



Central European Institute of Technology
BRNO | CZECH REPUBLIC

MUTANTS AND THEIR APPLICATION IN GENOMICS

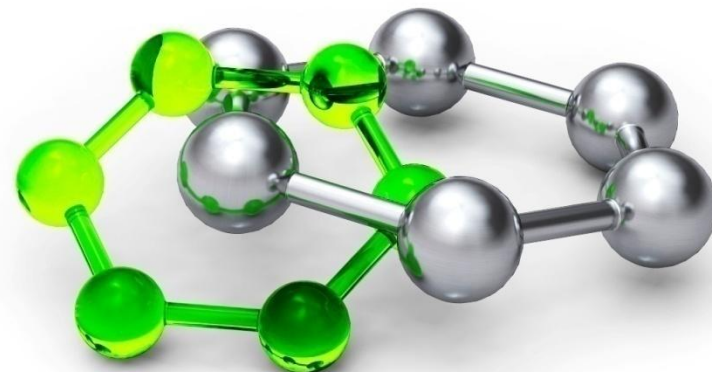
Methods in genomics and proteomics
(CG980)

Genomics – lesson 6

Mgr. Markéta Žďárská, Ph.D.



INVESTMENTS IN EDUCATION DEVELOPMENT



- Transgenic plants and functional genomics
- Mutagenesis, mutagen types
- Forward genetics approaches
 - EMS mutagenesis
 - Positional cloning
 - Sequencing
- Reverse genetics approaches
- Tilling
- Reporter genes

- **transgene** – a gene (genetic material) that has been transferred naturally, or by any of a number of genetic engineering techniques from one organism to another
- synthetic, modified or heterogeneous genes introduced into a different animals and plants
- „non-native“ segment of DNA
 - retains the ability to produce RNA/protein in the transgenic organism
 - alters the normal function of the transgenic organism's genetic code
- Transgenic organisms are used for the study of gene functions

Transgenic organisms



- GloFish are a type of transgenic zebrafish (*Danio rerio*) that have been modified through the insertion of a green fluorescent protein (*gfp*) gene.

- ◎ aim: search for genes and determination of their function in genome
- ◎ 2 different approaches:
 - > „**forward genetics**“
 - phenotype → gene
 - > „**reverse genetics**“
 - Sequence of DNA (gene) → phenotype

- mutants – a tool in both approaches
- different type of mutagens
 1. chemical
 2. physical (radiation)
 - a huge number of random mutations, affecting all genes
 3. biological
 - modern
 - insertion mutagenesis, lower number of mutations, leave a molecular marker behind

} **classical**

◎ cause point mutations

1. during DNA replication but also in non-replicating DNA

- > **alkylating agents** (transfer an alkyl group to nucleobases; random)
 - **Sulfur mustard** (mustard gas; Ch. Auerbach)
 - **ENU - N-ethyl-N-nitroso urea**
 - ethyl group of ENU interacts usually with thymine
 - Bill Russell (1951) mouse strain („T-test stock”) used in genetic screens for testing mutagens such as radiations and chemicals



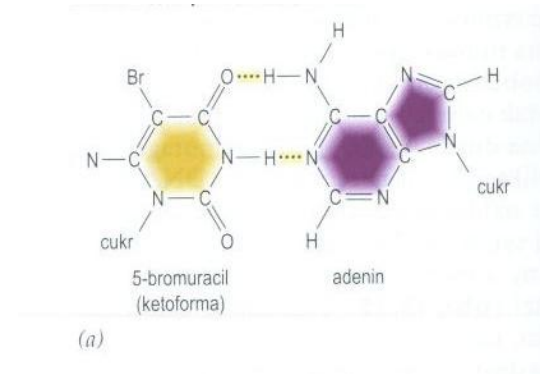
- **Ethyl methane sulfonate (EMS)**
 - ethyl group of EMS reacts with guanine in DNA, forming the abnormal base O-6-ethylguanine (original G:C can become A:T)
 - Maple J. & Moller SG, 2007 - Mutagenesis in *Arabidopsis*
- > **HNO₂** (deamination of amino groups)

Chemical mutagens

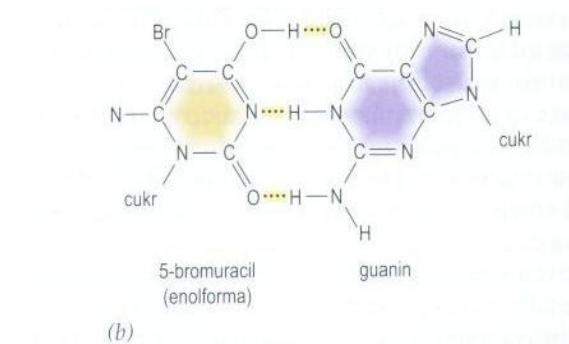
2. only during replication

- > **base analogs:** 5-bromuracil (5-BU), 2-aminopurine
 - cause transition mutations
- > **acridine dyes:** proflavine, acridine orange
 - cause addition or deletion of 1 or more bases → change in frame read → nonfunctional gene products
- > **hydroxylamine**
 - specific – transition induced only in direction G:C → A:T

5-BU pairing with adenine

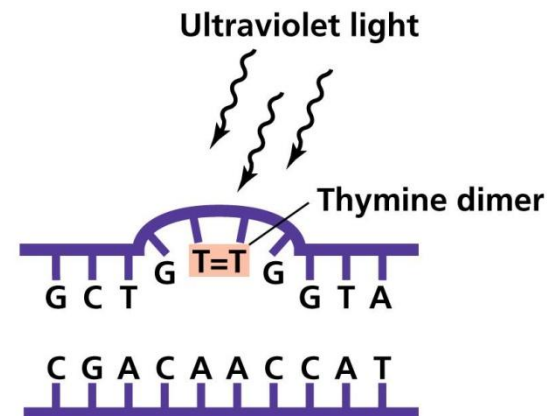


5-BU pairing with guanine



Physical mutagens

- ① ionizing: X-ray, gamma radiation, radioactive carbon ^{14}C
 - > Cause DNA molecules break downs
- ① nonionizing: UV light 254nm is absorbed by bases → pyrimidine dimers (distorting sugar phosphate backbone)



Copyright © 2006 Pearson Education, Inc., publishing as Benjamin Cummings.

- ① cause extensive insertions and chromosome alterations
- ① able to delete more genes or insert new regulations sequences
- ① improper for precise mutagenesis

◎ insertion mutagenesis

1. T-DNA

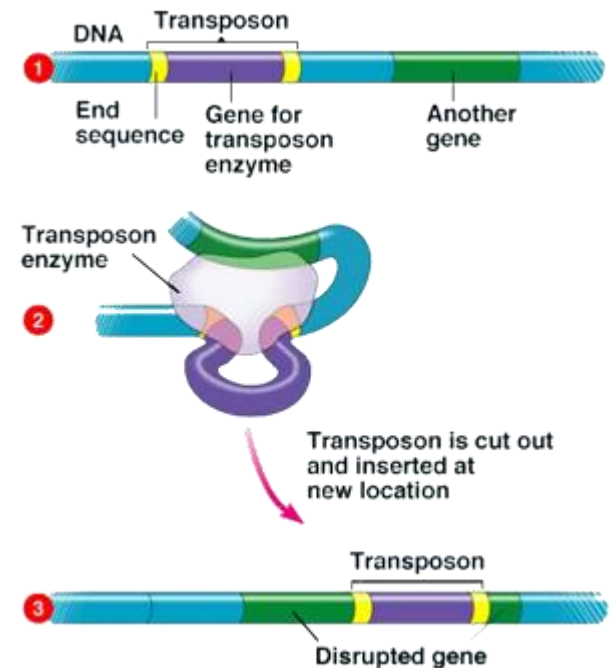
- > *Agrobacterium tumefaciens* is able to insert a part of DNA into plant genome
- > Results of T-DNA insertion into genome are various, caused by insert nature or location of insertion
- > effects:
 - Gene inactivation
 - activation (carrying a promoter, enhancer)

2. Transposons — transposable gene elements (TE or transposon)

- > Mobile genetic elements, less stable than T-DNA
- > Can jump from original location of insertion → recovery of normal phenotype
- > The footprint of insertion stays in the location also if the insertion is no longer in the genome; inappropriate for large mutagenesis (Petersen et al, 2000)

T-DNA a transposons

- ① Insertion mutagenesis
- ① Insertion into:
 - > coding region
 - > noncoding region – affecting intron splicing, gene expression
- ① 😊 pros:
 - > reversible mutation
 - > Easy to map and the region is easy for cloning
- ① ☹️ cons:
 - > Insertion is not random

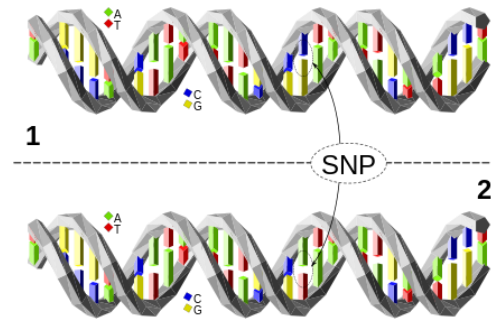


usually nonfunctional gene

How to search for mutations?

