New methodologies for gene identification and characterization:

Concepts and case studies
Course Outline for New Methodologies for gene identification and characterization (Functional Genomics)

• Introduction/Course Overview
  • Introduction
    • Concept of Functional Genomics
    • Goal of FG
• Forward/Reverse Genetic approaches
• Forward Genetics: Approaches and techniques
• Reverse Genetics: Approaches and techniques
• Prof Weisz lab visiting
• Lab training
• Seminar: Vincenzo D’amelio, UniNA
  Functional genomics: new genes for potato pigmentation.
• Seminar: Alessandra Rogato, CNR-IBBR
  Segreti molecolari dal fondo degli oceani: genetica e genomica funzionale delle diatomee marine
• Paper discussion
• Final test: Student Proposals
Genomics is a discipline in genetics that deals with the sequence, function and evolution of a genome (the complete set of DNA within a single cell of an organism):

• **Structural Genomics:** aimed at constructing genetic and physical maps and at the sequencing of entire genomes;

• **Functional Genomics:** aimed at assigning a function to genes with the ultimate goal of understanding how the different genes drive the development and how their malfunctioning can cause a pathological state.

The first step is bioinformatics to assign a functional annotation to genes.
Functional Genomics

Genome Sequence

Gene Discovery

In silico gene prediction

Experimental validation of gene sequence and structure

Gene Functional Annotation

Forward Genetics

Reverse Genetics
The new techniques of Next Generation Sequencing have allowed the sequencing of several plant genomes in the last few years.
### Table 1. Continued.

