Natural variation for seed dormancy in Arabidopsis is regulated by additive genetic and molecular pathways

Leónie Bentsink[a,b,1], Johannes Hanson[a,c], Corrie J. Hanhart[b], Hetty Blankestijn-de Vries[b], Colin Coltrane[d], Paul Keizer[c,d], Mohamed El-Lithy[b], Carlos Alonso-Blanco[e], M. Teresa de Andrés[e], Matthieu Reymond[f], Fred van Eeuwijk[c,d], Sjef Smeekens[a,c], and Maarten Koornneef[b,f,1]

[a]Department of Molecular Plant Physiology, Utrecht University, 3584 CH Utrecht, The Netherlands; [b]Laboratory of Genetics, Wageningen University, 6708 PB Wageningen, The Netherlands; [c]Centre for BioSystems Genomics, 6700 AB Wageningen, The Netherlands; [d]Biometris–Applied Statistics, Wageningen University and Research Centre, 6708 PB Wageningen, The Netherlands; [e]Department of Plant Molecular Genetics, Centro Nacional de Biotecnología (CNB) and Consejo Superior de Investigaciones Científicas (CSIC), E-28049 Madrid, Spain; and [f]Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany

Contributed by Maarten Koornneef, January 12, 2010 (sent for review December 2, 2009)

Timing of germination is presumably under strong natural selection as it determines the environmental conditions in which a plant germinates and initiates its postembryonic life cycle. To investigate how seed dormancy is controlled, quantitative trait loci (QTL) analyses has been performed in six *Arabidopsis thaliana* recombinant inbred line populations by analyzing them simultaneously using a mixed model QTL approach. The recombinant inbred line populations were derived from crosses between the reference accession Landsberg *erecta* (L*er*) and accessions from different world regions. In total, 11 *delay of germination* (*DOG*) QTL have been identified, and nine of them have been confirmed by near isogenic lines (NILs). The absence of strong epistatic interactions between the different *DOG* loci suggests that they affect dormancy mainly by distinct genetic pathways. This was confirmed by analyzing the transcriptome of freshly harvested dry seeds of five different *DOG* NILs. All five *DOG* NILs showed discernible and different expression patterns compared with the expression of their genetic background L*er*. The genes identified in the different *DOG* NILs represent largely different gene ontology profiles. It is proposed that natural variation for seed dormancy in Arabidopsis is mainly controlled by different additive genetic and molecular pathways rather than epistatic interactions, indicating the involvement of several independent pathways.

recombinant inbred lines | quantitative trait loci analyses | near isogenic lines | transcriptome analyses

Seed dormancy is an important adaptive trait that together with flowering time is a primary component of the different life history strategies of plants (1). Seasonal timing of germination might well be a stronger factor conditioning the flowering time of Arabidopsis in the field than variation in the genetic basis for flowering time itself (2). Seed dormancy controls the timing of germination by arresting growth and development, despite the presence of favorable environmental conditions to complete germination. Specific environmental and developmental triggers can overcome this arrest. Environmental factors can act during seed development on the mother plant, during seed storage (i.e., after-ripening; AR) and in mature imbibed seeds. The various aspects of seed dormancy and germination have been extensively reviewed recently (3–6). In addition, it has been shown that there is considerable variation for seed dormancy in nature (7–9). The identification of the genes underlying this natural variation for seed dormancy may help to further understand the mechanisms involved in this process. At the same time, it provides insight into the way nature shaped genetic variability for this trait during adaptive evolution. A common approach to discover genes that control quantitative traits is the use of whole-genome scans to identify quantitative trait loci (QTL). These analyses provide estimates of several genetic parameters that underlie phenotypic variation, including the number of loci, the type and magnitude of their effects, interactions between genes (i.e., epistasis), and gene-by-environment interactions when different environments are tested. Many seed dormancy QTL have been identified in crop and model plants (reviewed in ref. 9). In Arabidopsis, natural variation for seed dormancy has been studied in four recombinant inbred line (RIL) populations (10–14). The seed dormancy QTL called *Delay of Germination 1* (*DOG1*) was identified in three of the analyzed populations. *DOG1* is a member of a small gene family of unknown molecular function that had not been previously related to seed dormancy, thus illustrating that natural variation is a valuable resource to identify novel seed dormancy genes. *DOG1* was shown to be specifically expressed during seed development with detectable levels present in dry seeds (15). Global transcript analysis in Arabidopsis using microarrays indicated that the expression level of 30 genes including *DOG1* decreased during AR (16). The expression of *DOG1* was also reduced in the *hub1* mutant characterized by reduced dormancy (17) in agreement with a role of *DOG1* in regulating dormancy levels.

To identify loci that determine natural variation for seed dormancy, we have analyzed the seed dormancy behavior of six RIL populations by analyzing them simultaneously using a mixed-model QTL approach. All populations were derived from crosses between the reference accession Landsberg *erecta* (L*er*) and accessions with different levels of seed dormancy. QTL identified were confirmed by means of near isogenic lines (NILs), which were used to infer molecular pathways from transcriptome analyses. Results presented here indicate that distinct genetic and molecular pathways identified by QTL analyses and transcript profiling are involved in the control of the natural variation for seed dormancy.

Results

Seed Dormancy Behavior of Parental Accessions and Their RIL Populations. We have studied natural variation for seed dormancy defined as the days of seed dry storage required to reach 50% germination (DSDS50) present in six RIL populations. The populations were grown at different times of the year in different

Author contributions: L.B. and M.K. designed research; L.B., C.J.H., H.B.-d.V., M.E.-L., C.A.-B., M.T.d.A., and M.R. performed research; L.B., J.H., C.C., P.K., and F.v.E. analyzed data; and L.B., C.A.-B., F.v.E., S.S., and M.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. E-mail: l.bentsink@uu.nl or koornnee@mpiz-koeln.mpg.de.

