



OPEN Native alleles at *lhcb6* shape photosynthetic efficiency and early growth in maize

Sebastian Urzinger^{1,7}, Lukas Würstl^{1,7}, Viktoriya Avramova¹, Claude Urbany², Daniela Scheuermann², Thomas Presterl², Stefan Reuscher², Manfred Mayer^{1,3}, Sarah Brajkovic⁴, Bernhard Küster⁴, Milena Ouzunova², Bernardo Ordas⁵, Peter Westhoff⁶ & Chris-Carolin Schön¹✉

Maize is a globally important staple crop, providing food, animal feed, and industrial raw materials. Its extensive native diversity can be utilized to broaden the genetic basis for quantitative trait improvement and advance our understanding of the genotype-phenotype relationships underlying complex trait variation. In a forward genetic approach, we dissect a locus on chromosome 10 with major effects on Photosystem II (PSII) maximum quantum efficiency (F_v/F_m), non-photochemical quenching (NPQ) and biomass accumulation during early growth stages. Integrating molecular and physiological information we show that allelic variation at the gene encoding LIGHT HARVESTING CHLOROPHYLL A/B BINDING PROTEIN6 (LHCB6 also known as CP24), a component of the PSII light-harvesting complex (LHCII) antenna, underlies this locus. We demonstrate that the allelic variation results from a hAT transposon insertion at *lhcb6* and is associated with decreased accumulation of the LHCII antenna components LHCB6 and LHCB3. Based on proteomic analyses we propose candidate genes that partially compensate for the unfavorable early growth effects caused by impaired LHCII antenna assembly. Our work provides novel insights into the function of *lhcb6* in the C_4 crop maize and demonstrates the value of natural variation for the understanding and genetic improvement of complex photosynthetic processes.

†These authors contributed equally (SU and LW).

Introduction

Crop yields have increased steadily in recent decades, yet future genetic improvement will require broadening the genetic base of elite crop germplasm^{1,2}. In maize (*Zea mays* L.), multiple genetic bottlenecks—starting with its geographic expansion from its center of origin, followed by the establishment of heterotic groups using a small number of landraces, and subsequent cycles of intense selection—have significantly narrowed the genetic diversity of elite germplasm^{3–5}. Thus, beneficial alleles that influence (a)biotic stress tolerance, resource efficiency and photosynthetic traits, likely have been lost over time and undesirable alleles might have been fixed through genetic drift or linkage drag^{6–8}. Breakthroughs in genetic engineering have deepened our understanding of plant physiology and raised hopes for the development of improved crops. However, when it comes to quantitative traits, translating molecular advances to field performance often falls short of expectations^{9,10}. Alternatively, landraces offer a wealth of untapped allelic variation that can be leveraged to improve agronomically and ecologically significant traits^{1,11–13}. Despite their conceptual value, the incorporation of landraces into breeding programs is hindered by their yield gap compared to elite lines^{11,14,15} and the risk of linkage drag due to the co-inheritance of undesirable and beneficial alleles^{16,17}. One strategy to overcome these challenges is to extract clearly defined alleles associated with quantitative traits of interest from preselected landraces¹². In a previous study, we reported a successful example for such an approach by demonstrating that favorable allelic variation at

¹Plant Breeding, TUM School of Life Sciences, Technical University of Munich, 85354 Freising, Germany. ²Maize Breeding, KWS SAAT SE & Co. KGaA, 37574 Einbeck, Germany. ³Present address: Bayer Crop Science, 46325 Borken, Germany. ⁴Proteomics and Bioanalytics, TUM School of Life Sciences, Technical University of Munich, 85354 Freising, Germany. ⁵Misión Biológica de Galicia, Spanish National Research Council (CSIC), Pontevedra 36080, Spain. ⁶Institute of Molecular and Developmental Biology of Plants, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany. ⁷Sebastian Urzinger and Lukas Würstl contributed equally to this work. ✉email: chris.schoen@tum.de

the *ndhm1* locus, modulating cyclic electron transport, has pleiotropic positive effects on photosystem II (PSII) efficiency, cold tolerance and plant growth¹⁸.

Here, we focus on the target trait maximum quantum efficiency of PSII (F_v/F_m). F_v/F_m provides an estimate of how efficient light energy is used for photochemistry. In unstressed leaves F_v/F_m is stable, and a reduction in F_v/F_m is generally indicative of stress¹⁹. In maize, genetic variation in F_v/F_m has been shown to be correlated with multiple traits such as early plant development, chilling tolerance, PSII operating efficiency (Φ PSII), non-photochemical quenching (NPQ) and chlorophyll content^{11,20,21}. Thus, F_v/F_m can serve as a proxy for assessing genetic variation related to stress tolerance²², but it is also related to biomass production by modulation of light interception- and conversion-efficiency^{9,19,23,24}. Previous studies using biparental or multi-parent advanced generation inter-cross (MAGIC) populations have identified only a few quantitative trait loci (QTL) for F_v/F_m , each with relatively large effects^{20,25–28}. Candidate genes have been proposed for these QTL, but the physiological mechanisms underlying the variation of F_v/F_m in maize have remained largely unclear.

In this study we performed a comprehensive analysis of QTL for F_v/F_m in a population of 222 doubled-haploid (DH) lines derived from the maize landrace “Kemater Landmais Gelb” (Kemater). First, we conducted a genome-wide association study (GWAS), to explore the genetic architecture of F_v/F_m variation in the landrace. We then related the genetic variation in F_v/F_m to physiological parameters significantly influencing agronomically important traits. Our analysis led to the discovery of a genomic region on maize chromosome 10 harboring allelic variation at *light harvesting chlorophyll a/b binding protein6* (*lhcb6* also known as *CP24*), a gene encoding a component of the PSII light-harvesting complex, as a key determinant of F_v/F_m variation in Kemater. We subsequently characterized the impact of this allelic variation on transcript levels and the leaf proteome, as well as on the physiological trait NPQ and plant growth. We also investigated the prevalence of specific alleles at this gene in diverse maize inbred lines.

Results

Discovery of QTL affecting F_v/F_m in a maize landrace

In the genome-wide association study (GWAS), significant marker-trait associations were detected for maximum quantum efficiency of photosystem II (PSII, F_v/F_m ; Fig. 1A, Table 1) on chromosomes 1 (growth stage V6), 2 (V4, V6), 4 (V4), 8 (V4), and 10 (V4, V6). The QTL on chromosomes 1, 4 and 8 were detected in one growth stage only, while the QTL on chromosomes 2 and 10 were discovered in V4 and V6. Because the confidence intervals of the QTL on chromosomes 2 and 10 largely overlapped between the two growth stages, each genomic region was considered a single QTL affecting both growth stages in subsequent analyses. Together, the five QTL at the above-mentioned positions explained 50% and 59% of the genetic variance of F_v/F_m in growth stages V4 and V6, respectively (Table 1). Here, we focus on the QTL on chromosome 10 spanning a 3.9 Mb genomic region at the distal end of the chromosome (Fig. 1). This QTL had a substantial effect size on F_v/F_m of 0.45 (0.46) standard deviations (SD) and explained 35% (47%) of the genetic variance in V4 (V6). Due to a large effect in both growth stages, a biparental population was developed for fine-mapping and functional characterization of this region.

Genetic dissection of the genomic region on chromosome 10

Two Kemater DH lines, KE0095 and KE0109, were selected from the GWAS population as parents of a biparental population (Fig. 2A), based on their phenotypic difference in F_v/F_m (Supplementary Figure S1) and their genomic similarity assessed with the 600k Axiom™ Maize SNP Array (600k Array)²⁹. The two lines are polymorphic in the target region on chromosome 10 (Chr10:148,413,861–152,303,534; Fig. 1B), share the same SNP allele at the other four F_v/F_m QTL (Table 1) and have a similar genomic background with 419,014 out of 493,989 SNPs being monomorphic (Supplementary Figure S2). Chromosome 10 is nearly identical between KE0095 and KE0109 based on sequence alignment (Fig. 2B) and 30,120 out of 32,236 SNPs being monomorphic. Polymorphic SNPs on chromosome 10 were confined to a 22 Mb region (Chr10:130,130,769–152,357,073), covering the QTL for F_v/F_m . Thus, the biparental population allowed fine-mapping of the target region (Chr10:148,413,861–152,303,534) with no interference of other QTL. $F_{2,3}$ lines derived from the KE0095 x KE0109 cross, showing recombination events in the target region, were phenotyped for F_v/F_m at the three German field locations Freising (FRS), Roggenstein (ROG) and Einbeck (EIN) in 2024. Entry-mean heritability (h^2) for F_v/F_m was 0.63 (95% CI: 0.40–0.77), indicating a significant genetic component underlying trait expression in the $F_{2,3}$ recombinant lines. The association with F_v/F_m was assessed for 14 markers in the target region, of which eight were significantly associated with F_v/F_m (false discovery rate: FDR < 1%; Fig. 3A). The strongest association with F_v/F_m was observed for the most significant SNP from the GWAS analysis (AX-90599221; lead SNP; Supplementary Table S1). The KE0095 allele of the lead SNP had a trait increasing effect on F_v/F_m compared to the KE0109 allele, as expected from trait expression of the parents (Supplementary Figure S1). This marker explained 56% of the phenotypic and 98% of the genetic variance in the 52 $F_{2,3}$ lines, suggesting that the underlying gene is the major determinant of F_v/F_m segregation in this biparental population. The strength of the association with F_v/F_m diminished by several orders of magnitude up- and downstream of the lead SNP (Fig. 3A). Recombination events were observed between the lead SNP and neighboring SNPs (AX-91369070, AX-91196092; maximum $r^2=0.71$; Fig. 3B), allowing fine-mapping of the region. The fine-mapping narrowed the genomic region associated with F_v/F_m to a 154 kb fragment between the markers AX-91369070 and AX-91196092 (Chr10:150,109,398–150,262,989; blue box in Fig. 3A, Supplementary Table S1).

Molecular characterization of candidate genes

In the 154 kb region of interest, 13 genes are located in the B73v5 reference genome (Fig. 3C, Supplementary Table S2). Protein sequences of these genes were annotated and classified using MapMan4³⁰, and proteins belonging to category “Photosynthesis” were selected. One gene met the selection criterion: *Zm00001eb433540* (*light harvesting chlorophyll a/b binding protein6*; *lhcb6* also known as *CP24*). Since annotation alone reveals only

