ARABIDOPSIS THALIANA AS A MODEL FOR GENETIC ANALYSIS: A REVIEW

Sani, I*

S.B SADAU**

ABSTRACT

Arabidopsis thaliana is the most widely-studied plant today. Over 11 000 researchers and 4000 organizations around the world are generating a rich diversity and quantity of information and materials. The Flowering plant *Arabidopsis thaliana* is an important model system for identifying genes and determining their functions. Most biology students are familiar with thefruit fly (*Drosophila melanogaster*), adopted as a model system for genetic analysis byNobel laureate Thomas Hunt Morgan.*Arabidopsis* was not widely accepted as a model until relatively recently. *A. thaliana* was the first plant species for which a genome sequence became available. This initial sequence was from a single inbred strain (accession), and was of very high quality with each chromosome represented by merely two contigs, one for each arm. This review gives a comprehensive explanation of *Arabidopsis* as a model for genetic analysis.

* Department of Plant Biology,

** Department of Science and Technology Education Bayero University, Kano, NIGERIA.

INTRODUCTION

Many species of plants evolved on earth but few species are selected for scientific research. This is due to special characteristics features the plants posses. This species is called a model system. Information obtained from studying the model system can be applied to related organisms as needed. The choice of a representative species is typically driven by how easy it is to obtain, grow, and manipulate, and how well it has been studied. Most biology students are familiar with thefruit fly (*Drosophila melanogaster*), adopted as a model system for genetic analysis byNobel laureate Thomas Hunt Morgan. The plant species most commonly used for molecular genetics at present is the small member of the mustard family (Brassicaceae) *Arabidopsis thaliana* (fig. 1). Unlike the fruit fly adopted as a model atthe beginning of the twentieth century, *Arabidopsis*was not widely accepted as a model until relatively recently. *A. thaliana*, a tiny weed whose common name is "mouse ear cress," grows low to the ground and produces clusters of small white flowers in meadows and laboratories around the globe [15].

.

A. thaliana was the first plant species for which a genome sequence became available. This initial sequence was from a single inbred strain (accession), and was of very high quality with each chromosome represented by merely two contigs, one for each arm [17]. In addition to functional analyses, the 120Mb reference sequence of the Columbia (Col-0) accession proved to be a boon for evolutionary and ecological genetics. A particular advantage in this respect is that the species is mostly self-fertilizing, and most strains collected from the wild are homozygous throughout the genome. This distinguishes *A. thaliana* from other model organisms such as the mouse or the fruit fly. In these systems, inbred strains have been derived, but they do not represent any individuals actually found in nature[17].

FIG 1: *A. thaliana*in the palm ofa hand[17].

The ascendancy of *A. thaliana* to become one of the most popular species in basic plant research [2], despite its lack of economic value, is due to the favorable genetics of this plant. It has a diploid genome of only about 125 to 150Mbdistributed over five chromosomes, with fewer than 30,000 protein-coding genes. The ease with which it can be stably transformed is unsurpassed by any other multicellular organism [16]. Moreover, as flowering plants only appeared about 100 million years ago, they are all relatively closely related. Indeed, key aspects of plant physiology such as flowering are highly conserved between economically important grasses such as rice and *A. thaliana* [6].

Origin:

A. thaliana is thought to have originated in Central Asia and spread from there throughout Eurasia. During the last glaciations, *A. thaliana* was confined to the southern limit of its range and after the ice retreated, much of Europe was recolonized by different populations, resulting in complex ad mixture patterns. Today, *A. thaliana* occurs throughout the Northern Hemisphere, mostly in temperate regions, from the mountains of North Africa to the Arctic Circle. Like many other European plants, it has also invaded North America, most probably during historic times [14].

Taxonomic Hierarchy

Kingdom	Plantae
Subkingdom	Viridiplantae
Infrakingdom	Streptophyta
Superdivision	Embryophyta

Division	Tracheophyta
Subdivision	Spermatophytina
Class	Magnoliopsida
Superorder	Rosanae
Order	Brassicales
Family	Brassicaceae
Genus	Arabidopsis
Species	*Arabidopsis thaliana* (L.)

