Arabidopsis thaliana as a model species for studying plant biology

Lecture 1:	History Genome structure Genes that make a plant
Lecture 2:	Analysis of gene function in Arabidopsis

Why do we need a plant model?

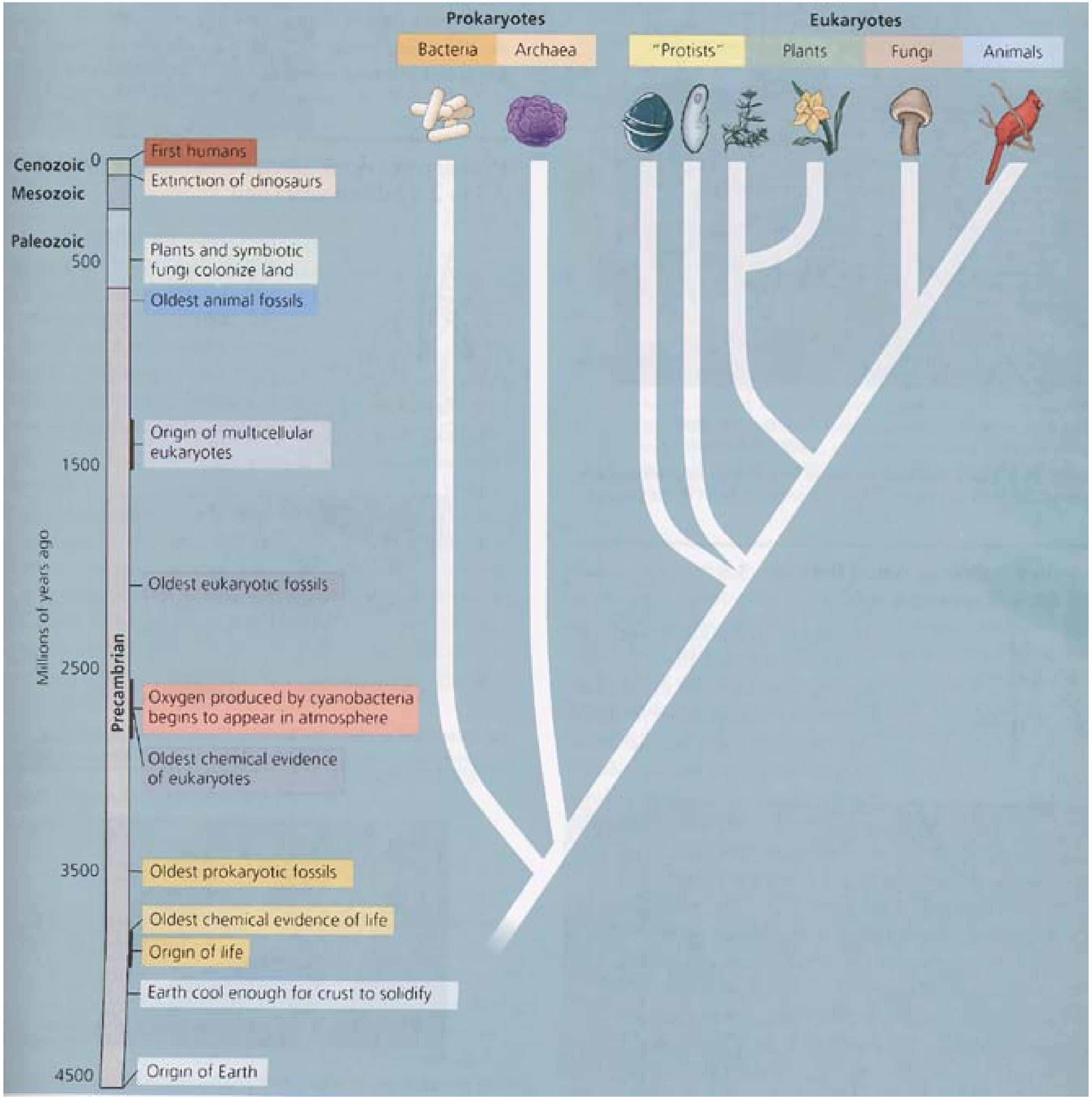
- Plants constitute over 90% of the world‘s biomass; 250,000 species of flowering plant

- Important biological processes are plant specific; photosynthesis that fixes carbon and produces oxygen

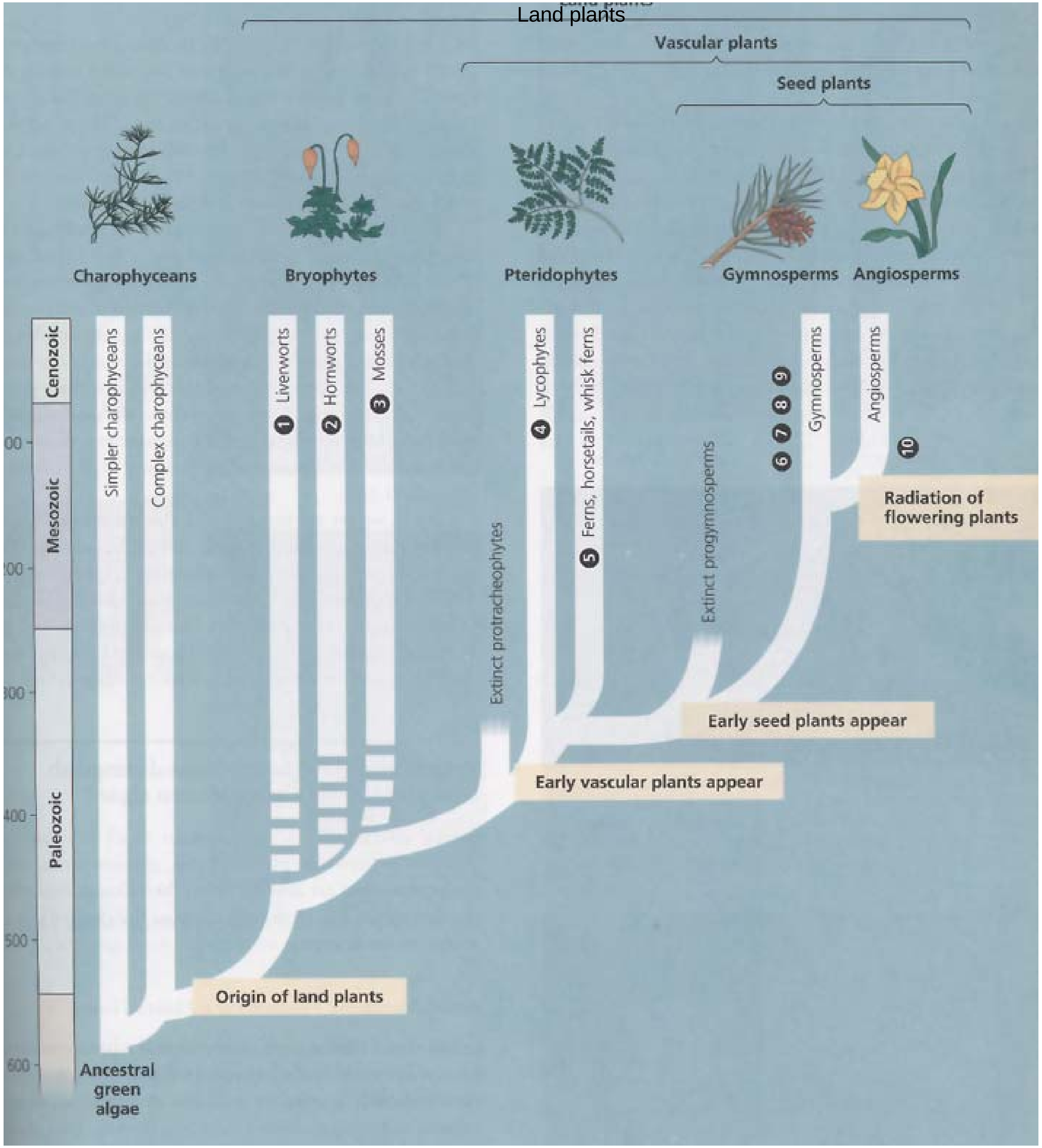
- Plants are economically important; in agriculture or in secondary metabolites as medicines and in nutrition

