M188	0.126	0.87	0
M96	0.03	0.838	0.132	M189	0.999	0	0
M97	0.186	0.814	0	M190	0.997	0	0.003
M98	0.343	0.637	0.02	M191	0	0.75	0.246
M99	0.109	0.745	0.145	M192	0.29	0.41	0.294

Supplementary Table 1. Individual admixture coefficients (Q-values) from STRUCTURE analysis with K = 3 for the 192 MAGIC lines. The three columns give the proportion of its genome assigned to Cluster 1, Cluster 2 or Cluster 3.

Label	Inferred clusters							
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**Cluster 7**	**Cluster 8**
M1	0.016	0.03	0.001	0	0.018	0.594	0.182	0.159
M2	0.094	0.136	0.001	0.004	0.189	0.517	0.056	0.002
M3	0.011	0	0	0	0.311	0.519	0.068	0.09
M4	0	0.066	0.031	0.001	0.151	0.747	0.003	0.001
M5	0.017	0	0.019	0	0.069	0.599	0.18	0.116
M7	0.001	0.011	0.001	0	0.002	0.916	0.068	0.001
M8	0.007	0.168	0.108	0.16	0.003	0.365	0.176	0.013
M9	0.057	0.166	0.107	0.029	0	0.355	0.156	0.129
M10	0.001	0.002	0.016	0	0.226	0.75	0.004	0.001
M11	0.001	0.001	0.002	0	0.126	0.869	0.001	0.001
M12	0	0.001	0.001	0	0.053	0.619	0.225	0.101
M13	0.001	0.001	0	0	0.33	0.608	0.027	0.032
M14	0.082	0.001	0	0	0.233	0.406	0.144	0.132
M15	0.103	0.003	0	0	0.342	0.505	0	0.047
M16	0.91	0	0	0	0.088	0.001	0	0
M17	0.496	0	0	0	0.502	0	0	0.001
M19	0.94	0.001	0	0.005	0.054	0	0	0
M20	0.843	0	0	0.001	0.155	0	0	0
M21	0.403	0.004	0.002	0.004	0.467	0.02	0.026	0.075
M22	0.286	0	0.251	0.001	0.219	0.156	0.086	0.001
M24	0.992	0	0.004	0	0.001	0.001	0.001	0
M25	0.231	0	0.021	0.081	0.577	0.074	0.014	0.001
M26	0.998	0	0	0	0	0	0	0
M27	0.797	0	0	0.001	0.201	0	0	0
M28	0.884	0	0	0	0.114	0	0	0
M29	0.282	0	0.127	0.003	0.472	0.09	0.025	0
M30	0.226	0	0.131	0.001	0.463	0.109	0.069	0.001
M31	0	0.001	0.843	0	0.154	0	0.001	0
M32	0	0	0.998	0	0	0	0	0
M33	0	0.092	0.435	0.002	0.245	0.117	0.109	0.001
M34	0	0	0.796	0	0.001	0	0.09	0.112
M35	0	0	0.998	0	0	0	0	0
M36	0	0	0.998	0.001	0	0	0	0
M38	0	0	0.917	0.003	0	0	0.017	0.062
M40	0	0	0.998	0	0	0	0	0
M42	0	0	0.999	0	0	0	0	0
M43	0.001	0.001	0.768	0	0.009	0.222	0.001	0
M44	0.012	0	0.978	0.004	0.001	0.004	0	0
M45	0.001	0	0.728	0	0.002	0.021	0.119	0.128
M46	0.005	0.075	0.119	0.071	0.121	0.399	0.12	0.089
M48	0	0	0	0.054	0.614	0.263	0.066	0.001
M49	0	0.166	0.266	0.01	0.074	0.419	0.064	0
M50	0	0.034	0.021	0.076	0.473	0.259	0.085	0.052
M51	0	0.002	0	0	0.676	0.32	0	0

M52	0	0.006	0	0	0.755	0.238	0	0
M53	0.101	0.033	0.137	0.046	0.133	0.455	0.093	0.001
M54	0.003	0.069	0	0.087	0.435	0.404	0	0.001
M55	0.002	0.001	0.117	0.066	0.142	0.591	0.079	0.002
M56	0	0.014	0	0.081	0.749	0.102	0.005	0.048
M58	0	0.019	0.001	0.014	0.664	0.173	0.071	0.058
M59	0	0	0	0	0.999	0	0	0
M60	0.053	0.005	0.001	0.352	0.25	0.289	0.049	0
M61	0	0.001	0	0.587	0.3	0.001	0.11	0
M62	0	0.002	0.002	0.767	0.088	0.081	0.06	0
M63	0	0	0	0.827	0.083	0	0.088	0
M64	0.003	0.042	0	0.5	0.445	0	0.008	0
M65	0.001	0.127	0	0.606	0.205	0.001	0.058	0.002
M66	0.002	0.021	0	0.537	0.438	0	0	0
M67	0.006	0	0.001	0.646	0.346	0	0.001	0
M68	0.001	0	0.009	0.8	0.184	0	0.005	0
M69	0	0	0.02	0.672	0.227	0	0.081	0
M71	0	0.001	0.001	0.265	0.679	0.017	0.037	0.001
M73	0.001	0.001	0	0.681	0.316	0	0.001	0
M74	0	0	0.101	0.892	0	0	0.005	0
M75	0	0	0	0	0.999	0	0	0
M76	0	0	0	0	0.996	0.001	0	0.001
M77	0	0.001	0	0.001	0.818	0.161	0.018	0
M78	0	0	0	0	0.998	0	0	0
M79	0.055	0.003	0.001	0	0.938	0.001	0	0
M80	0	0.012	0.062	0.001	0.899	0.011	0.013	0.002
M81	0	0	0.075	0	0.922	0.001	0	0
M82	0	0.002	0	0	0.996	0	0	0
M83	0	0.001	0.003	0	0.994	0.001	0	0.001
M84	0	0	0	0	0.999	0	0	0
M85	0	0	0.001	0.053	0.945	0.001	0	0
M86	0	0	0	0	0	0	0	0.998
M87	0	0	0	0	0.998	0	0	0
M88	0.001	0.016	0	0	0.982	0.001	0	0.001
M89	0	0	0.004	0.022	0.973	0	0	0
M90	0	0.001	0.081	0.001	0	0	0	0.917
M91	0.183	0.004	0.123	0.001	0.002	0.365	0.055	0.268
M92	0	0.001	0.002	0	0.073	0	0	0.924
M93	0	0.001	0.003	0.001	0.001	0.12	0.001	0.873
M94	0.047	0.114	0.184	0	0.004	0.085	0.014	0.553
M95	0.127	0.001	0.129	0	0.014	0.356	0.004	0.369
M96	0.055	0.092	0.075	0	0.001	0.375	0.064	0.338
M97	0.003	0.068	0.308	0.075	0.101	0.002	0	0.444
M98	0	0.051	0.062	0.07	0.268	0.005	0	0.544
M99	0	0	0	0.002	0.055	0	0.001	0.942
M100	0	0	0	0	0.001	0	0	0.998

M101	0.001	0.006	0.103	0.028	0.353	0.016	0	0.492
M102	0.296	0	0.023	0.003	0.398	0.134	0.146	0
M103	0.159	0.153	0.01	0.012	0.161	0.455	0.048	0.001
M105	0	0.001	0.001	0.318	0.002	0	0.678	0
M106	0	0.001	0.056	0	0.271	0.001	0.662	0.008
M107	0	0	0	0.011	0.093	0	0.895	0.001
M108	0	0	0	0	0	0	0.999	0
M109	0	0	0	0	0	0	0.999	0
M110	0	0	0.109	0.168	0.008	0	0.713	0.001
M111	0	0	0	0	0	0	0.999	0
M112	0	0	0	0	0	0	0.999	0
M113	0	0.877	0	0	0.121	0	0	0
M114	0	0.945	0	0.001	0.052	0	0	0
M115	0	0.814	0	0	0.184	0	0	0
M116	0.161	0.004	0.144	0.058	0.314	0.212	0.107	0
M117	0.26	0	0.156	0.001	0.446	0.133	0.001	0.003
M118	0.708	0.001	0	0	0.288	0.001	0.002	0
M119	0.001	0.48	0.033	0.133	0.156	0.171	0.008	0.019
M120	0	0.999	0	0	0	0	0	0
M121	0.126	0	0.131	0.373	0.126	0.066	0.176	0.001
M122	0	0.999	0	0	0	0	0	0
M123	0	0.999	0	0	0	0	0	0
M124	0	0.784	0	0	0.214	0	0	0
M125	0.001	0.358	0.07	0.221	0.006	0.085	0.035	0.224
M126	0.025	0.255	0.011	0.086	0.374	0.226	0.001	0.022
M127	0	0.932	0	0	0.066	0	0	0
M128	0	0.998	0	0	0	0	0	0
M129	0.052	0.297	0.008	0.085	0.035	0.246	0.14	0.135
M130	0	0.832	0	0.001	0.165	0.001	0	0
M131	0	0.82	0	0	0.179	0	0	0
M132	0	0.999	0	0	0	0	0	0
M133	0	0.995	0	0	0.003	0.001	0	0
M134	0	0	0	0	0	0	0.999	0
M135	0.001	0	0.117	0.146	0.043	0	0.693	0.001
M136	0	0	0	0	0	0	0.999	0
M137	0	0	0	0	0.001	0	0.998	0
M138	0.001	0	0	0	0.346	0	0.641	0.01
M139	0	0	0	0	0	0	0.999	0
M140	0.209	0.047	0.174	0.045	0.001	0.31	0	0.214
M141	0.12	0	0.166	0.004	0.001	0.707	0.001	0.001
M142	0.077	0.052	0.018	0.001	0.159	0.423	0.001	0.27
M143	0.113	0	0.028	0	0.342	0.515	0	0.001
M144	0.005	0	0.004	0	0.002	0.002	0.001	0.985
M145	0	0	0.001	0	0.876	0.121	0	0
M146	0.002	0.004	0.003	0.239	0.001	0.424	0.003	0.324
M147	0	0	0	0.244	0.001	0.694	0.059	0

M148	0.068	0.006	0.1	0.253	0.565	0.006	0	0.001
M149	0.001	0	0.069	0.213	0.007	0.709	0	0
M150	0.092	0	0	0	0.261	0.411	0.069	0.165
M151	0.001	0	0	0.113	0.138	0.747	0	0
M152	0.001	0.028	0.106	0.001	0.134	0.729	0.001	0
M153	0	0	0	0.457	0.001	0.538	0.004	0
M154	0.094	0	0	0.001	0.895	0	0.005	0.004
M156	0	0.075	0	0	0.8	0.091	0	0.033
M157	0.07	0	0.123	0.113	0.69	0.001	0.003	0
M158	0	0	0.002	0.001	0.808	0.001	0.127	0.061
M159	0.001	0	0.004	0.178	0.519	0.284	0.013	0.002
M160	0	0.002	0	0.005	0.771	0	0.052	0.169
M162	0	0.009	0.007	0.278	0.589	0	0.116	0.001
M163	0	0.015	0.016	0.002	0.62	0.001	0.181	0.166
M164	0.145	0	0.1	0.1	0.624	0.018	0.012	0
M165	0	0	0	0.204	0.014	0.781	0	0
M166	0.006	0.084	0.002	0.072	0.077	0.496	0.22	0.043
M167	0.125	0	0.192	0.004	0.001	0.677	0	0.001
M169	0.003	0.004	0.003	0.004	0.56	0.32	0	0.106
M170	0.036	0.002	0.005	0.099	0.001	0.581	0	0.276
M171	0.007	0.002	0.002	0.006	0.003	0.903	0.077	0.001
M172	0	0	0	0.001	0.169	0.829	0	0
M173	0	0	0	0	0.175	0.822	0	0
M174	0.1	0.04	0.187	0.106	0.001	0.464	0.001	0.101
M175	0	0	0	0.307	0.152	0.504	0.002	0.034
M176	0.005	0.086	0.069	0.001	0.005	0.686	0	0.148
M177	0	0	0.001	0.251	0.297	0.45	0	0
M178	0.095	0.002	0.063	0.001	0.827	0.001	0.002	0.01
M179	0.016	0.08	0.166	0.002	0.145	0.463	0.058	0.071
M180	0	0.001	0	0.04	0.595	0	0.203	0.159
M181	0	0	0	0.191	0.015	0.601	0.147	0.046
M182	0.002	0.001	0.012	0.001	0.837	0.146	0.001	0.001
M183	0.296	0.158	0.168	0	0.001	0.265	0	0.112
M184	0.032	0.018	0.372	0.002	0.001	0.096	0.001	0.48
M185	0.047	0.099	0.311	0.12	0.277	0.145	0	0.001
M186	0.066	0.001	0	0.001	0.692	0.238	0.001	0.001
M187	0.109	0.001	0.003	0.289	0.312	0.286	0	0
M188	0.001	0.11	0.187	0	0.001	0.256	0	0.445
M189	0	0.025	0	0.063	0.911	0	0	0
M190	0	0	0	0	0.998	0	0.001	0
M191	0.175	0.085	0.05	0.099	0	0.255	0.201	0.135
M192	0.093	0.011	0.181	0.082	0.153	0.001	0.222	0.258

Supplementary Table 2. Individual admixture coefficients (Q-values) from STRUCTURE analysis with K = 8 for the 192 MAGIC lines. The eigth columns give the proportion of its genome assigned to Cluster 1 to Cluster 8.

k3	Fixation Index
Fst_1	0.366
Fst_2	0.1211
Fst_3	0.83

k8	Fixation Index
Fst_1	0.6269
Fst_2	0.5971
Fst_3	0.5416
Fst_4	0.6765
Fst_5	0.4554
Fst_6	0.3326
Fst_7	0.8354
Fst_8	0.5322

Supplementary Table 3. Cluster-specific fixation indices from STRUCTURE analyses run with K = 3 and K = 8.

qH1; 98 SNPs					
SNP	p-value	SNP	p-value	SNP	p-value
S01_38246820	1.25E-06	S01_38845476	3.51E-06	S01_38844574	3.43E-06
S01_38250588	2.66E-06	S01_38845481	4.47E-06	S01_38845476	3.51E-06
S01_38255448	1.16E-06	S01_38845497	4.47E-06	S01_38845481	4.47E-06
S01_38256239	1.26E-06	S01_38845498	4.51E-06	S01_38845497	4.47E-06
S01_38259630	1.35E-06	S01_38845509	4.12E-06	S01_38845498	4.51E-06
S01_38280069	1.30E-06	S01_38845520	4.12E-06	S01_38845509	4.12E-06
S01_38292063	2.68E-06	S01_38862108	2.06E-05	S01_38845520	4.12E-06
S01_38292070	2.28E-06	S01_38862164	9.31E-06	S01_38862108	2.06E-05
S01_38292106	2.62E-06	S01_38864506	4.89E-06	S01_38862164	9.31E-06
S01_38333394	7.54E-07	S01_38884322	4.87E-06	S01_38864506	4.89E-06
S01_38360521	9.62E-07	S01_38884327	4.97E-06	S01_38884322	4.87E-06
S01_38360570	8.85E-07	S01_38886438	2.75E-06	S01_38884327	4.97E-06
S01_38363181	2.01E-06	S01_38887927	3.10E-06	S01_38886438	2.75E-06
S01_38363322	1.26E-06	S01_38888223	7.75E-07	S01_38887927	3.10E-06
S01_38363350	1.28E-06	S01_38903747	6.74E-07	S01_38888223	7.75E-07
S01_38366399	1.37E-06	S01_38903783	7.66E-07	S01_38903747	6.74E-07
S01_38376176	7.50E-07	S01_38927698	9.36E-07	S01_38903783	7.66E-07
S01_38393797	9.26E-07	S01_38927767	4.03E-06	S01_38927698	9.36E-07
S01_38396508	9.56E-07	S01_38947797	1.03E-06	S01_38927767	4.03E-06
S01_38427484	4.88E-06	S01_38950723	8.00E-07	S01_38947797	1.03E-06
S01_38461076	4.10E-06	S01_38950724	8.00E-07	S01_38950723	8.00E-07
S01_38477416	1.32E-06	S01_38952044	1.77E-06	S01_38950724	8.00E-07
S01_38524922	1.77E-06	S01_38994051	7.67E-07	S01_38952044	1.77E-06
S01_38529442	3.50E-06	S01_39008576	7.51E-07	S01_38994051	7.67E-07
S01_38560918	1.34E-06	S01_39011296	2.75E-06	S01_39008576	7.51E-07
S01_38560944	1.28E-06	S01_39024118	9.01E-07	S01_39011296	2.75E-06

qH5; 10 SNPs	
SNP	p-value
S05_18408140	4.68E-05
S05_18539722	2.01E-05
S05_18541982	1.98E-05
S05_18542763	2.37E-05
S05_18617899	1.79E-05
S05_18670300	6.23E-05
S05_18671553	3.73E-05
S05_18926193	2.41E-05
S05_18932162	3.72E-05
S05_19518444	9.22E-05

qH9; 20 SNPs	
SNP	p-value
S09_14659472	7.55E-05
S09_14699904	6.89E-05
S09_14814204	5.07E-05
S09_15202160	9.58E-05
S09_15219071	2.11E-05
S09_15225485	4.86E-05
S09_15239028	2.61E-05
S09_15239702	3.16E-05
S09_15251527	8.90E-05
S09_15261427	2.45E-05
S09_15266111	2.38E-05
S09_16515318	7.26E-07
S09_16527300	1.84E-05

S01_38560978	1.37E-06	S01_39035622	4.53E-07	S01_39024118	9.01E-07	S09_17104658	1.37E-07
S01_38658840	2.34E-06	S01_39035849	9.03E-07	S01_39035622	4.53E-07	S09_17115951	2.25E-08
S01_38659138	2.52E-06	S01_39039451	4.46E-06	S01_39035849	9.03E-07	S09_17115954	3.69E-08
S01_38666163	5.99E-06	S01_39041519	1.42E-06	S01_39039451	4.46E-06	S09_17116031	6.39E-08
S01_38682982	6.22E-06	S01_39048059	8.41E-07	S01_39041519	1.42E-06	S09_17116071	5.96E-07
S01_38694199	4.58E-06	S01_39210834	4.47E-05	S01_39048059	8.41E-07	S09_17402988	1.19E-05
S01_38694447	4.42E-06	S01_39210931	2.93E-05	S01_39210834	4.47E-05	S09_17424519	6.51E-06
S01_38694479	3.78E-06	S01_39221825	3.61E-05	S01_39210931	2.93E-05		
S01_38695376	2.58E-06	S01_39287321	1.79E-05	S01_39221825	3.61E-05		
S01_38697146	3.18E-06	S01_39288060	3.55E-05	S01_39287321	1.79E-05		
S01_38722518	5.40E-06	S01_39303635	1.61E-05	S01_39288060	3.55E-05		
S01_38757190	4.32E-06	S01_39317670	6.48E-05	S01_39303635	1.61E-05		
S01_38772039	2.45E-06	S01_39347304	3.08E-05	S01_39317670	6.48E-05		
S01_38787567	1.85E-06	S01_39385547	4.13E-05	S01_39347304	3.08E-05		
S01_38788831	3.11E-06	S01_39415573	5.62E-05	S01_39385547	4.13E-05		
S01_38799281	2.83E-06	S01_39536116	3.02E-05	S01_39415573	5.62E-05		
S01_38799320	2.17E-06	S01_39576118	3.96E-05	S01_39536116	3.02E-05		
S01_38804999	8.49E-06	S01_39682575	7.13E-05	S01_39576118	3.96E-05		
S01_38807136	7.99E-06	S01_39682583	7.36E-05	S01_39682575	7.13E-05		
S01_38819507	2.63E-06	S01_39682591	7.49E-05	S01_39682583	7.36E-05		
S01_38841537	6.94E-06	S01_39682666	3.25E-05	S01_39682591	7.49E-05		
S01_38844446	6.18E-06	S01_39695518	5.24E-05	S01_39682666	3.25E-05		
S01_38844490	8.25E-06	S01_38844446	6.18E-06	S01_39695518	5.24E-05		
S01_38844574	3.43E-06	S01_38844490	8.25E-06				

Supplementary Table 5. List of SNPs included in each QTL associated with **Heigth.** Each QTL is labelled with its name, and the total number of associated SNPs. SNPs are denoted by chromosome and physical position (chr_pos format), along with the p-value obtained from the GWAS. The SNP with the lowest p-value within each QTL region is highlighted as the lead SNP.

qAwn8; 7 SNPs	
SNP	**p-value**
S08_22480705	5.97E-04
S08_22922477	3.48E-04
S08_24102764	3.66E-04
S08_24174803	4.61E-04
S08_24855823	1.37E-04
S08_24857921	5.55E-05
S08_25297626	3.98E-04

qAwn12; 3 SNPs	
SNP	**p-value**
S12_808650	7.78E-04
S12_1039367	6.06E-04
S12_1835386	2.42E-05

Supplementary Table 6. List of SNPs included in each QTL associated with **Awn.** Each QTL is labelled with its name, and the total number of associated SNPs. SNPs are denoted by chromosome and physical position (chr_pos format), along with the p-value obtained from the GWAS. The SNP with the lowest p-value within each QTL region is highlighted as the lead SNP.

qGlb5; 4 SNPs	
SNP	p-value
S05_212132	3.58E-06
S05_841214	4.10E-05
S05_1440403	5.95E-07
S05_2240956	7.36E-07

qGlb7; 25 SNPs			
SNP	p-value	SNP	p-value
S07_1836005	7.00E-05	S07_19630036	4.65E-05
S07_19027731	7.44E-05	S07_19633513	9.03E-06
S07_19345161	4.40E-05	S07_19916604	9.03E-05
S07_19354433	5.39E-05	S07_19974715	8.61E-05
S07_19397119	6.04E-05	S07_20003931	6.49E-05
S07_19413896	3.28E-05	S07_20005699	8.60E-05
S07_19446758	5.32E-05	S07_20007651	9.51E-05
S07_19458249	3.93E-05	S07_20013577	8.34E-05
S07_19471091	3.33E-05	S07_20082072	8.90E-05
S07_19496651	6.75E-05	S07_20206574	8.67E-05
S07_19548816	7.52E-05	S07_20248514	6.49E-05
S07_19581886	7.65E-05	S07_20610304	8.50E-05
S07_19622996	8.18E-05		

Supplementary Table 7. List of SNPs included in each QTL associated with **Glb.** Each QTL is labelled with its name, and the total number of associated SNPs. SNPs are denoted by chromosome and physical position (chr_pos format), along with the p-value obtained from the GWAS. The SNP with the lowest p-value within each QTL region is highlighted as the lead SNP.

qFLA5; 12 SNPs	
SNP	p-value
S05_18971084	6.24E-04
S05_18983701	6.99E-04
S05_18983628	7.43E-04
S05_19102150	7.52E-04
S05_19045581	7.60E-04
S05_19105673	7.70E-04
S05_19033771	8.75E-04
S05_18961490	8.96E-04
S05_19033919	9.40E-04
S05_18806626	9.66E-04
S05_18961090	9.98E-04
S05_19108286	1.00E-03

qFLA; 113 SNPs					
SNP	p-value	SNP	p-value	SNP	p-value
S09_15777388	2.09E-05	S09_16719725	1.66E-04	S09_15631873	5.00E-04
S09_15675851	2.18E-05	S09_15857943	1.80E-04	S09_16527347	5.04E-04
S09_16005453	2.86E-05	S09_15926638	1.81E-04	S09_16353373	5.24E-04
S09_15982621	3.69E-05	S09_15890240	1.82E-04	S09_16577876	5.28E-04
S09_16268904	3.70E-05	S09_16231460	2.07E-04	S09_15551344	5.31E-04
S09_15971211	3.79E-05	S09_15920564	2.08E-04	S09_16494233	5.37E-04
S09_15929907	3.96E-05	S09_16346483	2.25E-04	S09_16405717	5.42E-04
S09_15762928	4.79E-05	S09_15913533	2.25E-04	S09_16385947	5.44E-04
S09_16084327	4.82E-05	S09_15773271	2.26E-04	S09_16420378	5.63E-04
S09_16110426	5.26E-05	S09_16315365	2.44E-04	S09_15234847	5.66E-04
S09_16084351	5.28E-05	S09_17281984	2.44E-04	S09_15651071	5.72E-04
S09_15768405	5.47E-05	S09_15890284	2.60E-04	S09_16515393	6.06E-04
S09_15746452	5.47E-05	S09_16527348	2.64E-04	S09_16476033	6.09E-04
S09_15747477	5.48E-05	S09_16643465	2.65E-04	S09_15616078	6.15E-04
S09_16012359	5.59E-05	S09_16670495	2.79E-04	S09_15551341	6.23E-04
S09_16012363	5.59E-05	S09_16476072	2.83E-04	S09_15416016	6.24E-04
S09_15825930	5.97E-05	S09_16715937	2.87E-04	S09_16487060	6.45E-04
S09_15839941	6.00E-05	S09_16279459	3.04E-04	S09_16527341	6.75E-04

