What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation?

Carlos Alonso-Blanco,a Mark G.M. Aarts,b Leonie Bentsink,c Joost J.B. Keurentjes,b,d,e Matthieu Reymond,f Dick Vreugdenhil,d,e and Maarten Koornneefb,f,1

a Centro Nacional de Biotecnologı´ a, Consejo Superior de Investigaciones Cientı´ ficas, Departamento de Gene´ tica Molecular de Plantas, Cantoblanco 28049 Madrid, Spain
b Laboratory of Genetics, Wageningen University, 6700 AH Wageningen, The Netherlands
c Molecular Plant Physiology Group, Institute of Environmental Biology, Utrecht University, 3584 CH Utrecht, The Netherlands
d Laboratory of Plant Physiology, Wageningen University, 6700 AR, Wageningen, The Netherlands
e Centre for Biosystems Genomics, 6708 PB Wageningen, The Netherlands
f Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany

Nearly 100 genes and functional polymorphisms underlying natural variation in plant development and physiology have been identified. In crop plants, these include genes involved in domestication traits, such as those related to plant architecture, fruit and seed structure and morphology, as well as yield and quality traits improved by subsequent crop breeding. In wild plants, comparable traits have been dissected mainly in Arabidopsis thaliana. In this review, we discuss the major contributions of the analysis of natural variation to our understanding of plant development and physiology, focusing in particular on the timing of germination and flowering, plant growth and morphology, primary metabolism, and mineral accumulation. Overall, functional polymorphisms appear in all types of genes and gene regions, and they may have multiple mutational causes. However, understanding this diversity in relation to adaptation and environmental variation is a challenge for which tools are now available.

INTRODUCTION

Intraspecific natural variation (hereinafter referred to as natural variation) may be broadly defined as the within-species phenotypic variation caused by spontaneously arising mutations that have been maintained in nature by any evolutionary process including, among others, artificial and natural selection. Thus, natural variation embraces the enormous diversity present within wild plant species as well as most of the genetic variants that are found in domesticated plants. Some of the phenotypic differences existing in wild or cultivated plants are due to single-gene (monogenic) allelic variants. However, most of the natural variation is quantitative and determined by molecular polymorphisms at multiple loci and genes (multigenic), which are referred to as quantitative trait loci (QTL) and quantitative trait genes (QTGs).

The natural variation present in crop plants has been exploited since their domestication thousands of years ago by the genetic manipulation of developmental traits and physiological features related to adaptation to agriculture. Methods for genetic analysis and mapping of natural quantitative variation were developed a few decades ago for crop species, in which many more studies have been performed than in wild plants. Currently, genomic resources have been developed for several herbaceous crop plants, such as rice (Oryza sativa), barley (Hordeum vulgare), and tomato (Solanum lycopersicum), enabling the identification of the genes and nucleotide polymorphisms underlying QTL involved in domestication, yield, biotic and abiotic stress, and quality traits (Doebley et al., 2006; Sang, 2009). In addition, the recent availability of the first genome sequences of cultivated perennial woody plants, such as poplar (Populus spp) and grapevine (Vitis vinifera; Tuskan et al., 2006; Jaillon et al., 2007), is increasing the scope of traits and processes whose natural variation can be dissected molecularly.

In addition, the analysis of natural variation in wild species has begun to elucidate the molecular bases of phenotypic differences related to plant adaptation to distinct natural environments and to determine the ecological and evolutionary processes that maintain this variation (Mitchell-Olds et al., 2007). The molecular approach to these questions has been applied mainly to the wild annual crucifer Arabidopsis thaliana, which has become a model plant for the study of natural variation (Mitchell-Olds and Schmitt, 2006). Its wide geographical and environmental distribution, combined with its small genome and the availability of unprecedented genetic and genome resources, have strongly facilitated its molecular analysis in the last decade. As a result, A. thaliana has provided the largest number of genes and nucleotide polymorphisms underlying natural variation of any plant species (Alonso-Blanco et al., 2005). However, the specific ecological niche and life history of A. thaliana limits the plant traits and

1 Address correspondence to maarten.koornneef@wur.nl.
www.plantcell.org/cgi/doi/10.1105/tpc.109.068114
processes that can be approached in a single species. Therefore, new plant models phylogenetically related to A. thaliana (e.g., Arabidopsis lyrata; Claus and Koch, 2006) as well as unrelated species (e.g., of the genera Aquilegia [Kramer, 2009], Mimulus [Wu et al., 2007], Ipomoea [Clegg and Durbin, 2003], and Helianthus [Rieseberg et al., 2003]) are beginning to be used in studies of natural variation and speciation.

Our understanding of plant biology has benefited tremendously from work done with artificially induced mutants in model species, as illustrated by A. thaliana mutants currently available for most genes, and mutant screenings continue to be useful in the identification of gene function (Page and Grossniklaus, 2002). Gene functions involved in plant survival and adaptation can be identified partially by induced mutant analyses of different wild genotypes, where mutants with reduced fitness are easily selected. However, current mutant collections have been obtained using a limited number of laboratory strains, which harbor only a small portion of A. thaliana natural variation. Interestingly, Clark et al. (2007) showed that ~9.4% of A. thaliana protein-coding genes are naturally absent or knocked out in wild accessions, limiting the mutant spectra that can be obtained from each accession. Therefore, natural variation provides a relevant complementary resource to discover novel gene functions as well as those allelic variants that specifically interact with the genetic background and/or the environment or alleles showing small effects on phenotype, particularly for traits related to plant adaptation (Benfey and Mitchell-Olds, 2008).

Genetic analyses of natural variation in plants currently are mainly performed by QTL mapping, often called linkage mapping, in which phenotypic variation is associated with allelic variation at molecular markers segregating in experimental mapping populations derived from directed crosses (Doerge, 2002). Thus, genomic regions accounting for trait variation are located in large physical intervals containing the causal QTLs. Further analyses of these regions with a combination of functional strategies allows the final identification of QTGs and nucleotide polymorphisms altering the function of those genes (reviewed in Koornneef et al., 2004; Alonso-Blanco et al., 2005; Doebley et al., 2007). Association mapping, which involves looking for phenotype-genotype associations in a general population of individuals whose degree of relatedness or pedigree is unknown, is also becoming more popular and useful in plant systems, owing to improvements in statistical and analytical tools and in gene sequencing technology (see Myles et al., 2009).

The information gleaned from the various types of QTL analysis is extending the current concept of gene function based on a single wild-type allele and provides a more realistic and dynamic view of gene function by estimating the effects of different natural alleles depending on genetic background, environment, and evolutionary temporal scale. In addition, recent genetic analyses of large data sets generated by high-throughput profiling of mapping populations with “-omics” procedures (genomics, transcriptomics, metabolomics, proteomics, ionomics, or epigenomics) is allowing the genetic integration of several levels of molecular regulation of phenotypic trait variation. Thus, QTL and gene networks controlling natural variation are being inferred (Jansen et al., 2009).

In this review, we discuss the major contributions of the analysis of natural variation to our understanding of plant development and physiology. Given the large number of studies performed in the past, we do not aim to completely cover all traits analyzed so far, but to focus on several relevant phenotypes recently or extensively dissected in A. thaliana and major crop species. These traits include the timing of germination and flowering, plant growth and morphology, primary metabolism, and mineral accumulation. Detailed discussion of natural variation for secondary metabolites is covered by Kliebenstein (2009a), and various reviews have specifically described the natural variation involved in crop domestication and diversification (Doebley et al., 2006; Pourkheirandish and Komatsuda, 2007; Purugganan and Fuller, 2009; Sang, 2009), plant development (Alonso-Blanco et al., 2005), or resistance to pathogens (Holub, 2007). Here, we summarize the major genes and allelic variants currently known to participate in plant adaptation to different natural and agricultural environments through some modifications of plant development and physiology. The kind of mutations, molecular mechanisms, and gene networks involved in microevolution and adaptation of different plant species to changing environments are beginning to be elucidated. Consequently, new perspectives in our understanding of phenotype and pan-genome variation and evolution are emerging (Morgante et al., 2007).

**DEVELOPMENTAL PROCESSES**

Domestication of crop plants mainly has resulted in changes in traits related to floral and seed morphology, such as shattering, free-threshing, and seed coat color in cereals (Salamin et al., 2002, Dubcovsky and Dvorak, 2007), plant architecture-like branching patterns in maize (Zea mays) and rice (Izawa et al., 2009), fruit morphology-like size and shape in Solanaceae crops (Tanksley, 2004), and seed dormancy, especially in domesticated cereals (Salamin et al., 2002). In addition, the analysis of natural variation has identified novel molecular mechanisms involved in the regulation of life history traits, such as seed dormancy and flowering time, by different environmental and endogenous clues. This section highlights some of the most important advances resulting from these studies.

