The Plant Cell, Vol. 25: 4827–4843, December 2013, www.plantcell.org

LARGE-SCALE BIOLOGY ARTICLE

Genome-Wide Analysis of Branched-Chain Amino Acid Levels in *Arabidopsis* Seeds[W]

Ruthie Angelovici,[a] Alexander E. Lipka,[b] Nicholas Deason,[a] Sabrina Gonzalez-Jorge,[a] Haining Lin,[c] Jason Cepela,[d] Robin Buell,[d] Michael A. Gore,[e] and Dean DellaPenna[a,1]

[a] Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824–1319
[b] Institute for Genomic Diversity, Cornell University, Ithaca, New York 14853
[c] Dupont Pioneer, Johnston, Iowa, 50131
[d] Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824
[e] Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York 14853

Branched-chain amino acids (BCAAs) are three of the nine essential amino acids in human and animal diets and are important for numerous processes in development and growth. However, seed BCAA levels in major crops are insufficient to meet dietary requirements, making genetic improvement for increased and balanced seed BCAAs an important nutritional target. Addressing this issue requires a better understanding of the genetics underlying seed BCAA content and composition. Here, a genome-wide association study and haplotype analysis for seed BCAA traits in *Arabidopsis thaliana* revealed a strong association with a chromosomal interval containing two *BRANCHED-CHAIN AMINO ACID TRANSFERASES*, *BCAT1* and *BCAT2*. Linkage analysis, reverse genetic approaches, and molecular complementation analysis demonstrated that allelic variation at *BCAT2* is responsible for the natural variation of seed BCAAs in this interval. Complementation analysis of a *bcat2* null mutant with two significantly different alleles from accessions Bayreuth-0 and Shahdara is consistent with BCAT2 contributing to natural variation in BCAA levels, glutamate recycling, and free amino acid homeostasis in seeds in an allele-dependent manner. The seed-specific phenotype of *bcat2* null alleles, its strong transcription induction during late seed development, and its subcellular localization to the mitochondria are consistent with a unique, catabolic role for BCAT2 in BCAA metabolism in seeds.

INTRODUCTION

The amino acids Ile, Leu, and Val are biosynthetically and chemically related, all having branched hydrocarbon side chains that are responsible for their aliphatic nature and typical location in transmembrane domains of proteins (Binder, 2010). Branched-chain amino acids (BCAAs) are essential nutrients for humans and animals, which cannot synthesize BCAAs and must obtain them in sufficient quantity through the diet. The homeostasis of BCAAs is highly regulated in humans and animals, and any surplus is degraded; the disruption of BCAA homeostasis is the basis of several inborn errors of metabolism in humans that have severe developmental and neurological consequences (Chuang et al., 2006). In addition to their roles as protein building blocks, BCAAs (mainly Leu) are important signaling molecules that can activate the target of rapamycin signaling pathway in animals, which in turn regulates several aspects of growth and development (Kimball and Jefferson, 2006a, 2006b; Hannah et al., 2010; Valerio et al., 2011). Studies of BCAA dietary supplements in diverse animal models support roles in cardiac and skeletal muscle biogenesis, mitochondrial biogenesis, limiting oxidative damage, and additive effects with other treatments for extended longevity (D'Antona et al., 2010; Valerio et al., 2011).

As in animals, BCAA homeostasis in plants is tightly controlled, with catabolism playing a central role. For example, mutation of the BCAA catabolic enzyme isovaleryl-coenzyme A (CoA) dehydrogenase (IVD) in *Arabidopsis thaliana* elevated levels of seed BCAAs and nine other amino acids (Gu et al., 2010). BCAA breakdown products are also involved in glutamate recycling, branched-chain ester formation (Kandra et al., 1990; Walters and Steffens, 1990; Kroumova et al., 1994; Däschner et al., 1999; Li et al., 2003; Gu et al., 2010), and the production of branched-chain volatiles in fruit (Tressl and Drawert, 1973; Rowan et al., 1996; Pérez et al., 2002; Matich and Rowan, 2007; Gonda et al., 2010). During carbon deficiency, BCAA breakdown products can provide an alternative energy source by being fed into the tricarboxylic acid cycle (Taylor et al., 2004; Engqvist et al., 2009) or donating electrons directly to the electron transport chain (Ishizaki et al., 2005, 2006; Araújo et al., 2010).

BCAA biosynthesis in plants occurs in the chloroplast (Ellerström et al., 1992; Diebold et al., 2002; Zybailov et al., 2008; Binder, 2010). Ile and Val synthesis requires a single set of four enzymes that use multiple substrates to form these two amino

[1] Address correspondence to dellapen@cns.msu.edu.
The author 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: Dean DellaPenna (dellapen@cns.msu.edu).
[W] Online version contains Web-only data.
www.plantcell.org/cgi/doi/10.1105/tpc.113.119370

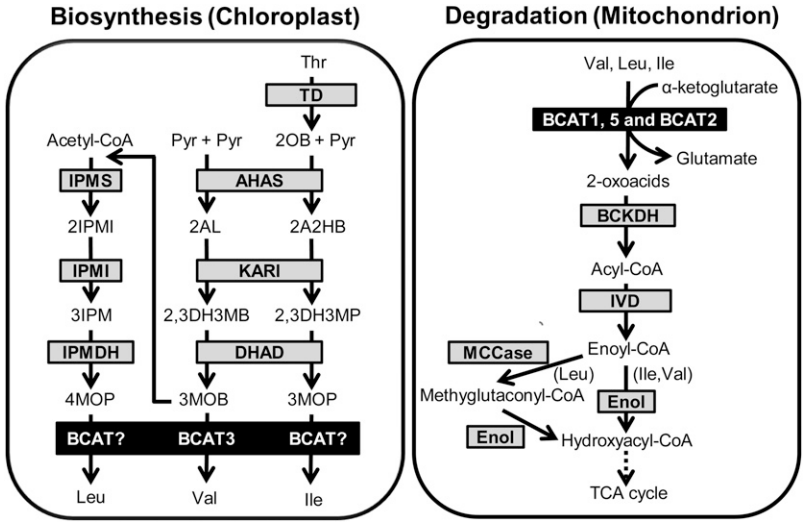

Figure 1. Compartmentalization of BCAA Biosynthesis and Degradation.

Key enzymes and metabolites of BCAA biosynthesis and degradation in the chloroplast and mitochondrion, respectively, are abbreviated and given in boxes. BCAT enzymes are highlighted by black boxes. AHAS, acetohydroxyacid synthase; 2A2HB, 2-aceto-2-hydroxybutanoate; 2AL, 2-acetolactate; BCKDH, branched-chain keto acid dehydrogenase; DHAD, dihydroxyacid dehydratase; 2,3DH3MB, 2,3-dihydroxy-3-methylbutanoate; 2,3DH3MP, 2,3-dihydroxy-3-methylpentanoate; Enol, enoyl-CoA hydratase; 2-IPM, 2-isopropylmalate; 3-IPM, 3-isopropylmalate; IPMDH, isopropylmalate dehydrogenase; IPMI, isopropylmalate isomerase; IPMS, isopropylmalate synthase; KARI, ketol acid reductoisomerase; 3MOB, 3-methyl-2-oxobutanoate; 3MOP, 3-methyl-2-oxopentanoate; 4MOP, 4-methyl-2-oxopentanoate; 2OB, 2-oxobutanoate; Pyr, pyruvate; TCA, tricarboxylic acid; TD, Thr deaminase. Known functions for specific BCATs are indicated, and steps without a specific associated BCAT are indicated by "BCAT?". BCAT2 localization and catabolic role were deduced in this article.

acids (Figure 1), the first three of which are also used for Leu synthesis. Leu synthesis branches off at 2-oxo-3-oxobutanoate and requires three additional, Leu-specific enzymatic activities (Figure 1; Binder, 2010). BCAA degradation takes place mostly in the mitochondria (Figure 1; Zolman et al., 2001; Beck et al., 2004) and hence is physically separated and compartmentalized from biosynthesis. However, the last reaction of BCAA biosynthesis is also, in the reverse direction, the first step of BCAA catabolism and catalyzed by a small gene family of branched-chain amino acid aminotransferases (BCATs; EC 2.6.1.42).

BCATs catalyze reversible reactions that convert 2-oxo acids into their corresponding amino acids and vice versa (Diebold et al., 2002; Schuster and Binder, 2005; Binder, 2010), and this class of enzymes has been studied biochemically and through transgenic approaches in several plant species (Campbell et al., 2001; Diebold et al., 2002; Schuster and Binder, 2005; Knill et al., 2008; Gao et al., 2009; Binder, 2010; Chen et al., 2010; Gonda et al., 2010; Maloney et al., 2010; Kochevenko and Fernie, 2011, 2012; Kochevenko et al., 2012a,b). These studies have defined roles for individual BCAT family members in amino acid and glucosinolate metabolism, hormone signaling, and volatile production (Binder et al., 2007; Knill et al., 2008; Gao et al., 2009; Binder, 2010). Linkage analysis was performed on BCAAs from tomato fruit pericarp using a population of the wild relative *Solanum pennellii* introgressed into the cultivated tomato (*Solanum lycopersicum*; Schauer et al., 2006, 2008; Kochevenko and Fernie, 2011, 2012). This analysis identified several quantitative trait loci (QTLs), some of which contained BCATs in their support intervals, suggesting a role for BCATs in the natural variation of BCAA content in ripening tomato fruit pericarp (Maloney et al., 2010; Kochevenko and Fernie, 2011).

In *Arabidopsis*, seven BCAT genes have been identified: *BCAT1* (At1g10060), *BCAT2* (At1g10070), *BCAT3* (At3g49680), *BCAT4* (At3g19710), *BCAT5* (At5g65780), *BCAT6* (At1g50110), and *BCAT7* (At1g50090). BCAT1 is active against all three BCAAs, and its mitochondrial localization suggests that it is a catabolic enzyme (Schuster and Binder, 2005; Binder, 2010). BCAT2 and BCAT3 are plastid localized and likely biosynthetic (Diebold et al., 2002), although expression correlation analysis grouped *BCAT2* with Leu degradation enzymes localized to the mitochondrion (Mentzen et al., 2008; Less and Galili, 2009; Less et al., 2010). BCAT4 and BCAT6 are localized in the cytosol (Diebold et al., 2002), with BCAT4 involved exclusively in glucosinolate synthesis (Schuster et al., 2006; Sawada et al., 2009), while BCAT3 is involved in both BCAA and glucosinolate synthesis (Knill et al., 2008). BCAT5 localization and function are still unclear, as green florescent protein tagging localized it to the chloroplast yet proteomic studies indicated mitochondrial localization (Diebold et al., 2002; Taylor et al., 2004; Zybailov et al., 2008). Finally, *BCAT7* is considered a pseudogene (Binder, 2010).

The genetic basis of natural variation for complex metabolic traits in plants is becoming increasingly accessible via association mapping (Zhu et al., 2008; Atwell et al., 2010; Chan et al., 2011). A genome-wide association study (GWAS) of maize (*Zea mays*) leaf metabolites showed an association of several lignin precursors with cinnamoyl-CoA reductase (Riedelsheimer et al., 2012). Similarly, GWAS of flowering time, pathogen resistance, and

glucosinolate levels in *Arabidopsis* leaves yielded important information on the genetic architecture of these traits, with several known genes validated and several new candidate genes identified (Zhu et al., 2008; Atwell et al., 2010; Chan et al., 2011). Recently, GWASs of maize kernel traits have been reported for starch, oil, and protein (Cook et al., 2012) as well as vitamin E (Li et al., 2012; Lipka et al., 2013), suggesting that beneficial alleles can be identified and deployed for improvement of grain quality.

Previous studies performed on seed free amino acid (FAA) pools successfully identified key genes regulating total Lys levels in *Arabidopsis* and tobacco (*Nicotiana tabacum*) seeds that served as a basis for engineering increased Lys in soybean (*Glycine max*), rapeseed (*Brassica napus*), and maize (Falco et al., 1995; Mazur et al., 1999; Zhu and Galili, 2003; Frizzi et al., 2008). The *opaque2* mutant in maize represents another example in which increasing free Lys levels enhanced the overall nutritional value of grain (Mertz et al., 1964). Thus, although seed storage proteins are the major contributors to total seed amino acid content (Millerd, 1975; Shewry and Halford, 2002) and FAAs generally account for 1 to 10% of the total (Muehlbauer et al., 1994), FAAs can make significant contributions to the levels of limiting essential amino acids in seeds. To better understand the genetic components underlying natural variation for seed BCAAs in *Arabidopsis*, we used liquid chromatography–tandem mass spectrometry (LC-MS/MS) to quantify FAAs from 313 accessions of an *Arabidopsis* diversity panel (Nordborg et al., 2005) and then conducted GWAS for free BCAA-related seed traits in combination with QTL analysis in a biparental population and reverse genetics approaches. These analyses defined *BCAT2* as a key locus associated with the natural variation of seed BCAAs. Metabolic, transgenic, and localization studies provided evidence that *BCAT2* is localized to the mitochondrion, where it plays an important role in BCAA catabolism in seeds.

