Institut für Nutzpflanzenwissenschaften und Ressourcenschutz (INRES)

Development of the *BonnMu* transposon resource in maize (*Zea mays* L.) and its application in characterizing the *magenta root dwarf 1* (*mrd1*) mutant

Dissertation

zur Erlangung des Grades

Doktor der Agrarwissenschaften (Dr. agr.)

der Landwirtschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

von

Yan Naing Win

aus

Nanmar, Kachin State, Myanmar

Bonn, 2024

Referentin: Priv.-Doz.in Dr. Caroline Marcon

Korreferent: Prof. Dr. Heiko Schoof

Tag der mündlichen Prüfung: 02.07.2024

Angefertigt mit Genehmigung der Landwirtschaftlichen Fakultät der Universität Bonn

Table of Contents

List of figures

List of tables

List of supplemental figures

List of supplemental tables

List of abbreviations

Ac ..... *Activator* element
An ..... Anthocyanin
ATAC-seq ..... Assay for Transposase-Accessible Chromatin using sequencing
AtCUl1 ..... *Arabidopsis thaliana* CULLIN 1
AUX ..... Auxin
AXR1 ..... AUXIN RESISTANT 1
BC ..... Before Christ
BiFC ..... Bimolecular Fluorescence Complementation
bp ..... base pair
BP reaction ..... Bacterial Phage (BP) recombination reaction
BSA ..... Bulked Segregant Analysis
BSR-seq ..... Bulked Segregant RNA Sequencing
Cas9 ..... CRISPR-associated protein 9
cDNA ..... complementary DNA
ChIP-seq ..... Chromatin Immunoprecipitation sequencing
chr ..... chromosome
Co ..... Cortex
cop/*det*/*fus* ..... *constitutive photomorphogenesis*/*de-etiolated*/*fusca*
COP9 ..... Constitutive Photomorphogenesis 9
CRISPR ..... Clustered Regularly Interspaced Short Palindromic Repeats
CRL ..... CULLIN-RING E3 LIGASE
CSN ..... Constitutive Photomorphogenesis 9 (COP9) SIGNALOSOME
csn4 ..... *constitutive photomorphogenesis 9* (*cop9*) *signalosome complex subunit 4*
CyG ..... Cyanidin 3-O-Glucoside
Dag ..... day after germination
DEG ..... Differentially Expressed Gene
DNA ..... Deoxyribonucleic Acid
Ds ..... *Dissociation* element
E1, 2 ..... ubiquitin-activating enzymes
ECR1 ..... E1 C-TERMINAL RELATED 1
Ed ..... Endodermis
Ep ..... Epidermis
FAO ..... The Food and Agriculture Organization of the United Nations
FDR ..... False Discovery Rate
FGS ..... Filtered Gene Set

FRET ..... Fluorescence Resonance Energy Transfer
GO ..... Gene Ontology
H ..... Histone
hb3 ..... *homeobox-transcription factor 3*
HMGA ..... Chromatin-Associated High Mobility Group Protein A
HPLC ..... High-Performance Liquid Chromatography
IAA ..... Indole-3-Acetic Acid
INDEL ..... Insertion/Deletion
ITS1 ..... Internal Transcribed Spacer 1
LR reaction ..... Lambda Recombinase (LR) recombination reaction
LTR ..... Long Terminal Repeats
lux ..... unit of illumination
Lys ..... Lysine
maizeGDB ..... Maize Genetics and Genomics Database
Mb ..... Mega base
MDS ..... Multidimensional Scaling
MeOH ..... Methanol
mrd1 ..... *magenta root dwarf 1*
MTM ..... Maize-Targeted Mutagenesis population
MUSCLE ..... MUltiple Sequence Comparison by Log-Expectation
Mu-seq ..... Mutant-seq
mut ..... mutant
MuWU ..... Mutant-seq Workflow Utility
Mu ..... *Mutator*
Mx ..... Metaxylem
NAM ..... Nested Association Mapping
NEDD8 ..... Neural Precursor Cell Expressed Developmentally Downregulated Protein 8
NGS ..... Next Generation Sequencing
OTUs ..... Operational Taxonomic Units
PAM ..... Protospacer Adjacent Motif
PBS ..... Phosphate Buffered Saline
PCA ..... Principal Component Analysis
PCI ..... Proteasome, COP9, Initiation factor 3 domain
PCR ..... Polymerase Chain Reaction
PeoG ..... Peonidin 3-O-Glucoside
Ph ..... Phloem
promoterCore ..... Core promoter

promoterProx ..... Proximal promoter
qRT-PCR ..... quantitative real-time reverse-transcription PCR
RCE1 ..... RUB1 Conjugating Enzyme 1
RFP ..... Red Fluorescent Proteins
RNA ..... Ribonucleic Acid
RNA-seq ..... Ribonucleic Acid Sequencing
rRNA ..... ribosomal Ribonucleic Acid
rtcl ..... *rootless concerning crown and seminal roots like*
rtcs ..... *rootless concerning crown and seminal root*
rth1 ..... *root hair defective 1*
rth3 ..... *root hair defective 3*
rth5 ..... *root hair defective 5*
rth6 ..... *root hair defective 6*
RUB1 ..... Related to Ubiquitin 1
rum1 ..... *rootless with undetectable meristem 1*
rum1-like1; *rul1* ..... *rootless with undetectable meristem like 1*
SCF ..... SKP1 (S-phase kinase-associated protein 1), CUL1 (Cullin 1), F-box proteins
sgRNA ..... single guide RNA
Skp1 ..... S-phase kinase-associated protein 1
SNP ..... Single Nucleotide Polymorphism
T-DNA ..... Transfer-DNA
TE ..... Transposable Element
TIDE ..... Tracking of Indels by Decomposition
TIR ..... Terminal Inverted Repeat
TIR1 ..... Transport Inhibitor Response 1
TSD ..... Target Site Duplication
TUSC ..... Trait Utility System for Corn
U adapter ..... Universal adapter
UMRs ..... Unmethylated regions
UTRs ..... Untranslated regions
WT ..... Wild type
Xy ..... Xylem

1 Zusammenfassung/ Summary

1.1 Zusammenfassung

Die genomweite Insertionsmutagenese ist eine Methode zur Erzeugung von Funktionsverlustmutationen für praktisch alle Gene in einem Genom. Bei diesem Verfahren werden Mutationen auf natürliche Weise durch zufällige Integration mobiler DNA-Elemente in das Genom erzeugt. Die *BonnMu* Ressource ist eine europäische Transposon-markierte Mutantensammlung, die für funktionelle Genomikstudien in Mais entwickelt wurde. In der vorliegenden Arbeit wurde die Mutantensammlung durch Kreuzung einer aktiven *Mutator* (*Mu*)-Linie mit nordamerikanischen Dent (B73, Co125) und europäischen Flint (DK105, EP1 und F7) Inzuchtlinien erweitert. Dadurch wurden insgesamt 8064 mutagenisierte *BonnMu* F_2-Familien erzeugt. Durch Sequenzierung dieser *Mu*-getaggten Familien wurden 425 924 vererbbare *Mu*-Insertionen identifiziert, die 36 612 (83 %) der 44 303 Genmodelle von Mais betreffen. Im Durchschnitt wurden 12 *Mu*-Insertionen pro Gen (425 924 Gesamtinsertionen/ 36 612 betroffene Gene) und 53 Insertionen pro *BonnMu* F_2-Familie (425 924 Gesamtinsertionen/ 8064 Familien) detektiert. Alle *Mu*-Insertionen und Fotos von Keimlingsphänotypen der segregierenden *BonnMu* F_2-Familien sind über die Maize Genetik und Genomik Datenbank (MaizeGDB) zugänglich. Die anschließende Untersuchung der Keimbahninsertionen erfolgte mit Hilfe des automatisierten Mutant-seq Workflow-Tools (MuWU). Insgesamt befanden sich 94 % aller Keimbahninsertionen in kodierenden Abschnitten des Genoms und nur ein kleiner Anteil von 6 % in nicht kodierenden, intergenischen Regionen. Damit übereinstimmend waren die *Mu*-Insertionen vor allem in den genreichen Chromosomenarmen zu finden. Insgesamt befanden sich 42 % aller *BonnMu*-Insertionen in der 5'-untranslatierten Region (UTR) von Genen, und damit im zugänglichen Chromatin. Darüber hinaus konnten für 38 % der Insertionen (163 843 von 425 924 Gesamtinsertionen) die ursächlichen *Mu*-Elemente *Mu1*, *Mu8* und *MuDR* bestätigt werden. Zusammenfassend archiviert die öffentlich zugängliche europäische *BonnMu* Ressource Insertionen für Dent- und Flintlinien und ermöglicht dadurch sowohl vorwärts- als auch rückwärtsgenetische Studien.

Im zweiten Projekt der vorliegenden Arbeit wurde die Mutante *magenta root dwarf 1* (*mrd1*), die in einer vorwärtsgenetischen Durchmusterung der *BonnMu* F_2-Familien identifiziert wurde, funktionell charakterisiert. Im Vergleich zum Wildtyp, reichert die *mrd1* Mutante Anthocyane in Primär-, Seminal-, und Kronenwurzeln an. Außerdem weisen *mrd1* Keimlinge ein reduziertes Sprosswachstum auf, unabhängig davon, ob sie unter Licht- oder Dunkelbedingungen gekeimt wurden. Die Identifizierung des ursächlichen

Kandidatengens wurde durch die Kombination einer bulked segregant RNA-seq (BSR-seq) Analyse und der zur Verfügung stehenden sequenzierten *BonnMu* F_2-Familien ermöglicht, die eine begrenzte Anzahl an *Mu*-getaggten Genen im BSR-seq Mapping-Intervall liefert. Dieser Kombinationsansatz identifizierte eine *konstitutive photomorphogenese 9 (cop9) signalosomkomplex-untereinheit 4 (csn4)* als ursächliches Kandidatengen, das dem *mrd1*-Phänotyp zugrunde liegt. Das COP9-Signalosom (CSN) ist ein Multiproteinkomplex, der ursprünglich in *Arabidopsis thaliana* als Repressor der Photomorphogenese identifiziert wurde. Die *mrd1* Mutanten weisen, beispielsweise mit ihrer konstitutiven Photomorphogenese, Ähnlichkeiten mit *csn* Mutanten in *Arabidopsis thaliana* auf, zeigen aber unterschiedliche Muster der Anthocyananreicherung. Eine PCR-basierte Co-Segregationsanalyse identifizierte *mrd1* Mutanten als homozygot für eine *Mu*-Insertion im *mrd1* Gen. Die Expression des *mrd1* Gens war in den Primärwurzeln der Mutanten im Vergleich zu heterozygoten und homozygoten Wildtyp-Pflanzen deutlich reduziert. Das Mikrobiom der Rhizosphäre von Kronenwurzeln zeigte eine Anreicherung spezifischer bakterieller Genera in *mrd1*-Mutanten und Wildtyp. Zusätzlich unterstrich die vergleichende RNA-seq-Analyse zwischen *mrd1* und Wildtyp-Kronenwurzeln eine Herunterregulierung des *mrd1*-Gens in Mutanten. Der Versuch, das Gen mit einem unabhängig mutierten Allel zu validieren, das durch Genom-Editierung mittels CRISPR/Cas9 erzeugt wurde, blieb erfolglos. In einem zukünftigen Ansatz könnte das *mrd1* Gen entweder durch eine Gen-Silencing-Strategie unter Verwendung von RNAi oder durch ein neuartiges Allel aus öffentlich zugänglichen Mutantenressourcen, wie z. B. der hauseigenen *BonnMu* Kollektion, die ständig erweitert wird, validiert werden.

1.2 Summary

Genome-wide insertional mutagenesis is a tool to generate loss-of-function mutations for virtually all genes in a genome. In this process, mutations are naturally generated by random integration of mobile DNA elements within a genome. The *BonnMu* resource is a European transposon-tagged mutant collection designed for functional genomics studies in maize. In this thesis the mutant repository was expanded by crossing an active *Mutator* (*Mu*) line with dent (B73, Co125) and flint (DK105, EP1 and F7) germplasms, resulting in the generation of 8,064 mutagenized *BonnMu* F_2-families. Sequencing of these *Mu*-tagged families identified 425,924 heritable *Mu* insertions affecting 36,612 (83%) of the 44,303 high-confidence gene models of maize. On average, 12 *Mu* insertions per gene (425,924 total insertions/ 36,612 affected genes) and 53 insertions per *BonnMu* F_2-family (425,924 total insertions/ 8,064 families) were observed. *Mu* insertions and photographs depicting seedling phenotypes from segregating *BonnMu* F_2-families are available for access through the Maize Genetics and Genomics Database (MaizeGDB). Downstream examination via the automated Mutant-seq Workflow Utility (MuWU) identified 94% of the germinal insertion sites in genic regions and only a small fraction of 6% inserting in non-coding intergenic sequences of the genome. Consistently, *Mu* insertions aligned with gene-dense chromosomal arms. In total, 42% of all *BonnMu* insertions were located in the 5' untranslated region (UTR) of genes, corresponding to accessible chromatin. Furthermore, for 38% of the insertions (163,843 of 425,924 total insertions) *Mu1*, *Mu8* and *MuDR* were confirmed to be the causal *Mu* elements. In summary, the publicly accessible European *BonnMu* resource has archived insertions covering two major germplasm groups, thus facilitating both forward and reverse genetics studies.

In the second project, the *magenta root dwarf 1* (*mrd1*) mutant, identified in a forward genetic screening of *BonnMu* F_2-families, was functionally characterized. In comparison to the wild type, *mrd1* mutants accumulate anthocyanins in primary, seminal and crown roots and exhibit a limited shoot growth under light and dark conditions. The identification of a causal candidate gene was facilitated by the combination of bulked segregant RNA-seq (BSR-seq) and the availability of the sequenced *Mu*-tagged *BonnMu* F_2-families, which provided a limited number of candidate mutations for the BSR-seq mapping interval. This combinatory approach identified a *constitutive photomorphogenesis 9* (*cop9*) *signalosome complex subunit 4* as the causal candidate gene that underlie the *mrd1* phenotype. The COP9 signalosome (CSN) is a multiprotein complex initially identified in *Arabidopsis thaliana* as a repressor of photomorphogenesis. Maize *mrd1* mutants show similarities to *csn* mutants in *Arabidopsis thaliana* such as the constitutive photomorphogenesis, but different patterns of anthocyanin accumulation. PCR-based co-segregation analysis

identified *mrd1* mutants as homozygous for a *Mu* insertion in the *mrd1* gene. The expression of the *mrd1* gene was significantly reduced in the primary roots of *mrd1* mutants, compared to heterozygous and wild type plants. Microbiome profiling of the rhizosphere from crown roots of *mrd1* and wild type indicated an enrichment of specific bacterial genera. Additionally, RNA-seq analysis between *mrd1* and wild type crown roots highlighted a down-regulation of the *mrd1* gene in the mutant. The attempt to validate the gene using an independent mutant allele generated through CRISPR/Cas9 genome editing was unsuccessful. In a future approach, the *mrd1* gene could be validated either by a gene silencing strategy using RNAi or by a novel allele from publicly available mutant resources, such as the in-house *BonnMu* collection, which is continuously expanding.

2 Introduction

2.1 Maize as an important crop plant

Maize (*Zea mays* L.) is a crucial cereal crop in the world, serving as a primary provider of food, feed, energy, and raw materials for industries. As a food crop, especially in developing countries, maize serves as a primary source of carbohydrates and essential nutrients (Shiferaw *et al.*, 2011). Moreover, maize plays a significant role in the biofuel industry, with over 90% of ethanol in the United States sourced from maize grain (Morris and Hill, 2006). It is also a vital component in livestock feed accounting ~60% of the global maize production (Grote *et al.*, 2021). Maize ranks second after wheat in terms of production area, with cultivation occurring on over 197 million hectares (Cooter *et al.*, 2017). It is projected to take over wheat as the most extensively grown crop by 2030 (Erenstein *et al.*, 2021). In Germany, maize is mainly cultivated for silage for both animal feed and biofuel on approximately two million hectares, and for grain on approximately 0.5 million hectares (DMK, 2022). Hybrid maize currently dominates the global maize production (Masuka *et al.*, 2017). Overall, the global maize production topped with 1.2 billion tons in 2021 with a record growth of over 100% among other cereals since 2000 (FAO, 2021). The world's population is projected to increase to approximately 8.5 billion by the year 2030, with an expected 9.7 billion by the year 2050 (UN, 2022). Currently, more than 50% of maize production occurs in developing countries, and it is projected to double the demand in these regions by 2050 (Rosegrant *et al.*, 2009; Shiferaw *et al.*, 2011). However, maize is vulnerable to extreme agro-climatic conditions, and thus ongoing research efforts are crucial for addressing global food security challenges in the face of changing environments (Tigchelaar *et al.*, 2018).

2.2 Maize: from domestication to a model species for plant genetics

Maize (*Zea mays* L.) belongs to the family Gramineae (Poaceae). It was domesticated from the wild grass teosinte (*Zea mays ssp. parviglumis*) approximately 9,000 years ago in the Central Balsas River Valley of southwestern Mexico (Matsuoka *et al.*, 2002; Piperno *et al.*, 2009; Hufford *et al.*, 2012). A recent study revealed that the origin of modern maize can be traced back to an additional admixture event between ancient maize and *Zea mays* ssp. *mexicana* in the highlands of Mexico some 4000 years after the initial domestication began (Yang *et al.*, 2023). Morphologically, teosinte and maize share a similar overall vegetative structure, but they have five main differences: maize has paired spikelets, a polystichous ear structure, a non-disarticulating rachis, softer and smaller glumes, and reduced axillary branching (Figure 1). However, the transformation from teosinte to maize involved multiple genetic interactions, including dominance relationships, epistatic

interactions, and pleiotropic constraints, contributing to the significant phenotypic evolution (Stitzer and Ross-Ibarra, 2018). Maize has been grown in the Southwestern United States no later than 2100 before christ (Merrill *et al.*, 2009) where the high- and low-altitude-adapted Flint maize were commonly cultivated (Pruitt, 2016). Flint maize got its name from its kernels which are as hard as flint. The Northern Flint maize (Hart and Lovis, 2013) was developed for early maturity and cold tolerance from the introduced Flint corn of the Southwestern United States 4,100 years ago (Troyer, 1999). Shortly after domestication, maize was spread from Mexico to North and South America, eventually to Europe in 1493 by Christopher Columbus with the first field planted in Spain in 1525 (Schlegel, 2017). These seeds were Northern Flints which became major progenitors of European maize and enabled to adapt to European climates (Rebourg *et al.*, 2003).

European maize germplasm is classified into five principal races based on latitude: the "German Flint" adapted to Central Europe's temperate climate, the "North-Eastern European Flint" suited for cooler and variable climates, the "Southern European Flint" for warmer, drier regions, the "Pyrenees-Galice Flint" reflecting the unique microclimates of the Pyrenees and Galicia, and the "Italian Orange Flint" distinguished by its kernel coloration and adaptation to Italian conditions. (Rebourg *et al.*, 2003). European Flint varieties serve as a crucial genetic resource, contributing traits that enhance the hardiness and adaptability of new maize hybrids (Haberer *et al.*, 2020). Recent advancements in genome sequencing, such as those conducted on France Flint line F7, German Flint DK105, and Spanish EP1 varieties, have unveiled detailed insights into the genetic structure of European Flint maize (Haberer et al., 2020).

The Corn Belt Dent maize, the second germplasm pool, was adapted from crossings of Flint maize shortly after the settlement of Europeans (Hallauer, 2000). Dent maize was named because of the dent appearance on top of its kernels. It was known for its plant height, late flowering time and especially white kernels, but higher yield compared to the Northern Flint maize (Troyer, 1999). The two germplasm pools, Flint and Dent, were then gradually used in breeding programs (Smith *et al.*, 2004). In the 1870s, Gregor Mendel, renowned for his breeding experiments with pea plants, also confirmed his proposed rules of inheritance through crossbreeding experiments involving maize (Rhoades, 1984; Coe, 2001). The studies on the diverse kernel characteristics of maize established it as a model system for plant genetics (Rhoades, 1984).

Maize is open-pollinated and a natural cross-pollinator. The male (staminate) flowers develop in the tassel located at the top of the main stem while female (pistillate) flowers are located in the ear emerged at approximately five to six nodes below the tassel (Figure

1B, Bonnett, 1954). Maize with its unique features has also been served as a model organism for fundamental biological research for many decades (Coe, 2001; Strable and Scanlon, 2009). These features include the ease of performing controlled crosses due to its separate male and female flowers, the production of hundreds of large seeds per ear that allows the observation of rare events in kernel color or composition and the deduction of segregation ratios, and the availability of induced mutations for genetic analysis (Hake and Ross-Ibarra, 2015). Meanwhile, several maize genomes of two major germplasms, such as B73 (Schnable *et al.*, 2009; Jiao *et al.,* 2017), W22 (Springer *et al.,* 2018) and Mo17 (Sun *et al*., 2018) of the Dent pool and F7, DK105, and EP1 of the European Flint pool (Haberer *et al.*, 2020) have been sequenced. While the maize genome sequencing allows for the identification of all gene sequences, the biological functions of the majority of these genes are still unknown.

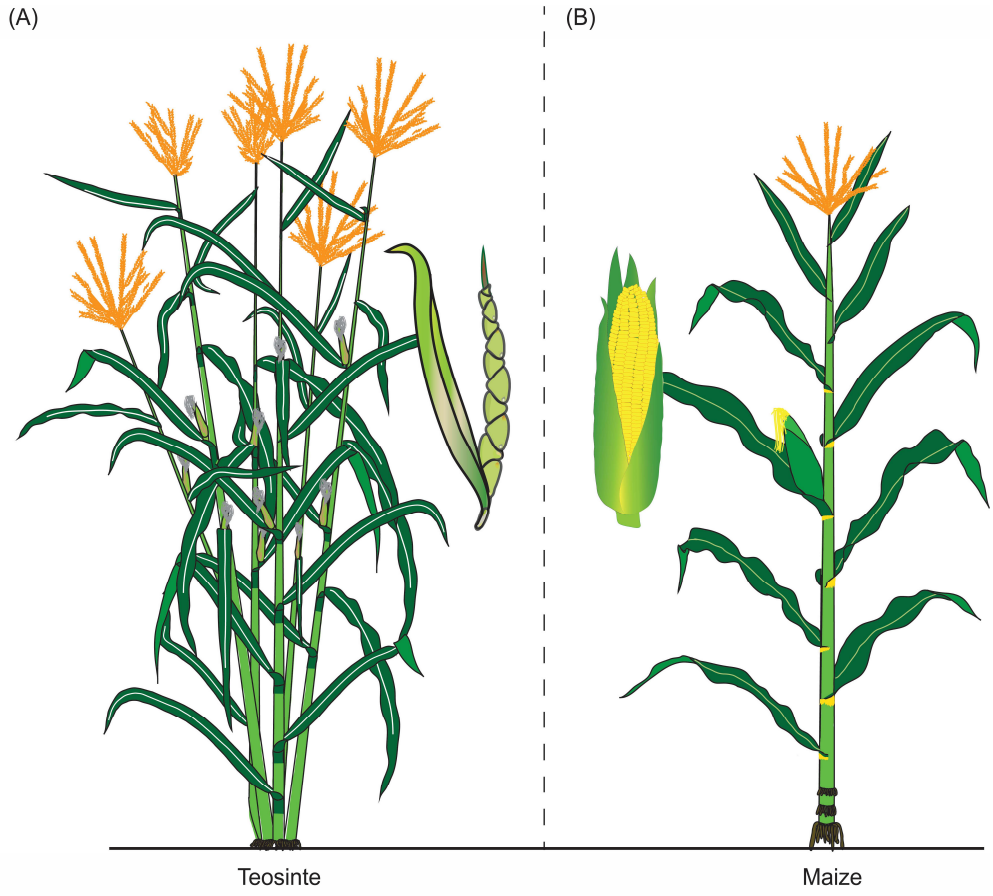


Figure 1. Aboveground phenotypes of teosinte and maize. (A) Teosinte and (B) maize plant architecture with magnified ear structures.

2.3 Functional genetics studies and genome-wide insertional mutagenesis in maize

The model organism maize has a long history of genetic investigation. Since the early 1900´s geneticists are collecting and describing maize mutations affecting a broad range of biological processes. However, to date only a few hundred genes have been functionally characterized based on visible morphological mutant phenotypes (Schnable and Freeling, 2011; https://www.maizegdb.org). Forward genetic experiments represent a classical method to unravel gene functions by cloning a gene based on a mutant phenotype. In contrast, reverse genetic screens allow for the identification of mutant phenotypes based on available disrupted gene sequences (Candela and Hake, 2008). Both approaches have been successfully applied in the past to identify and characterize multiple genes underlying maize root development (Nestler *et al.*, 2014; Xu *et al.*, 2015; Li *et al.*, 2016). For instance, in a forward genetic analysis, Nestler *et al*. (2014) identified the gene *root hair defective 5* (*rth5*), which plays a key role in root hair initiation and elongation in maize. Similarly, a reverse genetics study by Li *et al.* (2016) revealed the function of *root hair defective 6* (*rth6*), a gene encoding a D-type cellulose synthase, which is crucial for controlling the transition from bulge formation to tip growth in maize root hairs.

Transposons, also known as transposable elements (TEs), are mobile DNA elements first discovered in maize (McClintock, 1951). They are classified into two major classes: class I retrotransposons, which necessitate an RNA intermediate for transposition, and class II DNA transposons, which directly transpose via DNA transposase. In class I retrotransposons, there are two subclasses: LTR (long terminal repeat) retrotransposons, characterized by 4- to 6-base pair target site duplications (TSD), and non-LTR retrotransposons. LTR retrotransposons are prevalent in the B73 maize genome, constituting over three-quarters of its genetic makeup (Schnable *et al*., 2009; Haberer *et al*., 2020; Ou *et al*., 2022). In maize, Class II TEs serve as biological mutagens due to their ability to move within the genome, thereby potentially disrupting gene function (Feschotte *et al*., 2002). Autonomous TEs are capable of transposition on their own, as they produce their own transposase enzyme, while non-autonomous TEs rely on the transposase produced by autonomous TEs for their movement. These TEs have a tendency to insert in or near genes (Dietrich *et al*., 2002; Fernandes *et al*., 2004; Settles *et al*., 2004), which is beneficial for conducting genome-wide insertional mutagenesis screenings. *Mutator* (*Mu*) transposons, the most active class II transposon family in maize, comprise of an autonomous element (*MuDR*; Robertson, 1978), and multiple non-autonomous elements (Figure 2; Lisch, 2002; Tan *et al.*, 2011). All identified *Mu* transposons conserve highly similar ~215-bp terminal inverted repeats (TIRs) at both ends of the elements, and create

9-bp target site duplications (TSD) directly flanking the *Mu* transposon sequences upon insertion. *Mutator* transposons randomly target genes throughout the maize genome (Lisch, 2015). As such, *Mu* insertion site frequencies strongly correlate with gene density (Schnable *et al.*, 2009).

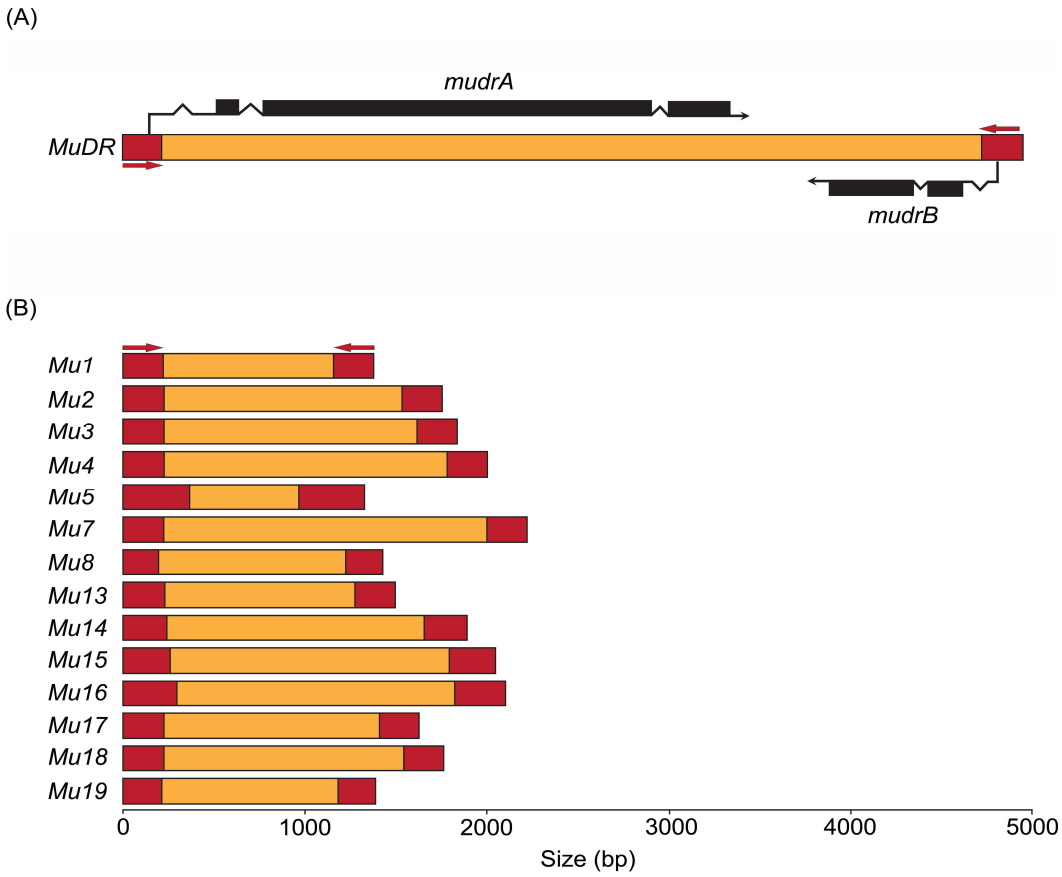


Figure 2. *Mutator* transposons in maize. (A) Model of an autonomous *Mutator* transposon *MuDR* (adapted from Lisch, 2015). The two transcripts, *mudrA* and *mudrB*, are shown above and below. Exons are depicted as black boxes. Introns are depicted as black lines. (B) The schematic structures of non-autonomous *Mutator* elements (adapted from *Tan et al.*, 2011; Lisch 2015). Red boxes and arrows depict terminal inverted repeats (TIRs) at the end of the transposons. Orange boxes represent *Mutator* internal sequences.

Genome-wide insertional mutagenesis, a powerful reverse genetics tool, involves the insertion of a DNA sequence, such as a transposon or T-DNA, to generate loss-of function mutations for virtually all genes of a genome. This method disrupts gene function, enabling the functional analysis of the resultant phenotypes and infer the roles of the affected genes. In *Arabidopsis thaliana*, T-DNA insertional mutagenesis has been extensively used to disrupt gene function (Alonso *et al.*, 2003). In maize, the two main DNA transposon families used are *Activator*/*Dissociation* (*Ac*/*Ds*) and *Mu*, with the latter being more widely used due to its higher transpositional activity and preference for genic regions (McCarty *et al.*, 2005; Vollbrecht *et al.*, 2010; Brutnell and Conrad, 2013). *Mutator*-based sequence-

indexed libraries such as *UniformMu*, *BonnMu*, *ChinaMu*, Maize-Targeted Mutagenesis population (MTM) and *Mu*-Illumina (Bensen *et al*., 1995; May *et al*., 2003; McCarty *et al*., 2005; Williams-Carrier *et al*., 2010; Liang *et al*., 2019; Marcon *et al*., 2020) and *Ac*/*Ds* based library such as Trait Utility System for Corn (TUSC; Briggs and Meeley, 1995) have been instrumental in maize functional genomics.

Recently, three sequence-indexed mutant libraries have been established by *Mu* transposon insertional mutagenesis as invaluable resources for functional genetics studies in maize (*UniformMu*: (McCarty, *et al*., 2013a); *ChinaMu*: (Liang *et al.*, 2019); *BonnMu*: (Marcon *et al.*, 2020)). These mutant collections are ideal starting points for forward and reverse genetic screens (1) to validate candidate genes by additional allelic mutations and (2) to functionally characterize novel mutants regulating various developmental processes (Hunter *et al.*, 2014; Dai *et al*., 2021; He *et al*., 2024).

2.4 *BonnMu – A* sequenced-indexed transposon-tagged mutant resource

The *BonnMu* library represents a novel and important addition to the existing resources for maize functional genetics studies (Marcon *et al.*, 2020). It was initially constructed in the B73 inbred background (Marcon *et al.*, 2020), while the *UniformMu* utilizes W22 inbred background (McCarty *et al.*, 2013a). The *BonnMu* resource utilized random *Mu* insertions to disrupt genes, based on a method called Mutant-seq (Mu-seq; McCarty *et al.*, 2013a). Mu-seq enables the identification of maize F_2-families carrying transposon insertions in a sequence indexed (i.e. transposon tagged) population, by high-throughput next generation sequencing (NGS). The analysis of this European-based sequence-indexed resource has been optimized and accelerated by the MuWU bioinformatic pipeline (Stöcker *et al.*, 2022), which facilitates unbiased and high-throughput sequencing of *Mu*-tagged families.

In a pilot study, sequencing of 1,152 *Mu*-tagged *BonnMu* F_2-families resulted in the identification of 225,936 genomic *Mu* insertion sites, with 41,086 high-quality germinal *Mu* insertions covering 16,392 annotated maize genes (37% of the B73v4 genome; Figure 3; Marcon *et al*., 2020)). *BonnMu* complemented the *UniformMu* (McCarty *et al*. 2013) and the *ChinaMu* (Liang *et al*. 2019) resources, and as such these three collections comprised *Mu* insertions in 57% (25,140 of 44,117 genes) of all genes in the reference genome. A unique feature of *BonnMu* is that insertions and associated phenotypic seedling pictures of *Mu*-tagged F_2-families are deposited at MaizeGDB.org to enable open access (Marcon *et al*., 2020). Here in this thesis, it is highlighted that *BonnMu* has expanded its collection to include two germplasm pools: the Dent germplasm (B73 and Co125) and the Flint germplasm (DK105, EP1, and F7).

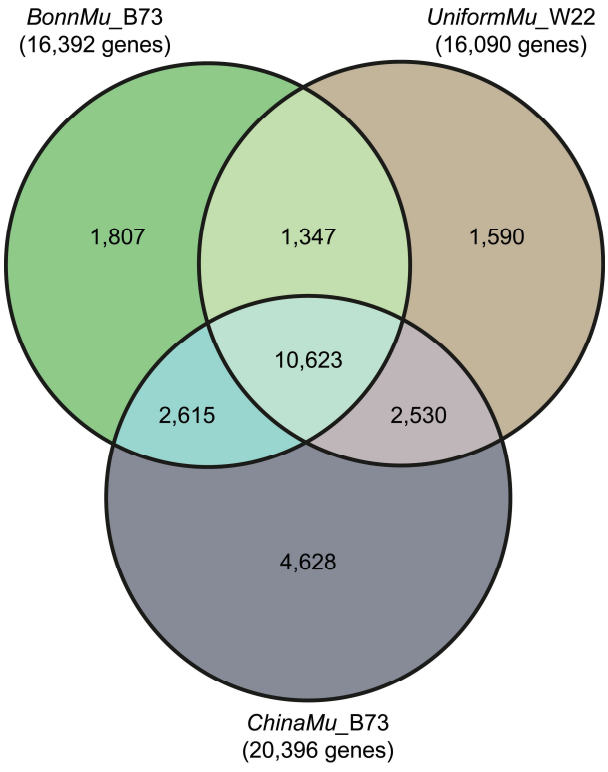


Figure 3. Number of *Mutator*-tagged genes in the *BonnMu* (B73 background), *ChinaMu* (B73 background) and *UniformMu* (W22 background) resources (adapted from Marcon *et al*., 2020)**.**

2.5 The maize root system

The maize root system plays a central role in water and nutrient uptake, rhizosphere microbial communication, and anchorage to the soil (Lynch, 1995). It consists of embryonic roots which are formed during embryogenesis, and postembryonic roots which are initiated after germination (Figure 4; Hochholdinger and Tuberosa, 2009). The embryonic roots include one single primary root which originates from the basal pole of the embryo and several seminal roots which develop at the scutellar node of the embryo. In contrast, postembryonic roots consist of several crown (below-ground) and brace (above-ground) roots which are born from the shoot nodes. Lateral roots emerge from both embryonic and postembryonic root-types. Similarly, root hairs, which emerge from epidermal cells appear along all root-types (Hochholdinger *et al.*, 2004, Hochholdinger *et al*., 2018a).

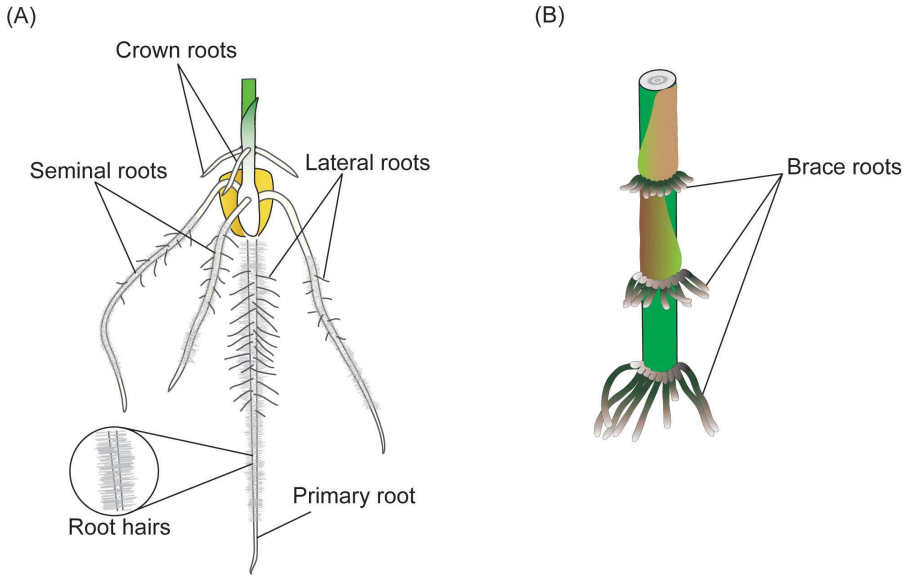


Figure 4. Schematic pictures of maize root system architecture. (A) Embryonic primary and seminal roots and postembryonic lateral and crown roots roots emerged in a 12-day-old wild type maize seedling. (B) Above ground shoot-borne brace roots of a 6-week-old maize plant.

In maize, several genes have been identified that affect various root types at different developmental stages, including root hair, shoot-borne roots, lateral, and seminal roots (Hochholdinger *et al.*, 2004; von Behrens *et al.*, 2011; Hochholdinger *et al.*, 2018b; Baer *et al.*, 2023). The *rootless concerning crown and seminal root* (*rtcs*) mutant lacks shoot-borne crown and brace roots, and seminal roots. Its paralog, *rootless concerning crown and seminal roots like* (*rtcl*) impacts shoot-borne root elongation and is involved in auxin signal transduction. While *rtcs* mutants show defective seminal root initiation, *rtcl* mutants have shorter crown roots but normal seminal roots (Taramino *et al.*, 2007; Xu *et al.*, 2015). The *rootless with undetectable meristem 1* (*rum1*) affects lateral root development in the primary root (Woll *et al.*, 2005). Its close relative, *rootless with undetectable meristem like 1* (*rum1-like1* or *rul1*), shares a high degree of amino acid sequence similarity (von Behrens *et al.*, 2011; Zhang *et al.*, 2016). Lastly, *root hair defective* (*rth1*, *rth3*, *rth5*, *rth6*) genes, crucial for root hair formation, have been identified. The gene *rth1* affects the exocytotic growth of root hair tips (Wen *et al.*, 2005), while *rth2* mutants have significantly shorter root hairs. The *rth3* gene, responsible for secondary cell wall organization, also impacts root hair elongation and interacts with *rth5* and *rth6* during this process (Hochholdinger *et al.*, 2008; Nestler *et al.*, 2014; Li *et al.*, 2016; Hochholdinger *et al.*, 2018). For the *BonnMu* resource as a forward genetic tool, we conducted phenotyping of all *BonnMu* families at the 10-12 days seedling stage. We identified and documented various

shoot and root mutant phenotypes, which led to the discovery of several novel root mutants.

2.6 The role of CSN in plants

The COP9 (constitutive photomorphogenesis 9) signalosome (CSN) is a multiprotein complex discovered in Arabidopsis through genetic screening of light-responsive mutants (Parks *et al.*, 1989). This complex predominantly consists of eight subunits (CSN1 to CSN8) and is conserved in higher eukaryotes including plants and animals. The subunits CSN1 to CSN4 and CSN7 to CSN8 contain the PCI (proteasome, COP9, initiation factor 3) domain, crucial for protein-protein interactions within the complex (Figure 5A). In contrast, subunits CSN5 and CSN6 are characterized by the MPN (Mov34 and Pad1 N-terminal) domain, which is implicated in the complex's metalloprotease activity. Additionally, comprehensive genetic analyses have been conducted across all subunits of the COP9 signalosome in Arabidopsis (Serino *et al*., 2003).

The CSN complex is involved in the regulation of protein degradation. It specifically modulates the activity of the SCF^{TIR1} ubiquitin ligase complex (SCF^{TIR1}: Skp1, Cullin, F-box containing complex with Transport Inhibitor Response 1, TIR1), a critical component of the auxin signaling pathway (Eckardt, 2003). The CSN complex facilitates the recycling and reassembly of the SCF^{TIR1} complex through the deneddylation of AtCUL1, a component of CULLIN-RING E3 ligase (CRL) complexes. This regulation is essential for the degradation of AUX/IAA proteins, which act as repressors.

Upon auxin binding to the SCF^{TIR1} complex, these proteins are targeted for ubiquitination and subsequent degradation, thereby influencing auxin-responsive gene expression and plant development (Serino *et al.*, 2003). The balance of AUX/IAA protein levels is further influenced by the action of AXR1 and ECR1, which are subunits of an E1-like RUB1-activating enzyme. This enzyme activates r (RUB1/NEDD8) that subsequently binds to the E2-like protein RCE1, promoting the neddylation of CUL1, a process vital to the function of the CSN complex. It has also been shown that the COP9 signalosome (CSN) plays a central role in the ubiquitin-proteasome pathway (Figure 5B; Schwechheimer, 2004; Wei, Serino and Deng, 2008).

