

A Sequence-Based Map of Arabidopsis Genes with Mutant Phenotypes^{1[w]}

David W. Meinke*, Laura K. Meinke², Thomas C. Showalter³, Anna M. Schissel⁴, Lukas A. Mueller, and Iris Tzafrir

Department of Botany, Oklahoma State University, Stillwater, Oklahoma 74078 (D.W.M., L.K.M., T.C.S., A.M.S., I.T.); and Department of Plant Biology, Carnegie Institution, 260 Panama Street, Stanford, California 94305 (L.A.M.)

The classical genetic map of Arabidopsis contains 462 genes with mutant phenotypes. Chromosomal locations of these genes have been determined over the past 25 years based on recombination frequencies with visible and molecular markers. The most recent update of the classical map was published in a special genome issue of *Science* that dealt with Arabidopsis (D.W. Meinke, J.M. Cherry, C. Dean, S.D. Rounsley, M. Koornneef [1998] *Science* 282: 662–682). We present here a comprehensive list and sequence-based map of 620 cloned genes with mutant phenotypes. This map documents for the first time the exact locations of large numbers of Arabidopsis genes that give a phenotype when disrupted by mutation. Such a community-based physical map should have broad applications in Arabidopsis research and should serve as a replacement for the classical genetic map in the future. Assembling a comprehensive list of genes with a loss-of-function phenotype will also focus attention on essential genes that are not functionally redundant and ultimately contribute to the identification of the minimal gene set required to make a flowering plant.

Before the advent of modern genomics, the existence of a gene was often first revealed when a mutant with a visible phenotype was recovered. From pea (*Pisum sativum*) plants with wrinkled seeds to fruitflies (*Drosophila melanogaster*) with altered eye pigmentation, mutants have long played a central role in genetic analysis. Recent advances in molecular biology have made it possible to identify large numbers of genes with mutant phenotypes in model organisms and to move toward a synthesis of classical genetics and structural genomics. This report provides one such synthesis for a model plant. The sequence-based map of genes with mutant phenotypes described here should provide a foundation for the long-term goal of determining which genes in Arabidopsis give a phenotype when disrupted by mutation. This information is needed to address from a genetic perspective the question of functional redundancy in Arabidopsis and to identify those genes capable of generating phenotypic diversity.

The first comprehensive map of Arabidopsis genes with mutant phenotypes was published 20 years ago

(Koornneef et al., 1983). Included on that map were 76 genes with phenotypes ranging from altered trichome morphology and seed coat pigmentation to reduced surface waxes and increased hypocotyl length. Most of the genes were assigned map locations based on analysis of F₂ plants produced following self-pollination of heterozygotes. Backcrosses of heterozygotes to parental homozygotes were avoided to minimize the total number of crosses performed. Precise gene orders were often not resolved because two-point crosses were involved, which meant that genes were placed on the map by comparing recombination frequencies obtained with different pairs of linked markers.

Large numbers of mutants were added to the classical map over the next 15 years, including embryo defectives with a seed phenotype that enabled heterozygous F₂ plants to be identified without progeny testing. This feature reduced the number of plants required to obtain accurate mapping data and resulted in further map enhancements (Patton et al., 1991). The most extensive study presented recombination data for 169 embryo-defective mutants and estimated locations of 110 *EMB* genes (Franzmann et al., 1995). Recombination percentages were converted into centiMorgans using the Kosambi (1944) mapping function. Maps with the most consistent gene order were constructed by determining the minimal chi-square for all recombination data combined (Jensen and Jorgensen, 1975). Several computer programs were used to facilitate map construction and integration (Patton et al., 1991; Stam, 1993). Despite these important advances, inconsistencies in recombination data soon made it necessary to place new

¹ This research was supported by the National Science Foundation Developmental Mechanisms and Arabidopsis 2010 Programs.

² Present address: Macalester College, St. Paul, MN 55105.

³ Present address: Swarthmore College, Swarthmore, PA 19081.

⁴ Present address: College of Osteopathic Medicine, Kirksville, MO 63501.

^[w]The online version of this article contains Web-only data. The supplemental material is available at www.plantphysiol.org.

* Corresponding author; e-mail Meinke@okstate.edu; fax 405-744-7074.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.014134.

genes on the map by hand. This approach was used to construct the map of 284 loci published by Koornneef (1994).

Genetic maps of molecular markers were also being constructed during this time. The types of markers involved quickly expanded to include restriction fragment length polymorphisms (Chang et al., 1988; Nam et al., 1989; Fabri and Schaffner, 1994), random-amplified polymorphic DNAs (Reiter et al., 1992), cleaved-amplified polymorphic sequences (Konieczny and Ausubel, 1993), simple sequence length polymorphisms (Bell and Ecker, 1994), and amplified fragment-length polymorphisms (Alonso-Blanco et al., 1998). A recombinant inbred map developed by crossing Columbia and Landsberg ecotypes (Lister and Dean, 1993) soon became established as the standard for genetic placement of molecular markers. Mutant genes could be placed on this map by determining recombination percentages with linked molecular markers. This resulted in the establishment of two parallel types of genetic maps with mutant genes: the classical map and the recombinant inbred map. Integration of these maps proved difficult because chromosome lengths and recombination estimates were not identical. Because many people were interested in gene isolation through map-based cloning, genes with mutant phenotypes were often mapped only in relation to linked molecular markers. Furthermore, tagged mutants identified from T-DNA insertion lines did not require mapping to clone the disrupted gene and therefore often did not contribute to map enhancements. As a result, the number of mutant genes added to the classical map began to diminish.

The most recent update to the classical genetic map contains 462 loci distributed over five chromosomes and 469 total centiMorgans (Meinke et al., 1998). This map includes 131 genes with a seed phenotype and 110 genes initially placed on the recombinant inbred map and then transferred to the classical map after adjusting for differences in estimated chromosome lengths. Problems with this combined map soon became apparent as more genes were cloned and their relative locations on the physical map contradicted their estimated locations on the genetic map. Resolving these inconsistencies was difficult and was not given a high priority, despite the importance of genetic maps in model systems. A fresh approach was therefore needed to update and correct the classical map.

We decided to address this problem by focusing on mutants disrupted in known genes that could be assigned a physical location on the sequenced chromosomes. Our goal was to construct a sequence-based map of genes with mutant phenotypes to replace the classical genetic map and to provide a new standard for dealing with the chromosomal locations of mutant genes. We started with an initial list of 115 genes already noted as cloned on the genetic map

(Meinke et al., 1998) and supplemented this with information from The Arabidopsis Information Resource (TAIR; <http://www.Arabidopsis.org>), extensive literature searches of publications listed in PubMed (<http://www.ncbi.nlm.nih.gov>), examination of abstracts from posters at recent Arabidopsis meetings, and requests for community additions and corrections to draft spreadsheets. The initial results of this effort are described here with the first sequence-based map of 620 mutant genes of Arabidopsis. Additional details collected during construction of this map document the methods used for gene isolation, the general phenotype of mutant alleles, the predicted function of protein products, and the dramatic increase in the number of genes identified in recent years. Although the total number of genes that can mutate to give a phenotype remains to be determined, the results presented here make it possible to begin comparisons with other model systems and to assess from a genetic perspective the extent of functional redundancy in the Arabidopsis genome.