**Seed Dormancy and Germination**

Most wild annual plant species, including A. thaliana, show substantial intraspecific natural variation for seed dormancy and germination properties, which is presumably involved in adaptation to different environments. Domestication has often led to a reduction of seed dormancy in crop plants, although considerable variation remains among cultivated varieties. The genetic bases of this variation have been approached mainly in A. thaliana and cereal plants, mostly in relation to environmental factors, such as temperature and light, and to endogenous hormonal signals, such as abscisic acid (ABA) and gibberellin (GA).

In A. thaliana, QTL mapping analyses of germination-related traits have been performed in relation to after-ripening
(Alonso-Blanco et al., 2003; Clerkx et al., 2004), light regimes (van der Schaar et al., 1997; Laserna et al., 2008; Meng et al., 2008), low temperature (Meng et al., 2008), or GA inhibitors (van der Schaar et al., 1997). Many of the loci are only detected in specific assays or conditions, indicating strong QTL by environment interactions. However, several QTLs collocate in different crosses, suggesting the presence of moderate frequency natural alleles with pleiotropic effects on dormancy and related traits, such as cold-tolerant germination and seed longevity (Clerkx et al., 2004; Meng et al., 2008). Particularly, the locus DOG1 has been detected in three experimental populations (Alonso-Blanco et al., 2003; Clerkx et al., 2004; Laserna et al., 2008) and has been isolated, identifying a member of a small gene family of unknown molecular function (Bentsink et al., 2006; Table 1). DOG1 induces seed dormancy and is specifically expressed during seed development. In addition, it has been shown that DOG1 functional nucleotide polymorphism(s) affect its expression, dormant accessions having higher RNA levels than nondormant accessions (Bentsink et al., 2006). Furthermore, DOG1 is involved in the ABA-mediated sugar signaling pathway in seedlings since glucose induces the expression of the DOG1/GSQ5 Cvi allele, whereas Landsberg erecta and Columbia alleles do not respond (Teng et al., 2008). Thus, DOG1 has provided the first gene involved in natural variation for seed dormancy and illustrates the value of this variation to identify previously unknown genes for this trait.

In cereals, several analyses have identified seed dormancy QTLs in crosses between cultivated varieties with low dormancy and between cultivated and dormant weedy accessions or wild relative species (Lin et al., 1998; Gu et al., 2005; Hori et al., 2007; Imtiaz et al., 2008). In particular, the large-effect locus Sdr1 of rice might be orthologous to barley SD4 and wheat (Triticum aestivum) 4AL loci (Lin et al., 1998). Detailed QTL characterizations have shown that some loci may control dormancy through seed maternal tissues, while others function in offspring tissue(s) (Gu et al., 2008). As in wheat, maternally inherited seed dormancy has been associated with the seed pigmentation color variation determined by the rice Rc locus (Gu et al., 2005), which encodes a bHLH transcription factor (Sweeney et al., 2006). In addition, rice QTLs have been found for germination at low temperature, and the first of such loci, qLTG3-1, has been recently isolated (Fujino et al., 2008). qLTG3-1 encodes a protein of unknown function that is expressed in the embryo during seed germination. It has been suggested that qLTG3-1 is involved in tissue weakening during seed germination, which is reduced in the nondormant rice accession bearing a loss-of-function deletion allele.

Genetic analyses of dormancy and germination also have been performed in other crop species, such as lettuce (Lactuca sativa; Argyris et al., 2005). Interestingly, an interspecific cross between L. sativa and L. serriola identified Htg6.1 as a major effect locus involved in germination at high temperature (thermo-inhibition), which colocalizes with germination QTLs for ABA sensitivity and GA response as well as with the ABA biosynthesis gene Ls-CE4. In addition, Ls-CE4 is induced by high temperature only in the thermoinhibited cultivated species, which strongly supports Ls-CE4 as candidate for Htg6.1 (Argyris et al., 2008).

### Flowering Time

Molecular studies of the natural variation in the annual species *A. thaliana*, pea (*Pisum sativum*), rice, barley, wheat, and maize have identified ~25 different causal genes and their nucleotide polymorphisms of major effect (Table 1). In addition, genetic analyses of flowering time or related traits have been performed in other annual plants, such as tomato (González-Martínez et al., 2006), biennial species like *Brassica* sp (Schranz et al., 2002), short-lived perennials like *Fragaria vesca* (Albani et al., 2004), and several long-lived woody trees (Frewen et al., 2000; Casasoli et al., 2006). These studies are identifying multiple candidate genes and polymorphisms, which allow comparative analyses among species with rather different environmental and ecological distributions as well as physiological and developmental behaviors.

*A. thaliana* is a facultative long-day species widely distributed in temperate regions of the Northern hemisphere. Two extreme flowering behaviors have been described among wild genotypes grown in laboratory conditions, which are considered winter (late) and spring (early) life habits because they show strong and weak vernalization responses, respectively. Nine genes have been isolated that contribute to A. *thaliana* natural variation for flowering time (Table 1). Among these, *FRI* and *FLC* are involved in the regulation of flowering by vernalization and were first identified because of this natural variation. *FLC* encodes a transcription factor of the MADS family that represses flowering (Michaels and Amasino, 1999). Its expression is downregulated by the low temperature of winter through epigenetic mechanisms leading to histone methylation and chromatin modifications (reviewed in Sung and Amasino, 2005). *FRI* encodes a protein of unknown function that also represses flowering initiation by activating the expression of *FLC* (Figure 1; Johanson et al., 2000). Most winter annual genotypes carry active alleles of both genes, which interact epistatically to delay flowering, while many summer annual accessions bear partial or total loss-of-function alleles in one or both genes (Werner et al., 2005a). More than 25 *FRI* independent loss-of-function alleles have been described, which mainly encode truncated proteins, indicating that *FRI* loss-of-function has evolved multiple times from a late flowering ancestor (Johanson et al., 2000; Le Corre et al., 2002; Shindo et al., 2005). By contrast, several natural loss-of-function alleles of *FLC* with reduced gene expression are caused by independent insertions of transposon elements within the first regulatory intron (Gazzani et al., 2003; Michaels et al., 2003). Other rare alleles are produced by nonsense or splicing mutations predicted to generate truncated proteins (Werner et al., 2005a). Interestingly, it has been suggested that additional allelic variation in *FLC* cis-regulatory sequences that alters the efficiency of *FLC* chromatin silencing might contribute to the quantitative variation for vernalization response present among late flowering genotypes (Shindo et al., 2006).