CSN exerts significant influence over key biological processes including gene expression, cell proliferation, and the cell cycle (reviewed in Qin *et al.*, 2020). The significance of CSN extends beyond its regulatory functions. It serves as a vital node in the plant's response to a range of external stimuli, such as light and temperature fluctuations, as well as internal cues, notably phytohormones. This broad spectrum of regulatory capabilities highlights the complexity and versatility of CSN in plant development and stress responses.

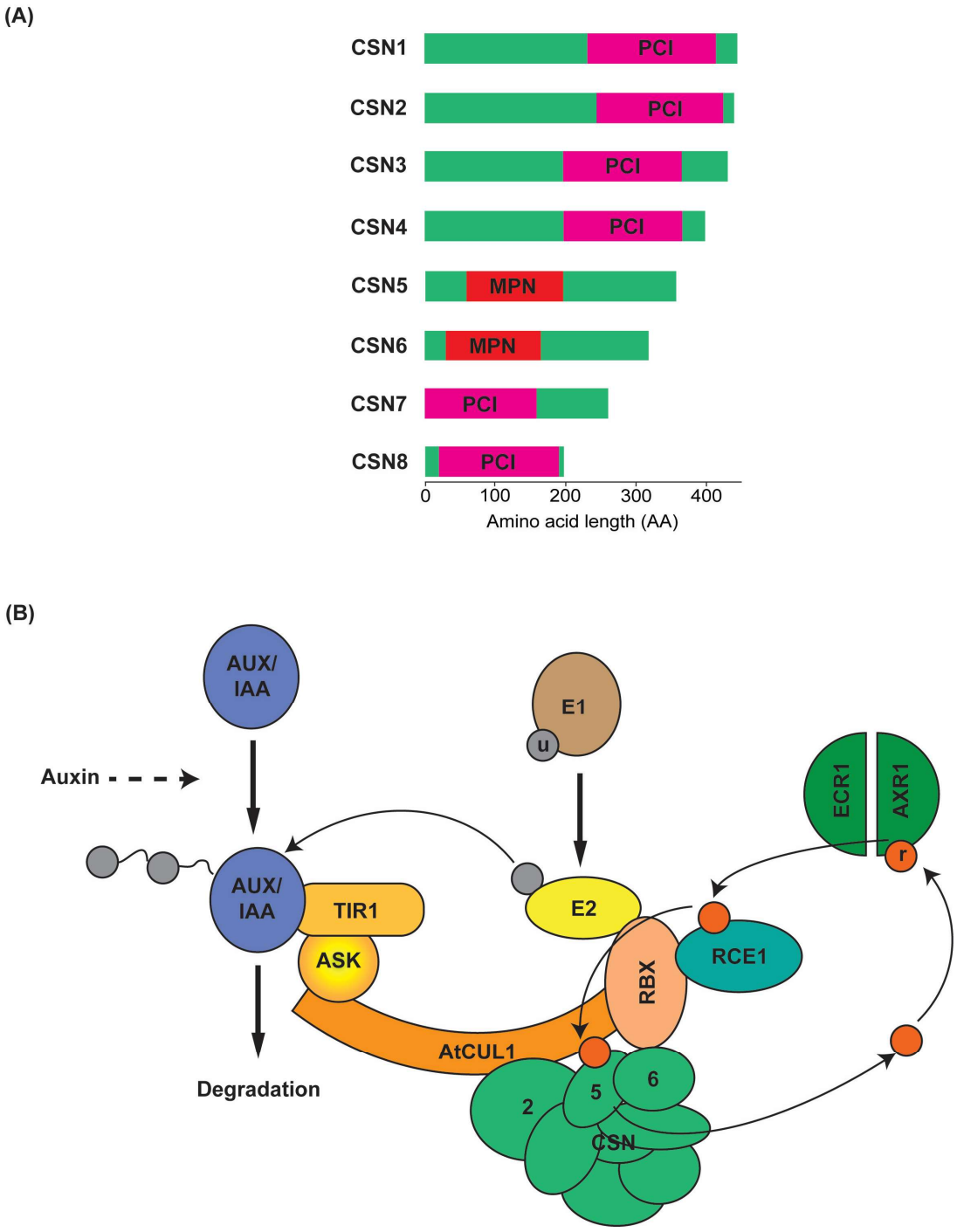


Figure 5. Eight subunits of CSN (retrieved from UniProt database) and a model of CSN function (adapted from Eckardt 2003). (A) Eight subunits of CSN in Arabidopsis. CSNs conserves Proteasome, COP9, Initiation factor 3(PCI) domain except CSN5 and CSN6 that contain the Mov34 and Pad1p N-terminal (MPN) domain. The UniProt accession numbers of each subunit are P45432 (CSN1), A0A178VSH3 (CSN2), Q8W575 (CSN3), Q8L5U0 (CSN4), Q8LAZ7 (CSN5), Q8W206 (CSN6), Q94JU3 (CSN7) and P43255 (CSN8). (B) A model of CSN complex. Ubiquitin-activating enzyme (E1); Ubiquitin-conjugating enzyme (E2); four parts of CULLIN-RING E3 ligase (AtCul1, RBX1, ASK, F-box protein TIR1); Subunits of E1-like RUB1-activiating enzyme (ECR1 and AXR1); E2-like protein (RCE1), Ubiquitin-like molecule RUB1/NEDD8 (r) and COP9 signalosome (CSN).

In general, seedlings exhibit photomorphogenic growth in light and skotomorphogenic growth in darkness. Photomorphogenesis (de-etiolation), the light-favored development, is characterized by the presence of short hypocotyls, open apical hooks, and expanded cotyledons, which accompany chloroplast development, cell differentiation, and anthocyanin accumulation. In contrast, skotomorphogenesis (etiolation) represents the dark-favored state, characterized by elongated hypocotyls, closed cotyledons, and the development of etioplasts — a precursor to chloroplasts in the absence of light (Wei and Deng, 2003). Arabidopsis *cop9* null mutants exhibit severe photomorphogenic development and excessive anthocyanin accumulation, leading to seedling death (Serino and Deng, 2003). Moreover, the *cop9* mutant exhibits greenish roots when grown under light, a phenomenon not observed in mutants grown in darkness or in wild type plants under the light condition (Wei and Deng, 1992). Similarly, the Arabidopsis *csn4* knockout mutants also show constitutive photomorphogenesis and severe developmental defect resulting in seedling lethality (Figure 6; Betsch *et al*., 2019).

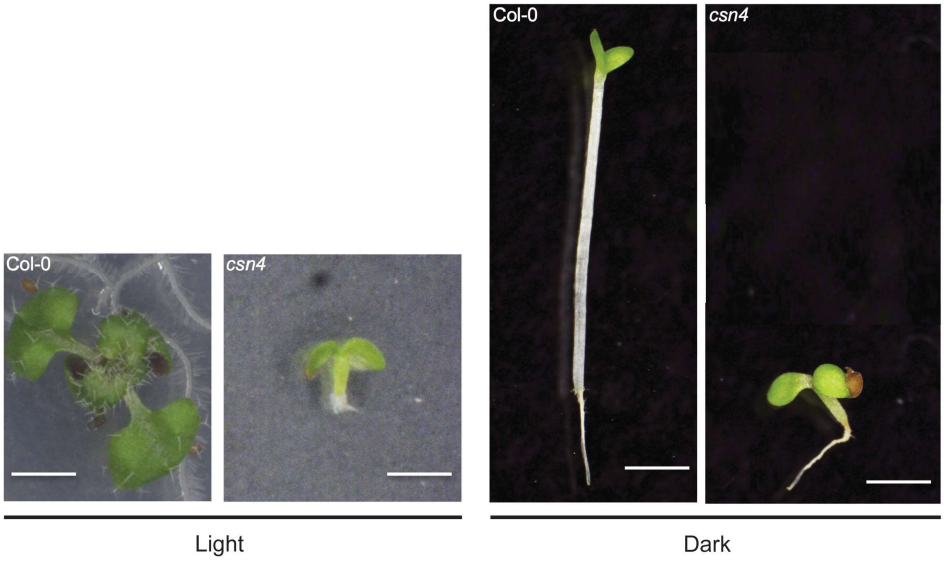


Figure 6. Photomorphogenic and skotomorphogenic growth of 10-day-old *csn4* mutant and Col-0 wild type plants under light and dark conditions, adapted from Betsch *et al*., (2019). The size of the bar represents 500µm.

Arabidopsis lines with reduced CSN function show altered photomorphogenesis, irregular cotyledon numbers, and disrupted leaf formation, indicating that CSN is crucial for multiple developmental pathways (Peng *et al*., 2001). These findings indicate the fundamental role of CSN in plant development. The functional characterization of the COP9 signalosome (CSN) has been studied well in Arabidopsis. The mutants displaying constitutive photomorphogenesis in the dark, resulted in the discovery of the *cop* (*constitutively*

photomorphogenic), *det* (*de-etiolated*), and *fus* (*fusca*) mutants. These mutants are collectively known as *cop*/*det*/*fus* mutants in Arabidopsis (Chory *et al.*, 1989; Wei *et al.*, 1996; Schwechheimer and Deng, 2000).

In this dissertation, we explore the new maize mutant, *magenta root dwarf 1* (*mrd1*), which was identified in a forward genetic screening of *BonnMu* F2-families. The mutant exhibits characteristics potentially indicative of a CSN complex disruption. The *mrd1* mutant is characterized by anthocyanin accumulation and de-etiolated growth in darkness, which are comparable to those observed in *csn4* mutants in *Arabidopsis thaliana*. In order to validate the *csn4* as causal for the *mrd1* phenotype, additional independent alleles were generated via CRISPR/Cas9 genome editing. In this dissertation, the identification and characterization of *mrd1* mutant is comprehensively presented including BSR-seq analysis, anthocyanin profiling, rhizosphere microbiome analysis, transcriptomic analysis via RNA-seq, phenotype to genotype co-segregation analysis, gene expression analysis via qRT-PCR, subcellular localization of MRD1 protein, and validation of gene function via CRISPR/Cas9.

3 Objectives

The overall objective of my dissertation was to expand the *BonnMu* resource by including the F_2-familes of mutagenized European inbred lines such as F7, DK105 and EP1 of the Flint pool. The following hypotheses were tested in this study.

Project I: *BonnMu* resource expansion

1. The *BonnMu* resource achieves extensive coverage of the maize genome, with a majority of *Mu* insertions occurring in gene-dense regions.
2. A significant fraction of the *Mu* insertions is located in genic regions, particularly the 5' untranslated region (UTR), indicating a pattern of mutagenesis that preferentially targets regulatory regions of genes.
3. The gene length correlates with the number of insertions.
4. The observed overlap of *Mu*-tagged genes across different *BonnMu* genetic backgrounds significantly deviates from the expected random distribution.

Project II: Identification and characterization of *mrd1* mutant

1. The *mrd1* mutant is a recessive mutant, segregating in a 3:1 Mendelian ratio.
2. The *mrd1* mutant is reduced in the overall growth compared to the wild type.
3. The *Mu* insertion in the *mrd1* gene co-segregates with the magenta-colored root dwarf phenotype in the *mrd1* mutant by PCR-based genotyping.
4. The *mrd1* expression is down-regulated in the primary root of the mutant compared to wild type.
5. The *mrd1* gene shows down-regulation in the crown root of the mutant compared to wild type by RNA-seq.
6. The MRD1 protein is localized in the nucleus.
7. A novel CRISPR/Cas9 edited allele of *mrd1* results in seed lethality and a knock-out mutation.

4 Material and methods

4.1 Plant material

Mutagenized *BonnMu* F_2-families were generated in field nurseries at the University of Bonn (Germany), Chile, Hawaii, and Mexico in the years 2014-2021 as previously described (Marcon *et al.*, 2020, 2024a, 2024b; Win *et al*., 2024b). Briefly, we obtained the F_1-population by crossing a *Mu*-active line (*Mu4* per se; Robertson, 1983) into five distinct inbred lines: B73, Co125, DK105, EP1, and F7 (Table 1). Then, the F_2-population, segregating for recessive mutations, was generated by selfing all plants of the F_1-generation. The *BonnMu* F_2-families comprise a genetic background of 50% inbred line (e.g. B73) and 50% *Mu*-active line.

Table 1. *BonnMu* F_2-families used for Mu-seq library construction.

Germplasm	Mu-seq library_ID	*BonnMu* F_2-families	Genetic background[1]
Dent pool	Museq_1	*BonnMu*-1-A-0001 - *BonnMu*-1-A-0576	(**B73** x *Mu[4] per se*) @
	Museq_2[2]	*BonnMu*-2-A-0577 - *BonnMu*-2-A-1152	
	Museq_3	*BonnMu*-3-E-0001 - *BonnMu*-3-E-0576	(**Co125** x *Mu[4] per se*) @
	Museq_4[2]	*BonnMu*-4-A-1153 - *BonnMu*-4-A-1655 & *BonnMu*-4-B-0001 - *BonnMu*-4-B-0073	(**B73** x *Mu[4] per se*) @
	Museq_5	*BonnMu*-5-B-0074 - *BonnMu*-5-B-0650[3]	
	Museq_6	*BonnMu*-6-B-0651 - *BonnMu*-6-B-0922 & *BonnMu*-6-C-0001 - *BonnMu*-6-C-0304	
	Museq_7	*BonnMu*-7-C-0305 - *BonnMu*-7-C-0805[3] & *BonnMu*-7-D-0001 - *BonnMu*-7-D-0077	
	Museq_8	*BonnMu*-8-D-0078 - *BonnMu*-8-D-0653	

Table 1. *BonnMu* F_2-families used for Mu-seq library construction (continued).

Germplasm	Mu-seq library_ID	*BonnMu* F_2-families	Genetic background[1]
Flint pool	Museq_DK105_EP1_1	*BonnMu*-9-G-0001 - *BonnMu*-9-G-0462 & *BonnMu*-9-H-0001 - *BonnMu*-9-H-0114	(**DK105** x *Mu*[4] *per se*) @
			(**EP1** x *Mu*[4] *per se*) @
	Museq_EP1_2	*BonnMu*-10-H-0115 - *BonnMu*-10-H-0690	(**EP1** x *Mu*[4] *per se*) @
	Museq_F7_1	*BonnMu*-F7-1-F-0001 - *BonnMu*-F7-1-F-0576	(**F7**x *Mu*[4] *per se*) @
	Museq_F7_2	*BonnMu*-F7-2-F-0577 - *BonnMu*-F7-2-F-1152	
	Museq_F7_3	*BonnMu*-F7-3-F-1153 - *BonnMu*-F7-3-F-1728	
	Museq_F7_4	*BonnMu*-F7-4-F-1729 - *BonnMu*-F7-4-F-2304	

[1] Flint and Dent lines were mutagenized by crossing with a *Mu*-active line (*Mu*[4] per se). Selfing (@) of the F_1-plants generated F_2-families segregating for recessive mutations in a 3:1 ratio. Therefore, the *BonnMu* F_2-families comprise a genetic background of 50% inbred line (e.g. B73) and 50% a *Mu*-active line. In the main text the genetic background of the different *BonnMu* F_2-families is referred to as B73, Co125, DK105, EP1, and F7, respectively.

[2] *Mu*-seq libraries used in previous publication (Marcon *et al.*, 2020).

[3] *BonnMu* F_2-families 5-B-0077, 7-C-0726 and 7-C-0738 failed to germinate. Consequently, they were not included in the respective Mu-seq libraries.

4.2 Chemicals

The chemicals used in the study are listed in Table 2.

Table 2. List of chemicals used in the study.

Name	Manufacturer
2-Mercaptoethanol	Carl Roth, Germany
2-Propanol	VWR, Germany
Acetic acid	
Agarose basic (A8963,0500)	AppliChem, Germany
Bromophenol blue - sodium salt	Merck, Germany
Chloroform	AppliChem, Germany
Ethanol	VWR, Germany
Ethylenediaminetetraacetic acid	
Glycerine	
Hydrogen peroxide	Merck, Germany
Isopropanol	VWR, Germany
Magnesium chloride	
Methanol	
Phenol/Chloroform/Isoamylalkohol (25:24:1)	Merck, Germany
Sodium chloride	Carl Roth, Germany
Sodium hydroxide	
Tris-base	AppliChem, Germany
Urea (U5378-1kg)	Merck, Germany

4.3 Reagents and kits

The reagents and kits used in the study are shown in Table 3.

Table 3. List of reagents and kits used in the study.

Name	Manufacturer
Agilent DNA 7500 Kit	Agilent Technologies, Germany
CleanNGS	CleanNA, The Netherlands
CloneJet PCR Cloning Kit	Thermo Fisher Scientific, Germany
Gateway™ BP Clonase™ II Enzyme mix	
Gateway™ LR Clonase™ II Enzyme mix	
Homemade Taq DNA Polymerase	AG Hochholdinger, University of Bonn, Germany
Monarch® Plasmid Miniprep Kit	NEB Ipswich, USA
PerfeCTa SYBR Green SuperMix	Quantabio, USA
Phusion fidelity DNA polymerase	AG Hochholdinger, University of Bonn, Germany
qScriptcDNA SuperMix	Quantabio, USA
Quick Blunting™ Kit, E12011	New England Biolabs, USA
Rapid DNA Ligation Kit	Merck, Germany
RNeasy Plant Mini Kit	Qiagen, Germany
T4 DNA Ligase	Thermo Fisher Scientific, Germany
Zymoclean-5 DNA Clean and Concentrator™ Kit (D4014)	Zymo Research, Germany
Zymoclean Gel DNA Recovery Kit (D4008)	
Zymoclean-25 DNA Clean and Concentrator™ Kit	

4.4 Bacterial strains

Table 4. List of bacterial strains used in the study.

Name	Manufacturer
Escherichia coli DH5α	Thermo Fisher Scientific, Germany
Agrobacterium tumefaciens AGL-1	Lazo *et al.,* 1991
Agrobacterium tumefaciens C58C1	Ashby *et al*., 1988
Agrobacterium tumefaciens EHA105 recA-	Rodrigues *et al.,* 2020

4.5 Vectors

Table 5. List of vectors used in the study.

Name	Bacterial selection	Application	Sources
pJET1.2/blunt	Ampicillin	PCR cloning	Thermo Fisher Scientific, Germany
pDONR™ 221 P1-P4	Kanamycin	Cloning for BiFC and FRET	
pDONR™ 221 P3-P2	Kanamycin	Cloning for BiFC and FRET	
pFRETvr-2in1-CC	Spectinomycin	Subcellular localization	
pGG-A-OsU3-BbsI-ccdB-CmR-BbsI-B	Ampicillin	CRISPR	www.gatewayvectors.vib.be
pGG-B-OsU3-BbsI-ccdB-CmR-BbsI-C	Ampicillin	CRISPR	
pGG-C-linker-G	Gentamicin	Linker/ CRISPR	
pG3HI-zCas9-tdT-AG	Gentamicin	CRISPR	

4.6 Software and tools

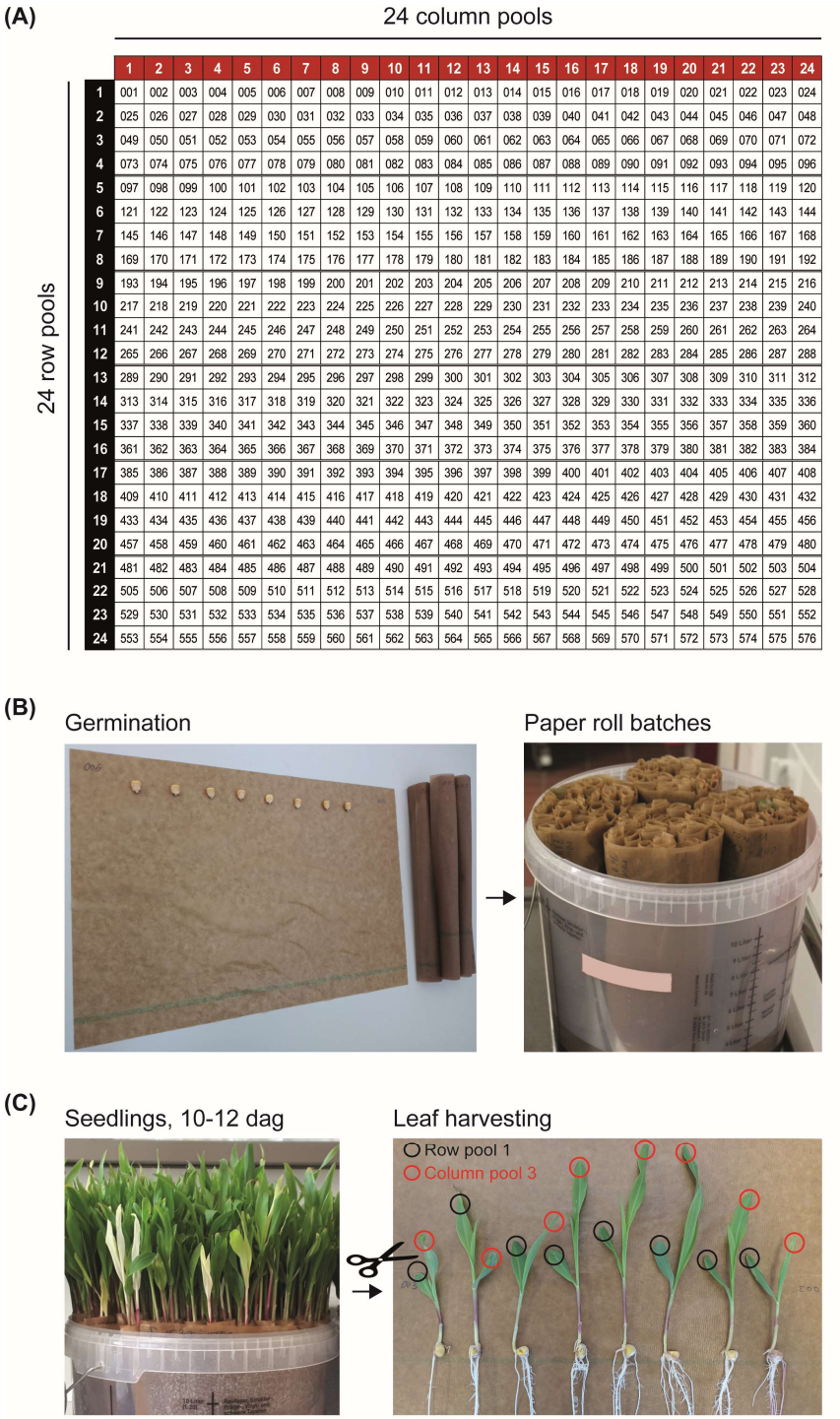
Table 6. List of software and online tools used in the study.

Name	Manufacturer/ Reference/ Website
BioEdit Sequence Alignment Editor	http://www.mbio.ncsu.edu/bioedit/bioedit.html
CLC Genomics Workbench 24	Qiagen, USA
CRISPOR	http://crispor.tefor.net/crispor.py
EnsemblPlants	https://plants.ensembl.org/index.html
FigTree v1.43	http://tree.bio.ed.ac.uk/software/figtree/
SYNTHEGO	https://tools.synthego.com/#/
Leica VT1200	Leica, Germany
Localizer	Sperschneider *et al.*, 2017
Maize Genetics and Genomics Database	http://www.maizegdb.org
Mega X	Kumar *et al.*, 2018
MrBayes	Ronquist and Huelsenbeck, 2003
National Center for Biotechnology Information (NCBI)	http://www.ncbi.nlm.nih.gov
Phytozome 13	https://phytozome-next.jgi.doe.gov/
R program	R Core Team, 2021
TIDE	https://tide.nki.nl/
UniProt	https://www.uniprot.org/
ZEN 3.4	Zeiss, Germany
Zeiss PALM MicroBeam microscope and software	Zeiss, Germany

4.7 Construction of Mu-seq libraries

BonnMu libraries were constructed using the Mu-seq method (Supplemental figure S1; McCarty *et al*., 2013a; Liu *et al*., 2016; Marcon *et al*., 2020, 2024a, 2024b; Win *et al*., 2024b). Briefly, we used a total of 6,912 *BonnMu* families for 12 Mu-seq libraries in the genetic backgrounds of B73, Co125, DK105, EP1, and F7 (Table 6). A 2-dimensional 24 x 24 grid design was utilized to pool 576 *BonnMu* F_2-families per library (Figure 7A; Marcon *et al*., 2024a). For each Mu-seq library construction, we germinated eight seeds per family using a paper roll system (Figure 7B; Hetz *et al*., 1996). We incubated the seedlings in a climate chamber with a photoperiod of 16 h (28 °C, 2700 lux) and a dark period of 8 h (21 °C) at 70% humidity. At 10-12 days after germination (dag), the leaf samples were harvested and pooled based on the 24 x 24 grid design. The samples were taken from at least 3 seedlings of each F_2-family to ensure the presence of at least one mutant allele per *Mu*-tagged gene within the 3-8 germinated plants per F_2-family. By applying the given formula through the dbinom() and dhyper() functions in R (R Core Team, 2021), the calculated probability was determined to be 99% (Supplemental table S1). To ensure the precise identification of heritable germinal insertions at the intersections of rows and columns in the grid, leaf samples from independent somatic cell lineages, i.e., alternate leaves of each seedling per family, were sampled in one distinct row and one distinct column pool (Figure 7 C). By using this method, somatic insertions appeared only in a single axis of the grid and were subsequently excluded from downstream analyses. The harvested samples were frozen in liquid nitrogen and kept at -80 C before use. For each library, the frozen leaf samples were ground manually using pre-cooled mortars and pestles.

Subsequently, genomic DNA was isolated from each pool according to Nalini *et al*., (2003). The genomic DNA was randomly sheared using a Bioruptor® Pico sonication device (Diagenode) at 2s-on/2s-off setting for 2-4 cycles to obtain the fragment sizes of about 1 kb (Supplemental figure S1). The size of fragmented genomic DNA was analyzed by agarose gel electrophoresis after sonication. The randomly sheared genomic DNA fragments had single-stranded overhangs, which were attentively filled in using an enzyme mix (Quick Blunting™ Kit, E1201L, New England Biolabs). This process generated blunt ends, facilitating the subsequent ligation of a double-stranded universal (U) adapter. All primers used in the library construction are listed in the Supplemental table S2).



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	001	002	003	004	005	006	007	008	009	010	011	012	013	014	015	016	017	018	019	020	021	022	023	024
2	025	026	027	028	029	030	031	032	033	034	035	036	037	038	039	040	041	042	043	044	045	046	047	048
3	049	050	051	052	053	054	055	056	057	058	059	060	061	062	063	064	065	066	067	068	069	070	071	072
4	073	074	075	076	077	078	079	080	081	082	083	084	085	086	087	088	089	090	091	092	093	094	095	096
5	097	098	099	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
6	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144
7	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168
8	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192
9	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216
10	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240
11	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264
12	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288
13	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312
14	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336
15	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360
16	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384
17	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408
18	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432
19	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456
20	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480
21	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504
22	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528
23	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552
24	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576

Figure 7. Multiplexing leaf samples according to a 24 x 24 grid design (Marcon *et al.*, 2024a). (A) A 24 row x 24 column grid is designed to position 576 individual F_2-families. (B) Per F_2-family eight kernels are germinated using a single germination paper (left). The 24 paper rolls of a row pool are bundled in one batch and four batches fit in one bucket (right) for germination of the enrolled kernels. (C) Seedlings are harvested 10-12 days after germination (dag, left), according to the 24 x 24 grid design illustrated in (A). Per F_2-family one leaf tip per plant is sampled into the respective row pool and an opposite leaf tip is harvested into the corresponding column pool (right).

The *Mu*-flanking amplicons were subsequently enriched through a ligation-mediated PCR (PCR-I), using a *Mu*-TIR-specific primer and a specific primer for the ligated U adapter (Supplemental figure S1). Then, the fragments were incorporated with a part of an Illumina sequencing adapter and a TIR sequence in the PCR-II. To minimize the number of very short *Mu*-flanking fragments, PCR-II products were purified using a CleanNGS magnetic bead-based clean-up system (CleanNA). The final PCR-III integrates the remaining sequencing adapters, and 6 bp barcodes which enabled multiplexing of the 48 pools. To assess library quality and quantity, each Mu-seq library was measured using a Bioanalyzer with a DNA 7500 chip (Agilent Technologies) to obtain the required concentrations for sequencing. The multiplexed Mu-seq libraries were subjected to paired-end sequencing with a read length of 150 base pairs (bp) using the HiSeq X Ten sequencing system. The raw sequencing data were stored at the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under BioProject accession number PRJNA914277. The detailed protocol for constructing Mu-seq libraries, including step-by-step instructions, has been published by Marcon *et al*., (2024). This protocol is available for reference to ensure precise replication of the methodology.

4.8 MuWU: Mu-seq library analysis and annotation

Mu-seq reads are processed using an automated processing pipeline, referred to as Mu-Seq Workflow Utility (MuWU; Stöcker *et al.*, 2022). Briefly, *Mu* insertion sites were detected based on the characteristic 9-bp target site duplications at the insertion flanking regions of *Mu* transposons. Combined with the grid design, this allowed the differentiation between germinal and somatic insertion events. After mapping all Mu-seq reads to the B73v5 reference genome (Zm-B73-REFERENCE-NAM-5.0; Zm00001eb.1; Gage *et al.*, 2020), insertions were associated with specific genomic loci. In detail, we considered *Mu* insertion sites in 5’ and 3’ untranslated regions (UTRs) of genes, exons and introns, the ≤2,101 bp upstream promoter regions, and ≤2,101 bp downstream regions of genes. With the release of MuWU v1.5, we added the capability to determine the specific classes of *Mu1*, *Mu8*, and *MuDR* elements for a detected insertion. Details of the implementation and required data are outlined in the software’s GitHub repository (https://github.com/tgstoecker/MuWU). Finally, we generated an output table of germinal insertion events, that included additional information on each event such as genomic location, associated information based on genome annotations and the most likely class of *Mu* element (Win *et al.,* 2024b).

Furthermore, we investigated germinal insertion sites by aligning Mu-seq reads from the 3,456 sequenced *BonnMu* F_2-libraries in the genetic backgrounds of the Flint lines DK105, EP1, and

F7 to their corresponding genomes: DK105 (Zm-DK105-REFERENCE-TUM-1.0; Zm00016a.1), EP1 (Zm-EP1-REFERENCE-TUM-1.0; Zm00010a.1), and F7 (Zm-F7-REFERENCE-TUM-1.0; Zm00011a.1; Haberer *et al.*, 2020). All *BonnMu* insertions and affected genes are listed in corresponding output tables (Supplemental data S2 on the attached CD-R; Win *et al*., 2024b).

4.9 Downstream analysis of *BonnMu* insertion sites

To investigate the presence of *Mu* insertions in different libraries/ genotypes, a presence/ absence intersection matrix was created using the merge() function in R. An Upset plot was generated using the UpSetR package (Conway *et al.*, 2017) in R, providing a visual representation of the intersections among the 14 Mu-seq libraries and five genotypes. To test the relationship between the number of insertions and the length of the affected genes we calculated the Pearson correlation coefficient (*r*) in R v4.3.1 (R Core Team, 2021). To further visualize the distribution of *Mu* insertions across various genomic partitions, i.e., exons, introns, UTRs, and promoter regions, we analyzed the eight Mu-seq libraries in B73 background. To determine if the *Mu* insertion sites align with gene density, the distribution of *BonnMu* insertions in B73 background were aligned with the genes in each chromosome using the MaizeGDB JBrowse genome browser (Woodhouse *et al.*, 2021). In addition, we employed the published ATAC-seq (Ricci *et al.*, 2019), ChIP-seq (Makarevitch *et al.*, 2013; Zhang *et al.*, 2015), and NAM-ATAC and NAM-UMRs datasets (Hufford *et al.*, 2021) to investigate chromatin accessibility and histone modifications in relation to *BonnMu* insertions.

4.10 Confirmation of *Mu* insertion

We performed PCR-based confirmation of *Mu* insertions using the following *BonnMu* F_2-families: *BonnMu*-2-A-0982, *BonnMu*-7-C-0458, and *BonnMu*-F7-2-F-1001. To this end, 12-30 seeds per segregating *BonnMu* F_2-family were germinated using the paper roll system (Hetz *et al.*, 1996). Leaf samples were harvested at 10 days after germination and gDNA was isolated according to Nalini *et al*., 2003. Gene-specific primers flanking the *Mu* insertion sites were designed using Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). To genotype the plants, three different combinations of primers in separate reactions were used: (1) gene-specific forward and reverse primers to detect the presence of the gene copy, (2) gene-specific forward and TIR6 and (3) gene-specific reverse primer and TIR6, to detect the presence of *Mu* insertions. Primer sequences are provided in Supplemental table S3. The PCR was performed using Phusion™ High-Fidelity DNA Polymerase (Thermo Fisher). The

PCR products from the individual plants that showed the presence of *Mu* insertions were subjected to Sanger sequencing (Sanger *et al.*, 1977). The resulting sequences were then analyzed using BioEdit software (Hall, 1999) to confirm the presence and specific locations of the *Mu* insertions.

4.11 Forward genetic analysis of *BonnMu*-F_2-stocks

Forward genetic screens not only identify novel genes but also unveil novel functions for known genes (Wen *et al*., 2005; Taramino *et al*., 2007; Nawy *et al*., 2010; Li *et al.*, 2016; Hochholdinger *et al*., 2018b). Forward genetics starts with an observable phenotype and seeks to identify the genetic basis of this trait. The *BonnMu* resource offers a comprehensive platform for forward and reverse genetic analyses in maize via MaizeGDB (Marcon *et al*., 2020). *BonnMu* comprises of 8,064 *Mu*-tagged F_2-stocks with identified 425,924 germinal *Mu* insertions covering 36,612 (83%) of all annotated maize genes (B73v5; Win *et al*. 2024). As a forward genetic resource, between 12 to 30 kernels from each of 7,679 families were germinated using a paper roll system as described in the method section 4.7. The remaining 385/ 8,064 families were not included in the analysis due to an insufficient amount of seeds. At 10-12 dag, the seedlings of each F_2-familiy were phenotyped and photographed using Canon digital camera (Canon EOS 500D). All the photographs are deposited on the MaizeGDB. The seedling phenotypes were screened and recorded for the characters associated with the primary root, seminal root, lateral root, root hair, crown root, shoot and leaf. Additionally, the mutation rate of *BonnMu* was calculated based on the occurrence of albino and pale green leaf phenotypes following 3:1 (wild type : leaf mutant) Mendalian segregation ratio.

4.12 Phenotype to genotype co-segregation analysis

A forward genetic screen of 10-day-old seedling phenotypes of the *BonnMu* F_2-families identified a *magenta root dwarf 1* (*mrd1*) mutant. For forward genetic screen of *BonnMu*, 12-20 kernels were germinated in a paper roll system as described previously (Win *et al*., 2024a). At 10 DAG, the segregation ratio of *BonnMu*-1-A-0438 segregating for *mrd1* mutation was analyzed using a χ^2 test. The individual plants from the *mrd1* segregating families were then subjected to genotyping by PCR. The genotyping and Sanger sequencing were performed as described in the previous section (4.7).

In addition, the mutant was phenotypically analyzed under both light and dark conditions based on the following parameters: primary root length (cm), shoot length (cm), mesocotyle

length (cm), coleoptile length (cm), number of seminal roots and lateral root density (number of lateral roots in 5 cm region of the primary root from coleorhiza). The kernels segregating for *mrd1* mutant were germinated in a paper roll system and incubated in the plant growth chamber (Gen2000, USA). The transparent buckets (12.5 L) used for this experiment were also wrapped around the side and bottom with aluminum foil to protect from the light. The chamber was divided into light and dark conditions by a light protected sheet and set to a photoperiod of 16 h (28 °C, 2700 lux) and a dark period of 8 h (21 °C) at 70% humidity. In total, 50 kernels each for light and dark conditions were used to obtain 15 representative mutant and 15 wild type plants. At 12 dag, the morphological characters mentioned above were measured accordingly.

Moreover, the differentiation zone of the primary roots of *mrd1* mutant and wild type plants were hand-sectioned at 6 dag. To prepare for sectioning, the primary roots of mutant and wild type plants were detached from the kernels by cutting with a sharp blade. The roots were placed on the slides with droplets of water using soft forceps, which were used to gently handle the roots and prevent any damage during the process. The differentiation zone was carefully identified under the Binocular, defined as the area starting from the side of the root tip where root hairs begin to emerge. After selecting the differentiation zone segment, ~1cm from the root tip, a sharp razor blade is utilized to cut it as thin as possible (<0.5 mm). The sections and the blade are kept wet to ensure the cut sections slide off smoothly into the water. Thin cross-sections are made with the root held gently in place with soft forceps, using a single downward motion with the blade kept perpendicular to the root's longitudinal axis to achieve uniform thickness. Blades are changed frequently to maintain sharpness. After many sections are cut, they are transferred to a new slide. Sections are then arranged on a slide, covered carefully with a coverslip to avoid air bubbles, and finally, the prepared slide is observed under Zeiss PALM MicroBeam microscope (Zeiss).

4.13 BSR-seq analysis

The bulked segregant RNA-seq (BSR-seq) is a modified method of the bulked segregant analysis (BSA) that uses RNA-seq reads (Figure 8; Li *et al.*, 2016; Win *et al*., 2024a). We used these reads to rapidly map a gene in a population with polymorphic markers, such as single nucleotide polymorphisms (SNPs). Briefly, we crossed the *mrd1* mutant, identified in a *BonnMu* F_2-stock harboring 64 germinal *Mu* insertions, with the inbred line Mo17. The F_1 plants were subsequently self-pollinated to generate multiple F_2-mapping populations. We germinated the kernels from an F_2-population segregating for the *mrd1* phenotype in a paper

roll system and incubated them in a climate chamber with 16 h (28 °C, 2700 lux) of light and 8 h (21 °C) of darkness at 70% humidity. The primary roots of both *mrd1* mutants and their wild type siblings were harvested at the size of 0.5 cm. In total, we collected 85 mutant and 85 wild type individuals and pooled them separately according to their phenotypes. We isolated RNA from each pool using the RNeasy mini kit (Qiagen, Hilden, Germany) and controlled its quality by measuring the RNA integrity number (RIN > 8) on a Bioanalyzer 2100 with an RNA 6000 Nano chip (Agilent Technologies). Finally, we constructed the RNA-seq libraries and sequenced them on the NovaSeq 6000 platform, obtaining paired-end 150 bp reads.

To map the gene, we detected allele frequencies in the wild type pool and the *mrd1* pool according to Mansfeld and Grumet (2018). When there is complete linkage of a SNP marker with a mutated gene, only one marker allele is expected to appear in the *mrd1* pool of the mapping population. We scanned the genome using a smoothing window size of 3.5 Mb. Using BSA (Michelmore *et al.*, 1991), we determined the linkage probability for each individual SNP through a G‘ analysis (Magwene *et al.*, 2011) and plotted it against the genomic position (Mb). BSR-seq mapped the *mrd1* gene to an approximately 70 Mb interval on the short arm of chromosome 1. Raw sequencing data were deposited in the SRA under BioProject accession number PRJNA926107.

4.14 Anthocyanin profiling

We germinated a *BonnMu* F_2-family segregating for the *mrd1* mutant in a paper roll system as described in the method section 4.7. At 10 days after germination, we harvested whole primary roots from three seedlings per replicate. We used three biological replicates each for both the *mrd1* mutant and wild type siblings. The anthocyanins were isolated and measured as previously described (Tohge *et al.*, 2015; Cenk *et al.*, 2021; Pahlke *et al.*, 2021). Finally, quantification was conducted using reference curves for cyanidin 3-O-glucoside and peonidin 3-O-glucoside. Anthocyanin profiling was performed at the IPK in Gatersleben.

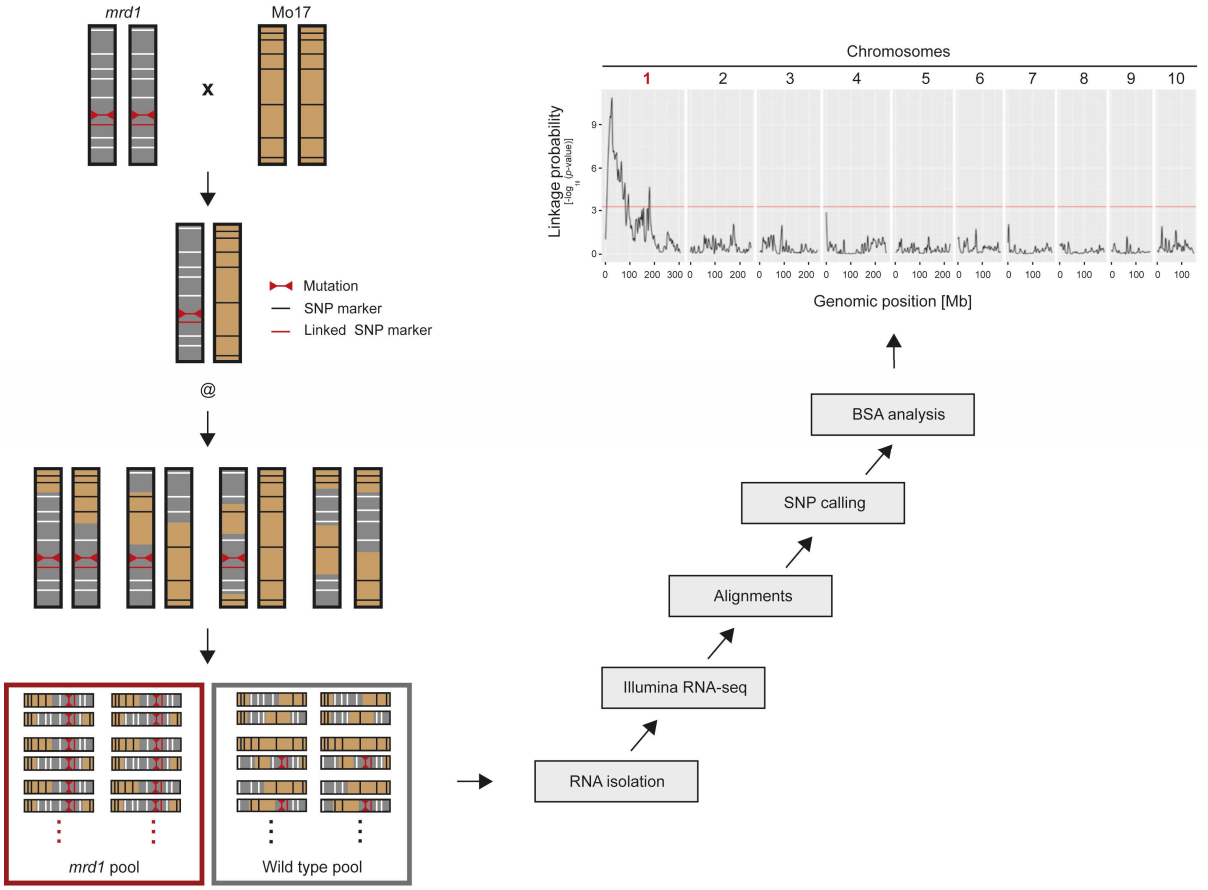


Figure 8. BSR-Seq workflow (Win *et al.*, 2024a). Generation of a mapping population by crossing an *mrd1* mutant in B73 background with a genetically distinct inbred line, Mo17. Mutation and linked single nucleotide polymorphism (SNP) marker are labelled in red on the simplified chromosome carrying the mutation, while additional polymorphic SNP markers are represented as black or white lines. After selfing the F_1-plants, *mrd1* mutants and wild types are respectively pooled from the segregating F_2-generation. For BSR-Seq analysis, RNA is isolated from bulks of mutant and wild type seedlings. The cDNA libraries for these mutant and wild type pools are constructed and sequenced using NovaSeq 6000 platform. Transcriptome profiling identifies polymorphic single nucleotide polymorphisms (SNPs) between the two samples to map the gene position on the chromosome. BSR-seq mapping narrowed down the candidate gene to an ~70 Mb interval on the short arm of chromosome 1. The genome was scanned using a smoothing window size of 3.5 Mb. The linkage probability (Mansfeld and Grumet, 2018) obtained by BSA for each individual SNP was determined by a G‘ analysis (Magwene *et al.*, 2011) and plotted against the genomic position (Mb). The mapping interval was used to prioritize a *Mu*-tagged candidate gene underlying *mrd1* mutant.

