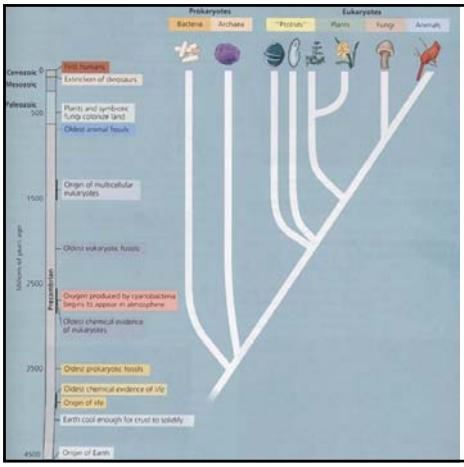
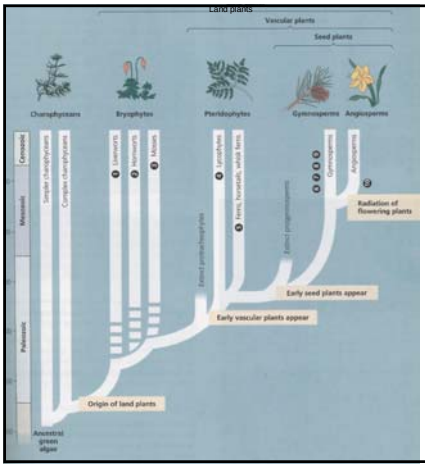
Arabidopsis thaliana as a model species for studying plant biology

- Why Arabidopsis was chosen as a model for plant biology
- History of Arabidopsis as a model
- What was learnt from the Arabidopsis genome sequence
- Arabidopsis transformation
- Determining functions of Arabidopsis genes
- Natural-genetic variation in Arabidopsis



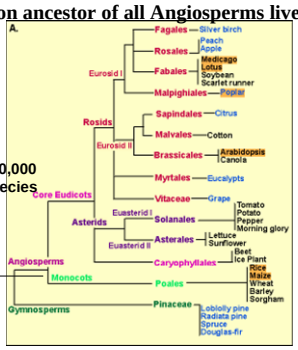
Plants diverged from animals around 1.5 billion years ago, before evolution of multicellularity. Have evolved different mechanisms of living as multicellular organisms.



Arabidopsis is an Angiosperm / flowering plant.

Arabidopsis is a model for seed plants (Angiosperms)

Flowering plants originated relatively recently and the last common ancestor of all Angiosperms lived around 180 mya.



Close relatives of Arabidopsis

Major reasons for the adoption of Arabidopsis as a model for plant molecular genetics

- Short-generation time; 8 weeks from seed to seed.
- Small (adult approximately 20 cms tall), easily grown at high density in glasshouse or culture room.
- Diploid genome, making analysis of recessive mutations easy.
- Self fertilizes, so can isolate seed from a single plant without need to cross fertilize. A single plant produces hundreds or thousands of seeds.
- Small genome size; around 130 Mb.
- Genome almost completely sequenced to high degree of accuracy.
- Efficient transformation by Agrobacterium tumefaciens.
- Forward genetics identified many mutants – over 1500 freely available from stock centre; Reverse genetic resources excellent – over 100,000 insertions at precise sequenced locations.

Timeline – important advances in Arabidopsis research (1)

1907. Arabidopsis first used by Laibach for cytology. Showed 5 chromosomes during his PhD in Bonn.

1943. Laibach described usefulness of Arabidopsis for studying genetics of phenotypes such as variation in flowering time, while Prof. of Botany in Frankfurt.

1947. Laibach's student, Erna Reinholz, isolated first mutants of Arabidopsis using X-rays.

1965. First Arabidopsis conference held in Göttingen, 25 people attended.

1983. First genetic map of Arabidopsis with genetic linkage groups covering all five chromosomes made by Maarten Koornneef.

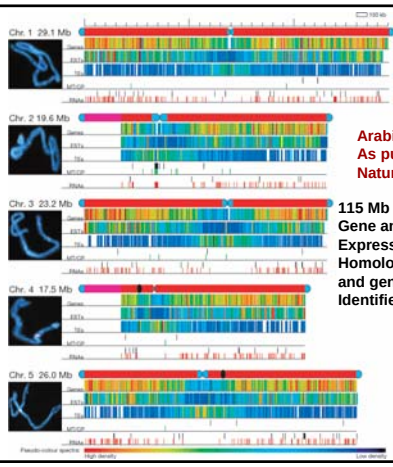
1984. Arabidopsis DNA characterized using lambda libraries. Genome size estimated at 70 Mb, by Elliot Meyerowitz.

