17

Natural variation among accessions of *Arabidopsis thaliana*: beyond the flowering date, what morphological traits are relevant to study adaptation?

X. Reboud, V. Le Corre, N. Scarcelli, F. Roux, J.L. David, T. Bataillon, C. Camilleri, D. Brunel, and H. McKhann

Abstract: A set of 240 wild type accessions of *Arabidopsis thaliana* were cultivated in two contrasting conditions, with or without application of a cold treatment prior to cotyledon emergence. In each condition, 20 primary and 3 derived variables were measured that describe the phenology and morphology of each plant. The cold pre-treatment greatly modified the time to flowering but diversely affected accessions (according to their need for vernalization).

Three criteria were then used to identify a minimum subset within the 23 variables to be measured under a time-saving constraint. We combined a criterion of highest genetic heritability so that environmental experimental variance is lowered, a criterion of highest Spearman coefficient so that the rank of any accessions remained stable across treatments and, last, a criterion of highest contribution of the variable to the seed production as estimated in a multilinear regression analysis. Applying this 'three criteria' procedure lead us finally to propose a minimum set of five variables: flowering precocity, maximum plant height, height to first flower, number of flowering heads and mean distance between siliques as best describing variation in life history traits expressed by the whole collection. We believe that these variables definitely affect the ability of *Arabidopsis* to adjust its life cycle to ecological conditions including the intensity of the interspecific competition prevailing in the environment. The core collection of 24 *Arabidopsis* accessions that was chosen for maximizing molecular diversity was confirmed here to also maximize most of each trait's variability. Some linkage disequilibrium expressed at the whole genome level when considering accessions from very diverse origin is probably responsible for the conservation of such a high diversity. These 24 accessions could thus provide an important resource for natural variation to be exploited in the identification of quantitative trait loci (QTL), in genotype/phenotype association studies or exploration of ecological and evolutionary relations.

Introduction

The collections maintained in stock centers often consist of a very large number of accessions making them hard to handle and sometimes even to exploit. Strategies to construct core collections have been designed to tackle this problem so that a restricted set of accessions is available that encompasses the range of diversity of the full collection. This strategy has been applied successfully to *Arabidopsis thaliana* by surveying genetic diversity using SNPs present at 10 fragments spread throughout the genome (two per chromosome) (McKhann et al. 2004). The resulting core collection consists of 24 accessions selected from a larger collection of 265 wild type *Arabidopsis* from diverse origins.

On the other hand, if we wish to characterize a large number of accessions in a comparative study, it would be advantageous to reduce the time spent on measuring each accession. This is clearly another optimization problem. Of all the morphological and phenological traits that could be measured, which ones best describe the variation between accessions? Of course, we do not expect a unique list of traits, as each specific study will require the precise analysis of a relevant set of characteristics. However, some traits are more sensitive to environmental noise than others and thus more difficult to analyze experimentally. Some traits are also highly correlated so that most of the information collected by measuring one trait can be used to infer the values of the

X. Reboud, V. Le Corre, and F. Roux. UMR 1210 'Biologie et Gestion des Adventices', INRA, B.P. 86510, 21065 Dijon cedex, France.
C. Camilleri, D. Brunel, and H. McKhann. Centre National de Recherche en Génomique, 2 rue Gaston Crémieux, 91057 Evry cedex, France et Station de Génétique et Amélioration des Plantes, Route de Saint-Cyr, INRA, 78026 Versailles cedex, France.
N. Scarcelli. IRD Montpellier, UMR 1097 — Equipe DYNADIV, 911, avenue Agropolis, BP 64501, 34394 Montpellier cedex 5, France.
J.L. David and T. Bataillon. UMR 1097 Diversité et Génomes des Plantes Cultivées, INRA, Domaine de Melgueil, 34130 Mauguio, France.

other. Moreover, some variables have more impact on the final morphology, phenology and fitness of accessions so that natural selection probably acted more intensely on these characters. For all these reasons, there will be sets of variables that are more informative than others. When the information on the precise ecological conditions prevailing in the area of origin of the samples is not available or when very diverse accessions are to be compared in a limited number of experiments under necessarily standardized conditions, what are the most appropriate set of variables to measure?

The sampling effort for *Arabidopsis* has been far from uniform over the species' geographic range, some areas providing many more accessions than others, and therefore, it does not reflect an underlying variation in species abundance (Alonso-Blanco and Koornneef 2000). The ecological characterization of the sampling origin also remains astonishingly scarce for most accessions and only few studies have provided some phenotypic characterization of *Arabidopsis* plants collected from different geographical origins. The traits most often analyzed are those affecting flowering time. The variation of this character is presumed to be of ecological significance for adaptation and highly variable in *Arabidopsis* (list of references at: http://www.dpw.wau.nl/natural/resources/literature.htm).

