GENOMICS ARTICLE

Growth Stage–Based Phenotypic Analysis of Arabidopsis: A Model for High Throughput Functional Genomics in Plants

Douglas C. Boyes, Adel M. Zayed, Robert Ascenzi, Amy J. McCaskill, Neil E. Hoffman, Keith R. Davis,[1] and Jörn Görlach

Department of Plant Research, Paradigm Genetics, Inc., Research Triangle Park, North Carolina 27709

With the completion of the Arabidopsis genome sequencing project, the next major challenge is the large-scale determination of gene function. As a model organism for agricultural biotechnology, Arabidopsis presents the opportunity to provide key insights into the way that gene function can affect commercial crop production. In an attempt to aid in the rapid discovery of gene function, we have established a high throughput phenotypic analysis process based on a series of defined growth stages that serve both as developmental landmarks and as triggers for the collection of morphological data. The data collection process has been divided into two complementary platforms to ensure the capture of detailed data describing Arabidopsis growth and development over the entire life of the plant. The first platform characterizes early seedling growth on vertical plates for a period of 2 weeks. The second platform consists of an extensive set of measurements from plants grown on soil for a period of ∼2 months. When combined with parallel processes for metabolic and gene expression profiling, these platforms constitute a core technology in the high throughput determination of gene function. We present here analyses of the development of wild-type Columbia (Col-0) plants and selected mutants to illustrate a framework methodology that can be used to identify and interpret phenotypic differences in plants resulting from genetic variation and/or environmental stress.

INTRODUCTION

The completion of the Arabidopsis genome sequence has paved the way for the development of numerous reverse genetics approaches for the determination of gene function. In most cases, the identification of mutations in the gene of interest is a straightforward, if sometimes laborious, process (Krysan et al., 1999). However, difficulties often arise in the identification of a phenotype that can be associated with the mutation. This is particularly true in plants, in which genes often exist as members of multigene families that exhibit redundant or highly specialized functions. Homology searching, microarray-based analysis of gene expression, and metabolic studies likely will provide some information regarding gene function. In most cases, however, a complete understanding of a gene's function will be realized only when this information can be associated with a phenotype at the organismal level. Many of these phenotypes may be manifested as subtle changes in growth or development, underscoring the need for a sensitive and robust methodology for their detection.

Our approach to this problem has been to develop an extensive phenotypic analysis process for capturing data describing growth and development during the entire life of the plant. Single-gene mutations or altered environmental conditions may affect any number of traits, resulting in morphological changes and/or altered timing of development. Morphological changes often can be identified readily and recorded outside the context of extensive temporal or growth stage information. In contrast, mutations that result in altered developmental progression without major morphological changes can be recorded only if a temporal component is included in the analysis. Such time-course analyses are inevitably resource intensive and must be restricted in scope to allow the analysis of many samples in parallel. We have addressed this issue by developing a method for phenotyping plants based on a series of defined growth stages. The growth stages serve both as developmental landmarks and as triggers for the collection of morphological data that are of interest at specific stages of development.

Developmental growth stages have been described at the organismal level for a variety of experimental models, including *Drosophila melanogaster* (Hartenstein, 1993) and *Caenorhabditis elegans* (Wilkins, 1993). However, with the

---

[1] To whom correspondence should be addressed. E-mail kdavis@paragen.com; fax 919-485-8789.

exception of definitions describing the phases of specific organs or tissue types (Smyth et al., 1990), very little unification in the stages of Arabidopsis growth and development has been achieved. Growth stage definitions for other plants have been developed, and these are used routinely in the breeding industry. One such example is the BASF, Bayer, Ciba-Geigy, Hoechst (BBCH) scale, which was proposed to provide a generic nomenclature for the assignment of growth stages in crop plants and weeds (Lancashire et al., 1991). We have used the BBCH scale as a basis to define a series of growth stages for use in the phenotypic analysis of Arabidopsis.

The analysis of Arabidopsis growth and development presented here provides a framework methodology for identifying and interpreting phenotypic differences in plants resulting from genetic variation and/or environmental stress. The utility of this methodology is validated through the discovery of new growth and development phenotypes for mutants identified previously as having defects in specific biochemical pathways but with subtle or no phenotypes observed at the organismal level. The phenotypic analysis platforms described here, in conjunction with metabolic and gene expression profiling analyses conducted in parallel, provide a robust method for the high throughput functional analysis of plant genes.

RESULTS

Growth Stages

Tables 1 and 2 list the growth stages that we have adapted from the BBCH scale (Lancashire et al., 1991) for use in the analysis of Arabidopsis phenotypes. Together, these 30 growth stages cover the development of the plant from seed imbibition through the completion of flowering and seed maturation. Tables 1 and 2 also show the time required for wild-type Columbia (Col-0) plants to reach each stage when grown under standard environmental conditions using a 16-hr daylength. As shown in Figure 1, the defined growth stages span the entire life cycle of the plant, thereby maximizing the ability to detect subtle changes that affect only a limited aspect of development. Furthermore, the coefficients of variation (CV) associated with these data generally are <15% (Tables 1 and 2), indicating that the developmental progression of wild-type plants is highly reproducible. Together, these findings suggest that this data set is a robust representation of wild-type development with which all mutants and environmentally stressed plants may be compared.

Data Collection Model

Our data collection model consists of two overlapping phases. In the first phase, a set of core measurements is performed at routine intervals during the course of development. The resulting data reflect the rate of plant growth and development and are used in growth stage assignment. Figure 2 illustrates the subset of growth stages used as key developmental landmarks determined during this process. The second phase of the data collection process is triggered periodically throughout development as these landmark growth stages are reached. Data collected during this phase reflect a wide range of morphological traits at multiple preflowering and postflowering growth stages.

We have developed two complementary phenotypic analysis platforms that are based on this data collection model.