Source: NODC Taxonomic Code, database (version 8.0)

Botanical description:

Like higher plants, the adult *A. thaliana* is composed of only a few relatively simple organs and tissues (**Fig. 2**). The aboveground shoot connected to the belowground rootsystem represents the plant's main axis. Both root and shoot contain several cell types organized into the outer epidermal layer; the middle ground or cortical, layer; and the inner vascular cylinder (which contains the water-conducting xylem vessels and the food-conducting phloem elements). Attached to the shoot are mature leaves also composed of three layers: an epidermal layer, containing both hair cells, or trichomes, and small pores, or stomata (singular, *stoma*) that regulate gas exchange; a mesophyll layer responsible for photosynthesis; and vascular tissues that conduct water and nutrients between the leaf and the stem[7].

FIG.2: Body anatomy of *A.thaliana*

Leaf: Theleaves form a rosette at the base of the plant. The basal leaves are green to slightly purplish 1.5-5cm long and 2–10cm broad and are covered with trichomes.

Flower: The*Arabidopsis*flower is the most complex set of organs in the plant**.**The flower consists of a modified stem with four concentricregions, or whorls,of modified leaves. The first whorl (whorl 1) consists of four green leaf like sepals;whorl 2 is composed of four white petalsthat are leaf like in shape but contain no photosynthetic cells; whorl3 is made of six stamens (four long and two short) bearing the male gametophytes in the form of pollen; and whorl 4is a cylinder composed of two fused carpels housing the female gametophytes in the form of embryo sacs enclosed with ovules. The gametophytesrepresent the haploid phase of the plant's life cycle. (They develop by mitotic cell divisions after meiosis and give rise to the "alternation of generations" and so on). The fused carpels are part of a cylinder known as the pistilthat consists of pollen-receptive stigma at the top and a short neck, or style,leading to the ovary,which houses roughly 50 gamete-bearing ovules[7].

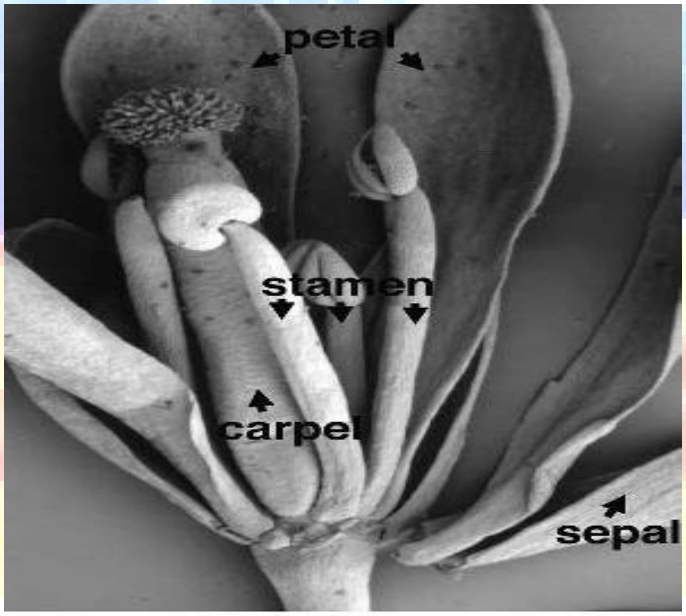


FIG 3: Showing flower parts of *A. thaliana*

Fruit: The fruit is siliqua, 5-20mm long containing 20-30 seeds**.**

Shoot and root: Cells at the growing points of both shoot and root are organized into a tissue called the apical meristem**,** a group of undifferentiated cells that divide continuously. The root apical meristem, which consists of a central zone of meristem cells and a peripheral zone of

differentiated cells, produces the cylindrical root through cell division, cell expansion, and cell differentiation. The roots are simple, later produce small lateral roots (this forms interaction with rhizosphere bacteria e.g. *Bacillus megaterium*). The shoot apical meristem, which consists of a central zone where cell divisions also maintain the meristem and a peripheral zone where cell divisions give rise to mature organs such as leaves and lateral shoots, generates the entire aboveground plant[6].