4.15 Rhizosphere microbiome analysis of the *mrd1* mutant

For microbiome profiling of the rhizosphere from crown roots, we conducted the experiment as described previously (Yu *et al.*, 2021). The kernels segregating for *mrd1* mutant were germinated in a paper roll system in a climate chamber. At 5 days after germination, we selected *mrd1* mutant and wild type siblings visually and then transplanted to soil-filled pots (7 × 7 × 20 cm^3) and grew in the climate chamber. We included four replicates each for both *mrd1* and wild type. Four pots without plants were included and treated the same for bulk soil samples. To ensure optimal growth, we watered the plants with distilled water every day until three days before harvesting. However, we did not treat them with additional fertilizers. We collected crown root and rhizosphere samples at 28 days after transplantation when the third whorl of crown roots has been initiated. To harvest the samples, we carefully removed the large soil particles from the whole root system while leaving the soil intact to the root which we defined as rhizosphere soil. We then selected the second whorl of shoot-borne crown roots at the length of about 7 cm from each replicate. For each biological replicate, one crown root for rhizosphere microbiome and another crown root for transcriptome analysis were taken, from wild type and mutant, respectively. The samples were placed in 15 ml reaction tubes and frozen immediately in liquid nitrogen. Samples were kept at –80 °C before use.

We extracted rhizosphere samples using PBS-Silwet washing method (Yu *et al.*, 2021). We isolated genomic DNA from bulk soil and rhizosphere samples using the Fast DNA Spin Kit for soil (MP Biomedicals, USA) and amplified the V4 region of the 16S rRNA gene for bacterial amplicon sequencing. We constructed sequencing libraries using the NEBNext Ultra DNA Library Prep Kit for Illumina and sequenced on an Illumina HiSeq 2500 platform, resulting in 250-bp paired-end reads. Using USEARCH v.11.0, we processed the sequences for quality, retaining only those with a mismatch ratio of less than 0.2. We assigned paired-end 16S rRNA amplicon sequencing reads to specific samples using unique barcodes and truncated them by removing the barcode and primer sequences. Using FLASH (v.1.2.7; Magoč & Salzberg, 2011), we merged the paired-end reads, naming the resulting sequences "raw tags." We applied specific quality filtering conditions to the raw tags to produce high-quality, clean tags as by the QIIME (v.1.7.0) quality control process (Caporaso *et al.*, 2011; Bokulich *et al.*, 2013). We employed the UCHIME algorithm (v.7; Edgar *et al.*, 2011) to match raw tags against the reference Gold database, identifying and removing chimera sequences to produce effective tags (Haas *et al.*, 2011). By using Uparse software (v.7.0.1001), we grouped sequences with 97% or greater similarity into the same Operational Taxonomic Units (OTUs) (Edgar *et al.*,

2013). We blasted each representative sequence of an OTU against the SSUrRNA database from the SILVA Database (release 128; Wang *et al*., 2007) using Mothur software for species-level annotation across various taxonomic ranks with a threshold of 0.8–1 (Quast *et al*., 2013). We removed OTUs identified as mitochondria or chloroplasts. We created an OTU table from the remaining sequences, keeping only those with more than two copies and removing unaligned sequences. We processed reads from the internal transcribed spacer amplicons by merging, de-multiplexing, filtering, de-replicating, and sorting them, keeping only those with at least two copies. Using UCLUST (v.7), we assigned sequences with 97% or more similarity to the same OTUs. We annotated taxonomic information for each OTU's representative sequence against the UNITE Database using a BLAST algorithm executed by QIIME (v.1.7.0; Kõljalg *et al*., 2013). We normalized bacterial OTU abundance to the sequence count of the least populous sample, followed by analyses of *β*-diversity and taxa differences among samples based on this normalized data. We employed the Phyloseq package (McMurdie and Holmes, 2013) to perform non-metric multidimensional scaling (NMDS) with two dimensions (K = 2) based on Bray–Curtis dissimilarity matrices to effectively visualize the differences among bulk soil, *mrd1* and wild type samples. Subsequently, we analyzed the relative abundance of enriched taxa in genus level in both wild type and *mrd1* mutant using Kruskal–Wallis and pairwise Wilcoxon tests ($p < 0.05$).

4.16 RNA-seq analysis of *mrd1* mutant

We used the crown roots of the plants grown in soil harvested within the rhizosphere microbiome experiment (see 4.2.3). There were four replicates for each genotype, wild type and *mrd1* mutant. Upon harvesting, we immediately froze the samples in liquid nitrogen, and later stored them at -80°C until further use. We ground the samples manually using mortars and pestles. Total RNA was isolated using the RNeasy Mini Kit (Qiagen, Hilden, Germany), following the manufacturer instructions. Subsequently, quality and integrity of the RNA was analyzed using a BioAnalyzer (Agilent RNA 6000 Nano Chip, Agilent Technologies). All the samples had a RIN value above 9.0. Finally, we conducted RNA sequencing on a NovaSeq 6000 platform, to obtain paired-end 150 bp reads. The raw sequencing data can be accessed using the SRA accession number PRJNA923594.

4.17 Analysis of differentially expressed gene

We processed the sequencing data with CLC GENOMICS WORKBENCH (v.23.0.4), adapting the methods from Osthoff *et al.*, 2019. After importing paired-end raw reads, we trimmed those

with a quality score below 0.05 and discarded reads shorter than 40 bp. The remaining reads were aligned to the B73 reference genome (*Zea mays*; Zm-B73-REFERENCE-NAM-5.0.57). We used R (v.4.3.1) for subsequent analysis. The raw reads were normalized by the cpm() function from the edgeR package. Using the limma package, we modeled the data, and shrank error variance empirically. We applied the estimateSizeFactors() from DESeq package to generate a principle component analysis (PCA) and visualized ggplot2. Outliers were detected by testing with Tukey's and Hubert's methods (Chen *et al.*, 2020). The differentially expressed genes were considered based on the FDR threshold of <0.05 and |log2 FC| ≥1. Gene Ontology (GO) analyses were performed by the R function TOPGO (Alexa and Rahnenfuhrer, 2023).

4.18 *mrd1* gene sequence analysis and phylogenetic tree construction

The predicted *mrd1* gene (Zm00001eb008060) has three transcripts according to MaizeGDB (https://www.maizegdb.org). The transcript Zm00001eb008060_T001 consists of 14 introns and 15 exons. We constructed the exon-intron gene model of *mrd1* gene (Zm00001eb008060) based on our Sanger sequencing result and the predicted model from MaizeGDB. The Blastp search of full MRD1 protein sequence was performed against *Arabidopsis thaliana*, *Brachypodium distachyon*, *Hordeum vulgare*, *Oryza sativa*, *Physcomitrella patens*, *Sorghum bicolor* and *Zea mays* in the Phytozome 13 (https://phytozome-next.jgi.doe.gov/blast-search). The retrieved homologous sequences were aligned using the MUSCLE algorithm in Mega10 (Kumar *et al*., 2018). We constructed the phylogenetic tree using the mixed model algorithm with four chains and 1 M generations using MrBayes (Ronquist & Huelsenbeck, 2003) and FigTree software (http://tree.bio.ed.ac.uk/software/figtree/).

4.19 qRT-PCR

To analyze the expression of *mrd1* in different genotypes, the *BonnMu* families segregating for *mrd1* (*BonnMu*-F7-3-F-1609, *BonnMu*-F7-3-F-1638, *BonnMu*-F7-4-F-1847 and 19-1013-05 (*Mu*^{4} per se)) were germinated in paper rolls. The plants were genotyped by PCR as described in the previous section (see 4.12). The primary roots of mutant, heterozygous, and wild type were harvested at the same developmental stage of 2-4 cm lengths. For each biological replicate, the primary roots from three plants were pooled, and each genotype was assessed with four biological replicates. All samples were frozen in liquid nitrogen and ground manually using mortars and pestles. To perform qRT-PCR, the total RNA was extracted from

each replicate using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). The RNA integrity was controlled and RNA concentration quantified using Bioanalyzer with RNA 7500 chip (Agilent Technologies). The complementary DNA (cDNA) was synthesized from 400 ng of total RNA (RIN > 8) using the qScript cDNA Synthesis Kit (Quantabio, Beverly, NJ, USA). The cDNA was diluted into 1:2 with nuclease free water before use. For the qRT-PCR reaction, a mixture containing 2 µL of PerfeCTa SYBR Green SuperMix (Quantabio, Beverly, NJ, USA), 1 µL of primer mix at a final concentration of 200 nM, and 1 µL of cDNA was prepared. The efficiency of each oligonucleotide was determined using a dilution series: 1, 1/2, 1/4, 1/8, 1/16, 1/32, 1/64, and 1/128. The *mrd1* transcript level was analyzed with respect to the expression of the housekeeping gene *homeobox-transcription factor 3* (*hb3*; Baer *et al*., 2022). Statistical analysis of gene expression levels was performed using a two-sided Student's t-test to determine significant differences. The primers used for qRT-PCR are listed in the Supplemental table S4.

4.20 Subcellular localization of MRD1

The subcellular localization of the MRD1 protein was examined using the 2in1 FRET vectors (Hecker *et al*., 2015) as described in Baer *et al.,* (2023). These vectors facilitate the study of co-localization and protein interactions, with one protein fused to tagRFP and another to mVenus. For nuclear localization control, HMGA (Launholt *et al*., 2006) was linked to tagRFP, paired with MRD1 fused to mVenus.

The *mrd1* gene's full coding sequence was PCR-amplified, purified, and cloned into pJet1.2/Blunt (Thermo Scientific). Following Sanger sequencing confirmation, entry clones were created using the BP-Clonase™ II enzyme mix (Invitrogen). LR reactions with these clones were performed using various pFRETvr-2in1 vectors (Hecker *et al*., 2015), followed by selection and verification of positive expression clones.

Transformed *Agrobacterium tumefaciens* with confirmed clones were used for transient expression in tobacco leaves. At three days post-infiltration, the expression and localization of MRD1 were analyzed via confocal laser scanning microscopy (Zeiss LSM 780 attached to an Axio Observer Z1). The fluorescence of TagRFP was excited using a Helium Neon laser at 543 nm, with its emission captured through a 597 to 641 nm bandpass filter. For mVenus, excitation was achieved at 488 nm using an argon laser, and its emission was collected using a 514 to 554 nm bandpass filter. Additionally, chlorophyll autofluorescence was excited at 488 nm by the same argon laser, though its emission was detected using a 686 to 711 nm

bandpass filter. Subsequent image processing was conducted using ZEN 3.4 software (Zeiss, Jena, Germany).

4.21 Validation of the *mrd1* candidate gene by CRISPR/Cas9

The *mrd1* gene has three transcripts in B73v5. We confirmed the sequences of transcripts Zm00001eb008060_T001 and Zm00001eb008060_T002 by PCR using the forward (5´-CAAAACCCTAGCTCGCCCTC-3´) and reverse (5´-CATCGACCGACAGCATGAGA-3´) primers and Sanger sequencing on cDNAs of B73 and B104 maize inbred lines. The transcript Zm00001eb008060_T001 consists of 15 exons. For genome editing via CRISPR/ Cas9, we targeted two regions of the *mrd1* gene using a dual single-guide RNA strategy (Pauwels *et al*., 2018). We designed and verified single guide RNAs (sgRNAs) on SYNTHEGO (https://tools.synthego.com/#/). The first pair of dual guides targeted a 441 bp region between exon 1 and intron 2. The second pair targeted a 592 bp region between exon 11 and exon 13 where the only PCI (Proteasome, COP9, Initiation factor 3) domain of the gene is located (Figure 9). We assessed the estimated specificity (Doench *et al*., 2016) and efficiency (Moreno-Mateos *et al*., 2015) of all sgRNAs using CRISPOR (http://crispor.tefor.net/).

Furthermore, we also verified the *mrd1* genomic DNA sequence of B104 at the CRISPR target sites using the primers CROPGEN962, CROPGEN963, CROPGEN964 and CROPGEN965 (Supplemental table S5). We next constructed sgRNA entry vectors and subsequently expression vectors using the shuttle vectors pGG-A-OsU3-BbsI-ccdB-CmR-BbsI-B and pGG-B-OsU3-BbsI-ccdB-CmR-BbsI-C with pGG-C-linker-G and the destination vector pG3HI-zCas9-tdT-AG (www.gatewayvectors.vib.be). The primers used for cloning gRNAs are listed in the Supplemental table S6.

The resulting expression vectors were transferred into the *Agrobacterium tumefaciens* strain EHA105 recA$^-$ (Rodrigues *et al.* 2020), containing the ternary vector pVS1-VIR2 (Zhang *et al*., 2019). Transformation of immature B104 embryos was carried out as described in Aesaert *et al*., (2021). We applied a genotyping approach according to Hunter (2021) to isolate the homozygous mutant plants. In brief, once we obtained T0 plants, leaf samples for gDNA isolation were harvested and genotyping were performed using the primers used for cloning of the sgRNAs. PCR products were confirmed by agarose gel electrophoresis, and sequenced via Sanger sequencing. Sequences were analyzed using TIDE (https://tide.nki.nl/) to track INDELS by Decomposition. The T0 plants with an edited *mrd1* gene were backcrossed with wild type B104. The presence/ absence of the CRISPR/Cas9 T-DNA in the F_1 plants were confirmed by PCR using the primers CROPGEN214 and CROPGEN215, targeting the Cas9 transgene. For the plants with Cas9 present, additional backcrossing to B104 was performed. The remaining plants, which are null segregants without Cas9 but with the edited *mrd1* gene were self-pollinated. Subsequently, we conducted phenotyping and genotyping of F_2 plants, confirming INDELs.

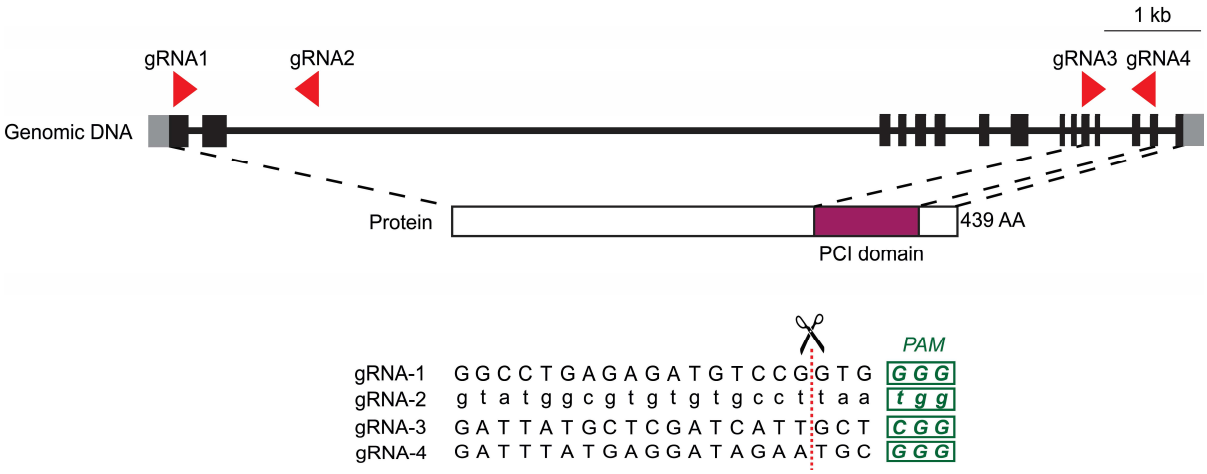


Figure 9. CRISPR/Cas9 target sites in the *mrd1* gene. This figure shows the gene and protein structure of the *mrd1* gene and the MRD1 protein. The regions targeted for CRISPR/Cas9-mediated genome editing using a dual single-guide RNA (sgRNA) approach are highlighted. Two distinct pairs of sgRNAs (gRNA1/2 and gRNA3/4) are designed to target upstream of the protospacer adjacent motifs (PAMs, red boxes) in two gene regions. The first sgRNA pair targets a 441 bp intergenic region between exons 1 and intron 2, while the second pair is directed at a 592 bp region where PCI domain is located between exons 11 and 13. The target sequences for each sgRNA are provided, with PAM (protospacer adjacent motif) sequences essential for Cas9 binding indicated in green boxes and cut sites with red dotted line. Exons are represented by black rectangles and introns are represented by connecting solid lines.

5 Results

5.1 Expanding the *BonnMu* sequence-indexed repository of transposon induced maize (*Zea mays* L.) mutations in Dent and Flint germplasm

5.1.1 *BonnMu* insertions cover 83% of all annotated B73v5 gene models

The *BonnMu* F_2-families were generated in different genetic germplasm backgrounds, i.e., B73 and Co125 from the Dent pool, and F7, EP1, and DK105 from the Flint pool (Table 1). In a previous study, we generated two Mu-seq libraries comprising 1,152 *BonnMu* F_2-families in B73 background (Marcon *et al.*, 2020). For the downstream bioinformatic analysis, we employed the Mu-seq method described by McCarty *et al.*, 2013 and Liu *et al.*, 2016. Here, we complement the two previous Mu-seq libraries by 12 additional libraries comprising 6,912 *BonnMu* F_2-families. To ensure equal analysis of all 14 datasets, we applied the bioinformatics method MuWU (Stöcker *et al.*, 2022) and integrated the first two libraries (Marcon *et al.*, 2020) into the analysis. Sequencing of 14 Mu-seq libraries yielded 1,59 billion raw read pairs (Table 7). After automated trimming of U-adapter and TIR sequences, 74% (1,17 billion read pairs; Table 7) of the read pairs remained. Among the remaining read pairs 90% were aligned to the B73 genome containing 44,303 gene models (Zm-B73-REFERENCE-NAM-5.0). After duplicate read removal, more than 235 million reads, i.e., read pairs and unpaired reads, remained and were used to identify germinal insertion sites at intersections of one row and one column pool. Subsequently, reads were counted in each of the 48 pools per library for insertion site identification. The MuWU analysis exclusively considered insertion sites supported by a minimum of four reads for subsequent analyses. In total, 425,924 distinct germinal insertion sites were detected in the 14 *Mu*-seq libraries tagging 36,612 (83%) of the 44,303 B73v5 genes (Table 7). In detail, we counted insertions affecting genic regions, defined as from the start of the 5’ UTR to the end of 3′ UTR of genes, including exons and introns. Additionally, *Mu* insertions in promoter regions – up to 2,100 bp upstream of the start of the 5' UTR – and close downstream regions of genes – up to 2,100 bp downstream of the end of the 3' UTR – were considered. Based on the number of 425,924 insertion sites, each of the 36,612 tagged genes carried on average 12 insertional alleles (425,924 insertion sites / 36,612 affected genes). The majority of 60% of the affected genes (21,812 of 36,612 B73v5 genes) harbored insertions in their coding sequence. The reverse genetic data are deposited on the Crop Functional Genomics website at (https://www.inres.uni-bonn.de/cfg/en/c-f-g/research/bonnmu) or on MaizeGDB (MaizeGDB.org). Among the 8,064 analyzed *BonnMu*

F_2-families, 98% (7,908 of 8,064 F_2-families) carried at least one germinal insertion. Only for a minority of 2% of the *BonnMu* F_2-families (156 of 8,064 F_2-families), no insertion was detected. Hence on average, each *BonnMu* F_2-family carried 53 heritable *Mu* insertions (425,924 insertion sites/ 8,064 F_2-families). An extreme example is the *BonnMu* F_2-family F7-4-F-1766 hosting 338 distinct *Mu* insertions in 457 different genes.

The genomes of the European Flint lines DK105, EP1, and F7 were recently sequenced (Haberer *et al*., 2020). In light of this, we explored the number of germinal insertions sites by mapping the Mu-seq reads of the 3,456 sequenced *BonnMu* F_2-libraries in DK105, EP1, and F7 genetic background (Table 1) to their respective genomes. If only genes that can be assigned to chromosomes are considered, there are 46,726, 43,375, and 44,043 genes in the DK105, EP1, and F7 genome, respectively (Haberer *et al*., 2020). Among the 462 *BonnMu* F_2-families in DK105 genetic background (Table 1) the MuWU analysis identified 29,986 unique insertions affecting 36% of the DK105 genes (16,704 of 46,726 genes; Win *et al*., 2024b). Furthermore, 57,806 distinct insertions in 53% of the EP1 genes (22,830 genes of 43,375) were detected, when the Mu-seq reads of 690 *BonnMu* F_2-families in EP1 genetic background were mapped to the EP1 genome (Win *et al*., 2024b). Finally, the 2,304 analyzed *BonnMu* F_2-families in F7 genetic background (Win *et al*., 2024b) carried 64,839 unique insertions affecting 42% of the F7 genes (18,548 genes of 44,043 genes of the F7 genome; Win *et al*., 2024b).

Table 7. Summary of Mu-seq libraries alignment to B73v5.

Description	14 Mu-seq libraries*
Raw read pairs	1,589,204,597
Read pairs after trimming	1,173,003,813
Average alignment rate	90%
Read pairs and unique unpaired reads after removing duplicates	235,240,526
Number of germinal *Mu* insertions (unique insertions)	425,924
Number of *Mu*-tagged genes (unique genes)	36,612

* Two previously published Mu-seq libraries (Marcon *et al.,* 2020) were included in the analysis.

5.1.2 *BonnMu* affected genes in Flint germplasms complement the set of tagged genes identified in Dent lines

Based on the analysis of 14 Mu-seq libraries in different genetic backgrounds, uniformly mapped to B73v5, insertions affected 36,612 genes (Table 8; Win *et al*., 2024b). The numbers of *Mu*-tagged genes in the five different germplasms varied from 8,396 mutated B73v5 genes in Co125 genetic background to 32,390 B73v5 genes in B73 genetic background (Figure 10A; Win *et al*., 2024b). This bias can be partially attributed to the fact that the Mu-seq analysis was conducted on only 576 mutagenized F_2-families in the Co125 genetic background, whereas it was performed on 4,032 mutagenized F_2-families in the B73 genetic background. Of the 36,612 affected genes, 5,502 (15%) were detected in *BonnMu* F_2-families of all 14 Mu-seq libraries used in this study (Figure 10A). The number of overlapping genes significantly exceeded the expected count of 975 genes by chance (Table 8). This finding indicates the preference of *Mu* transposons for tagging specific genes across diverse inbred lines.

A higher number of 8,800 overlapping genes (24%) were identified in *BonnMu* F_2-families of four different mutagenized germplasms: DK105, EP1, F7, and B73. A considerable proportion of *Mu*-tagged genes, that is, 16,827 (46%) of 36,612 were hit in *BonnMu* F_2-families of Mu-seq libraries in two or three different genetic backgrounds under analysis (Figure 10A). In summary, 85% of the *Mu*-tagged genes were identified across at least two different genetic backgrounds. This result further indicated that specific genes are more prone to *Mu* transposon insertions and are affected consistently across multiple genetic backgrounds. Finally, the remaining 5,483 genes (15%) of the affected 36,612 genes were exclusively detected in one of the mutagenized inbred lines. More precisely, the majority of 2,460 genes were uniquely identified in the Mu-seq libraries in the B73 genetic background, whereas only 58 of the tagged genes were uniquely detected in the single Mu-seq library in Co125 genetic background (Figure 10A).

Among the 36,612 tagged genes, 4,027 (11%) were identified in at least one Mu-seq library of the mutagenized Flint lines, i.e., DK105, EP1, and F7, but were not tagged in the set of affected genes identified in the Dent lines B73 and Co125. Almost half of these genes, specifically 1,969 of 4,027 (49%), were exclusively detected in the Mu-seq libraries in the F7 background (Figure 10). This result can be partially explained by the fact, that 2,304 *BonnMu* F_2-families in F7 background were used for the Mu-seq experiment, whereas only 462 and 690 *BonnMu* F_2-families in DK105 and EP1 background were analyzed, respectively (Table

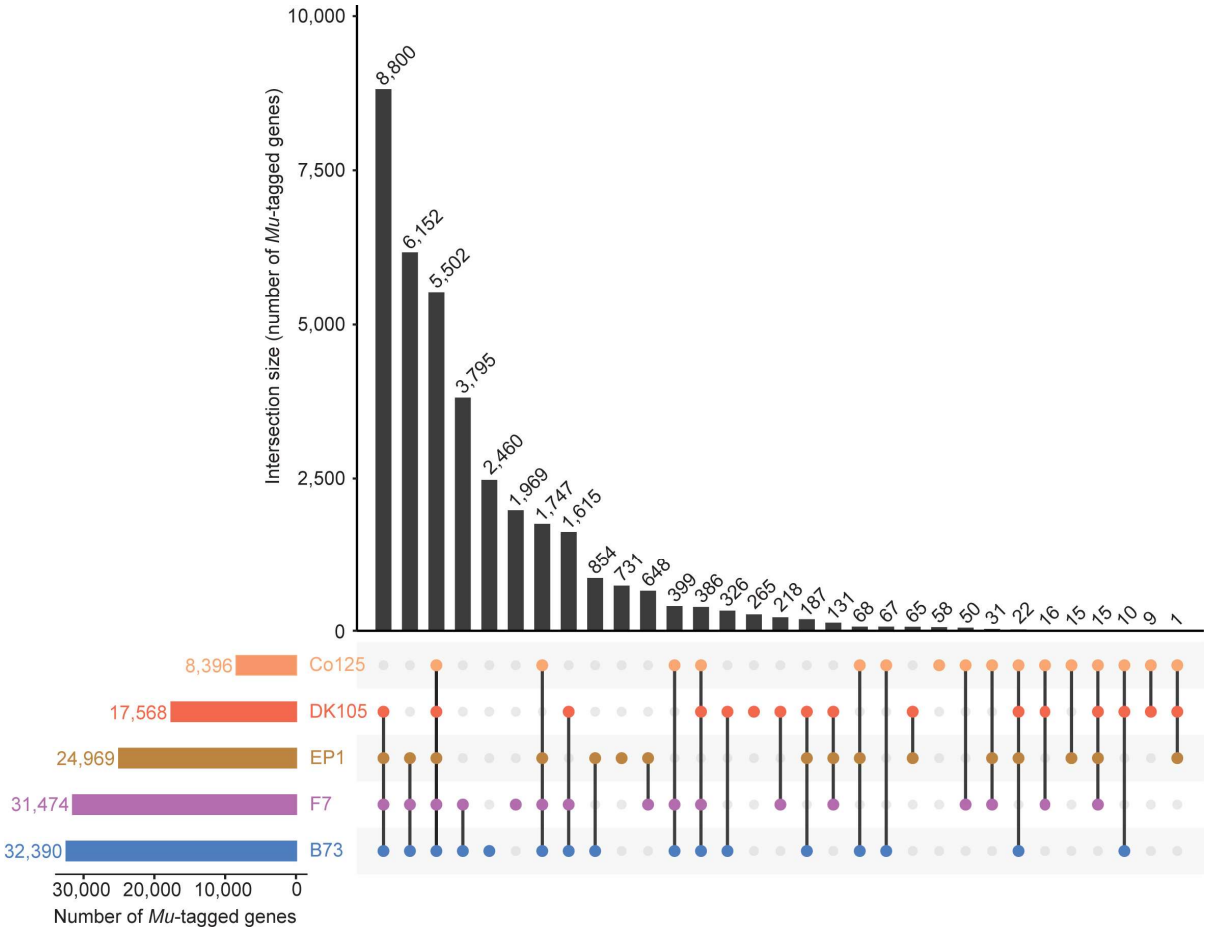


Figure 10. Overlap of genes affected by *Mu* insertions and distribution of insertions. Intersections of genes, tagged in *BonnMu* F_2-families of the two Dent lines B73 and Co125, and three Flint lines DK105, EP1, and F7, which have been mutagenized in this study. The UpSet plot displays 31 intersections. The lines connect overlapping genes among different genetic backgrounds. The total number of intersected genes is displayed above each bar.

1). Nevertheless, a considerable portion of 265 (7%) and 731 (18%) genes were uniquely tagged in *BonnMu* F_2-families in DK105 and EP1 background, respectively (Figure 10A). The remaining 1,062 of the affected genes (26%) were overlapping in *BonnMu* F_2-families of two or three of the Flint lines. Consequently, the genes affected by *Mu* insertions in Flint germplasms complement the set of tagged genes identified in dent lines.

Table 8. Expected and observed *Mu* tagged genes among the *Mu* insertional libraries in *BonnMu* inbred lines.

	***BonnMu* genetic background**					**Overlap**	
	B73	**Co125**	**DK105**	**EP1**	**F7**	**Expected[a]**	**Observed**
Total number of *Mu* tagged genes	32,390	8,396	17,568	27,969	31,474	975*	5,502

[a] The expected overlap of *Mu*-tagged genes in five *BonnMu* genetic backgrounds was calculated using a generalized linear model with a Poisson distribution and a log-link function in R. The initial model considered each genetic background (B73, Co125, DK105, EP1, F7) as independent for the expected values, then all possible interactions were included in a second model. A χ^2 test via the anova() function was conducted between the expected and observed numbers of *Mu*-tagged genes by comparing both models. Differences between expected and observed values were significant at $\alpha = 0.05$ (* = $p < 0.001$), marked by an asterisk.

5.1.3 The number of *BonnMu* insertions exhibit a weak to moderate correlation to gene length

We examined the distribution of all 425,924 *Mu* insertions within the 36,612 B73v5 genes, including insertions in promoter regions, within a 2,100 bp window upstream of genes and nearby downstream regions, within a 2,100 bp window downstream of genes. Among the tagged genes, nearly half of them, i.e. 49% (17,958 of 36,612) harbored 1-10 *Mu* insertions, 24% (8,908 of 36,612 genes) contained 11-20 insertions, and the remaining 27% (9,746 of 36,612) of the genes carried at least 21 insertions (Figure 11B). Among the latter group of genes 12% (1,160 of 9,746) carried at least 50 insertions. One extreme example is a 5,938 bp gene encoding a protein-serine/threonine phosphatase, Zm00001eb054350, carrying 303 unique insertions. Among these insertions, 38% (115 of 303 insertions) hit the coding sequence of the gene Zm00001eb054350 (Win *et al*., 2024b). Subsequently, we tested whether the number of *Mu* insertions was positively correlated with the length of the affected genes. To this end, we calculated the mean length of the affected genes which were grouped according to the number of insertions (Figure 11B). The five groups of genes harboring 1-50 insertions showed a comparable gene length, ranging between 7,931 bp and 10,321 bp, whereas the group of genes harboring >50 insertions exhibited an increased length of 12,595 bp. Consequently, the computed Pearson correlation across all groups of gene sizes showed

a weak positive correlation between gene size and the number of *Mu* insertions (r = 0.114; Figure 11C).

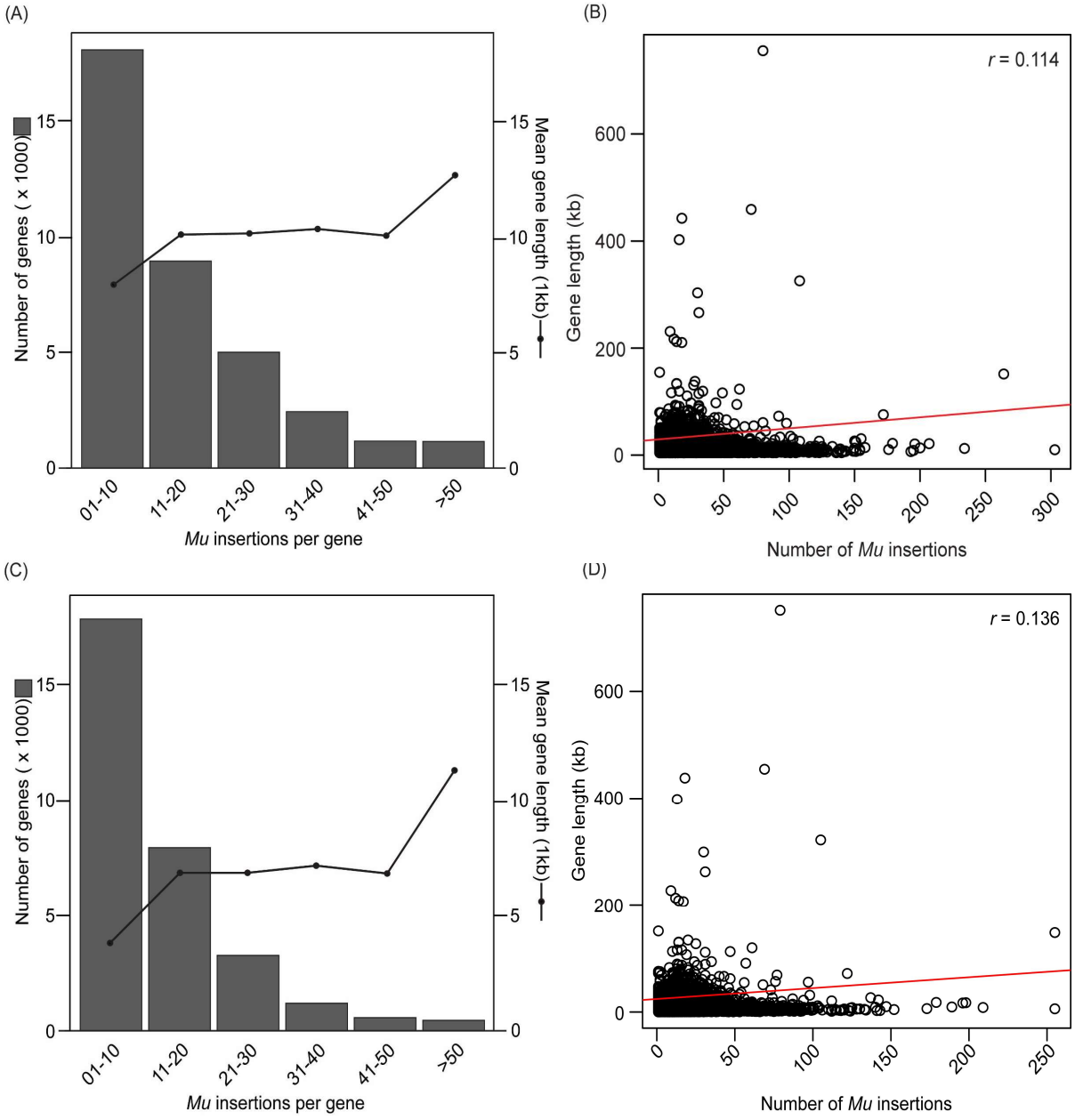


Figure 11. Distribution of *Mu* insertions against gene length. (A) Number of tagged genes and associated mean gene length plotted against the number of *Mu* insertions. (B) Distribution of the length of affected genes plotted against the number of individual *Mu* insertions. The calculated Pearson correlation coefficient is r = 0.114 (p<0.001). (A) and (B) consider all insertions within the gene body as well as promoter regions defined by a 2,100 bp window upstream and similar downstream regions. (C) Number of tagged genes and associated mean gene length plotted against the number of *Mu* insertions when only insertions in the genic regions, i.e., in 5' and 3' UTRs, exons and introns of genes were considered. (D) Distribution of affected gene lengths plotted against the number of individual *Mu* insertions. The calculated Pearson correlation coefficient is r = 0.147 (p<0.001).

We obtained consistent results when considering 87% of the insertions (370,396 of 425,924) affecting only genic regions, i.e., 5’ and 3’ UTRs, exons, and introns, (Figure 11C & 11D). These insertions covered 85% of the total genes (31,126 of 36,612). In this analysis, we calculated a moderate positive correlation of $r = 0.136$ between gene size and the number of insertion sites (Figure 11D).

5.1.4 *Mu* insertions preferentially target the 5’ UTR of genes

Next, we investigated whether *Mu* transposons exhibit preferences for insertion sites within the maize genome. Maize has a complex genome comprising over 80% repetitive intergenic non-coding sequences (Haberer *et al*., 2020; Hufford *et al*., 2021; Chen *et al*., 2023). As a result, coding sequences constitute less than 20% of the genome. Specifically, the B73v5 genome can be subdivided into gene coding regions including 5’ UTRs (0.5%), exons (2%), introns (5%), 3’ UTRs (1%) and promoter regions (3.5%; Supplemental table S7) which are located upstream of the 5’ UTRs of genes. We further divided the promoter region into a core promoter (0.2%) and a proximal promoter (3.3%), hereafter referred to as promoterCore and promoterProx, respectively (Supplemental table S7). While the promoterCore, including the transcription start site, is located directly (1 – 100 bp) upstream of the start of the 5’ UTR, the promoterProx is 101 – 2,100 bp upstream of the 5’ UTR. Consequently, gene-coding regions and their associated promoter segments account for only 12% of the total maize genome (Figure 12A). The remaining 88% is composed of non-coding repetitive sequences (Figure 12A).

To explore whether *Mu* transposons exhibit preferences for distinct categories of the coding regions or the intergenic region of the maize genome, we investigated the set of all *BonnMu* insertion sites identified in the subset of seven Mu-seq libraries consisting of 4,032 *BonnMu* F_2-families in B73 genetic background (Win *et al*., 2024b). Among all 774,692 somatic (not shown) and germinal insertions, a considerable proportion of 43% (331,169 insertions) affected the 5’ UTRs (Figure 12B). A comparable fraction of 15% (115,786 insertions) and 12% (95,371 insertions) tagged exons and introns of genes, while 13% (98,510 insertions) were incorporated in promoterProx sections of the maize genome. Minor fractions of insertions were identified in 3’ UTRs (4%; 27,189 insertions) and promoterCore regions (7%; 53,950 insertions). Interestingly, while 88% of the maize genome contains non-coding regions, only 6% (52,718 insertions) of the *BonnMu* insertions were detected in those regions (Figure 12B). In summary, these results indicate that intergenic insertions are underrepresented and genic regions, such as the 5’ UTRs of genes, are frequently targeted by *BonnMu* insertions.

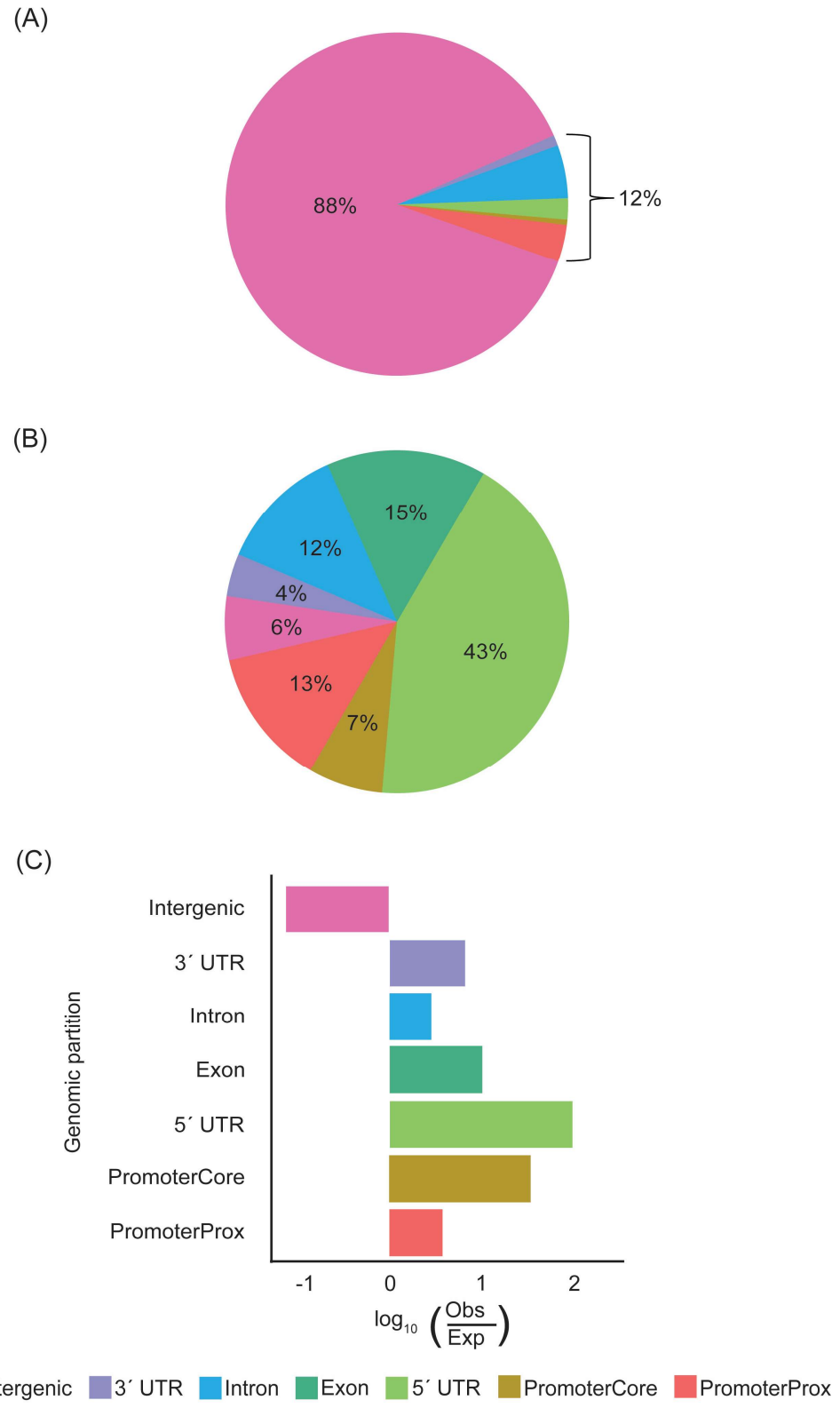


Figure 12. Distribution of *BonnMu* insertions across the maize genome. (A) Composition of the maize genome (B73v5). (B) *Mu* insertion sites across the genome. (C) Ratio of observed and expected *Mu* insertions across the genome.