RESULTS AND DISCUSSION

Criteria for Including Genes on the Map

Three questions had to be resolved before determining which genes should be included on a sequence-based map: (a) what constitutes a mutant phenotype; (b) should the map be limited to loss-of-function mutants; and (c) what level of confidence of gene identification should be required? We decided to include any gene with a dominant or recessive mutant phenotype that could be detected through visual inspection, cellular characterization, or biochemical analysis under standard greenhouse or specialized laboratory conditions. In this way, it was possible to include a broad range of mutants and to facilitate comparisons between the classical genetic and sequence-based maps. Dominant mutants for which a loss-of-function phenotype remains to be identified will need to be removed from future lists that are limited to genes with essential functions. We did not consider a change in gene expression pattern or metabolic profile alone sufficient to constitute a phenotype because such alterations may turn out to be characteristic of most gene disruptions in Arabidopsis. Including these genes might therefore overshadow more established mutants and result in a map filled with genes with subtle loss-of-function defects. Also excluded were genes that gave a phenotype only when inactivated by antisense or gene silencing or when overexpressed through activation tagging or introduction of a cloned wild-type allele. The reason was once again to focus on loci defined by mutation and not by experimental gene manipulation in order to present an updated map that was similar in scope to the classical genetic map but consistent with the genome sequence.

Suppressor and enhancer mutations were more problematic because it was difficult to establish consistent guidelines for what to exclude. We decided not to include cases where a double knockout in redundant genes was required to give a phenotype because these alterations could not be attributed to a single locus. But genes with mutant phenotypes detected only in specific ecotypes or genotypes were included provided the genetic lines involved did not appear to have a mutation in a redundant gene. Enhancers and suppressors that gave no phenotype by themselves but modified the phenotype of a non-redundant gene knockout were therefore included. Because we did not require that gene identities be confirmed through molecular complementation or sequencing of duplicate mutant alleles, the possibility exists that some genes listed here may later need to be removed. Genetic maps have also required consolidation when two mapped loci were later found to be allelic. We discovered at least nine such cases in the course of updating the classical map (*CBB3* and *DWF3*; *DET2* and *DWF6*; *FUS4* and *FUS8*; *DOC1* and *TIR3*; *PAS3* and *GK*; *ELL* and *FK*; *AGR* and *EIR*; *RPP11* and *RPP13*; *RPP4* and *RPP5*).

A Sequence-Based Map of Mutant Genes

The assembled list of 620 cloned genes with mutant phenotypes is presented in Figure 1. An expanded spreadsheet with full gene names, alias symbols, gene classes, mutant phenotypes, predicted functions, and reference labs is provided in Table S-I (supplemental data can be viewed at <http://www.plantphysiol.org>). Genes listed in Figure 1 are arranged by locus number, a unique identifier that corresponds to the physical location of a predicted gene along the length of the chromosome (Arabidopsis Genome Initiative, 2000). The gene order presented here is therefore consistent with the published genome sequence. Because adjacent genes are often assigned locus numbers that differ in value by 10, regardless of gene size or intergenic distances, the proximity of two genes can be estimated by comparing their locus numbers. The precise number of intervening genes and base pairs can then be obtained from current annotation (<http://www.Arabidopsis.org>). Locations of mutant genes are displayed on a sequence-based physical map in Figure 2. Enlarged versions that include gene symbols and locus numbers are given in Figures S-1 and S-2 (supplemental data can be viewed at <http://www.plantphysiol.org>).

Chromosomal Distribution of Mutant Genes

Three approaches were used to examine the chromosomal distribution of mutant genes. The first was to compare the relative numbers of genes on each chromosome. The results as shown in Table I are similar to those found with the classical genetic map

and the sequenced genome as a whole. Genes with an embryo-defective (*emb*) phenotype are the most common class included on both types of maps. This is consistent with the large number of genes known to have essential functions during seed development (McElver et al., 2001).

The second approach was to look for large gaps or clusters on the sequence-based map. As shown in Figure 2, mutant genes are distributed throughout the length of each chromosome with the notable exception of centromeric regions. We considered two alternative explanations for these gaps: (a) genes with mutant phenotypes might be preferentially excluded from centromeric regions; or (b) the absence of mutant genes in these regions might simply reflect the known scarcity of functional genes around the centromere. We attempted to distinguish between these models by looking at loci predicted to fall between the mutant genes located just above and below each centromere constriction (e.g. *At1g37130* and *At1g43170* for chromosome 1). The genetically defined centromere is positioned within these gaps for chromosomes 1, 2, 3, and 5 and extends somewhat north of the boundary defined by *At4g04890* and *At4g04770* on chromosome 4 (Copenhaver et al., 1999). Of the nearly 2,700 sequenced loci assigned to these gaps in the current genome annotation (<http://www.tigr.org>), approximately 50% appear to be pseudogenes, transposons, or repeat elements, 20% are annotated as encoding hypothetical proteins with no database matches, and another 20% correspond to putative proteins. Fewer than 300 loci in these combined regions appear to be promising candidates for active genes with defined functions. Based on the observed frequency of genes with mutant phenotypes elsewhere in the genome ($620/29,084 = 2.1\%$), six of these 300 genes should have already been found to give a mutant phenotype. The failure to identify such mutants, if confirmed in future studies, could reflect an overestimation of functional genes in the centromere, inaccessibility of these regions to traditional mutagens, or increased levels of functional redundancy. For the most part, however, the distribution of genes with mutant phenotypes in Arabidopsis mirrors the distribution of functional, protein-coding genes throughout the genome.

A final approach used to look at gene distribution was to determine how often genes with mutant phenotypes were positioned next to each other on the chromosome. In 95 cases of 620 examined, two mutant genes are separated by five genes or fewer based on current annotation: 21 gene pairs are physically adjacent, 15 are separated by a single gene, 23 by two genes, 10 by either three or five genes, and 16 by four genes. We identified one adjacent pair (*At4g03050* and *At4g03060*) that appears to represent a tandem duplication involving similar gene functions, and another cluster of three adjacent genes (*At1g08540*,