Table 1. Arabidopsis Growth Stages for the Plate-Based Phenotypic Analysis Platform

Stage	Description	Col-0 Data		
		Days[a]	SD	CV[b]
Principal growth stage 0	Seed germination			
0.10	Seed imbibition	3.0	NA[c]	NA
0.50	Radicle emergence	4.3	0.4	10.3
0.7	Hypocotyl and cotyledon emergence	5.5	0.6	11.2
Principal growth stage 1	Leaf development			
1.0	Cotyledons fully opened	6.0	0.5	8.5
1.02	2 rosette leaves >1 mm	10.3	0.6	5.8
1.04	4 rosette leaves >1 mm	14.4	0.5	3.4
Stage R6	More than 50% of the seedlings have primary roots ≥6 cm in length	ND[d]	ND	ND

[a] Average day from date of sowing, including a 3-day stratification at 4°C to synchronize germination.
[b] CV, coefficient of variation, calculated as (SD/days) × 100.
[c] NA, not applicable.
[d] ND, not determined (see text for details).

Table 2. Arabidopsis Growth Stages for the Soil-Based Phenotypic Analysis Platform

| Stage | Description | Col-0 Data | | |
		Days[a]	SD	CV[b]
Principal growth stage 1	Leaf development			
1.02	2 rosette leaves >1 mm in length	12.5	1.3	10.7
1.03	3 rosette leaves >1 mm in length	15.9	1.5	9.5
1.04	4 rosette leaves >1 mm in length	16.5	1.6	9.8
1.05	5 rosette leaves >1 mm in length	17.7	1.8	10.2
1.06	6 rosette leaves >1 mm in length	18.4	1.8	9.8
1.07	7 rosette leaves >1 mm in length	19.4	2.2	11.1
1.08	8 rosette leaves >1 mm in length	20.0	2.2	11.2
1.09	9 rosette leaves >1 mm in length	21.1	2.3	10.8
1.10	10 rosette leaves >1 mm in length	21.6	2.3	10.9
1.11	11 rosette leaves >1 mm in length	22.2	2.5	11.2
1.12	12 rosette leaves >1 mm in length	23.3	2.6	11.3
1.13	13 rosette leaves >1 mm in length	24.8	3.2	12.8
1.14	14 rosette leaves >1 mm in length	25.5	2.6	10.2
Principal growth stage 3	Rosette growth			
3.20	Rosette is 20% of final size	18.9	3.0	16.0
3.50	Rosette is 50% of final size	24.0	4.1	17.0
3.70	Rosette is 70% of final size	27.4	4.1	15.0
3.90	Rosette growth complete	29.3	3.5	12.0
Principal growth stage 5	Inflorescence emergence			
5.10	First flower buds visible	26.0	3.5	13.3
Principal growth stage 6	Flower production			
6.00	First flower open	31.8	3.6	13.3
6.10	10% of flowers to be produced have opened	35.9	4.9	13.6
6.30	30% of flowers to be produced have opened	40.1	4.9	12.3
6.50	50% of flowers to be produced have opened	43.5	4.9	11.2
6.90	Flowering complete	49.4	5.8	11.7
Principal growth stage 8	Silique ripening			
8.00	First silique shattered	48.0	4.5	9.3
Principal growth stage 9	Senescence			
9.70	Senescence complete; ready for seed harvest	ND[c]	ND	ND

[a] Average day from date of sowing, including a 3-day stratification at 4°C to synchronize germination.
[b] CV, coefficient of variation, calculated as (SD/days) × 100.
[c] ND, not determined (see text for details).

The early development platform encompasses the stages of seed imbibition through the early vegetative phase (stages 0.1 through 1.04), with data collected from seedlings grown on vertical plates. The soil-based phenotypic analysis platform encompasses the stages of leaf development, flower production, and seed maturity (stages 1.02 through 9.70), with data collected from plants grown in a horticultural potting medium.

The first-phase measurements for both platforms are listed in Tables 3 and 4. In both platforms, seed are stratified after sowing with a 4°C treatment for 3 days. For plate-based analysis, observations are made of the imbibed seed upon transfer to growing conditions and each day thereafter for 11 days. The plate-based analysis is terminated at this time (day 14 after sowing) or when growth stage R6 (>50% of the seedlings have primary roots ≥6 cm in length) is
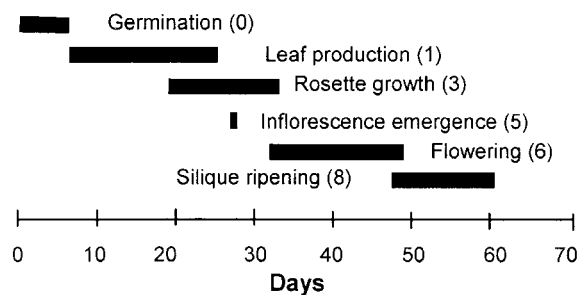

Figure 1. Scheme of the Chronological Progression of Principal Growth Stages in Arabidopsis.