Timeline – important advances in Arabidopsis research (2)

- 1986. Transgenic Arabidopsis plants generated. Regeneration of transformed plants from roots most widely used method.
- 1988. First restriction fragment length polymorphism map made.
- 1989. Cloning of first gene by insertional mutagenesis. T-DNA of Agrobacterium tumefaciens as the mutagen.
- 1992. First Arabidopsis genes isolated by positional cloning.
- 1993. High efficiency transformation established by vacuum infiltration of Agrobacterium cultures into plant tissues.
- 1997. Physical map of Arabidopsis genome completed. Whole genome in overlapping bacterial artificial chromosomes or yeast artificial chromosomes.

Timeline – important advances in Arabidopsis research (3)

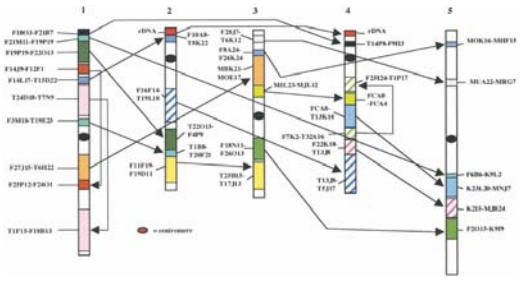
- 2000. Paper describing completion of main phase of sequencing the Arabidopsis genome appears in *Nature*.
- 2002. Availability of Affymetrix microarrays allowing the Simultaneous analysis of all known Arabidopsis genes.
- 2003. Availability of over 330,000 insertions at precisely sequenced locations. Provides genome-wide resources for reverse genetics with insertions in 90% of genes.
- 2004. 15th International Arabidopsis conference held in Berlin. 1100 people attended.



Arabidopsis genome sequence. As published in 2000. Nature 408, 796.

115 Mb of 125 Mb genome.
Gene annotation using Expressed sequence tags (ESTs)
Homology with cloned plant genes and genes of other organisms
Identified 25,500 genes.

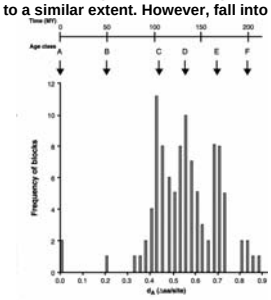
Large segments of the Arabidopsis genome are duplicated



Origin of genomic duplications in Arabidopsis

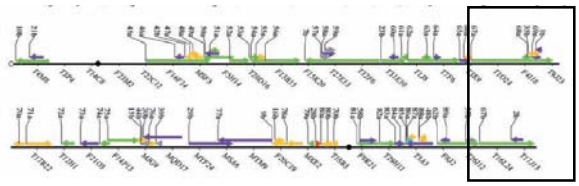
-103 duplicated blocks containing 7 or more genes
- Over 81% of ORFs fall within the bounds of a block, but only 28% of genes are present in duplicate due to extensive deletions extensive deletions of genes

Number of duplicated genes, suggests that the whole genome may have been duplicated, then expect all duplicated genes to have diverged to a similar extent. However, fall into three major age classes:



C – 48% of genes bounded; probably represents whole genome polyploidization.
D – 39% genes
E – 11% genes
F - 3% genes

Genetic redundancy can exist between genes in duplicated blocks



A duplicated block of genes exists on chromsomes 2 and 3.

One of the duplicated genes encodes a MADS box transcription factor, and the proteins encoded by the two genes are 87% identical at the amino acid level.