To respond or not to cold treatment: that is the question…

Variation in flowering responses after cold treatment has been the object of intense analysis. At the two extremes are plants that require a cold treatment to induce flowering and those that do not respond to cold. These two flowering "strategies" are thought to be adaptive. For the first group, cold sensitivity is considered as a means of synchronizing vegetative plant development during autumn and mild winter periods, with flowering occurring in early spring after a cold winter. With the ability to tolerate frost, these accessions remain at the rosette stage during the autumn when they set reserves to ensure the earliest possible resumption of growth before interspecific competition begins in the spring. Their ecological niche would thus be a very narrow window for rapid development and seed setting before taller and stronger plants species develop. Consequently, such accessions may suffer the fitness cost of being unable to complete a second generation within the same year even under favorable conditions because of lack of the required conditions to induce flowering.

The second group, being rather insensitive to cold treatment, is able to flower as soon as plants have reached a given biomass or accumulated day $\times$ degree threshold. In ecological niches where the favorable growing period is short, these cold-insensitive plants will also only reproduce once a year; whereas in optimal conditions they may complete more than one cycle per year and then outperform the genotypes requiring cold to induce flowering. There is however a second indirect expectation that these cold-insensitive genotypes could have only been selected under environmental conditions where virtually no local competition exists. Hence, flowering time, which has been shown to be under complex genetic determinism (Sheldon et al. 2000), has probably evolved under diversifying selection (Le Corre et al. 2002).

Other forces may act on the optimum time for flowering. Physiological constraints dictate that the smaller the plant, the smaller its pollen and seed output. In a rosette species such as *Arabidopsis thaliana* where bolting puts an end to further resource accumulation (there will be no more leaves), flowering is expected to be positively correlated with total seed production. Thus, there are probably opposing forces acting on optimal flowering period in any particular condition: completing more than one cycle and/or avoiding competition requires short generation times on one hand, while total yield requires bigger plants produced with long generation times on another hand. Indeed, the fact that the flowering trait is complex can be viewed as the consequence of the necessity to locally adjust the flowering time cycle in any possible way. Different *Arabidopsis* populations experiencing selection for early flowering would thus differ in their evolutionary response to the trade-off between total productivity and flowering time. This suggests that other genetically determined traits may be involved in ecological adaptation. The identification and study of these traits would allow us to better understand trade-offs between flowering time and other life history traits. Using both vernalized and non-vernalized treatments on a set of *Arabidopsis* accessions enables the identification of a set of traits having a high impact on fitness over a range of environmental conditions. We have therefore explored variation in morphological traits describing several aspects of inflorescence architecture among 240 accessions of *Arabidopsis thaliana* from stock center collections. Using the two contrasting conditions of three weeks vernalization at 4°C versus no cold treatment, we tested the stability of these traits in the two treatments, their genetic component (heritability), as well as their contribution to total seed production.

The work described here thus contributes to the selection of morphological and phenological traits that capture and describe the greatest proportion of variation in life history traits in *Arabidopsis* accessions. The identification of such traits will allow optimization of procedures that attempt to describe the variation in life histories between accessions. We then validate the core collection of 24 accessions by determining how well it captures the diversity in these chosen traits.

Material and methods

Choice of multiple criteria to optimize a set of informative variables

Contribution to seed production

Examining the most important components of the seed production such as vegetative biomass or number of flowering units is a common practice to estimate fitness values in selfing plant species in which male and female reproductive fitness are both reflected in seed set. Identification of factors contributing to seed productivity and quality is also a common exercise for plant breeders aiming to improve crop yield. However, identification of the most adaptive traits in wild plants like *Arabidopsis* is sometimes hindered by the plasticity of the response when measured in standard laboratory conditions. Under absence of cold treatment and infinite time to flower, the latest plants to flower would present ro-

settes with high number of leaves and astonishing large size, well outside the natural size range of rosettes observed under natural conditions. Having accumulated a lot of resources, these plants can be highly productive but only under those particular artificial conditions. Experimental measures may thus sometimes be only poorly informative on how accessions indeed vary under their natural ecological range. This fact is acknowledged in the "genotype × environment interaction, or reaction norms" literature (Pigliucci and Schlichting 1997; Sultan 2000 and references therein). It may therefore be misleading to rely on the single criterion of correlation with total seed production to decide what variables are more valuable when measured in a restricted subset. So another criterion also has to be applied when considering the most valuable traits to be measured.