Horizontal bars indicate the period during wild-type Col-0 plant development when the indicated trait can be used in growth stage determination. Numbers in parentheses correspond to principal growth stages listed in Table 1.
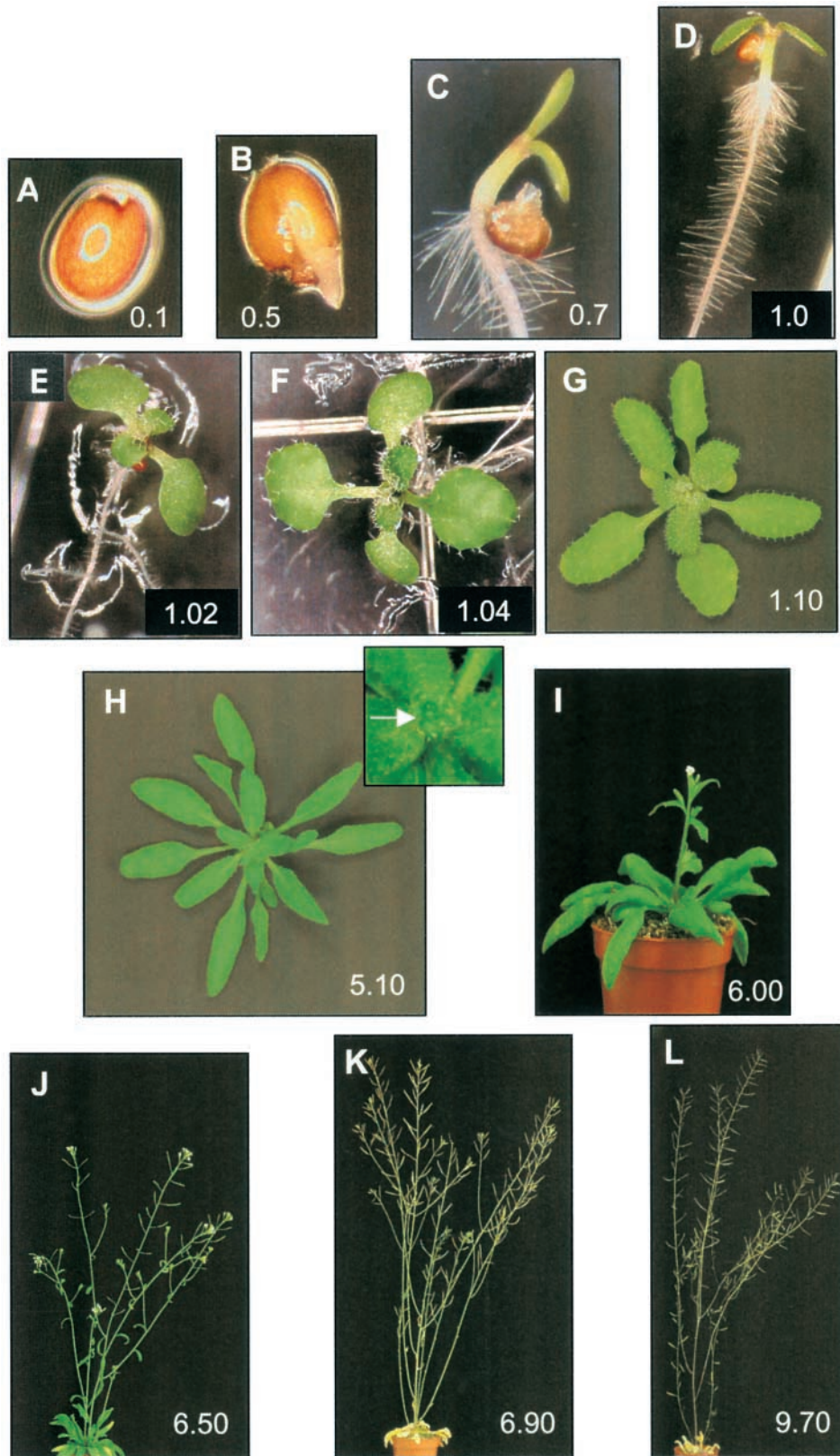
Figure 2. Arabidopsis Growth Stages.

Table 3. Measurements Performed during the Plate-Based Phenotypic Analysis

First-Phase Measurements					
Measurement/Query					**Growth Stage Defined**
Has radicle emergence been reached or passed?					Stage 0.5
Are the hypocotyl and cotyledons visible?					Stage 0.7
Have the cotyledons opened fully?					Stage 1.0
Number of rosette leaves >1 mm					Principal growth stage 1
Length of primary root ≥6 cm					Stage R6

Second-Phase Measurements					
			Col-0 Data		
Measurement	**Unit**	**Growth Stage**	**Average**	**SD**	**CV[a]**
Number of rosette leaves	Count	1.0[b]	3.3	0.5	13.7
Length of primary root	mm	1.0[b]	45.2	4.1	9.0
Number of secondary roots	Count	R6[c]	10.5	1.4	13.2
Rosette, total exposed leaf area	mm²	R6[c]	22.2	2.6	11.9
Rosette, perimeter	mm	R6[c]	42.1	5.2	12.2
Rosette, SD of radius	None	R6[c]	39.2	3.1	7.9
Rosette, major axis	mm	R6[c]	7.9	0.7	9.1
Rosette, minor axis	mm	R6[c]	5.9	0.6	9.5
Rosette, eccentricity	None	R6[c]	0.63	0.05	7.2

[a] CV, coefficient of variation, calculated as (SD/days) × 100.
[b] Data collection initiated at stage 1.0 and continued until the end of the experiment.
[c] R6 or 14 days after sowing, whichever comes first.

reached, which ever comes first. In the soil-based analysis platform, measurements are made every 2 days beginning 5 days after transfer to standard growing conditions and continue throughout the life of the plant. These core measurements are used in growth stage determinations and can be performed rapidly, requiring only a simple mechanical measurement with a caliper or ruler (e.g., maximum rosette radius) or a visual inspection (e.g., number of rosette leaves). Representative data taken from the analysis of wild-type Col-0 plants are given in Figure 3. In most cases, growth stages can be assigned at the time of data acquisition. An example of this class is principal growth stage 1, in which the number of rosette leaves is the growth stage–determining trait (Table 2, Figure 3A). Traits such as this allow a clear determination of the growth stage at the time they are reached and serve as developmental landmarks that can be used to trigger second-phase data collection activities.

Figure 2. (continued).

(A) Stage 0.1, imbibition.
(B) Stage 0.5, radicle emergence.
(C) Stage 0.7, hypocotyl and cotyledons emerged from seed coat.
(D) Stage 1.0, cotyledons opened fully.
(E) Stage 1.02, two rosette leaves >1 mm in length.
(F) Stage 1.04, four rosette leaves >1 mm in length.
(G) Stage 1.10, ten rosette leaves >1 mm in length.
(H) Stage 5.10, first flower buds visible (indicated by arrow in inset).
(I) Stage 6.00, first flower open.
(J) Stage 6.50, midflowering.
(K) Stage 6.90, flowering complete.
(L) Stage 9.70, senescent and ready for seed harvest.
(A) to **(F)** were determined in the early analysis platform. **(G)** to **(L)** were determined in the soil-based platform.