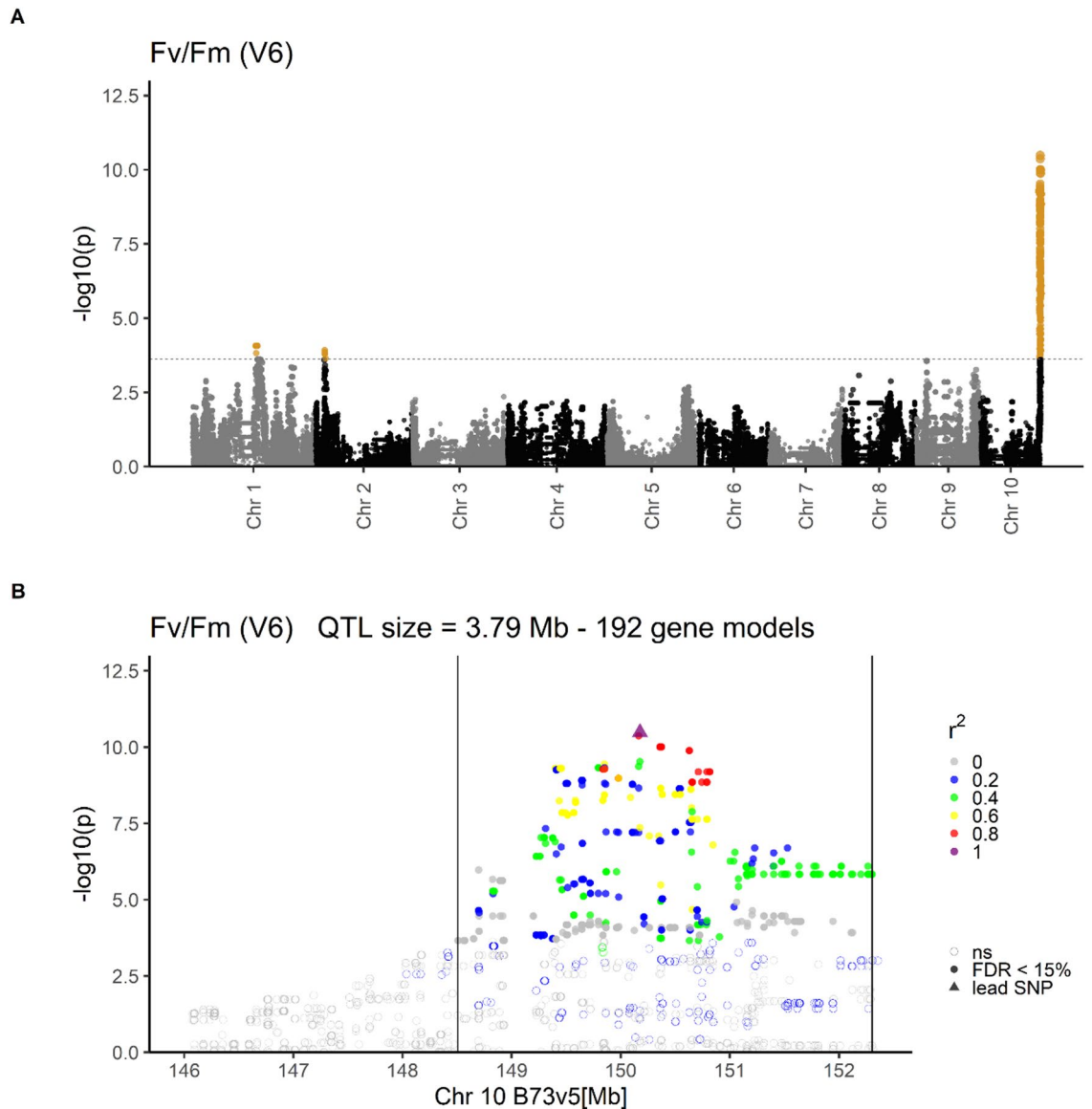


Fig. 1. Marker-trait associations from a genome-wide association study (GWAS) for maximum potential quantum yield of photosystem II (F_v/F_m) in growth stage V6. **(A)** Marker-trait associations across chromosomes. Phenotypic data were assessed on 211 DH lines derived from the landrace Kemater during early development in growth stage V6 in field experiments in two locations. The x-axis shows the ten chromosomes of maize and the y-axis the $-\log_{10}$ of the p-value of associations of SNPs with F_v/F_m . The significance threshold (false discovery rate FDR < 15%) is indicated by a dashed horizontal line. Marker-trait associations are color-coded: orange (significant); grey and black (not significant, ns). **(B)** Marker-trait associations for the QTL on chromosome 10. The x-axis indicates the physical position of SNPs on chromosome 10 in the reference sequence B73v5 and the y-axis the $-\log_{10}$ of the p-value of associations of SNPs with F_v/F_m . Significance of associations of SNPs with F_v/F_m (FDR < 15%) is indicated by symbols: empty circles (no significant association, ns); filled circles (significant association); triangle (SNP with highest significance, lead SNP, AX-90599221). Linkage disequilibrium (r^2) between the lead SNP and SNPs in the target region is color-coded. QTL boundaries are given by vertical lines.

genes with known functions, candidate gene prioritization was also conducted by comparing genomic sequences of KE0095 and KE0109.

Of the 13 genes in the target region, seven genes showed polymorphisms between KE0095 and KE0109, with five being non-synonymous leading to amino acid exchanges (Fig. 3C, Supplementary Table S3). Amino acid exchanges were non-conservative in two of these genes: *Zm00001eb433520* and *Zm00001eb433550*. The former encodes an uncharacterized protein and differs by multiple indels and four non-conservative amino acid exchanges between KE0095 and KE0109. The latter encodes PROTEIN S-ACYLTRANSFERASE38 and exhibits one non-conservative amino acid exchange (Supplementary Table S3). In addition, a 3.3 kb insertion directly upstream of the *lhcb6* translational start codon was detected in KE0109 which is not present in KE0095 and

QTL	Growth stage	Chr	Start	End	Genes	Effect SD	GVE	Total GVE
qV4FvFm2.1	V4	2	22,392,853	25,329,743	79	-0.43	7%	50%
qV4FvFm4.1	V4	4	244,037,990	244,041,963	1	-0.32	7%	
qV4FvFm8.1	V4	8	40,669,624	40,670,607	1	1.35	5%	
qV4FvFm10.1	V4	10	148,413,861	152,303,594	194	-0.45	35%	
qV6FvFm1.1	V6	1	160,150,163	163,660,341	25	0.33	30%	59%
qV6FvFm2.1	V6	2	24,398,414	27,269,195	81	-0.39	12%	
qV6FvFm10.1	V6	10	148,506,993	152,303,594	192	-0.46	47%	

Table 1. QTL for maximum potential quantum yield of photosystem II (F_v/F_m) in landrace Kemater. The genome-wide association study (GWAS) was conducted using approximately 500,000 SNPs and adjusted entry means of F_v/F_m in growth stages V4 and V6. Genomic positions are provided in B73v5 coordinates, and the number of genes within each QTL region is indicated. Effect SD: Absolute effect size from GEMMA output standardized by the standard deviation of the respective trait. GVE: Genetic variance explained by the respective QTL. Total GVE: Genetic variance explained by a model fitting all detected QTL in a growth stage simultaneously.

maize reference line B73 (Fig. 3D, File S1). The corresponding 5' sequence of *lhcb6* in B73 and KE0095 contains a sequence resembling a TATA box (TATTATTATTAAT; Chr10:150,168,719–150,168,732, B73v5; Fig. 3D, Supplementary Figure S3) and is highly conserved across 33 maize genomes (Supplementary Figure S4).

As a next step we investigated differences in transcript and protein abundance in leaves. In leaf proteomes of KE0095 and KE0109, 6,672 proteins were quantified. Of the 13 candidate genes two were represented (*Zm00001eb433540* [*lhcb6*], *Zm00001eb433610* [*Mg chelatase subunit H 1,chlH1*]; Supplementary Table S2). For *Zm00001eb433610* no difference in protein accumulation was observed between KE0095 and KE0109 (Supplementary Table S2), while protein accumulation of LHCB6 was reduced ~200 fold in KE0109 compared to KE0095 (Fig. 4B). The absence of the other 11 candidates in leaf proteomes of KE0095 and KE0109 cannot rule them out as candidate genes for the QTL but makes them less likely. For the two genes with non-synonymous amino acid exchanges (*Zm00001eb433520*, *Zm00001eb433550*) and *lhcb6*, transcript abundance was measured by RT-qPCR. No differences in transcript levels of *Zm00001eb433520* and *Zm00001eb433550* were observed between KE0095 and KE0109 (Supplementary Figure S5), while a 1000-fold reduction of transcript levels of *lhcb6* was detected in KE0109 compared to KE0095 (Fig. 4A).

Assuming that sequence variation at *lhcb6* determines differences in F_v/F_m , we further characterized variation at *lhcb6*. The 3.3 kb insertion is highly similar to a putative hAT transposon that is located approximately 600 kb downstream of *lhcb6* in B73 according to MaizeGDB (File S2, DTA_ZM00039). The insertion contains two long open reading frames (313 and 492 amino acids), encoding domains typically found in transposases—a TTF zinc finger domain, a Ribonuclease H-like superfamily domain, and a hAT C-terminal dimerization domain (InterPro accessions: IPR006580, IPR008906 and IPR012337). However, no terminal inverted repeats at the insertion site, which are characteristic for hAT transposons were detected in KE0109. Further, KE0095 contains a 14 bp insertion within the only intron of *lhcb6* compared to both B73 and KE0109 (Fig. 3D, File S1). Despite these differences in their DNA sequences, the amino acid sequences of LHCB6 in KE0095, KE0109, and B73 are identical. Hereafter, the KE0095 allele is referred to as *lhcb6-A* and the KE0109 allele with the 3.3 kb insertion is referred to as *lhcb6-B*.

Recently, *lhcb6* was proposed as a candidate gene for a QTL for F_v/F_m in a maize MAGIC population, with maize line F7 carrying a deficient allele at this locus²⁰. We compared the genomic sequences of F7, KE0095 and KE0109 and found that F7 carried an identical insertion in its putative promoter sequence of *lhcb6* as KE0109, corroborating an effect of the insertion on *lhcb6* transcript levels in a different genomic background.

To investigate if the variation in *lhcb6* transcript and protein levels is specific to KE0095 and KE0109, 27 Kemater lines were selected from the population of DH lines used in GWAS analyses for validation experiments. The DH lines were chosen to represent the two alleles of the lead SNP of the F_v/F_m QTL on chromosome 10 in equal proportion. The presence or absence of the *lhcb6-B* allele was determined by PCR analysis, with primers spanning the putative transposon insertion. The *lhcb6-B* allele was detected in 12 of the 27 Kemater lines and was in complete linkage disequilibrium (LD) with the lead SNP from the GWAS in this set of Kemater lines (Fig. 4C, Supplementary Table S4). Corroborating results obtained with the two Kemater lines KE0095 and KE0109, RT-qPCR analysis revealed that *lhcb6* transcript levels were strongly decreased in the 12 Kemater lines carrying *lhcb6-B* compared to the 15 lines carrying *lhcb6-A* (Fig. 4D). The 27 Kemater lines under study also harbor different alleles at *ndhml*, a gene which was shown to affect F_v/F_m and early plant height in maize¹⁸. Correspondingly, of the 12 lines with *lhcb6-B*, LHCB6 protein could either not be detected ($n=6$) or was strongly reduced ($n=6$) compared to the 15 lines with *lhcb6-A* (Fig. 4E).

As expected from GWAS findings and results from $F_{2,3}$ recombinant lines, mean F_v/F_m was significantly ($p < 0.01$) decreased in DH lines with reduced LHCB6 protein levels in growth stage V6 (Fig. 4F, Supplementary Table S5). Further, early plant height in growth stage V6 was also significantly ($p=0.02$) reduced in DH lines carrying the *lhcb6-B* allele (Fig. 4G, Supplementary Table S5). These results suggest that the transposon insertion in *lhcb6-B* diminishes *lhcb6* transcript abundance, leading to lower LHCB6 protein levels and thereby lower F_v/F_m .

