Arabidopsis semidwarfs evolved from independent mutations in *GA20ox1*, ortholog to green revolution dwarf alleles in rice and barley

Luis Barboza[a,b], Sigi Effgen[a], Carlos Alonso-Blanco[c], Rik Kooke[d], Joost J. B. Keurentjes[e], Maarten Koornneef[a,e,1], and Rubén Alcázar[a,f]

[a]Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, D-50829 Cologne, Germany; [b]Centro para Investigaciones en Granos y Semillas, Universidad de Costa Rica, 2060 San José, Costa Rica; [c]Departamento de Genética Molecular de Plantas, Centro Nacional de Biotecnología, 28049 Madrid, Spain; Laboratories of [d]Plant Physiology and [e]Genetics, Wageningen University, 6708 PB Wageningen, The Netherlands; and [f]Unitat de Fisiologia Vegetal, Facultat de Farmàcia, Universitat de Barcelona, 08028 Barcelona, Spain

Contributed by Maarten Koornneef, August 13, 2013 (sent for review June 20, 2013)

Understanding the genetic bases of natural variation for developmental and stress-related traits is a major goal of current plant biology. Variation in plant hormone levels and signaling might underlie such phenotypic variation occurring even within the same species. Here we report the genetic and molecular basis of semidwarf individuals found in natural *Arabidopsis thaliana* populations. Allelism tests demonstrate that independent loss-of-function mutations at *GA locus 5* (*GA5*), which encodes gibberellin 20-oxidase 1 (*GA20ox1*) involved in the last steps of gibberellin biosynthesis, are found in different populations from southern, western, and northern Europe; central Asia; and Japan. Sequencing of *GA5* identified 21 different loss-of-function alleles causing semidwarfness without any obvious general tradeoff affecting plant performance traits. *GA5* shows signatures of purifying selection, whereas *GA5* loss-of-function alleles can also exhibit patterns of positive selection in specific populations as shown by Fay and Wu's *H* statistics. These results suggest that antagonistic pleiotropy might underlie the occurrence of *GA5* loss-of-function mutations in nature. Furthermore, because *GA5* is the ortholog of rice *SD1* and barley *Sdw1/Denso* green revolution genes, this study illustrates the occurrence of conserved adaptive evolution between wild *A.thaliana* and domesticated plants.

Arabidopsis natural variation | dwarf accessions | gibberellin mutants

Bioactive gibberellins (GAs) are plant growth regulators involved in important traits such as seed germination, flowering time, flower development, and elongation growth (1). GA biosynthesis and signaling pathways are well defined (1, 2) and have been targeted in crop breeding. Modification of GA pathways was crucial in the green revolution because it conferred semidwarfness, thus reducing lodging and increasing crop yields (3–6). Green revolution semidwarf varieties in wheat are due to mutations in *DELLA* genes, whereas many short straw rice varieties carry a mutation in the *Semi-Dwarf-1* (*SD1*) locus. This locus codes for *GA 20-oxidase-2*, a GA biosynthesis gene that is also mutated in most modern barley varieties in which the gene was called *Denso* or *Semi-dwarf 1* (*Sdw1*) (7).

GA 20-oxidases are involved in the later steps of GA biosynthesis and belong to the group of 2-oxoglutarate–dependent dioxygenases that, together with GA 3-oxidases, form biologically active GA (8). *Arabidopsis thaliana* has five *GA20ox* paralogous genes. *AtGA20ox-1*, *AtGA20ox-2*, *AtGA20ox-3*, and *AtGA20ox-4* can catalyze the in vitro conversion of GA_{12} to GA_9. Therefore, *GA20ox* paralogs might have partial redundant functions (9). However, among paralog genes, only *AtGA20ox-1* (*GA5*), which was cloned on the basis of the *ga5* mutant (10), affected plant height (8).

Natural variation for GA biosynthesis has been previously described in *A. thaliana* because the Bur-0 accession carries a loss-of-function allele at *GA20ox4* (9), which does not result in a semidwarf phenotype. In addition, genetic variation in *GA1* has been associated with variation in floral morphology (11). Furthermore, the semidwarf phenotype (here defined as a plant height shorter than half the size of genetically related individuals) observed in the Kas-2 accession is due to a recessive allele at the *GA5* locus (12). The latter finding led to the questions of whether green revolution alleles, artificially selected in cereals, could also occur in natural populations of the wild species *A. thaliana*, and if so, how many different *GA5* loss-of-function alleles exist, how they are distributed, and why they occur in some populations.

Results

Identification, Characterization, and Geographic Distribution of Natural *ga5* Alleles. Phenotypic surveys for plant height in worldwide collections of *A. thaliana* accessions detected 97 individuals collected in 23 different locations showing semidwarf phenotypes. To determine the genetic basis of semidwarfness, we carried out allelism tests by crossing at least one semidwarf from each population to the recessive *ga5* (Ler) mutant (13) and to Ler "wild type" as control (Fig. 1 *A* and *B* and *SI Appendix*, Table S1). To discard the possibility that GA biosynthesis mutations other than *GA5* could account for the semidwarf phenotypes, we tested the complementation of the *ga5* (Ler) mutant by two other semidwarf mutant alleles also affecting GA biosynthesis: *ga4* (Ler),

Significance

Semidwarf accessions occur at low frequency across the distribution range of *Arabidopsis thaliana* and are mainly mutants of the *GA5* (*GA20ox1*) gene, mutations of which originate from wild-type alleles still present in the regions where the mutants were found. We identified the causal mutations by allelism tests and sequencing and performed a detailed population genetics analysis of this variation. Using Fay and Wu's *H* statistics, we obtained indications for local selection of the dwarf alleles. Mutants of functional orthologs of this gene have been selected as the so-called "green revolution genes" in rice and barley, thus indicating that *Arabidopsis* natural variation can be a source for the identification of useful genes for plant breeding.