- Plants evolved multicellularity independently, and use different mechanisms of cell to cell communication

- Plants represent important genetic model systems; transposons and gene silencing were first identified in plants.



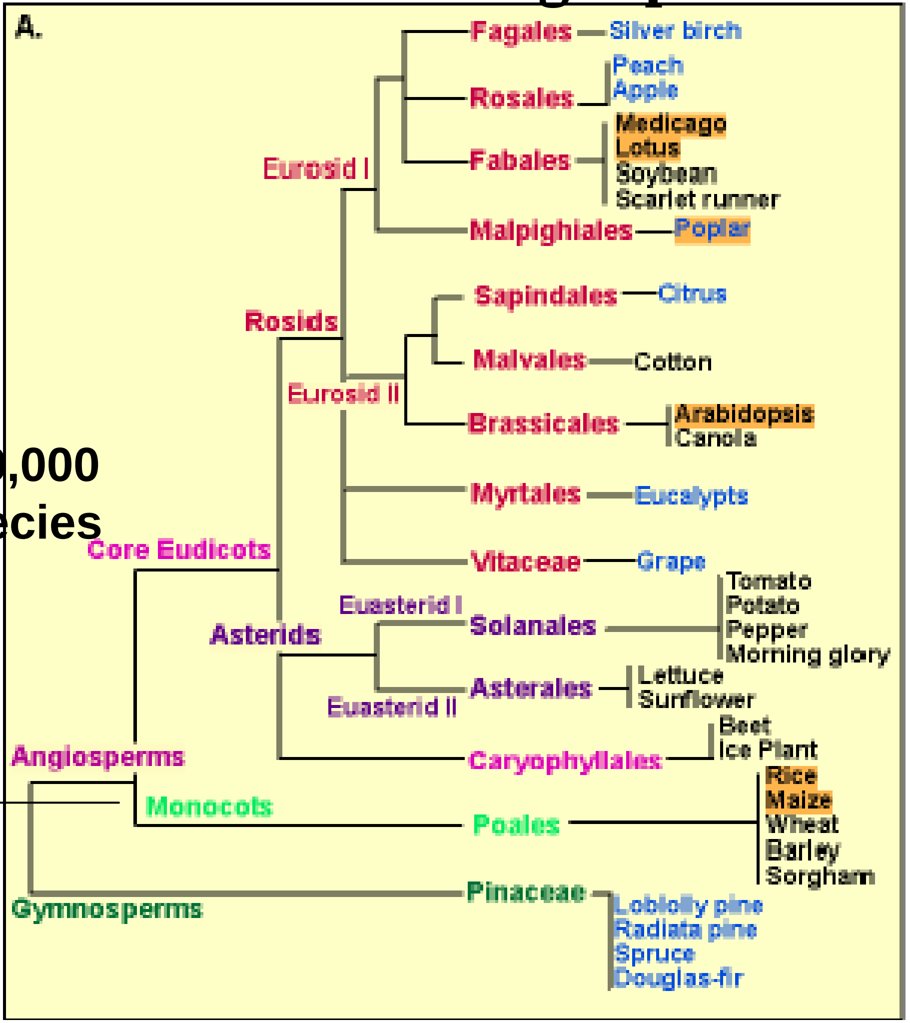
**Plants diverged from animals around 1.5 billion years ago.
Have evolved different mechanisms of living as multicellular organisms.**



Arabidopsis is an Angiosperm / flowering plant.

Arabidopsis is a model for Eudicot seed plants

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



170,000 species

65,000 species

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 140 Mb.

- Genome almost completely sequenced.

- 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, be 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.