Table 4. Measurements Performed during the Soil-Based Phenotypic Analysis

First-Phase Measurements	
Measurement/Query	Growth Stage Defined
Rosette radius	Principal growth stage 3
Are flower buds visible?	Stage 5.10
Is first flower open?	Stage 6.00
Length of stem	Stage 6.50[a]
Number of open flowers	Principal growth stage 6
Number of senescent flowers	Principal growth stage 6
Number of filled siliques	Principal growth stages 6 and 7
Number of shattered siliques	Principal growth stage 8
Is flower production complete?	Stage 6.90

Second-Phase Measurements					
			Col-0 Data		
Measurement	Unit	Growth Stage	Average	SD	CV[b]
Number of cotyledons	Count	1.04	2.0	0.1	5.0
Rosette, total exposed leaf area	mm^2	1.10	580.0	202.2	34.9
Rosette, perimeter	mm	1.10	418.0	119.1	28.5
Rosette, SD of radius	None	1.10	45.7	3.2	7.0
Rosette, major axis	mm	1.10	40.4	8.0	19.8
Rosette, minor axis	mm	1.10	34.6	7.5	21.7
Rosette, eccentricity	None	1.10	0.5	0.1	20.0
Rosette, total exposed leaf area	mm^2	6.00	3225.0	1088.3	33.7
Rosette, perimeter	mm	6.00	808.1	181.3	22.4
Rosette, SD of radius	None	6.00	36.7	3.5	9.5
Rosette, major axis	mm	6.00	82.3	15.3	18.6
Rosette, minor axis	mm	6.00	73.1	13.4	18.3
Rosette, eccentricity	None	6.00	0.4	0.1	25.0
Rosette, dry weight	mg	6.00	117.4	45.9	39.1
Number of stem branches on main bolt	Count	6.50	3.4	0.6	17.6
Number of side bolts >1 cm	Count	6.50	4.2	1.2	28.6
Length of peduncle of second flower on main bolt	mm	6.50	11.5	1.6	13.9
Distance across face of open flower	mm	6.50	3.9	0.3	7.7
Sepal length	mm	6.50	2.2	0.2	9.1
Pollen grain, area	μm^2	6.50	589.0	132.0	22.4
Pollen grain, perimeter	μm	6.50	114.5	13.0	11.2
Pollen grain, SD of radius	None	6.50	9.1	2.5	27.5
Pollen grain, major axis	μm	6.50	30.6	3.3	10.8
Pollen grain, minor axis	μm	6.50	24.3	3.0	12.3
Pollen grain, eccentricity	None	6.50	0.6	0.1	16.7
Silique, area	mm^2	6.50	10.6	1.9	17.9
Silique, perimeter	mm	6.50	40.9	6.1	14.9
Silique, SD of radius	None	6.50	55.8	0.5	0.9
Silique, major axis	mm	6.50	17.2	1.7	9.9
Silique, minor axis	mm	6.50	1.2	0.2	16.7
Silique, eccentricity	None	6.50	1.0	0.0	0.0
Total number of seeds per silique valve	Count	6.50	29.9	2.8	9.4
Number of abnormal seeds per silique valve	Count	6.50	0.2	0.4	200
Dry weight of stem	mg	6.50	188.8	39.3	20.8
Dry weight of rosette	mg	6.50	163.7	52.0	31.8
Total number of siliques	Count	6.90	160.4	60.7	37.8
Seed, area	mm^2	9.70	0.14	0.01	7.1
Seed, perimeter	mm	9.70	1.95	0.04	2.1
Seed, SD of radius	None	9.70	16.92	0.94	5.6
Seed, major axis	mm	9.70	0.53	0.03	5.7
Seed, minor axis	mm	9.70	0.33	0.02	6.1
Seed , eccentricity	None	9.70	0.78	0.02	2.6
Seed yield per plant (desiccated)	mg	9.70	127.9	52.7	41.2

[a] Used to define the working definition of stage 6.50 as described in the text.
[b] CV, coefficient of variation, calculated as (SD/days) $\times$ 100.

In contrast, other growth stage assignments are relative and can be determined unequivocally only in retrospect after the relevant trait has developed to completion. Examples of this class are principal growth stage 3, which is based on rosette radius (Table 2, Figure 3B), and principal growth stage 6, which is based on flower production (Table 2). Growth stages of this type are defined as a percentage of the final value of the trait being measured. For example, growth stage 6.50 is reached when 50% of the final number of flowers have been produced. Without previous knowledge of the number of flowers that will be produced by a given plant, growth stage 6.50 cannot be recognized at the time it occurs.

Growth stage definitions of this category present a challenge if they are to be used in real time to trigger additional data collection activities, particularly if the plant under study is being characterized for the first time. In the case of stage 6.50, we have been able to develop an alternative definition that approximates the midflowering stage and allows it to be recognized at the time it occurs. As depicted by the arrow in the graph of plant height over time (Figure 3C), the first 50% of flower production in wild-type Col-0 plants is completed coincident with a decrease in the rate of stem elongation. Extrapolating from this finding, we have established a working definition of stage 6.50 as the time at which the increase in stem elongation is <20% for two consecutive 2-day data collection cycles. This working definition allows stage 6.50 to be used as a real-time trigger for the activation of additional data collection activities.

Similarly, the BBCH definition for stage 9.70 (Table 2) also has been modified to accommodate the requirements of the data collection process. The inflorescence of Arabidopsis consists of a continuum of siliques at different stages of development. As such, the most mature siliques shatter to release their seed (stage 8.00) before the completion of flower production (stage 6.90; Table 2), resulting in seed loss and increasing the probability of inter-seed-lot contamination. Harvesting the inflorescence before stage 9.70 (complete senescence) is a means to minimize these potential problems. Thus, we harvest the inflorescence 2 to 4 days after the completion of flowering (stage 6.90) and store it in an envelope to complete the maturation process. The working definition of stage 9.70, therefore, is established as the time when the tissue is dried substantially to allow the seed to be released from the siliques.