S09_15839964	6.36E-05	S09_16695642	3.06E-04	S09_16476038	6.93E-04
S09_15771225	6.67E-05	S09_16363939	3.09E-04	S09_16807434	6.99E-04
S09_15789995	7.03E-05	S09_16367225	3.12E-04	S09_16315388	7.08E-04
S09_15857977	8.16E-05	S09_16405678	3.15E-04	S09_16390597	7.12E-04
S09_15780886	8.82E-05	S09_16670223	3.28E-04	S09_16527280	7.40E-04
S09_15742088	9.13E-05	S09_16737057	3.33E-04	S09_15470260	7.55E-04
S09_15773245	9.81E-05	S09_15651078	3.46E-04	S09_16527374	7.68E-04
S09_15725474	1.00E-04	S09_16672086	3.52E-04	S09_15251527	8.01E-04
S09_15742116	1.04E-04	S09_16672089	3.52E-04	S09_16300493	8.21E-04
S09_15929897	1.07E-04	S09_16273253	3.62E-04	S09_16807430	8.34E-04
S09_15825701	1.11E-04	S09_15551359	3.93E-04	S09_15461849	8.43E-04
S09_16265288	1.21E-04	S09_16577816	4.04E-04	S09_15489293	8.56E-04
S09_15987288	1.27E-04	S09_16578151	4.19E-04	S09_15605816	8.95E-04
S09_15890307	1.31E-04	S09_16392582	4.31E-04	S09_16515328	9.09E-04
S09_16338784	1.33E-04	S09_16475086	4.42E-04	S09_15655372	9.14E-04
S09_15890275	1.43E-04	S09_16428508	4.55E-04	S09_15490510	9.38E-04
S09_16249236	1.56E-04	S09_16467669	4.71E-04	S09_19595466	9.43E-04
S09_16652601	1.60E-04	S09_15596662	4.80E-04	S09_15482742	9.78E-04
S09_15918022	1.62E-04	S09_16848939	4.88E-04	S09_15621159	9.87E-04
S09_16250792	1.65E-04	S09_16539340	4.94E-04		

Supplementary Table 8. List of SNPs included in each QTL associated with **FLA.** Each QTL is labelled with its name, and the total number of associated SNPs. SNPs are denoted by chromosome and physical position (chr_pos format), along with the p-value obtained from the GWAS. The SNP with the lowest p-value within each QTL region is highlighted as the lead SNP.

qClm4; 2 SNPs	
SNP	**p-value**
S04_19546961	9.60E-07
S04_20315786	3.87E-05

qClm9; 97 SNPs					
SNP	**p-value**	**SNP**	**p-value**	**SNP**	**p-value**
S09_18053862	8.92E-06	S09_19472151	4.75E-09	S09_21352944	4.03E-10
S09_18474631	8.46E-06	S09_19593496	4.90E-09	S09_21358111	1.09E-09
S09_18474840	5.59E-05	S09_19595466	2.20E-09	S09_21413137	2.97E-09
S09_18488743	5.83E-05	S09_19698287	3.38E-09	S09_21413288	3.79E-08
S09_18501674	3.16E-05	S09_19747464	3.12E-09	S09_21816378	6.14E-07
S09_18511109	3.72E-05	S09_19805897	7.47E-11	S09_21851006	5.32E-07
S09_18549530	8.17E-05	S09_20030479	1.64E-10	S09_21851018	5.09E-07
S09_18554353	3.45E-05	S09_20217591	1.72E-10	S09_21887727	9.52E-07
S09_18608866	9.72E-05	S09_20520231	1.26E-12	S09_21887760	6.05E-07
S09_18652166	5.60E-05	S09_20563155	4.73E-12	S09_21887840	3.02E-07
S09_18711122	1.66E-05	S09_20580677	9.46E-13	S09_21893613	2.38E-06
S09_18715803	1.62E-05	S09_20722015	6.83E-12	S09_21893614	2.38E-06
S09_18744632	1.79E-05	S09_20722044	1.30E-12	S09_21902312	4.60E-06
S09_18744659	5.07E-06	S09_20728918	1.57E-12	S09_21953159	3.18E-07

S09_18745843	4.43E-06	S09_20756104	6.35E-13	S09_21979833	3.69E-07
S09_18747444	7.55E-06	S09_20756118	6.88E-13	S09_22016072	8.07E-08
S09_18772434	3.14E-06	S09_20756479	4.87E-13	S09_22190746	2.65E-07
S09_18822616	5.24E-06	S09_20756572	8.52E-13	S09_22236656	2.57E-07
S09_18914906	9.93E-05	S09_20820536	3.24E-12	S09_22248045	2.56E-08
S09_19037591	4.67E-07	S09_20820542	2.28E-11	S09_22280308	5.79E-08
S09_19037607	1.32E-07	S09_20820543	1.18E-11	S09_22292148	5.11E-08
S09_19099340	4.61E-08	S09_20847134	1.05E-11	S09_22307477	4.06E-08
S09_19142021	1.90E-07	S09_21036979	6.85E-11	S09_22314703	1.68E-08
S09_19165498	1.54E-07	S09_21036988	4.83E-11	S09_22314753	1.96E-08
S09_19168448	1.75E-07	S09_21176469	3.31E-11	S09_22359467	3.07E-07
S09_19313189	1.52E-08	S09_21212206	6.02E-11	S09_22387349	5.40E-08
S09_19401530	4.78E-08	S09_21238264	8.50E-12	S09_22486366	7.80E-08
S09_19401536	4.78E-08	S09_21248349	1.05E-10	S09_22689864	2.27E-07
S09_19401659	1.20E-07	S09_21281265	3.67E-11	S09_22799926	4.43E-07
S09_19401666	4.02E-07	S09_21306737	6.49E-10	S09_22802873	3.69E-06
S09_19408468	3.70E-08	S09_21307080	1.26E-10	S09_22802876	3.49E-07
S09_19438840	4.01E-08	S09_21327053	2.11E-10		
S09_19454656	3.21E-09	S09_21352924	1.13E-09		

Supplementary Table 9. List of SNPs included in each QTL associated with **Clm.** Each QTL is labelled with its name, and the total number of associated SNPs. SNPs are denoted by chromosome and physical position (chr_pos format), along with the p-value obtained from the GWAS. The SNP with the lowest p-value within each QTL region is highlighted as the lead SNP.

qHd3; 21 SNPs	
SNP	**p-value**
S03_1094461	9.34E-07
S03_8874823	9.58E-04
S03_9093488	7.48E-04
S03_9140692	9.99E-04
S03_16289975	9.80E-04
S03_17168984	1.65E-04
S03_17253276	1.00E-04
S03_17268333	3.70E-04
S03_17346596	4.99E-04
S03_17349682	1.23E-04
S03_17390209	1.37E-04
S03_17394895	3.49E-04
S03_17409272	7.09E-05
S03_17573702	1.20E-04

qHd4; 6 SNPs	
SNP	**p-value**
S04_33773923	5.29E-04
S04_33787498	2.79E-04
S04_33788988	7.29E-04
S04_33799481	3.61E-04
S04_33799486	5.74E-04
S04_35135602	6.48E-04

qHd6; 1 SNPs	
SNP	**p-value**
S06_14170873	9.20E-04

qHd7; 2 SNPs	
SNP	**p-value**
S07_19654246	8.92E-04

qHd8; 9 SNPs	
SNP	**p-value**
S08_4351455	2.24E-04
S08_4532920	3.55E-04
S08_4630253	8.78E-04
S08_4650515	8.02E-04
S08_4719380	7.59E-04
S08_4893202	7.93E-04
S08_4893204	8.71E-04
S08_4901936	8.55E-04
S08_5133187	5.32E-04

qHd9; 1 SNPs	
SNP	**p-value**
S09_7974566	7.30E-04

S03_17594381	2.25E-04	S07_19654269	6.85E-04
S03_19135756	2.88E-04		
S03_19451712	3.77E-04		
S03_20299395	1.34E-04		
S03_20357302	1.72E-04		
S03_20971667	2.27E-04		
S03_21288080	3.83E-04		

Supplementary Table 10. List of SNPs included in each QTL associated with **Hd.** Each QTL is labelled with its name, and the total number of associated SNPs. SNPs are denoted by chromosome and physical position (chr_pos format), along with the p-value obtained from the GWAS. The SNP with the lowest p-value within each QTL region is highlighted as the lead SNP.

qGF1; 5 SNPs	
SNP	**p-value**
S01_32481234	6.22E-04
S01_2345577	6.94E-04
S01_32498903	7.29E-04
S01_32450786	7.84E-04
S01_2374391	8.04E-04

qGF5; 4 SNPs	
SNP	**p-value**
S05_25541657	1.51E-04
S05_25577921	4.02E-04
S05_25928622	4.23E-04
S05_25571027	5.16E-04

qGF10; 10 SNPs	
SNP	**p-value**
S10_20860950	2.74E-06
S10_21124963	5.67E-06
S10_21136910	1.33E-05
S10_20110508	3.12E-05
S10_20110507	3.29E-05
S10_20110400	4.39E-05
S10_20206255	1.04E-04
S10_20112533	1.07E-04
S10_20331613	5.44E-04
S10_19994100	9.39E-04

qGF6; 29 SNPs	
SNP	**p-value**
S06_4330339	2.49E-05
S06_4553119	3.72E-05
S06_4241819	4.35E-05
S06_4553121	4.69E-05
S06_3468234	1.75E-04
S06_3463659	1.83E-04
S06_4540950	1.88E-04
S06_3463658	2.02E-04
S06_5544208	2.32E-04
S06_6143332	2.99E-04
S06_3457608	3.16E-04
S06_3510352	3.30E-04
S06_4241737	3.94E-04
S06_4237422	3.98E-04
S06_4244676	4.21E-04
S06_3477190	4.26E-04
S06_4244700	4.46E-04
S06_5384338	4.69E-04
S06_4241821	4.86E-04
S06_3475817	5.08E-04
S06_4105703	5.20E-04
S06_3505262	5.35E-04
S06_3457769	6.15E-04
S06_5531995	6.41E-04
S06_6165836	6.55E-04
S06_4417173	6.70E-04
S06_3458230	7.39E-04

qGF11; 27 SNPs	
SNP	**p-value**
S11_4906190	4.74E-05
S11_6403330	1.19E-04
S11_3529875	1.58E-04
S11_4844705	1.78E-04
S11_4931007	1.86E-04
S11_4892036	2.10E-04
S11_4871907	2.23E-04
S11_3505919	2.42E-04
S11_4882092	2.49E-04
S11_5330690	2.50E-04
S11_4769284	2.50E-04
S11_3533995	2.60E-04
S11_5335992	2.67E-04
S11_4852117	2.70E-04
S11_3532010	2.71E-04
S11_4839421	3.24E-04
S11_4632696	4.08E-04
S11_5360850	4.23E-04
S11_4842782	4.28E-04
S11_7620880	4.42E-04
S11_7568950	4.62E-04
S11_7477126	4.64E-04
S11_5358463	4.87E-04
S11_3412714	5.72E-04
S11_7583709	6.55E-04
S11_3380807	6.70E-04
S11_3505960	7.00E-04

S06_5538698	9.31E-04
S06_1158001	9.63E-04

Supplementary Table 11. List of SNPs included in each QTL associated with **GF.** Each QTL is labelled with its name, and the total number of associated SNPs. SNPs are denoted by chromosome and physical position (chr_pos format), along with the p-value obtained from the GWAS. The SNP with the lowest p-value within each QTL region is highlighted as the lead SNP.

qPL9; 86 SNPs							
SNP	**p-value**	**SNP**	**p-value**	**SNP**	**p-value**	**SNP**	**p-value**
S09_14198097	1.29E-05	S09_16110426	1.93E-05	S09_16420378	5.17E-05	S09_16562695	2.62E-05
S09_14401534	9.76E-06	S09_16231460	4.33E-05	S09_16428508	4.55E-05	S09_16562754	4.40E-05
S09_14423157	1.99E-05	S09_16249236	3.41E-05	S09_16467669	5.26E-05	S09_16577816	8.66E-05
S09_14423205	8.64E-06	S09_16250792	2.34E-05	S09_16475086	5.57E-05	S09_16577876	6.33E-05
S09_14423212	8.90E-06	S09_16265288	3.85E-05	S09_16476033	5.71E-05	S09_16578151	4.74E-05
S09_15152273	5.36E-05	S09_16268904	6.93E-05	S09_16476038	5.38E-05	S09_16652601	5.41E-05
S09_15202160	2.87E-05	S09_16273253	3.46E-05	S09_16476072	9.25E-05	S09_16670223	5.55E-05
S09_15219071	3.47E-05	S09_16277472	2.20E-05	S09_16487060	6.35E-05	S09_16672086	6.68E-05
S09_15225485	9.83E-06	S09_16277475	3.65E-05	S09_16487583	6.32E-05	S09_16672089	7.69E-05
S09_15234847	6.89E-05	S09_16279459	1.78E-05	S09_16491999	9.19E-05	S09_16715937	8.14E-05
S09_15239028	3.99E-05	S09_16315365	5.88E-05	S09_16494233	5.32E-05	S09_16719725	5.38E-05
S09_15239702	5.15E-05	S09_16315388	2.15E-05	S09_16504264	4.65E-05	S09_16807418	6.87E-05
S09_15243137	3.46E-05	S09_16338784	6.81E-05	S09_16515318	2.67E-11	S09_17104658	6.25E-11
S09_15251527	4.49E-06	S09_16346483	4.02E-05	S09_16515328	2.94E-05	S09_17114430	7.68E-06
S09_15261427	3.66E-05	S09_16353373	3.81E-05	S09_16515393	3.75E-05	S09_17115951	9.18E-12
S09_15266111	3.46E-05	S09_16363939	4.89E-05	S09_16527280	4.09E-05	S09_17115954	7.15E-12
S09_15655372	5.29E-09	S09_16367225	3.30E-05	S09_16527300	1.26E-10	S09_17116031	3.03E-11
S09_16005453	8.04E-05	S09_16385947	2.27E-05	S09_16527341	2.25E-05	S09_17116071	2.51E-10
S09_16012359	1.67E-05	S09_16390597	2.81E-05	S09_16527347	1.48E-05	S09_17144751	3.14E-06
S09_16012363	1.67E-05	S09_16392582	6.86E-05	S09_16527348	2.27E-05	S09_17281984	6.45E-05
S09_16084327	1.91E-05	S09_16405678	6.11E-05	S09_16527374	3.32E-05		
S09_16084351	7.92E-06	S09_16405717	5.18E-05	S09_16539340	4.44E-05		

Supplementary Table 12. List of SNPs included in each QTL associated with **PL.** Each QTL is labelled with its name, and the total number of associated SNPs. SNPs are denoted by chromosome and physical position (chr_pos format), along with the p-value obtained from the GWAS. The SNP with the lowest p-value within each QTL region is highlighted as the lead SNP.

Chapter IV

Genome-Wide Association Mapping in a MAGIC Rice Population Reveals QTLs and Candidate Genes Controlling Nitrogen Use Efficiency and Grain Yield under Low-Nitrogen Inputs

Julia García-Romeral and Concha Domingo

Departamento del Arroz, Centro de Genómica, Instituto Valenciano de Investigaciones Agrarias (IVIA)

Manuscript prepared for submission

Abstract

Plants require essential nutrients for optimal growth and development, which in agricultural systems, are usually supplied through the application of fertilisers. However, the excessive use of fertilisers, particularly nitrogen (N), has detrimental effects on the environment. To promote sustainable production practises, it is essential to reduce fertiliser inputs without compromising crop productivity. In this context, the development of high-yielding varieties that perform well under low nitrogen input conditions represents a promising and environmentally responsible strategy.

In this study, we conducted a genome-wide association studies (GWAS) using a Multi-parent Advanced Generation Inter-Cross (MAGIC) population to identify quantitative trait loci (QTLs), and SNP markers associated with nitrogen use efficiency (NUE) and yield in rice. The MAGIC population comprised 192 lines derived from intercrossing eight founder parents. These lines were evaluated for NUE, grain yield, and panicle number across three locations over two growing seasons. We detected 21 QTLs with the most promising loci located on chromosomes 5, 6 and 8. The associated SNP markers were found to co-locate with known major genes and QTLs such as *ERF142/NGR5, OsAAP12A* or *OsGATA8c*. The nature of the candidate genes involved illustrates the complexity of this trait, which encompasses N metabolism, assimilation, transport and remobilization as well as tillering and panicle architecture.

These findings offer new insights into the genetic control of NUE in rice and provide the genetic tools for future breeding programs that aim to reduce nitrogen inputs while maintaining high yields.

Introduction

Nitrogen (N) is one of the most critical macronutrients for soil fertility and crop productivity (Dobermann et al. 2002; Fageria 2007; Ladha et al., 2005). In high-performing cereal systems, including rice, N fertilization can account for as much as 40–50% of the total grain yield when comparing unfertilized conditions to well-managed N inputs (Fageria, 2007; Ladha et al., 2005). Global rice production has increased rapidly over the past decades due to population growth, leading to a significant rise in nitrogen fertilizer use (Godfray et al., 2010; Good et al., 2004; Muthayya et al., 2014). In addition, through years of breeding, rice varieties have been increasing their productivity at the cost of increasing the use of fertilizers. But a N excess can be harmful for the environment. It is a priority to reduce the use of fertilizers through coupling the best management practices with new varieties with superior grain yield and quality under low input of chemical fertilizers.

In this sense, one strategy to develop rice varieties capable of maintaining or improving yields while reducing fertilizer inputs consists in enhancing Nitrogen Use Efficiency (NUE), defined broadly as the ratio of biomass or grain yield to the amount of nitrogen supplied (Xu et al., 2012). NUE is a complicated trait as it is influenced by several environmental factors as water availability, light intensity, disease pressure. In addition, NUE depends on several genetics factors that modulate nitrogen uptake and assimilation as well as its remobilization and integration into development, including an increase in panicle number and grain yield (Liu et al. 2022).

Rice plants acquire nitrogen from the soil. In flooded paddy conditions of rice cultivation, ammonium (NH_4^+) and urea are the main nitrogen fertilizer applied in agricultural systems (Li et al. 2017). While nitrate (NO_3^-) becomes more frequent in aerobic soils (Girsang et al. 2020). After uptake, nitrogen is transported from roots to shoots via xylem and redistributed through the phloem. Nitrate is reduced by nitrate reductase (NR) and nitrite reductase (NiR), while ammonium is assimilated into amino acids via the glutamine synthetase (GS)–glutamate synthase (GOGAT) cycle. As the plant develops, nitrogen is remobilized from senescing tissues to support grain filling (Masclaux-Daubresse et al., 2010).

Many of these genes were first identified through genetic mapping approaches. Techniques such as QTL mapping and genome-wide association studies (GWAS) have been instrumental in uncovering novel loci associated with NUE-related traits. The combination of whole-genome sequencing (WGS) and GWAS now allows fine-scale mapping of trait-associated SNPs and candidate genes underlying quantitative traits. This approach is particularly powerful when applied to multi-parent advanced generation inter-cross (MAGIC) populations, as they capture greater allelic diversity and recombination than traditional biparental populations. This broader genetic variation improves the precision and robustness of QTL detection and facilitates the discovery of functionally relevant alleles (Bandillo et al., 2013).

NUE is different between the two main genetic groups in rice, with *indica* varieties generally having much higher NUE than *japonica* varieties (Zhang and Chu, 2020). The genetic diversity in rice cultivars already adapted to temperate climate is a useful resource to understand the genetic bases of adaptation of plants to specific environments and, consequently, to provide donors and molecular markers suitable for breeding under local agronomical conditions.

Building on these advances, our study uses a MAGIC population of rice to dissect the genetic architecture of NUE under contrasting nitrogen regimes. We evaluated 179 lines under two fertilization levels (N100 and N200), using panicle number and panicle weight as key agronomic indicators of nitrogen acquisition and utilization. Through a GWAS based on over 70,000 high-confidence SNPs, we identified several QTLs and candidate genes involved in NUE and yield under low N input. These findings provide novel insights into the genetic control of NUE in rice and lay the groundwork for future breeding efforts aimed at reducing nitrogen inputs while maintaining high yields.

Materials and Methods

Plant Materials and Field Trials

MAGIC population used in this study was developed following the intercrossing methodology described by (Bandillo et al., 2013), using eight elite inbred lines, , as explained in the Chapter III. This breeding strategy resulted in 500 recombinant inbred lines (RILs), from which 179 were randomly selected for phenotypic characterization. The development of the population, its advancement to a stable generation (F6–F8), its phenotypic evaluation for key agronomic traits, as well as its sequencing and population structure analysis, were all previously described in Chapter III.

The trials were conducted in Valencia, Spain (39°18'02.2"N, 0°19'16.4"W). An Augmented Block Design was implemented with two nitrogen fertilizer doses: N100 (100 kg/ha urea) and N200 (200 kg/ha urea). The experiment consisted into two blocks during the 2022 and 2023 growing seasons. Two local varieties, Argila and JSendra, were used as controls. The experimental design, which included 179 experimental lines, and 19 control replicates distributed across four blocks, was developed using SAS software (SAS Institute, Cary, NC, USA). This approach enabled the evaluation of experimental genotypes relative to the controls while accounting for environmental variability within and across blocks.

Seeds were initially sown in seedbeds until they reached the three to four-leaf stage, when they were transplanted into the experimental flooded fields. Nitrogen fertilizer was applied in urea form as it the most extended form in the region and was applied uniformly before transplanting. Each experimental plot for each experimental MAGIC line consisted of 30 plants arranged in 6 rows and 5 columns, occupying an area of 0.9 m × 0.75 m (width × length). Plots were arranged in rows separated by 0.5 m. The spacing between adjacent plots within a block was 0.3 m. The two nitrogen treatments (N100 and N200) were applied in separate sections of the field, with a 2 m buffer zone between them to prevent cross-contamination (**Figure 1**). Fields were kept flooded at a consistent depth of 10 cm. To ensure optimal crop development, calcium superphosphate (60 kg P2O5 ha-1) and potassium sulphate

(30 kg K2O ha-1) were incorporated into the soil before flooding, following the conventional fertilization practices commonly used in the region.

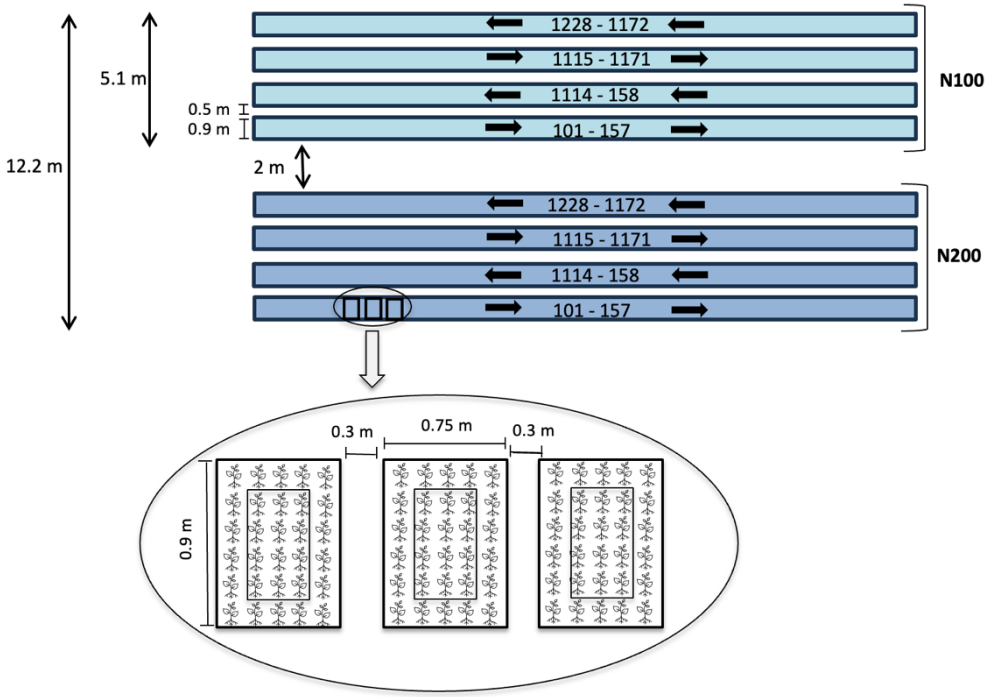


Figure 1. Field layout based on an augmented block design to evaluate the MAGIC rice population under two nitrogen levels (N100 and N200).