<table>
<thead>
<tr>
<th>Scientific name</th>
<th>Common name</th>
<th>Year</th>
<th>Type</th>
<th>Division or monocot/dicot</th>
<th>Chr (♯)</th>
<th>Size</th>
<th>Assembled</th>
<th>Assem</th>
<th>Gene (♯)</th>
<th>Repeat</th>
<th>scaffold N50</th>
<th>contig N50</th>
<th>Sequencer types</th>
<th>Journal</th>
<th>PMID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyrus betrakanideri</td>
<td>pear</td>
<td>2013</td>
<td>crop</td>
<td>dicot</td>
<td>17</td>
<td>527</td>
<td>512</td>
<td>97</td>
<td>4,218,12</td>
<td>53</td>
<td>541</td>
<td>36</td>
<td>1</td>
<td>Genome Research</td>
<td>23149293</td>
</tr>
<tr>
<td>Citrusus kanusus</td>
<td>watermelon</td>
<td>2012</td>
<td>crop</td>
<td>dicot</td>
<td>11</td>
<td>425</td>
<td>354</td>
<td>83</td>
<td>23,440</td>
<td>45</td>
<td>2380</td>
<td>26</td>
<td>1</td>
<td>Nature Genetics</td>
<td>23179023</td>
</tr>
<tr>
<td>Triticum aestivum</td>
<td>wheat</td>
<td>2012</td>
<td>crop</td>
<td>monocot</td>
<td>21</td>
<td>17,000</td>
<td>3800</td>
<td>22</td>
<td>94,000</td>
<td>80</td>
<td>NA</td>
<td>1</td>
<td>4</td>
<td>Nature</td>
<td>23192148</td>
</tr>
<tr>
<td>Gossypium raimondii</td>
<td>cotton D</td>
<td>2012</td>
<td>crop</td>
<td>dicot</td>
<td>13</td>
<td>880</td>
<td>738</td>
<td>84</td>
<td>37,505</td>
<td>61</td>
<td>18,800</td>
<td>136</td>
<td>Sa,4,1</td>
<td>Nature</td>
<td>23257886</td>
</tr>
<tr>
<td>Prunus mume</td>
<td>chinese plum</td>
<td>2012</td>
<td>crop</td>
<td>dicot</td>
<td>8</td>
<td>280</td>
<td>237</td>
<td>85</td>
<td>31,390</td>
<td>45</td>
<td>578</td>
<td>32</td>
<td>1</td>
<td>Nature</td>
<td>23271652</td>
</tr>
<tr>
<td>Cicer aninum</td>
<td>chickpea</td>
<td>2013</td>
<td>crop</td>
<td>dicot</td>
<td>8</td>
<td>738</td>
<td>532</td>
<td>72</td>
<td>28,269</td>
<td>49</td>
<td>39,990</td>
<td>24</td>
<td>Sa,1</td>
<td>Nature</td>
<td>23354103</td>
</tr>
<tr>
<td>Hovea brasiliensis</td>
<td>rubber tree</td>
<td>2013</td>
<td>crop</td>
<td>dicot</td>
<td>18</td>
<td>2150</td>
<td>1119</td>
<td>52</td>
<td>68,955</td>
<td>72</td>
<td>3</td>
<td>NA</td>
<td>4,5,1</td>
<td>BMC Genomics</td>
<td>23375136</td>
</tr>
<tr>
<td>Phylllostachys moso bamson</td>
<td>mono- model</td>
<td>2013</td>
<td>non- model</td>
<td>monocot</td>
<td>24</td>
<td>2075</td>
<td>2051</td>
<td>99</td>
<td>31,987</td>
<td>59</td>
<td>329</td>
<td>12</td>
<td>1</td>
<td>Nature Genetics</td>
<td>23435089</td>
</tr>
<tr>
<td>Oyza brachyantha</td>
<td>rice relative</td>
<td>2013</td>
<td>non- model</td>
<td>monocot</td>
<td>12</td>
<td>300</td>
<td>263</td>
<td>88</td>
<td>32,038</td>
<td>29</td>
<td>1013</td>
<td>20</td>
<td>1</td>
<td>Nature Communications</td>
<td>23481403</td>
</tr>
<tr>
<td>Prunus persica</td>
<td>peach</td>
<td>2013</td>
<td>crop</td>
<td>dicot</td>
<td>8</td>
<td>265</td>
<td>227</td>
<td>86</td>
<td>27,852</td>
<td>37</td>
<td>27,400</td>
<td>214</td>
<td>Sa</td>
<td>Nature Genetics</td>
<td>23525075</td>
</tr>
<tr>
<td>Aegilops tauschii</td>
<td>wheat DD</td>
<td>2013</td>
<td>crop</td>
<td>monocot</td>
<td>7</td>
<td>4360</td>
<td>4244</td>
<td>97</td>
<td>43,150</td>
<td>66</td>
<td>58</td>
<td>5</td>
<td>4,1</td>
<td>Nature</td>
<td>23535592</td>
</tr>
<tr>
<td>Triticum urartu</td>
<td>wheat AA</td>
<td>2013</td>
<td>crop</td>
<td>monocot</td>
<td>7</td>
<td>4940</td>
<td>4660</td>
<td>94</td>
<td>34,879</td>
<td>67</td>
<td>64</td>
<td>3</td>
<td>1</td>
<td>Nature</td>
<td>23535596</td>
</tr>
<tr>
<td>Nelumbo nucifera</td>
<td>ancient lotus</td>
<td>2013</td>
<td>crop</td>
<td>monocot</td>
<td>8</td>
<td>929</td>
<td>804</td>
<td>87</td>
<td>26,685</td>
<td>57</td>
<td>3400</td>
<td>39</td>
<td>1</td>
<td>Genome Biology</td>
<td>23663246</td>
</tr>
<tr>
<td>Utricularia gibba</td>
<td>bladderwort</td>
<td>2013</td>
<td>non- model</td>
<td>monocot</td>
<td>16</td>
<td>77</td>
<td>82</td>
<td>106</td>
<td>28,500</td>
<td>3</td>
<td>95</td>
<td>26</td>
<td>4,1</td>
<td>Nature</td>
<td>23665961</td>
</tr>
<tr>
<td>Picea abies</td>
<td>norway spruce</td>
<td>2013</td>
<td>crop</td>
<td>gymnosperm</td>
<td>12</td>
<td>19,600</td>
<td>12,000</td>
<td>61</td>
<td>28,354</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>Nature</td>
<td>23498360</td>
</tr>
<tr>
<td>Capsella rubella</td>
<td>capsella</td>
<td>2013</td>
<td>non- model</td>
<td>dicot</td>
<td>8</td>
<td>219</td>
<td>135</td>
<td>62</td>
<td>26,521</td>
<td>NA</td>
<td>15,100</td>
<td>134</td>
<td>Sa</td>
<td>Nature Genetics</td>
<td>23749190</td>
</tr>
</tbody>
</table>

1 Abbreviations: Sa, Sanger; 4, Roche/454; 5, SOLiD; I, Illumina; NA, not reported in primary publication; kb, kilobases; Mb, megabases; Chr, chromosome; PMID, PubMed ID.
It All Started with a Wild Mustard Plant....

