Title

Rapid identification of lettuce seed germination mutants by bulked segregant analysis and whole genome sequencing.

Journal

The Plant journal : for cell and molecular biology, 88(3)

ISSN

0960-7412

Authors

Huo, Heqiang
Henry, Isabelle M
Coppoolse, Eric R
et al.

Publication Date

2016-11-01

DOI

10.1111/tpj.13267

Peer reviewed

Received Date : 20-May-2016
Revised Date : 01-Jul-2016
Accepted Date : 06-Jul-2016
Article type : Original Article

Title: Rapid Identification of Lettuce Seed Germination Mutants by Bulked Segregant Analysis and Whole Genome Sequencing

Heqiang Huo[1], Isabelle M. Henry[2], Eric R. Coppoolse[3], Miriam Verhoef-Post[3], Johan W. Schut[3], Han de Rooij[3], Aat Vogelaar[3], Ronny V.L. Joosen[3], Leo Woudenberg[3], Luca Comai[2], Kent J. Bradford[1*]

[1]Seed Biotechnology Center, Department of Plant Sciences, University of California, Davis, CA, USA 95616

[2]Department of Plant Biology and Genome Center, University of California, Davis, CA, USA 95616

[3]Rijk Zwaan Breeding B.V., 2678 ZG De Lier, The Netherlands

* Corresponding author: kjbradford@ucdavis.edu

The author(s) responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Kent J. Bradford (kjbradford@ucdavis.edu).

Running title: Lettuce mutations identified by BSA sequencing

Key words: Bulked Segregant Analysis, Thermoinhibition, Seed, EMS Mutation, Whole Genome Sequencing, Lettuce (*Lactuca sativa*)

SUMMARY

Lettuce (*Lactuca sativa*) seeds exhibit thermoinhibition, or failure to complete germination when imbibed at warm temperatures. Chemical mutagenesis was employed to develop lettuce lines that exhibit germination thermotolerance. Two independent thermotolerant lettuce seed mutant lines, TG01 and TG10, were generated through ethyl methanesulfonate mutagenesis. Genetic and physiological analyses indicated that these two mutations were allelic and recessive. To identify the causal gene(s), we applied bulked segregant analysis by whole genome sequencing. For each mutant, bulked DNA samples of segregating thermotolerant (mutant) seeds were sequenced and analyzed for homozygous single-nucleotide polymorphisms. Two independent candidate mutations were identified at different physical positions in the zeaxanthin epoxidase gene (*ABSCISIC ACID DEFICIENT 1/ZEAXANTHIN EPOXIDASE*, or *ABA1/ZEP*) in TG01 and TG10. The mutation in TG01 caused an amino acid replacement, whereas the mutation in TG10 resulted in alternative mRNA splicing. Endogenous abscisic acid contents were reduced in both mutants, and expression of the *ABA1* gene from wild-type lettuce under its own promoter fully complemented the TG01 mutant. Conventional genetic mapping confirmed that the causal mutations were

located near the *ZEP/ABA1* gene, but the bulked segregant whole genome sequencing approach more efficiently identified the specific gene responsible for the phenotype.

INTRODUCTION

Lettuce (*Lactuca sativa* L.) is a valuable fresh vegetable with an annual value of production over $2.4 billion in the United States alone (NAAS 2015). The species originates from Mediterranean regions where the summers are long, hot, and dry (Lebeda *et al.,* 2004). In their natural setting, wild lettuce seedlings have the highest probability of survival when germination and emergence occur after rainfall in autumn or early spring. To avoid premature germination due to occasional rainfall during summer, lettuce seeds have evolved a high temperature-induced dormancy mechanism called thermoinhibition. When imbibition occurs at warm temperatures (above ~25 to 30°C for most commercial varieties), germination is prevented until the temperature is lowered (Argyris *et al.,* 2008; Argyris *et al.,* 2011). Major lettuce production areas in the United States include California, Arizona and Florida, which have warm autumn temperatures that can reduce field emergence and stand establishment for winter lettuce production (Cantliffe *et al.,* 2000; Valdes *et al.,* 1985).

Unlike seed thermodormancy that is induced by extended exposure of hydrated seeds to warm temperature and that inhibits seed germination even when temperature stress is removed, seed thermoinhibition is a temporary inhibition of germination that can be alleviated simply by lowering the temperature (Huo and Bradford 2015). The upper temperature limit for lettuce seed germination is genetically heritable and is influenced by environmental cues such as temperature and light during seed maturation and germination

(Argyris *et al.,* 2008; Argyris *et al.,* 2005; Contreras *et al.,* 2008; Contreras *et al.,* 2009). Molecular genetic studies have revealed that seed thermoinhibition is regulated by hormonal pathways, particularly abscisic acid (ABA), gibberellins (GA) and ethylene (Argyris *et al.,* 2008; Dong *et al.,* 2012; Huo *et al.,* 2013; Tamura *et al.,* 2006; Toh *et al.,* 2008; Yoong *et al.,* 2016). Among the hormonal regulators, ABA plays a leading role in induction and maintenance of primary seed dormancy (Bewley *et al.,* 2013). Defects in genes encoding ABA biosynthetic enzymes, including *ABA1*/*ZEP* (*ABSCISIC ACID DEFICIENT 1/ZEAXANTHIN EPOXIDASE*) (Koornneef *et al.,* 1982), *NCED*s (*9*-cis-*EPOXYCAROTENOID DIOXYGENASE*s) (Frey *et al.,* 2012; Lefebvre *et al.,* 2006), *ABA2* (*ABSCISIC ACID DEFICIENT 2*)*/GIN1*(*GLUCOSE INSENSITIVE* 1)*/SDR1*(*SHORT-CHAIN DEHYDROGENASE REDUCTASE* 1) (Gonzalez-Guzman *et al.,* 2002; Leon-Kloosterziel *et al.,* 1996) and *ABA3* (*ABSCISIC ACID DEFICIENT* 3)/*LOS5* (*LOW EXPRESSION OF OSMOTICALLY RESPONSIVE GENE 5*) (Leon-Kloosterziel *et al.,* 1996), all resulted in reduced seed dormancy, whereas disruption of ABA catabolism (e.g., defects in the CYP707A2 enzyme that inactivates ABA) led to higher levels of ABA and seed germination inhibition (Kushiro *et al.,* 2004; Okamoto *et al.,* 2006). Thermoinhibited seeds have elevated levels of ABA (Argyris *et al.,* 2008; Huo *et al.,* 2013; Toh *et al.,* 2008), suggesting that *de novo* ABA biosynthesis is induced in thermoinhibited seeds. Functional analyses demonstrated that most (but not all) of the ABA-deficient mutants exhibit germination thermotolerance due to lower endogenous ABA content (Huo and Bradford 2015). For example, *AtNCED9* in Arabidopsis or its homolog *LsNCED4* in lettuce (but not *AtNCED6* or *LsNCED2/3)* positively regulates seed thermoinhibition. Seeds of a T-DNA mutant *atnced9-1* in Arabidopsis or of a *LsNCED4*-RNAi silencing line in lettuce exhibited strong seed thermotolerance, while seeds of loss-of-function mutants of *AtNCED6* are still sensitive to high temperature during germination (Huo *et al.,* 2013; Toh *et al.,* 2008). In

addition, no phenotypic defect was observed in *LsNCED4*-RNAi silenced lines of lettuce for vegetative drought tolerance that is mainly regulated by *LsNCED2/3* (Huo *et al.,* 2013). *ABA1*/*ZEP* expression also responds to seed imbibition temperature and is associated with seed thermoinhibition (Argyris *et al.,* 2008; Chiu *et al.,* 2012; Huo *et al.,* 2013; Tamura *et al.,* 2006; Toh *et al.,* 2008). The *ABA1*/*ZEP* gene initiates the biosynthetic pathway leading to ABA by converting zeaxanthin to violaxanthin in a two-step epoxidation (Nambara and Marion-Poll 2005). Loss of function of *ABA1* in Arabidopsis caused a germination thermotolerance phenotype (Tamura *et al.,* 2006). Interestingly, seeds that are thermotolerant due to loss of function in other hormonal pathways or genetic factors generally also have altered expression of *ABA1* or *NCED*s (Cheng *et al.,* 2009; Chiu *et al.,* 2012; Gabriele *et al.,* 2010; Kim *et al.,* 2008; Piskurewicz *et al.,* 2008; Rizza *et al.,* 2011; Yoong *et al.,* 2016), further indicating that ABA indeed plays a primary role in regulating seed thermoinhibition.

To improve seed germination traits, quantitative trait locus (QTL) mapping and map-based cloning have been widely used to identify genetic loci underlying mutant phenotypes or desirable seed phenotypes (Alonso-Blanco *et al.,* 2003; Argyris *et al.,* 2005; Argyris *et al.,* 2011; Bentsink *et al.,* 2006; Gu *et al.,* 2011; Nakamura *et al.,* 2007; Osa *et al.,* 2003; Sugimoto *et al.,* 2010; Tamura *et al.,* 2006; Ye *et al.,* 2015; Yoong *et al.,* 2016). In addition to gene discovery from germplasm exhibiting natural variation, mutagenesis of commercial varieties or well-studied genotypes is widely used for crop improvement and functional analyses (Henikoff and Comai 2003; Parry *et al.,* 2009). Chemical mutagens such as ethyl methanesulfonate (EMS) are applied to induce single nucleotide polymorphisms (SNPs), a process that is simple and cost effective. However, subsequent identification and cloning of

causal mutations by conventional genetic mapping is a painstaking process, which often involves outcrossing mutants to divergent lines to generate a large mapping population followed by fine mapping. Alternatively, the combination of bulked segregant analysis (BSA) and next-generation sequencing could facilitate the identification of EMS-induced mutations corresponding to a specific phenotype (Michelmore *et al.,* 1991; Schneeberger 2014). The underlying principle of this method is that each EMS-treated mutant sample contains hundreds of specific mutations, only one or a few of which are linked to a specific observed phenotype. After backcrossing to the wild-type (non-mutated line) and then selfing, the segregating F2 seeds that exhibit a recessive mutant phenotype are selected and pooled. In the bulked DNA pool, the recessive causal EMS mutations are expected to be homozygous, whereas the non-causal mutations will segregate based on their linkage to the causal mutation. Whole genome sequencing can identify the genetic regions associated with the causal mutation by measuring allelic frequencies of these induced mutations across the genome in the phenotypically selected bulked DNA samples (Candela *et al.,* 2015; Schneeberger 2014).

Based on this principle, two main methods have been reported: 1) sequencing bulked DNA and mutation identification using crosses to polymorphic lines; and 2) sequencing bulked DNA and mutation identification using backcross isogenic lines (Candela *et al.,* 2015; Schneeberger 2014). The former method outcrosses the mutant to a polymorphic line to generate an F2 mapping population. The SNPs between the mutant and the divergent line are then used to identify homozygous genomic regions for mutation identification (Candela *et al.,* 2015; Schneeberger 2014). Because it provides ample polymorphic SNPs markers, this method can allow mutant phenotypes to be associated with small genetic intervals and to

eventually identify the candidate mutations. A drawback of this method is that the introduction of a distinct genetic background into the mutant background can alter the mutant phenotype, making the precise phenotyping of recombinant mutants challenging (Schneeberger 2014). In addition, the mapping power of this method depends primarily on the number of recombinants, although the depth of sequence coverage also had a small effect on the mapping intervals (James *et al.,* 2013); it therefore can require large numbers of F2 individuals to be generated and phenotyped in order to fine map the causal mutation. The second method is to backcross the mutant to the non-mutagenized parental wild- type, which can effectively remove background mutations to overcome the genetic complications due to outcrossing to a divergent polymorphic line and simplify the subsequent breeding process (Abe *et al.,* 2012; Candela *et al.,* 2015; Schneeberger 2014). Although the size of backcross (BC) mapping populations did not greatly affect the mapping by this method, mapping by sequencing using a backcross population is strongly affected by the sequence coverage owing to limited polymorphic markers, which can make it more difficult to unequivocally identify the causal mutation (Abe *et al.,* 2012; James *et al.,* 2013; Schneeberger 2014). Ultra-deep sequencing of the candidate region may be required to pinpoint the true causal mutation (Hartwig *et al.,* 2012).

In this study, we mutagenized the lettuce cultivar Yorvik and screened for seed germination capacity at high temperature. Genetic and physiological analyses revealed that two Yorvik EMS mutants (TG01 and TG10) derived from independent mutagenesis events were allelic, which enabled us to conduct separate bulked segregant analyses using two BC1/F2 populations and sequence them in parallel to simultaneously identify the causal mutation. As shown by Nordstrom *et al.* (2013), identification of the same locus for two independent

mutant alleles gives high probability of correct gene identification. BSA by sequencing and conventional genetic mapping both mapped the mutant germination phenotype to the same region on chromosome 3. The candidate gene was precisely identified, as mutations occurred in the same gene but at different positions in the two EMS-induced mutant lines. Functional analyses revealed that single base-pair mutations in *LsABA1*/*ZEP* caused the seed germination thermotolerance phenotype in both TG01 and TG10. Thus, we demonstrate that BSA by sequencing provides an efficient forward genetics method to identify mutations underlying specific phenotypes and can be applied in a large crop genome (2.7 Gb).