This article contains supporting information online at www.pnas.org/cgi/content/full/1000410107/DCSupplemental.

years (Table S1) but in the same greenhouse. These populations represent different levels of seed dormancy, varying from the rather nondormant L*er*/An-1 population to the L*er*/Kas-2 population, which segregates for strong dormancy (Fig. S1). In all populations, DSDS50 showed a high heritability (Table 1), indicating the presence of large genetic variation for this trait. Transgression beyond one or both parental values was observed in all populations. The L*er* parent was grown together with each population and showed DSDS50 values of 10.3 ± 1.1, 16.9 ± 1.4, 23.4 ± 1.4, 32.8 ± 2.4, 9.6 ± 0.5, and 11.0 ± 1.2 for L*er*/An-1, L*er*/Cvi, L*er*/Fei-0, L*er*/Kas-2, L*er*/Kond, and L*er*/Sha, respectively (Table S1). These differences indicate the existence of substantial environmental effects on the mother plants because plants were grown in different experiments in different seasons. Thus, the stronger dormancy in the L*er*/Kas-2 and L*er*/Fei-0 populations might partly result from unknown environmental differences between experiments, as L*er* seeds were most dormant in these two experiments.

Integrated QTL Analyses for Six RIL Populations. QTL analysis was performed using a mixed-model approach analyzing the six RIL populations simultaneously, and the allele substitution effects were estimated per QTL for each of the individual populations (Fig. 1 and Table 1). Eleven QTL were identified with an average total explained variance of 54% (Table 1). Seed dormancy QTL were named according to the nomenclature that has been previously described (11), whereby seven QTL were identified and named *Delay of Germination (DOG)* 1 through 7. Additional QTL that are not overlapping with already reported QTL have been named *DOG18* to *DOG*22 as *DOG*8 to *DOG*15 are already in use (13). As shown in Table 1, two to eight QTL had significant allele substitution effects in the individual populations. Small differences were observed when comparing the integrated analyses on all the populations versus earlier published data (11, 14). For the L*er*/Cvi population, for example, we identified significant effects for two QTL that were not identified before, *DOG19* and *DOG21*, as a result of increased power of detection by the combined analysis. We did not detect the unique QTL at marker msat2-5 that was identified in the L*er*/Sha population (14). Of the newly identified QTL, *DOG22* is population-specific and only showed a significant effect in the L*er*/An-1 population, whereas *DOG18-21* were identified to be significant in more populations. In the L*er*/Kond population we detected significant effects only at QTL in the *DOG1* and *DOG6* regions that were previously identified in the L*er*/Cvi and the L*er*/Sha populations (11, 14). These results indicate that variation for seed dormancy in different accessions is determined by allelic variation at different loci. However, part of the differences among populations might be a result of genotype-by-environment interactions, because the populations were grown in different experiments.

Validation of the *DOG* Loci. Introgression lines carrying single genomic fragments of the different accessions around eight of the *DOG* QTL regions (*DOG1*, *DOG3*, *DOG5*, *DOG6*, *DOG18*, *DOG19*, *DOG20*, *DOG22*) into an otherwise L*er* genetic background were developed using marker-assisted selection (Table S2). The dormancy behavior of 18 of these introgression lines and four other lines previously developed carrying *DOG*-Cvi alleles (11) was determined and DSDS50 values were calculated. Eighteen NILs significantly differed from L*er* ($P < 0.05$; Fig. 2), which confirmed the presence of QTL *DOG1*, *DOG2, DOG3, DOG6, DOG18*, *DOG19*, and *DOG22.* For *DOG1* we could study seven different alleles (including the L*er* allele). Five of these alleles increase the level of seed dormancy in comparison with the L*er* allele. The Fei-0 *DOG1* allele has an opposite allelic effect, indicating that this is even weaker than the L*er* allele, which is not a null allele (15). Five *DOG6* alleles were introgressed into L*er*, and in all cases the level of seed dormancy was increased in comparison with L*er*. The effects of QTL in these NIL are all in agreement with the results of the QTL analyses.

Table 1. QTL for DSDS50 in six RIL populations, as obtained after an integrated analysis comprising SIM, CIM, and backward selection