The second phase of the data collection process has been designed to identify phenotypes in traits other than those used in growth stage determination. As shown in Table 3, second-phase data collection in the early development platform first occurs at growth stage 1.0. This growth stage initiates the root length and the number of rosette leaves measurements as well as a series of descriptor choice lists to qualitatively describe seedling abnormalities. Collection of these data is performed daily until the completion of the experiment, which is triggered by growth stage R6, or on day 14, whichever comes first. At that time, the
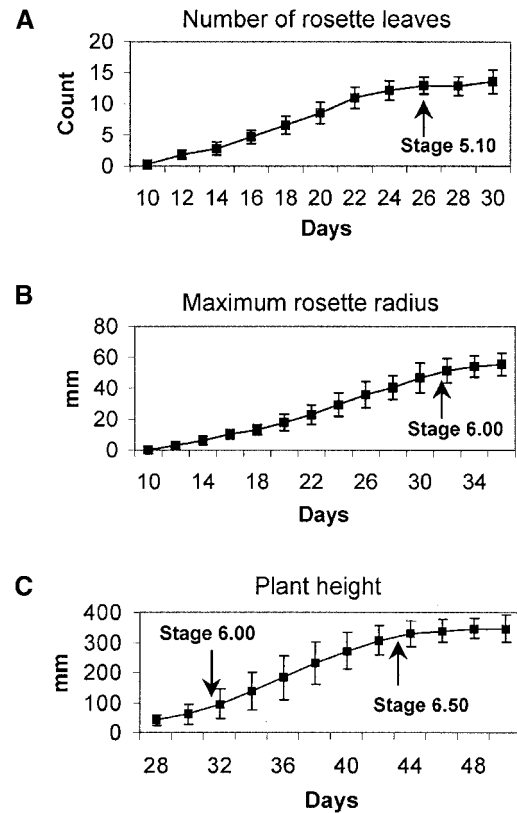

Figure 3. Representative Data from Wild-Type Col-0 Plants.

(A) Number of rosette leaves >1 mm in length produced over time.
(B) Maximum rosette radius (i.e., length of the longest rosette leaf) over time.
(C) Plant height over time.
Arrows indicate the time at which growth stages 5.10, 6.00, and 6.50 occur. Data are given as averages ±SD for >300 individual plants. Days are given relative to date of sowing, including a 3-day stratification at 4°C to synchronize seed germination.

root system is characterized by determining the final length of the primary root and the number of lateral roots. In addition, any abnormalities in root system morphology, color, or gravitropism are noted and recorded as selections from a series of descriptor choice lists. Also at the completion of the experiment, the number of rosette leaves is determined for a final time, and digital image analysis is performed to quantify total exposed leaf area, perimeter, size (i.e., major and minor axes), and shape (i.e., standard deviation of the radius and eccentricity). Abnormalities in rosette morphology or color are recorded as selections from descriptor choice lists.

Second-phase data collection in the soil-based phenotypic platform is triggered by growth stages 1.04, 1.10, 6.00,

6.50, 6.90, and 9.70. These growth stages represent key developmental points in the Arabidopsis life cycle and can be detected readily at the time they occur. Stages 1.04 and 1.10 represent early and late preflowering stages of growth, respectively, whereas stages 6.00, 6.50, and 6.90 span the period of flower production. Traits related to seed yield and

quality are assessed at stage 9.70. Table 4 lists the second-phase data collection steps triggered at each of these growth stages and provides representative data from the analysis of wild-type Col-0 plants. Of the 43 traits listed in Table 4, 15 have CVs that are <10%. Traits in this category include flower size, number of seed per silique, and seed size and shape characteristics. Twenty-one traits have CVs that are between 10 and 30%. These include size and shape estimates for rosettes and pollen grains, inflorescence structure characteristics (e.g., number of stem branches), and silique area and perimeter. Of the traits listed in Table 4, only seven exhibit CVs >30%. Notably, these include exposed rosette leaf area and dry weight taken at stages 1.10 and 6.0, total number of siliques, and mass of seed produced (yield). Thus, the total number of siliques produced by the plant (CV = 37.8%) plays a greater role in the variation associated with yield than does the number of seed per silique (CV = 9.4%). We have been unable to reduce the variation associated with yield through modified harvesting procedures designed to account for seed loss as a result of premature silique dehiscence.
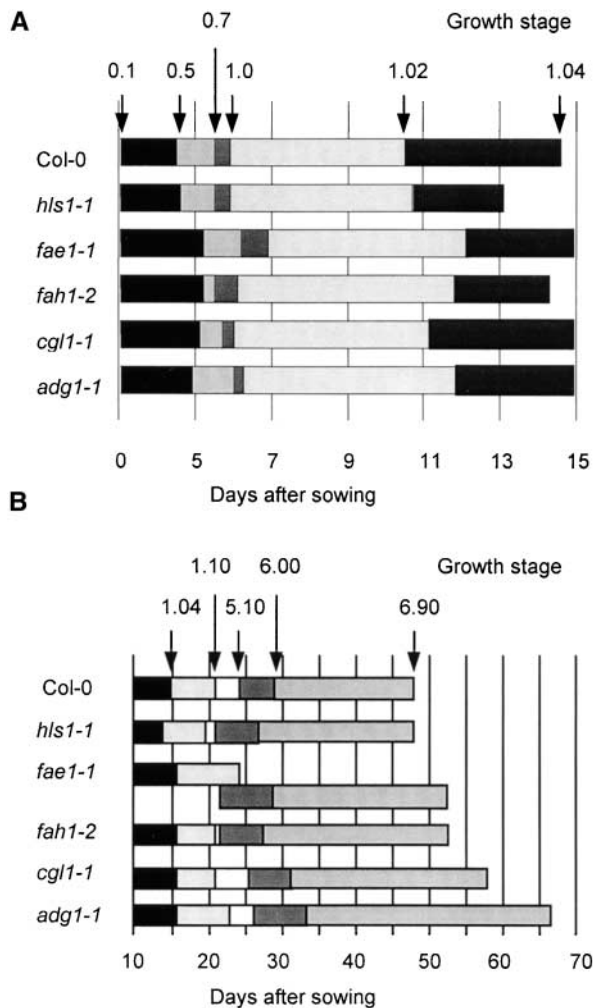
Validation of the Method

To determine whether the plate- and soil-based phenotypic analysis platforms captured subtle morphological differences, we analyzed mutant lines that were isolated previously by biochemical screening and had little to no reported morphological phenotypes. The mutants, obtained from the Arabidopsis Biological Resource Center, included *hls1-1*, which is deficient in ethylene production and apical hook formation when grown in the dark (Guzman and Ecker, 1990); *fae1-1*, which is deficient in long chain fatty acids in the seed (Lemieux et al., 1990); *fah1-2*, which is deficient in ferulate acid hydroxylase and affected in phenylpropanoid metabolism (Chapple et al., 1992); *cgl1-1*, which is deficient in complex glycan synthesis (von Schaewen et al., 1993); and *adg1-1*, which is deficient in an enzyme in starch biosynthesis (Lin et al., 1988).

