ATP-dependent Clp protease subunit C1, *HvClpC1*, is a strong candidate gene for barley variegation mutant *luteostrians* as revealed by genetic mapping and genomic re-sequencing

Mingjiu Li[a,1], Ganggang Guo[b], Hélène Pidon[a], Michael Melzer[a], Alberto R. Prina[c], Thomas Börner[d] and Nils Stein[a,e,1]

[a]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), 06466 Seeland, Germany

[b]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, 100081 Beijing, China

[c]Instituto de Genética "Ewald A. Favret", CICVyA (Centro de Investigación en Ciencias Veterinarias y Agronómicas), Instituto Nacional de Tecnología Agropecuaria, Nicolás Repetto y de los Reseros s/n (1686) Hurlingham, Buenos Aires, Argentina

[d]Molecular Genetics, Institute of Biology, Humboldt University, 10115 Berlin, Germany

[e]Center for Integrated Breeding Research (CiBreed), Department of Crop Sciences, Georg-August-University, Göttingen, Germany

Short title: *ClpC1* is a candidate gene for barley *luteostrians*

Key words: Chloroplast development, leaf variegation, genetic mapping, genomic re-sequencing, comparative analysis, *luteostrians*, *albostrians*, *HvClpC1*, *Hordeum vulgare*

[1]Corresponding authors:

Mingjiu Li: li@ipk-gatersleben.de; Nils Stein: stein@ipk-gatersleben.de

Abstract

Implementation of next-generation sequencing in forward genetic screens greatly accelerated gene discovery in species with larger genomes, including many crop plants. In barley, extensive mutant collections are available, however, the causative mutations for many of the genes remains largely unknown. Here we demonstrate how a combination of low-resolution genetic mapping, whole-genome resequencing and comparative functional analyses provides a promising path towards candidate identification of genes involved in plastid biology and / or photosynthesis, even if genes are located in recombination poor regions of the genome. As a proof of concept, we simulated the prediction of a candidate gene for the recently cloned variegation mutant *albostrians* (*HvAST* / *HvCMF7*) and adopted the approach for suggesting *HvClpC1* as candidate gene for the yellow-green variegation mutant *luteostrians*.

Author Summary

Forward genetics is an approach of identifying a causal gene for a mutant phenotype and has proven to be a powerful tool for dissecting the genetic control of biological processes in many species. A large number of barley mutants was generated in the 1940s to 1970s when mutation breeding programs flourished. Genetic dissection of the causative mutations responsible for the phenotype, however, lagged far behind, limited by lack of molecular markers and high-throughput genotyping platforms. Next-generation sequencing technologies have revolutionized genomics, facilitating the process of identifying mutations underlying a phenotype of interest. Multiple mapping-by-sequencing or cloning-by-sequencing strategies were established towards fast gene discovery. In this study, we used mapping-by-sequencing to identify candidate genes within coarsely delimited genetic intervals, for two variegation mutants in barley – *luteostrians* and *albostrians*. After testing the approach using the example of the previously cloned *albostrians* gene *HvAST*, the gene *HvClpC1* could be delimited as candidate gene for *luteostrians.* The mapping-by-sequencing strategy implemented here is generally suited for surveying barley mutant collections for phenotypes affecting fundamental processes of plant morphology, physiology and development.

Introduction

Barley mutagenesis was intensely studied in the mid twentieth century. These activities resulted in extensive mutant collections available through genebanks such as NordGen (https://www.nordgen.org/) and IPK (https://gbis.ipk-gatersleben.de/gbis2i/faces/index.jsf). Barley mutants served as valuable resource for dissecting the genetic basis of a wide range of complex biological processes. Their broad use, however, was impeded until recently by a lack of genomic resources and tools. This has been changed by the fast development of next-generation sequencing (NGS) based technologies and strategies for gene cloning. Due to diminishing sequencing costs, NGS can be applied for sequencing-based genotyping of whole mapping populations as well as for whole-genome resequencing to accelerate gene discovery even in large genome crop species (Candela et al., 2015; Jaganathan et al., 2020). Mapping-by-sequencing strategies initially were not always ready applicable to all species as they could be limited. For instance, the Mutmap (Abe et al., 2012) and Mutmap+ (Fekih et al., 2013) approaches, initially applied to rice, require access to a high-quality reference genome for the genotype used for mutagenesis; the homozygosity mapping approach requires availability of genotypes with a known pedigree (Singh et al., 2013). Special attention needs to be paid to experimental design and technical decisions in order to enforce that sequencing data will allow to map a mutation of interest (Wilson-Sanchez et al., 2019). In barley, *MANY-NODED DWARF* (*MND*) was cloned by a technique similar to SHOREmap (Schneeberger et al., 2009); *LAXATUM-a* (*LAX-a*) was isolated by exome-capture sequencing (Mascher et al., 2013) of several highly informative recombinant pools (Jost et al., 2016). Notably, identification of *MND* and *LAX-a* was even achieved while relying on a largely unordered draft genome of barley (International Barley Genome Sequencing Consortium, 2012). Recently, the release and improvement of a high-quality reference genome of barley (Mascher et al., 2017; Monat et al., 2019) has greatly facilitated forward genetic screens relying largely on mapping- / cloning-by-sequencing strategies (Candela et al., 2015) and thus paved the way towards systematic dissection of the genic factors underlying barley mutant resources.

Many mutants of genebank collections are affected by photosynthesis-related defects leading to aberrant coloration phenotypes. Photosynthesis-related mutants provide a highly valuable genetic tool for identification of nuclear genes involved in different aspects of chloroplast development. Among the distinct chlorophyll-deficient phenotypes, leaf variegation is a common phenomenon that has been observed for many plants in nature (Toshoji et al., 2012). Investigations with variegated mutants (patched in dicots / striping in monocots) generated insights into the molecular mechanisms of leaf variegation (Yu et al., 2007). A threshold-dependent genetic model was proposed as a mechanism underlying *var2* variegation. *Var2* encodes a chloroplast-localized protein AtFtsH2 which belongs to the *filamentation temperature sensitive* (*FtsH*) *metalloprotease* gene family (Chen et al., 2000; Takechi et al., 2000). In the model, two pairs of FtsH proteins, AtFtsH1/5 and AtFtsH2/8, form oligomeric complexes in the thylakoid membrane and a threshold level of oligomeric complexes is required for normal chloroplast function and green sector formation (Yu et al., 2004). In monocot species, *iojap* of maize (*Zea mays*) (Walbot and Coe, 1979) and *albostrians* of barley (Hagemann and Scholz, 1962; Hess et al., 1993) represent two examples of classical variegation mutation revealing that cells of albino sectors contain ribosome-free plastids. *iojap* encodes a component associated with the plastid ribosomal 50S subunit (Han et al., 1992). The *albostrians* gene *HvAST* / *HvCMF7* encodes a plastid-localized CCT MOTIF FAMILY (CMF) protein (Li et al., 2019c). Although a threshold-dependent mechanism is also very likely, the molecular mechanisms underlying the *iojap* and *albostrians* leaf variegation still remain

elusive. *luteostrians* is another barley mutant with a block in early chloroplast development, showing a yellow-green striped variegation phenotype that appears only if the mutant allele is inherited through the female gamete.

Here we report *HvClpC1* as a candidate gene for the variegation mutant *luteostrians* by applying a sequencing-based gene identification strategy that has great potential of systematical application to reveal causative mutations for photosynthesis-related mutants in barley collections.

Results

Inheritance of Leaf Variegation in the *luteostrians* Mutant

Phenotypically, progeny of *luteostrians* can be classified into three categories: 1) green-yellow striped, 2) completely yellow, and 3) albino (Figure 1A). The green-yellow striped seedlings were able to complete the life cycle while the yellow and albino plants do not survive beyond the three-leaf vegetative growth stage. Considering the variegated pattern to be analogous to the *albostrians* mutant of barley (Li et al., 2019c), we named the causal gene underlying the *luteostrians* striped phenotype as *HvLST*, representing *Hordeum vulgare LUTEOSTRIANS*. Selfing of heterozygous plants leads to about a quarter of aborted grains, indicating that homozygosity for the gene (*lst/lst*) is lethal at post-zygote or early embryonic stage (Table 1). Inheritance of the yellow striped phenotype was gamete-dependent (Figure 1B) since variegation is observed if the *lst* allele is inherited through the female gamete (i.e., *lst/LST*). This interpretation is supported by three lines of evidence: 1) when wild type (*LST/LST*) plants are fertilized with pollen from striped (*lst/LST*) or green heterozygous plants (*LST/lst*), in either case, only half of the F1 plants offspring showed phenotypic segregation in F2 generation; 2) a segregation ratio of 2:1 (green:striped) was observed in segregant offspring in F2 generation; and 3) heterozygous F2 plants (*LST/lst* or *lst/LST*) segregated with green (*LST/LST* or *LST/lst*) and striped (*lst/LST*) progenies with a ratio of 2:1 in subsequent F3 generation (Figure 1B). Thus, in addition to its essential function during embryogenesis, we postulate that *HvLST* plays an essential role in plastid differentiation / programming during gametic stage in the egg cell.

Chloroplast Translation is Abolished in the *luteostrians* Mutant

Defective chloroplasts do not contain 70S ribosomes in the *albostrians* mutant (Hess et al., 1993; Li et al., 2019c). In an attempt to check the function of the translation machinery in plastids of the *luteostrians* mutant, we initially examined accumulation of the rRNAs in wild type and *luteostrians* mutant. In analogy to the *albostrians* mutant, the 16S and 23S rRNA species are not observed in defective plastids of the *luteostrians* mutant (Figure 2), indicating the lack of 70S ribosomes and consequently missing chloroplast translation in plastids of the yellow leaf sectors. Next, we studied chloroplast/plastid ultrastructure by transmission electron microscopy. These analyses revealed normal chloroplast development in wild type and in green leaf sections of the *luteostrians* mutant; in both cases chloroplasts contained well-developed stroma and grana thylakoids (Figure 3A & 3B). In contrast, plastids in yellow sectors of the variegated *luteostrians* leaves contained no grana and only rudimentary stroma lamellae (Figure 3C). Notably, chloroplast 70S ribosomes were not detectable in yellow plastids of the *luteostrians* mutant (Figure 3). Altogether, based on the absence of chloroplast rRNA species and the lack of 70S ribosomes it can be postulated that chloroplast translation is abolished in the *luteostrians* mutant.

Genetic Mapping of *HvLST* to the Genetic Centromere of Chromosome 2H

Two F2 mapping populations, designated as 'BL' and 'ML', were constructed for the purpose of genetic mapping of the *HvLST* gene. The genotypic status of the *luteostrians* locus of F2 plants was determined by phenotypic segregation analysis of their respective F3 progenies. As homozygosity (*lst/lst*) leads to grain abortion, the BL population contained 95 wild type (*LST/LST*) and 172 heterozygotes (*LST/lst* and *lst/LST*), consistent with the segregation of a single recessive gene (Table 2). Genetic mapping in the ML population was affected by segregation distortion with genotype ratios deviating from the expected Mendelian ratios (Table 2). Genotyping-by-sequencing was performed for 267 and 269 F2 genotypes in BL and ML populations, respectively. Sequencing data was mapped to the reference genome of barley (Monat et al., 2019) for SNP calling. In total, 3,745 and 5,507 SNP markers were obtained genome-wide at a minimum sequencing coverage of six-fold for BL and ML populations, respectively. By applying a permissive threshold of 5% missing data for both molecular marker and F2 genotype, mapping could be performed in 124 F2 / 3369 SNPs and 146 F2 / 4854 SNPs for the BL and ML populations, respectively (Supplemental Table 1). Genetic maps for seven linkage groups, representing the seven barley chromosomes, were obtained for BL (LOD ≥ 6) and ML (LOD = 10) populations, comprising mapped markers for 66.4% (2238/3369) and 71.5% (3469/4584) of the originally defined SNPs, respectively (Supplemental Table 1). The accuracy of the linkage maps was consistency-checked by aligning genetic marker positions with their respective physical order in the reference genome of barley (Monat et al., 2019). The *HvLST* gene was assigned to the region of the genetic centromere region of chromosome 2H with a physical distance between the flanking markers of 461.7 and 499.9 Mbp in BL and ML populations, respectively (Supplemental Figure 1).

