Arabidopsis thaliana as an Experimental Organism

Maarten Koornneef, *Wageningen Agricultural University, Wageningen, The Netherlands*
Ben Scheres, *Utrecht University, Utrecht, The Netherlands*

Arabidopsis thaliana has become the model for plant genetics and plant development because of its genetic characteristics, such as short generation time, the small size of its genome and the availability of its complete DNA sequence.

Introductory article

Article Contents

Introduction

Arabidopsis thaliana (L.) Heyhn. is a small weed plant belonging to the mustard family (Brassicaceae or Cruciferae). The species can be found in nature almost everywhere in the northern hemisphere in ruderal sites, such as sandy patches along roads etc. *Arabidopsis* has been found from sea level up to high in the Himalayas and from northern Scandinavia to North Africa, including the Cape Verde Islands at 16° latitude. It also grows in North America, probably following introduction from Europe. This might also have been the origin for plants found in Australia and maybe even in Japan.

Arabidopsis was first suggested as a suitable model for plant biological studies, and especially genetics, in the 1940s. The reasons for this were its small size, the ease with which it can be grown, its self-fertilizing habit and the short generation time of many accessions (isolates), which are often called ecotypes in *Arabidopsis*. In greenhouse or in climate chambers 6–8 weeks is sufficient time to complete the entire life cycle from germination until seed set. Furthermore, *Arabidopsis* has one of the smallest genomes among higher plants. These factors and the ability to transform the plant, have made it the favourite plant model for molecular genetic studies to date. The research advantages of *Arabidopsis* became clear in the early 1980s and led to an upsurge in the international research effort on it. The large research community provides additional advantages for research nowadays; important resources are available and their distribution is well organized. Stock centres in the UK and USA provide seed stocks of mutants and wild accessions and DNA materials for research. Information is provided by the Arabidopsis database (TAIR) (see Further Reading for details). The presence of the complete genomic DNA sequence by the end of the year 2000 is another unique research resource. Until recently, *Arabidopsis* was considered to be less suitable for cytogenetic studies. However, when the large pachytene chromosomes are studied, in combination with *in situ* hybridization, detailed cytogenetic analysis can be performed.

Description of the Development

Arabidopsis seeds are small (0.5 mm), oval-shaped and produced in large numbers (up to a few thousand per plant). Seeds can be germinated easily, although freshly harvested seeds may need a cold treatment of a few days and/or a period of storage to germinate fully, because they are dormant. The seeds usually require light for germination. The degree of dormancy can differ between genotypes. Seedlings are small, with two cotyledons at opposite positions. They do not contain the small unicellular but branched hairs or trichomes, which appear on the surface of the true leaves. The latter are present on a nonelongating stem and thereby form a rosette of leaves. The first true leaves have trichomes only on the adaxial (upper) side; subsequent leaves have them on both sides. In general, the leaves are oval-shaped, but variations exist between accessions in petiole morphology, the degree to which leaves are round or elongated and the serration of the leaf margins. The number of rosette leaves that are formed depends on the genotype and environmental conditions and is strongly correlated with the time from germination to bolting and flowering.

Flowering starts with a change in the shape of the shoot apical meristem from flat to more rounded. Instead of leaf primordia, the meristem starts producing floral primordia (**Figure 1**). Soon thereafter the main stem elongates (bolt), which results in an inflorescence with a main stem that carries a number of (cauline) leaves with a reduced number of trichomes on the adaxial sides and having axillary buds that develop into secondary inflorescences. Higher on the inflorescence stems the typical crucifer flowers arise with four whorls of floral organs. The first whorl has four sepals, the second one has four white petals, the third whorl has six stamens and the fourth whorl or centre has two carpels,

Figure 1 An approximately 4-week-old plant of the frequently used laboratory accession Landsberg erecta. Total plant height is at this stage 15 cm. The insert shows a single flower.

which are fused into the pistil. The genetic control of flowering and especially the formation of floral organs has been studied extensively and resulted in the so-called ABC model for flower formation.