Author contributions: L.B., M.K., and R.A. designed research; L.B., S.E., and R.K. performed research; C.A.-B., J.J.B.K., and M.K. contributed new reagents/analytic tools; L.B., R.K., J.J.B.K., and R.A. analyzed data; and L.B., C.A.-B., M.K., and R.A. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. KF312645–KF312676).

[1] To whom correspondence should be addressed. E-mail: koornneef@mpipz.mpg.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1314979110/-/DCSupplemental.

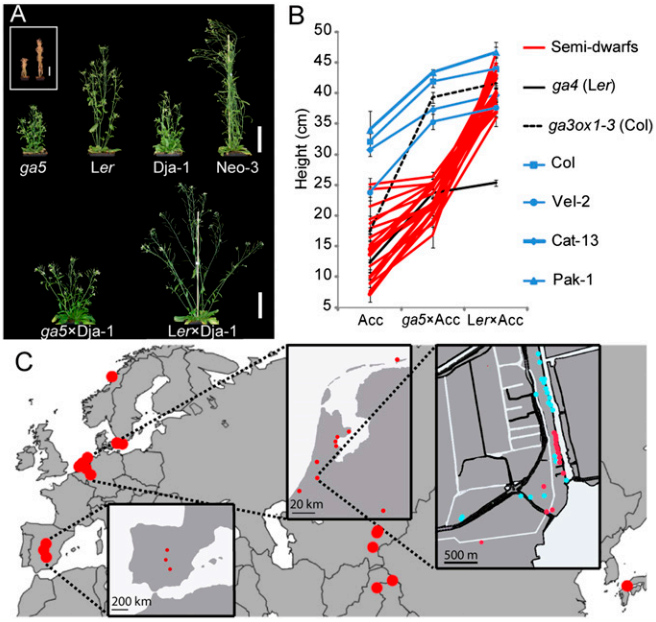

Fig. 1. Semidwarf genotypes allelic to *ga5* are present in nature. (*A*) Allelism test between the semidwarf mutant *ga5* (13) and the semidwarf central Asian accession Dja-1. Neo-3 (central Asia) shows the phenotype of a functional *GA5*. Pictures were taken 2 wk after flowering. The upper left image shows the phenotype of *ga5* and L*er* at harvesting time. (Scale bars, 7 cm.) (*B*) Mean values of stem height ± SE in F_1 plants derived from crosses between *ga5* or L*er* and twenty accessions (Acc) allelic to *ga5* (red), three nondwarf accessions (Col-0, Pak-1, and Cat-13), two semidwarf mutants (*ga4* and *ga3ox1-3*), and one semidwarf accession nonallelic to *ga5* (Vel-2). (*C*) Geographical distribution of semidwarf accessions in Europe, Scandinavia, and central Asia. Red marks indicate the locations of populations containing semidwarf accessions allelic to *ga5*. The right map shows the detailed local distribution of semidwarf (red) and wild-type (blue) individuals found in the Oude Wetering (OW) population from The Netherlands in 2012.

a mutant in the *GA3ox1* gene, and *ga3ox1-3* (Col-0) (14) (Fig. 1*B* and *SI Appendix*, Table S2). L*er* and Col mutants were used to test background effects. Control F_1 plants derived from crosses between nondwarf accessions and *ga5* mutants and F_1 plants grown from crosses with other GA mutants were all taller than their corresponding parents. The crosses *ga5* × *ga4* and L*er* × *ga4* yielded a low height due to the *erecta* mutation, which remained recessive in the F_1. In addition, three accessions showing a weaker semidwarf phenotype (Nfro, Norway; Kar, central Asia; and Vel, Spain) were not allelic to *ga5*, which indicated that other loci accounted for their plant height phenotype. However, for all of the remaining semidwarf accessions tested, the F_1 obtained from the cross to *ga5* exhibited the small size of the parents, whereas semidwarfism was lost in the cross with L*er*. This finding confirmed the recessiveness of the semidwarf alleles. Therefore, most semidwarf accessions were allelic to *ga5*.

To evaluate if there is any general negative pleiotropic effect on plant performance associated with natural *ga5* alleles, we measured several presumably adaptive traits in six wild *ga5* semidwarf accessions and in the *ga5* mutants in L*er* and Col genetic backgrounds (*SI Appendix*, Fig. S1). Consistent with previous studies (8), *ga5* mutants did not differ significantly from their wild types in the evaluated traits (*SI Appendix*, Fig. S1). However, natural *ga5* accessions strongly differed in flowering time, branch number, and silique number, indicating the absence of strong *ga5* effects on these traits but substantial contributions from other genes. Therefore, no major tradeoff on silique number, assumed to be a proxy for fitness, was found for these naturally occurring *ga5* alleles.

Semidwarf *ga5* accessions were found in 23 different populations distributed in western Europe, the Iberian Peninsula, Scandinavia, central Asia, and Japan (Fig. 1*C* and *SI Appendix*, Table S1). From our analysis, we roughly estimated that at worldwide scale, the frequency of wild populations containing semidwarf accessions allelic to *ga5* was at least 1%. However, these frequencies may be higher: because most populations segregate for *GA5* loss-of-function alleles, we cannot discard the possibility that some populations with a limited number of individuals may contain semidwarfs at low frequency not represented in the individuals studied. We also found a semidwarf frequency of 1% in the Hapmap experimental population consisting of 360 worldwide accessions with empirically reduced population structure (15). However, the frequency of *ga5* semidwarf-containing populations was not homogeneous throughout the *A. thaliana* geographic range because we did not find semidwarfs among the many central and east European accessions studied. By contrast, semidwarfism appeared more frequently in central Asia than elsewhere; 5 out of the 24 central Asian populations monitored in this and another study (16) carried semidwarf individuals (*SI Appendix*, Table S1). A ∼2% frequency was estimated for the Iberian Peninsula from the qualitative analysis of the intensive collection (17) used to select the Iberian accessions included in this study. In addition, detailed sampling and analysis of *ga5* semidwarfs in The Netherlands indicated a ∼5% frequency in this region. Dutch semidwarfs seemed to have spread mainly in the west of the country, although one population was found inland (Fig. 1*C*).