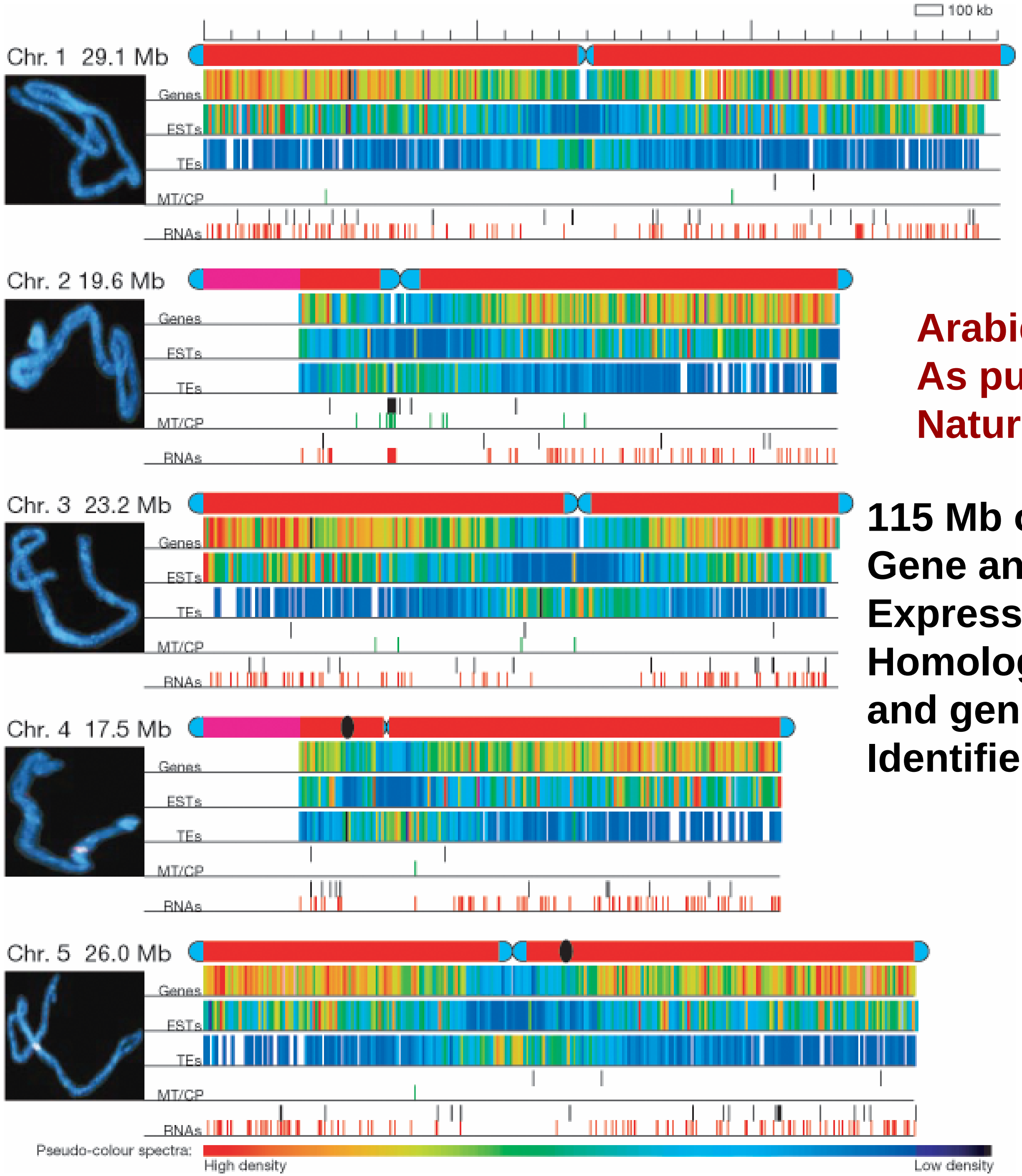
Timeline – important advances in Arabidopsis research (3)

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.

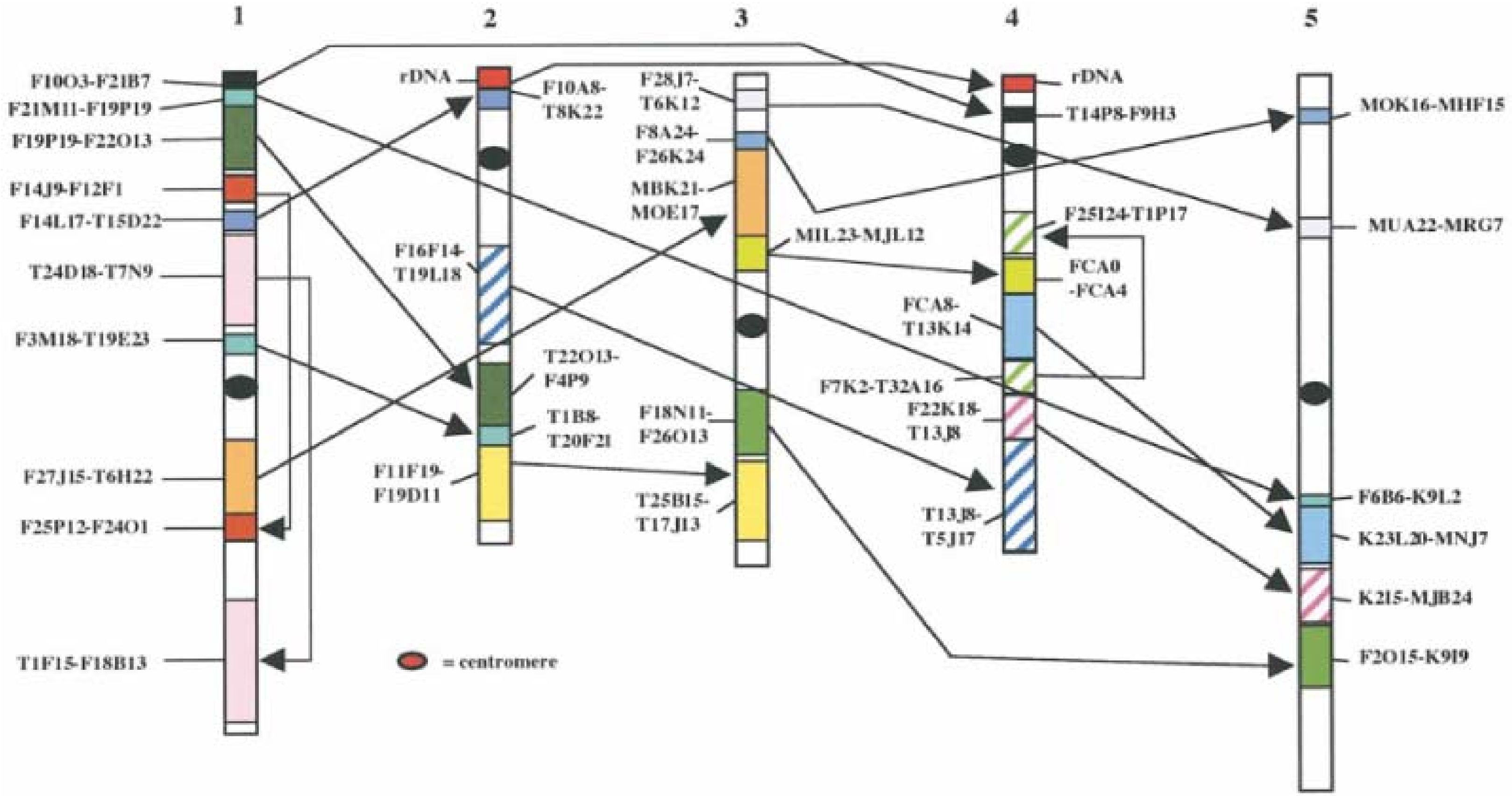
2004. 15th International Arabidopsis conference held in Berlin. 1100 people attended.