Phenotypic Evaluations and Nitrogen Use Efficiency Assays

Two yield-related traits, number of panicles per plant (PN) and weight of grain per plant (GW), were measured after harvest from the 12 inner plants in each plot and the mean was calculated. Nitrogen use efficiency (NUE) was then calculated as:

$$NUE = \frac{grain\ weight\ (12\ plants)}{Nitrogen\ dose}$$

The experimental design, as well as data analysis and normalization, were conducted using SAS software (SAS Institute, Cary, NC, USA)

Phenotype data evaluation under the Augmented Block Design was performed using the PROC MIXED procedure in SAS. Experimental and control data was used in the analysis to fit a mixed-effects model, which incorporated both, fixed and random effects to account for environmental and genetic variability. The statistical model used was:

$$Phenotype\ trait = \mu + entryc + block + (entry \times new) + e$$

Where μ is the overall mean; $entryc$ is the categorical variable included as a fixed effect to estimate the phenotypic differences between experimental and control entries; and $block$ and $(entry \times new)$ were modeled as random effects to account for variability due to environmental conditions and genotype classification variability. The residual error is denoted by e.

The least-squares means (LSMeans) of $entryc$ were calculated to provide adjusted phenotypic values. Additionally, empirical best linear unbiased predictors (EBLUPs) of the random effects were estimated to quantify the contribution of $block + (entry \times new)$ classification to the observed variability. These outputs were extracted using the "ods output" statement and for each plot, the adjusted phenotypic value was computed by summing: the intercept estimate from the "Solution for Fixed Effects" table and the random effect estimate for the specific plot. This adjustment ensured that the phenotypic values accounted for environmental variability across blocks and provided an accurate estimate of genetic performance under the Augmented Block Design.

Boxplots for each trait were generated to evaluate differences in trait values across years, with statistical significance determined by Student's t-test using ggplot2 and geom_signif.

Whole genome sequencing and data analysis

The analysis of the 179 MAGIC lines followed the methodology previously described in Chapters I and III. DNA was extracted, and Whole-Genome Sequencing

was performed at Novogene UK. Reads were quality-filtered and aligned to the Nipponbare reference genome (IRGSP-1.0). Variant calling was performed, followed by stringent filtering to retain finally a set of 70,289 high-confidence SNPs for subsequent analyses.

Genome Wide Association Study (GWAS)

Association analyses were performed in TASSEL v.5 (Bradbury et al., 2007) using the set of 70289 SNPs and the phenotypic data. To account for genetic relatedness, a kinship matrix (K) was generated based on the Centered IBS (Identity-By-State) method, which estimates the probability of shared alleles between individuals at a given locus. Additionally, a distance matrix (Q) was computed as 1—IBS similarity, and a Multidimensional Scaling (MDS) analysis, also referred to as Principal Coordinate Analysis (PCoA), was performed. The resulting MDS axes were used as covariates to adjust for population structure.

The GWAS incorporated genotype data, phenotype measurements, and the kinship matrix to conduct a phenotype-genotype association analysis. A Mixed Linear Model (MLM) was applied, using the K matrix to account for kinship and the Q matrix for population structure correction. To balance false positives and false negatives, a significance threshold of $P \leq 0.001$ ($-\log10(P) = 3$) was established.

Manhattan plots were generated using the R package CMplot (LiLin-Yin, 2024), while the phenotypic effects of the top-associated SNPs were visualized using the ggstatsplot package (Patil, 2021) in R.

Candidate Gene Mapping

SNPs exceeding the $P \leq 0.001$ ($-\log10(P) = 3$) threshold were further explored to identify candidate genes associated with NUE. Candidate regions were demarcated based on the linkage disequilibrium (LD) decay distance (879 kb), and in-house R scripts were used to annotate genes within these intervals using the Gramene and RapDB databases. Genes functionally linked to NUE were prioritized, and SNPs within the same LD-defined region were grouped as a single QTL, with the lead SNP being the one showing the lowest p-value.

Results

Phenotypic Characterization of the MAGIC Population

As described in Chapter III, a MAGIC rice population was developed by intercrossing eight elite breeding lines. Following the crossing methodology proposed by Bandillo et al. 2013, 500 recombinant inbred lines (RILs) were generated, from which 179 were randomly selected. These lines were assessed in three assays under field conditions over two years and using two nitrogen fertilization doses (N100 and N200). The analysis was focused on two agronomic traits related to NUE: Grain Weight (GW) and Panicle Number (PN) (**Table 1**).

		A1 (2022)			A2 (2023-1)			A3 (2023-2)		
Traits		**N100**	**N200**	**Ratio**	**N100**	**N200**	**Ratio**	**N100**	**N200**	**Ratio**
NUE	**Mean**	17,60	10,42	0,62	21,36	10,93	0,56	19,75	8,90	0,48
	Ds	4,63	1,88	0,16	7,16	1,03	0,18	7,03	3,14	0,15
	Max	34,02	17,97	1,31	63,29	14,66	1,49	51,10	26,43	1,00
	Min	8,00	5,68	0,36	6,25	7,82	0,17	5,16	1,54	0,12
Panicle Number	**Mean**	6,60	7,59	1,22	7,42	7,89	1,09	7,73	7,17	0,97
	Ds	1,99	2,06	0,45	1,30	0,61	0,15	2,36	1,85	0,24
	Max	12,25	15,13	4,66	12,74	9,74	1,64	18,98	11,78	1,90
	Min	1,08	1,24	0,31	4,04	6,07	0,62	1,73	1,58	0,16

Table 1. Variation in NUE and Panicle Number under low (N100) and high (N200) nitrogen levels. Values of the A1, A2 and A3 assays are indicated.

Grain yield per plant was recorded at both N doses and NUE was calculated, revealing significant variation. NUE under high nitrogen levels (NUE-200) decreased compared to NUE under low levels (NUE-100) in the three assays, indicating a general lower efficiency at high nitrogen availability (**Figure 2**). The NUE distributions revealed distinct differences between treatments.

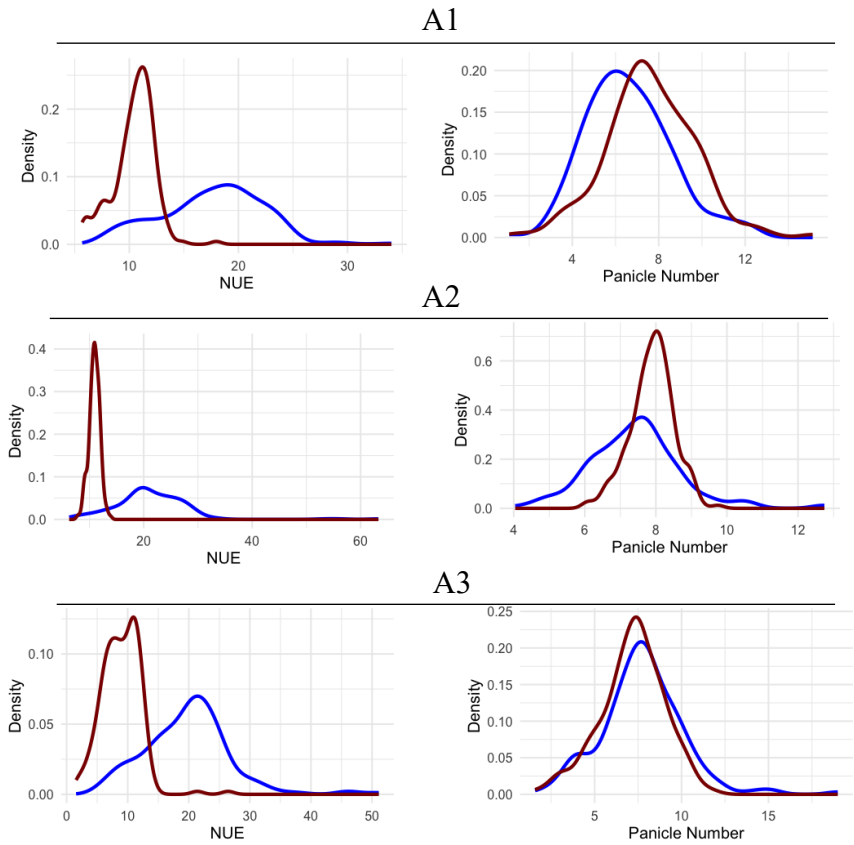


Figure 2. Density plots of nitrogen use efficiency (NUE) and panicle number across three field trials (A1, A2 and A3) under low (N100, blue) and high (N200, red) nitrogen conditions.

PN showed moderately higher values at N200 in the three assays, with a few lines showing comparable or slightly lower panicle numbers at N200. The PN distributions revealed distinct differences between treatments (**Figure 2**). To reflect the differences between the response of the plant to N200 and N100, the trait ratios (N200/N100) were used. NUE-ratios showed values predominantly below 1, indicating that most lines were more efficient in utilizing nitrogen at lower doses. However, the wide range of variation among lines (ratios ranging from 0.12 to 1.49) indicated strong differences among the varieties in the response to different N doses. Variability in PN-ratios (N200/N100) was particularly notable, ranging from 0.16 to 4.66, reflecting the diversity of responses within the population. The distribution of NUE-ratio and PN-ratio in the population was analyzed revealing differences among

the population (**Figure 3**). This observation allows for the identification of genotypes that could be prioritized for nitrogen-efficient breeding programs.

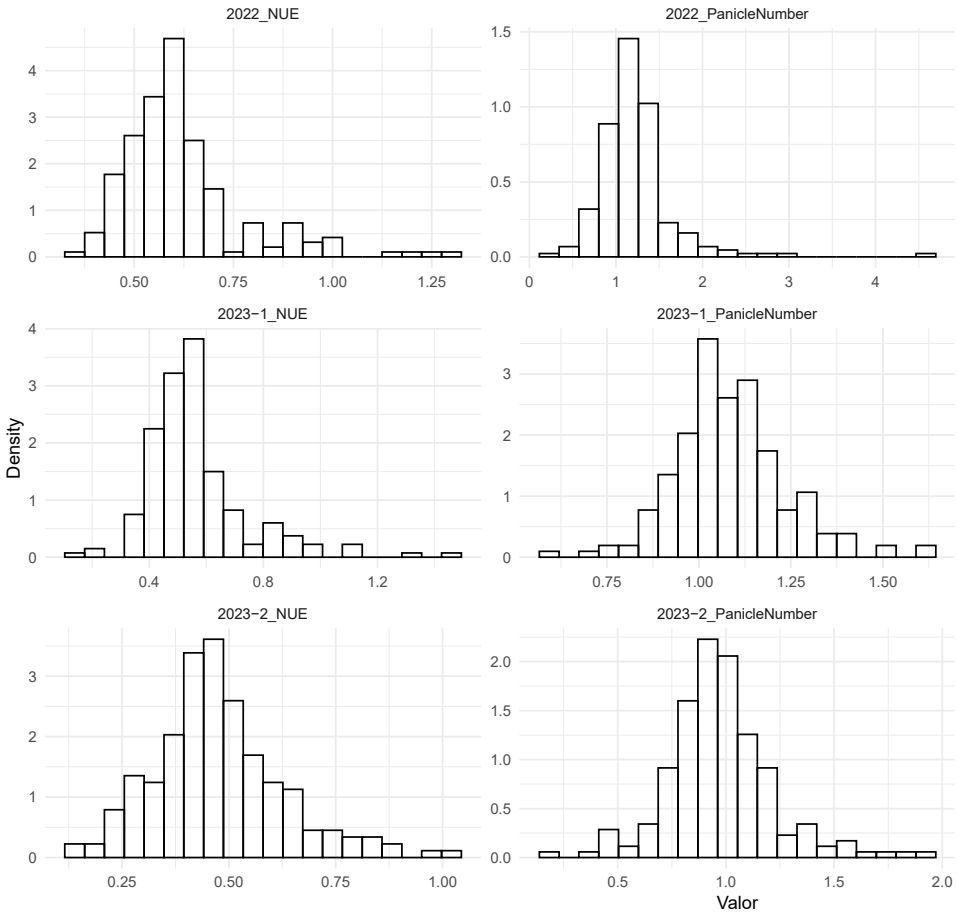


Figure 3. Frequency distributions of normalized trait values (N200/N100 ratios) for nitrogen use efficiency (NUE) and panicle number (PN) across the three trials (A1, A2, and A3).

Boxplot comparisons were conducted to statistically assess the differences between N100 and N200 for both NUE and PN across all years (**Figure 4**). Significant differences were observed, highlighting that nitrogen availability significantly affects plant development. The differences in NUE were particularly pronounced ($p < 0.001$), reflecting divergent nitrogen use strategies across genotypes. PN also

showed significant variation ($p < 0.001$ to $p < 0.05$), reinforcing that nitrogen influences in panicle production.

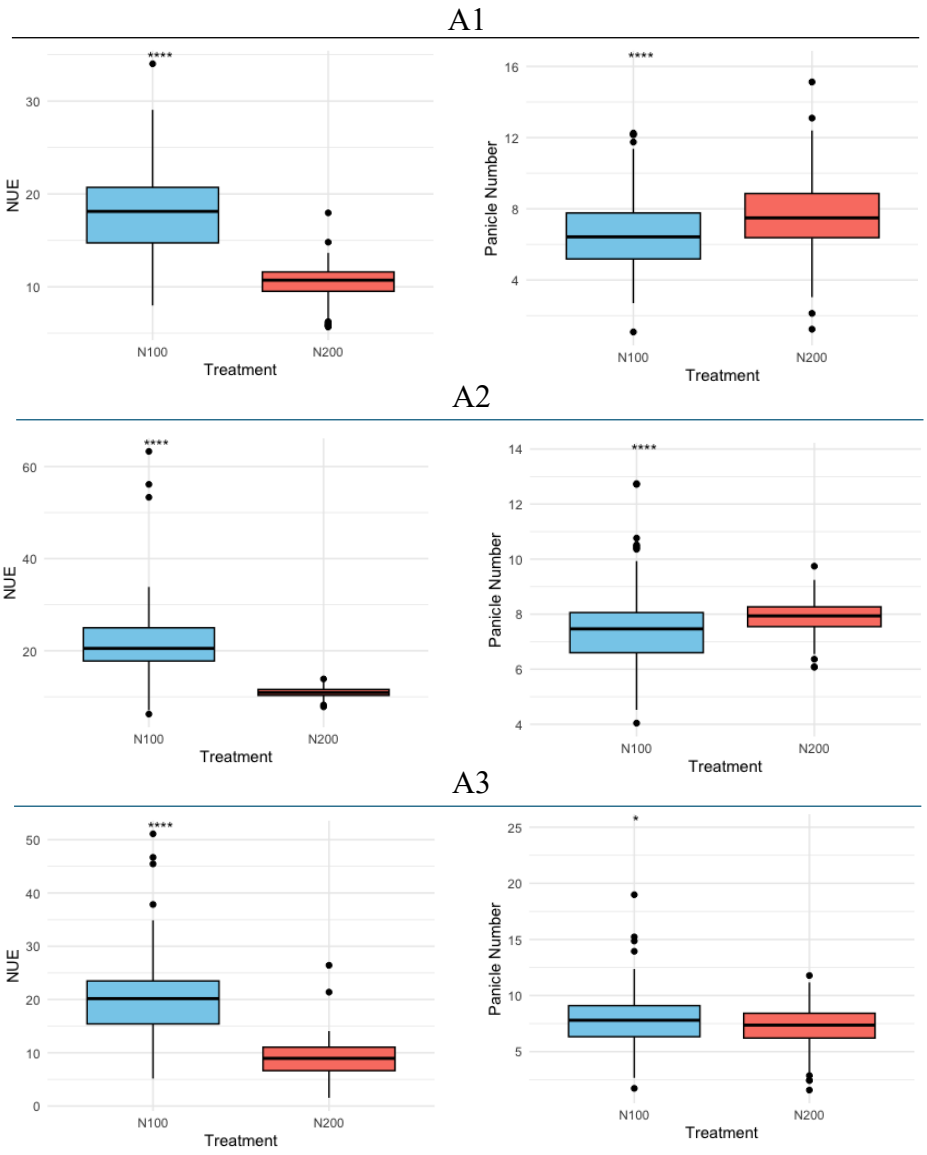


Figure 4. Boxplots displaying the distribution of nitrogen use efficiency (NUE) and panicle number (PN) under two nitrogen treatments (N100 and N200) across three trials (A1, A2, and A3). Asterisks above the boxes indicate the level of statistical significance from treatment comparisons (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

A year-to-year comparison of NUE and PN ratios was conducted to assess the consistency of the data across different growing seasons (**Figure 5**). Significant differences were found between years, illustrating that environmental factors contribute notably to phenotypic variation. Consequently, data from each year were analyzed separately, to ensure accurate interpretation and avoid potential biases.

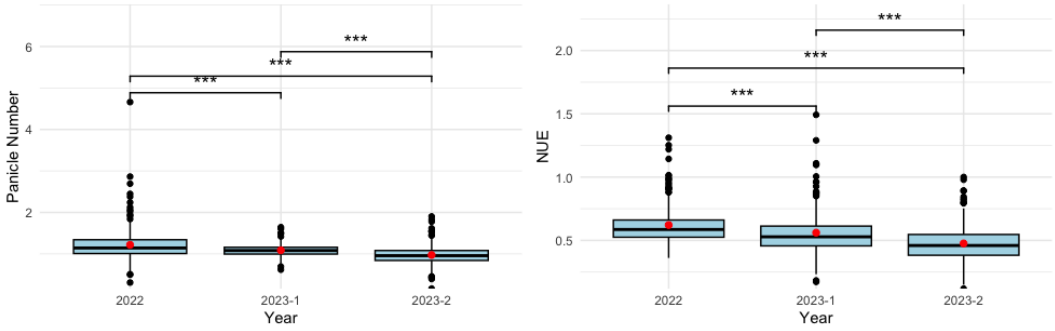


Figure 5. Boxplots showing the distribution of panicle number (PN) and nitrogen use efficiency (NUE) ratios (N200/N100) across three trials (A1, A2, and A3). Asterisks indicate statistically significant differences between years (***$p < 0.001$).

Correlations of PN and NUE across years and treatments were also explored (**Figure 6A**). Moderate to strong positive correlations ($r = 0.58–0.63$) indicate that N doses influence both NUE and panicle production. Given the influence of phenology, evaluating the N200/N100 ratios is particularly useful for minimizing phenological differences. Furthermore, the correlation of these ratios (**Figure 6B**) remained moderately to strongly positive ($r = 0.59–0.65$), reinforcing a consistent relationship between NUE and PN across environments.

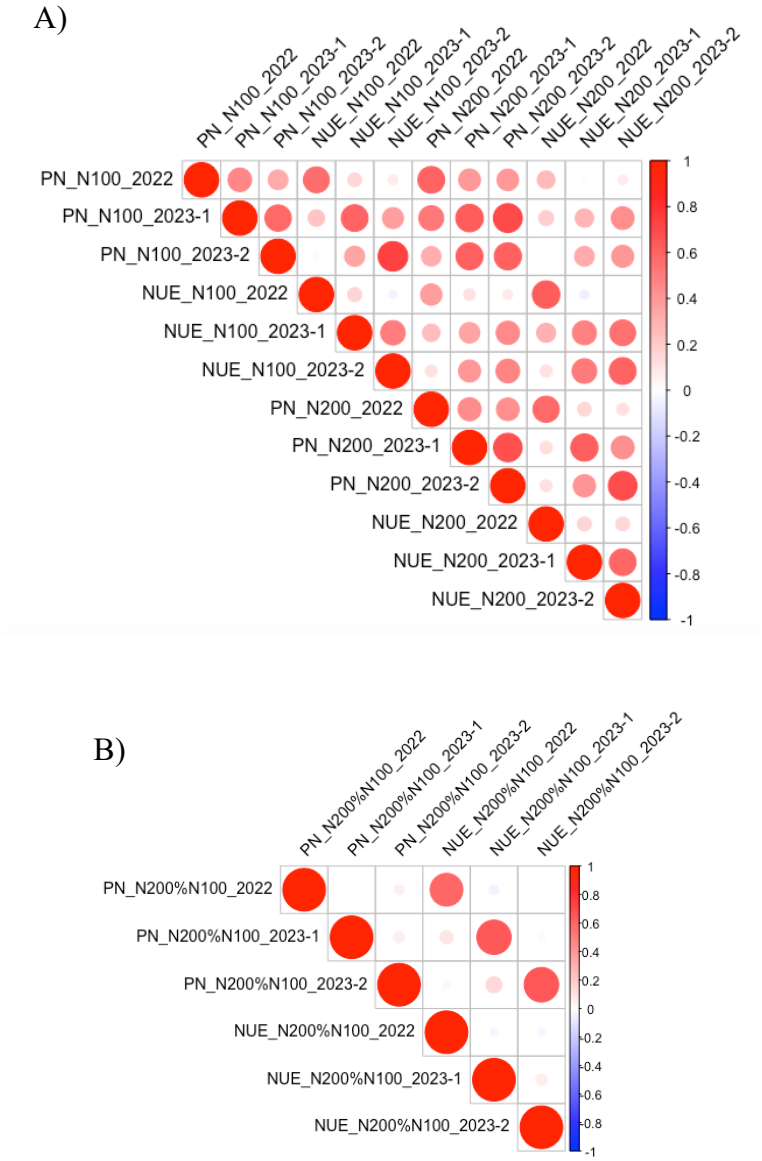


Figure 6. Pearson correlation matrices between panicle number (PN) and nitrogen use efficiency (NUE) measured across years and nitrogen treatments. **A)** Correlations between trait values under N100 and N200 treatments across all assays (A1, A2 and A3). **B)** Correlations based on N200/N100 ratios for PN and NUE across assays. Size and color intensity of the circles indicate the strength of the correlation (r), ranging from –1 (blue, strong negative) to +1 (red, strong positive)

Genome-wide association analysis

To dissect the genetic basis of yield when N availability is limited, we focused on NUE-100 and PN-100 as well as the NUE and PN ratios. While evaluating the traits at low N doses allows us to identify genetic components specifically associated with low N adaptation, comparing NUE and PN ratios between low and high N levels allowed us to assess the specific N responsiveness of varieties across environments. To detect QTLs associated with these traits, we conducted GWAS for NUE and PN-ratio and for NUE-100 and PN-100. The study was conducted in three different environmental scenarios (2022, 2023-1, and 2023-2) to account for annual variation.

A Mixed Linear Model (MLM) implemented in Tassel 5.0 was used to detect associations between SNP markers and variations across all years and traits. The significance threshold was set at $P \leq 0.001$ ($-\log10(P) = 3$). Manhattan and Q-Q plots (**Figure 7**) revealed clear significant peaks across multiple conditions and traits. Considering that the methodology may detect the false positive associations, we considered those QTLs containing more than significant SNP or that has been identified in more than one assay or trait. The Q-Q plots indicated that data distribution followed the expected model, with the observed p-values mostly uniformly distributed, but showing a slight deviation at high values in GW for 2022.

QTL associated with NUE and PN ratio

The statistical analysis revealed 268 SNPs associated with NUE-ratio, that according to the estimated linkage disequilibrium (LD), corresponded to 6 QTLs, located in chromosome 5, 6, 9 and 11 (**Table 2**). The Manhattan plot revealed three sharp peaks corresponding to qN5.2, qN6.4 and qN9.1. qN5.2 was detected in assays A1 and A3, with two leader peaks located at positions 19,085,653 and 19,587,320 with an explained variance of 16.14% and 7.18 % respectively. qN6.4 displayed high significance, and explained 14.55% of the variance, but was defined within a narrower region with 7 SNPs, in contrast with the 71 comprising qN5.2 (**Supplementary Table 1**). qN6.6, presented an interval of 8.2 kb. On chromosome 9, two separate peaks were observed: qN9.1a, with high significance and 10.2 % proportion of variance explained (PVE), and qN9.1b, less significant and with a broader LD span. Lastly, qN11.2, was detected in A1, accounting for 8.24% of the PVE, in NUE-ratio.