Next, PCR-based KASP (Kompetitive Allele-Specific PCR) markers were designed and employed to saturate the identified *HvLST* intervals in both mapping populations. Initially, genotyping all the 267 (BL population) and 269 (ML population) F2 individuals with two population-specific KASP markers (chr2H_178489464 and chr2H_461744325 for BL population; chr2H_200083049 and chr2H_463274491 for ML population) allowed to confirm and further narrow down the *HvLST* target region. Notably, the numerical value within each marker designation indicates the physical coordinate of the mapped SNP position on the reference genome of barley (Monat et al., 2019) (Figure 4). Subsequently, saturation mapping of the *HvLST* interval was performed in a total of 18 recombinants (13 in BL / 5 in ML) with additional KASP markers that were suitable for both mapping populations (Supplemental Table 2). Finally, we delimited the *HvLST* target region to a 0.74 centiMorgan (cM) interval between flanking markers chr2H_431057673 and chr2H_458001177, spanning a distance of ~26.94 Mbp (Figure 5 and Supplemental Table 2). A cluster of markers co-segregated with the *HvLST* locus, suggesting that maximum genetic resolution was achieved at the given size of both mapping populations (Figure 4).

Whole Genome Re-sequencing Identifies *HvClpC1* as *HvLST* Candidate Gene

Based on the annotated reference genome sequence of barley (Monat et al., 2019) the genetic interval of *HvLST* delimited by the closest flanking molecular markers is annotated with 284 genes (Figure 5A and 5C), a number too large for realistically spotting directly a candidate gene for *HvLST*. One option, higher resolution genetic mapping of the gene *HvLST* by screening for recombinants in a much larger mapping population was rejected due to the unfavorable genetic placement of the gene in a region of very low recombination. Instead, we evaluated the possibility to survey directly for mutations in any of the predicted genes in the *HvLST* interval by high-throughput re-sequencing.

First, we simulated the feasibility of the approach based on existing data for the variegation mutant *albostrians* (Li et al., 2019c). Previously, low-resolution mapping with 91 F2 genotypes delimited the gene *HvCMF7* to a 6.05 cM genetic interval between flanking markers Zip_2613 and 1_0169. Anchoring the flanking markers to the barley reference genome (Monat et al. 2019) revealed a 17.86 Mbp physical interval comprising 323 genes (Figure 5A-B); comparable to the situation for *HvLST*. Next, we used whole genome re-sequencing data for the mutant M4205 (original *albostrians* mutant) and barley cv. Haisa (genetic background used for induction of the *albostrians* mutant). In an initial screen we filtered M4205-specific SNPs in genic regions revealing nine candidate genes. The second step of filtering for functional SNPs (e.g., leading to non-synonymous exchange of amino acid, change of splice junction, or premature stop) retrieved four candidate genes (Figure 5A-B; Table 3 and Supplemental Table 3). Based on functional annotation of these four genes, a single gene, *HORVU.MOREX.r2.7HG0603920.1*, homolog to the Arabidopsis gene *CIA2*, was suggested as the most promising *albostrians* candidate, supported by photosynthesis-related pale green phenotype of the Arabidopsis *cia2* mutant (Sun et al., 2001). *HORVU.MOREX.r2.7HG0603920.1* in fact represents the genuine *albostrians* gene, as was verified by independent mutant analysis (Li et al., 2019c). In conclusion, based on a delimited genetic interval, survey sequencing of wild type and mutant genetic background may provide a direct path for candidate gene identification for photosynthesis related phenotype mutants in barley.

We adopted this approach to identify a *HvLST* candidate gene. A wild type (line luteostrians-2_2, *LST*/*LST*) and a mutant genotype (line luteostrians-3_6, *lst*/*LST*) were selected from the originally segregating *luteostrians* mutant, to ensure both analyzed genotypes sharing the same genetic background of line MC20. In contrast to the previously described simulation for the gene *albostrians*, the initial screening exercise for sequence polymorphisms between mutant and wild type had to consider the fact that homozygous *lst*/*lst* is embryo lethal, hence functional polymorphisms were expected to be present at heterozygous state. Consequently, 54 out of the 284 genes, identified for the *luteostrians* mapping interval, carried heterozygous SNPs specific to the *luteostrians* mutant. For a shortlist of eleven genes, the observed SNP was predicted to induce a putative functional change (Figure 5A and 5C; Table 3 and Supplemental Table 3). BLASTp search (Mount, 2007) against the Arabidopsis proteome revealed presence of orthologs for 8 of the 11 candidate genes. None of the remaining three genes likely represented a genuine *HvLST* candidate as they either encoded for a putative retrotransposon protein in barley or showed similarity to an organelle gene lacking essential function in chloroplast biogenesis in Arabidopsis (Table 3 and Supplemental Table 3). We then inspected functional annotation information and, if available, phenotype information for mutants of the narrowed shortlist of eight Arabidopsis genes. One of these genes *AT5G50920* encodes an ATP-dependent Clp protease ATP-binding subunit ClpC1. Mutants of this gene exhibited a pale-yellow phenotype with reduced photosynthetic performance (Sjogren et al., 2004). Notably, inactivation of the *ClpP* genes (e.g., *ClpP4*, *ClpP5*, *ClpP6*) in Arabidopsis resulted in embryo lethality and antisense repression lines exhibited a variegated 'yellow-heart' phenotype (Clarke et al., 2005). In conclusion, *HORVU.MOREX.r2.2HG0135340.1* (*HvClpC1*), homolog (putative ortholog) of Arabidopsis *ClpC1*, represented the most likely candidate for the gene *luteostrians*. The predicted heterozygous SNP in *HvClpC1* was confirmed by Sanger sequencing (Supplemental Table 4). The *luteostrians* mutant carries a heterozygous SNP at position 2,078 (i.e., G2078A; coordinate refers to the coding sequence), consequently, changing glycine into aspartic acid at position 693 (i.e., G693D) (Figure 6B).

The *HvLST* candidate *HvClpC1* has two close homologs in barley, *HORVU.MOREX.r2.5HG0373350.1* (homolog 1) and *HORVU.MOREX.r2.4HG0286520.1* (homolog 2), which share 85.57% and 74.26% amino acid identity with HvClpC1, respectively. All three homologs share the same gene structure with nine exons. In line with the chloroplast localization of ClpC1 in Arabidopsis (Sjogren et al., 2014), ChloroP predicts presence of a chloroplast transit peptide for all three homologs in barley (Figure 6A), potentially suggesting a chloroplast localization.

We looked up the expression profile of *HvClpC1* and its two close homologs in BARLEX database (Colmsee et al., 2015). *HvClpC1* shows ubiquitous expression in all examined tissues except young developing inflorescences (INF1) (Figure 6C). The expression pattern of homolog 1 resembles that of *HvClpC1* but in most tissues at a distinctly lower level. In contrast, expression of homolog 2 is barely detectable across all the samples.

Discussion

Induced mutants are important tools for elucidating basic principles in the biology of photosynthesis and organellogenesis in plants. Recently, we reported the cloning and characterization of the gene which bears a mutation in the well-known green-white striped mutant *albostrians* of barley (Li et al., 2019c). Here we introduce, based on low-resolution mapping, whole genome re-sequencing and homology search to photosynthesis or chloroplast related mutants in Arabidopsis, a highly promising candidate gene *HvClpC1* for the related yellow-green variegation mutant *luteostrians*.

Protease ATP-binding Subunit *HvClpC1* Represents the Perfect Candidate Gene for *HvLST*

Low-resolution mapping delimited *HvLST* to a large interval spanning a physical distance of 26.9 Mbp comprising a total of 284 genes. By comparative whole genome re-sequencing of wild type and mutant lines, a homolog of the Arabidopsis gene *AT5G50920* was identified as candidate *LUTEOSTRIANS* gene. *AT5G50920* belongs to the ATP-dependent protease family and encodes the ATP-binding subunit ClpC1. The *clpC1* null mutant exhibited a homogenous pale-yellow phenotype (Sjogren et al., 2004). The distinct phenotypes of *ClpC1* knockout lines in Arabidopsis and *luteostrians* barley are reminiscent of the recently reported orthologous pair of genes of the variegated *albostrians* barley mutant and its Arabidopsis homolog pale green mutant *cia2* (Sun et al., 2001; Li et al., 2019c).

The CLP protease system is a central component of the chloroplast protease network (Olinares et al., 2011). It plays essential roles in coordinating plastid proteome dynamics during developmental transitions such as embryogenesis and leaf development (Nishimura and van Wijk, 2015). Genetic studies demonstrated that members of the CLP protease family contribute to chloroplast biogenesis (Constan et al., 2004; Sjogren et al., 2004) and embryogenesis (Kovacheva et al., 2007; Kim et al., 2009). ClpC1 plays important roles on chlorophyll biosynthesis as it controls turnover of glutamyl-tRNA reductase (GluTR) and chlorophyllide *a* oxygenase (CAO); GluTR and CAO catalyse an early step of the tetrapyrrole biosynthesis pathway and the conversion of chlorophyllide *a* to *b*, respectively (Nakagawara et al., 2007; Apitz et al., 2016; Rodriguez-Concepcion et al., 2019). Although single mutants were viable, siliques of the double mutant *clpC1clpC2* (previously described as *hsp93-V / hsp93-III-2*) contained aborted seeds (because of a block in the zygote-embryo transition) and failed ovules (because of a moderate defect in female gametophytes) (Kovacheva et al., 2007). This suggested that both homologs *ClpC1* and *ClpC2* confer important functions to cell viability

during gamete stage and early zygotic stage during embryogenesis. Moreover, the proteome phenotype of *clpC1clpS1* suggested a ClpS1 and ClpC1 interaction effect on plastid gene expression components and nucleoid interactors, including RNA processing and editing, as well as 70S ribosome biogenesis (Nishimura et al., 2013). So far, biochemical and genetic evidences gathered for Arabidopsis single and double mutants related to *clpC1* were in line with the observed yellow-striped, chloroplast ribosome deficient and embryonic lethal phenotype (at homozygous state) of the *luteostrians* mutant. Mutants of related genes showed also variegation in Arabidopsis, e.g., Arabidopsis mutant *yellow variegated 2 (var2*) exhibited a patched phenotype (Chen et al., 2000). *VAR2* encodes an ATP-dependent metalloprotease FtsH2. Together with the CLP and the LON family proteases, FtsH2 belongs to the ATP-dependent protease AAA+ (ATPase associated with various cellular activities) superfamily (van Wijk, 2015). FtsH2 showed functional redundancy with FtsH8 and a threshold model (i.e., level of FtsH protein complexes formed in the thylakoid membrane) was proposed for the underlying mechanism of *var2* leaf variegation (Aluru et al., 2006). In analogy to the barley mutant *albostrians* (Li et al., 2019c), green and yellow sectors of *luteostrians* have the same genotype. Furthermore, no differences could be determined for the chloroplast ultrastructure and plastid rRNA content between wild-type leaves and green sectors of striped *luteostrians* leaves. Therefore, variegation of *luteostrians* leaves may be caused by a threshold-dependent mechanism.

Though *HvClpC1* represents the most promising candidate gene for *HvLST*, further experimental evidence is required to confirm this hypothesis. As previously demonstrated, reverse genetic approaches like TILLING or site-directed mutagenesis using Cas9 are viable options in barley to further check the identity of the candidate gene with *LUTEOSTRIANS* (Gottwald et al., 2009; Lawrenson et al., 2015; Li et al., 2019c).

Cloning-by-sequencing in Barley During the Post-NGS Era

In this study, we demonstrated that a combination of low-resolution genetic mapping and comparative whole-genome skim sequencing analysis of mutant and wild type can serve as an effective strategy for gene cloning in barley, especially for genes that share a highly conserved function in plants, including those involved in photosynthesis and chloroplast differentiation and development. The probability of having available functional characterization data for such genes in model plants is high and there is increased probability for seeing related or conserved phenotypic effects in both model and non-model plants. By adopting this strategy, we pinpointed the *albostrians* gene *HvCMF7* among 323 candidates within a 17.86 Mbp interval delimited with 91 F2 genotypes, circumventing the time-consuming and laborious fine-mapping process as performed previously (Li et al., 2019c). Implementation of the pipeline in case of the gene *luteostrians* suggested *HvClpC1* to be the likely causative gene underlying the yellow-stripe variegation phenotype. Notably, the cloning strategy is independent of the effect of genomic location. Compared to the barley genome-wide average ratio of physical to genetic distance of 4.4 Mb/cM (Kunzel et al., 2000; Mascher et al., 2017), *HvAST* and *HvLST* are located in genomic regions with either increased (2.95 Mb/cM) or suppressed (36.35 Mb/cM) rates of recombination, respectively. The success of candidate gene identification in both cases can be attributed to the achieved high-quality and completeness of the reference genome of barley (Mascher et al., 2017; Monat et al., 2019). Whole genome re-sequencing of wild type and mutant parental lines revealed four and eleven candidates for *HvAST* and *HvLST*, respectively. To further narrow down the shortlist to a single candidate gene, it was critical that functional analyses for the putative Arabidopsis homologs were available. In both cases, the

best candidate gene showed a chlorophyll-deficient phenotype for mutants of the respective Arabidopsis homolog.