QTL mapping studies have also shown that other loci, including homologs of *FRI* and *FLC*, affect the vernalization response (Figure 1). Natural loss-of-function alleles in two *FRI*-like genes (*FRL1* and *FRL2*; Schläppi, 2006) and one *FLC*-like gene named *FLM*/FLU1 (Werner et al., 2005b) have been shown to produce early flowering. Additional flowering time QTLs affecting this...
<table>
<thead>
<tr>
<th>Plant Species</th>
<th>Locus/Gene</th>
<th>Phenotypic Effects</th>
<th>Molecular Function</th>
<th>Gene Position</th>
<th>Mutation</th>
<th>Allelic Dysfunction</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. thaliana</td>
<td>DOG1</td>
<td>Seed dormancy, sugar sensing</td>
<td>Unknown</td>
<td>Promoter</td>
<td>Unknown</td>
<td>ME</td>
<td>Bentsink et al. (2006)</td>
</tr>
<tr>
<td>Thale cress</td>
<td>FRI</td>
<td>Flowering</td>
<td>Unknown</td>
<td>Promoter, coding</td>
<td>Ins, Del, nonsense Subs</td>
<td>TP, ME</td>
<td>Johanson et al. (2003); Shindo et al. (2005)</td>
</tr>
<tr>
<td></td>
<td>FFL1</td>
<td>Flowering</td>
<td>FRI-like</td>
<td>Coding</td>
<td>Nonsense Subs</td>
<td>TP</td>
<td>Schläppi (2006)</td>
</tr>
<tr>
<td></td>
<td>FFL2</td>
<td>Flowering</td>
<td>FRI-like</td>
<td>Coding</td>
<td>Missense Subs</td>
<td>AP</td>
<td>Schläppi (2006)</td>
</tr>
<tr>
<td></td>
<td>FLC</td>
<td>Flowering</td>
<td>MADS TF</td>
<td>Intron, coding</td>
<td>TE insertion, nonsense Subs</td>
<td>ME, TP</td>
<td>Michaels et al. (2003); Werner et al. (2005a)</td>
</tr>
<tr>
<td></td>
<td>FLW1/FLM</td>
<td>Flowering</td>
<td>MADS TF</td>
<td>Promoter + coding</td>
<td>Gene Del</td>
<td>Null expression</td>
<td>TP, AP</td>
</tr>
<tr>
<td>ART1/HUA2</td>
<td></td>
<td>Flowering, shoot morphology</td>
<td>RNA processing factor</td>
<td>Coding</td>
<td>Missense Subs</td>
<td>AP</td>
<td>El-Assal et al. (2001, 2004)</td>
</tr>
<tr>
<td></td>
<td>EDI/CRY2</td>
<td>Flowering, seedling growth, fruit size</td>
<td>Photoreceptor</td>
<td>Coding</td>
<td>Missense Subs</td>
<td>AP</td>
<td>Schlaeppi (2006); Schlaeppi (2006)</td>
</tr>
<tr>
<td></td>
<td>PHYD</td>
<td>Flowering, seedling growth</td>
<td>Photoreceptor</td>
<td>Coding</td>
<td>Del</td>
<td>TP</td>
<td>Schlaeppi (2006); Schlaeppi (2006)</td>
</tr>
<tr>
<td></td>
<td>PHYC</td>
<td>Flowering, seedling growth</td>
<td>Photoreceptor</td>
<td>Coding</td>
<td>Nonsense Subs</td>
<td>TP</td>
<td>Balasubramanian et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>PHYA</td>
<td>Seeding growth</td>
<td>Photoreceptor</td>
<td>Coding</td>
<td>Missense Subs</td>
<td>AP</td>
<td>Maloof et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>PHYB</td>
<td>Seeding growth</td>
<td>Photoreceptor</td>
<td>Coding</td>
<td>Missense Subs</td>
<td>AP</td>
<td>Filiault et al. (2008)</td>
</tr>
<tr>
<td></td>
<td>TZP/Light5</td>
<td>Seeding growth</td>
<td>Zn knuckle protein</td>
<td>Coding</td>
<td>Ins</td>
<td>TP</td>
<td>Loudet et al. (2008)</td>
</tr>
<tr>
<td>CAL</td>
<td></td>
<td>Flower morphology</td>
<td>MADS TF</td>
<td>Coding</td>
<td>Missense Subs</td>
<td>AP</td>
<td>Kempin et al. (1995)</td>
</tr>
<tr>
<td></td>
<td>IIL1</td>
<td>Leaf morphology</td>
<td>Leu biosynthesis enzyme</td>
<td>Intron</td>
<td>Ins, Del</td>
<td>ME</td>
<td>Sureshkumar et al. (2009)</td>
</tr>
<tr>
<td>GL1</td>
<td></td>
<td>Trichome formation</td>
<td>Myb TF</td>
<td>Coding</td>
<td>Del</td>
<td>ME</td>
<td>Hauser et al. (2001)</td>
</tr>
<tr>
<td>PUB8</td>
<td></td>
<td>Self-incompatibility</td>
<td>U box protein</td>
<td>Promoter</td>
<td>Unknown</td>
<td>ME</td>
<td>Liu et al. (2007)</td>
</tr>
<tr>
<td>MUM2</td>
<td></td>
<td>Seed mucilage</td>
<td>Galactosidase</td>
<td>Coding</td>
<td>Del</td>
<td>ME</td>
<td>Macquet et al. (2007)</td>
</tr>
<tr>
<td>BRX</td>
<td></td>
<td>Root growth</td>
<td>Novel TF</td>
<td>Coding</td>
<td>Nonsense Subs</td>
<td>TP</td>
<td>Mouchel et al. (2004)</td>
</tr>
<tr>
<td>LPR1</td>
<td></td>
<td>Root growth response to P</td>
<td>Multicopper oxidase</td>
<td>Promoter</td>
<td>Del</td>
<td>ME</td>
<td>Svistoonoff et al. (2007)</td>
</tr>
<tr>
<td></td>
<td>INV</td>
<td>Sugar composition, root growth</td>
<td>Invertase</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Sergeeva et al. (2006)</td>
</tr>
<tr>
<td>Small effect</td>
<td>QTL</td>
<td>Biomass accumulation</td>
<td>Protein kinase</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Kroyman and Mitchell-Olds (2005)</td>
</tr>
<tr>
<td>DM1</td>
<td></td>
<td>Hybrid incompatibility, disease resistance</td>
<td>NB-LRR proteins</td>
<td>Promoter + coding</td>
<td>Nonsense Subs, Del</td>
<td>AP</td>
<td>Bomblies et al. (2007)</td>
</tr>
<tr>
<td>QTL3</td>
<td></td>
<td>Hybrid incompatibility, disease resistance</td>
<td>NB-LRR proteins</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Alcázar et al. (2009)</td>
</tr>
<tr>
<td>LD1.1</td>
<td></td>
<td>Genetic incompatibility, embryogenesis</td>
<td>His biosynthesis enzyme</td>
<td>Promoter, coding</td>
<td>Nonsense Subs</td>
<td>ME, TP</td>
<td>Bikard et al. (2009)</td>
</tr>
<tr>
<td>LD1.5</td>
<td></td>
<td>Genetic incompatibility, embryogenesis</td>
<td>His biosynthesis enzyme</td>
<td>Promoter + coding</td>
<td>Gene Del</td>
<td>Null expression</td>
<td>TP, AP</td>
</tr>
<tr>
<td>APR2</td>
<td></td>
<td>Sulfate accumulation</td>
<td>Sulfate reductase</td>
<td>Coding</td>
<td>Nonsense Subs</td>
<td>AP</td>
<td>Loudet et al. (2007)</td>
</tr>
<tr>
<td>MOT1</td>
<td></td>
<td>Mo accumulation</td>
<td>Sulfate transporter</td>
<td>Promoter</td>
<td>Del</td>
<td>ME</td>
<td>Baxter et al. (2008)</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Plant Species</th>
<th>Locus/Gene</th>
<th>Phenotypic Effects</th>
<th>Molecular Function</th>
<th>Functional Nucleotide Polymorphism</th>
<th>Allelic Dysfunction</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gene Position</td>
<td>Mutation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMA5</td>
<td>Cu toxicity</td>
<td>Cu ATPase transporter</td>
<td>Coding</td>
<td>Missense Subs</td>
<td>AP</td>
<td>Kobayashi et al. (2008)</td>
</tr>
<tr>
<td>HKT1</td>
<td>Na accumulation</td>
<td>Na+ transporter</td>
<td>Promoter</td>
<td>Del</td>
<td>ME</td>
<td>Rus et al. (2006)</td>
</tr>
<tr>
<td>A. lyrata</td>
<td>FRI</td>
<td>Flowering</td>
<td>FRI-like</td>
<td>Coding</td>
<td>Indel</td>
<td>AP</td>
</tr>
<tr>
<td>B. rapa</td>
<td>BFLC1</td>
<td>Flowering</td>
<td>MADS TF</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>Turnip</td>
<td>BFLC2</td>
<td>Flowering</td>
<td>MADS TF</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Schranz et al. (2002)</td>
</tr>
<tr>
<td>S. lycopersicum</td>
<td>SUN</td>
<td>Fruit shape</td>
<td>IQ67 domain protein</td>
<td>Coding + promoter</td>
<td>Gene duplication</td>
<td>ME</td>
</tr>
<tr>
<td>Tomato</td>
<td>OvATE</td>
<td>Fruit shape</td>
<td>Nuclear protein</td>
<td>Coding</td>
<td>Nonsense Subs</td>
<td>AP</td>
</tr>
<tr>
<td>A. lyrata</td>
<td>FAS</td>
<td>Fruit shape</td>
<td>Yabb YY</td>
<td>Intron</td>
<td>Ins</td>
<td>ME</td>
</tr>
<tr>
<td>Tomato</td>
<td>Fw2.2</td>
<td>Fruit size</td>
<td>Cell cycle control</td>
<td>Promoter</td>
<td>Subs, indel</td>
<td>ME</td>
</tr>
<tr>
<td>Tomato</td>
<td>Sw4.1</td>
<td>Seed size</td>
<td>Transporter</td>
<td>Unknown</td>
<td>Unknown</td>
<td>ME</td>
</tr>
<tr>
<td>Tomato</td>
<td>Brx9-2-5/LIN5</td>
<td>Fruit sugar content</td>
<td>Invertase</td>
<td>Coding</td>
<td>Missense Subs</td>
<td>AP</td>
</tr>
<tr>
<td>Pisum sativum</td>
<td>Lf/PsTFL1c</td>
<td>Flowering</td>
<td>TFL1-like protein</td>
<td>Promoter</td>
<td>Gene Del</td>
<td>Null expression</td>
</tr>
<tr>
<td>Pisum sativum</td>
<td>Lycy</td>
<td>Flowering</td>
<td>TCP TF</td>
<td>Promoter + coding</td>
<td>DNA methylation</td>
<td>ME</td>
</tr>
<tr>
<td>Pisum sativum</td>
<td>T. aestivum</td>
<td>Flowering</td>
<td>AP1-like MADS TF</td>
<td>Promoter, intron</td>
<td>Del, Ins</td>
<td>ME</td>
</tr>
<tr>
<td>Pisum sativum</td>
<td>Vrn3</td>
<td>Flowering</td>
<td>FT-like protein</td>
<td>Promoter</td>
<td>TE insertion</td>
<td>ME</td>
</tr>
<tr>
<td>Pisum sativum</td>
<td>Q</td>
<td>Seed free-threshing</td>
<td>AP2-like TF</td>
<td>Coding</td>
<td>Missense Subs</td>
<td>AP</td>
</tr>
<tr>
<td>Pisum sativum</td>
<td>RhB1, RhsD1</td>
<td>Plant and leaf size</td>
<td>GAI-like protein</td>
<td>Coding</td>
<td>Nonsense Subs</td>
<td>TP</td>
</tr>
<tr>
<td>Pisum sativum</td>
<td>Gpc-B1/NAM-B1</td>
<td>Senescence; Zn, Fe, and protein content</td>
<td>NAC TF</td>
<td>Coding</td>
<td>Del, gen Del</td>
<td>ME</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>Vrn1</td>
<td>Flowering</td>
<td>AP1-like MADS TF</td>
<td>Promoter</td>
<td>Del</td>
<td>ME</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>Vrn2</td>
<td>Flowering</td>
<td>CCT domain protein</td>
<td>Promoter, coding</td>
<td>Missense Subs</td>
<td>AP</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>Vrn3</td>
<td>Flowering</td>
<td>FT-like protein</td>
<td>Promoter</td>
<td>Coding</td>
<td>ME</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>Ppd1</td>
<td>Flowering</td>
<td>Response regulator</td>
<td>Coding</td>
<td>Missense Subs</td>
<td>AP</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>Nud</td>
<td>Seed free-threshing</td>
<td>ERF TF</td>
<td>Promoter + coding</td>
<td>Gene Del</td>
<td>Null expression</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>Vrs1</td>
<td>Seed number, ear morphology</td>
<td>HD TF</td>
<td>Coding</td>
<td>Ins</td>
<td>Taketa et al. (2008)</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>Bm11</td>
<td>Boron toxicity tolerance</td>
<td>Boron efflux transporter</td>
<td>Promoter</td>
<td>Gene duplication</td>
<td>ME</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>qLTG3</td>
<td>Germination</td>
<td>Unknown</td>
<td>Coding</td>
<td>Del</td>
<td>ME</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>Rc</td>
<td>Germination, seed color</td>
<td>hHLH TF</td>
<td>Coding</td>
<td>Del</td>
<td>TP</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>Hd1</td>
<td>Flowering</td>
<td>CO-like TF</td>
<td>Coding</td>
<td>Del, TE insertion</td>
<td>TP, ME</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>H3sa</td>
<td>Flowering</td>
<td>FT-like protein</td>
<td>Unknown</td>
<td>Unknown</td>
<td>ME</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>H6</td>
<td>Flowering</td>
<td>Protein kinase CK2</td>
<td>Coding</td>
<td>Nonsense Subs</td>
<td>TP</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>Edh1</td>
<td>Flowering</td>
<td>Response regulator</td>
<td>Coding</td>
<td>Missense Subs</td>
<td>AP</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>Gh22</td>
<td>Flowering, plant height,</td>
<td>CCT domain protein</td>
<td>Coding</td>
<td>Gene Del, nonsense</td>
<td>Null expression</td>
</tr>
<tr>
<td>Triticum monoccocum</td>
<td>sh4</td>
<td>Seed yield</td>
<td>Myb TF</td>
<td>Coding</td>
<td>Missense Subs</td>
<td>AP</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Allelic Dysfunction</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>qSH1</td>
<td>Shattering HD TF Promoter Missense Subs ME Konishi et al. (2006)</td>
</tr>
<tr>
<td>qSW5</td>
<td>Seed size Unknown Coding, promoter Del ME Shomura et al. (2008)</td>
</tr>
<tr>
<td>GW2</td>
<td>Seed size RING protein Coding Del TP Song et al. (2007)</td>
</tr>
<tr>
<td></td>
<td>Seed size Unknown Coding Missense Subs TP Fan et al. (2006)</td>
</tr>
<tr>
<td>GS3</td>
<td>Seed number CK oxidase/Gn1a Promoter Del ME Ashikari et al. (2005)</td>
</tr>
<tr>
<td></td>
<td>Seed starch composition Starch biosynthesis Wx Intron Splice site Subs TP Isshiki et al. (1998)</td>
</tr>
<tr>
<td>SKC1</td>
<td>Flowering AP2-like TF Promoter enhancer TE insertion, indel ME Salvi et al. (2007)</td>
</tr>
<tr>
<td>Z. mays Vgt1</td>
<td>Inflorescence architecture TCP TF Unknown ME Doebley et al. (1997)</td>
</tr>
<tr>
<td>tga1</td>
<td>Seed morphology SBP TF Coding Missense Subs AP Wang et al. (2005)</td>
</tr>
<tr>
<td></td>
<td>Seed starch composition Starch biosynthesis Wx Coding TE insertion ME Varagona et al. (1992)</td>
</tr>
</tbody>
</table>