In spite of the fact that four of these five mutants had no reported morphological phenotypes, each was demonstrated to have a developmental alteration relative to Col-0 wild-type plants. As shown in Figure 4, these differences are manifested in both the plate-based analysis, in which the time to reach the growth stages 0.5, 0.7, 1.0, 1.02, and 1.04 was examined (Figure 4A), and in the soil-based analysis, in which the time to reach growth stages 1.04, 1.10, 5.10, 6.00, and 6.90 was examined (Figure 4B). For example, the *hls1-1* mutant, known to be an early flowering mutant (Arabidopsis Biological Resource Center seed stock record), reached growth stages 1.04 and 1.10 before Col-0, indicating that it produced rosette leaves more rapidly than did Col-0 (Figures 4A, 4B, and 5A). Although *hls1-1* mutants flowered early (stages 5.10 and 6.00), they flowered for a longer period, completing flower production (stage 6.90) at
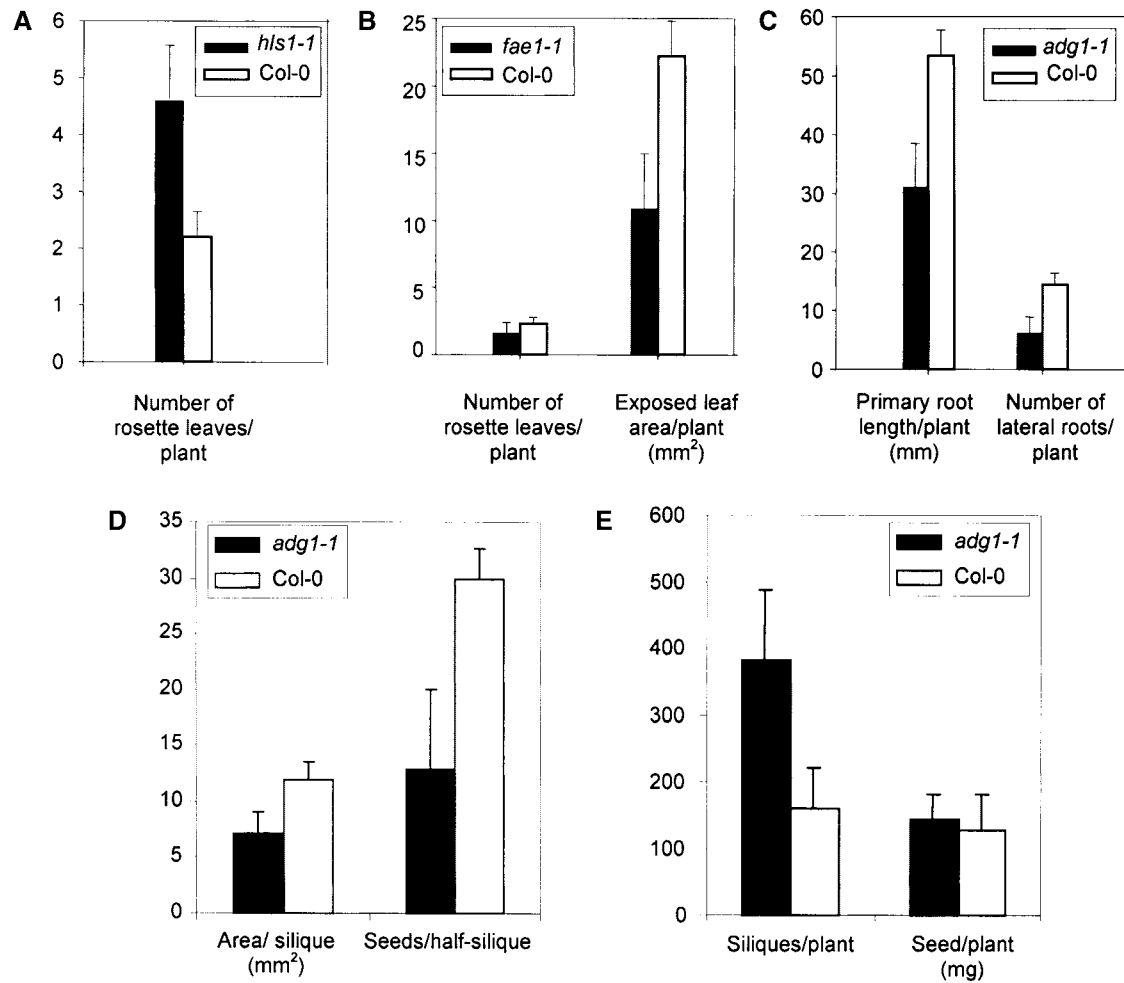

Figure 4. Growth Stage Progression for Wild-Type (Col-0) and Five Mutant Lines.

(A) Progression as determined in the plate-based early analysis platform.
(B) Progression as determined in the soil-based analysis platform. Arrows define the time (days after sowing) at which Col-0 plants reached the growth stages indicated. Boxes represent the time elapsed between the occurrence of successive growth stages. Junctions between boxes of different shading indicate the occurrence of a growth stage. In **(B)**, overlapping boxes for *fae1-1* indicate that growth stage 5.10 was reached before stage 1.10. Days are given relative to date of sowing, including a 3-day stratificaton at 4°C to synchronize seed germination.

Figure 5. Detection of Phenotypic Differences between Mutants and Wild-Type Col-0 Plants.

(A) Comparison of leaf initiation between *hls1-1* and Col-0. The number of leaves >1 mm were measured on day 14. Data are averages of 40 or >100 plants for *hls1-1* and Col-0, respectively.

(B) Comparison of shoot growth between *fae1-1* and Col-0. Total exposed leaf area was determined via a computerized analysis of digital images of intact rosettes (see Methods). Data are averages of 10 or >100 plants for *fae1-1* and Col-0, respectively.

(C) Comparison of root growth between *adg1-1* and Col-0. Both measurements were taken on day 14 (11 days after transfer to the growth chamber) and represent 20 seedlings grown on two plates.