Genetic mapping information is available for 881 barley mutants of the so-called Bowman near-isogenic lines (NILs) (Druka et al., 2011); among them, 142 mutants with pigmentation phenotype. Although the original genetic background of the Bowman NILs is represented by over 300 different barley genotypes, approximately one-half of the mutants (426/881) were identified in one of eight barley cultivars (i.e., Bonus, Foma, Betzes, Akashinriki, Morex, Steptoe, Volla and Birgitta) (Druka et al., 2011). Since leveraging of the high-quality barley reference genome sequence (Mascher et al., 2017; Monat et al., 2019), whole-genome resequencing of the parental lines would help to filter background mutations and facilitate candidate gene identification and isolation. Therefore, it is conceivable that gene identification in barley, given this pilot attempt based on chlorophyll / photosynthesis related mutants, reached to a point to become even faster and more systematic now.

Materials and Methods:

Plant Materials and Growth Conditions

The *luteostrians* mutant was derived from line MC20 [Institute of Genetics 'Ewald A. Favret' (IGEAF), INTA CICVyA, mutant accession number] by chemical mutagenesis using sodium azide. MC20 was derived from the spring barley cv. Maltería Heda by gamma ray mutagenesis. Two mapping populations were constructed. The six-rowed spring barley (*Hordeum vulgare*) cv. 'Morex' and two-rowed spring barley cv. 'Barke' were used as maternal parent in crossings with the variegated mutant line either *luteostrians*-1 (*lst/LST*) or *luteostrians*-3 (*lst/LST*), respectively. The derived F1 progeny were self-pollinated to generate F2 mapping populations designated as 'ML' (Morex x *luteostrians*) and 'BL' (Barke x *luteostrians*). All the F2 and F3 individuals were grown under the greenhouse condition with a day/night temperature and photoperiod cycle of 20°C/15°C and 16-h light/8-h darkness, respectively. Supplemental light (at 300 µmol photons m^{-2} s^{-1}) was used to extend the natural light with incandescent lamps (SON-T Agro 400, MASSIVEGROW).

For ribosomal RNA analysis, wild type (*luteostrians*-2; genotype *LST*/*LST*) and mutant (*luteostrians*-1; genotype *lst*/*LST*) lines were grown under the greenhouse light condition. For dark treatment, seeds were germinated within a carton box wrapped with aluminum foil under the greenhouse condition mentioned above.

Phenotyping

The trait 'leaf color' was scored for the F2 mapping populations at first, second and third leaf seedling stages. The phenotype of the seedlings was classified into two categories: green and variegated (i.e., green-yellow striped). Green phenotype defined all the three leaves of each seedling were purely green; and variegated phenotype defined green-white striped pattern can be observed on all or any of the three leaves. According to inheritance pattern of the causal gene *luteostrians* (Figure 1B), variegated plants are heterozygous for the mutant allele (*lst/LST*). Green plants, however, can be either wild type (*LST/LST*) or heterozygous (*LST/lst*). Therefore, the *luteostrians* genotype of all the F2 plants was determined by phenotyping 32 seedlings of each corresponding F3 family. A wild type (*LST/LST*) F2 would produce 100% green progeny, progeny of heterozygous (*LST/lst*) F2 green plant would follow a Mendelian segregation (green:variegated = 2:1).

Ribosomal RNA Analysis

Total RNA was extracted using TRIzol reagent (Invitrogen, Braunschweig, Germany) following manufacturer's instructions. Initially, concentration of the RNA was determined by help of a Qubit® 2.0 Fluorometer (Life Technologies, Darmstadt, Germany) according to manufacturer's instructions. RNA samples were further diluted within a quantitative range of 1 - 10 ng/µL. RNA quality and quantity were then measured using Agilent High Sensitivity RNA ScreenTape following the manufacturer's manual (Agilent, Santa Clara, USA).

Ultrastructural analysis

Primary leaves were collected from 7-day-old seedlings of wild type and variegated mutant (*luteostrians*-1). Preparation of 1-2 mm^2 leaf cuttings for ultrastructure analysis including aldehyde/osmium tetroxide fixation, dehydration, resin infiltration and ultramicrotomy were performed as previously described (Li et al., 2019b).

DNA Isolation

Genomic DNA was extracted from primary leaves of 7-day-old seedlings using a GTC-NaCL method in 96-well plate format. Frozen leaf samples were homogeneously crushed by help of a Mixer Mill MM400 (Retsch GmbH, Haan, Germany) for 30 s at 30 Hz. Add 600 µL of preheated (65°C for 1 hour) GTC extraction buffer (1 M guanidine thiocyanate, 2 M NaCl, 30 mM NaCOOH pH = 6.0, 0.2% Tween 20) to the frozen powder and mixed thoroughly for 30 s at 30 Hz under Mixer Mill MM400. Centrifugation at 2500 x *g* for 10 min at 4°C after incubation the samples at 65°C for 1 hour. Transfer 480 µL of supernatant to a 96-well EconoSpin plate (Epoch Life Science, Texas, USA). Vacuum the EconoSpin plate on the vacuum manifold (MACHEREY-NAGEL, Dueren, Germany) until no droplet drop down. Wash the DNA samples twice with 880 µL washing buffer (50 mM NaCl, 10 mM Tris-HCl pH = 8.0, 1 mM EDTA, 70% ethanol). Place the EconoSpin plate onto a 96-well Microtiter™ microplate (Thermo Fisher Scientific, Braunschweig, Germany) and centrifugation at 2500 x *g* for 3 min to remove residual wash solution. Dissolve DNA with 100 µL preheated (65°C for 1 hour) TElight buffer (0.1 mM EDTA, 10 mM Tris-HCl pH = 8.0). The isolated DNA were used for downstream analysis or put at -20°C for long term storage.

Genotyping-by-sequencing

For the genetic mapping of *HvLST*, GBS were prepared from genomic DNA extracted as described previously, digested with *Pst*I and *Msp*I (New England Biolabs, Frankfurt am Main, Germany) following published procedure (Wendler et al., 2015). DNA samples were pooled in an equimolar manner per lane and sequenced on Illumina HiSeq2500 for 107 cycles, single read, using a custom sequencing primer. The GBS reads were aligned to the reference genome of barley as described by Milner et al. (2019). Reads were trimmed using cutadapt (Martin, 2011), mapped with BWA-MEM version 0.7.12a (Li and Durbin, 2009) to barley reference genome (Monat et al., 2019) and the resulting bam files were sorted using Novosort (Novocraft Technologies Sdn Bhd, Selangor, Malaysia). Variants were called with samtools version 1.7 and bcftools version 1.6 (Li, 2011) and filtered following the protocol of Milner et al. (2019) for a minimum depth of sequencing of six to accept a genotype call, a maximum fraction of heterozygous call of 60% and a maximum fraction of 20% of missing data. In the case of 'BL' population, SNP calls were converted to reflect the polymorphisms between the two parents, using the calls for barley cv. Barke.

Genetic Linkage Map Construction

Genetic linkage groups were constructed by help of the JoinMap® 4.1 software (Van Ooijen, 2006) following the instruction manual. Homozygous wild type and heterozygous allele calls were defined as A and H, respectively; missing data was indicated by a dash. A permissive threshold of 5% missing data for both molecular marker and F2 genotype was applied to the datasets. Regression mapping algorithm and Kosambi's mapping function were chosen for building the linkage maps. Markers were grouped into seven groups based on Logarithm of Odds (LOD ≥ 6) groupings. The seven linkage groups were in corresponding to the barley chromosomes according to the locus coordinates determined during read mapping to the reference genome of barley (Monat et al., 2019). Visualization of maps derived from JoinMap® 4.1 was achieved by MapChart software (Voorrips, 2002).

KASP Assay

Sequence 50 bp upstream and downstream of the SNP was extracted from the barley reference genome (Monat et al., 2019). Allele-specific forward primers and one common reverse primer were designed by help of the free assay design service of 3CR Bioscience (https://3crbio.com/free-assay-design/) (Supplemental Table 4). KASP primer mix was prepared in a volume of 100 µL containing 12 µM of each allele-specific forward primer and 30 µM common reverse primer. KASP-PCR reactions were performed in a total volume of 5 µL containing 2.5 µL 2X PACE-IR™ Genotyping Master Mix (001-0010, 3CR Bioscience, Essex, UK), 0.07 µL KASP primer mix, and 40 ng of template DNA. PCR program was used with a HydroCycler (LGC, Teddington, UK): initial denaturation at 94°C for 15 min followed by 10 cycles at 94°C for 20 s, at 65 to 57°C (-0.8°C/cycle) for 1 min, and proceeded for 30 cycles at 94°C for 20 s, at 57°C for 1 min, and followed at 30°C for 30 s. In case of the genotyping clusters not well separated, an additional PCR was performed with 6 cycles at 94°C for 20 s, at 57°C for 1 min. Pre-read and post-read of the fluorescence signals were performed on ABI 7900HT instrument (Applied Biosystems, Thermo Fisher Scientific).

Whole-genome Shotgun Sequencing and Data Analysis

DNA was isolated according to the protocol of Doyle and Doyle (1990) from leaf samples collected from one-week-old, greenhouse-grown seedlings of barley cv. Haisa, *albostrians* mutant M4205, and lines luteostrians-2_2 and luteostrians-3_6. For lines luteostrians-2_2 and luteostrians-3_6**,** DNA was fragmented (200 - 500 bp) with ultrasounds, then 150 bp paired-end libraries were prepared according to Illumina standard protocol and sequenced with Illumina Hiseq X Ten. For Haisa and *albostrians* mutant M4205, 350 bp paired-end libraries were prepared and sequenced according to the protocol as described previously (International Barley Genome Sequencing Consortium, 2012). Reads mapping and variants calling to the barley reference genome were performed as above described for GBS data. Notably, re-sequencing data of barley cv. Barke, maternal parent of the BL mapping population (for *luteostrians*) and BM mapping population (for *albostrians*) (Li et al., 2019c) was also included for SNP calling for identification of mutant-specific SNPs. Functional effect of the mutant-specific SNPs was annotated by help of SNPeff version 4.3 (Cingolani et al., 2012).

Acknowledgements

We gratefully acknowledge Mary Ziems and Heike Harms for their technical support in maintaining plant material, performing crossings for construction of the mapping populations; Manuela Kretschmann for assistance in KASP genotyping; Claudia Riemey for help on

electron microscopy; Heike Mueller for photography. The work was supported by a fellowship of the China Agriculture Research System (CARS-05) and the Agricultural Science and Technology Innovation Program of CAAS to G.G. and by the Deutsche Forschungsgemeinschaft grant (DFG) 1102/13-1 to N.S.

Author Contributions

M.L. and N.S. conceived the research; A.R.P. performed initial genetic characterization of the *luteostrians* mutant; M.L., G.G. and M.M. performed experiments; M.L., H.P. and G.G. analyzed data; M.L. wrote the manuscript with contributions of H.P.; M.L., T.B. and N.S. revised the manuscript.

Supplemental Data

Supplemental Figure 1. Genetic mapping of the *HvLST* gene.
Supplemental Table 1. Summary of SNP markers derived from genotyping-by-sequencing.
Supplemental Table 2. Graphical genotype of selected F2 recombinants.
Supplemental Table 3. Summary of candidate genes for *HvAST* and *HvLST*.
Supplemental Table 4. Primers used in this study.

REFERENCES

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., Tamiru, M., Innan, H., Cano, L., Kamoun, S., and Terauchi, R. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. Nat. Biotechnol. **30,** 174-178.

Aluru, M.R., Yu, F., Fu, A., and Rodermel, S. (2006). Arabidopsis variegation mutants: new insights into chloroplast biogenesis. J Exp Bot **57,** 1871-1881.