To further support this finding and to account for the non-uniformity of intergenic *versus* genic space, we calculated the ratio between the number of observed and expected insertions per genomic partition using Pearson's χ^2 tests with Yates' continuity correction. Indeed, we observed more insertions than expected for all genic and promoter partitions of the genome (Supplemental table S7; Figure 12C). For the 5' UTRs a number of 3,824 insertions would be expected based on the genomic composition of the maize genome (Supplemental table S8). However, we discovered significantly more than expected, specifically 331,168 insertions, exceeding the anticipated count by over 86-fold. In contrast, we expected 684,938 *BonnMu* insertion sites in intergenic regions. However, only 52,718 of such insertion sites were detected (Supplemental table S8; Figure 12C), indicating approximately 13 times fewer than expected.

5.1.5 *BonnMu* insertions correspond with gene-dense telomeric regions

Previous studies described that *Mu* insertion site frequencies align with gene density (Schnable *et al*., 2009; Springer *et al*., 2018). Thus, we analyzed the distribution of *BonnMu* insertions in the B73 genetic background across all 10 chromosomes of maize by dividing each chromosome into 10k bins of 213,167 bp in size and counted the number of insertions per bin. At the heterochromatic centromeric regions of each chromosome, we predominantly detected 200 kbp bins containing less than 200 *BonnMu* insertions (Figure 13). In contrast, at the telomeres we frequently detected 250 – 500 insertions (Figure 13), indicating that *Mu* elements preferentially insert into gene-rich regions. However, there are some exceptions, i.e., bins which are in close proximity to the telomeres of the chromosomes 5-8, but lack insertions. One extreme example is the telomeric region at the short arm of chromosome 6, illustrating a 5-6 Mb window lacking insertions (Figure 13). We also observed that the distribution of *Mu* elements aligns with the chromosomal gaps in regions of chromosomes 5-8. We investigated this using the MaizeGDB JBrowse genome browser (Woodhouse *et al.*, 2021) and can conclude that these gaps correspond to highly heterochromatic knob regions located at the telomeric regions of these chromosomes (Ghaffari *et al.*, 2013).

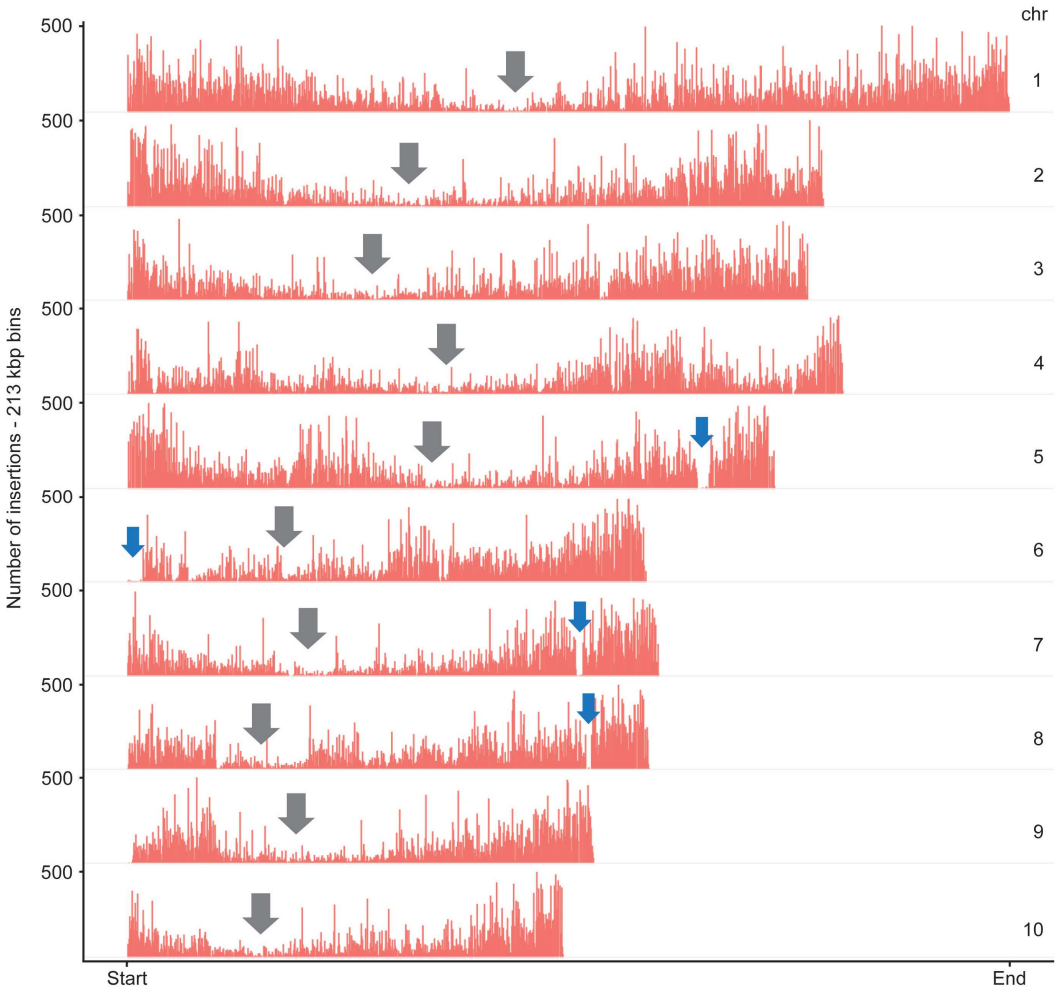


Figure 13. Distribution of *BonnMu* insertions across ten chromosomes. *Mu* insertions are predominantly found in the gene-rich telomer regions with fewer insertions in the heterochromatic centromere regions. Grey arrows indicate the centromeres. Blue arrows indicate gaps, i.e., regions with marginal numbers of *BonnMu* insertions, on the arms of chromosomes 5-8. These regions correspond to highly heterochromatic knob regions (Ghaffari *et al.*, 2013).

5.1.6 Spatial distribution of *BonnMu* insertions and chromatin accessibility across the maize genome

To gain a deeper understanding of the distribution patterns of *BonnMu* insertions within the genome, we investigated the relationship of the *BonnMu* families in B73 background to chromatin accessibility and histone modifications. To accomplish this, we utilized multiple ATAC-seq (Assay for Transposase-Accessible Chromatin using Sequencing) analyses, ChIP-seq data, and published datasets related to DNA acetylation and methylation processes. ATAC-seq is used to assess open chromatin regions on a genome-wide scale (Buenrostro *et al*., 2013). In this study, we visualized the chromatin accessibility within the maize NAM (Nested Association Mapping) population and various tissues, such as ear and leaf (Figure 14; Ricci *et al*., 2019). Additionally, the comparative analysis compared the frequency of *Mu* insertion sites with the genome-wide distribution of chromatin modifications, including trimethylation of Lys-27 of histone H3 (Makarevitch *et al*., 2013) and histone acetylation (Figure 14; Zhang *et al*., 2015).

In a 250 bp window around the midpoint of the maize gene models, we detected strong signals of unmethylated regions (UMRs; NAM_UMRs, Figure 14) which is in line with the high overlap of UMRs with accessible chromatin regions reported previously (Hufford *et al*., 2021). Similarly, distinct central enrichment at the gene model midpoints was observed for histone 3 modifications, such as trimethylations at Lys-4 and Lys-36 (H3K4me3_Leaf and H3K36me3_Leaf) or acetylation of Lys-56 (H3K56ac). The frequency of these histone modifications gradually diminishes in 250 bp windows both up- and downstream of the gene midpoint.

A contrasting pattern was observed for chromatin accessible signals, based on ATAC-seq datasets, and *BonnMu* insertion features. While gradually increasing frequencies of ATAC-seq signals and *Mu* insertions were identified in 250 bp windows flanking the gene midpoint, there are only a few such signals and insertions in the center of the gene. Hence, *BonnMu* insertions aligned well to transposase accessible chromatin signals detected in the following datasets: ATAC_Ear, ATAC_Leaf, and NAM_ATAC (Figure 14). According to Figure 14 *Mu* transposons predominantly insert at the start or end of genes, including UTRs and closely adjacent regulatory sequences. This finding is, at least in part, in line with the preference for *BonnMu* insertions targeting promoterCore and 5’ UTR regions of genes (Figure 12B). Aligning partly with the pattern observed for the *BonnMu* insertions, H3K27me3 modifications displayed tissue-specific differences in their distribution around gene midpoints. This can at

least partly be explained by the reported observation that H3K27me3 seems to be less coupled to chromatin accessibility than other modification – which on average deviate from open chromatin signals only in 15-21% of cases in a tissue-specific manner in maize (Ricci *et al.*, 2019).

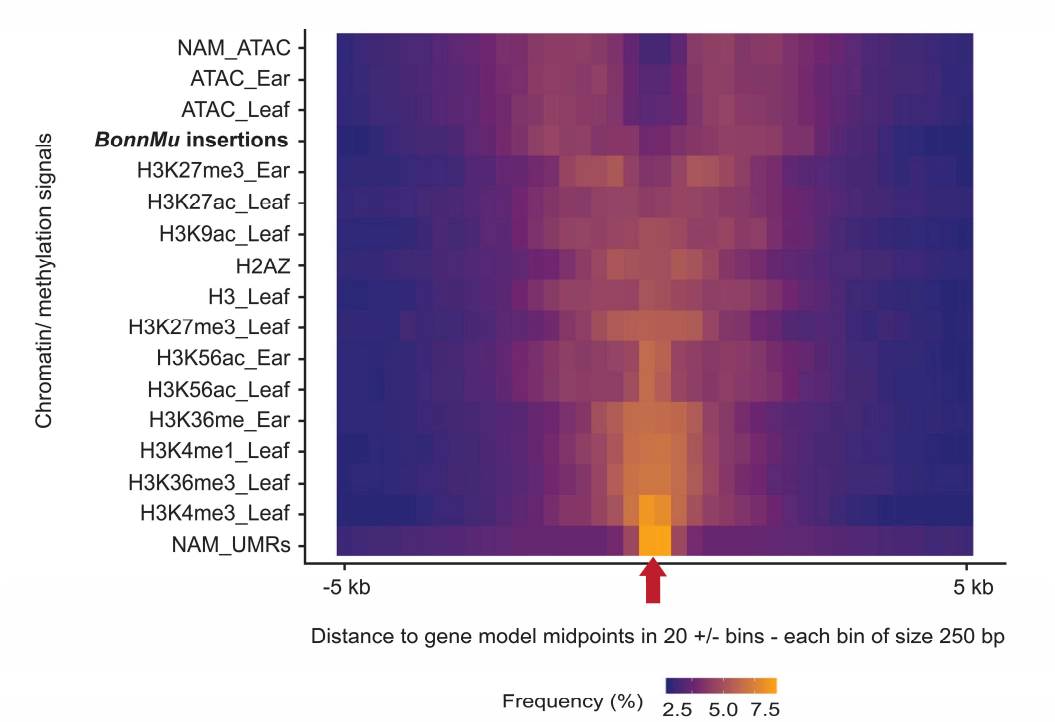
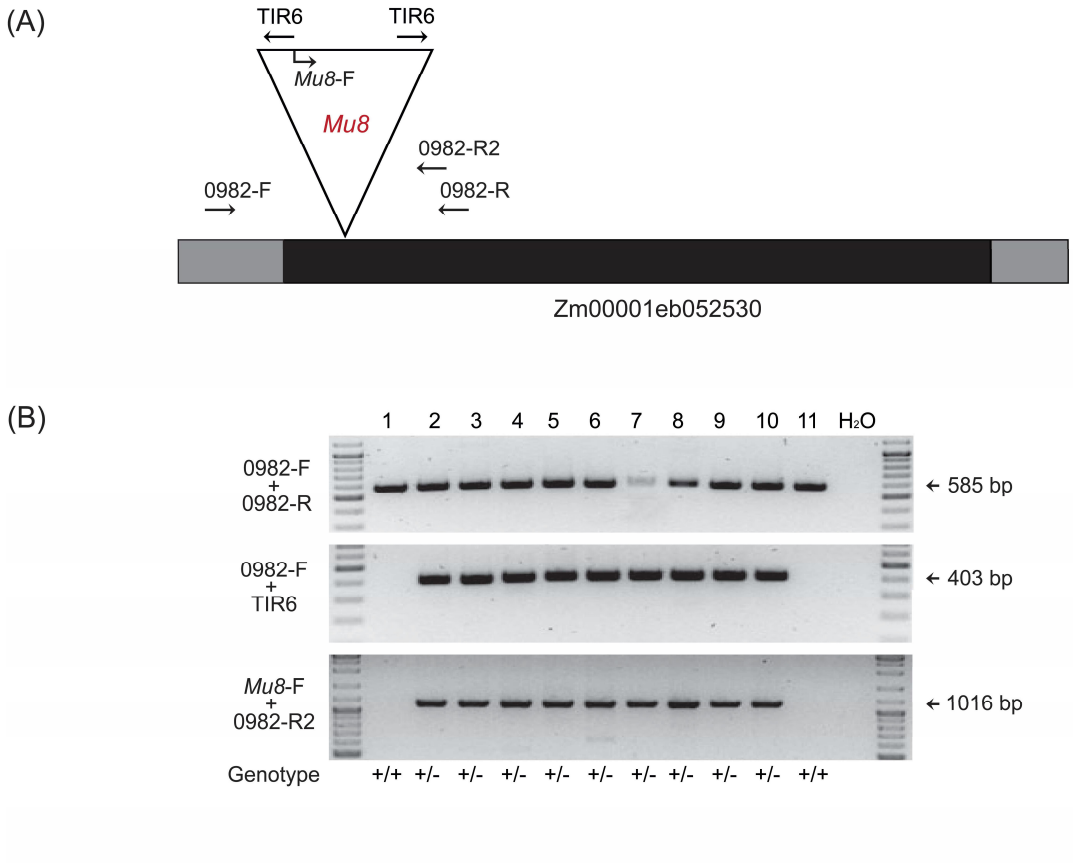


Figure 14. Epigenomic landscape surrounding *BonnMu* insertions in maize. The heatmap illustrates the frequency distribution of *BonnMu* insertions, chromatin modifications and chromatin accessibility in relation to the entire set of genes in the maize genome. The red arrow on the horizontal center of the plot indicates the midpoint of all maize gene models. Bins of 250 bp in size are presented on both sides of this midpoint, representing a 5 kb upstream and a 5 kb downstream region of the gene midpoint. The frequency of *Mu* insertions, chromatin marks or accessibility signals is color coded (yellow = higher frequency; blue = lower frequency). NAM, nested association mapping; UMR, unmethylated regions; H3K27me3, trimethylation of lysine 27 on histone H3; H3K27ac, acetylation of lysine 27 on H3; H3K9ac, acetylation of lysine 9 on H3; H3, histone H3; H3K56ac, acetylation of lysine 56 on H3; H3K36me, methylation of lysine 36 on H3; H3K4me1, monomethylation of lysine 4 on H3; H3K36me3, trimethylation of lysine 36 on H3; H3K4me3, trimethylation of lysine 4 on H3; ATAC, Assay for Transposase-Accessible Chromatin.

5.1.7 Mapping and validation of *Mu1* and *Mu8* transposons

The *Mu* transposon system is a powerful tool for large-scale mutagenesis in maize. Several non-autonomous *Mu* species have been comprehensively characterized (Lisch, 2015). In our study, we pinpointed the potential *Mu* species at the 425,924 distinct germinal insertion sites, tagging 36,612 B73v5 genes (Table 7; Win *et al*., 2024b). To identify these potential *Mu* species, we associated the *Mu*-TIR sequence which was part of each Mu-seq read (i.e., flanking the gene sequence of interest), to a list of known Mu species. Due to the highly conserved nature of the TIRs in all *Mu transposons* and the presence of non-specific, short TIR fragments in the Mu-seq reads, our reliable confirmation was limited to *Mu1*, *Mu8*, or *MuDR* transposons. Overall, we identified 82,285 (14.9%) *Mu1*, 1,211 (0.2%) *Mu8*, 80,347 (14.5%) *Mu8*|*MuDR* species. For the majority of 390,124 (70.4%) of the 425,924 insertion sites, we could not identify the respective *Mu* species.

For validating *BonnMu* insertions and corresponding *Mu* species, we randomly selected insertions in three distinct genes: (1) Zm00001eb052530 carrying a *Mu8* insertion in the F_2-family *BonnMu*-2-A-0982, (2) Zm00001eb280980, and (3) Zm00001eb030640 harboring a *Mu8* and *Mu1* insertion in the mutagenized F_2-families *BonnMu*-7-C-0459 and *BonnMu*-F7-2-F-1001, respectively. According to our *BonnMu* dataset, following insertion identifiers, i.e. distinct 7-digit numbers, identifying unique insertions, were assigned to the three insertions: (1) *BonnMu*0031087, (2) *BonnMu*0170576, and (3) *BonnMu*0446992. The insertion identifier *BonnMu*0031087 indicates a *Mu8* insertion in the single exon of the gene Zm00001eb052530 (Figure 15A), located 157 bp downstream of the A from the ATG start codon (Win *et al*., 2024b). A PCR-based co-segregation analysis of 11 individual plants of the segregating F_2-family *BonnMu*-2-A-0982, identified two plants (# 1 and # 11) being homozygous for the wild type allele. Mutants were identified as heterozygotes, so both wild type and mutant bands were observed (# 2 - # 10; Figure 15B). Sanger sequencing (Sanger *et al*., 1977) of the *Mu*-specific PCR products confirmed that the *Mutator* insertion in this gene was caused by a *Mu8* element (*Mu8*; Figure 15C).



(C)

```
_ _ _CCTACTCTCCCACGGTCTCTGCACGTCCCAAATGGCAAAAACGGACTGCGTAGTGC
GGATACAACTTGGCTACCCAAGGCAACGAACCCAGAGGAGATAATTGTCATTATAGACGA
AGAGCGGACGGGATTCGACGAAATAGAGGCGATGGCGTTGGCTTCTCTATCTTGCTGAAA
AACTCGAGCCATC_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
GCCGAGTTCTGGACGATGACGGAGATGCGTTCCCCGATGGCGGAGCGCACGTTCTGGGCG
TCCAGCGCGAGCCTGGCGGCCACGCGCGCGCTGGCGTTCTCGTCCGTGTCGAACCAGGCG
ATCTCGTTGCGGAGCACGGCGGCGAACATCTTCTCGCGCACCCGCTTGGTCAGGTTCTCG
CCCACCGTGTCCCAGAACACGTGCTGCACCGTGTTGAACAGCAGCGCCGCGGAGGACATG
CCGATGAGCAGGTAGCAGTATTTGGCGATCTCGCGCTTCATGTACCGCGGGTCCGGCGCG
TAGTACACGCTGAGCACGGCGCTGAGGATGTAGGCGAAGATGGCGCTGAAGGAGCCGCGG
ACCATGGAGCCGGCGAGCGCGTAGGCCCACTCGGGCGAGTTCATCCTGGCGAGGCGCAGG
AAGGAGCTGGCGCCGGCGCGGAACGCCAGCTGCTTGTCGGCCATGGTCCGGTGGTGGTGG
TGCGGGTCGTGGATGGAGAGGGTGAAGTCGGAGGTGGAGAAGTCGGAGAGGCGGCGGGAG
TAGGGGGAGCGGCCGTAGGAGGAGTTGCGCGTCATGATGGGCGAGCTGACGGAGTTGCGG
GCGCTGGAGGGCCTGGCGCTGCGCGCGTGCTGGTGCAATGTCGACCCCGAGAGCATGAAC
ACGAGAGCATGAACACGAAACGGCGGCTAGGGCAGCGTCTGCGCAGACGAACGGTAAACG
GGGACAGGAGACTAATACCTGTTTACGCAGCTCCAAGCGCTGTCTTCTTTCCGTTTTGGC
GTTTGGCTGTCGCCTGCGTCTCCAGAACAAGAGAAGCCAACGCCATCGCCTCCATTTCGT
CGAATCCCGTCCGCTCTTCGTCTATAATGAGAGAAGCCAACGCCATCGCCTCCATTTCGT
CGAATCCCGTCCGCTCTTCGTCTATAATGGCAATTATCTCCAGAGAGAGCACCAAACCGC
ACCGAAAAGCCAACACCCCCGGCCCGCGCGCCGGCGTCCTCCTATGCTCGCCGTCG_ _ _
```

Figure 15. PCR-based segregation analysis. (A) Structure of the gene Zm00001eb052530. The single exon is illustrated as a black box and UTRs as gray boxes. The *Mu8* insertion in the exon is shown as a triangle. Gene- and TIR-specific primer sites are indicated as arrows. (B) PCR segregation analysis of 11 individual plants of the segregating F_2-family *BonnMu*-2-A-0982. **(continued to page 47)**

(Figure 15. continued) Gene-specific primers (0982-F + 0982-R) flanking the insertion site were combined to detect the presence of a wild type copy of the gene. Additionally, one gene-specific primer along with a TIR-specific primer (0982-F + TIR6) were used to test for the presence of an insertion in the gene. To confirm the *Mu8* insertion, a combination of a *Mu8*-specific primer and a gene-specific primer (*Mu8*-F + 0982-R2) was used. H_2O was used as a negative control. (C) Confirmation of the *Mu8* insertion in the gene Zm00001eb052530 by Sanger sequencing. The sequence was obtained by sequencing of PCR products amplified using two primer pairs in separate reactions: 0982-F + TIR6 and *Mu8*-F + 0982-R2. The sequence represents a part of the gene with the *Mu8* insertion depicted in red letters and dashed lines. The black letters represent part of the gene sequence and dashed lines in black represents the remaining parts of the gene sequences.

Similarly, we confirmed the presence of a *Mu1* species in the gene Zm00001eb256020 (Figure 16) and another *Mu8* element in the gene Zm00001eb280980 (Supplemental figure S2) by genotyping individual plants from the F_2-families *BonnMu*-F7-2-F-1001 and *BonnMu*-7-C-0459, respectively. Subsequent confirmation involved sequencing of the *Mu*-specific PCR products to verify the corresponding *Mu* elements.

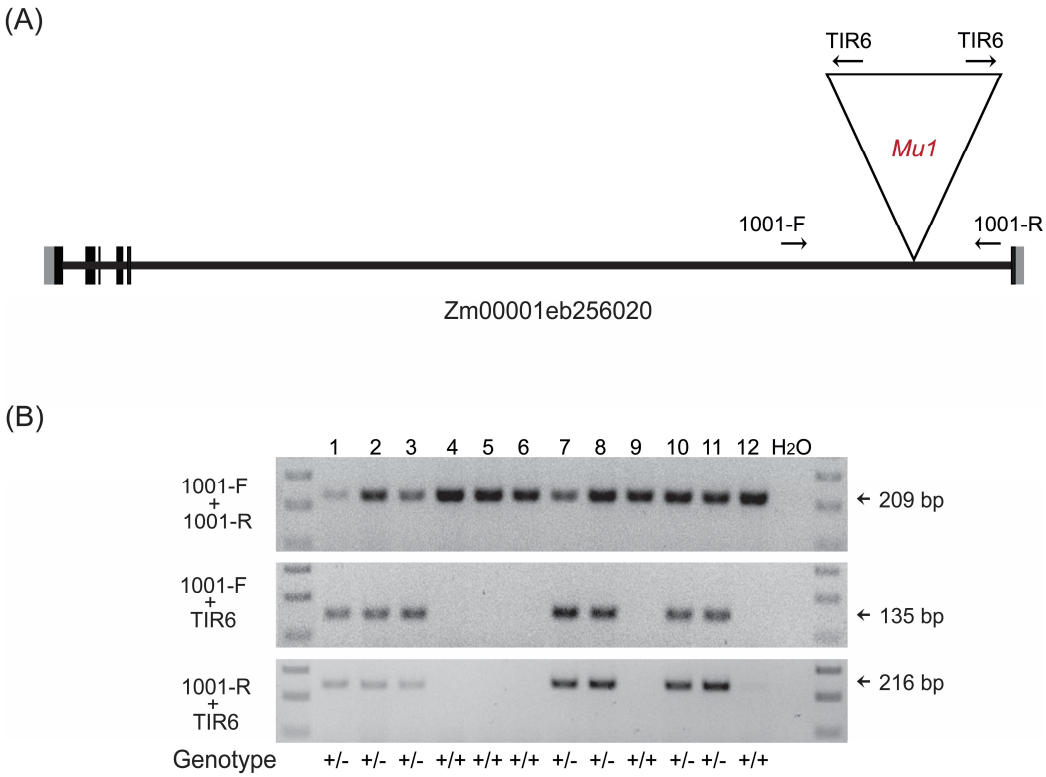


Figure 16. Confirmation of a *Mu1* element by PCR. (A) Simplified gene model of Zm00001eb256020 tagged by a *Mu1* element in an intron. Exons are illustrated as black boxes, introns as connecting line and UTRs as gray boxes. The *Mu* insertions are shown as triangles. Gene- and MuTIR-specific primer sites are indicated as arrows (F/R= gene-specific forward and reverse primers). (B) PCR segregation analysis of *BonnMu*-F7-2-F-1001 family identified 7 of the segregating plants as heterozygotes (+/-: # 1-3; # 7-8; # 10-11) and 5 plants as homozygous wild types (+/+: # 4-6; # 9; # 12).

5.1.8 Seedling images of *BonnMu* F_2-families as a forward genetic resource

The *BonnMu* resource provides easy-to-view photos of segregating F_2-families at the seedling stage accessible at the genome browser at https://jbrowse.maizegdb.org/ (Marcon *et al*., 2020; Win *et al*., 2024b). For the *BonnMu* reverse genetic resource, which contains 14 Mu-seq libraries, a total of 8,064 (576 x 14) *BonnMu* F_2-families were used. From these, 12-30 kernels per family for 7,679 families were germinated, and all the seedling images were phenotyped and photographed 10 dag. The remaining 385/ 8,064 families were not processed due to an insufficient amount of seeds.

The phenotyping included observations of leaf color and assessments of root and shoot phenotypes in seedlings aged 10-12 days. Among the *BonnMu* F_2-families analyzed, various mutants were identified, such as those affecting leaf color (e.g., *BonnMu*- F7-1-0481; Figure 17A) and root and shoot development (e.g., *BonnMu*-F7-3-1657; Figure 17B). Notably, the latter family displayed a segregating mutant characterized by magenta-colored roots and a dwarf phenotype, and its identification and characterization will be described in section (5.2).

The mutation rate, as determined by the albino and pale green leaf phenotypes, was 16% (1,219 of 7,679 F_2-families; Table 9). This mutation rate is consistent with previously published rates (Robertson, 1983; Marcon *et al.*, 2020), indicating high transposon activity in the *BonnMu* F_2-families.

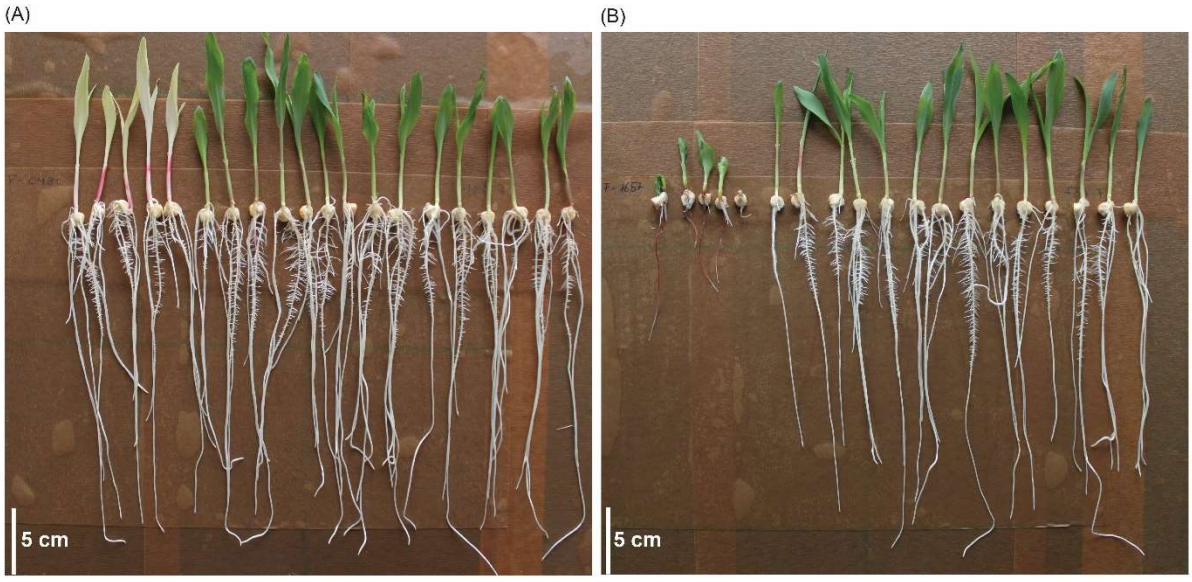


Figure 17. Phenotyping of two segregating *BonnMu* F_2-families. (A) *BonnMu*-F7-1-0481 segregating for a pale green leaf mutation and (B) *BonnMu*-F7-3-1657 segregating for the *magenta root dwarf 1* (*mrd1*) mutant, see chapter 5.2.

Table 9. Phenotyping and mutation rates of *BonnMu* F_2-families

BonnMu libraries	No. of phenotyped F_2-families	No. of families with leaf color mutation (3:1)	% leaf color mutation (3:1)
Museq_1	512	64	13
Museq_2	509*	28	6
Museq_3	522	24	5
Museq_4	524*	68	13
Museq_5	568	94	17
Museq_6	542	53	10
Museq_7	576[a]	48	8
Museq_8	576[a]	31	5
Museq_DK105_EP1_1	575	91	16
Museq_EP1_2	574	103	18
Museq_F7_1	567	188	33
Museq_F7_2	559	226	40
Museq_F7_3	562	100	18
Museq_F7_4	513	101	20
Total	**7,679**	**1,219**	**16**

* Two previously published Mu-seq libraries (Marcon *et al*., 2020) were included in the analysis.

[a] All 576 families per library were used for phenotyping for forward genetic resource. Not all 576 families per library were used in the remaining libraries due to insufficient seeds and ongoing propagation.

5.2 *BonnMu discovers a magenta root dwarf 1* (*mrd1*) mutant

5.2.1 Identification of *mrd1* in a forward genetic screen of *BonnMu*

A unique feature of *BonnMu* is that photographs of seedling phenotypes of segregating F_2-families are linked to the database entries of each transposon insertion at maizegdb.org. A forward genetic screen of 10-day-old seedling phenotypes of the *BonnMu* F_2-families allows to identify striking mutants affecting young root and shoot development. The mutant *magenta root dwarf 1* (*mrd1*) was first identified in the *BonnMu* F_2-family *BonnMu*-1-A-0438 (Supplemental figure S3). In this family, from 21 F_2-seedlings germinated, the χ^2 (Chi-squared) test showed a p-value of approximately 0.8997, with the observed 16 wild type to 5 mutant ratio which corresponds to the 3:1 Mendelian segregation expected for a monogenic recessive mutation (Table 10).

Table 10. χ^2 test for plants with wild type and *mrd1* phenotype in *BonnMu*-1-A-0438.

	Wild type	***mrd1***
Expected segregation	3:1	
Observed (O)	16	5
Expected (E)	16	5
O-E	0	0
χ^2	0.02	
Significance ($p \leq 0.05$, df=1)	*p*-value = 0.8997 (Not significant)	

In comparison to the wild type, *mrd1* mutants show a dwarfed primary root with only few lateral roots and a limited shoot growth under light and dark conditions (Supplemental figure S4). Furthermore, under both conditions (i.e. light or darkness), the primary root of *mrd1* accumulates anthocyanins in the cortex, resulting in the eponymous magenta root phenotype. Interestingly, the *mrd1* mutant was observed in 747 of 7,679 phenotyped *BonnMu* F_2-families. indicating a fixed mutation that was inherited from the *Mu*-active line (PS_07-1502-8[1st]/1505-5; Supplemental figure S5) used to mutagenize the inbred lines of our resource. We only detected a 3:1 Mendelian ratio in 7% (49 of 747 F_2-families) of the *BonnMu* families

segregating for *mrd1*. In addition, the *mrd1* mutant was observed in all genetic backgrounds used for generating *BonnMu* F_2-families.

The *mrd1* plants exhibit the "*magenta*" pigmentation in primary root, seminal roots and crown roots shortly after germination, suggesting an over-accumulation of anthocyanins. In maize, the first emergence of the primary root is typically observable approximately at 3 days after germination. Notably, magenta pigmentation in the primary root is appeared during the period of 4 to 5 days after germination. The pigmentation initiates at the basal part of the primary root adjacent to the kernel. At day 5 or 6 after germination, the whole primary root is pigmented, with the exception of the root tip which remains without coloration (Figure 18A). The cross sections of the differentiation zone (~1 cm from the root tip) of the primary root of *mrd1* mutant illustrates an increased accumulation of anthocyanins in the cortical parenchyma compared to the wild type (Figure 18B & 18C).

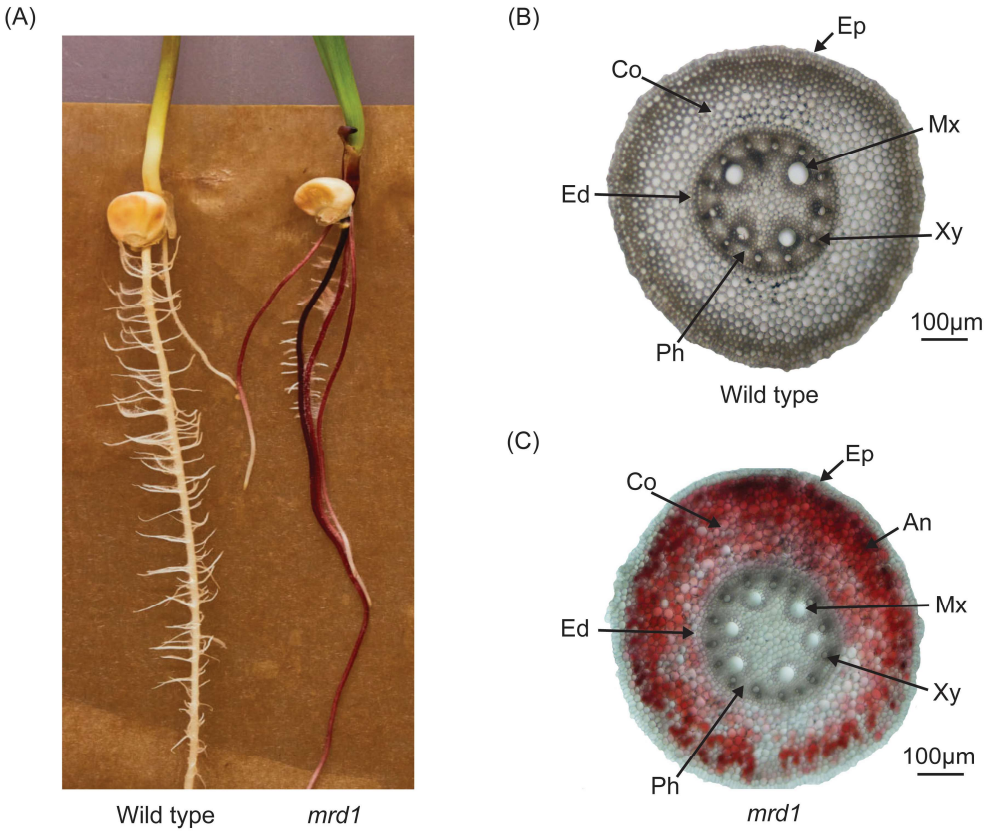


Figure 18. Phenotype of wild type and *mrd1* mutant seedlings. (A) The root phenotypes of wild type and *mrd1* mutant at 10 dag (taken by Volker Lannert, University of Bonn). (B) Cross section of the wild type primary root's differentiation zone (6 dag). (C) Cross section of the *mrd1* primary root's differentiation zone (6 dag). An: Anthocyanin; Co: Cortical parenchyma; Ed: Endodermis; Ep: Epidermis; Mx: Metaxylem; Ph: Phloem; Xy: Xylem.

5.2.2 Morphometric analysis of *mrd1* mutant and wild type seedlings reveals reduced growth of *mrd1*

To compare the phenotypic differences of the *mrd1* and wild type seedlings in more detail, following traits were analyzed under light and dark conditions: primary root length, lateral root density (number), seminal root number, crown root number, first leaf length, shoot length, mesocotyl length, and coleoptile length.

Under both light and dark conditions in a climate chamber *mrd1* mutants show a significantly reduced primary root length (Figure 19A). The density of lateral roots was lower in *mrd1* mutants, particularly under light conditions (Figure 19B). The number of seminal roots was also reduced in the *mrd1* mutant, with a significant reduction in the dark (Figure 19C). Crown root numbers in the *mrd1* mutants were not significantly different compared to the wild type in both conditions, but both *mrd1* and wild type showed more crown roots in the dark condition (Figure 19D).

In the case of shoot growth, the first leaf of the *mrd1* mutants was significantly shorter, especially when grown in the dark (Figure 19E). The shoot length did not differ between light conditions for *mrd1* mutant, in contrast to the wild type, which grew significantly longer in the dark (Figure 19F). In the dark, wild type seedlings undergo skotomorphogenesis characterized by an increase of mesocotyl length towards a potential light source (Supplemental figure S4). In contrast to wild type seedlings, *mrd1* develops a constitutive photomorphogenic phenotype in the dark comparable to the *mrd1* phenotype in light (Supplemental figure S4). The mesocotyl length of *mrd1* shows a significant difference to wild type under both conditions. The coleoptile length was also reduced in *mrd1* mutant in both conditions (Figure 19G&H). Across all these measurements, the *mrd1* mutant exhibited reduced growth in comparison to the wild type (Figure 19; Supplemental figure S6).

These findings suggest that the *mrd1* mutation has a profound impact on plant development, and the effects are exacerbated by the absence of light, implying a potential interaction between the *mrd1* gene function and light-mediated growth pathways.

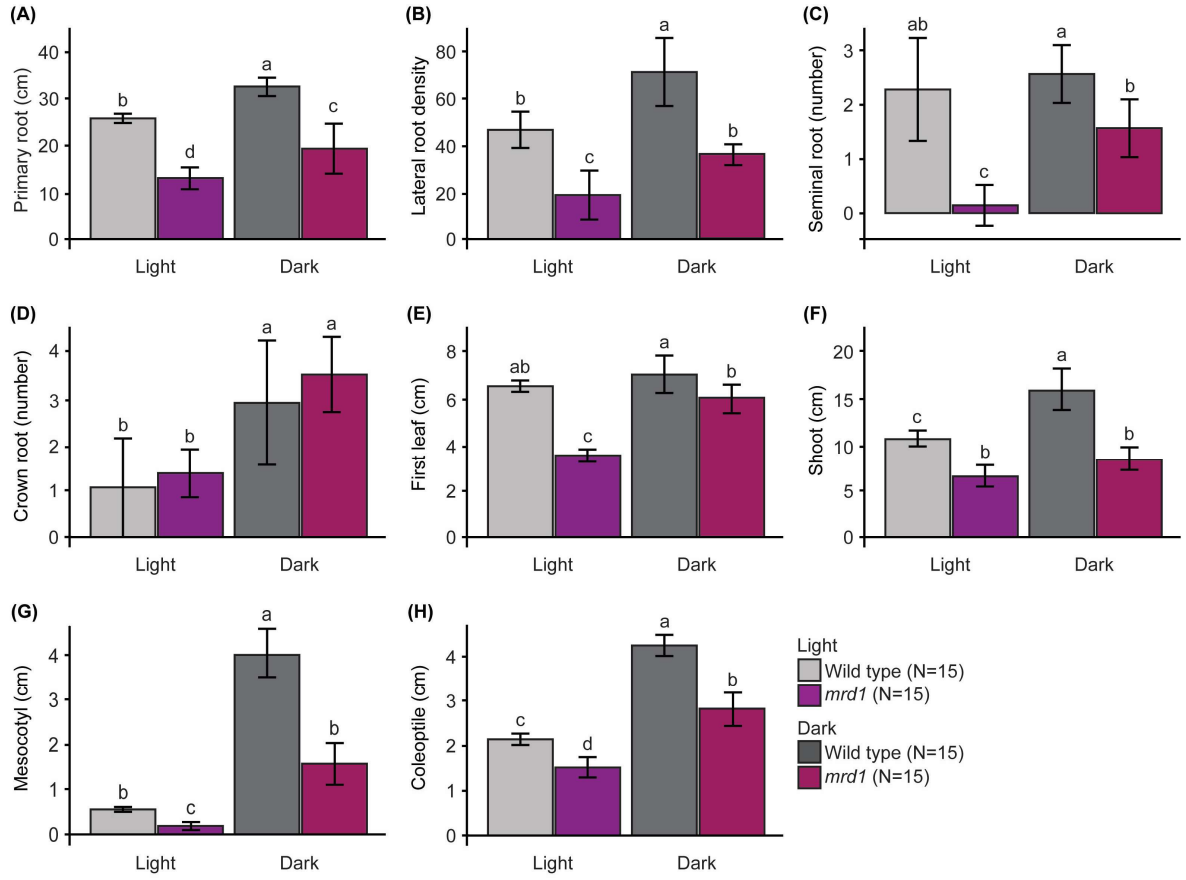


Figure 19. Phenotype characteristics of wild type and *mrd1* mutant seedlings at 10 days after germination. The measurement includes (A) primary root length, (B) lateral root density (number), (C) seminal root number, (D) crown root number, (E) length of first leaf, (F) shoot length, (G) mesocotyl length and (H) coleoptyle length. The phenotypes were measured under both light and dark condition. Error bars represent the standard error of the mean. An ANOVA test was performed for all the measurements and the statistical significance of these results was calculated at α = 0.05, with different letters (a, b, c, d) denoting statistically different groups. Bars with the same letter are not significantly different at α = 0.05 in all pairwise comparisons.