1. small mutations

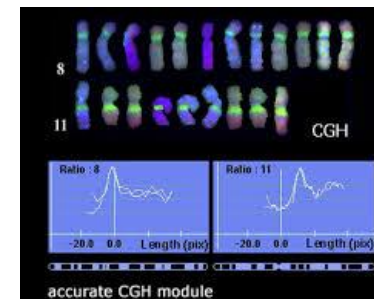
- a new **single nucleotide polymorphism (SNP)** at a specific position in the genome
- altered transcriptional profile



The upper DNA molecule differs from the lower DNA molecule at a single base-pair location (a C/A polymorphism).

2. chromosome alterations

- extensive reconstruction → *in situ* hybridization at metaphase chromosomes – resolution 5 Mbp
- reconstruction smaller → Comparative Genome Hybridization (CGH)
 - „DNA microarrays“ using DNA probe covering diverse sites of a genome
 - hybridization with genomic DNA – resolution 5-10 Mbp
 - a full genome analysis of in 1 experiment



- ① based on saturation mutagenesis „saturation screen“
 - > a mutagen induction to an organism → analysis of progeny for a specific phenotype
 - > identification of mutants → sorting into complementary groups
 - > mapping into general chromosomal locations employing known markers and then cloning, sequencing
- ① a mutation for every locus exists → possible to determine a groups of genes responsible for an exact trait
- ① aim is to achieve a saturation point → detect all genes responsible for a phenotype
- ① mutagens (RTG, EMS, transposons)
 - > Examples:
 - > plants missing the reaction to light
 - > bacteria unable to growth in the presence of some sugars,...

Identification of a mutated gene in a mutant line selected by its phenotype



◎ using genetic map

1. mapping - (co)segregation analysis

- > finding approximate position of a gene in a genetic map, based on genetic linkage with **genetic markers** (traits with polymorphism = they are divergent between parental genotypes)

2. search for an exact sequence carrying a mutation

- > **chromosome „walking“**
- > **sequencing**, matching with WT sequence

= trait with known (easily checked) position in the genetic map, featuring polymorphism (divergent between parental genotypes)

1. Morphologic

2. Molecular

- DNA markers – able to detect differences in a sequence
 - DNA with known and explicit location in genome
 - easily detected loci with known position at the chromosome
 - single nucleotide polymorphism (SNP)
 - ideal – equally localized

Natural morphologic variability of *Arabidopsis* – ecotypes „accessions“

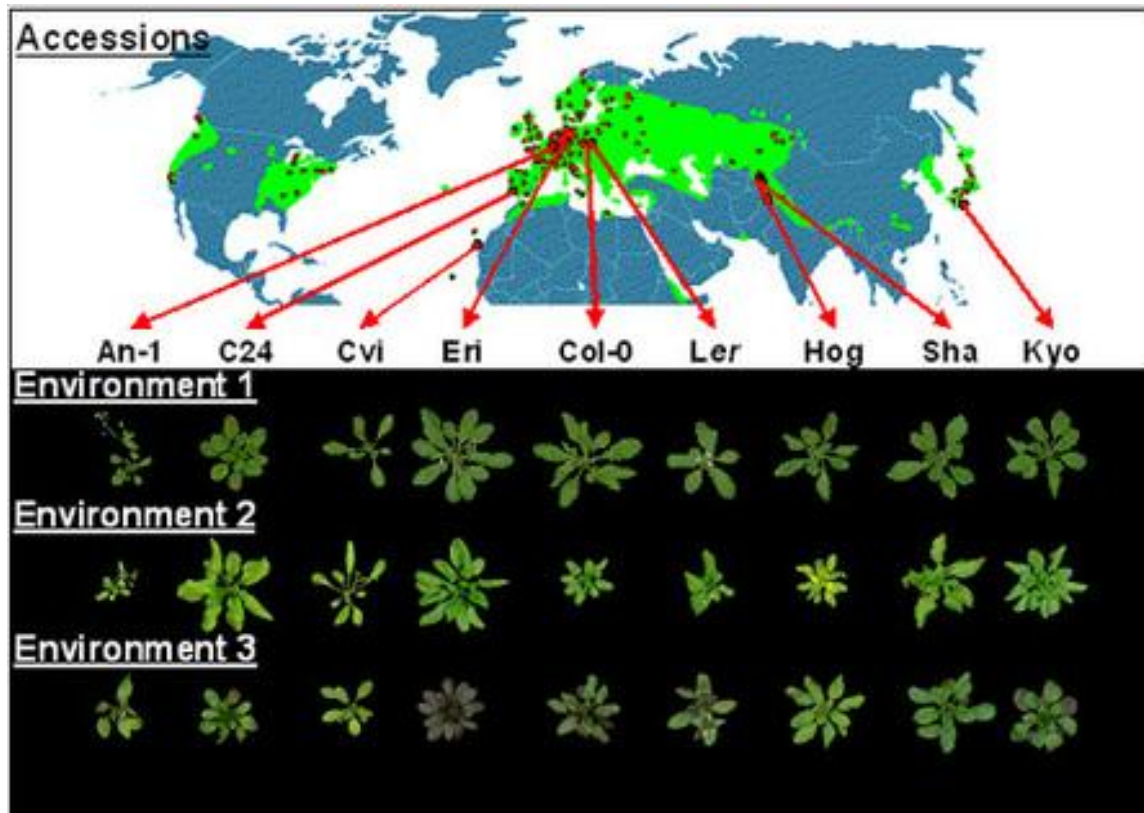


Fig. 1: Geographical distribution of *Arabidopsis thaliana* (green area on the world map) and overall phenotype of the rosette of a subset of accessions grown under 3 contrasted environment scenario.

<http://www.mpipz.mpg.de/102840/reymond>

DNA molecular markers

(= a visible band at electrophoretic gel or blot)



- ◎ **SSLP** (Simple Sequence Length Polymorphism)
 - > a genome length (PCR products) amplified using spec. primers
- ◎ **RFLP** (Restriction fragment length polymorphism) + Southern
 - > restriction fragments lengths of a genome segment, PCR followed by genome DNA cleavage and adaptors ligation
- ◎ **RAPD** (Random amplified polymorphism detection)
 - > a length of randomly amplified genome segments (short primers 8-10bp)
- ◎ **AFLP** (Amplified fragment length polymorphism)
 - > a length of genome fragments, PCR followed by cleavage of genome DNA a adaptors ligation

Other Markers	Acronym
Variable Number Tandem Repeat	VNTR
Oligonucleotide Polymorphism	OP
Inverse Sequence-tagged Repeats	ISTR
Inter-retrotransposon Amplified Polymorphism	IRAP

DNA molecular markers *Arabidopsis thaliana*

- crossing 2 ecotypes: Columbia a Landsberg *erecta* (Col X Ler)
- recombinant map contained originally 67 markers (Lister & Dean, 1993)
- today more than 1300 markers (Hou et al, 2010)



Arabidopsis thaliana physical map with indication of the positions of the markers.