Apitz, J., Nishimura, K., Schmied, J., Wolf, A., Hedtke, B., van Wijk, K.J., and Grimm, B. (2016). Posttranslational Control of ALA Synthesis Includes GluTR Degradation by Clp Protease and Stabilization by GluTR-Binding Protein. Plant Physiol. **170,** 2040-2051.

Benn, G., Wang, C.Q., Hicks, D.R., Stein, J., Guthrie, C., and Dehesh, K. (2014). A key general stress response motif is regulated non-uniformly by CAMTA transcription factors. Plant J. **80,** 82-92.

Candela, H., Casanova-Saez, R., and Micol, J.L. (2015). Getting started in mapping-by-sequencing. J. Integr. Plant Biol. **57,** 606-612.

Chen, M., Choi, Y., Voytas, D.F., and Rodermel, S. (2000). Mutations in the Arabidopsis *VAR2* locus cause leaf variegation due to the loss of a chloroplast FtsH protease. Plant J. **22,** 303-313.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) **6,** 80-92.

Clarke, A., Stanne, T., and Sjögren, L. (2005). The ATP-dependent Clp protease in chloroplasts of higher plants. Physiol. Plant. **123,** 406-412.

Colmsee, C., Beier, S., Himmelbach, A., Schmutzer, T., Stein, N., Scholz, U., and Mascher, M. (2015). BARLEX - the Barley Draft Genome Explorer. Mol Plant **8,** 964-966.

Constan, D., Froehlich, J.E., Rangarajan, S., and Keegstra, K. (2004). A stromal Hsp100 protein is required for normal chloroplast development and function in Arabidopsis. Plant Physiol. **136,** 3605-3615.

Daniel-Vedele, F., and Caboche, M. (1993). A tobacco cDNA clone encoding a GATA-1 zinc finger protein homologous to regulators of nitrogen metabolism in fungi. Mol. Gen. Genet. **240,** 365-373.

Doyle, J.J., and Doyle, J.L. (1990). Isolation of plant DNA from fresh tissue. Focus **12,** 13-15.

Druka, A., Franckowiak, J., Lundqvist, U., Bonar, N., Alexander, J., Houston, K., Radovic, S., Shahinnia, F., Vendramin, V., Morgante, M., Stein, N., and Waugh, R. (2011). Genetic dissection of barley morphology and development. Plant Physiol. **155,** 617-627.

Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci. **8,** 978-984.

Fekih, R., Takagi, H., Tamiru, M., Abe, A., Natsume, S., Yaegashi, H., Sharma, S., Sharma, S., Kanzaki, H., Matsumura, H., Saitoh, H., Mitsuoka, C., Utsushi, H., Uemura, A., Kanzaki, E., Kosugi, S., Yoshida, K., Cano, L., Kamoun, S., and Terauchi, R. (2013). MutMap+: genetic mapping and mutant identification without crossing in rice. PLoS One **8,** e68529.

Gottwald, S., Bauer, P., Komatsuda, T., Lundqvist, U., and Stein, N. (2009). TILLING in the two-rowed barley cultivar 'Barke' reveals preferred sites of functional diversity in the gene *HvHox1*. BMC Res. Notes **2,** 258.

Hagemann, R., and Scholz, F. (1962). A case of gene induced mutations of the plasmotype in barley. Theor Appl Genet **32,** 50-59.

Han, C.D., Coe, E.H., Jr., and Martienssen, R.A. (1992). Molecular cloning and characterization of *iojap* (*ij*), a pattern striping gene of maize. EMBO J. **11,** 4037-4046.

Hess, W.R., Prombona, A., Fieder, B., Subramanian, A.R., and Börner, T. (1993). Chloroplast *rps15* and the *rpoB/C1/C2* gene cluster are strongly transcribed in ribosome-deficient plastids: evidence for a functioning non-chloroplast-encoded RNA polymerase. EMBO J. **12,** 563-571.

International Barley Genome Sequencing Consortium. (2012). A physical, genetic and functional sequence assembly of the barley genome. Nature **491,** 711-716.

Jaganathan, D., Bohra, A., Thudi, M., and Varshney, R.K. (2020). Fine mapping and gene cloning in the post-NGS era: advances and prospects. Theor Appl Genet **133,** 1791-1810.

Jost, M., Taketa, S., Mascher, M., Himmelbach, A., Yuo, T., Shahinnia, F., Rutten, T., Druka, A., Schmutzer, T., Steuernagel, B., Beier, S., Taudien, S., Scholz, U., Morgante, M., Waugh, R., and Stein, N. (2016). A Homolog of Blade-On-Petiole 1 and 2 (BOP1/2) Controls Internode Length and Homeotic Changes of the Barley Inflorescence. Plant Physiol. **171,** 1113-1127.

Kim, J., Rudella, A., Ramirez Rodriguez, V., Zybailov, B., Olinares, P.D., and van Wijk, K.J. (2009). Subunits of the plastid ClpPR protease complex have differential contributions to embryogenesis, plastid biogenesis, and plant development in Arabidopsis. Plant Cell **21,** 1669-1692.

Kobayashi, M., Okada, K., and Ikeuchi, M. (2005). A suppressor mutation in the alpha-phycocyanin gene in the light/glucose-sensitive phenotype of the psbK-disruptant of the cyanobacterium Synechocystis sp. PCC 6803. Plant Cell Physiol. **46,** 1561-1567.

Kovacheva, S., Bedard, J., Wardle, A., Patel, R., and Jarvis, P. (2007). Further in vivo studies on the role of the molecular chaperone, Hsp93, in plastid protein import. Plant J. **50,** 364-379.

Kunzel, G., Korzun, L., and Meister, A. (2000). Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics **154,** 397-412.

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Ostergaard, L., Patron, N., Uauy, C., and Harwood, W. (2015). Induction of targeted, heritable mutations in barley and Brassica oleracea using RNA-guided Cas9 nuclease. Genome Biol. **16,** 258.

Li, B., Ferreira, M.A., Huang, M., Camargos, L.F., Yu, X., Teixeira, R.M., Carpinetti, P.A., Mendes, G.C., Gouveia-Mageste, B.C., Liu, C., Pontes, C.S.L., Brustolini, O.J.B., Martins, L.G.C., Melo, B.P., Duarte, C.E.M., Shan, L., He, P., and Fontes, E.P.B. (2019a). The receptor-like kinase NIK1 targets FLS2/BAK1 immune complex and inversely modulates antiviral and antibacterial immunity. Nat Commun **10,** 4996.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **27,** 2987-2993.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25,** 1754-1760.

Li, M., Hensel, G., Melzer, M., Junker, A., Tschiersch, H., Arend, D., Kumlehn, J., Börner, T., and Stein, N. (2019b). Mutation of the ALBOSTRIANS Ohnologous Gene HvCMF3 Impairs Chloroplast Development and Thylakoid Architecture in Barley due to Reduced Plastid Translation. bioRxiv.

Li, M., Hensel, G., Mascher, M., Melzer, M., Budhagatapalli, N., Rutten, T., Himmelbach, A., Beier, S., Korzun, V., Kumlehn, J., Borner, T., and Stein, N. (2019c). Leaf Variegation and Impaired Chloroplast Development Caused by a Truncated CCT Domain Gene in albostrians Barley. Plant Cell **31,** 1430-1445.

Manfield, I.W., Devlin, P.F., Jen, C.H., Westhead, D.R., and Gilmartin, P.M. (2007). Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the Arabidopsis GATA gene family. Plant Physiol. **143,** 941-958.

Martin, M. (2011). CUTADAPT removes adapter sequences from high-throughput sequencing reads. EMBnet.journal **17**.

Mascher, M., Richmond, T.A., Gerhardt, D.J., Himmelbach, A., Clissold, L., Sampath, D., Ayling, S., Steuernagel, B., Pfeifer, M., D'Ascenzo, M., Akhunov, E.D., Hedley, P.E., Gonzales, A.M., Morrell, P.L., Kilian, B., Blattner, F.R., Scholz, U., Mayer, K.F., Flavell, A.J., Muehlbauer, G.J., Waugh, R., Jeddeloh, J.A., and Stein, N. (2013). Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond. Plant J. **76,** 494-505.

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T., Radchuk, V., Dockter, C., Hedley, P.E., Russell, J., Bayer, M., Ramsay, L., Liu, H., Haberer, G., Zhang, X.Q., Zhang, Q., Barrero, R.A., Li, L., Taudien, S., Groth, M., Felder, M., Hastie, A., Simkova, H., Stankova, H., Vrana, J., Chan, S., Munoz-Amatriain, M., Ounit, R., Wanamaker, S., Bolser, D., Colmsee, C., Schmutzer, T., Aliyeva-Schnorr, L., Grasso, S., Tanskanen, J., Chailyan, A., Sampath, D., Heavens, D., Clissold, L., Cao, S., Chapman, B., Dai, F., Han, Y., Li, H., Li, X., Lin, C., McCooke, J.K., Tan, C., Wang, P., Wang, S., Yin, S., Zhou, G., Poland, J.A., Bellgard, M.I., Borisjuk, L., Houben, A., Dolezel, J., Ayling, S., Lonardi, S., Kersey, P., Langridge, P., Muehlbauer, G.J., Clark, M.D., Caccamo, M., Schulman, A.H., Mayer, K.F.X., Platzer, M., Close, T.J., Scholz, U., Hansson, M., Zhang, G., Braumann, I., Spannagl, M., Li, C., Waugh, R., and Stein, N. (2017). A chromosome conformation capture ordered sequence of the barley genome. Nature **544,** 427-433.

Milner, S.G., Jost, M., Taketa, S., Mazon, E.R., Himmelbach, A., Oppermann, M., Weise, S., Knupffer, H., Basterrechea, M., Konig, P., Schuler, D., Sharma, R., Pasam, R.K., Rutten, T., Guo, G., Xu, D., Zhang, J., Herren, G., Muller, T., Krattinger, S.G., Keller, B., Jiang, Y., Gonzalez, M.Y., Zhao, Y., Habekuss, A., Farber, S., Ordon, F., Lange, M., Borner, A., Graner, A., Reif, J.C., Scholz, U., Mascher, M., and Stein, N. (2019). Genebank genomics highlights the diversity of a global barley collection. Nat. Genet. **51,** 319-326.

Monat, C., Padmarasu, S., Lux, T., Wicker, T., Gundlach, H., Himmelbach, A., Ens, J., Li, C., Muehlbauer, G.J., Schulman, A.H., Waugh, R., Braumann, I., Pozniak, C., Scholz, U., Mayer, K.F.X., Spannagl, M., Stein, N., and Mascher, M. (2019). TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools. Genome Biol. **20,** 284.

Mount, D.W. (2007). Using the Basic Local Alignment Search Tool (BLAST). CSH Protoc **2007,** pdb top17.

Nakagawara, E., Sakuraba, Y., Yamasato, A., Tanaka, R., and Tanaka, A. (2007). Clp protease controls chlorophyll b synthesis by regulating the level of chlorophyllide a oxygenase. Plant J. **49,** 800-809.

Nishimura, K., and van Wijk, K.J. (2015). Organization, function and substrates of the essential Clp protease system in plastids. Biochim. Biophys. Acta **1847,** 915-930.

Nishimura, K., Asakura, Y., Friso, G., Kim, J., Oh, S.H., Rutschow, H., Ponnala, L., and van Wijk, K.J. (2013). ClpS1 Is a Conserved Substrate Selector for the Chloroplast Clp Protease System in Arabidopsis. Plant Cell **25,** 2276-2301.

Olinares, P.D., Kim, J., and van Wijk, K.J. (2011). The Clp protease system; a central component of the chloroplast protease network. Biochim. Biophys. Acta **1807,** 999-1011.

Perea-Resa, C., Hernandez-Verdeja, T., Lopez-Cobollo, R., del Mar Castellano, M., and Salinas, J. (2012). LSM proteins provide accurate splicing and decay of selected transcripts to ensure normal Arabidopsis development. Plant Cell **24,** 4930-4947.

Prigge, M.J., Platre, M., Kadakia, N., Zhang, Y., Greenham, K., Szutu, W., Pandey, B.K., Bhosale, R.A., Bennett, M.J., Busch, W., and Estelle, M. (2020). Genetic analysis of the Arabidopsis TIR1/AFB auxin receptors reveals both overlapping and specialized functions. Elife **9**.