RESULTS

Temperature sensitivities of four lettuce genotypes

Four lettuce cultivars, Apache, Yorvik, Senai and Troubadour, were initially used for EMS mutagenesis to screen for different phenotypic traits, including seed high temperature germination (HTG) or thermotolerance. The four lettuce genotypes exhibited distinct temperature sensitivities during seed germination. Seeds of all cultivars germinated well at 25 and 30°C (all germination tests were conducted in light unless noted otherwise), suggesting that they are more tolerant to temperature compared to the previously reported temperature-sensitive genotype Salinas in which germination is inhibited at 30°C (**Figure 1a**) (Argyris *et al.,* 2008; Argyris *et al.,* 2011; Huo *et al.,* 2013; Yoong *et al.,* 2016). However, only Apache and Yorvik seeds exhibited sensitivity to 32 and 35°C, whereas Troubadour and Sensai seeds could fully germinate at 32°C and were relatively tolerant to 35°C. Among the

four tested lettuce cultivars, Yorvik was the most sensitive to high temperature, followed by Apache (**Figure 1a**).

Germination responses of Apache and Yorvik EMS mutants to diverse conditions

Two Yorvik mutants (TG01 and TG10) and two Apache mutants (TG09 and ETR6) derived from four independent EMS mutagenesis events were further screened for seed thermotolerance. We first determined the median ceiling temperature for each mutant as the T_c50, or the temperature at which the germination percentage is reduced to 50%. TG01 and TG10 mutant seeds had the same T_c50 (36°C) for germination, while the ETR6 mutant was more tolerant to high temperature than the TG09 mutant (**Figure 1b**). Seeds of all four mutants could germinate fully at 35°C (**Figure 1c,d).**

Next, we examined how these mutant seeds responded to different stresses and hormones at 35°C, which totally inhibited germination of Yorvik and Apache wild-type seeds. To compare their responses to known HTG genotypes, we used *NCED4*-RNAi Salinas (RNAi) and US96UC23 (US96) as controls. As mentioned above, *LsNCED4* was identified through genetic mapping using a recombinant inbred population derived from the thermosensitive genotype Salinas and thermotolerant genotype US96 (Argyris *et al.,* 2008). RNAi is a thermotolerant line in which *LsNCED4* is suppressed by RNA interference in the Salinas cultivar, and both RNAi and US96 seeds can germinate well at 35°C in the light (Huo *et al.,* 2013). First, we tested how water stress affects germination of these mutants at high temperature (35°C). Seeds of the two Apache mutants were highly tolerant to moderate water stress (-0.5 MPa PEG8000), and although seeds of the two Yorvik mutants could complete germination, their

germination rate (speed) was much slower than for the Apache mutants; RNAi and US96 seeds were the most sensitive to water stress at this elevated temperature (**Figure 2a**). Under salt stress (150 mM NaCl; -0.74 MPa) at 35°C, ETR6 seeds germinated slightly faster than TG09, TG01 and TG10 seeds; again, the RNAi and US96 seeds were the most sensitive genotypes (**Figure 2b**). We next tested the response of different genotypes to paclobutrazol (PAC, a GA biosynthesis inhibitor), silver thiosulfate (STS, an ethylene action inhibitor) and ABA. With 100 µM PAC treatment, germination of US96 seeds was almost completely inhibited at 35°C; by contrast, seeds of the Apache and Yorvik mutants exhibited greater tolerance to PAC, consistent with their ability to germinate in the dark (**Figure S1**). Among all tested genotypes, only US96 seeds were unable to germinate in the dark (**Figure S1**), indicating that light stimulation of GA biosynthesis is essential for US96 seed germination (Argyris *et al.,* 2008). RNAi seeds exhibited tolerance to PAC, but with a slower germination rate (**Figure 2c**). When STS was applied to all genotypes, germination of US96 and RNAi seeds was greatly inhibited and Yorvik EMS mutants germinated better than Apache EMS mutants, although all four lettuce EMS mutants displayed strong tolerance to the ethylene perception inhibitor (**Figure 2d**).

Although the response patterns of the two Yorvik mutants and of the two Apache mutants to PEG, NaCl, PAC and STS treatment were very similar within each cultivar, distinct patterns were observed between the two Apache mutants in response to ABA treatments (**Figure 2e,f**). TG09 seeds exhibited greater sensitivity to 1 µM ABA than did ETR6 seeds (**Figure 2e**). This distinct response is even more apparent at 3 µM ABA, at which only 8% of TG09 seeds germinated while ETR6 seeds still germinated completely (**Figure 2f**). This result suggested that the causal mutation(s) in ETR6 might be different from the one(s) in TG09. On the other

hand, seeds of the two Yorvik EMS mutants displayed essentially identical responses to 1 μM and 3 μM ABA treatments (**Figure 2e,f**). RNAi seeds were fully inhibited by 1 μM ABA treatment (**Figure 2e**), indicating that the lower endogenous ABA content due to silencing of *LsNCED4* can be complemented by application of exogenous ABA. Based on these results, in which the two Yorvik mutants exhibited similar physiological responses to all treatments, we concluded that the causal mutations in these two mutants may have occurred in the same gene or regulatory pathway.

Genetic analysis of the Yorvik mutants

Since EMS can generate large numbers of random mutations that can cause defects in multiple pathways controlling seed germination, we next examined whether the phenotypically similar TG01 and TG10 mutants are due to a single allele. Germination tests on BC/F2 seeds showed that around 25% of seeds of the two Yorvik mutants could germinate at 35°C (**Figure S2a**), indicating that the causal mutations are recessive. *Chi*-square tests also confirmed that the germination percentages of Yorvik mutants fit the expectations of Mendelian segregation (P = 0.301 and P = 0.135 for TG01 and TG10), indicating that the mutant phenotype is attributed to a single locus in each case (**Figure S2a**). We crossed TG01 and TG10 with Yorvik wild-type (♂) and crossed TG10 (♀) with TG01 (♂). Germination of F1 seeds of TG01×Yorvik and of TG10×Yorvik were strongly inhibited by high temperature, confirming that these causal mutations are recessive (**Figure S2b**). Approximately 5% of F1 seeds could germinate at 35°C, possibly due to the presence of contaminant selfed seeds. In contrast, the TG10×TG01 F1 seeds germinated as well as their parents, suggesting that TG01 and TG10 are allelic (**Figure S2b)**. Given that these two Yorvik

mutants had been generated from independent mutagenesis events, the probability that both mutations result from the same base pair change is extremely low. Because pre-existing SNPs between Yorvik and Salinas cultivars may hinder the identification of EMS-induced mutations (Salinas is the reference lettuce genome sequence; Truco *et al.,* 2013), we then designed a parallel BSA whole genome sequencing strategy to identify the causal EMS mutations by comparing TG01 and TG10 sequences.

Parallel sequencing to identify the EMS mutations in allelic lettuce mutants

As described above, two mutant identification strategies have been described: mapping using crosses to polymorphic lines (i.e., SHOREmap; Schneeberger *et al.,* 2009), and mapping and mutation identification using backcross isogenic lines (i.e., MutMap; Abe *et al.,* 2012). Although the SHOREmap method can generate a clear sequence map with the candidate causal mutations at the peak of a mapped region, it requires large F2 populations and a reliable genome sequence of a distantly related wild accession of the same species (Schneeberger *et al.,* 2009). In our case, the genome of the distantly related lettuce US96 genotype is available, but it is highly tolerant to high temperature at seed germination (Argyris *et al.,* 2008; Argyris *et al.,* 2011; Huo *et al.,* 2013), making it impossible to phenotype the F2 homozygous mutant based on germination thermotolerance. In addition, crossing to the primitive US96 will not facilitate generation of horticulturally acceptable genetic material for subsequent breeding. On the other hand, natural SNPs between Yorvik and the Salinas reference genome sequence might generate noise that would compromise the power of mutation identification using this method. Thus, it could be hard to pinpoint the real candidate causal mutations without also sequencing the wild-type Yorvik genome.

Since TG01 and TG10 are allelic and derived from independent mutagenesis events, we opted for parallel sequencing of the two mutants to identify the causal mutations with high efficiency. First, the two mutants were backcrossed to their wild-type and the F2 seeds (from selfing of F1) were used for phenotyping (**Figure 3**). Because of the recessive nature of these mutations, only the homozygous mutant seeds can germinate at 35°C. Leaf tissues were pooled from these selected mutant seedlings to make separate TG01 and TG10 bulks from which to extract DNA for whole genome sequencing. Sequencing reads of TG01 and TG10 bulked DNA were mapped to the lettuce reference genome and positions that were polymorphic between the two bulks were identified as potential mutations. To reduce noise, only the canonical EMS mutations (C→T and G→A) were retained for further analysis (Henry *et al.,* 2014). Because the two bulked DNA samples are from phenotypically selected recessive mutant seedlings, the causal mutation was expected to be homozygous across all reads. Thus, a region with high mutant allele percentages should be evident in both bulks since the causal mutations in both TG10 and TG01 should be present within the same gene or nearby genetic loci. The non-EMS induced SNPs between Salinas and Yorvik could be easily identified and removed, even though these SNPs might be closely linked with the causal candidate EMS mutations (**Figure 3**). This strategy avoided the additional sequencing of the Yorvik wild-type genome and of the non-selected bulks for each genotype and enabled identification of the mutations in both mutants simultaneously.

Identification of genomic regions with mutant alleles

We performed bulked segregant analysis as follows. BC1/F2 TG01 and TG10 seeds were tested for germination at 35±0.5°C. Next, 72 TG01 and 48 TG10 seeds that germinated under those conditions were selected and allowed to grow into seedlings. DNA was extracted from leaf samples from the two mutant pools, and each pool was sequenced on two Illumina HiSeq 2000 lanes. Sequencing reads were aligned to the lettuce genome (v4.0) and sequence data from the two mutant pools were compared to identify mutations specific to each.

First, all positions that exhibited different allele calls between the two mutant pools were identified. Specifically, positions that were homozygous but different in the two mutant pools and positions for which one of the two mutant pools was heterozygous and carried an alternative allele in addition to an allele common to the two mutants were identified (see Methods). The positions that were polymorphic between the two mutant pools were enriched in the CG>TA canonical EMS mutations (**Figure S3a,b**), as expected from a segregating population of BC1/F2 plants each originating from a single EMS-induced mutant plant. In contrast, positions that were heterozygous in both mutant pools were not enriched in CG>TA transitions, as expected for natural variants (**Figure S3c**). The canonical mutations represented 65% and 64% of the heterozygous positions in TG10 and TG01, respectively, while they only represented 28% of the positions heterozygous in both mutants (**Figure S3**).

To identify the mutation(s) responsible for the germination thermotolerance phenotype or mutations closely linked to the causal mutations, we first only retained mutations that were consistent with EMS action (CG>TA transitions). Next, we preferentially selected mutations that were either homozygous or heterozygous but highly represented in the mutant bulks

("high heterozygous mutations", see Methods). A total of only 11 homozygous mutations (two for TG01 and nine for TG10) were detected, all located within 2 Mb of each other, on linkage group 3 (**Table S1 and Figure 4a,b**). To confirm the potential locations of the causal mutations, heterozygous mutations were utilized to identify 591 and 494 positions with high percentage mutant alleles for the TG10 and TG01 bulked mutants, respectively. To more easily visualize the trends in mutant allele percentages, the genome was partitioned into non-overlapping, consecutive 1 Mb bins and for each bin, the percentage of mutant alleles was averaged across all mutations detected in that bin. A single peak at the same location was identified for each bulk (**Figure 4a, b**), confirming linkage group 3 (LG3) as the best candidate for containing the causal mutation.

We next searched the list of candidate positions obtained within ~30 Mb on linkage group 3 for potential causal mutations (**Table S1**). Within this interval, we identified 28 heterozygous and 9 homozygous mutations for the TG10 mutant pool and 36 heterozygous and 2 homozygous mutations for the TG01 mutant pool. After testing seed germination of progenies from the individual plants that were bulked, we found that two plants had been mis-scored in the TG01 bulk and probably contributed to the wild-type alleles observed in the mutant bulks. Next, we searched for mutations affecting the same gene in each bulk because complementation analysis showed that the TG10 was allelic to TG01 (**Figure S2b**). Two mutations affected the same gene but at different locations, only 473 bp apart. We identified a G→A mutation at position 172,545,169 in the TG01 mutant pool and a second G →A mutation at position 172,544,696 in the TG10 mutant pool. Further examination of this result showed that the first G→A mutation in TG01 mutant occurred within the 8th exon of

the *ABA1/ZEP* gene and caused an amino acid change from glycine to aspartic acid (**Figure 5a; Figure S4**), while the second G→A mutation in TG10 mutant was located at the end of the 6th intron of *ABA1/ZEP* and likely caused mis-splicing of the mRNA (**Figure 5a**).