High genetic heritability

To determine the adaptive significance of a trait, another key criterion is its potential to respond to selection in a given environment, measured by its genetic variance in this environment. The higher this variance compared to that induced by environmental heterogeneity, the higher the heritability of the trait. Unfortunately, heritability also strongly depends on the choice of accessions as well as on the experimental conditions under which a given trait is measured. Further, a trait may be highly heritable but stable across a range of environmental conditions or highly heritable but conversely strongly affected by an environmental gradient (the genetic component building the phenotype in each particular condition). Therefore, the degree of plasticity of a trait may be unrelated to its heritability value and another criterion could still valuably be added.

Conservation of accessions ranking among environments

Finally, to progress in the understanding of plant adaptation, we could compare the same traits among several conditions in search for the most stable patterns. Doing so will not necessarily focus our choice on the ecologically most relevant traits because the interesting traits may be those showing high plasticity or because stability could first of all reflect the evenness of an underlying selective force across the natural species range that has built up that trait. Still, the underlying assumption in the choice of a stability criterion would be that this strategy will necessarily maximize the probability that a laboratory-measured response will reflect most precisely the response expected under natural conditions. For example, trichome number at the surface of the leaf is considered as stable across environmental variations such as degree of humidity or day length. Therefore, the trichome number as observed in greenhouse conditions reflects the value expected under natural conditions. As few data are available on the ecological origin of the accessions, we have no way to either reproduce the natural range of conditions or to get *a priori* knowledge of the most interesting variables, and thus, in these particular conditions, the quality of the measure to reflect the 'natural' value has also to be taken into account.

Hence, highest impact on fitness, highest heritability, and highest stability across variable conditions (here cold treatment) will be the three criteria under which we will compare the beneficial output for each of the 23 variables that have been measured here for more than 1400 plants (240 plants × 2 conditions × 3 replicates).

Plant material

The *Arabidopsis thaliana* collection used for morphological data consisted of 240 accessions, primarily obtained from the Arabidopsis Biological Resource Center (ABRC) at the Ohio State University and Nottingham (NASC) stock centers or collected by French (Lavigne et al. 2001; Le Corre et al. 2002) and Japanese (Todokoro et al. 1995) groups. From this set of 240 accessions plus an additional 25, a core collection has been generated (McKhann et al. 2004) using the maximization strategy of SNP allelic richness and Nei's diversity index employing the MSTRAT software (Gouesnard et al. 2001). This strategy has been shown to perform particularly well when accessions are from populations with restricted gene flow or when accessions are primarily selfing such as is the case of *Arabidopsis* (Bataillon et al. 1996). Twenty-four accessions could capture the majority (384 out of 399) of all the SNPs discovered by analyzing 10 fragments (two per chromosome) of 517–660 bp. This set of 24 accessions belonging to the so-called "core collection" is described in Table 1. All accessions were submitted to one round of multiplication prior to this experiment to homogenize seed age and maternal effects.

Collecting morphological data

All plants were placed in the greenhouse on the same day during Spring 2001 under the same growing conditions, but they differed in their pre-planting treatment: 3 weeks at 4°C for vernalization versus direct seed sowing for the non-vernalized plants. Each accession was represented by three plants in each treatment. The two sets of plants (vernalized versus non-vernalized) were grown separately according to an incomplete balanced randomized block design (the replicates are randomized within each block; blocks are not mixed but instead are maintained as group units so that two replicates can never be in the same tray). Eight trays were used as block units for each of the two treatments. On each tray, plants were separated from each other by 3 cm so that rosettes experienced some competition by the end of their development. The outermost line was planted with *A. thaliana* accession L*er* to limit possible border effects and was excluded from further analysis. During growth, several primary traits were measured: time to first flower (FLOR); time to first mature silique as monitored by silique yellowing (MATURE); rosette diameter after 21 days (DIAM21); height from soil to first flower (H1FL) and number of leaves on the rosette at bolting (LEAF). All other traits were measured at harvest. They were: the diameter of the first axis at rosette emergence (STEMDIA); the height from soil to first silique (H1SIL) as well as maximum plant height (HMAX); the cumulated length of all primary or secondary branches bearing siliques separately on the primary axis (LSILAP) and on the secondary axis (LSILAS); the numbers of cauline leaves (CAULIN), flowering axes (AXIS), primary and secondary branches on the first and secondary axes (NRAM1AP, NRAM1AS, NRAM2AP, NRAM2AS respectively), number of green and mature siliques on first and secondary axes (SILVAP, SILVAS, SILMAP, SILMAS respectively). Last, derived traits were the total number of

Table 1. Core collection of 24 accessions of *Arabidopsis thaliana*.