5.2.3 Analysis of the phenylpropanoid composition in roots of wild type and *mrd1* seedlings

To identify phenylpropanoid in roots of wild type and *mrd1* seedlings three biological replicates from each sample were analyzed via ultraperformance liquid chromatography coupled with a photodiode-array detector and high-resolution quadrupole time-of-flight mass spectrometry (UPLC-PDA-HR-QTOF-MS), with a special focus on anthocyanins. Among the phenylpropanoid compounds detected were anthocyanins, absorbing light at 500-550 nm (Figure 20). Commercial standards were used to confirm two anthocyanins, present in roots

of *mrd1* mutants: cyanidin 3-*O*-glucoside (CyG) and peonidin 3-*O*-glucoside (PeoG). Besides, other anthocynins or anthocyanin metabolites were detected but unknown and not confirmed by using standards (Figure 20).

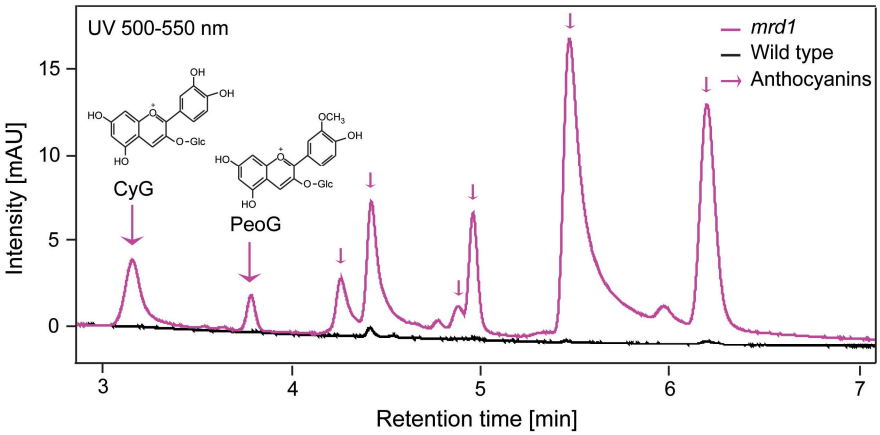


Figure 20. Phenylpropanoid profiles of primary roots of wild type and *mrd1* seedlings. Representative UPLC-UV chromatogram at 500-550 nm. The anthocyanins, cyanindin 3-O-glucoside (CyG) and peonidin 3-O-glucoside (PeoG), showing intensity peaks (milli-absorbance unit, mAU) at a retention time of 3.2 and 3.8 min, respectively, were identified. Magenta colored arrows in the chromatogram indicate *mrd1*-specific anthocyanins.

5.2.4 BSR-seq in combination with *Mu*-seq identified a *constitutive photomorphogenesis 9 signalosome complex subunit 4* as a candidate gene underlying the *mrd1* phenotype

The *mrd1* mutant was first identified in the *BonnMu* F_2-family 1-A-0438 harboring 64 germinal *Mu* insertions, which made it difficult to immediately identify the particular gene, associated with the phenotype. Therefore, we subjected the mutant to a bulked segregant RNA-seq analysis using the primary roots of *mrd1* mutant and wild type plants. To map the gene, allele frequencies were detected in the wild type pool against the *mrd1* pool (i.e. 85 bulked primary roots per pool; Mansfeld and Grumet, 2018). In case of a complete linkage of a SNP marker with a mutated gene, only one marker allele was expected to be detected in the *mrd1* pool of the mapping population. BSR-seq mapped the *mrd1* gene to a ~70 Mb interval of the short arm of chromosome 1 (Figure 21). The total number of differentially expressed genes on the chromosome 1 of the *mrd1* mutant pool compared to the wild type are shown in the Supplemental figure S7.

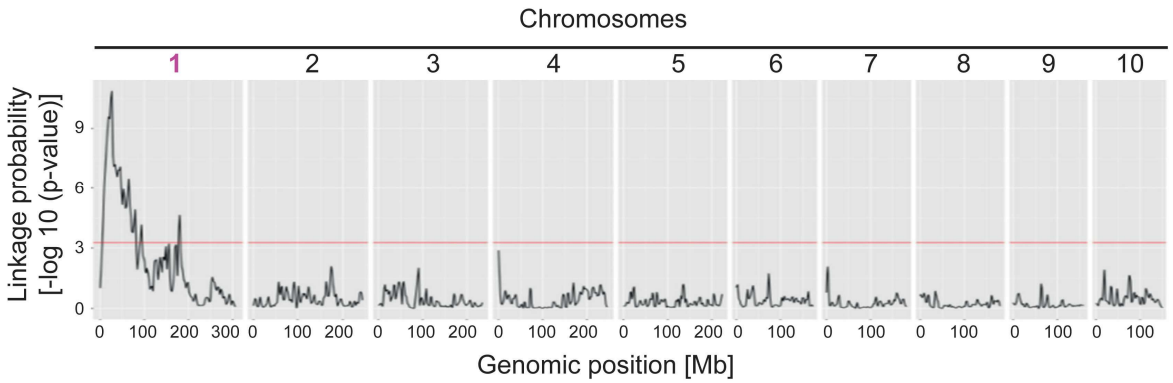


Figure 21. BSR-seq mapping confined *mrd1* to the short arm of chromosome 1. The genome was scanned using a smoothing window size of 3.5 Mb. The linkage probability obtained by BSA for each individual SNP was determined by a G‘ analysis and plotted against the genomic position (Mb).

Among the 64 germinal *Mu* insertions, detected in the F_2-family segregating for *mrd1*, five insertions were located in the mapping interval. Only the gene with the accession Zm00001eb008060 was significantly down-regulated based on the BSR-seq data (Supplemental table S9). Therefore, this gene, Zm00001eb008060, encoding a *constitutive photomorphogenesis 9 signalosome complex subunit 4* (*csn4*) was selected as the potential candidate gene underlying the *mrd1* phenotype.

Our Mu-seq detected a *Mu* insertion in the 2nd intron of the *mrd1* gene, 546 bp downstream of the start codon (Figure 22A). To verify if the *Mu* insertion in the *mrd1* gene co-segregates with the *mrd1* phenotype, PCR-based genotyping of segregating F_2-plants was performed. To this end we used gene-specific primers and primers specific for the terminal inverted repeat (TIR) sequence which is highly conserved among all *Mu* transposons. Indeed, genotyping of segregating plants identified all individuals with mutant phenotype as homozygous (-/-) for the insertion and wild types as heterozygous (+/-) or homozygous for the wild type allele (+/+; Figure 22B). Sequencing of the *Mu*-specific PCR-products demonstrated that the *Mutator* insertion in this gene was caused by the *Mu-element 1* (*Mu1*).

Expression of *mrd1* in young primary roots of homozygous mutant (-/-), heterozygous (+/-), and homozygous wild type was analyzed via qRT-PCR relative to the *homeobox-transcription factor 3* (*hb3*; Figure 22C). The expression of *mrd1* was significantly reduced in the primary roots of *mrd1* (-/-), compared to heterozygous (+/-) and wild type (+/+) plants. Precisely, *mrd1* expression in heterozygous (+/-) individuals was significantly higher than in *mrd1*

mutants (-/-), but significantly lower than in wild type (+/+) samples (Figure 22C). Hence, *csn4* was depicted as a likely candidate gene, underlying the *mrd1* phenotype.

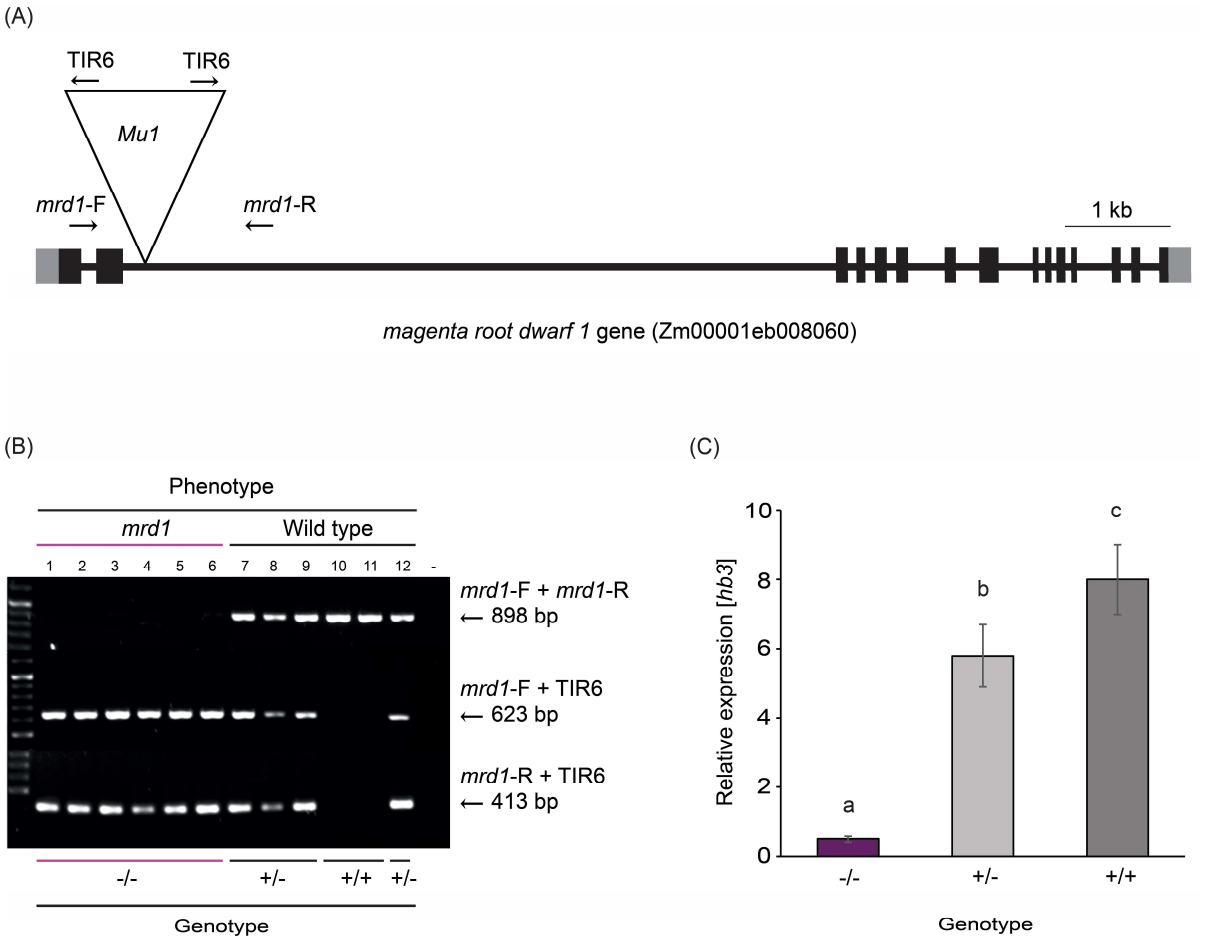


Figure 22. PCR-based co-segregation analysis to confirm *Mu* insertion in the *magenta root dwarf 1* (*mrd1*) gene, Zm00001eb008060. (A) Gene model of *mrd1*. The *Mu1* insertion in the 2nd intron is shown as a triangle. Gene- and MuTIR-specific primer sites are indicated as arrows. Exons are illustrated as black boxes and introns as lines. The gray boxes represent UTR regions. (B) Exemplary PCR-based co-segregation analysis of 12 individuals identified *mrd1* mutants as homozygous (-/-; samples 1-6) and wild types as either heterozygous (+/-; samples 7-9; 12) for the insertion or homozygous for wild type allele (+/+; samples 10-11). (C) Expression of *mrd1* analyzed via qRT-PCR relative to the *homeobox-transcription factor 3* (*hb3*) in pooled 2-4 cm primary roots of homozygous mutants, heterozygous, and homozygous wild type individuals. Small letters indicate significant differences of expression values between each sample as calculated by Tukey's test.

5.2.5 Phylogenetic tree analysis shows that MRD1 is a homolog to Arabidopsis CSN4

Phylogenetic analysis showed that the *Zea mays mrd1* gene (Zm00001eb008060) is part of a conserved genetic lineage in monocotyledonous and dicotyledonous plants, containing one *mrd1* homolog (Figure 23). It exhibits significant homology with the *csn4* gene from *Arabidopsis thaliana*, which possesses a similar PCI domain. The inclusion of Arabidopsis in the phylogenetic tree is particularly relevant given the phenotypic similarities observed in mutants of both species. The relatively small phylogenetic distance and relatively high posterior probability at the node encompassing *Zea mays* and *Arabidopsis thaliana* suggests a degree of confidence in their evolutionary relatedness and reflects the conserved functional aspects of the PCI domain across these divergent species.

The evolutionary relatedness between the *mrd1* gene in maize and *csn4* in Arabidopsis is also consistent with the functional similarities of the resulting phenotypes when mutated. As in Arabidopsis, the *mrd1* mutant shows a photomorphogenesis phenotype when it is grown in the dark. Unlike Arabidopsis, where anthocyanin accumulates in the cotyledons (Castle and Meinke, 1994; Pacurar *et al.*, 2017), in the *mrd1* mutant, anthocyanin accumulation occurs in the roots. Despite the evolutionary distance and the basal divergence of monocots and dicots, the presence of a common domain indicates a potentially similar functional role, which has been conserved across these two lineages. This functional conservation is further substantiated by the observed mutant phenotypes, suggesting that the role of the *mrd1* gene and its Arabidopsis homolog may be critical in similar biological processes, potentially relating to the accumulation of anthocyanins.

In addition, the identity and similarity metrics between the *mrd1* gene and its *Arabidopsis thaliana* homolog, at 84% and 90% respectively, further support the notion of functional conservation. Moreover, *mrd1* exhibits the highest homology with the gene in *Sorghum bicolor*, with 100% identity and similarity (Supplemental table S10).

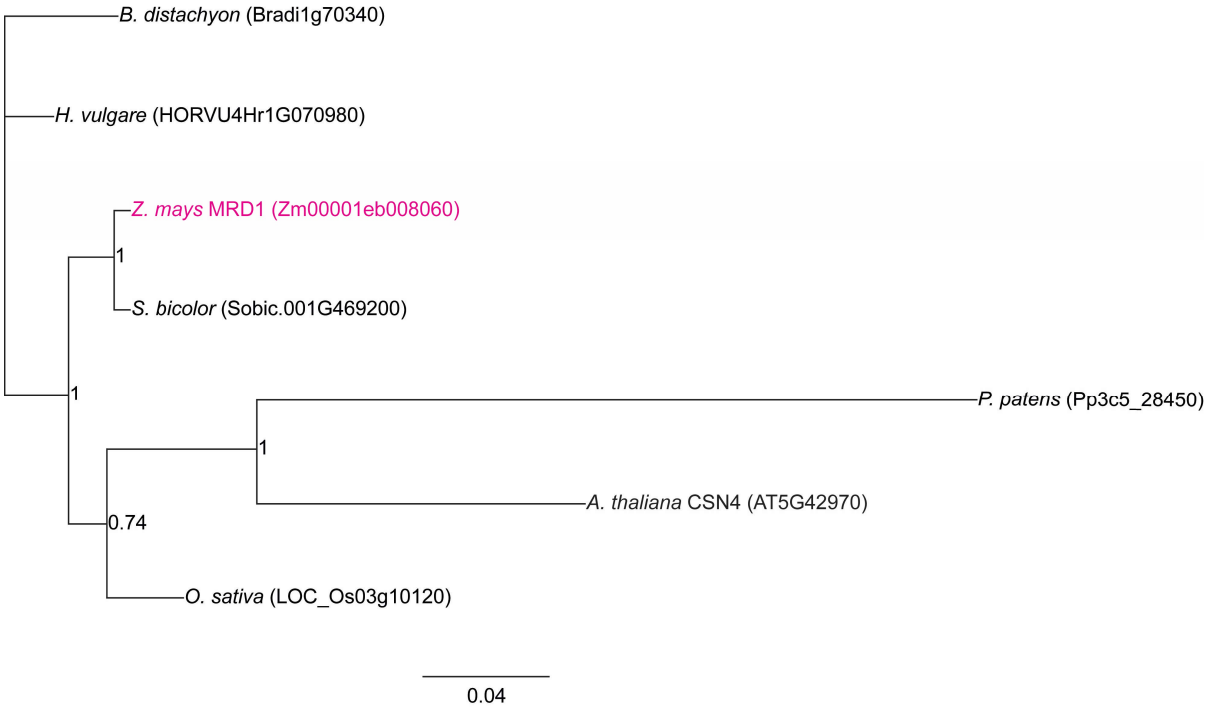


Figure 23. Phylogenetic tree of *Zea mays* MRD1 homologous proteins in selected plants. The protein sequences of MRD1 and the homologous sequences form other plants were aligned using the MUSCLE algorithm in MEGA X software (Kumar *et al.*, 2018). Subsequently, the tree was completed using the aligned sequences by MRBAYES software. The tree shows Bayesian posterior probability values at the nodes for each clade. The scale bar represents 0.04 estimated substitutions per site.

5.2.6 Microbiome profiling of the rhizosphere from crown roots shows enriched bacterial taxa in *mrd1* and wild type

Root-derived flavonoids promote predominantly the enrichment of bacteria of the taxon Oxalobacteraceae in the rhizosphere, which in turn promote maize growth and nitrogen acquisition (Yu *et al.*, 2021). Genetic experiments demonstrated that lateral root development coordinates the interactions of the root system with flavonoid-dependent Oxalobacteraceae under nitrogen deprivation. In order to analyze the impact of the altered root anthocyanin content and altered root architecture of *mrd1* on the rhizosphere microbiome, we performed the microbiome profiling experiment. The *mrd1* mutant showed a reduced number of lateral roots with a shortened primary root length and fewer seminal roots compared to the wild type plant (Supplemental figure S3). Since *mrd1* and wild type plants have a different growth rate, we selected crown roots of similar length for microbiome profiling experiment from *mrd1* and wild type to accurately compare these two genotypes. In maize, crown roots (CR, Figure 24A)

are post-embryonic shoot-borne roots emerging from the coleoptilar node and become dominant a few weeks after germination. We harvested the rhizosphere samples from one-month-old wild type and *mrd1* plants and conducted bacterial 16S rRNA DNA sequencing. Figure 24B displays a multidimensional scaling plot based on OTU (Operational Taxonomic Unit) level, illustrating the β (beta)-diversity or inter-sample variability in microbial community composition in wild type and *mrd1*. Bacterial diversity shows significant differences between *mrd1* and wild type rhizosphere as controlled by analysis of similarities (ANOSIM, R = 0.42, *P* = 0.03; Figure 24C). This result identified 10 high abundant bacterial genera to be significantly different between these two genotypes (Figure 24C). Three bacterial genera—*Luteimonas*, *Arenimonas*, and *Thermomonas*—were more abundant in the wild type, whereas seven other genera showed significant enrichment in the *mrd1* mutant (Figure 24C).

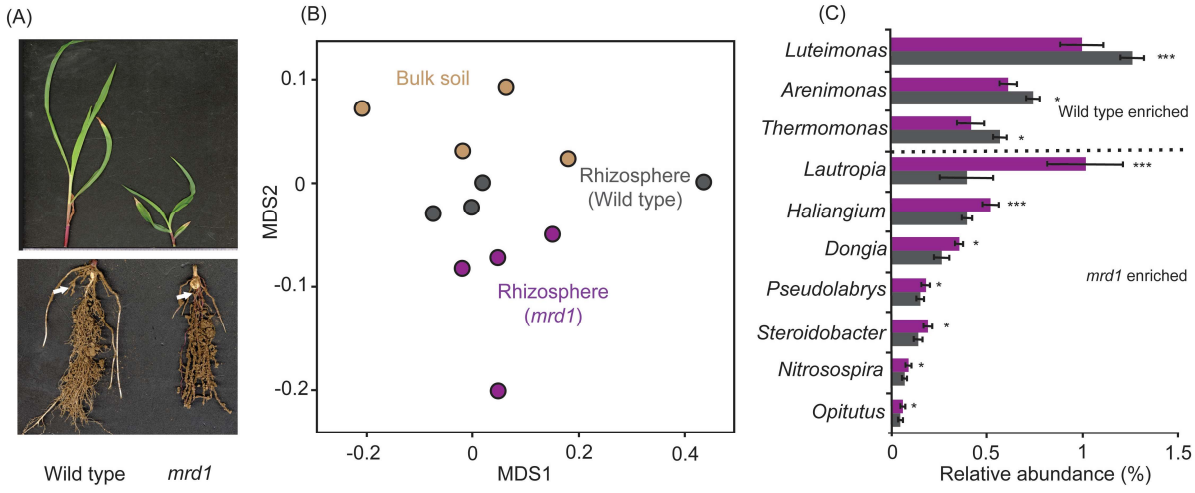


Figure 24. Dissimilarity of bacterial microbiome in the rhizosphere of wild type and *mrd1* mutant seedlings. (A) The rhizosphere samples were collected from one-month-old wild type and *mrd1* mutant plants. Top figure: One month after germination *mrd1* shows a shorter shoot system, compared to wild type. Bottom figure: Arrows indicate crown roots and attached soil, collected for bacterial microbiome analysis. (B) The microbiome of rhizosphere and bulk soil (i.e. control pots without plants) were profiled using 16S rRNA gene sequencing and the beta diversity was shown by multidimensional scaling (MDS) plot. (C) Significantly enriched bacterial genera in the rhizosphere of both wild type and *mrd1* mutant. Asterisks indicate the significant levels at the false discovery rate (FDR) 0.05 (*), 0.01 (**) and 0.001 (***).

5.2.7 Comparative RNA-seq analysis of *mrd1* and wild type crown roots

The *mrd1* mutant is distinguished by its magenta pigmentation in the roots and a dwarfed shoot phenotype. This mutant exhibits photomorphogenic growth in dark conditions similar to Arabidopsis. However, in contrast to Arabidopsis, where anthocyanin accumulation occurs in the cotyledons (Castle and Meinke, 1994; Pacurar *et al.*, 2017), the *mrd1* mutant demonstrates pigmentation predominantly in the roots. This analysis aimed to identify differentially expressed genes, particularly light-responsive ones hypothesized to be aberrantly activated in *mrd1* mutants even under dark conditions and to unravel the transcriptome alterations associated with the mutant's photomorphogenic growth. For the transcriptome analysis, we sampled crown roots from the same plants (wild type and *mrd1* mutant) used for microbiome profiling. For RNA-seq, we extracted total RNA from three replicates of each genotype. After sequencing, the paired raw reads were aligned to the reference genome of maize (*Zea mays*; Zm-B73-REFERENCE-NAM-5.0.57). To understand the transcriptomic relationships among the samples, we visualized them using a principle component analysis (PCA) plot (Figure 25A). In the PCA plot, the first two components, PC1 and PC2, represented 87% of the total variation. Specifically, PC1 accounted for 72% of the total variance and clearly differentiated between the wild type and *mrd1* mutant samples. One replicate from the mutant appeared visually separated from the primary group. Therefore, we assessed the significance of this separation through outlier detection tests, using Tukey's and Hubert's method. These outlier tests were applied to PCA results obtained from normalized gene expression data, reflecting the underlying transcriptomic relationships with adjustments for sequencing depth and compositional differences among samples. Despite the initial visual separation of one *mrd1* mutant replicate in the PCA plot, the outlier tests using both Tukey's and Hubert's methods, substantiated by p-values of 0.62 and 0.57 respectively, revealed no significant divergence from the other *mrd1* samples. Therefore, this sample was included in subsequent analyses.

To analyze genes that are differentially expressed between the two genotypes, we compared wild type and *mrd1* mutant samples. We identified differentially expressed genes (DEGs) with a false discovery rate (FDR) of less than 5. In total, we found 23,601 genes (59%) of 39,756 coding gene models (B73v5) in the RNA-seq dataset. Among those genes, 778 genes were differentially expressed in the mutant from which, 422 genes were upregulated, and 356 were downregulated. Notably, the *mrd1* gene was among the downregulated genes (Figure 25B).

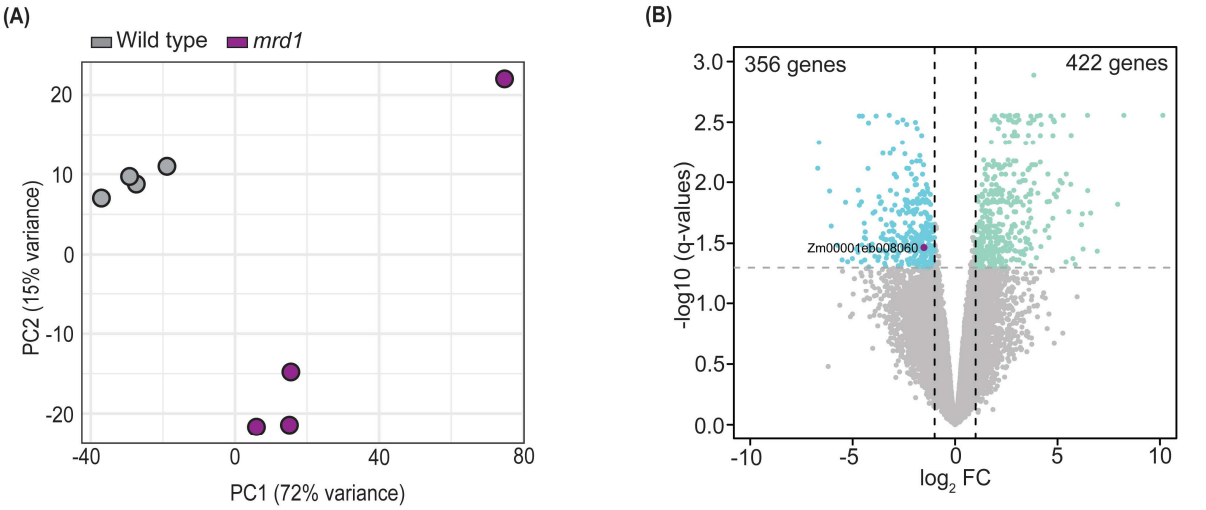


Figure 25. RNA-seq analysis of *mrd1* and wild type crown roots. (A) A principal component analysis (PCA) visualization of four wild type and four *mrd1* samples. In the PCA plot, two components (PC1 and PC2) accounted for 87% of the total variance. (B) A volcano plot highlighting differentially expressed genes (DEGs). Each dot represents an expressed gene. While green dots denote genes being up-regulated in *mrd1*, blue dots illustrate genes that are down-regulated in the mutant. The total number of significant up- and down-regulated DEGs are displayed in the top left and right corners, respectively. The *mrd1* gene is marked with a magenta dot. DEGs that do not surpass the threshold set at |log2FC| >1 and FDR ≤1% are shown in grey.

5.2.8 GO term analysis reveals upregulation of light responsive genes in *mrd1*

We performed Gene Ontology (GO) enrichment analyses using the genes which are up- and down-regulated in the *mrd1* mutant. The GO terms that are significantly affected in biological processes in the mutant are depicted in Figure 26. An interesting pattern emerged from our evaluation of the upregulated genes: a strong enrichment in GO terms linked to light interactions and light processes (Figure 26A). This was particularly evident from the overrepresentation of terms such as red, far-red light phototransduction (GO:0009585), regulation of photosynthesis via light reactions (GO:0042548), and the broader response to light stimulus (GO:0009416). Such enrichments elucidate a potentially altered or amplified interaction of the *mrd1* mutant with light.

When examining the down-regulated genes in the *mrd1* mutant through GO enrichment analyses, there was a distinct emphasis on processes tied to cell wall development and oxidative stress management (Figure 26B). The significantly enriched terms include cell wall organization (GO:0071555), plant-type secondary cell wall biogenesis (GO:0009834), and a

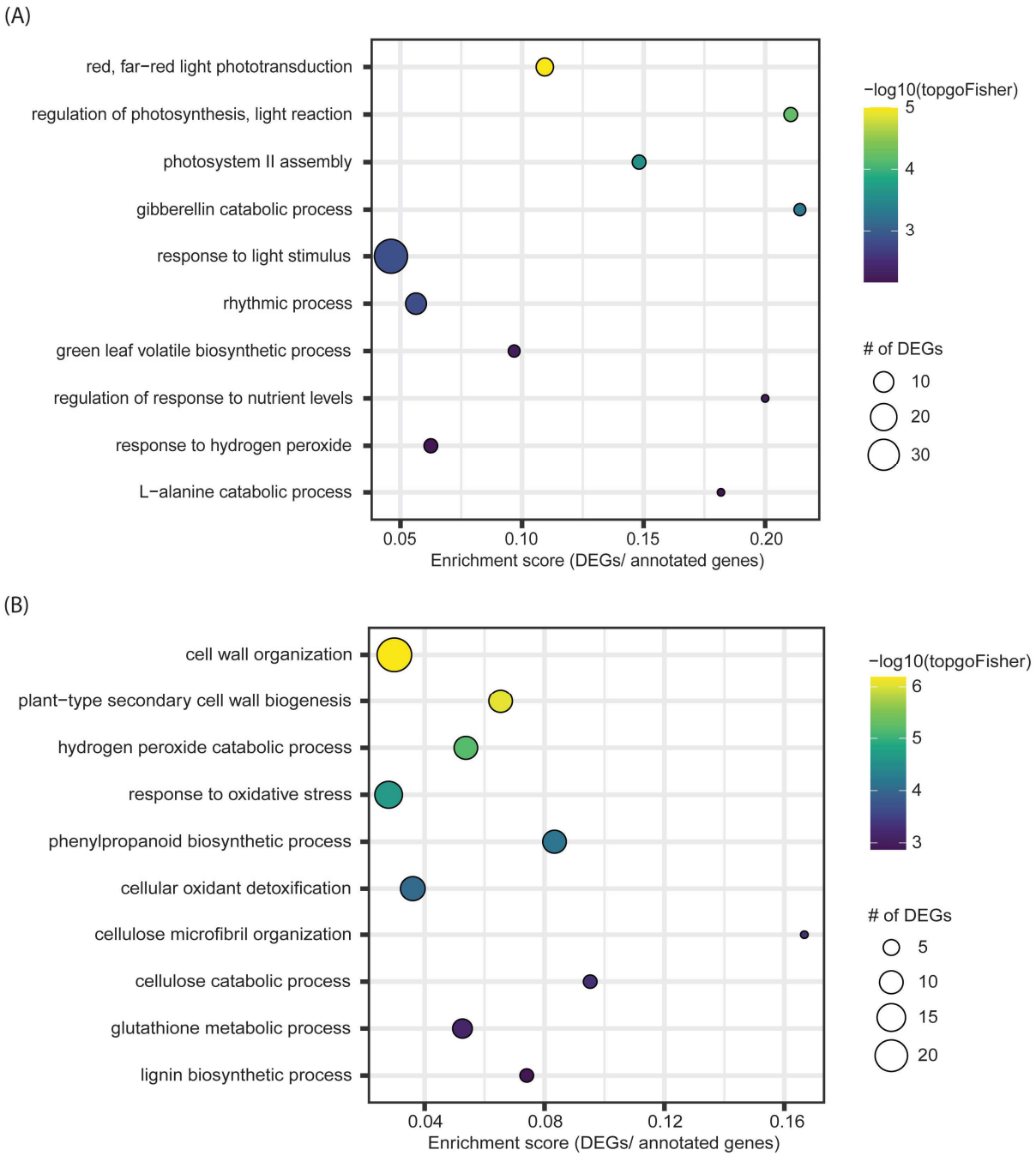


Figure 26. Enriched biological processes according to Gene Ontology (GO) enrichment analysis of up- and down-regulated DEGs. (A) Top ten GO terms of up-regulated DEGs. (B) Top ten GO terms of down-regulated DEGs. GO terms are displayed as circles. Circle size represents the number of DEGs associated with the respective term and the distance between those circles indicates relationship between terms. Color indicates significance of differential expression of an individual GO term. The circle represents the average $\log_2$ FC of the associated DEGs. Only the top 10 GO categories with significantly enriched terms with $p < 0.05$ based on Fisher's exact test are shown here.

cellulose microfibril organization (GO:0010215). This pattern indicates potential compromises in the structural integrity and composition of the cell wall in the mutant. Additionally, terms such as hydrogen peroxide catabolic process (GO:0042744), response to oxidative stress (GO:0006979), and cellular oxidant detoxification (GO:0098869) indicate potential vulnerabilities in the mutant's ability against oxidative stress. The GO analysis further revealed that among the down-regulated genes in the *mrd1* mutant, there was a notable association with pathways for responding to external stressors. Specifically, this includes pathways involved in the biosynthesis of key compounds such as lignin (GO:0009809) and glutathione (GO:0006749).

5.2.9 MRD1 is localized in the nucleus and endoplasmic reticulum (ER)

In Arabidopsis, the COP9 complex is localized in the nucleus (Chamovitz *et al*., 1996). The subcellular localization of the MRD1 protein was investigated in tobacco leaf epidermal cells utilizing the 2in1 FRET vectors, which allow for simultaneous visualization of two proteins of interest. This system was applied following the protocol described by Baer *et al*., (2022), enabling the study of co-localization and potential protein-protein interactions. The MRD1 protein was fused to the fluorescent reporter mVenus and HMGA (chromatin-associated high mobility group protein A) serving as a nuclear localization control, was tagged with tagRFP. Confocal imaging revealed a pronounced expression of the MRD1-mVenus fusion protein in nucleus and cytoplasm when MRD1 was fused to mVenus N-terminally (Figure 27A). When the MRD1-mVenus is localized C-terminally, the signal is clearly surrounding the nucleus which is typical for ER localization (Figure 27B). HMGA-tagRFP as nuclear localization control was expressed in the nucleus exclusively in both N- and C-terminal fusions (Figure 27).

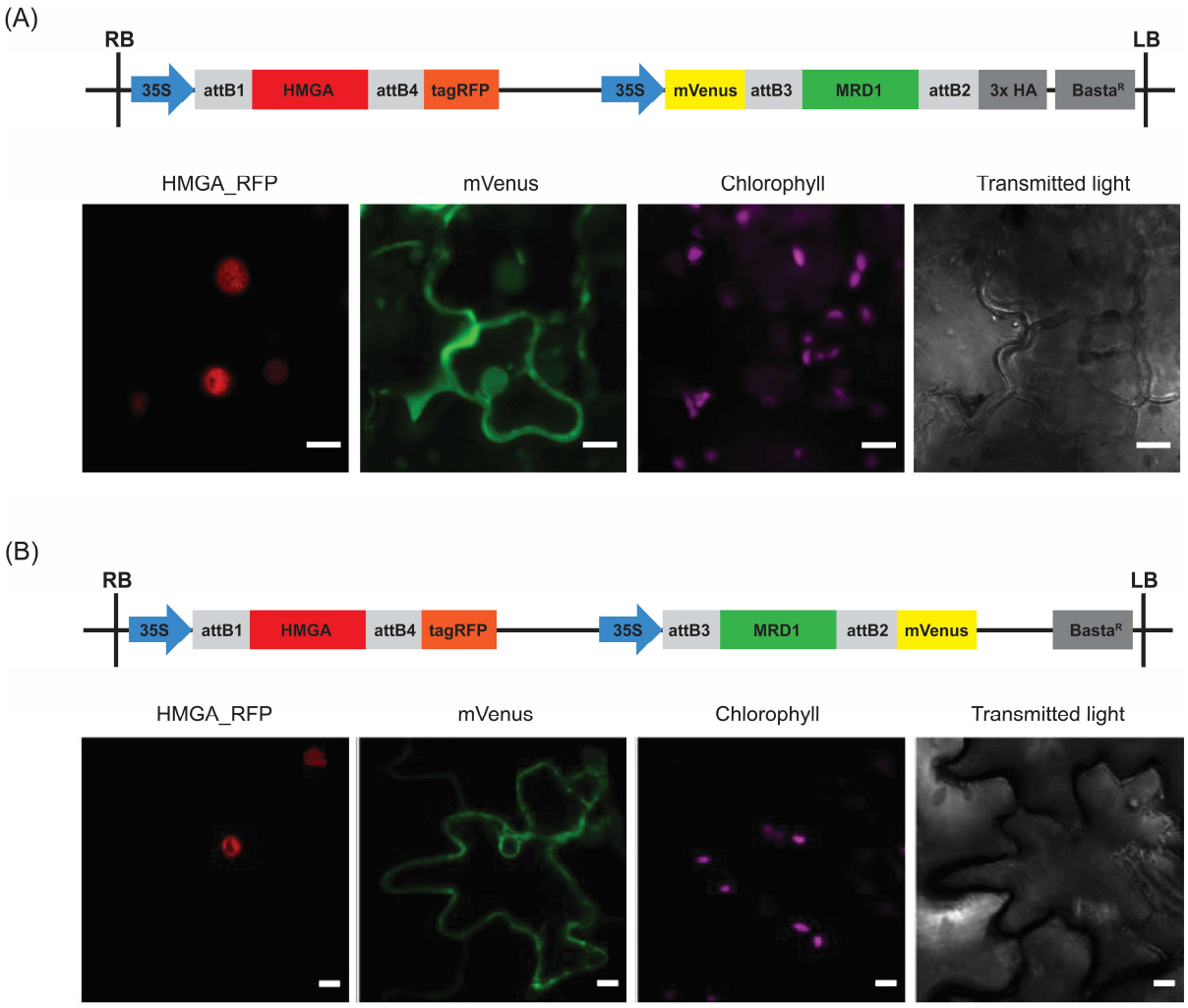


Figure 27. Subcellular localization of MRD1-mVenus fusion proteins transiently expressed in tobacco leaves. Subcellular localization of the MRD1-mVenus fusion protein; (A) the mVenus was fused N-terminally to MRD1 and (B) mVenus was fused C-terminally to MRD1. HMGA, utilized as a nuclear control, was fused to tagRFP. RB, right border; LB, left border. Scale bar, 10 µm.

5.2.10 Validation of *mrd1* gene using CRISPR/Cas9 genome editing

To confirm that *csn4* is the candidate gene responsible for the *mrd1* phenotype, we initially searched for an independent mutant allele in the *BonnMu* resource, but without success. Therefore, we wanted to create independent mutant alleles via CRISPR/Cas9 genome editing. We applied a dual single-guide RNA strategy to target two regions of the *mrd1* gene (Pauwels *et al*., 2018). The first target site is a 441 bp region between exon 1 and intron 2. The second target site is a 592 bp region between exon 11 and exon 13 where the only PCI domain of the gene is located. In total, nine T0 generation plants for the first target site and 24 plants for the second target site were regenerated from the callus.

Subsequently, all these plants were subjected to genotyping to analyze the INDELs (insertion/deletion). Among the plants targeted at the first target site, six plants were successfully edited. However, only five of these edited T0 plants survived and were backcrossed with the B104 line (Supplemental table S11). For the second target site, seven plants were successfully edited, however, all these died 3-4 weeks after transferring into the soil during the T0 stage (Supplemental table S12).

In T0 plants, the transgene is present as a single-copy and therefore, we expect that the backcrossed F_1 plants are segregating for 1:1 ratio for the absence and presence of transgene (Hunter, 2021). Consequently, the F_1 plants were examined for the presence of the transgene, as described in section (4.21). Those plants identified as free from the CRISPR constructs were then subjected to further genotyping the presence of mutant allele, followed by self-pollination to generate segregating F_2 generation. All plants within the F_2 generation were subjected to both genotyping and phenotyping analyses and the outcomes are shown in the Supplemental table S11. Homozygous mutants which carry different alleles in target region 1 are named *mrd1-2* to *mrd1-7* (Figure 28A) but they all did not exhibit the magenta root dwarf phenotype. Therefore, in the following part, the genotyping result of one allele (*mrd1-7*) with specific edits in the first target region is shown as an example. Genotypic analysis of the F_2 population highlighted specific edits within the *mrd1* gene: a 3 bp deletion in the first exon and a 1 base pair insertion within the second intron (Figure 28A, 28B and 28C). The 3 bp deletion leads to a change of one amino acid from P (proline) to H (histidine) and a deletion of the following amino acid (D, aspartic acid; Figure 28C). The 1 base pair insertion occurred within an intronic region, which is not translated into the protein sequence (Figure 28A and 28B). Given its location, this insertion is not expected to influence the protein's amino acid sequence, as introns are removed from the pre-mRNA during the splicing process before

translation. Consequently, no disruption to the protein structure was anticipated, and this was consistent with the observed absence of the magenta root dwarf phenotype (Figure 28A and 28B).

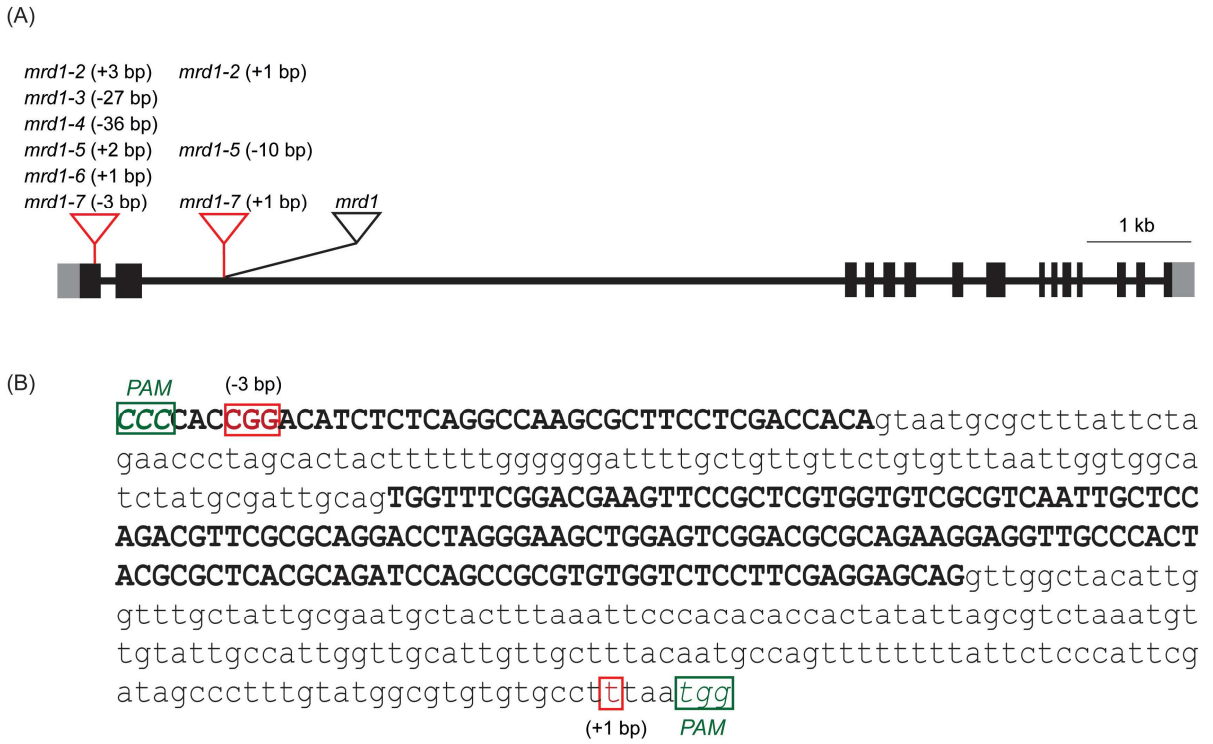


Figure 28. Validation of *mrd1* plants edited via CRISPR/Cas9. (A) The gene model of *mrd1* is depicted, with black boxes indicating exons and the connecting line representing introns. Mutations induced by CRISPR/Cas9 are denoted by red triangles, including six edited mutant alleles and one original *BonnMu mrd1* allele depicted by a black triangle. (B) Sanger sequencing of the *mrd1-7* allele confirms that the first guide RNA induced a deletion of 3 bp (depicted as -3 bp in red), and the second guide RNA facilitated the insertion of a single base pair (+1 bp in red) adjacent to the cut sites. Exons are shown in bold uppercase letters, whereas introns are represented in lowercase letters. The red letters with red boxes in the sequence highlight the deletion and insertion. (C) The genomic DNA and amino acid sequences of wild type and CRISPR edited *mrd1-7* allele are compared. The amino acid sequence of edited *mrd1-7* was obtained by NCBI translation tool. Green box indicates PAM sequence and red boxes indicate insertion and deletion.