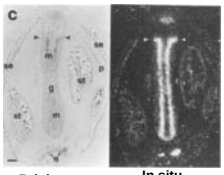
MADS — SHATTERPROOF 1

100% identical in MADS DNA binding domain

MADS — SHATTERPROOF2

SHP1 and SHP2 are expressed in similar patterns in the the developing Arabidopsis fruit

Expression of SHP1 in young flower bud in the developing fruit

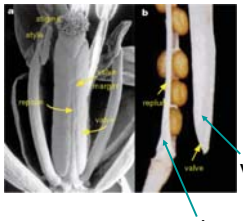


Bright field

In situ hybridization

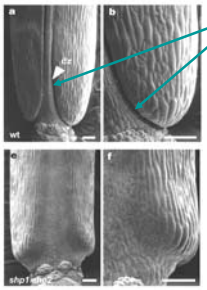
SHP2 expressed in a similar pattern

Wild-type flower



Valve separates from replum to release the seeds

SHP1 and *SHP2* are genetically redundant



shp1 or shp2 single mutants show no phenotype, but the fruit of the double mutant is impaired in dehiscence.

Improved annotation of the Arabidopsis genome:

Reportoire of gene families in Arabidopsis (11,000 – 15,000) similar to other sequenced multicellular eukaryotes.

However, gene number in Arabidopsis surprisingly high:

Arabidopsis – 30,700 genes. (Version 5 annotation).

Drosophila melanogaster – 13, 676 genes (Release 3).

Some of these extra genes are due to genome duplications, and other plants also seem to have high gene numbers. Rice the second plant genome to be sequenced Is estimated to contain around 40,000 genes in 15,000 families.

But there appear to be many genes that are unique to plants and not found in animals:

8,000 (25%) of Arabidopsis genes have homologues in the rice genome, but not In Drosophila, C.elegans or yeast.

What is unique about plants that can be inferred from the Arabidopsis genome?....

Enzymes involved in secondary metabolism

Arabidopsis genome contains many classes of enzymes involved in secondary metabolism that are required for the synthesis of specialized compounds.

An example, is the family of genes encoding the Cytochrome P450 monooxygenase enzymes.
Act as terminal oxidases in electron transport chains.

Mammals, C.elegans, Drosophila – 80 – 105 genes.

Arabidopsis – 246 genes.

In plants these enzymes are required for the synthesis of compounds such as growth regulators/plant hormones (gibberellic acid, Brassinosteroid), carotenoids (protect cell from oxidative damage) and phenylpropanoids that are present in plant cell walls.

Transcription factors

Arabidopsis contains around 1500 genes encoding transcription factors (approx. 5%)

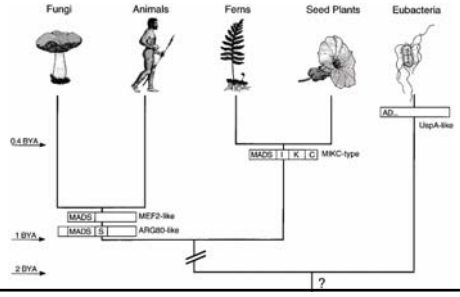
Drosophila contains around 640 genes encoding transcription factors, around 4.5%.

Many important animal transcription factor families are absent in plants, such as nuclear steroid receptors, NHR zinc finger proteins (252 in C. Elegans) and Fork head transcription factors (18 in Drosophila, 15 in C.elegans).

Each eukaryotic lineage has its own set of transcription factor families.

MADS box transcription factors are named after proteins found in yeast, humans and plants

M : MCM1 yeast
A : Agamous
D : Deficiens, Antirrhinum B function gene
S : serum response factor, humans



MADS box TFs have been amplified in the plant lineage

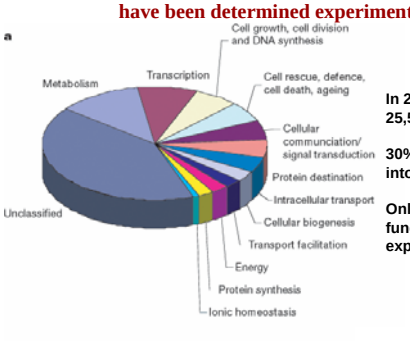
	MADS
Arabidopsis	107
Drosophila	2
C. Elegans	2
Yeast	4

agamous mutant

WT

Some have well defined roles in flower development, like AGAMOUS, but 84% are of unknown function

The functions of a minority of Arabidopsis genes have been determined experimentally



In 2000:
25,500 genes predicted

30% could not be classified into a functional group.