EMS mutation in TG10 causes mRNA mis-splicing

Given the position of the TG10 mutation adjacent to an intron (**Figure 5a**), we investigated whether it could cause alternative splicing or mis-splicing of the *ABA1/ZEP* transcript. We amplified the cDNA fragment from exon 4 to exon 16 with the primer set 1f/1r (**Figure 5a**). After gel-purification, cloning and sequencing, we observed a splicing variant with only 1 bp deletion in the TG10 *LsABA1* gene sequence that predicted an early stop codon and a truncated protein (**Figure 5b; Figure S4**). Results from multiple reverse transcription/PCR amplification and sequencing further confirmed that this 1 bp deletion variant was not due to a PCR amplification error. In addition, another splicing variant with a 39 nt unspliced intron was detected, which resulted in 13 additional amino acids (**Figure 5b; Figure S4)**. We next examined whether mutations in TG01 and TG10 affected *LsABA1* mRNA level. Using the primer set 2f/2r that binds to the 1st exon of *LsABA1* (**Figure 5a**), qPCR results showed that there is no significant difference in the abundance of *LsABA1* transcripts in dry seeds of different tested genotypes (**Figure 5c**). We also performed PCR using the primer set 3f/3r that binds the 14th and 15th exons (**Figure 5a**), and no significant differences were observed among the three tested lettuce genotypes (**Figure 5c**). In primer set 4f/4r, the SNP deletion in one cDNA variant is positioned at the 3' end of primer 4f, and it also spans the 6th and 7th exons; therefore it would not amplify a transcript variant with an indel or with an intervening intron. The relative expression of *LsABA1* in TG10 is only ~8.5% of that in Yorvik

wild-type, suggesting that the normal splicing of *LsABA1* transcripts was largely disrupted (**Figure 5c**).

Genetic mapping identified the same genomic region controlling high temperature germination

In parallel with the BSA method, we also conducted conventional genetic mapping of the mutant loci. For example, one segregating mapping population was developed from TG01 by crossing with the temperature-sensitive cultivar Apache. Genetic analysis using SNP markers (see Materials and Methods for details) found strong linkage on chromosome 3 at marker LS00899, with a 2 LOD drop interval spanning the region between markers LS00484 and LS00913, a genetic interval of 24.4 cM (~75.3 Mb physical region) (**Figure 6a; Figure S5a,b**). The detected linkage represented a recessive allele conferring the HTG trait. Further fine mapping of the candidate locus on linkage group 3 via F3 populations and recombinants (see Materials and Methods) indicated that the causal mutation was located between markers LS01045 and LS01079 (**Figures S5a,b**). This linkage analysis, combined with the phenotypic data (**Figure S4a; Table S2**), delimited the candidate area to a roughly 6.5 Mb area between markers Ls01079 and Ls00995 (**Figure 6b**). This region comprised 96 predicted gene sequences, 77 of which were associated with an annotation, including *LsABA1* (**Table S3**). Thus, the mapping results strongly supported the identification of the causal gene based on the BSA by parallel sequencing approach.

Validation of the identified EMS mutation in lettuce and Arabidopsis

Although we identified mutations at the same gene in both TG01 and TG10, we also observed other mutations under the peak and closely linked to the mutations in *LsABA1* (**Table S1**). This raised the possibility that the HTG phenotype may be caused by these mutations instead. We molecularly complemented the TG01 mutant with the wild-type *ABA1* gene from cv. Salinas including its own 2.3 kb promoter. F2 seeds segregating with the transgene (Pro*LsABA1*::*LsABA1*) were tested for their sensitivity to high temperature. Results showed that the ectopic expression of *LsABA1* in the TG01 mutant restored its thermosensitivity during germination (**Figure 7a**). For example, only 23%, 18% and 26% of seeds of transgenic lines *LsABA1OX*#1, #5 and #7, respectively, germinated at 35°C while the TG01 mutant seeds could fully germinate and Yorvik seeds were fully inhibited at 35°C. Genotyping results of the non-germinated seeds indicated that only 6 out of 188 (3.19%) non-germinated seeds from 3 different transgenic lines do not contain a transgene, suggesting that the thermotolerant phenotype of TG01 is caused by the single mutation in *LsABA1*. We next isolated the mutated *LsABA1* from TG01 for molecular complementation of the Arabidopsis *aba1-1* mutant to examine whether the single amino acid change caused a defect in *LsABA1* function. As reported previously (Tamura *et al.,* 2006), *aba1-1* seeds exhibited strong thermotolerance, and overexpression of wild-type *LsABA1* (under 35S promoter) fully restored the thermosensitivity of Arabidopsis *aba1-1* mutant seeds at both 32 and 34°C (**Figure 7b**). However, overexpression of the mutated gene (m*LsABA1*) from TG01 was not able to restore thermosensitivity of *aba1-1* seeds (**Figure 7b**). The mutated *LsABA1* from TG01 may have retained some enzyme activity since its overexpression could partially inhibit germination of *aba1-1* at 34°C (**Figure 7b**).

Mutation of *LsABA1* lowers endogenous ABA content

Given the known role of zeaxanthin epoxidase (encoded by *ABA1*) in ABA biosynthesis, we would predict that the TG01 mutation identified would reduce the activity of the expressed enzyme. We predicted the tertiary structures of LsABA1 of TG01 and Yorvik (**Figures S6; S7a, b;** complete views available in **Data S1** and **S2**, respectively). The replacement of glycine with aspartic acid in LsABA1 of TG01 caused significant changes in its tertiary structure. For example, compared to the Yorvik wild-type LsABA1, formation of several short additional helixes in TG01 LsABA1 was predicted due to this amino acid replacement (**Figures S6; S7a,b**). The effect of these changes was evident in the ABA contents of the dry seeds and high temperature-imbibed seeds of Yorvik, TG01, TG10 and Salinas. Seeds of Salinas and Yorvik were both sensitive to high temperature (**Figure 1a**) and also contained comparable levels of ABA in dry seeds (**Table S4**). ABA contents of dry seeds of TG01 and TG10 were only ~40% of that in Yorvik dry seeds (**Table S4)**. ABA contents decreased after 6 and 12 h of imbibition at 35°C in all genotypes, but to a greater extent in mutant seeds; after 12 h of imbibition, ABA contents in mutant seeds were only about 10% of that in wild-type seeds (**Table S4**). In addition, we observed pre-harvest sprouting of *LsABA1* mutant seeds (**Figure S8a-c**), as also occurs in Arabidopsis *aba1* mutants (Karssen *et al.,* 1983). Germination was observed prior to full seed maturation (**Figure S8a,c**). Seedlings with fully developed leaves were found growing vigorously in flower buds (**Figure S8b**). Although pre-harvest sprouting may occur occasionally in lettuce under warm temperatures and high humidity, this phenomenon was much more prevalent in both TG01 and TG10 mutants than in wild-type developing seeds under the same conditions.

DISCUSSION

Forward genetic approaches in crop breeding have long been hindered by the fact that the mutations underlying mutant phenotypes could not be detected rapidly using conventional mapping strategies. The conventional method requires a large mapping population that is commonly constructed by outcrossing the mutant to a line that has a polymorphic genetic background compared to the genetic background of the mutant, without causing suppression of recombination due to genetic incompatibility between both crossing partners. Although the polymorphism between the mutant and the divergent genetic line could increase the mapping power owing to the greater saturation of the map with polymorphic markers, developing a mapping population and genotyping of the recombinant lines are labor-intensive, time-consuming and costly, especially for non-model crop species that lack molecular marker information and have relatively long life cycles. The recent advances in next generation sequencing, combined with BSA (Michelmore *et al.,* 1991), greatly facilitate mapping and identification of EMS mutations. The strategy of mapping by sequencing of bulked recombinant lines derived from outcrossing of mutants to a divergent genetic line was first reported by Schneeberger *et al* (2009). Although this method was able to precisely detect the causal mutation for the mutant phenotype, a large mapping population was required for mapping a small DNA interval, which could hamper the application of this method to crop species requiring greater space and resources for production. Unlike the model plant species for which reference genomes of multiple genotypes are available, the genome sequences of divergent lettuce genotypes are limited; therefore, additional sequencing of this divergent genotype might be required for mapping and identification of mutations by BSA sequencing. In addition, the genetic interaction

between the mutant and the divergent line might alter the mutant phenotype and result in mis-scoring.

In contrast to the outcrossing/BSA sequencing strategy, backcrossing of a mutant line to the wild-type from which it is derived would not introduce a divergent background, which could expedite the production of breeding germplasm incorporating the selected mutant. In addition, the size of a backcrossing mapping population has less influence on the mapping power compared to an outcrossing mapping population (James *et al.,* 2013). For example, Abe *et al* (2012) and Takagi *et al* (2015) used as few as 20 BC1/F2 individuals for each bulked sample to successfully identify the EMS mutations underlying two phenotypes. In another study, bulked samples of 30 EMS BC1/F2 or γ-irradiated BC1/F2 sorghum seedlings were used to detect mutations that caused reduced density of leaf veins (Rizal *et al.,* 2015). In the present study, bulked DNA from 48 BC1/F2 seedlings of TG10 was sufficient to detect the causal mutations. Simulation tests showed that a bulked sample size of 20 is sufficient for detection of a useful mapping interval at a sequencing coverage of >10× (Abe *et al.,* 2012; James *et al.,* 2013). In contrast to the sample size, the inclusion of mis-scored plants in bulked samples could result in severe reduction in detection power since the mis-scored plants introduce wild-type alleles at the causal mutation locus (Abe *et al.,* 2012; James *et al.,* 2013). For example, the two mis-scored TG01 plants (2.78% of 72 sampled individuals) in one of our bulks resulted in not only a reduction in the number of detected homozygous mutations, but also caused altered allele frequency at the true causal mutation locus (*ABA1/ZEP*) (**Table S1**). The results from BSA sequencing results suggest that precisely phenotyped samples even with a small number of individuals would generate better output than larger bulked samples with a small percentage of mis-phenotyped individuals.

Mapping and identification of EMS mutations through backcrossing and BSA sequencing utilizes the SNPs generated by mutagenesis as markers to locate the region harboring the causal mutation. Although backcrossing can efficiently remove non-causal mutations and simplify the genetic background for phenotyping and sequence analysis, the mapping power also decreases due to lack of polymorphic markers; therefore a large mapping interval containing multiple candidate mutations would be expected, especially when sequence coverage is not deep due to a large genome size (Abe *et al.,* 2012; Ashelford *et al.,* 2011; Hartwig *et al.,* 2012; James *et al.,* 2013; Rizal *et al.,* 2015). For example, the mapping interval for TG10 in this study is relatively large (>30Mb) and contains more than 35 EMS candidate mutations, of which nine are homozygous mutations in addition to *ABA1/ZEP* (**Table S1**). Further effort such as deep candidate re-sequencing (dCARE) would be required to pinpoint the causal mutation from a strictly mapping-based approach (Hartwig *et al.,* 2012). However, the strategy developed in this study takes advantage of the hypothesis that two allelic mutations in TG01 and TG10 occurred at different positions within the same gene. As demonstrated in our sequencing results, two mutations happened at different loci within the *ABA1/ZEP* gene in TG01 and TG10, respectively. This strategy enabled us to precisely pinpoint the mutation without any further sequencing or mapping effort, in agreement with the results of (Nordstrom *et al.,* 2013), who used a similar strategy to identify two allelic mutants in *Arabis alpina*.

As the genome of lettuce is relatively large (~2.7 Gb), sequencing of only two bulked DNA samples containing distinct but allelic mutants not only improves the power of mapping-by-sequencing, but also reduces costs through avoiding the need for sequencing the wild-type parent or the non-selected phenotypic bulk (**Figure S9**). Our strategy might be extended to

the case of two allelic mutants with different genetic backgrounds. For example, two allelic sorghum mutants generated by EMS or γ-irradiation could have been subjected to our sequencing strategy to identify the mutation without requiring sequencing of the two parental wild-types (Rizal *et al.,* 2015). As demonstrated here, our approach identified the same genomic region as conventional mapping with much less time and effort than was required for population construction, phenotyping and genotyping (**Figure S9**). The fine-mapping interval could only be delimited to 6.5 Mb that harbors more than 70 candidate genes (**Table S3**), probably due to the limited numbers of polymorphic markers between Yorvik and Apache. Our results show that conducting forward genetics by BSA sequencing in parallel is highly cost-efficient and rapid compared to conventional mapping (**Figure S9**). On the other hand, there are limitations to the method, as it requires mutants with similar phenotypes from two or more independent mutagenesis events and crosses for allelism tests. Thus, considerable characterization of the mutants is required to determine whether this approach is suitable.

Sequencing of species with large genomes, such as wheat, poses additional challenges, such as the higher cost of sequencing, although the rapidly decreasing costs make it increasingly affordable. Recently, mRNA sequencing or exome-capture sequencing were applied to crop species to reduce sequencing costs and genomic complexity (Henry *et al.,* 2014; Liu *et al.,* 2012; Pankin *et al.,* 2014). However, these targeted-enrichment sequencing strategies could exclude the causal mutations from the sequencing data, such as for mutations in gene regulatory regions (James *et al.,* 2013; Schneeberger 2014). For example, the mutation at the 6th intron of *ABA1/ZEP* in TG10 might not have been detected by using the exome-capture sequencing method.

