Lecture 2: Functional analysis of Arabidopsis

Assigning functions to genes.

Forward and reverse gentics.
Transformation.
Natural Variation
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

Progenitor + Wild-type alleles at all genes
+ In all individuals

Treat seeds with EMS

M1 generation + + + + + + - + + + + .............to 40,000
+ + + + + + + +

Mutations heterozygous, only present in mutant sectors not in whole plant. Individual plants may have around 50 mutations at different positions. Only rare plants have a mutation in a gene of interest to the investigator.

Self-fertilize M1 plants

M2 generation + + + + + + - + + + +
+ + + + + + + +
+ + + + + + + +
+ + + + + + + +
+ + + + + + + +
+ + + + + + + +

Around 1:7 M2 progeny of M1 plants carrying mutation show recessive mutant phenotypes.
Examples of Arabidopsis mutants with visible phenotypes; 1500 available from stock centres

Embryo development

Flower development

Response to darkness

Response to light
Mapping an Arabidopsis mutation using molecular markers

Mutations identified in *Arabidopsis thaliana* (Columbia).

Another frequently used accession is *Arabidopsis thaliana* (Landsberg erecta).
Can use DNA markers to distinguish between Landsberg and Columbia.
There are more than 50,000 sequence polymorphisms between Landsberg and Columbia. The position of these markers in the genome sequence is known precisely.

These polymorphisms are detected in different ways; preferred markers are:
- Insertions/deletions that give different sized PCR fragments.
- Sequence polymorphisms that change a restriction enzyme recognition sequence so that in one accession the polymorphism is not cleaved with an enzyme.

Example of a marker that distinguishes Columbia and Landsberg after digestion with Dra1 restriction enzyme.

1. Size standard
2. PCR product
3. Landsberg PCR + Dra1
4. Columbia PCR + Dra1
5. F1 PCR + Dra1
Determining an approximate map position using molecular markers

Select for mutants based on phenotype. Every plant must be Columbia homozygous at site of mutation.
Test markers at 20 cM intervals; 25 markers covers the whole genome; Test DNA of 30 mutant plants.

- Common genotype; 50% chromosomes Landsberg for A; genetically unlinked to mutation.
- Rarer genotype; 20% chromosomes Landsberg for B; marker weakly genetically linked to mutation.
- Very rare genotype; 1% chromosomes Landsberg for C; marker strongly genetically linked to mutation.
Examples: linked and unlinked markers

F2 mutants

5 mutants tested; 9/10 chromosomes
Columbia; linked marker

F2 mutants

6 mutants tested; 5/12 chromosomes
Columbia; unlinked marker

By testing DNA from 800-1000 F2 mutant plants to identify crossovers close to the mutation and using markers closer to the mutation – candidates for the gene can be identified in the genome sequence. Mutant gene finally identified by making transgenic plants in which the wild-type gene is introduced into the mutant and shown to complement the mutation. Map-based cloning.
Transformation of Arabidopsis using *Agrobacterium tumefacies*

**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 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)
Arabidopsis thaliana is widely distributed around the world

Green : geographical distribution of Arabidopsis
Red   : accessions located and widely available

Analyzing natural genetic variation provides an important source of new genes and alleles that were difficult to identify by mutagenesis
Analysis of phenotypic differences between accessions can provide information on genetic variation important in nature.

**Flowering time**
- Santa Fe
- Columbia

**Resistance to fungal pathogen**
- Landsberg erecta (Columbia)
- Columbia

**Leaf shape**
- Bla-14
- CVI

Many other examples:
- Light responses,
- Root development,
- Secondary metabolism

Genes underlying natural variation can be isolated by map-based cloning using sequence polymorphism between accessions.
The Cape Verde Islands variety is earlier flowering than Landsberg *erecta*, especially under short days.

German variety Landsberg *erecta*

Variety from the Cape Verde islands
Construction of Recombinant inbred lines

Landsberg erecta \( \times \) Cape Verde Islands

F1

F2

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 → 163

Single seed descent to F8

Score pattern of 321 markers on all chromosomes in all 163 recombinant inbred lines.

Score phenotype in all 163 recombinant inbred lines.
CVI carries a dominant gene on chromosome 1 that promotes early flowering

Mapping of EDI
Construction of Near Isogenic lines (NILs)

parents

One Recombinant Inbred line

Leaf number

Near isogenic line
The **EARLY DAYLENGTH INSENSITIVE** locus is an allele of the **CRYPTOCHROME 2** gene.

Cryptochrome 2 is a blue light receptor and the proteins made in CVI and Ler differ at three amino acids.

<table>
<thead>
<tr>
<th></th>
<th>CVI</th>
<th>Ler</th>
<th>Columbia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pterin chromophore</td>
<td>Q</td>
<td>S</td>
<td>Q</td>
</tr>
<tr>
<td>Flavin chromophore</td>
<td>S</td>
<td>L</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>V</td>
<td>V</td>
</tr>
</tbody>
</table>

**CRYPTOCHROME 2** contains 612 amino acids.
Construction of transgenic plants carrying a single altered gene provides the final evidence that EDI is identified.
Identify genes for adaptive traits using Arabidopsis natural variation

- Collect Arabidopsis ecotypes in extreme places
- Make crosses and mapping populations
- Identify position of quantitative trait loci (QTL)
- Clone the QTL genes - identify quantitative trait nucleotide (QTN)
- Analyse variation for these genes in natural plant populations

For example analyzing genetic differences between CVI and Ler identified a new Cryptochrome 2 allele that promotes early flowering in CVI.