Figure 1. Timeline of *Arabidopsis* sequencing and resequencing efforts. SNPs are used as a proxy for all types of polymorphisms. The numbers of publicly available SNPs before 2003 and after 2010 are estimated.
But why all this focus on *Arabidopsis thaliana*?

- Small genome: 5 chromosomes;
- Short life cycle (8 weeks from seed to seed);
- Requires small spaces to grow;
- Easy to grow;
- High availability of mutants;
- No agronomical importance but advantageous to basic research.
Functional Genomics

Genome Sequence

Gene Discovery

In silico gene prediction

Experimental validation of gene sequence and structure

Gene Functional Annotation

Forward Genetics

Reverse Genetics
Functional Genomics

Gene Discovery and *in silico* gene prediction

1° step is **gene prediction**: a bioinformatic approach aimed at identifying genes encoded in the genome of interest; e.g.: looking for ORFs

![Graph showing the number of published plant genomes against predicted genes (thousands).]

- Arabidopsis
- Maize

Subject of other classes
Functional Genomics

Genome Sequence

Gene Discovery

In silico gene prediction

Experimental validation of gene sequence and structure

Gene Functional Annotation

Forward Genetics

Reverse Genetics
Experimental validation of gene sequence and structure

• Does the expressed sequence match the gene prediction?
• The intron/exon organization is confirmed?
• Are the UTRs confirmed? Etc etc.

Strategies used:
• ESTs;
• Full length cDNAs;
• Expression Arrays;
• RNAseq.

<table>
<thead>
<tr>
<th></th>
<th>Protein Coding Genes</th>
<th>Transposons and pseudogenes</th>
<th>Alternatively spliced genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAIR10 (11/10)</td>
<td>27,411</td>
<td>4,827</td>
<td>5,885</td>
</tr>
<tr>
<td>TAIR9 (6/09)</td>
<td>27,379</td>
<td>4,827</td>
<td>4,626</td>
</tr>
<tr>
<td>TAIR8 (4/08)</td>
<td>27,235</td>
<td>4,759</td>
<td>4,330</td>
</tr>
<tr>
<td>TAIR7 (4/07)</td>
<td>26,819</td>
<td>3,889</td>
<td>3,866</td>
</tr>
<tr>
<td>TAIR6 (11/05)</td>
<td>26,541</td>
<td>3,818</td>
<td>3,159</td>
</tr>
<tr>
<td>TIGR5 (1/04)</td>
<td>26,207</td>
<td>3,786</td>
<td>2,330</td>
</tr>
<tr>
<td>TIGR4 (4/03)</td>
<td>27,170</td>
<td>2,218</td>
<td>1,267</td>
</tr>
<tr>
<td>TIGR3 (8/02)</td>
<td>27,117</td>
<td>1,967</td>
<td>162</td>
</tr>
<tr>
<td>TIGR2 (1/02)</td>
<td>26,156</td>
<td>1,305</td>
<td>28</td>
</tr>
<tr>
<td>TIGR1 (8/01)</td>
<td>25,554</td>
<td>1,274</td>
<td>0</td>
</tr>
<tr>
<td>Nature (12/00)</td>
<td>25,498</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
Functional Genomics

Genome Sequence

Gene Discovery

In silico gene prediction

Experimental validation of gene sequence and structure

Gene Functional Annotation

Forward Genetics

Reverse Genetics
Gene Functional Annotation

In silico functional annotation

3 possible methods:

• Information intrinsic to the sequence
  (Presence of signal peptides, transmembrane domains etc);

• Gene context based
  (Functionally related genes are in some cases closely associated on the genome).

• Homology based: unknown proteins are compared with proteins stored in databases
Homology based methods: the principle

What is homology? Two proteins are homologous if they share a common ancestor.

If a newly sequenced gene is highly homologous to a known-function gene, we can assign to our unknown gene a predicted biological role.

Homology based methods: an example of a BLAST search
A word of caution

Analyses in silico are always predictions: to unequivocally assign gene functions we need «wet» work!

Gene Functional Annotation

Why mutant analysis? The inactivation of a gene is generally the most direct way to understand its function.

What is a mutant? In biology and especially genetics, a mutant is an individual, organism, or new genetic character, arising or resulting from an instance of mutation, which is a base-pair sequence change within the DNA of a gene or chromosome of an organism resulting in the creation of a new character or trait not found in the wild type.