ABA plays a pivotal role in regulating seed dormancy, and its biosynthesis is required for the induction of primary dormancy during seed development (Bewley *et al.,* 2013). *ABA1/ZEP* catalyzes conversion of zeaxanthin to violaxanthin in a two-step epoxidation that is the first key step in ABA biosynthesis (Nambara and Marion-Poll 2005). The *ABA1/ZEP* locus was first identified in an EMS-induced forward genetic screen for revertants in non-germinating gibberellin-deficient mutant lines and resulted in loss of seed dormancy (Koornneef *et al.,* 1982). It has been repeatedly identified in subsequent induced-mutation screens in multiple species based on germination or loss of dormancy (Nambara and Marion-Poll 2005). In lettuce and Arabidopsis, expression of *ABA1/ZEP* was upregulated by high temperature during seed imbibition (Argyris *et al.,* 2008; Chiu *et al.,* 2012; Huo *et al.,* 2013; Toh *et al.,* 2008); functional analyses showed that Arabidopsis *aba1-1* mutant seeds exhibit a greater tolerance to high germination temperature (**Figure 7b**) (Tamura *et al.,* 2006). Our present study also clearly demonstrated that mutation in *ABA1/ZEP* improved lettuce seed germination at high temperature (**Figure 1**). Similar to Arabidopsis, there is only one functional *ABA1/ZEP* gene in the lettuce genome, although one pseudogene is present. However, in contrast to Arabidopsis, in which wilting and other vegetative development phenotypes were evident in *aba1* mutant plants (Barrero *et al.,* 2005; Koornneef *et al.,* 1982), no obvious phenotypes were observed in lettuce, other than viviparous germination when grown in high humidity (**Figure S8**). This is probably due to either residual enzyme activity in TG01 or partially normally spliced mRNAs in TG10.

METHODS

Plant materials

Seeds of lettuce cultivars Apache, Yorvik, Sensai and Troubadour were subjected to EMS mutagenesis by Rijk Zwaan BV, The Netherlands (Woudenberg *et al.,* 2014). Germination screens at high temperature (above 30°C in darkness) identified more than 80 homozygous HTG mutants, of which eight lines consistently showed the largest improvement in HTG in darkness, as compared to their wild-type controls. Two lines (Yorvik TG01 and TG10) were analyzed in parallel by BSA sequencing, and one of them (TG01) was used for characterization by classical map-based cloning. For DNA sampling for BSA, lettuce plants were grown in a greenhouse in long-day conditions (16 h light/8 h dark) at Davis, California. Arabidopsis mutants *aba1-1* and L*er* were purchased from ABRC (www.arabidopsis.org). Arabidopsis plants were grown in a growth chamber under long days (16 h light/8 h dark) at 22°C.

Seed germination tests

All lettuce seed germination tests to characterize mutant lines or for BSA selection were conducted in 4.7-cm Petri dishes with one blotter paper that was moistened with 3 mL of deionized water or solutions at 35°C in the light. For Arabidopsis seed germination tests, seeds were placed in a 24-well plate with each well filled with 400 µL of deionized water. The plates were sealed with Parafilm. The deionized water or solutions were supplemented with 0.2% plant preservative mixture (PPM, Caisson Labs, USA) to prevent fungi

development. Three replications of 30 lettuce seeds or Arabidopsis seeds were utilized in all tests.

Selection of segregants and preparation of genomic DNA bulks for sequencing

Lettuce M5 mutants of TG01 and TG10 were backcrossed to Yorvik wild-type and progeny were grown in a greenhouse and allowed to self to produce F2 seeds. The seeds harvested from each individual F2 plant were first germinated at high temperature for segregation screening to confirm the backcross success. The BC1/F2 seeds were carefully screened for thermotolerance during germination at 35±0.5°C in the light. Seedlings were grown from 72 (TG01) and 48 (TG10) BC1/F2 thermotolerant seeds. Young leaves were pooled from 1-week-old seedlings and mixed for DNA extraction. All seedlings were grown to maturity to produce F3 seeds for further checking whether their parents were homozygous. High-quality genomic DNA was extracted using CTAB methods described by (Stoffel *et al.,* (2012)). The DNA purity and integrity were analyzed using an Agilent 2100 Bioanalyzer (Agilent Technologies, USA) prior to shipping to the Beijing Genomics Institute (BGI, Hong Kong, China) for DNA library construction and whole genome sequencing. Paired-end sequencing was performed on both samples in two Illumina HiSeq 2000 lanes.

Bulked segregant analysis

A total of 324 million and 297 million paired-end sequence reads were obtained from the TG01 and TG10 mutant pools, respectively. Raw sequencing reads were processed for quality and adapter contamination and demultiplexed based on index sequence using

custom Python scripts available at http://comailab.genomecenter.ucdavis.edu/index.php/Barcoded_data_ preparation_tools and as described previously (Henry *et al.,* 2015). Processed reads were aligned to the current version of the lettuce genome (v4.0; http://cgpdb.ucdavis.edu/database) using BWA and default parameters (Li and Durbin 2009). At the same time, clonal reads were excluded from further analysis using the –o option in a custom Python script available at http://comailab.genomecenter.ucdavis.edu/index.php/Bwa-doall. Sequence information in the two bulked mutant samples were compared to each other by creating a single mpileup file using samtools (Li *et al.,* 2009). Next, the mpileup file was parsed to extract the relevant information using a custom Python script available at http://comailab.genomecenter.ucdavis.edu/ index.php/Mpileup. Finally, only positions that were covered at least 10 times in both mutant pools were retained. This final file, containing information from ~945 Mb of the reference sequence, was used as the starting point for all subsequent mutation detection steps.

Next, positions were categorized depending on allele calls in the two mutant pools using custom Python scripts as follows: i) if the position was homozygous (one allele contributed >95% of the coverage) in both mutant pools but for a different allele, the position was called "homozygous mutant" for the mutant pool that exhibited an allele call different from the reference sequence; ii) positions that were heterozygous in both mutant pools (allele percentages between 40 and 60% for both alleles) were labeled "natural heterozygous position"; iii) positions that were heterozygous in one mutant but not the other were handled as follows. The allele that was common to the other mutant pool was referred to as the WT allele and the other allele was designated the mutant allele. If between 50 and 60%

of the coverage could be attributed to the mutant allele, that position was called "heterozygous mutation" for the mutant pool that exhibited that allele. If more than 60% of the coverage could be attributed to the mutant alleles, that position was called "high heterozygous mutation" for the mutant that exhibited that allele. All mutant positions were used to determine the percentage of EMS canonical mutations observed in the mutant pools (**Figure S3a, b**). All "natural heterozygous positions" (SNPs) were used for **Figure S3c**. Only the homozygous and high heterozygous mutations were used to identify the potential location of the causal mutation (**Figure 4a,b; Table S1**).

RNA splicing sequencing and gene expression analyses

Total RNA was isolated from dry seeds of lettuce using PureLink® Plant RNA Reagent following the instruction manual (Life Technologies, USA) or using the method previously described (Argyris *et al.,* 2008). Total RNAs were treated with *TURBO* DNAase I (Life Technologies, USA) prior to cDNA biosynthesis. For examination of mis-splicing of RNA, SuperScript III or IV First-Strand Synthesis kit (Thermo Fisher Scientific, USA) was used for cDNA biosynthesis with oligo-dt or random hexamer primers. Primer set 1f/1r was used for amplification of cDNA fragments. Amplified fragments were either gel-purified for sequencing or sub-cloned using pGEMT-Easy (Promega, USA) for sequencing. For mRNA quantitation, 1 µg of total RNA was used for cDNA synthesis using QuantiTect Reverse Transcription Kit (Qiagen, USA) and the real-time PCR was performed as described previously (Huo *et al.,* 2013).

Genetic mapping of the HTG trait in lettuce

A mapping population containing 100 plants was made by crossing the TG01 homozygous M5 HTG mutant to a contrasting, non-mutant cultivar Apache. The resulting 73 F3 families, harvested from individual F2 plants, were characterized by sowing 100 seeds at 32°C in the dark to differentiate among the homozygous HTG mutant families, the homozygous wild-type families and the heterozygous F3 families. DNA samples from about 30 seeds of only the homozygous HTG and homozygous wild-type plants were analyzed with a random set of proprietary SNP markers having good representation across the genome, which showed that approximately 50% of the markers were polymorphic. For the Yorvik TG01 mutant, coarse linkage mapping was performed on a small F2 population (33 individuals; 13 homozygous HTG, 20 homozygous wild-type) by crossing to the temperature-sensitive cultivar Apache. An additional set of proprietary SNP markers located in the selected interval were run on the same DNA samples as the first screening, which confirmed the coarse mapping results. These markers were used to identify F3 plants with recombination events at HTG candidate area within a selection of 27 heterozygous families that were grown in a greenhouse to create F3 families. In total, 706 F3 plants were genotyped (between 10 to 46 individual F3 plants per family) with the additional markers. In the linkage analysis of each F3 family, the non-segregating marker data were removed, resulting in a marker segregation of 1:1:2 (A:B:H; A allele: B allele: Heterozygous) as expected, as the F3 plants were created from F2 plants which were all selected to be heterozygous at the region of interest. Marker phase correction was applied by using the parental Yorvik line. Map distances were calculated using JoinMap 4.1 (www.kyazma.nl) (Stam 1993). After maximum likelihood mapping was performed to create a starting order, this order was subsequently used for regression

mapping using the Kosambi's mapping function. A subset of 197 F3 plants, selected from the 706 F3 plants, was genotyped with additional markers that were physically located between markers LS00484 and LS00913 (**Table S2**). The F4 families of the 197 individual F3 plants were used for phenotyping by germination at 32°C in darkness. Single-marker ANOVA was performed and showed the tightest linkage between HTG and markers LS00997, LS01045 and LS00995 (**Figure S5a,b**). The linkage pattern suggests that the causal mutation is located between markers LS00995 and LS01079 (**Figure.6b**).

Genetic and molecular complementation of lettuce and Arabidopsis *aba1* mutants

For genetic complementation of lettuce TG01 and TG10 mutants, Yorvik wild-type was used as male to pollinate both Yorvik mutants. The F1 seeds were tested for germination at 25 and 35°C in the light. Seeds failing to germinate at 35°C were grown for producing F2 seeds that were used for segregation analysis. In addition, TG10 and TG01 were crossed for epistatic analysis. For molecular complementation of TG01 mutant, we isolated the *LsABA1* gene from cv. Salinas using the primers listed in **Table S5**. The 2.3 kb promoter with 373 bp of the 1st exon was isolated from Salinas DNA using the primers *LsABA1*ProF1 and *LsABA1*ProR1 (**Table S5**), while the coding region of *LsABA1* was isolated from cDNA generated from total RNA of Salinas seeds using SuperScript III First-Strand Synthesis kit (Thermo Fisher Scientific, USA) and primers LsABAF3 and 1r listed in **Table S5**. Both amplicons were digested with *BglII* and gel-purified for ligation to form a *ProLsABA1*::*LsABA1* fragment that was used for end-to-end PCR. The amplified *ProLsABA1*::*LsABA1* fragment was gel-purified and ligated to pDONR221 through BP reaction, and subsequently ligated to binary Gateway expression vector pGWB401via LR

reaction. The Pro*LsABA1*:: *LsABA1*-pGWB401 construct was introduced into EHA105 *Agrobacterium tumefasciens* for lettuce transformation that was conducted at the UC Davis Ralph M. Parsons Foundation Plant Transformation Facility using a method described by Michelmore *et al.,* (1987). The transformants were identified using primers that bind to the *nptII* gene (**Table S5**). The F2 seeds from these hemizygous T1 plants were used for germination tests at 35°C in the light. Among ten independent transgenic lettuce lines generated, three stable independent lines with single transgene insertions were characterized for this study. For Arabidopsis molecular complementation, the *LsABA1* coding region was amplified from the pDONR221-Pro*LsABA1*::*LsABA1* plasmid DNA and also from TG01 mutant seed cDNA. Both amplified fragments were PCR-purified and cloned into pDONR207 via BP reaction, and then ligated to pGWB402 via LR reaction. Both expression constructs were transferred to GV3101 *Agrobacterium* for Arabidopsis transformation through floral dip (Zhang *et al.,* 2006). All gateway pGWB vectors in this study were kindly provided by Dr. Tsuyoshi Nakagawa (Nakagawa et al., 2007). We have observed that Arabidopsis *aba1-1* plants are sensitive to bacterial infection and die easily with or without plastic cover for maintaining high humidity after floral dips; therefore, plants were sprayed with100 µM ABA + 0.02% (w/v) Triton X-100 three times per week until ~3days before floral dip. Transgenic Arabidopsis plants were genotyped and advanced to the T3 generation to obtain stable insertion lines for germination tests.

Measurement of endogenous ABA content

ABA extraction from mature seeds was performed as previously described (Huo *et al.,* 2013). Seeds (~60 mg) were frozen in liquid nitrogen, ground into fine powder, and then lyophilized for 1 h using a SpeedVac at 25°C (Heto Vacuum Centrifuge; Thermo Scientific). For each 10 mg of dry tissue, 1 mL of methanol containing 500 mg L^{-1} citric acid monohydrate and 100 mg L^{-1} butylated hydroxytoluene was added. Samples were incubated with shaking (100 rpm) at 4°C for 24 h and centrifuged at 1500*g* for 10 min to collect the supernatant. The extracts were assayed using the Phytodetek ABA test kit (Agdia). Three biological samples with two technical replications each were assayed for each genotype.