Embryo and Seed Development

Embryo development (**Figure 2**) starts with the fusion of the egg cell present in the embryo sac of the polygonum type with a sperm cell (male gamete) deposited by the germinating pollen grain. Pollen develops from microspores within the two-lobed anther, which contains four locules surrounded by a tapetum layer. This does not differ

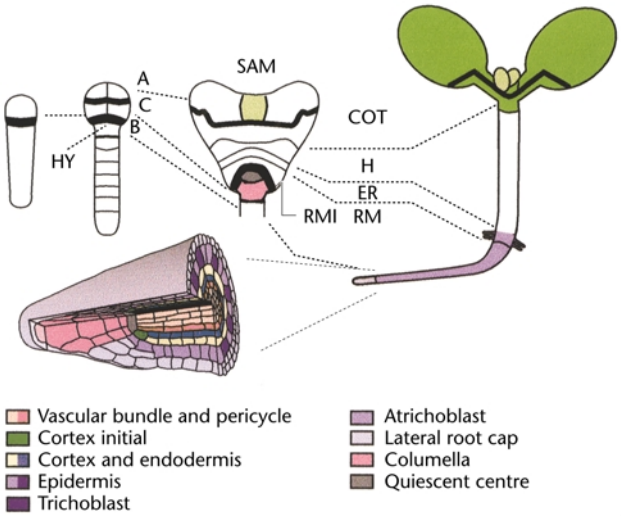

Figure 2 Establishment of the *Arabidopsis* body plan in the embryo and the structure of a primary root. A, apical region; C, central region; B, basal region; HY, hypophyseal cell; SAM, shoot apical meristem; COT, cotyledon; H, hypocotyl; ER, embryonic root; RM, root meristem; RMI, root meristem initials.

greatly from that in many other plant species. A number of male sterile mutants has been shown to be affected in anther and especially tapetum development. After fertilization the zygote follows a regular pattern of cell divisions, which correlate with morphologically defined stages. Fate mapping and cell ablation studies, however, show that throughout development these cell divisions do not cause the segregation of cell fates. Rather, cell identity is based on continuous positional information in the embryo and in the developing regions after embryogenesis, the meristems. The first division of the zygote results in an embryo with a small apical cell and a longer basal cell, from which the filamentous suspensor, supporting the embryo, will be formed. Specific stages that are distinguished thereafter are the octant stage (when the apical cell has given rise to two tiers of four cells), the dermatogen stage (when the protoderm is formed by periclinal cell divisions), the globular stage, the heart stage (when cotyledon primordia become visible) and the torpedo stage. At this stage, cell division is arrested and further growth, during the so-called bend cotyledon and walking stick stages, occurs mainly through cell expansion. Hereafter the seed-maturation phase starts. During this period the seeds further accumulate food reserves (lipids, sugars, proteins), develop desiccation tolerance and become dormant.

In the mature seed the embryo consists of two cotyledons derived from the upper tier, from which the shoot apical meristem (SAM) is also formed. Below this are present the cotyledon shoulders derived from the upper-lower tier and below these, the hypocotyl derived from the lower-lower tier from which the root meristem also originates. The tip of the root meristem contains the root cap and quiescent

centre which originate from the upper cell of the basal cell, called the hypophysis.

In addition to the embryo proper, the seed consists of a seed coat or testa and endosperm. The latter develops from the fertilization of the central cell, which contains two haploid nuclei, with the second male gamete and results in a triploid endosperm. This endosperm initially develops as a syncytium which thereafter undergoes cellularization. Later in seed development the endosperm dies, except the outermost layer, and is replaced by the growing embryo.

The testa consists of two layers derived from the outer and inner integument of the ovule. Early in seed development, the outer integument consists of two layers. Later on, the cells in the outer layer of this integument produce mucilage, which is secreted by the seeds when they imbibe water. The second layer of the outer integument, as well as the two outer layers of the inner integument, are compressed in mature seeds and the remnants of the cell walls are impregnated with flavonoid-derived polymers. The innermost layer of the inner integument, called endothelium, is clearly distinct in structure from the other layers but is also compressed at maturity and contains the brown tannin-like pigments that give *Arabidopsis* seeds their brown colour.

Upon germination the seedling grows and develops from its shoot apical and root meristems. The activities of these meristems shape the structure of the mature plant, and they are able to give rise to new structures throughout the lifespan of the plant.

The above-ground shoot apical meristem consists of a central zone with slowly dividing cells that replenish the cells leaving the neighbouring peripheral zone. In the peripheral zone, organ primordia arise. These primordia can give rise to leaves but also to modified leaf structures such as, for example, the different floral organs. In the axils of peripheral organs, new meristems are formed and the time at which these are activated is important in determining the architecture of the plant. Many genes have recently been identified that play a role in the continuous allocation of cells from the shoot apical meristem to newly formed organs.