DOG No.	Chrom	Position (cM)	−10Log(p)	Support interval	QTL effects (square root scale)					
					Ler/An-1	Ler/Cvi	Ler/Fei-0	Ler/Kas-2	Ler/Kond	Ler/Sha
Mean effect					3.29	7.44	6.54	9.14	5.4	5.07
DOG2	1	13.5	4.7	6.9–19.9	0.03	**0.62**	−0.10	−0.20	0.07	0.00
DOG3	1	54.3	5.7	49.4–63.1	**−0.43**	**−0.35**	−0.08	−0.11	−0.07	**−0.39**
DOG19	1	146.8	9.9	142.0–150.7	−0.21	**−0.41**	**−0.58**	**−0.35**	−0.01	0.11
DOG20	2	49.1	8.5	45.9–64.5	**−0.56**	0.12	**−0.38**	0.10	−0.12	−0.15
DOG22	3	3.9	2.7	3.9–3.9	**0.32**	0.04	0.23	0.01	0.02	0.15
DOG21	3	24.5	4.4	16.0–40.8	0.18	**0.25**	**−0.41**	0.00	0.17	0.21
DOG6	3	72.9	45.9	67.9–75.6	0.21	**−0.39**	**−0.98**	**−1.03**	**−0.77**	**−1.43**
DOG18	4	28.1	16.9	26.1–28.1	**0.75**	−0.16	**0.62**	**0.28**	−0.16	−0.06
DOG5	4	67	3.3	63.6–76.1	**−0.25**	**−0.39**	−0.05	−0.21	−0.02	−0.10
DOG4	5	7.5	3.5	3.9–32.5	−0.06	**−0.37**	**−0.26**	−0.01	0.07	−0.11
DOG1	5	91.3	53.8	91.3–91.3	0.18	**−1.31**	**0.33**	**−2.34**	**−0.48**	**−0.51**
Fraction of explained variance by main effect QTL					0.61	0.55	0.56	0.66	0.42	0.49
Fraction of explained variance by main effect QTL and epistatic interactions					0.69	0.62	0.56	0.66	0.42	0.51
Heritability					0.94	0.98	0.92	0.97	0.92	0.87
Average fraction of explained variance by main effect QTL					0.54	–	–	–	–	–
Average fraction of explained variance by main effect QTL plus epistatic interactions					0.57	–	–	–	–	–

Genome-wide threshold in SIM and CIM was 3.2, on the −log10(p) scale, with p the *P* value [*DOG22* was added to the model at a −log10(p) of 2.7]. QTL name, chromosome, position, significance expressed on a −10log(p) scale, and a ± 1.5 dropoff interval on the −10log(p) scale are presented. *P* values were taken from the final multi-QTL model after backward selection. Dropoff intervals were assessed on the CIM profile. QTL effects (DSDS50, square root scale) are given in the right part of the table. The first line contains the mean DSDS50 for each of the six populations. The allele substitution effects of the individual QTL are given below that. A negative value indicates that L*er* is decreasing the DSDS50, whereas a positive value indicates that L*er* increases the DSDS50 compared to the alleles of the other accessions. Significant effects are indicated in bold; these are the effects of the QTL indicated in Fig. 1. In the bottom part of the table, for each population is given the fraction of explained variance by main effect QTL, the fraction of explained variance by main effect QTL and epistatic interactions, and the heritability. Finally, the fraction of variance explained across the six populations is given for the model with main effect QTL only, and for main effect QTL plus epistatic interactions.

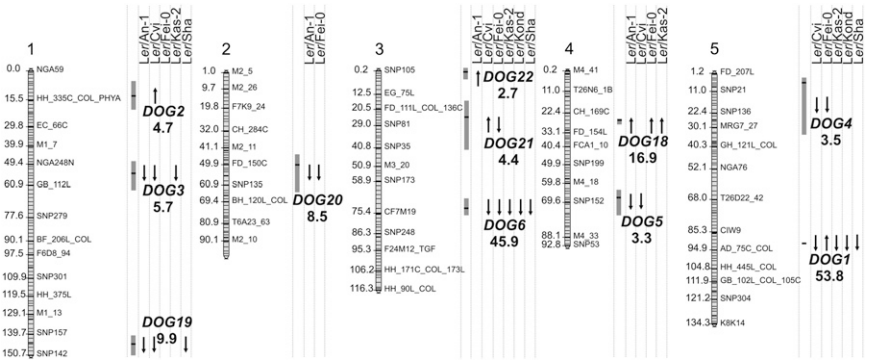

Fig. 1. Seed dormancy QTL identified in six RIL populations. Combined physical map of L*er*/An-1, L*er*/Cvi, L*er*/Fei-0, L*er*/Kond, L*er*/Kas-2, and the L*er*/Sha RIL populations. QTL are depicted as gray arrows along the chromosomes. Markers that were fixed as cofactors are indicated by the black horizontal bar in the gray arrows. The lengths of the arrows indicate the 1.5 [−log10(p)] unit dropoff intervals. −Log10(*P*) values that indicate the significance of the QTL are indicated below the DOG number. Allelic effects of the QTL in the different populations are indicated by the black arrows. Arrowheads pointing up represent L*er* increasing the DSDS50 value; arrowheads pointing down represent the non-L*er* allele increasing the DSDS50.

Epistatic Analyses. Epistasis analyses were performed by analyzing pair-wise QTL interactions. Two significant population-specific epistatic interactions have been identified for which the presence of L*er* alleles at both loci would decrease the DSDS50: *DOG3* × *DOG1*, in the L*er*/Cvi population, in agreement with Alonso-Blanco et al. (11); and *DOG18* × *DOG1*, in the L*er*/An-1 population. The latter probably prevented the detection of a significant effect for *DOG1* in this population. Over the six populations, the effect of epistatic interactions was negligible compared with the additive main effects of the QTL (the average total explained variance including epistatic interactions was 57% vs. 54% explained by main effect additive QTL; Table 1). It is possible that two-way interactions among other loci and/or higher order interactions contribute to the dormancy variation given the difference between the QTL explained variance and heritability estimates. However, unexplained or missing heritability is a general problem (18). These data indicate that natural variation for seed dormancy in the analyzed accessions is regulated by distinct genetic pathways, which probably show some downstream convergence.