Accession Numbers

Primer sequences and accession numbers in this study can be found in the **Table S5.** Sequences of lettuce genome can be accessed in the website of Compositate Genome Project (CGP: http://compgenomics.ucdavis.edu/).

ACKNOWLEDGEMENTS

The authors greatly thank Dr. Shouhui Wei, Dr. Danubia Nobre, and undergraduate students Alex Wai and Phat Hua for maintaining plants in greenhouse at Davis. We also thank Dr. Fei-Yian Yoong for lettuce crossing demonstration. We also thank Kathie J. Ngo for technical support in sequencing analysis. This work was supported by Rijk Zwaan Breeding B.V., 2678 ZG De Lier, The Netherlands. The authors declare that they have no conflict of interest.

Short supporting information legends

Figure S1. Germination of seeds of four mutants and two genotypes of lettuce at 25°C in the dark for 5 days.

Figure S2. TG01 and TG10 are allelic and controlled by a single recessive mutation.

Figure S3. Distribution of polymorphism types in the two mutant DNA bulks.

Figure S4. Partial alignment of *LsABA1* amino acid sequences from Yorvik wild-type (YORWT), TG01 and TG10.

Figure S5. Fine mapping and marker order around the high temperature germination (HTG) trait locus in Yorvik mutant TG01.

Figure S6. Predicted secondary structure of TG01 mutant and Yorvik wild-type ABA1 proteins.

Figure S7. Predicted tertiary structures of LsABA1 protein of Yorvik wild-type and TG01.

Figure S8. Pre-harvest sprouting of TG01 mutant seeds due to lower ABA content.

Figure S9. Comparison of timelines among different sequencing methods and conventional mapping to identify EMS mutations.

Table S1. Candidate mutations within the BSA mapping region in TG01and TG10 mutants.

Table S2. Graphical representation of the haplotypes of 197 recombinants.

Table S3. Candidate genes found in the genetically mapped region.

Table S4. ABA content of Yorvik, TG01, TG10 and Salinas dry seeds and seeds imbibed at 35°C.

Table S5. Primers used in this study.

Data S1. Predicted tertiary structures of LsABA1 protein of Yorvik wild-type in PDB format.

Data S2. Predicted tertiary structures of LsABA1 protein of TG01 mutant in PDB format.

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.

Alonso-Blanco, C., Bentsink, L., Hanhart, C.J., Blankestijn-de Vries, H. and Koornneef, M. (2003) Analysis of natural allelic variation at seed dormancy loci of *Arabidopsis thaliana*. *Genetics*, **164**, 711-729.

Argyris, J., Dahal, P., Hayashi, E., Still, D.W. and Bradford, K.J. (2008) Genetic variation for lettuce seed thermoinhibition is associated with temperature-sensitive expression of abscisic acid, gibberellin, and ethylene biosynthesis, metabolism, and response genes. *Plant Physiol*, **148**, 926-947.

Argyris, J., Truco, M.J., Ochoa, O., Knapp, S.J., Still, D.W., Lenssen, G.M., Schut, J.W., Michelmore, R.W. and Bradford, K.J. (2005) Quantitative trait loci associated with seed and seedling traits in Lactuca. *Theor Appl Genet*, **111**, 1365-1376.

Argyris, J., Truco, M.J., Ochoa, O., McHale, L., Dahal, P., Van Deynze, A., Michelmore, R.W. and Bradford, K.J. (2011) A gene encoding an abscisic acid biosynthetic enzyme (*LsNCED4*) collocates with the high temperature germination locus *Htg6.1* in lettuce (*Lactuca* sp.). *Theor Appl Genet*, **122**, 95-108.

Ashelford, K., Eriksson, M.E., Allen, C.M., D'Amore, R., Johansson, M., Gould, P., Kay, S., Millar, A.J., Hall, N. and Hall, A. (2011) Full genome re-sequencing reveals a novel circadian clock mutation in Arabidopsis. *Genome Biol*, **12**.

Barrero, J.M., Piqueras, P., Gonzalez-Guzman, M., Serrano, R., Rodriguez, P.L., Ponce, M.R. and Micol, J.L. (2005) A mutational analysis of the ABA1 gene of Arabidopsis thaliana highlights the involvement of ABA in vegetative development. *J Exp Bot*, **56**, 2071-2083.

Bentsink, L., Jowett, J., Hanhart, C.J. and Koornneef, M. (2006) Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. *Proc Natl Acad Sci USA*, **103**, 17042-17047.

Bewley, J.D., Bradford, K.J., Hilhorst, H.W.M. and Nonogaki, H. (2013) *Seeds: Physiology of Development, Germination and Dormancy* Third edn. New York: Springer.

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

Cantliffe, D.J., Sung, Y. and Nascimento, W.M. (2000) Lettuce seed germination. *Hortic Rev*, **224**, 229-275.

Cheng, W.H., Chiang, M.H., Hwang, S.G. and Lin, P.C. (2009) Antagonism between abscisic acid and ethylene in Arabidopsis acts in parallel with the reciprocal regulation of their metabolism and signaling pathways. *Plant Mol Biol*, **71**, 61-80.

Chiu, R.S., Nahal, H., Provart, N.J. and Gazzarrini, S. (2012) The role of the Arabidopsis FUSCA3 transcription factor during inhibition of seed germination at high temperature. *BMC Plant Biol*, **12**, 15.

Contreras, S., Bennett, M.A., Metzger, J.D. and Tay, D. (2008) Maternal light environment during seed development affects lettuce seed weight, germinability, and storability. *HortScience*, **43**, 845-852.

Contreras, S., Bennett, M.A., Metzger, J.D., Tay, D. and Nerson, H. (2009) Red to far-red ratio during seed development affects lettuce seed germinability and longevity. *HortScience*, **44**, 130-134.

Dong, T.T., Tong, J.H., Xiao, L.T., Cheng, H.Y. and Song, S.Q. (2012) Nitrate, abscisic acid and gibberellin interactions on the thermoinhibition of lettuce seed germination. *Plant Growth Regul*, **66**, 191-202.

Frey, A., Effroy, D., Lefebvre, V., Seo, M., Perreau, F., Berger, A., Sechet, J., To, A., North, H.M. and Marion-Poll, A. (2012) Epoxycarotenoid cleavage by NCED5 fine-tunes ABA accumulation and affects seed dormancy and drought tolerance with other NCED family members. *Plant J*, **70**, 501-512.

Gabriele, S., Rizza, A., Martone, J., Circelli, P., Costantino, P. and Vittorioso, P. (2010) The Dof protein DAG1 mediates PIL5 activity on seed germination by negatively regulating GA biosynthetic gene *AtGA3ox1*. *Plant J*, **61**, 312-323.

Gonzalez-Guzman, M., Apostolova, N., Belles, J.M., Barrero, J.M., Piqueras, P., Ponce, M.R., Micol, J.L., Serrano, R. and Rodriguez, P.L. (2002) The short-chain alcohol dehydrogenase ABA2 catalyzes the conversion of xanthoxin to abscisic aldehyde. *Plant Cell*, **14**, 1833-1846.

Gu, X.-Y., Foley, M.E., Horvath, D.P., Anderson, J.V., Feng, J., Zhang, L., Mowry, C.R., Ye, H., Suttle, J.C., Kadowaki, K.-i. and Chen, Z. (2011) Association between seed dormancy and pericarp color is controlled by a pleiotropic gene that regulates abscisic acid and flavonoid synthesis in weedy red rice. *Genetics*, **189**, 1515-1524.

Hartwig, B., James, G.V., Konrad, K., Schneeberger, K. and Turck, F. (2012) Fast isogenic mapping-by-sequencing of ethyl methanesulfonate-induced mutant bulks. *Plant Physiol*, **160**, 591-600.

Henikoff, S. and Comai, L. (2003) Single-nucleotide mutations for plant functional genomics. *Annu Rev Plant Biol*, **54**, 375-401.

Henry, I.M., Nagalakshmi, U., Lieberman, M., Ngo, K., Krasileva, K., Vasquez-Gross, H., Akhunova, A., Akunov, E., Dubcovsky, J., Tai, T.H. and Comai, L. (2014) Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. *Plant Cell*, **26**, 1382-1397.

Henry, I.M., Zinkgraf, M.S., Groover, A.T. and Comai, L. (2015) A system for dosage-based functional genomics in *Poplar*. *Plant Cell*, **27**, 2370-2383.

Huo, H. and Bradford, K.J. (2015) Molecular and hormonal regulation of thermoinhibition of seed germination. In *Advances in Plant Dormancy* (Anderson, J.V. ed. Switzerland: Springer.

Huo, H., Dahal, P., Kunusoth, K., McCallum, C.M. and Bradford, K.J. (2013) Expression of *9*-cis-*EPOXYCAROTENOID DIOXYGENASE4* is essential for thermoinhibition of lettuce seed germination but not for seed development or stress tolerance. *Plant Cell*, **25**, 884-900.

James, G.V., Patel, V., Nordstrom, K.J., Klasen, J.R., Salome, P.A., Weigel, D. and Schneeberger, K. (2013) User guide for mapping-by-sequencing in Arabidopsis. *Genome Biol*, **14**, R61.

Karssen, C.M., Brinkhorst-van der Swan, D.L.C., Breekland, A.E. and Koornneef, M. (1983) Induction of dormancy during seed development by endogenous abscisic acid : Studies on abscisic acid deficient genotypes of *Arabidopsis thaliana* (L) Heynh. *Planta*, **157**, 158-165.

Kim, D.H., Yamaguchi, S., Lim, S., Oh, E., Park, J., Hanada, A., Kamiya, Y. and Choi, G. (2008) SOMNUS, a CCCH-type zinc finger protein in Arabidopsis, negatively regulates light-dependent seed germination downstream of PIL5. *Plant Cell*, **20**, 1260-1277.

Koornneef, M., Jorna, M.L., Brinkhorst-van der Swan, D.L.C. and Karssen, C.M. (1982) The isolation of abscisic acid (ABA) deficient mutants by selection of induced revertants in non-germinating gibberellin sensitive lines of *Arabidopsis thaliana* (L) Heynh. *Theor Appl Genet*, **61**, 385-393.

Kushiro, T., Okamoto, M., Nakabayashi, K., Yamagishi, K., Kitamura, S., Asami, T., Hirai, N., Koshiba, T., Kamiya, Y. and Nambara, E. (2004) The Arabidopsis cytochrome P450 CYP707A encodes ABA 8'-hydroxylases: key enzymes in ABA catabolism. *EMBO J*, **23**, 1647-1656.

Lebeda, A., Dolezalova, I., Ferakova, V. and Astley, D. (2004) Geographical distribution of wild *Lactuca* species (Asteraceae, Lactuceae). *Bot Rev*, **70**, 328-356.

Lefebvre, V., North, H., Frey, A., Sotta, B., Seo, M., Okamoto, M., Nambara, E. and Marion-Poll, A. (2006) Functional analysis of Arabidopsis *NCED6* and *NCED9* genes indicates that ABA synthesized in the endosperm is involved in the induction of seed dormancy. *Plant J*, **45**, 309-319.

Leon-Kloosterziel, K.M., Gil, M.A., Ruijs, G.J., Jacobsen, S.E., Olszewski, N.E., Schwartz, S.H., Zeevaart, J.A. and Koornneef, M. (1996) Isolation and characterization of abscisic acid-deficient Arabidopsis mutants at two new loci. *Plant J*, **10**, 655-661.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

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

Michelmore, R., Marsh, E., Seely, S. and Landry, B. (1987) Transformation of Lettuce (Lactuca-Sativa) Mediated by Agrobacterium-Tumefaciens. *Plant Cell Rep*, **6**, 439-442.

Michelmore, R.W., Paran, I. and Kesseli, R.V. (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci USA*, **88**, 9828-9832.

NAAS (2015) Crop Values 2014 Summary: United States Department of Agriculture.

Nakagawa T, Suzuki T, Murata S, Nakamura S, Hino T, Maeo K, Tabata R, Kawai T, Tanaka K, Niwa Y, Watanabe Y (2007) Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. Biosci Biotechnol Biochem 71: 2095-2100.

Nakamura, S., Komatsuda, T. and Miura, H. (2007) Mapping diploid wheat homologues of Arabidopsis seed ABA signaling genes and QTLs for seed dormancy. *Theor Appl Genet*, **114**, 1129-1139.

Nambara, E. and Marion-Poll, A. (2005) Abscisic acid biosynthesis and catabolism. *Annu Rev Plant Biol*, **56**, 165-185.

Nordstrom, K.J., Albani, M.C., James, G.V., Gutjahr, C., Hartwig, B., Turck, F., Paszkowski, U., Coupland, G. and Schneeberger, K. (2013) Mutation identification by direct comparison of whole-genome sequencing data from mutant and wild-type individuals using k-mers. *Nat Biotechnol*, **31**, 325-330.