RESULTS

Natural Variation of Free BCAAs in *Arabidopsis* Seeds

To assess the extent of phenotypic variation for free BCAAs in dry *Arabidopsis* seeds, the absolute levels of 18 FAAs were quantified from three independent outgrowths (biological replicates) of an *Arabidopsis* diversity panel (see Supplemental Data Set 1 online; Atwell et al., 2010; Platt et al., 2010). Nine BCAA-related traits were determined: absolute BCAA levels (Ile, Leu, and Val in nmol/mg dry seed), relative BCAA levels (each BCAA as a percentage of total FAAs; e.g., Ile/total), and the ratio of each BCAA to the sum of all three BCAAs (e.g., Ile/BCAAs). Val was the most abundant of the three BCAAs, with Val/BCAAs averaging 53.6%, while Ile/BCAAs and Leu/BCAAs averaged 22 and 24.4%, respectively (Table 1). The relationship between phenotypic values of the nine traits was evaluated by Spearman's rank correlation, and strong positive correlations were identified between all three BCAA absolute traits and between all three relative level traits (see Supplemental Table 1 online). By contrast, a weak positive correlation between Ile/BCAAs and Leu/BCAAs and strong negative correlations between Val/BCAAs and Leu/BCAAs and Ile/BCAAs were identified (see Supplemental Table 1 online). Broad-sense heritabilities ranged from 0.55 to 0.82 for all traits (Table 1), suggesting that phenotypic variation is controlled primarily by QTLs.

Association Mapping of Seed BCAA Traits

To dissect the genetic basis of natural variation for BCAA levels in dry *Arabidopsis* seeds, GWAS was performed using a 360-ecotype *Arabidopsis* diversity panel previously genotyped with a 250K single nucleotide polymorphism (SNP) chip (Li et al., 2010; Horton et al., 2012). Because of extremely delayed flowering of some accessions, seeds could be harvested and phenotyped from only 313 of the 360 accessions (see Supplemental Data Set 1 online). Removal of low-quality SNPs and those with minor allele frequency (MAF) < 0.05 yielded a 170,344-SNP data set for association analysis using a unified mixed linear model that controls for population structure and familial relatedness (Yu et al., 2006). Significant SNP–trait associations at a 5% false discovery rate (FDR) were found for five traits: Ile, Ile/total, Ile/BCAAs, Leu/total, and Val/BCAAs (Figure 2; see Supplemental Data Set 2 online), with all showing a strong association signal on chromosome 1 (Figure 2). Of these five traits, the four Ile- and Leu-related traits demonstrated high positive correlations but had negative correlations to Val/BCAAs (see Supplemental Table 1 online).

Table 1. Means, Ranges, and Broad-Sense Heritabilities of Nine BCAA Traits in *Arabidopsis* Seeds

	Untransformed BLUPs				
Trait	Mean	SE	Range	Broad-Sense Heritability	SE
Ile (nmol/mg seed)	0.15	2.3E-03	0.074–0.33	0.76	0.02
Leu (nmol/mg seed)	0.16	1.6E-03	0.10–0.27	0.66	0.03
Val (nmol/mg seed)	0.36	3.9E-03	0.26–0.63	0.72	0.03
Ile/total (%)	1.13	1.3E-02	0.61–1.93	0.82	0.02
Leu/total (%)	1.19	9.4E-03	0.87–1.62	0.69	0.03
Val/total (%)	2.71	1.9E-02	2.11–3.73	0.78	0.02
Ile/BCAAs	0.22	1.0E-03	0.16–0.26	0.71	0.03
Leu/BCAAs	0.24	7.3E-04	0.20–0.27	0.55	0.04
Val/BCAAs	0.54	1.5E-03	0.49–0.59	0.70	0.03

Total is the sum of all 18 FAAs in nmol/mg dry seeds; BCAAs is the sum of Ile, Val, and Leu in nmol/mg dry seeds.

The peak SNP locus on chromosome 1 (SNP5373; 3,292,297 bp; P values ranged between 3.1×10^{-12} and 4.9×10^{-7}) for all five traits was located within a gene annotated as *EARLY-RESPONSIVE TO DEHYDRATION STRESS PROTEIN4* (*ERD4*; At1G10090; see Supplemental Data Set 2 online). This peak SNP locus, SNP5373, explained 19.2, 16.2, 15.2, 15, and 12.3% of the total phenotypic variance for Ile/total, Ile/BCAAs, Leu/BCAAs, Ile, and Val/BCAA, respectively. Notably, this SNP is located 2133 bp from the stop codon of *BCAT2* (At1G10070) and 5460 bp from the stop codon of *BCAT1* (At1G10060), which encode biosynthetic and catabolic BCAA enzymes. Depending on the trait, a total of 1 to 43 SNPs in the vicinity of SNP5373 had significant associations (see Supplemental Data Set 2 online). Taken together, these SNPs span a 119-kb interval that contains 20 open reading frames (see Supplemental Data Set 2 online).

In addition to the GWAS signal on chromosome 1, five chromosomal regions with significant SNP–trait associations were identified (Figure 2; see Supplemental Data Set 2 online). Ile/BCAAs had a significant association with SNP76015 on chromosome 2 (16,780,007 bp) within At2G40170, annotated as encoding a stress-induced Late Embryogenesis Abundant6 protein, GEA6. Ile and Ile/total had two associations on chromosome 3, SNP114591 and SNP114596, at 18,151,275 and 18,149,257 bp, respectively, in adjacent genes annotated as encoding a glycoside hydrolase family 28 protein (At3G48950) and a large subunit ribosomal protein, L13e (At3G48960). Chromosome 4 had four significant SNP associations: Ile/total and Ile/BCAAs with SNP149802 (11,776,120 bp) within a gene annotated as coding for an F-box/RNI-like superfamily protein (At4G22280), Ile/BCAAs with SNP146708 and SNP146715 (10,386,003 and 10,388,712 bp, respectively) within and near a gene for a MADS box transcription factor (At4G18960), and Ile/total with SNP139249 (7,535,127 bp) located within a gene of unknown function (At4G12840).

To further resolve the complex signals from the chromosome 1 region containing the most statistically significant SNPs, a multiple-locus mixed model (MLMM) approach that uses stepwise model selection (Segura et al., 2012) was implemented

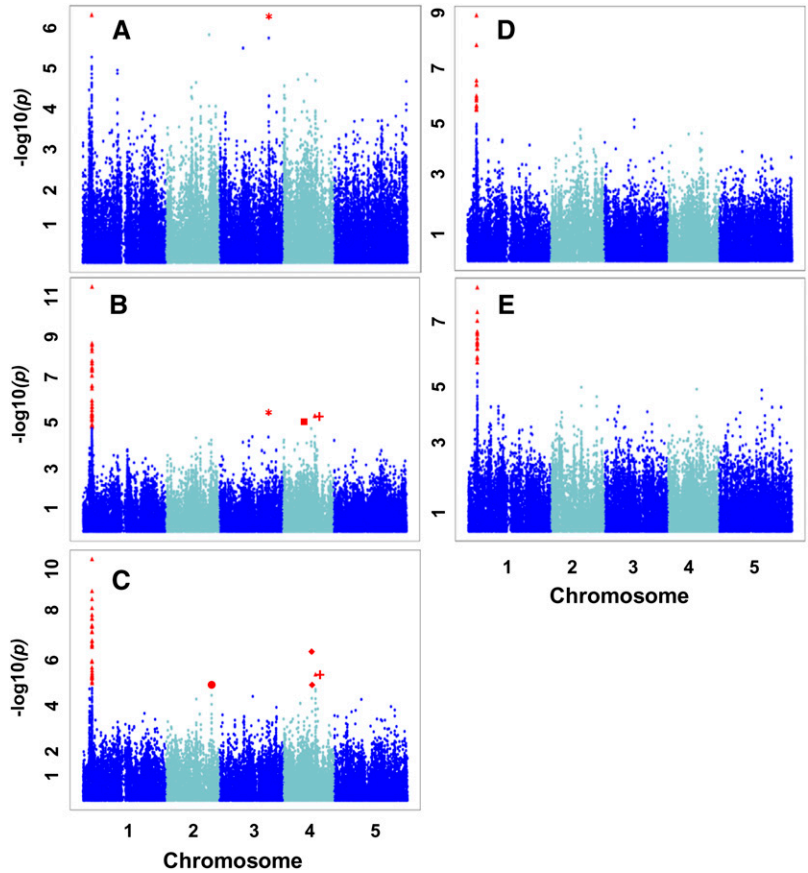

Figure 2. GWAS for the Five BCAA Traits with Significant SNP–Trait Associations.

Scatterplots of association results from a unified mixed-model analysis for five BCAA traits across the five *Arabidopsis* chromosomes are shown. Negative $\log_{10}$-transformed P values from the GWAS analysis are plotted against the genomic physical position. P values for SNPs that are statistically significant for a trait at 5% FDR are colored in red. Traits are as follows: Ile **(A)**, Ile/total **(B)**, Ile/BCAAs **(C)**, Leu/total **(D)**, and Val/BCAAs **(E)**. All five traits shared significant SNP–trait associations on chromosome 1. Five additional chromosomal regions with significant SNP–trait associations are marked by different red shapes, with the same shapes indicating a corresponding chromosomal region influencing multiple traits.

on the five traits in the vicinity of SNP5373. In this analysis, all 594 SNPs with a MAF $\geq$ 0.05 in the $\pm$100-kb region surrounding SNP5373 on chromosome 1 were considered for inclusion into the final model. For each trait, the optimal model contained only SNP5373 (the peak SNP locus identified in the initial GWAS) and explained 7 to 16% of the phenotypic variation for the five traits (see Supplemental Table 2 online). To validate these results, GWAS was reconducted using a unified mixed model that included SNP5373 as a covariate. With this conditional GWAS, no significant associations were detected for any of these traits, including the five additional genomic regions that were previously identified on chromosomes 2, 3, and 4 (Figure 2). Estimates of linkage disequilibrium (LD) between SNP5373 and other SNPs in the surrounding $\pm$100-kb region indicated that patterns of long-range LD were present (see Supplemental Figure 1 online) and may account for the multiple significant SNPs that were identified for this region. Taken together, these results indicate that SNP5373 is likely in LD with one or more causative polymorphisms that are responsible for the natural variation of BCAA-related traits in dry *Arabidopsis* seeds.

Haplotype Analysis

It has been shown previously (Atwell et al., 2010) that peak GWAS signals like SNP5373 are not always located within a causal gene but can be in LD with other causal polymorphisms nearby. In an attempt to increase our resolution and further characterize contributing polymorphisms in this region, haplotype analysis was performed. In contrast to GWAS, which considers individual SNPs in isolation, haplotype analysis allows one to consider multiple SNPs that segregate together due to LD and as a group may be most significant for a trait. Haplotype analysis was conducted on a 23.5-kb chromosome 1 region that includes *ERD4*, *BCAT1*, and *BCAT2* (3,274,080–3,297,645; see Supplemental Data Set 2 online) and encompasses all significant SNPs common to the four BCAA traits with multiple SNP associations. Fourteen haploblocks were identified across this region (Figure 3B; see Supplemental Figure 2 online; see Supplemental Table 3 online), including the 6,148-bp haploblock 10, which contains the peak SNP from GWAS, SNP5373, and encompasses portions of *ERD4* and *BCAT2*. When unified mixed linear models were fitted for each haploblock, haploblocks 8 and 9 were found to be the most significant of the seven haploblocks associated with all five BCAA traits (see Supplemental Table 3 online; see Supplemental Figure 2 online). By contrast, haploblock 10 was associated with only three traits and in each case was less significant than haploblocks 8 and 9.