Only 9% were assigned a function based on experimentation

Forward genetics: Isolation of mutants of Arabidopsis

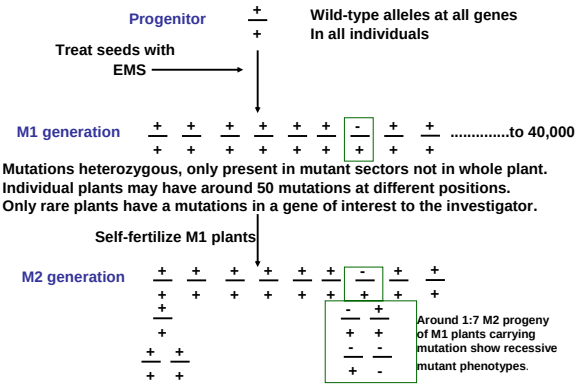
- Treat seeds of *Arabidopsis thaliana* (Columbia) with mutagen
  - chemical mutagen ethylmethane sulfonate (EMS) most common
  - radiation, X-rays or gamma rays also used.

Typically around 40,000 seeds treated with mutagen.

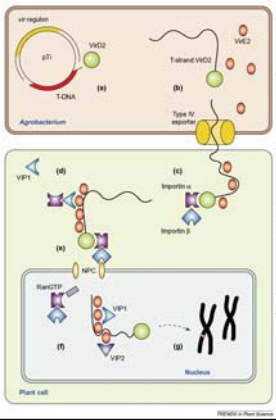
- Plant the seeds on soil, and grow the plants.
  This is the M1 generation.
  Mutations are heterozygous and not present in every cell, because a mutation occurs in only one cell in the embryo of the seed.
  M1 plants self fertilize and seeds are harvested, typically in pools of 1000 M1 plants.

- M2 seeds are sown and mutants with the phenotypes expected identified.

Genetics of mutant screening



Arabidopsis can be transformed using *Agrobacterium tumefaciens*



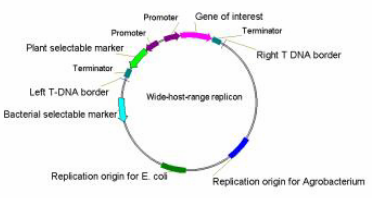
Agrobacterium cell

Agrobacterium in nature carries a Tumor-inducing (Ti) plasmid:

In response to plant signals VIR regulon expressed. Contains 8 genes.
Transferred-DNA (T-DNA) defined by border sequences. A single strand of DNA Is transferred to plant cell.
VirD2 protein is bound to 5'end.
VirE2 protein coats the T-strand.
VirD2 and VirE2 proteins both contain Nuclear localization sequences that Localize the T-DNA to the plant nucleus.
VirE2 binding proteins VIP1 and VIP2 Target the T-strand to regions of chromatin In the plant chromosome.

Plant Cell

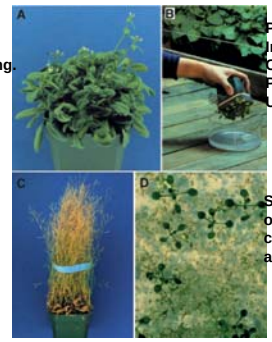
The Ti plasmid is modified to make binary vectors used for plant transformation



VIR genes are provided on a second helper plasmid, in the Agrobacterium cell.

Arabidopsis transformation by floral dipping

Arabidopsis plants are grown in pots until they start flowering. Around 10/pot.

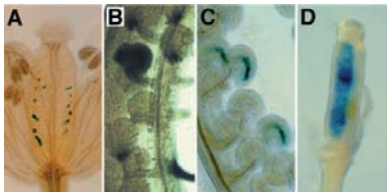


Plants are dipped in Agrobacterium Culture plus sucrose Plus surfactant for Up to 3 minutes.

Plants grown for a Few more weeks to Allow seed development. Seeds collected.

Seeds germinated on medium containing selectable agent e.g. Kanamycin

Expression in dipped plants of markers for plant gene expression present on the T-DNA



Agrobacterium enters the developing flower and the T-DNA is introduced into the developing female gametophyte, and transmitted Through the ovule to the next generation.