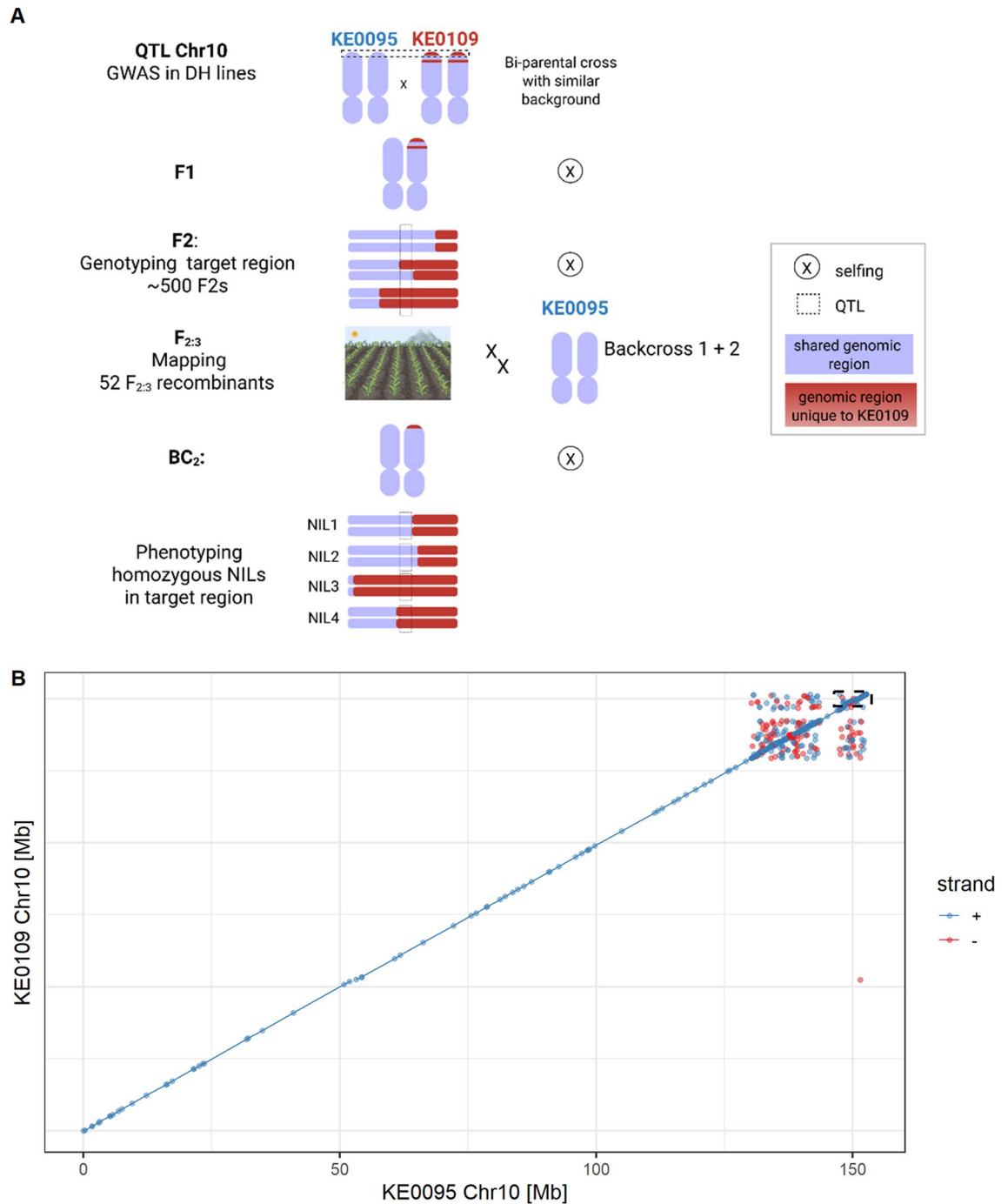
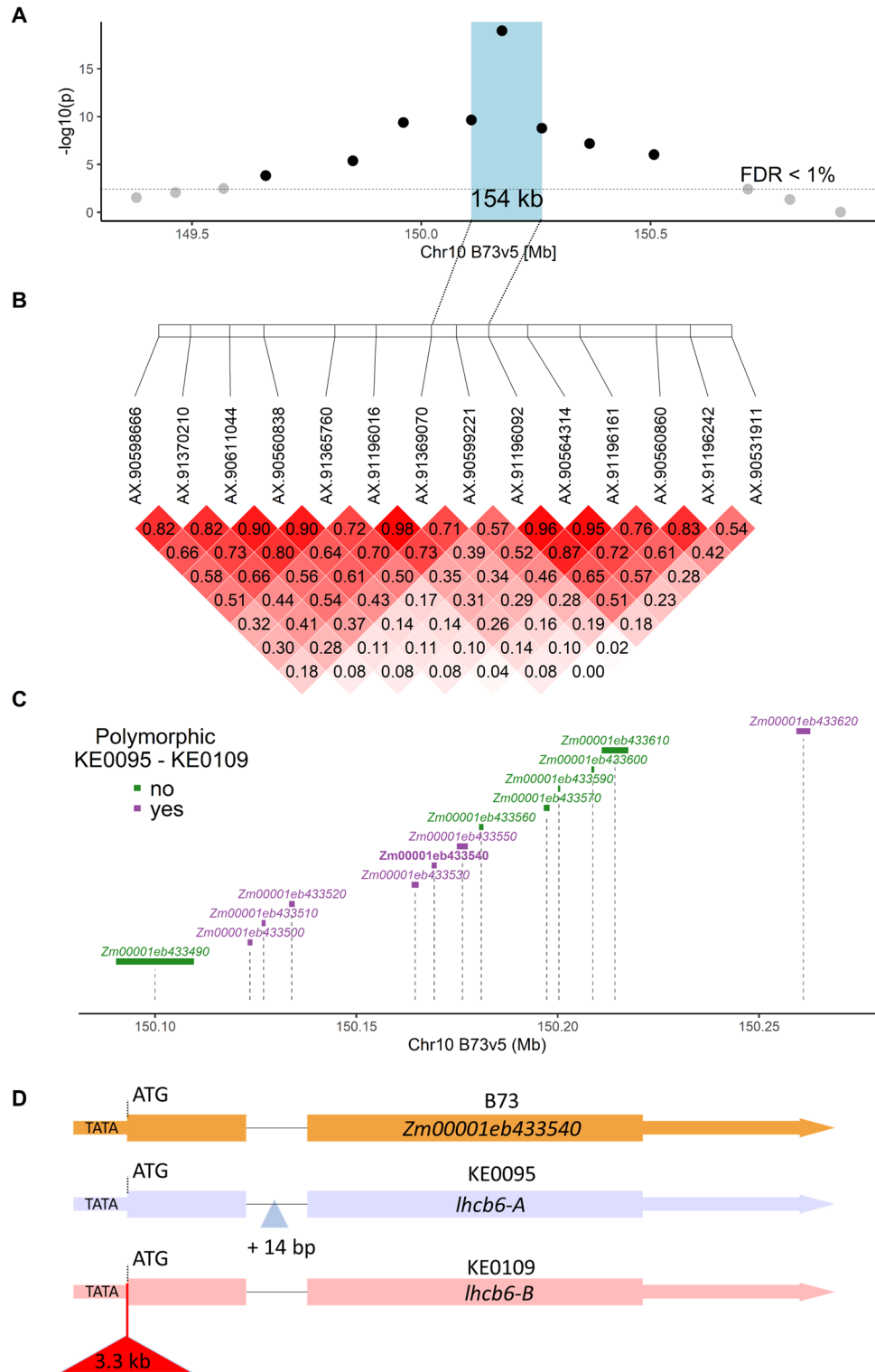


Fig. 2. Material development for fine-mapping of a QTL for F_v/F_m on chromosome 10. **A**) Schematic representation of the development of near isogenic lines segregating for a genomic region on chromosome 10 associated with F_v/F_m . Genomic fragments unique to the parental line KE0109 are highlighted in red. The QTL region is indicated by a dashed box. **B**) Pairwise sequence alignment of KE0095 and KE0109, the parents of the bi-parental population. The dashed box at the distal end of chromosome 10 marks the target region, which harbors the QTL for F_v/F_m . Alignments > 1 kb with sequence identity > 95% are shown: blue represents alignments on the forward strand, red indicates alignments on the reverse strand. Figure 2A was created with BioRender.com.

To determine whether *lhcb6-B* is specific to the Flint maize heterotic group, haplotype diversity at this locus was investigated in 136 diverse inbred maize lines analyzed by Unterseer et al.³¹ and 501 DH lines derived from the Kemater landrace³². Haplotypes were defined by concatenating the 15 SNPs closest to *lhcb6* (8 upstream, 7 downstream). We identified 19 haplotypes with absolute counts ranging from 3 to 237 in the entire dataset. The most common haplotype (Hap1) had frequencies of 45% and 21% in the Kemater landrace and Flint



breeding lines, respectively, but was absent in elite Dent lines (Supplementary Figure S6). Based on the results of PCR genotyping, Hap1 was in complete LD with the insertion in *lhcb6-B* in the subset of 27 Kemater lines (Supplementary Table S4). These findings can be extended from the landrace population, provided that the association between Hap1 and *lhcb6-B* is consistent in breeding lines.

Functional characterization of the *lhcb6-B* allele

To evaluate the effect of the *lhcb6-A* and *lhcb6-B* alleles on phenotypic traits other than F_v/F_m , we developed four near isogenic lines (NILs) with contrasting *lhcb6* alleles from the $F_{2,3}$ recombinant lines. NIL 1 and NIL 2 differ from NIL 3 and NIL 4 in a 154 kb genomic fragment harboring *lhcb6* and 12 other genes defined by fine-mapping in $F_{2,3}$ recombinant lines (Fig. 3). NIL 1 and NIL 2 carry *lhcb6-A*, while NIL 3 and NIL 4 carry *lhcb6-B* (Supplementary Figure S7, Supplementary Table S6). In addition to F_v/F_m , we measured the effective

◀ **Fig. 3.** Characterization of a QTL region associated with F_v/F_m in $F_{2:3}$ recombinant lines. **A)** Association of 14 SNP markers with F_v/F_m in $F_{2:3}$ recombinant lines ($n = 52$) based on adjusted entry means across three field locations. The horizontal line indicates the false discovery rate (FDR) of 1%. The fine-mapped QTL region between the flanking markers of the marker with highest significance is indicated by a blue-shaded box. **B)** Linkage disequilibrium (r^2) between 14 SNPs in the $F_{2:3}$ recombinant lines. Dashed lines indicate the fine-mapped QTL region. **C)** Gene models located between the flanking markers of the fine-mapped QTL region. The size of the gene models is indicated by rectangles and their position on chromosome 10 in the reference sequence B73v5 is indicated by dashed vertical lines. Gene models with polymorphisms between KE0095 and KE0109 are shown in purple, and monomorphic gene models between KE0095 and KE0109 are shown in green. **D)** Gene models of the gene light harvesting chlorophyll a/b binding protein6 (*lhcb6*) in B73 (*Zm00001eb433540*), KE0095 and KE0109. Exons are indicated with rectangles, introns with a thin line and UTRs with a thick line. KE0109 carries a 3.3 kb genomic insertion immediately upstream of the start codon (red triangle). KE0095 carries a 14 bp insertion in the only intron of the gene.

PSII antenna size, NPQ, and fresh and dry biomass of the NILs. As expected, F_v/F_m was lower in NILs carrying *lhcb6-B* compared to NILs carrying *lhcb6-A* (Fig. 5A), which was the result of a combination of higher minimal fluorescence (F_0 ; Supplementary Figure S8A) and lower maximal fluorescence (F_m ; Supplementary Figure S8B). Effective antenna size of PSII was determined from the half-time needed to reach the J level of the initial O-J rise in an OJIP chlorophyll induction curve³³. The half-time was significantly lower in KE0109 and NILs carrying the *lhcb6-B* allele indicating larger effective antenna size (Fig. 5B). During the light period, maximum NPQ in NILs with *lhcb6-B* and KE0109 was approximately half of that observed in NILs with *lhcb6-A* and KE0095 at the given actinic light intensity (Fig. 5C). Conversely, NPQ values at the end of the dark period were increased in NILs with *lhcb6-B* and KE0109 (Fig. 5C), suggesting impaired relaxation of NPQ. Fresh and dry biomass were reduced in NILs with *lhcb6-B* compared to NILs with *lhcb6-A*, but this reduction was not observed in KE0109 compared to KE0095 (Fig. 5D, E).

We further explored the consequences of reduced *lhcb6* expression on PSII light-harvesting complex (LHCII) antenna components in the subset of 27 Kemater lines described earlier. Based on sequence similarity to experimentally annotated Arabidopsis (*Arabidopsis thaliana*) LHCII components, 13 putative LHCII subunits were detected in the maize reference line B73 (Supplementary Table S7). One of these is an uncharacterized paralog of LHCB6 (*Zm00001eb066480*, LHCB6-2). LHCB6-2 has an amino acid sequence identity of 93% (235/253 amino acids) with LHCB6-1, including a four amino acid deletion at the N-terminus in LHCB6-2 compared to LHCB6-1. Of these 13 proteins, 12 were detected and quantified in the leaf proteomes of the 27 Kemater lines, including the three major LHCII antenna proteins (LHCB1, LHCB2 and LHCB3) and the three minor LHCII antenna proteins (LHCB4, LHCB5 and LHCB6). In addition to the expected reduction in LHCB6, LHCB3 (*Zm00001eb324240*) protein levels were decreased by ~50% in *lhcb6-B* lines. The abundance of the other 10 quantified LHCII proteins, including LHCB6-2, remained unchanged. We also analyzed differential accumulation among all 5,531 quantified proteins. Only six proteins exhibited differential accumulation based on *lhcb6* allele status (Supplementary Table S8, Fig. 6). The *lhcb6-B* allele was correlated with a reduced abundance of four proteins and an increased abundance of two proteins (Fig. 6, Supplementary Table S8). Notably, LHCB6 showed the most significant reduction, followed by LHCB3 (Fig. 6A).

Of the two other proteins with decreased accumulation in Kemater lines with *lhcb6-B* (*Zm00001eb068760*, *Zm00001eb289810*), only one has a clearly defined function: a 3-oxoacyl-[acyl-carrier-protein]-synthase III (KAS III; *Zm00001eb068760*). KAS III catalyzes the first step in the fatty acid synthesis pathway II, which is essential for the synthesis of membrane phospholipid acyl chains³⁴.