Ch. 1	Symbol	cM	At1g21970	LEC 1	29	At1g69180	CRC	100	At2g26300	GPA 1		At2g46340	SPA 1	70
At1g01040	SUS 1	0	At1g22090	EMB 2204		At1g70170	MMP		At2g26330	ER	48	At2g46370	FIN 219	
At1g01120	KCS 1		At1g22710	SUC 2		At1g70940	PIN 3		At2g26510	PDE 135		At2g46410	CPC	63*
At1g01510	AN	0	At1g22770	GI	33	At1g71190	TTN 4		At2g26550	HO 2		At2g46590	DAG 2	
At1g01860	PFC 1		At1g22780	PFL	32	At1g71440	PFI		At2g26650	AKT 1		At2g46720	HIC	
At1g02090	FUS 5	5*	At1g22940	TH 1	33	At1g72970	ACE	102*	At2g26670	HY 1	48	At2g46830	CCA 1	
At1g02200	CER 1	1	At1g23420	INO		At1g73590	PIN 1	103	At2g26830	EMB 1187		At2g47430	CKI 1	
At1g02280	PPI 1		At1g24490	ARTEMIS		At1g74310	HOT*		At2g26930	PDE 277		At2g47450	CAO	
At1g02340	HFR 1		At1g25220	TRP 4	35	At1g74330	SID 2		At2g26990	FUS 12	50	At2g48120	PAC	
At1g02580	MEA	6	At1g25490	RCN 1		At1g74960	FAB 1		At2g27100	SE	49	Ch. 3	Symbol	cM
At1g02780	EMB 2386		At1g26310	CAL	46	At1g75350	EMB 2184		At2g27150	AAO 3		At3g01120	MTO 1	4
At1g04110	SDD 1		At1g27450	APT 1	39	At1g75820	CLV 1	110	At2g27170	TTN 7		At3g01610	EMB 1354	
At1g04120	MRP 5		At1g28300	LEC 2		At1g75950	ASK 1		At2g27250	CLV 3	47	At3g02150	PTF 1	
At1g04240	SHY 2		At1g28490	OSM 1		At1g76060	EMB 1793		At2g28000	SLP		At3g02180	SPIL	
At1g04250	AXR 3	11	At1g30950	UFO	46	At1g77120	ADH 1	114	At2g28290	SYD	52	At3g02260	TIR 3	4
At1g04400	FHA	12	At1g31480	SGR 2		At1g77760	NIA 1	116	At2g28610	PRS		At3g02580	DWF 7	
At1g04550	BDL		At1g32900	PRPL 11		At1g77860	KOM		At2g28800	ALB 3		At3g02850	SKO	
At1g04640	EMB 1687		At1g34790	TT 1	55	At1g78000	SEL 1	108*	At2g28880	EMB 1997		At3g02875	ILR 1	2*
At1g05180	AXR 1	12	At1g37130	NIA 2	56	At1g78580	TPS*		At2g29980	FAD 3	58	At3g03050	KJK	
At1g05750	PDE 247			CEN		At1g78630	EMB 1473		At2g30410	KIS		At3g04400	EMB 2171	
At1g05760	RTM 1		At1g43170	EMB 2207		At1g79460	GA 2	119	At2g30510	RPT 2	46*	At3g04870	PDE 181	
At1g05850	ELP		At1g43710	EMB 1075		At1g79560	EMB 1047		At2g30950	VAR 2	75	At3g07040	RPM 1	8*
At1g06150	EMB 1444		At1g43850	SEU		At1g79810	TED 3		At2g31260	APG 9*		At3g07060	EMB 1974	
At1g06570	PDS 1	26	At1g44446	CH 1	58	At1g79840	GL 2	118	At2g31340	EMB 1381		At3g08040	FRD 3	
At1g08060	MOM		At1g48380	RHL 1		At1g79850	ORE 4	117*	At2g31530	EMB 2289		At3g08550	KOB 1	
At1g08090	NRT 2		At1g48410	AGO 1		At1g80070	SUS 2	116	At2g31870	TEJ		At3g09090	DEX 1	
At1g08260	EMB 2284		At1g49040	SCD 1		At1g80080	TMM		At2g31970	RAD 50*		At3g09150	HY 2	12
At1g08520	PDE 166		At1g49400	EMB 1129		At1g80260	EMB 1427		At2g32950	COP 1	60	At3g10050	OMR 1	17
At1g08540	ABC 1		At1g49510	EMB 1273		At1g80350	FRA 2		At2g33150	PED 1		At3g11170	FAD 7	8
At1g08550	NPQ 1		At1g49720	ABF 1		Ch. 2	Symbol	cM	At2g33860	ETT		At3g11540	SPY	12
At1g08560	KN		At1g50030	TOR*			NOR		At2g34650	PID	62	At3g11670	DGD 1	14*
At1g08720	EDR 1		At1g50430	DWF 5	66	At2g01420	PIN 4		At2g34690	ACD 11		At3g11940	AML 1	
At1g09530	POC 1		At1g51760	IAR 3	68*	At2g01830	CRE 1		At2g34780	EMB 1611		At3g11980	MS 2	27
At1g09540	MYB 61		At1g52340	ABA 2	70	At2g01980	SOS 1		At2g35630	MOR 1		At3g12120	FAD 2	20
At1g09570	FHY 2	16	At1g54040	TASTY		At2g02480	STI	8	At2g35670	FIS 2		At3g12680	HUA 1	17*
At1g09700	HYL 1		At1g55540	EMB 1011		At2g03150	EMB 1579		At2g36270	ABI 5	63	At3g13550	COP 10	15*
At1g10760	SEX 1	13	At1g55610	BRL 1		At2g03220	MUR 2	9*	At2g36530	LOS 2		At3g13870	RHD 3	
At1g11680	EMB 1738		At1g55900	EMB 1860		At2g03680	SPR 1*	8*	At2g37260	TTG 2		At3g14110	FLU	
At1g12040	LXR 1		At1g56075	LOS 1		At2g04030	EMB 1956		At2g37630	AS 1	64	At3g14225	EMB 1474	
At1g12110	CHL 1	14	At1g56580	ASB 3		At2g05990	MOD 1*		At2g37680	PAT 3		At3g15170	CUC 1	
At1g12220	RPS 5	10*	At1g58250	SAB	88*	At2g06050	OPR 3		At2g37860	SOZ 2		At3g15620	UVR 3	
At1g12360	KEU	17	At1g60170	EMB 1220			CEN		At2g37920	EMB 1513		At3g15850	FAD 5	16
At1g12370	UVR 2		At1g61720	BAN		At2g13540	ABH 1		At2g38020	VCL 1		At3g17390	MTO 3	
At1g12770	EMB 1586		At1g62340	ALE 1		At2g15790	SQN		At2g38050	DET 2	64	At3g18110	EMB 1270	
At1g12840	DET 3	12	At1g62360	STM*	75	At2g17310	SON 1		At2g38120	AUX 1	66	At3g18290	EMB 2454	
At1g13120	EMB 1745		At1g63700	YDA	82*	At2g17950	WUS		At2g38440	ASB 1		At3g18390	EMB 1865	
At1g13870	DRL 1	15*	At1g64280	NIM 1	89*	At2g18020	EMB 2296		At2g39290	PGP		At3g18730	TSK	
At1g13980	EMB 30	21	At1g64970	TMT 1		At2g18390	TTN 5		At2g39770	CYT 1	65	At3g18990	VRN 1	17*
At1g14350	FLP		At1g65380	CLV 2	89	At2g18510	EMB 2444		At2g39810	HOS 1	66	At3g19820	DWF 1	26
At1g14610	TWN 2	19*	At1g65470	FAS 1	88	At2g18790	PHYB	35	At2g39940	COI 1		At3g20070	TTN 9	
At1g14750	SDS		At1g65480	FT	88	At2g19450	TAG 1	28	At2g40080	ELF 4		At3g20440	EMB 2729	
At1g14920	GAI	22	At1g65620	AS 2	85	At2g19520	FVE	32	At2g40220	ABI 4	66	At3g20600	NDR 1	
At1g15550	GA 4	22	At1g66340	ETR 1	93	At2g19760	PRF 1		At2g40890	REF 8		At3g20630	TTN 6	
At1g15950	IRX 4		At1g66350	RGL 1		At2g20000	HBT		At2g42620	ORE 9	57*	At3g20740	FIE	
At1g16410	SPS		At1g66520	PDE 194		At2g20610	SUR 1	35	At2g43010	SRL 2		At3g20770	EIN 3	
At1g16540	ABA 3	10	At1g67370	ASY 1		At2g21710	EMB 2219		At2g43360	BIO 2	67	At3g20780	BIN 3	
At1g16970	KU 70		At1g67440	EMB 1688		At2g22540	SVP		At2g43410	FPA	67	At3g21640	UCU 2	
At1g17560	HLL		At1g67490	KNF		At2g22990	SNG 1		At2g43710	SSI 2		At3g22780	TSO 1	
At1g18370	HIK		At1g68050	FKF 1		At2g23380	CLF	30*	At2g44490	PEN 2		At3g22880	DMC 1	
At1g1850	MP	22	At1g68100	LAR 1	101*	At2g25170	PKL		At2g44810	DAD 1	60*	At3g23050	AXR 2	33
At1g20050	HYD 1		At1g68370	ARG 1	97	At2g25600	SPIK		At2g45190	FIL	55*	At3g23130	SUP	35
At1g20200	EMB 2719		At1g68530	CER 6	96	At2g25920	ELF 3	47	At2g45330	EMB 1067		At3g23150	ETR 2	
At1g20960	EMB 1507		At1g68560	XYL 1		At2g26060	EMB 1345		At2g45440	DHDPS 2		At3g23430	PHO 1	27
At1g21310	RSH		At1g68640	PAN		At2g26250	FDH		At2g45690	SSE		At3g23560	ALF 5	36