Positional cloning, map-based cloning



- gene isolation based on the **position** on the **map**
- gene function is usually unknown
- gene mapping based on **genetic linkage with molecular marker** followed by a gene isolation with approximate chromosomal location - this is known as the candidate region.
- need for a standard (WT) line, which is crossed with a mutant line to proceed with a recombinant analysis using markers (“linkage analysis”)
- **aim of positional cloning:**
 - find a gene with desired mutation located in the interval of two closest markers
 - region small enough → to choose a candidate gene and identify a mutation in it
 - generally – complicated and time consuming

Positional cloning - a scheme

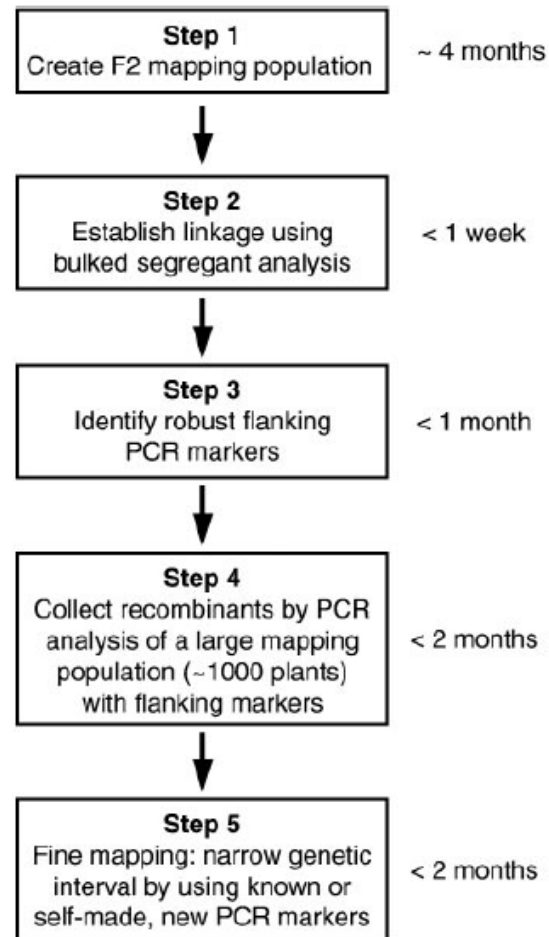


Figure 2. Procedure of a typical fine mapping experiment. The time that should be calculated to complete each of the steps, assuming no complications are encountered, is shown to the right.

Lukowitz et al, 2000

◎ ***Arabidopsis***

- > a map of positional cloned genes contains approx. 620 mutant genes with a phenotype (Meinke et al, 2003)
- > a number of genes gained by positional cloning is increasing every year → clarification of new gene functions in a model plant

◎ **Wheat (*Tritium aestivum*)- crops**

- > molecular markers:
 - SSR (Simple Sequence Repeat)
 - SCAR (Sequence – Characterized Amplified Region)
- > gene isolation with markers is more complicated – hexapodies
- > a lot of unknown genes coding crop related traits (up to now approx. 20), i.e. fungi and their resistance

◎ **Genetic disorders** (a genetic problem caused by one or more abnormalities in the genome, especially a condition that is present from birth (congenital))

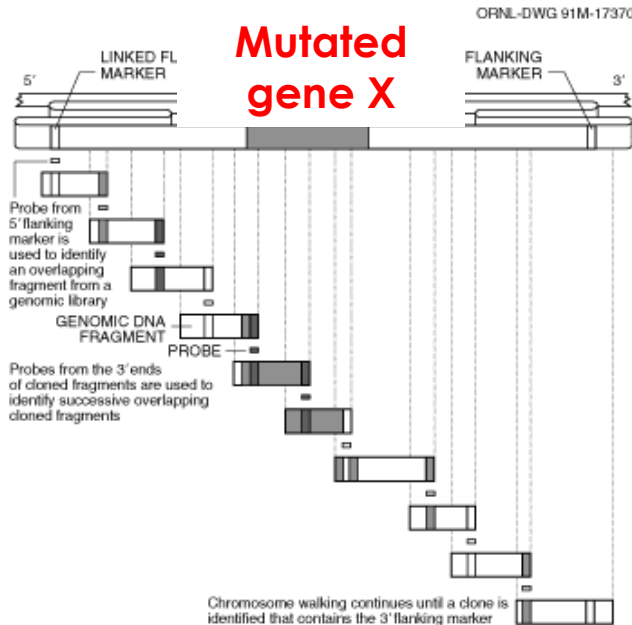
- > **Huntington's Disease** (no signs until adulthood; autosomal dominant)
- > **Down syndrome (DS or DNS)**, also known as **trisomy 21**
- > **Cystic fibrosis (CF)** (affects mostly the lungs but also the pancreas, liver, kidneys, and intestine; autosomal recessive)

Identification of mutant gene – an exact sequence carrying a mutation



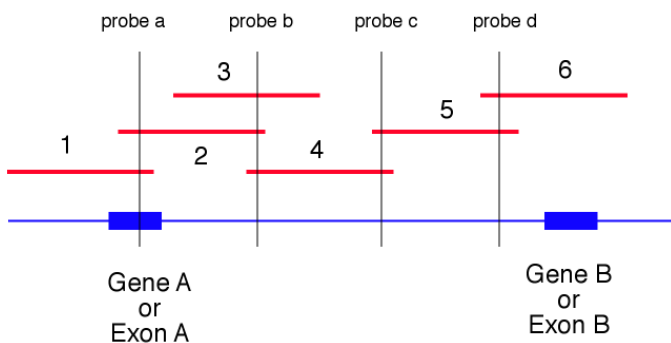
- ◎ **allelic test** - crossing a desired mutant line with a „knock-out“ candidate gene – check if the mutant phenotype is maintained
- ◎ **molecular complementation** (recessive mutations)
 - > transformation of mutant plant with sequences of WT DNA in restricted space in order to detect the one recovering the WT phenotype
- ◎ **analysis of entire DNA sequence in restricted genetic interval** → search for alterations causing the mutation
 - > SSCP(Single Stranded Conformational Polymorphism)
 - > HMA (Heteroduplex Mobility Assay)
 - > DGGE (Denaturing Gradient Gel Electrophoresis)
 - > dHPLC (denaturing High Performance Liquid Chromatography)
 - > a DNA chip hybridization
 - > “pyrosequencing”
 - > „chromosome walking“
- ◎ **sequencing**

„Chromosome walking“



- search for 2 optimal markers surrounding mutated gene
- libraries of huge genome DNA fragments (redundant; random):
 - YACs, BACs
 - = yeast (bacterial) artificial chromosome, ~ 300 (100) kbp
- search for overlaps based on hybridization

Walking from Gene A or Exon A to Gene B or Exon B



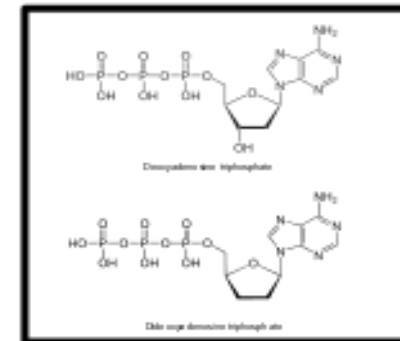
Genomic DNA is shown in blue. Selected clones from a library of cloned genomic DNA fragments are shown in red. The initial probe, probe a, is specific to gene A or exon A and allows identification of clones 1 and 2. A new probe, probe b, is prepared from one end of clone 2 and used to isolate new clones 3 and 4 from the genomic library. Probe c, prepared from clone 4 is used to identify clone 5, etc. The orientation of the clones is determined by restriction mapping of the clones. Clone 6 contains the desired gene B or exon B.