Rodriguez-Concepcion, M., D'Andrea, L., and Pulido, P. (2019). Control of plastidial metabolism by the Clp protease complex. J Exp Bot **70,** 2049-2058.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Nielsen, K.L., Jorgensen, J.E., Weigel, D., and Andersen, S.U. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat Methods **6,** 550-551.

Singh, R., Low, E.T., Ooi, L.C., Ong-Abdullah, M., Ting, N.C., Nagappan, J., Nookiah, R., Amiruddin, M.D., Rosli, R., Manaf, M.A., Chan, K.L., Halim, M.A., Azizi, N., Lakey, N., Smith, S.W., Budiman, M.A., Hogan, M., Bacher, B., Van Brunt, A., Wang, C., Ordway, J.M., Sambanthamurthi, R., and Martienssen, R.A. (2013). The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. Nature **500,** 340-344.

Sjogren, L.L., MacDonald, T.M., Sutinen, S., and Clarke, A.K. (2004). Inactivation of the clpC1 gene encoding a chloroplast Hsp100 molecular chaperone causes growth retardation, leaf chlorosis, lower photosynthetic activity, and a specific reduction in photosystem content. Plant Physiol. **136,** 4114-4126.

Sjogren, L.L., Tanabe, N., Lymperopoulos, P., Khan, N.Z., Rodermel, S.R., Aronsson, H., and Clarke, A.K. (2014). Quantitative analysis of the chloroplast molecular chaperone ClpC/Hsp93 in Arabidopsis reveals new insights into its localization, interaction with the Clp proteolytic core, and functional importance. J. Biol. Chem. **289,** 11318-11330.

Sun, C.W., Chen, L.J., Lin, L.C., and Li, H.M. (2001). Leaf-specific upregulation of chloroplast translocon genes by a CCT motif-containing protein, CIA2. Plant Cell **13,** 2053-2061.

Takechi, K., Sodmergen, Murata, M., Motoyoshi, F., and Sakamoto, W. (2000). The YELLOW VARIEGATED (VAR2) locus encodes a homologue of FtsH, an ATP-dependent protease in Arabidopsis. Plant Cell Physiol. **41,** 1334-1346.

Tehseen, M., Cairns, N., Sherson, S., and Cobbett, C.S. (2010). Metallochaperone-like genes in Arabidopsis thaliana. Metallomics **2,** 556-564.

Toshoji, H., Katsumata, T., Takusagawa, M., Yusa, Y., and Sakai, A. (2012). Effects of chloroplast dysfunction on mitochondria: white sectors in variegated leaves have higher mitochondrial DNA levels and lower dark respiration rates than green sectors. Protoplasma **249,** 805-817.

Van Ooijen, J.W. (2006). JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands.

van Wijk, K.J. (2015). Protein maturation and proteolysis in plant plastids, mitochondria, and peroxisomes. Annu. Rev. Plant Biol. **66,** 75-111.

Voorrips, R.E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. J. Hered. **93,** 77-78.

Walbot, V., and Coe, E.H. (1979). Nuclear gene *iojap* conditions a programmed change to ribosome-less plastids in *Zea mays*. Proc. Natl. Acad. Sci. U. S. A. **76,** 2760-2764.

Ward, J.M., Smith, A.M., Shah, P.K., Galanti, S.E., Yi, H., Demianski, A.J., van der Graaff, E., Keller, B., and Neff, M.M. (2006). A new role for the Arabidopsis AP2 transcription

factor, LEAFY PETIOLE, in gibberellin-induced germination is revealed by the misexpression of a homologous gene, SOB2/DRN-LIKE. Plant Cell **18,** 29-39.

Wendler, N., Mascher, M., Himmelbach, A., Johnston, P., Pickering, R., and Stein, N. (2015). Bulbosum to Go: A Toolbox to Utilize Hordeum vulgare/bulbosum Introgressions for Breeding and Beyond. Mol Plant **8,** 1507-1519.

Wilson-Sanchez, D., Lup, S.D., Sarmiento-Manus, R., Ponce, M.R., and Micol, J.L. (2019). Next-generation forward genetic screens: using simulated data to improve the design of mapping-by-sequencing experiments in Arabidopsis. Nucleic Acids Res. **47,** e140.

Yang, W., Dong, R., Liu, L., Hu, Z., Li, J., Wang, Y., Ding, X., and Chu, Z. (2016). A novel mutant allele of SSI2 confers a better balance between disease resistance and plant growth inhibition on Arabidopsis thaliana. BMC Plant Biol. **16,** 208.

Yu, F., Park, S., and Rodermel, S.R. (2004). The Arabidopsis FtsH metalloprotease gene family: interchangeability of subunits in chloroplast oligomeric complexes. Plant J. **37,** 864-876.

Yu, F., Fu, A.G., Aluru, M., Park, S., Xu, Y., Liu, H.Y., Liu, X.Y., Foudree, A., Nambogga, M., and Rodermel, S. (2007). Variegation mutants and mechanisms of chloroplast biogenesis. Plant Cell and Environment **30,** 350-365.

Table 1. Phenotypic segregation of zygotic lethality in F2 of BL and ML populations

	Number of scored heterozygous F2 plants	viable F3 grains	non-viable F3 grains	$\chi^2_{(df=1)}$	*p* value
BL population	39	609	210	0.18	0.67*
ML population	26	343	109	0.19	0.66*

* Statistically no significant derivation from the expected 3:1 (fertile:sterile) segregation ratio.

682

Table 2. Genotypic segregation of the *HvLST* locus in F2 of BL and ML populations

	Maternal	Paternal	Population size	Wild type	Heterozygote	$\chi^2_{(df=1)}$	*p* value
BL Population	Barke	*luteostrians (lst/LST)*	269	95	172	0.61	0.44*
ML Population	Morex	*luteostrians (lst/LST)*	271	117	152	12.47	0.0004[$]

* Statistically no significant derivation from the expected 1:2 (wild type:heterozygote) segregation ratio.

[$] Segregation distortion observed for ML population.

683

Table 3. Summary of candidate genes for *HvAST* and *HvLST*

Gene ID*	Confidence	Annotation	Homolog in Arabidopsis[$]	Phenotype / Function	Reference
Candidate genes for *HvAST*					
HORVU.MOREX.r2.7HG0603920.1	HC	Zinc finger protein CONSTANS-LIKE 2	*AT5G57180* (*CIA2*)	*cia2* mutant exhibits a pale green phenotype	(Sun et al., 2001)
HORVU.MOREX.r2.7HG0603970.1	LC	Retrotransposon protein, putative, unclassified	*AT1G43760*	Uncharacterized in Arabidopsis	n.a
HORVU.MOREX.r2.7HG0604040.1	HC	GATA transcription factor 27	*AT1G51600* (*ZML2*)	GATA transcription factors are known to be involved in light-dependent gene regulation and nitrate assimilation in plants	(Manfield et al., 2007); (Daniel-Vedele and Caboche, 1993)
HORVU.MOREX.r2.7HG0604110.1	HC	Receptor-like kinase	*AT5G16000* (*NIK1*)	Dwarfed morphology, enhanced disease resistance to bacteria and increased PAMP-triggered immunity responses	(Li et al., 2019a)
Candidate genes for *HvLST*					
HORVU.MOREX.r2.2HG0133160.1	HC	Acyl-[acyl-carrier-protein] desaturase	*AT2G43710* (*SS2*)	Decreased growth and increased disease resistance	(Yang et al., 2016)
HORVU.MOREX.r2.2HG0133350.1	HC	Retrotransposon protein, putative, unclassified	n.a	Absence in Arabidopsis	n.a
HORVU.MOREX.r2.2HG0133900.1	LC	Retrotransposon protein, putative, Ty3-gypsy subclass	*ATMG00860*	Mitochondria-specific gene *ORF158*	n.a
HORVU.MOREX.r2.2HG0134330.1	HC	Calmodulin-binding transcription activator	*AT1G67310* (*CAMTA4*)	Positive regulation of a general stress response	(Benn et al., 2014)
HORVU.MOREX.r2.2HG0134390.1	HC	U6 snRNA-associated Sm-like protein LSm6	*AT2G43810* (*LSM6B*)	*lsm6b* mutant exhibits wild-type phenotype	(Perea-Resa et al., 2012)
HORVU.MOREX.r2.2HG0134800.1	HC	Photosystem II reaction center protein K	*ATCG00070* (*psbK*)	Chloroplast-specific p*sb*K gene; *psb*K is not essential for PSII activity in cyanobacterium *Synechocystis* 6803	(Kobayashi et al., 2005)
HORVU.MOREX.r2.2HG0135080.1	HC	Transport inhibitor response 1-like protein	*AT3G26810* (*AFB2*)	Resistance to IAA; a role in shoot development	(Prigge et al., 2020)
HORVU.MOREX.r2.2HG0135330.1	HC	Heavy-metal-associated domain-containing family protein	*AT1G22990* (*HIPP22*)	*hipp22* mutant exhibits wild-type phenotype; a role in Cd-detoxification	(Tehseen et al., 2010)
HORVU.MOREX.r2.2HG0135340.1	HC	ATP-dependent Clp protease ATP-binding subunit ClpC1	*AT5G50920* (*ClpC1*)	Retarded growth; leaf chlorosis; lower photosynthetic activity; reduction in photosystem content	(Sjogren et al., 2004)
HORVU.MOREX.r2.2HG0135360.1	HC	Protein TSSC4	*AT5G13970*	Uncharacterized in Arabidopsis; *TSSC4* represents tumor suppressing subtransferable candidate 4 in *Homo sapiens*	n.a
HORVU.MOREX.r2.2HG0135690.1	HC	Ethylene-responsive transcription factor, putative	*AT5G13910* (*LEP*)	*lep-1* mutant exhibits short hypocotyls and small cotyledons	(Ward et al., 2006)

*The gene ID in bold indicates the *HvAST* locus and potential *HvLST* locus.

[$]The gene ID is included in parentheses.

Figure Legends

Figure 1. Phenotype and inheritance of variegation in the barley mutant *luteostrians*.

(A) Penetration of the mutant phenotype varies among seedlings, ranging from a narrow yellow stripe to complete yellowish or albino. Neither yellowish/albino plants survived beyond third-leaf stage.
(B) Inheritance pattern of the *luteostrians* mutant phenotype. Variegation only occurs in plants if the *lst* allele was transmitted through the female gamete. Upper panel: Heterozygous plants can be obtained by using either green or variegated plants (heterozygous for the *luteostrians* allele) as pollen donor. This will generate 50% F1-progeny heterozygous for *luteostrians* (panel F1). Progenies of selfed F1 heterozygotes will exhibit Mendelian segregation in F2; the variegated phenotype, however, will appear only in 50% of the heterozygous plants, carrying the mutant allele inherited from F1 female gamete. Zygotes homozygous for the *luteostrians* allele will be aborted as homozygosity of *luteostrians* early zygotic lethal. Lower panel: Green phenotype of homozygous wild type plants in F2 will be stably transmitted in F3; progenies of heterozygous F2 plants follow a Mendelian inheritance pattern in F3.

Figure 2. rRNA analysis of wild type and *HvIst* mutant.

(A) Separation of cytosolic and plastid rRNAs using the Agilent high sensitivity RNA ScreenTape assay.
(B) Analysis of rRNA from wild type and *HvIst* mutant using Agilent Tapestation 4200. The respective cytosolic and plastid rRNA species are indicated by arrows.

Figure 3. Ultrastructural analysis of chloroplasts of the wild type and *HvIst* mutant.

(A-C) Ultrastructural analysis of wild type (A), and green (B) and yellow (C) sectors of the *HvIst* mutant, respectively. Wild type and green leaves of the *HvIst* mutant contain chloroplasts with fully differentiated grana and stroma thylakoids. By contrast, in plastids of yellow leaves of the *HvIst* mutant only some pieces of undeveloped membranes were observed.
(D-F) Larger magnification of square areas of the corresponding plastid in the top panels A to C. GA, grana; ST, stroma thylakoid; PL, plastoglobuli.

Figure 4. Saturation mapping of the delineated *HvLST* genetic interval.

(A) Marker saturation around the *HvLST* locus in BL population.
(B) Marker saturation around the *HvLST* locus in ML population. Recombination events and genetic distance (recombinants/cM) between the neighboring markers are shown on the left of each genetic map. Markers co-segregating with the gene *HvLST* are indicated by a vertical green line. Red font is highlighting the closest flanking markers.