These results suggest that the specific genomic alterations introduced by the CRISPR/Cas9 system were not sufficient to alter the protein function, as evidenced by the preservation of the wild type phenotype (Figure 29).

Figure 29. The phenotype of a CRISPR/Cas9 edited *mrd1-7* mutant with a 3 bp deletion in the first exon and a 1 bp insertion in the second intron and a wild type plant at 14 dag.

6 Discussion

6.1 *BonnMu* resource expansion in different maize germplasm backgrounds

Sequencing of 8,064 *Mu* transposon-tagged *BonnMu* F_2-families identified 425,924 germinal insertions, representing 36,612 (83%; Table 7) of all annotated gene models of maize. This number marks a substantial increase from the prior 57% coverage (Marcon *et al*. in 2020), which incorporated data from the *UniformMu* (McCarty *et al.*, 2013a) and the *ChinaMu* (Liang *et al.*, 2019) resources. The *ChinaMu* database has been updated with 32,224 *Mu*-tagged genes, which covered 73% of the maize genes (Liang *et al*., 2024). Nearly approaching whole-genome coverage, the *BonnMu* collection in maize tagged more genes than those mutagenized in rice (60%; Wang *et al.*, 2013) but fewer than in Arabidopsis (>94% gene coverage; Alonso *et al.*, 2003). For rice and Arabidopsis different techniques were employed, such as the two-component transposon system *Ac*/*Ds*-based mutagenesis (van Enckevort *et al.*, 2005), the *Tos17* retrotransposon mutagenesis (Miyao *et al.*, 2003), and the transfer-DNA insertional mutagenesis (Alonso *et al.*, 2003; Toki *et al.*, 2006). To track *Mu*-induced insertions in the maize genome, the *BonnMu* library uses the high-throughput sequencing strategy Mu-seq (McCarty *et al.*, 2013a; Liu *et al*., 2016), which has been coupled to the robust automated downstream analysis MuWU (Stöcker *et al.*, 2022), accelerating the identification of germinal insertions.

A unique feature of the *BonnMu* resource is that different North American Dent and European Flint germplasms were mutagenized thereby expanding and complementing the available resource. Notably, the Flint lines DK105, EP1 and F7 are adapted to the climate of central Europe (Unterseer *et al.*, 2016), and they represent important founders for European breeding programs (Haberer *et al.*, 2020). In this study, we demonstrated that a substantial number of 4,027 genes were exclusively tagged in at least one Mu-seq library of the mutagenized Flint lines, but not in the Dent lines B73 and Co125 (Figure 10). In contrast, the mutagenized *BonnMu* F_2-families of the Dent pool contributed 2,585 *Mu*-tagged genes, which remained unaffected in any of the mutagenized Flint lines. It could be worth checking for genic presence-absence variations among the genes affected by *Mu* insertions. Moderate genic presence-absence variations exist among Flint and Dent germplasm. Some of these genes found in either the Dent or the Flint pool, are expressed at high levels (Haberer *et al.*, 2020), which could contribute to line-specific adaptations to environmental impacts and hence be of interest for maize improvement and breeding. Therefore, in future studies, the effect of *Mu*-tagged

presence-absence genes could be investigated to identify genotype-specific mutations and their impact on the mutant phenotype.

It has been reported that *Mu* elements exhibit a pronounced preference for 5' UTRs of genes and tend to concentrate in genomic regions with epigenetic marks of open chromatin near the transcription start site of genes (Liu *et al.*, 2009; Springer *et al.*, 2018; Marcon *et al.*, 2020; Zhang *et al.*, 2020). In support with this, we identified the vast majority of 94% of all *Mu* insertions in genic regions of the genome, while only 6% of the insertions targeted intergenic regions (Figure 12). Moreover, the distribution pattern of *BonnMu* insertions across the 10 maize chromosomes indicates that gene-dense chromosome arms are hotspots for *Mu* elements, whereas the heterochromatic centromere regions harbor fewer insertions (Figure 13). This finding is consistent with previous observations that gene-rich chromosome arms are associated with highly accumulated *Mu* elements, e.g. in the B73 (Schnable *et al.*, 2009) and W22 (Springer *et al.*, 2018) genomes.

Remarkably, we pinpointed chromosomal regions, specifically on chromosomes 5-8, that exhibit minimal occurrences of *BonnMu* insertions. These areas could represent highly heterochromatic non-accessible knob regions in the genome that suppress local recombination (Ghaffari *et al.*, 2013). Knobs are multi-megabase tandem repeat arrays, predominantly located in mid-arm positions of chromosomes (Dawe and Hiatt, 2004). They primarily consist of two tandemly repeated DNA sequences: the 180 bp knob repeat and the 350 bp tandemly repeated element TR-1 (Ananiev *et al.*, 1998). TR1 and 180-bp repeat knobs are known to contribute to meiotic segregation by pulling themselves toward spindle poles (Swentowsky *et al.*, 2020). However, they have not been fully sequenced or accurately represented in genome assemblies due to the challenge of assembling their long, repetitive structures. Karyotyping has identified knob loci in maize, but there is a significant discrepancy in their representation (Ghaffari *et al.*, 2013). While the physical size of knobs extends over a million base pairs, they are represented as only a few kilobases in genome assemblies. Advances in DNA sequencing technologies have significantly improved the maize B73 genome assembly, thereby reducing the initial count of over one hundred thousand gaps to just a few thousand (Schnable *et al.*, 2009; Jiao *et al.*, 2017; S. Sun *et al.*, 2018). In the genome of the maize inbred line Mo17, heterochromatic knob180 and TR-1 arrays were predominantly detected in the chromosome arms of chr1L, chr4L, chr6S, chr6L, chr8L and chr9S (Chen *et al.*, 2023). Similarly, the karyotype of B73 identified such tandem repeat DNA sequences, primarily on chr4L, chr5L, chr6S, chr7L, chr8L, chr9S (Ghaffari *et al.*, 2013).

These areas are roughly overlapping with the regions on chr5L, chr6S, chr7L, and chr8L in the B73 genome (Figure 13), showing a notable absence of *BonnMu* insertions. Since knobs represent heterochromatic repeat elements, it is likely that they are not accessible for *Mu* transposon integration. The phenomenon highlights a fundamental limitation in achieving 100% gene coverage in maize using *Mu* transposon-tagged mutagenesis, underscoring the complex interplay between chromatin accessibility, gene expression (CSHL, 2000; Carvalho *et al.*, 2022), and transposon insertion preferences. Therefore, while significant strides have been made in expanding the scope of gene coverage through Mu transposon tagging in maize, the inherent genomic architecture, particularly the presence and characteristics of knob regions, presents a barrier to complete genome-wide tagging.

As previously shown, *Mu* elements exhibit a preference for targeting the 5' UTRs and transcription start sites of genes (Liang *et al.*, 2019; Marcon *et al.*, 2020), which corresponds to open chromatin signals (Liu *et al.*, 2009; Springer *et al.*, 2018). To analyze the epigenomic landscape around *Mu* elements, we aligned *BonnMu* insertions with open chromatin signals and histone modifications, obtained from published datasets (Makarevitch *et al.*, 2013; Zhang *et al.*, 2015; Ricci *et al.*, 2019; Hufford *et al.*, 2021). *BonnMu* insertion patterns were in line with transposase-accessible chromatin, as demonstrated by a comparison with ATAC-seq datasets (Figure 14; Ricci *et al.*, 2019). Both transposase accessible chromatin and *BonnMu* insertions showed a depletion at gene model midpoints, gradually increasing in frequency in 250 bp bins away from these midpoints. The opposite frequency was found for most of the DNA methylation marks, which were predominantly found at midpoints of gene models. This finding was previously reported by a meta-analysis of ATAC-seq signals (Ricci *et al.*, 2019), and using Micrococcal Nuclease (Rodgers-Melnick *et al.*, 2016) and DNase based assays (Oka *et al.*, 2017), respectively. Generally, the relationship between TEs and chromatin in maize has been shown to be markedly variable, with a complex interplay between DNA methylation, histone modifications and TEs impacting gene expression in the maize genome (West *et al.*, 2014; Zhao *et al.*, 2016; Noshay *et al.,* 2019; Ricci *et al.,* 2019). The identification of a pronounced preference for *Mu* elements targeting the 5' UTRs and transcription start sites, regions characterized by open chromatin (Liu *et al.*, 2009; Springer *et al.*, 2018; Marcon *et al.*, 2020; Zhang *et al.*, 2020), highlights the complex relationship between transposons and the epigenomic landscape. This relationship is further complicated by the interplay between different transposable element families and DNA methylation patterns in maize (Noshay *et al.*, 2019). The dynamic interplay suggests that *Mu* elements, through their insertion patterns, may not only contribute to the physical tagging of genes but also potentially influence the

regulatory regions that control gene expression. This dual role of *Mu* transposons, as both mutagens and modulators of epigenetic states, highlights the need for a nuanced understanding of their impact on the maize genome. Future studies should therefore delve deeper into the epigenetic consequences of *Mu* insertions, exploring how these elements might influence maize phenotypes through alterations in gene expression patterns. Such investigations could provide valuable insights into the mechanisms by which transposable elements contribute to plant adaptability and evolution, offering new avenues for maize improvement and breeding.

Functional genetics experiments using mutagenized maize stocks, such as *BonnMu* F_2-families (Marcon *et al.*, 2020), which were extended within this study or *UniformMu* stocks (McCarty *et al.*, 2013a) require the validation of *Mu* insertions by PCR-based genotyping (Liu *et al*., 2016). This method specifically amplifies DNA located between the highly conserved TIR sequence of the *Mu* element and adjacent regions of the genome. The TIR6 primer, which was generated based on the TIR sequences of *Mu1*, *Mu7*, *Mu3*, *Mu8*, and their variants (Settles *et al.*, 2004), is typically used to confirm *Mu* insertions (Figure 16). The TIR6 primer is mostly effective to amplify the different classes of autonomous and nonautonomous *Mu* elements (Liu *et al*., 2016). Among the classes of *Mu* elements, *Mu1*, *Mu8* and *MuDR* exhibit the highest copy numbers in the maize genome (Liu *et al.*, 2009). Here, we identified *Mu1*, *Mu8,* and *MuDR* elements in 30% of the *BonnMu* insertions (Figure 15 & 16; Supplemental figure S2; Win *et al*., 2024b), facilitating future genotyping of *BonnMu* F_2-families. For these three classes of *Mu* elements, specific primers can be designed to validate the insertions.

In summary, the *BonnMu* resource has undergone significant expansion, providing a comprehensive assortment of Mu-seq libraries that encompass diverse genetic backgrounds. The genetic diversity represented by the different genetic backgrounds in *BonnMu* enables the identification of genotype-specific mutations in the future. The ability to tag and identify almost every gene in the maize genome is particularly noteworthy, as it enhances the resource's utility for researchers conducting functional genomic studies.

6.2 Application of *BonnMu* as a forward and reverse genetic resource and identification of *mrd1* mutant

Forward genetics starts by identifying a mutant phenotype and then identifying the underlying gene to examine the gene function. In contrast, reverse genetics begins with a known gene and looks for a mutant affected by the gene to confirm the function (Candela and Hake, 2008). The *BonnMu* repository, serving as both a forward and a reverse genetic tool, aids in conducting a thorough genetic analysis, ranging from elucidating gene function to identifying phenotypically uncharacterized mutants. Thus, this resource contributes to deepening our understanding of the genetic architecture and developmental processes of maize. Meanwhile, several *Mu*-induced mutants associated with maize root development have been identified and functionally characterized (Hochholdinger *et al.*, 2001, 2008; Wen *et al.*, 2005; Taramino *et al.*, 2007; von Behrens *et al.*, 2011; Li *et al.*, 2016). Mutant analyses have enabled the dissection of the genetic architecture underlying maize root development (Hochholdinger *et al*., 2018b).

We identified the *magenta root dwarf 1* (*mrd1*) mutant in a forward genetic screening of *BonnMu* F_2-families. In comparison to the wild type, *mrd1* mutants are characterized by a dwarfed shoot and an over-accumulation of magenta colored anthocyanins in primary, seminal, and crown roots (Figure 18). Only a small subset of the *BonnMu* F_2-families (49 out of 747) segregating for *mrd1* exhibited a Mendelian segregation ratio of 3:1. This could at least in part be explained by the fact that we phenotyped a limited number of seedlings per *BonnMu* F_2-family due to seed limitations. Moreover, several *mrd1 BonnMu* families showed poor or no germination during our forward genetic analysis. This is also consistent with the the delayed or poor germination observed in *csn* mutants in Arabidopsis (Jin *et al.*, 2018).

In this study, we have explored the morphological differences between *mrd1* mutants and wild type plants under both light and dark conditions. Our findings reveal a pronounced reduction in growth across several parameters in *mrd1* mutants, including primary root length, lateral root density, and shoot length, among others (Figure 19). Similarly, Arabidopsis *csn4* mutant also exhibit a reduced main root system, affecting the primary root and the lateral root formation compared to the wild type (Kwok *et al*., 1996; Pacurar *et al*., 2017). Similar to the Arabidopsis mutants which display smaller rosettes and dwarfism (Kwok *et al*., 1996; Pacurar *et al*., 2017), the *mrd1* maize mutants also showed significantly shorter first leaves with dwarf phenotype (Figure 19 E & F). In addition, the mutant displayed shorter hypocotyls in Arabidopsis (Pacurar *et al*. 2017). Similarly, in our study, the coleoptile and mesocotyl length

in maize *mrd1* mutants showed a comparable reduction (Figure G & H), highlighting a consistent effect across different monocot and dicot plant species.

Arabidopsis *cop9* null mutants exhibit a severe phenotype of constitutive photomorphogenesis and accumulation of anthocyanins in the embryo, culminating in early seedling lethality (Serino and Deng, 2003). Reduction-of-function lines for several CSN subunits have been generated in Arabidopsis. In addition to an altered photomorphogenic response, phenotypic alterations of CSN reduction-of-function plants include morphological abnormalities such as modified number of cotyledons, loss of leaf symmetry, and altered auxin responses (Peng *et al.*, 2001). The pleiotropic phenotypes in CSN reduction-of-function lines indicated a cellular activity of CSN affecting multiple pathways. Hence, reduced levels of CSN would impair many aspects of plant development. Maize *mrd1* mutants show similarities to Arabidopsis *csn* mutants such as the constitutive photomorphogenesis (Supplemental figure S4). However, anthocyanin accumulates predominantly in primary, seminal, and crown roots of *mrd1* mutants, shortly after their emergence (Figure 18A), while it occurs in the cotyledons in the dicot Arabidopsis mutant (Kwok *et al*., 1996; Pacurar *et al*., 2017). In the *mrd1* mutant, the root is pigmented with magenta color in both light and dark conditions, in contrast, the Arabidopsis mutant develops greenish coloration only in the light grown condition (Wei and Deng, 1992), which could be explained by the formation of chloroplasts instead of amyloplasts in the root cells (Chory and Peto, 1990; Deng and Quail, 1992).

Phenolic compounds, like phenylpropanoids, are widespread secondary metabolites in plants which are involved in key metabolic and physiological processes (Boudet, 2007). Large phenylpropanoid-derived groups of phenolic compounds are represented by lignols and colorless flavonoids. The biosynthesis of flavonoids, including anthocyanins, is induced in plants by exposure to abiotic and biotic stresses (Li and Ahammed, 2023). Several transcription factors that control the expression of genes involved in anthocyanin biosynthesis during stress conditions have been identified across various plant species (Sun *et al.*, 2018; An *et al.*, 2020). The color variation in plants, ranging from orange to blue-red, is determined by the type of anthocyanidin — such as pelargonidin, cyanidin, and malvidin — each modified differently and identified by specific absorption peaks within the visible spectrum from 494 nm to 510 nm (Ma *et al.*, 2021). In *mrd1*, the magenta color in the primary root appears 4-5 days after germination. The ultraperformance liquid chromatography coupled with a photodiode-array detector and high-resolution quadrupole time-of-flight mass spectrometry (UPLC-PDA-HR-QTOF-MS) confirmed that *mrd1* root presents cyanidin 3-O-glucoside and peonidin 3-O-

glucoside (Figure 20). These anthocyanins have been previously reported to be present in maize kernels (Castañeda-Ovando *et al*., 2009; Hu *et al*., 2020). Anthocyanins, belonging to the flavonoid family, are water-soluble pigments that impart purple, blue, and red hues to plant tissues, examples of which include cyanidin-3-O-glucoside (C3G), pelargonidin-3-O-glucoside (Pg-3-glc), peonidin-3-O-glucoside (P3G), and malvidin-3-O-glucoside (M3G). Not all anthocyanin compounds present could be identified through this preliminary analysis and the presence of additional, yet to be characterized, anthocyanins may also play significant roles in the mutant.

The presence of flavonoids in maize root exudates not only facilitates a beneficial microbial environment conducive to plant growth but also establishes a chemical barrier against pathogens (Cesco *et al*., 2010; Hu *et al*., 2018; Kudijordjie *et al*., 2019). This dual function underscores the potential of root exudates as a natural strategy for enhancing crop resilience and productivity. Understanding the role of anthocyanins, such as cyanidin 3-O-glucoside and peonidin 3-O-glucoside, in modulating root-microbe interactions and nutrient uptake efficiency could provide valuable insights into enhancing crop performance in the future.

The rhizosphere represents a unique soil environment in which complex biogeochemical processes and dynamic biological interactions occur that are mainly driven by plant roots via their secretory activities and soil microorganisms, designated the rhizosphere microbiota (Marschner, 2012). Maize root exudates contain a variety of metabolites, including substantial amounts of flavonoids that are secreted into the rhizosphere (Cesco *et al.*, 2010). Flavonoids have been considered as crucial root-released rhizosphere signal molecules modulating the interaction of legume roots with microbes (Hassan and Mathesius, 2012; Mierziak *et al*., 2014). Root flavonoids are essential for initiating symbiosis with rhizobia in legumes and also act as auxin transport regulators (Wasson *et al.*, 2006; Subramanian *et al*., 2007). This implies that flavonoid-mediated root–microbe interactions could modulate developmental processes also in the host plant (Mierziak *et al*., 2014). Thus, metabolic profiles of specific mutants and the analyses of the microbial composition in the respective rhizosphere could contribute to the mechanistic understanding of root–microbe interaction underlying root flavonoid homeostasis. Furthermore, such analyses could contribute to improve nutrient uptake and crop performance in the future. In recent years, next-generation sequencing technologies such as RNA sequencing (RNA-Seq) and 16S rRNA gene sequencing have been introduced to plant functional genomics and soil microbial metagenomics (Hatzenpichler *et al.*, 2020). Therefore, a combined strategy of reverse genetics and multi-omics study will pave a promising way to

identify the functional gene-encoded flavonoid homeostasis interacting with beneficial microbes underlying plant growth and fitness. Under nitrogen deficiency, lateral root development coordinates the interactions of the root system with flavonoid-dependent Oxalobacteraceae (Yu *et al*., 2021). It has been demonstrated that the flavonoid composition in maize roots, notably influenced by flavone synthase type I2 (FNSI2), correlates with variations in rhizosphere microbiota composition. The maize genotypes with root exudates rich in specific flavonoids like apigenin and luteolin modulate bacterial community structure in the rhizosphere (Yu *et al*., 2021). The investigation of the *mrd1* mutant's role in flavonoid biosynthesis and its subsequent impact on rhizosphere interactions adds a significant layer to our understanding of plant-microbe dynamics. We examined that some bacterial genera were enriched in the rhizosphere of *mrd1* mutant such as *Lautropia*, *Haliangium*, *Dongia*, *Pseudolabrys*, *Steroidobacter*, *Nitrosospira* and *Opitutus*. In contrast, three bacterial genera were enriched in the rhizosphere of the wild type such as *Luteimonas*, *Arenimonas* and *Thermomonas*. It has been reported that there was an increased denitrification with an increased *Lautropia* abundance (Pan *et al.*, 2020). However, the bacterial genera *Haliangium* were enriched which has an antifungal effect by producing haliangicin compound (Fudou *et al.*, 2001). The haliangicin is a novel beta-methoxyacrylate antibiotic isolated from *Haliangium luteum* that inhibits fungal growth through interference in the mitochondrial electron flow, particularly within the cytochrome b-c1 segment (Fudou *et al*., 2001). We also found the genera *Nitrosospira* to be enriched in the rhizosphere of the *mrd1* mutant. It was previously shown that the presence of *Nitrosospira* in the rhizosphere can increase nitrate nitrogen content, soil nitrification activity, and shoot dry weight, indicating its vital role in promoting nitrogen availability and plant growth (Li *et al.*, 2023). In comparison, the observed three genera *Luteimonas*, *Arenimonas* and *Thermomonas* in the wild type are from the same family Xanthomonadaceae. The enrichment of *Luteimonas* in the wild type indicates a distinct microbial interaction potentially influencing nitrogen cycling differently from the *mrd1* mutant. *Luteimonas*, comprised of many species in multiple habitats, is known for its unusual denitrification ability (Finkmann *et al*., 2000; Ulrich *et al*., 2022) and plant growth promotion (Xiao *et al*., 2017; Gu *et al*., 2020). *Arenimonas* can be found in various habitats including sand and soil substrates (Kwon *et al*., 2007; Han *et al*., 2020; Zhou *et al*., 2022). Little is known for *Thermomonas*-plant interaction but has been reported that those bacteria might inhibit the soybean cyst nematode (Yuan *et al*., 2021). Nonetheless, the accumulation of anthocyanins in the *mrd1* mutant may confer enhanced stress tolerance and influence rhizosphere microbial communities, potentially through anthocyanin-mediated signaling pathways. It would be

worthwhile to confirm and further analyze how *mrd1* root exudate-derived flavones contribute to changes in the rhizosphere microbiome in the future.

In addition, one of the most striking aspects of the *mrd1* mutant is its constitutive photomorphogenic phenotype (Supplemental figure S3). Typically, in many plants, including maize, dark-grown seedlings exhibit skotomorphogenesis. This developmental program is characterized by elongated hypocotyles/ mesocotyles, a phenotype attributed to the plant's natural instinct to reach light (Casal *et al*., 2014). In Arabidopsis, mutants in COP9 signalosome (CSN) subunits exhibit constitutive photomorphogenic phenotypes (Pacurar *et al.*, 2017) which is similar to the *mrd1* mutant, and this suggests that MRD1 can have a conserved role for the COP9 signalosome in modulating light responses across different plant species.

In order to identify the gene underlying the phenotype, we subjected *mrd1* mutant to BSR-seq mapping. The advantage of BSR-seq is that it does not only provide the map position of a gene underlying a mutant phenotype. It also provides the effect of a mutation in this gene on global gene expression patterns (Liu et al., 2012). In this study, we locate the causative gene to a 70 Mb interval on the short arm of chromosome 1 by allele frequency detection (Mansfeld and Grumet, 2018) in *mrd1* in comparison to wild type pool. To precisely identify the causative gene such as the *mrd1* gene identified within the large 70 Mb interval, it is essential to refine and narrow this interval by conducting backcrosses over 6-9 generations to reduce genetic background noise, increasing mutant and wild type pools for BSR-seq analysis (Liu *et al*., 2012; Li *et al*., 2016). Additionally, optimizing SNP window size could enhance mapping accuracy (Liu *et al*., 2012; Li *et al*., 2016).

In addition, the *BonnMu mrd1* family harbors 64 germinal *Mu* insertions, from which five insertions were detected to be located in the mapping interval. Remarkably, only one gene, Zm00001eb008060, exhibited significant downregulation in the BSR-seq data. The *mrd1* mutant phenotype was identified across 747 *BonnMu* insertional mutant families. Of these, only 49 families exhibited the expected 3:1 wild type to mutant segregation ratio. Notably, Zm00001eb008060 emerged as the gene most frequently tagged by *Mu* in 16 of these families. Additionally, out of all *BonnMu* families, 125 had insertions in this gene, regardless of whether the *mrd1* phenotype was observed or not. This could also be explained by the fact that the insertions in the intron of the *mrd1* gene were mostly confirmed with few Mu-seq reads (Marcon *et al*., 2020; Stöcker *et al*., 2022). In many cases only 4-5 reads support the insertion. In our read analysis we required at least 4 reads to call it a true *Mu*-insertion. Therefore, it is

likely that we omitted some of the insertions due to this cutoff. The consistency of this gene's association across multiple families and its significant downregulation in the mutant line provide a significant evidence that not only we confirmed the *mrd1* candidate gene but also highlight the utility of Mu-seq in combination with BSR-seq methods in genetic mapping and gene function studies.

The gene Zm00001eb008060 encodes a *cop9 signalosome complex subunit 4* (*csn4*). The COP9 signalosome (CSN) is a multiprotein complex, composed of eight distinct subunits, that was initially identified in *Arabidopsis thaliana* as a repressor of photomorphogenesis (Serino and Deng, 2003). The identification of Zm00001eb008060 as a candidate gene underlying the *mrd1* phenotype was verified through genotyping confirmation where the *mrd1* phenotype strictly co-segregates with the genotype. This genotypic confirmation is critical as it provides a direct link between the *Mu* insertion in the *csn4* gene and the *mrd1* phenotype. The precise co-segregation of the *Mu* insertion with the phenotype across the segregating population not only implicates *csn4* as the gene responsible for the *mrd1* phenotype but also demonstrates the utility of combining BSR-seq, Mu-seq, and traditional molecular genotyping techniques for genetic mapping and mutation validation. Additionally, the relative expression analysis of *mrd1* compared to the wild type shows a significant down-regulation of *mrd1* gene in the mutant which is usually the case if the gene is directly involved in the physiological process or pathway being disrupted. This observation in *mrd1* with *Mu* insertion can be possibly explained by the process of nonsense-mediated mRNA decay (NMD), a mechanism that detects premature stop codons, leading to the degradation of the affected mRNA and preventing the translation of shorter and non-functional proteins (Kervestin and Jacobson, 2012; Rayson *et al*., 2012). Moreover, the significant decrease in expression levels of the *mrd1* gene in mutants indicates that its *Mu* insertion in the *mrd1* gene leads to loss of function or reduction of function.

We constructed a phylogenetic tree of *mrd1* gene in maize against monocots like *Brachypodium distachyon*, *Hordeum vulgare*, *Oryza sativa*, and *Sorghum bicolor*, dicots such as *Arabidopsis thaliana* and the angiosperm *Physcomitrella patens*. The phylogenetic relationship exhibited the homology between maize MRD1 and Arabidopsis CSN4, providing a further confirmation that *csn4* is indeed the candidate gene of *mrd1* mutant as observed in the previous analyses. Moreover, identity and similarity analyses revealed that *mrd1* shares high identity and similarity with the homologous genes of the species examined (Supplemental

table S10). Shi *et al*., (2023) also showed that CSN genes are evolutionarily conserved in land plants.

Our transcriptome analysis identified 422 upregulated and 356 downregulated genes in the crown roots of the *mrd1* mutant. The upregulation of 422 genes suggests an activation of pathways that might compensate for or respond to the altered photomorphogenic growth observed in the dark. Conversely, the downregulation of 356 genes, including the *mrd1* gene, highlights potential targets of the mutation that directly contribute to the observed phenotype. In Arabidopsis, *cop9* mutant also exhibits high-level expression of light-inducible genes in the absence of light (Wei and Deng, 1992). The identification of light-responsive genes among the upregulated genes in *mrd1* plants supports the hypothesis that the *mrd1* mutation aberrantly activates certain pathways leading to the unique pigmentation and light responsive growth characteristics.

Previous studies on CSN have predominantly focused on model organisms like Arabidopsis (Wei and Deng, 2003), offering limited insights into the maize-specific genetic and molecular mechanisms. Therefore, this study not only contributes to filling this gap but also raises questions about the conservation and divergence of photomorphogenic pathways among different plant species. The impact of downregulated genes on the susceptibility to external stressors, such as an increased abundance in microbiome species (Figure 24) and the biosynthesis of crucial compounds like lignin and glutathione suggests a complex interplay between the *mrd1* mutation and the plant's ability to respond to environmental challenges. In conclusion, our comparative RNA-seq analysis has unveiled a complex network of transcriptional changes associated with the *mrd1* mutation in maize, offering novel insights into plant photomorphogenesis and pigmentation.

Furthermore, our preliminary subcellular localization experiments have indicated that MRD1 is localized in the nucleus and cytosol but also within the endoplasmic reticulum (ER). However, the COP9 complex in Arabidopsis resides in the nucleus (Chamovitz *et al.*, 1996). The ER localization of MRD1 suggests a possible involvement in processes such as protein folding, maturation, or trafficking, which are critical for cellular homeostasis and response to environmental stimuli (Liu and Li, 2019). This multifaceted localization implies that MRD1 could be playing a complex role in cellular signaling pathways, potentially gene regulation with cytoplasmic and ER-associated processes. However, it is crucial to consider that the fusion of fluorescent proteins at the N-terminus can obscure targeting signals, leading to erroneous localization (Wiemann *et al*., 2001; Palmer and Freemen, 2004). Similarly, C-terminal fusions

can disrupt localization signals at the protein's C-terminus (Huh *et al*., 2003). Thus, the precise localization of MRD1, particularly its association with the ER, necessitates further validation to rule out potential artifacts from fusion protein influences. The redox-based topology analysis (ReTA) visualizes the orientation of ER membrane proteins by fusing the roGFP2 to either the N- or the C-termini of the protein sequence (Brach *et al*., 2009). This will allow us to visualize the orientation and topology of ER membrane proteins of MRD1 through the transient expression of these fusion proteins in tobacco leaves.

In Arabidopsis, the loss of any subunit from the COP9 complex causes identical seedling lethal phenotypes and limits the analysis of its full physiological roles in adult plants (Gusmaroli *et al*., 2007). It has also been demonstrated that the two independent T-DNA insertions in the 1st intron and 12th exon of *csn4* gene in Arabidopsis led to seedling lethality (Pacurar *et al*., 2017). However, the viable *csn* mutants with only partial loss of the CSN function are available for CSN1, CSN2, CSN3, CSN4, CSN5 and CSN6 (Dohmann *et al*., 2005; Gusmaroli *et al*., 2007; Zhang *et al*., 2008; Stuttmann *et al*., 2009; Huang *et al*., 2013; Pacurar *et al*., 2017). The *mrd1* mutant allele with mutation in the PCI domain created by CRISPR/Cas9 genome editing in our study showed lethality, suggesting a possible knockout. This outcome underscores the potential to generating a new partially deficient knock-down mutant via RNA interference (RNAi; Kir *et al*., 2015) in the future to elucidate the role of this complex in maize.

7 Conclusion

Establishing the *BonnMu* sequence-indexed resource of transposon induced maize mutations identified 425,924 germinal *Mu* insertions affecting 36,612 (83%) of all annotated gene models. This surpasses prior gene coverage rates, firmly placing the *BonnMu* collection among the top reverse genetic resources. Mutagenizing North American Dent and European Flint germplasms underscores the potential to uncover genotype-specific mutations that could impact maize adaptation to various environmental conditions.

The analysis of *BonnMu* insertions revealed a clear preference for targeting 5' UTRs and regions near the transcription start sites of genes, commonly associated with open chromatin. This trend mirrors broader patterns observed in plant genomes, where insertion sites often correlate with epigenetic marks of transcriptionally active regions. In support with that *BonnMu* insertions align with open chromatin signals and mainly target gene-dense chromosome arms. In contrast, *Mu* insertions are less frequently identified in heterochromatic centromere regions, indicating that certain genomic regions, e.g. highly heterochromatic knob areas could inhibit *Mu* insertions due to their dense chromatin structure, which suppresses local recombination and transposon integration. Therefore, genes located in or near knob regions could be challenging or even impossible to tag with *Mu* transposons, explaining why the *BonnMu* resource is unlikely to achieve 100% gene coverage in the future.

In the second project, the potential of the *BonnMu* repository as a valuable forward genetic resource was demonstrated with the *mrd1* mutant, distinguished by magenta-colored roots and a dwarfed shoot stature. Although, pinpointing the causal gene underlying the mutant phenotype was challenging, due to the dozen insertions in the respective *BonnMu* F_2-family, the combined BSR-seq and Mu-seq approach identified *csn4* as a promising candidate gene. The functional plasticity of the *mrd1* mutant phenotype was extensively characterized, including anthocyanin profiling, comparative transcriptome and microbiome analyses. However, gene validation with an independent mutant allele generated by genome-editing via CRISPR/Cas9 failed initially, necessitating another attempt in the near future, such as employing a gene-silencing strategy via RNAi.

8 References

Aesaert, S., Impens, L., Coussens, G., Van Lerberge, E., Vanderhaeghen, R., Desmet, L., Vanhevel, Y., Bossuyt, S., Wambua, A. N., Van Lijsebettens, M., Inzé, D., De Keyser, E., Jacobs, T. B., Karimi, M., & Pauwels, L. (2022). Optimized transformation and gene editing of the B104 public maize inbred by improved tissue culture and use of morphogenic regulators. *Frontiers in Plant Science*, *13*, 883847.

Alexa, A., & Rahnenfuhrer, J. (2023). *topGO:* Enrichment analysis for gene ontology (p. doi:10.18129/B9.bioc.topGO, R package version 2.54).

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., Stevenson, D. K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C. C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D. E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W. L., Berry, C. C., & Ecker, J. R. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, *301*(5633), 653–657.

An, J. P., Wang, X. F., Zhang, X. W., Xu, H. F., Bi, S. Q., You, C. X., & Hao, Y. J. (2020). An apple MYB transcription factor regulates cold tolerance and anthocyanin accumulation and undergoes MIEL1-mediated degradation. *Plant Biotechnology Journal*, *18*(2), 337–353.

Ananiev, E. V., Phillips, R. L., & Rines, H. W. (1998). A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome knobs megatransposons? *Proceedings of the National Academy of Sciences of the United States of America*, *95*(18), 10785–10790.

Ashby, A. M., Watson, M. D., Loake, G. J., & Shaw, C. H. (1988). Ti plasmid-specified chemotaxis of *Agrobacterium tumefaciens* C58C^{1} toward vir-inducing phenolic compounds and soluble factors from monocotyledonous and dicotyledonous plants. *Journal of Bacteriology*, *170*(9), 4181–4187.

Baer, M., Taramino, G., Multani, D., Fengler, K., & Hochholdinger, F. (2023). Maize *lateral rootless 1* encodes a homolog of the DCAF protein subunit of the CUL4-based E3 ubiquitin ligase complex. *New Phytologist*, *237*(4), 1067–1473.

Bensen, R. J., Johal, G. S., Crane, V. C., Tossberg, J. T., Schnable, P. S., Meeley, R. B., & Briggs, S. P. (1995). Cloning and characterization of the maize *An1* gene. *The Plant Cell*, *7*(1), 75–84.

Betsch, L., Boltz, V., Brioudes, F., Pontier, G., Girard, V., Savarin, J., Wipperman, B., Chambrier, P., Tissot, N., Benhamed, M., Mollereau, B., Raynaud, C., Bendahmane, M., & Szécsi, J. (2019). TCTP and CSN4 control cell cycle progression and development by regulating CULLIN1 neddylation in plants and animals. *PLoS Genetics*, *15*(1), e1007899.

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., Mills, D. A., & Caporaso, J. G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, *10*(1), 57–59.

Bonnett, O. T. (1954). The inflorescences of maize. *Science*, *120*(3107), 77–87.

Boudet, A. M. (2007). Evolution and current status of research in phenolic compounds. *Phytochemistry*, *68*(22–24), 2722–2735.

Brach, T., Soyk, S., Müller, C., Hinz, G., Hell, R., Brandizzi, F., & Meyer, A. J. (2009). Non-invasive topology analysis of membrane proteins in the secretory pathway. *The Plant Journal*, *57*(3), 534–541.

Briggs, S., & Meeley, B. (1995). Reverse genetics for maize. *Maize Genetics Cooperation Newsletter*, 69, 67-82.

Brutnell, T. P., & Conrad, L. J. (2013). Transposon tagging using *Activator* (*Ac*) in maize. In E. Grotewold (Ed.), *Plant Functional Genomics, Methods in Molecular Biology™*, *236*, 157–175.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *10*(12), 1213–1218.

Carvalho, R. F., Aguiar-Perecin, M. L. R., Clarindo, W. R., Fristche-Neto, R., & Mondin, M. (2022). A heterochromatic knob reducing the flowering time in maize. *Frontiers in Genetics,* 12, 799681.

Camus-Kulandaivelu, L., Veyrieras, J. B., Madur, D., Combes, V., Fourmann, M., Barraud, S., Dubreuil, P., Gouesnard, B., Manicacci, D., & Charcosset, A. (2006). Maize adaptation

to temperate climate: relationship between population structure and polymorphism in the *Dwarf8* gene. *Genetics*, *172*(4), 2449–2463.

Candela, H., & Hake, S. (2008). The art and design of genetic screens: Maize. *Nature Reviews Genetics*, *9*(3), 192–203.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*, *108 Suppl 1*(Suppl 1), 4516–4522.

Carter, J. D., Pereira, A., Dickerman, A. W. & Veilleux R. E. (2013). An active *Ac*/*Ds* transposon systemfor activation tagging in tomato cultivar M82 using clonal propagation. *Plant Physiology*, *162*(1),145-156.

Casal, J. J., Candia, A. N., & Sellaro, R. (2014). Light perception and signalling by phytochrome A. *Journal of Experimental Botany*, *65*(11), 2835–2845.

Castañeda-Ovando, A., Pacheco-Hernández, M. d. L., Páez-Hernández, M. E., Rodríguez, J. A., & Galán-Vidal, C. A. (2009). Chemical studies of anthocyanins: A review. *Food Chemistry*, *113*(4), 859-871.

Castle, L. A., & Meinke, D. W. (1994). A FUSCA gene of Arabidopsis encodes a novel protein essential for plant development. *The Plant Cell*, *6*(1), 25–41.

Cenk, E., Schmutz, C., Pahlke, G., Oertel, A., Kollarova, J., Mock, H., Matros, A., & Marko, D. (2021). Immunomodulatory properties of blackberry anthocyanins in THP-1 derived macrophages. *International Journal of Molecular Sciences*, *22*(19), 10483.

Cesco, S., Neumann, G., Tomasi, N., Pinton, R., & Weisskopf, L. (2010). Release of plant-borne flavonoids into the rhizosphere and their role in plant nutrition. *Plant and Soil*, *329*(1), 1–25.

Chamovitz, D. A., Wei, N., Osterlund, M. T., Von Arnim, A. G., Staub, J. M., Matsui, M., & Deng, X. W. (1996). The COP9 complex, a novel multisubunit nuclear regulator involved in light control of a plant developmental switch. *Cell*, *86*(1), 115–121.

Chen, J., Wang, Z., Tan, K., Huang, W., Shi, J., Li, T., Hu, J., Wang, K., Wang, C., Xin, B., Zhao, H., Song, W., Hufford, M. B., Schnable, J. C., Jin, W., & Lai, J. (2023). A complete telomere-to-telomere assembly of the maize genome. *Nature Genetics*, *55*(7), 1221–

1231.

Chen, X., Zhang, B., Wang, T., Wang, T., Bonni, A., & Zhao, G. (2020). Robust principal component analysis for accurate outlier sample detection in RNA-Seq data. *BMC Bioinformatics*, *21*, 269 (2020).

Chory, J., Peto, C. A., Ashbaugh, M., Saganich, R., Pratt, L., & Ausubel, F. (1989). Different roles for phytochrome in etiolated and green plants deduced from characterization of *Arabidopsis thaliana* mutants. *The Plant Cell*, *1*(9), 867.

Coe, E. H. (2001). The origins of maize genetics. *Nature Reviews Genetics*, *2*(11), 898–905.

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938–2940.

Cooter, E., Dodder, R., Bash, J., Elobeid, A., & Yang, D. (2017). Exploring a united states maize cellulose biofuel scenario using an integrated energy and agricultural markets solution approach. *EPA Public Access*, *2*(2), 1031.

CSHL/ The Cold Spring Harbor Laboratory (2000).The complete sequence of a heterochromatic island from a higher eukaryote. *Cell*, *100*(3), 377–386.

Dai, D., Jin, L., Huo, Z., Yan, S., Ma, Z., Qi, W., & Song, R. (2021). Maize pentatricopeptide repeat protein DEK53 is required for mitochondrial RNA editing at multiple sites and seed development. *Journal of Experimental Botany*, *71*(20), 6246–6261.

Dawe, R. K., & Hiatt, E. N. (2004). Plant neocentromeres: fast, focused, and driven. *Chromosome Research*, *12*(6), 655–669.

DMK. (2022). Deutschen Maiskomitee e.V. (DMK). https://www.maiskomitee.de/

Dietrich, C. R., Cui, F., Packila, M. L., Li, J., Ashlock, D. A., Nikolau, B. J., & Schnable, P. S. (2002). Maize *Mu* transposons are targeted to the 5' untranslated region of the *gl8* gene and sequences flanking *Mu* target-site duplications exhibit nonrandom nucleotide composition throughout the genome. *Genetics*, *160*(2), 697–716.

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, *34*(2), 184–191.

Dohmann, E. M., Kuhnle, C., & Schwechheimer, C. (2005). Loss of the CONSTITUTIVE PHOTOMORPHOGENIC9 signalosome subunit 5 is sufficient to cause the *cop*/*det*/*fus* mutant phenotype in Arabidopsis. *The Plant Cell*, *17*(7), 1967–1978.