- ⦿ based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in vitro DNA replication
- ⦿ Developed by Frederick Sanger and colleagues in 1977
- ⦿ requires a single-stranded DNA template, a DNA primer, a DNA polymerase, normal dNTPs, and modified ddNTPs), the latter of which terminate DNA strand
- ⦿ The DNA sample is divided into four separate sequencing reactions, containing all four of the standard deoxynucleotides (dATP, dGTP, dCTP and dTTP) and the DNA polymerase
- ⦿ due to length of reading frames (**700 to 800 bp**) – still **commonly used and precise**

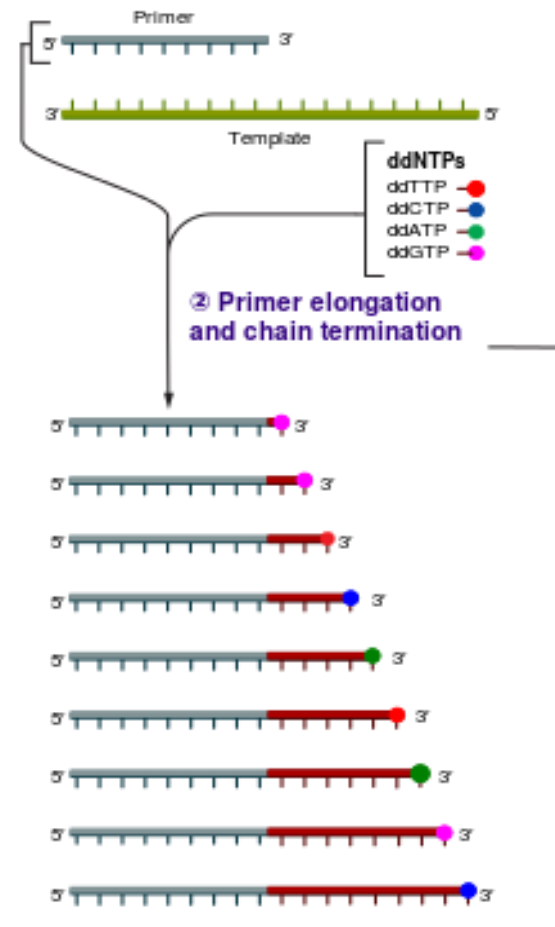
Sequencing - classic - Sanger

① Reaction mixture

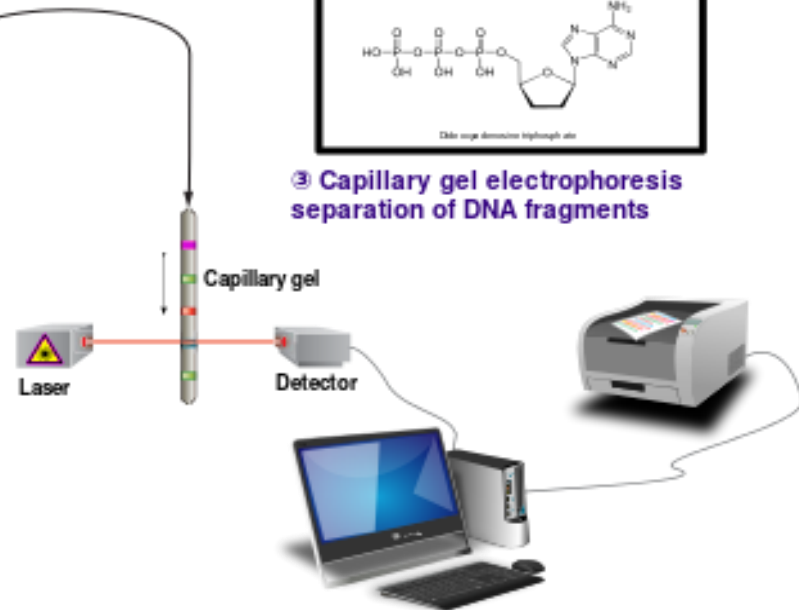
- ▶ Primer and DNA template
- ▶ DNA polymerase
- ▶ ddNTPs with flouochromes
- ▶ dNTPs (dATP, dCTP, dGTP, and dTTP)



③ Capillary gel electrophoresis separation of DNA fragments



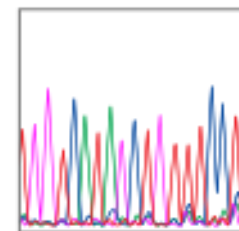
② Primer elongation and chain termination



Laser

Capillary gel

Detector



Chromatograph

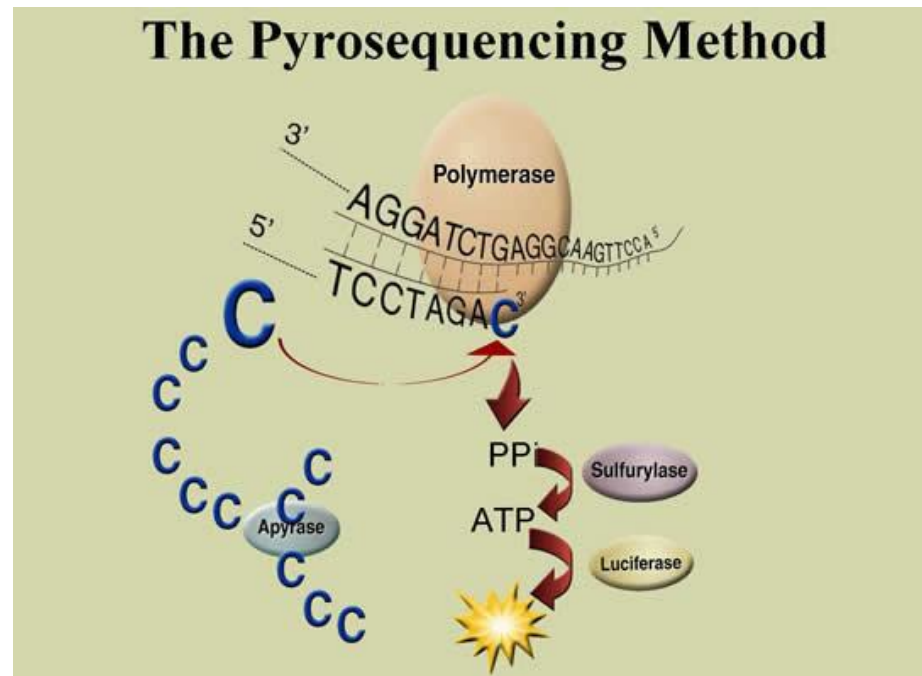
④ Laser detection of flouochromes and computational sequence analysis

Pyrosequencing –

„Next-generation sequencing“



- DNA sequencing (determining the order of nucleotides in DNA) based on the "**sequencing by synthesis**" principle
- relies on the **detection of pyrophosphate release on nucleotide incorporation**, rather than chain termination with dideoxynucleotides; no need for electrophoresis
- developed by **Mostafa Ronaghi** and **Pål Nyrén** at the Royal Institute of Technology in Stockholm in 1996