Stock center number	Versailles number	Name	Country	Latitude	Longitude
N1094	162AV	Ct-1	Italy	37.3	15.06
N1436	224AV	Oy-0	Norway	60.23	6.13
N929	236AV	Shahdara	Tadjikistan	37.29	71.3
N1030	180AV	Blh-1	Czech Republic	48.49	16.45
N1028	172AV	Bur-0	Eire	53.07	−9.04
	25AV	JEA	France	43.41	7.2
N902	166AV	Cvi-0	Cape Verde Islands	16	−24
N1244	157AV	Ita-0	Morocco	34.04	−4.12
N1186	101AV	Ge-0	Switzerland	46.12	6.1
N1656	178AV	Alc-0	Spain	40.29	−3.22
N1534	62AV	St-0	Sweden	59.19	18.03
N1064	163AV	Can-0	Canary Islands	28	−15.3
	8AV	PYL-1	France	44.39	−1.1
N22491	266AV	Konchezero	Russia	62.07	34.01
N1380	94AV	Mt-0	Libya	32.34	22.46
N1368	215AV	Mh-1	Poland	53.31	20.12
	257AV	Sakata	Japan	38.55	139.5
N1210	200AV	Gre-0	U.S.A	43.11	−85.15
N1122	83AV	Edi-0	United Kingdom*	50.57	−3.13
N1286	70AV	Kn-0	Lithuania	54.54	23.54
N1564	91AV	Tsu-0	Japan	34.19	129.19
	252AV	Akita	Japan	39.43	140.06
N968	42AV	Bl-1	Italy*	44.29	11.2
N1538	92AV	Stw-0	Russia	52.57	36.04

*Botanical garden.

flowering heads (HEADS) measured by summing the variable AXIS to the four variables NRAM1AP, NRAM1AS, NRAM2AP, NRAM2AS; the total silique production (TOT SIL), a sum of the four variables SILVAP, SILVAS, SILMAP, SILMAS; and the mean distance between siliques (LEN), as total length of silique areas (LSILAS + LSILAP) divided by adjusted number of branches (TOTSIL − HEAD) to take into account that there is always one more silique than inter-silique number on each branch.

These variables thus describe the height and volume occupied by the inflorescence, its degree of ramification and compactness.

Heritability of traits

The heritability of each trait was computed independently for each treatment using decomposition into several variance components as estimated by the VARCOMP procedure in SAS®. Here, the block (tray) effect was considered as fixed while the accessions were defined as a random factor. Assuming that accessions were completely homozygous, their associated variance component was thus considered an estimation for the genetic variance V(G), while the within-accession-within-tray component (error factor) was taken as an estimation of the environmental error V(E). Assuming that maternal effect was negligible, since all seeds came from mother plants previously multiplied under the same green house conditions, heritability, h^2, was then computed as the proportion of the variance under the genetic component V(G)/ (V(G)+V(E)). As in this experiment we could not further decompose the variance to access the additive part of the genetic variance, h^2 as presented here is a *sensu lato* heritability.

Analyzing stability between treatments

The Spearman's rank correlation was taken as a simple measure of the maintenance of the same accession ranking when changing from the vernalized to non-vernalized treatment. A variable with a high Spearman coefficient would ensure that an accession behaves similarly relative to other accessions irrespective of the cold pre-treatment; i.e., a small value in one environment would be matched by a small value in the other. Thus, variables showing opposite reaction

norms in different accessions according to the treatment would definitely not be favored by this criterion. We also plotted cluster trees of the 23 variables using the Pearson coefficient as a distance measure to illustrate the relationship among variables and the degree of pattern conservation among variables across treatments.

Extracting fitness components from the inflorescence architecture

With such a high number of accessions, nearly all traits were found to be significantly correlated to fitness. Thus, using a forward stepwise option, variables were ranked by the number of steps necessary to get them to enter into a multi-linear model estimating the total silique production (TOT SIL). This was carried out for each treatment independently using stepwise probabilities p_{enter} = 0.15, p_{remove} = 0.05 to enter or remove variables from the model, respectively. Under both treatments, nine variables introduced in the model were sufficient for the linear regression coefficient R^2 to reach high values. As there was no clear further improvement of the model fitting when entering supplementary variables, we stopped the sorting at step nine. In Table 2, only those nine variables kept in the model have their rank at entering (all others being labeled "*ne*" for not entered in the model). Ranking orders were then compared between the two treatments and the most often encountered variables were kept in the list of primary components of fitness. Of course, as they were directly correlated to total seed production, variables giving number of siliques or length of silique areas (which poorly represent architecture) were excluded from this analysis. Our approach therefore optimizes a complementary set of variables that best models the seed production.