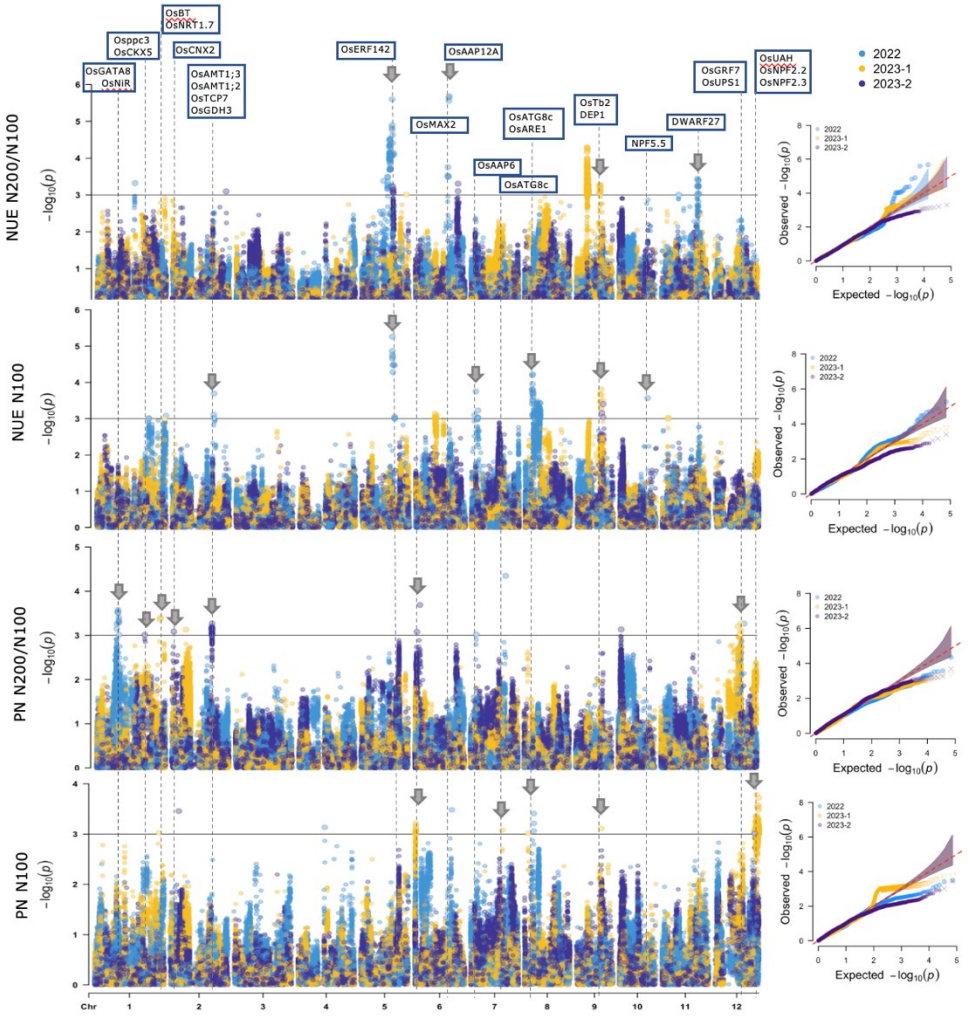


Figure 7. Manhattan plots (left) and quantile–quantile (Q-Q) plots (right) of NUE and PN under N100, and as trait ratios (N200/N100).Colours indicate the assay acroos the three years (2022, 2023-1, 2023-2). The horizontal line indicates the significance threshold ($-\log_{10}(p) = 3$). Arrows mark selected QTL regions supported by multiple significant SNPs, years, or traits. Candidate genes located near lead SNPs are highlighted.

The analysis of association with PN-ratio revealed 36 significant SNP above the threshold, showing weak association values, that corresponded to 6 QTLs, distributed across chromosomes 1, 2, 6, and 12 (**Table 2; Supplementary Table 2**). qN6.2b showed the strongest association and explained the largest proportion of phenotypic variance (8.78%), despite being defined by a single SNP. qN1.1 also showed a relatively high PVE (8.21%) and was mapped to a defined region of ~1.25 Mb. In contrast, qN1.4, was characterized by a smaller effect and narrower interval. Both detected in A1. The two QTLs on chromosome 2, qN2.2 and qN2.3, were detected in different assays and showed comparable statistical support, with qN2.3 having a better-defined interval (~350 kb). Finally, qN12.1 was the only region detected in two independent assays (A1 and A2), with nearly identical p-values and PVE.

QTL	Chr	Assay	Lead SNP Position	P-Value	PVE	Allele	Effect Size	Interval
NUE ratio								
qN5.2	5	A1	19.085.653	2,5E-06	16,14	A	0,22	17,440,428 - 20,422,515
		A3	19.587.320	6,2E-04	7,18	A	0,09	
qN6.4	6	A1	21.311.240	2,1E-06	14,55	A	-0,18	20,251,259 - 21,311,240
qN6.6	6	A3	26.734.351	4,9E-04	7,68	A	0,16	26,734,351 - 26,742,604
qN9.1a	9	A2	7.134.210	5,2E-05	10,25	C	-0,26	7,016,960 - 7,955,654
qN9.1b	9	A2	15.239.028	5,4E-04	7,53	C	-0,21	14,673,737 - 15,266,111
qN11.2	11	A1	22.645.067	3,6E-04	8,24	A	0,11	22,632,966 - 23,435,912
PN ratio								
qN1.1	1	A1	13.412.069	2,7E-04	8,21	A	0,68	12,687,073 - 13,942,728
qN1.4	1	A2	40.034.230	3,9E-04	7,96	A	-0,14	39,576,121 - 40,034,230
qN2.2	2	A2	8.970.127	7,4E-04	7,31	A	0,21	-
qN2.3	2	A3	25.315.944	5,4E-04	7,84	C	-0,23	24,971,352 - 25,325,091
qN6.2b	6	A3	3.312.938	2,1E-04	8,78	C	-0,27	-
qN12.1	12	A2	17.438.546	5,4E-04	7,37	C	0,10	17,066,507 - 19,538,903
		A1	19.538.903	5,4E-04	7,39	A	-0,39	

Table 2. List of lead-SNPs significantly associated with NUE and PN ratio. The proportion of variance explained (PVE) is indicated. The Effect Size is from the denoted allele.

QTL associated with NUE and PN at N100

We also examined the associations to NUE-100 to find QTLs associated with grain yield under low nitrogen conditions of fertilization and we identified 9 associated QTLs comprised of 95 significant SNPs (**Table 3; Supplementary Table 3**). Among all detected NUE-100 QTLs, qN5.2 and qN9.1b were also detected by GWAS of the NUE-ratio and corresponded to the most prominent peaks in the Manhattan plot. In addition to these, qN2.3, qN7.1, and qN8.1 showed the highest levels of significance. Also, qN9.1b was consistently detected in both A2 and A3 assays. With respect to the proportion of phenotypic variance explained, the highest values corresponded to qN5.2 and qN8.1 displaying 13.22 and 10.55 % respectively. Most QTL intervals were relatively, except for qN8.1, which spanned over 3.2Mb. In contrast, qN10.1 was defined by a single SNP.

QTLs for tiller/panicle number to increase grain yield

The association analysis also revealed 369 SNPs associated with PN-100, which could be grouped into 6 QTLs (**Table 3; Supplementary Table 4**). Overall, the associations displayed moderate levels of significance. As a result, qN6.4, qN7.2, qN8.1, and qN9.1b, were defined by a single SNP. Among the PN-100 QTLs, qN12.2 stands out showing the most prominent peak in the Manhattan plot and being consistently detected in two independent assays (A2 and A3). No common associations were found between PN-ratio and PN-100, although two QTL, qN8.1 and qN9.1b, coincided with those identified in the NUE-ratio and NUE-100 analyses.

QTL	Chr	Assay	Lead SNP Position	P-Value	PVE	Allele	Effect Size	Interval
NUE N100								
qN1.2b	1	A1	32.492.169	9,6E-04	6,82	G	5,04	32,491,201 - 32,492,169
qN2.3	2	A1	26.232.385	2,0E-04	8,62	C	-4,09	26,045,353 - 26,232,385
qN5.2	5	A1	19.085.653	2,5E-06	13,22	A	-5,31	19,053,547 - 20,576,620
qN6.3	6	A2	12.102.783	7,3E-04	7,18	A	-9,63	12,102,783 - 13,009,769
qN7.1	7	A1	2.672.533	1.70E-04	8,97	C	3,24	1,836,005 - 2,672,533
qN8.1	8	A1	4.584.816	6,0E-05	10,55	A	3,38	4,351,455 - 7,591,279
qN9.1b	9	A2	15.243.137	1,6E-04	8,95	C	-4,57	14,678,888 - 16,527,300
		A3	16.515.318	4,0E-04	7,75	A	3,79	
qN10.1	10	A1	17.673.046	2,7E-04	8,60	C	-2,36	-
qN11.1	11	A2	3.874.320	9,5E-04	6,91	C	3,70	3584189 - 3874320
PN N100								
qN6.1	6	A2	737.151	6,1E-04	7,37	C	0,48	688,641 - 786,784
qN6.4	6	A1	23.190.901	3,3E-04	7,83	A	-1,22	-
qN7.2	7	A2	19.613.498	8,4E-04	2,30	C	-0,99	-
qN8.1	8	A1	5.794.542	3,9E-04	8,56	C	1,55	-
qN9.1b	9	A2	15.655.372	7,7E-04	7,40	A	0,34	-
qN12.2	12	A3	24.896.521	9,7E-04	6,72	A	2,43	24,896,521 - 27,509,569
		A2	25.771.769	1,9E-04	8,747	C	1,86	

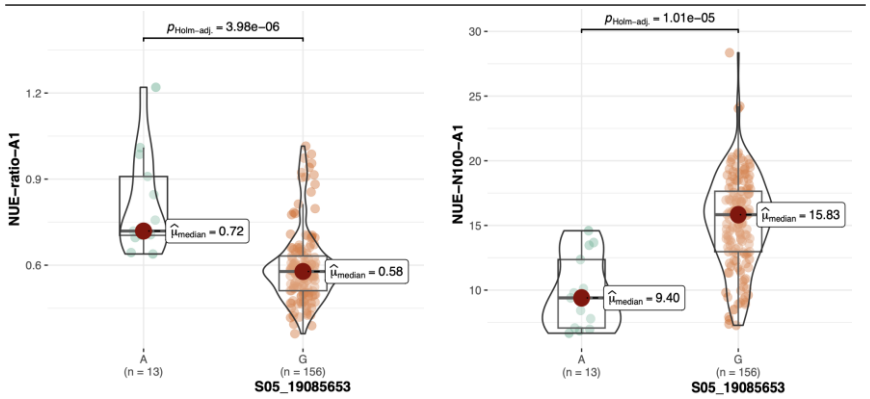
Table 3. List of lead-SNPs significantly associated with NUE and PN at N100 dose. The proportion of variance explained (PVE) is indicated. The Effect Size is from the denoted allele.

Among all detected QTLs, qN9.1b was detected in NUE-ratio, NUE-100 and PN-100. Some QTLS were detected in more than one trait as qN2.3 was detected in PN-ratio and NUE-100, qN5.2 was detected in both NUE-ratio and NUE-100, qN8.1 in common between NUE-100 and PN-100, qN6.4 was associated to NUE-ratio and PN-100.

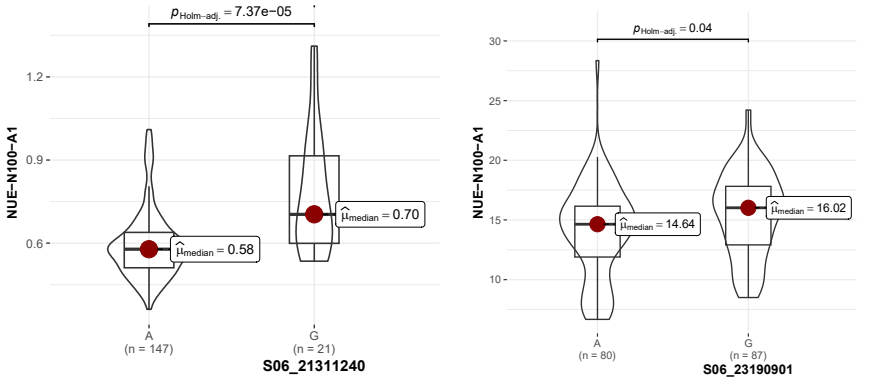
The allele content in the lead SNP displayed promising distribution among the MAGIC lines, as significant differences in the studied traits were found in several QTLs (**Figure 7; Supplementary Figure 1**). The most promising QTLs, with the lowest p-values and the sharpest Manhattan plot peaks were, qN5.2 (2.53E-06 for NUE-ratio and NUE-100), qN6.4a (2.10E-06, also for NUE-ratio), qN8.1 (6.00E-05,

for NUE-100), and qN9.1a (5,2E-05) at NUE-ratio. Accessions carrying different alleles of SNP leader in qN5.2, located at 19085653, showed a different NUE, in ratio and N100 analysis. Those carrying G, the mayor allele, displayed 15.83 g in contrast to those carrying A which displayed 9.40 g. Similarly, accessions with smaller NUE ratios tended to carry the G allele, suggesting higher NUE than those with the A allele.

qN5.2



qN6.4



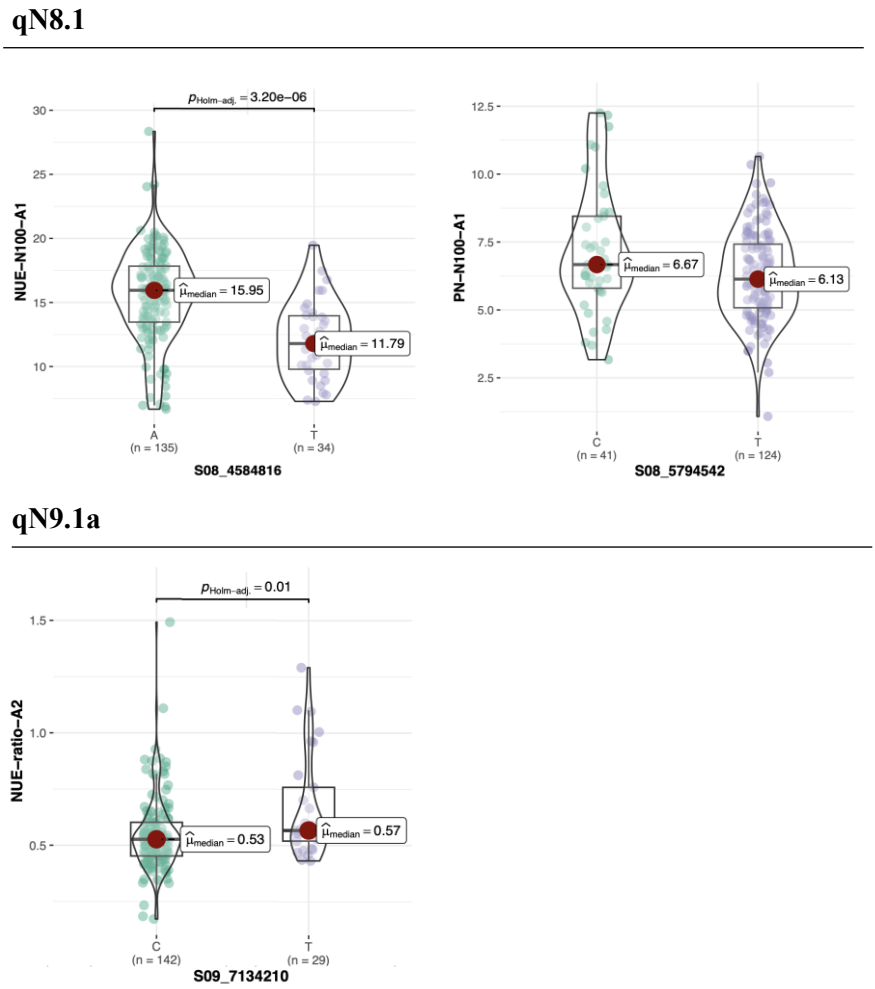


Figure 8. Violin plots showing the distribution of trait values (NUE and PN) according to allelic variation at lead SNPs for selected QTLs.

Accessions carrying A allele at the peak leader in q6.4a showed a mean value of 0.58 NUE-ratio, while those carrying allele G displayed a ratio of 0.7. For NUE-N100 within the same QTL, the G allele was associated with the highest panicle weight.

In qN8.1, the lead SNP for NUE-N100 (position 4,584,816) showed that lines carrying the A allele had higher values than those with the T allele. However, for the

lead SNP associated with PN-N100 within the same QTL (position 5,794,542), no significant differences were observed between alleles. In qN9.1a, accessions carrying the major C allele were identified as more nitrogen-efficient, exhibiting lower ratio than those carrying the alternative allele.

In addition to the lead SNPs of the major QTLs, the allele content of the minor peaks revealed promising SNPs (**Supplmentary Figure 1**). At qN1.4, the lead SNP for PN-ratio indicated that varieties carrying the A major allele had higher NUE and produced more panicles below N100 (7.58) compared with those carrying the G allele (6.44) (**Suppl. Fig. 1A**). The lead SNP at qN2.3 associated with PN-ratio showed lower ratios for the C allele relative to T, suggesting an improved nitrogen efficiency and higher productivity at N100 for those genotypes. At the same QTL, the minor T allele for NUE-100 peak increased panicle weight to 17 g (**Suppl. Fig. 1B, C**). At qN7.2, another minor allele (T) was associated with more panicles per plant at N100 (**Suppl. Fig. 1D**). In qN9.1, the SNPs with the lowest p-values were consistently associated with favourable alleles across all trials, with lines reaching 18 g panicle weight under N100 (**Suppl. Fig. 1E**). Finally, in qN12.2, both lead SNPs for PN-100 showed that minor alleles increased panicle number, with values of 9.17 and 8.25 (**Suppl. Fig. 1F**).

Mining candidate genes for NUE and yield at low N doses

In an attempt to find candidate genes involved in NUE and yield under low N doses, we looked for annotated genes that could be within the QTLs regions in an extended interval of 0.9 Mb, based on the estimated LD, using the annotated reference genome IRGSP-1.0 (RAP-DB). (**Table 4**). A total of 29 genes were detected from the significant SNPs, many of them associated to more than one of the evaluated traits. No candidate genes for N were found in the QTLs: qN6.1, qN9.1a and qN11.1

The candidate genes could be grouped according to functional categories of nitrogen metabolism, including uptake, transport, assimilation, and remobilization. Beyond genes directly involved in nitrogen metabolism, we also identified a set of genes that indirectly influence NUE through transcriptional and post-translational regulation, hormonal signaling, and root and shoot architecture. These findings further emphasize the complex genetic control of this trait in rice.

Locus	Gene Symbol	Description	Trait	QTL	Chr	Reference
Nitrogen uptake						
Os01g0908200	OsBT	Member of the Bric-a-Brac/Tramtrack/Broad (BTB) family	PN-ratio	qN1.4	1	Araus et al. 2016
Os02g0620500	OsAMT1;3	Uptake transporter 1;3 under low ammonium condition	NUE-N100 and PN-ratio	qN2.3	2	Konishi et al. 2021
Os02g0620600	OsAMT1;2	Uptake transporter 1;2 under low ammonium condition				Sonoda et al. 2003; Hao et al. 2016
Os12g0503000	OsUPS1	Ureide permease 1. Allantoin transporter	PN-ratio	q12.1b	12	Lee et al. 2018
Nitrogen transport						
Os01g0913300	OsNRT1.7	TGF-beta receptor, type I/II extracellular region family protein	PN-ratio	qN1.2	1	Fan et al. 2009; Fan et al .2017
Os06g0556000	OsAAP12A	Amino acid permease 12A. Transport of amino acids	NUE-ratio and PN-N100	qN6.4	6	Fang et al. 2021; Taylor et al. 2015; Zhao et al. 2012
Os07g0134000	OsAAP6	Amino acid permease 6. Transport of amino acids	NUE-N100	qN7.1	7	Tiong et al. 2021; Taylor et al. 2015; Zhao et al. 2012
Os10g0470700	OsNPF5.5	Nitrate transporter 1/Peptide transporter 5.5	NUE-N100	qN10.1	10	Zhu et al. 2023
Os12g0638200	OsNPF2.2	Nitrate transporter 1/peptide transporter family 2.2	PN-ratio and PN-N100	qN12.2	12	Li et al. 2015
Os12g0638300	OsNPF2.3	Nitrate transporter 1/peptide transporter family 2.3			12	
Nitrogen assimilation						
Os01g0357100	OsNiR	Ferredoxin-nitrite reductase	PN-ratio	qN1.1	1	Nishimura et al. 2005; Terada et al. 1995
Os01g0357500		Similar to Ferredoxin-nitrite reductase	PN-ratio			
Os01g0758300	Osppc3	Phosphoenolpyruvate carboxylase (PEPC)	NUE-N100	qN1.2b	1	Lian et al. 2021; Muramatsu et al. 2015; Masumoto et al. 2010
Os02g0258900		Similar to Molybdopterin biosynthesis CNX2 protein	PN-ratio	qN2.2	2	
Os02g0650900	OsGDH3	Similar to Glutamate dehydrogenase 2	NUE-N100	qN2.3	2	
Os12g0507000		Similar to Molybdopterin biosynthesis CNX2 protein	PN-ratio	qN12.1b	12	
Nitrogen remobilization						

Os07g0512200	OsATG8a	Core component of an autophagy gene	PN-N100	qN7.2	7	Yu et al. 2019; Zhen et al. 2019; Fan et al. 2020
Os08g0191600	OsATG8c	Core autophagy gene	NUE-ratio and PN-N100	qN8.1	8	Yu et al. 2019; Zhen et al. 2019; Fan et al. 2020
Os12g0597500	OsUAH	Peptidase M20 domain containing protein	PN-N100	qN12.2	12	
Indirect Nitrogen Related Factors						
Nitrogen Transcriptional or Postraductional Regulation						
Os08g0224300	OsARE1	Abnormal cytokinin response (abc) 1-1 repressor1	NUE-N100	qN8.2	8	Wang et al. 2018
Os12g0484900	OsGRF7	Growth-regulating factor 7	PN-ratio	qN12.1	12	Chen et al. 2020
Other Nitrogen Indirect Related Genes						
Os01g0775400	CKX5	Cytokinin oxidase/dehydrogenase 5	NUE-N100	qN1.2b	1	Zheng et al. 2023
Os09g0441900	OsDEP1	Phosphatidylethanolamine-binding protein (PEBP)	NUE-ratio; NUE-100 and PN-N100	qN9.1b	9	Huang et al. 2009; Zhou et al. 2009
Os11g0587000	DWARF27	Carotene isomerase, Strigolactones biosynthesis	NUE-ratio	qN11.2	11	Alder A et al. 2012; Lin et al. 2009
Tillering						
Os01g0343300	OsGATA 8	GATA transcription factor	PN-ratio	qN1.1	1	Wu et al. 2024
Os02g0635800	OsTCP7	Teosinte branched/cycloidea/proliferating cell factor 7. Transcription factor, Cass I PCF type protein	PN-ratio and NUE-N100	qN2.3	2	Zhang et al. 2024
Os05g0389000	OsERF142	Nitrogen-mediated tiller growth response 5, ethylene response factor 142	NUE-ratio and NUE-N100	qN5.2	5	Li et al. 2015; Aya et al. 2014; Wu K et al. 2020
Os06g0154200	D3/OsMAX2	F-box component of the SKP-Cullin-F box (SCF) E3 ubiquitin ligase complex	PN-ratio	qN6.2b	6	Sang et al. 2014
Os09g0410500	OsTb2/OsREP1	TCP family transcription factor. Retarded palea 1	NUE-ratio; NUE-100 and PN-N100	qN9.1b	9	Yuan et al. 2009; Yin et al. 2018

Table 4. List of candidate genes located within QTL intervals associated to NUE and PN ratios and at N100 dose. Candidate loci were identified within an extended interval of ±0.45 Mb (0.9 Mb total) from the lead SNPs, based on estimated linkage disequilibrium.