**TF:** Transcription factor; **TE:** Transposable element; **Subs:** substitution; **Ins:** insertion; **Del:** deletion; **TP:** truncated protein; **AP:** altered protein; **ME:** Misexpression

Response have been colocated with other *FLC* homologs in several crosses (Alonso-Blanco et al., 1998; O’Neill et al., 2008). Furthermore, loss- and gain-of-function alleles of another *FLC* regulator called *HUA2* likely contribute to this variation (Doyle et al., 2005; Wang et al., 2007).

In addition, natural variation for the flowering photoperiod response is caused by three photoreceptor genes, *CRY2, PHYC, and PHYD*, in *A. thaliana* (Table 1, Figure 1). Rare loss-of-function alleles of *PHYD* and *PHYC* as well as a gain-of-function of *CRY2* produce early flowering and reduced photoperiod response (Aukerman et al., 1997; El-Assal et al., 2001; Balasubramanian et al., 2006). Association mapping analyses suggest that haplotypes of *CRY2* and *PHYC* differing in eight amino acid substitutions might further contribute to this variation (Olsen et al., 2004; Balasubramanian et al., 2006).

QTL linkage mapping and association analyses of candidate genes also have been performed in other *Brassicaceae* species, showing that *FLC*-like genes probably contribute to the differentiation of annual and biennial habits in *Brassica rapa* (Schranz et al., 2002; Yuan et al., 2009) and that functional variants of a *FRI*-like gene of *A. lyrata* affects its vernalization requirement (Kuittinen et al., 2008). In addition, molecular analyses have identified the *LF* gene in another dicotyledonous crop species, pea, which is a long-day temperate legume that also responds to vernalization. Pea *LF* behaves as a flowering repressor that is homologous to *A. thaliana* *TFL1*, and a natural loss-of-function *LF* allele caused by a large deletion shows reduced photoperiod response (Foucher et al., 2003).

In the last few years, comparable progress has been achieved in understanding flowering time variation in several monocotyledonous grass species, including temperate cereals with similar photoperiod and vernalization responses to *A. thaliana* (Table 1). Two genes largely determine the life habit and vernalization response of barley and wheat. *VRN1* encodes a MADS box transcription factor homologous to the meristem identity protein *AP1* of *A. thaliana* (Yan et al., 2003), and *VRN2* encodes a grass-specific putative transcription factor with a CCT domain (Yan et al., 2004a). *VRN1* promotes flowering and is speculated to be directly downregulated by vernalization, whereas *VRN2* is a repressor that reduces the expression of *VRN1* mediated by *VRN3*, a feedback regulatory loop among these genes accounting for their epistatic interactions (Figure 1; Distelfeld et al., 2009a). The early flowering spring habit of wheat and barley has been associated with more than 15 independent insertion and deletions in the first regulatory intron of *VRN1* (Yan et al., 2004a; Fu et al., 2005; Szucs et al., 2007) and with several loss-of-function alleles of *VRN2* (Distelfeld et al., 2009b). In addition, natural variation for the photoperiod response and its interaction with vernalization is determined by *VRN3* and *Ppd1*. *VRN3*, which is a homolog of *A. thaliana* *FT* encoding a long-distance flowering signal, is upregulated by long-day photoperiod and vernalization and activates *VRN1* expression (Yan et al., 2006). *Ppd1* encodes a pseudoresponse regulator that mediates the photoperiod upregulation of *VRN3* (Turner et al., 2005). Dominant mutations in the promoter and first intron of wheat and barley *VRN3* show increased expression and response to vernalization (Yan et al., 2006). By contrast, missense recessive mutations in the CCT domain of barley *Ppd-H1* likely cause late flowering...
The availability of causal genes and polymorphisms enables the study of the evolutionary processes that generate and maintain flowering time variation within and among species (reviewed in...
Roux et al., 2006; Mitchell-Olds et al., 2007). In crop species, fitness effects and adaptive values of genetic variants are often documented because segregating populations are grown in their natural agricultural environments (Izawa, 2007). Detailed analyses indicate that some loci affect not only heading date but several other fitness components, as shown with the effects of Ghd7 on plant height, panicle size, and seed production (Xue et al., 2008). Geographical distributions of alleles of Hdh1, Ehd1, and Dgh7 from rice (Izawa, 2007; Xue et al., 2008) or Vgt1 and Dwarf8 from maize (Camus-Kulandaivelu et al., 2006; Ducrocq et al., 2008) show latitudinal and/or altitudinal clines, which suggests that they are involved in adaptation to different world regions. Current multilocus analyses of cultivated varieties are elucidating the role of flowering time allelic variation in regional adaptation and crop expansion during the few thousand years after domestication.