Gene Expression Profiles of Dry Dormant Seeds from *DOG* NILs. Regulation of seed dormancy by different genetic pathways should lead to distinct transcriptome profiles in the different *DOG* NILs. To test this hypothesis, expression profiling was performed in dry dormant seeds of five of the *DOG* NILs (*DOG1*-Cvi, *DOG2*-Cvi, *DOG3*-Cvi, *DOG6*-Kas-2, and *DOG22*-An-1). Expression levels of the different NILs were compared with that of L*er* representing the genetic background of the lines. A total of 640 genes have been identified as differently expressed in comparison with L*er* ($P < 0.05$), of which 342 are up-regulated and 298 are down-regulated in the different NILs (Table 2). Differentially expressed genes were identified to be located both within (i.e., local) and outside (i.e., distant to) the introgressed regions (Table 2). Most genes identified as being differentially expressed in the individual *DOG* NILs were not identified as differentially expressed in the other *DOG* NILs (Fig. 3). This was the case both for local and distantly regulated genes. Each *DOG* NIL therefore represents a unique pattern of gene expression differences (Dataset S1 and Table S3). Genes identified as being differentially expressed in the different *DOG*s also represented different Gene Ontology (GO) classes. To investigate if the GO profiles of the different *DOG* NILs were similar, we performed an equivalence test based on the squared Euclidean distance between the GO profiles (19) of lists of genes differentially regulated in the different *DOG* NILs (Fig. S2). Only the comparison *DOG1*-*DOG6* for the cellular compartment indicated that these GO profiles were significantly similar ($P < 0.0001$), but all other comparisons were significantly different. Therefore, we concluded that different processes are affected in the different *DOG* NILs. Major biological processes down-regulated in dormant seeds of *DOG* NILs are catabolic processes, hormone metabolic processes, maintenance of cellular localization and translation for NIL*DOG1*, NIL*DOG2*, NIL*DOG3*, and NIL*DOG6,* respectively. Expression differences between NIL*DOG22* and L*er* were very small and did not lead to specific GO classes being significantly down-regulated. Biological processes that were up-regulated in dormant *DOG* NILs seeds were indole

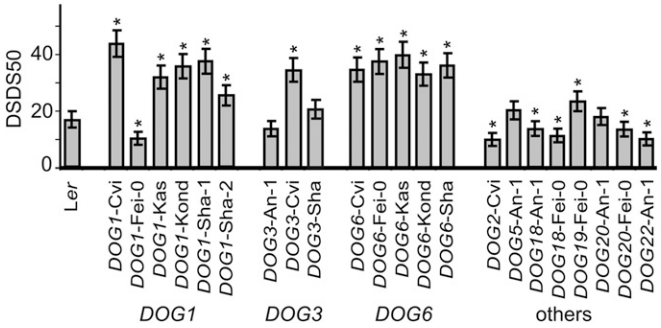

Fig. 2. Average DSDS50 values and CIs representing dormancy behavior of L*er* and *DOG* NILs (analysis was done on square root scale, back-transformed values are given). Asterisks indicate if the NILs differ significantly from L*er* at $P < 0.05$ (on square root scale).

Table 2. Numbers of genes differentially expressed in freshly harvested dry seeds of the different *DOG* NILs compared with L*er*

Gene	Up-regulated			Down-regulated		
	Total	Local	AR (%)	Total	Local	AR (%)
DOG1	85	33	3/8 (38)	40	26	2/12 (17)
DOG2	64	28	3/15 (20)	33	20	3/14 (21)
DOG3	25	22	1/14 (7)	35	30	12/15 (80)
DOG6	162	42	2/8 (25)	187	36	5/12 (42)
DOG22	6	2	NA	3	0	NA
Total	342	–	–	298	–	–

"Total" reflects differentially expressed genes; "Local" locally expressed genes (differentially expressed genes located within introgressed region); and AR the fraction of genes affected by AR (tested with QRT-PCR). NA, not analyzed.

metabolic processes, response to stress, viral genome replication, cell redox homeostasis and photosynthesis for NIL*DOG1*, NIL*DOG2*, NIL*DOG3,* NIL*DOG6*, and NIL*DOG22,* respectively.

Not all genes identified as differentially expressed may reflect the dormancy status of the seeds. Both local and distant gene expression differences can be caused by genes located in the introgressed region other than the gene responsible for the dormancy effect. Genes that are most likely involved in the control of seed dormancy are expected to differ in expression when dormant and nondormant seed batches are compared. Therefore the expression of 15 genes with the highest significances for up- and down-regulation of expression was analyzed by means of quantitative RT-PCR in dry dormant and dry AR seeds of *DOG* NILs in which they were detected (Fig. S3). In many cases a substantial proportion of the tested genes tested (7–80%) were differentially expressed when comparing dormant with AR seeds (Table 2).

The observation that *DOG1* (At5g45830) is up-regulated in freshly harvested dry (i.e., dormant) seeds of the NIL*DOG1* compared with L*er* indicates that microarray data can be used to identify genes that encode for the QTL when regulated at the transcription level. Other genes that are differentially expressed in our data that have been previously related to seed dormancy are the *ABI3* gene (up-regulated in NIL*DOG6*) involved in the network that controls various aspects of seed maturation (20); *NCED4* (up-regulated in NIL*DOG1*), one of the genes encoding the abscisic acid (ABA) biosynthetic enzyme NCED; *SOMNUS* (down-regulated NIL*DOG2*) encoding a CCCH-type zinc finger protein in Arabidopsis that negatively regulates light-dependent seed germination downstream of PIL5 (21); and *CYP707A2* (down-regulated in NIL*DOG6*), a gene that plays a major role in ABA degradation early during seed imbibition (22). We also identify *FLOWERING LOCUS C* (*FLC*) (dormancy down-regulated, which means up-regulated in NIL*DOG2* and down-regulated in NIL*DOG3*), which is a major regulator of flowering time that has been recently found to control germination behavior in a temperature-dependent way (23). In addition, we detected overlap between our set of differentially expressed genes with published global transcript analysis performed on Cvi seeds tested in different germination environments, which reflect different dormancy stages (16). Genes identified as related to dormancy release in Cvi seeds in our data are divided over the different *DOG* NILs. This was expected, as three of our NILs contain introgression fragments of Cvi, representing three of the seven dormancy loci previously identified in Cvi (11).