Okamoto, M., Kuwahara, A., Seo, M., Kushiro, T., Asami, T., Hirai, N., Kamiya, Y., Koshiba, T. and Nambara, E. (2006) CYP707A1 and CYP707A2, which encode abscisic acid 8'-hydroxylases, are indispensable for proper control of seed dormancy and germination in Arabidopsis. *Plant Physiol*, **141**, 97-107.

Osa, M., Kato, K., Mori, M., Shindo, C., Torada, A. and Miura, H. (2003) Mapping QTLs for seed dormancy and the *Vp1* homologue on chromosome 3A in wheat. *Theor Appl Genet*, **106**, 1491-1496.

Pankin, A., Campoli, C., Dong, X., Kilian, B., Sharma, R., Himmelbach, A., Saini, R., Davis, S.J., Stein, N., Schneeberger, K. and von Korff, M. (2014) Mapping-by-sequencing identifies *HvPHYTOCHROME C* as a candidate gene for the early maturity 5 locus modulating the circadian clock and photoperiodic flowering in barley. *Genetics*, **198**, 383-396.

Parry, M.A., Madgwick, P.J., Bayon, C., Tearall, K., Hernandez-Lopez, A., Baudo, M., Rakszegi, M., Hamada, W., Al-Yassin, A., Ouabbou, H., Labhilili, M. and Phillips, A.L. (2009) Mutation discovery for crop improvement. *J Exp Bot*, **60**, 2817-2825.

Piskurewicz, U., Jikumaru, Y., Kinoshita, N., Nambara, E., Kamiya, Y. and Lopez-Molina, L. (2008) The gibberellic acid signaling repressor RGL2 inhibits Arabidopsis seed germination by stimulating abscisic acid synthesis and ABI5 activity. *Plant Cell*, **20**, 2729-2745.

Rizal, G., Thakur, V., Dionora, J., Karki, S., Wanchana, S., Acebron, K., Larazo, N., Garcia, R., Mabilangan, A., Montecillo, F., Danila, F., Mogul, R., Pablico, P., Leung, H., Langdale, J.A., Sheehy, J., Kelly, S. and Quick, W.P. (2015) Two forward genetic screens for vein density mutants in sorghum converge on a cytochrome P450 gene in the brassinosteroid pathway. *Plant J*, **84**, 257-266.

Rizza, A., Boccaccini, A., Lopez-Vidriero, I., Costantino, P. and Vittorioso, P. (2011) Inactivation of the *ELIP1* and *ELIP2* genes affects Arabidopsis seed germination. *New Phytol*, **190**, 896-905.

Schneeberger, K. (2014) Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nat Rev Genet*, **15**, 662-676.

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.

Stam, P. (1993) Construction of integrated genetic-linkage maps by means of a new computer package - Joinmap. *Plant J*, **3**, 739-744.

Stoffel, K., van Leeuwen, H., Kozik, A., Caldwell, D., Ashrafi, H., Cui, X., Tan, X., Hill, T., Reyes-Chin-Wo, S., Truco, M.J., Michelmore, R.W. and Van Deynze, A. (2012) Development and application of a 6.5 million feature Affymetrix Genechip(R) for massively parallel discovery of single position polymorphisms in lettuce (*Lactuca* spp.). *BMC genomics*, **13**, 185.

Sugimoto, K., Takeuchi, Y., Ebana, K., Miyao, A., Hirochika, H., Hara, N., Ishiyama, K., Kobayashi, M., Ban, Y., Hattori, T. and Yano, M. (2010) Molecular cloning of *Sdr4*, a regulator involved in seed dormancy and domestication of rice. *Proc Natl Acad Sci USA*, **107**, 5792-5797.

Takagi, H., Tamiru, M., Abe, A., Yoshida, K., Uemura, A., Yaegashi, H., Obara, T., Oikawa, K., Utsushi, H., Kanzaki, E., Mitsuoka, C., Natsume, S., Kosugi, S., Kanzaki, H., Matsumura, H., Urasaki, N., Kamoun, S. and Terauchi, R. (2015) MutMap accelerates breeding of a salt-tolerant rice cultivar. *Nat Biotechnol*, **33**, 445-449.

Tamura, N., Yoshida, T., Tanaka, A., Sasaki, R., Bando, A., Toh, S., Lepiniec, L. and Kawakami, N. (2006) Isolation and characterization of high temperature-resistant germination mutants of Arabidopsis thaliana. *Plant Cell Physiol*, **47**, 1081-1094.

Toh, S., Imamura, A., Watanabe, A., Nakabayashi, K., Okamoto, M., Jikumaru, Y., Hanada, A., Aso, Y., Ishiyama, K., Tamura, N., Iuchi, S., Kobayashi, M., Yamaguchi, S., Kamiya, Y., Nambara, E. and Kawakami, N. (2008) High temperature-induced abscisic acid biosynthesis and its role in the inhibition of gibberellin action in Arabidopsis seeds. *Plant Physiol*, **146**, 1368-1385.

Truco, M.J., Ashrafi, H., Kozik, A., van Leeuwen, H., Bowers, J., Wo, S.R.C., Stoffel, K., Xu, H., Hill, T., Van Deynze, A. and Michelmore, R.W. (2013) An ultra-high-density, transcript-based, genetic map of lettuce. *G3: Genes|Genomes|Genetics*, **3**, 617-631.

Valdes, V.M., Bradford, K.J. and Mayberry, K.S. (1985) Alleviation of thermodormancy in coated lettuce seeds by seed priming. *HortScience*, **20**, 1112-1114.

Woudenberg, L.J., Coppoolse, E.R., Lenssen, G.M., Schut, J.W. and Smits, E.C.J. (2014) High temperature germinating lettuce seeds. Pat. Appl.: PCT Publ.No. WO 2014/140378.

Ye, H., Feng, J., Zhang, L., Zhang, J., Mispan, M.S., Cao, Z., Beighley, D.H., Yang, J. and Gu, X.-y. (2015) Map-based cloning of *Seed Dormancy1-2* identified a gibberellin synthesis gene regulating the development of endosperm-imposed dormancy in rice. *Plant Physiol*, **169**, 2152-2165.

Yoong, F.Y., O'Brien, L.K., Truco, M.J., Huo, H., Sideman, R., Hayes, R., Michelmore, R.W. and Bradford, K.J. (2016) Genetic variation for lettuce seed germination thermotolerance is associated

with temperature-sensitive regulation of *ETHYLENE RESPONSE FACTOR1* (*ERF1*). *Plant Physiol*, **70**, 472-488.

Zhang, X.R., Henriques, R., Lin, S.S., Niu, Q.W. and Chua, N.H. (2006) *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat Protoc*, **1**, 641-646.

Figure Legends

Figure 1. Temperature sensitivity of four wild-type lines and four EMS mutants during seed germination.

(**a**) Germination of seeds of four wild-type lettuce cultivars at different temperatures. (**b**) Temperatures at which seed germination was reduced to 50% for four mutant lines. (**c**) Seed germination of Yorvik wild-type (Yor) and its EMS mutants TG01 and TG10 and of (**d**) Apache wild-type (Apa) and its EMS mutants, TG09 and ETR6, at 25 and 35°C for 24 h and 35°C + 3 µM ABA for 60 h. Bars in (a) indicate the standard error calculated from three replicates with 30 seeds for each replicate; germination in (a) and (b) was scored after 120 h in the light.

Figure 2. Responses of lettuce EMS mutants to different stresses during imbibition at 35°C.

Germination of TG01, TG10, TG09, ETR6, RNAi and US96 seeds imbibed in: (**a**) -0.5 MPa PEG8000; (**b**) 150 mM NaCl; (**c**) 100 µM PAC; (**d**) 20 mM STS; (**e**) 1 µM ABA; and (**f**) 3 µM ABA. All tests were performed in continuous light.

Figure 3. Simplified scheme for parallel sequencing of two lettuce mutants to identify EMS mutations.

Homozygous TG01 and TG10 EMS mutants were generated through independent EMS mutagenesis and selfing. Two homozygous EMS mutants were backcrossed to Yorvik wild-type to generate BC/F2 seeds, which were screened for germination at 35°C. Seedlings from germinated seeds were grown for two weeks, and DNA was extracted from two bulked leaf samples for whole genome sequencing. Sequencing reads were aligned to the lettuce reference genome sequence (Salinas). SNPs observed between the reference genome and the two Yorvik mutants, but not between the two mutants, were considered non-EMS variants. Only SNPs observed between the two mutants were considered as potential EMS-induced mutations. SNPs that occurred at different loci within the same gene were considered to be candidate gene mutations.

Figure 4. Mapping of the causal genes based on EMS mutations selected in the two mutant pools.

(a) For each of the mutant pools, positions that were either homozygous mutant (one allele contributed >95% of the coverage in both mutant pools but for a different allele) or heterozygous (allele percentages between 40 and 60% for both alleles) with a high percentage of mutant alleles were selected. Next, the mean percentage of mutant alleles was calculated for each consecutive, non-overlapping 1 Mb bin across the current genomic reference sequence. (**b**) Close up of the peak obtained in (**a**) with data from both mutant pools overlapping. Lines: smoothed curves.

Figure 5. The TG10 mutation results in mRNA mis-splicing.

(a) Diagram of the *LsABA1* gene structure. The green boxes with numbers represent exons. The arrows above exon7 and exon8 indicate mutation positions present in TG10 and TG01, respectively. The positions of different primer sets are also indicated with different colors and labels. (**b**) Section of alignment of *LsABA1* cDNA sequences from Yorvik wild-type (YORWT), TG01 and TG10. The black line indicates the sequence for primer 4f that spans the 6^{th} exon and 7^{th} exon as shown in (**a**). TG10-DEL: TG10 cDNA with 1 nt deletion; TG10-INS: TG10 cDNA with 39 nt unspliced intron insertion. (**c**) *LsABA1* relative expression in dry seeds of YORWT, TG01 and TG10. Expression data were normalized to *LsACT7* reference gene expression. Levels were normalized relative to those in YORWT for each primer set (2f/2r, 3f/3r, 4f/4r) indicated in (**a**) and (**b**). Expression levels did not differ significantly except for TG10 with primer set 4. Bars in (c) indicate standard errors calculated from three biological and two technical replicates. Student's *t*-tests were performed for statistical separations at P=0.05 within each primer set.

Figure 6. Genetic map of the lettuce chromosome (linkage group) 3 with map positions of *LsABA1* and other genetic markers.

(a) The region flanked by markers Ls00484 and Ls00913 (black arrowed line) was identified to be closely linked to the HTG phenotype in TG01 using a population of 706 TG01 × Apache F3 progeny. The physical distances on the lettuce genome version 5.0 (Lsat V5) are also

shown for the HTG markers. (**b**) A physical map shows a 6.5 Mb HTG region after fine mapping with the additional markers between Ls00484 and Ls00898 in (**a**) using the same F3 mapping population. Polymorphic marker positions are shown on their corresponding scaffolds (grey boxes). The markers underlined were used to screen the F3 plants; part of the genotyping data of these polymorphic markers is shown in **Table S2**.

Figure 7. Molecular complementation of lettuce TG01 mutant and Arabidopsis *aba1-1*.

(**a**) The wild type lettuce *LsABA1* gene under its own 2.3 kb promoter was introduced to the TG01 mutant for molecular complementation. Germination of seeds from TG01 and Yorvik wild type (YORWT) and F2 seeds from three independent transgenic TG01 lines expressing Pro*LsABA1*::*LsABA1*was scored after 5 days at 25 and 35°C in the light. (**b**) Coding region of wild type *LsABA1* and mutated *LsABA1*(*mLsABA1*) from TG01 under CaMV35S promoter were introduced to the Arabidopsis mutant *aba1-1* for heterologous complementation. To test the sensitivity to high temperature during seed germination, seeds (3 weeks after harvest) of Arabidopsis *aba1-1* mutant and *aba1-1* mutant overexpressing wild-type *LsABA1* or *mLsABA1* were scored for germination after 5 days in the light at 32 or 34°C.