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.

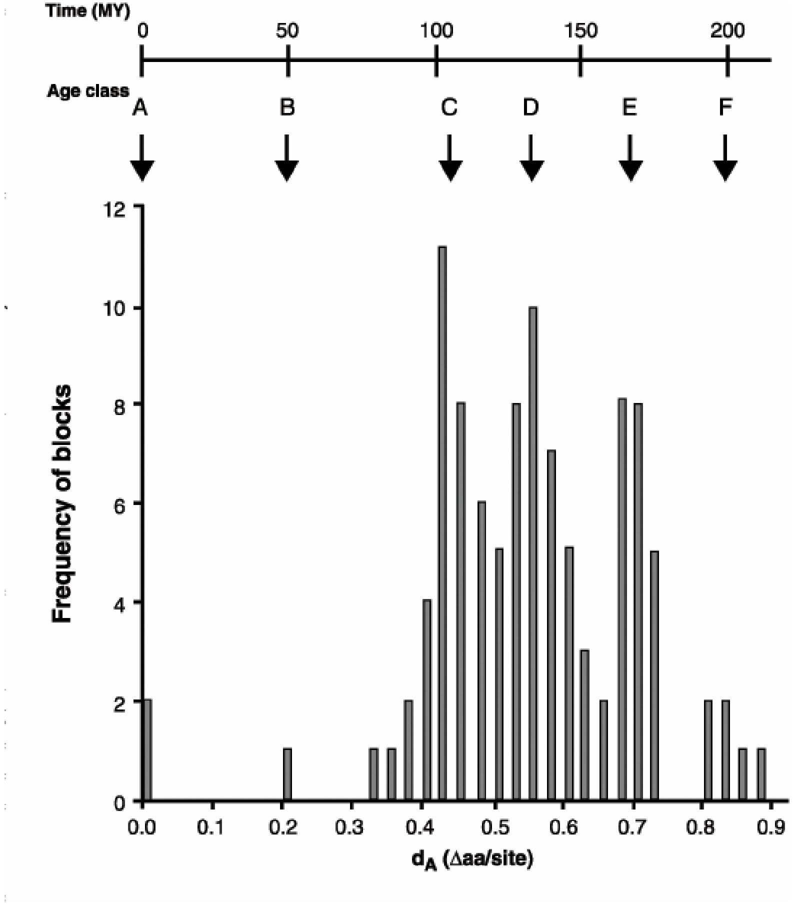
Large segments of the Arabidopsis genome are duplicated



Origin of genomic duplications in Arabidopsis

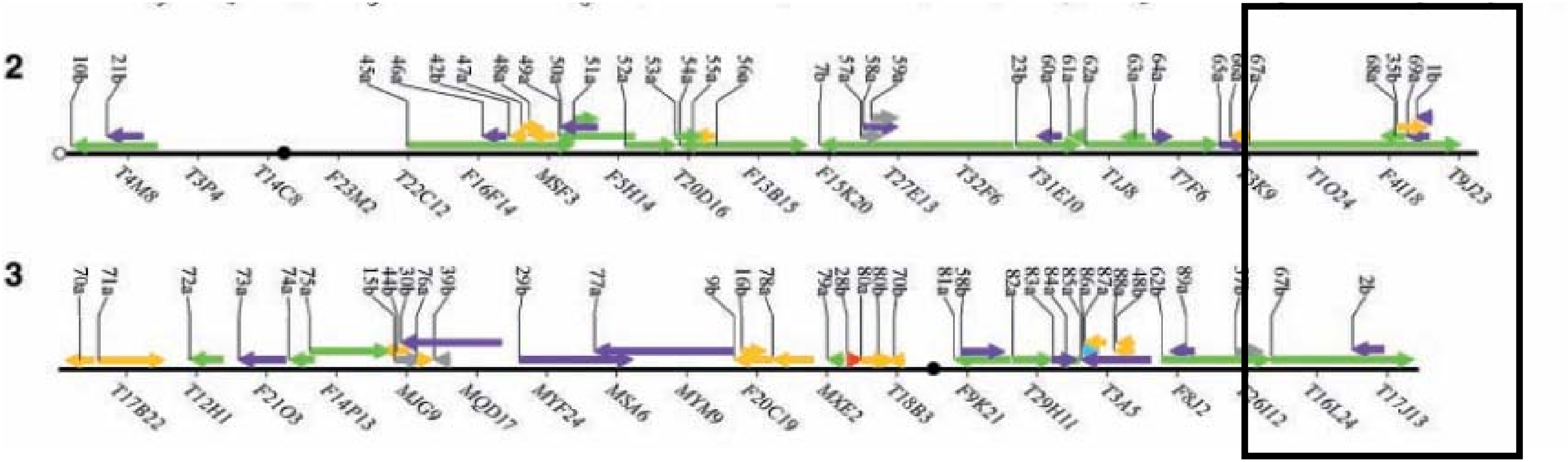
-103 duplicated blocks containing 7 or more genes
- Over 81% of ORFs fall within the bounds of a block,
  but only 28% of genes are present in duplicate
  due to extensive deletions extensive deletions of genes

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.