Figure 1. (Figure continues on facing page.)

At1g08550, and At1g08560) with different functions and mutant phenotypes. One interesting gene (*SIN1/SUS1/CAF/DCL1*; At1g01040) with an essential role in growth and development (Golden et al., 2002) was found at the extreme north end of chromosome 1. The presence of adjacent pairs of mutant genes is particularly intriguing in light of the frequent occur-

rence of tandem gene duplications and associated redundancy in Arabidopsis. Whether these clusters define small regions of the genome with unusual structural features or limited functional redundancy remains to be determined. We have nevertheless demonstrated that genes with mutant phenotypes in Arabidopsis are not always surrounded by dispens-

At3g24560	RSY 3		At4g03280	PGR 1		At4g31780	MGD 1		At5g15840	CO	13	At5g49820	EMB 1879		
At3g24650	ABI 3	38	At4g04770	LAF 6		At4g31990	IAT 3	52*	At5g16560	KAN		At5g49930	EMB 1441		
At3g24800	PRT 1	38	At4g04890	PDF 2		At4g32410	RSW 1		At5g16620	PDE 120		At5g50280	EMB 1006		
At3g25140	QUA 1			CEN		At4g32540	YUC		At5g16910	IXR 1		At5g51240	EMF 2	71*	
At3g26790	FUS 3	41	At4g08150	BP	15	At4g32551	LUG		At5g17420	IRX 3		At5g51330	SWI 1		
At3g26830	PAD 3	44	At4g08810	SUB 1*		At4g33030	SQD 1		At5g17690	TFL 2	15*	At5g51700	PBS 2	x	
At3g27920	GL 1	46	At4g08920	HY 4	10	At4g33430	BAK 1		At5g17990	TRP 1		At5g51820	PGM	63	
At3g28030	UVH 3		At4g09820	TT 8	22	At4g34460	AGB 1		At5g18170	GDH 1		At5g53950	CUC 2		
	CEN		At4g10180	DET 1	20	At4g34620	SSR 16	60	At5g18580	FS 1	23	At5g54510	DFL 1		
At3g43210	TES	57	At4g10310	SOS 3		At4g35900	FD	64	At5g19220	ADG 2		At5g54640	RAT 5		
At3g44310	NIT 1		At4g11260	EDM 1		At4g36380	ROT 3		At5g19530	ACL 5		At5g54770	TZ	77	
At3g44480	RPP 1	52	At4g12420	SKU 5		At4g36920	AP 2	68	At5g19550	AAT 2	20*	At5g54810	TRP 2		
At3g44880	ACD 1	68	At4g14110	COP 9	29	At4g36930	SPT	68	At5g20240	PI	21	At5g56930	EMB 1789		
At3g45780	NPH 1	61	At4g14210	PDE 226		At4g37000	ACD 2	68	At5g20730	NPH 4	28*	At5g57050	ABI 2		
At3g46530	RPP 13	64	At4g14550	SLR	30	At4g37200	HCF 164		At5g20930	TSL	20	At5g57090	AGR 1	80	
At3g46560	EMB 2474		At4g14750	FRC 3		At4g37580	HLS 1		At5g20990	CHL 6	22	At5g57180	CIA 2		
At3g47950	AHA 4		At4g15090	FAR 1		At4g37650	SHR	68	At5g22260	MS 1	23	At5g57360	ZTL		
At3g48090	EDS 1		At4g15210	RAM 1		At4g37750	ANT		At5g22370	EMB 1705		At5g57590	BIO 1	74	
At3g48110	EDD		At4g15560	CLA		At4g38160	PDE 191		At5g22640	EMB 1211		At5g57930	EMB 1629		
At3g48500	PDE 312		At4g15570	MAA 3		At4g38600	KAK		At5g230xx	MAM 1*		At5g58270	STARIK		
At3g48560	CSR 1	70	At4g16130	ARA 1	37	At4g38620	MYB 4		At5g23120	HCF 136		At5g60690	REV	87*	
At3g48930	EMB 1080		At4g16280	FCA	32	At4g39030	EDS 5		At5g23150	HUA 2		At5g60890	ATR		
At3g49170	EMB 2261		At4g16340	SPIKE 1		At4g39400	BRI 1	69*	At5g23570	SGS 3		At5g60910	AGL 8		
At3g49240	EMB 1796		At4g16420	PRZ 1		At4g39620	EMB 2453		At5g23880	EMB 1265		At5g60920	COB	81	
At3g49500	SGS 2		At4g16845	VRN 2	36*	At4g39850	PXA 1		At5g24314	PDE 225		At5g61150	VIP 4		
At3g50660	DWF 4		At4g16860	RPP 4	38	At4g39920	POR		At5g24400	EMB 2024		At5g61380	TOC 1*	79*	
At3g50870	MNP		At4g16890	BAL			Ch. 5	Symbol	cM	At5g24520	TTG 1	28	At5g61410	EMB 2728	
At3g51160	MUR 1	67*	At4g18460	SNI 1		At5g01220	SQD 2		At5g24860	FPF		At5g61460	MIM		
At3g51240	TT 6	96	At4g18480	CH 42	44	At5g02030	LSN		At5g25810	TNY	44*	At5g61850	LFY	82	
At3g52430	PAD 4		At4g18710	BIN 2		At5g02250	EMB 2730		At5g25900	GA 3	33	At5g61900	BON 1		
At3g52940	FK	77	At4g18780	IRX 1		At5g02310	CER 3	82	At5g26742	EMB 1138		At5g62310	IRE*		
At3g54010	PAS 1	72*	At4g18960	AG	42	At5g02820	BIN 5			CEN		At5g62410	TTN 3	82	
At3g54170	FIP 37*		At4g19690	IRT 1		At5g03280	EIN 2	1	At5g33320	PPT	39*	At5g62790	PDE 129		