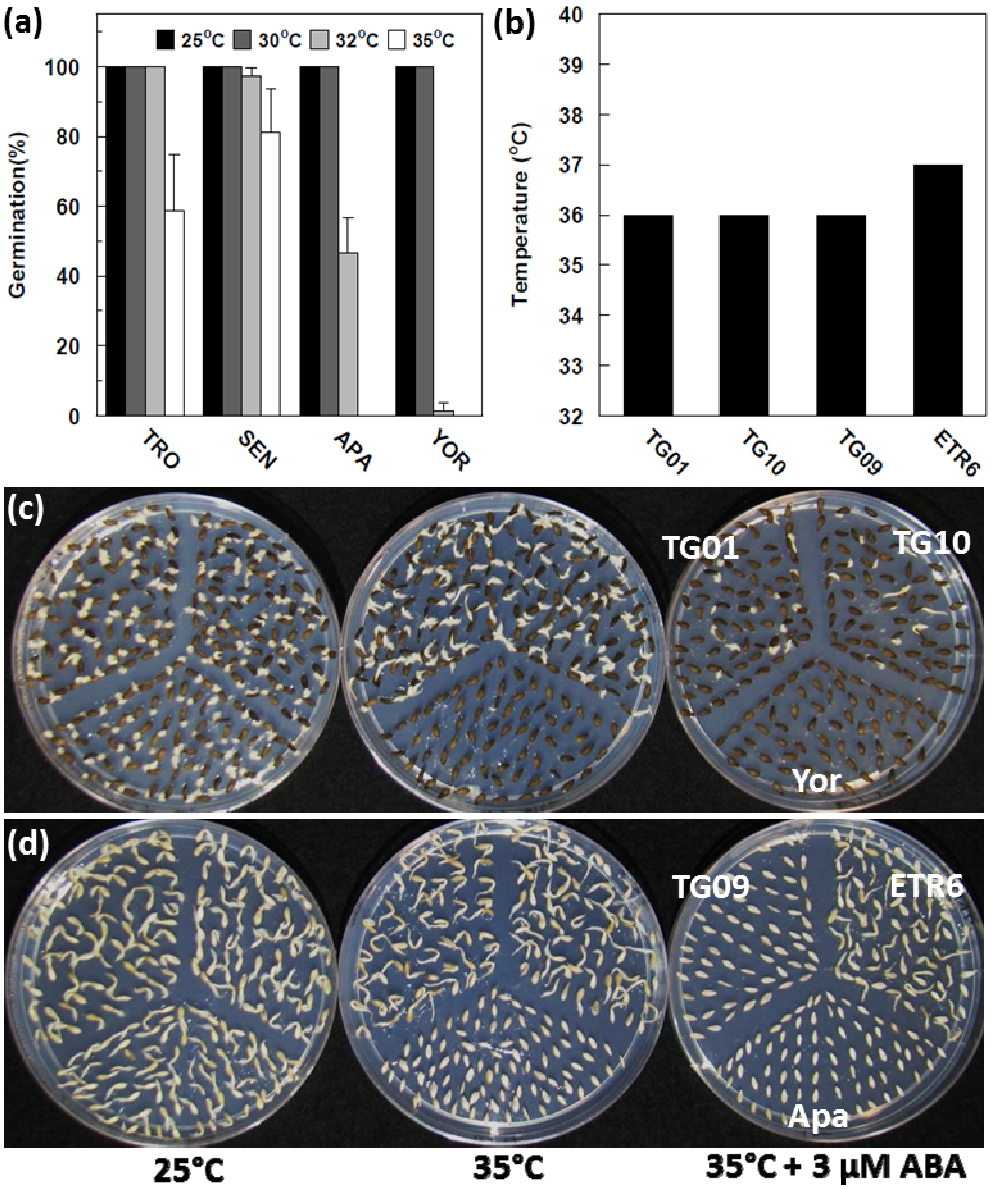

Figure 1. Temperature sensitivity of four wild-type lines and four EMS mutants during seed germination.

(**a**) Germination of seeds of four wild type lettuce cultivars at different temperatures. (**b**) Temperatures at which seed germination was reduced to 50% for four mutant lines. (**c**) Seed germination of Yorvik wild type (Yor) and its EMS mutants TG01 and TG10 and of (**d**) Apache wild type (Apa) and its EMS mutants, TG09 and ETR6, at 25 or 35°C for 24 h and 35°C + 3 µM ABA for 60 h. Bars in (a) indicate the standard error calculated from three replicates with 30 seeds for each replicate; germination in (a) and (b) was scored after 120 h in the light.

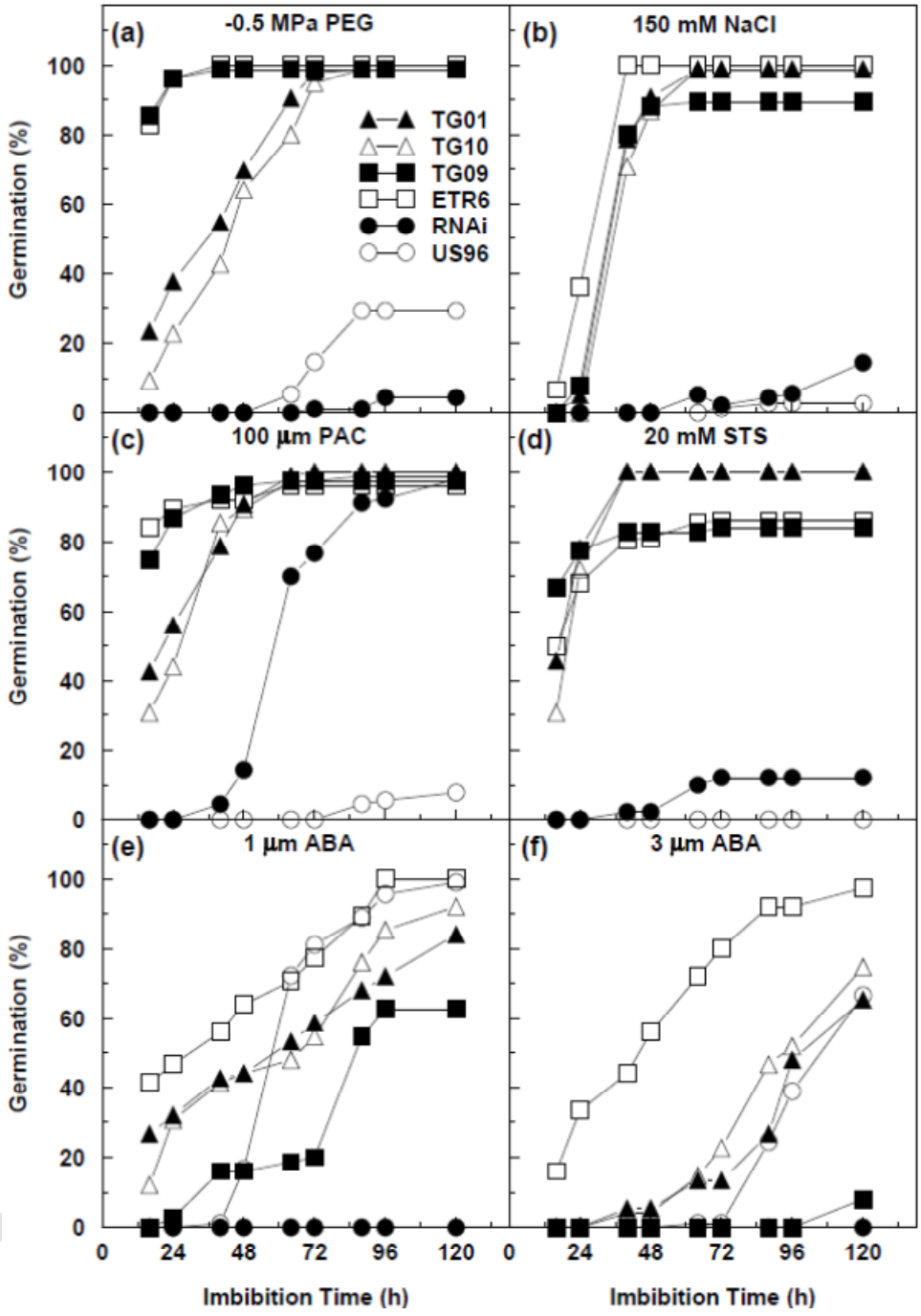

Figure 2. Responses of lettuce EMS mutants to different stresses during imbibition at 35°C.

Germination of TG01, TG10, TG09, ETR6, RNAi and US96 seeds imbibed in: (**a**) -0.5 MPa PEG8000; (**b**) 150 mM NaCl; (**c**) 100 µM PAC; (**d**) 20 mM STS; (**e**) 1 µM ABA; and (**f**) 3 µM ABA. All tests were performed in continuous light.

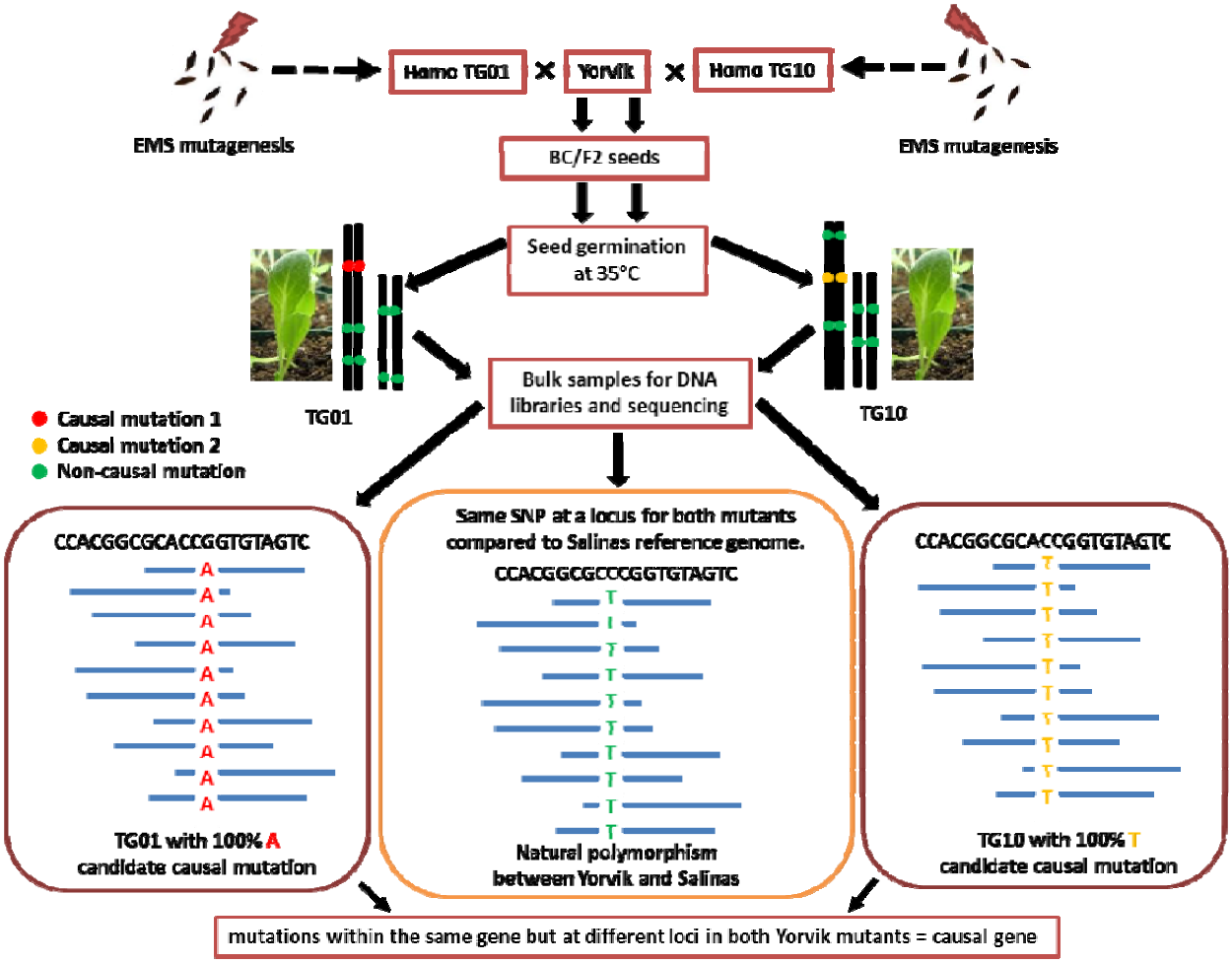

Figure 3. Simplified scheme for parallel sequencing of two lettuce mutants to identify EMS mutations.

Homozygous TG01 and TG10 EMS mutants were generated through independent EMS mutagenesis and selfing. Two homozygous EMS mutants were backcrossed to Yorvik wild type to generate BC/F2 seeds, which were screened for germination at 35°C. Seedlings from germinated seeds were grown for two weeks. DNA was extracted from separate bulked leaf samples for whole genome sequencing. Sequencing reads were aligned to the lettuce reference genome sequence (Salinas). SNPs observed between the reference genome and the two Yorvik mutants, but not between the two mutants, were considered non-EMS variants. Only SNPs observed between the two mutants were considered as potential EMS-induced mutations. SNPs that occurred at different loci within the same gene were considered to be candidate gene mutations.