Descriptions of the habitat of populations containing *ga5* semidwarf individuals show that they occur in multiple diverse environments where the species occurs. For instance, Dutch semidwarf accessions were found in the anthropoid environments where *A. thaliana* grows, including urban (street populations) and rural (road and field sides; *SI Appendix*, Fig. S2) habitats. However, in the Iberian Peninsula and central Asia, semidwarfs occurred in more natural environments, including Mediterranean forests and mountain wet grasslands (*SI Appendix*, Fig. S2). This wide geographic and ecological distribution indicates that *ga5* semidwarfism does not show a strong geographic structure and is not associated with a single and common climatic factor across its distribution range.

Identification of Multiple *GA5* Loss-of-Function Alleles. To determine the putative mutations causing semidwarf phenotypes, we sequenced the *GA5* gene (∼1.5 kb) in 59 semidwarf accessions collected worldwide and 135 nondwarf individuals, which were collected from the same population or geographic region as the semidwarfs identified. For the Dutch OW and Sch populations, the ∼1-kb *GA5* region, spanning semidwarf causal mutations, was sequenced in 16 semidwarfs and 77 wild-type individuals. Collectively, sequencing data identified 21 different mutations which were predicted to cause *GA5* loss-of-function alleles in semidwarf accessions (Fig. 2). These mutations were classified in six loss-of-function classes according to their nucleotidic nature. First, nonsense mutations causing premature stop codons were found in Kas-2 and Sparta. Second, missense mutations were found close to the conserved metal binding sites of *GA5* in an Iberian (Mar-1, Mar-3, and Mar-11) and a Scandinavian (Var 2–1 and Var 2–6) population, which might underlie their *ga5* phenotype. In addition, the Mdc-10 and Mdc-53 semidwarf accessions also carried missense mutations in GA5 conserved domains. Third, a single substitution in the donor splice site of the first intron was found in all Dutch semidwarf accessions. This affects normal *GA5* splicing generating a truncated GA5 protein. Fourth, seven small insertions (Cat-0, Dja-1, and Pak-3) or deletions (Cat-17, Cat-23, Cat-43, and Sus-1) were predicted to cause frame shifts and truncated GA5 proteins. Fifth, a transposon insertion, with high similarity to *At4g04410*, was identified

PLANT BIOLOGY

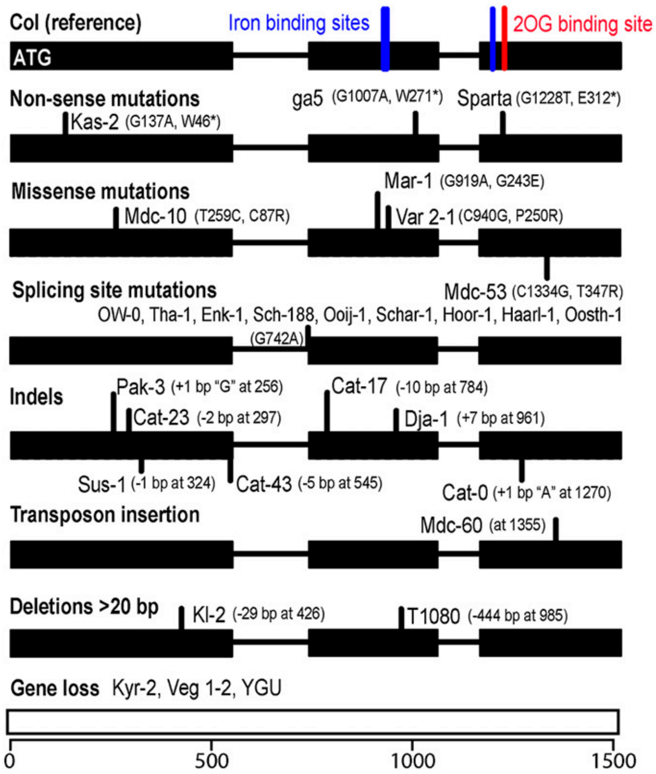

Fig. 2. Natural loss-of-function mutations in the *AtGA20ox1* (*GA5*) gene. The different natures and positions of mutations causing *GA5* loss-of-function alleles are shown in each row. Exons (black boxes) are connected with horizontal lines representing intronic regions of *GA5*. Iron and 2-oxoglutarate binding sites (18) are indicated at the top.

in the MdcA-60 accession. Sixth, several large deletions (>20 bp) were found in some accessions. These included a 29-bp deletion in the first exon of Kl-2 (Germany) and a 444-bp deletion spanning part of the second exon and the complete third exon of accession T1080 (Sweden) (*SI Appendix*, Figs. S3 and S4 and Table S3). This deletion was first detected by the absence of sequence coverage in the 1001 Genomes Project data (www.1001genomes.org) and further confirmed by extensive PCR amplifications (*SI Appendix*, Figs. S3 and S4 and Table S3). In addition, large *GA5* deletions of several kilobases were found in the Veg 1–1, Kyr-2, and YGU accessions. These deletions included not only the coding region but also the promoter (*SI Appendix*, Figs. S3 and S4 and Table S3) and were associated with absence of *GA5* expression in Kyr-2, Veg 1–2, and YGU.