Figure 5. Workflow of candidate gene identification in barley photosynthetic mutants exemplified for *HvAST* and *HvLST*.

(A) The initial step of the strategy built on low-resolution genetic mapping. Whole genome re-sequencing data for the mutant and its corresponding wild type genotype was then generated for mapping and variation calling against the Morex v2 reference sequence (Monat et al., 2019).
(B) Candidate gene identification was exemplified on the basis of the previously cloned gene *albostrians* (*HvAST)*. 323 genes were annotated for the physical interval of ~18 Mbp initially delimited by low-resolution genetic mapping. SNP variation was found in 9 genes, while only 4 genes carried non-synonymous or other deleterious mutations. A single candidate gene (the confirmed gene *HvAST*) could be delimited based on available functional annotation information indicating a role in plastid biology / photosynthesis.
(C) A similar strategy was applied to the cloning of *luteostrians*. 284 genes were annotated for the initial genetic interval. Eventually, a single candidate gene with predicted functional annotation for a role in plastid biology / photosynthesis and non-synonymous / deleterious mutation could be spotted. HC, high-confidence gene; LC, low-confidence genes. Criteria for classification of HC and LC refers to Mascher et al. (2017).

Figure 6. Validation of the heterozygous SNP of *HvClpC1* by Sanger sequencing and expression profiles of *HvClpC1* and homologs.

(A) Gene structure of *HvLST* candidate (*HvClpC1*) and its two closest homologs. Black boxes indicate exons and horizontal lines indicate introns. Green areas indicate chloroplast transit peptides as predicted by ChloroP (Emanuelsson et al., 1999). The first and second introns of homolog 1 are not drawn at scale as indicated by the interrupted lines; the actual length is shown above gaps, respectively. Red filled triangle indicates SNP position of *HvClpC1*.
(B) Chromatogram of Sanger sequencing. Red arrow indicates position of the heterozygous SNP of *HvClpC1* in the original mutant line luteostrians-1_1. Details of the SNP are illustrated in the table below.
(C) Expression profiles of *HvClpC1* and its two closest homologs. The expression levels are given as fragments per kilobase of exon per million reads mapped (FPKM) across sixteen different tissues or developmental stages. The data was taken from Mascher et al. (2017). EMB, 4-day embryos; ROO1, roots from seedlings (10 cm shoot stage); LEA, shoots from seedlings (10 cm shoot stage); INF1, young developing inflorescences (5 mm); INF2, developing inflorescences (1-1.5 cm); NOD, developing tillers, 3rd internode (42 DAP); CAR5, developing grain (5 DAP); CAR15, developing grain (15 DAP); ETI, etiolated seedling, dark condition (10 DAP); LEM, inflorescences, lemma (42 DAP); LOD, inflorescences, lodicule (42 DAP); PAL, dissected inflorescences, palea (42 DAP); EPI, epidermal strips (28 DAP); RAC, inflorescences, rachis (35 DAP); ROO2, roots (28 DAP); SEN, senescing leaves (56 DAP).

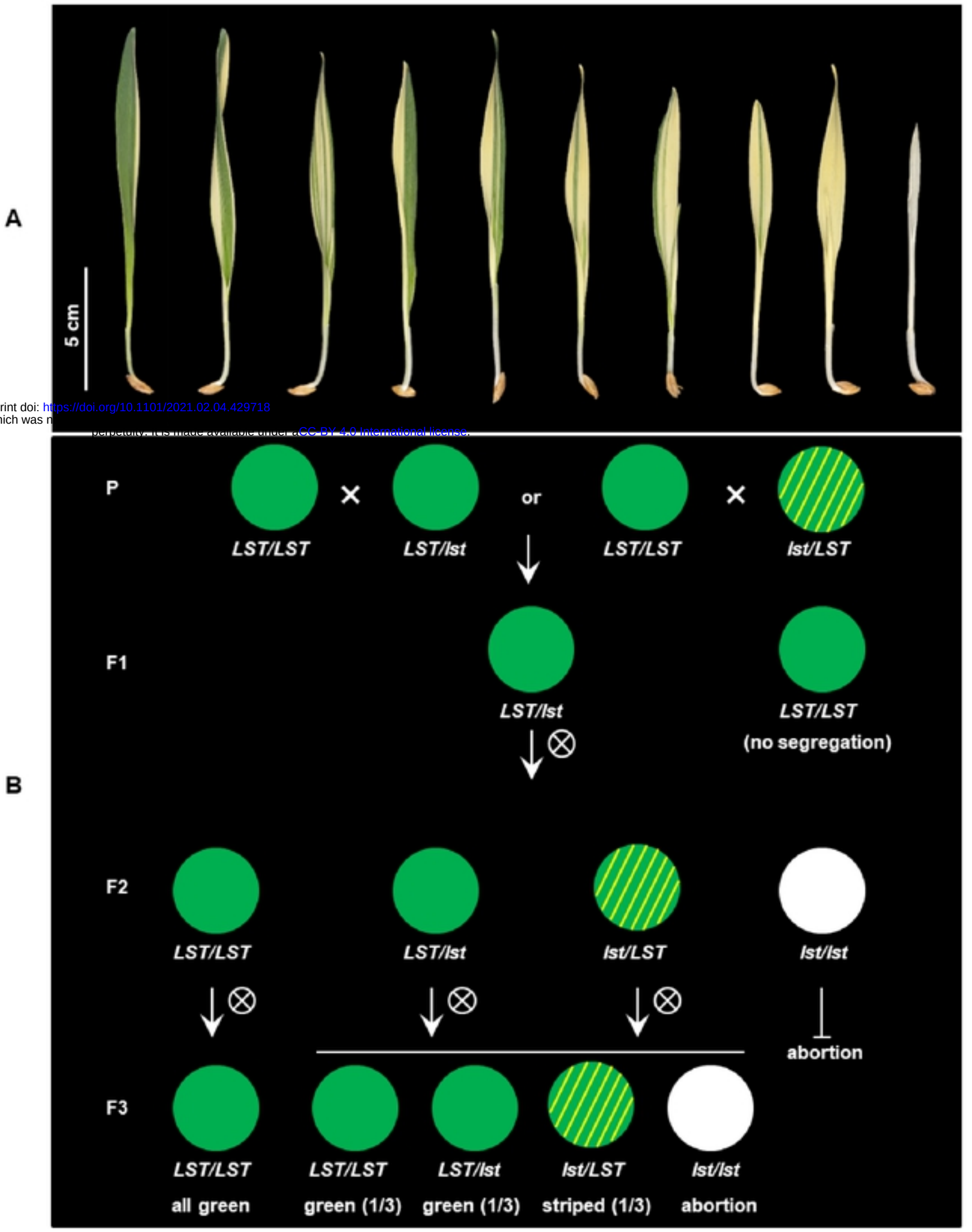


Figure 1. Phenotype and inheritance of variegation in the barley mutant *luteostrians*.

(A) Penetration of the mutant phenotype varies among seedlings, ranging from a narrow yellow stripe to complete yellowish or albino. Neither yellowish/albino plants survived beyond third-leaf stage.
(B) Inheritance pattern of the *luteostrians* mutant phenotype. Variegation only occurs in plants if the *lst* allele was transmitted through the female gamete. Upper panel: Heterozygous plants can be obtained by using either green or variegated plants (heterozygous for the *luteostrians* allele) as pollen donor. This will generate 50% F1-progeny heterozygous for *luteostrians* (panel F1). Progenies of selfed F1 heterozygotes will exhibit Mendelian segregation in F2; the

variegated phenotype, however, will appear only in 50% of the heterozygous plants, carrying the mutant allele inherited from F1 female gamete. Zygotes homozygous for the *luteostrians* allele will be aborted as homozygosity of *luteostrians* early zygotic lethal. Lower panel: Green phenotype of homozygous wild type plants in F2 will be stably transmitted in F3; progenies of heterozygous F2 plants follow a Mendelian inheritance pattern in F3.

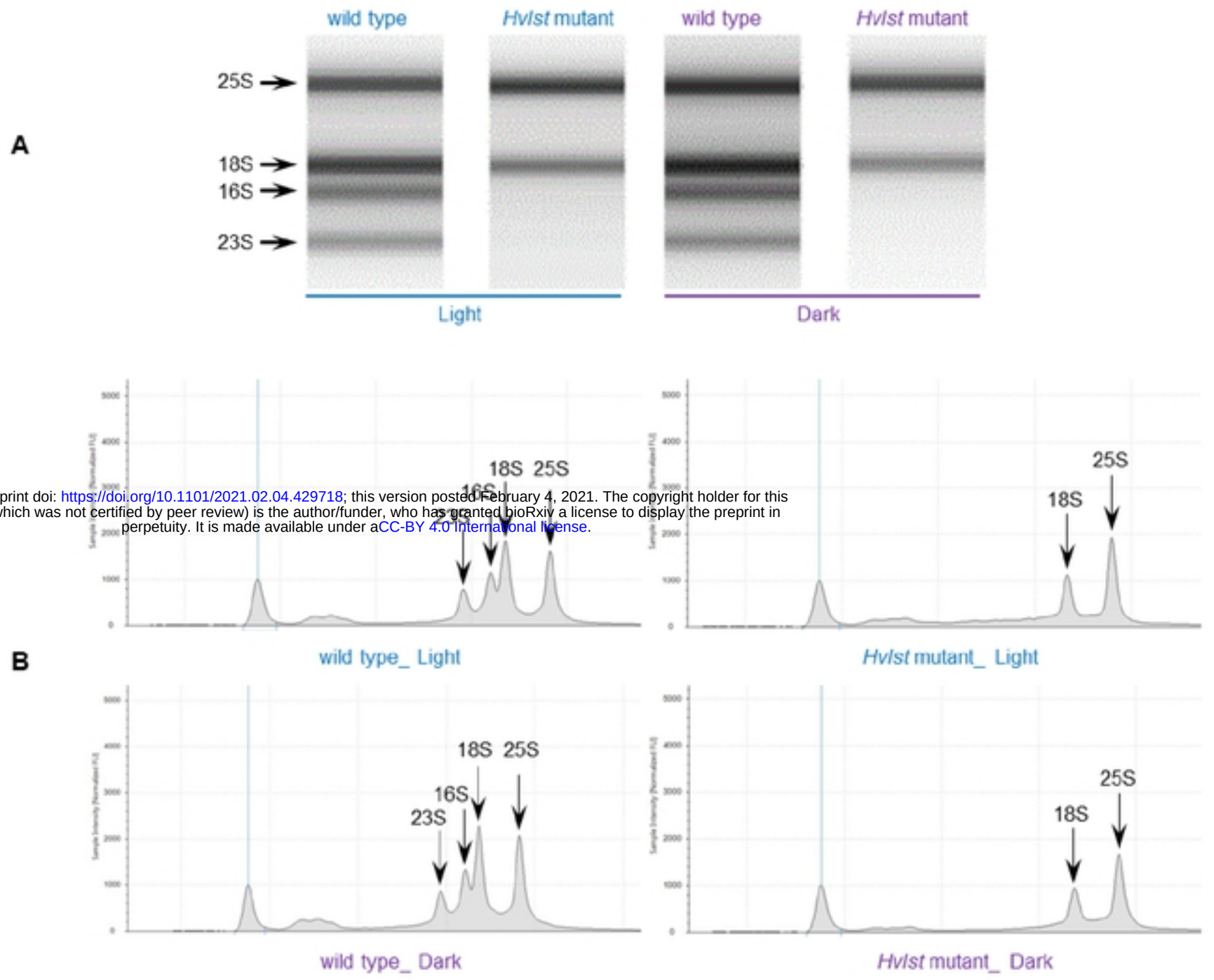


Figure 2. rRNA analysis of wild type and *Hvlst* mutant.

(A) Separation of cytosolic and plastid rRNAs using the Agilent high sensitivity RNA ScreenTape assay.
(B) Analysis of rRNA from wild type and *Hvlst* mutant using Agilent Tapestation 4200. The respective cytosolic and plastid rRNA species are indicated by arrows.