Source: wikipedia
Functional Genomics

- Genome Sequence
  - Gene Discovery
    - *In silico* gene prediction
    - Experimental validation of gene sequence and structure
  - Gene Functional Annotation
    - Forward Genetics
    - Reverse Genetics
Forward vs. Reverse genetics: What’s the difference?

Forward genetics – phenotype is selected and then the mutation in a specific locus is identified

Reverse genetics – mutation in a specific locus is identified and then the phenotype is determined

• Gene – functional DNA unit, promoter and coding sequence, i.e. locus
• Locus – usually two alleles at each, contributed by two homologous chromosomes in a diploid (2n) organism
Forward Genetics

Identification of mutations resulting in a particular phenotype of interest.

Steps:

a) Select a biological process;
b) Generate a mutant population;
c) Screen your mutants: phenotype;
d) Map and clone the mutation responsible for the observed phenotype.
a) Select a biological process: which one are you interested in???

etc.etc.
b) generate a *representative* mutant population;

3 requisites:

1. The number of mutations within our collection needs to be 5-10X the number of genes;

2. Every single mutant line has to be catalogued, propagated and grouped so that it gets actually screened

1. The methods of mapping need to be sensitive enough so that they can be identified within a pool of plants.
Several kinds of mutagens are available which differ for **efficiency** and **type** of mutations generated.

<table>
<thead>
<tr>
<th>Mutagen</th>
<th>Main characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chemical agents</strong> (for example, EMS)</td>
<td>Plotpoint mutations</td>
</tr>
<tr>
<td></td>
<td>Large range of mutations including loss and gain of function</td>
</tr>
<tr>
<td></td>
<td>Very high efficiency with hundreds of mutations per genome</td>
</tr>
<tr>
<td><strong>Physical agents</strong> (Fast neutrons, X-rays and accelerated ions)</td>
<td>Big insertions, deletions and rearrangements</td>
</tr>
<tr>
<td></td>
<td>Mostly loss-of-function mutants</td>
</tr>
<tr>
<td></td>
<td>Medium efficiency</td>
</tr>
<tr>
<td></td>
<td>Laborious identification of the mutations</td>
</tr>
<tr>
<td><strong>Biological agents</strong> (Transposons and T-DNA)</td>
<td>Insertion of specific DNA sequences</td>
</tr>
<tr>
<td></td>
<td>Mostly loss-of-function mutants</td>
</tr>
<tr>
<td></td>
<td>Low efficiency</td>
</tr>
<tr>
<td></td>
<td>Easy identification of the mutations</td>
</tr>
</tbody>
</table>

Unfortunately, no efficient homologous recombination strategies have been developed in plants to date...
EMS treatment in Arabidopsis

10 to 50 mM for few hours

EMS treatment

2-3000 M1 plants to obtain 3 alleles per locus

Collect M2 seeds in pools.
Pool size is variable.

Individual plants may have around 50 mutations at different positions.

Only rare plants have mutations in a gene of interest to the investigator.

Frequency of mutation is 1/300 to 1/30000 per locus.

Growing and phenotyping of 8 M2 plants per M1 plant

DNA extraction from plant A and collection of M3 seeds
Several EMS mutagenized populations ready-to-screen are available.
How does EMS mutagenesis work?

Mutations are heterozygous, only present in mutant sectors not in the whole plant. For the mutation to be transferred to the next generation, one of the 2 cells of the embryo that will give rise to reproductive tissues will have to be mutagenized.

Mutant showing desired phenotype
EMS mutagenesis delivers 99% G/C-to-A/T transition mutations:

- New stop codons: only 5%
- Introduction of missense mutations: ~65%
- Loss-, gain-of-function mutants
- Weak nonlethal mutants.
- No chromosome rearrangements
EMS-induced mutations are randomly distributed
Importance of EMS for obtaining weak alleles, when complete abolishment of a gene causes letality.