The unique gene expression pattern of the *DOG* NILs and that of the GO profiles, together with the lack of strong epistatic interactions, strongly indicate that natural variation for seed dormancy in Arabidopsis is regulated by distinct genetic and molecular pathways.

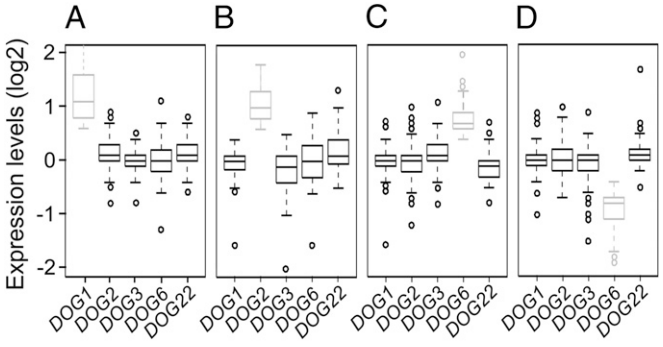

Fig. 3. Changes in expression of differentially expressed genes do not correlate between *DOG* NILs. (*A*) Dormancy up-regulated (Dup) genes in *DOG1* (n = 85), (*B*) Dup-regulated genes in *DOG2* (n = 64), (*C*) Dup-regulated genes in *DOG6* (n = 162), and (*D*) dormancy down-regulated genes in *DOG6* (n = 188). Expression distribution of genes determined as differentially expressed in one genotype. The log2 expression level difference of each NIL*DOG* compared with L*er* is depicted. The reference is indicated in light gray; the circles indicate the outliers. All expression levels differences are depicted except some outliers used for the selection of genes, one gene repressed more than fourfold in NIL*DOG6* (*B*), and one gene induced more that fourfold in NIL*DOG2* (*D*).

Discussion

To understand the genetic basis of the substantial natural variation that exists for seed dormancy in Arabidopsis (7, 8, 10, 11), we have performed an integrated QTL analyses for AR requirement in six RIL populations including two populations that were analyzed before (11, 14). This procedure allowed increasing the effective population size for QTL that segregate in several populations. This may lead to discovery of new QTL, as shown for *DOG19* and *21* in the L*er*/Cvi population, and/or to a more precise map position as demonstrated (e.g., for *DOG1* and *DOG6* that segregate in several populations). In total we identified 11 QTL, of which nine could be confirmed by new and previously developed NILs (11), as their dormancy behavior (DSDS50) was significantly different from L*er* ($P < 0.05$). We consider the colocation of QTL detected in different populations as allele effects from the same locus, although we can not rule out that different but linked genes are responsible in the different accessions. Epistatic relationships have been identified for only two combinations of loci, but their effects are relatively small. The absence of strong epistatic interactions between the different *DOG* loci and the transcriptomics analyses performed on freshly harvested dry seeds of the *DOG* NILs suggests that dormancy is affected through distinct genetic and molecular pathways. The genes identified in the transcriptomic experiments were generally specific for the individual *DOG* NILs and the GO analyses indicated that different processes are affected, including catabolic processes, hormone synthesis, translation, and photosynthesis. The observation that dormancy in Arabidopsis is primarily determined by additive effects is in agreement with data from barley (24). For weedy rice, epistatic interactions have been studied extensively. Although most QTL consisted of predominantly gene additive effects, significant interactions between dormancy loci were detected (25), which is probably a result of the involvement of other mechanisms in rice. For flowering time, another important life history trait, genetic variation in nature, is controlled by two major genes (*FRIGIDA* and *FLC)* of a pathway with strong epistatic interactions (26).

Presumably, the different pathways regulating seed dormancy are the result of adaptation. Although our data are based on laboratory experiments, recently Huang et al. (27) have shown that field and laboratory experiments with a different RIL population revealed the same dormancy QTL. These QTL, which were at the position of *DOG1* and *DOG6*, are the major loci responsible for dormancy in the more dormant accessions and segregate in three accessions originally collected in the same broad geographical region (Kond and Sha in Tajikistan and Kas-2 in India and Kashmir). For the two Tajikistan accessions, *DOG6* is the major QTL, whereas in Kas-2 *DOG1* shows the strongest effect. The other accessions analyzed in this work, Cvi, Fei-0, and An-1, revealed additional loci, and loci that were accession-specific. This indicates that seed dormancy analyses in accessions from different environments might lead to the identification of additional loci, indicating the complexity of pathways that control seed dormancy.

The QTL identified in this study might also play a role in the control of other germination related traits. In agreement with this hypothesis, Laserna et al. (13) recently reported on natural genetic variation for the light responses of seeds of L*er*/Cvi and Bay-0/Sha RIL populations and their interactions with AR and incubation temperature. Twelve loci were identified under red light, far-red light, or in darkness in both RIL populations, which

overlap in their locations with *DOG1*, *DOG2*, *DOG3*, and *DOG7*. Of the newly identified QTL, *DOG21* overlaps with *DOG8* and *DOG20* with *DOG11*. Furthermore, three major loci have been identified as responsible for the variability in cold-tolerant and dark germination in the Bay-0/Sha RIL population (12). One of these QTL, *CDG-1* (Cold-tolerant Dark Germination), was localized on the same genomic region as *DOG2* and *DOG3* on chromosome 1. *DOG20* might colocate with *CDG-2*. Recently, Kover et al. (28) identified two QTLs for days to germination using a multiparent advanced generation intercross population, one of them colocating with *DOG6*.