Underneath the soil, the embryonic ('primary') root meristem gives rise to various root tissues and, at a distance behind the meristem to new meristems that will form lateral roots. The structure of the primary root in *Arabidopsis* seedlings is very regular and consists above the root cap of an outer layer of epidermis cells, which share a common precursor cell with the lateral root cap. A single cortex and a single endodermis each with eight cells derive from another precursor cell. Within the endodermis, a single layer of pericycle cells surrounds the vascular bundle. In longitudinal section these layers form long and regular cell files. Root hairs, involved in solute uptake, develop above the elongation zone as outgrowths of those epidermal cells that are located in the clefts between adjacent cortical cells. Lateral roots are initiated from cells in the pericycle layer and repeat the pattern of the primary roots, although the cell number in the various layers is more variable.

Root development has been studied extensively by careful microscopy, cell lineage and cell-ablation analysis and a large number of mutants defective in root development have been described and are now being analysed at the molecular level.

Life Cycle

Depending on genotype and conditions, flower primordia become visible as early as 2 weeks after germination of the seeds and fertilization can take place 3 weeks after germination in early genotypes. *Arabidopsis* produces progeny almost exclusively by self-pollination and the seeds develop from the zygote within the ovule. This process takes 2–3 weeks, resulting in a total minimum generation time of approximately 6 weeks. Although in laboratory conditions six generations can be obtained per year, *Arabidopsis* in nature probably produces only one generation per year. In Europe, most *Arabidopsis* can be seen flowering in spring and early summer. These plants might have germinated in spring (summer annuals) or during the previous fall (winter annuals). The latter are probably those genotypes that are late flowering in greenhouse conditions, but which can respond strongly to a vernalization treatment that induces flowering. In addition to flowering, the presence of seed dormancy prevents several generations occurring in one year. There are large genetic variations in nature for both flowering time and seed dormancy. However, the number of field observations on the ecology of *Arabidopsis* is still limited. Very little is known about the distribution of the many small seeds produced by *Arabidopsis*. In nature, most seeds probably remain in the vicinity of the mother plant.

Genetic Characteristics

The size of the *Arabidopsis* genome is estimated to be approximately 130 megabases, organized into five chromosomes. The estimated total gene number is approximately 25 000, which is close to the number for the worm *Caenorhabditis elegans*. The chromosomes are built up of mainly unique sequences with, on average, one gene per 5 kb. Repeat sequences are clustered around the centromeres and constitute the centromeric heterochromatin, observed as condensed regions in chromosome preparations. The location of the centromeres is metacentric in chromosomes 1, 3 and 5 and submetacentric in chromosomes 2 and 4. The latter two chromosomes contain large arrays of repeated ribosomal DNA (rDNA) genes at the end of their short arms. The telomeres consist of specific 7-

bp repeats (TTTAGGG) characteristic for many plant species.

Genetic and Physical Maps

The first consistent map of the five linkage groups was published in 1983 and was constructed on the basis of extensive linkage analysis with 76 morphological markers (mutants). Restriction fragment length polymorphism (RFLP) maps became available in 1988 and 1989. Subsequently these RFLP markers and additional markers based on polymerase chain reaction (PCR) technology were mapped in a set of recombinant inbred lines (RILs) derived from a cross between the two most commonly used laboratory strains, Landsberg erecta (L*er*) and Columbia (Col). For the mapping of newly identified mutations sets of polymorphic PCR markers such as simple sequence length polymorphism (SSLPs or microsatellites) and cleaved amplified polymorphic sequences (CAPSs) and so-called amplified fragment length polymorphism (AFLPTM) markers are available.

The many molecular markers that were mapped served as anchors to construct the physical map of *Arabidopsis* based on genomic DNA cloned in yeast and bacterial artificial chromosomes (YACs and BACs). At present, almost complete contigs are available for all five chromosomes. Clones that make up these contigs are used for the internationally coordinated sequencing of the complete genome of *Arabidopsis*. The complete sequence of chromosomes 2 and 4 was published in 1999 and the sequence of the three remaining chromosomes is expected to be available on the internet by the end of 2000.