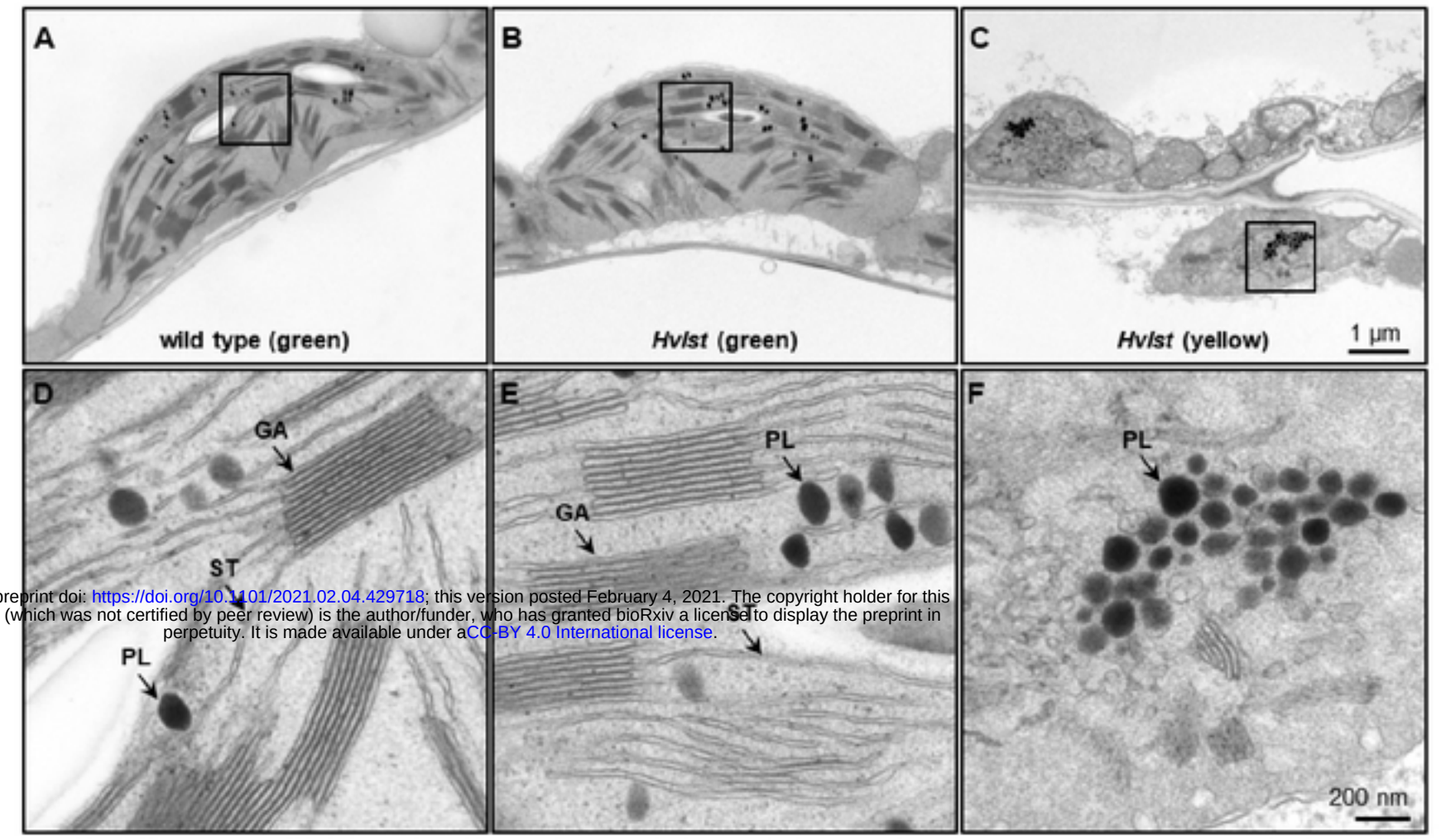


Figure 3. Ultrastructural analysis of chloroplasts of the wild type and *Hvlst* mutant.

(A-C) Ultrastructural analysis of wild type (A), and green (B) and yellow (C) sectors of the *Hvlst* mutant, respectively. Wild type and green leaves of the *Hvlst* mutant contain chloroplasts with fully differentiated grana and stroma thylakoids. By contrast, in plastids of yellow leaves of the *Hvlst* mutant only some pieces of undeveloped membranes were observed.
(D-F) Larger magnification of square areas of the corresponding plastid in the top panels A to C. GA, grana; ST, stroma thylakoid; PL, plastoglobuli.

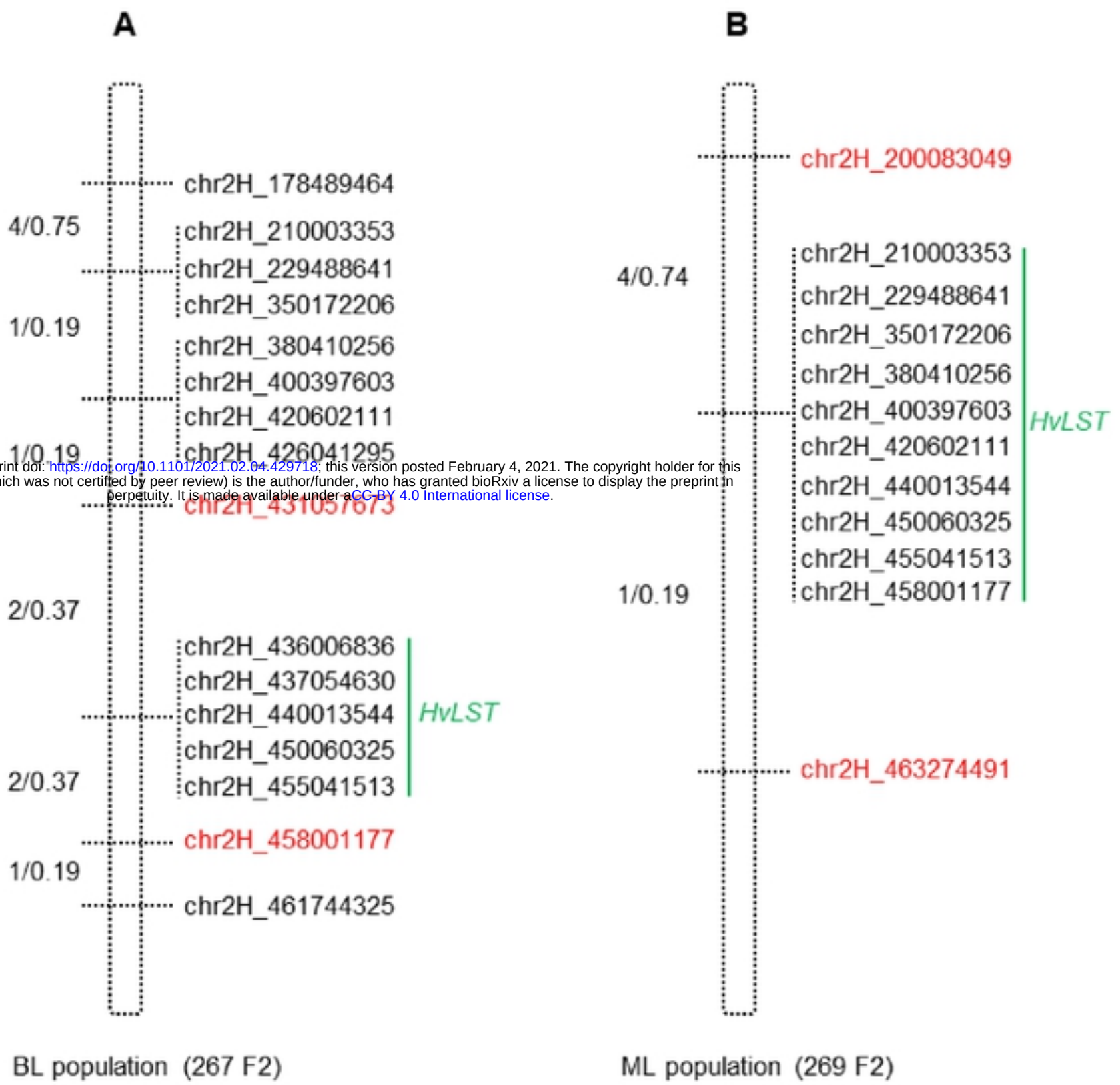


Figure 4. Saturation mapping of the delineated *HvLST* genetic interval.

(A) Marker saturation around the *HvLST* locus in BL population.
(B) Marker saturation around the *HvLST* locus in ML population. Recombination events and genetic distance (recombinants/cM) between the neighboring markers are shown on the left of each genetic map. Markers co-segregating with the gene *HvLST* are indicated by a vertical green line. Red font is highlighting the closest flanking markers.