Plant development:

*A.thaliana*undergoes both vegetative and reproductive development. Vegetative development involves the development of non-flowering parts of the plant. During the period before flowering- the shoot apical meristem initiates a spiral ofleaves in precise sequence. When the plant switches to reproductive development,the shoot apical meristem becomes an inflorescence meristem, which, in turn, produces a series of lateral floral meristems. Each floral meristem produces a floral primordium that gives rise to a flower. The inflorescence is the flower or group of flowers at the tip of a branch. In *Arabidopsis*, the flowers are arranged in a spiral around the inflorescence stalk, as depicted in **Fig. 3**[6].

Life cycle:

Arabidopsis, like a hermaphrodite animal, carries male and female gametes; it is thus capable of self-fertilization, although cross-fertilization by artificial means is easy to accomplish. Fertilization is the first step in a life cycle that also includes embryonic development, seed germination and vegetative growth, reproductive development, and senescence (**Fig. 4**). In *Arabidopsis*, each mature pollen grain contains two coupled haploid (1*n*) sperm cells plus a 1*n* vegetative cell, while each embryo sac includes six mononucleate cells, each with one 1*n* nucleus and a central cell containing two1*n* nuclei, all encased within a common cell wall. The landing of a pollen grain on the receptive tissue of the stigma initiates fertilization by triggering germination of the pollen grain. The emerging pollen tube migrates through the short neck of the pistil, also known as the *transmitting tract*, to an ovule in the ovary. There, one 1*n* sperm nucleus from the pollen fuses with a 1*n* egg nucleus from the embryo sac to form the 2*n*nucleus of the zygote; the second 1*n* sperm nucleus fuses with the two 1*n* nuclei within the embryo sac to form a 3*n* endosperm nucleus. After fertilization, the zygote divides mitotically to form the embryo, and the endosperm nucleus divides mitotically to form endosperm tissue that, like an animal yolk sac,

will nourish the developing embryo. Meanwhile, the remaining 1*n* cells of the ovule degenerate, and the outer layer of the ovule hardens to form a seed coat[6].