Sequencing analyses indicated that most populations containing semidwarf individuals carry a single loss-of-function mutation in all dwarf plants (e.g., OW-0 in Fig. 2). However, two Iberian populations (Cat and Mdc) segregated for four independent *GA5* loss-of-function mutations (Fig. 2). One allele appeared more frequently; it was present in eight Cat individuals out of 22 sequenced samples. On the other hand, most *GA5* loss-of-function alleles appeared distributed in a single wild population, with the exception of the splicing site mutation widely distributed across The Netherlands. Analysis of the sequence data from the 1001 Genomes Project detected four other putative semidwarf accessions from south Sweden (Sim-1, TV-22, TV-30, and TV-7) that carry the Var 2–1 missense mutation. This result suggests that Var-2 missense loss-of-function allele might be widely distributed at a local scale because Var, Sim, and TV accessions originate from the same south Swedish coastal area (*SI Appendix*, Table S4).

Genome-Wide Association Study for Plant Height. Because several of the *ga5* semidwarf accessions identified in this study (Tha-1, Sparta, Var 2–1, and T1080) were included in the *A. thaliana* Hapmap experimental population (15), we tested whether the *GA5* locus could be detected by genome-wide association study (GWAS) mapping. Measurements of plant height in 345 accessions of this collection showed a large amount of natural variation and high broad-sense heritability ($h_b^2 = 0.80$) (Fig. 3*A*). However, no marker was significantly associated ($P > 0.05$ with Bonferroni correction for 214,000 markers; *SI Appendix*, Fig. S5) with plant height; the largest association was detected on chromosome 4, ~0.3 Mb away from *GA5* ($P = 3 \times 10^{-5}$; Fig. 3*B*). Analysis of linkage disequilibrium (LD) showed a complete LD decay 10 kb upstream and downstream of *GA5* (Fig. 3*C*), thus excluding the linkage of the observed association with *GA5*. By contrast, a significant association was detected when all four *GA5* loss-of-function alleles were combined as a single nonfunctional haplotype ($P = 2.7 \times 10^{-14}$). Therefore, despite the strong effect of natural *GA5* loss-of-function alleles on plant height, GWAS was unable to detect this locus due to the low frequency of semidwarf accessions and their multiple independent causal mutations.

***GA5* Phylogeny and Population Structure.** We determined the genetic relationships among the semidwarf accessions using a structure analysis with 117 genome-wide SNP markers already available (19, 20) or developed in this work. Structure analysis of these accessions found five distinct genetic groups that closely corresponded to the geographic regions of origin of the semidwarf accessions (Fig. 3*D* and *SI Appendix*, Fig. S6) in agreement with the strong global geographic structure described in *A. thaliana* (20). In all cases, semidwarf accessions were genetically more related to the nondwarf individuals from the same population and region than to any other accession, indicating the independent origin and expansion of semidwarfs in these regions. In most populations containing *ga5* semidwarfs where five or more individuals were collected, wild-type *GA5* alleles were found within the populations, except for the central Asian populations Dja and Sus, in which all individuals were semidwarf. Dja-1 and Sus-1 accessions carried different *GA5* loss-of-function alleles (Fig. 3*D*) regardless of the overall low genetic variation present in central Asia (21). It is also remarkable that different *GA5* loss-of-function alleles were found in the Iberian Cat and Mdc populations together with wild-type alleles (Fig. 3*D*). In contrast, semidwarf genotypes in Dutch populations were very similar and carried the same loss-of-function mutation (*SI Appendix*, Fig. S7).

Network analysis of the 33 different *GA5* haplotypes detected within the genomic *GA5* sequence identified a common *GA5* functional haplotype which showed a worldwide distribution (Fig. 3*E* and *SI Appendix*, Table S5). Twenty other *GA5* haplotypes were connected to this frequent haplotype by fewer than five mutational steps and were distributed in all geographic regions. The central network position of the most frequent haplotype suggests that this is the oldest *GA5* allele, from which most other haplotypes may have derived by a small number of mutations (Fig. 3*E*). Furthermore, 14 additional low-frequency haplotypes, which include only Iberian and Asian haplotypes (Cat, Mdc, Mar, Kas, Pak, and Fuk), were separated from the main node of the network by two long related branches. Loss-of-function *GA5* haplotypes appeared evenly distributed within this network, and all but one of these alleles were connected by a single mutational step to their presumably ancestral haplotype. In addition, all loss-of-function haplotypes occupied branch-end positions in this network. Therefore, independent *GA5* loss-of-function alleles seem to be generated in multiple genetic backgrounds, but they have not produced derived haplotypes (Fig. 3*E*).