In addition to the sequencing of genomic clones, many cDNAs have been partially sequenced and provide a large collection of ESTs (expressed sequence tags).

Gene Transfer (Transformation) in *Arabidopsis*

Arabidopsis can be transformed relatively easy using the bacterium *Agrobacterium tumefaciens* containing the Ti plasmid from which a specific region (T-DNA) is transferred to the plant chromosome. Modern *Agrobacterium* vectors, called binary vectors, have a separate plasmid containing the virulence genes (necessary for the transfer of T-DNA) and a second plasmid with the T-DNA itself (containing the genes to be transferred). Antibiotic- and herbicide-resistance genes are suitable selection markers, which allow the identification of transformed plants. Although initially tissue culture methods were used to transform *Arabidopsis*, these procedures have been replaced by the so-called in-planta or vacuum transformation method in most laboratories. To apply this procedure agrobacteria are infiltrated (under vacuum) into plants that have just started bolting. Apparently the bacteria can infect cells that will go on to form gametes. This will result in the formation of seeds that are transformed and which can be selected for at the seedling stage using the selection markers mentioned above.

Recovery of Mutations and Functional Analysis of Genes

As in other model species, the function of genes in *Arabidopsis* is deduced from mutant versions of the genes and the DNA sequences. Mutants are used to clone the respective genes (forward genetics) and to connect the function (altered in the mutant) with the protein encoded by the gene. With the availability of the almost complete sequence of genes, sequence data can also become the starting material for the analysis of gene function (reverse genetics). Searches are done for plants in which a particular gene is disrupted. Thereafter these plants are analysed for their phenotype. For this procedure DNA is isolated from a collection of plants with inserts introduced in plants by transformation or transposable elements. Pooling strategies are used to avoid the need to isolate DNA from thousands of plants. In these DNA preparations, specific DNA is amplified using PCR in which one primer is specific for the target gene and the other one is specific for the insert. Only those pools containing plants with an insert in that specific gene allow the amplification of a fragment. When the plant with the insert is identified it can be analysed for its phenotype.

Forward genetics requires the isolation of mutants. The self-fertilizing character of *Arabidopsis* means that the starting material for most mutagenesis experiments is homozygous and therefore recessive mutants are not observed in the first generation of mutagenized plants. This so-called M_1 generation derives from seeds treated with either classical mutagens such as irradiation and chemicals (e.g. ethylmethane sulfonate, EMS). Selfing of these M_1 plants produces the M_2 generation, in which recessive mutants (the majority of all mutants) will segregate. The selection of mutants depends on the type of mutants the study is interested in and can involve visible selection for aberrant phenotypes or the ability/inability to grow under specific selection conditions or even large-scale chemical or biochemical screens. Presently, insertion mutagenesis, where a gene is mutated by the insertion of DNA sequences, is frequently used because such mutants facilitate gene cloning. For this, plants transformed with *Agrobacterium tumefaciens* vectors (T-DNAs) or transposable elements are used. Transposable elements (TEs) are introduced only once into the host genome and mutate genes by their ability to move from one site in the genome

to another. Mutations generated by TEs often are unstable because the element may excise from the gene in which it was located, thereby restoring the function of the gene. Modified Ac/Ds or En/I (= Spm/dSpm) elements from maize have been introduced into *Arabidopsis* and were shown to be active. An active endogenous *Arabidopsis* transposon, called Tag1, has been found in some *Arabidopsis* genotypes but is less frequently used in tagging experiments than maize TEs.

In addition to induced mutants the genetic variation present in different *Arabidopsis* accessions can be analysed and will allow the identification of specific genes, as has been demonstrated for genes conferring disease resistance, flowering time, etc.

Gene Cloning in *Arabidopsis*

When it is demonstrated, using co-segregation analysis, that a specific T-DNA or TE causes a mutation, then part of the genomic DNA flanking the insertion can be isolated. This flanking DNA represents part of the mutated gene and can be used to clone the complete genome or complementary DNA (cDNA) sequence of that gene from genomic or cDNA libraries. The final proof that the gene has been cloned can be obtained by the demonstration of complementation of the mutant by the complete gene and/or by expression and sequence analysis of several mutants in that same gene. In the case of TEs, the DNA sequence of revertants will show specific target site duplications at the position where the original element was excised. This can also be used as proof that the gene for which the mutant was found has been isolated.