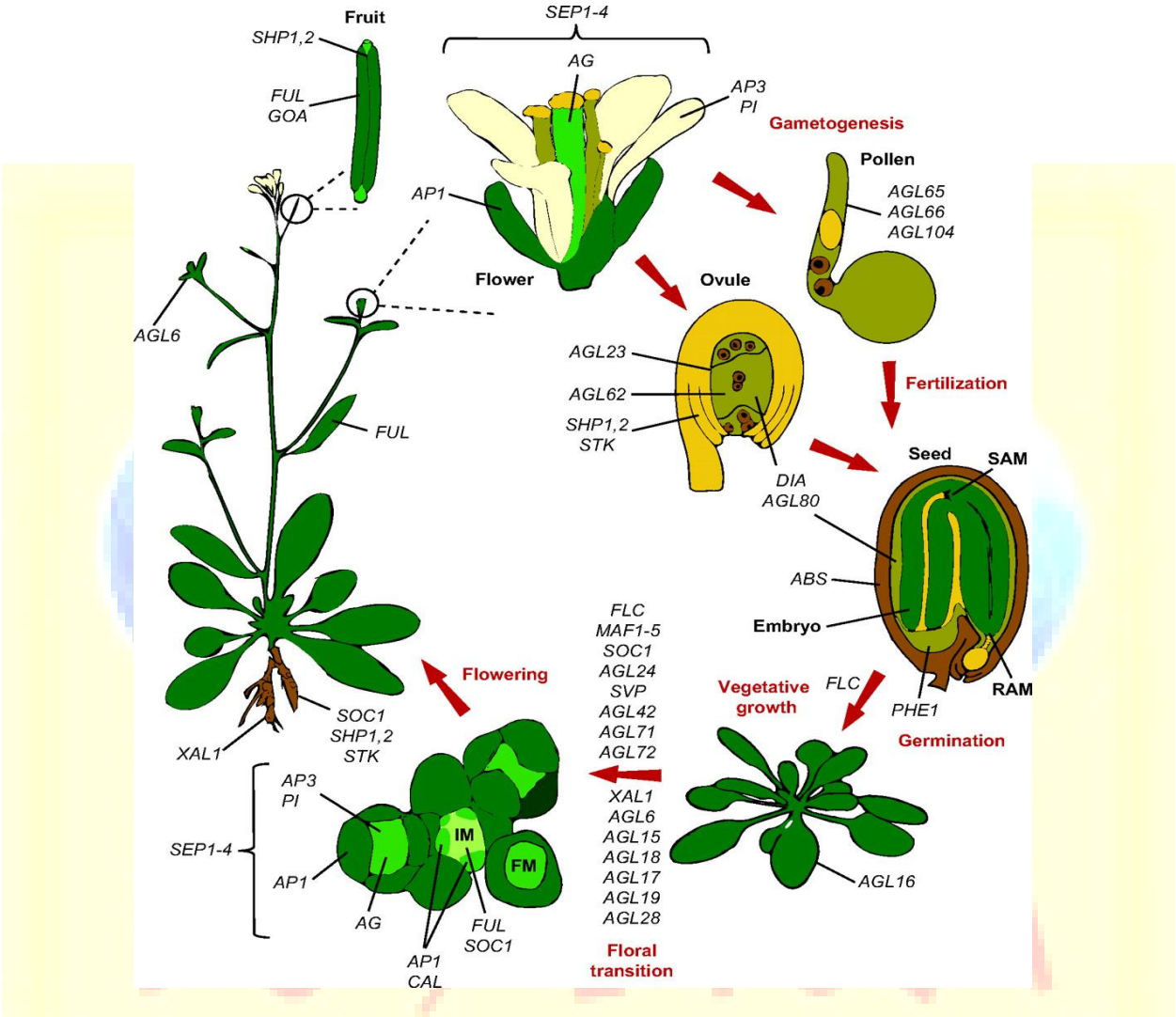


Fig 4: the life cycle of *A. thaliana*

Major reasons for the adoptionof Arabidopsis as a model for genomic analysis:

- Diploid genome, making analysis of recessive mutations easy.
- Self-fertilization so can isolate seed from a single plant without need tocross fertilize. A single plant produces hundred thousands of seeds.
- Small genome size; around140Mb.

- Genome almost completely sequenced.
- Efficient transformation by *Agrobacterium tumefaciens*.

Forward genetics identified many mutants–over1500 freely available fromstock Centre; Reverse genetic resources excellent–over 100, 000 insertions at precise sequenced locations[19].

Important advances in Arabidopsis research

- 1986. Transgenic *Arabidopsis*plantsgenerated. Regeneration of transformedplants from rootsmostwidelyusedmethod.
- 1988. Firstrestriction fragment length polymorphism made.
- 1989. Cloningof firstgeneby insertion mutagenesis. T-DNA of *Agrobacteriumtumefaciensas* themutagen.
- 1992. First *Arabidopsis*genes isolatedbypositionalcloning.
- 1993. High efficiencytransformationestablishedbyvacuumInfiltrationof *Agrobacterium* culturesintoplant tissues.
- 1997. Physicalmapof *Arabidopsis*genomecompleted. Wholegenomein overlappingbacterialartificialchromosomes of yeast.
- 2000. Paper describingcompletionof mainphaseof sequencingthe*Arabidopsis*genomeappearsin *Nature*.
- 2002. Availabilityof Affymetrixmicroarrays allowingtheSimultaneousanalysisof all known*Arabidopsis*genes.
- 2003. Availability of over 330,000 insertionsat preciselysequencedlocations. Providesgenome-wideresources for reversegenetics withininsertionsin 90% of genes.
- 2004. 15th International *Arabidopsis*conferenceheld in Berlin.1100 people attended[19].

Genome organization and duplication:

The *Arabidopsis* genome sequence provides a complete view ofchromosomal organization and clues to its evolutionary history.Gene families organized in tandem arrays of two or more units havebeen described in C. elegans1 and Drosophila2. Analysis of the*Arabidopsis* genome revealed 1,528 tandem arrays containing4, 140 individual genes, with arrays ranging up to 23 adjacentmembers. Thus, 17% of all genes of *Arabidopsis* are arrangedin tandem arrays.Large segmental duplications were identified either by directlyaligning chromosomal sequences or by aligning proteins andsearching for tracts of conserved gene order. All five chromosomeswere