Similar studies have been initiated in A. thaliana, where in contrast with crop plants, several of the natural alleles identified in CRY2, FLM, PhyC, PhyD, or HUA2 (Table 1) are found in a single accession; therefore, it is possible that they are rare deleterious alleles. However, analyses of nucleotide diversity of CRY2, PhyC, FRI, and FLC have detected highly differentiated haplotypic groups, suggesting that their allelic variation might be maintained by various types of selection (Olsen et al., 2004; Caicedo et al., 2004; Stinchcombe et al., 2004; Balasubramanian et al., 2006). Most of the large number of FRI loss-of-function alleles appear at very low frequency, suggesting that they have been generated recently, for example, after the last glaciation (Le Corre et al., 2002; Shindo et al., 2005). However, two of these alleles are widely distributed in Central Europe, supporting the notion that their frequency increased due to strong natural selection during the postglacial colonization of this region (Johanson et al., 2000; Toomajian et al., 2006). In addition, a Eurasian latitudinal cline of flowering time and vernalization sensitivity has been shown to be partly determined by interactions among FRI, FLC, and PhyC allelic variation, suggesting that these interactions contribute to latitudinal and local adaptation (Caicedo et al., 2004; Lempe et al., 2005; Balasubramanian et al., 2006). Supporting the adaptive value of this allelic variation, analysis of FRI and FLC alleles in field experiments performed under environments closer to their natural habitat than laboratory conditions show that the FRI/FLC epistatic interaction affects different fitness components depending on the season of germination (Korves et al., 2007). In addition, antagonistic pleiotropic effects of FRI allelic variation on different fitness components, such as flowering time and plant architecture, suggest that complex phenotypic pleiotype accounts for the adaptive value of FRI alleles (Scarcelli et al., 2007). Similarly, the large pleiotropic effects described for photoreceptor genes and HUA2 (Table 1) probably also contribute to the global adaptive value of their genetic variants. Thus, several demographical and selective processes seem to determine the natural variation for flowering time in crop and wild species, but their relative significance and molecular targets are just beginning to emerge.

**Plant Architecture and Morphology**

More than 20 loci of major effect have now been cloned for several domestication traits related to plant architecture, revealing that loss of functions of transcription factor genes (sometimes previously unknown) are often involved (Table 1). Examples are the rice shattering locus sh4 encoding a Myb factor (Li et al., 2006), the rice shattering locus qSH1, and the six/two row locus Vrn1 of barley encoding homeodomain transcription factors (Konishi et al., 2006; Komatsuda et al., 2007) and a SBP gene underlying the teosinte glume architecture1 (tg1) locus in maize (Wang et al., 2005). Gain-of-function alleles of genes encoding other transcription factors also have been identified. These are illustrated with the classical maize tb1 and wheat Rht1 genes involved in plant architecture, whose selected alleles show mutations in a cis-regulatory region or in the coding region, respectively (Doebley et al., 2006; Peng et al., 1999; Clark et al., 2006). In tomato, fruit size and shape strongly differ from the small and round fruits of ancestral wild species. Detailed QTL analyses (reviewed in Tanksley, 2004) have been followed up with the cloning of four major effect loci (Table 1). Thus, Fw2.2 and Ovate identified novel classes of genes (Frary et al., 2000; Liu et al., 2002), the FAS locus has recently found to be a Yabbtype cell fate gene (Cong et al., 2008), and Sun has been shown to encode an IQ67 domain protein (Xiao et al., 2008). An ovate nonsense allele causes fruits with a pear shape (Liu et al., 2002), while different regulatory mutations producing misexpression of Fw2.2, FAS, and SUN lead to large fruit phenotypes. Furthermore, seed size loci have been isolated in several species, such as the rice Gw2 and Gs3 loci encoding an ubiquitin-E3-ligase and a transmembrane protein, respectively (Fan et al., 2006; Song et al., 2007), or the tomato Sw4.1 locus encoding an ABC transporter (Orsi and Tanksley, 2009). On the other hand, a large number of QTL analyses have been described in many crops for other developmental and morphological traits related to yield and plant performance (e.g., http://www.gramene.org/ and http://www.sgn.cornell.edu/index.pl). The first genes underlying those loci have been identified mainly in rice (Sakamoto and Matsuoka, 2008; Weng et al., 2008) as illustrated with the Gln1a locus encoding a cytokinin dehydrogenase that affects grain number (Ashikari et al., 2005), or the pleiotropic Ghd7 locus with effects on various yield components (Xue et al., 2008).

Developmental traits related to plant architecture and organ morphology also have been studied in A. thaliana, where natural variation for seed size (Alonso-Blanco et al., 1999), inflorescence architecture (Juenger et al., 2000), leaf morphology (Perez-Perez et al., 2002), and flower size (Juenger et al., 2005) have been described. Although none of the underlying genes have been cloned, microsynteny comparisons between tomato and A. thaliana, combined with functional analyses, suggest that a Sw4.1 ortholog gene might underlie an A. thaliana seed size QTL (Orsi and Tanksley, 2009). Several QTLs have been detected for root growth and architecture (Mouchel et al., 2004; Loudet et al., 2005; Fitz Gerald et al., 2006), and the major effect locus BrrX has identified a novel gene with unknown molecular function (Mouchel et al., 2004). In addition, developmental features present as monogenic traits in some A. thaliana accessions have identified non-sense mutations in the Myb transcription factor gene GL1 causing glabrous phenotype (Hauser et al., 2001) and in the β-galactosidase gene corresponding to the MUM2 locus, which reduces seed mucilage liberation after imbibition (Macquet et al., 2007).
Different strategies are revealing cryptic developmental genetic variation of *A. thaliana* that is only apparent under specific conditions (Table 1). It has been shown that differential expansion of an intronic trinucleotide repeat affects the expression of a Leu biosynthesis enzyme gene, hence causing aberrant leaf morphology defects that are mainly visible under short-day photoperiod and high temperature conditions (Sureshkumar et al., 2009). In addition, *A. thaliana* transgenic lines carrying the SRK and SCR genes that determine self-incompatibility in *A. lyrata* have been used to reveal a gene underlying pseudo-self-compatibility. Thus, it has been shown that polymorphisms in the promoter region of the previously uncharacterized gene PLANT U-BOX8 regulating SRK transcript levels reduce its expression and cause such compatibility (Liu et al., 2007). These examples show how the analysis of morphological and developmental traits in crops and wild plants are identifying not only genes with allelic variants increasing fitness under particular conditions, but also cryptic genetic variation with unknown effects on plant adaptation.

VEGETATIVE GROWTH AND PHYSIOLOGY

Natural variation has been described extensively for general growth traits (e.g., biomass accumulation) as well as primary metabolism and mineral uptake and accumulation, and the complex molecular bases underlying these traits are beginning to be elucidated. Such studies are aiding our understanding of hybrid vigor (positive heterosis) and hybrid incompatibility (negative heterosis), which are of primary concern in agronomy, as well as specific metabolic pathways and physiological processes that affect crop growth and/or crop quality.

**Growth and Biomass Accumulation**

The complexity underlying the regulation of biomass accumulation is illustrated with the precise dissection of a small-effect QTL for this trait in *A. thaliana*, which includes two closely linked QTLs that have opposing effects on biomass accumulation within a small genomic region of 210 kb (Kroymann and Mitchell-Olds, 2005). In addition, both loci depend on the genetic background. Thus, a highly polygenic architecture involving genetic interactions has been speculated for biomass accumulation (Kroymann and Mitchell-Olds, 2005). Recently, significant progress has been made in understanding the genetic and molecular bases of growth traits that show positive or negative heterosis, as well as in finding genes contributing to the variation for growth interactions with abiotic environmental factors.