Results and discussion

Heritabilities and correlations among traits

The 20 primary and 3 indirect variables studied exhibited a wide range of heritability values from virtually zero for the number of first branches on the secondary axis (NRAM1AS) in the vernalized treatment to 0.87 for flowering time (FLOR) in the non-vernalized treatment. With about 240 degrees of freedom, all heritability values exceeding 0.04 were statistically highly significant (p < 0.001). In most cases, the cold treatment tended to homogenize phenotypic expression across accessions, leading to lower heritability values. It is easy to illustrate that choosing the most significant variables according to this single criterion would have been misleading. For example, the heritability index considered alone would lead to keeping both flowering date (FLOR) and number of rosette leaves at flowering (LEAF) because they had the highest h^2 values in both treatments. However, it has long been established that these two variables are so tightly correlated that many teams in fact use the number of leaves on the rosette as their index of flowering precocity. Part of the measuring effort would thus be lost by measuring both of these rather redundant variables.

With 23 variables measured under two conditions, we could compare the correlation between pairs of variables as well as the correlation for a single variable between treatments (paired by accession number). We found that, in al-

most all cases, the correlation between two traits measured in the same environment (treatment) was higher than the correlation for the same trait between the two treatments. This observation demonstrates that the necessary equilibrium between different plant compartments such as roots and leaf area or other physiological constraints as shaped by the growth environment prevails over the genetic background in determining the relationship among traits. The comparison between the two cluster trees presented in Fig. 1 illustrates the degree of conservation of the relationship between variables. Some clusters such as the one containing the flowering time FLOR, the time to silique maturation MATURE and the number of leaves on the rosette at bolting LEAF are indeed highly conserved. Note also that variables such as NRAM1AP and NRAM1AS, although referring to a same character (number of ramifications but on a different flowering axis) can belong to different clusters, suggesting that their timing, underlying physiological and/or genetic pathways are different.

Relationships between traits and fitness

Individual fitness, measured as the total number of siliques, was shown to result from somewhat different sets of explicative traits according to the treatment considered. Without vernalization, greater variation was observed among accessions in the length of their vegetative phase, mainly because more accessions were late flowering. In this case, more plants could accumulate enough resources to develop a complex architecture and the number of siliques produced then depended on the maximum plant height, the number of siliques on the main reproductive axis and the total number of reproductive axes. In contrast, under vernalization, the vegetative phase was reduced for most plants so that the production of siliques mainly depended on the number of ramifications on the first axis. Thus, the effective significance of the trade-off between precocity and seed production seems strongly shaped by environmental conditions. Whatever the treatment, all accessions flowering with less than 22 rosette leaves at bolting suffered a fitness cost that can only be compensated in their original environment by some other advantage conferred by earliness.

Stability of traits across vernalized versus non-vernalized conditions

As shown in Table 2, the most stable traits were also those having the highest heritability values, so that the two criteria of tttththehigh heritability and high ranking stability of the accessions across treatments, were found here to be very consistent. This result was not obvious, since the variation at regulatory genes underlying traits with a strong genetic basis may result in highly variable norms of reaction to environmental conditions, and thus to a high "instability" for the trait. However, in the case of flowering time (and related traits), the stability in the ranking of accessions may be explained by the underlying regulation mechanism. Flowering time depends on the repressing action of the MADS-box regulatory gene FLC (Michaels and Amasino 1999; Sheldon et al. 2000), whose level of expression is constitutively up-regulated by several genes (FRIGIDA and other genes in the "autonomous pathway") that determine flowering time in the absence of vernalization, and eventually, down-regulated by

Table 2. Ranking of 23 morphological variables under vernalized and non-vernalized treatments using their respective values calculated using three different criteria. See Materials and Methods for details on each criterion. Columns are coefficient and variable ranks under each treatment, respectively.