Haploblock 8 (321 bp) contains three haplotypes and is delineated by two SNPs located in the 3′ UTR of *BCAT1* and the promoter region of *BCAT2*. Haploblock 9 (708 bp) also contains three haplotypes delineated by two SNPs located in the promoter and first exon of *BCAT2* (Figure 3B; see Supplemental Figure 2 online; see Supplemental Table 3 online). The most significantly different haplotype pairs for haploblocks 8 and 9 were CA versus TT and GA versus AG, respectively (the combined haplotype frequencies of CA-GA and TT-AG were 72 and 12.3% respectively). The CA-GA haplotype pairs represent the high-level haplotypes for Ile, Ile/total, Leu/total, and Ile/BCAAs and the low-level haplotype for Val/BCAAs (Figure 3C; see Supplemental Table 3 online; see Supplemental Figure 2 online). The opposite effect of this haplotype pair on these BCAA traits (Figure 3C) is consistent with their divergent Pearson correlations (see Supplemental Table 1 online). Taken together, these results strongly suggest that either *BCAT1* or *BCAT2*, and not *ERD4*, is the causative gene, but the LD structure of this genomic region necessitates molecular dissection with mutagenesis and transgenic methods to define which *BCAT* is causal.

Bayreuth-0 × Shahdara QTL Analysis

To independently confirm the results predicted from GWAS and haplotype analyses, a biparental QTL analysis was performed using the Bayreuth-0 (Bay) × Shahdara (Sha) recombinant inbred population (Loudet et al., 2002). The Bay and Sha parents contain the haploblocks 8 and 9 low and high most significant haplotype pairs predicted by GWAS, respectively, and should segregate accordingly in the population for Ile-related traits (see Supplemental Figure 2 online; see Supplemental Table 3 online). Seed BCAA levels of 158 recombinant inbred lines of a Bay × Sha mapping population (Loudet et al., 2002) were quantified by LC-MS/MS, and composite interval mapping was performed for all nine BCAA traits using plant breeding and biology QTL analysis (PLABQTL) version 1.2 (Utz and Melchinger, 1996). In total, 12 chromosome regions with QTLs were identified, three of which contained known BCAA catabolism genes in their intervals (see Supplemental Data Set 3 online). No pairwise epistatic interactions were detected. Confidence intervals for QTLs that included the chromosomal 1 region defined by GWAS were identified for all five BCAA traits plus one additional trait, Val/total (Table 2). These composite interval mapping results indicate that, consistent with predictions from GWAS and haplotype analysis, Sha is the higher parent for Ile-related traits and Bay the higher parent for Val/BCAAs. These chromosome 1 QTLs explain 13.2 to 25.5% of the total variation and represent one of two major QTLs in the Bay × Sha mapping population that are shared with all five traits detected in GWAS (Table 2; see Supplemental Data Set 3 online).

Characterization of *BCAT1* and *BCAT2* T-DNA Insertion Mutants

A transgenic strategy was used to definitively test whether *BCAT1* or *BCAT2* is the causal gene underlying natural variation for BCAA levels in dry *Arabidopsis* seeds. Null, homozygous T-DNA insertion lines for *BCAT1*, *BCAT2*, and *ERD4* (*bcat1-1*, *bcat1-2*, *bcat2-1*, *bcat2-2*, and *erd4*) were isolated and characterized from the SALK and Versailles T-DNA collections (see Supplemental Figures 3 and 4 online). RT-PCR using primers designed to span the full *BCAT1*, *BCAT2*, and *ERD4* transcripts confirmed that all four mutant alleles lack transcripts derived from their respective genes in flowers and seeds (see Supplemental Figures 3C and 4B online).

Amino acid quantification of dry seeds from *bcat1*, *bcat2*, and *erd4* alleles showed that only the two *bcat2* alleles had a consistent and significant change in seed BCAA profiles compared with the wild type (Figure 4; Supplemental Data Set 4 online).

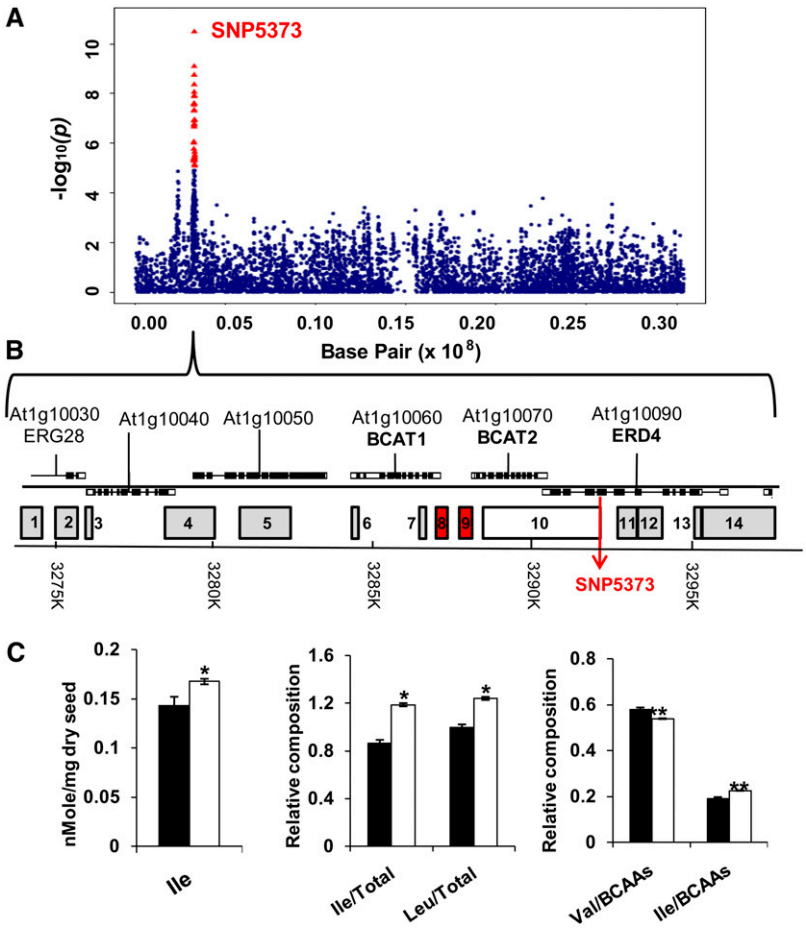

Figure 3. GWAS for Ile/BCAAs with Significant SNP–Trait Associations.

(A) Scatterplot of association results from a unified mixed-model analysis of Ile/BCAAs on chromosome 1. Negative $\log_{10}$-transformed P values from the GWAS analysis are plotted against the genomic physical position. P values for SNPs that are statistically significant at 5% FDR are colored in red.
(B) Graphical representation of genes and haploblocks within the genomic region spanning SNPs that showed significant associations across at least four BCAA traits (chromosome 1,3,274,080 to 3,297,645). Gene models are represented by black (exons) and white (untranslated regions) connected boxes. Haploblocks are represented with gray, red, and white squares; white indicates the haploblock containing SNP5373 that includes a large part of *BCAT2*; red represents the most significant haploblock associations across all five BCAA traits.
(C) Average levels of Ile, Ile/total, Leu/total, Val/BCAAs, and Ile/BCAAs from three biological replicates of the population. Black and white bars represent the trait average from accessions with the CA-GA (white) and TT-AG (black) haplotypes for haploblocks 8 and 9, respectively. In the case of Ile, Ile/total, Leu/total, and Ile/BCAAs, the CA-GA haplotype results in higher levels; lower levels are seen in the case of Val/BCAA. *P < 1.0E-04, **P < 1.5E-16 by Student's *t* test. Error bars represent SE ($n > 80$).

The Ile, Leu, and Val increases were 3.7-, 5.4-, and 2.6-fold relative to the wild type (Columbia [Col-0]) in *bcat2-1* and 5.2-, 7.7-, and 3.8-fold that of the wild type Wassilewskija-4 (WS-4) in *bcat2-2*. By contrast, consistent and significant effects in BCAA profiles were not observed in leaves from 4-week-old mutant plants (Figure 4; see Supplemental Data Set 4 online). These results indicate that mutant alleles of *bcat2*, but not *bcat1*, specifically influence seed BCAA levels and suggest that *BCAT2* is the causal locus for the natural variation of BCAAs in dry seeds. Accumulation of BCAAs in *bcat2* null mutants suggests that BCAT2 is a catabolic rather than a biosynthetic enzyme, as suggested previously (Diebold et al., 2002), as BCAAs would be expected to decrease from disruption of a biosynthetic enzyme.

Molecular Complementation of *bcat2-1* with the Bay and Sha Alleles

In order to directly test whether polymorphisms associated with *BCAT2* are the cause of natural variation for seed BCAA levels, molecular complementation of *bcat2-1* was performed with the Bay and Sha alleles, lines that harbor the two most significantly different allele pairs (high and low trait levels) of haploblocks 8 and 9. The null *bcat2-1* background was transformed with 3.81-kb *BCAT2* genomic DNA fragments from Bay and Sha that contain 1.5 kb of promoter, the *BCAT2* coding region, and 230 bp downstream of the *BCAT2* stop codon. The two alleles differ by 45 polymorphisms, including 10 insertions/deletions ranging

Table 2. QTL Analysis of BCAA Traits from the Bay × Sha Mapping Population

Trait	Chromosome	Position (centimorgan)	Supporting Interval (centimorgan)	Log of the Odds	High Parent	Allelic Effect Estimates	r^2 (%)	Pathway Genes in Supporting Intervals
Ile	1	8	4–14	6.06	Sha	−0.021	16.7	At1g10060, *BCAT1* At1g10070, *BCAT2*
Ile/total	1	8	6–12	7.09	Sha	−0.102	18.9	At1g10060, *BCAT1* At1g10070, *BCAT2*
Ile/BCAAs	1	8	6–12	9.3	Sha	−0.011	23.9	At1g10060, *BCAT1* At1g10070, *BCAT2*
Leu/total	1	8	6–14	9.91	Sha	−0.119	25.5	At1g10060, *BCAT1* At1g10070, *BCAT2*
Val/total	1	10	6–16	4.88	Sha	−0.13	13.5	At1g10060, *BCAT1* At1g10070, *BCAT2*
Val/BCAAs	1	8	6–14	4.77	Bay	0.013	13.2	At1g10060, *BCAT1* At1g10070, *BCAT2*

Only QTLs that overlap the ~6- to 12-centimorgan chromosome 1 interval previously detected by GWAS are shown. For full analysis, see Supplemental Data Set 3 online. Parameters were obtained from PLABQTL software using genotypic and phenotypic data from 158 recombinant inbred lines.

in size from 1 to 19 bp. Nine of the 35 SNP polymorphisms were scored in the association panel with the 250K SNP chip (Li et al., 2010; Horton et al., 2012), and seven were found to be significant for at least one BCAA trait. Five of these seven SNPs are located in the *BCAT2* promoter, one in the targeting sequence of the *BCAT2* coding region and one in the 3′ UTR (see Supplemental Table 4 online). Homozygous transgenic (T2) lines segregating for a single insertion locus were isolated, and the effect of each allele on seed amino acid phenotypes was determined by LC-MS/MS. Complementation with either transgenic allele caused a substantial reduction in the elevated BCAA levels of *bcat2-1*, with the average of 10 independent transgenic events containing the Sha allele [*bcat2-1*/*BCAT2*-(Sha)] being statistically indistinguishable from the average of 11 Col-0 plants (the *bcat2-1* background Col-0 and Sha have identical haplotypes for haploblocks 8 and 9; Figure 5). By contrast, the average of nine independent transgenic events containing the Bay allele [*bcat2-1*/*BCAT2*-(Bay)] resulted in significantly lower absolute levels of Ile and Val and significantly lower Ile/total and Ile/BCAAs ratios compared with Col-0. BCAA ratios of *bcat2-1*/*BCAT2*-(Sha) were indistinguishable from Col-0 with the exception of Ile/BCAAs, which was slightly lower (Figure 5; see Supplemental Table 5 online). These molecular complementation results confirm that the Bay *BCAT2* allele results in a significantly different seed BCAA profile (see Supplemental Table 5 online) that is consistent with results from the GWAS and QTL analyses. Interestingly, the *bcat2-1*/*BCAT2*-(Bay) FAA profile also showed significant reductions in Arg, Glu, Gly, Ser, and Met levels compared with Col-0 (see Supplemental Table 5 online).