Picture is an example of seedling lethal phenotypes.
c) Screen your mutants: phenotype

Phenotype needs to be clearly distinguishable, reproducible and easy to detect
Examples of EMS mutants affecting flower development

*Arabidopsis*
Examples of Arabidopsis EMS mutants affecting development

Embryo

Plant height
Examples of EMS mutants affecting vegetative and leaf development

*B. rapa*
Examples of EMS mutants affecting root development

Fig. 3. Mutations in the AtPIN2 gene alter root growth and gravitropism. Homozygous 5-day-old Columbia-0 wild-type seedlings (A) and Atpin2::En701 mutant seedlings (B) were grown vertically on agar plates.
Examples of EMS mutants affecting response to the environment

**Light**

- Response to darkness
  - WT
  - WT
  - cop1

- Response to light
  - Wildtype
  - phyB-4

**NaCl**

- Root Development (C24)
- Shoot Development (Col-0 sos3-1)
- Stress
- Stress
EMS mutagenesis can result in mutants of the same gene with phenotypes of different strength.
Examples of 3 different alleles of the same gene.
d) Map and clone the mutation responsible for the observed phenotype.

MAPPING CLASSICAL GENETICS

Mapping using classical genetics uses molecular markers **physically anchored** to the different chromosomes of the genomes. These markers need to be **polymorphic** in different ecotypes of Arabidopsis, such as Columbia (Col-0) and Landsberg erecta (Ler).
Figure 7. Map-based cloning of the VTC2 gene. A, First-pass mapping of the VTC2 identified flanking CAPS markers WU95 and PRHA (4 cM apart). B, Fine mapping of VTC2 using SNP and InDel markers identified markers 424439 and 424446 in the Cereon Arabidopsis Polymorphism Collection (20 kb apart) as the closest flanking markers based on the available recombinants. C, Nine candidate genes between the SNP markers 424439 and 424446 were identified from the Col-0 sequence in GenBank. D, Mutations in vtc2-1, vtc2-2, vtc2-3, and vtc2-4 were identified by sequencing. Staggered lines represent the predicted exons and introns of the VTC2 gene. The 5' end of the gene is at the bottom.
Schematic of the map-based cloning process.

Col-0: AACATTCCTCAAGTTTGGTTA

Ler: AACATTCCTCTAGTTTGGTTA

SNP 442795

ATTTTCTA

Col-0: TTGATTTTCTATAAAGTAACT

Ler: AGTGAGGGGTCTACCTCCTGC

InDel 448516
d) Map and clone the mutation responsible for the observed phenotype.

FAST-FORWARD GENETICS

MAPPING AND IDENTIFICATION OF MUTATIONS BY DEEP SEQUENCING
Precise mapping to a region small enough to contain only a handful of candidate mutation
Confirming a mutation after mapping

• Variants matching the canonical EMS-induced G/C > A/T transition;

• Confirming the existence of the mutation by manual sequencing;

• Complementation: does the selected gene rescue the mutant phenotype?!?!?

• Additional alleles
Pros:

High efficiency of mutant generation (in plants easy to transform...)

Wide range of mutations (loss and gain of function, reduction of function)

It has been the most widely used mutagen for basic research

Cons:

Mutations can be very hard to map! Several mutants could not be mapped at all. Fast-forward genetics is expensive and results in a range of candidates.
b) generate a **representative** mutant population;

Several kinds of mutagens are available which differ for **efficiency** and **type** of mutations generated

<table>
<thead>
<tr>
<th>Mutagen</th>
<th>Main characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Chemical agents</td>
<td>Point mutations</td>
</tr>
<tr>
<td></td>
<td>Large range of mutations including loss and gain of function</td>
</tr>
<tr>
<td>(for example, EMS)</td>
<td>Very high efficiency with hundreds of mutations per genome</td>
</tr>
<tr>
<td>b) Physical agents</td>
<td>Big insertions, deletions and rearrangements</td>
</tr>
<tr>
<td>(Fast neutrons, X-rays and accelerated ions)</td>
<td>Mostly loss-of-function mutants</td>
</tr>
<tr>
<td>c) Biological agents</td>
<td>Insertion of specific DNA sequences</td>
</tr>
<tr>
<td>(Transposons and T-DNA)</td>
<td>Mostly loss-of-function mutants</td>
</tr>
<tr>
<td></td>
<td>Low efficiency</td>
</tr>
<tr>
<td></td>
<td>Easy identification of the mutations</td>
</tr>
</tbody>
</table>

Unfortunately, no efficient homologous recombination strategies have been developed in plants to date...
b) generate a **representative** mutant population;