Several but not all *DOG* genomic regions identified in this work contain genes that previously have been associated with seed dormancy. For instance, *DOG3* colocates with *LEC2*, *DOG4* with *TT7*, *DOG5* with *ABI1*, *DOG19* with *GA2*, *DOG20* with *PIL5*, and *DOG22* with *RGL2*. However fine mapping excluded *LEC2* and *RGL2* as candidate genes (see Fig. 5.1 in ref. 9 for an overview). Several arguments indicate that genes identified in the study of natural variation are different from genes identified in mutant screens. The parental lines used for mutation experiments often are the low dormant standard laboratory accessions L*er* and Col. These might contain mutations in specific genes, e.g., those promoting seed dormancy. Furthermore, mutants showing strong pleiotropic effects, such as most ABA mutants, likely fail to survive in nature.

This integrative analysis of natural variation for seed dormancy reported here indicates that different genetic and molecular pathways control seed dormancy. Further investigations with the genotypes generated in this work will elucidate the molecular mechanisms underlying the different pathways.

Materials and Methods

Plant Material. The RIL populations L*er*/An-1, L*er*/Cvi, L*er*/Kond, L*er*/Kas-2, L*er*/Sha, and L*er*/Fei-0 were described before (14, 29–31). NILs with the Cvi introgression fragments (NIL*DOG*-Cvi) are described by Alonso-Blanco et al. (11).

Construction of dormancy NILs. Eighteen NILs were constructed by the introgression of the identified DSDS50 QTL regions (*DOG1*, *DOG3*, *DOG5*, *DOG6*, *DOG18*, *DOG19*, *DOG20*, and *DOG22*) of the different accessions into a L*er* genetic background. Specific RILs (Table S2) were backcrossed to L*er* and the BCF2 generation lines were identified that contained single introgressions of the donor accession by applying marker-assisted selection. The lines were called NILDOGx-accession, in which the x represents the number of the *DOG* QTL and the donor accession of the introgressed fragment.

Growth Conditions. The growth conditions used were described by El-Lithy et al. (30).

RIL analyses. RILs L*er*/An-1, L*er*/Fei-0, L*er*/Kas-2, and L*er*/Kond and their parental lines were grown in a randomized complete block design with two replicates. An experimental plot consisted of a row of six plants for the RILs and of 12 plants for the parental lines. To reduce developmental and environmental effects on seed dormancy, the onset of flowering was synchronized for the L*er*/Kond RIL population as this population shows large variation for flowering initiation (30). For that, RILs were planted at two consecutive weeks according to their flowering times. For the other three populations, all RILs were planted at the same time (per population) although the seeds of the L*er*/Kas-2 populations were harvested at two consecutive weeks, because not all seeds were synchronously ripened at the same time (Table S1). Seeds were harvested and a seed bulk of two plants was used for seed dormancy measurements.

NIL analyses. NILs were grown in a randomized complete block design with four and eight replicates, respectively, for the dormancy assays and the transcriptome analyses. An experimental plot consisted of a row of 12 plants. For the dormancy assays seeds of three plants per replicate were bulked, and for the transcriptome analyses seeds of eight plants per line were bulked.

Seed Dormancy Measurements and Germination Assays. Germination tests were performed as described by Alonso-Blanco et al. (11). In each experiment, germination was tested for the various genotypes in at least six different time points of dry storage from the harvest date until 100% of the seeds germinated in most genotypes. Specific details, i.e., number of data points and how many days of seed dry storage the germination experiments have been performed, are presented in Table S1.

Nonlinear Regression Analysis. Logistic curves were fitted to the germination data, for each RIL in each replicate, by GenStat version 9.0 (32), using the following function:

$$Y = A + \frac{C}{1 + e^{-B(X-M)}} + \text{error}, \qquad [1]$$

where Y is the response variable, percent germination, and X is the explanatory variable, days of seed dry storage. The lower asymptote and upper asymptote, A and C, were constrained to zero and 100%, respectively, and the slope, B, and inflection point, M, were estimated. M gives DSDS50. The average DSDS50 across replicates was used in QTL mapping. Variation in slope was found too small to be of interest (data not provided) and no results will be reported on it.

Analyses of Heritabilities. Heritabilities for individual trials were calculated by taking the ratio of between RIL variance and phenotypic variance, where the latter was the sum of the between RIL variance and the intra block variance. The variance components were estimated by the residual maximum likelihood implementation in Genstat (32).

Joint QTL Analyses of the Six RIL Populations. ***Genetic map.*** Linkage maps published before (14, 29–31) were recalculated using JoinMap version 4 (33). Additional markers were added to the genetic linkage maps of some populations to enable comparison among populations (Tables S4 and S5). All markers were located on a physical map, based on their position in the Columbia accession, which is referred to as the combined physical map. The order of the markers on this combined physical map was used as a fixed order within JoinMap to correct discrepancies between the marker arrangement of linkage maps of different populations and the physical map. Recalculated linkage maps were checked against the combined physical map to ensure correct marker order. Genetic distances on the linkage map were derived from the distances on the physical map by equating 200 kb to 1 cM.