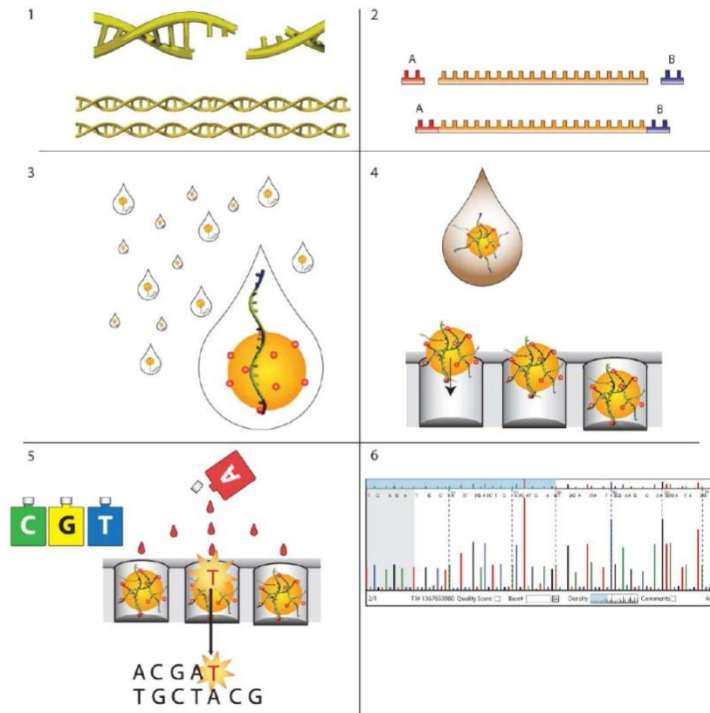


„Next-generation sequencing“



454 - 2005

- emulzní PCR
- pyrosekvenování



454 Genome Sequencers

FLX System

- 1 million of reads/run
- 400-650 bp/read
- 2 přístroje v ČR



GS Junior

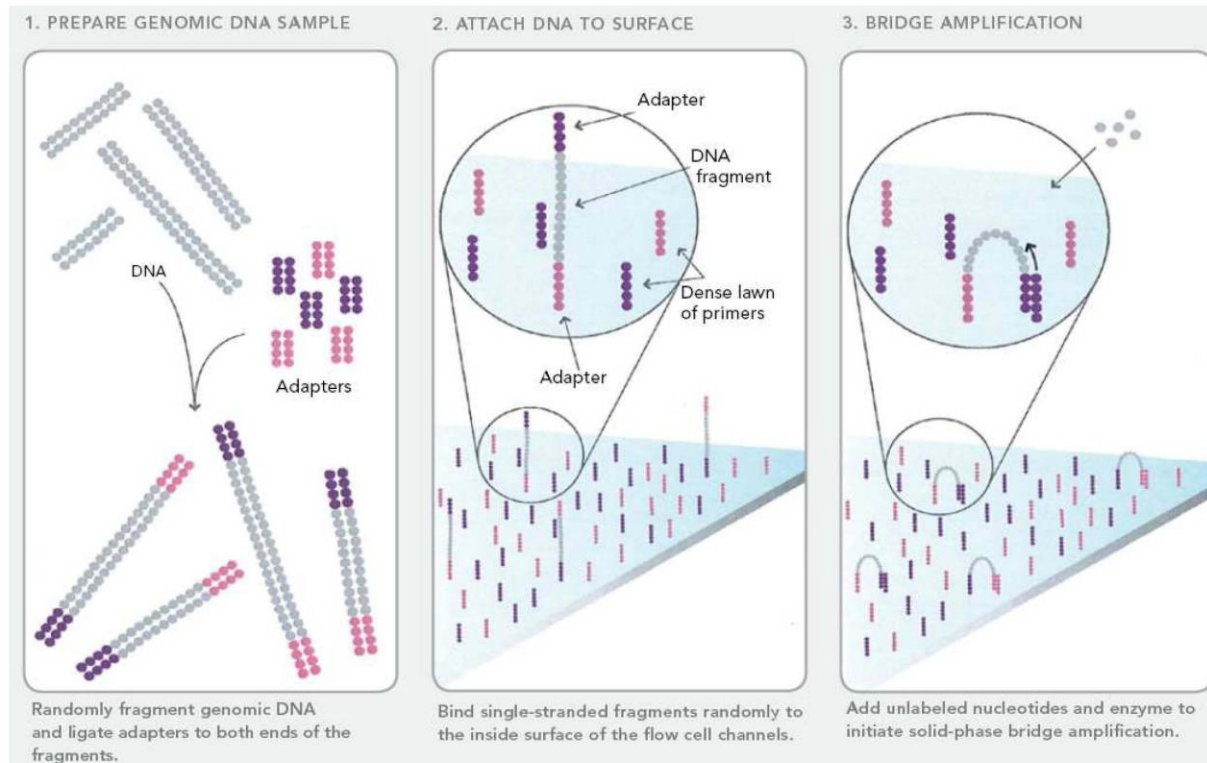
- 0.1 millions of reads/run
- 400 bp/read

„Next-generation sequencing“

Solexa (Illumina) - 2007



- můstková „bridge“ PCR
- sekvenování pomocí DNA syntézy



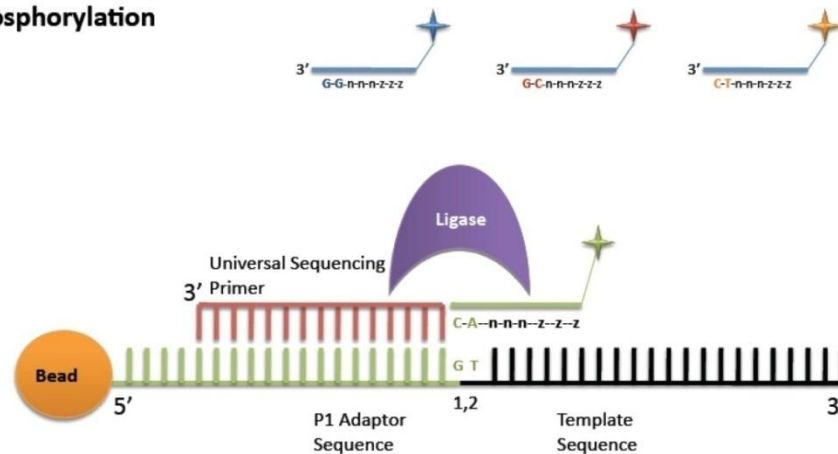
„Next-generation sequencing“

SOLiD (2008)



- emulzní PCR,
- sekvenování pomocí ligace

**First Cycle with primer 1-
Hybridization, Ligation, de-
Phosphorylation**



„Next-generation sequencing“



Platform	Year	Sequencing Method	Amplification	Detection	Features
454	2005	Pyro-sequencing	Emulsion PCR	Light	First NGS
Illumina	2007	Synthesis	Bridge PCR	Light	90% of Market
SOLiD	2008	Ligation	Emulsion PCR	Light	Lowest Error Rate
Ion Torrent	2010	Synthesis	Emulsion PCR	Hydrogen Ion	Semiconductor Chip
Pacific Biosciences	2010	Synthesis	None = Single Molecule	Light	Anchored Polymerases
Oxford Nanopore	2012	Nanopore	None = Single Molecule	Electrical Conductivity	“Run Until” Sequencing

Modified from T. C. Glenn. 2011. Field guide to next-generation DNA sequencers. Molecular Ecology Resources 11: 759-769.

„Next-generation sequencing“



Instrument	Run time	Millions of Reads/run	Bases / read	Yield MB/run
3730xl (capillary)	2 hrs	0.000096	650	0.06
PacBio RS	2 hrs	0.01	860 – 1,500	5-10
454 GS Jr. Titanium	10 hrs	0.1	400	50
Ion Torrent – 314 chip	2.5 hrs	0.25	200	50
454 FLX Titanium	10 hrs	1	400	400
454 FLX+	20 hrs	1	650	650
Ion Torrent – 316 chip	3 hrs	1.6	200	320
Illumina MiSeq	26 hrs	4	150+150	1200
Ion Torrent – 318 chip	4.5 hrs	4	200	800
Illumina GAIIx	14 days	300	150+150	96,000
SOLiD – 5500xl	8 days	>1,410 ^d	75+35	155,100
Illumina HiSeq 1000	8.5 days	≤1500	100+100	≤300,000
Illumina HiSeq 2000	11.5 days	≤3000	100+100	≤600,000

2012. NGS Field Guide (www.molecularecologist.com)

◎ **whole genomes**

- > resequencing → search for variability in human population
- > *de novo* sequencing → non-model organisms

◎ **transcriptome (RNA-seq)**

- > identification of unknown transcripts
- > more precise than „microarrays“

◎ **targeted** → to a specific part of a genome or focused on selected group of genes

- > random genome regions → based on length after restriction cleavage of genome DNA (RAD-seq)
- > hybridization to approx. 100bp probes with DNA fragments that are sequenced (Hyb-seq)

Large-scale random mutagenesis and „screening“



- systematic mutagenesis – gradual
- EMS mutagenesis
 - search for a phenotype (forward), gene of interest with **alterations in nucleotides**
 - common „screen“ of 1000 or 10000 individuals
 - use of PCR gene of interest → search for mild changes in **PCR product migration** on a gel or a column
 - be aware that all changes are not covered by a „knockout“ gene, some of them can be “silenced”- present at non-essential AK positions
 - methods commonly used:
 - **DHPLC** – “Denaturing High Performance Liquid Chromatography“
 - **DGGE** – „Denaturing Gradient Gel Electrophoresis“
 - **SSCP** – „Single-Stranded Conformation Polymorphism“

A scientific example of EMS mutagenesis (Feraru et al, 2010)