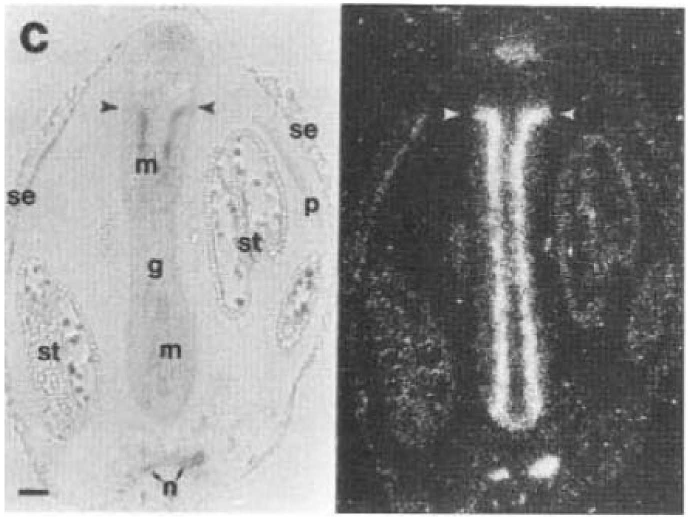
SHATTERPROOF 1

100% identical in MADS DNA binding domain

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

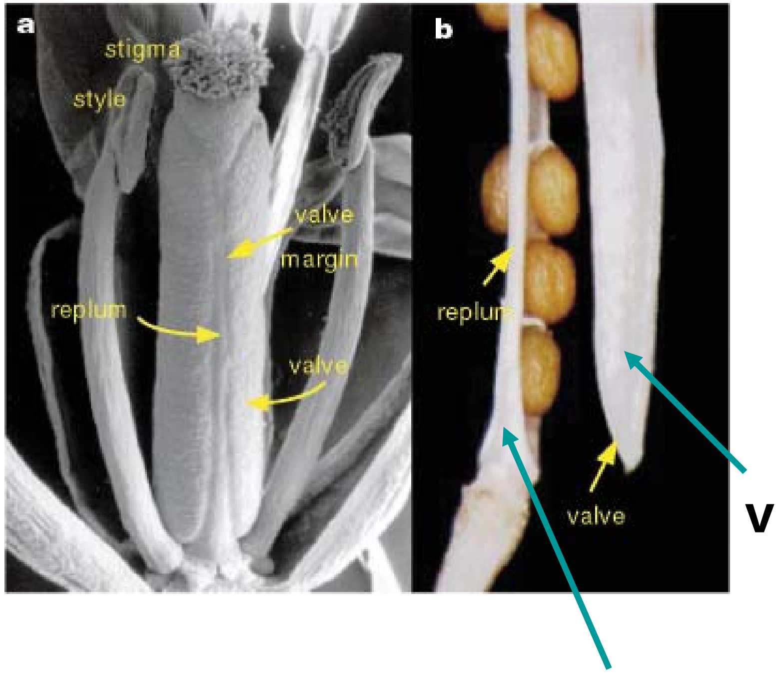


Bright field

In situ hybridization

SHP2 expressed in a similar pattern

Wild-type flower

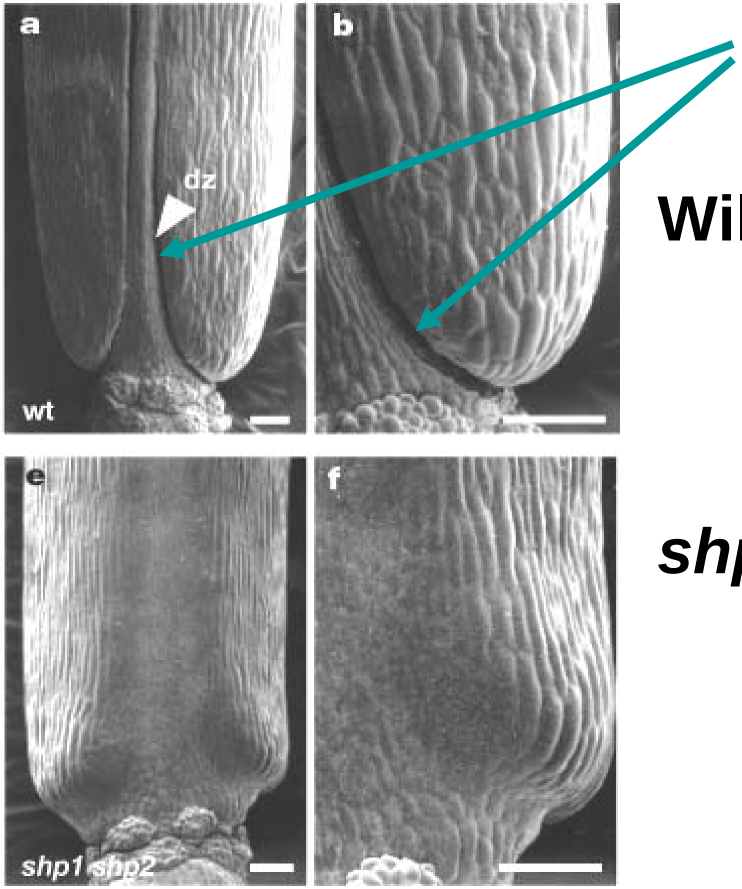


valve

replum

Valve separates from replum to release the seeds

SHP1 and *SHP2* are genetically redundant



Dehiscence zone

Wild-type

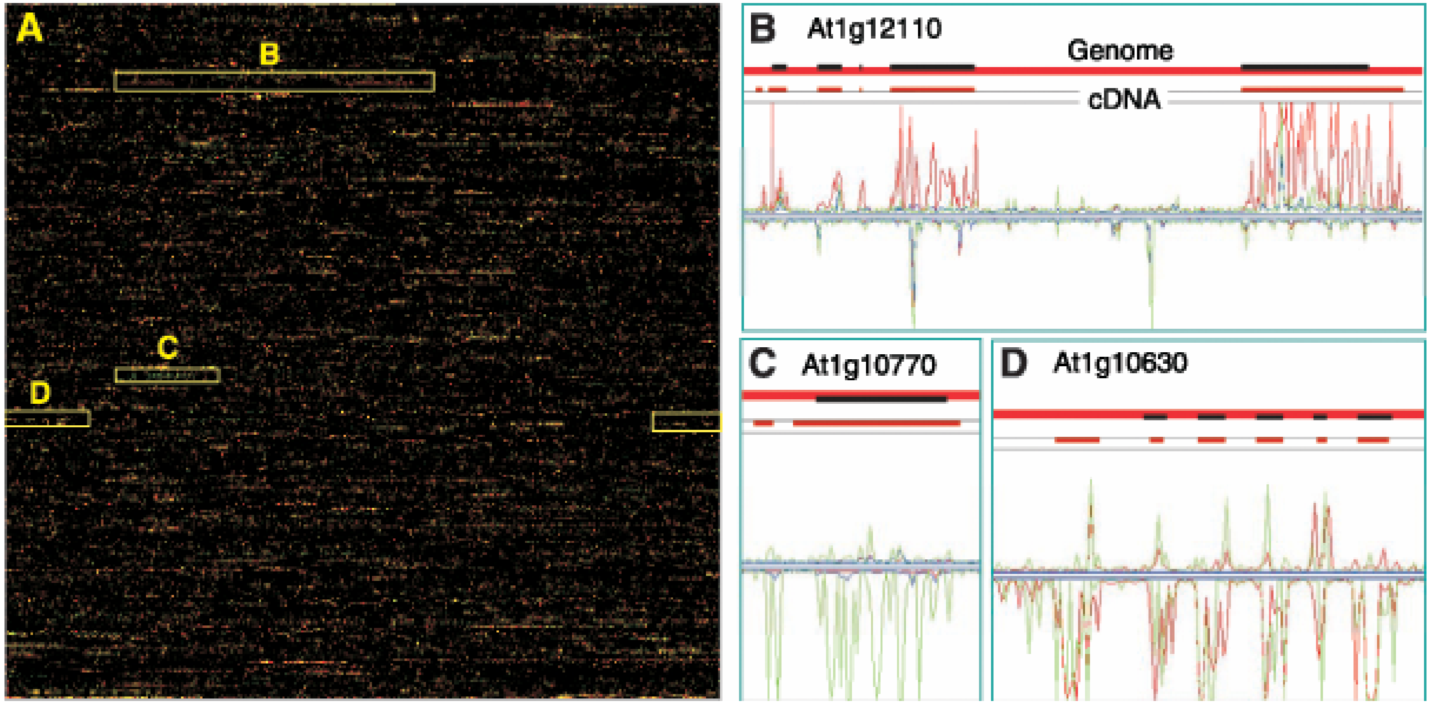
shp1 shp2

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 sequence

Initial sequence analysis relied heavily on expressed sequence tags (ESTs) and gene prediction programmes. Interpretation of genome sequence constantly improved.