Eckardt, N. A. (2003). Characterization of the last subunit of the Arabidopsis COP9 signalosome. *The Plant Cell*, *15*(3), 580–581.

Edgar R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*(10), 996–998.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*(16), 2194–2200.

Erenstein, O., Chamberlin, J., & Sonder, K. (2021). Estimating the global number and distribution of maize and wheat farms. *Global Food Security*, *30*(7), 100558.

FAO. (2021). FAOSTAT analytical brief. *Agricultural Production Statistics 2000-2021*, *60*, 1–17.

Fernandes, J., Dong, Q., Schneider, B., Morrow, D., Nan, G., Brendel, V. & Walbot, V. (2004). Genome-wide mutagenesis of *Zea mays* L. using *RescueMu* transposons. *Genome Biology*, *5*(10), pp.R82.1-R82.20.

Feschotte, C., Jiang, N., & Wessler, S. R. (2002). Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics*, *5*(3), 329–341.

Finkmann, W., Altendorf, K., Stackebrandt, E., & Lipski, A. (2000). Characterization of N_2O-producing *Xanthomonas*-like isolates from biofilters as *Stenotrophomonas nitritireducens* sp. nov., *Luteimonas mephitis* gen. nov., sp. nov. and *Pseudoxanthomonas broegbernensis* gen. nov., sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, *50*(1), 273–282.

Fudou, R., Iizuka, T., Sato, S., Ando, T., Shimba, N., & Yamanaka, S. (2001). Haliangicin, a novel antifungal metabolite produced by a marine myxobacterium. 2. Isolation and structural elucidation. *Journal of Antibiotics*, *54*(2), 153–156.

Gage, J. L., Monier, B., Giri, A., & Buckler, E. S. (2020). Ten years of the maize nested association mapping population: impact, limitations, and future directions. *The Plant Cell*, *32*(7), 2083–2093.

Ghaffari, R., Cannon, E. K. S., Kanizay, L. B., Lawrence, C. J., & Dawe, R. K. (2013). Maize chromosomal knobs are located in gene-dense areas and suppress local recombination. *Chromosoma*, *122*(1–2), 67–75.

Gould, K. S. (2004). Nature's Swiss army knife: the diverse protective roles of anthocyanins in leaves. *Journal of Biomedicine and Biotechnology*, *2004*(5), 314–320.

Grote, U., Fasse, A., Nguyen, T. T., & Erenstein, O. (2021). Food security and the dynamics of wheat and maize value chains in africa and asia. *Frontiers in Sustainable Food Systems*, *4*, 617009.

Gu, Y., Dong, K., Geisen, S., Yang, W., Yan, Y., Gu, D., Liu, N., Borisjuk, N., Luo, Y., & Friman, V.-P. (2020). The effect of microbial inoculant origin on the rhizosphere bacterial community composition and plant growth-promotion. *Plant Soil*, *452*, 105–117.

Gusmaroli, G., Figueroa, P., Serino, G., & Deng, X. W. (2007). Role of the MPN subunits in COP9 signalosome assembly and activity, and their regulatory interaction with Arabidopsis Cullin3-based E3 ligases. *The Plant Cell*, *19*(2), 564–581.

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S. K., Sodergren, E., Methé, B., DeSantis, T. Z., Human Microbiome Consortium, Petrosino, J. F., Knight, R., & Birren, B. W. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*, *21*(3), 494–504.

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

Hake, S., & Ross-Ibarra, J. (2015). The natural history of model organisms: genetic, evolutionary and plant breeding insights from the domestication of maize. *eLife*, *4*, e05861.

Hall, T. A. (1999). A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Hallauer, A. R. (2000). Specialty corns. In *Specialty Corns*. 289–303.

Han, D. M., Chun, B. H., Kim, H. M., Khan, S. A., & Jeon, C. O. (2020). *Arenimonas terrae* sp.

nov., isolated from orchard soil. *International Journal of Systematic and Evolutionary Microbiology*, *70*(1), 537–542.

Hart, J. P., & Lovis, W. A. (2013). Reevaluating what we know about the histories of maize in Northeastern North America: a review of current evidence. *Journal of Archaeological Research*, *21*(2), 175–216.

Hassan, S., & Mathesius, U. (2012). The role of flavonoids in root-rhizosphere signalling: Opportunities and challenges for improving plant-microbe interactions. *Journal of Experimental Botany*, *63*(9), 3429–3444.

Hatzenpichler, R., Krukenberg, V., Spietz, R. L., & Jay, Z. J. (2020). Next-generation physiology approaches to study microbiome function at single cell level. *Nature Reviews Microbiology*, *18*(4), 241–256.

He, X., Wang, D., Jiang, Y., Li, M., Delgado-Baquerizo, M., McLaughlin, C., Marcon, C., Guo, L., Baer, M., Moya, Y. A. T., von Wirén, N., Deichmann, M., Schaaf, G., Piepho, H.-P., Yang, Z., Yang, J., Yim, B., Smalla, K., Goormachtig, S., de Vries, F. T., Hüging, H., Baer, M., Sawers, R. J. H., Reif, J. C., Hochholdinger, F., Chen, X., & Yu, P. (2024). Heritable microbiome variation is correlated with source environment in locally adapted maize varieties. *Nature Plants*, in press.

Hecker, A., Wallmeroth, N., Peter, S., Blatt, M.R., Harter, K. & Grefen, C. (2015). Binary 2in1 vectors improve in planta (co)localization and dynamic protein interaction studies. *Plant Physiology*, *168*(3), 776–787.

Hetz, W., Hochholdinger, F., Schwall, M., & Feix, G. (1996). Isolation and characterization of rtcs, a maize mutant deficient in the formation of nodal roots. *The Plant Journal*, *10*(5), 845–857.

Hochholdinger, F., Park, W. J., Feix, H., & Feix, G. H. (2001). Cooperative action of *SLR1* and *SLR2* is required for lateral root-specific cell elongation in maize. *Plant Physiology*, *125*(3), 1529–1539.

Hochholdinger, F., & Tuberosa, R. (2009). Genetic and genomic dissection of maize root development and architecture. *Current Opinion in Plant Biology*, *12*(2), 172–177.

Hochholdinger, F., Wen, T. J., Zimmermann, R., Chimot-Marolle, P., Da Costa E Silva, O., Bruce, W., Lamkey, K. R., Wienand, U., & Schnable, P. S. (2008). The maize (*Zea mays*

L.) *roothairless3* gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. *Plant Journal*, *54*(5), 888–898.

Hochholdinger, F., Woll, K., Sauer, M., & Dembinsky, D. (2004). Genetic dissection of root formation in maize (*Zea mays*) reveals root-type specific developmental programmes. *Annals of Botany*, *93*(4), 359–368.

Hochholdinger, F., Marcon, C., Baldauf, J. A., Yu, P., & Frey, F. P. (2018a). Proteomics of maize root development. *Frontiers in Plant Science*, *9*, 143.

Hochholdinger, F., Yu, P., & Marcon, C. (2018b). Genetic control of root system development in maize. *Trends in Plant Science*, *23*(1), 79–88.

Hu L., Robert, C. A. M., Cadot, S., Zhang, X., Ye, M., Li, B., Manzo, D, Chervet, N., Steinger, T., van der Heijden, M. G. A., Schlaeppi, K. & Erb, M. (2018) Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota. *Nature Communication*, *9*(1), 2738.

Hu, X., Liu, J., Li, W., Wen, T., Li, T., Guo, X.-B., & Liu, R. H. (2020). Anthocyanin accumulation, biosynthesis and antioxidant capacity of black sweet corn (*Zea mays* L.) during kernel development over two growing seasons. *Journal of Cereal Science*, *95*, 103065.

Huang, H., Quint, M., & Gray, W. M. (2013). The *eta7*/*csn3-3* auxin response mutant of Arabidopsis defines a novel function for the CSN3 subunit of the COP9 signalosome. *PloS ONE*, *8*(6), e66578.

Hufford, M. B., Seetharam, A. S., Woodhouse, M. R., Chougule, K. M., Ou, S., Liu, J., Ricci, W. A., Guo, T., Olson, A., Qiu, Y., Della Coletta, R., Tittes, S., Hudson, A. I., Marand, A. P., Wei, S., Lu, Z., Wang, B., Tello-Ruiz, M. K., Piri, R. D., Wang, N., Kim, D. W., Zeng, Y., O'Connor, C. H., Li, X., Gilbert, A. M., Baggs, E., Krasileva, K. V., Portwood, J. L., 2nd, Cannon, E. K. S., Andorf, C. M., Manchanda, N., Snodgrass, S. J., Hufnagel, D. E., Jiang, Q., Pedersen, S., Syring, M. L., Kudrna, D. A., Llaca, V., Fengler, K., Schmitz, R. J., Ross-Ibarra, J., Yu, J., Gent, J. I., Hirsch, C. N., Ware, D., & Dawe, R. K. (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science*, *373*(6555), 655–662.

Hufford, M. B., Xu, X., van Heerwaarden, J., Pyhäjärvi, T., Chia, J. M., Cartwright, R. A., Elshire, R. J., Glaubitz, J. C., Guill, K. E., Kaeppler, S. M., Lai, J., Morrell, P. L., Shannon,

L. M., Song, C., Springer, N. M., Swanson-Wagner, R. A., Tiffin, P., Wang, J., Zhang, G., Doebley, J., McMullen, M. D., Ware, D., Buckler, E. S., Yang, S., & Ross-Ibarra, J. (2012). Population genomics of domestication and improvement in maize. *Nature Genetics*, *44*(7), 808–811.

Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., & O'Shea, E. K. (2003). Global analysis of protein localization in budding yeast. *Nature*, *425*(6959), 686–691.

Hunter C. T. (2021). CRISPR/Cas9 targeted mutagenesis for functional genetics in maize. *Plants* (*Basel, Switzerland*), *10*(4), 723.

Hunter, C. T., Suzuki, M., Saunders, J., Wu, S., Tasi, A., McCarty, D. R., & Koch, K. E. (2014). Phenotype to genotype using forward-genetic Mu-seq for identification and functional classification of maize mutants. *Frontiers in Plant Science*, *4*, 545.

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., Campbell, M. S., Stein, J. C., Wei, X., Chin, C. S., Guill, K., Regulski, M., Kumari, S., Olson, A., Gent, J., Schneider, K. L., Wolfgruber, T. K., May, M. R., Springer, N. M., Antoniou, E., McCombie, W. R., Presting, G. G., McMullen, M., Ross-Ibarra, J., Dawe, R. K., Hastie, A., Rank, D. R., & Ware, D. (2017). Improved maize reference genome with single-molecule technologies. *Nature*, *546*(7659), 524–527.

Jin, D., Wu, M., Li, B., Bücker, B., Keil, P., Zhang, S., Li, J., Kang, D., Liu, J., Dong, J., Deng, X. W., Irish, V., & Wei, N. (2018). The COP9 Signalosome regulates seed germination by facilitating protein degradation of RGL2 and ABI5. *PLoS Genetics*, *14*(2), e1007237.

Kervestin, S., & Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. Nature reviews. *Molecular Cell Biology*, *13*(11), 700–712.

Kir, G., Ye, H., Nelissen, H., Neelakandan, A. K., Kusnandar, A. S., Luo, A., Inzé, D., Sylvester, A. W., Yin, Y., & Becraft, P. W. (2015). RNA interference knockdown of BRASSINOSTEROID INSENSITIVE1 in maize reveals novel functions for brassinosteroid signaling in controlling plant architecture. *Plant Physiology*, *169*(1), 826–839.

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., Bates, S. T., Bruns, T. D., Bengtsson-Palme, J., Callaghan, T. M., Douglas, B., Drenkhan, T., Eberhardt, U., Dueñas, M., Grebenc, T., Griffith, G. W., Hartmann, M., Kirk, P. M.,

Kohout, P., Larsson, E., Lindahl, B. D., Lücking, R., Martín, M. P., Matheny, P. B., Nguyen, N. H., Niskanen, T., Oja, J., Peay, K. G., Peintner, U., Peterson, M., Põldmaa, K., Saag, L., Saar, I., Schüßler, A., Scott, J. A., Senés, C., Smith, M. E., Suija, A., Taylor, D. L., Telleria, M. T., Weiss, M., & Larsson, K. H. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology*, *22*(21), 5271–5277.

Kudjordjie, E. N., Sapkota, R., Steffensen, S. K., Fomsgaard, I. S. & Nicolaisen, M. (2019). Maize synthesized benzoxazinoids affect the host associated microbiome. *Microbiome*, *7*(1), 59.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X : Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549.

Kwok, S. F., Piekos, B., Misera, S., & Deng, X. W. (1996). A complement of ten essential and pleiotropic arabidopsis *COP*/*DET*/*FUS* genes is necessary for repression of photomorphogenesis in darkness. *Plant Physiology*, *110*(3), 731–742.

Kwon, S. W., Kim, B. Y., Weon, H. Y., Baek, Y. K., & Go, S. J. (2007). *Arenimonas donghaensis* gen. nov., sp. nov., isolated from seashore sand. *International Journal of Systematic and Evolutionary Microbiology*, *57*(5), 954–958.

Launholt, D., Merkle, T., Houben, A., Schulz, A. & Grasser, K.D. (2006). Arabidopsis chromatin-associated HMGA and HMGB use different nuclear targeting signals and display highly dynamic localization within the nucleus. *The Plant Cell*, *18*(11), 2904–2918.

Lazo, G. R., Stein, P. A. & Ludwig, R. A. (1991). A DNA transformation–competent *Arabidopsis* genomic library in *Agrobacterium*. *Nature Biotechnology*, *9*(10), 963–967.

Li, K., Hao, Z., Chen, L., Sha, Y., Wang, E., Sui, X., & Mi, G. (2023). Conservation strip-till modifies rhizosphere ammonia-oxidizing archaea and bacteria, increases nitrate accumulation and promotes maize growth at grain filling stage. *Soil and Tillage Research*, *234*(2), 105821.

Li, L., Hey, S., Liu, S., Liu, Q., McNinch, C., Hu, H. C., Wen, T. J., Marcon, C., Paschold, A., Bruce, W., Schnable, P. S., & Hochholdinger, F. (2016). Characterization of maize *roothairless6* which encodes a D-type cellulose synthase and controls the switch from bulge formation to tip growth. *Scientific Reports*, *6*, 34395.

Li, Z., & Ahammed, G. J. (2023). Plant stress response and adaptation via anthocyanins: A review. *Plant Stress*, *10*, 100230.

Liang, L., Wang, Y., Han, Y., Chen, Y., Li, M., Wu, Y., Ma, Z., Zhao, H., & Song, R. (2024). Expansion and improvement of *ChinaMu* by MuT-seq and chromosome-level assembly of the *Mu*-starter genome. J*ournal of Integrative Plant Biology*, 10.1111/jipb.13637.

Liang, L., Zhou, L., Tang, Y., Li, N., Song, T., Shao, W., Zhang, Z., Cai, P., Feng, F., Ma, Y., Yao, D., Feng, Y., Ma, Z., Zhao, H., & Song, R. (2019). A sequence-indexed *Mutator* insertional library for maize functional genomics study. *Plant Physiology*, *181*(4), 1404–1414.

Lisch, D. (2002). *Mutator* transposons. *Trends in Plant Science*, *7*(11), 498–504.

Lisch, D. (2015). *Mutator* and *MULE* transposons. *Microbiology Spectrum*, *3*(2), MDNA3-0032–2014.

Liu, L., & Li, J. (2019). Communications between the endoplasmic reticulum and other organelles during abiotic stress response in plants. *Frontiers in Plant Science*, *10*, 749.

Liu, P., McCarty, D. R., & Koch, K. E. (2016). Transposon mutagenesis and analysis of mutants in *UniformMu* maize (*Zea mays*). *Current Protocols in Plant Biology*, *1*(3), 451–465.

Liu, S., Yeh, C. T., Ji, T., Ying, K., Wu, H., Tang, H. M., Fu, Y., Nettleton, D., & Schnable, P. S. (2009). *Mu* transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS Genetics*, *5*(11), e1000733.

Liu, S., Yeh, C. T., Tang, H. M., Nettleton, D., & Schnable, P. S. (2012). Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PloS ONE*, *7*(5), e36406.

Lynch, J. (1995). Root architecture and plant productivity. *Plant Physiology*, *109*(1), 7–13.

Ma, Y., Ma, X., Gao, X., Wu, W., & Zhou, B. (2021). Light induced regulation pathway of anthocyanin biosynthesis in plants. *International Journal of Molecular Sciences*, *22*(20), 11116.

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957–2963.

Magwene, P. M., Willis, J. H., & Kelly, J. K. (2011). The statistics of bulk segregant analysis using next generation sequencing. *PLoS Computational Biology*, *7*(11), e1002255.

Makarevitch, I., Eichten, S. R., Briskine, R., Waters, A. J., Danilevskaya, O. N., Meeley, R. B., Myers, C. L., Vaughn, M. W., & Springer, N. M. (2013). Genomic distribution of maize facultative heterochromatin marked by trimethylation of H3K27. *The Plant Cell*, *25*(3), 780–793.

Mansfeld, B. N., & Grumet, R. (2018). QTLseqr : An R package for bulk segregant analysis with next-generation sequencing. *Plant Genome*, *11*(2), 1–5.

Marcon, C., Altrogge, L., Win, Y. N., Stöcker, T., Gardiner, J. M., Portwood, J. L., Opitz, N., Kortz, A., Baldauf, J. A., Hunter, C. T., McCarty, D. R., Koch, K. E., Schoof, H., & Hochholdinger, F. (2020). *BonnMu*: a sequence-indexed resource of transposon-induced maize mutations for functional genomics studies. *Plant Physiology*, *184*(2), 620–631.

Marcon, C., Brox, A., Win, Y. N., Stöcker, T., Du, X., Schoof, H., & Hochholdinger, F. (2024a). Identification of transposon insertion sites in F_2-stocks of the *BonnMu* resource. *Cold Spring Harbor Protocols*, *in press*.

Marcon, C., Win, Y. N., Du, X., & Hochholdinger, F. (2024b). *BonnMu*: a resource for functional genetics in maize (*Zea mays* L.) – Identification and molecular analysis of *Mutator* transposon insertions. *Cold Spring Harbor Protocols*, *in press.*

Marschner, P. (2012). Chapter 15 - Rhizosphere Biology. In Marschner's mineral nutrition of higher plants. In *Mineral nutrition of higher plants* (3rd ed.).

Martienssen, R. A. (1998). Functional genomics: probing plant gene function and expression with transposons. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(5), 2021–2026.

Masuka, B., Atlin, G. N., Olsen, M., Magorokosho, C., Labuschagne, M., Crossa, J., Bänziger, M., Pixley, K. V., Vivek, B. S., Von Biljon, A., Macrobert, J., Alvarado, G., Prasanna, B. M., Makumbi, D., Tarekegne, A., Das, B., Zaman-Allah, M., & Cairns, J. E. (2017). Gains in maize genetic improvement in eastern and southern Africa: I. CIMMYT hybrid breeding pipeline. *Crop Science*, *57*(1), 168–179.

Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez G., J., Buckler, E., & Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping.

Proceedings of the National Academy of Sciences, *99*(9), 6080–6084.

May, B. P., Liu, H., Vollbrecht, E., Senior, L., Rabinowicz, P. D., Roh, D., Pan, X., Stein, L., Freeling, M., Alexander, D., & Martienssen, R. (2003). Maize-targeted mutagenesis: A knockout resource for maize. *Proceedings of the National Academy of Sciences*, *100*(20), 11541–11546.

McCarty, D. R., Latshaw, S., Wu, S., Suzuki, M., Hunter, C. T., Avigne, W. T., & Koch, K. E. (2013a). Mu-seq: Sequence-based mapping and identification of transposon induced mutations. *PLoS ONE*, *8*(10), e77172.

McCarty, D. R., Mark Settles, A., Suzuki, M., Tan, B. C., Latshaw, S., Porch, T., Robin, K., Baier, J., Avigne, W., Lai, J., Messing, J., Koch, K. E., & Curtis Hannah, L. (2005). Steady-state transposon mutagenesis in inbred maize. *Plant Journal*, *44*(1), 52–61.

McCarty, D. R., Suzuki, M., Hunter, C., Collins, J., Avigne, W. T., & Koch, K. E. (2013b). Genetic and molecular analyses of *UniformMu* transposon insertion lines. *Methods in Molecular Biology*, *1057*, 157–166.

McClintock, B. (1951). Chromosome organization and genic expression. *Cold Spring Harbor Laboratory (CSHL) Symposium on Quantitative Biology*, *16*,13–47.

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS One*, *8*(4), e61217.

Merrill, W. L., Hard, R. J., Mabry, J. B., Fritz, G. J., Adams, K. R., Roney, J. R., & MacWilliams, A. C. (2009). The diffusion of maize to the southwestern United States and its impact. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(50), 21019–21026.

Michelomore, R., Paran, I., & Kesseli, R. V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the United States of America*, *88*, 9828–9832.

Mierziak, J., Kostyn, K., & Kulma, A. (2014). Flavonoids as important molecules of plant interactions with the environment. *Molecules*, *19*(10), 16240–16265.

Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K., & Hirochika, H. (2003). Target site specificity of the *Tos17* retrotransposon shows a

preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *The Plant Cell*, *15*(8), 1771–1780.

Mol, J., Grofewold, E., & Koes, R. (1998). How genes paint flowers and seeds. *Trends in Plant Science*, *3*(6), 212–217.

Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J. D., Fernandez, J. P., Mis, E. K., Khokha, M. K., & Giraldez, A. J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nature Methods*, *12*(10), 982–988.

Morris, M., & Hill, A. (2006). Ethanol opportunities and questions. *ATTRA, www.attra.ncat.org*, 1–16.

Nalini, E., Bhagwat, S., & Jawali, N. (2003). A simple method for isolation of DNA from plants suitable for long-term storage and DNA marker analysis. *BARC Newsletter*, *249*, 208–214.

Nestler, J., Liu, S., Wen, T. J., Paschold, A., Marcon, C., Tang, H. M., Li, D., Li, L., Meeley, R. B., Sakai, H., Bruce, W., Schnable, P. S., & Hochholdinger, F. (2014). *Roothairless5*, which functions in maize (*Zea mays* L.) root hair initiation and elongation encodes a monocot-specific NADPH oxidase. *Plant Journal*, *79*(5), 729–740.

Noshay, J. M., Anderson, S. N., Zhou, P., Ji, L., Ricci, W., Lu, Z., Stitzer, M. C., Crisp, P. A., Hirsch, C. N., Zhang, X., Schmitz, R. J., & Springer, N. M. (2019). Monitoring the interplay between transposable element families and DNA methylation in maize. *PLoS Genetics*, *15*(9), e1008291.

Nawy, T., Bayer, M., Mravec, J., Friml, J., Birnbaum, K. D., & Lukowitz, W. (2010). The GATA factor HANABA TARANU is required to position the proembryo boundary in the early Arabidopsis embryo. *Developmental Cell*, *19*(1), 103–113.

Oka, R., Zicola, J., Weber, B., Anderson, S. N., Hodgman, C., Gent, J. I., Wesselink, J. J., Springer, N. M., Hoefsloot, H. C. J., Turck, F., & Stam, M. (2017). Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biology*, *18*(1), 137.

Osthoff, A., Donà, P., Baldauf, J. A., Piepho, H., & Hochholdinger, F. (2019). Transcriptomic reprogramming of barley seminal roots by combined water deficit and salt stress. *BMC Genomics*, *20*(1), 325.

Ou S., Collins T., Qiu Y., Seetharam A., Menard C., Manchanda N., Gent J., Schatz M., Anderson S., Hufford M., & Hirsch C. (2022). Differences in activity and stability drive transposable element variation in tropical and temperate maize. *bioRxiv: 2022.10.09.511471v1*. In revision.

Pacurar, D. I., Pacurar, M. L., Lakehal, A., Pacurar, A. M., Ranjan, A., & Bellini, C. (2017). The Arabidopsis Cop9 signalosome subunit 4 (CNS4) is involved in adventitious root formation. *Scientific Reports*, *7*(1), 628.

Pahlke, G., Ahlberg, K., Oertel, A., Janson-schaffer, T., Grabher, S., Mock, H., Matros, A., & Marko, D. (2021). Antioxidant effects of elderberry anthocyanins in human colon carcinoma cells : A study on structure – activity relationships. *Molecular Nutrition*, *65*(17), e2100229.

Palmer, E., & Freeman, T. (2004). Investigation into the use of C- and N-terminal GFP fusion proteins for subcellular localization studies using reverse transfection microarrays. *Comparative and Functional Genomics*, *5*(4), 342–353.

Pan, H., Qin, Y., Wang, Y., Liu, S., Yu, B., Song, Y., Wang, X., & Zhu, G. (2020). Dissimilatory nitrate/nitrite reduction to ammonium (DNRA) pathway dominates nitrate reduction processes in rhizosphere and non-rhizosphere of four fertilized farmland soil. *Environmental Research*, *186*, 109612.

Parks, B. M., Shanklin, J., Koornneef, M., Kendrick, R. E., & Quail, P. H. (1989). Immunochemically detectable phytochrome is present at normal levels but is photochemically nonfunctional in the *hy 1* and *hy 2* long hypocotyl mutants of Arabidopsis. *Plant Molecular Biology*, *12*(4), 425–437.

Peng, Z., Serino, G., & Deng, X. W. (2001). Molecular characterization of subunit 6 of the COP9 signalosome and its role in multifaceted developmental processes in Arabidopsis. *The Plant Cell*, *13*(11), 2393–2407.

Piperno, D. R., Ranere, A. J., Holst, I., Iriarte, J., & Dickau, R. (2009). Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Sciences*, *106*(13), 5019–5024.

Pauwels, L., De Clercq, R., Goossens, J., Iñigo, S., Williams, C., Ron, M., Britt, A., & Goossens, A. (2018). A dual sgRNA approach for functional genomics in *Arabidopsis thaliana*. *G3 Genes|Genomes|Genetics*, *8*(8), 2603–2615.

Pruitt, J. D. (2016). A brief history of corn : Looking back to move forward. *UNL Doctor of Plant Health Doctoral Document*.

Qin, N., Xu, D., Li, J., & Deng, X. W. (2020). COP9 signalosome: Discovery, conservation, activity, and function. *Journal of Integrative Plant Biology*, *62*(1), 90–103.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(Database issue), D590–D596.

R Core Team. (2021). R: A language and environment for statistical computing. In *R Foundation for Statistical Computing, Vienna, Austria.*

Rayson, S., Arciga-Reyes, L., Wootton, L., De Torres Zabala, M., Truman, W., Graham, N., Grant, M., & Davies, B. (2012). A role for nonsense-mediated mRNA decay in plants: pathogen responses are induced in *Arabidopsis thaliana* NMD mutants. *PloS ONE*, *7*(2), e31917.

Rebourg, C., Chastanet, M., Gouesnard, B., Welcker, C., Dubreuil, P., & Charcosset, A. (2003). Maize introduction into Europe: The history reviewed in the light of molecular data. *Theoretical and Applied Genetics*, *106*(5), 895–903.

Rhoades, M. M. R. (1984). The early years of maize genetics. *Annual Review of Genetics*, *18*, 1–29.

Ricci, W. A., Lu, Z., Ji, L., Marand, A. P., Ethridge, C. L., Murphy, N. G., Noshay, J. M., Galli, M., Mejía-guerra, M. K., Johannes, F., Rowley, M. J., Corces, V. G., Scanlon, M. J., Buckler, E. S., Gallavotti, A., & Nathan, M. (2019). Widespread long-range cis-regulatory elements in the maize genome. *Nature Plants*, *5*(12), 1237–1249.

Robertson, D. S. (1978). Characterization of a mutator system in maize. *Mutation Research*, *51*(1), 21–28.

Robertson, D. S. (1983). A possible dose-dependent inactivation of *Mutator* (*Mu*) in maize. *Mol Gen Genet*, *191*, 86–90.

Rodgers-Melnick, E., Vera, D. L., Bass, H. W., & Buckler, E. S. (2016). Open chromatin reveals the functional maize genome. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(22), 3177–3184.

Rodrigues S. D., Karimi, M., Impens, L., Van Lerberge, E., Coussens, G., Aesaert, S., Rombaut, D., Holtappels, D., Ibrahim, H. M. M., Van Montagu, M., Wagemans, J., Jacobs, T. B., De Coninck, B. & Pauwels, L. (2021). Efficient CRISPR-mediated base editing in *Agrobacterium spp*. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(2):e2013338118..

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572–1574.

Rosegrant, M. W., Ringler, C., Sulser, T. B., Ewing, M., Palazzo, A., Zhu, T., Nelson, G. C., Koo, J., Robertson, R., Msangi, S., & Batka, M. (2009). Agriculture and food security under global change : Prospects for 2025 / 2050. *International Food Policy Research Institute, Washinton, DC.*

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467.

Schlegel, R. H. J. (2017). History of plant breeding. In *History of Plant Breeding* (1st ed.).

Schnable, J. C. & Freeling, M. (2011). Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. *PLoS ONE*, *6*(3), e17855.

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T. A., Minx, P., Reily, A. D., Courtney, L., Kruchowski, S. S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S. M., Belter, E., Du, F., Kim, K., Abbott, R. M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S. M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M. J., McMahan, L., Van Buren, P., Vaughn, M. W.,

Ying, K., Yeh, C. T., Emrich, S. J., Jia, Y., Kalyanaraman, A., Hsia, A. P., Barbazuk, W. B., Baucom, R. S., Brutnell, T. P., Carpita, N. C., Chaparro, C., Chia, J. M., Deragon, J. M., Estill, J. C., Fu, Y., Jeddeloh, J. A., Han, Y., Lee, H., Li, P., Lisch, D. R., Liu, S., Liu, Z., Nagel, D. H., McCann, M. C., SanMiguel, P., Myers, A. M., Nettleton, D., Nguyen, J., Penning, B. W., Ponnala, L., Schneider, K. L., Schwartz, D. C., Sharma, A., Soderlund, C., Springer, N. M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T. K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J. L., Dawe, R. K., Jiang, J., Jiang, N., Presting, G. G., Wessler, S. R., Aluru, S., Martienssen, R. A., Clifton, S. W., McCombie, W. R., & Wing, R. A. (2009). The B73 maize genome: Complexity, diversity, and dynamics. *Science*, *326*(5956), 1112–1115.

Schwechheimer, C. (2004). The COP9 signalosome (CSN): An evolutionary conserved proteolysis regulator in eukaryotic development. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1695*(1–3), 45–54.

Schwechheimer, C., & Deng, X. (2000). The COP/DET/FUS proteins — regulators of eukaryotic growth and development. *Seminars in Cell and Developmental Biology*, *11*(6), 495–503.

Serino, G., & Deng, X. W. (2003). The COP9 signalosome: Regulating plant development through the control of proteolysis. *Annual Review of Plant Biology*, *54*, 165–182.

Serino, G., Su, H., Peng, Z., Tsuge, T., Wei, N., Gu, H., & Deng, X. W. (2003). Characterization of the last subunit of the Arabidopsis COP9 signalosome: Implications for the overall structure and origin of the complex. *The Plant Cell*, *15*(3), 719–731.

Settles, A. M., Latshaw, S., & McCarty, D. R. (2004). Molecular analysis of high-copy insertion sites in maize. *Nucleic Acids Research*, *32*(6), e54.

Shi, B., Hou, J., Yang, J., Han, I. J., Tu, D., Ye, S., Yu, J., & Li, L. (2023). Genome-wide analysis of the CSN genes in land plants and their expression under various abiotic stress and phytohormone conditions in rice. *Gene*, *850*, 146905.

Shiferaw, B., Prasanna, B. M., Hellin, J., & Bänziger, M. (2011). Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Security*, *3*(3), 307–327.

Smith, C. W., Betrán, J., & Runge, E. C. A. (2004). *Corn : Origin, history, technology, and production*.

Sperschneider, J., Catanzariti, A., Deboer, K., Petre, B., Gardiner, D. M., Singh, K. B., Dodds, P. N., & Taylor, J. M. (2017). OPEN LOCALIZER : subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientific Reports*, *7*, 44598.

Springer, N. M., Anderson, S. N., Andorf, C. M., Ahern, K. R., Bai, F., Barad, O., Barbazuk, W. B., Bass, H. W., Baruch, K., Ben-Zvi, G., Buckler, E. S., Bukowski, R., Campbell, M. S., Cannon, E. K. S., Chomet, P., Dawe, R. K., Davenport, R., Dooner, H. K., Du, L. H., Du, C., Easterling, K. A., Gault, C., Guan, J. C., Hunter, C. T., Jander, G., Jiao, Y., Koch, K. E., Kol, G., Köllner, T. G., Kudo, T., Li, Q., Lu, F., Mayfield-Jones, D., Mei, W., McCarty, D. R., Noshay, J. M., Portwood, J. L., 2nd, Ronen, G., Settles, A. M., Shem-Tov, D., Shi, J., Soifer, I., Stein, J. C., Stitzer, M. C., Suzuki, M., Vera, D. L., Vollbrecht, E., Vrebalov, J. T., Ware, D., Wei, S., Wimalanathan, K., Woodhouse, M. R., Xiong, W., & Brutnell, T. P. (2018). The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nature Genetics*, *50*(9), 1282-1288.

Stitzer, M. C., & Ross-Ibarra, J. (2018). Maize domestication and gene interaction. *New Phytologist*, *220*(2), 395–408.

Stothard, P. (2000). The sequence manipulation suite: javascript programs for analyzing and formatting protein and DNA sequences. *BioTechniques*, *28*(6), 1102–1104.

Stöcker, T., Altrogge, L., Marcon, C., Win, Y. N., Hochholdinger, F., & Schoof, H. (2022). MuWU: Mutant-seq library analysis and annotation. *Bioinformatics*, *38*(3), 837–838.

Strable, J., & Scanlon, M. J. (2009). Maize (*Zea mays*): A model organism for basic and applied research in plant biology. *Cold Spring Harbor Protocols*, *2009*(10), pdb.emo132.

Stuttmann, J., Lechner, E., Guérois, R., Parker, J. E., Nussaume, L., Genschik, P., & Noël, L. D. (2009). COP9 signalosome- and 26S proteasome-dependent regulation of SCFTIR1 accumulation in Arabidopsis. *The Journal of Biological Chemistry*, *284*(12), 7920–7930.

Subramanian, S., Stacey, G., & Yu, O. (2007). Distinct, crucial roles of flavonoids during legume nodulation. *Trends in Plant Science*, *12*(7), 282–285.

Sun, S., Zhou, Y., Chen, J., Shi, J., Zhao, H., Song, W., Zhang, M., Cui, Y., Dong, X., Liu, H., Ma, X., Jiao, Y., Wang, B., Wei, X., Stein, J.C., Glaubitz, J.C., Lu, F., Yu, G., Liang, C., Fengler, K., Li, B., Rafalski, A., Schnable, P.S., Ware, D.H., Buckler, E.S., & Lai, J. (2018). Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nature Genetics*, *50*(9):1289-1295.

Sun, X., Jia, X., Huo, L., Che, R., Gong, X., Wang, P., & Ma, F. (2018). *MdATG18a* overexpression improves tolerance to nitrogen deficiency and regulates anthocyanin accumulation through increased autophagy in transgenic apple. *Plant Cell and Environment*, *41*(2), 469–480.

Swentowsky, K. W., Gent, J. I., Lowry, E. G., Schubert, V., Ran, X., Tseng, K.-F., Harkess, A. E., Qui, W., & Dawe, R. K. (2020). Distinct kinesin motors drive two types of maize neocentromeres. *Genes & Development*, *34*, 1239–1251.

Tan, B.-C., Chen, Z., Shen, Y., Zhang, Y., Lai, J., & Sun, S. S. M. (2011). Identification of an active new *Mutator* transposable element in maize . *G3 Genes*|*Genomes*|*Genetics*, *1*(4), 293–302.

Taramino, G., Sauer, M., Stauffer, J. L., Multani, D., Niu, X., Sakai, H., & Hochholdinger, F. (2007). The maize (*Zea mays* L.) *RTCS* gene encodes a LOB domain protein that is a key regulator of embryonic seminal and post-embryonic shoot-borne root initiation. *The Plant Journal*, *50*(4), 649–659.

Tenaillon, M. I., & Charcosset, A. (2011). A European perspective on maize history. *Comptes Rendus Biologies*, *334*(3), 221–228.

Tigchelaar, M., Battisti, D. S., Naylor, R. L., & Ray, D. K. (2018). Future warming increases probability of globally synchronized maize production shocks. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(26), 6644–6649.

Tohge, T., Zhang, Y., Peterek, S., Matros, A., Rallapalli, G., Tandrón, Y. A., Butelli, E., Kallam, K., Hertkorn, N., Mock, H., Martin, C., & Fernie, A. R. (2015). Ectopic expression of snapdragon transcription factors facilitates the identification of genes encoding enzymes of anthocyanin decoration in tomato. *Plant Journal*, *83*(4), 686–704.

Toki, S., Hara, N., Ono, K., Onodera, H., Tagiri, A., Oka, S., & Tanaka, H. (2006). Early infection of scutellum tissue with *Agrobacterium* allows high-speed transformation of rice. *Plant Journal*, *47*(6), 969–976.

Troyer, A. F. (1999). Background of U.S. Hybrid Corn. *Crop Science*, *39*(3), 601–626.

Ulrich, K., Becker, R., Behrendt, U., Kube, M., Schneck, V., & Ulrich, A. (2022). Physiological and genomic characterisation of *Luteimonas fraxinea* sp. nov., a bacterial species associated with trees tolerant to ash dieback. *Systematic and Applied Microbiology*, *45*(4), 126333.

United Nations Department of Economic and Social Affairs, Population Division (2022). World population prospects 2022: Summary of results. UN DESA/POP/2022/TR/NO. 3.

Unterseer, S., Pophaly, S. D., Peis, R., Westermeier, P., Mayer, M., Seidel, M. A., Haberer, G., Mayer, K. F. X., Ordas, B., Pausch, H., Tellier, A., Bauer, E., & Schön, C. C. (2016). A comprehensive study of the genomic differentiation between temperate Dent and Flint maize. *Genome Biology*, *17*(1), 137.

van Enckevort, L. J., Droc, G., Piffanelli, P., Greco, R., Gagneur, C., Weber, C., González, V. M., Cabot, P., Fornara, F., Berri, S., Miro, B., Lan, P., Rafel, M., Capell, T., Puigdomènech, P., Ouwerkerk, P. B., Meijer, A. H., Pe', E., Colombo, L., Christou, P., Guiderdoni, E., & Pereira, A. (2005). EU-OSTID: a collection of transposon insertional mutants for functional genomics in rice. *Plant Molecular Biology*, *59*(1), 99–110.

Vollbrecht, E., Duvick, J., Schares, J. P., Ahern, K. R., Deewatthanawong, P., Xu, L., Conrad, L. J., Kikuchi, K., Kubinec, T. A., Hall, B. D., Weeks, R., Unger-Wallace, E., Muszynski, M., Brendel, V. P., & Brutnell, T. P. (2010). Genome-wide distribution of transposed *Dissociation* elements in maize. *The Plant Cell*, *22*(6), 1667–1685.

von Arnim, A. G., Osterlund, M. T., Kwok, S. F., & Deng, X. W. (1997). Genetic and developmental control of nuclear accumulation of COP1, a repressor of photomorphogenesis in Arabidopsis. *Plant Physiology*, *114*(3), 779–788.

von Behrens, I., Komatsu, M., Zhang, Y., Berendzen, K. W., Niu, X., Sakai, H., Taramino, G., & Hochholdinger, F. (2011). *Rootless with undetectable meristem 1* encodes a monocot-specific AUX/IAA protein that controls embryonic seminal and post-embryonic lateral root initiation in maize. *Plant Journal*, *66*(2), 341–353.

Wang, N., Long, T., Yao, W., Xiong, L., Zhang, Q., & Wu, C. (2013). Mutant resources for the functional analysis of the rice genome. *Molecular Plant*, *6*(3), 596–604.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267.

Wang, X., Li, W., Piqueras, R., Cao, K., Deng, X. W., & Wei, N. (2009). Regulation of COP1 nuclear localization by the COP9 signalosome via direct interaction with CSN1. *Plant Journal*, *58*(4), 655–667.

Wasson, A. P., Pellerone, F. I., & Mathesius, U. (2006). Silencing the flavonoid pathway in *Medicago truncatula* inhibits root nodule formation and prevents auxin transport regulation by rhizobia. *The Plant Cell*, *18*(7), 1617–1629.

Wei, N., & Deng, X. W. (1992). COP9: a new genetic locus involved in light-regulated development and gene expression in Arabidopsis. *The Plant Cell*, *4*(12), 1507–1518.

Wei, N. & Deng, X. W. (1996). The role of the COP/DET/FUS genes in light control of Arabidopsis seedling development. *Plant Physiology*, *112*(3), 871–878.

Wei, N., & Deng, X. W. (2003). The COP9 Signalosome. *Annual Review of Cell and Developmental Biology*, *19*, 261–286.

Wei, N., Serino, G., & Deng, X. W. (2008). The COP9 signalosome: more than a protease. *Trends in Biochemical Sciences*, *33*(12), 592–600.

Wen, T., Hochholdinger, F., Sauer, M., Bruce, W., & Schnable, P. S. (2005). The *roothairless1* gene of maize encodes a homolog of *sec3*, which is involved in polar exocytosis. *138*(3), 1637–1643.

Wiemann, S., Weil, B., Wellenreuther, R., Gassenhuber, J., Glassl, S., Ansorge, W., Böcher, M., Blöcker, H., Bauersachs, S., Blum, H., Lauber, J., Düsterhöft, A., Beyer, A., Köhrer, K., Strack, N., Mewes, H.W., Ottenwälder, B., Obermaier, B., Tampe, J., Heubner, D., Wambutt, R., Korn, B., Klein, M., & Poustka, A. (2001). Toward a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs. *Genome Research*, *11*(3), 422–435.

Williams-Carrier, R., Stiffler, N., Belcher, S., Kroeger, T., Stern, D. B., Monde, R. A., Coalter, R., & Barkan, A. (2010). Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy *Mutator* lines of maize. *Plant Journal*, *63*(1), 167–177.

Win, Y. N., Pöschel, M., Stöcker, T., Du, X., Klaus, A., Braun, B. W., Lukas, L., Brox, A., Schoof, H., Hochholdinger, F., & Marcon, C. (2024a). Forward and reverse genetic analysis of *Mutator* transposon induced maize (*Zea mays* L.) mutants of the *BonnMu* resource. *Cold Spring Harbor Protocols*, *in press*.