(D) Comparison of silique area and the number of seed per half-silique between *adg1-1* and Col-0. Silique area was determined via computerized analysis of digital images of mature filled siliques (see Methods). The number of seed per half-silique was observed after removal of the outer layer of one valve of a mature filled silique. Both measures were averaged from three siliques per plant and are reported as the average of 10 or >300 plants for *adg1-1* and Col-0, respectively.

(E) Comparison of silique number and yield per plant between *adg1-1* and Col-0. The final number of siliques per plant was determined after the completion of flower production (stage 6.90). Yield is reported as the desiccated mass (mg) of seed produced per plant. Data are averages and standard deviations of 10 or >300 plants for *adg1-1* or Col-0, respectively.

(A) to **(C)** were obtained in the plate-based assay, and **(D)** and **(E)** were obtained in the soil-based assay.

Error bars indicate ±SD.

approximately the same time as did wild-type Col-0 plants (Figure 3B).

No vegetative phenotypes have been ascribed to the *fae1-1* mutation, and biochemical and mRNA expression studies have suggested that the activity of the *FAE1* gene is restricted to the seed (James et al., 1995). Surprisingly, our analysis revealed that *fae1-1* plants make the transition to flowering (stage 5.10) sooner than Col-0 plants do. In addition, *fae1-1* plants reached stage 5.10 before stage 1.10, indicating that *fae1-1* plants flower before producing 10 rosette leaves. *fae1-1* plants germinate more slowly (Figure 4A) and produce leaves more slowly than does the wild type when they are grown on soil (Figure 4B). The number of leaves produced by *fae1-1* and the wild-type Col-0 plants at the end of the plate-based assay was similar, although we detected a difference in total exposed leaf area (Figure 5B). Thus, *fae1-1* plants, although appearing visibly normal, are profoundly altered in many stages of development. *fah1-2*, *cgl1-1*, and *adg1-1* all exhibited subtle variations in their developmental progression, including somewhat delayed germination and leaf development and increased flowering period. In the case of *fah1-2*, initiation of flowering was early (Figure 4).

Our analysis of these mutants also revealed striking morphological differences compared with wild-type plants, an example of which is shown in Figure 5. Siliques produced by *adg1-1* mutants were smaller and contained fewer seed than those produced by wild-type plants (Figure 5D). It was expected, therefore, that the overall yield of the *adg1-1* plants would be reduced relative to the wild type. However, *adg1-1* plants produced a greater number of siliques than did Col-0 plants, resulting in a yield that was indistinguishable from that of Col-0 (Figure 5E). Thus, the phenotypic analysis method described here can be used to determine the contributions of different genetic loci to complex traits such as yield. Additional morphological variation also could be observed readily in the early analysis platform. Analysis of root growth, evaluated by the length of the primary roots and the degree of branching, revealed that the *adg1-1* mutant had an overall smaller root size than did the wild type (Figure 5C).

DISCUSSION

Historically, biological experiments have been designed to address a targeted research question. However, the nature of genome-scale research, coupled with the rapid development of databases to store the resulting data, now makes it possible for the same data to be used by scientists with completely different research objectives. Because the data in these databases often are integrated from many different sources, their ultimate utility is dependent on the uniformity of the methods used for its collection. In this light, we have developed a standardized process for the collection of phe-

notypic data that is built on a series of growth stage definitions. The growth stages serve both as developmental landmarks and as triggers for the collection of detailed morphological data. The process is generic and can be applied to any organism for which growth stages have been defined. Using this method, we have established a data set representative of wild-type Col-0 plants grown under standardized environmental conditions and validated our ability to detect novel phenotypes through the characterization of single-gene mutations (Figures 4 and 5). The same process can be used for the detection of phenotypic alterations resulting from more divergent genetic backgrounds (e.g., different ecotypes) as well as those resulting from the introduction of biotic or abiotic stress.

It is clear that not every phenotype can be represented adequately using the quantitative measurements presented here (Tables 3 and 4). As an example, qualitative characteristics such as altered leaf phyllotaxy and abnormal stem thickening are not represented directly in the quantitative data set. These data can be captured in images that are tagged with key word descriptors and stored for later reference.

Our analysis includes a plate-based platform to allow the detection of early phenotypes that may not be possible to observe in soil-based phenotypic analyses or at later stages of development. For example, some gene mutations or certain environmental stress conditions may severely inhibit radicle and hypocotyl emergence and/or elongation. Such seedling-lethal phenotypes are difficult to observe in a soil-based assay and are more readily characterized on artificial medium on plates. In addition, the early development analysis is rapid, requires minimal laboratory and growth space (we were able to grow 5000 plants per growth chamber occupying 1 m^2 of laboratory space), and provides a highly efficient means for examining root morphology. The results presented here demonstrate clearly that the early development analysis platform is capable of identifying mutations in root and shoot morphology and of detecting alterations in seedling growth rate and developmental progression similar to those observed in the soil-based platform.

We have emphasized the application of this method to the collection of morphological data. However, the same approach is applicable to the collection of tissue for subsequent molecular or biochemical analysis. It is likely that the regulation of the expression of nearly every gene and biochemical pathway is under some level of developmental control. Therefore, tissue harvests destined to supply material for metabolic or gene expression profiling will be more useful if the tissues are collected from plants at similar developmental stages rather than from plants of similar chronological age. Indeed, the analysis of the data collected in this manner may allow the identification of complex growth stage–specific "phenotypic fingerprints" consisting of morphological, biochemical, and molecular genetic markers. Analyzing the effect that mutations or environmental conditions have on these fingerprints will lead to a more rapid and complete understanding of gene function.

Integration of data from the plate- and soil-based analysis platforms into a single database enables the discovery of morphological and/or developmental mutations that occur at any growth stage during the life cycle of the plant. This unified data set, particularly when coupled with gene expression profiling and metabolic profiling data produced from tissues harvested at the corresponding growth stages, provides the opportunity to perform multivariate statistical analyses and positions the investigator to identify correlative relationships among highly diverse traits at different stages of growth. Thus, it may be possible to identify easily measured and highly reproducible traits that can serve as surrogates for more difficult to measure or more variable traits, such as yield. Ultimately, these surrogate traits may be tested for validity in crop species and developed as transgenic products or as tools to aid in conventional breeding programs.