aligned to each other in both orientations using MUMmer30, and the results were altered to identify all segments at least 1,000bpin length with at least 50% identity. These revealed 24 large duplicated segments of 100kb orlarger, comprising 65.6Mb or 58% of the genome. The onlyduplicated segment in the centromeric regions was a 375-kbsegment on chromosome 4. Much duplicationappear to haveundergone further shuffling, such as local inversions after theduplication event.TBLASTX5 is used to identify collinear clusters of genes residingin large duplicated chromosomal segments. The duplicated regionsencompass 67.9Mb, 60% of the genome, slightly more than was found in the DNA-based alignment and these data extendearlier findings [1][7][9]. The extent of sequence conservation of theduplicated genes varies greatly, with 6,303 (37%) of the 17,193 genesin the segments classified as highly conserved (E, 10-30) and afurther 1,705 (10%) showing less significant similarity up toE, 10-5. The proportion of homologous genes in each duplicatedsegment also varies widely, between 20% and 47% for the highlyconserved class of genes. In many cases, the number of copies of agene and its counterpart differ (forexample, one copy on onechromosome and multiple copies on the other; see Supplementary; this could be due to either tandem duplicationor gene loss after the segmental duplication. [8]

FIG 5: An Arabidopsis plant. An approximately four-weekold*Arabidopsis* plant with a rosette of leaves close to the soil and an inflorescence bearing flowers and seedpods.**Source:***Photo by Ljerka Kunst*.

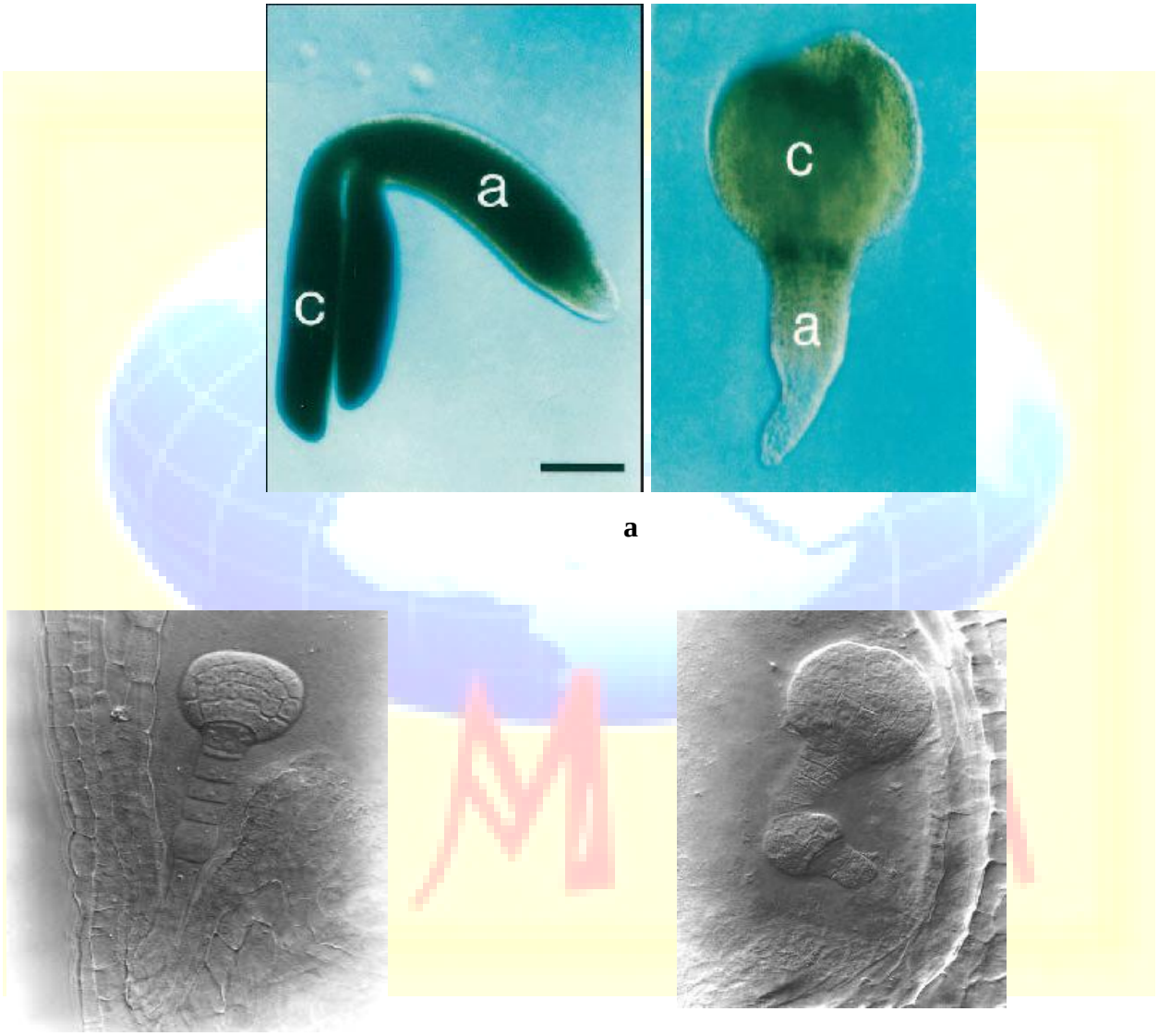
The Genetic Analysis ofDevelopment in Arabidopsis:

Although geneticists only partially understand the mechanisms of development in plants and animals, they do know that instructions for the formation of a body plan and the elaboration of specific functions are ultimately encoded in the DNA. Genetic analysis can be seen in the study of embryogenesis, hormone control systems, and responses to environmental signals[5].

The Genetic Analysis of Embryogenesis:

Mutations that are lethal to the embryo or that arrest embryonic development have provided a basis for understanding embryonic development in *Arabidopsis*. Screenings for mutations resulting in developmental arrest at a specific stage of embryogenesis have led to the identification of more than 500 embryo-lethal and embryo-defective mutations. A significant portion of the embryo-lethal mutants are arrested at the transient stage. Many of them carry mutations in housekeeping genesmutations occur in maternal genes whose proteinproducts are deposited in the egg during oogenesis. Maternal-effect mutations disrupt embryogenesis in allprogeny of a mutant mother, regardless of the zygote's genotype. The small number of maternal-effect mutations constitutes a major difference between plants and certainkinds of animals, such as *Drosophila*, that develop from large eggs[5]. (In *Drosophila*, numerous maternal-effect mutations alter the early stages of embryogenesis, as described in Portrait D.) A major reason for the near absence of maternal effects in *Arabidopsis* is that the plant's embryonic development does not depend on a large cytoplasm filled with maternal products (as happens in *Drosophila*). Instead, it depends from a very early stage on a cytoplasm newly produced from the zygote's own genome (as in mammals). Also unlike what happens in *Drosophila*, in *Arabidopsis*, very few zygotic mutations (that is, mutations in the zygote's own genes) affect embryogenesis before the globular stage[9]. At least part of this difference between *Arabidopsis a*nd*Drosophila* may derive from the fact that in plants many zygotic genes are expressed very early in embryogenesis (as they are in mammals), and these zygotic genes may encode protein products that are functionally redundant with those of maternal origin. In addition, many plant genes are members of multigene families in which each member encodes a protein whose function duplicates that of some of the other members. Hence, as in mammals, the scarcity of maternal-effect and early zygotic mutations inplants may simply reflect a large amount of functionalredundancy between maternally deposited and zygotically derivedproducts.Screenings for mutations that arrestdevelopment at a specific stage ofembryogenesis enables the

identification of important regulatory processesin plant development[9]. A large number of mutations affecting *Arabidopsis* embryo development have been isolated. Two of these mutations are highlighted to give a flavor of range of phenotypes and mechanisms involved. One type of embryo defective mutation known as *leafy cotyledon* results in embryos with cotyledons resembling mature leaves rather than wild-type cotyledons (**Fig.7a**) [11].



a

b. **c.**

FIG 7: Examples of two mutations that arrest development. **(a)** Micrograph of *lec*mutant (*right*) versus a plantcarrying*lec*wild type (*left*). The cotyledons in a *lec*mutant areabnormally expanded. **(b)** Micrograph of *twin* (*right*) versus wildtype*TWIN*embryos (*left*). *A*naxis; *(c)*cotyledons.