Insertional mutagenesis: T-DNA

**T-DNA as a mutagen**

**Agrobacterium-mediated**

**T-DNA transformation**

**A simple T-DNA example**

**Bar:** encodes resistance to the herbicide gluphosinate ammonium, also known as **Basta**
Floral transformation of *Arabidopsis* using *Agrobacterium tumefaciens*

1. **T-0 plants**
2. **Agrobacterium suspension**
3. **Floral dip**
4. **Incubation for 48h in the dark**

- Unwrap, allow plants to grow and set seeds
- Let the plants finish the cycle: seeds are mature
- T-1 seeds are ready to screen for antibiotic/herbicide resistance
Generation of T-DNA tagged population of plants

Herbicide selection with Basta of T-1 transformants

Propagation and collection of seed

T-2 seeds ready for phenotype screening!
The insertion of a T-DNA element into an Arabidopsis chromosome can lead to many different outcomes:

- **Knock-out**
  - Location: coding region or promoter
  - Result: null
  - Effect: loss-of-function

- **Knock-down**
  - Location: promoter or 3’ UTR
  - Result: reduced
  - Effect: expression

- **Knock-on**
  - Location: promoter
  - Result: increased
  - Effect: expression gain-of-function
Workflow of a T-DNA screening

**Phenotype?**

- Yes
  - Clean-up line
    - Outcross
    - Find single insert line
    - Sequence both T-DNA/plant junctions
    - Find Homozygotes
  - Test Co-segregation of T-DNA with phenotype
  - Test additional alleles for similar phenotype
  - Complement with wild type copy of gene
  - Link biological function to the observed phenotype
    - Biochemistry
    - Physiology
    - Genetics

- No
  - OK, forget it!

Figure 4. Using Insertional Mutations to Understand Biological Function.
d) Map and clone the mutation responsible for the observed phenotype.

**TAIL-PCR**

- Long specific primers: P1 → P2 → P3
- Short AD primer
- Vector: insert end
- Nontargeted sequence

**Primary PCR with P1 and an AD primer**

- 5 high-stringency cycles
- 1 low-stringency cycle (annealing at 25 °C)

**TAIL-cycling** (as supercycles)

- 1 reduced-stringency cycle (annealing at 45 °C)
- 2 high-stringency cycles (annealing at 65 °C)

**Product yield**

- Specific (I)
- Nontargeted (II)
- Nonspecific (III)

- 1000-fold dilution

**Secondary PCR with P2 and the same AD primers**

- Product yield

- Specific (I): high
- Nonspecific (III): very low

- 1000-fold dilution

**Tertiary PCR with P3 and the same AD primer**

- Specific (I)
TAIL-PCR

I. PCR with SP1 and AD primers
II. PCR with SP2 and AD primers using an aliquot of Iary PCR as template
III. PCR with SP3 and AD primers using an aliquot of Iary PCR as template

Sequence IIIary PCR product and Blast against the genome: the region amplified is flanking the insertion
Pros:

High efficiency of mutant generation (in plants easy to transform...)

Tagged mutant, easy to map

Insertions in around 90% of genes can be achieved

Cons:

Complex integration patterns: transfer of vector sequences that flank the T-DNA; Multiple insertions; Tandem insertions; Rearrangements; Often it does not result in abolished/reduced expression of the disrupted gene.
b) generate a **representative** mutant population;

Insertional mutagenesis: Transposons

Most famous systems exploited in Arabidopsis (which lacks functional transposons) for generating populations of mutants employ functional transposons originating in maize:

- **Ac/Ds system: Activator /dissociation**
- **En/Spm system: Enhancer/ suppressor mutator**

Transposable element (TE, transposon or retrotransposon): is a DNA sequence that can change its position within the genome, sometimes creating or reversing mutations and altering the cell's genome size.
Ac/Ds system

50% of the transposition events are close to the donor site.
4 possible transposition outcomes

Figure 1. Schematic Diagram of the Ds Donor Site and Possible Transposition Events.