Nickname	Variable	Criterion 1 Heritability Non-vernalized		Vernalized		Criterion 2 Regression stepwise entering rank Non-vernalized	Vernal-ized	Criterion 3 Spearman correlation Coeffi-cient	Variable rank
LEAF	Number of leaves	0.82	**2**	0.73	**1**	6	*ne*	0.602	**1**
FLOR	Flowering time	0.86	**1**	0.69	**2**	7	*ne*	0.583	**2**
NRAM1AP	Number of primary branches on first axis	0.72	**4**	0.63	**3**	*ne*	ne	0.548	**4**
CAULIN	Number of cauline leaves	0.75	**3**	0.62	**4**	*ne*	*ne*	0.580	**3**
MATURE	Mature silique time	0.71	**5**	0.61	**5**	*ne*	7	0.503	**7**
H1SIL	Height of first silique	0.59	**8**	0.56	**6**	9	5	0.511	**5**
HMAX	Maximum height	0.53	**9**	0.51	**7**	1	2	0.507	**6**
STEMDIA	Main stem diameter	0.44	13	0.51	**8**	8	6	0.406	13
LEN	Mean length between siliques	0.41	15	0.47	**9**	3	3	0.397	14
LSILAP	Cumulated length of siliques along first axis	0.43	14	0.15	10	*ne*	*ne*	0.498	**8**
SILVAP	Number of green siliques on first axis	0.40	16	0.38	11	*ne*	*ne*	0.367	19
NRAM2AP	Number of secondary branches on first axis	0.39	17	0.38	12	*ne*	**4**	0.378	16
SILMAP	Number of siliques on first axis	0.49	11	0.35	13	*ne*	*ne*	0.200	22
H1FL	Height of first flower	0.53	10	0.34	14	*ne*	8	0.481	**9**
HEAD	Total heads	0.38	19	0.33	15	**4**	1	0.451	10
TOTSIL	Total siliques	0.39	18	0.32	16	*ne*	*ne*	0.367	19
DIAM21	Rosette diameter at day 21	0.37	20	0.27	17	*ne*	*ne*	0.371	17
AXIS	Number of axes	0.09	22	0.20	18	*ne*	**9**	0.360	20
SILVAS	Number of green siliques on second axis	0.62	7	0.15	19	*ne*	*ne*	0.319	21
LSILAS	Cumulated length of siliques along second axis	0.47	12	0.33	20	*ne*	*ne*	0.368	18
SILMAS	Number of mature siliques on second axis	0.15	21	0.11	21	*ne*	*ne*	0.383	15
NRAM2AS	Number of secondary branches on second axis	0.69	**6**	0.00	22	**5**	*ne*	0.451	10
NRAM1AS	Number of primary branches on second axis	0.04	23	0.00	23	**2**	*ne*	0.411	12

ne = not entered into the multiple regression model.

vernalization. As effect of a given vernalization treatment probably depends on the constitutive level of expression of FLC, the more late flowering accessions in the absence of vernalization are also the more late-flowering ones after vernalization.

Here we propose a set of the most relevant variables based on the quality of the data collected as estimated via our multiple criteria applied here. These variables are flowering time (or the number of rosette leaves), maximum plant height, mean distance between siliques, total number of flowering heads and height to first silique (as well as main stem diameter to a lesser degree). It is clear from Fig. 1 that this set of variables covers different parts of the cluster trees. It demonstrates the low degree of redundancy between these variables (that otherwise would fall on the same branch of the cluster). Retaining only these variables allows accurate mod-

Fig. 1. Cluster trees between the 23 variables.

Trees were constructed separately under each condition using complete (conservative) clustering option and Pearson coefficient as distance measure. 100 re-sampling bootstraps on accessions were performed on each tree to test mean distance between variables, cluster branches and robustness of nodes. Bootstrap values exceeding 60% are added on graphs.