Among proteins with increased abundance, CHLOROPHYLLIDE A OXYGENASE1 (CAO1; *Zm00001eb039350*) showed the strongest upregulation—a threefold increase in *lhcb6-B* lines (Fig. 6, Supplementary Table S8). Another protein with increased accumulation is of unknown function (*Zm00001eb271080*).

Discussion

A profound understanding of the genetic basis of photosynthetic efficiency can leverage the improvement of agricultural productivity^{9,24,35}. Previous research has established a genetic correlation between the maximum quantum yield of photosystem II (PSII, F_v/F_m) and agronomically relevant traits like early development and chilling stress tolerance in maize, and links it to other photosynthetic traits, including PSII operating efficiency (ΦPSII), non-photochemical quenching (NPQ), and chlorophyll content^{11,18,20,21}. QTL with large effects on F_v/F_m have been mapped in maize and several candidate genes have been proposed, but their underlying genetic mechanisms have remained largely unknown^{20,25–28,36}.

In this study, we identified five QTL for F_v/F_m across two early growth stages segregating in a maize landrace. The causal gene underlying the QTL on chromosome 2 (*ndhm1*) with pleiotropic effects on early plant growth and F_v/F_m has been characterized by our group recently and encodes a subunit of the NADH-dehydrogenase like (NDH) complex¹⁸. Here, we could show that *lhcb6* (also known as *CP24*), a minor antenna component of the light-harvesting complex II (LHCII), is the causal factor underpinning the major QTL for F_v/F_m on chromosome 10. Both *lhcb6* and *ndhm1* play key roles in dynamic regulation of photosynthesis, by balancing photoprotection and efficiency via NPQ and cyclic electron transport, respectively^{37,38}. We therefore conclude that F_v/F_m can serve as a proxy trait for identifying meaningful genetic variation that contributes to photosynthetic regulation in changing environmental conditions.

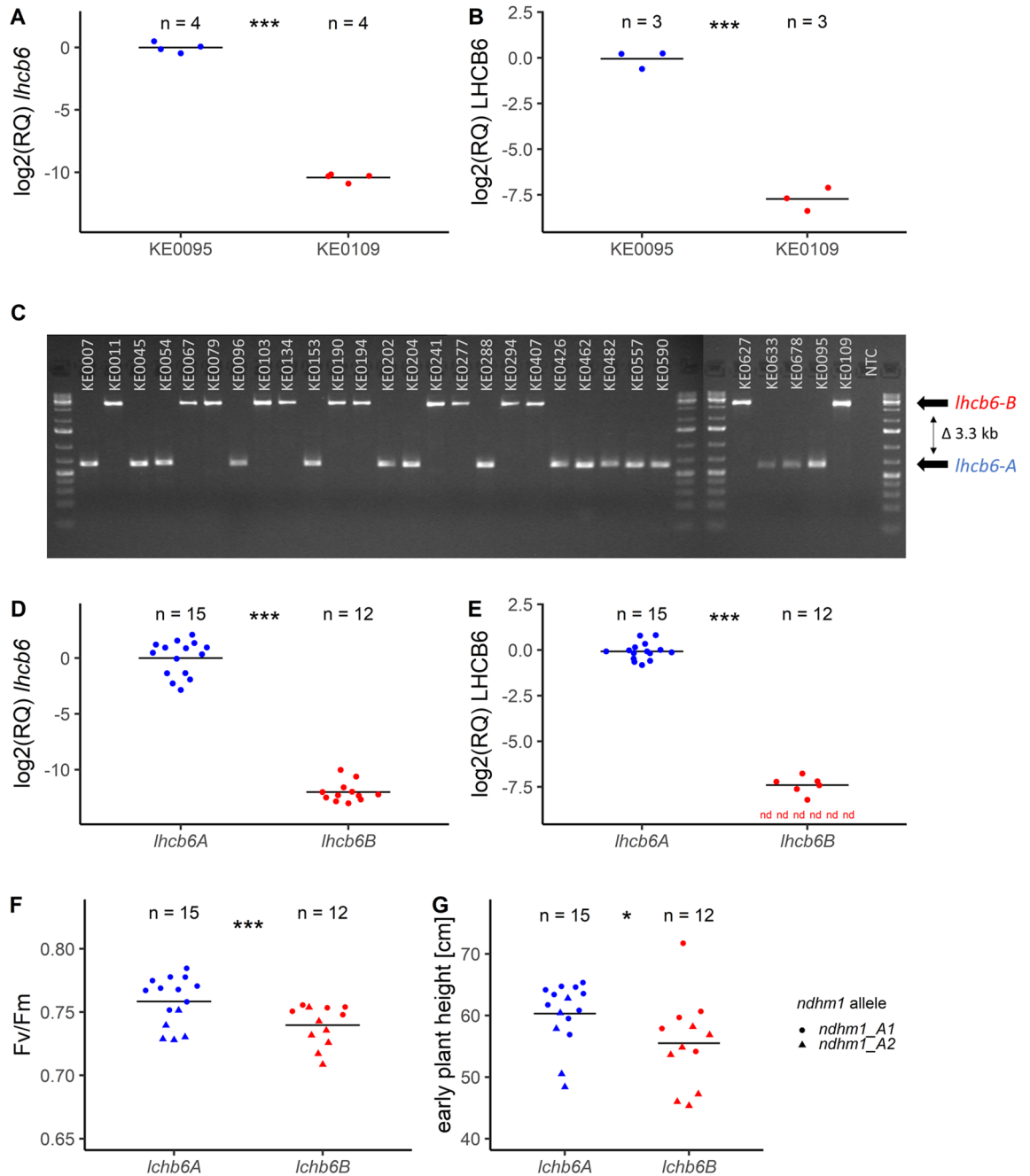


Fig. 4. Molecular characterization of the candidate gene *lhcb6*. **A, B** Log₂-transformed relative transcript (**A**) and protein (**B**) levels of *lhcb6* in KE0095 and KE0109, normalized to KE0095. **C** Genotyping of a 3.3 kb insertion in the genomic region upstream of the *lhcb6* start codon, assessed in a set of 27 Kemater lines and the parental lines KE0095 and KE0109. The lower band in the gel corresponds to the *lhcb6-A* allele and the higher band to the *lhcb6-B* allele, which contains the 3.3 kb genomic insertion. NTC: no template control **D-G** Differences in transcript (**D**) and protein (**E**) abundance, F_v/F_m (**F**) and early plant height (**G**) between 27 Kemater lines with contrasting *lhcb6* alleles. *lhcb6* (*lhcb6-A*, $n = 15$; *lhcb6-B*, $n = 12$) and *ndhm1* (*ndhm1-A1*, $n = 15$; *ndhm1-A2*, $n = 12$) segregate in the 27 Kemater lines. Genotypic data for *ndhm1* were derived from Urzinger et al.¹⁸. Significance was assessed by a two-way analysis of variance (ANOVA) including *lhcb6*, *ndhm1* and their interaction as factors. In **F** and **G** the *lhcb6* allele is indicated by color and *ndhm1* allele by shape of data points. Horizontal bars indicate group means and dots individual plant observations. Statistical differences were assessed by analyses of variance (ANOVA). *** $p < 0.001$; * $p < 0.05$. In **E** nd stands for not detected.

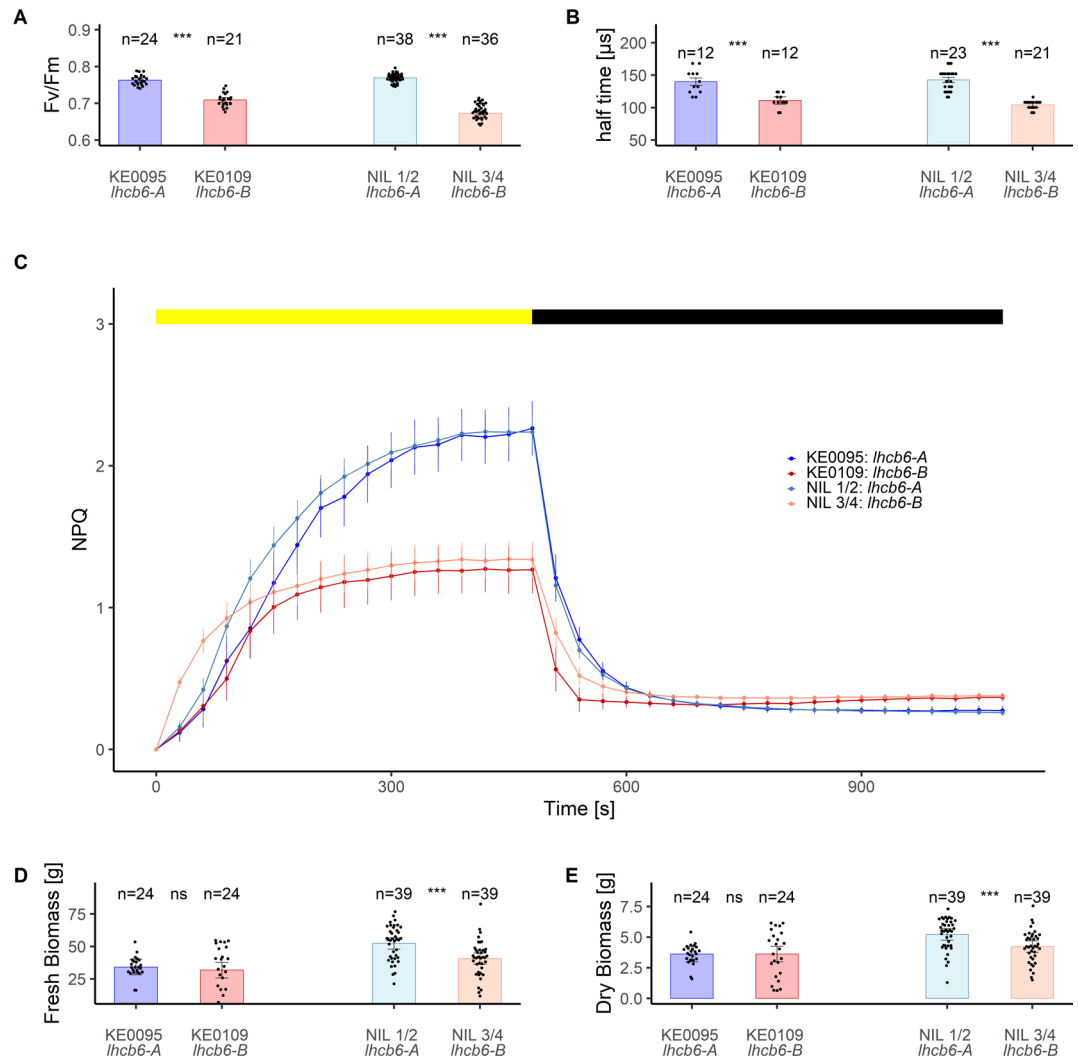


Fig. 5. Physiological and agronomic evaluation of near isogenic lines (NILs) with contrasting *lhcb6* alleles and of parental lines KE0095 and KE0109. Plants were grown in a growth chamber under fluctuating light intensity until growth stage V6. **A**) Maximum quantum yield of photosystem II (F_v/F_m). **B**) Functional antenna size estimation based on the half-time to the J level ($T_{1/2}$) during an OJIP fluorescence induction curve. **C**) Non-photochemical quenching (NPQ) dynamics during an eight-minute light phase (yellow bar; actinic light intensity: $1500 \mu\text{mol m}^{-2}\text{s}^{-1}$) followed by a ten-minute dark phase (black bar). Dots represent adjusted means, bars represent standard errors (SE; KE0095: $n = 4$, KE0109: $n = 4$, NIL1/2: $n = 8$, NIL3/4: $n = 7$). **D**) Fresh biomass. **E**) Dry biomass. Measurements were taken at growth stages V5–V6. Significant differences based on Student's t-tests are indicated by stars. *** $p < 0.001$; ns: not significant. Bars represent adjusted means \pm SE and dots represent individual plant observations (**A**, **B**, **D**, **E**).