At3g54220	SCR		At4g20090	EMB 1025		At5g03730	CTR 1	0	At5g35410	SOS 2		At5g62810	PED 2		
At3g54340	AP 3	81	At4g20380	LSD 1		At5g03800	EMB 1899		At5g35550	TT 2	44	At5g62990	EMB 1692		
At3g54350	EMB 1967		At4g20900	MS 5		At5g03840	TFL 1	2*	At5g35620	LSP 1		At5g63840	RSW 3		
At3g54640	TRP 3	87*	At4g20910	CRM 2		At5g03940	FFC		At5g35770	SAP		At5g63860	UVR 8	82	
At3g54670	TTN 8		At4g21190	EMB 1417		At5g04140	GLS 1		At5g35840	PHYC		At5g63980	FRY 1		
At3g54720	AMP 1		At4g21670	FRY 2		At5g04560	DME		At5g37850	SOS 4		At5g64330	NPH 3	88*	
At3g55120	TT 5	81	At4g21790	TOM 1		At5g04890	RTM 2		At5g39510	ZIG		At5g64630	FAS 2	56	
At3g55270	MKP 1		At4g22260	IM	48	At5g05170	CEV 1		At5g40160	EMB 506		At5g64740	IXR 2	86	
At3g59030	TT 12		At4g23100	RML 1	44*	At5g05490	SYN 1		At5g40280	ERA 1		At5g64860	DPE 1		
At3g60740	TTN 1	85	At4g23250	EMB 1290		At5g05690	CBB 3	15*	At5g40890	CLCA		At5g64930	CPR 5	86*	
At3g61110	ARS 27		At4g23640	TRH 1		At5g05730	TRP 5	11*	At5g41150	UVH 1		At5g65930	ZWI		
At3g61140	FUS 6	87	At4g24190	SHD		At5g07280	EMS 1		At5g41315	GL 3	53	At5g65940	CHY 1		
At3g61780	EMB 1703		At4g24220	VEP 1		At5g07990	TT 7	6	At5g41370	XPB 1		At5g66055	EMB 2036		
At3g61850	DAG 1		At4g24510	CER 2	52	At5g08170	EMB 1873		At5g41410	BEL 1	50	At5g66130	RAD 17*		
At3g62910	APG 3*		At4g24540	AGL 24		At5g09640	SNG 2		At5g42080	ADL 1A		At5g66350	SHI		
At3g62980	TIR 1	77*	At4g24620	PGI 1		At5g09810	ACT 7*		At5g42270	VAR 1	65	At5g66570	PSBO		
	Ch. 4	Symbol	cM			At5g09900	EMB 2107		At5g42650	AOS		At5g66750	DDM 1	97	
		NOR				At5g10140	FLC	8	At5g42800	TT 3	57	At5g67030	ABA 1	98	
At4g00100	PFL 2		At4g25530	FWA	54	At5g10410	HY 5	8	At5g42970	COP 8	64	At5g67590	FRO 1		
At4g00450	CRP		At4g25560	LAF 1		At5g11260	HY 5	8	At5g43470	RPP 8	58				
At4g00650	FRI	0	At4g26080	ABI 1	55	At5g11530	EMF 1	12*	At5g43810	ZLL	57				
At4g00730	ANL 2		At4g26090	RPS 2	55	At5g12130	PDE 149		At5g44070	CAD 1	57				
At4g01370	MPK 4		At4g26300	EMB 1027		At5g12840	EMB 2220		At5g44790	RAN 1					
At4g02060	PRL		At4g26500	EMB 1374		At5g13160	PBS 1		At5g45250	RPS 4					
At4g02510	PPI 2		At4g27330	SPL		At5g13480	FY	12	At5g45710	RHA 1					
At4g02560	LD		At4g27760	FEY		At5g13630	GUN 5	13	At5g46330	FLS 2					
At4g02570	AXR 6	7*	At4g28750	PSAE 1		At5g13710	SMT 1		At5g46800	BOU					
At4g02780	GA 1	5	At4g29010	AIM		At5g13930	TT 4	13	At5g46860	SGR 3					
At4g02980	ABP 1		At4g29060	EMB 2726		At5g13960	KYP		At5g48150	PAT 1					
At4g03050	AOP 3		At4g29840	MTO 2		At5g14250	FUS 11	22*	At5g48230	EMB 1276					
At4g03060	AOP 2		At4g30090	EMB 1353		At5g14750	WER 1	20*	At5g48300	ADG 1					
			At4g30950	FAD 6	54	At5g15140	DND 1	22*	At5g49720	KOR 1					
			At4g31500	SUR 2	50*										

Figure 1. An ordered list of 620 Arabidopsis genes with mutant phenotypes. Genes are arranged by locus number, starting with the top of chromosome 1. Sequences are available at TIGR (<http://www.tigr.org>) and TAIR (<http://www.Arabidopsis.org>). Gene symbols with an asterisk conflict with other registered symbols that correspond to different genes (see Table S-1). Numbers in the centiMorgan column represent estimated locations on the genetic map. Numbers with an asterisk designate genes placed on the recombinant inbred map and later transferred to the classical map. Sequence gaps are noted for the centromere (CEN) and nucleolar organizer (NOR) regions.

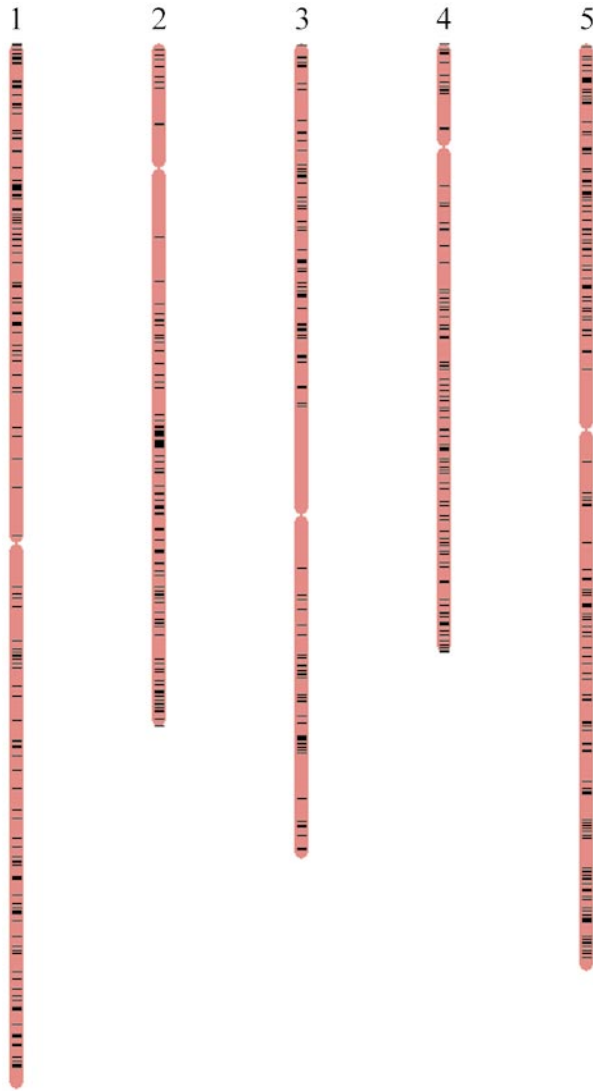


Figure 2. A sequence-based map of genes with mutant phenotypes. Gene locations are marked with horizontal lines. A single line at this scale may represent two or more neighboring genes. The length of each chromosome is proportional to its sequence. Centromeric gaps are marked by short constrictions.

able genes with little relevance to growth and development.