Gene Expression Analysis of *BCAT1* and *BCAT2* during Seed Development and Desiccation

To address whether BCAA natural variation is driven by differences in expression levels, *BCAT1* and *BCAT2* transcripts were quantified at different stages of seed development in Bay and Sha using quantitative RT-PCR. In both accessions, *BCAT2* transcript levels showed a rapid increase during maturation and a slight decrease during desiccation, while *BCAT1* transcript levels showed a substantial decrease during late maturation and desiccation to nondetectable levels in dry seeds (Figure 6). That *BCAT2* transcript levels showed only minor variation between the two accessions suggests that expression differences make, at best, modest contributions to the observed phenotypic differences. *BCAT1* transcript levels in Bay were higher than those in Sha at 12 d after flowering (DAF) but lower at all other time points, substantially so at 18 and 20 DAF. The *Arabidopsis* eFP browser (Winter et al., 2007; http://bbc.botany.utoronto.ca/efp/cgi-bin/efpWeb.cgi) indicates that *BCAT1* is expressed only in the seed coat (see Supplemental Figure 5 online); thus, the absence of *BCAT1* transcript in dry seeds is consistent with seed coat death during the later stages of desiccation. The expression profiles of *BCAT1* and *BCAT2* are consistent with molecular complementation studies (Figure 5), demonstrating that *BCAT2* is the causal gene driving seed BCAA natural variation.

BCAT2 Subcellular Localization Analysis

The accumulation of BCAAs in *BCAT2* null mutants implied that the BCAT2 enzyme functions as a catabolic rather than a biosynthetic enzyme, as was reported previously (Diebold et al., 2002). Catabolic BCATs are generally located in the mitochondrion (Binder, 2010), but *BCAT2* had been reported to be plastidic, based on transient expression in tobacco leaves of the first 126 of the 388 amino acids of the BCAT2 protein (Diebold et al., 2002). To clarify its subcellular localization, the C terminus of the full-length BCAT2 protein was tagged with yellow fluorescent protein (YFP) and stably transformed into *Arabidopsis* plants expressing the cyan fluorescent protein (CFP) mitochondrial marker mt-ck CS16262 (Nelson et al., 2007). A C-terminal YFP fusion was used to avoid disrupting any potential N-terminal targeting sequences, as reported for other BCAT family members (Maloney et al., 2010; Kochevenko et al., 2012a, b). Fluorescence microscopy of stably transformed *Arabidopsis* leaves indicated that the 35S:BCAT2-YFP signal colocalized with the

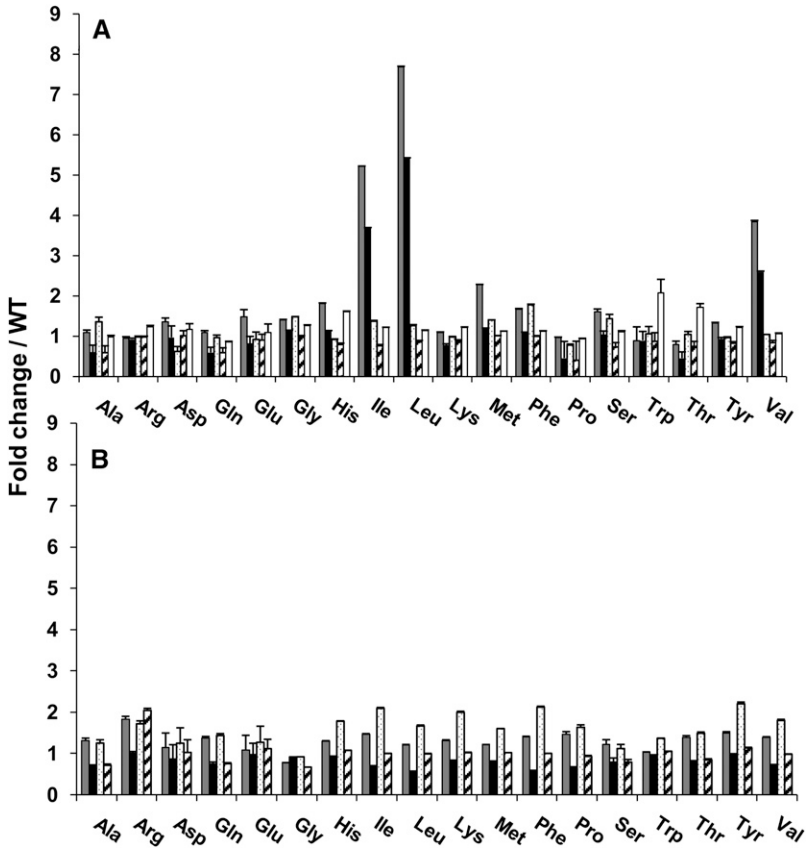

Figure 4. Seed and Leaf FAA Ratios in *BCAT* Mutants Compared with the Wild Type (Col-0 or WS-4).

FAA levels were measured from dry seeds **(A)** and leaves harvested from 4-week-old plants **(B)**. Values represent fold changes in FAA for *bcat2-2* (dark gray bars), *bcat2-1* (black bars), *bcat1-1* (dotted bars), *bcat1-2* (striped bars), and *erd4* (white bars; seeds only) relative to their respective wild type (WT). Averages and SE were calculated from five measurements. Statistical analyses are presented in Supplemental Data Set 4 online.

mitochondrial marker (Figure 7), with no signal detectable in chloroplasts. These results support in silico targeting predictions for *BCAT2* by TargetP (Emanuelsson et al., 2000) and iPSORT (Bannai et al., 2002) as well as observations that the *BCAT2* transcript is coexpressed with Leu catabolic enzymes, which are also located in the mitochondrion (Less and Galili, 2008, 2009; Mentzen et al., 2008). The observations are also consistent with the mitochondrial localization of cultivated tomato BCAT2, the closest tomato homolog to *Arabidopsis* BCAT2 (Maloney et al., 2010).

DISCUSSION

Understanding the genetic and biochemical regulation of seed BCAA levels is important because of their roles as essential amino acids in food and feed as well as additional effects on signaling and general metabolism in plants. Several studies in model plant systems show that seed development is associated with temporally distinct metabolic, proteomic, and transcriptomic switches that collectively accommodate the specialized metabolic needs of seeds during development and desiccation and in preparation for germination (Fait et al., 2006; Sreenivasulu et al., 2008; Xu et al., 2008; Angelovici et al., 2010). In *Arabidopsis* seeds, amino acid metabolism undergoes a pronounced metabolic switch during the transition from reserve accumulation to desiccation that results in the active biosynthesis and accumulation of many amino acids during desiccation (Fait et al., 2006). In addition, because of high metabolic rates and limited oxygen diffusion, developing seeds undergo hypoxia, which limits respiration, and as a consequence, several metabolic processes provide alternative energy sources (Rolletschek et al., 2002; Weber et al., 2005). BCAA catabolism is thought to be one such process, as the end products, acetyl-CoA, propionyl-CoA, and acetoacetate, can be tricarboxylic acid cycle substrates (Binder et al., 2007; Binder, 2010). BCAA catabolism can also donate electrons directly through electron transfer flavoprotein, as has been shown for IVD (Ishizaki et al., 2005, 2006; Engqvist et al., 2009; Araújo et al., 2010). Genetic disruption of BCAA catabolism has little effect on leaf amino acid profiles but severely influences amino acid metabolism in seeds. For example, an IVD null mutant showed elevated levels of BCAAs and nine other amino acids in seeds but not in leaves (Araújo et al., 2010; Gu et al., 2010). Broad effects on seed

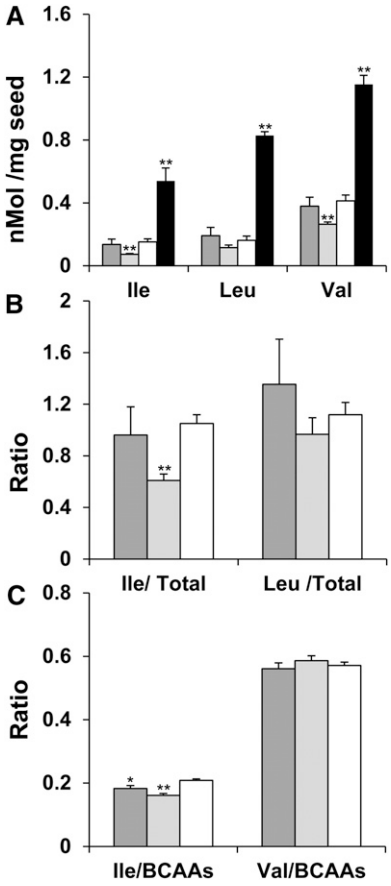

Figure 5. Complementation of *bcat2-1* with the Bay and Sha Alleles.

Absolute levels of free BCAAs **(A)** and ratios of Ile/total and Leu/total **(B)** and Ile/BCAAs and Val/BCAAs **(C)** are shown for dry seeds of Col-0 (white bars), *bcat2-1* (black bars), *bcat2-1*/*BCAT2*-(Sha) (dark gray bars), and *bcat2-1*/*BCAT2*-(Bay) (light gray bars). The BCAA levels of *bcat2-1*/*BCAT2*-(Bay) were significantly lower than those of Col-0, while *bcat2-1*/*BCAT2*-(Sha) levels were indistinguishable from those of Col-0 except for Ile/BCAAs. Means and SE were calculated from 9 independent transformation events for *bcat2-1*/*BCAT2*-(Bay), 10 for *bcat2-1*/*BCAT2*-(Sha), and 11 for Col-0 plants. *P <0.05, **P < 3E-03 calculated from Student's *t* test compared with Col-0.

amino acid profiles, development, and germination were also observed for the BCAA catabolic enzymes hydroxymethylglutaryl CoA lyase (HML1) and 3-methylcrotonyl-CoA (MCCase α- and β-subunits; Lu et al., 2011; Ding et al., 2012). It was even suggested that seed viability requires one of the BCAA catabolic products (Ding et al., 2012). These examples highlight the interconnectivity, importance, and unique role of BCAA metabolism in seeds versus in other tissues. To better understand the regulation of BCAA metabolism in seeds, we used a combination of association mapping, linkage analysis, and transgenic approaches to study these traits in *Arabidopsis* seeds.

Evaluation of natural variation for dry seed BCAAs and their relationships across an *Arabidopsis* diversity panel showed that Ile levels vary the most and Val levels vary the least, with ranges of 4.5- and 2.4-fold, respectively (Table 1). The same holds true for their relative compositions (as a percentage of total amino acids), although the range of variation is generally narrower. This may reflect the fact that all BCAAs share four biosynthetic enzymes and are coordinately influenced by regulatory feedback loops (Figure 1; Binder et al., 2007; Binder, 2010). All BCAA traits demonstrated moderate to high broad sense heritability, indicating that trait variation is mostly governed by genetic factors, rather than environmental factors during seed desiccation (Table 1). The strong positive correlations between absolute BCAA levels and their relative compositions (see Supplemental Table 1 online) are consistent with previous reports (Schauer et al., 2006; Gu et al., 2010; Toubiana et al., 2012), but the strong negative correlations between Val/BCAA and Ile/BCAA or Leu/BCAA (see Supplemental Table 1 online) had not been observed previously. This suggests that some genetic components

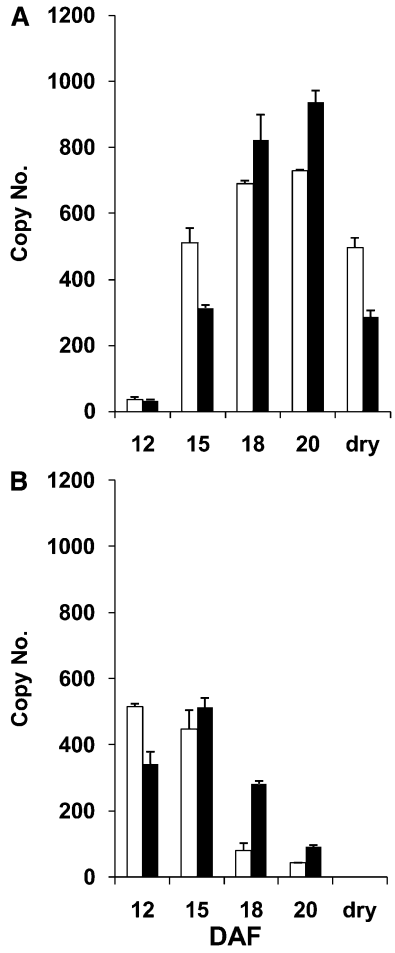

Figure 6. *BCAT1* and *BCAT2* Transcript Levels during Seed Maturation and Desiccation.

Transcript levels for *BCAT2* **(A)** and *BCAT1* **(B)** were measured from Bay (white bars) and Sha (black bars) accessions over five time points of seed maturation and desiccation using quantitative PCR with *ACTIN2* mRNA as an internal control. Error bars represent SE (independent biological replicates; $n = 3$).