In a recent publication, Ferguson et al.²⁰ proposed *lhcb6* as candidate gene for a QTL affecting F_v/F_m , ΦPSII and NPQ in a maize multi-parent advanced generation inter-cross (MAGIC) population. They showed that a specific haplotype found in the Flint line F7 was associated with reduced transcript and protein levels of *lhcb6* compared to the maize reference line B73. Our current study detected a QTL in the same genomic region in a Flint maize landrace. Ferguson et al.²⁰ based their evidence for *lhcb6* being the causal gene on functional annotation, physiological similarity to Arabidopsis mutants, and expression differences between B73 and F7, leading to impaired assembly of higher order PSII-LHCII supercomplexes. By mapping short-read sequencing data from F7 against the B73 reference, they identified three SNPs and one indel specific to the F7 haplotype altering computationally predicted transcription factor binding sites (TFBS). Here, we used a forward genetic approach with near isogenic lines (NILs) derived from a maize landrace in combination with long-read genome assemblies of their two parental lines KE0095 and KE0109 and we could pinpoint the causal polymorphism underlying this QTL to a 3.3 kb insertion of a hAT transposon directly upstream of the start codon of *lhcb6* in the *lhcb6-B* allele. The functional importance of the affected region as promoter of *lhcb6* is supported by its conservation across a diverse panel of maize lines, its overlap with a region of accessible chromatin, indicating its accessibility to transcription factors³⁹, computationally predicted TFBS²⁰, and a putative TATA-box. The insertion of a hAT transposon in *lhcb6-B* shifts the putative promoter region of *lhcb6* 3.3 kb upstream, extending

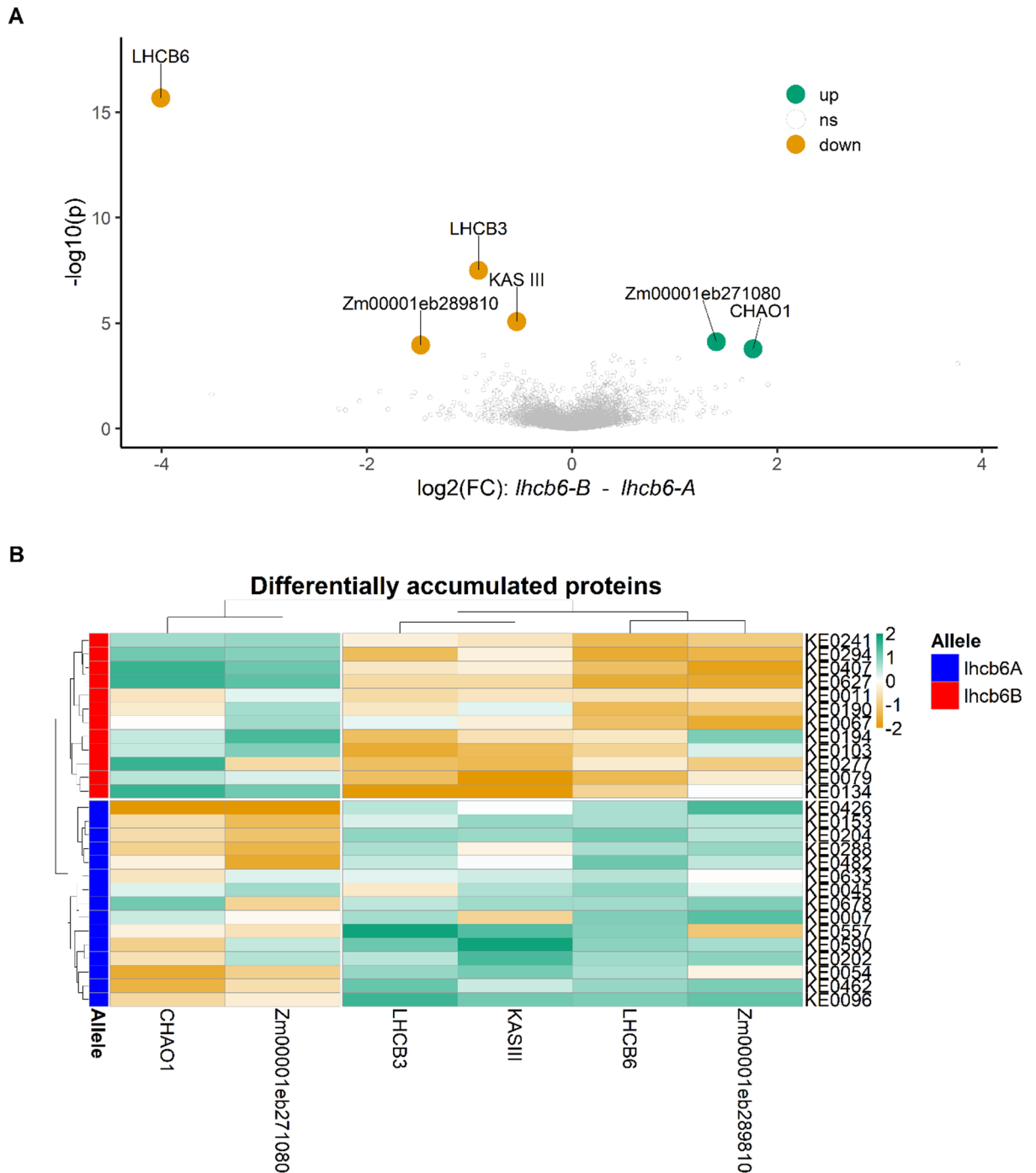


Fig. 6. Leaf proteome analysis of 27 Kemater lines segregating for *lchb6*. **A**) Differentially accumulated proteins in 27 Kemater lines grouped by their *lhcb6* allele. Proteins were considered differentially accumulated at a false discovery rate (FDR) < 20%. Proteins with lower abundance in lines carrying the *lhcb6-B* allele are shown in orange, while those with higher abundance are shown in green. **B**) Relative quantities of six significantly differentially accumulated proteins from (A) expressed as z-scores for each of the 27 Kemater lines. Lines carrying the *lhcb6-A* allele are indicated by the blue and those with the *lhcb6-B* allele with the red bar.

the typical distance of *cis*-regulatory elements to the transcription start site in maize⁴⁰. We identified the same mutation in the Flint line F7, linking our findings with those of Ferguson et al.²⁰. Thus, our results strongly support the hypothesis that *cis*-genic allelic variation at *lhcb6* is causative for differential gene expression, leading to significantly reduced RNA and protein levels, and further enabled us to identify the specific underlying structural variation.

In most vascular plants, the LHCII antenna consists of major antennae (homo- or heterotrimers of LHCb1, LHCb2, and LHCb3) and minor antennae (monomers of LHCb4, LHCb5, and LHCb6)^{41,42}. These

antennae assemble into PSII supercomplexes (e.g., $C_2S_2M_1$, $C_2S_2M_2$, C_2S_2), which are composed of two core (C) subunits of PSII, two strongly (S) bound trimers, and zero, one, or two moderately (M) bound trimers. In these supercomplexes, the M trimer is coordinated by LHCB3 and LHCB6⁴³. While nearly all mechanistic insights into LHCII organization and function have come from studies on *Arabidopsis thaliana* mutants it has remained unclear whether the proposed principles hold in a C_4 crop with different photosynthetic architecture. Here, we provide insights into LHCII function and its natural variation in maize. Our proteome analyses show both significantly reduced LHCB6 and LHCB3 levels in DH lines carrying the *lhcb6-B* allele reflecting the direct interaction between the two proteins. In *Arabidopsis*, a lack of LHCB6 protein impairs LHCII assembly, favoring the formation of C_2S_2 complexes⁴⁴. Conversely, we observed that estimates of the functional antenna size based on chlorophyll fluorescence induction curves were increased in NILs carrying the *lhcb6-B* allele and KE0109, matching observations in F7²⁰ and *lhcb6* and *lhcb6 lhcb3* double mutants of *Arabidopsis*⁴⁵. Ilíková et al.⁴⁵ proposed that despite the changes in LHCII assembly, the M-trimers remain functionally connected to PSII arrays in the *lhcb6* mutants, explaining the increased functional antenna size. Further, they suggest that the unbound domains of LHCII are responsible for increased F_0 values observed in the mutants. In accordance with these findings, we also observed no change in the abundance of light harvesting proteins (Supplementary Table S7) other than LHCB6 and LHCB3 in the subset of 27 Kemater lines and increased F_0 in NILs carrying *lhcb6-B* and in parent KE0109.

Ferguson et al.²⁰ proposed that *lhcb6-B* might be specific for the Flint heterotic group. Here, we could show that, despite its negative effect on LHCB6 abundance, it is present at high frequency in Flint breeding lines and is absent in Dent lines with publicly available genomic information. Whether the Flint specificity of *lhcb6-B* reflects beneficial effects of this allele in temperate climates or has arisen solely from random genetic drift has remained unclear and warrants further investigation.

LHCII plays a dual role in light harvesting and photoprotection. It captures and transfers light energy to PSII and is involved in photoprotection through NPQ, a mechanism that dissipates excess light energy as heat³⁷. Genetic variation for NPQ has been reported for maize and sorghum (*Sorghum bicolor*) populations^{46,47}. In maize, LHCII regulation responds to light and temperature⁴⁸, with implications for state transitions and energy balance between the two photosystems^{49,50}. Thus, targeting LHCII dynamics is a potential strategy to enhance stress resilience⁵¹ and maize landraces can enhance our understanding of these complex physiological processes^{9,13}.

Interestingly, although *lhcb6* mutants severely impair growth in *Arabidopsis*^{44,52,53}, the *lhcb6-B* allele in maize causes only mild growth reductions, suggesting compensatory mechanisms in the genome. To explore this hypothesis, we analyzed genomic and proteomic data in our landrace population with the aim of identifying factors that attenuate the effects of *lhcb6-B* in maize. Our analysis revealed a previously uncharacterized paralog of LHCB6 in maize, which we designated as LHCB6-2 (Zm00001eb066480). While paralogs for LHCB1, LHCB2, and LHCB4 have been reported in *Arabidopsis*⁵⁴, this is, to our knowledge, the first description of a LHCB6 paralog. The two maize LHCB6 paralogs are highly similar (File S3), with most differences occurring at the N-terminus, a region that includes a potential phosphorylation site⁴². Although phosphorylation is known to regulate LHC proteins in species such as spinach (*Spinacia oleracea*), *Arabidopsis*, and *Chlamydomonas reinhardtii*, it has not been documented for LHCB6^{55,56}. As LHCB6-2 cannot compensate for the deficiency of LHCB6-1 (Zm00001eb433540) in the Kemater DH lines, we propose that LHCB6-1 and LHCB6-2 may have distinct roles. While LHCB6-1 is important for photoprotection balanced by NPQ, LHCB6-2 may primarily facilitate efficient light capture for photosynthesis, partially compensating for negative effects on biomass accumulation. However, this hypothesis requires further analyses which are beyond the scope of this study.

Comparing the leaf proteome between contrasting Kemater lines carrying either the *lhcb6-A* or the *lhcb6-B* alleles, we observed co-expression of several proteins shedding light on the dynamic regulation of the LHCII antenna. For instance, CHLOROPHYLLIDE A OXYGENASE1 (CAO1; also known as CHAO1), which converts chlorophyll a to chlorophyll b and is essential for LHCII antenna assembly⁵⁷, shows elevated levels in lines carrying *lhcb6-B*. Increased CAO1 may serve as a compensatory feedback mechanism in response to LHCII antenna deficiency. A connection between chlorophyll b levels and LHCII efficiency has been shown in other species. In *Arabidopsis* and barley (*Hordeum vulgare*), deficiencies in chlorophyll b have been shown to lead to reduced accumulation of LHCII components^{58,59}, and overexpression of CAO1 in *Arabidopsis* enlarged the LHCII antenna⁶⁰. Further, an uncharacterized protein potentially involved in lipid transfer (Zm00001eb271080) has been linked to variation in maize tocopherol and tocotrienol content⁶¹. Given the protective role of tocopherols against photoinhibition and lipid photooxidation⁶², its increased protein levels might indicate increased oxidative stress associated with impaired NPQ, similar to the increased reactive oxygen species (ROS) observed in *Arabidopsis lhcb6* mutants⁶³. Reduced protein levels of KAS III, an enzyme involved in lipid biosynthesis, may reflect disturbances in the assembly of the PSII–LHCII supercomplex, where lipids play a stabilizing role⁶⁴.