- A fluorescence imaging-based forward genetic screen“
 - As a tool for the screening to identify novel components of plant intracellular trafficking, they used a well characterized plant cargo, **the auxin efflux carrier PIN1** (Petrasek et al., 2006).
 - With this strategy, they aimed to identify novel regulators at different stages of subcellular protein trafficking.
 - **EMS-mutagenized PIN1pro:PIN1-GFP** (for green fluorescent protein) population using epifluorescent microscopy for seedlings displaying aberrant PIN1-GFP distribution in the root.
 - From 1500 M1 families, they identified several **protein affected trafficking (pat)** mutants defining three independent loci (mapping with simple sequence length polymorphism (SSLP) and cleaved amplified polymorphic sequence (CAPS and dCAPS) markers
 - The At3G55480 candidate gene was sequenced and a point mutation that caused a stop codon was found at the position 705 downstream of ATG)

A scientific example of EMS mutagenesis (Feraru et al, 2010)

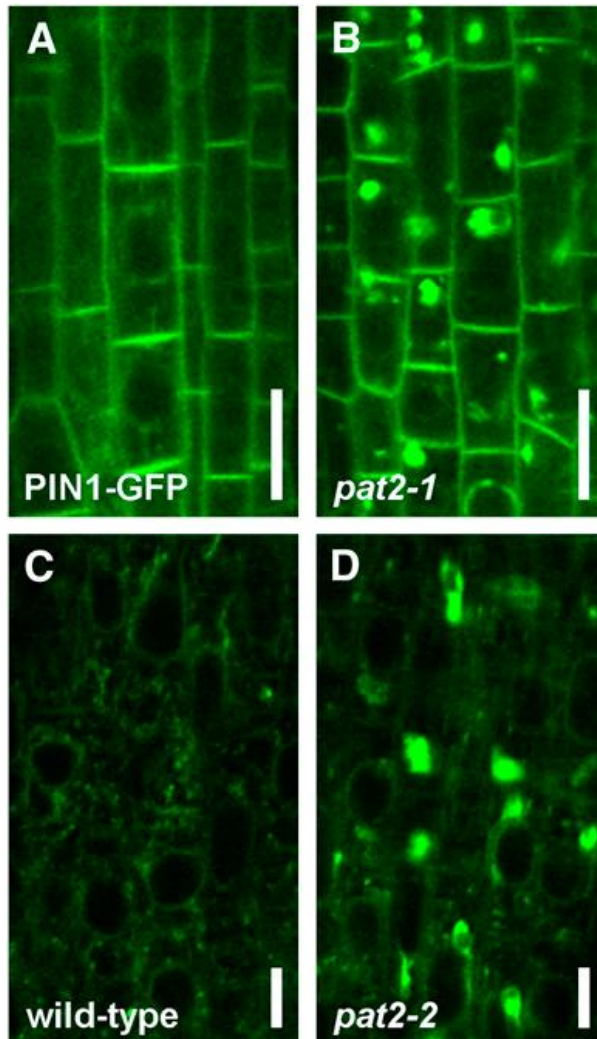


Figure 1. The *pat2* Mutant Displays Ectopic Intracellular Protein Accumulation.

(A) to (D) Both PIN1-GFP ([A] and [B]) and aleurain-GFP ([C] and [D]) accumulate intracellularly in *pat2-1* (B) or *pat2-2* (D) root cells compared with control ([A] and [C]).

- *pat2* mutant lytic vacuoles display altered morphology and accumulation of proteins
- unlike other mutants affecting the vacuole, *pat2* is specifically defective in the biogenesis, identity, and function of lytic vacuoles but shows normal sorting of proteins to storage vacuoles
- PAT2 encodes a putative β -subunit of adaptor protein complex 3 (AP-3)
- AP-3 β functions in mediating lytic vacuole performance and transition of storage into the lytic vacuoles independently of the main prevacuolar compartment-based trafficking route

„Reverse“ genetics



- today → post-genomic era
- genes are known (sequences)
- unknown
 - functions of genes
 - usually >50% predicted genes in eukaryotes
 - phenotypes causing mutations in these genes

TILLING

(Targeting induced local lesions in genomes)



- ◎ to create libraries of mutagenized individuals that are later subjected to high-throughput screens for the discovery of mutations
 - > potential changes in regulation, interaction, ...
- ◎ introduced on *Arabidopsis thaliana* (McCallum et al, 2000)
- ◎ Principle
 - > random induction of point mutations(EMS)
 - > followed by search of lines with mutations in targeted gene using PCR and heteroduplex analysis

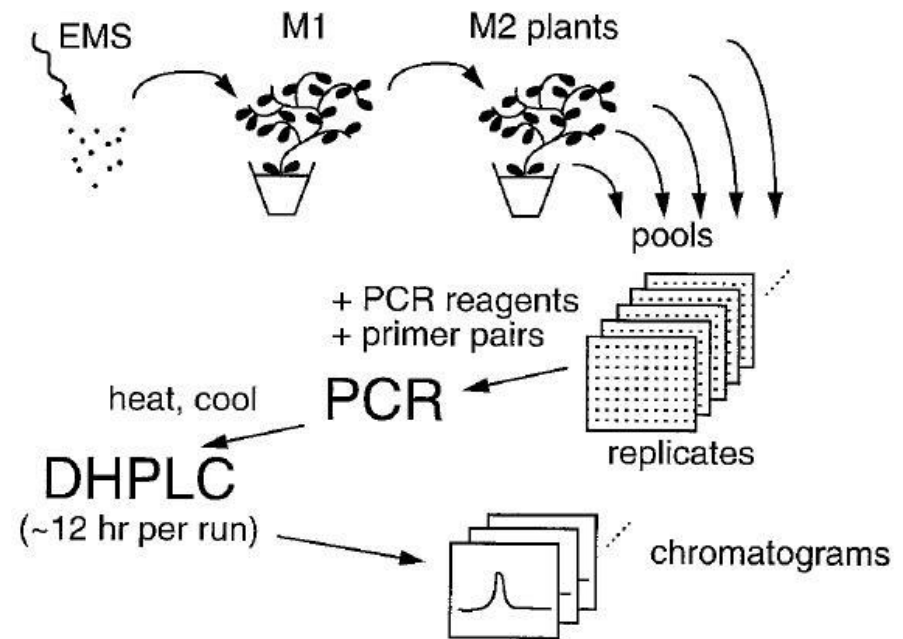


Figure 1. Schematic depicting the TILLING strategy applied to a plant such as Arabidopsis.

(McCallum et al, 2000)

TILLING – detection of mutations, strategy

1 Seeds are mutagenized to induce point mutations throughout the genome.



2 A founder population is grown from mutagenized seeds.



3 Founder population is self-fertilized to produce a crossed population.



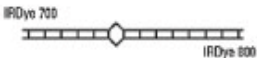
4 Seeds from the crossed population are stored and DNA samples are collected in 96-well plates.



5 Up to eight 96-well plates are pooled into one and the samples (768) subjected to PCR with two gene-specific primers labeled with different IRDye® infrared dyes.



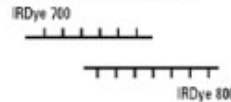
6 Resulting amplicons are heated and cooled, resulting in heteroduplexes between wild type and mutant samples.



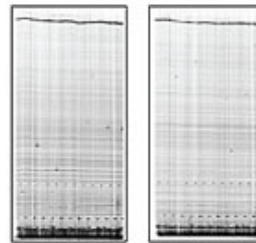
7 CEL I nuclease is used to cleave at base mismatches.



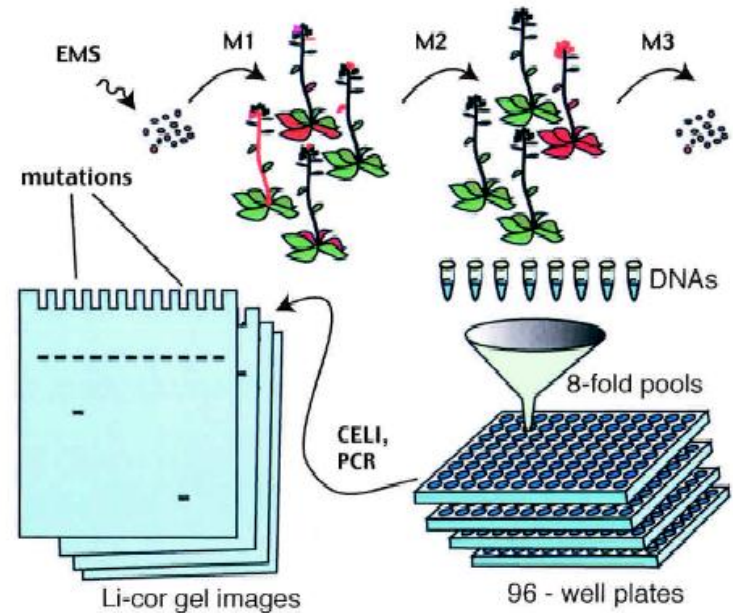
8 Samples are denatured and electrophoresed on a LI-COR 4300 DNA Analysis System.