Open arrowheads indicate the 5′ and 3′ ends of the transposon. The Ds element carries the NPTII gene, which confers resistance to kanamycin (KanR), and a modified GUS reporter gene (Sundaresan et al., 1995). Possible transposition events include the following: (1) unlinked or loosely linked transposition to the same chromosome; (2) transposition to a different chromosome; (3) closely linked transposition; and (4) closely linked transposition disrupting the IAAH gene. Hatched boxes represent left and right T-DNA borders (T-DNA-LB and T-DNA-RB, respectively).
Ds transposons are “randomly” distributed

The positions of 312 Ds insertions are shown. Small arrowheads represent insertion sites. Insertions in the same BAC or P1 clone (average size, \( z100 \) kb) are stacked together to form a single column. Therefore, this insertion map manifests resolution equal to the average size of one BAC clone. The three major hot spots correspond to NOR2- and NOR4-adjacent regions and the region surrounding the DsG1 donor site.
Ds transposons show a preference for insertion into the 5’ region of genes
Pros:

• Transposon can be excised from the disrupted gene in the presence of a transposase, resulting in the reversion of the mutation
• Most transposition events occur preferentially in linked sites, a desirable trait if we are trying to mutate a specific gene and we have a transposon mutant with an insertion site close to that gene (Screening by PCR).
• Can be applied to organisms difficult to transform (cereals)

Cons:

• Most transposition events occur preferentially in linked sites (random saturation of a genome is difficult to achieve)
A collection of sequenced and mapped *Ds* transposon insertion sites in *Arabidopsis thaliana*

Surabhi Raina, Ramamurthy Mahalingam, Fuqiang Chen and Nina Fedoroff
Several transposon mutagenized populations ready-to-screen are available

Rice *Tos17* Insertion Mutant Database

This page is an *in silico* rice mutant screening page by BLAST search against flanking sequences from our mutant lines induced by rice retrotrans the mutant line identified by the screening up to 20 seeds/line. To search our mutant lines, your registration is required. The cost for distribution of request is 100 lines per year.

*Tos17*, T-DNA, and Ds insertions on Rice Genome IRGSP Build 3 (New version with RAP annotations, beta test) New

Search (e.g. AK119573 or Os01g0113300 or Chlorophyll, Click vertical black line on the *Tos17*(salmon) area)
Two factors are most critical for a forward genetic screening:

- Use of simple and direct screening procedures.
- The genetic background chosen: which wild-type? Which mutant?

Mutagenized how?
How the background is important for your screening!

Two ecotypes of Arabidopsis with different tolerance levels to salt! At 120 mM NaCl, Ler would be good to identify negative regulators: is there any mutants that survive? Shakdara would be good for positive regulators: is there any mutants that die?
- Enhancer screen identifies mutations which exacerbate (or enhance) a phenotype of interest in an already mutant individual.
- Second site suppression identifies mutations which alleviate or revert the phenotype of the original mutation.

Genetic modifier screening
How the background is important for your screening!

**Mutagenizing a mutant:**
- Enhancer screen identifies mutations which exacerbate (or enhance) a phenotype of interest in an already mutant individual.

[Image of wild-type, 1st mutant, and enhancer line with annotations showing genetic interactions and phenotypic changes.]
Example of an enhancer screening in maize

1<sup>st</sup> mutant       enhancer line
Mutagenizing a mutant:
- Second site suppression identifies mutations which alleviate or revert the phenotype of the original mutation

How the background is important for your screening!

Wild-type  1\textsuperscript{st} mutant  suppressor line
How the background is important for your screening!

**Diagram:**
- **RD29A promoter:**
  - Active: WT
  - Silenced: ros

**RdDM pathway:**
- ROS1 & ROS3
- New enzymes & regulators?

**Images:**
- Luminescence (4°C, 1 day)
How the background is important for your screening!
Modified screenings: Activation Tagging

T-DNA insertion mutants that do not result in loss- or gain-of-function

An et al., 2005 PMB 59, 111
A mutant isolated through activation tagging
Modified screenings: Enhancer Trap Screening

T-DNA insertion mutants that do not result in loss- or gain-of-function
Modified screenings: Promoter Trap

T-DNA and transposon mutants that do not result in loss- or gain-of-function
Expression Patterns of Promoter Trap Lines

An example of screening: mutant that affect gametophyte will have a distorted marker segregation ratio.
An example of “screening”: natural variation

Natural variation can be exploited to identify alleles mutated in different ecotypes/variety which are responsible for different phenotypic traits.

Highly variable traits in Arabidopsis ecotypes
• Flowering time;
• Resistance to fungal pathogens;
• Leaf shape;
• Light responses;
• Root development;
• Secondary metabolism.

Genes underlying natural variation can be isolated by map-based cloning using sequence polymorphism between accessions.

One of the 2 paper discussions will be focused on the DRO1 gene of rice influencing drought tolerance, isolated through exploitation of natural variation.