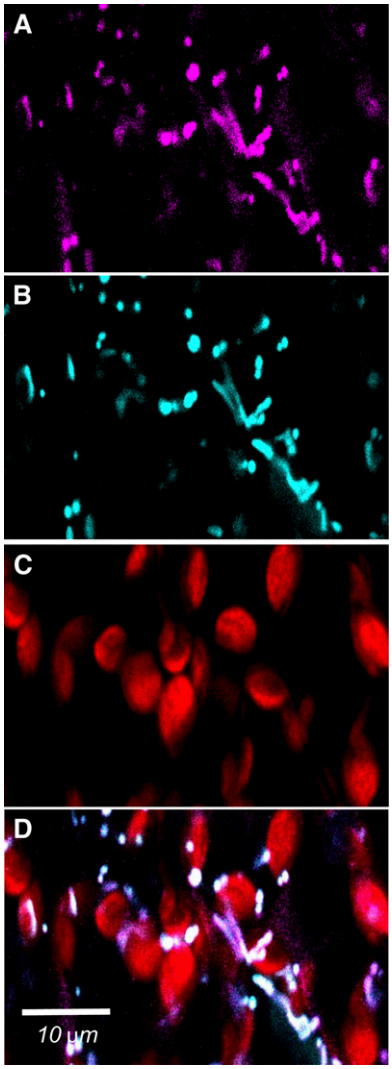

Figure 7. Colocalization of *BCAT2* with a Mitochondrial Organelle Marker.

YFP was fused in frame to the C terminus of *BCAT2*, and the construct was transformed into *Arabidopsis* plants expressing the CFP mitochondrial marker mt-ck CS16262 (Nelson et al., 2007).
(A) 35S:BCAT2-YFP is shown in magenta.
(B) Mitochondrial marker mt-ck CS16262 is shown in cyan.
(C) Chlorophyll autoflorescence is shown in red.
(D) Merge of **(A)**, **(B)**, and **(C)**.

influencing BCAA natural variation differentially regulate Ile and Leu relative to Val. Such observations demonstrate the usefulness of correlating relative compositions and absolute levels within an amino acid family, as it can allow a better understanding of the metabolic dynamics between different components.

Our GWAS of the nine BCAA traits identified five traits with significant associations at 5% FDR: Ile, Ile/BCAAs, Ile/total, Leu/total, and Val/BCAAs (Figure 2; see Supplemental Data Set 2 online). The SNP with the most significant association for all five traits was located on chromosome 1 in a gene annotated as *ERD4*, which is adjacent to two known BCAA pathway genes, *BCAT1* and *BCAT2* (Figure 3; see Supplemental Figure 2 online; see Supplemental Data Set 2 online). The subsequent LD analysis indicated that SNP5373 is in strong LD with many SNPs in its surrounding region, including SNPs located within *BCAT1* and *BCAT2* (see Supplemental Figure 1 online; see Supplemental Data Set 2 online). Atwell et al. (2010) previously demonstrated that the most significant associations are not necessarily situated in the actual causal loci. Haploblock analysis of the 23.5-kb region surrounding the most significant SNP from GWAS, SNP5373, showed that, although it is located in the *ERD4* gene, SNP5373 is part of a haploblock (haploblock 10) that encompasses most of *BCAT2* (Figure 3; see Supplemental Figure 2 online), suggesting that SNP5373 may be tagging a causal polymorphism in *BCAT2*. In support of this hypothesis, association analysis with the 14 haploblocks and their respective haplotypes across this region indicated that the two most significant haploblocks, haploblocks 8 and 9, are not situated in *ERD4* (Figure 3; see Supplemental Table 3 online) but rather in the 3′ UTR of *BCAT1* and the promoter and first exon of *BCAT2*. Averages of relative BCAA compositions from accessions with the most significantly different haplotype pair for haploblocks 8 and 9 showed divergent regulation of Val/BCAAs versus Ile/BCAAs and Leu/BCAAs (Figure 3C). This result suggests that all five BCAA traits are genetically driven by polymorphism(s) at this location. Interestingly, the MLMM analysis (Segura et al., 2012) showed that only SNP5373 tags the large-effect polymorphism(s) shared with all the BCAA traits (see Supplemental Table 2 online). When GWAS was reconducted with SNP5373 included as a covariate, all other associations in this chromosome 1 region were no longer significant. Our observations for BCAA traits in *Arabidopsis* seed are similar to the large-effect loci reported (Riedelsheimer et al., 2012) for other metabolic traits in association analysis of maize leaves, including absolute levels of leaf Lys and Tyr. These findings led to the hypothesis that metabolic traits have less complex genetic architecture than developmental traits such as flowering time.

The QTL tagged by SNP5373 in our GWAS was independently confirmed using a Bay × Sha recombinant inbred line mapping population whose parental accessions contain the most significantly different haplotype pairs for haploblocks 8 and 9. This linkage analysis identified two major BCAA QTLs, one of which included the same chromosome 1 interval identified by GWAS (Table 2; see Supplemental Data Set 3 online). These linkage results were also in agreement with respect to the high- and low-level predictions of BCAA phenotypes made by GWAS for the two QTL parental accessions. Interestingly, the second major QTL interval on chromosome 5 contains *METHYLTHIOALKYLMALATE SYNTHASE1* (*MAM1*) and *MAM3* (see Supplemental Data Set 3 online), which encode enzymes that catalyze the first part of aliphatic glucosinolate biosynthesis from Met (Textor et al., 2004, 2007). As MAM3 can also catalyze the chain elongation of Ile-derived glucosinolates (Field et al., 2004; Textor et al., 2007) and the chromosome 5 QTL is most significant for Ile traits (see Supplemental Data Set 3 online), *MAM3* seems a likely candidate gene for this QTL. This chromosome 5 QTL and several other BCAA trait QTLs identified by

linkage analysis were not detected in our GWAS, which may be due to these QTLs being rare alleles specific to the mapping population or to limitations inherent to the GWAS approach (Kennedy et al., 1992; Pritchard and Donnelly, 2001; Yu et al., 2006).

To identify the causal locus, amino acid analysis of null mutants for the three genes in the vicinity of SNP5373 (*bcat1*, *bcat2*, and *erd4*) was performed. Only *bcat2* alleles had a significant and consistent influence on seed BCAA levels, with no significant effect in leaves (Figure 4; see Supplemental Data Set 4 online). The metabolic profile of *bcat2* mutants showing increased seed BCAA levels (Figure 4), along with BCAT2-YFP localization to the mitochondrion (Figure 7), strongly support a catabolic function for *BCAT2*. A similar catabolic role has been proposed for the adjacent *BCAT1* gene, which is also targeted to the mitochondrion (Schuster and Binder, 2005). Transcriptional analysis of *BCAT1* and *BCAT2* showed strong induction of *BCAT2* expression during late seed maturation and dramatic reduction in *BCAT1* mRNA to nondetected levels upon desiccation, both of which support the hypothesis that *BCAT2*, and not *BCAT1*, is the causal locus (Figure 6). These expression data were consistent with the public *Arabidopsis* gene expression data set, which indicates that *BCAT2* is expressed throughout the embryo (Figure 6; see Supplemental Figure 5 online) while *BCAT1* is specifically expressed in the seed coat, which dies at the desiccation stage (Figure 6; see Supplemental Figure 5 online). These contrasting transcriptional profiles and mutant phenotypes indicate that although both are catabolic, *BCAT1* and *BCAT2* are not functionally redundant. These data also suggest that *BCAT2* plays a key role in the metabolic transition during late seed maturation (Angelovici et al., 2010) and that *BCAT2* is the causal locus for the natural variation in BCAAs in dry *Arabidopsis* seeds.

The contribution of *BCAT2* to natural variation in seed amino acid metabolism was further assessed by molecular complementation analysis of the null *bcat2-1* genotype with genomic sequences encompassing the Bay and Sha *BCAT2* alleles, both of which significantly reduced BCAA levels relative to *bcat2-1* (Figure 5). Complementation with the Sha allele, the predicted high-level trait allele with respect to Ile-related traits, gave rise to seed BCAA levels that were indistinguishable from the wild-type Col-0 background, while complementation with the Bay allele, the predicted low-level trait allele, had significantly lower seed BCAA levels versus Col-0 (Figure 5). These data confirm that the most significantly different *BCAT2* haplotype pairs confer significantly different phenotypes, allowing us to conclude that *BCAT2* is the causal locus underlying the chromosome 1 QTL identified by both GWAS and linkage analysis. In addition, complementation with the Bay allele caused a significant reduction in seed Arg, Glu, Gly, and Ser. Glu accounts for ~30% of the total FAA pool (see Supplemental Table 5 online), and a 16% reduction in pool size could be biologically significant, as Glu is utilized in various metabolic processes (Forde and Lea, 2007). Mitochondrial catabolism of BCAAs was previously implicated in Glu recycling in plants (Lapujade et al., 1998; Li et al., 2003). In animals, a direct link with Glu recycling has been established, as a mitochondrial BCAT is part of a complex that includes glutamate dehydrogenase and takes part in cycling ammonia through glutamate (Islam et al., 2010). These results are consistent with the catabolic function of *BCAT2* contributing to metabolic homeostasis during seed maturation in an allele-dependent manner.

The molecular mechanism(s) underlying the phenotypic differences in Bay and Sha *BCAT2* alleles are very hard to determine, as these accessions are polymorphic for 35 SNPs and 10 insertion/deletion events (see Supplemental Table 4 online). Seven of the nine SNPs scored in the association panel are significant for one or more BCAA traits and include the four SNPs that define the most significant haplotypes for haploblocks 8 and 9 (Figure 3; see Supplemental Table 4 online). Three of these four SNPs are located in the promoter region, and one encodes a nonsynonymous amino acid change in the mitochondrial targeting sequence of the encoded protein. Thus, the overall phenotypic differences between Bay and Sha are likely due to a combination of transcriptional and protein targeting/activity differences. It is also important to note that GWAS never scores all polymorphisms present in a panel or even between two ecotypes (e.g., Bay and Sha), and it is also likely that one or more of the remaining 36 unscored polymorphisms, including two that alter amino acids in the mature BCAT2 protein, also contribute to phenotypic differences between the two ecotypes (see Supplemental Table 4 online).

Previous QTL studies of BCAA pathway genes in cultivated tomato/*S. pennellii* introgression lines identified 26 intervals affecting one or more BCAA traits in tomato fruit pericarp, of which 10 intervals contained candidate BCAA biosynthetic and catabolic enzymes (Schuster et al., 2006; Schauer et al., 2008; Maloney et al., 2010; Kochevenko and Fernie, 2011). Of the candidates tested, two BCAA biosynthetic genes, *ISOPROPYLMALATE DEHYDRATASE* and *BCAT4*, and one catabolic BCAA gene, *BCAT1*, were demonstrated by overexpression or antisense experiments to affect BCAA levels in tomato fruit pericarp. Surprisingly, seed BCAA QTLs from the same introgression population did not overlap with pericarp BCAA QTLs (Toubiana et al., 2012), suggesting different genetic architectures for these traits in pericarp and seeds. Linkage analysis of the *Arabidopsis* Bay × Sha population yielded 12 intervals affecting one or more seed BCAA traits, of which three contained candidate BCAA catabolic enzymes and none contained BCAA biosynthetic enzymes (see Supplemental Data Set 3 online). Thus, GWAS results were also consistent with amino acid catabolism, rather than biosynthesis, playing a key role in determining free BCAA levels during seed maturation on a population-wide level (Zhu and Galili, 2003; Gu et al., 2010; Lu et al., 2011; Ding et al., 2012; Kirma et al., 2012).

The somewhat surprising evolutionary selection of BCAT2 as an important regulator of natural variation for seed BCAAs may be due to several factors. First, BCAT2 is located in the mitochondrion and, therefore, is very likely the first and regulatory step for seed BCAA catabolism, a key determinant of FAA levels in seed. Second, among the seven *Arabidopsis* BCAT family members, BCAT2 uniquely shows strongly induced expression in late seed maturation. Finally, unlike null mutants of other BCAA catabolic enzymes, which have major pleotropic effects on seed amino acid levels (e.g., IVD, HML1, and MCCase), *bcat2* null mutants do not show significant pleiotropy and specifically

affect seed BCAA levels. These unique characteristics of *Arabidopsis* BCAT2 suggest that BCAT2 orthologs from crop plants should be considered as candidates for transgenic and/or breeding approaches to positively affect the levels of these three essential amino acids in grain used for feed and food.