Time and Method of Gene Isolation

There has been a dramatic rise in recent years in the number of mutant genes cloned and characterized at a molecular level. This trend, documented in Figure 3, reflects improvements in methods for map-based cloning, widespread availability of transposon and T-DNA insertion lines, and longstanding efforts of the Arabidopsis community to characterize mutants obtained through forward genetics. At least 45 of the 76 loci (59%) found on the original classical map (Koornneef et al., 1983) and 237 of the 462 loci (51%) on the updated map (Meinke et al., 1998) have now

been cloned. One-half of the 235 mapped genes that remain to be identified are embryo defectives, many of which were given a low priority for gene isolation because they were not tagged (Franzmann et al., 1995). A current version of the classical map with cloned genes highlighted is presented in Figure S-3. Methods used to identify these genes are summarized in Figure 4A. Within the next several years, the number of genes identified through reverse genetics is likely to increase sharply as more emphasis is placed on screening insertion lines for knockouts in specific genes of interest. The rate-limiting step for characterizing genes with mutant phenotypes will then shift from trying to isolate the disrupted gene to searching for a subtle or conditional phenotype.

Phenotypes of Disrupted Genes

Mutant phenotypes should be associated with individual alleles rather than a single gene because weak and strong alleles may produce different types of abnormalities. We nevertheless decided to place the 620 mutant genes described here into broad phenotypic classes based on known gene disruptions in order to assess the diversity of genes and mutants included. The results of this subjective but informative effort are shown in Figure 4B. The relative frequency of each phenotype class reflects not only the number of genes available to be disrupted but also the amount of attention devoted to that class of mutants by members of the community. Dominant mutants have been placed in a separate category to highlight their inclusion and to acknowledge the absence of a knockout phenotype.

Comparison of Genetic and Sequence-Based Maps

Information presented in Figure 1 makes it possible for the first time to assess the physical accuracy of the classical genetic map on a global scale. Although gene positions based on recombination percentages are for the most part consistent with physical locations confirmed through genome sequencing, precise locations and orders of closely linked genes are sometimes incorrect. Many of these inaccuracies can be attributed to the subjective process of transferring genes from the recombinant inbred map to the classical genetic map. Similar problems were encountered in the past when genetic maps constructed under different conditions were combined (Stam, 1993). The level of inaccuracy within each type of map is nevertheless about the same. In approximately 75% of the cases where a gene from either the classical genetic or recombinant inbred map has been cloned and sequenced, the adjacent cloned gene from the same map was placed in the correct position. In other words, the centiMorgan values increase or stay the same going down the chromosome in 75% of the cases examined when genes from the two maps are

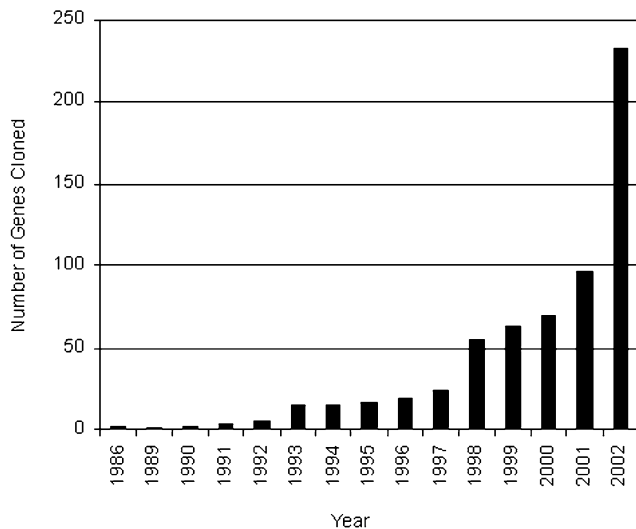


Figure 3. Date of initial publication or public release of gene identities associated with a mutant phenotype.

considered as separate groups. Some inconsistencies in gene placement are not surprising given the methods used to construct genetic maps. Other irregularities, such as the placement of *CER3* at the top of chromosome 5 (GenBank accession no. 1669654) despite extensive genetic evidence documenting its location at the bottom of chromosome 5 (Koornneef, 1994), are more problematic and remain to be resolved. We chose not to perform a genome-wide comparison of genetic and physical map distances because many of the genetic locations were estimated based on recombination percentages with different sets of distant markers and were therefore not an accurate reflection of genetic distance. Crosses between mutants listed here could be performed in the future to compare recombination percentages and physical distances on a global scale.

Diversity of Genes Identified

Mutant genes identified to date encode proteins with a wide range of biological functions. Characterizing the full spectrum of cellular processes involved will require a higher degree of saturation and functional characterization of the entire proteome using a standardized gene ontology (GO) system adopted for model eukaryotes (Gene Ontology Consortium, 2001). At the present time, 35% of the 620 genes listed here have no GO assignment, 17% have a single functional assignment, and the remaining 48% have multiple GO assignments. Additional information on predicted protein functions can be found in Table S-1 and at TAIR.

Mutant genes also differ in predicted length from start to stop codons. We reasoned that existing mutant collections might be biased toward large genes because they represent bigger targets for random mutagenesis. This model is supported by the results

presented in Table II. Small genes (<1 kb in length) are underrepresented in our collection (4% versus 25% of total) whereas large genes (>3 kb in length) are more common (42% versus 16%). These differences are also reflected in the average gene size: 3.2 kb for mutant genes and 1.9 kb for the entire genome. Large genes have already been shown to be preferred targets for T-DNA insertions that result in a seed phenotype (McElver et al., 2001). Results presented here document a similar trend for the entire collection of mutant genes identified through forward genetic screens. A significant challenge for the future is therefore to isolate and characterize large numbers of mutants disrupted in small genes with important biological functions.