QTL mapping. QTL were identified by fitting various mixed models to the estimated mean DSDS50 from the preliminary statistical analyses. The procedure that was followed was based on a mixed-model approach as described recently (34, 35). A set of so-called genetic predictors is created from marker information at a grid of evaluation points along the genome. Test for QTL follow from including genetic predictors in a model for the phenotypic response (i.e., DSDS50) as explanatory variables for which coefficients need to be estimated. Whenever the test statistic for the regression exceeds a critical threshold, the position corresponding to the particular genetic predictor provided evidence for the presence of a QTL.

In its simplest form, for additive QTL allele effects, at marker positions, genetic predictors simply count the number of alleles coming from the first parent (vs. those of the second parent), i.e., the genetic predictor takes the value 2, 1, or 0, when the marker genotype is AA, Aa, and aa, respectively, and we define the A-allele as the allele coming from the first parent. For positions between marker loci, specific algorithms exist to calculate genetic predictors (36), where the values of the genetic predictors become the conditional probabilities for a QTL genotype, QQ, Qq, and qq, given the genotypes at the flanking marker loci.

After calculation of genetic predictors, simple interval mapping (SIM) and composite interval mapping (CIM) strategies were used for QTL mapping. Tests for QTL were performed on a grid of genomic locations. The grid for evaluation consisted of 293 marker positions, complemented by 118 evaluations between marker positions, where the maximum distance between consecutive evaluations was 2.5 cM. The mixed models that were fitted to identify QTL contained a fixed effect for the cross, and random effects for QTL allele effects and the error term. The QTL allele effects were assumed to follow the same normal distribution across populations. The variance for the error was dependent on the cross. For SIM, we compared the following two models for each point on the evaluation grid: (*i*) response = cross + error; (*ii*) response = cross + genetic predictor + error. The difference in deviance (37) between the models provided a test for a possible QTL at that position. The test level for the deviance tests was corrected for multiple testing according to the procedure proposed by Li and Ji (38). Our threshold for declaring a QTL significant was 6.8×10^{-4}, or 3.2 on the 10log scale. (Except for *DOG22*, where we applied a slightly more liberal criterion of 2.7 on the -log10(p) scale.) A SIM scan was followed by a CIM scan in which all QTL identified in the previous SIM scan were included as cofactors, except for those that appeared in a window of 10 cM on either side of the evaluation point. The models that were compared by a deviance test in the CIM scan were 1)

response = cross + cofactors + error; 2) response = cross + cofactors + genetic predictor + error. Here, both the cofactors and the genetic predictor at the evaluation point were design matrices. Finally, after the CIM scan, a backward selection procedure was performed, at a test level of 0.05, to find out whether particular QTL could be dropped from the model resulting from the CIM scan. For the retained QTL, a rough estimate for a support interval for their position was obtained by using the test profiles of the CIM scan that were created by plotting the −10log value of the *P* value. Heuristic lower and upper bounds for QTL position were calculated as a 1.5-unit dropoff interval as described by Keurentjes et al. (39).

Epistatic analysis. Pair-wise additive by additive epistatic interactions were investigated for all QTL that were found to possess significant additive QTL main effects. Epistatic interactions were fitted as random effects with a proper variance. The test for a specific epistatic interaction was performed by comparing the deviance of a model with all main effect QTL and a particular epistatic interaction with the deviance of a model with just the main effect QTL and no epistatic interaction. The test level was chosen equal to that of the earlier genome wide scans for main effect QTL. All QTL analyses were done on square root–transformed DSDS50.

DNA Microarray Analysis, Quantitative RT-PCR, and Gene Ontology Analyses. For information on DNA microarray analysis, quantitative RT-PCR and gene ontology analyses, see *SI Materials and Methods*. RT-PCR primers are given in Table S6.

ACKNOWLEDGMENTS. The authors thank Joost Keurentjes for sharing marker information for the L*er*/Cvi population, Martin de Boer for assistance with the QTL analyses and Service XS for performing the microarray hybridizations. L.B. was supported by grants from The Earth and Life Sciences Foundation subsidized by the Netherlands Organization for Scientific Research. F.v.E. and P.K. were supported by the Generation Challenge Program (Molecular Breeding Platform activity 2.2.1) and the Centre of BioSystems Genomics (Projects BB9 and BB12).