9 In lanes that have a mutation in the pool, a band will be visible below the wild type band on the IRDye® 700 infrared dye image. A counterpart band will be visible in the same lane on the IRDye® 800 infrared dye image. This band is the cleavage product labeled with IRDye® 800 infrared dye from the complementary DNA strand. The sum of the length of the two counterpart bands is equal to the size of the amplicon, which makes it easy to distinguish mutations from amplification artifacts.



10 After detection of a mutation in a pool (lane), the individual DNA samples in the pool are screened again to find out which of the eight pooled samples from the crossed population has the mutation.



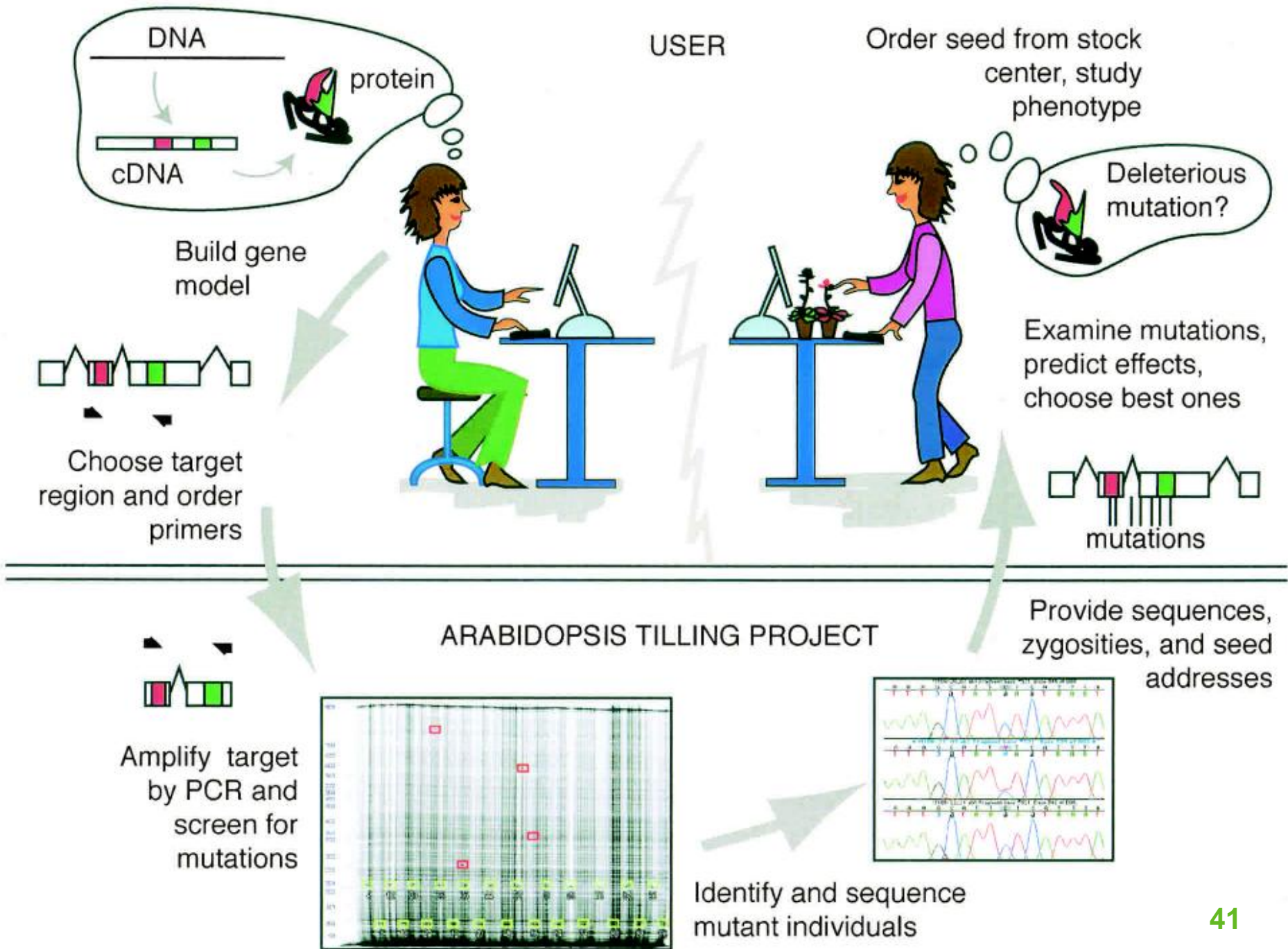
○ TILLING: A five step process

1. You decide whether your gene is worth TILLING.

- **"I have an insertion in my gene but the knockout phenotype is lethal."**TILLING can provide the sub-lethal phenotypes you want.
- ***"The knockout phenotype is interesting."***.....TILLING can provide an allelic series that may help you better ascertain the function of your gene.
- **"I have an insertion in my gene that knocks out gene function but my plants have no phenotype."**.....TILLING is not for you. In this scenario, a gain-of-function mutation is needed to investigate the potential *in vivo* role of this gene. The large majority of phenotypes arising from our populations will cause full or partial loss of function.
- ***"I have a candidate gene and I want to know the knockout phenotype."***....There are very good reasons why you should start by insertional mutagenesis rather than by TILLING. First, only a small percentage of EMS-induced mutations will yield a change likely to truncate the protein (~5%). Second, the Arabidopsis community has access to excellent insertional mutagenesis resources. Third, if a knockout mutation causes no phenotype, then the TILLING allelic series is not expected to either.

- TILLING: A five step process

2. You find the best the region to be targeted and place your order.
3. ATP screens the region for mutations.
4. ATP sequences the mutation and enters it in our public database.
5. ATP sends you a mutant report and you order seed.



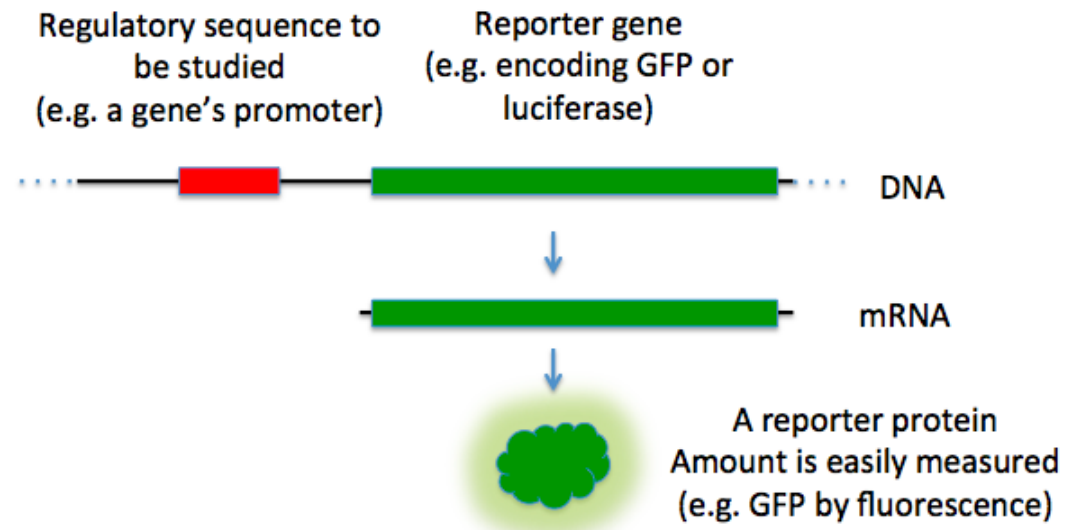
- Several TILLING centers exist over the world that focus on agriculturally important species:
 - Rice – UC Davis (USA)
 - Maize – Purdue University (USA)
 - *Brassica napus* – University of British Columbia (CA)
 - *Brassica rapa* – John Innes Centre (UK)
 - *Arabidopsis* – Fred Hutchinson Cancer Research
 - Soybean – Southern Illinois University (USA)
 - [Lotus and Medicago – John Innes Centre \(UK\)](#)
 - Wheat – UC Davis (USA)
 - Pea, Tomato - INRA (France)
 - Tomato - University of Hyderabad (India)