Our findings highlight the value of natural genetic variation for unraveling the complex regulation of photosynthesis and for enhancing crop performance under stress conditions with significant implications for breeding. The development of a diagnostic marker capturing the allelic state of *lhcb6*—including the hAT transposon insertion in the promoter—and using this marker alone or in combination with two or more flanking markers to define a diagnostic haplotype will facilitate the evaluation of the gene's effect in segregating populations and enable any type of marker-based selection including whole-genome selection. In addition, the screening of additional natural genetic variation and the generation of new variation at *lhcb6* via chemical induced mutations, genome editing, or classical transgenic approaches may allow engineering of the LHCII antenna and modification of NPQ in breeding material. Beneficial alleles detected or created in this manner can be introgressed into elite material through marker assisted backcrossing.

Materials and methods

Plant Material

The landrace “Kemater Landmais Gelb” (Kemater) was chosen as the base population for this study due to its phenotypic variation for early development, as well as its fast decay of linkage disequilibrium (LD) and absence of pronounced population structure. Kemater was selected from a set of 35 European maize landraces covering a broad geographical region described by Mayer et al.⁶⁵. Kemater represents 77% of the molecular variance of the entire collection¹¹ based on the 600k Axiom™ Maize Genotyping Array (600k Array)²⁹. Directly from the landrace, 516 doubled-haploid (DH) lines were produced and multiplied using the in vivo haploid induction method⁶⁶, and 222 (growth stage V4) and 211 (V6) DH lines were phenotyped for F_v/F_m in 2017 and 2018. A curated dataset comprising 501 Kemater DH lines genotyped with 501,124 markers from the 600k Array was obtained from Mayer et al.³².

A subset of 27 DH lines was selected from the full set of Kemater lines for functional validation of a QTL for F_v/F_m on chromosome 10 (Chr10:148,413,861–152,303,534; Table 1, Fig. 1), including phenotypic, transcript and proteome measurements. These lines were chosen to obtain a balanced frequency of two contrasting alleles at the QTL (validation set; Supplementary Table S4).

All plant-related experiments were conducted in accordance with the internal guidelines of the Technical University of Munich as well as all applicable national and international regulations. Compliance with the Nagoya Protocol for all material derived from the landrace Kemater was ensured through standard material transfer agreements.

Development of recombinant and near isogenic lines

A schematic overview of the material development for fine-mapping is presented in Fig. 2A. Two Kemater DH lines, KE0095 and KE0109, were chosen for their contrasting alleles at the QTL for F_v/F_m on chromosome 10, while being as similar as possible in their genomic background (Fig. 2B, Supplementary Figure S2) and crossed to produce a biparental population. KE0095 and KE0109 are polymorphic for 74,975 of 493,989 markers of the 600k Array²⁹. KASP markers polymorphic between the two parental lines and positioned in the QTL region on chromosome 10 were synthesized using probe sequences of the 600k Array. F_2 lines were genotyped with six KASP markers and individual plants showing recombination between markers AX-90560736 (Chr10:143,112,717; Supplementary Table S1) and AX-91196698 (Chr10:152,357,073; Supplementary Table S1) were self-pollinated. Resulting $F_{2,3}$ recombinant lines were genotyped with 14 KASP markers (Supplementary Table S1) and phenotyped for F_v/F_m in field experiments at three German locations in 2024. Selected $F_{2,3}$ recombinants were backcrossed two times to KE0095 as the recurrent parent. BC_2 was genotyped with a 15k SNP chip to select individual plants with minimum donor parent genome. These were selfed to obtain four near isogenic lines (NILs) which were phenotyped in growth chamber experiments.

Field experiments

The Kemater DH lines were evaluated in 2017 and 2018 as part of a broader study on landrace diversity¹¹. Briefly, in two locations, Golada and Tomeza, 222 Kemater DH lines were phenotyped for F_v/F_m . The experimental design was a 10×10 lattice design with two replications at each site. As checks, fourteen Flint and one Dent inbred line as well as the original landrace population (each replicated four times) were used. Further information on field trial locations, check varieties, and methods for phenotypic data analysis is available in the publications of Hölker et al.¹¹ and Mayer et al.³².

Phenotypic evaluation of the set of 27 Kemater DH lines and the population of recombinant $F_{2,3}$ lines used for fine-mapping was conducted in Einbeck (EIN, Germany, 51°49'05.9"N 9°52'00.3"E), Roggenstein (ROG, Germany, 48°10'47.5"N 11°19'12.9"E), Bernburg (BBG, Germany, 51°49'28.6"N 11°42'26.3"E) and Freising (FRS, Germany, 48°24'13.2"N 11°43'28.5"E). The subset of 27 Kemater DH lines was grown in a randomized complete block design (RCBD) with two replications in locations ROG, FRS and Oberer Lindenhof (OLI, Germany, 48°28'26.3"N 9°18'17.9"E) in 2023 together with four inbred checks. The 52 $F_{2,3}$ recombinant lines were evaluated in a RCBD with two replications in EIN, ROG and FRS in 2024 using the same inbred checks. Plots consisted of 20 plants grown in single rows of 3 m length with 0.75 m spacing between them (9 plants/m²). Field trials were subjected to standard agricultural practices. Phenotypic observations from plots containing at least five plants were filtered by a Grubbs' outlier test⁶⁷.

Growth conditions in growth chamber experiments

For phenotyping of the four NILs, alongside their recurrent parent (KE0095) and donor parent (KE0109) in growth chamber experiments, kernels were immersed in water for 5 min, placed on filter paper and allowed to pre-germinate in the dark at 28 °C for 72 h. After germination, samples from individual seedlings were collected for DNA extraction and the seedlings were transplanted to small pots filled with CL ED73 soil (Einheitserdewerke Patzer, Germany) as described in Urzinger et al.¹⁸. Plants were grown in growth chambers with 16/8 h day night [d/n], 25/20 °C d/n, 75% relative humidity [RH] until growth stage V4 to V6. To simulate fluctuations in light intensity occurring in the field, a light switch scheme was employed, alternating between 30 min high light (800 $\mu\text{mol m}^{-2}\text{s}^{-1}$ photosynthetically active radiation [PAR]) and 15 min low light (100 $\mu\text{mol m}^{-2}\text{s}^{-1}$ PAR). The four NILs and the two parental DH lines KE0095 and KE0109 were replicated 10 times in each of the two blocks in a RCBD. The developmental stages of maize were determined according to the leaf collar method⁶⁸. Phenotypic data from growth chamber experiments were filtered by a Grubbs' outlier test⁶⁷.

Measurement of photosynthetic traits

F_v/F_m was measured in growth chamber grown plants using a LI-600 (LI-COR Inc., Lincoln, NE, USA). Measurements were taken pre-dawn on the last fully developed leaf of plants that had been dark-adapted

overnight. In field experiments, the last fully developed leaf was cut from the plant and dark-adapted for at least 30 min before measuring. F_v/F_m was assessed at growth stage V6 with a LI-6800 Portable Photosynthesis System (LI-6800; LI-COR Inc., Lincoln, NE, USA) as described in Urzinger et al.¹⁸. The last fully developed leaf was clipped at the midpoint, avoiding the midvein. Chamber flow rate was set to high ($150 \mu\text{mol s}^{-1}$) with a match frequency of 10. Maximum fluorescence (F_m) of dark-adapted leaves was obtained using a rectangular saturating flash ($6000 \mu\text{mol m}^{-2} \text{s}^{-1}$, 800 ms) with fluorescence acquisition at a modulation rate of 5 Hz. Fluorescence constants used for electron transport calculations were set to a leaf light absorptance of 0.8 and a fraction of absorbed light partitioned to PSII of 0.5. Non-photochemical quenching (NPQ) was measured in growth chamber grown plants using a LI-6800. Measurements were taken pre-dawn on the last fully developed leaf of plants that had been dark-adapted overnight. Dark-adapted maximum fluorescence (F_m) was determined using a rectangular saturating flash ($6000 \mu\text{mol m}^{-2} \text{s}^{-1}$, 800 ms) with an output rate of 100 Hz, margin of 5 points and dark mode rate of 50 Hz. Light-adapted maximum fluorescence F_m' was determined after setting the actinic light to $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ by conducting 16 saturating flashes every 30 s from 30 s up to 480 s after illumination and after subsequently switching off the actinic light by another set of 20 saturating flashes every 30 s from 30 s to 600 s after dark. NPQ was calculated for each measurement timepoint separately as $(F_m - F_m') / F_m'$. For estimation of PSII antenna size, fast chlorophyll a fluorescence induction kinetics (OJIP transients) were measured by a flash with intensity of $20,000 \mu\text{mol m}^{-2} \text{s}^{-1}$ and a duration of 600 ms with a LI-6800. The Dark mode rate was set to 500 Hz, the Light mode rate to 1 kHz and the Flash mode rate to 250 kHz. The half-rise time $t_{1/2}$ from the minimal fluorescence F_0 to the fluorescence at time point J (F_J) was used as an estimate of the effective antenna size of PSII³³.

Plant height measurement in the field

Early plant height was measured by stretching all leaves of a plant to measure the maximum length between soil and the tip of the leaves in growth stages V4 and V6. The measurements of three individual plants per plot were averaged to obtain the plot level measurement.

Identification of trait associations

We conducted a genome-wide association study (GWAS) for F_v/F_m in growth stages V4 and V6 in the Kemater population with the software GEMMA (v 0.98.1)⁶⁹. The same population and phenotypic data as reported in Urzinger et al.¹⁸ were used. Genotypic data for the 222 Kemater DH lines was obtained from Hölker et al.⁷⁰ and filtered for polymorphic SNPs. Phenotypic data for F_v/F_m estimates were obtained from Hölker et al.¹¹. Briefly, F_v/F_m was assessed in a set of Kemater DH lines (V4: $n=222$; V6: $n=211$) using a fluorometer (OS-30p, Opti-Sciences Inc., USA) at growth stages V4 in both 2017 and 2018, and V6 in 2017. The following statistical model was used:

$$y = 1\alpha + x\beta + Zu + e$$

where y is the n -dimensional vector of adjusted entry means for F_v/F_m averaged across four (V4) and two (V6) environments respectively, with n being the number of DH lines; α is the intercept; β is the fixed effect of the tested SNP; x is the vector of corresponding genotype scores coded as 0 or 2; u is the n -dimensional vector of random genotypic effects, with $u \sim N(0, K\sigma_g^2)$; and e is the n -dimensional vector of random residual effects, with $e \sim N(0, I_n\sigma^2)$. K denotes the $(n \times n)$ genomic kinship matrix based on SNP markers, calculated with GEMMA ("gk-2"). I_n denotes the $(n \times n)$ identity matrix. σ_g^2 and σ^2 refer to the genetic and residual variance pertaining to the model, respectively. The matrix Z ($n \times n$) assigns adjusted means to the random genotypic effects. Significance of trait associations was assessed with a likelihood ratio test, as implemented in GEMMA, using a 15% false discovery rate (FDR)⁷¹. Pairwise r^2 between SNPs was calculated as their squared Pearson correlation coefficient⁷². SNPs with an $r^2 \geq 0.6$ and within a physical distance < 1 Mb were considered to belong to the same QTL. Physically overlapping QTL were merged and the resulting QTL was described by the start and end position of the first and last SNP from the merged QTL. The proportion of genetic variance explained by a single QTL was estimated by calculating the reduction in genetic variance between a full model including all QTL and a reduced model excluding the respective QTL, following Mayer et al.¹².

Transcript level measurements

RNA was extracted from leaves using a guanidine hydrochloride protocol⁷³, followed by DNase digestion and first-strand cDNA synthesis (Maxima H Minus Kit, random hexamer primers, Thermo Scientific K1652). RT-qPCRs were performed in technical triplicates with both primers binding to the 3' UTR of *lhcb6*, *Zm00001eb433520* and *Zm00001eb433550* (Supplementary Table S9). For normalization, RT-qPCR results of the house-keeping gene MEP served as reference (membrane protein PB1A10.07c, *Zm00001eb257640*).