A large number of QTL studies are currently dissecting the genetic bases of positive heterosis or hybrid vigor, which is a growth trait observed in many hybrids of most species and a major plant breeding objective. Numerous QTLs with different levels of dominant, overdominant, and epistatic effects have been mapped for heterosis in maize (Frascaroli et al., 2007), rice (Li et al., 2001), tomato (Semel et al., 2006), rapeseed (*Brassica napus*) (Radoev et al., 2008), and *A. thaliana* (Kusterer et al., 2007; Melchinger et al., 2007). Comparisons of genomic sequences among maize inbred lines have shown extensive within-species loss of gene colinearity, with a large portion of genes being absent in some individuals (Fu and Dooner, 2002; Song and Messing, 2003). This lack of colinearity has led to a molecular hypothesis for the dominance model of heterosis and its counterpart inbreeding depression, where the presence of distinct genes in different lines might cause hybrid vigor (Fu and Dooner, 2002). In addition, overdominance frequently has been detected at the level of gene expression in hybrids of maize and *A. thaliana* (Song and Messing, 2003; Vuylsteke et al., 2005).

On the other hand, analyses of *A. thaliana* hybrids are identifying the molecular bases of a related fundamental evolutionary phenomenon, namely, negative heterosis or hybrid incompatibility (Table 1). Nearly 2% of *A. thaliana* intraspecific crosses yield hybrid plants expressing severe growth defects and leaf necrosis, which are due to epistatic interactions between two or three genes fitting the Dobzhansky-Muller model of hybrid postzygotic isolation (Bombles et al., 2007; Alcázar et al., 2009). Alleles of two different tandem repeats of NB-LRR disease resistance genes differing in sequence and gene copy number have been shown to cause these hybrid incompatibilities (Bombles et al., 2007; Alcázar et al., 2009). Incompatible combinations are characterized by autoimmune-like responses and enhanced disease resistance, which depend on growth temperature and activation of the salicylic acid stress signaling pathway (Alcázar et al., 2009). In addition, embryo lethality segregating in crosses among wild *A. thaliana* genotypes is determined by an epistatic interaction between two paralogous genes encoding a key enzyme involved in His biosynthesis (Bikard et al., 2009). Thus, divergent evolution of particular duplicated genes appears as the main mechanism causing genetic incompatibilities at different postzygotic levels.

Several QTLs have been isolated that contribute to natural variation for growth responses to light or soil nutrient availability in *A. thaliana*. Analyses of light responses have shown that amino acid substitutions in photoreceptors, including phytochrome A (Malloof et al., 2001), phytochrome B (Filiault et al., 2008), cryptochrome 2 (Botto et al., 2003), and probably phytochrome C (Balsubramanian et al., 2006), affect hypocotyl elongation in different light conditions. In addition, a tandem zinc knuckle/PLU3 (TZP/LIGHT5) domain-encoding gene has been shown to promote growth, associated with an insertion mutation appearing as a rare natural loss-of-function allele (Loudet et al., 2008). Furthermore, it has been shown that natural polymorphisms affecting the expression of *LPR1*, which encodes a multicopper oxidase, strongly influences the primary root growth response to phosphate starvation (Svistoonoff et al., 2007). Hence, it appears that mainly signal perception and transduction components have been involved in natural variation for growth environmental responses.

**Primary Metabolism**

Primary metabolism has been defined as “those essential reactions involving compounds that are formed as part of the normal anabolic and catabolic processes, which result in assimilation, respiration, transport, and differentiation processes that take place in most, if not all, cells of an organism” (Fernie and Schauer, 2009). Besides a wide range of intermediate compounds (e.g.,
compounds from glycolysis or Krebs cycle), primary metabolites include end products of metabolic pathways that accumulate in (harvestable) sink organs (e.g., seeds, fruits, and tubers) determining relevant crop quality traits related to nutritional content and composition. Recent molecular analyses of natural variation have identified a large number of genes encoding enzymes and have associated primary metabolism with phenotypic variation for morphology and growth-related traits (Liseć et al., 2008; Schauer et al., 2008).

Two decades ago, analyses of starch content in seeds of pea and cereals identified the genes underlying classical loci, such as Mendel’s pea *rugosus* (*R*) and maize *Waxy*. These pioneering works illustrate the pleiotropic effects of primary metabolism on plant development. The *R* locus encodes a starch branching enzyme, whose recessive allele carrying a transposon insertion in the coding sequence has been widely spread in European cultivars due to the associated sweetness of the wrinkled pea phenotype (Bhattacharyya et al., 1990). Maize *Waxy* loci encode granule-bound starch biosynthesis enzymes, where multiple loss-of-function alleles caused by transposon insertions lead to a waxy appearance of kernels that reflect the lack of amylose in endosperm (Varagona et al., 1992). Moreover, a single splice donor site mutation in a homologous *Waxy* gene is responsible for the absence of amylose in glutinous rice varieties, which has been widely spread in East Asia associated with human cultural behaviors (Olson and Purugganan, 2002).

Recent quantitative analyses of natural variation for the content of primary metabolites in fruits of tomato interspecific crosses (Schauer et al., 2006) and in potato tubers (*Solanum tuberosum*; Menéndez et al., 2002) have found a large number of QTLs underlying the content of other carbohydrate metabolites. Isolation of a large-effect QTL for tomato sugar content identified the invertase encoding gene *LIN5*, where a single amino acid substitution alters its activity (Fridman et al., 2004). Based on colocation and genetic association, two tandemly repeated invertase genes, *invGE* and *invGF*, have been also suggested to underlie a cold-sweetening QTL in potato (Li et al., 2005).

Several studies have dissected the variation for primary metabolism of *A. thaliana* through the analysis of enzymatic activities and metabolite contents in different vegetative organs. QTLs have been found for the activity of multiple carbohydrate metabolism enzymes, some of them colocalizing with the corresponding structural genes, for example, phosphogluco-mutase (*PGM*) and hexokinase, which indicates cis-regulatory variation (Mitchell-Olds and Pedersen, 1998; Keurentjes et al., 2008b). Histochemical analyses further shows that certain *PGM* QTLs are tissue specific (Sergeeva et al., 2004). However, several enzymatic activity QTLs are located in different genomic regions, identifying trans-acting sequences. These loci might simultaneously coregulate several enzymes, hence revealing novel modes of carbohydrate metabolic regulation (Mitchell-Olds and Pedersen, 1998; Sergeeva et al., 2004; Keurentjes et al., 2008b). In addition, a vacuolar invertase has been found that contributes specifically to the natural variation for root elongation but not for hypocotyl length (Sergeeva et al., 2006). Similarly, QTLs have been identified for nitrogen- and sulfur-containing compounds, such as nitrate, sulfate, and free amino acids (Loudet et al., 2003, 2007). A large-effect QTL has been isolated for sulfate content, which corresponds to the APR2 sulfate reductase gene (Loudet et al., 2007). *APR2* activity is altered by a single amino acid substitution, and the effects of this polymorphism interact with nitrogen availability, hence illustrating the interaction between sulfate assimilation and nitrogen metabolism. Furthermore, as described in the previous section, genes involved in Leu or His biosynthesis account for natural variation for severe zygotic (embryo) and postzygotic morphological defects (Bikard et al., 2009; Sureshkumar et al., 2009).

*A. thaliana* natural variation for untargeted metabololites from vegetative tissues has been dissected by quantifying high-throughput metabolic profiles in different mapping populations (Keurentjes et al., 2006; Meyer et al., 2007; Rowe et al., 2008). A large number of metabolite content QTLs (mQTLs) have been mapped, demonstrating that two- and three-way epistatic interactions determine a substantial proportion of this variation (Liseć et al., 2008; Rowe et al., 2008). Candidate genes have been proposed for a large proportion of detected mQTL based on the enzymatic activities involved in different biochemical pathways (Liseć et al., 2008). In addition, comparative analyses have shown colocation of QTLs for biomass accumulation and mQTL, supporting a link between growth and metabolic profiles (Liseć et al., 2008).

**Mineral Accumulation**

Most nutrients that plants need for growth and development are supplied as minerals to the roots, and they are classified as macronutrients (Ca, K, Mg, N, P, and S) or micronutrients (B, Fe, Mn, Co, Cu, Mo, Ni, and Zn) depending on the necessary quantities. The composition of mineral nutrients and trace elements (i.e., the inorganic component of an organism) is now referred to as the ionome (Salt et al., 2008). There is substantial natural variation for mineral use efficiency, root uptake, translocation from roots to shoots, and accumulation in the seed as storage and supply for the germinating seedling. This variation has been reported in many species, leading to breeding programs such as those aiming to improve zinc and iron status of cereal grains or tuber crops (www.harvestplus.org). Recent QTL analyses have identified six genes involved in the complex mechanisms controlling the variation for plant nutrition and mineral homeostasis (Table 1; Ghaniyarian et al., 2006).