Estimating the Total Number of Mutant Genes

Despite the limitations of defining what constitutes a mutant phenotype, the question of how many *Arabidopsis* genes will be found to exhibit a phenotype when disrupted by mutation needs to be addressed to place the current study in perspective and to com-

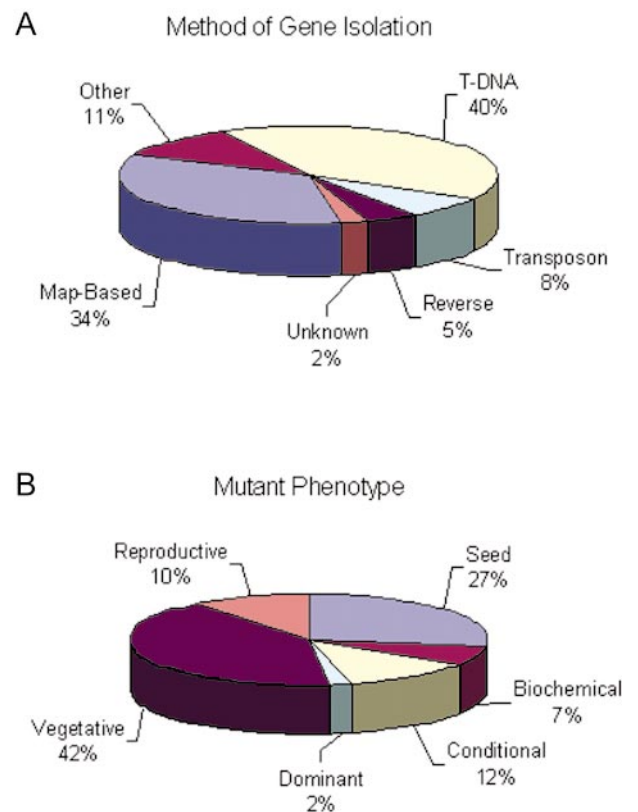


Figure 4. Classification of mapped genes according to method of identification and phenotype of mutant alleles. A, Method used to determine the identity of each mutant gene included on the sequence-based map. "Other" includes cases where promising candidate genes with consistent functions and map locations were analyzed directly for altered nucleotide sequence or protein function. B, Phenotype class of a representative mutant allele for each gene identified. Refer to "Materials and Methods" for definitions.

Table I. Distribution of genes on the classical genetic and sequence-based maps of *Arabidopsis*

Map Feature	Classical Genetic	Sequence-Based	Total Genome ^a
Total loci	462 (100.0%)	620 (100.0%)	29,084 (100.0%)
Chromosome 1	122 (26.4%)	156 (25.2%)	7,378 (25.4%)
Chromosome 2	70 (15.2%)	106 (17.1%)	4,814 (16.6%)
Chromosome 3	80 (17.3%)	102 (16.4%)	5,987 (20.6%)
Chromosome 4	77 (16.7%)	103 (16.6%)	4,319 (14.8%)
Chromosome 5	113 (24.4%)	153 (24.7%)	6,586 (22.6%)
Total <i>EMB</i>	131 (28.4%)	140 (22.6%)	Est. 500 to 750 ^b

^aBased on current annotation of the *Arabidopsis* genome (TIGR Release 3.0; <http://www.tigr.org>).
^bBased on relative frequency of embryo-defective mutants and duplicate alleles (Franzmann et al., 1995; McElver et al., 2001).

pare *Arabidopsis* with other model systems. The most definitive approach would be to perform a comprehensive phenotypic screen of a complete collection of individual knockout lines. This level of saturation has been reported for the yeast *Saccharomyces cerevisiae* (Giaever et al., 2002) but at present remains impractical for *Arabidopsis*, where efficient methods of gene replacement are not available and emphasis has been placed instead on screening for knockouts in existing collections of insertion lines, which are random, redundant, and incomplete.

An alternative approach would be to take several small regions of the genome, produce knockouts for each predicted gene, perform a comprehensive screen for phenotypes on each knockout, and extrapolate the percentage of genes found to have mutant phenotypes to the genome as a whole. Such an effort is certainly feasible, and if representative regions of the genome were chosen, could provide a direct estimate of the anticipated number of target genes. A related approach would be to take results from other model systems, compensate for differences in functional redundancy, and come up with an adjusted estimate for *Arabidopsis*. Recent studies have revealed that approximately 14% of the 2,400 genes on chromosome 1 of *C. elegans* disrupted with RNAi exhibit a phenotype (Fraser et al., 2000) and that 19% of the 5,900 knockouts of yeast analyzed in a genome-wide survey exhibit a growth defect (Giaever et al., 2002). Approximately 35% of all predicted genes in *Arabidopsis* appear to be unique based on sequence comparisons in which similarity is defined by

BLASTP value ($e < 10^{-20}$) and alignment (>80% of the protein). The proportion increases to 55% for *C. elegans* and 71% for yeast (*Arabidopsis* Genome Initiative, 2000). Although disrupting a duplicated gene can still result in a phenotype, increased functional redundancy should reduce the likelihood of a phenotype overall. Given this level of redundancy and phenotype detection in different model systems, we estimate that 10% of *Arabidopsis* genes (about 3,000 total) should give a loss-of-function phenotype that can be identified using current screening methods. This estimate may increase considerably in the future as visual and biochemical analyses of knockout mutants become more robust.

Information collected from large-scale screens of existing T-DNA collections should also provide insights into the level of saturation achieved. This requires estimating the average number of inserts per line, the percentage of inserts that fall within a single gene, the frequency of mutant phenotypes observed in the entire collection, and the percentage of mutant phenotypes associated with T-DNA insertion. Much of this information has already been obtained, particularly in relation to embryo defectives. There are about 1.5 inserts per line on average (Feldmann, 1991; Krysan et al., 1999; McElver et al., 2001), 30% of the phenotypes observed result from stable T-DNA integration (Castle et al., 1993; McElver et al., 2001), and 35% of T-DNA insertions appear to fall within an open reading frame (Krysan et al., 2002). The percentage of T-DNA insertions that disrupt the function of a typical *Arabidopsis* gene, however, remains to be

Table II. Predicted sizes of cloned mutant genes identified through forward genetic screens^a

Gene Class Analyzed	Number	Percentage of Genes Identified in Different Size Classes ^b							
		<1 kb	1+ kb	2+ kb	3+ kb	4+ kb	5+ kb	6+ kb	>7 kb
Total genome	27,288	25.3	37.9	20.9	8.1	3.9	1.9	0.9	1.1
Total forward genetics	590	4.1	24.7	28.8	17.8	12.4	4.9	3.2	4.1
Map-based cloning	208	3.4	23.6	34.1	18.3	8.6	6.2	2.4	3.4
T-DNA tagged	254	4.7	24.0	23.7	17.3	14.6	5.5	5.5	4.7
Other forward ^c	128	3.9	28.1	30.5	18.0	14.1	1.5	0.0	3.9

^aBased on recent annotation from TIGR (<http://www.tigr.org>). Gene boundaries correspond to start and stop codons. Gene classes are defined by method of gene isolation. ^bGenes are placed in size groups: 1+ kb includes all predicted genes between 1.00 and 1.99 kb in length; 2+ kb includes those between 2.00 and 2.99 kb; etc. ^cIncludes mutant genes identified by transposon tagging and other combinations of methods.

resolved. There are an estimated 500 to 750 *EMB* genes based on the frequency of duplicate mutant alleles identified to date (Franzmann et al., 1995; McElver et al., 2001). If we estimate that embryo defectives represent about 20% of all mutant phenotypes in T-DNA populations, then the total number of genes that can give a mutant phenotype is 2,500 to 3,750. The current estimated level of saturation for mapped genes with mutant phenotypes is therefore 15% to 25%. Although these numbers may need to be adjusted dramatically as additional details emerge from forward and reverse genetic screens, we believe that they represent a reasonable starting point for future experiments and an important first step in the analysis of *Arabidopsis* genes with mutant phenotypes.