The GWAS analysis identified 4 candidate genes associated with nitrogen uptake in rice from the soil via root absorption. Interestingly, these genes co-localized with QTLs significantly associated with PN-ratio. Two genes coding for high-affinity ammonium transporters (AMTs), *OsAMT1.2* (*Os02g0620600*) (D. Hao et al. 2016; Sonoda et al., 2003) and *OsAMT1.3* (*Os02g0620500*) (Konishi & Ma, 2021), could be found in qN2.3, which is also associated to NUE-100. High-affinity transporters play a crucial role in ammonium acquisition under low nitrogen conditions (N100), where efficient nitrogen utilization is essential. Similarly, *OsBT* (*Os01g0908200*), a member of the Bric-a-Brac Tamtrack Broad family, identified qN1.4, has been described as a negative regulator of nitrogen uptake, modulating the expression of nitrate transporter genes. Consistently with the application of urea as a source of N in the field trial, *Ureide permease 1* (*OsUPS1*, *Os12g0503000*), a gene encoding an allantonin transporter involved in ureide metabolism, was found in qN12.1. The role of *OsUPS1*, facilitating the root uptake and translocation of urea within the plant has been described previously, suggesting that urea uptake pathways also play a role in optimizing nitrogen acquisition (Redillas et al. 2019).

Several candidate genes have been implicated in nitrate transport from the root to the shoot. Most of these transporters belong to the low-affinity nitrate transporter (NRT1)/peptide transporter (PTR) family (NPF), although many can also facilitate the uptake of small peptides. *OsNPF2.2* (*Os12g0638200*) and *OsNPF2.3* (*Os12g0638300*), were located in qN12.2b, a QTL consistently detected over two years for PN-ratio and PN-100. *OsNPF2.2* is a low-affinity nitrate transporter involved in root-to-shoot nitrate transport and vascular development (Y. Li et al., 2015), while the function of *OsNPF2.3* function remains unclear beyond its similarity to peptide transporters. Given that low-affinity transporters are typically more active under high nitrogen conditions, their association with PN-100 suggests a possible role in nitrogen redistribution rather than direct uptake. Additionally, *OsNPF5.5* (*Os10g0470700*) was located in qN10.1 and was associated with NUE-N100. *OsNRT1.7* (*Os01g0913300*) appeared in both PN-ratio and PN-100.

Amino acid permease 6 and 12A (*OsAAP6*, *Os07g0134000* and *OsAAP12A*, *Os06g0556000*), two amino acid permeases involved in amino acid transport and nitrogen metabolism (Taylor et al. 2015; H. Zhao et al. 2012), were associated with distinct QTLs. While *OsAAP6* was detected in qN7.1, associated to NUE-100, *OsAAP12A* was found in qN6.4, a robust QTL, associated to NUE-ratio, and PN-100

Significant trait-associations were found with genes involved in nitrogen assimilation, mainly involved in nitrate reduction, glutamate metabolism, and molybdenum cofactor biosynthesis. *OsNiR* (*Os01g0357100*) and *Os01g0357500* which encode two ferredoxin-nitrite reductases were located in qN1.1, associated with PN-ratio. These enzymes catalyze the reduction of nitrite to ammonium and are involved in the nitrate assimilation (Nishimura et al. 2005; Terada et al. 1995). *Os02g0258900*, in qN2.1, and *Os12g0507000*, in qN12.1b, encode two proteins similar to *CNX2* (*Cofactor of Nitrate reductase and Xanthine dehydrogenase 2*), involved in the biosynthesis of molybdopterin, a molybdenum cofactor essential for nitrate reductase activity (Liu et al., 2019). Both QTLs are associated with PN-ratio. *Phosphoenolpyruvate carboxylase 3* (*Osppc3*, Os01g0758300), and *Glutamate Dehydrogenase 3* (*OsGDH3*, *Os02g0650900*), involved in glutamate metabolism, were located in qN1.2 and qN2.3, respectively, and emerged as a candidate genes associated with NUE-100.

Nitrogen remobilization, the process of reallocating stored nitrogen from senescing tissues to actively growing organs, plays a crucial role in plant adaptation to fluctuating nitrogen availability. *Autophagy-related Gene 8a* (*OsATG8a*, *Os07g0512200*), in qN7.2, and *Autophagy-related Gene 8c* (*OsATG8c*, *Os08g0191600*), in qN8.1, were found to be associated with PN-100. These two core autophagy genes encode ubiquitin-like proteins that anchor to autophagic membranes, facilitating the degradation and recycling of macromolecules in vacuoles. qN8.1 is also associated to NUE-100, showing a sharp peak in the Manhattan plot and *OsATG8c* is one of the most promising candidate genes.

In addition, *Ureidoglycolate Amidohydrolase* (*OsUAH*, *Os12g0597500*), a peptidase M20 domain containing protein that is involved in ureide metabolism, was found in qN12.2 associated with PN-100 in two assays.

Beyond the core pathways of nitrogen uptake, transport, assimilation, and remobilization, other genes exert indirect effects on NUE by modulating tillering, plant architecture, hormone signaling, and transcriptional/post-translational processes. These factors often converge to determine how effectively the plant acquires, utilizes and reallocates nitrogen under different environmental conditions.

Two transcription factors were identified. *Abnormal cytokinin response (abc) 1-1 repressor 1* (*OsARE1*, *Os08g0224300*), a chloroplast-localized protein, was found in qN8.1 in association with NUE-100. Loss-of-function or down-regulation of *OsARE1* has been shown to improve NUE and grain yield. *Growth-Regulating Factor 7* (*OsGRF7*, *Os12g0484900*), was found in qN12.1a, associated with PN-ratio. It plays a key role in plant architecture by modulating gibberellin and auxin metabolism, affecting leaf angle, culm thickness, and plant height (Chen et al. 2020). These traits can influence panicle weight by altering nutrient allocation, structural support, and hormonal regulation.

A further set of candidate genes, identified within the LD distance of significant SNPs, were associated with hormone biosynthesis, transport, stress response, and panicle and plant architecture, affecting tillering. Significant polymorphisms were identified in *DWARF27* (*Os11g0587000*), a carotene isomerase, under NUE ratio, which encodes key enzymes in the strigolactone biosynthesis pathway. A cytokinin oxidase/dehydrogenase gene, involved in cytokinin degradation, has been identified. *Cytokinin oxidase/dehydrogenase 5* (*OsCKX5*, *Os01g0775400*), in qN1.2 associated with NUE-100, regulates root development, plant height, and tillering through cytokinin degradation (Zheng et al., 2023). Finally, *Dense and erect panicle 1* (*OsDEP1*, *Os09g0441900*), encoding a phosphatidylethanolamine-binding protein (PEBP), was associated with NUE-100 over two years. This gene regulates panicle architecture, with loss-of-function mutations resulting in erect panicles, a trait selected during rice domestication for improved light interception, yield stability, and lodging resistance (Huang et al. 2009; Zhou et al. 2009). In addition to its role in panicle structure, recent studies suggest that *OsDEP1* may also influence NUE. Research on nitrogen-related pathways suggests that *OsDEP1* is involved in optimizing nitrogen uptake and assimilation, which may contribute to improved grain filling and panicle weight under varying nitrogen conditions (Zhao et al. 2019).

The search for tillering-related genes led us to the detection of two genes encoding TCP transcription factors were found, *Teosinte branched 2* (*OsTb2*, *Os09g0410500*), found in qN9.1b associated with NUE-ratio, NUE-100 and PN-100 and, in addition, *Teosinte branched/cycloidea/proliferating cell factor 7* (*OsTCP7*, *Os02g0635800*) in qN2.3 associated with NUE-100 and PN-ratio. TCP transcription factors have been implicated in the regulation of tillering in rice. Nitrogen-mediated tiller *Growth Response 5/Ethylene Response Factor 142* (*OsNGR5/OsERF142*, *Os05g0389000*),

which encodes an AP2/ERF transcription factor was found in qN5.2 and is strongly associated with NUE-ratio and NUE-100.

Os01g0343300, encoding *OsGATA8*, a GATA transcription factor was found in qN1.1, associated with PN-ratio. Recent studies have shown that OsGATA regulates nitrogen uptake by repressing the expression of the ammonium transporter genes *OsAMT1.2* and *OsAMT3.2*, thereby affecting NUE and tillering (Wu et al. 2024). *D3/OsMAX2* (*Os06g0154200*), which encodes a F-box leucine-trich repeat (LRR) protein, co-localized with qN6.2b in association with PN-ratio.

Discussion

Reducing the amount of fertilizer inputs without a penalty in yield has become a challenge in crops and has intensified efforts to find plants with an optimal response to N. In this study, in an attempt to find the genomic basis for high yield under low N doses, we have performed GWAS for NUE and PN at low N doses under field conditions. Evaluating these traits separately at moderate and high fertilizer rates allowed us to characterize the baseline performance of each genotype and, by comparing these values across doses, established a ratio-based assessment of how effectively each line converts available nitrogen into yield. To achieve this, we used a MAGIC population. Compared to whole variety collections, MAGIC populations offer several advantages providing a high level of recombination from a limited number of genetic variations as well as a broader genetic base and a higher recombination rate than biparental crosses. To perform GWAS, we used a robust panel of high-confidence SNPs generated by whole-genome sequencing. This comprehensive marker set enabled us to identify genomic regions and candidate genes that could potentially control the variation observed in PN and GW under low nitrogen regimes, thus facilitating the understanding of the genetic basis of NUE.

From three experimental trials, we identified 585 leading SNP peaks, grouped into 22 QTLs associated to NUE or PN under different N doses. Some of the loci identified in this study in association with NUE and PN appeared particularly promising based on differences in GW and PN between cultivars carrying different alleles. In particular, qN5.2, qN6.4, qN8.1, and qN9.1a showed the lowest p-values across all SNPs detected in the three trials. Among them, the lead SNP at qN5.2 exhibited the most pronounced allelic difference for NUE-100 and NUE-ratio. However, the grain weight of the higher-yielding allele remained moderate (~15 g), a trend also observed for the lead SNP at qN8.1 (~14.86 g). By contrast, at qN1.4, lines carrying the A allele exhibited both a high PN under N100 conditions (7.58) and a PN-ratio close to 1, suggesting stable productivity across nitrogen levels. Interestingly, in qN2.3 and qN7.2, the minor alleles were associated with increased panicle numbers and higher grain yield. Furthermore, the SNP chr09:14678888 at qN9.1b was identified as a promising candidate marker for breeding, as it was associated with NUE-100 in two independent trials and varieties carrying the favourable allele at this site displayed the highest GW in the population. Similarly,

qN12.2 was notable for its contribution to an increased panicle number under N100 conditions and for its consistent effects observed across trials.

Detailed annotation of these QTL intervals revealed 29 candidate genes associated with phenotypic variation in either NUE and PN ratio, and under N100 conditions. The nature of the different genes indicated the complexity of the yield under low N inputs and the NUE. Candidate genes involved in N metabolism could be grouped into functional categories including those related to nitrogen uptake, transport, assimilation, and remobilization. However, other functional categories were also identified, such as transcriptional regulation, hormone biosynthesis, tillering and other developmental processes. No candidate genes were found to be associated with qN6.1, qN9.1 and qN11.1.

Nitrogen is one of the limiting nutrients for plant growth and development. Rice farmers usually supply N in the soil in the form of ammonia or urea. In flooded rice paddies, where ammonium is the predominant nitrogen form, the *AMT* family, including *OsAMT1;2*, and *OsAMT1;3*, becomes especially relevant in the ammonium uptake (Bao et al. 2015). These genes appeared in qN2.3 associated to PN ratio and NUE-100. In addition to *OsAMT1;2* and *OsAMT1;3*, *OsUPS1*, encoding a urea transporter, emerged as candidate genes in the PN-ratio association analysis, indicating that urea uptake pathways contribute to optimizing nitrogen acquisition; this is in agreement with previous overexpression studies that shows that enhanced *OsUPS1* activity improves growth under limited nitrogen (Redillas et al. 2019). Furthermore, *OsUAH* which catalyzes the final step in ureide degradation by converting ureidoglycolate into glyoxylate and ammonium, an alternative nitrogen source under suboptimal conditions (Werner & Witte, 2011), was detected under low nitrogen treatment, PN-100, underscoring its role in providing an alternative nitrogen source when supply is limited.

Genes encoding enzymes implicated in ammonium assimilation, such as a *OsARE1* and *OsGDH3*, were also found in NUE-100. *OsARE1* encodes a chloroplast-localized protein that acts as a negative regulator of N assimilation, in part by repressing Fd-GOGAT activity within the GS/GOGAT cycle (Wang et al. 2018). Nitrogen assimilation is mediated by the coupled reactions catalyzed by glutamine synthetase (GS) and glutamate synthase (GOGAT). Previous studies have shown that loss-of-function mutations in *OsARE1* lead to enhanced NUE, delayed

senescence, and a 10–20% increase in grain yield under nitrogen-limiting conditions (Wang et al. 2018). *OsGDH3* encodes a glutamate deshydrogenase that catalyzes the reversible conversion of glutamate, offering metabolic flexibility. Despite its primary function *in vivo* appears to be glutamate catabolism and nitrogen remobilization, particularly during senescence or carbon limitation, there are evidence that it may also contribute to ammonium assimilation under stress or high ammonium conditions (Dubois et al., 2003). As with other plants, multiple *GDH* genes are present in rice, yet they appear to play a minor role in primary ammonium assimilation compared to the GS/GOGAT pathway. Nevertheless, QTL studies have linked GDH activity to grain yield and NUE, suggesting that GDH may play a role in optimizing nitrogen metabolism, particularly during phases of nitrogen remobilization and translocation (Dubois et al. 2003).

OsAAP3/12A and OsAAP6 are two amino acid permeases that have been reported to affect the distribution of various amino acids in plants, which is essential for optimal nitrogen allocation under varying supply conditions. Previous studies on related AAPs, such as OsAAP4, have linked efficient amino acid transport to improved NUE and grain yield (Taylor et al. 2015; H. Zhao et al. 2012). OsAAP6 acts as a positive regulator of the grain protein content and grain quality in rice (Fang Rice 2021). *OsAAP3* is located in qN6.4, a robust QTL associated to NUE-ratio, reinforcing its key role in nitrogen metabolism via amino acid transport.

Other gene involved in redistribution of N, *OsNRT1.7*, was found in qN1.4 associated to PN-ratio. In rice, *OsNRT1.7* plays a key role in reallocating nitrate within the plant, particularly under suboptimal nitrogen conditions, thereby supporting new growth and enhancing NUE (Fan et al. 2009; X. Fan et al. 2017).

Autophagy is involved in N remobilization process through the degradation and recycling of cellular components. Previous research has shown that enhanced autophagy improves nitrogen remobilization, plant fitness, and yield (Yu et al. 2019; Zhen et al. 2019). In line with these findings, overexpressing *OsATG8b* enhances nitrogen recycling to grains and overall grain quality, whereas autophagy-deficient mutants in Arabidopsis and maize exhibit reduced nitrogen remobilization, lower seed nitrogen content, and impaired yields (Fan et al. 2020). Genes involved in the autophagy regulation, such as *OsATG8a* and *OsATG8c*, were found in qN8.1 and qN7.2, both QTLs associated with PN-100, which aligns with their functions under

moderate nitrogen inputs and in accordance with previous studies. The role of *OsATG8a* and *OsATG8c* in autophagy under limited nitrogen conditions have been reported previously, establishing their role in buffering nitrogen fluctuations. Overexpression of *OsATG8a* and *OsATG8c* significantly enhanced the level of autophagy and the number of effective tillers in the transgenic rice (Yu et al. 2019; Zhen et al. 2019). It has been observed that the overexpression of both genes leads to an increase in N uptake, but while overexpression of *OsATG8a* under suboptimal N conditions did not seem to have any effect on yield and NUE, the overexpression of *OsATG8c* significantly increased yield and NUE in the transgenic rice under both optimal and suboptimal N conditions. qN8.1, co-localized with *OsATG8c*, is also a strong QTL associated to PN-ratio. In addition, regarding to genes involved in remobilization, *NPF5.5* was found in qN10.1 and encodes a peptide transporter capable of translocating small peptides and potentially nitrate across the cell membrane (Zhu et al. 2023).

As mentioned above, qN1.4 is one of the most promising QTLs due to the differences in PN-100 among varieties according to their allele content. The search for candidate genes within this QTLs led us to the detection of *OsBT* and *OsNRT1.7*. *OsBT* is a negative regulator of nitrogen uptake and it has been observed that the lack of function of *OsBT* produces a 20% increase in NUE under low-nitrogen conditions, likely due to increased nitrate transporter activity and improved root system expansion (Araus et al. 2016). *OsNRT1.7* facilitates nitrate redistribution within the plant, particularly under suboptimal nitrogen conditions (Fan et al. 2009; Fan et al. 2017). Given that these genes operate under suboptimal nitrogen conditions, their presence in qN1.4 reinforce its potentiality as candidate genes in modulating panicle development when nitrogen is limited.

Finally, we identified two *CNX2* homologs, located on different chromosomes, that are involved in the biosynthesis a molybdenum cofactor, which is essential for nitrate reductase (NR) activity. NR catalyzes the first and rate limiting step in nitrate assimilation by reducing nitrate to nitrite, a crucial process for converting inorganic nitrogen into organic forms that support plant growth. Without proper molybdenum cofactor synthesis, NR activity would be severely compromised, resulting in inefficient assimilation (Schwarz & Mendel, 2006).

Under low nitrogen conditions (N100), the analysis highlights several candidate genes that are associated with the plant's response to reduced nitrogen supply. *OsDEP1* encodes a phosphatidylethanolamine-binding protein (PEBP) that plays a critical role in regulating panicle architecture. Loss-of-function mutations in *OsDEP1* result in erect panicles, a trait that was favorably selected during rice domestication for improved light interception, yield stability, and lodging resistance (Huang et al. 2009; Zhou et al. 2009). Beyond its well-established role in shaping plant architecture, recent research indicates that allelic variation at the *OsDEP1* locus also impacts nitrogen metabolism. Specifically, the dep1 allele, which produces a truncated protein, has been linked to enhanced nitrogen uptake and assimilation, leading to improved grain filling and increased panicle weight under both low and high nitrogen conditions (Zhao et al. 2019).

The number of tillers rice is an important factor in the final yield as tiller determines the number of panicles. The production of tillers is affected by environmental factors, such as nutrient availability. Known genes play a dual role in regulating NUE and grain productivity as well as promoting tillering. This is the case of *OsGATA8-H*, where “H” designates a specific haplotype, acts as a transcription factor that represses key high-affinity ammonium transporters, particularly *OsAMT1;2* and *OsAMT1;3*. Knockout studies have revealed that loss of *OsGATA8-H* function leads to increased expression of these AMTs. This results in enhanced ammonium uptake and higher tillering (Wu et al. 2024). *OsGATA8*, is located in qN1.1 while *OsAMT1;2* and *OsAMT1;3*, are located in qN2.3 respectively, which are associated with PN-ratio. *OsERF142/NGR5* encides an AP2/ERF transcription factor has been shown to modulate root meristem proliferation and tillering in response to variations in nitrogen availability (Li et al. 2015; Wu et al. 2020). Previous studies have found that *ERF142/NGR5* regulates tillering and NUE participating in the interaction of carbon assimilation and the nitrate signal (Wen et al. 2024). *ERF142/NGR5* is a potential candidate gene found in the strong QTL qN5.2, which is associated with GW-ratio and GW-100.

TCP transcription factors participate in diverse developmental processes including branching (Manassero et al. 2013). Among them, *OsTb2* is known to play a role in regulating palea development and floral zygomorphy through vascular patterning (Yin et al. 2018; Z. Yuan et al. 2009). Despite *OsTb2* is primarily associated with floral organogenesis, its presence suggests that it may indirectly impact NUE

through developmental regulators, potentially by affecting spikelet architecture and resource allocation in response to different nitrogen conditions. *OsTb2* also promotes tillering, in contrast to the inhibitory effect of its homologous *OsTb1*. According to functional analyses, OsTb2 protein interacts with its homologous OsTb1 protein, counteracting its inhibitory effect (Lyu et al. 2020). *OsTb2* was detected in qN9.1b, which is consistently associated with various traits and assays. *OsTCP4*, which co-localized with qN2.3, encodes another TCP transcription factor that is known to have an inhibitory effect on tillering as the lack of function of *OsTCP4* produces plants with an increased number of tillers (Zhang et al. 2024)

Plant hormones play an important role in plant tillering response of plants to fertilizers. As an example, *OsGRF7*, which is co-localized with qN12.1a, modulates plant architecture by regulating gibberellin and auxin metabolism. This influences traits such as leaf angle, culm thickness, and plant height (Chen et al. 2020). Low-nitrogen and low-phosphorus environments significantly promote strigolactone biosynthesis and repress branching of plants by inhibiting lateral bud outgrowth (Umehara et al. 2008). *DWARF27*, in qN11.2 associated to GW-ratio, is a key gene encoding an enzyme of the strigolactone biosynthesis pathway, Molecular studies further confirm that DWARF27 functions as a β-carotene isomerase, catalyzing a key step in strigolactone biosynthesis (Alder et al. 2012; Hao et al., 2009). In addition, *OsD3*, encoding a F-box leucine-trich repeat (LRR) protein, is co-localized with qN6.2b which is associated with PN-ratio, OsD3 participates in strigolactone perception and signal activation through its interaction with DWARF14 (D14) (Tal et al. 2022). DWARF27 and OsD3 play roles in the control of tiller bud dormancy to suppress bud activity (Ishikawa et al. 2005).

Conclusions

The agronomic characterisation of a MAGIC rice population adapted to temperate regions, grown under different fertilisation doses, led to the identification of QTLs and candidate genes associated with yield under low nitrogen input. Using a high-density panel of markers, 22 QTLs were mapped, with the most promising loci located on chromosomes 5, 6, and 8. The associated SNP markers co-localized with known major genes and QTLs, such as *ERF142/NGR5*, *OsAAP12A*, and *OsGATA8c*. The functional diversity of these candidate genes underscores the complexity of NUE, which encompasses nitrogen metabolism, assimilation, transport, and remobilization, as well as developmental traits such as tillering and panicle architecture. These findings not only deepen our understanding of the genetic basis of NUE but also offer valuable molecular markers for breeding high-yielding rice cultivars under reduced nitrogen input, contributing to both agricultural sustainability and economic efficiency.