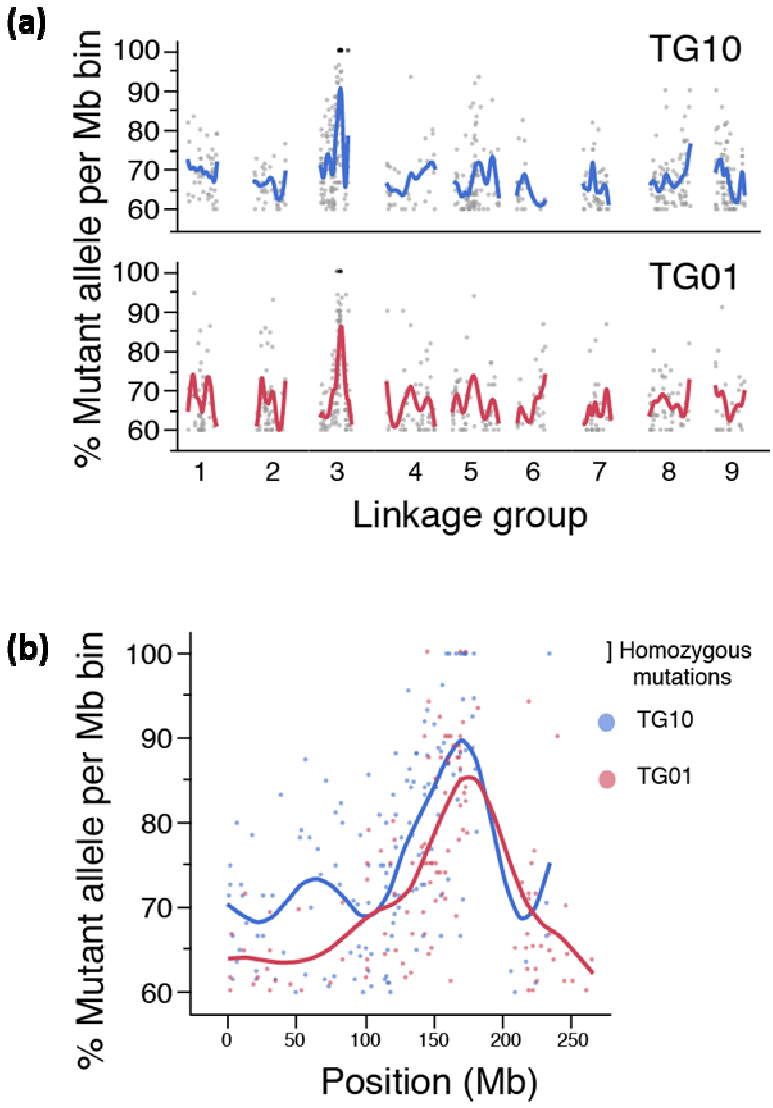

Figure 4. Mapping of the causal genes based on EMS mutations selected in the two mutant pools.

(**a**) For each of the mutant pools, positions that were either homozygous mutant (one allele contributed >95% of the coverage in both mutant pools but for a different allele) or heterozygous with a high percentage (allele percentages between 40 and 60% for both alleles) of mutant alleles were selected. Next, the mean percentage of mutant alleles was calculated for each consecutive, non-overlapping 1 Mb bin across the current genomic reference sequence. (**b**) Close up of the peak obtained in (**a**) with data from both mutant pools overlapping. Lines: smoothed curves.

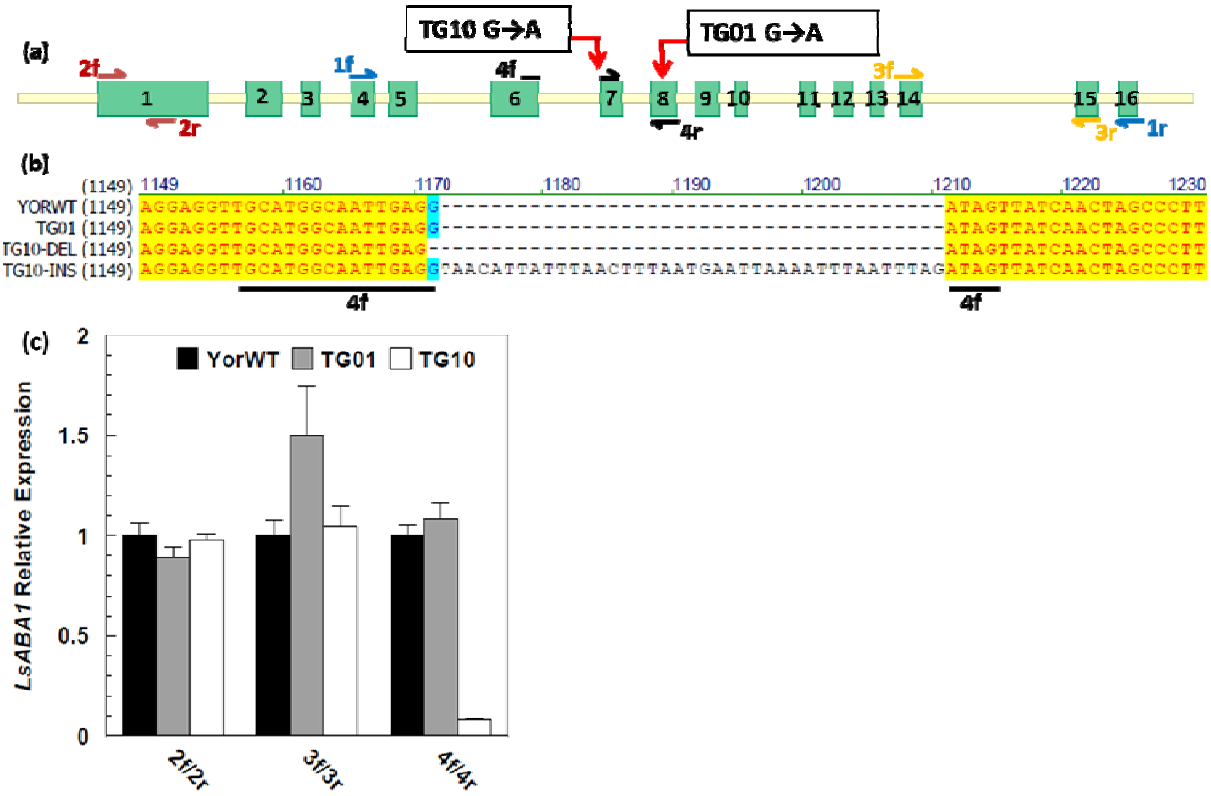

Figure 5. The TG10 mutation results in mRNA mis-splicing.

(**a**) Diagram of the *LsABA1* gene structure. The green boxes with numbers represent exons. The arrows above exon7 and exon8 indicate mutation positions present in TG10 and TG01, respectively. The positions of different primer sets are also indicated with different colors and labels. (**b**) Section of alignment of *LsABA1* cDNA sequences from Yorvik wild type (YORWT), TG01 and TG10. The black line indicates the sequence for primer 4f that spans the 6th exon and 7th exon as shown in (**a**). TG10-DEL: TG10 cDNA with 1 nt deletion; TG10-INS: TG10 cDNA with 39 nt unspliced intron insertion. (**c**) *LsABA1* relative expression in dry seeds of YORWT, TG01 and TG10. Expression data were normalized to *LsACT7* reference gene expression. Levels were normalized relative to those in YORWT for each primer set (2f/2r, 3f/3r, 4f/4r) indicated in (**a**) and (**b**). Expression levels did not differ significantly except for TG10 with primer set 4. Bars in (c) indicate standard errors calculated from three biological and two technical replicates. Student's *t* tests were performed for statistical separations at P=0.05 within each primer set.

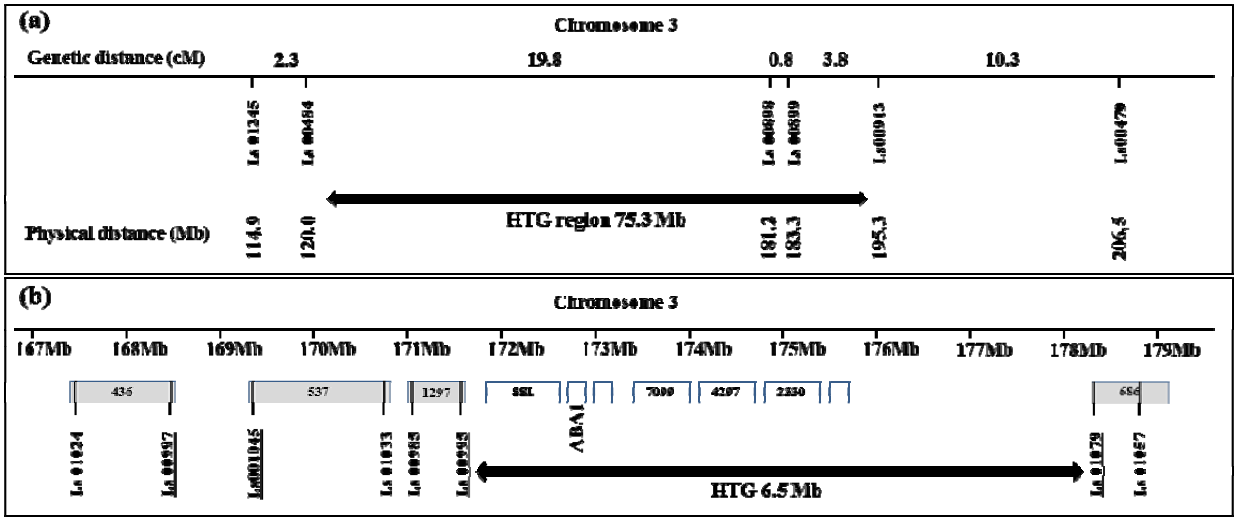

Figure 6. Genetic map of the lettuce chromosome (linkage group) 3 with map positions of *LsABA1* and other genetic markers.

(**a**) The region flanked by markers Ls00913 and Ls00484 (black arrowed line) was identified to be closely linked to the HTG phenotype in TG01 using a population of 706 TG01 × Apache F3 progeny. The physical distances on the lettuce genome version 5.0 (Lsat V5) are also shown for the HTG markers. (**b**) A physical map shows a 6.5 Mb HTG region after fine mapping with the additional markers between Ls00484 and Ls00898 in (**a**) using the same F3 mapping population. Polymorphic marker positions are shown on their corresponding scaffolds (grey boxes). The markers underlined were used to screen the F3 plants; part of the genotyping data of these polymorphic markers is shown in **Table S2**.

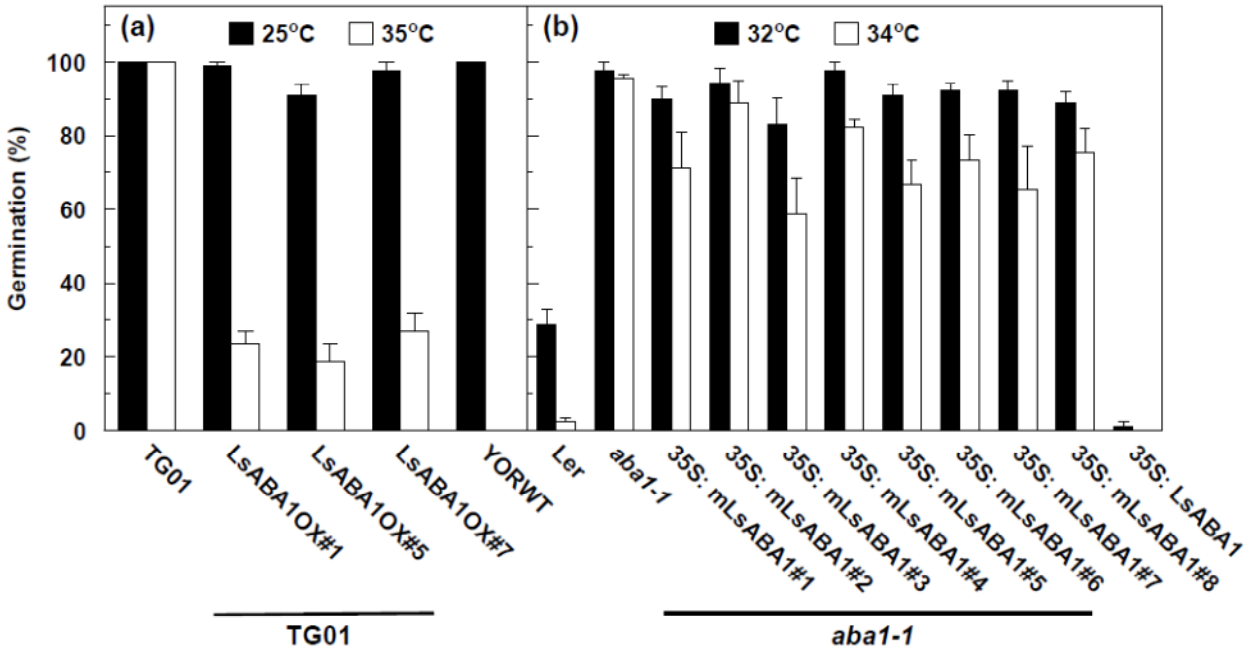

Figure 7. Molecular Complementation of Lettuce TG01 Mutant and Arabidopsis *aba1-1*.

(**a**) The wild type lettuce *LsABA1* gene under its own 2.3 kb promoter was introduced to the TG01 mutant for molecular complementation. Germination of seeds from TG01 and Yorvik wild type (YORWT) and F2 seeds from three independent transgenic TG01 lines expressing Pro*LsABA1*::*LsABA1*was scored after 5 days at 25 and 35°C in the light. (**b**) Coding region of wild type *LsABA1* and mutated *LsABA1*(*mLsABA1*) from TG01 under CaMV35S promoter were introduced to the Arabidopsis mutant *aba1-1* for heterologous complementation. To test the sensitivity to high temperature during seed germination, seeds (3 weeks after harvest) of Arabidopsis *aba1-1* mutant and *aba1-1* mutant overexpressing wild-type *LsABA1* or *mLsABA1* were scored for germination after 5 days in the light at 32 or 34°C.