The data collection process presented here is amenable to high throughput implementation. The limited amount of effort required for first-phase data collection facilitates the rapid determination of growth stages. Focusing on a subset of growth stages for the second-phase data collection reduces the overall effort while maintaining the ability to detect alterations in many different traits. Staggered sowings can be used to prevent all of the plants in a study from reaching growth stages for detailed analysis simultaneously, thereby allowing the analysis of many samples in parallel. In addition, the standardized rules used in growth stage determination can readily be modeled in a computer application, enabling the creation of an interface that can be used to streamline the data collection process.

The growth stages and data collection methodology presented here can serve as a powerful means to unify the collection of phenotypic data. Reporting the growth stage during which data were obtained can provide an explicit developmental context for comparative purposes and enhance the value of data for future investigations.

METHODS

All mutant lines of *Arabidopsis thaliana* were obtained from the Arabidopsis Biological Resource Center (Columbus, OH). For early analysis, seed were surface-sterilized with 70% ethanol for 5 min followed by 30% Clorox for 10 min and rinsed with sterile deionized water. Surface-sterilized seed were sown onto square Petri plates (10 × 10 cm) containing 40 mL of sterile medium consisting of 0.5 × Murashige and Skoog (1962) salts (Life Technologies) and 1% (w/v) phytagel (Sigma). The medium contained no supplemental sucrose. Plates were arranged vertically in plastic racks and placed in a cold room for 3 days at 4°C to synchronize germination. Racks with cold-stratified seed were then transferred into growth chambers (Percival, Perry, IA) with day and night temperatures of 22 and 20°C, respectively. The average light intensity at the level of the rosette was maintained at ~110 μmol·m^{-2}·sec^{-1} supplied by 2-foot fluorescent tubes (model SP41; General Electric, Fairfield, CT) during a 16-hr light cycle. Daily measurements were performed to assess seedling devel-

opment beginning at removal from the cold room (day 3 after sowing) until the seedlings were harvested on day 14 (or when >50% of the seedlings had primary roots ≥6 cm in length) for rosette image analysis. To control for position-dependent variation in the plate-based assay, the plates were relocated to a new position in the growth chamber after each measurement cycle.

For soil-based analysis, plants were grown in Metro-Mix 200 (Scott's Sierra Horticultural Products, Marysville, OH) in individual 2.5-inch pots. Pots were arrayed in a 4 × 8 grid in standard greenhouse flats (Kord Products, Lugoff, SC). Flats were grown on wire racks (Metro), each consisting of three 2- × 4-foot shelves. Each shelf provided space for four flats that were illuminated by 4-foot fluorescent tubes (model SP41; General Electric). Relative humidity was maintained at 60 to 70%. Plants were watered by subirrigation as needed, usually every 2 to 3 days, depending on growth stage. Day length was 16 hr, and daytime and nighttime temperatures were maintained at 22 and 20°C, respectively. The average light intensity at the top of the pots was ~175 μmol·m^{-2}·sec^{-1}. All measurements were as described in the figure legends and tables. Digital image analysis to obtain area, perimeter, standard deviation of the radius, major axis, minor axis, and eccentricity was performed using IP Lab software version 3.5 (Scanalytics, Fairfax, VA). An analysis of Columbia (Col-0) control data did not reveal substantial position-dependent bias in the growth rooms used for the soil-based assay. Therefore, the locations of the flats used in this study were held constant throughout the experiment.

Wild-type Col-0 data are presented as the average and standard deviation for a sample size of >300 individual plants. These plants were sown and analyzed in subgroups of 10 to 32 individuals each at periodic intervals during the course of 4 months. In all cases, the standard error among subgroups was similar to the standard deviation observed in the population as a whole. This finding supports the reproducibility of the assay over time.

REFERENCES

Chapple, C.C.S., Vogt, T., Ellis, B.E., and Somerville, C.R. (1992). An *Arabidopsis* mutant defective in the general phenylpropanoid pathway. Plant Cell **4,** 1413–1424.

Guzman, P., and Ecker, J.R. (1990). Exploiting the triple response of *Arabidopsis* to identify ethylene-related mutants. Plant Cell **2,** 513–523.

Hartenstein, V. (1993). Atlas of *Drosophila* Morphology and Development. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

James, D.W., Lim, E., Keller, J., Plooy, I., Ralston, E., and Dooner, H.K. (1995). Directed tagging of the *Arabidopsis FATTY ACID ELONGATION (FAE1)* gene with the maize transposon activator. Plant Cell **7,** 309–319.

Krysan, P.J., Young, J.C., and Sussman, M.R. (1999). T-DNA as an insertional mutagen. Plant Cell **11,** 2283–2290.

Lancashire, P.D., Bleiholder, H., van der Boom, T., Langeluddeke, P., Stauss, R., Weber, E., and Witzenberger, A. (1991). A uniform decimal code for growth stages of crops and weeds. Ann. Appl. Biol. **119,** 561–601.

Lemieux, B., Miquel, M., Somerville, C.R., and Browse, J. (1990). Mutants of *Arabidopsis* with alterations in seed lipid fatty acid composition. Theor. Appl. Genet. **80,** 234–240.

Lin, T.-P., Caspar, T., Somerville, C., and Preiss, J. (1988). Isolation and characterization of a starchless mutant of *Arabidopsis thaliana* (L.) Heynh lacking ADP glucose pyrophosphorylase activity. Plant Physiol. **86,** 1131–1135.

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue culture. Physiol. Plant. **15,** 473–497.

Smyth, D.R., Bowman, J.L., and Meyerowitz, E.M. (1990). Early flower development in *Arabidopsis*. Plant Cell **2**, 755–767.

von Schaewen, A., Sturm, A., O'Neil, J., and Chrispeels, M.J. (1993). Isolation of a mutant *Arabidopsis* plant that lacks *N*-acetyl glucosaminyl transferase I and is unable to synthesize Golgi-modified complex N-linked glycans. Plant Physiol. **102,** 1109–1118.

Wilkins, A.S. (1993). Genetic Analysis of Animal Development. (New York: Wiley-Liss).