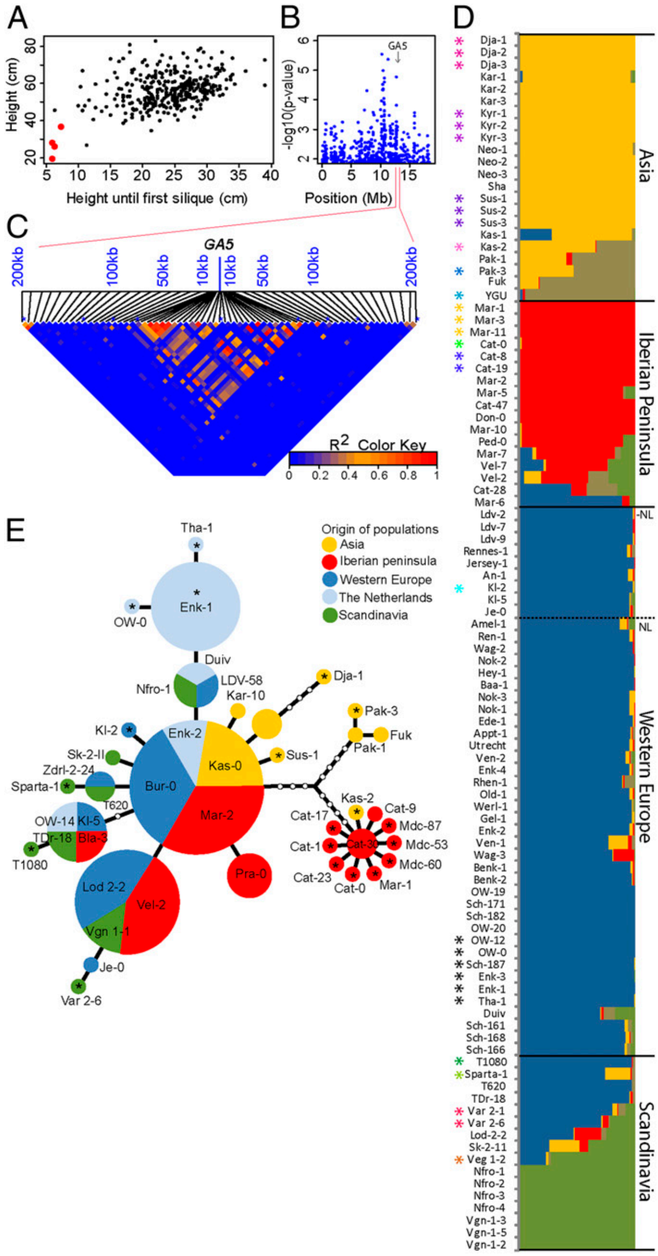

Fig. 3. GWAS analyses, population structure, and *GA5* diversity. (*A*) Correlations between height and height up to first silique. Red dots indicate the values from semidwarf accessions. (*B*) Genome-wide association mapping profile for plant height on chromosome 4. The *GA5* position is indicated by an arrow. (*C*) Linkage disequilibrium 200 kb upstream and downstream of the *GA5* locus. The heat color scale represents squared correlation (R^2) between pairs of SNPs. (*D*) Population structure of 100 accessions including nondwarf and *GA5* semidwarfs collected in different world regions at $K = 5$. Colored asterisks indicate accessions carrying different *GA5* loss-of-function alleles. (*E*) *GA5* haplotype network. Haplotypes are represented by circles with size proportional to the number of populations containing that haplotype. Each node represents a single mutation.

Signatures of Selection at the *GA5* Locus. To estimate the amount and pattern of nucleotide diversity in the *GA5* gene, we analyzed the SNP data from 512 accessions available from the 1001 Genomes Project. *GA5* shows lower nucleotide diversity within coding regions than introns (Fig. 4*A*). Total nucleotide diversity ($\pi = 0.0017$; *SI Appendix*, Table S6) was lower than the average nucleotide diversity reported in previous studies [0.0081 for centromeric and 0.0059 for noncentromeric regions (22)]. *GA5* also presents a low ratio of nonsilent to silent polymorphism [$\pi(\text{ns})/\pi(\text{s}) = 0.132$], which is consistent with a signature of purifying selection, as previously suggested for rice GA biosynthesis genes (23). In addition, significant negative values for Tajima's D at nonsynonymous sites (D_n) were detected in both the aforementioned 512 accessions ($D_n = -2.289$, $P < 0.01$) and the more than 100 accessions used in the present study ($D_n = -1.987$, $P < 0.05$), including semidwarf haplotypes (*SI Appendix*, Table S6). Overall, this pattern is compatible with the occurrence of purifying selection, in which polymorphisms leading to amino acid substitutions are maintained at low frequencies.

To test whether positive selection may have contributed to an increase of *GA5* loss-of-function alleles, we searched for molecular fingerprints of recent selective sweeps over a region of 80 kb upstream and downstream of *GA5* in two populations from two different regions. These Cat (Iberian Peninsula) and OW/Sch (The Netherlands) populations were selected because they contain a moderate frequency of *GA5* loss-of-function alleles. One additional population that does not contain semidwarf individuals from each of the regions was analyzed as control. Significant negative values of the normalized Fay and Wu's H statistics were found around *GA5* in the Cat and OW/Sch populations containing semidwarfs ($0.019 < P < 0.05$) (Fig. 4*B*), which is consistent with an excess of derived high-frequency mutations that commonly accompanies selective sweeps. We also detected negative values for the Fay and Wu's H_n statistics in the semidwarf Iberian Peninsula population around the *CBF* cluster involved in cold acclimation, for which natural variation has been reported (Fig. 4*B*) (24). This pattern was absent in populations without semidwarfs from the same regions (Fig. 4*B*). These results suggest that positive selection might contribute to increase the frequency of *GA5* loss-of-function mutations under particular environments, although drift and relaxed purifying selection could also contribute to a high frequency of *GA5* loss-of-function alleles in some other populations.

Discussion

In this study we have shown that *A. thaliana* semidwarf genotypes are relatively frequent in natural populations of different regions in the world and mostly caused by mutations in the *GA5/GA20ox1* gene. These results evidence a rather simple genetic basis for plant height, but its multiallelic bases hampered *GA5* detection by GWAS mapping. *GA5* behaves as a functional ortholog of the green revolution genes of rice *SD1* and barley *Sdw1/Denso*. This result points to a conserved evolution for this common trait in crop and wild plant species. Thus, GA 20-oxidase is identified as a hot spot for phenotypic variation in plants (25) and illustrates the usefulness of the analysis of *A. thaliana* natural variation to find genes of interest for plant breeding. The observation of major phenotypic changes caused by a large number of independent mutations resembles the situation found for the *FRIGIDA* (*FRI*) gene of *A. thaliana* involved in flowering time, another adaptive trait, which indicates that this pattern is not unique but rather common (17). As previously reported for *FRI* and *Flowering Locus C* (*FLC*) flowering genes, most *GA5* haplotypes show a subregional or local distribution, but the number of independent functional alleles was significantly larger in the Iberian Peninsula than in northern and central Europe, in agreement with the overall larger Iberian diversity (17, 21, 26).