Leaf proteome analysis

For proteome measurements the parental lines KE0095 and KE0109 were grown in a growth chamber as described above. Leaf material was sampled in frozen nitrogen for three biological replicates per genotype in growth stage V4. From the subset of 27 DH lines grown in field experiments, a pooled sample of six plants from two replicates of one genotype was taken from the last fully developed leaf in growth stage V6 and frozen in liquid nitrogen. Total proteome was extracted and measured following an established protocol⁷⁴. For quantification of peptides an additional labeling step was added after protein digestion. Samples of KE0095 and KE0109 were labeled using 16-plex tandem mass tag (TMT) reagent, while samples from the subset of 27 DH lines were labeled using 11-plex TMT reagent and processed in three batches. A detailed protocol for sample preparation and mass spectrometry was described by Urzinger et al.¹⁸. For normalization of batch effects in the subset of 27

Kemater lines, a reference was prepared by mixing equal peptide amounts of all samples which was measured twice in each of the three TMT batches. Proteins were identified using the proteome of B73v5 as reference⁷⁵ and quantified by MaxQuant⁷⁶. To avoid imputation of missing values in quantitative analyses, only proteins which were identified in all samples of an experiment were considered. Log₂ transformed raw intensities of identified proteins were median normalized to remove differences in summed intensities per sample. For the subset of 27 Kemater lines the two reference channels per TMT batch were used to calculate a correction factor for each identified protein per TMT batch.

Detection of the *lhcb6-B* allele

To determine presence or absence of a 3.3 kb insertion in the genomic sequence upstream of the *lhcb6* start codon (Fig. 3D) a PCR assay was developed. Primer sequences are listed in Supplementary Table S9. PCR reactions were performed using Q5 High-Fidelity DNA Polymerase (New England Biolabs Inc., Ipswich, MA, USA) according to the manufacturer's specifications at an annealing temperature of 65 °C for 15 s followed by elongation at 72 °C for 30 s. Amplicon lengths of genomic sequences were discriminated by gel electrophoresis of PCR products. Sanger sequencing of PCR amplicons was performed using Mix2seq kits at Eurofins Genomics Germany GmbH (Ebersberg, Germany).

Whole-genome sequencing

DH lines KE0095 and KE0109, from which recombinant lines and NILs were derived, were sequenced using PacBio HiFi long reads by CNRGV, INRA Occitanie Toulouse, France (cnrgv.toulouse.inra.fr). Circular consensus sequences were *de-novo* assembled in contigs using the hifiasm assembler with default parameters⁷⁷. Subsequently, the contigs were ordered in ALLMAPS⁷⁸ using a genetic map derived from the cross of inbred lines EP1xPH207 as reference⁷⁹.

Comparative genomic analyses

Genomic positions of KASP markers in the QTL region on chromosome 10 of KE0095 and KE0109 were obtained by mapping their probe sequences against KE0095 and KE0109 PacBio HiFi genomes using bwa-mem⁸⁰. Genomic positions of B73v5 genes⁷⁵ in KE0095 and KE0109 were identified using BLAST + 2.9.0⁸¹.

Pairwise sequence alignment of chromosome 10 between KE0095 and KE0109 was conducted using nucmer with a minimum length of a cluster set to 156 and a maximum gap between adjacent matches in a cluster set to 1000, and subsequently filtered using delta-filter with settings of minimum identity of 95% and minimum alignment length of 1000⁸².

To compare the amino acid sequences of candidate genes in the QTL region on chromosome 10 between KE0095 and KE0109, coding sequences of the candidate genes were extracted from the respective genome assembly. Coding sequences were translated to amino acids and aligned using Clustal Omega⁸³.

LHCII components of *Arabidopsis thaliana* were identified in Araport 11 (Supplementary Table S7)⁵⁴ and searched against B73v5 proteins⁷⁵ via BLAST + 2.9.0⁸¹. The search result was filtered for minimum sequence identity of 70% and minimum query coverage of 75%. For each query, the search result with the highest bit score was selected and reported (Supplementary Table S7).

Haplotype distribution in elite lines

Data for 136 diverse Dent and Flint inbred lines genotyped with the 600k Array were obtained from Unterseer et al.³¹ and merged with the 600k SNP data from the Kemater DH lines, retaining only SNPs present in both datasets. The 15 SNPs closest to *lhcb6* in B73v5 (8 upstream; 7 downstream), including the most significant SNP from the GWAS analysis (AX-90599221; lead SNP), were concatenated to define haplotype alleles. These haplotype alleles were then filtered to keep only those occurring at least three times in the combined dataset.

Statistical analyses

All statistical analyses were performed in R (R Core Team, 2013). The statistical model for estimating genotype and genotype by environment variance components of F_{2,3} recombinant lines derived from the cross of the DH lines KE0095 and KE0109 in field experiments 2024 was:

$$y_{tkos} = \mu + \eta_t + g_{k(t)} + l_o + g_{l_{ko}(t)} + r_{s(o)} + e_{tkos}$$

where y_{tkos} are the phenotypic observations; μ is the overall mean; η_t is the fixed effect of the group t ($t = 1$ for checks, $t = 2$ for F_{2,3} recombinant lines); $g_{k(t)}$ is the random effect of genotype k nested within group t ; l_o is the random effect of environment o ; $g_{l_{ko}(t)}$ is the random interaction effect for genotype k and environment o nested within group t ; $r_{s(o)}$ is the random effect of the block s nested within environment o and e_{tkos} is the residual error. Genotypic and genotype by environment variance components were modeled individually for recombinant lines and for checks. Restricted maximum-likelihood estimation implemented in 'ASRemlR' (v4.1.0) was used for estimating variance components and their standard errors. Trait heritability (h^2) and associated confidence intervals were calculated on an entry mean basis applying established procedures^{84,85}.

Adjusted entry means for the F_{2,3} recombinant lines, the subset of 27 Kemater lines, and respective checks were calculated using the same model but dropping the group effect (η_t) and treating genotype as a fixed effect. Adjusted means for KE0095, KE0109 and four NILs in the growth chamber experiments were calculated using a statistical model with genotype as fixed effect and block as random effect.

To evaluate the association between marker alleles and phenotypic data in the F_{2,3} recombinant lines, significance of marker effects was assessed for 14 markers (Supplementary Table S1) based on adjusted means with incremental Wald tests implemented in 'ASRemlR' (v4.1.0)⁸⁶. Marker alleles were coded as 0 and 2 and

marker effects were considered significant at a FDR of 1%. Phenotypic variance explained (PVE) by the markers was calculated as the coefficient of determination (R^2) of the regression of F_v/F_m adjusted means on marker alleles. Genetic variance explained (GVE) was calculated as PVE divided by the heritability estimate of F_v/F_m ⁸⁷. To assess the effect of the *lhcb6-B* allele on F_v/F_m and early plant height in the subset of 27 Kemater lines, a two-way analysis of variance (ANOVA) was conducted with the genes *lhcb6* and *ndhm1*¹⁸, and their interaction as factors. Significance for the traits F_v , F_m , F_v/F_m , NPQ, fresh and dry biomass and half rise time from F_0 to F_j in comparisons of KE0095, KE0109 and four NILs in the growth chamber experiments was tested with Student's *t*-tests with Benjamini-Hochberg correction to control for multiple testing.

To assess the significance of differential protein accumulation in the 27 Kemater lines, an ANOVA was conducted for each protein separately. To avoid false positive associations, proteins were considered differentially accumulated at a FDR of 20%.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE⁸⁸ partner repository with the dataset identifiers PXD069740 and PXD070951.

Received: 19 November 2025; Accepted: 25 February 2026

Published online: 06 March 2026

References

- McCouch, S. et al. Feeding the future. *Nature* **499**, 23–24 (2013).
- Swarup, S. et al. Genetic diversity is indispensable for plant breeding to improve crops. *Crop Sci.* **61**, 839–852 (2021).
- Mir, C. et al. Out of America: tracing the genetic footprints of the global diffusion of maize. *Theor. Appl. Genet.* **126**, 2671–2682 (2013).
- Gouesnard, B. et al. Genotyping-by-sequencing highlights original diversity patterns within a European collection of 1191 maize flint lines, as compared to the maize USDA genebank. *Theor. Appl. Genet.* **130**, 2165–2189 (2017).
- Reif, J. C. et al. Trends in genetic diversity among European maize cultivars and their parental components during the past 50 years. *Theor. Appl. Genet.* **111**, 838–845 (2005).
- Hartfield, M. & Otto, S. P. Recombination and hitchhiking of deleterious alleles. *Evolution* **65**, 2421–2434 (2011).
- Lobell, D. B. & Tebaldi, C. Getting caught with our plants down: the risks of a global crop yield slowdown from climate trends in the next two decades. *Environ. Res. Lett.* **9**, 074003 (2014).
- Sharwood, R. E. et al. Mining for allelic gold: finding genetic variation in photosynthetic traits in crops and wild relatives. *J. Exp. Bot.* **73**, 3085–3108 (2022).
- Theeuwens, T. P., Logie, L. L., Harbinson, J. & Aarts, M. G. Genetics as a key to improving crop photosynthesis. *J. Exp. Bot.* **73**, 3122–3137 (2022).
- Khaipho-Burch, M. et al. Genetic modification can improve crop yields—but stop overselling it. *Nature* **621**, 470–473 (2023).
- Hölker, A. C. et al. European maize landraces made accessible for plant breeding and genome-based studies. *Theor. Appl. Genet.* **132**, 3333–3345 (2019).
- Mayer, M. et al. Discovery of beneficial haplotypes for complex traits in maize landraces. *Nat. Commun.* **11**, 1–10 (2020).
- Liang, Y., Liu, H. J., Yan, J. & Tian, F. Natural variation in crops: realized understanding, continuing promise. *Annu. Rev. Plant Biol.* **72**, 357–385 (2021).
- Hallauer, A. R. & Sears, J. Integrating exotic germplasm into corn belt maize breeding programs 1. *Crop Sci.* **12**, 203–206 (1972).
- Strigens, A., Schipprack, W., Reif, J. C. & Melchinger, A. E. Unlocking the genetic diversity of maize landraces with doubled haploids opens new avenues for breeding. *PLoS One.* **8**, e57234 (2013).
- Sood, S., Flint-Garcia, S., Willcox, M. C. & Holland, J. B. Mining natural variation for maize improvement: selection on phenotypes and genes. *Genomics of plant genetic resources: Managing, sequencing and mining genetic resources.* **1**, 615–649 (2014).
- Böhm, J., Schipprack, W., Utz, H. F. & Melchinger, A. E. Tapping the genetic diversity of landraces in allogamous crops with doubled haploid lines: a case study from European flint maize. *Theor. Appl. Genet.* **130**, 861–873 (2017).
- Urzinger, S. et al. Embracing native diversity to enhance the maximum quantum efficiency of photosystem II in maize. *Plant Physiol.* **197**, kiae670 (2025).
- Murchie, E. H. & Lawson, T. Chlorophyll fluorescence analysis: a guide to good practice and understanding some new applications. *J. Exp. Bot.* **64**, 3983–3998 (2013).
- Ferguson, J. N. et al. A deficient CP24 allele defines variation for dynamic non-photochemical quenching and photosystem II efficiency in maize. *Plant. Cell.* <https://doi.org/10.1093/plcell/koaf063> (2025).
- Lainé, C. M., Abdelgawad, H. & Beemster, G. T. A meta-analysis reveals differential sensitivity of cold stress responses in the maize leaf. *Plant. Cell. Environ.* **46**, 2432–2449 (2023).
- Burnett, A. C. & Kromdijk, J. Can we improve the chilling tolerance of maize photosynthesis through breeding? *J. Exp. Bot.* **73**, 3138–3156 (2022).
- Long, S. P., Marshall-Colon, A. & Zhu, X. G. Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell* **161**, 56–66 (2015).
- Leister, D. Enhancing the light reactions of photosynthesis: Strategies, controversies, and perspectives. *Mol. Plant.* **16**, 4–22 (2023).
- Fracheboud, Y., Ribaut, J. M., Vargas, M., Messmer, R. & Stamp, P. Identification of quantitative trait loci for cold-tolerance of photosynthesis in maize (*Zea mays* L.). *J. Exp. Bot.* **53**, 1967–1977 (2002).
- Fracheboud, Y., Jompuk, C., Ribaut, J., Stamp, P. & Leipner, J. Genetic analysis of cold-tolerance of photosynthesis in maize. *Plant Mol. Biol.* **56**, 241–253 (2004).
- Hund, A. et al. QTL controlling root and shoot traits of maize seedlings under cold stress. *Theor. Appl. Genet.* **109**, 618–629 (2004).
- Jompuk, C., Fracheboud, Y., Stamp, P. & Leipner, J. Mapping of quantitative trait loci associated with chilling tolerance in maize (*Zea mays* L.) seedlings grown under field conditions. *J. Exp. Bot.* **56**, 1153–1163 (2005).
- Untersee, S. et al. A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. *BMC Genom.* **15**, 1–15 (2014).
- Schwacke, R. et al. MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. *Mol. Plant.* **12**, 879–892. <https://doi.org/10.1016/j.molp.2019.01.003> (2019). <https://doi.org/https://doi.org/>
- Untersee, S. et al. A comprehensive study of the genomic differentiation between temperate Dent and Flint maize. *Genome Biol.* **17**, 1–14 (2016).
- Mayer, M. et al. Genetic diversity of European maize landraces: Dataset on the molecular and phenotypic variation of derived doubled-haploid populations. *Data Brief.* **42**, 108164 (2022).