Win, Y. N., Stöcker, T., Du, X., Brox, A., Pitz, M., Klaus, A., Schoof, H., Hochholdinger, F., & Marcon, C. (2024b). Expanding the *BonnMu* sequence-indexed repository of transposon induced maize (*Zea mays* L.) mutations in dent and flint germplasm. *BioRixv*, 10.1101/2024.02.24.581857; *The Plant Journal*, *in revision*.

Woll, K., Borsuk, L. A., Stransky, H., Nettleton, D., Schnable, P. S., & Hochholdinger, F. (2005). Transcriptome analyses of the novel maize lateral and seminal root initiation mutant *rum1*. *Plant Physiology*, *139*(3), 1255–1267.

Woodhouse, M. R., Cannon, E. K., Portwood, J. L., Harper, L. C., Gardiner, J. M., Schaeffer, M. L., & Andorf, C. M. (2021). A pan-genomic approach to genome databases using maize as a model system. *BMC Plant Biology*, *21*(1), 1–10.

Xiao, X., Fan, M., Wang, E., Chen, W., & Wei, G. (2017). Interactions of plant growth-promoting rhizobacteria and soil factors in two leguminous plants. *Applied Microbiology and Biotechnology*, *101*(23-24), 8485–8497.

Xu, C., Tai, H., Saleem, M., Ludwig, Y., Majer, C., Kenneth, W., Nagel, K. A., Wojciechowski, T., Meeley, R. B., Taramino, G. & Hochholdinger, F. (2015). Cooperative action of the paralogous maize lateral organ boundaries (LOB) domain proteins RTCS and RTCL in shoot-borne root formation. *New Phytologist*, *207*(4),1123–1133.

Yang, N., Wang, Y., Liu, X., Jin, M., Vallebueno-Estrada, M., Calfee, E., Chen, L., Dilkes, B. P., Gui, S., Fan, X., Harper, T. K., Kennett, D. J., Li, W., Lu, Y., Ding, J., Chen, Z., Luo, J., Mambakkam, S., Menon, M., Snodgrass, S., Veller, C., Wu, S., Wu, S., Zhuo, L., Xiao, Y., Yang, X., Stitzer, M. C., Runcie, D., Yan, J., & Ross-Ibarra, J. (2023). Two teosintes made modern maize. *Science*, *382*(6674), eadg8940.

Yu, P., He, X., Baer, M., Beirinckx, S., Tian, T., Moya, Y. A. T., Zhang, X., Deichmann, M., Frey, F. P., Bresgen, V., Li, C., Razavi, B. S., Schaaf, G., von Wirén, N., Su, Z., Bucher, M., Tsuda, K., Goormachtig, S., Chen, X., & Hochholdinger, F. (2021). Plant flavones enrich rhizosphere Oxalobacteraceae to improve maize performance under nitrogen deprivation. *Nature Plants*, *7*(4), 481–499.

Yuan, M., Yu, T., Shi, Q., Han, D., Yu, K., Wang, L., Wang, S., Xiang, H., Wen, R., Nian, H., & Lian, T. (2021). Rhizosphere soil bacterial communities of continuous cropping-tolerant and sensitive soybean genotypes respond differently to long-term continuous cropping in Mollisols. *Frontiers in Microbiology*, *12*, 729047.

Zhang, J., Zhang, X., Chen, R., Yang, L., Fan, K., Liu, Y., Wang, G., Ren, Z., & Liu, Y. (2020). Generation of transgene-free semidwarf maize plants by gene editing of *Gibberellin-Oxidase20-3* using CRISPR/Cas9. *Frontiers in Plant Science*, *11*, 1048.

Zhang, Q., Zhang, Y., Lu, M. H., Chai, Y. P., Jiang, Y. Y., Zhou, Y., Wang, X. C., & Chen, Q. J. (2019). A novel ternary vector system united with morphogenic genes enhances CRISPR/Cas delivery in maize. *Plant Physiology*, *181*(4), 1441–1448.

Zhang, W., Garcia, N., Feng, Y., Zhao, H., & Messing, J. (2015). Genome-wide histone acetylation correlates with active transcription in maize. *Genomics*, *106*(4), 214–220.

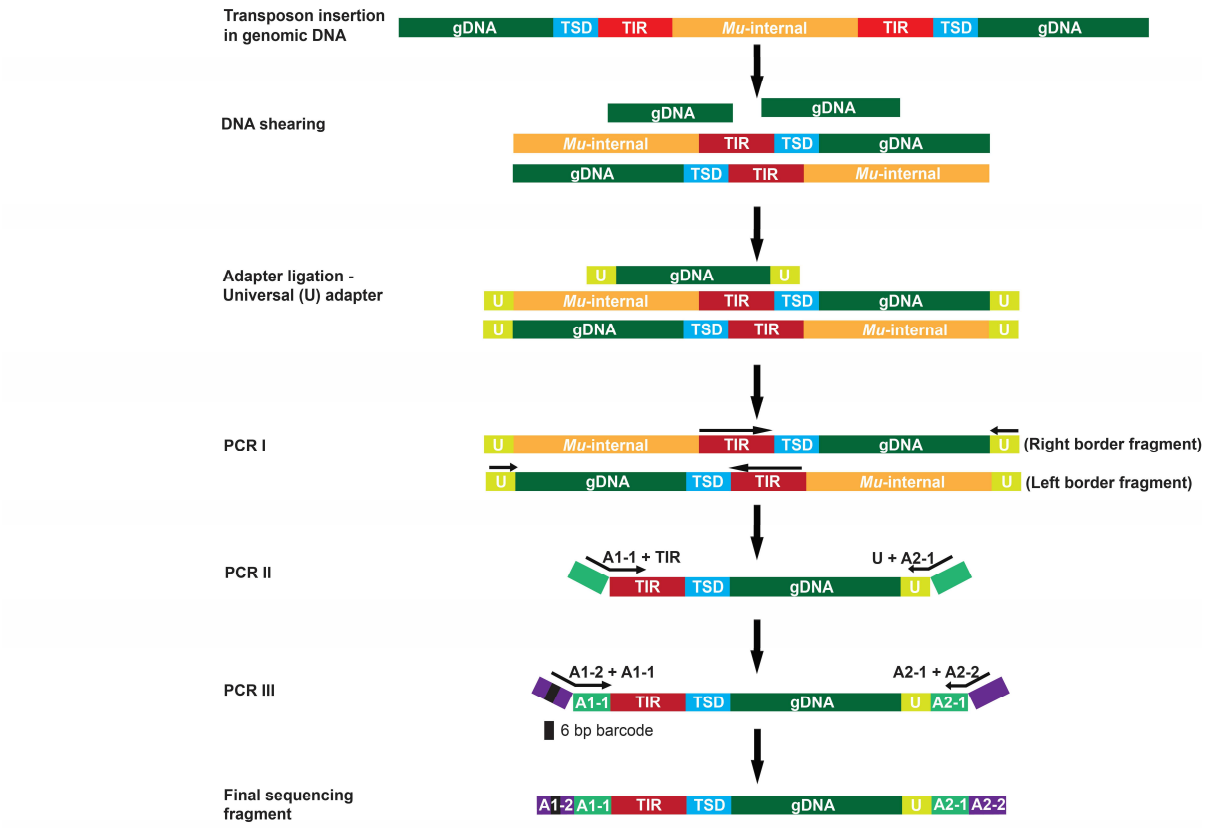
Zhang, W., Ito, H., Quint, M., Huang, H., Noël, L. D., & Gray, W. M. (2008). Genetic analysis of CAND1-CUL1 interactions in Arabidopsis supports a role for CAND1-mediated cycling of the SCFTIR1 complex. *Proceedings of the National Academy of Sciences of the United States of America, 105*(24), 8470–8475.

Zhang, Y., Marcon, C., Tai, H., von Behrens, I., Ludwig, Y., Hey, S., Berendzen, K. W., & Hochholdinger, F. (2016). Conserved and unique features of the homeologous maize Aux / IAA proteins ROOTLESS WITH UNDETECTABLE MERISTEM 1 and RUM1-like 1. *Journal of Experimental Botany*, *67*(4), 1137–1147.

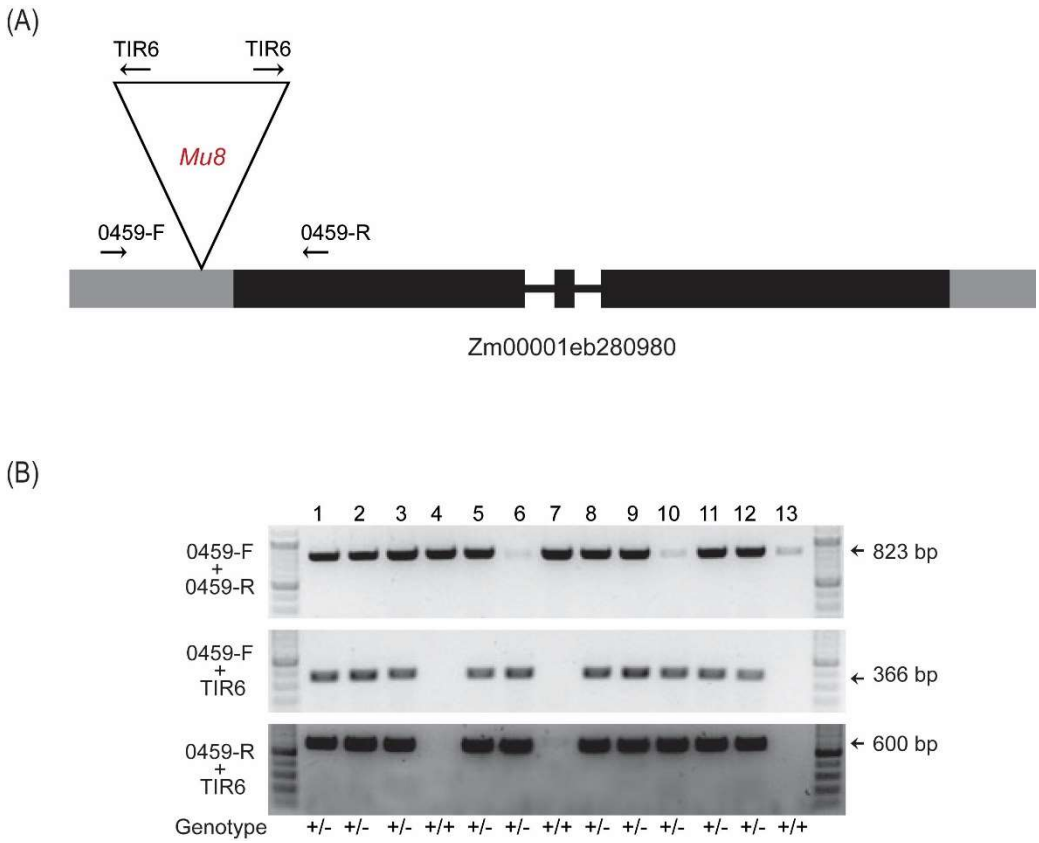
Zhou, F., Wu, X., Gao, Y., Fan, S., Zhou, H., & Zhang, X. (2022). Diversity shifts in the root microbiome of cucumber under different plant cultivation substrates. *Frontiers in Microbiology*, *13*, 878409.

9 Supplemental data files

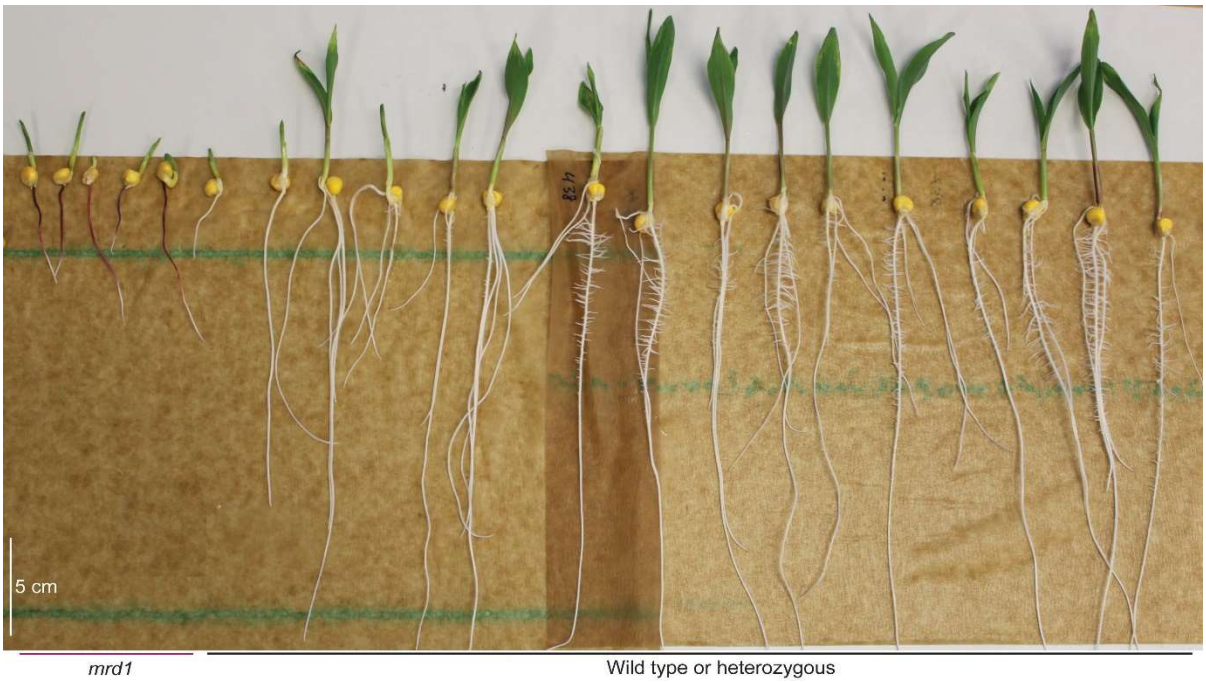
9.1 Supplemental figures



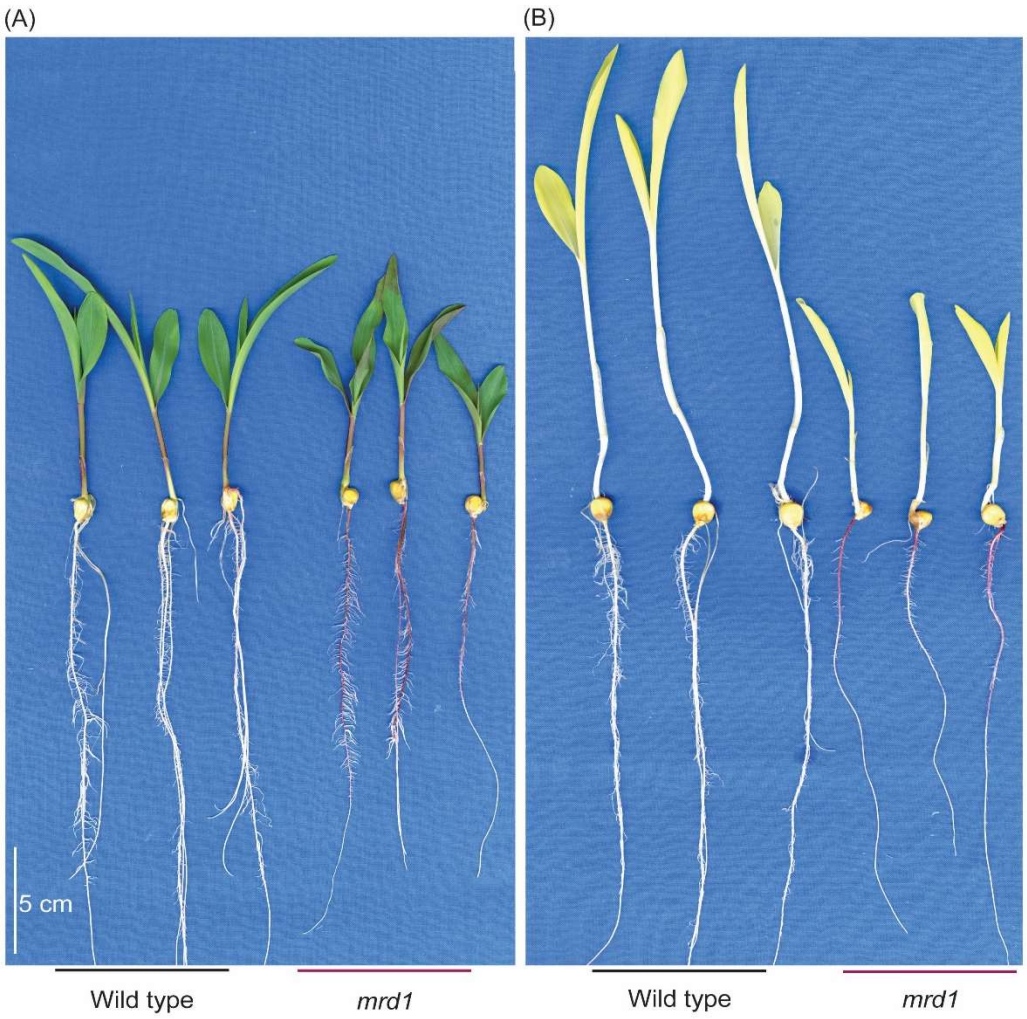
Supplemental figure S1. Overview of Mu-seq library construction (adapted from McCarty *et al*., 2013). This schematic depicts the Mu-seq process for sequencing and identifying *Mutator* (*Mu*) transposon insertions within genomic DNA. The procedure genomic DNA (green), which contains *Mu* elements characterized by terminal inverted repeats (TIRs in red), a *Mu* internal sequence (orange), and target site duplications (TSDs in blue), through bioruptor sonication. Subsequently, fragments are ligated to a universal adapter (U-adapter in yellowish green). The first PCR amplification (PCR-I) uses primers specific to the TIR and U-adapter to specifically amplify Mu-TIR flanking genomic DNA, enriching both the left and right border DNA fragments. In the second PCR (PCR-II), parts of the Illumina sequencing adapters (A1 and A2 in light green and purple) are attached to the ends of the enriched fragments. The third PCR (PCR-III) completes the remaining parts of Illumina sequencing adapters and adds a 6-base pair barcode (illustrated in black) to each fragment, allowing for the multiplexing of 48 samples. The prepared Mu-seq library is then measured for quality and quantity using a Bioanalyzer and sequenced using HiSeq 2500 or Illumina XTen platform.



Supplemental figure S2. Confirmation of *Mu8* insertion by PCR. (A) Simplified gene model of Zm00001eb280980 carrying a *Mu8* element in its 5' UTR. (B) PCR segregation analysis of 13 segregating plants of the *BonnMu*-7-C-0459 family (lower picture). Genotyping of the individual plants using gene-specific primers 0459-F + 0459-R and Mu-specific primers 0459-F + TIR6 and 0459-R + TIR6 identified 10 plants as heterozygotes (-/+: # 1-3; # 5-6; # 8-12) and three plants as homozygous wild types (+/+: # 4; # 7; # 13).



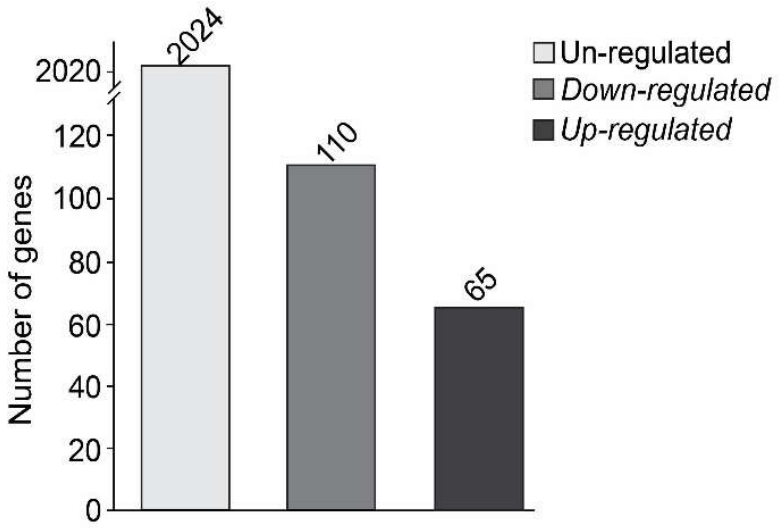
Supplemental figure S3. Phenotype of the 10-day-old *BonnMu* F_2-family (*BonnMu*-1-A-0438) segregating for *mrd1* mutation. The *mrd1* mutant was first discovered in this family, where χ^2 test showed 3:1 (wild type: *mrd1*) Mendelian segregation ratio (Table 11).



Supplemental figure S4. Phenotype of 10-day-old wild type and *mrd1* mutant seedlings under light and dark condition. (A) Under light condition (800 µmol/m2/s), *mrd1* displays a reduced primary root and shoot growth, compared to wild type. (B) In the dark, *mrd1* mutants are de-etiolated, comparable to the *mrd1* phenotype grown in light. Under both conditions (i.e. light or darkness) the primary root of *mrd1* accumulates anthocyanins, resulting in the magenta root phenotype.

Supplemental figure S5. Phenotype of the 10-day-old *Mu*-active line PS_07-1502-8[1st]/1505-5. This *Mu*-active line is segregating for *mrd1* mutant.

Supplemental figure S6. Phenotype of 45-day-old wild type and *mrd1* mutant plants.



Supplemental figure S7. Significant differentially expressed genes in *mrd1* mutant. The bar plot shows the number of un-regulated, down-regulated and up-regulated genes in the chromosome 1 of *mrd1* mutant via BSR-seq analysis.

9.2 Supplemental tables

Supplemental table S1. Calculated probability of obtaining at least one mutant allele per tagged gene among 3-8 germinated plants per *BonnMu* F_2-family. The left panel, highlighted in grey, shows probabilities for scenarios with eight germinated seedlings per F_2-family. The right panel represents probabilities when only three out of the eight seedlings per F_2-family germinated and were subsequently harvested. WT: wild type, mut.: mutant.

Number of assumed mutants among eight seedlings	Probability[1]	Probability, if 3 seedlings are harvested[2]				Probability of ≥ 1 mutant
		3 mut., 0 WT	2 mut., 1 WT	1 mut., 2 WT	0 mut., 3 WT	
0	0.00	NA[3]	NA	NA	1.00	0.00
1	0.00	NA	NA	0.50	0.50	0.50
2	0.00	NA	0.10	0.50	0.40	0.60
3	0.02	0.02	0.27	0.54	0.18	0.83
4	0.09	0.07	0.43	0.43	0.07	0.93
5	0.21	0.18	0.54	0.27	0.02	0.98
6	0.31	0.36	0.54	0.12	NA	1
7	0.27	0.50	0.50	NA	NA	1
8	0.10	1.00	NA	NA	NA	1

[1] The probability of having at least one mutant allele per tagged gene among eight seedlings was calculated using the binomial probability formula in R:

Formula 1:

$$Bi_{n;\,p}(k) = P\,(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

In this formula:

n represents the number of kernels: 8.

p is the probability of each kernel harboring a mutant allele, which is 0.75.

k is the number of mutants among 8 kernels, 0-8, representing the number of mutants among the 8 kernels.

$\binom{n}{k}$ is the binomial coefficient, representing the number of k mutants can occur in n trials.

The formula calculates the probability of having exactly k mutant alleles. To find the probability of having at least one mutant allele, we sum the probabilities for all cases where k ranges from 1 to 8

[2] The probabilities, when only three out of eight seedlings per F_2-family are harvest, are calculated using the following formula in R.

Formula 2:

$$P\,(X = k) = \frac{\binom{K}{k}\ \mathrm{X}\ \binom{N-K}{n-k}}{\binom{N}{n}}$$

[3] NA, combination not possible.

The final probability to have at least one mutant allele per tagged gene included among the 3-8 germinated plants per F_2-family was calculated as:

0 x 0 + 0 x 0.50 + 0 x 0.60 + 0.02 x 0.83 + 0.09 x 0.93 + 0.21 x 0.98 + 0.31 x 1 + 0.27 x 1 + 0.1 x 1 = 0,9861 (99%).

Supplemental table S2. Primers used in Mu-seq library construction

Name	Sequences (5′→3′)
Adapter_rc-tiB	CTGAGACTGCCAAGGCACAC
PCR-I_TIR6	AGAGAAGCCAACGCCAWCGCCTCYATTTCGTC
PCR-I_tiB	CCTATCCCCTGTGTGCCTTGGCAGTCTCAG
PCR-II_fwd	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCBCTCTTCKTCYATAATGGCAAT
PCR-II_rev	CTTTCCCTACACGACGCTCTTCCGATCTGCCTTGGCAGTCTCAG
PCR-III_rev	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC
PCR-III_RPI01	CAAGCAGAAGACGGCATACGAGAT **CGTGAT** GTGACTGGAGTTCAGACGT
PCR-III_RPI02	CAAGCAGAAGACGGCATACGAGAT **ACATCG** GTGACTGGAGTTCAGACGT
PCR-III_RPI03	CAAGCAGAAGACGGCATACGAGAT **GCCTAA** GTGACTGGAGTTCAGACGT
PCR-III_RPI04	CAAGCAGAAGACGGCATACGAGAT **TGGTCA** GTGACTGGAGTTCAGACGT
PCR-III_RPI05	CAAGCAGAAGACGGCATACGAGAT **CACTGT** GTGACTGGAGTTCAGACGT
PCR-III_RPI06	CAAGCAGAAGACGGCATACGAGAT **ATTGGC** GTGACTGGAGTTCAGACGT
PCR-III_RPI07	CAAGCAGAAGACGGCATACGAGAT **GATCTG** GTGACTGGAGTTCAGACGT
PCR-III_RPI08	CAAGCAGAAGACGGCATACGAGAT **TCAAGT** GTGACTGGAGTTCAGACGT
PCR-III_RPI09	CAAGCAGAAGACGGCATACGAGAT **CTGATC** GTGACTGGAGTTCAGACGT
PCR-III_RPI10	CAAGCAGAAGACGGCATACGAGAT **AAGCTA** GTGACTGGAGTTCAGACGT
PCR-III_RPI11	CAAGCAGAAGACGGCATACGAGAT **GTAGCC** GTGACTGGAGTTCAGACGT
PCR-III_RPI12	CAAGCAGAAGACGGCATACGAGAT **TACAAG** GTGACTGGAGTTCAGACGT
PCR-III_RPI13	CAAGCAGAAGACGGCATACGAGAT **TTGACT** GTGACTGGAGTTCAGACGT

Supplemental table S2. Primers used in Mu-seq library construction (continued)

Name	Sequences (5′→3′)
PCR-III_RPI14	CAAGCAGAAGACGGCATACGAGAT **GGAACT** GTGACTGGAGTTCAGACGT
PCR-III_RPI15	CAAGCAGAAGACGGCATACGAGAT **TGACAT** GTGACTGGAGTTCAGACGT
PCR-III_RPI16	CAAGCAGAAGACGGCATACGAGAT **GGACGG** GTGACTGGAGTTCAGACGT
PCR-III_RPI17	CAAGCAGAAGACGGCATACGAGAT **CTCTAC** GTGACTGGAGTTCAGACGT
PCR-III_RPI18	CAAGCAGAAGACGGCATACGAGAT **GCGGAC** GTGACTGGAGTTCAGACGT
PCR-III_RPI19	CAAGCAGAAGACGGCATACGAGAT **TTTCAC** GTGACTGGAGTTCAGACGT
PCR-III_RPI20	CAAGCAGAAGACGGCATACGAGAT **GGCCAC** GTGACTGGAGTTCAGACGT
PCR-III_RPI21	CAAGCAGAAGACGGCATACGAGAT **CGAAAC** GTGACTGGAGTTCAGACGT
PCR-III_RPI22	CAAGCAGAAGACGGCATACGAGAT **CGTACG** GTGACTGGAGTTCAGACGT
PCR-III_RPI23	CAAGCAGAAGACGGCATACGAGAT **CCACTC** GTGACTGGAGTTCAGACGT
PCR-III_RPI24	CAAGCAGAAGACGGCATACGAGAT **GCTACC** GTGACTGGAGTTCAGACGT
PCR-III_RPI25	CAAGCAGAAGACGGCATACGAGAT **ATCAGT** GTGACTGGAGTTCAGACGT
PCR-III_RPI26	CAAGCAGAAGACGGCATACGAGAT **GCTCAT** GTGACTGGAGTTCAGACGT
PCR-III_RPI27	CAAGCAGAAGACGGCATACGAGAT **AGGAAT** GTGACTGGAGTTCAGACGT
PCR-III_RPI28	CAAGCAGAAGACGGCATACGAGAT **CTTTTG** GTGACTGGAGTTCAGACGT
PCR-III_RPI29	CAAGCAGAAGACGGCATACGAGAT **TAGTTG** GTGACTGGAGTTCAGACGT
PCR-III_RPI30	CAAGCAGAAGACGGCATACGAGAT **CCGGTG** GTGACTGGAGTTCAGACGT
PCR-III_RPI31	CAAGCAGAAGACGGCATACGAGAT **ATCGTG** GTGACTGGAGTTCAGACGT
PCR-III_RPI32	CAAGCAGAAGACGGCATACGAGAT **TGAGTG** GTGACTGGAGTTCAGACGT
PCR-III_RPI33	CAAGCAGAAGACGGCATACGAGAT **CGCCTG** GTGACTGGAGTTCAGACGT

Supplemental table S2. Primers used in Mu-seq library construction (continued)

Name	Sequences (5′→3′)
PCR-III_RPI34	CAAGCAGAAGACGGCATACGAGAT **GCCATG** GTGACTGGAGTTCAGACGT
PCR-III_RPI35	CAAGCAGAAGACGGCATACGAGAT **AAAATG** GTGACTGGAGTTCAGACGT
PCR-III_RPI36	CAAGCAGAAGACGGCATACGAGAT **TGTTGG** GTGACTGGAGTTCAGACGT
PCR-III_RPI37	CAAGCAGAAGACGGCATACGAGAT **ATTCCG** GTGACTGGAGTTCAGACGT
PCR-III_RPI38	CAAGCAGAAGACGGCATACGAGAT **AGCTAG** GTGACTGGAGTTCAGACGT
PCR-III_RPI39	CAAGCAGAAGACGGCATACGAGAT **GTATAG** GTGACTGGAGTTCAGACGT
PCR-III_RPI40	CAAGCAGAAGACGGCATACGAGAT **TCTGAG** GTGACTGGAGTTCAGACGT
PCR-III_RPI41	CAAGCAGAAGACGGCATACGAGAT **GTCGTC** GTGACTGGAGTTCAGACGT
PCR-III_RPI42	CAAGCAGAAGACGGCATACGAGAT **CGATTA** GTGACTGGAGTTCAGACGT
PCR-III_RPI43	CAAGCAGAAGACGGCATACGAGAT **GCTGTA** GTGACTGGAGTTCAGACGT
PCR-III_RPI44	CAAGCAGAAGACGGCATACGAGAT **ATTATA** GTGACTGGAGTTCAGACGT
PCR-III_RPI45	CAAGCAGAAGACGGCATACGAGAT **GAATGA** GTGACTGGAGTTCAGACGT
PCR-III_RPI46	CAAGCAGAAGACGGCATACGAGAT **TCGGGA** GTGACTGGAGTTCAGACGT
PCR-III_RPI47	CAAGCAGAAGACGGCATACGAGAT **CTTCGA** GTGACTGGAGTTCAGACGT
PCR-III_RPI48	CAAGCAGAAGACGGCATACGAGAT **TGCCGA** GTGACTGGAGTTCAGACGT

Supplemental table S3. Primers used for genotyping

Name	Primer sequence (5′→3′)	Target sequence	*BonnMu* F_2-family	Insertion_Identifier	GeneID
A0982-F	GGCCTAAACTCGCAAAGGGAT	Gene-specific	*BonnMu*-2-A-0982	*BonnMu*0031087	Zm00001eb052530
A0982-R	ATGCTTCTTGAGGCGACCAT				
A0982-mu8-F	GCCGAGTTCTGGACGATGA	*Mu8*			
A0982-mu8-R	GACCATGGTTCTTGACGACG	Gene-specific			
C0459-F	TACTACGGTTTAAGGCGTGTGG	Gene-specific	*BonnMu*-7-C-0459	*BonnMu*0170576	Zm00001eb280980
C0459-R	AAGAGAGCAAGGGGTTTAGGC	Gene-specific			
F1001-F	ATATGCAGGTGAGCGGGTAG	Gene-specific	*BonnMu*-F7-2-F-1001	*BonnMu*0480236	Zm00001eb256020
F1001-R	AAATCGAGATCGCAAGGCCA				
TIR6	AGAGAAGCCAACGCCAWCGCCCYATTTCGTC	MuTIR			
mrd1_Fwd1	CCGGTAGAAGCCATGGACAG	Gene-specific	*BonnMu* resource	All *mrd1* families	Zm00001eb008060
mrd1_Rev1	TCGAACCGAGCGTGGAAATC				
mrd1_Fwd2	CTCTGCGCTTTGCAGTCATC	Gene-specific	*ChinaMu* resource	23809630	Zm00001eb008060
mrd1_Rev2	AACCAATGTAGCCAACCTGCT				

Supplemental table S4. Primers used for qRT-PCR

Name	Primer sequence (5′→3′)	Gene
mrd1-qRT-PCR-F	TTTGATGATGACACGGAAGAGC	Zm00001eb008060 (*mrd1*)
mrd1-qRT-PCR-R	TCAAGGATGTCGTTCAGGGC	
hb3-qRT-PCR-F	TCATGAAGCTTCCACTGAGATAC	Zm00001eb295800 (*hb3*)
hb3-qRT-PCR-R	CGGCCATTGCTGTTAGTTAGC	

Supplemental table S5. Primers used for validation of *mrd1* gene sequence of B104 genome

Name	Primer sequence (5′→3′)	Gene
CROPGEN962	AATGCCAAAACCCTAGCTCGC	GRMZM2G153769 (*mrd1*)
CROPGEN963	AAACTCACGCAGTACTAACAGGA	
CROPGEN964	GATGCCTGTTATTTTGTGTTGCC	
CROPGEN965	ATTGCTGCAGCTCTTCCGT	

Supplemental table S6. Primers used for cloning sgRNAs

Name	Primer sequence (5′→3′)	Gene
CROPGEN1002	GGCAGGCCTGAGAGATGTCCGGTG	GRMZM2G153769 (*mrd1*)
CROPGEN1003	AAACCACCGGACATCTCTCAGGCC	
CROPGEN1010	GGCAGTATGGCGTGTGTGCCTTAA	
CROPGEN1011	AAACTTAAGGCACACACGCCATAC	
CROPGEN1012	GGCAGATTATGCTCGATCATTGCT	
CROPGEN1013	AAACAGCAATGATCGAGCATAATC	
CROPGEN1014	GGCAGATTTATGAGGATAGAATGC	
CROPGEN1015	AAACGCATTCTATCCTCATAAATC	

Supplemental table S7. Partition of the B73v5 genome.

Partition	%
Intergenic	88,414
PromoterProx[1]	3,343
PromoterCore[2]	0,178
5‘ UTR	0,494
Exon	1,871
Intron	5,052
3‘ UTR	0,648
Total	**100%**

[1] PromoterProx: Proximal promoter region located 101 – 2,100 bp upstream of the start of the 5’ untranslated region (UTRs) of a gene.
[2] PromoterCore: Core promoter region located 1 – 100 bp upstream of the the start of the 5’ UTR of a gene.

Supplemental table S8. Number of observed and expected Mu insertions across the B73v5 genome.

Partition	Observed	Expected	$\mathbf{Log_{10}}\left(\frac{\mathbf{Obs.}}{\mathbf{Exp.}}\right)$
Intergenic	52,718	684,938	-1,114
PromoterProx	98,510	25,899	0,580
PromoterCore	53,950	1,380	1,592
5‘ UTR	331,168	3,824	1,938
Exon	115,786	14,491	0,903
Intron	95,371	39,141	0,387
3‘ UTR	27,189	5,020	0,734

We used Pearson's χ^2 test with Yates' continuity correction to calculate the ratio between the number of observed and expected insertions per genomic partition.

Supplemental table S9. Top five *Mu*-tagged genes identified to overlap in at least six *BonnMu*-F_2-families, segregating for the *mrd1* phenotype.

No. of F_2-families	Affected gene, overlapping in F_2-families	Location in mapping interval (Chr 1-70,000,000)	Description	Expression, according to BSR-seq data
7	Zm00001eb007200	21,551,923-21,555,136	*N/A*	Not regulated
38	Zm00001eb008060	24,000,830-24,011,578	*Cop9 signalosome complex subunit 4*	Downregulated
6	Zm00001eb011170	35,712,410-35,717,428	*N/A*	Not regulated
8	Zm00001eb013450	45,537,884-45,539,961	*Probable protein phosphatase 2C 30*	Not regulated
11	Zm00001eb015320	53,030,837-53,032,126	*WRKY55-superfamily of TFs having WRKY and zinc finger domains*	Not regulated

Supplemental table S10. The matrix displays the percent identity and similarity among MRD1 homologs. The full protein sequences of each organism were aligned using MUSCLE algorithm in Mega X (Kumar *et al*., 2018) and the percentages of identity and similarity of the aligned sequences were calculated using Ident and Sim (Storthard, 2000).

Identity (%) / Similarity (%)	***A. t.***	***B. d.***	***H. v.***	***O. s.***	***P. p.***	***S. b.***	***Z. m.***
A. t.		82	84	85	72	84	84
B. d.	90		95	92	72	94	94
H. v.	91	98		94	74	95	95
O. s.	91	95	97		74	96	96
P. p.	84	83	84	85		74	74
S. b.	90	96	98	98	84		100
Z. m.	90	96	97	98	84	100	

A. t., *Arabidopsis thialiana*; *B. d*., *Brachypodium distachyon*; *H. v*., *Hordeum vulgare*; *O. s*., *Oryza sativa*; *P. p*., *Physcomitrella patens*; *S. b*., *Sorghum bicolor*; *Z. m*., *Zea mays*.

The UniProt accession numbers used for each organism are Q8L5U0 (*A. t*.), I1H8K4 (*B. d*.), A0A8I6XN54 (*H. v*.), Q10QG2 (*O. s*.), A9TH34 (*P. p*.), C5WUN5 (*S. b*.) and B4FCA4 (*Z. m*.).

Supplemental table S11. CRISPR/Cas9-edited *mrd1* plants for the first target region (WT, Wild type; mut., mutant). The dual gRNA1 and gRNA2 were used to target a 441 bp region between exon 1 and intron 2 of the *mrd1* gene.

Label	Allele	T0 INDELs in nucleotide (nt)		F_1		F_2		Note
		gRNA1	gRNA2	Phenotype	Genotype	Phenotype	Genotype	
843-481-A	*mrd1-2*	+3	+1	WT	mut.	WT	mut.	Edited
843-481-Ba	*mrd1-3*	-27	No INDEL	WT	mut.	WT	mut.	Edited
843-481-Bb	*mrd1-4*	-36	No INDEL	WT	mut.	WT	mut.	Edited
843-481-E	*mrd1-5*	+2	-10	WT	mut.	WT	mut.	Edited
843-481-F	*mrd1-6*	+1	No INDEL	WT	mut.	WT	mut.	Edited
836-481-A	*mrd1-7*	-3	+1	WT	mut.	WT	mut.	Edited

Supplemental table S12. CRISPR/Cas9-edited *mrd1* plants for the second target region (WT, Wild type; mut., mutant). The dual gRNA3 and gRNA4 were used to target a 592 bp region between exon 11 and exon 13 of *mrd1* gene where the PCI domain of the gene is located.

Label	Allele	T0 INDELs in nucleotide (nt)		F_1		F_2		Note
		gRNA3	gRNA4	Phenotype	Genotype	Phenotype	Genotype	
837-482-Aa	*mrd1-8*	-4	-9					Died in T0 stage
837-482-Ab	*mrd1-9*	+1	No INDEL					
837-482-Ac	*mrd1-10*	No INDEL	-2, -44, -32					
837-482-Ae	*mrd1-11*	-1	NA					
837-482-Af	*mrd1-12*	-35	-1, -2					
837-482-Ag	*mrd1-13*	No INDEL	-1					
837-482-B*	*mrd1-14*	-11, -597	-597					

T0 transformants were chimeric for the target gene editing and therefore, there are more than one INDELs for the same gRNA in some plants.

Acknowledgement

First and foremost, my heartfelt thanks to Priv.-Doz.in Dr. Caroline Marcon, who welcomed me into her project within the Crop Functional Genomics research group as a Ph.D student. Her guidance, patience, and supportive presence have been invaluable throughout this research journey.

I extend my sincere appreciation to Prof. Dr. Frank Hochholdinger, for providing me the opportunity to be a part of his research team and complete this dissertation.

My gratitude also to Tyll Stöcker and Prof. Dr. Heiko Schoof from Crop Bioinformatics for their indispensable analysis of our Mu-seq data. Additionally, I am grateful to Prof. Dr. Heiko Schoof for his role as the second examiner, and to Prof. Gabriel Schaaf and Prof. Claudia Knief for their invaluable input as members of my thesis committee.

Special thanks to Alexa for her invaluable contributions in producing Mu-seq libraries, a key factor in the success of our *BonnMu* collection.

Special thanks to Priv.-Doz. Dr. Hans-Peter Mock (Leibniz Institute for Plant Genetics and Crop Plant Research, Gatersleben) for his significant contribution to anthocyanin profiling, and to Laurens Pauwels (VIB Center, Ghent, Belgium) for his work on the CRISPR/Cas9 constructs transformation.

I deeply appreciate the support of all the members of the Crop Functional Genomics lab and Prof. Dr. Andreas Meyer's Chemical Signaling lab, whose assistance has been crucial in my research. My gratitude extends to the Deutschen Forschungsgemeinschaft for their financial support of this project.

A special acknowledgment to my colleague and friend Marcel, whose assistance in both research and personal matters has been extraordinary, making my time in the lab truly enjoyable. I am also thankful to Mareike for her constant support and suggestions, as well as for the enjoyable lunch and fun times we shared.

My special thanks to Helmut and Britta, whose technical skills were invaluable both in the lab and in the field. I appreciate the assistance of Li in the subcellular localization experiment, and Xiaoming and Peng for their collaborative efforts in the soil microbiome experiment. Grateful acknowledgment to Alina and Marion for their exceptional assistance in data analysis and interpretation, greatly enhancing our research.

I owe a debt of gratitude to my high school teachers, Daw Than Than Aye (Physics, Kantkaw) and Daw Hnin Swe (Mathematics), for guiding me towards the study of agricultural science.

I am grateful to my colleagues for their camaraderie, knowledge sharing, and encouragement during my time in Bonn.

Lastly, my profound thanks to my parents, brothers, and sisters for their unwavering support and love.