Reporter genes



- a gene attached to a regulatory sequence of another gene of interest in bacteria, cell culture, animals or plants
- **reporters** – used as an indication of whether a certain gene has been taken up by or expressed in the cell or organism population

- fluorescent (GFP)
- β -galactosidase (GUS)



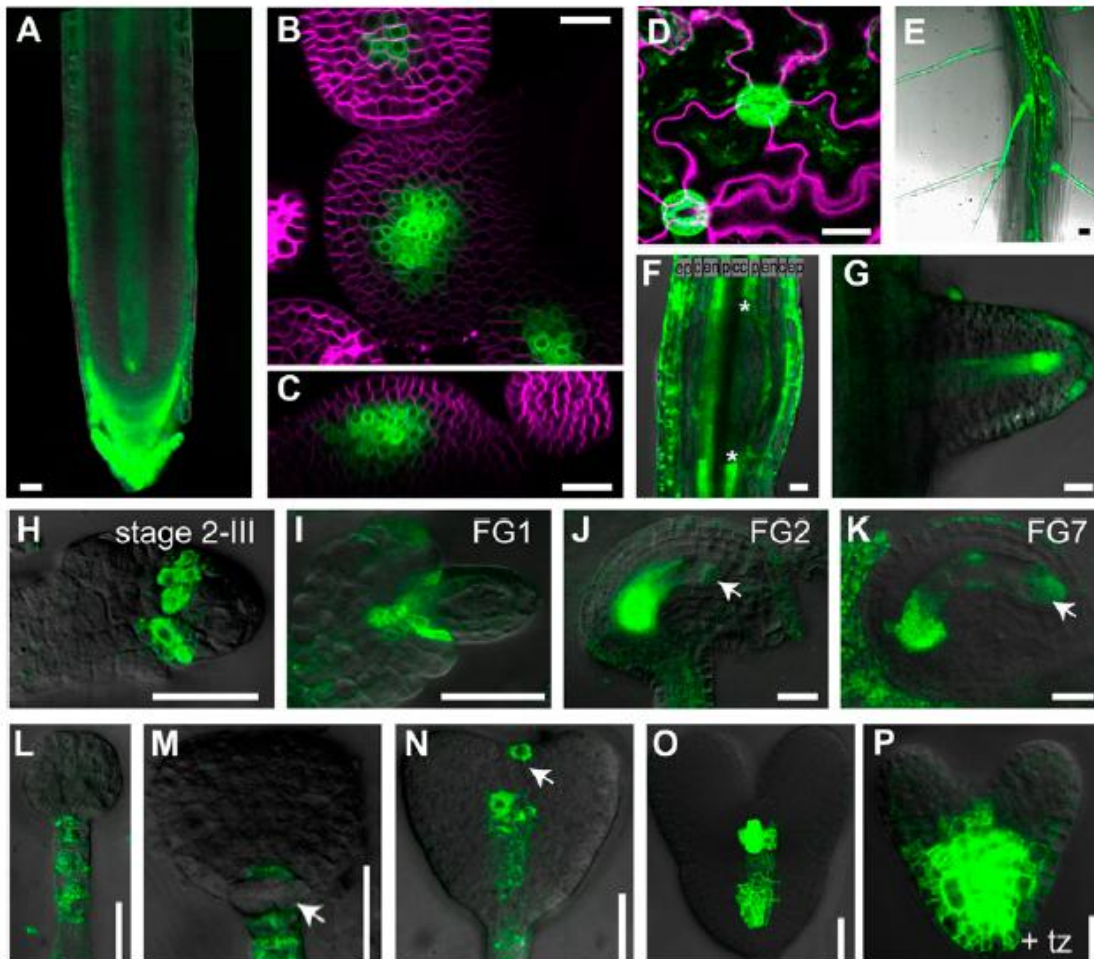


Figure 4. *TCSn::GFP* expression in different developmental contexts. A, Primary root meristem of 5-d-old seedling. B, Top view of shoot apical meristem. C, Side view of shoot apical meristem. D, Pavement cells and guard cells. E, Primary seedling root with root hairs. F, Lateral root primordium, early stage. Asterisks delineate lateral root primordium founder cells of pericycle that down-regulate MSP output. G, Emerging lateral root primordium. H, Ovule primordium after first mitotic division of megaspore mother cell stages, according to Schneitz et al. (1995). I to K, Embryo sac, stages according to Christensen et al. (1997). Arrows denote faint GFP signal in nuclei of embryo sac. L to P, Embryos. L, Globular stage. M, Transition stage, arrow denotes down-regulation of GFP in basal cell lineage. N, Heart stage, arrow denotes transient signal in the prospective shoot meristem. O, Late heart stage. P, Late heart stage, overnight incubation with 10 μM transzeatin. The signal from the membrane stain FM4-64 is shown in magenta. tz, Transzeatin; ep, epidermis; c, cortex; en, endodermis; p, pericycle cells; cc, central cylinder. Bars = 20 μm .

GUS reporter system

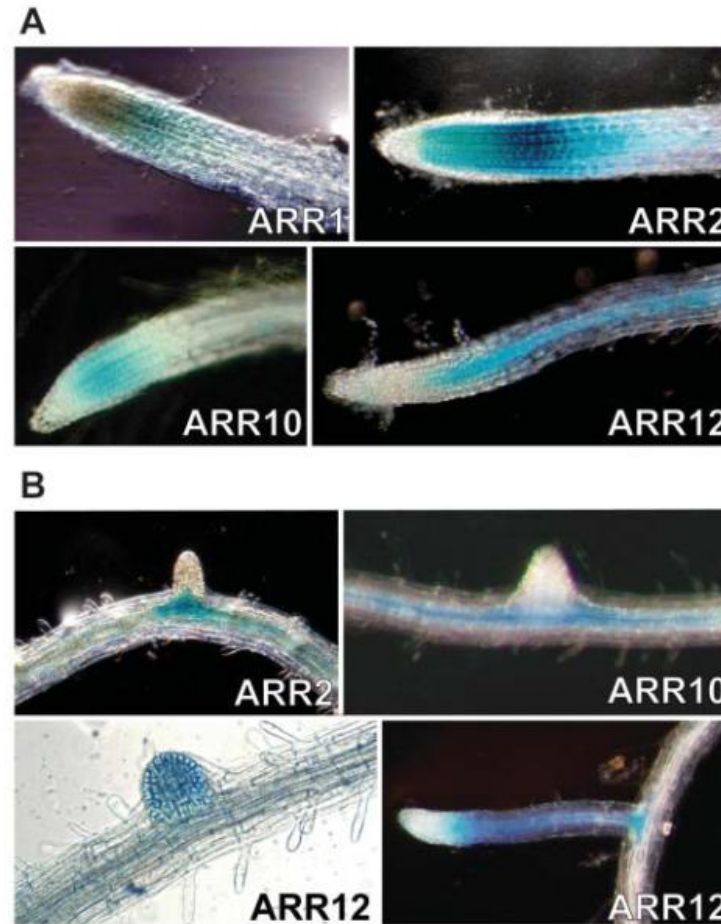


Figure 5. Histochemical localization of type-B ARRs in roots based on GUS reporter gene expression. **A**, Expression of ARR1, ARR2, ARR10, and ARR12 GUS-fusions in the root tip region. **B**, Expression of ARR2, ARR10, and ARR12 GUS-fusions at lateral root junctions.

(Mason et al, 2004)

Thank you for your attention 😊



Central European Institute of Technology
c/o Masaryk University
Žerotínovo nám. 9
601 77 Brno, Czech Republic

www.ceitec.eu | info@ceitec.cz



EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE



**OP Research and
Development for Innovation**