33. Strasser, R. J., Tsimilli-Michael, M. & Srivastava, A. in *Chlorophyll a fluorescence: a signature of photosynthesis* 321–362 (Springer, 2004).
34. White, S. W., Zheng, J., Zhang, Y. M. & Rock, C. O. The structural biology of type II fatty acid biosynthesis. *Annu. Rev. Biochem.* **74**, 791–831 (2005).
35. Kromdijk, J. et al. Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science* **354**, 857–861 (2016).
36. Ali, W. et al. Quantitative genetics of photosynthetic trait variation in maize. *J. Exp. Bot.* **76**, 4141–4153. <https://doi.org/10.1093/jxb/eraf198> (2025).
37. Murchie, E. H. & Ruban, A. V. Dynamic non-photochemical quenching in plants: from molecular mechanism to productivity. *Plant J.* **101**, 885–896 (2020).
38. Ishikawa, N. et al. NDH-Mediated Cyclic Electron Flow Around Photosystem I is Crucial for C4 Photosynthesis. *Plant Cell Physiol.* **57**, 2020–2028. <https://doi.org/10.1093/pcp/pcw127> (2016).
39. Ricci, W. A. et al. Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants.* **5**, 1237–1249. <https://doi.org/10.1038/s41477-019-0547-0> (2019).
40. Savadel, S. D. et al. The native cisrome and sequence motif families of the maize ear. *PLoS Genet.* **17**, e1009689. <https://doi.org/10.1371/journal.pgen.1009689> (2021).
41. Ballottari, M., Girardon, J., Dall’Osto, L. & Bassi, R. Evolution and functional properties of photosystem II light harvesting complexes in eukaryotes. *Biochim. et Biophys. Acta (BBA)-Bioenergetics.* **1817**, 143–157 (2012).
42. Iwai, M., Patel-Tupper, D. & Niyogi, K. K. Structural diversity in eukaryotic photosynthetic light harvesting. *Annual Rev. Plant. Biology* **75**, 119–152 (2024).
43. Su, X. et al. Structure and assembly mechanism of plant C2S2M2-type PSII-LHCII supercomplex. *Science* **357**, 815–820 (2017).
44. Kovacs, L. et al. Lack of the light-harvesting complex CP24 affects the structure and function of the grana membranes of higher plant chloroplasts. *Plant. Cell.* **18**, 3106–3120 (2006).
45. Ilkóvá, I. et al. Towards spruce-type photosystem II: consequences of the loss of light-harvesting proteins LHCB3 and LHCB6 in Arabidopsis. *Plant Physiol.* **187**, 2691–2715 (2021).
46. Sahay, S., Grzybowski, M., Schnable, J. C. & Glowacka, K. Genetic control of photoprotection and photosystem II operating efficiency in plants. *New Phytol.* **239**, 1068–1082. <https://doi.org/10.1111/nph.18980> (2023).
47. Sahay, S. et al. Nonphotochemical quenching kinetics GWAS in sorghum identifies genes that may play conserved roles in maize and Arabidopsis thaliana photoprotection. *Plant. J.* **119**, 3000–3014. <https://doi.org/10.1111/tpj.16967> (2024).
48. Caffarri, S., Frigerio, S., Olivieri, E., Righetti, P. G. & Bassi, R. Differential accumulation of Lhcb gene products in thylakoid membranes of Zea mays plants grown under contrasting light and temperature conditions. *Proteomics* **5**, 758–768 (2005).
49. Kargul, J. & Barber, J. Photosynthetic acclimation: Structural reorganisation of light harvesting antenna—role of redox-dependent phosphorylation of major and minor chlorophyll a/b binding proteins. *FEBS J.* **275**, 1056–1068 (2008).
50. Pan, X. et al. Structure of the maize photosystem I supercomplex with light-harvesting complexes I and II. *Science* **360**, 1109–1113 (2018).
51. Kromdijk, J. & Walter, J. *Relaxing non-photochemical quenching (NPQ) to improve photosynthesis in crops* (Burleigh Dodds Science Publishing, 2023).
52. Ganeteg, U., Kulheim, C., Andersson, J. & Jansson, S. Is each light-harvesting complex protein important for plant fitness? *Plant Physiol.* **134**, 502–509 (2004).
53. De Bianchi, S., Dall’Osto, L., Tognon, G., Morosinotto, T. & Bassi, R. Minor antenna proteins CP24 and CP26 affect the interactions between photosystem II subunits and the electron transport rate in grana membranes of Arabidopsis. *Plant. Cell.* **20**, 1012–1028 (2008).
54. Cheng, C. Y. et al. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J.* **89**, 789–804 (2017).
55. Pan, X. et al. Structural basis of LhcbM5-mediated state transitions in green algae. *Nat. Plants.* **7**, 1119–1131 (2021).
56. Dunahay, T. G., Schuster, G. & Staehelin, L. A. Phosphorylation of spinach chlorophyll-protein complexes CPII*, but not CP29, CP27, or CP24, is phosphorylated in vitro. *FEBS Lett.* **215**, 25–30 (1987).
57. Tanaka, R. & Tanaka, A. Chlorophyll cycle regulates the construction and destruction of the light-harvesting complexes. *Biochim. et Biophys. Acta (BBA)-Bioenergetics.* **1807**, 968–976 (2011).
58. Havaux, M., Dall’Osto, L. & Bassi, R. Zeaxanthin has enhanced antioxidant capacity with respect to all other xanthophylls in Arabidopsis leaves and functions independent of binding to PSII antennae. *Plant Physiol.* **145**, 1506–1520 (2007).
59. Bossmann, B., Knoetzel, J. & Jansson, S. Screening of chlorina mutants of barley (*Hordeum vulgare* L.) with antibodies against light-harvesting proteins of PS I and PS II: absence of specific antenna proteins. *Photosynth. Res.* **52**, 127–136 (1997).
60. Tanaka, R. et al. Overexpression of chlorophyllide a oxygenase (CAO) enlarges the antenna size of photosystem II in Arabidopsis thaliana. *Plant J.* **26**, 365–373 (2001).
61. Diepenbrock, C. H. et al. Novel loci underlie natural variation in vitamin E levels in maize grain. *Plant. Cell.* **29**, 2374–2392 (2017).
62. Havaux, M., Eymery, F., Porfirova, S., Rey, P. & Dörmann, P. Vitamin E protects against photoinhibition and photooxidative stress in Arabidopsis thaliana. *Plant. Cell.* **17**, 3451–3469 (2005).
63. Chen, Y. E. et al. The roles of Arabidopsis proteins of Lhcb4, Lhcb5 and Lhcb6 in oxidative stress under natural light conditions. *Plant Physiol. Biochem.* **130**, 267–276 (2018).
64. Sheng, X., Liu, X., Cao, P., Li, M. & Liu, Z. Structural roles of lipid molecules in the assembly of plant PSII-LHCII supercomplex. *Biophys. Rep.* **4**, 189–203 (2018).
65. Mayer, M. et al. Is there an optimum level of diversity in utilization of genetic resources? *Theor. Appl. Genet.* **130**, 2283–2295 (2017).
66. Röber, F., Gordillo, G. & Geiger, H. In vivo haploid induction in maize—performance of new inducers and significance of doubled haploid lines in hybrid breeding. *Maydica* **50**, 275–283 (2005).
67. Grubbs, F.E. Procedures for Detecting Outlying Observations in Samples. *Technometrics* **11**, 1–21. <https://doi.org/10.1080/00401706.1969.10490657> (1969).
68. Ritchie, S. W., Hanway, J. J., Benson, G. O. & Herman, J. C. *How a corn plant develops* (Iowa State University of Science and Technology, Cooperative Extension Service, 2005).
69. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
70. Hölker, A. C. et al. Theoretical and experimental assessment of genome-based prediction in landraces of allogamous crops. *Proceedings of the National Academy of Sciences.* **119**, e2121797119 (2022).
71. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* **57**, 289–300 (1995).
72. Hill, W. G. & Robertson, A. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**, 226–231. <https://doi.org/10.1007/BF01245622> (1968).
73. Logemann, J., Schell, J. & Willmitzer, L. Improved method for the isolation of RNA from plant tissues. *Anal. Biochem.* **163**, 16–20 (1987).
74. Brajkovic, S. et al. Getting ready for large-scale proteomics in crop plants. *Nutrients* **15**, 783 (2023).
75. Hufford, M. B. et al. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655–662 (2021).

76. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).
77. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods.* **18**, 170–175. <https://doi.org/10.1038/s41592-020-01056-5> (2021).
78. Tang, H. et al. ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biol.* **16**, 1–15 (2015).
79. Haberer, G. et al. European maize genomes highlight intraspecies variation in repeat and gene content. *Nat. Genet.* **52**, 950–957 (2020).
80. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
81. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421. <https://doi.org/10.1186/1471-2105-10-421> (2009).
82. Kurtz, S. et al. Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12. <https://doi.org/10.1186/gb-2004-5-2-r12> (2004).
83. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145. <https://doi.org/10.1002/pro.3290> (2018).
84. Knapp, S. J., Stroup, W. W. & Ross, W. M. Exact Confidence Intervals for Heritability on a Progeny Mean Basis. *Crop Sci.* **25**, crops19850011183X002500010046x. <https://doi.org/10.2135/cropsci1985.0011183X002500010046x> (1985).
85. Hallauer, A., Carena, M. & Filho, J. B. Means and Variances in Quantitative Genetics in Maize Breeding (eds Carena, M., Hallauer, R. & Filho, J.B.) 33–67 (Springer New York, 2010).
86. Kenward, M. G. & Roger, J. H. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**, 983–997 (1997).
87. Schön, C. C. et al. Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* **167**, 485–498. <https://doi.org/10.1534/genetics.167.1.485> (2004).
88. Perez-Riverol, Y. et al. The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res.* **53**, D543–D553. <https://doi.org/10.1093/nar/gkae1011> (2024).

Acknowledgements

We thank Brigitte Neuhauser, Iris Prücklmaier, Sylwia Schepella, Stefan Schwertfirm and Margot Siebler for technical assistance. We thank the Plant Technology Center (Technical University of Munich, Germany) for providing infrastructure and technical support during greenhouse, growth chamber and field experiments. We would like to acknowledge the support of William Marande and Caroline Callot from CNRGV, INRAE (<http://cnrgev.toulouse.inrae.fr/>) for providing assistance in PacBio long read sequencing and GENTYANE platform of Clermont-Ferrand INRAE Center (<http://gentyane.clermont.inra.fr/>) for providing access to PacBio sequencer.

Author contributions

C.-C.S., V.A., M.O., P.W., L.W. and S.U. designed the research and developed ideas; L.W., S.U., V.A. and B.O. designed and performed phenotyping experiments and analyzed the data; S.R., L.W. and S.U. analyzed whole genome sequencing data; S.B. and B.K. acquired and analyzed proteomics data. L.W. and S.U. conducted candidate gene analysis; L.W., S.U., M.M., M.O., T.P., D.S. and C.U. developed the plant material; L.W., S.U., V.A., P.W. and C.-C.S. wrote the manuscript; all authors read and approved the final manuscript; C.-C.S. agrees to serve as the author responsible for contact and to ensure communication.

Funding

Open Access funding enabled and organized by Projekt DEAL. This study was funded by the Federal Ministry of Education and Research (BMBF, Germany) within the scope of the funding initiative “Plant Breeding Research for the Bioeconomy” (Funding ID: 031B0195, 031B0882 and 031B1301) as part of the project MAZE (www.europeanmaize.net). A Ph.D. fellowship for L.W. was funded by the Elitenetzwerk Bayern in the scope of the project “The Proteomes That Feed the World”. KWS SAAT SE & Co. KGaA funded Ph.D. fellowships for S.U. and M.M.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-42348-8>.

Correspondence and requests for materials should be addressed to C.-C.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026