1. Donohue K, et al. (2005) Environmental and genetic influences on the germination of *Arabidopsis thaliana* in the field. *Evolution* 59:740–757.
2. Wilczek AM, et al. (2009) Effects of genetic perturbation on seasonal life history plasticity. *Science* 323:930–934.
3. Kucera B, Cohn MA, Leubner-Metzger G (2006) Plant hormone interactions during seed dormancy release and germination. *Seed Sci Res* 15:281–307.
4. Finch-Savage WE, Leubner-Metzger G (2006) Seed dormancy and the control of germination. *New Phytol* 171:501–523.
5. Holdsworth MJ, Bentsink L, Soppe WJJ (2008) Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. *New Phytol* 179:33–54.
6. Bentsink L, Koornneef M (2008) Seed dormancy and germination. *The Arabidopsis Book* (American Society of Plant Biologists, Rockville, MD). Available at www.aspb.org/publications/arabidopsis/.
7. Lawrence MJ, Vaughan JG, Macleod AJ, Jones BMG (1976) Variations in natural populations of *Arabidopsis thaliana* (L.) Heynh. *The biology and chemistry of the cruciferae* (Academic Press, London), pp 167–190.
8. Ratcliffe D (1976) Germination characteristics and their inter- and intra-population variability in Arabidopsis. *Arab Inf Serv* 13:34–45.
9. Bentsink L, Soppe WJJ, Koornneef M (2007) Genetic aspects of seed dormancy. *Seed Development, dormancy and Germination*, eds Bradford KJ, Nonogaki H (Blackwell Publishing, Oxford), pp 113–127.
10. Van der Schaar W, et al. (1997) QTL analysis of seed dormancy in Arabidopsis using recombinant inbred lines and MQM mapping. *Heredity* 79:190–200.
11. Alonso-Blanco C, Bentsink L, Hanhart CJ, Blankenstijn-de Vries H, Koornneef M (2003) Analysis of natural allelic variation at seed dormancy loci of *Arabidopsis thaliana*. *Genetics* 164:711–729.
12. Meng PH, Macquet A, Loudet O, Marion-Poll A, North H (2008) Analysis of natural allelic variation controlling *Arabidopsis thaliana* seed germinability in response to cold and dark: Identification of three major quantitative trait loci. *Mol Plant* 1:145–154.
13. Laserna MP, Sanchez RA, Botto JF (2008) Light-related loci controlling seed germination in L*er* x Cvi and Bay-0 x Sha recombinant inbred-line populations of *Arabidopsis thaliana*. *Ann Bot (Lond)* 102:631–642.
14. Clerkx EJM, et al. (2004) Analysis of natural allelic variation of Arabidopsis seed germination and seed longevity traits between the accessions Landsberg *erecta* and Shakdara, using a new recombinant inbred line population. *Plant Physiol* 135:432–443.
15. Bentsink L, Jowett J, Hanhart CJ, Koornneef M (2006) Cloning of *DOG1*, a quantitative trait locus controlling seed dormancy in Arabidopsis. *Proc Natl Acad Sci USA* 103:17042–17047.
16. Finch-Savage WE, Cadman CS, Toorop PE, Lynn JR, Hilhorst HW (2007) Seed dormancy release in Arabidopsis Cvi by dry after-ripening, low temperature, nitrate and light shows common quantitative patterns of gene expression directed by environmentally specific sensing. *Plant J* 51:60–78.
17. Liu Y, Koornneef M, Soppe WJ (2007) The absence of histone H2B monoubiquitination in the Arabidopsis *hub1* (*rdo4*) mutant reveals a role for chromatin remodeling in seed dormancy. *Plant Cell* 19:433–444.
18. Manolio TA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753.
19. Sanchez A, Salicru M, Ocana J (2007) Statistical methods for the analysis of high-throughput data based on functional profiles derived from the Gene Ontology. *J Stat Plann Infer* 137:3975–3989.
20. To A, et al. (2006) A network of local and redundant gene regulation governs Arabidopsis seed maturation. *Plant Cell* 18:1642–1651.
21. Kim DH, et al. (2008) *SOMNUS*, a CCCH-type zinc finger protein in Arabidopsis, negatively regulates light-dependent seed germination downstream of *PIL5*. *Plant Cell* 20:1260–1277.
22. Kushiro T, et al. (2004) The Arabidopsis cytochrome P450 CYP707A encodes ABA 8′-hydroxylases: key enzymes in ABA catabolism. *EMBO J* 23:1647–1656.
23. Chiang GC, Barua D, Kramer EM, Amasino RM, Donohue K (2009) Major flowering time gene, *FLOWERING LOCUS C*, regulates seed germination in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 106:11661–11666.
24. Takeda K, Hori K (2007) Geographical differentiation and diallel analysis of seed dormancy in barley. *Euphytica* 153:249–256.
25. Gu XY, Kianian SF, Foley ME (2004) Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (*Oryza sativa*). *Genetics* 166:1503–1516.
26. Sung S, Amasino RM (2005) Remembering winter: toward a molecular understanding of vernalization. *Annu Rev Plant Biol* 56:491–508.
27. Huang X, et al. The earliest stages of adaptation in an experimental plant population: strong selection of QTLs for seed dormancy. *Mol Ecol*, in press.
28. Kover PX, et al. (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet* 5:e1000551.
29. Alonso-Blanco C, et al. (1998) Development of an AFLP based linkage map of L*er*, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a L*er*/Cvi recombinant inbred line population. *Plant J* 14:259–271.
30. El-Lithy ME, et al. (2006) New Arabidopsis recombinant inbred line populations genotyped using SNPWave and their use for mapping flowering-time quantitative trait loci. *Genetics* 172:1867–1876.
31. Méndez-Vigo B, de Andrés MT, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C (2010) Temporal analysis of natural variation for the rate of leaf production and its relationship with flowering initiation in *Arabidopsis thaliana*. *J Exp Bot*, in press.
32. Payne RW, et al. (2006) *GenStat r Release 9 Reference Manual* (VSN International, Hertfordshire, UK).
33. Van Ooijen JW, Voorrips RE (2001) *Joinmap(R) version 3.0: Software for the calculation of genetic linkage maps.* (Plant Research International, Wageningen, The Netherlands). Available at www.kyazma.nl/index.php/mc.JoinMap/.
34. Boer MP, et al. (2007) A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. *Genetics* 177:1801–1813.
35. Paulo MP, Boer M, Huang X, Koornneef M, van Eeuwijk F (2009) A mixed model QTL analysis for a complex cross population consisting of a half diallel of two-way hybrids in *Arabidopsis thaliana*: analysis of simulated data. *Euphytica* 161:107–114.
36. Jiang C, Zeng ZB (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* 140:1111–1127.
37. Verbeke G, Molenberghs G (2000) *Linear mixed models for longitudinal data* (Springer-Verlag, New York).
38. Li J, Ji L (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 95:221–227.
39. Keurentjes JJB, et al. (2006) The genetics of plant metabolism. *Nat Genet* 38:842–849.