References

Adrian Alder, Muhammad Jamil, Mattia Marzorati, Mark Bruno, Martina Vermathen, Peter Bigler, Sandro Ghisla, Harro Bouwmeester, Peter Beyer, & Salim Al-Babili. (2012). The Path from b-Carotene to Carlactone, a Strigolactone-Like Plant Hormone. Science, 335(6074), 1344–1348. https://doi.org/10.1126/science.1215933

Araus, V., Vidal, E. A., Puelma, T., Alamos, S., Mieulet, D., Guiderdoni, E., & Gutiérrez, R. A. (2016). Members of BTB gene family of scaffold proteins suppress nitrate uptake and nitrogen use efficiency. Plant Physiology, 171(2), 1523–1532. https://doi.org/10.1104/pp.15.01731

Arite, T., Iwata, H., Ohshima, K., Maekawa, M., Nakajima, M., Kojima, M., Sakakibara, H., & Kyozuka, J. (2007). DWARF10, an RMS1/MAX4/DAD1 ortholog, controls lateral bud outgrowth in rice. Plant Journal, 51(6), 1019–1029. https://doi.org/10.1111/j.1365-313X.2007.03210.x

Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C. J., Tung, C. W., McCouch, S., Thomson, M., Mauleon, R., Singh, R. K., Gregorio, G., Redoña, E., & Leung, H. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: Progress and potential for genetics research and breeding. Rice, 6(1), 1–15. https://doi.org/10.1186/1939-8433-6-1

Bao, A., Liang, Z., Zhao, Z., & Cai, H. (2015). Overexpressing of OsAMT1-3, a high-affinity ammonium transporter gene, modifies rice growth and carbon-nitrogen metabolic status. International Journal of Molecular Sciences, 16(5), 9037-9063.Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, 23(19). https://doi.org/10.1093/bioinformatics/btm308

Cao, F. Q., Werner, A. K., Dahncke, K., Romeis, T., Liu, L. H., & Witte, C. P. (2010). Identification and characterization of proteins involved in rice urea and Arginine catabolism. Plant Physiology, 154(1), 98–108. https://doi.org/10.1104/pp.110.160929

Chen, Y., Dan, Z., Gao, F., Chen, P., Fan, F., & Li, S. (2020). Rice GROWTH-REGULATING FACTOR7 Modulates Plant Architecture through Regulating GA and Indole-3-Acetic Acid Metabolism1[OPEN]. Plant Physiology, 184(1), 393–406. https://doi.org/10.1104/PP.20.00302

Chen, Y., Fan, X., Song, W., Zhang, Y., & Xu, G. (2012). Over-expression of OsPIN2 leads to increased tiller numbers, angle and shorter plant height through suppression of OsLAZY1. Plant Biotechnology Journal, 10(2), 139–149. https://doi.org/10.1111/j.1467-7652.2011.00637.x

Dobermann, A., Cassman, & K. G. (2002). Plant nutrient management for enhanced productivity in intensive grain production systems of the United States and Asia. Plant and Soil, 247(1), 153–175. https://doi.org/10.1023/A:1021197525875

Du, H., Wu, N., Fu, J., Wang, S., Li, X., Xiao, J., & Xiong, L. (2012). A GH3 family member, OsGH3-2, modulates auxin and abscisic acid levels and differentially affects drought and cold tolerance in rice. Journal of Experimental Botany, 63(18), 6467–6480. https://doi.org/10.1093/jxb/ers300

Dubois, F., Tercé-Laforgue, T., Gonzalez-Moro, M. B., Estavillo, J. M., Sangwan, R., Gallais, A., & Hirel, B. (2003). Glutamate dehydrogenase in plants: Is there a new story for an old enzyme? Plant Physiology and Biochemistry, 41(6-7), 565-576. https://doi.org/10.1016/S0981-9428(03)00075-5

Duan, J., Yu, H., Yuan, K., Liao, Z., Meng, X., Jing, Y., Liu, G., Chu, J., & Li, J. (2019). Strigolactone promotes cytokinin degradation through transcriptional activation of CYTOKININ OXIDASE/ DEHYDROGENASE 9 in rice. Proceedings of the National Academy of Sciences of the United States of America, 116(28), 14319–14324. https://doi.org/10.1073/pnas.1810980116

Fageria, N. K. (2007). Yield physiology of rice. In Journal of Plant Nutrition, 30(6), 843–879. https://doi.org/10.1080/15226510701374831

Fan, S. C., Lin, C. S., Hsu, P. K., Lin, S. H., & Tsay, Y. F. (2009). The Arabidopsis nitrate transporter NRT1.7, expressed in phloem, is responsible for source-to-sink remobilization of nitrate. Plant Cell, 21(9), 2750–2761. https://doi.org/10.1105/tpc.109.067603

Fan, T., Yang, W., Zeng, X., Xu, X., Xu, Y., Fan, X., Luo, M., Tian, C., Xia, K., & Zhang, M. (2020). A Rice Autophagy Gene OsATG8b Is Involved in Nitrogen Remobilization and Control of Grain Quality. Frontiers in Plant Science, 11(June), 1–15. https://doi.org/10.3389/fpls.2020.00588

Fan, X., Naz, M., Fan, X., Xuan, W., Miller, A. J., & Xu, G. (2017). Plant nitrate transporters: From gene function to application. Journal of Experimental Botany, 68(10), 2463–2475. https://doi.org/10.1093/jxb/erx011

Fang, Z., Wu, B., & Ji, Y. (2021). The amino acid transporter OsAAP4 contributes to rice tillering and grain yield by regulating neutral amino acid allocation through two splicing variants. Rice, 14, 2. https://doi.org/10.1186/s12284-020-00446-9

Fu, J., Liu, H., Li, Y., Yu, H., Li, X., Xiao, J., & Wang, S. (2011). Manipulating broad-spectrum disease resistance by suppressing pathogen-induced auxin accumulation in rice. Plant Physiology, 155(1), 589–602. https://doi.org/10.1104/pp.110.163774

Girsang, S. S., Correa Jr., T. Q., Quilty, J. R., Sanchez, P. B., & Buresh, R. J. (2020). Soil aeration and relationship to inorganic nitrogen during aerobic cultivation of irrigated rice on a consolidated land parcel. Soil and Tillage Research, 202, 104647. https://doi.org/10.1016/j.still.2020.104647

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food security: The challenge of feeding 9 billion people. Science, 327(5967), 812–818. https://doi.org/10.1126/science.1185383

Good, A. G., Shrawat, A. K., & Muench, D. G. (2004). Can less yield more? Is reducing nutrient input into the environment compatible with maintaining crop production? In Trends in Plant Science, 9(12), 597–605. https://doi.org/10.1016/j.tplants.2004.10.008

Guevara, D. R., El-Kereamy, A., Yaish, M. W., Mei-Bi, Y., & Rothstein, S. J. (2014). Functional characterization of the rice UDP-glucose 4-epimerase 1, OsUGE1: A potential role in cell wall carbohydrate partitioning during limiting nitrogen conditions. PLoS ONE, 9(5). https://doi.org/10.1371/journal.pone.0096158

Hao, D., Yang, S., Huang, Y., & Su, Y. (2016). Identification of structural elements involved in fine-tuning of the transport activity of the rice ammonium transporter OsAMT1;3. Plant Physiology and Biochemistry, 108, 99–108. https://doi.org/10.1016/j.plaphy.2016.07.003

Hao, L., Renxiao, W., Qian, Q., Meixian, Y., Xiangbing, M., Zhiming, F., Cunyu, Y., Biao, J., Zhen, S., Jiayang, L., & Yonghong, W. (2009). DWARF27,an iron-containing protein required for the biosynthesis of strigolactones,regulates rice tiller bud outgrowth. Plant Cell, 21(5), 1512–1525. https://doi.org/10.1105/tpc.109.065987

Hou, M., Yu, M., Li, Z., Ai, Z., & Chen, J. (2021). Molecular regulatory networks for improving nitrogen use efficiency in rice. International Journal of Molecular Sciences, 22(16), 1–15. https://doi.org/10.3390/ijms22169040

Huang, W., Bai, G., Wang, J., Zhu, W., Zeng, Q., Lu, K., Sun, S., & Fang, Z. (2018). Two splicing variants of OsNPF7.7 regulate shoot branching and nitrogen utilization efficiency in rice. Frontiers in Plant Science, 9. https://doi.org/10.3389/fpls.2018.00300

Huang, X., Qian, Q., Liu, Z., Sun, H., He, S., Luo, D., Xia, G., Chu, C., Li, J., & Fu, X. (2009). Natural variation at the DEP1 locus enhances grain yield in rice. Nature Genetics, 41(4), 494–497. https://doi.org/10.1038/ng.352

Ishikawa, S., Maekawa, M., Arite, T., Onishi, K., Takamure, I., & Kyozuka, J. (2005). Suppression of tiller bud activity in tillering dwarf mutants of rice. Plant & Cell Physiology, 46(1), 79–86. https://doi.org/10.1093/pcp/pci022

Jain, M., Kaur, N., Tyagi, A. K., & Khurana, J. P. (2006). The auxin-responsive GH3 gene family in rice (Oryza sativa). Functional and Integrative Genomics, 6(1), 36–46. https://doi.org/10.1007/s10142-005-0142-5

Konishi, N., & Ma, J. F. (2021). Three polarly localized ammonium transporter 1 members are cooperatively responsible for ammonium uptake in rice under low ammonium condition. New Phytologist, 232(4), 1778–1792. https://doi.org/10.1111/nph.17679

Ladha, J. K., Pathak, H., Krupnik, T. J., Six, J., & van Kessel, C. (2005). Efficiency of Fertilizer Nitrogen in Cereal Production: Retrospects and Prospects. In Advances in Agronomy 87, 85–156. https://doi.org/10.1016/S0065-2113(05)87003-8

Lee, D.-K., Redillas, M. C. F. R., Jung, H., Choi, S., Kim, Y. S., & Kim, J.-K. (2018). A nitrogen molecular sensing system, comprised of the ALLANTOINASE and UREIDE PERMEASE 1 genes, can be used to monitor N status in rice. Frontiers in Plant Science, 9, 444. https://doi.org/10.3389/fpls.2018.00444

Li, H., Hu, B., & Chu, C. (2017). Nitrogen use efficiency in crops: Lessons from Arabidopsis and rice. Journal of Experimental Botany, 68(10), 2477–2488. https://doi.org/10.1093/jxb/erx101

Li, J., Zhao, Y., Chu, H., Wang, L., Fu, Y., Liu, P., Upadhyaya, N., Chen, C., Mou, T., Feng, Y., Kumar, P., & Xu, J. (2015). SHOEBOX Modulates Root Meristem Size in Rice through Dose-Dependent Effects of Gibberellins on Cell Elongation and Proliferation. PLoS Genetics, 11(8). https://doi.org/10.1371/journal.pgen.1005464

Li, K., Zhang, S., Tang, S., Zhang, J., Dong, H., Yang, S., Qu, H., Xuan, W., Gu, M., & Xu, G. (2022). The rice transcription factor Nhd1 regulates root growth and nitrogen uptake by activating nitrogen transporters. Plant Physiology, 189(3), 1608–1624. https://doi.org/10.1093/plphys/kiac178

Li, S. (2024). Is auxin the key to improve crop nitrogen use efficiency for greener agriculture? New Phytologist, 2170–2175. https://doi.org/10.1111/nph.20066

Li, Y., Ouyang, J., Wang, Y. Y., Hu, R., Xia, K., Duan, J., Wang, Y., Tsay, Y. F., & Zhang, M. (2015). Disruption of the rice nitrate transporter OsNPF2.2 hinders root-to-shoot nitrate transport and vascular development. Scientific Reports, 5. https://doi.org/10.1038/srep09635

Lian, L., Lin, Y., Wei, Y., He, W., Cai, Q., Huang, W., Zheng, Y., Xu, H., Wang, F., Zhu, Y., Luo, X., Xie, H., & Zhang, J. (2021). PEPC of sugarcane regulated glutathione S-transferase and altered carbon–nitrogen metabolism under different N source concentrations in Oryza sativa. BMC Plant Biology, 21(1), 1–15. https://doi.org/10.1186/s12870-021-03071-w

LiLin-Yin. (2024). CMplot: Circle Manhattan Plot (R package version 4.5.1).

Liu, Q., Wu, K., Song, W., Zhong, N., Wu, Y., & Fu, X. (2022). Improving crop nitrogen use efficiency toward sustainable green revolution. Annual Review of Plant Biology, 73, 523–551. https://doi.org/10.1146/annurev-arplant-070121-015752

Liu, X., Wang, J., Yu, Y., Kong, L., Liu, Y., Liu, Z., Li, H., Wei, P., Liu, M., Zhou, H., Bu, Q., & Fang, J. (2019). Identification and characterization of the rice pre-harvest sprouting mutants involved in molybdenum cofactor biosynthesis. New Phytologist, 222(1), 275–285. https://doi.org/10.1111/nph.15607

Liu, Y., Wang, H., Jiang, Z., Wang, W., Xu, R., Wang, Q., Zhang, Z., Li, A., Liang, Y., Ou, S., Liu, X., Cao, S., Tong, H., Wang, Y., Zhou, F., Liao, H., Hu, B., & Chu, C. (2021). Genomic basis of geographical adaptation to soil nitrogen in rice. Nature, 590(7847), 600–605. https://doi.org/10.1038/s41586-020-03091-w

Lyu, J., Huang, L., Zhang, S., Zhang, Y., He, W., Zeng, P., Zeng, Y., Huang, G., Zhang, J., Ning, M., Bao, Y., Zhao, S., Fu, Q., Wade, L. J., Chen, H., Wang, W., & Hu, F. (2020). Neo-functionalization of a Teosinte branched 1 homologue mediates adaptations of upland rice. Nature Communications, 11, 725. https://doi.org/10.1038/s41467-019-14264-1

Manassero, N. G. U., Viola, I. L., Welchen, E., & Gonzalez, D. H. (2013). TCP transcription factors: Architectures of plant form. Biomolecular Concepts, 4(2), 111–127. https://doi.org/10.1515/bmc-2012-0051

Masclaux-Daubresse, C., Daniel-Vedele, F., Dechorgnat, J., Chardon, F., Gaufichon, L., & Suzuki, A. (2010). Nitrogen uptake, assimilation and remobilization in plants: Challenges for sustainable and productive agriculture. Annals of Botany, 105(7), 1141–1157. https://doi.org/10.1093/aob/mcq028

Masumoto, C., Miyazawa, S.-I., Ohkawa, H., Fukuda, T., Taniguchi, Y., & Mae, T. (2010). Phosphoenolpyruvate carboxylase intrinsically located in the chloroplast of rice plays a crucial role in ammonium assimilation. Proceedings of the National Academy of Sciences of the USA, 107(11), 5226–5231. https://doi.org/10.1073/pnas.0913127107

Muramatsu, M., Suzuki, R., Yamazaki, T., & Miyao, M. (2015). Comparison of plant-type phosphoenolpyruvate carboxylases from rice: Identification of two plant-specific regulatory regions of the allosteric enzyme. Plant & Cell Physiology, 56(3), 468–480. https://doi.org/10.1093/pcp/pcu189

Muthayya, S., Sugimoto, J. D., Montgomery, S., & Maberly, G. F. (2014). An overview of global rice production, supply, trade, and consumption. Annals of the New York Academy of Sciences, 1324(1), 7–14. https://doi.org/10.1111/nyas.12540

Nishimura, A., Ashikari, M., Lin, S., Takashi, T., Angeles, E. R., Yamamoto, T., & Matsuoka, M. (2005). Isolation of a rice regeneration quantitative trait loci gene and its application to transformation systems. www.pnas.orgcgidoi10.1073pnas.0504220102

Patil, I. (2021). Visualizations with statistical details: The "ggstatsplot" approach. Journal of Open Source Software, 6(61), 3167. https://doi.org/10.21105/joss.03167

Qi, Z., & Xiong, L. (2013). Characterization of a Purine Permease Family Gene OsPUP7 Involved in Growth and Development Control in Rice. Journal of Integrative Plant Biology, 55(11), 1119–1135. https://doi.org/10.1111/jipb.12101

Redillas, M. C. F. R., Bang, S. W., Lee, D. K., Kim, Y. S., Jung, H., Chung, P. J., Suh, J. W., & Kim, J. K. (2019). Allantoin accumulation through overexpression of ureide permease1 improves rice growth under limited nitrogen conditions. Plant Biotechnology Journal, 17(7), 1289–1301. https://doi.org/10.1111/pbi.13054

Riemann, M., Riemann, M., & Takano, M. (2008). Rice JASMONATE RESISTANT 1 is involved in phytochrome and jasmonate signalling. Plant, Cell and Environment, 31(6), 783–792. https://doi.org/10.1111/j.1365-3040.2008.01790.x

Sade, N., Umnajkitikorn, K., Rubio Wilhelmi, M. D. M., Wright, M., Wang, S., & Blumwald, E. (2018). Delaying chloroplast turnover increases water-deficit stress tolerance through the enhancement of nitrogen assimilation in rice. Journal of Experimental Botany, 69(4), 867–878. https://doi.org/10.1093/jxb/erx247

Schwarz, G., & Mendel, R. R. (2006). Molybdenum cofactor biosynthesis and molybdenum enzymes. In Annual Review of Plant Biology (Vol. 57, pp. 623–647). https://doi.org/10.1146/annurev.arplant.57.032905.105437

Sonoda, Y., Ikeda, A., Saiki, S., Von Wirén, N., Yamaya, T., & Yamaguchi, J. (2003). Distinct Expression and Function of Three Ammonium Transporter Genes (OsAMT1;1-1;3) in Rice. Plant and Cell Physiology, 44(7), 726-734. https://doi.org/10.1093/pcp/pcg083

Suenaga, A., Moriya, K., Sonoda, Y., Ikeda, A., Von Wirén, N., Hayakawa, T., Yamaguchi, J., & Yamaya, T. (2003). Constitutive Expression of a Novel-Type Ammonium Transporter OsAMT2 in Rice Plants. In Plant Cell Physiol (Vol. 44, Issue 2). http://pcp.oxfordjournals.org/

Tabuchi, M., Abiko, T., & Yamaya, T. (2007). Assimilation of ammonium ions and reutilization of nitrogen in rice (Oryza sativa L.). Journal of Experimental Botany, 58(9), 2319–2327. https://doi.org/10.1093/jxb/erm016

Takayanagi, S., Takagi, Y., Araki, R., & Hasegawa, H. (2011). High-affinity nitrate uptake by rice (Oryza sativa) coleoptiles. Journal of Plant Research, 124(2), 305–309. https://doi.org/10.1007/s10265-010-0375-9

Tal, L., Palayam, M., Ron, M., Young, A., Britt, A., & Shabek, N. (2022). A conformational switch in the SCF-D3/MAX2 ubiquitin ligase facilitates strigolactone signalling. Nature Plants, 8(5), 561–573. https://doi.org/10.1038/s41477-022-01145-7

Tamura, W., Kojima, S., Toyokawa, A., Watanabe, H., Tabuchi-Kobayashi, M., Hayakawa, T., & Yamaya, T. (2011). Disruption of a novel NADH-glutamate synthase2 gene caused marked reduction in spikelet number of rice. Frontiers in Plant Science, 2(SEP), 1–9. https://doi.org/10.3389/fpls.2011.00057

Taylor, M. R., Reinders, A., & Ward, J. M. (2015). Transport function of rice amino acid permeases (AAPs). Plant and Cell Physiology, 56(7), 1355–1363. https://doi.org/10.1093/pcp/pcv053

Terada, Y., Aoki, H., Ida, S., Tanaka, T., & Morikawa, H. (1995). Cloning and Nucleotide Sequence of a Leaf Ferredoxin-Nitrite Reductase cDNA of Rice. Bioscience, Biotechnology, and Biochemistry, 59(11), 2183–2185. https://doi.org/10.1271/bbb.59.2183

Terol, J., Domingo, C., & Talón, M. (2006). The GH3 family in plants: Genome wide analysis in rice and evolutionary history based on EST analysis. Gene, 371(2), 279–290. https://doi.org/10.1016/j.gene.2005.12.014

Tiong, J. W., Sharma, N., Sampath, R., MacKenzie, N., Watanabe, S., Metot, C., Lu, Z., Skinner, W., Lu, Y., Kridl, J., Baumann, U., Heuer, S., Kaiser, B., & Okamoto, M. (2021). Improving nitrogen use efficiency through overexpression of alanine aminotransferase in rice, wheat, and barley. Frontiers in Plant Science, 12, 628521. https://doi.org/10.3389/fpls.2021.628521

Umehara, M., Hanada, A., Yoshida, S., Akiyama, K., Arite, T., Takeda-Kamiya, N., Magome, H., Kamiya, Y., Shirasu, K., Yoneyama, K., Kyozuka, J., & Yamaguchi, S. (2008). Inhibition of shoot branching by new terpenoid plant hormones. Nature, 455(7210), 195–200. https://doi.org/10.1038/nature07272

Wakuta, S., Suzuki, E., Saburi, W., Matsuura, H., Nabeta, K., Imai, R., & Matsui, H. (2011). OsJAR1 and OsJAR2 are jasmonyl-l-isoleucine synthases involved in wound- and pathogen-induced jasmonic acid signalling. Biochemical and Biophysical Research Communications, 409(4), 634–639. https://doi.org/10.1016/j.bbrc.2011.05.055

Wang, J. R., Hu, H., Wang, G. H., Li, J., Chen, J. Y., & Wu, P. (2009). Expression of PIN genes in rice (Oryza sativa L.): Tissue specificity and regulation by hormones. Molecular Plant, 2(4), 823–831. https://doi.org/10.1093/mp/ssp023

Wang, Q., Nian, J., Xie, X., Yu, H., Zhang, J., Bai, J., Dong, G., Hu, J., Bai, B., Chen, L., Xie, Q., Feng, J., Yang, X., Peng, J., Chen, F., Qian, Q., Li, J., & Zuo, J. (2018). Genetic variations in ARE1 mediate grain yield by modulating nitrogen utilization in rice. Nature Communications, 9(1). https://doi.org/10.1038/s41467-017-02781-w

Wang, S., Chen, A., Xie, K., Yang, X., Luo, Z., Chen, J., Zeng, D., Ren, Y., Yang, C., Wang, L., Feng, H., Lizbeth López-Arredondo, D., Rafael Herrera-Estrella, L., Xu, G., Gojon, A., Harrison, M. J., & Wang, E. (2020). Functional analysis of the OsNPF4.5 nitrate transporter reveals a conserved mycorrhizal pathway of nitrogen acquisition in plants. PNAS, 117(28), 16649–16659. https://doi.org/10.1073/pnas.2000926117/-/DCSupplemental

Wang, W., Hu, B., Yuan, D., Liu, Y., Che, R., Hu, Y., Ou, S., Liu, Y., Zhang, Z., Wang, H., Li, H., Jiang, Z., Zhang, Z., Gao, X., Qiu, Y., Meng, X., Liu, Y., Bai, Y., Liang, Y., … Chu, C. (2018). Expression of the nitrate transporter gene OsNRT1.1A/ OsNPF6.3 confers high yield and early maturation in rice. Plant Cell, 30(3), 638–651. https://doi.org/10.1105/tpc.17.00809

Wen, R., Zhu, M., Yu, J., Kou, L., Ahmad, S., Wei, X., Jiao, G., Hu, S., Sheng, Z., Zhao, F., Tang, S., Shao, G., Yu, H., & Hu, P. (2024). Photosynthesis regulates tillering bud elongation and nitrogen-use efficiency via sugar-induced NGR5 in rice. New Phytologist, 243(4), 1440–1454. https://doi.org/10.1111/nph.19921

Werner, A. K., & Witte, C. P. (2011). The biochemistry of nitrogen mobilization: Purine ring catabolism. In Trends in Plant Science (Vol. 16, Issue 7, pp. 381–387). https://doi.org/10.1016/j.tplants.2011.03.012

Witte, C. P. (2011). Urea metabolism in plants. In Plant Science (Vol. 180, Issue 3, pp. 431–438). https://doi.org/10.1016/j.plantsci.2010.11.010

Wu, K., Wang, S., Song, W., Zhang, J., Wang, Y., Liu, Q., Yu, J., Ye, Y., Li, S., Chen, J., Zhao, Y., Wang, J., Wu, X., Wang, M., Zhang, Y., Liu, B., Wu, Y., Harberd, N. P., & Fu, X. (2020). Enhanced sustainable green revolution yield via nitrogen-responsive chromatin modulation in rice. Science, 367(6478). https://doi.org/10.1126/science.aaz2046

Wu, W., Dong, X., Chen, G., Lin, Z., Chi, W., Tang, W., Yu, J., Wang, S., Jiang, X., Liu, X., Wu, Y., Wang, C., Cheng, X., Zhang, W., Xuan, W., Terzaghi, W., Ronald, P. C., Wang, H., Wang, C., & Wan, J. (2024). The elite haplotype OsGATA8-H coordinates nitrogen uptake and productive tiller formation in rice. Nature Genetics, 56(7), 1516–1526. https://doi.org/10.1038/s41588-024-01795-7

Yamaya, T., & Kusano, M. (2014). Evidence supporting distinct functions of three cytosolic glutamine synthetases and two NADH-glutamate synthases in rice. In Journal of Experimental Botany (Vol. 65, Issue 19, pp. 5519–5525). Oxford University Press. https://doi.org/10.1093/jxb/eru103

Yan, M., Fan, X., Feng, H., Miller, A. J., Shen, Q., & Xu, G. (2011). Rice OsNAR2.1 interacts with OsNRT2.1, OsNRT2.2 and OsNRT2.3a nitrate transporters to provide uptake over high and low concentration ranges. Plant, Cell and Environment, 34(8), 1360–1372. https://doi.org/10.1111/j.1365-3040.2011.02335.x

Yin, D., Liu, X., Shi, Z., Li, D., & Zhu, L. (2018). An AT-hook protein DEPRESSED PALEA1 physically interacts with the TCP Family transcription factor RETARDED PALEA1 in rice. Biochemical and Biophysical Research Communications, 495(1), 487–492. https://doi.org/10.1016/j.bbrc.2017.11.031

Yu, J., Zhen, X., Li, X., Li, N., & Xu, F. (2019). Increased autophagy of rice can increase yield and nitrogen use efficiency (NUE). Frontiers in Plant Science, 10(May), 1–11. https://doi.org/10.3389/fpls.2019.00584

Yuan, S. jiang, Wang, T., Yin, L., Zhao, J. feng, Wan, J. min, & Li, X. yong. (2013). Cloning and expression of gene responsible for high-tillering dwarf phenotype in indica rice mutant gsor23. Rice Science, 20(5), 320–328. https://doi.org/10.1016/S1672-6308(13)60134-1

Yuan, Z., Fan, K., Wang, Y., Tian, L., Zhang, C., Sun, W., He, H., & Yu, S. (2021). Osgretchenhagen3-2 modulates rice seed storability via accumulation of abscisic acid and protective substances. Plant Physiology, 186(1), 469–482. https://doi.org/10.1093/plphys/kiab059

Yuan, Z., Gao, S., Xue, D. W., Luo, D., Li, L. T., Ding, S. Y., Yao, X., Wilson, Z. A., Qian, Q., & Zhang, D. B. (2009). Retarded Palea1 controls palea development and floral zygomorphy in rice. Plant Physiology, 149(1), 235–244. https://doi.org/10.1104/pp.108.128231