Our study supports the possible contribution of different evolutionary forces to the occurrence of *GA5* loss-of-function alleles in nature. The relatively high frequency of several *GA5* loss-of-function alleles in central Asia and within local populations in The Netherlands, central Asia, and Iberian Peninsula suggests an advantage or neutrality. This is especially the case in some populations where multiple mutations have occurred and are still present. The wide geographic distribution of the same

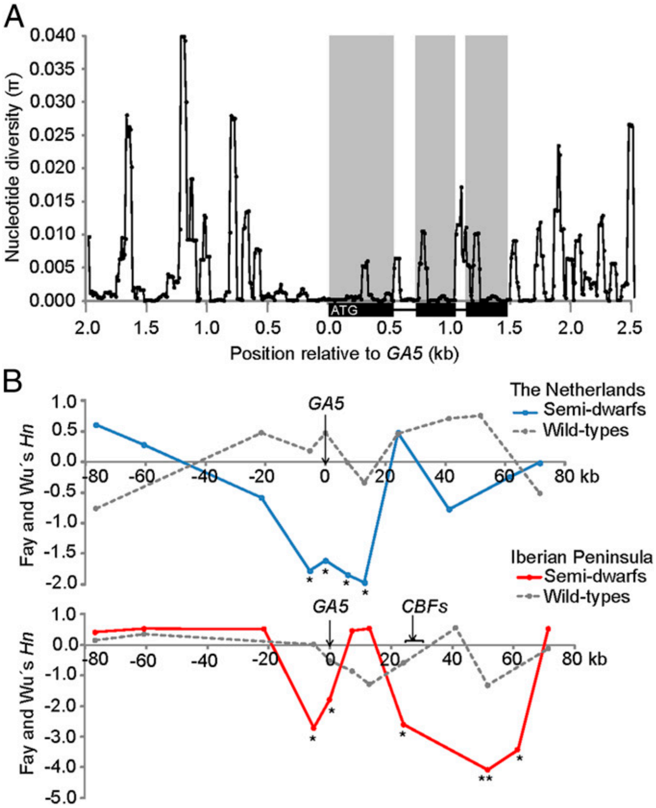

Fig. 4. The *GA5* locus shows signatures of natural selection. (*A*) Nucleotide-sliding window analysis of nucleotide diversity (π) across the *GA5* locus in 512 *A.thaliana* wild accessions. (*B*) Fay and Wu's H_n analysis across the *GA5* genomic region in populations containing semidwarfs from The Netherlands (blue) and the Iberian Peninsula (red) and populations of normal size (gray). Asterisks denote statistical significance: $*P < 0.05$ and $**P < 0.01$.

GA5 allele found in many locations of The Netherlands separated by more than 100 km indicates that this allele is spreading, further indicating the absence of deleterious effects. In addition, phenotypic characterization of *GA5* semidwarf accessions did not detect any strong negative effect on adaptive and fitness traits, which suggests that these alleles do not display any general obvious negative pleiotropic effects or tradeoffs. This result is in agreement with the phenotypes described for artificially induced *GA5* loss-of-function mutants, which show similar seed yields to those of wild-type accessions (8). This lack of effect on seed production is probably due to expression of *GA20ox* paralogs, mainly *GA20ox2* (8). Similarly, *GA20ox2* mutations in rice and barley do not display tradeoffs (5, 6, 27). By contrast, mutations in early steps of GA biosynthesis have been associated with negative pleiotropic effects, such as the absence of seed germination shown by *ga1* null mutants or the reduced fertility and altered flower development observed even in leaky *GA1* alleles (13). A similar situation has been reported in rice, where the effects derived from mutations on genes involved in early steps of GA biosynthesis were less favorable for crop production compared with mutations on rice *GA20ox2* (28).

Our analyses suggest that both negative and positive selection may act on *GA5* loss-of-function alleles. The conditional negative effect of these alleles is suggested by the low frequency of most loss-of-function alleles and by the fact that they are not maintained long enough to derive new haplotypes. Hence, such alleles seem to be transiently maintained in nature. In addition, such a potential negative effect of *GA5* loss-of-function alleles is also suggested by purifying selection inferred from the low ratio of replacement to silent polymorphisms and negative Tajima's D_n values, in agreement with previous reports in rice (23). In contrast, positive selection might contribute to transient increases in the frequency of loss-of-function alleles in certain populations, as suggested by the negative values of Fay and Wu's H_n tests across the *GA5* locus for the two tested populations segregating for semidwarf individuals. Remarkably, this pattern is absent in populations of normal-sized plants from the same regions. Therefore, we reason that allelic variation at *GA5* locus might be maintained in nature by antagonist pleiotropy, (i.e., reversed fitness effects in different environments) (29). However, we cannot discard the possibility that *GA5* variation shows conditional neutrality in other populations (i.e., loss-of-function alleles might be neutral in some environments but deleterious in others). Neutrality tests should be considered carefully due to the complex demographic history of *A. thaliana* populations in the wild. Furthermore, the population genetic analysis is agnostic to the local extinction or recolonization dynamics of populations. The identification of signatures for selection using genome-wide screens may be hampered by the occurrence of different loss-of-function *GA5* alleles under positive selection, a situation that also affected GWA mapping.