Isolation of full-length cDNAs and use of whole genome oligonucleotide tiling arrays greatly improved the annotation. (32% of predicted gene models incorrect).



Arabidopsis genome represented on 12 oligonucleotide arrays.
Each array 834,000 25-mer oligonucleotides.
Is hybridized with cDNA made from 4 RNA populations; seedlings, roots, flowers, cultured cells. Sequences that hybridize are mapped directly onto genome sequence.

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 monoxygenase enzymes.

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 (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 (aprox. 5%)

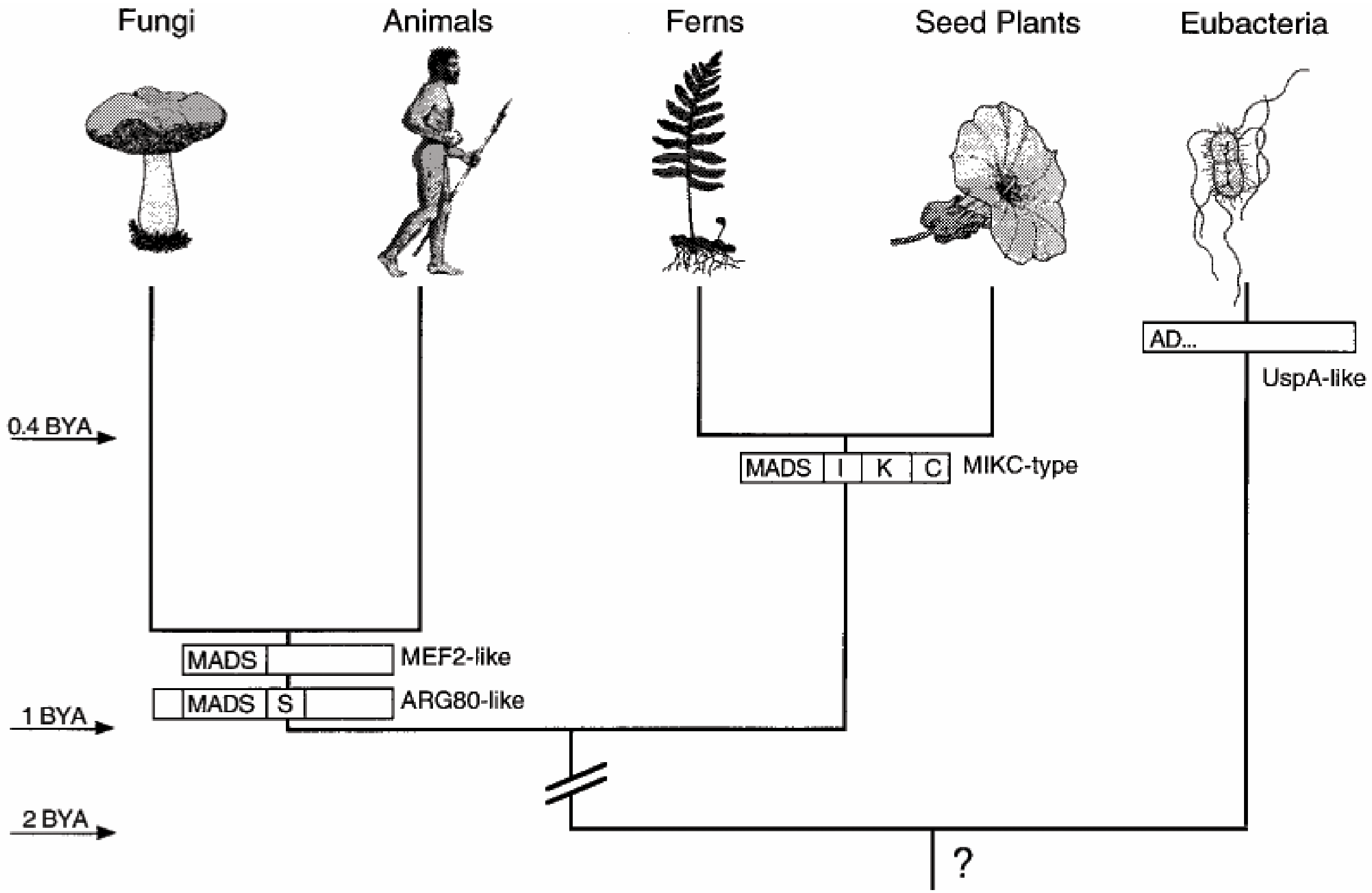
Drosophila contains around 640 genes encoding transcription factors, around 4.5%.