Xu, G., Fan, X., & Miller, A. J. (2012). Plant nitrogen assimilation and use efficiency. Annual Review of Plant Biology, 63, 153–182. https://doi.org/10.1146/annurev-arplant-042811-105532

Zhang, Z., & Chu, C. (2020). Nitrogen-use divergence between indica and *japonica* rice: Variation at nitrate assimilation. Molecular Plant, 13(1), 6–7. https://doi.org/10.1016/j.molp.2019.11.011

Zhang, H., Zhang, J., Xu, P., Li, M., & Li, Y. (2024). Insertion of a miniature inverted-repeat transposable element into the promoter of OsTCP4 results in more tillers and a lower grain size in rice. Journal of Experimental Botany, 75(5), 1421–1436. https://doi.org/10.1093/jxb/erad467

Zhao, H., Ma, H., Yu, L., Wang, X., & Zhao, J. (2012). Genome-Wide Survey and Expression Analysis of Amino Acid Transporter Gene Family in Rice (Oryza sativa L.). PLoS ONE, 7(11). https://doi.org/10.1371/journal.pone.0049210

Zhao, M., Zhao, M., Gu, S., Sun, J., Ma, Z., Wang, L., Zheng, W., & Xu, Z. (2019). DEP1 is involved in regulating the carbon–nitrogen metabolic balance to affect grain yield and quality in rice (Oriza sativa L.). PLoS ONE, 14(3). https://doi.org/10.1371/journal.pone.0213504

Zhen, X., Xu, F., Zhang, W., Li, N., & Li, X. (2019). Overexpression of rice gene OsATG8b confers tolerance to nitrogen starvation and increases yield and nitrogen use efficiency (NUE) in Arabidopsis. PLoS ONE, 14(9). https://doi.org/10.1371/journal.pone.0223011

Zheng, X., Zhang, S., Liang, Y., Zhang, R., Liu, L., Qin, P., Zhang, Z., Wang, Y., Zhou, J., Tang, X., & Zhang, Y. (2023). Loss-function mutants of OsCKX gene family based on CRISPR-Cas systems revealed their diversified roles in rice. Plant Genome, 16(2). https://doi.org/10.1002/tpg2.20283

Zhou, Y., Zhu, J., Li, Z., Yi, C., Liu, J., Zhang, H., Tang, S., Gu, M., & Liang, G. (2009). Deletion in a quantitative trait gene qPE9-1 associated with panicle erectness improves plant architecture during rice domestication. Genetics, 183(1), 315–324. https://doi.org/10.1534/genetics.109.102681

Zhu, X., Chen, L., Zhang, Z., Li, J., Zhang, H., Li, Z., Pan, Y., & Wang, X. (2023). Genetic-based dissection of resistance to bacterial leaf streak in rice by GWAS. BMC Plant Biology, 23(1). https://doi.org/10.1186/s12870-023-04412-7

Supplementary Data

Tables

qN5.2; A1; 71 SNPs	
SNP	p-value
S05_17440428	7.52E-05
S05_17450599	3.28E-05
S05_17457559	4.67E-05
S05_17469028	1.01E-04
S05_17497185	9.58E-05
S05_17526580	9.16E-05
S05_17526604	1.01E-04
S05_17529248	1.06E-04
S05_17533106	6.94E-05
S05_17566946	1.09E-04
S05_17600737	7.35E-05
S05_17601084	9.81E-05
S05_17601126	2.06E-04
S05_17601467	8.58E-05
S05_17615486	1.17E-04
S05_17642740	6.00E-04
S05_17650627	4.72E-05
S05_17657070	9.39E-05
S05_17666171	4.37E-05
S05_17666176	5.00E-05
S05_17667835	1.03E-04
S05_17679327	9.13E-05
S05_17686465	3.14E-05
S05_17696709	1.05E-04
S05_17708670	9.82E-05
S05_17718128	7.74E-05
S05_17739086	8.89E-05
S05_17775702	1.27E-04
S05_17810196	2.06E-04
S05_17810688	7.69E-05
S05_17818503	1.42E-04
S05_17824775	2.08E-04
S05_17824854	8.43E-05
S05_17839889	4.23E-05
S05_17840482	8.43E-05
S05_17845586	3.00E-05
S05_17845591	3.00E-05

qN5.2; A3; 6 SNPs	
SNP	p-value
S05_19587320	6.17E-04
S05_19628090	8.36E-04
S05_19644955	7.05E-04
S05_19694480	8.49E-04
S05_19816533	7.27E-04
S05_19933023	8.76E-04

qN6.4; A1; 7 SNPs	
SNP	p-value
S06_20251259	2.97E-04
S06_20406562	1.75E-04
S06_20468029	3.18E-04
S06_20807407	3.15E-04
S06_20973930	2.29E-06
S06_21029697	2.77E-06
S06_21311240	2.10E-06

qN6.6; A3; 2 SNPs	
SNP	p-value
S06_26734351	4.93E-04
S06_26742604	7.87E-04

qN9.1b; A2; 7 SNPs	
SNP	p-value
S09_15202160	9.68E-04
S09_15225485	7.02E-04
S09_15239028	5.37E-04
S09_15243137	5.41E-04
S09_15251527	6.36E-04
S09_15261427	8.02E-04
S09_15266111	8.02E-04

qN11.2; A1; 12 SNPs	
SNP	p-value
S11_22632966	9.95E-04
S11_22645067	3.63E-04
S11_22651757	5.74E-04

qN9.1a; A2; 163 SNPs			
SNP	p-value	SNP	p-value
S09_7016960	9.83E-04	S09_7197112	9.04E-04
S09_7017502	3.40E-04	S09_7198005	2.55E-04
S09_7022245	6.32E-04	S09_7217513	7.02E-04
S09_7036452	6.59E-04	S09_7230378	5.47E-04
S09_7036453	6.47E-04	S09_7238416	5.24E-04
S09_7041873	9.11E-04	S09_7242003	6.58E-04
S09_7044840	2.22E-04	S09_7246526	7.64E-04
S09_7046374	7.63E-04	S09_7254113	7.52E-04
S09_7047664	5.36E-04	S09_7254242	8.85E-04
S09_7048851	7.00E-04	S09_7254249	8.79E-04
S09_7049055	5.75E-04	S09_7254545	9.73E-04
S09_7055398	7.50E-04	S09_7254919	7.25E-04
S09_7056847	6.48E-04	S09_7264422	3.73E-04
S09_7057644	5.66E-04	S09_7264429	6.12E-05
S09_7057681	6.24E-04	S09_7264672	7.45E-04
S09_7057865	2.60E-04	S09_7282505	6.06E-04
S09_7059999	8.37E-04	S09_7282513	5.14E-04
S09_7060009	8.82E-04	S09_7282518	5.00E-04
S09_7062272	6.34E-04	S09_7283536	9.78E-04
S09_7074597	6.38E-04	S09_7283549	3.41E-04
S09_7074640	5.69E-04	S09_7289874	5.26E-04
S09_7076288	2.44E-04	S09_7289880	7.79E-04
S09_7080358	6.87E-04	S09_7289907	6.13E-04
S09_7080834	8.26E-04	S09_7289946	3.66E-04
S09_7080849	8.18E-04	S09_7289968	5.64E-04
S09_7081609	6.30E-04	S09_7292912	6.78E-04
S09_7082470	3.42E-04	S09_7296789	8.11E-04
S09_7082484	2.87E-04	S09_7308568	6.75E-04
S09_7082513	3.89E-04	S09_7308594	7.96E-04
S09_7085984	6.85E-04	S09_7308658	3.52E-04
S09_7086680	6.01E-04	S09_7312269	5.37E-04
S09_7089027	4.22E-04	S09_7312329	5.11E-04
S09_7094450	9.41E-04	S09_7312336	6.83E-04
S09_7097077	8.80E-04	S09_7312341	6.80E-04
S09_7101240	1.07E-04	S09_7312395	5.19E-04
S09_7102024	6.40E-04	S09_7313715	5.32E-04
S09_7104777	9.05E-04	S09_7315943	5.50E-04

S05_17849015	6.62E-05	S11_23082085	3.94E-04	S09_7104820	9.74E-04	S09_7320554	7.85E-04
S05_17850698	1.01E-04	S11_23084754	5.50E-04	S09_7104974	3.79E-04	S09_7324303	5.70E-04
S05_17854881	1.04E-04	S11_23169437	5.65E-04	S09_7107436	7.33E-04	S09_7333342	4.44E-04
S05_17856103	1.94E-04	S11_23172685	4.02E-04	S09_7117895	1.05E-04	S09_7338408	2.80E-04
S05_17869975	8.48E-05	S11_23172685	4.02E-04	S09_7118481	1.69E-04	S09_7380916	2.67E-04
S05_17875247	1.05E-04	S11_23184180	6.08E-04	S09_7118485	1.55E-04	S09_7390590	6.87E-04
S05_17910037	4.39E-05	S11_23197717	5.82E-04	S09_7118488	1.68E-04	S09_7396477	4.97E-04
S05_17911593	1.12E-04	S11_23199393	8.82E-04	S09_7119573	6.90E-05	S09_7408814	5.74E-04
S05_17911622	1.06E-04	S11_23435912	9.47E-04	S09_7120351	1.61E-04	S09_7411114	9.26E-04
S05_17911703	9.09E-05			S09_7120574	1.78E-04	S09_7411848	6.19E-04
S05_17932306	9.84E-05			S09_7127451	1.26E-04	S09_7411853	3.78E-04
S05_17933723	9.69E-05			S09_7127892	5.35E-05	S09_7433198	5.94E-04
S05_17961983	7.39E-05			S09_7128144	1.46E-04	S09_7433648	7.24E-04
S05_18010964	8.85E-05			S09_7129648	1.58E-04	S09_7434926	8.02E-04
S05_18027330	8.35E-05			S09_7129661	1.92E-04	S09_7438372	7.24E-04
S05_18028483	1.00E-04			S09_7129678	1.02E-04	S09_7447584	4.48E-04
S05_18028924	6.03E-04			S09_7130875	1.75E-04	S09_7469726	2.54E-04
S05_18030885	1.79E-04			S09_7134210	5.21E-05	S09_7478703	7.19E-04
S05_18041108	7.26E-05			S09_7134211	5.21E-05	S09_7478720	7.32E-04
S05_18045915	9.47E-05			S09_7138386	1.35E-04	S09_7483954	8.25E-04
S05_18091137	8.95E-05			S09_7139080	1.32E-04	S09_7487540	9.23E-04
S05_18102777	1.99E-04			S09_7147437	2.48E-04	S09_7491291	5.23E-04
S05_18922003	5.62E-04			S09_7149348	1.96E-04	S09_7491294	5.80E-04
S05_18990123	5.44E-04			S09_7149389	5.69E-04	S09_7491308	6.11E-04
S05_19053547	1.48E-05			S09_7150380	5.12E-04	S09_7492561	1.85E-04
S05_19069296	8.60E-05			S09_7152482	8.00E-04	S09_7520649	8.24E-04
S05_19085653	2.53E-06			S09_7157442	1.60E-04	S09_7633606	9.15E-04
S05_19103798	1.68E-05			S09_7157510	1.85E-04	S09_7638425	6.97E-04
S05_19119014	1.26E-05			S09_7157513	1.77E-04	S09_7638426	6.97E-04
S05_19397673	3.91E-05			S09_7163328	5.35E-04	S09_7639477	5.86E-04
S05_19411757	2.35E-05			S09_7165882	2.64E-04	S09_7724893	7.83E-04
S05_19439837	8.12E-05			S09_7165904	3.25E-04	S09_7741146	9.58E-04
S05_19756023	7.15E-05			S09_7166222	8.24E-04	S09_7741173	3.65E-04
S05_20422515	7.83E-05			S09_7166231	7.86E-04	S09_7748839	5.25E-04

Supplementary Table 1. List of SNPs included in each QTL associated with **NUE ratio.** Each QTL is labelled with its name, the field trial (assay) in which it was detected, and the total number of associated SNPs. SNPs are denoted by chromosome and physical position (chr_pos format), along with the p-value obtained from the GWAS. The SNP with the lowest p-value within each QTL region is highlighted as the lead SNP.

qN1.1; A1; 9 SNPs	
SNP	**p-value**
S01_12687073	7.11E-04
S01_12732698	5.49E-04
S01_13272051	2.98E-04
S01_13412069	2.66E-04
S01_13418138	3.09E-04
S01_13916091	4.62E-04
S01_13922872	3.97E-04
S01_13925709	5.12E-04
S01_13942728	4.96E-04

qN1.4; A2; 2 SNPs	
SNP	**p-value**
S01_39576121	4.40E-04
S01_40034230	3.92E-04

qN2.2; A2; 1 SNPs	
SNP	**p-value**
S02_8970127	7.40E-04

qN2.3; A2; 19 SNPs	
SNP	**p-value**
S02_24971352	7.05E-04
S02_25143543	6.41E-04
S02_25149160	9.99E-04
S02_25152030	8.24E-04
S02_25153253	8.65E-04
S02_25202669	6.79E-04
S02_25204785	9.25E-04
S02_25206733	9.75E-04
S02_25217011	8.93E-04
S02_25225272	7.47E-04
S02_25228825	6.96E-04
S02_25231082	7.26E-04
S02_25231119	9.05E-04
S02_25231342	7.25E-04
S02_25238707	7.20E-04
S02_25253444	8.16E-04
S02_25284906	6.57E-04
S02_25315944	5.38E-04
S02_25325091	8.62E-04

qN6.2; A3; 1 SNPs	
SNP	**p-value**
S06_3312938	2.05E-04

qN12.1; A2; 3 SNPs	
SNP	**p-value**
S12_17066507	9.27E-04
S12_17438546	5.44E-04
S12_17654889	7.84E-04

qN12.1; A1; 1 SNPs	
SNP	**p-value**
S12_19538903	5.41E-04

Supplementry Table 2. List of SNPs included in each QTL associated with **PN ratio.** Each QTL is labelled with its name, the field trial (assay) in which it was detected, and the total number of associated SNPs. SNPs are denoted by chromosome and physical position (chr_pos format), along with the p-value obtained from the GWAS. The SNP with the lowest p-value within each QTL region is highlighted as the lead SNP.

qN1.2b; A1; 2 SNPs	
SNP	**p-value**
S01_32491201	9.97E-04
S01_32492169	9.57E-04

qN2.3; A1; 3 SNPs	
SNP	**p-value**
S02_26045353	9.54E-04
S02_26103439	7.76E-04
S02_26232385	2.02E-04

qN8.1; A1; 44 SNPs	
SNP	**p-value**
S08_4351455	1.55E-04
S08_4511882	3.91E-04
S08_4532920	6.23E-05
S08_4563249	1.97E-04
S08_4564851	1.55E-04
S08_4584816	6.00E-05
S08_4607800	1.10E-04
S08_4630253	7.63E-04

qN9.1b; A2 and A3; 9 SNPs	
SNP	**p-value**
S09_14673737	3.40E-04
S09_14678888	2.14E-04
S09_15202160	4.43E-04
S09_15225485	3.52E-04
S09_15234847	5.16E-04
S09_15243137	1.55E-04
S09_15251527	2.23E-04
S09_15655372	7.70E-04

qN5.2; A1; 12 SNPs	
SNP	**p-value**
S05_19053547	2.18E-05
S05_19085653	5.48E-06
S05_19103798	1.36E-05
S05_19119014	1.74E-05
S05_19397673	5.23E-05
S05_19411757	1.95E-05
S05_19439837	3.31E-05
S05_19756023	3.25E-05
S05_20422515	3.38E-05
S05_20563724	9.77E-04
S05_20574367	9.85E-04
S05_20576620	8.82E-04

qN6.3; A2; 10 SNPs	
SNP	**p-value**
S06_12102783	7.25E-04
S06_12136452	9.42E-04
S06_12143917	9.85E-04
S06_12196024	9.75E-04
S06_12232824	9.18E-04
S06_12232843	9.48E-04
S06_12263875	9.89E-04
S06_12291797	8.56E-04
S06_12765240	9.23E-04
S06_13009769	8.05E-04

qN7.1; A1; 2 SNPs	
SNP	**p-value**
S07_1836005	7.95E-04
S07_2672533	1.76E-04

S08_4650515	3.14E-04
S08_4699109	8.50E-04
S08_4719380	1.71E-04
S08_4893202	3.18E-04
S08_4893204	2.49E-04
S08_4901936	2.17E-04
S08_5133187	1.95E-04
S08_5407414	7.12E-04
S08_5429929	7.44E-04
S08_5599220	9.25E-04
S08_5769073	7.78E-04
S08_5790018	7.97E-04
S08_5790607	9.80E-04
S08_6183424	4.50E-04
S08_6186697	5.00E-04
S08_6186702	6.74E-04
S08_6218378	7.45E-04
S08_6247068	6.12E-04
S08_6248063	6.65E-04
S08_6273423	5.79E-04
S08_6287036	9.75E-04
S08_7150474	3.90E-04
S08_7225932	7.43E-04
S08_7278126	6.96E-04
S08_7278464	6.02E-04
S08_7321825	5.16E-04
S08_7333135	9.13E-04
S08_7450520	6.54E-04
S08_7451234	3.71E-04
S08_7452053	3.95E-04
S08_7452062	3.75E-04
S08_7461396	9.09E-04
S08_7482165	3.33E-04
S08_7512886	7.86E-04
S08_7567091	7.39E-04
S08_7591279	5.51E-04
S09_16515318	3.19E-04

qN10.1b; A1; 1 SNPs	
SNP	**p-value**
S10_17673046	2.69E-04

qN11.1; A2; 2 SNPs	
SNP	**p-value**
S11_3584189	9.96E-04
S11_3874320	9.46E-04

Supplementry Table 3. List of SNPs included in each QTL associated with **NUE N100.** Each QTL is labelled with its name, the field trial (assay) in which it was detected, and the total number of associated SNPs. SNPs are denoted by chromosome and physical position (chr_pos format), along with the p-value obtained from the GWAS. The SNP with the lowest p-value within each QTL region is highlighted as the lead SNP.

qN6.1; A2; 11 SNPs	
SNP	**p-value**
S06_688641	9.87E-04
S06_698307	9.27E-04
S06_698314	8.96E-04
S06_737151	6.11E-04
S06_758598	6.85E-04
S06_758937	9.87E-04
S06_763524	7.59E-04
S06_775360	9.61E-04
S06_775382	8.48E-04
S06_775478	8.73E-04
S06_786784	8.51E-04

qN6.4; A1; 1 SNP	
SNP	**p-value**
S06_23190901	3.30E-04

qN7.2; A2; 1 SNP	
SNP	**p-value**
S07_19613498	8.37E-04

qN8.1; A1; 2 SNPs	
SNP	**p-value**
S08_5794542	3.90E-04
S08_5790169	6.13E-04

qN9.1b; A2; 1 SNP	
SNP	**p-value**
S09_15655372	7.71E-04

qN12.2; A2 and A3; 353 SNPs			
SNP	**p-value**	**SNP**	**p-value**
S12_24896521	8.20E-04	S12_26495354	8.83E-04
S12_24913573	8.97E-04	S12_26503466	9.81E-04
S12_25154542	9.30E-04	S12_26506314	8.68E-04
S12_25720719	9.85E-04	S12_26550666	8.02E-04
S12_25742608	8.74E-04	S12_26562047	6.12E-04
S12_25742624	7.72E-04	S12_26567938	9.90E-04
S12_25749552	9.63E-04	S12_26567943	9.90E-04
S12_25757104	7.61E-04	S12_26573614	7.30E-04
S12_25771769	4.49E-04	S12_26573701	7.35E-04
S12_25775617	5.68E-04	S12_26603902	7.91E-04
S12_25777875	5.77E-04	S12_26616726	9.05E-04
S12_25777943	7.60E-04	S12_26622931	9.61E-04
S12_25780208	7.49E-04	S12_26625248	6.18E-04
S12_25780221	8.59E-04	S12_26632913	9.10E-04
S12_25788685	7.00E-04	S12_26632933	8.10E-04
S12_25811914	9.26E-04	S12_26633208	8.72E-04
S12_25837531	5.67E-04	S12_26633209	9.56E-04
S12_25841545	7.77E-04	S12_26641121	9.58E-04
S12_25842834	9.31E-04	S12_26644105	9.59E-04
S12_25842836	9.31E-04	S12_26644425	7.34E-04
S12_25842852	8.52E-04	S12_26647149	9.97E-04
S12_25842853	8.52E-04	S12_26660566	6.85E-04
S12_25869968	9.82E-04	S12_26660571	8.37E-04
S12_25881176	9.55E-04	S12_26660686	9.77E-04
S12_25885756	7.26E-04	S12_26664975	8.99E-04
S12_25886468	7.36E-04	S12_26666671	9.39E-04
S12_25886993	8.14E-04	S12_26671773	5.65E-04
S12_25886997	9.11E-04	S12_26671786	7.13E-04
S12_25888266	8.36E-04	S12_26671842	9.20E-04
S12_25888574	9.50E-04	S12_26671858	7.31E-04
S12_25898944	5.82E-04	S12_26671910	9.73E-04
S12_25899016	9.34E-04	S12_26675037	9.94E-04
S12_25913186	7.57E-04	S12_26675042	9.97E-04
S12_25913413	9.90E-04	S12_26678736	6.98E-04
S12_25915760	8.69E-04	S12_26679275	9.65E-04
S12_25916160	6.77E-04	S12_26679409	8.37E-04
S12_25918156	9.73E-04	S12_26684274	8.59E-04
S12_25924656	8.68E-04	S12_26686543	7.69E-04
S12_25928651	9.64E-04	S12_26692035	8.11E-04
S12_25928774	8.11E-04	S12_26697108	2.94E-04
S12_25932794	8.18E-04	S12_26699702	6.47E-04
S12_25933173	9.96E-04	S12_26705827	8.19E-04
S12_25934999	7.09E-04	S12_26708382	9.27E-04
S12_25935018	6.85E-04	S12_26715736	7.84E-04
S12_25951458	9.71E-04	S12_26719742	8.60E-04
S12_25958092	8.46E-04	S12_26751119	9.84E-04
S12_25958304	9.42E-04	S12_26762561	6.94E-04
S12_25958328	9.60E-04	S12_26762588	9.05E-04
S12_25971478	9.35E-04	S12_26762750	7.79E-04
S12_25974344	9.77E-04	S12_26762753	8.97E-04

S12_25974407	9.80E-04	S12_26762770	9.33E-04
S12_25985496	9.65E-04	S12_26762801	7.49E-04
S12_25988171	9.29E-04	S12_26762830	8.54E-04
S12_25988809	9.11E-04	S12_26766583	6.08E-04
S12_25989264	8.97E-04	S12_26787859	9.52E-04
S12_25989531	7.56E-04	S12_26787949	7.35E-04
S12_25992337	6.07E-04	S12_26788883	9.04E-04
S12_25992339	6.07E-04	S12_26789602	8.99E-04
S12_25992341	6.18E-04	S12_26789653	7.91E-04
S12_25992451	8.60E-04	S12_26790283	5.59E-04
S12_25994366	8.50E-04	S12_26791159	9.18E-04
S12_25996649	7.85E-04	S12_26792480	9.74E-04
S12_25997909	9.00E-04	S12_26795089	4.89E-04
S12_26001088	9.88E-04	S12_26798771	6.31E-04
S12_26009814	7.78E-04	S12_26875690	9.16E-04
S12_26010100	7.34E-04	S12_26875691	9.01E-04
S12_26010164	7.88E-04	S12_26876313	9.20E-04
S12_26010176	8.71E-04	S12_26880178	9.39E-04
S12_26010322	9.60E-04	S12_26880570	5.06E-04
S12_26010325	9.41E-04	S12_26880629	7.38E-04
S12_26010381	9.39E-04	S12_26880988	8.20E-04
S12_26012751	8.49E-04	S12_26891968	9.74E-04
S12_26013641	9.25E-04	S12_26892040	8.86E-04
S12_26013648	7.80E-04	S12_26893633	7.50E-04
S12_26013704	9.34E-04	S12_26901965	6.82E-04
S12_26014519	9.72E-04	S12_26906382	8.56E-04
S12_26014886	5.96E-04	S12_26906417	8.66E-04
S12_26014892	6.09E-04	S12_26906855	7.73E-04
S12_26015049	9.76E-04	S12_26907156	8.72E-04
S12_26015064	9.64E-04	S12_26907535	9.18E-04
S12_26019055	6.61E-04	S12_26910649	9.83E-04
S12_26021348	7.87E-04	S12_26911765	9.19E-04
S12_26024868	7.17E-04	S12_26911824	6.82E-04
S12_26027776	7.64E-04	S12_26911865	8.99E-04
S12_26028095	8.57E-04	S12_26952856	8.73E-04
S12_26029911	7.53E-04	S12_26952859	9.19E-04
S12_26029931	7.37E-04	S12_26952908	9.79E-04
S12_26041109	8.87E-04	S12_26952937	9.69E-04
S12_26043248	8.49E-04	S12_26960366	8.94E-04
S12_26054270	8.59E-04	S12_26960393	8.30E-04
S12_26054276	7.79E-04	S12_26979888	4.96E-04
S12_26057770	9.58E-04	S12_27006278	2.62E-04
S12_26064319	6.76E-04	S12_27012996	7.25E-04
S12_26066483	9.77E-04	S12_27013066	8.41E-04
S12_26066491	9.62E-04	S12_27038428	9.76E-04
S12_26066495	9.18E-04	S12_27038515	8.00E-04
S12_26067154	9.29E-04	S12_27038549	7.56E-04
S12_26067343	9.47E-04	S12_27045621	8.88E-04
S12_26068389	9.61E-04	S12_27062802	7.19E-04
S12_26071311	9.89E-04	S12_27064556	8.39E-04
S12_26074820	4.34E-04	S12_27081814	9.50E-04
S12_26074877	6.77E-04	S12_27081825	9.05E-04