QTL studies in crop species have focused on seed or leaf mineral concentrations, mostly in rice, wheat, barley, *Brassica* species, and soybean (*Glycine max*). In cereals, mostly rice, QTLs have been mapped for grain minerals (*Joppa* et al., 1997; Garcia-Oliveira et al., 2009), flag leaf nitrogen (*Ishimaru* et al., 2001) or phosphorous (*P*) uptake (*Wissuwa* et al., 2002), salt tolerance related to Na/K homeostasis (*Lin* et al., 2004), and grain accumulation of toxic Cd (*Ueno* et al., 2009). In *Brassica* crops, loci have been reported for phosphate concentrations in leaves and seeds of *B. rapa* (*Zhao* et al., 2008), for Ca and Mg concentrations in leaves of *B. oleracea* (*Broadley* et al., 2008), and for leaf mineral concentrations in *B. rapa* (*Wu* et al., 2008). In soybean, QTLs have been found for Ca and S concentrations in seeds (*Panthee* et al., 2006; *Zhang* et al., 2009). Three large-effect QTLs have been isolated in crops, which have identified several molecular mechanisms linking mineral content and other
physiological traits. The rice SKC gene encodes a Na\textsuperscript{+} selective transporter, whose more active allele determined by missense mutations increases salt tolerance (Ren et al., 2005). A boron tolerance locus of barley corresponds to the boron efflux transporter gene Bot1 that appears duplicated and overexpressed in a B tolerant accession (Sutton et al., 2007). In addition, the Gpc-B1/NAM-B1 locus of wheat encodes a NAC transcription factor that regulates nutrient remobilization from leaves to developing grains (Uauy et al., 2006). It has been shown that nonfunctional alleles caused by different deletions are present in cultivated varieties, leading to delayed senescence and reduced protein, Zn, and Fe grain content.

Detailed analyses of the ionome in A. thaliana have shown considerable variation for leaf mineral concentrations under various mineral/metal supply conditions (Salt et al., 2008). QTLs have been identified for accumulation of different elements (Ca, Cu, K, Mg, Mn, Na, P, S, and Zn) in seeds, siliques, leaves, and roots under different growth conditions (Vreugdenhil et al., 2004; Waters and Grusak, 2008; Ghandilyan et al., 2009). Overall, these analyses show that mineral concentrations are controlled by a large number of loci with mostly moderate effect, which often interact strongly with growth environment. Interestingly, a strong P QTL has been colocated with a previously identified locus for phytate (myo-inositol-1,2,3,4,5,6-hexakisphosphate or IP6), the storage compound for P in seeds (Bentsink et al., 2003). However, only the ERECTA gene, encoding a receptor-like kinase for which a mutant allele segregates in mapping populations, has been demonstrated to affect mineral concentrations in these studies (Waters and Grusak, 2008; Ghandilyan et al., 2009).

In addition, A. thaliana QTL analyses have been focused on accumulation of specific minerals, including N (as nitrate; Loudet et al., 2003; Harada et al., 2004), K (Harada and Leigh, 2006), Cu (Kobayashi et al., 2008), Mo (Baxter et al., 2008), and Na (Rus et al., 2006). These studies have led to the isolation of three genes underlying large-effect QTLs, which also encode different mineral transport components. A root copper tolerance locus corresponds to the HMA5 gene encoding a Cu-transporting ATPase. Several natural alleles differing in missense mutations in conserved motifs show lower activity and Cu translocation to the shoot (Kobayashi et al., 2008). A mitochondrial molybdenum transporter encoded by the nuclear MOT1 gene underlies a shoot Mo concentration QTL. A deletion in MOT1 promoter region has been associated with low expression and low shoot Mo concentration, suggesting that this regulatory mutation is the causal nucleotide polymorphism (Baxter et al., 2008). Finally, HKT1 encodes a Na\textsuperscript{+} transporter for which two loss-of-function alleles associated with promoter deletions produce lower root expression and enhanced shoot Na\textsuperscript{+} levels in two coastal accessions (Rus et al., 2006).

Studies of A. thaliana mineral accumulation have inspired research on natural variation for exceptional zinc and/or cadmium accumulation in leaves of the related metal hyperaccumulator species Arabidopsis halleri and Thlaspi caerulescens. Several loci have been mapped in populations derived from crosses between A. halleri and A. lyrata (Courbot et al., 2007), one of them colocaling with HMA4, a locus containing three copies of the Ah-HMA4 gene that is essential for Zn hyperaccumulation and Cd tolerance in A. halleri (Hanikenne et al., 2008). In addition, multiple QTLs have been located for root or shoot Zn or Cd accumulation in F3 populations of T. caerulescens (Assunção et al., 2006; Deniau et al., 2006). Thus, comparative analyses with A. thaliana are not only assisting the isolation of genes underlying those loci but will also enable identification of the molecular bases of interspecific variation present in wild crucifers.

**REVEALING THE VARIOUS LEVELS OF PHENOTYPIC REGULATION: NATURAL VARIATION FOR MOLECULAR TRAITS**

The path from genotype to phenotype includes several molecular intermediate steps in the translation of genetic information, with regulation occurring at each of these levels (Keurentjes et al., 2008a). Extensive natural variation has been described for the amount of molecular components, such as gene transcripts (Vuylstekte et al., 2005; Kliebenstein et al., 2006a), protein abundance and activity (Chevalier et al., 2004; Cross et al., 2006), and metabolites (Kliebenstein et al., 2001; Cross et al., 2006). Quantitative genetic analyses of “-omics” data collected from mapping populations by so-called genetical genomics approaches (Jansen and Nap, 2001) have shown that much of this natural variation is genetically determined by specific expression QTL (eQTLs), mQTL, or protein QTL (pQTL) (Keurentjes et al., 2008a). In addition, a portion of this variation is determined by epigenetic mechanisms of regulation, such as methylation and chromatin remodeling (He et al., 2004; Vaughn et al., 2007), which in turn can also result from DNA sequence polymorphisms (Johannes et al., 2008).

The genetic bases of natural variation for genotype-wide gene expression have been approached systematically in A. thaliana, barley, and maize (Schadt et al., 2003; Vuylstekte et al., 2006; Keurentjes et al., 2007; West et al., 2007; Potokina et al., 2008b). Expression variation of ~30 to 40% of the segregating genes is regulated by an eQTL colocating with the gene encoding the mRNA, most of those local eQTLs being the result of cis-regulated genes (i.e., allele-specific expression). However, most variability expressed genes are regulated in trans by eQTLs located in distant genomic regions (trans-eQTLs) that often appear concentrated in a small number of regulatory hot spots (Keurentjes et al. 2007; West et al., 2007; Fu et al., 2009). cis-eQTLs show larger average effects than trans-eQTLs probably due to the stronger impact of direct cis-regulation on the target gene. In addition, it has been shown that QTLs for gene expression variation might depend on spatial and temporal regulation (Potokina et al., 2008a). The molecular identification of regulatory eQTLs is currently facilitated by statistical methods that combine the functional annotation of biological relevant gene sets, the genomic position of genes and eQTLs, and the coexpression of candidate regulators and target genes. In this way, complex genetic networks are constructed, which reveal many novel putative regulatory steps together with strong interconnectivity and epistasis (Kliebenstein et al., 2006b; Keurentjes et al., 2007; Druka et al., 2008). As described above, large-scale untargeted analyses of metabolic profiles have also revealed extensive
genetic variation for metabolite content in *A. thaliana* and tomato (Keurentjes et al., 2006; Schauer et al., 2006; Lisec et al., 2008). Large numbers of mQTLs have been mapped, although they show a lower average heritability than eQTL (Rowe et al., 2008).

The genetic dissection of natural variation for developmental and physiological traits is beginning to link phenotypic QTLs for complex traits with eQTL, mQTL, and pQTL (Keurentjes, 2009). Several studies, mostly in *A. thaliana*, have integrated information from different levels of molecular regulation to improve the elucidation of genetic control of quantitative traits (Hirai et al., 2004; Tohge et al., 2005; Meyer et al., 2007; Wentzell et al., 2007; Druka et al., 2008; Lisec et al., 2008). Thus, it has been shown that plant growth and development are interwoven with primary metabolism, regulation occurring at different levels, including gene expression, enzyme activity, and metabolite abundance (Keurentjes et al., 2008b). Although these studies often reveal direct relationships between the various levels of biological information, such a relationship is not always linear or present. This was illustrated recently with a system-wide analysis, wherein much of the natural genetic variation affecting gene expression or functional protein differences did not appear to lead to phenotypic variation, a phenomenon known as phenotypic buffering (Fu et al., 2009). Interestingly, several loci have been identified as contributing to natural variation for the ability to buffer developmental processes through environmental canalization (Hall et al., 2007; Sangster et al., 2008). Hence, integrative analyses of multiple phenotypic and molecular regulatory levels are elucidating the mechanisms controlling developmental buffering and canalization, which are presumed to contribute to biological robustness of plant species and facilitate evolvability in changing environments (Queitsch et al., 2002; Kitano, 2004).