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

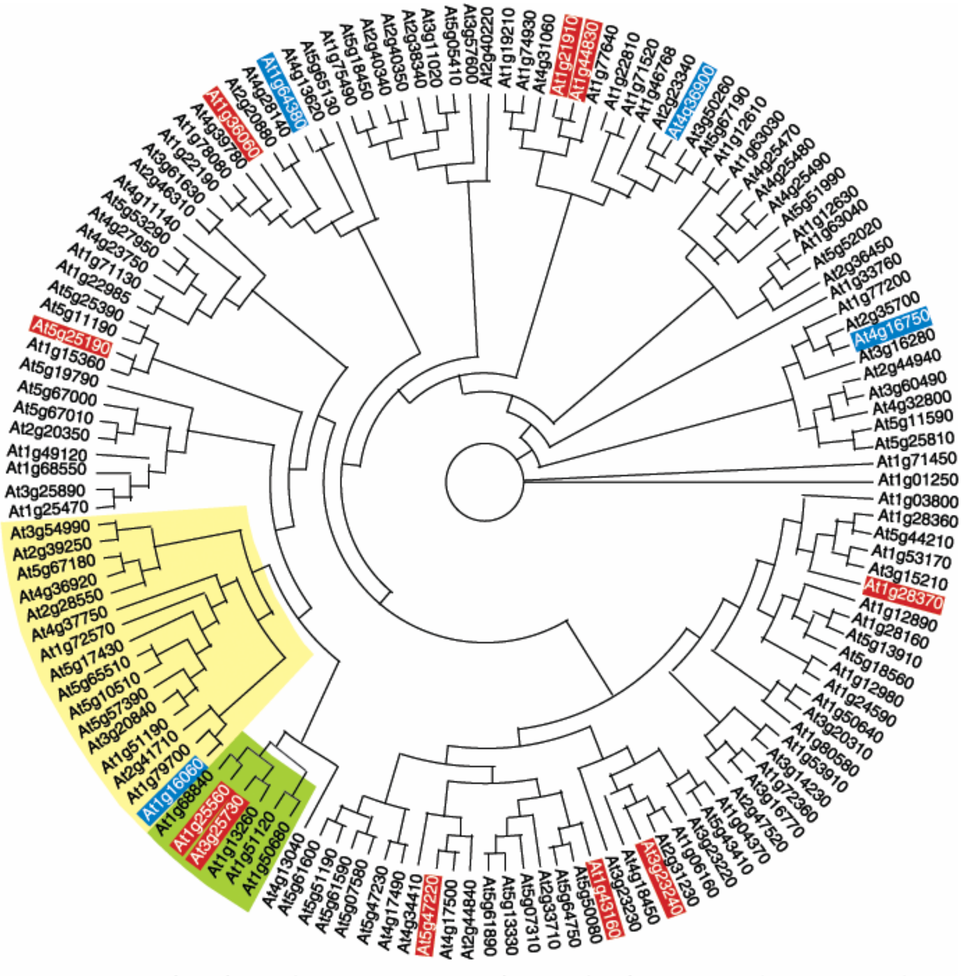
agamous mutant

WT

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

APETALA 2-like transcription factors are unique to plants

This is one of several families of transcription factor that are only found in plants. May have evolved from bacterial and bacteriophage endonucleases



141 AP2-domain TFs in Arabidopsis

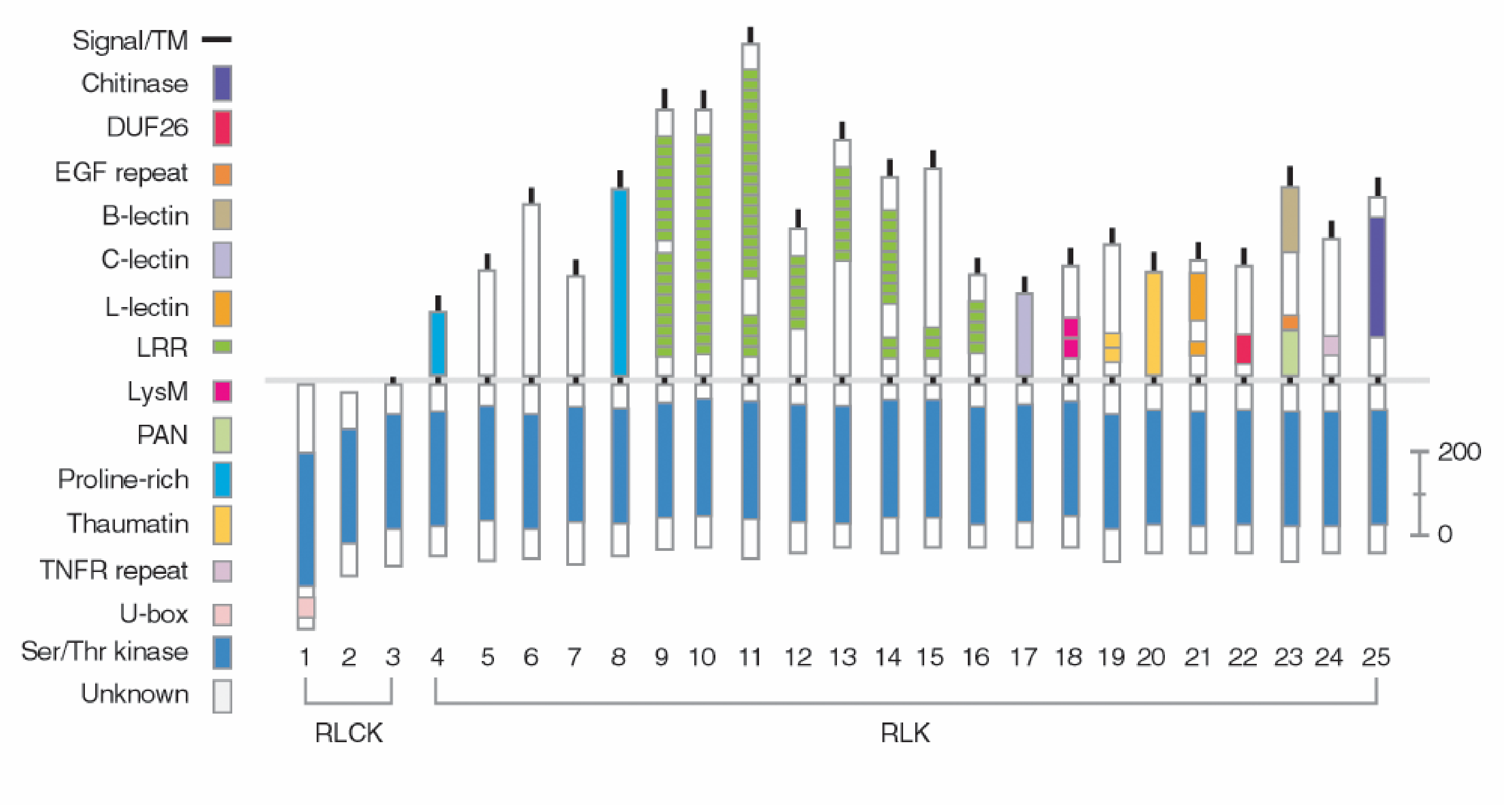
-Some related to APETALA2, Involved in flower development (yellow)