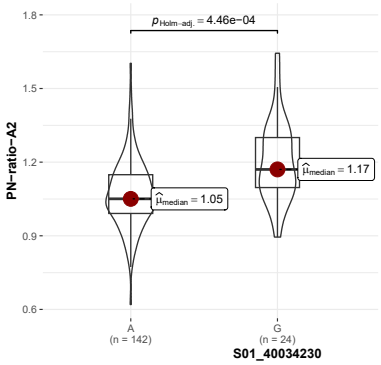
S12_26077590	9.38E-04	S12_27092603	7.95E-04
S12_26077909	8.99E-04	S12_27094738	8.67E-04
S12_26080106	9.71E-04	S12_27097566	6.51E-04
S12_26098014	9.12E-04	S12_27097583	8.07E-04
S12_26104526	7.94E-04	S12_27097795	9.97E-04
S12_26141922	7.89E-04	S12_27098514	9.94E-04
S12_26147502	7.46E-04	S12_27099337	9.55E-04
S12_26147884	9.49E-04	S12_27099352	9.73E-04
S12_26150815	8.54E-04	S12_27112459	8.78E-04
S12_26151037	6.68E-04	S12_27121897	8.34E-04
S12_26159543	8.82E-04	S12_27127107	6.62E-04
S12_26162223	9.12E-04	S12_27131893	9.76E-04
S12_26162224	9.18E-04	S12_27136424	8.70E-04
S12_26162325	9.27E-04	S12_27174994	7.81E-04
S12_26162698	9.04E-04	S12_27194705	7.55E-04
S12_26163856	7.49E-04	S12_27195402	7.78E-04
S12_26163914	8.44E-04	S12_27196704	9.72E-04
S12_26164244	7.22E-04	S12_27198279	8.76E-04
S12_26164268	5.42E-04	S12_27204000	8.48E-04
S12_26164299	7.93E-04	S12_27206180	8.55E-04
S12_26167936	8.69E-04	S12_27206193	9.09E-04
S12_26172367	8.95E-04	S12_27239913	9.45E-04
S12_26173831	8.71E-04	S12_27243046	8.67E-04
S12_26174991	8.57E-04	S12_27246847	6.36E-04
S12_26179655	8.97E-04	S12_27262689	7.63E-04
S12_26181054	9.28E-04	S12_27262744	8.90E-04
S12_26182350	9.18E-04	S12_27280650	7.93E-04
S12_26183066	9.53E-04	S12_27303403	8.75E-04
S12_26183187	9.10E-04	S12_27305478	9.88E-04
S12_26183620	9.37E-04	S12_27305498	5.49E-04
S12_26183690	9.38E-04	S12_27318279	8.80E-04
S12_26183874	9.74E-04	S12_27323070	9.65E-04
S12_26198042	5.63E-04	S12_27328894	9.46E-04
S12_26198931	8.18E-04	S12_27332223	9.44E-04
S12_26207325	9.50E-04	S12_27333792	6.74E-04
S12_26207485	7.92E-04	S12_27348453	9.97E-04
S12_26207790	1.65E-04	S12_27348474	9.58E-04
S12_26208956	6.53E-04	S12_27351035	6.62E-04
S12_26209110	7.40E-04	S12_27351036	6.79E-04
S12_26209887	5.33E-04	S12_27358631	7.68E-04
S12_26209970	6.73E-04	S12_27365872	9.10E-04
S12_26230924	7.77E-04	S12_27367223	9.71E-04
S12_26230952	6.19E-04	S12_27367251	8.07E-04
S12_26235733	5.15E-04	S12_27372737	3.35E-04
S12_26240122	9.28E-04	S12_27372821	2.85E-04
S12_26241229	8.02E-04	S12_27374100	8.38E-04
S12_26241818	9.73E-04	S12_27374320	8.72E-04
S12_26242543	9.19E-04	S12_27379646	9.60E-04
S12_26243328	9.32E-04	S12_27398003	9.46E-04
S12_26244356	9.76E-04	S12_27405716	6.92E-04
S12_26244896	9.62E-04	S12_27406839	6.69E-04
S12_26248406	4.63E-04	S12_27411159	8.48E-04

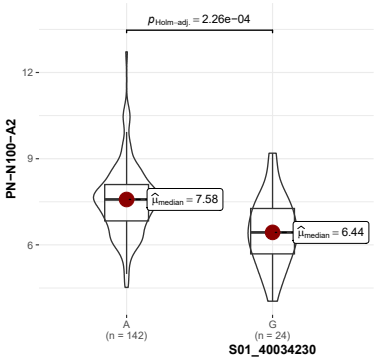
S12_26256252	7.76E-04	S12_27411255	9.13E-04
S12_26260657	9.11E-04	S12_27411461	9.82E-04
S12_26449925	9.22E-04	S12_27414764	8.84E-04
S12_26455984	8.79E-04	S12_27414800	9.91E-04
S12_26462785	8.83E-04	S12_27416800	9.94E-04
S12_26462790	8.01E-04	S12_27416820	7.01E-04
S12_26462813	8.63E-04	S12_27424727	8.16E-04
S12_26462955	7.15E-04	S12_27428875	9.72E-04
S12_26467166	8.97E-04	S12_27439457	9.64E-04
S12_26468972	9.46E-04	S12_27440508	9.08E-04
S12_26470282	8.56E-04	S12_27440621	9.38E-04
S12_26470649	9.94E-04	S12_27448764	9.65E-04
S12_26477118	9.14E-04	S12_27450459	9.57E-04
S12_26477190	8.96E-04	S12_27459879	8.93E-04
S12_26477196	5.50E-04	S12_27459944	7.59E-04
S12_26477206	9.04E-04	S12_27472385	7.51E-04
S12_26485377	8.77E-04	S12_27473223	7.70E-04
S12_26485930	9.27E-04	S12_27474085	1.92E-04
S12_26488470	8.19E-04	S12_27477837	8.16E-04
S12_26488511	9.16E-04	S12_27495273	9.60E-04
S12_26490134	8.53E-04	S12_27509489	9.43E-04
S12_26495244	9.60E-04	S12_27509569	9.96E-04
S12_26495353	8.83E-04		

Supplementry Table 4. List of SNPs included in each QTL associated with **PN N100.** Each QTL is labelled with its name, the field trial (assay) in which it was detected, and the total number of associated SNPs. SNPs are denoted by chromosome and physical position (chr_pos format), along with the p-value obtained from the GWAS. The SNP with the lowest p-value within each QTL region is highlighted as the lead SNP.

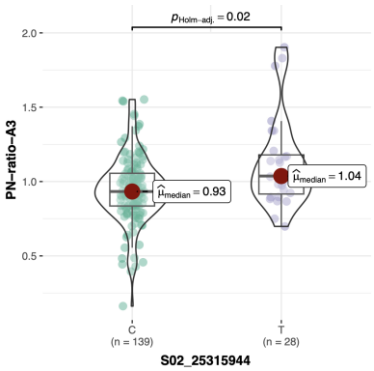
Figures

A) qN1.4

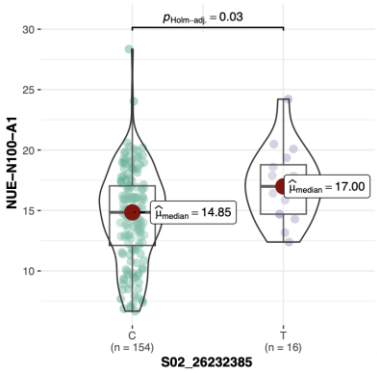




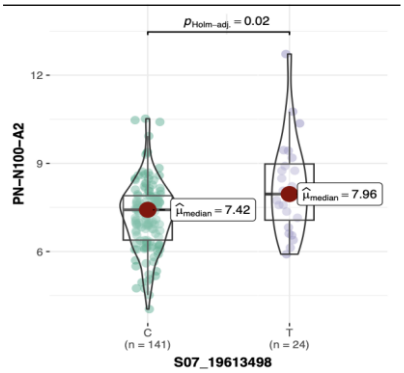
B) qN2.3



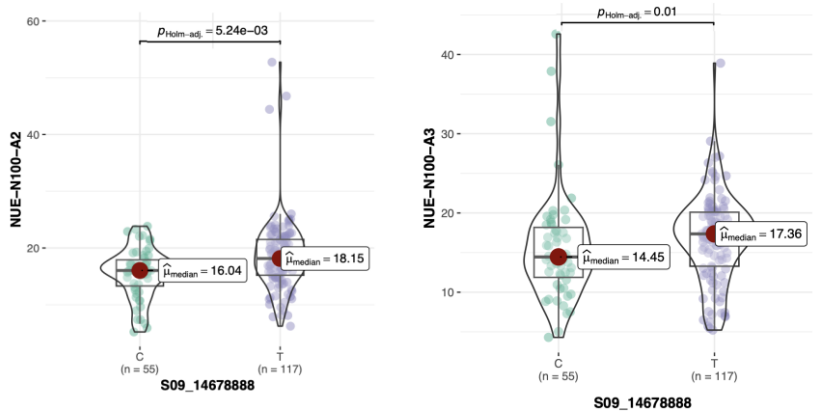
C) qN2.3



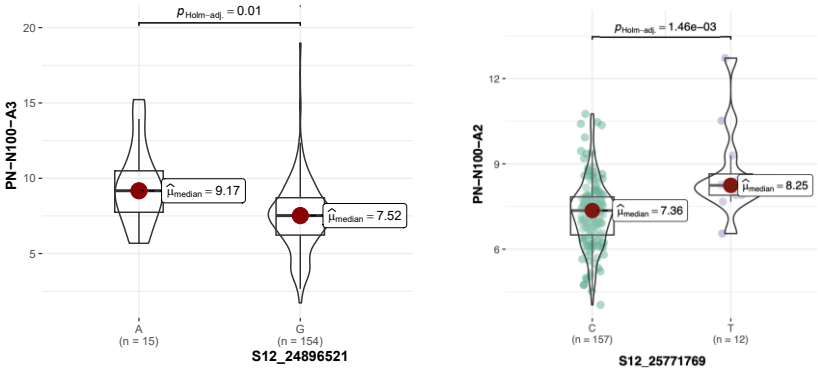
D) qN7.2



E) qN9.1b



F) qN12.2



Supplementary Figure1. Violin plots showing the distribution of trait values (NUE and PN) according to allelic variation at lead SNPs for selected QTLs.

General Discussion

The publication of the first draft rice genome, for the *indica* cultivar 93-11 and *japonica* Nipponbare in 2002, followed by the improved and high-quality reference genome of Nipponbare in 2005, laid the foundation of modern rice genomics. This initial step was significantly expanded with the advent of next-generation sequencing (NGS) platforms, which enabled the large-scale sequencing of the entire genome and facilitated population-scale initiatives, such as the 3,000 Rice Genomes Project (Li et al., 2014). More recently, the construction of platinum-standard pan-genomes (Zhou et al., 2020; Liu et al. (2022); Nature, 2025) has captured the genetic variations across cultivated and wild rice pools (Zhou et al., 2020; Liu et al., 2022; Dongling Guo 2025). In parallel, the conservation and curation of germplasm banks have been improved considerable in the last years and, today, more than 780,000 cultivated and wild rice accessions are safeguarded, including more than 130,000 in the International Rice Genebank (IRRI). Together, these deep sequence resources and vast germplasm repositories now enable an efficient and directed breeding.

Since the genetic improvement of rice is carried out locally due to its susceptibility to the agro-climatic conditions of each area, translating all of the knowledge and resources into cultivars adapted to specific local agro-climatic conditions, particularly temperate environments, such as in Spain, remains limited. Also, the temperate *japonica* germplasm available adapted to these areas tends to show narrow genetic variation, thereby restricting breeding potential. In this context, generating local, structured genetic resources, both in the form of varietal panels and experimental populations, is essential. Such resources enable the study of a wide range of agronomic traits and provide a foundation for genomics-assisted breeding adapted to local challenges.

Accordingly, this thesis focused its efforts on the development, phenotypic characterization, and genomic sequencing of two key genetic resources: (1) A diverse collection of temperate *japonica* rice varieties used in Chapters I and II, and (2) A Multi-parent Advanced Generation Inter-Cross (MAGIC) population, studied in Chapters III and IV.

Each resource offers distinct advantages. MAGIC design allows the simultaneous analysis of multiple traits, the exploration of pleiotropic effects, and the dissection of complex features that involve multiple contributing factors. Moreover, the

extensive recombination and allelic diversity captured through many intercrosses make MAGIC populations valuable not only for QTL discovery but also as a source of novel breeding material, yielding elite lines that meet both agronomic and market demands.

On the other hand, varietal panels offer access to a broad phenotypic and genetic spectrum in temperate *japonica* diversity. These panels are particularly well-suited to the study of underexplored or emerging traits, that may not be easily detectable in conventional breeding programs.

Among the needs that arise at the local level, adaptation and improvement of existing rice varieties to meet emerging climate-related challenges, particularly in sensitive environments, is becoming increasingly urgent. This is the case in Spain, where a significant portion of rice production takes place within a protected area with strict environmental regulations, where the use of chemical inputs, as pesticides and synthetic fertilizers, must be drastically reduced. As a result, traits such as allelopathy have gained interest as a potential strategy for sustainable weed management. Similarly, the excessive application of nitrogen fertilizers is a major source of environmental pollution, particularly of soil and water systems. Developing rice varieties capable of maintaining high productivity under low-nitrogen input conditions is therefore a priority.

For this reason, the genetic resources developed in this thesis were used to investigate basic agronomic traits (Chapter III) and emerging traits that are strategically relevant for sustainable agriculture, the allelopathic potential (Chapter I) and NUE (Chapter IV). The ultimate goal was to genetically dissect these traits and enable their application in varietal improvement through marker-assisted selection.

Chapters I and II employed a panel of 171 temperate *japonica* rice varieties to gain deeper insight into the genetic basis of allelopathy, adopting a holistic perspective by integrating both genomic and transcriptomic analyses. On the other hand, Chapters III and IV made use of a MAGIC population composed of 192 advanced inbred lines derived from eight temperate *japonica* founders, to study multiple agronomic traits and NUE at field conditions.

In the case of the varietal temperate *japonica* panel, the phenotypic evaluation for allelopathic potential against *E. crus-galli* revealed varieties such as Ganao, Katy and Senia with strong inhibiting potential of barnyardgrass root growth (aprox 50%), confirming the presence of allelopathic potential in elite *japonica* varieties beyond the widely described PI312777 line.

This panel was partially sequenced in a previous study by our group (Reig-Valiente et al.). But, for this work, was upgraded to full resequencing, yielding over 124,000 SNPs, greatly enhancing its resolution for downstream applications. As a result, four QTLs significantly associated with allelopathic potential against *E. crus-galli* were identified in Chapter I. The strongest signals on chromosomes 2 and 5 co-localized with a cluster of four genes encoding PALs, two genes corresponding to R2R3-type MYB transcription factors, a gene encoding a ferroportin transporter and several lipid-acyl-transferase genes. These loci extend the allelopathic repertoire beyond the canonical momilactone biosynthesis genes from the diterpenoid pathway, and suggest that phenylpropanoid derivatives may be decisive as allelochemicals in *japonica* backgrounds.

Building upon these findings, we compared the root transcriptomes of the most suppressive variety found in the study, Ganao, and the non-allelopathic control Bomba during co-culture with *E. crus-galli* in Chapter II. RNA-seq analyses at 0, 3, and 36 hours revealed an early and rapid transcriptional activation, progressively increasing over time and culminating in the full activation of the phenylpropanoid biosynthesis pathway in Ganao. While diterpenoid phytoalexin genes, previously assumed to be central in other studies, were down-regulated. The exception was the oryzalexin S which emerged as the only diterpenoid candidate. Also, the two *R2R3 MYBs*, *OsMYB30* and *OsMYB52*, were upregulated. The convergence of GWAS and RNA-seq, therefore, points to a phenolic strategy of weed suppression, supported by detoxification enzymes (cytochrome P450s, GSTs, UGTs) and potential wall remodeling through lignin reinforcement.

Given the limited prior knowledge on allelopathy addressed in Chapters I and II, significant progress has been made in identifying both rice varieties with allelopathic potential and the underlying genetic mechanisms involved. These positive findings encourage future work efforts, particularly towards the identification of key phenolic compounds using metabolomics, the assessment of oryzalexin S bioactivity, and the

introgression of Ganao alleles into elite backgrounds through marker-assisted backcrossing, with subsequent evaluation under field conditions.

Chapters III and IV pivoted to the construction and use of a MAGIC population composed of 192 advanced inbred lines derived from eight elite temperate *japonica* founders. This population was genotyped by whole-genome sequencing yielding 70,289 high-confidence SNPs uniformly distributed across the genome. Phenotyping was carried out for multiple agronomic traits grown in pools and under field conditions. This resource enabled us to study not only classical agronomic traits such as plant architecture and phenology, key components in the development of new rice varieties, but also complex traits influenced by multiple factors, such as NUE. Remarkably, the simultaneous evaluation of all these traits within the same population also provides a valuable opportunity to detect potential pleiotropic effects, offering deeper insights into the genetic architecture underlying trait correlations.

The traits evaluated in plants grown in open-air pools (plant height, panicle length, awning, lemma pubescence, flag leaf and tiller angle, heading date and grain filling period) were assessed over three growing seasons (Chapter III). No significant differences were observed between years, indicating a strong genetic basis with minimal environmental influence. Moreover, most of these traits exhibited high heritability, facilitating the genetic characterisation of their underlying determinants. This was not the case, however, for NUE, which was evaluated under field conditions in Chapter IV. In this case, the environment had a strong impact on the measured parameters, with significant differences observed between years. Nevertheless, the use of an augmented block design enabled us to partially account for annual environmental effects.

In the association analysis of the traits evaluated in plants grown in pools, 35 QTLs with strong associations were identified, many of which corresponded to previously characterised genomic loci, as several candidate genes located within these regions had already been reported, including *SD1, SDF5, RAE2, GL1, TAC1, EHD4, Ghd8, OsGW10* and *OsGS10*. In addition, we detected 7 novel QTLs of interest (qAwn12, qGlb7, qFLA12, qHd6, qHd7, qGR1, qGR6), which may represent adaptation-specific loci relevant under our local climatic conditions. Furthermore, a pleiotropic hotspot was detected on chromosome 9 (qPL9, qH9). This overlap points to a single

regulatory block, harbouring *DEP1* and *LP1*. The identification of strong and consistent associations validated the effectiveness of the MAGIC population for genetic dissection and yielded a valuable set of markers for marker-assisted selection in breeding programs.

Chapter IV expanded on this resource to investigate the complex trait of NUE under field conditions with contrasting nitrogen doses (N100 and N200). A total of 22 QTLs were associated with yield, particularly PN and NUE ratios at low nitrogen levels. Some QTLs were consistently detected across different assays, such as qN5.2 and qN12.1, identified in NUE ratio and PN analyses, respectively. This is also the case of qN9.1b and qN12.2, detected under N100 condition. The most promising loci, displaying the lowest p-values, were located on chromosomes 5, 6, 8, and 9. In total, 29 candidate genes were identified across all detected QTLs, including genes involved in nitrogen uptake (*OsNRT1.1B*), assimilation (*OsGS1;2*, *OsAAP1*), signaling and regulation (*OsNLP4*, *OsWRKY23*), and traits related to tillering and panicle development (*ERF142/NGR5*, *OsGATA8*). Notably, the response of the MAGIC lines varied widely across nitrogen regimes, revealing genotypes with superior performance under low-input conditions.

A logical extension of this work, based on the results obtained in Chapter III and IV, would be the development of a minimal marker set targeting the most reliable QTLs with unambiguous candidate genes associated with core architectural traits for early-generation screening. Implementing such a small and robust marker panel would enable breeders to discard lines lacking the desired baseline morphology before allocating resources to evaluate more complex and resource-intensive traits, such as NUE. In this way, the secure 'first filter' provided by high-confidence markers could enhance the efficiency of the breeding pipeline, ensuring that selection for multifactorial traits builds upon a genetically and agronomically solid foundation.

In sum, the results of this thesis not only provide molecular markers and candidate genes for use in future breeding efforts, but also lay the groundwork for new research directions. The integration of multi-omic data, as WGS and RNA-seq, multi-season phenotyping, and diverse population designs, exemplifies a holistic strategy to address both classical and emerging targets for rice improvement. This integrated approach not only accelerates the translation of genomic knowledge into elite

varieties but also sets the stage for future research that couples functional validation, metabolomics and multi-environment trials to deliver truly sustainable genetic gain.

Conclusions

This doctoral thesis highlights the value and potential of developing and using advanced genetic resources and multi-omic tools to dissect complex agronomic traits in temperate *japonica* rice. The main conclusions derived from this thesis are as follows:

1.- The panel of 171 temperate *japonica* rice accessions represents the genetic and phenotypic diversity of cultivated varieties in temperate regions, and displayed substantial natural variation in allelopathic potential, indicating the genetic basis of allelophaty in rice.

2.- Ganao, Katy, and Senia are varieties with allelopathic potential, that can inhibit more than 50% of *E crus-galli* root growth under controlled conditions.

3.- GWAS proved to be a powerful tool for dissecting allelopathy, a complex trait, enabling the identification of four associated QTLs. Two of these QTLs, located on chromosomes 2 and 5, colocalized with strong candidate genes encoding PAL, R2R3-MYB transcription factors, ferroportin, and lipid-acyl-transferases. This study represents the first application of GWAS to investigate allelopathy in rice.

4.- The root transcriptional profiles of the allelopathic Ganao and non-allelopathic Bomba varieties showed distinct differences at both early and late stages of interaction with *E. crus-galli*, indicating divergent responses.

5.- The interaction between *E. crus-galli* and Ganao leads to the upregulation in Ganao of key genes in the phenylpropanoid biosynthetic pathway. However, the synthesis of diterpenes, including momilactones, is not induced. This suggests that phenolic compounds play a central role in the observed allelopathic phenomenon.

6.- The convergence of GWAS and RNA-seq results supports a phenolic-based weed suppression strategy, potentially involving detoxification enzymes (P450s, GSTs, UGTs) and lignin pathway activation for cell wall reinforcement.

7.- The characterized MAGIC population exhibited high phenotypic diversity among its 192 lines, enabling the mapping of QTLs associated with plant architecture, phenology, and NUE.

8.- Traits such as plant height, heading date, awning, lemma pubescence and panicle length showed high heritability with a strong genetic basis, allowing the identification of 35 stable QTLs and potential candidate genes, including *SD1, SDF5, RAE2, GL1, TAC1, EHD4, Ghd8, OsGW10*, and *OsGS10*. Some loci, such as qPL9 and qH9, showed pleiotropic effects.

9.- A total of 22 QTLs and 29 candidate genes associated with NUE or tillering were identified through a GWAS using the MAGIC population. These candidate genes are involved in key biological processes such as nitrogen uptake, assimilation, signaling, and regulation, as well as related developmental pathways including tillering and panicle formation. These findings provide valuable insights into the genetic basis of NUE and tillering and offer potential targets for improving yield under low nitrogen inputs in rice breeding programs.