Variables: The four variables are highlighted in grey (LEN: mean distance between siliques; H1SIL: height from soil to first silique; HEAD: total number of flowering heads; HMAX: maximum plant height) and flowering time (FLOR) in italic. Belonging to different cluster branches, these four variables have limited redundancy (especially under the non-vernalized treatment). Note that NRAM1AP and NRAM1AS do not cluster tightly although they both refer to ramification but on a different flowering axis.
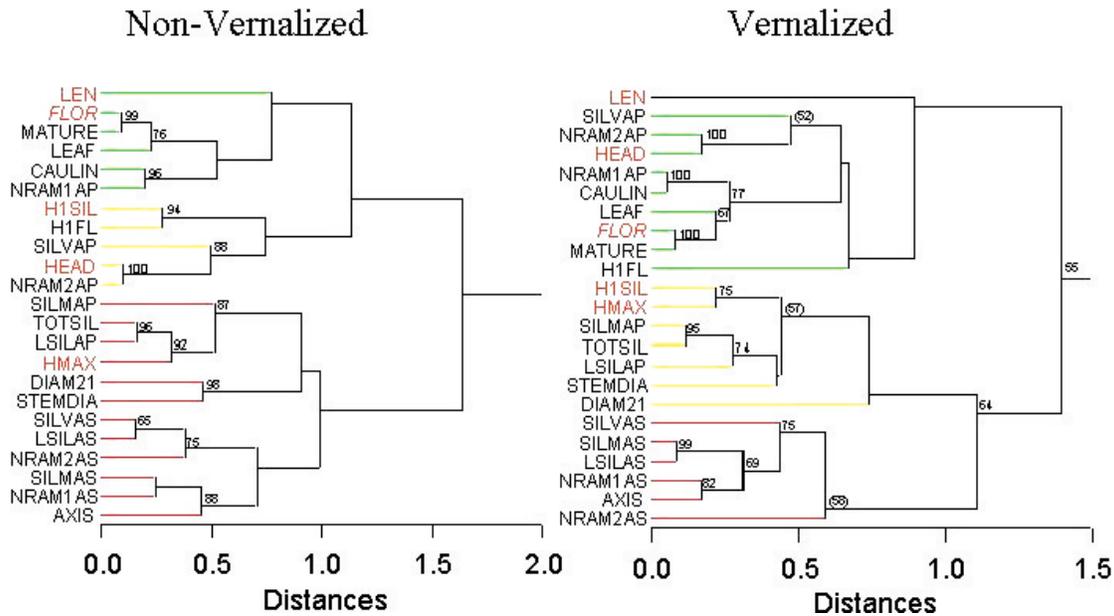
eling of total seed production in both treatments with $R^2 = 0.948$ and 0.952 for vernalized and non-vernalized treatments, respectively. With the exception of H1SIL under non-vernalized conditions ($p = 0.005$), all variables have high significant impact on the regression ($p < 0.001$) and thus all affect seed production.

**Diversity in a core collection of 24 accessions**

The core collection of 24 accessions selected for maximizing molecular diversity was then tested for its ability to cover the total range of morphological variation as well as its distribution evenness for each of the five more interesting variables. Because both vernalized and non-vernalized data sets are very similar, the illustration presented in Fig. 2 as a bar chart for each variable separately or Scatterplot matrix for pairs of variables is from the non-vernalized data set only.

We demonstrated that the core collection of 24 accessions contained significantly more variation for morphological traits than random samples of the same size (drawn from the full collection of 240 accessions) when analyzed globally on a set of morphological characters (data not shown). This is also illustrated here for individual trait distribution (bar charts in Fig. 2). For the core collection, the distribution for each trait is flattened compared to that of the whole sample and extreme values are often represented (as seen for FLOR, HMAX and HEAD). At the same time, the core collection is not only a subset of the most extreme accessions but rather includes accessions that regularly cover the whole range. In

other words, the core collection contains accessions that occupy most of the morphological space and not a restricted set of accessions belonging only to the external envelope bordering that morphological space. Similarly, the areas covered by the core collection in the two by two plots are close to the maximum area of the whole sample. With 10% of the total number of the accessions (24 out of 240), these regions always exceed 120% of the mean areas covered by 24 accessions taken at random.

Many scientific teams in the field of population genetics and evolution have developed programs to sample, describe and relate the variation present in natural populations to adaptation (e.g., Van Tienderen et al. 2002). In the model plant *Arabidopsis thaliana*, such diversity is often found in accessions collected over a wide range of ecological and geographic conditions along the species distribution. An approach using quantitative trait loci (QTL) on a subset of crosses between some of these natural accessions can allow the identification and isolation of genes responsible for the control of a particular trait. This is not possible when using the mutant approach because of the narrow genetic background of these lines. Natural accessions submitted to contrasting environmental conditions over long time periods have often developed a unique original combination of genes or alleles to fit a local optimum. Analyzing the diversity of the genetic responses developed by different accessions may allow the identification of different (parts of) metabolic pathways affecting a trait. It is therefore a complementary approach to the mutant analysis strategy for establishing gene functions.