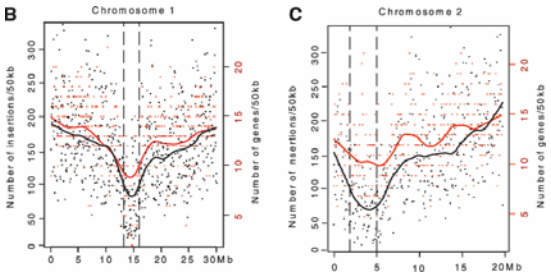
The ease of Agrobacterium-mediated transformation allows the T-DNA to be used as a mutagen for insertional mutagenesis and reverse genetics

Over 300,000 transformants were made, DNA isolated and the junction fragment between the T-DNA and the plant DNA recovered. This allows the precise location of the T-DNA in the genome to be assessed. Called Flanking Sequence Tag (FST)

Insertions in around 90% of genes are present.

These FSTs are present in databases, so insertions in a gene of interest can be recovered by searching the database.

Insertions are distributed non-randomly in the genome

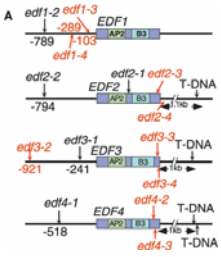


T-DNAs in black; genes in red
Region between vertical lines corresponds to the predicted centromeres.

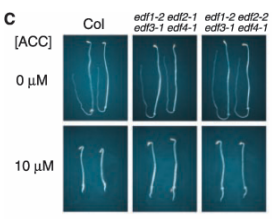
Libraries of FSTs and T-DNA insertions can be used for Reverse genetics to assign functions to Arabidopsis genes

Assigning functions to four AP2-like transcription factors called ETHYLENE RESPONSE DNA BINDING FACTORS 1-4.

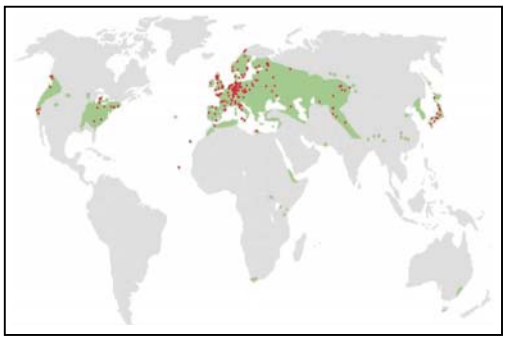
FSTs identify T-DNAs in genes



The quadruple mutant is insensitive to the growth regulator ethylene (ACC)



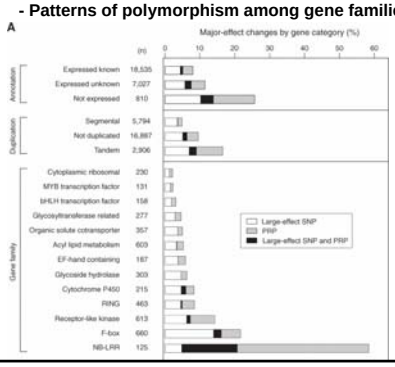
Global distribution of Arabidopsis



Arabidopsis accessions collected from positions across the range

Genome sequences of 20 diverse accessions were recently compared to the reference sequence (Columbia)

- Approx. 4% of the genome highly dissimilar or deleted relative to the reference sequence (Columbia)
- Patterns of polymorphism among gene families highly non random



Clark et al. (2007) Science **317**, 338.

Allelic variation at *FRIGIDA* and *FLOWERING LOCUS C* affects the time of flowering of many Arabidopsis accessions

San Feliu

Landsberg erecta

FRIGIDA
↓
FLOWERING LOCUS C
⊥
Flowering

frigida - deletion

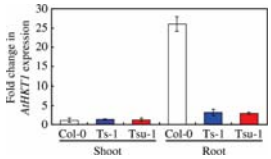
flowering locus C - insertion

Two Arabidopsis strains from coastal regions of Spain (Ts-1) and Japan (Tsu-1) are more resistant to NaCl

Col-0

Tsu-1

Plants treated with 100 microM NaCl



This difference is due to changes in the promoter region of a gene (AtHKT1) that encodes a sodium transporter.

Presentation appears as PDF on Max Planck web site

WEB ADDRESS:

www.mpiz-koeln.mpg.de

-Forschung
-Abt. Entwicklungsbiologie de Pflanzen
-George Coupland (bottom of page)
- Volesungsreihe

Also PDF files from papers used as references