MATERIALS AND METHODS

Establishing a List of Genes with Mutant Phenotypes

Gene identities associated with mutant phenotypes were identified in part by searching PubMed (<http://www.ncbi.nlm.nih.gov>) for relevant publications using different combinations of keywords (*Arabidopsis*, gene, mutant, mutation, and protein). Abstracts of papers describing the initial cloning of a mutant gene were saved for future reference. Publications were examined for details on gene symbols, mutant phenotypes, predicted functions, and methods of gene isolation. Reference laboratories responsible for identifying the disrupted gene were noted. Chromosome locus numbers maintained at The Institute for Genomic Research (TIGR; <http://www.tigr.org>) were identified using BLASTP (Altschul et al., 1997) accessed through TAIR (Huala et al., 2001) to compare published sequence information with the entire *Arabidopsis* proteome. Additional genes were found by scanning abstracts of recent *Arabidopsis* meetings in Madison, Wisconsin (June, 2001) and Seville, Spain (June, 2002). Direct requests for information were made to the *Arabidopsis* community through TAIR and the electronic *Arabidopsis* newsgroup (arab-gen@net.bio.net). Symbols of mapped and well-characterized mutants not included on initial lists of cloned genes were also used to search PubMed and GenBank. Information presented here was obtained through August 15, 2002.

Classification of Mutant Phenotypes

Six phenotype classes were used to document the diversity of genes identified: seed (embryo- or endosperm-defective or seed pigment mutant), vegetative (altered germination, seedling, root, rosette, or transition to flowering), reproductive (abnormal flower, silique, seed coat, or gamete), biochemical (altered enzyme activity, product accumulation, or cellular function without other striking defects), conditional (phenotype only in certain genetic backgrounds or in response to pathogen or unusual treatment), and dominant (phenotype observed only with dominant allele). Genes with variable mutant phenotypes were assigned to the first relevant class in the order listed above. Assignments were designed to be informative and representative but could not always be definitive given the complexity of some mutant phenotypes.

Drawing the Sequence-Based Maps

Maps were drawn with the "Chromosome Map Tool" available at TAIR, which queries a database with the supplied locus numbers to obtain assignment information. The locus name and coordinate information are then sent to an applet, which draws all five chromosomes on the browser screen. The user can choose a zoom level to scale the picture. Figure 2 was drawn at the 100% zoom level, where one pixel on the screen equals 50 kb. Figures S-1 and S-2 were drawn at the 600% zoom level, where one pixel equals approximately 8 kb.

ACKNOWLEDGMENTS

We thank Tanya Berardini at TAIR for providing a draft list of putative mutants associated with gene identifiers, Brian Haas at TIGR for information on gene size distributions genome wide, members of the Meinke laboratory for helpful comments on the manuscript, and the entire *Arabidopsis* community for providing information on gene identities.

Received September 6, 2002; returned for revision October 9, 2002; accepted November 18, 2002.

LITERATURE CITED

- Alonso-Blanco C, Peeters AJ, Koornneef M, Lister C, Dean C, van den Bosch N, Pot J, Kuiper MT (1998) Development of an AFLP based linkage map of Ler, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a Ler/Cvi recombinant inbred line population. *Plant J* 14: 259–271
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402
- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815
- Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* 19: 137–144
- Castle LA, Errampalli D, Atherton TL, Franzmann LH, Yoon ES, Meinke DW (1993) Genetic and molecular characterization of embryonic mutants identified following seed transformation in *Arabidopsis*. *Mol Gen Genet* 241: 504–514
- Chang C, Bowman JL, DeJohn AW, Lander ES, Meyerowitz EM (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 85: 6856–6860
- Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD et al. (1999) Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* 286: 2468–2474
- Fabri CO, Schaffner AR (1994) An *Arabidopsis thaliana* RFLP mapping set to localize mutations to chromosomal regions. *Plant J* 5: 149–156
- Feldmann KA (1991) T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. *Plant J* 1: 71–82
- Franzmann LH, Yoon ES, Meinke DW (1995) Saturating the genetic map of *Arabidopsis thaliana* with embryonic mutations. *Plant J* 7: 291–300
- Fraser AG, Kamath RS, Zipperlin P, Martinez-Campos M, Sohrmann M, Ahringer J (2000) Functional genomic analysis of *C. elegans* chromosome 1 by systematic RNA interference. *Nature* 408: 325–330
- Gene Ontology Consortium (2001) Creating the gene ontology resource: design and implementation. *Genome Res* 11: 1425–1433
- Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B et al. (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418: 387–391
- Golden TA, Schauer SE, Lang JD, Pien S, Mushegian AR, Grossniklaus U, Meinke DW, Ray A (2002) *SHORT INTEGUMENTS1/SUSPENSOR1/CARPEL FACTORY*, a dicer homolog, is a maternal effect gene required for embryo development in *Arabidopsis*. *Plant Physiol* 130: 808–822
- Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, Reiser L, LaFond F, Hanley D, Kiphart D, Zhuang M, Huang W et al. (2001) The *Arabidopsis* Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res* 29: 102–105
- Jensen J, Jorgensen JH (1975) The barley chromosome 5 linkage map: I. Literature survey and map estimation procedure. *Hereditas* 80: 5–16
- Konieczny A, Ausubel FM (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant J* 4: 403–410
- Koornneef M (1994) *Arabidopsis* genetics. In E Meyerowitz, C Somerville, eds, *Arabidopsis*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 89–120
- Koornneef M, van Eden J, Hanhart CJ, Stam P, Braaksma FJ, Feenstra WJ (1983) Linkage map of *Arabidopsis thaliana*. *J Hered* 74: 265–272
- Kosambi DD (1944) The estimation of map distances from recombination values. *Ann Eugen* 12: 172–175
- Krysan PJ, Young JC, Jester PJ, Monson S, Copenhaver G, Preuss D, Sussman MR (2002) Characterization of T-DNA insertion sites in *Arabi-*

- dopsis thaliana* and the implications for saturation mutagenesis. OMICS 6: 163–174
- Krysan PJ, Young JC, Sussman MR** (1999) T-DNA as an insertional mutagen in *Arabidopsis*. Plant Cell 11: 2283–2290
- Lister C, Dean C** (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J 4: 745–750
- McElver J, Tzafrir I, Aux G, Rogers R, Ashby C, Smith K, Thomas C, Schetter A, Zhou Q, Cushman MA et al.** (2001) Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*. Genetics 159: 1751–1763
- Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M** (1998) *Arabidopsis thaliana*: a model plant for genome analysis. Science 282: 662–682
- Nam HG, Giraudat J, den Boer B, Moonan F, Loos WDB, Hauge BM, Goodman HM** (1989) Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. Plant Cell 1: 699–705
- Patton DA, Franzmann LH, Meinke DW** (1991) Mapping genes essential for embryo development in *Arabidopsis thaliana*. Mol Gen Genet 227: 337–347
- Reiter RS, Williams JGK, Feldmann KA, Rafalski JA, Tingey SV, Scolnik PA** (1992) Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc Natl Acad Sci USA 89: 1477–1481
- Stam P** (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. Plant J 3: 739–744