Fig. 2. Representativeness of the 24 accessions belonging to the core collection for the flowering time and each of the four added variables. Evenness of the distribution is depicted on the bar charts, and the surrounding shapes demonstrate the diversity covered by the core collection for each pair of variables as compared to a world-wide set of accessions.
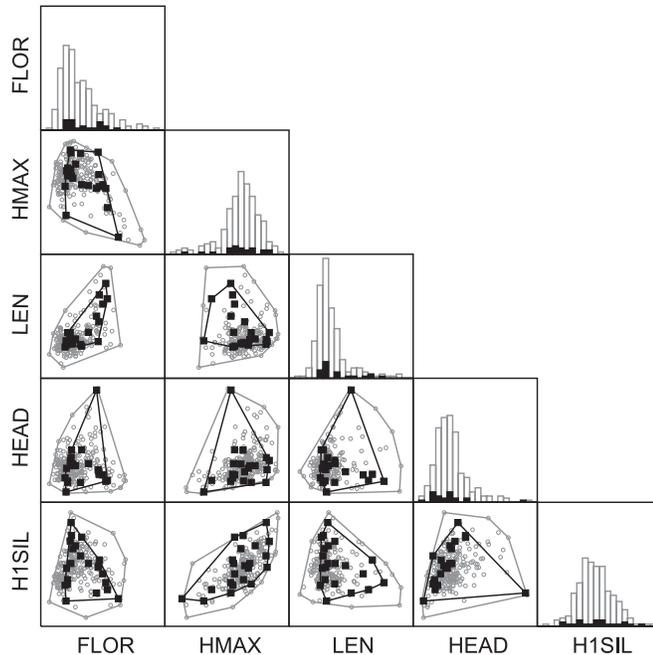
Conclusion

We have used multiple approaches to analyze 23 morphological characteristics from 240 accessions grown under two different cold treatments to search for environmentally stable and heritable fitness components in wild accessions of *Arabidopsis thaliana*. Four traits were selected in addition to the already well-established flowering precocity: maximum plant height, mean distance between siliques, number of flowering heads and height to first silique. All of these traits are easy to measure and could significantly reduce the effort needed to characterize natural diversity and to obtain a rough estimation of fitness. We believe that these traits reflect the phenotypic ability to thrive in a more or less competitive environment as determined by contrasting climatic conditions including the winter cold period.

This study also demonstrated the validity of the core collection strategy as a key method to seek natural diversity. Indeed, the core collection of 24 accessions maximizing the molecular diversity at the DNA level was confirmed here to cover most of each variable range. A reasonable amount of diversity was thus available in such a small sample and the core collection could be used to study the way genetic variation modifies the physiological relationships between traits

in different contrasting environments, to search for allelic polymorphism in some candidate genes known to be involved in a trait and eventually to produce segregating populations for QTL mapping and cloning.

References

Alonso-Blanco, C., and Koornneef, M. 2000. Naturally occurring variation in *Arabidopsis*: an underexploited resource for plant genetics. Trends Plant Sci. **5**: 22–29.

Bataillon, T.M., David, J.L., and Schoen, D.J. 1996. Neutral genetic markers and conservation genetics: simulated germplasm collections. Genetics **144**: 409–417.

Gouesnard, B., Bataillon, T.M., Decoux, G., Rozale, C., Schoen, D.J., and David, J.L. 2001. MSTRAT: An algorithm for building germ plasm core collections by maximizing allelic or phenotypic richness. J. Hered. **92**: 93–94.

Lavigne, C., Reboud, X., Lefranc, M., Porcher, E., Roux, F., Olivieri, I., and Godelle, B. 2001. Evolution of genetic diversity in metapopulations: *Arabidopsis thaliana* as an experimental model. Genet. Select. Evol. **33** (Suppl. 1): S399–S423.

Le Corre, V., Roux, F., and Reboud, X. 2002. DNA polymorphism at the FRIGIDA gene in *Arabidopsis thaliana*: extensive non-synonymous variation is consistent with local selection for flowering time. Mol. Biol. Evol. **19**: 1261–1271.

McKhann, H.I., Camilleri, C., Berard, A., Bataillon, T., David, J.L., Reboud, X., Le Corre, V., Caloustian, C., Gut, I.G., and Brunel, D. 2004. Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. Plant J. **38**: 193–202.

Michaels, S.D., and Amasino, R.M. 1999. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell **11**: 949–956.

Pigliucci, M., and Schlichting, C.D. 1997. Reaction norms of *Arabidopsis* V. Flowering time controls phenotypic architecture in response to nutrient stress. J. Evol. Biol. **3**: 285–301.

Sheldon, C.C., Finnegan, E.J., Rouse, D.T., Tadege, M., Bagnall, D.J., Helliwell, C.A., Peacock, W.J., and Dennis, E.S. 2000. The control of flowering by vernalization. Curr. Opin. Plant Biol. **3**: 418–422.

Sultan, S.E. 2000. Phenotypic plasticity for plant development, function and life history. Trends Plant Sci. **5**: 537–542.

Todokoro, S.R., Terauchi, R., and Kawano, S. 1995. Microsatellite polymorphisms in natural populations of *Arabidopsis thaliana* in Japan. Jap. J. Genet. **70**: 543–554.

Van Tienderen, P.H., de Haan, A., van der Linden, C.G., and Vosman, B.L. 2002. Biodiversity assessment using markers for ecologically relevant traits. TREE **17**: 577–582.