The genetic analysis ofhormonal control systems:

The control of plant development requires complex communicationswithin cells, between neighboring cells, and over long distances between different organ systems. Plant hormonesare the main internal chemical signals that influence plant development. One group of hormones, the *auxins* (from the Greek *auxin*, to increase or augment), promotes shoot growth through cell elongation and lateral root formation and helps prevent senescence. The auxins also inhibit growth in lateral buds in favor of growth in the apical meristem (apical dominance), and they inhibit cell elongation in roots[9]. A second set of hormones, the *cytokinins*, promotes growth by stimulating cell division, including cytokinesis (hence their name). The cytokinins also play a role in the prevention of senescence but, unlike the auxins, promote the growth of lateral buds. Together, the auxins and cytokinins help regulate a range of processes including meristem induction, apical dominance, and root-shoot communication. A third type of hormone, *abscisic acid*, controls aspects of seed development and dormancy; it also mediates responses to water stress (too much or too little water). Yet another group of plant hormones, the *gibberellins*, regulates seed germination and stem elongation. The *brassinosteroids*enhance cell growth and expansion, and they suppress leaf organogenesis in dark-grown seedlings. *Ethylene* promotes maturation of the seedling and influences the timing of leaf senescence. Finally, *salicylicacid*, *jasmonic acid*, and *methyl-jasmonate*are associated with disease resistance. Jasmonic acid also inhibits seed and pollen germination, inhibits seedling growth, and induces fruit ripening and abscission (detachment) of flowers[11]. Mutations that render *Arabidopsis*insensitive to a hormone help reveal themechanisms by which plants perceive andtransduce hormone signals. In the most advanced studies of this type, investigators look at ethylene signaling, and using ethylene's effects on early seedling development as the basis for mutant screens. The exposure to ethylene of seedlings grown in the dark produces an inhibition of shoot and root elongation and an accentuation of the apical hook. Mutant seedlings that are insensitive to ethylene grow tall in the presence of ethylene. Several loci that affect responsiveness to ethylene have been identified in *Arabidopsis* (**Fig. 8**). For example, dominant mutations at the *ETR1* locus (including *Ein1*) cause insensitivity to ethylene, as do mutations at *EIN2* and *EIN3*. By contrast, mutations at the *CTR1* locus result in the constitutive activation of ethylene responses. All these mutations affect a wide range of ethylene responses in various tissues, suggesting that the genes influence the primary steps of ethylene signal processing. Analysis of various combinations of double mutants

reveals that *CTR1*is epistatic to *ETR1*, while *EIN2* and *EIN3* are epistatic to *CTR1*.*Arabidopsis* mutants that are insensitive to other plant hormones have also been isolated; the genemutant discussed earlier is an example. The isolation of genes affected in the various mutants demonstrates that protein phosphatases, farnesyl transferases, transcription factors, and the ubiquitin protein degradation pathway play a role in plant hormonesignaling, while genes for membrane-based permeases and transporters contribute to hormone transport. The fact that many of the genes identified are homologous to genes in yeast and mammals are evidence that the basic information-processing systems arose very early in the evolution of eukaryotes [2]. On the other hand, receptors for ethylene, cytokinins, and red light (the phytochromes) are related to bacterial sensory receptors. These systems were likely inherited from the bacterial endosymbiont ancestor of the chloroplasts

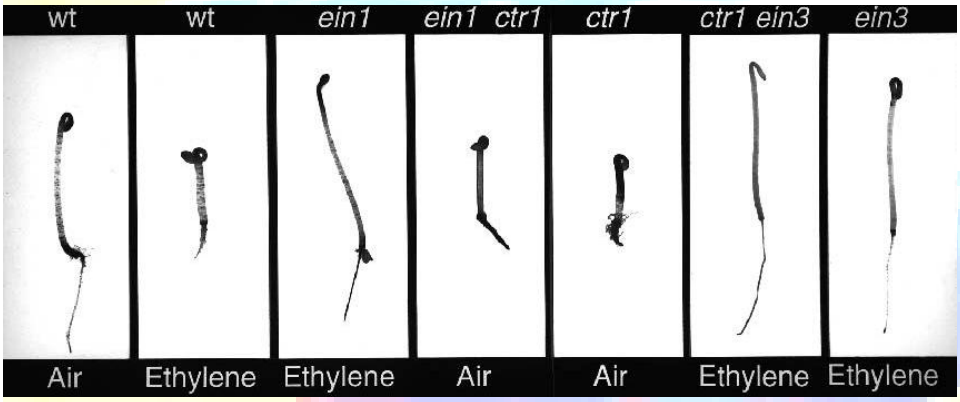


FIG 9:the effects of ethylene exposure on wild-type and mutant plants.Ethylene inhibits shoot and root elongation in wild-type seedlings. Single- and double-mutant analyses allow the ordering of genes in the signal pathway. *Ein1* is a mutation at the *ETR1* locus. The EIN3 protein acts before CTR1, which acts before *ETR1*.