METHODS

Plant Growth and Seed Collection

All *Arabidopsis thaliana* genotypes were grown at 18 to 21°C (day/night) under long-day conditions (16 h of light/8 h of dark). Dry seeds were collected at the end of the desiccation period and stored in a desiccator at room temperature for at least 6 weeks prior to analysis to ensure full desiccation. The 360-member *Arabidopsis* diversity panel (Nordborg et al., 2005) was grown as three independent replicates under the conditions above. Seeds from a total of 313 accessions with reasonable flowering times were used for association analysis. For the developing seed analysis, flowers of Bay and Sha were marked, and at specific times (12 ± 1, 15 ± 1, 18 ± 1, and 20 ± 1 DAF) seeds were collected and stored at −80°C. All accessions in this panel were genotyped with the 250K SNP data set version 3.06, which is available for download at https://cynin.gmi.oeaw.ac.at/home/resources/atpolydb/.

Isolation of T-DNA Insertion Mutants and Genotypic Characterization

The T-DNA mutant lines FLAG_093BO2 (*bcat1-2*) and FLAG_353G11 (*bcat2-2*) were obtained from the Versailles collection (Institut National de la Recherche Agronomique). The T-DNA mutant lines SALK_13630 (*bcat1-1*), SALK_037854 (*bcat2-1*), and SALK_131951C (*erd4*) were obtained from the ABRC stock center. *bcat2-1*, *bcat1-1*, and *erd4* are in the Col-0 background, and *bcat1-2* and *bcat2-2* are in the WS-4 background. Homozygous mutant lines were validated by genomic PCR using gene-specific primers in combination with the T-DNA left border primer. Lack of transcripts from *BCAT1*, *BCAT2*, and *ERD4* was validated by RT-PCR using flower and seed RNA from the respective mutants and primers designed to amplify the full-length transcript. Primers used for this study are summarized in Supplemental Table 6 online.

Complementation of *bcat2-1* with Genomic Fragments of *BCAT2*

A 3.81-kb region of genomic DNA containing the *BCAT2* gene (chromosome 1; 3,286,586 to 3,290,396) with 1500 bp prior to the start codon and 230 bp after the stop codon were amplified (primers are summarized in Supplemental Table 6 online) from the Bay and Sha accessions and cloned into the pMDC99 vector using the Gateway recombination system (Invitrogen). Homozygous *bcat2-1* plants were transformed with either the Bay or Sha allele by *Agrobacterium tumefaciens*–mediated gene transfer using the floral dip method (Clough and Bent, 1998). Homozygous lines segregating for a single insertion were isolated by selection on the appropriate antibiotics (9 independent lines for the Bay genomic allele and 10 for the Sha allele).

BCAT1 and *BCAT2* Transcript Analysis

Total RNA was isolated from developing and dry seeds using the hot-borate method (Birtić and Kranner, 2006) and from flowers using RNA-easy (Qiagen). DNase treatment was conducted using TURBO DNA-free DNase (Ambion). First-strand cDNA was synthesized from 1 μg of total RNA with SuperScript II H^- reverse transcriptase (Invitrogen) and oligo(dT) primer. Transcript levels were determined by quantitative PCR using SYBR green real-time PCR mix (Applied Biosystems), with *ACTIN2* (At3G18780) mRNA as an internal control (for primers, see Supplemental Table 6 online).

Subcellular Localization of BCAT2

BCAT2 mRNA was amplified (primers are summarized in Supplemental Table 6 online) from cDNA, cloned into the pEarleyGate101 vector (Earley et al., 2006), and mobilized into a vector driven by the 35S promoter that allowed in-frame fusion of the *BCAT2* coding sequence to the YFP coding region. This vector was used to transform plants containing a CFP mitochondrial marker (CS16262; Nelson et al., 2007) by *A. tumefaciens*–mediated gene transfer using the floral dip method (Clough and Bent, 1998). Localization of expressed proteins was visualized in T1 transformants with a 100XUPLSAPO oil objective NAL4 using an Olympus Fluview FV1000 confocal laser-scanning microscope. YFP was excited by a 515-nm argon laser line, and the emission range was set to 530 to 580 nm; CFP was excited by a 458-nm argon laser line, and the emission range was set to 475 to 500 nm. Chlorophyll autoflorescence was excited by a 515-nm argon laser line, with the emission range set to 655 to 755 nm. After background subtraction to remove camera noise, contrast enhancement was performed to increase signal intensity and remove low-level background. For colocalization experiments, YFP signals were false colored magenta, CFP signals were false colored cyan, and chlorophyll signals were false colored red.

Extraction and Analysis of *Arabidopsis* Seed Amino Acids Using LC- MS/MS

Plant Extraction and Analysis

FAA extraction for LC-MS/MS analysis was performed by modifying a previously described method (Gu et al., 2007). Briefly, ~5-mg *Arabidopsis* seed samples were homogenized in 400-μL heavy standards buffer (see components below) by vigorous shaking with two 3-mm glass beads for 5 min on an S2200 paint shaker (Hero Products Group) and then incubated for 10 min at 90°C. After incubation, extracts were centrifuged at 3700*g* and the supernatants were filtered by centrifugation at 3000*g* through a 0.45-μm low binding hydrophilic polytetrafluoroethylene (Millipore). Filtered samples were then diluted twofold with the heavy standards buffer and analyzed by LC-MS/MS. The LC-MS/MS method was slightly modified from Gu et al. (2007) to include selected ion pairs for the additional heavy amino acid standards. The method was composed of three electrospray photoionization (positive ion mode) functions (0 to 1.6, 1.61 to 2.3, and 2.31 to 5.6 min) to allow sufficient dwell time for each analyte. Multiple reaction monitoring transitions, optimized source cone voltages, collision cell voltages, gradient, functions, and selected ion monitoring pair details are described in detail in Supplemental Table 7 online. Data were acquired and processed with MassLynx 4.0.

Amino Acid Heavy Standards Buffer

The buffer contained 19 μM DTT with 8 μM Ala-D4 and Asp-D4, 12 μM Gly-D2-$^{15}N_1$, and 4 μM each Ala-D4, Asp-D3, Gln-U-$^{13}C_5$, Glu-D3, Gly-D2-$^{15}N_1$, Leu-D10, Lys-$^{15}N_1$, His-D3, Met-D3, Phe-D8, Ser-D3, Trp-D5, and Val-D8.

All the heavy amino acids were purchased from Cambridge Isotope Laboratories and C/D/N Isotope and will be described as AA* hereafter; normal isotope amino acids will be described as AA.

Standard Curve

Endogenous concentrations of the 18 AAs were calculated according to standard curves and the peak area ratios of each amino acid to its respective heavy internal standard (performed in MassLynx 4.0). Hence, a series of working standards with concentrations ranging from 0 to 100 μM (with the exception of 300 μM in the case of Gly) of 18 AAs were

prepared with the AA* heavy standard buffer (all 18 standard AAs were purchase from Sigma-Aldrich). Calibration curves for each amino acid were generated by plotting amino acid concentrations on the *x* axis versus responses (AA peak area or peak area ratio of AA to AA*) on the *y* axis using regression analysis. Measurement of triplicates demonstrated a high accuracy of ~15% relative SD (SD × 100/repeats average measurement), high repeatability (~90–95% repeatability for FAAs across the measured *Arabidopsis* association panel), and a wide linear range (1 to 1000 ng). This high accuracy of the method allowed only a single measurement per each accession per growth period for GWAS, although five replicates were routinely performed for the characterization of mutants.

QTL Analysis

Amino acid levels were quantified, different relative levels were calculated as described above, and raw values were used as quantitative values in QTL analysis. PLABQTL was used for composite interval mapping (Utz and Melchinger, 1996). Permutation analysis (1000) was performed to calculate the critical log of the odds score ($\alpha = 0.05$) of 3.11. Genotypic data used for the analysis were obtained from Loudet et al. (2002; http://dbsgap.versailles.inra.fr/vnat/Documentation/33/DOC.html). Cofactors used for calculations were automatically chosen by the PLABQTL program.

Data Analysis and GWAS

For each of the nine BCAA traits, samples below detection were assigned uniform random variables ranging from 0 to *min*, where *min* is the minimum detection level of a given trait (Lubin et al., 2004). Studentized deleted residuals (Kutner et al., 2004) were then calculated from a mixed model fitted in SAS version 9.2 (SAS Institute) that included replicate and accessions as random effects. Accessions with extreme Studentized deleted residuals were designated as outliers and removed from the analysis. Upon removal of the outliers, the procedure of Box and Cox (1964) was implemented to find the optimal transformation of each trait, thus ensuring that the model assumptions of normally distributed error terms and constant variance were not being violated. Finally, a best linear unbiased predictor (BLUP) for each accession was predicted for all transformed traits using mixed models fitted across all outgrowths. These BLUPs were used as the phenotype in all subsequent analyses. The variance component estimates from each fitted model was used to estimate the broad-sense heritability of each trait (Holland et al., 2003; Hung et al., 2012). SE values of heritability estimates were approximated using the delta method (Holland et al., 2003).

A GWAS of the nine BCAA traits was then conducted using 170,344 SNPs with MAF $\geq$ 0.05 among 313 accessions from the diversity panel with adequate flowering time to allow timely seed production. At each of these SNPs, a unified mixed linear model (Yu et al., 2006) was fitted for each trait in the Genome Association and Prediction Integrated Tool R package (Lipka et al., 2012). This model included principal components as fixed effects to account for population structure (Price et al., 2006) and a kinship matrix (Loiselle et al., 1995) to account for familial relatedness among the accessions. The principal components and kinship matrix were both calculated from 214,052 SNPs with MAF $\geq$ 0.10 among the 1307 accessions genotyped with the 250K SNP data set version 3.06. For each BCAA trait, the Bayesian information criterion (Schwarz, 1978) was used to determine the optimal number of principal components to include in the MLM. The efficient population parameters previously determined (Zhang et al., 2010) were used to eliminate the need of reestimating the variance components at each SNP. Finally, we used the procedure to control for the FDR at 5% (Benjamini and Hochberg, 1995).

In addition to the GWAS, we implemented the MLMM (Segura et al., 2012) to identify the key SNPs responsible for signals at major-effect loci. To account for multiple markers associated with a trait, the MLMM approach conducts stepwise model selection within a mixed-model framework. In contrast to the population parameters previously determined, the MLMM approach reestimates the variance components at each step of model fitting. All 170,344 SNPs from the 250K chip with MAF $\geq$ 0.05 among the 313 accessions were considered for inclusion in the final model. The extended Bayesian information criterion (Chen and Chen, 2008) was used to determine the number of SNPs in this optimal model.

Correlation Analysis

To assess the correlation between the BCAA traits in their original units of measurement, the BLUPs of each trait were back transformed. The Spearman's rank correlation coefficient was then used to estimate the correlation between each pair of traits. The statistical significance of each correlation was assessed by conducting a Student's *t* test.

LD

LD was investigated using previously described methods. Briefly, the squared allele frequency correlation (r^2) between pairs of SNPs was used to quantify LD in a given genomic region (Weir and Hill, 1986) . All SNPs with MAF $\geq$ 0.05 were considered for this analysis.

Haplotype Analysis

For the 23.5-kb region surrounding *BCAT1* and *BCAT2*, haploblocks were created using the confidence interval method in Haploview version 4.2 (Barrett et al., 2005). The associations between haploblocks and BCAA content were then analyzed by fitting the unified MLM (Yu et al., 2006) in the MIXED procedure available in SAS version 9.3 (SAS Institute). For this analysis, the same principal components and kinship matrix used in our GWAS were used to account for population structure and familial relatedness. Within each haploblock, the LSMEANS statement was used to compare BCAA levels at each haplotype. The Tukey–Kramer procedure (Kutner et al., 2004) was used to adjust for the multiple testing problem at $\alpha = 0.05$.

Accession Numbers

Sequence data from this article can be found in the Arabidopsis Genome Initiative or GenBank/EMBL databases under the following accession numbers: *BCAT2*, At1G10070; *BCAT1*, At1G10060; *ERD4*, At1G10090.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Pairwise LD r^2 Estimates of SNP5373 with SNPs across a 200-kb Chromosomal Interval Centered on SNP5373 and Plotted against Chromosomal Position.

Supplemental Figure 2. Haplotype Analysis of Chromosome 1, 3,274,080 to 3,297,645.

Supplemental Figure 3. *BCAT1* and *BCAT2* Genomic Structures and Transcript Levels in Seeds and Flowers of *bcat1* and *bcat2* Mutants.

Supplemental Figure 4. *ERD4* Genomic Structure and Transcripts in *erd4*.

Supplemental Figure 5. Expression in Absolute Levels for *BCAT1* and *BCAT2* as Reported at the *Arabidopsis* eFP Browser.

Supplemental Table 1. Spearman's Correlation Analysis of BCAA Traits across the *Arabidopsis* Diversity Panel.