It remains to be determined which are the environmental cues that could contribute to an increase in the frequency of *GA5* loss-of-function alleles because these mutations appear distributed in a wide range of anthropoid and natural environments. It has been previously shown that the short-plant height phenotype caused by the well-known *erecta* loss-of-function mutation provides fitness advantage in static landscapes. On the contrary, the *erecta* frequency was reduced under disturbed environments (30). Analogously, it can be speculated that environmental stability might favor *GA5* semidwarf individuals. Conclusive demonstration about positive, negative, or neutral fitness effects of *GA5* loss-of-function alleles depending on the environment will require further analyses under different natural conditions to elucidate the evolutionary forces driving *GA5* variation and its ecological significance.

Materials and Methods

Plant Material and Growth Conditions. Stock numbers and detailed information for accessions used in this work are listed in *SI Appendix*, Table S1. For allelism tests, semidwarf accessions were crossed with L*er* and *ga5* (13). To facilitate the allelism tests, male sterility based on the *ms1* mutant (31) was introgressed into the *ga5* background. Plants were grown under greenhouse conditions at 16 h light, 22 °C/18 °C day/night cycles. For all experiments, seeds were stratified in water at 4 °C for 4–6 d before germination. Ten repetitions per genotype (cross) were conducted. All crossed accessions are listed in *SI Appendix*, Table S2. The Ooij, Schar, Hoor, Haarl, and Oosth Dutch semidwarf populations and the Mdc Iberian semidwarf population were found in the course of our studies. Allelism was concluded based on sequence data that correlated with the semidwarf phenotypes and haplotypes tested before in allelism tests. Phenotyping for plant height and height up to first silique was conducted 2 wk after flowering because both traits did not change after that date (*SI Appendix*, Fig. S8). In cases of extreme flowering lateness, plants were vernalized for 6 wk. Flowering time was recorded as days after germination until the first opened flower. Branch number was scored as the number of axillary stems grown from the rosette.

Sequencing of *GA5* Gene and Genotyping. Genomic DNA was isolated from leaf material using the BioSprint workstation (Qiagen). Primers used for *GA5* sequencing are detailed in *SI Appendix*, Table S7. PCR reactions were performed using LA Taq DNA polymerase (Takara) following manufacturer's instructions. Sanger sequencing of purified PCR products was done by the Max Planck Genome Center Cologne (Cologne, Germany). GenBank accession numbers of DNA sequences generated in this work are listed in *SI Appendix*, Table S5. SNP genotyping of new accessions collected in this study was done as described in previous works (19, 20) using the genotyping facility service of the University of Chicago (Chicago).

Statistical Analysis. Descriptive statistics, *t* tests, Tukey tests, and principal component analyses were conducted with R. The method of efficient mixed-model association expedited was used for GWAS (32) using kinship matrix to correct for population structure. LD analysis was performed with the R package LD heatmap (33).

Structure Analysis. Population structure was inferred using model-based clustering algorithms implemented in the software STRUCTURE, using the haploid setting and running 20 replicates with 50,000 and 20,000 MCMC (Markov chain Monte Carlo) iterations of burn-in and after-burning length, respectively (34). To determine the *K* number of significantly different genetic clusters, we applied the ΔK method in combination with the absolute value of ln $P(X|K)$ (35).

Population Genetics: Fay and Wu's *H* Statistics and Haplotype Network. Population genetics analyses were conducted with the software DnaSP (DNA Sequence Polymorphism version 5.10) (36). The normalized Fay and Wu's *H* was performed as described (16) in populations containing semidwarfs from The Netherlands and the Iberian Peninsula (*SI Appendix*, Table S8). Representative accessions of different populations from central Spain, with no semidwarfs, were used as control (*SI Appendix*, Table S8). For the Dutch control population, accessions from a rural area northeast of Wageningen were collected with no prior knowledge of semidwarfness occurring in this population (*SI Appendix*, Table S8). The sequences of *GA5* (*At4g25420*) and flanking genes were obtained after specific PCR amplification from genomic DNA and sequencing using Applied Biosystems 3730XL automated sequencers (*SI Appendix*, Table S7). Sequences were aligned with ClustalW (37) and manually inspected. *Arabidopsis lyrata* sequences were obtained by BLAST search (www.phytozome.net) and used as out-group to assign ancestral and derived states to SNP variants. To assess the statistical significance of Fay and Wu's *H*, we computed 10,000 coalescent simulations in DnaSP version 5.10 (36). The haplotype network of *GA5* was constructed using TCS1.21 (38) that implements a maximum parsimony method and excluding gaps as events in the analysis. Insertions and deletions in the semidwarf accessions were considered as single events and added manually to the haplotype network.

ACKNOWLEDGMENTS. We thank Peter Hedden for providing mutant seeds relevant to our experiments. We also thank the collectors in The Netherlands who provided semidwarf seeds, Olivier Loudet for the central Asian accessions, and Odd Arne Rognli and Ilkka Kronholm for the Norwegian material. We also thank Frank Becker (Wageningen University and Research Centre) for providing crosses and information regarding the accessions present in the Hapmap, Matthew Horton and Odd Arne Rognli for providing SNP marker data, Marianne Harperscheidt and Regina Gentges for plant handling, and the Max Planck Genome Centre Cologne for dealing with sequencing. R.A. acknowledges support from the Ramón y Cajal Programme of the Ministerio de Ciencia e Innovación (RYC-2011-07847) and the Deutsche Forschungsgemeinschaft Grant SFB680. L.B. was supported by an International Max Planck Research School PhD Fellowship. C.A.-B. was funded by Grant BIO2010-15022 from the Ministerio de Ciencia e Innovación of Spain. R.K. acknowledges funding from the Centre for Biosystems Genomics. Funding was provided by the Max Planck Gesellschaft.