**MOLECULAR TARGETS OF SELECTION FOR PLANT ADAPTATION: CURRENT AND FUTURE PROSPECTS**

Nearly 100 genes underlying natural variation for plant development and growth have been identified in crop plants and *A. thaliana*. Most of them correspond to QTLs with large effects on trait variation and provide a first inventory of mechanisms, genes, and functional nucleotide polymorphisms that might be involved in plant adaptation to different agricultural or natural environments. They belong to all types of ontological classes, including transcription factors, signal transduction components, enzymes of primary and secondary metabolism, hormone metabolism and signaling, metal transporters, as well as genes with unknown functions. Interestingly, some of these genes had not been found previously in mutant screenins because common laboratory strains carry natural loss-of-function alleles, such as *FLC, FRI*, or *DOG1* in *A. thaliana*.

Nucleotide polymorphisms affecting the function of these genes are identifying the mutational mechanisms that generate natural variation. A large proportion of natural alleles carry null loss-of-function mutations, which are often produced by structural nonsense or insertion and deletion (indel) mutations (Alonso-Blanco et al., 2005). A particular class of these alleles corresponds to complete gene deletions or duplications caused by retrotransposition, recombination, and other molecular mechanisms, which may play a major role in heterosis of crop and wild species (Fu and Dooner, 2002; Song and Messing, 2003). A second type of common allele is a change-of-function allele produced by a missense mutation altering protein structure and function, such as described for photoreceptor genes (Table 1). In addition, multiple change-of-function alleles carrying regulatory mutations in three types of genomic regions affecting gene expression have been described for many traits and species. Most of these alleles show indels in their nearby promoter regions, in agreement with the large number of cis-eQTLs identified (Keurentjes et al., 2007; West et al., 2007). Second, gene expression changes are caused by unknown mutations located in distant 5'-regions of the genes as illustrated for maize *Tb1* (Clark et al., 2006) and *Vgt1* (Salvi et al., 2007). Third, expression alterations also appear to be produced by intronic mutations, including insertions of transposons or differential expansion of microsatellite repeated sequences (Sureshkumar et al., 2009). Finally, although most natural alleles that have been described are caused by variation at the level of nucleotide sequence, natural epigenetic variation also contributes to heritable developmental and physiological variation, as indicated by stably inherited epialleles in the *LCyc* gene involved in *Linaria* flower morphology (Cubas et al., 1999) and in an SBP-box gene affecting tomato fruit ripening (Manning et al., 2006). This is supported by the considerable amount of natural variation found for DNA methylation (Riddle and Richards, 2005; Zhang et al., 2006b; Vaughn et al., 2007), which is partly mediated by small interfering RNAs also differing among accessions (Zhai et al., 2008). It is hypothesized that methylation and heterochromatinization mediated by different kinds of small interfering RNAs might cis- and trans-regulate gene expression and function in a more quantitative and reversible manner than sequence variation (Zhai et al., 2008).

All of these studies provide information on the particular classes of genes and alleles that might have been naturally or artificially selected for different traits (reviewed in Alonso-Blanco et al., 2005). For instance, as described by Purugganan and Fuller (2009), it seems that domestication is mainly associated with transcriptional activators and changes in transcriptional networks. By contrast, crop diversification after domestication involves a large proportion of genes encoding enzymes, with loss-of-function alleles differently contributing to crop domestication and diversification. Thus, we are beginning to understand how natural variation is generated and maintained depending on plant species, traits, environments, and selective forces. However, conclusive generalizations on evolutionary processes await systematic extensions of the following aspects of these analyses: (1) the number of species analyzed, particularly of wild plants; (2) the number of genes identified in each species; (3) the spectrum of allelic effects per gene; and (4) the estimates of genetic and environmental interactions.

Understanding the ecological and evolutionary significance of natural variation requires precise evaluation of their role in adaptation to particular environments, which is just beginning to be approached. Based on their presence in a single wild genotype and their relative large effects, some of the *A. thaliana* natural alleles are probably deleterious variants segregating in natural populations. This has been suggested for loss-of-function
alleles of HUA2 and LIGHT5, found only in a subset of lines derived from the same parental accession (Doyle et al., 2005; Loudet et al., 2008). Similarly, the BRX loss-of-function allele has not been found in genotypes collected in the same geographical site where this variant was originally sampled many years before (Shindo et al., 2008). As described above for flowering time, demonstrating that natural variation in specific genes is involved in adaptation is currently addressed by combining population and evolutionary genetic analyses with detailed phenotypic studies of fitness-related traits. Analysis of the amount and pattern of nucleotide diversity may show potential signatures of selection, hence suggesting different evolutionary mechanisms to maintain such variation. This is illustrated with the identification of selective sweeps in several domestication genes of crop plants (Purugganan and Fuller, 2009) and with the highly differentiated haplotypes found in some A. thaliana photoreceptor genes (Olsen et al., 2004; Balasubramanian et al., 2006). However, demographical history also affects the patterns of DNA sequence diversity; therefore, other studies are necessary to test adaptive effects of such alleles. On one hand, detailed phenotypic characterizations under natural and agricultural conditions are detecting fitness effects of particular alleles or genotypes (Tian et al., 2003; Shindo et al., 2008; Wilczek et al., 2009). In addition, comparisons of differentiation among local populations for quantitative traits (QST) and nucleotide polymorphisms (FST) might further assist the detection of selection signatures on particular genes (Le Corre, 2005; Stenoien et al., 2005). On the other hand, association analyses of nucleotide and phenotypic diversity with environmental variation might identify environmental factors driving phenotypic selection. It is expected that new extensive collections of A. thaliana wild genotypes recently developed from several world regions (Jorgensen and Mauricio, 2004; Stenoien et al., 2005; Schmid et al., 2006; Beck et al., 2008; Picó et al., 2008) will facilitate estimations of allele frequencies, geographical distributions, and genetic associations with environmental factors.

In the past decade, the analysis of natural variation in crop plants and A. thaliana has provided an unprecedented amount of information on the genetic and molecular mechanisms that determine intraspecific variation and adaptation. It can be anticipated that this trend will continue in the next decade, especially with the broad implementation of “-omics” technologies for the precise analysis of natural variation at different levels. First, high-throughput genotyping will enable the development of new genetic mapping strategies, such as genome-wide association (Azanza et al. 2005; Nordborg and Weigel, 2008; Waugh et al., 2009), which is expected to provide higher mapping resolution. Second, comparisons of thousands of whole-genome sequences from genotypes of the same species will provide a dynamic perspective of a species’ genome in terms of pan-genomes (Morgante et al., 2007). Third, common integration of QTLs and genes causing phenotypic variation with those affecting the various levels of molecular variation will provide complete (genome-wide and whole organisms) phenotypic regulatory networks. The latter should include not only eQTL, mQTL, and pQTL estimated at particular points but also molecular fluxes in those networks. Finally, extension of the analysis of natural variation to other crop and wild species will allow deep comparisons among species. Such comparative intraspecific and interspecific studies will elucidate the mechanisms involved in the evolution of plant development and physiology at higher levels, including those of speciation and species diversification in relation to earth changing environments.

ACKNOWLEDGMENTS

We apologize to those authors whose work could not be discussed due to space limitations. Research in our laboratories was supported by the European Research Area-Plant Genomics program Grant 03482 ARA-BRAS and the Max Planck Society (M.R. and M.K.), Grants GEN2006-27786-E/VEG and BIO2007-62632 from the Ministerio de Ciencia e Innovacion de Spain (C.A.-B.), the Netherlands Organization for Scientific Research, VENI scheme (JK and LB), and the Centre for Biosystems Genomics (Netherlands Genomics Initiative; J.J.B.K. and D.V.).

Received April 21, 2009; revised June 10, 2009; accepted June 11, 2009; published July 2, 2009.

REFERENCES


Dubcovsky, J., and Dvorak, J. (2007). Genome plasticity a key factor in...


What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation?
Carlos Alonso-Blanco, Mark G.M. Aarts, Leonie Bentsink, Joost J.B. Keurentjes, Matthieu Reymond,
Dick Vreugdenhil and Maarten Koornneef
Plant Cell 2009;21;1877-1896; originally published online July 2, 2009;
DOI 10.1105/tpc.109.068114

This information is current as of October 5, 2017

| References | This article cites 240 articles, 90 of which can be accessed free at: /content/21/7/1877.full.html#ref-list-1 |
| eTOCs | Sign up for eTOCs at: http://www.plantcell.org/cgi/alerts/ctmain |
| CiteTrack Alerts | Sign up for CiteTrack Alerts at: http://www.plantcell.org/cgi/alerts/ctmain |
| Subscription Information | Subscription Information for The Plant Cell and Plant Physiology is available at: http://www.aspb.org/publications/subscriptions.cfm |