Supplemental Table 2. Summary of MLMM Analysis for the Five BCAA Traits Using All SNPs within ±100 kb of the SNP5373 Genomic Region.

Supplemental Table 3. Haploview-Generated Haplotypes with Their Respective Association P Values and Significantly Different Haplotypes That Were Found across All Five Traits.

Supplemental Table 4. Summary of the Polymorphisms between the Two Alleles from Bay and Sha That Were Used for the *bcat2-1* Complementation, as Determined by Sanger Sequence.

Supplemental Table 5. FAA Profile of *bcat2-1* Complementation with the Two Significantly Different Alleles from Bay and Sha.

Supplemental Table 6. List of the Primers Used for This Study.

Supplemental Table 7. Details of LC-MS/MS Gradient and Functions.

Supplemental Data Set 1. *Arabidopsis* Accessions Used in This Study.

Supplemental Data Set 2. GWAS Results Summary of all SNPs with Significant Associations at 5% FDR for All the BCAA Traits.

Supplemental Data Set 3. QTL Analysis of the Bay × Sha Mapping Population.

Supplemental Data Set 4. Seed and Leaf FAA Profiles of *bcat1-1*, *bcat1-2*, *bcat2-1*, *bcat2-2*, *erad4*, and the Wild Type (Col-0).

ACKNOWLEDGMENTS

We thank Maria Magallanes-Lundback for growth and handling of the *Arabidopsis* association panels and QTL populations. This research was supported by the U.S.–Israel Binational Agricultural Research and Development Fund (postdoctoral award ALTF 29–2011 to R.A.) and the National Science Foundation (grant 0922493 to D.D.).

AUTHOR CONTRIBUTIONS

R.A. designed and performed the research and wrote the article. N.D. performed research. A.E.L. and S.G.-J. performed data analysis. R.B., H.L., and J.C. contributed new computational tools for data analysis. M.A.G. designed research and performed data analysis. D.D. designed research and wrote the article.

Received October 6, 2013; revised November 26, 2013; accepted December 10, 2013; published December 24, 2013.

REFERENCES

Angelovici, R., Galili, G., Fernie, A.R., and Fait, A. (2010). Seed desiccation: A bridge between maturation and germination. Trends Plant Sci. **15:** 211–218.

Araújo, W.L., Ishizaki, K., Nunes-Nesi, A., Larson, T.R., Tohge, T., Krahnert, I., Witt, S., Obata, T., Schauer, N., Graham, I.A., Leaver, C.J., and Fernie, A.R. (2010). Identification of the 2-hydroxyglutarate and isovaleryl-CoA dehydrogenases as alternative electron donors linking lysine catabolism to the electron transport chain of Arabidopsis mitochondria. Plant Cell **22:** 1549–1563.

Atwell, S., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature **465:** 627–631.

Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., and Miyano, S. (2002). Extensive feature detection of N-terminal protein sorting signals. Bioinformatics **18:** 298–305.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics **21:** 263–265.

Beck, H.C., Hansen, A.M., and Lauritsen, F.R. (2004). Catabolism of leucine to branched-chain fatty acids in *Staphylococcus xylosus*. J. Appl. Microbiol. **96:** 1185–1193.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. B **57:** 289–300.

Binder, S. (2010). Branched-chain amino acid metabolism in *Arabidopsis thaliana*. The Arabidopsis Book **8:** e0137. doi/10.1199/tab.0137.

Binder, S., Knill, T., and Schuster, J. (2007). Branched-chain amino acid metabolism in higher plants. Physiol. Plant. **129:** 68–78.

Birtić, S., and Kranner, I. (2006). Isolation of high-quality RNA from polyphenol-, polysaccharide- and lipid-rich seeds. Phytochem. Anal. **17:** 144–148.

Box, G.E.P., and Cox, D.R. (1964). An analysis of transformations. J. R. Stat. Soc. Ser. B Stat. Methodol. **26:** 211–252.

Campbell, M.A., Patel, J.K., Meyers, J.L., Myrick, L.C., and Gustin, J.L. (2001). Genes encoding for branched-chain amino acid aminotransferase are differentially expressed in plants. Plant Physiol. Biochem. **39:** 855–860.

Chan, E.K.F., Rowe, H.C., Corwin, J.A., Joseph, B., and Kliebenstein, D.J. (2011). Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. PLoS Biol. **9:** e1001125.

Chen, H., Saksa, K., Zhao, F., Qiu, J., and Xiong, L. (2010). Genetic analysis of pathway regulation for enhancing branched-chain amino acid biosynthesis in plants. Plant J. **63:** 573–583.

Chen, J., and Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. Biometrika **95:** 759–771.

Chuang, D.T., Chuang, J.L., and Wynn, R.M. (2006). Lessons from genetic disorders of branched-chain amino acid metabolism. J. Nutr. **136** (suppl.)**:** 243S–249S.

Clough, S.J., and Bent, A.F. (1998). Floral dip: A simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J. **16:** 735–743.

Cook, J.P., McMullen, M.D., Holland, J.B., Tian, F., Bradbury, P., Ross-Ibarra, J., Buckler, E.S., and Flint-Garcia, S.A. (2012). Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol. **158:** 824–834.

D'Antona, G., Ragni, M., Cardile, A., Tedesco, L., Dossena, M., Bruttini, F., Caliaro, F., Corsetti, G., Bottinelli, R., Carruba, M.O., Valerio, A., and Nisoli, E. (2010). Branched-chain amino acid supplementation promotes survival and supports cardiac and skeletal muscle mitochondrial biogenesis in middle-aged mice. Cell Metab. **12:** 362–372.

Däschner, K., Thalheim, C., Guha, C., Brennicke, A., and Binder, S. (1999). In plants a putative isovaleryl-CoA-dehydrogenase is located in mitochondria. Plant Mol. Biol. **39:** 1275–1282.

Diebold, R., Schuster, J., Däschner, K., and Binder, S. (2002). The branched-chain amino acid transaminase gene family in Arabidopsis encodes plastid and mitochondrial proteins. Plant Physiol. **129:** 540–550.

Ding, G., Che, P., Ilarslan, H., Wurtele, E.S., and Nikolau, B.J. (2012). Genetic dissection of methylcrotonyl CoA carboxylase indicates a complex role for mitochondrial leucine catabolism during seed development and germination. Plant J. **70:** 562–577.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. Plant J. **45:** 616–629.

Ellerström, M., Josefsson, L.G., Rask, L., and Ronne, H. (1992). Cloning of a cDNA for rape chloroplast 3-isopropylmalate dehydrogenase by genetic complementation in yeast. Plant Mol. Biol. **18:** 557–566.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. **300:** 1005–1016.

Engqvist, M., Drincovich, M.F., Flügge, U.-I., and Maurino, V.G. (2009). Two D-2-hydroxy-acid dehydrogenases in *Arabidopsis thaliana* with catalytic capacities to participate in the last reactions of the methylglyoxal and beta-oxidation pathways. J. Biol. Chem. **284:** 25026–25037.

Fait, A., Angelovici, R., Less, H., Ohad, I., Urbanczyk-Wochniak, E., Fernie, A.R., and Galili, G. (2006). Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. Plant Physiol. **142:** 839–854.

Falco, S.C., Guida, T., Locke, M., Mauvais, J., Sanders, C., Ward, R.T., and Webber, P. (1995). Transgenic canola and soybean seeds with increased lysine. Biotechnology (N. Y.) **13:** 577–582.

Field, B., Cardon, G., Traka, M., Botterman, J., Vancanneyt, G., and Mithen, R. (2004). Glucosinolate and amino acid biosynthesis in Arabidopsis. Plant Physiol. **135:** 828–839.

Forde, B.G., and Lea, P.J. (2007). Glutamate in plants: Metabolism, regulation, and signalling. J. Exp. Bot. **58:** 2339–2358.

Frizzi, A., Huang, S., Gilbertson, L.A., Armstrong, T.A., Luethy, M.H., and Malvar, T.M. (2008). Modifying lysine biosynthesis and catabolism in corn with a single bifunctional expression/silencing transgene cassette. Plant Biotechnol. J. **6:** 13–21.

Gao, F., Wang, C., Wei, C., and Li, Y. (2009). A branched-chain aminotransferase may regulate hormone levels by affecting KNOX genes in plants. Planta **230:** 611–623.

Gonda, I., Bar, E., Portnoy, V., Lev, S., Burger, J., Schaffer, A.A., Tadmor, Y., Gepstein, S., Giovannoni, J.J., Katzir, N., and Lewinsohn, E. (2010). Branched-chain and aromatic amino acid catabolism into aroma volatiles in Cucumis melo L. fruit. J. Exp. Bot. **61:** 1111–1123.

Gu, L., Jones, A.D., and Last, R.L. (2007). LC-MS/MS assay for protein amino acids and metabolically related compounds for large-scale screening of metabolic phenotypes. Anal. Chem. **79:** 8067–8075.

Gu, L., Jones, A.D., and Last, R.L. (2010). Broad connections in the Arabidopsis seed metabolic network revealed by metabolite profiling of an amino acid catabolism mutant. Plant J. **61:** 579–590.

Hannah, M.A., Caldana, C., Steinhauser, D., Balbo, I., Fernie, A.R., and Willmitzer, L. (2010). Combined transcript and metabolite profiling of Arabidopsis grown under widely variant growth conditions facilitates the identification of novel metabolite-mediated regulation of gene expression. Plant Physiol. **152:** 2120–2129.

Holland, J. B., Nyquist, W. E., and Cervantes-Martínez, C. T. (2003). Estimating and interpreting heritability for plant breeding: An update. Plant Breeding Reviews **22:** 9–112.

Horton, M.W., et al. (2012). Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. Nat. Genet. **44:** 212–216.

Hung, H.Y., et al. (2012). The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. Heredity **108:** 490–499.

Ishizaki, K., Larson, T.R., Schauer, N., Fernie, A.R., Graham, I.A., and Leaver, C.J. (2005). The critical role of *Arabidopsis* electron-transfer flavoprotein:ubiquinone oxidoreductase during dark-induced starvation. Plant Cell **17:** 2587–2600.

Ishizaki, K., Schauer, N., Larson, T.R., Graham, I.A., Fernie, A.R., and Leaver, C.J. (2006). The mitochondrial electron transfer flavoprotein complex is essential for survival of Arabidopsis in extended darkness. Plant J. **47:** 751–760.

Islam, M.M., Nautiyal, M., Wynn, R.M., Mobley, J.A., Chuang, D.T., and Hutson, S.M. (2010). Branched-chain amino acid metabolon: Interaction of glutamate dehydrogenase with the mitochondrial branched-chain aminotransferase (BCATm). J. Biol. Chem. **285:** 265–276.

Kandra, G., Severson, R., and Wagner, G.J. (1990). Modified branched-chain amino acid pathways give rise to acyl acids of sucrose esters exuded from tobacco leaf trichomes. Eur. J. Biochem. **188:** 385–391.

Kennedy, B.W., Quinton, M., and van Arendonk, J.A. (1992). Estimation of effects of single genes on quantitative traits. J. Anim. Sci. **70:** 2000–2012.

Kimball, S.R., and Jefferson, L.S. (2006a). New functions for amino acids: Effects on gene transcription and translation. Am. J. Clin. Nutr. **83:** 500S–507S.

Kimball, S.R., and Jefferson, L.S. (2006b). Signaling pathways and molecular mechanisms through which branched-chain amino acids mediate translational control of protein synthesis. J. Nutr. **136** (suppl.)**:** 227S–231S.

Kirma, M., Araújo, W.L., Fernie, A.R., and Galili, G. (2012). The multifaceted role of aspartate-family amino acids in plant metabolism. J. Exp. Bot. **63:** 4995–5001.

Knill, T., Schuster, J., Reichelt, M., Gershenzon, J., and Binder, S. (2008). Arabidopsis branched-chain aminotransferase 3 functions in both amino acid and glucosinolate biosynthesis. Plant Physiol. **146:** 1028–1039.

Kochevenko, A., and Fernie, A.R. (2011). The genetic architecture of branched-chain amino acid accumulation in tomato fruits. J. Exp. Bot. **62:** 3895–3906.

Kochevenko, A., Klee, H.J., Fernie, A.R., and Araújo, W.L. (2012a). Molecular identification of a further branched-chain aminotransferase 7 (BCAT7) in tomato plants. J. Plant Physiol. **169:** 437–443.

Kochevenko, A., Araújo, W.L., Maloney, G.S., Tieman, D.M., Do, P.T., Taylor, M.G., Klee, H.J., and Fernie, A.R. (2012b). Catabolism of branched chain amino acids supports respiration but not volatile synthesis in tomato fruits. Mol. Plant **5:** 366–375.