1. Hedden P, Thomas SG (2012) Gibberellin biosynthesis and its regulation. *Biochem J* 444(1):11–25.
2. Yamaguchi S (2008) Gibberellin metabolism and its regulation. *Annu Rev Plant Biol* 59(1):225–251.
3. Hedden P (2003) The genes of the Green Revolution. *Trends Genet* 19(1):5–9.
4. Salamini F (2003) Plant biology. Hormones and the green revolution. *Science* 302(5642):71–72.
5. Sasaki A, et al. (2002) Green revolution: A mutant gibberellin-synthesis gene in rice. *Nature* 416(6882):701–702.
6. Spielmeyer W, Ellis MH, Chandler PM (2002) Semidwarf (*sd-1*), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. *Proc Natl Acad Sci USA* 99(13): 9043–9048.
7. Jia Q, et al. (2009) GA-20 oxidase as a candidate for the semidwarf gene *sdw1/denso* in barley. *Funct Integr Genomics* 9(2):255–262.
8. Rieu I, et al. (2008) The gibberellin biosynthetic genes *AtGA20ox1* and *AtGA20ox2* act, partially redundantly, to promote growth and development throughout the Arabidopsis life cycle. *Plant J* 53(3):488–504.
9. Plackett AR, et al. (2012) Analysis of the developmental roles of the Arabidopsis gibberellin 20-oxidases demonstrates that *GA20ox1*, *-2*, and *-3* are the dominant paralogs. *Plant Cell* 24(3):941–960.
10. Xu YL, et al. (1995) The *GA5* locus of *Arabidopsis thaliana* encodes a multifunctional gibberellin 20-oxidase: Molecular cloning and functional expression. *Proc Natl Acad Sci USA* 92(14):6640–6644.
11. Brock MT, Kover PX, Weinig C (2012) Natural variation in GA1 associates with floral morphology in *Arabidopsis thaliana*. *New Phytol* 195(1):58–70.
12. 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(3):1867–1876.
13. Koornneef M, van der Veen J (1980) Induction and analysis of gibberellin sensitive mutants in *Arabidopsis thaliana* (L.)Heynh. *Theor Appl Genet* 58(6):257–263.
14. Mitchum MG, et al. (2006) Distinct and overlapping roles of two gibberellin 3-oxidases in Arabidopsis development. *Plant J* 45(5):804–818.
15. Li Y, Huang Y, Bergelson J, Nordborg M, Borevitz JO (2010) Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 107(49):21199–21204.
16. Alcázar R, et al. (2010) Natural variation at Strubbelig Receptor Kinase 3 drives immune-triggered incompatibilities between *Arabidopsis thaliana* accessions. *Nat Genet* 42(12):1135–1139.
17. Méndez-Vigo B, Picó FX, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C (2011) Altitudinal and climatic adaptation is mediated by flowering traits and *FRI*, *FLC*, and *PHYC* genes in Arabidopsis. *Plant Physiol* 157(4):1942–1955.
18. Wilmouth RC, et al. (2002) Structure and mechanism of anthocyanidin synthase from *Arabidopsis thaliana*. *Structure* 10(1):93–103.
19. Lewandowska-Sabat AM, Fjellheim S, Rognli OA (2010) Extremely low genetic variability and highly structured local populations of *Arabidopsis thaliana* at higher latitudes. *Mol Ecol* 19(21):4753–4764.
20. Platt A, et al. (2010) The scale of population structure in *Arabidopsis thaliana*. *PLoS Genet* 6(2):e1000843.
21. Cao J, et al. (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* 43(10):956–963.
22. Schmid KJ, Ramos-Onsins S, Ringys-Beckstein H, Weisshaar B, Mitchell-Olds T (2005) A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics* 169(3):1601–1615.
23. Yang YH, Zhang FM, Ge S (2009) Evolutionary rate patterns of the Gibberellin pathway genes. *BMC Evol Biol* 9:206.
24. Alonso-Blanco C, et al. (2005) Genetic and molecular analyses of natural variation indicate *CBF2* as a candidate gene for underlying a freezing tolerance quantitative trait locus in Arabidopsis. *Plant Physiol* 139(3):1304–1312.
25. Martin A, Orgogozo V (2013) The Loci of repeated evolution: A catalog of genetic hotspots of phenotypic variation. *Evolution* 67(5):1235–1250.
26. Picó FX, Méndez-Vigo B, Martínez-Zapater JM, Alonso-Blanco C (2008) Natural genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian peninsula. *Genetics* 180(2):1009–1021.
27. Jia Q, et al. (2011) Expression level of a gibberellin 20-oxidase gene is associated with multiple agronomic and quality traits in barley. *Theor Appl Genet* 122(8):1451–1460.
28. Itoh H, et al. (2004) A rice semi-dwarf gene, Tan-Ginbozu (D35), encodes the gibberellin biosynthesis enzyme, ent-kaurene oxidase. *Plant Mol Biol* 54(4):533–547.
29. Anderson JT, Lee CR, Rushworth CA, Colautti RI, Mitchell-Olds T (2013) Genetic trade-offs and conditional neutrality contribute to local adaptation. *Mol Ecol* 22(3): 699–708.
30. Fakheran S, et al. (2010) Adaptation and extinction in experimentally fragmented landscapes. *Proc Natl Acad Sci USA* 107(44):19120–19125.
31. van der Veen JH, Wirtz P (1968) EMS-induced genic male sterility in *Arabidopsis thaliana*: A model selection experiment. *Euphytica* 17(3):371–377.
32. Kang HM, et al. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42(4):348–354.
33. Shin J-H, Blay S, McNeney B, Graham J (2006) LDheatmap: An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J Stat Software* 16:Code Snippet 3.
34. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959.
35. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol* 14(8):2611–2620.
36. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451–1452.
37. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673–4680.
38. Clement M, Posada D, Crandall KA (2000) TCS: A computer program to estimate gene genealogies. *Mol Ecol* 9(10):1657–1659.