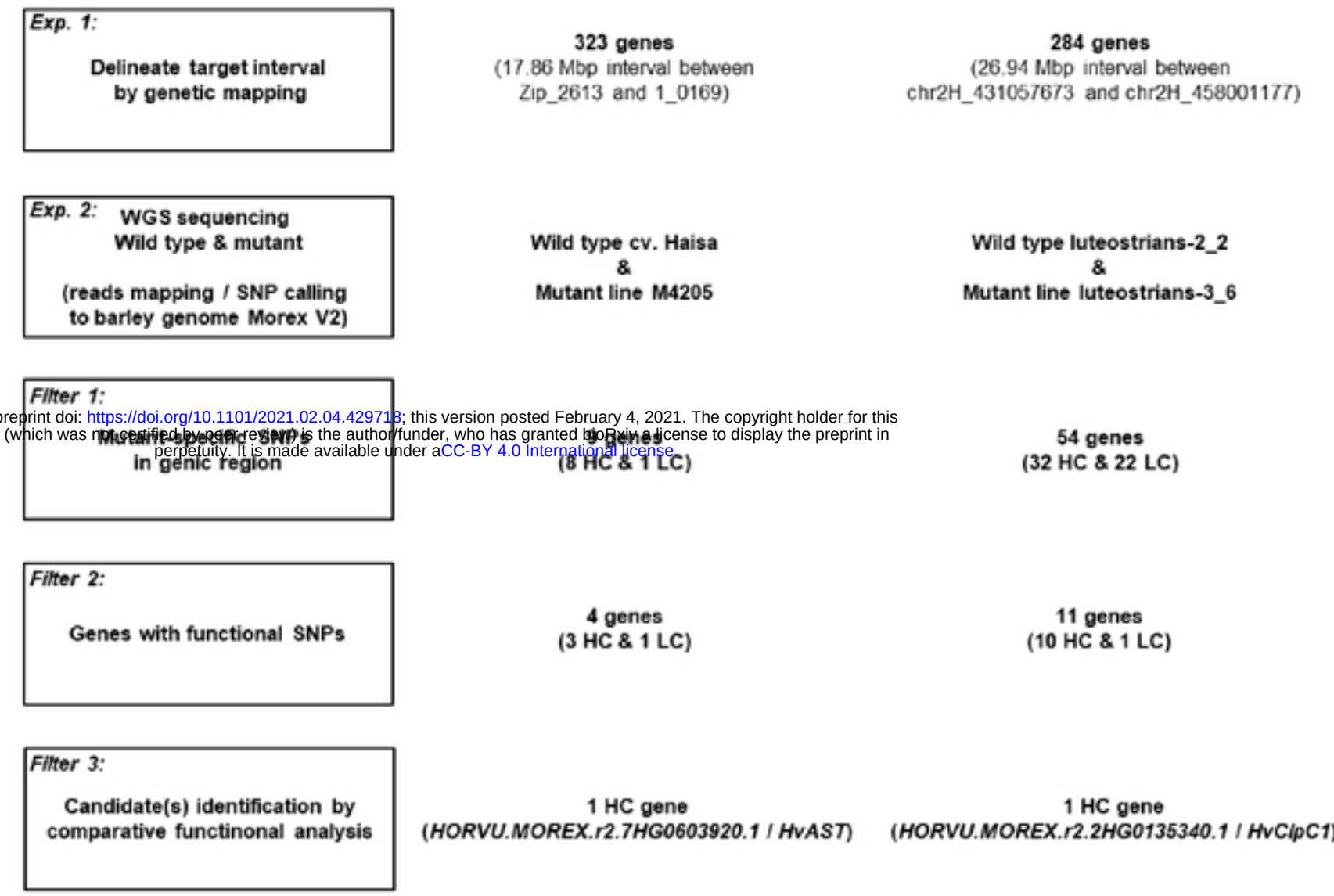
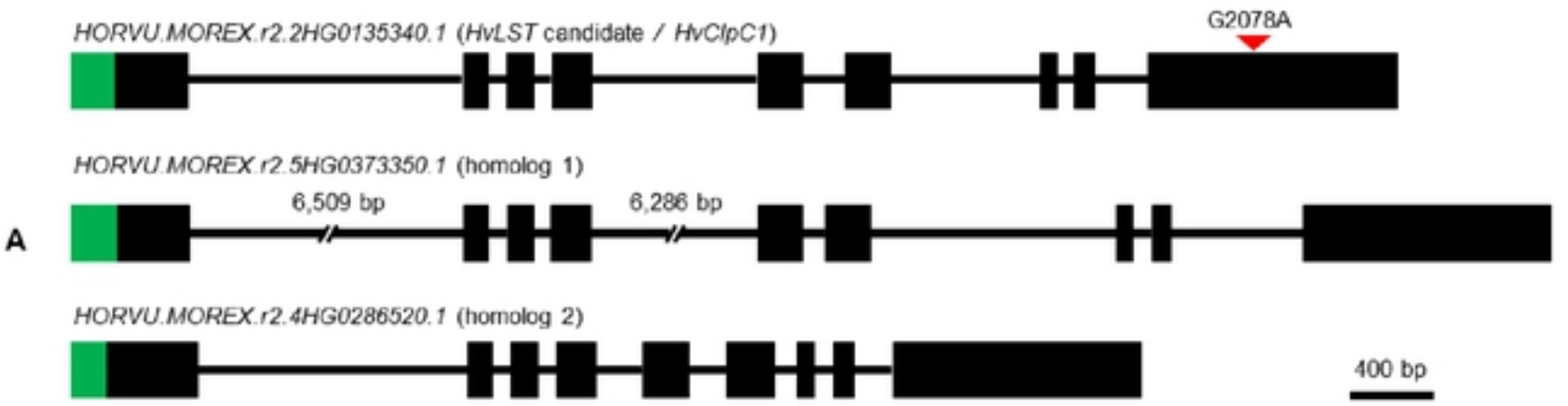
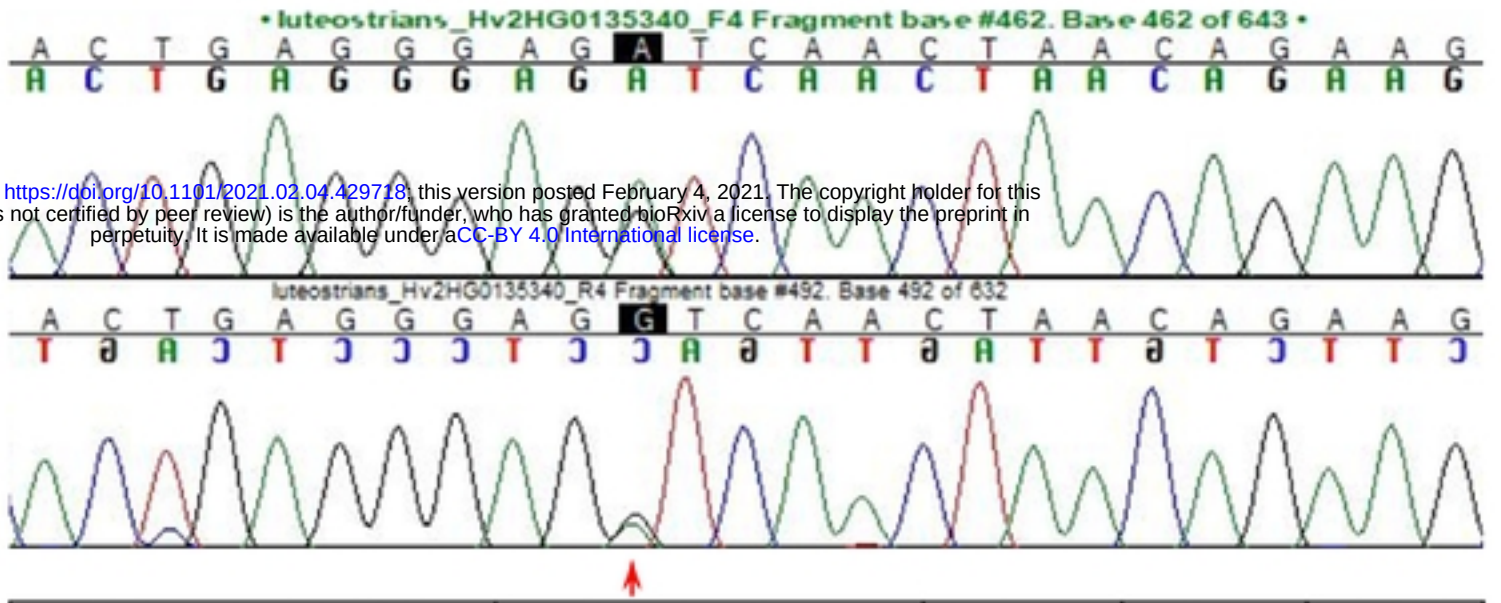


Figure 5. Workflow of candidate gene identification in barley photosynthetic mutants exemplified for *HvAST* and *HvLST*.

(A) The initial step of the strategy built on low-resolution genetic mapping. Whole genome re-sequencing data for the mutant and its corresponding wild type genotype was then generated for mapping and variation calling against the Morex v2 reference sequence (Monat et al., 2019).
(B) Candidate gene identification was exemplified on the basis of the previously cloned gene *albostrians* (*HvAST*). 323 genes were annotated for the physical interval of ~18 Mbp initially delimited by low-resolution genetic mapping. SNP variation was found in 9 genes, while only 4 genes carried non-synonymous or other deleterious mutations. A single candidate gene (the confirmed gene *HvAST*) could be delimited based on available functional annotation information indicating a role in plastid biology / photosynthesis.
(C) A similar strategy was applied to the cloning of *luteostrians*. 284 genes were annotated for the initial genetic interval. Eventually, a single candidate gene with predicted functional annotation for a role in plastid biology / photosynthesis and non-synonymous / deleterious mutation could be spotted. HC, high-confidence gene; LC, low-confidence genes. Criteria for classification of HC and LC refers to Mascher et al. (2017).





SNP position (CDS)	Exon location	Original	Mutant	Effect
2078	9th	G**G**T	G**A**T	G693D

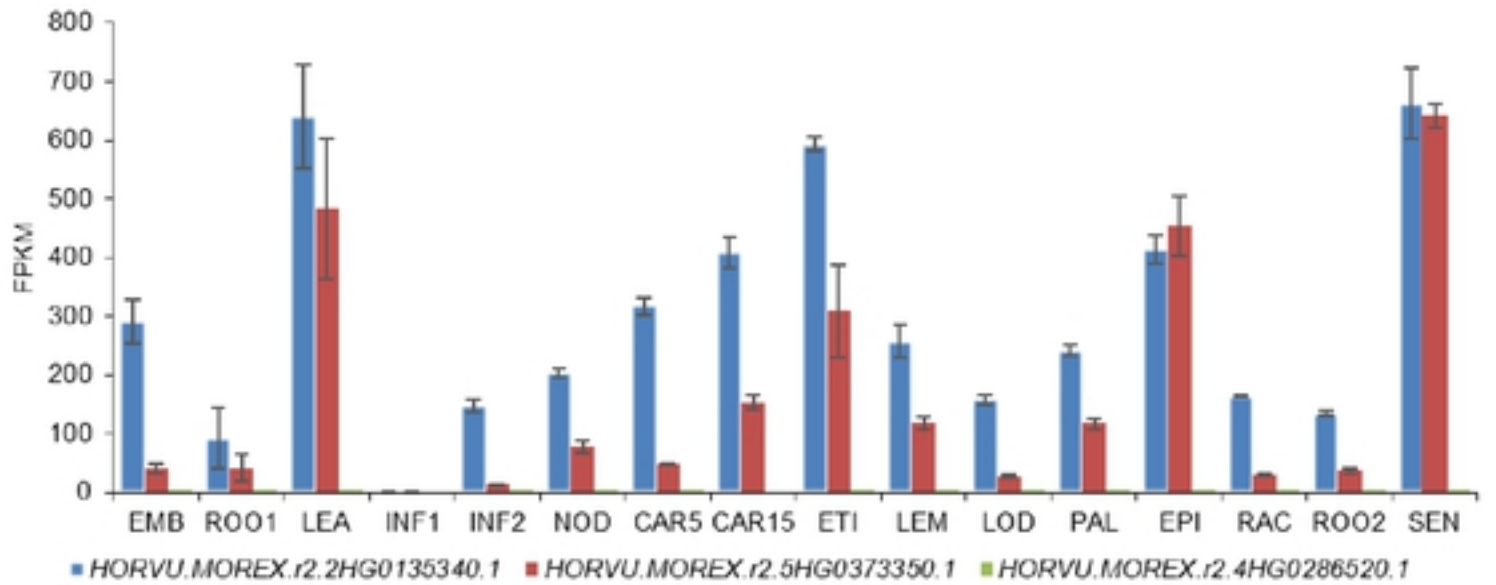


Figure 6. Validation of the heterozygous SNP of *HvClpC1* by Sanger sequencing and expression profiles of *HvClpC1* and homologs.

(A) Gene structure of *HvLST* candidate (*HvClpC1*) and its two closest homologs. Black boxes indicate exons and horizontal lines indicate introns. Green areas indicate chloroplast transit peptides as predicted by ChloroP (Emanuelsson et al., 1999). The first and second introns of homolog 1 are not drawn at scale as indicated by the interrupted lines; the actual length is shown above gaps, respectively. Red filled triangle indicates SNP position of *HvClpC1*.
(B) Chromatogram of Sanger sequencing. Red arrow indicates position of the heterozygous SNP of *HvClpC1* in the original mutant line luteostrians-1_1. Details of the SNP are illustrated in the table below.
(C) Expression profiles of *HvClpC1* and its two closest homologs. The expression levels are given as fragments per kilobase of exon per million reads mapped (FPKM) across sixteen different tissues or developmental stages. The data was taken from Mascher et al. (2017). EMB, 4-day embryos; ROO1, roots from seedlings (10 cm shoot stage); LEA, shoots from

seedlings (10 cm shoot stage); INF1, young developing inflorescences (5 mm); INF2, developing inflorescences (1-1.5 cm); NOD, developing tillers, 3rd internode (42 DAP); CAR5, developing grain (5 DAP); CAR15, developing grain (15 DAP); ETI, etiolated seedling, dark condition (10 DAP); LEM, inflorescences, lemma (42 DAP); LOD, inflorescences, lodicule (42 DAP); PAL, dissected inflorescences, palea (42 DAP); EPI, epidermal strips (28 DAP); RAC, inflorescences, rachis (35 DAP); ROO2, roots (28 DAP); SEN, senescing leaves (56 DAP).

References

Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci. **8**, 978-984.

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T., Radchuk, V., Dockter, C., Hedley, P.E., Russell, J., Bayer, M., Ramsay, L., Liu, H., Haberer, G., Zhang, X.Q., Zhang, Q., Barrero, R.A., Li, L., Taudien, S., Groth, M., Felder, M., Hastie, A., Simkova, H., Stankova, H., Vrana, J., Chan, S., Munoz-Amatriain, M., Ounit, R., Wanamaker, S., Bolser, D., Colmsee, C., Schmutzer, T., Aliyeva-Schnorr, L., Grasso, S., Tanskanen, J., Chailyan, A., Sampath, D., Heavens, D., Clissold, L., Cao, S., Chapman, B., Dai, F., Han, Y., Li, H., Li, X., Lin, C., McCooke, J.K., Tan, C., Wang, P., Wang, S., Yin, S., Zhou, G., Poland, J.A., Bellgard, M.I., Borisjuk, L., Houben, A., Dolezel, J., Ayling, S., Lonardi, S., Kersey, P., Langridge, P., Muehlbauer, G.J., Clark, M.D., Caccamo, M., Schulman, A.H., Mayer, K.F.X., Platzer, M., Close, T.J., Scholz, U., Hansson, M., Zhang, G., Braumann, I., Spannagl, M., Li, C., Waugh, R., and Stein, N. (2017). A chromosome conformation capture ordered sequence of the barley genome. Nature **544**, 427-433.

Monat, C., Padmarasu, S., Lux, T., Wicker, T., Gundlach, H., Himmelbach, A., Ens, J., Li, C., Muehlbauer, G.J., Schulman, A.H., Waugh, R., Braumann, I., Pozniak, C., Scholz, U., Mayer, K.F.X., Spannagl, M., Stein, N., and Mascher, M. (2019). TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools. Genome Biol. **20**, 284.