Kochevenko, A.S., and Fernie, A.R. (2012). [Investigation of tomato (*Solanum lycopersicum*) aminotransferases involved in biosynthesis of branched-chain-amino-acids]. Ukrain. Biochem. J. **84:** 18–25.

Kroumova, A.B., Xie, Z.Y., and Wagner, G.J. (1994). A pathway for the biosynthesis of straight and branched, odd- and even-length, medium-chain fatty acids in plants. Proc. Natl. Acad. Sci. USA **91:** 11437–11441.

Kutner, M.H., Nachtsheim, C.J., and Neter J., Li W. (2004). Applied Linear Statistical Models, 4th ed.(Boston, MA: McGraw Hill).

Lapujade, P., Cocaign-Bousquet, M., and Loubiere, P. (1998). Glutamate biosynthesis in *Lactococcus lactis* subsp. *lactis* NCDO 2118. Appl. Environ. Microbiol. **64:** 2485–2489.

Less, H., and Galili, G. (2008). Principal transcriptional programs regulating plant amino acid metabolism in response to abiotic stresses. Plant Physiol. **147:** 316–330.

Less, H., and Galili, G. (2009). Coordinations between gene modules control the operation of plant amino acid metabolic networks. BMC Syst. Biol. **3:** 14.

Less, H., Angelovici, R., Tzin, V., and Galili, G. (2010). Principal transcriptional regulation and genome-wide system interactions of the Asp-family and aromatic amino acid networks of amino acid metabolism in plants. Amino Acids **39:** 1023–1028.

Li, L., Thipyapong, P., Breeden, D.C., and Steffens, J.C. (2003). Overexpression of a bacterial branched-chain alpha-keto acid dehydrogenase complex in Arabidopsis results in accumulation of

branched-chain acyl-CoAs and alteration of free amino acid composition in seeds. Plant Sci. **165:** 1213–1219.

Li, Q., Yang, X., Xu, S., Cai, Y., Zhang, D., Han, Y., Li, L., Zhang, Z., Gao, S., Li, J., and Yan, J. (2012). Genome-wide association studies identified three independent polymorphisms associated with α-tocopherol content in maize kernels. PLoS ONE **7:** e36807.

Li, Y., Huang, Y., Bergelson, J., Nordborg, M., and Borevitz, J.O. (2010). Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **107:** 21199–21204.

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S., and Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. Bioinformatics **28:** 2397–2399.

Lipka, A.E., Gore, M.A., Magallanes-Lundback, M., Mesberg, A., Lin, H., Tiede, T., Chen, C., Buell, C.R., Buckler, E.S., Rocheford, T., and DellaPenna, D. (2013). Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. G3 (Bethesda) **3:** 1287–1299.

Loiselle, B.A., Sork, V.L., Nason, J., and Graham, C. (1995). Spatial genetic-structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). Am. J. Bot. **82:** 1420–1425.

Loudet, O., Chaillou, S., Camilleri, C., Bouchez, D., and Daniel-Vedele, F. (2002). Bay-0 x Shahdara recombinant inbred line population: A powerful tool for the genetic dissection of complex traits in Arabidopsis. Theor. Appl. Genet. **104:** 1173–1184.

Lu, Y., Savage, L.J., Larson, M.D., Wilkerson, C.G., and Last, R.L. (2011). Chloroplast 2010: a database for large-scale phenotypic screening of Arabidopsis mutants. Plant Physiol. **155:** 1589–1600.

Lubin, J.H., Colt, J.S., Camann, D., Davis, S., Cerhan, J.R., Severson, R.K., Bernstein, L., and Hartge, P. (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. Environ. Health Perspect. **112:** 1691–1696.

Maloney, G.S., Kochevenko, A., Tieman, D.M., Tohge, T., Krieger, U., Zamir, D., Taylor, M.G., Fernie, A.R., and Klee, H.J. (2010). Characterization of the branched-chain amino acid aminotransferase enzyme family in tomato. Plant Physiol. **153:** 925–936.

Matich, A., and Rowan, D. (2007). Pathway analysis of branched-chain ester biosynthesis in apple using deuterium labeling and enantioselective gas chromatography-mass spectrometry. J. Agric. Food Chem. **55:** 2727–2735.

Mazur, B., Krebbers, E., and Tingey, S. (1999). Gene discovery and product development for grain quality traits. Science **285:** 372–375.

Mentzen, W.I., Peng, J., Ransom, N., Nikolau, B.J., and Wurtele, E.S. (2008). Articulation of three core metabolic processes in Arabidopsis: Fatty acid biosynthesis, leucine catabolism and starch metabolism. BMC Plant Biol. **8:** 76.

Mertz, E.T., Bates, L.S., and Nelson, O.E. (1964). Mutant gene that changes protein composition and increases lysine content of maize endosperm. Science **145:** 279–280.

Millerd, A. (1975). Biochemistry of legume seed proteins. Annu. Rev. Plant Physiol. Plant Mol. Biol. **26:** 53–72.

Muehlbauer, G.J., Gengenbach, B.G., Somers, D.A., and Donovan, C.M. (1994). Genetic and amino-acid analysis of two maize threonine-overproducing, lysine-insensitive aspartate kinase mutants. Theor. Appl. Genet. **89:** 767–774.

Nelson, B.K., Cai, X., and Nebenführ, A. (2007). A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. Plant J. **51:** 1126–1136.

Nordborg, M., et al. (2005). The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol. **3:** e196.

Pérez, A.G., Olías, R., Luaces, P., and Sanz, C. (2002). Biosynthesis of strawberry aroma compounds through amino acid metabolism. J. Agric. Food Chem. **50:** 4037–4042.

Platt, A., et al. (2010). The scale of population structure in *Arabidopsis thaliana*. PLoS Genet. **6:** e1000843.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. **38:** 904–909.

Pritchard, J.K., and Donnelly, P. (2001). Case-control studies of association in structured or admixed populations. Theor. Popul. Biol. **60:** 227–237.

Riedelsheimer, C., Lisec, J., Czedik-Eysenberg, A., Sulpice, R., Flis, A., Grieder, C., Altmann, T., Stitt, M., Willmitzer, L., and Melchinger, A.E. (2012). Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proc. Natl. Acad. Sci. USA **109:** 8872–8877.

Rolletschek, H., Borisjuk, L., Koschorreck, M., Wobus, U., and Weber, H. (2002). Legume embryos develop in a hypoxic environment. J. Exp. Bot. **53:** 1099–1107.

Rowan, D.D., Lane, H.P., Allen, J.M., Fielder, S., and Hunt, M.B. (1996). Biosynthesis of 2-methylbutyl, 2-methyl-2-butenyl and 2-methylbutanoate esters in Red Delicious and Granny Smith apples using deuterium-labeled substrates. J. Agric. Food Chem. **44:** 3276–3285.

Sawada, Y., Kuwahara, A., Nagano, M., Narisawa, T., Sakata, A., Saito, K., and Hirai, M.Y. (2009). Omics-based approaches to methionine side chain elongation in Arabidopsis: Characterization of the genes encoding methylthioalkylmalate isomerase and methylthioalkylmalate dehydrogenase. Plant Cell Physiol. **50:** 1181–1190.

Schauer, N., Semel, Y., Balbo, I., Steinfath, M., Repsilber, D., Selbig, J., Pleban, T., Zamir, D., and Fernie, A.R. (2008). Mode of inheritance of primary metabolic traits in tomato. Plant Cell **20:** 509–523.

Schauer, N., et al. (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat. Biotechnol. **24:** 447–454.

Schuster, J., and Binder, S. (2005). The mitochondrial branched-chain aminotransferase (AtBCAT-1) is capable to initiate degradation of leucine, isoleucine and valine in almost all tissues in *Arabidopsis thaliana*. Plant Mol. Biol. **57:** 241–254.

Schuster, J., Knill, T., Reichelt, M., Gershenzon, J., and Binder, S. (2006). Branched-chain aminotransferase4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in Arabidopsis. Plant Cell **18:** 2664–2679.

Schwarz, G. (1978). Estimating the dimension of a model. Ann. Stat. **6:** 461–464.

Segura, V., Vilhjálmsson, B.J., Platt, A., Korte, A., Seren, U., Long, Q., and Nordborg, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat. Genet. **44:** 825–830.

Shewry, P.R., and Halford, N.G. (2002). Cereal seed storage proteins: Structures, properties and role in grain utilization. J. Exp. Bot. **53:** 947–958.

Sreenivasulu, N., Usadel, B., Winter, A., Radchuk, V., Scholz, U., Stein, N., Weschke, W., Strickert, M., Close, T.J., Stitt, M., Graner, A., and Wobus, U. (2008). Barley grain maturation and germination: metabolic pathway and regulatory network commonalities and differences highlighted by new MapMan/PageMan profiling tools. Plant Physiol. **146:** 1738–1758.

Taylor, N.L., Heazlewood, J.L., Day, D.A., and Millar, A.H. (2004). Lipoic acid-dependent oxidative catabolism of alpha-keto acids in mitochondria provides evidence for branched-chain amino acid catabolism in Arabidopsis. Plant Physiol. **134:** 838–848.

Textor, S., de Kraker, J.-W., Hause, B., Gershenzon, J., and Tokuhisa, J.G. (2007). MAM3 catalyzes the formation of all aliphatic

glucosinolate chain lengths in Arabidopsis. Plant Physiol. **144:** 60–71.

Textor, S., Bartram, S., Kroymann, J., Falk, K.L., Hick, A., Pickett, J.A., and Gershenzon, J. (2004). Biosynthesis of methionine-derived glucosinolates in *Arabidopsis thaliana*: Recombinant expression and characterization of methylthioalkylmalate synthase, the condensing enzyme of the chain-elongation cycle. Planta **218:** 1026–1035.

Toubiana, D., Semel, Y., Tohge, T., Beleggia, R., Cattivelli, L., Rosental, L., Nikoloski, Z., Zamir, D., Fernie, A.R., and Fait, A. (2012). Metabolic profiling of a mapping population exposes new insights in the regulation of seed metabolism and seed, fruit, and plant relations. PLoS Genet. **8:** e1002612.

Tressl, R., and Drawert, F. (1973). Biogenesis of banana volatiles. J. Agric. Food Chem. **21:** 560–565.

Utz, H.F., and Melchinger, A.E. (1996). PLABQTL: A program for composite interval mapping of QTL. Journal of Agricultural Genomics **2:** 1–6.

Valerio, A., D'Antona, G., and Nisoli, E. (2011). Branched-chain amino acids, mitochondrial biogenesis, and healthspan: An evolutionary perspective. Aging (Albany, N. Y.) **3:** 464–478.

Walters, D.S., and Steffens, J.C. (1990). Branched chained amino acid metabolism in biosynthesis of *Lycopersicon pennelli* glucose esterase. Plant Physiol. **93:** 1544–1551.

Weber, H., Borisjuk, L., and Wobus, U. (2005). Molecular physiology of legume seed development. Annu. Rev. Plant Biol. **56:** 253–279.

Weir, B.S., and Hill, W.G. (1986). Nonuniform recombination within the human beta-globin gene cluster. Am J Hum Genet. **5:** 776–781.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., and Provart, N.J. (2007). An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. PLoS ONE **2:** e718.

Xu, S.B., Li, T., Deng, Z.Y., Chong, K., Xue, Y., and Wang, T. (2008). Dynamic proteomic analysis reveals a switch between central carbon metabolism and alcoholic fermentation in rice filling grains. Plant Physiol. **148:** 908–925.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., and Buckler, E.S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38:** 203–208.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., and Buckler, E.S. (2010). Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. **42:** 355–360.

Zhu, C., Gore, M., Buckler, E.S., and Yu, J. (2008). Status and prospects of association mapping in plants. Plant Genome **1:** 5–20.

Zhu, X.H., and Galili, G. (2003). Increased lysine synthesis coupled with a knockout of its catabolism synergistically boosts lysine content and also transregulates the metabolism of other amino acids in *Arabidopsis* seeds. Plant Cell **15:** 845–853.

Zolman, B.K., Monroe-Augustus, M., Thompson, B., Hawes, J.W., Krukenberg, K.A., Matsuda, S.P.T., and Bartel, B. (2001). chy1, an Arabidopsis mutant with impaired beta-oxidation, is defective in a peroxisomal beta-hydroxyisobutyryl-CoA hydrolase. J. Biol. Chem. **276:** 31037–31046.

Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., and van Wijk, K.J. (2008). Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. PLoS ONE **3:** e1994.