In situations where only mutants induced by classical mutagens are available, the respective genes can be isolated by map-based cloning. This procedure requires that the map position of the locus is determined in detail in relation to molecular markers that are related with the physical map. This map position then allows the identification of specific (YAC, BAC or P1) clones that should contain the gene. The confirmation that the right gene has been cloned can be obtained from the complementation, by transformation, of the mutant phenotype with specific subclones that contain only one specific open reading frame. Additional arguments come from the analysis of the expression of the candidate gene in mutant and wild-type phenotypes and from the sequence of mutant alleles. Irradiation-induced mutants are useful in this respect because this mutagen, more often than chemical mutagens, produces deletions, which are easier to detect in molecular analyses than single base pair changes.

Examples of Plant Research where the Genetic and Molecular Approaches using *Arabidopsis* were Important

The isolation of mutants and the subsequent cloning and study of the genes have led to important new insights in plant developmental biology. This includes the genetic dissection of the pathways of flowering, flower development, root development, trichome (hair) formation, etc. Interestingly, a number of developmental principles appear to be shared with the animal kingdom, although plants and animals evolved independently from different unicellular ancestors. Recently, evidence has surfaced that plants also deploy new developmental strategies; this is not surprising given the fact that plant cells do not move relative to one another and that they have to cope with environmental stresses during their continuous development.

In addition, *Arabidopsis* has been used for biochemical genetics. Pathways that have not been genetically studied before, such as photorespiration, cell wall synthesis, lipid synthesis, etc., have been analysed. Important contributions have also been made to the field of hormone research and have led to the isolation of many genes controlling the biosynthesis and modes of action of various plant hormones. A fascinating challenge that is now ahead is to combine our emerging understanding of plant hormones with that of the genes that control development. The signal transduction of ethylene has been analysed almost exclusively by using *Arabidopsis* mutants defective in ethylene action or those showing constitutive ethylene responses. Although some plant growth regulatory properties were known before, the essential role of brassinosteroids recently became fully appreciated when it was shown that some extreme dwarf mutants of *Arabidopsis* were brassinosteroid deficient. These mutants were subsequently used to clone the genes of the various biosynthetic steps.

Photomorphogenesis research has also benefited from the genetic approach. This has made it possible to analyse the functions of the various phytochrome types and also to clone the blue light photoreceptors called cryptochromes which mediate hypocotyl inhibition and flowering. Recently, it has been shown that mammals and *Drosophila* also possess cryptochromes which entrain the circadian clock in these animals, probably using a mechanism similar to that used by plants, for instance to perceive daylength signals.

Within a period of less than 20 years *Arabidopsis* has become the favourite model plant because of its amenability to genetics and molecular biology. It will undoubtedly remain the favourite model plant in future years because of the vast amount of basic knowledge that has accumulated. The availability of the complete DNA sequence does not only greatly facilitate gene cloning and

analysis, but it can also be used to construct microarrays or DNA chips with all genes of *Arabidopsis*. These can then be used to study the expression of these genes in various conditions, tissues, mutants, etc. This combination of advantages will not be easily surpassed by other plants and therefore the status of 'model plant' is likely to be appropriate for *Arabidopsis* for a long time, as is the case for *Drosophila melanogaster* and *Caenorhabditis elegans*, the flies and worms that continue to serve in unravelling many mysteries of development in the animal kingdom.

Further Reading

The Arabidopsis Information Resource (TAIR) (1997) [http://www.arabidopsis.org]

Martinez-Zapater JM and Salinas J (eds) (1998) *Arabidopsis Protocols*. Totowa, NJ: Humana Press.

Meinke DW, Cherry JM, Dean C, Rounsley SD and Koornneef M (1998) *Arabidopsis thaliana*: a model plant for genome analysis. *Science* **282**: 662–682.

Meyerowitz EM and Somerville CR (1994) *Arabidopsis*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Westhoff P, Jeske H, Jürgens G, Kloppstech K and Link G (1998) *Molecular Plant Development from Gene to Plant*. Oxford: Oxford University Press.

Wolpert L, Beddington R, Brockes J, Jessell T, Lawrence P and Meyerowitz E (1998) *Principles of Development*. Oxford: Current Biology Ltd, Oxford University Press.