CONCLUSIONS

Arabidopsis makes an excellent model organism for genetic analysis because of its diploid nature, its capacity for prolific reproduction, and the relatively small size of its genome. Thechallenge of determining the function of the large set of predicted genes, many of which areplant-specific, is now a clear priority, and multinational programs have been initiated to accomplish this goal using site-selected mutagenesis among the necessary tools.

REFERENCES

1. Bent, E.*et al.*, (1998*).Representation of two low-copy regions of the genome ofArabidopsis thaliana*. Plant J. 13, 849±855
2. Copenhaver, G. P. *et al.*, (1999). Genetic definition and sequence analysis of *Arabidopsis* centromeres. Science 286, 246-2474
3. Hoffmann, M. H. (2002): *Biogeography of Arabidopsis thaliana* (L.) Heynh. (Brassicaceae). J Biogeography 2002, 29:125-134
4. Johanson, U.*et al.*, (2000): Molecular analysis of FRIGIDA, *a major determinant of naturalvariation in Arabidopsis flowering time*. Science, 290:344-347.
5. Koornneef, M, Alonso-Blanco, C and Vreugdenhil, D. (2004): *Naturally occurring genetic variation inArabidopsis thaliana*. *Annu Rev Plant Biol*, 55:141-172.
6. Kobayashi, Y. andWeigel,D.(2000): The *Arabidopsis* Genome Initiative: *Analysis of thegenomesequence of the flowering plant Arabidopsis thaliana*. Nature 408:796- 815.
7. Le, Q. -H. *et al.*, (2000). *Transposon diversity in Arabidopsis thaliana*. *Proc. Natl. Acad. Sci*. USA 97, 7376-7381.
8. Marra, M. *et al.*, (1999). *A map or sequence analysis of the Arabidopsis thaliana genome*. *Nature Genet*. 22, 265-270.
9. Mayer, K. et al. Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana. Nature 402,769±777 (
10. Meyerowitz, E. M. & Somerville, C. R.(1994*). Arabidopsis* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
11. Meinke, D. W. *et al.*, (1998). "*Arabidopsis thaliana: a model plant for genome analysis*" *Sci*. 282, 662-665.
12. Mozo, T. *et al.*, (2009). "*A complete BAC-based physical map of the Arabidopsis thalianagenome". Nature Genet*. 22, 271-275.
13. Mitchell-Olds T, and Schmitt, J. (2006). *Genetic mechanisms and evolutionary significance ofnatural variation in Arabidopsis*. 441:947- 952
14. Sato, S. *et al.*,(2000). *Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence features of the 1.6Mb regions covered by twenty physically assigned P1 clones*. DNA Res. 4, 215-230

15. Schmid*et al.*, (2003): *Evidence for a large-scale population structure of Arabidopsis thaliana from genome-wide single* nucleotide polymorphism markers. *Theor. Appl. Genet.* 2006, **112**:1104-1114
16. Sharbel, T. F, andHaubold, B. (2000): *Genetic isolation by distance in Arabidopsisthaliana*: biogeography and postglacial colonization of Europe. *Mol. Ecol*, **9**:2109-2118
17. Somerville, C, andKoornneef, M. (2002): *A fortunate choice: the history of Arabidopsis as a model plant. Nat Rev Genet*, **3**:883-889.
18. The *Arabidopsis* Genome Initiative: Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000, 408:796-815.
19. Yu, Z., Wright, S. & Bureau, T.(1999). Mutator-like elements (MULEs) in *Arabidopsis thaliana*: Structure diversity and evolution Genetics (in the press).
20. Http//www.mpiz-koeln.mpg.de