-Some involved in ethylene response and their transcription is changed by ethylene (red/blue)

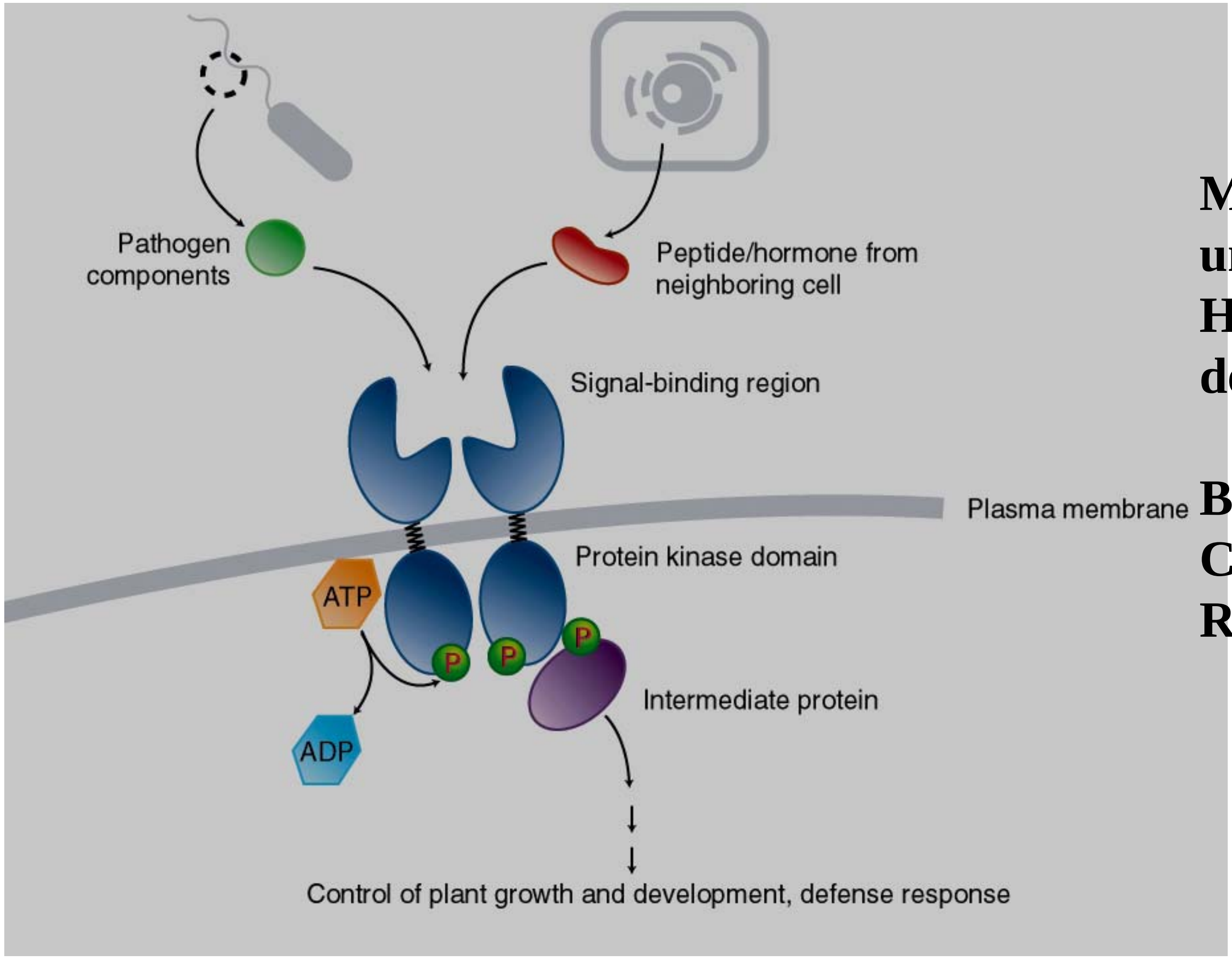
-Others are involved in stress responses to drought

Receptor-like kinases

600 Arabidopsis genes encode receptor-like kinases predicted to be located in the membrane.
These are similar in domain organization to animal receptor tyrosine kinases, such as epidermal growth factor.
However, are predicted to be serine/threonine kinases, and have divergent ligand binding domains.



Some Receptor Like Kinases have important functions, but for most their function is unknown



Many RLKs are of unknown function. However, some have defined roles:

Brassinosteroid receptor
Clavata 1
Resistance to pathogens

Other plant-specific processes

Hundreds of genes involved in photosynthesis

- light harvesting
- chlorophyll biosynthesis
- carbon dioxide fixation
- energy generating photosystems

Transporters

- plants mainly use proton-type ATPases, whereas in animals transport is usely coupled with sodium ions via sodium-type ATPases.

No major histocompatibility complex, however 100s of genes
Encoding nucleotide binding site leucine rich repeat proteins
Involved in pathogen resistance.

Presentation appears as PDF on Max Planck web site

WEB ADDRESS:

www.mpiz-koeln.mpg.de

-Forschung
-Abt. Entwicklungsbiologie der Pflanzen
-George Coupland (top of page)
- English version
- Teaching
- October 2005

Also PDF files from papers used as references